Networks, Decisions, and Outcomes: Coordination with Local Information and the Value of Temporal Data for Learning Influence Networks

by

Spyridon Ilias Zoumpoulis
S.B., Massachusetts Institute of Technology (2008)
M.Eng., Massachusetts Institute of Technology (2009)

Submitted to the Department of Electrical Engineering and Computer Science
in partial fulfillment of the requirements for the degree of
Doctor of Philosophy
at the
MASSACHUSETTS INSTITUTE OF TECHNOLOGY
June 2014

Author . . . . . .
Department of Electrical Engineering and Computer Science
May 21, 2014

Certified by . . . . . .
Munther A. Dahleh
Professor of Electrical Engineering
Thesis Supervisor

Certified by . . . . . .
John N. Tsitsiklis
Professor of Electrical Engineering and Computer Science
Thesis Supervisor

Accepted by . . . . . .
Leslie A. Kolodziejski
Chairman, Committee on Graduate Theses

Networks, Decisions, and Outcomes: Coordination with Local Information and the Value of Temporal Data for Learning Influence Networks

by

Spyridon Ilias Zoumpoulis

Submitted to the Department of Electrical Engineering and Computer Science
on May 21, 2014, in partial fulfillment of the
requirements for the degree of
Doctor of Philosophy

Abstract

We study decision making by networked entities and the interplay between networks and outcomes under two different contexts: in the first part of the thesis, we study how strategic agents that share local information coordinate; in the second part of the thesis, we quantify the gain of having access to temporally richer data for learning of influence networks.

In the first part of the thesis, we study the role of local information channels in enabling coordination among strategic agents. Building on the standard finite-player global games framework, we show that the set of equilibria of a coordination game is highly sensitive to how information is locally shared among agents. In particular, we show that the coordination game has multiple equilibria if there exists a collection of agents such that (i) they do not share a common signal with any agent outside of that collection; and (ii) their information sets form an increasing sequence of nested sets, referred to as a filtration. Our characterization thus extends the results on the uniqueness and multiplicity of equilibria in global games beyond the well-known case in which agents have access to purely private or public signals. We then provide a characterization of how the extent of equilibrium multiplicity is determined by the extent to which subsets of agents have access to common information: we show that the size of the set of equilibrium strategies is increasing with the extent of variability in the size of the subsets of agents who observe the same signal. We study the set of equilibria in large coordination games, showing that as the number of agents grows, the game exhibits multiple equilibria if and only if a non-vanishing fraction of the agents have access to the same signal. We finally consider an application of our framework in which the noisy signals are interpreted to be the idiosyncratic signals of the agents, which are exchanged through a communication network.

In the second part of the thesis, we quantify the gain in the speed of learning of parametric models of influence, due to having access to richer temporal information. We infer local influence relations between networked entities from data on outcomes

and assess the value of temporal data by characterizing the speed of learning under three different types of available data: knowing the set of entities who take a particular action; knowing the order in which the entities take an action; and knowing the times of the actions. We propose a parametric model of influence which captures directed pairwise interactions and formulate different variations of the learning problem. We use the Fisher information, the Kullback-Leibler (KL) divergence, and sample complexity as measures for the speed of learning. We provide theoretical guarantees on the sample complexity for correct learning based on sets, sequences, and times. The asymptotic gain of having access to richer temporal data for the speed of learning is thus quantified in terms of the gap between the derived asymptotic requirements under different data modes. We also evaluate the practical value of learning with richer temporal data, by comparing learning with sets, sequences, and times given actual observational data. Experiments on both synthetic and real data, including data on mobile app installation behavior, and EEG data from epileptic seizure events, quantify the improvement in learning due to richer temporal data, and show that the proposed methodology recovers the underlying network well.

Thesis Supervisor: Munther A. Dahleh
Title: Professor of Electrical Engineering

Thesis Supervisor: John N. Tsitsiklis
Title: Professor of Electrical Engineering and Computer Science

Acknowledgments

My deepest gratitude goes to my advisors, Munzer Dahleh and John Tsitsiklis. They have taught me how to learn and how to teach. They have taught me how to pick a relevant problem, set it up, think about it, solve it, and present it. The most important lesson I have learned during my Ph.D. is how to contrive and tell a relevant and rigorous story, and I owe this to John and Munzer. I greatly acknowledge their mentorship, guidance, and support, as well as the time and effort they have devoted. I am thankful for their insightful advice on research and academic plans, and their support on the academic job market. The researcher and teacher model that I look up to as I start my academic career is one that combines my advisors' rigor and brain power along with their humbleness, genuineness, and humor.

I would like to extend special thanks to my thesis readers, Marios Angeletos and Devavrat Shah, who have shared with me their insights and have contributed to this thesis through numerous discussions. I am grateful to Marios for his detailed feedback and suggestions on the part of this thesis on coordination with local information, as well as his support on the academic job market.

For the part of the thesis on coordination with local information, I want to thank my collaborator Alireza Tahbaz-Salehi for all of his work and guidance, as well as his help with the academic job market. I also thank Daron Acemoglu, George Mailath, Antonio Penta, and participants at the Workshop on the Economics of Networks, Systems, and Computation (NetEcon 2011, 2013), the Interdisciplinary Workshop on Information and Decision in Social Networks (WIDS 2011, 2012), the SENSEable City Lab at MIT, and the Cambridge Area Economics and Computation Day (CAEC 2013) for useful comments.

For the part of the thesis on inferring influence networks with temporal data, I want to thank Sridevi Sarma and her group for providing the EEG data on epileptic seizures, and hosting me for a visit at the Johns Hopkins Institute for Computational Medicine. I also want to thank Sandy Pentland and his group (Human Dynamics Laboratory at the MIT Media Lab) for sharing the mobile app installations data. I

thank Anima Anandkumar and participants at the following seminars/meetings for useful comments and discussions: the Marketing group at MIT Sloan, the INFORMS Annual Meeting (2013), the Neuromedical Control Systems Lab at Johns Hopkins, the Conference on Web and Internet Economics (WINE 2013), the Army Research Office Multidisciplinary University Research Initiative review, the New England Machine Learning Day (NEML 2014), the Engineering Systems and Design group at the Singapore University of Technology and Design, the Management Science and Operations group at London Business School, the Management Science and Innovation group at University College London, the Decision Sciences group at INSEAD, the Information, Operations, and Management Sciences group at New York University Stern, the Decision Sciences group at the IESE Business School, the Statistics, Econometrics and Quantitative Methods group at the Universitat Pompeu Fabra, the Networks, Economics and Algorithms group at Microsoft Research Cambridge, and the Data Analytics group at Mitsubishi Electric Research Laboratories.

The Laboratory for Information and Decision Systems has been a great home for my academic development. I would like to thank Dimitri Bertsekas, Vincent Blondel, Patrick Jaillet, Sanjoy Mitter, Asu Ozdaglar, Yury Polyanskiy, Moe Win, John's group (Kimon Drakopoulos, Hoda Eydgahi, Ali Kazerani, Jimmy Li, Mihalis Markakis, Alex Olshevsky, Theo Weber, Kuang Xu, Yunjian Xu, Yuan Zhong), Munzer's group (Elie Adam, Giancarlo Baldan, Saverio Bolognani, Ali Faghih, Rose Faghih, Diego Feijer, Qingqing Huang, Yola Katsargyri, Donatello Materassi, Noele Norris, Mardavij Roozbehani, Ketan Savla, Shreya Saxena), and Ammar Ammar, for giving useful comments on parts of this thesis and for stimulating discussions. I want to also thank the supportive and helpful LIDS or LIDS-related staff, and especially Albert Carter, Jennifer Donovan, Brian Jones, Lynne Dell, and Debbie Wright, and my officemates Diego Feijer, Yola Katsargyri, Mitra Osqui, and Rose Faghih.

I am grateful to Duncan Simester for his mentorship and his support and guidance on the academic job market. I also want to thank my other collaborators, Jimmy Li and Paat Rusmevichientong, from whom I have learnt a lot.

Throughout my years at MIT I have interacted with several academics who have

directly influenced my way of thinking. Apart from those already mentioned, I would like to thank Itai Ashlagi, Dimitri Bertsimas, Richard Dudley, Vivek Farias, Tom Leighton, Silvio Micali, Marvin Minsky, Alan Oppenheim, Michael Piore, Daniel Stroock, and John Wyatt.

Special thanks go to Eva for her support, patience, and companionship, as well as my closest friends Giorgos Angelis, Giorgos Faldamis, Anurupa Ganguly, Mihalis Markakis, Giorgos Papachristoudis, and Giorgos Samartzopoulos for all the great moments of the last few years in Cambridge.

My deepest appreciation goes to my parents, Vaso and Ilias, who have stood by me all throughout, and whose thirst for knowledge has been inspiring. This thesis is dedicated to them. I also thank my grandmother Angeliki and my sister Tzelika for their unconditional love and support.

My doctoral studies at MIT have been supported by the following grants, which I gratefully acknowledge: AFOSR FA9550-09-1-0420, AFOSR MURI/GaTech R6756-G2, ONR/GaTech R0897-G1.

Contents

1 Introduction **21**

1.1 Questions, Positioning, and Contributions 22

1.2 Overview 24

2 Coordination with Local Information: Introduction **29**

2.1 Related Literature 32

2.2 Model 34

2.2.1 Agents and Payoffs 34

2.2.2 Information and Signals 35

3 Coordination with Local Information: Results **39**

3.1 A Three-Agent Example 39

3.2 Local Information and Equilibrium Multiplicity 42

3.3 Local Information and the Extent of Multiplicity 48

4 Coordination with Exchange of Information **53**

4.1 Model 53

4.2 Characterization of Strategy Profiles that Survive Iterated Elimination of Strictly Dominated Strategies (IESDS) for Finite Unions of Cliques 54

4.2.1 Generic Characterization 54

4.2.2 Characterization in the Case of Cliques of Equal Size 57

4.3 The Case of Asymptotically Many Agents 58

4.3.1 Sufficient Conditions for Asymptotic Uniqueness 59

4.3.2 Interpretation 60

5 The Value of Temporal Data for Learning of Influence Networks: Introduction 63

5.1 Background and Related Literature 65

5.2 Overview 67

5.3 The Influence Model 69

6 Characterization of the Speed of Learning Using the Fisher Information and the Kullback-Leibler Divergence 71

6.1 Characterization of the Speed of Learning Using the Fisher Information 72

6.1.1 The Gap between Sets and Sequences, and a Network where It Is Zero 73

6.1.2 Comparing Learning with Sets and Learning with Sequences in an Example Network 76

6.2 Characterization of the Speed of Learning Using the Kullback-Leibler Divergence 78

6.2.1 Which of Two Peers Influences You Crucially? 79

6.2.2 Are You Influenced by Your Peer, or Do You Act Independently? 85

6.2.3 Does Your Peer Influence You a Lot or a Little? 90

6.2.4 Discussion 94

7 Theoretical Guarantees for Learning Influence with Zero/Infinity Edges 97

7.1 Conditions on Topology for Learnability 98

7.2 Learning Influence in the Star Network 102

7.2.1 The Bayesian Setting 104

7.2.2 The Worst-Case Setting 114

7.2.3 The Worst-Case Setting with Known Scaling of Agents with Influence Rate Infinity to $n+1$ 116

7.3 Learning Influence in the Star Network with Small Horizon 120

7.3.1 The Bayesian Setting 120
7.3.2 The Worst-Case Setting 122
7.3.3 The Worst-Case Setting with Known Scaling of Agents with Influence Rate Infinity to $n+1$ 123
7.4 Discussion 125

8 Theoretical Guarantees for General Networks 127
8.1 Learning Between the Complete Graph and the Complete Graph that Is Missing One Edge 127
8.1.1 An Algorithm for Learning 128
8.1.2 Learning with Sequences 129
8.1.3 Learning with Sets 132
8.2 Discussion 133

9 Learning Influence with Synthetic and Observational Data 135
9.1 Synthetic Data 135
9.2 Real Data: Mobile App Installations 139
9.2.1 The Dataset 139
9.2.2 Network Inference 139
9.3 Real Data: Epileptic Seizures 141
9.3.1 The Dataset 141
9.3.2 Network Inference 142
9.3.3 Discussion 146

10 Conclusion 147
10.1 Summary 147
10.2 Directions for Future Research 149

A Proofs for Coordination with Local Information 151

B Technical Appendix: Proof of Lemma 4 for Coordination with Local Information 165

List of Figures

4-1 A simple network. 56

4-2 A simple network with cliques of equal sizes. 57

6-1 A simple network of influence. No edge between agents 1 and 2 means that 1 does not influence 2, and vice versa. 76

6-2 Scaling of $\frac{\mathcal{J}^{\text{sequence}}_{13,13}}{\mathcal{J}^{\text{set}}_{13,13}}$ with κ_{13}, κ_{23}. 77

6-3 The hypothesis testing problem: which of agents 1, 2 crucially influences agent 3? 80

6-4 Which of two peers influences you? Plots of KL_{set} (circles), KL_{sequence} (crosses), KL_{time} (squares) against influence rate α for different horizon rates. 83

6-5 Which of two peers influences you? Plots of $KL_{\text{sequence}}/KL_{\text{set}}$ (crosses), $KL_{\text{time}}/KL_{\text{sequence}}$ (squares) against influence rate α for different horizon rates. 84

6-6 The hypothesis testing problem: is agent 2 influenced by agent 1, or does she have a high individual rate? 85

6-7 Are you influenced by your peer, or do you act independently? Plots of KL_{set} (circles), KL_{sequence} (crosses), KL_{time} (squares) against influence rate α for different horizon rates. 87

6-8 Are you influenced by your peer, or do you act independently? Plots of $KL_{\text{sequence}}/KL_{\text{set}}$ (crosses), $KL_{\text{time}}/KL_{\text{sequence}}$ (squares) against influence rate α for different horizon rates. 88

6-9 The hypothesis testing problem: is agent 2 influenced by agent 1 a lot or a little? 90

6-10 Does your peer influence you a lot or a little? Plots of KL_{set} (circles), KL_{sequence} (crosses), KL_{time} (squares) against influence rate α for different horizon rates. 91

6-11 Does your peer influence you a lot or a little? Plots of $KL_{\text{sequence}}/KL_{\text{set}}$ (crosses), $KL_{\text{time}}/KL_{\text{sequence}}$ (squares) against influence rate α for different horizon rates. 92

6-12 Does your peer influence you a lot or a little? $KL_{\text{sequence}}/KL_{\text{set}}$ in the limit of large influence rate α against horizon rate λ_{hor}. In the regime of high influence rate α, learning with sequences yields no significant gain over learning with sets, regardless of the horizon rate. 94

7-1 The hypothesis testing problem: what influence does each link carry to the star agent $(n+1)$: infinite or zero? 103

9-1 Learning a network of five agents with five high influence rates and all other influence rates zero (left) and four high influence rates and all other influence rates zero (right). We plot the ℓ_1 estimation error calculated on both the influence rates and the idiosyncratic rates. Learning is more accurate and faster with times than with sequences, and with sequences than with sets, with about an order of magnitude difference in the ℓ_1 error between sets and sequences, and between sequences and times. 136

9-2 Learning a network of 50 (left) and 100 (right) agents, generated from an Erdős-Rényi model with probability $p = 0.1$ that each directed edge is present. We plot the ℓ_1 estimation error calculated on both the influence rates and the idiosyncratic rates. Learning is more accurate and faster with times than with sequences, with a difference of one to two orders of magnitude in the ℓ_1 error between sequences and times. 137

9-3 Learning a network of ten agents, generated from an Erdős-Rényi model with probability $p = 0.1$ that each directed edge is present. We plot the ℓ_1 estimation error calculated on both the influence rates and the idiosyncratic rates, when the data is noisy. Learning is more accurate and faster with noisy times than with noisy sequences, with a difference of two orders of magnitude in the ℓ_1 error between sequences and times. 138

9-4 Concurrent visualization of the realized network of calls (the color of each edge denotes the number of calls between the users: closer to blue for lower number of calls, closer to red for higher number of calls) and the inferred network of influence using information on sequences of adoptions (the thickness of edge $i - j$ is proportional to the sum of the inferred influence rates $\lambda_{ij} + \lambda_{ji}$). We observe that edges with higher number of calls (red) are more likely to carry higher influence (thick). 141

List of Tables

4.1 Thresholds for strategies that survive IESDS for the network in Figure 4-1, for $h(x) = x$, and different values of σ. 56

5.1 Toy example with four records of adoption decisions by five agents, and how the same events are encoded in each of three different databases: a database that stores times, a database that stores sequences, and a database that stores sets. 64

7.1 Matching lower and upper bounds for the minimum number of i.i.d. products required to learn the influence model in terms of n, in the Bayesian setting when $p = 1/2$, for the two cases of learning the influence between one agent and the star agent and of learning the influence between all agents and the star agent, and for the two cases of learning based on sets of adoptions or sequences of adoptions. 104

7.2 Matching lower and upper bounds for the minimum number of i.i.d. products required to learn the influence model, in terms of n, in the Bayesian setting when $p = 1/n$, when learning the influence between all agents and the star agent, for the two cases of learning based on sets of adoptions or sequences of adoptions. 112

7.3 Matching lower and upper bounds for the minimum number of i.i.d. products required to learn the influence model in terms of n, in the worst-case setting, for the two cases of learning the influence between one agent and the star agent and of learning the influence between all agents and the star agent, and for the two cases of learning based on sets of adoptions or sequences of adoptions. 115

7.4 Matching lower and upper bounds for the minimum number of i.i.d. products required to learn the influence model in terms of n, in the worst-case setting when the scaling of agents ℓ with infinite influence rate to agent $n+1$ is known, for the two cases of learning based on sets of adoptions or sequences of adoptions. 117

7.5 Matching lower and upper bounds for the minimum number of i.i.d. products required to learn the influence model in terms of n, in the Bayesian setting when $p = 1/2$, for the two cases of learning the influence between one agent and the star agent and of learning the influence between all agents and the star agent, and for the two cases of learning based on sets of adoptions or sequences of adoptions. 121

7.6 Matching lower and upper bounds for the minimum number of i.i.d. products required to learn the influence model, in terms of n, in the Bayesian setting when $p = 1/n$, when learning the influence between all agents and the star agent, for the two cases of learning based on sets of adoptions or sequences of adoptions. 122

7.7 Matching lower and upper bounds for the minimum number of i.i.d. products required to learn the influence model, in terms of n, in the worst-case setting, for the two cases of learning the influence between one agent and the star agent and of learning the influence between all agents and the star agent, and for the two cases of learning based on sets of adoptions or sequences of adoptions. 123

7.8 Matching lower and upper bounds for the minimum number of i.i.d. products required to learn the influence model, in terms of n, in the worst-case setting when the scaling of agents ℓ with influence rate infinity to agent $n+1$ is known, for the two cases of learning based on sets of adoptions or sequences of adoptions. 124

9.1 Running time in seconds for learning based on sets, sequences, and times of adoptions, with and without the heuristic, for a network of five agents, with five influence rates equal to 10000 and all other equal to zero, idiosyncratic rates equal to 6, and horizon rate 30. The experiments were run with MATLAB on a 2.53 GHz processor with 4 GB RAM. 138

9.2 There is higher probability of friendship in the edges where we detect influence using sequences. A communication edge exists between two randomly selected nodes in the dataset with probability 0.3508. . . . 140

Chapter 1

Introduction

We study the interplay between networks, decisions, information, and inference, focusing on the intimate two-way connection between networks and outcomes of phenomena where networked entities interact. While the first part of the thesis[1] studies how networks shape equilibria, and therefore outcomes, of coordination games, the second part[2] studies the inference of influence networks from knowledge of outcomes. In particular, in the first part of this thesis, we characterize decision making of rational agents in coordination phenomena from a game-theoretic perspective; in the second part of this thesis, we instead posit an underlying mechanism for decision making, and develop theory for how to learn its parameters from data on agents' actions. The overarching theme is *locality* in the context of complex *networks*: we study local interactions between networked entities and how they shape aggregate outcomes. In the first part of this thesis, we study how the equilibria of coordination games depend on the patterns of *local* information sharing between the agents. In the second part of this thesis, we recover the network of *local* interactions or *local* influence between agents when we have access only to records of their behavior.

[1]Some of the material in the first part of the thesis appears in Dahleh, Tahbaz-Salehi, Tsitsiklis, and Zoumpoulis (2011, 2013a,b).

[2]Some of the material in the second part of the thesis appears in Dahleh, Tsitsiklis, and Zoumpoulis (2014).

1.1 Questions, Positioning, and Contributions

Our work on coordination with local information, presented in the first part of this thesis (Chapters 2, 3, 4), is part of the global games literature. Global games are "incomplete information games where the actual payoff structure is determined by a random draw from a given class of games and where each player makes a noisy observation of the selected game[3]" and have been used to study economic situations with an element of coordination and strategic complementarities, such as currency attacks, bank runs, debt crises, political protests and social upheavals, and partnership investments. While common knowledge of the fundamentals leads to the standard case of multiple equilibria due to the self-fulfilling nature of agents' beliefs, the global-games framework has been used extensively as a toolkit for arriving at a unique equilibrium selection. The endogeneity of the information structure can restore multiplicity; examples are the signaling role of the actions of a policy maker, the role of prices as an endogenous public signal, propaganda, and learning over time.

We study another natural setting in which correlation in the information between different agents is introduced, restoring multiplicity, while keeping the agents information solely exogenous: the (exogenous) information structure per se, namely, the details of who observes what noisy observation. The aim of our work is to understand the role of the exogenous information structure and, in particular, local information channels in the determination and characterization of equilibria in the coordination game. We answer the question of how the equilibria of the coordination game depend on the details of local information sharing. Our main contribution is to provide conditions for uniqueness and multiplicity that pertain solely to the details of information sharing.

Our key finding is that the number of equilibria is highly sensitive to the details of information locality. As a result, a new dimension of indeterminacy regarding the outcomes is being introduced: not only may the same fundamentals well lead to different outcomes in different societies, due to different realizations of the noisy observations;

[3] As defined in Carlsson and van Damme (1993).

the novel message of this work is that the same realization of the noisy observations is compatible with different equilibrium outcomes in societies with different structures of local information sharing.

In the second part of the thesis (Chapters 5, 6, 7, 8, 9), we assume a different stance to modeling decision making: no utilities are defined for agents; instead, we posit a parametric model for agents' actions that incorporates influence among agents and learn the model's parameters based on records of agents' actions.

Our work's overarching theme is to quantify the gain in the speed of learning of parametric models of influence, due to having access to richer temporal information. The question is highly relevant as many phenomena of interest in various contexts can be described with temporal processes governed by local interactions between networked entities that influence one another in their decisions or behaviors: examples abound in retailing (consumers sequentially purchasing a new product); social media (a topic trending on an online social network); public health (an epidemic spreading across a population); finance (systemic risk spreading across firms/sectors/countries); biology (a cellular process during which the expression of a gene affects the expression of other genes); computer networks (computer malware spreading across a network).

Due to the increasing capability of data acquisition technologies, rich data on the outcomes of such processes are oftentimes available (possibly with time stamps), yet the underlying network of local interactions is hidden. Untangling and quantifying such influences in a principled manner, based on observed behaviors, and thus identifying *causal* relations, can be a very challenging task. This is because there are many different confounding factors that may lead to seemingly similar phenomena. For example, distinguishing whether two individuals exhibit similar behaviors (i) because one influences another, or (ii) because they have similar tastes or live in the same environment, is not easy and may even be impossible. Our work proposes a parametric model of influence to infer who influences whom in a network of interacting entities based on data on their actions/decisions, and quantifies the gain in the speed of learning that is obtained when (i) the data provides the times when agents/entities took an action; versus when (ii) the data provides the (ordered) sequence of agents/entities

who took an action, but not the times; versus when (iii) the data provides merely the set of agents/entities who took an action.

Our main contributions are the quantification of the comparison in the rate of learning when having access to sets, sequences, and times of actions, both theoretically (for different measures of the speed of learning) and experimentally, and the identification of conditions under which the poor data mode of sets provides almost all the information needed for learning.

Our work is part of the literature on the problem of influence discovery using timed cascade data. Our novel perspective is the assessment of the value of having access to data of richer temporal detail for the speed of learning of the influence network. This is a relevant question with important practical implications: access to richer temporal data is at times expensive or even impossible, and richer temporal data is at times unreliable, while temporally poor data may suffice to learn key network relations. The decision maker wants to know what her learning potential is with the available data, or what the worth is of an investment in gaining access to richer data.

1.2 Overview

The first part of this thesis (Chapters 2, 3, 4) studies the role of local information channels in enabling coordination among strategic agents. Building on the standard finite-player global games framework, we show that the set of equilibria of a coordination game is highly sensitive to how information is locally shared among agents. More specifically, rather than restricting our attention to cases in which information is either purely public or private, we assume that there may be *local* signals that are only observable to (proper, but not necessarily singleton) subsets of the agents. The presence of such local sources of information guarantees that some (but not necessarily all) information is common knowledge among a group of agents, with important implications for the determinacy of equilibria. Our main contribution is to provide conditions for uniqueness and multiplicity of equilibria based solely on the pattern of information sharing among agents. Our findings thus provide a characterization of the

extent to which a coordination may arise as a function of what piece of information is available to each agent.

As the main result, we show that the coordination game has multiple equilibria if there exists a collection of agents such that (i) they do not share a common signal with any agent outside of that collection; and (ii) their observation sets form an increasing sequence of nested sets, referred to as a *filtration*. This result is a consequence of the fact that agents in such a collection face only limited strategic uncertainty regarding each others' actions. By the means of a few examples, we then show that the presence of common signals *per se* does not necessarily lead to equilibrium multiplicity. Rather, the condition that the observation sets of a collection of agents form a filtration, with no overlap of information within and without the filtration, plays a key role in introducing multiple equilibria.

We then focus on a special case in which each agent observes a single signal and provide an explicit characterization of the set of equilibria in terms of the commonality in agents' observation sets. We show that the size of the set of equilibrium strategies is increasing in the extent of variability in the size of the subsets of agents who observe the same signal. More specifically, we show that the distance between the two extremal equilibria in the coordination game is increasing in the standard deviation of the fractions of agents with access to the same signal.

Furthermore, we use our characterization to study the set of equilibria in large coordination games. We show that as the number of agents grows, the game exhibits multiple equilibria if and only if a non-trivial fraction of the agents have access to the same signal. Our result thus establishes that if the size of the subset of agents with common knowledge of a signal does not grow at the same rate as the number of agents, the information structure is asymptotically isomorphic to a setting in which all signals are private, which induces uniqueness.

Finally, we consider an application of our framework in which the noisy signals are interpreted to be the idiosyncratic signals of the agents, which are exchanged through a communication network.

The second part of this thesis (Chapters 5, 6, 7, 8, 9) quantifies the gain in speed

of learning of parametric models of influence, due to having access to richer temporal information. We seek to compare the speed of learning under three different cases of available data: (i) the data provides merely the set of agents/entities who took an action; (ii) the data provides the (ordered) sequence of agents/entities who took an action, but not the times; and (iii) the data provides the times of the actions. It is clear that learning is no slower with times than it is with sequences, and no slower with sequences than with sets; yet, what can we say about *how much* faster learning is with times than with sequences, and with sequences than with sets? This is, to the best of our knowledge, a comparison that has not been studied systematically before.

We propose a parametric model of influence which captures directed pairwise interactions. We first use the Fisher information and the Kullback-Leibler (KL) divergence as measures for the speed of learning.

Using the concept of Fisher information, we compute the gap between learning with sets and learning with sequences and provide an example of a network where the gap is zero, and an example of a network where the gain can be arbitrarily large. Using the concept of Kullback-Leibler divergence, we focus on learning the influence model in three particular instances, which we cast as respective binary hypothesis testing problems: Which of two agents influences a third agent? Is an agent influenced by another agent, or are her decisions self-induced? And is the influence between two agents large or small? We view these three questions as building blocks for understanding complex interactions in general networks. For each of the proposed hypothesis testing problems, we compare the Kullback-Leibler divergences in the cases of learning based on data of sets of decisions; learning based on data of sequences of decisions; and learning based on data of times of decisions.

We then provide theoretical guarantees on the sample complexity for correct learning with sets, sequences, and times. Our results characterize the sufficient and necessary scaling of i.i.d. samples required for correct learning. The asymptotic gain of having access to richer temporal data à propos of the speed of learning is thus quantified in terms of the gap between the derived asymptotic requirements under different data modes.

We also evaluate learning with sets, sequences, and times *with experiments*. Given data on outcomes, we learn the parametric influence model by maximum likelihood estimation. On both synthetic and real datasets, the value of learning with data of richer temporal detail is quantified, and our methodology is shown to recover the underlying network structure well. The real data come from (i) observations of mobile app installations of users, along with data on their communications and social relations; (ii) observations of levels of neuronal activity in different regions of the brain during epileptic seizure events, for different patients.

Chapter 2

Coordination with Local Information: Introduction

Coordination problems lie at the center of many economic or social phenomena, such as bank runs, currency attacks and social uprisings. The common feature of these phenomena (and many others) is that the decision of any given individual to take a specific action is highly sensitive to her expectations about whether others would take the same action. Such strong strategic complementarities in the agents' actions and the self-fulfilling nature of their expectations may lead to coordination failures: individuals may fail to take the action that is in their best collective interest.[1]

Since the seminal work of Carlsson and van Damme (1993), the machinery of global games has emerged as the workhorse model for understanding and analyzing various applications that exhibit an element of coordination. Carlsson and van Damme (and many authors since) showed that even though the complete information version of coordination games may have multiple equilibria, the same is not true if agents face some strategic uncertainty regarding each other's actions. In particular, they established that introducing minimal uncertainty about some payoff-relevant parameter of the game may lead to the selection of a unique equilibrium.

The global games literature, for the most part, has focused on the role of pub-

[1]For example, in the context of bank runs, depositors may decide to withdraw their deposits from a bank simply as a result of the fear that other depositors may do the same, leading to a self-fulfilling run on the bank.

lic and private information in coordination games, while ignoring the effects of local information channels in facilitating coordination. This is despite the fact that, in many real world scenarios, local information channels may play a key role as a means of enabling agents to coordinate on different actions. For instance, it is by now conventional wisdom that the protesters in many recent anti-government uprisings throughout the Middle East used decentralized modes of communication (such as word-of-mouth communications and Internet-based social media platforms) to coordinate on the time and location of street protests (Ali, 2011).

Motivated by these observations, this work studies the role of local information channels in enabling coordination among strategic agents. Building on the standard finite-player global games framework, we show that the set of equilibria of a coordination game is highly sensitive to how information is locally shared among agents. More specifically, rather than restricting our attention to cases in which information is either purely public or private, we assume that there may be *local* signals that are only observable to (proper, but not necessarily singleton) subsets of the agents. The presence of such local sources of information guarantees that some (but not necessarily all) information is common knowledge among a group of agents, with important implications for the determinacy of equilibria. Our main contribution is to provide conditions for uniqueness and multiplicity of equilibria based solely on the pattern of information sharing among agents. Our findings thus provide a characterization of the extent to which a coordination may arise as a function of what piece of information is available to each agent.

Strategic uncertainty, that is, uncertainty about others' actions at a given equilibrium, means inability to forecast one another's actions. We explain what patterns of local information sharing heighten strategic uncertainty, i.e., make it more difficult to predict the actions of others, thus making coordination harder and shrinking the set of equilibria (possibly to the extent of uniqueness); and what patterns of local information sharing reduce strategic uncertainty, i.e., permit better forecasting of one another's actions, thus allowing agents to coordinate their actions along any one of a larger set of equilibria.

As the main result of the work, we show that the coordination game has multiple equilibria if there exists a collection of agents such that (i) they do not share a common signal with any agent outside of that collection; and (ii) their observation sets form an increasing sequence of nested sets, referred to as a *filtration*. This result is a consequence of the fact that agents in such a collection face only limited strategic uncertainty regarding each others' actions. In fact, not only do agents with the larger observation sets face no uncertainty in predicting the actions of agents with smaller observation sets, but the latter group can also forecast the equilibrium actions of the former. The underlying reason for the limited strategic uncertainty among the agents belonging to a filtration is that the signals that are only observed by agents with larger observation sets cannot be used as a basis for coordination. Therefore, when these signals are close to the signals commonly observed by both agents with smaller and larger observation sets, their marginal effect on the updating of the mean of the posterior beliefs is only small; as a result, agents with smaller observation sets face only limited strategic uncertainty about the actions of agents with larger observation sets, leading to multiplicity. By the means of a few examples, we then show that the presence of common signals *per se* does not necessarily lead to equilibrium multiplicity. Rather, the condition that the observation sets of a collection of agents form a filtration, with no overlap of information within and without the filtration, plays a key role in introducing multiple equilibria.

We then focus on a special case in which each agent observes a single signal and provide an explicit characterization of the set of equilibria in terms of the commonality in agents' observation sets. We show that the size of the set of equilibrium strategies is increasing in the extent of variability in the size of the subsets of agents who observe the same signal. More specifically, we show that the distance between the two extremal equilibria in the coordination game is increasing in the standard deviation of the fractions of agents with access to the same signal.

Furthermore, we use our characterization to study the set of equilibria in large coordination games. We show that as the number of agents grows, the game exhibits multiple equilibria if and only if a non-trivial fraction of the agents have access to

the same signal. Our result thus establishes that if the size of the subset of agents with common knowledge of a signal does not grow at the same rate as the number of agents, the information structure is asymptotically isomorphic to a setting in which all signals are private. Under such conditions, any agent faces some uncertainty regarding the behavior of almost every other agent, even as all the fundamental uncertainty is removed. The presence of such strategic uncertainty implies that the set of rationalizable strategies of each agent collapses to a single strategy, as the number of agents grows.

Finally, we consider an application of our framework in which the noisy signals are interpreted to be the idiosyncratic signals of the agents, which are exchanged through a communication network.

2.1 Related Literature

Our work is part of the by now large literature on global games. The global games literature, which was initiated by the seminal work of Carlsson and van Damme (1993) and later expanded by Frankel, Morris, and Pauzner (2003) and Morris and Shin (2003), mainly focuses on how the presence of strategic uncertainty may lead to the selection of a unique equilibrium in coordination games. This machinery has since been used extensively to analyze various applications that exhibit an element of coordination. Examples include, currency attacks (Morris and Shin, 1998), bank runs (Goldstein and Pauzner, 2005), debt crises (Morris and Shin, 2004), political protests (Edmond, 2013), and partnership investments (Dasgupta, 2007).

Within the global games literature, Hellwig (2002) studies the set of equilibria of coordination games in the presence of both public and private information. He shows that the game exhibits an asymptotically unique equilibrium if the variance of the public signal converges to zero at a rate faster than one half the rate of convergence of the variance of private signals. Our work, on the other hand, focuses on more general information structures by allowing for signals that are neither completely public nor private, and provides conditions under which such signals may lead to equilibrium

multiplicity.

Our work is also related to the works of Morris and Shin (2007) and Mathevet (2012) who characterize the set of equilibria of coordination games in terms of the agents' types, while abstracting away from the information structure of the game. Our work, on the other hand, provides a characterization of the set of equilibrium strategies in terms of the details of the game's signal structure. Despite being more restrictive in scope, our results shed light on the role of local information in enabling coordination.

A different set of papers study how the endogeneity of agents' information structure in coordination games may lead to equilibrium multiplicity. For example, Angeletos and Werning (2006), and Angeletos, Hellwig, and Pavan (2006, 2007) show how prices, the action of a policy-maker, or past outcomes can function as endogenous public signals that may restore equilibrium multiplicity in settings that would have otherwise exhibited a unique equilibrium.[2] We, in contrast, study another natural setting in which correlation in the information between different agents is introduced, resulting in multiplicity, while keeping the agents' information exogenous.

Finally, our work is also related to the strand of literature that focuses on the role of local information channels and social networks in shaping economic outcomes. Examples include Acemoglu, Dahleh, Lobel, and Ozdaglar (2011), Dessein, Galeotti, and Santos (2013) and Galeotti, Ghiglino, and Squintani (forthcoming). Within this literature, our work is closest to Chwe (2000), who studies the role of communication networks in coordination games. His main focus is in characterizing the set of networks for which, regardless of the agents' prior beliefs, there exists an equilibrium in which agents can coordinate on a specific action. Our work, in contrast, characterizes the set of all equilibria of the coordination game as a function of the agents' observation sets.

[2] Yildiz and Weinstein (2007) and Penta (2012), on the other hand, show that, for sufficiently rich belief structures, multiplicity is degenerate and fragile: by introducing sufficiently rich perturbations to the information structure, we can always relax the implicit assumptions of the model and obtain an open set of situations in which there is a unique rationalizable outcome.

2.2 Model

Our model is a finite-agent variant of the canonical model of global games studied by Morris and Shin (2003).

2.2.1 Agents and Payoffs

Consider a *coordination game* played by n agents whose set we denote by $N = \{1, 2, \ldots, n\}$. Each agent can take one of two possible actions, $a_i \in \{0, 1\}$, which we refer to as the *safe* and *risky* actions, respectively. The payoff of taking the safe action is normalized to zero, regardless of the actions of other agents. The payoff of taking the risky action, on the other hand, depends on (i) the number of other agents who take the risky action, and (ii) an underlying *state* of the world $\theta \in \mathbb{R}$, which we refer to as *the fundamental*. In particular, the payoff function of agent i is

$$u_i(a_i, a_{-i}, \theta) = \begin{cases} \pi(k, \theta) & \text{if } \ a_i = 1 \\ \\ 0 & \text{if } \ a_i = 0, \end{cases}$$

where $k = \sum_{j=1}^{n} a_j$ is the number of agents who take the risky action and $\pi : N \times \mathbb{R} \to \mathbb{R}$ is a function that is continuous in its second argument. We impose the following assumptions on the agents' payoff function.

Assumption 1. The payoff function $\pi(k, \theta)$ is strictly increasing in k for all θ. Furthermore, there exists a constant $\rho > 0$ such that $\pi(k, \theta) - \pi(k - 1, \theta) > \rho$ for all θ and all k.

The above assumption captures the presence of an element of coordination between agents: taking either action becomes more attractive the more other agents take that action. For example, in the context of a bank run, each depositor has more incentives not to withdraw her deposits from a bank the higher the number of other depositors who also do not withdraw is. Hence, the game exhibits strategic complementarities. The second part of Assumption 1 is made for technical reasons and states that the

payoff of switching to the risky action when one more agent takes action 1 is uniformly bounded from below, regardless of the value of θ.

Assumption 2. $\pi(k, \theta)$ is strictly decreasing in θ for all k.

That is, any given individual has less incentive to take the risky action if the fundamental takes a higher value. Thus, taking the other agents' actions as given, each agent's optimal action is decreasing in the state. Once again, in the bank run context, the fundamental value θ may capture the extent of insolvency of the bank which determines the incentives of the depositors in not withdrawing their funds. Finally, we impose the following assumption:

Assumption 3. There exist constants $\underline{\theta}, \bar{\theta} \in \mathbb{R}$ satisfying $\underline{\theta} < \bar{\theta}$ such that,

(i) $\pi(k, \theta) > 0$ for all k and all $\theta < \underline{\theta}$.

(ii) $\pi(k, \theta) < 0$ for all k and all $\theta > \bar{\theta}$.

Thus, each agent strictly prefers to take the safe (risky) action for sufficiently high (low) states of the world, irrespective of the actions of other agents. If, on the other hand, the underlying state belongs to the so-called *critical region* $[\underline{\theta}, \bar{\theta}]$, then the optimal behavior of each agent depends on her beliefs about the actions of other agents.

In summary, the agents face a coordination game with strong strategic complementarities in which the value of coordinating on the risky action depends on the underlying state of the world. Furthermore, particular values of the state make either action strictly dominant for all agents. As already mentioned, classical examples of such games studied in the literature include currency attacks, bank runs, debt crises, social upheavals, and partnership investments.

2.2.2 Information and Signals

Similar to the canonical global games model, information about the fundamental is assumed to be incomplete and asymmetric. Agents hold a common prior belief on $\theta \in \mathbb{R}$, which for simplicity we assume to be the (improper) uniform distribution over

the real line. Conditional on the state θ, a collection $(x_1, \ldots, x_m) \in \mathbb{R}^m$ of noisy signals is generated, where $x_r = \theta + \xi_r$. We assume that the noise terms $(\xi_1, \ldots, \xi_m)$ are independent from θ and are drawn from a continuous joint probability density function with full support over $\mathbb{R}^m$ (e.g., a non-degenerate multivariate normal).

Not all agents can observe all realized signals. Rather, agent i has access to a non-empty subset of the realized signals, $I_i \subseteq \{x_1, \ldots, x_m\}$, which we refer to as the *observation set* of agent i.[3][4] The collection of observation sets $\{I_i\}_{i \in N}$, which we refer to as the *information structure* of the game, is common knowledge among all agents. Thus, a pure strategy of agent i is a mapping $s_i : \mathbb{R}^{|I_i|} \to \{0, 1\}$, where $|I_i|$ denotes the cardinality of agent i's observation set.

The possibility that the agents' information sets are distinct implies that, in general, agents may hold asymmetric information about the underlying state. Furthermore, the fact that agent i's observation set can be any subset of $\{x_1, \ldots, x_m\}$ means that the extent to which any given signal x_r is observed may vary across signals. For example, if $x_r \in I_i$ for all i, then x_r is in essence a *public signal* observed by all agents. On the other hand, if $x_r \in I_i$ for some i but $x_r \notin I_j$ for all $j \neq i$, then x_r is a *private signal* of agent i. Any signal which is neither private nor public can be interpreted as a *local* source of information observed only by a proper subset of the agents. The potential presence of such local signals is our point of departure from the canonical global games literature, which for the main part focuses on games with only private and public signals.

Finally, we impose the following mild technical assumption on the payoffs and on the distribution of the signals.

Assumption 4. $\pi(k, \theta)$ has bounded derivative with respect to θ, and for any col-

[3]With some abuse of notation, we use I_i to denote both the set of actual signals observed by agent i, as well as the set of indices of the signals observed by her.

[4]In an important special case, the agents are organized in a network. Each agent i observes a signal x_i and in addition observes the signals x_j of all its neighbors. Clearly, such a network structure induces an information structure of the type we are considering in this work. We discuss this special case in Chapter 4.

lection of signals $\{x_r\}_{r\in H}$, $H \subseteq \{1,\ldots,m\}$,

$$\mathbb{E}\big[|\theta|\big|\{x_r\}_{r\in H}\big] < \infty.$$

This assumption guarantees in particular that the agents' payoff function is integrable.

Chapter 3

Coordination with Local Information: Results

3.1 A Three-Agent Example

Before proceeding to our general results, we present a simple example and show how the presence of local signals determines the set of equilibria. Consider a game consisting of three agents $N = \{1, 2, 3\}$ and suppose that $\pi(k, \theta) = (k-1)/2 - \theta$. This payoff function, which is similar to the one in the canonical finite-player global game model studied in the literature, satisfies Assumptions 1–3 with $\underline{\theta} = 0$ and $\bar{\theta} = 1$. We consider three different information structures, contrasting the cases in which agents have access to only public or private information to a case with a local signal. Throughout this section, we assume that the noise terms ξ_r are mutually independent and are normally distributed with mean zero and variance $\sigma^2 > 0$.

Public Information First, consider the case where all agents observe the same public signal x, that is, $I_i = \{x\}$ for all i. Thus, no agent has any private information about the state. It is easy to verify that under such an information structure, the coordination game has multiple Bayesian Nash equilibria. In particular, for any $\tau \in [0, 1]$, the strategy profile in which all agents take the risky action if and only if $x < \tau$ is an equilibrium, regardless of the value of σ. Consequently, as the public

signal becomes infinitely accurate (i.e., as $\sigma \to 0$), the underlying game has multiple equilibria as long as the underlying state θ belongs to the critical region $[0, 1]$.

Private Information Next, consider the case where all agents have access to a different private signal. In particular, suppose that three signals $(x_1, x_2, x_3) \in \mathbb{R}^3$ are realized and that x_i is privately observed by agent i; that is, $I_i = \{x_i\}$ for all i. As is well-known from the global games literature, the coordination game with privately observed signals has an essentially unique Bayesian Nash equilibrium. To verify that the equilibrium of the game is indeed unique, it is sufficient to focus on the set of equilibria in threshold strategies, according to which each agent takes the risky action if and only if her private signal is smaller than a given threshold.[1] In particular, let τ_i denote the threshold corresponding to the strategy of agent i. Taking the strategies of agents j and k as given, agent i's expected payoff of taking the risky action is equal to $\mathbb{E}[\pi(k, \theta)|x_i] = \frac{1}{2}[\mathbb{P}(x_j < \tau_j|x_i) + \mathbb{P}(x_k < \tau_k|x_i)] - x_i$. For τ_i to correspond to an equilibrium strategy of agent i, she has to be indifferent between taking the safe and the risky actions when $x_i = \tau_i$. Hence, the collection of thresholds (τ_1, τ_2, τ_3) corresponds to a Bayesian Nash equilibrium of the game if and only if for all permutations of i, j and k, we have

$$\tau_i = \frac{1}{2}\Phi\left(\frac{\tau_j - \tau_i}{\sigma\sqrt{2}}\right) + \frac{1}{2}\Phi\left(\frac{\tau_k - \tau_i}{\sigma\sqrt{2}}\right),$$

where $\Phi(\cdot)$ denotes the cumulative distribution function of the standard normal.[2] It is then immediate to verify that $\tau_1 = \tau_2 = \tau_3 = 1/2$ is the unique solution of the above system of equations. Thus, in the (essentially) unique equilibrium of the game, agent i takes the risky action if she observes $x_i < 1/2$, whereas she takes the safe action if $x_i > 1/2$.[3] Consequently, in contrast to the game with public information, as $\sigma \to 0$,

[1] This is a consequence of the fact that the underlying game is a Bayesian game with strategic complementarities, and hence, the extremal equilibria are monotone in types. For a detailed study of Bayesian games with strategic complementarities, see Van Zandt and Vives (2007).

[2] Note that $\theta|x_i \sim \mathcal{N}(x_i, \sigma^2)$, and as a result, $x_j|x_i \sim \mathcal{N}(x_i, 2\sigma^2)$.

[3] Following standard arguments from Carlsson and van Damme (1993) or Morris and Shin (2003), one can also show that this strategy profile is also the (essentially) unique strategy profile that survives the iterated elimination of strictly dominated strategies.

all agents choose the risky action if and only if $\theta < 1/2$. This observation shows that, even in the limit as signals become arbitrarily precise and all the fundamental uncertainty is removed, the presence of strategic uncertainty among the agents still leads to the selection of a unique equilibrium. We remark that the equilibrium would remain unique even if the agents' private observations were (imperfectly) correlated.

Local Information Finally, consider the case where only two signals (x_1, x_2) are realized and the agents' observation sets are $I_1 = \{x_1\}$ and $I_2 = I_3 = \{x_2\}$; that is, agent 1 observes a private signal whereas agents 2 and 3 have access to the same local source of information. The fact that all the information available to agents 2 and 3 is common knowledge between them distinguishes this case from the canonical global game model with private signals.

To determine the extent of equilibrium multiplicity, once again it is sufficient to focus on the set of equilibria in threshold strategies. Let τ_1 and $\tau_2 = \tau_3$ denote the thresholds corresponding to the strategies of agents 1, 2 and 3, respectively.[4] If agent 1 takes the risky action, she obtains an expected payoff of $\mathbb{E}[\pi(k,\theta)|x_1] = \mathbb{P}(x_2 < \tau_2|x_1) - x_1$. On the other hand, the expected payoff of taking the risky action to agent 2 (and by symmetry, agent 3) is given by

$$\mathbb{E}\left[\pi(k,\theta)|x_2\right] = \begin{cases} \frac{1}{2}\mathbb{P}(x_1 < \tau_1|x_2) - x_2 + \frac{1}{2} & \text{if } \ x_2 < \tau_2 \\ \\ \frac{1}{2}\mathbb{P}(x_1 < \tau_1|x_2) - x_2 & \text{if } \ x_2 > \tau_2. \end{cases}$$

For thresholds τ_1 and τ_2 to correspond to a Bayesian Nash equilibrium, agent 1 has to be indifferent between taking the risky action and taking the safe action when $x_1 = \tau_1$, therefore it must be the case that

$$\tau_1 = \Phi\left(\frac{\tau_2 - \tau_1}{\sigma\sqrt{2}}\right). \tag{3.1}$$

Furthermore, agents 2 and 3 should not have an incentive to deviate, which requires

[4] Given that the set of equilibria constitute a lattice, agents 2 and 3 play the same strategy in both extremal equilibria.

that their expected payoffs have to be nonnegative for any $x_2 < \tau_2$, and nonpositive for any $x_2 > \tau_2$. It can be shown that this leads to the condition

$$2\tau_2 - 1 \leq \Phi\left(\frac{\tau_1 - \tau_2}{\sigma\sqrt{2}}\right) \leq 2\tau_2. \tag{3.2}$$

It is easy to verify that the pair of thresholds (τ_1, τ_2) simultaneously satisfying (3.1) and (3.2) is not unique. In particular, as $\sigma \to 0$, for every $\tau_1 \in [1/3, 2/3]$, there exists some τ_2 with $\tau_2 \approx \tau_1$, such that the profile of threshold strategies (τ_1, τ_2, τ_2) is a Bayesian Nash equilibrium. Consequently, as the signals become very precise, the underlying game has multiple equilibria as long as $\theta \in [\frac{1}{3}, \frac{2}{3}]$.

Thus, even though there are no public signals, the presence of common knowledge between a proper subset of the agents restores the multiplicity of equilibria. In other words, the local information available to agents 2 and 3 serves as a coordination device, enabling them to predict one another's actions. The presence of strong strategic complementarities in turn implies that agent 1 will use his private signal as a predictor of how agents 2 and 3 coordinate their actions. Nevertheless, due to the presence of some strategic uncertainty between agents 2 and 3 on the one hand, and agent 1 on the other, the set of rationalizable strategies is strictly smaller relative to the case where all three agents observe a public signal. Therefore, the local signal makes smaller the set of equilibria of the coordination game, yet not to the extent of uniqueness.

3.2 Local Information and Equilibrium Multiplicity

The example in the previous section shows that the set of equilibria in the presence of local signals may not coincide with the set of equilibria under purely private or public signals. In this section, we provide a characterization of the role of local information channels in determining the equilibria of the coordination game presented in Section 2.2.

As our first result, we show that the presence of locally common, but not neces-

sarily public, signals leads to equilibrium multiplicity.

Theorem 1. *Suppose that there exists a non-singleton subset of agents $C \subseteq N$ such that*

(a) $I_i = I_\ell = \tilde{I}$ *for all* $i, \ell \in C$.

(b) $I_i \cap I_j = \varnothing$ *for all* $i \in C$ *and* $j \notin C$.

Then, the coordination game has multiple Bayesian Nash equilibria.

The above result shows that common knowledge among a subset of agents is sufficient for equilibrium multiplicity. The intuition behind Theorem 1 is analogous to the simple example studied in Section 3.1 and is best understood under the limiting case in which all signals are arbitrarily precise. The absence of any informational asymmetry between agents in C guarantees that they do not face any strategic uncertainty regarding each other's behavior. Furthermore, taking the strategies of the agents outside of C as given, there is a subset of the critical region $\Theta \subset [\underline{\theta}, \bar{\theta}]$ over which the optimal action of each agent $i \in C$ depends on the actions of other agents in C. In other words, if $\theta \in \Theta$, agents in C effectively face an induced coordination game among themselves, with the collection of signals in $\tilde{I}$ serving as public signals on which they can coordinate. Any equilibrium of this induced, smaller coordination game played by agents in C then translates to a Bayesian Nash equilibrium of the original game.

We remark that even though the presence of local signals may lead to equilibrium multiplicity, the set of Bayesian Nash equilibria of the game does not necessarily coincide with that for the case in which all signals are public. Rather, as we show in the next section, the set of equilibria crucially depends on the number of agents in C, as well as the detailed information structure of other agents.

As already explained, Theorem 1 highlights that common knowledge among a subset of agents is sufficient for equilibrium multiplicity. Our next result shows that the game may exhibit multiple equilibria even under less stringent common knowledge assumptions. To simplify exposition, we restrict our attention to the case in which

each agent's payoff for taking the risky action is linear in the state as well as the aggregate action; that is,

$$\pi(k, \theta) = \frac{k-1}{n-1} - \theta, \tag{3.3}$$

where n is the number of agents and k is the number of agents who take the risky action. We also assume that the error terms in the signals ξ_r are independent and normally distributed. Before presenting our next result, we define the following concept:

Definition 1. The observation sets of agents $i_1, \ldots, i_c \in N$ form a *filtration* if $I_{i_1} \subseteq I_{i_2} \subseteq \cdots \subseteq I_{i_c}$.

Thus, the observation sets of agents in C constitute a filtration if they form a nested sequence of increasing sets. This immediately implies that the signals of the agent with the smallest observation set is common knowledge among all agents in C. We have the following result:

Theorem 2. *Suppose that there exists a subset of agents $C \subseteq N$ such that*

(a) The observation sets of agents in C form a filtration.

(b) $I_i \cap I_j = \varnothing$ for any $i \in C$ and $j \notin C$.

Then, the coordination game has multiple Bayesian Nash equilibria.

Thus, the presence of a cascade of increasingly rich observations is sufficient for guaranteeing equilibrium multiplicity. The above result clearly reduces to Theorem 1 when all observation sets in the filtration are identical.

Under an information structure that exhibits a filtration, it is immediate that, for any given equilibrium, agents in C with larger observation sets do not face any uncertainty (strategic or otherwise) in predicting the equilibrium actions of agents with smaller observation sets. After all, all signals available to the latter are also observed by the former. The above result suggests that, for certain realizations of the signals, agents with smaller observation sets can also forecast the equilibrium actions of agents with the larger observation sets, leading to the possibility of coordination

and equilibrium multiplicity. This is more clearly understood by considering a collection C with two agents, 1 and 2, with observation sets $I_1 \subset I_2$. The signals observed by agent 2 which are not observable by agent 1, cannot be used as a basis for coordination between the two. Hence, when these signals are close to the signals commonly observed by both agents, their marginal effect on the updating of the mean of agent 2's posterior beliefs is only small, and as a result, agent 1 faces no strategic uncertainty about the actions of agent 2. We remark that, assumption (b) of Theorem 2 plays a crucial role for the above argument to work: it is what causes agents in C to effectively face an induced coordination game among themselves, in which they can forecast each other's actions, as explained above.

The constructive proof of Theorem 2 yields insight into the nature of the multiplicity that emerges due to the presence of a filtration. Focusing on threshold strategies, each agent decides how to play by comparing the sum of his signals[5] to a threshold, using different thresholds for different cases for the observations of agents whose signals are contained in his observation set: if an agent observes all the information that another agent observes, and the latter agent switches from playing the risky action to playing the safe action, then the former agent's threshold below which he wants to play the risky action becomes lower (i.e., his strategy becomes more conservative). Starting from the agent with the largest observation set of the filtration, we iteratively construct allowed ranges for the thresholds of all agents in the collection so that every agent best-responds. Every selection of thresholds within the (non-trivial) constructed ranges is an equilibrium; we thus construct all possible equilibria in threshold strategies.

We remark that the mere presence of a collection of signals that are common knowledge between a subset of agents is not sufficient for equilibrium multiplicity. Rather, the occurrence of a filtration plays a key role in generating multiple equilibria. In particular, in the absence of a filtration, or if the condition of no overlap of information within and without the filtration is violated, then, despite the presence

[5]Due to our assumption of normality, the sum of the signals is a sufficient statistic with respect to θ.

of some signals that are common knowledge between some agents, we may obtain a unique equilibrium.

To illustrate this point, recall the three-agent coordination game presented in Section 3.1, in which $\pi(k, \theta) = (k-1)/2 - \theta$. Also, suppose that three signals $(x_1, x_2, x_3) \in \mathbb{R}^3$ are realized in which the noise terms ξ_r are mutually independent and normally distributed with mean zero and variance $\sigma^2 > 0$. The agents' observation sets are given by

$$I_1 = \{x_2, x_3\}, \quad I_2 = \{x_3, x_1\}, \quad I_3 = \{x_1, x_2\}. \tag{3.4}$$

Thus, even though one signal is common knowledge between any pair of agents, the observation sets of no subset of agents form a filtration. We have the following result:

Proposition 1. *Suppose that agents' observation sets are given by* (3.4)*. There exists* $\bar{\sigma}$ *such that if* $\sigma > \bar{\sigma}$*, then the game has an essentially unique*[6] *equilibrium.*

Hence, the coordination game has a unique equilibrium despite the fact that for any pair of agents, the two agents share a common signal. Theorem 2 shows that, a filtration, along with the condition of no overlap of information within and without the filtration, provides sufficient grounds for coordination between agents, leading to multiplicity. Proposition 1 highlights that the mere presence of some signals that are common knowledge between agents in some subset does not necessarily lead to multiple equilibria, as such signals may not provide enough grounds for coordination.

In addition to the presence of a subset of agents C whose observation sets form a filtration, another key assumption for Theorems 1 and 2 to hold is that the observation set of no agent outside C contains any of the signals observable to agents in C. To clarify the role of this assumption in generating multiple equilibria, once again consider the three-agent example of the previous section, and assume that the agents'

[6]The equilibrium is *essentially* unique in the sense that any two strategy profiles that are equilibria differ on a set in the signal space of Lebesgue measure zero.

observation sets are given by

$$I_1 = \{x_1\}, \quad I_2 = \{x_2\}, \quad I_3 = \{x_1, x_2\}. \tag{3.5}$$

It is immediate that assumption (a) of Theorem 2 is satisfied, whereas assumption (b) is violated. In particular, even though the observation sets of agents in $C = \{2, 3\}$ form a filtration, $I_1 \cap I_3 \neq \varnothing$.

Proposition 2. *Suppose that agents' observation sets are given by* (3.5)*. Then, the game has an essentially unique equilibrium. Furthermore, as $\sigma \to 0$, all agents choose the risky action if and only if $\theta < 1/2$.*

Thus, as fundamental uncertainty is removed, information structure (3.5) induces the same (essentially) unique equilibrium as in the case where all signals are private. This is despite the fact that the observation sets of agents in $C = \{2, 3\}$ form a filtration. To understand the intuition behind the uniqueness of equilibrium in the above game, we contrast it with a simple game consisting of two agents whose observation sets are given by

$$I_1 = \{x_1\}, \quad I_2 = \{x_1, x_2\}. \tag{3.6}$$

By Theorem 2, and in contrast with the game described by (3.5), this two-agent game has multiple equilibria. Notice that in either game, the agents with the larger observation sets do not face any uncertainty (strategic or otherwise) in predicting the actions of agents with smaller observation sets. In particular, in the game with information structure (3.5), agent 3 faces no strategic uncertainty regarding the equilibrium actions of agents 1 and 2. Similarly, with information structure (3.6), agent 2 can perfectly predict the equilibrium action of agent 1. Thus, the divergence in the determinacy of equilibria in the two examples is solely due to the form of uncertainty that agents with smaller observation sets face regarding the behavior of agents who observe more signals.

More specifically, under information structure (3.6), agent 2 uses signal x_1 to predict the equilibrium action of agent 1, whereas she uses x_2 only as a means for

obtaining a better estimate of the underlying state θ. In other words, with information structure (3.6), x_2 is not used as a basis for coordination with agent 1. Thus, whenever observing the extra signal x_2 does not change the mean of agent 2's posterior belief about θ to the extent that taking either action becomes dominant, agent 2 can base her action on her expectation of agent 1's equilibrium action. This in turn means that, for certain realizations of the signals, agent 1 does not face any strategic uncertainty about agent 2's action either, leading to equilibrium multiplicity.

This, however, is in sharp contrast with what happens under information structure (3.5). In that case, agent 3 uses signal x_2 not only to estimate θ, but also as a perfect predictor of agent 2's equilibrium action. Thus, even if observing x_2 does not change the posterior mean of agent 3's belief about θ, it provides her with information about the action that agent 2 is expected to take in equilibrium. This extra piece of information, however, is not available to agent 1. Thus, even as $\sigma \to 0$, agent 1 faces strategic uncertainty regarding the equilibrium action of agent 3. The presence of such strategic uncertainty implies that the game with information structure (3.5) has a unique equilibrium.

3.3 Local Information and the Extent of Multiplicity

Our analysis so far was focused on the dichotomy between multiplicity and uniqueness of equilibria. However, as the example in Section 3.1 shows, even when the game exhibits multiple equilibria, the set of equilibria depends on how information is locally shared between different agents. In this section, we provide a characterization of the set of all Bayesian Nash equilibria as a function of the observation sets of different agents. Our characterization quantifies the dependence of the set of rationalizable strategies on the extent to which agents observe common signals.

In order to explicitly characterize the set of equilibria, we restrict our attention to a game with linear payoff functions given by (3.3). We also assume that $m \leq n$

signals, denoted by $(x_1, \ldots, x_m) \in \mathbb{R}^m$, are realized where the noise terms $(\xi_1, \ldots, \xi_m)$ are mutually independent and normally distributed with mean zero and variance $\sigma^2 > 0$. Each agent observes only one of the realized signals; that is, given each agent $i \in N$, her observation set is $I_i = \{x_r\}$ for some $1 \leq r \leq m$. Finally, we denote the fraction of agents that observe signal x_r by c_r, and let $c = [c_1, \ldots, c_m]$.[7] The set of rationalizable strategies are the strategies that play the risky action if the observed signal is below a lower threshold, and that play the safe action if the observed signal is above an upper threshold. Our result characterizes the thresholds in the limit as $\sigma \to 0$.

Theorem 3. *As $\sigma \to 0$, the strategy s_i of agent i is rationalizable if and only if*

$$s_i(x) = \begin{cases} 1 & \text{if } x < \underline{\tau} \\ 0 & \text{if } x > \bar{\tau}, \end{cases}$$

where

$$\underline{\tau} = 1 - \bar{\tau} = \frac{n}{2(n-1)}\Big(1 - \|c\|_2^2\Big) \leq \frac{1}{2}.$$

The above theorem shows that the distance between the thresholds of the "largest" and "smallest" rationalizable strategies depends on how information is locally shared between different agents. More specifically, a smaller $\|c\|_2$ implies that the set of rationalizable strategies would shrink. Note that $\|c\|_2$ is essentially a proxy for the extent to which agents observe common signals: it takes a smaller value whenever any given signal is observed by fewer agents. Hence, a smaller value of $\|c\|_2$ implies that agents would face higher strategic uncertainty about one another's actions, even when all the fundamental uncertainty is removed as $\sigma \to 0$. As a consequence, the set of rationalizable strategies shrinks as the norm of c decreases. In the extreme case that agents' information is only in the form of private signals (that is, when $m = n$ and $c_r = 1/n$ for all r), then $\|c\|_2^2 = 1/n$ and the upper and lower thresholds coincide ($\underline{\tau} = \bar{\tau} = 1/2$), implying the generic uniqueness of rationalizable strategies. This is indeed the case that corresponds to maximal level of strategic uncertainty. On the

[7] Since each agent observes a single signal, $c_1 + \cdots + c_m = 1$.

other hand, when all agents observe the same public signal (i.e., when $m = 1$ and $\|c\|_2^2 = 1$), they face no strategic uncertainty about each other's actions and hence, all undominated strategies are rationalizable.

We remark that given that the payoff functions have non-decreasing differences in the actions and the state, the Bayesian game under consideration is monotone supermodular in the sense of Van Zandt and Vives (2007). Therefore, there exist a greatest and a smallest Bayesian Nash equilibrium, both of which are in threshold strategies. Moreover, by Milgrom and Roberts (1990), all profiles of rationalizable strategies are "sandwiched" between these two equilibria. Therefore, Theorem 3 also provides a characterization of the set of equilibria of the game, showing that a higher level of common knowledge, captured via a larger value for $\|c\|_2$, implies a larger set of equilibria. These observations are consistent with the equilibrium uniqueness and multiplicity results (under private and public signals, respectively), already known in the global games literature.[8]

A simple corollary to Theorem 3 is that with $m < n$ sources of information, the set of Bayesian Nash equilibria is largest when $m - 1$ agents each observe a private signal and $n-m+1$ agents have access to the remaining signal. In this case, common knowledge between $n-m+1$ agents minimizes the extent of strategic uncertainty, and hence, leads to the largest set of equilibria. On the other hand, the set of equilibria becomes smaller whenever the sizes of the sets of agents with access to the same signal are more equalized. In particular, the case in which $c_r = 1/m$ for all r corresponds to the highest level of inter-group strategic uncertainty, leading to the greatest extent of refinement of rationalizable strategies.

Equilibrium Multiplicity in Large Coordination Games Recall from Theorem 1 that the existence of two agents i and j with identical observation sets is sufficient to guarantee equilibrium multiplicity, irrespective of the number of agents in the game or how much other agents care about coordinating with i and j. In

[8] Also note that if $m < n$, by construction, there are at least two agents with identical information sets. Theorem 3 then implies that $\underline{\tau} < 1/2 < \bar{\tau}$, which means that the game exhibits multiple equilibria, an observation consistent with Theorem 1.

particular, no matter how insignificant and uninformed the two agents are, the mere fact that i and j face no uncertainty regarding each other's behavior leads to equilibrium multiplicity. On the other hand, as Theorem 3 and the preceding discussion show, even under information structures that lead to multiplicity, the set of equilibria depends on the extent to which agents observe common signals. To further clarify the role of local information in determining the size of the equilibrium set, we next study large coordination games.

Formally, consider a sequence of games $\{G(n)\}_{n\in\mathbb{N}}$ parametrized by the number of agents. The payoff functions in the game with n agents are given by (3.3). We assume that each agent i in $G(n)$ can observe a single signal and that the noise terms in the signals are mutually independent and normally distributed with mean zero and variance $\sigma^2 > 0$. Remember that Theorem 3 characterizes the thresholds that define the set of rationalizable strategies in the limit as $\sigma \to 0$. Our following result considers the thresholds in the (second) limit as $n \to \infty$.

Proposition 3. *The sequence of games $\{G(n)\}_{n\in\mathbb{N}}$ exhibits an (essentially) unique equilibrium asymptotically as $n \to \infty$ and in the limit as $\sigma \to 0$ if and only if the size of the largest set of agents with a common observation grows sublinearly in n.*

Thus, as the number of agents grows, the game exhibits multiple equilibria if a non-trivial fraction of the agents have access to the same signal. Even though such a signal is not public — in the sense that it is not observed by all agents — the fact that it is common knowledge among a non-zero fraction of the agents implies that it can function as a coordination device, and hence, induce multiple equilibria. On the other hand, if the size of the largest subset of agents with common knowledge of a signal does not grow at the same rate as the number of agents, information is diverse and effectively private: any agent faces strategic uncertainty regarding the behavior of most other agents, even as all the fundamental uncertainty is removed ($\sigma \to 0$). Consequently, the set of rationalizable strategies of each agent collapses to a single strategy as the number of agents grows.

Chapter 4

Coordination with Exchange of Information

In this chapter, we consider an important special case of the model presented in Section 2.2, in which the agents are organized in a communication network. Each agent i observes an idiosyncratic signal x_i and in addition observes the idiosyncratic signals x_j of all her neighbors. Clearly, such a network structure induces an information structure of the type we considered in Section 2.2.

4.1 Model

We restrict our attention to the following form of payoff function:

$$u_i(a_i, a_{-i}, \theta) = \begin{cases} h(k/n) - \theta & \text{if } \ a_i = 1 \\ \\ 0 & \text{if } \ a_i = 0, \end{cases}$$

where $k = \sum_{j=1}^{n} a_j$ is the number of agents who take the risky action and $h : [0,1] \longrightarrow [0,1]$ is a non-decreasing function such that $h(0) = 0$ and $h(1) = 1$. We assume that h is common knowledge among all agents.

Agents hold a common prior belief on $\theta \in \mathbb{R}$, which for simplicity we assume to be the (improper) uniform distribution over the real line. Each agent i receives

an idiosyncratic noisy signal x_i about the state. Conditional on the state θ, agents' idiosyncratic signals are independent and identically distributed: $x_i = \theta + \xi_i$, where $\xi_i \sim \mathcal{N}(0, \sigma^2)$ are i.i.d. normal random variables, independent of θ, with mean zero and variance $\sigma^2 > 0$.

In addition to her idiosyncratic signal, each agent i observes the signals of a subset $\mathcal{N}_i \subseteq N$ of the other agents, called her *neighbors*. We specify this neighborhood relation by an undirected graph G, where each vertex corresponds to an agent, and where an edge (i, j) indicates that individuals i and j are neighbors. Throughout this chapter, we assume that G is common knowledge among all agents. We also use the convention that $i \in \mathcal{N}_i$ for all agents i. We use $\mathcal{V}(G)$ to denote the set of nodes of graph G.

4.2 Characterization of Strategy Profiles that Survive Iterated Elimination of Strictly Dominated Strategies (IESDS) for Finite Unions of Cliques

In this section, we characterize the set of strategy profiles that survive IESDS for finite networks that are unions of disconnected cliques. We first provide a generic solution, which we then use to come up with a closed-form solution for the special case of cliques of equal size. Our results establish multiplicity of equilibria for the case of finitely many agents. This multiplicity arises because agents in the same clique can use their shared information to coordinate their actions in multiple ways.

We note that the network without edges and the complete network correspond respectively to the cases of private and public information, discussed in Section 3.1.

4.2.1 Generic Characterization

Assume that the network consists of M disconnected cliques; clique $i \in \{1, \ldots, M\}$ has n_i nodes, and $\sum_{i=1}^{M} n_i = n$. We have the following result, which characterizes the

set of rationalizable strategies[1]:

Proposition 4 (Characterization for finite unions of cliques)**.** *There exist thresholds* $\{t_R^c, t_S^c\}_{c=1}^M$ *such that a strategy profile survives IESDS if and only if it satisfies the following: each agent* i *in clique* c *chooses the risky action if* $\frac{\sum_{j\in c} x_j}{n_c} < t_R^c$ *and chooses the safe action if* $\frac{\sum_{j\in c} x_j}{n_c} > t_S^c$*. Furthermore, the thresholds* $\{t_R^c, t_S^c\}_{c=1}^M$ *solve the following system of equations (here a choice of* l *corresponds to selecting* r *out of the* $M-1$ *cliques):*

$$
\begin{aligned}
t_R^c &= h\left(\frac{n-n_c+1}{n}\right)\mathbb{P}\left(\forall d\neq c, \frac{\sum_{j\in d} x_j}{n_d} < t_R^d \;\middle|\; \frac{\sum_{j\in c} x_j}{n_c} = t_R^c\right) \\
&\quad + \sum_{r=1}^{M-2}\sum_{l=1}^{\binom{M-1}{r}} h\left(\frac{n-\sum_{d\neq c, d \text{ selected by } l} n_d - n_c + 1}{n}\right) p_R^{r,l} \\
&\quad + h\left(\frac{1}{n}\right)\mathbb{P}\left(\forall d\neq c, \frac{\sum_{j\in d} x_j}{n_d} \geq t_R^d \;\middle|\; \frac{\sum_{j\in c} x_j}{n_c} = t_R^c\right), \qquad \forall c\in\{1,\dots,M\}, \\
t_S^c &= h(1)\mathbb{P}\left(\forall d\neq c, \frac{\sum_{j\in d} x_j}{n_d} \leq t_S^d \mid \frac{\sum_{j\in c} x_j}{n_c} = t_S^c\right) \\
&\quad + \sum_{r=1}^{M-2}\sum_{l=1}^{\binom{M-1}{r}} h\left(\frac{n-\sum_{d\neq c, d \text{ selected by } l} n_d}{n}\right) p_S^{r,l} \\
&\quad + h\left(\frac{n_c}{n}\right)\mathbb{P}\left(\forall d\neq c, \frac{\sum_{j\in d} x_j}{n_d} > t_S^d \mid \frac{\sum_{j\in c} x_j}{n_c} = t_S^c\right), \qquad \forall c\in\{1,\dots,M\},
\end{aligned}
$$

and where

$$
p_R^{r,l} = \mathbb{P}\left(\text{ only for the } r \text{ cliques selected by } l \;\frac{\sum_{j\in d} x_j}{n_d} \geq t_R^d \middle| \frac{\sum_{j\in c} x_j}{n_c} = t_R^c\right)
$$

and

$$
p_S^{r,l} = \mathbb{P}\left(\text{only for the } r \text{ cliques selected by } l \;\frac{\sum_{j\in d} x_j}{n_d} > t_S^d \mid \frac{\sum_{j\in c} x_j}{n_c} = t_S^c\right).
$$

Notice that due to our normality assumptions, $\frac{\sum_{j\in c} x_j}{n_c}$ is a sufficient statistic for

[1]Theorem 3 also characterizes the set of rationalizable strategies, although for the case where each agent receives exactly one signal, and in the limit as $\sigma \to 0$.

$\{x_j\}_{j\in c}$ with respect to θ, and hence with respect to the signals of other cliques, for all cliques c.

Example 1. We showcase the characterization for the simple network of Figure 4-1. The relevant thresholds, $\{t_R^{\{1,2\}}, t_S^{\{1,2\}}, t_R^{\{3\}}, t_S^{\{3\}}\}$, satisfy the following system of

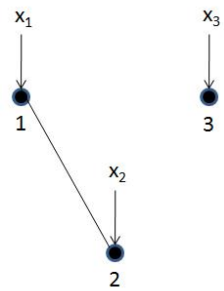


Figure 4-1: A simple network.

equations:

$$
\begin{aligned}
t_R^{\{1,2\}} &= h\left(\frac{2}{3}\right)\mathbb{P}\left(x_3 < t_R^{\{3\}} \;\middle|\; \frac{x_1+x_2}{2} = t_R^{\{1,2\}}\right) + h\left(\frac{1}{3}\right)\mathbb{P}\left(x_3 \geq t_R^{\{3\}} \;\middle|\; \frac{x_1+x_2}{2} = t_R^{\{1,2\}}\right)\\
t_S^{\{1,2\}} &= h(1)\mathbb{P}\left(x_3 \leq t_S^{\{3\}} \;\middle|\; \frac{x_1+x_2}{2} = t_S^{\{1,2\}}\right) + h\left(\frac{2}{3}\right)\mathbb{P}\left(x_3 > t_S^{\{3\}} \;\middle|\; \frac{x_1+x_2}{2} = t_S^{\{1,2\}}\right)\\
t_R^{\{3\}} &= h(1)\mathbb{P}\left(\frac{x_1+x_2}{2} < t_R^{\{1,2\}} \;\middle|\; x_3 = t_R^{\{3\}}\right) + h\left(\frac{1}{3}\right)\mathbb{P}\left(\frac{x_1+x_2}{2} \geq t_R^{\{1,2\}} \;\middle|\; x_3 = t_R^{\{3\}}\right)\\
t_S^{\{3\}} &= h(1)\mathbb{P}\left(\frac{x_1+x_2}{2} \leq t_S^{\{1,2\}} \;\middle|\; x_3 = t_S^{\{3\}}\right) + h\left(\frac{1}{3}\right)\mathbb{P}\left(\frac{x_1+x_2}{2} \geq t_S^{\{1,2\}} \;\middle|\; x_3 = t_S^{\{3\}}\right).
\end{aligned}
$$

The system can be solved numerically. The solutions for different values of σ, for the case when $h(x) = x$, are shown in Table 3. For small values of σ, the thresholds for

Table 4.1: Thresholds for strategies that survive IESDS for the network in Figure 4-1, for $h(x) = x$, and different values of σ.

	$\sigma = 0.01$	$\sigma = 1$	$\sigma = 100$
Agents $1, 2$	$t_R = 0.5474, t_S = 0.7859$	$t_R = 0.5132, t_S = 0.8202$	$t_R = 0.5002, t_S = 0.8332$
Agent 3	$t_R = 0.5518, t_S = 0.7815$	$t_R = 0.6345, t_S = 0.6989$	$t_R = 0.6662, t_S = 0.6671$

the clique of two agents and the single agent are close. As the noise increases, the extent of multiplicity (more formally, the difference $t_S - t_R$) grows larger for the two agents in the clique $\{1, 2\}$, but smaller for the single agent 3. For the clique of two

agents, less precise signals lead to multiplicity of greater extent; on the contrary, for the single agent, a less precise signal leads to more refined multiplicity, which gets very close to uniqueness for large σ. We notice that the difference $t_S - t_R$ is more sensitive to changes in the noise for the single agent than it is for the clique of two agents.

4.2.2 Characterization in the Case of Cliques of Equal Size

In the special case of M equally sized cliques, we can write down closed-form expressions for t_R and t_S. Notice that each clique has size n/M.

Corollary 1. *A strategy profile survives IESDS if and only if it satisfies the following. Each agent i in clique c chooses R if $\frac{\sum_{j\in c} x_j}{n/M} < t_R$, and S if $\frac{\sum_{j\in c} x_j}{n/M} > t_S$, where*

$$t_R = \frac{1}{M}\left(h\left(\frac{n-n/M+1}{n}\right) + \sum_{r=1}^{M-2} h\left(\frac{n-r(n/M)-(n/M)+1}{n}\right) + h\left(\frac{1}{n}\right)\right),$$

and

$$t_S = \frac{1}{M}\left(h(1) + \sum_{r=1}^{M-2} h\left(\frac{n-r(n/M)}{n}\right) + h\left(\frac{n/M}{n}\right)\right).$$

Example 2. We showcase the characterization for the simple network of Figure 4-2. The thresholds are given by

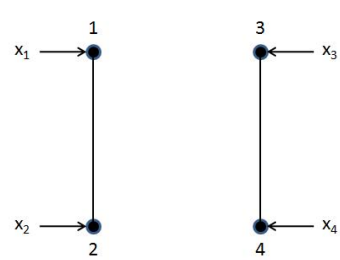


Figure 4-2: A simple network with cliques of equal sizes.

$$t_R = \frac{1}{2}\left(h\left(\frac{3}{4}\right) + h\left(\frac{1}{4}\right)\right)$$

$$t_S = \frac{1}{2}\left(h(1) + h\left(\frac{1}{2}\right)\right).$$

Notice that $t_R \leq \frac{1}{4}\left(h(1) + h\left(\frac{3}{4}\right) + h\left(\frac{2}{4}\right) + h\left(\frac{1}{4}\right)\right) \leq t_S$, where

$$\frac{1}{4}\left(h(1) + h\left(\frac{3}{4}\right) + h\left(\frac{2}{4}\right) + h\left(\frac{1}{4}\right)\right)$$

is the threshold pertaining to the network of four disconnected agents. Thus, there exist equilibria involving a threshold which is larger than that for the disconnected network (i.e., the society is "braver") as well as equilibria involving a threshold which is smaller than that for the disconnected network (i.e., the society is "less brave"). In particular, more communication can make society either more or less brave.

4.3 The Case of Asymptotically Many Agents

We have seen that for a finite number of agents, links induce multiplicity of strategy profiles that survive IESDS: in a network consisting of finitely many disconnected agents, there is a unique strategy profile that survives iterated elimination of strictly dominated strategies, and therefore a unique Bayesian Nash equilibrium; in sharp contrast, introducing a single link between any two agents induces multiplicity. The natural question that arises is under what conditions on the network there exists asymptotically a unique strategy profile that survives IESDS, and thus a unique Bayesian Nash equilibrium. In this section we provide sufficient conditions for uniqueness in the case of an asymptotically large number of agents.

We consider a growing sequence of graphs G_k, $k = 1, 2, \ldots$, with $G_k \subseteq G_{k+1}$. The graph G_k consists of $g(k)$ cliques of equal size $f(k) \geq 1$, for a total of $n(k) = f(k)g(k)$ nodes. We assume that the function f is nondecreasing in k. For example, the graph could grow by adding more cliques of the same size, or by merging existing cliques to form larger cliques. For any k, the strategy profiles that survive IESDS are described by the thresholds in Corollary 1. We denote these thresholds by t_R^k and t_S^k to indicate explicitly the dependence on k.

The proposition that follows shows that if the number of cliques grows to infinity (equivalently, if the clique size grows sublinearly with the number of agents), then t_R^k and t_S^k converge to a common value, as k increases. Thus, loosely speaking, in the limit, there is an essentially unique strategy that survives IESDS, and therefore a unique Bayesian Nash equilibrium.

4.3.1 Sufficient Conditions for Asymptotic Uniqueness

Proposition 5 (Asymptotic uniqueness)**.** *Suppose that* $\lim_{m\to\infty} g(m) = \infty$*. Then,*

$$\lim_{k\to\infty} t_R^k = \lim_{k\to\infty} t_S^k = \int_0^1 h(x)dx.$$

Proof. By Corollary 1, we have

$$\begin{aligned}
\lim_{k\to\infty} t_S^k &= \lim_{k\to\infty} \frac{1}{g(k)} \sum_{j=1}^{g(k)} h\left(\frac{jf(k)}{n(k)}\right) \\
&= \lim_{k\to\infty} \frac{1}{g(k)} \sum_{j=1}^{g(k)} h\left(\frac{j}{g(k)}\right) \\
&= \int_0^1 h(x)dx.
\end{aligned}$$

Again by Corollary 1, we have

$$\lim_{k\to\infty} t_R^k = \lim_{k\to\infty} \frac{1}{g(k)} \sum_{j=0}^{g(k)-1} h\left(\frac{jf(k)+1}{n(k)}\right).$$

Notice that for all k,

$$\lim_{k\to\infty} \frac{1}{g(k)} \sum_{j=0}^{g(k)-1} h\left(\frac{jf(k)}{n(k)}\right) \leq \lim_{k\to\infty} \frac{1}{g(k)} \sum_{j=0}^{g(k)-1} h\left(\frac{jf(k)+1}{n(k)}\right) \leq \lim_{k\to\infty} \frac{1}{g(k)} \sum_{j=1}^{g(k)} h\left(\frac{jf(k)}{n(k)}\right).$$

We showed above that

$$\lim_{k\to\infty} \frac{1}{g(k)} \sum_{j=1}^{g(k)} h\left(\frac{jf(k)}{n(k)}\right) = \int_0^1 h(x)dx.$$

Similarly, we have

$$\begin{aligned}
\lim_{k\to\infty} \frac{1}{g(k)} \sum_{k=0}^{g(k)-1} h\left(\frac{jf(k)}{n(k)}\right) &= \lim_{k\to\infty} \frac{1}{g(k)} \left(\sum_{k=0}^{g(k)} h\left(\frac{jf(k)}{n(k)}\right) - h(1)\right) \\
&= \lim_{k\to\infty} \frac{1}{g(k)} \sum_{k=1}^{g(k)} h\left(\frac{j}{g(k)}\right) \\
&= \int_0^1 h(x)dx.
\end{aligned}$$

By a standard sandwich argument, it follows that

$$\lim_{k\to\infty} t_R^k = \int_0^1 h(x)dx.$$

□

We note that we can use a similar argument to establish that if the growth of each clique is linear in n, then we have multiplicity: the thresholds satisfy

$$\lim_{k\to\infty} t_R^k < \lim_{k\to\infty} t_S^k.$$

4.3.2 Interpretation

Proposition 5 extends to the case of unions of cliques of unequal sizes, when the fastest-growing clique grows sublinearly with the total number of agents n. We also note that the case, in Proposition 5, of sublinearly growing cliques leads to the same asymptotic equilibrium analysis as the case of disconnected agents. Loosely speaking, in general we can view the properties of equilibria for the case of asymptotically many agents as being governed by two competing effects: the network is growing, and the sharing of information among agents is also growing. An intuitive explanation why the equilibrium analysis for the two aforementioned sequences of networks is asymptotically the same, and yields uniqueness, is the following: the two sequences of networks, in the limit of large n, are informationally equivalent; precisely, for both sequences of networks the growth of information sharing is insignificant compared to

the growth of the network, and this gap induces a unique equilibrium. In turn, we can view uniqueness of equilibria as predictability of individual behavior.

On the other hand, in the case of unions of disconnected cliques, when the fastest-growing clique grows linearly with the total number of agents n, there are asymptotically infinitely many strategy profiles that survive IESDS. The intuitive interpretation is that the growth of the sharing of information among agents is comparable to the growth of the network; the excess in communication is what breaks uniqueness (and predictability of individual behavior).

There are thus no sequences of networks that are unions of disconnected cliques for which uniqueness of Bayesian Nash equilibrium is obtained asymptotically, at a unique threshold other than $\int_0^1 h(x)dx$. Therefore, we cannot come up with sequences of networks for which the unique threshold shifts, signifying a societal shift in favor of or against taking a risky action. (A shift of the threshold to higher values would signify that rational individuals are willing to play the risky action over a signal space that includes higher observations, corresponding to higher realizations of the fundamentals. A shift of the threshold to lower values would signify that rational individuals are only willing to play the risky action over a signal space that is limited to lower observations, corresponding to lower realizations of the fundamentals.)

Chapter 5

The Value of Temporal Data for Learning of Influence Networks: Introduction

Consumers adopting a new product (Kempe, Kleinberg, and Tardos, 2003); an epidemic spreading across a population (Newman, 2002); a sovereign debt crisis hitting several countries (Glover and Richards-Shubik, 2013); a cellular process during which the expression of a gene affects the expression of other genes (Song, Kolar, and Xing, 2009); an article trending in the blogosphere (Lerman and Ghosh, 2010), a topic trending on an online social network (Zhou, Bandari, Kong, Qian, and Roychowdhury, 2010), computer malware spreading across a network (Kephart and White, 1991); all of these are temporal processes governed by local interactions of networked entities, which influence one another. Due to the increasing capability of data acquisition technologies, rich data on the outcomes of such processes are oftentimes available (possibly with time stamps), yet the underlying network of local interactions is hidden. In this work, we infer who influences whom in a network of interacting entities based on data of their actions/decisions, and quantify the gain of learning based on times of actions, versus sequences of actions, versus sets of actions. We answer the following question: how much faster can we learn influences with access to increasingly informative temporal data (sets versus sequences versus times)?

We start with a toy example to motivate the main question. Consider agents $1, \ldots, 5$, who sequentially make a decision (such as adopting a new product). We have access to a database that contains information on adoptions for each product[1]. We consider three different possibilities for the database: the database could be storing the set of adopters per product; it could be storing the sequence of adopters per product; or it could be storing the time stamps of the adoptions (and the adopters) per product. Table 5.1 shows what records would look like for each of the three types of database, for four products.

Table 5.1: Toy example with four records of adoption decisions by five agents, and how the same events are encoded in each of three different databases: a database that stores times, a database that stores sequences, and a database that stores sets.

	Times	**Sequences**	**Sets**
Record 1	2 1 5	$2, 1, 5$	$\{1, 2, 5\}$
Record 2	1 5 4	$1, 5, 4$	$\{1, 4, 5\}$
Record 3	5 3	$5, 3$	$\{3, 5\}$
Record 4	4 1 5 2	$4, 1, 5, 2$	$\{1, 2, 4, 5\}$

Assume that part of the ground truth is that there is high influence from agent 1 to agent 5. Does each type of database provide evidence for the high (directed) influence between these two agents, and how? If we have access to times, then it is easy to argue that evidence exists: whenever agent 1 adopts, agent 5 adopts shortly afterwards, consistently. If we have access to sequences, again evidence of the high influence from agent 1 to agent 5 exists: whenever agent 1 adopts, agent 5 is consistently the next agent to adopt. Surprisingly, it can be argued that evidence exists even with access to mere sets: whenever agent 1 adopts, agent 5 also adopts, consistently; yet, there are instances when agent 5 adopts, but agent 1 does not.

Clearly, having access to richer temporal information allows, in general, for faster and more accurate learning. Nevertheless, in some contexts, the temporally poor data mode of sets could provide almost all the information needed for learning, or

[1]The word "product", which we use throughout, could be interchanged by any of the following: information, behavior, opinion, virus, bankruptcy, etc., depending on the context.

at least suffice to learn key network relations. In addition, collecting, organizing, storing, and processing temporally richer data may require more effort and more cost. In some contexts, data on times of actions, or even sequences of actions, is noisy and unreliable; for example, the time marking of epilepsy seizure events, which are studied in Chapter 9, is done by physicians on an empirical basis and is not exact. In some other contexts, having access to time stamps or sequences of actions is almost impossible. For example, in the context of retailing, data exist on sets of purchased items per customer (and are easily obtained by scanning the barcodes at checkout); however, no data exist on the order in which the items a customer checked out were picked up from the shelf (and obtaining such data would be practically hard). In this light, the question of quantifying the gain of learning with increasingly informative temporal data, and understanding in what scenarios learning with temporally poor data modes is good enough, is highly relevant in various contexts.

5.1 Background and Related Literature

Untangling and quantifying local influences in a principled manner, based on observed outcomes, is a challenging task, as there are many different confounding factors that may lead to seemingly similar phenomena. In recent work, inference of causal relationships has been possible from multivariate time-series data (Lozano and Sindhwani, 2010; Materassi and Salapaka, 2012; Kolar, Song, Ahmed, and Xing, 2010). Solutions for the influence discovery problem have been proposed, which, similarly to this work, treat time explicitly as a continuous random variable and infer the network through cascade data, e.g., Du, Song, Smola, and Yuan (2012); Myers and Leskovec (2010); Gomez-Rodriguez, Leskovec, and Krause (2010); Gomez-Rodriguez, Balduzzi, and Schölkopf (2011). However, the focus of our work is not just to infer the underlying network, but rather to quantify the gain in speed of learning, due to having access to richer temporal information.

Most closely related to this work are Abrahao, Chierichetti, Kleinberg, and Panconesi (2013); Netrapalli and Sanghavi (2012); Gripon and Rabbat (2013). The first

two works present quantitative bounds on sample/trace complexity for various epidemic models; the models assumed by Abrahao, Chierichetti, Kleinberg, and Panconesi (2013), as well as Netrapalli and Sanghavi (2012), differ from the model we study, mainly in that we allow for self-induced infections (not just in the initial seeding), which makes the inference problem harder. Similarly to Abrahao, Chierichetti, Kleinberg, and Panconesi (2013), we assume exponentially distributed infection times, yet we are after the influence rate of each edge, rather than learning whether each edge exists or not. Gripon and Rabbat (2013) share with us the question of reconstructing a graph from traces defined as sets of unordered nodes, but they restrict their attention to the case where each observation/trace consists of nodes getting infected along a single path. Our scope differs from the works mentioned above, as we wish to compare explicitly the speed of learning when having access to datasets with times of actions, versus just sequences of actions, versus just sets of actions.

Recent research has focused on learning graphical models (which subsumes the question of identifying the connectivity in a network), either allowing for latent variables (e.g., Chandrasekaran, Parrilo, and Willsky, 2012; Choi, Tan, Anandkumar, and Willsky, 2011) or not (e.g., Anandkumar, Tan, Huang, and Willsky, 2012). Instead of proposing and learning a general graphical model, we focus on a simple parametric model that can capture the sequence and timing of actions naturally, without the descriptive burden of a standard graphical model.

Of relevance is also Shah and Zaman (2011), in which knowledge of both the graph and the set of infected nodes is used to infer the original source of an infection. In contrast, and somewhat conversely, we use knowledge of the set, order, or times of infections to infer the graph.

Last, economists have addressed the problem of identification in social interactions (e.g., Manski, 1993; Brock and Durlauf, 2001; Blume, Brock, Durlauf, and Ioannides, 2011; Bramoullé, Djebbari, and Fortin, 2009; Durlauf and Ioannides, 2010) focusing on determining aggregate effects of influence in a group; they classify social interactions into an endogenous effect, which is the effect of group members' behaviors on individual behavior; an exogenous (contextual) effect, which is the effect of group

members' observable characteristics on individual behavior; and a correlated effect, which is the effect of group members' unobservable characteristics on individual behavior. In sharp contrast, our approach identifies influence at the individual, rather than the aggregate, level.

5.2 Overview

The overarching theme of our work is to quantify the gain in speed of learning of parametric models of influence, due to having access to richer temporal information. We seek to compare the speed of learning under three different cases of available data: (i) the data provides merely the set of agents/entities who took an action; (ii) the data provides the (ordered) sequence of agents/entities who took an action, but not the times; and (iii) the data provides the times of the actions. It is clear that learning is no slower with times than it is with sequences, and no slower with sequences than with sets; yet, what can we say about *how much* faster learning is with times than with sequences, and with sequences than with sets? This is, to the best of our knowledge, a comparison that has not been studied systematically before.[2]

We propose a parametric model of influence which captures directed pairwise interactions. We first use the Fisher information and the Kullback-Leibler (KL) divergence as measures for the speed of learning.

Using the concept of Fisher information, we compute the gap between learning with sets and learning with sequences and provide an example of a network where the gap is zero, and an example of a network where the gap can be arbitrarily large. The inverse of the Fisher information matrix is the covariance matrix of the multivariate normal to which the normalized maximum likelihood estimates converge in distribution (under technical conditions); thus the Fisher information is a measure of how fast the estimates converge to the true parameters, and therefore a measure of the speed of learning.

Using the concept of Kullback-Leibler divergence, we focus on learning the in-

[2]Netrapalli and Sanghavi (2012) find such a comparison highly relevant.

fluence model in three particular instances, which we cast as respective binary hypothesis testing problems: Which of two agents influences a third agent? Is an agent influenced by another agent, or are her decisions self-induced? And is the influence between two agents large or small? We view these three questions as building blocks for understanding complex interactions in general networks. Given a hypothesis testing problem, the Kullback-Leibler divergence between the distributions pertaining to the two competing hypotheses yields the best achievable asymptotic exponent for the decaying of the probability of error in a Neyman-Pearson setting, and is therefore a measure for the speed of learning. For each of the proposed hypothesis testing problems, we compare the Kullback-Leibler divergences in the cases of learning based on data of sets of decisions; learning based on data of sequences of decisions; and learning based on data of times of decisions.

We then provide theoretical guarantees on the sample complexity for correct learning with sets, sequences, and times. Our results characterize the sufficient and necessary scaling of the number of i.i.d. samples required for correct learning. The asymptotic gain of having access to richer temporal data à propos of the speed of learning is thus quantified in terms of the gap between the derived asymptotic requirements under different data modes. We first restrict to the case where each edge carries an influence rate of either zero or infinity; we provide sufficient and necessary conditions on the graph topology for learnability, and we come up with upper and lower bounds for the minimum number of i.i.d. samples required to learn the correct hypothesis for the star topology, for different variations of the learning problem: learning one edge or learning all the edges, under different prior knowledge over the hypotheses, under different scaling of the horizon rate, and learning with sets or with sequences. We then relax the assumption that each edge carries an influence rate of either zero or infinity, and provide a learning algorithm and theoretical guarantees on the sample complexity for correct learning between the complete graph and the complete graph that is missing one edge.

We also evaluate learning with sets, sequences, and times *experimentally*. Given data on outcomes, we learn the parametric influence model by maximum likelihood

estimation. On both synthetic and real datasets, the value of learning with data of richer temporal detail is quantified, and our methodology is shown to recover the underlying network structure well. The real data come from (i) observations of mobile app installations of users, along with data on their communications and social relations; (ii) observations of levels of neuronal activity in different regions of the brain during epileptic seizure events, for different patients.

5.3 The Influence Model

A product becomes available at time $t = 0$ and each of a collection of agents may adopt it or not. Agent i adopts it at a time that is exponentially distributed with rate[3] $\lambda_i \geq 0$. After agent i adopts, the rate of adoption for all other agents $j \neq i$ increases by $\lambda_{ij} \geq 0$. The overall time horizon is modeled as an exponentially distributed random variable with rate λ_{hor}. No adoptions are possible after the end of the horizon. We study the adoption decisions for a collection of products, assuming that the parameters are static across products, and adoptions across products are independent.

[3]The probability density function we use for the exponential distribution with rate λ is $f(x) = \lambda e^{-\lambda x}, x \geq 0$, with expectation $1/\lambda$.

Chapter 6

Characterization of the Speed of Learning Using the Fisher Information and the Kullback-Leibler Divergence

In this chapter we characterize the speed of learning of the parameters of the influence model using the Fisher information and the Kullback-Leibler divergence for specific simple networks, and compare the speed of learning under each of the following scenarios:

- the available data provides information on who adopted each product;
- the available data provides information on who adopted each product, and in what order;
- the available data provides information on who adopted each product, and at what exact time.

Using the concept of Fisher information, we compute the gap between learning with sets and learning with sequences and provide an example of a network where the gap is zero, and an example of a network where the gain can be arbitrarily large.

Using the concept of Kullback-Leibler divergence, we compare learning with sets, sequences, and times for three binary hypothesis testing problems; for each problem, we characterize the gain of learning with temporally richer data for small, moderate, and large horizon.

6.1 Characterization of the Speed of Learning Using the Fisher Information

To learn a parametric model from data, we need to propose an estimator that satisfies desired properties. A standard selection for the estimator is the maximum-likelihood (ML) estimator, which, under mild assumptions, is parametrically consistent, in the sense that the ML estimates converge in probability to the true parameters, as the number of observations grows. We are interested in the speed of convergence, and in comparing the speed when learning with sets of adoptions to the speed when learning with sequences of adoptions.

Under technical conditions[1], the ML estimator also satisfies asymptotic normality[2]: the normalized vector of ML estimates $\sqrt{k}\left(\hat{\boldsymbol{\theta}}_{ML} - \boldsymbol{\theta}_{truth}\right)$, where k is the number of observations/products, converges in distribution to a multivariate normal, with mean equal to the vector of zeros and covariance matrix $\mathcal{J}^{-1}$, where

$$\mathcal{J}_{ij}(\boldsymbol{\theta}) = \mathbb{E}\left[\left(\frac{\partial}{\partial\theta_i}\log f(X;\boldsymbol{\theta})\right)\left(\frac{\partial}{\partial\theta_j}\log f(X;\boldsymbol{\theta})\right)\middle|\,\boldsymbol{\theta}\right]$$

are the elements of the Fisher information matrix $\mathcal{J}(\boldsymbol{\theta})$, X the observable random variable corresponding to outcomes[3], and $f(X;\boldsymbol{\theta})$ the probability mass or probability density[4] function of X under a given value of $\boldsymbol{\theta}$.

[1]See Newey and McFadden (1994). Most importantly, the first and second derivative of the log-likelihood function must be defined; the Fisher information matrix must be nonzero, and must be continuous as a function of the parameter; and the maximum likelihood estimator must be consistent.

[2]The ML estimator is also efficient: it achieves the Cramér-Rao bound asymptotically, as the sample size tends to infinity.

[3]The space of outcomes will be different, depending on whether we are learning with sets, sequences, or times.

[4]Probability density functions arise when learning with times. We do not consider learning with

In this light, the Fisher Information matrix $\mathcal{J}(\boldsymbol{\theta})$ is a measure of how fast the estimates converge to the true parameters, therefore a measure of the speed of learning.

6.1.1 The Gap between Sets and Sequences, and a Network where It Is Zero

We fix the number of agents n, and index the subsets of n agents using $i = 1, \ldots, 2^n$. For fixed i, we index the permutations of agents in set i using $j = 1, \ldots, |i|!$. Define p_{ij} to be the probability that we see a sequence of adopters that correspond to permutation j of set i. Define p_i to be the probability that the set of adopters is set i. Then,

$$p_i = \sum_{j=1}^{|i|!} p_{ij}.$$

For example, for $n = 2$, the probability of the set of agents $\{1, 2\}$ adopting is the sum of the probabilities of the two following events: agent 1 adopts first, agent 2 adopts second; and agent 2 adopts first, agent 1 adopts second; defining $\kappa_i = \frac{\lambda_i}{\lambda_{hor}}, \kappa_{ij} = \frac{\lambda_{ij}}{\lambda_{hor}}$, the probability becomes

$$\frac{\kappa_1}{\kappa_1 + \kappa_2 + 1} \cdot \frac{\kappa_2 + \kappa_{12}}{\kappa_2 + \kappa_{12} + 1} + \frac{\kappa_2}{\kappa_1 + \kappa_2 + 1} \cdot \frac{\kappa_1 + \kappa_{21}}{\kappa_1 + \kappa_{21} + 1}.$$

Letting $\boldsymbol{\kappa} = (\kappa_1, \kappa_2, \ldots, \kappa_n, \kappa_{12}, \ldots, \kappa_{n,n-1})$, we want to determine the $(n+1)$st row, $(n+1)$st column element of each of the Fisher information matrices $\mathcal{J}^{\text{set}}(\boldsymbol{\kappa}), \mathcal{J}^{\text{sequence}}(\boldsymbol{\kappa})$, for which we use the notation $\mathcal{J}^{\text{set}}_{12,12}(\boldsymbol{\kappa}), \mathcal{J}^{\text{sequence}}_{12,12}(\boldsymbol{\kappa})$, respectively. Our focus is on the ratio $\frac{\mathcal{J}^{\text{sequence}}_{12,12}}{\mathcal{J}^{\text{set}}_{12,12}}$. Of course, the ratio is at least 1. We want to characterize explicitly how much larger $\mathcal{J}^{\text{sequence}}_{12,12}$ is.

We have

$$\begin{aligned}
\mathcal{J}^{\text{set}}_{12,12}(\boldsymbol{\kappa}) &= \sum_{i=1}^{2^n} p_i \left(\frac{\partial}{\partial \kappa_{12}} \log p_i \right)^2 \\
&= \sum_{i=1}^{2^n} \frac{1}{p_i} \left(\frac{\partial p_i}{\partial \kappa_{12}} \right)^2,
\end{aligned}$$

times using Fisher information, but we return to learning with times in the next section.

and

$$\begin{aligned}\mathcal{J}_{12,12}^{\text{sequence}}(\boldsymbol{\kappa}) &= \sum_{i=1}^{2^n}\sum_{j=1}^{|i|!} p_{ij}\left(\frac{\partial}{\partial\kappa_{12}}\log p_{ij}\right)^2 \\ &= \sum_{i=1}^{2^n}\sum_{j=1}^{|i|!}\frac{1}{p_{ij}}\left(\frac{\partial p_{ij}}{\partial\kappa_{12}}\right)^2.\end{aligned}$$

The difference between the ith summand of $\mathcal{J}_{12,12}^{\text{sequence}}(\boldsymbol{\kappa})$ and the ith summand of $\mathcal{J}_{12,12}^{\text{set}}(\boldsymbol{\kappa})$ can be written as:

$$\begin{aligned}&i\text{th summand of } \mathcal{J}_{12,12}^{\text{sequence}}(\boldsymbol{\kappa}) - i\text{th summand of } \mathcal{J}_{12,12}^{\text{set}}(\boldsymbol{\kappa}) \\ &= \textstyle\sum_{j=1}^{|i|!}\frac{1}{p_{ij}}\left(\frac{\partial p_{ij}}{\partial\kappa_{12}}\right)^2 - \frac{1}{p_i}\left(\frac{\partial p_i}{\partial\kappa_{12}}\right)^2 \\ &= \textstyle\sum_{j=1}^{|i|!}\frac{1}{p_{ij}}\left(\frac{\partial p_{ij}}{\partial\kappa_{12}}\right)^2 - \frac{1}{\sum_{j=1}^{|i|!}p_{ij}}\left(\sum_{j=1}^{|i|!}\frac{\partial p_{ij}}{\partial\kappa_{12}}\right)^2 \\ &= \textstyle\frac{1}{\sum_{j=1}^{|i|!}p_{ij}}\sum_{j=1}^{|i|!-1}\sum_{k=j+1}^{|i|!}\left(\sqrt{\frac{p_{ik}}{p_{ij}}}\frac{\partial p_{ij}}{\partial\kappa_{12}} - \sqrt{\frac{p_{ij}}{p_{ik}}}\frac{\partial p_{ik}}{\partial\kappa_{12}}\right)^2,\end{aligned}$$

while the ratio can be written as

$$\begin{aligned}\frac{i\text{th summand of } \mathcal{J}_{12,12}^{\text{sequence}}(\boldsymbol{\kappa})}{i\text{th summand of } \mathcal{J}_{12,12}^{\text{set}}(\boldsymbol{\kappa})} &= 1 + \frac{i\text{th summand of } \mathcal{J}_{12,12}^{\text{sequence}}(\boldsymbol{\kappa}) - i\text{th summand of } \mathcal{J}_{12,12}^{\text{set}}(\boldsymbol{\kappa})}{i\text{th summand of } \mathcal{J}_{12,12}^{\text{set}}(\boldsymbol{\kappa})} \\ &= 1 + \frac{\frac{1}{\sum_{j=1}^{|i|!}p_{ij}}\sum_{j=1}^{|i|!-1}\sum_{k=j+1}^{|i|!}\left(\sqrt{\frac{p_{ik}}{p_{ij}}}\frac{\partial p_{ij}}{\partial\kappa_{12}} - \sqrt{\frac{p_{ij}}{p_{ik}}}\frac{\partial p_{ik}}{\partial\kappa_{12}}\right)^2}{\frac{1}{\sum_{j=1}^{|i|!}p_{ij}}\left(\sum_{j=1}^{|i|!}\frac{\partial p_{ij}}{\partial\kappa_{12}}\right)^2} \\ &= 1 + \frac{\sum_{j=1}^{|i|!-1}\sum_{k=j+1}^{|i|!}\left(\sqrt{\frac{p_{ik}}{p_{ij}}}\frac{\partial p_{ij}}{\partial\kappa_{12}} - \sqrt{\frac{p_{ij}}{p_{ik}}}\frac{\partial p_{ik}}{\partial\kappa_{12}}\right)^2}{\left(\sum_{j=1}^{|i|!}\frac{\partial p_{ij}}{\partial\kappa_{12}}\right)^2}. \qquad (6.1)\end{aligned}$$

Can we think of a case where there is no gain in having information on the order of adoptions per product, beyond knowing just who adopted what? Consider the case where for all agents i, j, we have $\kappa_i = \kappa_{ij} = \kappa > 0$. Then the influence model has only one parameter, κ, and the Fisher information is a scalar, $\mathcal{J}(\kappa)$. For all $i = 1, \ldots, 2^n$

and $j = 1, \ldots, |i|!$, we have

$$p_{ij} = \frac{\kappa}{n\kappa + 1} \cdot \frac{2\kappa}{2(n-1)\kappa + 1} \cdot \frac{3\kappa}{3(n-2)\kappa + 1} \cdots \frac{|i|\kappa}{|i|(n - (|i| - 1))\kappa + 1} \cdot \frac{1}{(|i| + 1)(n - |i|)\kappa + 1}. \tag{6.2}$$

Notice that p_{ij} only depends on the size of the set, $|i|$. It does not depend on the permutation j and the only way it depends on the set is through the set's size. Because each of the squared differences in Equation (6.1) will be zero, we conclude that

$$\frac{\mathcal{J}^{\text{sequence}}(k)}{\mathcal{J}^{\text{set}}(\kappa)} = 1.$$

Having information on the sequence of adoptions provides no gain asymptotically over having information merely on the set of adopters.

A stronger statement can be made: in this case, *the set of adopters is a sufficient statistic for the sequence of adopters relative to k.* Indeed, we have

$$\begin{aligned} \mathbb{P}\left(\text{sequence} = j \mid \text{set} = i; \kappa\right) &= \frac{\mathbb{P}\left(\text{sequence} = j, \text{ set} = i; \kappa\right)}{\mathbb{P}\left(\text{set} = i; \kappa\right)} \\ &= \begin{cases} \frac{\mathbb{P}(\text{sequence} = j; \kappa)}{\mathbb{P}(\text{set} = i; \kappa)} = \frac{1}{|i|!} & \text{if } j \text{ is valid permutation of } i \\ 0 & \text{otherwise,} \end{cases} \end{aligned}$$

i.e., the conditional distribution of the sequence of adopters (the data), given the set of adopters (the statistic), does not depend on the parameter κ.

In fact, the set of adopters is not a minimal sufficient statistic, because it is not a function of just the number of adopters, which is another sufficient statistic:

$$\begin{aligned} \mathbb{P}\left(\text{sequence} = j \mid \text{number} = \ell; \kappa\right) &= \frac{\mathbb{P}\left(\text{sequence} = j, \text{ number} = \ell; \kappa\right)}{\mathbb{P}\left(\text{number} = \ell; \kappa\right)} \\ &= \begin{cases} \frac{\mathbb{P}(\text{sequence} = j; \kappa)}{\mathbb{P}(\text{number} = \ell; \kappa)} = \frac{1}{\binom{N}{\ell}\ell!} & \text{if } \ell = |j| \\ 0 & \text{otherwise,} \end{cases} \end{aligned}$$

i.e., the conditional distribution of the sequence of adopters (the data), given the number of adopters (the statistic), does not depend on the parameter κ. In fact, the

number of adopters is a minimal sufficient statistic:

Proposition 6. *In the case where $\kappa_i = \kappa_{ij} = \kappa > 0$ for all agents i, j, the number of adopters is a minimal sufficient statistic for the sequence of adopters.*

Proof. We need to show that

$$\frac{\mathbb{P}\left(\text{sequence} = x\right)}{\mathbb{P}\left(\text{sequence} = y\right)} \text{ is independent of } \kappa \iff |x| = |y|.$$

$\Longleftarrow$: If $|x| = |y|$, then by Equation (6.2),

$$\frac{\mathbb{P}\left(\text{sequence} = x\right)}{\mathbb{P}\left(\text{sequence} = y\right)} = 1.$$

$\Longrightarrow$: By Equation (6.2), if, without loss of generality, $|x| > |y|$, then

$$\frac{\mathbb{P}\left(\text{sequence} = x\right)}{\mathbb{P}\left(\text{sequence} = y\right)} = \frac{|y|+1}{(|y|+1)(n-|y|)\kappa+1} \cdots \frac{|x|}{(|x|)(n-(|x|-1))\kappa+1} \cdot \frac{(|y|+1)(n-|y|)\kappa+1}{(|x|+1)(n-|x|)\kappa+1},$$

which depends on κ. □

6.1.2 Comparing Learning with Sets and Learning with Sequences in an Example Network

We showcase the comparison between $\mathcal{J}^{\text{set}}$ and $\mathcal{J}^{\text{sequence}}$ for the simple network of Figure 6-1, where agent 3 does not adopt unless he is influenced by agent 1 or 2 or both.

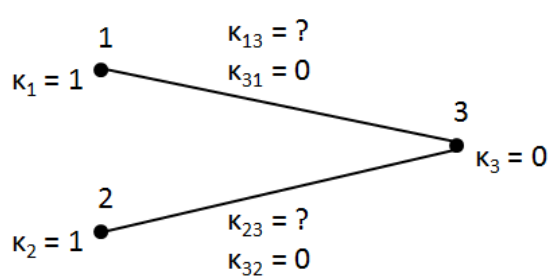


Figure 6-1: A simple network of influence. No edge between agents 1 and 2 means that 1 does not influence 2, and vice versa.

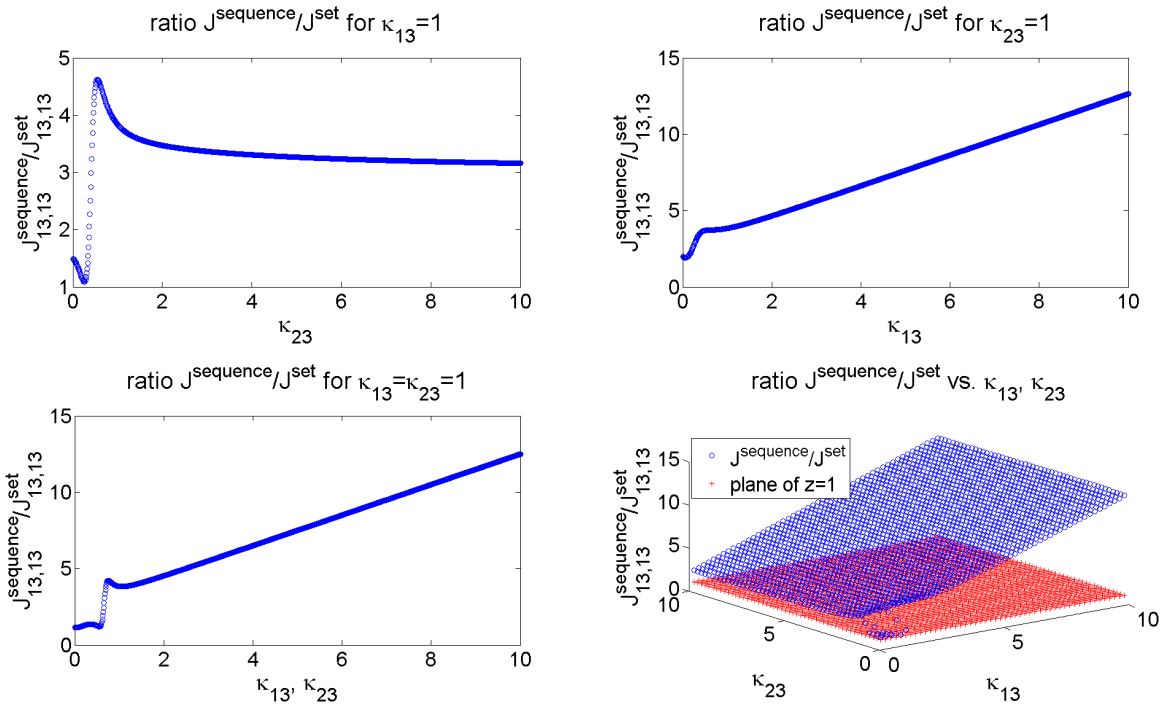


Figure 6-2: Scaling of $\frac{\mathcal{J}^{\text{sequence}}_{13,13}}{\mathcal{J}^{\text{set}}_{13,13}}$ with κ_{13}, κ_{23}.

We note that we have defined $\kappa_i = \frac{\lambda_i}{\lambda_{hor}}, \kappa_{ij} = \frac{\lambda_{ij}}{\lambda_{hor}}$. We assume $\kappa_1 = \kappa_2 = 1, \kappa_3 = 0, \kappa_{31} = \kappa_{32} = \kappa_{12} = \kappa_{21} = 0$, and the parameters to be estimated are κ_{13}, κ_{23}. We are interested in how the ratio $\frac{\mathcal{J}^{\text{sequence}}_{13,13}(\boldsymbol{\kappa})}{\mathcal{J}^{\text{set}}_{13,13}(\boldsymbol{\kappa})}$ scales with κ_{13}, κ_{23}.

The ratio of interest is strictly greater than 1 for $\kappa_{13}, \kappa_{23} > 0$, and satisfies

$$\begin{aligned}
\lim_{\kappa_{23} \longrightarrow \infty} \frac{\mathcal{J}^{\text{sequence}}_{13,13}(\boldsymbol{\kappa})}{\mathcal{J}^{\text{set}}_{13,13}(\boldsymbol{\kappa})} &< \infty \qquad \text{for fixed } \kappa_{13}, \\
\lim_{\kappa_{13} \longrightarrow \infty} \frac{\mathcal{J}^{\text{sequence}}_{13,13}(\boldsymbol{\kappa})}{\mathcal{J}^{\text{set}}_{13,13}(\boldsymbol{\kappa})} &= \infty \qquad \text{for fixed } \kappa_{23}.
\end{aligned}$$

Figure 6-2 illustrates. The interpretation is that as the rate κ_{13} grows, the gain in learning κ_{13} with sequences of adoptions over learning it with sets of adoptions becomes arbitrarily large, because sequences contain increasingly more information about the rate κ_{13} than mere sets.

6.2 Characterization of the Speed of Learning Using the Kullback-Leibler Divergence

We propose three binary hypothesis testing problems:

1. Which of two peers influences you crucially?

2. Are you influenced by your peer, or do you act independently?

3. Does your peer influence you a lot or a little?

In the context of binary hypothesis testing in the Neyman-Pearson setting, the Chernoff-Stein lemma yields the asymptotically optimal exponent for the probability of error of one type, under the constraint that the probability of error of the other type is less than ϵ. More specifically, given hypotheses H_1 and H_2, and corresponding probability measures P_1, P_2, the best achievable exponent for the probability of error of deciding in favor of H_1 when H_2 is true, given that the probability of deciding in favor of H_2 when H_1 is true is less than ϵ, is given by the negative Kullback-Leibler divergence between the two measures $-D(P_1||P_2) \equiv -\mathbb{E}_{P_1}\left[\log \frac{dP_1}{dP_2}\right]$, where $\frac{dP_1}{dP_2}$ denotes the Radon-Nikodym derivative of the two measures (see, for example, Cover and Thomas, 2006).

For each hypothesis testing problem, we observe i.i.d. observations drawn from the true distribution; the observations can be a collection of sets of adopters, a collection of sequences of adopters, or a collection of times of adoptions, depending on how much information is provided in the available data. We use $KL_{\text{set}}, KL_{\text{sequence}}, KL_{\text{time}}$ to denote the Kullback-Leibler divergence of the two distributions pertaining to the two hypotheses under the cases of learning with data which only provides the sets of adopters, learning with data which provides the sequence of adopters but not exact times, and learning with data which provides exact times of adoptions, respectively[5]. A greater Kullback-Leibler divergence implies a faster decaying probability of error, which in turn means that fewer i.i.d. observations are needed in order for the

[5]Abrahao, Chierichetti, Kleinberg, and Panconesi (2013) also make use of the Kullback-Leibler divergence, in order to quantify the sample complexity when time stamps are disregarded.

probability of error to become sufficiently small. We are interested in the relation between the Kullback-Leibler divergences for the cases of sets, sequences, and times of adoptions;[6] this relation reveals how faster learning becomes asymptotically with temporally richer data.

We show that:

1. When the data for each independent observation is collected over a small horizon, the sets of decisions provide almost all the necessary information for learning, and there is no value in richer temporal data for moderate values of the influence parameter.

2. When the data for each independent observation is collected over a large horizon, then sequences have a large gain over sets, and times have smaller gain over sequences for moderate values of the influence parameter, for the first and second hypothesis testing problems; for the third problem, in which the two hypotheses are asymmetric with respect to the rate with which the influenced agent adopts, sequences have no gain over sets, while times have a large gain even for small values of the influence parameter.

3. When the data for each independent observation is collected over a moderate horizon, times have some gain over sequences and sets for moderate values of the influence parameter.

6.2.1 Which of Two Peers Influences You Crucially?

We consider the hypothesis testing problem illustrated in Figure 6-3. In words, we fix some known α. According to Hypothesis I, agent 1 influences agent 3 with rate α, while agent 2 influences agent 3 with rate 1; according to Hypothesis II, agent 1 influences agent 3 with rate 1, while agent 2 influences agent 3 with rate α.

[6] It is clear that $KL_{\text{set}} \leq KL_{\text{sequence}} \leq KL_{\text{time}}$; we are interested in the relative scaling of the KL divergences and how it changes with the parameters.

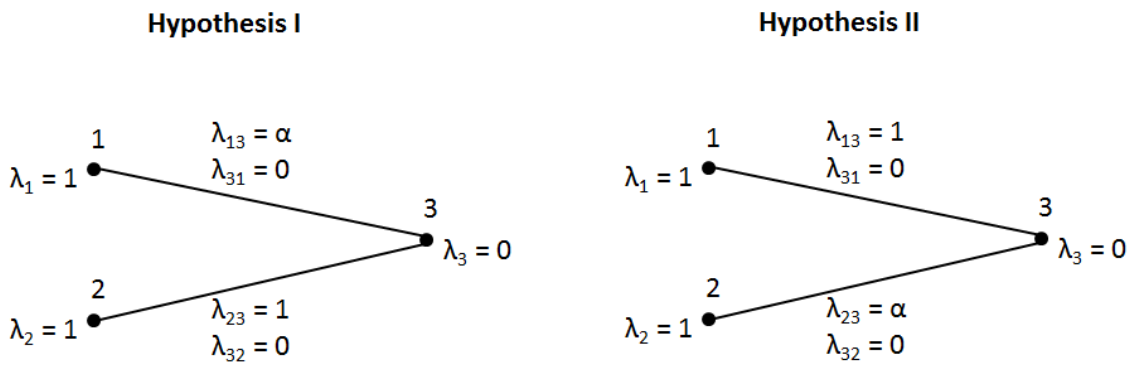


Figure 6-3: The hypothesis testing problem: which of agents 1, 2 crucially influences agent 3?

The probability mass functions needed for the calculation of $KL_{\text{set}}, KL_{\text{sequence}}$ are straightforward to compute. For example, the probability of the *sequence* of agents $\{1, 2\}$ occurring is

$$\frac{\lambda_1}{\lambda_1 + \lambda_2 + \lambda_{hor}} \cdot \frac{\lambda_2}{\lambda_2 + \lambda_{13} + \lambda_{hor}} \cdot \frac{\lambda_{hor}}{\lambda_{13} + \lambda_{23} + \lambda_{hor}},$$

the probability of the *sequence* of agents $\{2, 1\}$ is

$$\frac{\lambda_2}{\lambda_1 + \lambda_2 + \lambda_{hor}} \cdot \frac{\lambda_1}{\lambda_1 + \lambda_{23} + \lambda_{hor}} \cdot \frac{\lambda_{hor}}{\lambda_{13} + \lambda_{23} + \lambda_{hor}},$$

and the probability of the *set* of agents $\{1, 2\}$ is the sum of the two. Denoting with p^i_A the probability of a sequence of adopters given by ordered set A according to Hypothesis i, we can write

$$\begin{aligned} KL_{\text{set}} \;=\; & p^I_\emptyset \log \frac{p^I_\emptyset}{p^{II}_\emptyset} + p^I_{\{1\}} \log \frac{p^I_{\{1\}}}{p^{II}_{\{1\}}} + p^I_{\{2\}} \log \frac{p^I_{\{2\}}}{p^{II}_{\{2\}}} \\ & + \left(p^I_{\{1,2\}} + p^I_{\{2,1\}}\right) \log \frac{p^I_{\{1,2\}} + p^I_{\{2,1\}}}{p^{II}_{\{1,2\}} + p^{II}_{\{2,1\}}} + p^I_{\{1,3\}} \log \frac{p^I_{\{1,3\}}}{p^{II}_{\{1,3\}}} + p^I_{\{2,3\}} \log \frac{p^I_{\{2,3\}}}{p^{II}_{\{2,3\}}} \\ & + \left(p^I_{\{1,2,3\}} + p^I_{\{2,1,3\}} + p^I_{\{1,3,2\}} + p^I_{\{2,3,1\}}\right) \log \frac{p^I_{\{1,2,3\}} + p^I_{\{2,1,3\}} + p^I_{\{1,3,2\}} + p^I_{\{2,3,1\}}}{p^{II}_{\{1,2,3\}} + p^{II}_{\{2,1,3\}} + p^{II}_{\{1,3,2\}} + p^{II}_{\{2,3,1\}}} \end{aligned}$$

and

$$
\begin{aligned}
KL_{\text{sequence}} &= p^I_{\emptyset}\log\frac{p^I_{\emptyset}}{p^{II}_{\emptyset}} + p^I_{\{1\}}\log\frac{p^I_{\{1\}}}{p^{II}_{\{1\}}} + p^I_{\{2\}}\log\frac{p^I_{\{2\}}}{p^{II}_{\{2\}} \\
&+p^I_{\{1,2\}}\log\frac{p^I_{\{1,2\}}}{p^{II}_{\{1,2\}}} + p^I_{\{2,1\}}\log\frac{p^I_{\{2,1\}}}{p^{II}_{\{2,1\}}} + p^I_{\{1,3\}}\log\frac{p^I_{\{1,3\}}}{p^{II}_{\{1,3\}}} + p^I_{\{2,3\}}\log\frac{p^I_{\{2,3\}}}{p^{II}_{\{2,3\}} \\
&+p^I_{\{1,2,3\}}\log\frac{p^I_{\{1,2,3\}}}{p^{II}_{\{1,2,3\}}} + p^I_{\{2,1,3\}}\log\frac{p^I_{\{2,1,3\}}}{p^{II}_{\{2,1,3\}}} + p^I_{\{1,3,2\}}\log\frac{p^I_{\{1,3,2\}}}{p^{II}_{\{1,3,2\}}} + p^I_{\{2,3,1\}}\log\frac{p^I_{\{2,3,1\}}}{p^{II}_{\{2,3,1\}}}.
\end{aligned}
$$

From the log-sum inequality $\sum_{i=1}^n a_i \log\frac{a_i}{b_i} \geq (\sum_{i=1}^n a_i)\log\frac{\sum_{i=1}^n a_i}{\sum_{i=1}^n b_i}$ for nonnegative numbers, $a_1, a_2, \ldots, a_n$ and $b_1, b_2, \ldots, b_n$, it is clear that $KL_{\text{set}} \leq KL_{\text{sequence}}$.

For the probability density functions[7] needed for the calculation of KL_{time}, we consider T_1, T_2, T_{13}, T_{23}, T_{hor}, the exponentially distributed random variables modeling the time of adoption by agent 1, the time of adoption by agent 2, the time of adoption by agent 3 due to agent 1, the time of adoption by agent 3 due to agent 2, and the end of the horizon, respectively. Then $T_3 = \min(T_{13}, T_{23})$ models the time of adoption by agent 3 (due to either agent 1 or agent 2). One can then consider the joint density of $T_1, T_2, T_{13}, T_{23}, T_{hor}$. Nevertheless, the available data only captures the realizations of random variables that occurred before the realization of T_{hor}. Therefore, the calculation of the joint densities should account only for observed outcomes.

In particular, we write down the densities relating to each possible sequence of adopters:

$$
\begin{aligned}
f_{\emptyset}(t_{hor}) &= e^{-\lambda_1 t_{hor}} e^{-\lambda_2 t_{hor}} \lambda_{hor} e^{-\lambda_{hor} t_{hor}} \\
f_{\{1\}}(t_1, t_{hor}) &= \lambda_1 e^{-\lambda_1 t_1} e^{-\lambda_2 t_{hor}} e^{-(\lambda_{13}(t_{hor}-t_1))} \lambda_{hor} e^{-\lambda_{hor} t_{hor}} \\
f_{\{2\}}(t_2, t_{hor}) &= \lambda_2 e^{-\lambda_2 t_2} e^{-\lambda_1 t_{hor}} e^{-(\lambda_{23}(t_{hor}-t_2))} \lambda_{hor} e^{-\lambda_{hor} t_{hor}} \\
f_{\{1,2\}}(t_1, t_2, t_{hor}) &= \lambda_1 e^{-\lambda_1 t_1} \lambda_2 e^{-\lambda_2 t_2} e^{-\lambda_{13}(t_2-t_1)} e^{-(\lambda_{13}+\lambda_{23})(t_{hor}-t_2)} \lambda_{hor} e^{-\lambda_{hor} t_{hor}} \\
f_{\{2,1\}}(t_2, t_1, t_{hor}) &= \lambda_1 e^{-\lambda_1 t_1} \lambda_2 e^{-\lambda_2 t_2} e^{-\lambda_{23}(t_1-t_2)} e^{-(\lambda_{13}+\lambda_{23})(t_{hor}-t_1)} \lambda_{hor} e^{-\lambda_{hor} t_{hor}} \\
f_{\{1,3\}}(t_1, t_3, t_{hor}) &= \lambda_1 e^{-\lambda_1 t_1} e^{-\lambda_2 t_{hor}} \lambda_{13} e^{-(\lambda_{13}(t_3-t_1))} \lambda_{hor} e^{-\lambda_{hor} t_{hor}}
\end{aligned}
$$

[7] Although these likelihood functions are not densities in the strict sense, we abuse terminology and use the term "density functions" to refer to them.

$$
\begin{aligned}
f_{\{2,3\}}(t_2,t_3,t_{hor}) &= \lambda_2 e^{-\lambda_2 t_2} e^{-\lambda_1 t_{hor}} \lambda_{23} e^{-(\lambda_{23}(t_3-t_2))} \lambda_{hor} e^{-\lambda_{hor} t_{hor}} \\
f_{\{1,2,3\}}(t_1,t_2,t_3,t_{hor}) &= \lambda_1 e^{-\lambda_1 t_1} \lambda_2 e^{-\lambda_2 t_2} e^{-\lambda_{13}(t_2-t_1)} (\lambda_{13}+\lambda_{23}) e^{-(\lambda_{13}+\lambda_{23})(t_3-t_2)} \lambda_{hor} e^{-\lambda_{hor} t_{hor}} \\
f_{\{2,1,3\}}(t_2,t_1,t_3,t_{hor}) &= \lambda_1 e^{-\lambda_1 t_1} \lambda_2 e^{-\lambda_2 t_2} e^{-\lambda_{23}(t_1-t_2)} (\lambda_{13}+\lambda_{23}) e^{-(\lambda_{13}+\lambda_{23})(t_3-t_1)} \lambda_{hor} e^{-\lambda_{hor} t_{hor}} \\
f_{\{1,3,2\}}(t_1,t_3,t_2,t_{hor}) &= \lambda_1 e^{-\lambda_1 t_1} \lambda_2 e^{-\lambda_2 t_2} \lambda_{13} e^{-\lambda_{13}(t_3-t_1)} \lambda_{hor} e^{-\lambda_{hor} t_{hor}} \\
f_{\{2,3,1\}}(t_2,t_3,t_1,t_{hor}) &= \lambda_1 e^{-\lambda_1 t_1} \lambda_2 e^{-\lambda_2 t_2} \lambda_{23} e^{-\lambda_{23}(t_3-t_2)} \lambda_{hor} e^{-\lambda_{hor} t_{hor}}
\end{aligned}
$$

where t_1, t_2, t_3, t_{hor} denote the realized times of adoption by agents 1, 2, 3, and the realized time of the end of the horizon, respectively. Denoting with f_A^i the density relating to ordered set A under Hypothesis i, we have

$$
\begin{aligned}
KL_{\text{time}} = & \int_0^\infty f_\emptyset^I(t_{hor}) \log \frac{f_\emptyset^I(t_{hor})}{f_\emptyset^{II}(t_{hor})} dt_{hor} \\
& + \int_{t_1=0}^\infty \int_{t_{hor}=t_1}^\infty f_{\{1\}}^I(t_1,t_{hor}) \log \frac{f_{\{1\}}^I(t_1,t_{hor})}{f_{\{1\}}^{II}(t_1,t_{hor})} dt_{hor} dt_1 \\
& + \int_{t_2=0}^\infty \int_{t_{hor}=t_2}^\infty f_{\{2\}}^I(t_2,t_{hor}) \log \frac{f_{\{2\}}^I(t_2,t_{hor})}{f_{\{2\}}^{II}(t_2,t_{hor})} dt_{hor} dt_2 \\
& + \int_{t_1=0}^\infty \int_{t_2=t_1}^\infty \int_{t_{hor}=t_2}^\infty f_{\{1,2\}}^I(t_1,t_2,t_{hor}) \log \frac{f_{\{1,2\}}^I(t_1,t_2,t_{hor})}{f_{\{1,2\}}^{II}(t_1,t_2,t_{hor})} dt_{hor} dt_2 dt_1 \\
& + \int_{t_2=0}^\infty \int_{t_1=t_2}^\infty \int_{t_{hor}=t_1}^\infty f_{\{2,1\}}^I(t_2,t_1,t_{hor}) \log \frac{f_{\{2,1\}}^I(t_2,t_1,t_{hor})}{f_{\{2,1\}}^{II}(t_2,t_1,t_{hor})} dt_{hor} dt_1 dt_2 \\
& + \int_{t_1=0}^\infty \int_{t_3=t_1}^\infty \int_{t_{hor}=t_3}^\infty f_{\{1,3\}}^I(t_1,t_3,t_{hor}) \log \frac{f_{\{1,3\}}^I(t_1,t_3,t_{hor})}{f_{\{1,3\}}^{II}(t_1,t_3,t_{hor})} dt_{hor} dt_3 dt_1 \\
& + \int_{t_2=0}^\infty \int_{t_3=t_2}^\infty \int_{t_{hor}=t_3}^\infty f_{\{2,3\}}^I(t_2,t_3,t_{hor}) \log \frac{f_{\{2,3\}}^I(t_2,t_3,t_{hor})}{f_{\{2,3\}}^{II}(t_2,t_3,t_{hor})} dt_{hor} dt_3 dt_2 \\
& + \int_{t_1=0}^\infty \int_{t_2=t_1}^\infty \int_{t_3=t_2}^\infty \int_{t_{hor}=t_3}^\infty f_{\{1,2,3\}}^I(t_1,t_2,t_3,t_{hor}) \log \frac{f_{\{1,2,3\}}^I(t_1,t_2,t_3,t_{hor})}{f_{\{1,2,3\}}^{II}(t_1,t_2,t_3,t_{hor})} dt_{hor} dt_3 dt_2 dt_1 \\
& + \int_{t_2=0}^\infty \int_{t_1=t_2}^\infty \int_{t_3=t_1}^\infty \int_{t_{hor}=t_3}^\infty f_{\{2,1,3\}}^I(t_2,t_1,t_3,t_{hor}) \log \frac{f_{\{2,1,3\}}^I(t_2,t_1,t_3,t_{hor})}{f_{\{2,1,3\}}^{II}(t_2,t_1,t_3,t_{hor})} dt_{hor} dt_3 dt_1 dt_2 \\
& + \int_{t_1=0}^\infty \int_{t_3=t_1}^\infty \int_{t_2=t_3}^\infty \int_{t_{hor}=t_2}^\infty f_{\{1,3,2\}}^I(t_1,t_3,t_2,t_{hor}) \log \frac{f_{\{1,3,2\}}^I(t_1,t_3,t_2,t_{hor})}{f_{\{1,3,2\}}^{II}(t_1,t_3,t_2,t_{hor})} dt_{hor} dt_2 dt_3 dt_1 \\
& + \int_{t_2=0}^\infty \int_{t_3=t_2}^\infty \int_{t_1=t_3}^\infty \int_{t_{hor}=t_1}^\infty f_{\{2,3,1\}}^I(t_2,t_3,t_1,t_{hor}) \log \frac{f_{\{2,3,1\}}^I(t_2,t_3,t_1,t_{hor})}{f_{\{2,3,1\}}^{II}(t_2,t_3,t_1,t_{hor})} dt_{hor} dt_1 dt_3 dt_2.
\end{aligned}
$$

In Figures 6-4 and 6-5 we plot the KL divergences and their ratios against influence rate α, for different horizon rates.

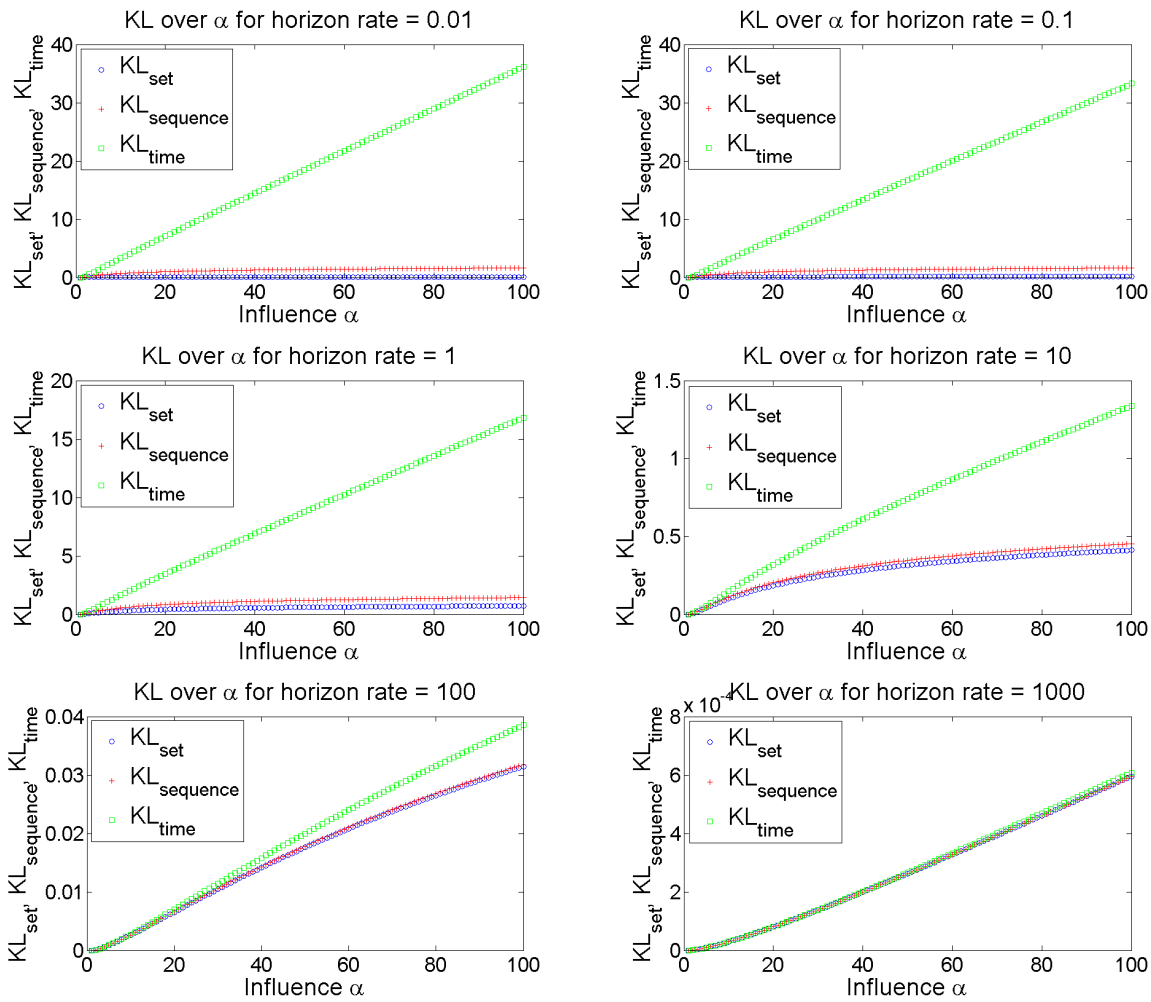


Figure 6-4: Which of two peers influences you? Plots of KL_{set} (circles), KL_{sequence} (crosses), KL_{time} (squares) against influence rate α for different horizon rates.

In the large horizon regime (i.e., when λ_{hor} is small), knowing the sequences of adoptions has a large gain over knowing the sets of adoptions. On the other hand, knowing the times of adoptions has large gain over knowing the sequences of adoptions only for large enough values of the influence rate α. For small values of α, the gain of times over sequences is small compared to the gain of sequences over sets.

In the small horizon regime (i.e., when λ_{hor} is large), the sets of adoptions give almost all the information for learning, and there is no much further value in richer

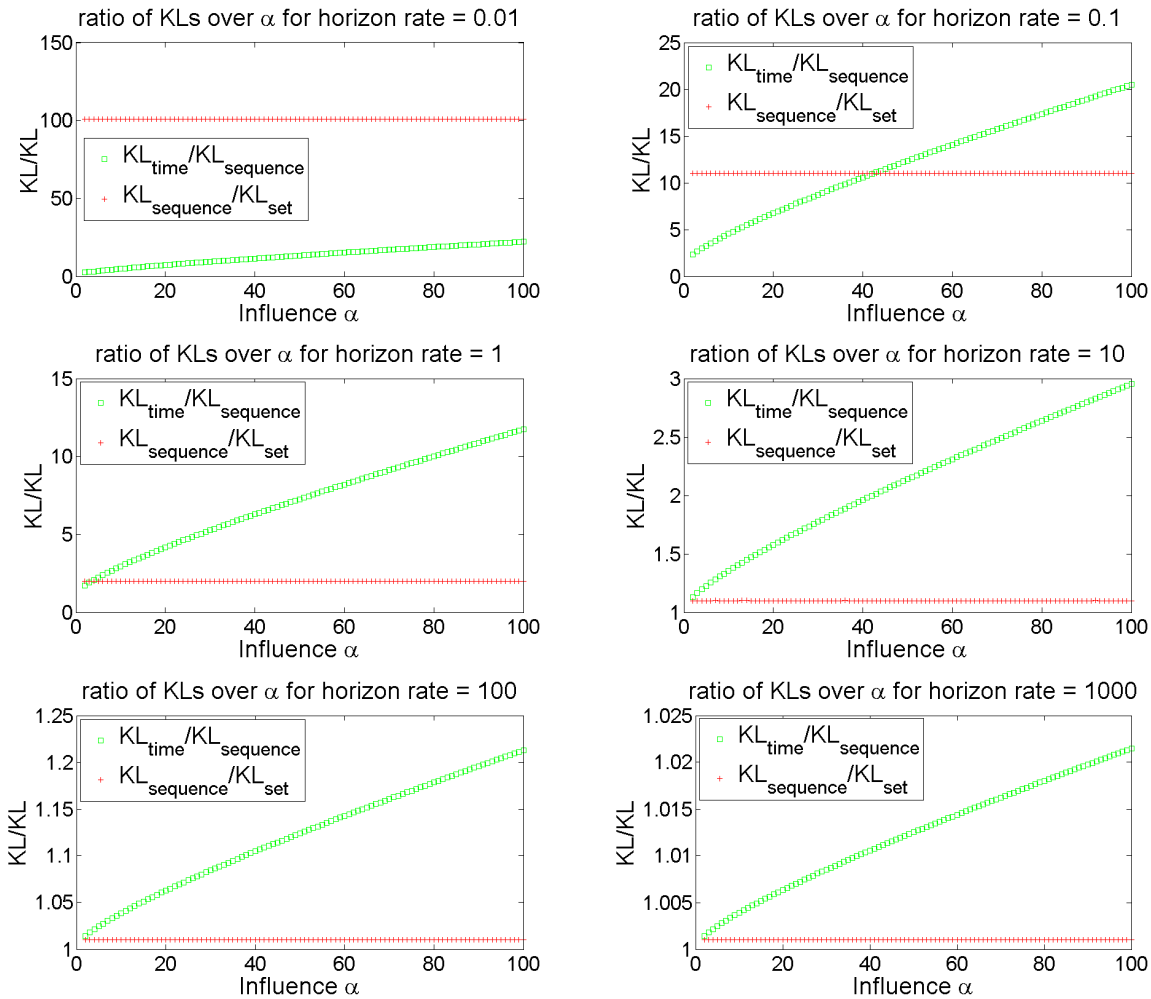


Figure 6-5: Which of two peers influences you? Plots of $KL_{\text{sequence}}/KL_{\text{set}}$ (crosses), $KL_{\text{time}}/KL_{\text{sequence}}$ (squares) against influence rate α for different horizon rates.

temporal data. Sequences have no gain (in the limit as $\lambda_{hor} \longrightarrow \infty$) over sets, while for times to have significant gain over sequences, the rate of influence α has to be large.

In the moderate horizon regime, knowing the times of adoptions has some value over knowing merely sequences or sets of adoptions even for small values of the influence rate α.

6.2.2 Are You Influenced by Your Peer, or Do You Act Independently?

We consider the hypothesis testing problem illustrated in Figure 6-6. In words, according to Hypothesis I, agent 1 influences agent 2 with rate α, while agent 2 adopts herself with rate 1; according to Hypothesis II, agent 1 influences agent 2 with rate 1, while agent 2 adopts herself with rate α.

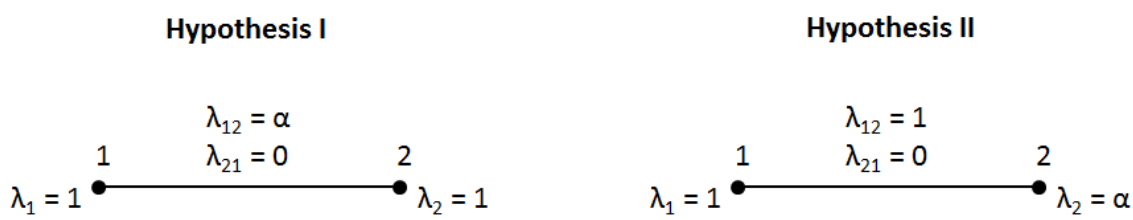


Figure 6-6: The hypothesis testing problem: is agent 2 influenced by agent 1, or does she have a high individual rate?

The probability mass functions needed for the calculation of $KL_{\text{set}}, KL_{\text{sequence}}$ are straightforward to write down. Denoting with p_A^i the probability of a sequence of adopters given by ordered set A according to Hypothesis i, we can write

$$\begin{aligned} KL_{\text{set}} &= p_\emptyset^I \log \frac{p_\emptyset^I}{p_\emptyset^{II}} + p_{\{1\}}^I \log \frac{p_{\{1\}}^I}{p_{\{1\}}^{II}} + p_{\{2\}}^I \log \frac{p_{\{2\}}^I}{p_{\{2\}}^{II}} \\ &\quad + \left(p_{\{1,2\}}^I + p_{\{2,1\}}^I\right) \log \frac{p_{\{1,2\}}^I + p_{\{2,1\}}^I}{p_{\{1,2\}}^{II} + p_{\{2,1\}}^{II}} \end{aligned}$$

and

$$\begin{aligned} KL_{\text{sequence}} &= p_\emptyset^I \log \frac{p_\emptyset^I}{p_\emptyset^{II}} + p_{\{1\}}^I \log \frac{p_{\{1\}}^I}{p_{\{1\}}^{II}} + p_{\{2\}}^I \log \frac{p_{\{2\}}^I}{p_{\{2\}}^{II}} \\ &\quad + p_{\{1,2\}}^I \log \frac{p_{\{1,2\}}^I}{p_{\{1,2\}}^{II}} + p_{\{2,1\}}^I \log \frac{p_{\{2,1\}}^I}{p_{\{2,1\}}^{II}}. \end{aligned}$$

For the calculation of KL_{time}, we write down the probability density function

relating to each possible sequence of adopters:

$$
\begin{aligned}
f_{\emptyset} &= e^{-\lambda_1 t_{hor}} e^{-\lambda_2 t_{hor}} \lambda_{hor} e^{-\lambda_{hor} t_{hor}} \\
f_{\{1\}} &= \lambda_1 e^{-\lambda_1 t_1} e^{-\lambda_2 t_{hor}} e^{-(\lambda_{12}(t_{hor}-t_1))} \lambda_{hor} e^{-\lambda_{hor} t_{hor}} \\
f_{\{2\}} &= \lambda_2 e^{-\lambda_2 t_2} e^{-\lambda_1 t_{hor}} \lambda_{hor} e^{-\lambda_{hor} t_{hor}} \\
f_{\{1,2\}} &= \lambda_1 e^{-\lambda_1 t_1} e^{-\lambda_2 t_1} (\lambda_2 + \lambda_{12}) e^{-(\lambda_2+\lambda_{12})(t_2-t_1)} \lambda_{hor} e^{-\lambda_{hor} t_{hor}} \\
f_{\{2,1\}} &= \lambda_1 e^{-\lambda_1 t_1} \lambda_2 e^{-\lambda_2 t_2} \lambda_{hor} e^{-\lambda_{hor} t_{hor}},
\end{aligned}
$$

where t_1, t_2, t_{hor} denote the realized times of adoption by agents 1, 2, and the realized time of the end of the horizon, respectively. Denoting with f_A^i the density relating to ordered set A under Hypothesis i, we have

$$
\begin{aligned}
KL_{\text{time}} &= \int_0^\infty f_{\emptyset}^I(t_{hor}) \log \frac{f_{\emptyset}^I(t_{hor})}{f_{\emptyset}^{II}(t_{hor})} dt_{hor} \\
&+ \int_{t_1=0}^\infty \int_{t_{hor}=t_1}^\infty f_{\{1\}}^I(t_1, t_{hor}) \log \frac{f_{\{1\}}^I(t_1, t_{hor})}{f_{\{1\}}^{II}(t_1, t_{hor})} dt_{hor} dt_1 \\
&+ \int_{t_2=0}^\infty \int_{t_{hor}=t_2}^\infty f_{\{2\}}^I(t_2, t_{hor}) \log \frac{f_{\{2\}}^I(t_2, t_{hor})}{f_{\{2\}}^{II}(t_2, t_{hor})} dt_{hor} dt_2 \\
&+ \int_{t_1=0}^\infty \int_{t_2=t_1}^\infty \int_{t_{hor}=t_2}^\infty f_{\{1,2\}}^I(t_1, t_2, t_{hor}) \log \frac{f_{\{1,2\}}^I(t_1, t_2, t_{hor})}{f_{\{1,2\}}^{II}(t_1, t_2, t_{hor})} dt_{hor} dt_2 dt_1 \\
&+ \int_{t_2=0}^\infty \int_{t_1=t_2}^\infty \int_{t_{hor}=t_1}^\infty f_{\{2,1\}}^I(t_2, t_1, t_{hor}) \log \frac{f_{\{2,1\}}^I(t_2, t_1, t_{hor})}{f_{\{2,1\}}^{II}(t_2, t_1, t_{hor})} dt_{hor} dt_1 dt_2.
\end{aligned}
$$

In Figures 6-7 and 6-8 we plot the KL divergences and their ratios against influence rate α, for different horizon rates.

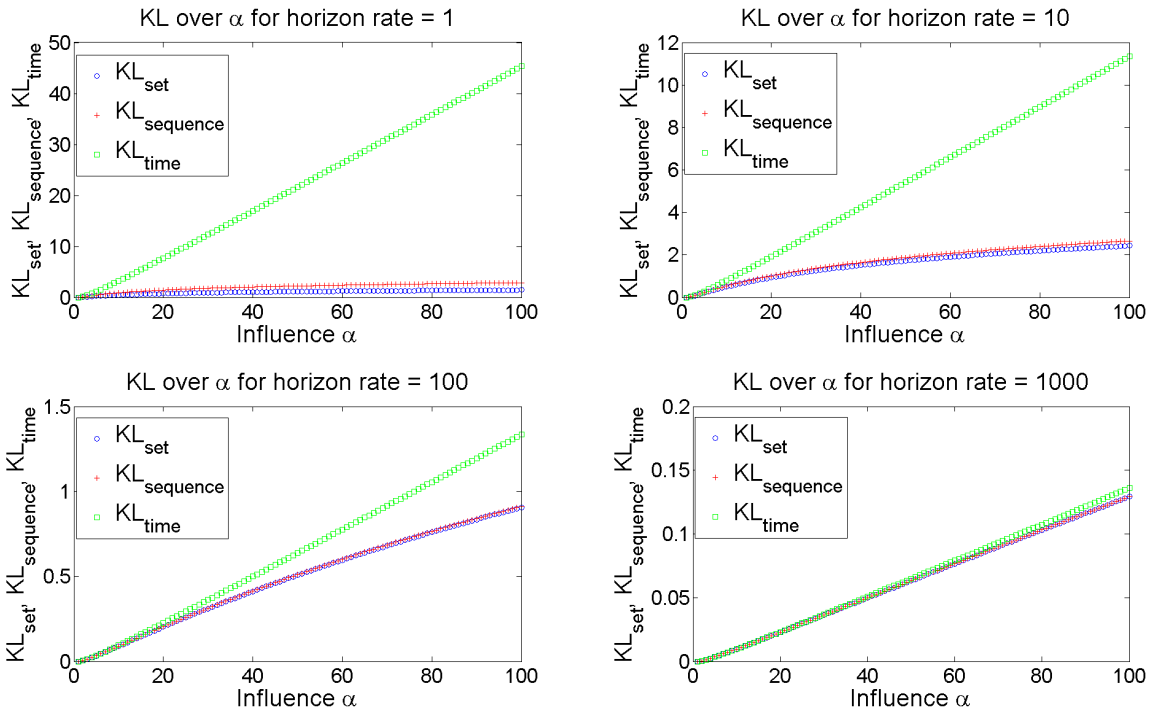


Figure 6-7: Are you influenced by your peer, or do you act independently? Plots of KL_{set} (circles), KL_{sequence} (crosses), KL_{time} (squares) against influence rate α for different horizon rates.

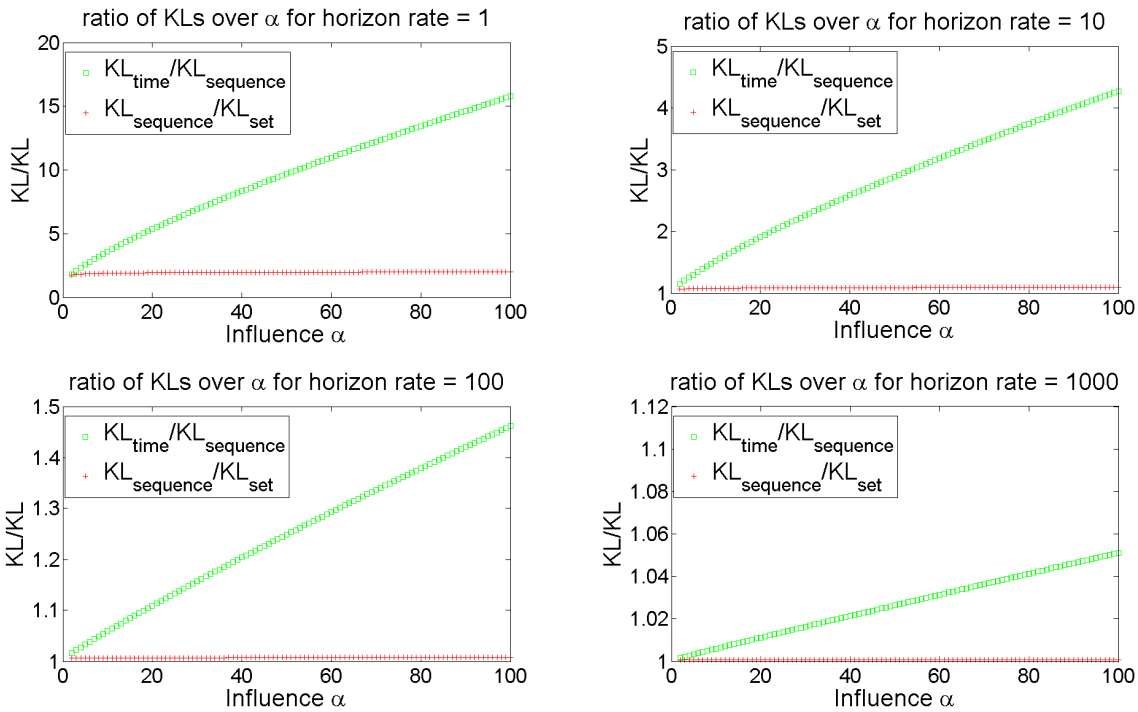


Figure 6-8: Are you influenced by your peer, or do you act independently? Plots of $KL_{\text{sequence}}/KL_{\text{set}}$ (crosses), $KL_{\text{time}}/KL_{\text{sequence}}$ (squares) against influence rate α for different horizon rates.

The qualitative analysis is similar to the analysis of the hypothesis testing problem of which of two peers influences crucially another agent. In the large horizon regime (i.e., when λ_{hor} is small), knowing the sequences of adoptions has a large gain over knowing the sets of adoptions, and knowing the times of adoptions yields much smaller gain over knowing just sequences for small values of the rate α. In the small horizon regime (i.e., when λ_{hor} is large), the sets of adoptions give almost all the information for learning, and there is no much further value in richer temporal data. In the moderate horizon regime, knowing the times of adoptions has some value over knowing merely sequences or sets of adoptions even for small values of the rate α.

The value of learning with sequences of adoptions over learning with sets of adoptions can be readily decided analytically by looking at the relevant limits. First we

write

$$
\begin{aligned}
KL_{\text{set}}(\alpha, \lambda_{hor}) &= \frac{\lambda_{hor}}{2+\lambda_{hor}} \log \frac{1+\alpha+\lambda_{hor}}{2+\lambda_{hor}} \\
&+ \frac{1}{2+\lambda_{hor}} \frac{\lambda_{hor}}{1+\alpha+\lambda_{hor}} \log \frac{1+\alpha+\lambda_{hor}}{2+\lambda_{hor}} + \frac{1}{2+\lambda_{hor}} \frac{\lambda_{hor}}{1+\lambda_{hor}} \log \frac{1+\alpha+\lambda_{hor}}{\alpha(2+\lambda_{hor})} \\
&+ \left(\frac{1}{2+\lambda_{hor}} \frac{1+\alpha}{1+\alpha+\lambda_{hor}} + \frac{1}{2+\lambda_{hor}} \frac{1}{1+\lambda_{hor}} \right) \\
&\qquad \cdot \log \frac{\frac{1}{2+\lambda_{hor}} \frac{1+\alpha}{1+\alpha+\lambda_{hor}} + \frac{1}{2+\lambda_{hor}} \frac{1}{1+\lambda_{hor}}}{\frac{1}{1+\alpha+\lambda_{hor}} \frac{1+\alpha}{1+\alpha+\lambda_{hor}} + \frac{\alpha}{1+\alpha+\lambda_{hor}} \frac{1}{1+\lambda_{hor}}}
\end{aligned}
$$

and

$$
\begin{aligned}
KL_{\text{sequence}}(\alpha, \lambda_{hor}) &= \frac{\lambda_{hor}}{2+\lambda_{hor}} \log \frac{1+\alpha+\lambda_{hor}}{2+\lambda_{hor}} \\
&+ \frac{1}{2+\lambda_{hor}} \frac{\lambda_{hor}}{1+\alpha+\lambda_{hor}} \log \frac{1+\alpha+\lambda_{hor}}{2+\lambda_{hor}} \\
&+ \frac{1}{2+\lambda_{hor}} \frac{\lambda_{hor}}{1+\lambda_{hor}} \log \frac{1+\alpha+\lambda_{hor}}{\alpha(2+\lambda_{hor})} \\
&+ \frac{1}{2+\lambda_{hor}} \frac{1+\alpha}{1+\alpha+\lambda_{hor}} \log \frac{1+\alpha+\lambda_{hor}}{2+\lambda_{hor}} \\
&+ \frac{1}{2+\lambda_{hor}} \frac{1}{1+\lambda_{hor}} \log \frac{1+\alpha+\lambda_{hor}}{\alpha(2+\lambda_{hor})}.
\end{aligned}
$$

For fixed α, we have

$$
\begin{aligned}
\lim_{\lambda_{hor} \longrightarrow 0} KL_{\text{set}}(\alpha, \lambda_{hor}) &= 0 \\
\lim_{\lambda_{hor} \longrightarrow 0} KL_{\text{sequence}}(\alpha, \lambda_{hor}) &= \frac{1}{2} \log \frac{1+\alpha}{2} + \frac{1}{2} \log \frac{1+\alpha}{2\alpha},
\end{aligned}
$$

which implies

$$
\lim_{\lambda_{hor} \longrightarrow 0} \frac{KL_{\text{sequence}}(\alpha, \lambda_{hor})}{KL_{\text{set}}(\alpha, \lambda_{hor})} = \infty,
$$

for $\alpha \neq 1$, establishing that in the large horizon regime, learning with sequences yields significant gain over learning with sets.

For fixed α, we have[8]

$$\lim_{\lambda_{hor}\longrightarrow\infty} \frac{KL_{\text{sequence}}(\alpha,\lambda_{hor})}{KL_{\text{set}}(\alpha,\lambda_{hor})} = \frac{\frac{\lambda_{hor}}{2+\lambda_{hor}}\log\frac{1+\alpha+\lambda_{hor}}{2+\lambda_{hor}}}{\frac{\lambda_{hor}}{2+\lambda_{hor}}\log\frac{1+\alpha+\lambda_{hor}}{2+\lambda_{hor}}} = 1,$$

which establishes that in the small horizon regime, learning with sequences has insignificant gain over learning with sets.

We can reach the same conclusions by looking at the limit as $\alpha \longrightarrow \infty$, for fixed λ_{hor}. Indeed, for fixed λ_{hor}, we have

$$\lim_{\alpha\longrightarrow\infty} \frac{KL_{\text{sequence}}(\alpha,\lambda_{hor})}{KL_{\text{set}}(\alpha,\lambda_{hor})} = \lim_{\alpha\longrightarrow\infty} \frac{\frac{\lambda_{hor}}{2+\lambda_{hor}}\log\frac{1+\alpha+\lambda_{hor}}{2+\lambda_{hor}} + \frac{1}{2+\lambda_{hor}}\frac{1+\alpha}{1+\alpha+\lambda_{hor}}\log\frac{1+\alpha+\lambda_{hor}}{2+\lambda_{hor}}}{\frac{\lambda_{hor}}{2+\lambda_{hor}}\log\frac{1+\alpha+\lambda_{hor}}{2+\lambda_{hor}}} = \frac{\lambda_{hor}+1}{\lambda_{hor}},$$

which in turn becomes arbitrarily large for $\lambda_{hor} \longrightarrow 0$, and converges to 1 for $\lambda_{hor} \longrightarrow \infty$.

6.2.3 Does Your Peer Influence You a Lot or a Little?

We consider the hypothesis testing problem illustrated in Figure 6-9. In words, according to Hypothesis I, agent 1 influences agent 2 with rate α; according to Hypothesis II, agent 1 influences agent 2 with rate 1.

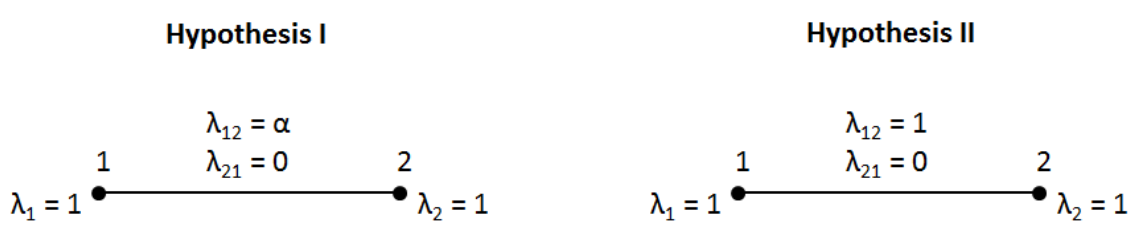


Figure 6-9: The hypothesis testing problem: is agent 2 influenced by agent 1 a lot or a little?

[8] We adopt the definition of the Kullback-Leibler divergence which uses the convention that $0\log\frac{0}{0} = 0$.

The calculation of $KL_{\text{set}}, KL_{\text{sequence}}, KL_{\text{time}}$ is identical to that of the hypothesis testing problem of whether an agent is influenced by her peer, or she has a high individual rate of adoption.

In Figures 6-10 and 6-11 we plot the KL divergences and their ratios against influence rate α, for different horizon rates.

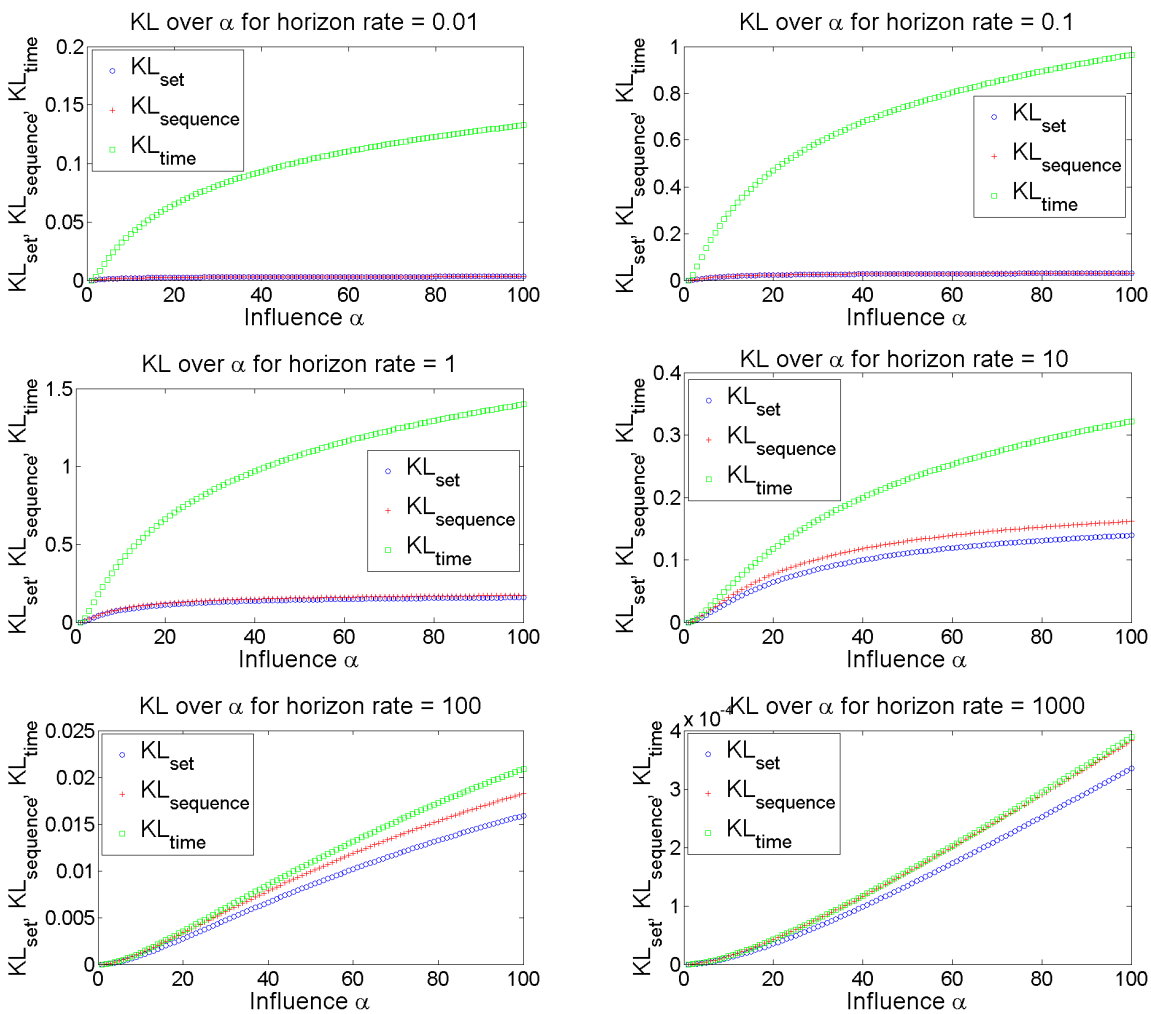


Figure 6-10: Does your peer influence you a lot or a little? Plots of KL_{set} (circles), KL_{sequence} (crosses), KL_{time} (squares) against influence rate α for different horizon rates.

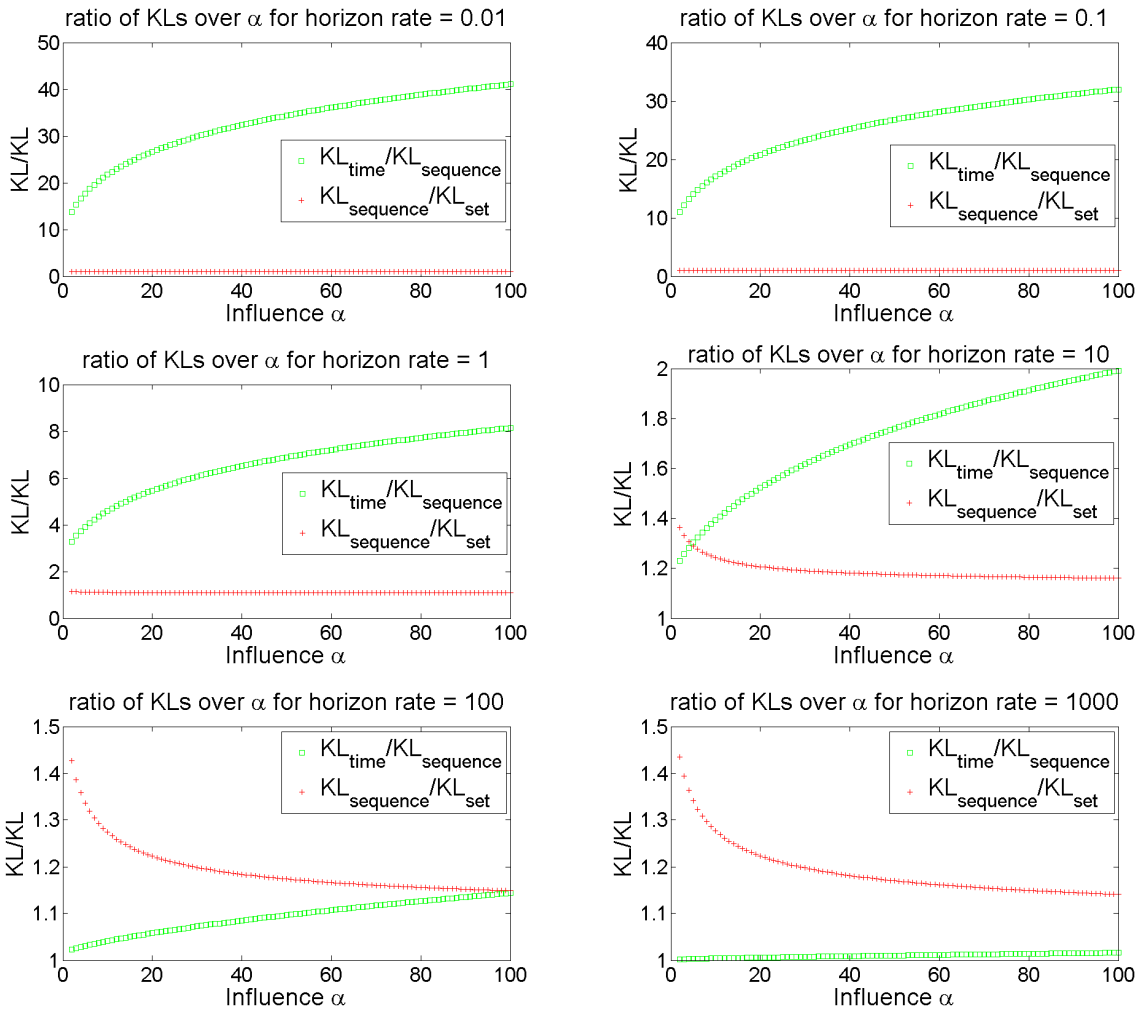


Figure 6-11: Does your peer influence you a lot or a little? Plots of $KL_{\text{sequence}}/KL_{\text{set}}$ (crosses), $KL_{\text{time}}/KL_{\text{sequence}}$ (squares) against influence rate α for different horizon rates.

In the large horizon regime (i.e., when λ_{hor} is small), knowing the times of adoptions has a large gain over knowing the sequences or sets of adoptions. On the other hand, knowing the sequences of adoptions does not have value over knowing just the sets of adoptions.

In the small horizon regime (i.e., when λ_{hor} is large), the sets of adoptions give almost all the information for learning, and there is no much further value in richer temporal data. Sequences have no significant gain over sets, while times have even less gain over sequences for moderate values of the influence rate α.

In the moderate horizon regime, knowing the times of adoptions has some value over knowing merely sequences or sets of adoptions even for small values of the influence rate α. For constant α, the gain becomes larger for larger horizon.

The reason why time has significant gain in the large horizon regime, even for small values of the influence rate, is the difference between the two hypotheses with respect to the total rate with which agent 2 adopts, after agent 1 has adopted (which is $1+\alpha$ under Hypothesis I, and 2 under Hypothesis II). When the horizon is long, having time information allows for more accurate learning of the rate with which agent 2 adopts after agent 1 has adopted, and therefore for better asymptotic optimal error exponent in the hypothesis test.

The value of learning with sequences of adoptions over learning with sets of adoptions can be readily decided analytically by looking at the relevant limits. First we write

$$\begin{aligned} KL_{\text{set}}(\alpha, \lambda_{hor}) &= \frac{1}{2+\lambda_{hor}} \frac{\lambda_{hor}}{1+\alpha+\lambda_{hor}} \log \frac{2+\lambda_{hor}}{1+\alpha+\lambda_{hor}} \\ &\quad + \left(\frac{1}{2+\lambda_{hor}} \frac{1+\alpha}{1+\alpha+\lambda_{hor}} + \frac{1}{2+\lambda_{hor}} \frac{1}{1+\lambda_{hor}} \right) \\ &\qquad \cdot \log \frac{\frac{1+\alpha}{1+\alpha+\lambda_{hor}} + \frac{1}{1+\lambda_{hor}}}{\frac{2}{2+\lambda_{hor}} + \frac{1}{1+\lambda_{hor}}} \end{aligned}$$

and

$$\begin{aligned} KL_{\text{sequence}}(\alpha, \lambda_{hor}) &= \frac{1}{2+\lambda_{hor}} \frac{\lambda_{hor}}{1+\alpha+\lambda_{hor}} \log \frac{2+\lambda_{hor}}{1+\alpha+\lambda_{hor}} \\ &\quad + \frac{1}{2+\lambda_{hor}} \frac{1+\alpha}{1+\alpha+\lambda_{hor}} \log \frac{(1+\alpha)(2+\lambda_{hor})}{2(1+\alpha+\lambda_{hor})}. \end{aligned}$$

For fixed λ_{hor}, we have

$$\begin{aligned} \lim_{\alpha \longrightarrow \infty} \frac{KL_{\text{sequence}}(\alpha, \lambda_{hor})}{KL_{\text{set}}(\alpha, \lambda_{hor})} &= \frac{\frac{1}{2+\lambda_{hor}} \log \frac{2+\lambda_{hor}}{2}}{\frac{1}{1+\lambda_{hor}} \log \frac{1+\frac{1}{1+\lambda_{hor}}}{\frac{2}{2+\lambda_{hor}} + \frac{1}{1+\lambda_{hor}}}} \\ &= \frac{1+\lambda_{hor}}{2+\lambda_{hor}} \frac{\log \frac{2+\lambda_{hor}}{2}}{\log \frac{\frac{2+\lambda_{hor}}{1+\lambda_{hor}}}{\frac{3\lambda_{hor}+4}{(2+\lambda_{hor})(1+\lambda_{hor})}}} \end{aligned}$$

$$= \frac{1+\lambda_{hor}}{2+\lambda_{hor}} \frac{\log \frac{2+\lambda_{hor}}{2}}{\log \frac{(2+\lambda_{hor})^2}{3\lambda_{hor}+4}},$$

which in turn converges to 1 both for $\lambda_{hor} \longrightarrow 0$ and for $\lambda_{hor} \longrightarrow \infty$, using l'Hôpital's rule. In fact, the gain of learning with sequences over learning with sets is insignificant asymptotically as $\alpha \longrightarrow \infty$, for all horizon rates (Figure 6-12).

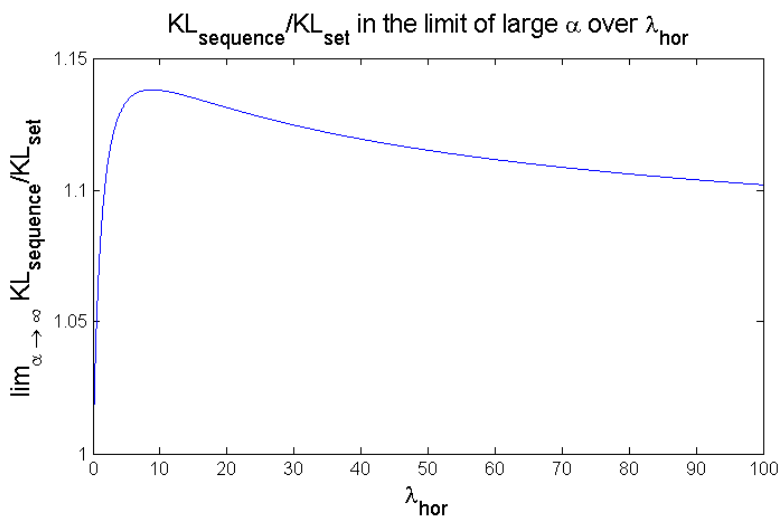


Figure 6-12: Does your peer influence you a lot or a little? $KL_{\text{sequence}}/KL_{\text{set}}$ in the limit of large influence rate α against horizon rate λ_{hor}. In the regime of high influence rate α, learning with sequences yields no significant gain over learning with sets, regardless of the horizon rate.

6.2.4 Discussion

What have we learnt from the characterization of the speed of learning using the Kullback-Leibler divergence for the three proposed hypothesis testing problems? First, we showcase a method to compare the speed of learning between learning based on sets, sequences, and times of adoptions. Although the computation of the KL divergences can get tedious, the method can be easily applied to simple hypothesis testing problems, such as the ones proposed here. Second, our characterization allows for more general conclusions on conditions under which the poor data mode of sets provides almost all the information needed for learning, or suffices for learning

key network relations. When actions are only observed over a short horizon window, then sets contain most of the information for learning (and, asymptotically as the horizon gets smaller, sets contain all the information for learning), and there is no much further value in richer temporal data. In contrast, when actions are observed over a long horizon window, then differences in the time of adoptions, or the order of adoptions, can be increasingly informative; equivalently, sets can be increasingly uninformative: not much information is extracted when adoption records consist of large sets of agents consistently, while learning with sets. Therefore, in the long horizon regime, learning with temporally richer data (i.e., times or sequences) has a large gain over learning merely with sets of adoptions. Our characterization thus reaches conclusions on the value of temporally richer data, or lack thereof, that agree with intuition.

Chapter 7

Theoretical Guarantees for Learning Influence with Zero/Infinity Edges

A directed[1] graph $G = (\mathcal{V}, \mathcal{E})$ is a priori given and $\lambda_{ij} = 0$ if edge (i, j) is not in $\mathcal{E}$. We provide theoretical guarantees for learning for the case where each edge in $\mathcal{E}$ carries an influence rate of either zero or infinity, casting the decision problem as a hypothesis testing problem. We restrict to influence rates that are either zero or infinite in order to simplify the analysis and derive crisp and insightful results. Given a graph G, lower and upper bounds for the number of i.i.d. products required to learn the correct hypothesis can be sought for different variations of the problem, according to the following axes:

- **Learning one edge versus learning all edges:** We pose two decision problems: learning the influence rate λ_{ij} between two specified agents i, j; and learning *all* the influence rates $\lambda_{ij}, i \neq j$.

- **Different prior knowledge over the hypotheses:** We study this question in the Bayesian setting of assuming a prior on the hypotheses, in the worst case over the hypotheses, as well as in the setting in which we know how many

[1] We allow bi-directed edges.

edges carry infinite influence rate. In general, a high prior probability of each edge carrying infinite influence rate, or knowing that a high number of edges carry infinite influence rate, correspond to the case of dense graphs; a low prior probability of each edge carrying infinite influence rate, or knowing that a low number of edges carry infinite influence rate, correspond to the case of sparse graphs, with few influence relations.

- **Data of different temporal detail:** We characterize the growth of the minimum number of i.i.d. products required for learning with respect to the number of agents n, when the available data provides information on sets, sequences, or times of adoptions.

- **Different scaling of the horizon rate with respect to the idiosyncratic rates:** We consider different scalings of λ_{hor} with respect to the idiosyncratic rates $\lambda_1, \ldots, \lambda_n$. Small values of λ_{hor} correspond to large horizon windows, during which many agents get to adopt; large values of λ_{hor} correspond to small horizon windows, during which only few agents get to adopt.

We first discuss conditions on the graph topology that guarantee learnability, and then we carry out the proposed program for the star topology. The star topology is one of the simplest non-trivial topologies, and is illustrative of the difference in the sample complexity between learning scenarios with information of different temporal detail.

7.1 Conditions on Topology for Learnability

We say that a graph is *learnable* if there exists an algorithm that learns all edges with probability of error that decays to zero in the limit of many samples. We show what graphs are learnable when learning with sets and when learning with sequences, assuming all edges carry influence zero or infinity. Adopting a Bayesian approach, we assume that if an edge exists in the sets of edges $\mathcal{E}$ of graph G, then the edge carries infinite influence with probability $q, 0 < q < 1$, and zero influence with

probability $1-q$. We also assume that the realization of each edge is independent of the realizations of all other edges. Last, we assume all idiosyncratic rates and the horizon rate to be equal to some $\lambda > 0$, which can be known or unknown.

Proposition 7. *When learning with sets, if the graph G has distinct nodes i, j, h such that*

(i) $(i,j) \in \mathcal{E}$, and

(ii) there exists a directed path from i to j through h,

then the graph is not learnable. If such triplet of distinct nodes does not exist, then the graph G is learnable, using $O(n^2 \log n)$ products. In particular, any polytree[2] *is learnable with sets, using $O(n^2 \log n)$ products.*

Proof. We focus on learning edge i, j. We first show the first half of the proposition.

We show that there is a region (which we call BAD) of large probability, where it is not clear what a good estimator should decide. No matter how this event is split between the competing hypotheses, the probability of error will be large.

We use X_p to denote the outcome of product p. We say the outcome X_p of product p is in BAD_p if one of the following happens: both agents i, j adopt; agent i does not adopt. We say the outcome $X_1, \ldots, X_k$ is in BAD if $X_p \in \text{BAD}_p$ for all products $p = 1, \ldots, k$.

We can write

$$\begin{aligned}
\mathbb{P}\left((X_1, \ldots, X_k) \in \text{BAD} \mid \lambda_{ij} = 0\right) &\geq \mathbb{P}\left(\text{path from } i \text{ to } j \text{ realized} \mid \lambda_{ij} = 0\right) \\
&\geq \mathbb{P}\left(\text{paths from } i \text{ to } h \text{ and from } h \text{ to } j \text{ realized} \mid \lambda_{ij} = 0\right) \\
&= q^{\ell_{ih} + \ell_{hj}} \\
&> 0,
\end{aligned}$$

where $\ell_{ih}(\ell_{hj})$ is the number of edges along a path from i to h (from h to j). Note that this is independent of the number of products k.

[2] A *polytree* is a directed acyclic graph (DAG) whose underlying undirected graph is a tree.

To show the second half of the proposition, we consider the following estimator: after k products, decide $\hat{\lambda}_{ij} = 0$ if there is a product such that agent i adopts and agent j does not adopt; otherwise, decide $\hat{\lambda}_{ij} = \infty$. Conditioning on the subset of agents $\mathcal{L}$ for which there is a directed path of edges carrying infinite influence to j, we can write

$$\begin{aligned}
\mathbb{P}(\text{error}) &= \mathbb{P}(\lambda_{ij} = 0) \cdot \mathbb{P}\left(\hat{\lambda}_{ij} = \infty \mid \lambda_{ij} = 0\right) \\
&= (1-q) \cdot \sum_{\mathcal{L} \subseteq \{1,\ldots,n\} \setminus \{i,j\}} \mathbb{P}\left(\hat{\lambda}_{ij} = \infty \mid \lambda_{ij} = 0, \mathcal{L}\right) \mathbb{P}\left(\mathcal{L} \mid \lambda_{ij} = 0\right). \quad (7.1)
\end{aligned}$$

Assuming $|\mathcal{L}| = m$, we can write for a given product:

$$\begin{aligned}
\mathbb{P}\left(i \text{ adopts, } j \text{ does not} \mid \lambda_{ij} = 0, \mathcal{L}\right) &\geq \mathbb{P}\left(i \text{ adopts first, } j \text{ does not adopt} \mid \lambda_{ij} = 0, \mathcal{L}\right) \\
&= \frac{\lambda}{n\lambda + \lambda} \cdot \frac{\lambda}{\lambda + m\lambda + \lambda} \\
&= \frac{1}{n+1} \cdot \frac{1}{m+2}.
\end{aligned}$$

Denoting with M the random variable which is the number of agents for which there is a directed path of edges carrying infinite influence to j (i.e., $M = |\mathcal{L}|$), we can now rewrite Equation (7.1) as

$$\begin{aligned}
\mathbb{P}(\text{error}) &\leq (1-q) \cdot \sum_{m=0}^{n-2} \left(1 - \frac{1}{(n+1)(m+2)}\right)^k p_{M|\lambda_{ij}=0}(m) \\
&\leq (1-q) \left(1 - \frac{1}{(n+1)n}\right)^k \\
&= (1-q) \left(\frac{n(n+1)-1)}{n(n+1)}\right)^k \longrightarrow 0 \text{ as } k \longrightarrow \infty.
\end{aligned}$$

In fact, assuming $q = \Theta(1)$, to ensure an accurate estimate for λ_{ij} with probability at least $1 - \delta$, for given $\delta \in (0,1)$, it suffices that $k \geq \frac{\log \frac{1-q}{\delta}}{\log \frac{n(n+1)}{n(n+1)-1}} = O(n^2)$.

Using the union bound, we relate the probability of error in learning all the edges

of the graph, to the probability of error in learning a single coefficient λ_{ij}:

$$\mathbb{P}(\text{error}) \quad \leq \quad n(n-1) \cdot (1-q) \left(\frac{n(n+1)-1)}{n(n+1)} \right)^k .$$

Again, assuming $q = \Theta(1)$, to ensure accurate estimates for all the edges with probability at least $1-\delta$, for given $\delta \in (0,1)$, it suffices that $k \geq \frac{\log \frac{n(n-1)(1-q)}{\delta}}{\log \frac{n(n+1)}{n(n+1)-1}} = O(n^2 \log n)$. □

We now define the data mode of enhanced sequences, which we consider in the setting where all edges carry influence zero or infinity.

Definition 2. With access to *enhanced sequences*, the observer has the following information for each product:

- the connected components of adopters;
- the order in which the connected components adopted;
- what agent adopted first in a connected component;
- who was influenced directly by the first adopter in a connected component.

This data mode provides knowledge of the agents that adopted or did not adopt immediately following an adoption by some agent. This knowledge can be leveraged to decide that a given edge carries infinite or zero influence.

Proposition 8. *When learning with enhanced sequences, all the graphs G are learnable, using $O(n \log n)$ products.*

Proof. We consider the following estimator: after k products, decide $\hat{\lambda}_{ij} = 0$ if there is a product such that agent i adopts before agent j, and agent j does not adopt immediately after; otherwise, decide $\hat{\lambda}_{ij} = \infty$. We can write

$$\begin{aligned} \mathbb{P}(\text{error}) &= \mathbb{P}(\lambda_{ij} = 0) \cdot \mathbb{P}\left(\hat{\lambda}_{ij} = \infty \mid \lambda_{ij} = 0\right) \\ &= (1-q) \cdot \mathbb{P}\left(\hat{\lambda}_{ij} = \infty \mid \lambda_{ij} = 0\right). \end{aligned} \tag{7.2}$$

We can also write for a given product:

$$\begin{aligned}
&\mathbb{P}\left(i \text{ adopts before } j,\ j \text{ does not adopt immediately after} \mid \lambda_{ij} = 0\right) \\
&\qquad \geq \mathbb{P}\left(i \text{ adopts first} \mid \lambda_{ij} = 0\right) \\
&\qquad = \tfrac{\lambda}{n\lambda+\lambda} \\
&\qquad = \tfrac{1}{n+1}.
\end{aligned}$$

We can now rewrite Equation (7.2) as

$$\mathbb{P}(\text{error}) \quad \leq \quad (1-q)\cdot\left(1-\frac{1}{n+1}\right)^k \longrightarrow 0 \text{ as } k \longrightarrow \infty.$$

In fact, assuming $q = \Theta(1)$, to ensure an accurate estimate for λ_{ij} with probability at least $1-\delta$, for given $\delta \in (0,1)$, it suffices that $k \geq \frac{\log \frac{1-q}{\delta}}{\log \frac{n+1}{n}} = O(n)$.

Using the union bound, we relate the probability of error in learning all the edges of the graph, to the probability of error in learning λ_{ij}:

$$\mathbb{P}(\text{error}) \quad \leq \quad n(n-1)\cdot(1-q)\left(1-\frac{1}{n+1}\right)^k.$$

Again, assuming $q = \Theta(1)$, to ensure accurate estimates for all the edges with probability at least $1-\delta$, for given $\delta \in (0,1)$, it suffices that $k \geq \frac{\log \frac{n(n-1)(1-q)}{\delta}}{\log \frac{n+1}{n}} = O(n \log n)$. $\square$

7.2 Learning Influence in the Star Network

We consider the hypothesis testing problem in which each of n agents influence agent $n+1$ either with rate zero or infinity. (A rate of $\lambda_{i,n+1} = \infty$ signifies that agent $n+1$ adopts right when agent i adopts.) Each of agents $1,\ldots,n$ adopts with rate $\lambda > 0$, which can be known or unknown, while agent $n+1$ does not adopt unless she is triggered to. There is no influence from agent $n+1$ to any of the agents $1,\ldots,n$, or from any of the agents $1,\ldots,n$ to any of the agents $1,\ldots,n$. Figure 7-1 illustrates this model.

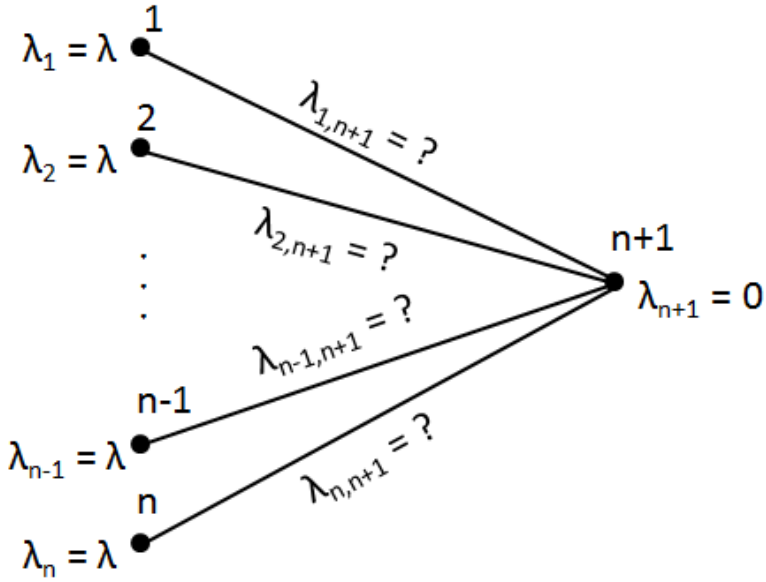


Figure 7-1: The hypothesis testing problem: what influence does each link carry to the star agent $(n+1)$: infinite or zero?

We assume that the horizon rate is equal to the agents' idiosyncratic rate of adoption, that is, $\lambda_{hor} = \lambda$. We pose two decision problems: learning the influence rate between a specified agent i and agent $n+1$; and learning all the influence rates $\lambda_{1,n+1}, \ldots, \lambda_{n,n+1}$. In each case, our goal is to come up with upper and lower bounds for the minimum number of i.i.d. products required to learn the correct hypothesis. We study this question in the Bayesian setting of assuming a prior on the hypotheses, in the worst case over the hypotheses, as well as in the setting in which we know how many agents have infinite influence rate and how many have zero influence. We characterize the growth of the minimum number of i.i.d. products required for learning with respect to n, both when the available data provides information on sets of adopters, and when the available data provides information on sequences of adopters. (Of course, knowledge of times of adoptions will not induce a gain over knowledge of sequences, because of our assumption that the influence rates are either zero or infinite, and $\lambda_{n+1} = 0$.)

7.2.1 The Bayesian Setting

In the Bayesian setting, we assume that the influence rate on each link is infinite, with probability p, and zero, with probability $1-p$, and that the selection of the rate for each link is independent of the selection for other links.

The Case $p = 1/2$. We assume that the influence rate on each link will be zero or infinite with equal probability. Table 7.1 summarizes the results on the necessary and sufficient number of i.i.d. products for learning.

Table 7.1: Matching lower and upper bounds for the minimum number of i.i.d. products required to learn the influence model in terms of n, in the Bayesian setting when $p = 1/2$, for the two cases of learning the influence between one agent and the star agent and of learning the influence between all agents and the star agent, and for the two cases of learning based on sets of adoptions or sequences of adoptions.

	Sets	Sequences
Learn one	$\Theta(n^2)$	$\Theta(n)$
Learn all	$\Theta(n^2 \log n)$	$\Theta(n \log n)$

We set to prove the results in the table, starting with the case where we observe sets.

Proposition 9. *To ensure correct learning of $\lambda_{1,n+1}$ with probability $1-\delta$ based on sets of adopting agents, it is sufficient for the number of i.i.d. products to be $O(n^2)$, and necessary for the number of i.i.d. products to be $\Omega(n^2)$. To ensure correct learning of $\lambda_{1,n+1}, \ldots, \lambda_{n,n+1}$ with probability $1-\delta$ based on sets of adopting agents, it is sufficient for the number of i.i.d. products to be $O(n^2 \log n)$, and necessary for the number of i.i.d. products to be $\Omega(n^2 \log n)$.*

Proof. To prove the upper bounds, we consider the following estimator: after k products, decide $\hat{\lambda}_{1,n+1} = 0$ if and only if there exists a product such that agent 1 adopts and agent $n+1$ does not adopt (and decide $\hat{\lambda}_{1,n+1} = \infty$ otherwise). For this estimator, we write the probability of error, conditioning on the subset $\mathcal{L}$ of agents $2, \ldots, n$

whose influence rate on agent $n+1$ is infinite:

$$\begin{aligned}
\mathbb{P}(\text{error}) &= \mathbb{P}\left(\hat{\lambda}_{1,n+1} = \infty \mid \lambda_{1,n+1} = 0\right) \mathbb{P}(\lambda_{1,n+1} = 0) \\
&\quad + \mathbb{P}\left(\hat{\lambda}_{1,n+1} = 0 \mid \lambda_{1,n+1} = \infty\right) \mathbb{P}(\lambda_{1,n+1} = \infty) \\
&= \mathbb{P}\left(\hat{\lambda}_{1,n+1} = \infty \mid \lambda_{1,n+1} = 0\right) \mathbb{P}(\lambda_{1,n+1} = 0) \\
&= \frac{1}{2} \sum_{\mathcal{L} \subseteq \{2,\ldots,n\}} \mathbb{P}\left(\hat{\lambda}_{1,n+1} = \infty \mid \mathcal{L}, \lambda_{1,n+1} = 0\right) \mathbb{P}\left(\mathcal{L} \mid \lambda_{1,n+1} = 0\right) \\
&= \frac{1}{2} \sum_{m=0}^{n-1} \left(1 - \frac{\lambda}{m\lambda + \lambda + \lambda} \cdot \frac{\lambda}{m\lambda + \lambda}\right)^k \binom{n-1}{m} \left(\frac{1}{2}\right)^{n-1} \\
&\leq \frac{1}{2} \left(\frac{n(n+1) - 1}{n(n+1)}\right)^k,
\end{aligned}$$

where the last inequality follows by setting all summands equal to the largest summand, which is the summand that corresponds to $m = n - 1$. To ensure accurate estimates with probability at least $1 - \delta$, for given $\delta \in (0, 1)$, it suffices that $k \geq \frac{\log \frac{1}{2\delta}}{\log \frac{n(n+1)}{n(n+1)-1}} = O(n^2)$.

Using the union bound, we relate the probability of error in learning all of $\lambda_{1,n+1}$, $\ldots$, $\lambda_{n,n+1}$ to the probability of error in learning $\lambda_{1,n+1}$:

$$\begin{aligned}
\mathbb{P}(\text{error}) &\leq n \cdot \mathbb{P}\left(\hat{\lambda}_{1,n+1} = 0 \mid \lambda_{1,n+1} = \infty\right) \mathbb{P}(\lambda_{1,n+1} = \infty) \\
&\leq n \cdot \frac{1}{2} \left(\frac{n(n+1) - 1}{n(n+1)}\right)^k.
\end{aligned}$$

To ensure accurate estimates with probability at least $1 - \delta$, for given $\delta \in (0, 1)$, it suffices that $k \geq \frac{\log \frac{n}{2\delta}}{\log \frac{n(n+1)}{n(n+1)-1}} = O(n^2 \log n)$.

To prove the lower bounds, we show that if k is small, there is a high probability event, where it is not clear what a good estimator should decide. In this region, the likelihood ratio of the two hypotheses is bounded away from zero and also bounded above. No matter how this event is split between the competing hypotheses, the probability of error will be large.

We use X_i to denote the outcome of product i. We say the outcome X_i of product i is in BAD_i if one of the following happens: (i) both agents 1 and $n+1$ adopt; (ii)

agent 1 does not adopt. We say the outcome $X_1, \ldots, X_k$ is in BAD if $X_i \in \mathrm{BAD}_i$ for all products $i = 1, \ldots, k$.

We first show the lower bound for learning just $\lambda_{1,n+1}$. First notice that for all $(x_1, \ldots, x_k) \in \mathrm{BAD}$, the likelihood ratio $\frac{\mathbb{P}(x_1,\ldots,x_k|\lambda_{1,n+1}=0)}{\mathbb{P}(x_1,\ldots,x_k|\lambda_{1,n+1}=\infty)}$ is bounded away from zero, and is smaller than one. Conditioning on the subset $\mathcal{L}$ of agents $2, \ldots, n$ whose influence rate on agent $n+1$ is infinite and using the total probability theorem, we write

$$\begin{aligned}
\mathbb{P}\left(X_i \in \mathrm{BAD}_i^{\complement} \mid \lambda_{1,n+1} = 0\right) &= \sum_{\mathcal{L} \subseteq \{2,\ldots,n\}} \mathbb{P}\left(X_i \in \mathrm{BAD}_i^{\complement} \mid \mathcal{L}, \lambda_{1,n+1} = 0\right) \mathbb{P}(\mathcal{L} \mid \lambda_{1,n+1} = 0) \\
&= \sum_{m=0}^{n-1} \frac{\lambda}{m\lambda + \lambda + \lambda} \cdot \frac{\lambda}{m\lambda + \lambda} \binom{n-1}{m} \left(\frac{1}{2}\right)^{n-1} \\
&= \frac{1}{2^{n-1}} \sum_{m=0}^{n-1} \frac{1}{(m+1)(m+2)} \binom{n-1}{m} \\
&= \frac{1}{n(n+1)} \cdot \frac{1}{2^{n-1}} \sum_{m=0}^{n-1} \frac{(n+1)!}{(m+2)!(n-1-m)!} \\
&= \frac{1}{n(n+1)} \cdot \frac{1}{2^{n-1}} \sum_{m=0}^{n-1} \binom{n+1}{m+2} \\
&= \frac{1}{n(n+1)} \cdot \frac{1}{2^{n-1}} \sum_{m'=2}^{n+1} \binom{n+1}{m'} \\
&= \frac{1}{n(n+1)} \cdot \frac{1}{2^{n-1}} \left(2^{n+1} - 1 - (n+1)\right) \\
&= \frac{1}{n(n+1)} \left(4 - \frac{n+2}{2^{n-1}}\right),
\end{aligned}$$

and thus, using the union bound,

$$\begin{aligned}
\mathbb{P}\Big((X_1, \ldots, X_k) \in \mathrm{BAD} \mid \lambda_{1,n+1} = 0\Big) &= \mathbb{P}\left(\bigcap_{i=1}^{k} (X_i \in \mathrm{BAD}_i) \mid \lambda_{1,n+1} = 0\right) \\
&= 1 - \mathbb{P}\left(\bigcup_{i=1}^{k} \left(X_i \in \mathrm{BAD}_i^{\complement}\right) \mid \lambda_{1,n+1} = 0\right) \\
&\geq 1 - k \cdot \mathbb{P}\left(X_i \in \mathrm{BAD}_i^{\complement} \mid \lambda_{1,n+1} = 0\right) \\
&\geq 1 - k \cdot \frac{1}{n(n+1)} \left(4 - \frac{n+2}{2^{n-1}}\right),
\end{aligned}$$

establishing the $\Omega(n^2)$ lower bound for the number of products k.

The proof for the lower bound for learning all of $\lambda_{1,n+1}, \ldots, \lambda_{n,n+1}$ relies on a coupon collector's argument, similarly to the proof for the lower bound of $\Omega(n \log n)$ when learning with sequences of adoptions in Proposition 10. $\square$

Proposition 10. *To ensure correct learning of $\lambda_{1,n+1}$ based on sequences of adoptions, it is sufficient for the number of i.i.d. products to be $O(n)$, and necessary for the number of i.i.d. products to be $\Omega(n)$. To ensure correct learning of $\lambda_{1,n+1}, \ldots, \lambda_{n,n+1}$ with probability $1-\delta$ based on sequences of adoptions, it is sufficient for the number of i.i.d. products to be $O(n \log n)$, and necessary for the number of i.i.d. products to be $\Omega(n \log n)$.*

Proof. To prove the upper bounds, we consider the following estimator: after k products, decide $\hat{\lambda}_{i,n+1} = \infty$ if and only if there exists a product such that agent i adopts and agent $n+1$ adopts immediately after (and decide $\hat{\lambda}_{i,n+1} = 0$ otherwise). For this estimator, we write the probability of error, conditioning on the subset $\mathcal{L}$ of agents $2, \ldots, n$ whose influence rate on agent $n+1$ is infinite:

$$\begin{aligned}
\mathbb{P}(\text{error}) &= \mathbb{P}\left(\hat{\lambda}_{1,n+1} = \infty \mid \lambda_{1,n+1} = 0\right) \mathbb{P}(\lambda_{1,n+1} = 0) \\
&\quad + \mathbb{P}\left(\hat{\lambda}_{1,n+1} = 0 \mid \lambda_{1,n+1} = \infty\right) \mathbb{P}(\lambda_{1,n+1} = \infty) \\
&= \mathbb{P}\left(\hat{\lambda}_{1,n+1} = 0 \mid \lambda_{1,n+1} = \infty\right) \mathbb{P}(\lambda_{1,n+1} = \infty) \\
&= \frac{1}{2} \sum_{\mathcal{L} \subseteq \{2,\ldots,n\}} \mathbb{P}\left(\hat{\lambda}_{1,n+1} = 0 \mid \mathcal{L}, \lambda_{1,n+1} = \infty\right) \mathbb{P}\left(\mathcal{L} \mid \lambda_{1,n+1} = \infty\right) \\
&= \frac{1}{2} \sum_{m=0}^{n-1} \left(1 - \frac{\lambda}{m\lambda + \lambda + \lambda}\right)^k \binom{n-1}{m} \left(\frac{1}{2}\right)^{n-1} \\
&\leq \frac{1}{2} \left(\frac{n}{n+1}\right)^k
\end{aligned}$$

To ensure accurate estimates with probability at least $1-\delta$, for given $\delta \in (0,1)$, it suffices that $k \geq \frac{\log \frac{1}{2\delta}}{\log \frac{n+1}{n}} = O(n)$.

Using the union bound, we relate the probability of error in learning all of $\lambda_{1,n+1}$,

$\ldots,\ \lambda_{n,n+1}$ to the probability of error in learning $\lambda_{1,n+1}$:

$$\begin{aligned}\mathbb{P}(\text{error}) &\leq n \cdot \mathbb{P}\left(\hat{\lambda}_{1,n+1} = 0 \mid \lambda_{1,n+1} = \infty\right) \mathbb{P}(\lambda_{1,n+1} = \infty) \\ &\leq n \cdot \frac{1}{2}\left(\frac{n}{n+1}\right)^k.\end{aligned}$$

To ensure accurate estimates with probability at least $1-\delta$, for given $\delta \in (0,1)$, it suffices that $k \geq \frac{\log \frac{n}{2\delta}}{\log \frac{n+1}{n}} = O(n \log n)$.

To prove the lower bounds, we show that if k is small, there is a high probability event, where it is not clear what a good estimator should decide. In this region, the likelihood ratio of the two hypotheses is bounded away from zero and also bounded above. No matter how this event is split between the competing hypotheses, the probability of error will be large.

We use X_i to denote the outcome of product i. Having fixed agent j, we say the outcome X_i of product i is in BAD_i^j if one of the following happens: (i) agent j adopts, but agent $n+1$ adopts before her; (ii) agent j does not adopt. We say that the outcome $X_1, \ldots, X_k$ is in BAD^j if $X_i \in \text{BAD}_i^j$ for all products $i = 1, \ldots, k$.

We first show the lower bound for learning just $\lambda_{1,n+1}$. First notice that for all $(x_1, \ldots, x_k) \in \text{BAD}^1$, we have $\frac{\mathbb{P}(x_1,\ldots,x_k|\lambda_{1,n+1}=0)}{\mathbb{P}(x_1,\ldots,x_k|\lambda_{1,n+1}=\infty)} = 1$. Conditioning on the subset $\mathcal{L}$ of agents $2, \ldots, n$ whose influence rate on agent $n+1$ is infinite and using the total probability theorem, we write

$$\begin{aligned}\mathbb{P}\left(X_i \in \text{BAD}_i^{1\complement} \mid \lambda_{1,n+1} = 0\right) &= \sum_{\mathcal{L} \subseteq \{2,\ldots,n\}} \mathbb{P}\left(X_i \in BAD_i^{1\complement} \mid \mathcal{L}, \lambda_{1,n+1} = 0\right) \mathbb{P}(\mathcal{L} \mid \lambda_{1,n+1} = 0) \\ &= \sum_{m=0}^{n-1} \left(1 - \left(\frac{\lambda}{2\lambda} + \frac{\lambda}{2\lambda} \cdot \frac{m\lambda}{m\lambda + \lambda + \lambda}\right)\right) \binom{n-1}{m} \left(\frac{1}{2}\right)^{n-1} \\ &= \frac{1}{2^{n-1}} \sum_{m=0}^{n-1} \frac{1}{m+2} \binom{n-1}{m} \\ &\leq \frac{1}{2^{n-1}} \sum_{m=0}^{n-1} \frac{1}{m+1} \binom{n-1}{m} \\ &= \frac{1}{n} \cdot \frac{1}{2^{n-1}} \sum_{m=0}^{n-1} \frac{n!}{(m+1)!(n-m-1)!}\end{aligned}$$

$$
\begin{aligned}
&= \frac{1}{n} \cdot \frac{1}{2^{n-1}} \sum_{m=0}^{n-1} \binom{n}{m+1} \\
&= \frac{1}{n} \cdot \frac{1}{2^{n-1}} \sum_{m'=1}^{n} \binom{n}{m'} \\
&= \frac{1}{n} \cdot \frac{1}{2^{n-1}} (2^n - 1) \\
&= \frac{1}{n} \left(2 - \frac{1}{2^{n-1}} \right),
\end{aligned}
$$

and thus, using the union bound,

$$
\begin{aligned}
\mathbb{P}\left((X_1, \ldots, X_k) \in \mathrm{BAD}^1 \mid \lambda_{1,n+1} = 0\right) &= \mathbb{P}\left(\bigcap_{i=1}^{k} \left(X_i \in \mathrm{BAD}_i^1\right) \mid \lambda_{1,n+1} = 0 \right) \\
&= 1 - \mathbb{P}\left(\bigcup_{i=1}^{k} \left(X_i \in \mathrm{BAD}_i^{1\complement}\right) \mid \lambda_{1,n+1} = 0 \right) \\
&\geq 1 - k \cdot \mathbb{P}\left(X_i \in \mathrm{BAD}_i^{1\complement} \mid \lambda_{1,n+1} = 0\right) \\
&\geq 1 - k \cdot \frac{1}{n} \left(2 - \frac{1}{2^{n-1}} \right),
\end{aligned}
$$

establishing the $\Omega(n)$ lower bound for the number of products k.

We now show the lower bound for learning all of $\lambda_{1,n+1}, \ldots, \lambda_{n,n+1}$. We are interested in the probability that for some agent j, it is the case that $X_1, \ldots, X_k \in \mathrm{BAD}^j$. We define A to be the event that each of the agents $1, \ldots, n$ adopts some product before all (other) agents $1, \ldots, n$ with links of rate infinity to agent $n+1$ adopt that product. We define B to be the event that all agents with links of rate infinity to agent $n+1$ adopt some product first among other agents with links of rate infinity to agent $n+1$. Then, we can write

$$
\begin{aligned}
\mathbb{P}\left(\exists j : (X_1, \ldots, X_k) \in \mathrm{BAD}^j\right) &= 1 - \mathbb{P}(A) \\
&\geq 1 - \mathbb{P}(B).
\end{aligned}
$$

Let random variable S be the number of i.i.d. products until event A occurs. Let random variable T be the number of i.i.d. products to obtain event B. Then $S \geq T$. The calculation of the expectation of T is similar to the calculation for the coupon

collector's problem, after conditioning on the subset of agents $\mathcal{L} \subseteq \{1, \ldots, n\}$ whose influence rate on agent $n+1$ is infinite:

$$\begin{aligned}
\mathbb{E}[T] &= \sum_{\mathcal{L} \subseteq \{1,\ldots,n\}} \mathbb{P}(\mathcal{L}) \mathbb{E}[T \mid \mathcal{L}] \\
&= \sum_{m=0}^{n} \mathbb{E}[T \mid \mathcal{L}] \binom{n}{m} \left(\frac{1}{2}\right)^n \\
&= \frac{1}{2^n} \sum_{m=0}^{n} \left(\left(\frac{m\lambda}{m\lambda+\lambda}\right)^{-1} + \left(\frac{(m-1)\lambda}{m\lambda+\lambda}\right)^{-1} + \ldots + \left(\frac{\lambda}{m\lambda+\lambda}\right)^{-1} \right) \binom{n}{m} \\
&= \frac{1}{2^n} \sum_{m=0}^{n} \left(\frac{m+1}{m} + \frac{m+1}{m-1} + \ldots + \frac{m+1}{1} \right) \binom{n}{m} \\
&= \frac{1}{2^n} \sum_{m=0}^{n} (m+1) H_m \binom{n}{m} \\
&= \Omega(n \log n),
\end{aligned}$$

where H_m is the mth harmonic number, i.e., $H_m = \sum_{k=1}^{m} \frac{1}{k}$ (and we define $H_0 = 0$), and where the last step follows because, defining $f(m) = (m+1)H_m, m \geq 0$, and using Jensen's inequality, we have

$$\begin{aligned}
\frac{1}{2^n} \sum_{m=0}^{n} (m+1) H_m \binom{n}{m} &\geq f\left(\left\lfloor \frac{1}{2^n} \sum_{m=0}^{n} m \binom{n}{m} \right\rfloor \right) \\
&= f\left(\left\lfloor \frac{1}{2^n} n \sum_{m=1}^{n} \binom{n-1}{m-1} \right\rfloor \right) \\
&= f\left(\left\lfloor \frac{1}{2^n} n \sum_{m'=0}^{n-1} \binom{n-1}{m'} \right\rfloor \right) \\
&= f\left(\left\lfloor \frac{1}{2^n} n 2^{n-1} \right\rfloor \right) \\
&= f\left(\left\lfloor \frac{n}{2} \right\rfloor \right) \\
&= \left(\left\lfloor \frac{n}{2} \right\rfloor + 1 \right) H_{\left\lfloor \frac{n}{2} \right\rfloor} \\
&= \Theta(n \log n).
\end{aligned}$$

Similarly, for the variance we can write

$$
\begin{aligned}
var\,(T) &= \sum_{\mathcal{L}\subseteq\{1,\ldots,n\}} \mathbb{P}(\mathcal{L}) var\,(T \mid \mathcal{L}) \\
&= \sum_{m=0}^{n} var\,(T \mid \mathcal{L}) \binom{n}{m} \left(\frac{1}{2}\right)^n \\
&= \frac{1}{2^n} \sum_{m=0}^{n} \left(\frac{1-\frac{m\lambda}{m\lambda+\lambda}}{\left(\frac{m\lambda}{m\lambda+\lambda}\right)^2} + \frac{1-\frac{(m-1)\lambda}{m\lambda+\lambda}}{\left(\frac{(m-1)\lambda}{m\lambda+\lambda}\right)^2} + \ldots + \frac{1-\frac{\lambda}{m\lambda+\lambda}}{\left(\frac{\lambda}{m\lambda+\lambda}\right)^2} \right) \binom{n}{m} \\
&= \frac{1}{2^n} \sum_{m=0}^{n} \left(\frac{\frac{1}{m+1}}{\left(\frac{m}{m+1}\right)^2} + \frac{\frac{2}{m+1}}{\left(\frac{m-1}{m+1}\right)^2} + \ldots + \frac{\frac{m}{m+1}}{\left(\frac{1}{m+1}\right)^2} \right) \binom{n}{m} \\
&\leq \frac{1}{2^n} \sum_{m=0}^{n} \left(\frac{1}{\left(\frac{m}{m+1}\right)^2} + \frac{1}{\left(\frac{m-1}{m+1}\right)^2} + \ldots + \frac{1}{\left(\frac{1}{m+1}\right)^2} \right) \binom{n}{m} \\
&= \frac{1}{2^n} \sum_{m=0}^{n} \left((m+1)^2 \left(\frac{1}{1^2} + \frac{1}{2^2} + \ldots + \frac{1}{m^2} \right) \right) \binom{n}{m} \\
&\leq (n+1)^2 \left(\frac{1}{1^2} + \frac{1}{2^2} + \ldots \right) \\
&= (n+1)^2 \cdot \frac{\pi^2}{6} \\
&\leq 2(n+1)^2.
\end{aligned}
$$

By Chebyshev's inequality,

$$\mathbb{P}\left(|T - \mathbb{E}[T]| \geq c(n+1)\right) \leq \frac{2}{c^2}.$$

Therefore, with $k = o(n \log n)$ products, there is a very small probability that event B will occur, and therefore a very large probability that the event $\{\exists j : (X_1, \ldots, X_k) \in \mathrm{BAD}^j\}$ will occur, which establishes the $\Omega(n \log n)$ lower bound for the number of products k. □

The Case $p = 1/n$. We assume that the influence rate on each link will be infinite with probability $p = 1/n$. (In this case, the expected number of agents who can influence agent $n+1$ is $\Theta(1)$.) Table 7.2 summarizes the results on the necessary and sufficient number of i.i.d. products for learning.

Table 7.2: Matching lower and upper bounds for the minimum number of i.i.d. products required to learn the influence model, in terms of n, in the Bayesian setting when $p = 1/n$, when learning the influence between all agents and the star agent, for the two cases of learning based on sets of adoptions or sequences of adoptions.

	Sets	Sequences
Learn all	$\Theta(\log n)$	$\Theta(1)$

We do not consider learning the influence between one agent and the star agent, because an algorithm that simply guesses that the influence is zero, has probability of error $1/n$, without using any products.

We set to prove the results in the table, starting with the case where we observe sets.

Proposition 11. *(Assume that $p = 1/n$.) To ensure correct learning of $\lambda_{1,n+1}, \ldots, \lambda_{n,n+1}$ with probability $1 - \delta$ based on sets of adopting agents, it is sufficient for the number of i.i.d. products to be $O(\log n)$, and necessary for the number of i.i.d. products to be $\Omega(\log n)$.*

Proof. Define A to be the event that every agent out of the agents with influence rate zero to agent $n + 1$ adopts some product without any of the agents with influence rate infinity to agent $n + 1$ adopting. For learning, we need event A to occur; and if event A occurs, then we have learning.

To calculate the expected number of i.i.d. products to obtain event A, we first note that the expected number of agents with infinite influence rate to agent $n+1$ is 1. Given a product, we define X to be the random variable which is the number of agents with influence rate zero to agent $n+1$ who adopt. We also define C to be the event that, given a product, none of the agents with influence rate infinity to agent $n + 1$ adopts. We are interested in $\mathbb{E}[X \mid C]$. The latter will grow asymptotically as fast as the expectation of X conditioned on the event C, assuming knowledge of the fact that exactly one agent has influence rate infinity to agent $n + 1$.

Assuming we know that exactly one agent has influence rate infinity to agent $n+1$,

we calculate the probability mass function of X, conditioned on the event C:

$$\begin{aligned}
p_{X|C}(m) &= \frac{\mathbb{P}\left(\{X=m\}\cap C\right)}{\mathbb{P}(C)} \\
&= \frac{\binom{n-1}{m}\cdot\frac{m\lambda}{n\lambda+\lambda}\cdot\frac{(m-1)\lambda}{(n-1)\lambda+\lambda}\cdots\frac{\lambda}{(n-(m-1))\lambda+\lambda}\cdot\frac{\lambda}{(n-m)\lambda+\lambda}}{1/2} \\
&= \frac{\frac{(n-1)!}{m!(n-m-1)!}\cdot\frac{m}{n+1}\cdot\frac{m-1}{n}\cdots\frac{1}{n-m+2}\cdot\frac{1}{n-m+1}}{1/2} \\
&= \frac{2(n-m)}{n(n+1)}, \quad m=0,\ldots,n-1.
\end{aligned}$$

We now calculate the conditional expectation of X:

$$\begin{aligned}
\mathbb{E}[X \mid C] &= \sum_{m=0}^{n-1} m\frac{2(n-m)}{n(n+1)} \\
&= \frac{2}{n(n+1)}\left(n\sum_{m=0}^{n-1} m - \sum_{m=0}^{n-1} m^2\right) \\
&= \frac{2}{n(n+1)}\left(n\left(\frac{1}{2}(n-1)^2+\frac{1}{2}(n-1)\right)-\left(\frac{1}{3}(n-1)^3+\frac{1}{2}(n-1)^2+\frac{1}{6}(n-1)\right)\right) \\
&= \frac{1}{3}n + o(n)
\end{aligned}$$

In expectation, half the products will not be adopted by the agent with influence rate infinity to agent $n+1$, and for the products that are not adopted by the agent with influence rate infinity to agent $n+1$, the number of adopters among the agents with influence rate zero to agent $n+1$ is $\frac{1}{3}n$. It follows that $O(\log n)$ products suffice for correct learning with high probability. Furthermore, with $o(\log n)$ products, there is a very small probability that event A will occur, and therefore a very small probability that learning will occur, which establishes the $\Omega(\log n)$ lower bound for the number of products.

It follows that the required number of i.i.d. products to obtain event A grows as $\Theta(\log n)$. □

Proposition 12. *(Assume that $p = 1/n$.) To ensure correct learning of $\lambda_{1,n+1},\ldots,\lambda_{n,n+1}$ with probability $1-\delta$ based on sequences of adoptions, it is sufficient for the number*

of i.i.d. products to be $O(1)$, and necessary for the number of i.i.d. products to be $\Omega(1)$.

Proof. Define A to be the event that every agent with influence rate infinity to agent $n+1$ is the first one to adopt among other agents with influence rate infinity to agent $n+1$ for some product. If event A occurs, then learning occurs. Let random variable T be the number of i.i.d. products until event A occurs. We calculate the expected value of T, conditioning on the subset of agents $\mathcal{L} \subseteq \{1, \ldots, n\}$ who have influence rate infinity to agent $n+1$:

$$\begin{aligned}
\mathbb{E}[T] &= \sum_{\mathcal{L} \subseteq \{1,\ldots,n\}} \mathbb{P}(\mathcal{L}) \mathbb{E}[T \mid \mathcal{L}] \\
&= \sum_{m=0}^{n} \left(\left(\frac{m\lambda}{m\lambda + \lambda} \right)^{-1} + \left(\frac{(m-1)\lambda}{m\lambda + \lambda} \right)^{-1} + \ldots + \left(\frac{\lambda}{m\lambda + \lambda} \right)^{-1} \right) \binom{n}{m} \left(\frac{1}{n} \right)^{m} \left(1 - \frac{1}{n} \right)^{n-m} \\
&= \sum_{m=0}^{n} \left(\frac{m+1}{m} + \frac{m+1}{m-1} + \ldots + \frac{m+1}{1} \right) \binom{n}{m} \left(\frac{1}{n} \right)^{m} \left(1 - \frac{1}{n} \right)^{n-m} \\
&= \sum_{m=0}^{n} (m+1) H_m \binom{n}{m} \left(\frac{1}{n} \right)^{m} \left(1 - \frac{1}{n} \right)^{n-m} \\
&= \Theta(1),
\end{aligned}$$

where H_m is the mth harmonic number, i.e., $H_m = \sum_{k=1}^{m} \frac{1}{k}$ (and we define $H_0 = 0$). By Markov's inequality, we have

$$\mathbb{P}\left(T \geq \frac{\mathbb{E}[T]}{\delta} \right) \leq \delta,$$

therefore with $\frac{\mathbb{E}[T]}{\delta} = O(1)$ products, we can learn correctly with probability at least $1 - \delta$. □

7.2.2 The Worst-Case Setting

In the worst-case setting, we assume that each of the influence rates $\lambda_{1,n+1}, \ldots, \lambda_{n,n+1}$ can be either zero or infinity, but we assume no prior over the hypotheses. We provide upper and lower bounds for the minimum number of i.i.d. products required to

learn the correct hypothesis, for the problem of learning the influence rate between a specified agent i and agent $n+1$, and for the problem of learning all the influence rates $\lambda_{1,n+1}, \ldots, \lambda_{n,n+1}$, assuming that the influence rates on the links are such that the minimum number of i.i.d. products required for learning is maximized. Table 7.3 summarizes our results.

Table 7.3: Matching lower and upper bounds for the minimum number of i.i.d. products required to learn the influence model in terms of n, in the worst-case setting, for the two cases of learning the influence between one agent and the star agent and of learning the influence between all agents and the star agent, and for the two cases of learning based on sets of adoptions or sequences of adoptions.

	Sets	Sequences
Learn one	$\Theta(n^2)$	$\Theta(n)$
Learn all	$\Theta(n^2 \log n)$	$\Theta(n \log n)$

We set to prove the results in the table.

Proposition 13. *To ensure correct learning of $\lambda_{1,n+1}$ with probability $1-\delta$ based on sets of adopting agents, it is sufficient for the number of i.i.d. products to be $O(n^2)$, and necessary for the number of i.i.d. products to be $\Omega(n^2)$. To ensure correct learning of $\lambda_{1,n+1}, \ldots, \lambda_{n,n+1}$ with probability $1-\delta$ based on sets of adopting agents, it is sufficient for the number of i.i.d. products to be $O(n^2 \log n)$, and necessary for the number of i.i.d. products to be $\Omega(n^2 \log n)$.*

Proof. For the upper bounds, we consider the following estimator: after k products, decide $\hat{\lambda}_{1,n+1} = 0$ if and only if there exists a product such that agent 1 adopts and agent $n+1$ does not adopt (and decide $\hat{\lambda}_{1,n+1} = \infty$ otherwise). The worst-case error for such estimator will be attained in the case when all of agents $2, \ldots, n$ have links of rate infinity to agent $n+1$. We write the worst-case probability of error, which was calculated in the proof of Proposition 9:

$$\mathbb{P}\left(\hat{\lambda}_{1,n+1} = \infty \mid \lambda_{1,n+1} = 0\right) = \left(\frac{n(n+1)-1}{n(n+1)}\right)^k$$

To ensure accurate estimates with probability at least $1-\delta$, for given $\delta \in (0,1)$, it suffices that $k \geq \frac{\log \frac{1}{\delta}}{\log \frac{n(n+1)}{n(n+1)-1}} = O(n^2)$. The result for learning all of $\lambda_1, \ldots, \lambda_n$ follows from a union bound.

For the lower bound, we mirror the proof of Proposition 9, noticing that the smallest $\mathbb{P}\left(X_i \in \mathrm{BAD}_i^{\complement}\right)$ can be is $\frac{1}{n(n+1)}$, which occurs in the case when all of agents $2, \ldots, n$ have infinite influence rate on agent $n+1$. The union bound then establishes the $\Omega(n^2)$ lower bound for the number of products k. Finally, a coupon collector's argument establishes the $\Omega(n^2 \log n)$ lower bound for learning all of $\lambda_{1,n+1}, \ldots, \lambda_{n,n+1}$.

□

Proposition 14. *To ensure correct learning of $\lambda_{1,n+1}$ based on sequences of adoptions, it is sufficient for the number of i.i.d. products to be $O(n)$, and necessary for the number of i.i.d. products to be $\Omega(n)$. To ensure correct learning of $\lambda_{1,n+1}, \ldots, \lambda_{n,n+1}$ with probability $1-\delta$ based on sequences of adoptions, it is sufficient for the number of i.i.d. products to be $O(n \log n)$, and necessary for the number of i.i.d. products to be $\Omega(n \log n)$.*

The proofs rely on arguments already presented in the corresponding propositions in Subsection 7.2.1 and are omitted.

7.2.3 The Worst-Case Setting with Known Scaling of Agents with Influence Rate Infinity to $n+1$

We denote the number of agents with influence rate infinity to agent $n+1$ by ℓ. Table 7.4 summarizes the results on the necessary and sufficient number of i.i.d. products for learning.

Table 7.4: Matching lower and upper bounds for the minimum number of i.i.d. products required to learn the influence model in terms of n, in the worst-case setting when the scaling of agents ℓ with infinite influence rate to agent $n+1$ is known, for the two cases of learning based on sets of adoptions or sequences of adoptions.

	Sets	Sequences
$\ell = 1$	$\Theta(\log n)$	$\Theta(1)$
$\ell = \alpha n, \alpha \in (0,1)$	$\Theta(n^2 \log n)$	$\Theta(n \log n)$
$\ell = n-1$	$\Theta(n^2)$	$\Theta(n)$

We set to prove the results in the table.

The Case $\ell = 1$.

Proposition 15. *To ensure correct learning of $\lambda_{1,n+1}, \ldots, \lambda_{n,n+1}$ with sets of adoptions, it is sufficient for the number of i.i.d. products to be $O(\log n)$, and necessary for the number of i.i.d. products to be $\Omega(\log n)$.*

Proof. We use AD_i to denote the set of agents in $1, \ldots, n$ who adopt product i along with agent $n+1$; we define $\mathrm{AD}_i = \emptyset$ if agent $n+1$ does not adopt product $n+1$. Define $S_k = \bigcap_{i \in \{1,\ldots,k\}: \mathrm{AD}_i \neq \emptyset} \mathrm{AD}_i$. For learning, we need S_k to be a singleton. And if S_k is a singleton, then learning occurs.

We use j^* to denote the unique agent with influence rate infinity to agent $n+1$. Given a product, we define A to be the event that j^* adopts. Given a product, we denote by X the random variable which is the number of agents in $\{1, \ldots, j^*-1, j^*+1, \ldots, n\}$ who adopt. We calculate the probability mass function of X, conditioned on the event A:

$$\begin{aligned}
p_{X|A}(m) &= \frac{\mathbb{P}\left(\{X = m\} \cap A\right)}{\mathbb{P}(A)} \\
&= \frac{\binom{n-1}{m} \cdot \frac{(m+1)\lambda}{n\lambda+\lambda} \cdot \frac{m\lambda}{(n-1)\lambda+\lambda} \cdots \frac{\lambda}{(n-m)\lambda+\lambda} \cdot \frac{\lambda}{(n-(m+1))\lambda+\lambda}}{1/2} \\
&= \frac{\frac{(n-1)!}{m!(n-m-1)!} \cdot \frac{m+1}{n+1} \cdot \frac{m}{n} \cdot \frac{m-1}{n-1} \cdots \frac{1}{n-m+1} \cdot \frac{1}{n-m}}{1/2}
\end{aligned}$$

$$= \frac{2(m+1)}{n(n+1)}, \quad m = 0, \ldots, n-1.$$

We now calculate the conditional expectation of X:

$$\begin{aligned}
\mathbb{E}[X \mid A] &= \sum_{m=0}^{n-1} m \frac{2(m+1)}{n(n+1)} \\
&= \frac{2}{n(n+1)} \sum_{m=0}^{n-1} m(m+1) \\
&= \frac{2}{n(n+1)} \left(\frac{1}{3}(n-1)^3 + \frac{1}{2}(n-1)^2 + \frac{1}{6}(n-1) + \frac{1}{2}(n-1)^2 + \frac{1}{2}(n-1) \right) \\
&= \frac{2}{n(n+1)} \left(\frac{1}{3}(n-1)^3 + (n-1)^2 + \frac{2}{3}(n-1) \right) \\
&= \frac{2}{n(n+1)} (n-1) \frac{1}{3} \left((n-1)^2 + 3(n-1) + 2 \right) \\
&= \frac{2}{n(n+1)} (n-1) \frac{1}{3} (n^2 + n) \\
&= \frac{2}{3}(n-1)
\end{aligned}$$

Therefore, in expectation half the products will be adopted by j^*, and for the products that are adopted by j^*, the size of AD_i is $\frac{2}{3}(n-1)+1 = \frac{2}{3}n + \frac{1}{3}$ in expectation. It follows that the required number of i.i.d. products to make the intersection of the AD_i's a singleton grows as $\Theta(\log n)$. □

Proposition 16. *To ensure correct learning of $\lambda_{1,n+1}, \ldots, \lambda_{n,n+1}$ with sequences of adoptions, it is sufficient for the number of i.i.d. products to be $O(1)$, and necessary for the number of i.i.d. products to be $\Omega(1)$.*

Proof. If the unique agent with infinite influence rate to agent $n+1$ adopts once, then learning occurs. For each i.i.d. product, the probability of her adopting is $\frac{\lambda}{2\lambda} = \frac{1}{2}$. □

The Case $\ell = \alpha n, \alpha \in (0,1)$.

Proposition 17. *To ensure correct learning of $\lambda_{1,n+1}, \ldots, \lambda_{n,n+1}$ with sets of adoptions, it is sufficient for the number of i.i.d. products to be $O(n^2 \log n)$, and necessary for the number of i.i.d. products to be $\Omega(n^2 \log n)$.*

Proof. Define A to be the event that every agent out of the agents with influence rate infinity to agent $n+1$ is the only one to adopt some product among agents with influence rate infinity to agent $n+1$. For learning, we need event A to occur; and if event A occurs, then we have learning. A coupon collector's argument establishes that the required number of i.i.d. products to obtain event A grows as $\Theta(n^2 \log n)$. □

Proposition 18. *To ensure correct learning of $\lambda_{1,n+1}, \ldots, \lambda_{n,n+1}$ with sequences of adoptions, it is sufficient for the number of i.i.d. products to be $O(n \log n)$, and necessary for the number of i.i.d. products to be $\Omega(n \log n)$.*

Proof. Define A to be the event that each of the agents with influence rate infinity to agent $n+1$ adopts a product first among agents with influence rate infinity to agent $n+1$. For learning, we need event A to occur; and if A occurs, then we have learning. A coupon collector's argument establishes that the required number of i.i.d. products to obtain event A grows as $\Theta(n \log n)$. □

The Case $\ell = n - 1$.

Proposition 19. *To ensure correct learning of $\lambda_{1,n+1}, \ldots, \lambda_{n,n+1}$ with sets of adoptions, it is sufficient for the number of i.i.d. products to be $O(n^2)$, and necessary for the number of i.i.d. products to be $\Omega(n^2)$.*

Proof. For learning, the unique agent with influence rate zero to agent $n+1$ needs to be the only one to adopt for some product; and if she is the only to adopt for some product, then learning occurs. Given a product, the probability of her being the only one to adopt is

$$\frac{\lambda}{n\lambda + \lambda} \cdot \frac{\lambda}{(n-1)\lambda + \lambda} = \frac{1}{n(n+1)},$$

implying that the minimum required number of i.i.d. products for learning grows as $\Theta(n^2)$. □

Proposition 20. *To ensure correct learning of $\lambda_{1,n+1}, \ldots, \lambda_{n,n+1}$ with sequences of adoptions, it is sufficient for the number of i.i.d. products to be $O(n)$, and necessary for the number of i.i.d. products to be $\Omega(n)$.*

Proof. For learning, the unique agent with influence rate zero to agent $n+1$ needs to adopt first once, and if she adopts first once, then learning occurs. For each i.i.d. product, the probability of her adopting first is $\frac{\lambda}{n\lambda+\lambda} = \frac{1}{n+1}$. □

7.3 Learning Influence in the Star Network with Small Horizon

In this section, we consider the same hypothesis testing setting for the star network as in Section 7.2, though assuming a small horizon: the rate of the horizon satisfies $\lambda_{hor} = n\lambda$, where n is the number of agents who adopt with rate λ.

As in Section 7.2, we pose two decision problems: learning the influence rate between a specified agent i and agent $n+1$; and learning all the influence rates $\lambda_{1,n+1}, \ldots, \lambda_{n,n+1}$. In each case, our goal is to come up with upper and lower bounds for the minimum number of i.i.d. products required to learn the correct hypothesis. We study this question in the Bayesian setting where we assume a prior on the hypotheses, in the worst case over the hypotheses, as well as in the setting in which we know how many agents have infinite influence rate and how many have zero influence. We characterize the growth of the minimum number of i.i.d. products required for learning with respect to n, both when the available data provides information on sets of adopters, and when the available data provides information on sequences of adopters. (Of course, knowledge of times of adoptions would not induce a gain over knowledge of sequences, because of our assumption that the influence rates are either zero or infinite, and $\lambda_{n+1} = 0$.)

For the results that follow, the arguments in the proofs mirror the corresponding arguments for the case $\lambda_{hor} = \lambda$ in Section 7.2, and the proofs are thus omitted.

7.3.1 The Bayesian Setting

The Case $p = 1/2$. We assume that the influence rate on each link will be zero or infinite with equal probability. Table 7.5 summarizes the results on the necessary

and sufficient number of i.i.d. products for learning, both for the case $\lambda_{hor} = \lambda$, and for the case $\lambda_{hor} = n\lambda$.

Table 7.5: Matching lower and upper bounds for the minimum number of i.i.d. products required to learn the influence model in terms of n, in the Bayesian setting when $p = 1/2$, for the two cases of learning the influence between one agent and the star agent and of learning the influence between all agents and the star agent, and for the two cases of learning based on sets of adoptions or sequences of adoptions.

	$\lambda_{hor} = \lambda$		$\lambda_{hor} = n\lambda$	
	Sets	Sequences	Sets	Sequences
Learn one	$\Theta(n^2)$	$\Theta(n)$	$\Theta(n)$	$\Theta(n)$
Learn all	$\Theta(n^2 \log n)$	$\Theta(n \log n)$	$\Theta(n \log n)$	$\Theta(n \log n)$

Proposition 21. *To ensure correct learning of $\lambda_{1,n+1}$ with probability $1 - \delta$ based on sets of adopting agents, it is sufficient for the number of i.i.d. products to be $O(n)$, and necessary for the number of i.i.d. products to be $\Omega(n)$. To ensure correct learning of $\lambda_{1,n+1}, \ldots, \lambda_{n,n+1}$ with probability $1 - \delta$ based on sets of adopting agents, it is sufficient for the number of i.i.d. products to be $O(n \log n)$, and necessary for the number of i.i.d. products to be $\Omega(n \log n)$.*

Proposition 22. *To ensure correct learning of $\lambda_{1,n+1}$ based on sequences of adoptions, it is sufficient for the number of i.i.d. products to be $O(n)$, and necessary for the number of i.i.d. products to be $\Omega(n)$. To ensure correct learning of $\lambda_{1,n+1}, \ldots, \lambda_{n,n+1}$ with probability $1 - \delta$ based on sequences of adoptions, it is sufficient for the number of i.i.d. products to be $O(n \log n)$, and necessary for the number of i.i.d. products to be $\Omega(n \log n)$.*

The Case $p = 1/n$. We assume that the influence rate on each link will be infinite with probability $p = 1/n$. (In this case, the expected number of agents who can influence agent $n+1$ is $\Theta(1)$.) Table 7.6 summarizes the results on the necessary and

sufficient number of i.i.d. products for learning, both for the case $\lambda_{hor} = \lambda$, and for the case $\lambda_{hor} = n\lambda$.

Table 7.6: Matching lower and upper bounds for the minimum number of i.i.d. products required to learn the influence model, in terms of n, in the Bayesian setting when $p = 1/n$, when learning the influence between all agents and the star agent, for the two cases of learning based on sets of adoptions or sequences of adoptions.

	$\lambda_{hor} = \lambda$		$\lambda_{hor} = n\lambda$	
	Sets	Sequences	Sets	Sequences
Learn all	$\Theta(\log n)$	$\Theta(1)$	$\Theta(n)$	$\Theta(n)$

Proposition 23. *To ensure correct learning of $\lambda_{1,n+1}, \ldots, \lambda_{n,n+1}$ with probability $1-\delta$ based on sets of adopting agents, it is sufficient for the number of i.i.d. products to be $O(n)$, and necessary for the number of i.i.d. products to be $\Omega(n)$.*

Proposition 24. *To ensure correct learning of $\lambda_{1,n+1}, \ldots, \lambda_{n,n+1}$ with probability $1-\delta$ based on sequences of adoptions, it is sufficient for the number of i.i.d. products to be $O(n)$, and necessary for the number of i.i.d. products to be $\Omega(n)$.*

7.3.2 The Worst-Case Setting

Table 7.7 summarizes the results on the necessary and sufficient number of i.i.d. products for learning, both for the case $\lambda_{hor} = \lambda$, and for the case $\lambda_{hor} = n\lambda$.

Table 7.7: Matching lower and upper bounds for the minimum number of i.i.d. products required to learn the influence model, in terms of n, in the worst-case setting, for the two cases of learning the influence between one agent and the star agent and of learning the influence between all agents and the star agent, and for the two cases of learning based on sets of adoptions or sequences of adoptions.

	$\lambda_{hor} = \lambda$		$\lambda_{hor} = n\lambda$	
	Sets	Sequences	Sets	Sequences
Learn one	$\Theta(n^2)$	$\Theta(n)$	$\Theta(n)$	$\Theta(n)$
Learn all	$\Theta(n^2 \log n)$	$\Theta(n \log n)$	$\Theta(n \log n)$	$\Theta(n \log n)$

Proposition 25. *To ensure correct learning of $\lambda_{1,n+1}$ with probability $1-\delta$ based on sets of adopting agents, it is sufficient for the number of i.i.d. products to be $O(n)$, and necessary for the number of i.i.d. products to be $\Omega(n)$. To ensure correct learning of $\lambda_{1,n+1}, \ldots, \lambda_{n,n+1}$ with probability $1-\delta$ based on sets of adopting agents, it is sufficient for the number of i.i.d. products to be $O(n \log n)$, and necessary for the number of i.i.d. products to be $\Omega(n \log n)$.*

Proposition 26. *To ensure correct learning of $\lambda_{1,n+1}$ based on sequences of adoptions, it is sufficient for the number of i.i.d. products to be $O(n)$, and necessary for the number of i.i.d. products to be $\Omega(n)$. To ensure correct learning of $\lambda_{1,n+1}, \ldots, \lambda_{n,n+1}$ with probability $1-\delta$ based on sequences of adoptions, it is sufficient for the number of i.i.d. products to be $O(n \log n)$, and necessary for the number of i.i.d. products to be $\Omega(n \log n)$.*

7.3.3 The Worst-Case Setting with Known Scaling of Agents with Influence Rate Infinity to $n+1$

Table 7.8 summarizes the results on the necessary and sufficient number of i.i.d. products for learning, both for the case $\lambda_{hor} = \lambda$, and for the case $\lambda_{hor} = n\lambda$.

Table 7.8: Matching lower and upper bounds for the minimum number of i.i.d. products required to learn the influence model, in terms of n, in the worst-case setting when the scaling of agents ℓ with influence rate infinity to agent $n+1$ is known, for the two cases of learning based on sets of adoptions or sequences of adoptions.

	$\lambda_{hor} = \lambda$		$\lambda_{hor} = n\lambda$	
	Sets	Sequences	Sets	Sequences
$\ell = 1$	$\Theta(\log n)$	$\Theta(1)$	$\Theta(n)$	$\Theta(n)$
$\ell = \alpha n, \alpha \in (0,1)$	$\Theta(n^2 \log n)$	$\Theta(n \log n)$	$\Theta(n \log n)$	$\Theta(n \log n)$
$\ell = n-1$	$\Theta(n^2)$	$\Theta(n)$	$\Theta(n)$	$\Theta(n)$

The Case $\ell = 1$.

Proposition 27. *To ensure correct learning of $\lambda_{1,n+1}, \ldots, \lambda_{n,n+1}$ with sets of adoptions, it is sufficient for the number of i.i.d. products to be $O(n)$, and necessary for the number of i.i.d. products to be $\Omega(n)$.*

Proposition 28. *To ensure correct learning of $\lambda_{1,n+1}, \ldots, \lambda_{n,n+1}$ with sequences of adoptions, it is sufficient for the number of i.i.d. products to be $O(n)$, and necessary for the number of i.i.d. products to be $\Omega(n)$.*

The Case $\ell = \alpha n, \alpha \in (0,1)$.

Proposition 29. *To ensure correct learning of $\lambda_{1,n+1}, \ldots, \lambda_{n,n+1}$ with sets of adoptions, it is sufficient for the number of i.i.d. products to be $O(n \log n)$, and necessary for the number of i.i.d. products to be $\Omega(n \log n)$.*

Proposition 30. *To ensure correct learning of $\lambda_{1,n+1}, \ldots, \lambda_{n,n+1}$ with sequences of adoptions, it is sufficient for the number of i.i.d. products to be $O(n \log n)$, and necessary for the number of i.i.d. products to be $\Omega(n \log n)$.*

The Case $\ell = n - 1$.

Proposition 31. *To ensure correct learning of $\lambda_{1,n+1}, \ldots, \lambda_{n,n+1}$ with sets of adoptions, it is sufficient for the number of i.i.d. products to be $O(n)$, and necessary for the number of i.i.d. products to be $\Omega(n)$.*

Proposition 32. *To ensure correct learning of $\lambda_{1,n+1}, \ldots, \lambda_{n,n+1}$ with sequences of adoptions, it is sufficient for the number of i.i.d. products to be $O(n)$, and necessary for the number of i.i.d. products to be $\Omega(n)$.*

7.4 Discussion

We characterize the scaling of the number of samples required for learning with sets and sequences, thus theoretically quantifying the gain of learning with sequences over learning with sets in regard to the speed of learning. Our inference algorithms look for signature events, and attain optimal sample complexity, as long as the signature events are reasonably chosen. Depending on the setting, learning with sets can take a multiplicative factor of $\Theta(n)$ more samples than learning with sequences, when the horizon rate is moderate (i.e., as large as the idiosyncratic rates of adoption). With much smaller horizon, learning with sequences has no gain asymptotically over learning with mere sets, across all the settings we study; when the observation window (i.e., the horizon) is small, then the sets of adoptions provide asymptotically all the information pertinent to learning that sequences provide. This insight is in agreement with our findings for the value of richer temporal information over mere sets in the small horizon regime, when using the Kullback-Leibler divergence as a measure for the speed of learning, as discussed in Chapter 6.

Chapter 8

Theoretical Guarantees for General Networks

In this chapter, we provide theoretical guarantees for more general networks than the star topology, which was considered in Chapter 7, and we relax the assumption that each edge carries an influence rate of either zero or infinity. In particular, we focus on the question of deciding between the complete graph, and the complete graph that is missing one directed edge, which we cast as a binary hypothesis testing problem[1]. Because of its nature, sample complexity results for this hard problem entail sample complexity results for broader families of networks.

8.1 Learning Between the Complete Graph and the Complete Graph that Is Missing One Edge

For ease of exposition, we assume that all from a collection of n agents have the same idiosyncratic rate $\lambda > 0$, and that all directed edges carry the same influence rate, which is equal to λ. λ can be known or unknown. We are learning between two hypotheses for the underlying influence graph:

- The complete graph, P_1;

[1] Abrahao, Chierichetti, Kleinberg, and Panconesi (2013) study the same binary hypothesis testing problem.

- The complete graph minus the directed edge (i, j), P_2.

8.1.1 An Algorithm for Learning

We propose a simple algorithm for deciding between the two hypotheses. The sample complexity of our algorithm gives an upper bound for the number of i.i.d. products required for learning. The algorithm is the following:

- We first choose an event of interest A in an appropriate manner.
- For each new product ℓ, we define an indicator variable

$$I_\ell = \begin{cases} 1 & \text{if event } A \text{ obtained in product } \ell \\ 0 & \text{otherwise} \end{cases}$$

- After k i.i.d. products, we compute $\hat{p} = \frac{1}{k}\sum_{\ell=1}^{k} I_k$.
- Choose the hypothesis with the smallest deviation from the empirical probability, $|P_i(A) - \hat{p}|, i = 1, 2$.

By the concentration inequality

$$\mathbb{P}\left(|\hat{p} - \mathbb{E}[I]| \geq t\right) \leq 2e^{-2kt^2},$$

and setting $2e^{-2kt^2} \leq \delta$, $0 < \delta < 1$, we obtain

$$k \geq \frac{\log\left(\frac{2}{\delta}\right)}{2t^2}.$$

Therefore, setting $t = 0.5 \cdot |\mathbb{E}_1[I] - \mathbb{E}_2[I]| = 0.5 \cdot |P_1(A) - P_2(A)|$, the proposed algorithm learns the true hypothesis correctly with probability at least $1 - \delta$.

The sample complexity of the proposed learning algorithm is given effectively by the inverse square of the distance $|\mathbb{E}_1[I] - \mathbb{E}_2[I]| = |P_1(A) - P_2(A)|$, which scales with the number of agents n.

An alternative derivation of an upper bound is through obtaining a lower bound on the Kullback-Leibler divergence between the two distributions P_1, P_2. In particular, in a Neyman-Pearson setting, the best achievable exponent for the probability of error of deciding in favor of the first hypothesis when the second is true, given that the probability of deciding in favor of the second hypothesis when the first is true is less than ϵ, is given by the negative Kullback-Leibler (KL) divergence, $-D(P_1||P_2)$. In turn, Pinsker's inequality bounds the KL divergence from below:

$$\begin{aligned} D(P_1||P_2) &\geq \frac{1}{2\log 2}||P_1 - P_2||_1^2 \\ &= \frac{1}{2\log 2}\Big(2\big(P_1(B) - P_2(B)\big)\Big)^2, \end{aligned}$$

where $B = \{x : P_1(x) > P_2(x)\}$. Therefore, the larger is the event A of interest in the algorithm proposed above, i.e., the closer it gets to the event B, the tighter upper bound we achieve for the number of i.i.d. products required for learning.

8.1.2 Learning with Sequences

Proposition 33. *To ensure correct learning of the true hypothesis with sequences, it is sufficient for the number of i.i.d. products to be $O(n^2)$.*

Proof. We focus on the event that both i and j adopt, with i adopting before j. We compute the probability for all the cases in this event, under each of the two hypotheses.

We have

$$\mathbb{P}(i \text{ first}, j \text{ second} \mid i \to j) = \frac{\lambda}{n\lambda + \lambda} \cdot \frac{2\lambda}{(n-1)\lambda + (n-1)\lambda + \lambda} = \frac{1}{n+1} \cdot \frac{2}{2n-1},$$

while

$$\mathbb{P}(i \text{ first}, j \text{ second} \mid i \nrightarrow j) = \frac{\lambda}{n\lambda + \lambda} \cdot \frac{\lambda}{(n-1)\lambda + (n-2)\lambda + \lambda} = \frac{1}{n+1} \cdot \frac{1}{2n-2},$$

and thus the difference of the two is

$$\frac{1}{n+1}\left(\frac{2}{2n-1}-\frac{1}{2n-2}\right)=\frac{1}{n+1}\cdot\frac{2n-3}{(2n-1)(2n-2)}\sim\frac{1}{2n^2}.$$

Similarly,

$$\begin{aligned}\mathbb{P}(i \text{ first}, j \text{ third} \mid i\to j) &= \frac{\lambda}{n\lambda+\lambda}\cdot\frac{(n-2)2\lambda}{(n-1)2\lambda+\lambda}\cdot\frac{3\lambda}{(n-2)\lambda+(n-2)2\lambda+\lambda}\\ &= \frac{1}{n+1}\cdot\frac{2n-4}{2n-1}\cdot\frac{3}{3n-5},\end{aligned}$$

while

$$\begin{aligned}\mathbb{P}(i \text{ first}, j \text{ third} \mid i\nrightarrow j) &= \frac{\lambda}{n\lambda+\lambda}\cdot\frac{(n-2)2\lambda}{(n-2)2\lambda+\lambda+\lambda}\cdot\frac{2\lambda}{(n-2)\lambda+(n-3)2\lambda+\lambda+\lambda}\\ &= \frac{1}{n+1}\cdot\frac{2n-4}{2n-2}\cdot\frac{2}{3n-6},\end{aligned}$$

with a difference of

$$\begin{aligned}\frac{2n-4}{n+1}\left(\frac{3}{(2n-1)(3n-5)}-\frac{2}{(2n-2)(3n-6)}\right) &= \frac{2n-4}{n+1}\cdot\frac{6n^2-28n+26}{(2n-1)(3n-5)(2n-2)(3n-6)}\\ &\sim \frac{1}{3n^2}.\end{aligned}$$

Similarly, we can show that

$$\mathbb{P}(i \text{ first}, j \text{ fourth} \mid i\to j)-\mathbb{P}(i \text{ first}, j \text{ fourth} \mid i\nrightarrow j)=\frac{1}{4n^2},$$

and in general, for $2\le\ell\le n$

$$\mathbb{P}(i \text{ first}, j\ \ell\text{th} \mid i\to j)-\mathbb{P}(i \text{ first}, j\ \ell\text{th} \mid i\nrightarrow j)=\frac{1}{\ell n^2}.$$

We now focus on the events in which i adopts second. We have

$$\mathbb{P}(i \text{ second}, j \text{ third} \mid i\to j)=\frac{(n-2)\lambda}{n\lambda+\lambda}\cdot\frac{2\lambda}{(n-1)2\lambda+\lambda}\cdot\frac{3\lambda}{(n-2)3\lambda+\lambda}=\frac{n-2}{n+1}\cdot\frac{2}{2n-1}\cdot\frac{3}{3n-5},$$

while

$$\mathbb{P}(i \text{ second}, j \text{ third} \mid i \nrightarrow j) = \frac{(n-2)\lambda}{n\lambda+\lambda} \cdot \frac{2\lambda}{(n-1)2\lambda+\lambda} \cdot \frac{2\lambda}{(n-3)3\lambda+2\lambda+\lambda} = \frac{n-2}{n+1} \cdot \frac{2}{2n-1} \cdot \frac{2}{3n-6},$$

and thus the difference of the two is

$$\frac{n-2}{n+1} \cdot \frac{2}{2n-1} \left(\frac{3}{3n-5} - \frac{2}{3n-6} \right) = \frac{n-2}{n+1} \cdot \frac{2}{2n-1} \cdot \frac{3n-8}{(3n-5)(3n-6)} \sim \frac{1}{3n^2}.$$

In general, for $3 \leq \ell \leq n$, the difference is

$$\mathbb{P}(i \text{ second}, j\ \ell\text{th} \mid i \to j) - \mathbb{P}(i \text{ second}, j\ \ell\text{th} \mid i \nrightarrow j) = \frac{1}{\ell n^2}.$$

We can sum up all the differences between the two hypotheses for the events in which both i and j adopt with i adopting before j, to get asymptotically

$$\begin{aligned}
\frac{1}{n^2} \cdot \Big[\frac{1}{2} + \frac{1}{3} + \frac{1}{4} + \ldots\ldots\ldots\ldots\ldots + \frac{1}{n} \\
+ \frac{1}{3} + \frac{1}{4} + \ldots\ldots\ldots\ldots\ldots + \frac{1}{n} \\
+ \frac{1}{4} + \ldots\ldots\ldots\ldots\ldots + \frac{1}{n} \\
\vdots \\
\vdots \\
+ \frac{1}{n} \Big]
\end{aligned}$$

which can be written as

$$\begin{aligned}
\frac{1}{n^2} \left(\frac{1}{2} + \frac{2}{3} + \frac{3}{4} + \ldots + \frac{n-1}{n} \right) &= \frac{1}{n^2} \left((1 - \frac{1}{2}) + (1 - \frac{1}{3}) + (1 - \frac{1}{4}) + \ldots + (1 - \frac{1}{n}) \right) \\
&\sim \frac{1}{n^2} \left(n - 1 - (\log n - 1) \right) \\
&= \frac{n - \log n}{n^2}.
\end{aligned}$$

The sample complexity is therefore

$$\frac{1}{\left(\frac{n-\log n}{n^2}\right)^2} = \frac{n^4}{(n-\log n)^2} = \Theta(n^2).$$

□

We now argue that there is a matching lower bound of $\Omega(n^2)$ for learning with sequences. Indeed, assuming we are learning based on not just sequences, bur rather times of adoptions, then the ratio of the likelihoods for the time of adoption for agent j between hypotheses P_1 and P_2, assuming everybody else has adopted, is

$$\frac{n\lambda e^{-n\lambda t}}{(n-1)\lambda e^{-(n-1)\lambda t}} = \left(1 + \frac{1}{n-1}\right) e^{-\lambda t},$$

resulting in a KL divergence of $\Theta\left(\frac{1}{n^2}\right)$, which in turn implies a $\Omega(n^2)$ complexity for the number of i.i.d. products required for learning. This is in agreement with the $\Omega(\frac{n^2}{\log^2 n})$ lower bound proven in Abrahao, Chierichetti, Kleinberg, and Panconesi (2013), for a model with exponential infection times, but no idiosyncratic adoptions.

8.1.3 Learning with Sets

Proposition 34. *To ensure correct learning of the true hypothesis with sets, it is sufficient for the number of i.i.d. products to be $O(n^6)$.*

Proof. We focus on the event that only i, j adopt. We have

$$\begin{aligned}
\mathbb{P}(\text{only } i,j \text{ adopt} \mid i \to j) &= 2\left(\frac{\lambda}{n\lambda+\lambda} \cdot \frac{2\lambda}{(n-1)\lambda+(n-1)\lambda+\lambda}\right) \frac{\lambda}{(n-2)\lambda+(n-2)2\lambda+\lambda} \\
&= 2 \cdot \frac{2}{(n+1)(2n-1)} \cdot \frac{1}{3n-5},
\end{aligned}$$

while

$$\begin{aligned}
\mathbb{P}(\text{only } i,j \text{ adopt} \mid i \nrightarrow j) &= \left(\frac{\lambda}{n\lambda+\lambda} \cdot \frac{\lambda}{(n-1)\lambda+(n-2)\lambda+\lambda}\right. \\
&\qquad \left. + \frac{\lambda}{n\lambda+\lambda} \cdot \frac{2\lambda}{(n-1)\lambda+(n-1)\lambda+\lambda}\right) \frac{\lambda}{(n-2)\lambda+(n-2)2\lambda+\lambda}
\end{aligned}$$

$$= \frac{1}{n+1}\left(\frac{1}{2n-2}+\frac{2}{2n-1}\right)\frac{1}{3n-5},$$

and thus the difference of the two is

$$\frac{1}{n+1}\cdot\frac{1}{3n-5}\left(\frac{4}{2n-1}-\frac{1}{2n-2}-\frac{2}{2n-1}\right)=\frac{1}{n+1}\cdot\frac{1}{3n-5}\cdot\frac{2n-3}{(2n-1)(2n-2)}\sim\frac{1}{6n^3}.$$

The sample complexity is therefore $\Theta(n^6)$. □

8.2 Discussion

We have proposed a simple algorithm for deciding between the complete graph (with all directed edges present), and the complete graph that is missing one directed edge. The algorithm relies on using samples to estimate the probability of an event of interest under each of the two hypotheses. Our algorithm results in sample complexity that is given by the inverse of the square of the difference between the probabilities of the chosen event of interest under each of the two hypotheses. This difference scales with the number of agents n. The dependence on the inverse of the square of the difference can also be derived from a lower bound on the Kullback-Leibler divergence via Pinsker's inequality. One would choose the event of interest in the algorithm so as to maximize the difference between the event's probabilities under each of the two hypotheses, resulting in a tighter upper bound.

When learning with sequences, we propose an implementation of our algorithm that can learn with $O(n^2)$ samples, and we argue that learning is not possible with $o(n^2)$ samples. When learning with sets, we propose an implementation of our algorithm that can learn with $O(n^6)$ samples, and although we are missing a lower bound, we conjecture that $O(n^2)$ samples do not suffice for correct learning with sets.

Chapter 9

Learning Influence with Synthetic and Observational Data

Given observational data on adoptions, we estimate the idiosyncratic rates of adoption as well as the influence rates (i.e., the network structure) using maximum likelihood (ML) estimation. We consider three cases of different temporal detail in the available data (sets, sequences, times), which correspond to different calculations of the log likelihood of the data. We do not restrict the λ_{ij}'s to be zero or infinite.

We evaluate our methodology on both synthetic and real data, showing that it recovers the network structure well, and quantifying the improvement in learning due to the availability of richer temporal data. The real data come from (i) observations of mobile app installations of users, along with data on their communications and social relations; (ii) observations of epileptic seizure events in patients, including levels of neuronal activity in different regions of the brain.

9.1 Synthetic Data

We generate a directed network according to the Erdős-Rényi model and assign some arbitrary influence rates λ_{ij}, idiosyncratic rates λ_i, and the horizon rate λ_{hor}. We generate data on adoptions of a number of products using the generative model of Section 5.3, and learn the parameters using maximum likelihood estimation for an

increasing sequence of sets of products.

Sets vs. Sequences vs. Times Figure 9-1 showcases the recovery of two different sparse network structures on a network of five agents. We plot the ℓ_1 estimation error, calculated on both the influence rates and the idiosyncratic rates, for different number of products, for the cases of learning with sets, sequences, and times of adoptions. The sparsity pattern is in general recovered more accurately with times than with

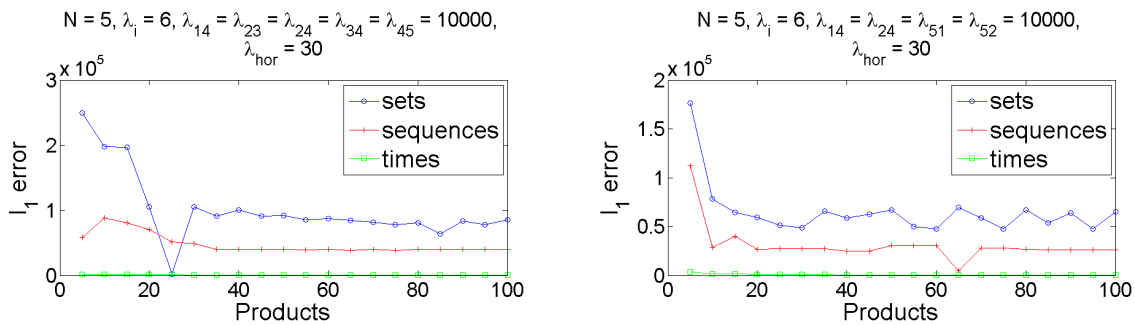


Figure 9-1: Learning a network of five agents with five high influence rates and all other influence rates zero (left) and four high influence rates and all other influence rates zero (right). We plot the ℓ_1 estimation error calculated on both the influence rates and the idiosyncratic rates. Learning is more accurate and faster with times than with sequences, and with sequences than with sets, with about an order of magnitude difference in the ℓ_1 error between sets and sequences, and between sequences and times.

sequences, and with sequences than with sets; learning with times actually gives accurate estimates of the network influence parameters themselves.

Sequences vs. Times Optimizing the log likelihood function when learning with sets is computationally heavy, as the likelihood term for a specific set is the sum over the likelihood terms for all possible permutations (i.e., orders) associated with the set. We therefore present experiments on networks of larger scale only for learning with sequences and times. Figure 9-2 showcases the recovery of sparse network structures on networks of 50 and 100 agents. We plot the ℓ_1 estimation error, calculated on both the influence rates and the idiosyncratic rates, for an increasing sequence of products, for the cases of learning with sequences and times of adoptions. Timing information

yields significantly better learning.

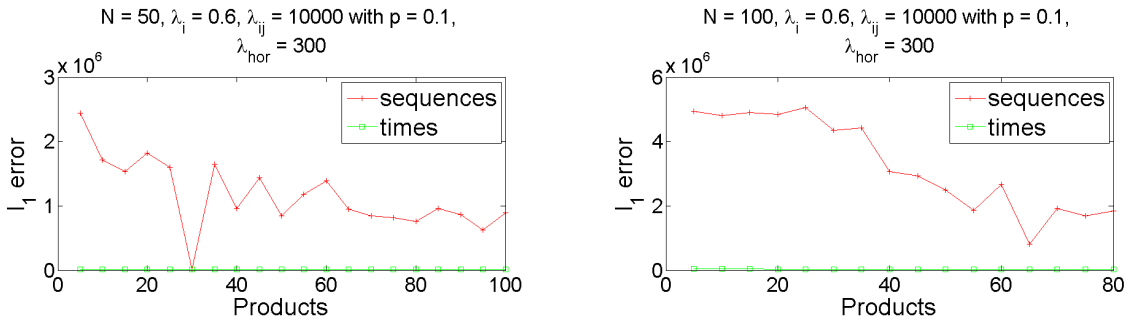


Figure 9-2: Learning a network of 50 (left) and 100 (right) agents, generated from an Erdős-Rényi model with probability $p = 0.1$ that each directed edge is present. We plot the ℓ_1 estimation error calculated on both the influence rates and the idiosyncratic rates. Learning is more accurate and faster with times than with sequences, with a difference of one to two orders of magnitude in the ℓ_1 error between sequences and times.

Learning under Model Mismatch We have generated data on adoptions using the generative model of Section 5.3 but adding noise to the realizations of the times of adoptions; this affects also the sets and the sequences of adoptions. In particular, we multiply the realized time stamp of each event (i.e., of each adoption or the end of the horizon) by a multiplier drawn from the uniform distribution on $[0.5, 1.5]$, and then generate information on sequences of adoptions and sets of adoptions accordingly.

Figure 9-3 showcases the recovery of sparse network structures on a network of ten agents. We plot the ℓ_1 estimation error, calculated on both the influence rates and the idiosyncratic rates, for an increasing sequence of products, for the cases of learning with noisy sequences and noisy times of adoptions. Again, we are able to recover the underlying sparsity pattern well, with timing information yielding significantly better learning.

A Heuristic We propose a heuristic that makes estimation significantly faster for sets, sequences, and times: instead of estimating the collection of λ_i's and λ_{ij}'s, we estimate instead only one parameter for each agent, $\hat{\lambda}_i$, and find the parameters $\hat{\lambda}_i$ that maximize the likelihood of the observed adoptions, assuming that each agent i adopts

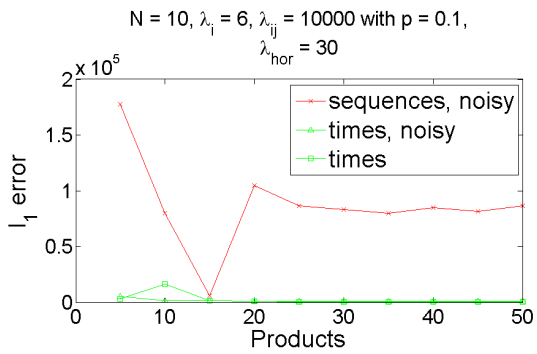


Figure 9-3: Learning a network of ten agents, generated from an Erdős-Rényi model with probability $p = 0.1$ that each directed edge is present. We plot the ℓ_1 estimation error calculated on both the influence rates and the idiosyncratic rates, when the data is noisy. Learning is more accurate and faster with noisy times than with noisy sequences, with a difference of two orders of magnitude in the ℓ_1 error between sequences and times.

at a time that is exponentially distributed with parameter $\hat{\lambda}_i$. The clear disadvantage of the proposed heuristic is that it does not estimate pairwise influences between agents. Table 9.1 shows the running time in seconds to estimate the parameters for a specific network, learning based on sets, sequences, and times of adoptions, with and without the heuristic. Although the proposed heuristic does not estimate the network, it can be used as a computationally cheap pre-processing stage for parameter initialization, prior to estimating the network.

Table 9.1: Running time in seconds for learning based on sets, sequences, and times of adoptions, with and without the heuristic, for a network of five agents, with five influence rates equal to 10000 and all other equal to zero, idiosyncratic rates equal to 6, and horizon rate 30. The experiments were run with MATLAB on a 2.53 GHz processor with 4 GB RAM.

	Running time in seconds		
	Without heuristic	With heuristic	Speed-up
Learn with sets	1535.3	178.0	**8.6**×
Learn with sequences	24.0	6.7	**3.6**×
Learning with times	65.0	6.2	**10.6**×

9.2 Real Data: Mobile App Installations

9.2.1 The Dataset

We use data[1] obtained from an experiment (Pan, Aharony, and Pentland, 2011) for which an Android mobile phone is given to each of 55 participants, all residents of a graduate student dormitory at a major US university campus, and the following information is tracked during the experimental period of four months:

- installations of mobile apps, along with associated time stamps;
- calls among users (number of calls for each pair of users);
- Bluetooth hits among users (number of Bluetooth radio hits for each pair of users);
- declared affiliation (in terms of academic department) and declared friendship among users (binary value denoting affiliation for each pair of users, and similarly for friendship).

9.2.2 Network Inference

We are interested in recovering patterns of influence among the 55 participants based solely on the mobile app installation data. Under the premise that influence travels through communication and social interaction, we expect a network inference algorithm that does well to recover a network of influence that is highly correlated to the realized network of social interaction. We therefore separate the available data into the mobile app installation data (i.e., the "actions") and the communication/interaction data (i.e., the social data). We learn the influence network using only the actions, and we then validate using the social data.

We employ ML estimation based on sequences of mobile app installations. The inferred influence rates are highly correlated with the realized communication net-

[1] We are thankful to Sandy Pentland and the Human Dynamics Laboratory at the MIT Media Lab for sharing the data with us.

works, providing evidence for the soundness of the proposed inference methodology, as we proceed to show.

For each edge (i, j), we add the inferred rates $\lambda_{ij} + \lambda_{ji}$, and rank all edges based on joint inferred influence. We then choose the top ranked edges based on joint influence, and we report the percentage of edges for which friendship was reported. A friendship edge exists between two randomly selected nodes with probability 0.3508, which is less than the percentage corresponding to the $10, 20, 50, 100$ edges that carry the highest inferred influence. Table 9.2 shows the results.

Table 9.2: There is higher probability of friendship in the edges where we detect influence using sequences. A communication edge exists between two randomly selected nodes in the dataset with probability 0.3508.

Out of top	10	joint influence edges, friendship exists in	70%
	20		65%
	50		54%
	100		38%

The correlation coefficient between the observations of calls and the inferred (joint) influence (using information on sequences of installations) per edge is 0.3381. The positive correlation between calls and (joint) influence inferred from sequences is visualized in Figure 9-4.

We finally also estimate influence based solely on sets of mobile app installations. We restrict our attention to the most active users out of the 55 participants, and learn influence among them using sets and sequences. When we learn influence based on sets of mobile app installations, as opposed to sequences, the correlation between the inferred rates and the realized communication/interaction networks is only slightly lower (or even higher) than when learning with sequences of mobile app installations. This is an instance where learning with temporally poor data modes is good enough.

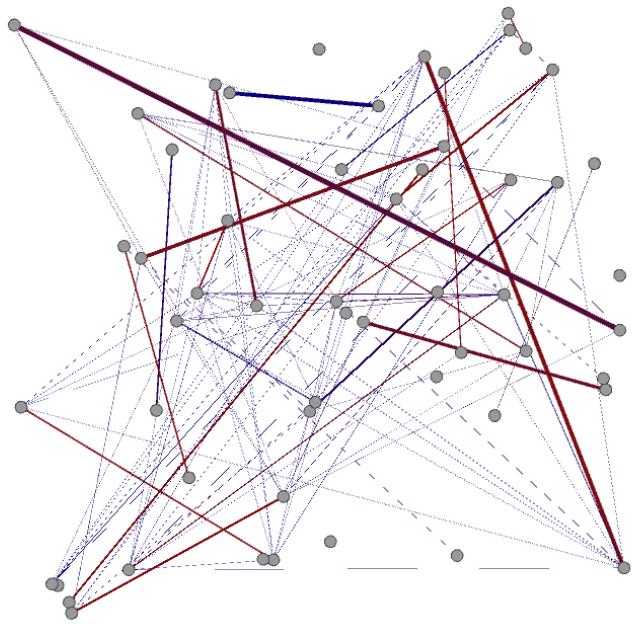


Figure 9-4: Concurrent visualization of the realized network of calls (the color of each edge denotes the number of calls between the users: closer to blue for lower number of calls, closer to red for higher number of calls) and the inferred network of influence using information on sequences of adoptions (the thickness of edge $i-j$ is proportional to the sum of the inferred influence rates $\lambda_{ij}+\lambda_{ji}$). We observe that edges with higher number of calls (red) are more likely to carry higher influence (thick).

9.3 Real Data: Epileptic Seizures

9.3.1 The Dataset

We use electroencephalography (EEG) data[2] obtained from epileptic patients. Between 10 and 15 electrodes are inserted along each subject's scalp; each electrode has 10 endings (channels), each of which is measuring voltage fluctuations resulting from local neuronal activity. Different electrodes, and possibly even different channels on the same electrode, monitor different anatomical regions of the brain. Each channel provides a voltage measurement per millisecond. We have access to EEG data for

[2]We are thankful to Sridevi Sarma and the Neuromedical Control Systems Lab at the Johns Hopkins Institute for Computational Medicine for sharing the data with us.

five epileptic patients, all of whom suffer from mesial temporal sclerosis[3].

For each subject, we have data from a number of seizure events[4] that ranges from three to six. For each separate seizure event, our data is the time-series voltage measurements at millisecond-resolution across all the channels from all the implanted electrodes. For each seizure event, the available data spans a window of 4-7 minutes that contains the time interval from the marked seizure initiation time through the marked seizure ending time.

9.3.2 Network Inference

Temporal lobe epilepsy is oftentimes amenable to temporal lobectomy, a surgical procedure of resection for which the part of the brain containing the point at which the seizures start is removed[5]. It is therefore crucial to identify the source of the seizures correctly for effective treatment. Furthermore, physicians are interested in understanding causal relations of influence between anatomical regions of the brain: how likely is it that a given region that is under seizure, infects another given region? A methodology that uses seizure cascades to infer idiosyncratic rates for seizure per brain region, as well as rates of seizure infection between pairs of brain regions, can provide answers for both the question of identifying the source of a seizure event, and

[3]According to Columbia Neurosurgeons (2014a), "the condition called mesial temporal sclerosis is closely related to temporal lobe epilepsy, a type of partial (focal) epilepsy in which the seizure initiation point can be identified within the temporal lobe of the brain. Mesial temporal sclerosis is the loss of neurons and scarring of the deepest portion of the temporal lobe and is associated with certain brain injuries.

Brain damage from traumatic injury, infection, a brain tumor, a lack of oxygen to the brain, or uncontrolled seizures is thought to cause the scar tissue to form, particularly in the hippocampus, a region of the temporal lobe. The region begins to atrophy; neurons die and scar tissue forms. This damage is thought to be a significant cause of temporal lobe epilepsy. In fact, 70 percent of temporal lobe epilepsy patients have some degree of mesial temporal sclerosis. It also appears that the mesial temporal sclerosis can be worsened by additional seizures.

Mesial temporal sclerosis usually results in partial (focal) epilepsy. This seizure disorder can cause a variety of symptoms such as strange sensations, changes in behavior or emotions, muscle spasms, or convulsions. The seizures usually are localized in the brain, but they may spread to become generalized seizures, which involve the entire brain and may cause a sudden loss of awareness or consciousness."

[4]Physicians mark the initiation and ending of a seizure event based on the EEG, combined with the synchronized video recording of the subject - a technique known as video-EEG monitoring.

[5]Temporal lobectomy has high success rates (60 percent of patients are seizure-free at one year, compared to 8 percent among patients given medication alone), but also side effects, which include loss of memory, visual disturbances, and emotional change (Columbia Neurosurgeons, 2014b).

for the question of identifying causal relations between regions.

The time marking of the seizures in each channel is done by physicians on an empirical basis; although software exists that marks a seizure on a channel based on the EEG time series of that channel, deciding upon the exact seizure starting time depends on the selected parameters of the marking algorithm. It follows that the time marking can be arbitrary and noisy. Nevertheless, temporally poorer modes of information (i.e., knowing the sequence in which channels/brain regions are affected for each seizure event, or only knowing the sets of brain regions that are affected for each seizure event) may be less noisy. This raises the question of whether we can recover causality patterns with just sequences or even sets.

We show three implementations of network inference, all based on ML estimation:

- unconstrained optimization of the likelihood;
- unconstrained optimization of the likelihood, regularized with an ℓ_1 term;
- constrained optimization of the likelihood over specific classes of graphs (trees).

Both regularizing with an ℓ_1 term, and constraining over trees, are approaches to learn a sparsified network model. In general, the advantage of learning a sparse model, over learning a dense model, is that learning a sparse model can help generate hypotheses about the system of interest which are easily understood; in other words, a sparse model can elucidate in a better way key properties of the system of interest.

We note that the network of influence among brain regions can be different across patients, therefore we cannot learn one model using the combined data from all patients. We infer the network and showcase results for one patient with four seizure events, whom we choose because the number of available seizure events is larger than for most other patients for whom we have data, and because the doctors have clearly annotated the order (and timing) of the spread of the seizure to different brain regions for some of the seizure events of that patient. For the selected patient, we indicate the sequence of affected regions for each seizure event:

Seizure 1 Hippocampus Head (HH), Hippocampus Tail (HT), Parahippocampus Gyrus/Fusiform Gyrus I/Inferior Temporal Gyrus I (PG), Temporal Pole (TP),

Mesial Frontal Gyrus/Inferior Frontal Gyrus (FG), Fusiform Gyrus II/Inferior Temporal Gyrus II (GY)

Seizure 2 Hippocampus Head (HH), Hippocampus Tail (HT), Parahippocampus Gyrus/Fusiform Gyrus I/Inferior Temporal Gyrus I (PG), Temporal Pole (TP), Fusiform Gyrus II/Inferior Temporal Gyrus II (GY), Mesial Frontal Gyrus/Inferior Frontal Gyrus (FG)

Seizure 3 Hippocampus Head (HH), Hippocampus Tail (HT)

Seizure 4 Hippocampus Head (HH), Hippocampus Tail (HT), Parahippocampus Gyrus/Fusiform Gyrus I/Inferior Temporal Gyrus I (PG), Temporal Pole (TP)

We report results on network inference based on sequences of affected regions. Instead of estimating λ_i's and λ_{ij}'s, we are after the collection of $\kappa_i = \lambda_i/\lambda_{hor}$ and $\kappa_{ij} = \lambda_{ij}/\lambda_{hor}$, where i, j index the brain regions.

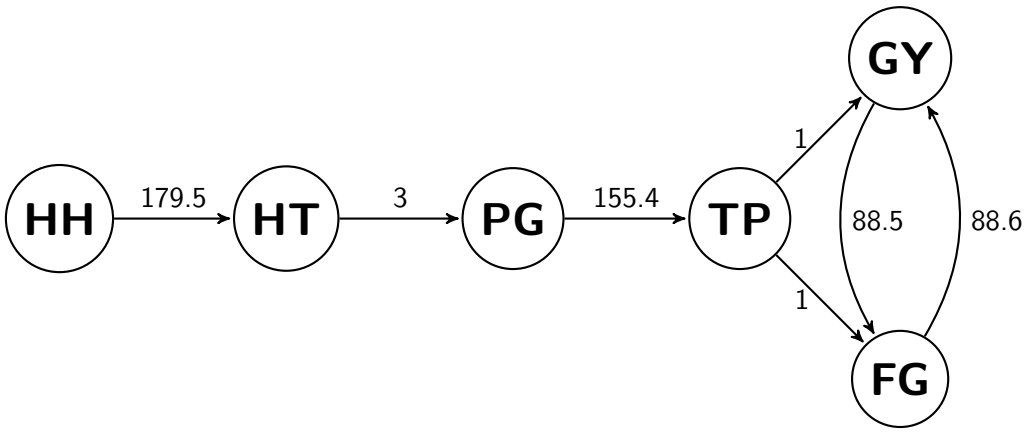
The unconstrained ML estimation estimates a large idiosyncratic rate for region HH, idiosyncratic rates that are close to zero for all other regions, and a dense network, and yields a log likelihood of -5.5461.

The ML estimation, regularized with an ℓ_1 term, minimizes the following objective function:

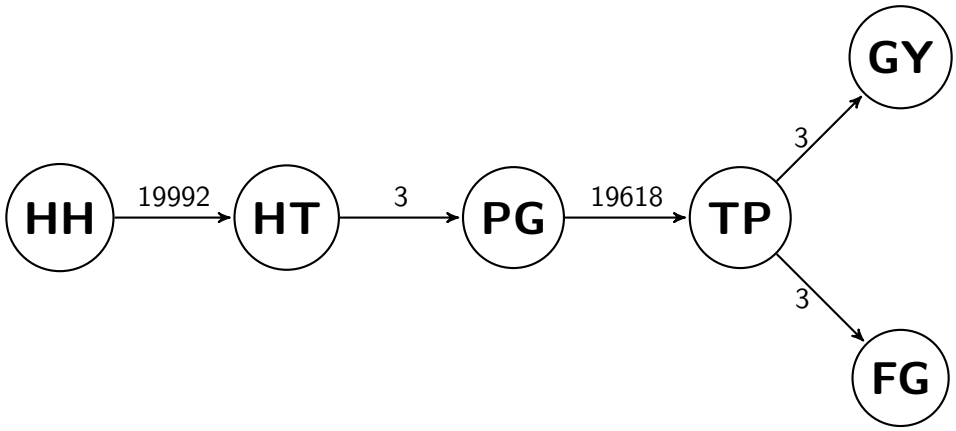
$$-\log \mathcal{L}\left(\{\kappa_i\}_{i=1}^6, \{\kappa_{ij}\}_{i,j=1}^6 \mid \text{sequences for seizures 1, 2, 3, 4}\right) + \rho \cdot \sum_{i,j=1}^{6} \kappa_{ij},$$

where $\mathcal{L}$ denotes the likelihood function, and we index the brain regions HH, HT, PG, TP, FG, GY with $i = 1, \ldots, 6$. $\rho > 0$ is a tuning parameter that controls the level of sparsity of the influence matrix with entries κ_{ij}. High values of ρ are likely to result in sparse inferred networks with high precision (i.e., fraction of retrieved instances that are relevant) and low recall (i.e., fraction of relevant instances that are retrieved); low values of ρ are likely to result in dense inferred networks with low precision and high recall. We start at a high ρ and decrease it with steps of size 1 in the exponent (i.e., $\rho = 1, e^{-1}, e^{-2}, \ldots$) until the inferred network is connected. The largest value of the regularizing parameter ρ for which the estimated network is

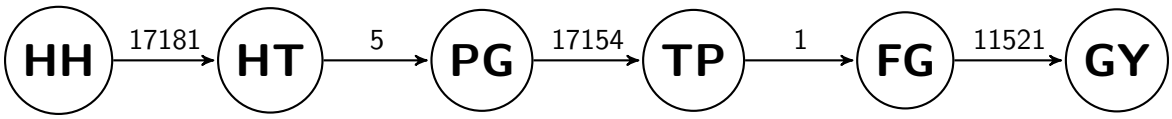
connected is $\rho = e^{-9}$, which results in a large estimated idiosyncratic rate for region HH, idiosyncratic rates that are close to zero for all other regions, the network drawn below, and a log likelihood of -5.6091 (-5.6729 after adding the regularization term).

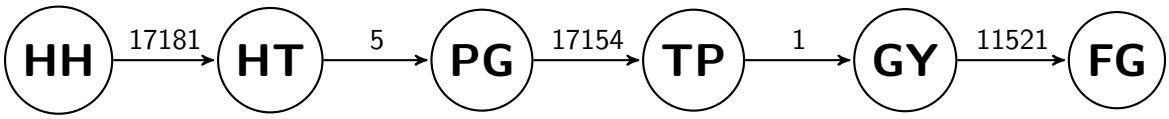


The ML estimation constrained over trees, yields the network drawn below as the maximum likelihood tree among the trees whose nodes have at most two children. The estimated idiosyncratic rate is very large for region HH, and close to zero for other regions. The log likelihood is -6.4625.



We also estimate the best serial model of influence that explains the data. The ML optimization program estimates a large idiosyncratic rate for region HH, idiosyncratic rates that are close to zero for all other regions, and the networks drawn below (both are optimal), resulting in a log likelihood of -6.7288.





We note that the worst serial model yields a log likelihood of -11.4896. It is worth observing that the best tree model (among trees whose nodes have at most two children), which is a sparse model, and even the best serial model, which is a sparse and very restrictive model, attain a log likelihood that is quite close to the log likelihood obtained by the model estimated through unconstrained ML estimation, which is dense.

9.3.3 Discussion

Physicians need to identify the source of the seizures, and are interested in understanding causal relations of influence between regions of the brain, based only on little data: EEG data is oftentimes available for only a handful of seizure events for each patient. Our methodology infers idiosyncratic rates of susceptibility to seizure, as well as pairwise rates of influence. Clearly, we cannot reach conclusive results through estimating our model with only a handful of events; nevertheless, we can generate informed hypotheses that shed light to the key properties of the system of interest: what regions are more likely to start the seizure, and how likely is it that a given region that is under seizure, infects another given region? By encouraging or constraining the inferred model to be sparse, we hope to generate parsimonious hypotheses that are easily understood and that can inform physicians' decisions on therapy.

Chapter 10

Conclusion

We have studied decision making in the context of networks of interconnected, interacting agents from two different angles: utility maximizers receive local information and want to coordinate, subject to the underlying fundamentals; and networked agents make decisions (or adopt a behavior) while influencing one another, and we seek to recover the network of influence from (possibly timed) data on the agents' behavior. We have thus studied the two-way link between networks and outcomes, using game theory to understand how networks shape outcomes, and using estimation and learning to infer networks from outcomes.

10.1 Summary

Coordination with Local Information In the first part of the thesis (Chapters 2, 3, 4), we studied the role of local information channels in enabling coordination among strategic agents. Building on the standard finite-player global games framework, we showed that the set of equilibria of a coordination game is highly sensitive to how information is locally shared among agents. In particular, we showed that the coordination game has multiple equilibria if there exists a collection of agents such that (i) they do not share a common signal with any agent outside of that collection; and (ii) their information sets form an increasing sequence of nested sets, referred to as a filtration. Our characterization thus extends the results on the uniqueness and

multiplicity of equilibria in global games beyond the well-known case in which agents have access to purely private or public signals. We then provided a characterization of how the extent of equilibrium multiplicity is determined by the extent to which subsets of agents have access to common information: we showed that the size of the set of equilibrium strategies is increasing with the extent of variability in the size of the subsets of agents who observe the same signal. We studied the set of equilibria in large coordination games, showing that as the number of agents grows, the game exhibits multiple equilibria if and only if a non-vanishing fraction of the agents have access to the same signal. We finally considered an application of our framework in which the noisy signals are interpreted to be the idiosyncratic signals of the agents, which are exchanged through a communication network.

The Value of Temporal Data for Learning of Influence Networks In the second part of the thesis (Chapters 5, 6, 7, 8, 9), we quantified the gain in the speed of learning of parametric models of influence, due to having access to richer temporal information. We inferred local influence relations between networked entities from data on outcomes and assessed the value of temporal data by characterizing the speed of learning under three different types of available data: knowing the set of entities who take a particular action; knowing the order in which the entities take an action; and knowing the times of the actions. We proposed a parametric model of influence which captures directed pairwise interactions and formulated different variations of the learning problem. We used the Fisher information, the Kullback-Leibler (KL) divergence, and sample complexity as measures for the speed of learning. We provided theoretical guarantees on the sample complexity for correct learning based on sets, sequences, and times. The asymptotic gain of having access to richer temporal data for the speed of learning was thus quantified in terms of the gap between the derived asymptotic requirements under different data modes. We also evaluated the practical value of learning with richer temporal data, by comparing learning with sets, sequences, and times given actual observational data. Experiments on both synthetic and real data, including data on mobile app installation behavior,

and EEG data from epileptic seizure events, quantified the improvement in learning due to richer temporal data, and showed that the proposed methodology recovers the underlying network well.

10.2 Directions for Future Research

Coordination with Local Information On the theoretical side, although we provide conditions on local information sharing among agents for uniqueness versus multiplicity, a complete characterization of sufficient and necessary conditions remains elusive. Furthermore, we have characterized the set of equilibria (i.e., the width of multiplicity) for the case where each agent's observation set is a singleton; for the general case, our Theorems 1 and 2 establish multiplicity when the observation sets form a nested sequence (and there is no overlap between signals within and without the cascade), yet we do not characterize the dependence of the width of multiplicity on the information structure. Such characterization would shed more light on the role of local information channels in enabling coordination among strategic agents. Finally, we are assuming common knowledge of the information structure for our results. Extending the model to incorporate asymmetric information about the information structure à la Galeotti, Goyal, Jackson, Vega-Redondo, and Yariv (2010), and checking whether our results are robust to a small amount of noise in the information structure of the game, would be a promising path for future research.

On the more applied side, our results show that incorporating local communication channels between agents may be of first order importance in understanding many phenomena that exhibit an element of coordination. Further exploring the impact of such channels in specific applications remains an important direction for future research.

The Value of Temporal Data for Learning of Influence Networks We hope to generalize our learning results for broader families of networks; formalizing learning guarantees on sample complexity for trees is the natural next step from our current

results. Furthermore, we are currently looking into the problem of reconstructing trees with latent nodes from data on sequences of actions taken by the observed nodes.

In addition, one can study models of influence other than the exponential delay and random observation window model used for our results so far. Another promising path ahead is to study experimentally the rate of convergence to the truth (i.e., the rate of error decay) as the number of products increases, and connect the findings with the developed theory, completing our understanding of how the network topology affects learning based on different data modes.

Appendix A

Proofs for Coordination with Local Information

Proof of Theorem 1

Before presenting the proof, we introduce some notation and prove some lemmas. Recall that a pure strategy of agent i is a mapping $s_i : \mathbb{R}^{|I_i|} \to \{0, 1\}$, where I_i denotes i's observation set. Thus, a pure strategy s_i can equivalently be represented by the set $A_i \subseteq \mathbb{R}^{|I_i|}$ over which agent i takes the risky action, i.e.,

$$A_i = \{y_i \in \mathbb{R}^{|I_i|} : s_i(y_i) = 1\},$$

where $y_i = (x_r)_{r \in I_i}$ denotes the collection of the realized signals in the observation set of agent i. Hence, a strategy profile can equivalently be represented by a collection of sets $A = (A_1, \ldots, A_n)$ over which agents take the risky action. We denote the set of all strategies of agent i by $\mathcal{A}_i$ and the set of all strategy profiles by $\mathcal{A}$. Given the strategies of other agents, A_{-i}, we denote the expected payoff to agent i of the risky action when she observes y_i by $V_i(A_{-i}|y_i)$. Thus, a best response mapping $\mathrm{BR}_i : \mathcal{A}_{-i} \to \mathcal{A}_i$ is naturally defined as

$$\mathrm{BR}_i(A_{-i}) = \left\{y_i \in \mathbb{R}^{|I_i|} : V_i(A_{-i}|y_i) > 0\right\}.$$

Finally, we define the mapping $\mathrm{BR} : \mathcal{A} \to \mathcal{A}$ as the product of the best response mappings of all agents, that is,

$$\mathrm{BR}(A) = \mathrm{BR}_1(A_{-1}) \times \cdots \times \mathrm{BR}_n(A_{-n}). \tag{A.1}$$

The $\mathrm{BR}(\cdot)$ mapping is monotone and continuous. More formally, we have the following lemmas:

Lemma 1 (Monotonicity)**.** *Consider two strategy profiles A and A' such that $A \subseteq A'$.*[1] *Then,* $\mathrm{BR}(A) \subseteq \mathrm{BR}(A')$.

Proof. Fix an agent i and consider $y_i \in \mathrm{BR}_i(A_{-i})$, which by definition satisfies $V_i(A_{-i}|y_i) > 0$. Due to the presence of strategic complementarities (Assumption 1), we have, $V_i(A'_{-i}|y_i) \geq V_i(A_{-i}|y_i)$, and as a result, $y_i \in \mathrm{BR}_i(A'_{-i})$. □

Lemma 2 (Continuity)**.** *Consider a sequence of strategy profiles $\{A^k\}_{k\in\mathbb{N}}$ such that $A^k \subseteq A^{k+1}$ for all k. Then,*

$$\bigcup_{k=1}^{\infty} \mathrm{BR}(A^k) = \mathrm{BR}(A^\infty),$$

where $A^\infty = \bigcup_{k=1}^{\infty} A^k$.

Proof. Clearly, $A^k \subseteq A^\infty$ and by Lemma 1, $\mathrm{BR}(A^k) \subseteq \mathrm{BR}(A^\infty)$ for all k. Thus,

$$\bigcup_{k=1}^{\infty} \mathrm{BR}(A^k) \subseteq \mathrm{BR}(A^\infty).$$

To prove the reverse inclusion, suppose that $y_i \in \mathrm{BR}_i(A^\infty_{-i})$, which implies that $V_i(A^\infty_{-i}|y_i) > 0$. On the other hand, for any θ and any observation profile $(y_1, \ldots, y_n)$, we have

$$\lim_{k\to\infty} u_i(a_i, s^k_{-i}(y_{-i}), \theta) = u_i(a_i, s^\infty_{-i}(y_{-i}), \theta),$$

where s^k and s^∞ are strategy profiles corresponding to sets A^k and A^∞, respectively.

[1] We write $A \subseteq A'$ whenever $A_i \subseteq A'_i$ for all i.

Thus, by the dominated convergence theorem,

$$\lim_{k\to\infty} V_i(A^k_{-i}|y_i) = V_i(A^\infty_{-i}|y_i), \tag{A.2}$$

where we have used Assumption 4. Therefore, there exists $r \in \mathbb{N}$ large enough such that $V_i(A^r_{-i}|y_i) > 0$, implying that $y_i \in \bigcup_{k=1}^\infty \mathrm{BR}(A^k_i)$. This completes the proof. $\square$

Next, let $\{R^k\}_{k\in\mathbb{N}}$ denote the sequence of strategy profiles defined recursively as

$$\begin{aligned} R^1 &= \varnothing, \\ R^{k+1} &= \mathrm{BR}(R^k). \end{aligned} \tag{A.3}$$

Thus, any strategy profile A that survives k rounds of iterated elimination of strictly dominated strategies must satisfy $R^{k+1} \subseteq A$. Consequently, A survives the iterated elimination of strictly dominated strategies only if $R \subseteq A$, where

$$R = \bigcup_{k=1}^\infty R^k. \tag{A.4}$$

We have the following lemma:

Lemma 3. *Suppose that there exists a non-singleton subset of agents $C \subseteq N$ such that $I_i = I_j$ for all $i, j \in C$. Then, there exists $\delta \geq \rho$ such that $V_i(R_{-i}|y_i) > \delta$ for all $y_i \in R_i$ and all $i \in C$.*

Proof. Consider agent $i \in C$. For any $y_i \in R_i$, there exists an integer k such that

$$y_i \notin R^k_i \quad \text{and} \quad y_i \in R^{k+1}_i.$$

Since $y_i \in R^{k+1}_i$, we have, $V_i(R^k_{-i}|y_i) > 0$. Furthermore, it must be the case that $y_i \in R^{k+1}_j$ for all $j \in C$. This is a consequence of the symmetry between agents in C which implies that $R^k_i = R^k_j$ for all $i, j \in C$ and all k. As a result, under the strategy profile R^{k+1}_{-i}, the realization of y_i implies that all agents $j \in C \setminus \{i\}$ take the risky action. A similar argument establishes that under the strategy profile R^k_{-i},

the realization of y_i implies that all agents $j \in C \setminus \{i\}$ take the safe action. Thus, conditional on the realization of y_i, at least $|C| - 1$ more agents take the risky action under R_{-i}^{k+1} compared to R_{-i}^{k}. Consequently, by Assumption 1,

$$V_i(R_{-i}^{k+1}|y_i) \geq V_i(R_{-i}^{k}|y_i) + (|C| - 1)\rho.$$

Letting $\delta = (|C|-1)\rho$ and using the fact that $V_i(R_{-i}^{k}|y_i) > 0$ completes the proof. $\square$

Proof of Theorem 1 First note that Lemma 2 implies that strategy profile R defined in (A.4) is a fixed-point of the best-response operator, that is, $R = \mathrm{BR}(R)$. Therefore, R is a Bayesian Nash equilibrium of the game. Next, suppose that there exists a strategy profile $R' \supset R$ such that the following properties are satisfied:

(i) $R'_i = R_i$ for all $i \notin C$.

(ii) $V_i(R'_{-i}|y_i) > 0$ for all $y_i \in R'_i$ and all $i \in C$.

(iii) $\lambda_{\mathbb{R}^{|\tilde{I}|}}(R'_i \setminus R_i) > 0$ for all $i \in C$, where λ is the Lebesgue measure.

If such a strategy profile R' exists, then (i) and (ii) immediately imply that $R' \subseteq \mathrm{BR}(R')$. Define the sequence of strategy profiles $\{Q^k\}_{k\in\mathbb{N}}$ as

$$Q^1 = R'$$
$$Q^{k+1} = \mathrm{BR}(Q^k).$$

Given that $Q^1 \subseteq Q^2$, Lemma 1 implies that $Q^k \subseteq Q^{k+1}$ for all k. Thus, $Q = \bigcup_{k=1}^{\infty} Q^k$ is well-defined, and by continuity of the BR operator (Lemma 2), satisfies $Q = \mathrm{BR}(Q)$. As a consequence, Q is also a Bayesian Nash equilibrium of the game which, in light of (iii), is distinct from R.

Thus, the proof is complete once we show that one can construct a strategy profile R' that satisfies properties (i)–(iii) above. To this end, let $R'_i = R_i$ for all $i \notin C$, whereas for $i \in C$, define

$$R'_i = R_i \cup \{y_i \notin R_i \text{ s.t. } \exists \tilde{y}_i \in R_i \text{ s.t. inequality (A.5) is satisfied}\},$$

where (A.5) is

$$\mathbb{E}\Big[\pi\big(|C| + \mu(y_{-C}, R_{-C}), \theta\big) \mid \tilde{y}_i\Big] - \mathbb{E}\Big[\pi\big(|C| + \mu(y_{-C}, R_{-C}), \theta\big) \mid y_i\Big] \leq \frac{\rho}{2}, \qquad \text{(A.5)}$$

and

$$\mu(y_{-C}, A_{-C}) = \sum_{i \notin C} \mathbb{1}_{\{y_i \in A_i\}}$$

is the number of agents not belonging to C that take the risky action under strategy profile A_{-C} and conditional on the realization of signals y_{-C}. Note that given assumption (b) of Theorem 1, the value of $\mu(y_{-C}, A_{-C})$ does not depend on the realization of the signals observed by agents in C.

Property (i) is satisfied by definition. To verify that R' satisfies (ii), first suppose that $y_i \in R_i$. This implies that

$$V_i(R'_{-i}|y_i) \geq V_i(R_{-i}|y_i).$$

By Lemma 3, the right-hand side of the above inequality is strictly positive, and hence, $V_i(R'_{-i}|y_i) > 0$. On the other hand, if $y_i \in R'_i \setminus R_i$, then

$$\begin{aligned} V_i(R'_{-i}|y_i) &= \mathbb{E}\Big[\pi\big(|C| + \mu(y_{-C}, R'_{-C}), \theta\big) \mid y_i\Big] \\ &= \mathbb{E}\Big[\pi\big(|C| + \mu(y_{-C}, R_{-C}), \theta\big) \mid y_i\Big]. \end{aligned}$$

Therefore,

$$V_i(R'_{-i}|y_i) = V_i(R_{-i}|\tilde{y}_i) + \mathbb{E}\Big[\pi\big(|C| + \mu(y_{-C}, R_{-C}), \theta\big) \mid y_i\Big] - \mathbb{E}\Big[\pi\big(|C| + \mu(y_{-C}, R_{-C}), \theta\big) \mid \tilde{y}_i\Big]$$

for some $\tilde{y}_i \in R_i$. Thus, by (A.5) and Lemma 3,

$$V_i(R'_{-i}|y_i) \geq \rho - \frac{\rho}{2} = \frac{\rho}{2}.$$

Thus, $V_i(R'_{-i}|y_i) > 0$ for all $y_i \in R'_i$.

Finally, to verify (iii), note that, by assumption (b) of the theorem, $I_i \cap I_j = \varnothing$ for all $i \in C$ and $j \notin C$, and therefore the joint density of θ, y_{-C}, conditioned on y_i, changes continuously with y_i. Thus, the expectations in (A.5) are continuous in agent i's observation. The continuity of the expectations in (A.5) then implies that $R'_i \setminus R_i$ has positive measure. $\square$

Proof of Theorem 2

Without loss of generality, we assume that $C = \{1, \ldots, \ell\}$, with $I_1 \subset I_2 \ldots \subset I_\ell$, and $\ell \geq 2$. Before proving the theorem, we introduce the induced coordination game among agents $\{1, \ldots, \ell\}$, and prove a lemma.

Consider the following ℓ-agent incomplete information game, played between agents 1, ..., ℓ. Agents $\ell+1, \ldots, n$ of the game of interest in the proposition are fixed to strategy profile R defined in (A.4), agent 1 (2, ..., ℓ)'s observation set is $I_1(I_2, \ldots, I_\ell)$, and payoffs to agents 1, ..., ℓ are

$$u_i(a_i, a_{-i}, \theta, x_{-I_\ell}) = \begin{cases} \Pi(k, \theta, x_{-I_\ell}) & \text{if } \ a_i = 1 \\ 0 & \text{if } \ a_i = 0, \end{cases}$$

where $k = \sum_{j=1}^{\ell} a_j$ is the number of agents out of $\{1, \ldots, \ell\}$ who take the risky action and

$$\Pi(k, \theta, x_{-I_\ell}) = \pi(k + g(x_{-I_\ell}), \theta) = \frac{k + g(x_{-I_\ell}) - 1}{n - 1} - \theta,$$

where the function $g : \mathbb{R}^{|I_{\ell+1}|} \times \ldots \times \mathbb{R}^{|I_n|} \to \{0, \ldots, n-\ell\}$, given by $g(x_{-I_\ell}) = \sum_{i=\ell+1}^{n} \mathbb{1}_{\{x_{I_i} \in R_i\}}$, maps the signals received by agents $\ell+1, \ldots, n$ to the number of agents from the collection $\{\ell+1, \ldots, n\}$ that play the risky action, given that they all play according to strategy profile R.

For this game, define the sequence of strategy profiles $\{R'^k\}_{k \in \mathbb{N}}$ as

$$R'^1 \ = \ \varnothing,$$

$$R'^{k+1} \quad = \quad \mathrm{BR}(R'^{k}).$$

Set $R' = \bigcup_{k=1}^{\infty} R'^{k}$. We have that $R'_1 = R_1, \ldots, R'_\ell = R_\ell$.

We have the following lemma, the proof of which is presented in the technical appendix of the paper.

Lemma 4. *For the ℓ-agent incomplete information game proposed above, there exists a strategy profile $R'' = (R''_1, \ldots, R''_\ell)$ such that the following properties are satisfied:*

(i) $R''_1 \supset R'_1 = R_1, \ldots, R''_\ell \supset R'_\ell = R_\ell$.

(ii) $V_1(R''_{-1} \mid y) > 0$ *for any* $y \in R''_1, \ldots, V_\ell(R''_{-\ell} \mid y) > 0$ *for any* $y \in R''_\ell$.

(iii) $\lambda_{\mathbb{R}^{|I_1|}}(R''_1 \setminus R'_1) > 0, \ldots, \lambda_{\mathbb{R}^{|I_\ell|}}(R''_\ell \setminus R'_\ell) > 0$, *where λ is the Lebesgue measure.*

We now present the proof of Theorem 2.

Proof of Theorem 2 Lemma 4 establishes a mutually profitable deviation (for agents $1, \ldots, \ell$) from strategy profile R for the game of interest in the lemma, and thus also for the game of interest in the proposition: namely, $(R''_1, \ldots, R''_\ell, R_{\ell+1} \ldots, R_n)$. By iteratively applying the BR operator on strategy profile $(R''_1, \ldots, R''_\ell, R_{\ell+1} \ldots, R_n)$, we converge to an equilibrium different from the equilibrium $(R_1, R_2, \ldots, R_n)$. $\square$

Proof of Proposition 1

Recall the sequence of strategy profiles R^k and its limit R defined in (A.3) and (A.4), respectively. By Lemma 2, $R = \mathrm{BR}(R)$, which implies that $y_i \in R_i$ if and only if $V_i(R_{-i}|y_i) > 0$. We have the following lemma.

Lemma 5. *There exists a strictly decreasing function $h : \mathbb{R} \to \mathbb{R}$ such that $V_i(R_{-i}|y_i) = 0$ if and only if $x_j = h(x_l)$, where $y_i = (x_j, x_l)$ and i, j and l are different.*

Proof. Using an inductive argument, we first prove that for all k:

(i) $V_i(R^k_{-i}|y_i)$ is continuously differentiable in y_i,

(ii) $V_i(R^k_{-i}|y_i)$ is strictly decreasing in both arguments, $(x_j, x_l) = y_i$

(iii) and $|\partial V_i(R^k_{-i}|y_i)/\partial x_j| \in [1/2, Q]$,

where $j \neq i$ and

$$Q = \frac{\sigma\sqrt{3\pi} + 1}{2\sigma\sqrt{3\pi} - 2}.$$

The above clearly hold for $k = 1$, as $V_i(\varnothing|y_i) = -(x_j + x_l)/2$. Now suppose that (i)–(iii) are satisfied for some $k \geq 1$. By the implicit function theorem[2], there exists a continuously differentiable function $h_k : \mathbb{R} \longrightarrow \mathbb{R}$ such that

$$V_i\left(R^k_{-i}\middle|x_j, h_k(x_j)\right) = 0,$$

and $-2Q \leq h'_k(x_j) \leq -1/2Q$.[3] The monotonicity of h_k implies that $V_i(R^k_{-i}|y_i) > 0$ if and only if $x_j < h_k(x_l)$. Therefore,

$$\begin{aligned}
V_i(R^{k+1}_{-i}|y_i) &= \frac{1}{2}\left[\mathbb{P}\left(y_j \in R^{k+1}_j|y_i\right) + \mathbb{P}\left(y_l \in R^{k+1}_l|y_i\right)\right] - \frac{1}{2}(x_j + x_l) \\
&= \frac{1}{2}\left[\mathbb{P}\left(x_i < h_k(x_l)|y_i\right) + \mathbb{P}\left(x_i < h_k(x_j)|y_i\right)\right] - \frac{1}{2}(x_j + x_l) \\
&= \frac{1}{2}\left[\Phi\left(\frac{h_k(x_l) - (x_j + x_l)/2}{\sigma\sqrt{3/2}}\right) + \Phi\left(\frac{h_k(x_j) - (x_j + x_l)/2}{\sigma\sqrt{3/2}}\right)\right] - \frac{1}{2}(x_j + x_l),
\end{aligned}$$

which immediately implies that $V_i(R^{k+1}_{-i}|y_i)$ is continuously differentiable and strictly decreasing in both arguments. Furthermore,

$$\frac{\partial}{\partial x_j} V_i(R^{k+1}_{-i}|y_i) = -\frac{1}{2\sigma\sqrt{6}}\phi\left(\frac{h_k(x_l) - (x_j + x_l)/2}{\sigma\sqrt{3/2}}\right) + \frac{2h'_k(x_j) - 1}{2\sigma\sqrt{6}}\phi\left(\frac{h_k(x_j) - (x_j + x_l)/2}{\sigma\sqrt{3/2}}\right) - \frac{1}{2},$$

which guarantees

$$-\frac{1}{2} - \frac{1 + 2Q}{2\sigma\sqrt{3\pi}} \leq \frac{\partial}{\partial x_j} V_i(R^{k+1}_{-i}|y_i) \leq -\frac{1}{2},$$

[2]See, for example, Hadamard's global implicit function theorem in Krantz and Parks (2002).
[3]Given the symmetry between the three agents, we drop the agent's index for function h_k.

completing the inductive argument, because

$$\frac{1}{2} + \frac{1+2Q}{2\sigma\sqrt{3\pi}} = Q.$$

Now using (A.2) and the implicit function theorem once again completes the proof.

□

Proof of Proposition 1 By definition,

$$\begin{aligned} V_i(R_{-i}|y_i) &= \frac{1}{2}\Big[\mathbb{P}\big(y_j \in R_j\big|y_i\big) + \mathbb{P}\big(y_l \in R_l\big|y_i\big)\Big] - \frac{1}{2}(x_j + x_l) \\ &= \frac{1}{2}\left[\mathbb{P}\left(V_j(R_{-j}|y_j) > 0\Big|y_i\right) + \mathbb{P}\left(V_l(R_{-l}|y_l) > 0\Big|y_i\right)\right] - \frac{1}{2}(x_j + x_l), \end{aligned}$$

where we used the fact that $R = \mathrm{BR}(R)$. By Lemma 5,

$$V_i(R_{-i}|y_i) = \frac{1}{2}\Phi\left(\frac{h(x_l) - (x_j + x_l)/2}{\sigma\sqrt{3/2}}\right) + \frac{1}{2}\Phi\left(\frac{h(x_j) - (x_j + x_l)/2}{\sigma\sqrt{3/2}}\right) - \frac{1}{2}(x_j + x_l).$$

Setting $V_i(R_{-i}|y_i) = 0$ and using the fact that any solution $y_i = (x_j, x_l)$ of $V_i(R_{-i}|y_i) = 0$ satisfies $h(x_l) = x_j$ and $h(x_j) = x_l$, imply that $x_j + x_l = 1$. Therefore,

$$R_i = \left\{(x_j, x_l) \in \mathbb{R}^2 : \frac{1}{2}(x_j + x_l) < \frac{1}{2}\right\}.$$

Hence, in any strategy profile that survives iterated elimination of strictly dominated strategies, an agent takes the risky action whenever the average of the two signals she observes is less than 1/2. A symmetrical argument implies that in any strategy profile that survives the iterated elimination of strictly dominated strategies, the agent takes the safe action whenever the average of her signals is greater than 1/2. Thus, the game has an essentially unique rationalizable strategy profile, and hence, an essentially unique Bayesian Nash equilibrium. □

Proof of Proposition 2

It is sufficient to show that there exists an essentially unique equilibrium in monotone strategies. Note that due to symmetry, agent 1 and 2's strategies in the extremal equilibria of the game are identical. Denote the (common) equilibrium threshold of agents 1 and 2's threshold strategies by $\tau \in [0, 1]$, in the sense that they take the risky action if and only if their observation is less than τ. Then, the expected payoff of taking the risky action to agent 3 is equal to

$$\mathbb{E}\left[\pi(k, \theta)|x_1, x_2\right] = \frac{1}{2}\big[\mathbb{1}_{\{x_1<\tau\}} + \mathbb{1}_{\{x_2<\tau\}} - (x_1 + x_2)\big].$$

First suppose that $\tau > 1/2$. Then, the best response of agent 3 is to take the risky action if either (i) $x_1 + x_2 \leq 1$, or (ii) $x_1, x_2 \leq \tau$ hold. On the other hand, for τ to correspond to the threshold of an equilibrium strategy of agent 1, her expected payoff of taking the risky action has to be positive whenever $x_1 < \tau$. In other words,

$$\frac{1}{2}\mathbb{P}(x_2 < \tau|x_1) + \frac{1}{2}\big[\mathbb{P}(x_1 + x_2 \leq 1|x_1) + \mathbb{P}(1 - x_1 \leq x_2 \leq \tau|x_1)\big] \geq x_1$$

for all $x_1 < \tau$. As a result, for any $x_1 < \tau$, we have

$$\frac{1}{2}\Phi\left(\frac{\tau - x_1}{\sigma\sqrt{2}}\right) + \frac{1}{2}\Phi\left(\frac{1 - 2x_1}{\sigma\sqrt{2}}\right) + \frac{1}{2}\left[\Phi\left(\frac{\tau - x_1}{\sigma\sqrt{2}}\right) - \Phi\left(\frac{1 - 2x_1}{\sigma\sqrt{2}}\right)\right] \geq x_1.$$

which simplifies to

$$\Phi\left(\frac{\tau - x_1}{\sigma\sqrt{2}}\right) \geq x_1.$$

Taking the limit of the both sides of the above inequality as $x_1 \uparrow \tau$ implies that $\tau \leq 1/2$. This, however, is in contradiction with the original assumption that $\tau > 1/2$. A similar argument would also rule out the case that $\tau < 1/2$. Hence, τ corresponds to the threshold of an equilibrium strategy of agents 1 and 2 only if $\tau = 1/2$. As a consequence, in the essentially unique equilibrium of the game agent 3 takes the risky action if and only if $x_1 + x_2 < 1$. This proves the first part of the proposition. The proof of the second part is immediate. □

Proof of Theorem 3

As already mentioned, the Bayesian game under consideration is monotone supermodular in the sense of Van Zandt and Vives (2007), which ensures that the set of equilibria has well-defined maximal and minimal elements, each of which is in threshold strategies. Moreover, by Milgrom and Roberts (1990), all profiles of rationalizable strategies are "sandwiched" between these two equilibria. Hence, to characterize the set of rationalizable strategies, it suffices to focus on threshold strategies and determine the smallest and largest thresholds that correspond to Bayesian Nash equilibria of the game.

Denote the threshold corresponding to the strategy of an agent who observes signal x_r with τ_r. The profile of threshold strategies corresponding to thresholds $\{\tau_r\}$ is a Bayesian Nash equilibrium of the game if and only if for all r,

$$\begin{aligned} \frac{n}{n-1}\sum_{j\neq r} c_j \Phi\left(\frac{\tau_j - x_r}{\sigma\sqrt{2}}\right) - x_r \leq 0 \qquad & \text{for} \quad x_r > \tau_r \\ \frac{nc_r - 1}{n-1} + \frac{n}{n-1}\sum_{j\neq r} c_j \Phi\left(\frac{\tau_j - x_r}{\sigma\sqrt{2}}\right) - x_r \geq 0 \qquad & \text{for} \quad x_r < \tau_r \end{aligned}$$

where the first (second) inequality guarantees that the agent has no incentive to deviate to the risky (safe) action when the signal she observes is above (below) threshold τ_r. Taking the limit as x_r converges to τ_r from above in the first inequality implies that in any Bayesian Nash equilibrium of the game in threshold strategies,

$$(n-1)\tau \geq nHc$$

where $\tau = [\tau_1, \ldots, \tau_m]$ is the vector of thresholds and $H \in \mathbb{R}^{m\times m}$ is a matrix with zero diagonal entries, and off-diagonal entries given by $H_{jr} = \Phi((\tau_j - \tau_r)/\sigma\sqrt{2})$. Therefore,

$$\begin{aligned} 2(n-1)c'\tau &\geq nc'(H' + H)c \\ &= nc'(\mathbf{1}\mathbf{1}')c - nc'c, \end{aligned}$$

where $\mathbf{1}$ is the vector of all ones and we used the fact that $\Phi(z) + \Phi(-z) = 1$. Consequently,

$$2(n-1)c'\tau \geq n(1 - \|c\|_2^2).$$

Finally, given that $|\tau_r - \tau_j| \to 0$ as $\sigma \to 0$, and using the fact that $\|c\|_1 = n$, the left-hand side of the above inequality converges to $2(n-1)\tau^*$ for some constant τ^*, and as a result, the smallest possible threshold that corresponds to a Bayesian Nash equilibrium is equal to

$$\underline{\tau} = \frac{n}{2(n-1)}\Big(1 - \|c\|_2^2\Big).$$

The expression for $\bar{\tau}$ is derived analogously. □

Proof of Proposition 3

We first prove that if the size of the largest set of agents with a common observation grows sublinearly in n, then, asymptotically as $n \to \infty$, the game has a unique equilibrium. We denote the vector of the fractions of agents observing each signal by $c(n)$, to make explicit the dependence on the number of agents n. Note that if the size of the largest set of agents with a common observation grows sublinearly in n, then

$$\lim_{n\to\infty} \|c(n)\|_\infty = 0,$$

where $\|z\|_\infty$ is the maximum element of vector z. On the other hand, it holds that $\|c(n)\|_2^2 \leq \|c(n)\|_1 \|c(n)\|_\infty$. Given that the elements of vector $c(n)$ add up to one, $\lim_{n\to\infty} \|c(n)\|_2 = 0$. Consequently, Theorem 3 implies that thresholds $\underline{\tau}(n)$ and $\bar{\tau}(n)$ characterizing the set of rationalizable strategies of the game of size n satisfy

$$\lim_{n\to\infty} \underline{\tau}(n) = \lim_{n\to\infty} \bar{\tau}(n) = 1/2.$$

Thus, asymptotically, the game has an essentially unique Bayesian Nash equilibrium.

To prove the converse, suppose that the size of the largest set of agents with a common observation grows linearly in n, which means that $\|c(n)\|_\infty$ remains bounded

away from zero as $n \to \infty$. Furthermore, the inequality $\|c(n)\|_\infty \leq \|c(n)\|_2$ immediately implies that $\|c(n)\|_2$ also remains bounded away from zero as $n \to \infty$. Hence, by Theorem 3,

$$\limsup_{n\to\infty} \underline{\tau}(n) < 1/2$$
$$\liminf_{n\to\infty} \bar{\tau}(n) > 1/2,$$

guaranteeing asymptotic multiplicity of equilibria as $n \to \infty$. $\square$

Appendix B

Technical Appendix: Proof of Lemma 4 for Coordination with Local Information

This technical appendix contains the proof of Lemma 4. The proof constructs all possible equilibria in threshold strategies for the induced ℓ-agent incomplete information game.

Proof. It suffices to show that the game has multiple Bayesian Nash equilibria that are distinct. The existence of strategy profile R'' satisfying the desired properties then follows, in light of the definition of the BR operator (A.1).

Suppose that, at an equilibrium, agent 1 uses threshold τ_1 in the sense that she plays risky if and only if she observes $\sum_{x_i \in I_1} x_i < \tau_1$; agent 2 uses the following policy: if $\sum_{x_i \in I_1} x_i < \tau_1$, then she plays risky if and only if $\sum_{x_i \in I_2} x_i < \tau_2^{(1)}$, while if $\sum_{x_i \in I_1} x_i \geq \tau_1$, then she plays risky if and only if $\sum_{x_i \in I_2} x_i < \tau_2^{(2)}$, with $\tau_2^{(1)} > \tau_2^{(2)}$; and in general, agent j, $1 \leq j \leq \ell - 1$ uses a policy with thresholds $\tau_j^{(1)} > \tau_j^{(2)} \ldots > \tau_j^{(j)}$.

Then the expected payoff to agent ℓ of taking the risky action is

$$\mathbb{E}[\Pi(k,\theta,x_{-I_\ell}) \mid x_{I_\ell}] = \begin{cases} \frac{\ell-1}{n-1} + f_\ell\left(\sum_{x_i\in I_\ell} x_i\right) - \frac{\sum_{x_i\in I_\ell} x_i}{|I_\ell|}, & \text{if everybody in } \{1,\ldots,\ell-1\} \text{ plays risky} \\ \frac{\ell-2}{n-1} + f_\ell\left(\sum_{x_i\in I_\ell} x_i\right) - \frac{\sum_{x_i\in I_\ell} x_i}{|I_\ell|}, & \text{if exactly one agent in } \{1,\ldots,\ell-1\} \text{ plays safe} \\ \vdots & \\ \frac{1}{n-1} + f_\ell\left(\sum_{x_i\in I_\ell} x_i\right) - \frac{\sum_{x_i\in I_\ell} x_i}{|I_\ell|}, & \text{if exactly } \ell-2 \text{ agents in } \{1,\ldots,\ell-1\} \text{ play safe} \\ f_\ell\left(\sum_{x_i\in I_\ell} x_i\right) - \frac{\sum_{x_i\in I_\ell} x_i}{|I_\ell|}, & \text{if everybody in } \{1,\ldots,\ell-1\} \text{ plays safe} \end{cases}$$

where function f_ℓ is continuous and strictly decreasing in its argument, with $\lim_{x\to\infty} f_\ell(x) = 0$ and $\lim_{x\to-\infty} f_\ell(x) = \frac{n-\ell}{n-1}$.

Each of the equations

$$\begin{aligned} \frac{\ell-1}{n-1} + f_\ell\left(\sum_{x_i\in I_\ell} x_i\right) - \frac{\sum_{x_i\in I_\ell} x_i}{|I_\ell|} &= 0 \\ \frac{\ell-2}{n-1} + f_\ell\left(\sum_{x_i\in I_\ell} x_i\right) - \frac{\sum_{x_i\in I_\ell} x_i}{|I_\ell|} &= 0 \\ &\vdots \\ \frac{1}{n-1} + f_\ell\left(\sum_{x_i\in I_\ell} x_i\right) - \frac{\sum_{x_i\in I_\ell} x_i}{|I_\ell|} &= 0 \\ f_\ell\left(\sum_{x_i\in I_\ell} x_i\right) - \frac{\sum_{x_i\in I_\ell} x_i}{|I_\ell|} &= 0 \end{aligned}$$

has a unique solution, which we denote respectively $\sum_{x_i\in I_\ell} x_i = \tau_\ell^{(1)}$, $\sum_{x_i\in I_\ell} x_i = \tau_\ell^{(2)}$, $\ldots$, $\sum_{x_i\in I_\ell} x_i = \tau_\ell^{(\ell)}$. It holds that $\tau_\ell^{(1)} > \tau_\ell^{(2)} > \ldots > \tau_\ell^{(\ell)}$.

The thresholds for agent ℓ determine bounds for the thresholds of agent $\ell-1$; a selection of thresholds for agent $\ell-1$, along with the thresholds for agent ℓ, determine bounds for the thresholds of agent $\ell-2$, and so on.

We next show this transition for agents $j, j-1$, with $2 \le j \le \ell$. Having determined

thresholds

$$
\begin{aligned}
&\tau_\ell^{(1)} > \tau_\ell^{(2)} \ldots > \tau_\ell^{(\ell)} \\
&\tau_{\ell-1}^{(1)} > \tau_{\ell-1}^{(2)} \ldots > \tau_{\ell-1}^{(\ell-1)} \\
&\quad\vdots \\
&\tau_j^{(1)} > \tau_j^{(2)} \ldots > \tau_j^{(j)},
\end{aligned}
\tag{B.1}
$$

for agents $\ell, \ell-1, \ldots, j$ respectively, we study the best response of agent $j-1$.

For $\sum_{x_i \in I_1} x_i < \tau_1, \sum_{x_i \in I_2} x_i < \tau_2^{(1)}, \ldots, \sum_{x_i \in I_{j-2}} x_i < \tau_{j-2}^{(1)}, \sum_{x_i \in I_{j-1}} x_i < \tau_{j-1}^{(1)}$, the expected payoff to agent $j-1$ of taking the risky action has to be positive. We show that this inequality has a unique solution, which we denote by $\sum_{x_i \in I_{j-1}} x_i < b_{j-1}^{(1)}$.

For ease of exposition, we define $y_m = \sum_{x_i \in I_m \setminus I_{m-1}} x_i$, $S_m = \sum_{x_i \in I_m} x_i$. We have

$$
\begin{aligned}
&(j-2)\frac{1}{n-1} \\
&\quad + \frac{1}{n-1}\Phi\left(\frac{\tau_j^{(1)} - |I_j|\frac{\sum_{x_i \in I_{j-1}} x_i}{|I_{j-1}|}}{\sigma\sqrt{(|I_j| - |I_{j-1}|)\frac{|I_{j-1}|+1}{|I_{j-1}|}}}\right) \\
&\quad + \frac{1}{n-1}\left(\int_{y_j=-\infty}^{y_j=\tau_j^{(1)}-S_{j-1}} \Phi\left(\frac{\tau_{j+1}^{(1)} - S_{j-1} - y_j - (|I_{j+1}| - |I_j|)\frac{S_{j-1}}{|I_{j-1}|}}{\sigma\sqrt{(|I_{j+1}| - |I_j|)\frac{|I_{j-1}|+1}{|I_{j-1}|}}}\right) dF_{y_j}(y_j)\right) \\
&\quad + \frac{1}{n-1}\left(\int_{y_j=\tau_j^{(1)}-S_{j-1}}^{y=+\infty} \Phi\left(\frac{\tau_{j+1}^{(2)} - S_{j-1} - y_j - (|I_{j+1}| - |I_j|)\frac{S_{j-1}}{|I_{j-1}|}}{\sigma\sqrt{(|I_{j+1}| - |I_j|)\frac{|I_{j-1}|+1}{|I_{j-1}|}}}\right) dF_{y_j}(y_j)\right) \\
&\quad + \frac{1}{n-1}\left(\int_{y_j=-\infty}^{y_j=\tau_j^{(1)}-S_{j-1}} \int_{y_{j+1}=-\infty}^{y_{j+1}=\tau_{j+1}^{(1)}-y_j-S_{j-1}} \Phi\left(\frac{\tau_{j+2}^{(1)} - S_{j-1} - y_j - y_{j+1} - (|I_{j+2}| - |I_{j+1}|)\frac{S_{j-1}}{|I_{j-1}|}}{\sigma\sqrt{(|I_{j+2}| - |I_{j+1}|)\frac{|I_{j-1}|+1}{|I_{j-1}|}}}\right)\right. \\
&\qquad\qquad \left. dF_{y_{j+1}}(y_{j+1}) dF_{y_j}(y_j)\right) \\
&\quad + \frac{1}{n-1}\left(\int_{y_j=-\infty}^{y_j=\tau_j^{(1)}-S_{j-1}} \int_{y_{j+1}=\tau_{j+1}^{(1)}-y_j-S_{j-1}}^{y_{j+1}=+\infty} \Phi\left(\frac{\tau_{j+2}^{(2)} - S_{j-1} - y_j - y_{j+1} - (|I_{j+2}| - |I_{j+1}|)\frac{S_{j-1}}{|I_{j-1}|}}{\sigma\sqrt{(|I_{j+2}| - |I_{j+1}|)\frac{|I_{j-1}|+1}{|I_{j-1}|}}}\right)\right. \\
&\qquad\qquad \left. dF_{y_{j+1}}(y_{j+1}) dF_{y_j}(y_j)\right) \\
&\quad + \frac{1}{n-1}\left(\int_{y_j=\tau_j^{(1)}-S_{j-1}}^{y_j=+\infty} \int_{y_{j+1}=-\infty}^{y_{j+1}=\tau_{j+1}^{(1)}-y_j-S_{j-1}} \Phi\left(\frac{\tau_{j+2}^{(2)} - S_{j-1} - y_j - y_{j+1} - (|I_{j+2}| - |I_{j+1}|)\frac{S_{j-1}}{|I_{j-1}|}}{\sigma\sqrt{(|I_{j+2}| - |I_{j+1}|)\frac{|I_{j-1}|+1}{|I_{j-1}|}}}\right)\right. \\
&\qquad\qquad \left. dF_{y_{j+1}}(y_{j+1}) dF_{y_j}(y_j)\right) \\
&\quad + \frac{1}{n-1}\left(\int_{y_j=\tau_j^{(1)}-S_{j-1}}^{y_j=+\infty} \int_{y_{j+1}=\tau_{j+1}^{(1)}-y_j-S_{j-1}}^{y_{j+1}=+\infty} \Phi\left(\frac{\tau_{j+2}^{(3)} - S_{j-1} - y_j - y_{j+1} - (|I_{j+2}| - |I_{j+1}|)\frac{S_{j-1}}{|I_{j-1}|}}{\sigma\sqrt{(|I_{j+2}| - |I_{j+1}|)\frac{|I_{j-1}|+1}{|I_{j-1}|}}}\right)\right. \\
&\qquad\qquad \left. dF_{y_{j+1}}(y_{j+1}) dF_{y_j}(y_j)\right) \\
&\quad \vdots \\
&\quad + f_{j-1}\left(\sum_{x_i \in I_{j-1}} x_i\right) - \sum_{x_i \in I_{j-1}} x_i > 0,
\end{aligned}
\tag{B.2}
$$

where function f_{j-1} is continuous and strictly decreasing, with $\lim_{x\to\infty} f_{j-1}(x) = 0$ and $\lim_{x\to-\infty} f_{j-1}(x) = \frac{n-\ell}{n-1}$.

We argue that the left-hand side of expression (B.2) is monotonically decreasing in $S_{j-1} = \sum_{x_i \in I_{j-1}} x_i$. Clearly, the term in the second line is decreasing in S_{j-1}. Using

$\phi(\cdot)$ to denote the probability density function of the standard normal, the derivative of the terms in the third and fourth lines, multiplied by $n-1$, is

$$
\begin{aligned}
&-\Phi\left(\frac{\tau_{j+1}^{(1)}-S_{j-1}-(\tau_j^{(1)}-S_{j-1})-(|I_{j+1}|-|I_j|)\frac{S_{j-1}}{|I_{j-1}|}}{\sigma\sqrt{(|I_{j+1}|-|I_j|)\frac{|I_{j-1}|+1}{|I_{j-1}|}}}\right)\mathrm{pdf}_{y_j}(\tau_j^{(1)}-S_{j-1})\\
&\quad-\frac{1+\frac{|I_{j+1}|-|I_j|}{|I_{j-1}|}}{\sigma\sqrt{(|I_{j+1}|-|I_j|)\frac{|I_{j-1}|+1}{|I_{j-1}|}}}\int_{y_j=-\infty}^{y_j=\tau_j^{(1)}-S_{j-1}}\phi\left(\frac{\tau_{j+1}^{(1)}-S_{j-1}-y_j-(|I_{j+1}|-|I_j|)\frac{S_{j-1}}{|I_{j-1}|}}{\sigma\sqrt{(|I_{j+1}|-|I_j|)\frac{|I_{j-1}|+1}{|I_{j-1}|}}}\right)dF_{y_j}(y_j)\\
&\quad+\Phi\left(\frac{\tau_{j+1}^{(2)}-S_{j-1}-(\tau_j^{(1)}-S_{j-1})-(|I_{j+1}|-|I_j|)\frac{S_{j-1}}{|I_{j-1}|}}{\sigma\sqrt{(|I_{j+1}|-|I_j|)\frac{|I_{j-1}|+1}{|I_{j-1}|}}}\right)\mathrm{pdf}_{y_j}(\tau_j^{(1)}-S_{j-1})\\
&\quad-\frac{1+\frac{|I_{j+1}|-|I_j|}{|I_{j-1}|}}{\sigma\sqrt{(|I_{j+1}|-|I_j|)\frac{|I_{j-1}|+1}{|I_{j-1}|}}}\int_{y_j=-\infty}^{y_j=\tau_j^{(1)}-S_{j-1}}\phi\left(\frac{\tau_{j+1}^{(2)}-S_{j-1}-y_j-(|I_{j+1}|-|I_j|)\frac{S_{j-1}}{|I_{j-1}|}}{\sigma\sqrt{(|I_{j+1}|-|I_j|)\frac{|I_{j-1}|+1}{|I_{j-1}|}}}\right)dF_{y_j}(y_j),
\end{aligned}
$$

which is negative for all S_{j-1}, because $\tau_{j+1}^{(1)} > \tau_{j+1}^{(2)}$. The derivative of the terms in

lines 5-8 of expression (B.2), multiplied by $n-1$, is[1]:

$$
\begin{aligned}
&-\int_{y_{j+1}=-\infty}^{y_{j+1}=\tau_{j+1}^{(1)}-(\tau_j^{(1)}-S_{j-1})-S_{j-1}} \Phi\left(\frac{\tau_{j+2}^{(1)}-S_{j-1}-(\tau_j^{(1)}-S_{j-1})-y_{j+1}-(|I_{j+2}|-|I_{j+1}|)\frac{S_{j-1}}{|I_{j-1}|}}{\sigma\sqrt{(|I_{j+2}|-|I_{j+1}|)\frac{|I_{j-1}|+1}{|I_{j-1}|}}}\right) dF_{y_{j+1}}(y_{j+1})\cdot \mathrm{pdf}_{y_j}(\tau_j^{(1)}-S_{j-1}) \\
&-\int_{y_j=-\infty}^{y_j=\tau_j^{(1)}-S_{j-1}} \Phi\left(\frac{\tau_{j+2}^{(1)}-S_{j-1}-y_j-(\tau_{j+1}^{(1)}-y_j-S_{j-1})-(|I_{j+2}|-|I_{j+1}|)\frac{S_{j-1}}{|I_{j-1}|}}{\sigma\sqrt{(|I_{j+2}|-|I_{j+1}|)\frac{|I_{j-1}|+1}{|I_{j-1}|}}}\right) \mathrm{pdf}_{y_{j+1}}(\tau_{j+1}^{(1)}-y_j-S_{j-1})dF_{y_j}(y_j) \\
&-\frac{1+\frac{|I_{j+2}|-|I_{j+1}|}{|I_{j-1}|}}{\sigma\sqrt{(|I_{j+2}|-|I_{j+1}|)\frac{|I_{j-1}|+1}{|I_{j-1}|}}}\int_{y_j=-\infty}^{y_j=\tau_j^{(1)}-S_{j-1}}\int_{y_{j+1}=-\infty}^{y_{j+1}=\tau_{j+1}^{(1)}-y_j-S_{j-1}} \phi\left(\frac{\tau_{j+2}^{(1)}-S_{j-1}-y_j-y_{j+1}-(|I_{j+2}|-|I_{j+1}|)\frac{S_{j-1}}{|I_{j-1}|}}{\sigma\sqrt{(|I_{j+2}|-|I_{j+1}|)\frac{|I_{j-1}|+1}{|I_{j-1}|}}}\right) \\
&\qquad\qquad dF_{y_{j+1}}(y_{j+1})dF_{y_j}(y_j) \\
&-\int_{y_{j+1}=\tau_{j+1}^{(1)}-(\tau_j^{(1)}-S_{j-1})-S_{j-1}}^{y_{j+1}=+\infty} \Phi\left(\frac{\tau_{j+2}^{(2)}-S_{j-1}-(\tau_j^{(1)}-S_{j-1})-y_{j+1}-(|I_{j+2}|-|I_{j+1}|)\frac{S_{j-1}}{|I_{j-1}|}}{\sigma\sqrt{(|I_{j+2}|-|I_{j+1}|)\frac{|I_{j-1}|+1}{|I_{j-1}|}}}\right) dF_{y_{j+1}}(y_{j+1})\cdot \mathrm{pdf}_{y_j}(\tau_j^{(1)}-S_{j-1}) \\
&+\int_{y_j=-\infty}^{y_j=\tau_j^{(1)}-S_{j-1}} \Phi\left(\frac{\tau_{j+2}^{(2)}-S_{j-1}-y_j-(\tau_{j+1}^{(1)}-y_j-S_{j-1})-(|I_{j+2}|-|I_{j+1}|)\frac{S_{j-1}}{|I_{j-1}|}}{\sigma\sqrt{(|I_{j+2}|-|I_{j+1}|)\frac{|I_{j-1}|+1}{|I_{j-1}|}}}\right) \mathrm{pdf}_{y_{j+1}}(\tau_{j+1}^{(1)}-y_j-S_{j-1})dF_{y_j}(y_j) \\
&-\frac{1+\frac{|I_{j+2}|-|I_{j+1}|}{|I_{j-1}|}}{\sigma\sqrt{(|I_{j+2}|-|I_{j+1}|)\frac{|I_{j-1}|+1}{|I_{j-1}|}}}\int_{y_j=-\infty}^{y_j=\tau_j^{(1)}-S_{j-1}}\int_{y_{j+1}=\tau_{j+1}^{(1)}-y_j-S_{j-1}}^{y_{j+1}=+\infty} \phi\left(\frac{\tau_{j+2}^{(2)}-S_{j-1}-y_j-y_{j+1}-(|I_{j+2}|-|I_{j+1}|)\frac{S_{j-1}}{|I_{j-1}|}}{\sigma\sqrt{(|I_{j+2}|-|I_{j+1}|)\frac{|I_{j-1}|+1}{|I_{j-1}|}}}\right) \\
&\qquad\qquad dF_{y_{j+1}}(y_{j+1})dF_{y_j}(y_j) \\
&+\int_{y_{j+1}=-\infty}^{y_{j+1}=\tau_{j+1}^{(1)}-(\tau_j^{(1)}-S_{j-1})-S_{j-1}} \Phi\left(\frac{\tau_{j+2}^{(2)}-S_{j-1}-(\tau_j^{(1)}-S_{j-1})-y_{j+1}-(|I_{j+2}|-|I_{j+1}|)\frac{S_{j-1}}{|I_{j-1}|}}{\sigma\sqrt{(|I_{j+2}|-|I_{j+1}|)\frac{|I_{j-1}|+1}{|I_{j-1}|}}}\right) dF_{y_{j+1}}(y_{j+1})\cdot \mathrm{pdf}_{y_j}(\tau_j^{(1)}-S_{j-1}) \\
&-\int_{y_j=\tau_j^{(1)}-S_{j-1}}^{y_j=+\infty} \Phi\left(\frac{\tau_{j+2}^{(2)}-S_{j-1}-y_j-(\tau_{j+1}^{(1)}-y_j-S_{j-1})-(|I_{j+2}|-|I_{j+1}|)\frac{S_{j-1}}{|I_{j-1}|}}{\sigma\sqrt{(|I_{j+2}|-|I_{j+1}|)\frac{|I_{j-1}|+1}{|I_{j-1}|}}}\right) \mathrm{pdf}_{y_{j+1}}(\tau_{j+1}^{(1)}-y_j-S_{j-1})dF_{y_j}(y_j) \\
&-\frac{1+\frac{|I_{j+2}|-|I_{j+1}|}{|I_{j-1}|}}{\sigma\sqrt{(|I_{j+2}|-|I_{j+1}|)\frac{|I_{j-1}|+1}{|I_{j-1}|}}}\int_{y_j=\tau_j^{(1)}-S_{j-1}}^{y_j=+\infty}\int_{y_{j+1}=-\infty}^{y_{j+1}=\tau_{j+1}^{(1)}-y_j-S_{j-1}} \phi\left(\frac{\tau_{j+2}^{(2)}-S_{j-1}-y_j-y_{j+1}-(|I_{j+2}|-|I_{j+1}|)\frac{S_{j-1}}{|I_{j-1}|}}{\sigma\sqrt{(|I_{j+2}|-|I_{j+1}|)\frac{|I_{j-1}|+1}{|I_{j-1}|}}}\right) \\
&\qquad\qquad dF_{y_{j+1}}(y_{j+1})dF_{y_j}(y_j) \\
&+\int_{y_{j+1}=\tau_{j+1}^{(1)}-(\tau_j^{(1)}-S_{j-1})-S_{j-1}}^{y_{j+1}=+\infty} \Phi\left(\frac{\tau_{j+2}^{(3)}-S_{j-1}-(\tau_j^{(1)}-S_{j-1})-y_{j+1}-(|I_{j+2}|-|I_{j+1}|)\frac{S_{j-1}}{|I_{j-1}|}}{\sigma\sqrt{(|I_{j+2}|-|I_{j+1}|)\frac{|I_{j-1}|+1}{|I_{j-1}|}}}\right) dF_{y_{j+1}}(y_{j+1})\cdot \mathrm{pdf}_{y_j}(\tau_j^{(1)}-S_{j-1}) \\
&+\int_{y_j=\tau_j^{(1)}-S_{j-1}}^{y_j=+\infty} \Phi\left(\frac{\tau_{j+2}^{(3)}-S_{j-1}-y_j-(\tau_{j+1}^{(1)}-y_j-S_{j-1})-(|I_{j+2}|-|I_{j+1}|)\frac{S_{j-1}}{|I_{j-1}|}}{\sigma\sqrt{(|I_{j+2}|-|I_{j+1}|)\frac{|I_{j-1}|+1}{|I_{j-1}|}}}\right) \mathrm{pdf}_{y_{j+1}}(\tau_{j+1}^{(1)}-y_j-S_{j-1})dF_{y_j}(y_j) \\
&-\frac{1+\frac{|I_{j+2}|-|I_{j+1}|}{|I_{j-1}|}}{\sigma\sqrt{(|I_{j+2}|-|I_{j+1}|)\frac{|I_{j-1}|+1}{|I_{j-1}|}}}\int_{y_j=\tau_j^{(1)}-S_{j-1}}^{y_j=+\infty}\int_{y_{j+1}=\tau_{j+1}^{(1)}-y_j-S_{j-1}}^{y_{j+1}=+\infty} \phi\left(\frac{\tau_{j+2}^{(3)}-S_{j-1}-y_j-y_{j+1}-(|I_{j+2}|-|I_{j+1}|)\frac{S_{j-1}}{|I_{j-1}|}}{\sigma\sqrt{(|I_{j+2}|-|I_{j+1}|)\frac{|I_{j-1}|+1}{|I_{j-1}|}}}\right) \\
&\qquad\qquad dF_{y_{j+1}}(y_{j+1})dF_{y_j}(y_j),
\end{aligned}
$$

which is negative for all S_{j-1}, because $\tau_{j+2}^{(1)} > \tau_{j+2}^{(2)} > \tau_{j+2}^{(3)}$, and therefore each of the pairs of terms in lines 1-7, 2-5, 4-10, 8-11 adds up to a negative number.

We can proceed in a similar fashion to show that the derivative of the entire left-hand side of expression (B.2) with respect to S_{j-1} is negative.

[1]In the expression that follows, the first three lines are the derivative of the term in line 5 of (B.2), the next three lines are the derivative of the term in line 6 of (B.2), and so on; we use Leibniz's rule for differentiation under the integral sign.

For $\sum_{x_i \in I_1} x_i < \tau_1, \sum_{x_i \in I_2} x_i < \tau_2^{(1)}, \ldots, \sum_{x_i \in I_{j-2}} x_i < \tau_{j-2}^{(1)}, \sum_{x_i \in I_{j-1}} x_i > \tau_{j-1}^{(1)}$, the expected payoff to agent $j-1$ of taking the risky action has to be negative. Let the unique solution of this inequality be $\sum_{x_i \in I_{j-1}} x_i > b_{j-1}^{(2)}$.

Similarly, for $\sum_{x_i \in I_1} x_i < \tau_1, \sum_{x_i \in I_2} x_i < \tau_2^{(1)}, \ldots, \sum_{x_i \in I_{j-2}} x_i > \tau_{j-2}^{(1)}, \sum_{x_i \in I_{j-1}} x_i < \tau_{j-1}^{(2)}$, the expected payoff to agent $j-1$ of taking the risky action has to be positive; and for $\sum_{x_i \in I_1} x_i < \tau_1, \sum_{x_i \in I_2} x_i < \tau_2^{(1)}, \ldots, \sum_{x_i \in I_{j-2}} x_i > \tau_{j-2}^{(1)}, \sum_{x_i \in I_{j-1}} x_i > \tau_{j-1}^{(2)}$, the expected payoff to agent $j-1$ of taking the risky action has to be negative; and so on.

There is a total of $2(j-1)$ inequalities, with solutions

$$\begin{aligned}
\sum_{x_i \in I_{j-1}} x_i &< b_{j-1}^{(1)} \\
\sum_{x_i \in I_{j-1}} x_i &> b_{j-1}^{(2)} \\
\sum_{x_i \in I_{j-1}} x_i &< b_{j-1}^{(3)} \\
\sum_{x_i \in I_{j-1}} x_i &> b_{j-1}^{(4)} \\
&\vdots \\
\sum_{x_i \in I_{j-1}} x_i &< b_{j-1}^{(2(j-1)-1)} \\
\sum_{x_i \in I_{j-1}} x_i &> b_{j-1}^{(2(j-1))},
\end{aligned}$$

with $b_{j-1}^{(1)} > b_{j-1}^{(2)} > \ldots > b_{j-1}^{(2(j-1))}$, because of Equation (B.1). It follows that any collection of thresholds $\tau_{j-1}^{(1)}, \ldots, \tau_{j-1}^{(j-1)}$ satisfying

$$b_{j-1}^{(1)} \geq \tau_{j-1}^{(1)} \geq b_{j-1}^{(2)} > b_{j-1}^{(3)} \geq \tau_{j-1}^{(2)} \geq b_{j-1}^{(4)} > \ldots > b_{j-1}^{(2(j-1)-1)} \geq \tau_{j-1}^{(j-1)} \geq b_{j-1}^{(2(j-1))}$$

can be sustained at an equilibrium. Multiplicity of Bayesian Nash equilibria follows, establishing the lemma. □

Bibliography

Abrahao, B., F. Chierichetti, R. Kleinberg, and A. Panconesi (2013), "Trace complexity of network inference." In *Proceedings of the 19th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 491–499.

Acemoglu, Daron, Munther Dahleh, Ilan Lobel, and Asuman Ozdaglar (2011), "Bayesian learning in social networks." *The Review of Economic Studies*, 78, 1201–1236.

Ali, Amir Hatem (2011), "The power of social media in developing nations: New tools for closing the global digital divide and beyond." *Harvard Human Rights Journal*, 24, 185–219.

Anandkumar, A., V.Y.F Tan, F. Huang, and A.S. Willsky (2012), "High-dimensional structure learning of Ising models: Local separation criterion." *Annals of Statistics*, 40, 1771–1812.

Angeletos, George-Marios, Christian Hellwig, and Alessandro Pavan (2006), "Signaling in a global game: Coordination and policy traps." *Journal of Political Economy*, 114, 452–484.

Angeletos, George-Marios, Christian Hellwig, and Alessandro Pavan (2007), "Dynamic global games of regime change: Learning, multiplicity, and the timing of attacks." *Econometrica*, 75, 711–756.

Angeletos, George-Marios and Ivan Werning (2006), "Crises and prices: Information aggregation, multiplicity, and volatility." *The American Economic Review*, 96, 1720–1736.

Blume, L. E., W. A. Brock, S. N. Durlauf, and Y. M. Ioannides (2011), *Identification of social interactions*, volume 1B of *Handbook of Social Economics*, Chapter 18, 853–964. Elsevier B.V., The Netherlands: North-Holland.

Bramoullé, Y., H. Djebbari, and B. Fortin (2009), "Identification of peer effects through social networks." *Journal of Econometrics*, 150, 41–55.

Brock, W. A. and S. N. Durlauf (2001), *Interaction-based models*, first edition, volume 5 of *Handbook of Econometrics*, Chapter 54, 3297–3380. Elsevier, Amsterdam: North-Holland.

Carlsson, Hans and Eric van Damme (1993), "Global games and equilibrium selection." *Econometrica*, 61, 989–1018.

Chandrasekaran, V., P. A. Parrilo, and A. S. Willsky (2012), "Latent variable graphical model selection via convex optimization." *Annals of Statistics*, 40, 1935–1967.

Choi, M.J., V. Tan, A. Anandkumar, and A. Willsky (2011), "Learning latent tree graphical models." *Journal of Machine Learning Research*, 12, 1771–1812.

Chwe, Michael Suk-Young (2000), "Communication and coordination in social networks." *The Review of Economic Studies*, 67, 1–16.

Columbia Neurosurgeons, Columbia University Medical Center, Department of Neurosurgery (2014a), "Mesial temporal sclerosis." Website. http://www.columbianeurosurgery.org/conditions/mesial-temporal-sclerosis/.

Columbia Neurosurgeons, Columbia University Medical Center, Department of Neurosurgery (2014b), "Temporal lobe epilepsy." Website. http://www.columbianeurosurgery.org/conditions/temporal-lobe-epilepsy/.

Cover, T. M. and J. A. Thomas (2006), *Elements of Information Theory*, second edition. Wiley-Interscience, New Jersey.

Dahleh, Munther A., Alireza Tahbaz-Salehi, John N. Tsitsiklis, and Spyros I. Zoumpoulis (2011), "On global games in social networks of information exchange." In *Proceedings of the Sixth Workshop on the Economics of Networks, Systems, and Computation (NetEcon), ACM Federated Computing Research Conference*.

Dahleh, Munther A., Alireza Tahbaz-Salehi, John N. Tsitsiklis, and Spyros I. Zoumpoulis (2013a), "Coordination with local information." In *ACM Performance Evaluation Review, Proceedings of the joint Workshop on Pricing and Incentives in Networks and Systems (W-PIN+NetEcon), ACM Sigmetrics*.

Dahleh, Munther A., Alireza Tahbaz-Salehi, John N. Tsitsiklis, and Spyros I. Zoumpoulis (2013b), "Coordination with local information." Submitted.

Dahleh, Munther A., John N. Tsitsiklis, and Spyros I. Zoumpoulis (2014), "The value of temporal data for learning of influence networks." Working paper.

Dasgupta, Amil (2007), "Coordination and delay in global games." *Journal of Economic Theory*, 134, 195–225.

Dessein, Wouter, Andrea Galeotti, and Tano Santos (2013), "Rational inattention and organizational focus." CEPR Discussion Paper No. DP9395.

Du, N., L. Song, A. Smola, and M. Yuan (2012), "Learning networks of heterogeneous influence." *Advances in Neural Information Processing Systems*, 25.

Durlauf, S. N. and Y. M. Ioannides (2010), "Social interactions." *Annual Review of Economics*, 2, 451–478.

Edmond, Chris (2013), "Information manipulation, coordination and regime change." Working paper.

Frankel, D., S. Morris, and A. Pauzner (2003), "Equilibrium selection in global games with strategic complementarities." *Journal of Economic Theory*, 108, 1–44.

Galeotti, A., S. Goyal, M. O. Jackson, F. Vega-Redondo, and L. Yariv (2010), "Network games." *Review of Economic Studies*, 77, 218–244.

Galeotti, Andrea, Christian Ghiglino, and Francesco Squintani (forthcoming), "Strategic information transmission networks." *Journal of Economic Theory*.

Glover, Brent and Seth Richards-Shubik (2013), "Sovereign debt crises and financial contagion." Carnegie Mellon University, preliminary draft.

Goldstein, I. and A. Pauzner (2005), "Demand deposit contracts and the probability of bank runs." *Journal of Finance*, 60, 1293–1328.

Gomez-Rodriguez, M., D. Balduzzi, and B. Schölkopf (2011), "Uncovering the temporal dynamics of diffusion networks." In *Proceedings of the 28th International Conference on Machine Learning*.

Gomez-Rodriguez, M., J. Leskovec, and A. Krause (2010), "Inferring networks of diffusion and influence." In *Proceedings of the 16th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 1019–1028.

Gripon, V. and M. Rabbat (2013), "Reconstructing a graph from path traces." In *Proceedings of 2013 IEEE International Symposium on Information Theory*.

Hellwig, Christian (2002), "Public information, private information, and the multiplicity of equilibria in coordination games." *Journal of Economic Theory*, 107, 191–222.

Kempe, D., J. Kleinberg, and E. Tardos (2003), "Maximizing the spread of influence through a social network." In *Proceedings of the 9th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 137–146.

Kephart, J. O. and S. R. White (1991), "Directed-graph epidemiological models of computer viruses." In *Proceedings of IEEE Symposium on Security and Privacy*, 343–359.

Kolar, M., L. Song, A. Ahmed, and E.P. Xing (2010), "Estimating time-varying networks." *Annals of Applied Statistics*, 4, 94–123.

Krantz, S. G. and H. R. Parks (2002), *The Implicit Function Theorem: History, Theory, and Applications*, first edition. Birkhauser, Boston, Massachusetts.

Lerman, K. and R. Ghosh (2010), "Information contagion: An empirical study of the spread of news on Digg and Twitter social networks." In *Proceedings of the International AAAI Conference on Weblogs and Social Media*.

Lozano, A. C. and V. Sindhwani (2010), "Block variable selection in multivariate regression and high-dimensional causal inference." *Advances in Neural Information Processing Systems*, 23.

Manski, C. (1993), "Identification of endogenous social effects: The reflection problem." *Review of Economic Studies*, 60, 531–542.

Materassi, D. and M.V. Salapaka (2012), "On the problem of reconstructing an unknown topology via locality properties of the Wiener filter." *IEEE Transactions on Automatic Control*, 57, 1765–1777.

Mathevet, Laurent (2012), "Beliefs and rationalizability in games with complementarities." Working paper.

Milgrom, Paul and John Roberts (1990), "Rationalizability, learning, and equilibrium in games with strategic complementarities." *Econometrica*, 58, 1255–1277.

Morris, Stephen and Hyun Song Shin (1998), "Unique equilibrium in a model of self-fulfilling currency attacks." *The American Economic Review*, 88, 587–597.

Morris, Stephen and Hyun Song Shin (2003), *Global Games: Theory and Applications*, 56–114. Advances in Economics and Econometrics, Cambridge University Press, Cambridge, United Kingdom.

Morris, Stephen and Hyun Song Shin (2004), "Coordination risk and the price of debt." *European Economic Review*, 48, 133–153.

Morris, Stephen and Hyun Song Shin (2007), "Common belief foundations of global games." Working paper.

Myers, S. and J. Leskovec (2010), "On the convexity of latent social network inference." *Advances in Neural Information Processing Systems*, 23.

Netrapalli, P. and S. Sanghavi (2012), "Finding the graph of epidemic cascades." In *Proceedings of the 12th ACM SIGMETRICS/PERFORMANCE Joint International Conference on Measurement and Modeling of Computer Systems*, 211–222.

Newey, W. K. and D. McFadden (1994), *Large sample estimation and hypothesis testing*, volume 4 of *Handbook of Econometrics*, Chapter 35, 2111–2245. Elsevier Science.

Newman, M. E. J. (2002), "The spread of epidemic disease on networks." *Physical Review E*, 66, 016128.

Pan, W., N. Aharony, and A. "S." Pentland (2011), "Composite social network for predicting mobile apps installation." In *Proceedings of 25th Conference on Artificial Intelligence, AAAI*.

Penta, Antonio (2012), "Higher order uncertainty and information: Static and dynamic games." *Econometrica*, 80, 631–660.

Shah, D. and T. Zaman (2011), "Rumors in a network: Who's the culprit?" *IEEE Transactions on Information Theory*, 57, 5163–5181.

Song, Le, Mladen Kolar, and Eric P. Xing (2009), "Keller: Estimating time-varying interactions between genes." *Bioinformatics*, 25, i128–i136.

Van Zandt, Timothy and Xavier Vives (2007), "Monotone equilibria in Bayesian games of strategic complementarities." *Journal of Economic Theory*, 134, 339–360.

Yildiz, Muhamet and Jonathan Weinstein (2007), "A structure theorem for rationalizability with application to robust predictions of refinements." *Econometrica*, 75, 365–400.

Zhou, Z., R. Bandari, J. Kong, H. Qian, and V. Roychowdhury (2010), "Information resonance on Twitter: Watching Iran." In *Proceedings of the First Workshop on Social Media Analytics, ACM*, 123–131.