Complex Networks: New Models and Distributed Algorithms

Alireza Tahbaz Salehi
University of Pennsylvania, atahbaz@seas.upenn.edu

Follow this and additional works at: http://repository.upenn.edu/edissertations
Part of the Controls and Control Theory Commons

Recommended Citation
http://repository.upenn.edu/edissertations/84

This paper is posted at ScholarlyCommons. http://repository.upenn.edu/edissertations/84
For more information, please contact libraryrepository@pobox.upenn.edu.
Abstract
Over the past few years, a consensus has emerged among scientists and engineers that net-centric technology can provide unprecedented levels of performance, robustness, and efficiency. Success stories such as the Internet, distributed sensor networks, and multi-agent networks of mobile robots are only a few examples that support this view. The important role played by complex networks has been widely observed in various physical, natural, and social systems. Given the complexity of many of these systems, it is important to understand the fundamental rules that govern them and introduce appropriate models that capture such principles, while abstracting away the redundant details.

The main goal of this thesis is to contribute to the emerging field of "network science" in two ways. The first part of the thesis focuses on the question of information aggregation over complex networks. The problem under study is the asymptotic behavior of agents in a network when they are willing to share information with their neighbors. We start by focusing on conditions under which all agents in the network will asymptotically agree on some quantity of interest, what is known as the consensus problem. We present conditions that guarantee asymptotic agreement when inter-agent communication links change randomly over time. We then propose a distributed (non-Bayesian) algorithm that enables agents to not only agree, but also learn the true underlying state of the world. We prove that our proposed learning rule results in successful information aggregation, in the sense that all agents asymptotically learn the truth as if they were completely informed of all signals and updated their beliefs rationally. Moreover, the simplicity of our local update rule guarantees that agents eventually achieve full learning, while at the same time, avoiding highly complex computations that are essential for full Bayesian learning over networks.

The second part of this thesis focuses on presenting a new modeling paradigm that greatly expands the tool set for mathematical modeling of networks, beyond graphs. The approach taken is based on using simplicial complexes, which are objects of study in algebraic topology, as generalizations of graphs to higher dimensions. We show how simplicial complexes serve as more faithful models of the network and are able to capture many of its global topological properties. Furthermore, we develop distributed algorithms for computing various topological invariants of the network. These concepts and algorithms are further explored in the context of a specific application: coverage verification in coordinate-free sensor networks, where sensor nodes have no access to location, distance, or orientation information. We propose real-time, scalable, and decentralized schemes for detection of coverage holes, as well as computation of a minimal set of sensors required to monitor a given region of interest. The presented algorithms clarify the benefits of using simplicial complexes and simplicial homology, instead of applying tools from graph theory, in modeling and analyzing complex networks.

Degree Type
Dissertation

Degree Name
Doctor of Philosophy (PhD)

Graduate Group
Electrical & Systems Engineering

This dissertation is available at ScholarlyCommons: http://repository.upenn.edu/edissertations/84
First Advisor
Ali Jadbabaie

Keywords
Complex networks, distributed algorithms, distributed control, multi-agent systems

Subject Categories
Controls and Control Theory
COMPLEX NETWORKS: NEW MODELS AND DISTRIBUTED ALGORITHMS

Alireza Tahbaz Salehi

A DISSERTATION

in

Electrical and Systems Engineering

Presented to the Faculties of the University of Pennsylvania

in

Partial Fulfillment of the Requirements for the

Degree of Doctor of Philosophy

2009

Ali Jadabala, Supervisor of Dissertation

Roch Guerin, Graduate Group Chair

Dissertation Committee

Robert Ghrist, Professor
Daniel Koditschek, Professor
George Pappas, Professor
Asuman Ozdaglar, Associate Professor
Acknowledgements

This thesis could not have been written without the support of my advisor, Professor Ali Jadbabaie. Ali taught me how to think independently while providing me with guidance on all aspects of research. His encouragement during these years will always be of immense value to me.

I am also greatly indebted to Professor Robert Ghrist for the many helpful conversations that we had over the past years. The second half of this thesis would have been impossible without his guidance. I am thankful to the other members of my committee: Professors George Pappas, Daniel Koditschek, and Asuman Ozdaglar. Their valuable suggestions and considerable advice have greatly improved the quality of this thesis. I would also like to thank Professor Alvaro Sandroni for his contributions in Chapter 3.

An important role was played by my many friends at Penn. Special thanks to Elena, Nima, Agung, Michael, Farhad, Gil, Antonio, Andreas, and many others. Without them, my years in Philadelphia would not have been as enjoyable.

Last but not least, my deepest thanks to my parents, Farzaneh and Ali, and my sister, Ghazaleh, who gave me their unconditional love and support. I owe them much more than I would ever be able to express.
ABSTRACT

COMPLEX NETWORKS: NEW MODELS AND DISTRIBUTED ALGORITHMS

Alireza Tahbaz Salehi

Supervisor: Ali Jadbabaie

Over the past few years, a consensus has emerged among scientists and engineers that net-centric technology can provide unprecedented levels of performance, robustness, and efficiency. Success stories such as the Internet, distributed sensor networks, and multi-agent networks of mobile robots are only a few examples that support this view. The important role played by complex networks has been widely observed in various physical, natural, and social systems. Given the complexity of many of these systems, it is important to understand the fundamental rules that govern them and introduce appropriate models that capture such principles, while abstracting away the redundant details.

The main goal of this thesis is to contribute to the emerging field of “network science” in two ways. The first part of the thesis focuses on the question of information aggregation over complex networks. The problem under study is the asymptotic behavior of agents in a network when they are willing to share information with their neighbors. We start by focusing on conditions under which all agents in the network will asymptotically agree on some quantity of interest, what is known as the consensus problem. We present conditions that guarantee asymptotic agreement when inter-agent communication links change randomly over time. We then propose a distributed (non-Bayesian) algorithm that enables agents to not only agree, but also learn the true underlying state of the world. We prove that our proposed learning rule results in successful information aggregation, in the sense that all agents asymptotically learn the truth as if they were completely informed of all signals and updated their beliefs rationally. Moreover, the simplicity of our local update rule guarantees that agents eventually achieve full learning, while at the same time, avoiding highly
complex computations that are essential for full Bayesian learning over networks.

The second part of this thesis focuses on presenting a new modeling paradigm that greatly expands the tool set for mathematical modeling of networks, beyond graphs. The approach taken is based on using simplicial complexes, which are objects of study in algebraic topology, as generalizations of graphs to higher dimensions. We show how simplicial complexes serve as more faithful models of the network and are able to capture many of its global topological properties. Furthermore, we develop distributed algorithms for computing various topological invariants of the network. These concepts and algorithms are further explored in the context of a specific application: coverage verification in coordinate-free sensor networks, where sensor nodes have no access to location, distance, or orientation information. We propose real-time, scalable, and decentralized schemes for detection of coverage holes, as well as computation of a minimal set of sensors required to monitor a given region of interest. The presented algorithms clarify the benefits of using simplicial complexes and simplicial homology, instead of applying tools from graph theory, in modeling and analyzing complex networks.
## Contents

Acknowledgements ii

Abstract iii

Contents v

List of Figures viii

1 Overview 1

  1.1 Information Aggregation over Complex Networks . . . . . . . . . . . . . . 3
  1.2 New Models of Networks: Beyond Graphs . . . . . . . . . . . . . . . . . 7

I Information Aggregation Over Complex Networks 9

2 Network Consensus Algorithms 10

  2.1 Distributed Consensus Algorithms . . . . . . . . . . . . . . . . . . . . . . . 11
    2.1.1 Ergodicity . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 13
    2.1.2 Coefficients of Ergodicity . . . . . . . . . . . . . . . . . . . . . . . . . . 15
    2.1.3 Consensus and Joint Connectivity . . . . . . . . . . . . . . . . . . . . . 18
  2.2 Consensus over Random Networks . . . . . . . . . . . . . . . . . . . . . . . 21
    2.2.1 Independent and Identically Distributed Random Networks . . . . . . 22
2.2.2 Ergodic Stationary Random Networks ..................................... 32
2.3 Related Literature ................................................................. 37

3 Social Learning ....................................................................... 39
  3.1 Summary of Results and Related Literature ............................. 40
    3.1.1 Summary of Results .......................................................... 41
    3.1.2 Related Literature ............................................................. 44
  3.2 A Model of Non-Bayesian Social Learning ............................... 46
    3.2.1 Agents and Observations .................................................... 46
    3.2.2 Social Structure ................................................................. 47
    3.2.3 Belief Updates ................................................................. 48
  3.3 Social Learning .................................................................... 50
    3.3.1 Correct Forecasts in Strongly Connected Societies .............. 52
    3.3.2 Asymptotic Beliefs ............................................................ 58
  3.4 Concluding Remarks ............................................................. 66

II New Models of Networks: Beyond Graphs ................................. 67

4 From Graphs to Simplicial Complexes ...................................... 68
  4.1 Simplicial Complexes ............................................................. 70
    4.1.1 Boundary Homomorphism ................................................ 73
    4.1.2 Simplicial Homology ........................................................ 74
    4.1.3 Relative Homology .......................................................... 76
  4.2 Combinatorial Laplacians ...................................................... 76

5 Application: Coverage Verification in Coordinate-Free Sensor Networks ........................................ 81
  5.1 Motivation and Related Literature ........................................ 81
List of Figures

3.1 The figure illustrates a strongly connected social network of 7 agents, which is of the form of a directed cycle. ........................................ 63

4.1 A simplicial complex. .................................................. 71

4.2 Relative homology: $\xi_1$ is the representative of a non-trivial element in $H_1(X,A)$. ......................................................... 77

4.3 A simplicial complex of dimension 2. The numbers labeled on the edges correspond to the components of the eigenvector corresponding to 0 eigenvalue, i.e. the kernel, of the first combinatorial Laplacian matrix. ........... 79

5.1 The simplicial complexes corresponding to the above figures are combinatorially equivalent. Therefore, in both cases, the shortest cycle encircling the hole has length 4. ......................................................... 93

5.2 The sparsest generator of the homology class of the cycles that encircle each hole once clockwise is not necessarily tight around both. ........... 94

5.3 If the first homology of $\mathcal{R}$ is non-trivial, then the second relative homology $H_2(\mathcal{R},\mathcal{F})$ has no generator with values on the boundary. Conversely, if the second homology relative to the boundary has a non-trivial generator with a non-vanishing boundary, then $H_1(\mathcal{R}) = 0$. ....................... 104
5.4 The eight faces of the octahedron form a non-trivial 2-cycle $\alpha$ such that $[\alpha] \in H_2(\mathcal{R})$. However, $\alpha$ has a vanishing boundary $\partial_2 \alpha = 0$, and therefore, does not correspond to a true relative 2-cycle.

5.5 Subgradient methods can be used to localize the holes in a distributed fashion.

5.6 Finding the minimal generator of the second relative homology $H_2(\mathcal{R}, \mathcal{F})$ leads to a minimal cover. 32 of the sensors can be turned off without generating any coverage holes.
Chapter 1

Overview

Over the past few years, a consensus has emerged among scientists and engineers that net-centric technology can provide unprecedented levels of performance, robustness, and efficiency. Success stories such as the Internet, distributed sensor networks, and multi-agent networks of mobile robots are only a few examples that support this view. The important role played by complex networks has been widely observed in various physical, natural, and social systems. For example, it is widely believed that many phenomena in the natural world such as swarming in locusts, flocking of birds, and synchronous flashing of fireflies are due to interaction of these species over a network. Similarly, in social sciences, the central role played by social and economics networks is well-documented in many scenarios including consumer choice, spread of rumors, and emergence of fashion and trends.

The wide range of network phenomena and applications, such as the ones mentioned above, have sparked a widespread interest in understanding, modeling, and analyzing networks in various fields such as engineering, economics, biology, and ecology. In the controls and robotics community, this interest has been coupled with the need for developing algorithms for decentralized decision-making in the absence of centralized coordination,
with real-life engineering applications ranging from cooperative control of multiple autonomous agents to mobile sensor networks.

Given the complexity of many networked-systems, the main challenge is to understand fundamental rules that govern them and introduce appropriate models that capture such principles, while abstracting away the redundant details. The goal of this thesis is to contribute to the emerging field of complex networks by providing new models and tools for analysis of network behavior, as well as proposing algorithms and methods for distributed decision-making and control of dynamical systems.

To this end, the first part of the thesis addresses the question of information aggregation in complex networks. The problem under study is analyzing asymptotic behavior of agents in a network when they are willing to share information with their neighbors. We start by focusing on conditions under which all agents in the network will asymptotically agree on some quantity of interest, what is known as the consensus problem. We present conditions that guarantee asymptotic agreement when inter-agent communication links change randomly over time. We then propose a distributed (non-Bayesian) algorithm that enables agents to not only agree on some state, but also learn the true underlying state of the world. We prove that our proposed learning rule results in successful information aggregation, in the sense that all agents asymptotically learn the truth as if they were completely informed of all signals and updated their beliefs rationally. Moreover, the simplicity of our local update rule guarantees that agents eventually achieve full learning, while avoiding highly complex computations that are essential for full Bayesian learning over the network.

The second part of this thesis focuses on presenting a new modeling paradigm that greatly expands the tool set for mathematical modeling of networks, beyond graphs. The approach taken is based on using simplicial complexes, which are objects of study in algebraic topology, as generalizations of graphs to higher dimensions. We show how simplicial complexes serve as more faithful models of the network and can capture many of its
global topological properties. Furthermore, we develop distributed algorithms for computing various topological invariants of the network. These concepts and algorithms are further explored in the context of a specific application: coverage verification in coordinate-free sensor networks, where sensor nodes have access to no location, distance, or orientation information. We propose real-time, scalable, and decentralized schemes for detection of coverage holes, as well as computation of a minimal set of sensors required to monitor a given region of interest. The presented algorithms clarify the benefits of using simplicial complexes and simplicial homology, instead of applying tools from graph theory, in modeling and analyzing complex networks.

1.1 Information Aggregation over Complex Networks

Consider a group of agents who are willing to share some privately available information with one another. An example would be a group of individuals in a social network, each with limited personal experience about a new product in the market. By communicating and sharing their information, the individuals expect to reach a better judgment about the actual quality of the product. As another example, consider a group of simple sensors deployed over a region of interest, who are responsible for collecting information on room temperatures in a building, potential fires in a forest, or enemy installations in a battlefield. In the absence of a centralized information center, and due to their limited sensing range, it is important that the sensors share their information with one another properly.

The important underlying characteristic of the examples presented above (as of many others) is that any given agent does not necessarily have access to information held by all others. Instead, in almost all scenarios, each agent can only communicate with a small subset of other agents. This restriction signifies the importance of studying distributed algorithms.
algorithms that guarantee successful information aggregation over networks, when each agent can only communicate locally with its neighbors.

**Distributed Consensus Algorithms**

One of the earliest examples of distributed algorithms for sharing information over networks goes back to DeGroot [33]. In the model studied by DeGroot, information held by each agent, called its belief or state value, is captured by a scalar. He assumes that agents update their beliefs in each period simply by taking weighted averages of their neighbors’ opinions from the previous period, possibly placing some weight on their own previous beliefs. This linear update model captures repeated communication and interactions among agents and their neighbors. In spite of its simplicity, DeGroot shows that his model guarantees that all agents will have asymptotically equal state values, as long as the network representing the communication links is connected. He refers to this outcome as *consensus* over the network.

Similar models have been studied by researchers in the control theory and robotics community. The main goal of this line of research is to analyze and design distributed protocols for motion coordination that use local information and result in a desired global behavior. A well-known example of such a model is Vicsek’s model of alignment for self-propelled particles [107], first appeared in the statistical physics literature. The model is a distributed, iterative scheme in which a set of self-propelled particles moving with constant velocity update their headings based on the average headings of their neighbors: agents within a certain pre-specified distance from them. One feature that sharply distinguishes Vicsek’s model from DeGroot’s is that, due to their mobility, the neighbors of each agent might change over time. In spite of this, simulations performed by Vicsek [107] clearly show that, in the absence of noise, all particles will asymptotically align with one another, the given sensor.
that is, a consensus in their headings is achieved. Vicsek’s model was first analyzed by Jadabaie, Lin, and Morse [64], who show that if the proximity graphs representing the interconnection of agents are infinitely-often jointly connected, then all headings converge to a common value asymptotically. Their work initiated a tremendous amount of interest in the controls and robotics community. A non-exhaustive list of relevant research in this field includes [4, 17, 66, 72, 78, 84, 85].

Chapter 2 of this thesis is dedicated to the study of distributed consensus algorithms over networks. It contains a review of the main known results on necessary and sufficient conditions for asymptotic agreement over dynamic networks. The review is followed by the study of consensus algorithms over randomly changing networks, which is the main contribution of the chapter. The importance of studying conditions for consensus over random networks lies in the fact that, in many real-world network applications, the availability of communication links between different entities is usually random. This randomness can be due to link failures, node failures, interference, or existence of physical obstacles interrupting communication, which are natural features of wireless communications. We conclude the chapter by computing the statistical properties of the value that all agents agree upon, when the network changes randomly.

Social Learning

One of the shortcomings of DeGroot’s model and other similar distributed consensus algorithms is that they can only guarantee asymptotic agreement, with limited amount of control over the value that the agents agree upon. Although sufficient in some applications, in many scenarios it is important to have the information aggregated correctly, in the sense that all agents agree on the right value, an outcome known as social learning.

Belief update rules that result in asymptotic learning can be categorized to two general classes depending on whether the agents update their beliefs rationally or not. The Bayesian
learning literature focuses on scenarios that agents incorporate information provided to them rationally into their beliefs. Although optimal, rational learning rules require agents to perform complicated and often not computationally feasible updates. Therefore, for large and complex networks, one needs to consider coarser learning rules according to which, agents use simple rule-ofthumbs to incorporate the information provided to them through their neighbors.

The main contribution of Chapter 3 is to propose a simple non-Bayesian and distributed belief update rule that results in asymptotic learning over networks. In the presented model, at every time period, each agent receives a private noisy signal about the true state of the world and updates her beliefs as a convex combination of the Bayesian posterior belief generated by the observed signal and the prior beliefs of her neighbors. We show that under such a boundedly rational model of opinion formation, individuals will eventually hold correct forecasts about the observations of all agents, as long as the social network is strongly connected; that is, if there exists an information path connecting any two agents in the society. Moreover, we show that if the individuals’ observations are conditionally independent, they asymptotically learn the true underlying state of the world as if they were completely informed of all signals and updated their beliefs according to Bayes’ rule. Our main result therefore establishes that the agents can successfully aggregate information held by all other agents in the network, even if they are not in direct communication with one another, or are not aware of each other’s existence. Finally, the simplicity of our local update rule guarantees that the agents eventually achieve full learning, while at the same time, avoiding the highly complex computations that are essential for full Bayesian learning over the network.
1.2 New Models of Networks: Beyond Graphs

A key step in understanding the common underlying principles of networked systems is to apply the proper level of abstraction in modeling them. Although graphs are often adequate discrete abstractions for network phenomena, there are plenty of challenging applications and scenarios where the abstraction based on binary relations is too crude. More specifically, in many applications, instead of only looking at pairwise relations between agents, one might find it useful (or even necessary) to consider 3-way or 4-way relations as well. For example, a coalition of 3 agents is different from the 3 pairwise relations between any two of them. In such cases, using a simplicial complex to model the network can potentially lead to extraction of the information buried in higher order relations between different entities. To this end, the second part of this thesis is focused on presenting the theories of simplicial complexes and simplicial homology, as well as developing distributed protocols for extracting topological properties of the network.

Simplicial complexes, which are generalization of graphs to higher dimensions, can serve as general geometric representation for a broad spectrum of modeling problems [41]. For example, in many situations that a geometrical space is represented by a finite set of points sampled from it - what is known as point cloud data - it is natural to use the machinery of simplicial complexes and homology to recover the attributes of the original space from the sampled data [20, 28]. More recently, simplicial (and other type of) complexes have been used in the study of proteins and other molecules: the protein is modeled as a union of balls, one ball per atom, and the complex used is dual to this union [40–42]. Other applications range from modeling dynamical systems [5] to configuration spaces of graphs in robotics [53], as well as compression of data sets and coverage verification in sensor networks [30, 79].

Chapter 4 is dedicated to formally defining simplicial complexes. It also provides a
brief review of simplicial homology theory, which studies the algebraization of the first layer of geometry in simplicial complexes [61, 112]. Furthermore, we present the theory of combinatorial Laplacians corresponding to simplicial complexes and show their application in implementing distributed algorithms. The other goal of the chapter is to relate the presented concepts to their more well-known counterparts in graph theory.

In Chapter 5, we apply tools and concepts from algebraic topology presented in Chapter 4 to design and analyze distributed algorithms over networks. More specifically, we develop distributed algorithms for computation of sparse generators of homologies. Moreover, by focusing on the problem of coverage verification in coordinate-free sensor networks, we show how capturing higher order relations (beyond the pairwise relation, as in graphs) between different entities can lead to extraction of more information from the network. The chapter contains a distributed algorithm, built on the concept of combinatorial Laplacians, that verifies whether the coverage is successful and “hole-free”. We also present two new distributed algorithms which provide further information about the cover. The first algorithm, based on the ideas of de Silva and Ghrist [30] and Muhammad and Egerstedt [79], is capable of “localizing” coverage holes in a network of sensors without any metric information. Our second algorithm, also based on the tools presented in Chapter 4, is a novel approach for detecting redundancies in the sensor network. The presented algorithm is a distributed method for computing a minimal set of sensors required to cover the entire domain.
Part I

Information Aggregation Over Complex Networks
Chapter 2

Network Consensus Algorithms

Distributed consensus algorithms are one of the most well-known and widely studied distributed algorithms for information sharing and aggregation over networks. These algorithms were first introduced by DeGroot in a 1974 paper [33] and were later generalized by Tsitsiklis [105], Jadbabaie, Lin, and Morse [64] and many others. In consensus algorithms, as in any information aggregation algorithm, a group of agents each with some partial knowledge about a quantity of interest, share their information with their neighbors and update their state values accordingly. However, in benchmark consensus algorithms, agents update their information simply as a linear combination of the information held by their neighbors at any given time step. The central question in the study of distributed consensus algorithms is whether such simple linear updates can result in asymptotic consensus; that is, a state in which all agents eventually agree on the same value.

Distributed consensus algorithms have wide-spread applications, ranging from distributed and parallel computing [105, 106] to control of robotic and sensor networks [26]. These algorithms have also been suggested as possible explanation of different phenomena as diverse as flocking in birds [91] and belief formation in social networks [56, 62].

In this chapter, we first focus on the benchmark discrete-time consensus algorithm over
a deterministic switching network sequence. The first section contains necessary and sufficient conditions that guarantee asymptotic consensus. The presented conditions are based on various concepts borrowed from theory of non-homogeneous Markov chains, such as weak and strong ergodicity, and coefficients of ergodicity. This is followed by Section 2.2, where discrete-time consensus algorithms over randomly switching networks are studied. The contents of Section 2.2 are mainly from the works of Tahbaz-Salehi and Jadbabaie [101–103]. Our results show that for i.i.d. and ergodic stationary random graph processes, consensus is achieved if and only if the expected network is connected.

2.1 Distributed Consensus Algorithms

Consider a group of $n$ agents, labeled 1 through $n$, each with an initial state value that can be updated over discrete time steps $t \in \{0, 1, 2, \ldots \}$. Depending on the application, this state may correspond to the belief of an individual, the heading of a robot, or an unknown quantity that needs to be estimated. We denote the state of agent $i$ at time $t \geq 0$ with $x_i(t)$, and the vector of all states with $x(t) = [x_1(t) \ldots x_n(t)]^T$. At a given time step $t$, each agent has access to the state value of a subset of other agents called its neighbors. This neighborhood relation can be captured by a sequence of directed graphs $G(t) = \{V, E(t)\}$ indexed by time $t$: the $i$-th vertex of graph $G(t)$ corresponds to agent $i$, and the directed edge $(i, j)$ belongs to edge set $E(t)$, if agent $j$ has access to information held by agent $i$ at time $t$. By definition, the set of neighbors of agent $i$ coincides with the set of vertices adjacent to vertex $i$ in $G(t)$, denoted by $\mathcal{N}_i(t) = \{j \in V | (j, i) \in E(t)\}$.

In the benchmark network consensus algorithm, each agent updates its state value as a convex combination of its own and its neighbors’ state values in the previous time step.
More formally, the state value of agent \(i\) follows the linear iterative update

\[
x_i(t + 1) = W_{ii}(t)x_i(t) + \sum_{j \in \mathcal{N}_i(t)} W_{ij}(t)x_j(t),
\]

where \(W_{ij}(t) \geq 0\) is the weight that agent \(i\) assigns to the state value of agent \(j\) at time \(t\).

Note that the above update is local and distributed in the sense that agent \(i\)'s state value at time \(t + 1\) is only a function of the state values of its neighbors at time \(t\). Moreover, since it is assumed that the update in (2.1) is a convex combination, it has to be the case that

\[
\sum_{j \in \mathcal{N}_i(t)} W_{ij}(t) = 1.
\]

The central question regarding equation (2.1) is whether such a local update results in asymptotic consensus over the network. Clearly, it is as important to understand how this consensus value is related to the initial information held by each agent at time \(t = 0\).

Before investigating the answer to these questions, in order to simplify the notation, we set \(W_{ij}(t) = 0\) for \(j \not\in \mathcal{N}_i(t) \cup \{i\}\), and write the state update equation (2.1) in the more compact matrix form

\[
x(t + 1) = W(t)x(t).
\]

In the above equation, \(W(t)\) is an \(n \times n\) non-negative matrix whose \((i, j)\) entry is equal to \(W_{ij}(t)\). Note that \(W(t)\) has an entry equal to zero if the corresponding entry in the matrix \(A(t) + I_n\) is zero, where \(A(t)\) is the adjacency matrix of \(G(t)\) and \(I_n\) is the identity matrix of size \(n\). It is also useful to note that \(W(t)\) is stochastic for all \(t\).\(^1\)

We now formally define the concept of reaching asymptotic consensus for the dynamical system given in (2.2).

**Definition 1.** Dynamical system (2.2) reaches asymptotic *consensus*, if for any initial state value \(x(0) \in \mathbb{R}^n\), there exists \(x^* \in \mathbb{R}\) such that \(x_i(t) \to x^*\) as \(t \to \infty\) for all \(i \in \{1, \ldots, n\}\).

\(^1\)A matrix is said to be *stochastic* if it is entry-wise non-negative and all its row sums are equal to one. A stochastic matrix is called *doubly stochastic* if all its column sums are equal to one, as well. For more on the properties of stochastic matrices see, e.g., Berman and Plemmons [10].
As the name suggests, reaching asymptotic consensus requires the state values of all agents in the system to converge to the same common value \( x^* \) as \( t \) grows. Note that this is stronger than requiring the difference between state values of any two agents to converge to zero. It might be the case that the differences between the state values of any two agents converge to zero, while the state values themselves do not converge.

### 2.1.1 Ergodicity

If we denote the vector of initial state values by \( x(0) \), (2.2) implies

\[
x(t + 1) = W(t) \cdot \ldots \cdot W(1)W(0)x(0).
\]  

Therefore, in order to study conditions for asymptotic consensus, we need to investigate the behavior of infinite products of stochastic matrices. For this purpose, we borrow the concept of weak ergodicity of a sequence of stochastic matrices from the theory of Markov chains. Weak ergodicity, first introduced by Kolmogorov [68] in 1931, is the main concept that is used in proving convergence of infinite products of stochastic matrices.

**Definition 2.** The sequence \( \{W(t)\}_{t=0}^{\infty} = W(0), W(1), \ldots \) of \( n \times n \) stochastic matrices is weakly ergodic, if for all \( i, j, s = 1, \ldots, n \) and all integer \( p \geq 0 \)

\[
U_{i,s}(t,p) - U_{j,s}(t,p) \to 0
\]

as \( t \to \infty \), where \( U(t,p) = W(p+t) \cdot \ldots \cdot W(p+1)W(p) \) is the left product of the matrices in the sequence.

As the definition suggests, a sequence of stochastic matrices is weakly ergodic if the difference between any two rows of the product matrix converges to zero, as the number of terms in the product grows. A closely related concept is strong ergodicity of a matrix.
sequence.

**Definition 3.** A sequence of \( n \times n \) stochastic matrices \( \{W(t)\}_{t=0}^{\infty} \) is strongly ergodic, if for all \( i, s = 1, \ldots, n \) and all integer \( p \geq 0 \), \( U_{i,s}(t, p) \to d_s(p) \) as \( t \to \infty \), where \( U(t, p) \) is the left product and \( d_s(p) \) is a constant not depending on \( i \).

Strong ergodicity of a matrix sequence means that not only the difference between any two rows of the product matrix converges to zero, but also all entries of \( U(t, p) \) converge to some limit. Therefore, weak ergodicity of a matrix sequence is a necessary condition for strong ergodicity and is implied by it. The following proposition, however, shows that the two concepts are equivalent [23,97].

**Proposition 1.** Given a matrix sequence \( \{W(t)\}_{t=0}^{\infty} \) and their left products \( U(t, p) = W(t + p) \cdots W(p + 1) \), weak and strong ergodicity are equivalent.\(^2\)

*Proof:* Clearly, any strongly ergodic sequence is also weakly ergodic. So, it is sufficient to prove that weak ergodicity implies strong ergodicity. For any \( \epsilon > 0 \), weak ergodicity implies that for large enough \( t \), we have 

\[-\epsilon \leq U_{i,s}(t, p) - U_{j,s}(t, p) \leq \epsilon \]

uniformly for all \( i, j, s = 1, \ldots, n \). Since \( U(t + 1, p) = W(t + p + 1)U(t, p) \), we have

\[U_{i,s}(t, p) - \epsilon \leq U_{h,s}(t + 1, p) \leq U_{i,s}(t, p) + \epsilon,\]

which by induction implies that

\[U_{i,s}(t, p) - \epsilon \leq U_{h,s}(t + r, p) \leq U_{i,s}(t, p) + \epsilon,\]

\(^2\)To be more precise, Definitions 2 and 3 correspond to weak and strong ergodicity in the backward direction. Since this is the only type of ergodicity used in this thesis, we simply refer to these properties as ergodicity. It is important to note that weak and strong ergodicity are equivalent only for left (backward) products of stochastic matrices. This equivalence breaks down in the case of right (forward) products of stochastic matrices. More on these and similar concepts can be found in the book of Seneta [97] and references therein.
for all \(i, s, h = 1, \ldots, n\) and \(r \geq 0\). By setting \(h = i\), it is evident that \(U_{i,s}(t, p)\) is a Cauchy sequence and therefore, \(\lim_{t \to \infty} U_{i,s}(t, p)\) exists.

As the above proposition suggests, weak ergodicity describes a tendency to consensus as defined in Definition 1. Under weak ergodicity, matrix product \(U(t, p)\) converges to some rank one matrix \(1d(p)^T\), where \(d(p)\) is a normalized vector and \(1\) is the vector with all entries equal to one. Therefore, linear dynamical system (2.2) reaches asymptotic consensus for all initial conditions \(x(0)\) if matrix sequence \(\{W(t)\}_{t=0}^{\infty}\) is weakly ergodic. Note that the converse of this statement is not true in general. In other words, convergence of dynamical system (2.2) to consensus does not necessarily imply weak ergodicity of the matrix sequence. For instance, existence of a rank one matrix in the sequence implies asymptotic consensus, while it does not guarantee weak ergodicity. Nevertheless, in what follows, we investigate conditions for weak ergodicity of the sequence as it provides sufficient conditions for asymptotic consensus.

### 2.1.2 Coefficients of Ergodicity

We now present the definition of coefficients of ergodicity, which are key concepts in proving weak ergodicity results.

**Definition 4.** The scalar continuous function \(\tau(\cdot)\) defined on the set of \(n \times n\) stochastic matrices is called a *coefficient of ergodicity* if it satisfies \(0 \leq \tau(\cdot) \leq 1\). A coefficient of ergodicity is said to be *proper* if, \(\tau(W) = 0 \iff W = 1d^T\), where \(d\) is a vector of size \(n\) satisfying \(d^T1 = 1\).

The notion of coefficient of ergodicity was first introduced in a paper of Doeblin in 1937 and was later developed in the works of Dobrushin [35] and Hajnal [58]. It is an extremely useful and effective tool in dealing with infinite products of stochastic matrices. In fact, it is straightforward to show that weak ergodicity of a matrix sequence is equivalent
to
\[ \tau(U(t, p)) \longrightarrow 0 \quad \forall p \in \mathbb{N} \cup \{0\} \]
as \( t \to \infty \) for some proper coefficient of ergodicity \( \tau \).

Some of the more widely used examples of coefficients of ergodicity are:

\[ \tau_1(W) = \frac{1}{2} \max_{i,j} \sum_{s=1}^{n} |W_{is} - W_{js}|, \]
\[ a(W) = \max_{s} \max_{i,j} |W_{is} - W_{js}|, \]
\[ b(W) = 1 - \sum_{j=1}^{n} \left( \min_{i} W_{ij} \right), \]
\[ c(W) = 1 - \max_{j} (\min_{i} W_{ij}). \]

Note that \( c(\cdot) \) is an improper coefficient of ergodicity, while \( \tau_1(\cdot), a(\cdot), \) and \( b(\cdot) \) are proper.

It is also straightforward to verify that for any stochastic matrix \( W \),

\[ a(W) \leq \tau_1(W) \leq b(W) \leq c(W) \]

More examples can be found in the works of Seneta [96, 97].

The next proposition outlines necessary and sufficient conditions for weak ergodicity of a sequence of stochastic matrices, in terms of the coefficients of ergodicity of the matrices in the sequence. Moreover, since weak ergodicity implies consensus, it also provides a sufficient condition for asymptotic consensus of linear dynamical system (2.2). The proof of the theorem is from [23].

**Proposition 2.** Suppose \( \tau(\cdot) \) is a proper coefficient of ergodicity that for any \( m \geq 1 \) stochastic matrices \( F_k, k = 1, 2, \ldots, m \) satisfies

\[ \tau(F_m \ldots F_2 F_1) \leq \prod_{k=1}^{m} \tau(F_k). \]  \( (2.4) \)
Then the sequence \( \{W(t)\}_{t=0}^{\infty} \) is weakly ergodic if and only if there exists a strictly increasing sequence of non-negative integers \( \{t_r\}, r = 0, 1, 2, \ldots \) such that

\[
\sum_{r=1}^{\infty} \left[ 1 - \tau(W(t_r+1) \cdots W(t_r + 1)) \right] = \infty. \tag{2.5}
\]

**Proof:** Suppose that there exists a strictly increasing sequence of positive integers \( t_r \) such that (2.5) holds. Then, the inequality \( \log x \leq x - 1 \) implies that

\[
\sum_{r=1}^{\infty} \log \left[ \tau(W(t_{r+1}) \cdots W(t_r + 1)) \right] = -\infty,
\]

and as a result, \( \prod_{r=1}^{\infty} \tau(W(t_{r+1}) \cdots W(t_r + 1)) = 0 \). Since \( \tau \) is proper, (2.4) guarantees that the sequence is weakly ergodic. This proves that (2.5) is a sufficient condition for weak ergodicity of \( \{W(t)\} \).

If we assume weak ergodicity, then it must be the case that \( \tau[U(t,p)] \to 0 \) as \( t \to 0 \), for \( p \geq 0 \). Let \( 0 < \epsilon < 1 \) be fixed. Then define the sequence \( \{t_r\} \) recursively by choosing \( t_0 \) arbitrarily, and \( t_{r+1} \) once \( t_r \) has been determined so that

\[
\tau[U(t_{r+1} - t_r - 1, t_r + 1)] \leq \epsilon.
\]

Note that since \( \tau[U(t,p)] \) converges to zero for all \( p \), given \( t_r \), one can always find \( t_{r+1} \) such that the above inequality is satisfied. Therefore, for the given sequence of integer \( \{t_r\} \), the sum in (2.5) diverges whenever the matrix sequence \( \{W(t)\} \) is weakly ergodic.

The above proposition, by stating a necessary and sufficient condition for weak ergodicity, provides a sufficient condition for asymptotic consensus in terms of the coefficient of ergodicity of the matrices in the sequence. However, it is silent on the properties of the consensus value that all agents eventually agree on. In fact, except for a few special cases, characterizing the consensus value in terms of the matrix sequence \( \{W(t)\} \) is still an open
problem. The special case for which one can analytically compute the consensus value is when all weight matrices \( W(t) \) share the same left eigenvector corresponding to their unit eigenvector; that is, there exists \( a \in \mathbb{R}^n \) such that \( a^T W(t) = a^T \) for all \( t \geq 0 \). In such a case, \( a^T U(t, 0) = a^T \) for all \( t \). Therefore, if the matrix sequence satisfies (2.5) for some proper coefficient of ergodicity \( \tau \), then, as \( t \) grows to infinity, \( x(t) \to \left[ a^T x(0) / (a^T 1) \right] 1 \). Clearly, when all weight matrices are doubly stochastic, the consensus value will be equal to the average of the initial conditions.

2.1.3 Consensus and Joint Connectivity

Proposition 2 provides a necessary and sufficient condition for the weak ergodicity of the sequence \( \{W(t)\} \), as well as a sufficient condition for asymptotic consensus, when the agents update their state values according to (2.1). However, it is important to interpret condition (2.5) in terms of the properties of the matrix sequence and eventually (and more importantly), in terms of the properties of the underlying network.

In order to explore this relationship, we focus on a specific coefficient of ergodicity \( \tau_1 \), defined as \( \tau_1(W) = \frac{1}{2} \max_{i,j} \sum_{s=1}^n |W_{is} - W_{js}| \), which is proper and satisfies the submultiplicative property (2.4). We also impose a simplifying assumption on the weights that each agent assigns to the state of its neighbors.

**Assumption 1.** The weight matrices in the sequence \( \{W(t)\} \) are such that for all \( t > 0 \), \( W_{ij}(t) \in \{0\} \cup [\alpha, 1] \) and \( W_{ii}(t) > \alpha \) for some \( \alpha > 0 \).

This assumption is meant to guarantee that whenever \( \tau_1(W) < 1 \) for some stochastic matrix \( W \), then, \( \tau_1(W) < 1 - \epsilon \) for some \( \epsilon > 0 \). Therefore, under Assumption 1, for condition (2.5) to hold, it is sufficient that there are infinitely many contractive matrices with respect to \( \tau_1 \) in the sequence \( \{W(t)\} \). Note that since \( \tau_1(W) \in [0, 1] \), the rest of the matrices in the sequence are non-expansive with respect to \( \tau_1 \). As Proposition 2 suggests,
this idea of contraction can be taken one step further, in the sense that as long as there are infinitely many intervals over which the product matrix is contractive with respect to $\tau_1$, then the sum in (2.5) diverges and therefore, $U(t, p)$ converges for all $p$.

Notice that for a given stochastic matrix $W$, $\tau_1(W)$ is subunit if and only if $W$ is scrambling. Moreover, it is well-known that a weight matrix is scrambling if and only if its corresponding graph is neighbor-shared, that is, every pair of vertices $i, j \in V$ are either neighbors, or there exists $k \in V$ which is a neighbor of both $i$ and $j$. Therefore, if the graph representing the underlying network of the system is neighbor-shared infinitely often, and Assumption 1 holds, then the agents in the network reach asymptotic consensus. However, as stated earlier, it is not necessary to have scrambling matrices (and hence neighbor-shared graphs) appearing infinitely often: as long as there are infinitely many intervals over which the matrix products are scrambling, weak ergodicity is implied. This statement can be interpreted in terms of the graph sequence as well. For this purpose, we need the following two definitions [64].

**Definition 5.** The union of a collection of (possibly directed) graphs $\{G_1, G_2, \ldots, G_m\}$ each with the vertex set $V$, is a graph $G$ with the same vertex set $V$, and edge set equaling the union of the of the edge sets of all the graphs in the collection; i.e., $E = \bigcup_{k=1}^{m} E_k$, where $E_k$ is the edge set of graph $G_k$.

**Definition 6.** A collection of graphs is said to jointly satisfy some property $\mathcal{P}$, if the union of its members satisfies that property $\mathcal{P}$.

The following lemma relates the above definitions with the scrambling property of the product matrices. The proof is omitted and can be found in [64].

**Lemma 1.** The product $W_m \ldots W_2 W_1$ of stochastic matrices is scrambling if and only if

\[ A \text{ matrix } W \text{ is scrambling if for each pair of indices } (i, j) \text{ there exists } k \text{ such that } W_{ik} \text{ and } W_{jk} \text{ are both nonzero.} \]
Thus, one can state the following proposition:

**Proposition 3.** Suppose that the weight matrices in the sequence \( \{W(t)\} \) satisfy Assumption 1. Then the sequence is weakly ergodic if and only if there exists a sequence of positive integers \( \{t_r\}, r = 0, 1, 2, \ldots \) such that the collection of graphs \( \{G(t_r + 1), G(t_r + 2), \ldots, G(t_{r+1})\} \) are jointly neighbor-shared.

To summarize, as long as the collection of graphs are infinitely often neighbor-shared, the sequence of weight matrices representing them is weakly ergodic, and therefore, asymptotic consensus is guaranteed. The following theorem, which is simply a consequence of Proposition 3, formalizes this statement as a sufficient condition for asymptotic consensus in terms of the properties of the underlying network.

**Theorem 1.** Consider the network of agents that update their states based on (2.1). Also suppose that the weights assigned by the agents satisfy Assumption 1. Then the agents reach asymptotic consensus, if there exists a sequence of positive integers \( \{t_r\}, r = 0, 1, 2, \ldots \) such that the collection of graphs \( \{G(t_r + 1), G(t_r + 2), \ldots, G(t_{r+1})\} \) are jointly neighbor-shared.

As a final remark, note that most results in the literature, such as [64, 78, 85], state the existence of infinitely often jointly strongly rooted graphs as a sufficient condition for asymptotic consensus. It is trivial to verify that a sequence of graphs is infinitely often jointly strongly rooted, if and only if it is infinitely often jointly neighbor-shared.
2.2 Consensus over Random Networks

In the previous section, we studied the network consensus problem, when the underlying communication network of the system changes over time. The main standing assumption was that changes in the network topology were deterministic. The relation between convergence of consensus problems and weak ergodicity was discussed, and Proposition 2 provided a sufficient condition for reaching asymptotic consensus in terms of the coefficients of ergodicity of the matrices in the sequence. Moreover, the interpretation of weak ergodicity in terms of joint connectivity of the graph sequence was explored.

In this section, we present conditions under which a group of agents who update their state values according to dynamical system in (2.2) reach a consensus, when the network changes randomly over time. The importance of studying such problems is due to the fact that, in many real-world network applications, the availability of communication links between different entities is usually random. This randomness, in a sensor network scenario, can be due to link failures, node failures, interference, or existence of physical obstacles interrupting communication. Due to the sensors’ reliance on battery power and their typically minimal communication capabilities, distributed ad hoc sensor networks are prone to such random failures. Similarly, in social networks, communication and information exchange between individuals is subject to the randomness inherent to social interactions.

The results of this section are presented in two parts: The first part, mainly based on Tahbaz-Salehi and Jadbabaie [102], focuses on convergence of consensus algorithms under the condition that the random process defining the underlying network of the system is an i.i.d. process. The presented results generalize the works of Hatano and Mesbahi [60], Wu [109], and Porfiri and Stilwell [90]. The second part is dedicated to deriving necessary and sufficient conditions for asymptotic consensus when the graph sequence is ergodic and stationary.
2.2.1 Independent and Identically Distributed Random Networks

We start the formal analysis of consensus algorithms over randomly switching graphs by defining the underlying stochastic process. Let \((\Omega_0, B, \mu)\) be a probability space, where \(\Omega_0 = S_n = \{\text{set of stochastic matrices of order } n \text{ with strictly positive diagonal entries}\}\), \(B\) is the Borel \(\sigma\)-algebra of \(\Omega_0\), and \(\mu\) is a probability measure defined on \(\Omega_0\). Define the product probability space as \((\Omega, F, P) = \prod_{t=0}^{\infty} (\Omega_0, B, \mu)\). By definition, the elements of the product space are of the following form:

\[
\Omega = \{ (\omega_0, \omega_1, \cdots) : \omega_t \in \Omega_0 \}
\]

\[
F = B \times B \times \cdots
\]

\[
P = \mu \times \mu \times \cdots
\]

The above equations mean that the coordinates of the infinite dimensional vector \(\omega \in \Omega\) are independent and identically distributed (i.i.d.) stochastic matrices with positive diagonals.

Now consider the following random discrete-time dynamical system:

\[
x(t+1) = (W(t)[\omega]) x(t),
\]

(2.6)

where \(t \in \{0, 1, 2, \cdots\}\) is the discrete time index, \(x(t) \in \mathbb{R}^n\) is the state vector at time \(t\) and the mapping \(W(t) : \Omega \to S_n\) is the \(t\)-th coordinate function, which for all \(\omega = (\omega_0, \omega_1, \cdots) \in \Omega\) is defined as \(W(t)[\omega] = \omega_t\). As a result, similar to update (2.2) in the previous section, the above equation defines a stochastic linear dynamical system, in which the weight matrices are drawn independently from the common distribution \(\mu\). For notational simplicity, \(W(t)[\omega] \) is denoted by \(W(t)\).

The next two definitions state what we mean by convergence to consensus in probabilistic terms.

**Definition 7.** Dynamical system (2.6) reaches consensus in probability, if for any initial
state value $x(0) \in \mathbb{R}^n$, there exists a random variable $x^* \in \mathbb{R}$ such that $x_i(t) \to x^*$ in probability for all $i \in \{1, \ldots, n\}$. That is, for any $\epsilon > 0$, $\mathbb{P}(|x_i(t) - x^*| > \epsilon) \to 0$, as $t \to \infty$ for all $i$.

This notion of reaching state agreement asymptotically, which is addressed in [109], is weaker than reaching consensus almost surely, defined below.

**Definition 8.** Dynamical system (2.6) reaches **consensus almost surely**, if for any initial state value $x(0) \in \mathbb{R}^n$ there exists a random variable $x^* \in \mathbb{R}$ such that $x_i(t) \to x^*$ almost surely.

Reaching almost sure consensus is stronger than reaching consensus in probability. In this case, not only the probability of the events $\{|x_i(t) - x^*| > \epsilon\}$ goes to zero for an arbitrary $\epsilon > 0$ as $t \to \infty$, but also such events occur only finitely many times [37]. Note that even under almost sure consensus, there are realizations of the matrix sequence that do not lead to asymptotic consensus, but such sample paths are of measure zero.

As stated in Section 2.1.1, asymptotic consensus is closely related to the concept of weak ergodicity of the matrix sequence capturing the weights used in the belief update equations. In what follows, we examine conditions for weak ergodicity of a random matrix sequence. However before presenting our results, it is constructive to note that weak ergodicity of a sequence is a **tail event**; that is, changing finitely many matrices in the sequence does not affect its ergodicity properties. To see this, notice that by definition, $\{W(t)\}$ is weakly ergodic if $U(t, p)$ converges to a rank one matrix for all $p \geq 0$. Clearly, if a finite subset of these matrices are changed and new products $\tilde{U}(t, p)$ are formed, there still exists a large enough $\bar{p}$ for which $U(t, \bar{p}) = \tilde{U}(t, \bar{p})$. Therefore, the new sequence is weakly ergodic, if only if the original one is.

The fact that weak ergodicity is a tail event has an important consequence for i.i.d. matrix sequences. Kolmogorov’s 0-1 Law [37] states that for i.i.d. sequences, any tail
event $B$ is trivial; i.e., $\mathbb{P}(B) \in \{0, 1\}$. As a result, a sequence \{\(W(t)\)\} of independent and identically distributed matrices is weakly ergodic either almost surely or almost never. In the next theorem, we state the criterion that distinguishes these two cases. We show that information in the average weight matrix suffices to predict the long-run behavior of the left product matrices $U(t, p)$.

**Theorem 2.** Let \{\(W(t) : t \geq 0\)\} = \(W(0), W(1), \ldots\) denote a sequence of stochastic matrices with positive diagonals which are independent and identically distributed. This random sequence is weakly ergodic almost surely, if and only if $|\lambda_2[\mathbb{E}W(1)]| < 1$, where $\lambda_2$ is the eigenvalue with the second largest modulus.\(^4\)

**Proof:** First, we prove the necessity. Suppose $|\lambda_2[\mathbb{E}W(1)]| = 1$. Since all weight matrices have positive diagonals, $\mathbb{E}W(1)$ has strictly positive diagonal entries as well. Hence, if $\mathbb{E}W(1)$ is irreducible, then it is primitive and as a result of the Perron-Frobenius theorem [10], $|\lambda_2[\mathbb{E}W(1)]| < 1$, which contradicts our assumption. Therefore, $|\lambda_2[\mathbb{E}W(1)]| = 1$ implies reducibility of $\mathbb{E}W(1)$. As a result, without loss of generality, one can label the vertices of the graph such that $\mathbb{E}W(1)$ gets the following block triangular form

$$
\mathbb{E}W(1) = 
\begin{bmatrix}
Q_{11} & 0 & \cdots & 0 \\
Q_{21} & Q_{22} & \cdots & 0 \\
\vdots & \vdots & \ddots & \vdots \\
Q_{s1} & Q_{s2} & \cdots & Q_{ss}
\end{bmatrix},
$$

where each $Q_{ii}$ is an irreducible matrix corresponding to the vertices in the $i$-th communication class, which we denote by $\alpha_i$.\(^5\) Since $\lambda_1[\mathbb{E}W(1)] = |\lambda_2[\mathbb{E}W(1)]| = 1$, submatrices corresponding to at least two of the classes have unit spectral radii (because of irreducibility).

\(^4\)Note that $\mathbb{E}W(1) = \mathbb{E}W(t)$ for all $t$, because of the i.i.d. assumption.

\(^5\)A set of vertices form a communication class if there exists a directed path connecting every vertex in that class to every other vertex in that class. For more see the Appendix.
ity and aperiodicity of $Q_{ii}$’s, the multiplicity of the unit-modulus eigenvalue of each one cannot be more than one). Therefore, Lemma 7 in the Appendix implies,

$$\exists i \neq j \text{ s.t. } \alpha_i \text{ and } \alpha_j \text{ are both initial classes,}^6$$

or equivalently, $Q_{ir} = 0$ for all $r \neq i$ and $Q_{jl} = 0$ for all $l \neq j$. In other words, matrix $\mathbb{E}W(1)$ has two orthogonal rows. Since all weight matrices are non-negative, $\mathbb{E}W(1)$ has the same type$^7$ as $W(t)$ does for all time $t$, with probability one. Therefore, $U(t, 0) = W(t) \cdots W(1)W(0)$ has two orthogonal rows almost surely for any $t$.$^8$ This means that there are initial conditions for which random discrete-time dynamical system (2.2) reaches consensus with probability zero. Since weak ergodicity of $\{W(t)\}$ is a subset of convergence of (2.2) to consensus, the random sequence of weight matrices is weakly ergodic almost never.

Now consider the reverse implication. When $|\lambda_2 [\mathbb{E}W(1)]| < 1$, Lemma 7 in the Appendix implies that the graph of the network has exactly one initial class. In other words, there exists an agent $i$, such that for all $j \neq i$, there exists a sequence $i = j(1), \ldots, j(s_j) = j$ of agents for which $[\mathbb{E}W(1)]_{j(q+1), j(q)} > 0$. Equivalently, there exists a path of length $s_j - 1$ from some node $i$ to any other node $j$ in the expected graph of the network. As a result, there exists $\epsilon > 0$ such that

$$\mathbb{P} \left[ (W(1))_{j(q+1), j(q)} > \epsilon \right] > 0 \quad \forall q = 1, 2, \ldots, s_j - 1.$$

$^6$A set of vertices form an initial class if there is no directed path connecting any vertex outside of that set to the vertices of the class. For a more detailed treatment, see Section A.2 in the Appendix.

$^7$The type of a matrix refers to its pattern of zero and non-zero elements.

$^8$This is because of the fact that this event is the intersection of countably many unit-measure events.
for all vertices \( j \). Hence, the second Borel-Cantelli lemma [37, p.49] implies

\[
P \left[ (W(t))_{j(q+1), j(q)} > \epsilon \text{ infinitely often} \right] = 1 \quad 0 < q < s_j
\]

for all \( j \neq i \). Any countable intersection of these events also occurs with probability one. As a result, there exists a deterministic time \( T \) for which

\[
P \left[ \delta (W(T - 1) \ldots W(1)W(0)) > \zeta \right] > 0
\]

for some \( \zeta > 0 \), where \( \delta (W) = \max_j (\min_i W_{ij}) \). In other words, there exists a deterministic time \( T \), for which all the entries of at least one column of the matrix product \( U(T - 1, 0) = W(T - 1) \ldots W(1)W(0) \) is bounded away from zero with positive probability. Now, once again the second Borel-Cantelli lemma implies that such an event occurs infinitely often almost surely, i.e.,

\[
P \left[ \delta (W((r + 1)T) \ldots W(rT + 1)) > \zeta \quad \text{for infinitely many } r \right] = 1.
\]

Consequently, by defining \( t_r = rT \), we have

\[
\delta (W(t_{r+1}) \ldots W(t_r + 1)) > \zeta \quad \text{infinitely often, almost surely}
\]

Since \( \delta(W) = 1 - c(W) \leq 1 - \tau_1(W) \) for any stochastic matrix \( W \), we have,

\[
\sum_{r=1}^{\infty} \left[ 1 - \tau_1 (W(t_{r+1}) \ldots W(t_r + 1)) \right] = \infty \quad \text{almost surely},
\]

which is exactly (2.5), the sufficient condition for weak ergodicity. Therefore, the sequence is weakly ergodic almost surely. Note that, as stated earlier, \( \tau_1(\cdot) \) is a proper coefficient of ergodicity which satisfies the submultiplicative property (2.4).
Theorem 2 suggests that the information in the average weight matrix $\mathbb{E} W(1)$ suffices to predict the long-run behavior of left product matrices $U(t,p)$. The following corollary states that the same information is sufficient to extract the asymptotic convergence properties of linear dynamical system (2.6).

**Corollary 1.** Distributed update (2.6) reaches consensus almost surely, if and only if $|\lambda_2(\mathbb{E} W(1))| < 1$. Otherwise, it reaches asymptotic consensus almost never.

**Proof:** According to Theorem 2, $|\lambda_2(\mathbb{E} W(1))| < 1$ guarantees weak ergodicity with probability one. As a result, the event of asymptotic consensus occurs on a full-measure set, as it is a superset of the weak ergodicity event. To prove the reverse implication, note that when $\mathbb{E} W(1)$ has more than one unit-modulus eigenvalues, as in the proof of Theorem 2, its corresponding graph has more than one initial class, which implies that $\mathbb{E} W(1)$ has two orthogonal rows. Since $\Omega_0$ is a subset of nonnegative matrices, $W(t)$ has the same type as $\mathbb{E} W(1)$ for all time $t$, with probability one. Therefore, $U(t,0) = W(t) \ldots W(1)W(0)$ has two orthogonal rows almost surely for any $t$. This means that the random discrete-time dynamical system (2.2) reaches a consensus with probability zero.

In summary, $|\lambda_2(\mathbb{E} W(1))| < 1$ provides a necessary and sufficient condition for almost sure asymptotic consensus in (2.6). It is important to note the difference between the deterministic and the i.i.d. case. While in the i.i.d. case asymptotic consensus occurs almost surely if and only if weak ergodicity occurs almost surely, in the deterministic case weak ergodicity is only a sufficient condition for consensus. This difference is due to the fact that although the set of matrix sequences which result in consensus but are not weakly ergodic is non-empty, it has zero measure when weight matrices are independent and identically distributed.

As a final remark, note that condition $|\lambda_2(\mathbb{E} W(1))| < 1$ as a necessary and sufficient condition for almost sure asymptotic consensus is quite intuitive. In fact, when
|λ_2 (E W(1))| is sub-unit, as was shown in the proof of Theorem 2, there exists a sequence of integer numbers t_r, r = 0, 1, 2, ... such that \( \tau_1 (W(t_{r+1}) \ldots W(t_r + 1)) < 1 - \zeta \) infinitely often with probability one. As a consequence, the product of matrices \{W(t_{r+1}), \ldots, W(k_r + 1)\} is scrambling infinitely often. This means that the collection of graphs \{G(t_r + 1), \ldots, G(t_{r+1})\} is jointly neighbor-shared for infinitely many \( r \), almost surely. This infinite often connectivity over time guarantees the possibility of information flow in the network over time, and therefore results in asymptotic consensus with probability one. On the other hand, when \( \lambda_2 (E W(1)) \) is on the unit circle, no such sequence exists and therefore, there are at least two classes of vertices in the network which never have access to each other. Clearly, in such a case, reaching consensus is not in general possible.

**Asymptotic Mean of the Consensus Value**

As stated in Theorem 2, the i.i.d. process \( \{W(t) : t \geq 0\} \) of stochastic matrices is weakly ergodic almost surely, if and only if, \( |\lambda_2 [E W(1)]| < 1 \). In other words, if the expected weight matrix has a unique unit-modulus eigenvalue, then left products of the stochastic matrices in the sequence converge to a rank one matrix on almost all sample paths; that is, there exists a random non-negative vector \( d \) satisfying \( 1^T d = 1 \) such that \( U(t, 0) \to 1d^T \) almost surely, as \( t \to \infty \). Thus, the asymptotic consensus value of linear dynamical system (2.6) is the random variable \( x^* = d^Tx(0) \).

A natural question to ask is whether one can determine the distribution of this random consensus value. Unfortunately, except for some very special cases, computing the distribution of the consensus value is far from trivial. Nonetheless, it is possible to compute its first two moments. In the remainder of this section, we compute the mean and variance of the random consensus value \( x^* = d^Tx(0) \) for a general i.i.d. process.

Computing the mean of the consensus value is straightforward. We showed that whenever \( |\lambda_2 [E W(t)]| \) is subunit, then \( W(t) \ldots W(1)W(0) \to 1d^T \) almost surely, for some
random stochastic vector $d$. By taking expectations and applying the dominated convergence theorem [37], one obtains

$$
\mathbb{E} [W(t) \ldots W(1)W(0)] \rightarrow \mathbb{E} [1d^T],
$$

which implies $[\mathbb{E}W(1)]^t+1 \rightarrow 1(\mathbb{E}d^T)$, due to independence. Therefore, by the Perron-Frobenius theorem, $\mathbb{E}d$ is simply equal to the normalized left eigenvector of $\mathbb{E}W(1)$, corresponding to its unit eigenvalue. Thus, the mean of the asymptotic consensus value $x^*$ conditional on the initial condition $x(0)$ is given by $\mathbb{E} [x^*|x(0)] = x(0)^T v_1 [\mathbb{E}W(1)]$, where $v_1(\cdot)$ denotes the normalized left eigenvector corresponding to the unit eigenvalue. For example, if the expected weight matrix is symmetric (and hence, doubly stochastic), then the expected consensus value is equal to the average of the initial conditions.

**Asymptotic Variance**

In order to compute the variance, first note that

$$
\frac{1}{n} [W(t) \ldots W(0)]^T [W(t) \ldots W(0)] \rightarrow dd^T \ a.s.,
$$

which can be rewritten as

$$
\frac{1}{n} \text{vec} \left([W(t) \ldots W(0)]^T(W(t) \ldots W(0))\right) =
$$

$$
= \frac{1}{n} \left[W(0)^T \otimes W(0)^T\right] \left[W(1)^T \otimes W(1)^T\right] \ldots \left[W(t)^T \otimes W(t)^T\right] \text{vec}(I_n)
$$

$$
\rightarrow \text{vec}(dd^T) \ a.s.,
$$

---

9The assumption $|\lambda_2 [\mathbb{E}W(1)]| < \lambda_1 [\mathbb{E}W(1)] = 1$ guarantees that such an eigenvector exists and is unique.
where vec is the vectorization operator, ⊗ denotes the Kronecker product and $I_n$ is the identity matrix of size $n$.\textsuperscript{10} By applying the dominated convergence theorem once again, and using the assumption that the weight matrices are independent, we get
\[
\frac{1}{n} \left[ \mathbb{E}(W(1)^T \otimes W(1)^T) \right]^{t+1} \text{vec}(I_n) \longrightarrow \mathbb{E}[^{\text{vec}}(dd^T)] = \mathbb{E}(d \otimes d).
\]
Hence, the Perron-Frobenius theorem implies
\[
\mathbb{E}(d \otimes d) = \frac{1}{n} \mathbf{v}_1 \left( \mathbb{E} \left[ W(1) \otimes W(1) \right] \right) (1_{n^2} \text{vec}(I_n)) = \mathbf{v}_1 \left( \mathbb{E} \left[ W(1) \otimes W(1) \right] \right),
\]
where again, $\mathbf{v}_1(\cdot)$ denotes the normalized left eigenvector corresponding to the unit eigenvalue. Therefore, the covariance matrix of the random vector $d$ satisfies
\[
\text{vec}(\text{cov}(d)) = \text{vec} \left( \mathbb{E}dd^T \right) - \text{vec} \left( \mathbb{E}d \mathbb{E}d^T \right) = \mathbb{E}(d \otimes d) - \mathbb{E}d \otimes \mathbb{E}d = \mathbf{v}_1 \left( \mathbb{E} \left[ W(1) \otimes W(1) \right] \right) - \mathbf{v}_1 \left[ \mathbb{E}W(1) \right] \otimes \mathbf{v}_1 \left[ \mathbb{E}W(1) \right].
\]
By combining all the above, one can compute the conditional variance of the random consensus value $x^* = d^T x(0)$ in terms of the moments of the weight matrices:
\[
\text{var} \left[ x^* | x(0) \right] = [x(0) \otimes x(0)]^T \mathbf{v}_1 \left( \mathbb{E} \left[ W(1) \otimes W(1) \right] \right) - [x(0)^T \mathbf{v}_1 \left( \mathbb{E}W(1) \right)]^2 \quad (2.7)
\]
Given the expression for variance in (2.7), we can now derive conditions under which the asymptotic consensus value is deterministic, even though the network switches randomly over time. We have the following proposition.

**Proposition 4.** The consensus value $x^*$ has a degenerate distribution, if and only if all
\textsuperscript{10}In deriving the above expression, we have used the identity $\text{vec}(ABC) = (C^T \otimes A) \text{vec}(B)$.\noindent
weight matrices share the same left eigenvector corresponding to the unit eigenvector with probability one.

Proof: We first prove the sufficiency. Suppose that all weight matrices share the same common left eigenvector $a$, with probability one; that is, $a^T W(1) = a^T$ almost surely. Therefore, $a^T E W(1) = a^T$. Moreover,

$$(a^T \otimes a^T) [W(1) \otimes W(1)] = a^T \otimes a^T \quad \text{almost surely.}$$

Thus, $(a^T \otimes a^T) E [W(1) \otimes W(1)] = a^T \otimes a^T$. Consequently, $\text{vec}(\text{cov}(d)) = 0$, implying that the conditional variance of the consensus value in (2.7) is zero. Therefore, $P \left( x^* = a^T x(0) \mid x(0) \right) = 1$.

To prove the reverse implication, assume that $\text{var} \left[ x^* \mid x(0) \right] = 0$ for all initial conditions $x(0)$. We denote the left eigenvector of $E W(1)$ by $a$. Equation (2.7) implies that

$$v_1 (E [W(1) \otimes W(1)]) = v_1 [E W(1)] \otimes v_1 [E W(1)] = a \otimes a.$$ 

Therefore,

$$[a^T E W(1)] \otimes [a^T E W(1)] = (a \otimes a)^T E [W(1) \otimes W(1)]
= E \left[ (a^T W(1)) \otimes (a^T W(1)) \right],$$

which implies that

$$E \left[ \left( a^T W(1) - E (a^T W(1)) \right) \otimes \left( a^T W(1) - E (a^T W(1)) \right) \right] = 0,$$

and as a result, $a^T W(t) = E [a^T W(1)] = a^T$ almost surely.

One special case of interest is when weight matrices are doubly stochastic almost surely.
In such a case, all matrices have vector 1 as their common left eigenvector at all times and therefore, all entries of the state vector converge to the deterministic value $\frac{1}{n} [1^T x(0)]$, the average of the initial state values, with probability one. This special case, known as average consensus, is studied by Boyd et al. in [15].

### 2.2.2 Ergodic Stationary Random Networks

Theorem 2 and the results presented so far concentrated on consensus algorithms over randomly switching graphs, with the assumption that the graphs are independent and identically distributed over time. However, in most realistic cases, there exists a strong correlation between the realizations of the graph capturing the network’s topology over time. For example, the existence of a communication link in a wireless network at a given time instance is highly correlated with the existence of that link at the previous time steps. Similarly, in a social network scenario, it is plausible to assume that the probability of two individuals communicating on a given instance depends on whether they have communicated before or not. In what follows, we relax the i.i.d. assumption and assume that the weight matrices $W(t)$ are generated by an ergodic and stationary process. The reason for concentrating on ergodic and stationary models of network evolution is that not only these models can capture correlations over time, but also the fact that they contain a lot of important frameworks, such as i.i.d., Markovian and ARMA models, as special cases. The results of this section are mainly from the paper of Tahbaz-Salehi and Jadbabaie [103] which is based on the work of Picci and Taylor [88].

Similar to the i.i.d. case, let $(\Omega_0, B)$ be a measurable space, where $\Omega_0 = \{\text{set of stochastic matrices of order } n \text{ with strictly positive diagonal entries}\}$ and $B$ is the Borel $\sigma$-algebra.
on $\Omega_0$. Consider a probability measure $\mathbb{P}$ defined on the sequence space $(\Omega, \mathcal{F})$ defined as

$$\Omega = \{(\omega_0, \omega_1, \ldots) : \omega_k \in \Omega_0\}$$

$$\mathcal{F} = \mathcal{B} \times \mathcal{B} \times \ldots,$$

so that $(\Omega, \mathcal{F}, \mathbb{P})$ forms a probability space. If we let $\varphi : \Omega \to \Omega$ be the shift operator defined as $\varphi(\omega_0, \omega_1, \ldots) = (\omega_1, \omega_2, \ldots)$ and let $W(\omega) = \omega_0$, then we can define a sequence of stochastic matrices as $[W(t)](\omega) = W(\varphi^t \omega)$. For notational simplicity, we denote $[W(t)](\omega)$ by $W(t)$.

**Definition 9.** The sequence of random stochastic matrices $W(0), W(1), \ldots$ is stationary if the families $\{W(t_1), W(t_2), \ldots, W(t_r)\}$ and $\{W(t_1 + h), W(t_2 + h), \ldots, W(t_r + h)\}$ have the same joint distribution for all $t_1, t_2, \ldots, t_r$ and all $h > 0$.

The above definition states that process $\{W(t) : t \geq 0\}$ is stationary if all of its finite dimensional distributions are invariant under time shifts. Equivalently, one can define stationarity as the case that the shift operator is a measure-preserving transformation, i.e., $\mathbb{P}(\varphi B) = \mathbb{P}(B)$ for all sets $B \in \mathcal{F}$. Clearly, any i.i.d. sequence of random matrices is stationary.

**Definition 10.** Consider the probability space $(\Omega, \mathcal{F}, \mathbb{P})$ and suppose that the shift operator $\varphi : \Omega \to \Omega$ is measure-preserving. $\varphi$ is said to be ergodic if every invariant set $B \in \mathcal{F}$ is trivial.

In other words, the transformation $\varphi$ is ergodic if for any event $B \in \mathcal{F}$ satisfying $\mathbb{P}(B \Delta \varphi^{-1} B) = 0$, we have $\mathbb{P}(B) \in \{0, 1\}$, where $\Delta$ denotes the symmetric difference between the two sets.\(^{11}\) Given the above definitions, we say random matrix process $\{W(t) : t \geq 0\}$ is ergodic stationary, if the shift operator defined over $(\Omega, \mathcal{F}, \mathbb{P})$ is measure-preserving and ergodic. For example, a time-invariant Markov chain with its\(^{11}\)The symmetric difference between two sets $X$ and $Y$ is defined as $X \Delta Y = (X \setminus Y) \cup (Y \setminus X)$.\(^{11}\)
unique stationary distribution as the initial distribution forms a stationary and ergodic process. Clearly, any i.i.d. sequence of matrices is also ergodic stationary.

For a sequence of stationary and ergodic random graph process we have the following theorem, which is a generalization of Theorem 2.

**Theorem 3.** Let \( \{W(t) : t \geq 0\} = W(0), W(1), \ldots \) denote a sequence of stochastic matrices with positive diagonals generated by an ergodic stationary process. This random sequence is weakly ergodic almost surely, if and only if \( |\lambda_2[\mathbb{E}W(1)]| < 1 \), where \( \lambda_2 \) is the eigenvalue with the second largest modulus.\(^{\text{12}}\)

This theorem states that, exactly similar to the i.i.d. case studied in the previous section, as long as the expected graph representing the network is strongly rooted, all agents reach asymptotic consensus. The proof is also very similar to the proof of Theorem 2 already stated. The only step that needs modification is the application of the second Borel-Cantelli lemma which requires independence. This issue can be resolved by using the following lemma which is a consequence of Birkhoff’s ergodic theorem. The lemma states that any event with positive probability at a given time will occur infinitely often almost surely.

**Lemma 2.** Suppose \( W(0), W(1), \ldots \) is an ergodic stationary process of stochastic \( n \times n \) matrices. If the event \( \{W(t) \in A\} \) has positive probability \( p > 0 \), then such events occur infinitely often almost surely; that is, \( \mathbb{P}[W(t) \in A \text{ for infinitely many } t] = 1 \).

**Proof:** Since the process \( \{W(t) : t \geq 0\} \) is ergodic stationary, so is the process \( \{\mathbb{I}_{\{W(t) \in A\}} : t \geq 0\} \), where \( \mathbb{I} \) is the indicator function. Therefore, by Birkhoff’s ergodic theorem [25,37],

\[
\frac{1}{T} \sum_{t=1}^{T} \mathbb{I}_{\{W(t) \in A\}} \rightarrow \mathbb{P}\{W(0) \in A\} = p \quad \text{almost surely},
\]

\(^{\text{12}}\)Note that, due to the stationarity assumption, \( \mathbb{E}W(t) \) is time-invariant. Therefore, \( \mathbb{E}W(t) = \mathbb{E}W(1) \) for all \( t \geq 0 \).
which implies
\[ P \left[ \sum_{t=1}^{\infty} I\{ W(t) \in A \} = \infty \right] = 1. \]

Thus, the events \{ W(t) \in A \} occur infinitely often almost surely.

The ergodicity of the graph process can be interpreted as the property that the ensemble average coincides with the time average. In other words, when the expected graph of the network contains a directed spanning tree, that is, when \( |\lambda_2 [E(W(1))]| < 1 \), then there exists a time sequence \{ t_r : r \geq 0 \} such that collection of graphs \{ G(t_r + 1), \ldots, G(t_{r+1}) \} are infinitely often jointly strongly rooted with probability one, and therefore, asymptotic consensus is guaranteed almost surely.

Theorem 3 also states that depending on the second largest eigenvalue modulus of the expected weight matrix, weak ergodicity occurs with either probability 1 or 0. This was to be expected, as the event \( B = \{ W(0), W(1), \ldots \text{ is weakly ergodic} \} \) satisfies \( B = \varphi B \) and therefore, is invariant, i.e., \( P(B \Delta \varphi B) = 0 \). Due to ergodicity of \( \varphi \), such an event must be trivial. This is similar to the 0-1 property mentioned for independent and identically distributed matrices.

**Asymptotic Distribution of the Consensus Value**

As stated for i.i.d. graph processes in Section 2.2.1 the consensus value that the agents reach asymptotically is a random variable, whose distribution depends on the distribution of the weight matrices \( \{ W(t) \}_{t=0}^{\infty} \). Even though expressing this relationship analytically is still an open problem, one can characterize necessary and sufficient conditions under which the distribution of the consensus value is degenerate. The next proposition generalizes the statement of Proposition 4 of Section 2.2.1.

**Proposition 5.** Let \( \{ W(t) \} \) be a sequence of stochastic matrices with positive diagonals generated by an ergodic stationary process with \( |\lambda_2 [E(W(1))]| < 1 \). Also consider the
deterministic vector \(y\) satisfying \(1^T y = 1\). Then, the left product \(U(t, 0) = W(t) \ldots W(1)\) converges to \(1y^T\) almost surely, if and only if \(y\) is a left eigenvector of \(W(1)\) corresponding to the unit eigenvalue, with probability one.

**Proof:** The sufficiency proof is trivial and quite well-known [23, 110]: since \(|\lambda_2 (E W_k)| < 1\), Theorem 3 guarantees that the product in (2.3) converges to a rank one matrix with probability one, i.e., \(W(t) \ldots W(1)W(0) \to 1d^T\) almost surely, for some random vector \(d\).

In the case that almost all weight matrices share the same left eigenvector \(y\) corresponding to the unit eigenvalue,\(^{13}\) any product \(U(t, 0)\) has also the same left eigenvector, and so does its limit as \(t \to \infty\). Therefore, \(W(t) \ldots W(1)W(0) \to 1y^T\) almost surely, or in other words, \(P(d = y) = 1\).

To prove the reverse implication assume \(|\lambda_2 [E W(1)]| < 1\). Also, suppose that there exists a non-random stochastic vector \(y\) such that \(U(t, 0) = W(t) \ldots W(0) \to 1y^T\) almost surely. Since the sequence \(\{W(t) : t \geq 0\}\) is stationary, \(U(t, 1) = W(t) \ldots W(2)W(1)\) should also converge to \(1y^T\) almost surely. Combining the above, we have,

\[
U(t, 0) = U(t, 1)W(0) \to 1y^TW(0) \quad \text{almost surely},
\]

and therefore,

\[
y^T U(t, 1) \to y^TW(0) = y^T \quad \text{almost surely}.
\]

As a consequence, \(P(y^TW_1 = y^T) = 1\), which means that almost all weight matrices have the same common left eigenvector \(y\) corresponding to the unit eigenvalue.

\[\blacksquare\]

As a final remark, note that stationarity of the matrix process plays a crucial role in proving the necessity part of the above theorem. In fact, if the weight matrix process is not stationary, having a common left eigenvector corresponding to the unit eigenvalue is not necessary.\(^{13}\)

\(^{13}\)Note that since \(|\lambda_2 [E W(1)]|\) is subunit, there is only one such vector \(y\).
necessary anymore. As an example, consider the following two stochastic matrices:

\[
W = \begin{bmatrix}
\frac{3}{4} & \frac{1}{4} \\
\frac{1}{3} & \frac{2}{3}
\end{bmatrix}, \quad W' = \begin{bmatrix}
\frac{1}{3} & \frac{4}{3} \\
\frac{3}{5} & \frac{2}{5}
\end{bmatrix}.
\]

It is easy to verify that neither matrix is doubly stochastic. However, the product \(W'W\) is a doubly stochastic matrix. Therefore, if matrix sequence \(\{W(t)\}_{t=0}^{\infty}\) is such that \(W(0) = W, W(1) = W'\), and \(W(t)\) is doubly stochastic for \(t \geq 2\), then the agents reach an asymptotic average consensus, even though \(W(0)\) and \(W(1)\) are not doubly stochastic.

### 2.3 Related Literature

**Deterministic Consensus Algorithms**

Formal studies of linear iterative updates, now known as consensus algorithms, goes back to the work of DeGroot [33], who assumes a fixed network over time. The case of time-varying networks was first studied by Tsitsiklis [105] and Tsitsiklis, Bertsekas, and Athens [106], in the context of distributed and parallel computing. In the context of coordination in multi-agent networks, Jadabaie, Lin, and Morse [64] and Olfati-Saber and Murray [85] studied similar problems in discrete and continuous-time, respectively. In [92], Ren and Beard extended the results of [64, 85] and presented some improved conditions for state consensus under dynamically changing directed interaction topology. These results are further generalized to non-linear updates by Moreau [78] and Lin, Francis, and Maggiore [72], networks with communication delays by Papachristodoulou and Jadabaie [86] and Angeli [4], consensus with quantized values by Kashyap *et al.* [66] and Carli *et al.* [19], and asynchronous information consensus by Fang and Antsaklis [49]. For more details, see the survey of Ren *et al.* [93] and references therein.
Random Consensus Algorithms

One of the first results in random consensus algorithms goes back to Hatano and Mesbani [60], who assume independent undirected edges over time. A more general model is studied by Wu [109], who assumes that information links in the network are directed and not necessarily independent. However, he only proves convergence to consensus in probability, rather than the more general notion of almost sure convergence. These results are further extended by Porfiri and Stilwell [90], and Tahbaz-Salehi and Jadbabaie [102] to general i.i.d. matrix sequences. In [101], Tahbaz-Salehi and Jadbabaie investigate the continuous-time variant of the problem. The speed of convergence to consensus and some concentration results for the general i.i.d. case is studied by Fagnani and Zampieri in [48]. These results are applied by Lobel and Ozdaglar [73] in the study of distributed subgradient methods over randomly switching networks. In a more recent work, Acemoglu, Ozdaglar, and ParandehGheibi [3] use the random consensus model to study the tension between information aggregation and spread of misinformation in large social networks.

The common crucial assumption of all the above works is that the weight matrices (and hence, the underlying graphs of the network) are independent and identically distributed. The first departure from this framework is the model of Picci and Taylor [88], where undirected and unweighted information links are generated by an ergodic and stationary process. Their work is extended by Tahbaz-Salehi and Jadbabaie [103] to general ergodic stationary graph processes. Finally, Matei, Martins, and Baras [74] study a Markovian (but not necessarily ergodic) model of graph evolution.
Chapter 3

Social Learning

Network consensus algorithms, studied in Chapter 2, are amongst the most well-known distributed update rules that lead to a simple form of information aggregation over networks. These algorithms are specifically useful in applications that require an eventual agreement over the network on some parameter of interest, such as headings of robots, frequencies of inter-connected oscillators, etc. They can also be used in explaining different interesting natural and social phenomena, such as flocking in birds or spread of gossips in social networks.

Although useful and simple, the applicability of consensus algorithms can be quite limited. One of the shortcomings of consensus algorithms is that they are tailored for fusing the initial information held by agents over the network, and, at least at their basic level, are not capable of handling an ever increasing flow of information over time. In fact, it is not immediately clear how agents should incorporate the new information provided to them into their beliefs. A more important shortcoming is that consensus algorithms can only guarantee agreement with limited control over the value that the agents agree upon. However, in many applications, it is not sufficient for agents to simply reach an agreement over the network, and it is equally important for them to agree on the right value; that is,
agents need to learn the unknown parameter of interest through sharing information with one another. In such applications, one needs algorithms that can guarantee a controllable and robust convergence to both consensus and learning.

The main contribution of this chapter is to propose a simple, distributed, and non-Bayesian belief update rule that results in asymptotic learning over the network. Our proposed learning rule enables agents to correctly aggregate information, that is, to recover an unknown parameter as if they were completely informed of all signals observed across the network. Moreover, we show that under our non-Bayesian social learning rule, agents can achieve full learning, while avoiding the highly complex computations that are essential for full Bayesian learning over networks. The first section of this chapter contains a survey of related works on social learning, and clarifies the contributions of our results. The proposed model of non-Bayesian social learning is presented in Section 3.2. This is followed by Section 3.3, in which we present and prove conditions that guarantee asymptotic consensus and learning.

### 3.1 Summary of Results and Related Literature

Individuals, for the most part, form and update their opinions based on their observations and signals. For example, a physician’s personal experience with a new medical treatment might serve as a basis for her beliefs regarding its effectiveness. Such private signals, however, are not the only factors that determine the opinions, and hence, the behavior of an individual. In fact, a variety of evidence suggests that in many cases, individuals’ beliefs on various economic, political, and social variables are based on information they receive from other agents in their social clique; people such as friends, colleagues and family members. For instance, Hagerstrand [57] and Rogers [94] document such a phenomenon in the choice of new agricultural techniques by various farmers, while Kotler [69] shows the
importance of learning from others in the purchase of consumer products. Similarly, Ioannides and Loury [63] document how social networks and “neighborhood effects” influence employment outcomes.

Even though individuals learn from each other through social interactions, the complexity of social networks makes it difficult to incorporate information provided by others in a fully Bayesian manner. One reason is that agents often do not have complete information about the structure of the network. In order to form opinions rationally, individuals need to hold beliefs not only about the states of the world, but also about the topology of the social network. This significantly complicates the required calculations of individuals, well beyond their computational capabilities, even in networks of moderate size. However, the problem with Bayesian learning persists even when individuals have complete information about the network structure, as they still need to perform rational deductions about the beliefs of every other individual in the network while only observing the evolution of the beliefs of their neighbors. Such fully Bayesian agents need to form beliefs about who the source of each piece of information is, how information is spread around the network, and how every agent’s beliefs affect everyone else’s. The computational burden of these calculations are prohibitive for adopting rational learning in complex networks. Therefore, in large and complex networks, agents need to apply coarser learning rules in order to incorporate information provided to them by their neighbors.

3.1.1 Summary of Results

In this chapter, our focus is on understanding the evolution of beliefs in a society where agents use simple update rules to incorporate the beliefs of other agents in their social clique. We base our study on a model according to which, at every time period, each agent receives a private noisy signal about the true state of world and updates her beliefs as a convex combination of the Bayesian posterior belief generated by her private signals
This second term captures a notion of aggregate belief held in the local neighborhood of the agent. As stated earlier, agents in this world are boundedly rational and fail to incorporate the information provided to them by their neighbors in a Bayesian manner. Instead, their opinion is simply swayed in the direction of the average belief in their neighborhood.

First, we show that all agents will eventually hold correct forecasts about their signals, provided that the social network is strongly connected; that is, there exists either a direct or an indirect information path between any two agents. By the means of an example we show that the assumption of strong connectivity cannot be disposed of. We further show that in strongly connected networks, these social interactions, not only enable agents to accurately predict their future observations, but also lead to information aggregation about the true underlying state of the world. In other words, even if the combined signals of each agent are not informative enough to completely reveal the state of world, the non-Bayesian learning rule successfully provides them with information available to others. In particular, we show that, in a strongly connected society, all agents are also capable of correctly forecasting the signals observed by others. An important consequence of this phenomenon is that if the signals observed privately by agents are conditionally independent, they asymptotically learn the true underlying state of the world as if they were completely informed of all signals and updated their beliefs according to Bayes’ rule. Hence, agents can successfully aggregate information, while avoiding the computational cost and complexity involved in Bayesian update. On the other hand, if the cross-section of signals are correlated, they are only guaranteed to eventually learn the marginal distributions of the observations and might fail in learning the true joint distribution. This can be regarded as a cost that agents

\footnote{We abstract from any strategic interaction among the agents. In the model, agents are assumed to report their beliefs truthfully. This is a reasonable assumption in many settings, such as sharing political opinions among friends or family members. Nevertheless, there are also many interactions that create incentives for the agents to strategically misreport their beliefs.}
have to pay for not being Bayesian.

The important feature of our main result is that it highlights the role of social networks in the process of information aggregation. In essence, an agent can learn from individuals whom she is neither in direct contact with nor has any knowledge about their signal structure. In fact, it might even be the case that she is not aware of their existence altogether. Nevertheless, the indirect communication path in the social network guarantees that she will eventually incorporate their information correctly into her beliefs. Moreover, this successful aggregation of information occurs despite the fact that truth is not immediately recognizable to any of the agents, and neither one of them could have learned it by herself in isolation. On the other hand, while the role played by the network is crucial, its topology only affects the speed of learning. In other words, as long as strong connectivity holds, the structure of the network and the level of influence of different individuals does not prevent them from correctly aggregating information. For example, it might be the case that the most influential agents are the ones with the least informative signals. Nonetheless, as long as informed agents also have some (even arbitrarily small) influence on others, the learning process is successful.

Another distinctive feature of our results is the absence of absolute continuity of the true measure with respect to all prior beliefs, as a requirement for social learning. In standard Bayesian learning literature, an agent will have accurate predictions, as long as she assigns a positive prior belief on the true parameter. On the other hand, in the absence of absolute continuity, no information can persuade a Bayesian agent to update her belief on a parameter from zero to some positive number, and hence, such an agent fails to predict future correctly.\footnote{For example, see [77] and [95].} In a network of Bayesian agents, the existence of an agent with zero prior belief on the true parameter can also affect the learning of others. For instance, such an agent located at the network’s bottleneck, functioning as the only link connecting
two components of the network, might prevent agents on each component from obtaining valuable information from the other side. In contrast, in our model, we show that it is sufficient to have one agent with a positive initial belief assigned to the true parameter: as long as the social network is strongly connected and the true measure is absolutely continuous with respect to prior belief of some agent, complete learning is achieved.

3.1.2 Related Literature

Bayesian Learning

There exists a vast literature on learning over social networks, both boundedly and fully rational. The Bayesian learning literature mainly focuses on formulating the problem as a dynamic game over the network and characterizing its equilibria. However, since characterizing such equilibria in complex networks is generally an intractable problem, the existing literature focuses on relatively simple and stylized environments [2]. For example, Bikchandani, Hirshleifer, and Welch [12] and Banerjee [8] consider models where each individual takes a single action and observes all past actions. Similarly, Banerjee and Fudenberg [9] and Simith and Sørensen [100] focus on models where agents make decisions sequentially, but instead, only observe a representative sample of the past actions. In a more recent paper, Acemoglu et al. [1] generalize these results to an arbitrary network structure. Nevertheless, the assumption of a single decision for each agent remains in place. Another example of Bayesian learning over networks is Gale and Kariv [51], who show how the complexity of a rational agent’s decision-problem increases over time, as she has to hold beliefs about her neighbors’ knowledge of their neighbors’ actions and the private information they reveal. Gale and Kariv also show the intuitive result that all individuals in a connected component converge to having the same beliefs about an optimal decision. They

3Notice that Bayesian literature is silent on how agents form their prior beliefs. Moreover, it is not clear how a Bayesian agent must interpret the prior beliefs of her neighbors.
argue that if that is not the case, then the agents can simply “imitate” their neighbors and increase their payoffs.

**Non-Bayesian Learning**

The model presented in this chapter is closely related to another collection of works in the social learning literature that focus on non-Bayesian learning models; models such as Ellison and Fudenberg [44, 45], and Bala and Goyal [6, 7] in which the agents are not fully rational, and use simple rule-of-thumb methods to incorporate the information of their neighbors. In the same spirit are DeMarzo, Vayanos, and Zwiebel [34], Golub and Jackson [56] and Acemoglu, Ozdaglar, and ParandehGheibi [3], which are all based on the opinion formation model of DeGroot [33], studied in Chapter 2. These papers, unlike most of Bayesian learning literature, consider repeated communication and interactions among agents. More specifically, in DeGroot-style models, which in essence are consensus algorithms over the network, each individual initially receives one signal about the state of the world, and at the consequent time steps, updates her beliefs as a weighted average of the beliefs of her neighbors. Their main focus is on conditions under which individuals in the connected components of the social network converge to similar beliefs.\(^4\) Golub and Jackson further show that if the size of the network grows unboundedly, this asymptotic consensus belief converges to the true state of the world, provided that there are not overly influential agents in the society.

One main feature that distinguishes our results from the works that are based on DeGroot’s consensus model, such as Golub and Jackson [56], is the existence of time dynamics. Whereas in DeGroot’s model each agent has only a single observation, the individuals in our model receive information in small bits over time, and therefore, need to incorporate new privately observed signals into their beliefs. This continuous flow of information, as

\(^4\)For more on DeGroot’s model and consensus algorithms see Chapter 2.
well as the repeated social interaction and information exchange are also aspects that make our model different from the Bayesian learning literature. Another crucial difference between Golub and Jackson [56] and our model is the role played by the social network in successful information aggregation. They show that existence of “influential” individuals, in the sense of being connected to a large number of people, makes learning impossible. In contrast, in our environment, as discussed earlier, strong connectivity is the only requirement on the network for successful learning, and neither the network topology nor the influence level of different agents can prevent learning.

Another closely related paper is Epstein, Noor, and Sandroni [46], who consider a single non-Bayesian agent with a bias towards her own prior beliefs, a model similar to the dynamics of belief update presented in this chapter. Nevertheless, the focus of our analysis is not only on the dynamics of the beliefs, but also on the process of information aggregation over a network consisting of many agents.

### 3.2 A Model of Non-Bayesian Social Learning

This section contains the formal description of the proposed non-Bayesian social learning model.

#### 3.2.1 Agents and Observations

Let \( \Theta \) denote a finite set of possible states of the world and let \( \theta^* \in \Theta \) denote the true underlying state of the world. We consider a set \( \mathcal{N} = \{1, 2, \ldots, n\} \) of agents interacting over a social network who do not know the true state of the world. Each agent \( i \) starts with a prior belief about the true state, represented by \( \mu_{i,0} \in \Delta \Theta \) which is a probability distribution over the set \( \Theta \). \( \mu_{i,0}(\theta) \) denotes the initial probability that agent \( i \) assigns to the state \( \theta \in \Theta \) to be the true underlying state of the world.
Conditional on the state of the world $\theta$, at each time period $t \geq 1$, an observation profile $s_t = (s^1_t, \ldots, s^n_t) \in S_1 \times \cdots \times S_n \equiv S$ is generated according to the likelihood function $\ell(s_t|\theta)$. $s^i_t \in S_i$ denotes a signal that is privately observed by agent $i$. These privately observed signals are independent over time, but might be correlated among agents at the same time period. Without much loss of generality, we assume that $\ell(s|\theta) > 0$ for all $(s, \theta) \in S \times \Theta$. The signal space $S_i$ is assumed to be finite for all $i$. We use $\ell_i(\cdot|\theta)$ to denote the $i$th marginal of $\ell(\cdot|\theta)$. We further assume that every agent $i$ knows the conditional likelihood function $\ell_i(\cdot|\theta)$, known as her signal structure.

For a fixed $\theta \in \Theta$, we define a probability triple $(\Omega, \mathcal{F}, \mathbb{P}_\theta)$, where $\Omega$ is the space containing sequences of realizations of the signals $s_t \in S$ over time, and $\mathbb{P}_\theta$ is the probability measure induced over sample paths in $\Omega$ by the state $\theta$. In other words, $\mathbb{P}_\theta = \otimes_{t=1}^\infty \ell(\cdot|\theta)$. We use $\mathbb{E}_\theta[\cdot]$ to denote the expectations with respect to the measure $\mathbb{P}_\theta$. Define $\mathcal{F}_{i,t}$ as the $\sigma$-field generated by the past history of agent $i$’s observations up to time period $t$, and let $\mathcal{F}_t$ be the smallest $\sigma$-field containing all $\mathcal{F}_{i,t}$ for $1 \leq i \leq n$.

### 3.2.2 Social Structure

When updating their beliefs about the true state of the world, agents have access to the beliefs currently held by other individuals in their social clique. We capture the social interaction structure between agents by a directed graph $G = (V, E)$, where each vertex in $V$ corresponds to an agent, and an edge connecting vertex $i$ to vertex $j$, denoted by the ordered pair $(i, j) \in E$, captures the fact that agent $j$ has access to the belief held by agent $i$. Note that because of the way we have defined the social network, the beliefs of agent $i$ might be accessible to agent $j$, but not the other way around.

For each agent $i$, define $\mathcal{N}_i = \{j \in V : (j, i) \in E\}$, called the set of neighbors of agent $i$. The elements of this set are agents whose beliefs are available to agent $i$ at each time period. The important underlying assumption here is that agent $i$ can observe the exact
beliefs of her neighbors. In other words, we are assuming that either agent \( i \) can directly observe her neighbors’ beliefs, or that there is no strategic interaction between the agents, and therefore, every agent reports her beliefs truthfully.

A directed path in \( G = (V, E) \) from vertex \( i \) to vertex \( j \), is a sequence of vertices starting with \( i \) and ending with \( j \) such that each vertex is a neighbor of the next vertex in the sequence. The social network is strongly connected, if there exists a directed path from each vertex to any other vertex. We say the social network is connected, whenever such a path exists ignoring the direction of the edges. We refer to any network that is connected but not strongly connected as a weakly connected network.\(^5\)

### 3.2.3 Belief Updates

Given the social structure described above, the agents update their beliefs based on the signal that they privately observe and the beliefs held by their neighbors. In particular, each agent’s belief over \( \Theta \) is a linear combination of her Bayesian posterior belief and her neighbors’ priors. That is, if we let \( \mu_{i,t}(\theta) \) denote the belief that agent \( i \) assigns to parameter \( \theta \in \Theta \) at time period \( t \) after observing \( (s_{i,1}, s_{i,2}, \ldots, s_{i,t}) \), then

\[
\mu_{i,t+1} = a_{ii} BU(\mu_{i,t}; s_{i,t+1}) + \sum_{j \in N_i} a_{ij} \mu_{j,t},
\]

where \( a_{ij} \in \mathbb{R}_+ \) captures the weight that agent \( i \) assigns to the belief held by agent \( j \) in her neighborhood, \( BU(\mu_{i,t}; s_{i,t+1})(\cdot) \) is the Bayesian update of \( \mu_{i,t} \) when signal \( s_{i,t+1} \) is observed, and \( a_{ii} \) is the weight assigned to the Bayesian update by agent \( i \), which we call her self-confidence.\(^6\) Note that the weights \( a_{ij} \) must satisfy \( \sum_{j \in N_i} a_{ij} = 1 \), in order for the

\(^5\)For more on the strong connectivity and related concepts see the Appendix.

\(^6\)One can modify this belief update model and assume that agent \( i \)'s belief update also depends on his own beliefs at the previous time step, \( \mu_{i,t} \). Such an assumption is equivalent to adding a prior-bias to the model, as stated in Epstein, Noor, and Sandroni [46]. Since this generality does not change the results or the economic intuitions, we assume that the agents have no prior bias.
period $t + 1$ beliefs to form a well-defined probability distribution. Therefore, at each time period, the posterior belief of agent $i$ is a convex combination of her Bayesian update and the priors of her neighbors.

Given agent $i$’s beliefs at time period $t$, her time $t + 1$ forecast of the next observation is given by

$$m_{i,t}(s_{t+1}^i) = \int_{\Theta} \ell_i(s_{t+1}^i | \theta) d\mu_{i,t}(\theta), \quad (3.2)$$

where the forecasts $m_{i,t}(\cdot)$ form a probability measure on $S_i$. Similarly, if agent $i$ knows the signal structure of agent $j \neq i$, she would also be able to form forecasts for $j$’s observations. Such forecasts are given by

$$m_{j,i,t}(s_{t+1}^j) = \int_{\Theta} \ell_j(s_{t+1}^j | \theta) d\mu_{i,t}(\theta).$$

Given the above definitions, the law of motion for the beliefs about the parameters can be written as

$$\mu_{i,t+1}(\theta) = a_{ii}\mu_{i,t}(\theta) \frac{\ell_i(s_{t+1}^i | \theta)}{m_{i,t}(s_{t+1}^i)} + \sum_{j \in N_i} a_{ij}\mu_{j,t}(\theta), \quad (3.3)$$

for all $\theta \in \Theta$. Note that the dynamics of belief update in our model is local, in the sense that each individual only uses the beliefs of her immediate neighbors to form her opinions, ignores the structure of the network, and does not make any inferences about the beliefs of other individuals in the society. This boundedly rational dynamics for opinion formation greatly reduces the computational burden of the individuals. Moreover, equation (3.3) suggests that an individual does not need to keep track of the identities of her neighbors and the exact information provided by them. In fact, she only needs to know the “average belief” held in her neighborhood, given by the term $\sum_{j \in N_i} a_{ij}\mu_{j,t}(\cdot)$. In the special case that the signals observed by an agent are non-informative, equation (3.3) reduces to the belief update model of DeGroot [33], studied in Chapter 2 and used by Golub and Jackson [56].
When analyzing the asymptotic behavior of the beliefs, sometimes it is more convenient to use a matrix notation. Define $A$ to be a real $n \times n$ matrix which captures the social interaction of the agents as well as the weight that each agent assigns to her neighbors. More specifically, we let the $ij$ element of the matrix $A$ be $a_{ij}$ when agent $j$ is a neighbor of agent $i$, and zero otherwise. Thus, equation (3.3) can be rewritten as

$$
\mu_{t+1}(\theta) = A\mu_t(\theta) + \text{diag}
\left[
\frac{\ell_1(s_{t+1}^1|\theta)}{m_1, t(s_{t+1}^1)} - 1, \ldots, \frac{\ell_n(s_{t+1}^n|\theta)}{m_n, t(s_{t+1}^n)} - 1
\right] \mu_t(\theta)
$$

where $\mu_t(\cdot) = [\mu_{1,t}, \ldots, \mu_{n,t}]^T(\cdot)$, and diag of a vector is a diagonal matrix which has the entries of the vector as its diagonal. Note that $A$ is an irreducible matrix if and only if graph $G$ is strongly connected.\(^7\) Moreover, since at each time period the beliefs of all agents are convex combinations of their Bayesian posteriors and the priors of their neighbors, $A$ is a stochastic matrix.\(^8\) In the special case that $A$ is the identity matrix, our model reduces to the benchmark Bayesian case, in which the society consists of $n$ Bayesian agents who do not have access to the beliefs of other members of the society, and only observe their own private signals.

### 3.3 Social Learning

Given the model described above, we are interested in the question of what is learned in the long run. Learning may either signify learning the true parameter or learning to forecast future outcomes. These two notions of learning are distinct and might not occur simultaneously. We start this section by specifying what we exactly mean by either type.

Suppose that $\theta^* \in \Theta$ is the true state of the world and thus, the measure $\mathbb{P}^* = \otimes_{t=1}^\infty \ell(\cdot|\theta^*)$.

---

\(^7\)An $n \times n$ matrix $A$ is said to be reducible, if for some permutation matrix $P$, the matrix $P'AP$ is block upper triangular. If a square matrix is not reducible, it is said to be irreducible. For more on this, see e.g., [10].

\(^8\)Recall that a matrix is said to be stochastic if it is entry-wise non-negative and all its row sums are equal to one. A stochastic matrix is called doubly stochastic if all its column sums are equal to one, as well.
is the probability law describing the process \((s_t)\).

**Definition 11.** The forecasts of agent \(i\) are *eventually correct* on a path \(\{s_t\}_{t=1}^{\infty}\) if, along that path,

\[
m_{i,t}(\cdot) \rightarrow \ell_i(\cdot|\theta^*) \quad \text{as} \quad t \rightarrow \infty.
\]

Similarly, provided that agent \(i\) has access to the signal structure of agent \(j\neq i\), her forecasts for the observations of agent \(j\) are eventually correct, if along that path, \(m_{i,t}^j(\cdot) \rightarrow \ell_j(\cdot|\theta^*)\) as \(t \rightarrow \infty\).

This notion of learning, which Kalai and Lehrer [65] call *weak merging* of opinions, captures the ability of agents to correctly forecast events in near future. It is well-known, that repeated applications of the Bayes’ rule leads to eventually correct forecasts with probability 1 under the truth, given suitable conditions, the key condition being absolute continuity of the true measure with respect to initial beliefs.\(^9\) The implication is that the mere repetition of Bayes’ rule eventually transforms the historical record into a near perfect guide for the future. However, predicting future observations accurately is not the same as learning the underlying state of the world. In fact, depending on the signal structure of each agent, there might be an “identification problem” which can potentially prevent the agent from learning the true (payoff-relevant) parameter \(\theta^*\). The other type of learning that we are concerned with, precisely captures this notion:

**Definition 12.** Agent \(i \in \mathcal{N}\) *asymptotically learns* the true parameter \(\theta^*\) on a path \(\{s_t\}_{t=1}^{\infty}\) if, along that path,

\[
\mu_{i,t}(\theta^*) \rightarrow 1 \quad \text{as} \quad t \rightarrow \infty.
\]

Asymptotic learning occurs when the agent assigns probability one to the true parameter. As mentioned earlier, having eventually correct forecasts does not guarantee asymptotic

\(^9\)Lehrer and Smorodinsky [70] show that an assumption weaker than absolute continuity, known as accommodation, is sufficient for weak merging of the opinion.
learning of the true parameter. In general, the converse is not true either. However, it is straightforward to show that in the absence of time correlations, as in our model, asymptotically learning $\theta^*$ implies eventually correct forecasts.\(^{10}\)

### 3.3.1 Correct Forecasts in Strongly Connected Societies

We now turn to the main question of this chapter: under what circumstances does learning occur over the social network, when agents update their beliefs according to (3.3)? Note that without loss of generality, we can limit our focus to two cases of strongly connected and weakly connected networks. This is due to the fact that in the case that the network is disconnected, our results can be applied to each connected component separately.

Our first result shows that under very mild assumptions, in spite of local interactions and the non-Bayesian belief update, agents will eventually forecast their private signals correctly.

**Proposition 6.** Suppose that the social network is strongly connected, and all agents have strictly positive self-confidence. Then, all agents eventually forecast their private observations accurately with $\mathbb{P}^*$-probability one, provided that there exists an agent in the social network with strictly positive prior belief on the true parameter $\theta^*$.

The proposition relies on three main assumptions: strictly positive weights on new observations by all agents, a strongly connected social network, and finally, a strictly positive prior belief on the true parameter by at least one agent in the network. If either of these assumptions are dropped, the forecasts might either diverge or converge to wrong values. Before presenting the proof of the above proposition, we briefly discuss each assumption.

The first assumption on strictly positive self-confidences is quite intuitive: it prohibits agents from completely discarding information provided to them through their observa-

\(^{10}\)See Lehrer and Smorodinsky [70], for an example of the case that learning the true parameter does not guarantee merging.
tions. Clearly, if all agents discard their private signals, no new information is incorporated to their beliefs, and (3.3) simply turns into a diffusion of prior beliefs. An interesting case is when only some of them discard their private signals. In such a case, only agents with no self-confidence fail to predict their signals accurately. This failure, however, does not prevent the ones with positive self-confidence from forecasting their own observations correctly.

The second requirement for accurate predictions is strong connectivity of the social network. To understand why strong connectivity is the key, we present a simple example of a weakly connected social network, in which an agent fails in forecasting future correctly.

**Example 1.** Consider a society consisting of two agents, \( \mathcal{N} = \{1, 2\} \), and assume that \( \Theta = \{\theta_1, \theta_2\} \) with the true state being \( \theta^* = \theta_1 \). Both agents have non-degenerate prior beliefs over \( \Theta \). The signals observed by agents are independent conditional on the state of the world, and belong to the set \( S_1 = S_2 = \{H, T\} \). We further assume that the signals observed by agent 2 are non-informative, while agent 1’s observations are perfectly informative about the state; that is, \( \ell_1(H|\theta_1) = \ell_1(T|\theta_2) = 1 \), and \( \ell_2(s|\theta_1) = \ell_2(s|\theta_2) \) for \( s \in \{H, T\} \). As for the social structure, we assume that agent 1 has access to the beliefs of agent 2, while agent 2 cannot (or chooses not to) use the beliefs held by agent 1. Therefore, the social interaction matrix is given by

\[
A = \begin{bmatrix}
1 - \alpha & \alpha \\
0 & 1
\end{bmatrix},
\]

where \( \alpha \in (0, 1) \) is the weight that agent 1 assigns to the beliefs of agent 2, when updating her beliefs using equation (3.3). Since the private signals observed by the latter are non-informative, her beliefs, at all times, remain equal to her prior. Clearly, she has correct forecasts at all times. Agent 1, on the other hand, will not have eventually correct forecasts.
To see this, notice that agent 1 has eventually correct forecasts, only if she eventually assigns probability 1 to the true state, $\theta_1$. This is due to the fact that her observations are perfectly informative. However, the belief she assigns to $\theta_2$ follows the law of motion

$$
\mu_{1,t+1}(\theta_2) = (1 - \alpha)\mu_{1,t}(\theta_2)\frac{\ell_1(s_{t+1}^1|\theta_2)}{m_1(s_{t+1}^1)} + \alpha\mu_{2,t}(\theta_2)
$$

which cannot converge to zero, due to the fact that $\mu_{2,t}(\theta_2) = \mu_{2,0}(\theta_2)$ is strictly positive.

The intuition for the failure of learning in the above example is simple. First of all, notice that the two agents have different signal structures, which means that they interpret the states differently. Moreover, agent 1, in essence, is following the beliefs of agent 2 without considering the fact that agent 2 is less informed than herself, while at the same time she does not influence her back. This one-way influence and non-identical interpretations of signals (due to non-identical signal structures) result in confusion on the part of agent 1, and hence incorrect forecasts. Clearly, if agent 1 were Bayesian and capable of incorporating the information provided to her by agent 2 rationally, she would have learned the true parameter. As a final remark on the second assumption, note that in the special case that all social interactions are bidirectional, the social network is trivially strongly connected.

Finally, in order to have accurate predictions, Proposition 6 requires the existence of at least one agent with a positive prior belief on the true state $\theta^*$. Note that correct forecasts are achieved, even if the true measure, $\mathbb{P}^*$, is not absolutely continuous with respect to the prior beliefs of many agents in the network. This is in contrast to the standard Bayesian learning literature, which requires absolute continuity in order to guarantee accurate predictions. This feature of our model is significant: as long as some agent assigns a positive prior belief to the true state, all agents in the network will be able to correctly forecast their observations, even if she is located at the fringe of the society and has very small influence on her neighbors. Clearly, if $\mu_{i,0}(\theta^*) = 0$ for all $i \in \mathcal{N}$, then the belief assigned to the true
parameter by all agents will remain equal to zero over time, and no learning can happen.

**Proof of Proposition 6**

We now turn to proving Proposition 6. The following two lemmas, both consequences of the martingale convergence theorem, are used in the proof.

**Lemma 3.** If $A$ is stochastic, then the sequence $\sum_{i=1}^{n} v_i \mu_{i,t}(\theta^*)$ converges $\mathbb{P}^*$-almost surely as $t \to \infty$, where $v$ is any non-negative left eigenvector of $A$ corresponding to its unit eigenvalue.

**Proof:** First, note that since $A$ is stochastic, it always has at least one eigenvalue equal to 1. Moreover, there exists a non-negative left eigenvector corresponding to this eigenvalue. We denote such a vector by $v$.

We evaluate equation (3.4) at the true parameter $\theta^*$ and pre-multiply both sides by $v^T$:

$$v^T \mu_{t+1}(\theta^*) = v^T A \mu_t(\theta^*) + \sum_{i=1}^{n} v_i \mu_{i,t}(\theta^*) a_{ii} \left[ \frac{\ell_i(s_{t+1}^i|\theta^*)}{m_i(s_{t+1}^i)} - 1 \right].$$

Thus,

$$\mathbb{E}^* \left[ \sum_{i=1}^{n} v_i \mu_{i,t+1}(\theta^*) | \mathcal{F}_t \right] = \sum_{i=1}^{n} v_i \mu_{i,t}(\theta^*) + \sum_{i=1}^{n} v_i a_{ii} \mu_{i,t}(\theta^*) \mathbb{E}^* \left[ \frac{\ell_i(s_{t+1}^i|\theta^*)}{m_i(s_{t+1}^i)} - 1 | \mathcal{F}_t \right],$$

(3.5)

where $\mathbb{E}^*$ denotes expectation with respect to the measure $\mathbb{P}^*$. Since the function $f(x) = 1/x$ is convex, Jensen’s inequality implies $\mathbb{E}^* \left[ \frac{\ell_i(s_{t+1}^i|\theta^*)}{m_i(s_{t+1}^i)} | \mathcal{F}_t \right] \geq \left( \mathbb{E}^* \left[ \frac{m_i(s_{t+1}^i)}{\ell_i(s_{t+1}^i|\theta^*)} | \mathcal{F}_t \right] \right)^{-1} = 1$, and therefore,

$$\mathbb{E}^* \left[ \sum_{i=1}^{n} v_i \mu_{i,t+1}(\theta^*) | \mathcal{F}_t \right] \geq \sum_{i=1}^{n} v_i \mu_{i,t}(\theta^*).$$

The last inequality is due to the fact that $v$ is element-wise non-negative. As a result,

---

11This is a consequence of the Perron-Frobenius theorem. For more on the properties of non-negative and stochastic matrices, see [10].
\( \sum_{i=1}^{n} v_i \mu_{i,t}(\theta^*) \) is a submartingale with respect to the filtration \( \mathcal{F}_t \), which is also bounded above by \( \|v\|_1 \). Hence, it converges \( \mathbb{P}^* \)-almost surely.

**Lemma 4.** Suppose that there exists an agent \( i \) such that \( \mu_{i,0}(\theta^*) > 0 \). Then, whenever \( A \) is stochastic and irreducible, the sequence \( \sum_{i=1}^{n} v_i \log \mu_{i,t}(\theta^*) \) converges \( \mathbb{P}^* \)-almost surely as \( t \to \infty \), where \( v \) is the positive left eigenvector of \( A \) corresponding to its unit eigenvalue.

**Proof:** Similar to the proof of the previous lemma, we show that \( \sum_{i=1}^{n} v_i \log \mu_{i,t}(\theta^*) \) is a bounded submartingale and invoke the martingale convergence theorem to obtain almost sure convergence.

First, note that since \( A \) is a stochastic matrix, the right hand side of equation (3.3) is a convex combination for all \( i \). Therefore, by evaluating the law of motion at \( \theta^* \) and taking \( \log \) from both sides, we obtain

\[
\log \mu_{i,t+1}(\theta^*) \geq a_{ii} \log \mu_{i,t}(\theta^*) + a_{ii} \log \left( \frac{\ell_i(s_{i+1}^t)}{m_{i,t}(s_{i+1}^t)} \right) + \sum_{j \in N_i} a_{ij} \log \mu_{j,t}(\theta^*),
\]

where we have used the concavity of the logarithm function. Note that since \( A \) is irreducible, the social network is strongly connected. Thus, the existence of one agent with a positive prior on \( \theta^* \) guarantees that after at most \( n \) time periods all agents assign a strictly positive probability to the true parameter, which means that \( \log \mu_{i,t}(\theta^*) \) is well-defined for large enough \( t \) for all \( i \).

Our next step is to show that \( \mathbb{E}^* \left[ \log \frac{\ell_i(s_{i+1}^t)}{m_{i,t}(s_{i+1}^t)} | \mathcal{F}_t \right] \geq 0 \). To obtain this,

\[
\mathbb{E}^* \left[ \log \frac{\ell_i(s_{i+1}^t)}{m_{i,t}(s_{i+1}^t)} | \mathcal{F}_t \right] = -\mathbb{E}^* \left[ \log \frac{m_{i,t}(s_{i+1}^t)}{\ell_i(s_{i+1}^t)} | \mathcal{F}_t \right] \\
\geq - \log \left( \mathbb{E}^* \left[ \frac{m_{i,t}(s_{i+1}^t)}{\ell_i(s_{i+1}^t)} | \mathcal{F}_t \right] \right) = 0.
\]
Thus,
\[ \mathbb{E}^* \left[ \log \mu_{i,t+1}(\theta^*)|\mathcal{F}_t \right] \geq a_{ii} \log \mu_{i,t}(\theta^*) + \sum_{j \in N_i} a_{ij} \log \mu_{j,t}(\theta^*), \]
which can be rewritten in matrix form as \[ \mathbb{E}^* \left[ \log \mu_{t+1}(\theta^*)|\mathcal{F}_t \right] \geq A \log \mu_{t}(\theta^*), \] where by the logarithm of a vector, we mean its entry-wise logarithm. Multiplying both sides by the \( A \)'s non-negative left eigenvector \( v^T \) leads to
\[ \mathbb{E}^* \left[ \sum_{i=1}^{n} v_i \log \mu_{i,t+1}(\theta^*)|\mathcal{F}_t \right] \geq \sum_{i=1}^{n} v_i \log \mu_{i,t}(\theta^*). \]
Thus, the non-negative term \( \sum_{i=1}^{n} v_i \log \mu_{i,t}(\theta^*) \) is a submartingale with respect to the filtration \( \mathcal{F}_t \), and therefore, it converges with \( \mathbb{P}^* \)-probability one.

With these lemmas in hand, we can prove Proposition 6.

**Proof of Proposition 6:** According to Lemma 3, \( \sum_{i=1}^{n} v_i \mu_{i,t}(\theta^*) \) converges with \( \mathbb{P}^* \)-probability one, where \( v \) is the non-negative left eigenvector of \( A \) corresponding to its unit eigenvalue. Therefore, equation (3.5) implies that
\[ \sum_{i=1}^{n} v_i a_{ii} \mu_{i,t}(\theta^*) \left( \mathbb{E}^* \left[ \frac{\ell_i(s_{i,t+1}^i|\theta^*)}{m_{i,t}(s_{i,t+1}^i)}|\mathcal{F}_t \right] - 1 \right) \longrightarrow 0 \quad \mathbb{P}^* - \text{a.s.} \]
Since the term \( v_i a_{ii} \mu_{i,t}(\theta^*) \mathbb{E}^* \left[ \frac{\ell_i(s_{i,t+1}^i|\theta^*)}{m_{i,t}(s_{i,t+1}^i)} - 1|\mathcal{F}_t \right] \) is non-negative for all \( i \), each such term converges to zero with \( \mathbb{P}^* \)-probability one. Moreover, the assumptions that all the diagonal entries of \( A \) are strictly positive and that of its irreducibility (which means that \( v \) is entry-wise positive) lead to
\[ \mu_{i,t}(\theta^*) \left( \mathbb{E}^* \left[ \frac{\ell_i(s_{i,t+1}^i|\theta^*)}{m_{i,t}(s_{i,t+1}^i)}|\mathcal{F}_t \right] - 1 \right) \longrightarrow 0 \quad \text{for all } i \quad \mathbb{P}^* - \text{a.s.} \quad (3.6) \]
Furthermore, Lemma 4 guarantees that \( \sum_{i=1}^{n} v_i \log \mu_{i,t}(\theta^*) \) converges almost surely, meaning that \( \mu_{i,t}(\theta^*) \) is uniformly bounded away from zero for all \( i \) with probability one. Note
that, again we are using the fact that $v$ is a positive vector. Hence, $E^* \left[ \frac{\ell_i(s_{t+1}^i|\theta^*)}{m_i(t,s_{t+1}^i)} | \mathcal{F}_t \right] \to 1$ almost surely. Thus,

$$
E^* \left[ \frac{\ell_i(s_{t+1}^i|\theta^*)}{m_i(t,s_{t+1}^i)} | \mathcal{F}_t \right] - 1 = \sum_{s \in S_i} \ell_i(s|\theta^*) \left( \frac{\ell_i(s|\theta^*)}{m_i(t)(s)} - 1 \right) \\
= \sum_{s \in S_i} \left( \ell_i(s|\theta^*) \frac{\ell_i(s|\theta^*) - m_i(t)(s)}{m_i(t)(s)} + m_i(t)(s) - \ell_i(s|\theta^*) \right) \\
= \sum_{s \in S_i} \frac{[\ell_i(s|\theta^*) - m_i(t)(s)]^2}{m_i(t)(s)} \to 0 \quad \mathbb{P}^* \text{ -- a.s.,}
$$

where the second equality is due to the fact that both $\ell_i(\cdot|\theta^*)$ and $m_i(\cdot)$ are measures on $S_i$, and therefore, $\sum_{s \in S_i} \ell_i(s|\theta^*) = \sum_{s \in S_i} m_i(t)(s) = 1$.

In the last expression, the term in the braces and the denominator are always non-negative and therefore,

$$
m_i(t)(s) \to \ell_i(s|\theta^*) \quad \mathbb{P}^* \text{ -- a.s.}
$$

for all $s \in S_i$ and all $i \in \mathcal{N}$.

### 3.3.2 Asymptotic Beliefs

The key implication of Proposition 6 is that as long as the social network is strongly connected, all agents will eventually forecast their future observations correctly. In this section, we characterize the asymptotic beliefs held by the agents. Such characterization is required, in order to be able to predict what each agent asymptotically learns about the true state from her neighbors.

Our next result shows that all agents will have asymptotically equal beliefs.

**Proposition 7.** Under the assumptions of Proposition 6, all agents have asymptotically equal beliefs with $\mathbb{P}^*$-probability 1. That is, $\mu_{i,\infty}(\theta) = \lim_{t \to \infty} \mu_{i,t}(\theta)$ does not depend on $i$ for all $\theta \in \Theta$. 

58
Proof: Notice that \( \mu_{i,t}(\theta) \frac{\ell_i(s_{i+1}^t | \theta)}{m_{i,t}(s_{i+1}^t)} \) in the belief update (3.3) converges to some limit for all agents \( i \). This can happen only if

\[
\mu_{i,t}(\theta) \left[ \frac{\ell_i(s_{i+1}^t | \theta)}{m_{i,t}(s_{i+1}^t)} - 1 \right] \to 0 \quad \mathbb{P}^* - \text{a.s.}
\]

for all \( \theta \in \Theta \) and all \( i \in \mathcal{N} \). As a consequence, equation (3.4) implies that with \( \mathbb{P}^* \)-probability one, and for all parameters \( \theta \), \( \mu_{t+1}(\theta) - A\mu_t(\theta) \to 0 \). That is, on almost all sample paths, for any \( \epsilon > 0 \) there exists a large enough time \( T \) such that for all \( t \geq T \),

\[
|\mu_{i,t+1}(\theta) - \sum_{k=1}^n a_{ik}\mu_{k,t}(\theta)| < \frac{\epsilon}{2} \quad \forall i \in \mathcal{N}
\]

Therefore, given any two agents \( i \) and \( j \),

\[
|(\mu_{i,t+1}(\theta) - \mu_{j,t+1}(\theta)) - \sum_{k=1}^n \mu_{k,t}(\theta)(a_{ik} - a_{jk})| < \epsilon,
\]

and hence,

\[
|\mu_{i,t+1}(\theta) - \mu_{j,t+1}(\theta)| < \epsilon + \left| \sum_{k=1}^n \mu_{k,t}(\theta)(a_{ik} - a_{jk}) \right|.
\]

Since \( A \) is a stochastic matrix, \( \sum_{k=1}^n (a_{ik} - a_{jk}) = 0 \). Therefore, we can use Paz’s inequality to find an upper bound for the right hand side of the above inequality.\(^{12}\)

\[
|\mu_{i,t+1}(\theta) - \mu_{j,t+1}(\theta)| < \epsilon + \frac{1}{2} \max_{p,q} |\mu_{p,t}(\theta) - \mu_{q,t}(\theta)| \sum_{k=1}^n |a_{ik} - a_{jk}|.
\]

Thus,

\[
\max_{i,j} |\mu_{i,t+1}(\theta) - \mu_{j,t+1}(\theta)| < \epsilon + \tau_1(A) \max_{i,j} |\mu_{i,t}(\theta) - \mu_{j,t}(\theta)|,
\]

\(^{12}\)Paz’s inequality states that if \( d \) is a vector with an entry-wise sum of zero, then for any arbitrary vector \( z \) of the same size, \( |d^T z| \leq \frac{1}{2} \|d\|_1 \max_{i,j} |z_i - z_j| \). This inequality can be found in the book of Paz [87] or Kirkland, Neumann, and Shader [67].
where $\tau_1(A) = \frac{1}{2} \max_{i,j} \sum_{k=1}^{n} |a_{ik} - a_{jk}|$, is the coefficient of ergodicity of matrix $A$, defined in Chapter 2. As was shown earlier, the coefficient of ergodicity of any stochastic matrix lies in the interval $[0, 1]$. This along with the fact that $\epsilon > 0$ is arbitrary imply that $\max_{i,j} |\mu_{i,t}(\theta) - \mu_{j,t}(\theta)|$ is a non-increasing sequence and hence, converges. We claim that the limit is in fact zero. To show this, we consider two cases.

First suppose that $\tau_1(A) < 1$. In that case, for any positive integer $p$, we have

$$\max_{i,j} |\mu_{i,t+p}(\theta) - \mu_{j,t+p}(\theta)| < \sum_{k=0}^{p-1} [\tau_1(A)]^k \epsilon + [\tau_1(A)]^p \max_{i,j} |\mu_{i,t}(\theta) - \mu_{j,t}(\theta)|$$

$$= \frac{1 - [\tau_1(A)]^p}{1 - \tau_1(A)} \epsilon + [\tau_1(A)]^p \max_{i,j} |\mu_{i,t}(\theta) - \mu_{j,t}(\theta)|$$

$$\rightarrow \frac{\epsilon}{1 - \tau_1(A)} \quad \text{as } p \rightarrow \infty.$$

Since $\epsilon > 0$ is arbitrary, the right hand side can be made arbitrarily small, and as a result, $\max_{i,j} |\mu_{i,t}(\theta) - \mu_{j,t}(\theta)|$ must converge to zero; i.e., all agents have asymptotically equal belief.

We now consider the case that $\tau_1(A) = 1$. In this case, as shown by Seneta [97] and discussed in Chapter 2, since $A$ is irreducible and corresponds to a strongly connected graph, there exists a positive integer $r$ such that $\tau_1(A^r) < 1$. Therefore, using a similar argument as above for the matrix $A^r$, one can show that the convergent sequence $\{\max_{i,j} |\mu_{i,t}(\theta) - \mu_{j,t}(\theta)|\}^{\infty}_{i=1}$ has a subsequence that converges to zero. Therefore, on any connected network, if all agents have eventually correct forecasts, then $\mu_{i,t}(\theta) - \mu_{j,t}(\theta) \rightarrow 0$ for all $i, j \in \mathcal{N}$. This completes the proof.

The above proposition states that under the assumptions of strong connectivity, positive self-confidence, and positive prior beliefs, all agents will have asymptotically equal beliefs $P^*$-almost surely. The intuition behind this result is simple. Recall that the belief of each individual is a function of two terms: her own posterior belief after observing a private
signal and the priors of her neighbors. Since the beliefs of all agents are converging, the
correlation contribution of the private signals to the belief update equation must vanish asymptotically.
Thus, at a large enough $t$, the belief of each agent lies simply in the convex hull of the
beliefs held by her neighbors. Therefore, strong connectivity guarantees asymptotically
equal beliefs over the network.

The asymptotic agreement of beliefs guaranteed by Proposition 7 has important conse-
quences for learning. Recall that Proposition 6 shows that all agents can correctly forecast
their observations. Therefore, the asymptotic agreement over the network implies that each
agent, not only can forecast her own signals correctly, but also can correctly forecast the
signals of any other agent in the network as well, even if they are not neighbors. In other
words, the interaction over the network provides the neighbors with potentially much more
information compared to the case that they ignore the beliefs held by their neighbors. The
significance of this result lies in the fact that the agents can extract information from any
other individual, even if they do not necessarily communicate with one another directly.
Notice that this information extraction is occurring while the agents do not update their
beliefs rationally, and instead use a simple and computationally tractable update.

The following theorem, which is the main result of this chapter, summarizes the results
stated so far and characterizes the level of learning achieved in the social network.

**Theorem 4.** Suppose that $A$ is irreducible, stochastic and all its diagonal elements are
uniformly bounded away from zero. Also suppose that there exists an agent $k$ such that
$\mu_{k,0}(\theta^*) > 0$. Then, every agent can eventually forecast the signals of every other agent
correctly with $P^*$-probability one. Moreover, if there exists a state $\theta \in \Theta$ and an agent $i$
such that $\ell_i(\tilde{s}^i|\theta) \neq \ell_i(\tilde{s}^i|\theta^*)$ for some $\tilde{s}^i \in S_i$, then $\mu_j(\theta) \rightarrow 0$ for all $j \in N$.

**Proof:** Proving the first statement is trivial, as it is a simple corollary to Propositions 6
and 7. We turn to the proving the second part of the theorem. In the course of the proof of
Proposition 7, we showed that

\[ \mu_{i,t}(\theta) \left[ \frac{\ell_i(s_{i+1}^i|\theta)}{m_{i,t}(s_{i+1}^i)} - 1 \right] \to 0 \quad (3.7) \]

for all \( i \in \mathcal{N} \) on almost all sample paths. Now suppose there exists an agent \( i \) and a signal \( \tilde{s}^i \in S_i \), such that \( \ell_i(\tilde{s}^i|\theta) \neq \ell_i(\tilde{s}^i|\theta^*) \). Moreover, agent \( i \) observes signal \( \tilde{s}^i \) infinitely often on almost all sample paths, which means that the term \( \frac{\ell_i(s_{i+1}^i|\theta)}{m_{i,t}(s_{i+1}^i)} \) is bounded away from zero infinitely often. Also, notice that \( \mu_{i,t}(\theta) \) converges to some random variable. Therefore, for (3.7) to hold, it must be case that the limit of \( \mu_{i,t}(\theta) \) is zero. As a result, \( \mu_{j,t}(\theta) \) should also converge to zero for all \( j \neq i \). \( \blacksquare \)

This theorem captures how information gets aggregated over the social network when individuals update their beliefs according to the law of motion (3.3). It guarantees that as long as there exists an agent \( i \) whose signals are informative enough to distinguish the state \( \theta \) from the true state \( \theta^* \), everybody else will be able to eventually distinguish the two states. Hence, every individual, regardless of her position in the network, assigns a belief zero to \( \theta \) asymptotically almost surely. The next example shows the power of the Theorem 4.

**Example 2.** Consider the collection of agents \( \mathcal{N} = \{1, 2, \ldots, 7\} \) who are located in a social network as depicted in Figure 3.1: agent \( i \leq 6 \) can observe the beliefs of agent \( i + 1 \) and agent 7 has access to the beliefs of agent 1 at all times. Clearly, this is a strongly connected social network.

Assume that the set of possible states of the world is given by \( \Theta = \{\theta^*, \theta_1, \theta_2, \ldots, \theta_7\} \), where \( \theta^* \) is the true underlying state of the world. We also assume that the signal observed by the agents can only take values in the set \( S_i = \{H, T\} \) for all \( i \), are conditionally
Figure 3.1: The figure illustrates a strongly connected social network of 7 agents, which is of the form of a directed cycle.

The observations of agent $i$ are such that she can only distinguish state $\theta_i$ from the rest of the elements of $\Theta$. Nevertheless, for any given state $\theta \neq \theta^*$, there exists an agent whose signals are informative enough to distinguish the two. Therefore, Theorem 4 implies that if the prior beliefs are in the interior of the belief simplex and all agents have strictly positive self-confidence when applying (3.3), then, $\mu_{i,t}(\theta^*) \to 1$, as $t \to \infty$ for all agents $i$. In other words, all agents will asymptotically learn the true underlying state of the world. Clearly, if the agents discard the information provided to them by their neighbors, they have no means of learning the true state.

Theorem 4 is quite powerful as it completely characterizes the asymptotic beliefs of all individuals in the society, and demonstrates how our non-Bayesian learning model leads to
information aggregation over the network. However it is important to notice that it does not
guarantee complete learning in the sense of Definition 12. Clearly, if there exists a global
identification problem (i.e., there exist $\theta \neq \theta^*$ such that $\ell(s|\theta) = \ell(s|\theta^*)$ for all $s \in S$),
then the two states $\theta$ and $\theta^*$ are observationally equivalent, and therefore, undistinguish-
able. In such cases even a collection of completely rational agents fail in learning the true
state. However, when updating their beliefs based on (3.3), the agents might fail to achieve
complete learning, even in the absence of global identification problem. Such failure is due
to the fact that when updating their beliefs, an individual only considers the marginal distri-
bution of her own signal, and does not take into account the possible correlations between
her own observations and those of her neighbors. Put differently, even if the individuals can
correctly forecast the signals of every other agent, as suggested by Theorem 4, they might
not still be able to learn the joint distribution of the signals. The next example is meant to
clarify this point.

Example 3. Consider a strongly connected social network consisting of two individuals
$\mathcal{N} = \{1, 2\}$. Assume that $\Theta = \{\theta_1, \theta_2\}$, and $S_1 = S_2 = \{H, T\}$. Also assume that the
distribution function describing the random private observations of the agents conditional
on the underlying state of the world is given by the tables below:

\[
\begin{array}{c|cc}
\ell(s_1s_2|\theta_1) & H & T \\
\hline
H & 1/2 & 0 \\
T & 0 & 1/2 \\
\end{array}
\quad
\begin{array}{c|cc}
\ell(s_1s_2|\theta_2) & H & T \\
\hline
H & 0 & 1/2 \\
T & 1/2 & 0 \\
\end{array}
\]

In other words, when the underlying state of the world is $\theta_1$, private observations of the
two agents are perfectly correlated, while in state $\theta_2$, their observations are perfectly nega-
tively correlated.

It is an easy exercise to show that if both agents are completely rational, they can learn
the true underlying state of the world in exactly two time period, simply because the information in $\mu_{i,0}$ and $\mu_{i,1}$ for $i = 1, 2$ is sufficient to completely learn the true state. On the other hand, consider the case that both individuals apply the non-Bayesian learning rule in (3.3). Since $\ell_i(H|\theta_1) = \ell_i(H|\theta_2) = \frac{1}{2}$ for $i = 1, 2$, the agents’ beliefs remain constant over time. This implies that learning as defined in Definition 12 does not occur, while each agent can correctly forecast her own and her neighbor’s signals correctly, at all time steps.

The example above clearly shows the shortcoming of the non-Bayesian belief update (3.3) compared to the case that all agents are completely rational. Nevertheless, the non-Bayesian agents are still capable of extracting valuable information from the information provided to them by their neighbors, while avoiding the highly complex computations that are essential for Bayesian learning: they can learn the true state of the world up to its marginal distributions.

One interesting special case to consider is when the observations of different individuals are conditionally independent from one another for all states of the world, i.e., $\ell(s_1, \ldots, s_n|\theta) = \ell_1(s_1|\theta) \ldots \ell_n(s_n|\theta)$ for all $\theta \in \Theta$. Clearly, in such a case, forecasting the observations of all agents correctly is equivalent to learning the true underlying state of the world. Therefore, as a corollary to Theorem 4, we can conclude that under independent observations, all individuals asymptotically learn the true state of the world as if they were completely rational. In other words, asymptotically, Bayesian and non-Bayesian agents will extract the same amount of information from their neighbors. However, it is important to note that this equivalence does not hold for finite $t$. Clearly, rational agents achieve a higher level of learning at every finite time step compared to their non-Bayesian counterparts. Nevertheless, as long as the agents have a high discount factor and care about their long-run payoffs much more than their transient payoffs, it is preferable for them to use the non-Bayesian belief update in (3.3), given its computational simplicity and tractability.
3.4 Concluding Remarks

In this chapter, we studied a boundedly rational model of dynamic opinion formation in a social network. The individuals in our model fail to incorporate the information provided to them by their neighbors in a fully Bayesian manner, and instead, use a simple, local belief update rule. More specifically, we assumed that at every time period, the belief of each individual is a convex combination of her Bayesian posterior belief and her neighbors’ priors. We showed that under such a boundedly rational model of opinion formation, the individuals will eventually hold correct forecasts about the observations of all agents, as long as the social network is strongly connected. We also showed that if individuals’ observations are conditionally independent, they asymptotically learn the true underlying state of the world as if they were completely informed of all signals and updated their beliefs according to Bayes’ rule. Furthermore, we showed that, in contrast to Bayesian learning, absolute continuity of the true measure with respect to all prior beliefs is not necessary condition for social learning: as long as there is some individual with a positive prior belief on the true parameter, social learning is achieved.

The other characteristic of our model is that aggregation of information is achieved with very low computational costs for the agents. First, individuals do not need to have any information about the global structure of the social network, as they only update their beliefs locally and do not make any deductions beyond their immediate neighbors. Second, they do not need to know the signal structure of any other agent in the network, besides their own. the simplicity of the local update rule guarantees that the agents eventually achieve full learning, while at the same time, avoiding highly complex computations that are essential for full Bayesian learning over the network.
Part II

New Models of Networks: Beyond Graphs
Chapter 4

From Graphs to Simplicial Complexes

The first step in understanding the common underlying principles of complex networks and their various applications is to apply the proper level of abstraction in modeling them. As in almost all scientific studies, the key step is to capture the right level of structure in the model, while abstracting away small scale details that are redundant to the specific problem at hand.

A useful abstraction for modeling and analysis of complex networks has been developed using graph theory. Typically interaction among agents is modeled with graphs in which vertices represent agents (e.g. unmanned vehicles, sensors, individuals) and edges represent some form of proximity or other binary relationships (e.g., proximity in distance, communication link, friendship) [81]. Although graphs are often adequate discrete abstractions for network phenomena, there are plenty of challenging applications and scenarios where the abstraction based on binary relations is too crude. More specifically, in many applications, instead of only looking at pairwise relations between agents, one might find it useful (or even necessary) to consider 3-way or 4-way relations as well. For example, a coalition of 3 agents is different from the 3 pairwise relations between any two of them. In such cases, using a simplicial complex as the network model can potentially lead to the
extraction of more information, as it takes the higher order relations between different entities into account. In this modeling paradigm, extraction of information from higher order relations can then be achieved by the means of the tools in the field of algebraic topology, especially homology theory which deals with global topological properties of the space under study.

The application of the theory of simplicial complexes and simplicial homology goes well beyond networks. In fact, simplicial complexes can serve as general geometric representations for a broad spectrum of modeling problems [41]. For example, in many situations that a geometrical space is represented by a finite set of points sampled from it - what is known as point cloud data - it is natural to use the machinery of simplicial complexes and homology to recover the attributes of the original space from the sampled data [20,28]. More recently, simplicial (and other type of) complexes and homology theory have been used in the study of proteins and other molecules: the protein is modelled as a union of balls, one ball per atom, and the complex used is dual to this union [40–42]. Other applications range from modeling dynamical systems [5] to configuration spaces of graphs in robotics [53], as well as compression of data sets and coverage verification in sensor networks [30, 79].

In this chapter, we present the theory of simplicial complexes as a generalization of graphs and an object of study in algebraic topology, as a more faithful modeling framework for networks. The chapter can be considered as a brief review of the main concepts of simplicial complexes and the theory of simplicial homology. Moreover, it presents the theory of combinatorial Laplacians corresponding to simplicial complexes and shows their potential benefit in implementing distributed algorithms. The other goal of the chapter is to relate the concepts presented to their more well-known counterparts in graphs theory. A thorough treatment of the subject can be found in [61] and [82].

The benefits of the concepts and tools presented in this chapter, in the context of mod-
eling complex networks, are highlighted through a specific application in Chapter 5. The presented application, on coverage verification in coordinate-free sensor networks, depicts how this new modeling paradigm can extract more information from the network than simply using the standard graph model.

4.1 Simplicial Complexes

Given a set of points $V$, a $k$-simplex (or a simplex of dimension $k$) is an unordered set $\{v_0, v_1, \ldots, v_k\} \subseteq V$ where $v_i \neq v_j$ for all $i \neq j$. A face of the $k$-simplex $\{v_0, v_1, \ldots, v_k\}$ is a $(k - 1)$-simplex of the form $\{v_0, \ldots, v_{i-1}, v_{i+1}, \ldots, v_k\}$ for some $0 \leq i \leq k$. Clearly, any $k$-simplex has exactly $k + 1$ faces.

Definition 13. A simplicial complex $X$ is a finite collection of simplices which is closed with respect to inclusion of faces, i.e., if $\sigma \in X$, then all faces of $\sigma$ are also in $X$.

Roughly speaking, a simplicial complex is a generalization of a graph, in the sense that in addition to binary relations between the elements of $V$, it captures higher order relations between them as well. Note that due to the requirement of closure with respect to the inclusion of the faces, a simplicial complex is different from a hyper-graph, in which any subset of the power set of $V$ can be considered as a hyper-edge. Figure 4.1 depicts a simplicial complex consisting of 11 vertices (0-simplices), 14 edges (1-simplices), five 2-simplices and one 3-simplex.

The dimension of a simplicial complex is the maximum dimension of any of its simplices. Clearly, any undirected graph can be considered as a simplicial complex of dimension one. A subcomplex of $X$ is a simplicial complex $Y \subseteq X$. A particular subcomplex of $X$ is its $k$-skeleton consisting of all simplices of dimension $k$ or less, denoted by $X^{(k)} = \{\sigma \in X : \dim \sigma \leq k\}$. Therefore, the 1-skeleton of any non-empty simplicial com-
plex is a graph. Given a graph $G$, its flag complex $F(G)$ is the largest simplicial complex whose 1-skeleton is $G$; every $(k + 1)$-clique in $G$ defines a $k$-simplex in $F(G)$.

Given a simplicial complex $X$, two $k$-simplices $\sigma_i$ and $\sigma_j$ are upper adjacent (denoted by $\sigma_i \sim \sigma_j$) if both are faces of a $(k + 1)$-simplex in $X$. The two $k$-simplices are said to be lower adjacent (denoted by $\sigma_i \sim \sigma_j$) if both have a common face. These notions of adjacency are simply higher order variants of the concept of neighborhood defined for graphs. Having defined the concept of adjacency, one can define the upper and lower adjacency matrices, $A^{(k)}_u$ and $A^{(k)}_l$ respectively, in order to book keep the adjacency relations between the $k$-simplices. More specifically, for any $k \geq 0$, $A^{(k)}_u = [a^{(k)}_{ij}]$ is an $n_k \times n_k$ matrix, with $n_k$ being equal to the number of $k$-simplices in the complex, whose rows and columns are indexed by the $k$-simplices of $X$ such that $a^{(k)}_{ij} = 1$ if $k$-simplices $\sigma_i$ and $\sigma_j$ are upper adjacent, and $a^{(k)}_{ij} = 0$ otherwise. The definition of the lower adjacency matrix $A^{(k)}_l$ for $k \geq 1$ is similar. Note that the upper adjacency matrix of order zero of a simplicial complex, $A^{(0)}_u$, coincides with the well-known notion of the adjacency matrix of the graph capturing its 1-skeleton.

One can also generalize the concept of degree for graphs. The upper degree of a $k$-simplex $\sigma$ in $X$, denoted $\deg_u(\sigma)$, is the number of $(k + 1)$-simplices in $X$ of which $\sigma$ is a face. Similarly, the lower degree of $\sigma$, denoted $\deg_l(\sigma)$, is the number of nonempty $(k - 1)$-
simplices in $X$ that are faces of $\sigma$. We can now define the upper and lower degree matrices of order $k$, $D_{u}^{(k)}$ and $D_{l}^{(k)}$ respectively, as diagonal matrices indexed with the number of $k$-simplices in $X$, with $i$-th diagonal entry equal to $\deg_{u}(\sigma_{i})$ and $\deg_{l}(\sigma_{i})$, respectively.

**Examples: Rips and Čech complexes**

In order to clarify the concepts defined so far, we define two well-known simplicial complexes, known as the Rips and Čech complexes corresponding to a set of points in a Euclidean space. These complexes are widely used in different applications, such as topological data analysis\(^1\) and coverage verification in coordinate-free sensor networks.

**Definition 14.** Given a set of points $V = \{v_{1}, \cdots, v_{n}\}$ in a finite dimensional Euclidean space and a fixed radius $\epsilon$, the *Vietoris-Rips complex* of $V$, $R_{\epsilon}(V)$, is the abstract simplicial complex whose $k$-simplices correspond to unordered $(k + 1)$-tuples of points in $V$ which are pairwise within Euclidean distance $\epsilon$ of each other [108].

The Rips complex can be considered as the generalization of the geometric graph defined over $V$ to higher dimensions. In fact, to be more precise, the Rips complex is simply the flag complex of the proximity graph of $V$, whose edges are pairs of points $v_{i}, v_{j} \in V$ with $\|v_{i} - v_{j}\| \leq \epsilon$. A different choice of how to to fill in the higher dimensional simplices of the proximity graph is to construct the Čech or Nerve complex corresponding to the set of points $V$, defined below.

**Definition 15.** Given a finite collection of points $V = \{v_{1}, \cdots, v_{n}\}$, the *Čech complex*, $C_{\epsilon}(V)$, is the abstract simplicial complex whose $k$-simplices are determined by unordered $(k + 1)$-tuples of points $v_{i} \in V$ whose closed $\epsilon$-ball neighborhoods have a point of common intersection.

---

\(^1\)For example, see the works of Carlsson, de Silva, Edelsbrunner, and Zomorodian [29, 43, 113, 114].
4.1.1 Boundary Homomorphism

In this subsection, we define the boundary map, which is the key in formally defining homologies. Let $X$ denote a simplicial complex. Similar to graphs, an orientation can be defined for $X$ by defining an ordering on all of its $k$-simplices. We denote the $k$-simplex $\{v_0, \cdots, v_k\}$ with an ordering by $[v_0, \cdots, v_k]$. For each $k \geq 0$, define $C_k(X)$ to be the vector space whose basis is the set of oriented $k$-simplices of $X$, where a change in the orientation corresponds to a change in the sign of the coefficient as $[v_0, \cdots, v_i, \cdots, v_j, \cdots, v_k] = -[v_0, \cdots, v_j, \cdots, v_i, \cdots, v_k]$. We let $C_k(X) = 0$, if $k$ is larger than the dimension of $X$.

Therefore, by definition, elements of $C_k(X)$, called $k$-chains, can be written as finite formal sums $\sum_j \alpha_j \sigma_j^{(k)}$ where the coefficients $\alpha_j \in \mathbb{R}$ and $\sigma_j^{(k)}$ are the oriented $k$-simplices of $X$.\(^2\) Note that $C_k$ is a finite-dimensional vector space with the number of $k$-simplices as its dimension. We now define the boundary map.

**Definition 16.** For an oriented simplicial complex $X$, the $k$-th simplicial boundary map is a homomorphism $\partial_k : C_k(X) \rightarrow C_{k-1}(X)$, which acts on the basis elements of its domain via

$$\partial_k[v_0, \cdots, v_k] = \sum_{j=0}^k (-1)^j[v_0, \cdots, v_{j-1}, v_{j+1}, \cdots, v_k]. \quad (4.1)$$

Intuitively, the above operator maps a $k$-chain to its faces. For example, the boundary of a directed path in a graph (which is an oriented 1-chain) is simply the difference between its two end points,

Since for any finite simplicial complex $C_k(X)$ is a finite dimensional vector space for all $k$, $\partial_k$ has a matrix representation. We denote the matrix representation of the $k$-th boundary map relative to the bases of $C_k$ and $C_{k-1}$ by $B_k \in \mathbb{R}^{n_{k-1} \times n_k}$, where $n_k$ is the number of

---

\(^2\)To be more precise, this is the definition of $k$-chains with coefficients in $\mathbb{R}$. In most algebraic topology texts such as [61], the $k$-chains are defined over integers rather than reals. In such a case, $C_k(X)$ is defined as a free abelian group with the set of oriented $k$-simplices as its basis. However, as in [38], we find it more convenient to define the chains over $\mathbb{R}$.
$k$-simplices of $X$. In particular, the matrix representation of the first boundary map $\partial_1$ is nothing but the edge-vertex incidence matrix of a graph which maps edges (1-simplices) to vertices (0-simplices).

Using (4.1), it is an easy exercise to prove the following lemma.

**Lemma 5.** The map $\partial_k \circ \partial_{k+1} : C_{k+1}(X) \to C_{k-1}(X)$ is uniformly zero for all $k \geq 1$.

In other words, the boundary of any $k$-chain has no boundary.

### 4.1.2 Simplicial Homology

In this subsection we define the concept of homology for simplicial complexes. Homology is a topological invariant that is quite popular in computational topology as it is easily computable. Homologies may be regarded as an algebraization of the first layer of geometry in simplicial complexes: how simplices of dimension $k$ attach to simplices of dimension $k - 1$ [61, 112].

Let $X$ denote a simplicial complex. Consider the following two subspaces of $C_k(X)$:

$$
\begin{align*}
\text{k-cycles} & : \ker \partial_k = \{ x \in C_k(X) : \partial_k x = 0 \} \\
\text{k-boundaries} & : \text{img} \partial_{k+1} = \{ x \in C_k(X) : \exists y \text{ s.t. } x = \partial_{k+1} y \}
\end{align*}
$$

An element in $\ker \partial_k$ is a subcomplex without a boundary and therefore represents a $k$-dimensional cycle, while the elements in $\text{img} \partial_{k+1}$ are boundaries of higher dimensional chains and are known as $k$-boundaries. For example, any directed path on a graph, whose two end points coincide is a 1-cycle. On the other hand, the difference of two vertices forms a 0-boundary if there exists a path on the complex connecting the two.

The $k$-cycles are the basic objects that count the presence of “$k$-dimensional holes” in the simplicial complex [30]. But, certainly, many of the $k$-cycles in $X$ are measuring the same hole; still other cycles do not really detect a hole at all — they bound a subcomplex
of dimension \( k + 1 \) in \( X \). In fact, we say two \( k \)-cycles \( \xi \) and \( \eta \) are homologous if their difference is a boundary: \( \xi - \eta \in \text{img} \, \partial_{k+1} \). Therefore, as far as measuring holes is concerned, homologous cycles are equivalent. Consequently, it makes sense to define the quotient vector space

\[
H_k(X) = \ker \partial_k / \text{img} \, \partial_{k+1},
\]

known as the \( k \)-th homology of \( X \), as the proper vector space for distinguishing homologous cycles. Note that according to Lemma 5, we have \( \partial_k \circ \partial_{k+1} = 0 \), implying that \( \text{img} \, \partial_{k+1} \) is a subspace of \( \ker \partial_k \), and therefore, making \( H_k(X) \) a well-defined vector space.\(^3\)

Roughly speaking, when constructing the homology, we are removing the cycles that are boundaries of a higher order subcomplex from the set of all \( k \)-cycles, so that the remaining ones carry information about the \( k \)-dimensional holes of the complex. A more precise way of interpreting (4.2) is that any element of \( H_k(X) \) is an equivalence class of homologous \( k \)-cycles. Moreover, it inherits the structure of a vector space in the natural way: \([\xi] + [\eta] = [\xi + \eta]\) and \( c[\xi] = [c\xi] \) for \( c \in \mathbb{R} \), where \([\xi]\) represents the equivalence class of all \( k \)-cycles homologous to \( \xi \). Therefore, each non-trivial homology class\(^4\) in a certain dimension identifies a corresponding “hole” in that dimension. In fact, the dimension of the \( k \)-th homology of \( X \) (known as its \( k \)-th Betti number) identifies the number of \( k \)-dimensional holes in \( X \). For example, the dimension of \( H_0(X) \) is the number of connected components of \( X \), whereas the dimension of \( H_1(X) \) is equal to the number of holes in its 2-skeleton. Similarly, \( H_2(X) \) identifies the number of 3-dimensional voids in \( X \) and so on.

\(^3\)If we define the \( k \)-chains over integers, then \( \text{img} \, \partial_{k+1} \) becomes a normal subgroup of \( \ker \partial_k \). In that case, the homology is defined as the quotient group \( H_k = \ker \partial_k / \text{img} \, \partial_{k+1} \).

\(^4\)By the trivial homology class, we mean the equivalence class of all null-homologous \( k \)-cycles on the simplicial complex.
4.1.3 Relative Homology

In some applications, one may need to compute the holes modulo some region of space, such as the boundary. The concept of relative homology is defined for this purpose.

Given a simplicial complex $X$ and a subcomplex $A \subset X$, let $C_k(X, A)$ be the quotient vector space $C_k(X) / C_k(A)$. In other words, the chains in $A$ are trivial in $C_k(X, A)$. Since the boundary map $\partial_k : C_k(X) \to C_{k-1}(X)$ takes $C_k(A)$ to $C_{k-1}(A)$, it induces a quotient boundary map $\bar{\partial}_k : C_k(X, A) \to C_{k-1}(X, A)$. One can verify that the subspaces defined by the kernel and image of the quotient map are well-defined and satisfy $\text{img } \bar{\partial}_{k+1} \subset \text{ker } \bar{\partial}_k \subset C_k(X, A)$. Therefore, similar to before, one can define the $k$-th relative homology as the quotient vector space $\[ 61 \]

$$H_k(X, A) = \ker \bar{\partial}_k / \text{img } \bar{\partial}_{k+1}. \hspace{1cm} (4.3)$$

Given the above definition, one can interpret elements of $H_k(X, A)$ as representatives for relative cycles: $k$-chains $\xi \in C_k(X)$ such that $\partial_k \xi \in C_{k-1}(A)$. Moreover, such a relative cycle $\xi$ is trivial in $H_k(X, A)$ if and only if it is a relative boundary: $\xi = \partial_{k+1} \eta + \gamma$ for some $\eta \in C_{k+1}(X)$ and $\gamma \in C_k(A)$. Figure 4.2 is meant to clarify this concept. It depicts the 2-skeleton of a simplicial complex $X$ and the subcomplex $A \subset X$ consisting of all the boundary vertices and edges (heavy lines). Both $\xi_1$ and $\xi_2$ (highlighted) are relative 1-cycles, but only $\xi_1$ is the representative of a non-trivial element in $H_1(X, A)$.

4.2 Combinatorial Laplacians

The graph Laplacian [76] has various applications in image segmentation, graph embedding, dimensionality reduction for large data sets, machine learning, and more recently in consensus and agreement problems in distributed control of multi-agent systems [64, 84].
Figure 4.2: Relative homology: $\xi_1$ is the representative of a non-trivial element in $H_1(X, A)$.

For a graph $G$, the Laplacian matrix is defined as $L = BB^T$ where $B$ is the vertex-by-edge-dimensional incidence matrix of $G$. As it is evident from the definition, $L$ is a positive semi-definite matrix. Also it is well-known that the Laplacian matrix can be written in terms of the adjacency and degree matrices of $G$ as well: $L = D - A$, which implies that the $i$-th row of the Laplacian matrix only depends on the local interactions between vertex $i$ and its neighbors. The goal of this subsection is to present the generalization of this matrix to simplicial complexes and investigate its properties. The importance of these generalized Laplacian matrices (known as combinatorial Laplacians) lies in an observation made by Eckemman [39]; the fact that when working with real coefficients, the kernel of such a matrix spans a subspace isomorphic to the homologies.

The definitions and results of this subsection can be found in [39] and [38].

**Definition 17.** Let $X$ be a finite oriented simplicial complex. The $k$-th combinatorial Laplacian of $X$ is the homomorphism $L_k : C_k(X) \to C_k(X)$ given by

$$L_k = \partial^*_k \circ \partial_k + \partial_k \circ \partial^*_{k+1}$$

(4.4)
where $\partial_k^*$ is the adjoint of the operator $\partial_k$ with respect to the inner product that makes the basis orthonormal.

The Laplacian operator, as defined above, is the sum of two positive semi-definite operators and therefore, any $k$-chain $x \in \ker L_k$ satisfies

$$x \in \ker \partial_k, \quad x \perp \text{img } \partial_{k+1}$$

In other words, the kernel of the $k$-th combinatorial Laplacian consists of $k$-cycles which are orthogonal to the subspace $\text{img } \partial_{k+1}$, and therefore, are not $k$-boundaries. This implies that the non-zero elements in the kernel of $L_k$ are representatives of the non-trivial equivalence classes of cycles in the $k$-th homology. This property was first observed by Eckmann [39] and is formalized in the following theorem [38].

**Theorem 5.** *If the vector spaces $C_k(X)$ are defined over $\mathbb{R}$, then for all $k$ there is an isomorphism*

$$H_k(X) \cong \ker L_k \quad (4.5)$$

*where $H_k(X)$ is the $k$-th homology of $X$ and $L_k$ is its $k$-th combinatorial Laplacian. Moreover, there is an orthogonal direct sum decomposition of the vector space $C_k(X)$ in the form of*

$$C_k(X) = \text{img } \partial_{k+1} \oplus \ker L_k \oplus \text{img } \partial_k^*,$$

*in which the first two summands comprise the set of $k$-cycles $\ker \partial_k$, and the first summand is the set of $k$-boundaries.*

The immediate implication of the above theorem is that the dimension of the subspace in the kernel of the $k$-th combinatorial Laplacian operator is equal to the $k$-th Betti number of the simplicial complex. The next example is meant to clarify the statement of Theorem 5. But first, note that, as stated earlier, the boundary operators defined over a finite simplicial
Figure 4.3: A simplicial complex of dimension 2. The numbers labeled on the edges correspond to the components of the eigenvector corresponding to 0 eigenvalue, i.e. the kernel, of the first combinatorial Laplacian matrix.

complex have matrix representations with respect to the bases of vector spaces $C_k(X)$. Therefore, one can use matrices to represent the combinatorial Laplacian operators in a similar manner: define the $k$-th combinatorial Laplacian matrix as

$$L_k = B_k^T B_k + B_{k+1}^T B_{k+1} \in \mathbb{R}^{n_k \times n_k}$$

(4.6)

where $B_k$ is the matrix representation of $\partial_k$ and $n_k$ is the number of $k$-simplices of $X$. Note that the expression for $L_0$ reduces to the well-known graph Laplacian matrix. Similarly, the combinatorial Laplacian matrices can be represented in terms of the adjacency and degree matrices [55, 79] of the simplicial complex. More precisely, for $k > 0$,

$$L_k = D^{(k)}_u - A^{(k)}_u + (k + 1) I_{n_k} + A^{(k)}_l,$$

(4.7)

where $A^{(k)}_u$ and $A^{(k)}_l$ are the upper and lower adjacency matrices, respectively and $D^{(k)}_u$ represents the upper degree matrix. (4.7) implies that the $i$-th row of $L_k$ only depends on the local interactions between $i$-th $k$-simplex and its upper and lower adjacent $k$-simplices. This property is also an extension of the locality property of the graph Laplacian matrix.

**Example 4.** Consider the oriented simplicial complex depicted in Figure 4.3, which consists of 6 vertices, 8 edges and 2 triangles. It is an easy exercise to show that the first
combinatorial Laplacian matrix is given by

\[
L_1 = \begin{pmatrix}
2 & -1 & 0 & 0 & 0 & -1 & 0 & 0 \\
-1 & 2 & -1 & 0 & 0 & 0 & -1 & -1 \\
0 & -1 & 3 & 0 & 0 & 0 & 0 & 1 \\
0 & 0 & 0 & 3 & -1 & 0 & 0 & 0 \\
0 & 0 & 0 & -1 & 3 & -1 & 0 & 0 \\
-1 & 0 & 0 & 0 & -1 & 2 & 0 & -1 \\
0 & -1 & 0 & 0 & 0 & 0 & 0 & 4 & 0 \\
0 & -1 & 1 & 0 & 0 & -1 & 0 & 3
\end{pmatrix}
\]

where the edges are ordered as \([v_1v_2], [v_2v_3], [v_3v_4], [v_4v_5], [v_5v_6], [v_6v_1], [v_3v_5], \) and \([v_3v_6]\).

The kernel of \(L_1\) is a one dimensional subspace spanned by the vector \([8 \ 8 \ 1 \ 1 \ 3 \ 8 \ 2 \ 5]^T\). In Figure 4.3, these values are depicted as flows on the edges of the simplicial complex. One can make the following observations based on the above computation: first, the dimension of \(\ker L_1\) is equal to the number of 1-dimensional holes in the simplicial complex, as suggested by Theorem 5. Moreover, for any \(x \in \ker L_1\) the value of the algebraic sum of the flows entering each vertex is equal to zero. This is a consequence of the fact that any element in \(\ker L_1\) is also in \(\ker B_1\). Finally, note that the algebraic sum of the flows over any filled-in region is equal to zero as well. This is due to the fact that if \(x\) is in \(\ker L_1\), then \(B_2^T x = 0\) and therefore, \(x\) is orthogonal to \(\text{img } B_2\).
Chapter 5

Application: Coverage Verification in Coordinate-Free Sensor Networks

In Chapter 4, we presented a brief review on the theory of simplicial complexes and some related concepts such as simplicial homology, relative homology, and the combinatorial Laplacians. We also argued how these concepts generalize similar concepts in graph theory to higher dimensions; concepts such as edges, connectivity, and the graph Laplacian.

In this chapter, we show how the concepts and tools presented in the previous chapter can be used as more faithful models of networks. More specifically, by focusing on the problem of coverage verification in coordinate-free sensor networks, we show the benefits of capturing higher order relations (beyond the pairwise relation, as in graphs) between different entities in the network.

5.1 Motivation and Related Literature

Recent advances in computing, communication, sensing and actuation technologies, have brought networks composed of hundreds or even thousands of inexpensive mobile sensing
platforms closer to reality. This has induced a significant amount of interest in development of analytical tools for predicting the behavior, as well as controlling the complexities of such large-scale sensor networks. Designing algorithms for deployment, localization, duty-cycling, communication and coverage verification in sensor networks form the core of this active area of research.

Of the most fundamental problems in this domain is the *coverage problem*. In general, this reflects how well an area of interest is monitored or tracked by sensors. In most applications, we are interested in a reliable coverage of the environment in such a way that there are no gaps left in the coverage. Algorithms for this purpose have been extensively studied [50]. One of the most prominent approaches for addressing the coverage problem has been the ‘computational geometry’ approach, in which one uses the coordinates of the nodes and standard geometric tools (such as Delaunay triangulations or Voronoi diagrams) to determine coverage [27, 71, 75, 111]. One very well-known example of utilizing this geometric approach is in solving the *Art Gallery Problem*, in which one determines the number of observers necessary to cover an art gallery such that every point in the gallery is monitored by at least one observer [52, 98].

Such geometrical approaches often suffer from the drawback that they can be too expensive to compute in real-time. Moreover, in most applications, they require exact knowledge of the locations of the sensors. Although, this information can be made available in real-time by a localization algorithm or by the means of localization devices (such as GPS), it can only be used most effectively in an off-line pre-deployment analysis for large networks or when there are strong assumptions on geometrical structure of the network and the environment [81]. This drawback becomes more evident if the network topology changes due to node mobility or sensor failure. In such cases, a continuous monitoring of the network coverage becomes prohibitive if the algorithm is too expensive to run or is sensitive to location uncertainty. Finally, localization equipments add to the cost of the network, which can
be a limiting factor as the size of the network grows. Consequently, a minimal geometry approach for addressing these issues becomes essential.

In this chapter, we show how topological spaces and their topological invariants can be used in addressing the coverage problem in the absence of geometric data, such as location or orientation information: simplicial complexes, as more general models of networks, are used in modeling the sensor network in order to preserve many global geometric properties of the network, while abstracting away the small scale redundant details. Moreover, we show how homology theory can be used for inferring the properties of the sensor network.

The chapter also contains three distributed algorithms that can be used in verifying coverage, localizing coverage holes, and detecting redundant sensors in the network. We first present a distributed algorithm that by using the machinery of the combinatorial Laplacians, can verify whether coverage is successful and “hole-free”. Our second algorithm, which is based on the ideas of de Silva and Ghrist [30] and Muhammad and Egerstedt [79], is capable of “localizing” coverage holes in a network of sensors without any metric information. We show that given a generator in the first homology of the Rips complex corresponding to the sensor network, the problem of finding the “tightest” cycle encircling the hole represented by that homology class can be formulated as an integer programming problem. Furthermore, we present conditions under which the linear programming relaxation of this integer programming problem is exact and therefore, its solution provides the location of the coverage holes in the simplicial complex without use of any coordinate information. This optimization-based approach is a direct generalization of network flow algorithms on graphs to simplicial complexes. Finally, we show that if subgradient methods [11, 89, 99] are used for solving this relaxation, the updates are distributed in nature and therefore, the computation of the tightest cycle around the holes can be implemented in a distributed fashion.

The last algorithm presented in this chapter, concerns detecting redundancies in the
sensor network. Once again, based on the ideas and tools introduced in Chapter 4, we present a novel approach for computing a minimal set of sensors required to cover the entire domain. The problem of computing the sparsest generator of the second homology of the Rips complex with respect to its boundary is formulated as an integer-programming problem. By solving the LP relaxation of this integer programming problem in a distributed way, we compute a sparse set of sensors that cover the region of interest.

The results presented in this chapter are based on multiple works, where topological spaces and their invariants are used for addressing the coverage problem [30–32, 54, 79–81]. The idea of using simplicial complexes and homology theory for the study of coverage in sensor networks goes back to the work of Ghrist and Muhammad [54]. Their work was extended by de Silva, Ghrist, and Muhammad [32] and de Silva and Ghrist [31], who present a relative homological criterion for coverage. These results are further extended by de Silva and Ghrist [30] to networks without boundary, the pursuit-evasion problem, and barrier coverage in 3-D. The first steps for implementing the above mentioned ideas as distributed algorithms are taken by Muhammad and Egerstedt [79] and Muhammad and Jadabaie [80]. Muhammad and Egersdet use combinatorial Laplacians to implement a coverage verification algorithm in a distributed way. The chapter also contains two algorithms, for hole localization and detection of redundant sensors, that are based on the work of Tahbaz-Salehi and Jadabaie [104].

5.2 Problem Formulation

Consider a collection of $n$ stationary sensors, denoted by $V$, deployed over a region of interest $D \subset \mathbb{R}^2$. These sensors are equipped with local communication and sensing capabilities: each sensor is only capable of communicating with a limited number of other sensors in its proximity, and has a limited sensing range. Furthermore, assume a complete
absence of localization capabilities and metric information, in the sense that sensors in this network can determine neither distance nor direction. Under these assumptions, we are interested in distributed algorithms for coverage verification. In particular, we are interested in verifying the existence of coverage holes, compute their locations, and detect redundancies in the network.

The following two frameworks are adopted as the coverage models for which we present our coverage verification algorithms:

### 5.2.1 Simplicial Coverage Framework

In this framework, we assume that each sensor is capable of communicating with other sensors within a radially symmetric domain of radius $r_b$, called the broadcast disk. As for the coverage, we assume a “capture” modality in which any subset of nodes which are in pairwise communication cover their entire convex hull. In other words, the region covered by the sensors is given by

$$A(V) = \bigcup \{ \text{conv}(Q) | Q \subseteq V, \max_{v_i, v_j \in Q} \| v_i - v_j \|_2 \leq r_b \}$$

where $V$ is the set of sensor locations and $v_i$ represents the location of the $i$-th sensor. This model, which is inspired by [21], guarantees that the coverage and communication capabilities of the sensors are limited and based on proximity.

### 5.2.2 Symmetric Coverage Framework

Similar to the previous framework, we assume that each sensor is capable of communicating with other sensors within a distance $r_b$. However, unlike the simplicial coverage model, we assume that each sensor is capable of covering a radially symmetric area with radius $r_c$, known as the coverage radius. We refer to the disk covered by a sensor located at point
$v \in V$ as its coverage disk. The region covered by the sensors is given by $U(V) = \bigcup_{v \in V} U_v$, where $U_v = \{x \in \mathbb{R}^2 : \|x - v\| \leq r_c\}$ is the coverage disk corresponding to the sensor located at point $v$. Clearly, region of interest $\mathcal{D}$ is completely covered if it is a subset of $U(V)$. For technical reasons that will become clear in the following sections, we assume that $r_b \leq r_c \sqrt{3}$. The study of this framework is motivated by networks consisting of sensors with omnidirectional communication and sensing capabilities.

In order to be able to address the coverage problem for the above frameworks, we assume that the domain $\mathcal{D}$ is connected and compact and its boundary $\partial \mathcal{D}$ is connected and piecewise linear. Moreover, to avoid boundary effects, it is necessary to assume that there are sensors, known as fence nodes, located on $\partial \mathcal{D}$ such that each such sensor is capable of communicating with its two closest neighbors on $\partial \mathcal{D}$ on either side.

### 5.3 Distributed Coverage Verification

In this section, we present a distributed coverage verification algorithm that can be used in the absence of any metric information. Unlike computational geometry approaches for coverage, this algorithm is based on computational algebraic topology which does not depend on location and orientation information. In essence, we compute the kernel of the first combinatorial Laplacian of a simplicial complex corresponding to the cover and use the fact that the first homology of the cover is trivial, if and only if the coverage is hole-free. The contents of this section are mainly based on the works of de Silva and Ghrist [30] and Muhammad and Egerstedt [79].

**Simplicial Coverage Framework**

We first investigate the simplicial coverage framework: Let $V = \{v_1, \cdots, v_n\}$ denote the locations of $n$ sensors deployed over a region $\mathcal{D} \subset \mathbb{R}^2$, satisfying the assumptions presented
in Section 5.2. These sensors are equipped with local communication capabilities, which enables them to exchange information with other sensors in their proximity: two sensors are capable of communicating with each other if the distance between them is less than or equal to \( r_b \). As for coverage, we assume that any subset of the nodes in pairwise communication can cover their entire convex hull. This implies that the region covered by the sensors is given by

\[ \mathcal{A}(V) = \bigcup \{ \text{conv}(Q) \mid Q \subseteq V, \max_{v_i, v_j \in Q} \| v_i - v_j \|_2 \leq r_b \}. \]

We are interested in verifying whether all points within \( D \) are monitored by the sensors, i.e., whether \( D \subseteq \mathcal{A}(V) \). Our assumptions of Section 5.2 regarding the fence nodes guarantee that \( \partial D \subseteq \mathcal{A}(V) \).

Since no location information is available to the sensors, we need to capture their communication and coverage properties combinatorially. In the simplicial coverage framework, the communication links uniquely determine the coverage pattern provided by the sensors. Also recall from Chapter 4 that the Rips complex of a set of points is simply the flag complex of its proximity graph. Therefore, it is reasonable to expect that the Rips complex corresponding to the set of sensors to contain some information about set \( \mathcal{A}(V) \). In fact, the covered region \( \mathcal{A}(V) \) is nothing but the image of the canonical projection map \( p : \mathcal{R}_{r_b}(V) \to \mathbb{R}^2 \) that maps each simplex in the Rips complex affinely onto the convex hull of its vertices in \( \mathbb{R}^2 \). We refer to the image of map \( p \) as the Rips shadow. The following theorem, proved by Chambers et. al [21], indicates that the Rips complex is rich enough to contain the required topological and geometric properties of its shadow.

**Theorem 6.** Let \( V \) denote a finite set of points in the plane, with the corresponding Rips complex \( \mathcal{R}_e(V) \). Then the induced homomorphism \( p_* : \pi_1(\mathcal{R}_e(V)) \to \pi_1(\mathcal{A}(V)) \) between the fundamental groups of the Rips complex and its shadow is an isomorphism.

Equivalently, Theorem 6 states that a cycle \( \gamma \) in the Rips complex is contractible if
and only if its projection $p(\gamma)$ is contractible in the Rips shadow [22]. The important implication of this theorem is that the first homology groups of the Rips complex and its shadow are also isomorphic. Therefore, the triviality of the first homology of the Rips complex provides a necessary and sufficient condition for a hole-free coverage of $\mathcal{D}$.

Another desirable property of the Rips complex is that it can be formed just by using communication among nearest neighbors. This is due to the fact that Rips complex is the flag complex of the proximity graph and as a result, solely depends on connectivity information. This property makes the Rips complex a desirable combinatorial abstraction of the sensor network, which can be used for distributed coverage verification in the absence of location information. Also, as stated in Chapter 4, the combinatorial Laplacians carry valuable information about the topological properties of a simplicial complex. In particular, $\ker L_1(\mathcal{R}_{rb}) = \{0\}$ guarantees that $H_1(\mathcal{R}_{rb})$ is trivial and as a result, all 1-cycles over the Rips complex are null-homologous. Therefore, according to Theorem 6, $\ker L_1(\mathcal{R}_{rb}) = \{0\}$ serves as a necessary and sufficient condition for the Rips shadow to be hole-free.

One way to compute a generic element in the kernel of the Laplacian matrix is through dynamical system $\dot{x}(t) = -L_1x(t)$ which asymptotically converges to such an element. This implies the following theorem which was first stated and proved in [79].

**Theorem 7.** The linear dynamical system

$$\dot{x}(t) = -L_1x(t), \quad x(0) = x_0 \in \mathbb{R}^{n_1} \quad (5.1)$$

is globally asymptotically stable if and only if $H_1(\mathcal{R}) = 0$, where $x(t)$ is a vector of dimension $n_1$ (the number of 1-simplices of the simplicial complex) and $L_1$ is the first combinatorial Laplacian matrix of the Rips complex $\mathcal{R}_{rb}$.

Note that for any initial condition $x(0)$, the trajectory $x(t); t \geq 0$ always converges to a point in $\ker L_1$. Thus, the asymptotic stability of the system is an indicator of an underlying
trivial homology. In different terms, since \( x^* = \lim_{t \to \infty} x(t) \) is an element in the null space of \( L_1 \), it is a representative of a homology class of the Rips complex. Clearly, if \( x^* = 0 \) for all initial conditions, then the first homology of the simplicial complex consists of only a trivial class and therefore, the simplicial complex is hole-free.

The importance of using the first combinatorial Laplacian of the Rips complex is not limited to the above theorem. Its very specific structure guarantees that (5.1) is effectively a local update rule; that is, the local state of an edge is updated using elements from edges that are lower adjacent to it. In fact, this update is very similar to the distributed, continuous-time consensus algorithms, a variant of which was studied in Chapter 2. In continuous-time consensus algorithms the graph Laplacian is used in the update equation \( \dot{x}(t) = -L_0x(t) \), in order to reach a consensus (which is a point in the kernel of \( L_0 \)) over a connected graph [85].

In summary, in order to verify coverage in a network of fixed sensors, it is sufficient to setup distributed linear dynamical system (5.1) for a random initial condition and observe the asymptotic state value as \( t \to \infty \). If this distributed dynamical system converges to zero, then the first Betti number of the Rips complex is zero, and therefore, the Rips shadow (which is the actual region covered by the sensors) is hole-free. Conversely, if the asymptotic value of (5.1) is non-zero, then the first homology of the Rips complex is non-trivial and therefore, Theorem 6 implies the existence of a non-contractible 1-cycle in the Rips shadow and hence, the presence of holes in the cover. Note that our assumptions on the fence nodes located on the boundary of \( \mathcal{D} \) is crucial in avoiding boundary effects. These fence nodes guarantee that if a coverage hole exists, it is located in the interior of the domain, \( \mathcal{D} \).\(^1\)

\(^1\)This assumption on strong degree of control along the boundary is not strictly required and can be relaxed. See [30] for more details.
Symmetric Coverage Framework

We now consider the symmetric coverage framework, in which each sensor is capable of covering a disk of radius $r_c$ and communicate with other sensors within distance $r_b \leq r_c \sqrt{3}$. In this case, the region covered by the sensors is the union of disks of radius $r_c$ centered at the location of the sensors: $\mathcal{U}(V) = \bigcup_{v_i \in V} \{ x \in \mathbb{R}^2 : \| x - v_i \| \leq r_c \}$. Similar to the previous framework, we need to define a combinatorial object that can capture the topological properties of set $\mathcal{U}(V)$. Given the fact that the covered region is simply a union of disks of radius $r_c$ (which are contractible sets), the Čech complex of $V$, defined in Chapter 4, is a natural candidate for capturing the topological properties of the sensor network. Recall that this complex is simply formed by associating a vertex to each disk, and then adding edges and other higher order simplices based on the overlap of the disks.

The following theorem, known as the Čech Theorem or the Nerve Lemma, indicates why the Čech complex captures the topological properties of the region covered by the sensor footprints [13, 14].

**Theorem 8 (The Čech Theorem).** *Given a finite collection of disks $U_v$ of radius $\epsilon$ and centered at points $v \in V$, the Čech complex $\mathcal{C}_\epsilon(V)$ has the homotopy type of the union of the disks in the collection, $\bigcup_{v \in V} U_v$. *

The above theorem implies that the Čech complex carries the same homological properties of the union of the sets. In particular, both objects have isomorphic homologies in all dimensions.\(^3\) Therefore, in order to verify coverage in a given domain by a set of sensors, one only needs to look at the homologies of the underlying Čech complex. If this simplicial complex has no holes, neither does the sensor cover. However, computation of this simplicial complex has no holes, neither does the sensor cover. However, computation of this simplicial complex has no holes, neither does the sensor cover. However, computation of this simplicial complex has no holes, neither does the sensor cover. However, computation of this simplicial complex has no holes, neither does the sensor cover. However, computation of this simplicial complex has no holes, neither does the sensor cover. However, computation of this simplicial complex has no holes, neither does the sensor cover. However, computation of this simplicial complex has no holes, neither does the sensor cover. However, computation of this simplicial complex has no holes, neither does the sensor cover. However, computation of this simplicial complex has no holes, neither does the sensor cover. However, computation of this simplicial complex has no holes, neither does the sensor cover. However, computation of this simplicial complex has no holes, neither does the sensor cover.

\(^2\)The statement of this theorem holds for any collection of contractible sets when all nonempty intersections of all subcollections are contractible.

\(^3\)Note that homotopy equivalence of two topological spaces is much stronger than having isomorphic fundamental groups, as was the case in Theorem 6. In other words, the Čech complex contains much more information about the union of disks than the Rips complex does about its shadow.
plicial complex and hence, its homologies is not an easy task, as it requires localization of each sensor as well as distance measurements in order to verify overlaps of footprints. Furthermore, as shown in [54], the Čech complex is very fragile with respect to uncertainties in distance and location measurements. In the absence of location information, an alternative would be to use the Rips complex instead, which can be formed uniquely from the communication graph of the network. Unfortunately, the Rips complex is not rich enough to contain all the topological and geometric information of the Čech complex and in general does not provide much information about coverage holes. Recently, it is shown that in certain cases, the Rips complex does in fact carry the necessary information to extract the homological properties of the cover. Namely, de Silva and Ghrist [30] show that a Rips complex with parameter $\epsilon$, $\mathcal{R}_\epsilon$, is a subcomplex of a Čech complex corresponding to disks of radius $\epsilon/\sqrt{3}$ centered at the location of vertices of the Rips complex. As a result, our assumption of $r_b \leq r_c \sqrt{3}$ leads to

$$\mathcal{R}_{r_b}(V) \subseteq \mathcal{C}_{r_c}(V),$$

which implies $\mathcal{A}(V) \subseteq \mathcal{U}(V)$, where $\mathcal{A}(V)$ is the shadow of the Rips complex with parameter $r_b$ and $\mathcal{U}(V)$ is the actual region covered by the sensors in the symmetric coverage framework with coverage radius $r_c$. Hence, if the Rips complex with broadcast disks of radius $r_b$ is hole-free, then so is the sensor coverage. This result would serve as a sufficient homological criterion for coverage verification. Note that the case of $r_b = \sqrt{3}r_c$ corresponds to the tightest such sufficient condition for planar networks.

In summary, in order to verify successful coverage in a distributed fashion, the sensors need to compute the first homology of the Rips complex $\mathcal{R}_{r_b}$ using local neighborhood information available to them. The triviality of the first homology of this simplicial complex provides a sufficient condition for a hole-free coverage of $\mathcal{D}$. Therefore, one can set up

91
linear dynamical system (5.1) corresponding to the Rips complex with parameter $r_b$ and observe its asymptotic behavior. Similar to the simplicial coverage framework, the asymptotic stability of this dynamical system is a sufficient condition for a hole-free coverage, although it is not necessary anymore.

As a last remark note that (5.1) is an edge-dimensional dynamical system, where each element of vector $x(t)$ corresponds to a 1-simplex. However, in both frameworks, edges and all other higher order simplices are simply combinatorial objects; the only real physical entities capable of computation are the sensors themselves. Therefore, in order to implement (5.1) in a sensor network, we need a protocol to assign the computation required by each edge to its adjacent nodes. One such algorithm is presented by Muhammad and Jadbabaie [80], who obtain a local representation of the Rips complex and implement the dynamical system in Theorem 7 at the node level. They also show that this local implementation at the node level can be achieved by using at most 2-hops of communications between neighboring vertices.

5.4 Distributed Hole Localization

In the previous section, we presented a coverage verification algorithm for a sensor network in which the nodes have no location or distance information. As noted before, this distributed algorithm is based on the close topological relationship between the actual cover and the Rips complex as its combinatorial representation. Unfortunately, this verification algorithm is not powerful enough to provide any further information on the cover. All it is capable of is verifying whether the coverage is successful (hole-free) or not. In most practical scenarios, however, one’s interest is not simply limited to coverage verification. In fact, we are as much interested in the location, number, and the size of the coverage holes (if they exist). Therefore, the algorithm of Section 5.3 needs to be followed by algorithms
that can reveal further information about the cover.

In this section, we present a distributed algorithm which is capable of “localizing” coverage holes in a sensor network with no location or metric information. By hole localization, we mean detecting cycles over the proximity graph of the network that encircle the coverage holes. The tightest of such cycles provides information on the location and the size of the hole in the Rips shadow.\(^4\) Similar to the previous algorithm, the results of this section are also based on the algebraic topological invariants, namely the homology, of the cover and the Rips complex of the network. In essence, in order to find the coverage holes, our algorithm computes the sparsest generator of a non-trivial class of homologous 1-cycles in the first homology of the simplicial complex, which corresponds to the shortest possible cycle around the holes. Our method is more general than the algorithms presented in [18, 47], where it is explicitly assumed that the simplicial complex is embedded on an orientable surface. It is also different from the results of [22] in the sense that it is not limited to Rips complexes, is distributed in nature, and does not use node coordinates.

Before presenting the algorithm, we state a few remarks regarding the relationship between the sparsest generator of the homology classes and the location of the holes. It is

\(^4\)Note that in the simplicial framework, the Rips shadow coincides with the actual cover, whereas in the symmetric framework it is only a subset of the region covered by the sensors.
important to keep in mind that we are using simplicial complexes which are combinatorial objects. Therefore, for hole localization in the absence of metric information, the best we can hope for is computing the shortest cycle encircling a hole, which is also a combinatorial object. For instance, consider two different sensor configurations and the region covered by them as depicted in Figure 5.1. Although the region covered is different, they are combinatorially equivalent as far as the Rips complex is concerned. Therefore, in both cases, any hole localization algorithm leads to the same result.

Another case that is worth mentioning is the case that the simplicial complex contains multiple holes. It is quite possible that in the case that two holes are “close” relative to their “sizes”,\(^5\) the sparsest generator of the homology class encircles both of them simultaneously, rather than each hole individually. Figure 5.2 is meant to clarify this case. In either case, the sparsest 1-cycle provides valuable information on the location and size of the holes.

With the above in mind, we present an algorithm which is capable of finding a short non-trivial cycle in a homology class. Intuitively, given a representative cycle of a non-trivial homology class, our algorithm computes a sparse generator of that homology class.

\(^5\)By terms such as close or big, we simply mean combinatorially close (in terms of hop count) and combinatorially big (in terms of the length of the shortest cycle).
in a distributed fashion, simply by removing components corresponding to cycles that are boundaries of 2-chains in the complex, and hence “tightening” the representative cycle around the holes. Therefore, in order to find a short cycle in a homology class, the algorithm needs an initial non-trivial 1-cycle in that class. Clearly, any non-zero point in $\ker L_1$ can potentially serve as such an initial 1-cycle. The immediate advantage of using $x \in \ker L_1$ is that one can easily compute such a point in a distributed manner as the limit of linear dynamical system (5.1). The following example clarifies the idea behind our algorithm.

**Example 5.** Consider the 2-dimensional simplicial complex depicted in Figure 4.3. As was shown in Example 4, the kernel of the first combinatorial Laplacian of this complex is one-dimensional. Therefore, the distributed linear dynamical system (5.1) converges to a non-zero vector in the span of $[8 \ 8 \ 1 \ 3 \ 8 \ 2 \ 5]^T$ for almost all initial conditions. Notice that all edges, including edges $[v_3v_4]$, $[v_3v_5]$, $[v_4v_5]$, and $[v_5v_6]$ that are not adjacent to the hole, have non-zero values asymptotically. In other words, no element of $\ker L_1$ is “tight” around the hole of the simplicial complex. Another key observation is that any $x \in \ker L_1$ can be written as a linear combination of three fundamental cycles in the 1-skeleton of the simplicial complex:

$$x = 8\alpha c_1 + 3\alpha c_2 + \alpha c_3$$

where

$$c_1 = \begin{bmatrix} 1 & 1 & 0 & 0 & 0 & 1 & 0 \end{bmatrix}^T$$

$$c_2 = \begin{bmatrix} 0 & 0 & 0 & 0 & 1 & 0 & -1 \end{bmatrix}^T$$

$$c_3 = \begin{bmatrix} 0 & 0 & 1 & 1 & 0 & 0 & -1 & 0 \end{bmatrix}^T,$$

and $\alpha$ is some real number. Among these cycles, only the first one corresponds to the

---

6To recap, the edges of the simplicial complex in Figure 4.3 are ordered as $[v_1v_2]$, $[v_2v_3]$, $[v_3v_4]$, $[v_4v_5]$, $[v_5v_6]$, $[v_6v_1]$, $[v_3v_5]$, and $[v_3v_6]$. 95
hole, while the other two are simply contractible cycles corresponding to boundaries of 2-simplices. Therefore, in order to find a tight cycle around the hole, one needs to subtract the right amount of null-homologous 1-cycles encircling 2-simplices (in this example $3\alpha$ and $\alpha$, respectively) from $x$. What remains is simply a 1-cycle with non-zero values only over the edges that are adjacent to the hole. Note that this cycle is also the sparsest generator of the non-trivial element of the first homology of the simplicial complex.

Computing the tightest cycle around the hole in the above example is simple, due to the fact that the simplicial complex only consists of very few simplices. Unfortunately, once the simplicial complex becomes large, it is not an easy task to compute the right amount of null-homologous cycles to subtract from an element in $\ker L_1$, and find a sparse representative of the homology class. Moreover, in the absence of a centralized scheme, it is reasonable to assume that elements of $x \in \ker L_1$ are only known locally to the nodes. In fact, this is the case if the kernel element is computed in a distributed fashion using dynamical system (5.1). Therefore, we need an algorithm which is capable of finding the sparsest non-trivial generator of the homology classes of a simplicial complex by using only local information.

### 5.4.1 Computing the Sparsest Generator: IP Formulation

Consider a simplicial complex $X$ with the first combinatorial Laplacian $L_1$. By construction, any element in the null space of $L_1$ is a 1-cycle that is orthogonal to the subspace spanned by the boundaries of the 2-simplices. In other words, $x \in \ker L_1 \subset \mathbb{R}^{n_1}$ implies $x \in \ker B_1$ and $x \perp \text{img} B_2$. Therefore, as stated in Chapter 4, any non-zero $x$ in the kernel of the first combinatorial Laplacian is a representative element of a non-trivial homology class of $X$. However, as in Example 5, $x$ is not necessarily the sparsest representative of the homology class it belongs to. In general, given a generator $x$ of a homology class, the
sparsest generator of that class can be computed as the solution to the following integer programming optimization problem:

\[
\begin{align*}
\text{Minimize} & \quad \|y\|_0 \\
\text{subject to} & \quad y = x + B_2z
\end{align*}
\] (5.2)

where \(\|\cdot\|_0\) is the \(\ell_0\)-norm of a vector, equal to the number of its non-zero elements, and \(B_2\) is the matrix representation of the second boundary operator \(\partial_2\). Note that if \(x\) is a 1-cycle, then the minimizer \(y^*\) is also a 1-cycle in the kernel of \(B_1\). Moreover, the constraint \(y - x \in \text{img } B_2\) guarantees that both \(x\) and \(y^*\) are representatives of the same homology class, simply because adding and subtracting null-homologous cycles does not change the homology class. Therefore, any solution of the above optimization problem is the sparsest generator of the homology class that \(x\) belongs to, and has the desired property that it is the tightest possible cycle (in terms of the length) around the holes represented by that homology class.

### 5.4.2 LP Relaxation

Optimization problem (5.2) has a very simple formulation. However, due to the 0-1 combinatorial element in the problem statement, solving it is not, in general, computationally tractable. In fact, Chen and Freedman [24] show that computing the sparsest generator of an arbitrary homology class is NP-hard.

A popular relaxation for solving such a problem is to minimize the \(\ell_1\)-norm of the objective function rather than its \(\ell_0\)-norm [36]:

\[
\begin{align*}
\text{Minimize} & \quad \|y\|_1 \\
\text{subject to} & \quad y = x + B_2z
\end{align*}
\] (5.3)
This relaxation\(^7\) is a linear programming (LP) problem and can be solved quite efficiently. An argument similar to before shows that the minimizer of the above optimization problem is also a 1-cycle homologous to the initial \(x\), as their difference is simply a null-homologous cycle in the image of \(B_2\).

In general, the minimizer of (5.3) is simply an approximation to the minimizer of (5.2) and has a larger \(\ell_0\)-norm. However, in certain cases the solutions of the two problems coincide. In the next theorem, we present conditions under which the two minimizers have the same zero/non-zero pattern. Under such conditions, we would be able to compute the sparsest generator of the homology class of \(x\) efficiently.

**Exact Relaxation**

Before formally presenting the theorem, we need to define some notation and present a lemma. Consider an oriented Rips complex \(\mathcal{R}\) with first Betti number \(b\), where the holes are labeled 1 through \(b\). By \(h(\alpha_1, \ldots, \alpha_b)\) we denote the class of homologous 1-cycles that encircle the \(i\)-th hole \(\alpha_i\) many times in a given direction. Assume that the shortest representative cycle that encircles the \(i\)-th hole is unique for all \(i\), and is denoted by \(c_i^*\). In other words,

\[
c_i^* = \arg \min \|c\|_0 \\
\text{s.t. } c \in h(e_i)
\]

where \(e_i\) is the \(i\)-th coordinate vector. Since \(c_i^*\) is the sparsest 1-cycle that encircles the \(i\)-th hole only once, we have the following lemma.

**Lemma 6.** \(c_i^*\) is a 1-cycle which only has value in \(\{0, 1, -1\}\).

\(^7\)Strictly speaking, (5.3) is not a relaxation of (5.2), as the feasibility sets of the two problems coincide. However, one can show, [59], that there exists an LP equivalent to (5.3) which is a relaxation of an IP equivalent to (5.2).
We now present the main theorem of this section that characterizes sufficient conditions for the exactness of the relaxation problem.

**Theorem 9.** Given a Rips complex $\mathcal{R}$, suppose that the shortest representative cycle that encircles the $i$-th hole, denoted by $c_i^*$, is unique for all $i$. Also assume that any simple loop $c \in h(\mu)$ satisfies $\|c\|_0 \geq \sum_{i=1}^b |\mu_i| \|c_i^*\|_0$ for all $\mu \in \mathbb{Z}^b$. Then, for all $\alpha \in \mathbb{R}^b$ we have,

$$\arg \min_{c \in h(\alpha)} \|c\|_0 = \arg \min_{c \in h(\alpha)} \|c\|_1.$$  

**Proof:** First we prove that the two minimizers have the same zero/non-zero pattern. Given a class $h(\alpha)$, suppose that the $\ell_1$-minimizer, denoted by $y$, does not have the same pattern as the $\ell_0$-minimizer. This means that there exists an edge $\sigma_1$ in the simplicial complex such that $y$ has a positive value on, but the $\ell_0$-minimizer does not. Since $y$ is a 1-cycle, there exists another edge $\sigma_2$ lower-adjacent to $\sigma_1$ with a non-zero value. Without loss of generality, we assume that the directions are defined such that all values are positive. Reapplying the same argument implies that $\sigma_1$ belongs to a set $E$ of edges, all with positive values and forming a simple loop over the simplicial complex. Moreover, it implies that $\tilde{c}_j = \mathbb{I}_{\{\sigma_j \in E\}}$ is a 1-cycle. Note that $\tilde{c}$ is a simple 1-cycle which only takes values in $\{0, 1, -1\}$. Finally, define $\gamma > 0$ to be the smallest value that the edges in $E$ take in the $\ell_1$-minimizer $y$.

The 1-cycle $\tilde{c}$ belongs to some homology class $h(\mu)$, that is, the class of 1-cycles that encircle the $i$-th hole $\mu_i$ many times. Without loss of generality, we can assume that $\mu_i \geq 0$
for all $0 \leq i \leq b$. Define $y' = y - \gamma \tilde{c} + \gamma (\sum_{i=1}^{b} \mu_i c^*_i)$ for which we have,

$$
\|y'\|_1 \leq \|y - \gamma \tilde{c}\|_1 + \gamma \sum_{i=1}^{b} \mu_i \|c^*_i\|_1
$$

$$
= \|y\|_1 - \gamma \|\tilde{c}\|_1 + \gamma \sum_{i=1}^{b} \mu_i \|c^*_i\|_1
$$

$$
= \|y\|_1 - \gamma \|\tilde{c}\|_0 + \gamma \sum_{i=1}^{b} \mu_i \|c^*_i\|_0 < \|y\|_1,
$$

The first inequality is a consequence of the triangular inequality. The following equality is due to the fact that we assumed that $\gamma$ is the smallest value on the edges of $\tilde{c}$ at $y$. In the next equality, we use that fact that $\tilde{c}$ and all $c^*_i$ are 1-cycles with values in $\{0, 1, -1\}$, which means that their $\ell_1$ and $\ell_0$-norms are equal. Finally, the last inequality is due to the assumption of the theorem.

In summary, there exists a 1-cycle $y'$ homologous to $y$ with a smaller $\ell_1$-norm, which contradicts the fact that $y$ is the $\ell_1$-minimizer. Therefore, the two 1-cycles \[ \arg \min_{c \in h(\alpha)} \|c\|_0 \] and \[ \arg \min_{c \in h(\alpha)} \|c\|_1 \] have the same zero/non-zero pattern for all $\alpha$. Also note that both minimizers belong to the same homology class $h(\alpha)$. As a result, the two must be equal.

**A Few Remarks**

The above theorem states that, under the given conditions, the $\ell_1$-minimizer is the sparsest generator of its homology class as well, and therefore, its non-zero entries indicate the edges of the 1-cycle that are tight around the holes. As a consequence, one can compute this sparse generator efficiently, using methods known for solving LPs.

It is important to notice that Theorem 9 requires the uniqueness of the sparsest generator of each homology class in order to guarantee that the minimizers of the two problems coincide. When the $\ell_0$-minimizer is not unique, not only every $\ell_0$-minimizer is a solution to (5.3), but so is any convex combination of those minimizers. Note that if two vectors
have the same $\ell_1$-norm, then any vector in their convex hull cannot have a larger $\ell_1$-norm. In such cases, solving (5.3) results in a 1-cycle in the convex hull of the minimizers of (5.2).

The intuition behind the condition of Theorem 9 is also worth exploring. One very important case for which the condition holds, is the case that the simplicial complex has only one hole. Another is the case that the holes in the simplicial complex are far from each other relative to their sizes. In either case, the shortest representative cycle of any homology class is simply a linear combination of the shortest cycles encircling the holes separately. Note that even when the condition does not hold, the solution of (5.3) is a relatively sparse (although not necessarily the sparsest) 1-cycle, and therefore, can be used as a good approximation for hole localization.

### 5.4.3 Decentralized Computation: The Subgradient Method

As mentioned before, unlike the original IP problem (5.2), one can convert (5.3) to a linear programming problem and solve it efficiently using methods known for solving LPs. However, applying the subgradient method [16, 83] enables us to compute the $\ell_1$-minimizer in a distributed manner. Even though convergence would be slower than usual methods for solving linear programs, the added value of decentralization makes the method worthwhile.

The subgradient method is a simple algorithm for minimizing non-differentiable convex functions. Given such a function $f : \mathbb{R}^n \to \mathbb{R}$, the subgradient method minimizes $f(z)$ by simply using the iteration

$$z^{(k+1)} = z^{(k)} - \alpha_k g^{(k)}$$

where $z^{(k)}$ is the $k$-th iterate, $g^{(k)}$ is any subgradient of $f$ at $z^{(k)}$ and $\alpha_k > 0$ is the $k$-th step size [16]. Recall that a subgradient of $f$ at $z$ is any vector $g$ that satisfies the inequality $f(z') \geq f(z) + g^T(z' - z)$ for all $z'$. Since the above update is not a descent method, one needs to keep track of the best point found so far $f_{\text{best}}^{(k)} = \min\{f_{\text{best}}^{(k-1)}, f(z^{(k)})\}$. For
the proper choice of the step sizes and large enough $k$, $f_{best}^{(k)}$ gets arbitrarily close to the minimum value of the objective function.

Now consider the optimization problem (5.3), which can be rewritten as

$$
\begin{align*}
\text{Minimize} & \quad \|x + B_2 z\|_1 \\
\text{subject to} & \quad z \in \mathbb{R}^{n_2}
\end{align*}
$$

where $n_2$ is the number of the 2-simplices of the simplicial complex. A subgradient for the objective function in the above problem is the sign function. Therefore, the subgradient update can be written as

$$
z^{(k+1)} = z^{(k)} - \alpha_k B_2^T \text{sgn}(B_2 z^{(k)} + x)
$$

with the initial condition $z^{(0)} = 0$. Note that $z$ is a face-dimensional vector and the iteration updates an evaluation on the 2-simplices of the simplicial complex. The most important characteristic of (5.5) is that, due to the local structure of $B_2$, the subgradient update is implementable in a distributed manner, if the initial $x$ is known locally. By picking a small enough constant step size $\alpha_k$, (5.5) is guaranteed to get arbitrarily close to the solution of 5.4 [16]. By choosing more sophisticated dynamic step sizes we can improve the convergence properties of the above algorithm to the optimal solution, which is a sparse generator of the homology class that $x$ belongs to. We provide non-trivial simulations of this algorithms in Section 5.6.

### 5.5 Distributed Detection of Redundant Sensors

In the previous sections we presented a homological criterion for coverage. Namely, based on the results of de Silva, Ghrist, and Muhammad [32], we argued that a sufficient condition
for successful coverage is to have no holes in the flag complex of the proximity graph, i.e., the Rips complex of the network. This condition is translated into algebraic topological terms as $H_1(\mathcal{R}_{r_b}) = 0$, which means that every 1-cycle in the communication graph can be realized as the boundary of a surface built from the 2-simplices of $\mathcal{R}_{r_b}$. Furthermore, through Theorem 5, we showed that the first combinatorial Laplacian can be used to verify this homological criterion in a distributed manner.

In this section, we present a distributed algorithm which is capable of computing a sparse cover of domain $\mathcal{D}$ and detect redundancies in the sensor network, in the absence of location information. In other words, the algorithm enables us to “turn off” redundant sensors without impinging upon the coverage integrity. As in section 5.4, we formulate the problem of finding a sparse cover as an optimization problem to compute the sparsest generator of a certain homology class, and use subgradient methods to solve it in a distributed way. However, unlike previous sections, we use the second homology of the Rips complex relative to its boundary, rather than its first homology. The advantage of the second relative homology lies in the fact that it is more robust to extensions and therefore, yields stronger information about the actual cover [30].

Consider the Rips complex $\mathcal{R}$ corresponding to network of the sensors deployed over region $\mathcal{D}$ and consider $\mathcal{F} \subset \mathcal{R}$ to be the subcomplex that is canonically identified with the fence nodes over $\partial \mathcal{D}$. If this cycle is null-homologous - that is, if $[\mathcal{F}] = 0$ in $H_1(\mathcal{R})$ - then, the coverage is hole-free. In such a case, there exists a 2-chain which bounds $\mathcal{F}$:

$$\forall \, \text{1-cycle} \, \beta \in C_1(\mathcal{F}), \quad \exists \alpha \in C_2(\mathcal{R}) \, \text{ s.t. } \beta = \partial_2 \alpha$$

Therefore, when translated into the language of algebraic topology, such a 2-chain $\alpha$, which is not necessarily unique, represents a relative 2-dimensional homology class, a certain generator in $H_2(\mathcal{R}, \mathcal{F})$. As a result, the condition for a hole-free successful coverage can be
Figure 5.3: If the first homology of $\mathcal{R}$ is non-trivial, then the second relative homology $H_2(\mathcal{R}, \mathcal{F})$ has no generator with values on the boundary. Conversely, if the second homology relative to the boundary has a non-trivial generator with a non-vanishing boundary, then $H_1(\mathcal{R}) = 0$.

rewritten in terms of the second relative homology classes:

**Theorem 10.** For a set of nodes $V$ in a domain $\mathcal{D} \subset \mathbb{R}^2$ satisfying the assumptions of Section 5.2, the sensor cover contains $\mathcal{D}$ if there exists $[\alpha] \in H_2(\mathcal{R}, \mathcal{F})$ such that $\partial_2 \alpha \neq 0$.

This theorem is first stated and proved by de Silva and Ghrist [30]. Intuitively, the 2-chain $\alpha$ has the appearance of “filling in” $\mathcal{D}$ with triangles composed of projected 2-simplices from $\mathcal{R}$. Note that the relative group $H_2(\mathcal{R}, \mathcal{F})$ captures the second homology of the quotient space $\mathcal{R}/\mathcal{F}$, in which all the simplices in $\mathcal{F}$ are identified. This can be done by adding a “super node” to the complex, as depicted in Figure 5.3. If the Rips complex is hole-free, then the topology of this quotient space is that of a sphere, and therefore, the relative homology $H_2(\mathcal{R}, \mathcal{F})$ has a non-trivial generator. On the other hand, if the 1-cycles defined over subcomplex $\mathcal{F}$ are not contractible, then the second relative homology has no generator with non-zero values on the boundary [30].

Note that the dimension of the second relative homology $H_2(\mathcal{R}, \mathcal{F})$ may be greater than one. This can happen if there exists a 2-cycle which is a generator of $H_2(\mathcal{R})$ as well as $H_2(\mathcal{R}, \mathcal{F})$, as depicted in Figure 5.4. Such 2-cycles do not represent a true relative class, as they may still exist, even if the fence cycle $\mathcal{F}$ is not the boundary of any 2-chain. Hence, Theorem 10 requires the existence of a relative 2-cycle $\alpha$ with a non-zero boundary.

Given the above, it is easy to see that the minimal cover is simply the sparsest gen-
erator of a second homology class of $\mathcal{R}$ relative to $\mathcal{F}$. Therefore, one can formulate the problem of finding the sparsest cover over $\mathcal{D}$ as an optimization problem, simply by extending the results of the previous section to a higher dimension. The only difference lies in the fact that, instead of the Rips complex corresponding to the network, one needs to use the quotient complex $\overline{\mathcal{R}} = \mathcal{R}/\mathcal{F}$ which is obtained by identifying all the simplices of $\mathcal{F}$ with a supper node. Once this quotient simplicial complex is formed, we compute its second combinatorial Laplacian in a distributed manner, and by running the decentralized linear dynamical system $\dot{x}(t) = -L_2 x(t)$, with a random initial condition, obtain a point $x \in \ker L_2$ asymptotically. The limit of this dynamical update is a relative 2-cycle which does not vanish on the boundary, for almost all initial conditions. Once such a 2-cycle $x$ is computed, the minimizer of the optimization problem

$$\begin{align*}
\text{Minimize} & \quad \|y\|_0 \\
\text{subject to} & \quad y = x + B_3 z
\end{align*} \quad (5.6)$$

represents the sparsest generator of the relative homology class that $x$ belong to. In the above problem, $B_3$ is the triangle-by-tetrahedron incidence matrix of the quotient complex $\mathcal{R}/\mathcal{F}$, $x$ and $y$ are 2-cycles and $z$ is a 3-chain. Similar to problem (5.2), the constraint $y - x \in \text{img } B_3$ guarantees that $y$ and $x$ are homologous 2-cycles. Since the above problem is again NP-hard and cannot be solved efficiently, one can instead solve its $\ell_1$-relaxation:

$$\text{Minimize} \quad \|x + B_3 z\|_1 \quad (5.7)$$

which can also be solved by the means of the distributed subgradient update

$$z^{(k+1)} = z^{(k)} - \alpha_k B_3^T \text{sgn}(B_3 z^{(k)} + x). \quad (5.8)$$

\footnote{Note that this object can be formed in a distributed fashion. All that is required is that the fence nodes take the local neighborhood relations of each other into account and update their values together.}
Figure 5.4: The eight faces of the octahedron form a non-trivial 2-cycle $\alpha$ such that $[\alpha] \in H_2(R)$. However, $\alpha$ has a vanishing boundary $\partial_2 \alpha = 0$, and therefore, does not correspond to a true relative 2-cycle.

in a distributed manner.

The above distributed iteration leads to a sparse generator of the second relative homology, in which most 2-simplices have a corresponding value equal to zero. In the optimal solution, any vertex that only belongs to 2-simplices with zero valuations can be removed from the network, without generating a coverage hole. The next section contains simulations that demonstrate the performance of the presented algorithm.

5.6 Simulations

This section contains non-trivial simulations of the algorithms we presented in Sections 5.4 and 5.5, for hole localization and computation of the minimal cover, respectively.

5.6.1 Hole Localization

We demonstrate the performance of our distributed hole localization algorithm with a randomly generated numerical example. Figure 5.5(a) depicts the Rips shadow of a simplicial complex on $n = 81$ vertices distributed over $\mathbb{R}^2$. The 2-skeleton of this simplicial complex consists of 81 vertices, 372 edges, and 66 triangles (2-simplices). As Figure 5.5(a) suggests, the null space of the first combinatorial Laplacian of this Rips complex has di-
mension 2. The two non-trivial homology classes correspond to two zero eigenvectors of the Laplacian matrix. We generated a point in $x \in \ker L_1$ by running the distributed linear dynamical system (5.1) with a random initial condition $x(0)$. The edge-evaluation of the limiting $x \in \ker L_1$ is depicted in Figure 5.5(b), where the thickness of an edge is directly proportional to the magnitude of its corresponding component in $x$. It can be seen that all components of the generated 1-cycle in the null space of $L_1$ are more or less of the same order of magnitude. In order to localize the two holes, we ran subgradient update (5.5) with a diminishing, square summable, but not summable step size. The edge evaluation of the 1-cycles after 1000 and 4000 iterations are depicted in Figures 5.5(c) and 5.5(d). These figures illustrate that after enough iterations, the subgradient method converges to a 1-cycle that has non-zero values only over the cycles that are tight around the holes. Therefore, the algorithm is capable of localizing the coverage holes. In Figure 5.5(d), the valuations of the 12 edges adjacent to the holes are 3 orders of magnitude higher than the rest.

Note that our algorithm is only capable of finding tight minimal-length cycles surrounding the holes, which do not necessarily coincide with the cycles that are closer in distance to the holes. As stated before, after all, we are not using any metric information and the combinatorial relations between vertices is the only information available. Moreover, in the case that there are two minimal-length cycles surrounding the same hole (as in the upper hole in Figure 5.5), any convex combination of those is also a minimizer of the LP relaxation problem (5.3). In such cases, the subgradient method in general converges to a point in the convex hull of the solutions, rather than a corner solution. Also note that the holes in the Rips complex are far relative to their sizes and therefore, Theorem 9 guarantees that the solution obtained by the $\ell_1$-minimization lies in the convex hull of the $\ell_0$-minimizers.
Figure 5.5: Subgradient methods can be used to localize the holes in a distributed fashion.

5.6.2 Computing a Sparse Cover

Figure 5.6 illustrates the performance of the algorithm we presented in section 5.5. The randomly generated Rips complex used for this simulation is made up of 66 vertices, 22 of which function as fence nodes (Figure 5.6(a)). The second relative homology of this simplicial complex consists of only one non-trivial class of relative 2-cycles. In order to compute a non-trivial representative of the second relative homology, we introduced an extra node, connected to all the fence nodes. We computed the second combinatorial Laplacian of the resulting complex and used the linear update \( \dot{x}(t) = -L_2 x(t) \) to obtain a point in the null space of \( L_2 \). Subgradient update (5.8) is used to solve the optimization problem (5.6). The minimizer 2-cycle is depicted in Figure 5.6(b). Any vertex which does not belong to a 2-simplex with a non-zero evaluation at the optimal can be removed,
Figure 5.6: Finding the minimal generator of the second relative homology $H_2(R, F)$ leads to a minimal cover. 32 of the sensors can be turned off without generating any coverage holes.

without impinging upon the coverage integrity. As illustrated in Figure 5.6(b), 32 of the vertices can be removed, while the Rips shadow remains hole-free.

As a last remark, note that either one of the vertices $a$ or $b$ in Figure 5.6(b) can also be removed from the network without generating a coverage hole. In fact, the removal of either one, would lead to an even sparser solution than the one obtained by the subgradient update. This is due to the fact that the generator depicted in Figure 5.6(b) is in fact in the convex hull of two distinct solutions of the original integer programming problem (5.6). As in the earlier example, since the original problem has more than one minimizer, any convex combination of them is also a minimizer of the LP relaxation problem (5.7).
6.1 Thesis Summary

In this thesis, we studied two different problems in the field of complex networks. The first part of the thesis was focused on the question of information aggregation over networks, whereas in the second part, we suggested new paradigms for modeling and analysis of networks, namely simplicial complexes and the machinery of simplicial homology theory. We also provided a specific application to highlight the benefits of such tools in analysis of complex networks. In both parts, we focused on algorithms that can be implemented in a distributed fashion, in the absence of a centralized computational entity or centralized coordination among agents. We will summarize the presented results in the following.

The first part of the thesis was presented in Chapters 2 and 3. Chapter 2 was dedicated to the study of discrete-time distributed consensus algorithms over networks. We first presented a brief review on the known results regarding these algorithms, followed by the study of their behavior over randomly switching networks. We proved that as long as the graphs representing the random changes in the network are independent and identically distributed, the benchmark consensus algorithm converges if and only if the network is con-
nected in expectation. We further generalized this result to ergodic and stationary random network processes. We concluded the chapter by computing the mean and variance of the consensus value, in terms of the statistical properties of the graph process.

Even though useful in many applications, distributed consensus algorithms, studied in Chapter 2, only result in an agreement over the network, and do not necessarily guarantee agreement on the correct parameter, what is known as social learning. In Chapter 3, we presented a non-Bayesian belief update rule which under very mild conditions on the network and signal structures guarantees complete learning by all agents. The presented belief update rule, unlike rational learning, has the advantage of being local and computationally tractable.

Chapters 4 and 5 formed the second part of this thesis. In Chapter 4, we presented a brief review on the concept of simplicial complex, a generalization of graphs to higher order relations between entities, as a more faithful model of the network. The section also contained the basics of simplicial homology theory, which studies how $k$-simplices of a simplicial complex are attached to its $(k-1)$-simplices [61]. This review was followed by Chapter 5, in which we highlighted the benefits of modeling networks with simplicial complexes, instead of graphs, in the context of a specific application: the problem of coverage verification in coordinate-free sensor networks. We showed that homologies of different dimensions can provide valuable information on the properties of the cover in a network of sensors with access to no location, distance, or orientation information. The chapter also contained three real-time, scalable, and decentralized schemes for coverage verification, hole detection, and computation of a minimal set of sensors required to monitor the region of interest.
6.2 Future Research Directions

Random Consensus in Dynamic Networks

As mentioned in Chapter 2, characterizing the distribution of the final agreement value in distributed consensus algorithms over general i.i.d. random networks is still an open problem. Nevertheless, we showed that it is possible to compute the mean and variance of the random consensus value. A next step would be to further characterize the asymptotic consensus distribution by computing its higher order moments.

A different direction to pursue on this topic is to study the behavior of random consensus algorithms in large dynamic networks. The idea would be to use the asymptotic properties of very large graphs to characterize the distribution of the random variable that all agents agree upon. One natural and simple candidate is to study the distribution of the consensus value for an i.i.d. sequence of large Erdős-Rényi random graphs, where agents communicate with one another independently with some pre-specified probability.

Non-Bayesian Learning in Social Networks

In Chapter 3, we proved convergence results for a non-Bayesian learning rule over social networks. Our main results highlighted that regardless of the topology of the network and the influence level of different individuals, as long as the social network is strongly connected, all agents will asymptotically learn the true underlying state of the world. Nonetheless, the topology of the network, the influence level of different agents, and their signal structures determine the speed at which social learning is achieved. Therefore, a next step would be to compute the learning rate as a function of the fundamentals of the environment. Such a study leads to a quantitative characterization of the roles played by the graph topology and the influence level of different agents in the learning process.

One fundamental assumption of the social learning model we studied in Chapter 3
was the observability of the beliefs. We assumed that each agent has access to the beliefs held by her neighbors. However, in many realistic scenarios, agents can only observe the actions, and not the beliefs, of others. In such cases, the challenge would be the design of decentralized belief update rules for social learning, where each agent updates her beliefs based on her private observations and the actions of her neighbors. This is a particularly interesting and challenging problem, given the fact that information contained in one’s actions is often much narrower than information one can extract from her beliefs.

Another potential future research direction is to investigate the behavior of our non-Bayesian learning rule in dynamic social networks. In many applications, it is plausible to assume that agents communicate with potentially distinct sets of agents over time. They might also assign different levels of confidence to their neighbors’ beliefs at different time periods. Understanding the effect of such dynamic changes in network topology and confidence levels provides a path for future research. A further challenge in this area would be to investigate conditions for asymptotic social learning under endogenous changes in network topology and confidence levels; that is, when agents endogenously decide whom to communicate with based on their beliefs, rather than some exogenously specified process dictating the evolution of the network.

**Distributed Computation of Homologies**

In Chapter 5, we showed how the problem of computing the sparsest generator of a given homology class can be formulated as an integer programming (IP) problem. Furthermore, we provided a condition to guarantee the exactness of the linear programming (LP) relaxation of the problem; that is, a sufficient conditions for the solutions of the two problems to coincide. One aspect that was not addressed in this thesis is the complete characterization of a necessary and sufficient condition for the exactness of the relaxation. Such a necessary and sufficient condition would characterize cases that one can efficiently compute the
sparsest generator of a homology class, using linear programming. Exploring the similarity between our problem and the $\ell_0$ and $\ell_1$-minimization problems in the compress sensing is yet another direction one can take for future research.

Another aspect that we did not address is the performance of the LP relaxation of the sparse generator computation problem (which is an IP), when the relaxation is not exact. In particular, how well does the linear programming relaxation of the original problem perform when the solutions of the two problems do not coincide? Answering such a question is especially important in the context of computing a sparse cover for the sensor network, discussed in Section 5.5.
Appendix A

Appendix: Graph Theory

A.1 Undirected Graphs

A (finite) undirected graph $G$ consists of a finite vertex set $V$ and an edge set $E$ where an edge is an unordered pair of vertices in $V$. If $i, j \in V$ and $(i, j) \in E$, then $i$ and $j$ are said to be adjacent or neighbors. This relationship is denoted by $i \sim j$. We use $\mathcal{N}_i = \{ j \in V | (i, j) \in E \}$ to denote the set of neighbors of vertex $i$. The size of this set which represents the number of vertex $i$’s neighbors is called the degree of vertex $i$. A path in $G = (V, E)$ from vertex $i$ to vertex $j$ is a sequence of vertices starting with $i$ and ending with $j$ such that each vertex is a neighbor of the next vertex in the sequence. An undirected graph is said to be connected if there exists a path from every vertex in $V$ to any other vertex. The set of vertices $j \in V$ for which there exists a path from vertex $i$ to is called the connected component that $i$ belongs to.

The adjacency matrix $A(G) = [a_{ij}]$ of an undirected graph $G$ is a symmetric matrix with rows and columns indexed by the vertices of the graph, such that $a_{ij} = 1$ if vertices $i$ and $j$ are neighbors, and $a_{ij} = 0$ otherwise. The degree matrix of graph $G$ denoted by $D(G)$ is a diagonal matrix with rows and columns indexed by the vertices of $G$, in which
its \((i, i)\) entry is the degree of vertex \(i\).

Given the definitions of the adjacency and degree matrices of a graph, one can define the matrix

\[
L(G) = D(G) - A(G)
\]

known as the Laplacian matrix of graph \(G\). It is straightforward to verify that the Laplacian is a symmetric and singular matrix. The importance of Laplacian \(L(G)\) is that it captures many topological properties of its corresponding graph \((G)\). The algebraic multiplicity of the zero eigenvalues of this positive semi-definite M-matrix is equal to the number of the connected components of \(G\).

The incidence matrix of the graph \(G\) is a \(|V| \times |E|\) matrix \(B(G) = [b_{ij}]\), where \(|V|\) and \(|E|\) are the numbers of vertices and edges of \(G\), respectively. The rows and columns of \(B\) are indexed by the vertices and the columns of the graph, respectively. The elements of the incidence matrix are such that \(b_{ij} = 1\) if the vertex \(i\) and edge \(j\) are incident and \(0\) otherwise. It is also well-known that the Laplacian of an undirected graph can be written in terms of its incidence matrix as well:

\[
L(G) = B(G)B(G)^T
\]

where the superscript \(T\) denotes the transpose of the matrix.

**A.2 Directed Graphs**

Similar to undirected graphs, a directed graph consists of a vertex set \(V\) and an edge set \(E\). However, unlike undirected graphs, every element of \(E\) is an ordered pair of elements of \(V\), called a directed edge. The set of neighbors of vertex \(i\) in \(G\) is denoted by \(\mathcal{N}_i = \{j \in V : (j, i) \in E\}\). A subgraph \(G_s\) of a directed graph \(G\) is a directed graph such that its vertex
set $V_s \subseteq V$ and the edge set $E_s \subseteq E$. If $V_s = V$, we call $G_s$ a spanning subgraph of $G$.

A (directed) path in a directed graph $G$ is a finite sequence of edges $(v_1, v_2), (v_2, v_3), \ldots, (v_{k-1}, v_k)$. A directed tree is a directed graph, where every vertex, except one vertex which is called the root, has exactly one neighbor, and the root vertex has no neighbors and can be connected to any other vertex through paths. A spanning tree of $G$ is a directed tree that is a spanning subgraph of $G$. A graph is said to contain a directed spanning tree if a subset of the edges forms a spanning tree. Such a graph is also called strongly rooted. A directed graph $G$ is strongly connected, if between every pair of distinct vertices $i, j$ in $V$, there is a path that begins at $i$ and ends at $j$.

The adjacency matrix $A(G)$ of a directed graph $G$ is a square matrix with rows and columns indexed by the vertices of the graph, such that $a_{ij} = 1$ if there exists a directed edge connecting vertex $j$ to vertex $i$, and $a_{ij} = 0$, otherwise.

For any given $n \times n$ non-negative matrix $W$, one one can define a corresponding graph denoted by $G(W)$ on $n$ vertices, which correspond to the rows and columns of $W$, and an edge set $E$, such that $(i, j) \in E$ if and only if $W_{ji} \neq 0$. In this case, we say vertex $j$ has access to vertex $i$. We say vertices $i$ and $j$ communicate if both $(i, j)$ and $(j, i)$ are edges of $G(W)$. The communication relation is an equivalence relation and defines equivalence classes on the set of vertices. If no vertex in a specific communication class has access to any vertex outside that class, such a class is called initial. For a given stochastic matrix $W$ and its corresponding graph $G(W)$, we have the following lemma, the proof of which can be found in [10].

**Lemma 7.** Suppose that $W$ is a stochastic matrix for which its corresponding graph has $s$ communication classes named $\alpha_1, \cdots, \alpha_s$. Class $\alpha_r$ is initial, if and only if the spectral radius of $\alpha_r[W]$ equals to one, where $\alpha_r[W]$ is the submatrix of $W$ corresponding to the vertices in the class $\alpha_r$. 

117
Bibliography


129


