# Statistical Signal Processing on Dynamic Graphs with Applications in Social Networks 

by<br>Maziyar Hamdi<br>M.A.Sc., The University of British Columbia, 2010

A THESIS SUBMITTED IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE DEGREE OF

DOCTOR OF PHILOSOPHY
in

The Faculty of Graduate and Postdoctoral Studies
(Electrical and Computer Engineering)

THE UNIVERSITY OF BRITISH COLUMBIA
(Vancouver)
December 2015
(c) Maziyar Hamdi 2015


#### Abstract

Due to the proliferation of social networks and their significant effects on our day-to-day activities, there has been a growing interest in modeling and analyzing behavior of agents in social networks over the past decade. The unifying theme of this thesis is to develop a set of mathematical theories and algorithmic tools for different estimation and sensing problems over graphs with applications to social networks.

The first part of this dissertation is devoted to multi-agent Bayesian estimation and learning problem in social networks. We consider a set of agents that interact over a network to estimate an unknown parameter called state of nature. As a result of the recursive nature of Bayesian models and the correlation introduced by the structure of the underlying communication graph, information collected by one agent can be mistakenly considered independent, that is, mis-information propagation, also known as data incest arises. This part presents data incest removal algorithms that ensure complete mitigation of the mis-information associated with the estimates of agents in two different information exchange patterns: First, a scenario where beliefs (posterior distribution of state of nature) are transmitted over the network. Second, a social learning context where agents map their local beliefs into a finite set of actions and broadcast their actions to other agents. We also present a necessary and sufficient condition on the structure of information flow graph to mitigate mis-information propagation.

The second part of the thesis considers a Markov-modulated duplication-deletion random graph where at each time instant, one node can either join or leave the network; the probabilities of joining or leaving evolve according to the realization of a finite state Markov chain. This part presents two results. First, motivated by social network applications, the asymptotic behavior of the degree distribution is analyzed. Second, a stochastic approximation algorithm is presented to track empirical degree distribution as it evolves over time. The tracking performance of the algorithm is analyzed in terms of mean square error and a functional central limit theorem is presented for the asymptotic tracking error.


## Preface

The work presented in this thesis is based on the research and development conducted in the Statistical Signal Processing Laboratory at the University of British Columbia (Vancouver). The research work presented in the chapters of this dissertation is performed by the author with feedback and assistance provided by Prof. Vikram Krishnamurthy. The author is responsible for writeup, problem formulation, research development, data analyses and numerical studies presented in this dissertation with frequent suggestions, technical and editorial feedback from Prof. Vikram Krishnamurthy. Hilbert space analysis of Chapter 4 are due in part to Prof. George Yin. The dataset for the psychology experiment presented in Chapter 3 is obtained from Prof. Alan Kingstone and Dr. Grayden Solman. For the psychology experiment, informed consent was obtained from all participants, and all experimental procedures and protocols were reviewed and approved by the University of British Columbia Behavioral Research Ethics Board (UBC BREB). The UBC BREB approval number is: H10-00527. The work presented in different chapters of this thesis has been appeared in several publications which are listed below. In these publications, all co-authors contributed to the editing of the manuscript.

- The work of Chapter 2 has been presented in the following publications:
- [Journal Paper] V. Krishnamurthy, M. Hamdi, Mis-information Removal Algorithms on Social Networks: Constrained Estimation on Random Graphs, IEEE Journal of Selected Topics in Signal Processing, vol:7, issue: 2, pp.333-346.
- [Conference Paper] V. Krishnamurthy and M. Hamdi, Data Fusion and Mis-information Removal in Social Networks, IEEE conference on Information Fusion, 2012, Singapore, July 2012.
- Materials in Chapter 3 have been appeared in the following publications and preprints for possible publications:
- [Book] V. Krishnamurthy, O. N. Gharehshiran, and M. Hamdi. Interactive Sensing and Decision Making in Social Networks. Now Publishers Inc., Hanover, MA, 2014.
- [Journal Paper] M. Hamdi, V. Krishnamurthy Removal of data incest in multi-agent social learning in social networks, preprint: arXiv:1309.6687.
- [Report] M Hamdi, G Solman, A Kingstone, V Krishnamurthy Social learning in a human society: An experimental study, preprint: arXiv:1408.5378.
- The work of Chapter 4 has been presented in the following publications:
- [Book] V. Krishnamurthy, O. N. Gharehshiran, and M. Hamdi. Interactive Sensing and Decision Making in Social Networks. Now Publishers Inc., Hanover, MA, 2014.
- [Journal Paper] M. Hamdi, V. Krishnamurthy, G. Yin Tracking a Markov-Modulated Stationary Degree Distribution of a Dynamic Random Graph, IEEE Transaction on Information Theory, vol 60, issue 10, pp. 6609-6625.
- [Conference Paper] M. Hamdi and V. Krishnamurthy, A Novel Use of Stochastic Approximation Algorithms for Estimating Degree of Each Node in Social Networks, IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP'12), Kyoto, Japan, March 2012.
- [Conference Paper] M.Hamdi, V. Krishnamurthy, The Asymptotic of Duplication-deletion Random Graphs, accepted for publication in 44th Asilomar Conference on Signals, Systems and Computers. Nov. 2010.
- Although not being presented in this thesis, the discussion and results about social networks that has appeared in the following journal article was inspired by the work presented in this thesis:
- [Journal Paper] A. Leshem, M. Hamdi, V. Krishnamurthy, Boundary Value Problems in Consensus Networks, Submitted to IEEE Transactions on Signal Processing, arXiv preprint arXiv:1407.7170.


## Table of Contents

Abstract ..... ii
Preface ..... iii
Table of Contents ..... v
List of Tables ..... ix
List of Figures ..... X
Acknowledgments ..... xiv
Dedication ..... xv
1 Introduction ..... 1
1.1 Overview ..... 1
1.1.1 Bayesian Estimation over Social Networks ..... 3
1.1.2 Interactive Social Learning over Networks ..... 5
1.1.3 Tracking Degree Distribution of Social Networks ..... 6
1.2 Main Contributions ..... 9
1.2.1 Bayesian Estimation over Social Networks ..... 9
1.2.2 Interactive Social Learning over Networks ..... 9
1.2.3 Tracking Degree Distribution of Social Networks ..... 10
1.3 Related Works ..... 11
1.3.1 Bayesian Estimation over Social Networks ..... 12
1.3.2 Interactive Social Learning over Networks ..... 12
1.3.3 Tracking Degree Distribution of Social Networks ..... 14
1.4 Thesis Outline ..... 15
Part I Estimation and Learning Over Directed Acyclic Graphs ..... 18
2 Constrained Estimation Over Random Graphs ..... 19
2.1 Introduction ..... 19
2.1.1 Chapter Goals ..... 21
2.1.2 Main Results and Organization of Chapter ..... 22
2.2 Modeling Information Flow in Social Networks ..... 22
2.2.1 Constrained Information Flow Protocol ..... 23
2.2.2 Benchmark Full Information Flow Protocol ..... 24
2.2.3 Modeling Time Evolution of the Information Flow ..... 25
2.3 Optimal Mis-information Propagation Removal Algorithm ..... 25
2.3.1 Optimal Combination Scheme in Constrained Information Flow Protocol ..... 26
2.4 Sub-optimal Mis-information Removal Algorithm Without Complete Knowledge of Information Flow Graph ..... 29
2.4.1 Sub-optimal Combination Scheme ..... 30
2.5 Numerical Examples ..... 30
2.5.1 Example and Intuition on Theorems 2.2.1 and 2.3.2 ..... 31
2.5.2 Numerical Examples Illustrating Alg. $\mathscr{A}$ in Estimation Problem (2.5) ..... 32
2.5.3 Numerical Examples Illustrating Alg. $\mathscr{B}$ in Estimation Problem (2.14) ..... 34
2.6 Closing Remarks ..... 38
2.7 Proof of Theorems ..... 39
2.7.1 Proof of Theorem 2.2.1 ..... 40
2.7.2 Proof of Theorem 2.3.1 ..... 40
2.7.3 Proof of Theorem 2.3.2 ..... 41
3 Mis-information Management Problem in Social Learning Over Directed Graphs ..... 42
3.1 Introduction ..... 42
3.1.1 Social Learning Protocol on Network ..... 43
3.1.2 Chapter Goals ..... 45
3.1.3 Main Results and Organization of Chapter ..... 47
3.2 Social Learning Over Social Networks ..... 48
3.2.1 Constrained Social Learning in Social Networks ..... 49
3.3 Data Incest Removal Algorithm ..... 51
3.3.1 The Idealized Benchmark for Data Incest Free Social Learning in Social Networks ..... 51
3.3.2 The Data Incest Free Belief in the Idealized Social Learning Protocol 2 ..... 52
3.3.3 Data Incest Removal Algorithm for Problem (3.9) With Constrained Social Learning Protocol 1 ..... 53
3.3.4 Discussion of Data Incest Removal in Social Learning ..... 57
3.4 Numerical Examples ..... 58
3.5 Psychology Experiment ..... 65
3.5.1 Experiment Setup ..... 65
3.5.2 Experimental Results ..... 68
3.6 Closing Remarks ..... 71
3.7 Proof of Results ..... 73
3.7.1 Proof of Lemma 3.2.1 ..... 73
3.7.2 Proof of Theorem 3.3.1 ..... 75
3.7.3 Proof of Theorem 3.3.2 ..... 76
Part II Tracking Degree Distribution in Dynamic Social Networks ..... 78
4 Tracking a Markov Modulated Degree Distribution ..... 79
4.1 Introduction ..... 79
4.1.1 Chapter Goals ..... 79
4.1.2 Main Results and Organization of Chapter ..... 81
4.2 Markov-modulated Dynamic Random Graph of Duplication-deletion Type ..... 82
4.3 Asymptotic Degree Distribution Analysis for Non-Markov Modulated case ..... 84
4.3.1 Fixed Size Random Graph ..... 84
4.3.2 Power Law Exponent for Infinite Duplication-deletion Random Graph ..... 87
4.4 Tracking the Degree Distribution of the Fixed Size Markov-modulated Random Graph ..... 88
4.4.1 Tracking Error of the Stochastic Approximation Algorithm ..... 91
4.4.2 Limit System of Regime-Switching Ordinary Differential Equations ..... 92
4.4.3 Scaled Tracking Error ..... 92
4.5 Estimating the Degree Distribution of Infinite Duplication-deletion Random Graphs ..... 93
4.5.1 Infinite Random Graphs without Markovian Dynamics ..... 93
4.5.2 Markov-modulated Probability Mass Functions with Denumerable Support ..... 96
4.6 Numerical Examples ..... 98
4.7 Closing Remarks ..... 102
4.8 Proof of Results ..... 104
4.8.1 Proof of Theorem 4.3.1 ..... 104
4.8.2 Proof of Theorem 4.3.2 ..... 109
4.8.3 Proof of Lemma 4.8.1 ..... 112
4.8.4 Proof of Lemma 4.8.2 ..... 112
4.8.5 Proof of Theorem 4.4.1 ..... 116
4.8.6 Sketch of the Proof of Theorem 4.4.2 ..... 118
4.8.7 Sketch of the Proof of Theorem 4.4.3 ..... 119
4.8.8 Proof of Theorem 4.5.2 ..... 120
4.8.9 Proof of Theorem 4.5.3 ..... 122
4.8.10 Proof of Theorem 4.5.4 ..... 125
5 Conclusions ..... 127
5.1 Summary of Findings in Part I ..... 127
5.2 Summary of Findings in Part II ..... 129
5.3 Directions for Future Research and Development ..... 130
Bibliography ..... 133
A Some Graph Theoretic Definitions ..... 145
Appendices ..... 145
B A Note on Degree-based Graph Construction ..... 147

## List of Tables

3.1 The frequency of the internal and the boundary nodes in a community of 3316 undergraduate students of the University of British Columbia along with the statistics of the time required by participants (of both types) to make their judgments in milliseconds.

## List of Figures

1.1 Example of a network of six agents (social sensors) aims to estimate a parameter
(state of nature) interactively; each edge depicts a communication link between two
sensors. ..... 4
1.2 Tracking the underlying state of nature using a Markov-modulated random graph as a social sensor. ..... 8
1.3 Main results and organization of thesis. ..... 16
2.1 Example of constrained information flow network, $S=2$ and $K=3$. Circles repre- sent a social group at a specific time indexed by (2.4) in the social network and each edge depicts a communication link between two nodes. ..... 31
2.2 The conditional mean of the state of nature given the observation in estimation with optimal mis-information removal algorithm compared to the full information net- work. ..... 33
2.3 Comparison of the the mean squared errors of the estimates obtained by optimal mis-information removal algorithm, Bayesian estimator in full information flow network (free of mis-information), and standard Bayesian estimator in constrained information flow network (with mis-information propagation). ..... 35
2.4 Comparison of the conditional mean of the state of nature $x$ given the observa- tions obtained by sub-optimal mis-information removal algorithm, optimal mis- information removal algorithm (knowing the exact information flow graph), Bayesian estimator in full information flow network (free of mis-information), and standard Bayesian estimator in constrained information flow network (with mis-information propagation) in "accurate estimation" scenario ( $\beta=0.2$ ). ..... 36
2.5 Comparison of the mean squared errors of the estimates obtained by sub-optimalmis-information removal algorithm, optimal mis-information removal algorithm (know-ing the exact information flow graph), Bayesian estimator in full information flownetwork (free of mis-information), and standard Bayesian estimator in constrainedinformation flow network (with mis-information propagation) in "accurate estima-tion" scenario ( $\beta=0.2$ ).37
2.6 Comparison of the conditional mean of the state of nature $x$ given the observa- tions obtained by sub-optimal mis-information removal algorithm, optimal mis- information removal algorithm (knowing the exact information flow graph), Bayesian estimator in full information flow network (free of mis-information), and standard Bayesian estimator in constrained information flow network (with mis-information propagation) in "inaccurate estimation" scenario ( $\beta=0.8$ ). ..... 38
2.7 Comparison of the the mean squared errors of the estimates obtained by sub-optimal mis-information removal algorithm, optimal mis-information removal algorithm (know- ing the exact information flow graph), Bayesian estimator in full information flow network (free of mis-information), and standard Bayesian estimator in constrained information flow network (with mis-information propagation) "inaccurate estima- tion" scenario ( $\beta=0.8$ ). ..... 39
3.1 Two examples of multi-agent social learning in social networks: (i) target localiza- tion, and (ii) online rating and review systems. ..... 46
3.2 Example of communication graph, with two agents $(S=2)$ and over three event epochs ( $K=3$ ). The arrows represent exchange of information regarding actions taken by agents. ..... 47
3.3 Protocol 1: Constrained social learning in social networks described in Section 3.1.1.
As a result of random (unknown) communication delays, data incest arises. ..... 49
3.4 Protocol 2: Idealized benchmark social learning in social networks. In this protocol, the complete history of actions chosen by agents and the communication graph are known. Hence, data incest does not arise. This benchmark protocol will be used to design the data incest removal protocol. ..... 52
3.5 Two examples of networks: (a) satisfies the topological constraint, and (b) does not satisfy the topological constraint. ..... 57
3.6 Data incest removal algorithm employed by network administrator in the state es- timation problem over social network. The underlying state of nature could be ge- ographical coordinates of an event (target localization problem) or reputation of a social unit (online rating and review systems). ..... 58
3.7 Three different communication topologies: (a) the communication graph with 41 nodes, (b) agents interact on a fully interconnected graph and the information from one agent reach other agents after a delay chosen randomly from $\{1,2\}$ with the same probabilities, (c) star-shaped communication topology with random delay cho- sen from $\{1,2\}$ ..... 59
3.8 Actions of agents obtained with social learning over social networks in three differ- ent scenarios described in Section 3.4 with communication graph depicted in Fig.3.7a. 61
3.9 Mean of the estimated state of nature in the state estimation problem with social learning over social networks in three different scenarios described in Section 3.4 with communication graph depicted in Fig.3.7a. ..... 61
3.10 Actions of agents obtained with social learning over social networks in three differ- ent scenarios described in Section 3.4 with communication graph depicted in Fig.3.7c. 62
3.11 Mean of the estimated state of nature in the state estimation with social learning over social networks in three different scenarios described in Section 3.4 with com- munication graph depicted in Fig.3.7c. ..... 62
3.12 Actions of agents obtained with social learning over social networks in three differ- ent scenarios described in Section 3.4 with communication graph depicted in Fig.3.7b. 63
3.13 Mean of the estimated state of nature in the state estimation problem with social learning over social networks in three different scenarios described in Section 3.4 with communication graph depicted in Fig.3.7b. ..... 63
3.14 Actions of agents obtained with social learning over social networks in three differ- ent scenarios described in Section 3.4 with arbitrary communication graph. ..... 64
3.15 Mean of the estimated state of nature in the state estimation problem with social learning over social networks in three different scenarios described in Section 3.4 with arbitrary communication graph. ..... 64
3.16 Mean squared error of estimates (of state of nature) obtained with social learning with communication graph depicted in Fig.3.7a. ..... 66
3.17 Mean squared error of estimates (of state of nature) obtained with social learning with communication graph depicted in Fig.3.7b (complete fully interconnected graph). 66
3.18 Mean squared error of estimates (of state of nature) obtained with social learning with communication graph depicted in Fig.3.7c (star-shaped communication graph). ..... 67
3.19 Mean squared error of estimates (of state of nature) obtained with social learning with arbitrary communication graph. ..... 67
3.20 Two arrays of circles were given to each pair of participants on a screen. Their task is to interactively determine which side (either left or right) had the larger average diameter. The partner's previous decision was displayed on screen prior to the stimulus. ..... 69
3.21 Actions of two participants in a group at different trials in one experiment. ..... 70
3.22 Two scenarios where data incest arose in our experimental studies. ..... 71
3.23 Social learning with data incest that is exercised by groups of students who were asked to perform a conceptual task in our experimental study. ..... 72
3.24 Actions of two participants in a group at different epochs. Participant 1 can be considered as an internal node and Participant 2 can be viewed as a boundary node. ..... 73
3.25 Optimal weights (which depends on the topology of the communication graph) and set of available public belief are computed in separate units. The user, then, can compute the most updated data-incest free network belief. ..... 74
4.1 The power law component for the non-Markovian random graph generated accord- ing to Procedure 4.5 obtained by (4.11) for different values of $p$ and $q$ in Proce- dure 4.5. ..... 89
4.2 Degree distribution of the duplication-deletion random graph satisfies a power law. The parameters are specified in Example 4.6.1 of Section 4.6. ..... 100
4.3 Degree distribution of the fixed size duplication-deletion random graph. The pa- rameters are specified in Example 4.6.2 of Section 4.6. ..... 100
4.4 Degree distribution of the fixed size duplication-deletion random graph satisfies a power law when $N_{0}$ is sufficiently large. The parameters are specified in Exam- ple 4.6.2 of Section 4.6. ..... 101
4.5 The estimates obtained by SA algorithm (4.14) follows the expected PMF precisely with no knowledge of the Markovian dynamics. The parameters are specified in Example 4.6.3 ..... 101
4.6 The average degree of nodes (as a measure of connectivity) of the fixed size Markov- modulated duplication-deletion random graph obtained by Procedure 4.5 for differ- ent values of the probability of connection, $p$, in Algorithm 4.5. The parameters are specified in Example 4.6.4 of Section 4.6. ..... 102
4.7 Trace of the covariance matrix of scaled tracking error, trace $(\Sigma(\theta))$, versus the av- erage degree of nodes as a measure of connectivity of the network. The parameters are specified in Example 4.6.3 of Sec.4.6. ..... 103
4.8 Trace of the covariance matrix of the scaled tracking error, trace $(\Sigma(\theta))$, versus the order of delay in the searching problem as a measure of searchability of the network. The parameters are specified in Example 4.6.3 of Sec.4.6. ..... 103

## Acknowledgments

I owe my sincere gratitude to my academic advisor Prof. Vikram Krishnamurthy. I am deeply indebted to him for every bit of guidance, encouragement and support that I received during the completion of this work. This feat was not possible without his consistent support, fruitful discussions, valuable feedbacks, and constructive suggestions. I hope that someday I could be as enthusiastic, and energetic as him and be able to advise an audience as well as he can.

I would like to express my thanks to Prof. George Yin for his time, effort, suggestions and helpful comments during our collaboration in the past years which resulted in much improvement in our work. I would also like to extend my thanks to my friends and colleagues in the Statistical Signal Processing Lab for creating a friendly and productive environment.

Furthermore, heartfelt and deep thanks go to my family, and particularly my parents, for all the love, patience, sacrifices, assistance, and support. They have always encouraged me and given their unconditional support throughout my studies at the University of British Columbia. Without their support I could not have completed this thesis and it is to them that I dedicate this thesis.

Maziyar Hamdi<br>UBC, Vancouver<br>October 2015

## Dedication

To my beloved family: Mohammadreza, Haydeh, and Mazda.

## Introduction

### 1.1 Overview

Social networks are crucial to the modern society: they permeate our social, personal, and economic lives. They have changed the way that people connect and communicate, vote, select items to purchase, pick hotels to stay in, and adapt a new technology or behavior. For example, in the new "social media" era, Angry Birds ${ }^{1}$ required only 35 days to obtain 50 million users. For comparison purposes and to get a better idea, radio reached the same milestone in 38 years, telephone in 75 years [5]. Social networks facilitate transmission of information, communication, and interaction among people and have grown steadily in size, complexity, and importance [ 28,80 ]. One of the reasons behind this huge growth of social networks, both in science and society, is the concept of a "network" as a structural pattern of interactions among social actors. Such a social structure appears on a wide range of contexts in sociology, biology, economics, computer science and electrical engineering, whenever a set of dyadic ties between objects are observed. In all these topics, an ability to model and analyze those-often huge and complex-social structures, is fundamental for the computational purposes and it answers questions about the behavior of the social actors-individually as an agent or collectively as a team-in a data-driven manner.

As a result of the enormous effects and applications of social networks in community and in science, there has been a growing interest in modeling and analysis of social networks over the past decade. Due to the constraints imposed by the structure of networks and the nature of humans (as interacting sensors), research in social networks requires borrowing techniques from complex networks (dynamics of random graphs) and social analysis (mostly used in the areas of economics and sociology)[80, 146]. Statistical inference using social networks is an area that has witnessed marvelous progress recently. Such systems comprising of humans acting as sensors, also called participatory sensing, has received wide attention recently in the research communities of computer science, economics, marketing, social sciences and electrical engineering. The proliferation of social media such as real-time microblogging services (Twitter ${ }^{2}$ ) and online rating and review systems (Yelp) make real time sensing of social activities, social patterns, and behavior easier. In

[^0]signal processing community, the term social sensor is used to denote an agent (or a group of agents) that provides information about its environment (state of nature) to others (possibly via social media channel) after interaction with other agents in a social network. Examples of such social media channels include Twitter, Facebook, online rating and review systems like Yelp and Tripadvisor, e-commerce platforms such as Amazon. The ability of social sensors in sensing social activities, patterns and behaviors is beyond physical sensors. For example, level of satisfaction of attendees of a concert revealed by sentiment analysis on Facebook statuses and tweets are impossible to predict using physical sensors, or as opposed to physical sensors, social sensors can be used to find out how the quality of food is in a restaurant from the reviews on online rating system such as Yelp [91, 92].

Statistical signal processing in social networks (inference using social sensors) appears in enormous range of applications across different industry sectors including marketing and advertisement, health and medicine, and financial technology. For example, $[38,101]$ used content of tweets to provide Geo-location services which are useful in targeting and event advertising. Other examples include detecting influential users with applications in marketing [144], localization of natural disasters [132], and predicting stock markets [26]. It is shown in [11] that a simple model built from the rate of tweets casted about particular topics can outperform market-based predictors.

With the above applications of social sensors, there is a strong motivation to develop a set of mathematical models and algorithmic tools to understand the effects of interactions among social sensors on estimation problem (interactive sensing). The majority of this thesis is devoted to development of algorithms and procedures that are aimed at multi-agent estimation, tracking, and decision making where agents acts as social sensors of an underlying state of nature in the presence of uncertainty. Such problems are non-standard in two ways: First, in social networks, agents interact with and influence other agents. For example, ratings posted on online rating and review systems strongly influence the behavior of individuals ${ }^{3}$. Such interactions can result in non-standard information patterns as a results of constraints imposed and correlations introduced by the structure of the underlying social network (communication topology among social sensors). Second, due to privacy reasons and time constraints, social sensors typically do not reveal raw observations of the underlying state of nature. Instead, they reveal their beliefs (opinions), or actions (tweet/re-tweet of common trends, thumbs-up on rating and review systems, purchase an item on e-commerce platforms) which can be viewed as a quantized version of their knowledge about state of nature formed by raw measurements and interactions with other social sensors. These together with the uncertainty involved in observations of state of nature, result in non-standard estimation and tracking problem which are the major topics of this thesis. The unifying theme of this thesis is to develop a set of theory and methods for statistical signal processing on graphs (social sensors) which involves adaptive filtering and stochastic approximation, dynamics of random graphs, multi-agent Bayesian estima-

[^1]tion, and social learning to understand how sensors interact. We employ social learning [16, 24, 35], graph theoretic tools, and stochastic approximation $[99,157]$ as useful mathematical abstractions for modeling the interaction of social sensors. The rest of this section is devoted to an overview of these topics along with motivation and research goals that have been addressed in this thesis.

### 1.1.1 Bayesian Estimation over Social Networks

Bayesian filtering, which is a recursive form of the famous Bayes' rule, has been used extensively in the traditional signal processing literature for logical inference on uncertain parameters from a prior knowledge and new observations[21]. What statisticians and mathematicians call "Bayesian theory" was originally developed by Thomas Bayes in a famous paper which was presented at a meeting of the Royal Society of London [18, 138]. The well-known Bayes theorem describes the fundamental probability law in order to perform logical inference. He states[138]
"If there be two subsequent events, the probability of the second $\frac{b}{N}$ and the probability of both together $\frac{P}{N}$, and it being first discovered that the second event has also happened, from hence I guess that the first event has also happened, the probability I am right is $\frac{P}{b}$.,

Bayesian inference is devoted to applying Bayes' rule to statistical inference in order to update the probability of a hypothesis as new observations are obtained [21, 125, 130]. Bayesian models for inference have been applied in decision theory, detection and estimation, communications theory, pattern recognition, machine learning and artificial intelligence, and filtering and parameter estimation [39, 64, 75, 105, 112].

Chapter 2 of this dissertation considers a multi-agent sensing problem where agents (social sensors) interact over a random graph and evaluate their belief (opinion) about an economic or a social parameter, namely state of nature. This state of nature can be, for example, quality of food in a restaurant, occurrence of earthquake, or Geo-location of a target. The evaluation of belief about state of nature can be made using Bayesian [1, 64] or non-Bayesian models [57, 81]. In this work, we focus on Bayesian inference in social networks. In the setup considered in Chapter 2, each agent acts a social sensor: (i) she records her observation about state of nature in uncertainty (by sampling from a conditional probability distribution), (ii) interact with other agents and receives estimates from other agents, (iii) updates her belief using recursive Baysian models and transmits her updated belief over the network, and this repeats on.

As a result of the recursive nature of the Bayesian models and the correlation introduced by the structure of the communication graph between agents, "mis-information propagation" also known as data incest ${ }^{4}$ arises in such networks. In a general sense, mis-information propagation involves in-

[^2]advertent re-use of identical information (which are naively considered to be independent) in formation of belief about state of nature [71, 89]. The following example illustrates how mis-information can propagate and affect the estimates of social sensors in the multi-agent state estimation over graphs.

## Example to Illustrate Mis-information Propagation:

Consider six social sensors that aim to estimate an underlying state of nature interactively with their communication topology illustrated in Figure 1.1. The graph in Figure 1.1 shows the information exchange protocol among social sensors, for example, the link between node 1 and node 3 shows that the estimation of node 1 is available at node 3 . Each node records its own observation and combines it with the available estimates received from other nodes in the network (which are available due to the network structure) in order to form its estimates of state of nature. Then, the updated estimates are broadcasted over the network.


Figure 1.1: Example of a network of six agents (social sensors) aims to estimate a parameter (state of nature) interactively; each edge depicts a communication link between two sensors.

To understand what can go wrong with the above example, note that the estimate of node 1 is used at nodes 3 and 4 , therefore, the estimates of these nodes are both functions of the estimate of node 1 . Thus, if node 5 naively combines the information received from node 3 and 4 , it would have double counted the estimate of node 1 ; this results in mis-information propagation.

## Research Goals

As illustrated with the above example, mis-information propagation (also called data incest) arises in interactive sensing over networks due to the recursive nature of Bayesian estimators and the correlation imposed by the structure of the underlying communication graph among agents (social sensors). Mis-information propagation results in a bias in estimates of agents and over-confidence phenomena, i.e., the variance is underestimated. Therefore, in the presence of mis-information propagation, Bayesian estimators require careful design. Chapter 2 is devoted to design and development of algorithms that mitigate the effect of mis-information propagation in multi-agent estimation problem over the networks.

Remark 1.1.1. Our implicit assumption throughout Chapter 2 is that, agents do not directly broadcast their raw observations over the network. First, (private) raw observations of the other individuals are not typically available because of the privacy concerns or time constraints. For example, let $x \in(\{1,2, \ldots, 5\} \times\{1,2, \ldots, 5\})$ denote quality and affordability of a restaurant ( $x$ is a two dimensional vector). Assume that an individual in social network (social sensor) checks out the restaurant and based on her observations, she estimates the quality and cost of that restaurant (the estimated value of $x$ ). But at the time of her future social interactions, she usually does not provide (or even does not remember) details of the "raw" observations and she only shares her belief with others. Second, the dimension of observations, is typically much larger than the dimension of the state of nature. In the example of restaurant, the observation vector can include many elements such as quality of food, music, lighting, price of food, staff, neighborhood, cost of beverages, etc. Instead of broadcasting all these raw observations, it is more common in social networks to share only the beliefs about quality and cost of that restaurants.

### 1.1.2 Interactive Social Learning over Networks

Social learning which has been applied to understand, model, and predict the behavior of agents in economics, financial markets, political sciences, and social networks [16, 35, 36, 100] seeks to answer the following question: How do actions (decisions) of other agents affect actions of subsequent agents? Social learning model comprises of a set of agents seeking to estimate the underlying state of nature not only from their private observations, but also from the actions of previous agents. All agents know the structure of the model, they know that the action of each agent is a rational response to its private observations, and thus, convey information about state of nature. Social learning can be considered as the diffusion of (private) information about state of nature to all agents through the intercommunication of actions in a set of (finite, or infinite) agents [35].

When a human learns form another person's behavior, decision, or action, social learning occurs. In social learning, the actions of one agent affects the behavior of others, since they know that those actions are motivated by some type of information that other agents have about the state of nature. To better understand this, let's borrow the umbrella example from [35]. When someone sees other people going out with an umbrella, she also takes an umbrella without checking the weather forecast. That happens because people know that the actions or behavior of others have some information about the state of nature; this results in rational herding. Social learning is a useful approximation to ordinal human behavior. Classical social learning is used to model the behavior of expected cost minimizer agents. Also, social learning can be generalized to of risk averse minimizers, see [95, 96].

Chapter 3 considers a social learning model comprising of a set of agents (social sensors) that are interacting over a network to estimate an underlying state of nature. As opposed to the classical social learning model (where agents act once in a pre-determined order), in the social learning model considered in Chapter 3, the structure of social network dictates who interacts with whom.

We use social learning as a mathematical abstraction to model interactions between agents (social sensors) in interactive sensing problem. From a statistical signal processing point of view, the interactive sensing using social learning models-estimating state of nature via social sensors-is non-standard in two ways: First, agents are influenced by the rating of other agents, this can result in interesting phenomenon where rational agents can all end up making the same decision (herding and information cascades; [35]). Second, (and this effect is more complex), an agent might be influenced by his own rating leading to data incest (mis-information propagation). In the following, the main results of Chapter 3 is described briefly.

## Research Goals

As is apparent from the above discussion, mis-information can propagate in interactive sensing with social learning due to the correlations imposed by the structure of the underlying social networks (existence of multiple paths between agents in the graph which represents the underlying social network) and the recursive nature of Bayesian inference, see Section 1.1.1. To give an example, suppose an agent wrote a poor rating of the restaurant on a rating and review site. Another agent is influenced by this rating and also gives the restaurant a poor rating. The first agent visits the rating and review again and finds out that another agent has also given the restaurant a poor rating-this double confirms his rating and he enters another poor rating. In a fair system, the first agent should have been aware that the rating of the second agent was influenced by his rating-so that first agent has effectively double counted his first rating by casting the second poor rating. Mis-information propagation cause the over-confidence problem and results in a bias in the estimate of state of nature.

Motivated by online rating and review systems, our aim is to manage mis-information propagation problems associated with interactive sensing using social learning. In particular, our goal is to design and develop a protocol for the administrator of a rating and review system such that it can automatically maintain a fair (data incest free) rating and review system which results in a system with a higher trust rating.

### 1.1.3 Tracking Degree Distribution of Social Networks

Tracking a time-varying parameter that evolves according to a finite-state Markov chain (state of nature) is a problem of much interest in signal processing [20, 63, 153]. In the context of social networks, this parameter can be, for example, the level of happiness in a community, the tendency of individuals to expand their networks, the strength of social links between individuals, or searchability of network which cannot be sensed by physical sensors. In such cases, social sensors can go beyond physical sensors and can be used to track those parameters of social networks. A social network with a large number of individuals can be viewed as an interactive sensing tool to obtain information about individuals or state of nature; this is a social sensor. Motivated by social network
applications, a social sensor based framework is presented in Chapter 4 to track the degree distribution of Markov-modulated dynamic networks whose dynamics evolve over time according to a finite-state Markov chain.

The question that may arise here is: "What is the importance of degree distribution of a social network?" The most important parameter of a network that characterizes its structure is the degree distribution. It yields useful information about the connectivity of the random graph [10, 86, 116]. For example, if a majority of nodes in the random graph have relatively high degrees, the graph is highly connected and a message can be transferred between two arbitrary nodes with shorter paths. However, if a majority of nodes have smaller degrees then for transmitting a message throughout the network, longer paths are needed, see [80]. The degree distribution can further be used to investigate the diffusion of information or disease through social networks [108, 146]. The existence of a "giant component" ${ }^{5}$ in complex networks can be studied using the degree distribution of the graph that models that social network. The size and existence of a giant component has important implications in social networks in terms of modeling information propagation and spread of human disease, see $[62,115,118]$. The degree distribution is also used to analyze the "searchability" of a network. The "search" problem arises when a specific node in a network faces a problem (request) whose solution is at other node, namely, destination (e.g., delivering a letter to a specific person, or finding a web page with specific information) [4, 146]. The searchability of a social network [146] is the average number of nodes that need to be accessed to reach the destination. Degree distribution is also used to investigate the robustness and vulnerability of a network in terms of the network response to attacks on its nodes or links [33, 76]. The papers [148, 149] further use degree-dependent tools for classification of social networks.

Chapter 4 considers a dynamic social network where the interactions between nodes evolve over time according to a Markov process that undergoes infrequent jumps (the state of nature). An example of such social networks is the friendship network among residents of a city, where the dynamics of the network change in the event of a large festival. In this chapter, we propose Markovmodulated random graphs to mimic social networks where the interactions among nodes evolve over time due to the underlying dynamics (state of nature). For example, state of nature can be level of happiness in the society which is impossible to measure using participatory sensors. Here, social networks can be used as social sensors for tracking the underlying state of nature. That is, using noisy measurements of the degree distribution of the network, the jumps in the underlying state of nature can be tracked.

[^3]

Figure 1.2: Tracking the underlying state of nature using a Markov-modulated random graph as a social sensor.

## Research Goals

Chapter 4 considers a Markov-modulated duplication-deletion random graph where, at each time instant, one node can either join or leave the network with probabilities that evolve according to the realization of a finite state Markov chain (state of nature). This chapter deals with the following questions: How can one estimate the state of nature using noisy observations of nodes' degrees in a social network? and How good are these estimates? By tracking the degree distribution of a Markov-modulated random graph, we can design a social sensor to track the underlying state of nature using the noisy measurements of nodes' connectivity; see Figure 3.20.

Chapter 4 comprises of two results. First, motivated by social network applications, we analyze the asymptotic behavior of the degree distribution of the Markov-modulated random graph. From this degree distribution analysis, we can study the connectivity of the network, the size and the existence of a large connected component, the delay in searching such graphs, etc. [62, 80, 115, 118]. Second, a stochastic approximation algorithm is presented to track the empirical degree distribution as it evolves over time. We further show that the stationary degree distribution of Markov-modulated duplication-deletion random graphs depends on the dynamics of such graphs and, thus, on the state of nature. This means that, by tracking the empirical degree distribution, the social network can be viewed as a social sensor to track the state of nature. The tracking performance of the proposed stochastic approximation algorithm is analyzed in terms of mean square error. A functional central limit theorem is further presented for the asymptotic tracking error.

### 1.2 Main Contributions

In the subsequent sections, a brief summary of major novel contributions of the chapters that constitute this thesis is provided in order that they appear in the thesis. More detailed description of the contributions and findings of each chapter is provided in individual chapters.

### 1.2.1 Bayesian Estimation over Social Networks

As briefly described in Section 1.1.1, Chapter 2 considers multi-agent Bayesian estimation problem with constraints imposed by the structure of the underlying social network. In such problems, as a result of the recursive nature of Bayesian estimators and the correlation imposed by the communication topology of social sensors, mis-information propagation arises.

The main contributions of Chapter 2 are summarized below:

1. Mis-information propagation problem in interactive-sensing problem over social networkswhere agents transmit their beliefs about state of nature instead of raw observations over the network-is mathematically formulated using a family of directed acyclic graphs.
2. A necessary and sufficient condition on the information flow graph ${ }^{6}$ is presented for the exact mis-information removal problem. It is shown in Section 2.3 that under Constraint 2.3.1, mis-information associated with estimates of agents can be completely removed.
3. An optimal information aggregation algorithm is proposed for multi-agent estimation problem over networks which mitigates the mis-information associated with estimates of agents (social sensors) when the information flow graph is known.
4. A sub-optimal mis-information removal algorithm is presented for scenarios where the the information flow graph is not completely known.

### 1.2.2 Interactive Social Learning over Networks

Motivated by online rating and review systems, Chapter 3 considers social learning as a mathematical abstraction to model the interactions among agents (social sensors) in state estimation problem using interactive sensing. Agents record their own private observations, then, update their private beliefs about the state of nature using Bayes' rule. Based on their belief each agent, then, chooses an action (rating) from a finite set and transmits this action over the social network. An important consequence of such social learning over a network is the ruinous multiple re-use of information

[^4]known as data incest (or mis-information propagation). In this chapter, the data incest management problem in social learning context is formulated on a family directed acyclic graphs.

The main contributions of Chapter 3 are summarized below:

1. A social learning model is presented to mimic the behavior of agents in online rating and review systems that aim to estimate a state of nature (for example quality of a restaurant on Yelp).
2. A fair rating and review protocol is presented and the criterion for achieving a fair rating is defined. This protocol is used as a benchmark in the data incest management problem in social learning over social networks.
3. An automated incest removal protocol is developed for the administrator of online rating and review system, to deploy such that the system maintains a fair rating of its entities. Such algorithm can easily be applied to any interactive-sensing system that involves transmission of actions and Bayesian inference.
4. Necessary and sufficient conditions on the graph topology of social interactions between social sensors are presented to eliminate data incest.

### 1.2.3 Tracking Degree Distribution of Social Networks

Chapter 4 considers dynamical random graphs. The most important measure that characterizes the structure of a network (specially when the size of the network is large and the connectionsadjacency matrix of the underlying graph-are not given) is the degree distribution of that network. The degree of a node in a network (also known as the connectivity) is the number of connections the node has in that network. In this chapter, motivated by social network applications, we consider a class of stochastic approximation algorithms to track a time-varying probability mass function that evolves according to a finite-state Markov chain whose transition matrix is close to identity. In the context of social network analysis, the time-varying probability mass function which we aim to track is the expected degree distribution of a dynamic random graph.

The main contributions of Chapter 4 are summarized below:

1. A family of Markov-modulated duplication-deletion random graphs are introduced in Chapter 4 that mimic social networks where interaction among agents are varying over time according to realization of a finite-state Markov chain. We consider two categories of such graphs: (i) fixed size duplication-deletion random graph, and (ii) infinite duplication-deletion random graph.
2. An asymptotic degree distribution analysis is presented for the fixed size Markov-modulated random graph. In particular, it is shown that the expected degree distribution of such graphs
at each time can be computed in terms of the expected degree distribution at the previous time and the dynamics of the graph via a recursive equation.
3. We extend the degree distribution analysis to infinite random graphs and prove that the degree distribution of such graphs satisfy a power law with the exponent depending on the dynamics of the graph. An expression is presented to compute the power law exponent in terms of the dynamics of the duplication-deletion model.
4. Chapter 4, further, considers the adaptive estimation problem of degree distribution for a fixed size Markov-modulated duplication-deletion random graph given noisy observations. A stochastic approximation algorithm is presented for tracking the degree distribution as it evolves over time. In the following, the results related to the tracking performance of stochastic approximation algorithm are presented.

- Mean square error analysis: Using error bounds on two-time scale Markov chains and perturbed Lyapunov function methods, the asymptotic mean square error between the expected degree distribution and the estimates obtained via the stochastic approximation algorithm is computed.
- Weak convergence analysis: We show that the asymptotic behavior of the stochastic approximation algorithm converges weakly to the solution of a switched Markovian ordinary differential equation.
- Functional central limit theorem for scaled tracking error: Finally, Chapter 4 investigates the asymptotic behavior of the scaled tracking error of stochastic approximation algorithm. Similar to [94], it is shown that the interpolated scaled tracking error converges weakly to the solution of a switching diffusion process.

5. Chapter 4, further, investigates infinite (denumerable) duplication-deletion random graphs where the number of nodes in the graph (and so the support of degree distribution) is no longer fixed and increases over time. A Hilbert-space-valued stochastic approximation algorithm is proposed to track the degree distribution of the infinite graph with support on the set of non-negative integers. To study the tracking performance of such a Hilbert-space-valued stochastic approximation algorithm, limit system characterization and asymptotic analysis of scaled tracking error are provided.

### 1.3 Related Works

This section is devoted to the literature review of topics and advances in the fields related to this dissertation.

### 1.3.1 Bayesian Estimation over Social Networks

Bayesian inference deals with the problem of inferring knowledge about unknown parameters using Bayes' rule [21, 125, 130]. Bayesian theory has many application in decision theory, detection and estimation, communications theory, pattern recognition, machine learning and artificial intelligence, and filtering and parameter estimation [39, 58, 64, 75, 77, 105, 112]. For a comprehensive survey on Bayesian inference and estimation theory, we refer the interested reader to books [130, 133, 141]. Bayesian networks and different inference methods in Bayesian networks are investigated in [120]. A model of Bayesian social learning where agents receive a private information about state of nature and observe the actions of their neighbors is investigated in [83]. They proposed an algorithm to compute actions of agents on tree-based social networks and analyzed their algorithm in terms of efficiency and convergence.

There are several papers discussing the spread of information in social networks, see [7, 31, 36] for a comprehensive survey and tutorial on different methods for diffusion of information in social networks. Applications of gossip algorithms, which is a protocol based on communication of agents with their local neighbors, in signal processing are studied in [65]. A type of mis-information propagation in social networks caused by "influential" or " forceful" agents is investigated in [3]. Viral propagation of faulty information (for example, mis-information of swine flu) through social media channels is studied in [30, 119]. For the motivation of the mis-information problem addressed in this work, we refer to [29, 49, 78, 106, 128] in sensor networks. Data incest in sensor network context happens in distributed tracking systems where sensors locally integrate the estimates received from other sensors through a (possibly loopy) communication graph with random delays. The key requirement is to fuse estimates that share a common information set. An optimal solution for the case of connected tree networks by combining a decentralized information filter and a channel filter is presented in [49].

In this thesis, we consider mis-information propagation through a social network with arbitrary network topologies. Each agent records its observation of state of nature in presence of noise. We used a combination of graph theoretic tools and Bayesian estimation to remove the mis-information removal generated by different delays in links.

### 1.3.2 Interactive Social Learning over Networks

Social learning theory is used to investigate the learning behavior of agents in social and economic networks [2]. There are several papers in the literature discussing Bayesian models [1, 43, 92, 120] and non-Bayesian models [14, 48, 56, 57, 81] for social learning. Different models for diffusion of beliefs in social networks are presented in [36]. For a comprehensive survey on herding and information cascade in social learning, see [37]. Stochastic control with social learning for sequential change detection problems is considered in [87].

Mis-information in the signal processing literature refers to faulty or inaccurate information which is broadcasted unintentionally. A brief summary of works related to misinformation propagation and removal is presented in Section 1.3.1. Mis-information in the context of this chapter is motivated by sensor networks where the term "data incest" is used [89]. In multi-agent social learning in networks, data incest occurs when information (action) of one agent is double-counted by other agents (due to the lack of information about the topology of the communication graph); this yields to overconfidence. The overconfidence phenomena (caused by data incest) also arises in Belief Propagation (BP) algorithms [113, 123] which are used in various fields such as graphical models for learning, computer vision, and error-correcting coding theory. The aim of BP algorithms is to solve inference problems over graphical models such as Bayesian networks (where nodes represent random variables and edges depict dependencies among them) by computing a marginal distribution. BP algorithms require passing local messages over the graph (Bayesian network) at each iteration. These algorithms converge to the exact marginal distribution when the factor graph is a tree (loop free). But for graphical models with loops, BP algorithms are only approximate due to the over-counting of local messages [113, 152] (which is similar to data incest in multi-agent social learning) ${ }^{7}$.

In Chapter 2 and papers [89, 90], data incest is considered in a network where agents exchange their private belief states-that is, no social learning is considered. In a social network, agents rarely exchange private beliefs, they typically broadcast actions (votes) over the network. Motivated by trustable online rating and review systems, we consider data incest in a social learning context with social network structure where actions (or equivalently public belief of the social learning) are transmitted over the network. This is quite different from private belief propagation in social networks. Simpler versions of this information exchange process and estimation were investigated by Aumann [12] and Geanakoplosand and Polemarchakis [66]. The results derived in this chapter extend theirs.

Finally, the methodology of Chapter 3 can be interpreted in terms of the recent Time magazine article [145] which provides interesting rules for online rating and review systems. These include: (i) review the reviewers, (ii) censor fake (malicious) reviewers. The data incest removal algorithm proposed in this chapter can be viewed as "reviewing the reviews" of other agents to see if they are associated with data incest or not.

[^5]
### 1.3.3 Tracking Degree Distribution of Social Networks

With a large number of rational agents, social networks can be viewed as social sensors for extracting information about the world or people. For example, the paper [132] presents a social sensor (based on Twitter) for real-time event detection of earthquakes in Japan, namely, the target event. They perform semantic analysis on tweets (which are related to the target event) in order to detect the occurrence and the location of the target event (earthquake). Another example is the use of the social network as a sensor for early detection of contagious outbreaks [40]. Using the fact that central individuals in a social network are likely to be infected sooner than others, a social sensor is designed for the early detection of contagious outbreaks in [40]. The performance of this social sensor was verified during a flu outbreak at Harvard College in 2009-the experimental studies showed that this social sensor provides a significant additional time to react to epidemics in the society. Social sensing in the context where physical sensors present in mobile devices such as GPS or Bluetooth are used to infer social interactions is studied in [6, 32, 34, 54] . Here, we consider a scenario that a social network considered as a sensor of social interactions, human activities, or behavior and aims to track the degree distribution of a random graph via stochastic approximation algorithms.

Stochastic approximation algorithms have several applications in diverse areas such as system identification, control theory, adaptive filtering, state estimation, wireless communications, target tracking, change detection, and economics [20,50-52, 93, 99, 110, 111, 150, 153]. The ubiquitous use of stochastic approximation algorithms is mainly due to their ability to track a time-varying unknown parameter of a system; this is called "tracking capability", see [20]. Tracking a timevarying parameter that evolves according to a finite-state Markov chain has several applications in target tracking [63], change detection [20], multi-user detection in wireless systems [153], and economics [93]. Tracking capability of regime switching stochastic approximation algorithms is further investigated in [154] in terms of the mean squared error. The interested reader is referred to [99] for a comprehensive development of stochastic approximation algorithms.

For the background and fundamentals on social and economic networks, we refer to [80]. Here, the related literature on dynamic social networks is reviewed briefly. The book [53] provides a detailed exposition of random graphs. The dynamics of random graphs are investigated in the mathematics literature, for example, see $[41,60,104]$ and the reference therein. In [121], a duplication model is proposed where at each time step a new node joins the network. However, the dynamics of this model do not evolve over time. In [41], it is shown that the degree distribution of such networks satisfies a power law. In random graphs which satisfy the power law, the number of nodes with a specific degree depends on a parameter called "power law exponent" [82, 147]. A generalized Markov graph model for dynamic social networks along with its application in social network synthesis and classification is also presented in [149]. The degree distribution analysis of real-world
networks has attracted much attention recently, $[8,45,55,67,85,114,118]$. A large network dataset collection can be found in [103], which includes datasets from social networks, web graphs, road, internet, citation, collaboration, and communication networks. The paper [114] investigates the structure of scientific collaboration networks in terms of degree distribution, existence of giant component, and the average degree of separation. In the scientific collaboration networks, two scientists are connected if they have co-authored a paper. Another example is the degree distribution of the affiliation network of actors ${ }^{8}$, which is studied in [118] based on real data from IMDb. In [8], the structure and characteristics of three different online social networks (Cywork, Myspace, and Orkut) are investigated. The authors use snowball sampling method ${ }^{9}[22,102]$ to estimate the degree distribution of the network when it is not possible to access all nodes in the network (specially when the size of the network is large). It is further shown in [55] that the degree distribution of email networks satisfies a power law.

Finally, different applications of social sensors in detection and estimation are investigated in [9, 40, 132]. The differences between social sensors, social sensing, and pervasive sensors along with challenges and open areas in social sensors are further presented in [88, 131].

### 1.4 Thesis Outline

In this section, we present the organization of this dissertation which is illustrated in Figure 1.3. The rest of this thesis is divided into two parts and four chapters as outlined below:

Motivated by different information diffusion patterns in interactive sensing over social networks, Part I considers multi-agent state estimation and learning problem over directed acyclic graphs in social networks and comprises of two chapters:

- Chapter 2 considers Bayesian estimation over directed acyclic graphs where agents transmit their beliefs about state of nature (the posterior distribution of state of nature given private observations and beliefs of other agents which are available due to the structure of the underlying social networks) instead of raw observations. It then formulates data incest (also known as mis-information propagation) that arises in such estimation problems using a graph-theoretic setup. It is, then, shown that under some necessary and sufficient conditions on the topology of the communication graph among agents, mis-information can completely be removed from the estimates of agents. Assuming that the communication graph is known, an optimal mis-information removal algorithm is proposed. We also provide a sub-optimal algorithm for reducing the effect of mis-information when the communication graph is not completely

[^6]
## Signal Processing Methods for Interactive Sensing Using Social Sensors

Part I: Estimation and Learning Over Directed Acyclic Graphs:

- Bayesian estimation for interactive sensing

Part II: Tracking the Degree

- Multi-agent social learning
- Degree distribution analysis of Markov-modulated graphs
- Data incest in social learning
- Social sensor of Markovian dynamics (state of nature)

Algorithmic Tools: Stochastic Approximation, Bayesian Filtering
Analysis Tools: Weak Convergence analysis, Graph Theory

Figure 1.3: Main results and organization of thesis.
known. This chapter is concluded with a numerical study that illustrates the excellent performance of the proposed algorithms.

- Motivated by online rating and review systems, Chapter 3 employs social learning to model the interactions among agents in a multi-agent estimation problem where the actions of agents are transmitted over the network (instead of raw observations or private beliefs). In such a setup, each agent-in order to estimate an underlying state of nature-chooses an action from a finite set of actions to minimize a local cost function and then transmits this action over the network. We give necessary and sufficient conditions on the graph topology of social interactions to eliminate data incest. A data incest removal algorithm is, then, proposed in this chapter such that the public belief of social learning (and, hence, the actions of agents) is not affected by data incest propagation. This results in an online rating and review system with a higher trust rating. This chapter then presents an actual psychology experiment that was conducted by our colleagues at the Department of Psychology of University of British Columbia in September and October, 2013, to illustrate social learning, data incest and social influence. Finally, numerical examples are provided to illustrate the performance of the proposed optimal data incest removal algorithm.

Motivated by applications of degree distribution in social network analysis and tracking a Markovian dynamics of graphs, Part II considers Markov modulated duplication-deletion random graphs
and is comprised of one chapter:

- Chapter 4 considers a Markov-modulated duplication-deletion random graph where at each time instant, one node can either join or leave the network; the probabilities of joining or leaving evolve according to the realization of a finite state Markov chain. This chapter comprises of two inter-related research problems. First, motivated by social network applications, the asymptotic behavior of the degree distribution is analyzed. Second, a stochastic approximation algorithm is presented to track empirical degree distribution as it evolves over time. The tracking performance of the algorithm is analyzed in terms of mean square error and a functional central limit theorem is presented for the asymptotic tracking error. Chapter 4, then, presents a Hilbert-space-valued stochastic approximation algorithm that tracks a Markovmodulated probability mass function with support on the set of nonnegative integers. Finally, this chapter is concluded with some numerical example that illustrates the performance of tracking algorithms and corroborate the findings of this chapter.

Chapter 5 briefly outlines a summary of findings in these two parts and provides a direction for future research and development in the fields related to this dissertation.

A brief review of some graph theoretic definitions and tools that have been used throughout this dissertation is presented in Appendix A.

An important associated problem to numerical studies of social networks is how to actually construct random graphs via simulation algorithms. In particular, for large social networks, only the degree sequence is available, and not the adjacency matrix. (The degree sequence is a nonincreasing sequence of vertex degrees.) Does a simple graph exist that realizes a particular degree sequence? How can all graphs that realize a degree sequence be constructed? Appendix B presents a discussion of these issues.

## Part I

## Estimation and Learning Over Directed Acyclic Graphs

## 2

## Constrained Estimation Over Random Graphs

### 2.1 Introduction

This chapter deals with Bayesian estimation problem over networks and considers a social network where each group of individuals use received information from the other social groups and employs Bayesian models of information aggregation to evaluate their belief about state of nature. In this context, each group of individuals form a social sensor of a an economic or a social parameter. The process of exchanging information between social groups is crucial for individuals to evaluate their beliefs. An important parameter that characterizes how the belief evolves is the delay in this information exchange. This delay can be extrinsic - individuals take different amounts of time to form beliefs and communicate them, or intrinsic to the network; highly connected nodes exchange information faster compared to nodes that have fewer connections. The most important consequence of this delay is mis-information (or incest) propagation as we will explain shortly. Let us first formulate the observation and information exchange using a graph-theoretic notation.

State of Nature: Let $x$ represent a state of nature that individuals in the social network aim to estimate such as quality of a restaurant. Assume that $x$ belongs to a finite set $\mathbf{X}=\left\{\bar{x}_{1}, \bar{x}_{2}, \ldots, \bar{x}_{N}\right\}$. Here, $\bar{x}_{i}$ (for $1 \leq i \leq N$ ) is in $\mathbb{R}^{d}$ or $\mathbb{R}^{+} d$ or $\mathbb{N}^{d}$ where $d$ is a positive integer number. Assume that $x$ has prior distribution $\pi_{0}$.

Observation Protocol: To estimate $x$, each individual in the social network obtains an $M$ dimensional observation vector where $M$ is a positive integer number. To simplify the analysis, we assume that the set of individuals in the social network is partitioned into $S$ social groups such that within each social group individuals record the same observations. At time $k$, the noisy observation of social group $s, z_{\lfloor s, k\rfloor}$ has conditional probability distribution

$$
\begin{equation*}
p\left(z_{\lfloor s, k\rfloor} \leq \bar{z} \mid x=\bar{x}_{i}\right)=\sum_{z \leq \bar{z}} B_{i z}, \quad 1 \leq i \leq N . \tag{2.1}
\end{equation*}
$$

Here $\sum_{z}$ denotes integration with respect to Lebesgue measure (in which case $B_{i z}$ is the conditional probability distribution function) or the counting measure (in which case $B_{i z}$ is the conditional probability mass function). Assume that $z_{n}$ given $x$ for different values of " $s$ " and " $k$ " are independent
random variables with respect to $s$ and $k$. Each social group $s$ combines its private observation $z_{\lfloor s, k\rfloor}$, with information received from other groups in social network to update its belief about state of nature $x$. Then, it communicates this updated belief to other groups in the social network.

Information Exchange Protocol: Let $G_{\lfloor s, k\rfloor}=\left(V_{\lfloor s, k\rfloor}, E_{\lfloor s, k\rfloor}\right), k=1,2,, \ldots, s=1,2, \ldots, S$ denote a sequence of time-dependent directed graphs of information flow in the social network until and including time $k$. Here $V_{\lfloor s, k\rfloor}$ denotes the set of vertices,

$$
\begin{equation*}
V_{\lfloor s, k\rfloor}=\left\{\left(s, k^{\prime}\right) \mid k^{\prime} \leq k, s \in\{1,2, \ldots, S\}\right\} \tag{2.2}
\end{equation*}
$$

and $E_{[s, k]} \subseteq V_{[s, k]} \times V_{\lfloor s, k]}$ is the set of edges which depicts the connections between vertices in $G_{\lfloor s, k]}$. For example if $\left(\left(s, k^{\prime}\right),\left(s^{\prime}, k^{\prime \prime}\right)\right) \in E_{\lfloor s, k\rfloor}$, it means that the information from social group $s$ at time $k^{\prime}$ is available at social group $s^{\prime}$ at time $k^{\prime \prime}\left(k^{\prime} \leq k^{\prime \prime} \leq k\right)$. Each social group uses Bayesian model to estimate the underlying state of nature $x$.

As a result of the recursive nature of Bayesian estimators, mis-information propagation can arise in a social network with the above information exchange protocol. For example, assume that the estimates of social group 1 at time $1, \theta_{[1,1]}$, reach social group 2 at time 2 . Also suppose the estimates from social group $2, \theta_{[2,2]}$, reach social group 1 at time 3 . Since social group 2 used $\theta_{\lfloor 1,1]}$, the estimate generated by social group 2 is a function of the $\theta_{[1,1]}$. Therefore, if social group 1 naively combines the estimate of social group 2 received at time $3, \theta_{[2,2]}$, with its own private estimates, it would have double counted its estimate at time $1, \theta_{[1,1]}$. In the above graph theoretic notation, we can depict graph $G_{\lfloor 2,3\rfloor}$ as

$$
\begin{array}{rlll}
(1,1) & \rightarrow & (1,2) & \rightarrow \\
& (1,3)  \tag{2.3}\\
\searrow & \nearrow & \\
(2,1) & \rightarrow(2,2) & \rightarrow & (2,3)
\end{array}
$$

where the two-tuples denote vertices defined in (2.2) and the arrows denote edges of directed acyclic graph. The fact that there exists two distinct paths between $(1,1)$ and $(1,3)$ in the graph of $(2.3)$ shows that information in $(1,1)$ is double counted leading to mis-information propagation. As the above example shows, the mis-information (rumor) propagation can be viewed as the destructive re-use of observation information. It leads to an overconfidence phenomenon i.e the variance is underestimated. The recursive nature of Bayesian estimation requires careful design to cope with the possible ruinous re-use of information. In more realistic problems considered in this chapter, there are multiple groups in social network (and thus an arbitrarily complex network topology) together with random delays in the network. For such cases, mis-information management is a non-trivial problem.

### 2.1.1 Chapter Goals

A network of social groups is considered in this chapter that aim to estimate the underlying state of nature $x$. Before proceeding, let us introduce the following scalar index $n$ instead of $\lfloor s, k\rfloor$ for the sake of notational simplicity:

$$
\begin{equation*}
n \triangleq s+S(k-1), \quad s \in\{1, \ldots, S\}, k \in\{1,2,3, \ldots\} . \tag{2.4}
\end{equation*}
$$

Notice that $n$ is a composite of time $k$ and social group $s$. Subsequently, we will refer to $n$ as a "node" of a time dependent graph namely information flow graph. This estimation problem can be expressed in the following abstract form:

Estimate $x$ with prior $\pi_{0}$ subject to:

$$
\left\{\begin{array}{l}
G_{n}=\left(V_{n}, E_{n}\right) \text { is given. }  \tag{2.5}\\
z_{n} \sim B_{i z}, \quad x=\bar{x}_{i}, \quad \text { (observation process) } \\
\theta_{n}=\mathscr{A}\left(\Theta_{n}, z_{n}\right), \quad \text { (filter constraint) }
\end{array}\right.
$$

Here, $\Theta_{n}$ denotes the set of beliefs from nodes (social groups at previous times) available at node $n$ (social group $s$ at time $k$ ) which depends on the information flow network $G_{n}$. Let $\theta_{n}$ denote the posterior distribution ${ }^{10}$ of $x$ given $\Theta_{n}$ and $z_{n}$. In $2.5, \mathscr{A}$ denotes the algorithm used by each node to update the belief $\theta_{n}$. The aim of this chapter is to construct the information aggregation algorithm $\mathscr{A}$ such that estimates $\theta_{n}$ are not affected by mis-information propagation. If $\mathscr{A}$ is not constructed properly, then mis-information can propagate in social network as explained earlier in this chapter. Thus, from an abstract point of view, mis-information removal can be interpreted as optimal Bayesian estimation on a directed acyclic graph with information exchange constraints. This chapter aims to address the following questions:

1. Existence Problem: Under what constraints on the information flow, is complete mis-information removal possible?
2. Design Problem: Synthesize an algorithm $\mathscr{A}$ such that mis-information propagation is prevented.
3. Reconstruction Problem: If the information flow graph, $G_{n}$, is not completely known at each time, design an algorithm to mitigate the mis-information propagation.
[^7]
### 2.1.2 Main Results and Organization of Chapter

This chapter considers mis-information propagation through a social network with arbitrary network topologies. Each social group records their observation of state of nature with any arbitrary aposteriori probability distribution and any arbitrary observation noise. The recursive nature of Bayesian estimation in decentralized fusion requires careful design to cope with the possible re-use of information such that the estimates $\theta_{n}$ are equal to mis-information free estimates of the optimal scenario which is described in more detail in Section 2.2.2. In more realistic problems considered in this chapter, there are multiple groups together with random delays in the network. For such cases, mis-information management is a non-trivial problem. A combination of graph theory and Bayesian estimation is employed to remove the mis-information removal generated by different delays in links. The rest of the chapter is organized as follows:

- We represent information flow in a social network by a family of directed acyclic graphs in Section 2.2. The communication among social groups is modeled by information exchange Protocol 2.1 where mis-information propagation may arise. Information exchange Protocol 2.2 is introduced to benchmark against Protocol 2.1. From the benchmark Protocol 2.2, the mis-information removal algorithm can be specified.
- Section 2.3 presents a necessary and sufficient condition (called Constraint 2.3.1) on topology of the network that guarantees optimal, mis-information free estimates. It is shown that with the full knowledge of information flow graph, Constraint 2.3.1 leads to an algorithm for exact mis-information removal using optimal Bayesian estimation defined in Section 2.3.
- A sub-optimal algorithm is proposed in Section 2.4 to remove the mis-information associated with estimates of social groups when the information flow graph is not completely known at each time.
- Numerical results that show the effect of mis-information propagation and also the excellent performance of the proposed mis-information removal algorithms are presented in Section 2.5 .


### 2.2 Modeling Information Flow in Social Networks

Section 2.1 outlined the goals and main results of this chapter and informally described constrained estimation over social networks. In this section, first, using a graph-theoretic notation, the following types of communication protocols are presented:

- Constrained information flow protocol: This protocol mimics information exchange, and inference in a social network where beliefs are communicated among groups (nodes). As stated
in Section 2.1, mis-information propagation arises in this protocol due to the abusive repetition of information.
- Full information flow protocol: This protocol is an ideal (and, thus, impractical) communication protocol that prevents mis-information propagation. To devise an algorithm to mitigate mis-information propagation, this protocol is used as a benchmark against the constrained information flow protocol, as we explain shortly.

Then, we assert via Theorem 2.2.1 that the flow of information in a social network can be represented by a family of time dependent Directed Acyclic Graphs (DAGs). Some essential graph theoretic tools that will be used to formulate the mis-information propagation problem are outlined in Appendix A.

### 2.2.1 Constrained Information Flow Protocol

This protocol refers to a social network where nodes aim to estimate an underlying state of nature As described in Section 2.1, instead of raw observations, the posterior distribution of state of nature (beliefs) are broadcasted over the network. Due to the information exchange constraint in this protocol, the complete history of beliefs are not available at each node. It is in such a constrained information flow network that mis-information propagation arises.

The information exchange protocol in constrained estimation over social networks described in Section 2.1 can be summarized as following:

Protocol 2.1 Constrained Information Flow Network Protocol at each node $n$
Step 1. Observation: Node $n$ (social group $s$ at time $k$ ) records its private observation vector $z_{n}$ according to (2.1), that is,

$$
z_{n} \sim B_{i z}, \quad x=\bar{x}_{i},
$$

where $n=s+S(k-1)$, see (2.4).
Step 2. Interaction with other social groups: Node $n$, then, accesses the network for beliefs from other social groups at previous time instants $\Theta_{n}$.
Step 3. Mis-information removal and Bayesian data fusion: Node $n$ uses mis-information removal algorithm together with Bayesian data fusion to combine $\Theta_{n}$ with its private observation $z_{n}$ and updates its belief $\theta_{n}$.
Step 4. Transmit the updated belief: Node $n$, then, broadcasts the updated belief over the network.

Remarks: We assume a reasonable degree of flexibility that each node deploys for broadcasting its information over the network. In Step 2 above we assumed for simplicity (to avoid collision of information) that only one social group is allowed to transmit information at each time instant.

### 2.2.2 Benchmark Full Information Flow Protocol

The goal of this chapter is to solve estimation problem (2.5) subject to the information exchange Protocol 2.1. We now describe an idealized (and therefore impractical) Protocol 2.2 that will be used as a benchmark against Protocol 2.1. In the benchmark protocol, we assume that instead of transmitting posterior distribution $\theta_{n}$, each node transmits its own private observations $z_{n}$ and all raw observations received over the network. In this protocol, since each node has the entire available observation history from previous nodes, there is no room for mis-information propagation, i.e., there is no chance for inadvertent re-use of private observations by any node. Let $Z_{n}$ be the set of observations from previous nodes (recorded or received until time $k$ at social group $s$ where $n=s+S(k-1))^{11}$. The benchmark protocol proceeds as follows:

Protocol 2.2 Benchmark Full Information Flow Network Protocol at each node $n$
Step 1. Observation: Node $n$ records its private observation vector $z_{n}$ according to (2.1).
Step 2. Interaction with other social groups: Node $n$ then accesses the network and receives the private observations from other nodes, $Z_{n}$.
Step 3. Information aggregation and Bayesian data fusion: Node $n$ uses $z_{n}, Z_{n}$ to compute it's belief $y_{n}=p\left(x \mid z_{n}, Z_{n}\right)$.
Step 4. Transmit augmented data: Node $n$, then, broadcasts the set of observations $Z_{n} \cup\left\{z_{n}\right\}$ over the network.

Since Protocol 2.2 serves as an idealized benchmark for designing mis-information removal algorithms, its efficiency is irrelevant. However, it can be made more efficient by requiring each node to only broadcast the observations which have not been already integrated in the belief computed by the other nodes in the network. By comparing the posterior distribution of state of nature in Protocol 2.1 with the same in benchmark Protocol 2.2, mis-information removal algorithm is specified in Section 2.3.

With Protocol 2.2 defined above, the benchmark estimation problem can be summarized as:

Estimate $x$ with prior $\pi_{0}$ subject to:

$$
\left\{\begin{array}{l}
G_{n}=\left(V_{n}, E_{n}\right) \text { is given. }  \tag{2.6}\\
z_{n} \sim B_{i z}, \quad x=\bar{x}_{i}, \quad \text { (observation process) } \\
y_{n}=\mathscr{F}\left(Z_{n}, z_{n}\right), \text { (standard estimation problem) }
\end{array}\right.
$$

The estimates $y_{n}$ are free of mis-information because node $n$ uses all available raw observations (and not estimates) form other previous nodes. Note that estimation problem (2.5) is a dynamic constrained estimation on the directed acyclic graphs. One set of constraints are on the topology of the information flow graph (which are also valid for the estimation problem (2.6) in benchmark protocol). However, there exists another constraint on the algorithm $\mathscr{A}$ (which does not hold for

[^8]the benchmark scenario). As will show in Section 2.3, the algorithm $\mathscr{A}$ has a specific linear form which is depicted by (2.11) in Section 2.3.1.

### 2.2.3 Modeling Time Evolution of the Information Flow

Before proceeding, we refer the interested reader to Appendix A for a summary of graph theoretic definitions that will be used to model information flow graph in Protocols 2.1 and 2.2. Recall from Section 2.1, $G_{n}$ denotes the time-dependent information flow graph of the social network. Each node $n^{\prime}$ in $G_{n}$ represent a social group $s^{\prime}$ at time $k^{\prime}$ such that $n^{\prime}=s^{\prime}+S\left(k^{\prime}-1\right)$, see (2.4). Each directed edge of $G_{n}$ between node $i$ and node $j$ shows that the information (belief in Protocol 2.1 or observation in Protocol 2.2) of node $i$ is available at node $j$ in the social network represented by $G_{n}$. Note that $G_{n}$ is always a sub-graph of $G_{n+1}$. Therefore we can use a family of time dependent Directed Acyclic Graphs (DAGs) ${ }^{12}$ to model the time evolution of the information flow in the social network. Indeed, the following proposition shows that information flow in a (group-based) social network can always be represented by a family of DAGs.

Theorem 2.2.1. The information flow in a social network defined in Protocol 2.1 and Protocol 2.2 comprising of $S$ groups up and until time $k$ can be represented by a family of DAGs $\mathscr{G}=$ $\left\{G_{n}\right\}_{n \in\{1, \ldots, N\}}$ where $N=$ Sk. Each DAG $G_{n}=\left(V_{n}, E_{n}\right)$ represents the information flow between the $n$ first nodes, where the generic node $n$ is defined by (2.4).

Proof. The proof is presented in in Section 2.7.1.
The Adjacency and the Transitive Closure matrices of $G_{n}$ are denoted by $A_{n}$ and $T_{n}$, respectively (see Appendix A for detail). Because of the fact that the information of each node cannot travel backwards in time, $A_{n}$ and $T_{n}$ are upper triangular matrices.
Memory Requirement: In this chapter, we assume that beliefs are valid for duration of $K$ timeinstants, where $K$ is a positive integer, i.e., social groups at time $k$ do not remember beliefs generated before time $k-K$. This means that the size of the adjacency and transitive closure matrices of each graph in $\mathscr{G}$ is limited to $N=S K$.

### 2.3 Optimal Mis-information Propagation Removal Algorithm

This section considers the estimation problem (2.5) with constrained information flow Protocol 2.1 in Section 2.2. The aim is to devise the information aggregation algorithm $\mathscr{A}$ in (2.5) such that the estimates of (2.5) with Protocol 2.1 are equal to those of (2.6) with the benchmark Protocol 2.2, i.e., $y_{n}=\theta_{n}$. Also, we provide necessary and sufficient conditions on the information flow graph under which the mis-information removal is possible.

[^9]
### 2.3.1 Optimal Combination Scheme in Constrained Information Flow Protocol

Consider the estimation problem (2.5) on directed acyclic graphs where social groups deploy Protocol 2.1. In this section, we address the following question: How should each social group combine its private observation with the received information (beliefs) from the network so that its updated belief is misinformation free?
To answer this question, consider estimation problem (2.6) with the idealized benchmark Protocol 2.2, where the set of raw observations are transmitted over the network and, thus, the estimates are mis-information free. In this scenario, since the history of all observations are available at each node and these observations are independent, the standard Bayesian update is used to evaluate

$$
y_{n}=p\left(x \mid Z_{n}, z_{n}\right) .
$$

Estimates $y_{n}$ are free of mis-information, therefore, to prevent mis-information propagation in constrained information flow Protocol 2.1, the information aggregation algorithm $\mathscr{A}$ should devised such that

$$
\begin{equation*}
p\left(x \mid \Theta_{u\left(G_{n}\right)}, z_{n}\right)=p\left(x \mid Z_{v\left(G_{n}\right)}, z_{n}\right), \quad \text { for } n=1,2, \ldots \tag{2.7}
\end{equation*}
$$

So, the first step in building the optimal information aggregation algorithm is to compute the estimates $y_{n}$ of the benchmark protocol in terms of raw observations and the information flow graph. Before proceeding, let us define

$$
\begin{equation*}
\hat{y}_{n}^{\text {full }}=\log \left(y_{n}\right)=\log \left(p\left(x \mid Z_{n}, z_{n}\right)\right) \quad \text { for } n=1,2, \ldots \tag{2.8}
\end{equation*}
$$

Since the logarithm is monotonically increasing, surjective function, we can work with logarithm of estimates $\widehat{y}_{n}^{\text {full }}$ instead of $y_{n}{ }^{13}$. The following proposition gives an expression for the estimates $\widehat{y}_{n}^{\text {full }}$ in full information flow network.

Theorem 2.3.1. Consider estimation problem (2.6) with information exchange Protocol 2.2 of Section 2.2. The mis-information free estimate at node $n$ is:

$$
\begin{equation*}
\widehat{y}_{n}^{\text {full }}=\left(t_{n} \otimes \mathbf{I}_{d}\right) l_{1: n-1}+l_{n}, \tag{2.9}
\end{equation*}
$$

where $\boldsymbol{\imath}_{n}$ denotes $\log \left(p\left(z_{n} \mid x\right)\right)$ and $\imath_{1: n-1} \triangleq\left[\imath_{1}^{\prime}, \ldots, \imath_{n-1}^{\prime}\right]^{\prime} \in \mathbb{R}^{(n-1) d \times 1}$. Here $\otimes$ denotes Kronecker (tensor) product and $\mathbf{I}_{d}$ denotes the $d \times d$ identity matrix. Recall that $t_{n}$ defined in (A.6) as the first $n-1$ elements of the $n^{\text {th }}$ column of $T_{n}$.

Proof. The proof is in Section 2.7.2

[^10]According to Theorem 2.3.1, the optimal mis-information free estimates can be expressed as a linear combinations of $t_{i}=\log \left(p\left(z_{i} \mid x\right)\right)$ in terms of Transitive Closure Matrix $T_{n}$ of graph $G_{n}$. Eq. (2.9) is quite intuitive. In information exchange Protocol 2.2, a node broadcasts its own raw observations and also passes the observations received from others nodes so that each node has the entire history of all possible observations, i.e., if there exists a path from node $i$ to node $n$, the observation of node $i, z_{i}$, is available at node $n$. Therefore, the estimate $\widehat{y}_{n}^{\text {full }}$ of node $n$ is sum of the information from all nodes that are connected to node $n$ (by single-hop or multi-hop paths) in information flow graph $G_{n}$.

Evaluating the estimates of benchmark Protocol 2.2, we are now ready to tackle the algorithm design problem. Consider the estimation problem (2.5) with information exchange Protocol 2.1 of Section 2.2. The aim is to devise algorithm $\mathscr{A}$ such that (2.7) holds. Define

$$
\begin{equation*}
\widehat{y}_{n}=\log \theta_{n} \quad \text { and } \quad \widehat{y}_{1: n-1} \triangleq\left[\hat{y}_{1}^{\prime}, \ldots, \hat{y}_{n-1}^{\prime}\right]^{\prime} . \tag{2.10}
\end{equation*}
$$

With the $n-1$ dimensional vector $w_{n}$ below denoting a weight vector (a more precise construction is given in Eq. (2.11) below), and $l_{n}$ defined in (2.9), we propose the following optimal combination scheme:

$$
\begin{equation*}
\widehat{y}_{n}=\left(w_{n} \otimes \mathbf{I}_{d}\right) \widehat{y}_{1: n-1}+t_{n}, \tag{2.11}
\end{equation*}
$$

Before describing why estimations $\widehat{y}_{n}$ in (2.11) are free of mis-information, we introduce the following constraints on the $n-1$ dimensional weight vector $w_{n}$.

Constraint 2.3.1. Consider the estimation problem (2.5) with information exchange Protocol 2.1. Then the set of weights $\left\{w_{n}\right\}_{n \in\{1, \ldots, N\}}$ in (2.11) satisfies the topological constraint for constrained flow network if $\forall j \in\{1, \ldots, n-1\}$ and $\forall n \in\{1, \ldots, N\}$

$$
\begin{equation*}
a_{n}(j)=0 \quad \Longrightarrow w_{n}(j)=0, \tag{2.12}
\end{equation*}
$$

where $a_{n}$ is defined in (A.6).
Constraint 2.3.1 imposes a topological condition on the weight vector $w_{n}$ Assuming that Constraint 2.3.1 holds, Theorem 2.3.2 below asserts that estimates computed from (2.3.1) are identical to the optimal, mis-information free estimates of information exchange Protocol 2.2.

Theorem 2.3.2. Consider estimation problem (2.5) with information exchange Protocol 2.1 of Section 2.2. Then the set of weights $\left\{w_{n}\right\}_{n \in\{1, \ldots, N\}}$ in (2.11) satisfies the topological constraint for constrained flow network if $\forall j \in\{1, \ldots, n-1\}$ and $\forall n \in\{1, \ldots, N\}$. Then the following optimality
property holds for the estimates $\widehat{y}_{n}$ in the constrained information flow network:

$$
\widehat{y}_{n}=\widehat{y}_{n}^{\text {full }} \Longleftrightarrow\left\{\begin{array}{l}
w_{n}=t_{n}\left(\left(T_{n-1}\right)^{\prime}\right)^{-1}  \tag{2.13}\\
\text { and } w_{n} \text { satisfy Constraint 2.3.1 }
\end{array}\right.
$$

where $\widehat{y}_{n}$ and $\widehat{y}_{n}^{\text {full }}$ are defined in (2.10) and (2.8) respectively. Recall that $t_{n}$ defined in (A.6) as the first $n-1$ elements of the $n^{\text {th }}$ column of $T_{n}$.

Proof. The proof is presented in Appendix 2.7.3
In words: A necessary and sufficient condition for $\widehat{y}_{n}=\widehat{y}_{n}^{f}$ full to be held is that the weight vector $w_{n}$ satisfies $w_{n}=t_{n}\left(\left(T_{n-1}\right)^{\prime}\right)^{-1}$. As a result of constrained information flow, $w_{n}$ should simultaneously satisfy topological constraint (2.12) to ensure that required information for mis-information removal is available at node $n$.

Discussion: According to Theorem 2.3.1, the mis-information free estimates of full information flow protocol at node $n$ is linear in the estimates computed by the previous nodes i.e. $l_{1: n-1}$ and the information collected $l_{n}$ at node $n$. This linearity enables us to remove the mis-information in the constrained information flow protocol by employing the optimal combination scheme (2.11) with $n-1$ dimensional weight vector $w_{n}$ defined in Theorem 2.3.2. However, according to the communication topology described by Adjacency Matrix $A_{n}$, some nodes do not transmit their estimates to node $n$. Consequently, the estimates of those nodes are not available at node $n$ and must not be used to compute the estimate $\widehat{y}_{n}$. An obvious way to introduce this constraint in the estimation of $\widehat{y}_{n}$ is to set the weight related to an unavailable estimate to zero. This is Constraint 2.3.1. Therefore, it is also clear that (2.12) is a necessary condition for exact mis-information removal. Now, assuming that all the estimates of the $n-1$ latest nodes are optimal estimates free of mis-information i.e. $\widehat{y}_{1: n-1}=\widehat{y}_{1: n-1}^{\text {full }}$ and are available at node $n$, is it possible to find a vector $w_{n}$ so that $\widehat{y}_{n}$ is equal to the optimal estimates free of mis-information i.e. $\widehat{y}_{n}^{\text {full }}$ ? Theorem 2.3.2 provides the answer, with $w_{n}=t_{n}\left(\left(T_{n-1}\right)^{\prime}\right)^{-1}$. The non-zero elements of $w_{n}$ show the nodes whose estimates should be available at node $n$ to remove the mis-information. But we know that due to the topology of the graph some of the estimates form previous $n-1$ nodes are not available at node $n$. Constraint 2.3.1 basically ensures that the essential estimates (to remove the mis-information) from previous nodes are available at node $n$ and gives a necessary and sufficient condition on the topology of the graph for $\widehat{y}_{n}=\widehat{y}_{n}^{\text {full }}$. In Section 2.5.1, we give more intuition on Constraint 2.3.1 by a simple example.

### 2.4 Sub-optimal Mis-information Removal Algorithm Without Complete Knowledge of Information Flow Graph

So far in this chapter, an optimal information aggregation scheme is proposed for estimation problem (2.5) with Protocol 2.1. Theorem 2.3.2 asserts that the estimates obtained via optimal aggregation scheme (2.11) are equal to those of the benchmark Protocol 2.2. Recall that mis-information propagation arises due to the abusive repetition of information received from other nodes. This happens at one node, for example node $n$, when there exists a node in information flow graph $G_{n}$ with two or more links to node $n$. When, the information flow graph is known, each node can identify the origin of mis-information propagation and, thus, remove that (under the assumptions of Constraint 2.3.1) via optimal information aggregation scheme (2.11). However, this scheme requires the full knowledge of the information flow graph $G_{n}$. In this section, we relax that assumption and propose a mis-information removal algorithm for the arbitrary case that each node does not know the path of the received message. Instead, the 'expected adjacency matrix' of the information flow graph is known at all nodes. Our goal here, is to devise a sub-optimal information aggregation scheme based on (2.11) to reduce the effect of mis-information propagation where the information flow graph is unknown. Before presenting the estimation problem when information flow is unknown, let's take a closer look into the expected adjacency matrix of a graph. Recall that the adjacency (or connectivity) matrix of a graph shows the connections among nodes in a graph, i.e., the element on row $i$ and column $j$ of the adjacency matrix is equal to one if there exists a (single-hop) link from node $i$ to node $j$, otherwise, it is zero. Similarly, the expected adjacency matrix of a graph is defined as follows:

$$
\widetilde{A}_{n}=\left[\widetilde{a}_{i, i}\right] \text { for } 1 \leq i, j \leq n,
$$

where $\widetilde{a}_{i, j}$ denotes the probability of having a link from node $i$ to node $j$. We assume that instead of adjacency matrix, $\widetilde{A}_{n}$ is known at node $n$. The estimation problem when $G_{n}$ is unknown can be summarized in the following abstract form:

Estimate $x$ with prior $\pi_{0}$ subject to:

$$
\left\{\begin{array}{l}
\widetilde{A}_{n} \text { is given },  \tag{2.14}\\
z_{n} \sim B_{i z}, \quad x=\bar{x}_{i} \\
\theta_{n}=p\left(x \mid \Theta_{n}, z_{n}, \widetilde{A}_{n}\right)=\mathscr{B}\left(\Theta_{n}, z_{n}, \widetilde{A}_{n}\right)
\end{array}\right.
$$

The aim is to devise an algorithm to reduce the effect of mis-information propagation in estimation problem (2.14). Knowing $\widetilde{A}_{n}$, nodes which are more likely to have multiple paths to node $n$ can be identified. Note that in the optimal information aggregation scheme (2.11), the weight vector $w_{n}$ is the only term which depends on the information flow graph. More specifically, Theorem 2.3.2 computes the optimal weight vector in terms of the transitive closure matrix of the information flow
graph $T_{n}$. In estimation problem (2.14), $G_{n}$ (and consequently $T_{n}$ ) is unknown as opposed to the estimation problem (2.5). Therefore, the main challenge in the scenario where $G_{n}$ is unknown, is to approximate $T_{n}$ and, then, compute $w_{n}$ in terms of the expected transitive closure matrix.

### 2.4.1 Sub-optimal Combination Scheme

Having known the expected adjacency matrix, our sub-optimal approach to reduce the effect of mis-information at each node, for example node $n$, consists of three steps:

- First, we approximate the transitive closure matrix (probability of having a path between each pair of nodes in the graph) from the expected adjacency matrix $\widetilde{A}_{n}$, that is,

$$
\widetilde{T}_{n}(i, j)=p\left(T_{n}(i, j)=1\right), \quad 1 \leq i, j \leq n,
$$

where $p(\cdot)$ is used to denote probability of an event.

- Second, From $\widetilde{T}_{n}$, nodes that are more likely to have a path to node $n$ can be identified. From this, the hard estimation of the transitive closure matrix can be constructed, that is,

$$
\bar{T}_{n}(i, j)= \begin{cases}1 & \widetilde{T}_{n}(i, j)>\lambda_{t h}  \tag{2.15}\\ 0 & \text { otherwise }\end{cases}
$$

where $\lambda_{t h}$ is threshold and $\bar{T}_{n}$ is the hard estimate of transitive closure matrix of $G_{n}$.

- Having computed $\bar{T}_{n}$, Algorithm $\mathscr{A}$, (in the estimation problem (2.5)) is used to reduce effect of the mis-information propagation.

$$
\begin{equation*}
\bar{y}_{n}=\left(\bar{w}_{n} \otimes \mathbf{I}_{d}\right) \bar{y}_{1: n-1}+t_{n}, \tag{2.16}
\end{equation*}
$$

where weight vector $\bar{w}_{n}$ satisfies $\bar{w}_{n}=\bar{t}_{n}\left(\left(\bar{T}_{n-1}\right)^{\prime}\right)^{-1}$ and simultaneously satisfies topological Constraint 2.3.1. Here, $\bar{T}_{n}$ is defined in (2.15) and $\bar{t}_{n}$ is the first $n-1$ elements of " $n$ "th column of $\bar{T}_{n}$ and $l_{n}=\log \left(p\left(z_{n} \mid x\right)\right)$.

Algorithm 2.3 summarizes the sub-optimal mis-information removal problem without knowledge of information flow network.

### 2.5 Numerical Examples

In this section, we first provide an example to give more intuition on the topological Constraint 2.3.1 which is required for exact mis-information removal. Then, the performance of the optimal misinformation removal algorithm presented in Theorem 2.3.2 is compared with that of the full infor-

## Algorithm 2.3 Algorithm for mis-information removal in step 3 of Protocol. 2.1

For $n=1,2, \ldots$

1. Reconstruct the weighted adjacency matrix of the information flow, $\widetilde{A}_{n}$.
2. Compute $\widetilde{T}_{n}$ (2.4.1) and $\bar{T}_{n}$ using threshold, $\lambda_{t h}$ (2.15).
3. Using $\bar{t}_{n}$, compute $\bar{w}_{n}=\bar{t}_{n}\left(\left(\bar{T}_{n-1}\right)^{\prime}\right)^{-1}$.
4. Update the estimates $\bar{y}_{n}=\left(\bar{w}_{n} \otimes \mathbf{I}_{d}\right) \bar{y}_{1: n-1}+t_{n}$.
mation flow communication protocol where instead of the beliefs about the state of nature, all raw observations are transmitted over the network. Finally, the performance of the sub-optimal misinformation removal algorithm (Algorithm 2.3) is investigated in two scenarios: (i) accurate and (ii) inaccurate estimation of the information flow graph.

### 2.5.1 Example and Intuition on Theorems 2.2.1 and 2.3.2

In this subsection, we provide an example that shows the propagation of mis-information in a simple social network and the mis-information removal algorithm proposed in Section 2.3 to prevent that. Consider a social network consisting of two groups with the following information flow graph until time $K=3$


Figure 2.1: Example of constrained information flow network, $S=2$ and $K=3$. Circles represent a social group at a specific time indexed by (2.4) in the social network and each edge depicts a communication link between two nodes.

There are $S=2$ groups and the total time duration $K=3$. From (2.4), the element indexed by $n=s+2(k-1)$ in Fig.2.1, represents node $s$ at time $k$. According to Theorem 2.2.1, we can build the family of $N=S K=6 \mathrm{DAGs}$, namely, $\left\{G_{1}, G_{2}, G_{3}, G_{4}, G_{5}, G_{6}\right\}$. Based on the information flow in Fig.2.1, since nodes 1 and 2 do not communicate (see Fig.2.1), clearly $A_{1}$ and $A_{2}$ are zero matrices. Nodes 1 and 3 and Nodes 2 and 3 communicate, hence $A_{3}$ has two ones; and so on. The adjacency matrix associated with graphs $G_{1}, G_{2}, G_{3}, G_{4}$ and $G_{5}$ are:
$A_{1}=[0], A_{2}=\left[\begin{array}{ll}0 & 0 \\ 0 & 0\end{array}\right], A_{3}=\left[\begin{array}{lll}0 & 0 & 1 \\ 0 & 0 & 1 \\ 0 & 0 & 0\end{array}\right], A_{4}=\left[\begin{array}{llll}0 & 0 & 1 & 1 \\ 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0\end{array}\right], A_{5}=\left[\begin{array}{lllll}0 & 0 & 1 & 1 & 1 \\ 0 & 0 & 1 & 1 & 1 \\ 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0\end{array}\right]$.
The transitive closure matrices $T_{n}$ are obtained using (A.4). Using (A.4), we derive the transitive closure matrices from the adjacency matrices associated with graph $G_{1}, G_{2}, G_{3}, G_{4}$ and $G_{5}$ :
$T_{1}^{-1}=[1], T_{2}^{-1}=\left[\begin{array}{ll}1 & 0 \\ 0 & 1\end{array}\right], T_{3}^{-1}=\left[\begin{array}{ccc}1 & 0 & -1 \\ 0 & 1 & -1 \\ 0 & 0 & 1\end{array}\right], T_{4}^{-1}=\left[\begin{array}{cccc}1 & 0 & -1 & -1 \\ 0 & 1 & -1 & -1 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1\end{array}\right], T_{5}^{-1}=\left[\begin{array}{ccccc}1 & 0 & -1 & -1 & 1 \\ 0 & 1 & -1 & -1 & 1 \\ 0 & 0 & 1 & 0 & -1 \\ 0 & 0 & 0 & 1 & -1 \\ 0 & 0 & 0 & 0 & 1\end{array}\right]$.
Note that $T_{n}(i, j)$ is non-zero only for $i \geq j$ due to the causality-since information sent by a social group can only arrive at another social group at a later time instant. Also note that since $G_{n}$ is the subgraph of $G_{n+1}$ with node $n+1$ removed, the adjacency matrix and transitive closure matrix $A_{n+1}$ and $T_{n+1}$ contain $A_{n}$ and $T_{n}$, respectively. Also, they are upper left $n \times n$ matrices, see Remark A. 2 in Section A. The weight vectors are derived from the Transitive Closure Matrices via (2.13):
$w_{2}=[0]$,
$w_{3}=\left[\begin{array}{ll}1 & 1\end{array}\right]$,
$w_{4}=\left[\begin{array}{lll}1 & 1 & 0\end{array}\right]$,
$w_{5}=\left[\begin{array}{llll}-1 & -1 & 1 & 1\end{array}\right]$.
Let us examine these weight vectors. $w_{2}$ means that node 2 does not use estimate from node 1 . This formula is consistent with the constraint information flow because estimate from node 1 is not available to node 2; see Fig.2.1. $w_{3}$ means that node 3 uses estimates from node 1 and 2; $w_{4}$ means that node 4 only uses estimates from node 1 and node 2. The estimate from node 3 is not available at node 4. As shown in Fig.2.1, the mis-information propagation occurs at node 5. The vector $w_{5}$ says that node 5 adds estimates from nodes 3 and 4 and removes estimates from nodes 1 and 2 to avoid double counting of these estimates already integrated in estimates from node 3 and 4. Indeed, using the algorithm and the weight vector proposed in Theorem 2.3.2, the mis-information propagation is completely prevented in this example. Now consider the case that the edge between node 3 and node 5 does not exist. In this scenario $a_{5}(2)=0$ while $w_{5}(2) \neq 0$, therefore Constraint 2.3.1 does not hold and exact mis-information removal is not possible.

### 2.5.2 Numerical Examples Illustrating Algorithm $\mathscr{A}$ in Estimation Problem (2.5)

In this section, numerical results are given to illustrate the effect of mis-information propagation on the performance of multi-agent Bayesian estimation. The excellent efficacy of the optimal misinformation removal algorithm proposed in Section 2.3 is also corroborated in this section. We con-
sider a social network consisting of $S=2$ different groups that aim to estimate a scalar state of nature $x$ in the network with a given prior distribution. We simulate a social network with communication delays between different groups chosen randomly from $\{1,2, \ldots, 10\}$. The prior distribution of $x$ is uniform distribution $U[0,4]$ and the observation noise is zero-mean normal distribution $N(0,1)$. The sample of state of nature is $x^{*}=2.78$. The simulation is repeated $M=100$ times. At each iteration $i$, estimated value of state of nature for node $n$ is computed via

$$
\begin{equation*}
x_{n}^{i}=\sum_{j=1}^{N} \bar{x}_{j} p\left(x=\bar{x}_{j} \mid \Theta_{n}, z_{n}\right) . \tag{2.17}
\end{equation*}
$$

Then, the results of all 100 iterations $x_{n}^{l}$ are averaged to find the conditional mean of the state of nature $x$ for node $n$ as $\frac{1}{M} \sum_{l=1}^{M} x_{n}^{l}$.


Figure 2.2: The conditional mean of the state of nature given the observation in estimation with optimal mis-information removal algorithm compared to the full information network.

Fig.2.2 illustrates the effect of mis-information propagation in the Bayesian estimators. We consider three scenarios:
(i) Full information flow protocol which is free of mis-information as discussed in Section 2.2,
(ii) Constrained information flow protocol with standard Bayesian filter (naive mixing of observations and received information) which may contain mis-information,
(iii) Constrained information flow protocol with optimal mis-information removal algorithm proposed in Section 2.3.
As can be seen in this figure, the performance of the Bayesian estimator is ruined in the existence of the mis-information propagation. The dash-dash line, which represents the standard Bayesian estimation without mis-information removal algorithm, converges to a slightly different value than 2.78. Fig.2.2 also shows the excellent performance of the mis-information removal algorithm presented in Theorem 2.3.2. This figure shows that, the expected value of $x$ (unknown state of nature) given estimates obtained by the optimal information aggregation scheme (2.11)-depicted with dashdot line marked with " $\diamond$ "-is similar to that of the optimal mis-information free estimate of the full information flow protocol depicted by the solid line. This verifies the results of Theorem 2.3.2. We also investigate the performance of the mis-information removal algorithm in terms of the mean squared error in estimation of state of nature, namely $\sigma_{n}=\frac{1}{M} \sum_{l=1}^{M}\left(x_{n}^{l}-x^{*}\right)^{2}$ where $x^{*}$ is the true state of nature and $x_{n}^{l}$ is the estimates of node $n$ at iteration $l$ as in (2.17).

We can see in Fig. 2.3 that mean squared error associated with the estimates obtained by misinformation removal algorithm proposed in Section 2.3 (dash-dot line marked with " $\diamond$ ") is lower than that of the constrained information flow protocol without mis-information removal proposal (dash-dash line ).

### 2.5.3 Numerical Examples Illustrating Algorithm $\mathscr{B}$ in Estimation Problem (2.14)

Performance of sub-optimal Algorithm 2.3 is studied in a social network consisting of two groups $S=2$ over a duration of $K=25$ instants. From (2.4), social group $s$ at time $k$ is represented by node $n=s+2(k-1)$ in information flow graph. Information from one social group reaches the others after a random delay. In our numerical study, the prior of $x$ is uniform distribution $U[0,4]$. Furthermore, the observation noise is zero-mean normal distribution $N(0,1)$. The sample of state of the nature is $x=2.78$. In this study, we assume that information flow graph is not known at each node, but a weighted adjacency matrix (which can be considered as a noisy version of true adjacency matrix) is available at each node. This is motivated by the fact that nodes are able to construct the adjacency matrix $A_{n}$ from the additional information they have about communication topology. This information can be, for example, the distribution of communication delays. Therefore, each node can construct the weighted adjacency matrix of the information flow graph with elements $(i, j)$ which depicts the probability that the information of a node $i$ reaches node $j$. To investigate the performance of the sub-optimal information removal algorithm, we choose the elements of the


Figure 2.3: Comparison of the the mean squared errors of the estimates obtained by optimal misinformation removal algorithm, Bayesian estimator in full information flow network (free of misinformation), and standard Bayesian estimator in constrained information flow network (with misinformation propagation).
weighted adjacency matrix, $\widetilde{A}_{n}$ as follows:

$$
\widetilde{a}_{i j}= \begin{cases}a_{i j}-\beta u_{i j}, & a_{i j} \neq 0  \tag{2.18}\\ 0, & a_{i j}=0\end{cases}
$$

where $u_{i j}$ has a uniform distribution $U[0,1]$ and $\beta$ is a positive real number in $(0,1)$. We choose $\beta=0.2$ for the "accurate estimation" and $\beta=0.8$ for the "inaccurate estimation" of the information flow graph. The estimation problem (2.14) is investigated in the following four cases:
(i) Full information flow network with full knowledge of information flow graph at each time (optimal mis-information free scenario) which is shown by the solid line.
(ii) Standard (naive in this context) Bayesian estimation in constrained information flow network with full knowledge of information flow graph at each time which is depicted by the dash-dash line. (iii) Optimal mis-information removal algorithm in constrained information flow network with full knowledge of the information flow graph at each time which is shown via the dotted line marked with " $\diamond$ ".
(iv) Sub-optimal mis-information removal algorithm in constrained information flow network without knowledge of the information flow graph (using the weighted adjacency matrix with $\lambda_{t h}=0.6$ ) which is shown by the dash-dot line marked with " $\times$ ".
Fig.2.4 illustrates the expected value of state of nature in the above four scenarios. Similar to


Figure 2.4: Comparison of the conditional mean of the state of nature $x$ given the observations obtained by sub-optimal mis-information removal algorithm, optimal mis-information removal algorithm (knowing the exact information flow graph), Bayesian estimator in full information flow network (free of mis-information), and standard Bayesian estimator in constrained information flow network (with mis-information propagation) in "accurate estimation" scenario ( $\beta=0.2$ ).

Section 2.5.2, the estimates are found by means of the Monte-Carlo simulations with $M=100$ iterations. As can be seen in this figure, the estimations of the state of nature with employing suboptimal mis-information removal algorithm in "good estimation" scenario ( $\beta=0.2$ in (2.18)) is very close to the mis-information free estimates in full information flow network and those obtained by the optimal mis-information removal, knowing the exact adjacency matrix of information flow graph.

Mean squared errors associated with four scenarios studied in this section, are compared in Fig.2.5. As can be seen in the figure, the mean squared error of estimates obtained by the suboptimal mis-information removal algorithm is lower than the mean squared error of the estimates


Figure 2.5: Comparison of the mean squared errors of the estimates obtained by sub-optimal mis-information removal algorithm, optimal mis-information removal algorithm (knowing the exact information flow graph), Bayesian estimator in full information flow network (free of misinformation), and standard Bayesian estimator in constrained information flow network (with misinformation propagation) in "accurate estimation" scenario ( $\beta=0.2$ ).
obtained by the standard Bayesian estimation without mis-information removal algorithm. Figures 2.4 and 2.5 show the excellent performance of the proposed sub-optimal algorithm when the information flow graph is not completely known but a good approximation of it is available at each time.

To study the effect of the estimated adjacency matrix of the information flow graph on the performance of the sub-optimal algorithm for mitigating the mis-information propagation, we repeat the simulation for $\beta=0.8$ in (2.18) (inaccurate estimation scenario). The expected value of the state of nature given the available information in four different scenarios described above are depicted in Fig.2.6.

Fig. 2.7 shows the means squared errors of estimates for state of nature in four scenarios under investigation with $\beta=0.8$ in (2.18). As can be inferred from Figures 2.6 and 2.7, the performance of the sub-optimal depends on the estimated adjacency matrix of information flow graph. When an accurate approximation is available the performance of the sub-optimal algorithm is very close


Figure 2.6: Comparison of the conditional mean of the state of nature $x$ given the observations obtained by sub-optimal mis-information removal algorithm, optimal mis-information removal algorithm (knowing the exact information flow graph), Bayesian estimator in full information flow network (free of mis-information), and standard Bayesian estimator in constrained information flow network (with mis-information propagation) in "inaccurate estimation" scenario ( $\beta=0.8$ ).
to the performance of the optimal mis-information removal algorithm. However, in the presence of high variance noise in the estimation of the adjacency matrix of the information flow graph, the performance of the sub-optimal algorithm for mitigation mis-information drops.

### 2.6 Closing Remarks

In this chapter, the problem of mis-information propagation among different groups in social networks is addressed. We considered the most general scenario with arbitrary observation noise and any priori distribution of state of nature. A sufficient and necessary condition for mis-information removal problem is derived based on the topology of the information flow network. Also the performance of the proposed mis-information removal algorithm is illustrated in numerical examples. We also proposed a sub-optimal algorithm to mitigate the mis-information propagation when the information flow network is not known. Numerical results are presented to illustrate the performance of


Figure 2.7: Comparison of the the mean squared errors of the estimates obtained by sub-optimal mis-information removal algorithm, optimal mis-information removal algorithm (knowing the exact information flow graph), Bayesian estimator in full information flow network (free of misinformation), and standard Bayesian estimator in constrained information flow network (with misinformation propagation) "inaccurate estimation" scenario ( $\beta=0.8$ ).
the proposed algorithms.
Note that data incest considered in this chapter may also arise in any set of sensors that interact over graphs (possibly with random communication delays) and employ Bayesian models for information aggregation. Examples of such sensor network setup includes decentralized target tracking, localization, and fault detection. Although this chapter is motivated by social networks, but the data incest removal algorithm presented in Section 2.3 can applied to remove the mis-information associated with estimates of sensors in such setups.

### 2.7 Proof of Theorems

Here, we present proof for propositions and results of this chapter in the order they appeared.

### 2.7.1 Proof of Theorem 2.2.1

To prove that the graph $G_{n}=\left(V_{n}, E_{n}\right)$ from family $\mathscr{G}_{n}$ is a directed acyclic graph, we only need to show that the adjacency matrix of $G_{n}$ is an upper triangular matrix. Then from Lemma A.1, the graph $G_{n}$ is a directed acyclic graph. Suppose that $v_{i}$ and $v_{j}$ are two vertices of $G_{n}$, that is $v_{i}, v_{j} \in V_{n}$. From re-indexing scheme (3.7), $v_{i}$ and $v_{j}$ represents agents $s_{i}$ and $s_{j}$ at time instants $k_{i}$ and $k_{j}$, respectively. We have $v_{i}=s_{i}+S\left(k_{i}-1\right)$ and $v_{j}=s_{j}+S\left(k_{j}-1\right)$. Because of the information flow, information of each agent may become available at other agents at later time instants, a message cannot travel back in the time! This means that if $k_{i}<k_{j}$, there should not be an edge from $v_{j}$ to $v_{i}$, $\left(v_{j}, v_{i}\right) \notin E_{n}$. Using re-indexing scheme if $k_{i}<k_{j}$, then $v_{i}<v_{j}$ (because $k_{i}$ and $k_{j}$ are integers and $s_{i}, s_{j} \leq S$ ). Therefore, we deduce that

$$
\begin{equation*}
i<j \Rightarrow\left(v_{j}, v_{i}\right) \notin E_{n} \tag{2.19}
\end{equation*}
$$

Consequently, the adjacency Matrix is a strictly upper triangular matrix so that $G_{n}$ is a DAG. Then it follows from the construction of the DAGs that $\mathscr{G}_{N}$ is a family of DAGs.

### 2.7.2 Proof of Theorem 2.3.1

In the full information flow protocol, each node uses Bayesian estimation to update the probability distribution of $x$ given the set of available observations. The following recursive equation is used at each node to update the estimation of the probability distribution at node $n+1$ :

$$
\begin{equation*}
p\left(x \mid Z_{v\left(G_{n+1}\right)}, z_{n+1}\right)=\pi_{0} p\left(z_{n+1} \mid x\right) \prod_{i \in Z_{v\left(G_{n+1}\right)}} p\left(z_{i} \mid x\right) \tag{2.20}
\end{equation*}
$$

There is a path from each node $j \in v\left(G_{n}\right)$ to node $n$, therefore from (A.6), $t_{n}(j)=1$. Thus, $\left(t_{n} \otimes \mathbf{I}_{d}\right) l_{1: n-1}$ can be written as

$$
\begin{equation*}
\left(t_{n} \otimes \mathbf{I}_{d}\right) \iota_{1: n-1}=\sum_{j \in v\left(G_{n}\right)} \imath_{j}=\sum_{j \in v\left(G_{n}\right)} \log \left(p\left(z_{j} \mid x\right)\right)=\sum_{j \in v\left(G_{n}\right)} \log \left(p\left(z_{j} \mid x\right)\right) . \tag{2.21}
\end{equation*}
$$

Taking logarithms of (2.20), yields:

$$
\begin{equation*}
\left.\log \left(p\left(x \mid Z_{v\left(G_{n}\right)}, z_{n}\right)\right)=\sum_{j \in v\left(G_{n}\right)} \log \left(p\left(z_{j}\right) \mid x\right)\right) . \tag{2.22}
\end{equation*}
$$

Using (2.22) and (2.21), we can rewrite $\hat{y}_{n}^{\text {full }}$ and complete the proof,

$$
\begin{equation*}
\left.\left.\left.\widehat{y}_{n}^{\text {full }}=\sum_{j \in v\left(G_{n}\right)} \log \left(p\left(z_{j}\right) \mid x\right)\right)=\sum_{j \in v\left(G_{n}\right)} \log \left(p\left(z_{j}\right) \mid x\right)\right)+\log \left(p\left(z_{n}\right) \mid x\right)\right)=\left(t_{n} \otimes \mathbf{I}_{d}\right) l_{1: n-1}+t_{n} . \tag{2.23}
\end{equation*}
$$

(Note that we omit $\log \pi_{0}$ which is the same for both full and constrained information flow protocols for simplicity.)

### 2.7.3 Proof of Theorem 2.3.2

We, first, show that the left hand side of (2.13) (i.e., $\widehat{y}_{n}=\widehat{y}_{n}^{\text {full }}$ ) implies the right hand side of (2.13). Start with (2.11) for $\widehat{y}_{n}$ and replacing $\widehat{y}_{n}$ with $\widehat{y}_{n}^{\text {full }}$ yields

$$
\begin{equation*}
\widehat{y}_{n}^{\text {full }}=\left(w_{n} \otimes \mathbf{I}_{d}\right) \widehat{y}_{1: n-1}^{\text {full }}+v_{n} . \tag{2.24}
\end{equation*}
$$

Using Theorem 2.3.1 it follows that $y_{1: n-1}^{\text {full }}=\left(T_{n-1}^{\prime} \otimes \mathbf{I}_{d}\right) \boldsymbol{l}_{1: n-1}$. Then, incorporating this into (2.24) yields

$$
\begin{equation*}
\widehat{y}_{n}^{\text {full }}=\left(w_{n} \otimes \mathbf{I}_{d}\right)\left(T_{n-1}^{\prime} \otimes \mathbf{I}_{d}\right) l_{1: n-1}+l_{n} . \tag{2.25}
\end{equation*}
$$

From Theorem 2.3.1, we have $\hat{y}_{n}^{\text {full }}=\left(t_{n} \otimes \mathbf{I}_{d}\right) \boldsymbol{l}_{1: n-1}+\boldsymbol{t}_{n}$. Equating the right hand sides of this equation and (2.25) yields

$$
\begin{equation*}
\left.\left(t_{n} \otimes \mathbf{I}_{d}\right) \iota_{1: n-1}=\left(w_{n} \otimes \mathbf{I}_{d}\right)\left(T_{n-1}^{\prime} \otimes \mathbf{I}_{d}\right) \iota_{1: n-1}=\left(\left(w_{n} T_{n-1}^{\prime}\right) \otimes \mathbf{I}_{d}\right)\right) l_{1: n-1} \tag{2.26}
\end{equation*}
$$

The last equality above follows from the distributive property of tensor products. (2.26) is true for any information vector $t_{1: n-1}$. This implies $t_{n}=w_{n} T_{n-1}^{\prime} \Longrightarrow w_{n}=t_{n}\left(T_{n-1}^{\prime}\right)^{-1}$, since $T_{n}$ is an upper triangular matrix with ones on the diagonal ${ }^{14}$ and so invertible. To complete the proof that the left hand side of (2.13) implies the right hand side, recall that the information structure for constrained flow is such that if $a_{n}(j)=0$ then certain components of the vector $\widehat{y}_{1: n-1}$ are not available to node $n$. If the corresponding weight $w_{n}(j)$ is non-zero it is impossible to reconstruct $\widehat{y}_{n}$ according to (2.11) to be equal to $\widehat{y}_{n}^{\text {full }}$. This is simply the topological constraint (2.12). Showing that the right hand side of (2.13) implies the left hand side is very similar to the above proof and is omitted.

[^11]
## 3

## Mis-information Management Problem in Social Learning Over Directed Acyclic Graphs

### 3.1 Introduction

Motivated by online rating and review systems, we investigate social learning in a network where agents interact on a time dependent graph to estimate an underlying state of nature. Agents record their own private observations, then update their private beliefs about the state of nature using Bayes' rule. Based on their belief, each agent, then, chooses an action (rating) from a finite set and transmits this action over the social network. An important consequence of such social learning over a network is the ruinous multiple re-use of information known as data incest (or mis-information propagation). In this chapter, the data incest management problem in social learning context is formulated on a directed acyclic graph. We give necessary and sufficient conditions on the graph topology of social interactions to eliminate data incest. A data incest removal algorithm is proposed such that the public belief of social learning (and hence the actions of agents) is not affected by data incest propagation. This results in an online rating and review system with a higher trust rating. Numerical examples are provided to illustrate the performance of the proposed optimal data incest removal algorithm.

In social learning, agents aim to estimate the state of nature using their private observations and actions from other agents [2]. The process of updating belief by agents can be done using Bayesian models [1,64] or non-Bayesian models [56, 57]. Classical social learning is used to model the behavior of expected cost minimizer agents. Also, social learning can be generalized to of risk averse minimizers. The resulting risk-averse social learning filter is studied in [95].

In this chapter, we consider Bayesian social learning that models expected cost minimizers along with data incest (mis-information propagation). This results in a non-standard information pattern for Bayesian estimation. Before proceeding to the formal definition of data incest in learning over social networks, let us describe the social learning model.

### 3.1.1 Social Learning Protocol on Network

Consider a social network comprising of $S$ agents that aim to estimate (localize) an underlying state of nature (a random variable). Let $x \in\left\{\bar{x}_{1}, \bar{x}_{2}, \cdots, \bar{x}_{X}\right\}$ represent a state of nature (such as quality of a hotel) with known prior distribution $\pi_{0}$ where $X$ denotes the dimension of the state space. Let $k=1,2,3, \ldots$ depict epochs at which events occur. These events comprise of taking observations, evaluating beliefs and choosing actions as described below. The index $k$ depicts the historical order of events and not necessarily absolute time. However, for simplicity, we refer to $k$ as "time" in the rest of this chapter. Assume that there exists a network administrator who provides the network belief $\pi_{-\lfloor s, k\rfloor}$ defined in Step 5 to node $s$ at time $k$. Network belief can be considered as a summary of information received from nodes whose actions are available at node $[s, k]$ due to the constraints imposed by the structure of social network. The agents use the following Bayesian social learning protocol to estimate the state of nature:

Step 1. Private observations: To estimate the state of nature $x$, each agent records its $M$ dimensional private observation vector. At each time $k=1,2,3, \ldots$, each agent $s(1 \leq s \leq S)$ obtains a noisy private observation $z_{[s, k]}$ from the finite $\operatorname{set}^{15} \mathbf{Z}=\left\{\bar{z}_{1}, \bar{z}_{2}, \ldots, \bar{z}_{Z}\right\}$ with conditional probability

$$
\begin{equation*}
B_{i j}=p\left(z_{[s, k]}=\bar{z}_{j} \mid x=\bar{x}_{i}\right) \tag{3.1}
\end{equation*}
$$

It is assumed that the observations $z_{[s, k]}$ are independent random variables with respect to agent $s$ and time $k^{16}$.

Step 2. Private belief: After obtaining its private observation, each agent combines its private observation with the network belief to evaluate its private belief about state of nature. Each agent $s$ combines its private observation $z_{[s, k]}$ with the network belief (which is provided by the network administrator) and evaluates its private belief of state of nature ${ }^{17}$. Private belief, $\mu_{[s, k]}$, is evaluated via Bayesian models from the network belief and private observations, that is

$$
\begin{equation*}
\mu_{[s, k]}=\left(\mu_{[s, k]}(i), 1 \leq i \leq X\right), \text { where } \quad \mu_{[s, k]}(i)=p\left(x=\bar{x}_{i} \mid \pi_{-[s, k]}, z_{[s, k]}\right) \tag{3.2}
\end{equation*}
$$

Note that private belief of each agent is only available to herself and not to the other agents or the network administrator, (that is why the term "private" is used).

[^12]Step 3. Myopic action: Based on its private belief $\mu_{[s, k]}$, agent $s$ at time $k$ chooses an action $a_{[s, k]}$ from a finite set $\mathbf{A}=\{1,2, \ldots, A\}$ to minimize its expected cost function (based on the current information available on the network). That is

$$
\begin{equation*}
a_{[s, k]}=\underset{a \in \mathbb{A}}{\operatorname{argmin}} \mathbf{E}\left\{C(x, a) \mid \mu_{[s, k]}\right\} . \tag{3.3}
\end{equation*}
$$

Here $\mathbf{E}$ denotes expectation and $C(x, a)$ denotes the cost incurred by the agent if action $a$ is chosen when the state of nature is $x$. In the context of rating and review systems, cost function can be considered as the cost of loosing reputation in that review system. For example, if one under-rate a good restaurant, her reputation will be affected negatively and this is costly for her. After agent $s$ at time $k$ records its action $a_{[s, k]}$, the network administrator automatically computes the updated "public belief" at this node by combining action $a_{[s, k]}$ with network belief $\pi_{-[s, k]}$; that is, the public belief $\pi_{[s, k]}$ is

$$
\begin{equation*}
\pi_{[s, k]}=\left(\pi_{[s, k]}(i), 1 \leq i \leq X\right), \text { where } \quad \pi_{[s, k]}(i)=p\left(x=\bar{x}_{i} \mid \pi_{-[s, k]}, a_{[s, k]}\right) . \tag{3.4}
\end{equation*}
$$

Step 4. Social network: Unlike private beliefs, public beliefs are visible to the other agents and are broadcasted over the social network ${ }^{18}$. These public beliefs are observed by other agents after a random delay (communication delay). We model this information exchange using a family of directed acyclic graphs. Let

$$
\begin{equation*}
G_{[s, k]}=\left(V_{[s, k]}, E_{[s, k]}\right), \quad k=1,2,3, \ldots, s=1,2, \ldots, S, \tag{3.5}
\end{equation*}
$$

denote a sequence of time-dependent graphs of information flow in the social network until and including time $k$. Each vertex in $V_{[s, k]}$ represents an agent $s$ in the social network at time $k$ and each edge $\left(\left[s^{\prime}, k^{\prime}\right],\left[s^{\prime \prime}, k^{\prime \prime}\right]\right)$ in $E_{[s, k]} \subseteq V_{[s, k]} \times V_{[s, k]}$ shows that the public belief (or action) of agent $s^{\prime}$ at time $k^{\prime}$ reaches agent $s^{\prime \prime}$ at time $k^{\prime \prime}$.

Step 5. Network belief: For the past actions (those from other agents at previous time instants), the network administrator has already computed the public beliefs, see Step 3. Define

$$
\Theta_{[s, k]}=\left\{\pi_{[i, j]} ; \text { for all }[i, j] \in V_{[s, k]} \text { where }([i, j],[s, k]) \in E_{[s, k]}\right\} .
$$

[^13]At each node, the automated network administrator fuses all the available public beliefs (for example $\left.\Theta_{[s, k]}\right)$ into a single network belief; that is, the network belief is

$$
\begin{equation*}
\boldsymbol{\pi}_{-[s, k]}=p\left(x \mid \Theta_{[s, k]}\right)=\mathscr{A}\left(\Theta_{[s, k]}\right), \tag{3.6}
\end{equation*}
$$

where $\mathscr{A}$ denotes the information fusion algorithm used to aggregate the public beliefs received from the network. If algorithm $\mathscr{A}$ is not constructed properly, mis-information propagation (data incest) occurs.

As we will see shortly, a major issue with the above protocol with naive information aggregation in Step 5, is the inadvertent reuse of information (actions of previous agents) which makes the estimates of state of nature biased; that is data incest.

### 3.1.2 Chapter Goals

The above protocol models the interaction of agents in a social network that aim to estimate the underlying state of nature $x$. An example is where users aim to localize a target event by tweeting the location of the detected "target" on Twitter [132]. Another example is where the state of nature is the true quality of a social unit (such as restaurant). Online rating and review systems such as Yelp or Tripadvisor maintain logs of votes by agents (customers). Each agent visits a restaurant based on reviews on a review website such as Yelp. The agent then obtains private noisy measurement of the state (quality of food in a restaurant). The agent then reviews the restaurant on that review website. Such a review typically is a quantized version (for example, rating) of the total information (private belief) gathered by the agent ${ }^{19}$. With such a protocol, how can agents obtain a fair (unbiased) estimate of the underlying state? ${ }^{20}$. The aim of this chapter is for the network administrator to maintain an unbiased rating and review system, or alternatively modify the actions of agents, to avoid incest.

From a statistical signal processing point of view, estimating the state of nature $x$ using the above five-step protocol is non-standard in two ways: First, agents are influenced by the rating of other agents, this is prior influences their posterior and hence their rating. This effect of agents learning from the actions (ratings) of other agents along with their own private observation is termed "social learning" in the economics literature. Social learning can result in an interesting phenomenon where rational agents can all end up making the same decision (herding and information cascades; [35]). Second, (and this effect is more complex), an agent might be influenced by his own rating leading

[^14]
## Examples of Social Learning in Social Networks

| Target Localization | Online Rating and Review System |
| :--- | :--- |
| Aim: To estimate location of a target Aim: To estimate quality of a social unit <br> $x:$ Geographical coordinates of the target $x:$ (State of nature) quality of the social unit <br> $a_{n}:$ (action) region of detected target $a_{n}:$ (action) rating of the social unit |  |

Figure 3.1: Two examples of multi-agent social learning in social networks: (i) target localization, and (ii) online rating and review systems.
to data incest.
To explain what can go wrong with the above protocol, suppose an agent wrote a poor rating of the restaurant on a social media site at time 1 . Another agent is influenced by this rating and also gives the restaurant a poor rating at time 2. Assume that the information exchange is modeled by the graph depicted in Fig.3.2. The first agent visits the social media site at time 3 and sees that another agent has also given the restaurant a poor rating - this double confirms his rating and he enters another poor rating. In a fair system, the first agent should have been aware that the rating of the second agent was influenced by his rating - so that first agent has effectively double counted his first rating by casting the second poor rating. Data incest is a consequence of the recursive nature of Bayesian social learning and the communication graph. The data incest in a social network is defined as the naive re-use of actions of other agents in the formation of the belief of an agent when these actions could have been initiated by the agent. In Figure 3.2, the fact that there exist two distinct paths between Agent 1 at time 1 and Agent 1 at time 3 (depicted in red) implies that the information of Agent 1 at time 1 is double counted, thereby leading to a data incest event.

The twin effects of social learning and data incest lead to non-standard information patterns in state estimation. Herd occurs when the public belief overrides the private observations and thus actions of agents are independent of their private observations. An extreme case of this is an information cascade when the public belief of social learning hits a fixed point and does not evolve any longer. Each agent in a cascade acts according to the fixed public belief and social learning stops $[35]^{21}$. Data incest results in bias in the public belief as a consequence of the unintentional re-use of identical actions in the formation of public belief of social learning; the information gathered by

[^15]

Figure 3.2: Example of communication graph, with two agents ( $S=2$ ) and over three event epochs ( $K=3$ ). The arrows represent exchange of information regarding actions taken by agents.
each agent is mistakenly considered to be independent. This results in over confidence and bias in estimates of state of nature. Due to the lack of information about the topology of the communication graph, data incest arises in Bayesian social learning in social networks. Therefore, the Bayesian social learning protocol requires a careful design to ensure that data incest is mitigated. The aim of this chapter is to modify the five-step protocol presented in Section 3.1.1 such that data incest does not arise. As we will see in Section 3.3.4, the proposed data incest removal algorithm can be applied to the state estimation problems shown in Fig.3.1.

### 3.1.3 Main Results and Organization of Chapter

With the above five-step social learning protocol in social networks, we are now ready to outline the main results of this chapter:

1. In Section 3.2, the data incest problem is formulated on a family of time dependent directed acyclic graphs
2. In Section 3.3, a necessary and sufficient condition on the graph is provided for exact data incest removal. This constraint is on the topology of communication delays (communication graph). Also examples where exact incest removal is not possible, are illustrated.
3. A data incest removal algorithm is proposed for the five-step social learning protocol in Section 3.3. The data incest removal algorithm is employed by the network administrator to update the network belief in Step 5 of the social learning protocol of Section 3.1.1 ${ }^{22}$.
[^16]Finally in Section 3.4, numerical examples are provided which illustrate the data incest removal algorithm.

### 3.2 Social Learning Over Social Networks

The five-step social learning protocol is introduced in Section 3.1.1. We also discussed that as a result of the loopy information exchange graph, data incest (or mis-information propagation) arises because of the abusive re-use of information of the other agents. Here, with the graph theoretic definitions provided in Appendix A, we discuss the diffusion of information in the social network in more details. Before proceeding, for notational simplicity, instead of $[s, k]$, the following scalar index $n$ is used:

$$
\begin{equation*}
n \triangleq s+S(k-1), \quad s \in\{1, \ldots, S\}, k \in\{1,2,3, \ldots\} \tag{3.7}
\end{equation*}
$$

Note that, in the social learning model considered in this chapter, the historical order of events is important and $k$ is used to denote the order of occurrence of events in real time. Subsequently, we will refer to $n$ as a "node" of the time dependent communication graph $G_{n}$. Recall from Section 3.1, $G_{n}=\left(V_{n}, E_{n}\right)$ denotes the time-dependent communication graph of the social network. Each node $n^{\prime}$ in $G_{n}$ represents an agent $s^{\prime}$ at time $k^{\prime}$ such that $n^{\prime}=s^{\prime}+S\left(k^{\prime}-1\right)$, see (3.7). Each directed edge of $G_{n}$ shows a communication link in the social network represented by $G_{n}$. This means that if $\left(n, n^{\prime}\right) \in E_{n}$, agent $s^{\prime}$ at time $k^{\prime}$ uses the information of agent $s$ at time $k$ to update his private belief about the underlying state of nature $x$. Note that with the way we defined the communication graph, $G_{n}$ is always a sub-graph of $G_{n+1}$. Therefore, as the following theorem proves, diffusion of actions can be modeled via a family of time-dependent Directed Acyclic Graphs (DAGs) ${ }^{23}$.

Theorem 3.2.1. The information flow in a social learning over social networks comprising of $S$ agents for $k=1,2,3, \ldots, K$ can be represented by a family of DAGs $\mathscr{G}=\left\{G_{n}\right\}_{n \in\{1, \ldots, N\}}$ where $N=S K$. Each $D A G G_{n}=\left(V_{n}, E_{n}\right)$ represents the information flow between the $n$ first nodes, where the generic node $n$ is defined by (3.7).

Proof. The proof is similar to that presented in Section 2.7.1.
The adjacency and the transitive closure matrices of $G_{n}$ are denoted by $\bar{A}_{n}$ and $T_{n}$, respectively (see Appendix A). Using the adjacency and transitive closure matrices of $G_{n}$, the following two sets which have involved in formulation of data incest problem in social learning over networks, are introduced:

$$
\begin{align*}
& \mathscr{F}_{n}=\left\{k, \quad \bar{A}_{n}(k, n)=1\right\}, \\
& \mathscr{H}_{n}=\left\{k, \quad T_{n}(k, n)=1\right\} . \tag{3.8}
\end{align*}
$$

[^17]Protocol 1: Constrained Social Learning in Social Networks


Figure 3.3: Protocol 1: Constrained social learning in social networks described in Section 3.1.1. As a result of random (unknown) communication delays, data incest arises.

In words, $\mathscr{F}_{n}$ consists of all nodes who have a single-hop link (edge) to node $n$ and $\mathscr{H}_{n}$ includes the ones with either a single-hop or multi-hop (path) link to node $n$.

### 3.2.1 Constrained Social Learning in Social Networks

The five-step constrained social learning protocol introduced in Section 3.1.1, is illustrated in Fig.3.3. Note that in the constrained social learning problem, agents do not have information about the communication graph. This is why the term "constrained" is used. The constrained social learning in social networks can be summarized in an abstract form as

$$
\begin{align*}
& \text { Choose action } a_{n}=\underset{a \in \mathbb{A}}{\operatorname{argmin}} \mathbf{E}\left\{C_{a}^{\prime} \mu_{n}\right\} \text { subject to: }  \tag{3.9}\\
& \left\{\begin{array}{l}
z_{n} \sim B_{i z}, \quad x=i, \quad \text { (observation process) } \\
\Theta_{n}=\left\{\pi_{i} ; \quad i \in \mathscr{F}_{n}\right\}, \quad \text { (network constraint) } \\
\pi_{-n}=\mathscr{A}\left(\Theta_{n}\right), \quad \text { (network belief formation) } \\
\mu_{n}=p\left(x \mid \pi_{-n}, z_{n}\right), \quad \text { (private belief evaluation) }
\end{array}\right.
\end{align*}
$$

In (3.9) Algorithm $\mathscr{A}$ is the automated information fusion algorithm employed by the network administrator to compute the network belief. Due to the loopy communication graph and the recursive nature of Bayesian models, data incest (mis-information propagation) arises in constrained social learning if algorithm $\mathscr{A}$ is not designed properly. The aim of this chapter is to devise the algorithm $\mathscr{A}$ such that the public belief of social learning (and consequently, actions $a_{n}$ for all $n=1,2, \ldots$ ) are not affected by data incest.

The following lemma summarizes the social learning filters in (3.9).

Lemma 3.2.1. Consider the five-step social learning protocol presented in Section 3.1.1 with $S$ agents and the communication graph $G_{n}$. Let $\pi_{-n}$ denote the network belief of social network at this node. Then, the social learning elements (private belief, action, and public belief) of node $n$ with observation vector $z_{n}=\bar{z}_{l}$ can be computed from $(1 \leq m \leq X)$

$$
\begin{align*}
& \mu_{n}(m)=p\left(x=\bar{x}_{m} \mid \Theta_{n}, z_{n}\right) \propto c \pi_{-n}(m) B_{m l}, \\
& a_{n}=\underset{a \in \mathbb{A}}{\operatorname{argmin}} \mathbf{E}\left\{C(x, a) \mid \Theta_{n}, z_{n}\right\}=\underset{a \in \mathbb{A}}{\operatorname{argmin}} \mathbf{E}\left\{C_{a}^{\prime} \mu_{n}\right\}, \\
& \pi_{n}(m) \propto c \pi_{-n}(m) \sum_{j=1}^{Z}\left[\prod_{\widehat{a} \in \mathbf{A}-\left\{a_{n}\right\}} \mathbb{I}\left(C_{a_{n}}^{\prime} B_{j} \pi_{-n}<C_{\widehat{a}}^{\prime} B_{j} \pi_{-n}\right)\right] B_{m j}, \tag{3.10}
\end{align*}
$$

where $c$ is a generic normalizing constant, $B_{j}=\operatorname{diag}\left(B_{1} j, \ldots, B_{X j}\right)$, and $\mathbb{I}(\cdot)$ is indicator function. Here, $C_{a}$ is the cost vector defined as $C_{a}=\left[\begin{array}{lll}C(1, a) & C(2, a) & \ldots C(X, a)] \text {. Using a matrix notation, }\end{array}\right.$

$$
\mu_{n}=\frac{B_{\overline{\bar{z}}} \pi_{-n}}{\mathbf{1}^{\prime} \cdot B_{\bar{z}} \pi_{-n}}
$$

where $\mathbf{1}$ denotes a all-one vector. Also, the public belief $\pi_{n}$ can be written as

$$
\pi_{n}=\frac{R_{a_{n}}^{\pi_{-n}} \pi_{-n}}{\mathbf{1}^{\prime} \cdot R_{a_{n}}^{\pi_{n}} \pi_{-n}}
$$

Here, $R_{a_{n}}^{\pi_{-n}}=\operatorname{diag}\left(r_{1}, \ldots, r_{X}\right)$, where $r_{m}=\sum_{j=1}^{Z}\left[\prod_{\hat{a} \in \mathbf{A}-\left\{a_{n}\right\}} \mathbb{I}\left(C_{a_{n}}^{\prime} B_{j} \pi_{-n}<C_{\widehat{a}}^{\prime} B_{j} \pi_{-n}\right)\right] B_{m j}$.
Proof. The proof is presented in Section 3.7.1.
Lemma 3.2.1 summarizes the social learning problem considered in this chapter. Each node combines the network belief with its private observation to evaluate its private belief. Based on this private belief, action $a_{n}$ is chosen such that a local cost function is minimized. Action $a_{n}$ is used by the network administrator to automatically update the public belief of social learning. Then, the public belief is transmitted over the network. As described in Section 3.1, a major issue with the above protocol is data incest. The aim of this chapter is to devise a data incest removal algorithm for the network administrator to deploy such that the estimates of agents are unbiased.

Remark 3.2.1. In order to choose an action from the finite set of all possible actions, agents minimize a cost function. This cost function can be interpreted in terms of the reputation of agents in online rating and review systems. For example if the quality of a restaurant is good and an agent wrote a bad review for it in Yelp and he continues to do so for other restaurants, his reputation becomes lower among the users of Yelp. Consequently, other people ignore reviews of that (low-reputation) agent in evaluation of their opinion about the social unit under study (restaurant).

Therefore, agents minimize the penalty of writing inaccurate reviews (or equivalently increase their reputations) by choosing proper actions. This behavior is modeled by minimizing a cost function in our social learning model.

Remark 3.2.2. In comparison to the public belief which can be computed by the network administrator (who monitors the agents' actions and communication graph), the agents' private beliefs cannot be computed by the network administrator. The private belief depends on the local observation which is not available to the network. Note that in Step 2 of the constrained social learning Protocol 1, the results of Lemma 3.2.1 are used to compute $\mu_{n}$ using $z_{n}$ and $\pi_{-n}$.

Remark 3.2.3. The constrained social learning protocol is practiced in many online rating and review systems such as Yelp or Tripadvisor ${ }^{24}$

### 3.3 Data Incest Removal Algorithm

So far in this chapter, Bayesian social learning model and communication amongst agents in social networks have been described. This section presents the main result of this chapter, namely the solution to the constrained social learning problem (3.9). We propose a data incest removal algorithm such that the public belief of social learning (and consequently the chosen action) is not affected by data incest. To devise the data incest removal algorithm, an idealized framework is presented that prevents data incest as we will describe shortly. Comparing the public belief of the idealized framework with the same of the constrained social learning, the data incest removal algorithm is specified. This data incest removal algorithm is used by the network administrator and replaces Step 5 of the social learning protocol presented in Section 3.1.1. A necessary and sufficient condition for the data incest removal problem is also presented in this section.

### 3.3.1 The Idealized Benchmark for Data Incest Free Social Learning in Social Networks

In this subsection, an idealized (and therefore impractical) framework that will be used as a benchmark to derive the constrained social learning protocol, is described. In the idealized protocol, it is assumed that the entire history of actions along with the communication graph are known at each node. Due to the knowledge about the entire history of actions and the communication graph (dependencies among actions) in the idealized framework, data incest does not arise ${ }^{25}$. Define

$$
\begin{equation*}
\Theta_{n}^{\text {full }}=\left\{a_{i} ; \quad i \in \mathscr{H}_{n}\right\}, \tag{3.11}
\end{equation*}
$$

[^18]Protocol 2: Idealized Social Learning in Social Networks


Figure 3.4: Protocol 2: Idealized benchmark social learning in social networks. In this protocol, the complete history of actions chosen by agents and the communication graph are known. Hence, data incest does not arise. This benchmark protocol will be used to design the data incest removal protocol.
where, $\mathscr{H}_{n}$ is defined in (3.8). In the idealized framework, the network belief can be written as

$$
\begin{equation*}
\pi_{-n}^{\mathrm{full}}=p\left(x \mid \Theta_{n}^{\mathrm{full}}\right) \propto \pi_{0} \prod_{a_{i} \in \Theta_{n}^{\text {full }}} p\left(a_{i} \mid x, S_{i}\right) \tag{3.12}
\end{equation*}
$$

where $S_{i} \subset \Theta_{n}^{\text {full }}$ denotes the set of actions that $a_{i}$ depends on them. The public belief in the idealized social learning is free of data incest, as it can be inferred from (4.37). The idealized social learning in social networks (Protocol 2) is illustrated in Fig.3.4. The private belief of node $n$ in the idealized social learning is denoted by $\widehat{\mu}_{n}$.

Note that if there exists a path between node $i$ and node $n$, then action $a_{i} \in \Theta_{n}^{\text {full }}$. Since the history of actions and the dependencies among them (communication topology) are available in the idealized social learning, $\pi_{-n}^{\text {full }}$ is free of data incest.

### 3.3.2 The Data Incest Free Belief in the Idealized Social Learning Protocol 2

The goal of this chapter is to replace Step 5 of the five-step constrained social learning protocol with an algorithm that mitigates data incest. As described earlier, to solve the data incest management problem, we introduced the idealized social learning that prevents data incest. By comparing the network belief (or equivalently the public beliefs of agents) in the idealized social learning with that in the constrained social learning Protocol 1 , the data incest removal algorithm can be invented. Our aim is to devise algorithm $\mathscr{A}$ in (3.9)-also in Step 5 of the five-step social learning protocol in

Section 3.1- such that

$$
\begin{equation*}
p\left(x \mid \Theta_{n}^{\text {full }}\right)=\pi_{-n} \tag{3.13}
\end{equation*}
$$

Here, first, an expression is derived for the public beliefs of agents in the idealized social learning Protocol 2. Then, using that, algorithm $\mathscr{A}$ is constructed such that (3.13) holds; that is, data incest is mitigated. Let $\theta_{n}^{\text {full }}$ denote the logarithm of public belief of node $n$ in the idealized social learning Protocol 2, that is

$$
\begin{equation*}
\theta_{n}^{\text {full }}=\log \left(p\left(x \mid \Theta_{n}^{\text {full }}, a_{n}\right)\right) \tag{3.14}
\end{equation*}
$$

Theorem 3.3.1 below gives an expression for $\theta_{n}^{\text {full }}$ in the idealized social learning Protocol 2.
Theorem 3.3.1. Consider problem (2.6) with the idealized social learning Protocol 2. The data incest free public belief of node $n$ (which represents agent $s$ at time $k$ according to re-indexing equation (3.7)) is:

$$
\begin{equation*}
\theta_{n}^{\text {full }}=\sum_{i=1}^{n-1} t_{n}(i) v_{i}+v_{n} \tag{3.15}
\end{equation*}
$$

where $v_{k}$ denotes $\log \left(p\left(a_{k} \mid x, S_{k}\right)\right)$.Recall that $t_{n}$ defined in (A.6) in Appendix $A$ as the first $n-1$ elements of the $n^{\text {th }}$ column of $T_{n}$.

Proof. The proof is presented in Section 3.7.2.
As can be seen in (3.15), the (logarithm of the) public belief of node $n$ can be written as a linear function in terms of $v_{i}$ using $t_{n}$. Due to this linearity, the data incest removal algorithm can be constructed as we will explain later in this section. Also (3.15) implies that the optimal data incest free public beliefs of agents in the idealized social learning Protocol 2 depend on the communication graph explicitly in terms of the transitive closure matrix ${ }^{26}$. Basically the non-zero elements of $t_{n}$ show all nodes who have a path to node $n$ and thus their actions contribute in the formation of the private belief of node $n$. Eq. (3.15) is quite intuitive from the fact that each agent employs a recursive Bayesian filter to combine its private observation with the information received from the network.

### 3.3.3 Data Incest Removal Algorithm for Problem (3.9) With Constrained Social Learning Protocol 1

Given the expression for the public belief of the idealized social learning Protocol 2, the aim here is to propose an optimal information aggregation scheme (that replaces Step 5) such that the public belief of the constrained social learning Protocol 1 is equal to the same of the idealized social

[^19]learning Protocol 2 (which is free of data incest). That is, (3.13) holds or equivalently
\[

$$
\begin{equation*}
p\left(x \mid a_{i} ; \quad i \in \mathscr{H}_{n}\right)=p\left(x \mid \pi_{i} ; \quad i \in \mathscr{F}_{n}\right) \tag{3.16}
\end{equation*}
$$

\]

Similar to $\theta_{n}^{\text {full }}$, let $\widehat{\theta}_{n}$ denote the logarithm of the after action public belief of node $n$,

$$
\begin{equation*}
\widehat{\theta}_{n}=\log \left(p\left(x \mid \Theta_{n}, a_{n}\right)\right) \tag{3.17}
\end{equation*}
$$

We propose the following optimal information aggregation scheme to evaluate the public belief using a $n-1$ dimensional weight vector $w_{n}$ as follows,

$$
\begin{equation*}
\widehat{\theta}_{n}=\sum_{i=1}^{n-1} w_{n}(i) \widehat{\theta}_{i}+v_{n} \tag{3.18}
\end{equation*}
$$

where $w_{n}$ with elements $w_{n}(i)(1 \leq i \leq n-1)$ is defined more precisely in (3.20). Using optimal information aggregation scheme (3.18) and (3.10) in Lemma 3.2.1, algorithm $\mathscr{A}$ in (3.9) can be specified.

Remark 3.3.1. The optimal information aggregation scheme (3.18) is deployed by the automated network administrator in Step 5 of the social learning protocol presented in Section 3.1.1 to combine the received information (beliefs or equivalently actions) form other nodes and compute $\sum_{i=1}^{n-1} w_{n}(i) \widehat{\theta}_{i}$, this is the network belief at node n. Then, node $n$ updates its private belief based on the most updated network belief (provided by the network administrator) and chooses its action $a_{n}$ accordingly and then transmits it over the network. Then, the network administrator automatically evaluates $v_{n}$ and updates public belief by computing $\widehat{\theta}_{n}=\sum_{i=1}^{n-1} w_{n}(i) \widehat{\theta}_{i}+v_{n}$.

The weight vector $w_{n}$ depends on the communication graph and can be computed simply by (3.20). Theorem 3.3.2 below asserts that by using the optimal information aggregation scheme (3.18) with $w_{n}$ defined in (3.20), data incest can be completely mitigated. However, for some network topologies, it is not possible to remove data incest completely. The following constraint presents the necessary and sufficient condition on the network for the exact data incest removal.

Constraint 3.3.1. Consider the constrained social learning problem (3.9) with Protocol 1. Then, the weight vector $w_{n}$ used in optimal information aggregation scheme (3.18) satisfies the topological constraints if $\forall j \in\{1, \ldots, n-1\}$ and $\forall n \in\{1, \ldots, N\}$

$$
\begin{equation*}
b_{n}(j)=0 \quad \Longrightarrow w_{n}(j)=0 \tag{3.19}
\end{equation*}
$$

where $b_{n}$ is defined in (A.6) and denotes the $n$-th column of the adjacency matrix of $G_{n}$. Basically

Constraint 3.3.1 puts the "availability constraint" on the communication graph.
This means that if information of node $j$ is required at node $n\left(w_{n}(j) \neq 0\right)$, there should be a communication link between node $j$ and node $n\left(b_{n}(j) \neq 0\right)$. Assuming that Constraint 1 holds, Theorem 3.3.2 below ensures that the public belief of nodes in problem (3.9) with the constrained social learning Protocol 1 is identical to the same of the problem (2.6) with the idealized social learning Protocol 2.

Theorem 3.3.2. Consider problem (3.9) with the constrained social learning Protocol 1 of Section 3.2. Then using the optimal information aggregation scheme (3.18), data incest can be mitigated by using the optimal set of weights $\left\{w_{n}\right\}_{n \in\{1, \ldots, N\}}$ given that the topological Constraint 3.3.1 is satisfied. The optimal weight vector is

$$
\begin{equation*}
w_{n}=t_{n}\left(\left(T_{n-1}\right)^{\prime}\right)^{-1} . \tag{3.20}
\end{equation*}
$$

By using the optimal combination scheme (3.18) and optimal weight vector defined in (3.20), the data incest in social learning problem (3.9) is completely mitigated, that is $\widehat{\theta}_{n}=\theta_{n}^{\text {full }}$ if $w_{n}$ satisfies topological Constraint 3.3.1 where $\widehat{\theta}_{n}$ and $\theta_{n}^{\text {full }}$ are defined in (3.17) and (3.14) respectively. Recall that $t_{n}$ is defined in (A.6) as the first $n-1$ elements of the $n^{\text {th }}$ column of $T_{n}$.

Proof. The proof is presented in Section 3.7.3.
Theorem 3.3.2 proves that if $w_{n}=t_{n}\left(\left(T_{n-1}\right)^{\prime}\right)^{-1}$ then $\widehat{\theta}_{n}=\theta_{n}^{\text {full }}$, that is, data incest (misinformation propagation mitigated).

Using the optimal information aggregation scheme (3.18), the five-step Bayesian social learning protocol in Section 3.1.1 with data incest removal algorithm can be summarized as

```
Algorithm 3.4 Constrained Bayesian social learning with data incest removal algorithm at each
node \(n\)
Step 1. Observation process: Private observation vector \(z_{n}\) is obtained according to (3.1).
Step 2. Private belief: Node \(n\) accesses the network and evaluates its private belief according to
```

(3.2) using the most recent network belief at node $n$,

$$
\mu_{n}=\frac{B_{\bar{z}_{l}} \pi_{-n}}{\mathbf{1}^{\prime} \cdot B_{\bar{z}_{l}} \pi_{-n}}
$$

Step 3. Myopic action: Action $a_{n}$ is chosen via (3.3), that is

$$
a_{n}=\underset{a \in \mathbb{A}}{\operatorname{argmin}} \mathbf{E}\left\{C_{a}^{\prime} \mu_{n}\right\} .
$$

Then, the automated network administrator compute the public belief at node $n$ which is defined in (3.4), that is

$$
\pi_{n}=\frac{R_{a_{n}}^{\pi_{-n}} \pi_{-n}}{\mathbf{1}^{\prime} \cdot R_{a_{n}}^{\pi_{n}} \pi_{-n}}
$$

Step 4. Social network: Social network model is similar to the same in Step 4 of the protocol presented in Section 3.1.1.
Step 5. Network belief update: The automated network administrator evaluates the network belief using $\Theta_{n}=\left\{\pi_{i}, \quad i \in \mathscr{H}_{n}\right\}$ and the optimal weight vector

$$
w_{n}=t_{n}\left(\left(T_{n-1}\right)^{\prime}\right)^{-1} .
$$

Discussion of topological constraint (3.19): The non-zero elements of $w_{n}$ are corresponding to the nodes whose information are required at node $n$ to remove data incest. This imposes a topological constraint on the communication graph. If $w_{n}(j)$ is non-zero, this means that information of node $j$ is needed at node $n$ and there should be an edge in $G_{n}$ that connects node $j$ to node $n$, this is the topological Constraint 3.3.1. Constraint 3.3.1 ensures that the essential elements for data incest removal are available at node $n$ and Theorem 3.3.2 specifies the exact data incest removal algorithm. From Theorem 3.3.2, it is simple to show that Constraint 3.3.1 is a necessary and sufficient condition for data incest removal in learning problem (3.9). Consider two examples of communication graph shown in Figure 3.5.

The optimal weight vector at node 5 for both networks of Fig. 3.5 computed from (3.20) is $w_{5}=[-1,-1,1,1]$. This means that there should be a link between node 2 and node 5 for exact data incest removal according to the topological constraint (3.19). Hence, Constraint 3.3.1 does not hold for the network of Figure 3.5b, while the topological constraint is satisfied in network depicted in Fig.3.5a. Also as it is clear from the network shown in Fig.3.5a, there is no need for the communication graph to be a tree.

(a)

(b)

Figure 3.5: Two examples of networks: (a) satisfies the topological constraint, and (b) does not satisfy the topological constraint.

### 3.3.4 Discussion of Data Incest Removal in Social Learning

Here, we discuss the application of data incest removal Algorithm 3.4 (presented in Theorem 3.3.2) in two examples of multi-agent state estimation problem which are presented in the introductory part of this chapter (i) online rating and review systems, and (ii) target localization using social networks, see Fig.3.1. Both problems can be formulated using the five-step constrained social learning protocol presented in Section 3.1.1. As illustrated in Fig.3.6, agents observe the underlying state of nature in noise and practice social learning to choose an action such that a local cost function is minimized. But as a result of unknown communication graph and the recursive nature of Bayesian estimators, data incest or abusive re-use of information occurs. To mitigate data incest, the network administrator plays an intermediating role. Instead of transmitting the communication graph and complete history of actions, the network administrator monitors all the information exchanges and provides the data incest free network belief of social learning at each node. To compute the data incest free public belief, the network administrator uses the optimal information aggregation scheme (3.18) with the optimal weight vector $w_{n}$, see (3.20). Using the most updated public belief and its own private observation $z_{n}$, node $n$ evaluates its private belief. Based on this private belief (which is free of data incest), action $a_{n}$ is chosen and transmitted it over the network. Given that the communication graph satisfies the topological Constraint 1, Theorem 3.3.2 ensures that by means


Figure 3.6: Data incest removal algorithm employed by network administrator in the state estimation problem over social network. The underlying state of nature could be geographical coordinates of an event (target localization problem) or reputation of a social unit (online rating and review systems).
of the optimal weight vector $w_{n}$, action $a_{n}$ is not affected by data incest and, therefore, performance of state estimation process is improved.

### 3.4 Numerical Examples

In this section, numerical examples are given to illustrate the performance of data incest removal Algorithm 3.4 presented in Section 3.3. As described in the five-step protocol of Section 3.1, agents interact on a graph to estimate an underlying state of nature (which represents the location of a target event in target localization problem, or the reputation of a social unit in online rating and review systems). The underlying state of nature $x$ is a random variable uniformly chosen from $\mathbf{X}=\{1,2, \cdots, 20\}$, and actions are chosen from $\mathbf{A}=\{1,2, \ldots 10\}$. We consider the following three scenarios for each of four different types of social networks:

1. Constrained social learning without data incest removal algorithm (data incest occurs) depicted with dash-dot line
2. Constrained social learning with Protocol 1 with data incest removal algorithm depicted with dashed line
3. Idealized framework where each node has the entire history of raw observations and thus data incest cannot propagate. This scenario is only simulated for comparison purposes and is depicted by solid line.

The effect of data incest on estimation problem and the performance of the data incest removal algorithm, proposed in Section 3.3, is investigated for the networks shown in Fig.3.7.

We first consider a communication graph with 41 nodes. The communication graph under study, which is shown in Fig.3.7a, satisfies the topological constraint (3.19). The action of node 1 reaches all other nodes and node 41 receives all actions of previous 40 nodes (some edges are omitted from the figure to make it more clear).


Figure 3.7: Three different communication topologies: (a) the communication graph with 41 nodes, (b) agents interact on a fully interconnected graph and the information from one agent reach other agents after a delay chosen randomly from $\{1,2\}$ with the same probabilities, (c) star-shaped communication topology with random delay chosen from $\{1,2\}$.

As can be seen in Fig.3.8, data incest makes agents' actions in the constrained social learning without data incest removal different from the same in the idealized framework. Also Fig.3.8 corroborates the excellent performance of data incest removal Algorithm 3.4. As illustrated in Fig.3.8, the actions of agents in social learning with data incest removal algorithm are exactly similar to those of the idealized framework without data incest. The social learning problem over the graph shown in Fig.3.7a is simulated 100 times to investigate the difference between the estimated state of nature with the true one $(x=10)$. The estimates of state of nature (obtained in three different scenarios discussed in the beginning of the section) are depicted in Fig.3.9. As can be seen from the figure, the estimates obtained with data incest removal algorithm are very close to data incest
free estimates of Scenario (iii). The bias in estimates in presence of data incest is also clear in this figure.

In the next simulation, a different communication topology is considered. We repeat the simulation for a star-shaped communication graph comprising of six agents $(S=6)$ at four time instants, $K=4$, so the total number of nodes in the communication graph is 24 , see Fig.3.7c. The communication delay is randomly chosen from $\{1,2\}$ with the same probabilities. We simulated the social learning in three different scenarios discussed above, to investigate the effect of data incest on the actions and the estimates of agents in the star-shaped social network. The actions chosen by nodes are depicted in Fig.3.10. As can be seen from Fig.3.10, using the data incest removal algorithm, the agents' actions in the constrained social learning with Protocol 1 are very close to those of the idealized social learning with Protocol 2 which are free of data incest. Also the estimates of state of nature are very close to the true value of state of nature compared to the constrained social learning without data incest removal algorithm. Also note that the effect of data incest, as expected, in this communication topology is different for each agent; the agent who communicates with all other nodes is affected more by data incest. This fact is verified in Figures 3.10 and 3.11.

In the third example, a complete fully interconnected graph (where agents communicate with all other agents) is considered. In this example, action of each agent becomes available at all other agents after a random delay chosen from $\{1,2\}$ with the same probabilities. The agents' actions are shown in Fig.3.12. Similar to the star-shaped graph, using data incest removal Algorithm 3.4 makes the agents' actions in the constrained social learning very similar to those of the idealized (data incest free) framework. Also, the excellent performance of data incest removal Algorithm 3.4 in the estimation problem is depicted in Fig.3.13.

We also extend our numerical studies to an arbitrary random network with five agents, $S=$ $5, K=4$. We consider a fully connected network and assume that the interaction between two arbitrary agents (say agent $i$ and agent $j$ ) at time $k$ has four (equiprobable) possible statuses: (i) connected with delay 1 , (ii) connected with delay 2 , (iii) connected with delay 3 , and (iv) not connected. If the link is connected with delay $\tau$, this means that the information from agent (i) at time $k$ becomes available at agent $j$ at time $k+\tau$. If the link is not connected, the information of agent $i$ at time $k$ never reaches agent $j$. We verify that the underlying communication graph, $G_{n}$, satisfies the topological Constraint 3.3.1. Fig.3.14 depicts the agents' actions in three different scenarios (with data incest, without data incest, and with data incest removal algorithm). The simulation results show that, even in this case with arbitrary network (that satisfies topological constraint), the actions obtained by the constrained social learning with data incest removal algorithm is very close to those in the idealized social learning. As expected, using the data incest removal algorithm, the data incest associated with the estimates of agents can be mitigated completely, as shown in Fig.3.15.

Here, we also present numerical studies to investigate the accuracy of the state estimation considered in this chapter in terms of mean squared error. The mean squared error of estimates obtained


Figure 3.8: Actions of agents obtained with social learning over social networks in three different scenarios described in Section 3.4 with communication graph depicted in Fig.3.7a.


Figure 3.9: Mean of the estimated state of nature in the state estimation problem with social learning over social networks in three different scenarios described in Section 3.4 with communication graph depicted in Fig.3.7a.


Figure 3.10: Actions of agents obtained with social learning over social networks in three different scenarios described in Section 3.4 with communication graph depicted in Fig.3.7c.


Figure 3.11: Mean of the estimated state of nature in the state estimation with social learning over social networks in three different scenarios described in Section 3.4 with communication graph depicted in Fig.3.7c.


Figure 3.12: Actions of agents obtained with social learning over social networks in three different scenarios described in Section 3.4 with communication graph depicted in Fig.3.7b.


Figure 3.13: Mean of the estimated state of nature in the state estimation problem with social learning over social networks in three different scenarios described in Section 3.4 with communication graph depicted in Fig.3.7b.


Figure 3.14: Actions of agents obtained with social learning over social networks in three different scenarios described in Section 3.4 with arbitrary communication graph.


Figure 3.15: Mean of the estimated state of nature in the state estimation problem with social learning over social networks in three different scenarios described in Section 3.4 with arbitrary communication graph.
in social learning with three different scenarios discussed in the beginning of this section (with data incest, with data incest removal algorithm, and the idealized framework) is computed for each of four scenarios considered in our numerical studies. Fig. 3.16 depicts the mean squared error of estimates obtained in the first example with social learning over the communication graph of Fig. 3.7a. As can be seen from this figure, the mean squared error associated with the estimates of the constrained social learning with data incest removal Algorithm 3.4 are lower than the those of the constrained social learning in presence of data incest. This means that the performance of estimation problem with social learning is improved using data incest removal algorithm proposed in this chapter.

Figures 3.17 and 3.18 show the mean squared error of estimation with communication graphs presented in Figures 3.7 b and 3.7 c , respectively, and Figure 3.19 depicts the same for the arbitrary random network with five agents and random communication delays described earlier in this section. As can be seen in these figure, as a result of herding, in star shaped and random communication topologies, the mean squared error of estimates is slightly (compared to the scenario without data incest removal algorithm) more than the idealized framework at each time.

### 3.5 Psychology Experiment

This section presents an experimental study to investigate the learning and decision making behavior of individuals in a human society. Social learning is used as the mathematical basis for modeling interaction of individuals that aim to perform a perceptual task interactively. A psychology experiment was conducted on a group of undergraduate students at the University of British Columbia to examine whether the decision (action) of one individual affects the decision of the subsequent individuals. The major experimental observation that stands out here is that the participants of the experiment (agents) were affected by decisions of their partners in a relatively large fraction ( $60 \%$ ) of trials. We fit a social learning model that mimics the interactions between participants of the psychology experiment. Mis-information propagation (also known as data incest) within the society under study is further investigated in this experiment.

### 3.5.1 Experiment Setup

Here, a detailed description of the psychology experiment we carried out to study the learning behavior of individuals in a human society is presented:

- Experiment Date: The psychology experiment was conducted in September and October 2013.
- Society under study: The participants were 36 undergraduate students of the department of Psychology of the University of British Columbia who completed the experiment for course


Figure 3.16: Mean squared error of estimates (of state of nature) obtained with social learning with communication graph depicted in Fig.3.7a.


Figure 3.17: Mean squared error of estimates (of state of nature) obtained with social learning with communication graph depicted in Fig.3.7b (complete fully interconnected graph).


Figure 3.18: Mean squared error of estimates (of state of nature) obtained with social learning with communication graph depicted in Fig.3.7c (star-shaped communication graph).


Figure 3.19: Mean squared error of estimates (of state of nature) obtained with social learning with arbitrary communication graph.
credit.

- Experiment Setup: Participants were asked to perform a perceptual task interactively. Two arrays of circles were given to each pair of participants, then, they were asked to judge which array had the larger average diameter; that is, picking their actions. On each trial, two $4 \times 4$ grids of circles were generated by randomly drawing from the radii: $\{20,24,29,35,42\}$ (in pixels). The average diameter of each grid was computed, and if the means differed by more than $8 \%$ or less than $4 \%$, new grids were made, i.e., each trial had arrays of circles differing in the average diameter length by $4-8 \%^{27}$. One participant was chosen randomly and started the experiment by choosing an action according to his observation. Thereafter, each member saw their partner's previous response (action) and his own previous action prior to making their own judgment; this is social learning. The participants continued choosing actions until their responses stabilized for a run of at least three (two participants did not necessarily agree, but each was fixed in her responses). In this experimental study, each participant chose an action in $\mathbb{A}=\{0,1\} ; a=0$ when she judged that the left array of circles had the larger diameter and $a=1$ when her judgments was that the right array of circles had the larger diameter. In each experiment, judgments (actions) of participants are recorded along with the amount of time taken to make that judgment. Fig. 3.24 shows the judgments of two participants within a group at different trials in one experiment. In this experiment, the average diameter of the left array of circles was 32.1875 and the right array was 30.5625 (in pixels).


### 3.5.2 Experimental Results

The results of our experimental study, which are summarized in Fig.3.23, are as follows:

- Social learning Model: As mentioned above, the experiment for each pair of participants was continued until both participants' responses stabilized. A question that may arise here is: In what percentage of these experiments, an agreement is made between two participants? The answer to this question unveils that whether in our experiments "herding" occurred or not. In other words, it reveals that if participants exercised a social learning (influenced by their partners) or not. Interestingly, our experimental study shows that in $66 \%$ of total experiments (1102 among 1658), participants reached an agreement; that is herding occurred. Further, our experimental studies show that in $32 \%$ of experiments, the social learning was successful and both participants made the right judgment after few interactions. To find a proper social learning model, we focus on the experiments where both participants reached an agreement. Define the social learning (SL) success rate as

SL Success Rate: $\frac{\text { No. of experiments where both participants chose the correct side }}{\text { No. of experiments where both participants reached an agreement }}$.

[^20]
## Pick the side with the largest average size

$\bigcirc$ (size $=$ diameter $)$



Figure 3.20: Two arrays of circles were given to each pair of participants on a screen. Their task is to interactively determine which side (either left or right) had the larger average diameter. The partner's previous decision was displayed on screen prior to the stimulus.

In this experimental study, the state of nature belongs to $x \in\{0,1\}$ where $x=0$, when the left array of circles has the larger diameter and $x=1$, when the right array has the larger diameter. The initial belief for both participants is considered to be $\pi_{0}=[0.5,0.5]$. The observation state is assumed to be $z \in\{0,1\}$. We fit a social learning model to our experimental data which gives the same success rate as the experimental study. The social learning parameters (probability observations, $B_{i z}=p\left(z_{k}=z \mid x=i\right), i \in\{0,1\}, z \in\{0,1\}$ and the cost function $C(i, a), i \in\{0,1\}, a \in \mathbf{A})$, obtained by exhaustive search, are as follows:

$$
\begin{aligned}
& B_{i y}=\left[\begin{array}{ll}
0.61 & 0.39 \\
0.41 & 0.59
\end{array}\right], \\
& C(i, a)=\left[\begin{array}{ll}
0 & 2 \\
2 & 0
\end{array}\right]
\end{aligned}
$$

- Data incest: Here, we study the effect of data incest on the judgments of participants in our experimental study. Since we do not have access to the private observations of individuals (almost no one has such information!), we cannot exactly verify that whether data incest changed the judgment of an individual in each trial of the experiment. However, two scenarios which are depicted in Fig.3.22 are used to find data incest events in the experiments. In these


Figure 3.21: Actions of two participants in a group at different trials in one experiment.
two events, as can be seen in Fig.3.22, the action of the first participant at time $k$ influences the action of the second participant at time $k+1$, and thus, is double counted by the first participant at time $k+2$. As discussed above, since we do not have access to the private observations of participant, we cannot exactly say that data incest affects the action of the first participant at time $k+2$ or not. However, it is clear that in these events, data incest occurs. As we expect from the communication topology between partners, data incest occurred in relatively large percentage of trials in the experiment. More precisely, in $79 \%$ of experiments one of the data incest events shown in Fig. 3.22 occurred (1303 experiments with data incest among 1658 experiments). Our experimental study further shows that in $21 \%$ of experiments, data incest resulted in changing the decision of one of the participants in the group, i.e., the judgment of participant at time $k+1$ differed from her judgments at time $k+2$ and $k$ in the events shown in Fig.3.22. This experimental study reveals that data incest is quite common in social learning in human societies (happened frequently even in our simple social learning setup) and, therefore, social learning protocols require a careful design to handle and mitigate data incest.

- Discussion: Among 3316 (non-unique) participants of this experiment, 1336 participants (around $40 \%$ ) did not change their judgments after observing the action of their partners, while the other $60 \%$ changed their initial judgment and were influenced by the action of their partners. An experimental observation that stands out here is that the individuals can be divided into two types: (i) boundary agents who stand firm on their decisions during the trial,


Figure 3.22: Two scenarios where data incest arose in our experimental studies.
i.e., their decisions are independent of decisions of the other agents and (ii) internal agents who are affected by decisions of the other agents. Fig. 3.24 shows the sample path of two participants in a group, Participant 1 is an internal node while Participant 2 is a boundary node.

To study the decision making behavior of individuals of each type, we investigate the time taken by each participant to make his judgment. Let $\mu_{\mathrm{judg}}$. and $\sigma_{\mathrm{judg} \text {. denote the mean and the }}$ standard deviation of the time taken by participants to make their judgments in milliseconds. The results of our experimental study, which are presented in Table I, show that the internal nodes, in average, required more time to make their judgments compared to the boundary nodes; this is quite intuitive from the fact that the boundary nodes stood firm on their decisions and ignored the judgment of their partners and thus required less time to make their judgments.

| Type of nodes | relative frequency | $\mu_{\text {judg. }}$ | $\sigma_{\text {judg. }}$ |
| :---: | :---: | :--- | :--- |
| Internal | $40 \%$ | 1058 ms | 315 ms |
| Boundary | $60 \%$ | 861 ms | 403 ms |

Table 3.1: The frequency of the internal and the boundary nodes in a community of 3316 undergraduate students of the University of British Columbia along with the statistics of the time required by participants (of both types) to make their judgments in milliseconds.

### 3.6 Closing Remarks

In this chapter, the state estimation problem in social networks with social learning is investigated. State of nature could be geographical coordinates of an event (target localization problem) or quality


Figure 3.23: Social learning with data incest that is exercised by groups of students who were asked to perform a conceptual task in our experimental study.
of a social unit (online rating and review system). In online rating and review systems, privacy concerns impose a constraint on the resolution of information that users reveal to other people. People are more likely to share a lower resolution action to others rather than their detailed private observations ${ }^{28}$.

As discussed in the chapter, data incest arises in this setup as a result of the recursive nature of Bayesian estimation and random communication delays in social networks. We proposed a data incest removal algorithm for the multi-agent social learning in social networks in this chapter along with a topological necessary and sufficient condition for data incest free estimation. The main difference of this work with the data incest removal algorithms in Chapter 2 is that in this chapter we considered a social learning context where only public belief of agents (which can be computed directly from actions) is transmitted over the network while in the previous chapter, the private belief of agents which depends on their private observations are transmitted through the network.

The results of this chapter can be applied to a scenario where the network administrator provides the public beliefs to agents (instead of the updated network belief). In this scenario, agents combine the received public beliefs using the optimal weight vector $w_{n}$ to compute the updated public belief

[^21]

Figure 3.24: Actions of two participants in a group at different epochs. Participant 1 can be considered as an internal node and Participant 2 can be viewed as a boundary node.
and then evaluate their private belief accordingly. Optimal weights (which depends on the topology of the communication graph) and set of available public beliefs are the essential ingredients that one needs in order to compute the network belief. As can be seen in Figure 3.25, these ingredients can be computed by separate units (that do not communicate with each other) and the user can aggregate these two and compute the most updated network belief. This mean that only he has access to the updated network belief and, thus, his privacy is preserved ${ }^{29}$.

### 3.7 Proof of Results

Here, we present proof for propositions and results of this chapter in the order of their appearance.

### 3.7.1 Proof of Lemma 3.2.1

Proof. We assume that each node has the most up-to-date public belief of social learning, $\pi_{-n}=$ $p\left(x \mid \boldsymbol{\Theta}_{n}\right)$. This node records its own private observation $z_{n}=\bar{z}_{l}$. The private belief is

$$
\begin{equation*}
\mu_{n}(m)=p\left(x=\bar{x}_{m} \mid \Theta_{n}, z_{n}\right) . \tag{3.21}
\end{equation*}
$$

[^22]

Figure 3.25: Optimal weights (which depends on the topology of the communication graph) and set of available public belief are computed in separate units. The user, then, can compute the most updated data-incest free network belief.

Using Bayes' theorem, (3.22) can be written as

$$
\begin{align*}
\mu_{n}(m)=p\left(x=\bar{x}_{m} \mid \Theta_{n}, z_{n}\right)= & c p\left(z_{n} \mid x\right) p\left(x \mid \Theta_{n}\right) \\
& =c \pi_{-n}(m) B_{m l} . \tag{3.22}
\end{align*}
$$

The normalizing factor $c$ is used to make $\mu_{n}$ a true probability mass function, that is $\sum_{m=1}^{X} \mu_{n}(m)=1$. Expected cost given $\mu_{n}$ is equal to $C_{a}^{\prime} \mu_{n}$ thus the action $a_{n}$ is $a_{n}=\operatorname{argmin}_{a \in \mathbb{A}}\left\{C_{a}^{\prime} \mu_{n}\right\}$. To complete the proof, we need to compute the public belief, $\pi_{+n}=p\left(x \mid \Theta_{n}, a_{n}\right)$. Applying Bayes' theorem, the after action public belief can be written as

$$
\begin{align*}
\pi_{+n}(m) & =p\left(x=\bar{x}_{m} \mid \Theta_{n}, a_{n}\right)=c p\left(a_{n} \mid \Theta_{n}, x\right) p\left(x=\bar{x}_{m} \mid \Theta_{n}\right) \\
& =c p\left(a_{n} \mid x, \pi_{-n}\right) \pi_{-n}(m)=c \sum_{j=1}^{Z} p\left(a_{n} \mid x, z=\bar{z}_{j}, \pi_{-n}\right) p\left(z=\bar{z}_{j}\right) . \tag{3.23}
\end{align*}
$$

Knowing observations and public belief, the private belief can be computed. From the private belief, the action $a_{n}$ is speified. Thus

$$
p\left(a_{n} \mid x, z=\bar{z}_{j}, \pi_{-n}\right)= \begin{cases}1 & \text { if } a_{n}=\operatorname{argmin}_{a \in \mathbb{A}}\left\{C_{a}^{\prime} B_{j} \pi_{-n}\right\}  \tag{3.24}\\ 0 & \text { if } a_{n} \neq \operatorname{argmin}_{a \in \mathbb{A}}\left\{C_{a}^{\prime} B_{j} \pi_{-n}\right\}\end{cases}
$$

where $B_{j}=\operatorname{diag}\left(B_{1} j, \ldots, B_{X}\right)$. Using indicator function $\mathbb{I}(\cdot)$, Eq. (3.24) can be reorganized as

$$
\begin{equation*}
p\left(a_{n} \mid x, z=\bar{z}_{j}, \pi_{-n}\right)=\prod_{\widehat{a} \in \mathbf{A}-\left\{a_{n}\right\}} \mathbb{I}\left(C_{a_{n}}^{\prime} B_{j} \pi_{-n}<C_{\widehat{a}}^{\prime} B_{j} \pi_{-n}\right) \tag{3.25}
\end{equation*}
$$

Substituting (3.25) in (3.23) completes the proof as follows

$$
\begin{equation*}
\pi_{+n}(m)=c \pi_{-n}(m) \sum_{j=1}^{Z}\left[\prod_{\widehat{a} \in \mathbf{A}-\left\{a_{n}\right\}} \mathbb{I}\left(C_{a_{n}}^{\prime} B_{j} \pi_{-n}<C_{\widehat{a}}^{\prime} B_{j} \pi_{-n}\right)\right] B_{m j} \tag{3.26}
\end{equation*}
$$

### 3.7.2 Proof of Theorem 3.3.1

Proof. The logarithm of the public belief of learning problem (2.6) with benchmark information exchange Protocol 2 , $\theta_{n}^{\text {full }}$, is $\log \left(p\left(x \mid \Theta_{n}^{\text {full }}, G_{n}\right)\right)$. Recall that $\Theta_{n}^{\text {full }}$ denotes the entire history of actions from previous nodes who have a path to node $n$ and $S_{i}$ denotes the set of all actions that $a_{i}$ depends on them. Also from definition of the transitive closure matrix (A.3) and $t_{n}$ in (A.6), the nodes who have a path to node $n$ are corresponding to non-zero elements of $t_{n}$. Because if $t_{n}(i)=1$, then there exists a path from node $i$ to node $n$. Therefore, the public belief can be written as

$$
\begin{align*}
p\left(x \mid \Theta_{n}^{\text {full }}, a_{n}, G_{n}\right) & =c p\left(a_{n} \mid S_{n}, x\right) p\left(x \mid\left\{a_{i} ; a_{i} \in \Theta_{n}^{\text {full }}\right\}\right) \\
& =c \pi_{0} p\left(a_{n} \mid S_{n}, x\right) \prod_{a_{i} \in \Theta_{n}^{\text {fill }}} p\left(a_{i} \mid S_{i}, x\right) . \tag{3.27}
\end{align*}
$$

Note that Bayes' theorem is used recursively to expand $p\left(x \mid\left\{a_{i} ; a_{i} \in \Theta_{n}^{\text {full }}\right\}\right)$ and $S_{i}$ includes actions(from $\Theta_{n}^{\text {full }) ~ i n t o ~ a c c o u n t ~ t h a t ~} a_{i}$ depends on them. Taking the logarithm of both sides of (3.27) yields

$$
\begin{align*}
\theta_{n}^{\text {full }} & =\log \left(p\left(x \mid \Theta_{n}^{\text {full }}, a_{n}, G_{n}\right)\right) \\
& =\log \left(c \pi_{0} p\left(a_{n} \mid S_{n}, x\right) \prod_{a_{i} \in \Theta_{n}^{\text {fill }}} p\left(a_{i} \mid S_{i}, x\right)\right) \\
& =\log \left(p\left(a_{n} \mid S_{n}, x\right)\right)+\sum_{t_{n}(i) \neq 0} \log \left(p\left(a_{i} \mid S_{i}, x\right)\right), \\
& =\sum_{i=1}^{n-1} t_{n}(i) v_{i}+v_{n}, \tag{3.28}
\end{align*}
$$

where $v_{i}$ denotes $\log \left(p\left(a_{i} \mid x, S_{i}\right)\right)$. Note that the normalizing constant $c$ and $\pi_{0}$ are omitted for the sake of simplicity as they are the same for both learning problems (3.9) with the constrained social learning Protocol 1 and (2.6) with the benchmark Protocol 2.

### 3.7.3 Proof of Theorem 3.3.2

Proof. The aim here is to show that if $w_{n}=t_{n}\left(T_{n-1}^{\prime}\right)^{-1}$ then $\widehat{\theta}_{n}$ defined in (3.18) is exactly equal to $\theta_{n}^{\text {full }}$ in (3.14). Before proceeding, let us first rewrite (3.18) and (3.14) using the following notations

$$
\begin{gather*}
\theta_{n}^{\text {full }}=v_{n}+\left(t_{n} \otimes \mathbf{I}_{d}\right) v_{1: n-1}, \\
\widehat{\theta}_{n}=v_{n}+\left(w_{n} \otimes \mathbf{I}_{d}\right) \widehat{\theta}_{1: n-1}, \tag{3.29}
\end{gather*}
$$

where $\widehat{\theta}_{1: n-1} \triangleq\left[\widehat{\theta}_{1}^{\prime}, \ldots, \widehat{\theta}_{n-1}^{\prime}\right]^{\prime}, v_{1: n-1} \triangleq\left[v_{1}^{\prime}, \ldots, v_{n-1}^{\prime}\right]^{\prime} \in \mathbb{R}^{(n-1) d \times 1}$. Here $\otimes$ denotes Kronecker (tensor) product and $\mathbf{I}_{d}$ denotes the $d \times d$ identity matrix.

To prove Theorem 3.3.2, we first start from

$$
\begin{equation*}
\widehat{\theta}_{n}=\theta_{n}^{\text {full }} \tag{3.30}
\end{equation*}
$$

Assume that (3.30) holds for all $i$ where $1 \leq i \leq n$. From (3.18), $\theta_{n}^{\text {full }}$ can be written as (given that Eq. (3.30) holds)

$$
\begin{align*}
\theta_{n}^{\text {full }}=\widehat{\theta}_{n} & =\left(w_{n} \otimes \mathbf{I}_{d}\right) \widehat{\theta}_{1: n-1}+v_{n} \\
& =\left(w_{n} \otimes \mathbf{I}_{d}\right) \theta_{1: n-1}^{\text {full }}+v_{n} . \tag{3.31}
\end{align*}
$$

Eq. (3.30) holds for all $i$ where $1 \leq i \leq n$. Therefore, $\widehat{\theta}_{1: n-1}=\theta_{1: n-1}^{\text {full }}$. From (3.14) in Theorem 3.3.1, $\theta_{1: n-1}^{\text {full }}$ can be expressed as

$$
\begin{equation*}
\theta_{1: n-1}^{\text {full }}=\left(T_{n-1}^{\prime}\right) v_{1: n-1} . \tag{3.32}
\end{equation*}
$$

Note that in the derivation of (3.32), we use the definition of $t_{n-1}$ in (A.6) as the first $n-2$ elements of $T_{n-1}$ and so on. Using (3.32), (3.31) can be written as

$$
\begin{equation*}
\theta_{n}^{\text {full }}=\left(w_{n} \otimes \mathbf{I}_{d}\right)\left(T_{n-1}^{\prime}\right) v_{1: n-1}+v_{n} . \tag{3.33}
\end{equation*}
$$

From (3.14) in Theorem 3.3.1, we have another expression for $\theta_{n}^{\text {full }}$. Comparing (3.33) and (3.14) yields

$$
\begin{align*}
\left(t_{n} \otimes \mathbf{I}_{d}\right) v_{1: n-1} & =\left(w_{n} \otimes \mathbf{I}_{d}\right)\left(T_{n-1}^{\prime}\right) v_{1: n-1} \\
& =\left(\left(w_{n} T_{n-1}^{\prime}\right) \otimes \mathbf{I}_{d}\right) v_{1: n-1} . \tag{3.34}
\end{align*}
$$

Note that in going from the first line to the second line in (3.34), the distributive property of tensor products is used. From (3.34) it can be inferred that $t_{n}=w_{n} T_{n-1}^{\prime}$. As presented in Appendix A, $T_{n}$ is upper triangular matrix with ones in the diagonal. Therefore $T_{n}$ is invertible and $w_{n}=\left(t_{n} T_{n-1}^{\prime}\right)^{-1}$. To complete the proof we need to start from $w_{n}=\left(t_{n} T_{n-1}^{\prime}\right)^{-1}$ and obtain $\widehat{\theta}_{n}=\theta_{n}^{\text {full }}$. This part of proof
is straightforward and thus omitted from the chapter. Note that the topological Constraint 3.3.1 says that if $b_{n}(j)=0$ then the $j$-th entry of $v_{1: n-1}$ is not available to the node $n$ and thus the corresponding element of the weight vector $w_{n}(j)$ should be equal to zero as well. Also note that $v_{n}$ is computed by the network administrator and the data incest free public belief, $\pi_{-n}$, is available to the network administrator.

## Part II

## Tracking Degree Distribution in Dynamic Social Networks

# Tracking a Markov Modulated Degree Distribution 

### 4.1 Introduction

Dynamic random graphs have been widely used to model social networks, biological networks [42] and Internet graphs [41]. Such dynamic models can be viewed as a sequence of graphs where the random graph at each time may depend on all the earlier graphs (snapshots of the evolving graph at earlier times) [41]. Motivated by analyzing social networks, we introduce Markov-modulated duplication-deletion random graphs ${ }^{30}$ where at each time instant, nodes can either be added to or eliminated from the graph with probabilities that change according to a finite-state Markov chain. Such graphs mimic social networks where the interactions between nodes evolve over time according to a Markov process that undergoes infrequent jumps. An example of such a social network is the friendship network among residents of a city, where the dynamics of the network change in the event of a large festival.

Social networks can be viewed as complex sensors that provide information about interacting individuals and an underlying state of nature ${ }^{31}$. In this chapter, we consider a dynamic social network where at each time instant one node can join or leave the network. The probabilities of joining or leaving evolve according to the realization of a finite state Markov chain that represents the state of nature. This chapter presents two results. First, motivated by social network applications, the asymptotic behavior of the degree distribution of the Markov-modulated random graph is analyzed. Second, using noisy observations of nodes' connectivity, a "social sensor" is designed for tracking the underlying state of nature as it evolves over time.

### 4.1.1 Chapter Goals

As explained above, in this chapter, Markov-modulated dynamic random graphs are introduced to mimic social networks where the evolution of the network is varying over time. The most important parameter of a network that characterizes its structure is the degree distribution. It yields useful

[^23]information about the connectivity of the random graph [10, 86, 116]. For example, if a majority of nodes in the random graph have relatively high degrees, the graph is highly connected and a message can be transferred between two arbitrary nodes with shorter paths. However, if a majority of nodes have smaller degrees then for transmitting a message throughout the network, longer paths are needed, see [80]. The degree distribution can further be used to investigate the diffusion of information or disease through social networks [108, 146]. The existence of a "giant component"32 in complex networks can be studied using the degree distribution. The size and existence of a giant component has important implications in social networks in terms of modeling information propagation and spread of human disease $[62,115,118]$. The degree distribution is also used to analyze the "searchability" of a network. The "search" problem arises when a specific node in a network faces a problem (request) whose solution is at other node, namely, destination (e.g., delivering a letter to a specific person, or finding a web page with specific information) [4, 146]. The searchability of a social network [146] is the average number of nodes that need to be accessed to reach the destination. Degree distribution is also used to investigate the robustness and vulnerability of a network in terms of the network response to attacks on its nodes or links [33, 76]. The papers [148, 149] further use degree-dependent tools for classification of social networks.

The fist goal of this chapter is to provide a degree distribution analysis that allows us to determine the relation between the structure of the network (in terms of connectivity) and the underlying state of nature. Indeed, it will be shown in Section 4.3 that there exists a unique stationary degree distribution for the Markov-modulated graph for each state of the underlying Markov chain. It thus suffices to estimate the degree distribution in order to track the underlying state of nature. The second goal of the chapter is to propose a stochastic approximation algorithm to track the empirical degree distribution of the Markov-modulated random graph. In particular, our goals are to address the following two questions in Section 4.4:

- How can a social sensor estimate (track) the empirical degree distribution using a stochastic approximation algorithm with no knowledge of the Markovian dynamics?
- How accurate are the estimates generated by the stochastic approximation algorithm when the random graph evolves according to the duplication-deletion model with Markovian switching?

By tracking the degree distribution of a Markov-modulated random graph, we can design a social sensor to track the underlying state of nature using the noisy measurements of nodes' connectivity.

[^24]
### 4.1.2 Main Results and Organization of Chapter

Section 4.2 describes the construction of Markov-modulated duplication-deletion random graphs. Section 4.3 provides an asymptotic degree distribution analysis for the non-Markov modulated case of two different scenarios: (i) fixed size duplication-deletion random graph, and (ii) infinite duplication-deletion random graph. Theorem 4.3.1 in Section 4.3.1 asserts that the expected degree distribution of the fixed size Markov-modulated random graph at each time can be computed in terms of the expected degree distribution of the graph at the previous time and the dynamics of the graph via recursive equation (4.9). Section 4.3.2 extends the results of Section 4.3.1 to infinite random graphs. Theorem 4.3.2 parameterizes the degree distribution of such a graph by the power law exponent which depends on the dynamics of the graph.

Section 4.4 considers the problem of adaptively estimating the degree distribution of a fixed size Markov-modulated duplication-deletion random graph given observations of the degree distribution. A stochastic approximation algorithm is presented for tracking the degree distribution as it evolves over time. In particular, Section 4.4 presents three results regarding the tracking performance of the stochastic approximation algorithm:

- Mean square error analysis: Theorem 4.4.1 analyzes the asymptotic mean square error between the expected degree distribution and the estimate obtained via the stochastic approximation algorithm. Deriving this result uses error bounds on two-time scale Markov chains and perturbed Lyapunov function methods.
- Weak convergence analysis: Theorem 4.4.2 shows that the asymptotic behavior of the stochastic approximation algorithm converges weakly to the solution of a switched Markovian ordinary differential equation.
- Functional central limit theorem for scaled tracking error: Finally, Theorem 4.4.3 investigates the asymptotic behavior of the scaled tracking error. Similar to [94], it is shown that the interpolated scaled tracking error converges weakly to the solution of a switching diffusion process.

Section 4.5 extends the results of Section 4.4 to infinite (denumerable) duplication-deletion random graphs where the number of nodes in the graph (and so the support of degree distribution) is no longer fixed and increases over time. A Hilbert-space-valued stochastic approximation algorithm is proposed to track the degree distribution of the infinite graph with support on the set of nonnegative integers. To study the tracking performance of such a Hilbert-space-valued stochastic approximation algorithm, limit system characterization and asymptotic analysis of scaled tracking error are provided. Numerical examples are presented in Section 4.6.

### 4.2 Markov-modulated Dynamic Random Graph of Duplication-deletion Type

This section outlines the construction of Markov-modulated dynamic random graphs of duplicationdeletion type. Let $n=0,1,2, \ldots$ denote discrete time. Denote by $\theta_{n}$ a discrete-time Markov chain with state space

$$
\begin{equation*}
\mathscr{M}=\{1,2, \ldots, M\} \tag{4.1}
\end{equation*}
$$

and initial probability distribution $\pi_{0}$.
Assumption 4.2.1. The Markov chain $\theta_{n}$ evolves according to the transition matrix

$$
\begin{equation*}
A^{\rho}=I+\rho Q \tag{4.2}
\end{equation*}
$$

Here, $I$ is an $M \times M$ identity matrix, $\rho$ is a small positive real number, and $Q=\left[q_{i j}\right]$ is an irreducible ${ }^{33}$ generator of a continues-time Markov chain satisfying

$$
\begin{equation*}
q_{i j}>0 \text { for } i \neq j, \text { and } Q \mathbf{1}=\mathbf{0} \tag{4.3}
\end{equation*}
$$

where $\mathbf{1}$ and $\mathbf{0}$ represent column vectors of ones and zeros, respectively. The transition probability matrix $A^{\rho}$ is therefore close to identity matrix. Here and henceforth, we refer to such a Markov chain $\theta_{n}$ as a "slow" Markov chain. The initial distribution $\pi_{0}$ is assumed independent of $\rho$.

A Markov-modulated duplication-deletion dynamic random graph is parameterized by the 7tuple $\left(M, A^{\rho}, \pi_{0}, r, p, q, G_{0}\right)$. Here, $p$ and $q$ are $M$-dimensional vectors with elements $p(i)$ and $q(i) \in$ $[0,1], i=1, \ldots, M$, where $p(i)$ denotes the connection probability, and $q(i)$ denotes the deletion probability. Also, $r \in[0,1]$ denotes the probability of the duplication step, and $G_{0}$ denotes the initial graph at time 0 . In general, $G_{0}$ can be any finite simple connected graph. For simplicity, assume that $G_{0}$ is a simple connected graph with size $N_{0}$. The duplication-deletion random graph is constructed via the duplication-deletion Procedure $4.5^{34}$.

The Markov-modulated random graph generated by the duplication-deletion Procedure $4.5 \mathrm{mim}-$ ics social networks where the interactions between nodes evolve over time due to the underlying dynamics (state of nature) such as seasonal variations (e.g., the high school friendship social network evolving over time with different winter/summer dynamics). In such cases, the connection/deletion

[^25]Procedure 4.5 Markov-modulated Graph parameterized by $\left(M, A^{\rho}, \pi_{0}, r, p, q, G_{0}\right)$
At time $n$, given the graph $G_{n}$ and Markov chain state $\theta_{n}$, simulate the following events:
Step 1: Duplication step: With probability $r$ implement the following steps:

- Choose node $u$ from graph $G_{n}$ randomly with uniform distribution.
- Vertex-duplication: Generate a new node $v$.
- Edge-duplication:
- Connect node $u$ to node $v$. (A new edge between $u$ and $v$ is added to the graph.)
- Connect each neighbor of node $u$ with probability $p\left(\theta_{n}\right)$ to node $v$. These connection events are statistically independent.

Step 2: Deletion Step: With probability $q\left(\theta_{n}\right)$ implement the following steps:

- Edge-deletion: Choose node $w$ randomly from $G_{n}$ with uniform distribution. Delete node $w$ along with the connected edges in graph $G_{n}$.
- Duplication Step: Choose a node from graph $x$ from $G_{n}$ randomly and implement Vertexduplication and Edge-duplication processes as described in Step 1.

Step 3: Denote the resulting graph by $G_{n+1}$. Generate $\theta_{n+1}$ (Markov chain) using transition matrix $A^{\rho}$. Set $n \rightarrow n+1$ and go to Step 1 .
probabilities $p, q$ evolve with time. Procedure 4.5 models these time variations as a finite state Markov chain $\theta_{n}$ with transition matrix $A^{\rho}$.

The Markov-modulated random graph generated by Procedure 4.5 mimics social networks where the interactions between nodes evolve over time due to the underlying dynamics (state of nature) such as seasonal variations (e.g., the high school friendship social network evolving over time with different winter/summer dynamics). In such cases, the connection/deletion probabilities $p, q$ depend on the state of nature and evolve with time. Procedure 4.5 models these time variations as a finite state Markov chain $\theta_{n}$ with transition matrix $A^{\rho}$.

## Discussion:

The connection/deletion probabilities $p, q$ can be determined by the solution of a utility maximization problem. Let $U^{\text {join }}:[0,1] \times \mathscr{M} \rightarrow \mathbb{R}$ denote a utility function that gives payoff to an individual who considers to expand his neighbors in "Edge-duplication step" of Procedure 4.5 as a function of $(p, \theta)$. Similarly, let $U^{\text {leave }}:[0,1] \times \mathscr{M} \rightarrow \mathbb{R}$ denote a utility function that pays off to an individual who considers to leave the network in "Deletion step" of Procedure 4.5 as a function of $(q, \theta)$. With the above utility functions, the probabilities of connection/deletion when the state of nature is $\theta$ can
be viewed as the solutions of the following maximization problems:

$$
\begin{align*}
& p(\theta)=\underset{p}{\operatorname{argmax}}\left\{U^{\text {join }}(p, \theta)\right\},  \tag{4.4}\\
& q(\theta)=\underset{q}{\operatorname{argmax}}\left\{U^{\text {leave }}(q, \theta)\right\} .
\end{align*}
$$

These utility functions can be interpreted in terms of mutual benefits and privacy concerns. One example could be $U^{\text {join }}(p, \theta)=b^{\text {join }}(p, \theta)-v$, where $b^{\text {join }}(p, \theta)$ is the benefit one obtains by expanding his network with probability $p$ when the underlying state of nature is $\theta$, and $v$ is the cost incurred by sacrificing his "privacy". In this example, when an individual decides to leave the network, the utility he obtains will be $U^{\text {leave }}(q, \theta)=b^{\text {leave }}(q, v)-c(\theta)$, where $b^{\text {leave }}(q, v)$ is the benefit he earns by preserving privacy and $c(\theta)$ is the benefit he loses by leaving the network when the underlying state of nature is $\theta$.

### 4.3 Asymptotic Degree Distribution Analysis for Non-Markov Modulated case

This section presents degree distribution analysis for duplication-deletion random graphs generated according Procedure 4.5 for the non-Markov modulated case, i.e., $M=1$. The stationary degree distribution obtained in Section 4.3.1 below will be used in the Markov modulated case. The results in this section constitute a minor extension of [41] to the duplication-deletion random graphs.

Notation At each time $n$, let $N_{n}$ denote the number of nodes of graph $G_{n}$. Also, let $f_{n}$ be a $N_{n}$ dimensional vector such that its $i$-th element, $f_{n}^{i}$, denotes the number of vertices of graph $G_{n}$ with degree $i$. Clearly $f_{n}^{\prime} \mathbf{1}=N_{n}$ where $\mathbf{1}$ denotes the vector of ones. Here, ${ }^{\prime}$ is used to denote the transpose of a vector or matrix. Define the "empirical vertex degree distribution" as

$$
\begin{equation*}
g_{n}=\left(g_{n}^{i}, i=1,2, \ldots\right), \text { where } g_{n}^{i}=\frac{f_{n}^{i}}{N_{n}} . \tag{4.5}
\end{equation*}
$$

Note that $g_{n}$ can be viewed as a probability mass function since all of its elements are non-negative and $g_{n}^{\prime} \mathbf{1}=1$.

### 4.3.1 Fixed Size Random Graph

This subsection analyzes the evolution of the expected degree distribution for a fixed size duplicationdeletion random graph generated according to Procedure 4.5 with $r=0, M=1$. (Recall $r$ denotes the probability of Step 1 in Procedure 4.5.) Therefore, the number of vertices in the graph remains
fixed, i.e., $N_{n}=N_{0}$ for $n=0,1,2, \ldots$. Theorem 4.3.1 below gives a recursion for the expected degree distribution of the fixed size Markov-modulated duplication-deletion random graph.

Theorem 4.3.1. Consider the fixed size duplication-deletion random graph generated according to Procedure 4.5, where $r=0, M=1$. Let $\bar{g}_{n}$ denote the expected degree distribution of nodes at time $n$. Then, $\bar{g}_{n}$ satisfies the recursion

$$
\begin{equation*}
\bar{g}_{n+1}=\left(I+\frac{1}{N_{0}} L^{\prime}\right) \bar{g}_{n}, \tag{4.6}
\end{equation*}
$$

where $L$ is a generator matrix ${ }^{35}$ with elements (for $1 \leq i, j \leq N_{0}$ ):

$$
l_{j i}= \begin{cases}0, & j<i-1,  \tag{4.7}\\ q p^{i-1}+q(1+p(i-1)), & j=i-1, \\ i q p^{i-1}(1-p)-q(i+2+p i), & j=i, \\ q\binom{i+1}{i-1} p^{i-1}(1-p)^{2}+q(i+1), & j=i+1, \\ q\binom{j}{i-1} p^{i-1}(1-p)^{j-i+1}, & j>i+1 .\end{cases}
$$

Proof. The proof is presented in Appendix 4.8.1.
Theorem 4.3.1 shows that evolution of the expected degree distribution in a fixed size Markovmodulated duplication-deletion random graph satisfies (4.6). One can rewrite (4.6) as

$$
\begin{equation*}
\bar{g}_{n+1}=B_{N_{0}}^{\prime} \bar{g}_{n}, \text { where } \quad B_{N_{0}}=I+\frac{1}{N_{0}} L . \tag{4.8}
\end{equation*}
$$

Since $L$ is a generator matrix, $B_{N_{0}}$ can be considered as the transition matrix of a slow Markov chain. It is also straightforward to show that $B_{N_{0}}$ is irreducible and aperiodic ${ }^{36}$. Hence, there exists a unique stationary distribution $\bar{g}=\left(\bar{g}^{i}, i=1,2, \ldots\right)$ such that

$$
\begin{equation*}
\bar{g}=B_{N_{0}}^{\prime} \bar{g} . \tag{4.9}
\end{equation*}
$$

The stationary distribution $\bar{g}$ is the stationary expected degree distribution of a fixed size duplicationdeletion random graph generated according to Procedure 4.5 where $r=0$. Note that the underlying Markov chain $\left\{\theta_{n}\right\}$ depends on the small parameter $\rho$. The main idea is that, although $\theta_{n}$ is timevarying but it is piecewise constant (since $\rho$ is small parameter)-it changes slowly over time. Further, in light of (4.6), the evolution of $\bar{g}_{n}$ depends on $\frac{1}{N_{0}}$. Our assumption throughout this chapter is that $\rho \ll \frac{1}{N_{0}}$. Therefore, the evolution of $\bar{g}_{n}$ is faster than the evolution of $\theta_{n}$. That is, $\bar{g}_{n}$ reaches its stationary distribution $\bar{g}$ before the state of $\theta_{n}$ changes. From (4.9), the expected degree distribution

[^26]of the fixed size Markov-modulated duplication-deletion random graph can be uniquely computed for each state of the underlying Markov chain $\theta_{n}=\theta$.

## Example: Searchability of a Network

So far in this section, an asymptotic analysis of the degree distribution was presented for a random graph generated according to Procedure 4.5. We now comment briefly on how the degree distribution can be used to investigate the searchability of the network. This also motivates the stochastic approximation algorithm presented in Section 4.4 as will be described below. The search problem arises in a network when a specific node faces a problem (request) whose solution is at other node (e.g., delivering a letter to a specific person or finding a web page with specific information). Assume [146] that on receiving a search request, each node follows the following protocol: (a) It address the request if it or its neighbors have the solution; otherwise (b) it relays the request to one of its neighbors chosen uniformly. The objective is to find the expected search delay, that is, the expected number of steps until the request is addressed.

Lemma 4.3.1. Consider the sequence of fixed size Markov-modulated duplication-deletion random graph obtained by Procedure 4.5, $\left\{G_{n}\right\}$, with $\left(M, A^{\rho}, \pi_{0}, I, q, p, G_{0}\right)$ where $A^{\rho}=I+\rho Q$ and $p=0$ and expected degree distribution $\bar{g}_{n}$. The expected search delay is

$$
\begin{equation*}
\lambda\left(N_{0}\right)=O\left(\frac{N_{0} \delta}{\bar{d}_{2}-\delta}\right) \tag{4.10}
\end{equation*}
$$

as $n \rightarrow \infty$ where $\delta=\sum_{i=1}^{N_{0}} i \bar{g}_{n}(i)$ and $\bar{d}_{2}=\sum_{i=1}^{N_{0}} i^{2} \bar{g}_{n}(i)$.
Proof. See Chapter 5 of [146] and recall that size of the considered random graph is $N_{0}$.
Lemma 4.3.1 implies that, if the empirical degree distribution of the possibly time-varying network can tracked accurately, then such an estimate can be used to track the searchability of the network. Also, using the estimated degree distribution and Lemma 4.3.1, we can address the following design problem as: How can $p$ and $q$ in Algorithm 4.5 be chosen so that the average delay does not exceed a threshold?
Using the stochastic approximation algorithm in (4.14) (see Section 4.4 below for the convergence proof), we can estimate the expected degree distribution, $\widehat{g}_{n}$, and from that, we can compute $\delta$ and $\bar{d}_{2}$. Then, from Lemma 4.3 .1 we can find the measure of searchability and compare it with the maximum acceptable average delay and modify the parameters of Procedure 4.5 accordingly. We illustrate searchability in numerical examples given in Section 4.6.

### 4.3.2 Power Law Exponent for Infinite Duplication-deletion Random Graph

The degree distribution analysis provided in the previous subsection was for a fixed size random graph generated according to the duplication-deletion Procedure 4.5 with $r=0$. This section extends this analysis to infinite duplication-deletion random graphs (obtained by choosing $r=1$ ). Assume that $G_{0}$ is an empty set. Since $r=1$, at time $n$, the graph $G_{n}$ has $n$ nodes. By employing the same approach as in the proof of Theorem 4.3.1, it will be shown that the infinite duplication-deletion random graph without Markovian dynamics satisfies a power law. An expression is further derived for the power law exponent.

Definition 4.3.1 (Power Law Distribution). The degree distribution $g=\left(g^{i}, i=1,2, \ldots\right)$, of a graph $\mathscr{G}$ has a power law distribution ${ }^{37}$ if there exists an integer $i^{*}$ such that for all $i \geq i^{*}$,

$$
\log g^{i}=\alpha-\beta \log i
$$

where $\alpha$ is a constant ${ }^{38}$ and $\beta>1$. Parameter $\beta$ is called the power law exponent.
The power law is satisfied in many networks such as WWW-graphs, peer-to-peer networks, phone call graphs, co-authorship graph and various massive online social networks (e.g. Yahoo, MSN, Facebook) [17, 27, 44, 135, 139]. The following theorem states that the graph generated according to Procedure 4.5 with $r=1$ and $M=1$ satisfies a power law.

Theorem 4.3.2. Consider the infinite random graph with Markovian dynamics $G_{n}$ obtained by Procedure 4.5 with 7 -tuple $\left(1,1,1,1, p, q, G_{0}\right)$ with the expected degree distribution $\bar{g}_{n}$. Then, if $\log p+p<\frac{q}{1+q}<p$, the expected degree of nodes in $G_{n}$ has a power law distribution with exponent $\beta>1$. The power law exponent is computed from

$$
\begin{equation*}
(1+q)\left(p^{\beta-1}+p \beta-p\right)=1+\beta q . \tag{4.11}
\end{equation*}
$$

Here, $p$ and $q$ are the probabilities defined in duplication and deletion steps, respectively.
Proof. The proof is similar to that of Theorem 4.3.1 with some modifications, see [72]. Here, we only present an outline of the proof which is comprised of two steps: (i) finding the power law exponent, and (ii) showing that the degree distribution converges to a power law with the computed exponent as $n \rightarrow \infty$. To find the power law exponent, we derive a recursive equation for the number of nodes with degree $i+1$ at time $n+1$, denoted by $f_{n+1}^{i+1}$, in terms of the degrees of nodes in

[^27]graph $G_{n}$. Then, rearranging this recursive equation yields an equation for the power law exponent. To prove that the degree distribution satisfies a power law, we show that $\lim _{n \rightarrow \infty} \sum_{k=1}^{i} \mathbf{E}\left\{f_{n}^{k}\right\}=$ $\sum_{k=1}^{i} c k^{-\beta}$, where $\beta>1$ is the power law exponent computed in the first step and $f_{n}^{k}$ is the $k$-th element of $f_{n}$.

Theorem 4.3.2 asserts that the infinite duplication-deletion random graph without Markovian dynamics generated by Procedure 4.5 satisfies a power law and provides an expression for the power law exponent. The significance of this theorem is that it ensures, with use of one single parameter (the power law exponent), we can describe the degree distribution of graphs with relatively large number of nodes. The above result slightly extends [42, 121], where only a duplication model was considered. Theorem 4.3.2 allows us to explore characteristics (such as searchability, diffusion, and existence/size of the giant component) of large networks which can be modeled with the infinite duplication-deletion random graphs. Remark 1. Outline of Proof: The proof of Theorem 4.3.2, which is presented in Appendix 4.8.2, consists of two steps: (i) finding the power law component and (ii) showing that the degree distribution converges to a power law as $n \rightarrow \infty$. To find the power law component, we derive a recursive equation for the number of nodes with degree $i+1$ at time $n+$ 1 , $f_{n+1}(i+1)$, in terms of degree of nodes in graph $G_{n}$. Then, this recursive equation is rearranged to equation for the power law component. To prove that the degree distribution satisfies a power law, we define a new parameter $h_{n}(i)=\frac{1}{n} \sum_{k=1}^{i} \mathbf{E}\left\{f_{n}(k)\right\}$ and we show that $\lim _{n \rightarrow \infty} h_{n}(i)=\sum_{k=1}^{i} C k^{-\beta}$ where $\beta$ is the power law component computed by the solving the recursive equation. Theorem 4.3.2 asserts that the infinite duplication-deletion random graph without Markovian dynamics generated by Procedure 4.5 satisfies a power law and provides an expression for the power law component. The significance of this theorem is that it ensures that with use of one single parameter (the power law component), we can describe the degree distribution of large numbers of nodes in graphs that model social networks.
Remark 2. Power Law Component: Let $\beta^{*}$ denote the solution of (4.11). Then the power law component is defined as $\beta=\max \left\{1, \beta^{*}\right\}$. Fig.4.1 shows the the power law component and $\beta^{*}$ versus $p$ for different values of probability of deletion, $q$. As can be seen in Fig.4.1, the power law component is increasing in $q$ and decreasing in $p$.

### 4.4 Estimating (Tracking) the Degree Distribution of the Fixed Size Markov-modulated Duplication-deletion Random Graph

In Section 4.2, an expression was given for the unique stationary degree distribution $\bar{g}$ for the nonMarkov modulated case, see (4.9). In this section, we consider fixed size Markov modulated duplication deletion random graphs. Consider Procedure 4.5 and assume that there are $M$ possible stationary degree distributions, namely $\mathbf{g}=\{\bar{g}(1), \bar{g}(2), \ldots, \bar{g}(M)\}$ corresponding to the $M$ states


Figure 4.1: The power law component for the non-Markovian random graph generated according to Procedure 4.5 obtained by (4.11) for different values of $p$ and $q$ in Procedure 4.5.
of a Markov chain. Here each $\bar{g}(i)$ is computed using (4.9) where the corresponding parameters $p(i), q(i)$ are used. At each time $n$, a stationary distribution $\bar{g}\left(\theta_{n}\right) \in \mathbf{g}$ is chosen where $\theta_{n}$ evolves according to an $M$-state Markov chain as described in Section 4.2. We assume that the stationary degree distribution of the graph is sampled by a network administrator. How can the network administrator track the expected degree distribution of the fixed size Markov-modulated duplication deletion random graph without knowing the dynamics of the graph? The motivation for tracking the stationary degree distribution stems from social networks where the dynamics of the degree distribution evolve on a faster time scale than the Markov chain $\theta_{n}$. Therefore, it suffices to track $\bar{g}\left(\theta_{n}\right)$ given observations.

At each time $n$, the network administrator samples a node from the graph based on degree distribution $\bar{g}\left(\theta_{n}\right)$ and records its degree $y_{n}\left(\theta_{n}\right)$. Let $z_{n}\left(\theta_{n}\right)=\mathbf{e}_{y_{n}\left(\theta_{n}\right)}$ denote the observation vector where $\mathbf{e}_{i} \in \mathbb{R}^{N_{0} \times 1}$ is the $i$-th standard unit vector. Such a sampling procedure can be time correlated. Therefore, we allow $z_{n}\left(\theta_{n}\right)$ to be a mixing process with thee following assumption:

Assumption 4.4.1. For each $\theta \in \mathscr{M}$, the sequence $\left\{y_{n}(\theta)\right\}$ is stationary $\phi$-mixing with sufficiently fast mixing rate such that the sequence $\left\{y_{n}(1), \ldots, y_{n}(M)\right\}$ is independent of $\left\{\theta_{n}\right\}$ and that for each $\theta \in \mathscr{M},\left\{z_{n}(\theta)\right\}$ is a stationary $\phi$-mixing sequence with mixing rate $\psi_{n}$ satisfying $\sum_{j=0}^{\infty} \psi_{j}^{1 / 2}<\infty$.

Remark 4.4.1. Because $\left\{y_{n}(\theta)\right\}$ is a stationary $\phi$-mixing sequence for each $\theta \in \mathscr{M},\left\{z_{n}(\theta)\right\}$ is
a bounded sequence of $\phi$-mixing process for each $\theta \in \mathscr{M}$ [97, p. 82] (see also [25, p.170]). The stationarity implies that

$$
\begin{equation*}
\mathbf{E} z_{n}(\theta)=\mathbf{E} z_{1}(\theta)=\sum_{i=0}^{\infty} \mathbf{e}_{i} P\left(y_{1}(\theta)=i\right)=\sum_{i=0}^{\infty} \bar{g}^{i}(\theta) \mathbf{e}_{i}=\bar{g}(\theta) \tag{4.12}
\end{equation*}
$$

The mixing rate given requires that for any positive integers $i$ and $j$,

$$
\begin{align*}
& \left\|\mathbf{E}_{k} I\left\{y_{n}(\theta)=i\right\}-\bar{g}^{i}(\theta)\right\| \leq \psi_{n-k} \text { for } n \geq k \\
& \left\|\mathbf{E}\left[I\left\{y_{l}(\theta)=j\right\}-\bar{g}^{j}(\theta)\right]\left[I_{\left\{y_{n}=i\right\}}-\bar{g}^{i}(\theta)\right]\right\| \leq \psi_{n-l}^{1 / 2} \psi_{l-k}^{1 / 2} \text { for any } k<l<n \tag{4.13}
\end{align*}
$$

where $\mathbf{E}_{k}$ denotes the conditional expectation on the past data up to time $k$ (i.e., condition on the $\sigma$-algebra $\mathscr{F}_{k}$ generated by $\left\{z_{j}(\theta): j \leq k\right\}$ ) and $I\{\cdot\}$ denotes the indicator function. Here, $\|\cdot\|$ is used to denote the Euclidean norm.

The analysis in this chapter can be generalized to include certain non-stationary cases for the observation process $\left\{y_{n}(\theta)\right\}$. For example, for each $\theta \in \mathscr{M}$, suppose $\left\{\zeta_{n}(\theta)\right\}$ is an ergodic finite state Markov chain ${ }^{39}$. Let $y_{n}(\theta)=\widetilde{f}\left(\zeta_{n}(\theta)\right)$. The $n$-step transition probability matrix of the Markov chain converges to a matrix (with identical rows consisting of its stationary distribution) with exponential rate. Then it can be verified similar to [25, pp.178] that $y_{n}(\theta)$ is mixing. Although (4.12) does not hold, the analysis using mixing inequalities can still be obtained.

Given the observation sequence $z_{n}\left(\theta_{n}\right), n=0,1,2, \ldots$, the aim is to adaptively estimate $\bar{g}\left(\theta_{n}\right)$ via the following stochastic approximation algorithm with (small positive) constant step-size $\varepsilon$ :

$$
\begin{equation*}
\widehat{g}_{n+1}=\widehat{g}_{n}+\varepsilon\left(z_{n}\left(\theta_{n}\right)-\widehat{g}_{n}\right), \quad \widehat{g}_{0}=\mathbf{e}_{1} \tag{4.14}
\end{equation*}
$$

To summarize, the evolution of the slow Markov chain $\theta_{n}$ and stochastic approximation algorithm (4.14) form a two-time-scale Markovian system as follows when $\rho, \varepsilon=o\left(\frac{1}{N_{0}}\right)$

$$
\left\{\begin{array}{l}
\text { True system: } \bar{g}\left(\theta_{n}\right) \in\{\bar{g}(1), \ldots, \bar{g}(M)\}, \text { where } \theta_{n} \text { evolves according to } A^{\rho}=I+\rho Q  \tag{4.15}\\
\text { Algorithm: } \widehat{g}_{n+1}=\widehat{g}_{n}+\varepsilon\left(z_{n}\left(\theta_{n}\right)-\widehat{g}_{n}\right), \quad z_{n}\left(\theta_{n}\right)=\mathbf{e}_{y_{n}\left(\theta_{n}\right)}, \text { where } y_{n}\left(\theta_{n}\right) \sim \bar{g}\left(\theta_{n}\right)
\end{array}\right.
$$

Note that the stochastic approximation algorithm (4.14) does not assume any knowledge of the Markov-modulated dynamics of the graph. The Markov chain assumption for the random graph dynamics is only used in our convergence and tracking analysis. By means of the stochastic approximation (4.14), the network administrator can track the stationary expected degree distribution $\bar{g}\left(\theta_{n}\right)$.

[^28]
### 4.4.1 Tracking Error of the Stochastic Approximation Algorithm

The goal here is to analyze how well algorithm (4.14) tracks the empirical degree distribution of the fixed size Markov-modulated duplication-deletion graph. Define the tracking error as $\widetilde{g}_{n}=\widehat{g}_{n}-$ $\bar{g}\left(\theta_{n}\right)$. Theorem 4.4.1 below shows that the difference between the sample path and the stationary degree distribution is small-implying that the stochastic approximation algorithm can successfully track the Markov-modulated node distribution given the noisy measurements. We again emphasize that no knowledge of the Markov chain parameters are required in the algorithm. It also finds the order of this difference in terms of $\varepsilon$ and $\rho$.

Theorem 4.4.1. Consider the random graph $\left(M, A^{\rho}, \pi_{0}, p, q, r, G_{0}\right)$. Suppose $\rho^{2} \ll \varepsilon$ and Assumptions 4.2.1 and 4.4.1 hold ${ }^{40}$. Then, for sufficiently large $n$, the tracking error of the stochastic approximation algorithm (4.14) is

$$
\begin{equation*}
\mathbf{E}\left\|\widetilde{g}_{n}\right\|^{2}=O\left(\varepsilon+\rho+\frac{\rho^{2}}{\varepsilon}\right) . \tag{4.16}
\end{equation*}
$$

Proof. The proof uses the perturbed Lyapunov function method and is provided in Appendix 4.8.5.

Remark 4.4.2. Most existing literature analyzes stochastic approximation algorithms for tracking a parameter that evolves according to a "slowly time-varying" sample path of a continuous-valued process so that the parameter changes by small amounts over small intervals of time. When the rate of change of the underlying parameter is slower than the adaptation rate of the stochastic approximation algorithm (e.g., a slow random walk), the mean square tracking error can be analyzed as in [19, 69, 99, 111, 127, 137]. In comparison, our analysis covers the case where the underlying parameter evolves with discrete jumps that can be arbitrarily large in magnitude on short intervals of time. Also, the jumps occur on the same time scale as the speed of adaptation of the stochastic approximation algorithm. We explicitly consider this Markovian time-varying parameter in our mean square error and weak convergence analysis.

As a corollary of Theorem 4.4.1, we obtain the following mean square error convergence result.
Corollary 4.4.1. Under the conditions of Theorem 4.4.1, if $\rho=O(\varepsilon)$,

$$
\lim _{n \rightarrow \infty} \mathbf{E}\left\|\widetilde{g}_{n}\right\|^{2}=O(\varepsilon)
$$

Therefore,

$$
\lim _{\varepsilon \rightarrow 0} \lim _{n \rightarrow \infty} \mathbf{E}\left\|\widetilde{g}_{n}\right\|^{2}=0
$$

[^29]
### 4.4.2 Limit System of Regime-Switching Ordinary Differential Equations

The following theorem asserts that the sequence of estimates generated by the stochastic approximation algorithm (4.14) follows the dynamics of a Markov-modulated ordinary differential equation (ODE).

Before proceeding with the main theorem below, let us recall a Definition.
Definition 4.4.1 (Weak Convergence). Let $Z_{k}$ and $Z$ be $\mathbb{R}^{r}$-valued random vectors. We say $Z_{k}$ converges weakly to $Z\left(Z_{k} \Rightarrow Z\right)$ if for any bounded and continuous function $f(\cdot), E f\left(Z_{k}\right) \rightarrow E f(Z)$ as $k \rightarrow \infty$.

Weak convergence is a generalization of convergence in distribution to a function space ${ }^{41}$.
Theorem 4.4.2. Consider the Markov-modulated random graph generated according to Procedure 4.5, and the sequence of estimates $\left\{\widehat{g}_{n}\right\}$, generated by the stochastic approximation algorithm (4.14). Suppose Assumptions 4.2.1 and 4.4.1 hold and $\rho=O(\varepsilon)$. Define the continuous-time interpolated process

$$
\begin{equation*}
\widehat{g}^{\varepsilon}(t)=\widehat{g}_{n}, \theta^{\varepsilon}(t)=\theta_{n} \text { for } t \in[n \varepsilon,(n+1) \varepsilon) \text {. } \tag{4.17}
\end{equation*}
$$

Then, as $\varepsilon \rightarrow 0$, $\left(\widehat{g}^{\varepsilon}(\cdot), \theta^{\varepsilon}(\cdot)\right)$ converges weakly to $(\widehat{g}(\cdot), \theta(\cdot))$, where $\theta(\cdot)$ is a continuous-time Markov chain with generator $Q, \widehat{g}(\cdot)$ satisfies the Markov-modulated ODE

$$
\begin{equation*}
\frac{d \widehat{g}(t)}{d t}=-\widehat{g}(t)+\bar{g}(\theta(t)), \quad \widehat{g}(0)=\widehat{g}_{0} \tag{4.18}
\end{equation*}
$$

and $\bar{g}(\theta) \in \mathbf{g}$.
The above theorem asserts that the limit system associated with the stochastic approximation algorithm (4.14) is a Markovian switched ODE (4.18). As mentioned in Section 4.1, this is unusual since typically in averaging of stochastic approximation algorithms, convergence occurs to a deterministic ODE. The intuition behind this is that the Markov chain evolves on the same time-scale as the stochastic approximation algorithm. If the Markov chain evolved on a faster time-scale, then the limiting dynamics would be a deterministic ODE weighed by the stationary distribution of the Markov chain. If the Markov chain evolved slower than the dynamics of the stochastic approximation algorithm, then the asymptotic behavior would also be a deterministic ODE with the Markov chain being a constant.

### 4.4.3 Scaled Tracking Error

Next, we study the behavior of the scaled tracking error between the estimates generated by the stochastic approximation algorithm (4.14) and the expected degree distribution. The following theorem states that the tracking error should also satisfy a switching diffusion equation and provides a

[^30]functional central limit theorem for this scaled tracking error. Let $v_{k}=\frac{\widehat{\hat{g}_{k}}-\overline{\bar{z}}\left(\theta_{k}\right)}{\sqrt{\varepsilon}}$ denote the scaled tracking error.

Theorem 4.4.3. Suppose Assumptions 4.2.1 and 4.4.1 hold. Define $v^{\varepsilon}(t)=v_{k}$ for $t \in[k \varepsilon,(k+1) \varepsilon)$. Then, $\left(v^{\varepsilon}(\cdot), \theta^{\varepsilon}(\cdot)\right)$ converges weakly to $(v(\cdot), \theta(\cdot))$ such that $v(\cdot)$ is the solution of the following Markovian switched diffusion process

$$
\begin{equation*}
v(t)=-\int_{0}^{t} v(s) d s+\int_{0}^{t} \Sigma^{\frac{1}{2}}(\theta(\tau)) d \omega(\tau) . \tag{4.19}
\end{equation*}
$$

Here, $\omega(\cdot)$ is an $\mathbb{R}^{N_{0}}$-dimensional standard Brownian motion. The covariance matrix $\Sigma(\theta)$ in (4.19) can be explicitly computed as

$$
\begin{equation*}
\Sigma(\theta)=Z(\theta)^{\prime} D(\theta)+D(\theta) Z(\theta)-D(\theta)-\bar{g}(\theta) \bar{g}^{\prime}(\theta) \tag{4.20}
\end{equation*}
$$

Here, $D(\theta)=\operatorname{diag}(\bar{g}(\theta))$ and $Z(\theta)=\left(I-B_{N_{0}}(\theta)+\mathbf{1}^{\prime}(\theta)\right)^{-1}$, where $\bar{g}(\theta) \in \mathbf{G}$. For each $\theta \in \mathscr{M}$, $B_{N_{0}}(\theta)$ is computed using (4.8) where the corresponding parameters $p(i), q(i)$ are used.

For general switching processes, we refer to [157]. In fact, more complex continuous-state dependent switching rather than Markovian switching are considered there. Equation (4.20) reveals that the covariance matrix of the tracking error depends on $B_{N_{0}}(\theta)$ and $\bar{g}(\theta)$ and, consequently, on the parameters $p$ and $q$ of the random graph. Recall from Section 4.2 that $B_{N_{0}}(\theta)$ is the transition matrix of the Markov chain which models the evolution of the expected degree distribution in duplication-deletion random graphs and can be computed from Theorem 4.3.1.

### 4.5 Estimating the Degree Distribution of Infinite Duplication-deletion Random Graphs

This section has two results: First, the results of Section 4.4 are extended to infinite random graphs without Markovian dynamics generated according to Procedure 4.5. Second, we show how this analysis can be extended to Markov-modulated probability mass functions with denumerable support. The analysis is non-standard, since it is formulated on a Hilbert space.

### 4.5.1 Infinite Random Graphs without Markovian Dynamics

Consider the infinite duplication-deletion random graph without Markovian dynamics generated according to Procedure 4.5 with 7 -tuple ( $1,1,1,1, p, q, G_{0}$ ). In this section, let $g_{n}$ represent the degree distribution of the infinite graph with support on the set of non-negative integers; its elements are denoted by $g_{n}^{i}, i=0,1,2, \ldots$. Recall from Section 4.3 .2 that, the size of such a graph increments at each time by one and thus the size of the graph at time $n$ is equal to $n$; that is $N_{n}=n$. Therefore,
the maximum degree of the graph at time $n$ cannot exceed $n-1$ and $g_{n}^{j}=0$ for $j \geq n$. Similar to the proof of Theorem 4.3.1, the following theorem asserts that the expected degree distribution of the infinite duplication-deletion random graph satisfies a recursive equation.

Theorem 4.5.1. Consider the infinite duplication-deletion random graph without Markovian dynamics generated according to Procedure 4.5 with 7 -tuple $\left(1,1,1,1, p, q, G_{0}\right)$, Let $\bar{g}_{n}=\mathbf{E}\left\{g_{n}\right\}$ denote the expected degree distribution of nodes with support on the set of non-negative integers. Then, $\bar{g}_{n}$ satisfies the following recursion

$$
\begin{equation*}
\bar{g}_{n+1}=\bar{g}_{n}+\frac{1}{n} L^{(n) /} \bar{g}_{n} \tag{4.21}
\end{equation*}
$$

where $L^{(n)}$ is a generator matrix of infinite size with elements:

$$
l_{j i}^{(n)}= \begin{cases}(1+q)\left(p^{i-1}+1+p(i-1)\right), & j=i-1 \text { and } 1 \leq i, j \leq n  \tag{4.22}\\ (1+q)\left(i i^{i-1}(1-p)+1+p i\right)-q(i+1), & j=i \text { and } 1 \leq i, j \leq n \\ (1+q)\binom{i+1}{i-1} p^{i-1}(1-p)^{2}+q(i+1), & j=i+1 \text { and } 1 \leq i, j \leq n \\ (1+q)\binom{j}{i-1} p^{i-1}(1-p)^{j-i+1}, & j>i+1 \text { and } 1 \leq i, j \leq n, \\ 0, & \text { otherwise }\end{cases}
$$

Proof. The proof is similar to the proof of Theorem 4.3.1 and is omitted due to the lack of space.

Remark 4.5.1. Theorem 4.3.2 in Section 4.3.2 asserts that the expected degree distribution converges to a power law probability distribution $\bar{g}$ with exponent $\beta>1$, if $\log p+p<\frac{q}{1+q}<p$; that is $\lim _{n \rightarrow \infty} \bar{g}_{n}^{i}=\frac{i^{-\beta}}{\zeta(\beta)}$. We assume that the dynamics of the degree distribution evolve on a faster time scale than the stochastic approximation algorithm. Therefore, it suffices to track the stationary degree distribution $\bar{g}$ given observations.

At each time $n$, the network administrator samples from the graph and records the degree of a randomly chosen vertex of the graph which is denoted by $y_{n}$. Let $z_{n}=\mathbf{e}_{y_{n}}$ denote the observation vector. Here, $\mathbf{e}_{i}$ is the $i$-th standard unit vector with support on the set of non-negative integers (i.e., $\left.\mathbf{e}_{i}=(0, \ldots, 1, \ldots)^{\prime} \in \mathbb{R}^{\infty}\right)$. The following stochastic approximation algorithm is used to estimate the expected degree distribution of the graph from such samples.

$$
\begin{equation*}
\widehat{g}_{n+1}=\widehat{g}_{n}+\varepsilon\left(z_{n}-\widehat{g}_{n}\right) \tag{4.23}
\end{equation*}
$$

Here, $\varepsilon>0$ denote a small positive step size and $\widehat{g}_{0}=\mathbf{e}_{1}$. Therefore, (4.23) is a Hilbert-spacevalued stochastic approximation algorithms. By means of the stochastic approximation (4.23), the network administrator can track the expected degree distribution of the infinite graph size increases over time.

Define

$$
\widehat{g}^{\varepsilon}(t)=\widehat{g}_{n} \text { for } t \in[n \varepsilon, n \varepsilon+\varepsilon) .
$$

Then $\widehat{g}^{\varepsilon}(\cdot) \in D\left([0, \infty): \ell_{2}\right)$ the space of functions defined on $[0, \infty)$ taking values in $\ell_{2}=\left\{z \in \mathbb{R}^{\infty}\right.$ : $\left.\sum_{i=0}^{\infty}\left\|z_{i}\right\|^{2}<\infty\right\}$ such that the functions are right continuous and have left limits endowed with the Skorohod topology. Here, we obtain a weak convergence result of the interpolated sequence of iterates. Theorem 4.5.2 below asserts that the mean square tracking error is bounded and shows that the sequnce of estimates obtained by (4.23) converge to the solution of an ODE. Before proceeding to the main theorem, we shall use the following conditions.

Theorem 4.5.2. Suppose Assumption 4.4.1 holds with the modification that $M=1$, i.e., there is no Markovian dynamics. Define $\widetilde{g}_{n}=\bar{g}-\widehat{g}_{n}$ Then, $\lim _{n \rightarrow \infty} \mathbf{E}\left\|\widetilde{g}_{n}\right\|^{2}=O(\varepsilon)$. Also, $\widehat{g}^{\varepsilon}(\cdot)$ is tight in $D\left([0, \infty): \ell_{2}\right)$. Any convergent subsequence has a limit $\widehat{g}(\cdot)$ that is the solution of the differential equation

$$
\begin{equation*}
\frac{d \widehat{g}(t)}{d t}=\bar{g}-\widehat{g}(t), \widehat{g}(0)=\mathbf{e}_{1} . \tag{4.24}
\end{equation*}
$$

Proof. The proof is presented in Appendix 4.8.8. The proof of the theorem is divided into several steps, which uses techniques in stochastic approximation [99] but with the modification that $\ell_{2}$ is a Hilbert space (see [61, 98]).

The above result concerns $n \rightarrow \infty$ and $\varepsilon \rightarrow 0, \varepsilon n$ remains to be bounded. We next obtain a result with $\varepsilon \rightarrow 0, n \rightarrow \infty, \varepsilon n \rightarrow \infty$.

Corollary 4.5.1. Consider $\widehat{g}^{\varepsilon}\left(\cdot+t_{\varepsilon}\right)$, where $t_{\varepsilon} \rightarrow \infty$ as $\varepsilon \rightarrow 0$. Under the condition of Theorem 4.5.2, $\widehat{g}^{\varepsilon}\left(\cdot+t_{\varepsilon}\right) \rightarrow \bar{g}$ in probability as $\varepsilon \rightarrow 0$.

Proof Note that $\left\{\widehat{g}_{k}\right\}$ is tight. Define $\widehat{g}^{\varepsilon, \text { large }}(\cdot)=\widehat{g}^{\varepsilon}\left(\cdot+t_{\varepsilon}\right)$. Using the same approach, we can show that $\left\{\widehat{g}^{\varepsilon, \text { large }}(\cdot)\right\}$ is tight. We extract a weakly convergent subsequence of $\left(\widehat{g}^{\varepsilon, \text { large }}(\cdot), \widehat{g}^{\varepsilon, \text { large }}(\cdot-\right.$ $T)$ ) with limit denoted by $\left(\widehat{g}(\cdot), \widehat{g}_{T}(\cdot)\right)$. We note that $\widehat{g}(0)=\widehat{g}_{T}(T)$ and that $\widehat{g}_{T}(0)$ belongs to a set that is bounded in probability. Writing it in variational form, we obtain

$$
\begin{aligned}
\widehat{g}_{T}(T) & =e^{-T} \widehat{g}_{T}(0)+\int_{0}^{T} e^{-(T-t)} \bar{g} d t=e^{-T} \widehat{g}_{T}(0)+\bar{g}-\int_{T}^{\infty} e^{-t} d t \\
& \rightarrow \bar{g} \text { as } T \rightarrow \infty .
\end{aligned}
$$

The desired result then follows.
To study the rate of variation of estimation error, we define the sequence of scaled estimation error $v_{n}=\left(\widehat{g}_{n}-\bar{g}\right) / \sqrt{\varepsilon}$. Theorem 4.5.3 asserts that the scaled estimation error satisfy a differential equation and provides a weak convergence results for it.

Theorem 4.5.3. Suppose assumptions of Theorem 4.5 .2 hold. Then, for sufficiently small $\varepsilon$ there is an $N_{\varepsilon}$ such that $\mathbf{E}\left\{\left\langle v_{n}, v_{n}\right\rangle\right\}=O(1)$ for all $n>N_{\varepsilon}$. Define the sequence of continuous-time interpolation of estimation error as

$$
v^{\varepsilon}(t)=v_{n} \text { for } t \in\left[\left(n-N_{\varepsilon}\right) \varepsilon,\left(n-N_{\varepsilon}+1\right) \varepsilon\right) .
$$

Under the assumptions of Theorem 4.5.2, $\left\{v^{\varepsilon}(\cdot)\right\}$ is tight in $D\left([0, \infty) ; \ell_{2}\right)$. Moreover, suppose that $v^{\varepsilon}(0)$ converges weakly to $v(0), v^{\varepsilon}(\cdot)$ converges weakly to $v(\cdot)$ such that $v(\cdot)$ is the solution of the following stochastic differential equation

$$
\begin{equation*}
d v(t)=-v(t) d t+d W(t) \tag{4.25}
\end{equation*}
$$

Here, $W(t)=\sum_{i=0}^{\infty} W_{i}(t) \mathbf{e}_{i}$ and the covariance operator is given by

$$
\begin{equation*}
\mathbf{E}\langle W(t), v\rangle\langle W(t), z\rangle=t\langle z, \Gamma v\rangle=t \sum_{i=0}^{\infty} \sigma_{i}^{2}\left\langle\mathbf{e}_{i}, v\right\rangle\left\langle\mathbf{e}_{i}, z\right\rangle \text { for } v, z \in \ell_{2}, \tag{4.26}
\end{equation*}
$$

where $W_{i}(\cdot)$ is a real-valued Wiener process with covariance $t \sigma_{i}^{2}$ and

$$
\sigma_{i}^{2}=\mathbf{E}\left[\left\langle z_{0}-\bar{g}, \mathbf{e}_{i}\right\rangle\right]^{2}+2 \sum_{j=1}^{\infty} \mathbf{E}\left\langle z_{0}-\bar{g}, \mathbf{e}_{i}\right\rangle\left\langle z_{j}-\bar{g}, \mathbf{e}_{i}\right\rangle .
$$

Proof. The proof is presented in Appendix 4.8.9.
Note that the covariance $\sigma_{i}^{2}$ depends on the stationary expected degree distribution $\bar{g}$ and thus is a function of the power law exponent $\beta$.

### 4.5.2 Markov-modulated Probability Mass Functions with Denumerable Support

Here, we extend the above results to the problem of tracking a time-varying probability mass function with infinite support. The aim is to track a probability mass function with support on the set of non-negative integers that evolves according to a slow Markov chain $\theta_{n}$ with $M$ states and initial probability distribution $\pi_{0}$. The state space $\mathscr{M}$, and the transition probability matrix $A^{\rho}$ of the underlying Markov chain $\theta_{n}$ are defined in (4.1) and (4.2), respectively. For each $\theta \in \mathscr{M}$, let

$$
\begin{equation*}
\bar{g}(\theta)=\left[\bar{g}^{1}(\theta), \bar{g}^{2}(\theta), \ldots\right]^{\prime} \tag{4.27}
\end{equation*}
$$

be a probability mass function with support on the set of non-negative integers such that $\sum_{i=1}^{\infty} \bar{g}^{i}(\theta)=$ 1 and $\bar{g}^{i}(\theta) \propto i^{-\beta_{\theta}}$, where $\beta_{\theta}>1$. When the underlying Markov chain $\theta_{n}$ jumps from one state to another within $\mathscr{M}, \bar{g}\left(\theta_{n}\right)$ switches accordingly.

At each time $n$, we sample $y_{n}\left(\theta_{n}\right)$ from PMF $\bar{g}\left(\theta_{n}\right)$; that is $y_{n}\left(\theta_{n}\right) \sim \bar{g}\left(\theta_{n}\right)$. Let $z_{n}\left(\theta_{n}\right)=\mathbf{e}_{y_{n}\left(\theta_{n}\right)}$ denote the observation vector. To estimate $\bar{g}\left(\theta_{n}\right)$, the following constant step size stochastic approximation algorithm is deployed

$$
\begin{equation*}
\widehat{g}_{n+1}=\widehat{g}_{n}+\varepsilon\left(z_{n}\left(\theta_{n}\right)-\widehat{g}_{n}\right) . \tag{4.28}
\end{equation*}
$$

Here $\varepsilon>0$ denotes a small positive step size and $\widehat{g}_{0}=\mathbf{e}_{1}$. We further assume that the Markov chain is slowly changing in that the rate of changes is an order slower than that of adaptation (4.28); that is $\rho=\varepsilon^{2}$.

To analyze the asymptotic properties of the stochastic approximation algorithm, we define the sequence of continuous-time interpolation $\widehat{g}^{\varepsilon}(t)=\widehat{g}_{n}$ for $t \in[n \varepsilon, n \varepsilon+\varepsilon)$. Similar to what have been obtained thus far for the non-Markovian case, with the details omitted, we obtain the following weak convergence results. Theorem 4.5 .4 states that the sequence of estimates obtained via Hilbert-space-valued stochastic approximation algorithm (4.28) converges weakly to the solution of an ODE which depends on the initial distribution of the underlying Markov chain.

Theorem 4.5.4. Suppose Assumptions 4.2 .1 and 4.4.1 hold. Then $\widehat{g}^{\varepsilon}(\cdot)$ is tight in $D\left([0, \infty): \ell_{2}\right)$. Any convergent subsequence has a limit $\widehat{g}(\cdot)$ that is the solution of the differential equation

$$
\begin{equation*}
\frac{d \widehat{g}(t)}{d t}=\sum_{\theta=1}^{M} \bar{g}(\theta) p_{\theta}-\widehat{g}(t), \widehat{g}(0)=\mathbf{e}_{1} \tag{4.29}
\end{equation*}
$$

where

$$
\bar{g}(\theta)=\sum_{i=0}^{\infty} \bar{g}^{i}(\theta) \mathbf{e}_{i}, \text { and } \quad\left(p_{\theta}: \theta \leq M\right)=\pi_{0}
$$

is the initial probability distribution of Markov chain.
Proof. The proof is presented in Appendix 4.8.10.
Furthermore, we can obtain the following corollary. The proof is similar to that of Corollary 4.5.1 and thus omitted.

Corollary 4.5.2. Consider $\widehat{g}^{\varepsilon}\left(\cdot+t_{\varepsilon}\right)$, where $t_{\varepsilon} \rightarrow \infty$ as $\varepsilon \rightarrow 0$. Under the condition of Theorem 4.5.4, $\widehat{g}^{\varepsilon}\left(\cdot+t_{\varepsilon}\right) \rightarrow \underline{g}=\sum_{\theta=1}^{M} p_{\theta} \bar{g}(\theta)$ in probability as $\varepsilon \rightarrow 0$.

Redefine $v_{n}=\left(\widehat{g}_{n}-\underline{g}\right) / \sqrt{\varepsilon}$. It can be shown that there exists $N_{\varepsilon}$ such that the sequence $\left\{v_{n}: n \geq\right.$ $\left.N_{\varepsilon}\right\}$ is tight. Next, redefine $v^{\varepsilon}(t)=v_{n}$ for $t \in\left[\varepsilon\left(n-N_{\varepsilon}\right), \varepsilon(n-N-\varepsilon)+\varepsilon\right)$. With a little more effort, we can also obtain the associated rates of convergence result, which is stated in the next theorem.

Theorem 4.5.5. Suppose Assumptions 4.2.1 and 4.4.1 hold. Then, $\left\{v^{\varepsilon}(\cdot)\right\}$ is tight in $D\left([0, \infty) ; \ell_{2}\right)$. Moreover, suppose that $v^{\varepsilon}(0)$ converges weakly to $v(0)$, then $v^{\varepsilon}(\cdot)$ converges weakly to $v(\cdot)$ such
that $v(\cdot)$ is the solution of the following stochastic differential equation (SDE)

$$
\begin{equation*}
d u(t)=-v(t) d t+\sum_{\theta=1}^{M} p_{\theta} d W(\theta, t) \tag{4.30}
\end{equation*}
$$

where for each $\theta \in \mathscr{M}, W(\theta, \cdot)$ is a Wiener process as given in Theorem 4.5.3.
Proof. The proof is similar to the proof of Theorem 4.5.3 with modifications similar to those of the proof of Theorem 4.5.4.

### 4.6 Numerical Examples

In this section, numerical examples are given to illustrate the results from Section 4.2, and Section 4.4.
The main conclusions are:

1. The infinite duplication-deletion random graph without Markovian dynamics generated by the duplication-deletion Procedure 4.5 satisfies a power law as stated in Theorem 4.3.2; see Example 4.6.1.
2. The degree distribution of the fixed size duplication-deletion random graph generated by the duplication-deletion Procedure 4.5 can be computed from Theorem 4.3.1. When $N_{0}$ (the size of the random graph) is sufficiently large, numerical results show that the degree distribution satisfies a power law as well; see Example 4.6.2.
3. The estimates obtained by stochastic approximation algorithm (4.14) follow the expected probability distribution precisely without information about the Markovian dynamics; see Example 4.6.3.
4. The larger the trace of the asymptotic covariance of the scaled tracking error, the greater the average degree of nodes and the searchability of the graph. This is illustrated in Example 4.6.4 below.

Example 4.6.1. Consider an infinite duplication-deletion random graph without Markovian dynamics (so $M=1$ ) generated by Procedure 4.5 with $p=0.5$ and $q=0.1$. Theorem 4.3.2 implies that the degree sequence of the resulting graph satisfies a power law with exponent computed using (4.11). Fig. 4.2 displays the un-normalized degree distribution on a logarithmic scale. The linearity in Fig. 4.2 (excluding the nodes with very small degree), implies that the resulting graph from duplication-deletion process satisfies a power law. As can be seen in Fig.4.2, the power law is a better approximation for the middle points compared to both ends.

Example 4.6.2. Consider the fixed size duplication-deletion random graph generated by Procedure 4.5 with $r=0, N_{0}=10, p=0.4$, and $q=0.1$. We consider $M=1$ (no Markovian dynamics) to illustrate Theorem 4.3.1. Fig. 4.3 depicts the normalized degree distribution of the fixed size duplication-deletion random graph obtained by Theorem 4.3.1. As can be seen in Fig. 4.3, the computed degree distribution is close to that obtained by simulation. The numerical results show that the degree distribution of the fixed size random graph also satisfies a power law for some values of $p$ when the size of random graph is sufficiently large. Fig. 4.4 shows the number of nodes with specific degree for the fixed size random graph obtained by Procedure 4.5 with $r=0, N_{0}=1000$, $p=0.4$, and $q=0.1$ on a logarithmic scale for both horizontal and vertical axes.

Example 4.6.3. Consider the fixed size Markov-modulated duplication-deletion random graph generated by Procedure 4.5 with $r=0$ and $N_{0}=500$. Assume that the underlying Markov chain has three states, $M=3$. We choose the following values for probabilities of connection and deletion: state (1): $p=q=0.05$, state (2): $p=0.2$ and $q=0.1$, and state (3): $p=0.4, q=0.15$. The sample path of the Markov chain jumps at times $n=3000$ from state (1) to state (2) and $n=6000$ from state (2) to state (3). As the state of the Markov chain evolves, the expected degree distribution, $\bar{g}(\theta)$, obtained by (4.9) evolves over time. The corresponding values for the expected degree distribution for nodes of degree $i=3$ are displayed in Fig.4.5 using a dotted line. The estimated probability mass function, $\widehat{g}_{n}$, obtained by the stochastic approximation algorithm (4.14) is plotted in Fig.4.5 using a solid line. The figure shows that the estimates using by the stochastic approximation algorithm (4.14) follow the expected degree distribution (4.9) satisfactorily even though the algorithm has no information about the underlying Markovian dynamics.

Example 4.6.4. Consider the fixed size Markov-modulated duplication-deletion random graph obtained by Procedure 4.5 with $M=91$ and $r=0$ and $N_{0}=1000$. For each value of $p(\theta)=$ $0.04+\theta \times 0.01, \theta \in\{1,2, \ldots, 91\}$ and $q \in\{0.05,0.1,0.15,0.2\}$, we compute $L(\theta)$ from (4.7) and consequently the stationary distribution, $\bar{g}(\theta)$, from (4.9). As expected, the stationary distribution does not depend on $q$ because only the deletion step in Procedure 4.5 occurs with probability $q$. From $\bar{g}(\theta)$, we compute the average degree of nodes, $\delta$. Fig.4.6 shows the average degree of nodes versus the probability of the connection in Procedure 4.5. As can be seen in Fig.4.6, with increasing the probability of connection in Procedure 4.5, the average degree of nodes in the graph (which is a measure for the connectivity of the graph, see [41]). Then for each value of $p(\theta)=0.04+\theta \times 0.01, \theta \in\{1,2, \ldots, 91\}$ and $q \in\{0.05,0.1,0.15,0.2\}$, the covariance matrix is computed using (4.6). Fig.4.7 depicts the trace of the covariance matrix, trace $(\Sigma(\theta))$, for each value of $p$ and $q$ versus the corresponding average degree of nodes (for each value of $p$ ). As can be seen in Fig.4.7, the trace of the covariance matrix is larger when the average degree of nodes is higher (the graph is highly connected).

Recall from Lemma 4.3.1, the order of delay in the searching problem can be computed by


Figure 4.2: Degree distribution of the duplication-deletion random graph satisfies a power law. The parameters are specified in Example 4.6.1 of Section 4.6.


Figure 4.3: Degree distribution of the fixed size duplication-deletion random graph. The parameters are specified in Example 4.6.2 of Section 4.6.


Figure 4.4: Degree distribution of the fixed size duplication-deletion random graph satisfies a power law when $N_{0}$ is sufficiently large. The parameters are specified in Example 4.6.2 of Section 4.6.


Figure 4.5: The estimates obtained by SA algorithm (4.14) follows the expected PMF precisely with no knowledge of the Markovian dynamics. The parameters are specified in Example 4.6.3


Figure 4.6: The average degree of nodes (as a measure of connectivity) of the fixed size Markovmodulated duplication-deletion random graph obtained by Procedure 4.5 for different values of the probability of connection, $p$, in Algorithm 4.5. The parameters are specified in Example 4.6.4 of Section 4.6.
$\lambda\left(N_{0}\right)=O\left(\frac{N_{0} \delta}{\bar{d}_{2}-\delta}\right)$. Knowing the degree distribution $\bar{g}(\theta), \delta$ and $\bar{d}_{2}$ can be computed for each value of $p \in\{0.05,0.06, \ldots, 0.95\}$. Fig. 4.8 shows the trace of the covariance matrix versus $\left(\frac{\delta}{\bar{d}_{2}-\delta}\right)$ as a measure of the searchability for each value of $q \in\{0.05,0.1,0.15,0.2\}$. As can be seen in Fig.4.8, the trace of covariance matrix is larger when the order of delay in the search problem in (4.10) is smaller ${ }^{42}$.

### 4.7 Closing Remarks

Markov-modulated duplication-deletion random graphs are analyzed in terms of their degree distribution. When the size of graph is fixed $(r=0)$ and $\rho$ is small, the expected degree distribution of the Markov-modulated duplication-deletion random graph can be computed from (4.6) for each state of the underlying Markov chain. This result allows us to express the structure of network (degree distribution) in terms of the dynamics of the model. We also showed that, the infinite duplication-deletion random graph without Markovian dynamics generated according to Procedure 4.5 ( $r=1, M=1$ ) satisfies a power law with component computed from (4.11). The importance of this result is that a single parameter (power law component) characterizes the structure of a possibly very large dy-

[^31]

Figure 4.7: Trace of the covariance matrix of scaled tracking error, trace $(\Sigma(\theta))$, versus the average degree of nodes as a measure of connectivity of the network. The parameters are specified in Example 4.6.3 of Sec.4.6.


Figure 4.8: Trace of the covariance matrix of the scaled tracking error, trace $(\Sigma(\theta))$, versus the order of delay in the searching problem as a measure of searchability of the network. The parameters are specified in Example 4.6.3 of Sec.4.6.
namic network.
Also a stochastic approximation algorithm was presented to adaptively estimate the degree distribution of random graphs. The stochastic approximation algorithm (4.14) does not assume knowledge of the Markov-modulated dynamics of the graph. Theorem 4.4.1 showed that the tracking error of the stochastic approximation algorithm is small and is in order of $O(\varepsilon)$. As a result of this bound, we showed that the scaled tracking error weakly converges to a diffusion process. Motivated by the analysis of social networks, we presented a Hilbert-space-valued stochastic approximation algorithm to estimate the expected degree distribution of the infinite duplication-deletion random graph without Markovian dynamics. The asymptotic behaviour of such an algorithm is analyzed in terms of the power law degree distribution. Finally, we extended the analysis to a Hilbert-space-valued stochastic approximation algorithm that aims to track a Markov-modulated probability mass function with denumerable support. Using weak convergence methods, it was shown that the estimates obtained via such an algorithm converge weakly to the solution of an ordinary differential equation. It was also shown that the interpolated sequence of scaled tracking error converges weakly to the solution of a stochastic differential equation.

### 4.8 Proof of Results

### 4.8.1 Proof of Theorem 4.3.1

The proof is based on the proof of Lemma 4.1 in [41, Chapter 4, p79]. To compute the expected degree distribution of the Markov-modulated random graph, we find a relation between the number of nodes with specific degree at time $n$ and the degree distribution of the graph at time $n-1$. Recall that the $i$-th element of $f_{n}, f_{n}^{i}$, denotes the number of vertices with degree $i$ at time $k$. Given the resulting graph at time $n$, the aim is to find the expected number of nodes with degree $i+1$ at time $n+1$. The following events can occur that result in a node with degree $i+1$ at time $n+1$ :

- Degree of a node with degree i increments by one in the duplication step (Step 1 of the duplication-deletion Procedure 4.5) and remains unchanged in the deletion step (Step 2):
- A node with degree $i$ is chosen at the duplication step as a parent node and remains unchanged in the deletion step. The probability of occurrence of such an event is

$$
r\left(1-\frac{q(i+1)+q(1+p i)-q(1+p i)(i+1) / N_{n}}{N_{n}}\right) \frac{f_{n}^{i}}{N_{n}} ;
$$

the probability of choosing a node with degree $i$ is $\frac{f_{n}^{i}}{N_{n}}$ and the probability of the event
that this node remains unchanged in the deletion step is ${ }^{43}$

$$
1-\frac{q(i+1)+q(1+p i)-q(1+p i)(i+1) / N_{n}}{N_{n}} .
$$

- One neighbor of a node with degree $i$ is selected as a parent node; the parent node connects to its neighbors (including the node with degree $i$ ) with probability $p$ in the edge-duplication part of Step 1. The probability of such an event is

$$
r \frac{f_{n}^{i} p i}{N_{n}}\left(1-\frac{q(i+2)+q(1+p(i+1))-q(1+p(i+1))(i+2) / N_{n}}{N_{n}}\right) .
$$

Note that the node whose degree is incremented by one in this event should remain unaffected in Step 2; the probability of being unchanged in Step 2 for such a node is

$$
1-\frac{q(i+2)+q(1+p(i+1))-q(1+p(i+1))(i+2) / N_{n}}{N_{n}} .
$$

- A node with degree $i+1$ remains unchanged in both Step 1 and Step 2 of Procedure 4.5:
- Using the same argument as above, the probability of such an event is

$$
f_{n}^{i+1}\left(1-q \frac{i+3+p(i+1)-\frac{(1+p(i+1))(i+2)}{N_{n}}}{N_{n}}\right)\left(1-r \frac{p(i+1)+1}{N_{n}}\right) .
$$

- A new node with degree $i+1$ is generated in Step 1:
- The degree of the most recently generated node (in the vertex- duplication part of Step 1) increments to $i+1$; the new node connects to " $i$ " neighbors of the parent node and remains unchanged in Step 2. The probability of this scenario is

$$
r\left(1-q \frac{i+3+p(i+1)-\frac{(1+p(i+1))(i+2)}{N_{n}}}{N_{n}}\right) \sum_{j \geq i} \frac{f_{n}^{j}}{N_{n}}\binom{j}{i} p^{i}(1-p)^{j-i} .
$$

- Degree of a node with degree $i+2$ decrements by one in Step 2:

[^32]- A node with degree $i+2$ remains unchanged in the duplication step and one of its neighbors is eliminated in the deletion step. The probability of this event is

$$
q\left(\frac{i+2}{N_{n}}\right)\left(1-\frac{p(i+2)+1}{N_{n}}\right) .
$$

- A node with degree $i+1$ is generated in Step 2:
- The degree of the node generated in the vertex-duplication part of duplication step within Step 2 increments to $i+1$. The probability of this event is

$$
q \sum_{j \geq i} \frac{1}{N_{n}} f_{n}^{j}\binom{j}{i} p^{i}(1-p)^{j-i} .
$$

- Degree of a node with degree i increments by one in Step 2:
- A node with degree $i$ remains unchanged in Step 1 and its degree increments by one in the duplication part of Step 2. The corresponding probability is

$$
\frac{q(1+p i)}{N_{n}}\left(1-\frac{1+p i}{N_{n}}\right) .
$$

Let $\Omega$ denote the set of all arbitrary graphs and $\mathscr{F}_{n}$ denote the sigma algebra generated by graphs $G_{\tau}, \tau \leq n$. Considering the above events that result in a node with degree $i+1$ at time $n+1$, the following recurrence formula can be derived for the conditional expectation of $f_{n+1}^{i+1}$ :

$$
\begin{align*}
& \mathbf{E}\left\{f_{n+1}^{i+1} \mid \mathscr{F}_{n}\right\}= \\
& \left(1-q \frac{i+3+p(i+1)-\frac{(1+p(i+1))(i+2)}{N_{n}}}{N_{n}}\right)\left(1-r \frac{p(i+1)+1}{N_{n}}\right) f_{n}^{i+1} \\
& +r\left(1-\frac{q(i+1)+q(1+p i)-q(1+p i)(i+1) / N_{n}}{N_{n}}\right)\left(\frac{1+p i}{N_{n}}\right) f_{n}^{i} \\
& +r\left(1-q \frac{i+3+p(i+1)-\frac{(1+p(i+1))(i+2)}{N_{n}}}{N_{n}}\right) \sum_{j \geq i} \frac{f_{n}^{j}}{N_{n}}\binom{j}{i} p^{i}(1-p)^{j-i} \\
& +q \sum_{j \geq i} \frac{f_{n}^{j}}{N_{n}}\binom{j}{i} p^{i}(1-p)^{j-i}+q\left(\frac{i+2}{N_{n}}\right)\left(1-\frac{p(i+2)+1}{N_{n}}\right) f_{n}^{i+2} \\
& +\frac{q(1+p i)}{N_{n}}\left(1-\frac{1+p i}{N_{n}}\right) f_{n}^{i} . \tag{4.31}
\end{align*}
$$

Let $\bar{f}_{n}^{i}=\mathbf{E}\left\{f_{n}\right\}$. By taking expectation of both sides of (4.31) with respect to trivial sigma algebra $\{\Omega, \emptyset\}$, the smoothing property of conditional expectations yields:

$$
\begin{align*}
& \bar{f}_{n+1}^{i+1}= \\
& \quad\left(1-q \frac{i+3+p(i+1)-\frac{(1+p(i+1))(i+2)}{N_{n}}}{N_{n}}\right)\left(1-r \frac{p(i+1)+1}{N_{n}}\right) \bar{f}_{n}^{i+1} \\
& +r\left(1-\frac{q(i+1)+q(1+p i)-\frac{q(1+p i)(i+1)}{N_{n}}}{N_{n}}\right)\left(\frac{1+p i}{N_{n}}\right) \bar{f}_{n}^{i} \\
& +r\left(1-q \frac{i+3+p(i+1)-\frac{(1+p(i+1))(i+2)}{N_{n}}}{N_{n}}\right) \sum_{j \geq i} \frac{\bar{f}_{n}^{j}}{N_{n}}\binom{j}{i} p^{i}(1-p)^{j-i} \\
& +q \sum_{j \geq i} \frac{1}{N_{n}} \bar{f}_{n}^{j}\binom{j}{i} p^{i}(1-p)^{j-i}+q\left(\frac{i+2}{N_{n}}\right)\left(1-\frac{p(i+2)+1}{N_{n}}\right) \bar{f}_{n}^{i+2} \\
& +\frac{q(1+p i)}{N_{n}}\left(1-\frac{1+p i}{N_{n}}\right) \bar{f}_{n}^{i} . \tag{4.32}
\end{align*}
$$

Assuming that size of the graph is sufficiently large, each term like $\frac{\bar{f}_{n}^{i}}{N_{n}^{2}}$ can be neglected. Eq. (4.32) can be written as

$$
\begin{align*}
\bar{f}_{n+1}^{i+1} & =\left(1-\frac{q(i+2)+(r+q)(p(i+1)+1)}{N_{n}}\right) \bar{f}_{n}^{i+1} \\
& +\left(\frac{(1+p i)(r+q)}{N_{n}}\right) \bar{f}_{n}^{i}+q\left(\frac{i+2}{N_{n}}\right) \bar{f}_{n}^{i+2} \\
& +q \sum_{j \geq i} \frac{1}{N_{n}} \bar{f}_{n}^{j}\binom{j}{i} p^{i}(1-p)^{j-i} \tag{4.33}
\end{align*}
$$

Using (4.32), we can write the following recursion for the $(i+1)$-th element of $\bar{g}_{n+1}$ :

$$
\begin{align*}
\bar{g}_{n+1}^{i+1} & =\left(\frac{N_{n}-(q(i+2)+(r+q)(p(i+1)+1))}{N_{n+1}}\right) \bar{g}_{n}^{i+1} \\
& +\left(\frac{(1+p i)(r+q)}{N_{n+1}}\right) \bar{g}_{n}^{i}+q\left(\frac{i+2}{N_{n+1}}\right) \bar{g}_{n}^{i+2} \\
& +q \sum_{j \geq i} \frac{1}{N_{n+1}} \bar{g}_{n}^{j}\binom{j}{i} p^{i}(1-p)^{j-i} \tag{4.34}
\end{align*}
$$

Since the probability of duplication step $r=0$, the number of vertices does not increase. Thus, $N_{n}=N_{0}$ and (4.34) can be written as

$$
\begin{align*}
\bar{g}_{n+1}^{i+1}= & \left(1-\frac{1}{N_{0}}(q(i+2)+q(p(i+1)+1))\right) \bar{g}_{n}^{i+1} \\
& +\frac{1}{N_{0}}\left((1+p i) q \bar{g}_{n}^{i}+\frac{1}{N_{0}} q(i+2) \bar{g}_{n}^{i+2}\right) \\
& +\frac{1}{N_{0}} q \sum_{j \geq i} \bar{g}_{n}^{j}\binom{j}{i} p^{i}(1-p)^{j-i} . \tag{4.35}
\end{align*}
$$

It is clear in (4.35) that the vector $\bar{g}_{n+1}$ depends on elements of $\bar{g}_{n}$. Using matrix notation, (4.35) can be expressed as

$$
\begin{equation*}
\bar{g}_{n+1}=\left(I+\frac{1}{N_{0}} L^{\prime}\right) \bar{g}_{n} \tag{4.36}
\end{equation*}
$$

where $L$ is defined as (4.7).
To prove that $L$ is a generator, we need to show that $l_{i i}<0$ and $\sum_{i=1}^{N_{0}} l_{k i}=0$. Accordingly,

$$
\begin{align*}
\sum_{i=1}^{N_{0}} l_{k i} & =-(q(k+1)+q(1+p k))+(1+p k) q \\
& +q k+q \sum_{k \leq i-1}\binom{k}{i-1} p^{i-1}(1-p)^{k-i+1} \\
& =-q+q \sum_{k \leq i-1}\binom{k}{i-1} p^{i-1}(1-p)^{k-i+1} . \tag{4.37}
\end{align*}
$$

Let $m=i-1$. Then, (4.37) can be rewritten as

$$
\begin{align*}
\sum_{i=1}^{N_{0}} l_{i k} & =-q+q \sum_{m=0}^{k}\binom{k}{m} p^{m}(1-p)^{k-m} \\
& =-q+q(1-p)^{k} \sum_{m=0}^{k}\binom{k}{m}\left(\frac{p}{1-p}\right)^{m} . \tag{4.38}
\end{align*}
$$

Knowing that $\sum_{m=0}^{k}\binom{k}{m} a^{m}=(1+a)^{k}$, (4.38) can be written as

$$
\begin{equation*}
\sum_{i=1}^{N_{0}} l_{i k}=-q+q(1-p)^{k}\left(\frac{1}{1-p}\right)^{k}=0 \tag{4.39}
\end{equation*}
$$

Also, it can be shown that $l_{i i}<0$. Since $p^{i-1} \leq 1, p^{i-1}<1+\frac{2}{i}+p+p^{i}$. Consequently, iqp ${ }^{i-1}(1-$ $p)-q(i+2+i p)<0$. Therefore, $t_{i i}<0$ and the desired result follows.

### 4.8.2 Proof of Theorem 4.3.2

Proof. To prove Theorem 4.3.2, we first compute the power law component, $\beta$, and then we prove that the expected degree distribution converges to the power law distribution with component $\beta$. Let $\bar{f}_{n}(i)=\mathbf{E}\left\{f_{n}(i)\right\}$. Similar to (4.31), $\bar{f}_{n}\left(\theta_{n}, i\right)$ can be written as

$$
\begin{align*}
\bar{f}_{n+1}(i+1) & =\left(1-q \frac{(i+2)+(1+p(i+1))}{N_{n}}\right)\left(1-\frac{p(i+1)+1}{N_{n}}\right) \bar{f}_{n}(i+1) \\
& +\left(1-q \frac{(i+1)+(1+p i)}{N_{n}}\right)\left(\frac{1+p i}{N_{n}}\right) \bar{f}_{n}(i) \\
& +\left(1-q \frac{(i+2)+(1+p(i+1))}{N_{n}}\right) \sum_{j \geq i} \frac{1}{N_{n}} \bar{f}_{n}(j)\binom{j}{i} p^{i}(1-p)^{j-i} \\
& +q \sum_{j \geq i} \frac{1}{N_{n}} \bar{f}_{n}(j)\binom{j}{i} p^{i}(1-p)^{j-i} \\
& +q\left(\frac{i+2}{N_{n}}\right)\left(1-\frac{p(i+2)+1}{N_{n}}\right) \bar{f}_{n}(i+2) \\
& +\frac{q(1+p i)}{N_{n}}\left(1-\frac{1+p i}{N_{n}}\right) \bar{f}_{n}(i) \\
& +q\left(\frac{i+2}{N_{n}}\right)\left(\frac{p((i+1)+1)}{N_{n}}\right) \bar{f}_{n}(i+1) . \tag{4.40}
\end{align*}
$$

To compute the power law component, we can heuristically assume that $\bar{f}_{n}(i)=a_{i} t$ as $N_{n}=n$ goes to infinity (we will prove this precisely later on this section). Therefore, each term like $\frac{\bar{f}_{n}\left(i^{\prime}\right)}{N_{n}^{2}}$ can be neglected as $n$ approaches infinity. So (4.40) can be re-written as

$$
\begin{align*}
\bar{f}_{n+1}(i+1)= & \left(1-\frac{q(i+2)+(1+q)(p(i+1)+1)}{N_{n}}\right) \bar{f}_{n}(i+1)+\left(\frac{(1+p i)(1+q)}{N_{n}}\right) \bar{f}_{n}(i) \\
& +q\left(\frac{i+2}{N_{n}}\right) \bar{f}_{n}(i+2)+(1+q) \sum_{j \geq i} \frac{1}{N_{n}} \bar{f}_{n}(j)\binom{j}{i} p^{i}(1-p)^{j-i} . \tag{4.41}
\end{align*}
$$

Substituting $\bar{f}_{\tau}(j)=a_{j} \tau$ and $N_{n}=n$ in (4.98) yields

$$
\begin{align*}
a_{i+1}(n+1)= & a_{i+1} n-a_{i+1}((1+p(i+1))(1+q)+q(i+2))+(1+q)(1+p i) a_{i}+q(i+2) a_{i+2} \\
& +(1+q) \sum_{j \geq i} a_{j}\binom{j}{i} p^{i}(1-p)^{j-i} . \tag{4.42}
\end{align*}
$$

Taking all terms with $a_{i+1}$ to the left hand side, we have

$$
\begin{align*}
\left.a_{i+1}(1+(1+q)(1+p(i+1))+q(i+2))\right)= & (1+q)\left((1+p i) a_{i}+\sum_{j \geq i} a_{j}\binom{j}{i} p^{i}(1-p)^{j-i}\right) \\
& +q(i+2) a_{i+2} . \tag{4.43}
\end{align*}
$$

Dividing both sides of (4.43) by $a_{i}$ yields

$$
\begin{align*}
\left.\frac{a_{i+1}}{a_{i}}(1+(1+q)(1+p(i+1))+q(i+2))\right)= & (1+q)\left((1+p i)+\sum_{j \geq i} \frac{a_{j}}{a_{i}}\binom{j}{i} p^{i}(1-p)^{j-i}\right) \\
& +q(i+2) \frac{a_{i+2}}{a_{i}} \tag{4.44}
\end{align*}
$$

Solving Equation (4.43) for $a_{i}$, we can complete the proof of Theorem 4.3.2 The following lemma whose proof can be found in [42] is used to solve the recurrence relation for $a_{i}$.

Lemma 4.8.1.

$$
\begin{equation*}
\sum_{j \geq i} \frac{a_{j}}{a_{i}}\binom{j}{i} p^{i}(1-p)^{j-i}=p^{\beta-1}+\mathrm{O}\left(\frac{1}{i}\right) \tag{4.45}
\end{equation*}
$$

Proof. The proof is presented in Appendix 4.8.3.
To solve (4.43) for $a_{i}$, we can further assume that $a_{i}=C i^{-\beta}$ [41]. Therefore, $\frac{a_{i+\alpha}}{a_{i}}=\left(\frac{i+\alpha}{i}\right)^{-\beta}$

$$
\begin{align*}
\left(1-\frac{\beta}{i}\right)(1+(1+q)(1+p(i+1))+q(i+2))= & (1+q)\left(1+p i+p^{\beta-1}\right) \\
& +\mathrm{O}\left(\frac{1}{i}\right)+q(i+2)\left(1-\frac{2 \beta}{i}\right) \tag{4.46}
\end{align*}
$$

Neglecting the $\mathrm{O}\left(\frac{1}{i}\right)$ terms, yields

$$
\begin{equation*}
(1+q)\left(p^{\beta-1}+p \beta-p\right)=1+\beta q \tag{4.47}
\end{equation*}
$$

Note that the proof presented above depends on few assumptions. To give a rigorous proof, the succeeding steps should be followed as described in [41]:

- First, we need to show that the $\operatorname{limit} \lim _{n \rightarrow \infty} \frac{1}{n} \mathbf{E}\left\{f_{n}(i)\right\}$ exists.
- Let $a_{i}$ be the solution of (4.43) such that $\sum_{i=1}^{\infty} a_{i}=1$ and $a_{0}=0$, then it is needed to show that

$$
\begin{equation*}
\lim _{n \rightarrow \infty} \frac{1}{n} \mathbf{E}\left\{f_{n}(i)\right\}=a_{i} \tag{4.48}
\end{equation*}
$$

- Finally, we should show that $a_{i}$ is proportional to $i^{-\beta}$, where $\beta$ is the root of (4.47).

To complete the proof we define new function as follows $h_{n}(i)=\frac{1}{n} \sum_{k=1}^{i} \mathbf{E}\left\{f_{n}(k)\right\}$ which can be described as CDF of degree of each node in random graph. It is sufficient to show that for all $i>0$,

$$
\begin{equation*}
\lim _{n \rightarrow \infty} h_{n}(i)=\sum_{k=1}^{i} a_{k} \tag{4.49}
\end{equation*}
$$

where $a_{i}$ is the solution of (4.43). It is obvious if (4.49) holds, $h_{n}(i)-h_{n}(i-1)=a_{i}$ and thus

$$
\lim _{n \rightarrow \infty} \frac{1}{n} \mathbf{E}\left\{f_{n}(i)\right\}=a_{i}
$$

(as presented in (4.48)). The following lemma gives a recurrence formula to compute the value of $h(n+1, i)$.

## Lemma 4.8.2.

$$
\begin{equation*}
h_{n+1}(i)=D_{n+1}(i) h_{n}(i)+B_{n+1}(i) h_{n}(i-1)+C_{n+1}(i) h_{n}(i+1)+\frac{1+q}{n+1} \sum_{j \geq i-1} h_{n}(j) F(j, i-1, p), \tag{4.50}
\end{equation*}
$$

where

$$
\begin{aligned}
D_{n+1}(i) & =\left(\frac{n-(q(i+2)+(1+q)(p i+1))}{n+1}\right) \\
B_{n+1}(i) & =\frac{(1+q)(1+p i)}{n+1} \\
C_{n+1}(i) & =\frac{q(i+1)}{n+1} \\
F(j, i, p) & =\sum_{k=0}^{i}\binom{j}{k} p^{k}(1-p)^{j-k}-\sum_{k=0}^{i}\binom{j+1}{k} p^{k}(1-p)^{j+1-k}
\end{aligned}
$$

This lemma can be proved by induction. The complete proof can be found in Appendix 4.8.4. The recursive equation presented in Lemma 4.8.2 is used later to prove that the degree distribution converges to a power law.

Lemma 4.8.3. Let $s_{i}=\sum_{k=1}^{i} a_{i}$ and

$$
\begin{equation*}
\omega(n)=\sup _{i \geq 1} \frac{h_{n}(i)}{s_{i}}, \tag{4.51}
\end{equation*}
$$

where $h_{n}(i)$ satisfies (4.50). Then the limit $\lim _{n \rightarrow \infty} \omega(n)$ exists and we have $\lim _{n \rightarrow \infty} \omega(n)=1$.

Sketch of the proof Knowing that $h_{n}(i)$ satisfies the recurrence formula (4.50), the proof is similar to [41]. Plugging $i=n$ in (4.51) yields $\omega(n) \geq \frac{h_{n}(n)}{s_{n}} \geq \frac{1}{s_{n}} \geq 1$. Using the Lemma 4.8.2 and similar to [41], it can be shown that $\omega(n+1) \leq \omega(n) . \omega(n)$ is bounded and decreasing, so the limit of $\lim _{n \rightarrow \infty} \omega(n)$ exists. To show $\lim _{n \rightarrow \infty} \omega(n)=1$, we assume that $\lim _{n \rightarrow \infty} \omega(n)=c$. It can be shown that if $c \neq 1, \omega(n) \leq 1$ is violated. Thus $c=1$ and the proof is complete.

### 4.8.3 Proof of Lemma 4.8.1

Proof.

$$
\begin{align*}
\sum_{j \geq i} \frac{a_{j}}{a_{i}}\binom{j}{i} p^{i}(1-p)^{j-i} & =\sum_{j \geq i}\left(\frac{i}{j}\right)^{\beta}\binom{j}{i} p^{i}(1-p)^{j-i} \\
& =\sum_{j \geq i}\left(\frac{i}{j}\right)^{\beta}\binom{j}{j-i} p^{i}(1-p)^{j-i} \\
& =\left(1+O\left(\frac{1}{i}\right)\right) \sum_{j \geq i}\binom{j-\beta}{j-i} p^{i}(1-p)^{j-i} \\
& =\left(1+O\left(\frac{1}{i}\right)\right) p^{i} \sum_{k=0}\binom{k+i-\beta}{k}(1-p)^{k} \\
& =\left(1+O\left(\frac{1}{i}\right)\right) p^{i} \sum_{k=0}\binom{\beta-i-1}{k}(-1)^{k}(1-p)^{k} \\
& =\left(1+O\left(\frac{1}{i}\right)\right) p^{i} p^{\beta-i-1}=\left(1+O\left(\frac{1}{i}\right)\right) p^{\beta-1} . \tag{4.52}
\end{align*}
$$

### 4.8.4 Proof of Lemma 4.8.2

Proof. We prove the lemma by induction on $i$ :

For $i=1 \quad$ It is sufficient to show that:
$h(n+1,1)=D_{n+1}(1) h(n, 1)+C_{n+1}(1) h(n, 2)+\frac{1}{n+1} \sum_{j \geq 1} h(n, j) F(j, 0, p)$. Also using the definition of $F(j, i, p)$, we can rewrite $F(j, 0, p)$ as $(1-p)^{j}-(1-p)^{j+1}$. The number of nodes with degree
one at time $n+1$ can be written as following

$$
\begin{align*}
\mathbf{E}\{f(n+1,1)\}= & \left(1-\frac{(1+q)(1+p)+q}{n}\right) \mathbf{E}\left\{f_{n}(1)\right\}+\frac{2 q}{n} \mathbf{E}\left\{f_{n}(2)\right\} \\
& +(1+q) \sum_{j \geq 1} \frac{1}{n} \mathbf{E}\left\{f_{n}(j)\right\}(1-p)^{j} . \tag{4.53}
\end{align*}
$$

Note that (4.53) is slightly different from the general equation for each $i$, (4.98). Because as described in Section 4.1, neighbors of a node with degree one cannot be eliminated from the graph to maintain the connectivity in the graph. Therefore, a node with degree one can change in the deletion step if that node is selected in the deletion step (with probability $q$ ). Using (4.53), $h(n+1,1)$ can be written as

$$
\begin{align*}
h(n+1,1) & =\frac{1}{n+1} \mathbf{E}\{f(n+1,1)\} \\
& =\frac{1}{n+1}\left(\left(1-\frac{(1+q)(1+p)+q}{n}\right) \mathbf{E}\left\{f_{n}(1)\right\}+\frac{2 q}{n} \mathbf{E}\left\{f_{n}(2)\right\}\right) \\
& +\frac{1}{n+1} \sum_{j \geq 1} \frac{1+q}{n} \mathbf{E}\left\{f_{n}(j)\right\}(1-p)^{j} . \tag{4.54}
\end{align*}
$$

We know that $h(n, 0)=0$ for all $n$. Using the definition of $h(\cdot, \cdot)$ and (4.53), (4.54) can be re-arranged as follows

$$
\begin{align*}
h(n+1,1)= & \frac{1}{n+1}\left((n-((1+q)(1+p)+q)) h(n, 1)+\frac{2 q}{n}(h(n, 2)-h(n, 1))\right. \\
& \left.+(1+q) \sum_{j \geq 1}(h(n, j)-h(n, j-1))(1-p)^{j}\right) \\
= & \frac{1}{n+1}\left((n-(3 q+(1+q)(1+p))) h(n, 1)+\frac{2 q}{n} h(n, 2)\right) \\
& +\frac{1+q}{n+1} \sum_{j \geq 1}(h(n, j)-h(n, j-1))(1-p)^{j} \tag{4.55}
\end{align*}
$$

$\sum_{j \geq 1}(h(n, j)-h(n, j-1))(1-p)^{j}$ can be written in terms of the $F(j, i, p)$.

$$
\begin{align*}
\sum_{j \geq 1}(h(n, j)-h(n, j-1))(1-p)^{j} & =\sum_{j \geq 1} h(n, j)(1-p)^{j}-\sum_{j \geq 1}\left(h(n, j-1)(1-p)^{j}\right. \\
& =\sum_{j \geq 1} h(n, j)(1-p)^{j}-\sum_{j \geq 1}\left(h(n, j)(1-p)^{j}+1\right. \\
& =\sum_{j \geq 1} h(n, j)\left((1-p)^{j}-(1-p)^{j+1}\right) \\
& =\sum_{j \geq 1} h(n, j) F(j, 0, p) . \tag{4.56}
\end{align*}
$$

Substituting (4.56) in (4.55) yields

$$
\begin{align*}
h(n+1,1) & =\frac{1}{n+1}\left((n-((1+q)(1+p)+3 q)) h(n, 1)+\frac{2 q}{n} h(n, 2)+(1+q) \sum_{j \geq 1} h(n, j) F(j, 0, p)\right) \\
& =D_{n+1}(1) h(n, 1)+C_{n+1}(1) h(n, 2)+\frac{1+q}{n+1} \sum_{j \geq 1} h(n, j) F(j, 0, p) . \tag{4.57}
\end{align*}
$$

Thus (4.50) holds for $i=1$. Now it is assumed that (4.50) holds for $i=k$, we want to show that it also holds for $i=k+1$.

$$
\begin{align*}
\mathbf{E}\{f(n+1, k+1)\}= & \left(1-\frac{q(k+2)+(1+q)(p(k+1)+1)}{n}\right) \mathbf{E}\{f(n, k+1)\} \\
& +\left(\frac{(1+q)(1+p k)}{n}\right) \mathbf{E}\left\{f_{n}(k)\right\}+\left(\frac{q(k+2)}{n}\right) \mathbf{E}\left\{f_{n}(k+2)\right\} \\
& +(1+q) \sum_{j \leq k} \frac{f_{n}(j)}{n}\binom{j}{k} p^{k}(1-p)^{j-k} . \tag{4.58}
\end{align*}
$$

from definition of $h(n, k)$, we have : $\mathbf{E}\left\{f_{n}(k)\right\}=n(h(n, k)-h(n, k-1))$. Eq. (4.58) can be rewritten as follows

$$
\begin{align*}
\mathbf{E}\{f(n+1, k+1)\}= & (n-(q(k+2)+(1+q)(p(k+1)+1)))(h(n, k+1)-h(n, k)) \\
& +(1+q)(1+p k)(h(n, k)-h(n, k-1))+q(k+2)(h(n, k+2)-h(n, k+1)) \\
& +(1+q) \sum_{j \leq k}(h(n, j)-h(n, j-1))\binom{j}{k} p^{k}(1-p)^{j-k} . \tag{4.59}
\end{align*}
$$

Using the Abel summation identity, and knowing that

$$
F(j, k, p)=\sum_{k=0}^{k}\binom{j}{k} p^{k}(1-p)^{j-k}-\sum_{k=0}^{k}\binom{j+1}{k} p^{k}(1-p)^{j+1-k}
$$

the last term can be written as

$$
\begin{align*}
& \sum_{j \leq k}(h(n, j)-h(n, j-1))\binom{j}{k} p^{k}(1-p)^{j-k}  \tag{4.60}\\
& \quad=\sum_{j \geq k}\left(\binom{j}{k} p^{k}(1-p)^{j-k}-\binom{j+1}{k} p^{k}(1-p)^{j+1-k}\right)-p^{k} h(n, k-1) \\
& \quad=-p^{k} h(n, k-1)+\sum_{j \geq k} h(n, j)(F(j, k, p)-F(j, k-1, p)) . \tag{4.61}
\end{align*}
$$

Substituting (4.60) in (4.59) yields

$$
\begin{align*}
\mathbf{E}\{f(n+1, k+1)\}= & h(n, k+2)(q(k+2))+h(n, k+1)(n-(2 q(k+2)+(1+q)(p(k+1)+1))) \\
& +h(n, k)((1+q)(2+p(2 k+1)) \\
& +q(k+2)-n)+h(n, k-1)(1+q)\left(-1-p k-p^{k}\right) \\
& +(1+q) \sum_{j \geq k} h(n, j)(F(j, k, p)-F(j, k-1, p)) . \tag{4.62}
\end{align*}
$$

The value of $h(n+1, k+1)$ can be computed using $h(n, k+1)$ and $\mathbf{E}\left\{f_{n}(k+1)\right\}$ as follows

$$
\begin{equation*}
h(n+1, k+1)=h(n+1, k)+\frac{1}{n+1} \mathbf{E}\{f(n+1, k+1)\} . \tag{4.63}
\end{equation*}
$$

Eq.(4.62) gives an expression for $\mathbf{E}\{f(n+1, k+1)\}$ in terms of the value of $h(\cdot, \cdot)$ at time $n$. Substituting (4.62) in (4.63) gives a recursive equation for computing $h(n+1, k+1)$ :

$$
\begin{align*}
h(n+1, k+1)= & h(n+1, k)+\frac{1}{n+1} \mathbf{E}\{f(n+1, k+1)\} \\
= & D_{n+1}(k) h(n, k)+B_{n+1}(k) h(n, k-1)+C_{n+1} h(n, k+1) \\
& +\frac{1+q}{n+1} \sum_{j \geq k-1} h(n, j) F(j, k-1, p) \\
& +\frac{1}{n+1}(h(n, k+2)(q(k+2))+h(n, k+1) \\
& (n-(2 q(k+2)+(1+q)(p(k+1)+1))) \\
& +h(n, k)\left((1+q)(2+p(2 k+1))+h(n, k-1)(1+q)\left(-1-p k-p^{k}\right)\right. \\
& \left.+(1+q) \sum_{j \geq k} h(n, j)(F(j, k, p)-F(j, k-1, p))\right) . \tag{4.64}
\end{align*}
$$

We assume that (4.98) holds for $i=k$ so substituting the values for $D_{n+1}(k), B_{n+1}(k)$, and $C_{n+1}(k)$ from (4.98) in (4.64) yields

$$
\begin{align*}
h(n+1, k+1)= & h(n, k+2)\left(\frac{q(k+2)}{n+1}\right)+h(n, k+1)\left(\frac{n-(q(k+3)+(1+q)(p(k+1)+1))}{n+1}\right) \\
& +h(n, k)\left(\frac{(1+q)(1+p(k+1))}{n+1}\right)+\frac{1+q}{n+1} \sum_{j \geq k} h(n, j)(F(j, k, p)) . \tag{4.65}
\end{align*}
$$

(4.65)can be written as follows

$$
\begin{align*}
h(n+1, k+1)= & D_{n+1}(k+1) h(n, k+1)+B_{n+1}(k+1) h(n, k)+C_{n+1}(k+1) h(n, k+2) \\
& +\frac{1+q}{n+1} \sum_{j \geq k} h(n, j) F(j, k, p) . \tag{4.66}
\end{align*}
$$

Thus, (4.98) holds for $i=k+1$ and the proof is completed by induction.

### 4.8.5 Proof of Theorem 4.4.1

Proof. Define the Lyapunov function $V(x)=\left(x^{\prime} x\right) / 2$ for $x \in \mathbb{R}^{N_{0}}$. Use $\mathbf{E}_{n}$ to denote the conditional expectation with respect to the $\sigma$-algebra $\mathscr{H}_{n}$ generated by $\left\{z_{j}\left(\theta_{j}\right), \theta_{j}, \quad j \leq n\right\}$. Then,

$$
\begin{align*}
& \mathbf{E}_{n}\left\{V\left(\widetilde{g}_{n+1}\right)-V\left(\widetilde{g}_{n}\right)\right\}=\mathbf{E}_{n}\left\{\widetilde{g}_{n}\left[-\varepsilon \widetilde{g}_{n}+\varepsilon\left(z_{n}\left(\theta_{n}\right)-\bar{g}\left(\theta_{n}\right)\right)+\bar{g}\left(\theta_{n}\right)-\bar{g}\left(\theta_{n+1}\right)\right]\right\} \\
& \quad+\mathbf{E}_{n}\left\{\left\|-\varepsilon \widetilde{g}_{n}+\varepsilon\left(z_{n}\left(\theta_{n}\right)-\bar{g}\left(\theta_{n}\right)\right)+\bar{g}\left(\theta_{n}\right)-\bar{g}\left(\theta_{n+1}\right)\right\|^{2}\right\} \tag{4.67}
\end{align*}
$$

where $z_{n}\left(\theta_{n}\right)$ and $\bar{g}\left(\theta_{n}\right)$ are vectors in $\mathbb{R}^{N_{0}}$ with elements $z_{n}^{i}\left(\theta_{n}\right)$ and $\bar{g}\left(\theta_{n}\right)^{i}, \quad 1 \leq i \leq N_{0}$, respectively. Due to the Markovian assumption and the structure of the transition matrix of $\theta_{n}$, defined in (4.2),

$$
\begin{gather*}
\mathbf{E}_{n}\left\{\bar{g}\left(\theta_{n}\right)-\bar{g}\left(\theta_{n+1}\right)\right\}=\mathbf{E}\left\{\bar{g}\left(\theta_{n}\right)-\bar{g}\left(\theta_{n+1}\right) \mid \theta_{n}\right\}=\sum_{i=1}^{M} \mathbf{E}\left\{\bar{g}(i)-\bar{g}\left(\theta_{n+1}\right) \mid \theta_{n}=i\right\} I\left\{\theta_{n}=i\right\} \\
=\sum_{i=1}^{M}\left[\bar{g}(i)-\sum_{j=1}^{M} \bar{g}(j) A_{i j}^{\rho}\right] I\left\{\theta_{n}=i\right\}=-\rho \sum_{i=1}^{M} \sum_{j=1}^{M} \bar{g}(j) q_{i j} I\left\{\theta_{n}=i\right\}=O(\rho) \tag{4.68}
\end{gather*}
$$

where $I\{\cdot\}$ denotes the indicator function. Similarly, it is easily seen that

$$
\begin{equation*}
\mathbf{E}_{n}\left\{\left\|\bar{g}\left(\theta_{n}\right)-\bar{g}\left(\theta_{n+1}\right)\right\|^{2}\right\}=O(\rho) . \tag{4.69}
\end{equation*}
$$

Using $K$ to denote a generic positive value (with the notation $K K=K$ and $K+K=K$ ), a familiar inequality $a b \leq \frac{a^{2}+b^{2}}{2}$ yields

$$
\begin{equation*}
O(\varepsilon \rho)=O\left(\varepsilon^{2}+\rho^{2}\right) \tag{4.70}
\end{equation*}
$$

Moreover, we have $\left\|\widetilde{g}_{n}\right\|=\left\|\widetilde{g}_{n}\right\| \cdot 1 \leq\left(\left\|\widetilde{g}_{n}\right\|^{2}+1\right) / 2$. Thus,

$$
\begin{equation*}
O(\rho)\left\|\widetilde{g}_{n}\right\| \leq O(\rho)\left(V\left(\widetilde{g}_{n}\right)+1\right) . \tag{4.71}
\end{equation*}
$$

Then, detailed estimates lead to

$$
\begin{align*}
& \mathbf{E}_{n}\left\{\left\|-\varepsilon \widetilde{g}_{n}+\varepsilon\left(z_{n}\left(\theta_{n}\right)-\bar{g}\left(\theta_{n}\right)\right)+\bar{g}\left(\theta_{n}\right)-\bar{g}\left(\theta_{n+1}\right)\right\|^{2}\right\} \\
& \quad \leq K \mathbf{E}_{n}\left\{\varepsilon^{2}\left\|\widetilde{g}_{n}\right\|^{2}+\varepsilon^{2} \|\left(z_{n}\left(\theta_{n}\right)-\bar{g}\left(\theta_{n}\right)\left\|^{2}+\varepsilon^{2}\right\| \widetilde{g}_{n}^{\prime}\left(z_{n}\left(\theta_{n}\right)-\bar{g}\left(\theta_{n+1}\right)\right) \|\right.\right. \\
& \left.\quad+\varepsilon\left\|\widetilde{g}_{n}\left(\bar{g}\left(\theta_{n}\right)-\bar{g}\left(\theta_{n+1}\right)\right)\right\|+\varepsilon\left\|\left(z_{n}\left(\theta_{n}\right)-\bar{g}\left(\theta_{n}\right)\right)^{\prime}\left(\bar{g}\left(\theta_{n}\right)-\bar{g}\left(\theta_{n+1}\right)\right)\right\|\right\} \\
&  \tag{4.72}\\
& \quad+\mathbf{E}_{n}\left\{\left\|\bar{g}\left(\theta_{n}\right)-\bar{g}\left(\theta_{n+1}\right)\right\|\right\}^{2} .
\end{align*}
$$

It follows that

$$
\begin{equation*}
\mathbf{E}_{n}\left\{\left\|-\varepsilon \widetilde{g}_{n}+\varepsilon\left(z_{n}\left(\theta_{n}\right)-\bar{g}\left(\theta_{n}\right)\right)+\bar{g}\left(\theta_{n}\right)-\bar{g}\left(\theta_{n+1}\right)\right\|^{2}\right\}=O\left(\varepsilon^{2}+\rho^{2}\right)\left(V\left(\widetilde{g}_{n}\right)+1\right) . \tag{4.7.7}
\end{equation*}
$$

Furthermore,

$$
\begin{align*}
\mathbf{E}_{n}\{V & \left.\left(\widetilde{g}_{n+1}\right)-V\left(\widetilde{g}_{n}\right)\right\}=-2 \varepsilon V\left(\widetilde{g}_{n}\right)+\varepsilon \mathbf{E}_{n}\left\{\widetilde{g}_{n}^{\prime}\left[z_{n}\left(\theta_{n}\right)-\bar{g}\left(\theta_{n}\right)\right]\right\}  \tag{4.74}\\
& +\mathbf{E}_{n}\left\{\widetilde{g}_{n}\left[\bar{g}\left(\theta_{n+1}\right)-\bar{g}\left(\theta_{n}\right)\right]\right\}+O\left(\varepsilon^{2}+\rho^{2}\right)\left(V\left(\widetilde{g}_{n}\right)+1\right) .
\end{align*}
$$

To obtain the desired bound, define $V_{1}^{\rho}$ and $V_{2}^{\rho}$ as follows:

$$
\begin{align*}
& V_{1}^{\rho}(\widetilde{g}, n)=\varepsilon \sum_{j=n}^{\infty} \widetilde{g}^{\prime} \mathbf{E}_{n}\left\{z_{j}\left(\theta_{j}\right)-\bar{g}\left(\theta_{j}\right)\right\}, \\
& V_{2}^{\rho}(\widetilde{g}, n)=\sum_{j=n}^{\infty} \widetilde{g}^{\prime} \mathbf{E}_{n}\left\{\bar{g}\left(\theta_{j}\right)-\bar{g}\left(\theta_{j+1}\right)\right\} . \tag{4.75}
\end{align*}
$$

It can be shown that

$$
\begin{align*}
& \left|V_{1}^{\rho}(\widetilde{g}, n)\right|=O(\varepsilon)(V(\widetilde{g})+1),  \tag{4.76}\\
& \left|V_{2}^{\rho}(\widetilde{g}, n)\right|=O(\rho)(V(\widetilde{g})+1) .
\end{align*}
$$

Define $W(\widetilde{g}, n)$ as

$$
\begin{equation*}
W(\widetilde{g}, n)=V(\widetilde{g})+V_{1}^{\rho}(\widetilde{g}, n)+V_{2}^{\rho}(\widetilde{g}, n) . \tag{4.77}
\end{equation*}
$$

This leads to

$$
\begin{align*}
& \mathbf{E}_{n}\left\{W\left(\widetilde{g}_{n+1}, n+1\right)-W\left(\widetilde{g}_{n}, n\right)\right\}=\mathbf{E}_{n}\left\{V_{1}^{\rho}\left(\widetilde{g}_{n+1}, n+1\right)-V_{1}^{\rho}\left(\widetilde{g}_{n}, n\right)\right\} \\
& \quad+\mathbf{E}_{n}\left\{V\left(\widetilde{g}_{n+1}\right)-V\left(\widetilde{g}_{n}\right)\right\}+\mathbf{E}_{n}\left\{V_{2}^{\rho}\left(\widetilde{g}_{n+1}, n+1\right)-V_{2}^{\rho}\left(\widetilde{g}_{n}, n\right)\right\} . \tag{4.78}
\end{align*}
$$

Moreover,

$$
\begin{equation*}
\mathbf{E}_{n}\left\{W\left(\widetilde{g}_{n+1}, n+1\right)-W\left(\widetilde{g}_{n}, n\right)\right\}=-2 \varepsilon V\left(\widetilde{g}_{n}\right)+O\left(\varepsilon^{2}+\rho^{2}\right)\left(V\left(\widetilde{g}_{n}\right)+1\right) \tag{4.79}
\end{equation*}
$$

Equation (4.79) can be rewritten as

$$
\begin{equation*}
\mathbf{E}_{n}\left\{W\left(\widetilde{g}_{n+1}, n+1\right)-W\left(\widetilde{g}_{n}, n\right)\right\} \leq O\left(\varepsilon^{2}+\rho^{2}\right)\left(W\left(\widetilde{g}_{n}, n\right)+1\right)-2 \varepsilon W\left(\widetilde{g}_{n}, n\right) \tag{4.80}
\end{equation*}
$$

If $\varepsilon$ and $\rho$ are chosen small enough, then there exists a small $\lambda$ such that $-2 \varepsilon+O\left(\rho^{2}\right)+O\left(\varepsilon^{2}\right) \leq$ $-\lambda \varepsilon$. Therefore, (4.80) can be rearranged as

$$
\begin{equation*}
\mathbf{E}_{n}\left\{W\left(\widetilde{g}_{n+1}, n+1\right)\right\} \leq(1-\lambda \varepsilon) W\left(\widetilde{g}_{n}, n\right)+O\left(\varepsilon^{2}+\rho^{2}\right) \tag{4.81}
\end{equation*}
$$

Taking expectation of both sides yields

$$
\begin{equation*}
\mathbf{E}\left\{W\left(\widetilde{g}_{n+1}, n+1\right)\right\} \leq(1-\lambda \varepsilon) \mathbf{E}\left\{W\left(\widetilde{g}_{n}, n\right)\right\}+O\left(\varepsilon^{2}+\rho^{2}\right) \tag{4.82}
\end{equation*}
$$

Iterating on (4.82) then results

$$
\begin{equation*}
\mathbf{E}\left\{W\left(\widetilde{g}_{n+1}, n+1\right)\right\} \leq(1-\lambda \varepsilon)^{n-N_{\rho}} \mathbf{E}\left\{W\left(\widetilde{g}_{N_{\rho}}, N_{\rho}\right)\right\}+\sum_{j=N_{\rho}}^{n} O\left(\varepsilon^{2}+\rho^{2}\right)(1-\lambda \varepsilon)^{j-N_{\rho}} \tag{4.83}
\end{equation*}
$$

As the result,

$$
\begin{equation*}
\mathbf{E}\left\{W\left(\widetilde{g}_{n+1}, n+1\right)\right\} \leq(1-\lambda \varepsilon)^{n-N_{\rho}} \mathbf{E}\left\{W\left(\widetilde{g}_{N_{\rho}}, N_{\rho}\right)\right\}+O\left(\varepsilon+\rho^{2} / \varepsilon\right) \tag{4.84}
\end{equation*}
$$

If $n$ is large enough, one can approximate $(1-\lambda \varepsilon)^{n-N_{\rho}}=O(\varepsilon)$. Therefore,

$$
\begin{equation*}
\mathbf{E}\left\{W\left(\widetilde{g}_{n+1}, n+1\right)\right\} \leq O\left(\varepsilon+\frac{\rho^{2}}{\varepsilon}\right) \tag{4.85}
\end{equation*}
$$

Finally, using (4.76) and replacing $W\left(\widetilde{g}_{n+1}, n+1\right)$ with $V\left(\widetilde{g}_{n+1}\right)$, we obtain

$$
\begin{equation*}
\mathbf{E}\left\{V\left(\widetilde{g}_{n+1}\right)\right\} \leq O\left(\rho+\varepsilon+\frac{\rho^{2}}{\varepsilon}\right) \tag{4.86}
\end{equation*}
$$

### 4.8.6 Sketch of the Proof of Theorem 4.4.2

Proof. Since the proof is similar to [154, Theorem 4.5], we only indicate the main steps in what follows and omit most of the verbatim details.

Step 1: First, we show that the two component process $\left(\widehat{g}^{\varepsilon}(\cdot), \theta^{\varepsilon}(\cdot)\right)$ is tight in $D\left([0, T]: \mathbb{R}^{N_{0}} \times\right.$ $M)$. Using techniques similar to [156, Theorem 4.3], it can be shown that $\theta^{\varepsilon}(\cdot)$ converges weakly to a continuous-time Markov chain generated by $Q$. Thus, we mainly need to consider $\widehat{g}^{\varepsilon}(\cdot)$. We show that

$$
\begin{equation*}
\lim _{\Delta \rightarrow 0} \limsup _{\varepsilon \rightarrow 0} \mathbf{E}\left[\sup _{0 \leq s \leq \Delta} \mathbf{E}_{t}^{\varepsilon}\left\|\widehat{g}^{\varepsilon}(t+s)-\widehat{g}^{\varepsilon}(t)\right\|^{2}\right]=0 \tag{4.87}
\end{equation*}
$$

where $\mathbf{E}_{t}^{\varepsilon}$ denotes the conditioning on the past information up to $t$. Then, the tightness follows from the criterion [97, p. 47].

Step 2: Since $\left(\widehat{g}^{\varepsilon}(\cdot), \theta^{\varepsilon}(\cdot)\right)$ is tight, we can extract weakly convergent subsequence according to the Prohorov theorem; see [99]. To figure out the limit, we show that $\left(\widehat{g}^{\varepsilon}(\cdot), \theta^{\varepsilon}(\cdot)\right)$ is a solution of the martingale problem with operator $L_{0}$. For each $i \in \mathscr{M}$ and continuously differential function with compact support $f(\cdot, i)$, the operator is given by

$$
\begin{equation*}
L_{0} f(\widehat{g}, i)=\nabla f^{\prime}(\widehat{g}, i)[-\widehat{g}+\bar{g}(i)]+\sum_{j \in \mathscr{M}} q_{i j} f(\widehat{g}, j), i \in \mathscr{M} . \tag{4.88}
\end{equation*}
$$

We can further demonstrate the martingale problem with operator $L_{0}$ has a unique solution (in the sense of in distribution). Thus, the desired convergence property follows.

### 4.8.7 Sketch of the Proof of Theorem 4.4.3

Proof. The proof comprises of four steps as described below:
Step 1: First, note

$$
\begin{equation*}
v_{n+1}=v_{n}-\varepsilon v_{n}+\sqrt{\varepsilon}\left(y_{n+1}-\mathbf{E} \bar{g}\left(\theta_{n}\right)\right)+\frac{\mathbf{E}\left[\bar{g}\left(\theta_{n}\right)-\bar{g}\left(\theta_{n+1}\right]\right.}{\sqrt{\varepsilon}} . \tag{4.89}
\end{equation*}
$$

The approach is similar to that of [154, Theorem 5.6]. Therefore, we will be brief.
Step 2: Define an operator

$$
\begin{equation*}
\mathscr{L} f(v, i)=-\nabla f^{\prime}(v, i) v+\frac{1}{2} \operatorname{tr}\left[\nabla^{2} f(v, i) \Sigma(i)\right]+\sum_{j \in \mathscr{M}} q_{i j} f(v, j), i \in \mathscr{M}, \tag{4.90}
\end{equation*}
$$

for function $f(\cdot, i)$ with compact support that has continuous partial derivatives with respect to $v$ up to the second order. It can be shown that the associated martingale problem has a unique solution (in the sense of in distribution).

Step 3: It is natural now to work with a truncated process. For a fixed, but otherwise arbitrary $r_{1}>0$, define a truncation function

$$
q^{r_{1}}(x)= \begin{cases}1, & \text { if } x \in S^{r_{1}} \\ 0, & \text { if } x \in \mathbb{R}^{N_{0}}-S^{r_{1}}\end{cases}
$$

where $S^{r_{1}}=\left\{x \in \mathbb{R}^{N_{0}}:\|x\| \leq r_{1}\right\}$. Then, we obtain the truncated iterates

$$
\begin{equation*}
v_{n+1}^{r_{1}}=v_{n}^{r_{1}}-\varepsilon v_{n}^{r_{1}} q^{r_{1}}\left(v_{n}^{r_{1}}\right)+\sqrt{\varepsilon}\left(y_{n+1}-\mathbf{E} \bar{g}\left(\theta_{n}\right)\right)+\frac{\mathbf{E}\left[\bar{g}\left(\theta_{n}\right)-\bar{g}\left(\theta_{n+1}\right]\right.}{\sqrt{\varepsilon}} q^{r_{1}}\left(v_{n}^{r_{1}}\right) \tag{4.91}
\end{equation*}
$$

Define $v^{\varepsilon, r_{1}}(t)=v_{n}^{r_{1}}$ for $t \in[\varepsilon n, \varepsilon n+\varepsilon)$. Then, $v^{\varepsilon, r_{1}}(\cdot)$ is an $r$-truncation of $v^{\varepsilon}(\cdot)$; see [99, p. 284] for a definition. We then show the truncated process $\left(v^{\varepsilon, r_{1}}(\cdot), \theta^{\varepsilon}(\cdot)\right)$ is tight. Moreover, by Prohorov's theorem, we can extract a convergent subsequence with limit $\left(v^{r_{1}}(\cdot), \theta(\cdot)\right)$ such that the limit $\left(v^{r_{1}}(\cdot), \theta(\cdot)\right)$ is the solution of the martingale problem with operator $\mathscr{L}^{r_{1}}$ defined by

$$
\begin{equation*}
\mathscr{L}^{r_{1}} f^{r_{1}}(v, i)=-\nabla \prime f^{r_{1}}(v, i) v+\frac{1}{2} \operatorname{tr}\left[\nabla^{2} f^{r_{1}}(v, i) \Sigma(i)\right]+\sum_{j \in \mathscr{M}} q_{i j} f^{r_{1}}(v, j) \tag{4.92}
\end{equation*}
$$

for $i \in \mathscr{M}$, where $f^{r_{1}}(v, i)=f(v, i) q^{r_{1}}(v)$.
Step 4: Letting $r_{1} \rightarrow \infty$, we show that the un-truncated process also converges and the limit, denoted by $(v(\cdot), \theta(\cdot))$, is precisely the martingale problem with operator $\mathscr{L}$ defined in (4.90). The limit covariance can further be evaluated as in [154, Lemma 5.2].

### 4.8.8 Proof of Theorem 4.5.2

Proof. The proof of the theorem is divided into several steps and uses techniques in stochastic approximation [99] but with the modification that $\ell_{2}$ is a Hilbert space (see [61, 98]). Whenever possible, we only indicate the main idea and refer to the literature of stochastic approximation.

Step 0: Note that (4.24) has a unique solution for each initial condition since it is linear in $\widehat{g}(\cdot)$.
Step 1: Preliminary estimates. From (4.23), we obtain that for $0<\varepsilon<1$, the elements of $\widehat{g}_{n}$ are non-negative and add up to one. Thus, $\widehat{g}_{n}$ is bounded.

In addition, define $V(\widehat{g})=\frac{1}{2}\langle\widehat{g}-\bar{g}, \widehat{g}-\bar{g}\rangle$, which can be thought of as a Lyapunov function. Then using perturbed Lyapunov function argument [99], it can be shown

$$
\begin{equation*}
\mathbf{E} V\left(\widehat{g}_{n}\right)=O(\varepsilon) \tag{4.93}
\end{equation*}
$$

Step 2: Tightness of $\left\{\widehat{g}^{\varepsilon}(\cdot)\right\}$. Henceforth, we often use $t / \varepsilon$ and $(t+s) / \varepsilon$ to denote $\lfloor t / \varepsilon\rfloor$ and $\lfloor(t+s) / \varepsilon\rfloor$, the integer parts of $t / \varepsilon$ and $(t+s) / \varepsilon$, respectively. By using the boundedness of $\left\{\widehat{g}_{n}\right\}$ established in the first step together with the Hölder inequality, we have for each $0<T<\infty$, any
$t \geq 0$, any $0<\delta$, any $0<s \leq \delta$, and $\varepsilon>0$,

$$
\begin{aligned}
& \mathbf{E}_{t}^{\varepsilon}\left\|\widehat{g}^{\varepsilon}(t+s)-\widehat{g}^{\varepsilon}(t)\right\|^{2} \\
& \quad \leq \mathbf{E}_{t}^{\varepsilon}\left|\varepsilon \sum_{j=t / \varepsilon}^{(t+s) / \varepsilon-1}\left[z_{j}-\widehat{g}_{j}\right]\right|^{2} \\
& \leq K \varepsilon\left(\frac{t+s}{\varepsilon}-\frac{t}{\varepsilon}\right) \leq K s,
\end{aligned}
$$

where $K>0$ is independent of $\varepsilon$ and $\mathbf{E}_{t}^{\varepsilon}$ denotes the conditional expectation with respect to $\mathscr{F}_{t}^{\varepsilon}$. Thus

$$
\begin{align*}
& \mathbf{E}_{t}^{\varepsilon}\left\|\widehat{g}^{\varepsilon}(t+s)-\widehat{g}^{\varepsilon}(t)\right\|^{2} \leq K s, \\
& \lim _{\delta \rightarrow 0} \limsup _{\varepsilon \rightarrow 0} \mathbf{E}\left[\sup _{0<s \leq \delta} \mathbf{E}_{t}^{\varepsilon}\left\|\widehat{g}^{\varepsilon}(t+s)-\widehat{g}^{\varepsilon}(t)\right\|^{2}\right]=0 . \tag{4.94}
\end{align*}
$$

The tightness criterion (see [97, Theorem 3, p. 47] with $\mathbb{R}^{r}$ replaced by $\ell_{2}$; see also [61]) enables us to conclude that $\left\{\widehat{g}^{\varepsilon}(\cdot)\right\}$ is tight in $D\left([0, \infty): \ell_{2}\right)$.

Step 3: Characterization of the limit process. Since $\left\{\widehat{g}^{\varepsilon}(\cdot)\right\}$ is tight, by Prohorov's theorem, we can extract a convergent subsequence. Select such a sequence and still denote it by $\widehat{g}^{\varepsilon}(\cdot)$ with limit denoted by $\widehat{g}(\cdot)$. By using the Skorohod representation, with a slight abuse of notation, we may assume that $\widehat{g}^{\varepsilon}(\cdot)$ converges to $\widehat{g}(\cdot)$ w.p. 1 and the convergence is uniform on any bounded time interval. We shall show that $\widehat{g}(\cdot)$ is a solution of the martingale problem with operator

$$
\mathscr{L} f(\widehat{g})=\langle\nabla f(\widehat{g}),[\bar{g}-\widehat{g}]\rangle
$$

for any $f(\cdot) \in C_{0}^{1}\left(\ell_{2}: \mathbb{R}\right)$ (collection of real-valued $C^{1}$ functions defined on $\ell_{2}$ with compact support). We need to show that

$$
f(\widehat{g}(t))-f(\widehat{g}(0))-\int_{0}^{t} \mathscr{L} f(\widehat{g}(\tau)) d \tau \text { is a marginale. }
$$

To prove the martingale property, we pick out any bounded and continuous function $h(\cdot)$ defined on $\ell_{2}$, any $T<\infty$, any $0<t, s \leq T$, any positive integer $\kappa$, and $t_{l_{1}} \leq t$ for any $l \leq \kappa$. To derive the desired property, we need only show that

$$
\begin{equation*}
\mathbf{E} h\left(\widehat{g}\left(t_{l_{1}}\right): l_{1} \leq \kappa\right)\left(f(\widehat{g}(t+s))-f(\widehat{g}(t))-\int_{t}^{t+s} \mathscr{L} f(\widehat{g}(\tau)) d \tau\right)=0 . \tag{4.95}
\end{equation*}
$$

To prove (4.95), we work with the process indexed by $\varepsilon$. First, by the weak convergence and the Skorohod representation,

$$
\begin{align*}
& \lim _{\varepsilon \rightarrow 0} \mathbf{E} h\left(\widehat{g}^{\varepsilon}\left(t_{l_{1}}\right): l_{1} \leq \kappa\right)\left[f\left(\widehat{g}^{\varepsilon}(t+s)\right)-f\left(\widehat{g}^{\varepsilon}(t)\right)\right]  \tag{4.96}\\
& =\mathbf{E} h\left(\widehat{g}\left(t_{l_{1}}\right): l_{1} \leq \kappa\right)[f(\widehat{g}(t+s))-f(\widehat{g}(t))] .
\end{align*}
$$

Choose a sequence of integers $\left\{m_{\varepsilon}\right\}$ such that $m_{\varepsilon} \rightarrow \infty$ as $\varepsilon \rightarrow 0$ but $\Delta_{\varepsilon}=\varepsilon m_{\varepsilon} \rightarrow 0$. Next, we note

$$
\begin{aligned}
f\left(\widehat{g}^{\varepsilon}(t+s)\right)-f\left(\widehat{g}^{\varepsilon}(t)\right) & =f\left(\widehat{g}_{(t+s) / \varepsilon}\right)-f\left(\widehat{g}_{t / \varepsilon}\right) \\
& =\sum_{\substack{(t s) / \varepsilon-1}}^{l m_{\varepsilon}=t / \varepsilon}\left[f\left(\widehat{g}_{l m_{\varepsilon}+m_{\varepsilon}}\right)-f\left(\widehat{g}_{l m_{\varepsilon}}\right)\right] \\
& =\varepsilon \sum_{l(t+s) / \varepsilon-1}^{t+s=t / \varepsilon}\left\langle\nabla f\left(\widehat{g}_{l m_{\varepsilon}}\right), \sum_{j=l m_{\varepsilon}+m_{\varepsilon}-1}^{l m_{\varepsilon}}\left[z_{j}-\widehat{g}_{j}\right]\right\rangle+o(1) \\
& =\sum_{l m_{\varepsilon}=t / \varepsilon}^{(t+s) / \varepsilon-1} \Delta_{\varepsilon}\left\langle\nabla f\left(\widehat{g}_{l m_{\varepsilon}}\right), \frac{1}{m_{\varepsilon}} \sum_{j=l m_{\varepsilon}}^{l m_{\varepsilon}+m_{\varepsilon}-1}\left[z_{j}-\widehat{g}_{j}\right]\right\rangle+o(1),
\end{aligned}
$$

where $o(1) \rightarrow 0$ in probability as $\varepsilon \rightarrow 0$. The stationarity and the mixing condition imply that

$$
\frac{1}{m_{\varepsilon}} \sum_{j=l m_{\varepsilon}}^{l m_{\varepsilon}+m_{\varepsilon}-1} \mathbf{E}_{l m_{\varepsilon}} z_{j} \rightarrow \mathbf{E} z_{0}=\sum_{i=0}^{\infty} \mathbf{e}_{i} P\left(y_{0}=i\right)=\bar{g} \text { in probability }
$$

Therefore,

$$
\begin{align*}
& \mathbf{E} h\left(\widehat{g}^{\varepsilon}\left(t_{l_{1}}\right): l_{1} \leq \kappa\right)\left[\sum_{l m_{\varepsilon}=t / \varepsilon}^{(t+s) / \varepsilon-1} \Delta_{\varepsilon}\left\langle\nabla f\left(\widehat{g}_{l m_{\varepsilon}}\right), \frac{1}{m_{\varepsilon}} \sum_{j=l m_{\varepsilon}}^{l m_{\varepsilon}+m_{\varepsilon}-1} z_{j}\right\rangle\right] \\
& =\mathbf{E} h\left(\widehat{g}^{\widehat{g}}\left(t_{l_{1}}\right): l_{1} \leq \kappa\right)\left[\sum_{l m_{\varepsilon}=t / \varepsilon}^{(t+s) / \varepsilon-1} \Delta_{\varepsilon}\left\langle\nabla f\left(\widehat{g}{ }_{l m_{\varepsilon}}\right), \frac{1}{m_{\varepsilon}} \sum_{j=l m_{\varepsilon}}^{l m_{\varepsilon}+m_{\varepsilon}-1} \mathbf{E}_{l m_{\varepsilon}} z_{j}\right\rangle\right]  \tag{4.97}\\
& \rightarrow \mathbf{E} h\left(\widehat{g}\left(t_{l_{1}}\right): l_{1} \leq \kappa\right)\left(\int_{t}^{t+s}\langle\nabla f(\widehat{g}(\tau)), \bar{g}\rangle d \tau\right) \text { as } \varepsilon \rightarrow 0 .
\end{align*}
$$

Likewise,

$$
\begin{align*}
& \mathbf{E} h\left(\widehat{g}^{\varepsilon}\left(t_{l_{1}}\right): l_{1} \leq \kappa\right)\left[-\sum_{l m_{\varepsilon}=t / \varepsilon}^{(t+s) / \varepsilon-1} \Delta_{\varepsilon}\left\langle\nabla f\left(\widehat{g}_{l m_{\varepsilon}}\right), \frac{1}{m_{\varepsilon}} \sum_{j=l m_{\varepsilon}}^{l m_{\varepsilon}+m_{\varepsilon}-1} \widehat{g}_{j}\right\rangle\right]  \tag{4.98}\\
& \rightarrow \mathbf{E} h\left(\widehat{g}\left(t_{l_{1}}\right): l_{1} \leq \kappa\right)\left[-\int_{t}^{t+s}\langle\nabla f(\widehat{g}(\tau)), \widehat{g}(\tau)\rangle d \tau\right] .
\end{align*}
$$

Combing (4.96)-(4.98), (4.95) follows.

### 4.8.9 Proof of Theorem 4.5.3

Proof. In the proof of Theorem 4.5.3, we use several lemmas and propositions described below. From (4.23),

$$
\begin{equation*}
v_{n+1}=v_{n}-\varepsilon v_{n}+\sqrt{\varepsilon}\left(z_{n}-\bar{g}\right) . \tag{4.99}
\end{equation*}
$$

Lemma 4.8.1. Under assumption Theorem 4.5.2, for sufficiently small $\varepsilon$, there is an $N_{\varepsilon}$ such that $\mathbf{E} V\left(v_{n}\right)=O(1)$ for all $n \geq N_{\varepsilon}$.

Proof The proof uses a perturbed Lyapunov function argument.
To proceed, recall the definition of covariance operator and Wiener process [46, 98] on $\ell_{2}$. A covariance $\Gamma$ of an $\ell_{2}$-valued random variable $y$ is an operator from $\ell_{2}$ to $\ell_{2}$ defined by $\Gamma v=E Y\langle v, y\rangle$ for any $v \in \ell_{2}$. A process $W(\cdot)$ is a zero mean (stationary increment) $\ell_{2}$-valued Wiener process if there are mutually independent real-valued, zero mean, Wiener processes $\left\{W_{i}(\cdot)\right\}$ with covariances $t \rho_{i}$ satisfying $\sum_{i=0}^{\infty} \rho_{i}<\infty$ and there is an orthonormal sequence $\left\{\beta_{i}\right\}$ with $\beta_{i} \in \ell_{2}$ such that $W(t)=$ $\sum_{i=0}^{\infty} W_{i}(t) \beta_{i}$. For $v, z \in \ell_{2}$, the covariance operator of $W(t)$ is defined by

$$
\begin{equation*}
\mathbf{E}\langle W(t), v\rangle\langle W(t), z\rangle=t\langle z, \Gamma v\rangle=t \sum_{i=0}^{\infty} \rho_{i}\left\langle\beta_{i}, v\right\rangle\left\langle\beta_{i}, z\right\rangle . \tag{4.100}
\end{equation*}
$$

Lemma 4.8.2. Assume the conditions of Theorem 4.5.2. For any natural number $i \in \mathbb{N}$, define

$$
W_{i}^{\varepsilon}(t)=\sqrt{\varepsilon} \sum_{j=0}^{t / \varepsilon-1}\left\langle z_{j}-\bar{g}, \mathbf{e}_{i}\right\rangle .
$$

Then $W_{i}^{\varepsilon}(\cdot)$ converges weakly to a real-valued Wiener process $W_{i}(\cdot)$ with covariance $t \sigma_{i}^{2}$, where

$$
\begin{equation*}
\sigma_{i}^{2}=\mathbf{E}\left[\left\langle z_{0}-\bar{g}, \mathbf{e}_{i}\right\rangle\right]^{2}+2 \sum_{j=1}^{\infty} \mathbf{E}\left\langle z_{0}-\bar{g}, \mathbf{e}_{i}\right\rangle\left\langle z_{j}-\bar{g}, \mathbf{e}_{i}\right\rangle . \tag{4.101}
\end{equation*}
$$

Proof Note that with the use of inner product in $\ell_{2},\left\{\left\langle z_{n}-\bar{g}, \mathbf{e}_{i}\right\rangle\right\}$ is a real-valued mixing sequence with mean 0 . The desired convergence follows from the functional invariance principle for mixing process; see [99, Chapter 7] (see also [25, 61]).

Lemma 4.8.3. Under the conditions of Lemma 4.8.2, for $i \neq l, E W_{i}^{\varepsilon}(t) W_{l}^{\varepsilon}(t)=0$. As a result, the limit Wiener processes $W_{i}(\cdot)$ and $W_{l}(\cdot)$ are independent.

Proof It is straightforward that

$$
\begin{aligned}
\mathbf{E} W_{i}^{\varepsilon}(t) W_{l}^{\varepsilon}(t) & =\varepsilon \mathbf{E} \sum_{k=0}^{t / \varepsilon-1} \sum_{j=0}^{t / \varepsilon-1}\left\langle z_{j}-\bar{g}, \mathbf{e}_{i}\right\rangle\left\langle z_{k}-\bar{g}, \mathbf{e}_{l}\right\rangle \\
& =\varepsilon \mathbf{E} \sum_{k=0}^{t / \varepsilon-1 t / \varepsilon-1} \sum_{j=0}^{t}\left\langle z_{j}-\bar{g}, \mathbf{e}_{i} \mathbf{e}_{l}^{\prime}\left(z_{k}-\bar{g}\right)\right\rangle \\
& =0 \text { since } \mathbf{e}_{i} \mathbf{e}_{l}^{\prime}=0 \in \mathbb{R}^{\infty \times \infty} .
\end{aligned}
$$

Since $\mathbf{E} W_{i}^{\varepsilon}(t)=0$, we conclude that $\Sigma\left(W_{i}^{\varepsilon}(t), W_{l}^{\varepsilon}(t)\right)=0$. Consequently, $\Sigma\left(W_{i}(t), W_{l}(t)\right)=0$, and as a result $W_{i}(t)$ and $W_{l}(t)$ are independent Wiener processes.

Proposition 4.8.4. Under the conditions of Lemma 4.8.2, define

$$
\begin{equation*}
W^{\varepsilon}(t)=\sqrt{\varepsilon} \sum_{j=0}^{t / \varepsilon-1}\left[z_{j}-\bar{g}\right] . \tag{4.102}
\end{equation*}
$$

Then $W^{\varepsilon}(\cdot)$ converges weakly to $W(\cdot)$ such that

$$
\begin{equation*}
W(t)=\sum_{i=0}^{\infty} W_{i}(t) \mathbf{e}_{i} \tag{4.103}
\end{equation*}
$$

and the covariance operator is given by

$$
\begin{equation*}
\mathbf{E}\langle W(t), v\rangle\langle W(t), z\rangle=t\langle z, \Gamma v\rangle=t \sum_{i=0}^{\infty} \sigma_{i}^{2}\left\langle\mathbf{e}_{i}, v\right\rangle\left\langle\mathbf{e}_{i}, z\right\rangle \text { for } v, z \in \ell_{2}, \tag{4.104}
\end{equation*}
$$

where $\sigma_{i}^{2}$ is defined in (4.101).

Proof In view of the definition of (4.102), for any $\delta>0, t>0,0<s \leq \delta$, with $\mathbf{E}_{t}^{\varepsilon}$ denotes the conditional expectation with respect to $\mathscr{F}_{t}^{\varepsilon}$, using the mixing properties, we can show that

$$
\lim _{\delta \rightarrow 0} \limsup _{\varepsilon \rightarrow 0}\left[\sup _{0 \leq \delta \leq s} \mathbf{E}_{t}^{\varepsilon}\left\langle W^{\varepsilon}(t+s)-W^{\varepsilon}(t), W^{\varepsilon}(t+s)-W^{\varepsilon}(t)\right\rangle\right]=0 .
$$

Thus $W^{\varepsilon}(\cdot)$ is tight in $D\left([0, \infty) ; \ell_{2}\right)$. We can extract any weakly convergent subsequence and denote the limit by $W(\cdot)$. We next characterize its limit.

Again, using (4.102)

$$
W^{\varepsilon}(t)=\sum_{i=0}^{\infty} W_{i}^{\varepsilon}(t) \mathbf{e}_{i}=\sqrt{\varepsilon} \sum_{i=0}^{\infty} \sum_{j=0}^{t / \varepsilon-1}\left\langle z_{j}-\bar{g}, \mathbf{e}_{i}\right\rangle \mathbf{e}_{i} .
$$

Therefore, for each $l \in \mathbb{N}$,

$$
\mathbf{E}\left[\left\langle W^{\varepsilon}(t), \mathbf{e}_{l}\right\rangle\right]^{2}=\mathbf{E}\left\|W_{l}^{\varepsilon}(t)\right\|^{2}=t \sigma_{l}^{2} .
$$

By virtue of exponential decay property of $\bar{g}^{i} \propto i^{-\beta}, \sum_{l=0}^{\infty} \sigma_{l}^{2}<\infty$. By Lemma 4.8.2, $W_{i}^{\varepsilon}(\cdot)$ converges weakly to $W_{i}(\cdot)$. By virtue of Lemma 4.8.3, $W_{i}(\cdot)$ are independent Wiener processes. In view of the definition of Wiener process on $\ell_{2}$, we conclude that $W^{\varepsilon}(\cdot)$ converges weakly to $W(\cdot)$ such that (4.103) holds. In addition, the structure of the covariance operator (4.104) is obtained.

We proceed to obtain the desired weak convergence of $v^{\varepsilon}(\cdot)$. Since the stochastic differential equation (4.25) is linear, there is a unique solution for each initial condition. The rest of the proof is similar to the finite dimensional counter part with necessary modifications similar to that of the proof of Theorem 4.5.2.

### 4.8.10 Proof of Theorem 4.5.4

Proof. Before proceeding to the main proof, we first state a preliminary result. The proofs of the assertions below can be found in [156, Theorem 3.6 and Theorem 4.3, respectively] and are thus omitted.

Lemma 4.8.5. Under Assumption 4.2.1, the following claims hold:
(a) Denote $p_{n}^{\rho}=\left[P\left(\theta_{n}^{\rho}=1\right), \ldots, P\left(\theta_{n}^{\rho}=M\right)\right]$ and the $n$-step transition probability by $\left(A^{\rho}\right)^{n}$ with $A^{\rho}$ given in (4.2) with $\rho=\varepsilon^{2}$. Then

$$
\begin{align*}
& p_{n}^{\rho}=p(\rho n)+O\left(\rho+\rho^{-k_{0} t / \rho}\right) \\
& \left(A^{\rho}\right)^{n-n_{0}}=\Xi\left(\rho n, \rho n_{0}\right)+O\left(\rho+e^{-k_{0}\left(t-t_{0}\right) / \rho}\right) \tag{4.105}
\end{align*}
$$

where $p(t) \in \mathbb{R}^{1 \times M}$ and $\Xi\left(t, t_{0}\right) \in \mathbb{R}^{M \times M}$ are the continuous-time probability vector and transition matrix satisfying

$$
\begin{align*}
& \frac{d p(t)}{d t}=p(t) Q, p(0)=p_{0},  \tag{4.106}\\
& \frac{d \Xi\left(t, t_{0}\right)}{d t}=\Xi\left(t, t_{0}\right) Q, \quad \Xi\left(t_{0}, t_{0}\right)=I,
\end{align*}
$$

with $t_{0}=\rho n_{0}$ and $t=\rho n$.
(b) $\theta^{\rho}(\cdot)$ converges weakly to $\theta(\cdot)$, a continuous-time Markov chain generated by $Q$.

To analyze the algorithm, the techniques developed in the proof of Theorem 4.5.2 are used along with the ideas and methods developed in [155]. The developments are similar in the approach and the results, but are more complex due to the added switching process. For example, with modifications, Step 1 in the proof of Theorem 4.5 .2 can still be carried out. Also Step 2 can be proved. So the sequence $\left\{\widehat{g}^{\varepsilon}(\cdot)\right\}$ is tight.

To characterize the limit, we still use martingale averaging techniques. We shall only highlight the main difference here. In carrying out the analysis similar to that of Step 3 in the proof of Theorem 4.5.2, we will encounter the following term

$$
\begin{align*}
& \mathbf{E} h\left(\widehat{g}^{\varepsilon}\left(t_{l_{1}}\right): l_{1} \leq \kappa\right)\left[\sum_{l m_{\varepsilon}=t / \varepsilon}^{(t+s) / \varepsilon-1} \Delta_{\varepsilon}\left\langle\nabla f\left(\widehat{g}_{l m_{\varepsilon}}\right), \frac{1}{m_{\varepsilon}} \sum_{j=l m_{\varepsilon}}^{l m_{\varepsilon}+m_{\varepsilon}-1} Y_{j}\left(\theta_{j}\right)\right\rangle\right] \\
& =\mathbf{E} h\left(\widehat{g}^{\varepsilon}\left(t_{l_{1}}\right): l_{1} \leq \kappa\right)\left[\sum_{l m_{\varepsilon}=t / \varepsilon}^{(t+s) / \varepsilon-1} \Delta_{\varepsilon}\left\langle\nabla f\left(\widehat{g}_{l m_{\varepsilon}}\right), \frac{1}{m_{\varepsilon}} \sum_{j=l m_{\varepsilon}}^{l m_{\varepsilon}+m_{\varepsilon}-1} \mathbf{E}_{l m_{\varepsilon}} Y_{j}\left(\theta_{j}\right)\right\rangle\right]  \tag{4.107}\\
& =\mathbf{E} h\left(\widehat{g}^{\varepsilon}\left(t_{l_{1}}\right): l_{1} \leq \kappa\right)\left[\sum_{l m_{\varepsilon}=t / \varepsilon}^{(t+s) / \varepsilon-1} \Delta_{\varepsilon}\left\langle\nabla f\left(\widehat{g}_{l m_{\varepsilon}}\right), \frac{1}{m_{\varepsilon}} \sum_{\theta=1}^{M} \sum_{j=l m_{\varepsilon}}^{l m_{\varepsilon}+m_{\varepsilon}-1} \mathbf{E}_{l m_{\varepsilon}} Y_{j}(\theta) I_{\left\{\theta_{j}=\theta\right\}}\right\rangle\right] .
\end{align*}
$$

Since $Y_{j}(\theta)$ and $\theta_{j}$ are independent, we have

$$
\begin{align*}
& \frac{1}{m_{\varepsilon}} \sum_{\theta=1}^{M} \sum_{j=l m_{\varepsilon}}^{l m_{\varepsilon}+m_{\varepsilon}-1} \mathbf{E}_{l m_{\varepsilon}} Y_{j}(\theta) I_{\left\{\theta_{j}=\theta\right\}}  \tag{4.108}\\
& \quad=\frac{1}{m_{\varepsilon}} \sum_{l_{0}=1}^{M} \sum_{\theta=1}^{M} \sum_{j=l m_{\varepsilon}}^{l m_{\varepsilon}+m_{\varepsilon}-1} \mathbf{E}_{l m_{\varepsilon}} Y_{j}(\theta) P\left(\theta_{j}=\theta \mid \theta_{l m_{\varepsilon}}=\imath_{0}\right) I_{\left\{\theta_{l m_{\varepsilon}}=l_{0}\right\}}
\end{align*}
$$

For each $\theta \in \mathscr{M}$, the averaging of $Y_{j}(\theta)$ can be carried out as in Case 1 . We concentrate on the term involving Markov chain. By virtue of Lemma 4.8.5, noting $\rho=\varepsilon^{2}$ and using (4.105), we have

$$
\left[A^{\rho}\right]^{j-l m_{\varepsilon}}=\Xi\left(\varepsilon^{2} j, \varepsilon^{2} \operatorname{lm}_{\varepsilon}\right)+O\left(\varepsilon^{2}+e^{-k_{0}\left(\varepsilon^{2} j-\varepsilon^{2} l m_{\varepsilon}\right) / \varepsilon^{2}}\right)
$$

Because we are working with (4.28) and the stepsize is $\varepsilon$. In the interval $\left[l \Delta_{\varepsilon}, l \Delta_{\varepsilon}+\Delta_{\varepsilon}\right)$ with $\Delta_{\varepsilon}=\varepsilon m_{\varepsilon}$, it is readily seen that $\Xi\left(\varepsilon^{2} j, \varepsilon^{2} l m_{\varepsilon}\right) \rightarrow \Xi(0,0)=I$ as $\varepsilon \rightarrow 0$. As a result,

$$
P\left(\theta_{j}=\theta \mid \theta_{l m_{\varepsilon}}=\imath_{0}\right)+o_{\varepsilon}(1)=\delta_{l_{0}, \theta}+o_{\varepsilon}(1)=\left\{\begin{array}{ll}
1, & \text { if } \imath_{0}=\theta \\
0, & \text { otherwise }
\end{array}+o_{\mathcal{\varepsilon}}(1)\right.
$$

where $o(1) \rightarrow 0$ as $o_{\varepsilon}(1) \rightarrow 0$ as $\varepsilon \rightarrow 0$. Putting the above estimates in (4.108), we obtain the limit in probability of

$$
\frac{1}{m_{\varepsilon}} \sum_{j=l m_{\varepsilon}}^{l m_{\varepsilon}+m_{\varepsilon}-1} \mathbf{E}_{l m_{\varepsilon}} Y_{j}(\theta) P\left(\theta_{j}=\theta \mid \theta_{l m_{\varepsilon}}=\boldsymbol{l}_{0}\right) I_{\left\{\theta_{l m_{\varepsilon}}=l_{0}\right\}}
$$

is the same as that of

$$
\sum_{i=1}^{\infty} \mathbf{e}_{i} \theta_{i}(\theta) \delta_{l_{0} \theta} I_{\left\{\theta^{\varepsilon^{2}}\left(\varepsilon^{2} l m_{\varepsilon}\right)=\theta_{0}\right\}}
$$

This further leads to that as $\varepsilon \rightarrow 0$,

$$
\begin{aligned}
& \mathbf{E} h\left(\widehat{g}^{\varepsilon}\left(t_{l_{1}}\right): l_{1} \leq \kappa\right)\left[\sum_{l m_{\varepsilon}=t / \varepsilon}^{(t+s) / \varepsilon-1} \Delta_{\varepsilon}\left\langle\nabla f\left(\widehat{g}_{l m_{\varepsilon}}\right), \frac{1}{m_{\varepsilon}} \sum_{j=l m_{\varepsilon}}^{l m_{\varepsilon}+m_{\varepsilon}-1} Y_{j}\left(\theta_{j}\right)\right\rangle\right] \\
& \rightarrow \mathbf{E} h\left(\widehat{g}\left(t_{l_{1}}\right): l_{1} \leq \kappa\right)\left[\int_{t}^{t+s}\left\langle\nabla f(\widehat{g}(\tau)), \theta_{i}(\theta) P(\theta(0)=\theta)\right\rangle d \tau\right] \\
& =\mathbf{E} h\left(\widehat{g}\left(t_{l_{1}}\right): l_{1} \leq \kappa\right)\left[\int_{t}^{t+s}\left\langle\nabla f(\widehat{g}(\tau)), \theta_{i}(\theta) \mathbf{e}_{i} p_{\theta}\right\rangle d \tau\right]
\end{aligned}
$$

## 5

## Conclusions

The unifying theme of this thesis was to devise a set of theories and methods for statistical signal processing on graphs (possibly random) which involves multi-agent Bayesian estimation, social learning, stochastic approximation algorithms and adaptive filtering, and dynamics of random graphs to understand the effects of the interactions among agents on the estimation/tracking problem. Part I of this dissertation was devoted to mis-information management problem in multi-agent state estimation over social networks. Part II of this thesis deals with tracking a time-varying degree distribution of a dynamic social network using noisy observations. This chapter concludes this work and presents a summary of findings along with some some directions for future research and development.

### 5.1 Summary of Findings in Part I

Over the last decade there has been a growing interest in social networks which facilitate our day-to-day activities such as our decision-makings, social communications, and sharing news or stories. Many of these activities involve estimation, learning or decision making using social networks such as rating and review systems, micro-blogging platforms, and online social networks. In these estimation problems, structure of the underlying social network imposes a communication constraint and dictates who talks to whom. First part of this thesis was motivated by such social networks that comprises of a set of agents (social sensors) that seek to estimate an underlying state of nature interactively. Part I dealt with mis-information management problem in two different information exchange protocols:

- Chapter 2 considered an information exchange protocol where agents broadcast their (private) beliefs over the network. In such a protocol, each agent records its (private) observations, then, it combines it with the information received from other agents to form its belief about state of nature. Finally, it transmits the updated belief over the network. Note that this is not social learning, since (private) beliefs of agents are transmitted.
- Chapter 3 used social learning to model the interactions among agents. In this model, each agent computes its private belief using the local observation and the information received from the network. Then, based on its private belief, it chooses an action from a finite set such
that a local cost function is minimized. In the social learning context considered in Chapter 3, as opposed to the one in Chapter 3, private beliefs of agents which depend on the private observations are not communicated to other agents. Instead, public beliefs of them, which can be computed directly from their actions, are broadcasted over the network.

In both of the above protocols, mis-information propagation arises as a result of the correlation introduced by the loops in the communication graph and the recursive nature of Bayesian models. In Chapter 2, we present an optimal information aggregation scheme that completely removes the mis-information associated with estimates of agents under some conditions on the topology of the communication graph. The optimal mis-information removal algorithm proposed in Chapter 2 requires knowledge of transitive closure matrix of the communication graph. For the scenarios where due to a random delay in communications among agents the transitive closure matrix is not completely known, a sub-optimal algorithm is proposed to mitigate "double counting" events which are more likely to happen.

The social learning model considered in Chapter 3 results in two interesting phenomena: (i) herding where all rational agents end up choosing the same action and (ii) mis-information propagation which produces a bias in the public belief and results in overconfidence. Inspired by online rating and review systems, Chapter 3 presented a 5 -step protocol to mimic the interactions among agents (social sensors) that aim to estimate an underlying state of nature. It then introduced a fair protocol that prevents mis-information propagation and was used as a benchmark. Using that, an algorithm is invented for the administrator of the rating system to deploy such that it maintains fair ratings. The results of a psychology experiment on a group which has been conducted by our colleagues at the University of British Columbia on a group of undergraduate students to study the learning behavior of humans in a society. The experiment showed that the interactions of the agents can be modeled using a social learning model. We further showed that, as a result of the information exchange protocol (between individuals within a group) and the recursive nature of decision making process, data incest arises in a large fraction of trials in the experiment.

## Tools

Graph theory first started 250 years ago with a paper written by Leonhard Eüler on the Seven Bridges of Knigsberg published in 1763 [23]. Since then, it has became a powerful tool to model several networks. Each vertex denotes an agent (or group of individuals) in the social network and each edge depicts a relationship between different agents in the social network. In this work, graph theoretic tools and definitions are used to model the flow of information through the network. Also, the necessary and sufficient condition for complete mis-information removal is presented in terms of adjacency and transitive closure matrices of the underlying communication graph.

Social learning is another mathematical abstraction which is used in this work to model the
interactions among agents (social sensors) in social networks.

### 5.2 Summary of Findings in Part II

The second part of this thesis was motivated by the importance of the degree distribution in analysis of social networks. The interaction between nodes in dynamic social networks is not always fixed and may evolve over time. An example of such time-varying dynamics is the seasonal variations in friendships among college students. Chapter 4 considered social networks where dynamics of the underlying graph is evolving according to realization of a Markov chain. The Markov-modulated random graph generated by Algorithm 4.5 mimics such networks where the dynamics (the connection/deletion probabilities $p, q$ ) depend on the state of nature and evolve over time. Algorithm 4.5 models these time variations as a finite state Markov chain $\left\{\theta_{n}\right\}$. This model forms our basis for analysis of social networks.

Markov-modulated duplication-deletion random graphs are analyzed in terms of degree distribution. When the size of graph is fixed $(r=0)$ and $\rho$ is small, the expected degree distribution of the Markov-modulated duplication-deletion random graph can be uniquely computed from (4.6) for each state of the underlying Markov chain. This result allows us to express the structure of network (degree distribution) in terms of the dynamics of the model. We also showed that, when the size of the graph is fixed and there is no Markovian dynamics, the random graph generated according to Algorithm 4.5 satisfies a power law with exponent computed from (4.11). The importance of this result is that a single parameter (power law exponent) characterizes the structure of a possibly very large dynamic network.

Moreover, a stochastic approximation algorithm is presented to estimate the empirical degree distribution of the finite duplication-deletion random graph using noisy observations. The proposed stochastic approximation algorithm (4.14) does not assume any knowledge of the Markovmodulated dynamics of the graph (state of nature). Since the expected degree distribution can be uniquely computed for each state of underlying Markov chain, a social sensor can be designed based on (4.14) to track the state of nature using the noisy observations of nodes' degrees. These noisy observations can be samples of the degree sequence of each node; that is, some nodes are randomly chosen and inquired about the number of connections that they have. Using perturbed Lyapunov function, we showed that the tracking error of the stochastic approximation algorithm is small and bounded.

Then, in Chapter 4 a Hilbert-space-valued stochastic approximation algorithm is presented to track the expected degree distribution of the infinite duplication-deletion random graph without Markovian dynamics. The asymptotic behavior of such an algorithm is analyzed in terms of the power law degree distribution. Finally, we extended the analysis to a Hilbert-space-valued stochastic approximation algorithm that aims to track a Markov-modulated probability mass function with
denumerable support. Using weak convergence methods, it was shown that the estimates obtained via such an algorithm converge weakly to the solution of an ordinary differential equation. It was also shown that the interpolated sequence of scaled tracking error converges weakly to the solution of a stochastic differential equation.

## Tools

We borrowed techniques from graph theory to perform the degree distribution analysis provided in Chapter 4. Such techniques were previously used in analysis of complex networks in the social and economic networks literature, see [41, 80]. Adaptive filtering, stochastic approximation, and Markov-switched systems are among the tools which are employed in this chapter in order to estimate the expected degree distribution and consequently the state of underlying Markov chain. Weak convergence analysis and functional central limit theorem are used as mathematical abstractions to analyze the performance of such tracking algorithms.

### 5.3 Directions for Future Research and Development

There is clearly much work to be done in the area of signal processing on complex (and possibly random) networks to understand the behavior of agents in social networks. In this section we present some of the immediate extensions of the work presented in this dissertation.

## Mis-information removal algorithms in Bayesian quickest-time change detection:

In change detection problem, the objective is to detect a random change time by optimizing the trade-off between number of observations and delay penalty [ 124,134$]$. This problem is very similar to the sensing problems that have been considered in the first part of this dissertation and involve interactive sensing with the goal of detecting a random change in state of nature. Multi-agent Bayesian change detection involves a set of agents (sensors) where each agent estimates an underlying state and then, using Bayesian models, updates the posterior distribution of the change. Then it sends this updated posterior distribution (or a myopic action obtained based on a local cost optimization [87]) over the network. This process repeats until a global decision maker detects a change and all agents stops making observations. In other words, using all the local information (decisions or distributions), the goal in the quickest-time change detection is to detect a change and make a global decision. Because of the recursive nature of the Bayesian estimators and the information exchange protocol, mis-information propagation can arise in such system. One immediate extension of the work presented in this dissertation is to investigate the effect of mis-information propagation in such interactive sensing scenarios and possibly to devise a mis-information removal algorithm
for each agent to employ, such that the decision of the global decision maker is not affected by the mis-information propagation.

## Analyzing spread of contagious disease through network

Diffusion of information and disease through society has been studied extensively in the literature of social network analysis, see [80, 107, 108, 122, 146]. One of the models that used to model the diffusion of information in the social network is called the Susceptible-Infected-Susceptible (SIS) model [122]. Consider a social network where agents interacts with other nodes that are dictated by the structure of the network. Each agent in such a social network can be in two states: (i) infected, or (ii) not infected but susceptible to becoming infected. In SIS model, infected nodes can recover and become susceptible again. In the model considered in [107, 108], the assumption is that the degree of nodes remain fixed. One extension or the work presented in the second part of this thesis, is to study the diffusion of information in dynamic graphs (where the graph is not fixed and is evolving according to Procedure 4.5 presented in Chapter 4). Investigating contagion in such dynamic networks is non-standard in two ways: First, the spread of disease ${ }^{44}$ is dynamic and depends on the number and the distribution of infected nodes over the network. Second, the structure of the underlying graph (which dictates who talks to whom) is evolving over time. Analyzing diffusion process in such networks (both in stationary state and in transient phase) can be a possible avenue for further research and development.

## Mis-information removal algorithms in non-Bayesian models

In our work, we addressed mis-information management problem in two scenarios: (i) constrained filtering and (ii) social learning over social networks. In both of these scenarios, Bayesian models for information aggregation have been used. There exists a large body of works in the literature that consider Bayesian models to formulate the learning behavior of humans [1, 13, 16, 24, 136]. On the other hand, a body of social learning literature focuses on non-Bayesian social learning models, see $[3,14,15,56,81]$. In these models, agents use a simple rule-of-thumb to update their beliefs from their private information about state of nature and those received from other agents. Similar to the learning problems considered in Part I of this dissertation, as a result of the recursive nature of the information aggregation schemes and possible loops in communication graph, mis-information propagation may arise in such non-Bayesian learnings over social networks. Mis-information management in learning with non-Bayesian models over social networks can be a possible direction for further research and development.

[^33]
## Mis-information propagation mitigation for specific network structures

In the proposed mis-information algorithm presented in this thesis, we used a combination of the previous estimates in order to completely mitigate data incest. However, the optimality of this approach demands a topological constraint to hold. An extension to the work presented in this dissertation is to investigate mis-information management problem in graphs with specific structures that do not necessarily satisfy such a constraint. An alternative approach is to "censor" the estimates of some nodes that are most likely to be polluted with mis-information. This can be viewed as deliberately cutting some edges in order to lower the risk of mis-information propagation. Cutting these edges can be useful for some specific graph structures (like bi-bipartite graphs) specially when the topological constraints is not satisfied and, thus, optimal mis-information removal is not possible.

## Bibliography

[1] D. Acemoglu, M. A. Dahleh, I. Lobel, and A. Ozdaglar. Bayesian learning in social networks. The Review of Economic Studies, 78(4):1201-1236, 2011.
[2] D. Acemoglu and A. Ozdaglar. Opinion dynamics and learning in social networks. Dynamic Games and Applications, 1(1):3-49, Mar. 2011.
[3] D. Acemoglu, A. Ozdaglar, and A. ParandehGheibi. Spread of (mis) information in social networks. Games and Economic Behavior, 70(2):194-227, 2010.
[4] L. Adamic and E. Adar. How to search a social network. Social Networks, 27(3):187-203, Jul. 2005.
[5] T. Aeppel. 50 million users: The making of an angry birds internet meme. Wall Street Journal, Mar. 2015.
[6] C. C. Aggarwal and T. Abdelzaher. Integrating sensors and social networks. Social Network Data Analytics, pages 397-412, 2011.
[7] D. Agrawal, C. Budak, and A. El Abbadi. Information diffusion in social networks: Observing and affecting what society cares about. In Proceedings of the 20th ACM International Conference on Information and Knowledge Management, CIKM '11, pages 2609-2610, New York, NY, USA, 2011. ACM.
[8] Y. Y. Ahn, S. Han, H. Kwak, S. Moon, and H. Jeong. Analysis of topological characteristics of huge online social networking services. In Proceedings of 16th International Conference on World Wide Web, pages 835-844, Banff, AB, Canada, May 2007.
[9] M. Albakour, C. Macdonald, and I. Ounis. Identifying local events by using microblogs as social sensors. In Proceedings of 10th Conference on Open Research Areas in Information Retrieval, pages 173-180, Lisbon, Portugal, May 2013.
[10] R. Albert and A.-L. Barabási. Statistical mechanics of complex networks. Reviews of Modern Physics, 74(1):47-97, Jan. 2002.
[11] S. Asur and B. A. Huberman. Predicting the future with social media. In 2010 IEEE/WIC/ACM International Conference on Web Intelligence and Intelligent Agent Technology, volume 1, pages 492-499, Toronto, ON, Canada, 2010.
[12] R. J. Aumann. Agreeing to disagree. The Annals of Statistics, 4(6):1236-1239, Nov. 1976.
[13] R. J. Aumann. Correlated equilibrium as an expression of bayesian rationality. Econometrica, 55(1):1-18, Jan. 1987.
[14] V. Bala and S. Goyal. Learning from neighbours. The review of economic studies, 65(3):595621, 1998.
[15] V. Bala and S. Goyal. Conformism and diversity under social learning. Economic theory, 17(1):101-120, 2001.
[16] A. Banerjee. A simple model of herd behavior. Quaterly Journal of Economics, 107(3):797817, Aug. 1992.
[17] A.-L. Barabási and A. Reka. Emergence of scaling in random networks. Science, 286(5439):509, 1999.
[18] T. Bayes. An essay toward solving a problem in the doctrine of chances. Philosophical Transactions of the Royal Society of London, 53, 1764.
[19] A. Benveniste, M. Métivier, and P. P. Priouret. Adaptive algorithms and stochastic approximations. Springer-Verlag, 1990.
[20] A. Benveniste, M. Métivier, and P. Priouret. Adaptive Algorithms and Stochastic Approximations. Springer, 2012.
[21] J. M. Bernardo and A. Smith. Bayesian theory, volume 405. John Wiley \& Sons, 2009.
[22] P. Biernacki and D. Waldorf. Snowball sampling: Problems and techniques of chain referral sampling. Sociological Methods \& Research, 10(2):141-163, Nov. 1981.
[23] N. L. Biggs, E. K. Lloyd, and R. J. Wilson. Graph Theory. Oxford University Press, 1986.
[24] S. Bikchandani, D. Hirshleifer, and I. Welch. A theory of fads, fashion, custom, and cultural change as information cascades. Journal of Political Economy, 100(5):992-1026, Oct. 1992.
[25] P. Billingsley. Convergence of probability measures, volume 493. John Wiley \& Sons, 2009.
[26] J. Bollen, H. Mao, and X. Zeng. Twitter mood predicts the stock market. Journal of Computational Science, 2(1):1-8, Mar. 2011.
[27] B. Bollobás, O. Riordan, J. Spencer, and G. Tusnády. The degree sequence of a scale-free random graph process. Random Structures and Algorithms, 18(3):279-290, 2001.
[28] P. Bonhard and M. A. Sasse. knowing me, knowing you using profiles and social networking to improve recommender systems. BT Technology Journal, 24(3):84-98, 2006.
[29] T. Brehard and V. Krishnamurthy. Optimal data incest removal in Bayesian decentralized estimation over a sensor network. In Proceedings of IEEE International Conference on Acoustics, Speech and Signal Processing, pages 173-176, Honolulu, Hawaii, Apr. 2007.
[30] Ceren Budak, Divyakant Agrawal, and Amr El Abbadi. Limiting the spread of misinformation in social networks. In Proceedings of the 20th International Conference on World Wide Web, WWW '11, pages 665-674. ACM, 2011.
[31] D. Budak and A. El Abbadi. Information diffusion in social networks: Observing and influencing societal interests. Proceedings of the VLDB Endowment, 4(12), 2011.
[32] J. Burke, D. Estrin, M. Hansen, A. Parker, N. Ramanathan, S. Reddy, and M. B. Srivastava. Participatory sensing. In Workshop on World-Sensor-Web: Mobile Device Centric Sensor Networks and Applications, Boulder, CO, 2006.
[33] D. S. Callaway, M. E. J. Newman, S. H. Strogatz, and D. J. Watts. Network robustness and fragility: Percolation on random graphs. Physical Review Letters, 85(25):5468-5471, Dec. 2000.
[34] A. T. Campbell, S. B. Eisenman, N. D. Lane, E. Miluzzo, R. Peterson, H. Lu, X. Hong, X. Zheng, M. Musolesi, K. Fodor, and G. Ahn. The rise of people-centric sensing. IEEE Internet Computing, 12(4):12-21, 2008.
[35] C. Chamley. Rational Herds: Economic Models of Social Learning. Cambridge University Press, 2004.
[36] C. Chamley, A. Scaglione, and L. Li. Models for the diffusion of beliefs in social networks: An overview. IEEE Signal Processing Magazine, 30(3):16-29, May 2013.
[37] C. P. Chamley. Rational Herds: Economic Models of Social Learning. Cambridge University Press, 2003.
[38] Z. Cheng, J. Caverlee, and K. Lee. You are where you tweet: A content-based approach to geo-locating twitter users. In Proceedings of 19th ACM International Conference on Information and Knowledge Management, pages 759-768, Toronto, ON, Canada, Oct. 2010.
[39] W. H. Chin, D. B. Ward, and A.G. Constantinides. Semi-blind mimo channel tracking using auxiliary particle filtering. In Global Telecommunications Conference, 2002. GLOBECOM'02. IEEE, volume 1, pages 322-325. IEEE, 2002.
[40] N. A. Christakis and J. H. Fowler. Social network sensors for early detection of contagious outbreaks. PLoS ONE, 5(9), Sep. 2010.
[41] F. Chung and L. Lu. Complex Graphs and Networks. Conference Board of the Mathematical Sciences, National Science Foundation (U.S.), 2006.
[42] F. Chung, L. Lu, T. G. Dewey, and D. G. Galas. Duplication models for biological networks. Journal of Computational Biology, 10(5):677-687, Oct. 2003.
[43] T. Conley and C. Udry. Social learning through networks: The adoption of new agricultural technologies in Ghana. American Journal of Agricultural Economics, 83(3):668-673, Aug. 2001.
[44] C. Cooper and A. Frieze. A general model of web graphs. Random Structures and Algorithms, 22(3):311-335, 2003.
[45] R. Corten. Composition and structure of a large online social network in the netherlands. PLoS ONE, 7(4), Apr. 2012.
[46] R. F. Curtain and A. J. Pritchard. Infinite dimensional linear systems theory. Lecture notes in control and information sciences, 8, 1978.
[47] B. Debatin, J. P. Lovejoy, A. K. Horn, and B. N. Hughes. Facebook and online privacy: Attitudes, behaviors, and unintended consequences. Journal of Computer-Mediated Communication, 15(1):83-108, 2009.
[48] M. H. DeGroot. Reaching a consensus. Journal of the American Statistical Association, 69(345):118-121, 1974.
[49] P. Dodin and V. Nimier. Distributed tracking systems and their optimal inference topology. In Proceedings of 5th International Conference on Information Fusion, volume 1, pages 585592, Annapolis, MD, Jul. 2002.
[50] X. G. Doukopoulos and G. V. Moustakides. Adaptive power techniques for blind channel estimation in cdma systems. IEEE Transactions on Signal Processing, 53(3):1110-1120, Mar. 2005.
[51] X. G. Doukopoulos and G. V. Moustakides. Blind adaptive channel estimation in OFDM systems. IEEE Transactions on Wireless Communications, 5(7):1716-1725, Jul. 2006.
[52] X. G. Doukopoulos and G. V. Moustakides. Fast and stable subspace tracking. IEEE Transactions on Signal Processing, 56(4):1452-1465, Apr. 2008.
[53] R. Durrett. Random Graph Dynamics. Cambridge Series on Statistical and Probabilistic Mathematics. Cambridge University Press, 2007.
[54] N. Eagle and A. Pentland. Reality mining: Sensing complex social systems. Personal and Ubiquitous Computing, 10(4):255-268, May 2006.
[55] H. Ebel, L. I. Mielsch, and S. Bornholdt. Scale-free topology of e-mail networks. Physical Review E, 66, 2002.
[56] G. Ellison and D. Fudenberg. Rules of thumb for social learning. Journal of Political Economy, 101(4):612-643, Aug. 1993.
[57] G. Ellison and D. Fudenberg. Word-of-mouth communication and social learning. The Quarterly Journal of Economics, 110(1):93-125, 1995.
[58] Y. Ephraim. A bayesian estimation approach for speech enhancement using hidden markov models. IEEE Transactions on Signal Processing, 40(4):725-735, 1992.
[59] P. Erdös and T. Gallai. Graphs with given degrees of vertices. Mat Lapok, 11:264-274, 1960.
[60] P. Erdös and A. Rényi. On the evolution of random graphs. Publ. Math. Inst. Hung. Acad. Sci, 5:17-61, 1960.
[61] S. N. Ethier and T. G. Kurtz. Markov processes: characterization and convergence, volume 282. John Wiley \& Sons, 2009.
[62] S. Eubank, H. Guclu, V. S. Anil Kumar, M. V. Marathe, A. Srinivasan, Z. Toroczkai, and N. Wang. Modelling disease outbreaks in realistic urban social networks. Nature, 429:180184, May 2004.
[63] R. Evans, V. Krishnamurthy, G. Nair, and L. Sciacca. Networked sensor management and data rate control for tracking maneuvering targets. IEEE Transactions on Signal Processing, 53(6):1979-1991, Jun. 2005.
[64] D. Gale and S. Kariv. Bayesian learning in social networks. Games and Economic Behavior, 45(2):329-346, Nov. 2003.
[65] A. G.Dimakis, S. Kar, J. M. F. Moura, M. G. Rabbat, and A. Scaglione. Gossip algorithms for distributed signal processing. Proceedings of the IEEE, 98(11):1847 -1864, Nov. 2010.
[66] J. D. Geanakoplos and H. M. Polemarchakis. We can't disagree forever. Journal of Economic Theory, 28(1):192-200, Oct. 1982.
[67] M. Girvan and M. E. J. Newman. Community structure in social and biological networks. Proceedings of the National Academy of Sciences, 99(12):7821-7826, 2002.
[68] S. Goel and M. J. Salganik. Respondent-driven sampling as markov chain Monte Carlo. Statistics in Medicine, 28(17):2209-2229, Jul 2009.
[69] L. Guo, L. Ljung, and G. J. Wang. Necessary and sufficient conditions for stability of lms. IEEE Transactions on Automatic Control,, 42(6):761-770, 1997.
[70] S. L. Hakimi. On realizability of a set of integers as degrees of the vertices of a linear graph. I. Journal of the Society for Industrial and Applied Mathematics, 10(3):496-506, 1962.
[71] M. Hamdi and V. Krishnamurthy. Removal of data incest in multi-agent social learning in social networks. arXiv:1309.6687 [cs.SI], Sep. 2013.
[72] M. Hamdi, V. Krishnamurthy, and G. Yin. Tracking the empirical distribution of a Markovmodulated duplication-deletion random graph. arXiv:1303.0050 [cs.IT], Feb. 2013.
[73] V. Havel. A remark on the existence of finite graphs. Časopis Pěst Mat 80:477, 1955.
[74] D. D. Heckathorn. Respondent-driven sampling: A new approach to the study of hidden populations. Social Problems, 44(2):174-199, May 1997.
[75] Y.C. Ho and RCKA Lee. A bayesian approach to problems in stochastic estimation and control. Automatic Control, IEEE Transactions on, 9(4):333-339, 1964.
[76] P. Holme, B. J. Kim, C. N. Yoon, and S. K. Han. Attack vulnerability of complex networks. Physical Review E, 65(5), May 2002.
[77] B. Huang and T. Jebara. Maximum likelihood graph structure estimation with degree distributions. In Analyzing Graphs: Theory and Applications, NIPS Workshop, volume 14, 2008.
[78] M. B. Hurley. An information theoretic justification for covariance intersection and its generalization. In Proceedings of the 5th International Conference on Information Fusion, volume 1, pages 505-511, Annapolis, MD, Jul. 2002.
[79] B. Ifrach, C. Maglaras, and M. Scarsini. Monopoly pricing in the presence of social learning. NET Institute Working Paper No. 12-01, 2011.
[80] M. O. Jackson. Social and Economic Networks. Princeton University Press, 2008.
[81] A. Jadbabaie, P. Molavi, A. Sandroni, and A. Tahbaz-Salehi. Non-bayesian social learning. Games and Economic Behavior, 76(1):210-225, 2012.
[82] H. Jeong, S. P. Mason, A. L. Barabási, and Z. N. Oltvai. Lethality and centrality in protein networks. Nature, 411(6833):41-42, 2001.
[83] Y. Kanoria and O. Tamuz. Tractable bayesian social learning on trees. In Proceedings of IEEE International Symposium on Information Theory, pages 2721-2725, Cambridge, MA, Jul. 2012.
[84] H. Kim, Z. Toroczkai, P. L. Erdös, I. Miklós, and L. A. Székely. Degree-based graph construction. Journal of Physics A: Mathematical and Theoretical, 42(39), Oct. 2009.
[85] G. Kossinets and D. J. Watts. Empirical analysis of an evolving social network. Science, 311(5757):88-90, Jan. 2006.
[86] P. L. Krapivsky, G. J. Rodgers, and S. Redner. Degree distributions of growing networks. Physical Review Letters, 86(23):5401-5404, Jun. 2001.
[87] V. Krishnamurthy. Quickest detection POMDPs with social learning: Interaction of local and global decision makers. IEEE Transactions on Information Theory, 58(8):5563-5587, 2012.
[88] V. Krishnamurthy, O. Namvar Gharehshiran, and M. Hamdi. Interactive sensing and decision making in social networks. Foundations and Trends in Signal Processing, 7(1-2):1-196, 2013.
[89] V. Krishnamurthy and M. Hamdi. Data fusion and mis-information removal in social networks. In Proceedings of 15th International Conference on Information Fusion, pages 11421149, Singapore, Jul. 2012.
[90] V. Krishnamurthy and M. Hamdi. Mis-information removal in social networks: Constrained estimation on dynamic directed acyclic graphs. IEEE Journal of Selected Topics in Signal Processing, 7(2):333-346, Apr. 2013.
[91] V. Krishnamurthy and H. V. Poor. Social learning and Bayesian games in multiagent signal processing: How do local and global decision makers interact? IEEE Signal Processing Magazine, 30(3):43-57, May 2013.
[92] V. Krishnamurthy and H. V. Poor. A tutorial on interactive sensing in social networks. IEEE Transactions on Computational Social Systems, 1(1), 2014.
[93] V. Krishnamurthy and T. Ryden. Consistent estimation of linear and non-linear autoregressive models with Markov regime. Journal of Time Series Analysis, 19(3):291-307, May 1998.
[94] V. Krishnamurthy, K. Topely, and G. Yin. Consensus formation in a two-time-scale markovian system. SIAM Journal on Multiscale Modeling and Simulation, 7(4):1898-1927, 2009.
[95] Vikram Krishnamurthy and Sujay Bhatt. Sequential detection of market shocks using riskaverse agent based models. arXiv preprint arXiv:1511.01965, 2015.
[96] Vikram Krishnamurthy and William Hoiles. Online reputation and polling systems: Data incest, social learning, and revealed preferences. Computational Social Systems, IEEE Transactions on, 1(3):164-179, 2014.
[97] H. J. Kushner. Approximation and Weak Convergence Methods for Random Processes, with Applications to Stochastic Systems Theory. MIT Press, Cambridge, MA, 1984.
[98] H. J. Kushner and A. Shwartz. Stochastic approximation in hilbert space: Identification and optimization of linear continuous parameter systems. SIAM journal on control and optimization, 23(5):774-793, 1985.
[99] H. J. Kushner and G. Yin. Stochastic Approximation and Recursive Algorithms and Applications, volume 37 of Stochastic Modeling and Applied Probability. Springer-Verlag, 2nd edition, 2003.
[100] P. J. Lamberson. Social learning in social networks. The BE Journal of Theoretical Economics, 10(1), 2010.
[101] R. Lee and K. Sumiya. Measuring geographical regularities of crowd behaviors for twitterbased geo-social event detection. In Proceedings of 2nd ACM SIGSPATIAL International Workshop on Location Based Social Networks, pages 1-10, San Jose, CA, Nov. 2010.
[102] S. H. Lee, P.-J. Kim, and H. Jeong. Statistical properties of sampled networks. Physical Review E, 73(1), Jan. 2006.
[103] J. Leskovec. SNAP library. http://snap.stanford.edu/data/index.html.
[104] E. Lieberman, C. Hauert, and M. A. Nowak. Evolutionary dynamics on graphs. Nature, 433(7023):312-316, Jan. 2005.
[105] T. T. Lin and S. S. Yau. Bayesian approach to the optimization of adaptive systems. Systems Science and Cybernetics, IEEE Transactions on, 3(2):77-85, 1967.
[106] C. Lingji, P. O. Arambel, and R. K. Mehra. Estimation under unknown correlation: covariance intersection revisited. IEEE Transactions on Automatic Control, 47(11):1879-1882, Nov. 2002.
[107] D. López-Pintado. Contagion and coordination in random networks. International Journal of Game Theory, 34(3):371-381, Oct. 2006.
[108] D. López-Pintado. Diffusion in complex social networks. Games and Economic Behavior, 62(2):573-590, Mar. 2008.
[109] M. Luca. Reviews, reputation, and revenue: The case of Yelp.com, Technical Report 12-016. Harvard Business School, 2011.
[110] R. Maben, G. V. Moustakides, and J. S. Baras. Adaptive sampling for linear state estimation. SIAM Journal on Control and Optimization, 50(2):672-702, 2012.
[111] G. V. Moustakides. Exponential convergence of products of random matrices: Application to adaptive algorithms. International Journal of Adaptive Control and Signal Processing, 12(7):579-597, 1998.
[112] K. P. Murphy. Machine learning: a probabilistic perspective. MIT press, 2012.
[113] K. P. Murphy, Y. Weiss, and M. I. Jordan. Loopy belief propagation for approximate inference: an empirical study. In Proceedings of 15th Conference on Uncertainty in Artificial Intelligence, pages 467-475, 1999.
[114] M. E. J. Newman. The structure of scientific collaboration networks. Proceedings of the National Academy of Sciences, 98(2):404-409, 2001.
[115] M. E. J. Newman. Assortative mixing in networks. Physical Review Letters, 89, Oct. 2002.
[116] M. E. J. Newman. The structure and function of complex networks. SIAM Review, 45(2):167256, 2003.
[117] M. E. J. Newman. Power laws, pareto distributions and zipf's law. Contemporary physics, 46(5):323-351, 2005.
[118] M. E. J. Newman, D. J. Watts, and S. H. Strogatz. Random graph models of social networks. Proceedings of the National Academy of Sciences, 99(S1):2566-2572, 2002.
[119] N. P. Nguyen, G. Yan, M. T. Thai, and S. Eidenbenz. Containment of misinformation spread in online social networks. In Proceedings of the 4th Annual ACM Web Science Conference, WebSci '12, pages 213-222. ACM, 2012.
[120] T. D. Nielsen and F. V. Jensen. Bayesian networks and decision graphs. Information Science and Statistics. Springer, 2007.
[121] R. Pastor-Satorras, E. Smith, and R. V. Solé. Evolving protein interaction networks through gene duplication. Journal of Theoretical Biology, 222(2):199-210, May 2003.
[122] R. Pastor-Satorras and A. Vespignani. Epidemic spreading in scale-free networks. Physical Review Letters, 86(14):3200, 2001.
[123] J. Pearl. Fusion, propagation, and structuring in belief networks. Artificial Intelligence, 29(3):241-288, Sep. 1986.
[124] H. V. Poor and O. Hadjiliadis. Quickest Detection. Cambridge University Press, 2008.
[125] S. J. Press. Subjective and objective Bayesian statistics: principles, models, and applications, volume 590. John Wiley \& Sons, 2009.
[126] Naren Ramakrishnan, Benjamin J Keller, Batul J Mirza, Ananth Y Grama, and George Karypis. Privacy risks in recommender systems. IEEE Internet Computing, 5(6):54-63, 2001.
[127] J. Reason and W. Ren. Estimating the optimal adaptation gain for the lms algorithm. In Proceedings of the 32nd IEEE Conference on Decision and Control,, pages 1587-1588. IEEE, 1993.
[128] S. Reece and S. Roberts. Robust, low-bandwidth, multi-vehicle mapping. In Proceedings of 8th International Conference on Information Fusion, volume 2, Philadelphia, PA, Jul. 2005.
[129] F. Ricci, L. Rokach, and B. Shapira. Introduction to recommender systems handbook. Springer, 2011.
[130] C. Robert. The Bayesian choice: from decision-theoretic foundations to computational implementation. Springer Science \& Business Media, 2007.
[131] A. Rosi, M. Mamei, F. Zambonelli, S. Dobson, G. Stevenson, and Y. Juan. Social sensors and pervasive services: Approaches and perspectives. In Proceedings of IEEE International Conference on Pervasive Computing and Communications Workshops, pages 525-530, Seattle, WA, Mar. 2011.
[132] T. Sakaki, M. Okazaki, and Y. Matsuo. Earthquake shakes twitter users: Real-time event detection by social sensors. In Proceedings of 19th International Conference on World Wide Web, pages 851-860, Raleigh, NC, Apr. 2010.
[133] L. Scharf. Statistical signal processing, volume 98. Addison-Wesley Reading, MA, 1991.
[134] A. N. Shiryaev. On optimum methods in quickest detection problems. Theory of Probability and its Applications, 8(1):22-46, 1963.
[135] J. Shrager, T. Hogg, and B. A. Huberman. A graph-dynamic model of the power law of practice and the problem-solving fan-effect. Science, 242(4877):414-416, 1988.
[136] L. Smith and P. Sørensen. Pathological outcomes of observational learning. Econometrica, pages 371-398, 2000.
[137] V. Solo and X. Kong. Adaptive signal processing algorithms: stability and performance. Prentice-Hall information and system sciences series. Prentice Hall, 1995.
[138] S. M. Stigler. The true title of bayess essay. Statistical Science, 28(3):283-288, 2013.
[139] S. H. Strogatz. Exploring complex networks. Nature, 410(6825):268, 2001.
[140] R. Taylor. Contrained switchings in graphs. In K. L. McAvaney, editor, Combinatorial Mathematics VIII, volume 884 of Lecture Notes in Mathematics, pages 314-336. Springer Berlin Heidelberg, 1981.
[141] H. L. Van Trees. Detection, estimation, and modulation theory. John Wiley \& Sons, 2004.
[142] Anne M Treisman and Garry Gelade. A feature-integration theory of attention. Cognitive psychology, 12(1):97-136, 1980.
[143] A. Tripathi and S. Vijay. A note on a theorem of erdös and gallai. Discrete Mathematics, 265(1-3):417-420, Apr. 2003.
[144] M. Trusov, A. V. Bodapati, and R. E. Bucklin. Determining influentual users in internet social networks. Journal of Marketing Research, XLVII:643-658, Aug. 2010.
[145] B. Tuttle. Fact-checking the crowds: How to get the most out of hotel-review sites. Time magazine, Jul. 2013.
[146] F. Vega-Redendo. Complex Social Networks. Economic Society Monographs. Cambridge University Press, 2007.
[147] A. Wagner. The yeast protein interaction network evolves rapidly and contains few redundant duplicate genes. Molecular Biology and Evolution, 18(7):1283-1292, 2001.
[148] T. Wang and H. Krim. Statistical classification of social networks. In Proceedings of IEEE International Conference on Acoustics, Speech and Signal Processing, pages 3977-3980, Kyoto, Japan, Mar. 2012.
[149] T. Wang, H. Krim, and Y. Viniotis. A generalized Markov graph model: Application to social network analysis. IEEE journal of Selected Topics in Signal Processing, 7(2):318-332, Apr. 2013.
[150] X. Xie and R. J. Evans. Multiple target tracking and multiple frequency line tracking using hidden markov models. IEEE Transactions on Signal Processing, 39(12):2659-2676, Dec. 1991.
[151] Y. Xin and T. Jaakkola. Controlling privacy in recommender systems. In Advances in Neural Information Processing Systems, pages 2618-2626, 2014.
[152] J. S. Yedidia, W. T. Freeman, and Y. Weiss. Constructing free-energy approximations and generalized belief propagation algorithms. IEEE Transactions on Information Theory, 51(7):2282-2312, Jul. 2005.
[153] G. Yin and V. Krishnamurthy. LMS algorithms for tracking slow Markov chains with applications to hidden Markov estimation and adaptive multiuser detection. IEEE Transactions on Information Theory, 51(7):2475-2490, Jul. 2005.
[154] G. Yin, V. Krishnamurthy, and C. Ion. Regime switching stochastic approximation algorithms with application to adaptive discrete stochastic optimization. SIAM Journal on Optimization, 14(4):1187-1215, 2004.
[155] G. Yin, Y. Sun, and L. Y. Wang. Asymptotic properties of consensus-type algorithms for networked systems with regime-switching topologies. Automatica, 47(7):1366-1378, 2011.
[156] G. Yin and Q. Zhang. Discrete-Time Markov Chains: Two-Time-Scale Methods and Applications, volume 55 of Applications of Mathematics. Srpinger-Verlag, New York, NY, 2005.
[157] G. Yin and C. Zhu. Hybrid Switching Diffusions: Properties and Applications, volume 63 of Stochastic Modeling and Applied Probability. Springer, 2010.

## Appendix A

## Some Graph Theoretic Definitions

## Graph, Directed Graph, Path and Directed Acyclic Graph (DAG):

- A graph $G_{N}$ comprising of $N$ nodes is a pair $(V, E)$, where $V=\left\{v_{1}, \ldots, v_{N}\right\}$ is a set of nodes (also called vertices), and $E \subset V \times V$ is a set of edges between the nodes.
- Graph $G_{N}$ is an undirected graph if for any $\left(v_{i}, v_{j}\right) \in E$ then $\left(v_{j}, v_{i}\right) \in E$ and a graph is said to be directed if $\left(v_{j}, v_{i}\right) \in E$ is not a consequence of $\left(v_{i}, v_{j}\right) \in E$.
- A path is an alternating sequence of nodes and edges, beginning and ending with an edge, in which each node is incident to the two edges that precede and follow it in the sequence.
- A Directed Acyclic Graph (DAG) is a directed graph with no path that starts and ends at the same node.
- A family of DAGs $\mathscr{G}_{N}$ is defined as a set of DAGs $\left\{G_{1}, \ldots, G_{N}\right\}$ where $G_{n}$ is the sub-graph of $G_{n+1}$ such that for $n=1, \ldots, N-1$

$$
\left\{\begin{array}{l}
V_{n}=V_{n+1} / v_{n+1},  \tag{A.1}\\
E_{n}=E_{n+1} /\left\{\left(v_{i}, v_{n+1}\right) \in E_{n+1} \mid v_{i} \in V_{n+1}\right\} .
\end{array}\right.
$$

## Adjacency and Transitive Closure matrices:

Let $G_{N}=(V, E)$ denote a graph with $N$ nodes $V=\left\{v_{1}, \ldots, v_{N}\right\}$.

- The Adjacency Matrix $\bar{A}$ of $G_{N}$ is an $N \times N$ matrix whose elements $\bar{A}(i, j)$ are given by

$$
\bar{A}(i, j)=\left\{\begin{array}{ll}
1 & \text { if }\left(v_{j}, v_{i}\right) \in E,  \tag{A.2}\\
0 & \text { otherwise }
\end{array} . \quad . A(i, i)=0 .\right.
$$

- The Transitive Closure Matrix $T$ of $G_{N}$ is an $N \times N$ matrix whose elements $T(i, j)$ are given by $T(i, i)=1$, and

$$
T(i, j)=\left\{\begin{array}{ll}
1 & \text { if there is a path between } v_{j} \text { and } v_{i},  \tag{A.3}\\
0 & \text { otherwise }
\end{array} .\right.
$$

The following shows the special form of the adjacency matrix of the directed acyclic graph and provides a closed form expression to compute the transitive closure matrix from the adjacency matrix of a directed acyclic graph.

Lemma A.1. A sufficient condition for a graph $G_{N}$ to be a DAG is that the Adjacency matrix $A$ is an upper triangular matrix. For a DAG $G_{N}$, the Transitive Closure Matrix $T$ is related to the Adjacency matrix by

$$
\begin{equation*}
T=Q\left(\left\{\mathbf{I}_{N}-A\right\}^{-1}\right) \tag{A.4}
\end{equation*}
$$

Here, $\mathbf{I}_{N}$ is the $N \times N$ identity matrix, and $Q$ denote the matrix valued "quantization" function so that for any $N \times N$-matrix $B, Q(B)$ is the $N \times N$ matrix with elements

$$
Q(B)(i, j)= \begin{cases}0 & \text { if } B(i, j)=0  \tag{A.5}\\ 1 & \text { if } B(i, j) \neq 0\end{cases}
$$

Proof: This result is derived from the classical interpretation of matrix $\left\{\mathbf{I}_{N}-A\right\}^{-1}$. The entry in row $i$ and column $j$ of this matrix gives the number of paths from node $i$ to node $j$.

To deal with information flow in a social network, we now introduce the concept of a family of DAGs.

Remark A.1. For the sake of simplicity in notations, let us define two vector representatives of adjacency and transitive closure matrices of directed acyclic graph. For each graph $G_{n} \in \mathscr{G}_{N}$, let the $n \times n$ matrices $\bar{A}_{n}$ and $T_{n}$, respectively, denote the adjacency matrix and transitive closure matrix. Define the following:

$$
\left\{\begin{array}{l}
t_{n} \in\{0,1\}^{1 \times(n-1)}: \text { transpose of first } n-1 \text { elements of nth column of } T_{n},  \tag{A.6}\\
b_{n} \in\{0,1\}^{1 \times(n-1)}: \text { transpose of first } n-1 \text { elements of nth column of } \bar{A}_{n} .
\end{array}\right.
$$

Remark A.2. As can be straightforwardly followed from the construction of adjacency and transitive closure matrices in (A.1), for a family of DAGs $\mathscr{G}_{N}=\left\{G_{1}, \ldots, G_{N}\right\}$, for any $n \in\{1, \ldots, N-1\}$, the adjacency matrix $\bar{A}_{n}$ and transitive closure matrix $T_{n}$ of graph $G_{n}$ are respectively the $n \times n$ left upper matrices of the adjacency matrix $A_{n+1}$ and transitive closure matrix $T_{n+1}$ of graph $G_{n+1}$.

## Appendix B

## A Note on Degree-based Graph Construction

The first step in numerical studies of social networks is the graphical modeling of such networks. A graph can be uniquely determined by the adjacency matrix (also known as the connectivity matrix) of the graph. However, in the graphical modeling of social networks (specially when the size of the network is relatively large), the only available information is the degree sequence of nodes, and not the adjacency matrix of the network.

Definition B.1. The degree sequence, denoted by $\mathbf{d}$, is a non-increasing sequence comprising of the vertex degrees of the graph vertices.

The degree sequence, in general, does not specify the graph uniquely; there can be a large number of graphs that realize a given degree sequence. It is straightforward to show that not all integer sequences represent a true degree sequence of a graph. For example, sequence $\mathbf{d}=\{2,1,1\}$ represents a tree with two edges, but $\mathbf{d}=\{3,2,1\}$ cannot be realized as the degree sequence of a simple graph. Motivated by social network applications, this section addresses the following two questions given a degree sequence $\mathbf{d}$ :

- Existence Problem: Is there any simple graph that realizes d?
- Construction Problem: How can we construct all simple graphs that realize a true degree sequence d?

There are two well-known results that address the existence problem: (i) the Erdös-Gallai theorem [59] and the Havel-Hakimi theorem [70, 73]. These theorems provide necessary and sufficient conditions for a sequence of non-negative integers to be a true degree sequence of a simple graph. Here, we recall these results without proofs.

Theorem B. 1 (Erdös-Gallai, [59]). Let $d_{1} \geq d_{2} \geq \cdots \geq d_{n}>0$ be integers. Then, the degree sequence $\mathbf{d}=\left\{d_{1}, \cdots, d_{n}\right\}$ is graphical if and only if

1. $\sum_{i=1}^{n} d_{i}$ is even;
2. for all $1 \leq k<n$ :

$$
\begin{equation*}
\sum_{i=1}^{k} d_{i} \leq k(k-1)+\sum_{i=k+1}^{n} \min \left\{k, d_{i}\right\} . \tag{B.1}
\end{equation*}
$$

It is shown in [143] that there is no need to check (B.1) for all $1 \leq k \leq n-1$; it suffices to check (B.1) for $1 \leq k \leq s$, where $s$ is chosen such that $d_{s} \geq s$ and $d_{s+1}<s+1$. Note that, in degree-based graph construction, we only care about nodes of degree greater than zero; zero-degree nodes are isolated nodes which can be added to the graph consisting of nodes of degree greater than zero.

The Havel-Hakimi theorem also provides necessary and sufficient conditions for a degree sequence to be graphical. It also gives a greedy algorithm to construct a graph from a given graphical degree sequence.

Theorem B. 2 (Havel-Hakimi, [70, 73]). Let $d_{1} \geq d_{2} \geq \cdots \geq d_{n}>0$ be integers. Then, the degree sequence $\mathbf{d}=\left\{d_{1}, \cdots, d_{n}\right\}$ is graphical if and only if the degree sequence $\mathbf{d}^{\prime}=\left\{d_{2}-1, d_{3}-\right.$ $\left.1, \cdots, d_{d_{1}+1}-1, d_{d_{1}+2}, \cdots, d_{n}\right\}$ is graphical.

In the following, we provide algorithms to construct a simple graph from a true degree sequence. In the construction problem, the degree sequence is treated as a collection of half-edges; a node with degree $d_{i}$ has $d_{i}$ half-edges. One end of these half-edges are fixed at node $i$, but the other ends are free. An edge between node $i$ and node $j$ is formed by connecting a half-edge from node $i$ to a half-edge from node $j$. The aim is to connect all these half edges such that no free half-edge is left. The Havel-Hakimi theorem provides a recursive procedure to construct a graph from a graphical degree sequence. This procedure is presented in Algorithm B. 6

Using Algorithm B.6, one can sample from graphical realizations of a given degree sequence. In this algorithm, each vertex is first connected to nodes with lower degrees. Therefore, Algorithm B. 6 generates graphs where high-degree nodes tend to connect to low-degree nodes; the resulting graph has assortative property [84, 115]. To overcome this problem, one way is to perform edge swapping repeatedly such that the final graph looses its assortative property. In the edge swapping method, two edges (for example $(1,2)$ and $(3,4)$ ) can be swapped (to $(1,4)$ and $(2,3)$ ) without changing the degree sequence. Edge swapping method is also used to generate all samples from a given degree sequence; one sample is generated via Algorithm B. 6 and then, by use of Markov chain MonteCarlo algorithm based on edge swapping [140], other samples from the graphical realizations of the degree sequence are obtained.

In [84] a swap-free algorithm is proposed to generate all graphical realizations of a true degree sequence. Before proceeding to Algorithm B.7, we first provide definitions which will be used in this algorithm.

Definition B.2. Let $\mathbf{d}=\left\{d_{1}, \cdots, d_{n}\right\}$ be a degree sequence of a simple graph and $N(i)$ be the set of adjacent nodes of node $i$. Then, the degree sequence reduced by $N(i)$ is denoted by $\left.\mathbf{d}\right|_{N(i)}=$

```
Algorithm B. 6 Creating a sample graph from a given degree sequence
Given a graphical sequence \(d_{1} \geq d_{2} \geq \cdots \geq d_{n}>0\) :
Start from \(i=1\)
(i) Initialize \(k=n\).
(ii) Connect (one half-edge of) node \(i\) to (a half-edge of) node \(k\)
(iii) Check that the resulting degree sequence is graphical
```


## - if Yes:

1. Let $k=k-1$.
2. Repeat (i).

## - if No:

1. Save the connection between node $i$ and node $k$
2. If node $i$ has any half-edges left, let $k=k-1$ and repeat (i)
(iv) If $i<n$, then, $i \leftarrow i+1$ and repeat (i).
$\left\{\left.d_{1}\right|_{N(i)}, \cdots,\left.d_{n}\right|_{N(i)}\right\}$ with elements defined as follows

$$
\left.d_{k}\right|_{N(i)}=\left\{\begin{array}{lc}
d_{k}-1, & \text { if } k \in N(i)  \tag{B.2}\\
0, & \text { if } k=i, \\
d_{k}, & \text { otherwise }
\end{array}\right.
$$

Definition B.3. Let $\left(a_{1}, a_{2}, \ldots, a_{n}\right)$ and $\left(b_{1}, b_{2}, \ldots, b_{n}\right)$ be two sequences. Then, $\left(a_{1}, a_{2}, \ldots, a_{n}\right)<_{C R}$ $\left(b_{1}, b_{2}, \ldots, b_{n}\right)$ if and only if there exists an index $m$ such that $1 \leq m \leq n$ and $a_{m}<b_{m}$ and $a_{i}=b_{i}$ for all $m<i \leq n$.

Let $\mathbf{d}$ be a non-increasing graphical degree sequence. In order to construct the graph, we need to find all possible neighbors $N(i)$ ("allowed set") of each node $i$ such that if we connect this node to its allowed set, then the resulting reduced degree sequence $\left.\mathbf{d}\right|_{N(i)}$ is also graphical, i.e., the graphicality is preserved. Algorithm B. 7 provides a systematic way (swap-free) to generate all graphical realizations of a true degree sequence (by means of finding all possible neighbors of each node).

```
Algorithm B. 7 Constructing all graphs from a graphical degree sequence [84]
Given a graphical sequence \(d_{1} \geq d_{2} \geq \cdots \geq d_{n}>0\)
Start from \(i=1\)
```


## Step 1: Find neighbors with highes index of node $i$

```
The aim is to find \(A_{R}(i)\) :
(i) Initialize \(k=n\).
(ii) Connect node \(i\) to node \(k\)
(iii) Check that the resulting degree sequence is graphical
- if Yes:
1. Let \(k=k-1\)
2. Repeat (i).
- if No:
1. Save the connection between node \(i\) and node \(k\)
2. If node \(i\) has any stubs left, let \(k=k-1\) and repeat (i)
```


## Step 2: Find all possible neighbors of node $i$

```
With \(<_{C R}\) defined in (B.3), the aim is to find
\(\mathscr{A}(i)=\left\{N(i)=\left\{v_{1}, \cdots, v_{d_{i}}\right\} ; N(i)<_{C R} A_{R}(i)\right.\) and \(\left.\mathbf{d}\right|_{N(i)}\) is graphical \(\}\)
```

(i) Find all sets of nodes who are colexicographically smaller than $A_{R}(1)$ (prospective neighbor sets).
(ii) Connect node $i$ to those neighbors and check if the resulting degree sequence is graphical.

Step 3: For every $N(i) \in \mathscr{A}(i)$ :

- Connect node $i$ to $N(i)$
- Discard node $i$ and compute the reduced degree sequence $\left.\mathbf{d}\right|_{N(i)}$
- Create all graphs from degree sequence $\left.\mathbf{d}\right|_{N(i)}$ using this algorithm


[^0]:    ${ }^{1}$ https://www.angrybirds.com.
    ${ }^{2}$ About 6 thousand tweets are sent on average in a second. The tweet-per-second number on US Presidential election day in 2012, was 15 thousand tweets per second resulting in 500 million tweets in the day. Twitter can be considered as a real-time human-based sensor of social situations.

[^1]:    ${ }^{3}$ It is reported in [79] that $81 \%$ of hotel managers regularly check Tripadvisor reviews. It is reported in [109] that a one-star increase in the Yelp rating maps to $5-9 \%$ revenue increase.

[^2]:    ${ }^{4}$ These terms are used interchangeably throughout this dissertation.

[^3]:    ${ }^{5}$ A giant component is a connected component with size $O(n)$, where $n$ is the total number of vertices in the graph. If the average degree of a random graph is strictly greater than one, then there exists a unique giant component with probability one [41], and the size of this component can be computed from the expected degree sequence.

[^4]:    ${ }^{6}$ Information flow graph is a directed acyclic graph that models the flow of information among social sensors, for example, a directed edge from Sensor 1 to Sensor 2 means that the information- or beliefs in the context of Chapter 2-of Sensor 1 is available at Sensor 2, see Section 2.2 for more details.

[^5]:    ${ }^{7}$ There exists some similarities between BP and social learning in the sense that they are both systematic structures to perform Bayesian inference over graphs. However, they are not related in principle. While graphs represent social interactions among agents in social learning, graphical models in BP depict the conditional dependency between nodes (random variables)- they do not imply the actual communications, for more detail see[36].

[^6]:    ${ }^{8}$ Collaboration network of movie actors.
    ${ }^{9}$ There are different methods of sampling a network, for example, link sampling, node sampling, and snowball sampling. For a complete survey, we refer to [102]. In link or node sampling, a given fraction of links or nodes are sampled. In snowball sampling method, one node is chosen randomly and the next samples are chosen from its neighbors.

[^7]:    ${ }^{10}$ For some distributions, instead of transmitting the posterior distribution, it is sufficient to broadcast the sufficient statistic. In the finite case, the sufficient statistic and posterior distribution are similar.

[^8]:    ${ }^{11}$ In the example (3) in Section 2.1, $Z_{\lfloor 1,3\rfloor}=\left\{z_{\lfloor 1,2\rfloor}, z_{\lfloor 2,2\rfloor}, z_{\lfloor 1,1\rfloor}, z_{\lfloor 2,1\rfloor}\right\}$.

[^9]:    ${ }^{12}$ see Appendix A

[^10]:    ${ }^{13}$ Because the logarithm of product is the sum of individual logarithms, it is more convenient to use logarithm of estimates in Bayesian estimation context.

[^11]:    ${ }^{14}$ See Appendix A.

[^12]:    ${ }^{15}$ The results of this chapter also apply to continuous-valued observations. We consider discrete-valued observations since humans typically record discrete observations.
    ${ }^{16}$ It is not necessary for agents to record observations at each time $k$ and this does not interfere with the common knowledge assumption in social learning where agents all know about the structure of social learning model. Agents at different time instants are treated as different nodes in our graphical model. The assumption that agents record observation at each time $k$ simplifies notation.
    ${ }^{17}$ The scenario where agents choose their actions according to the network belief is similar to the classical social learning formulation [35] where actions are transmitted over the network.

[^13]:    ${ }^{18}$ We assume that multiple agents can transmit simultaneously over the network without interfering with each other. This is realistic in a social network, since the time required to exchange (broadcast) information is substantially smaller than the time to record observations, update beliefs or take actions.

[^14]:    ${ }^{19}$ The dimension of private beliefs is typically larger than that of actions. Also, individuals tend not to provide their private beliefs at the time of their further social interactions. Therefore, agents map their beliefs to a finite set of actions which are easier to broadcast.
    ${ }^{20}$ Having fair estimates of quality of a social unit is a problem of much interest in business. Most of hotel managers $(81 \%)$ regularly check the reviews on Tripadvisor [79]. In [109], it is found that a one-star increase in the average rating of users in Yelp is mapped to about 5-9 \% revenue increase.

[^15]:    ${ }^{21}$ There are subtle differences between an individual agent herding, a herd of agents and an information cascade; see for example $[35,91]$.

[^16]:    ${ }^{22}$ In this chapter we consider Bayesian estimation over a finite time horizon. We do not consider the asymptotic agreement of social learning or consensus formation in social networks. Consensus formation is asymptotic and typically non-Bayesian. From a practical point of view, information exchange in a social network is typically over a finite horizon.

[^17]:    ${ }^{23}$ See (A.3) in Appendix A.

[^18]:    ${ }^{24}$ In Section 3.3, a discussion is presented on the data incest removal algorithms in consumer rating web sites such as Yelp ${ }^{\circledR}$ - www.yelp.com.
    ${ }^{25}$ In the constrained social learning algorithm, each node receives the most recent public beliefs of its neighbors or equivalently the updated public belief.

[^19]:    ${ }^{26}$ See (A.3) in Appendix A.

[^20]:    ${ }^{27}$ These numbers are based on the work in Treisman et. al. paper [142].

[^21]:    ${ }^{28}$ One of the issues that comes with proliferation of online social networks (especially content-aware recommender systems), due to the astronomical amounts of information that these sites have about their users, is privacy. People are not usually willing to disclose their private information to a large group of audience. That's the reason why finding a trade-off between accuracy and privacy of the users have been studied widely in the literature of recommender systems [47, 126, $129,151]$.

[^22]:    ${ }^{29}$ Another type of privacy concerns that may arise here is the fact that the network administrator can compute the network belief for users, and thus, can predict the actions that users are about to make.

[^23]:    ${ }^{30}$ The duplication-deletion procedure for Markov-modulated random graphs is described in Section 4.2.
    ${ }^{31}$ For example, real-time event detection from Twitter posts is investigated in [132] or the early detection of contagious outbreaks via social networks is studied in [40].

[^24]:    ${ }^{32} \mathrm{~A}$ giant component is a connected component with size $O(n)$, where $n$ is the total number of vertices in the graph. If the average degree of a random graph is strictly greater than one, then there exists a unique giant component with probability one [41], and the size of this component can be computed from the expected degree sequence.

[^25]:    ${ }^{33}$ The irreducibility assumption implies that there exists a unique stationary distribution $\pi \in \mathbb{R}^{M \times 1}$ for this Markov chain such that $\pi^{\prime}=\pi^{\prime} A^{\rho}$.
    ${ }^{34}$ In Procedure 4.5, Step 1 is executed with probability $r$. Then, regardless of execution of Step 1 , Step 2 is implemented. For convenience in the analysis, assume that a node generated in the duplication step cannot be eliminated in the deletion step immediately after its generation. Also, nodes whose degrees change in the edge-deletion part of Step 2 , remain unchanged in the duplication part of Step 2 at that time instant. Finally, to prevent formation of isolated nodes, assume that the neighbor of a node with degree one cannot be eliminated in the deletion step. Note also that the duplication step in Step 2 ensures that the graph size does not decrease.

[^26]:    ${ }^{35}$ That is, each row adds to zero and each non-diagonal element of $L$ is positive.
    ${ }^{36}$ It is straightforward to show that all elements of $\left(B_{N_{0}}\right){ }^{N_{0}}$ are strictly greater than zero. Therefore, $B_{N_{0}}$ is irreducible and aperiodic.

[^27]:    ${ }^{37}$ There is a difference between "power law" and "power law distribution". Power law is a functional relationship between two parameters where one parameter is proportional to the power of another, i.e., $x \propto y^{-\beta}$, where $\beta$ can be any real number. In comparison, the exponent of a power law distribution is strictly greater that one [117]. Otherwise, the probability distribution does not add up to one.
    ${ }^{38}$ The normalization constant $\alpha$ is computed from $\alpha=-\log \left[\zeta\left(\beta, i^{*}\right)\right]$, where $\zeta\left(\beta, i^{*}\right)=\sum_{k=i^{*}}^{\infty} k^{-\beta}$ denotes the incomplete Riemann $\zeta$-function.

[^28]:    ${ }^{39}$ Respondent driven sampling (RDS) was introduced in [74] as an approach for sampling from hidden populations in social networks. RDS has been selected by the U.S. Centers for Disease Control and Prevention as part of the HIV behavioral surveillance system. RDS can be viewed as a form of Markov Chain Monte Carlo sampling [68].

[^29]:    ${ }^{40}$ In this work, we assume that $\rho=O(\varepsilon)$. Therefore, $\rho^{2} \ll \varepsilon$.

[^30]:    ${ }^{41}$ We refer the interested reader to [99, Chapter 7] for further details on weak convergence and related matters.

[^31]:    ${ }^{42}$ This means that the target node can be found in the search problem with smaller number of steps.

[^32]:    ${ }^{43}$ The deletion step (Step 2 of Procedure 4.5) comprises an edge-deletion step and a duplication step. The probability that the degree of node with degree $i$ changes in the edge-deletion step is $\frac{q(i+1)}{N_{n}}$; either this node or one of its neighbors should be selected in the edge-deletion step. Also given that the degree of this node dose not change in the edge-deletion step, if either this node or one of its neighbor is selected in the duplication step (within Step 2) then the degree of this node increments by one with probability $\frac{1+p i}{N_{n}}$. Therefore, the probability that the degree of a node of degree $i$ remains unchanged in Step 2 is

    $$
    1-\frac{q(i+1)+q(1+p i)-q(1+p i)(i+1) / N_{n}}{N_{n}} .
    $$

    Note that for simplicity in our analysis, it is assumed that the nodes whose degrees change in the edge-deletion part of Step 2, remain unchanged in the duplication part of Step 2 at that time instant. Also, the new node, which is generated in the vertex-duplication step of Step 1, remains unchanged in Step 2.

[^33]:    ${ }^{44}$ Instead of spread of disease or information, we can also consider a scenario where nodes are interactively deciding whether to adapt a technology (infected) or not (remain susceptible). The aim now is to study the rate of adaptation of new technology in the network.

