The Dissertation Committee for Akhil Jalan
certifies that this is the approved version of the following dissertation:

Heterogeneous and Dynamic Network Modeling and Statistical Inference with Provable Guarantees

Committee:

Purnamrita Sarkar, Supervisor

Deepayan Chakrabarti (Co-Supervisor)

Shuchi Chawla

Adam Klivans

Arya Mazumdar

Heterogeneous and Dynamic Network Modeling and Statistical Inference with Provable Guarantees

by

Akhil Jalan

Dissertation

Presented to the Faculty of the Graduate School of

The University of Texas at Austin

in Partial Fulfillment

of the Requirements

for the Degree of

Doctor of Philosophy

The University of Texas at Austin

May 2025

Dedication

To my parents.

Epigraph

O God,
thy sea is so great,
and my boat is so small.
—Fisherman's Prayer

Acknowledgments

I am deeply thankful to all of the people who have supported me throughout my doctoral work. I cannot hope to adequately convey my thanks to them in this short acknowledgments file, but I will make an attempt.

First, I sincerely thank my advisors, Dr. Purnamrita Sarkar and and Dr. Deepayan Chakrabarti. Working with them has been an amazing privilege and joy, and I have learned an incredible amount from both of them in what feels like a few short years. One thing that I cannot emphasize enough is how kind and supportive both are as people; through all of my many mistakes, blunders, and missteps, both of them have been unfailingly supportive and positive, and this has made all the difference. Their support, kindness, and generosity is really beyond what I can put into words.

Next, I thank all of my collaborators. These include Drs. Yassir Jedra, Arya Mazumdar, Soumendu Sundar Mukherjee, and (soon to be) Dr. Marios Papachristou, who each contributed to one or more of the works appearing in this thesis. Learning from one's colleagues is one of the great joys of science, and in this respect I could not have been more fortunate; a huge fraction of what I know is thanks to these people, and I could not done any of this work without them. I also wish to thank collaborators that I had earlier in my PhD, when I was focusing on computational complexity theory: Drs. Dana Moshkovitz, David Zuckerman, Zeyu Guo, and Ben Lee Volk. I consider myself extremeley lucky to have had the chance to work with each of them, and they undoubtedly influenced my development as a researcher in a major way.

Next, I sincerely thank Drs. Shuchi Chawla and Adam Klivans for their guidance and feedback as members of my PhD committee. They have been very generous with their feedback and time, and their perspectives have improved this thesis tremendously. Dr. Klivans has also supported me financially, and I am very grateful for the opportunities this provided me.

Besides academic work I have been fortunate to collaborate with industry researchers in biotechnology during my time at UT. These hands-on experiences not only helped me learn more about this important area, but directly inspired the works on transfer learning that

appear in this thesis. I want to thank the team members at Ark Biotech and Van Heron Labs, including Drs. Kai Hoeffner, Zheng Huang, Rebecca Vaught, as well as Yossi Quint, Dustin Sands, Damien Waits, Nickellaus Roberts, and all the other people at these companies.

I also would like to thank Drs. Nikhil Srivastava, Gireeja Ranade, and Brett Kolesnik for their early mentorship and encouraging me to pursue graduate school during my undergraduate years.

My friends have been a huge source of support and encouragement during my time at UT. The happiness and levity that they brought into my life was a constant reminder that, crazy as it may seem, there's more to life than research. To all of my friends, whom I will not bother naming because I hope they never bother to read this — thank you, from the bottom of my heart.

Finally, I want to thank my brothers and parents for their unconditional love and support throughout my life. Being an academic at all is a privilege reserved for the very luckiest people on the planet, and I am forever grateful to my parents for sacrificing so much to give me the opportunities that I have had. This thesis is dedicated to them.

Abstract

Heterogeneous and Dynamic Network Modeling and Statistical Inference with Provable Guarantees

Akhil Jalan, PhD
The University of Texas at Austin, 2025

SUPERVISOR: Purnamrita Sarkar

Networks, or graphs, are fundamental objects used to model a huge range of phenomena such as social interactions, biological processes, and the global economy. Due to this broad applicability, both the development of network models and statistical inference problems related to networks are major areas of research. In this thesis, we propose new models of networks both capture real-world phenomena and enable learning algorithms with provable guarantees.

First, we introduce new models of distributional shifts for network models, and give *transfer learning* methods to estimate a target network with limited and noisy data. For latent variable networks (Chapter 2), which generalize common network models such as Stochastic Block Models and Graphons, we give a transfer learning algorithm for combinatorial distributional shifts. In this setup, we observe an $o(1)$ fraction of the target data for a graph Q, as well as side information in the form of a source graph P. We give an efficient algorithm to estimate Q that achieves vanishing error with high probability. Moreover, we give minimax lower bounds for the special case of Stochastic Block Models, and give an efficient algorithm to achieve the minimax rate in this setting. Furthermore, we validate our results on real-world transfer learning problems in cell biology and dynamic social networks.

Next, we study transfer learning for *matrix completion*, which generalizes the problem of network estimation with missing data (Chapter 3). We consider low-rank source matrix P and target matrix Q which are related via a linear shift in their row and column singular subspaces, which is a commonly studied geometric model of distributional shift. The target

matrix Q is noisily observed in a Missing Not-at-Random (MNAR) setting that is motivated by biological problems; entire rows and columns missing, making estimation impossible without side information. Unlike our work on latent variable models in Chapter 2, we consider both the *active* and *passive* sampling of rows and columns. We establish minimax lower bounds for entrywise estimation error in each setting. Further, we give a computationally efficient estimation framework to achieve the lower bound for the active setting, which leverages the source data to query the most informative rows and columns of Q. This avoids the need for *incoherence* assumptions required for rate optimality in the passive sampling setting. We demonstrate the effectiveness of our approach through comparisons with existing algorithms on real-world biological datasets.

Second, we study various network models for heterogeneous and dynamic real-world settings in economics and sociology. We first propose a network model of bilateral contracts between heterogeneous, mean-variance optimizing agents (Chapter 4). Our model applies to several important classes of economic networks, such as the multi-trillion dollar market of derivatives contracts between large financial institutions. We give an efficient algorithm for honest agents to find a stable network from iterative pairwise negotiations, and prove that it converges to a strong (coalitional) Nash equilibrium. This algorithm is decentralized, and only requires that agents communicate with their neighbors in order to myopically update their preferred contract sizes based on their own utility functions. Moreover, we give a learning algorithm that recovers network parameters from time-series data using Semidefinite Programming. Further, we empirically demonstrate how an external observer can learn the source of a network shock based on observing the equilibrium before and after the shock. We verify our findings with experiments on real-world international trade networks, and networks constructed from real-world portfolio data.

Next, we study a model of strategic negotiations in which agents can manipulate the pairwise negotiation algorithm of Chapter 4 by misrepresenting their true preferences (Chapter 5). By negotiating strategically, agents can obtain better contracts. Unlike prior works on strategic behavior in network games, which consider honest behavior or a single strategic agent, we allow any subset of agents to be strategic. We provide an efficient algorithm for finding the set of Nash equilibria of the game played by the strategic agents, if any exist, and certify their nonexistence otherwise. We also show that when several strategic agents

are present, their utilities can increase or decrease compared to when they are all honest. Small changes in the inter-agent correlations can cause such shifts. Finally, we develop an algorithm by which new agents can learn the information needed for strategic behavior. Our algorithm works even when the (unknown) strategic agents deviate from the Nash-optimal strategies. We verify these results on both simulated networks and a real-world dataset on international trade.

Finally, we introduce a model of opinion formation in social networks where strategic agents can manipulate publicly expressed opinions to further their own narratives (Chapter 6). This captures real-world manipulation of social networks, such as during the 2016 US elections and the 2019 Hong Kong protests. As in Chapter 5, we go beyond prior works by considering multiple strategic actors, who can have conflicting goals. Unlike Chapter 5, our focus is not on the formation of a network of contracts, but rather on the equilibrium opinions expressed in an exogenous social network, such as Twitter. We characterize the Nash Equilibrium of the resulting meta-game played by the strategic actors. Experiments on real-world social network datasets from Twitter, Reddit, and Political Blogs show that strategic agents can significantly increase polarization and disagreement, as well as increase the "cost" of the equilibrium. To this end, we give worst-case upper bounds on the Price of Misreporting (analogous to the Price of Anarchy). Finally, we give efficient learning algorithms for the platform to (i) detect whether strategic manipulation has occurred, and (ii) learn who the strategic actors are. Our algorithms are accurate on the same real-world datasets, suggesting how platforms can take steps to mitigate the effects of strategic behavior.

Contents

Chapter 1: Introduction

1.1 A Brief Introduction to Network Models

Networks, or graphs[1], are fundamental objects in computer science, mathematics, and statistics. Networks represent pairwise relationships within a set of entities. Given their flexibility and simplicity, networks are ubiquitous in statistical and mathematical modeling (Newman, 2018). Some of their applications include protein-protein interactions (Fan et al., 2019), opinion formation in online communities (De et al., 2016), wireless networks (Page et al., 1999), and economic supply chains (Elliott et al., 2022a).

Formally, a network is a tuple $G = (V, E)$, where V is a finite set of vertices and $E \subseteq V \times V$ is a set of edges. We denote the number of vertices as $n = |V|$ and the number of edges as $m = |E|$. In many settings, one has additional information regarding vertices, edges, or both. To incorporate such information, researchers have proposed several extensions to the basic graph model. Some prominent examples are graphons, which study a "graph limit" as $|V| \to \infty$ (Gao et al., 2015), dynamic graphs that evolve their edge structure over time (Trivedi et al., 2019), and network games that model interactions of n agents with their neighbors (Tardos, 2004).

Our work involves each of the settings listed above. Broadly speaking, we consider network models that go beyond the basic graph model in two ways. First, we consider *heterogeneity* among the nodes, meaning that different nodes in V have different feature information which determines how they form edges. These features may or may not be known. Second, we investigate *dynamic* networks that change due to new external information or more general kinds of distributional shifts.

Within the setting of heterogeneous and dynamic network modeling, a major focus of our work is on statistical inference with provable guarantees. Networks are widely studied in statistics, usually within the *random graph* model. In this setup, there is a *population* network represented as an adjacency matrix $G \in \mathbb{R}^{n \times n}$, and one wishes to estimate this given some data A. The Erdos-Renyi model (Erdos et al., 1960) considers the setting where

[1]Throughout this thesis we will use the terms network and graph interchangeably. Statisticians seem to prefer "network" whereas computer scientists say "graph."

$G_{ij} = p$ for some fixed $p \in [0, 1]$ and all distinct $i, j \in [n]$. The data are then a random matrix $A \in \{0, 1\}^{n \times n}$, where $A_{ij} \overset{\text{iid}}{\sim} \text{Bernoulli}(p)$. The problem of network estimation then reduces to estimating the single parameter p. Several generalizations of the Erdos-Renyi model are studied, such as latent distance models, stochastic block models, random dot product graphs and mixed membership block models (Hoff et al., 2002b; Hoff, 2007; Handcock et al., 2007; Holland et al., 1983; Rubin-Delanchy et al., 2022; Airoldi et al., 2008). The idea for each of these models is that all vertices have some feature information $\boldsymbol{x}_i \in \mathbb{R}^d$, and then the existence of a (possibly weighted) edge $\{i, j\}$ in the population graph G is a function of $\boldsymbol{x}_i$ and $\boldsymbol{x}_j$. For example in the Stochastic Block Model (Holland et al., 1983), there is a $k \geq 2$ and the feature vectors $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n \in \{0, 1\}^k$ have only a single nonzero entry each. The nonzero entry of $\boldsymbol{x}_i$ indicates the membership of node i in one of k distinct communities, and there is a connectivity matrix $B \in [0, 1]^{k \times k}$ where B_{st} indicates the probability of an edge between communities s, t. Notice that the Erdos-Renyi is an SBM with $k = 1$.

In these random graph models, recovering the population graph G depends on the properties of the function describing edge probabilities, as well as the noise model for observations. For example, one might (noisily) observe only a small fraction of all edges, rather than the $n \times n$ observation matrix A in the usual Erdos-Renyi model. Moreover, the noise may be additive rather than Bernoulli, meaning we observe $G_{ij} + \eta_{ij}$ for random η_{ij}.

More recently, these distinct models have been studied under the common framework of graph limits or graphons (Lovász, 2012; Bickel and Chen, 2009), which provide a natural representation of vertex exchangeable graphs (Aldous, 1981; Hoover, 1979). The feature data $\boldsymbol{x}_i$ for vertices i can be considered as unseen latent variables, so graphon models are also called latent variable models. These network models have found applications in real-world settings such as neuroscience Ren et al. (2023), ecology Trifonova et al. (2015), international relations Cao and Ward (2014), political pscyhology Barberá et al. (2015), and education research Sweet et al. (2013). We defer a more detailed discussion to Chapter 2.

Across these settings in network modeling and statistical inference, we focus on two main questions, which we discuss next.

1.2 Our Main Questions

The core questions of this work are the following.

(Q1) How can we appropriately model real-world phenomena with networks?

We are of course motivated by real-world applications. At the same time, models should be parsimonious and analytically tractable, and these desiderata may conflict with applicability. As we will argue, network models should balance these considerations, with careful attention to the specific application at hand.

Next, given the ubiquitous nature of statistical techniques in modern mathematical modeling, we emphasize the importance of learning such models from data. This leads to our second main question.

(Q2) How can we provably learn network models from data in a computationally efficient way?

Our **(Q2)** is more technically well-defined than **(Q1)**. By comptuational efficiency we will use the standard notion of polynomial-time algorithms (the complexity class P) (Arora and Barak, 2009). By "provable" we mean theoretical upper bounds on e.g. estimation error, ideally with matching minimax lower bounds (Tsybakov, 2009). However, theoretical bounds are not enough. Because of the need to build useful models for practitioners, we will seek to compare the guarantees of our theory with experimental results on both simulated and real-world data.

The advantage of simulated data is that one can carefully engineer conditions so that theoretical assumptions are met or violated, as in an ablation study. For real-world data, theoretical assumptions are rarely met perfectly; therefore, real-world experimental results have the additional advantage of illustrating the extent to which theoretical assumptions fail to match reality, as well as how much this affects the modeling outcomes.

The answer to our **(Q1)** depends both on what real-world phenomena are under consideration, as well as what we mean by "appropriate." The question of appropriateness of models touches on fundamental issues of mathematical and statistical modeling, which we briefly discuss here before moving to our contributions in Section 1.4.

What makes a model appropriate? The appropriateness of a model depends on the phenomenon being studied. In this thesis, we study *network* models that are both mathematical and statistical in nature. By *mathematical* modeling (also called *mechanistic* modeling), we simply mean the use of mathematical constructs (such as functions, variables, sets) to make statements about a real-world phenomenon (such as proteins, businesses, or social media). For example, Maxwell's laws are a mathematical model used to describe the activity of magnets (Fitzpatrick, 2008). By *statistical* modeling we refer to the class of mathematical models that incorporate randomness in some way; for example, the use of Stochastic Differential Equations to model stock prices (Kou, 2007).

We will demonstrate applications of our network models to areas including biology (Chapters 2 and 3), economics (Chapters 4 and 5), and sociology (Chapter 6). Each of these fields has its own methodological considerations. At the same time, the use of mathematical and statistical models, and network models in particular, raises general questions that apply to all fields.

A major conceptual issue is the role of theory versus empirics (Fang and Casadevall, 2023). Theorems are abstract statements; what can they tell us about the real world? The straightforward answer is that theorems describe *models*, and to the extent that the models match reality, the theorems do as well. Of course, common wisdom tells us that "all models are wrong, but some are useful" (Box, 1976). We certainly agree, but in *what sense* are the models wrong? How can a wrong model be useful; put differently, how wrong can the model be before it stops being useful? Moreover, if a model is wrong, how can we know this, except by referring to a more accurate model? These questions have been debated in applied mathematics and physics for centuries, dating at least as far back as Newton (Newton, 1687). Going back even further, Plato discussed the question of how mathematics might describe physical reality in the *Timaeus* (circa 360 B.C.E.) (Gill, 1987).

For network models specifically, a particularly relevant issue is that of *reductionism* (Thurner et al., 2018). Any mathematical model can be viewed as reductionist, since it seeks to reduce a system's behavior to a set of assumptions (axioms). For example, microeconomics models might assume utility-maximizing behavior of agents (Browning and Zupan, 2020), and physics models might assume that a dynamical system is linear (Thirring, 2013). But problems of a "complexity science" or "emergence" flavor seem to elude reductionist

models (Thurner et al., 2018), and this motivates the use of network models to capture emergent phenomena in fields such as biology (Kauffman, 2019), economics (Arthur, 2021), sociology (Castellani and Hafferty, 2009).

Statistical models also raise important questions regarding interpretation and applicability. It is said that a "random variable is neither random nor variable" (Hedderich and Sachs, 2024). This aphorism is typically used to motivate the definition of random variables as measurable functions, but a measure-theoretic perspective does not clarify matters to the practitioner either. The practitioner wishes to know: is reality stochastic or not? This question hinges not only on probability theory but also an exact theory of physics, which we currently lack. The question of what randomness and probability mean has been debated since the founding of probability theory more than 3 centuries ago (Diaconis and Skyrms, 2018). This informs not only the interpretation of statistical practice but also frequentism versus Bayesianism in statistics, which is widely debated in its own right (Bland and Altman, 1998; Samaniego, 2010). Resolving these questions is well beyond the scope of our work, but should be kept in mind when we consider how theoretical statements with probabilities should be used for practical problems.

Finally, how should we think of statistical modeling in the era of deep learning? Our work on statistical modeling with networks largely involves "traditional" statistical models, rather then deep learning methods such as Graph Neural Networks (Wu et al., 2020; Zhu et al., 2021). We note that some of our models can accept inputs from deep learning models as part of a larger multi-scale framework (Weinan, 2011). For example, the source information for the transfer learning algorithm of Chapter 3 might be an estimate from a deep learning method. Similarly, the mean-variance beliefs of economic agents in Chapters 4, and 5 might be the outcome of a deep learning model.

Still, we largely focus on more traditional methods. The main advantage of these approaches is that one can provide theoretical guarantees. Furthermore, these traditional algorithms are more interpretable than deep learning systems, which is critical for fields such as healthcare (Vellido, 2020). However, these models might have more restrictive assumptions, and might be less effective in general. How should practitioners navigate these chioces?

Deep learning advocates might advance the "theory free ideal," which combines massive datasets and massive neural networks to simply examine the results (Andrews, 2023). While

this is tempting, it is known that deep learning has fundamental limits and cannot learn functions of certain types (Abbe and Sandon, 2018). In fact, there are information-theoretic limits to learning closed-form models from data for *any method*, based on thermodynamic arguments (Fajardo-Fontiveros et al., 2023).

Moreover, in data-bottlenecked domains such as biology and chemistry, one cannot build a "perpetual motion machine" that generates ever-more synthetic data to train ever-larger models (Listgarten, 2024). While some have proposed using synthetic data from generative models to replace real data, recent work on representation learning suggests this may be of limited usefulnes, as "[g]enerative models can be viewed as a compressed and organized copy of a dataset" (Jahanian et al., 2022).

We emphasize that the question of what makes an appropriate model (related to our **(Q1)**) is not merely an academic concern. The choice of conceptual framework of modeling is critical to any real-world scientific or engineering context (Fang and Casadevall, 2023). Model choices drive real-world decisions, and in real-world controls systems the decision-maker may be an algorithm rather than a human (de Cañete et al., 2018). Therefore, understanding the implications of model selection and the broader context of usage is critical for researchers, and this informs our **(Q1)**.

1.3 Notation

For an integer $r \geq 1$, we use $[r]$ to denote the set of integers $[r] := \{1, 2, \ldots, r\}$.

We use lowercase letters, with or without subscripts, to denote scalars (e.g., c, γ_i). Let $a \vee b := \max\{a, b\}$ and $a \wedge b := \min\{a, b\}$. Lowercase bold letters denote vectors ($\boldsymbol{\mu}_i, \boldsymbol{w}$), and uppercase letters denote matrices (W, P, Σ_i).

For vector $\boldsymbol{\mu}$ and matrix Σ we use $\boldsymbol{\mu}_i$ to denote the i^{th} entry of $\boldsymbol{\mu}$ and Σ_{jk} to denote the (j, k) cell of Σ. If $\boldsymbol{\mu}_i$ refers instead to the i^{th} member of a collection of vectors $\boldsymbol{\mu}_1, \ldots, \boldsymbol{\mu}_k$ then we use $\boldsymbol{\mu}_{i;j}$ to refer to the j^{th} component of the vector $\boldsymbol{\mu}_i$, and similarly $\Sigma_{i;jk}$ for the (j, k) cell of matrix Σ_i. The absence of a semicolon in the subscript indicates the former case always.

For multisets S, T and $A \in \mathbb{R}^{m \times n}$, let $A[S, T] \in \mathbb{R}^{|S| \times |T|}$ be the submatrix with row and column indices in S, T respectively, possibly with repeated entries from A if S, T have

repeated elements.

We use $\boldsymbol{v}^T$ to denote the transpose of a vector $\boldsymbol{v}$, and $\|\cdot\|_p$ to denote the ℓ_p norm of a vector or matrix. For vectors $\boldsymbol{x}, \boldsymbol{y}$ we use $\langle \boldsymbol{x}, \boldsymbol{y} \rangle$ to denote the standard inner product (the dot product). Let $\otimes$ denote the tensor (Kronecker) product: for $A \in \mathbb{R}^{m \times n}, B \in \mathbb{R}^{s \times t}$, we have $(A \otimes B) \in \mathbb{R}^{ms \times nt}$ with $(A \otimes B)_{i(r-1)+v, j(s-1)+w} = A_{ij} B_{vw}$. Furthermore, for a matrix A we denote the Frobenius norm as $\|A\|_F$, max norm as $\|A\|_{\max} := \max_{i,j} |A_{ij}|$, and $2 \to \infty$ norm as $\|A\|_{2 \to \infty} := \max_i \|A^T \boldsymbol{e}_i\|_2 = \sup_{\|\boldsymbol{x}\|_2 = 1} \|A\boldsymbol{x}\|_\infty$.

We say the matrix $A \succeq 0$ if A is positive semidefinite, $A \succ 0$ if it is positive definite, and $A \succeq B$ if $A - B \succeq 0$. The vectors $\boldsymbol{e}_1, \ldots, \boldsymbol{e}_n$ denote the standard basis in $\mathbb{R}^n$, and I_n is the $n \times n$ identity matrix. For integer n, d such that $d \leq n$, the *Stiefel manifold* $\mathcal{O}^{n \times d}$ Hatcher (2002) consists of all $U \in \mathbb{R}^{n \times d}$ such that $U^T U = I_d$.

For an appropriate matrix M, $\mathrm{tr}(M)$ calculates its trace, $\mathrm{vec}(M)$ vectorizes M by stacking its columns into a single vector, and $\mathrm{uvec}(M)$ vectorizes the upper-triangular off-diagonal entries of M.

For functions $f, g : \mathbb{N} \to \mathbb{R}$ we let $f \lesssim g$ denote $f = O(g)$ and $f \gtrsim g$ denote $f = \Omega(g)$.

1.4 Overview of Our Contributions

This thesis is organized into separate chapters, and can be read in roughly two parts. First, we study transfer learning for latent variable network models (Chapter 2) and for matrix completion (Chapter 3). These chapters mainly concern statistical modeling (our **(Q2)**), and specifically on statistical inference with provable guarantees in the face of heterogeneous data. The heterogeneity can come from a distribution shift, or from dynamics (changes in time). The applications are mainly focused on biological problems, although we also give applications to dynamic social networks in Chapter 2. While the theoretical results in these chapters are general, we will also discuss the appropriateness of our modeling frameworks for each application as well (our **(Q1)**).

Second, we study specific network models for financial networks (Chapter 4), strategic negotiations during network formation games (Chapter 5), and manipulation of social networks (Chapter 6). These chapters concern both mathematical and statistical modeling (our **(Q1)** and **(Q2)**). The specificity of the application domains will allow us to examine the

appropriateness of our network models in depth (our **(Q1)**).

We now give an overview of each chapter. For each work, the dissertator was the first author, and was responsible for discussing, deriving and writing up the detailed theoretical analysis, conducting and discussing the experiments, and writing and revising the paper.

1.4.1 Transfer Learning for Latent Variable Network Models[2]

In machine learning and statistics, *transfer learning* is a paradigm in which data from a source distribution P is exploited to improve estimation of a target distribution Q for which a small amount of data is available. Transfer learning is well-studied in learning theory, starting with works such as Ben-David et al. (2006); Cortes et al. (2008); Crammer et al. (2008), and at the same time has found applications in areas such as computer vision Tzeng et al. (2017b) and speech recognition Huang et al. (2013). A fairly large body of work in transfer learning considers different types of relations that may exist between P and Q, for example, Mansour et al. (2009); Hanneke and Kpotufe (2019, 2022), with emphasis on model selection, multitask learning and domain adaptation. On the other hand, optimal nonparametric rates for transfer learning have very recently been studied, both for regression and classification problems Cai and Wei (2021a); Cai and Pu (2024).

In Chapter 2, we study transfer learning in the context of *random network/graph models*. In our setting, we observe Bernoulli samples from the full $n \times n$ edge probability matrix for the source P and only a $n_Q \times n_Q$ submatrix of Q for $n_Q \ll n$. We would like to estimate the full $n \times n$ probability matrix Q, using the full source data and limited target data, i.e., we are interested in the task of estimating Q in the partially observed target network, utilizing information from the fully observed source network. This is a natural extension of the transfer learning problem in classification/regression to a network context. Moreover, the particular formulation of observing only a vertex-induced subgraph of Q is motivated by practical settings, such as biological network estimation (Figure 1.1). However, it is to be noted that network transfer is a genuinely different problem owing to the presence of edge correlations.

To study transfer learning on networks, one needs to fix a general enough class of

[2]This work appeared in Advances in Neural Information Processing Systems (NeurIPS) 2024 and can be cited as Jalan et al. (2024b).

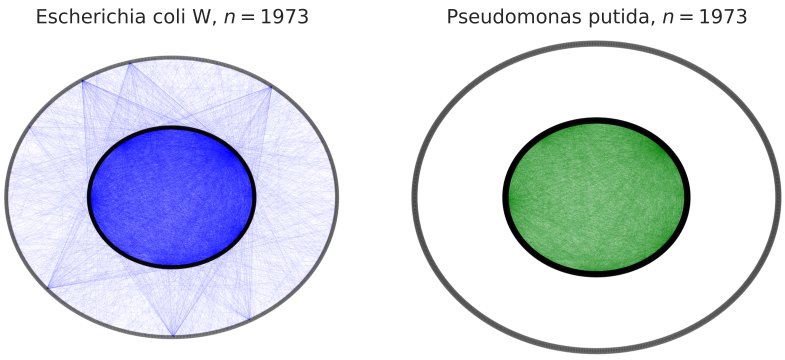


Figure 1.1: In metabolic networks, nodes are metabolites (such as amino acids), and edges are between metabolites that occur in the same reaction. *In vivo* methods for testing edges in metabolic networks require metabolite balancing and labeling experiments, so only the edges whose endpoints are *both* incident to the experimentally chosen metabolites are observed Christensen and Nielsen (2000). Therefore, for a model organism (left) we may see edges incident to all n metabolites, but for non-model organisms (right) we may only see edges incident to the center $n_Q \ll n$ metabolites.

networks that is appropriate for the applications (such as the biological networks mentioned above) and also suitable to capture the transfer phenomenon. We study latent variable models, which generalize many classes of networks in the statistics literature such as latent distance models, stochastic block models, random dot product graphs, and graphons Hoff et al. (2002b); Hoff (2007); Handcock et al. (2007); Holland et al. (1983); Rubin-Delanchy et al. (2022); Airoldi et al. (2008); Lovász (2012). For unseen latent variables $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n \in \mathcal{X} \subset \mathbb{R}^d$ and unknown function $f_Q : \mathcal{X} \times \mathcal{X} \to [0, 1]$ where $\mathcal{X}$ is a compact set and d an arbitrary fixed dimension, the edge probabilities are $Q_{ij} = f_Q(\boldsymbol{x}_i, \boldsymbol{x}_j)$.

The transfer learning problem is to to estimate the population matrix Q, given observations $A_{Q;ij} \sim \text{Bernoulli}(Q_{ij})$, as well as observations $A_{P;st} \sim \text{Bernoulli}(P_{st})$ for a source matrix P. The source data A_Q are missing in all but an $S \times S$ submatrix, for a small $S \subset [n]$ that is sampled from all n_Q-sized subsets uniformly at random. The goal is to output a $\hat{Q} \in \mathbb{R}^{n \times n}$ which minimizes the squared error $\|\hat{Q} - Q\|_F^2$.

To enable transfer learning, one must assume some relationship between the source and target data. One natural assumption is to consider pairs (f_P, f_Q) such that for all $\boldsymbol{x}, \boldsymbol{y} \in \mathcal{X}$, the difference $(f_P(\boldsymbol{x}, \boldsymbol{y}) - f_Q(\boldsymbol{x}, \boldsymbol{y}))$ is small. For example, Cai and Pu (2024) study transfer learning for nonparametric regression when $f_P - f_Q$ is close to a polynomial in $\boldsymbol{x}, \boldsymbol{y}$. But,

requiring $f_P - f_Q$ to be pointwise small does not capture a broad class of pairs in the network setting. For example, if $f_P = \alpha f_Q$. Then $f_P - f_Q = (\alpha - 1) f_Q$ can be far from all polynomials if f_Q is, e.g. a Hölder -smooth graphon.[3] However, under the network model, this means A_P and A_Q are stochastically identical modulo one being α times denser than the other.

We will therefore consider pairs (f_P, f_Q) that are close in some measure of local graph structure. In particular, we use a graph distance measure introduced in Mao et al. (2021) for a different inference problem. Informally, in graph P the distance between nodes i and j measures the similarity of the 2-hop neighborhoods of i, j.

We will require that f_P, f_Q satisfy a local similarity condition on the relative rankings of nodes with respect to this graph distance. Since we only estimate the probability matrix of Q, the condition is on the latent variables $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n$ of interest. The hope is that the proximity in graph distance reflects the proximity in latent positions.

With this relationship between the source and target, the main contributions of Chapter 2 are as follows.

Algorithm for Latent Variable Models. We provide an efficient algorithm for latent variable models with Hölder-smooth f_P, f_Q. The benefit of this algorithm is that it does not assume a parametric form of f_P and f_Q. We prove a guarantee on its error in terms of the dimension d, the dataset sizes n_Q, n, and the smoothness levels of f_P, f_Q.

Minimax Rates. We prove a minimax lower bound for Stochastic Block Models (SBMs). Moreover, we provide a simple Algorithm that attains the minimax rate for this class.

Experimental Results on Real-World Data. We test both of our algorithms on real-world metabolic networks and dynamic email networks, as well as synthetic data from well-studied classes of networks such as latent distance models and mixed-membership stochastic block models.

[3] In fact, Cai and Pu (2024) highlight this exact setting as a direction for future work.

1.4.2 Optimal Transfer Learning for Missing Not-at-Random Matrix Completion[4]

In Chapter 3, we study a natural generalization of the network estimation problem via matrix completion. A weighted graph can be viewed as an adjaceny matrix $Q \in \mathbb{R}^{n \times n}$. Therefore in the transfer learning setting of Chapter 2, we can view estimation of the target matrix Q given only a subset of the edges as a special case of matrix completion, when the population matrix is a latent variable network model. Matrix completion, in addition to generalizing network estimation, is a fundamental problem in its own right, and is well-motivated by theory Candès and Recht (2009); Candès and Tao (2010) and practice Fernández-Val et al. (2021); Einav and Cleary (2022); Gao et al. (2022).

Classically, matrix completion is studied in the Missing Completely-at-Random (MCAR) setting Jain et al. (2013); Chatterjee (2015a); Chen et al. (2020b), where each entry of Q is observed i.i.d. with probability p (possibly with additional noise). However, the MCAR assumption may not necessarily hold in practice, which motivates more general missingness patterns called Missing-Not-at-Random (MNAR). Instead of a single parameter p, MNAR works often consider an underlying *propensity matrix* p_{ij} so that each $\tilde{Q}_{ij}$ is observed independently with probabilty p_{ij} Ma and Chen (2019); Bhattacharya and Chatterjee (2022). Various MNAR models have been formulated based on missigness structures in panel data Agarwal et al. (2023b), recommender systems Jedra et al. (2023), and electronic health records Zhou et al. (2023).

Motivated by biological problems, in Chapter 3 we consider a challenging MNAR structure where most rows and columns of $\tilde{Q}$ (a noisy version of Q) are entirely missing. This is similar to the missigness model for network estimation we studied in Chapter 2, but has important differences as well.

In particular, we consider both the *active sampling* and *passive sampling* settings for $\tilde{Q}$, whereas Chapter 2 only considers the passive sampling setting. In active sampling, a practitioner can *choose* rows R and columns C so that entries in $R \times C$ are observed. This follows experimental design constraints in metabolite balancing experiments Christensen and Nielsen (2000), marker selection for single-cell RNA sequencing Vargo and Gilbert

[4]This work is under review at the 42nd International Conference on Machine Learning (ICML 2025), and can be cited as Jalan et al. (2025).

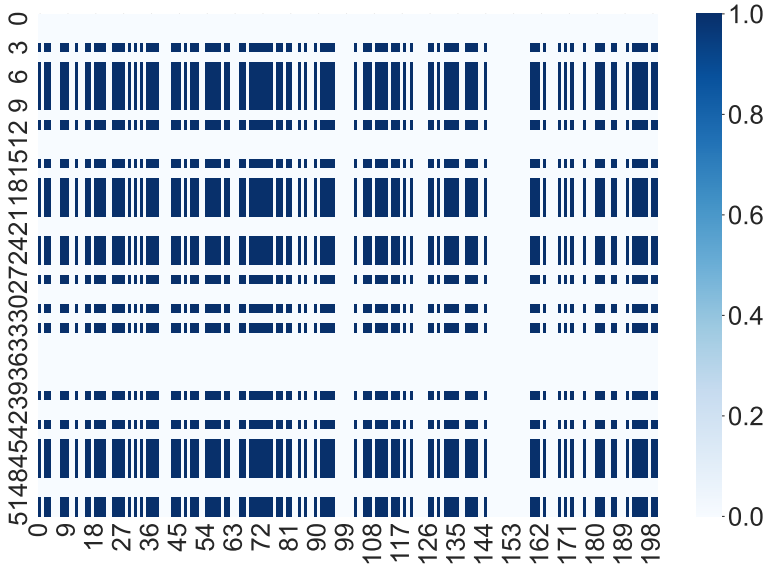


Figure 1.2: The missingness matrix for gene expression levels on Day 2 of a sepsis study Parnell et al. (2013) shows entire rows (patients) and columns (genes) as missing, due to e.g. probe-target hybridization failure of the Illumina HT-12 gene expression microarray Hu et al. (2021). We mark missing entries as 0 (white) and present entries as 1 (blue). This motivates our missingness model in Chapter 3.

(2020), patient selection for companion diagnostics Huber et al. (2022), and gene expression microarrays Hu et al. (2021).

In the *passive sampling* setting, the practitioner cannot choose the experiments. We model this by sampling each row (column) with probability p_{Row} (p_{Col}). This passive sampling setup captures observation patterns in settings such as gene microarrays. For an illustration, see Figure 1.2. Note that in Chapter 2, we instead consider a passive sampling setup in which a set of nodes $S \subset [n]$ is chosen uniformly at random from all subsets of size $|S|$.

In both the active and passive sampling settings, estimation of Q is impossible without additional information. Hence, we consider transfer learning in a setting where one has a noisy and masked $\tilde{P}$ corresponding to a source matrix P. P and Q are related by a distribution shift in their latent singular subspaces (see Chapter 3 for precise definitions), which is a common model in e.g. Genome-Wide Association Studies McGrath et al. (2024) and Electronic Health Records Zhou et al. (2023). This model of distribution shift is a *geometric* one, as opposed to the *combinatorial* distribution shift model of Chapter 2.

The main contributions of Chapter 3 are as follows.

Minimax lower bounds. We obtain *minimax lower bounds* for entrywise estimation error for both the active and passive sampling settings.

Computationally efficient estimation framework and upper bounds. We give a *computationally efficient* estimation framework for both sampling settings. Our procedure is *minimax optimal* for the active setting. We also establish minimax optimality for the passive setting under *incoherence* assumptions.

Real world experiments. We compare the performance of our algorithm with existing algorithms on *real-world datasets* for gene expression microarrays and metabolic modeling.

1.4.3 Dynamic, Incentive-Aware Models of Financial Networks[5]

After the financial crisis of 2008, a major body of work in economics and related fields found that the financial collapse was in part due to the network structure of the financial system Elliott et al. (2014); Glasserman and Young (2015, 2016); Birge (2021); Jackson and Pernoud (2021). The network of interconnections between firms (through e.g. debt obligations) caused problems at one firm spread to others. If one firm defaulted on its debt, its creditors suffered losses. Some creditors would then be forced into default, triggering a default cascade Eisenberg and Noe (2001). Besides the network of debt obligations, an important financial network is the implicit network between firms holding similar assets. Sales by one firm can cause the valuations of neighboring firms to decline. These can snowball into fire sales and rapid, correlated declines in overall market valuations Caballero and Simsek (2013); Cont and Minca (2016); Feinstein (2020); Feinstein and Søjmark (2021).

This motivates the study of financial networks, and in particular the development of models which capture systemic risk. For example, network density, diversification, and inter-firm cross-holdings can affect how robust the networks are to shocks and how such shocks propagate Elliott et al. (2014); Acemoglu et al. (2015); Eisfeldt et al. (2023). The network

[5]This work appeared in Operations Research 2024, and can be cited as Jalan et al. (2024a).

structure also affects the design of regulatory interventions Papachristou and Kleinberg (2022); Amini et al. (2015); Calafiore et al. (2022); Galeotti et al. (2020).

In Chapter 4, we propose a model of financial networks in which agents (firms) optimize mean-variance utility by forming contracts with their neighbors. Unlike previous works, our model does not assume that the network is fixed and observable, but rather models the network formation process and assumes that agents only know their own edges. Moreover, we make no assumptions about the network topology, unlike previous works that assume a ring Caballero and Simsek (2013) or core-periphery Amini et al. (2015) topology. This is important since real-world financial networks exhibit complex structure depending on inter-agent heterogeneity and other factors Peltonen et al. (2014); Glasserman and Young (2016); Eisfeldt et al. (2023).

In our model, there is a set of n agents, and an underlying undirected graph given by some $E \subseteq V \times V$. A pair (i, j) are allowed to form a contract iff $\{i, j\} \in E$. Agent i forms a vector of contracts $\boldsymbol{w}_i \in \mathbb{R}^n$, where $\boldsymbol{w}_{i;j}$ is the size of their contract with j. Notice that $\boldsymbol{w}_{i;j} < 0$ is allowed and corresponds to swapping the roles of two parties in a contract (e.g. the role of lender & borrower). Further, $\boldsymbol{w}_{i;j}$ is nonzero iff $\{i, j\} \in E$.

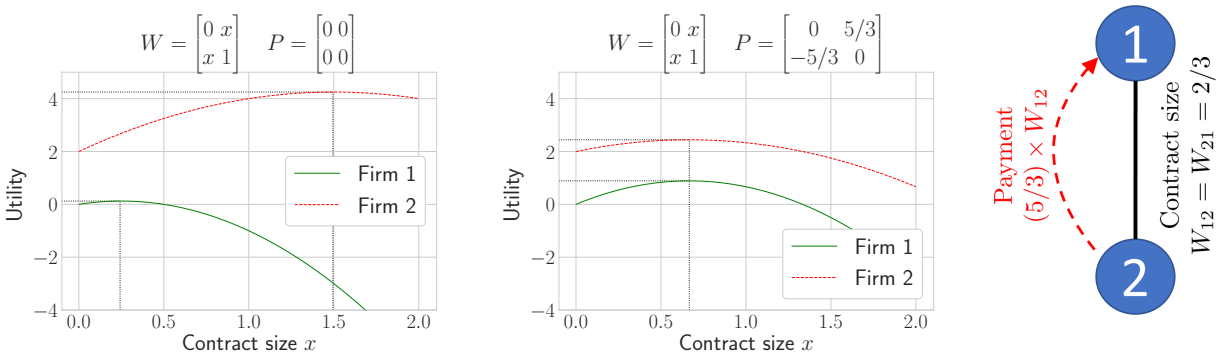
Agent i has private beliefs $(\boldsymbol{\mu}_i, \gamma_i, \Sigma_i) \in \mathbb{R}^n \times \mathbb{R} \times \mathbb{R}^{n \times n}$ regarding the returns of a vector of contracts, where $\boldsymbol{\mu}_{i;j}$ is the expected reward of a unit-sized contract with j, $\Sigma_{i;jk}$ is the covariance of contracts that i might form with j and k, and $\gamma_i > 0$ is a risk-aversion parameter. Then their utility is:

$$\text{agent } i\text{'s utility } g_i(\boldsymbol{w}_i) := \boldsymbol{w}_i^T \boldsymbol{\mu}_i - \gamma_i \cdot \boldsymbol{w}_i^T \Sigma_i \boldsymbol{w}_i, \tag{1.1}$$

We assume $\Sigma_i \succ 0$, so Eq. (1.1) has a unique optimum. However, agents i, j may disagree about the preferred size of a contract between them. If agent j wants a larger contract, they can pay agent i a price $P_{ij} > 0$ per unit contract. The overall network is then a pair (W, P) depending on the private preferences $(\boldsymbol{\mu}_i, \gamma_i, \Sigma_i)_{i \in [n]}$ and the underlying set of allowed edges E. Here $W_{ij} = W_{ji}$ is the size of the contract, and $P_{ij} \cdot W_{ij}$ is the payment that j makes to i. Figure 1.3 gives an illustration for $n = 2$.

With the model defined as above, the main contributions of Chapter 4 are as follows.

Stable Networks and Strong Nash Equilibria. When $n = 2$, it is clear that the two parties can agree to a contract size through payment (Figure 4.1). However, for $n > 2$ it



(a) No payments allowed

(b) Firm 2 pays 5/3 per contract

(c) Network

Figure 1.3: *Example of a stable point for a borrower (Firm 1) and a lender (Firm 2):* (a) When the borrower cannot pay the lender an additional payment, the firms may be unable to agree to a contract, even if trading improves their utilities. (b) By allowing for contract-specific payments, both firms can agree on a contract size. In effect, the borrower (Firm 2) shares its utility with the lender (Firm 1) to achieve agreement. (c) The stable network is shown.

is not clear that all parties can agree to a set of contracts. The reason is that the network structure introduces dependencies between non-neighboring agents. So that Alice's contract with Bob depends not just on their preferences, but also the contracts that Bob forms with his neighbors, and so on.

Without higher-level coordination, is it possible for all agents to form contracts that are optimal for each person? We call such a network of contracts stable, and show that indeed agents can find it with an efficient algorithm that only involves negotiations with one's neighbors.

Moreover, the stable point is *Higher-Order Nash Stable*: for any *cartel* $S \subset [n]$ and *deviation* (W', P'), there exists some cartel member $i \in S$ who is not better off at (W', P'). This is a strengthening of solution concepts such as Nash equilibria and pairwise-stable Nash equilibria Sadler and Golub (2021).

Learning from time-series network data. Learning the parameters of real-world financial networks is an important problem for government regulators Arora et al. (2021). In our model, we consider how an external observer might learn agent preferences from time-series data consisting of stable networks $(W_t)_{t=1,2,\ldots,T}$. Under the assumption that the $(\boldsymbol{\mu}_i)_{i\in[n]}$ evolve according to a Brownian motion, and that agents share the same risk

assessment Σ (e.g. from a credit ratings agency Lopatta et al. (2019)), we give prove that a Semidefinite Programming algorithm can recover the risk asssessment matrix Σ.

Identifying the source of a network shock. Suppose that the change from W_t to W_{t+1} is due to a single changed preference, e.g. an update for $\boldsymbol{\mu}_{i;j}$. If all other preferences are unchanged, we say that the update to $\boldsymbol{\mu}_{i;j}$ is a *shock*. Identifying the sources of shocks is important both for government regulators and agents who might wish to update their own mean/covariance beliefs regarding i, j. However, we show empirically that in realistic settings, the indirect effects (changes not affect i or j) can be as significant as the direct effects. In such cases, a regulator cannot infer the underlying cause of changes in the network.

Outlier detection by agents. A firm i can observe its contracts with neighbors but not the entire network. Suppose another firm j (say, a real-estate firm) has beliefs that are very different from its peers. Then, we prove that under certain conditions, j's contract size with i is also an outlier compared to other real-estate firms. So, firm i can use the network to detect outliers and update its beliefs. But suppose all real-estate firms change their beliefs. This changes all their contract sizes without creating outliers. We show that i cannot determine the cause of this change. For example, firm i would observe the same change whether all real-estate firms had become more risk-seeking or profitable. Since the data cannot distinguish between these two scenarios, firm i remains uncertain about what to do.

1.4.4 Strategic Negotiations in Endogenous Network Formation[6]

The network formation process of Chapter 4 can be viewed as a special case of a network formation game. In statistics and machine learning, as well as economics and computing, there has been increasing interest in network games, which describe n-player games in which each agent plays a game with each of their neighbors on a network (Leng et al., 2020a; Rossi et al., 2022; Wang and Kleinberg, 2024; Park et al., 2024). Besides the pairwise negotations used to form the contract network of Chapter 4, other examples of network games include opinion spreading in a social network De et al. (2016); Gaitonde et al. (2020b); Chen and Rácz (2021b) or firm-level investment choices in a competitive market

[6]This work is under review at the 26th ACM Conference on Economics and Computation (ACM EC 2025), and can be cited as Jalan and Chakrabarti (2024).

Sadler and Golub (2021).

In such network games, an agent may wish to strategically mislead or manipulate their neighbors to obtain a better payoff. In Chapter 5, we consider how agents might mislead their neighbors in the model of Chapter 4 to achieve better utility at the resulting stable point. We call agents who mislead their neighbors *strategic*, and propose a model in which any subset of agents can be strategic.

Prior works on such network games only consider a single strategic node or external actor Gaitonde et al. (2020b); Galeotti et al. (2020); Chen and Rácz (2021b). Interactions between several strategic agents are not explored. Also, in these models, the network is pre-specified (exogenous), whereas we study strategic behavior during the network formation process. Within network formation games, researchers largely assume that all agents report their actual preferences Acemoglu and Azar (2020); Sadler and Golub (2021); Wang and Kleinberg (2024). In our work on contract formation, we also assumed that agents negotiate honestly during pairwise negotations (Chapter 4). Hence, prior works do not consider strategic interactions between agents.

In Chapter 5, we present, to our knowledge, the first results for a multi-agent network formation game with an arbitrary set of strategic agents. Our model can be viewed as a meta-game with respect to the (honest) pairwise negotiations of Chapter 4. Fix a set $S \subset [n]$ of strategic agents. In the meta-game, each strategic agent $i \in S$ chooses some $\boldsymbol{\mu}_i' \in \mathbb{R}^n$ different from their true preference $\boldsymbol{\mu}_i$. We call $\boldsymbol{\mu}_i'$ the *negotiating position*. Non-strategic $j \notin S$ have honest negotiating positions $\boldsymbol{\mu}_j' = \boldsymbol{\mu}_j$.

Once the negotiating positions are chosen, the stable network $(W_{strategic}, P_{strategic})$ is chosen *as if* all agents negotiated honestly (e.g. through the first Algorithm of Chapter 4). The goal of each strategic $i \in S$ in the meta-game is to choose negotiating positions $\boldsymbol{\mu}_i'$ so that their mean-variance utility at this stable point is maximized.

Within this model, the main contributions of Chapter 5 are as follows.

Finding the optimal negotiating position of each agent in the network. At first sight, it is not obvious that an optimal position exists, especially when there are several strategic agents. For instance, consider two hedge funds competing against each other to trade with an investor. It may seem that the two funds would be locked in an arms race,

leading to unbounded negotiating positions. Furthermore, each fund must pick its position before seeing the other fund's choice. This uncertainty makes the problem even more difficult. Nevertheless, we show that if agents know the preferences of all network members, they can find optimal negotiating positions. We prove that there exists an efficient algorithm to find the set of optimal negotiating positions an arbitrary set $S \subseteq [n]$ of strategic agents, or report when no optimal solution exists. Note that "optimal" is with respect to the choices of other agents as well; in particular, the negotiating positions found by our algorithm are Nash equilibria.

Learning algorithm for agents. The algorithm for strategic negotiations requires that each strategic agent knows the true preferences $(\boldsymbol{\mu}_i)_{i\in[n]}$. How can an agent learn these preferences? They may observe the network from previous timesteps, but they cannot directly infer other agents' beliefs since the network was formed from strategic negotiations. Moreover, agents perceive correlations between their contracts and want to minimize their risk. Hence, agent i's contract with j can depend on j's contract with k, which depends on k's contract with ℓ, and so on. So, an edge between (i, j) can depend on the beliefs of all agents (including the strategic ones), not just i and j.

We given algorithm that we present an algorithm to learns the other agents' true beliefs and the set of strategic agents from a single observation of a stable network W. Our algorithm is robust to strategic agents playing non-Nash-optimal strategies to fool the learner. Deviations from Nash equilibria are known to be strategic in certain games against learning agents Assos et al. (2024).

Experiments on simulated and real-world data. We simulate Nash-optimal strategic negotiations on real-world international trade data OECD (2022). Our experiments confirm that the utilities of agents are sensitive to the set of strategic agents. We also show that our learning algorithm recovers the parameters needed for strategic negotiations for a broad range of networks.

1.4.5 Opinion Dynamics with Multiple Adversaries[7]

In the past two decades, social media has grown rapidly. Online social networks, which allow users to share updates about their lives and opinions with a broad audience instantaneously, are now utilized by billions of people globally. These platforms crucially serve as a medium of information exchange, for topics including politics, news, health-related updates, consumer products, and many more (Backstrom et al., 2012; Young, 2006; Banerjee et al., 2013; Shearer and Mitchell, 2021).

In sociology, the *filter-bubble theory* (Pariser, 2011) argues that personalized algorithms used by online platforms, such as search engines and social media,selectively display content that aligns with a user's past behaviors, preferences, and beliefs. Therefore social networks can incude polarization and social discord (Musco et al., 2018b; Chen and Rácz, 2021b; Wang and Kleinberg, 2024; Gaitonde et al., 2020a). This has major real-world consequences, as malicious entities can exploit social networks in order to create discord and cause disagreement. This has already occurred in the 2016 U.S. presidential election (Mueller, 2018), and the 2019 Hong Kong Protests (Twitter, Inc., 2019). These manipulation efforts are sometimes coordinated; however, malicious actors have also sought to target users in conflicting ways, such as when Facebook pages have targeted Americans with sports betting scams and conspiracy theories (Bjork-James and Donovan, 2024).

To model the evolution of opinions in social networks, computer scientists, sociologists, and statisticians have relied on the framework of *opinion dynamics*, where the users' opinions coevolve according to a weighted network $G = (V, E, w)$. Each user updates their opinion as a combination of their own intrinsic opinion as well as the opinions of their neighbors (Friedkin and Johnsen, 1990), capturing the effect of social pressure on opinion formation and expression. This is called the Friedkin-Johnson (FJ) model. So far, all of the existing works on manipulation of opinion dynamics consider a single actor who has the ability to act on the network to induce disagreement or polarization (Musco et al., 2018b; Chen and Rácz, 2021b; Wang and Kleinberg, 2024; Ristache et al., 2024; Gaitonde et al., 2020a; Rácz and Rigobon, 2023; Chitra and Musco, 2020).

[7]This work is under review at the 26th ACM Conference on Economics and Computation (ACM EC 2025), and can be cited as Jalan and Papachristou (2025).

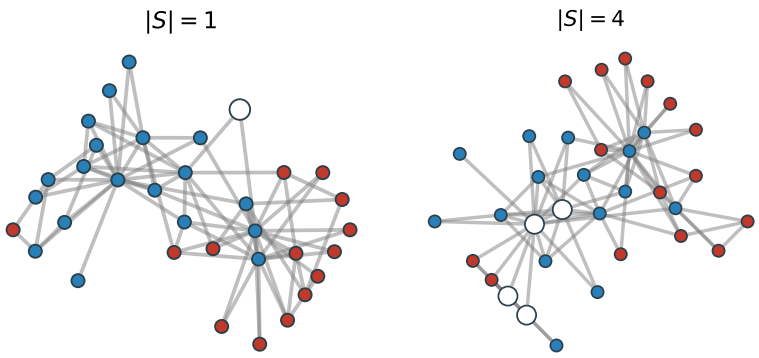


Figure 1.4: Visualization of the strategically manipulated equilibrium on the Karate Club Graph for two different choices of S (the set of strategic agents). White nodes correspond to the nodes in S. For the other nodes, the nodes colored in blue (resp. red) correspond to nodes whose $\mathbb{R}$-valued public opinion increased (resp. decreased), as a result of strategic behavior.

In Chapter 6, we lift the assumption that only a single actor manipulates the network, and consider the case of several decentralized actors. This is motivated by real-world social networks, which involve *multiple* malicious actors, who use different levels of manipulation and hate speech based on their individual goals (Bjork-James and Donovan, 2024). In our setting, we assume that there is a set $S \subseteq V$ of strategic agents whose goal is to report false intrinsic opinions that are different from their true intrinsic opinions.

This strategic misreporting can be viewed as a meta-game played by the members of S, where the base game is the ordinary Friedkin-Johnson opinion dynamics model. The goal of strategic agents in S is to influence others while not deviating much from their neighbors; namely, they want to reach an equilibrium where their neighbors agree with them. Such adversarial behavior can result in significantly different (cf. Figure 1.4) and highly polarized equilibria, where the strategic agents' opinions appear dominant despite not reflecting the actual intrinsic views of the majority.

Our work investigates the conditions under which these strategic manipulations are successful, the extent of their impact on network-wide opinion dynamics, and how platforms can learn from observing these manipulated equilibria to mitigate such impacts. Our main contributions are as follows.

Characterizing Nash Equilibria with Multiple Adversaries. We give the Nash equilibrium of the meta-game defined by strategic misreporting, and show that all Nash-optimal strategies are pure. The Pure Strategy Nash Equilibrium (PSNE) that is given by solving a constrained linear system. Given the PSNE of the game, we characterize the actors who can have the most influence in strategically manipulating the network.

Real-World Experiments to Understand Properites of Equilibria. We apply our framework to real-world social network data from Twitter and Reddit (Chitra and Musco, 2020), and data from the Political Blogs (Polblogs) dataset (Adamic and Glance, 2005). We find that the influence of strategic agents can be rather significant as they can significantly increase polarization and disagreement, as well as increase the overall "cost" of the consensus.

Analysis of Equilibrium Outcomes Under Different Sets of Strategic Actors. Various metrics for network polarization and disagreement are sensitive to the choice of *who* acts strategically, in nontrivial ways. For example, adding more strategic agents can sometimes *decrease* the Disagreement Ratio at equilibrium, due to counterbalancing effects. To address the effects of manipulation, we give worst-case upper bounds on the *Price of Misreporting* (PoM), which is analogous to well-studied Price of Anarchy bounds (see, for example, Bhawalkar et al. (2013); Roughgarden and Schoppmann (2011)), and suggest ways that the platform can be used to mitigate the effect of strategic behavior on their network.

Learning Algorithms for the Platform. We give an efficient algorithm for the platform to detect if manipulation has occurred, based on a hypothesis test with the publicly reported opinions. Next, we give an algorithm to infer *who* manipulated the network (the set of strategic agents S), based on observing the publicly observed equilibrium opinions at a previous timestep. As in Chapter 5, this latter algorithm relies on robust regression to correct for the corruption of the data due to strategic behavior. Our algorithm is practical, requiring node embeddings which are computable even in billion-scale networks such as Twitter (El-Kishky et al., 2022), and runs in polynomial time. Our algorithms have high accuracy on real-world datasets from Twitter, Reddit, and Polblogs.

Chapter 2: Transfer Learning for Latent Variable Network Models

2.1 Introduction

Within machine learning and statistics, the paradigm of *transfer learning* describes a setup where data from a source distribution P is exploited to improve estimation of a target distribution Q for which a small amount of data is available. Transfer learning is quite well-studied in learning theory, starting with works such as Ben-David et al. (2006); Cortes et al. (2008); Crammer et al. (2008), and at the same time has found applications in areas such as computer vision (Tzeng et al., 2017b) and speech recognition (Huang et al., 2013). A fairly large body of work in transfer learning considers different types of relations that may exist between P and Q, for example, Mansour et al. (2009); Hanneke and Kpotufe (2019, 2022), with emphasis on model selection, multitask learning and domain adaptation. On the other hand, optimal nonparametric rates for transfer learning have very recently been studied, both for regression and classification problems (Cai and Wei, 2021a; Cai and Pu, 2024).

In this paper, we study transfer learning in the context of *random network/graph models*. In our setting, we observe Bernoulli samples from the full $n \times n$ edge probability matrix for the source P and only a $n_Q \times n_Q$ submatrix of Q for $n_Q \ll n$. We would like to estimate the full $n \times n$ probability matrix Q, using the full source data and limited target data, i.e., we are interested in the task of estimating Q in the partially observed target network, utilizing information from the fully observed source network. This is a natural extension of the transfer learning problem in classification/regression to a network context. However, it is to be noted that network transfer is a genuinely different problem owing to the presence of edge correlations.

While transfer learning in graphs seems to be a fundamental enough problem to warrant attention by itself, we are also motivated by potential applications. For example, metabolic networks model the chemical interactions related to the release and utilization of energy within an organism (Christensen and Nielsen, 2000). Existing algorithms for

The content of this chapter appeared in Advances in Neural Information Processing Systems (NeurIPS) 2024 and can be cited as Jalan et al. (2024b).

metabolic network estimation (Sen et al., 2018; Baranwal et al., 2020) and biological network estimation more broadly (Fan et al., 2019; Li et al., 2022) typically assume that some edges are observed for every node in the target network. One exception is Kshirsagar et al. (2013), who leverage side information for host-pathogen protein interaction networks. For the case of metabolic networks, determining interactions *in vivo*[1] requires metabolite balancing and labeling experiments, so only the edges whose endpoints are *both* incident to the experimentally chosen metabolites are observed (Christensen and Nielsen, 2000). For a non-model organism, the experimentally tested metabolites may be a small fraction of all metabolites believed to affect metabolism. However, data for a larger set of metabolites might be available for a model organism.

To study transfer learning on networks, one needs to fix a general enough class of networks that is appropriate for the applications (such as the biological networks mentioned above) and also suitable to capture the transfer phenomenon. The latent variable models defined below appear to be a natural candidate for that.

Latent Variable Models. Latent variable network models consist of a large class of models whose edge probabilities are governed by the latent positions of nodes. This includes latent distance models, stochastic block models, random dot product graphs and mixed membership block models (Hoff et al., 2002b; Hoff, 2007; Handcock et al., 2007; Holland et al., 1983; Rubin-Delanchy et al., 2022; Airoldi et al., 2008). They can also be unified under graph limits or graphons (Lovász, 2012; Bickel and Chen, 2009), which provide a natural representation of vertex exchangeable graphs (Aldous, 1981; Hoover, 1979). In addition to their theoretical breadth and usefulness, latent variable models are relevant and applicable to real-world settings such as neuroscience Ren et al. (2023), ecology Trifonova et al. (2015), international relations Cao and Ward (2014), political pscyhology Barberá et al. (2015), and education research Sweet et al. (2013).

For unseen latent variables $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n \in \mathcal{X} \subset \mathbb{R}^d$ and unknown function $f_Q : \mathcal{X} \times \mathcal{X} \to [0, 1]$ where $\mathcal{X}$ is a compact set and d an arbitrary fixed dimension, the edge probabilities are:

$$Q_{ij} = f_Q(\boldsymbol{x}_i, \boldsymbol{x}_j). \tag{2.1}$$

[1] In the organism, as opposed to *in vitro* (in the lab).

Typically, in network estimation, one observes adjacency matrix $\{A_{ij}\}$ distributed as $\{\text{Bernoulli}(Q_{ij})\}$, and either has to learn $\boldsymbol{x}_i$ or directly estimate f_Q. There has been much work in the statistics community on estimating $\boldsymbol{x}_i$ for specific models (usually up to rotation). For stochastic block models, see the excellent survey in Abbe (2017).

Estimating f_Q can be done with some additional assumptions (Chatterjee, 2015b). When f_Q has appropriate smoothness properties, one can estimate it by a histogram approximation (Olhede and Wolfe, 2014; Chan and Airoldi, 2014). This setting has also been compared to nonparametric regression with an unknown design (Gao et al., 2015). Methods for network estimation include Universal Singular Value Thresholding (Chatterjee, 2015b; Xu, 2018), combinatorial optimization (Gao et al., 2015; Klopp et al., 2017), and neighborhood smoothing (Zhang et al., 2017; Mukherjee and Chakrabarti, 2019).

Transfer Learning on Networks. We wish to estimate the target network Q. However, we only observe f_Q on $\binom{n_Q}{2}$ pairs of nodes, for a uniformly random subset of variables $S \subset \{1, 2, \ldots, n\}$. We assume S is vanishingly small, so $n_Q := |S| = o(n)$.

Absent additional information, we cannot hope to achieve $o(1)$ mean-squared error. To see this, suppose f_Q is a stochastic block model with 2 communities of equal size. For a node $i \notin S$, no edges incident to i are observed, so its community cannot be learned. Since $n_Q \ll n$, we will attain $\Omega(1)$ error overall. To attain error $o(1)$, we hope to leverage transfer learning from a source P if available. In fact, we give an efficient algorithm to achieve $o(1)$ error, formally stated in Section 2.2.

Theorem 2.1.1 (Theorem 2.2.3, Informal)**.** *There exists an efficient algorithm such that, if given source data $A_P \in \{0,1\}^{n \times n}$ and target data $A_Q \in \{0,1\}^{n_Q \times n_Q}$ coming from an* appropriate *pair (f_P, f_Q) of latent variable models, outputs $\hat{Q} \in \mathbb{R}^{n \times n}$ such that*

$$\mathbb{P}\left[\frac{1}{n^2}\|Q - \hat{Q}\|_F^2 \leq o(1)\right] \geq 1 - o(1).$$

There must be a relationship between P and Q for them to be an *appropriate* pair for transfer learning. We formalize this relationship below.

Relationship Between Source and Target. It is natural to consider pairs (f_P, f_Q) such that for all $\boldsymbol{x}, \boldsymbol{y} \in \mathcal{X}$, the difference $(f_P(\boldsymbol{x}, \boldsymbol{y}) - f_Q(\boldsymbol{x}, \boldsymbol{y}))$ is small. For example, Cai

and Pu (2024) study transfer learning for nonparametric regression when $f_P - f_Q$ is close to a polynomial in $\boldsymbol{x}, \boldsymbol{y}$. But, requiring $f_P - f_Q$ to be pointwise small does not capture a broad class of pairs in the network setting. For example, if $f_P = \alpha f_Q$. Then $f_P - f_Q = (\alpha - 1) f_Q$ can be far from all polynomials if f_Q is, e.g. a Hölder -smooth graphon.[2] However, under the network model, this means A_P and A_Q are stochastically identical modulo one being α times denser than the other.

We will therefore consider pairs (f_P, f_Q) that are close in some measure of local graph structure. With this in mind, we use a graph distance introduced in Mao et al. (2021) for a different inference problem.

Definition 2.1.2 (Graph Distance)**.** *Let $P \in [0,1]^{n \times n}$ be the probability matrix of a graph. For $i, j \in [n], i \neq j$, we define the graph distance between them as follows:*

$$d_P(i,j) := \|(\boldsymbol{e}_i - \boldsymbol{e}_j)^T P^2 (I - \boldsymbol{e}_i \boldsymbol{e}_i^T - \boldsymbol{e}_j \boldsymbol{e}_j^T)\|_2^2,$$

where $\boldsymbol{e}_i, \boldsymbol{e}_j \in \mathbb{R}^n$ are standard basis vectors.

Intuitively, this first computes the matrix P^2 of common neighbors, and then computes the distance between two rows of the same (ignoring the diagonal elements). We will require that f_P, f_Q satisfy a local similarity condition on the relative rankings of nodes with respect to this graph distance. Since we only estimate the probability matrix of Q, the condition is on the latent variables $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n$ of interest. The hope is that the proximity in graph distance reflects the proximity in latent positions.

Definition 2.1.3 (Rankings Assumption at Quantile h_n)**.** *Let (P, Q) be a pair of graphs evaluated on n latent positions. We say (P, Q) satisfy the rankings assumption at quantile $h_n \leq 1$ if there exists constant $C > 0$ such that for all $i \in [n]$ and all $j \neq i$, if j belongs to the bottom h_n-quantile of $d_P(i, \cdot)$, then j belongs to the bottom Ch_n-quantile of $d_Q(i, \cdot)$.*

To further motivate Definition 2.1.3, recall our motivating example of biological network estimation. Previous works require some form of similarity between networks to enable transfer Sen et al. (2018); Fan et al. (2019); Baranwal et al. (2020). For example, Kshirsagar et al. (2013) require a *commonality hypothesis*: if pathogens A, B target the same

[2] In fact, Cai and Pu (2024) highlight this exact setting as a direction for future work.

neighborhoods in a protein interaction network, one can transfer from A to B. Our rankings assumption similarly posits that to transfer knowledge from A to B, A and B have similar 2-hop neighborhood structures.

Note that Definition 2.1.3 involves quantiles of graph distances; therefore it is a *relative* condition, because it depends on a rank-ordering within both graphs P, Q before comparison. On the other hand, an *absolute* condition would require that for nodes $i, j \in [n]$, if e.g. $d_P(i,j) < 100$ then $d_Q(i,j) < C \cdot 100$. Our condition is more flexible and will hold for a larger set of graph pairs (P, Q), such as pairs where one graph is much more dense than the other.

Finally, to illustrate Definition 2.1.3, consider stochastic block models f_P, f_Q with $k_P \geq k_Q$ communities respectively. If nodes i, j are in the same communities then $P\boldsymbol{e}_i = P\boldsymbol{e}_j$, so $d_P(i,j) = 0$. We require that j minimizes $d_Q(i, \cdot)$. This occurs if and only if $d_Q(i,j) = 0$. Hence if i, j belong to the same community in P, they are in the same community in Q. Note that the converse is not necessary; we could have Q with 1 community and P with arbitrarily many communities.

With the relationship between the source and target defined by the rankings assumption, our contributions are as follows.

(1) Algorithm for Latent Variable Models. We provide an efficient Algorithm 1 for latent variable models with Hölder-smooth f_P, f_Q. The benefit of this algorithm is that it does not assume a parametric form of f_P and f_Q. We prove a guarantee on its error in Theorem 2.2.3.

(2) Minimax Rates. We prove a minimax lower bound for Stochastic Block Models (SBMs) in Theorem 2.3.2. Moreover, we provide a simple Algorithm 2 that attains the minimax rate for this class (Proposition 2.3.4).

(3) Experimental Results on Real-World Data. We test both of our algorithms on real-world metabolic networks and dynamic email networks, as well as synthetic data (Section 2.4).

All proofs are deferred to the Section 2.6.

Next, we review some additional related work.

Transfer learning has recently drawn a lot of interest both in applied and theoretical communities. The notion of transferring knowledge from one domain with a lot of data to

another with less available data has seen applications in epidemiology Apostolopoulos and Bessiana (2020), computer vision Long et al. (2015); Tzeng et al. (2017a); Huh et al. (2016); Donahue et al. (2014); Neyshabur et al. (2020), natural language processing Daumé (2007), etc. For a comprehensive survey see Zhuang et al. (2019); Weiss et al. (2016); Kim et al. (2022). Recently, there have also been advances in the theory of transfer learning Yang et al. (2013); Tripuraneni et al. (2020); Agarwal et al. (2023a); Cai and Wei (2021a); Cai and Pu (2024); Cody and Beling (2023).

In the context of networks, transfer learning is particularly useful since labeled data is typically hard to obtain. Tang et al. (2016) develop an algorithmic framework to transfer knowledge obtained using available labeled connections from a source network to do link prediction in a target network. Lee et al. (2017) proposes a deep learning framework for graph-structured data that incorporates transfer learning. They transfer geometric information from the source domain to enhance performance on related tasks in a target domain without the need for extensive new data or model training. The SGDA method Qiao et al. (2023) introduce adaptive shift parameters to mitigate domain shifts and propose pseudo-labeling of unlabeled nodes to alleviate label scarcity. Zou et al. (2021) proposes to transfer features from the previous network to the next one in the dynamic community detection problem. Simchowitz et al. (2023a) work on combinatorial distribution shift for matrix completion, where only some rows and columns are given. A similar setting is used for link prediction in egocentrically sampled networks in Wu et al. (2018). Zhu et al. (2021) train a graph neural network for transfer based on an ego-graph-based loss function. Learning from observations of the full network and additional information from a game played on the network Leng et al. (2020b); Rossi et al. (2022). Wu et al. (2024) study graph transfer learning for node regression in the Gaussian process setting, where the source and target networks are fully observed.

Levin et al. (2022) proposes an inference method from multiple networks all with the same mean but different variances. While our work is related, we do not assume $\mathbb{E}[P_{ij}] = \mathbb{E}[Q_{ij}]$. Cao et al. (2010) do joint link prediction on a collection of networks with the same link function but different parameters.

Another line of related but different work deals with multiplex networks (Lee et al., 2014b, 2015; Iacovacci and Bianconi, 2016; Cozzo et al., 2018) and dynamic networks Sarkar and Moore (2005); Kim et al. (2018); Sewell and Chen (2015); Sarkar et al. (2012); Chang

et al. (2024); Wang et al. (2023). One can think of transfer learning in clustering as clustering with side information. Prior works consider stochastic block models with noisy label information (Mossel and Xu, 2016; Mazumdar and Saha, 2017b) or oracle access to the latent structure (Mazumdar and Saha, 2017a).

Notation. Throughout this chapter, all asymptotics $O(\cdot), o(\cdot), \Omega(\cdot), \omega(\cdot)$ are with respect to n_Q unless specified otherwise.

2.2 Estimating Latent Variable Models with Rankings

In this section, we present a computationally efficient transfer learning algorithm for latent variable models. Algorithm 1 learns the local structure of P based on graph distances (Definition 2.1.2). For each node i of P, it ranks the nodes in S with respect to the graph distance $d_P(i, \cdot)$. For most nodes $i, j \in [n]$, none of the edges incident to i or j are observed in Q. Therefore, we estimate $\hat{Q}_{ij}$ by using the edge information about nodes $r, s \in S$ such that $d_P(i, r)$ and $d_P(j, s)$ are small.

Formally, we consider a model as in Eq. (2.1) with a compact latent space $\mathcal{X} \subset \mathbb{R}^d$ and latent variables sampled i.i.d. from the normalized Lebesgue measure on $\mathcal{X}$. We set $\mathcal{X} = [0, 1]^d$ without loss of generality and assume that functions $f : \mathcal{X} \times \mathcal{X} \to [0, 1]$ are α-Hölder-smooth.

Definition 2.2.1. *Let $f : \mathcal{X} \times \mathcal{X} \to \mathbb{R}$ and $\alpha > 0$. We say f is α-Hölder -smooth if there exists $C_\alpha > 0$ such that for all $\boldsymbol{x}, \boldsymbol{x}', \boldsymbol{y} \in \mathcal{X}$,*

$$\sum_{\kappa \in \mathbb{N}^d : \sum_i \kappa_i = \lfloor \alpha \rfloor} \left| \frac{\partial^{\sum_i \kappa_i} f}{\partial_{x_1}^{\kappa_1} \cdots \partial_{x_d}^{\kappa_d}}(\boldsymbol{x}, \boldsymbol{y}) - \frac{\partial^{\sum_i \kappa_i} f}{\partial_{x_1}^{\kappa_1} \cdots \partial_{x_d}^{\kappa_d}}(\boldsymbol{x}', \boldsymbol{y}) \right| \leq C_\alpha \|\boldsymbol{x} - \boldsymbol{x}'\|_2^{\alpha \wedge 1}.$$

To exclude degenerate cases where a node may not have enough neighbors in latent space, we require the following assumption.

Assumption 2.2.2 (Assumption 3.2 of Mao et al. (2021))**.** *Let G be a graph on $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n$. There exist $c_2 > c_1 > 0$ and $\Delta_n = o(1)$ such that for all $\boldsymbol{x}_i, \boldsymbol{x}_j$,*

$$c_1 \|\boldsymbol{x}_i - \boldsymbol{x}_j\|^{\alpha \wedge 1} - \Delta_n \leq \frac{1}{n^3} d_G(i, j) \leq c_2 \|\boldsymbol{x}_i - \boldsymbol{x}_j\|^{\alpha \wedge 1}.$$

The second inequality follows directly from Hölder-smoothness, and the first is shown to hold for e.g. Generalized Random Dot Product Graphs, among others (Mao et al., 2021).

We establish the rate of estimation for Algorithm 1 below.

Algorithm 1 $\hat{Q}$-Estimation for Latent Variable Models

Input: $A_P \in \{0,1\}^{n\times n}, A_Q \in \{0,1\}^{n_Q \times n_Q}, S \subset [n]$ s.t. $|S| = n_Q$

Initialize $\hat{Q} \in \mathbb{R}^{n\times n}$ to be all zeroes

For all i, all $j \neq i$, compute graph distances:

$$d_{A_P}(i,j) := \|(\boldsymbol{e}_i - \boldsymbol{e}_j)^T (A_P)^2 (I - \boldsymbol{e}_i\boldsymbol{e}_i^T - \boldsymbol{e}_j\boldsymbol{e}_j^T)\|_2^2$$

Fix a bandwidth $h \in (0,1)$ based on n, n_Q

for $i = 1$ **to** n **do**
  Let $T_i^{A_P}(h) \subset S$ be bottom h-quantile of S with respect to $d_{A_P}(i,\cdot)$ **if** $i \in S$ **then**
    Update $T_i^{A_P}(h) \leftarrow T_i^{A_P}(h) \cup \{i\}$
  end
end
for $i = 2$ **to** n **do**
  for $j = 1$ **to** $i-1$ **do**
    Compute $\hat{Q}_{ij} = \hat{Q}_{ji}$ by averaging:

$$\hat{Q}_{ij} := \frac{1}{|T_i^{A_P}(h)||T_j^{A_P}(h)|} \sum_{r\in T_i^{A_P}(h)} \sum_{s\in T_j^{A_P}(h)} A_{Q;rs}$$

  end
end
return $\hat{Q}$

Theorem 2.2.3. *Let $\hat{Q}$ be as in Algorithm 1. Let f_P be α-Hölder-smooth and f_Q be β-Hölder-smooth for $\beta \geq \alpha > 0$, and let c be an absolute constant. Suppose (P,Q) satisfy Definition 2.1.3 at $h_n = c\sqrt{\frac{\log n_Q}{n_Q}}$ and P satisfies Assumption 2.2.2 with $\Delta_n = O((\frac{\log n}{n_Q})^{\frac{1}{2}\vee\frac{\alpha\wedge 1}{d}})$. Then there exists an absolute constant $C > 0$ such that*

$$\mathbb{P}\left[\frac{1}{n^2}\|\hat{Q} - Q\|_F^2 \lesssim \left(\frac{d}{2}\right)^{\frac{\beta\wedge 1}{2}} \left(\frac{\log n}{n_Q}\right)^{\frac{\beta\wedge 1}{2d}}\right] \geq 1 - n_Q^{-C}.$$

To parse Theorem 2.2.3, consider the effect of various parameter choices. First, observe that our upper bound scales quite slowly with n. Even if n is superpolynomial in n_Q, e.g. $n = n_Q^{\log n_Q}$, then $\log n = O((\log n_Q)^2) = n_Q^{o(1)}$, so the overall effect on the error is dominated by the n_Q term.

Second, the bound is worse in large dimensions, and scales exponentially in $\frac{1}{d}$. This kind of scaling also occurs in minimax lower bounds for nonparametric regression (Tsybakov,

2009), and upper bounds for smooth graphon estimation (Xu, 2018). However, we caution that nonparametric regression can be quite different from network estimation; it would be very interesting to know the dependence of dimension on minimax lower bounds for network estimation, but to the best of our knowledge this is an open problem. Finally notice that a greater smoothness β results in a smaller error, up to $\beta = 1$, exactly as in (Gao et al., 2015; Klopp et al., 2017; Xu, 2018).

2.3 Minimax Rates for Stochastic Block Models

In this section, we will show matching lower and upper bounds for a very structured class of latent variable models, namely, Stochastic Block Models (SBMs).

Definition 2.3.1 (SBM)**.** *Let $P \in [0,1]^{n\times n}$. We say P is an (n,k)-SBM if there exist $B \in [0,1]^{k\times k}$ and $z : [n] \to [k]$ such that for all i, j, $P_{ij} = B_{z(i)z(j)}$. We refer to $z^{-1}(\{j\})$ as community $j \in [k]$.*

We first state a minimax lower bound, proved via Fano's method.

Theorem 2.3.2 (Minimax Lower Bound for SBMs)**.** *Let $k_P \geq k_Q \geq 1$ with k_Q dividing k_P. Let $\mathcal{F}$ be the family of pairs (P,Q) where P is an (n,k_P)-SBM, Q is an (n,k_Q)-SBM, and (P,Q) satisfy Definition 2.1.3 at $h_n = 1/k_P$. Moreover, suppose $S \subset [n]$ is restricted to contain an equal number of nodes from communities $1, 2, \ldots, k_P$ of P. Then the minimax rate of estimation is:*

$$\inf_{\hat{Q}\in[0,1]^{n\times n}} \sup_{(P,Q)\in\mathcal{F}} \mathbb{E}\left[\frac{1}{n^2}\|\hat{Q} - Q\|_F^2\right] \gtrsim \frac{k_Q^2}{n_Q^2}.$$

Note that Definition 2.1.3 at $h_n = 1/k_P$ implies that the true community structure of Q coarsens that of P. The condition that k_Q divides k_P is merely a technical one that we assume for simplicity.

We remark that minimax lower bounds for smooth graphon estimation are established by first showing lower bounds for SBMs, and then constructing a graphon with the same block structure using smooth mollifiers (Gao et al., 2015). Therefore, we expect that Theorem 2.3.2 can also be extended to the graphon setting, using the same techniques. However, sharp lower

bounds for other classes such as Random Dot Product Graphs will likely require different techniques (Xie and Xu, 2020; Yan and Levin, 2023).

Remark 2.3.3 (Clustering Regime)**.** *In Appendix 2.6.4 we also prove a minimax lower bound of* $\frac{\log k_Q}{n_Q}$ *in the regime where the error of recovering the true clustering* z *dominates. This matches the rate of Gao et al. (2015), but for estimating all* n^2 *entries of* Q*, rather than just the* n_Q^2 *observed entries.*

Theorem 2.3.2 suggests that a very simple algorithm might achieve the minimax rate. Namely, use both A_P, A_Q to learn communities, and then use only A_Q to learn inter-community edge probabilities. If (P, Q) are in the nonparametric regime where regression error dominates clustering error (called the *weak consistency* or *almost exact recovery* regime), then the overall error will hopefully match the minimax rate.

We formalize this approach in Algorithm 2, and prove that it does achieve the minimax error rate in the weak consistency regime. To this end, we define the signal-to-noise ratio of an SBM with parameter $B \in [0,1]^{k\times k}$ as follows:

$$s := \frac{p-q}{\sqrt{p(1-q)}},$$

where $p = \min_i B_{ii}, q = \max_{i\neq j} B_{ij}$.

Algorithm 2 $\hat{Q}$-Estimation for Stochastic Block Models

Input: $A_P \in \{0,1\}^{n\times n}, A_Q \in \{0,1\}^{n_Q\times n_Q}, S \subset [n]$ s.t. $|S| = n_Q$

Estimate clusterings $\hat{Z}_P \in \{0,1\}^{n\times k_P}, \hat{Z}_Q \in \{0,1\}^{n_Q\times k_Q}$ using Chen et al. (2014) on A_P, A_Q respectively Let $\hat{W}_Q \in \mathbb{R}^{k_Q\times k_Q}$ be diagonal with

$$\hat{W}_{Q;ii} = (\mathbf{1}^T \hat{Z}_Q \boldsymbol{e}_i)^{-1}$$

Initialize $\hat{\Pi} \in \{0,1\}^{k_P\times k_Q}$ to be all zeroes **for** $i \in S$ **do**

  Let $j_P \in [k_P], j_Q \in [k_Q]$ be the unique column indices at which row i of $\hat{Z}_P, \hat{Z}_Q$ respectively are nonzero Let $\hat{\Pi}_{j_P,j_Q} = 1$

end

Let $\hat{B}_Q \in [0,1]^{k_Q\times k_Q}$ be the block-average:

$$\hat{B}_Q = \hat{W}_Q \hat{Z}_Q^T A_Q \hat{Z}_Q \hat{W}_Q$$

return $\hat{Q} := \hat{Z}_P \hat{\Pi} \hat{B}_Q \hat{\Pi}^T \hat{Z}_P^T$

Proposition 2.3.4 (Error Rate of Algorithm 2)**.** *Suppose $P, Q \in [0,1]^{n\times n}$ are $(n, k_P), (n, k_Q)$-SBMs with minimum community sizes $n^{(P)}_{\min}, n^{(Q)}_{\min}$ respectively. Suppose also that (P, Q) satisfy Definition 2.1.3 at $h_n = n^{(P)}_{\min}/n$. Then if the signal-to-noise ratios are such that: $s_P \geq C(\frac{\sqrt{n}}{n^{(P)}_{\min}} \vee \frac{\log^2(n)}{\sqrt{n^{(P)}_{\min}}})$ and $s_Q \geq C(\frac{\sqrt{n_Q}}{n_{\min(Q)}} \vee \frac{\log^2(n_Q)}{\sqrt{n^{(Q)}_{\min}}})$ for large enough constant $C > 0$, Algorithm 2 returns $\hat{Q}$ such that*

$$\mathbb{P}\left[\frac{1}{n^2}\|\hat{Q} - Q\|_F^2 \lesssim \frac{k_Q^2 \log(n^{(Q)}_{\min})}{n_Q^2}\right] \geq 1 - O\left(\frac{1}{n_Q}\right).$$

2.4 Experiments

In this section, we test Algorithm 1 against several classes of simulated and real-world networks. We use quantile cutoff of $h_n = \sqrt{\frac{\log n_Q}{n_Q}}$ for Algorithm 1 in all experiments.

Baselines. To the best of our knowledge, our exact transfer formulation has not been considered before in the literature. Therefore, we implement two algorithms as alternatives to Algorithm 1.

(1) Algorithm 2. Given $A_P \in \{0,1\}^{n\times n}, A_Q \in \{0,1\}^{n_Q \times n_Q}$, let $k_P = \lceil\sqrt{n}\rceil, k_Q = \lceil\sqrt{n_Q}\rceil$. Compute spectral clusterings $\hat{Z}_P, \hat{Z}_Q$ with k_P, k_Q clusters respectively. Let $J_S \in \{0,1\}^{n_Q \times n}$ is such that $J_{S;ij} = 1$ if and only if $i = j$ and $i \in S$. The projection $\hat{\Pi} \in \mathbb{R}^{k_P \times k_Q}$ solves the least-squares problem $\min_{\Pi \in \mathbb{R}^{k_P \times k_Q}} \|J_S \hat{Z}_P \Pi - \hat{Z}_Q\|_F^2$. We compute the $\hat{\Pi}$ differently from steps 4-7 in Algorithm 2 to account for cases where Q is not a true coarsening of P. When Q is a true coarsening of P, this reduces to the procedure in steps 4-7. Given $\hat{Z}_P, \hat{\Pi}$ we return $\hat{Q}$ as in Algorithm 2.

(2) Oracle. Suppose that an oracle can access data for Q on *all* $n \gg n_Q$ nodes as follows. Fix an error probability $p \in (0,1)$. The oracle is given symmetric $A'_Q \in \{0,1\}^{n\times n}$ with independent entries following a mixture distribution. For all $i, j \in [n]$ with $i < j$ let $X_{ij} \sim \text{Bernoulli}(p)$ and $Y_{ij} \sim \text{Bernoulli}(Q(\boldsymbol{x}_i, \boldsymbol{x}_j))$. Then:

$$A'_{Q;ij} = \mathbb{1}_{i\in S, j\in S} Y_{ij} + (1 - \mathbb{1}_{i\in S, j\in S})((1 - X_{ij})Y_{ij} + X_{ij}(1 - Y_{ij})).$$

Given A'_Q, the oracle returns the estimate from Universal Singular Value Thresholding on A'_Q Chatterjee (2015b). As $p \to 0$, the error will approach $O(n^{\frac{-2\beta}{2\beta+d}})$ for a β-smooth

network on on d-dimensional latent variables (Xu, 2018), so the oracle will outperform any transfer algorithm.

Simulations. We first test on several classes of simulated networks. For $n_Q = 50, n = 200$, we run 50 independent trials for each setting. We report results for each setting in Table 2.1, and visualize estimates for stylized examples in Figure 2.1.

At a glance, Figure 2.1 shows that Algorithms 1 and 2 both work well on Stochastic Block Models (first row), that only Algorithm 1 works well on graphons (second and third rows), and that the Oracle performs well in all cases.

Smooth Graphons. The latent space is $\mathcal{X} = [0, 1]$. We consider graphons of the form $f_\gamma(x, y) = \frac{x^\gamma + y^\gamma}{2}$ where P, Q have different γ. We denote this the γ-Smooth Graphon.

Mixed-Membership Stochastic Block Model. Set $k_P = \lfloor \sqrt{n} \rfloor, k_Q = \lfloor \sqrt{n_Q} \rfloor$. The latent space $\mathcal{X}$ is the probability simplex $\mathcal{X} = \Delta_{k_P} := \{x \in [0, 1]^{k_P} : \sum_i x_i = 1\} \subset \mathbb{R}^{k_P}$. The latent variables $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n$ are iid-Dirichlet distributed with equal weights $\frac{1}{k_P}, \ldots, \frac{1}{k_P}$. Then $P_{ij} = \boldsymbol{x}_i^T B_P \boldsymbol{x}_j$ and $Q_{ij} = \Pi(\boldsymbol{x}_i)^T B_Q \Pi(\boldsymbol{x}_j)$, for connectivity matrices $B_P \in [0, 1]^{k_P \times k_P}, B_Q \in [0, 1]^{k_Q \times k_Q}$, and projection $\Pi : \Delta_{k_P} \to \Delta_{k_Q}$ for a fixed subset of $[k_P]$. For parameters $a, b, \epsilon \in [0, 1]$ we generate $B \in [0, 1]^{k \times k}$ by sampling $E \in \text{Uniform}(-\epsilon, \epsilon)^{k \times k}$ and set $B = \text{clip}((a - b)I + b\mathbf{1}\mathbf{1}^T + E, 0, 1)$. We call this Noisy-MMSB(a, b, ϵ).

Latent Distance Model. The latent space is the unit sphere $\mathcal{X} = \mathbb{S}^{d-1} \subset \mathbb{R}^d$. For scale parameter $s > 0$, we call $f_s(\boldsymbol{x}, \boldsymbol{y}) = \exp(-s\|\boldsymbol{x} - \boldsymbol{y}\|_2)$ the $\mathbb{R}^d$-Latent(s) model.

Discussion. When the latent dimension is larger than 1 (the Noisy MMSB and Latent Variable Models), our Algorithm 1 is better than both Algorithm 2 and the Oracle with $p = 0.1$. Note that Algorithms 1 and 2 use $\frac{n_Q^2}{n^2} \approx 0.06$ unbiased edge observations from Q, while the Oracle with $p = 0.1$ observes $(1 - p)\frac{n^2 - n_Q^2}{n^2} \approx 0.9$ unbiased edge observations in expectation.

Real-World Data. Next, we test on two classes of real-world networks. We summarize our dataset characteristics in Table 2.2. See Appendix 2.8 for further details.

Transfer Across Species in Metabolic Networks. For a fixed organism, a metabolic network has a node for each metabolite, and an edge exists if and only if two metabolites co-occur in a metabolic reaction in that organism. We obtain the unweighted metabolic networks for multiple gram-negative bacteria from the BiGG genome-scale metabolic

Source	Target	Alg. 1	Alg. 2	Oracle ($p = 0.1$)	Oracle ($p = 0.3$)	Oracle ($p = 0.5$)
Noisy-MMSB $(0.7, 0.3, 0.01)$	Noisy-MMSB $(0.9, 0.1, 0.01)$	**0.7473**± **0.0648**	1.3761± 1.1586	*0.9556* ± *0.0633*	2.2568± 0.3107	4.2212± 0.2825
0.1-Smooth Graphon	0.5-Smooth Graphon	*1.7656* ± *0.7494*	4.5033± 1.5613	**0.5016** ± **0.0562**	2.4423± 0.4574	5.7774± 0.7126
$\mathbb{R}^{10}$ Latent(2.5)	$\mathbb{R}^{10}$ Latent(1.0)	**0.5744** ± **0.1086**	1.1773± 1.0481	*0.7715* ± *0.0456*	2.1822± 0.2741	4.3335± 0.3476

Table 2.1: Comparison of different algorithms on simulated networks. Each cell reports $\hat{\mu} \pm 2\hat{\sigma}$ of the mean-squared error over 50 independent trials. Error numbers are all scaled by $1e2$ for ease of reading. Bold: Best algorithm. Emphasis: Second-best algorithm.

Table 2.2: Dataset Characteristics

Name	n	Median Degree	Type
BiGG Model iWFL1372	251	15.00	Source
BiGG Model iPC815	251	12.00	Source
BiGG Model iJN1463	251	14.00	Target
Email-EU Days 1-80	1005	6.92	Source
Email-EU Days 81-160	1005	7.35	Target
Email-EU Days 561-640	1005	7.66	Target

model dataset (King et al., 2016; Norsigian et al., 2020). In the left half of Figure 2.2, we compare two choices of source organism in estimating the network for BiGG model iJN1463 (*Pseudomonas putida*). For a good choice of source, Algorithm 1 is competitive with the Oracle at $p = 0.1$.

Transfer Across Time in the Email Interaction Networks. We use the Email-EU interaction network between $n = 1005$ members of a European research institution across 803 days Leskovec and Krevl (2014); Paranjape et al. (2017). The source graph A_P is the network from day 1 to ≈ 80 ($[1, 80]$). In Figure 2.2 we simulate transfer with targets $[81, 160]$ (left) and $[561, 640]$ (right). We visualize results for arbitrary target periods; similar results hold for other targets. Unlike metabolic networks, Algorithm 2 has comparable performance to both our Algorithm 1 and the oracle algorithm with $p \in \{0.01, 0.05\}$. Compared to

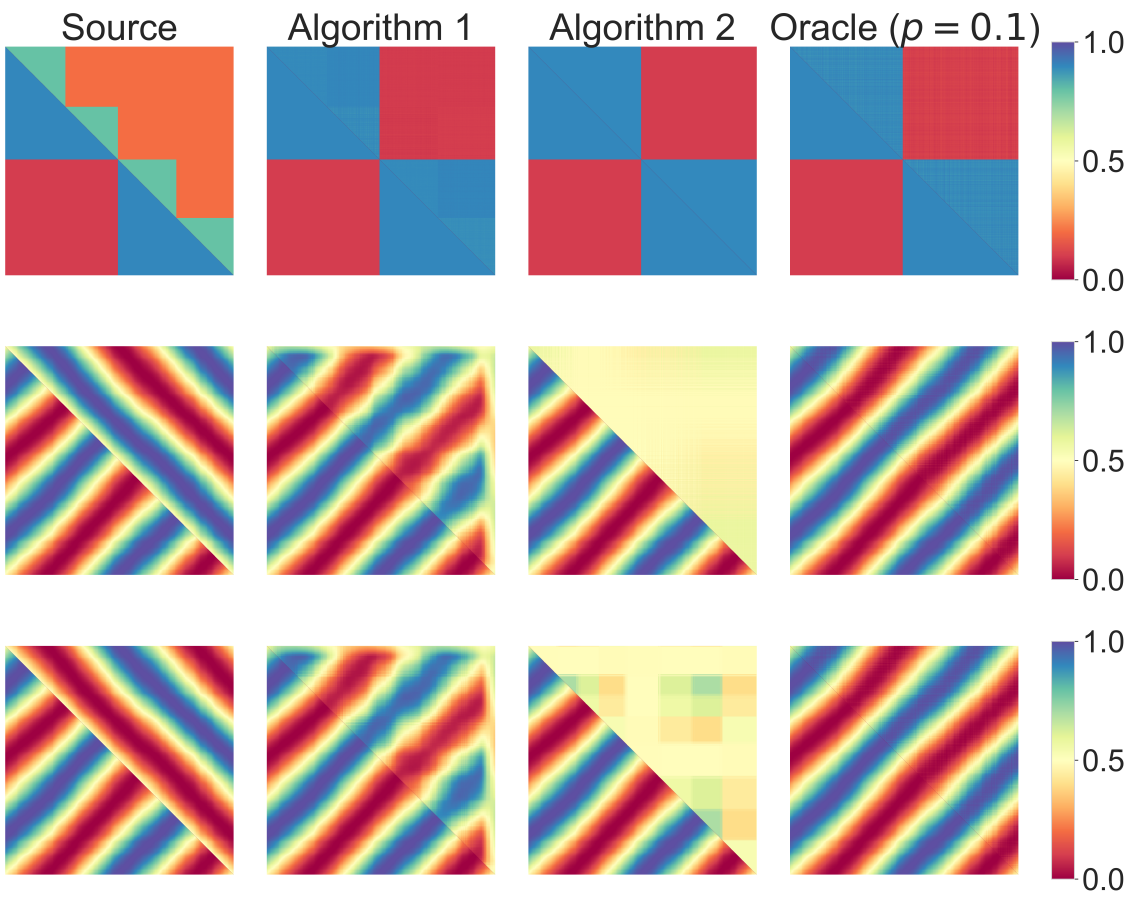


Figure 2.1: Comparison of algorithms on three source-target pairs ($n = 2000, n_Q = 500$). Each row corresponds to a different source/target pair (P, Q). For a fixed row, the upper triangular part on columns 2, 3, 4 corresponds a $\hat{Q}$ for a different algorithm. The upper triangular part of column 1 shows the true P. The lower triangular part of columns 1, 2, 3, and 4 is identical for a fixed row, and shows the true Q. In each heatmap, the lower triangle is the target Q. Algorithm 2 performs best when (P, Q) are SBMs (top), while Algorithm 1 is better for smooth graphons (2nd and 3rd rows).

the metabolic networks, this indicates that the email interaction networks are relatively well-approximated by SBMs, although Algorithm 1 is still the best.

Additional Experiments and Baseline. In Appendix 2.7.1, we present additional ablation experiments that test the dependence of Algorithms 1 and 2 on all relevant parameters. We compare their performance to the Oracle baseline with $p = 0.0$ (the non-transfer setting), and an additional baseline adapted from Levin et al. (2022). We find that our Algorithms outperform this new baseline but are worse than the Oracle with $p = 0.0$, as expected. Further, in Appendix 2.7.2, we test our Algorithms and original baselines on a link

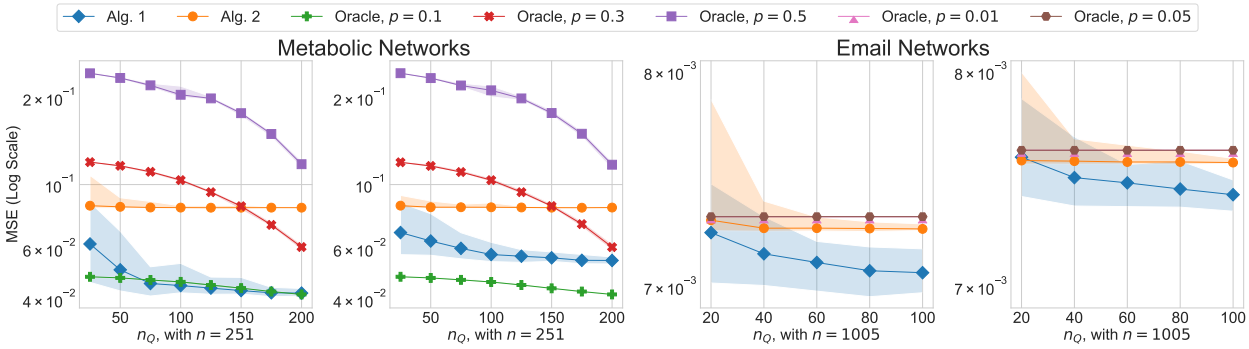


Figure 2.2: Results of network estimation on real-world data. Shaded regions denote $[1, 99]$ percentile outcomes from 50 trials.
Left half: Estimating metabolic network of iJN1463 (*Pseudomonas putida*) with source iWFL1372 (*Escherichia coli W*) leftmost, and source iPC815 (*Yersinia pestis*) second-left.
Right half: Using source data from days $1-80$ of EMAIL-EU to estimate target days $81-160$ (third-left) and target days $561-640$ (rightmost). Note that we use smaller values of p for the Oracle in EMAIL-EU.

prediction task in the setting of Figure 2.2. We find that the relative accuracy of the methods for link prediction is qualitatively similar to that of Figure 2.2, and the Oracle performs even better with sparsity tuning.

2.5 Conclusion

In this paper, we study transfer learning for network estimation in latent variable models. We show that there exists an efficient Algorithm 1 that achieves vanishing error even when $n \geq n_Q^{\omega(1)}$, and a simpler Algorithm 2 for SBMs that achieves the minimax rate.

There are several interesting directions for future work.

First, we believe that Algorithm 1 works for moderately sparse networks with population edge density $\Omega(\frac{1}{\sqrt{n}})$. This is because the concentration of empirical graph distance (Algorithm 1 line 3) requires expected edge density $\tilde{\Omega}(n^{-1/2})$ Mao et al. (2021). It would be interesting to see if a similar approach can work for edge density $\Omega(\frac{\log n}{n})$. For example, in the aforementioned paper it is shown that a variation of the graph distance of Definition 2.1.2 concentrates at expected edge density $\tilde{\Omega}(n^{-2/3})$. While is this still far from the $\Omega(\frac{\log n}{n})$ regime, it suggests that variations on the graph distance might ensure our Algorithm 1 works for sparser graphs.

Second, the case of multiple sources is also interesting. We have focused on the case

of one source distribution, as in Cai and Wei (2021a); Cai and Pu (2024), but expect that our algorithms can be extended to multiple sources as long as they satisfy Definition 2.1.3.

2.6 Proofs

2.6.1 Preliminaries

Recall Hoeffding's inequality.

Lemma 2.6.1 (Hoeffding (1994))**.** *Let $X_1, \dots, X_n$ be independent random variables such that $a_i \le X_i \le b_i$ almost surely for all $i \in [n]$. Then*

$$\mathbb{P}\left[|\sum_{i=1}^{n}(X_i - \mathbb{E}[X_i])| \ge t\right] \le 2\exp\left(\frac{-2t^2}{\sum_{i=1}^{n}(b_i - a_i)^2}\right).$$

We also need Bernstein's inequality.

Lemma 2.6.2 (Bernstein's Inequality)**.** *Let $X_1, \dots, X_n$ be independent mean-zero random variables with $|X_i| \le 1$ for all i and $n \ge 5$. Then*

$$\mathbb{P}\left[|\frac{1}{n}\sum_{i=1}^{n} X_i| \ge t\right] \le 2\exp\left(\frac{-nt^2}{2(1+\frac{t}{3})}\right) \le 2\exp\left(-\frac{nt^2}{4}\right).$$

2.6.2 Proof of Theorem 2.2.3

Throughout this section, let $\mathcal{X} = [0,1]^d$ and $\mu : \mathcal{X} \to [0,1]$ be the normalized Lebesgue measure.

We require the following Lemmata.

Lemma 2.6.3. *Let $\upsilon \in (0,1)$ and $\mu : \mathcal{X} \to [0,1]$ be the normalized Lebesgue measure. Then for all $\boldsymbol{x} \in \mathcal{X}$,*

$$\mu(\mathrm{Ball}(\boldsymbol{x}, 2\upsilon) \cap \mathcal{X}) \ge \mu(\mathrm{Ball}(\mathbf{0}, \upsilon) \cap \mathcal{X}).$$

Proof. Recall $\mathcal{X} = [0,1]^d$. Fix $\boldsymbol{x} \in \mathcal{X}, \upsilon > 0$. Note that $\mu(\mathrm{Ball}(\boldsymbol{x}, \upsilon) \cap \mathcal{X})$ is smallest when $\boldsymbol{x}$ is a vertex of the hypercube; therefore take $\boldsymbol{x} \in \{0,1\}^d$ without loss of generality. Then, note that for each $\boldsymbol{z} \in \mathrm{Ball}(\boldsymbol{x}, \upsilon) \cap \mathcal{X}$, we can find $(2^d - 1)$ other points $\boldsymbol{z}' \in \mathrm{Ball}(\boldsymbol{x}, \upsilon) \setminus \mathcal{X}$ by reflecting

subsets of coordinates of $\boldsymbol{z}$ about $\boldsymbol{x}$. There are $2^d - 1$ such nonempty subsets of coordinates. This shows that $\mu(\text{Ball}(\boldsymbol{x}, \upsilon) \cap \mathcal{X}) \geq \mu(\text{Ball}(\boldsymbol{x}, \upsilon))/2^d$ for all $\boldsymbol{x}$. Since $\mu(\text{Ball}(\boldsymbol{x}, \upsilon)) \asymp \upsilon^d$, the conclusion follows. □

We will repeatedly make use of the concentration of latent positions.

Lemma 2.6.4 (Latent Concentration)**.** *Let $\mathcal{X} = [0,1]^d$ and μ denote the normalized Lebesgue measure on $\mathcal{X}$. Suppose $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n \sim \mathcal{X}$ are sampled iid and uniformly at random from μ. Fix some $T \subset \mathcal{X}$ such that $\mu(T) = v$. Then*

$$\mathbb{P}\left[\left|vn - |\{j \in [n] : \boldsymbol{x}_j \in T\}|\right| \geq 10\sqrt{\frac{\log n}{n}}\right] \leq n^{-10}.$$

Proof. Let X_i be an indicator variable that equals 1 if $\boldsymbol{x}_i \in T$ and zero otherwise. Notice the X_i are iid and bounded within $[0,1]$. Moreover, $\sum_i \mathbb{E}[X_i] = n\mu(T)$. Therefore by Hoeffding's inequality, for any $t > 0$,

$$\mathbb{P}[|vn - |\{j \in [n] : \boldsymbol{x}_j \in T\}|| \geq t] \leq 2\exp\left(\frac{-2t^2}{n}\right).$$

Setting $t = 10\sqrt{\frac{\log n}{n}}$ gives the result. □

Corollary 2.6.5. *Let $\epsilon > 0$. For $i \in [n]$ let $\epsilon_i' > 0$ be $\epsilon_i' := \sup\{\upsilon > 0 : \mu(\text{Ball}(\boldsymbol{x}_i, \upsilon) \cap \mathcal{X}) \leq \epsilon$. Let $T_i := \text{Ball}(\boldsymbol{x}_i, \epsilon_i') \cap \mathcal{X}$. Let $u_i(S) := |\{j \in S : \boldsymbol{x}_j \in T_i\}|$ denote the number of members of S landing in T_i. Then*

$$\mathbb{P}\left[\forall i \in [n] : |u_i(S) - n_Q\epsilon| \leq 10\sqrt{\frac{\log n}{n_Q}}\right] \geq 1 - n^{-8}.$$

Proof. Notice that each T_i has Lebesgue measure ϵ by definition. Therefore $\mathbb{E}[u_i(S)] = n_Q\epsilon$. Since S has n_Q members, setting $t = 10\sqrt{\frac{\log n}{n_Q}}$ in the statement of Lemma 2.6.4 and taking a union bound over all $i \in [n]$ gives the conclusion. □

We will decompose the error of Algorithm 1 into two parts.

Proposition 2.6.6. *Let $\hat{Q} \in [0,1]^{n\times n}$ be the estimator from Algorithm 1. Then*

$$\frac{1}{n^2}\|Q-\hat{Q}\|_F^2 \le \frac{2}{n^2}\sum_{i,j\in[n]}(J_S(i,j)+J_B(i,j)),$$

where J_S, J_B are the smoothing and Bernoulli errors respectively:

$$J_S(i,j) := \frac{1}{|T_i|^2|T_j|^2}\Big(\sum_{r\in T_i, s\in T_j} Q_{ij} - Q_{rs}\Big)^2;$$
$$J_B(i,j) := \frac{1}{|T_i|^2|T_j|^2}\Big(\sum_{r\in T_i, s\in T_j} Q_{rs} - A_{Q;rs}\Big)^2.$$

Controlling the Bernoulli errors is relatively straightforward.

Proposition 2.6.7. *Let h be the bandwidth of Algorithm 1. The Bernoulli error is at most $O(\frac{\log n}{m})$ with probability $\ge 1-n^{-8}$, where $m = h^2 n_Q^2$.*

Proof. Fix $i,j \in [n]$. We will bound the maximum Bernoulli error $J_S(i,j)$ over i,j, which suffices to bound the average. Let $m = |T_i||T_j|$. We want to bound:

$$|\frac{1}{|T_i||T_j|}\sum_{r\in T_i, s\in T_j}(Q_{rs} - A_{Q;rs})|^2.$$

Notice each summand is bounded within $\pm\frac{1}{m}$. Bernstein's inequality gives:

$$\mathbb{P}\left[\left(\frac{1}{|T_i||T_j|}\sum_{r\in T_i, s\in T_j} Q_{rs} - A_{Q;rs}\right)^2 \ge t^2\right] \le 2\exp(-0.5t^2 m).$$

Setting $t = C\sqrt{\frac{\log n}{m}}$ for large enough $C = O(1)$, a union bound tells us that with probability $\ge 1-n^{-8}$, the Bernoulli error is bounded by t^2. □

Corollary 2.6.8. *The Bernoulli error is at most $O(\sqrt{\frac{\log n_Q}{n_Q}})$ with probability $\ge 1-n_Q^{-4}$.*

The rest of this section is devoted to bounded the smoothing errors $J_S(i,j)$.

2.6.2.1 Latent Distance to Graph Distance

We claim that if nodes are close in the latent space then they are close in graph distance.

Proposition 2.6.9. *Suppose that* $\|\boldsymbol{x}_i - \boldsymbol{x}_r\| \leq \epsilon$ *and* Q *is* β*-smooth. Then* $d_Q(i,r) \leq C_\beta^2 n^3 \epsilon^{2(\beta \wedge 1)}$.

Proof. We the use smoothness of Q. By definition there exists $C_\beta > 0$ such that $Q_{ki} - Q_{kr} \leq C_\beta \|\boldsymbol{x}_i - \boldsymbol{x}_r\|^{\beta \wedge 1}$. Therefore,

$$\begin{aligned} d_Q(i,r) &= \sum_{\ell \neq i,r} |(Q^2)_{\ell i} - (Q^2)_{\ell r}|^2 \\ &= \sum_{\ell \neq i,r} \left(\sum_{k \in [n]} Q_{\ell k}(Q_{ki} - Q_{kr}) \right)^2 \\ &\leq \sum_{\ell \neq i,r} \sum_{k \in [n]} Q_{\ell k}^2 C_\beta^2 \epsilon^{2(\beta \wedge 1)} \\ &\leq n^3 C_\beta^2 \epsilon^{2(\beta \wedge 1)}. \end{aligned}$$

□

We can now bound the minimum sizes of the neighborhoods using the concentration of latent positions and the smoothness of the graphon.

Lemma 2.6.10 (Vershynin (2018b))**.** *The volume of a ball of radius* $r > 0$ *in* $\mathbb{R}^d$ *is* $\frac{\sqrt{\pi}^d}{\Gamma(d/2+1)} r^d$*, where* $\Gamma(\cdot)$ *is the* Γ *function.*

Proposition 2.6.11. *Let* $C_d = (\Gamma(\frac{d}{2}+1))^{1/d}$*. Let* C_0, C' *be constants. If* $\upsilon_n \geq C \cdot C_d (\sqrt{\frac{\log n}{n_Q}})^{1/d}$ *for large enough constant* $C > 0$*, and* $g_n = C_0 C_\beta^2 n^2 (\upsilon_n)^{2(\beta \wedge 1)}$*, then with probability* $\geq 1 - n^{-6}$ *for all* $i \in [n]$ *the neighborhood size is* $|\{r : d_Q(i,r) \leq g_n\}| \geq C' n_Q \sqrt{\frac{\log n}{n_Q}}$.

Proof. Fix $i \in [n]$ and $\upsilon_n > 0$. Let ϵ_i denote the Lebesgue measure of $\text{Ball}(\boldsymbol{x}_i, \upsilon_n) \cap \mathcal{X}$. By Lemma 2.6.3 and Lemma 2.6.10, for all i, $\epsilon_i \geq (\frac{\sqrt{\pi}\upsilon_n}{2C_d})^d = (\frac{0.5\sqrt{\pi}\upsilon_n}{C_d})^d$. Let $\epsilon = \min_{i \in [n]} \epsilon_i$.

By Corollary 2.6.5, with probability $\geq 1 - n^{-8}$, there are $n_Q \epsilon - C\sqrt{\frac{\log n}{n_Q}}$ members j of S such that $\|\boldsymbol{x}_i - \boldsymbol{x}_j\| \leq \upsilon_n$. A union bound over i gives the result simultaneously for all i with probability $\geq 1 - n^{-6}$.

From Proposition 2.6.9, it follows that for all $i \in [n]$,

$$|\{r \in S : d_Q(i,r) \leq C_\beta^2 n^2 (2\upsilon_n')^{2(\beta \wedge 1)}\}| \geq n_Q \epsilon - 10\sqrt{\frac{\log n}{n_Q}}.$$

Choosing $\upsilon_n \geq C \cdot C_d(\frac{\log n}{n_Q})^{\frac{1}{2d}}$ for large enough $C > 0$ gives the conclusion. □

2.6.2.2 Graph Distance Concentration

Next, we show that the empirical graph distance concentrates to the population distance.

Proposition 2.6.12. *For any arbitrary symmetric $P \in [0,1]^{n\times n}$, we have, for all i, j simultaneously with probability at least $\geq 1 - O(n^{-8})$, that*

$$|d_{A_P}(i,j) - d_P(i,j)| \leq O(n^2 \log n) + O(n^{2.5}\sqrt{\log n}).$$

Proof. Fix i, j. Let $C_{ij} := (A_P^2)_{ij}$. By Mao et al. (2021) A.1, we have $C_{ij} = (P^2)_{ij} + t_{ij}$ for an error term t_{ij} such that $\mathbb{P}[\forall i, j : |t_{ij}| \leq 10\sqrt{n \log n}] \geq 1 - n^{-10}$. Then,

$$\begin{aligned}
|d_{A_P}(i,j) - d_P(i,j)| &= |\sum_{\ell \neq i,j} \left((C_{i\ell} - C_{j\ell})^2 - ((P^2)_{i\ell} - (P^2)_{j\ell})^2\right)| \\
&= \sum_{\ell \neq i,j} |(t_{i\ell} + t_{j\ell})^2 + 2(t_{i\ell} + t_{j\ell})((P^2)_{i\ell} - (P^2)_{j\ell})| \\
&\leq O(n^2 \log n) + O\left(\sqrt{n \log n} \sum_{\ell \neq i,j} \left((P^2)_{i\ell} - (P^2)_{j\ell}\right)\right).
\end{aligned}$$

Finally, notice that all entries of P^2 are of size $O(n)$, so the conclusion follows. □

Finally, we will show that taking the restriction of the graph distance T_i^P to nodes in $S \subset [n]$ does not incur too much error.

Proposition 2.6.13. *Suppose $n = n_Q^{O(1)}$. Then there exists a constant C such that if $h_0 \geq C\sqrt{\frac{\log n}{n_Q}} + \Delta_n$, then for all i, r simultaneously, $r \in T_i^{A_P}(h_0)$ implies $r \in T_i^P(h_2)$ for some $h_2 = O(h)$ with probability $\geq 1 - O(n^{-5})$.*

Proof. Let us introduce the notation $T_i^{P,S}(h)$ to denote the bottom h-quantile of $\{d_P(i,j) : j \in S\}$. In this notation, $T_i^{A_P}(h) := T_i^{A_P,S}(h)$ since we restrict the quantile to nodes in S. From Proposition 2.6.12 and Assumption 2.2.2, we know that if $n \geq n_Q$ then for $h_0 \leq h_1 - 20\sqrt{\frac{\log n}{n}} - \Delta_n$ we have $T_i^{A_P}(h_0) \subseteq T_i^{P,S}(h_1)$ simultaneously for all $i \in [n]$ with probability $\geq 1 - O(n^{-8})$. It remains to compare $T_i^{P,S}(h_1)$ with $T_i^P(h_2)$ for some h_2.

We claim that if $h_2 \geq 30\sqrt{\frac{\log n_Q}{n_Q}}$ then $\mathbb{P}[\forall i |T_i^P \cap S| \geq h_2 n_Q - 3\sqrt{n_Q \log n_Q}] \geq 1 - O(n_Q^{-2})$. To see this, fix $i \in [n]$ and consider $T_i^P(h_2)$. For $j \in S$, let X_j be the indicator variable:

$$X_j = \begin{cases} 1 & \text{if } j \in T_i^P(h_2), \\ 0 & \text{otherwise.} \end{cases}$$

Notice that $|T_i^P(h_2) \cap S| = \sum_{j \in S} X_j$. By Hoeffding's inequality, since $\mathbb{E}[\sum_{j \in S} X_j] = h_2 n_Q$ and $|X_j - h_2| \leq 1$ for all j, we have

$$\mathbb{P}\left[||T_i^P(h_2) \cap S| - h_2 n_Q| \geq 3\sqrt{n_Q \log n} \right] \leq 2 \exp\left(- \frac{6 n_Q^2 \log n}{n_Q^2} \right) \leq 2n^{-6}.$$

Taking a union bound over all $i \in [n]$ shows the claim holds with probability $\geq 1 - O(n^{-5})$. Therefore we set $h_1 \leq h_2 - 3.1\sqrt{\frac{\log n}{n_Q}}$ then $j \in T_i^{P,S}(h_1)$ implies $j \in T_i^P(h_2)$.

The conclusion follows with $C = 24\sqrt{\frac{\log n}{\log n_Q}} = O(1)$. □

The ranking condition (Definition 2.1.3) then allows us to translate between graph distances in A_P and Q.

Corollary 2.6.14. *Suppose that Definition 2.1.3 holds for (P, Q) at $h_n = c\sqrt{\frac{\log n_Q}{n_Q}} + \Delta_n$, for large enough constant $c > 0$. Suppose $n_Q \leq n \leq n_Q^{O(1)}$. Then for $h > h_n$ and $r \in T_i^{A_P}(h)$, it follows that $r \in T_i^Q(h_3)$ for some $h_3 = O(h)$. The statement holds simultaneously for all i, r with probability $\geq 1 - O(n^{-5})$.*

2.6.2.3 Control of Smoothing Error

We will decompose smoothing error into a sum of two terms called $E_{S,1}$ and $E_{S,2}$. The control of $E_{S,1}$ is relatively straightforward.

Lemma 2.6.15. *The total smoothing error can be bounded with two terms:*

$$\frac{2}{n^2} \sum_{i,j \in [n]} J_S(i,j) \leq E_{S,1} + E_{S,2},$$

where

$$E_{S,1} := \frac{C}{n} \max_{j \in [n], s \in T_j} \|Q(e_j - e_s)\|_2^2;$$

$$E_{S,2} := \frac{4}{n^2} \sum_{i \in [n]} \frac{1}{|T_i|} \mathbb{E}\left[\sum_{r \in T_i} \sum_{j \in [n]} \sum_{s \in T_j} (Q_{rj} - Q_{rs})^2. \right]$$

Proof. Note that

$$\begin{aligned}\frac{2}{n^2}\sum_{i,j\in[n]} J_S(i,j) &= \frac{2}{n^2}\sum_{i,j\in[n]} \frac{1}{|T_i|^2|T_j|^2}\,\mathbb{E}\left[\left(\sum_{r\in T_i, s\in T_j} Q_{ij}-Q_{rs}\right)^2\right] \\ &\le \frac{2}{n}\sum_{i\in[n]}\frac{1}{n|T_i|}\sum_{j\in[n]}\frac{2}{|T_j|}\,\mathbb{E}\left[\sum_{r\in T_i, s\in T_j}(Q_{ij}-Q_{rj})^2+(Q_{rj}-Q_{rs})^2\right] \\ &= \frac{4}{n}\sum_{i\in[n]}\frac{1}{n|T_i|}\,\mathbb{E}\left[\sum_j \frac{1}{|T_j|}\left(\sum_{r\in T_i}(Q_{ij}-Q_{rj})^2+\sum_{r\in T_i}\sum_{s\in T_j}(Q_{rj}-Q_{rs})^2\right)\right].\end{aligned}$$

The second inner summand is precise $E_{S,2}$. For $E_{S,1}$, notice that $|T_i| = |T_j| = h(n_Q - 1)$ by definition. Therefore

$$\sum_j \frac{1}{|T_j|}\sum_{r\in T_i}(Q_{ij}-Q_{rj})^2 = \frac{1}{h(n_Q-1)}\sum_{r\in T_i}\sum_j (Q_{ij}-Q_{rj})^2 \le 2\max_{r\in T_i}\|(e_i-e_r)^T Q\|_2^2. \qquad \square$$

We can now bound $E_{S,1}$ in terms of graph distances.

Lemma 2.6.16. *The smoothing error term $E_{S,1}$ can be bounded as follows:*

$$E_{S,1} \le \frac{2}{n}\max_{i\in[n], r\in T_i}\sqrt{d_Q(i,r)} + \frac{2c}{\sqrt{n}}$$

for some constant $c > 0$.

Proof. Fix $i \in [n]$ and $r \in T_i$. We have

$$\begin{aligned}\|Q(e_i-e_r)\|_2^2 &\le \|e_i-e_r\|_2\|Q^TQ(e_i-e_r)\|_2 \\ &\le 2\|Q^2(e_i-e_r)\|_2.\end{aligned}$$

Now we will pass to graph distances. Let $e_{ab} := ((Q^2)_{aa} - (Q^2)_{ab})^2$ for $a, b \in [n]$. Notice that $\|Q^2(e_i - e_r)\|_2 = \sqrt{d_Q(i,r) + e_{ir} + e_{ri}}$. Moreover, $\sqrt{e_{ir}+e_{ri}} \le 2\sqrt{n}$ since the entries of Q^2 are individually bounded by $O(n)$. The conclusion follows. $\square$

Proposition 2.6.17. *Suppose $\Delta_n = O(\sqrt{\frac{\log n}{n_Q}})$. Let C_d be the constant of Proposition 2.6.11. Then if the bandwidth of Algorithm 1 is $h_n = C\sqrt{\frac{\log n}{n_Q}}$, for a constant $C = O(1)$, then the smoothing error $E_{S,1}$ is at most*

$$E_{S,1} \le C_2 C_d^{\beta\wedge 1}\left(\sqrt{\frac{\log n_Q}{n_Q}}\right)^{\frac{\beta\wedge 1}{d}}$$

for some $C_2 = O(1)$, with probability $\ge 1 - O(n^{-6})$.

Proof. Fix $i \in [n]$ and $r \in T_i^{A_P}(h_n)$. By Corollary 2.6.14, if $h_n \geq C\sqrt{\frac{\log n}{n_Q}} + \Delta_n$ for a large enough constant $C > 0$, then there exists constant $C_2 > 0$ such that the following holds. With probability $\geq 1 - O(n^{-5})$, for all $i \in [n]$ and $r \in S$, $r \in T_i^Q(C_2 h_n)$,

Let $\upsilon_n = CC_d(\sqrt{\frac{\log n}{n_Q}})^{1/d}$ for C_d as in Proposition 2.6.11 and $C > 0$ large enough constant. Then by Proposition 2.6.11 the set of $s \in S$ such that $d_Q(i, r) \leq C_0 C_\beta^2 n^2 (\upsilon_n)^{2(\beta \wedge 1)}$ has size at least $C_2 n_Q \sqrt{\frac{\log n}{n_Q}}$.The statement holds for all i simultaneously with probability at least $1 - O(n^{-6})$. Therefore for all $i \in [n]$ and $r \in T_i^{A_P}(h_n)$, we have

$$d_Q(i, r) \leq C_0 C_\beta^2 n^2 (\upsilon_n)^{2(\beta \wedge 1)}$$

for some $C_0, C_\beta = O(1)$, with probability $\geq 1 - O(n^{-6})$. By Lemma 2.6.16 we conclude that $E_{S,1}$ is bounded by $2\upsilon_n^{\beta \wedge 1} + \frac{2}{\sqrt{n}}$ with the same probability. □

2.6.2.4 Control of the Second Smoothing Error

In this section, we show that the second smoothing error can be controlled in terms of $E_{S,1}$. We will need to track the following quantity.

Definition 2.6.18 (Membership Count)**.** *For $r \in S$ and bandwidth h, distance cutoff ϵ, the P-neighborhood count of r is $\psi_P(r) := |\{j \in [n] : r \in T_j^P(h, \epsilon)\}|$.*

In words, $\psi_P(r)$ counts the number of nodes $j \in [n]$ such that r lands in the neighborhood of j in our algorithm. While we know that $|T_j^P(h)| \leq h n_Q$ always, simply applying the pigeonhole principle gives too weak of a bound on membership counts. The base case is that there may be a "hub" node r lands in $T_j^P(h)$ for all j. We will show that there can be no such hub node.

Supposing that we can control of the empirical count ψ_{A_P}, we show that the smoothing error can be bounded.

Proposition 2.6.19. *Let h_n be the bandwidth. Then*

$$E_{S,2} \leq O\left(\frac{E_{S,1}}{h_n n}\right) \cdot \max_{r \in [n]}(\psi_{A_P}(r)).$$

Proof. Rearranging terms, we have

$$\begin{aligned}
E_{S,2} &= \frac{1}{n^2h^2n_Q^2} \sum_{i,j\in[n], r\in T_i, s\in T_j} (Q_{rj} - Q_{rs})^2 \\
&= \frac{1}{n^2h^2n_Q^2} \sum_{r\in S} \psi_{A_P}(r) \sum_{j,s} (Q_{rj} - Q_{rs})^2 \\
&= \frac{n_Q}{n^2h^2n_Q^2} \mathop{\mathbb{E}}_{r\in S}\left[\psi_{A_P}(r) \sum_{j,s} (Q_{rj} - Q_{rs})^2\right] \\
&= \frac{n_Q}{n^2h^2n_Q^2} \mathop{\mathbb{E}}_{r\in [n]}\left[\psi_{A_P}(r) \sum_{j,s} (Q_{rj} - Q_{rs})^2\right],
\end{aligned}$$

where the last step follows because j, s do not depend on i, r and because $S \subset [n]$ is chosen uniformly at random. Now, we will control the expectation by passing to a row sum, which is handled by $E_{S,1}$.

$$\mathop{\mathbb{E}}_{r\in[n]}\left[\psi_{A_P}(r) \sum_{j,s} (Q_{rj} - Q_{rs})^2\right] \le \max_{r\in[n]} \left(\frac{\psi_{A_P}(r)}{n}\right) \cdot \sum_{j\in[n]} \sum_{s\in T_j} \|Q(e_j - e_s)\|_2^2.$$

Recall that $n^2 n_Q h_n E_{S,1} = \Omega\Big(\sum_{j\in[n]} \sum_{s\in T_j} \|Q(e_j - e_s)\|_2^2\Big)$. Hence we conclude that

$$E_{S,2} \le O\left(\frac{E_{S,1}}{h_n n}\right) \cdot \max_{r\in[n]}(\psi_{A_P}(r)). \qquad \square$$

We therefore must show that $\max_{r\in S} \psi_{A_P}(r) \le O(hn)$ with high probability.

Proposition 2.6.20 (Population Version)**.** *Suppose Assumption 2.2.2 holds for P with $c_1 < c_2$ and $\Delta_n = O((\frac{\log n}{n_Q})^{\frac{1}{2}\vee\frac{\alpha\wedge 1}{d}})$. Then if $h \le C\sqrt{\frac{\log n}{n_Q}}$ for large enough constant $C > 0$, then we have $\max_{r\in S} \psi_P(r) \le O(hn)$ with probability at least $1 - O(n_Q^{-8})$.*

Proof. Fix $r \in S$. Let C_d be as in Proposition 2.6.11.Suppose that $\epsilon = C_d(C+10)\sqrt{\frac{\log n_Q}{n_Q}}^{1/d}$ and $h = C\sqrt{\frac{\log n_Q}{n_Q}}$. Now, we will claim that for large enough constant $c > 0$, that $\psi_P(r)$ is at most the size of $\text{Ball}(\boldsymbol{x}_r, c\epsilon) \cap \{\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n\}$.

Suppose that $c > 0$ is a large enough constant. Now suppose that $\boldsymbol{x}_j$ is such that $\|\boldsymbol{x}_j - \boldsymbol{x}_r\| \ge c\epsilon$. We can lower bound the graph distance using Assumption 2.2.2, as:

$$d_P(r, j) := \|(e_r - e_j)^T P^2 (I - e_r e_r^T - e_j e_j^T)\|_2^2 \ge c_1 n^3 (c\epsilon)^{2(\alpha\wedge 1)} - n^3\Delta_n.$$

On the other hand, suppose that $i \in S$ is such that $\|\boldsymbol{x}_i - \boldsymbol{x}_j\| \le \epsilon$. Then $d_P(i,j) \le C_\alpha^2 n^3 \epsilon^{2(\alpha\wedge 1)}$ by Proposition 2.6.9. Therefore since $\epsilon = C_d(C+10)\sqrt{\frac{\log n_Q}{n_Q}}^{1/d}$ and $\Delta_n = O((\frac{\log n}{n_Q})^{\frac{1}{2}\vee\frac{\alpha\wedge 1}{d}})$, for large enough $c_1 > 0$ we have

$$d_P(r,j) := \|(e_r - e_j)^T P^2 (I - e_r e_r^T - e_j e_j^T)\|_2^2 \ge \frac{c_1}{2} n^3 (c\epsilon)^{2(\alpha\wedge 1)}.$$

Then, if we choose $c > 0$ such that $c^{2(\alpha\wedge 1)} > \frac{2C_\alpha^2}{c_1}$, then $d_P(i,j) < d_P(r,j)$.

Next, from our choices of h, ϵ, by Corollary 2.6.5, simultaneously for all $i \in [n]$ there are at least hn_Q nodes in S that have distance $\le \epsilon$ in latent space from $\boldsymbol{x}_i$, with probablity $\ge 1 - O(n_Q^{-6})$.

Therefore, if $\boldsymbol{x}_r \notin \text{Ball}(\boldsymbol{x}_j, c\epsilon) \cap \{\boldsymbol{x}_1, \dots, \boldsymbol{x}_n\}$ then $r \notin T_j^P(h)$. This implies that $\psi_P(r) \le |\{\text{Ball}(\boldsymbol{x}_r, 2c\epsilon) \cap \{\boldsymbol{x}_1, \dots, \boldsymbol{x}_n\}|$. We can bound the size of this ball with Lemma 2.6.4. Notice the Lebesgue measure of $\text{Ball}(\boldsymbol{x}_r, 2c\epsilon) \cap [0,1]$ is at most $(\frac{4c\epsilon}{C_d})^d$. Therefore, since $\boldsymbol{x}_i$ are chosen iid from the Lebesgue measure on $\mathcal{X}$, with probability at least $\ge 1 - O(n_Q^{-10})$, we have

$$\frac{1}{n}|\text{Ball}(\boldsymbol{x}_r, 2c\epsilon) \cap \{\boldsymbol{x}_1, \dots, \boldsymbol{x}_n\}| \le 2c\epsilon + 10\sqrt{\frac{\log n}{n}}.$$

The right-hand side is bounded by $O(h)$ if $n \ge n_Q$. Taking a union bound over all $r \in S$ gives the conclusion. □

We conclude with the desired upper bound.

Proposition 2.6.21 (Bound on $\psi_{A_P}(r)$)**.** *Suppose Assumption 2.2.2 holds for P with $c_1 < c_2$ and $\Delta_n = O((\frac{\log n}{n_Q})^{\frac{1}{2}\vee\frac{\alpha\wedge 1}{d}})$. Then if $h \le C_0\sqrt{\frac{\log n_Q}{n_Q}}$ for small enough constant C_0, then we have $\max_{r\in S} \psi_{A_P}(r) \le O(hn)$ with probability at least $1 - O(n_Q^{-8})$.*

Proof. By Proposition 2.6.12, with probability at least $1 - O(n_Q^{-8})$, we have for all $r \in S, j \in [n]$ simultaneously that

$$\begin{aligned} d_{A_P}(r,j) &\ge d_P(r,j) - O(n^{2.5}\sqrt{\log n}) \\ &\ge (1 - O(\frac{1}{\sqrt{n}}))d_P(r,j). \end{aligned}$$

Similarly, $d_{A_P}(r,j) \le (1 + O(\frac{1}{\sqrt{n}}))d_P(r,j)$. We conclude that $\psi_{A_P}(r) \le 2\psi_P(r) = O(hn)$ with probability $\ge 1 - O(n_Q^{-8})$. □

2.6.2.5 Overall Error

We can bound $C_d := \Gamma(\frac{d}{2}+1)^{1/d}$ with the elementary inequality.

Lemma 2.6.22. *Let $C_d := \Gamma(\frac{d}{2}+1)^{1/d}$. Then $C_d \le \sqrt{d/2}$.*

Proof of Theorem 2.2.3. By Proposition 2.6.21 and Prop 2.6.19, we have that $E_{S,1} \le O(E_{S,1})$ with probability at least $1 - O(n_Q^{-8})$. Therefore by Proposition 2.6.17,

$$\mathbb{P}\left[E_{S,1} + E_{S,2} \le O\left(C_d^{\beta\wedge 1}\left(\frac{\log n}{n_Q}\right)^{\frac{\beta\wedge 1}{2d}}\right)\right] \ge 1 - O(n_Q^{-6}).$$

By Lemma 2.6.22, $C_d \le \sqrt{d/2}$. Finally, by Corollary 2.6.8, the Bernoulli error is bounded by $O(\sqrt{\frac{\log n_Q}{n_Q}})$ with probability $\ge 1 - O(n_Q^{-4})$. Applying a union bound over the two kinds of error and Lemma 2.6.15 gives the result. □

2.6.3 Proof of Theorem 2.3.2

Recall the Gilbert-Varshamov code (Guruswami et al., 2019).

Theorem 2.6.23 (Gilbert-Varshamov)**.** *Let $q \ge 2$ be a prime power. For $0 < \epsilon < \frac{q-1}{q}$ there exists an ϵ-balanced code $C \subset \mathbb{F}_q^n$ with rate $\Omega(\epsilon^2 n)$.*

We will use the following version of Fano's inequality.

Theorem 2.6.24 (Generalized Fano Method, Yu (1997))**.** *Let $\mathcal{P}$ be a family of probability measures, $(\mathcal{D}, d)$ a pseudo-metric space, and $\theta : \mathcal{P} \to \mathcal{D}$ a map that extracts the parameters of interest. For a distinguished $P \in \mathcal{P}$, let $X \sim P$ be the data and $\hat{\theta} := \hat{\theta}(X)$ be an estimator for $\theta(P)$.*

Let $r \ge 2$ and $\mathcal{P}_r \subset \mathcal{P}$ be a finite hypothesis class of size r. Let $\alpha_r, \beta_r > 0$ be such that for all $i \ne j$, and all $P_i, P_j \in \mathcal{P}_r$,

$$d(\theta(P_i), \theta(P_j)) \ge \alpha_r;$$
$$KL(P_i, P_j) \le \beta_r.$$

Then

$$\max_{j\in[r]} \mathbb{E}_{P_j}[d(\hat{\theta}(X), \theta(P_j))] \ge \frac{\alpha_r}{2}\left(1 - \frac{\beta_r + \log 2}{\log r}\right).$$

Definition 2.6.25 (Relative Hamming Distance)**.** *For $\boldsymbol{x}, \boldsymbol{y} \in \{0,1\}^m$, we define their* relative Hamming distance *as follows:*

$$d_H(\boldsymbol{x}, \boldsymbol{y}) := \frac{1}{m}|\{i \in [m] : x_i \neq y_i\}|.$$

We will need the following construction of coupled codes.

Proposition 2.6.26. *Let $m_P, m_Q \geq 2$ and m_Q divide m_P. There exists a code $C \subset \{0,1\}^{m_P}$ and a projection map $\Pi : \{0,1\}^{m_P} \to \{0,1\}^{m_Q}$ such that if $C' = \{\Pi(w) : w \in C\}$ then C' is a code with relative Hamming distance $\Omega(1)$. Moreover, $|C| = |C'| \geq 2^{0.1 m_Q}$*

Throughout the proof, we will identify the community assignment function $z : [n] \to [k]$ of an SBM (Definition 2.3.1) with the matrix $Z \in \{0,1\}^{n \times k}$ where $Z_{ij} = 1$ if and only if $z(i) = j$.

Proof. Begin with a Gilbert-Varshamov code $B \subset \{0,1\}^{m_Q}$ as in Theorem 3.6.19. We can "lift" B to a code on $\{0,1\}^{m_P}$ simply by concatenation. If $w \in B$, then the corresponding $w' \in C$ is just $w' = (w, w, \ldots, w) \in \{0,1\}^{m_P}$. Let $\Pi : \{0,1\}^{m_P} \to \{0,1\}^{m_Q}$ simply select the first m_Q bits of a word. It is clear that $B = \{\Pi(w) : w \in C\}$, so we are done. □

Now we are ready to prove Theorem 2.3.2.

Proof of Theorem 2.3.2. Let $m_P = \binom{n}{2}$, $m_Q = \binom{n_Q}{2}$, and $m = m_P$. Let $C \subset \{0,1\}^{m_P}$ be the code and $\Pi : \{0,1\}^{m_P} \to \{0,1\}^{m_Q}$ the projection map of Prop 2.6.26. For each $w \in C$, we construct a pair of SBMs $P_w, Q_w \in \mathbb{R}^{n \times n}$ as follows.

Each P_w, Q_w is a stochastic block model with k_P, k_Q classes respectively. All the P_w share the same community structure, namely the lexicographic assignment where nodes $1, 2, \ldots, \frac{n}{k_P}$ are assigned to community 1, and so on. Similarly all the Q_w share the same lexicographic community structure with nodes $1, 2, \ldots, \frac{n}{k_Q}$ assigned to community 1, and so on. Therefore, there are fixed $Z_P \in \{0,1\}^{n \times k_P}, Z_Q \in \{0,1\}^{n \times k_Q}$, such that for all $w \in C$, there exist $A_w \in \mathbb{R}^{k_P \times k_P}, B_w \in \mathbb{R}^{k_Q \times k_Q}$ with

$$P_w = Z_P A_w Z_P^T,$$
$$Q_w = Z_Q B_w Z_Q^T.$$

The A_w, B_w are defined as follows. Let $i, j \in [k_P]$ and $i', j' \in [k_Q]$ be such that $i < j$ and $i' < j'$. Since $m_P = \binom{k_P}{2}$ and $m_Q = \binom{k_Q}{2}$, we can identify (i, j) and (i', j') with indices of $[m_P], [m_Q]$ respectively. Then for fixed $\delta_P, \delta_Q > 0$, the edge connectivity probabilities are

$$A_w(i,j) = A_w(j,i) := \begin{cases} 1/2 & \text{if } w_{ij} = 0, \\ 1/2 + \delta_P & \text{if } w_{ij} = 1; \end{cases}$$
$$B_w(i',j') = B_w(j',i') := \begin{cases} 1/2 & \text{if } \Pi(w)_{i'j'} = 0, \\ 1/2 + \delta_Q & \text{if } \Pi(w)_{i'j'} = 1. \end{cases}$$

We can set the diagonals of A_w, B_w to be $1/2$ as well.

Next, let $\mathcal{P}_r$ be a family of $r = |C|$ probability measures. For fixed $w \in C$, the corresponding measure is the distribution over data $(A_P, A_Q) \in \{0,1\}^{n \times n} \times \{0,1\}^{n_Q \times n_Q}$ sampled from $(P_w, Q_w[S, S])$. Note that we restrict S to be a fixed subset of $[n]$.

Next, let $\theta((P_w, Q_w)) := Q_w$, and let $d(\theta((P_w, Q_w)), \theta((P_{w'}, Q_{w'}))) := \frac{1}{n}\|Q_w - Q_{w'}\|_F$. We will show that for all $w, w' \in C$ with $w \neq w'$,

$$\begin{aligned} KL((P_w, Q_w), (P_{w'}, Q_{w'})) &\leq KL(P_w, P_{w'}) + KL(Q_w, Q_{w'}) \\ &\leq O(n^2\delta_P^2 + n_Q^2\delta_Q^2) \\ &=: \beta, \\ d((P_w, Q_w), (P_{w'}, Q_{w'})) &:= \frac{1}{n}\|Q_w - Q_{w'}\|_F \\ &\geq \Omega(\delta_Q) \\ &=: \alpha. \end{aligned}$$

For the β claim, by Proposition 4.2 of Gao et al. (2015), if $\delta_P, \delta_Q \in (0, 1/4)$, we have

$$\begin{aligned} KL((P_w, Q_w), (P_{w'}, Q_{w'})) &\leq KL(P_w, P_{w'}) + KL(Q_w, Q_{w'}) \\ &\lesssim \sum_{i,j \in [n]} (P_w(i,j) - P_{w'}(i,j))^2 + (Q_w(i,j) - Q_{w'}(i,j))^2. \end{aligned}$$

Next, notice that $A_w(i,j) \neq A_{w'}(i,j)$ if and only if $w_{ij} \neq w'_{ij}$. Then for distinct $w, w' \in C$, we have $d_H(w, w') = \Omega(m_P)$, so

$$\sum_{i,j \in [n]} (P_w(i,j) - P_{w'}(i,j))^2 \lesssim \delta_P^2 \frac{n^2}{k_P^2} d_H(w, w') \binom{k_P}{2} \lesssim \delta_P^2 n^2.$$

The bound for Q_w is similar, so this verifies the β claim.

Similarly, for the α claim, notice that

$$\frac{1}{n}\|Q_w - Q_{w'}\|_F \gtrsim \frac{1}{k_Q}\sqrt{\delta_Q^2 d_H(\Pi(w), \Pi(w'))} \geq \frac{\delta_Q}{k_Q}\sqrt{d_H(\Pi(w), \Pi(w'))}.$$

By Prop 2.6.26, $d_H(\Pi(w), \Pi(w')) = \Omega(m_Q) = \Omega(k_Q^2)$. Therefore $\alpha \leq \Omega(\delta_Q)$.

Next, by Prop 2.6.27, the pair (P_w, Q_w) satisfies Definition 2.1.3 for all $w \in C$. Moreover, $\log|C| \geq 0.1 m_Q$ by Prop 2.6.26.

Combining these results, by Theorem 3.6.20 the overall lower bound is

$$\begin{aligned}\inf_{\hat{Q}} \sup_w \frac{1}{n}\|\hat{Q} - Q_w\|_F &\gtrsim \alpha\left(1 - \frac{\beta + \log 2}{0.1\binom{k_Q}{2}}\right) \\ &\geq \delta_Q\left(1 - \frac{30n^2\delta_P^2}{k_Q^2} - \frac{30n_Q^2\delta_Q^2}{k_Q^2} - o(1)\right).\end{aligned}$$

If we choose $\delta_P = 0.01(\frac{k_Q}{n})$ and $\delta_Q = 0.01\frac{k_Q}{n_Q}$, then

$$\begin{aligned}\inf_{\hat{Q}} \sup_w \frac{1}{n^2}\|\hat{Q} - Q_w\|_F^2 &\gtrsim \delta_Q^2 \\ &\gtrsim \frac{k_Q^2}{n_Q^2}.\end{aligned}$$

Note that $k_Q \leq n_Q \leq n$, so $\delta_P, \delta_Q \in (0, 1/4)$ as desired. □

Proposition 2.6.27. *If $h_n = \min\{\frac{1}{k_P}, \frac{1}{k_Q}\}$ then for all $w \in C$, the pair (P_w, Q_w) satisfies Defn 2.1.3 at h_n.*

Proof. Consider $h = h_n$ and some node $i \in [n]$. Suppose that $j \neq i$ is in the same P_w-community as i, and that $\ell \neq i$ is in a different community. Then notice that $d_{P_w}(i, \ell) \geq d_{P_w}(i, j)$. Therefore $j \in T_i^{P_w}(h)$. Moreover, since $h \leq \frac{1}{k_P}$ and since the nodes of $S \subset [n]$ are equidistributed among the communities $1, 2, \ldots, k_P$, it follows that all members of $T_i^{P_w}(h)$ must belong to the same P_w-community as i.

Therefore, since the communities of Q_w are a coarsening of the communities of P_w, $j \in T_i^{Q_w}(\frac{1}{k_Q})$. Since $h \leq \frac{1}{k_Q}$, we are done. □

2.6.4 SBM Clustering Error

In this section, we prove a minimax lower bound in the clustering regime for stochastic block models.

Theorem 2.6.28. *Let Π denote the parameter space of pairs of SBMs (P, Q) on n nodes with k_P, k_Q communities respectively, such that the cluster structure of Q is a coarsening the cluster structure of P. Then*

$$\inf_{\hat{Q}} \sup_{(P,Q)\in\Pi} \mathbb{E}[\frac{1}{n^2}\|\hat{Q} - Q_i\|_F^2] \gtrsim \frac{\log k_Q}{n_Q}.$$

Proof. Let $H_m \in [0,1]^{m\times m}$ be the Hadamard matrix of order m modified to replace all entries -1 with 0. If m is not a power of two, let H_m be defined as follows. Let $\ell = \lfloor \log_2 m \rfloor$ and let $H_{m'} \in \mathbb{R}^{m/2\times m/2}$ contain $H_{2^{\ell-1}}$ on its top left block and zeroes elsewhere. Let

$$H_m = \begin{bmatrix} \mathbf{0}\mathbf{0}^T & H_{m'} \\ H_{m'}^T & \mathbf{0}\mathbf{0}^T \end{bmatrix}.$$

Notice that at most $\frac{7}{8}$ fraction of the entries of H_m are zero-padded, for any m. Now, let $B_P = \frac{1}{2}\mathbf{1}\mathbf{1}^T + \delta_P H_{k_P}$ and $B_Q = \frac{1}{2}\mathbf{1}\mathbf{1}^T + \delta_Q H_{k_Q}$ for some $\delta_P, \delta_Q \in (0, 1/4)$ to be chosen later.

We will define two families of matrices indexed by a finite set T. For $i \in T$, there are some $Z_i \in \{0,1\}^{n\times k_P}$ and $Y_i \in \{0,1\}^{n\times k_Q}$ to be specified later. Then

$$P_i = Z_i B_P Z_i^T,$$
$$Q_i = Y_i B_Q Y_i^T.$$

Now, we define Y_i as follows. Let Z_{n,k_Q} denote the set of balanced clusterings $z : [n] \to [k_Q]$ such that for all $i, j \in [k_Q]$, $|z^{-1}(\{i\})| = |z^{-1}(\{j\})|$. Let $Z \subset Z_{n,k_Q}$ select the z such that for all $j \le k_Q/2$, $z^{-1}(j) = \{\left\lfloor \frac{n(j-1)}{k_Q} \right\rfloor, \ldots, \left\lfloor \frac{nj}{k_Q} \right\rfloor\}$. Define a distance on Z as follows. For $y, y' \in Z$ let $Y, Y' \in \{0,1\}^{n\times k_Q}$ be the corresponding cluster matrices and let $d(y, y') := \frac{1}{n}\|YB_QY^T - Y'B_Q(Y')^T\|_F$. By Theorem 2.2 of Gao et al. (2015), there exists a packing $T_0 \subset Z$ with respect to d such that for all $y, y' \in T_0$, we have $|\{j : y'(j) \neq y(j)\}| \ge n/6$. Moreover, $\log|T_0| \ge \frac{1}{12} n \log k_Q$. Set $T = T_0$. For any $y_i \in T_0$, let $Y_i \in \{0,1\}^{n\times k_Q}$ be the corresponding cluster matrix and then $Q_i = Y_i B_Q Y_i^T$.

Now, to define Z_i, take $a \in [k_Q]$ and partition $y_i^{-1}(\{a\}) \subset [n]$ into k_P/k_Q equally sized communities in a uniformly random way. Number these $1, \ldots, \frac{k_P}{k_Q}$. In this way, we split community 1 of y_i into communities $1, \ldots, \frac{k_P}{k_Q}$ of z_i, and so on. Define Z_i to be the matrix corresponding to z_i. Notice that Z_i, Y_i are both balanced clusterings and that the

clustering Y_i coarsens that of Z_i. Therefore (P_i, Q_i) are a pair of heterogeneous symmetric SBMs satisfying Definition 2.1.3 at $h = 1/k_Q$.

Next, we apply Fano's Inequality (Theorem 3.6.20). Recall $\log|T| \geq \frac{1}{12} n \log k_Q$. Now, for $i, j \in T$ distinct, Prop 4.2 of Gao et al. (2015) gives

$$D_{KL}((P_i, Q_i), (P_j, Q_j)) \leq D_{KL}(P_i, P_j) + D_{KL}(Q_i, Q_j) \leq O(n^2\delta_P^2 + n_Q^2\delta_Q^2) =: \gamma_1.$$

Finally, we can bound:

$$\begin{aligned}\frac{1}{n^2}\|Q_i - Q_{i'}\|_F^2 &\geq \frac{1}{n^2} \sum_{n/2<j\leq n} \frac{n}{k_Q} \|(e_{y_i(j)} - e_{y_i'(j)})B_Q\|^2 \\ &\geq c_0\delta_Q^2 =: \gamma_2^2,\end{aligned}$$

where $c_0 > 1$ is some constant. This follows because there are a constant fraction of $j > n/2$ such that $y_i(j) \neq y_i'(j)$, and any two rows of the Hadamard matrix differ on half their entries.

Now, set $\delta_Q^2 = \frac{n_Q \log k_Q}{10n_Q^2}$ and $\delta_P^2 = \frac{\log k_Q}{10n^2}$. Since $n \geq n_Q$, we conclude that

$$\begin{aligned}\inf_{\hat{Q}} \sup_{i\in T} \mathbb{E}\left[\frac{1}{n^2}\|\hat{Q} - Q_i\|_F^2\right] &\gtrsim \gamma_2^2\left(1 - \frac{\gamma_1 + \log 2}{(1/12)n\log k_Q}\right) \\ &\gtrsim \frac{\log k_Q}{n_Q}.\end{aligned}$$

□

2.6.5 Proof of Proposition 2.3.4

We first argue that Algorithm 2 perfectly recovers Z_P, Z_Q with high probability.

Theorem 2.6.29 (Implicit in Chen et al. (2014))**.** *Let $M = ZBZ^T$ be an (n, n_{min}, s)-HSBM. Then there exists absolute constant $C > 0$ such that the Algorithm of Chen et al. (2014) can recover Z, up to permutation, with zero error with probability $\geq 1 - O(n^{-8})$ if*

$$s \geq C\left(\frac{\sqrt{n}}{n_{\min}} \vee \frac{\log^2(n)}{\sqrt{n_{\min}}}\right).$$

Proof. The algorithm of Chen et al. (2014) returns a matrix $Y \in \{0,1\}^{n\times n}$ such that $Y_{ij} = 1$ if and only if i, j are in the same community, with probability $\geq 1 - O(n^{-8})$. Therefore, to construct a clustering from Y, simply assign the cluster of node 1 to all $j \in [n]$ such that

$Y_{1j} = 1$, and so on. This returns the true $Z \in \{0,1\}^{n\times k}$ up to permutation with probability $\geq 1 - O(n^{-8})$. Note that k is correctly chosen because Y is equal to a block-diagonal matrix of ones up to permutation, with k blocks. □

Theorem 2.6.29 implies the following.

Proposition 2.6.30. *Let $\hat{Z}_P, \hat{Z}_Q$ be as in Algorithm 2. Let s_P, s_Q be the signal to noise ratios of P, Q respectively. If s_P, s_Q satisfy the conditions of Theorem 2.6.29 with respect to $(n, n_{\min}^{(P)})$ and $(n_Q, n_{\min}^{(Q)})$ respectively, then then with probability $\geq 1 - O(n_Q^{-8})$, there are permutation matrices $U_P \in \{0,1\}^{k_P\times k_P}, U_Q \in \{0,1\}^{k_Q\times k_Q}$ such that $\hat{Z}_P = Z_P U_P$ and $\hat{Z}_Q = Z_Q U_Q$.*

Next, we want to recover the clustering of Q on all n nodes, not just the n_Q nodes that we observe in A_Q. This is given by the following.

Proposition 2.6.31. *1. If $h_n = 1/k_P$ and $k_Q \leq k_P$ then there exists a unique $\Pi \in \{0,1\}^{k_P\times k_Q}$ such that $Z_P\Pi$ contains the Q-clustering of all nodes in $[n]$. Let $\tilde{Z}_Q := Z_P\Pi$.*

2. Let $\hat{\Pi}$ be as in Algorithm 2 and U_P, U_Q be as in Proposition 2.6.30. Then with probability $1 - O(\frac{1}{n_Q})$, $Z_P U_P \hat{\Pi} = \tilde{Z}_Q U_Q$.

Proof. Part (1) follows immediately from the SBM structure of P, Q and definition of Definition 2.1.3.

For Part (2), first notice that by Proposition 2.6.30, with probability at least $1 - O(\frac{1}{n_Q})$, Algorithm 2 returns the true clusterings $\hat{Z}_P = Z_P \in \{0,1\}^{n\times k_P}$ and $\hat{Z}_Q = Z_Q \in \{0,1\}^{n_Q\times k_Q}$, up to permutation.

Now, Algorithm 2 simply takes unions of the clusters of Z_P to learn $\hat{\Pi}$. Therefore, let $V : \mathbb{R}^n \to \mathbb{R}^{n_Q}$ project onto coordinates in S. Then $V\hat{Z}_P\hat{\Pi} = \hat{Z}_Q$. Moreover, by Proposition 2.6.30, $\hat{Z}_P = Z_P U_P$ and $\hat{Z}_Q = Z_Q U_Q$. Hence $V Z_P U_P \hat{\Pi} = V\tilde{Z}_Q U_Q$. To remove dependence on V, we need to argue that each Q-cluster has a reprensentatve in S.

Let E be the event that at least one Q-cluster has no representative in S. For a fixed $j \in [k_Q]$, cluster j has no representative in S with probability $\leq \left(1 - \frac{n_{\min}^{(Q)}}{n_Q}\right)^{n_Q}$. A union bound implies that

$$\mathbb{P}[E] \leq k_Q\left(1 - \frac{n_{\min}^{(Q)}}{n_Q}\right)^{n_Q} \leq k_Q \exp(-n_{\min}^{(Q)}) \leq O(n_Q^{-1}).$$

The last inequality holds because the condition of Theorem 2.6.29 implies that $n_{\min}^{(Q)} \geq \Omega(\sqrt{n_Q})$ and $k_Q \leq \frac{n_Q}{n_{\min}^{(Q)}}$.

Finally, we proceed by conditioning on $\neg E$. Since $\hat{Z}_P = Z_P U_P$, we know that for all $i \in S$, the unique $j_P \in [k_P]$ such that row i, column j_P of Z_P is nonzero contains its true P-community up to U_P. Similarly since $\hat{Z}_Q = Z_Q U_Q$, the the unique $j_Q \in [k_Q]$ such that row i, column j_Q of Z_P is nonzero contains its true Q-community up to U_Q. Therefore the nodes in community j_P in P are in community j_Q in Q. So, up to permutations U_P and U_Q, we have $\Pi_{j_P, j_Q} = 1$. Since we condition on $\neg E$, each cluster of Q has at least one representative in S, so each columns of Π is nonzero. We conclude that $Z_P U_P \hat{\Pi} = \tilde{Z}_Q U_Q$ with probability at least $1 - O(n_Q^{-1})$. □

We are ready to give the overall error of Proposition 2.

Proposition 2.6.32. *Suppose that $\hat{Z}_P = Z_P, \hat{\Pi} = \Pi$ in Algorithm 2. Then with probability $\geq 1 - O(\frac{1}{n_Q})$, Algorithm 2 returns a $\hat{Q} \in [0,1]^{n \times n}$ such that*

$$\frac{1}{n^2} \|\hat{Q} - Q\|_F^2 \lesssim \frac{k_Q^2 \log(n_{\min}^{(Q)})}{n_Q^2}.$$

Proof. By Proposition 2.6.31, with probability $\geq 1 - O(\frac{1}{n_Q})$, we have $\hat{Z}_P = Z_P U_P$, $\hat{Z}_Q = Z_Q U_Q$, and $\tilde{Z}_Q U_Q = Z_P U_P \hat{\Pi}$. We proceed by conditioning on these events.

Next, let $W_Q \in \mathbb{R}^{k_Q \times k_Q}$ be the population version of $\hat{W}_Q$ with $W_{Q;ii} = (\mathbf{1}^T Z_Q \boldsymbol{e}_i)^{-1}$. Then since $\hat{Z}_Q = Z_Q U_Q$ we have $\hat{W}_Q = U_Q^T W_Q U_Q$. Hence

$$\begin{aligned}\hat{Q} &= (Z_P U_P \hat{\Pi})(U_Q^T W_Q U_Q^T)(Z_Q U_Q)^T A_Q (Z_Q U_Q)(U_Q^T W_Q U_Q)(Z_P U_P \hat{\Pi})^T \\ &= \tilde{Z}_Q (W_Q Z_Q^T A_Q Z_Q W_Q) \tilde{Z}_Q^T.\end{aligned}$$

Next, let $z_Q : [n] \to [k_Q]$ be the ground truth clustering map given by $\tilde{Z}_Q \in \{0,1\}^{n \times k_Q}$. Let B_Q be defined analogously to $\hat{B}_Q$ in Algorithm 2, but using $W_Q, Z_Q, \mathbb{E}[A_Q]$ in place of $\hat{W}_Q, \hat{Z}_Q, A_Q$. Let $m_i := W_{Q;ii}^{-1}$ be the the number of nodes in S belong to community i, and let $n_i$be the the number of nodes in $[n]$ belonging to community i of Q. Then the error of

Algorithm 2 is then

$$\frac{1}{n^2}\|\tilde{Z}_Q(\hat{B}_Q - B_Q)\tilde{Z}_Q^T\|_F^2 = \frac{1}{n^2}\Bigg(\sum_{i,j\in[k_Q]} n_i n_j \Bigg(\sum_{\substack{r\in z_Q^{-1}(\{i\})\cap S\\ s\in z_Q^{-1}(\{j\})\cap S}} \frac{B_{Q;ij} - A_{Q;rs}}{m_i m_j}\Bigg)^2\Bigg)$$
$$= \frac{1}{n^2}\sum_{i,j\in[k_Q]} \frac{n_i n_j}{m_i^2 m_j^2}\Bigg(\sum_{\substack{r\in z_Q^{-1}(\{i\})\cap S\\ s\in z_Q^{-1}(\{j\})\cap S}} B_{Q;ij} - A_{Q;rs}\Bigg)^2.$$

Next, fix $i, j \in [k_Q]$ and let

$$X_{ij} = \sum_{\substack{r\in z_Q^{-1}(\{i\})\cap S\\ s\in z_Q^{-1}(\{j\})\cap S}} B_{Q;ij} - A_{Q;rs}.$$

If we condition on the clusterings of P, Q being correct then $\mathbb{E}[B_{Q;ij} - A_{Q;rs}] = 0$. Therefore by Hoeffding's inequality,

$$\mathbb{P}(X_{ij} \geq t^2) \leq 2\exp\Bigg(-\frac{2t^2}{m_i m_j}\Bigg).$$

Setting $t^2 = 10\log(m_i m_j) m_i m_j$ implies that with probability at least $1 - k_Q^2 \min_i (m_i)^{-20}$, that the overall error is

$$\frac{1}{n^2}\|\hat{Q} - Q\|_F^2 \leq \frac{1}{n^2}\sum_{i,j\in[k_Q]} \frac{10\log(m_i m_j) n_i n_j}{m_i m_j}.$$

Finally, note that there exists a constant $c_0 > 0$ such that for all $i \in [k_Q]$, $m_i \geq c_0\sqrt{n_Q}$ and $n_i \geq c_0\sqrt{n}$, by assumption. Note that each m_i is a random quantity depending on the choice of $S \subset [n]$ such that $\mathbb{E}[m_i] = \frac{n_Q}{n} n_i$. Hoeffding's inequality and a union bound over all $i \in [k_Q]$ imply that that with probability at least $\geq 1 - O(n_Q^{-8})$ that $m_i \geq \mathbb{E}[m_i] - 10\sqrt{\log n_Q} \geq \Omega(\mathbb{E}[m_i])$. We conclude that

$$\frac{1}{n^2}\|\hat{Q} - Q\|_F^2 \leq O\Bigg(\frac{1}{n_Q^2}\sum_{i,j\in[k_Q]} 10\log(m_i m_j)\Bigg)$$
$$\leq O\Bigg(\frac{k_Q^2\log(n_{\min}^{(Q)})}{n_Q}\Bigg). \qquad \square$$

2.7 Additional Experiments

2.7.1 Ablation Experiments

In this section, we discuss additional experiments that quantify the dependence of our algorithms on all relevant parameters. Our experiments also include a new baseline adapted from the estimator of Levin et al. (2022).

Description of New Baseline. Levin et al. (2022) assumes that full edge data from both P and Q are observed, and $P = Q$. Since this is not true for us, we instead compute the following modified MLE based on their estimator from Section 3.3 of Levin et al. (2022).

$$\tilde{Q}_{ij} = \begin{cases} \frac{w_P}{w_P+w_Q} A_{P;ij} + \frac{w_Q}{w_P+w_Q} A_{Q;ij} & \text{if } i,j \in S, \\ A_{P;ij} & \text{otherwise.} \end{cases}$$

Here w_P, w_Q are computed as in their paper, based on estimated sub-gamma parameters of the noise for A_P, A_Q. Akin to their adjacency spectral embedding, which assumes known rank of Q, we use Universal Singular Value Thresholding to obtain $\hat{Q}$ from $\tilde{Q}$ Chatterjee (2015b).

Oracle with $p = 0.0$. In addition to testing the new baseline from Levin et al. (2022), we also test the Oracle baseline with $p = 0.0$. As noted in Section 2.4, this corresponds to the non-transfer setting where all edges from the target graph Q are observed. Note that in this case, the value of n_Q does not matter because edges incident to nodes outside of S never get flipped. The Oracle error for β-smooth graphons on d-dimensional latent variables will therefore be $O(n^{-\frac{2\beta}{2\beta+d}})$ Xu (2018), which is less than the error bound of Theorem 2.2.3. Indeed, we will find that the Oracle our transfer algorithms in the regimes where its theoretical upper bound is better than our theoretical upper bounds.

Next, we describe the experimental results.

Figure 2.3 tests Algorithm 1 for general latent variable models. The error (Theorem 2.2.3) depends on the smoothness β of the target graph, the number of observed target nodes n_Q, and the dimension of the latent variables d.

Figure 2.4 tests Algorithm 2 for Stochastic Block Models. The error (Proposition 2.3.4) depends on the number of communities k_Q in the target graph, and the number of observed target nodes n_Q. Note that Proposition 2.3.4 also depends logarithmically on the minimum community size of Q, but this is less significant.

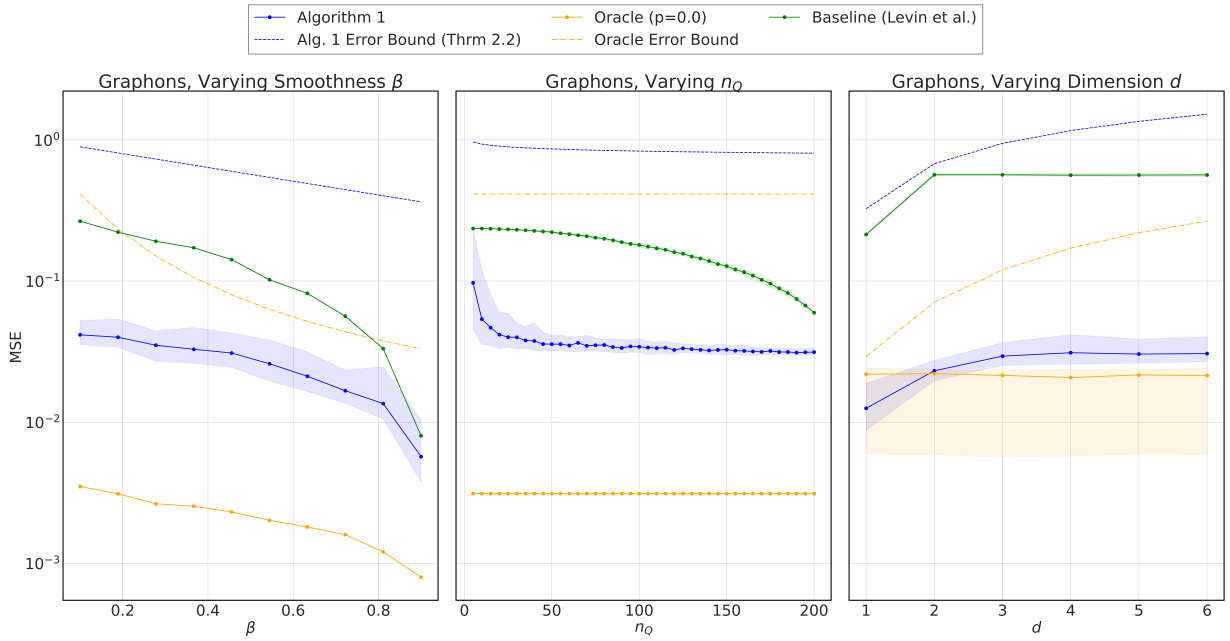


Figure 2.3: Testing parameters of Algorithm 1 (Transfer for Latent Variable Models). For most parameter settings, our method is better than the baseline and worse than the Oracle. **Left**: Testing Hölder -smoothness of f_Q with $n = 200, n_Q = 25, d = 1$. All methods improve as $\beta \to 1$. Here $f_P(x, y) = \frac{x^\alpha + y^\alpha}{2}$, $f_Q(x, y) = \frac{x^\beta + y^\beta}{2}$ with $\alpha = 0.01$ and β varying. **Middle**: Testing number of observed target nodes n_Q with $n = 200, d = 1$. Here $f_P(x, y) = \frac{x^\alpha + y^\alpha}{2}$, $f_Q(x, y) = \frac{x^\beta + y^\beta}{2}$ with $\alpha = 0.01, \beta = 0.1$. Note that the oracle does not depend on n_Q because it observes the full adjacency matrix $A_Q \in \{0, 1\}^{n \times n}$. **Right**: Testing dimension d of latent positions $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n \in [0, 1]^d$ (i.i.d. Lebesgue) with $n = 200, n_Q = 25$. Here $f_P(\boldsymbol{x}, \boldsymbol{y}) = \exp(-6\|\boldsymbol{x} - \boldsymbol{y}\|_2)$ and $f_Q(\boldsymbol{x}, \boldsymbol{y}) = \exp(-|x_1 - y_1|)$. Points are the median MSE across 50 trials, with with $[5, 95]$ percentile outcomes shaded.

Note that while we can plot theoretical guarantees for the mean squared error $\frac{1}{n^2}||\hat{Q} - Q||_F^2$ of both our algorithms' $\hat{Q}$ and the oracle's $\hat{Q}$, Levin et al. (2022) only give theoretical guarantees on the spectral norm $||\hat{Q} - Q||_2$ for their estimator $\hat{Q}$. Analyzing the stronger metric of mean-squared error would require different techniques than their paper.

2.7.2 Link Prediction Experiments

In this section, we present additional link prediction experiments on the EMAIL-EU and BIGG MODELS datasets. Unlike Section 2.4, we tune the sparsity estimate $\hat{\rho} \in (0, 1)$ used in the Universal Singular Value Thresholding step of the Oracle baseline. In particular, we set $\hat{\rho} \in (0, 1)$ to be the mean of the entries of the ground truth target matrix $Q \in [0, 1]^{n \times n}$. Note that this value is inaccessible to other algorithms since it requires knowing all the edges

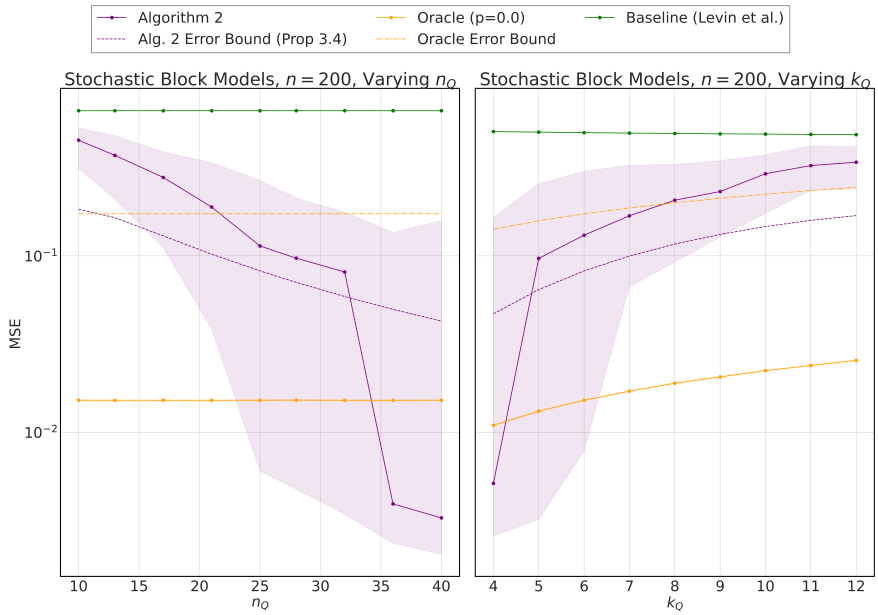


Figure 2.4: Testing parameters of Algorithm 2 (Transfer for SBMs). For most parameter settings, our method is better than the baseline and worse than the Oracle.
Left: $n = 200, k_P = 12, k_Q = 6$. Note that the oracle does not depend on n_Q because it observes the full adjacency matrix $A_Q \in \{0,1\}^{n \times n}$.
Right: $n = 200, n_Q = 25, k_P = 2k_Q$.
For both experiments, the intra-community edge probabilities are $0.2, 0.9$ for P, Q respectively, while the inter-community edge probabilities are $0.1, 0.8$ respectively. Points are the median MSE across 50 trials, with with $[5, 95]$ percentile outcomes shaded.

of Q.

Figures 2.7 and 2.8 show the performance of our Algorithms on the EMAIL-EU dataset, and Figures 2.5 and 2.6 for the BIGG MODELS dataset. As in the mean-squared error setting (Figure 2.2), we find that Algorithm 1 outperforms Algorithm 2, and that the Oracle baseline outperforms both for small p. Moreover, we find that the choice of source & target affects the performance of both of our algorithms. Hence Figure 2.7 shows better performance than Figure 2.8 for the same source but different targets, and Figure 2.5 shows better performance than Figure 2.6 for the same target but different sources.

2.8 Experimental Details

In this section, we give further details on the experiments of Section 2.4.

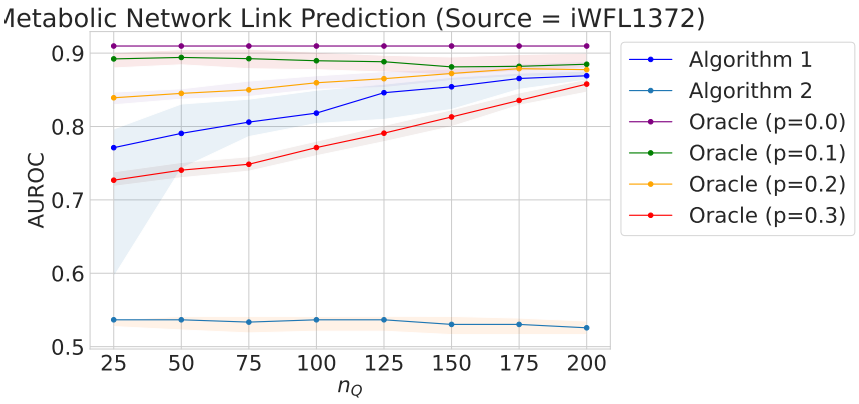


Figure 2.5: Link prediction results with the metabolic network of BiGG model iWFL1372 (*Escherichia coli W*) as the source and iJN1463 (*Pseudomonas putida*) the target. Shaded regions denote $[5, 95]$ percentile outcomes from 50 independent trials.

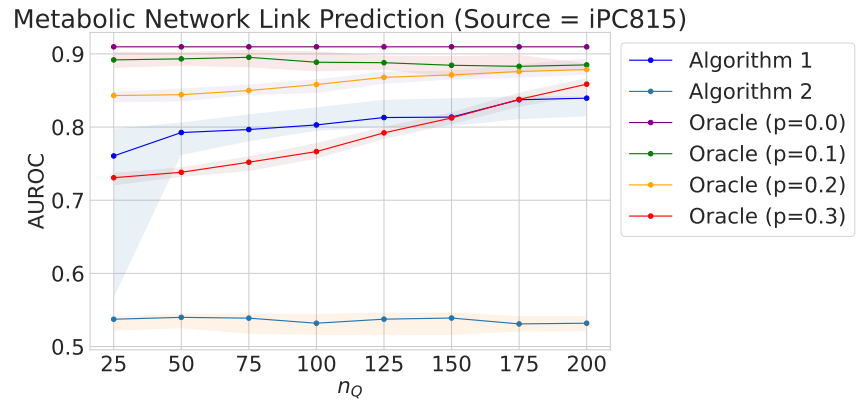


Figure 2.6: Link prediction results with the metabolic network of BiGG model iPC815 (*Yersinia pestis*) as the source and iJN1463 (*Pseudomonas putida*) the target. Shaded regions denote $[5, 95]$ percentile outcomes from 50 independent trials.

Compute Environment. We run all experiments on a personal Linux machine with 378GB of CPU/RAM. The total compute time across all results in the paper was less than 2 hours.

Functions for Figure 2.1. For the top row, the source is an $(n, 4)$-SBM with 0.8 on the diagonal and 0.2 on the off-diagonal of $B \in \mathbb{R}^{4\times4}$. The target is an $(n, 2)$-SBM with 0.9 on the diagonal and 0.1 on the off-diagonal of $B \in \mathbb{R}^{2\times2}$.

For the second and third rows, the source function is $Q(x, y) = \frac{1+\sin(\pi(1+3(x+y-1)))}{2}$ (modified from Zhang et al. (2017)). The sources are $P(x, y) = 1 - Q(x, y)$ and $P(x, y) =$

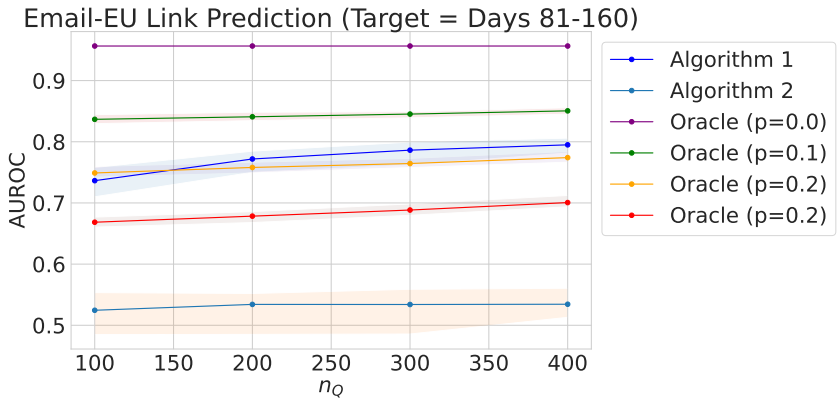


Figure 2.7: Link prediction results with Days 1-80 of EMAIL-EU as the source, and Days 81-160 as target. Shaded regions denote $[5, 95]$ percentile outcomes from 50 independent trials.

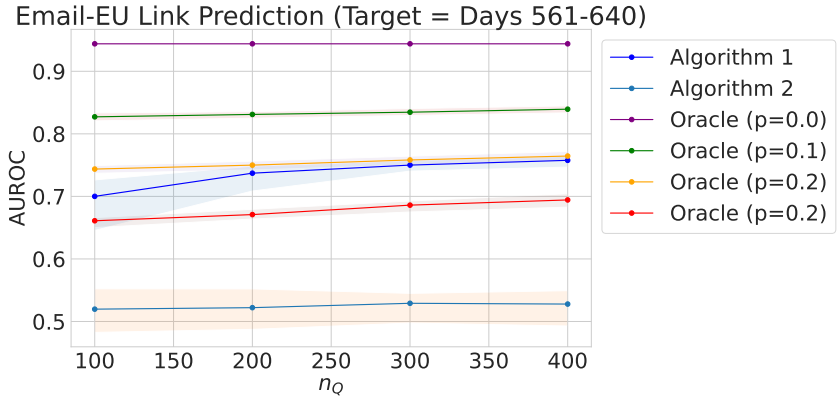


Figure 2.8: Link prediction results with Days 1-80 of EMAIL-EU as the source, and Days 561-640 as target. Shaded regions denote $[5, 95]$ percentile outcomes from 50 independent trials.

$Q(\phi(x), y)$, where $\phi(x) = 0.5 + |x - 0.5|$ if $x < 0.5$, and $0.5 - |x - 0.5|$ otherwise.

Metabolic Networks. We access metabolic models from King et al. (2016) at `http://bigg.ucsd.edu`. To construct a reasonable set of shared metabolites across the networks, we take the intersection of the node sets for the following BiGG models: iCHOv1, IJN1463, iMM1415, iPC815, iRC1080, iSDY1059, iSFxv1172, iYL1228, iYS1720, and Recon3D. We obtain a set of $n = 251$ metabolites that are present in all of the listed models.

The resulting networks are undirected, unweighted graphs on 251 nodes. We construct the matrix $A_P \in \{0, 1\}^{n \times n}$ for species P by setting $A_{P;uv} = 1$ if and only if u and v co-occur

in a metabolic reaction in the BiGG model for P.

Email-EU. We use the "email-EU-core-temporal" dataset at `https://snap.stanford.edu/data/email-Eu-core-temporal.html`, as introduced in Paranjape et al. (2017). Note that we do not perform any node preprocessing, so we use all $n = 1005$ nodes present in the data, as opposed to Leskovec and Krevl (2014); Paranjape et al. (2017) who use only 986 nodes.

Data consist of triples (u, v, t) where u, v are anonymized individuals and $t > 0$ is a timestamp. We split the data into 10 bins based on equally spaced timestamp percentiles. For simplicity we refer to these time periods as consisting of 80 days each in Section 2.4, but technically there are 803 days total. The network at time period ℓ consists of an unweighted undirected graph with adjacency matrix entry $A_{uv} = 1$ if and only if (u, v, t) or (v, t, u) occurred in the data for an appropriate timestamp t.

Hyperparameters. We do not tune any hyperparameters. For Algorithm 1 we use the quantile cutoff of $h_n = \sqrt{\frac{\log n_Q}{n_Q}}$ in all experiments.

Chapter 3: Optimal Transfer Learning for Missing Not-at-Random Matrix Completion

3.1 Introduction

We study transfer learning in the context of matrix completion, a fundamental problem motivated by theory Candès and Recht (2009); Candès and Tao (2010) and practice Fernández-Val et al. (2021); Einav and Cleary (2022); Gao et al. (2022).

A major body of work studies matrix completion in the Missing Completely-at-Random (MCAR) setting Jain et al. (2013); Chatterjee (2015a); Chen et al. (2020b), where each entry is observed i.i.d. with probability p. A more general missingness pattern, known as Missing Not-at-Random (MNAR), considers an underlying *propensity matrix* p_{ij} so that the $(i,j)^{th}$ entry is observed independently with probability p_{ij} Ma and Chen (2019); Bhattacharya and Chatterjee (2022). Various MNAR models have been formulated based on missingness structures in panel data Agarwal et al. (2023b), recommender systems Jedra et al. (2023), and electronic health records Zhou et al. (2023).

Motivated by biological problems, we consider a challenging MNAR structure where most rows and columns of $\tilde{Q}$ (a noisy version of Q) are entirely missing. Specifically, we consider both the *active sampling* and *passive sampling* settings for $\tilde{Q}$. In active sampling, a practitioner can choose rows R and columns C so that entries in $R \times C$ are observed. This follows experimental design constraints in metabolite balancing experiments Christensen and Nielsen (2000), marker selection for single-cell RNA sequencing Vargo and Gilbert (2020), patient selection for companion diagnostics Huber et al. (2022), and gene expression microarrays Hu et al. (2021).

In the *passive sampling* setting, the practitioner cannot choose the experiments. We model this by sampling each row (column) with probability p_{Row} (p_{Col}). For example, microarray analysis detects RNA segments corresponding to known genes by using chemical hybridization. However, rows may be missing because of a patient sample failing to hybridize,

The content of this chapter is under review at the 42nd International Conference on Machine Learning (ICML 2025), and can be cited as Jalan et al. (2025).

and columns may be missing because of gene probe failure Hu et al. (2021). For an illustration, see Figure 3.1.

This setting is inherently difficult because there are many entries (i, j) for which row i and column j are *both* missing in $\tilde{Q}$. Clearly, even when Q is low-rank and incoherent, estimation is impossible without side information (Proposition 3.3.1). Transfer learning is *necessary* to achieve vanishing estimation error since no information about Q_{ij} is known. Hence, we consider transfer learning in a setting where one has a noisy and masked $\tilde{P}$ corresponding to a source matrix P. P and Q are related by a distribution shift in their latent singular subspaces (Definition 3.1.2), which is a common model in e.g. Genome-Wide Association Studies McGrath et al. (2024) and Electronic Health Records Zhou et al. (2023).

Contributions. Below, we list our contributions:

(i) We obtain **minimax lower bounds** for entrywise estimation error for both the active (Theorem 3.3.2) and passive sampling settings (Theorem 3.3.12).

(ii) We give a **computationally efficient** estimation framework for both sampling settings. Our procedure is **minimax optimal** for the active setting (Theorem 3.3.6). We also establish minimax optimality for the passive setting under *incoherence* assumptions (Theorem 3.3.9).

(iii) We compare the performance of our algorithm with existing algorithms on **real-world datasets** for gene expression microarrays and metabolic modeling (Section 3.4).

Setup. $P, Q \in \mathbb{R}^{m \times n}$ are the underlying source and target matrices, related by a distributional shift in their latent singular subspaces (Definition 3.1.2). We observe a noisy and possibly masked $\widetilde{P}$. The observation model of $\tilde{Q}$ depends on which setting below we consider:

(i) Active Sampling Setting. We have a budget of T_{row} rows and T_{col} columns. We select rows $i_1, \ldots, i_{T_{\text{row}}}$ and columns $j_1, \ldots, j_{T_{\text{col}}}$, possibly at random, and with repeats allowed. Let $n_{ij} \geq 0$ be the number of times *both* row i and column j are chosen. Then, we have n_{ij} independent noisy observations $\tilde{Q}_{i,j}^{(1)}, \ldots, \tilde{Q}_{i,j}^{(n_{ij})}$ such that:

$$\tilde{Q}_{i,j}^{(t)} = \begin{cases} Q_{ij} + \zeta_{i,j}^{(t)} & \text{if } n_{ij} > 0, \\ \star & \text{otherwise,} \end{cases} \tag{3.1}$$

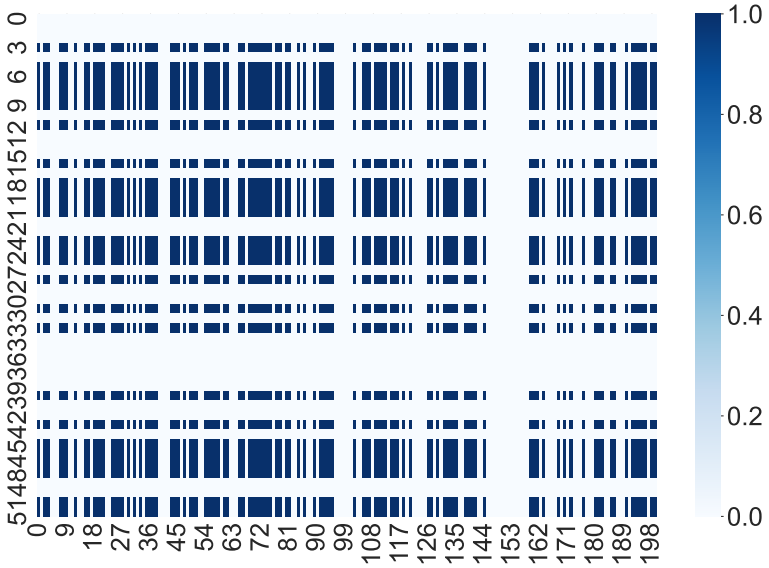


Figure 3.1: The missingness matrix for gene expression levels on Day 2 of a sepsis study Parnell et al. (2013) shows entire rows (patients) and columns (genes) as missing, due to e.g. probe-target hybridization failure of the Illumina HT-12 gene expression microarray Hu et al. (2021). We mark missing entries as 0 (white) and present entries as 1 (blue). This motivates our missingness model (Eq. (3.1) and Eq. (3.2)).

For $\zeta_{i,j}^{(t)} \overset{\text{iid}}{\sim} \mathcal{N}(0, \sigma_Q^2)$.

(ii) Passive Sampling Setting. Instead of row and column budgets, there are probabilities $p_{\text{Row}}, p_{\text{Col}} \in [0,1]$ corresponding to the random row mask $\eta_1, \ldots, \eta_m \overset{i.i.d.}{\sim} \text{Ber}(p_{\text{row}})$ and column mask $\nu_1, \ldots, \nu_n \overset{\text{iid}}{\sim} \text{Ber}(p_{\text{col}})$. Entry (i,j) of Q is noisily observed if $\eta_i = \nu_j = 1$, and missing otherwise.

$$\tilde{Q}_{ij} = \begin{cases} Q_{ij} + \zeta_{i,j} & \text{if } \eta_i = \nu_j = 1, \\ \star & \text{otherwise,} \end{cases} \tag{3.2}$$

where $\zeta_{i,j} \overset{\text{iid}}{\sim} \mathcal{N}(0, \sigma_Q^2)$.

3.1.1 Organization of the Chapter

We give our main theoretical findings, including lower and upper bounds for the active and passive sampling settings, in Section 3.3. Next, we compare our methods against existing algorithms on real-world and synthetic datasets in Section 3.4. Finally, we discuss related work in Section 3.2 and conclusions in Section 3.5.

3.1.2 Problem Setup

Notation. We will consider $P, Q \in \mathbb{R}^{m \times n}$ throughout. Asymptotics $O(\cdot), o(\cdot), \Omega(\cdot), \omega(\cdot)$ are with respect to $m \wedge n$ unless specified otherwise.

We first define matrix incoherence, which measures how concentrated the entries of the singular vectors are.

Definition 3.1.1 (Incoherence)**.** *Let M be an $m \times n$ matrix of rank d, and write its SVD as $M = U\Sigma V^\top$. The left (resp. right) incoherence parameter of M is defined as $\mu_U = m\|U\|_{2\to\infty}^2/d$ (resp. $\mu_V = n\|V\|_{2\to\infty}^2/d$). The incoherence parameter of M is defined as $\mu(M) := \max\{\mu_U, \mu_V\}$.*

We now formally define the distribution shift from P to Q, which generalizes the latent space rotation model Xu et al. (2013); McGrath et al. (2024).

Definition 3.1.2 (Matrix Transfer Model)**.** *In the matrix transfer model, we have source and target matrices $P, Q \in \mathbb{R}^{m \times n}$ such that:*

(i) *(Low-Rank) Let $P = U_P \Sigma_P V_P^\top$ for some $d \le m \wedge n$ where $U_P \in \mathcal{O}^{m\times d}, V_P \in \mathcal{O}^{n\times d}$, and $\Sigma_P \succeq 0$ is diagonal $d \times d$.*

(ii) *(Distribution shift) There exist $T_1, T_2, R \in \mathbb{R}^{d\times d}$ such that $Q = U_P T_1 R T_2^T V_P^T$, and $\|T_i\|_2 = O(1)$ for $i = 1, 2$.*

We will define the parameter space as:

$$\mathcal{F}_{m,n,d} = \Big\{ (P, Q) \in \mathbb{R}^{m\times n} \times \mathbb{R}^{m\times n} : P = U\Sigma_P V^T, \\ Q = UT_1 R T_2^T V^T, U \in \mathcal{O}^{m\times d}, V \in \mathcal{O}^{n\times d}, \\ T_1, T_2, R \in \mathbb{R}^{d\times d}, \Sigma_P \succeq 0 \Big\} \tag{3.3}$$

Definition 3.1.2 requires that the d-dimensional features of rows and columns lie in a shared subspace for P, Q. Consider the matrix of associations between m genetic variants (e.g. the MC1R gene) and n phenotypes (e.g. dark hair) for different populations P, Q (e.g. England and Spain) McGrath et al. (2024). The above model ensures that the latent feature vector for a genotype (resp. phenotype) in Q is a linear combination of those in P.

Note that T_1, T_2 are not necessarily rotations and can even be singular. We set $\|T_i\|_2 = O(1)$ to simplify theorem statements, but it is not required.

3.2 Related Work

We review the most relevant literature here. For additional discussion, we refer to the surveys De Handschutter et al. (2021); Jafarov (2022) for matrix completion and Zhuang et al. (2019); Kim et al. (2022) for transfer learning.

Matrix Completion. Most matrix completion algorithms require a Missing Completely at Random (MCAR) assumption Candès and Recht (2009); Chatterjee (2015a); Davenport et al. (2014); Zhong et al. (2019), where each Q_{ij} is observed with probability p independently of all others. The Missing Not-at-Random setting allows the masking probability of Q_{ij} to depend on the value of Q_{ij} itself Ma and Chen (2019); Bhattacharya and Chatterjee (2022); Jedra et al. (2023), but still assumes that entries are masked independently of one another. If masking variables are dependent, then authors assume identifiability of the matrix conditioned on the masking Agarwal et al. (2023b), or that entries in every row and column are observed Simchowitz et al. (2023b). By contrast, we study one of the simplest possible MNAR models in which entries of $\tilde{Q}$ are *not* independent and entire rows and columns can be missing. This MNAR model is motivated by biological problems Christensen and Nielsen (2000); Hu et al. (2021); Einav and Cleary (2022).

Transfer learning. Transfer learning has been well-studied in learning theory Ben-David et al. (2006); Cortes et al. (2008); Crammer et al. (2008). Recent works address various supervised learning Reeve et al. (2021); Cai and Wei (2021b); Ma et al. (2023a); Cai and Pu (2024) and unsupervised learning settings Gu et al. (2024); Ding and Ma (2024). Statistical works consider minimax rates of estimation, and computationally efficient estimators to achieve such rates Tripuraneni et al. (2020); Agarwal et al. (2023a); Cai and Wei (2021a); Ma et al. (2023a); Cody and Beling (2023); Cai and Pu (2024). In applications, transfer learning from data-rich to data-poor domains has applications in biostatistics Kshirsagar (2015); Datta et al. (2021), epidemiology Apostolopoulos and Bessiana (2020), computer vision Tzeng et al. (2017a); Neyshabur et al. (2020), language models Han et al. (2021), and other areas.

Transfer learning for matrix completion typically assumes the source P and target Q are observed in an MCAR fashion, and are related through a rotation in latent space Xu et al. (2013); McGrath et al. (2024); He et al. (2024). Rotational shift is a special case of our distribution shift model (Definition 3.1.2), which allows for any linear shift in latent space.

On the other hand, works that study transfer learning for specific classes of matrices typically assume distributional shifts that are unique to those structures, such as in latent variable networks Jalan et al. (2024b) or the log-linear word production model Zhou et al. (2023).

Optimal experimental design. Choosing a set of maximally informative experiments is a classical problem in statistics Smith (1918); Pukelsheim (2006) with connections to active learning Dasgupta (2011), bandits Abbasi-Yadkori et al. (2011), and reinforcement learning Lattimore et al. (2020). Optimal designs have been studied for domain adaptation Rai et al. (2010); Xie et al. (2022), misspecified regression Lattimore et al. (2020), and linear Markov Decision Processes Jedra et al. (2023). In our active sampling setting, we *jointly* query rows and columns to observe the corresponding submatrix of $\tilde{Q}$, rather than one entry at a time Chakraborty et al. (2013); Ruchansky et al. (2015); Bhargava et al. (2017). But, the optimal row queries depend on column queries (and vice versa) – so we use the tensorization property of G-optimal designs (Proposition 3.3.4) to prove global optimality with respect to joint row/column samplers.

3.3 Main Findings

We first show that without transfer – side information from the source data P – completing the target matrix Q is impossible. To this end, we present a minimax lower bound on the expected prediction error. First, we define the parameter space of matrices with bounded incoherence:

$$\mathcal{T}_{mn}^{(d)} = \Big\{ Q \in \mathbb{R}^{m \times n} : \operatorname{rank}(Q) \le d, \\ \mu(Q) \le O\big(\log(m \vee n)\big) \Big\}. \tag{3.4}$$

Proposition 3.3.1 (Minimax Error of MNAR Matrix Completion Without Transfer)**.** *Let $m, n \ge 1$ and $d \le m \wedge n$. Let $\Psi = (Q, \sigma, p_{\text{Row}}, p_{\text{Col}})$ where $Q \in \mathcal{T}_{mn}^{(d)}$, $\sigma^2 > 0$, and $p_{\text{Row}}, p_{\text{Col}} \in [0, 1]$. Let $\mathbb{P}_\Psi$ denote the law of the random matrix $\tilde{Q}$ defined as in Eq. (3.2) with $\sigma_Q = \sigma$, and denote the expectation under this law as $\mathbb{E}_\Psi$. The minimax rate of estimation is:*

$$\inf_{\hat{Q}} \sup_{Q \in \mathcal{T}_{mn}^{(d)}} \inf_{\substack{p_{\text{Row}} \le .99 \\ p_{\text{Col}} \le .99}} \mathbb{E}_\Psi \left[\frac{1}{mn} \|Q - \hat{Q}\|_F^2 \right] \ge \Omega(d\sigma^2).$$

An immediate consequence of the above proposition is that the minimax rate for max squared error $\|\hat{Q} - Q\|_{\max}^2$ is also $\Omega(d\sigma^2)$. We see that in both error metrics, vanishing estimation error is impossible without transfer learning.

3.3.1 Lower Bound for Active Sampling Setting

We now give a minimax lower bound for Q estimation in the active sampling setting.

Theorem 3.3.2 (Minimax Lower Bound for Q-estimation with Active Sampling)**.** *Fix m, n and $2 \le d \le m \wedge n$. Fix $\sigma^2 > 0$ and let $|\Omega| = T_{\text{row}} \cdot T_{\text{col}}$.*

Let $\mathbb{P}_{P,Q,\sigma^2}$ be the distribution of $(\tilde{P}, \tilde{Q})$ where $\tilde{P} := P$ and $\tilde{Q} := Q + G$ where $G_{ij} \overset{\text{iid}}{\sim} N(0, \sigma^2)$.

Let $\mathcal{Q}$ be the **class of estimators** *which observe $\tilde{P}$, and choose row and column samples according to the budgets $T_{\text{row}}, T_{\text{col}}$ as in Eq.* (3.1)*, and then return some estimator $\hat{Q} \in \mathbb{R}^{m \times n}$. Then, there exists absolute constant $C > 0$ such that minimax rate of estimation is:*

$$\inf_{\hat{Q} \in \mathcal{Q}} \sup_{(P,Q) \in \mathcal{F}_{m,n,d}} \mathbb{E}_{\mathbb{P}_{P,Q,\sigma^2}}[\|\hat{Q} - Q\|_{\max}^2] \ge \frac{Cd^2\sigma^2}{|\Omega|}.$$

We prove Theorem 3.3.2 using a generalization of Fano's method Verdú et al. (1994). We construct a family of distributions indexed by d^2 source/target pairs $(P^{(s)}, Q^{(s)})_{s=1}^{d^2}$. The source P is the same for all s, while each pair of target matrices $Q^{(s)}, Q^{(s')}$ differs in at most 2 entries. For example, say entries $(5, 6)$ and $(8, 7)$ are different between $Q^{(1)}$ and $Q^{(2)}$. Regardless of the choice of row/column samples, the *average* KL divergence of a pair of targets is small. If e.g. the entries $(5, 6), (8, 7)$ are heavily sampled, then the estimator can distinguish $Q^{(1)}, Q^{(2)}$ well, but cannot distinguish $Q^{(t)}, Q^{(t')}$ for all t, t' pairs that are equal on $(5, 6)$ and $(8, 7)$.

3.3.2 Estimation Framework

Next, we describe our estimation framework. Given $\tilde{P}$ and $\tilde{Q}[R, C]$, where R, C can come from either the active (Eq. (3.1)) or passive sampling (Eq. (3.2)) setting, we estimate $\hat{Q}$ via the least-squares estimator.

Least Squares Estimator.

1. Extract features via SVD from $\tilde{P} = \widehat{U}_P \widehat{\Sigma}_P \widehat{V}_P^T$.

2. Let Ω be the multiset of observed entries. Then solve

$$\hat{\Theta}_Q := \arg\min_{\Theta \in \mathbb{R}^{d \times d}} \sum_{(i,j) \in \Omega} |\tilde{Q}_{ij} - \hat{\boldsymbol{u}}_i^\top \Theta \hat{\boldsymbol{v}}_j|^2, \tag{3.5}$$

   where $\hat{\boldsymbol{u}}_i := \hat{U}_P^T \boldsymbol{e}_i, \hat{\boldsymbol{v}}_j := \hat{V}_P^T \boldsymbol{e}_j$.

3. Estimate $\hat{Q}$:

$$\hat{Q}_{ij} = \hat{\boldsymbol{u}}_i^\top \hat{\Theta}_Q \hat{\boldsymbol{v}}_j. \tag{3.6}$$

This fully specifies $\hat{Q}$ in the passive sampling setting (Eq. (3.2)). For the active sampling setting, we must also specify how rows and columns are chosen.

Active sampling poses two main challenges. First, it is not clear how to leverage $\tilde{P}$ for sampling $\tilde{Q}$ because samples are chosen *before* observing $\tilde{Q}$, so the distribution shift from P to Q is unknown. Second, the best design depends on the choice of estimator and vice versa.

Surprisingly, we show that for the right choice of experimental design, the optimal estimator is precisely the least-squares estimator $\hat{Q}$ as in Eq. (3.6). We use the classical G-optimal design Pukelsheim (2006), which has been used in reinforcement learning to achieve minimax optimal exploration Lattimore and Szepesvári (2020b) and optimal policies for linear Markov Decision Processes Taupin et al. (2023).

Definition 3.3.3 (ϵ-approximate G-optimal design)**.** *Let $\mathcal{A} \subset \mathbb{R}^d$ be a finite set. For a distribution $\pi : \mathcal{A} \to [0, 1]$, its G-value is defined as*

$$g(\pi) := \max_{\boldsymbol{a} \in \mathcal{A}} \left[\boldsymbol{a}^T \left(\sum_{\boldsymbol{a} \in \mathcal{A}} \pi(\boldsymbol{a}) \boldsymbol{a}\boldsymbol{a}^T \right)^{-1} \boldsymbol{a} \right].$$

For $\epsilon > 0$, we say $\hat{\pi}$ is ϵ-approximately G-optimal if

$$g(\hat{\pi}) \leq (1 + \epsilon) \inf_{\pi} g(\pi).$$

If $\epsilon = 0$, we say $\hat{\pi}$ is simply G-optimal.

Notice that in Eq. (3.5), the covariates are tensor products ($\hat{\boldsymbol{v}}_j \otimes \hat{\boldsymbol{u}}_i$) of column and row features. The G-optimal design is useful because it respects the tensor structure of the least-squares estimator. We prove this via the Kiefer-Wolfowitz Theorem Lattimore and Szepesvári (2020b).

Proposition 3.3.4 (Tensorization of G-optimal design)**.** *Let $U \in \mathbb{R}^{m\times d_1}, V \in \mathbb{R}^{n\times d_2}$. Let ρ be a G-optimal design for $\{U^T\boldsymbol{e}_i : i \in [m]\}$ and ζ be a G-optimal design for $\{V^T\boldsymbol{e}_j : j \in [n]\}$. Let $\pi(i,j) = \rho(i)\zeta(j)$ be a distribution on $[m]\times[n]$. Then π is a G-optimal design on $\{V^T\boldsymbol{e}_j \otimes U^T\boldsymbol{e}_i : i \in [m], j \in [n]\}$.*

Consider a maximally coherent P that is nonzero at entry $(3,5)$ and zero elsewhere. Then Q is also zero outside $(3,5)$. By the Kiefer-Wolfowitz Theorem, the G-optimal design for rows (resp. columns) samples row 3 (resp. column 5) with probability 1. So, if $\tilde{P}$ is not too noisy, then the G-optimal design samples *precisely the useful rows/columns.*

In light of Proposition 3.3.4, we leverage the tensorization property to sample rows and columns as follows.

Active Sampling. Given $\hat{U}, \hat{V}$, and budget $T_{\text{row}}, T_{\text{col}}$,

1. Compute ϵ-approximate G-optimal designs $\hat{\rho}, \hat{\zeta}$ for $\{\hat{U}_P^T\boldsymbol{e}_i : i \in [m]\}$ and $\{\hat{V}_P^T\boldsymbol{e}_j : j \in [n]\}$ respectively, with the Frank-Wolfe algorithm Lattimore and Szepesvári (2020b).
2. Sample $i_1, \dots i_{T_{\text{row}}} \overset{\text{iid}}{\sim} \hat{\rho}$ and $j_1, \dots j_{T_{\text{col}}} \overset{\text{iid}}{\sim} \hat{\zeta}$.

Finally, we specify the assumption we need on the source data $\tilde{P}$, called Singular Subspace Recovery (SSR).

Assumption 3.3.5 (ϵ-SSR)**.** *Given $\tilde{P} \in (\mathbb{R}\cup\{\star\})^{m\times n}$, we have access to a method that outputs estimates $\widehat{U}_P \in \mathcal{O}^{m\times d}$ and $\widehat{V}_P \in \mathcal{O}^{n\times d}$, such that:*

$$\begin{aligned} &\inf_{W_U\in\mathcal{O}^{d\times d}} \|\hat{U} - UW_U\|_{2\to\infty} \le \epsilon_{\text{SSR}}, \\ \textit{and} \qquad &\inf_{W_V\in\mathcal{O}^{d\times d}} \|\hat{V} - VW_V\|_{2\to\infty} \le \epsilon_{\text{SSR}} \end{aligned} \tag{3.7}$$

for some $\epsilon_{\text{SSR}} > 0$.

This assumption holds for a number of models. For instance, recent works in both MCAR Chen et al. (2020b) and MNAR Agarwal et al. (2023b); Jedra et al. (2023) settings give estimation methods for $\hat{P}$ with entry-wise error bounds. In Appendix 3.6.2, we prove that these entry-wise guarantees, combined with standard theoretical assumptions such as incoherence, imply Assumption 3.3.5.

We now give our main upper bound.

Theorem 3.3.6 (Generic error bound for active sampling)**.** *Let $\hat{Q}$ be the active sampling estimator with $T_{\text{row}}, T_{\text{col}} \geq 20d\log(m+n)$. Then, for absolute constants $C, C' > 0$, and all $\epsilon < \frac{1}{10}$,*

$$\begin{aligned}\mathop{\mathbb{P}}_{\tilde{P},\tilde{Q}}\left[\|\hat{Q}-Q\|_{\max}^2 \leq C(1+\epsilon)\left(\frac{d^2\sigma_Q^2\log(m+n)}{|T_{\text{col}}||T_{\text{row}}|}+d^2\epsilon_{\text{SSR}}^2\|Q\|_2^2\right)\right]\\ \geq 1-C'(m+n)^{-2}.\end{aligned}$$

We will discuss implications of Theorem 3.3.6 in Remark 3.3.7. First, we give some intuition. Notice that Theorem 3.3.6 (and Theorem 3.3.9) gives an error bound as a sum of two terms, which depend on the sample size and ϵ_{SSR} respectively. To see why, let Ω be the set of observed entries, either in a passive or active sampling setting. Let $\hat{\boldsymbol{u}}_i, \hat{\boldsymbol{v}}_j$ be the covariates as in Eq. (3.5). The observation $\tilde{Q}_{ij}$ can be decomposed:

$$\begin{aligned}\tilde{Q}_{ij} &= Q_{ij} + (\tilde{Q}_{ij} - Q_{ij}) \\ &= \hat{\boldsymbol{u}}_i^\top \Theta_Q \hat{\boldsymbol{v}}_j + \underbrace{\epsilon_{ij}}_{\text{misspecification } \tilde{P}} + \underbrace{(\tilde{Q}_{ij} - Q_{ij})}_{\text{noise}}\end{aligned} \tag{3.8}$$

The population estimand $\Theta_Q \in \mathbb{R}^{d\times d}$, which is estimated in Eq. (3.5), is:

$$\Theta_Q := W_U^T T_1 R T_2^T W_V,$$

where T_1, T_2 are the distribution shift matrices as in Definition 3.1.2, and $W_U, W_V \in \mathcal{O}^{d\times d}$ are some rotations. The misspecification error is due to the estimation error of the singular subspaces of P and depends on ϵ_{SSR} as follows:

$$\begin{aligned}\epsilon_{ij} := &\boldsymbol{e}_i^T(\hat{U}-UW_U)\Theta_Q\hat{V}\boldsymbol{e}_j \\ &+ \boldsymbol{e}_i^T\hat{U}\Theta_Q(\hat{V}-VW_V)\boldsymbol{e}_j \\ &+ \boldsymbol{e}_i^T(\hat{U}-UW_U)\Theta_Q(\hat{V}-VW_V)\boldsymbol{e}_j\end{aligned}$$

Therefore $\epsilon_{ij}^2 = O(\epsilon_{\text{SSR}}^2 \|Q\|_2^2)$ for all i, j.[1] Notice the misspecification error is independent of the estimator $\widehat{\Theta}_Q$, so it will not depend on sample size. This explains the appearance of the two summands in our upper bounds. The first term depends on estimation error $\Theta_Q - \hat{\Theta}_Q$, which is unique to the sampling method. The second depends on misspecification, which is common to both.

Remark 3.3.7 (Minimax Optimality for MNAR and MCAR Source Data)**.** *The rate of Theorem 3.3.6 is minimax-optimal in the usual transfer learning regime when target data is noisy (σ_Q large) and limited ($|\Omega| := |T_{\text{row}}||T_{\text{col}}|$ small).*

Suppose P is rank d, μ-incoherent, with singular values $\sigma_1 \geq \cdots \geq \sigma_d$, condition number κ and $m = n$. For the MNAR $\tilde{P}$ setting, suppose each $\tilde{P}_{ij}$ has i.i.d. additive noise $\mathcal{N}(0, \sigma_P^2)$ with sampling sparsity factor $n^{-\beta}$ for $\beta \in [0, 1]$ and $\sigma_P = O(1)$. By Jedra et al. (2023), $\hat{Q}$ is minimax-optimal if

$$\frac{4\mu^3 d^3 \kappa^2 \|Q\|_2^2}{n^{1+\frac{2-\beta}{d}}} \lesssim \frac{\sigma_Q^2}{|\Omega|},$$

where $\lesssim$ ignores $\log(m+n)^{O(1)}$ factors. For the MCAR $\tilde{P}$ setting, suppose $\tilde{P}$ has additive noise $\mathcal{N}(0, \sigma_P^2)$ and observed entries i.i.d. with probability $p \gtrsim \frac{\kappa^4 \mu^2 d^2}{n}$, with $\sigma_P \sqrt{\frac{n}{p}} \lesssim \frac{\sigma_d(P)}{\sqrt{\kappa^4 \mu d}}$. Letting $|\Omega| = n^2 p_{\text{Row}} p_{\text{Col}}$, by Chen et al. (2020b), $\hat{Q}$ is minimax-optimal if

$$\frac{\mu^6 d^4 \|Q\|_2^2}{n^2} \lesssim \frac{\sigma_Q^2}{|\Omega|}.$$

While the results of Jedra et al. (2023); Chen et al. (2020b) used in Remark 3.3.7 require incoherence, recent work also gives guarantees on ϵ_{SSR} without incoherence assumptions, although in limited settings.

Remark 3.3.8 (Incoherence-free minimax optimality)**.** *Let $P \in \mathbb{R}^{n \times n}$ be rank-1 and Hermitian, and $\tilde{P} = P + W$ where W is Hermitian with i.i.d. $\mathcal{N}(0, \sigma_P^2)$ noise on the upper triangle. Under the assumptions of Yan and Levin (2024), for constant $C > 0$, $\hat{Q}$ is minimax optimal if*

$$\frac{C\sigma_P^2 (\log n)^{O(1)} \|Q\|_2^2}{\|P\|_2^2} \leq \frac{\sigma_Q^2}{|\Omega|}.$$

Taking $|\Omega| = O(\log n)$ since $d = 1$, and $\|Q\|_2 = O(\|P\|_2)$, we require

$$C\sigma_P^2 (\log n)^{O(1)} \leq \sigma_Q^2.$$

[1] In fact $\epsilon_{ij}^2 = O(\epsilon_{\text{SSR}}^2 \|R\|_2^2)$, but we report bounds with the weaker $O(\epsilon_{\text{SSR}}^2 \|Q\|_2^2)$ for ease of reading.

3.3.3 Passive Sampling

We next give the estimation error for the passive sampling setting. The rate almost exactly matches Theorem 3.3.6, but we pay an extra factor due to incoherence. This is because unlike the active sampling setting, if ℓ_2 mass of the features is highly concentrated in a few rows and columns, then the passive sample will simply miss these with constant probability. To give a high probability guarantee, we require that features cannot be too highly concentrated.

Theorem 3.3.9 (Generic Error Bound for $\hat{Q}$)**.** *Let $\hat{Q}$ be as in Eq.* (3.6) *and $C > 0$ an absolute constant. Suppose P has left/right incoherence μ_U, μ_V respectively, and $p_{\text{Row}}, p_{\text{Col}}$ are such that $\frac{p_{\text{Row}} m}{Cd \log m} \geq \mu_U + \frac{\epsilon_{\text{SSR}}^2 m}{d}$, $\frac{p_{\text{Col}} n}{Cd \log n} \geq \mu_V + \frac{\epsilon_{\text{SSR}}^2 n}{d}$. Let $\mu = \mu_U \mu_V$. Then,*

$$\begin{aligned}\mathbb{P}\left[\|\hat{Q} - Q\|_{\max}^2 \leq C\mu\left(\frac{d^2\sigma_Q^2 \log(m+n)}{p_{\text{Row}} p_{\text{Col}} mn} + d^2\epsilon_{\text{SSR}}^2 \|Q\|_2^2\right)\right] \\ \geq 1 - O((m \wedge n)^{-2}).\end{aligned}$$

If P is coherent, the sample complexity $|\Omega| \approx p_{\text{Row}} p_{\text{Col}} mn$ needed to achieve vanishing estimation error in Theorem 3.3.9 may be large. By contrast, our active sampling with G-optimal design requires only $|\Omega| \gtrsim d^2\sigma_Q^2$ (Theorem 3.3.6). This shows the advantage of active sampling, which can query the most informative rows/columns when P is coherent.

3.3.4 Lower Bound for Passive Sampling

We give a lower bound for the passive sampling setting in terms of a fixed, arbitrary mask. To exclude degenerate cases such as all entries being observed, we require the following definition.

Definition 3.3.10 (Nondegeneracy)**.** *Let $p > 0$ and $\eta_1, \ldots, \eta_m \overset{\text{iid}}{\sim} \text{Ber}(p)$. Let $D \in \{0,1\}^{m \times m}$ be diagonal with $D_{ii} = \eta_i$. We say $(\eta_i)_{i=1}^m$ is p-nondegenerate for $U \in \mathcal{O}^{n \times d}$ if $|\|DU\|_2 - \sqrt{p}| \leq \frac{\sqrt{p}}{10}$.*

The Matrix Bernstein inequality Chen et al. (2021) implies that masks are nondegenerate with high probability.

Proposition 3.3.11. *Under the conditions of Theorem 3.3.9, the event that both $(\eta_i)_{i=1}^m$ is p_{Row}-nondegenerate for $\hat{U}_P$ and that $(\nu_j)_{j=1}^n$ is p_{Col}-nondegenerate for $\hat{V}_P$ holds with probability $\geq 1 - 2(m \wedge n)^{-10}$.*

We can now state our lower bound, proved via Fano's method.

Theorem 3.3.12 (Minimax Lower Bound for Passive Sampling)**.** *Let $\mathcal{F}_{m,n,d}$ be the parameter space of Theorem 3.3.2. Let*

$$\mathcal{G}_{m,n,d} := \left\{ (P, Q) \in \mathcal{F}_{m,n,d} : P, Q \text{ are } O(1) - \text{incoherent} \right\}$$

Suppose $(\eta_i)_{i=1}^m, (\nu_j)_{j=1}^n$ are nondegenerate with respect to U, V respectively. Let $\mathbb{P}_{Q,\sigma^2,p_{\text{Row}},p_{\text{Col}}}$ be the law of the random matrix $\tilde{Q}$ generated as in Eq. (3.2) with $\sigma = \sigma_Q$.

There exists absolute constant $C > 0$ such that minimax rate of estimation is:

$$\inf_{\hat{Q}} \sup_{(P,Q) \in \mathcal{G}_{m,n,d}} \mathbb{E}_{\mathbb{P}_{(Q,\sigma^2,p_{\text{Row}},p_{\text{Col}})}} \left[\frac{1}{mn} \|\hat{Q} - Q\|_F^2 \middle| (\eta_i)_{i=1}^m, (\nu_j)_{j=1}^n \right] \geq \frac{C d^2 \sigma_Q^2}{p_{\text{Row}} p_{\text{Col}} mn}$$

We immediately obtain the same lower bound for max squared error.

We see that our error rate for passive sampling in Theorem 3.3.9 is minimax-optimal when $\mu = O(1)$, modulo bounds on ϵ_{SSR} as in Remark 3.3.7.

Unlike the lower bound for max squared error in active sampling (Theorem 3.3.2), Theorem 3.3.12 gives a lower bound for the mean-squared error, which is strictly stronger. An interesting question is whether Theorem 3.3.12 can be generalized to incoherence greater than a constant. We leave this for future work.

3.4 Experiments

In this section, we compare both our active and passive sampling estimators against existing methods on real-world and simulated datasets.

Experimental setup. We compare against two baselines from the matrix completion literature. First, we use the MNAR matrix completion method of Bhattacharya and Chatterjee (2022). We tune the method by passing in the true rank of Q as well as the rank of the mask matrix. Second, we use the transfer learning method of Levin et al. (2022). This

Table 3.1: Summary of real-world datasets. The $2 \to \infty$ norms are for U_P, V_P, U_Q, V_Q respectively. Notice these are within $[0, 1]$ always, and $2 \to \infty$ norm of 1 implies maximal coherence.

Dataset	Shape	Rank	$2 \to \infty$ Norms
Gene Expr.	31×300	4	0.55, 0.30, 0.64, 0.38
Metabolic	251×251	8	0.99, 0.99, 0.99, 0.99

method is designed for matrix completion, but in a missingness structure different from our MNAR setting. For shorthand, we will refer to these as *BC22* and *LLL22* respectively. See Section 3.7 for precise details of our implementations.

The input to each of these, as well as our passive sampling method, is the pair $\tilde{P}, \tilde{Q}$. The method of Bhattacharya and Chatterjee (2022) requires input matrices to have entries in $[-1, 1]$ so we normalize all $\tilde{P}, \tilde{Q}$ by their maximum entry in absolute value, for all methods. We also compute the active sampling estimator by fixing the budgets $T_{\text{row}} = m \cdot p_{\text{Row}}, T_{\text{col}} = n \cdot p_{\text{Col}}$ throughout.

3.4.1 Real World Experiments

In this section we study real-world datasets on gene expression microarrays in a whole-blood sepsis study Parnell et al. (2013), and weighted metabolic networks of gram-negative bacteria King et al. (2016). Table 3.1 summarizes the datasets, and Appendix 3.7 gives more details on our data preparation.

Patient Gene Expression Matrices. The matrices P, Q represent the gene expression for patients in a sepsis study Parnell et al. (2013). Here $P, Q \in \mathbb{R}^{31 \times 300}$ where P_{ij} measures the expression level of gene j in patient i on day 1 of the study, and Q corresponds to day 2 of the study.

Figure 3.2 displays the maximum squared error for a range of masking probabilities on $\tilde{Q}$. We see that both active and passive sampling perform well even at small sample sizes, while the transfer baseline method Levin et al. (2022) achieves a worse but nontrivial maximum error.

Notably, active sampling is no better than passive sampling here. This makes sense because P, Q are relatively incoherent (Table 3.1), so our theoretical guarantees are the same.

In fact, active sampling displays higher variation in error, due to the variability in random sampling from the G-optimal design. It is known that the G-optimal design for any $\mathcal{A} \subset \mathbb{R}^d$ has support size $O(d^2)$ Lattimore and Szepesvári (2020a), so the sampled set of rows and columns will vary somewhat from one experiment to the next.

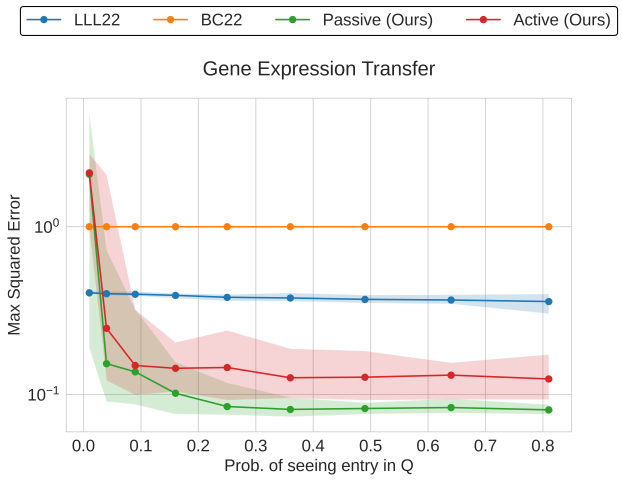


Figure 3.2: Max-squared error of $\hat{Q} - Q$. Here, $\tilde{Q}$ has $p_{\text{Row}} = p_{\text{Col}}$ varying along the x-axis, which displays p^2_{Row}. We set $\sigma_Q = 0.1$, and P is fully observed. For each method, we show the median of the errors across 50 independent runs, as well as the $[10, 90]$ percentile.

Weighted Metabolic Network Adjacency Matrices. We collect weighted metabolic networks from the BiGG Genome Scale Metabolic Models repository King et al. (2016), consistent with recent work on transfer learning for network estimation Jalan et al. (2024b). Specifically, $P, Q \in \mathbb{R}^{251 \times 251}$ where $P_{ij} \geq 0$ counts the number of co-occurrences of metabolites i and j in a reaction for organism P. Q_{ij} represents the same quantity in a different organism Q. We use the gram-negative bacteria *E. coli W* and *P. putida* for P, Q respectively. Unlike Jalan et al. (2024b), we do not need to truncate the adjacency matrices to $\{0, 1\}$, allowing us to handle edge weights. This makes a difference, because without truncation the edge weights distribution is highly skewed for both P, Q (see Section 3.7).

Figure 3.3 shows max squared error for a range of masking probabilities on $\tilde{Q}$. We see that active sampling does well, while passive sampling is very poor (note however, that passive sampling does relatively well for mean-squared error - Figure 3.12). This is because P, Q are almost maximally coherent (Table 3.1), so the assumptions of our guarantee for passive sampling (Theorem 3.3.9) do not hold. By contrast, active sampling performs well even in this highly coherent setting.

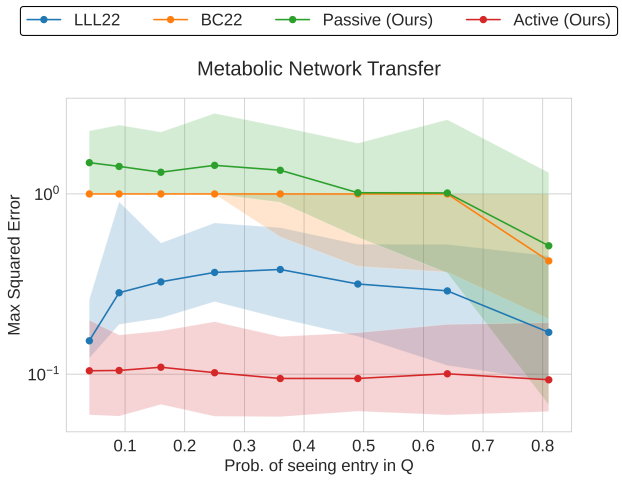


Figure 3.3: Max-squared error of $\hat{Q} - Q$, with the same experimental parameters as Figure 3.2.

3.4.2 Simulations

In this section, we further probe the effects of incoherence by testing on two highly coherent synthetic datasets (described below). Table 3.2 displays our results, with $p_{\text{Row}} = p_{\text{Col}} = 0.1, \sigma_Q = 0.1$, and P fully observed. Note that $0.1 \approx \frac{2d \log n}{n}$ here, so $p_{\text{Row}}, p_{\text{Col}}$ are near the theoretical limit of our guarantees even for incoherent matrices.

Each table entry shows $\hat{\mu} \pm 2\hat{\sigma}$ for mean-squared error across 50 independent trials. We find that for a stylized example of maximally coherent P, Q, active sampling is much better than all other methods. However, for less stylized P, Q that are still not incoherent, active and passive sampling are comparable, and outperform both baselines.

Stylized Coherent Model. For $n = 200, d = 5$ we generate $U_P, V_P \in \{0, 1\}^{n \times d}$ via $(U_P)_{ii} = 1$, $(V_P)_{(n-i),i} = 1.0$ and the other entries zero. We sample the diagonal entries of $\Sigma_P, \Sigma_Q \in \mathbb{R}^{d \times d}$ iid uniformly at random from $[0.5, 1]$. Then $P = U_P \Sigma_P V_P^T$ and $Q = U_P \Sigma_Q V_P^T$. We call this class "Coherent."

Matrix Partition Model. For a less stylized class, let $m = 300, n = 200, d = 5, a = 0.1, b = 0.8$. We generate partitions $U_P \in \{0, 1\}^{m \times d}, V_P \in \{0, 1\}^{n \times d}$ where each row is uniformly at random from $\{\boldsymbol{e}_1, \ldots, \boldsymbol{e}_d\}$. Then, $B_P \in [0, 1]^{d \times d}$ is generated by sampling $C \in [0, 1]^{d \times d}$ with $C_{ij} \overset{\text{iid}}{\sim} \text{Unif}([0, b])$ and $(B_P)_{ij} = C_{ij} + \mathbf{1}_{i=j} a$. Finally, we sample permutations $\Pi_1, \Pi_2 \in \{0, 1\}^{d \times d}$ uniformly at random from all such permutations. Then, $P = U_P B_P V_P^T$ and $Q = U_P \Pi_1 B_P \Pi_2^T V_P^T$. We call this class "Matrix Partition Model" in analogy with the Planted Partition Model Abbe (2017). Spectral arguments show that such matrices are

somewhat coherent Lee et al. (2014a), although not maximally so.

Table 3.2: Comparison of the errors of different approaches on synthetic data.

	Coherent	Partition
Passive (Ours)	$0.084 \pm 0.039 \times 10^{-3}$	$\mathbf{0.040 \pm 0.090}$
Active (Ours)	$\mathbf{0.009 \pm 0.015 \times 10^{-3}}$	0.046 ± 0.074
LLL22	$0.061 \pm 0.037 \times 10^{-3}$	0.134 ± 0.011
BC22	$0.789 \pm 0.644 \times 10^{-3}$	0.305 ± 0.002

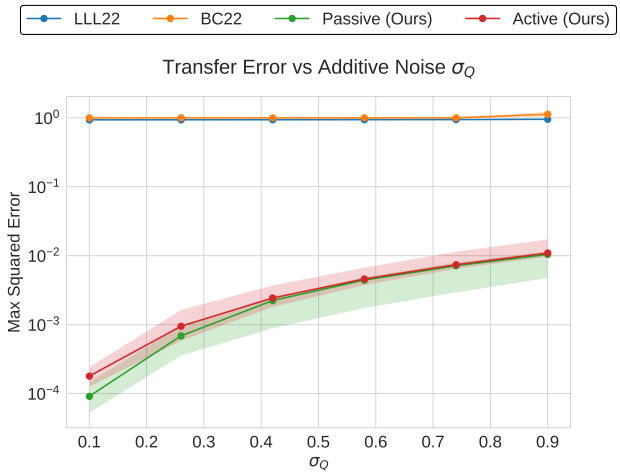


Figure 3.4: Ablation study for the effect of additive target noise in the Matrix Partition Model. For each method, we display the median max-squared error across 10 independent runs, as well as the $[10, 90]$ percentile.

3.4.3 Ablation Studies

Our main focus is to understand how sample budgets $T_{\text{row}}, T_{\text{col}}$, or probabilities $p_{\text{Row}}, p_{\text{Col}}$ affects the estimation error for transfer learning. We also perform ablation studies to test the effect of other model parameters, such as rank, dimension, noise variance, etc. Figure 3.4 shows the effect of target noise variance on maximum error in the Matrix Partition Model with $m = 300, n = 200, d = 5, a = 0.1, b = 0.8, p_{\text{Row}} = 0.5, p_{\text{Col}} = 0.5$. We defer our additional ablation studies to Section 3.7.

3.5 Conclusion and Future Work

We study transfer learning for a challenging MNAR model of matrix completion. We obtain minimax lower bounds for entrywise estimation of Q in both the active (Theorem 3.3.2) and passive sampling settings (Theorem 3.3.12). We give a computationally efficient minimax-optimal estimator that uses tensorization of G-optimal designs in the active setting (Theorem 3.3.6). Further, in the passive setting, we give a rate-optimal estimator under incoherence assumptions (Theorem 3.3.9). Finally, we experimentally validate our findings on data from gene expression micoarrays and metabolic modeling.

Future work could consider even more difficult missingness structures, such as when the masks $(\eta_i)_{i=1}^m, (\nu_j)_{j=1}^n$ are dependent. If the mask can be partitioned into subsets whose mutual dependencies are small, an Efron-Stein argument Paulin et al. (2016) may work. Is bounded dependence necessary? Moreover, one can consider other kinds of side information, such as gene-level features in Genome-Wide Association Studies McGrath et al. (2024). Finally, there can be other interesting nonlinear models for transfer between source and target matrices.

3.6 Proofs and Additional Results

3.6.1 Preliminaries

We will repeatedly make use of the vectorization operator.

Definition 3.6.1 (Vectorization)**.** *For $X \in \mathbb{R}^{n\times d}$, the vectorization $\mathrm{vec}(X) \in \mathbb{R}^{nd}$ is the vector whose first n entries correspond to the first column of X, and next n entries correspond to the second column of X, and so on.*

We can vectorize matrix products as follows.

Lemma 3.6.2 (Horn and Johnson (2012))**.** *Let A, B, X be matrices of shapes such that AXB is well-defined. Then:*

$$\mathrm{vec}(AXB) = (B^T \otimes A)\mathrm{vec}(X).$$

3.6.2 From Entrywise Guarantees to SSR

We prove that Assumption 3.3.5 follows from entrywise estimation guarantees on the source.

Proposition 3.6.3. *Let P an $m \times n$ matrix of rank r. Let $\epsilon > 0$, and $\widehat{P}$ be a rank-r estimate of P, satisfying*

$$\|\widehat{P} - P\|_{\max} \leq \epsilon \|P\|_{\max}. \tag{3.9}$$

Consider the SVDs $P = U\Sigma V^\top$, and $\widehat{P} = \widehat{U}\widehat{\Sigma}\widehat{V}^\top$. Then, it holds that

$$\begin{aligned}
&\min_{W \in \mathcal{O}^{r\times r}} \|U - \widehat{U}R\|_{2\to\infty} \\
&\qquad \leq \frac{(2\sqrt{n} + (2+\sqrt{2})\sqrt{mn}\|UU^\top\|_{2\to\infty})\|P - \widehat{P}\|_{\max}}{\sigma_r(P)} \\
&\min_{W \in \mathcal{O}^{r\times r}} \|V - \widehat{V}W\|_{2\to\infty} \leq \\
&\qquad \leq \frac{(2\sqrt{m} + (2+\sqrt{2})\sqrt{mn}\|VV^\top\|_{2\to\infty})\|P - \widehat{P}\|_{\max}}{\sigma_r(P)}
\end{aligned}$$

provided that $\sqrt{mn}\epsilon\|P\|_{\max} \leq \frac{\sigma_r(P)}{2}$.

Below, we give a result showing that entry-wise guarantees imply subspace recovery in the two-to-infinity guarantee.

Proof. We will only prove the result concerning the left subspaces U and $\widehat{U}$. Our first step is to relate the errors $\widehat{U}R - U$ and $UU^\top\widehat{U} - U$. We will introduce in our computations the sign matrix[2] of $U^\top\widehat{U}$, namely $\operatorname{sgn}(U^\top\widehat{U})$ which is a rotation matrix. We have

$$\begin{aligned}
\min_{W \in \mathcal{O}^{r\times r}} \|UW - \widehat{U}\|_{2\to\infty} &\leq \|U\operatorname{sgn}(U^\top\widehat{U}) - \widehat{U}\|_{2\to\infty} \\
&\leq \|U(U^\top\widehat{U}) - \widehat{U}U\|_{2\to\infty} + \|U\|_{2\to\infty}\|U^\top\widehat{U} - \operatorname{sgn}(U^\top\widehat{U})\|_{\mathrm{op}}.
\end{aligned}$$

Moreover, we also know (e.g., see Lemma 4.15 Chen et al. (2021)) that

$$\|\widehat{U}^\top U - \operatorname{sgn}(\widehat{U}^\top U)\|_{\mathrm{op}} \leq \|\sin(\Theta)\|_{\mathrm{op}},$$

and using the Davis-Kahan Theorem (Chen et al., 2021) we obtain

$$\|\widehat{U}^\top U - \operatorname{sgn}(\widehat{U}^\top U)\|_{\mathrm{op}} \leq \|\sin(\Theta)\|_{\mathrm{op}} \leq \frac{\sqrt{2}\|M - \widehat{M}\|_{\mathrm{op}}}{\sigma_r(M)}.$$

[2]The sign matrix of an $n \times n$ matrix Z with SVD $U_Z\Sigma_Z V_Z^\top$ is given by $\operatorname{sgn}(Z) = U_Z V_Z^\top \in \mathcal{O}^{n\times n}$.

Thus, we conclude that

$$\min_{W\in\mathcal{O}^{r\times r}} \|UW-\widehat{U}\|_{2\to\infty} \le \|U(U^\top\widehat{U})-\widehat{U}U\|_{2\to\infty} + \frac{\sqrt{2}\|U\|_{2\to\infty}\|M-\widehat{M}\|_{\mathrm{op}}}{\sigma_r(M)}. \tag{3.10}$$

Next, we show that $\min_{W\in\mathcal{O}^{r\times r}} \|UW-\widehat{U}\|_{2\to\infty}$ can be well controlled by the error $M-\widehat{M}$. On the one hand, we have triangular inequality, and noting that $UU^\top M = M$ and $\widehat{U}\widehat{U}^\top\widehat{M}=\widehat{M}$ that

$$\begin{aligned}\|(UU^\top-\widehat{U}\widehat{U}^\top)\widehat{M}\|_{2\to\infty} &\le \|UU^\top M-\widehat{U}\widehat{U}^\top\widehat{M}\|_{2\to\infty} + \|UU^\top(M-\widehat{M})\|_{2\to\infty}\\ &\le \|M-\widehat{M}\|_{2\to\infty} + \|UU^\top\|_{2\to\infty}\|M-\widehat{M}\|_{\mathrm{op}}\end{aligned}$$

On the other hand, we have

$$\begin{aligned}\|(UU^\top-\widehat{U}\widehat{U}^\top)\widehat{M}\|_{2\to\infty} &= \|(U(U^\top\widehat{U})-\widehat{U})\widehat{\Sigma}\widehat{V}^\top\|_{2\to\infty}\\ &= \|(U(U^\top\widehat{U})-\widehat{U})\widehat{\Sigma}\|_{2\to\infty}\\ &\ge \|U(U^\top\widehat{U})-\widehat{U}\|_{2\to\infty}\sigma_r(\widehat{M})\\ &\ge \|U(U^\top\widehat{U})-\widehat{U}\|_{2\to\infty}\sigma_r(M) - \|U(U^\top\widehat{U})-\widehat{U}\|_{2\to\infty}\|M-\widehat{M}\|_{\mathrm{op}},\end{aligned}$$

where in the last inequality we used Weyl's inequality: $|\sigma_r(M)-\sigma_r(\widehat{M})|\le\|M-\widehat{M}\|_{\mathrm{op}}$. We combine the above inequalities to obtain

$$\|U(U^\top\widehat{U})-\widehat{U}\|_{2\to\infty} \le \frac{\|M-\widehat{M}\|_{2\to\infty} + \|UU^\top\|_{2\to\infty}\|M-\widehat{M}\|_{\mathrm{op}} + \|U(U^\top\widehat{U})-\widehat{U}\|_{2\to\infty}\|\widehat{M}-M\|_{\mathrm{op}}}{\sigma_r(M)}$$

If the following condition holds

$$\|M-\widehat{M}\|_{\mathrm{op}} \le \sqrt{mn}\|M-\widehat{M}\|_{\max} \le \frac{\sigma_r(M)}{2},$$

then

$$\|U(U^\top\widehat{U})-\widehat{U}\|_{2\to\infty} \le \frac{\|M-\widehat{M}\|_{2\to\infty} + \|UU^\top\|_{2\to\infty}\|M-\widehat{M}\|_{\mathrm{op}}}{\sigma_r(M)} + \frac{1}{2}\|U(U^\top\widehat{U})-\widehat{U}\|_{2\to\infty}$$

which in turn gives

$$\|U(U^\top\widehat{U})-\widehat{U}\|_{2\to\infty} \le \frac{2\|M-\widehat{M}\|_{2\to\infty} + 2\|UU^\top\|_{2\to\infty}\|M-\widehat{M}\|_{\mathrm{op}}}{\sigma_r(M)} \tag{3.11}$$

In summary we conclude that

$$\min_{W\in\mathcal{O}^{r\times r}} \|UW-\widehat{U}\|_{2\to\infty} \le \frac{2\|M-\widehat{M}\|_{2\to\infty} + (2+\sqrt{2})\|UU^\top\|_{2\to\infty}\|M-\widehat{M}\|_{\mathrm{op}}}{\sigma_r(M)} \tag{3.12}$$

Using the inequalities

$$\|M - \widehat{M}\|_{2\to\infty} \le \sqrt{n}\|M - \widehat{M}\|_{\max} \qquad \text{and} \qquad \|M - \widehat{M}\|_{\text{op}} \le \sqrt{mn}\|M - \widehat{M}\|_{\max},$$

we can express our bounds as

$$\min_{W\in\mathcal{O}^{r\times r}} \|UW - \widehat{U}\|_{2\to\infty} \le \frac{(2\sqrt{n} + (2+\sqrt{2})\sqrt{mn}\|UU^\top\|_{2\to\infty})\|M - \widehat{M}\|_{\max}}{\sigma_r(M)}. \tag{3.13}$$

□

A simple calculation also gives the following.

Proposition 3.6.4. *Suppose $\hat{U} \in \mathcal{O}^{m\times r}$ satisfies Assumption 3.3.5 with bound ϵ_{SSR}, and the population incoherence is $\mu_U := \frac{m\|U\|_{2\to\infty}^2}{d}$. Then $\hat{U}$ is γ-incoherent for $\gamma \le 2\mu_U + \frac{2\epsilon_{\text{SSR}}^2 m}{d}$.*

3.6.3 Proof of Proposition 3.3.1

We require the following special case of Hoeffding's inequality.

Lemma 3.6.5. *Let $X_1, \dots, X_n \overset{\text{iid}}{\sim}$ Bernoulli(p). Then:*

$$\mathbb{P}\left[|\frac{1}{n}\sum_i X_i - p| \ge \sqrt{\frac{\log n}{n}}\right] \le 2n^{-2}$$

The following concentration is standard.

Lemma 3.6.6. *Let $\boldsymbol{x} \sim S^{n-1}$. Then:*

$$\mathbb{P}[\|\boldsymbol{x}\|_\infty \ge C\sqrt{\frac{\log n}{n}}] \le 1 - O(n^{-1/2})$$

Proof. By Hoeffding's inequality,

$$\mathbb{P}\left[|\sum_i X_i - np| \ge t\right] \le 2\exp\left(-\frac{2t^2}{n}\right)$$

Let $t = \sqrt{n\log n}$. The conclusion follows. □

Finally, we require the following version of the Hanson-Wright inequality.

Theorem 3.6.7 (Rudelson and Vershynin (2013) Theorem 2.1)**.** *Let $A \in \mathbb{R}^{m\times n}$ be fixed and $\boldsymbol{x} \in \mathbb{R}^n$ a random vector with i.i.d. mean zero entries with variance 1 and $\|\boldsymbol{x}_i\|_{\psi_2} \leq K$ for all i. Then there exists constant $c > 0$ such that for any $t > 0$,*

$$\mathbb{P}\left[|\|A\boldsymbol{x}\|_2 - \|A\|_F| > t\right] \leq 2\exp\left(-\frac{ct^2}{K^4\|A\|^2}\right)$$

We are ready to state our lower bound.

Proof of Proposition 3.3.1. Let $\boldsymbol{u}_1, \ldots, \boldsymbol{u}_d \in \mathbb{R}^m$ be generated with iid $N(0, \frac{1}{m})$ entries and $\boldsymbol{v}_1, \ldots, \boldsymbol{v}_d \in \mathbb{R}^m$ be generated with iid $N(0, \frac{1}{n})$ entries. Let $Q = \sum_{i=1}^d \boldsymbol{u}_i\boldsymbol{v}_i^T$.

We first analyze the incoherence of Q. We analyze the left-incoherence. Fix $i \in [m]$ and let $\boldsymbol{y} = (U^T\boldsymbol{e}_i)$. Then we apply Theorem 3.6.7 with $\boldsymbol{x} = \sqrt{m}\boldsymbol{y}$ and $A = V$, to obtain that $\|A\boldsymbol{x}\| = \|\sqrt{m}VU^T\boldsymbol{e}_i\| \leq \|V\|_F + C'K^2\|V\|_2\sqrt{\log n}$ with probability $\geq 1 - n^{-10}$ for absolute constant $C' > 0$. Since $\boldsymbol{x}$ has iid $N(0,1)$ entries, the Orlicz norm constant is at most $K \leq 2$. Taking a union bound over all i, it follows that:

$$\mathbb{P}\left[\|\sqrt{m}VU^T\|_{2\to\infty} \leq \|V\|_F + 4C'\|V\|_2\sqrt{\log n}\right] \geq 1 - O(n^{-9})$$

It follows that the left incoherence is at most $O(\log n)$ with high probability. An identical application of Theorem 3.6.7 with $A = U$ implies that the right-incoherence is at most $O(\log m)$. Let $\mathcal{E}'$ be the event that Q is $O(\log(n \vee m))$ incoherent. Let Q be the random matrix generated as above, conditioned on $\mathcal{E}'$. Note that $\mathbb{P}[\mathcal{E}'] \geq 1 - o(1)$.

Next, let $I \subset [m], J \subset [n]$ be the rows and columns of Q that are seen in $\tilde{Q}$. Then by Lemma 3.6.5, $|I| \leq 0.99m + \sqrt{m\log m}$and $|J| \leq 0.99n + \sqrt{n\log n}$ with probability $\geq 1 - 2n^{-2} - 2m^{-2}$. Let $\mathcal{E}$ be the event that the bounds on I and J both hold.

Consider $k \in [m] \setminus I, \ell \in [n] \setminus J$. None of the entries of Q in the k^{th} row or ℓ^{th} column are seen. Therefore, since $m - |I| \geq \Omega(m)$ and $n - |J| \geq \Omega(n)$, and since $\mathbb{P}[\mathcal{E}'] \geq 1 - o(1)$, there exists a constant C such that for all $i \in [d]$, $Var(\boldsymbol{u}_{i;k}\boldsymbol{v}_{i;\ell}|\tilde{Q}) \geq C$. Therefore, since $\boldsymbol{u}_1, \ldots, \boldsymbol{u}_d, \boldsymbol{v}_1, \ldots, \boldsymbol{v}_d$ are independent, for any $\hat{Q}$, we have:

$$\begin{aligned}
\mathbb{E}[(\hat{Q}_{k\ell} - Q_{k\ell})^2|\tilde{Q}] &\geq Var(Q_{k\ell}|\tilde{Q}) \\
&\geq \sum_{i=1}^d Var(\boldsymbol{u}_{i;k}\boldsymbol{v}_{i;\ell}|\tilde{Q}) \\
&\geq Cd
\end{aligned}$$

Therefore, if we condition on $\mathcal{E}$, then $|[m] \setminus I| \geq \Omega(m)$ and $|[n] \setminus J| \geq \Omega(n)$, so $\mathbb{E}[\frac{1}{mn}\|\hat{Q} - Q\|_F^2|\tilde{Q}] \geq cd$ for a constant $c > 0$. Since $1 - 2n^{-2} - 2m^{-2} \geq \frac{1}{2}$, we conclude that:

$$\begin{aligned}\mathbb{E}[\frac{1}{mn}\|\hat{Q} - Q\|_F^2|\tilde{Q}] &\geq \frac{1}{2}\mathbb{E}[\frac{1}{mn}\|\hat{Q} - Q\|_F^2|\tilde{Q}, \mathcal{E}] \\ &\geq \frac{cd}{2}\end{aligned}$$

□

3.6.4 Proof of Theorem 3.3.2

We require a version of Fano's theorem given in Theorem 7 of Verdú et al. (1994).

Theorem 3.6.8 (Generalized Fano)**.** *Let $\mathcal{P}$ be a family of probability measures, $(\mathcal{D}, d)$ a metric space, and $\theta : \mathcal{P} \to \mathcal{D}$ a map that extracts the parameters of interest. Let $\mathcal{H} \subset \mathcal{P}$ be a finite subset of size M. Suppose $\alpha > 0$ is such that for any distinct $H_i, H_j \in \mathcal{H}$,*

$$d(\theta(H_i), \theta(H_j)) \geq \alpha.$$

And, suppose that $\beta > 0$ is such that:

$$\log 2 + \frac{1}{M^2}\sum_{i=1}^{M}\sum_{j=1}^{M} KL(H_i, H_j) \leq \beta \log M.$$

Then,

$$\inf_{\hat{\theta}} \sup_{P \in \mathcal{P}} \mathbb{E}[d(\theta(P), \hat{\theta})] \geq \alpha(1 - \beta).$$

We also require a standard expression for the KL divergence of a pair of multivariate Gaussians.

Lemma 3.6.9. *Let $\boldsymbol{\mu}, \boldsymbol{\mu}' \in \mathbb{R}^d$ be distinct and $\Sigma \succ 0$. The KL divergence of two multivariate Gaussians sharing the same covariance is given as:*

$$KL(\mathcal{N}(\boldsymbol{\mu}, \Sigma), \mathcal{N}(\boldsymbol{\mu}', \Sigma)) = (\boldsymbol{\mu} - \boldsymbol{\mu}')^T\Sigma^{-1}(\boldsymbol{\mu} - \boldsymbol{\mu}')$$

We now prove our lower bound.

Proof of Theorem 3.3.2. Let $U \in \mathbb{R}^{m\times d}, V \in \mathbb{R}^{n\times d}$ be such that $U_{ii} = 1$ and $V_{ii} = 1$ for $i \in [d]$, and all other entries are zero. Let $P = UV^T$. We construct a hypothesis space $\mathcal{H} = \{(P^{(ij)}, Q^{(ij)} : i, j \in [d]\}$ of size d^2 where $P^{(ij)}, Q^{(ij)} \in \mathbb{R}^{m\times n}$ as follows. For all members ij, we set $P^{(ij)} = P$. Next, let $R^{(ij)} = \gamma \boldsymbol{e}_i \boldsymbol{e}_j^T$ for $\gamma > 0$ to be specified later. We set $Q^{(ij)} = UR^{(ij)}V^T$.

First, notice for any $(r, s) \neq (i, j)$ that:

$$\|Q^{(ij)} - Q^{(rs)}\|_{\max}^2 = \gamma^2$$

Next, consider the KL divergences between a pair of hypotheses. Let $(\tilde{P}^{(ij)}, \tilde{Q}^{(ij)})$ be the distribution of the data under hypothesis $(P^{(ij)}, Q^{(ij)})$. Since $\tilde{P}^{(ij)} = P^{(ij)} = P$ for all (i, j), we must simply bound $KL(\tilde{Q}^{(ij)}, \tilde{Q}^{(rs)})$ for each pair (ij, rs). Now, let $\pi_R^{(ij)}, \pi_C^{(ij)}$ be the row and column sampling distributions (possibly deterministic) respectively, based on the source data $\tilde{P}^{(ij)}$. Since $\tilde{P}^{(ij)} = P^{(ij)} = P$ for all (i, j) we know that there is a pair of distributions π_R, π_C such that $\pi_R^{(ij)} = \pi_R, \pi_C^{(ij)} = \pi_C$ for all (i, j). In other words the sampling cannot depend on the hypothesis index (i, j).

Next, we analyze $KL(\tilde{Q}^{(ij)}, \tilde{Q}^{(rs)})$. Each distribution depends on the randomness of π_R, π_C as well as the Gaussian noise. Let R, C be the random multisets of rows and columns generated by π_R, π_C according to the prescribed row/column budgets. By the chain rule for KL divergences (Theorem 2.15 of Polyanskiy and Wu (2024)), we have:

$$KL(\tilde{Q}^{(ij)}, \tilde{Q}^{(rs)}) = \mathop{\mathbb{E}}_{R,C}\left[KL\Big((\tilde{Q}^{(ij)}|R, C), (\tilde{Q}^{(rs)}|R, C)\Big)\right]$$

Note that the marginal term involving $\pi_R^{(ij)}, \pi_C^{(ij)}$ versus $\pi_R^{(rs)}, \pi_C^{(rs)}$ is zero, because the distributions are equal for all ij, rs.

Next, for $u \in [m], v \in [n]$, let $n_{uv}(R, C)$ be the number of times that (u, v) is sampled in R, C. Notice that $\mathbb{E}_{R,C}[n_{uv}(R, C)] = |\Omega|\pi_R(u)\pi_C(v)$. So, by Lemma 3.6.9,

$$\begin{aligned}
\mathop{\mathbb{E}}_{R,C}\left[KL\Big((\tilde{Q}^{(ij)}|R, C), (\tilde{Q}^{(rs)}|R, C)\Big)\right] &= \mathop{\mathbb{E}}_{R,C}\left[\sum_{u\in[m],v\in[n]} \frac{n_{uv}(R, C)}{\sigma_Q^2}(Q_{uv}^{(ij)} - Q_{uv}^{(rs)})^2\right] \\
&= \mathop{\mathbb{E}}_{R,C}\left[\frac{\gamma^2}{\sigma_Q^2}(n_{ij}(R, C) + n_{rs}(R, C))\right] \\
&= \frac{\gamma^2|\Omega|}{\sigma_Q^2}(\pi_R(i)\pi_C(j) + \pi_R(r)\pi_C(s))
\end{aligned}$$

Hence, the average KL divergence for all pairs is:

$$\begin{aligned}\frac{1}{d^4}\sum_{(i,j)\in[d]^2}\sum_{(r,s)\in[d]^2} KL(\tilde{Q}^{(ij)},\tilde{Q}^{(rs)}) &= \frac{\gamma^2|\Omega|}{\sigma_Q^2 d^4}\sum_{(i,j)\in[d]^2}\sum_{(r,s)\in[d]^2}(\pi_R(i)\pi_C(j)+\pi_R(r)\pi_C(s)) \\ &\le \frac{\gamma^2|\Omega|}{\sigma_Q^2 d^4}\sum_{(i,j)\in[d]^2}(1+d^2\pi_R(i)\pi_C(j)) \\ &\le \frac{\gamma^2|\Omega|}{\sigma_Q^2 d^4}\cdot 2d^2 \\ &= \frac{2\gamma^2|\Omega|}{\sigma_Q^2 d^2}\end{aligned}$$

Let $\gamma^2 = \frac{1}{10}\frac{\sigma_Q^2 d^2}{|\Omega|}$. By Theorem 3.6.8, we conclude that for $d \ge 2$, the minimax rate of estimation is at least $\frac{1}{10}\gamma^2 = \frac{1}{100}\frac{\sigma_Q^2 d^2}{|\Omega|}$ □

3.6.5 Proof of Proposition 3.3.4

We use the classical characterization of G-optimal designs due to Kiefer and Wolfowitz.

Theorem 3.6.10 (Kiefer and Wolfowitz (1960))**.** *Let π be a distribution on a finite space $\mathcal{A} \subset \mathbb{R}^d$. The following are equivalent:*

- *π is G-optimal.*
- *$g(\pi) = d$.*
- *For $V(\pi) := \sum_{\boldsymbol{a}\in\mathcal{A}} \pi(\boldsymbol{a})\boldsymbol{a}\boldsymbol{a}^T$, π maximizes $\log\det V(\pi)$.*

We now prove the tensorization of G-optimal designs.

Proposition 3.6.11 (Restatement of Proposition 3.3.4)**.** *Let ρ be a G-optimal design for $\{\hat{U}_P^T\boldsymbol{e}_i : i \in [m]\}$ and ζ be a G-optimal design for $\{\hat{V}_P^T\boldsymbol{e}_j : j \in [n]\}$. Let $\pi(i,j) = \rho(i)\zeta(j)$ be a distribution on $[m]\times[n]$. Then π is a G-optimal design on $\{\hat{V}_P^T\boldsymbol{e}_j \otimes \hat{U}_P^T\boldsymbol{e}_i : i \in [m]\}$.*

Proof. Let $i \in [m], j \in [n]$. Then by the Kiefer-Wolfowitz theorem,

$$\begin{aligned}
g(\pi) &= \max_{i,j} \left[(\hat{V}_P^T \boldsymbol{e}_j \otimes \hat{U}_P^T \boldsymbol{e}_i)^T \left(\sum_{i,j} \pi(i,j) (\hat{V}_P^T \boldsymbol{e}_j \otimes \hat{U}_P^T \boldsymbol{e}_i)(\hat{V}_P^T \boldsymbol{e}_j \otimes \hat{U}_P^T \boldsymbol{e}_i)^T \right)^{-1} (\hat{V}_P^T \boldsymbol{e}_j \otimes \hat{U}_P^T \boldsymbol{e}_i) \right] \\
&= \max_{i,j} \left[(\hat{V}_P^T \boldsymbol{e}_j \otimes \hat{U}_P^T \boldsymbol{e}_i)^T \left(\left(\sum_j \zeta(j) \hat{V}_P^T \boldsymbol{e}_j \boldsymbol{e}_j^T \hat{V}_P^T \right) \otimes \left(\sum_i \rho(i) \hat{U}_P^T \boldsymbol{e}_i \boldsymbol{e}_i^T \hat{U}_P^T \right) \right)^{-1} (\hat{V}_P^T \boldsymbol{e}_j \otimes \hat{U}_P^T \boldsymbol{e}_i) \right] \\
&= \max_{i,j} \left[(\hat{V}_P^T \boldsymbol{e}_j \otimes \hat{U}_P^T \boldsymbol{e}_i)^T \left[\left(\sum_j \zeta(j) \hat{V}_P^T \boldsymbol{e}_j \boldsymbol{e}_j^T \hat{V}_P^T \right)^{-1} \otimes \left(\sum_i \rho(i) \hat{U}_P^T \boldsymbol{e}_i \boldsymbol{e}_i^T \hat{U}_P^T \right)^{-1} \right] (\hat{V}_P^T \boldsymbol{e}_j \otimes \hat{U}_P^T \boldsymbol{e}_i)^T \right] \\
&= \max_{i,j} \left[(\hat{V}_P^T \boldsymbol{e}_j)^T \left(\sum_j \zeta(j) \hat{V}_P^T \boldsymbol{e}_j \boldsymbol{e}_j^T \hat{V}_P^T \right)^{-1} (\hat{V}_P^T \boldsymbol{e}_j)(\hat{U}_P^T \boldsymbol{e}_i)^T \left(\sum_i \rho(i) \hat{U}_P^T \boldsymbol{e}_i \boldsymbol{e}_i^T \hat{U}_P^T \right)^{-1} (\hat{U}_P^T \boldsymbol{e}_i) \right] \\
&= g(\rho) g(\zeta) \\
&= d^2
\end{aligned}$$

Where the last step follows from G-optimality of ρ and ζ. By Theorem 3.6.10, π is G-optimal. □

3.6.6 Proof of Theorem 3.3.6

We first prove a useful error decomposition.

Proposition 3.6.12 (Decomposition)**.** *Let $\hat{U}_P \in \mathcal{O}^{m \times d}, \hat{V}_P \mathcal{O}^{n \times d}$ be the estimates of the left/right singular vectors of P. Then there exist matrices $W_U, W_V \in \mathcal{O}^d$ such that if T_1, T_2 are the distribution shift matrices as in Definition 3.1.2, and if $M = (W_U^T T_1) R (T_2^T W_V)$, then:*

$$Q = \hat{U}_P (W_U^T T_1) R (T_2^T W_V) \hat{V}_P^T + E$$

Where the E-error depends on the estimator error of $\hat{P}$.

$$E := (\hat{U}_P - U_P W_U) M \hat{V}_P^T + \hat{U}_P M (\hat{V}_P - V_P W_V)^T + (\hat{U}_P - U_P W_U) M (\hat{V}_P - V_P W_V)^T$$

Proof. Let $T_1, T_2 \in \mathbb{R}^{d \times d}$ be the distributional shift matrices from Definition 3.1.2 such that $U_Q = U_P T_1, V_Q = V_P T_2$.

Let W_U be the solution to the Procrustes problem:

$$W_U := \arg \inf_{W \in \mathcal{O}^{d \times d}} \|U_P W - \hat{U}_P\|_{2 \to \infty}$$

And similarly,

$$W_V := \arg \inf_{W \in \mathcal{O}^{d \times d}} \|V_P W - \hat{V}_P\|_{2 \to \infty}$$

Next, let $Z = T_1 R T_2^T$ and $M = W_U^T Z W_V$. Further, let $\Delta_U = \hat{U}_P - \hat{U}_P W_U$ and $\Delta_V = \hat{V}_P - \hat{V}_P W_V$. Then, we can write Q as:

$$\begin{aligned} Q &= U_P T_1 R (V_P T_2)^T \\ &= U_P Z V_P^T \\ &= U_P W_U W_U^T Z W_V W_V^T V_P^T \\ &= (\hat{U}_P + \Delta_U) W_U^T Z W_V (\hat{V}_P + \Delta_V)^T \\ &= \hat{U}_P M \hat{V}_P^T + E \end{aligned}$$

Where E contains the cross-terms:

$$E = \Delta_U M \hat{V}_P^T + \hat{U}_P M \Delta_V^T + \Delta_U M \Delta_V^T$$

So we are done. □

We require a strong form of matrix concentration due to Taupin et al. (2023).

Lemma 3.6.13 (Design Matrix Concentration)**.** *Let $\hat{\pi}$ be an ϵ-approximate G-optimal design on a finite set $\mathcal{A} \subset \mathbb{R}^d$. Let $\rho, \delta > 0$ and $t \geq 2(1+\epsilon)(\frac{1}{\rho^2} + \frac{1}{3\rho}) d \log(\frac{2d}{\delta})$. Suppose $\Omega = \{\boldsymbol{a}_1, \ldots, \boldsymbol{a}_t\}$ is the multiset of t samples drawn i.i.d. from $\hat{\pi}$, and let $W_t = \frac{1}{t} \sum_{i=1}^t \boldsymbol{a}_i \boldsymbol{a}_i^T$. Then:*

$$\mathbb{P}\left[(1-\rho) \sum_{\boldsymbol{a} \in A} \hat{\pi}(\boldsymbol{a}) \boldsymbol{a}\boldsymbol{a}^T \preceq W_t \preceq (1+\rho) \sum_{\boldsymbol{a} \in A} \hat{\pi}(\boldsymbol{a}) \boldsymbol{a}\boldsymbol{a}^T\right] \geq 1 - \delta$$

In particular, since $\hat{\pi}$ is ϵ-approximately G-optimal,

$$\mathbb{P}\left[\frac{d}{(1+\rho)} \leq \max_{\boldsymbol{a} \in A} \|\boldsymbol{a}\|_{W_t^{-1}}^2 \leq \frac{(1+\epsilon)d}{(1-\rho)}\right] \geq 1 - \delta$$

We also require the following standard bound on the maximum of Gaussians.

Lemma 3.6.14 (Vershynin (2018b) 2.5.10)**.** *Let $X_1, \ldots, X_n \overset{\text{iid}}{\sim} N(0, \sigma^2)$. Then for all $u > 0$,*

$$\mathbb{P}[\max_i X_i^2 \geq 4\sigma^2 \log(n) + 2u^2] \leq \exp(-\frac{u^2}{2\sigma^2}).$$

Proof of Theorem 3.3.6. We first introduce some notation. Let S_r, S_c be the multisets of rows/columns sampled and $\Omega = S_r \times S_c$.

Let $\boldsymbol{\psi}_j = \hat{V}_P^T \boldsymbol{e}_j$ and $\boldsymbol{\varphi}_i = \hat{U}_P^T \boldsymbol{e}_i$. Then, let $\hat{\boldsymbol{\phi}}_{ij} = \hat{V}_P^T \boldsymbol{e}_j \otimes \hat{U}_P^T \boldsymbol{e}_i = \boldsymbol{\psi}_j \otimes \boldsymbol{\varphi}_k$, and $W = \sum_{ij \in \Omega} \hat{\boldsymbol{\phi}}_{ij} \hat{\boldsymbol{\phi}}_{ij}^T$. Notice that:

$$W = \left(\sum_{j \in S_c} \boldsymbol{\psi}_j \boldsymbol{\psi}_j^T \right) \otimes \left(\sum_{i \in S_r} \boldsymbol{\varphi}_i \boldsymbol{\varphi}_i^T \right)$$

Therefore, let $W_1 = \sum_{j \in S_c} \boldsymbol{\psi}_i \boldsymbol{\psi}_j^T$ and $W_2 = \sum_{i \in S_r} \boldsymbol{\varphi}_i \boldsymbol{\varphi}_i^T$ for shorthand. Then W^{-1} exists iff W_1^{-1}, W_2^{-1} exist. By Lemma 3.6.13, both W_1^{-1}, W_2^{-1} exist with probability at least $1 - (m+n)^{-2}$, since S_r, S_c are both large enough by assumption.

Therefore, conditioning on the inverses existing, if we solve the least-squares system, we obtain $\hat{M} \in \mathbb{R}^{d \times d}$ such that:

$$\text{vec}(\hat{M}) = (\sum_{ij \in \Omega} \hat{\boldsymbol{\phi}}_{ij} \hat{\boldsymbol{\phi}}_{ij}^T)^{-1} \sum_{ij \in \Omega} \hat{\boldsymbol{\phi}}_{ij} \tilde{Q}_{ij}$$

Recall from Proposition 3.6.12 that $Q = \hat{U}_P M \hat{V}_P^T + E$, where $E_{ij} = \epsilon_{ij}$ is the misspecification error. Therefore, we can bound the error of $\hat{Q} = \hat{U}_P \hat{M} \hat{V}_P^T$ as:

$$\begin{aligned}
\hat{Q}_{ij} - Q_{ij} &= \boldsymbol{e}_i^T \hat{U}_P (\hat{M} - M) \hat{V}_P^T \boldsymbol{e}_j - \epsilon_{ij} \\
&= \hat{\boldsymbol{\phi}}_{ij}^T \text{vec}(\hat{M} - M) + \epsilon_{ij} \\
&=: E_{1;ij} + E_{2;ij} \\
E_{1;ij} &:= \hat{\boldsymbol{\phi}}_{ij}^T \text{vec}(\hat{M} - M) \\
E_{2;ij} &:= \epsilon_{ij}
\end{aligned}$$

Let $G_{ij} \overset{\text{iid}}{\sim} N(0, \sigma_Q^2)$ be the additive noise for $\tilde{Q}_{ij}$. Then, $\tilde{Q}_{ij} = \hat{\boldsymbol{\phi}}_{ij}^T \text{vec}(M) + \epsilon_{ij} + G_{ij}$.

Hence we can write E_1 as:

$$\begin{aligned}
E_{1;k\ell} &= \Big(\hat{\boldsymbol{\phi}}_{k\ell}^T(\sum_{ij\in\Omega}\hat{\boldsymbol{\phi}}_{ij}\hat{\boldsymbol{\phi}}_{ij}^T)^{-1}\sum_{ij\in\Omega}\hat{\boldsymbol{\phi}}_{ij}\tilde{Q}_{ij}\Big) - \hat{\boldsymbol{\phi}}_{k\ell}^T\mathrm{vec}(M) \\
&= \hat{\boldsymbol{\phi}}_{k\ell}^T\Big((\sum_{ij\in\Omega}\hat{\boldsymbol{\phi}}_{ij}\hat{\boldsymbol{\phi}}_{ij}^T)^{-1}\sum_{ij\in\Omega}\hat{\boldsymbol{\phi}}_{ij}\big(\hat{\boldsymbol{\phi}}_{ij}^T\mathrm{vec}(M)+\epsilon_{ij}+G_{ij}\big)\Big) - \hat{\boldsymbol{\phi}}_{k\ell}^T\mathrm{vec}(M) \\
&= \hat{\boldsymbol{\phi}}_{k\ell}^T\Big((\sum_{ij\in\Omega}\hat{\boldsymbol{\phi}}_{ij}\hat{\boldsymbol{\phi}}_{ij}^T)^{-1}\sum_{ij\in\Omega}\hat{\boldsymbol{\phi}}_{ij}\big(\epsilon_{ij}+G_{ij}\big)\Big) \\
&= \hat{\boldsymbol{\phi}}_{k\ell}^T\Big((\sum_{ij\in\Omega}\hat{\boldsymbol{\phi}}_{ij}\hat{\boldsymbol{\phi}}_{ij}^T)^{-1}\sum_{ij\in\Omega}\hat{\boldsymbol{\phi}}_{ij}\epsilon_{ij}\Big) + \hat{\boldsymbol{\phi}}_{k\ell}^T\Big((\sum_{ij\in\Omega}\hat{\boldsymbol{\phi}}_{ij}\hat{\boldsymbol{\phi}}_{ij}^T)^{-1}\sum_{ij\in\Omega}\hat{\boldsymbol{\phi}}_{ij}G_{ij}\Big) \\
&=: E_{3;k\ell} + E_{4;k\ell}
\end{aligned}$$

We analyze E_4 first. Let $\boldsymbol{x} = W^{-1}\sum_{ij\in\Omega}\hat{\boldsymbol{\phi}}_{ij}G_{ij}$. For any k,ℓ, we wish to bound $\hat{\boldsymbol{\phi}}_{k\ell}^T\boldsymbol{x}$. Notice that $\boldsymbol{x}$ is a multivariate Gaussian with mean $\mathbf{0}$. Its covariance is therefore:

$$\mathbb{E}[\boldsymbol{x}\boldsymbol{x}^T] = \sum_{ij\in\Omega}\sum_{i'j'\in\Omega}W^{-1}\hat{\boldsymbol{\phi}}_{ij}\hat{\boldsymbol{\phi}}_{i'j'}^TW^{-1}\,\mathbb{E}[G_{ij}G_{i'j'}] = \sigma_Q^2W^{-1}\big(\sum_{ij\in\Omega}\hat{\boldsymbol{\phi}}_{ij}\phi_{ij}^T\big)W^{-1} = \sigma_Q^2W^{-1}$$

Hence $\hat{\boldsymbol{\phi}}_{k\ell}^T\boldsymbol{x}$ is a scalar Gaussian with mean zero and variance $\hat{\boldsymbol{\phi}}_{k\ell}^T\sigma_Q^2W^{-1}\hat{\boldsymbol{\phi}}_{k\ell}$. We next bound this quadratic form. Notice that we can tensorize the quadratic form as:

$$\begin{aligned}
\boldsymbol{\phi}_{k\ell}^TW^{-1}\boldsymbol{\phi}_{k\ell} &= (\boldsymbol{\psi}_\ell\otimes\boldsymbol{\varphi}_k)^T(W_1\otimes W_2)^{-1}(\boldsymbol{\psi}_\ell\otimes\boldsymbol{\varphi}_k) \\
&= (\boldsymbol{\psi}_\ell W_1^{-1}\boldsymbol{\psi}_\ell)(\boldsymbol{\varphi}_kW_2^{-1}\boldsymbol{\varphi}_k)
\end{aligned}$$

We apply Lemma 3.6.13 to each term in the product. With probability $1-2(m+n)^{-2}$, for S_r, S_c both of size at least $20d\log(\frac{2d}{m+n})$,

$$\|\boldsymbol{\psi}_\ell\|_{W_1^{-1}}^2\|\boldsymbol{\varphi}_k\|_{W_2^{-1}}^2 \le \frac{(2+2\epsilon)d^2}{|S_r||S_c|}$$

Conditioning on this event, the variance of $\hat{\boldsymbol{\phi}}_{k\ell}^T\boldsymbol{x}$ is at most $\frac{(1+\epsilon)d^2\sigma_Q^2}{|\Omega|(1-\rho)}$, for $|\Omega| = |S_r||S_c|$. Therefore, by Lemma 3.6.14,

$$\mathbb{P}\left[\max_{k\in[m],\ell\in[n]}|\hat{\boldsymbol{\phi}}_{k\ell}^T\boldsymbol{x}|^2 \le 20\log(mn)\frac{(2+2\epsilon)\sigma_Q^2d^2}{|\Omega|}\right] \le \delta + (mn)^{-2}$$

Finally, we analyze the error term $E_{3;k\ell}$. Let $a_{ij} = \hat{\boldsymbol{\phi}}_{k\ell}^TW^{-1}\hat{\boldsymbol{\phi}}_{ij}$. By the Cauchy-Schwarz inequality,

$$|E_{3;k\ell}| \le \big(\sum_{ij\in\Omega}a_{ij}^2\big)^{1/2}\big(\sum_{ij\in\Omega}\epsilon_{ij}^2\big)^{1/2}$$

First,

$$\begin{aligned}\sum_{ij\in\Omega} a_{ij}^2 &= \sum_{ij\in\Omega} \hat{\boldsymbol{\phi}}_{ij}^T W^{-1} \hat{\boldsymbol{\phi}}_{k\ell} \hat{\boldsymbol{\phi}}_{k\ell}^T W^{-1} \hat{\boldsymbol{\phi}}_{ij} \\ &= \sum_{ij\in\Omega} \mathrm{tr}\Big(\hat{\boldsymbol{\phi}}_{ij} \hat{\boldsymbol{\phi}}_{ij}^T W^{-1} \hat{\boldsymbol{\phi}}_{k\ell} \hat{\boldsymbol{\phi}}_{k\ell}^T W^{-1} \Big) \\ &= \mathrm{tr}\Big(\sum_{ij\in\Omega} \hat{\boldsymbol{\phi}}_{ij} \hat{\boldsymbol{\phi}}_{ij}^T W^{-1} \hat{\boldsymbol{\phi}}_{k\ell} \hat{\boldsymbol{\phi}}_{k\ell}^T W^{-1} \Big) \\ &= \mathrm{tr}\Big(\hat{\boldsymbol{\phi}}_{k\ell} \hat{\boldsymbol{\phi}}_{k\ell}^T W^{-1} \Big) \\ &= |\hat{\boldsymbol{\phi}}_{k\ell}^T W^{-1} \hat{\boldsymbol{\phi}}_{k\ell}| \\ &\leq \frac{(2+2\epsilon)d^2}{|\Omega|}\end{aligned}$$

For the other term,

$$\big(\sum_{ij\in\Omega} \epsilon_{ij}^2 \big)^{1/2} \leq |\Omega|^{1/2} \max_{ij\in\Omega} |\epsilon_{ij}|$$

It follows that $\max_{k,\ell} |E_{3;k\ell}| \leq \sqrt{2+2\epsilon} \cdot d \max_{i,j\in\Omega} |\epsilon_{ij}|$. The conclusion follows. □

3.6.7 Proof of Theorem 3.3.9

We require the following concentration result to control the sizes of masks.

Lemma 3.6.15 (Bernoulli Concentration)**.** *Let* $X_1, \ldots, X_n \overset{\text{iid}}{\sim}$ *Bernoulli*(p) *for* $p \in (0,1)$. *Then if* $p \geq 10 \log n$,

$$\mathbb{P}[|\sum_i (X_i - p)| \geq \frac{np}{2}] \leq n^{-\omega(1)}$$

Proof. By the scalar Bernstein inequality (Lemma 3.6.16), we have for $B = 1$ and $\zeta = np$ that:

$$\mathbb{P}[|\sum_i (X_i - p)| \geq \tau] \leq 2\exp(-\frac{\tau^2/2}{\zeta + (B\tau/3)})$$

Let $\tau = np/2$. Then

$$\begin{aligned}\mathbb{P}[|\sum_i (X_i - p)| \geq \tau] &\leq 2\exp(\frac{-10}{8}\log n) \\ &\leq 2n^{-(\log n)^{1/4}}\end{aligned}$$

□

We are ready to prove the estimation error for passive sampling.

Proof of Theorem 3.3.9. Following the notation of the proof of Theorem 3.3.6, we want to bound $E_{3;k\ell}$ and $E_{4;k\ell}$. However, rather than using G-optimality to bound quadratic forms of the type $\hat{\boldsymbol{\phi}}_{k\ell} W^{-1} \hat{\boldsymbol{\phi}}_{ij}$, we will apply spectral concentration via Proposition 3.6.17.

To this end, we condition on the events that $\hat{V}_P^T \Pi_R \hat{V}_P \succeq \frac{p_{\text{Row}}}{2}$ and $\hat{U}_P^T \Pi_C \hat{U}_P \succeq \frac{p_{\text{Col}}}{2}$. By Proposition 3.6.4 and Proposition 3.6.17, the two events occur simultaneously with probability $\geq 1 - 2(m \wedge n)^{-10}$. Then W^{-1} exists and $W^{-1} \preceq \frac{4}{p_{\text{Row}} p_{\text{Col}}} I$. Therefore, for all i, j, k, ℓ, by incoherence,

$$\begin{aligned}
|\hat{\boldsymbol{\phi}}_{k\ell}^T W^{-1} \hat{\boldsymbol{\phi}}_{ij}| &\leq \frac{4}{p_{\text{Row}} p_{\text{Col}}} \|\hat{\boldsymbol{\phi}}_{k\ell}\| \|\hat{\boldsymbol{\phi}}_{ij}\| \\
&= \frac{4}{p_{\text{Row}} p_{\text{Col}}} \|\boldsymbol{\varphi}_k\| \|\boldsymbol{\varphi}_i\| \|\boldsymbol{\psi}_\ell\| \|\boldsymbol{\psi}_j\| \\
&\leq \frac{4}{p_{\text{Row}} p_{\text{Col}}} \Big(\sqrt{\frac{\mu_U^2 \mu_V^2 d^4}{m^2 n^2}}\Big) \\
&= \frac{4}{p_{\text{Row}} p_{\text{Col}}} \frac{\mu d^2}{mn}
\end{aligned}$$

Hence, by Lemma 3.6.14,

$$\mathbb{P}\left[\max_{k \in [m], \ell \in [n]} |E_{4;k\ell}|^2 \leq 20 \log(mn) \sigma_Q^2 \frac{4}{p_{\text{Row}} p_{\text{Col}}} \frac{\mu d^2}{mn}\right] \leq 2(m \wedge n)^{-10} + (mn)^{-2}.$$

Next, we analyze E_3. Let $a_{ij} = \hat{\boldsymbol{\phi}}_{k\ell}^T W^{-1} \hat{\boldsymbol{\phi}}_{ij}$. Let $p = q = 2$. By the Cauchy-Schwarz inequality,

$$|E_{3;k\ell}| \leq \Big(\sum_{ij \in \Omega} a_{ij}^p\Big)^{1/p} \Big(\sum_{ij \in \Omega} \epsilon_{ij}^q\Big)^{1/q}$$

First, we have:

$$\begin{aligned}
\sum_{ij\in\Omega} a_{ij}^2 &= \sum_{ij\in\Omega} \hat{\boldsymbol{\phi}}_{ij}^T W^{-1} \hat{\boldsymbol{\phi}}_{k\ell} \hat{\boldsymbol{\phi}}_{k\ell}^T W^{-1} \hat{\boldsymbol{\phi}}_{ij} \\
&= \sum_{ij\in\Omega} \mathrm{tr}\Big(\hat{\boldsymbol{\phi}}_{ij} \hat{\boldsymbol{\phi}}_{ij}^T W^{-1} \hat{\boldsymbol{\phi}}_{k\ell} \hat{\boldsymbol{\phi}}_{k\ell}^T W^{-1} \Big) \\
&= \mathrm{tr}\Big(\sum_{ij\in\Omega} \hat{\boldsymbol{\phi}}_{ij} \hat{\boldsymbol{\phi}}_{ij}^T W^{-1} \hat{\boldsymbol{\phi}}_{k\ell} \hat{\boldsymbol{\phi}}_{k\ell}^T W^{-1} \Big) \\
&= \mathrm{tr}\Big(\hat{\boldsymbol{\phi}}_{k\ell} \hat{\boldsymbol{\phi}}_{k\ell}^T W^{-1} \Big) \\
&= |\hat{\boldsymbol{\phi}}_{k\ell}^T W^{-1} \hat{\boldsymbol{\phi}}_{k\ell}| \\
&\le \frac{4\mu d^2}{p_{\mathrm{Row}} p_{\mathrm{Col}} mn}
\end{aligned}$$

On the other hand,

$$\big(\sum_{ij\in\Omega} \epsilon_{ij}^q \big)^{1/2} \le |\Omega|^{1/2} \max_{ij\in\Omega} |\epsilon_{ij}|$$

Notice $\mathbb{E}[|\Omega|] = mnp_{\mathrm{Row}}p_{\mathrm{Col}}$. By Lemma 3.6.15, with probability $\ge 1 - 4(m \wedge n)^{-\omega(1)}$,

$$|\Omega| \le \frac{9}{4} p_{\mathrm{Row}} p_{\mathrm{Col}} mn$$

Therefore, with probability $\ge 1 - 4(m \wedge n)^{-2}$,

$$\frac{\sqrt{|\Omega|}}{p_{\mathrm{Row}} p_{\mathrm{Col}} mn} \le \frac{3}{2} \frac{1}{\sqrt{p_{\mathrm{Row}} p_{\mathrm{Col}} mn}}$$

The conclusion follows. □

3.6.8 Proof of Proposition 3.3.11

We require the following version of the Matrix Bernstein Inequality Chen et al. (2021).

Lemma 3.6.16 (Matrix Bernstein Inequality)**.** *Suppose that* $\{Y_i : i = 1, \dots, n\}$ *are independent mean-zero random matrices of size* $d_1 \times d_2$*, such that* $\|Y_i\|_2 \le B$ *almost surely for all* i*, and* $\zeta \ge \max\{\|\mathbb{E}[\sum_i Y_i Y_i^T]\|_2, \|\mathbb{E}[\sum_i Y_i^T Y_i]\|_2\}$*. Then,*

$$\mathbb{P}\left[\Big\| \sum_{i=1}^n Y_i \Big\|_2 \ge \tau \right] \le (d_1 + d_2) \exp\left(- \frac{\tau^2/2}{\zeta + B\tau/3} \right)$$

We now prove nondegeneracy of masks with high probability.

Proposition 3.6.17 (Spectral Concentration)**.** *Suppose that $\hat{V}_P$ and $\hat{U}_P$ are μ_V, μ_U-incoherent respectively. Let $\Pi_C \in \{0,1\}^{n\times n}$ be the random matrix with diagonal entries $\nu_1, \dots, \nu_n$ and similarly let $\Pi_R \in \{0,1\}^{m\times m}$ have diagonal entries $\eta_1, \dots, \eta_m$. Then, assuming that $\mu_V \le \frac{p_{\text{Col}} n}{400 d \log n}$ and $\mu_U \le \frac{p_{\text{Row}} n}{400 d \log n}$, we have:*

$$\mathbb{P}[\hat{U}_P^T \Pi_R \hat{U}_P \succeq p_{\text{Row}}/2] \ge 1 - m^{-10}$$
$$\mathbb{P}[\hat{V}_P^T \Pi_C \hat{V}_P \succeq p_{\text{Col}}/2] \ge 1 - n^{-10}$$

Proof. Suppose that $\hat{V}_P$ has rows $\boldsymbol{y}_1, \dots, \boldsymbol{y}_n \in \mathbb{R}^d$. Then,

$$\hat{V}_P^T \Pi_C \hat{V}_P = \sum_{i=1}^n \nu_i \boldsymbol{y}_i \boldsymbol{y}_i^T.$$

Let $\boldsymbol{v}_i = \sqrt{n}\boldsymbol{y}_i$. Let $p_{\text{Col}} = \mathbb{E}[\nu_i]$. We use $p = p_{\text{Col}}$ for shorthand. Notice $\mathbb{E}[\hat{V}_P^T \Pi_C \hat{V}_P] = \sum_i p \boldsymbol{y}_i \boldsymbol{y}_i^T = p I_d$, since $\hat{V}_P^T \hat{V}_P = I_d$. Therefore,

$$\left\| \sum_i \nu_i \boldsymbol{v}_i \boldsymbol{v}_i^T - pn I_d \right\|_2 = \left\| \sum_i (\nu_i - p) \boldsymbol{v}_i \boldsymbol{v}_i^T \right\|_2$$

Let $Y_i = (\nu_i - p)\boldsymbol{v}_i \boldsymbol{v}_i^T$. Note that $\mathbb{E}[Y_i] = 0$. Next, let $\mu := \mu_V$. By incoherence, $\|Y_i\|_2 \le \|\boldsymbol{v}_i\|_2^2 \le \mu d$ for all i. Further,

$$\begin{aligned}
\max\{\|\mathbb{E}[\sum_i Y_i Y_i^T]\|_2, \|\mathbb{E}[\sum_i Y_i^T Y_i]\|_2\} &= \|\mathbb{E}[\sum_i Y_i^2]\|_2 \\
&= p(1-p)\|\sum_i \|\boldsymbol{v}_i\|_2^2 \boldsymbol{v}_i \boldsymbol{v}_i^T\|_2 \\
&\le p(1-p) n \mu d \|\sum_i \boldsymbol{y}_i \boldsymbol{y}_i^T\|_2 \\
&= p(1-p) n \mu d
\end{aligned}$$

Thus, by Lemma 3.6.16, for $B = \mu d$ and $\zeta = p(1-p)n\mu d$, we have:

$$\mathbb{P}\left[\left\|\sum_{i=1}^n Y_i\right\|_2 \ge \tau\right] \le 2n \exp\left(-\frac{\tau^2/2}{\zeta + B\tau/3}\right)$$

Setting $\tau = 10\sqrt{p(1-p)n\mu d \log n} \vee 10\mu d\sqrt{\log n}$ implies that:

$$\mathbb{P}\left[\sum_i \nu_i \boldsymbol{v}_i \boldsymbol{v}_i^T \succeq pn - \tau\right] \ge 1 - n^{-10}$$

If $\mu \leq \frac{pn}{400d\log n}$, then $\tau \leq pn/2 = p_{\text{Col}} \cdot n/2$. We conclude that $\mathbb{P}[\hat{V}_P^T \Pi_C \hat{V}_P \succeq p_{\text{Col}}/2] \geq 1 - n^{-10}$. An identical argument gives $\mathbb{P}[\hat{U}_P^T \Pi_R \hat{U}_P \succeq p_{\text{Row}}/2] \geq 1 - m^{-10}$. □

Corollary 3.6.18. *Under the assumptions of Proposition 3.6.17, the design matrix for passive sampling has rank d^2 with probability at least $1 - 2(m \wedge n)^{-10}$.*

Proof. Let $\Omega \subset [m] \times [n]$ be the set of indices corresponding to the observed entries of $\tilde{Q}$. Let $P_\Omega \in \{0,1\}^{|\Omega| \times mn}$ be the coordinate projection. The design matrix is precisely $P_\Omega(\hat{V}_P \otimes \hat{U}_P)$. Then, notice that:

$$\begin{aligned}
\big(P_\Omega(\hat{V}_P \otimes \hat{U}_P)\big)^T \big(P_\Omega(\hat{V}_P \otimes \hat{U}_P)\big) &= (\hat{V}_P \otimes \hat{U}_P)^T P_\Omega^T P_\Omega (\hat{V}_P \otimes \hat{U}_P) \\
&= (\hat{V}_P \otimes \hat{U}_P)^T (\Pi_C \otimes \Pi_R)(\hat{V}_P \otimes \hat{U}_P) \\
&= \hat{V}_P^T \Pi_C \hat{V}_P \otimes \hat{U}_P^T \Pi_R \hat{U}_P
\end{aligned}$$

By Proposition 3.6.17, this matrix has rank at least d^2 with probability $\geq 1 - 2(m \wedge n)^{-10}$. □

3.6.9 Proof of Theorem 3.3.12

We require the the Gilbert-Varshamov code Guruswami et al. (2019).

Theorem 3.6.19 (Gilbert-Varshamov)**.** *Let $q \geq 2$ be a prime power. For $0 < \epsilon < \frac{q-1}{q}$ there exists an ϵ-balanced code $C \subset \mathbb{F}_q^n$ with rate $\Omega(\epsilon^2 n)$.*

We will use the following version of Fano's inequality.

Theorem 3.6.20 (Generalized Fano Method, Yu (1997))**.** *Let $\mathcal{P}$ be a family of probability measures, $(\mathcal{D}, d)$ a pseudo-metric space, and $\theta : \mathcal{P} \to \mathcal{D}$ a map that extracts the parameters of interest. For a distinguished $P \in \mathcal{P}$, let $X \sim P$ be the data and $\hat{\theta} := \hat{\theta}(X)$ be an estimator for $\theta(P)$.*

Let $r \geq 2$ and $\mathcal{P}_r \subset \mathcal{P}$ be a finite hypothesis class of size r. Let $\alpha_r, \beta_r > 0$ be such that for all $i \neq j$, and all $P_i, P_j \in \mathcal{P}_r$,

$$\begin{aligned}
d(\theta(P_i), \theta(P_j)) &\geq \alpha_r; \\
KL(P_i, P_j) &\leq \beta_r.
\end{aligned}$$

Then

$$\max_{j \in [r]} \mathbb{E}_{P_j}[d(\hat{\theta}(X), \theta(P_j))] \geq \frac{\alpha_r}{2}\left(1 - \frac{\beta_r + \log 2}{\log r}\right).$$

We can now prove Theorem 3.3.12.

Proof of Theorem 3.3.12. Let $C \subset \{0,1\}^{d^2}$ be the 0.1-balanced Gilbert-Varshmaov code as in Theorem 3.6.19. Let $U, V \in \mathbb{R}^{n\times d}$ be Stiefel matrices with incoherence parameter $\mu = O(1)$. Let $P = U\Sigma_P V^T$ for a diagonal $\Sigma_P \succ 0$ to be specified later. Let $\delta_Q > 0$ be a positive real to be specified later.

We will construct a family of source/target pairs indexed by C similar to Jalan et al. (2024b). For $w \in C$, let $B_w \in \mathbb{R}^{d\times d}$ be defined as:

$$B_{w;ij} := \begin{cases} \frac{\sqrt{mn}}{2d} & w_{ij} = 0 \\ \frac{\sqrt{mn}}{d}(\frac{1}{2} + \delta_Q) & w_{ij} = 1 \end{cases}$$

Then define $(P_w, Q_w) = (P, UB_wV^T)$.

For a fixed $w \in C$, the distribution of the data $(A_P, \tilde{Q})$ depends on the random noise and masking of both $A_P, \tilde{Q}$. Let $D_R \in \{0,1\}^{m\times m}$ and $D_C \in \{0,1\}^{n\times n}$ be the diagonal matrices corresponding to the row/column masks for Q, and let $G \in \mathbb{R}^{m\times n}$ have iid $N(0, \sigma_Q^2)$ entries. Then $\tilde{Q} = D_R(Q + G)D_C$.

Now, we will apply Theorem 3.6.20 to lower bound $\mathbb{E}\left[\frac{1}{mn}\|\hat{Q} - Q_w\|_F^2 \middle| D_R, D_C\right]$. Fix any $D_R \in \text{supp}(\mathcal{E}_1), D_C \in \text{supp}(\mathcal{E}_2)$. Let $\tilde{P}_w, \tilde{Q}_w$ denote the distribution of the data when the population matrices are P_w, Q_w and we condition on the Q-mask matrices D_R, D_C.

By Theorem 3.6.19, the hypothesis space indexed by C is such that $\log(|C|) \geq C_1 d^2$ for absolute constant $C_1 > 0$. Next, for distinct $w, w' \in C$,

$$\begin{aligned} KL((\tilde{P}_w, \tilde{Q}_w), (\tilde{P}_{w'}, \tilde{Q}_{w'})) &= KL(\tilde{P}_{w'}, \tilde{P}_w) + KL(\tilde{Q}_w, \tilde{Q}_{w'}) \\ &\leq KL(\tilde{Q}_w, \tilde{Q}_{w'}) \\ &= KL((D_C \otimes D_R)\text{vec}(Q_w + G), (D_C \otimes D_R)\text{vec}(Q_{w'} + G)) \end{aligned}$$

Notice that we do not use any properties of $\tilde{P}_w, \tilde{P}_{w'}$, and in particular allow for deterministic $\tilde{P}_w = P_w = P$.

Since D_C, D_R are fixed, this is simply the KL divergence of two multiariate Gaussians

with the same covariance but different means. Therefore, by Lemma 3.6.9, we have that:

$$\begin{aligned}
KL((\tilde{P}_w, \tilde{Q}_w), (\tilde{P}_{w'}, \tilde{Q}_{w'})) &\leq \frac{1}{\sigma_Q^2}\text{vec}(Q_w - Q_{w'})^T(D_C \otimes D_R)^T(D_C \otimes D_R)^{-1}(D_C \otimes D_R)\text{vec}(Q_w - Q_{w'}) \\
&= \frac{1}{\sigma_Q^2}\|D_R(Q_w - Q_{w'})D_C\|_F^2 \\
&= \frac{1}{\sigma_Q^2}\|D_R U(B_w - B_{w'})V^T D_C\|_F^2 \\
&\leq \frac{1}{\sigma_Q^2}\|D_R U\|_2^2\|D_C V\|_2^2\|B_w - B_{w'}\|_F^2 \\
&\leq \frac{5p_{\text{Row}}p_{\text{Col}}}{\sigma_Q^2}(\delta_Q^2\frac{mn}{d^2})d^2 \\
&= \frac{5p_{\text{Row}}p_{\text{Col}}mn\delta_Q^2}{\sigma_Q^2}.
\end{aligned}$$

In the penultimate step, we used the fact that $D_R \in \text{supp}(\mathcal{E}_1), D_C \in \text{supp}(\mathcal{E}_2)$.

Next, for any distinct $w, w' \in C$, by Theorem 3.6.19 we have that $\mathbb{P}_{i,j\in[d]}[w_{ij} \neq w'_{ij}] \geq 0.1$. Therefore,

$$\begin{aligned}
\|Q_w - Q_{w'}\|_F &= \|U(B_w - B_{w'})V^T\|_F \\
&= \|(B_w - B_{w'})\|_F \\
&= \left(\sum_{i,j\in[d]:w_{ij}\neq w'_{ij}} \delta_Q^2\frac{mn}{d^2}\right)^{1/2} \\
&\geq \frac{1}{10}\delta_Q\sqrt{mn}
\end{aligned}$$

In the notation of Theorem 3.6.20, we have:

$$\begin{aligned}
\alpha_r &:= \frac{1}{10}\delta_Q\sqrt{mn} \\
\beta_r &= \frac{5p_{\text{Row}}p_{\text{Col}}mn\delta_Q^2}{\sigma_Q^2}
\end{aligned}$$

Since $\log(|C|) \geq C_1 d^2$, we set $\delta_Q = \sqrt{\frac{C_1 d^2 \sigma_Q^2}{10p_{\text{Row}}p_{\text{Col}}mn}}$ so that that $\beta_r = \frac{C_1 d^2}{2}$. Therefore, by Theorem 3.6.20, for absolute constants $C_2, C_3, C_4 > 0$,

$$\begin{aligned}
\min_{D_R\in\text{supp}(\mathcal{E}_1), D_C\in\text{supp}(\mathcal{E}_2)} \mathbb{E}\left[\frac{1}{mn}\|\hat{Q} - Q_w\|_F^2 \middle| D_R, D_C\right] &\geq \frac{C_2\alpha_r^2}{mn} \\
&\geq C_3\delta_Q^2 \\
&\geq \frac{C_4 d^2\sigma_Q^2}{p_{\text{Row}}p_{\text{Col}}mn}
\end{aligned}$$

The conclusion follows. □

3.7 Additional Experiments and Details

Compute environment. We run all experiments on a Linux machine with 378GB of CPU/RAM. The total compute time across all results in the paper was less than 4 hours.

Dataset details. For the gene expression experiments, we gather whole-blood sepsis gene expression data sampled by Parnell et al. (2013), available at `https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=gse54514`. We take the intersection of rows and columns present on days 1 and 2 of the study, and then filter by the 300 most expressed columns (genes) on day 1, to obtain $P, Q \in \mathbb{R}^{31\times 300}$. Here P_{ij} is the expression level of gene j for patient i on day 1, and Q_{ij} is the same on day 2.

For the metabolic networks experiments, we access the BiGG genome-scale metabolic models datasets King et al. (2016) at `http://bigg.ucsd.edu`. We use the same set of shared metabolites for iWFL1372 (the source species P) and IJN1463 (the target species Q) as Jalan et al. (2024b). The resulting networks are weighted undirected graphs with adjacency matrices $P, Q \in \mathbb{R}^{251\times 251}$ where P_{ij} counts the number of co-occurrences of metabolites i, j in iWFL1372, and Q_{ij} does the same for IJN1463.

Details of the baselines. For the method of Bhattacharya and Chatterjee (2022), we use the estimator from their Section 2.2, but modify step (3) to truncate to the true rank d, and in step (6) truncate to the true rank of the propensity matrix whose (i, j) entry is $\eta_i \nu_j$. The propensity rank is always 1 in our case. This is the estimator $\hat{Q}_{\mathrm{BC22}} \in \mathbb{R}^{m\times n}$.

For the method of Levin et al. (2022), we use the estimator from their Section 3.3, with weights w_P, w_Q based on estimated sub-gamma parameters of the noise for $\tilde{P}, \tilde{Q}$. Then, let $Q' \in \mathbb{R}^{m\times n}$ be:

$$Q'_{ij} := \begin{cases} \frac{w_P}{w_P+w_Q}\tilde{P}_{ij} + \frac{w_Q}{w_P+w_Q}\tilde{Q}_{ij} & \tilde{Q}_{ij} \neq \star \\ \tilde{P}_{ij} & \text{otherwise} \end{cases}$$

We return the rank-d SVD truncation of Q' as $\hat{Q}_{\mathrm{LLL22}} \in \mathbb{R}^{m\times n}$.

We will discuss additional ablation experiments in Section 3.7.1, and experiments on the real-world data in Section 3.7.2.

3.7.1 Ablation Studies

Throughout this section we use the Partitioned Matrix Model with $a = 0.1, b = 0.8$ from Section 5.6. For each setting, we hold all parameters fixed and vary one parameter pto observe the effect of all algorithms on both Max Sqaured Error and Mean Squared Error. The default settings are:

- Matrices $P, Q \in \mathbb{R}^{m \times n}$ with $m = 300, n = 200$.
- The parameters $a = 0.8, b = 0.1$ in the Partitioned Matrix Model.
- Additive noise for $\tilde{Q}$ is iid $\mathcal{N}(0, \sigma_Q^2)$ with $\sigma_Q = 0.1$.
- The rank is $d = 5$.
- $p_{\text{Row}} = p_{\text{Col}} = 0.5$, so the probability of seeing any entry of Q is 0.25.

For all experiments, we test for 10 independent trials at each parameter setting and display the median error of each method, along with the $[10, 90]$ percentile.

Figure 3.9 shows that all methods do poorly in max errow when P is masked. Our methods are best in mean-squared error. This is because the Matrix Partition Model is highly coherent, as can be shown from spectral partitioning arguments Lee et al. (2014a). Therefore, the max-squared error is high, as we would expect from Remark 3.3.7 and the results of Chen et al. (2020b).

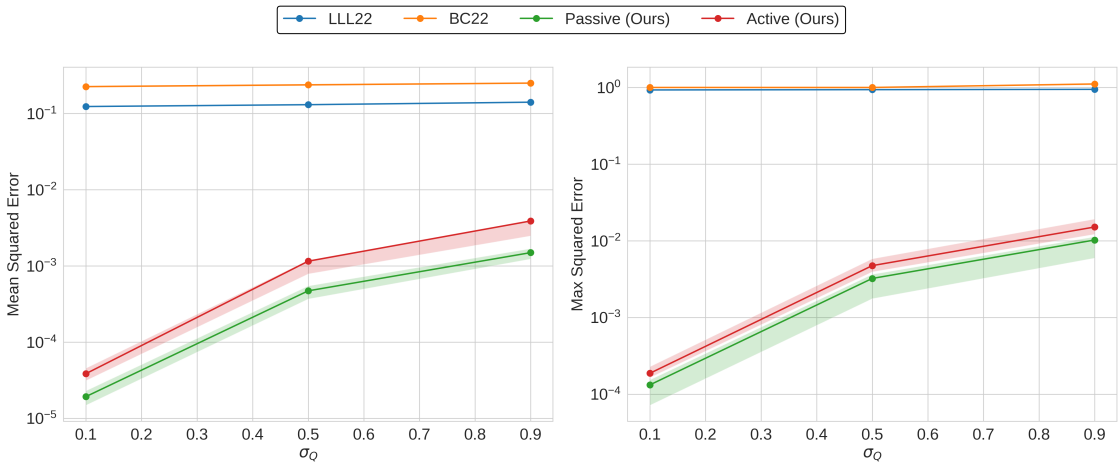


Figure 3.5: We test the effect of growing the target additive noise parameter σ_Q.

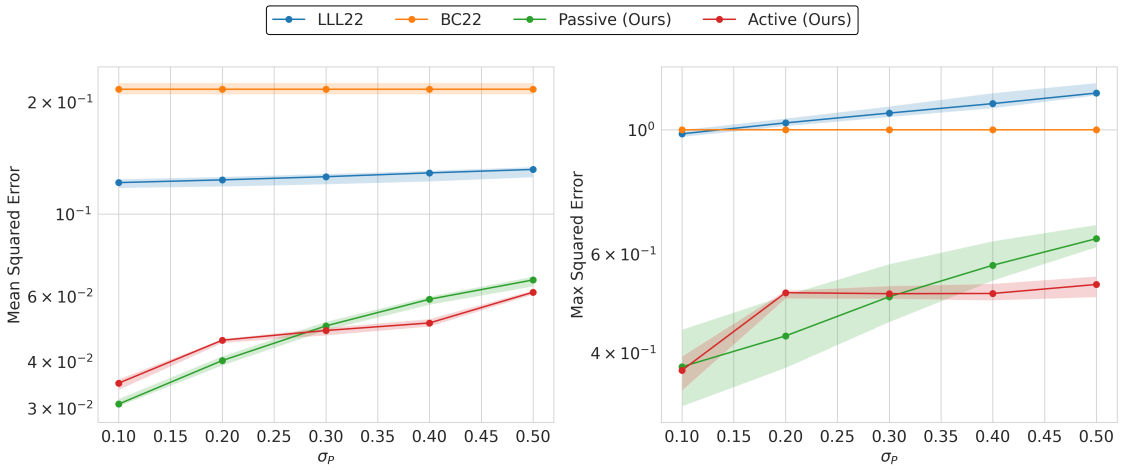


Figure 3.6: We test the effect of growing the target additive noise parameter σ_P. Each entry of P is observed with i.i.d. additive noise $\mathcal{N}(0, \sigma_P^2)$.

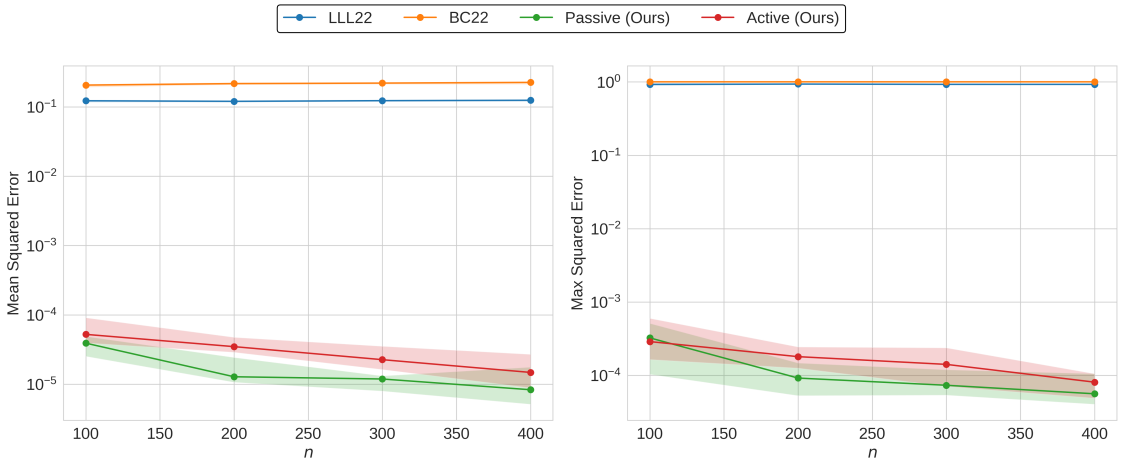


Figure 3.7: We test the effect of growing n for $P, Q \in \mathbb{R}^{300 \times n}$.

3.7.2 Additional Real-World Experiments

We first display the weighted adjacency matrices for P, Q for the metabolic networks setting of Section 5.6 as Figure 3.10 and Figure 3.11. It is evident that the edge weights show significant skew. Note that the colorbar for both visualizations is logarithmically scaled.

Next, we report mean-squared error for the same experimental settings discussed in Section 5.6. Figure 3.13 shows the results for gene expression. Figure 3.12 shows the results for metabolic data; notably, despite poor performance in max-squared error, the passive sampling estimator is reasonably good in mean-squared error, although not as good as the active sampling estimator.

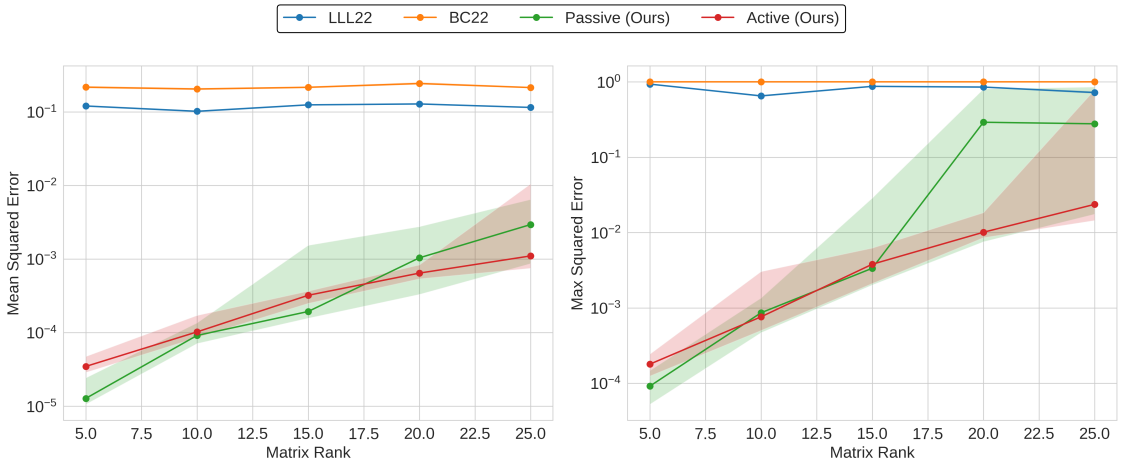


Figure 3.8: We test the effect of rank.

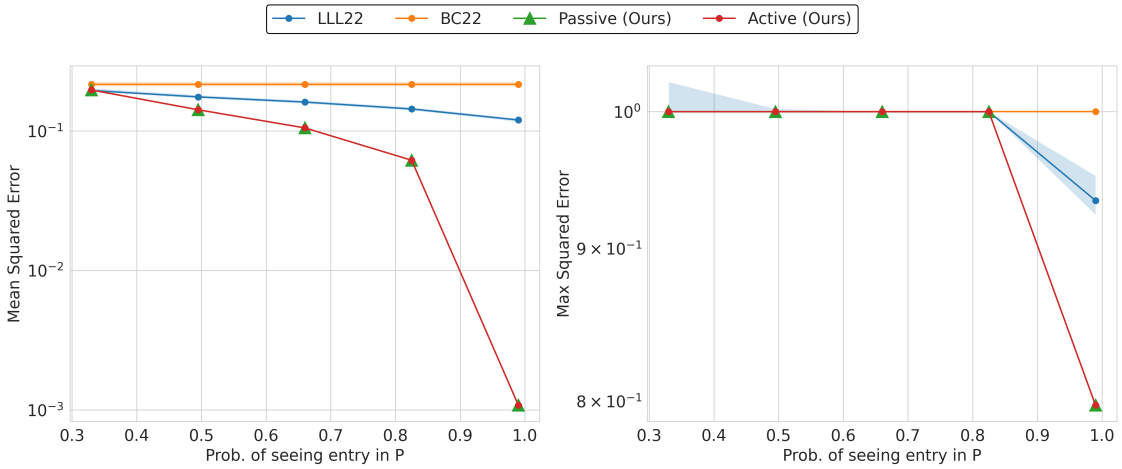


Figure 3.9: We test the effect of masking entries of P in a Missing Completely-at-Random setup with probability p. Note that the errors for active and passive sampling are almost identical, so we use different markers (circle and triangle resp.) to distinguish them. We see that our methods do better in mean-squared error (left) while max error is poor for all methods (right).

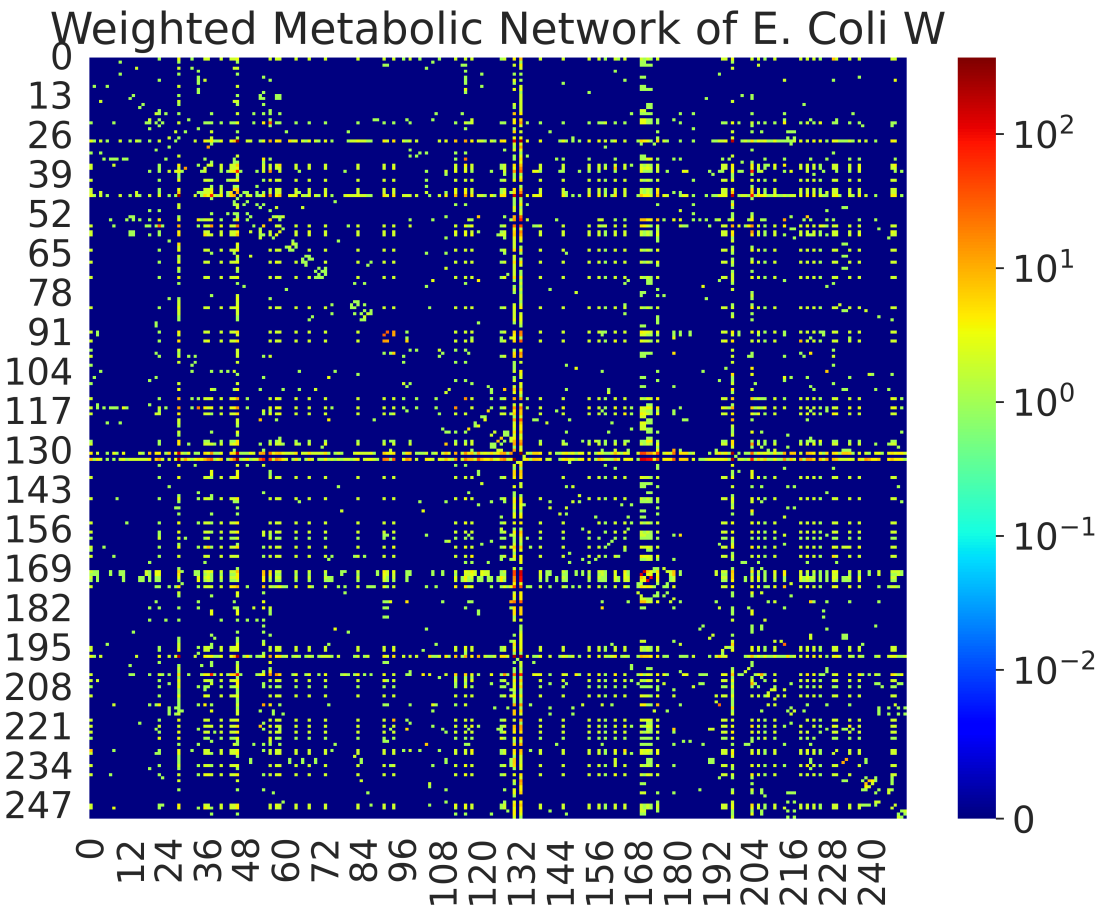


Figure 3.10: The source matrix P in the setting of Figure 3.3.

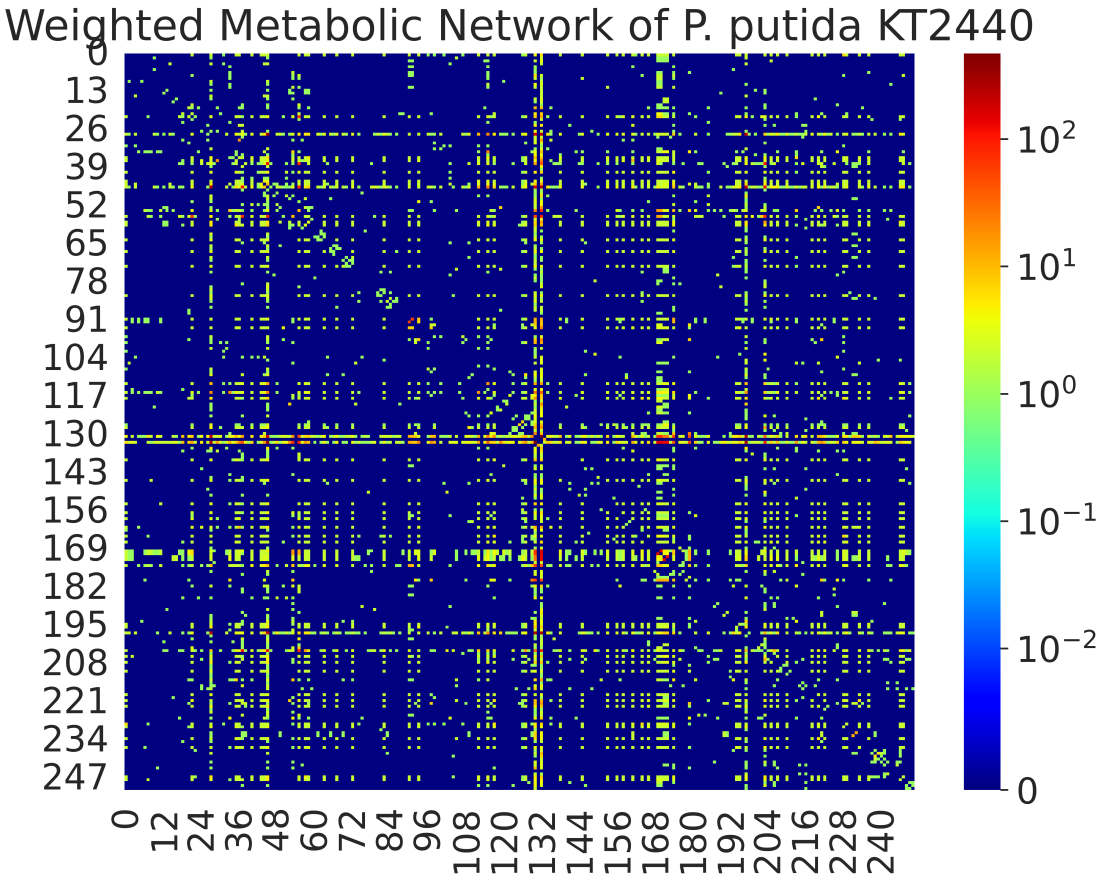


Figure 3.11: The target matrix Q in the setting of Figure 3.3.

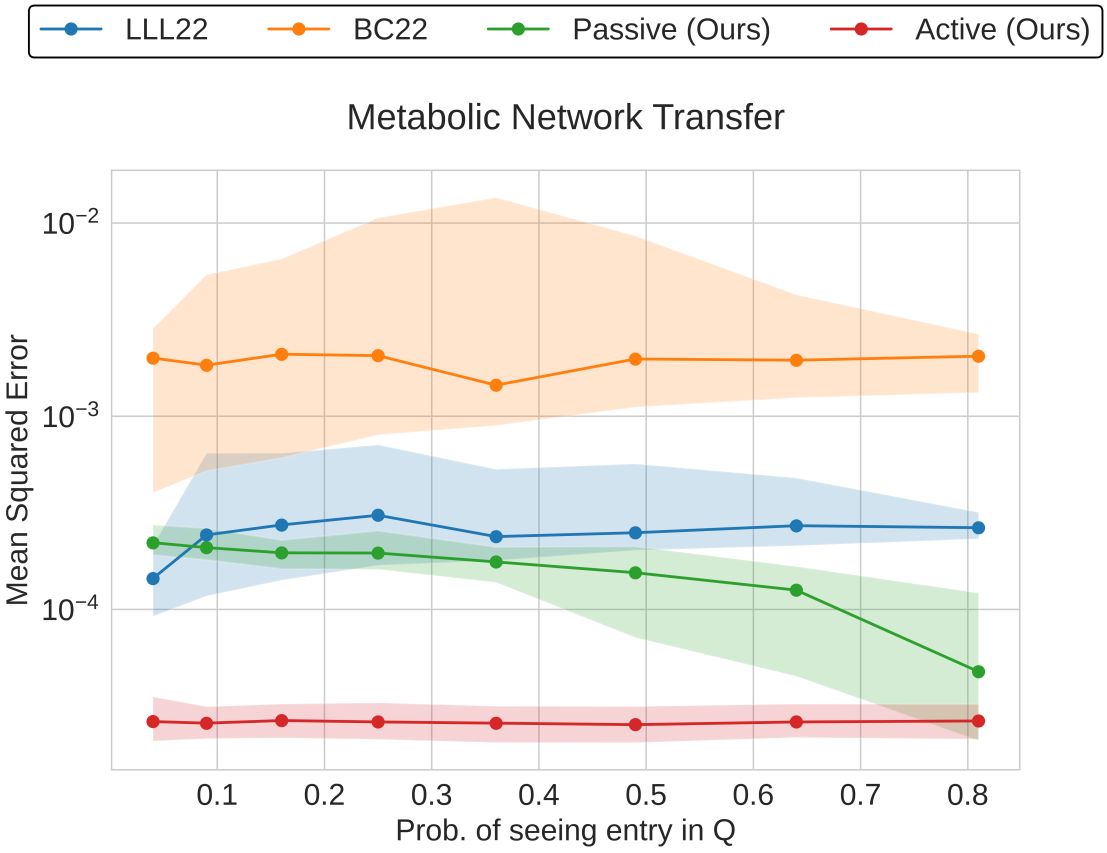


Figure 3.12: The mean-squared error of each $\hat{Q} - Q$ in the setting of Figure 3.3.

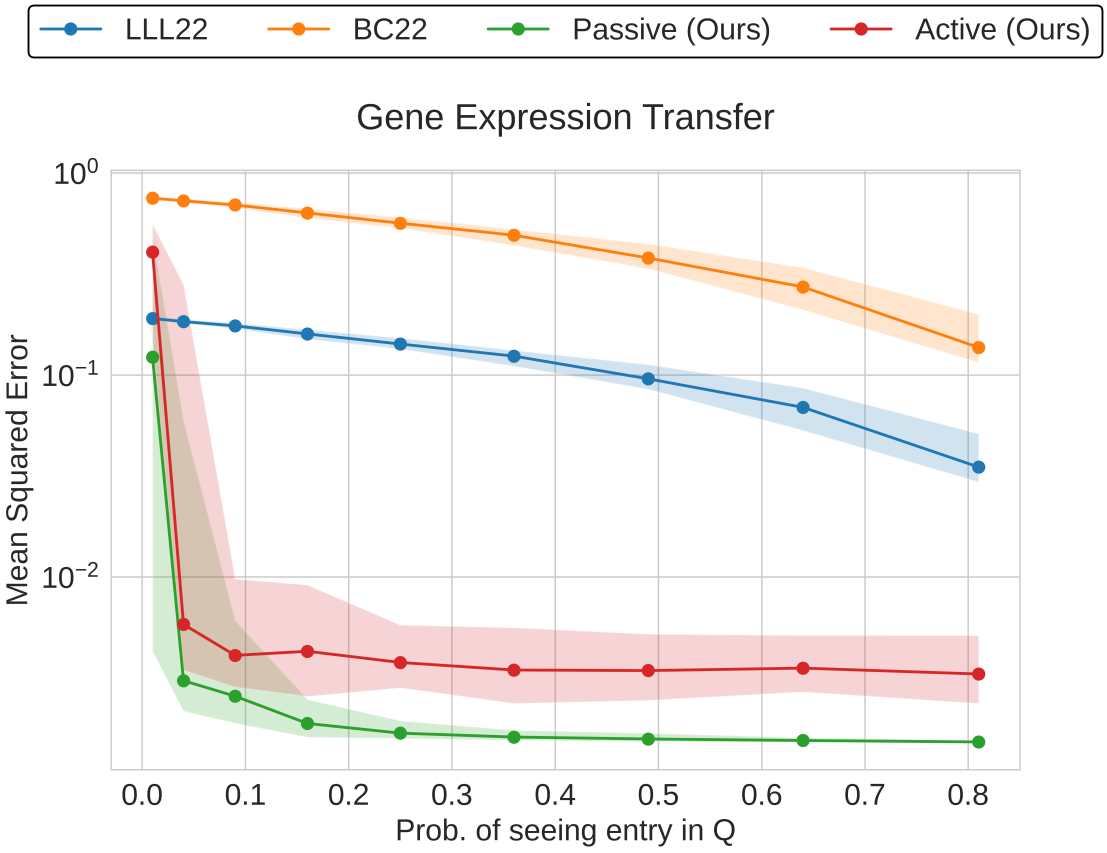


Figure 3.13: The mean-squared error of each $\hat{Q} - Q$ in the setting of Figure 3.2.

Chapter 4: Dynamic, Incentive-Aware Models of Financial Networks

4.1 Introduction

The financial crisis of 2008 showed the need for mitigating systemic risks in the financial system. There has been much recent work on categorizing such risks (Elliott et al., 2014; Glasserman and Young, 2015, 2016; Birge, 2021; Jackson and Pernoud, 2021). While the causes of systemic risk are varied, they often share one feature. This shared feature is the network of interconnections between firms via which problems at one firm spread to others. One example is the weighted directed network of debt between firms. If one firm defaults on its debt, its creditors suffer losses. Some creditors may be forced into default, triggering a default cascade (Eisenberg and Noe, 2001). Another example is the implicit network between firms holding similar assets. Sales by one firm can lead to mark-to-market valuation losses at other firms. These can snowball into fire sales (Caballero and Simsek, 2013; Cont and Minca, 2016; Feinstein, 2020; Feinstein and Søjmark, 2021).

The structure of inter-firm networks plays a vital role in the financial system. Small changes in network structure can lead to jumps in credit spreads in Over-The-Counter (OTC) markets (Eisfeldt et al., 2021). Network density, diversification, and inter-firm cross-holdings can affect how robust the networks are to shocks and how such shocks propagate (Elliott et al., 2014; Acemoglu et al., 2015). The network structure also affects the design of regulatory interventions (Papachristou and Kleinberg, 2022; Amini et al., 2015; Calafiore et al., 2022; Galeotti et al., 2020).

Despite its importance, many prior works use simplistic descriptions of the network structure. For instance, they often assume that the network is fixed and observable. But only regulators may have access to the entire network. Furthermore, shocks or regulatory interventions can change the network. Others assume that the network belongs to a general class. For instance, Caballero and Simsek (2013) assume a ring network between banks. Amini et al. (2015) derive tractable optimal interventions for core-periphery networks. But financial

The content of this chapter appeared in Operations Research 2024, and can be cited as Jalan et al. (2024a).

networks can exhibit complex structure (Peltonen et al., 2014; Eisfeldt et al., 2021). Leverage levels, size heterogeneity, and other factors can affect the network topology (Glasserman and Young, 2016). Hence, there is a need for models to help reason about financial networks.

In this paper, we design a model for a weighted network of contracts between agents, such as firms, countries, or individuals. The contracts can be arbitrary, and the edge weights denote contract sizes. In designing the model, we have two main desiderata. First, the model must account for heterogeneity between firms. This follows from empirical observations that differences in dealer characteristics lead to different trade risk exposures in OTC markets (Eisfeldt et al., 2021). Second, each firm seeks to maximize its utility and selects its contract sizes accordingly. In effect, each firm tries to optimize its portfolio of contracts (Markowitz, 1952). The model must reflect this behavior. From this starting point, we ask the following questions:

1. How does a network emerge from interactions between heterogeneous utility-maximizing firms?
2. How does the network respond to regulatory interventions?
3. How can the network structure inform the beliefs that firms hold about each other?

Next, we review the relevant literature.

Imputing financial networks. We often have only partial information about the structure of a financial network. For example, we may know the total liability of each bank in a network. From this, we want to reconstruct all the inter-bank liabilities (Squartini et al., 2018). One approach is to pick the network that minimizes the Kullback-Leibler divergence from a given input matrix (Upper and Worms, 2004). Mastromatteo et al. (2012) use message-passing algorithms, while Gandy and Veraart (2017) use a Bayesian approach. But such random graph models often do not reflect the sparsity and power-law degree distributions of financial networks (Upper, 2011). Furthermore, these models do not account for the utility-maximizing behavior of firms.

General-purpose network models. The simplest and most well-explored network model is the random graph model (Gilbert, 1959; Erdös and Rényi, 1959). Here, every pair of nodes

is linked independently with probability p. Generalizations of this model allow for different degree distributions and edge directionality (Aiello et al., 2000). Exponential random graph models remove the need for independence, but parameter estimation is costly (Frank and Strauss, 1986; Wasserman and Pattison, 1996; Hunter and Handcock, 2006; Caimo and Friel, 2011). Several models add node-specific latent variables to model the heterogeneity of nodes. For example, in the Stochastic Blockmodel and its variants, nodes are members of various latent communities. The community affiliations of two nodes determine their probability of linkage (Holland et al., 1983; Chakrabarti et al., 2004; Airoldi et al., 2008; Mao et al., 2018). Instead of latent communities, Hoff et al. (2002a) assign a latent location to each node. Here, the probability of an edge depends on the distance between their locations.

All the latent variable models assume conditional independence of edges given the latent variables. But in financial networks, contracts between firms are not independent. Two firms will sign a contract only if the marginal benefit of the new contract is higher than the cost. This cost/benefit tradeoff depends on all other contracts signed with other firms. Unlike our model, existing general-purpose models do not account for such utility-maximization behavior.

Network games. Here, the payoffs of nodes are dependent on the actions of their neighbors (Tardos, 2004). One well-studied class of network games is linear-quadratic games, with linear dynamics and quadratic payoff functions. Prior work has explored the stability of Nash equilibria (Guo and De Persis, 2021) and algorithms to learn the agents' payoff functions (Leng et al., 2020a). But our model does not yield a linear-quadratic game except in exceptional cases. Instead, our process involves non-linear rational functions of the beliefs of firms. Thus, our setting differs from linear-quadratic games. Recently, network games have been extended to settings where the number of players tends to infinity (Carmona et al., 2022). However, we only consider finite networks.

Games to form networks. Several works study the stability of networks. In a pairwise-stable network, no pair of agents want to form or sever edges. This may be achieved via side-payments between agents, which our model also uses (Jackson and Wolinsky, 2003). Pairwise stability has been extended to strong stability for networks (Jackson and Van den Nouweland, 2005), and also to weighted networks with edge weights in $[0, 1]$ (Bich and Morhaim, 2020; Bich and Teteryatnikova, 2022). We introduce an analogous notion called

Higher-Order Nash stability against any deviating coalition. However, the weights in our network are not bounded in $[0, 1]$ and can be negative. Furthermore, our edge weights denote contract size, requiring agreement from both parties. In contrast, prior works typically interpret edge weights as the engagement level in an ongoing interaction.

Sadler and Golub (2021) study a network game with endogenous network formation, whose stable points are both pairwise stable and Nash equilibria. We show similar results for our model. But they consider unweighted networks and focus on the case of separable games. In our setting, this corresponds to the case where all firms are uncorrelated. But in financial networks, correlations are widespread and help firms diversify their contracts.

Several authors study the effect of exogenous inputs on production networks (Herskovic, 2018; Elliott et al., 2022a). Acemoglu and Azar (2020) also model endogenous network formation but differ from our approach. Prices in their model equal the minimum unit cost of production. For us, prices are determined by pairwise negotiations between firms. Also, each firm in their model only considers a discrete set of choices among possible suppliers. In our model, firms can choose both their counterparties and the contract sizes.

Risk-sharing and exchange economies. The pricing of risk is a well-studied problem (Arrow and Debreu, 1954; Bühlmann, 1980, 1984; Tsanakas and Christofides, 2006; Banerjee and Feinstein, 2022). Most models typically price risk via a global market. However, in our model, all contracts are pairwise, and the contract terms and payments between a buyer and seller are bespoke. There is no global contract or global market price. Since contracts are pairwise, each firm under our model must consider counterparty risks and the correlations between them. A firm i may make large payments and accept a *negative* reward for a contract with firm j to diversify the risk from contracts with other firms. Finally, in our model, agents can hedge their risk by betting against one another. In contrast, Bühlmann equilibria always result in comonotonic endowments, which firms cannot use as hedges for each other (Banerjee and Feinstein, 2022; Yaari, 1987).

Network valuation adjustment. Some recent works price the risk due to exposure to the entire financial network (Banerjee and Feinstein, 2022; Feinstein and Søjmark, 2022). The network is usually treated as exogenous and fully known to all firms. In contrast, we consider endogenous network formation resulting from pairwise interactions between firms. The network valuation algorithm of Barucca et al. (2020) works with incomplete information,

but is not designed for network formation, and it needs firms to share information not required to form their contracts.

Properties of equilibria. Another line of work considers the efficiency or social welfare of equilibria (Jackson and Pernoud, 2021; Elliott and Golub, 2022). Galeotti et al. (2020) show that welfare-maximizing interventions rely mainly on the top or bottom eigenvectors of the network. Elliott et al. (2022a) show an efficiency-stability tradeoff for their model of supply network formation. Like prior work, we show that stable equilibria exist and are non-dominated. But our emphasis is on potentially valuable insights for regulators and firms. For instance, we show a negative result about the ability of regulators to infer the causes of changes to the network structure. The linkage between firms' utilities and their beliefs, and its effect on stability, is not considered in prior work.

4.1.1 Our Contributions

We develop a new network model of contracts between heterogeneous agents, such as firms, countries, or individuals. Each agent aims to maximize a mean-variance utility parametrized by its beliefs. But for two agents to sign a contract, both must agree to the contract size. For a stable network, all agents must agree to all their contracts. We show that such constraints are solvable by allowing agents to pay each other. By choosing prices appropriately, every agent maximizes its utility in a stable network.

Characterization of stable networks (Section 4.2): We show that unique stable networks exist for almost all choices of agents' beliefs. These networks are robust against actions by cartels, a condition that we call Higher-Order Nash Stability. The agents can also converge to the stable network via iterative pairwise negotiations. The convergence is exponential in the number of iterations. Hence, the stable network can be found quickly. Finally, we show how to infer the agents' beliefs by observing network snapshots over time, under certain conditions.

The limits of regulation (Section 4.3): A financial regulator can observe the entire network but not the agents' beliefs. Suppose firm i changes its beliefs about firm j. Then the contract size between i and j will change. Indirectly, other contracts will change too. We show empirically that in realistic settings, the indirect effects can be as significant as the direct effects. In such cases, the regulator cannot infer the underlying cause of changes in the

network. Similarly, suppose the regulator intervenes with one firm, affecting its beliefs. The resulting network changes need not be localized to that firm's neighborhood in the network. Thus, targeted interventions can have strong ripple effects. Broad-based interventions aimed at increasing stability can also have adverse effects. For instance, increasing margin requirements on contracts may even increase some contract sizes.

Outlier detection by firms (Section 4.4): A firm i can observe its contracts with counterparties but not the entire network. Suppose another firm j (say, a real-estate firm) has beliefs that are very different from its peers. Then, we prove that under certain conditions, j's contract size with i is also an outlier compared to other real-estate firms. So, firm i can use the network to detect outliers and update its beliefs. But suppose all real-estate firms change their beliefs. This changes all their contract sizes without creating outliers. We show that i cannot determine the cause of this change. For example, firm i would observe the same change whether all real-estate firms had become more risk-seeking or profitable. However, firm i may want to increase its exposure if they are more profitable but reduce exposure if they are more risk-seeking. Since the data cannot identify the proper action, firm i remains uncertain. Exogenous, seemingly insignificant information may persuade firm i one way or another. Thus, minor news may trigger drastic changes in the network.

4.2 The Proposed Model

We consider a *weighted* network $W \in \mathbb{R}^{n\times n}$ between n agents (such as firms, countries, or individuals). The element W_{ij} represents the size of a contract between agents i and j. We make no assumptions about the content of the contract. For instance, the contract could be a interest rate swap, a stock swap, or an insurance contract. We assume that each pair of firms can form a contract of a standard type, and negotiate only on the contract size and price (discussed below). Since contracts need mutual agreement, $W_{ij} = W_{ji}$. We take W_{ii} to represent i's investment in itself. Note that a negative contract ($W_{ij} = W_{ji} < 0$) is a valid contract that reverses the content of a positive contract. For example, if a positive contract is a derivative trade between two firms, the negative contract swaps the roles of the two firms.

Let $\boldsymbol{w}_i$ denote the i^{th} column of W (i.e., $\boldsymbol{w}_{i;j} = W_{ji}$ for all j). Each agent i would prefer to set its contract sizes $\boldsymbol{w}_i$ to maximize its utility. But other agents will typically have different preferences. So, to achieve an agreement about the contract size W_{ij}, agents i and j

can agree on a price for the contract. For example, i may agree to pay j an amount $P_{ji} \cdot W_{ji}$ in cash at the beginning of the contract. Since payments are zero-sum and $W_{ji} = W_{ij}$, we must have $P_{ji} = -P_{ij}$. We do not model how firms raise funds to pay the price.

Each contract yields a stochastic payout, and agents have beliefs about these payouts. We represent agent i's beliefs by a vector $\boldsymbol{\mu}_i$ of expected returns and a covariance matrix $\Sigma_i \succ 0$. Thus, Σ_i represents firm i's perceived risk of trading with other firms, and includes both contract-specific risk and counterparty risk. Note that we do *not* assume that the contracts are zero-sum or that the beliefs are correct, even approximately. Thus, the overall expected return from all contracts of i is $\boldsymbol{w}_i^T(\boldsymbol{\mu}_i - P\boldsymbol{e}_i)$, and the variance of the overall return is $\boldsymbol{w}_i^T \Sigma_i \boldsymbol{w}_i$. We assume that each agent has a mean-variance utility (Markowitz, 1952):

$$\text{agent } i\text{'s utility } g_i(W, P) := \boldsymbol{w}_i^T(\boldsymbol{\mu}_i - P\boldsymbol{e}_i) - \gamma_i \cdot \boldsymbol{w}_i^T \Sigma_i \boldsymbol{w}_i, \tag{4.1}$$

where $\gamma_i > 0$ is a risk-aversion parameter. In practice, we expect the set $\{\gamma_i\}_{i \in [n]}$ to be not too heterogeneous (Metrick, 1995; Kimball et al., 2008; Ang, 2014; Paravisini et al., 2017). Note that Eq. (5.1) ignores costs for contract formation; we will consider these in Section 4.3.1. Also, we assume that P_{ji} does not change the perceived risk.

Example 4.2.1 (Insurance Contract)**.** *Suppose firm i buys fire insurance from insurer j. Then, $\boldsymbol{\mu}_{i;j}$ is the buyer's expected insurance payout minus the insurance premium. The expected payout depends on the probability of a fire, for which the buyer and insurer may have different estimates. Also, the insurance contract is negatively correlated with the buyer's other contracts (reflected in Σ_i). This is because the buyer gains a payout from the insurer in case of a fire, but incurs losses on other contracts. Hence, the buyer i may be willing to accept a contract with negative expected reward, and even pay a higher-than-usual premium P_{ji} per contract.*

Example 4.2.2 (Interest rate swap contract)**.** *Suppose firm i makes fixed-rate payments to firm j, and receives floating-rate payments in return. Then, $\boldsymbol{\mu}_{i;j}$ is the expected net present value of these payments for i from a standard unit-sized contract. This value depends on i's forecast of future interest rates and need for floating-rate income, e.g., to match future liabilities. Hence, it may be quite different from $\boldsymbol{\mu}_{j;i}$. Also, the firms agree to a price $P_{ij} = -P_{ji}$ per contract. If $P_{ij} > 0$, then firm j must pay firm i the price $P_{ij} \cdot W_{ij}$; if $P_{ij} < 0$, then firm i makes the payment.*

Example 4.2.3 (Loan contract)**.** *Suppose borrower i takes a loan of size W_{ij} from lender j. Then, $\boldsymbol{\mu}_{j;i} \cdot W_{ij}$ represents the lender j's expected value for this loan. The expected value depends on the repayment schedule, the collateral, j's estimate of the probability of default, the recovery rate in case of default, etc. The borrower's expected value $\boldsymbol{\mu}_{i;j} \cdot W_{ij}$ depends on the planned use of this loan. For example, if the borrower wants the loan to purchase equipment, $\boldsymbol{\mu}_{i;j}$ is the net present value of expected extra profits due to that equipment. Hence, $\boldsymbol{\mu}_{i;j}$ may not be a function of $\boldsymbol{\mu}_{j;i}$. Now, the borrower and lender must settle on a contract price to reach an agreement on the contract size. If the standard loan contract requires the lender to give cash to the borrower at the beginning of the contract, this loan amount can be adjusted for the price. Otherwise, if the borrower firm needs to pay the price, it must arrange a separate bridge loan.*

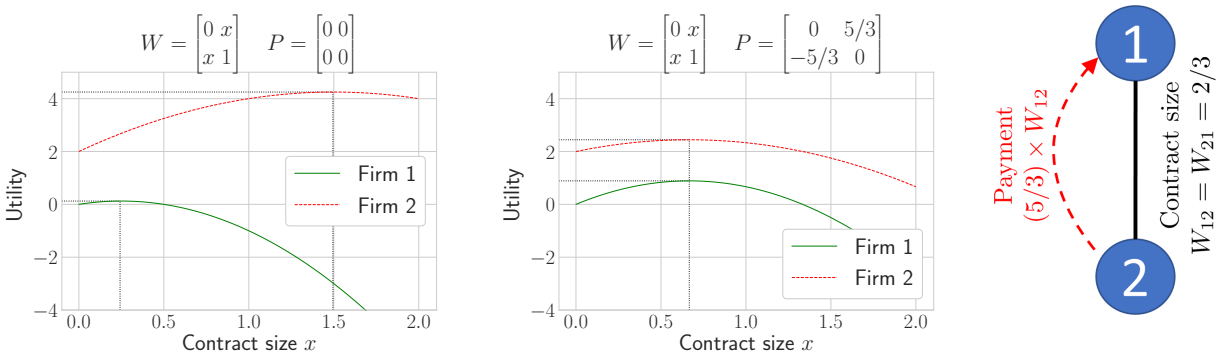
The model above allows contracts between all pairs of agents. But some edges may be prohibited due to logistical or legal reasons. For each agent i, let $J_i \subseteq [n]$ denote the ordered set of agents with whom i can form an edge. So, if $k \notin J_i$ (and hence $i \notin J_k$), we have $W_{ik} = W_{ki} = P_{ik} = P_{ki} = 0$. Similarly, if $i \notin J_i$, then self-loops are prohibited ($W_{ii} = P_{ii} = 0$). We will encode these constraints in the binary matrix $\Psi_i \in \mathbb{R}^{|J_i| \times n}$ where $\Psi_{i;jk} = 1$ if k is the j^{th} element of J_i, and $\Psi_{i;jk} = 0$ otherwise. In other words, Ψ_i is obtained from I_n by deleting the rows corresponding to the prohibited counterparties of i. Thus, for any $\boldsymbol{v} \in \mathbb{R}^n$, $\Psi_i \boldsymbol{v}$ selects the elements of $\boldsymbol{v}$ corresponding to J_i. If all edges are allowed, we have $\Psi_i = I_n$ for all i.

Definition 4.2.4 (Network Setting)**.** *A* network setting $(\boldsymbol{\mu}_i, \gamma_i, \Sigma_i, \Psi_i)_{i \in [n]}$ *captures the beliefs and constraints of n agents. When there are no constraints (i.e., all edges are allowed), we drop the $\Psi_i = I_n$ terms to simplify the exposition. Finally, we will use $M \in \mathbb{R}^{n \times n}$ to denote a matrix whose i^{th} column is $\boldsymbol{\mu}_i$, and Γ to denote a diagonal matrix with $\Gamma_{ii} = \gamma_i$.*

4.2.1 Characterizing Stable Points

In the above model, every agent tries to optimize its own utility (Eq.(5.1)). We now characterize the conditions under which selfish utility-maximization leads to a stable network.

Definition 4.2.5 (Feasibility)**.** *A tuple (W, P) is feasible if $W = W^T$, $P = -P^T$, and W and P obey the constraints encoded in $(\Psi_i)_{i \in [n]}$.*



(a) No payments allowed (b) Firm 2 pays 5/3 per contract (c) Network

Figure 4.1: *Example of a stable point for a borrower (Firm 1) and a lender (Firm 2):* (a) When the borrower cannot pay the lender an additional payment, the firms may be unable to agree to a contract, even if trading improves their utilities. (b) By allowing for contract-specific payments, both firms can agree on a contract size. In effect, the borrower (Firm 2) shares its utility with the lender (Firm 1) to achieve agreement. (c) The stable network is shown.

Definition 4.2.6 (Stable point)**.** *A feasible (W, P) is stable if each agent achieves its maximum possible utility given prices P:*

$$g_i(W, P) = \max_{\text{feasible}(W', P) \text{ under } \{\Psi_i\}} g_i(W', P) \quad \forall i \in [n].$$

Example 4.2.7. *Suppose we only have two firms with the following setting:*

$$\text{mean beliefs } M = \begin{bmatrix} 0 & 3 \\ 1 & 4 \end{bmatrix}$$
$$\text{covariance } \Sigma_1 = \Sigma_2 = \begin{bmatrix} 1 & 0 \\ 0 & 2 \end{bmatrix}$$
$$\text{risk aversion } \gamma_1 = \gamma_2 = 1.$$

So, both firms perceive a benefit from trading ($M_{12} > 0, M_{21} > 0$). If trading is disallowed, the optimum W is diagonal with $W_{11} = 0$ and $W_{22} = 1$ (and P is the zero matrix). The corresponding utilities are 0 for firm 1 and 2 for firm 2. Suppose we allow trading but do not allow pricing (Figure 4.1a). Then, the two firms can each improve their utility by trading, but achieve their optimum utilities at different contract sizes. Hence, they may be unable to agree to a contract. In Figure 4.1b, firm 2 pays firm 1 a specially chosen price of 5/3 per unit contract. At this price, both firms achieve their optimum utilities at the same contract size

$W_{12} = W_{21} = 2/3$. *Hence, they can agree to a contract. By paying the price, firm* 2 *shares some of its utility with firm* 1 *to achieve agreement on the contract. This choice of* W *and* P *is a stable point (Figure 4.1c). The following results show that this is the* only *stable point.*□

Define $Q_i = \Psi_i^T(2\gamma_i\Psi_i\Sigma_i\Psi_i^T)^{-1}\Psi_i$. When all edges are allowed, $\Psi_i = I_n$ and $Q_i = (2\gamma_i\Sigma_i)^{-1}$. Let $F = \{(i,j) : 1 \leq i < j \leq n, \Psi_i \boldsymbol{e}_j \neq \mathbf{0}\}$ denote the ordered pairs $i < j$ where P_{ij} is allowed to be non-zero. Note that $|F| \leq n(n-1)/2$. For any $n \times n$ matrix X, let $\text{uvec}(X)_F \in \mathbb{R}^{|F|}$ be a vector whose entries are the ordered set $\{X_{ij} \mid (i,j) \in F\}$.

Theorem 4.2.8 (Existence and Uniqueness of Stable Point)**.** *Define* $n \times n$ *matrices* A, $B_{(i,j)}$, *and* $C_{(i,j)}$ *as follows:*

$$A_{ij} = \boldsymbol{e}_i^T Q_j M \boldsymbol{e}_j, B_{(i,j)} = \boldsymbol{e}_i \boldsymbol{e}_j^T Q_i,$$
$$C_{(i,j)} = (B_{(i,j)} - B_{(j,i)}) - (B_{(i,j)} - B_{(j,i)})^T.$$

Let Z_F *be the* $|F| \times |F|$ *matrix whose rows are the ordered sets* $\{\text{uvec}(C_{(i,j)})_F \mid (i,j) \in F\}$. *Then, we have the following:*

1. *A stable point* (W, P) *under* $\{\Psi_i\}$ *exists if and only if* $\text{uvec}(A - A^T)_F$ *lies in the column space of* Z_F.
2. *If a stable point* (W, P) *exists, then* $Z_F \text{uvec}(P)_F = \text{uvec}(A - A^T)_F$.
3. *A unique stable point always exists if* Z_F *is full rank.*

Theorem 4.2.8 is proved in Section 4.6.1. When the Σ_i are random variables, we give a simple sufficient condition that a stable point exists and is unique with probability 1 (see Sections 4.6.3 and 4.6.4). Also, Appendix 4.6.2 provides closed-form formulas for the stable point when all agents have the same covariance ($\Sigma_i = \Sigma$ for all $i \in [n]$). This occurs when the risk of a contract is primarily counterparty risk (so $\Sigma_{i;jk}$ depends on j and k, not i) and there is reliable public data on such risks (say, via credit rating agencies).

Next, we consider some properties of the stable point. For two feasible tuples (W_1, P_1) and (W_2, P_2), let (W_2, P_2) *dominate* (W_1, P_1) if for all $i \in [n], g_i(W_1, P_1) \leq g_i(W_2, P_2)$, with at least one inequality being strict.

Theorem 4.2.9 (Stable points cannot be dominated)**.** *Suppose a stable point* (W, P) *exists. Then, there is no feasible* (W', P') *that dominates* (W, P)*.*

The proofs of Theorem 4.2.9 and all subsequent claims are provided in Section 4.6.

The stable point obeys a strong form of robustness that we call *Higher-Order Nash Stability*. This strengthens the notions of *pairwise stability* (Hellmann, 2013) and *pairwise Nash* (Calvó-Armengol and Ilkılıç, 2009; Sadler and Golub, 2021) by allowing for agent coalitions, instead of just considering pairs of agents. It is also closely related to the concept of *Strong Nash equilibrium*, which strengthens Nash equilibrium by requiring that no subset of agents can deviate at equilibrium without at least one agent being worse off (Mazalov and Chirkova, 2019).

Definition 4.2.10 (Agent Action)**.** *At a given feasible point* (W, P)*, an "action" by agent* i *is the ordered set* $(w'_{i,j}, p'_{i,j})_{j \in J_i}$*, where* $J_i \subseteq [n]$ *is the set of permissible edges for agent* i*. The action represents a set of proposed changes to* i*'s existing contracts. Each agent* $j \in J_i$ *responds as follows:*

1. *If the new* (w'_{ij}, p'_{ij}) *raises* j*'s utility, then* j *agrees to the revised contract and price.*
2. *Otherwise,* i *must either keep the existing contract or cancel it (*$w_{ij} = p_{ij} = 0$*). We assume that* i *cancels the contract if and only if this strictly increases* i*'s utility.*

We call the shifted (W', P') *the* resulting network*.*

Definition 4.2.11 (Higher-Order Nash Stability)**.** *A feasible* (W, P) *is Higher-Order Nash Stable if:*

1. Nash equilibrium*: No agent* i *has an action such that the resulting network* (W', P') *is strictly better for* i*.*
2. Cartel robustness*: For any proper subset* $S \subset [n]$ *of agents, there is no feasible point* (W', P') *that differs from* (W, P) *only for indices* $\{i, j\}$ *with* $i \in S, j \in S$ *such that all agents in* S *have higher utility under* (W', P') *than* (W, P)*.*

Theorem 4.2.12 (Higher-Order Nash Stability)**.** *Any stable point* (W, P) *is Higher-Order Nash Stable.*

4.2.2 Finding the Stable Point via Pairwise Negotiations

To compute the stable point in Theorem 4.2.8, we must know the beliefs of all agents. But in practice, contracts are set iteratively by negotiations among pairs of agents. We will now formalize the process of pairwise negotiations and characterize the conditions under which such negotiations can converge to the stable point.

We propose a multi-round pairwise negotiation process. In round $t+1$, every pair of agents i and j update the price $P_{ij}(t)$ to $P_{ij}(t+1)$ (and hence $P_{ji}(t)$ to $P_{ji}(t+1)$) as follows. First, they agree to a price P'_{ij} between themselves, *assuming optimal contract sizes with all other agents at the current prices* $P(t)$. In other words, we assume that the other agents will accept the prices in $P(t)$ and the contract sizes preferred by i and j. Under this condition, P'_{ij} is the price at which i's optimal contract size with j is also j's optimal size with i. We provide an explicit formula for P'_{ij} in Section 4.6.7. All pairs of agents calculate these prices *simultaneously*. We create a new price matrix P' from these prices. Then, we set $P(t+1) = (1-\eta)P(t) + \eta P'$, where $\eta \in (0,1)$ is a dampening factor chosen to achieve convergence. Algorithm 3 shows the details.

Algorithm 3 Pairwise Negotiations

Input: $\eta \in (0,1)$

$t \leftarrow 0$

$P(0) \leftarrow$ any skew-symmetric matrix

while $P(t)$ *has not converged* **do**

  $\forall i, j \in [n], P'_{ij} \leftarrow$ pairwise-negotiated price for (i,j) (Section 4.6.7)

  $P(t+1) \leftarrow (1-\eta)P(t) + \eta P'$

  $t \leftarrow t+1$

end

Example 4.2.13 (Pairwise negotations for loan contracts.)**.** *Consider a* 3*-firm loans network containing a national bank (firm 1), local bank (firm 2), and local firm (firm 3). Suppose that the local firm cannot access the national bank, so the edge between firms* 1 *and* 3 *is prohibited.*

The other parameters are:

$$\Sigma_1 = \Sigma_2 = \Sigma_3 = \begin{bmatrix} 1 & 0.25 & 0.75 \\ 0.25 & 1 & 0.6 \\ 0.75 & 0.6 & 1 \end{bmatrix},$$

$$M = \begin{bmatrix} 0 & 0.9 & 0.9 \\ 0.75 & 0 & 0.95 \\ 0.5 & 0.8 & 0 \end{bmatrix}, \gamma_1 = \gamma_2 = \gamma_3 = 1.$$

Figure 4.2 shows how pairwise negotiations via Algorithm 3 converge to the stable network.

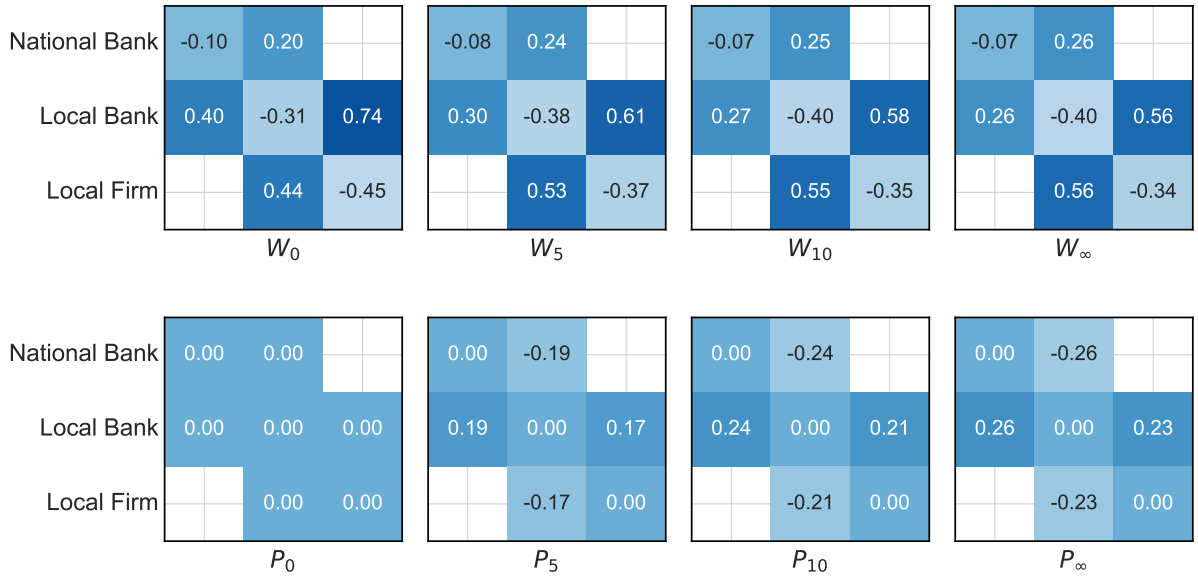


Figure 4.2: *Pairwise negotiations for the setting of Example 4.2.13:* The contracts matrix W_t and payments matrix P_t after $t = 0, 5, 10$ steps of Algorithm 3 ($\eta = 0.5$) converge to the stable point $(W, P) = (W_\infty, P_\infty)$. Cells corresponding to forbidden edges are empty.

Now, we will show that Algorithm 3 converges. First, we define *global asymptotic stability* (following Callier and Desoer (1994)).

Definition 4.2.14 (Global Asymptotic Stability)**.** *The pairwise negotiation process is globally asymptotically stable for a given network setting and dampening factor η if, for any initial price matrix $P(0)$, there exists a matrix $P^\star$ such that the sequence of price matrices $P(t)$ converges to $P^\star$ in Frobenius norm:* $\lim_{t\to\infty} \|P(t) - P^\star\|_F = 0$.

When pairwise negotiations are globally asymptotically stable, the limiting matrix $P^\star$ must be skew-symmetric since each $P(t)$ is skew-symmetric. Also, since prices are updated

whenever two agents disagree on the size of the contract between them, all agents agree on their contract sizes at $P^\star$. Hence, $P^\star$ must be a stable point for the given network setting.

Now, we show that for a range of η, pairwise negotiations are globally asymptotically stable (Section 4.6.9 presents an example).

Theorem 4.2.15 (Convergence Conditions and Rate)**.** *Let Q_i be defined as in Theorem 4.2.8. Define the following $n^2 \times n^2$ matrices:*

$$K := \sum_{r=1}^{n} \boldsymbol{e}_r \boldsymbol{e}_r^T \otimes Q_r + Q_r \otimes \boldsymbol{e}_r \boldsymbol{e}_r^T$$

$$L_{(i-1)n+j,(i-1)n+j} = Q_{i;j,j} + Q_{j;i,i} \quad \forall i, j \in [n]$$

(L is diagonal).

Let $L^\dagger$ denote the pseudoinverse of L, and $(L^\dagger K)\,|_R$ denote the principal submatrix of $L^\dagger K$ containing the rows/columns $(i-1)n+j$ such that the edge (i,j) is not prohibited. Let $\lambda_{max}, \lambda_{min}$ be the largest and smallest eigenvalues of the matrix $(L^\dagger K)\,|_R$ respectively. Let $\eta^* = \frac{2}{\lambda_{max}}$. Then, we have:

1. *For all $\eta \in (0, \eta^*)$, pairwise negotiations with η are globally asymptotically stable.*
2. *For such an η, the convergence is exponential in the number of rounds t:*

$$\|P(t) - P^\star\|_F \leq \frac{\alpha^t}{1-\alpha} \cdot \|P(1) - P(0)\|_F,$$
$$\text{where } \alpha = \max\{|1 - \eta\lambda_{min}|, |1 - \eta\lambda_{max}|\}.$$

Here, $P^\star$ is the stable point to which the negotiation converges.

Remark 4.2.16. *For clarity of exposition we restrict $\eta \in (0,1)$ in Algorithm 3. However, Theorem 4.2.15 shows that we only need $\eta < \eta^*$ for convergence to the stable point.*

4.2.3 Pairwise Negotiations under Random Covariances

So far we have made no assumptions about agents' beliefs. In this section, we analyze the convergence of pairwise negotiations for "data-driven" agents. Specifically, each agent i now *estimates* its covariance matrix. For this section only, we will call the covariance matrix $\hat{\Sigma}_i$ instead of Σ_i to emphasize that it is an estimated quantity.

Suppose each agent i observes m independent data samples. Each sample is a vector of the returns of unit contracts with all n agents. The samples for agent i are collected in a matrix $X_i \in \mathbb{R}^{n\times m}$, with one column per sample. The sample covariance of this data is $\hat{\Sigma}_i$.

We assume that all agents observe samples from the same return distribution, which has covariance Σ. Under a wide range of conditions, $\|\hat{\Sigma}_i - \Sigma\| \to 0$ in probability (Vershynin, 2018a). Hence, at convergence, the maximum allowed dampening rate $\eta^\star$ in Theorem 4.2.15 would be a function of Σ. But for finite sample sizes, each agent's $\hat{\Sigma}_i$ can be different. Hence, the maximum dampening $\hat{\eta}^\star$ may be less than $\eta^\star$. The smaller the $\hat{\eta}^\star$, the worse the rate of convergence of pairwise negotiations. However, even with a few samples, $\hat{\eta}^\star$ is close to $\eta^\star$, as the next theorem shows.

Theorem 4.2.17 (Small Sample Sizes are Sufficient for Fast Convergence)**.** *Suppose that* $\|\Sigma\|, \|\Sigma^{-1}\|, \|\Gamma\|$, *and* $\|\Gamma^{-1}\|$ *are* $O(1)$ *with respect to* n *and all edges are allowed. Also, suppose that each sample column of* X_i *is drawn independently from a* $\mathcal{N}(\mathbf{0}, \Sigma)$ *distribution, and let* $\hat{\boldsymbol{\mu}} = \frac{1}{m}\sum_i X_i$ *and* $\hat{\Sigma}_i := \frac{1}{m-1}\sum_i (X_i - \hat{\boldsymbol{\mu}})(X_i - \hat{\boldsymbol{\mu}})^T$. *Let* $\hat{\eta}^\star$ *be the maximum dampening factor using* $(\hat{\Sigma}_i)_{i\in[n]}$ *as defined in Theorem 4.2.15. Let* $\eta^\star$ *be the dampening factor if* $\hat{\Sigma}_i$ *were replaced by* Σ *for all* i. *If* $m = \lceil n \log n \rceil$, *then for large enough* n, $\hat{\eta}^\star \geq (1 - o(1))\eta^\star$ *with probability at least* $1 - \exp(-\Omega(n))$.

Theorem 4.2.17 shows that data-driven agents using a broad range of dampening factors are still likely to find the stable point via pairwise negotiations. Furthermore, the amount of data they need is comparable to the number of agents (up to a logarithmic factor). We note that if firms use datasets of fixed sizes $m_1, \ldots, m_n$, then the conclusion of Theorem 4.2.17 still holds, as long as $\min_i m_i \geq \lceil n \log n \rceil$. For example, firms might use different look-back periods for covariance estimation.

4.2.4 Inferring Beliefs from the Network Structure

Suppose we are given a network that lies at a unique stable point as defined in Theorem 4.2.8. How can we infer the beliefs of the agents?

Non-identifiability of beliefs. Suppose we are given a network W that is generated using a single covariance $\Sigma_i = \Sigma \succ 0$. We want to infer the agents' beliefs (M, Γ, Σ). By

Corollary 4.6.1,

$$\frac{1}{2}\text{vec}(M + M^T) = (\Gamma \otimes \Sigma + \Sigma \otimes \Gamma)\text{vec}(W).$$

Clearly, the agents' beliefs can only be specified up to an appropriate scaling of M, Γ, and Σ. But even if we specify a scale (e.g., $\text{tr}[\Gamma] = \text{tr}[\Sigma] = 1$), for any valid choice of Γ and Σ we can find a corresponding M. Thus, even in the simple setting of identical covariance and fixed scale, the network W cannot be used to select a unique combination of the parameters (M, Γ, Σ). By a similar argument, we cannot identify the underlying beliefs even if we observe multiple networks generated using the same Σ and Γ (but different M). Thus, we need further assumptions in order to infer beliefs.

Assumption 4.2.18. *Consider a sequence of networks $W(t)$ over timesteps $t \in [T]$. We assume that (a) $\Gamma(t) = I$ and $\Sigma_i(t) = \Sigma$ for all $t \in [T]$, (b) for all $i, j \in [n]$, $M_{ij}(t)$ varies independently according to a Brownian motion with the same parameters for all (i, j), and (c)* $\text{tr}\Sigma = 1$.

The first assumption is motivated by the observations in portfolio theory that errors in mean estimation are far more significant than covariance estimation errors (Chopra and Ziemba, 2013). So, accounting for variations in Σ may be less important than variations in M (but see Remark 4.2.20 below). The homogeneity of risk aversion was noted in Section 4.2, and this justifies setting $\Gamma = I$. The second assumption is common in the literature on pricing models (Geman et al., 2001; Bianchi et al., 2013). The third assumption fixes the scale, as discussed above.

Proposition 4.2.19. *Finding the maximum likelihood estimator of Σ under Assumption 4.2.18 is equivalent to the following Semidefinite Program (SDP):*

$$\begin{aligned} \min_{\Sigma} \sum_{t=1}^{T-1} \left\| \Sigma(W(t+1) - W(t)) + (W(t+1) - W(t))\Sigma \right\|_F^2 \\ s.\ t.\ \Sigma \succeq 0, \text{tr}(\Sigma) = 1. \end{aligned}$$

Remark 4.2.20 (Generalization to time-varying Σ)**.** *Instead of a constant covariance Σ, the time range may be split into intervals, with covariance $\Sigma_{(j)}$ in interval j. Then, we can add a regularizer $\nu \cdot \sum_j \|\Sigma_{(j+1)} - \Sigma_{(j)}\|$ for some $\nu > 0$ to the objective of the SDP to penalize differences between successive covariances. This allows the covariance to evolve while keeping the objective convex. The time intervals can be tuned based on heuristics or prior information.*

4.3 Insights for Regulators

A financial regulator can observe the network but does not know the firms' beliefs. The regulator may ask: what changes in beliefs caused recently observed changes in the network? What are the side effects of different regulatory interventions? To answer these questions, we need to know how changes in firms' beliefs or utility functions affect the network. That is the subject of this section.

4.3.1 Effect of Friction in Contract Formation

Our model imposes no costs for contract formation. This is reasonable for large firms where the fixed costs associated with contract negotiations may be small relative to the contract sizes. However, in an overheating market, a regulator may impose frictions by penalizing large contracts, for example by increasing margin requirements.

We model contract costs via an adding a penalty term $F_i(\boldsymbol{w}_i)$ to the utility of agent i in Eq. (5.1):

$$\text{agent } i\text{'s utility } g_i(W,P) := \boldsymbol{w}_i^T(\boldsymbol{\mu}_i - P\boldsymbol{e}_i) - \gamma_i \cdot \boldsymbol{w}_i^T \Sigma_i \boldsymbol{w}_i - F_i(\boldsymbol{w}_i). \tag{4.2}$$

Theorem 4.3.1. *Consider a network setting where $\Sigma_i = \Sigma$ and all edges are allowed. Suppose that for each firm $i \in [n]$, the function $F_i : \mathbb{R}^n \to \mathbb{R}$ is twice differentiable, and there exist strictly increasing functions $f_{ji} : \mathbb{R} \to \mathbb{R}$ such that for all $\boldsymbol{x} \in \mathbb{R}^n$, $\nabla F_i(\boldsymbol{x}) = [f_{1i}(x_1), \ldots, f_{ni}(x_n)]^T$. Then, there exists a unique stable point.*

Example 4.3.2. *By imposing frictions, the regulator may* increase *the sizes of certain contracts. For example, let $F_i(\boldsymbol{w}_i) = \epsilon \cdot w_{i;i}^2 + \lambda \cdot \sum_{j \neq i} w_{i;j}^2$ for some $\lambda > \epsilon > 0$. Thus, the cost of inter-firm trades scales with the square of the contract size (we assume $\epsilon \approx 0$). Consider a network setting with 3 firms, with $\gamma_i = 1$, $\Sigma_i = \Sigma = \begin{bmatrix} 0.1 & 0.1 & 0.1 \\ 0.1 & 1 & 0.5 \\ 0.1 & 0.5 & 1 \end{bmatrix}$, and $M = \begin{bmatrix} 0 & 1000 & 111.233 \\ 1000 & 1 & 0.1 \\ 1000 & 0.1 & 1 \end{bmatrix}$. Then, $W_{23} = W_{32} \approx 0$ without frictions (when $F_i(\boldsymbol{w}_i) = 0$) but $|W_{23}| > 0$ for $\lambda > 0$.*

4.3.2 Effect of Changes in Firms' Beliefs

Regulatory actions can change the risk and expected return perceptions of firms. The next theorem shows the effect of such belief changes on the stable point.

Theorem 4.3.3. *Suppose $\Sigma_i = \Sigma$ for all firms, and let M be the matrix of expected returns.*

1. **Change in beliefs about expected returns:** *Let Σ have the eigendecomposition $\Sigma = V\Lambda V^T$. Then for $i, j, k, \ell \in [n]$,*

$$\frac{\partial W_{ij}}{\partial M_{k\ell}} = \frac{1}{2\sqrt{\gamma_i\gamma_j\gamma_k\gamma_\ell}} \cdot \sum_{s,t\in[n]} \frac{V_{is}V_{ks}V_{jt}V_{\ell t} + V_{is}V_{\ell s}V_{jt}V_{kt}}{\lambda_s + \lambda_t}. \tag{4.3}$$

   In particular, W_{ij} is monotonically increasing with respect to M_{ij}.

2. **Risk scaling:** *If the covariance Σ changes to $c\Sigma$ ($c > 0$), then W changes to $(1/c)W$.*

3. **Increase in perceived risk:** *Suppose $\gamma_i = \gamma$ for all i, and the covariance Σ increases to $\Sigma' \succ \Sigma$. Let W and W' be the stable points under Σ and Σ' respectively. Then, $\mathrm{tr}(M^T(W' - W)) < 0$.*

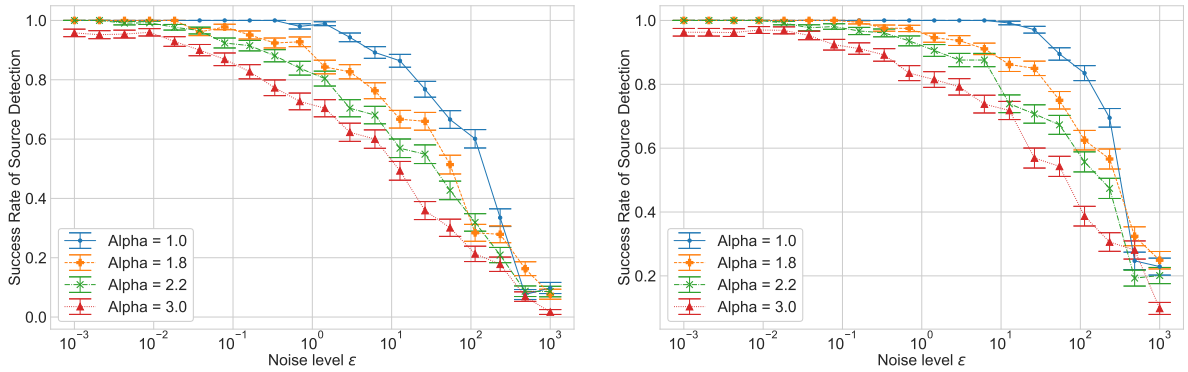
This shows that, in general, an increase in risk leads to a decrease in the weighted average of the contract sizes. The weights are given by the expected return beliefs of the firms. However, individual contracts between firms can increase, as can the norm $\|W\|_F$. This is because increases in the covariance Σ may also increase correlations, which can offer better hedging opportunities. By hedging some risks, larger contract sizes can be supported.

Theorem 4.3.3 also shows that a change in the perceived expected return $M_{k\ell}$ affects all contracts W_{ij}. Can we trace the changes in W back to the underlying changes in M? For instance, consider the following problem.

Definition 4.3.4 (Source Detection Problem)**.** *Suppose that a financial regulator observes two networks W and W', with the only difference being a small change in a single entry of M (say, M_{ij}). Can the regulator identify the pair (i, j)?*

One approach is to try to infer all beliefs of all firms, and then identify the changed belief. But, as discussed in Section 4.2.4, the beliefs are only identifiable under extra assumptions and more data. An alternative approach for the source detection problem is to find the entry (i, j) with the largest change $|W_{ij} - W'_{ij}|$. The intuition is that a change in M_{ij} has a direct effect on W_{ij} and (hopefully weaker) indirect effects on other contracts. Thus, the source detection problem is closely tied to the following:

Definition 4.3.5 (Targeted Intervention Problem)**.** *Can a regulator induce a small change in a single entry of M (say, M_{ij}) such that the change in W_{ij} is significantly larger than changes in other entries of W?*



(a) Predict most shifted contract as source

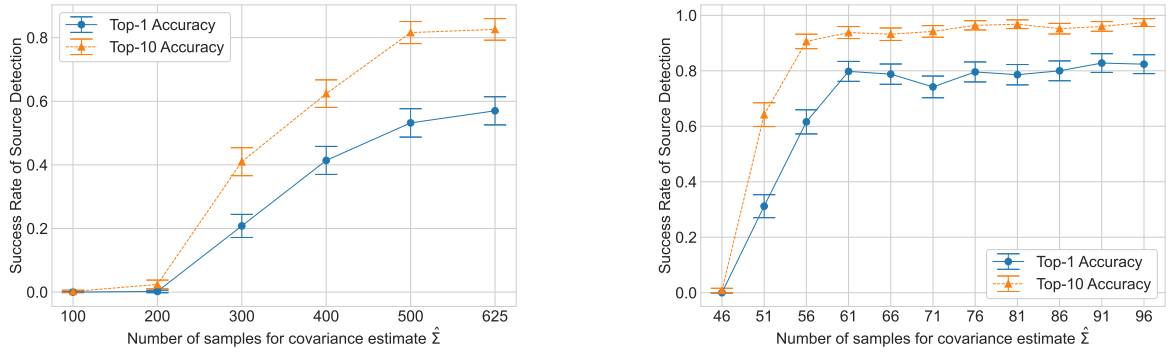
(b) Predict top-10 most shifted contracts

Figure 4.3: *Source Detection Problem in a noisy scaled equi-correlation model of* Σ*:* We rank the entries of W by the magnitude of change induced by a change in one entry of M (M_{ij}). Plot (a) shows the fraction of times W_{ij} is most-changed entry of W. Plot (b) shows the fraction of times W_{ij} is among the top-10 most changed entries of W. The success rate goes to zero as α and ϵ increase.

When all eigenvalues of Σ are equal (that is, $\Sigma \propto I_n$), a change in $M_{k\ell}$ only affects $W_{k\ell} (= W_{\ell k})$, as can be seen from Corollary 4.6.1. But when the eigenvalues are skewed, the terms in Eq. (4.3) corresponding to the smallest eigenvalues have greater weight. In such circumstances, the indirect effect of a change in $M_{k\ell}$ on other W_{ij} can be significant. The following empirical results show that this is indeed the case.

Empirical Results for the Source Detection Problem (Simulated Data). Here, we set the covariance $\Sigma = D^{1/2}(R + \mathcal{E})D^{1/2}$, where D is a diagonal matrix, R a correlation matrix, and $\mathcal{E}$ a noise matrix. If $\mathcal{E} = 0$, then D_{ii} would be the variance of firm i. We set D_{ii} according to a power law: $D_{ii} = i^{-\alpha}$ for an $\alpha > 0$. Larger values of α correspond to greater skew in the variances. We choose R to be an equi-correlation matrix with 1 along the diagonal and $\rho \in (0, 1)$ everywhere else. We draw the error matrix $\mathcal{E}$ from a scaled Wishart distribution: $\mathcal{E} = \|R\|_2 \cdot \mathcal{W}(\sqrt{\epsilon} \cdot I_n, n)/n$ for some chosen the noise level ϵ. As ϵ increases, the noise $\mathcal{E}$ dominates R.

Figure 4.3 shows the success rate of source detection over 1000 experiments for various values of (ϵ, α) for $\rho = 0.1$ and $n = 50$. As α increases, the variances become more skewed and the source detection can fail even with $\epsilon = 0$ noise. When ϵ grows, the success rate for the source detection problem goes to zero. This suggests that skew combined with noise makes source detection difficult. These trends occur even if we only test whether the source



(a) Simulated network of 96 portfolio managers.

(b) 46-country (OECD) trade network.

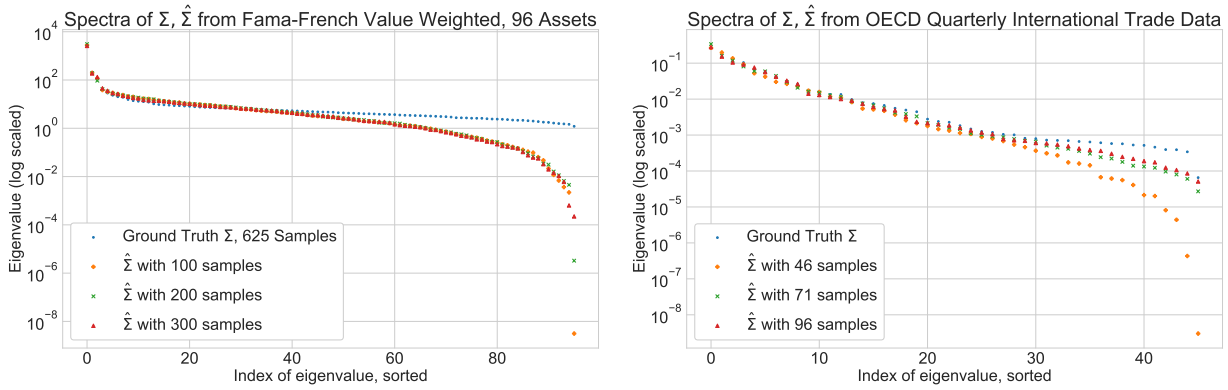
Figure 4.4: *Source Detection Problem on real-world data:* The success rate scales monotonically with the number of samples used to construct the data-driven covariance matrix $\hat{\Sigma}$.

belongs to the 10 most changed contracts (Figure 4.3b), as opposed to single largest change (Figure 4.3a). We observe similar results for real-world choices of Σ, as we show next.

Empirical Results for the Source Detection Problem (Real-World Data). We consider two datasets: (a) a trade network between 46 large economies (OECD, 2022), and (b) a simulated network between 96 portfolio managers following various Fama-French strategies Fama and French (2015). For each dataset, we construct a "ground-truth" covariance Σ using all available data (the details are in Section 4.7 of the supplementary materials). Then, using m independent samples $\boldsymbol{x}_i \sim \mathcal{N}(0, \Sigma)$, we build a "data-driven" covariance $\hat{\Sigma} = (1/(m-1)) \sum_{i=1}^{m} (\boldsymbol{x}_i - \hat{\boldsymbol{\mu}})(\boldsymbol{x}_i - \hat{\boldsymbol{\mu}})^T$, where $\hat{\boldsymbol{\mu}} = (1/m) \sum_{i=1}^{m} \boldsymbol{x}_i$ is the sample mean. We use this $\hat{\Sigma}$ to construct the financial network.

Figure 4.4 shows the success rate over 500 experiments for various choices of the sample size m. The success rate increases monotonically with m. The reason for this behavior lies in the spectra of Σ and $\hat{\Sigma}$. We find that in both datasets, the largest and smallest eigenvalues of Σ are separated by several orders of magnitude. This gap becomes even more extreme in the data-driven $\hat{\Sigma}$; the fewer the samples m, the greater the gap (see Figure 4.5). In fact, we observe that the smallest eigenvalue of $\hat{\Sigma}$ is much smaller than the second-smallest eigenvalue: $\lambda_n \ll \lambda_{n-1}$. Zhao et al. (2019) make similar observations.

In summary, the experiments on both simulated and real-world datasets highlight the difficulty of source detection and targeted intervention in realistic networks. The reason is the skew in the eigenvalues coupled with noise, which affects the eigenvectors. Skewed eigenvalues



(a) Simulated network of 96 portfolio managers

(b) 46-country (OECD) trade network

Figure 4.5: The eigenvalues of estimated covariance matrices are skewed, and the degree of skew depends on the number of samples m. As m decreases, so does the smallest eigenvalue λ_n and the ratio λ_n/λ_{n-1}.

correspond to trade combinations (eigenvectors) that are seemingly low-risk. Hence, firms use such trades to diversify. This implies that these eigenvectors have an outsized effect on the network, and how it responds to local changes. Intuitively, if these eigenvectors are "random," the effect of a changed belief $M_{k\ell}$ affects the rest of the network randomly. Hence, the direct effects on $W_{k\ell}$ may be less than the indirect effects on other W_{ij}. We explore this theoretically in Section 4.6.14 of the supplementary material.

4.4 Insights for Firms

Until now, we have treated the beliefs of firms as fixed and exogenous. In this section, we consider how a firm can use its contracts to gain insights into other firms and update its beliefs.

For instance, suppose a firm j faces a crisis, e.g., a looming debt payment that may make it insolvent. The firm may then become risk-seeking (i.e., lower its γ_j), hoping that the risks pay off. Another firm i may be unaware of the crisis, so i's risk perceptions (perhaps based on historical data) would be outdated. Can firm i *infer* the lower γ_j, solely from i's contracts $\boldsymbol{w}_i$ with all firms? What if a group of firms become risk-seeking, and not just one firm?

4.4.1 Detecting Outlier Firms

Intuitively, firm i will try to answer these questions by comparing the behavior of firm j against other similar firms. We formalize this by assuming that each firm j belongs to a community θ_j, e.g., banking, or real-estate, or insurance, etc. The community of each firm is publicly known. Firms in the same community are perceived to have similar return distributions:

$$M_{ij} = f(\theta_i, \theta_j) + \epsilon'_{\theta_i,j}, \quad \Sigma_{ij} = g(\theta_i, \theta_j), \quad \gamma_i = h(\theta_i) + \epsilon_i \tag{4.4}$$

for some unknown deterministic functions $f(.)$, $g(.)$, and $h(.)$ and random error terms ϵ_i and $\epsilon'_{\theta_i,j}$. We also assume that all firms use the same covariance Σ.

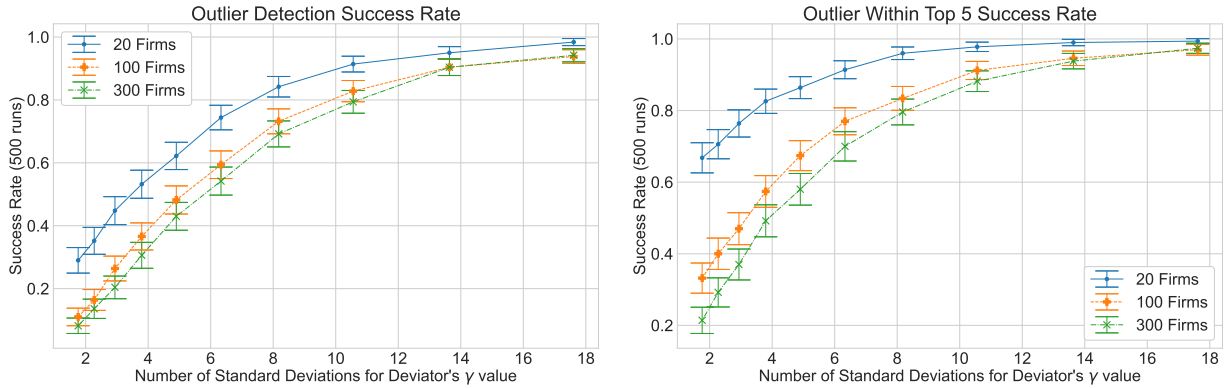
Now, suppose one firm j is an outlier, with very different beliefs from other firms in its community. For firm i to detect the outlier firm j, the contract size W_{ij} should deviate from a cluster of contracts $\{W_{ij'} \mid \theta_{j'} = \theta_j\}$ of other firms from the same community as firm j. Now, outlier detection methods often assume independent datapoints. In our model, all contracts are dependent. But we can still do outlier detection if the contracts are appropriately exchangeable. We prove below this is the case.

Definition 4.4.1. *An intra-community permutation is a permutation $\pi : [n] \to [n]$ such that $\pi(i) = j$ implies that $\theta_i = \theta_j$.*

Proposition 4.4.2. *Suppose M, Σ, Γ exhibit community structure (Eq. (4.4)), and all the error terms $(\epsilon_i)_{i\in[n]}$ and $(\epsilon'_{\theta_i,j})_{i,j\in[n]}$ are independent and identically distributed. Let $\pi : [n] \to [n]$ be any intra-community permutation, and let $\Pi : \mathbb{R}^n \to \mathbb{R}^n$ be the corresponding column-permutation matrix: $\Pi(\boldsymbol{e}_i) = \boldsymbol{e}_{\pi(i)}$. Then, W and $\Pi^T W \Pi$ are identically distributed.*

Corollary 4.4.3. *Let $j_1, \ldots, j_m \in [n]$ belong to the same community: $\theta_{j_1} = \cdots = \theta_{j_m}$. Suppose the conditions of Proposition 4.4.2 hold. Then, for any $i \in [n]$, the joint distribution of $(W_{i,j_1}, \ldots, W_{i,j_m})$ is exchangeable.*

Empirical Results for Outlier Detection. We generate community-based networks (Eq. (4.4)) such that $\gamma_i \sim N(1, \sigma^2)$ truncated to $[0.5, 1.5]$. The smaller the σ, the more closely the γ_i values cluster around 1. For the outlier risk-seeking firm, we set $\gamma_{\text{outlier}} = 0.5$. For clarity of exposition, we set $\epsilon' = 0$ everywhere.



(a) Predict largest contract as outlier

(b) Predict the top-5 largest contracts

Figure 4.6: *Success rate for detecting outlier risk-seeking firms:* Detection is easier when there are fewer firms and when the risk-seeking firm's γ_{outlier} is more standard deviations away from the γ of the normal firms.

To detect outliers under exchangeability (Corollary 4.4.3), we can use methods based on conformal prediction (Guan and Tibshirani, 2022). Here, we use a simpler approach: pick the firm j with the largest contract size as the outlier; $\hat{j} := \arg\max_{j \in \{j_1,\ldots,j_m\}} |W_{i,j}|$. To test sensitivity to false negatives, we also test whether the outlier is among the 5 largest contracts in $\{|W_{i,j}| : j = j_1, \ldots, j_m\}$. We run 500 experiments for each choice of σ, and count the frequency with which the outlier firm is detected via its contract size. Further details are presented in Section 4.7.3 of the supplementary material.

Figure 4.6 shows the results. We characterize the degree of outlierness by how many standard deviations away γ_{outlier} is from the baseline of 1. The smaller the σ, the more the outlierness. The success rate increases with increasing outlierness, as expected. It also increases when the number of firms n is reduced. This is because contract sizes depend on the γ values of all firms; fewer firms reduces the chances of any one firm attaining large contract sizes due to randomness.

4.4.2 Risk-Aversion versus Expected Returns

The discussion above shows that a firm can detect outlier counterparties. However, the firm cannot determine *why* the counterparty is an outlier, as the following theorem shows.

Theorem 4.4.4 (Non-identifiability of risk-aversion versus expected returns)**.** *Consider two network settings* $S = (\boldsymbol{\mu}_i, \Sigma, \gamma_i)_{i\in[n]}$ *and* $S' = (\boldsymbol{\mu}_i, \Sigma, \gamma'_i)_{i\in[n]}$ *which differ only in the risk-*

aversions of firms $J = \{j \mid \gamma_j \neq \gamma'_j\} \subseteq [n]$. *Then, there exists a setting* $S^\dagger = (\boldsymbol{\mu}_i^\dagger, \Sigma, \gamma_i)_{i\in[n]}$ *such that* $\boldsymbol{\mu}_i = \boldsymbol{\mu}_i^\dagger$ *for all* $i \notin J$ *and the stable networks under* $S^\dagger$ *and* S' *are identical.*

Thus, one cannot determine if an outlier is more risk-seeking than its community or expects higher profits. But risk-seeking behavior may be indicative of stress, while higher profits than similar firms are unlikely. Hence, in either case, the firm detecting the outlier may choose to reduce its exposure to the outlier. However, this approach fails if an entire community shifts its behavior. The following example illustrates the problem.

Example 4.4.5. *Consider two communities numbered* 1 *and* 2*, with* n_1 *and* n_2 *firms respectively. Let the setting* S *of Theorem 4.4.4 correspond to*

$$M_{ij} = \begin{cases} a & \text{if } \theta_i = \theta_j = 1 \\ b & \text{if } \theta_i = \theta_j = 2 \\ c/2 & \text{otherwise} \end{cases}$$

$$\Sigma_{ij} = \begin{cases} 1 & \text{if } \theta_i = \theta_j = 1 \\ 1 & \text{if } \theta_i = \theta_j = 2 \\ 0 & \text{otherwise} \end{cases} \qquad \gamma_i = 1.$$

Now, suppose that under setting S', $\gamma_i \mapsto \gamma_i + \delta$ *for some small* δ *for all nodes* i *in community* 1*. The change in the network would be the same if we had updated the columns corresponding to community* 1 *in the* M *matrix instead (setting* $S^\dagger$*):*

$$M_{ij}^\dagger = M_{ij} + \Delta(\theta_i, \theta_j)$$

$$\Delta(\theta_i, \theta_j) + O(\delta^2) = \begin{cases} -\delta a/2 & \text{if } \theta_i = \theta_j = 1 \\ -\delta b \cdot n_2/(n_1 + n_2) & \text{if } \theta_i = 2, \theta_j = 1 \\ 0 & \text{if } \theta_j = 2 \end{cases}$$

Thus, a firm from community 2 *cannot determine if the network change was due to a change in* $(\gamma_i)_{\theta_i=1}$ *or* $(\boldsymbol{\mu}_i)_{\theta_i=1}$*. For instance, when* $b > 0$*, an increase in risk-seeking* ($\delta < 0$) *looks the same as an increase in trading benefits* ($\Delta(1,2) > 0$)*. In the former case, firms in community* 2 *should* reduce *their exposure to community* 1 *firms. But in the latter case, they should* increase *exposure. Since the data cannot be used to choose the appropriate action, the behaviors of firms may be guided by their prior beliefs or inertia. When such beliefs change due to external events (e.g., due to news about one firm in community* 1*), the resulting change in the network may be drastic.* □

4.5 Conclusions

We have proposed a model of a weighted undirected financial network of contracts. The network emerges from the beliefs of the participant firms. The link between the two is utility maximization coupled with pricing. For almost all belief settings, our approach yields a unique network. This network satisfies a strong Higher-Order Nash Stability property. Furthermore, the firms can converge to this stable network via iterative pairwise negotiations.

The model yields two insights. First, a regulator is unable to reliably identify the causes of a change in network structure, or engage in targeted interventions. The reason is that firms seek to diversify risk by exploiting correlations. We find that in realistic settings, there are often combinations of trades that offer seemingly low risk. Hence, all firms aim to use such trades. The over-dependence on a few such combinations leads to a pattern of connections between firms that thwarts targeted regulatory interventions.

The second insight is that firms can use the network to update their beliefs. For instance, they can identify counterparties that behave very differently from their peers. However, the cause of the outlierness remains hidden. If all firms in one line of business become more risk-seeking, the result is indistinguishable from that business becoming more profitable. Innocuous events (such as a news story) may cause beliefs to change suddenly, leading to drastic changes in the network. In addition to identifying risky counterparties, firms may use the network to update their mean and covariance beliefs. For example, a firm that suffers significant losses on its current trades may be judged by others to be a riskier counterparty for future trades. We leave this for future work.

Our work focuses on mean-variance utility, but some of our results are applicable in other settings too. A second-order Taylor approximation of a twice-differentiable concave utility matches the form of a mean-variance utility. Hence, results based on mean-variance utility can be useful guides for small perturbations around a stable point. Some of our results for pairwise negotiations and targeted interventions are based on such perturbation arguments.

Finally, contract formation under budget constraints is an important direction for future work. In Theorem 4.3.1, we only consider contract frictions that depend on a firm's contract sizes. To model budget constraints, we must also consider the contract prices. These

require different techniques than our approach, which is based on results from Sandberg and Willson (1972) (see Section 4.6.17 in the supplementary material).

4.6 Proofs and Additional Results

4.6.1 Proof of Theorem 4.2.8

Recall that $Q_i = \Psi_i^T(2\gamma_i\Psi_i\Sigma_i\Psi_i^T)^{-1}\Psi_i$, $F = \{(i,j) : 1 \leq i < j \leq n, \Psi_i \boldsymbol{e}_j \neq \boldsymbol{0}\}$, and $\text{uvec}(X)_F \in \mathbb{R}^{|F|}$ is a vector whose entries are the ordered set $\{X_{ij} \mid (i,j) \in F\}$. Note that $\Psi_i\Sigma_i\Psi_i^T$ is positive definite, since it is a principal submatrix of the positive definite matrix Σ_i.

Proof. Proof of Theorem 4.2.8. For clarity of exposition, we first prove the result when all edges are allowed, and then consider the case of disallowed edges.

(1) All edges allowed. Here, $E = \{i, j \mid 1 \leq i < j \leq n\}$, and we use uvec(.) and Z to refer to $\text{uvec}(.)_E$ and Z_E in the theorem statement. For any price matrix P with $P = -P^T$, consider the matrix W whose j^{th} column has the utility-maximizing contract sizes for agent j:

$$\begin{aligned} W_{ij} &= \boldsymbol{e}_i^T\Psi_j^T(2\gamma_j\Psi_j\Sigma_j\Psi_j^T)^{-1}\Psi_j(M-P)\boldsymbol{e}_j \\ &= \boldsymbol{e}_i^T Q_j(M-P)\boldsymbol{e}_j. \end{aligned}$$

The tuple (W, P) is stable if $W = W^T$. So, for all $i < j$, we require

$$\begin{aligned} W_{ij} &= W_{ji} \qquad (4.5) \\ \Leftrightarrow \boldsymbol{e}_i^T Q_j(M-P)\boldsymbol{e}_j &= \boldsymbol{e}_j^T Q_i(M-P)\boldsymbol{e}_i \\ \Leftrightarrow \boldsymbol{e}_i^T Q_j M\boldsymbol{e}_j - \boldsymbol{e}_j^T Q_i M\boldsymbol{e}_i &= \boldsymbol{e}_i^T Q_j P\boldsymbol{e}_j - \boldsymbol{e}_j^T Q_i P\boldsymbol{e}_i \\ \Leftrightarrow \boldsymbol{e}_i^T(A - A^T)\boldsymbol{e}_j &= \boldsymbol{e}_i^T(Q_jP - (Q_iP)^T)\boldsymbol{e}_j. \qquad (4.6) \end{aligned}$$

Since $P = -P^T$, we must have $P = R - R^T$, where R is upper-triangular with zero on the diagonal. Hence, using $Q_i = Q_i^T$, we have

$$\begin{aligned} \boldsymbol{e}_i^T(Q_jP - (Q_iP)^T)\boldsymbol{e}_j &= \boldsymbol{e}_i^T(Q_jP + PQ_i)\boldsymbol{e}_j \\ &= trP(\boldsymbol{e}_j\boldsymbol{e}_i^TQ_j + Q_i\boldsymbol{e}_j\boldsymbol{e}_i^T) \\ &= tr(R - R^T)(B_{(j,i)} + B_{(i,j)}^T) \\ &= trR^TC_{(i,j)} \\ &= \text{uvec}(R)^T\text{uvec}(C_{(i,j)}), \end{aligned}$$

where we used the upper-triangular nature of R in the last step. Plugging into Eq. (4.6), a stable point exists if and only if there is an appropriate vector $\boldsymbol{p} := \text{uvec}(R) \in \mathbb{R}^{n(n-1)/2}$ such that for all $1 \leq i < j \leq n$, $\boldsymbol{e}_i^T(A - A^T)\boldsymbol{e}_j = \text{uvec}(C_{(i,j)})^T\boldsymbol{p}$. This is equivalent to $\text{uvec}(A - A^T) = Z\boldsymbol{p}$. If such a solution vector $\boldsymbol{p}$ exists, then by definition it corresponds to a matrix $P = -P^T$ via $P = R - R^T$ and $\boldsymbol{p} = \text{uvec}(R)$.

(2) Disallowed edges. If $\{i, j\}$ is a prohibited edge then $\Psi_i\boldsymbol{e}_j = \Psi_j\boldsymbol{e}_i = \boldsymbol{0}$, so $B_{(i,j)} = B_{(j,i)} = 0$, so $\boldsymbol{e}_{ij}^T Z = \boldsymbol{0}^T$. Also, $A_{ij} = A_{ji} = 0$ so $\text{uvec}(A - A^T)_{ij} = 0$. Therefore, the equality $\boldsymbol{e}_i^T(A - A^T)\boldsymbol{e}_j = \text{uvec}(C_{(i,j)})^T\boldsymbol{x}$ is achieved for any solution vector $\boldsymbol{x}$ if $\{i, j\}$ is a prohibited edge. We can therefore reduce the linear system $Z\boldsymbol{p} = \text{uvec}(A - A^T)$ from part (1) by deleting rows of Z corresponding to prohibited edges.

Similarly, since the system is constrained by $\boldsymbol{p}_{ij} = 0$ for prohibited edges $\{i, j\}$, the columns of Z corresponding to such edges have no effect on the solution set.

We conclude that the linear system in (1) is equivalent to the (unconstrained) reduced system $Z_F\boldsymbol{p}_F = \text{uvec}(A - A^T)_F$. Each solution $\boldsymbol{p}_F$ corresponds to a skew-symmetric P by construction. Finally, if Z_F has full rank then the unique reduced solution is $\boldsymbol{p}_F = Z_F^{-1}\text{uvec}(A - A^T)_F$. □ □

4.6.2 Stable Network for the Shared Covariance Case

In the case of a shared covariance matrix for all agents, we can give a closed form expression for the stable network.

Corollary 4.6.1 (Shared Σ, all edges allowed)**.** *Suppose $\Sigma_i = \Sigma$ and $\Psi_i = I_n$ for all $i \in [n]$. Let $(\lambda_i, \boldsymbol{v}_i)$ denote the i^{th} eigenvalue and eigenvector of $\Gamma^{-1/2}\Sigma\Gamma^{-1/2}$. Then, the network W can be written in two equivalent ways:*

$$\text{vec}(W) = \frac{1}{2}(\Gamma \otimes \Sigma + \Sigma \otimes \Gamma)^{-1}\text{vec}(M + M^T),$$
$$W = \Gamma^{-1/2}\left(\sum_{i=1}^{n}\sum_{j=1}^{n}\frac{\boldsymbol{v}_i^T\Gamma^{-1/2}}{2(\lambda_i + \lambda_j)}(M + M^T)\Gamma^{-1/2}\boldsymbol{v}_j\boldsymbol{v}_i\boldsymbol{v}_j^T\right)\Gamma^{-1/2}.$$

The prices can be written as:

$$\text{vec}(P) = (\Gamma^{-1} \otimes \Sigma^{-1} + \Sigma^{-1} \otimes \Gamma^{-1})^{-1}\text{vec}(\Sigma^{-1}M\Gamma^{-1} - \Gamma^{-1}M^T\Sigma^{-1})$$
$$P = \Gamma^{1/2}\left(\sum_{i=1}^{n}\sum_{j=1}^{n}\frac{\boldsymbol{v}_i^T\Gamma^{1/2}}{\lambda_i^{-1} + \lambda_j^{-1}}(\Sigma^{-1}M\Gamma^{-1} - \Gamma^{-1}M^T\Sigma^{-1})\Gamma^{1/2}\boldsymbol{v}_j\boldsymbol{v}_i\boldsymbol{v}_j^T\right)\Gamma^{1/2}.$$

Proof. Proof. We first prove the identity with $\text{vec}(W)$.

For each agent i the optimal set of contracts is given as $\boldsymbol{w}_i = (2\gamma_i\Sigma_i)^{-1}(M-P)\boldsymbol{e}_i$. Since $\Sigma_i = \Sigma$ for all i, we obtain $W = \frac{1}{2}\Sigma^{-1}(M-P)\Gamma^{-1}$. Hence $M - P = 2\Sigma W\Gamma$. Using $W = W^T$ and $P^T = -P$ for a stable feasible point (W, P), we obtain $\Sigma W\Gamma + \Gamma W\Sigma = \frac{1}{2}(M + M^T)$.

Vectorization implies $(\Gamma \otimes \Sigma + \Sigma \otimes \Gamma)\text{vec}(W) = \frac{1}{2}\text{vec}(M + M^T)$. It remains to show that $(\Gamma \otimes \Sigma + \Sigma \otimes \Gamma)$ is invertible.

Let $K := (\Gamma \otimes \Sigma + \Sigma \otimes \Gamma)$ for shorthand. Notice $K = (\Gamma^{1/2} \otimes \Gamma^{1/2})(I \otimes \Gamma^{-1/2}\Sigma\Gamma^{-1/2} + \Gamma^{-1/2}\Sigma\Gamma^{-1/2} \otimes I)(\Gamma^{1/2} \otimes \Gamma^{1/2})$. Let $K' = (I \otimes \Gamma^{-1/2}\Sigma\Gamma^{-1/2} + \Gamma^{-1/2}\Sigma\Gamma^{-1/2} \otimes I)$. Since $(\Gamma^{1/2} \otimes \Gamma^{1/2})$ is invertible it suffices to show K' is invertible.

Properties of Kronecker products imply that if a matrix $A \in \mathbb{R}^{n\times n}$ has strictly positive eigenvalues then $\sigma(I \otimes A + A \otimes I) = \{\lambda + \mu : \lambda, \mu \in \sigma(A)\}$ counting mutiplicities (Horn and Johnson, 1994). Let $\boldsymbol{v} \neq \boldsymbol{0}$. Then, since $\Sigma \succ 0$ and $\Gamma^{-1/2} \succ 0$ we obtain $\boldsymbol{v}^T\Gamma^{-1/2}\Sigma\Gamma^{-1/2}\boldsymbol{v} = (\Gamma^{-1/2}\boldsymbol{v})^T\Sigma(\Gamma^{-1/2}\boldsymbol{v}) > 0$. Hence $\Gamma^{-1/2}\Sigma\Gamma^{-1/2} \succ 0$, so K' is invertible and hence K is invertible. This proves the first identity.

Next, we prove the second identity. Properties of Kronecker products imply that $(K')^{-1}$ has eigendecomposition $(K')^{-1} = \sum_{i=1}^n \sum_{j=1}^n \frac{1}{\lambda_i+\lambda_j}(\boldsymbol{v}_i \otimes \boldsymbol{v}_j)(\boldsymbol{v}_i \otimes \boldsymbol{v}_j)^T$.

Therefore, since $(\Gamma^{1/2} \otimes \Gamma^{1/2})^{-1} = (\Gamma^{-1/2} \otimes \Gamma^{-1/2})$ we obtain:

$$
\begin{aligned}
\text{vec}(W) &= (\Gamma^{-1/2} \otimes \Gamma^{-1/2}) \sum_{i=1}^n \sum_{j=1}^n \frac{1}{\lambda_i + \lambda_j}(\boldsymbol{v}_i \otimes \boldsymbol{v}_j)(\boldsymbol{v}_i \otimes \boldsymbol{v}_j)^T(\Gamma^{-1/2} \otimes \Gamma^{-1/2})\text{vec}\Big(\frac{M + M^T}{2}\Big) \\
&= (\Gamma^{-1/2} \otimes \Gamma^{-1/2}) \sum_{i=1}^n \sum_{j=1}^n \frac{1}{2(\lambda_i + \lambda_j)}\text{vec}\big(\Gamma^{-1/2}(M + M^T)\Gamma^{-1/2}\big) \\
&= (\Gamma^{-1/2} \otimes \Gamma^{-1/2})\text{vec}\bigg(\sum_{i=1}^n \sum_{j=1}^n \frac{\boldsymbol{v}_i^T\Gamma^{-1/2}}{2(\lambda_i + \lambda_j)}(M + M^T)\Gamma^{-1/2}\boldsymbol{v}_j\boldsymbol{v}_i\boldsymbol{v}_j^T\bigg) \\
W &= \Gamma^{-1/2}\bigg(\sum_{i=1}^n \sum_{j=1}^n \frac{\boldsymbol{v}_i^T\Gamma^{-1/2}}{2(\lambda_i + \lambda_j)}(M + M^T)\Gamma^{-1/2}\boldsymbol{v}_j\boldsymbol{v}_i\boldsymbol{v}_j^T\bigg)\Gamma^{-1/2}
\end{aligned}
$$

Finally, the formulas for $\text{vec}(P)$ and P follow from similar reasoning, using $W = W^T$ and $W = \frac{1}{2}\Sigma^{-1}(M-P)\Gamma^{-1}$. □ □

4.6.3 Example of Stable Network

To illustrate Theorem 4.2.8, consider the following example.

Example 4.6.2 (Stable points)**.** *Consider a 3-firm network where the only allowed edges are given by* $F = \{(1,2),(1,3)\}$*. Suppose firms share the same covariance belief matrix* $\Sigma_1 = \Sigma_2 = \Sigma_3 = \Sigma$*, but have different mean beliefs* $M = \begin{bmatrix}\boldsymbol{\mu}_1 & \boldsymbol{\mu}_2 & \boldsymbol{\mu}_3\end{bmatrix}$ *and risk aversions. The firms' beliefs are:*

$$M = \begin{bmatrix} 0 & 2/3 & 1/2 \\ 1 & 0 & 0 \\ 1 & 0 & 0 \end{bmatrix},$$
$$\Sigma = \begin{bmatrix} 1 & 1/2 & 1/2 \\ 1/2 & 1 & 1/2 \\ 1/2 & 1/2 & 1 \end{bmatrix}, \gamma_1 = 1, \gamma_2 = 1/2, \gamma_3 = 1/4$$

Then the $A, B_{(i,j)}$ *matrices in Theorem 4.2.8 are given as:*

$$A = \begin{bmatrix} 0 & 2/3 & 3/4 \\ 1/2 & 0 & 0 \\ 1/2 & 0 & 0 \end{bmatrix}, B_{(1,2)} = \begin{bmatrix} 0 & 3/4 & -1/4 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix},$$
$$B_{(1,3)} = \begin{bmatrix} 0 & -1/4 & 3/4 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}, B_{(2,1)} = \begin{bmatrix} 0 & 0 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix},$$
$$B_{(3,1)} = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 3/2 & 0 & 0 \end{bmatrix}$$

Hence

$$C_{(1,2)} = \frac{1}{4}\begin{bmatrix} 0 & 7 & -1 \\ -7 & 0 & 0 \\ 1 & 0 & 0 \end{bmatrix}, C_{(1,3)} = \frac{1}{4}\begin{bmatrix} 0 & -1 & 9 \\ 1 & 0 & 0 \\ -9 & 0 & 0 \end{bmatrix},$$

Therefore, $Z_F = \frac{1}{4}\begin{bmatrix} 7 & -1 \\ -1 & 9 \end{bmatrix}$ *and* $\operatorname{uvec}(A - A^T)_F = (1/6, 1/4)^T$*. Since* Z_F *is full-rank, there exists a unique stable point for this network setting.*

4.6.4 Stable Points are Common

Lemma 4.6.3. *Define* F*,* Z_F *and* Q_i *as in Theorem 4.2.8. Let* Q_i' *be such that*

$$(Q_i')_{j,k} = \begin{cases} (Q_i)_{j,k} + \beta & \text{if } j = k, (i,j) \in F \\ (Q_i)_{j,k} & \text{otherwise} \end{cases}$$

Then, the corresponding Z_F' *has the form* $Z_F' = Z_F + \beta I$*.*

Proof. Proof of Lemma 4.6.3. This follows from the form of the matrices $B_{(i,j)}$ and $C_{(i,j)}$ in the statement of Theorem 4.2.8. □ □

Now, we consider the Σ_i's (and hence the Q_i's) to be random variables. Any distribution of $\{\Sigma_i\}_{i\in[n]}$ induces a distribution on $\{Q_i\}_{i\in[n]}$, where $Q_i \succ 0$. Define $\tilde{Q}_i := Q_i - \delta I$, where $\delta > 0$ is the minimum of union of the (nonzero) eigenvalues of all the Q_i's. A distribution over $\{Q_i\}$ corresponds to a distribution over $(\{\tilde{Q}_i\}, \delta)$.

Proposition 4.6.4. *If the distribution of δ given $\{\tilde{Q}_i\}_{i\in[n]}$ is continuous, then a unique stable point exists with probability* 1*.*

Proof. Proof of Proposition 4.6.4. Let $\tilde{Z}_F$ be the $|F| \times |F|$ matrix generated from $\{\tilde{Q}_i\}_{i\in[n]}$, and Z_F the corresponding matrix for $\{Q_i\}_{i\in[n]}$. By Lemma 4.6.3, $Z_F = \tilde{Z}_F + \delta I$. Hence, $\sigma(Z_F) = \sigma(\tilde{Z}_F) + \delta$, where $\sigma(M)$ denote the set of eigenvalues of M. Since $\sigma(\tilde{Z}_F)$ is a function of $\{\tilde{Q}_i\}$ and δ is continuous given $\{\tilde{Q}_i\}$, the eigenvalues of Z_F are non-zero with probability 1. Hence, by Theorem 4.2.8, a unique stable point exists for $\{Q_i\}$ with probability 1. □ □

Note that we require no condition on the distribution of $\{\tilde{Q}_i\}$. The condition of Proposition 4.6.4 is satisfied if the joint distribution of the $\{\Sigma_i\}_{i\in[n]}$ is continuous and all edges are permitted, as shown in the following example.

Example 4.6.5. *Fix some $n \geq 2$. Suppose the joint distribution of the $\{\Sigma_i\}_{i\in[n]}$ is continuous and all edges are permitted. Then $Q_i = (2\gamma_i)^{-1}\Sigma_i^{-1}$ so the joint distribution of $\{Q_i\}_{i\in[n]}$ is continuous. By Bayes' rule, $\mathbb{P}[\delta|\tilde{Q}_1, \ldots, \tilde{Q}_n] \propto \mathbb{P}[\delta, \tilde{Q}_1, \ldots, \tilde{Q}_1] = \mathbb{P}[Q_1, \ldots, Q_n]$. Since $\mathbb{P}[Q_1, \ldots, Q_n]$ is continuous, we conclude $\mathbb{P}[\delta|\tilde{Q}_1, \ldots, \tilde{Q}_n]$ is continuous.*

4.6.5 Proof of Theorem 4.2.9

Proof. Proof of Theorem 4.2.9. *Case 1: $P = P^*$.* First, consider a feasible (W, P) such that $P = P^*$. Then $W \neq W^*$. Since W^* is stable, by definition each agent optimizes contracts with respect to P^*, so no agent is worse off under (W^*, P^*) then (W, P^*). Hence $(W, P) \not\succ (W^*, P^*)$.

Case 2: $P \neq P^*$. Second, suppose that $P \neq P^*$. Let $\Delta_i := g_i(W, P) - g_i(W, P^*)$. It follows that $\Delta_i = (W\boldsymbol{e}_i)^T((P^* - P)\boldsymbol{e}_i)$. Let $A \in \mathbb{R}^{n\times n}$ be defined as $A_{ij} = W_{ij}(P^*_{ij} - P_{ij})$. Then $\Delta_i = \boldsymbol{e}_i^T A\mathbf{1}$.

Next, notice that $A_{ji} = -A_{ij}$. Therefore, $\sum_i \Delta_i = \mathbf{1}^T A \mathbf{1} = 0$. Hence, either $\Delta_i = 0$ for all i, or there exists k such that $\Delta_k < 0$.

Case 2(i). Suppose there exists k such that $\Delta_k < 0$. Then $g_k(W, P) < g_k(W, P^*)$. By case 1, we have $g_k(W, P^*) \leq g_k(W^*, P^*)$. Therefore agent k is strictly worse off, so $(W, P) \not\succ (W^*, P^*)$.

Case 2(ii). Suppose $\Delta_i = 0$ for all i. Then $g_i(W, P) = g_i(W, P^*)$ for all i. By case 1, we have $g_i(W, P^*) \leq g_i(W^*, P^*)$. Therefore no agent is better off, so $(W, P) \not\succ (W^*, P^*)$. □ □

4.6.6 Proof of Theorem 4.2.12

Proof. Proof of Theorem 4.2.12. First, we argue (W, P) is a Nash equilibrium. Suppose that agent i wants to shift some of their contracts at the stable feasible point (W, P). Suppose they propose $(w'_{i,j_1}, p'_{i,j_1}), \ldots, (w'_{i,j_m}, p'_{i,j_m})$ for $j_1, \ldots, j_m \in [n]$. Let (W', P') denote the new feasible point that occurs if all changes are accepted. By Theorem 4.2.9 we know that $(W', P') \not\succ (W, P)$, so at least one agent does not prefer (W', P'). Since the only changes are to edges $\{i, j_1\}, \ldots, \{i, j_m\}$, there must exist a $j \in \{j_1, \ldots, j_m\}$ who does not prefer (W', P'). Therefore, they will reject the proposal of agent i to shift to (w'_{ij}, p'_{ij}).

Then, agent i can choose to either maintain the existing contract (w_{ij}, p_{ij}) or delete the edge $\{i, j\}$. We claim that agent i prefers to keep the edge, since they could have chosen to set $W_{ij} = 0$ during the network formation process, no matter what price was offered. But $W_{ij} \neq 0$ at equilibrium (W, P). By stability of (W, P) we know W_{ij} is the optimal choice for agent i at prices P. Therefore, after agent j rejects (w'_{ij}, p'_{ij}), it follows that the edge remains at (w_{ij}, p_{ij}).

Since (W', P') was arbitrary, we conclude that at equilibrium, agent i cannot propose any set of changes that result in a strictly better network for them. Therefore, their optimal action at (W, P) is to not deviate from the equilibrium.

Next, we show cartel robustness. Suppose $S \subset [n]$ is a strict subset and $(W', P') \neq (W, P)$ is a feasible point differing only at indices $\{i, j\}$ such that $i, j \in S$. By Theorem 4.2.9, we know (W', P') cannot dominate (W, P), so there is some agent $i \in [n]$ that does not prefer (W', P') to (W, P). Since (W', P') only changes contracts where both members are in S, the utility of agents in $[n] \setminus S$ must be unchanged. Therefore $i \in S$, and hence not all members of the cartel have higher utility under (W', P'). □ □

4.6.7 Price Update Rule for Pairwise Negotiations

We give an explicit formula for the updated price of a unit contract after a pairwise negotiation.

Proposition 4.6.6 (Price after Pairwise Negotiation)**.** *Consider a network setting* $(\boldsymbol{\mu}_i, \gamma_i, \Sigma_i, \Psi_i)_{i\in[n]}$. *Let* Q_i *be as in Theorem 4.2.8. Given a price matrix* $P = -P^T$ *and a pair of firms* (i, j) *that are permitted to trade, let* P' *be another skew-symmetric price matrix such that (a)* P' *differs from* P *only in the cells* (i, j) *and* (j, i)*, (b)* i *and* j *both maximize their utility at the same contract size under* P'*, and (c)* i *and* j *can choose their optimal contract sizes with all other agents given these prices. Then,*

$$P'_{ij} = \frac{1}{Q_{i;j,j} + Q_{j;i,i}} \Big(\boldsymbol{e}_i^T Q_j (M - P)\boldsymbol{e}_j - \boldsymbol{e}_j^T Q_i (M - P)\boldsymbol{e}_i \Big) + P_{ij}$$

Proof. Proof. Let $A_i := \gamma_i Q_i$ for $i \in [n]$. Since $\Sigma_i \succ 0$ and $\Psi_i \Sigma_i \Psi_i^T$ is a principal submatrix, we know $\Psi_i \Sigma_i \Psi_i^T$ is real symmetric and positive definite, and hence its inverse is as well. Therefore A_i is real symmetric and PSD. (It is not full rank in general, unless $\Psi_i = I$). Since $\{i, j\}$ is a permitted edge, $\Psi_i \boldsymbol{e}_j \neq \boldsymbol{0}$ and $\Psi_j \boldsymbol{e}_i \neq \boldsymbol{0}$. Therefore $A_{i;j,j} = \boldsymbol{e}_j^T A_i \boldsymbol{e}_j = (\Psi_i \boldsymbol{e}_j)^T (2\Psi_i \Sigma_i \Psi_i^T)^{-1} (\Psi_i \boldsymbol{e}_j) > 0$ since $(2\Psi_i \Sigma_i \Psi_i^T)^{-1}$ is positive definite. So, $A_{i;j,j} > 0$ and similarly $A_{j;i,i} > 0$.

Now, the optimal contracts for agent i under prices P' are given by $\boldsymbol{w}_i = A_i(M - P')\Gamma^{-1}\boldsymbol{e}_i$. Note that $P' = P + (P'_{ij} - P_{ij})(\boldsymbol{e}_i \boldsymbol{e}_j^T - \boldsymbol{e}_j \boldsymbol{e}_i^T)$. Since both i and j maximize their utility at the same contract size, we have:

$$\begin{aligned}
\boldsymbol{w}_{i;j} &= \boldsymbol{w}_{j;i} \\
\Rightarrow \boldsymbol{e}_j^T \boldsymbol{w}_i &= \boldsymbol{e}_i^T \boldsymbol{w}_j \\
\Rightarrow \boldsymbol{e}_j^T (A_i(M - P')\Gamma^{-1})\boldsymbol{e}_i &= \boldsymbol{e}_i^T (A_j(M - P')\Gamma^{-1})\boldsymbol{e}_j \\
\Rightarrow \gamma_j \boldsymbol{e}_j^T A_i M \boldsymbol{e}_i - \gamma_i \boldsymbol{e}_i^T A_j M \boldsymbol{e}_j &= \gamma_j \boldsymbol{e}_j^T A_i P' \boldsymbol{e}_i - \gamma_i \boldsymbol{e}_i^T A_j P' \boldsymbol{e}_j
\end{aligned}$$

The last line can written:

$$\gamma_j \boldsymbol{e}_j^T A_i P \boldsymbol{e}_i - \gamma_i \boldsymbol{e}_i^T A_j P \boldsymbol{e}_j - (P'_{ij} - P_{ij}) \left(\gamma_j \boldsymbol{e}_j^T A_i \boldsymbol{e}_j + \gamma_i \boldsymbol{e}_i^T A_j \boldsymbol{e}_i \right).$$

Hence, we can write $(P'_{ij} - P_{ij})$ as:

$$\begin{aligned} P'_{ij} - P_{ij} &= \frac{1}{\gamma_j A_{i;j,j} + \gamma_i A_{j;i,i}} \Big(\boldsymbol{e}_i^T \Gamma A_j (M - P) \boldsymbol{e}_j - \boldsymbol{e}_j^T \Gamma A_i (M - P) \boldsymbol{e}_i \Big) \\ &= \frac{1}{Q_{i;j,j} + Q_{j;i,i}} \Big(\boldsymbol{e}_i^T Q_j (M - P) \boldsymbol{e}_j - \boldsymbol{e}_j^T Q_i (M - P) \boldsymbol{e}_i \Big) \end{aligned}$$

□ □

4.6.8 Proof of Theorem 4.2.15

First, we characterize pairwise negotiation dynamics as linear in the price updates.

Theorem 4.6.7. *Consider a network setting $(\boldsymbol{\mu}_i, \gamma_i, \Sigma_i, \Psi_i)_{i \in [n]}$. Define Q_i as in Theorem 4.2.8. Let $s_{ij} = 1$ if $\{i, j\}$ is a permitted edge and 0 otherwise. Let $L, R \in \mathbb{R}^{n^2 \times n^2}$ be diagonal matrices such that $L_{(i-1)n+j,(i-1)n+j} = Q_{i;jj} + Q_{j;ii}$ and $R_{(i-1)n+j,(i-1)n+j} = s_{ij}$, and $L^\dagger$ be the pseudoinverse of L. Let $\Delta_{(t+1)} = P(t+1) - P(t)$, where $P(t)$ is the price matrix at time step t of pairwise negotiations. Then,*

$$\begin{aligned} \mathrm{vec}(\Delta_{(t+1)}) &= R\Big(I_{n^2} - \eta L^\dagger K \Big) \mathrm{vec}(\Delta_{(t)}), \\ \textit{where } K &= \sum_{r=1}^{n} \big(\boldsymbol{e}_r \boldsymbol{e}_r^T \otimes Q_r + Q_r \otimes \boldsymbol{e}_r \boldsymbol{e}_r^T \big). \end{aligned}$$

Proof. Proof. Let $\{i, j\}$ be a permitted edge. From Proposition 4.6.6, we obtain:

$$\begin{aligned} (\Delta_{(t+1)})_{ij} &= \frac{\eta}{Q_{i;j,j} + Q_{j;i,i}} \Big(\boldsymbol{e}_i^T Q_j (M - P(t)) \boldsymbol{e}_j - \boldsymbol{e}_j^T Q_i (M - P(t)) \boldsymbol{e}_i \Big) \\ \Rightarrow (\Delta_{(t+1)})_{ij} - (\Delta_{(t)})_{ij} &= \frac{\eta}{Q_{i;j,j} + Q_{j;i,i}} \Big(\boldsymbol{e}_i^T Q_j (-\Delta_{(t)}) \boldsymbol{e}_j - \boldsymbol{e}_j^T Q_i (-\Delta_{(t)}) \boldsymbol{e}_i \Big) \\ &= \frac{-\eta}{Q_{i;j,j} + Q_{j;i,i}} \Big(\boldsymbol{e}_i^T Q_j \Delta_{(t)} \boldsymbol{e}_j - \boldsymbol{e}_j^T Q_i \Delta_{(t)} \boldsymbol{e}_i \Big) \\ &= \frac{-\eta}{Q_{i;j,j} + Q_{j;i,i}} \boldsymbol{e}_i^T \Big(Q_j \Delta_{(t)} - (Q_i \Delta_{(t)})^T \Big) \boldsymbol{e}_j \end{aligned}$$

Hence,

$$(Q_{i;j,j} + Q_{j;i,i}) \Big((\Delta_{(t+1)})_{ij} - (\Delta_{(t)})_{ij} \Big) = -\eta s_{ij} \cdot \boldsymbol{e}_i^T \Big(Q_j \Delta_{(t)} + \Delta_{(t)} Q_i \Big) \boldsymbol{e}_j.$$

We assumed that $\{i, j\}$ was a permitted edge above, but notice the identity is also true for prohibited $\{i, j\}$ since both the numerator and denominator become 0, and we can define

their ratio to be 0. Defining $Y_{ij} = \boldsymbol{e}_i^T \Big(Q_j \Delta_{(t)} + \Delta_{(t)} Q_i \Big) \boldsymbol{e}_j$, and recalling the definitions of L and R from the theorem statement, the above formula becomes

$$L\text{vec}(\Delta_{(t+1)} - \Delta_{(t)}) = -\eta R \text{vec}(Y). \tag{4.7}$$

We show next that $\text{vec}(Y) = K\text{vec}(\Delta_{(t)})$, where K is defined in the theorem statement. Let tr denote the trace operator. Then $(\boldsymbol{e}_j^T \otimes \boldsymbol{e}_i^T)\text{vec}(Y) = Y_{ij}$. Hence,

$$\begin{aligned}
Y_{ij} &= \boldsymbol{e}_i^T \Big(Q_j \Delta_{(t)} + \Delta_{(t)} Q_i \Big) \boldsymbol{e}_j \\
&= tr\Big(\boldsymbol{e}_i^T Q_j \Delta_{(t)} \boldsymbol{e}_j \Big) + tr\Big(\boldsymbol{e}_i^T \Delta_{(t)} Q_i \boldsymbol{e}_j \Big) \\
&= tr\Big(\boldsymbol{e}_j^T \Delta_{(t)}^T Q_j^T \boldsymbol{e}_i \Big) + tr\Big(\boldsymbol{e}_i^T \Delta_{(t)} Q_i \boldsymbol{e}_j \Big) \\
&= tr\Big(\Delta_{(t)}^T Q_j^T \boldsymbol{e}_i \boldsymbol{e}_j^T \Big) + tr\Big(Q_i \boldsymbol{e}_j \boldsymbol{e}_i^T \Delta_{(t)} \Big) \\
&= \text{vec}(\Delta_{(t)})^T \text{vec}(Q_j^T \boldsymbol{e}_i \boldsymbol{e}_j^T + (Q_i \boldsymbol{e}_j \boldsymbol{e}_i^T)^T) \\
&= \text{vec}(Q_j \boldsymbol{e}_i \boldsymbol{e}_j^T + \boldsymbol{e}_i \boldsymbol{e}_j^T Q_i)^T \text{vec}(\Delta_{(t)}),
\end{aligned}$$

where we used $Q_i = Q_i^T$.

Hence we need to show $(\boldsymbol{e}_j^T \otimes \boldsymbol{e}_i^T)K = \text{vec}(Q_j \boldsymbol{e}_i \boldsymbol{e}_j^T + \boldsymbol{e}_i \boldsymbol{e}_j^T Q_i)^T$. Letting δ denote the Kronecker delta, we obtain:

$$\begin{aligned}
(\boldsymbol{e}_j^T \otimes \boldsymbol{e}_i^T)K &= (\boldsymbol{e}_j^T \otimes \boldsymbol{e}_i^T) \Big(\sum_{r=1}^{n} \boldsymbol{e}_r \boldsymbol{e}_r^T \otimes Q_r + Q_r \otimes \boldsymbol{e}_r \boldsymbol{e}_r^T \Big) \\
&= \sum_{r=1}^{n} \Big(\delta_{jr} (\boldsymbol{e}_j^T \otimes \boldsymbol{e}_i^T Q_r) + \delta_{ir} (\boldsymbol{e}_j^T Q_r \otimes \boldsymbol{e}_i^T) \Big) \\
&= (\boldsymbol{e}_j^T \otimes \boldsymbol{e}_i^T Q_j) + (\boldsymbol{e}_j^T Q_i \otimes \boldsymbol{e}_i^T) \\
&= (\boldsymbol{e}_j \otimes Q_j \boldsymbol{e}_i + Q_i \boldsymbol{e}_j \otimes \boldsymbol{e}_i)^T .
\end{aligned} \tag{4.8}$$

Now, we observe that $\boldsymbol{e}_j \otimes Q_j \boldsymbol{e}_i$ is the vectorization of a matrix whose j^{th} column is $Q_j \boldsymbol{e}_i$, i.e., the matrix $Q_j \boldsymbol{e}_i \boldsymbol{e}_j^T$. Similarly, $Q_i \boldsymbol{e}_j \otimes \boldsymbol{e}_i$ is the vectorization of a matrix whose i^{th} row is $(Q_i \boldsymbol{e}_j)^T$, i.e., the matrix $\boldsymbol{e}_i \boldsymbol{e}_j^T Q_i$. Hence, $(\boldsymbol{e}_j^T \otimes \boldsymbol{e}_i^T)K = \text{vec}(Q_j \boldsymbol{e}_i \boldsymbol{e}_j^T + \boldsymbol{e}_i \boldsymbol{e}_j^T Q_i)^T$, as desired.

Plugging into Eq. (4.7),

$$\begin{aligned} L\text{vec}(\Delta_{(t+1)} - \Delta_{(t)}) &= -\eta RK\text{vec}(\Delta_{(t)}) \\ \Rightarrow L\text{vec}(\Delta_{(t+1)}) &= L\text{vec}(\Delta_{(t)}) - \eta RK\text{vec}(\Delta_{(t)}) \\ \Rightarrow \text{vec}(\Delta_{(t+1)}) &= \Big(L^\dagger L - \eta L^\dagger RK\Big)\text{vec}(\Delta_{(t)}) \\ \Rightarrow \text{vec}(\Delta_{(t+1)}) &= \Big(R - \eta RL^\dagger K\Big)\text{vec}(\Delta_{(t)}) \\ &= R\Big(I_{n^2} - \eta L^\dagger K\Big)\text{vec}(\Delta_{(t)}), \end{aligned}$$

where we used the facts that $(\Delta_t)_{ij} = (\Delta_{(t+1)})_{ij} = 0$ for disallowed edges, and $L^\dagger L = R$ and $LR = RL = L$, which can be easily confirmed by inspection of these diagonal matrices. □ □

We use Lyapunov theory to analyze the convergence of pairwise negotiation dynamics. In particular, we need the the discrete Lyapunov equation, also called the Stein equation.

Theorem 4.6.8 (Callier and Desoer (1994) 7.d)**.** *For the discrete-time dynamical system $\boldsymbol{x}_{t+1} = A\boldsymbol{x}_t$, with $\boldsymbol{x}_t \in \mathbb{R}^n$, the following are equivalent:*

1. *The system is globally asymptotically stable towards $\mathbf{0}$.*
2. *For any positive definite $R \in \mathbb{R}^{n\times n}$, there exists a unique solution $X \succ 0$ to the equation*

$$AXA^T - X = -R$$

3. *For any eigenvalue λ of A, $|\lambda| < 1$.*

Pairwise negotiation dynamics can be described as a discrete-time linear system in $\text{vec}(\Delta_t)$, where Δ_t is the price difference at time t. Clearly, the system converges iff Δ_t approaches zero. Therefore, we can use the Stein equation to prove global asymptotic stability conditions.

We will also need the *commutation matrix*.

Lemma 4.6.9 (Horn and Johnson (1994))**.** *Let $\Pi^{(n,n)} : \mathbb{R}^{n^2} \to \mathbb{R}^{n^2}$ be a permutation matrix (called the (n,n)* commutation matrix*) defined as $\Pi^{(n,n)} = \sum_{i=1}^{n}\sum_{j=1}^{n} \boldsymbol{e}_i\boldsymbol{e}_j^T \otimes \boldsymbol{e}_j\boldsymbol{e}_i^T$. Then for any $A, B \in \mathbb{R}^{n\times n}$, we have*

$$A \otimes B = \Pi^{(n,n)}(B \otimes A)(\Pi^{(n,n)})^T$$

Recall that for a linear operator T that $\sigma(T)$ denotes the eigenvalues of T. We are ready to prove Part 1 of Theorem 4.2.15.

Proposition 4.6.10 (Part 1 of Theorem 4.2.15)**.** *Let L, R, K be defined as in Theorem 4.6.7. For a matrix $X \in \mathbb{R}^{n^2 \times n^2}$ let $X\,|_R$ denote the principal submatrix of X corresponding to the nonzero rows/columns of R. Define $\eta^\star = \min_{\lambda \in \sigma((L^\dagger K)|_R)} \frac{2}{\lambda}$. Then, for any $\eta \in (0, \eta^\star)$, $\mathrm{vec}(\Delta_{(t)})$ is globally asymptotically stable towards $\mathbf{0}$.*

Proof. Proof of Proposition 4.6.10. Let $T = R(I - \eta L^\dagger K)$. By Theorem 4.6.8, the dynamics are globally asymptotically stable towards $\mathbf{0}$ iff for all $\lambda \in \sigma(T)$, we have $|\lambda| < 1$.

From Eq. (4.8) for a prohibited edge (i, j), we see that $(\boldsymbol{e}_j^T \otimes \boldsymbol{e}_i^T)K = \mathbf{0}^T$, since $Q_i \boldsymbol{e}_j = \mathbf{0} = Q_j \boldsymbol{e}_i$. Hence, $K = RK$. Taking transposes and noting that both K and R are symmetric, we find $KR = K$. Hence, $T = R(I - \eta L^\dagger K) = R(I - \eta L^\dagger K)R$, where we used $R^2 = R$. Thus, T is zero except for the principal submatrix corresponding to the nonzero columns of R. So, to apply Theorem 4.6.8, we only require $|\lambda| < 1$ for $\lambda \in \sigma(T\,|_R)$.

For clarity of exposition we will first consider the case where $R = I$ (no prohibited edges). Then, the eigenvalues of $T\,|_R = T$ equal $1 - \eta\lambda$, where $\lambda \in \sigma(L^{-1}K) = \sigma(L^{-1/2}KL^{-1/2})$ by a similarity transformation. Also, $K = U_1 + U_2$, where $U_1 = \sum_{r=1}^{n} \left(\boldsymbol{e}_r \boldsymbol{e}_r^T \otimes Q_r\right)$ and $U_2 := \sum_{r=1}^{n} \left(Q_r \otimes \boldsymbol{e}_r \boldsymbol{e}_r^T\right)$. The matrix U_1 is block diagonal with positive-definite blocks $Q_r \succ 0$, so $U_1 \succ 0$. By Lemma 5.8.5, U_2 is similar to U_1 via a permutation matrix, so $U_2 \succ 0$. Hence, $K \succ 0$, and $L^{-1/2}KL^{-1/2} \succ 0$. So, the eigenvalues of $L^{-1}K$ are real and positive. Hence, we have convergence iff for all $\lambda \in \sigma(L^{-1}K)$, we have $1 > (1 - \eta\lambda)^2 = 1 - 2\eta\lambda + \eta^2\lambda^2$. i.e., $\lambda < 2/\eta$. Hence, $\eta^\star = 2/\|L^{-1}K\|$ as required.

Now we consider the prohibited edges setting ($R \neq I$). Here, convergence occurs iff $|1 - \eta\lambda| < 1$ for all $\lambda \in \sigma((L^\dagger K)\ |_R)$. Since $RL^\dagger R = L^\dagger$ and $RKR = K$, we have $(L^\dagger K)\ |_R = L^\dagger\ |_R\ K\ |_R = (L\ |_R)^{-1} K\ |_R$. Arguing as above, it suffices to show that $K\ |_R \succ 0$. We claim $K\ |_R = V_1 + V_2$ where V_1 is a block diagonal matrix with i^{th} block equal to $(2\gamma_i \Psi_i \Sigma_i \Psi_i^T)^{-1} \succ 0$, and V_2 is similar to V_1 via Lemma 5.8.5. Hence $K\ |_R \succ 0$ and the expression for $\eta^\star$ follows. □ □

Proposition 4.6.11 (Part 2 of Theorem 4.2.15)**.** *We define $\eta^\star$ as in Proposition 4.6.10, and*

L, R, K, α as in Theorem 4.6.7. Let $\eta \in (0, \eta^\star)$. Then,

$$\|P(t) - P^\star\|_F \leq \frac{\alpha^t}{1-\alpha} \cdot \|P(1) - P(0)\|_F$$

Here, $P^\star$ is the stable point to which the negotiation converges.

Proof. Proof. Let β denote the greatest eigenvalue in absolute value of $R(I_{n^2} - \eta L^\dagger K)$. From Theorem 4.6.7, we have $\|\Delta_{t+1}\|_F \leq |\beta| \|\Delta_t\|_F$. Recall that $\lambda_{\max}, \lambda_{\min}$ denote largest and smallest eigenvalues of the matrix $(L^\dagger K)\,|_R$ respectively. Since $\|R\| = 1$, it follows that $|\beta| = \max\{|1 - \eta\lambda_{\min}|, |1 - \eta\lambda_{\max}|\} = \alpha$.

Then,

$$\begin{aligned}
\|P^\star - P(t)\|_F &\leq \sum_{i>t} \|\Delta_i\|_F \\
&\leq \|\Delta_t\|_F(\alpha + \alpha^2 + \ldots) \\
&\leq \|\Delta_t\|_F \frac{\alpha}{1-\alpha} \\
&\leq (\alpha^{t-1}\|\Delta_1\|_F)\frac{\alpha}{1-\alpha} \\
&= \|\Delta_1\|_F \frac{\alpha^t}{1-\alpha}
\end{aligned}$$

Since $\|\Delta_1\|_F = \|P(1) - P(0)\|_F$ we are done. □ □

4.6.9 Example of Convergence Conditions and Rate

The following example illustrates Theorem 4.2.15 in the setting of Example 4.6.2 (Appendix 4.6.3).

Example 4.6.12 (Convergence Conditions and Rate)**.** *In the setting of Example 4.6.2, we have*

$$Q_1 = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 2/3 & -1/3 \\ 0 & -1/3 & 2/3 \end{bmatrix}, Q_2 = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix},$$
$$Q_3 = \begin{bmatrix} 2 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix},$$

Hence

$$K = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & \frac{2}{3}+1 & \frac{-1}{3} & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & \frac{-1}{3} & \frac{2}{3}+2 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1+\frac{2}{3} & 0 & 0 & \frac{-1}{3} & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & \frac{-1}{3} & 0 & 0 & 2+\frac{2}{3} & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

Also, L is the diagonal matrix with $L_{i,i} = K_{i,i}$ for $i \in [9]$. Since the permitted edges are $\{(1,2),(1,3)\}$, $R = \{2,3\}$ and so $(L^\dagger K)_R = \begin{bmatrix} 1 & \frac{-1}{5} \\ \frac{-1}{8} & 1 \end{bmatrix}$. Hence $\lambda_{min} = 1 - \frac{1}{2\sqrt{10}}, \lambda_{max} = 1 + \frac{1}{2\sqrt{10}}$, and $\eta^* = \frac{2}{1+(40)^{-1/2}} \approx 1.727$.

It follows that pairwise negotiations with $\eta \in (0, \frac{2}{1+(40)^{-1/2}})$ are globally asymptotically stable. Suppose that $\eta = 0.99$. Then $\alpha = (1 - \eta \cdot (1 - \frac{1}{2\sqrt{10}})) \approx 0.17$. Hence after t rounds, the distance of $P(t)$ to P^* shrinks by a factor of $\approx \frac{0.17^t}{0.83}$.

4.6.10 Proof of Theorem 4.2.17

We will use a series of Lemmas to reduce the result of Theorem 4.2.17 to a matrix concentration inequality in each of the $\hat{\Sigma}_i$.

Lemma 4.6.13. *Let $\widehat{\eta^*}, \eta^*$ be as in Theorem 4.2.17. Suppose all edges are permitted.*

Suppose that for all $i \in [n]$, we have $\|\hat{\Sigma}_i^{-1} - \Sigma^{-1}\| = o(1)$. Then, $\hat{\eta}^\star > \eta^\star(1 - o(1))$.

Proof. Proof. Let $\hat{L}, \hat{K} \in \mathbb{R}^{n^2 \times n^2}$ be as in Theorem 4.2.17, but built using $\hat{\Sigma}_1, \ldots, \hat{\Sigma}_n$ instead of $\Sigma, \ldots, \Sigma$. Let L, K be defined similarly to $\hat{L}, \hat{K}$ but using Σ in place of all $\hat{\Sigma}_i$.

Then $\widehat{\eta^*} := \frac{2}{\max \sigma(\hat{L}^{-1}\hat{K})}$ and $\eta^* := \frac{2}{\max \sigma(L^{-1}K)}$.

Let $\epsilon_L, \epsilon_K \in \mathbb{R}^{n^2 \times n^2}$ be such that $\hat{L}^{-1} = L^{-1} + \epsilon_L$ and $\hat{K} = K + \epsilon_K$. We will bound $\|\epsilon_L\|, \|\epsilon_K\|$.

Let $Q_i, \hat{Q}_i$ be defined as in Theorem 4.2.8, so $Q_i := (2\gamma_i \Sigma)^{-1}$ and $\hat{Q}_i := (2\gamma_i \hat{\Sigma}_i)^{-1}$. Let $\alpha = \max_{i \in [n]} \|\hat{Q}_i - Q_i\|$. Notice $\|\Gamma^{-1}\| = O(1)$, so $\alpha = o(1)$.

First, since L is diagonal, $\|\epsilon_L\| \leq \max_{i,j \in [n]} \left((\hat{Q}_{i;jj} - Q_{i;jj}) + (\hat{Q}_{j;ii} - Q_{j;ii})\right) \leq 2 \max_{i,j \in [n]} \left(\hat{Q}_{i;jj} - Q_{i;jj}\right) \leq 2 \max_{i \in [n]} \|\hat{Q}_i - Q_i\| = 2a$.

Second, let $\hat{K} := \hat{U}_1 + \hat{U}_2$ where $\hat{U}_1, \hat{U}_2$ are defined analogously to U_1, U_2 in the proof of Theorem 4.2.15. Letting Π be the (n, n) commutation matrix of Lemma 5.8.5, we know $\hat{U}_2 = \Pi \hat{U}_1 \Pi^T$, so $\|\epsilon_K\| \leq 2\|\hat{U}_1 - U_1\|$. Since $U_1, \hat{U}_1$ are block diagonal with i^{th} blocks $Q_i, \hat{Q}_i$ respectively, it follows $\|\hat{U}_1 - U_1\| = \max_{i \in [n]} \|\hat{Q}_i - Q_i\| = \alpha$. Hence $\|\epsilon_K\| \leq 2\alpha$.

Third, notice that since $\|\Sigma\|$ and $\|\Gamma\|$ are assumed to be $O(1)$ that $\|L^{-1}\| = O(\max_i \|Q_i\|) = O(1)$ and $\|K\| = O(\max_i \|Q_i\|) = O(1)$. So,

$$\begin{aligned}
\|\hat{L}^{-1}\hat{K} - L^{-1}K\|_2 &\leq \|\epsilon_L\|\|K\| \\
&\quad + \|L^{-1}\|\|\epsilon_K\| + \|\epsilon_L\|\|\epsilon_K\| \\
&\leq 2\alpha(\|K\| + 2\alpha) \\
&\quad + 4\alpha(\|L^{-1}\| + \alpha) \\
&= 4\alpha(\|K\| + \|L^{-1}\|) + 8\alpha^2 \\
&\leq o(1)
\end{aligned}$$

We conclude that $\|\hat{L}^{-1}\hat{K}\|_2 \leq \|L^{-1}K\|_2 + o(1)$, so $\widehat{\eta^*} \geq \frac{\eta^*}{1+(o(1)/\|L^{-1}K\|)} \geq (1 - o(1))\eta^*$. □ □

Lemma 4.6.14. *Suppose for $i \in [n]$, we have $\delta_i := \|\hat{\Sigma}_i - \Sigma\| = o(1)$. Then $\|\hat{\Sigma}_i^{-1} - \Sigma_i^{-1}\| = o(1)$.*

Proof. Proof. Weyl's inequality implies that $\lambda_{min}(\hat{\Sigma}_i) \geq \lambda_{min}(\Sigma) - \|\hat{\Sigma}_i - \Sigma\|$. Therefore,

$$\begin{aligned}
\|\hat{\Sigma}_i^{-1}\| &= \frac{1}{\lambda_{min}(\hat{\Sigma}_i)} \\
&\leq \frac{1}{\lambda_{min}(\Sigma) - \delta_i} \\
&= \frac{1}{\lambda_{min}(\Sigma)}\Big(1 + \frac{\delta_i}{\lambda_{min}(\Sigma)} + O\Big(\Big(\frac{\delta_i}{\lambda_{min}(\Sigma)}\Big)^2\Big)\Big) \\
&= \|\Sigma^{-1}\|(1 + o(1)) \\
\Rightarrow \|\hat{\Sigma}_i^{-1} - \Sigma^{-1}\| &= \|\Sigma^{-1}(\Sigma_i - \hat{\Sigma}_i)\hat{\Sigma}_i^{-1}\| \\
&\leq (1 + o(1))\|\Sigma^{-1}\|^2\delta_i \\
&\leq o(1)
\end{aligned}$$

The last step follows from the fact $\|\Sigma^{-1}\| = O(1)$. □ □

The hypothesis of Lemma 4.6.14 follows from a standard argument on the concentration of random covariance matrices.

Theorem 4.6.15. *Under the setting of Theorem 4.2.17, with probability at least* $1 - e^{-\Omega(n)}$, *we have* $\|\hat{\Sigma}_i - \Sigma\| = o(1)$ *for all* $i \in [n]$.

Proof. Proof of Theorem 4.6.15. Let $\boldsymbol{X}_1, \ldots, \boldsymbol{X}_m \overset{\text{iid}}{\sim} \mathcal{N}(\mathbf{0}, \Sigma)$ be the samples. Let $\hat{\boldsymbol{\mu}} = \frac{1}{m}\sum_{i=1}^{m} \boldsymbol{X}_i$, and $\tilde{\Sigma}_i := \frac{1}{m}\sum_{i=1}^{m} \boldsymbol{X}_i \boldsymbol{X}_i^T$. Then, $\hat{\Sigma}_i = m/(m-1) \cdot (\tilde{\Sigma}_i - \hat{\boldsymbol{\mu}}\hat{\boldsymbol{\mu}}^T)$. Hence,

$$\|\hat{\Sigma}_i - \Sigma\| \leq m/(m-1) \cdot \left(\|\tilde{\Sigma}_i - \Sigma\| + \|\hat{\boldsymbol{\mu}}\hat{\boldsymbol{\mu}}^T\|\right) = m/(m-1)\left(\|\tilde{\Sigma}_i - \Sigma\| + \|\hat{\boldsymbol{\mu}}\|^2\right).$$

Now, $\hat{\boldsymbol{\mu}} \sim \mathcal{N}(0, \frac{1}{m}\Sigma)$, so $\sqrt{m}\Sigma^{-1/2}\hat{\boldsymbol{\mu}} \sim \mathcal{N}(0, I_n)$. By Vershynin (2018a) (4.7.3 and 2.8.3), there exist constants $c, c_2 > 0$ such that for any $u, \epsilon > 0$,

$$\begin{aligned}
&\mathbb{P}\left[\|\tilde{\Sigma}_i - \Sigma\|_2 \leq c\|\Sigma\|_2\left(\sqrt{\frac{n+u}{m}} + \frac{n+u}{m}\right)\right] \\
&\geq 1 - 2e^{-u}, \\
&\mathbb{P}\left[\left|\frac{1}{n}\|\sqrt{m}\Sigma^{-1/2}\hat{\boldsymbol{\mu}}\|_2^2 - 1\right| \leq \epsilon\right] \\
&\geq 1 - 2e^{-c_2 n \min(\epsilon, \epsilon^2)}
\end{aligned}$$

Now we set $\epsilon > 1$ and $u = c_3 n$ for some constant $c_3 > 0$. Then, when $m = \lceil n \log n \rceil$, we have $(n+u)/m = o(1)$ Then, with probability at least $1 - 2e^{-c_3 n} - 2e^{-c_2 \epsilon n}$, we have

$$\begin{aligned}
\|\tilde{\Sigma}_i - \Sigma\|_2 &\leq \|\Sigma\| \cdot o(1), \\
\text{and } \|\Sigma^{-1/2}\hat{\boldsymbol{\mu}}\|_2^2 &\leq \frac{(1+\epsilon)n}{m} \\
\Rightarrow \|\hat{\boldsymbol{\mu}}\|^2 &\leq \frac{(1+\epsilon)n\|\Sigma\|}{m} = \|\Sigma\| \cdot o(1), \\
\Rightarrow \|\hat{\Sigma}_i - \Sigma\| &\leq \|\Sigma\| \cdot o(1).
\end{aligned}$$

Choosing large enough c_3 and ϵ, this statement holds for all $i \in [n]$ with probability greater than $1 - e^{logn - c_4 n} = 1 - e^{-\Omega(n)}$. □ □

Theorem 4.2.17 follows easily.

Proof. Proof of Theorem 4.2.17 When all edges are permitted, the proof follows from Theorem 4.6.15, Lemma 4.6.13, and Lemma 4.6.14.

If there are prohibited edges, then we must use matrix concentration to bound $\max \sigma(\hat{L}^{\dagger}\hat{K})$ instead of $\max \sigma(\hat{L}^{-1}\hat{K})$. Notice that prohibited edges have the effect of simply

zeroing out certain rows and columns of Q_i, so that $Q_i := \Psi_i(2\gamma_i \Psi_i^T \Sigma_i \Psi_i)^{-1}\Psi_i^T$, rather than $(2\gamma_i\Sigma_i)^{-1}$. Therefore, we can use Theorem 4.6.15 to bound $\|\Psi_i^T\hat{\Sigma}_i\Psi_i - \Psi_i^T\Sigma\Psi_i\|$ for all i, and then prove the appropriate analogue of Lemma 4.6.13. In particular, the sample size requirement remains the same. □ □

4.6.11 Proof of Proposition 4.2.19

Recall that in Assumption 4.2.18 we assumed that $M_{ij}(t)$ varies independently according to a Brownian motion with the same parameters for all (i, j). To avoid ambiguity, we recall the definition of a standard Brownian motion as follows.

Definition 4.6.16 (Brownian Motion)**.** *For $d \geq 1$, a d-dimensional Brownian motion with scale parameter $\sigma > 0$ is a stochastic process $\{\boldsymbol{X_t} : t \geq 0\}$ such that $\boldsymbol{X_t} \in \mathbb{R}^d$ for all t, the components of $\boldsymbol{X_t}$ are independent, and for all $j \in [d]$,*

i) The process $\{(\boldsymbol{X_t})_j : t \geq 0\}$ has independent increments.

ii) For $r > 0$, the increment $(\boldsymbol{X_{t+r}})_j - (\boldsymbol{X_t})_j$ is distributed as $N(0, r\sigma^2)$.

iii) With probability 1, the function $t \mapsto \boldsymbol{X_t}$ is continuous on $[0, \infty)$.

We can derive the SDP of Proposition 4.2.19 as follows.

Proposition 4.6.17. *Under Assumption 4.2.18, the maximum likelihood estimator for Σ is the unique $\Sigma \succ 0$ such that $tr\Sigma = 1$ and*

- **Consistency***: For all $t \in [T]$,*

$$W(t)\Sigma + \Sigma W(t) = \frac{1}{2}(M(t) + M(t)^T)$$
$$\text{for some } M(1), M(2), \ldots M(t)$$

- **Minimum mean shift***: The resulting $M(1), \ldots, M(T)$ minimize the objective*

$$\sum_{t=1}^{T-1} \|M(t+1) - M(t)\|_F^2$$

Proof of Proposition 4.6.17.

$$\begin{aligned}
&\mathbb{P}(M(1),\dots,M(T)\mid W(1),\dots,W(T),\Sigma)\\
\propto&\,\mathbb{P}(W(1),\dots W(T)\mid M(1),\dots,M(T),\Sigma)\cdot\mathbb{P}(M(1),\dots,M(T)\mid\Sigma)\\
=&\left(\prod_{t=1}^{T}1_{W(t)\Sigma+\Sigma W(t)=0.5(M(t)+M(t)^T)}\right)\cdot\left(\prod_{t=1}^{T-1}\mathbb{P}(M(t+1)-M(t))\right)\\
=&\left(\prod_{t=1}^{T}1_{\mathrm{vec}(W(t))=0.5(\Sigma\otimes I+I\otimes\Sigma)\mathrm{vec}(M(t)+M(t)^T)}\right)\cdot\left(\prod_{t=1}^{T-1}\exp(-\frac{\|\mathrm{vec}(M(t+1)-M(t))\|^2}{2\sigma^2})\right),
\end{aligned}$$

where the first step follows from Bayes' Rule, the second step from Corollary 4.6.1, and the third from Assumption 4.2.18. The theorem follows from the observation that for any matrix X, we have $\|\mathrm{vec}(X)\|^2=\|X\|_F^2$. □

The proof of Proposition 4.2.19 follows easily.

Proof. Proof of Proposition 4.2.19. By Proposition 4.6.17, we obtain the SDP

$$\begin{aligned}
&\min_{\Sigma}\sum_{t=1}^{T-1}\|M(t+1)-M(t)\|_F^2\\
&\qquad\forall t\in[T]:W(t)\Sigma+\Sigma W(t)=\frac{1}{2}(M(t)+M(t)^T)
\end{aligned}$$

under the assumptions of $\Sigma\succ 0$ and $\mathrm{tr}(\Sigma)=1$. Since the Frobenius norm is invariant under transposes, we have

$$\sum_{t=1}^{T-1}\|M(t+1)-M(t)\|_F^2\propto\sum_{t=1}^{T-1}\|(M(t+1)+M(t+1)^T)-(M(t)+M(t)^T)\|_F^2.$$

We can replace $M(t)+M(t)^T$ with $2W(t)\Sigma+2\Sigma W(t)$ for all $t\in[T]$ to obtain the equivalent objective function $\sum_{t=1}^{T-1}\|(W(t+1)-W(t))\Sigma+\Sigma(W(t+1)-W(t))\|_F^2$ (up to a constant). This substitution enforces the fixed point equation $W(t)\Sigma+\Sigma W(t)=\frac{1}{2}(M(t)+M(t)^T)$ for all $t\in[T]$, so the conclusion follows. □ □

Remark 4.6.18 (The prohibited edges setting.)**.** *Proposition 4.2.19 generalizes straightforwardly to the setting of prohibited edges. Let E denote the set of permitted edges. Then minimum mean shift assumption is equivalent to minimizing $\sum_{t=1}^{T-1}\sum_{\{i,j\}\in E}\big(M(t+1)+M(t+1)^T-M(t)-M(t)^T\big)_{ij}^2$. In words, the objective just zeroes out prohibited edges, since mean estimates*

for prohibited edges have no effect on the network. For a network setting $(\boldsymbol{\mu}_j, \Sigma, \gamma_j, \Psi_j)_{j\in[n]}$*, some algebra gives* $M(t)_{ij} = \boldsymbol{e}_i^T 2\gamma_j(\Psi_j^T\Psi_j)\Sigma(\Psi_j^T\Psi_j)W(t)\boldsymbol{e}_j$*. Notice* $\Psi_j^T\Psi_j \in \mathbb{R}^n$ *is a diagonal matrix with* $(\Psi_j^T\Psi_j)_{ii} = 1$ *if* $\{i,j\} \in E$ *and zero otherwise. Therefore, it is clear that upon substitution, the objective is an SDP in* Σ *with the same constraints.*

4.6.12 Proof of Theorem 4.3.1

Proof. Proof of Theorem 4.3.1. Note that the Hessian of $F_i(\boldsymbol{w}_i)$ is a positive diagonal matrix due to the conditions on $F_i(.)$. So, any stationary point is a local maximum. Hence, it suffices to show the existence of a unique stationary point.

Let $R(W)$ be an $n \times n$ matrix whose $(i,j)^{th}$ entry $R(W)_{ij} := f_{ij}(W_{ij})$. If a stable point (W, P) exists, it must satisfy $W = W^T$, $P = -P^T$, and

$$M - P = 2\Sigma W\Gamma + R, \tag{4.9}$$

following the same steps as the proof for Corollary 4.6.1. Adding this equation to its transpose, the stable point must satisfy

$$(M + M^T)/2 = (\Sigma W\Gamma + \Gamma W\Sigma) + (R(W) + R(W)^T)/2.$$

For a stable point, $[R(W)+R(W)^T]_{ij} = f_{ij}(W_{ij})+f_{ji}(W_{ji}) = (f_{ij}+f_{ji})(W_{ij})$, using $W = W^T$. Define $S(W)$ to be an $n \times n$ matrix with $S(W)_{ij} = (1/2)\cdot(f_{ij} + f_{ji})(W_{ij})$. Hence, the stable point must satisfy

$$\begin{aligned} (M + M^T)/2 &= S(W) + (\Sigma W\Gamma + \Gamma W\Sigma) \\ \Leftrightarrow \mathrm{vec}((M + M^T)/2) &= \mathrm{vec}(S(W)) + \underbrace{(\Gamma \otimes \Sigma + \Sigma \otimes \Gamma)}_{Q}\mathrm{vec}(W). \end{aligned} \tag{4.10}$$

Note that Q is positive-definite (from the proof of Corollary 4.6.1), and each entry of $\mathrm{vec}(S(W))$ is a function of the corresponding entry of $\mathrm{vec}(W)$. By Theorems 1 and 2 of Sandberg and Willson (1972), Eq. (4.10) has a unique solution if (1) for all diagonal $D \succ 0$, $\det(D+Q) > 0$ and (2) for any $\boldsymbol{x}, \boldsymbol{y} \in \mathbb{R}^{n^2}$ such that $\boldsymbol{x} = Q\boldsymbol{y}$, we have $\boldsymbol{x}^T\boldsymbol{y} \geq 0$. The first condition holds because $\det(D+Q) = \det(D^{1/2}(I+D^{-1/2}QD^{-1/2})D^{1/2}) = \det(D)\cdot\det(I+D^{-1/2}QD^{-1/2}) > 0$. The second condition is true because $\boldsymbol{x}^T\boldsymbol{y} = \boldsymbol{y}^TQ\boldsymbol{y} \geq 0$. Hence, Eq. (4.10) has a unique solution $\mathcal{W}$.

We now show that this solution satisfies the conditions of the stable point, that is, $\mathcal{W} = \mathcal{W}^T$, and there exists a skew-symmetric P which satisfies Eq. (4.9). Observe that

$$\begin{aligned}[S(\mathcal{W})^T]_{ij} &= S(\mathcal{W})_{ji} \\ &= (1/2) \cdot (f_{ij} + f_{ji})(\mathcal{W}_{ji}) \\ &= S(\mathcal{W}^T)_{ij},\end{aligned}$$

so $S(\mathcal{W})^T = S(\mathcal{W}^T)$. Taking the transpose of Eq. (4.10) and using $\Sigma = \Sigma^T$, $\Gamma = \Gamma^T$, and $S(\mathcal{W})^T = S(\mathcal{W}^T)$, we find

$$(M + M^T)/2 = (\Sigma\mathcal{W}^T\Gamma + \Gamma\mathcal{W}^T\Sigma) + S(\mathcal{W}^T).$$

But since there is only one solution to Eq. (4.10), we must have $\mathcal{W} = \mathcal{W}^T$.

Finally, we choose

$$\begin{aligned}P &= M - 2\Sigma\mathcal{W}\Gamma - R \\ \Rightarrow P + P^T &= (M + M^T) - 2(\Sigma\mathcal{W}\Gamma + \Gamma\mathcal{W}\Sigma) - 2S(W) \\ &= 0,\end{aligned}$$

using the fact that $\mathcal{W} = \mathcal{W}^T$ is a solution for Eq. (4.10). Hence, this choice of P is both skew-symmetric and satisfies Eq. (4.9). □ □

4.6.13 Proof of Theorem 4.3.3

Proof. Proof of Theorem 4.3.3. 1. Let $(\lambda_i, \boldsymbol{v}_i)$ denote the i^{th} eigenvalue and eigenvector of $\Gamma^{-1/2}\Sigma\Gamma^{-1/2}$, and let $V_{ij} = \boldsymbol{e}_i^T\boldsymbol{v}_j$. By Corollary 4.6.1,

$$\begin{aligned}W &= \Gamma^{-1/2}\left(\sum_{s=1}^{n}\sum_{t=1}^{n}\frac{\boldsymbol{v}_s^T\Gamma^{-1/2}(M + M^T)\Gamma^{-1/2}\boldsymbol{v}_t}{2(\lambda_r + \lambda_s)}\boldsymbol{v}_s\boldsymbol{v}_t^T\right)\Gamma^{-1/2} \\ \Rightarrow \frac{\partial W_{ij}}{\partial M_{k\ell}} &= \boldsymbol{e}_i^T\Gamma^{-1/2}\left(\sum_{s=1}^{n}\sum_{t=1}^{n}\frac{\boldsymbol{v}_s^T\Gamma^{-1/2}(\boldsymbol{e}_k\boldsymbol{e}_\ell^T + \boldsymbol{e}_\ell\boldsymbol{e}_k^T)\Gamma^{-1/2}\boldsymbol{v}_t}{2(\lambda_s + \lambda_t)}\boldsymbol{v}_s\boldsymbol{v}_t^T\right)\Gamma^{-1/2}\boldsymbol{e}_j \\ &= \frac{1}{2\sqrt{\gamma_i\gamma_j\gamma_k\gamma_\ell}}\left(\sum_{s=1}^{n}\sum_{t=1}^{n}\frac{\boldsymbol{v}_s^T(\boldsymbol{e}_k\boldsymbol{e}_\ell^T + \boldsymbol{e}_\ell\boldsymbol{e}_k^T)\boldsymbol{v}_t}{(\lambda_s + \lambda_t)}(\boldsymbol{e}_i^T\boldsymbol{v}_s)(\boldsymbol{v}_t^T\boldsymbol{e}_j)\right) \\ &= \frac{1}{2\sqrt{\gamma_i\gamma_j\gamma_k\gamma_\ell}}\sum_{s=1}^{n}\sum_{t=1}^{n}\left(\frac{V_{is}V_{ks}V_{jt}V_{\ell t} + V_{is}V_{\ell s}V_{jt}V_{kt}}{\lambda_s + \lambda_t}\right)\end{aligned}$$

This proves Eq. (4.3). If $i = k, j = \ell$, we have:

$$\begin{aligned}
\frac{\partial W_{ij}}{\partial M_{ij}} &= (2\gamma_i\gamma_j)^{-1}\left(\sum_{s=1}^{n}\sum_{t=1}^{n}\underbrace{\frac{V_{is}^2V_{jt}^2 + V_{is}V_{js}V_{jt}V_{it}}{\lambda_s+\lambda_t}}_{Z_{st}}\right)\\
&= (4\gamma_i\gamma_j)^{-1}\left(\sum_{s=1}^{n}\sum_{t=1}^{n}Z_{st} + \sum_{t=1}^{n}\sum_{s=1}^{n}Z_{ts}\right)\\
&= (4\gamma_i\gamma_j)^{-1}\sum_{s=1}^{n}\sum_{t=1}^{n}(Z_{st}+Z_{ts})\\
&= (4\gamma_i\gamma_j)^{-1}\sum_{s=1}^{n}\sum_{t=1}^{n}\frac{(V_{is}V_{jt}+V_{js}V_{it})^2}{\lambda_r+\lambda_s} > 0.
\end{aligned}$$

Hence, W_{ij} is monotonically increasing with respect to M_{ij}.

2. This follows from Corollary 4.6.1.

3. By Corollary 4.6.1, $\mathrm{vec}(W) = \gamma^{-1}(\Sigma\otimes I + I\otimes\Sigma)^{-1}\mathrm{vec}(\frac{M+M^T}{2})$. Let $K = \gamma(\Sigma\otimes I + I\otimes\Sigma)$ and $K' = \gamma(\Sigma'\otimes I + I\otimes\Sigma')$. Since $\Sigma' \succ \Sigma$ it follows that $K' \succ K$. Therefore $K^{-1} \succ (K')^{-1}$.

So, since $\mathrm{vec}(W'-W) = ((K')^{-1}-K^{-1})\mathrm{vec}(\frac{M+M^T}{2})$, we immediately obtain $\frac{1}{2}\mathrm{vec}(M+M^T)^T\mathrm{vec}(W'-W) < 0$. Since W, W' are symmetric it follows that $\mathrm{vec}(M^T)^T\mathrm{vec}(W'-W) = \mathrm{vec}(M)^T\mathrm{vec}(W'-W)$. So we have $\mathrm{vec}(M)^T\mathrm{vec}(W'-W) < 0$.

Since $\mathrm{vec}(M)^T\mathrm{vec}(W'-W) = \mathrm{tr}(M^T(W'-W))$, the conclusion follows. □ □

4.6.14 Hardness of Source Detection

We begin by defining

$$\left|\frac{\partial W_{ij}}{\partial M_{k\ell}}\right|_{approx} := \frac{|V_{in}V_{kn}V_{jn}V_{\ell n}|}{2\lambda_n}. \tag{4.11}$$

This approximates the right hand side of Eq. (4.3) when the term corresponding to the smallest eigenvalue λ_n dominates the sum. We now show that if the corresponding eigenvector $\boldsymbol{v}_n$ is random, source detection becomes difficult.

Proposition 4.6.19 (Hardness of Source Detection)**.** *Suppose $\boldsymbol{v}_n$ is Haar-distributed, that is, $\boldsymbol{v}_n$ is distributed uniformly on the unit sphere S^{n-1}. Then, if $\Sigma = V\Lambda V^T$ and $\Gamma = I$,*

$$\mathbb{P}\left[\max_{i,j\in[n]:(i,j)\neq(k,\ell)}\left|\frac{\partial W_{ij}}{M_{k\ell}}\right|_{approx} < \left|\frac{\partial W_{k\ell}}{M_{k\ell}}\right|_{approx}\right] \leq O\left(\frac{1}{n}\right).$$

Proof. Proof of Proposition 4.6.19. Without loss of generality we can set $k = 1, \ell = 2$ (the analysis of $k = \ell$ is identical). Notice that $\left|\frac{\partial W_{ij}}{M_{k\ell}}\right|_{approx}$ is maximized at the (i, j) that maximizes $|V_{in}V_{jn}|$.

Now, consider $(i, j) \in \{(1, 2), (3, 4), \ldots, (n-1, n)\}$. Notice the distribution of $\boldsymbol{v}_n$ is permutation-invariant by assumption. Hence the joint distribution of (V_{in}, V_{jn}) is the same for all such pairs (i, j). Hence the distribution of $|V_{in}V_{jn}|$ is also the same for all such (i, j). Therefore,

$$\mathbb{P}\left[\arg\max_{(i,j)\in\{(1,2),(3,4),\ldots,(n-1,n)\}} \left|\frac{\partial W_{ij}}{M_{12}}\right|_{approx} = (1, 2)\right] \leq O(1/n). \qquad \square$$

$\square$

4.6.15 Proof of Proposition 4.4.2

Proof. Proof of Proposition 4.4.2. Let $H = \frac{1}{2}(M + M^T)$. The fixed point equation for W is given by Corollary 4.6.1 as $\Sigma W\Gamma + \Gamma W\Sigma = H$. Vectorization implies $(\Gamma \otimes \Sigma + \Sigma \otimes \Gamma)\text{vec}(W) = \text{vec}(H)$. Let $X \sim Y$ denote that a pair of random variables X, Y are identically distributed. We want to show $\Pi^T W\Pi \sim W$. Vectorization gives $\text{vec}(\Pi^T W\Pi) = (\Pi^T \otimes \Pi^T)\text{vec}(W)$. Let $P = (\Pi^T \otimes \Pi^T)$ and $K = (\Gamma \otimes \Sigma + \Sigma \otimes \Gamma)$ for shorthand.

In this notation, we want to show that $PK^{-1}\text{vec}(H) \sim K^{-1}\text{vec}(H)$. Since P is a permutation, we have $PK^{-1}\text{vec}(H) = PK^{-1}P^T P\text{vec}(H) = (PKP^T)^{-1}P\text{vec}(H)$. Since the collections of random variables $\{\epsilon_i\}_i$ and $\{\epsilon'_{\theta_{i,j}}\}_{i,j}$ are independent, we know $\text{vec}(H)$ and K are independent. So to show $(PKP^T)^{-1}P\text{vec}(H) \sim K^{-1}\text{vec}(H)$ it suffices to show that $P\text{vec}(H) \sim \text{vec}(H)$ and $PKP^T \sim K$.

Notice $P\text{vec}(H) = \text{vec}(\Pi^T H\Pi)$. Hence, we want to show $\Pi^T H\Pi \sim H$, which holds iff $\Pi^T(M + M^T)\Pi \sim M + M^T$. Notice that $\Pi^T M^T\Pi = (\Pi^T M\Pi)^T$, so if $\Pi^T M\Pi \sim M$ then we obtain $\Pi^T M^T\Pi \sim M^T$ as well. It suffices to show $\Pi^T M\Pi \sim M$.

Similarly, we can simplify $PKP^T = \Pi^T\Sigma\Pi \otimes \Pi^T\Gamma\Pi + \Pi^T\Gamma\Pi \otimes \Pi^T\Sigma\Pi$. It suffices to show $\Pi^T\Gamma\Pi \sim \Gamma$ and $\Pi^T\Sigma\Pi = \Sigma$.

We are left to show that $\Pi^T\Sigma\Pi = \Sigma$ and $\Pi^T A\Pi \sim A$ for $A \in \{\Gamma, M\}$.

Proof of $\Pi^T \Sigma \Pi = \Sigma$. Let $i, j \in [n]$. Then $(\Pi^T \Sigma \Pi)_{ij} = \Sigma_{\pi(i),\pi(j)} = g(\theta_{\pi(i)}, \theta_{\pi(j)})$. Since π only commutes members within communities, $g(\theta_{\pi(i)}, \theta_{\pi(j)}) = g(\theta_i, \theta_j) = \Sigma_{ij}$. So $\Pi^T \Sigma \Pi = \Sigma$.

Proof of $\Pi^T \Gamma \Pi \sim \Gamma$. Notice $\Pi^T \Gamma \Pi$ and Γ are both diagonal. Let $i \in [n]$. Then $(\Pi^T \Gamma \Pi)_{ii} = \Gamma_{\pi(i),\pi(i)} = h(\theta_{\pi(i)}) + \epsilon_{\pi(i)} = h(\theta_i) + \epsilon_{\pi(i)}$. Since $\theta_i = \theta_{\pi(i)}$, we know $\epsilon_i \sim \epsilon_{\pi(i)}$. The conclusion follows.

Proof of $\Pi^T M \Pi \sim M$. Let $i, j \in [n]$. Then $(\Pi^T M \Pi)_{ij} = M_{\pi(i),\pi(j)} = f(\theta_{\pi(i)}, \theta_{\pi(j)}) + \epsilon'_{\theta_{\pi(i)},\pi(j)} = f(\theta_i, \theta_j) + \epsilon'_{\theta_i,\pi(j)}$. Since $\theta_j = \theta_{\pi(j)}$, we know that $\epsilon'_{\theta_i,\pi(j)} \sim \epsilon'_{\theta_i,j}$, and the conclusion follows. □ □

4.6.16 Proof of Theorem 4.4.4

Proof. Proof of Theorem 4.4.4. First, consider the network settings S and S'. Let $\Gamma \in \mathbb{R}^{n \times n}$ be a diagonal matrix with $\Gamma_{i,i} = \gamma_i$; define Γ' similarly under S'. Let the corresponding networks be W and W', and let $\Delta_W = W' - W$ and $\Delta_\Gamma = \Gamma' - \Gamma$. By Corollary 4.6.1, we have

$$\begin{aligned}
\Sigma W \Gamma + \Gamma W \Sigma &= \frac{M + M^T}{2} \\
&= \Sigma W' \Gamma' + \Gamma' W' \Sigma \\
\Rightarrow \frac{M + M^T}{2} &= \Sigma(W + \Delta_W)(\Gamma + \Delta_\Gamma) + (\Gamma + \Delta_\Gamma)(W + \Delta_W)\Sigma \\
&= \Sigma W \Gamma + \Gamma W \Sigma + \Sigma \Delta_W \Gamma + \Gamma \Delta_W \Sigma + \Sigma W \Delta_\Gamma + \Delta_\Gamma W \Sigma \\
&\quad + \Sigma \Delta_W \Delta_\Gamma + \Delta_\Gamma \Delta_W \Sigma \\
\Rightarrow \Sigma \Delta_W \Gamma + \Gamma \Delta_W \Sigma &= -(\Sigma W \Delta_\Gamma + \Delta_\Gamma W \Sigma + \Sigma \Delta_W \Delta_\Gamma + \Delta_\Gamma \Delta_W \Sigma) \\
= -(\Sigma W' \Delta_\Gamma + \Delta_\Gamma W' \Sigma) &
\end{aligned} \tag{4.12}$$

Next, consider S versus $S^\dagger$. Suppose that $M^\dagger$ has columns $\boldsymbol{\mu}_1^\dagger, \ldots, \boldsymbol{\mu}_n^\dagger$ and let $\Delta_M = M^\dagger - M$. Let $W^\dagger$ be the fixed point network under $S^\dagger$, given by $\Sigma W^\dagger \Gamma + \Gamma W^\dagger \Sigma = \frac{M^\dagger + (M^\dagger)^T}{2}$. Let $\Delta_W^\dagger = W^\dagger - W$. Then a similar argument gives:

$$\frac{\Delta_M + \Delta_M^T}{2} = \Sigma \Delta_W^\dagger \Gamma + \Gamma \Delta_W^\dagger \Sigma \tag{4.13}$$

Therefore, from Eq (4.12) and (4.13), it follows that $W' = W^\dagger$ if

$$\frac{\Delta_M + \Delta_M^T}{2} = -(\Sigma W' \Delta_\Gamma + \Delta_\Gamma W' \Sigma).$$

Hence, $W' = W^\dagger$ if we set $\Delta_M = -\Sigma W' \Delta_\Gamma$.

It remains to show that $M^\dagger$ differs from M only in columns corresponding to J. Suppose that $i \notin J$. Then $\gamma_i = \gamma_i'$, so $\Delta_\Gamma \boldsymbol{e}_i = \mathbf{0}$. We conclude that $\Delta_M \boldsymbol{e}_i = \mathbf{0}$ and hence $M\boldsymbol{e}_i = M^\dagger \boldsymbol{e}_i$. □ □

4.6.17 Additional Discussion of Theorem 4.3.1

Theorem 4.3.1 considers budget constraints or penalties of the form $F_i(\boldsymbol{w}_i)$, where $\boldsymbol{w}_i$ is the vector of contracts for agent i. Consider the more general setting of $F_i(\boldsymbol{w}_i^T P e_i)$ or $F_i(\boldsymbol{w}_i \odot P\boldsymbol{e}_i)$. Using the techniques from Sandberg and Willson (1972), we cannot prove the existence and uniqueness of stable points in the general setting, except in trivial cases.

To see this, note that we must impose conditions on the first derivative $f_{ji} = \frac{\partial F_i}{\partial W_{ij}}$ of the penalty function F_i. Specifically, we need $S_{ij} := f_{ij} + f_{ji}$ to be a function of W_{ij} alone. But if F_i were to depend on P, so would f_{ij}. Each entry of P depends on all entries of W in general, not just W_{ij}. Hence, we cannot handle general forms of $F_i(W, P)$.

In the special case where $F_i(W, P) := \boldsymbol{w}_i^T P \boldsymbol{e}_i$, we have $S_{ij} = P_{ij} + P_{ji} = 0$, and Theorem 4.3.1 still applies. However, this case is trivial since it amounts to modifying the payments matrix by a factor of 2:

$$\begin{aligned}\text{agent } i\text{'s utility } g_i(W, P) &:= \boldsymbol{w}_i^T(\boldsymbol{\mu}_i - P\boldsymbol{e}_i) - \gamma_i \cdot \boldsymbol{w}_i^T \Sigma_i \boldsymbol{w}_i - F_i(\boldsymbol{w}_i) \\ &= \boldsymbol{w}_i^T(\boldsymbol{\mu}_i - 2P\boldsymbol{e}_i) - \gamma_i \cdot \boldsymbol{w}_i^T \Sigma_i \boldsymbol{w}_i.\end{aligned}$$

If we instead have a positive penalty only when the total payment is positive (say, $F_i(W, P) := \max(0, \boldsymbol{w}_i^T P \boldsymbol{e}_i)$), the approach no longer works.

4.7 Experimental Details

4.7.1 Fama-French Stock Market Data

We use the Fama-French value-weighted asset returns dataset, for 96 assets over 625 months (Fama and French, 2015).

4.7.2 OECD International Trade Data

We use international trade statistics from the OECD to get quarterly measurements of bilateral trade between 46 large economies, including the top 15 world nations by GDP OECD (2022). The data are available at the OECD Statistics webpage (`https://stats.oecd.org/`). The data are measured quarterly from Q1 2010 to Q2 2022. We take the sum of trade flows $i \to j$ and $j \to i$ to measure the weight of an edge $\{i, j\}$.

To obtain the corresponding Σ, we run our inference procedure (Section 4.2.4). Since there is no data for within-country trade, the network has no self-loops ($W_{ii} = 0$). So we modify the inference according to Remark 4.6.18 in Appendix 4.6.11.

4.7.3 Outlier Detection Simulation

The experiments in Figure 4.6 proceed as follows. Fix a number of communities k and number of firms n. Fix a value of $\sigma > 0$. For us, $k = 2$, $n \in \{20, 100, 300\}$, and $\sigma \in \{\sigma_1, \ldots, \sigma_{10}\}$, where the σ_i are logarithmically spaced on the interval $[0.1, 1]$, so that

$$\begin{aligned}\sigma \in \{&0.1, 0.12915497, 0.16681005,\\ &0.21544347, 0.27825594, 0.35938137,\\ &0.46415888, 0.59948425, 0.77426368, 1.0\}\end{aligned}$$

For a setting of n, k, σ, we perform the following simulation $m = 500$ times.

Generate communities. Generate the community membership matrix $\Theta \in \{0, 1\}^{n \times k}$ with rows independently and uniformly at random from $\{\boldsymbol{e_1}, \ldots, \boldsymbol{e_k}\}$.

Generate the network setting. The deterministic functions f, g, h for M, Σ, Γ respectively are as follows. First $f(\theta_1, \theta_2) = f(\theta_2, \theta_1) = 1$ and $f = 0$ otherwise. Next, let $G \in \mathbb{R}^{k \times k}$ be the matrix $G_{ij} = g(\theta_i, \theta_j)$. Then G is generated from a normalized Wishart distribution centered at I_k and with 5 degrees of freedom. Finally, $h(\theta_i) = 1$ for all i.

The noise variables for agent beliefs are as follows. Sample i.i.d. ϵ_i according to a $N(0, \sigma^2)$ distribution truncated to $[-0.5, 0.5]$ for all i. Sample $\epsilon'_{\theta_i, j} \overset{\text{iid}}{\sim} N(0, \sigma^2)$ for all i, j.

Designate an outlier. Set the the noise parameter $\epsilon_1 = -0.5$ for firm 1 (the risk-seeker), so as $\sigma \to 0$, γ_1 gets further separated from all other γ_i.

Outlier detection simulation. Then for a random firm i such that $\theta_i \neq \theta_1$, we test whether the outlier $\hat{j} := \arg\max_{j:\theta_j=\theta_1} |W_{i,j}|$ is equal to the true outlier firm 1.

Collate results. Once the $m = 500$ runs are completed for a single setting of n, k, σ, we obtain an estimate $\hat{p}$ for the probability of successful deviator detection at this setting of parameters. We plot a confidence interval $[p - 2\sqrt{\frac{\hat{p}(1-\hat{p})}{m}}, p + 2\sqrt{\frac{\hat{p}(1-\hat{p})}{m}}]$. This is plotted on the y-axis. The x-axis quantifies how much γ_1 deviates from the mean, in terms of the number of standard deviations of the truncated normal distribution ϵ_i.

Chapter 5: Strategic Negotiations in Endogenous Network Formation

5.1 Introduction

In computer science and economics, network games (Tardos, 2004; Kearns et al., 2001) model pairwise interactions of agents. Many important processes can be modeled as a network game. For example, rumors and information spread in a social network via interactions between pairs of friends De et al. (2016); Gaitonde et al. (2020b); Chen and Rácz (2021b). Similarly, infectious diseases can spread through a physical contact network via face-to-face interactions Huang and Zhu (2022).

We study network games involving the formation of bilateral contracts between agents such as firms, nations, or individuals. Networks of contracts are widely studied, with applications to supply chain modeling (Acemoglu and Azar, 2020; Elliott et al., 2022b), international trade (Jalan et al., 2024a), and financial contagion (Eisenberg and Noe, 2001; Feinstein et al., 2018). For example, the US over-the-counter (OTC) market for financial derivative products (e.g. credit default swaps) is a network of bespoke bilateral contracts among large firms with a notional value of over 600 trillion USD as of 2022 (ISDA, 2023). Despite their significance, financial networks are not fully understood (Glasserman and Young, 2016).

Network games, both for contract networks and for other settings such as social networks, typically assume that agents play utility-maximizing actions based on the actions of their neighbors. These models assume *honesty*, meaning that agents play according to the pre-specified dynamics of the network game. For example, previous work on contract networks assumes that agents form edges by reporting the true parameters of their utility functions in negotiations (Acemoglu and Azar, 2020; Gaitonde et al., 2020b; Elliott et al., 2022b; Jalan et al., 2024a). However, recent works show that agents can manipulate network games by acting *strategically*. For example, agents in a social network can deviate from ordinary opinion dynamics to spread misinformation (Gaitonde et al., 2020b; Chen and Rácz,

The content of this chapter is under review at the 26th ACM Conference on Economics and Computation (ACM EC 2025), and can be cited as Jalan and Chakrabarti (2024).

2021b). More generally, strategic agents can manipulate the dynamics of a known game for various purposes (Galeotti et al., 2020; Kolumbus and Nisan, 2022a; Kolumbus et al., 2024), such as maximizing utility, gaining information, or deceiving others.

All of these works assume that only one actor is strategic. Interactions between several strategic agents are not explored. We present, to our knowledge, the first results for a multi-agent network formation game with an *arbitrary set of strategic agents*.

In our model, n heterogeneous agents negotiate to form a network of bilateral contracts. Each agent wants contracts to maximize their utility (Eq. (5.1)), which follows the classical mean-variance utility model for a portfolio of contracts (Markowitz, 1952). But contract details must be agreed to by both parties, so they negotiate during network formation. An arbitrary subset S of the n agents negotiates *strategically*, and the rest are honest. Specifically, each strategic agent chooes a negotiating strategy privately, before negotiations. Then, during network formation they must play according to this strategy, and cannot adjust in response to others. For this game among the strategic actors, we ask the following question:

(Q1) How should each agent negotiate optimally in a network?

At first sight, it is not obvious that an optimal strategy exists, especially when there are several strategic agents. For instance, consider two hedge funds competing against each other to get a contract with an investor. Each fund wants to offer more favorable terms than the other, but must pick its position before seeing the other fund's choice. This uncertainty makes the problem even more difficult.

Next, even for a single strategic agent, optimal negotiations requires some knowledge of other agents' preferences. This leads to our second question:

(Q2) How can an agent learn the parameters needed for optimal negotiations?

If the network is changing in time (e.g. due to agents entering and leaving), then before entering the network agents may observe an equilibrium from previous strategic negotiations. But, they cannot directly infer other agents' beliefs since the network was not formed truthfully. Moreover, because i and j account for variance in their utility, a contract between (i, j) can depend on depend on j's contract with k, which depends on k's contract with ℓ, and so on. This makes learning from previously observed edges difficult.

Our main contributions are as follows.

Efficient algorithm to find Nash equilibria. Our main result is an efficient algorithm (Algorithm 4) to find the optimal negotiating positions for an arbitrary set of strategic agents, or report when no optimal solution exists (Theorem 5.3.7). We show that all Nash equilibria are pure, and give extensions of Algorithm 4 to the case where agents are uncertain about the preferences of their neighbors.

Learning algorithm for agents. Given the edges formed from previous strategic negotiations, we present an algorithm (Algorithm 5) to learn the other agents' true beliefs and the set S of strategic agents. Our algorithm is robust to strategic agents playing non-Nash-optimal strategies to fool the learner.

Analysis of international trade networks. We simulate Nash-optimal strategic negotiations on real-world international trade data OECD (2022). Our experiments confirm that the utilities of agents, as well as the "Price of Strategy" (analogous to the Price of Anarchy), are sensitive to the set of strategic agents. We also show that our learning algorithm recovers the parameters needed for strategic negotiations for a broad range of networks.

5.2 Background and Related Work

Network games are widely studied in both economics (Glasserman and Young, 2016; Elliott and Golub, 2022) and computer science (Leng et al., 2020a; Rossi et al., 2022). Our work closely relates to endogenous network formation models for economic networks (Acemoglu and Azar, 2020; Sadler and Golub, 2021; Jalan et al., 2024a). Note that the contracts in our model can include but are not limited to principal-agent contracts, which are studied in their own right (Papireddygari and Waggoner, 2022; Alon et al., 2023). As in these works on network formation, we study stable points of a network formation process in which each agent wants to form edges to maximize its utility. But, our work focuses on the effects of strategic manipulation of the network formation process, and learning algorithms that observe the outcomes of this manipulated process.

Prior work on steering the outcomes of network games (Galeotti et al., 2020; Gaitonde et al., 2020b; Chen and Rácz, 2021b; Wang and Kleinberg, 2024) and on learning from observations of network games (Irfan and Ortiz, 2014; Garg and Jaakkola, 2016; De et al.,

2016; Leng et al., 2020a; Rossi et al., 2022) largely consider games with one-dimensional action spaces and one strategic actor. Instead, we focus on consider arbitrary sets of strategic actors. Moreover, each of the n agents in our model has an n-dimensional action space. This results in a vector of contracts $\boldsymbol{w}_k \in \mathbb{R}^n$ for each agent k. The correlations between entries of $\boldsymbol{w}_k$ can have strategic consequences, as we show in Section 5.4. Some recent works study multiple strategic actors in repeated auctions (Kolumbus and Nisan, 2022b) and Fisher markets with linear utilities Kolumbus et al. (2024). Our work has a similar thrust, but we focus on network formation.

The effect of strategic behavior is often studied through the Price of Anarchy (Roughgarden, 2005, 2015; Christodoulou et al., 2017; Gkatzelis et al., 2022), which measures the welfare under a strategic equilibrium versus welfare under a socially optimal equilibrium. We similarly give a worst-case upper bound on the "Price of Strategy" (Proposition 5.4.2), which is the analogue of the Price of Anarchy in our setting. However, in addition to overall welfare, we also examine how the set of strategic agents affects individual welfare, and find that the behavior is complex. For example, there are strategic sets $S \subseteq [n]$ in which *all* members of S are worse off than if they had all negotiated honestly. This sensitivity to S is similar in spirit to Christodoulou and Sgouritsa (2019), although there is no central designer in our setting (see Remark 5.3.10).

Moreover, the the data used in our learning algorithms are strategically manipulated by agents, resulting in a "strategic source." Recently, there has been a growing interest in developing learning algorithms for such sources. Chen et al. (2020a) study strategy-awareness for linear classification, but assume that agents can only misreport data up to an ϵ-ball. Our setting is closer to that of Ghalme et al. (2021), who show that agents who are evaluated by a third-party classifier (e.g. for approval for a bank loan) can strategically modify their features to game the classifier, even if the classifier used is strategy-robust in the sense of Hardt et al. (2016). Harris et al. (2023) give a $\tilde{O}(n^{(d+1)/(d+2)})$-regret algorithm for online binary classification against n strategic agents with d-dimensional features, but in a linear reward model. Finally, Cai et al. (2015) give a mechanism to encourage strategic data providers to report truthful data, but in a model where the data providers have no incentive to hurt the classifier's accuracy (e.g. crowdsourcing).

Next, we discuss notation and present a background on network formation without

strategic agents.

Background: network formation without strategy. We use a network model with side-payments between agents Jackson and Wolinsky (2003) and mean-variance utility, which is a widely used model of risk-aware utility Harrison and Qin (2009); Li et al. (2014); Simaan (2014); Zhang et al. (2021); Ma et al. (2023b). This network model has been shown to provide closed-form solutions for truthful network formation Jalan et al. (2024a). We summarize this model below.

Let $W = W^T \in \mathbb{R}^{n\times n}$ denote an undirected weighted network of contracts between n agents, with W_{ij} being the size of the contract between i and j and W_{ii} representing self-investment. A negative contract $W_{ij} < 0$ is valid and represents a reversed version of a positive contract; for example, in a derivative contract, $W_{ij} < 0$ swaps the roles of the long and short position holders. During contract negotiations, agent i can pay P_{ji} per unit contract to agent j to get j to agree to the contract size. Since payments are zero-sum, $P^T = -P$. The contracts size and payments (W, P) together give the network. At (W, P), agent i has contracts $\boldsymbol{w}_i := W\boldsymbol{e}_i$. Agent i wants to optimize the utility of their contracts and believes that contracts have mean return $\boldsymbol{\mu}_i \in \mathbb{R}^n$ and covariance $\Sigma \succ 0$. Moreover, they have a risk-aversion parameter $\gamma_i > 0$. Their utility is then:

$$\text{agent } i\text{'s utility } g_i(W, P) := \boldsymbol{w}_i^T(\boldsymbol{\mu}_i - P\boldsymbol{e}_i) - \gamma_i \cdot \boldsymbol{w}_i^T\Sigma\boldsymbol{w}_i. \tag{5.1}$$

Note that beliefs do not have to be accurate or follow a particular distribution.

Definition 5.2.1 (Stable point)**.** *A feasible (W, P) is stable if each agent achieves its maximum possible utility given prices P:*

$$g_i(W, P) = \max_{(W',P):W'=W'^T,P^T=-P} g_i(W', P) \quad \forall i \in [n].$$

Stable points for truthful network formation are as follows.

Theorem 5.2.2 (Stable network without strategy Jalan et al. (2024a))**.** *Let M be such that $M\boldsymbol{e}_i = \boldsymbol{\mu}_i$. Let Γ be a diagonal matrix with $\Gamma_{ii} = \gamma_i$. Note that $\Gamma \succ 0$ and $\Sigma \succ 0$. There exists a unique stable point (W, P):*

$$\begin{aligned}\text{vec}(W) &= \frac{1}{2}(\Gamma \otimes \Sigma + \Sigma \otimes \Gamma)^{-1}\text{vec}(M + M^T),\\ \text{vec}(P) &= (\Gamma^{-1} \otimes \Sigma^{-1} + \Sigma^{-1} \otimes \Gamma^{-1})^{-1} \cdot \text{vec}(\Sigma^{-1}M\Gamma^{-1} - \Gamma^{-1}M^T\Sigma^{-1}).\end{aligned}$$

Furthermore, agents can efficiently find the stable point through honest pairwise negotiations.

To illustrate the network model, we consider the example of trade networks. We analyze real-world trade networks in Section 5.6.

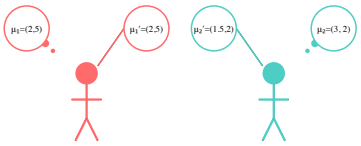
Example 5.2.3 (Trade Networks Without Strategy)**.** *A set of n nations forms bilateral trades with contracts W, payments P, and $(\boldsymbol{\mu}_i)_{i\in[n]}, \Sigma$ are beliefs regarding the mean and covariances of the contract returns as in Theorem 5.2.2.*

- *For $i < j$, the pair $\{i, j\}$ trade a fixed good depending on $\{i, j\}$ (e.g. food, energy, manufacturing equipment).*
- *If $W_{ij} > 0$ then i is the seller, otherwise j is the seller.*
- *Each contract $\{i, j\}$ has a base price. For example, nation 1 sells wheat to nation 2 at a base price of \$10 per bushel, and they agree to $W_{12} = 200$ bushels.*
- *The quantity P_{ij} is a negotiated adjustment to the base price (e.g. nation 1 gives nation 2 a discount of $P_{21} = 3$ and charges them \$7 per bushel of wheat).*
- *The quantities $\boldsymbol{\mu}_{i;j}, \boldsymbol{\mu}_{j;i}$ represent the perceptions that agents i, j have about the expected return per unit of contract (e.g. based on the base price, the demand for wheat in each country, etc.).*
- *Finally, Σ_{11} is the perceived risk of trading with nation 1 (due to e.g. wheat price volatility, political instability in agent 1's country, etc). If agent 3 sells a complementary good for wheat (e.g. sugar), then agent 2 might perceive $\Sigma_{13} > 0$, because the value of wheat would positively correlate with that of sugar. Nations who trade with they trade with both 1 and 3 account for this correlation in Eq.* (5.1)*.*

5.3 Strategic Negotiations

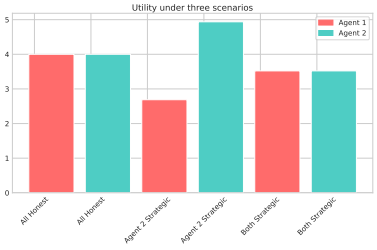
We now formalize the contract negotiation process. Figure 5.1 illustrates a toy example.

Definition 5.3.1 (Our Model of Strategic Contract Negotiation (M, Γ, Σ, S))**.** *There is a set $S \subseteq [n]$ of strategic agents who know the (M, Γ, Σ) defined in Theorem 5.2.2. An honest agent $i \notin S$ only knows $(\boldsymbol{\mu}_i, \gamma_i, \Sigma)$. The contract negotiation is a two-stage process:*



(a) Only agent 2 (right stick figure) is strategic, so $\boldsymbol{\mu}_1' = \boldsymbol{\mu}_1$ and $\boldsymbol{\mu}_2' \neq \boldsymbol{\mu}_2$.

(b) Both agents are strategic, so $\boldsymbol{\mu}_1' \neq \boldsymbol{\mu}_1$ and $\boldsymbol{\mu}_2' \neq \boldsymbol{\mu}_2$.



(c) Utilities.

Figure 5.1: Toy illustration of our model (Definition 5.3.1) for a network with $n = 2$ agents and 1 edge, with $\Gamma = I$ and $\Sigma_{12} = \Sigma_{21} = \frac{1}{2}, \Sigma_{11} = \Sigma_{22} = 1$. We may have only one strategic agent (left), or multiple (middle). Which subset of agents is strategic affects utility for everyone (right). The case of *All Honest* corresponds to prior work on truthful network formation (Jalan et al., 2024a).

1. **Strategy Phase**: *Each strategic agent $k \in S$ independently and privately chooses a negotiating position $\boldsymbol{\mu}_k' \in \mathbb{R}^n$. For honest agents $k \notin S$, $\boldsymbol{\mu}_k' = \boldsymbol{\mu}_k$. Let M' be a matrix whose i^{th} column is $\boldsymbol{\mu}_i'$.*

2. **Contract Formation Phase**: *The network is formed as if every agent's negotiating position was their true belief. Specifically, (W, P) is formed according to Theorem 5.2.2 with (M', Σ, Γ).*

The network (W, P), and the true beliefs (M, Γ, Σ) determine each agent's utility (Eq. (5.1)).

The above definition assumes that strategic agents know the true beliefs (M, Γ, Σ) (Definition 5.3.1). Our approach generalizes to the case where agents have a distribution over M, and each agent aims to maximize its expected utility. For ease of exposition, we focus on the fixed M setting here, with the general setting deferred to the Appendix. Also, we do not consider strategic choices for the risk aversion γ_i and covariance matrix Σ. The former is typically similar for all agents Paravisini et al. (2017), while the latter is often known from public sources such as credit rating agencies White (2010).

We first prove a general result that characterizes an agent's utility given *arbitrary* negotiating positions for all other agents. It also shows that no agent can gain unbounded utility by being strategic.

Theorem 5.3.2 (Concave utility given others' choices)**.** *Given M, Γ, Σ and any set of negotiating positions $\{\boldsymbol{\mu}_i'; i \neq k\}$ for all agents except k, agent k's utility is a quadratic*

Algorithm 4 Nash Equilibria Computation

Input: (M, Γ, Σ) as in Definition 5.3.1, strategic agent set $S \subseteq [n]$.
Output: Nash equilibria set $\mathcal{E}$.
$T, K, L, \{\boldsymbol{y}_k\} \leftarrow$ as in Definition 5.3.6.
$T_S \leftarrow$ submatrix of T with blocks $\{T^{(i,j)} : i, j \in S\}$
$\boldsymbol{y}_S \leftarrow$ concatenation of $\{\boldsymbol{y}_k : k \in S\}$
$\mathcal{F} \leftarrow \{X \in \mathbb{R}^{n \times |S|} : T_S \mathrm{vec}(X) = \boldsymbol{y}_S\}$
$\mathcal{E} \leftarrow \left\{ \Delta \in \mathbb{R}^{n \times n} : \begin{Bmatrix} \Delta|_S = X, \\ \Delta|_{[n] \setminus S} = 0 \end{Bmatrix}, X \in \mathcal{F} \right\}$
return $\mathcal{E}$ *if* $|\mathcal{E}| > 0$ *else "No Nash Equilibrium"*

function of $\boldsymbol{\mu}_k'$ *with a negative definite Hessian.*

Theorem 5.3.2 immediately implies the following.

Corollary 5.3.3 (Strategy yields bounded utility)**.** *No choice of negotiating position lets agent k achieve unbounded utility, even if agent k has full information about other agents' beliefs and choices.*

Next, we turn to the case of multiple strategic agents. In strategic contract negotiations, all agents make strategic choices independently and cannot adapt their strategy to the others' choices ex post. They can choose a strategy ex ante based on a Nash equilibrium, defined below.

Definition 5.3.4. *A Nash Equilibrium for a Strategic Contract Negotiation (M, Γ, Σ, S) is a matrix $\Delta \in \mathbb{R}^{n \times n}$ with the following property. For each $k \in S$, if all other strategic agents $j \in S$ choose negotiating position $(M + \Delta)\boldsymbol{e}_j$, then agent k gains the highest utility by choosing negotiating position $(M + \Delta)\boldsymbol{e}_k$ in the Strategy Phase.*

Notice that $\Delta = M' - M$, so for a fixed M the negotiating positions are determined by the columns of Δ.

We allow the matrix Δ to be random, corresponding to mixed strategies. However, we will see that even if others play mixed strategies, the optimal choice for an agent is to play a pure strategy (Theorem 5.3.7).

To give an explicit solution for optimal negotiating positions, we require the following definitions.

Definition 5.3.5 (Commutator Matrix)**.** *We define the commutator matrix* $\Pi : \mathbb{R}^{n^2} \to \mathbb{R}^{n^2}$ *and the projection matrices* $\Pi_k : \mathbb{R}^{n^2} \to \mathbb{R}^n$ *such that* $\Pi \text{vec}(X) = \text{vec}(X^T)$ *and* $\Pi_k \text{vec}(X) = X\boldsymbol{e}_k$ *for all* $X \in \mathbb{R}^{n \times n}$*. For any* $Z \in \mathbb{R}^{n^2} \to \mathbb{R}^{n^2}$*, we define* $Z^{(p,q)}$ *as the* $n \times n$ *block at position* (p, q)*.*

We also require the following definitions.

Definition 5.3.6. *For* (M, Γ, Σ) *as in Definition 5.3.1, we define the following:*

$$
\begin{aligned}
K &= (\Gamma \otimes \Sigma + \Sigma \otimes \Gamma), \\
L &= \frac{1}{2}(K^{-1} + K^{-1}\Pi), \\
T^{(k,j)} &= \begin{cases} L^{(k,k)} + (L^{(k,k)})^T - 2\gamma_k (L^{(k,k)})^T \Sigma L^{(k,k)} & k = j \\ (I - 2\gamma_k (L^{(k,k)})^T \Sigma) L^{(k,j)} & k \neq j \end{cases} \\
\boldsymbol{y}_k &= \frac{1}{2}(2\gamma_k (L^{(k,k)})^T \Sigma - I)\Pi_k K^{-1} \text{vec}(M + M^T).
\end{aligned}
$$

Note that K^{-1} exists because $\Sigma, \Gamma \succ 0$.

We can now fully characterize the optimal negotiating position of an agent.

Theorem 5.3.7. *Suppose a strategic agent* k *knows* $S \subseteq [n]$ *and* $M \in \mathbb{R}^{n \times n}$*. The negotiating position* $\boldsymbol{\mu}'_k$ *(or, equivalently, the* $\boldsymbol{\delta}_k$*) that optimizes* k*'s utility is given by the solution(s) to the following linear system, if any exist.*

$$
T^{(k,k)}\boldsymbol{\delta}_k + \sum_{j \in S : j \neq k} T^{(k,j)}\boldsymbol{\delta}_j = \boldsymbol{y}_k \tag{5.2}
$$

Thus, the optimal negotiating position of an agent $k \in S$ is a fixed $\boldsymbol{\delta}^*_k \in \mathbb{R}^n$ that solves a deterministic linear system. In a Nash equilibrium, every strategic agent solves their corresponding equation. Algorithm 4 explicitly describes how a strategic agent can solve for the equilibrium Δ.

Corollary 5.3.8 (Nash Equilibria)**.** *The Nash equilibria correspond to solving the system of* $n|S|$ *linear equations in the fixed vectors* $\{\boldsymbol{\delta}_i \mid i \in S\}$ *given by taking Eq. 5.6 for each* $k \in S$*.*

Corollary 5.3.9 (All Equilibria are Pure)**.** *All Nash equilibria are pure-strategy Nash equilibria.*

Finally, we comment briefly on mechanism design.

Remark 5.3.10 (VCG Mechanism)**.** *It is natural to ask whether a mechanism designer can mitigate the effects of strategic behavior. While incentive-compatible mechanisms such as the Vickrey-Clarke-Groves (VCG) mechanism are known for bilateral trade, which is a special case of our model, the VCG mechanism requires a subsidy if the buyer values the good more than the seller (Nisan et al., 2007). In real-world settings such as international trade networks (Section 5.6), it is not clear who would provide this subsidy. If we consider a mechanism with neither subsidies nor taxes, the Gibbard-Satterthwaite theorem (Gibbard, 1973; Satterthwaite, 1975) forbids the existence of non-trivial decision rules in dominant strategies for many settings (Jackson, 2000). In general, there is tension between the twin goals of (i) a dominant strategy incentive-compatible mechanism, and (ii) subsidies and taxes summing to zero (Jackson, 2000).*

In summary, we see that strategic agents can negotiate optimally and find Nash equilibria. This motivates two questions that we will address in turn. First, what are the effects of strategic negotiations on both overall utility, and individual utilities (Section 5.4)? Second, strategic agents need to know the matrix M and the set of other strategic agents S. Can agents learn these from observing the network (Section 5.5)?

5.4 Winners and Losers with Multiple Strategic Agents

The motivation for negotiating strategically, rather than honestly, is that an agent might achieve better terms for their contracts and hence more utility. However, if multiple strategic agents are present, their conflicting goals may result in an overall worse equilibrium. In this section, we study how strategic negotiations affect agents' utilities. We want to understand how both the *overall* welfare (the sum of all utilities) changes, as well as how *individual* welfare changes.

To study the overall welfare, we introduce the Price of Strategy[1], which is analogous to the Price of Anarchy.

[1]Not to be confused with the Price of Stability.

Definition 5.4.1 (Price of Strategy)**.** *For fixed* (M, Γ, Σ, S)*, let* (W', P') *be the equilibrium of strategic negotiations and* (W, P) *be the equilibrium under honest negotiations. Then the Price of Strategy is:*

$$\text{PoS} := \frac{\sum_{i=1}^{n} g_i(W, P)}{\sum_{i=1}^{n} g_i(W', P')}$$

Notice that the numerator of the Price of Strategy measures welfare at the equilibrium without strategy ($\Delta = 0$). This is unlike the Price of Anarchy, which measures the *socially optimal* equilibrium (for a possibly nonzero Δ^*). In our setting, agents do not cooperate and do not care about other agents' welfare, so it is more appropriate to study the Price of Strategy. Note that PoS $\leq$ PoA.

We now present an upper bound, which shows that the Price of Strategy is $O(1)$ when the norm of the strategically chosen deviation Δ can be bounded. When $\|\Delta\|_F \gg \|M\|_F$, the bound no longer holds, and the Price of Strategy may be unbounded. The proof is deferred to the Appendix.

Proposition 5.4.2 (Price of Strategy with All Agents Strategic)**.** *Let* $(M, I, \Sigma, [n])$ *be as in Definition 5.3.1. Suppose* Σ *has least eigenvalue* $\lambda_n > 0$*. Let* L, T *be as in Definition 5.3.6, and* $C \in \mathbb{R}^{n^2 \times n^2}$ *be the matrix (implicit in Definition 5.3.6) such that* $\text{vec}(\Delta) = CL\text{vec}(M)$*. If* $\|C\|_2 < c\lambda_n$ *for a constant* $c > 0$*, then:*

$$\text{PoS} \leq O(1).$$

Proposition 5.4.2 gives a *global* guarantee on how welfare changes due to strategic negotiations. We can also ask more fine-grained questions. In particular, we ask:

1. Who "pays the price" for strategy? In particular, if some actors are honest, are only the honest actors worse off, or can strategic actors be worse off as well?

2. How does the PoS change when only a subset of actors S are strategic?

Who pays the price of strategy? If $|S| = 1$, it is clear from Theorem 5.3.2 that the lone strategic agent does not pay the price for strategy. They can do no worse by negotiating strategically. However, when $|S| > 1$, there are three possibilities.

1. All strategic agents in S are better off than if they had all negotiated honestly.
2. Some members of S are worse off.
3. All members of S are worse off.

We will show that all three possibilities can occur, even in a simple network with $n = 3$ agents. Outcome (3) is especially interesting, as every strategic agent would be better off if all of them were honest. However, our model does not allow agents to coordinate. Hence, they are stuck in a lose-lose Nash equilibrium, akin to the Prisoner's Dilemma.

How does the PoS change as S grows? One might expect that as more agents become strategic, the PoS may grow because there are more strategic manipulations occurring. However, we show that in the same example network, the PoS is *not* mononotonic in $|S|$. As more agents become strategic, they counterbalance against the negotiations of other strategic agents, "taking back" some of the utility that they lost when they were honest.

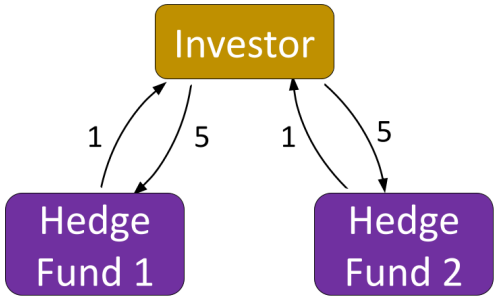
We now introduce the following example network. More details and other examples are presented in the Appendix.

Example 5.4.3 (Two Hedge Funds, One Investor (Figure 5.2))**.** *For $m, a \in \mathbb{R}$ and $\rho \in (-1, 1)$, define:*

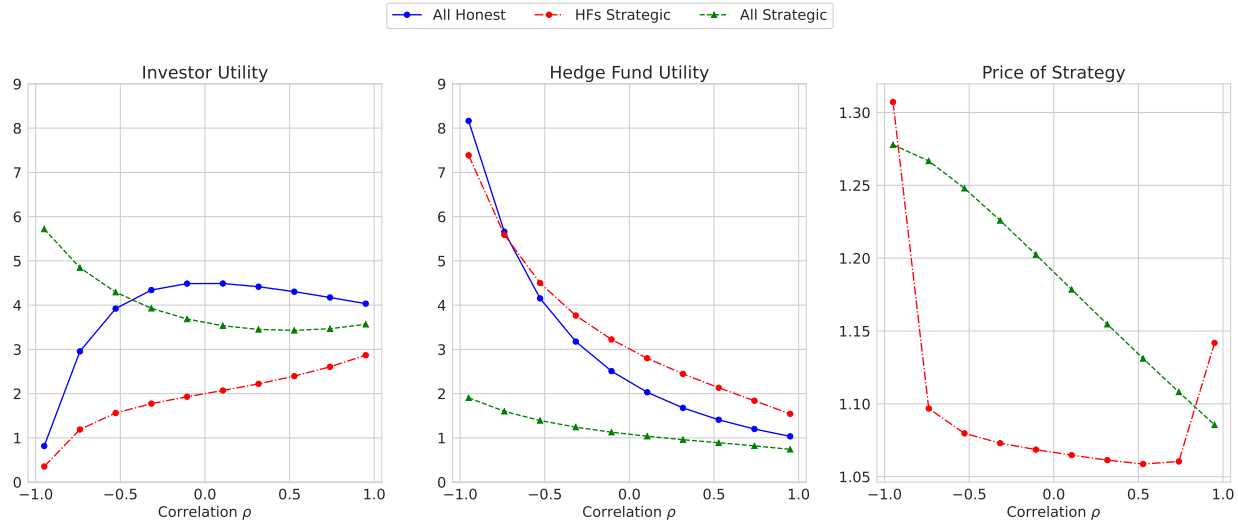
$$M = \begin{bmatrix} 0 & a & a \\ m & 0 & 0 \\ m & 0 & 0 \end{bmatrix}, \quad \Sigma = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & \rho \\ 0 & \rho & 1 \end{bmatrix}, \quad \Gamma = I. \tag{5.3}$$

The first column corresponds to the investor, and the others to the hedge funds. Under this setting, the hedge funds do not want to trade with each other, and none of the agents want to self-invest. Also, the hedge funds are correlated with each other (via ρ), and uncorrelated with the investor.

We will consider two cases: (a) the investor is honest and the hedge funds are strategic ($S = \{2, 3\}$), and (b) all agents are strategic ($S = \{1, 2, 3\}$).



(a) Network diagram with entries of entries of M marked on edges.



(b) Utilities of the Investor and a single Fund under Proposition 5.9.3. Note that the two Funds achieve the same utility.

Figure 5.2: *Network with three agents:* (a) An investor trades with two hedge funds, with the investor gaining 5 per unit contract while the hedge funds gain 1. (b) We show the utility for the investor (left) and either hedge fund (right) for various strategic behaviors. The investor has low utility when she is honest, but is better off than the hedge funds when she is strategic.

From Proposition 5.9.3 we can compute the utilities at equilibrium for any particular choice of M and Σ. Figure 5.2b shows the utilities for a specific M and varying ρ. Note that $\kappa = \frac{1+|\rho|}{1-|\rho|}$, so our bound on the Price of Strategy diverges as $\rho \to -1$ or $\rho \to 1$.

We observe the following.

When only the hedge funds are strategic, they can be both better off or both worse off (Outcomes 1 and 3). The specifics depend on the perceived correlation ρ. As $\rho \approx -1$, the hedge funds are worse off being strategic than if they were both honest (Figure 5.2b). This is due to the hedging behavior of the investor.

As $\rho \to -1$, the investor wishes to invest almost equally in both funds to reduce her overall risk. But the hedge funds only form one contract each. Since they cannot hedge their risk, they prefer much smaller contracts than the investor. If both funds are honest, they can negotiate contract sizes to match their risk preference. However, if both are strategic, each fund worries about its competitor. So, both funds end up taking on more risk than they would prefer, and are worse off.

On the other hand, for $\rho \gg -1$, the investor will not seek such large contracts, since she cannot hedge as well. So the hedge funds are both better off being strategic.

When all agents are strategic ($S = [n]$), the investor can be better off while the funds are both worse off (Outcome 2). When $S = [n]$, there is no setting in which all agents are better off. However, the investor is better off as $\rho \to -1$. As before, the funds are worse off because they are forced to take large contracts. When the investor is also strategic, she can force the funds to compete for her investment and obtain better terms from both. She will obtain large contracts with both, which enable low risk due to hedging and have better terms than if she was honest.

5.5 Learning from Strategic Negotiation Outcomes

Algorithm 4 describes how a strategic agent should choose its Nash-optimal negotiating position. However, to run the algorithm, the agent must know

1. the matrix $H := (M + M^T)$ of the beliefs of all agents, and

2. the set $S \subseteq [n]$ of strategic agents.

Suppose an agent (called the *learner*) lacks this information but can observe the entire network W'. The learner wishes to learn M and S for use in future negotiations. From W', the learner can recover the negotiating positions $M' + M'^T$ via Theorem 5.2.2. But she cannot infer the true beliefs $M + M^T$.

We therefore consider a model in which the learner also knows extra information in the form of a feature matrix X.

Definition 5.5.1 (Network Setting with Agent Features)**.** *Let $X \in \mathbb{R}^{n\times d}$ have rows $\boldsymbol{x}_i \in \mathbb{R}^d$ for agent $i \in [n]$, and $B \in \mathbb{R}^{d\times d}$. A network setting with agent features (B, Γ, Σ, X) is such that, for $i, j \in [n]$*

$$M_{ij} = \boldsymbol{x}_i^T B \boldsymbol{x}_j$$

for all $i, j \in [n]$. Hence, $\mathrm{vec}(M) = (X \otimes X)\mathrm{vec}(B)$.

For simplicity, we focus on the case of $d \ll n$ and $\Gamma = I$. The latter corresponds (up to a constant) to homogeneous risk aversions, which are commonly observed Ang (2014); Paravisini et al. (2017).

Given W' and X, the learner wishes to learn B. Now, W' depends on the negotiating positions M', which can differ from M in an unknown way. For instance, some strategic agents may occasionally deviate from the Nash strategy to throw off the learner.

We therefore formulate our learning problem as a robust regression.

Definition 5.5.2 (Robust Regression)**.** *Given features X and a stable network W' arising from negotiating positions M', the robust regression problem with covariates $(X \otimes X)$, response $\boldsymbol{y} \in \mathbb{R}^{n^2}$, and corruption threshold $\beta \in [0, 1]$ is to solve:*

$$\min_{\hat{B}\in\mathbb{R}^{d\times d}} \|(X \otimes X)\mathrm{vec}(\hat{B} + \hat{B}^T) - \boldsymbol{y}\|_2^2,$$

assuming that $\boldsymbol{y}$ differs from $\mathrm{vec}(M + M^T)$ *arbitrarily at up to βn^2 entries. We write* $\mathrm{RR}(X \otimes X, \boldsymbol{y}, \beta)$ *for shorthand.*

Note that we do not require that M' corresponds to a Nash equilibrium. We assume an adaptive adversary as opposed to the simpler oblivious adversary setting. Hence, the

Algorithm 5 Estimation of mean beliefs B and strategic agents S

Input: Network $W' \in \mathbb{R}^{n\times n}$, agent features $X \in \mathbb{R}^{n\times d}$, corruption threshold $\beta \in (0,1)$
Output: Estimates $\hat{B} \in \mathbb{R}^{d\times d}$ and $\hat{S} \subseteq [n]$.
$\text{vec}(H') \leftarrow K\text{vec}(W')$
$\text{vec}(\hat{B}) \leftarrow$ solve $\text{RR}(X \otimes X, \text{vec}(H'), \beta)$
$R_{ij} \leftarrow |\boldsymbol{e}_i^T(H' - \frac{1}{2}X(\hat{B} + \hat{B}^T)X^T)\boldsymbol{e}_j|$
$A_{ij} = \begin{cases} 1 & \text{if } R_{ij} \text{ belongs to top } \beta n^2 \text{ entries of } R \\ 0 & \text{otherwise} \end{cases}$
$S_1, S_2 \leftarrow$ spectral clustering on A with 2 clusters
$\hat{S} \leftarrow S_i$ such that $|S_i|$ is nearest to $\frac{\sqrt{8\beta n}+1}{4}$
return $\hat{B}, \hat{S}$.

learner must learn in the face of complex counter-strategies by other agents who have full knowledge of the network setting and features.

We use Torrent Bhatia et al. (2015) to solve the robust regression problem. This algorithm has provable guarantees even when an $\Omega(1)$ fraction of the response vector $\boldsymbol{y}$ is adversarially corrupted. However, we emphasize that the choice of robust regression algorithm is independent of our method, and alternative algorithms can also be used.

Our Algorithm 5 first learns a $\hat{B}$ using Torrent. Then, it computes a matrix of residuals R and constructs an unweighted graph $G = ([n], E)$ such that $(i,j) \in E$ iff R_{ij} is one of the βn^2 largest residuals. If $\hat{B} \approx B$, then these edges should all be incident to the strategic nodes $S \subset [n]$. Therefore, a consistent clustering algorithm such as Rohe et al. (2011) can recover S.

Next, we discuss the recovery guarantee for $\hat{B}$. We defer the theoretical guarantee for $\hat{S}$ to Section 5.8.6. We recount the following technical condition of Bhatia et al. (2015). For a matrix $X \in \mathbb{R}^{n\times d}$ with n samples in $\mathbb{R}^d$ and $S \subset [n]$ let $X_S \in \mathbb{R}^{|S|\times d}$ select rows in S.

Definition 5.5.3 (SSC and SSS Conditions)**.** *Let $\gamma \in (0,1)$. A design matrix $X \in \mathbb{R}^{n\times d}$ satisfies the Subset Strong Convexity Property at level $1-\gamma$ and Subset Strong Smoothness Property at level γ with constants $\lambda_{1-\gamma}, \Lambda_\gamma$ respectively if:*

$$\lambda_{1-\gamma} \leq \min_{S\subset[n]:|S|=(1-\gamma)n} \lambda_{min}(X_S^T X_S)$$
$$\Lambda_\gamma \geq \max_{S\subset[n]:|S|=\gamma n} \lambda_{max}(X_S^T X_S)$$

We will give a concrete example of the SSC and SSS constants for a feature matrix X in Example 5.5.4.

Example 5.5.4 (Balanced Stochastic Block Model)**.** *Suppose that $X \in \{0,1\}^{n\times 2}$ describes community memberships for a Stochastic Block Model with equally sized communities. Then $(X \otimes X)^T(X \otimes X) = \frac{n^2}{4} I_4$. It can be shown that for any $\gamma \in (0,1)$ that the corresponding constants are $\lambda_{1-\gamma} = n^2(\frac{1}{4} - \gamma)$ and $\Lambda_\gamma = \gamma n^2$. Therefore the condition of Proposition 6.4.2 holds for $\gamma < \frac{1}{68}$. A sufficient condition is $|S| \leq \frac{n}{136}$.*

Proposition 5.5.5. *Suppose $S \subset [n]$ is the strategic set, and $M' \in \mathbb{R}^{n\times n}$ is the matrix of negotiating positions that results in a stable network W'. Let $\beta \geq \frac{2n|S|-|S|^2}{n^2}$. Suppose $X \otimes X$ satisfies the SSC condition at level $1-\beta$ with constant $\lambda_{1-\beta}$, and SSS condition at level β with constant Λ_β (Definition 5.5.3). Then, there exists a constant $C > 0$ such that if $|S| \leq Cn$ and $4\frac{\sqrt{\Lambda_\alpha}}{\sqrt{\lambda_{1-\alpha}}} < 1$, Algorithm 5 with threshold parameter β and T iterations of* Torrent *returns $\hat{B}$ such that:*

$$\|\hat{B} + \hat{B}^T - (B + B^T)\|_F \leq \exp(-cT) \cdot \left(\frac{\|M' + (M')^T - (M + M^T)\|_F}{n} \right)$$

for an absolute constant $c > 0$, and n large enough.

Remark 5.5.6 (Random design)**.** *The SSC and SSS conditions are known to hold for a sub-Gaussian design Bhatia et al. (2015). Our Proposition 5.5.5 concerns arbitrary fixed design, but can be easily extended to the random design setting with similar techniques.*

5.6 Experiments

We show experiments for learning the network parameters on a simulated dataset, and then test the effects of strategic negotiations on the OECD international trade network.

5.6.1 Learning Experiments

We validate our learning approach on networks where agent i has d-dimensional features $\boldsymbol{x}_i$ sampled independently from Dirichlet$(1/d, 1/d, \ldots, 1/d)$, and M has a bilinear form with a random symmetric matrix $B \in \mathbb{R}^{d\times d}$ with upper triangular entries $N(5,1)$ (Definition 5.5.1). Then, we sample $S \subseteq [n]$ uniformly from all subsets of a certain size, and

compute the stable network W' with Algorithm 4. All experiments use $n = 100$ agents and $\beta := \frac{2|S|n-|S|^2}{n^2}$. See the Appendix for full experimental details.

Figures 5.3 and 5.4 show the accuracy of the recovered matrix $\hat{B}$ and the strategic subset of agents $\hat{S}$, respectively. We find that Algorithm 5 performs well for a broad range of parameter settings. Unsurprisingly, it is best for small d and $|S|$. We find that the regression error is low even with a large number of strategic actors (Figure 5.3). This suggests that the condition of $|S| \leq Cn$ in Proposition 5.5.5, which is required to handle the worst-case S, may be relaxed if we are willing to accept an average-case guarantee.

Figure 5.3: Normalized regression error for $\hat{B}$ estimation, with shaded regions denoting $[10, 90]$-percentile outcomes across 10 independent trials. Left: $|S| = 0.1 \times n$. Right: $d = 2$.

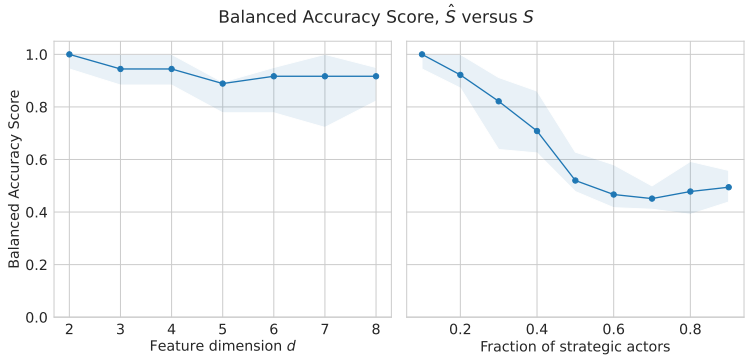


Figure 5.4: Balanced accuracy (the mean of the true positive rate and true negative rate) of $\hat{S}$ estimation, in the same settings as Figure 5.3. Shaded regions denote $[10, 90]$-percentile outcomes across 10 independent trials.

5.6.2 Negotiations on International Trade Networks

In this section, we analyze strategic negotiations on a dataset of international trade among $n = 46$ large economies OECD (2022) across $T = 49$ time periods.

Nodes represent nations, and edge W_{ij}^t at time t is the total recorded trade between i and j during a fiscal quarter. Following Jalan et al. (2024a), we infer the $(M^t, \Sigma^t, \Gamma^t)$ (Definition 5.3.1) from the networks W^t over the period 2010-2020 (see Appendix for full details).

Trade networks arise from complex strategic considerations Carlson and Dacey (2013). We perform a counterfactual analysis by comparing actual trade volumes from 2010-2020 (the *Observed Network*) to counterfactual networks that would have arisen from strategic behavior. Counterfactual analysis is a common experimental tool in economics (Chudik et al., 2021; Arca et al., 2023; Bogetoft et al., 2024), and has been used in the network games literature to study manipulation of opinion dynamics (Chen and Rácz, 2021b).

In our counterfactual analysis, we model what would have happened if, in addition to their usual strategies, certain countries used Algorithm 4 in their trade negotiations. Figure 5.5 shows whether agents gain or lose utility under strategic negotiations, for different choices of S. Each row corresponds to a a time period, and each column to a country. Cell (t, i) is positive (red) if agent i is better off at time t under the given choice of S than they would have been if all agents had negotiated honestly. We note the following.

(1) Strategic behavior does not always help. When all countries are strategic (left), agents can be better or worse off depending on the time step. Interestingly, the smallest countries by trade volume are usually better off under both scenarios. This is because they deal in such small volumes (a factor of 10^3 difference from the largest countries) that offering them good deals does not hurt much. Moreover, the structure of the covariance matrix Σ matters. Small countries may trade in unique goods, offer sources of uncorrelated returns and therefore incentivizing larger countries to deal with them.

(2) Sensitivity to choice of S. The outcomes for certain countries are notably different under the scenario where $S = \{\text{US}, \text{UK}\}$ as opposed to $S = [n]$. In particular, the column corresponding to the UK (6th largest degree) shows that they are usually worse off when all are strategic (left), but are better off when only the US and UK are strategic (right). Conversely, the US is typically better off when all are strategic (left) and worse off when only the US and UK are strategic. So, despite being much larger than the UK, the US cannot

sway outcomes as much in the case of $S = \{\text{US}, \text{UK}\}$. Moreover, as in Example 5.4.3, there are time periods where all members of S are worse off (right).

(3) Honest actors can gain due to the strategic choices of others. When only the US and UK are strategic, various honest actors are *better off* under the strategic equilibrium than the honest equilibrium (right). Despite not intending to, the US and UK may inadvertently offer better terms to some countries (such as the smallest economies) as a result of their competition with each other.

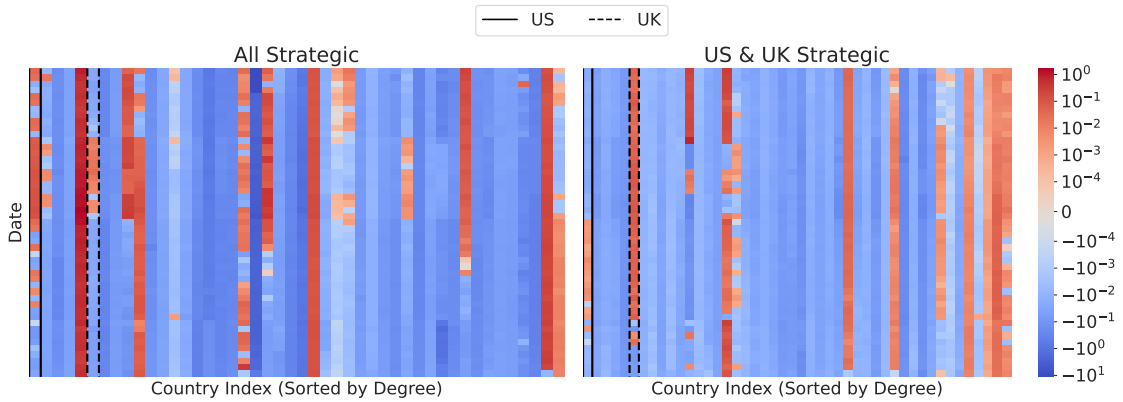


Figure 5.5: The effect of strategic negotiations on each country's utility at each timestep, for $S = [n]$ (left), and $S = \{\text{United States}, \text{United Kingdom}\}$ (right). In each heatmap, if the cell (t, i) is positive (red), then agent i gains at time t, while if it negative (blue), then agent i loses at time t. Specifically, for a fixed S, let $g_i^t(S)$ be the utility of country i and time t when S is the set of strategic agents. Note that $g_i^t(\emptyset)$ is the utility when all countries negotiate honestly. Then each cell (t, i) displays $\frac{g_i^t(S) - g_i^t(\emptyset)}{g_i^t(\emptyset)}$ with respect to the particular choice of S.

Next, Figure 5.6 displays the Price of Strategy for three scenarios: the worst-case choice of S when $|S| = 1$ (left), the worst choice of S among five random choices with $|S| = 5$ (the middle), and when $S = [n]$ (right). Surprisingly, the Price of Strategy is much higher for $|S| = 5$ than when $S = [n]$ or $|S| = 1$, showing that the Price of Strategy is *not* monotonic in $|S|$. We note that the Price of Strategy is $O(1)$ for all time periods with $S = [n]$.

5.7 Conclusions and Future Work

In this paper, we propose a model of network formation with multiple strategic actors. Strategic agents manipulate the network formation process by using negotiating positions different from their true preferences. We give an efficient algorithm to find the set of all Nash

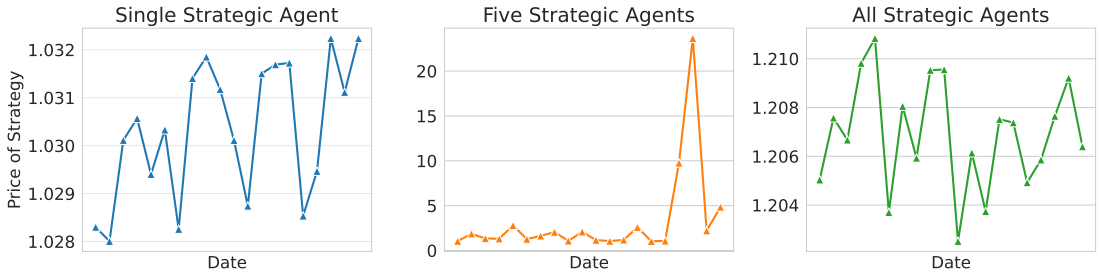


Figure 5.6: Price of strategy for the first 20 time periods, with multiple choices of $|S|$. For $|S| = 1$ and $|S| = n$ we can compute the PoS exactly. For $|S| = 5$ we generate 5 choices of S uniformly at random and report the maximum. For the worst S among all $\binom{n}{5}$ choices, the PoS could be even higher.

equilibria, and show that they are all pure-strategy equilibria. The resulting equilibrium can result in a loss of utility compared to honest negotiations, even for agents who are strategic. This is in contrast to the single strategic agent setting where the strategic agent can do no worse than if she was honest. When all actors are strategic, we show that the Price of Strategy can be bounded. However, this is not conclusive, because we also show that the Price of Strategy can be higher when some agents are honest. Next, we show that agents can learn the true preferences of others from historical network data, even if the others had negotiated strategically or sought to fool the learner. Finally, experimental results on real-world and simulated networks validate our approach.

Future work could include different noise models for learning, such as partially adaptive Mukhoty et al. (2023) or oblivious noise d'Orsi et al. (2021). Further, one could extend our strategic negotiations model to a repeated games setting. To our knowledge, optimal negotiating positions in repeated games against learning agents have only been studied in non-network settings Deng et al. (2019); Assos et al. (2024). Finally, in light of the results of Section 5.4, it would be interesting to fully determine when strategic actors are better or worse off than if they had all negotiated honestly.

5.8 Proofs and Additional Theoretical Results

In this section we give proofs for the results of the paper, and also include additional theoretical results (Section 5.8.4 and Section 5.8.6) discussed in the paper.

5.8.1 Proof of Theorem 5.3.2

The proof of Theorem 5.3.2 is through Lyapunov analysis. We break up the proof into a series of Propositions. First, we characterize how strategic negotiation affects the contracts of a strategic actor $k \in [n]$.

Throughout this section, let $i \in [n]$ let $\boldsymbol{\delta}_i := \boldsymbol{\mu}_i' - \boldsymbol{\mu}_i$. Choosing the optimal $\boldsymbol{\mu}_i'$ is equivalent to $\boldsymbol{\delta}_i$, so we give results in terms of $\boldsymbol{\delta}_i$.

Proposition 5.8.1. *Let $k \in [n]$. Let $\boldsymbol{\delta}_{-k} := \{\boldsymbol{\delta}_i : i \neq k\}$ and let (W, P) be the stable point if k reports honestly and all others report according to $\boldsymbol{\delta}_{-k}$. Next, consider some $\boldsymbol{\delta}_k \neq \mathbf{0}$ and let (W', P') be the stable point if k reports $\boldsymbol{\delta}_k$ and all others report according to $\boldsymbol{\delta}_{-k}$. Then $W'\boldsymbol{e}_k = W\boldsymbol{e}_k + B\boldsymbol{\delta}_k$ for a matrix B defined as follows. Let $\Gamma^{-1/2}\Sigma\Gamma^{-1/2} = V\Lambda V^T$ be the eigendecomposition of $\Gamma^{-1/2}\Sigma\Gamma^{-1/2} \succ 0$. Let $A \in \mathbb{R}^{n\times n}$ be a symmetric matrix such that:*

$$A_{ij} = \begin{cases} \frac{V_{ki}^2}{4\lambda_i} + \sum_{\ell=1}^{n} \frac{V_{k\ell}^2}{2(\lambda_i+\lambda_\ell)} & i = j \\ \frac{V_{ki}V_{kj}}{2(\lambda_i+\lambda_j)} & i \neq j \end{cases}$$

Then $B = \gamma_k^{-1}\Gamma^{-1/2}VAV^T\Gamma^{-1/2}$.

Proof. Let $\Delta_W = W' - W$. By Theorem 5.2.2, $2(\Sigma\Delta_W\Gamma + \Gamma\Delta_W\Sigma) = \boldsymbol{e}_k\boldsymbol{\delta}_k^T + \boldsymbol{\delta}_k\boldsymbol{e}_k^T$. Therefore $\mathrm{vec}(\Delta_W) = \frac{1}{2}(\Sigma\otimes\Gamma + \Gamma\otimes\Sigma)^{-1}(\boldsymbol{e}_k\boldsymbol{\delta}_k^T + \boldsymbol{\delta}_k\boldsymbol{e}_k^T)$.

Next, let $\boldsymbol{v}_i := V\boldsymbol{e}_i$. Using the eigendecomposition properties of Kronecker sums Horn and Johnson (2008) as in Corollary 1 of Jalan et al. (2024a),

$$\Gamma^{1/2}\Delta_W\Gamma^{1/2} = \left(\sum_{i=1}^{n}\sum_{j=1}^{n} \frac{\boldsymbol{v}_i^T\Gamma^{-1/2}\big(\boldsymbol{e}_k\boldsymbol{\delta}_k^T\big)\Gamma^{-1/2}\boldsymbol{v}_j}{2(\lambda_i+\lambda_j)}\boldsymbol{v}_i\boldsymbol{v}_j^T\right) + \left(\sum_{i=1}^{n}\sum_{j=1}^{n} \frac{\boldsymbol{v}_i^T\Gamma^{-1/2}\big(\boldsymbol{\delta}_k\boldsymbol{e}_k^T\big)\Gamma^{-1/2}\boldsymbol{v}_j}{2(\lambda_i+\lambda_j)}\boldsymbol{v}_i\boldsymbol{v}_j^T\right)$$

Let $G := \Gamma^{-1/2}V$ and $\boldsymbol{g}_i := G\boldsymbol{e}_i$. Then $\boldsymbol{v}_i^T\Gamma^{-1/2}\boldsymbol{e}_k = G_{ki}$ and $\boldsymbol{e}_k^T\Gamma^{-1/2}\boldsymbol{v}_j = G_{kj}$. Hence:

$$\begin{aligned}
\Delta_W\boldsymbol{e}_k &= \Gamma^{-1/2}\left(\sum_{i=1}^{n}\sum_{j=1}^{n} \frac{G_{ki}\boldsymbol{g}_j^T\boldsymbol{\delta}_k + G_{kj}\boldsymbol{g}_i^T\boldsymbol{\delta}_k}{2(\lambda_i+\lambda_j)}\boldsymbol{v}_i\boldsymbol{v}_j^T\right)\Gamma^{-1/2}\boldsymbol{e}_k \\
&= \Gamma^{-1/2}\left(\sum_{i=1}^{n}\sum_{j=1}^{n} \frac{G_{ki}\boldsymbol{g}_j^T\boldsymbol{\delta}_k + G_{kj}\boldsymbol{g}_i^T\boldsymbol{\delta}_k}{2(\lambda_i+\lambda_j)}G_{kj}\boldsymbol{v}_i\right) \\
&= \Gamma^{-1/2}\left(\sum_{i=1}^{n}\sum_{j=1}^{n} \frac{G_{ki}G_{kj}}{2(\lambda_i+\lambda_j)}\boldsymbol{v}_i\boldsymbol{v}_j^T + \sum_{j=1}^{n} G_{kj}^2\cdot\left(\sum_{i=1}^{n}\left(\frac{1}{2(\lambda_i+\lambda_j)}\boldsymbol{v}_i\boldsymbol{v}_i^T\right)\right)\right)\Gamma^{-1/2}\boldsymbol{\delta}_k
\end{aligned}$$

Hence $W'\boldsymbol{e}_k - W\boldsymbol{e}_k = B\boldsymbol{\delta}_k$ for a matrix B defined as above. We can further simplify B as

$$B = \Gamma^{-1/2}\Big(\sum_{i=1}^{n}\Big(\frac{G_{ki}^2}{4\lambda_i} + \sum_{j=1}^{n}\frac{G_{kj}^2}{2(\lambda_i+\lambda_j)}\Big)\boldsymbol{v}_i\boldsymbol{v}_i^T + \sum_{i=1}^{n}\sum_{j\neq i}\frac{G_{ki}G_{kj}}{2(\lambda_i+\lambda_j)}\boldsymbol{v}_i\boldsymbol{v}_j^T\Big)\Gamma^{-1/2}$$

Notice that $G_{ki} = \gamma_k^{-1/2}V_{ki}$, and that B depends only on the k^{th} row of G so we can factor out γ_k^{-1} and replace the entries G_{ki} with V_{ki}.

Finally, let $A \in \mathbb{R}^{n\times n}$ be defined as in the statement of this Proposition. Then:

$$\begin{aligned} B &= \gamma_k^{-1}\Gamma^{-1/2}\big(\sum_{i=1}^{n}A_{ii}\boldsymbol{v}_i\boldsymbol{v}_i^T + \sum_{i=1}^{n}\sum_{j\neq i}A_{ij}\boldsymbol{v}_i\boldsymbol{v}_j^T\big)\Gamma^{-1/2} \\ &= \gamma_k^{-1}\Gamma^{-1/2}VAV^T\Gamma^{-1/2}. \end{aligned}$$

The conclusion follows. □

The next Proposition gives the core idea, which is to prove that $\Lambda^{1/2}A\Lambda^{1/2}$ is a contraction via a Lyapunov argument.

Proposition 5.8.2. *Let $F = \Lambda^{1/2}A\Lambda^{1/2}$. The eigenvalues of F are all real and contained in $(0,1)$.*

Proof. Notice that F is symmetric, since Λ is diagonal and A is symmetric. By the Spectral Theorem, F has an eigendecompsition with real eigenvalues.

Next, notice $A = C + D$ for $C_{ij} = C_{ji} = \frac{V_{ki}V_{kj}}{2(\lambda_i+\lambda_j)}$ and D a diagonal matrix with $D_{ii} = \sum_{\ell=1}^{n}\frac{V_{k\ell}^2}{2(\lambda_i+\lambda_\ell)}$. Then $F = \tilde{C} + \tilde{D}$ for $\tilde{C} = \Lambda^{1/2}C\Lambda^{1/2}$ and $\tilde{D} = \Lambda^{1/2}D\Lambda^{1/2}$.

Let $\boldsymbol{x} = \frac{1}{\sqrt{2}}\Lambda^{1/2}V^T\boldsymbol{e}_k$. Then $\tilde{C}$ satisfies the Lyapunov equation:

$$\Lambda\tilde{C} + \tilde{C}\Lambda = \boldsymbol{x}\boldsymbol{x}^T$$

Since $\tilde{C}$ is self-adjoint it has an eigenbasis with real eigenvalues. Let $\boldsymbol{y}$ be an eigenvector of $\tilde{C}$ with eigenvalue μ. Then $(\boldsymbol{y}^T\boldsymbol{x})^2 = \boldsymbol{y}^T(\tilde{C}\Lambda + \Lambda\tilde{C})\boldsymbol{y} = 2\mu\boldsymbol{y}^T\Lambda\boldsymbol{y}$. By the Cauchy-Schwarz

inequality,

$$\begin{aligned}
\mu &= \frac{(\boldsymbol{y}^T\boldsymbol{x})^2}{2\boldsymbol{y}^T\Lambda\boldsymbol{y}} \\
&= \frac{1}{4}\frac{(\sum_{i=1}^{n}\sqrt{\lambda_i}y_iV_{ki})^2}{\sum_{i=1}^{n}\lambda_iy_i^2} \\
&\leq \frac{1}{4}\frac{(\sum_{i=1}^{n}\lambda_iy_i^2)\sum_{i=1}^{n}V_{ki}^2)}{\sum_{i=1}^{n}\lambda_iy_i^2} \\
&\leq \frac{1}{4}
\end{aligned}$$

Further, since $\Lambda \succ 0$, $\mu \geq 0$. So the eigenvalues of $\tilde{C}$ are all within $[0, \frac{1}{4}]$. Next, the eigenvalues of $\tilde{D}$ are simply its diagonal entries. Recall that

$$\begin{aligned}
\tilde{D}_{ii} &= \sum_j \frac{\lambda_iV_{kj}^2}{2(\lambda_i+\lambda_j)} \\
&> \frac{\lambda_i}{2(\lambda_i+\max_j\lambda_j)}\sum_j V_{kj}^2 \\
&> 0,
\end{aligned}$$

since V is orthonormal and $\lambda_j > 0$. By similar reasoning, $\tilde{D}_{ii} < \frac{1}{2}$. We conclude that the eigenvalues of F are contained in $(0, \frac{3}{4})$. □

We are ready to prove Theorem 5.3.2.

Proof of Theorem 5.3.2. Fix the index k of the strategic actor. Suppose k reports $\boldsymbol{\delta}_k$ and each $i \neq k$ reports some $\boldsymbol{\delta}_i$. Let (W', P') be the resulting stable point.

By Lemma 5.8.3, the utility of k at (W', P') is $-\langle\boldsymbol{\delta}_k, \boldsymbol{w}_k'\rangle + \gamma_k\langle\boldsymbol{w}_k', \Sigma\boldsymbol{w}_k'\rangle$ where $\boldsymbol{w}_k'$ is the k^{th} column of W'.

By Proposition 5.8.1, $\boldsymbol{w}_k' = \boldsymbol{w}_k + B\boldsymbol{\delta}_k$ for $B = \gamma_k^{-1}\Gamma^{-1/2}VAV^T\Gamma^{-1/2}$ with V and A as in Proposition 5.8.1. Hence the utility of k is quadratic in $\boldsymbol{\delta}_k$, and the quadratic term is $-\boldsymbol{\delta}_k^T\big(B - \gamma_kB\Sigma B\big)\boldsymbol{\delta}_k$. A straightforward calculation gives:

$$B - \gamma_kB\Sigma B = \gamma_k^{-1}\Gamma^{-1/2}V\Lambda^{-1/2}\Big(\Lambda^{1/2}A\Lambda^{1/2} - \big(\Lambda^{1/2}A\Lambda^{1/2}\big)^2\Big)\Lambda^{-1/2}V^T\Gamma^{-1/2}$$

Let $F := \Lambda^{1/2} A \Lambda^{1/2}$. To show that the Hessian of the utility of k is negative definite in $\boldsymbol{\delta}_k$, we need to show $F - F^2 \succ 0$. Since the spectrum of F is contained in $(0, 1)$ by Proposition 5.8.2, $F \succ F^2$ and the conclusion follows. □

5.8.2 Proof of Theorem 5.3.7

To prove Theorem 5.3.7, we first state an easy Lemma.

Lemma 5.8.3 (Utility from Strategy)**.** *If agent k reports $\boldsymbol{\delta}'_k$ resulting in (W', P'), then k's utility is*

$$g_k(W', P') = -\langle \boldsymbol{\delta}_k, \boldsymbol{w}'_k \rangle + \gamma_k \langle \Sigma \boldsymbol{w}'_k, \boldsymbol{w}'_k \rangle,$$

where $\boldsymbol{w}'_k = W' \boldsymbol{e}_k$.

We now prove Theorem 5.3.7.

Proof of Theorem 5.3.7. We begin by proving that agent k's optimal negotiating position $\boldsymbol{\delta}^\star_k$ given $\boldsymbol{\delta}^*_{j \in S: j \neq k}$ is the solution to the linear system:

$$(2\gamma_k (L^{(k,k)})^T \Sigma - I)\boldsymbol{v}_k = T^{(k,k)} \boldsymbol{\delta}^\star_k + \sum_{j \in S: j \neq k} T^{(k,j)} \boldsymbol{\delta}^\star_j \tag{5.4}$$

where $T^{(p,q)}$ and L are defined as in Algorithm 4.

Let $\Delta_M \in \mathbb{R}^{n \times n}$ have i^{th} column $\boldsymbol{\delta}^\star_i$ if $i \in S$ and zero otherwise. Let (W', P') be the stable point resulting from a choice of $M' := M + \Delta_M$ as the agents' negotiating positions in the Strategy Phase. From Theorem 5.2.2, we have

$$\begin{aligned} \mathrm{vec}(W') &= \mathrm{vec}(W) \\ &\quad + 0.5(\Sigma \otimes \Gamma + \Gamma \otimes \Sigma)^{-1} \mathrm{vec}(\Delta_M + \Delta_M^T) \\ \Rightarrow \mathrm{vec}(W' - W) &= L \mathrm{vec}(\Delta_M) \\ \Rightarrow \boldsymbol{w}'_k - \boldsymbol{w}_k &= L^{(k,k)} \boldsymbol{\delta}_k + \sum_{j \in S: j \neq k} L^{(k,j)} \boldsymbol{\delta}_j, \end{aligned}$$

where the second line follows from the commutation property of Π, and the matrix L is defined as in Algorithm 4.

Now, fix an agent $k \in S$. They want to choose $\boldsymbol{w}'_k$ optimally based on the above equation, but are uncertain about the value of $\boldsymbol{w}_k$.

Let $A_k := L^{(k,k)}$ and $\boldsymbol{b}_k := \sum_{j \in S: j \neq k} L^{(k,j)} \boldsymbol{\delta}_j$. By Lemma 5.8.3, we have:

$$\begin{aligned}
g_k &= -\langle \boldsymbol{\delta}_k, \boldsymbol{w}'_k \rangle + \gamma_k \langle \boldsymbol{w}'_k, \Sigma \boldsymbol{w}'_k \rangle \\
&= -\langle \boldsymbol{\delta}_k, (\boldsymbol{w}_k + A_k \boldsymbol{\delta}_k + \boldsymbol{b}_k) \rangle + \gamma_k \langle (\boldsymbol{w}_k + A_k \boldsymbol{\delta}_k + \boldsymbol{b}_k), \Sigma (\boldsymbol{w}_k + A_k \boldsymbol{\delta}_k + \boldsymbol{b}_k) \rangle \\
&= -\langle \boldsymbol{\delta}_k, A_k \boldsymbol{\delta}_k \rangle - \langle \boldsymbol{\delta}_k, \boldsymbol{b}_k \rangle - \langle \boldsymbol{\delta}_k, \boldsymbol{w}_k \rangle + 2\gamma_k \langle \Sigma \boldsymbol{w}_k, A_k \boldsymbol{\delta}_k + \boldsymbol{b}_k \rangle \\
&+ \gamma_k \langle A_k \boldsymbol{\delta}_k + \boldsymbol{b}_k, \Sigma (A_k \boldsymbol{\delta}_k + \boldsymbol{b}_k) \rangle + \gamma_k \langle \boldsymbol{w}_k, \Sigma \boldsymbol{w}_k \rangle
\end{aligned}$$

Agent k wants to optimize g_k by choosing $\boldsymbol{\delta}_k$. By Theorem 5.3.2, the optimal negotiating position $\boldsymbol{\delta}^\star$ is the critical point of g_k with respect to $\boldsymbol{\delta}_k$. Notice that the Hessian of g_k with respect to $\boldsymbol{\delta}_k$ does not depend on $\boldsymbol{\delta}_j$ for any $j \neq k$, so the critical point gives the optimal negotiating position for g_k. Setting $\nabla_{\delta_k} g_k = \mathbf{0}$, we obtain:

$$(A_k + A_k^T - 2\gamma_k A_k^T \Sigma A_k) \boldsymbol{\delta}_k^\star = (2\gamma_k A_k^T \Sigma - I) \boldsymbol{w}_k$$

Rearranging terms, we obtain the linear system:

$$\begin{aligned}
(2\gamma_k A_k^T \Sigma - I) \boldsymbol{w}_k = \Bigg[& (L^{(k,k)} + (L^{(k,k)})^T - 2\gamma_k (L^{(k,k)})^T \Sigma L^{(k,k)}) \boldsymbol{\delta}_k^\star \\
& + \sum_{j \in S: j \neq k} L^{(k,j)} \boldsymbol{\delta}_j^* - 2\gamma_k (L^{(k,k)})^T \Sigma \sum_{j \in S: j \neq k} L^{(k,j)} \boldsymbol{\delta}_j^* \Bigg] \\
\Rightarrow \boldsymbol{y}^{(k)} = T^{(k,k)} \boldsymbol{\delta}_k^\star & + \sum_{j \in S: j \neq k} T^{(k,j)} \boldsymbol{\delta}_j^* \qquad (5.5)
\end{aligned}$$

Where $T^{(k,k)} = (L^{(k,k)} + (L^{(k,k)})^T - 2\gamma_k (L^{(k,k)})^T \Sigma L^{(k,k)})$ and $T^{(k,j)} = (I - 2\gamma_k (L^{(k,k)})^T \Sigma) L^{(k,j)}$ for $j \neq k$. □

5.8.3 Generalization of Theorem 5.3.7 to Distribution on Negoiating Positions

We require the following Lemmata.

Lemma 5.8.4 (Seber and Lee (2012))**.** *Let $\boldsymbol{x} \in \mathbb{R}^n$ be a random vector and $A \in \mathbb{R}^{n \times n}$ a symmetric matrix. Then, if $\mathbb{E}[\boldsymbol{x}] = \boldsymbol{\mu}$ and $\boldsymbol{x}$ has covariance Σ, then:*

$$\mathbb{E}[\boldsymbol{x}^T A \boldsymbol{x}] = tr(A\Sigma) + \boldsymbol{\mu}^T A \boldsymbol{\mu}$$

Next, we need the standard property of the commutation matrix.

Lemma 5.8.5 (Horn and Johnson (2008))**.** *There exists a permutation matrix* $\Pi : \mathbb{R}^{n^2} \to \mathbb{R}^{n^2}$ *such that for* $X \in \mathbb{R}^{n \times n}$, $\Pi \text{vec}(X) = \text{vec}(X^T)$. *We call* Π *the commutation matrix.*

We are ready to prove a generalization of Theorem 5.3.7. The following statement generalizes to the case where agents have distributions over the negotiating positions of other strategic agents.

Theorem 5.8.6 (Generalization of Theorem 5.3.7.)**.** *Suppose a strategic agent* $k \in [n]$ *knows* S*, and has a distribution* $\mathcal{D}_k$ *over the negotiation positions* $\{\boldsymbol{\mu}_i' : i \in S \setminus \{k\}\}$ *with finite first and second moments. Define* $\boldsymbol{\delta}_i := \boldsymbol{\mu}_i' - \boldsymbol{\mu}_i$*. The negotiating position* $\boldsymbol{\mu}_k'$ *(or, equivalently, the* $\boldsymbol{\delta}_k$*) that optimizes* k*'s expected utility with respect to* $\mathcal{D}_k$ *is given by the solution(s) to the following linear system, if any exist.*

$$T^{(k,k)}\boldsymbol{\delta}_k + \sum_{j \in S: j \neq k} T^{(k,j)} \mathop{\mathbb{E}}_{\mathcal{D}_k} \boldsymbol{\delta}_j = \boldsymbol{y}_k \tag{5.6}$$

Proof of Theorem 5.8.6. We begin by proving that agent k's optimal negotiating position $\boldsymbol{\delta}_k^\star$ given $\boldsymbol{\delta}^*_{j \in S: j \neq k}$ is the solution to the linear system:

$$(2\gamma_k (L^{(k,k)})^T \Sigma - I)\boldsymbol{v}_k = T^{(k,k)}\boldsymbol{\delta}_k^\star + \sum_{j \in S: j \neq k} T^{(k,j)}\boldsymbol{\delta}_j^\star \tag{5.7}$$

where $T^{(p,q)}$ and L are defined as in Algorithm 4.

Let $\Delta_M \in \mathbb{R}^{n \times n}$ have i^{th} column $\boldsymbol{\delta}_i^\star$ if $i \in S$ and zero otherwise. Let (W', P') be the stable point resulting from a choice of $M' := M + \Delta_M$ as the agents' negotiating positions in the Strategy Phase. From Theorem 5.2.2, we have

$$\begin{aligned} \text{vec}(W') &= \text{vec}(W) \\ &\quad + 0.5(\Sigma \otimes \Gamma + \Gamma \otimes \Sigma)^{-1}\text{vec}(\Delta_M + \Delta_M^T) \\ \Rightarrow \text{vec}(W' - W) &= L\text{vec}(\Delta_M) \\ \Rightarrow \boldsymbol{w}_k' - \boldsymbol{w}_k &= L^{(k,k)}\boldsymbol{\delta}_k + \sum_{j \in S: j \neq k} L^{(k,j)}\boldsymbol{\delta}_j, \end{aligned}$$

where the second line follows from Lemma 5.8.5, and the matrix L is defined as in Algorithm 4. Notice that there is a distribution on $\boldsymbol{w}_k$ induced by $\mathcal{D}_k$.

Now, fix an agent $k \in S$. They want to choose $\boldsymbol{w}'_k$ optimally based on the above equation, but are uncertain about the value of $\boldsymbol{w}_k$.

Let $A_k := L^{(k,k)}$ and $\boldsymbol{b}_k := \sum_{j \in S: j \neq k} L^{(k,j)} \boldsymbol{\delta}_j$. By Lemma 5.8.3, we have:

$$\begin{aligned}
g_k &= -\langle \boldsymbol{\delta}_k, \boldsymbol{w}'_k \rangle + \gamma_k \langle \boldsymbol{w}'_k, \Sigma \boldsymbol{w}'_k \rangle \\
&= -\langle \boldsymbol{\delta}_k, (\boldsymbol{w}_k + A_k \boldsymbol{\delta}_k + \boldsymbol{b}_k) \rangle \\
&+ \gamma_k \langle (\boldsymbol{w}_k + A_k \boldsymbol{\delta}_k + \boldsymbol{b}_k), \Sigma (\boldsymbol{w}_k + A_k \boldsymbol{\delta}_k + \boldsymbol{b}_k) \rangle \\
&= -\langle \boldsymbol{\delta}_k, A_k \boldsymbol{\delta}_k \rangle - \langle \boldsymbol{\delta}_k, \boldsymbol{b}_k \rangle - \langle \boldsymbol{\delta}_k, \boldsymbol{w}_k \rangle + 2\gamma_k \langle \Sigma \boldsymbol{w}_k, A_k \boldsymbol{\delta}_k + \boldsymbol{b}_k \rangle \\
&+ \gamma_k \langle A_k \boldsymbol{\delta}_k + \boldsymbol{b}_k, \Sigma (A_k \boldsymbol{\delta}_k + \boldsymbol{b}_k) \rangle + \gamma_k \langle \boldsymbol{w}_k, \Sigma \boldsymbol{w}_k \rangle
\end{aligned}$$

Agent k wants to optimize $\mathbb{E}_{\boldsymbol{b}_k}[g_k]$ by choosing $\boldsymbol{\delta}_k$. Notice $\boldsymbol{b}_k$ is a linear function of the vectors $\boldsymbol{\delta}_j$. Let $\boldsymbol{\mu} = \mathbb{E}[\boldsymbol{b}_k]$ and $Q = \mathbb{E}[(\boldsymbol{b}_k - \boldsymbol{\mu})(\boldsymbol{b}_k - \boldsymbol{\mu})^T]$.

Therefore agent $k \in S$ wants to optimize:

$$\begin{aligned}
\mathop{\mathbb{E}}_{\boldsymbol{b}_k}[g_k] &= \mathop{\mathbb{E}}_{\boldsymbol{b}_k} \Big[- \langle \boldsymbol{\delta}_k, A_k \boldsymbol{\delta}_k \rangle - \langle \boldsymbol{\delta}_k, \boldsymbol{b}_k \rangle - \langle \boldsymbol{\delta}_k, \boldsymbol{w}_k \rangle + 2\gamma_k \langle \Sigma \boldsymbol{w}_k, A_k \boldsymbol{\delta}_k + \boldsymbol{b}_k \rangle \\
&+ \gamma_k \langle A_k \boldsymbol{\delta}_k + \boldsymbol{b}_k, \Sigma (A_k \boldsymbol{\delta}_k + \boldsymbol{b}_k) \rangle + \gamma_k \langle \boldsymbol{w}_k, \Sigma \boldsymbol{w}_k \rangle \Big] \\
&= -\langle \boldsymbol{\delta}_k, A_k \boldsymbol{\delta}_k \rangle - \langle \boldsymbol{\delta}_k, \boldsymbol{\mu} \rangle - \langle \boldsymbol{\delta}_k, \boldsymbol{w}_k \rangle + 2\gamma_k \langle \Sigma \boldsymbol{w}_k, A_k \boldsymbol{\delta}_k \rangle + 2\gamma_k \langle \Sigma \boldsymbol{w}_k, \boldsymbol{\mu} \rangle + \gamma_k \langle \boldsymbol{w}_k, \Sigma \boldsymbol{w}_k \rangle \\
&+ \gamma_k \langle \Sigma A_k \boldsymbol{\delta}_k, A_k \boldsymbol{\delta}_k \rangle + 2\gamma_k \langle \boldsymbol{\mu}, \Sigma A_k \boldsymbol{\delta}_k \rangle + \gamma_k \mathrm{tr}(\Sigma Q) + \gamma_k \langle \boldsymbol{\mu}, \boldsymbol{\mu} \rangle
\end{aligned}$$

Where the last step is by Lemma 5.8.4.

Next, by Theorem 5.3.2, the optimal negotiating position $\boldsymbol{\delta}^\star$ is the critical point of g_k with respect to $\boldsymbol{\delta}_k$. Notice that the Hessian of $\mathbb{E}[g_k]$ with respect to $\boldsymbol{\delta}_k$ does not depend on $\boldsymbol{\delta}_j$ for any $j \neq k$, so the critical point gives the optimal negotiating position for $\mathbb{E}[g_k]$. Setting $\nabla_{\delta_k} \mathbb{E}_{\boldsymbol{b}_k}[g_k] = 0$, we obtain:

$$(A_k + A_k^T - 2\gamma_k A_k^T \Sigma A_k) \boldsymbol{\delta}_k^\star = (2\gamma_k A_k^T \Sigma - I) \boldsymbol{w}_k + (2\gamma_k A_k^T \Sigma - I) \boldsymbol{\mu}$$

Notice that the gradients of the quadratic terms $\mathrm{tr}(\Sigma R_k)$ and $\gamma_k^T \boldsymbol{v}_k^T \Sigma \boldsymbol{v}_k$ with respect to $\boldsymbol{\delta}_k$ are zero.

Rearranging terms, we obtain the linear system:

$$(2\gamma_k(L^{(k,k)})^T\Sigma - I)\boldsymbol{v}_k = \Big[(L^{(k,k)} + (L^{(k,k)})^T - 2\gamma_k(L^{(k,k)})^T\Sigma L^{(k,k)})\boldsymbol{\delta}_k^\star + \sum_{j\in S:j\neq k} L^{(k,j)}\,\mathbb{E}[\boldsymbol{\delta}_j] \quad -2\gamma_k(L^{(k,k)})^T\Sigma \Big]_j$$

$$\Rightarrow \boldsymbol{z}^{(k)} = T^{(k,k)}\boldsymbol{\delta}_k^\star + \sum_{j\in S:j\neq k} T^{(k,j)}\,\mathbb{E}[\boldsymbol{\delta}_j] \tag{5.8}$$

Where $T^{(k,k)} = (L^{(k,k)} + (L^{(k,k)})^T - 2\gamma_k(L^{(k,k)})^T\Sigma L^{(k,k)})$ and $T^{(k,j)} = (I - 2\gamma_k(L^{(k,k)})^T\Sigma)L^{(k,j)}$ for $j \neq k$.

□

5.8.4 Generalization of Theorem 5.3.7 to Stochastic M

In this section we will describe optimal negotiating positions in the setting where each strategic agent does not know the true matrix $M \in \mathbb{R}^{n\times n}$ but instead has a probability distribution for it. The proof is similar to that of Theorem 5.3.7.

Theorem 5.8.7. *Suppose agent $i \in [n]$ believes $M \in \mathbb{R}^{n\times n}$ follows $M \sim \mathcal{D}_i$, and seeks to maximize its expected utility $\mathbb{E}_{\mathcal{D}_i}[g_i]$. We assume that all distributions $\mathcal{D}_i$ have finite first and second moments. Let $V_i \in \mathbb{R}^{n\times n}$ be such that $\mathrm{vec}(V_i) = \frac{1}{2}K^{-1}(\mathbb{E}_{\mathcal{D}_i}[M] + \mathbb{E}_{\mathcal{D}_i}[M]^T)$, where K is defined in Algorithm 4. Let $\boldsymbol{v}_i = V_i\boldsymbol{e}_i$. Suppose each strategic agent $k \in S$ knows $\{\boldsymbol{v}_j \mid j \in S\}$ (they can compute it from the network setting $(\{\mathcal{D}_i\}, \Gamma, \Sigma)$) which is known to all strategic agents). Modify the linear system of Algorithm 4 so that:*

$$\forall k \in S : \boldsymbol{y}^{(k)} \leftarrow (2\gamma_k(L^{(k,k)})^T\Sigma - I)\boldsymbol{v}_k$$

This modified version of Algorithm 4 returns the set of Nash equilibria if they exist, and otherwise returns "No Nash Equilibrium."

Proof. We begin by proving that agent k's optimal negotiating position $\boldsymbol{\delta}_k^\star$ given $\boldsymbol{\delta}^*_{j\in S:j\neq k}$ is the solution to the linear system:

$$(2\gamma_k(L^{(k,k)})^T\Sigma - I)\boldsymbol{v}_k = T^{(k,k)}\boldsymbol{\delta}_k^\star + \sum_{j\in S:j\neq k} T^{(k,j)}\boldsymbol{\delta}_j^\star \tag{5.9}$$

where $T^{(p,q)}$ and L are defined as in Algorithm 4.

Let $\Delta_M \in \mathbb{R}^{n\times n}$ have i^{th} column $\boldsymbol{\delta}_i^\star$ if $i \in S$ and zero otherwise. Let (W', P') be the stable point resulting from a choice of $M' := M + \Delta_M$ as the agents' negotiating positions in

the Strategy Phase. From Theorem 5.2.2, we have

$$\begin{aligned}
\text{vec}(W') &= \text{vec}(W) \\
&\quad + 0.5(\Sigma \otimes \Gamma + \Gamma \otimes \Sigma)^{-1}\text{vec}(\Delta_M + \Delta_M^T) \\
\Rightarrow \text{vec}(W' - W) &= L\text{vec}(\Delta_M) \\
\Rightarrow \boldsymbol{w}'_k - \boldsymbol{w}_k &= L^{(k,k)}\boldsymbol{\delta}_k + \sum_{j \in S: j \neq k} L^{(k,j)}\boldsymbol{\delta}_j,
\end{aligned}$$

where the second line follows from Lemma 5.8.5, and the matrix L is defined as in Algorithm 4. Notice that there is a distribution on $\boldsymbol{w}_k$ induced by $\mathcal{D}_k$.

Now, fix an agent $k \in S$. They want to choose $\boldsymbol{w}'_k$ optimally based on the above equation, but are uncertain about the value of $\boldsymbol{w}_k$.

Let $A_k := L^{(k,k)}$ and $\boldsymbol{b}_k := \sum_{j \in S: j \neq k} L^{(k,j)}\boldsymbol{\delta}^*_j$. By Lemma 5.8.3, we have:

$$\begin{aligned}
g_k &= -\langle \boldsymbol{\delta}_k, \boldsymbol{w}'_k \rangle + \gamma_k \langle \boldsymbol{w}'_k, \Sigma \boldsymbol{w}'_k \rangle \\
&= -\langle \boldsymbol{\delta}_k, (\boldsymbol{w}_k + A_k\boldsymbol{\delta}_k + \boldsymbol{b}_k) \rangle \\
&+ \gamma_k \langle (\boldsymbol{w}_k + A_k\boldsymbol{\delta}_k + \boldsymbol{b}_k), \Sigma(\boldsymbol{w}_k + A_k\boldsymbol{\delta}_k + \boldsymbol{b}_k) \rangle \\
&= -\langle \boldsymbol{\delta}_k, A_k\boldsymbol{\delta}_k \rangle - \langle \boldsymbol{\delta}_k, \boldsymbol{b}_k \rangle - \langle \boldsymbol{\delta}_k, \boldsymbol{w}_k \rangle + 2\gamma_k \langle \Sigma \boldsymbol{w}_k, A_k\boldsymbol{\delta}_k + \boldsymbol{b}_k \rangle \\
&+ \gamma_k \langle A_k\boldsymbol{\delta}_k + \boldsymbol{b}_k, \Sigma(A_k\boldsymbol{\delta}_k + \boldsymbol{b}_k) \rangle + \gamma_k \langle \boldsymbol{w}_k, \Sigma \boldsymbol{w}_k \rangle
\end{aligned}$$

Therefore agent $k \in S$ wants to optimize:

$$\begin{aligned}
\mathop{\mathbb{E}}_{M \sim \mathcal{D}_k}[g_k] &= \mathop{\mathbb{E}}_{M \sim \mathcal{D}_k}\Big[- \langle \boldsymbol{\delta}_k, A_k\boldsymbol{\delta}_k \rangle - \langle \boldsymbol{\delta}_k, \boldsymbol{b}_k \rangle - \langle \boldsymbol{\delta}_k, \boldsymbol{w}_k \rangle + 2\gamma_k \langle \Sigma \boldsymbol{w}_k, A_k\boldsymbol{\delta}_k + \boldsymbol{b}_k \rangle \\
&\quad + \gamma_k \langle A_k\boldsymbol{\delta}_k + \boldsymbol{b}_k, \Sigma(A_k\boldsymbol{\delta}_k + \boldsymbol{b}_k) \rangle + \gamma_k \langle \boldsymbol{w}_k, \Sigma \boldsymbol{w}_k \rangle \Big] \\
&= -\langle \boldsymbol{\delta}_k, A_k\boldsymbol{\delta}_k \rangle - \langle \boldsymbol{\delta}_k, \boldsymbol{b}_k \rangle + \gamma_k \langle A_k\boldsymbol{\delta}_k + \boldsymbol{b}_k, \Sigma(A_k\boldsymbol{\delta}_k + \boldsymbol{b}_k) \rangle \\
&\quad + \mathop{\mathbb{E}}_{M \sim \mathcal{D}_k}\Big[- \langle \boldsymbol{\delta}_k, \boldsymbol{w}_k \rangle + 2\gamma_k \langle \Sigma \boldsymbol{w}_k, A_k\boldsymbol{\delta}_k + \boldsymbol{b}_k \rangle + \gamma_k \langle \boldsymbol{w}_k, \Sigma \boldsymbol{w}_k \rangle \Big]
\end{aligned}$$

Next, recall that $\mathbb{E}_{M \sim \mathcal{D}_k}[\boldsymbol{w}_k] = \boldsymbol{v}_k$. Let $\boldsymbol{w}_k$ have covariance R_k. Then, by Lemma 5.8.4, we have:

$$\begin{aligned}
\mathop{\mathbb{E}}_{M \sim \mathcal{D}}[g_k] &= -\langle \boldsymbol{\delta}_k, A_k\boldsymbol{\delta}_k \rangle - \langle \boldsymbol{\delta}_k, \boldsymbol{b}_k \rangle + \gamma_k \langle A_k\boldsymbol{\delta}_k + \boldsymbol{b}_k, \Sigma(A_k\boldsymbol{\delta}_k + \boldsymbol{b}_k) \rangle \\
&\quad + \Big(- \boldsymbol{\delta}_k^T \boldsymbol{v}_k + 2\gamma_k \boldsymbol{v}_k^T \Sigma A_k \boldsymbol{\delta}_k + 2\gamma_k \boldsymbol{v}_k^T \Sigma \boldsymbol{b}_k + \gamma_k tr(\Sigma R_k) + \gamma_k \boldsymbol{v}_k^T \Sigma \boldsymbol{v}_k \Big)
\end{aligned}$$

By Theorem 5.3.2, the optimal negotiating position $\boldsymbol{\delta}^\star$ is the critical point of g_k with respect to $\boldsymbol{\delta}_k$. Notice that the Hessian of $\mathbb{E}[g_k]$ with respect to $\boldsymbol{\delta}_k$ does not depend on $\mathcal{D}$, so the critical point gives the optimal negotiating position for $\mathbb{E}[g_k]$. Setting $\nabla_{\delta_k} \mathbb{E}_{M\sim\mathcal{D}}[g_k] = 0$, we obtain:

$$\begin{aligned}(A_k + A_k^T - 2\gamma_k A_k^T \Sigma A_k)\boldsymbol{\delta}_k^\star = (2\gamma_k A_k^T \Sigma - I)\boldsymbol{v}_k \\ + (2\gamma_k A_k^T \Sigma - I)\boldsymbol{b}_k\end{aligned}$$

Notice that the gradients of the quadratic terms $\mathrm{tr}(\Sigma R_k)$ and $\gamma_k^T \boldsymbol{v}_k^T \Sigma \boldsymbol{v}_k$ with respect to $\boldsymbol{\delta}_k$ are zero.

Rearranging terms, we obtain the linear system:

$$\begin{aligned}(2\gamma_k (L^{(k,k)})^T \Sigma - I)\boldsymbol{v}_k = \Bigg[(L^{(k,k)} + (L^{(k,k)})^T - 2\gamma_k (L^{(k,k)})^T \Sigma L^{(k,k)})\boldsymbol{\delta}_k^\star + \sum_{j\in S: j\neq k} L^{(k,j)}\boldsymbol{\delta}_j^\star \\ - 2\gamma_k (L^{(k,k)})^T \Sigma \sum_{j\in S: j\neq k} L^{(k,j)}\boldsymbol{\delta}_j^\star\Bigg] \\ \Rightarrow \boldsymbol{z}^{(k)} = T^{(k,k)}\boldsymbol{\delta}_k^\star + \sum_{j\in S: j\neq k} T^{(k,j)}\boldsymbol{\delta}_j^\star, \qquad (5.10)\end{aligned}$$

where $T^{(p,q)}$ is defined as in Algorithm 4 and and $\boldsymbol{z}^{(k)} = (2\gamma_k (L^{(k,k)})^T \Sigma - I)\boldsymbol{v}_k$. This proves Eq. 5.9.

Having verified Eq. 5.9, it follows that a Nash equilibrium exists, the modified Algorithm 4 finds it. Conversely, a tuple $(\boldsymbol{\delta}_i^\star)_{i\in S}$ that solves Eq. (5.10) for all k is such that $\boldsymbol{\delta}_i^\star$ is the optimal $\boldsymbol{\delta}_i$ for all $i \in S$ given that other agents report $(\boldsymbol{\delta}_j^\star)_{j\in S\setminus\{i\}}$. If no Nash equilibrium exists, Eq. (5.10) cannot be simultaneously satisfied for all k, so the modified Algorithm 4 returns "No Nash Equilibrium." □

5.8.5 Proof of Proposition 5.5.5

For completeness, we first state the techincal result of Bhatia et al. (2015) that we require.

Theorem 5.8.8 (Bhatia et al. (2015))**.** *Let $X \in \mathbb{R}^{n\times d}$ be a design matrix and $C > 0$ an absolute constant. Let $\{0,1\}$ be a corruption vector with $\|\{0,1\}\|_0 \leq \alpha n$, $\alpha \leq C$.*

Let $\boldsymbol{y} = X\boldsymbol{w}^* + \{0,1\}$ be the observed responses, and $\beta \geq \alpha$ be the active set threshold given to the Algorithm 2 of Bhatia et al. (2015).

Suppose X satisfies the SSC property at level $1-\beta$ and SSS property at level β, with constants $\lambda_{1-\beta}$ and and Λ_β respectively. If the data $(X, \boldsymbol{y})$ are such that $\frac{4\sqrt{\Lambda_\beta}}{\sqrt{\lambda_{1-\beta}}} < 1$, then after t iterations, Algorithm 2 of Bhatia et al. (2015) with active set threshold $\beta \geq \alpha$ obtains a solution $\boldsymbol{w}^t \in \mathbb{R}^d$ such that

$$\|\boldsymbol{w}^t - \boldsymbol{w}^*\|_2 \leq \frac{\|\{0,1\}\|_2}{\sqrt{n}} \exp(-ct)$$

for large enough n.

We are ready to prove Proposition 5.5.5.

Proof of Proposition 6.4.2. Let $\text{vec}(\widehat{H'})$ be as in Algorithm 5, and $\boldsymbol{y} = \text{vec}(\widehat{H'})$. Notice $\boldsymbol{y} = \text{vec}(H) + \text{vec}(H' - H) + \text{vec}(\widehat{H'} - H')$. Let $\{0,1\} := \text{vec}(H' - H)$ be the corruption vector due to strategic negotiations and $\boldsymbol{r} = \text{vec}(\widehat{H'} - H')$ be the residual vector. Recall:

$$\text{vec}(\widehat{H'}) = \arg\min_{\boldsymbol{v} \in \mathbb{R}^{n^2}} \|\text{vec}(W') - K^{-1}\boldsymbol{v}\|_2^2.$$

Since K^{-1} is full rank, $\text{vec}(\widehat{H'}) = K\text{vec}(W') = \text{vec}(H')$, so $\boldsymbol{r} = 0$.

Next, we apply Theorem 5.8.8. Let $\tilde{X} = X \otimes X$. Notice that $\|\{0,1\}\|_0 \leq 2ns - s^2 = \beta$ since $(H' - H)_{i,j}$ is zero if $i, j \notin S$. Therefore the fraction of corrupted entries is at most $\beta = \frac{2|S|}{n} - \frac{|S|^2}{n^2} \leq 1$. Therefore, if C is the constant in Theorem 5.8.8, then $\alpha \leq C$ if and only if $|S| \leq C'n$ for some constant C' depending on C.

Further, the design matrix $\tilde{X}$ satisfies the required SSC and SSS conditions. Therefore, after T iterations, Algorithm 5 obtains $\hat{B}$ such that:

$$\|\text{vec}(\hat{B} + \hat{B}^T - (B + B^T))\|_2 \leq \frac{\exp(-cT)}{n} \|\text{vec}(H' - H)\|_F$$

□

5.8.6 Estimating the set of strategic agents

Proposition 5.8.9. *Under the conditions of Proposition 6.4.2, let b_{min} be the least nonzero entry of $(H' - H)$ in absolute value and $T > 0$. Then there exist constants $\rho, C > 0$ such that if b_{min} satisfies:*

$$|b_{min}| > \exp(-\rho T)\|\{0,1\}\|_2$$

then Algorithm 5 with threshold parameter $\beta = \frac{2n|S| - |S|^2}{n^2}$ and T iterations of TORRENT *recovers S exactly.*

Proof. We proceed by analyzing the residual matrix R of Algorithm 5.

Let $\eta = \frac{4\sqrt{\Lambda_\beta}}{\sqrt{\lambda_{1-\beta}}}$. We have $\eta < 1$ by assumption, so let $\rho = 1 - \eta > 0$. Recall that TORRENT maintains a set $S_t \subset [n^2]$ called the *active set*, which is its guess at iteration t for what indices of the response vector $\text{vec}(H')$ are non-corrupted. Let $\hat{B}^{(t)} \in \mathbb{R}^{d\times d}$ be the estimate of TORRENT at iteration t. Let $\boldsymbol{b}^{(t)} = \text{vec}(H') - \frac{1}{2}(X \otimes X)\text{vec}(\hat{B}^{(t)} + \hat{B}^{(t)})$ be the residual at iteration t, and $\boldsymbol{b}_{S_t} \in \mathbb{R}^{n^2}$ be the coordinate projection vector such that:

$$\boldsymbol{b}_{S_t;i} = \begin{cases} \boldsymbol{e}_i^T \boldsymbol{b}^{(t)} & i \in S_t \\ 0 & \text{otherwise} \end{cases}$$

From the proof of Bhatia et al. (2015) Theorem 10, we obtain that if S_{t+1} is the active set at time $t+1$, then:

$$\|\boldsymbol{b}_{S_{t+1}}\|_2 \leq \eta \|\boldsymbol{b}_{S_t}\|_2,$$

Successively applying the inequality and noting that the first estimated active set $S_0 = [n]^2$, we have that:

$$\begin{aligned} \|\boldsymbol{b}_{S_{t+1}}\|_2 &\leq \eta^{t+1}\|\boldsymbol{b}\|_2 \\ &\leq \exp(-\rho t)\|\boldsymbol{b}\|_2 \end{aligned}$$

By assumption on b_{min}, the above event can only occur if $\|\boldsymbol{b}_{S_{t+1}}\|_2 = 0$. Hence $\|\boldsymbol{b}_{S_T}\|_2 = 0$, so the final active set S_T must be a subset of the non-corrupted entries of $\boldsymbol{b}$. Hence $R_{ij} = 0$ if $i \notin S, j \notin S$. Further, since S_T is the output of a hard-thresholding operation, $|S_T| = (1-\beta)n^2 = 2sn - n^2$. Therefore, after a permutation, the residual matrix is precisely

$$R = \begin{bmatrix} \mathbf{1}_s\mathbf{1}_s^T & \mathbf{1}_s\mathbf{1}_{n-s}^T \\ \mathbf{1}_{n-s}\mathbf{1}_s^T & \mathbf{0}_{n-s}\mathbf{0}_{n-s}^T \end{bmatrix}.$$

A calculation shows that R is rank-two with nonzero eigenvalues $\frac{s(1\pm\sqrt{(4n/s)-3})}{2}$. Let λ_2 be the lesser eigenvalue. The corresponding eigenvector $\boldsymbol{v}_2$ has entries a, b at indices $\{1, 2, \ldots, s\}$ and indices $\{s+1, \ldots, n\}$ respectively, where $a = \frac{1-\sqrt{(4n/s)-3}}{2}, b = 1$. Let S_1, S_2 be as in Algorithm 5. We have that $S_1 = S$ and $S_2 = [n] \setminus S$ by Rohe et al. (2011).

Since $\beta = \frac{2n|S|-|S|^2}{n^2}$, it follows that $\frac{\sqrt{8\beta n}+1}{4}$ is closer to $|S|$ than $n - |S|$, so the output $\hat{S}$ is $S_1 = S$. □

5.8.7 Proof of Proposition 5.4.2

Proof of Proposition 5.4.2. We first analyze the global welfare under W' versus W.

Let W' be the outcome under strategic negotiations with $S = [n]$ and W be the outcome under honest negotiations. Let Δ be such that $M' = M + \Delta$. Let L, K, T, Π be as in Definition 5.3.6. Notice that $\text{vec}(W') = L\text{vec}(M')$ and $\text{vec}(W) = L\text{vec}(M)$. Moreover, we can simplify $L\text{vec}(M)$ as follows. Let $M = H + \tilde{H}$ where $H := \frac{1}{2}(M + M^T)$ is the symmetric part of M and $\tilde{H}$ is the skew-symmetric part. Notice that $L\text{vec}(H + \tilde{H}) = K^{-1}(\frac{1}{2}I + \frac{1}{2}\Pi)\text{vec}(H + \tilde{H}) = K^{-1}\text{vec}(H)$.

Therefore, by Lemma 5.8.3, we see that:

$$\begin{aligned}
\sum_{k=1}^{n} g_k(W', P') &= \sum_{k=1}^{n} g_k(W', P') \\
&= \sum_{k=1}^{n} \Big[- \langle \boldsymbol{\delta}_k, \boldsymbol{w}'_k \rangle + \langle \boldsymbol{w}'_k, \Sigma \boldsymbol{w}'_k \rangle \Big] \\
&= \langle -\Delta, W' \rangle_F + \text{tr}(W' \Sigma W') \\
&= \langle \Sigma W' - \Delta, W' \rangle_F
\end{aligned}$$

In the honest case we have $\Delta = 0$, so:

$$\begin{aligned}
\sum_{k=1}^{n} g_k(W, P) &= \langle \Sigma W, W \rangle_F \\
&= \langle (I \otimes \Sigma)\text{vec}(W), \text{vec}(W) \rangle \\
&= \langle K^{-1}\text{vec}(H), (I \otimes \Sigma)K^{-1}\text{vec}(H) \rangle
\end{aligned}$$

For the strategically negotiated equilibrium, we first simplify Δ. Let $\boldsymbol{y}$ be as in Algorithm 4, so that $\text{vec}(\Delta) = T^{-1}\boldsymbol{y}$. Let $D_L \in \mathbb{R}^{n^2 \times n^2}$ be block diagonal with $n \times n$ diagonal blocks equal to those of L, and zero elsewhere. From inspection, we see that:

$$\begin{aligned}
T &= D_L + D_L^T + 2D_L^T(I \otimes \Sigma)D_L + (L - D_L) - 2D_L^T(I \otimes \Sigma)(L - D_L) \\
&= D_L^T + L - 2D_L^T(I \otimes \Sigma)L
\end{aligned}$$

Similarly,

$$\begin{aligned}
\boldsymbol{y} &= (2D_L^T(I \otimes \Sigma) - I_{n^2})\text{vec}(W) \\
&= (2D_L^T(I \otimes \Sigma) - I_{n^2})L\text{vec}(M)
\end{aligned}$$

Therefore, let $A_2 = (2D_L^T(I \otimes \Sigma) - I_{n^2})$ and $A = T^{-1}A_2L$ so that:

$$\text{vec}(\Delta) = A\text{vec}(M) = T^{-1}A_2L\text{vec}(M)$$

Then, let $A_3 = T^{-1}A_2$. Note that A_3 corresponds to the matrix C in the statement of the Proposition. The total utility at (W', P') is:

$$\begin{aligned}
\sum_{k=1}^{n} g_k(W', P') &= \langle \Sigma W' - \Delta, W' \rangle_F \\
&= \text{vec}(M)^T \Big((I + A)^T L^T (I \otimes \Sigma) L (I + A) - A^T (L + LA) \Big) \text{vec}(M) \\
&= (K^{-1}\text{vec}(H))^T \Big((I + A_3^T L^T)(I \otimes \Sigma)(I + LA_3) - A_3^T (I + LA_3) \Big) K^{-1}\text{vec}(H) \\
&= K^{-1}\text{vec}(H))^T (I \otimes \Sigma) K^{-1}\text{vec}(H) + K^{-1}\text{vec}(H))^T Y K^{-1}\text{vec}(H))^T,
\end{aligned}$$

where we define Y as:

$$Y := 2(I \otimes \Sigma)LA_3 + A_3^T L^T (I \otimes \Sigma) LA_3 - A_3^T (I + LA_3).$$

Then, let:

$$\gamma := \frac{K^{-1}\text{vec}(H)^T Y K^{-1}\text{vec}(H)^T}{K^{-1}\text{vec}(H)^T (I \otimes \Sigma) K^{-1}\text{vec}(H)^T}.$$

We see that:

$$\text{PoS} = \frac{1}{1 + \gamma}$$

Note that $1 + \gamma \geq 0$ always, since utilities cannot be negative. However, we may have $\gamma \in (-1, 0)$.

We will upper bound $|\gamma|$. The matrix $I \otimes \Sigma$ has eigenvalues $\lambda_1, \ldots, \lambda_n$ each with multiplicity n. Therefore:

$$|\gamma| \leq \frac{\|Y\|_2}{\lambda_n}$$

It remains to upper bound $\|Y\|_2$.

First, notice that $\Pi(\boldsymbol{v} \otimes \boldsymbol{w}) = \boldsymbol{w} \otimes \boldsymbol{v}$ for any $\boldsymbol{v}, \boldsymbol{w} \in \mathbb{R}^n$. Let $\Sigma = \sum_{i=1}^n \lambda_i \boldsymbol{u}_i \boldsymbol{u}_i^T$ for $\lambda_1 \geq \cdots \geq \lambda_n > 0$. By properties of Kronecker products Horn and Johnson (2008),

$$K^{-1} = \sum_{i=1}^{n} \sum_{j=1}^{n} \frac{1}{2(\lambda_i + \lambda_j)} (\boldsymbol{u}_i \otimes \boldsymbol{u}_j)(\boldsymbol{u}_i \otimes \boldsymbol{u}_j)^T.$$

Therefore, since $L = \frac{1}{2}(K^{-1} + K^{-1}\Pi)$, we have the following.

$$L = \sum_{i=1}^{n}\sum_{j=1}^{n} \frac{1}{2(\lambda_i + \lambda_j)}(\boldsymbol{u}_i \otimes \boldsymbol{u}_j)(\boldsymbol{u}_i \otimes \boldsymbol{u}_j + \boldsymbol{u}_j \otimes \boldsymbol{u}_i)^T$$

$$(I \otimes \Sigma)L = \sum_{i=1}^{n}\sum_{j=1}^{n} \frac{\lambda_j}{2(\lambda_i + \lambda_j)}(\boldsymbol{u}_i \otimes \boldsymbol{u}_j)(\boldsymbol{u}_i \otimes \boldsymbol{u}_j + \boldsymbol{u}_j \otimes \boldsymbol{u}_i)^T$$

$$L^T(I \otimes \Sigma)L = \sum_{i=1}^{n}\sum_{j=1}^{n} \frac{\lambda_j}{4(\lambda_i + \lambda_j)^2}(\boldsymbol{u}_i \otimes \boldsymbol{u}_j + \boldsymbol{u}_j \otimes \boldsymbol{u}_i)(\boldsymbol{u}_i \otimes \boldsymbol{u}_j + \boldsymbol{u}_j \otimes \boldsymbol{u}_i)^T$$

We see that $\|L\| \leq \frac{1}{2\lambda_n}$, $\|(I \otimes \Sigma)L\|_2 \leq 1$, and $\|L^T(I \otimes \Sigma)L\|_2 \leq \frac{1}{4\lambda_n}$. Therefore, by triangle inequality and sub-multiplicativity of operator norms,

$$\begin{aligned}\|Y\|_2 &\leq \|A_3\|_2\Big(\|2(I \otimes \Sigma)L\|_2 + 1\Big) + \|A_3\|_2^2\Big(\|L^T(I \otimes \Sigma)L\| + \|L\|\Big) \\ &\leq 3\|A_3\|_2 + \frac{3}{4\lambda_n}\|A_3\|_2^2\end{aligned}$$

If $\|A_3\|_2 \leq \frac{1}{10}\lambda_n$, then $|\gamma| < \frac{1}{2}$. Therefore,

$$\begin{aligned}\text{PoS} &= \frac{1}{1+\gamma} \\ &\leq \frac{1}{1-|\gamma|} \\ &\leq O(1).\end{aligned}$$

□

5.9 Analysis of Model Networks

In this section we will prove Proposition 5.9.3 that characterizes Example 5.4.3, and also analyze an additional model network.

5.9.1 Two Agents That Can Self-Invest

Consider a network with two agents (P1 and P2) who can form a contract with each other, and each can invest in themselves ("self-invest"). The network setting is as follows.

$$M = \begin{bmatrix} M_{11} & M_{12} \\ M_{21} & M_{22} \end{bmatrix}, \quad \Sigma = \begin{bmatrix} 1 & \rho \\ \rho & 1 \end{bmatrix}, \quad \Gamma = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \tag{5.11}$$

The parameter $\rho \in (-1, 1)$ is the correlation between an agent's returns from self-investment versus trading. For $\rho \approx 1$, returns from self-investing and trading move in lockstep. So, each

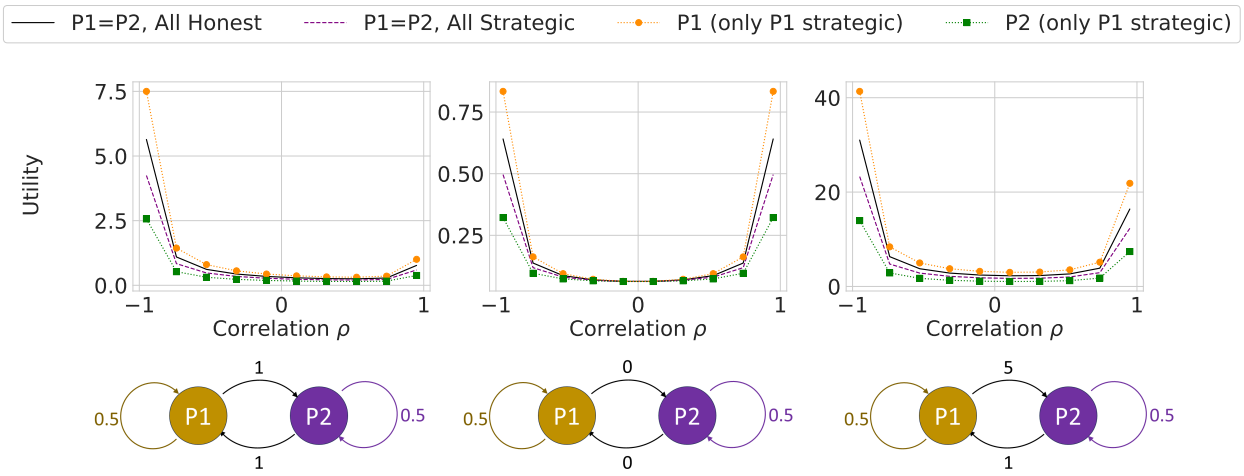


Figure 5.7: *Nash equilibria for two agents:* The utility for either agent when both are honest (solid line) is higher than when both are strategic (dashed line). When only agent P1 is strategic, P1 gains the highest utility (dotted circles) while P2's utility is lowest (dotted squares). The network settings are shown in the bottom row, with an arrow from i to j corresponding to M_{ji}.

agent must hedge between self-investing and trading, hoping to benefit from any differences in their returns. But as ρ goes to -1, the risks from self-investing and trading offset. If both offer positive returns, an agent can gain nearly risk-free reward. Hence, negative correlations can lead to higher utility for agents.

Theorem 5.9.1. *Consider the network setting of Eq. 5.11. Let* $\kappa := \rho(M_{11} + M_{22}) - (M_{12} + M_{21})$*, and let* Δ *be the negotiation positions (Definition 5.3.1) for all agents at the Nash equilibrium.*

1. *If only agent* k *is strategic, then*

$$\Delta_{ij} = \begin{cases} \kappa/3 & \text{if } i \neq k, j = k \\ 0 & \text{otherwise} \end{cases}$$

2. *If both agents are strategic, then*

$$\Delta_{ij} = \begin{cases} \kappa/4 & \text{if } i \neq j \\ 0 & \text{otherwise} \end{cases}$$

Remark 5.9.2. *We can show that it is strategic for* i *to report* $\Delta_{ii} = 0$ *even if* $\Sigma_{1;1} \neq \Sigma_{2;2}$ *and* $\Gamma_{11} \neq \Gamma_{22}$.

Figure 5.7 shows the agents' utility for honest versus strategic negotiating positions over a range of ρ. We can make several observations.

Strategic agents report self-investing returns truthfully. Suppose agent i claims that her self-investments have higher returns than in reality (that is, $M'_{ii} > M_{ii}$). If agent j wants to trade with i, then j will have to offer better trading terms via better prices. Thus, high self-investing returns are a plausible negotiating strategy. However, Theorem 5.9.1 shows that $\Delta_{ii} = 0$ at the Nash equilibrium, so $M'_{ii} = M_{ii}$. This is because if both agents make untrue claims about self-investing, they get smaller contracts, lowering utility.

Payments to others can increase when the agent becomes strategic. It may appear that strategic agents can only increase their utility by extracting higher payments from others. However, this need not be true. Suppose both agents have a utility of 1 from self-investing and 0 from trading. By symmetry, if both agents are honest, they make no payments. Now, suppose only agent P1 is strategic and $\rho \approx 1$. By Theorem 5.9.1, P1 will claim to have *higher* returns from trading than her actual returns. This implies that P1 pays P2 during contract formation. But the contract size also changes. With the new contract size, P1 still gains utility at the expense of P2.

Utility is lower when both agents are strategic. Figure 5.7 shows several instances where the agents are worse off when both are strategic versus when both are honest. This is because the agents face a Prisoner's Dilemma. If both are honest, they cooperate, and both gain high utility. However, being strategic is a dominant strategy. This forces both to be strategic, leading to lower utility for both.

Negative correlations amplify the effect of negotiating positions. Suppose self-investing and trading both have positive expected returns. When correlations are negative, their risks cancel while their returns add. So, an agent can take large positions and achieve high utility. But, as noted in the previous paragraph, there is a drop in utility when both agents are honest versus strategic. We find a larger drop for negative correlations. Hence, under negative correlations, the effect of strategic behavior is also more pronounced.

5.9.2 One Investor and Two Hedge Funds (Example 5.4.3)

Recall Example 5.4.3. We have a 3-agent network where one investor interacts with two hedge funds under the following network setting.

$$M = \begin{bmatrix} 0 & a & a \\ m & 0 & 0 \\ m & 0 & 0 \end{bmatrix}, \quad \Sigma = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & \rho \\ 0 & \rho & 1 \end{bmatrix}, \quad \Gamma = I. \tag{5.12}$$

The first column corresponds to the investor, and the others to the hedge funds. Under this setting, the hedge funds do not want to trade with each other, and none of the agents want to self-invest. Also, the hedge funds are correlated with each other (via ρ), and uncorrelated with the investor.

The optimal negotiating positions are as follows.

Proposition 5.9.3 (Restatement of Proposition 5.9.3)**.** *Consider the the network setting of Eq. 5.3, where strategic agents can only modify the non-zero entries in their column of M. Define*

$$\nu = \frac{1}{2}\left(\frac{1}{2-\rho} + \frac{1}{2+\rho}\right), \qquad \eta = \frac{1}{2}\left(\frac{1}{2-\rho} - \frac{1}{2+\rho}\right),$$
$$\zeta = \frac{\nu - \eta}{\nu + (\nu-\eta)(1-\nu)}.$$

1. Honest investor and strategic hedge funds ($S = \{2, 3\}$):

$$M'_{21} = M'_{31} = m, \qquad M'_{12} = M'_{13} = \frac{a\nu - m(1-\nu)(\nu-\eta)}{\nu + (1-\nu)(\nu-\eta)}.$$

2. All agents strategic ($S = [n]$):

$$M'_{21} = M'_{31} = \frac{m - a\zeta}{1+\zeta},$$
$$M'_{12} = M'_{13} = \frac{a\nu - M'_{21}(1-\nu)(\nu-\eta)}{\nu + (1-\nu)(\nu-\eta)}.$$

We first discuss insights from Proposition 5.9.3 and Figure 5.2b of the main text, and then give the proof.

The investor's utility is very sensitive to her negotiating position. Suppose the investor is honest and both hedge funds are strategic. Then, the investor will accept worse terms from the funds and achieve less utility. But the situation is reversed if the investor is

also strategic (Figure 5.2b). The investor now achieves higher utility than either fund. Thus, the investor's outcome is very sensitive to her negotiating position.

The sensitivity to negotiating positions increases as $\rho \to -1$. Figure 5.2b shows that when ρ decreases, the investor loses utility if she is honest but gains utility if she is strategic. The reason is that as $\rho \to -1$, the investor wishes to invest almost equally in both funds to reduce her risk. The hedge funds only form one contract each. Since they cannot hedge their risk, they prefer much smaller contracts than the investor. The investor can extract higher payments for this, increasing her utility.

Strategic behavior can reduce utility. Suppose the investor is honest. As $\rho \approx -1$, the hedge funds are worse off being strategic than if they were both honest (Figure 5.2b). If both funds are honest, their contract sizes match their risk preference. However, if both are strategic, each fund worries about its competitor. So, both funds end up with worse terms.

We prove Proposition 5.9.3 Part 1 in Corollary 5.9.8, and Part 2 in Proposition 5.9.9 below. Throughout the remainder of this section, we will refer to the investor as P1 and the hedge funds as P2 and P3. Hence P1 has beliefs according to $M\boldsymbol{e}_1$, and so on.

Proposition 5.9.4. *Assume that P1 reports M_{21}, M_{31} as $\tilde{m}$, P2 reports M_{12} as $\tilde{a}$, and P3 reports M_{13} as $\tilde{b}$. Then $w_{21} = 0.5 \cdot (\tilde{\alpha} + \tilde{a}\nu - \tilde{b}\eta)$ and $w_{31} = 0.5 \cdot (\tilde{\alpha} + \tilde{b}\nu - \tilde{a}\eta)$ for $\tilde{\alpha} = \frac{\tilde{m}}{2+\rho}$.*

Proof. Notice that Σ has eigenvalues $\lambda_1 = 1, \lambda_2 = 1 + \rho, \lambda_3 = 1 - \rho$ and corresponding eigenvectors $\boldsymbol{v}_1 = (1,0,0)^T, \boldsymbol{v}_2 = \frac{1}{\sqrt{2}}(0,1,1)^T, \boldsymbol{v}_3 = \frac{1}{\sqrt{2}}(0,1,-1)^T$. Therefore $2^{-1/2}(\boldsymbol{w}_2 + \boldsymbol{w}_3) = W\boldsymbol{v}_2$. Let $\tilde{M}$ be the matrix of reported values, so $\tilde{M}_{21} = \tilde{M}_{31} = \tilde{m}$, $\tilde{M}_{12} = \tilde{a}$, and $\tilde{M}_{13} = \tilde{b}$. All other entries of $\tilde{M}$ are zero.

From Theorem 5.2.2, the resulting network is $W = \sum_{i,j \in [3]} \frac{\boldsymbol{v}_i^T(\tilde{M} + \tilde{M}^T)\boldsymbol{v}_j}{2(\lambda_i + \lambda_j)} \boldsymbol{v}_i \boldsymbol{v}_j$. By orthogonality of eigenvectors, we have:

$$W\boldsymbol{v}_2 = \sum_i \frac{\boldsymbol{v}_i^T(\tilde{M} + \tilde{M}^T)\boldsymbol{v}_2}{2(\lambda_i + \lambda_2)} \boldsymbol{v}_i$$

Only the term at $i = 1$ is nonzero, and therefore $2^{-1/2}(\boldsymbol{w}_2 + \boldsymbol{w}_3) = \frac{2\tilde{m} + \tilde{a} + \tilde{b}}{2\sqrt{2}(2+\rho)} \boldsymbol{e}_1$. Similarly, $2^{-1/2}(\boldsymbol{w}_2 - \boldsymbol{w}_3) = \frac{\tilde{a} - \tilde{b}}{2\sqrt{2}(2-\rho)} \boldsymbol{e}_1$. The conclusion follows. □

Proposition 5.9.5. *Let $\tilde{\alpha}$ be as in Proposition 5.9.4. Assume P1 reports $\tilde{\mu}$, and P3 reports $\tilde{b}$. The optimal choice of reported $\tilde{a}$ for P2 is:*

$$\tilde{a}^* = c_a + s\tilde{b}$$

For $c_a = \frac{a\nu - \tilde{\alpha}(1-\nu)}{\nu(2-\nu)}$ and $s = \frac{\eta(1-\nu)}{\nu(2-\nu)}$.

Proof. Let $w := w_{21}$ for shorthand. The utility of firm 2, by Lemma 5.8.3, is given by:

$$\begin{aligned} g_2 &= -w(\tilde{a} - a) + w^2 \\ &= w(a - \tilde{a} + w) \\ 2g_2 &= (\tilde{\alpha} + \tilde{a}\nu - \tilde{b}\eta)(a + (0.5\nu - 1)\tilde{a} + 0.5\tilde{\alpha} - 0.5\tilde{b}\eta) \end{aligned}$$

The coefficient of $\tilde{a}^2$ in g_2 is $\nu(0.5\nu - 1)$. Since $\nu > 0$ and $0.5\nu < 1$ for all $\rho \in (-1, 1)$, the Hessian of g_2 with respect to $\tilde{a}$ is negative definite, and so the optimal choice of a is at the critical point $\frac{\partial g_2}{\partial \tilde{a}} = 0$. Solving for $\tilde{a}$ gives:

$$\tilde{a}^* = c_a + s\tilde{b},$$

with c_a, s as in the statement of the proposition. □

A symmetric argument gives the following.

Proposition 5.9.6. *Let $\tilde{\alpha}$ be as in Proposition 5.9.4. Assume P1 reports $\tilde{\mu}$, and P2 reports $\tilde{a}$. The optimal choice of reported $\tilde{b}$ for P3 is:*

$$\tilde{b}^* = c_b + s\tilde{a}$$

For $c_a = \frac{b\nu - \tilde{\alpha}(1-\nu)}{\nu(2-\nu)}$ and $s = \frac{\eta(1-\nu)}{\nu(2-\nu)}$.

Next, we can solve for the Nash equilibria given the reported $\tilde{m}$ of the investor.

Proposition 5.9.7. *If $M_{12} = M_{13} = a$, then let $c := c_a = c_b$ and s be as in Proposition 5.9.5 and 5.9.6. Assume P1 reports $\tilde{\mu}$. The Nash equilibrium for P2, P3 is to report:*

$$M'_{12} = M'_{13} = \frac{c}{1 - s}$$

Corollary 5.9.8 (Proposition 5.9.3 Part 1)**.** *Assume P1 reports $\tilde{m}$. If both hedge funds are strategic, then the Nash equilibrium for P2, P3 is to report:*

$$M'_{12} = M'_{13} = \frac{a\nu - \tilde{m}(1-\nu)(\nu-\eta)}{\nu + (1-\nu)(\nu-\eta)}$$

Hence if $\tilde{m} = m$, then $M'_{12} = M'_{13}$ are as in Proposition 5.9.3.1.

Proof. We simplify $a_{NS} = \frac{c}{1-s}$ as follows.

$$\begin{aligned}
\frac{c}{1-s} &= \frac{a\nu - \tilde{\alpha}(1-\nu)}{\nu(2-\nu)(1-s)} \\
&= \frac{a\nu - \tilde{\alpha}(1-\nu)}{\nu(2-\nu)\left(1 - \frac{\eta(1-\nu)}{\nu(2-\nu)}\right)} \\
&= \frac{a\nu - \tilde{m}(1-\nu)(\nu-\eta)}{\nu(2-\nu) - \eta(1-\nu)} \\
&= \frac{a\nu - \tilde{m}(1-\nu)(\nu-\eta)}{\nu + (1-\nu)(\nu-\eta)}
\end{aligned}$$

In the setting of Proposition 5.9.3.1, the investor reports $\tilde{m} = m$. The conclusion follows. □

Next, we solve for the optimal report of the investor if all agents are strategic.

Proposition 5.9.9 (Proposition 5.9.3 Part 2)**.** *Let $y = \frac{1}{2(2+\rho)}$. If all agents are strategic, then the optimal reported $\tilde{m}$ for the investor is:*

$$M'_{21} = M'_{31} = \frac{m - a\zeta}{1+\zeta}$$

For $\zeta = \frac{\nu-\eta}{\nu+(\nu-\eta)(1-\nu)} = (1 - 2y(1+\rho))$. The optimal report for the hedge funds is:

$$M'_{12} = M'_{13} = \frac{a\nu - M'_{21}(1-\nu)(\nu-\eta)}{\nu + (1-\nu)(\nu-\eta)}$$

Proof. From Proposition 5.9.7 and 5.9.4, we know $w_{12} = w_{13}$. Therefore by Lemma 5.8.3, if P1 reports $\tilde{m}$ then the investor utility is:

$$g_1(\tilde{m}) = 2(m - \tilde{m})w_{12} + w_{12}^2(2 + 2\rho)$$

Let $w_{12} = (c_2 + \tilde{m}y)$ for shorthand. Then g_1 is quadratic in $\tilde{m}$ and the coefficient of $\tilde{m}^2$ is $2y^2 + 2\rho y^2 - 2y = \frac{-(\rho+3)}{2(\rho+2)^2} < 0$. Therefore, the optimal $\tilde{m}$ is at the critical point $\frac{\partial g_1}{\partial \tilde{m}} = 0$. This is given as:

$$\begin{aligned}
\tilde{m} &= \frac{-c_2 + my + 2c_2y(1+\rho)}{2y(1 - y(1+\rho)} \\
&= \frac{m - (c_2/y)(1 - 2y(1+\rho))}{2(1 - y(1+\rho))} \\
&= \frac{m - (c_2/y)(1 - 2y(1+\rho))}{1 + (1 - 2y(1+\rho))}
\end{aligned}$$

We simplify the terms $(c_2/y), 2y(1+\rho)$ as follows. First, notice that $c_2 = \frac{\nu-\eta}{2} \cdot \frac{a\nu}{\nu(1-\nu)(1-s)}$, where a is true value of M_{12} and M_{13} for the hedge funds. Let $c_1 := \frac{a\nu}{\nu(1-\nu)(1-s)}$ for shorthand, so that $c_2 = \frac{\nu-\eta}{2}c_1$. Next,

$$\begin{aligned}
\frac{c_2}{y} &= \frac{c_1}{\nu - \eta} \\
&= \frac{a}{(2-\nu)(1-s)} \cdot \frac{2(\nu-\eta)}{2(2+\rho)^{-1}(1 - (\nu-\eta)x)} \\
&= \frac{a}{(2-\nu)(1-s)} \\
&\quad \cdot \left(1 - (\nu - \eta) \cdot \frac{1-\nu}{\nu(2-\nu)(1-s)}\right)^{-1} \\
&= \frac{a\nu}{\nu(2-\nu)(1-s) - (\nu-\eta)(1-\nu)} \\
&= \frac{a\nu}{\nu(2-\nu) - \eta(1-\nu) - (\nu-\eta)(1-\eta)} \\
&= \frac{a\nu}{\nu} \\
&= a
\end{aligned}$$

Moreovoer,

$$\begin{aligned}
1 - 2y(1+\rho) &= 1 - 2(c_2/a)(1+\rho) \\
&= 1 - (1+\rho)\frac{c_1(\nu-\eta)}{a} \\
&= 1 - (1+\rho)\cdot\frac{a}{(2-\nu)(1-s)}\frac{(\nu-\eta)}{a} \\
&= 1 - \frac{(\nu-\eta)(1+\rho)\nu}{\nu(2-\nu)-\eta(1-\nu)} \\
&= 1 - \frac{\nu(\nu-\eta)(1+\rho)}{\nu+(\nu-\eta)(1-\nu)} \\
&= 1 - \frac{\nu(\nu-\eta)((2+\rho)-1)}{\nu+(\nu-\eta)(1-\nu)} \\
&= 1 - \frac{\nu-\nu(\nu-\eta)}{\nu+(\nu-\eta)(1-\nu)} \\
&= \frac{\nu-\eta}{\nu+(\nu-\eta)(1-\nu)}
\end{aligned}$$

Let $\zeta := \frac{\nu-\eta}{\nu+(\nu-\eta)(1-\nu)} = 1 - 2y(1+\rho)$. The optimal $\tilde{m}^* = \frac{m-a\zeta}{1+\zeta}$.

Finally, substituting this $\tilde{m}^*$ into Corollary 5.9.8 gives the optimal reports for the hedge funds. □

5.10 Additional Experiments

In this section, we describe additional results from negotiations on international trade networks. We visualize the utility of specific agents at all timesteps under various scenarios, and discuss the implications. The main takeaway is that when a *lone* actor is strategic, then they are better off than when they were honest. Moreover, the strategic actor gains utility at the expense of others.

Figure 5.8 illustrates strategic behavior by the UK, and its effect on UK and another large economy (the Netherlands) which is generally worse off. We find that the Netherlands is worst-off when all are strategic, and best off if all are honest. Moreover, each strategic actor is best-off when they are the lone strategic actor.

For different choices of lone strategic actor, such as the US (Figure 5.9) and India (Figure 5.10), we obtain analogous results to Figure 5.8.

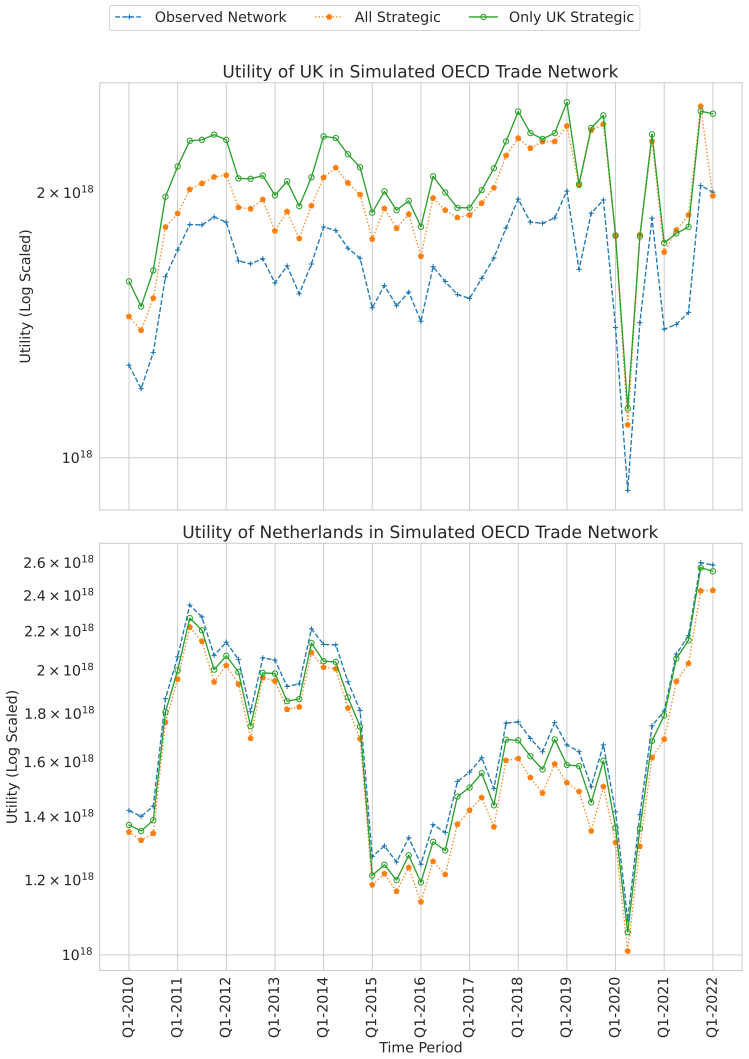


Figure 5.8: *Effect of strategic trading on the UK (top) and Netherlands (bottom):* The UK's utility is highest when it is the only strategic agent, and lowest when all are honest (the observed network). However, the pattern changes in the last quarter, where the utility when all are strategic is worse than when all are honest. The Netherlands has the highest utility when all others are honest, unlike the UK.

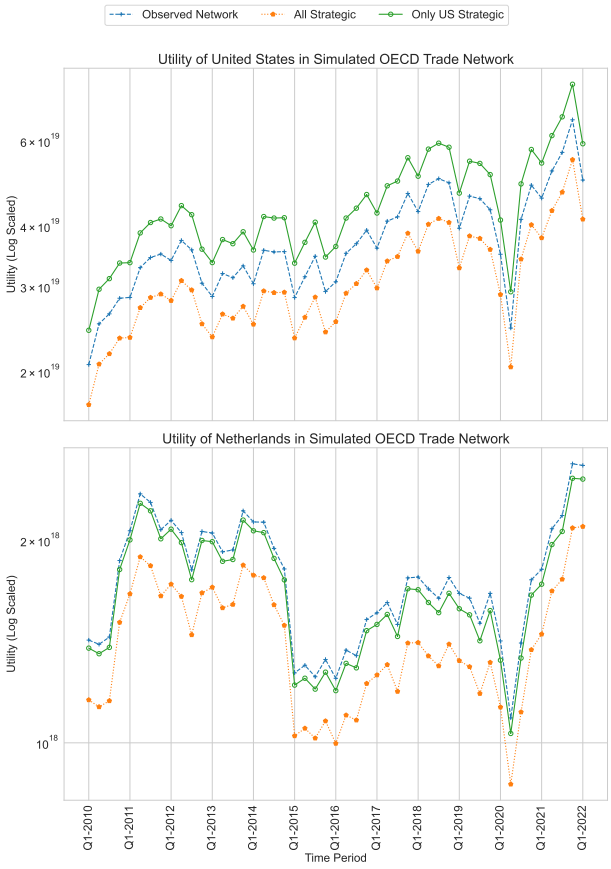


Figure 5.9: *Effect of strategic trading on the US (top) and Netherlands (bottom):* The US's utility is highest. The Netherlands has the highest utility when all others are honest, unlike the UK.

5.11 Experimental Details

5.11.1 Learning Experiments

In all learning experiments, we generate random X, B, Σ as follows. $X \in \mathbb{R}^{n \times d}$ has rows that are iid Dirichlet$(1/d, \ldots, 1/d)$. $B \in \mathbb{R}^{d \times d}$ is symmetric, with $B_{ij} \sim N(5, 1)$ for $i \geq j$ and $B_{ji} = B_{ij}$ for $i < j$. $\Sigma = XB_\Sigma X^T + \epsilon I_n$ for random $B_\Sigma \in \mathbb{R}^{d \times d}$ and $\epsilon = 10^{-3}$. The B_Σ is generated as $B_\Sigma = UDU^T$ for $U \in \mathbb{R}^{d \times d}$ having d random orthonormal columns, and D diagonal with iid Uniform$(1, \sqrt{n})$ entries on the diagonal.

Given the network setting (B, I, Σ, X) with some number of strategic agents $s \in [n]$, we generate random $S \subset [n]$ of size s, uniformly at random from all subsets of size s. Then, we generate negotiating positions M' based on Algorithm 4.

Our impelementation of spectral clustering is as follows. We compute the eigenvector

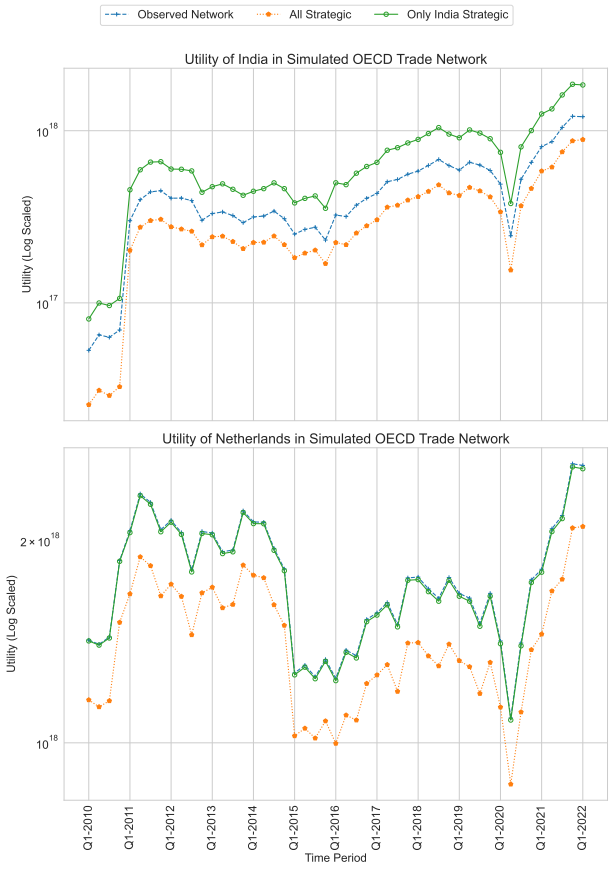


Figure 5.10: *Effect of strategic trading on India (top) and Netherlands (bottom):* India's utility is highest. The Netherlands has the highest utility when all others are honest, unlike the UK.

$\boldsymbol{v}_2 \in \mathbb{R}^n$ corresponding to the least nonzero eigenvalue λ_2 of R, and then assigns S_1, S_2 to the positive and negative indices of $\boldsymbol{v}_2$ respectively. We solve for s from β, n, and then return $\hat{S} = S_1$ if $|S_1| > s$ and $\hat{S} = S_2$ otherwise.

5.11.2 Negotiations on International Trade Networks

We use the same international trade dataset as Jalan et al. (2024a). Specifically, we use international trade statistics from the OECD to get quarterly measurements of bilateral trade between 46 large economies, including the top 15 world nations by GDP OECD (2022) Jalan et al. (2024a). The data are available at the OECD Statistics webpage (`https://stats.oecd.org/`). The data are measured quarterly from Q1 2010 to Q2 2022. We take the sum of trade flows $i \to j$ and $j \to i$. The diagonals $W_{ii} = 0$ for all i.

Given these measurements, which we denote as $W_t \in \mathbb{R}^{46\times 46}$ for $t = 1, 2, \ldots, 46$, we

solve for a covariance matrix $\Sigma \succ 0$ using the Semidefinite Programming algorithm of Jalan et al. (2024a). This Σ is fixed and used throughout the experiments of Section 5.6.

5.11.3 Compute Environment

All experiments were performed on a Linux machine with 48 cores running Ubunutu 20.04.4 OS. Each CPU core is an Intel(R) Xeon(R) CPU E5-2695 v2 (2.40GHz). The architecture was x86. Total RAM was 377 GiB.

The total time to generate experimental results was approximately 24 hours of wall time with parallelization. See our code submission for the exact scripts.

Chapter 6: Opinion Dynamics with Multiple Adversaries

6.1 Introduction

Over the past decade, social media has experienced rapid growth in both usage and significance. Online social networks, which allow users to share updates about their lives and opinions with a broad audience instantaneously, are now utilized by billions of people globally. These platforms serve various purposes, such as being informed about politics, news, health-related updates, products, and many more (Backstrom et al., 2012; Young, 2006; Banerjee et al., 2013; Shearer and Mitchell, 2021).

Unfortunately, networks can induce polarization, with the network connections serving as a pathway for social discord to increase Musco et al. (2018b); Chen and Rácz (2021b); Wang and Kleinberg (2024); Gaitonde et al. (2020a). This is a well-studied sociological phenomenon called the *filter-bubble theory* (Pariser, 2011). The filter-bubble theory argues that personalized algorithms used by online platforms, such as search engines and social media, selectively display content that aligns with a user's past behaviors, preferences, and beliefs. This customization creates an *"invisible algorithmic editing"* of the web, isolating individuals within their own ideological bubbles where they encounter only information that reinforces their existing views. As a result, users are less likely to be exposed to diverse perspectives, potentially narrowing their worldview and fostering polarization. Pariser (2011) warns that such bubbles undermine democratic discourse by limiting opportunities for individuals to engage with challenging or unfamiliar ideas.

Additionally, social networks can be manipulated by malicious entities in order to create discord and cause disagreement. For instance, the 2017 indictment of the Russian Internet Research Agency (IRA) by the U.S. Department of Justice Special Counsel's Office alleged that the IRA leveraged multiple social media accounts and targeted advertising to achieve *"a strategic goal to sow discord in the U.S. political system, including the 2016 U.S. presidential election"* (Mueller, 2018). In 2019, Twitter, Inc. (2019) disclosed that at least 936 accounts attempted to induce discord in Hong Kong, to e.g. hinder protesters' ability

The content of this chapter is under review at the 26th ACM Conference on Economics and Computation (ACM EC 2025), and can be cited as Jalan and Papachristou (2025).

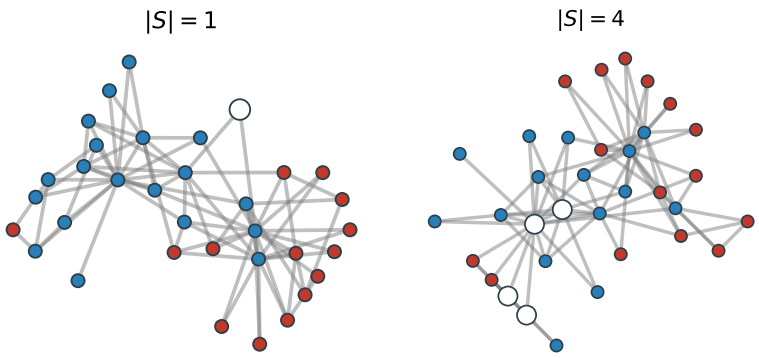


Figure 6.1: Visualization of the strategic equilibrium ($\boldsymbol{z}'$) on the Karate Club Graph for two different choices of S. The truthful intrinsic opinions have been taken to be $\boldsymbol{s} = \boldsymbol{u}_2$ where $\boldsymbol{u}_2$ is the Fiedler eigenvector of G. The white nodes correspond to the nodes in S. For the other nodes, the nodes colored in blue (resp. red) correspond to nodes whose public opinion $\boldsymbol{z}'_i$ increased (resp. decreased), i.e., $(\boldsymbol{z}'_i - \boldsymbol{z}_i)/\boldsymbol{z}_i \geq 0$ (resp. $(\boldsymbol{z}'_i - \boldsymbol{z}_i)/\boldsymbol{z}_i < 0$) after $\boldsymbol{s}'$ was chosen.

to organize effectively during the independence movement. As social media continues to proliferate, it is likely that these types of external interferences will become increasingly common. Additionally, networks of Facebook pages have targeted Americans with sports betting scams, amplifying their reach by disseminating provocative conspiracy theories about political figures and natural disasters (Bjork-James and Donovan, 2024). These schemes leverage the economics of the internet, where engagement with inflammatory content is monetized, and social media algorithms inadvertently amplify such content, enabling bad actors to exploit audiences for profit.

To model the opinions' evolution, computer scientists, sociologists, and statisticians have relied on the framework of *opinion dynamics* where the users' opinions coevolve according to a weighted network $G = (V, E, w)$, and each user updates their opinion as a combination of their own intrinsic opinion as well as the opinions of their neighbors (Friedkin and Johnsen, 1990). This model of opinion exchange has the advantage of taking into account both network interactions and their own intrinsic opinion. So far, all of the existing works consider a single actor who has the ability to act on the network to induce disagreement or polarization Musco et al. (2018b); Chen and Rácz (2021b); Wang and Kleinberg (2024); Ristache et al. (2024); Gaitonde et al. (2020a); Rácz and Rigobon (2023); Chitra and Musco (2020).

In this work, we lift the assumption of requiring a single actor (such as the platform) to act as an adversary to induce polarization or disagreement and consider the case of several

decentralized actors. It is known that empirically, a very small percentage (25%) of the users in a network need to disagree to sway consensus (Centola et al., 2018). Moreover, real-world social networks involve *multiple* malicious actors, who use different levels of manipulation and hate speech based on their individual goals (Bjork-James and Donovan, 2024). In this paper, we attempt to provide a theoretical basis for this phenomenon: specifically, in our setting, we assume that there is a set $S \subseteq V$ of strategic agents whose goal is to report false intrinsic opinions ($\boldsymbol{s}'$) that are different from their true intrinsic opinions ($\boldsymbol{s} \neq \boldsymbol{s}'$). Their goal is to influence others while not deviating much from their neighbors; namely, they want to reach an equilibrium where their neighbors agree with them.

For instance, assume a social network where a set of S of political actors want the network to believe that their stance on a topic (e.g., abortion, elections, drug legalization, etc.) is the best. They achieve this by adversarially reporting different intrinsic opinions. This ensures that their influence is both persuasive and credible within the local network context. Such adversarial behavior can result in significantly different (cf. Figure 6.1) and highly polarized equilibria, where the strategic agents' opinions appear dominant despite not reflecting the actual intrinsic views of the majority.

Our work investigates the conditions under which these strategic manipulations are successful, the extent of their impact on network-wide opinion dynamics, and how platforms can learn from observing these manipulated equilibria to mitigate such impacts.

6.1.1 Our Contributions

In this paper, we ask the following research question (RQ):

> ***(RQ)*** *What if a set of strategic actors with* ***possibly conflicting goals*** *tries to manipulate the consensus by strategically reporting beliefs different than their true beliefs?*

We rely on the Friedkin-Johnsen (FJ) model (Friedkin and Johnsen, 1990), where the opinions of agents coevolve via the help of a weighted undirected network $G = (V = [n], E, w)$ with non-negative weights. The intrinsic opinions are given by $\boldsymbol{s} \in \mathbb{R}^n$, where $\boldsymbol{s}_i \in \mathbb{R}$ is the intrinsic opinion of agent i. According to the FJ model, the agents possess intrinsic opinions $\boldsymbol{s}$ and express opinions $\boldsymbol{z} \in \mathbb{R}^n$, which they update via the following rule for each agent i:

$$\boldsymbol{z}_i(t+1) = \frac{\alpha_i \boldsymbol{s}_i + (1-\alpha_i)\sum_{i\sim j} w_{ij}\boldsymbol{z}_j(t)}{1+\sum_{i\sim j} w_{ij}}. \tag{6.1}$$

where $\alpha_i \in (0,1)$ is i's susceptibility to persuation (Abebe et al., 2018). The scalar $\boldsymbol{z}_i$ is the expressed opinion of agent i, which can be different from their intrinsic opinion $\boldsymbol{s}_i$.

We additionally define $\tilde{\alpha}_i = \alpha_i/(1-\alpha_i)$ to be the normalized susceptibility parameter corresponding to i. The update rule of Eq. (6.1) corresponds to the best-response dynamics arising from minimizing the quadratic cost function for each i (Bindel et al., 2011; Abebe et al., 2018):

$$c_i(\boldsymbol{z}_i, \boldsymbol{z}_{-i}) = (1-\alpha_i)\sum_{i\sim j} w_{ij}(\boldsymbol{z}_i - \boldsymbol{z}_j)^2 + \alpha_i(\boldsymbol{z}_i - \boldsymbol{s}_i)^2. \tag{6.2}$$

The Pure Strategy Nash Equilibrium (PSNE) can be written as $z = ((I-A)L + A)^{-1}A\boldsymbol{s} = B\boldsymbol{s}$ where L is the Laplacian of graph G, $A = \mathrm{diag}(\alpha_1, \ldots, \alpha_n)$ is the diagonal matrix of susceptibilities. When an external *single actor aims to induce disagreement or polarization* – see, e.g., Gaitonde et al. (2020a); Racz and Rigobon (2022); Musco et al. (2018b) – the adversary is tasked with optimizing the objective function

$$\sum_{i\in[n]} c_i(\boldsymbol{z}_i, \boldsymbol{z}_{-i}) = \boldsymbol{s}^T((I-A)L+A)^{-1}Af(L)((I-A)L+A)^{-1}A\boldsymbol{s},$$

where $f(L)$ is a function of the Laplacian of G, either with optimizing towards $\boldsymbol{s}$ (Gaitonde et al., 2020a), or the graph itself (Musco et al., 2018b; Racz and Rigobon, 2022).

Usually, as we also discussed earlier, many adverse actions on social networks come from *several independent strategic adversaries* who try to manipulate the network by infiltrating intrinsic opinions $\boldsymbol{s}'_i$, which are different from their true stances $\boldsymbol{s}_i$ but are simultaneously close to $\boldsymbol{s}_i$. Unlike previous works, these "adversaries" can have conflicting goals.

Concretely, the true opinions of the agents are $\boldsymbol{s}_1, ..., \boldsymbol{s}_n \in \mathbb{R}$, and there is a set S of deviating agents who report $\{\boldsymbol{s}'_i\}_{i\in S}$. The goal of the strategic agents is to minimize the cost function of Equation (6.2) at consensus $\boldsymbol{z}' = ((I-A)L+A)^{-1}A\boldsymbol{s}'$ where $\boldsymbol{s}'$ is the vector which has entries $\boldsymbol{s}_i$ for all $i \notin S$ and $\boldsymbol{s}'_i$ for all $i \in S$. The local optimization of agent i becomes:

$$\min_{s_i' \in \mathbb{R}} c_i \left(\boldsymbol{z}' = ((I - A)L + A)^{-1} A \boldsymbol{s}' \right). \tag{6.3}$$

Our contributions are as follows.

Characterizing Nash Equilibria with Multiple Adversaries. We give the Nash equilibrium of the game defined by Equation (6.3), and show that all Nash-optimal strategies are pure. The Pure Strategy Nash Equilibrium (PSNE) that is given by solving a constrained linear system. Given the PSNE of the game, we characterize the actors who can have the most influence in strategically manipulating the network.

Real-World Experiments to Understand Properites of Equilibria. We apply our framework to real-world social network data from Twitter and Reddit (Chitra and Musco, 2020), and data from the Political Blogs (Polblogs) dataset (Adamic and Glance, 2005). We find that the influence of strategic agents can be rather significant as they can significantly increase polarization and disagreement, as well as increase the overall "cost" of the consensus.

Analysis of Equilibrium Outcomes Under Different Sets of Strategic Actors. Various metrics for network polarization and disagreement are sensitive to the choice of *who* acts strategically, in nontrivial ways. For example, adding more strategic agents can sometimes *decrease* the Disagreement Ratio at equilibrium (Figure 6.5), due to counterbalancing effects. To address the effects of manipulation, we give worst-case upper bounds on the *Price of Misreporting* (PoM), which is analogous to well-studied Price of Anarchy bounds (see, for example, Bhawalkar et al. (2013); Roughgarden and Schoppmann (2011)), and suggest ways that the platform can be used to mitigate the effect of strategic behavior on their network.

Learning Algorithms for the Platform. We give an efficient algorithm for the platform to detect if manipulation has occurred (Algorithm 6), based on a hypothesis test with the publicly reported opinions $\boldsymbol{z}'$. Next, we give an algorithm to infer *who* manipulated the network (the set of strategic agents S) from $\boldsymbol{z}'$, as long as the size of S is sufficiently small. Our algorithm is inspired by the robust regression algorithm of Bhatia et al. (2015), and is practical for real-world networks. It *(i)* requires the platform to have access to node

embeddings X which have been shown computable even in billion-scale networks such as Twitter (El-Kishky et al., 2022), and *(ii)* can be computed in time $(n+m)^{O(1)}$, where n is the number of nodes and m is the number of edges of the network. Our algorithms have high accuracy on real-world datasets from Twitter, Reddit, and Polblogs.

6.1.2 Preliminaries and Notations

The Laplacian of the graph G is denoted by $L = D - W$ where W is the weight matrix of the graph, which has entries $w_{ij} \geq 0$, and D is the diagonal degree matrix with diagonal entries $D_{ii} = \sum_{i \sim j} w_{ij}$. The Laplacian has eigenvalues $0 = \lambda_1 \leq \lambda_2 \leq \cdots \leq \lambda_n$. For any undirected and connected graph G, L is symmetric and PSD, so we can write the eigendecomposition of L as:

$$L = \sum_{i \in [n]} \lambda_i \boldsymbol{u}_i \boldsymbol{u}_i^T \succeq 0, \tag{6.4}$$

where $\boldsymbol{u}_1, \ldots, \boldsymbol{u}_n$ are orthonormal eigenvectors. Moreover, $\boldsymbol{u}_1 = (1/\sqrt{n})\mathbf{1}$, where $\mathbf{1}$ is the column vector of all 1s. U denotes the matrix which has the eigenvectors of L as columns; i.e., such that $L = U^T \Lambda U$ where $\Lambda = \text{diag}(\lambda_1, \ldots, \lambda_n)$ is the diagonal matrix of L's eigenvalues. L_i denotes the i-restricted Laplacian which corresponds to the Laplacian of the graph with all edges that are non-adjacent to i being removed, and, similarly, $L_{\{u,v\}}$ corresponds to the Laplacian of an edge $\{u, v\}$. Note that $L_i = \sum_{i \sim j} L_{\{i,j\}}$. For a function $f(L)$ of the Laplacian we write $f(L) = U^T f(\Lambda) U$ where $f(\Lambda) = \text{diag}(f(\lambda_1), \ldots, f(\lambda_n))$. For brevity, regarding the equilibrium $\boldsymbol{z}$ of the FJ model, we write $B = ((I - A)L + A)^{-1} A$, such that $\boldsymbol{z} = B\boldsymbol{s}$ and $\boldsymbol{z}' = B\boldsymbol{s}'$.

We define the total cost of an equilibrium $\boldsymbol{z}$ to be

$$C(\boldsymbol{z}) = \sum_{i \in [n]} c_i(\boldsymbol{z}). \tag{6.5}$$

We define the platform-wide metrics to be

$$\text{Polarization Ratio} \qquad \mathcal{P}(\boldsymbol{z}) = \sum_{i \in [n]} (\boldsymbol{z}_i - \bar{z})^2, \text{ where } \bar{z} = \frac{1}{n} \sum_{i \in [n]} \boldsymbol{z}_i, \tag{6.6}$$

$$\text{Disagreement Ratio} \qquad \mathcal{D}(\boldsymbol{z}) = \sum_{i,j \in [n]} w_{ij} (\boldsymbol{z}_i - \boldsymbol{z}_j)^2 = \boldsymbol{z}^T L \boldsymbol{z}. \tag{6.7}$$

Finally, we define the *"Price of Misreporting"* (PoM), which is analogous to the Price of Anarchy Roughgarden (2005). The PoM is the ratio of the cost $C(\boldsymbol{z}')$ when the agents are deviating, and the cost $C(\boldsymbol{z})$ when the agents are reporting truthfully, i.e.,

$$\mathsf{PoM} := \frac{C(\boldsymbol{z}')}{C(\boldsymbol{z})}. \tag{6.8}$$

Unlike the Price of Anarchy (PoA), the equilibrium $\boldsymbol{z}$ in the denominator of Eq. (6.8) is the Nash equilibrium for the Friedkin-Johnson dynamics without manipulation. In the PoA, the denominator would be $C(\boldsymbol{z}^*)$, where $\boldsymbol{z}^*$ is a socially optimal equilibrium Bindel et al. (2011). Since we study strategic manipulations as a meta-game with respect to the base game of FJ dynamics, it is more relevant for us to compare $\boldsymbol{z}'$ with $\boldsymbol{z}$ than with $\boldsymbol{z}^*$. Note that $C(\boldsymbol{z}^*) \leq C(\boldsymbol{z}) \leq C(\boldsymbol{z}')$, so $\mathsf{PoA} \geq \mathsf{PoM} \geq 1$ always.

6.1.3 Related Work

Opinion Dynamics Opinion dynamics are well-studied in computer science and economics, as well as sociology, political science, and related fields. There have been many models proposed for opinion dynamics, such as with network interactions as we study in this paper (FJ model) (Friedkin and Johnsen, 1990; Bindel et al., 2015), bounded confidence dynamics (Hegselman-Krausse Model) (Hegselmann et al., 2002), coevolutionary dynamics (Bhawalkar et al., 2013) as well as many variants of them; see, for example Abebe et al. (2018); Hązła et al. (2019); Fotakis et al. (2016, 2023); Ristache et al. (2024). The work of (Bindel et al., 2011) shows bounds on the Price of Anarchy (PoA) between the PSNE and the welfare-optimal solution for the FJ model, and the subsequent work of Bhawalkar et al. (2013) shows PoA bounds for the coevolutionary dynamics. Additionally, the opinion dynamics have been modeled by the control community; see, for example, (Nedić and Touri, 2012; De Pasquale and Valcher, 2022; Bhattacharyya et al., 2013; Chazelle, 2011).

As in these works, we treat the FJ model as a basis. However, our work is significantly different as it studies a framework where any subset $S \subseteq [n]$ of strategic agents can *deviate* from their truthful intrinsic opinions, as opposed to studying the evolution of the expressed opinions and their PSNE in the FJ model. In our model, each strategic agent $i \in S$ can only choose a single entry $\boldsymbol{s}'_i$ of the overall deviation $\boldsymbol{s}'$, but pays a cost based on the resulting equilibrium ($\boldsymbol{z}' = B\boldsymbol{s}'$), which depends on the choices of other members of S.

Disagreement and Polarization in Social Networks Motivated by real-world manipulation of social networks in, e.g., the 2016 US election, a recent line of work studies polarization and strategic behavior in opinion dynamics Gaitonde et al. (2020a, 2021); Chen and Rácz (2021a); Wang and Kleinberg (2024); Ristache et al. (2024, 2025). Chen and Rácz (2021a) consider a model in which an adversary can control $k \leq n$ nodes' internal opinions and seeks to maximize polarization at equilibrium. Similarly, Gaitonde et al. (2020a) considers a single adversary who can modify intrinsic opinions $\boldsymbol{s}$ belonging to an ℓ_2-ball. More recent work also studies modification of agents' susceptibility parameters α_i to alter the median opinion at equilibrium Ristache et al. (2025). By contrast, we study a setting in which any subset $S \subseteq [n]$ can be strategic. Unlike previous works, these "adversaries" can have conflicting goals in our model.

Manipulation of Dynamic Games. Opinion dynamics are a widely studied instance of a network game, which is a game played by nodes in a network with payoffs depending on the actions of their neighbors Kearns et al. (2001); Tardos (2004). In addition to the manipulation of opinion dynamics, researchers have studied strategic manipulation of financial network formation Jalan and Chakrabarti (2024). In the non-network setting, researchers have studied the manipulation of recommendation systmes from a game-theoretic perspective Ben-Porat and Tennenholtz (2018), as well as security games Nguyen et al. (2019), repeated auctions Kolumbus and Nisan (2022b) and Fisher markers with linear utilities Kolumbus et al. (2023).

Learning from Strategic Data. We develop learning algorithms which observe the (possibly manipulated) equilibrium $\boldsymbol{z}'$ to detect if manipulation occurred, and if so who was responsible. The former problem relates to anomaly detection in networks. Chen and Tsourakakis (2022) develop a hypothesis test to detect such fraud in financial transaction neteworks, by testing if certain subgraphs deviate from Benford's Law. Similarly, Agarwal et al. (2020) propose a framework based on a χ^2-statistic to perform graph similarity search.

The problem of recovering the set of deviators relates to the broader literature of learning from observations of network games. Most works give learning algorithms for games *without* manipulation Irfan and Ortiz (2014); Garg and Jaakkola (2016); De et al. (2016); Leng et al. (2020a); Rossi et al. (2022); Jalan et al. (2024a). But our data $\boldsymbol{z}'$ can be a

manipulated equilibrium, which is a *strategic sources* of data Zampetakis (2020). Learning algorithms for strategic sources are known for certain settings such as linear classifiers with small-deviation assumptions (Chen et al., 2020a), or binary classifiers in a linear reward model (Harris et al., 2023). When agents can modify their features to fool a known algorithm, even strategy-robust classifiers such as Hardt et al. (2016) can be inaccurate Ghalme et al. (2021). Since agents can deviate arbitrarily in our model, we use a robust regression method with guarantees against *adversarial* corruptions (Bhatia et al., 2015), similar to the learning algorithms in (Kapoor et al., 2019; Russo, 2023). The work of Jalan and Chakrabarti (2024) studies learning from financial networks with strategic manipulations, which is in a similar spirit to our work but differs significantly in the application domain and context.

6.1.4 Real-world Datasets

To support our results, we use data grounded in practice, which have also been used in previous studies to study polarization and disagreement (cf. Musco et al. (2018a); Chitra and Musco (2020); Wang and Kleinberg (2024); Adamic and Glance (2005)). Specifically, we use Twitter, Reddit, and Political blog networks, summarized in Table 6.1 summarizes these. Both the Twitter and Reddit datasets are due to Chitra and Musco (2020). The vectors $\boldsymbol{s}$ of initial opinions for both are obtained via sentiment analysis and also follow the post-processing of Wang and Kleinberg (2024).

(1) Twitter dataset. These data correspond to debate over the Delhi legislative assembly elections of 2013. Nodes are Twitter users, and edges refer to user interactions.

(2) Reddit dataset. These data correspond to political discussion on the `r/politics` subreddit. Nodes are who posted in the `r/politics` subreddit, and there is an edge between two users i, j if two subreddits (other than `r/politics`) exist that both i, j posted on during the given time period.

(3) Political Blogs (Polblogs) dataset. These data, due to Adamic and Glance (2005), contain opinions from political blogs (liberal and conservative). Edges between blogs were automatically extracted from a crawl of the front page of the blog. Each blog is either liberal, where we assign a value $\boldsymbol{s}_i = -1$, or conservative, where we assign $\boldsymbol{s}_i = +1$.

Network	Nodes (n)	Edges (m)	Description
Twitter	548	3,638	User interactions during 2013 Delhi elections.
Reddit	556	8,969	User interactions in r/politics subreddit
Polblogs	1,490	16,178	Liberal and conservative blog network

Table 6.1: Summary of the social network datasets we use.

6.2 Strategic Opinion Formation

The opinion formation game has two phases. First, strategic agents privately choose a strategic intrinsic opinion according to Equation (6.3). Second, agents exchange opinions and reach consensus *as if* they were in the Friedkin-Johnson dynamics, except the strategic opinions are used in place of the true intrinsic opinions.

1. *Strategy Phase.* Each strategic agent $i \in S$ independently and privately chooses a fictitious strategic opinion $\boldsymbol{s}'_i \in \mathbb{R}$. For honest agents ($i \notin S$) we have $\boldsymbol{s}'_i = \boldsymbol{s}_i$.

2. *Opinion Formation Phase.* Reach equilibrium $\boldsymbol{z}' = B\boldsymbol{s}'$ as if $\boldsymbol{s}'$ were the true intrinsic opinions $\boldsymbol{s}$.

The network G and the true beliefs $\boldsymbol{s}$ determine each agent's utility. We pose the following problem:

Definition 6.2.1 (Instrinsic belief lying problem.)**.** *Let $S \subseteq [n]$ be a set of strategic agents. If agent $i \in S$ wants network members to express opinions close to $\boldsymbol{s}_i$, what choice of $\boldsymbol{s}'_i$ is optimal and minimizes the cost function of Equation* (6.3)*?*

The following theorem characterizes the Nash Equilibria of the Intrinsic Belief Lying Problem.

Theorem 6.2.2 (Nash Equilibrium)**.** *Let $\mathbb{T}_i = (1-\alpha_i)(B^T L_i B) + \alpha_i(B^T \boldsymbol{e}_i \boldsymbol{e}_i^T B) \in \mathbb{R}^{n\times n}$ and $\boldsymbol{y}_i = \alpha_i B_{ii} \boldsymbol{s}_i$. The Nash equilibria, if any exist, are given by solutions $s' \in \mathbb{R}^n$ to the following constrained linear system:*

$$\forall i \in S : \boldsymbol{e}_i^T \mathbb{T}_i \boldsymbol{s}' = \boldsymbol{y}_i,$$
$$\forall j \notin S : \boldsymbol{s}'_j = \boldsymbol{s}_j.$$

To illustrate the Theorem, we consider a toy example.

Example 6.2.3 (Two-Node Graph)**.** *Consider a graph with 2 nodes and one edge with weight $w > 0$. We set $\alpha_1 = \alpha_2 = 0.5$ for simplicity. Suppose that both agents deviate, i.e., $S = [2]$. Then, we can calculate B to be*

$$B = \frac{1}{2w+1}\begin{pmatrix} w+1 & w \\ w & w+1 \end{pmatrix}$$

and

$$\boldsymbol{z}'_0 = \frac{(w+1)\boldsymbol{s}'_0 + w\boldsymbol{s}'_1}{2w+1}, \quad \boldsymbol{z}'_1 = \frac{w\boldsymbol{s}'_0 + (w+1)\boldsymbol{s}'_1}{2w+1}, \tag{6.9}$$

yielding the two cost functions

$$c_0(\boldsymbol{s}') = \frac{1}{2}w\left(\frac{\boldsymbol{s}'_0 - \boldsymbol{s}'_1}{2w+1}\right)^2 + \frac{1}{2}\left(\frac{(w+1)\boldsymbol{s}'_0 + w\boldsymbol{s}'_1}{2w+1} - \boldsymbol{s}_0\right)^2$$
$$c_1(\boldsymbol{s}') = \frac{1}{2}w\left(\frac{\boldsymbol{s}'_0 - \boldsymbol{s}'_1}{2w+1}\right)^2 + \frac{1}{2}\left(\frac{(w+1)\boldsymbol{s}'_1 + w\boldsymbol{s}'_0}{2w+1} - \boldsymbol{s}_1\right)^2.$$

Taking the first order conditions $\frac{\partial c_0}{\partial \boldsymbol{s}'_0} = 0$ and $\frac{\partial c_1}{\partial \boldsymbol{s}'_1} = 0$ we get a linear system whose solutions are:

$$\boldsymbol{s}'_0 = \frac{w^2(s_0 - s_1) + (3w+1)\boldsymbol{s}_0}{3w+1}, \quad \boldsymbol{s}'_1 = \frac{w^2(\boldsymbol{s}_0 - \boldsymbol{s}_1) + (3w+1)s_1}{3w+1}.$$

Replacing these values back to the costs we get that

$$\forall i : c_i(s'_0, s'_1) = \frac{1}{2}\frac{w(w^2+3w-1)(\boldsymbol{s}_0 - \boldsymbol{s}_1)^2}{9w^2+6w+1},$$

On the other hand, if all agents are honest, then the cost for each is:

$$\forall i : c_i(\boldsymbol{s}_0, \boldsymbol{s}_1) = \frac{1}{2}\frac{w(w+1)(\boldsymbol{s}_0 - \boldsymbol{s}_1)^2}{(2w+1)^2}.$$

and the ratio of the two costs is at least $\max\{1, w/3\}$.

Next, we discuss some consequences of Theorem 6.2.2. First, we characterize $\boldsymbol{s}'$ as the solution to a linear system.

Corollary 6.2.4. *Let $T \in \mathbb{R}^{|S|\times n}$ have rows $\{\boldsymbol{e}_i^T \mathbb{T}_i\}_{i\in S}$ given by Theorem 6.2.2. Let $\tilde{T} \in \mathbb{R}^{|S|\times|S|}$ be the submatrix of T selecting columns belonging to S. Let $\boldsymbol{y} \in \mathbb{R}^{|S|}$ have entries $\boldsymbol{y}_i = \alpha_i B_{ii} \boldsymbol{s}_i$ as above. Let $\tilde{\boldsymbol{y}} = \boldsymbol{y} - \sum_{j\notin S} \boldsymbol{s}_j T \boldsymbol{e}_j$. Then the set of Nash equilibria, if any exist, are given by the solutions to the unconstrained linear system*

$$\tilde{T}\boldsymbol{x} = \tilde{\boldsymbol{y}}. \tag{6.10}$$

The resulting opinions vector $\boldsymbol{s}'$ is given by $\boldsymbol{s}'_i = \boldsymbol{x}_i$ if $i \in S$ and $\boldsymbol{s}'_i = \boldsymbol{s}_i$ otherwise.

Thus, in a Nash equilibrium, every strategic agent solves their corresponding equation given by Equation (6.10). The explicit characterization of equilibria also implies that Nash equilibria cannot be mixed.

Corollary 6.2.5 (Pure Strategy Nash Equilibria)**.** *The Nash equilibrium corresponds to solving the system of $|S|$ linear equations in the scalars $\{\boldsymbol{s}'_i | i \in S\}$ given by Equation* (6.10)*. Also, all Nash equilibria are pure-strategy Nash equilibria.*

Optimal Deviation for One Agent and All Agents Assuming that we have one strategic agent, what is the change in their opinion? We can show that the new opinion is a scalar multiple of the initial opinion plus a bias term, where neither the scalar multiple nor the bias term can be zero.

Corollary 6.2.6 (Deviation for One Agent)**.** *Let $S = \{i\}$. Then, $\boldsymbol{s}'_i = \theta_i \boldsymbol{s}_i + \beta_i$ where*

$$\theta_i = \frac{\alpha_i B_{ii}}{(1-\alpha_i)\sum_{i\sim j} w_{ij}(B_{ii} - B_{ij})^2 + \alpha_i B_{ii}^2} > 0,$$
$$\beta_i = -\frac{\alpha_i \sum_{j\neq i} B_{ij} s_j}{(1-\alpha_i)\sum_{i\sim j} w_{ij}(B_{ii} - B_{ij})^2 + \alpha_i B_{ii}^2}.$$

Similarly, we can relate the maximum deviation of $\boldsymbol{s}'$ from $\boldsymbol{s}$ in the other extreme case, i.e., when all agents are deviating ($S = [n]$).

Corollary 6.2.7. *When all agents are deviating ($S = [n]$), and $\alpha_i = \alpha$, then $\boldsymbol{s}'$ satisfies:*

$$\frac{\|\boldsymbol{s}'\|_2}{\|\boldsymbol{s}\|_2} \leq \frac{\lambda_n + \tilde{\alpha}}{\tilde{\alpha}}.$$

The proof of Corollary 6.2.7 shows that the adjusted susceptibility ($\tilde{\alpha}$) and the maximum eigenvalue of the Laplacian (λ_n) are responsible for changes in the norm of $\boldsymbol{s}'$. From classic spectral graph theory, we know that $\lambda_n = \Theta(d_{\max})$ where $d_{\max}$ is the maximum degree of the graph; therefore, graphs with a lower maximum degree experience smaller distortions. Also, regarding the susceptibility to persuasion, the distortion becomes $1 + o(1)$ as long as $\tilde{\alpha} = \omega(d_{\max})$.

Equilibria for real-world datasets. Next, we discuss the results of experiments simulating the strategically manipulated equilibria for our real-world datasets.

Effect of Susceptibility to Persuasion in Real-world Data Regarding real-world data, Figure 6.2 shows the relationship between the truthful opinions ($\boldsymbol{s}$ and $\boldsymbol{z}$) and the strategic ones ($\boldsymbol{s}'$ and $\boldsymbol{z}'$) for the datasets, along with the corresponding correlation coefficient R^2, assuming that S consists of the top-50% nodes in terms of their eigenvector centrality, for susceptibility parameters set to $\alpha_i = 0.5$ (equal self-persuasion and persuasion due to others) and $\alpha_i = 0.25$ (higher persuasion due to others).

Regarding the public opinions, even though in the Reddit dataset, the strategic opinions seem to be correlated with the truthful ones ($R^2 = 0.78$ for $\alpha_i = 0.25$ and $R^2 = 0.94$ for $\alpha_i = 0.5$ respectively), in the Twitter dataset, we do not get the same result (i.e., $R^2 < 0.25$). Finally, in the Polblogs dataset, the situation is somewhere in the middle; when $\alpha_i = 0.25$ we get a low R^2 ($R^2 = 0.18$) where for $\alpha_i = 0.5$ we get a high R^2 ($R^2 = 0.74$). Additionally, in all cases except Twitter, we get that the effect is significant ($P < 0.01$).

Regarding the relationship between the intrinsic opinions, we do not detect any significant effect in most cases except Reddit with $\alpha_i = 0.5$ ($P < 0.01$) and Twitter with $\alpha_i = 0.5$ ($P < 0.05$).

Asymmetric Effects of Strategic Behavior on Liberals and Conservatives. Figure 6.3 analyzes the opinions of the strategic set S on the Polblogs dataset. Specifically, we find that larger changes in sentiment happen across liberal outlets compared to conservative

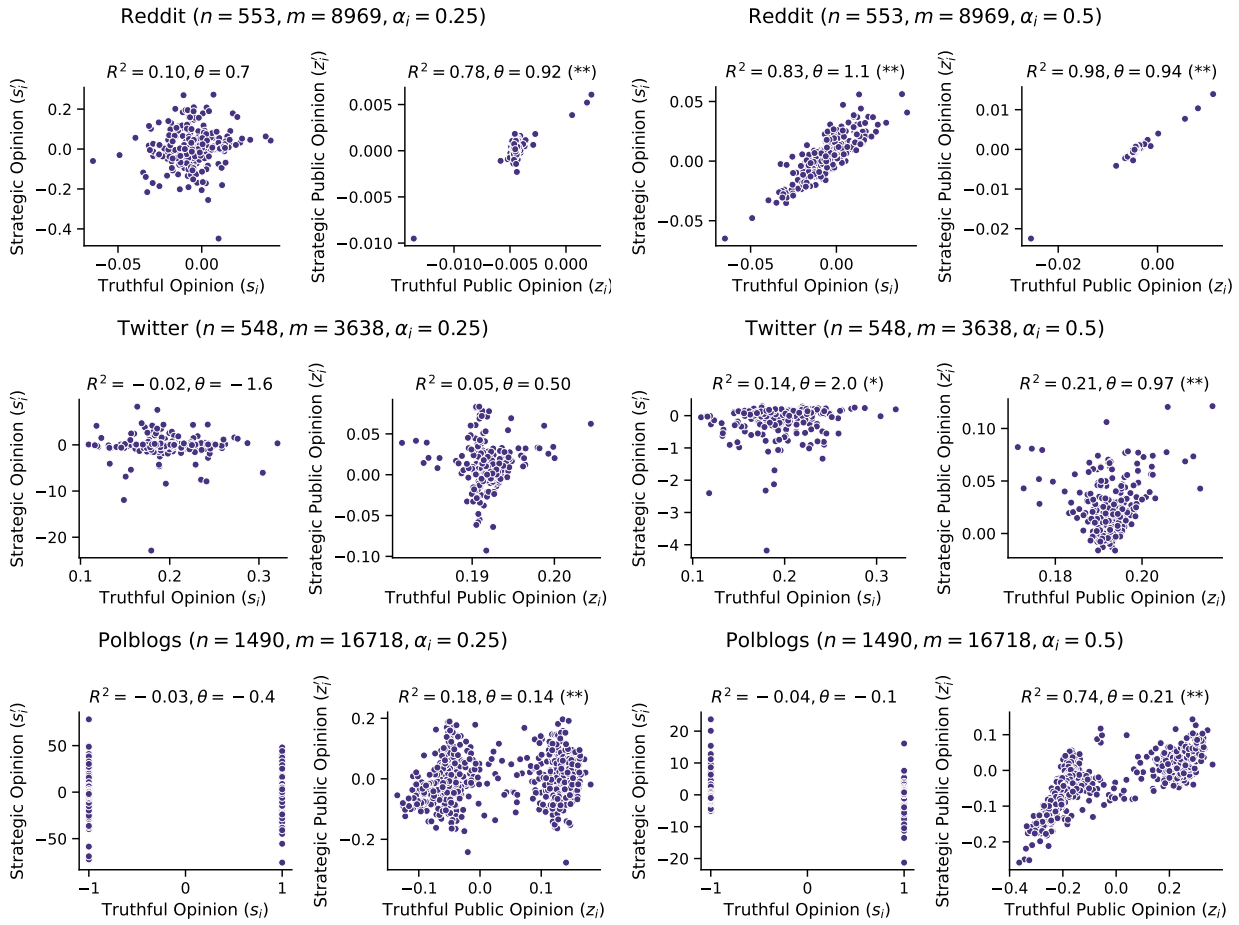


Figure 6.2: Plot of truthful intrinsic opinions (s) and strategic opinions (s'), and truthful public opinions (z) compared to the strategic public opinions (z') for the nodes belonging to S. S is taken to be the top-50% in terms of their eigenvector centrality. In both cases we have taken $\alpha_i \in \{0.25, 0.5\}$ for all nodes. We fit a linear regression between s' and s (resp. between z and z'). We report the effect size θ which corresponds to the slope of the linear regression and the P-value with respect to the null hypothesis ($\theta = 0$). *** stands for $P < 0.001$, ** stands for $P < 0.01$ and * stands for $P < 0.05$.

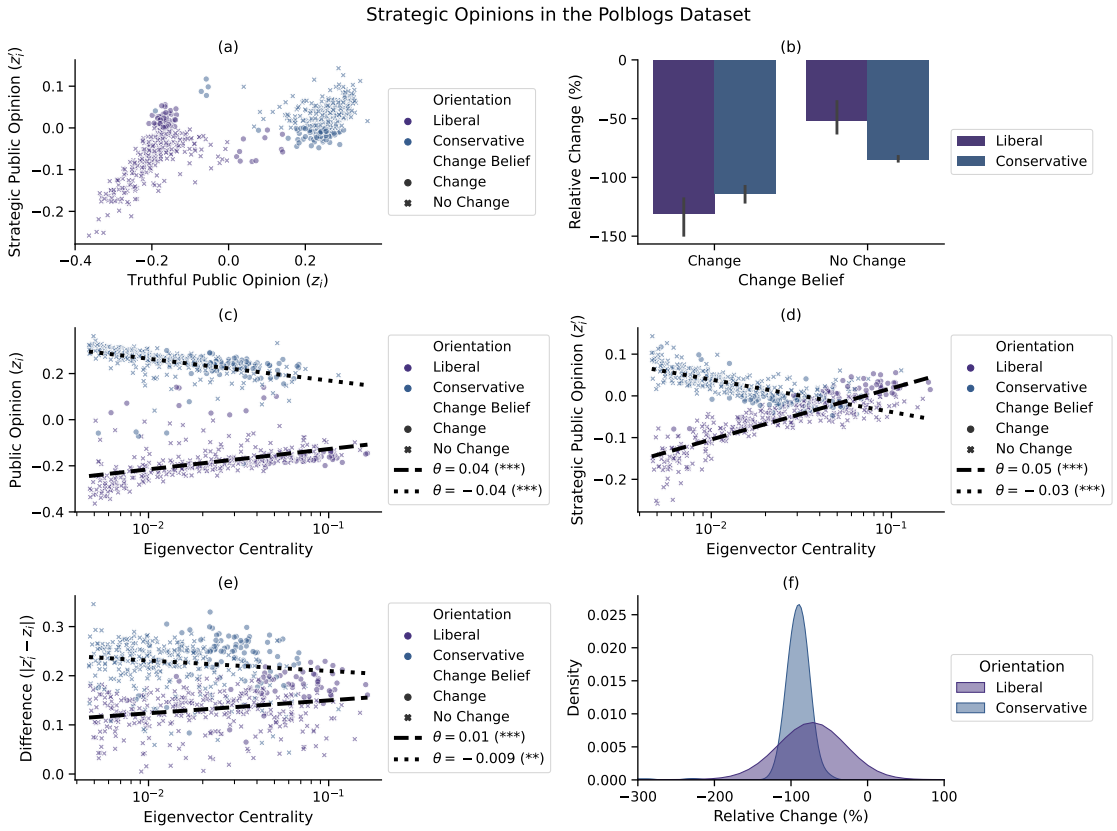


Figure 6.3: Strategic misreports for the Polblogs dataset where S is taken to be the top-50% of the agents in terms of their eigenvector centralities. The nodes are labeled either as liberal ($\boldsymbol{s}_i = -1$) or conservative ($\boldsymbol{s}_i = +1$), and we consider the nodes that change their beliefs as the nodes for which $\boldsymbol{z}'_i$ and $\boldsymbol{z}_i$ do not have the same sign. In the scatterplots (a), (c), (d), (e), the shape of each point indicates whether that user changed belief or not, and the color indicates their true (intrinsic) opinion. Overall, we discover a higher amount of change among liberal blogs compared to conservative ones (panel (b)). Additionally, we report the truthful/strategic public opinion as a function of the logarithm of the eigenvector centrality $\boldsymbol{\pi}_i$ (cf. panels (c, d)) for each node, as well as the absolute change $|\boldsymbol{z}'_i - \boldsymbol{z}_i|$ (cf. panel (e)). We fit a regression model, and we detect significant effects ($^{***}: P < 0.001$, $^{**}: P < 0.01$, $^{*}: P < 0.05$; effects denoted by θ) of the logarithm of the centrality to the truthful equilibrium $\boldsymbol{z}$, the strategic equilibrium $\boldsymbol{z}'$, and the change $|\boldsymbol{z}' - \boldsymbol{z}|$, revealing the structure of a power law. Finally, we observe that relative changes are more dispersed along liberal sources compared to conservative sources (c.f. panel F.

ones. Additionally, the changes in the truthful/strategic opinions are related to the eigenvector centrality $\boldsymbol{\pi}_i$ as a power law, i.e., $\boldsymbol{z}'_i \propto \boldsymbol{\pi}_i^{\theta}$ ($P < 0.001$; linear regression between the log centralities $\log \boldsymbol{\pi}_i$ and $\boldsymbol{z}'_i$). The same finding holds for $|\boldsymbol{z}'_i - \boldsymbol{z}_i|$ and $\boldsymbol{z}_i$.

At this point, one may wonder whether the eigenvector centrality really influences the strategic opinions $\boldsymbol{z}'_i$ for $i \in S$. Our answer is negative. We repeat the same experiment with the Twitter and Reddit datasets, where we find no effects ($P > 0.1$; linear regression between the log centralities $\log \boldsymbol{\pi}_i$ and $\boldsymbol{z}'_i$). Due to space limitations, the corresponding figures are deferred to Section 6.7.

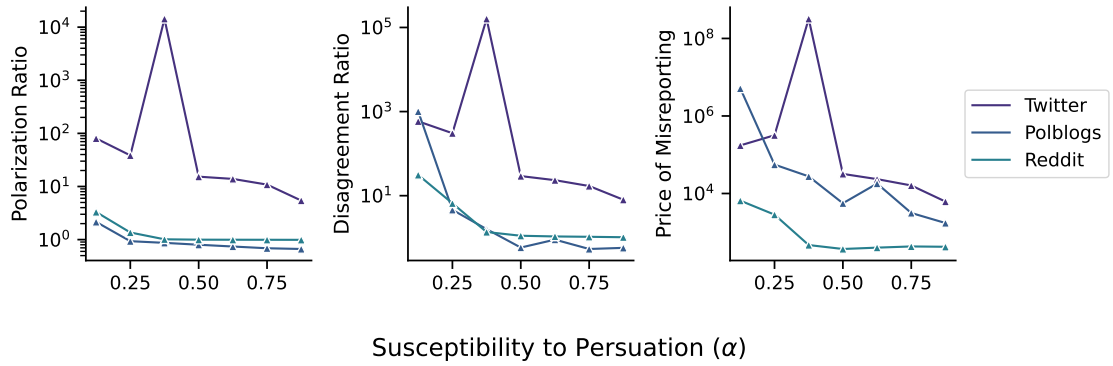


Figure 6.4: Polarization ratio ($\mathcal{P}(z')/\mathcal{P}(z)$), disagreement ratio ($\mathcal{D}(z')/\mathcal{D}(z)$), and price of misreporting ($C(z')/C(z)$) for the three datasets for varying susceptibility to persuasion values. We have set all susceptibilities α_i to the same value α. The Twitter dataset has the largest variation in all three ratios compared to the others. S is taken to be the top-50% nodes in terms of their eigenvector centrality.

Polarization and Disagreement. Figure 6.4 shows how the polarization, disagreement, and cost change as a function of the susceptibility parameter α_i. Except for $\alpha_i \approx 0.3$, the polarization ratio, disagreement ratio, and the price of misreporting experience a downward trend as α_i increases. This indicates that as as users prioritize their own opinions more than their neighbors, they are less susceptible to strategic manipulation.

Effect of the number of deviators ($|S|$) Next, we study the effect of the number of deviators, which corresponds to $|S|$, on the changes in polarization, disagreement, and the total cost (through the price of misreporting). Figure 6.5 shows how the polarization and disagreement when S consists of the top-1-10% most central agents with respect to eigenvector

centrality. We show that even if only 1% of agents are strategic, this can impact consensus by several orders of magnitude.

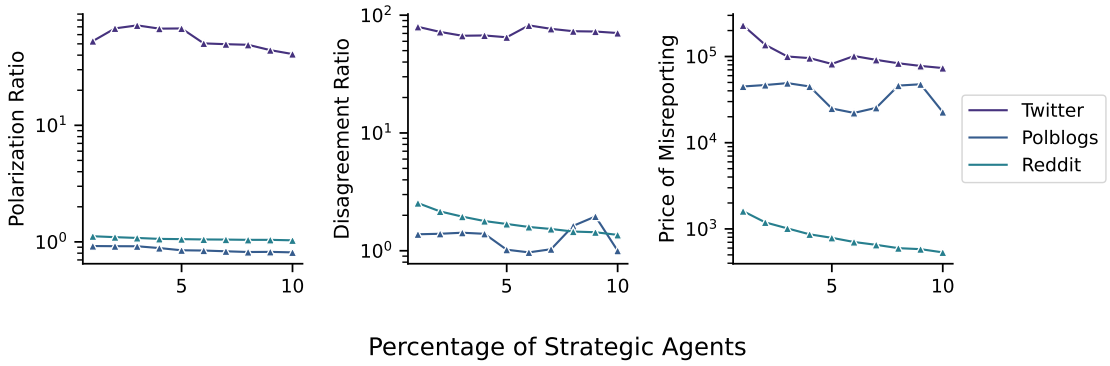


Figure 6.5: Polarization ratio ($\mathcal{P}(\boldsymbol{z}')/\mathcal{P}(\boldsymbol{z})$), disagreement ratio ($\mathcal{D}(\boldsymbol{z}')/\mathcal{D}(\boldsymbol{z})$), and price of misreporting ($C(\boldsymbol{z}')/C(\boldsymbol{z})$) for the three datasets for varying the size of $|S|$. The size of $|S|$ corresponds to the top p percent of the actors ($|S| = \lceil pn \rceil$) based on their eigenvector centrality (in decreasing order), for $p \in [0.01, 0.1]$. The susceptibility parameter is set to $\alpha_i = 0.5$.

6.3 Price of Misreporting

In Section 6.2, we saw that strategic manipulation can substantially affect network outcomes via the Polarization Ratio and Disruption Ratio. We now give an upper bound for the Price of Misreporting (Eq. (6.8)), which is the analogue of the Price of Anarchy in our setting. The PoM measures the total cost paid by agents under the corrupted equilibrium $\boldsymbol{z}'$, versus the total cost under the non-corrupted $\boldsymbol{z}$. Since the cost captures an agent's deviation from her *truthful* intrinsic opinion as well as her deviation from the expressed opinions of her neighbors, it is a natural measure of the network's discord at equilibrium.

Theorem 6.3.1 shows that the PoM is small when the spectral radius of the Laplacian is small, and when agents are somewhat susceptible to their neigbhors ($\alpha \not\to 0$). Note that the spectral radius can be replaced by a degree bound: if $d_{\max}$ is the maximum degree of the graph, then $\lambda_n \leq 2d_{\max}$. So the PoM is small if the maximum degree is small.

Theorem 6.3.1. *Suppose all agents deviate ($S = [n]$) and there exists α such that $\alpha_i = \alpha$ for all i. Let $\tilde{\alpha} = \alpha/(1-\alpha)$, and λ_n be the spectral radius of the Laplacian. Then the price*

of misreporting is bounded as:

$$\mathsf{PoM} \leq \frac{(\lambda_n + 4\tilde{\alpha})(\lambda_n + \tilde{\alpha})^2}{\tilde{\alpha}^5} = O\left(\max\left\{\frac{\lambda_n}{\tilde{\alpha}^5}, \frac{1}{\tilde{\alpha}^2}\right\}\right).$$

From Theorem 6.3.1, we can show that the upper bound is minimized when $\lambda_n = \Theta(\tilde{\alpha}^3)$ and has a value of $O(1/\tilde{\alpha}^2)$. As we noted, Theorem 6.3.1 can be written with $d_{\max}$ in the place of λ_n as well.

Next, we give an easy generalization to the case of differing susceptibility.

Corollary 6.3.2 (Price of Misreporting for Heterogeneous Susceptibility)**.** *If the α_i are differing, let $\alpha_{\min} = \min_i \alpha_i$ and $\alpha_{\max} = \max_j \alpha_j$. Define $\tilde{\alpha}_{\min} = \frac{\alpha_{\min}}{1-\alpha_{\max}}, \tilde{\alpha}_{\max} = \frac{\alpha_{\max}}{1-\alpha_{\min}}$. The Price of Misreporting is bounded as:*

$$\mathsf{PoM} \leq \frac{1-\alpha_{\min}}{1-\alpha_{\max}} \frac{(\lambda_n + 4\tilde{\alpha}_{\max})(\lambda_n + \tilde{\alpha}_{\max})^2}{\tilde{\alpha}_{\min}}.$$

Finally, we discuss how one may generalize Theorem 6.3.1 to the case where some agents are honest.

Towards fine-grained PoM guarantees. Figure 6.4 shows that the PoM is *not* monotonic in $|S|$. As the number of strategic agents grows, the PoM can fall or grow, depending on the choice of S, network parameters, and so on. Therefore, we would like to give a version of Theorem 6.3.1 for *any* set of strategic agents $S \subset [n]$, not just the case of $S = [n]$. However, proving such a bound would require analyzing $S \times S$ principal submatrices of B, L to obtain characterizations of the cost at the corrupted equilibrium $\boldsymbol{z}'$. In particular, we would require a *restricted invertibility* estimate to prove the analogue of Eq. (6.14). To our knowledge, the best such estimates (Marcus et al., 2022) are too lossy when $n - |S|$ is large. We leave this question to future work.

6.4 Learning from Network Outcomes

To mitigate the effects of strategic behavior, a platform must understand whether manipulation has occurred, and who the strategic actors are. In this section, we give

Algorithm 6 Learning from Misreporting Equilibrium with Hypothesis Testing

Input: Estimated graph information $\widehat{L}$ and $\widehat{A}$, observed equilibrium $\boldsymbol{z}'$
Output: "Manipulation" or "No Manipulation"
Observe corrupted equilibrium $\boldsymbol{z}'$.
Solve for $\widehat{\boldsymbol{s}'}$ (Eq. (6.11)).
Perform one-sample t test on the entries of $\widehat{\boldsymbol{s}'}$ with a population mean μ_0 under the null hypothesis.
return *If the t-test rejects, return "Manipulation." Otherwise, return "No Manipulation."*

computationally efficient methods to do so based on knowledge of the network edges and observing the corrupted equilibrium $\boldsymbol{z}'$. The latter can be found, for example, by performing sentiment analysis on the users' posts.

6.4.1 Detecting Manipulation with a Hypothesis Test

In many real-world networks, the distribution of truthful opinions follows a Gaussian distribution (Figure 6.6). Given estimates $(\widehat{L}, \widehat{A})$ for the graph Laplacian and susceptibility matrix, the platform can observe the corrupted equilibrium $\boldsymbol{z}'$ and solve for the strategic opinions $\boldsymbol{s}'$ via:

$$\widehat{\boldsymbol{s}'} := \widehat{A}^{-1}((I - \widehat{A})\widehat{L} + \widehat{A})\boldsymbol{z}'. \tag{6.11}$$

We propose that the platform perform a one-sample t-test with the entries $\boldsymbol{s}'$, with a population mean $\mu_0 \in \mathbb{R}$ based on e.g. historical data. Under the null hypothesis in which no manipulation has occurred, $\boldsymbol{s}' = \boldsymbol{s}$, so the test should fail to reject the null hypothesis. However, when agents $S \subset [n]$ deviate, then $\boldsymbol{s}'_i \neq \boldsymbol{s}_i$ for $i \in S$, so the test should reject the null for large enough deviations. The test is simple, and described in Algorithm 6. Figure 6.6 shows the results of the test for varying choices of S. We see that at significance level 0.05, the test has low Type I error, as it will return "No Manipulation" when $S = \emptyset$, and low Type II error as it will return "Manipulation" when $S \neq \emptyset$.

For the Political Blogs dataset, intrinsic opinions belong to $\{\pm 1\}$, so the null hypothesis should be a biased Rademacher distribution. In this case, one should use a χ^2-test, as in Agarwal et al. (2020); Chen and Tsourakakis (2022).

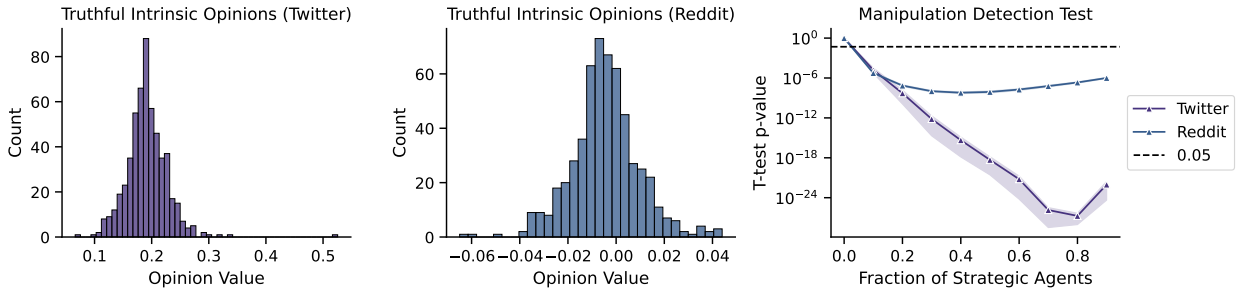


Figure 6.6: The true opinions for Twitter (left) and Reddit (middle) both follow a normal distribution. When simulating strategic manipulation with random choices of S (right), the detection test (Algorithm 6) has no Type I or Type II error at significance level 0.05. The tester uses $\widehat{L} = L, \widehat{A} = A = \frac{1}{2}I$, and μ_0 equal to the mean of the true intrinsic opinions. Shaded regions are 95% confidence intervals for p-values of the test across 5 independent runs.

6.4.2 Learning the Strategic Actors with Robust Regression

The algorithm described in the previous section (Algorithm 6) can be used to detect whether there exists manipulation in the network. However, the set S is unknown, and therefore the platform cannot target the deviators to perform interventions to mitigate strategic behavior.

It is, therefore, essential for the platform to be able to identify the set of deviators S, in case the platform needs to take regulatory actions. While at first, it may seem that finding the set of deviators S is a hard task, it turns out that under mild assumptions on the intrinsic opinion formation process, we can learn the set of deviators S from observing the strategically corrupted equilibrium $\boldsymbol{z}'$ via Algorithm 7 in polynomial time, described in Algorithm 7. Our algorithm is based on robust regression leveraging the TORRENT algorithm developed by Bhatia et al. (2015) and requires access to a node embedding matrix $X \in \mathbb{R}^{n \times d}$, and the size $|S|$ of the set of deviators.

The key idea of Algorithm 7 is that if the size of the strategic set S is sufficiently small, in general, $|S| \leq Cn$ for some small constant C, then we can view the misreported intrinsic opinions $\boldsymbol{s}'$ as a *perturbation* of the truthful opinion vector $\boldsymbol{s}$, and then use a robust regression algorithm to estimate $\boldsymbol{s}$. We assume that the embedding matrix $X \in \mathbb{R}^{n \times d}$ determines intrinsic beliefs: for example, demography, geographic location, etc. Node-level features can be learned by a variety of methods, such as spectral embeddings on the graph Laplacian or graph neural networks. Previous works have used the framework of combining a robust

Algorithm 7 Learning from Misreporting Equilibrium

Input: Features $X \in \mathbb{R}^{n\times d}$, graph information L and A, observed equilibrium $\boldsymbol{z}'$, set size $|S|$
Output: Set of strategic agents $\hat{S}$, estimated intrisic beliefs $\hat{\boldsymbol{s}}$.
$\widehat{\boldsymbol{s}'} \leftarrow A^{-1}((I-A)L+A)\boldsymbol{z}'$
$\hat{v} \leftarrow$ Robust Regression (TORRENT) with design matrix X, response vector $\widehat{\boldsymbol{s}'}$
$\hat{\boldsymbol{s}} \leftarrow X\hat{\boldsymbol{v}}$
diffs $\leftarrow |\hat{\boldsymbol{s}} - \widehat{\boldsymbol{s}'}|$
$\hat{S} \leftarrow$ indices of top k largest values in diffs
return $\hat{\boldsymbol{s}} \in \mathbb{R}^n, \hat{S} \subseteq [n]$

estimator with model-specific information to learn from "strategic sources" of data, such as in bandits Kapoor et al. (2019), controls Russo (2023), and network formation games Jalan and Chakrabarti (2024).

In the sequel, we give the precise technical condition of the features required for robust regression to work (Bhatia et al., 2015), which is based on conditions on the minimum and maximum eigenvalues of the correlation matrix determined by the features corresponding to agents in S. Specifically, for a matrix $X \in \mathbb{R}^{n\times d}$ with n samples in $\mathbb{R}^d$ and $S \subset [n]$ let $X_S \in \mathbb{R}^{|S|\times d}$ select rows in S. Note that $\lambda_{\min}(\cdot), \lambda_{\max}(\cdot)$ are the min/max eigenvalues respectively.

Definition 6.4.1 (SSC and SSS Conditions)**.** *Let $\gamma \in (0,1)$. The features matrix $X \in \mathbb{R}^{n\times d}$ satisfies the Subset Strong Convexity Property at level $1-\gamma$ and Subset Strong Smoothness Property at level γ with constants $\xi_{1-\gamma}, \Xi_\gamma$ respectively if:*

$$\xi_{1-\gamma} \leq \min_{S\subset[n]:|S|=(1-\gamma)n} \lambda_{\min}(X_S^T X_S),$$
$$\Xi_\gamma \geq \max_{S\subset[n]:|S|=\gamma n} \lambda_{\max}(X_S^T X_S).$$

We give our guarantee for Algorithm 7.

Proposition 6.4.2. *Let X be as in Algorithm 7, and suppose that $X\boldsymbol{v} = \boldsymbol{s}$ for some $\boldsymbol{v} \in \mathbb{R}^d$, and that X satisfies the SSC condition at level $1-\gamma$ with constant $\xi_{1-\gamma}$, and SSS condition at level γ with constant Ξ_γ (Definition 6.4.1). Then, there exist absolute constants $C, C' > 0$ such that if $|S| \leq Cn$ and $4\frac{\sqrt{\Xi_\gamma}}{\sqrt{\xi_{1-\gamma}}} < 1$, Algorithm 7 returns $\hat{\boldsymbol{s}}$ such that:*

$$\|\hat{\boldsymbol{s}} - \boldsymbol{s}\|_2 \leq \|X\|_2 n^{-\omega(1)},$$

using $T = C'(\log n)^2$ *iterations of* TORRENT *for the Robust Regression step. Moreover, if for all* $j \in S$ *we have* $|\boldsymbol{s}_j - \boldsymbol{s}'_j| \gg \|X\|_2 n^{-\omega(1)}$*, then* $\hat{S} = S$*.*

It is interesting to investigate what an upper bound on the size of S is when nodes have community memberships, such that recovery is possible, as the SSC and SSS conditions determine it. Specifically, we show the following for a blockmodel graph (proof deferred in the Appendix).

Proposition 6.4.3. *If* G *has two communities with* n_1 *and* n_2 *nodes respectively such that* $n_1 \geq n_2 \geq 1$*, and* $X \in \{0,1\}^{n\times 2}$ *is an embedding vector where each row* $\boldsymbol{x}_i$ *corresponds to a one-hot vector for the community of node* i*, then Algorithm 7 can recover* S *perfectly as long as* $|S| < \frac{n_1}{17}$*.*

For instance, when Proposition 6.4.3 is applied to the Polblogs dataset, it shows that S can be fully recovered as long as $|S| \leq 9$. We can obtain a slightly worse bound and extend the result to a blockmodel graph with K communities (proof in the Appendix).

Proposition 6.4.4. *If* G *has* $K \geq 2$ *communities with sizes* $n_1 \geq n_2 \geq \cdots \geq n_K$ *with* $\left(\frac{16K}{16K+1}\right)\frac{n}{K} < n_K \leq \frac{n}{K}$ *and* $X \in \{0,1\}^{n\times K}$ *is an embedding vector where* $\boldsymbol{x}_i$ *corresponds to an one-hot encoding of the community membership, then Algorithm 7 can recover* S *perfectly as long as* $|S| < \frac{1}{16K+1}n$*.*

In detail, Proposition 6.4.4 states that as long as the smallest community of the graph has size $\Theta(n/K)$ then the recovery of a set of size $|S| = O(n/K)$ is possible. If $|S| \gg n/K$ and S contains all members of the smallest community, then robust regression can fail.

We show that in real-world datasets, Algorithm 7 can identify the set of deviators with high accuracy (cf. Figure 6.7). Specifically, in the real-world datasets we take S to be a randomly sampled set of size $\lceil pn \rceil$ for $p \in \{0.05, 0.1, 0.15, 0.2\}$ and the embeddings to be 128-dimenaional Node2Vec embeddings for Twitter and Reddit and community membership embeddings for the Pollblogs dataset. Algorithm 7 achieves low recovery error as well as high balanced accuracy score.

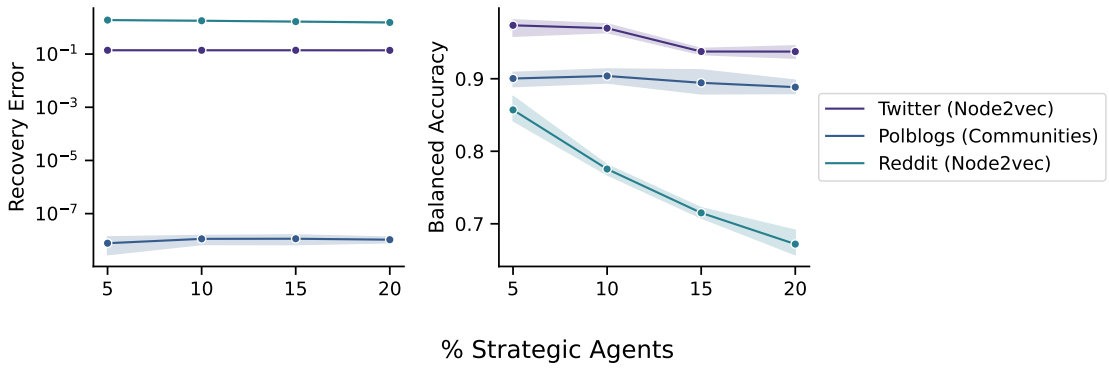


Figure 6.7: Reconstruction error and balanced accuracy for the robust regression problem presented in Algorithm 7. The x-axis shows the percentage of strategic agents. The left subfigure shows the recovery error, measured as $\frac{1}{n}\sum_{i\in[n]}\left|\frac{\hat{\boldsymbol{s}}_i-\boldsymbol{s}_i}{\boldsymbol{s}_i}\right|$, and the right subfigure measures the balanced accuracy between the recovered $\hat{S}$ and the true S. To construct the confidence intervals, for each size $|S|$ of the set S, we draw S five times randomly from the vertex set $[n]$. For the Twitter and Reddit datasets, we have used 128-dimensional Node2vec embeddings. For the Polblogs dataset we have used the community membership (which corresponds to the political orientation) of each node, such that $\boldsymbol{x}_i = (1, 0)$ corresponds to liberal and $\boldsymbol{x}_i = (0, 1)$ corresponds to conservative, and (the true) $\boldsymbol{s}$ is such that $\boldsymbol{s} = X\boldsymbol{v}$ for $\boldsymbol{v} = (1, -1)^T$. We have also provided results using 128-dimensional spectral embeddings. We have set the recovery threshold for TORRENT to be $|S|/n$, and the step size to be $\eta = 1/\|\bar{X}\|_2^2$ where $\bar{X}$ is the min-max normalized embedding matrix.

6.5 Discussion and Conclusion

In this paper, we examine how opinions evolve in social networks, where individuals adjust their publicly stated views based on interactions with others and their inherent beliefs. In our model, strategically motivated users can distort these dynamics by misrepresenting their intrinsic opinions, often to advance conflicting objectives or promote rival narratives. We analyze the Nash Equilibrium of the resulting strategic interactions and empirically show — using diverse datasets from Twitter, Reddit, and Political Blogs — that such deceptive behavior intensifies polarization, fuels disagreement, and increases equilibrium costs. Additionally, we establish worst-case guarantees on the Price of Misreporting, akin to the Price of Anarchy, and introduce scalable learning algorithms to help platforms (i) detect opinion manipulation and (ii) identify the users responsible. Our algorithms perform effectively on real-world data, suggesting how platforms might mitigate the effects of strategic opinion shaping.

We conclude with some discussion of the implications of our work, and directions for future work.

Structural Platform Interventions. We give algorithms for platforms to detect in strategic manipulation has occurred, and who is responsible (Section 6.4). Having done so, a platform might seek to mitigate the effects of strategic behavior. There are multiple plausible avenues to do so.

First, they may seek to reduce degree disparities in the network by algorithmically encouraging balanced connections, such as by suggesting users connect with those who have fewer connections, or reducing the visibility of central nodes (hubs). As we saw, both the upper bound on the PoM and the ratio $\|\boldsymbol{s}'\|_2/\|\boldsymbol{s}\|_2$ depend on the largest eigenvalue λ_n of the Laplacian, which scales with the maximum degree. This motivates interventions to balance the degree distribution.

Second, platforms can design strategy-proof mechanisms to incentivize the agents to report their true opinions. For resource allocation games on networks, it is known that the classical Vickrey–Clarke–Groves (VCG) mechanism is susceptible to adversarial behaviors such as collusion, motivating the need for different mechanisms (Chorppath et al., 2015). In

the case of social networks, platforms have unique tools such as fines or banning of accounts to modify agents' utility functions.

Future Work First, it is not clear which measure of centrality should be used to identify the users who are most capable of manipulating others. As can be seen from Theorem 6.2.2, the influence of agent $i \in S$ should depend on the other members of S, as well as the spectral properties of the localized Laplacian matrices L_j for $j \in S$ and susceptibility parameters α_k for $k \in [n]$.

Second, providing PoM bounds where S can be any set of agents constitutes another interesting research direction, especially if platforms can assume that S is a small fraction of all users. As we noted after Theorem 6.3.1, we believe that this would require restricted invertibility analysis of the matrices determining Nash equilibria.

Third, future work might consider different models of strategic manipulation. For example, one could consider a "feedback equilibrium" model (in the sense of dynamic games Li et al. (2024)), in which agents $i \in S$ can report arbitrary $\boldsymbol{z}'_i(t)$ at each timestep t, rather than following Eq. (6.1). This flexbility may give strategic agents more power to influence outcomes.

6.6 Proofs

6.6.1 Proof of Theorem 6.2.2

Proof of Theorem 6.2.2. Consider agent $i \in S$. To calculate the best-response $\boldsymbol{s}'_i$ of i in response to $\boldsymbol{s}'_{-i}$, we analyze derivatives of its cost function with respect to $\boldsymbol{s}'$. Since the

equilibrium $\boldsymbol{z}'$ is $\boldsymbol{z}' = B\boldsymbol{s}'$, we have:

$$\begin{aligned}
c_i(\boldsymbol{z}') &= (1-\alpha_i)\sum_{j\sim i} w_{ij}(\boldsymbol{z}'_i - \boldsymbol{z}'_j)^2 + \alpha_i(\boldsymbol{z}'_i - \boldsymbol{s}_i)^2 \\
c_i(\boldsymbol{s}') &= (1-\alpha_i)\sum_{j\sim i} w_{ij}((\boldsymbol{e}_i - \boldsymbol{e}_j)^T B\boldsymbol{s}')^2 + \alpha_i(\boldsymbol{e}_i^T(B\boldsymbol{s}' - \boldsymbol{s}))^2 \\
&= (1-\alpha_i)\sum_{j\sim i} w_{ij}(\boldsymbol{s}')^T(B^T(\boldsymbol{e}_i - \boldsymbol{e}_j)(\boldsymbol{e}_i - \boldsymbol{e}_j)^T B)(\boldsymbol{s}') \\
&\quad + \alpha_i((\boldsymbol{s}')^T B^T e_i e_i^T B\boldsymbol{s}' - 2(\boldsymbol{s}')^T B^T e_i e_i^T \boldsymbol{s} + \boldsymbol{s}^T e_i e_i^T \boldsymbol{s}) \\
\nabla_{\boldsymbol{s}'} c_i(\boldsymbol{s}') &= (1-\alpha_i)\sum_{j\sim i} w_{ij} 2(B^T(\boldsymbol{e}_i - \boldsymbol{e}_j)(\boldsymbol{e}_i - \boldsymbol{e}_j)^T B)(\boldsymbol{s}') + \alpha_i(2B^T\boldsymbol{e}_i\boldsymbol{e}_i^T B\boldsymbol{s}' - 2B^T\boldsymbol{e}_i\boldsymbol{e}_i^T\boldsymbol{s}), \\
\nabla^2_{\boldsymbol{s}'} c_i(\boldsymbol{s}') &= 2(1-\alpha_i)B^T\left[\sum_{j\sim i} w_{ij} 2(\boldsymbol{e}_i - \boldsymbol{e}_j)(\boldsymbol{e}_i - \boldsymbol{e}_j)^T\right]B + 2\alpha_i B^T\boldsymbol{e}_i\boldsymbol{e}_i^T B.
\end{aligned}$$

Let $L_i \in \mathbb{R}^{n\times n}$ be:

$$L_i := \sum_{j\sim i} w_{ij}(\boldsymbol{e}_i - \boldsymbol{e}_j)(\boldsymbol{e}_i - \boldsymbol{e}_j)^T.$$

Notice that L_i is precisely the Laplacian of the graph when all edges not incident to i are equal to zero. Therefore $L_i \succeq 0$. Since $\boldsymbol{e}_i\boldsymbol{e}_i^T \succeq 0$, the Hessian of c_i with respect to $\boldsymbol{s}'$ is PSD. In particular, its (i,i) entry is non-negative, so $\frac{\partial^2 c_i(\boldsymbol{s}')}{\partial(\boldsymbol{s}'_i)^2} \geq 0$, and hence the optimal $\boldsymbol{s}'_i$ is at the critical point. This is given as:

$$\begin{aligned}
0 &= \frac{1}{2}\frac{\partial}{\partial \boldsymbol{s}'_i} c_i(\boldsymbol{s}') \\
&= \boldsymbol{e}_i^T(1-\alpha_i)B^T\left[\sum_{j\sim i} w_{ij}(\boldsymbol{e}_i - \boldsymbol{e}_j)(\boldsymbol{e}_i - \boldsymbol{e}_j)^T\right]B\boldsymbol{s}' + \boldsymbol{e}_i^T\alpha_i(B^T\boldsymbol{e}_i\boldsymbol{e}_i^T Bs' - B^T\boldsymbol{e}_i\boldsymbol{e}_i^T\boldsymbol{s}) \\
&= (1-\alpha_i)\boldsymbol{e}_i^T B^T L_i B\boldsymbol{s}' + \boldsymbol{e}_i^T\alpha_i(B^T\boldsymbol{e}_i\boldsymbol{e}_i^T B\boldsymbol{s}' - B^T\boldsymbol{e}_i\boldsymbol{e}_i^T\boldsymbol{s}).
\end{aligned}$$

The above display gives the solution for $\boldsymbol{s}'_i$ in terms of all entries of s'. Assembling the critical points into a linear system, we obtain precisely that for all $i \in S$, $\boldsymbol{e}_i^T\mathbb{T}_i\boldsymbol{s}' = \boldsymbol{y}_i$. Since $\boldsymbol{s}'_j = \boldsymbol{s}_j$ for $j \notin S$, the overall linear system describes the Nash equilibria. □

6.6.2 Proof of Corollary 6.2.7

Proof of Corollary 6.2.7. When all agents are deviating, it is straightforward to show that $\tilde{T} = \tilde{\alpha}B$ with minimum eigenvalue $\tilde{\alpha}^2/(\lambda_n + \tilde{\alpha}) > 0$. Thus $\tilde{T}$ is invertible, and therefore $\boldsymbol{s}' = \frac{1}{\tilde{\alpha}}B^{-1}\widetilde{\text{diag}(B)}\boldsymbol{s}$, where $\widetilde{\text{diag}(B)}$ is a diagonal matrix with entries B_{ii}. Then

$$\begin{aligned}\|\boldsymbol{s}'\|_2 &\leq \frac{1}{\tilde{\alpha}}\|B^{-1}\|_2\|\widetilde{\mathrm{diag}(B)}\|_2\|\boldsymbol{s}\|_2 \\ &= \left(\max_i B_{ii}\right)\left(\max_i \frac{\lambda_i+\tilde{\alpha}}{\tilde{\alpha}}\right)\|\boldsymbol{s}\|_2 \\ &\leq \frac{\tilde{\alpha}}{\lambda_1+\tilde{\alpha}}\frac{\lambda_n+\tilde{\alpha}}{\alpha}\|\boldsymbol{s}\|_2 \\ &= \frac{\lambda_n+\tilde{\alpha}}{\tilde{\alpha}}\|\boldsymbol{s}\|_2.\end{aligned}$$

□

6.6.3 Proof of Theorem 6.3.1

Proof of Theorem 6.3.1. First, we set $\tilde{\alpha} = \alpha/(1-\alpha)$. By substituting $\boldsymbol{z} = B\boldsymbol{s}$ we can show by straightforward algebra that $C(\boldsymbol{z})/(1-\alpha) = \boldsymbol{s}^T Q\boldsymbol{s}$ where $Q \succeq 0$ with

$$Q = BLB + \tilde{\alpha}(I - 2B + B^2) \tag{6.12}$$

Since $Q \succeq 0$, it has eigendecomposition $Q = U\Lambda_Q U^T$. Moreover, U is precisely the matrix of eigenvectors for the Laplacian. The eigenvalues of Q can be shown to be $\tilde{\alpha}^2/(\lambda_i+\tilde{\alpha})$. Therefore, $C(z) = \boldsymbol{s}^T Q\boldsymbol{s} \geq \frac{\tilde{\alpha}^2}{\lambda_n+\tilde{\alpha}}\|\boldsymbol{s}\|_2^2$.

Next, let $\mathrm{diag}(B)$ be the diagonal matrix with entries B_{ii} and $\widetilde{\mathrm{diag}(B)}$ be as in Corollary 6.2.7. In the proof of Corollary 6.2.7, we show that $\boldsymbol{s}' = (1/\tilde{\alpha})B^{-1}\widetilde{\mathrm{diag}(B)}\boldsymbol{s}$ and $\boldsymbol{z}' = (1/\tilde{\alpha})\widetilde{\mathrm{diag}(B)}\boldsymbol{s}$, which similarly implies (after algebraic operations) that $C(\boldsymbol{z}')/(1-\alpha) = \boldsymbol{s}^T Q'\boldsymbol{s}$ where:

$$Q' := \frac{1}{\tilde{\alpha}^2}\widetilde{\mathrm{diag}(B)}L\widetilde{\mathrm{diag}(B)} + \frac{1}{\tilde{\alpha}}B^{-1}\left(\widetilde{\mathrm{diag}(B)}\right)^2 B^{-1} - 2\frac{1}{\tilde{\alpha}}B^{-1}\left(\widetilde{\mathrm{diag}(B)}\right)^2 + \frac{1}{\tilde{\alpha}}\left(\widetilde{\mathrm{diag}(B)}\right)^2$$

Note that Q' cannot be diagonalized since, in general, $\widetilde{\mathrm{diag}(B)}$ has a different eigenbasis than L. However, we note that:

$$\|\widetilde{\mathrm{diag}(B)}\|_2 = \max_i B_{ii} \leq \|B\|_2 = 1, \tag{6.13}$$

$$\|B^{-1}\|_2 = \max_i \frac{\lambda_i+\tilde{\alpha}}{\tilde{\alpha}} = \frac{\lambda_n+\tilde{\alpha}}{\tilde{\alpha}}. \tag{6.14}$$

By the triangle inequality, the Cauchy-Schwarz inequality, and Equations (6.13) and (6.14), we have that:

$$\|Q'\|_2 \leq \frac{1}{\tilde{\alpha}^2}\|L\|_2\left(\|\widetilde{\text{diag}(B)}\|_2\right)^2 + \frac{1}{\tilde{\alpha}}\left(\|\widetilde{\text{diag}(B)}\|_2\right)^2\|B^{-1}\|_2^2 + \frac{2}{\tilde{\alpha}}\|B^{-1}\|_2 + \frac{1}{\tilde{\alpha}}\left(\|\widetilde{\text{diag}(B)}\|_2\right)^2$$
$$\leq \frac{(\lambda_n + 4\tilde{\alpha})(\lambda_n + \tilde{\alpha})}{\tilde{\alpha}^3}.$$

Therefore $C(\boldsymbol{z}')/(1-\alpha) \leq \frac{(\lambda_n+4\tilde{\alpha})(\lambda_n+\tilde{\alpha})}{\tilde{\alpha}^3}\|\boldsymbol{s}\|_2^2$. Hence,

$$\frac{C(\boldsymbol{z}')}{C(\boldsymbol{z})} \leq \frac{(\lambda_n + 4\tilde{\alpha})(\lambda_n + \tilde{\alpha})^2}{\tilde{\alpha}^5}. \tag{6.15}$$

Finally, we can simplify:

$$\frac{(\lambda_n + 4\tilde{\alpha})(\lambda_n + \tilde{\alpha})^2}{\tilde{\alpha}^5} \leq \frac{64(\lambda_n + \tilde{\alpha})^3}{\tilde{\alpha}^5} \leq \frac{128(\max\{\lambda_n, \tilde{\alpha}\})^3}{\tilde{\alpha}^5}. \tag{6.16}$$

□

6.6.4 Proof of Corollary 6.3.2

Proof of Corollary 6.3.2. Note that $C(\boldsymbol{z}') \leq (1-\alpha_{\min})(\boldsymbol{z}')^T L\boldsymbol{z}' + \alpha_{\max}\|\boldsymbol{z}' - \boldsymbol{s}\|_2^2 = \overline{C}(\boldsymbol{z}')$, and $C(\boldsymbol{z}) \geq (1-\alpha_{\max})\boldsymbol{z}^T L\boldsymbol{z} + \alpha_{\min}\|\boldsymbol{z} - \boldsymbol{s}\|_2^2 = \underline{C}(\boldsymbol{z})$ where $\alpha_{\min} = \min_{i\in[n]}\alpha_i$, and $\alpha_{\max} = \max_{i\in[n]}\alpha_i$. Then, the same analysis of Theorem 6.3.1 can be applied, since $\overline{C}(\boldsymbol{z}')/\underline{C}(\boldsymbol{z}) \geq C(\boldsymbol{z}')/C(\boldsymbol{z})$. □

6.6.5 Proof of Proposition 6.4.2

We are ready to prove Proposition 6.4.2.

Proof of Proposition 6.4.2. Let $\widehat{\boldsymbol{s}'}$ be as in Algorithm 7, and $\boldsymbol{y} = \widehat{\boldsymbol{s}'}$. Notice $\boldsymbol{y} = \boldsymbol{s} + (\boldsymbol{s}' - \boldsymbol{s}) + (\widehat{\boldsymbol{s}'} - \boldsymbol{s}')$. Let $\boldsymbol{w} := (\boldsymbol{s}' - \boldsymbol{s})$ be the corruption vector due to strategic negotiations and $\boldsymbol{r} = (\widehat{\boldsymbol{s}'} - \boldsymbol{s}')$ be the residual vector due to least-squares regression. We claim that $\boldsymbol{r} = \boldsymbol{0}$, because A^{-1} is full rank and $((I-A)L + A)$ is full rank, so $\widehat{\boldsymbol{s}'} = A^{-1}((I-A)L + A)\boldsymbol{z}' = \boldsymbol{s}'$.

Next, we apply the main Theorem of Bhatia et al. (2015) (Theorem 5.8.8). Notice that $\|\boldsymbol{w}\|_0 \leq Cn$ by assumption. Moreover, X satisfies the SSC and SSS coniditons. Therefore, after T iterations, Algorithm 7 obtains $\hat{\boldsymbol{s}}$ such that:

$$\|\hat{\boldsymbol{v}} - \boldsymbol{v}\|_2 \leq \frac{\exp(-cT)}{\sqrt{n}}\|\boldsymbol{s}' - \boldsymbol{s}\|_2.$$

Therefore, letting $T = C'(\log n)^2$ for large enough constant $C' > 0$, we see that $\|\hat{\boldsymbol{v}} - \boldsymbol{v}\|_2 \leq O(n^{-\omega(1)})$. Hence $\|\hat{\boldsymbol{s}} - \boldsymbol{s}\|_2 = \|X\hat{\boldsymbol{v}} - X\boldsymbol{v}\|_2 \leq \|X\|_2 \cdot n^{-\omega(1)}$. Now, let $\boldsymbol{u} = \hat{\boldsymbol{s}} - \boldsymbol{s}'$. If $i \in [n] \setminus S$, then $|\boldsymbol{u}_i| \leq \|X\|_2 \cdot n^{-\omega(1)}$. On the other hand for $j \in S$, $|\boldsymbol{u}_j| \geq |\boldsymbol{s}_j - \boldsymbol{s}'_j| - \|X\|_2 n^{-\omega(1)}$. Therefore the top-$|S|$ entries of $\boldsymbol{u}$ recover S. □

6.6.6 Proof of Proposition 6.4.3

Let V_1 correspond to the vertex set for community 1 and V_2 correspond to the vertex set for community 2. Let I_q denote the $q \times q$ identity matrix, and $a_i = |V_i \cap S|$ for $i = 1, 2$. Then

$$X_S^T X_S = \begin{pmatrix} I_{a_1} & 0 \\ 0 & I_{a_2} \end{pmatrix}.$$

Therefore $\lambda_{\max}(X_S^T X_S) = \max\{|V_1 \cap S|, |V_2 \cap S|\}$ and $\lambda_{\min}(X_S^T X_S) = \min\{|V_1 \cap S|, |V_1 \cap S|\}$.

First, we determine sufficient ranges of γ and the value of Ξ_γ: Let S be such that $|S| = \gamma n$. We have the following options:

- $S \subseteq V_1$. Then $\lambda_{\max}(X_S^T X_S) = \gamma n$.
- $S \subseteq V_2$. Then $\lambda_{\max}(X_S^T X_S) = \gamma n$.
- $V_1 \subseteq S$. Then $\lambda_{\max}(X_S^T X_S)$ lies between $\gamma n/2$ and γn.
- $V_2 \subseteq S$. Then $\lambda_{\max}(X_S^T X_S)$ lies between $\gamma n/2$ and γn.
- If S lies partially in V_1 and V_2, then $\lambda_{\max}(X_S^T X_S) = \max\{(1-t)\gamma n, t\gamma n\}$ for some $t \in [0, 1]$. Again, this is upper bounded by γn.

The above yield $\Xi_\gamma = \gamma n$. To determine $\xi_{1-\gamma}$ we let S be such that $|S| = (1-\gamma)n$. We have the following options:

- $S \subseteq V_1$. Then $\lambda_{\min}(X_S^T X_S) = 0$.
- $S \subseteq V_2$. Then $\lambda_{\min}(X_S^T X_S) = 0$.
- $V_1 \subseteq S$. Then $\lambda_{\min}(X_S^T X_S) = \min\{n_1, (1-\gamma)n - n_1\} = n_1 \geq (1-\gamma)n - n_2$ since always $n_1 \geq (1-\gamma)n/2$.

- $V_2 \subseteq S$. Then $\lambda_{\min}(X_S^T X_S) = \min\{n_2, (1-\gamma)n - n_2\} = (1-\gamma)n - n_2$ since $n_2 \leq (1-\gamma)n/2$.

- If S lies partially in V_1 and V_2, then $\lambda_{\min}(X_S^T X_S) = \min\{(1-t)(1-\gamma)n, t(1-\gamma)n\}$ for some $t \in [0,1]$. Again, this is lower bounded by $(1-\gamma)n - n_2$.

Therefore, for either $1-\gamma \leq n_1/n$ or $1-\gamma \leq n_2/n$ we have that $\xi_{1-\gamma} = (1-\gamma)n - n_2$. The final inequality corresponds to

$$4\sqrt{\frac{\Xi_\gamma}{\xi_{1-\gamma}}} < 1 \iff \gamma < \frac{1}{17} - \frac{n_2}{17n}$$

Combining the above we get two systems of inequalities. The first one corresponds to $1 - \frac{n_1}{n} \leq \gamma < \frac{1}{17} - \frac{n_2}{17n}$ which holds for $n_2 < 1/18$ which is impossible since $n_2 \geq 1$. The second one corresponds to $1 - \frac{n_2}{n} \leq \gamma < \frac{1}{17} - \frac{n_2}{17n}$ which holds for $n_2 > 16/18$, which is always true.

6.6.7 Proof of Proposition 6.4.4

First, we note that if $V_1, \ldots, V_K$ are the vertex sets and S is a set of size γn the maximum eigenvalue equals to $\lambda_{\max}(X_S^T X_S) = \max_{i \in [K]} |V_i \cap S|$ and is always at most γn. So $\Xi_\gamma = \gamma n$. If S is a set of size $(1-\gamma)n$, then the minimum eigenvalue $\lambda_{\min}(X_S^T X_S) = \min_{i \in [K} |V_i \cap S|$ is maximized when $|V_1 \cap S| = \cdots = |V_K \cap S| = (1-\gamma)n/K$ which holds as long as $(1-\gamma)n/K \leq n_K$, so $\xi_{1-\gamma} = (1-\gamma)n/K$ as long as $\gamma \geq 1 - n_K/nK$. Also the other condition is

$$4\sqrt{\frac{\Xi_\gamma}{\xi_{1-\gamma}}} < 1 \iff \gamma < \frac{1}{16K+1}$$

Finally, we must have $1/(16K+1) > 1 - Kn_K/K$ which yields $n_K > n/K(16K/(16K+1))$.

6.7 Additional Figures

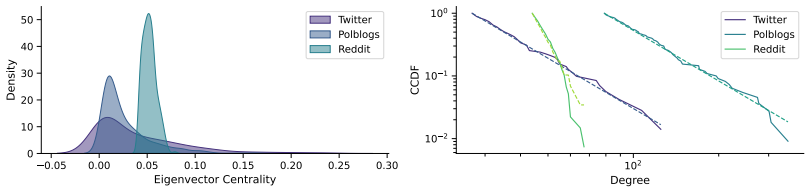


Figure 6.8: Distribution of centralities and degrees for the datasets

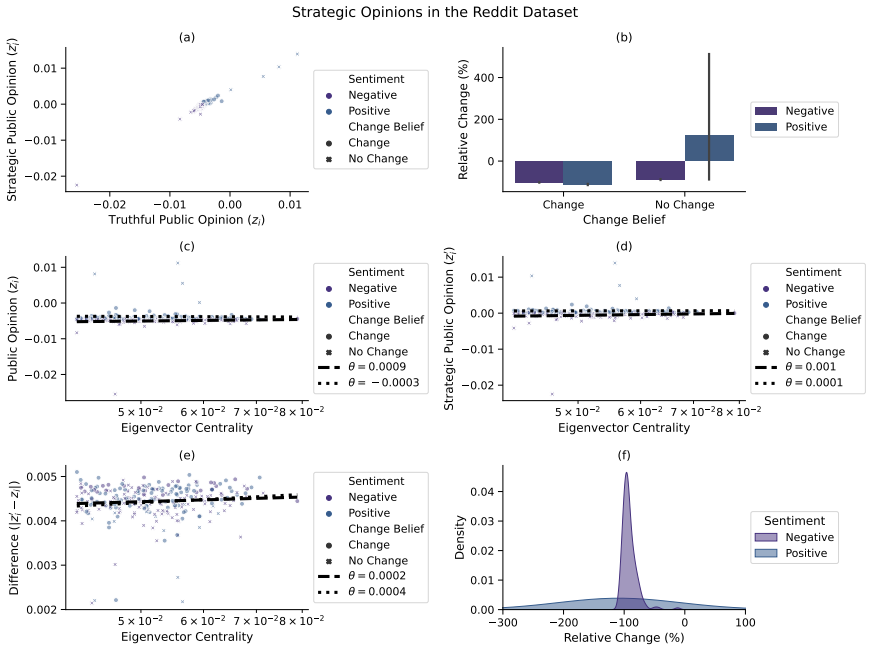


Figure 6.9: Running the experiments of Figure 6.3 for the Reddit dataset.

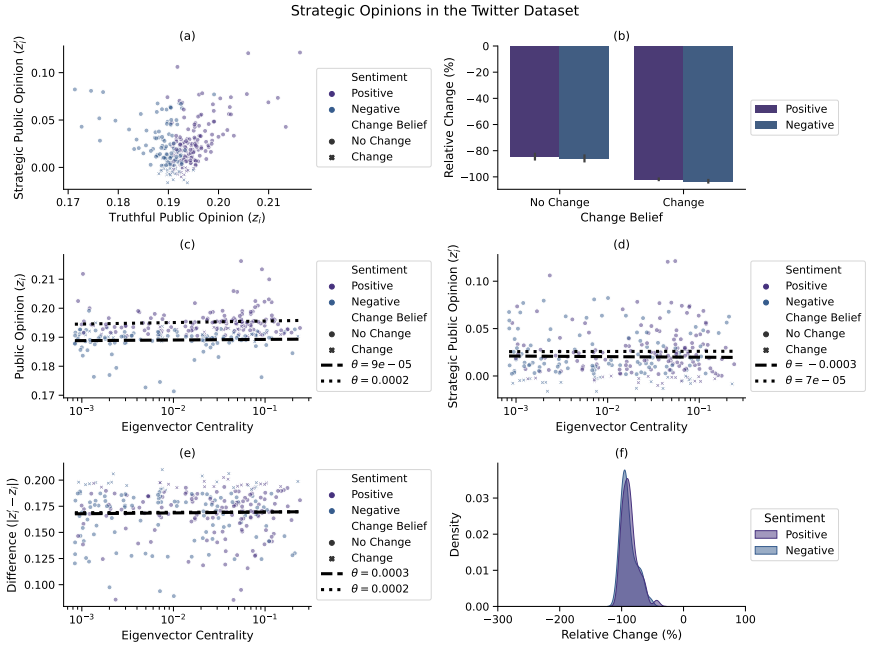


Figure 6.10: Running the experiments of Figure 6.3 for the Twitter dataset.

References

Yasin Abbasi-Yadkori, Dávid Pál, and Csaba Szepesvári. Improved algorithms for linear stochastic bandits. *Advances in neural information processing systems*, 24, 2011.

Emmanuel Abbe. Community detection and stochastic block models. *arXiv preprint arXiv:1703.10146*, 2017.

Emmanuel Abbe and Colin Sandon. Provable limitations of deep learning. *arXiv preprint arXiv:1812.06369*, 2018.

Rediet Abebe, Jon Kleinberg, David Parkes, and Charalampos E Tsourakakis. Opinion dynamics with varying susceptibility to persuasion. In *Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, pages 1089–1098, 2018.

Daron Acemoglu and Pablo D Azar. Endogenous production networks. *Econometrica*, 88(1):33–82, 2020.

Daron Acemoglu, Asuman Ozdaglar, and Alireza Tahbaz-Salehi. Systemic risk and stability in financial networks. *American Economic Review*, 105(2):564–608, 2015.

Lada A Adamic and Natalie Glance. The political blogosphere and the 2004 us election: divided they blog. In *Proceedings of the 3rd international workshop on Link discovery*, pages 36–43, 2005.

Alekh Agarwal, Yuda Song, Wen Sun, Kaiwen Wang, Mengdi Wang, and Xuezhou Zhang. Provable benefits of representational transfer in reinforcement learning. In Gergely Neu and Lorenzo Rosasco, editors, *Proceedings of Thirty Sixth Conference on Learning Theory*, volume 195 of *Proceedings of Machine Learning Research*, pages 2114–2187. PMLR, 12–15 Jul 2023a. URL `https://proceedings.mlr.press/v195/agarwal23b.html`.

Anish Agarwal, Munther Dahleh, Devavrat Shah, and Dennis Shen. Causal matrix completion. In *The thirty sixth annual conference on learning theory*, pages 3821–3826. PMLR, 2023b.

Shubhangi Agarwal, Sourav Dutta, and Arnab Bhattacharya. Chisel: Graph similarity search using chi-squared statistics in large probabilistic graphs. *Proceedings of the VLDB Endowment*, 13(10):1654–1668, 2020.

William Aiello, Fan Chung, and Linyuan Lu. A random graph model for massive graphs. In *Proceedings of the 32nd Annual ACM Symposium on Theory of Computing (STOC '00)*, pages 171–180, 2000.

Edoardo M. Airoldi, David M. Blei, Stephen E. Fienberg, and Eric P. Xing. Mixed membership stochastic blockmodels. *Journal of Machine Learning Research*, 9:1981–2014, 2008.

David J. Aldous. Representations for partially exchangeable arrays of random variables. *Journal of Multivariate Analysis*, 11:581–598, 1981.

Tal Alon, Inbal Talgam-Cohen, Ron Lavi, and Elisheva Shamash. Incomplete information vcg contracts for common agency. *Operations Research*, 2023.

Hamed Amini, Andreea Minca, and Agnes Sulem. Control of interbank contagion under partial information. *SIAM Journal on Financial Mathematics*, 6(1):1195–1219, 2015.

Mel Andrews. The immortal science of ml: Machine learning & the theory-free ideal. *Preprint at https://rgdoi. net/10.13140/RG*, 2(28311.75685), 2023.

Andrew Ang. *Asset management: A systematic approach to factor investing*. 2014.

Ioannis D. Apostolopoulos and Tzani Bessiana. Covid-19: automatic detection from x-ray images utilizing transfer learning with convolutional neural networks. *Physical and Engineering Sciences in Medicine*, 43:635 – 640, 2020. URL `https://api.semanticscholar.org/CorpusID:214667149`.

Pasqualina Arca, Gianfranco Atzeni, and Luca Deidda. The signaling role of trade credit: Evidence from a counterfactual analysis. *Journal of Corporate Finance*, 80: 102414, 2023.

Rohit Arora, Rui Gao, and Stathis Tompaidis. Extreme yet plausible: Choosing scenarios to stress test financial institutions. *Available at SSRN 3803263*, 2021.

Sanjeev Arora and Boaz Barak. *Computational complexity: a modern approach*. Cambridge University Press, 2009.

Kenneth J Arrow and Gerard Debreu. Existence of an equilibrium for a competitive economy. *Econometrica: Journal of the Econometric Society*, pages 265–290, 1954.

W Brian Arthur. Foundations of complexity economics. *Nature Reviews Physics*, 3(2): 136–145, 2021.

Angelos Assos, Yuval Dagan, and Constantinos Daskalakis. Maximizing utility in multi-agent environments by anticipating the behavior of other learners. *arXiv preprint arXiv:2407.04889*, 2024.

Lars Backstrom, Paolo Boldi, Marco Rosa, Johan Ugander, and Sebastiano Vigna. Four degrees of separation. In *Proceedings of the 4th annual ACM Web science conference*, pages 33–42, 2012.

Abhijit Banerjee, Arun G Chandrasekhar, Esther Duflo, and Matthew O Jackson. The diffusion of microfinance. *Science*, 341(6144):1236498, 2013.

Tathagata Banerjee and Zachary Feinstein. Pricing of debt and equity in a financial network with comonotonic endowments. *Operations Research*, 70(4):2085–2100, 2022.

Mayank Baranwal, Abram Magner, Paolo Elvati, Jacob Saldinger, Angela Violi, and Alfred O Hero. A deep learning architecture for metabolic pathway prediction. *Bioinformatics*, 36(8):2547–2553, 2020.

Pablo Barberá, John T Jost, Jonathan Nagler, Joshua A Tucker, and Richard Bonneau. Tweeting from left to right: Is online political communication more than an echo chamber? *Psychological science*, 26(10):1531–1542, 2015.

Paolo Barucca, Marco Bardoscia, Fabio Caccioli, Marco D'Errico, Gabriele Visentin, Guido Caldarelli, and Stefano Battiston. Network valuation in financial systems. *Mathematical Finance*, 30(4):1181–1204, 2020.

Shai Ben-David, John Blitzer, Koby Crammer, and Fernando Pereira. Analysis of representations for domain adaptation. *Advances in neural information processing systems*, 19, 2006.

Omer Ben-Porat and Moshe Tennenholtz. A game-theoretic approach to recommendation systems with strategic content providers. *Advances in Neural Information Processing Systems*, 31, 2018.

Aniruddha Bhargava, Ravi Ganti, and Rob Nowak. Active positive semidefinite matrix completion: Algorithms, theory and applications. In *Artificial Intelligence and Statistics*, pages 1349–1357. PMLR, 2017.

Kush Bhatia, Prateek Jain, and Purushottam Kar. Robust regression via hard thresholding. *Advances in neural information processing systems*, 28, 2015.

Sohom Bhattacharya and Sourav Chatterjee. Matrix completion with data-dependent missingness probabilities. *IEEE Transactions on Information Theory*, 68(10):6762–6773, 2022.

Arnab Bhattacharyya, Mark Braverman, Bernard Chazelle, and Huy L Nguyen. On the convergence of the hegselmann-krause system. In *Proceedings of the 4th conference on Innovations in Theoretical Computer Science*, pages 61–66, 2013.

Kshipra Bhawalkar, Sreenivas Gollapudi, and Kamesh Munagala. Coevolutionary opinion formation games. In *Proceedings of the forty-fifth annual ACM symposium on Theory of computing*, pages 41–50, 2013.

Sergio Bianchi, Alexandre Pantanella, and Augusto Pianese. Modeling stock prices by multifractional brownian motion: an improved estimation of the pointwise regularity. *Quantitative finance*, 13(8):1317–1330, 2013.

Philippe Bich and Lisa Morhaim. On the existence of pairwise stable weighted networks. *Mathematics of Operations Research*, 45(4):1393–1404, 2020.

Philippe Bich and Mariya Teteryatnikova. On perfect pairwise stable networks. *Journal of Economic Theory*, page 105577, 2022.

Peter J Bickel and Aiyou Chen. A nonparametric view of network models and newman–girvan and other modularities. *Proceedings of the National Academy of Sciences*, 106 (50):21068–21073, 2009.

David Bindel, Jon Kleinberg, and Sigal Oren. How bad is forming your own opinion? In *2011 IEEE 52nd Annual Symposium on Foundations of Computer Science*, pages 57–66. IEEE, 2011.

David Bindel, Jon Kleinberg, and Sigal Oren. How bad is forming your own opinion? *Games and Economic Behavior*, 92:248–265, 2015.

John R Birge. Modeling investment behavior and risk propagation in financial networks. *Available at SSRN 3847443*, 2021.

Sophie Bjork-James and Joan Donovan. Profiteers are exploiting us election conspiracies and hate to make millions, 2024. URL `https://www.wired.com/story/2024-election-profiteers?utm_source=chatgpt.com`. Accessed: 2024-12-27.

J Martin Bland and Douglas G Altman. Bayesians and frequentists. *Bmj*, 317(7166): 1151–1160, 1998.

Peter Bogetoft, Jasone Ramírez-Ayerbe, and Dolores Romero Morales. Counterfactual analysis and target setting in benchmarking. *European Journal of Operational Research*, 315(3):1083–1095, 2024.

George EP Box. Science and statistics. *Journal of the American Statistical Association*, 71(356):791–799, 1976.

Edgar K Browning and Mark A Zupan. *Microeconomics: Theory and applications*. John Wiley & Sons, 2020.

Hans Bühlmann. An economic premium principle. *ASTIN Bulletin: The Journal of the IAA*, 11(1):52–60, 1980.

Hans Bühlmann. The general economic premium principle. *ASTIN Bulletin: The Journal of the IAA*, 14(1):13–21, 1984.

Ricardo J Caballero and Alp Simsek. Fire sales in a model of complexity. *The Journal of Finance*, 68(6):2549–2587, 2013.

T Tony Cai and Hongming Pu. Transfer learning for nonparametric regression: Non-asymptotic minimax analysis and adaptive procedure. *arXiv preprint arXiv:2401.12272*, 2024.

T Tony Cai and Hongji Wei. Transfer learning for nonparametric classification: Minimax rate and adaptive classifier. *The Annals of Statistics*, 49(1), 2021a.

T Tony Cai and Hongji Wei. Transfer learning for nonparametric classification: Minimax rate and adaptive classifier. *The Annals of Statistics*, 49(1), 2021b.

Yang Cai, Constantinos Daskalakis, and Christos Papadimitriou. Optimum statistical estimation with strategic data sources. In *Conference on Learning Theory*, pages 280–296. PMLR, 2015.

Alberto Caimo and Nial Friel. Bayesian inference for exponential random graph models. *Social Networks*, 33(1):41–55, 2011.

Giuseppe C Calafiore, Giulia Fracastoro, and Anton V Proskurnikov. Control of dynamic financial networks. *IEEE Control Systems Letters*, 2022.

Frank M Callier and Charles A Desoer. *Linear System Theory*. Springer Verlag (Springer Texts in Electrical Engineering), 1994.

Antoni Calvó-Armengol and Rahmi Ilkılıç. Pairwise-stability and nash equilibria in network formation. *International Journal of Game Theory*, 38(1):51–79, 2009.

Emmanuel J Candès and Benjamin Recht. Exact matrix completion via convex optimization. *Found Comput Math*, 9:717–772, 2009.

Emmanuel J Candès and Terence Tao. The power of convex relaxation: Near-optimal matrix completion. *IEEE transactions on information theory*, 56(5):2053–2080, 2010.

Bin Cao, Nathan N Liu, and Qiang Yang. Transfer learning for collective link prediction in multiple heterogenous domains. In *Proceedings of the 27th international conference on machine learning (ICML-10)*, pages 159–166. Citeseer, 2010.

Xun Cao and Michael D Ward. Do democracies attract portfolio investment? transnational portfolio investments modeled as dynamic network. *International Interactions*, 40(2):216–245, 2014.

Lisa Carlson and Raymond Dacey. Game theory: International trade, conflict and co-operation. In *Global Political Economy*, pages 91–103. Routledge, 2013.

René Carmona, Daniel B Cooney, Christy V Graves, and Mathieu Lauriere. Stochastic graphon games: I. The static case. *Mathematics of Operations Research*, 47(1):750–778, 2022.

Brian Castellani and Frederic William Hafferty. Sociology and complexity science a new field of inquiry. 2009.

Damon Centola, Joshua Becker, Devon Brackbill, and Andrea Baronchelli. Experimental evidence for tipping points in social convention. *Science*, 360(6393):1116–1119, 2018.

Deepayan Chakrabarti, Yiping Zhan, and Christos Faloutsos. R-MAT: A recursive model for graph mining. In *Proceedings of the 4th SIAM International Conference on Data Mining (SDM '04)*, pages 442–446, 2004.

Shayok Chakraborty, Jiayu Zhou, Vineeth Balasubramanian, Sethuraman Panchanathan, Ian Davidson, and Jieping Ye. Active matrix completion. In *2013 IEEE 13th international conference on data mining*, pages 81–90. IEEE, 2013.

Stanley Chan and Edoardo Airoldi. A consistent histogram estimator for exchangeable graph models. In *International Conference on Machine Learning*, pages 208–216. PMLR, 2014.

Serina Chang, Frederic Koehler, Zhaonan Qu, Jure Leskovec, and Johan Ugander. Inferring dynamic networks from marginals with iterative proportional fitting. *arXiv preprint arXiv:2402.18697*, 2024.

Sourav Chatterjee. Matrix estimation by universal singular value thresholding. *The Annals of Statistics*, pages 177–214, 2015a.

Sourav Chatterjee. Matrix estimation by universal singular value thresholding. *The Annals of Statistics*, pages 177–214, 2015b.

Bernard Chazelle. The total s-energy of a multiagent system. *SIAM Journal on Control and Optimization*, 49(4):1680–1706, 2011.

Mayee F Chen and Miklós Z Rácz. An adversarial model of network disruption: Maximizing disagreement and polarization in social networks. *IEEE Transactions on Network Science and Engineering*, 9(2):728–739, 2021a.

Mayee F Chen and Miklós Z Rácz. An adversarial model of network disruption: Maximizing disagreement and polarization in social networks. *IEEE Transactions on Network Science and Engineering*, 9(2):728–739, 2021b.

Tianyi Chen and Charalampos Tsourakakis. Antibenford subgraphs: Unsupervised anomaly detection in financial networks. In *Proceedings of the 28th ACM SIGKDD Conference on Knowledge Discovery and Data Mining*, pages 2762–2770, 2022.

Yiling Chen, Yang Liu, and Chara Podimata. Learning strategy-aware linear classifiers. *Advances in Neural Information Processing Systems*, 33:15265–15276, 2020a.

Yudong Chen, Sujay Sanghavi, and Huan Xu. Improved graph clustering. *IEEE Transactions on Information Theory*, 60(10):6440–6455, 2014.

Yuxin Chen, Yuejie Chi, Jianqing Fan, Cong Ma, and Yuling Yan. Noisy matrix completion: Understanding statistical guarantees for convex relaxation via nonconvex optimization. *SIAM journal on optimization*, 30(4):3098–3121, 2020b.

Yuxin Chen, Yuejie Chi, Jianqing Fan, Cong Ma, et al. Spectral methods for data science: A statistical perspective. *Foundations and Trends® in Machine Learning*, 14 (5):566–806, 2021.

Uthsav Chitra and Christopher Musco. Analyzing the impact of filter bubbles on social network polarization. In *Proceedings of the 13th International Conference on Web Search and Data Mining*, pages 115–123, 2020.

Vijay K Chopra and William T Ziemba. The effect of errors in means, variances, and covariances on optimal portfolio choice. In *Handbook of the fundamentals of financial decision making: Part I*, pages 365–373. 2013.

Anil Kumar Chorppath, Tansu Alpcan, and Holger Boche. Adversarial behavior in network games. *Dynamic Games and Applications*, 5:26–64, 2015.

Bjarke Christensen and Jens Nielsen. *Metabolic Network Analysis*, pages 209–231. Springer Berlin Heidelberg, Berlin, Heidelberg, 2000. ISBN 978-3-540-48773-9. doi: 10.1007/3-540-48773-5_7. URL https://doi.org/10.1007/3-540-48773-5_7.

George Christodoulou and Alkmini Sgouritsa. Designing networks with good equilibria under uncertainty. *SIAM Journal on Computing*, 48(4):1364–1396, 2019.

Giorgos Christodoulou, Vasilis Gkatzelis, and Alkmini Sgouritsa. Cost-sharing methods for scheduling games under uncertainty. In *Proceedings of the 2017 ACM conference on economics and computation*, pages 441–458, 2017.

Alexander Chudik, Kamiar Mohaddes, M Hashem Pesaran, Mehdi Raissi, and Alessandro Rebucci. A counterfactual economic analysis of covid-19 using a threshold augmented multi-country model. *Journal of International Money and Finance*, 119:102477, 2021.

Tyler Cody and Peter A. Beling. A systems theory of transfer learning. *IEEE Systems Journal*, 17(1):26–37, 2023. doi: 10.1109/JSYST.2022.3224650.

Rama Cont and Andreea Minca. Credit default swaps and systemic risk. *Annals of Operations Research*, 247(7):523–547, 2016.

Corinna Cortes, Mehryar Mohri, Michael Riley, and Afshin Rostamizadeh. Sample selection bias correction theory. In *International conference on algorithmic learning theory*, pages 38–53. Springer, 2008.

Emanuele Cozzo, Guilherme Ferraz De Arruda, Francisco Aparecido Rodrigues, and Yamir Moreno. *Multiplex networks: basic formalism and structural properties*, volume 10. Springer, 2018.

Koby Crammer, Michael Kearns, and Jennifer Wortman. Learning from multiple sources. *Journal of Machine Learning Research*, 9(8), 2008.

Sanjoy Dasgupta. Two faces of active learning. *Theoretical computer science*, 412(19): 1767–1781, 2011.

Abhirup Datta, Jacob Fiksel, Agbessi Amouzou, and Scott L Zeger. Regularized bayesian transfer learning for population-level etiological distributions. *Biostatistics*, 22(4):836–857, 2021.

Hal Daumé. Frustratingly easy domain adaptation. *ArXiv*, abs/0907.1815, 2007. URL `https://api.semanticscholar.org/CorpusID:5360764`.

Mark A Davenport, Yaniv Plan, Ewout Van Den Berg, and Mary Wootters. 1-bit matrix completion. *Information and Inference: A Journal of the IMA*, 3(3):189–223, 2014.

Abir De, Isabel Valera, Niloy Ganguly, Sourangshu Bhattacharya, and Manuel Gomez Rodriguez. Learning and forecasting opinion dynamics in social networks. *Advances in neural information processing systems*, 29, 2016.

J Fernández de Cañete, C Galindo, J Barbancho, and A Luque. *Automatic control systems in biomedical engineering*. Springer, 2018.

Pierre De Handschutter, Nicolas Gillis, and Xavier Siebert. A survey on deep matrix factorizations. *Computer Science Review*, 42:100423, 2021.

Giulia De Pasquale and Maria Elena Valcher. Multi-dimensional extensions of the hegselmann-krause model. In *2022 IEEE 61st Conference on Decision and Control (CDC)*, pages 3525–3530. IEEE, 2022.

Yuan Deng, Jon Schneider, and Balasubramanian Sivan. Prior-free dynamic auctions with low regret buyers. *Advances in Neural Information Processing Systems*, 32, 2019.

Persi Diaconis and Brian Skyrms. *Ten great ideas about chance*. Princeton University Press, 2018.

Xiucai Ding and Rong Ma. Kernel spectral joint embeddings for high-dimensional noisy datasets using duo-landmark integral operators. *arXiv preprint arXiv:2405.12317*, 2024.

Jeff Donahue, Yangqing Jia, Oriol Vinyals, Judy Hoffman, Ning Zhang, Eric Tzeng, and Trevor Darrell. Decaf: A deep convolutional activation feature for generic visual recognition. In Eric P. Xing and Tony Jebara, editors, *Proceedings of the 31st International Conference on Machine Learning*, volume 32 of *Proceedings of Machine Learning Research*, pages 647–655, Bejing, China, 22–24 Jun 2014. PMLR. URL `https://proceedings.mlr.press/v32/donahue14.html`.

Tommaso d'Orsi, Gleb Novikov, and David Steurer. Consistent regression when oblivious outliers overwhelm. In *International Conference on Machine Learning*, pages 2297–2306. PMLR, 2021.

Tal Einav and Brian Cleary. Extrapolating missing antibody-virus measurements across serological studies. *Cell Systems*, 13(7):561–573, 2022.

Larry Eisenberg and Thomas H Noe. Systemic risk in financial systems. *Management Science*, 47(2):236–249, 2001.

Andrea L. Eisfeldt, Bernard Herskovic, Sriram Rajan, and Emil Siriwardane. OTC intermediaries. Research Paper 18-05, Office of Financial Research, August 2021.

Andrea L Eisfeldt, Bernard Herskovic, Sriram Rajan, and Emil Siriwardane. Otc intermediaries. *The Review of Financial Studies*, 36(2):615–677, 2023.

Ahmed El-Kishky, Thomas Markovich, Serim Park, Chetan Verma, Baekjin Kim, Ramy Eskander, Yury Malkov, Frank Portman, Sofía Samaniego, Ying Xiao, et al. Twhin: Embedding the twitter heterogeneous information network for personalized recommendation. In *Proceedings of the 28th ACM SIGKDD conference on knowledge discovery and data mining*, pages 2842–2850, 2022.

Matthew Elliott and Benjamin Golub. Networks and economic fragility. *Annual Review of Economics*, 14:665–696, 2022.

Matthew Elliott, Benjamin Golub, and Matthew O Jackson. Financial networks and contagion. *American Economic Review*, 104(10):3115–53, 2014.

Matthew Elliott, Benjamin Golub, and Matthew V Leduc. Supply network formation and fragility. *American Economic Review*, 112(8):2701–47, 2022a.

Matthew Elliott, Benjamin Golub, and Matthew V Leduc. Supply network formation and fragility. *American Economic Review*, 112(8):2701–2747, 2022b.

P. Erdös and A. Rényi. On random graphs I. *Publicationes Mathematicae*, 6:290–297, 1959.

Paul Erdos, Alfréd Rényi, et al. On the evolution of random graphs. *Publ. math. inst. hung. acad. sci*, 5(1):17–60, 1960.

Oscar Fajardo-Fontiveros, Ignasi Reichardt, Harry R De Los Ríos, Jordi Duch, Marta Sales-Pardo, and Roger Guimerà. Fundamental limits to learning closed-form mathematical models from data. *Nature Communications*, 14(1):1043, 2023.

Eugene F Fama and Kenneth R French. A five-factor asset pricing model. *Journal of financial economics*, 116(1):1–22, 2015.

Jason Fan, Anthony Cannistra, Inbar Fried, Tim Lim, Thomas Schaffner, Mark Crovella, Benjamin Hescott, and Mark DM Leiserson. Functional protein representations from biological networks enable diverse cross-species inference. *Nucleic acids research*, 47(9): e51–e51, 2019.

Ferric C Fang and Arturo Casadevall. *Thinking about science: good science, bad science, and how to make it better*. John Wiley & Sons, 2023.

Zachary Feinstein. Capital regulation under price impacts and dynamic financial contagion. *European Journal of Operational Research*, 281(2):449–463, 2020.

Zachary Feinstein and Andreas Søjmark. Dynamic default contagion in heterogeneous interbank systems. *SIAM Journal on Financial Mathematics*, 12(4):SC83–SC97, 2021.

Zachary Feinstein and Andreas Søjmark. Endogenous network valuation adjustment and the systemic term structure in a dynamic interbank model. *arXiv preprint arXiv:2211.15431*, 2022.

Zachary Feinstein, Weijie Pang, Birgit Rudloff, Eric Schaanning, Stephan Sturm, and Mackenzie Wildman. Sensitivity of the eisenberg–noe clearing vector to individual interbank liabilities. *SIAM Journal on Financial Mathematics*, 9(4):1286–1325, 2018.

Iván Fernández-Val, Hugo Freeman, and Martin Weidner. Low-rank approximations of nonseparable panel models. *The Econometrics Journal*, 24(2):C40–C77, 2021.

Richard Fitzpatrick. *Maxwell's Equations and the Principles of Electromagnetism*. Jones & Bartlett Publishers, 2008.

Dimitris Fotakis, Dimitris Palyvos-Giannas, and Stratis Skoulakis. Opinion dynamics with local interactions. In *IJCAI*, pages 279–285, 2016.

Dimitris Fotakis, Vardis Kandiros, Vasilis Kontonis, and Stratis Skoulakis. Opinion dynamics with limited information. *Algorithmica*, 85(12):3855–3888, 2023.

Ove Frank and David Strauss. Markov Graphs. *Journal of the American Statistical Association*, 81(395):832–842, 1986.

Noah E Friedkin and Eugene C Johnsen. Social influence and opinions. *Journal of Mathematical Sociology*, 15(3-4):193–206, 1990.

Jason Gaitonde, Jon Kleinberg, and Eva Tardos. Adversarial perturbations of opinion dynamics in networks. In *Proceedings of the 21st ACM Conference on Economics and Computation*, pages 471–472, 2020a.

Jason Gaitonde, Jon Kleinberg, and Eva Tardos. Adversarial perturbations of opinion dynamics in networks. In *Proceedings of the 21st ACM Conference on Economics and Computation*, pages 471–472, 2020b.

Jason Gaitonde, Jon Kleinberg, and Éva Tardos. Polarization in geometric opinion dynamics. In *Proceedings of the 22nd ACM Conference on Economics and Computation*, pages 499–519, 2021.

Andrea Galeotti, Benjamin Golub, and Sanjeev Goyal. Targeting interventions in networks. *Econometrica*, 88(6):2445–2471, 2020.

Axel Gandy and Luitgard AM Veraart. A bayesian methodology for systemic risk assessment in financial networks. *Management Science*, 63(12):4428–4446, 2017.

Chao Gao, Yu Lu, and Harrison H Zhou. Rate-optimal graphon estimation. *The Annals of Statistics*, pages 2624–2652, 2015.

Yuan Gao, Laurence T Yang, Jing Yang, Dehua Zheng, and Yaliang Zhao. Jointly low-rank tensor completion for estimating missing spatiotemporal values in logistics systems. *IEEE Transactions on Industrial Informatics*, 19(2):1814–1822, 2022.

Vikas Garg and Tommi Jaakkola. Learning tree structured potential games. In D. Lee, M. Sugiyama, U. Luxburg, I. Guyon, and R. Garnett, editors, *Advances in Neural Information Processing Systems*, volume 29. Curran Associates, Inc., 2016. URL `https://proceedings.neurips.cc/paper_files/paper/2016/file/22ac3c5a5bf0b520d281c122d1490650pdf`.

Helyette Geman, Dilip B Madan, and Marc Yor. Asset prices are brownian motion: only in business time. In *Quantitative Analysis In Financial Markets: Collected Papers of the New York University Mathematical Finance Seminar (Volume II)*, pages 103–146, 2001.

Ganesh Ghalme, Vineet Nair, Itay Eilat, Inbal Talgam-Cohen, and Nir Rosenfeld. Strategic classification in the dark. In *International Conference on Machine Learning*, pages 3672–3681. PMLR, 2021.

Allan Gibbard. Manipulation of voting schemes: a general result. *Econometrica: journal of the Econometric Society*, pages 587–601, 1973.

E. N. Gilbert. Random Graphs. *The Annals of Mathematical Statistics*, 30(4): 1141–1144, 1959.

Mary Louise Gill. Matter and flux in plato's" timaeus". *Phronesis*, pages 34–53, 1987.

Vasilis Gkatzelis, Kostas Kollias, Alkmini Sgouritsa, and Xizhi Tan. Improved price of anarchy via predictions. In *Proceedings of the 23rd ACM Conference on Economics and Computation*, pages 529–557, 2022.

Paul Glasserman and H Peyton Young. How likely is contagion in financial networks? *Journal of Banking & Finance*, 50:383–399, 2015.

Paul Glasserman and H Peyton Young. Contagion in financial networks. *Journal of Economic Literature*, 54(3):779–831, 2016.

Yuqi Gu, Zhongyuan Lyu, and Kaizheng Wang. Adaptive transfer clustering: A unified framework. *arXiv preprint arXiv:2410.21263*, 2024.

Leying Guan and Robert Tibshirani. Prediction and outlier detection in classification problems. *Journal of the Royal Statistical Society. Series B, Statistical Methodology*, 84(2):524, 2022.

Meichen Guo and Claudio De Persis. Linear quadratic network games with dynamic players: Stabilization and output convergence to Nash equilibrium. *Automatica*, 130: 109711, 2021.

Venkatesan Guruswami, Atri Rudra, and Madhu Sudan. Essential coding theory. 2019.

Wenjuan Han, Bo Pang, and Ying Nian Wu. Robust transfer learning with pretrained language models through adapters. In *Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing (Volume 2: Short Papers)*, pages 854–861, 2021.

Mark S. Handcock, Adrian E. Raftery, and Jeremy M. Tantrum. Model-based clustering for social networks. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, 170(2):301–354, March 2007. ISSN 1467-985X. doi: 10.1111/j.1467-985X.2007.00471.x. URL `http://onlinelibrary.wiley.com/doi/10.1111/j.1467-985X.2007.00471.x/abstract`.

Steve Hanneke and Samory Kpotufe. On the value of target data in transfer learning. *Advances in Neural Information Processing Systems*, 32, 2019.

Steve Hanneke and Samory Kpotufe. A no-free-lunch theorem for multitask learning. *The Annals of Statistics*, 50(6):3119–3143, 2022.

Moritz Hardt, Nimrod Megiddo, Christos Papadimitriou, and Mary Wootters. Strategic classification. In *Proceedings of the 2016 ACM conference on Innovations in Theoretical Computer Science (ITCS)*, pages 111–122, 2016.

Keegan Harris, Chara Podimata, and Steven Z Wu. Strategic apple tasting. *Advances in Neural Information Processing Systems*, 36:79918–79945, 2023.

Christopher A Harrison and S Joe Qin. Minimum variance performance map for constrained model predictive control. *Journal of Process Control*, 19(7):1199–1204, 2009.

Allen Hatcher. *Algebraic Topology*. Cambridge University Press, 2002.

Jan Hązła, Yan Jin, Elchanan Mossel, and Govind Ramnarayan. A geometric model of opinion polarization. *arXiv preprint arXiv:1910.05274*, 2019.

Yong He, Zeyu Li, Dong Liu, Kangxiang Qin, and Jiahui Xie. Representational transfer learning for matrix completion. *arXiv preprint arXiv:2412.06233*, 2024.

Jürgen Hedderich and Lothar Sachs. Random variables, distributions. In *Applied Statistics*, pages 201–324. Springer, 2024.

Rainer Hegselmann, Ulrich Krause, et al. Opinion dynamics and bounded confidence models, analysis, and simulation. *Journal of artificial societies and social simulation*, 5 (3), 2002.

Tim Hellmann. On the existence and uniqueness of pairwise stable networks. *International Journal of Game Theory*, 42(1):211–237, 2013.

Bernard Herskovic. Networks in production: Asset pricing implications. *The Journal of Finance*, 73(4):1785–1818, 2018.

Wassily Hoeffding. Probability inequalities for sums of bounded random variables. *The collected works of Wassily Hoeffding*, pages 409–426, 1994.

Peter Hoff. Modeling homophily and stochastic equivalence in symmetric relational data. *Advances in neural information processing systems*, 20, 2007.

Peter D. Hoff, Adrian E. Raftery, and Mark S. Handcock. Latent space approaches to social network analysis. *Journal of the American Statistical Association*, 97(460): 1090–1098, December 2002a.

Peter D Hoff, Adrian E Raftery, and Mark S Handcock. Latent Space Approaches to Social Network Analysis. *Journal of the American Statistical Association*, 97(460):1090–1098, December 2002b. ISSN 0162-1459, 1537-274X. doi: 10.1198/016214502388618906. URL `http://www.tandfonline.com/doi/abs/10.1198/016214502388618906`.

Paul W. Holland, Kathryn Blackmond Laskey, and Samuel Leinhardt. Stochastic blockmodels: First steps. *Social Networks*, 5(2):109–137, June 1983.

Douglas N. Hoover. *Relations on probability spaces arrays of random variables*. Institute for Advanced Study,RI, 1979.

Roger Horn and Charles Johnson. *Topics in matrix analysis*. Cambridge University Press, 2008.

Roger A Horn and Charles R Johnson. *Topics in Matrix Analysis*. Cambridge University Press, 1994.

Roger A Horn and Charles R Johnson. *Matrix analysis*. Cambridge university press, 2012.

Taishan Hu, Nilesh Chitnis, Dimitri Monos, and Anh Dinh. Next-generation sequencing technologies: An overview. *Human Immunology*, 82(11):801–811, 2021.

Jui-Ting Huang, Jinyu Li, Dong Yu, Li Deng, and Yifan Gong. Cross-language knowledge transfer using multilingual deep neural network with shared hidden layers. In *2013 IEEE international conference on acoustics, speech and signal processing*, pages 7304–7308. IEEE, 2013.

Yunhan Huang and Quanyan Zhu. Game-theoretic frameworks for epidemic spreading and human decision-making: A review. *Dynamic Games and Applications*, 12(1):7–48, 2022.

Cynthia Huber, Tim Friede, Julia Stingl, and Norbert Benda. Classification of companion diagnostics: a new framework for biomarker-driven patient selection. *Therapeutic Innovation & Regulatory Science*, pages 1–11, 2022.

Minyoung Huh, Pulkit Agrawal, and Alexei A Efros. What makes imagenet good for transfer learning? *arXiv preprint arXiv:1608.08614*, 2016.

David R. Hunter and Mark S. Handcock. Inference in Curved Exponential Family Models for Networks. *Journal of Computational and Graphical Statistics*, 15(3):565–583, 2006.

Jacopo Iacovacci and Ginestra Bianconi. Extracting information from multiplex networks. *Chaos: An Interdisciplinary Journal of Nonlinear Science*, 26(6), 2016.

Mohammad T. Irfan and Luis E. Ortiz. On influence, stable behavior, and the most influential individuals in networks: A game-theoretic approach. *Artificial Intelligence*, 215:79–119, 2014. ISSN 0004-3702. doi: https://doi.org/10.1016/j.artint.2014.06.004. URL `https://www.sciencedirect.com/science/article/pii/S0004370214000812`.

ISDA. Key trends in the size and composition of otc derivatives markets in the second half of 2022. Technical report, International Swaps and Derivatives Association, June 2023.

Matthew O Jackson. Mechanism theory, 2000.

Matthew O Jackson and Agathe Pernoud. Systemic risk in financial networks: A survey. *Annual Review of Economics*, 13:171–202, 2021.

Matthew O Jackson and Anne Van den Nouweland. Strongly stable networks. *Games and Economic Behavior*, 51(2):420–444, 2005.

Matthew O Jackson and Asher Wolinsky. A strategic model of social and economic networks. In *Networks and groups*, pages 23–49. 2003.

Jafar Jafarov. Survey of matrix completion algorithms. *arXiv preprint arXiv:2204.01532*, 2022.

Ali Jahanian, Xavier Puig, Yonglong Tian, and Phillip Isola. Generative models as a data source for multiview representation learning. In *International Conference on Learning Representations*, 2022.

Prateek Jain, Praneeth Netrapalli, and Sujay Sanghavi. Low-rank matrix completion using alternating minimization. In *Proceedings of the forty-fifth annual ACM symposium on Theory of computing*, pages 665–674, 2013.

Akhil Jalan and Deepayan Chakrabarti. Strategic negotiations in endogenous network formation, 2024. URL `https://arxiv.org/abs/2402.08779`.

Akhil Jalan and Marios Papachristou. Opinion dynamics with multiple adversaries. *arXiv preprint arXiv:2502.15931*, 2025.

Akhil Jalan, Deepayan Chakrabarti, and Purnamrita Sarkar. Incentive-aware models of financial networks. *Operations Research*, 72(6):2321–2336, 2024a.

Akhil Jalan, Arya Mazumdar, Soumendu Sundar Mukherjee, and Purnamrita Sarkar. Transfer learning for latent variable network models. In *The Thirty-eighth Annual Conference on Neural Information Processing Systems*, 2024b. URL `https://openreview.net/forum?id=PK8xOCBQRO`.

Akhil Jalan, Yassir Jedra, Arya Mazumdar, Soumendu Sundar Mukherjee, and Purnamrita Sarkar. Optimal transfer learning for missing not-at-random matrix completion, 2025. URL `https://arxiv.org/abs/2503.00174`.

Yassir Jedra, Sean Mann, Charlotte Park, and Devavrat Shah. Exploiting observation bias to improve matrix completion. *arXiv preprint arXiv:2306.04775*, 2023.

Sayash Kapoor, Kumar Kshitij Patel, and Purushottam Kar. Corruption-tolerant bandit learning. *Machine Learning*, 108(4):687–715, 2019.

Stuart A Kauffman. *A world beyond physics: the emergence and evolution of life*. Oxford University Press, 2019.

Michael Kearns, Michael L Littman, and Satinder Singh. Graphical models for game theory. In *Proceedings of the Seventeenth conference on Uncertainty in artificial intelligence*, pages 253–260, 2001.

Jack Kiefer and Jacob Wolfowitz. The equivalence of two extremum problems. *Canadian Journal of Mathematics*, 12:363–366, 1960.

Bomin Kim, Kevin H. Lee, Lingzhou Xue, and Xiaoyue Niu. A review of dynamic network models with latent variables. *Statistics Surveys*, 12(none):105 – 135, 2018. doi: 10.1214/18-SS121. URL `https://doi.org/10.1214/18-SS121`.

H.E. Kim, A. Cosa-Linan, N. Santhanam, and et al. Transfer learning for medical image classification: a literature review. *BMC Medical Imaging*, 22(1):69, 2022. doi: 10.1186/s12880-022-00793-7. URL `https://doi.org/10.1186/s12880-022-00793-7`.

Miles S Kimball, Claudia R Sahm, and Matthew D Shapiro. Imputing risk tolerance from survey responses. *Journal of the American statistical Association*, 103(483): 1028–1038, 2008.

Zachary A King, Justin Lu, Andreas Dräger, Philip Miller, Stephen Federowicz, Joshua A Lerman, Ali Ebrahim, Bernhard O Palsson, and Nathan E Lewis. Bigg models: A platform for integrating, standardizing and sharing genome-scale models. *Nucleic acids research*, 44(D1):D515–D522, 2016.

Olga Klopp, Alexandre B Tsybakov, and Nicolas Verzelen. Oracle inequalities for network models and sparse graphon estimation. *Annals of Statistics*, 45(1):316–354, 2017.

Yoav Kolumbus and Noam Nisan. How and why to manipulate your own agent: On the incentives of users of learning agents. *Advances in Neural Information Processing Systems*, 35:28080–28094, 2022a.

Yoav Kolumbus and Noam Nisan. Auctions between regret-minimizing agents. In *Proceedings of the ACM Web Conference 2022*, pages 100–111, 2022b.

Yoav Kolumbus, Menahem Levy, and Noam Nisan. Asynchronous proportional response dynamics: convergence in markets with adversarial scheduling. *Advances in Neural Information Processing Systems*, 36:25409–25434, 2023.

Yoav Kolumbus, Menahem Levy, and Noam Nisan. Asynchronous proportional response dynamics: convergence in markets with adversarial scheduling. *Advances in Neural Information Processing Systems*, 36, 2024.

Steve G Kou. Jump-diffusion models for asset pricing in financial engineering. *Handbooks in operations research and management science*, 15:73–116, 2007.

Meghana Kshirsagar. *Combine and conquer: methods for multitask learning in biology and language*. PhD thesis, Carnegie Mellon University, 2015.

Meghana Kshirsagar, Jaime Carbonell, and Judith Klein-Seetharaman. Multitask learning for host–pathogen protein interactions. *Bioinformatics*, 29(13):i217–i226, 2013.

Tor Lattimore and Csaba Szepesvári. *Bandit algorithms*. Cambridge University Press, 2020a.

Tor Lattimore and Csaba Szepesvári. *Bandit algorithms*. Cambridge University Press, 2020b.

Tor Lattimore, Csaba Szepesvari, and Gellert Weisz. Learning with good feature representations in bandits and in rl with a generative model. In *International conference on machine learning*, pages 5662–5670. PMLR, 2020.

Jaekoo Lee, Hyunjae Kim, Jongsun Lee, and Sungroh Yoon. Transfer learning for deep learning on graph-structured data. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 31, 2017.

James R Lee, Shayan Oveis Gharan, and Luca Trevisan. Multiway spectral partitioning and higher-order cheeger inequalities. *Journal of the ACM (JACM)*, 61(6):1–30, 2014a.

Kyu-Min Lee, Jung Yeol Kim, Sangchul Lee, and K-I Goh. Multiplex networks. *Networks of networks: The last frontier of complexity*, pages 53–72, 2014b.

Kyu-Min Lee, Byungjoon Min, and Kwang-Il Goh. Towards real-world complexity: an introduction to multiplex networks. *The European Physical Journal B*, 88:1–20, 2015.

Yan Leng, Xiaowen Dong, Junfeng Wu, and Alex Pentland. Learning quadratic games on networks. In *International Conference on Machine Learning*, pages 5820–5830, 2020a.

Yan Leng, Xiaowen Dong, Junfeng Wu, and Alex Pentland. Learning quadratic games on networks. In *International Conference on Machine Learning*, pages 5820–5830. PMLR, 2020b.

Jure Leskovec and Andrej Krevl. SNAP Datasets: Stanford large network dataset collection. `http://snap.stanford.edu/data`, June 2014.

Keith Levin, Asad Lodhia, and Elizaveta Levina. Recovering shared structure from multiple networks with unknown edge distributions. *Journal of machine learning research*, 23(3):1–48, 2022.

Jingqi Li, Somayeh Sojoudi, Claire J Tomlin, and David Fridovich-Keil. The computation of approximate feedback stackelberg equilibria in multiplayer nonlinear constrained dynamic games. *SIAM Journal on Optimization*, 34(4):3723–3749, 2024.

Lechuan Li, Ruth Dannenfelser, Yu Zhu, Nathaniel Hejduk, Santiago Segarra, and Vicky Yao. Joint embedding of biological networks for cross-species functional alignment. *bioRxiv*, pages 2022–01, 2022.

YZ Li, QH Wu, MS Li, and JP Zhan. Mean-variance model for power system economic dispatch with wind power integrated. *Energy*, 72:510–520, 2014.

Jennifer Listgarten. The perpetual motion machine of ai-generated data and the distraction of chatgpt as a 'scientist'. *Nature Biotechnology*, 42(3):371–373, 2024.

Jonathan Long, Evan Shelhamer, and Trevor Darrell. Fully convolutional networks for semantic segmentation. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 3431–3440, 2015.

Kerstin Lopatta, Magdalena Tchikov, and Finn Marten Körner. The impact of market sectors and rating agencies on credit ratings: global evidence. *The Journal of Risk Finance*, 20(5):389–410, 2019.

László Lovász. *Large Networks and Graph Limits.*, volume 60 of *Colloquium Publications*. American Mathematical Society, 2012. ISBN 978-0-8218-9085-1.

Cong Ma, Reese Pathak, and Martin J Wainwright. Optimally tackling covariate shift in rkhs-based nonparametric regression. *The Annals of Statistics*, 51(2):738–761, 2023a.

Shuai Ma, Xiaoteng Ma, and Li Xia. A unified algorithm framework for mean-variance optimization in discounted markov decision processes. *European Journal of Operational Research*, 311(3):1057–1067, 2023b. ISSN 0377-2217. doi: https://doi.org/10.1016/j.ejor.2023.06.022. URL https://www.sciencedirect.com/science/article/pii/S0377221723004757.

Wei Ma and George H Chen. Missing not at random in matrix completion: The effectiveness of estimating missingness probabilities under a low nuclear norm assumption. *Advances in neural information processing systems*, 32, 2019.

Yishay Mansour, Mehryar Mohri, and Afshin Rostamizadeh. Domain adaptation: Learning bounds and algorithms. *arXiv preprint arXiv:0902.3430*, 2009.

Xueyu Mao, Purnamrita Sarkar, and Deepayan Chakrabarti. Overlapping clustering models, and one (class) SVM to bind them all. In *Advances in Neural Information Processing Systems*, volume 31, 2018.

Xueyu Mao, Deepayan Chakrabarti, and Purnamrita Sarkar. Consistent nonparametric methods for network assisted covariate estimation. In *International Conference on Machine Learning*, pages 7435–7446. PMLR, 2021.

Adam W Marcus, Daniel A Spielman, and Nikhil Srivastava. Interlacing families iii: Sharper restricted invertibility estimates. *Israel Journal of Mathematics*, pages 1–28, 2022.

Harry Markowitz. Portfolio selection. *Journal of Finance*, 7(1):77–91, 1952. URL https://EconPapers.repec.org/RePEc:bla:jfinan:v:7:y:1952:i:1:p:77-91.

I. Mastromatteo, E. Zarinelli, and M. Marsili. Reconstruction of financial networks for robust estimation of systemic risk. *Journal of Statistical Mechanics: Theory and Experiment*, 2012.

Vladimir Mazalov and Julia V Chirkova. *Networking games: network forming games and games on networks*. Academic Press, 2019.

Arya Mazumdar and Barna Saha. Query complexity of clustering with side information. *Advances in Neural Information Processing Systems*, 30, 2017a.

Arya Mazumdar and Barna Saha. Clustering with noisy queries. *Advances in Neural Information Processing Systems*, 30, 2017b.

Sean McGrath, Cenhao Zhu, Min Guo, and Rui Duan. Learner: A transfer learning method for low-rank matrix estimation. *arXiv preprint arXiv:2412.20605*, 2024.

Andrew Metrick. A natural experiment in "Jeopardy!". *The American Economic Review*, pages 240–253, 1995.

Elchanan Mossel and Jiaming Xu. Local algorithms for block models with side information. In *Proceedings of the 2016 ACM Conference on Innovations in Theoretical Computer Science*, pages 71–80, 2016.

Robert S Mueller. United states of america v. *Internet Research Agency. Case*, 2018.

Soumendu Sundar Mukherjee and Sayak Chakrabarti. Graphon estimation from partially observed network data. *arXiv preprint arXiv:1906.00494*, 2019.

Bhaskar Mukhoty, Debojyoti Dey, and Purushottam Kar. Corruption-tolerant algorithms for generalized linear models. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 37, pages 9243–9250, 2023.

Cameron Musco, Christopher Musco, and Charalampos E Tsourakakis. Minimizing polarization and disagreement in social networks. In *Proceedings of the 2018 World Wide Web Conference*, pages 369–378, 2018a.

Cameron Musco, Christopher Musco, and Charalampos E Tsourakakis. Minimizing polarization and disagreement in social networks. In *Proceedings of the 2018 world wide web conference*, pages 369–378, 2018b.

Angelia Nedić and Behrouz Touri. Multi-dimensional hegselmann-krause dynamics. In *2012 IEEE 51st IEEE Conference on Decision and Control (CDC)*, pages 68–73. IEEE, 2012.

Mark Newman. *Networks*. Oxford university press, 2018.

Isaac Newton. *Philosophiae naturalis principia mathematica*. 1687.

Behnam Neyshabur, Hanie Sedghi, and Chiyuan Zhang. What is being transferred in transfer learning? In H. Larochelle, M. Ranzato, R. Hadsell, M.F. Balcan, and H. Lin, editors, *Advances in Neural Information Processing Systems*, volume 33, pages 512–523. Curran Associates, Inc., 2020. URL `https://proceedings.neurips.cc/paper_files/paper/2020/file/0607f4c705595b911a4f3e7a127b44e0-Paper.pdf`.

Thanh H Nguyen, Yongzhao Wang, Arunesh Sinha, and Michael P Wellman. Deception in finitely repeated security games. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 33, pages 2133–2140, 2019.

Noam Nisan, Tim Roughgarden, Eva Tardos, and Vijay V Vazirani. *Algorithmic game theory*. Cambridge University Press, 2007.

Charles J Norsigian, Neha Pusarla, John Luke McConn, James T Yurkovich, Andreas Dräger, Bernhard O Palsson, and Zachary King. Bigg models 2020: multi-strain genome-scale models and expansion across the phylogenetic tree. *Nucleic acids research*, 48(D1):D402–D406, 2020.

OECD. OECD statistics, 2022. URL `https://stats.oecd.org/`.

Sofia C Olhede and Patrick J Wolfe. Network histograms and universality of blockmodel approximation. *Proceedings of the National Academy of Sciences*, 111(41):14722–14727, 2014.

Larry Page, Sergey Brin, Rajeev Motwani, and Terry Winograd. The pagerank citation ranking: Bringing order to the web. Technical report, Technical Report, Stanford InfoLab, 1999.

Marios Papachristou and Jon Kleinberg. Allocating stimulus checks in times of crisis. In *Proceedings of the ACM Web Conference 2022*, pages 16–26, 2022.

Maneesha Papireddygari and Bo Waggoner. Contracts with information acquisition, via scoring rules. In *Proceedings of the 23rd ACM Conference on Economics and Computation*, pages 703–704, 2022.

Ashwin Paranjape, Austin R Benson, and Jure Leskovec. Motifs in temporal networks. In *Proceedings of the tenth ACM international conference on web search and data mining*, pages 601–610, 2017.

Daniel Paravisini, Veronica Rappoport, and Enrichetta Ravina. Risk aversion and wealth: Evidence from person-to-person lending portfolios. *Management Science*, 63 (2):279–297, 2017.

Eli Pariser. *The filter bubble: How the new personalized web is changing what we read and how we think*. Penguin, 2011.

Chanwoo Park, Kaiqing Zhang, and Asuman Ozdaglar. Multi-player zero-sum markov games with networked separable interactions. *Advances in Neural Information Processing Systems*, 36, 2024.

Grant P Parnell, Benjamin M Tang, Marek Nalos, Nicola J Armstrong, Stephen J Huang, David R Booth, and Anthony S McLean. Identifying key regulatory genes in the whole blood of septic patients to monitor underlying immune dysfunctions. *Shock*, 40(3):166–174, 2013.

Daniel Paulin, Lester Mackey, and Joel A Tropp. Efron–stein inequalities for random matrices. 2016.

Tuomas A. Peltonen, Martin Scheicher, and Guillaume Vuillemey. The network structure of the CDS market and its determinants. *Journal of Financial Stability*, 13: 118–133, 2014.

Yury Polyanskiy and Yihong Wu. *Information theory: From coding to learning*. Cambridge university press, 2024.

Friedrich Pukelsheim. *Optimal design of experiments*. SIAM, 2006.

Ziyue Qiao, Xiao Luo, Meng Xiao, Hao Dong, Yuanchun Zhou, and Hui Xiong. Semi-supervised domain adaptation in graph transfer learning. *ArXiv*, abs/2309.10773, 2023. URL `https://api.semanticscholar.org/CorpusID:260859484`.

Miklos Z Racz and Daniel E Rigobon. Towards consensus: Reducing polarization by perturbing social networks. *arXiv preprint arXiv:2206.08996*, 2022.

Miklos Z Rácz and Daniel E Rigobon. Towards consensus: Reducing polarization by perturbing social networks. *IEEE Transactions on Network Science and Engineering*, 10(6):3450–3464, 2023.

Piyush Rai, Avishek Saha, Hal Daumé III, and Suresh Venkatasubramanian. Domain adaptation meets active learning. In *Proceedings of the NAACL HLT 2010 Workshop on Active Learning for Natural Language Processing*, pages 27–32, 2010.

Henry WJ Reeve, Timothy I Cannings, and Richard J Samworth. Adaptive transfer learning. *The Annals of Statistics*, 49(6):3618–3649, 2021.

Mingyang Ren, Sanguo Zhang, and Junhui Wang. Consistent estimation of the number of communities via regularized network embedding. *Biometrics*, 79(3):2404–2416, 2023.

Dragos Ristache, Fabian Spaeh, and Charalampos E Tsourakakis. Wiser than the wisest of crowds: The asch effect and polarization revisited. In *Joint European Conference on Machine Learning and Knowledge Discovery in Databases*, pages 440–458. Springer, 2024.

Dragos Ristache, Fabian Spaeh, and Charalampos E Tsourakakis. Countering election sway: Strategic algorithms in friedkin-johnsen dynamics. *arXiv preprint arXiv:2502.01874*, 2025.

Karl Rohe, Sourav Chatterjee, and Bin Yu. Spectral clustering and the high-dimensional stochastic blockmodel. *The Annals of Statistics*, 39(4):1878–1915, 2011.

Emanuele Rossi, Federico Monti, Yan Leng, Michael Bronstein, and Xiaowen Dong. Learning to infer structures of network games. In *International Conference on Machine Learning*, pages 18809–18827. PMLR, 2022.

Tim Roughgarden. *Selfish routing and the price of anarchy*. MIT press, 2005.

Tim Roughgarden. The price of anarchy in games of incomplete information. *ACM Transactions on Economics and Computation (TEAC)*, 3(1):1–20, 2015.

Tim Roughgarden and Florian Schoppmann. Local smoothness and the price of anarchy in atomic splittable congestion games. In *Proceedings of the Twenty-Second Annual ACM-SIAM Symposium on Discrete Algorithms*, pages 255–267. SIAM, 2011.

Patrick Rubin-Delanchy, Joshua Cape, Minh Tang, and Carey E Priebe. A statistical interpretation of spectral embedding: the generalised random dot product graph. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, 84(4):1446–1473, 2022.

Natali Ruchansky, Mark Crovella, and Evimaria Terzi. Matrix completion with queries. In *Proceedings of the 21th ACM SIGKDD international conference on knowledge discovery and data mining*, pages 1025–1034, 2015.

Mark Rudelson and Roman Vershynin. Hanson-Wright inequality and sub-gaussian concentration. *Electronic Communications in Probability*, 18(none):1 – 9, 2013. doi: 10.1214/ECP.v18-2865. URL `https://doi.org/10.1214/ECP.v18-2865`.

Alessio Russo. Analysis and detectability of offline data poisoning attacks on linear dynamical systems. In *Learning for Dynamics and Control Conference*, pages 1086–1098. PMLR, 2023.

Evan Sadler and Benjamin Golub. Games on endogenous networks. *arXiv preprint arXiv:2102.01587*, 2021.

Francisco J Samaniego. *A comparison of the Bayesian and frequentist approaches to estimation*, volume 24. Springer, 2010.

IW Sandberg and AN Willson, Jr. Existence and uniqueness of solutions for the equations of nonlinear DC networks. *SIAM Journal on Applied Mathematics*, 22(2): 173–186, 1972.

Purnamrita Sarkar and Andrew Moore. Dynamic social network analysis using latent space models. In Y. Weiss, B. Schölkopf, and J. Platt, editors, *Advances in Neural Information Processing Systems*, volume 18. MIT Press, 2005. URL https://proceedings.neurips.cc/paper_files/paper/2005/file/ec8b57b0be908301f5748fb04b0714c7-Paper.pdf.

Purnamrita Sarkar, Deepayan Chakrabarti, and Michael I. Jordan. Nonparametric link prediction in dynamic networks. In *Proceedings of the 29th International Coference on International Conference on Machine Learning*, ICML'12, page 1897–1904, Madison, WI, USA, 2012. Omnipress. ISBN 9781450312851.

Mark Allen Satterthwaite. Strategy-proofness and arrow's conditions: Existence and correspondence theorems for voting procedures and social welfare functions. *Journal of economic theory*, 10(2):187–217, 1975.

George AF Seber and Alan J Lee. *Linear regression analysis*. John Wiley & Sons, 2012.

Rishika Sen, Somnath Tagore, and Rajat K De. Asapp: Architectural similarity-based automated pathway prediction system and its application in host-pathogen interactions. *IEEE/ACM Transactions on Computational Biology and Bioinformatics*, 17(2):506–515, 2018.

Daniel K Sewell and Yuguo Chen. Latent space models for dynamic networks. *Journal of the american statistical association*, 110(512):1646–1657, 2015.

Elisa Shearer and Amy Mitchell. News use across social media platforms in 2020. 2021.

Yusif Simaan. The opportunity cost of mean–variance choice under estimation risk. *European Journal of Operational Research*, 234(2):382–391, 2014.

Max Simchowitz, Abhishek Gupta, and Kaiqing Zhang. Tackling combinatorial distribution shift: A matrix completion perspective. In Gergely Neu and Lorenzo Rosasco, editors, *Proceedings of Thirty Sixth Conference on Learning Theory*, volume 195 of *Proceedings of Machine Learning Research*, pages 3356–3468. PMLR, 12–15 Jul 2023a. URL https://proceedings.mlr.press/v195/simchowitz23a.html.

Max Simchowitz, Abhishek Gupta, and Kaiqing Zhang. Tackling combinatorial distribution shift: A matrix completion perspective. In *The Thirty Sixth Annual Conference on Learning Theory*, pages 3356–3468. PMLR, 2023b.

Kirstine Smith. On the standard deviations of adjusted and interpolated values of an observed polynomial function and its constants and the guidance they give towards a proper choice of the distribution of observations. *Biometrika*, 12(1/2):1–85, 1918.

Tiziano Squartini, Guido Caldarelli, Giulio Cimini, Andrea Gabrielli, and Diego Garlaschelli. Reconstruction methods for networks: the case of economic and financial systems. *Physics Reports*, 757:1–47, 2018.

Tracy M Sweet, Andrew C Thomas, and Brian W Junker. Hierarchical network models for education research: Hierarchical latent space models. *Journal of Educational and Behavioral Statistics*, 38(3):295–318, 2013.

Jie Tang, Tiancheng Lou, Jon Kleinberg, and Sen Wu. Transfer learning to infer social ties across heterogeneous networks. *ACM Trans. Inf. Syst.*, 34(2), apr 2016. ISSN 1046-8188. doi: 10.1145/2746230. URL `https://doi.org/10.1145/2746230`.

Eva Tardos. Network games. In *Proceedings of the thirty-sixth annual ACM symposium on Theory of computing*, pages 341–342, 2004.

Jerome Taupin, Yassir Jedra, and Alexandre Proutiere. Best policy identification in linear mdps. In *2023 59th Annual Allerton Conference on Communication, Control, and Computing (Allerton)*, pages 1–8. IEEE, 2023.

Walter Thirring. *Classical mathematical physics: dynamical systems and field theories*. Springer Science & Business Media, 2013.

Stefan Thurner, Rudolf Hanel, and Peter Klimek. *Introduction to the theory of complex systems*. Oxford University Press, 2018.

Neda Trifonova, Andrew Kenny, David Maxwell, Daniel Duplisea, Jose Fernandes, and Allan Tucker. Spatio-temporal bayesian network models with latent variables for revealing trophic dynamics and functional networks in fisheries ecology. *Ecological Informatics*, 30:142–158, 2015.

Nilesh Tripuraneni, Michael I. Jordan, and Chi Jin. On the theory of transfer learning: the importance of task diversity. In *Proceedings of the 34th International Conference on Neural Information Processing Systems*, NIPS '20, Red Hook, NY, USA, 2020. Curran Associates Inc. ISBN 9781713829546.

Rakshit Trivedi, Mehrdad Farajtabar, Prasenjeet Biswal, and Hongyuan Zha. Dyrep: Learning representations over dynamic graphs. In *International conference on learning representations*, 2019.

Andreas Tsanakas and Nicos Christofides. Risk exchange with distorted probabilities. *ASTIN Bulletin: The Journal of the IAA*, 36(1):219–243, 2006.

Alexandre B Tsybakov. *Introduction to Nonparametric Estimation*. Springer series in statistics. Springer, Dordrecht, 2009. doi: 10.1007/b13794. URL `https://cds.cern.ch/record/1315296`.

Twitter, Inc. Information operations directed at hong kong, 2019. URL `https://blog.x.com/en_us/topics/company/2019/information_operations_directed_at_Hong_Kong`. Accessed: 2024-12-27.

Eric Tzeng, Judy Hoffman, Kate Saenko, and Trevor Darrell. Adversarial discriminative domain adaptation. *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 2962–2971, 2017a. URL `https://api.semanticscholar.org/CorpusID:4357800`.

Eric Tzeng, Judy Hoffman, Kate Saenko, and Trevor Darrell. Adversarial discriminative domain adaptation. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 7167–7176, 2017b.

C. Upper and A. Worms. Estimating bilateral exposures in the German interbank market: Is there a danger of contagion? *European Economic Review*, 48:827–849, 2004.

Christian Upper. Simulation methods to assess the danger of contagion in interbank markets. *Journal of financial stability*, 7(3):111–125, 2011.

Alexander HS Vargo and Anna C Gilbert. A rank-based marker selection method for high throughput scrna-seq data. *BMC bioinformatics*, 21:1–51, 2020.

Alfredo Vellido. The importance of interpretability and visualization in machine learning for applications in medicine and health care. *Neural computing and applications*, 32 (24):18069–18083, 2020.

Sergio Verdú et al. Generalizing the fano inequality. *IEEE Transactions on Information Theory*, 40(4):1247–1251, 1994.

Roman Vershynin. *High-dimensional probability: An introduction with applications in data science*, volume 47. Cambridge University Press, 2018a.

Roman Vershynin. *High-dimensional probability: An introduction with applications in data science*, volume 47. Cambridge university press, 2018b.

Haohui Wang, Yuzhen Mao, Jianhui Sun, Si Zhang, Yonghui Fan, and Dawei Zhou. Dynamic transfer learning across graphs. *arXiv preprint arXiv:2305.00664*, 2023.

Yanbang Wang and Jon Kleinberg. On the relationship between relevance and conflict in online social link recommendations. *Advances in Neural Information Processing Systems*, 36, 2024.

Stanley Wasserman and Philippa Pattison. Logit models and logistic regressions for social networks: I. An introduction to Markov graphs and p^*. *Psychometrika*, 61(3): 401–425, 1996.

E Weinan. *Principles of multiscale modeling*. Cambridge University Press, 2011.

Karl R. Weiss, Taghi M. Khoshgoftaar, and Dingding Wang. A survey of transfer learning. *Journal of Big Data*, 3:1–40, 2016. URL `https://api.semanticscholar.org/CorpusID:3761015`.

Lawrence J White. Markets: The credit rating agencies. *Journal of Economic Perspectives*, 24(2):211–226, 2010.

Jun Wu, Lisa Ainsworth, Andrew Leakey, Haixun Wang, and Jingrui He. Graph-structured gaussian processes for transferable graph learning. *Advances in Neural Information Processing Systems*, 36, 2024.

Yun-Jhong Wu, Elizaveta Levina, and Ji Zhu. Link prediction for egocentrically sampled networks. *Journal of Computational and Graphical Statistics*, 32:1296 – 1319, 2018. URL `https://api.semanticscholar.org/CorpusID:3859958`.

Zonghan Wu, Shirui Pan, Fengwen Chen, Guodong Long, Chengqi Zhang, and S Yu Philip. A comprehensive survey on graph neural networks. *IEEE transactions on neural networks and learning systems*, 32(1):4–24, 2020.

Binhui Xie, Longhui Yuan, Shuang Li, Chi Harold Liu, Xinjing Cheng, and Guoren Wang. Active learning for domain adaptation: An energy-based approach. In *Proceedings of the AAAI conference on artificial intelligence*, volume 36, pages 8708–8716, 2022.

Fangzheng Xie and Yanxun Xu. Optimal bayesian estimation for random dot product graphs. *Biometrika*, 107(4):875–889, 2020.

Jiaming Xu. Rates of convergence of spectral methods for graphon estimation. In *International Conference on Machine Learning*, pages 5433–5442. PMLR, 2018.

Miao Xu, Rong Jin, and Zhi-Hua Zhou. Speedup matrix completion with side information: Application to multi-label learning. *Advances in neural information processing systems*, 26, 2013.

Menahem E Yaari. The dual theory of choice under risk. *Econometrica: Journal of the Econometric Society*, pages 95–115, 1987.

Hao Yan and Keith Levin. Minimax rates for latent position estimation in the generalized random dot product graph. *arXiv preprint arXiv:2307.01942*, 2023.

Hao Yan and Keith Levin. Coherence-free entrywise estimation of eigenvectors in low-rank signal-plus-noise matrix models. In *The Thirty-eighth Annual Conference on Neural Information Processing Systems*, 2024.

Liu Yang, Steve Hanneke, and Jaime Carbonell. A theory of transfer learning with applications to active learning. *Machine learning*, 90:161–189, 2013.

H Peyton Young. The diffusion of innovations in social networks. *The economy as an evolving complex system III: Current perspectives and future directions*, 267:39, 2006.

Bin Yu. Assouad, fano, and le cam. In *Festschrift for Lucien Le Cam: research papers in probability and statistics*, pages 423–435. Springer, 1997.

Emmanouil Zampetakis. *Statistics in high dimensions without IID samples: truncated statistics and minimax optimization.* PhD thesis, Massachusetts Institute of Technology, 2020.

Shangtong Zhang, Bo Liu, and Shimon Whiteson. Mean-variance policy iteration for risk-averse reinforcement learning. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 35, pages 10905–10913, 2021.

Yuan Zhang, Elizaveta Levina, and Ji Zhu. Estimating network edge probabilities by neighbourhood smoothing. *Biometrika*, 104(4):771–783, 2017.

Long Zhao, Deepayan Chakrabarti, and Kumar Muthuraman. Portfolio construction by mitigating error amplification: The bounded-noise portfolio. *Operations Research*, 67(4):965–983, 2019.

Kai Zhong, Zhao Song, Prateek Jain, and Inderjit S Dhillon. Provable non-linear inductive matrix completion. *Advances in Neural Information Processing Systems*, 32, 2019.

Doudou Zhou, Tianxi Cai, and Junwei Lu. Multi-source learning via completion of block-wise overlapping noisy matrices. *Journal of Machine Learning Research*, 24(221): 1–43, 2023.

Qi Zhu, Carl Yang, Yidan Xu, Haonan Wang, Chao Zhang, and Jiawei Han. Transfer learning of graph neural networks with ego-graph information maximization. *Advances in Neural Information Processing Systems*, 34:1766–1779, 2021.

Fuzhen Zhuang, Zhiyuan Qi, Keyu Duan, Dongbo Xi, Yongchun Zhu, Hengshu Zhu, Hui Xiong, and Qing He. A comprehensive survey on transfer learning. *Proceedings of the IEEE*, 109:43–76, 2019. URL `https://api.semanticscholar.org/CorpusID:207847753`.

Jungang Zou, Fan Lin, Siyu Gao, Gaoshan Deng, Wenhua Zeng, and Gil Alterovitz. Transfer learning based multi-objective genetic algorithm for dynamic community detection. *arXiv preprint arXiv:2109.15136*, 2021.

Vita

Akhil Jalan was born in California in 1997. He earned his B.A. in Applied Mathematics with highest honors from the University of California, Berkeley in 2019, where he was a Regents and Chancellor's scholar. He enrolled as a PhD student at The University of Texas at Austin in 2020. In addition to the contents of this dissertation, he has contributed to theoretical research on computational complexity theory, and applied research in bioprocessing and bioinformatics.

Address: akhiljalan@utexas.edu