Characterizing Diverse Link Patterns in Complex Networks: Theory and Applications

A DISSERTATION
SUBMITTED TO THE FACULTY OF THE GRADUATE SCHOOL
OF THE UNIVERSITY OF MINNESOTA
BY

Yanhua Li

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR THE DEGREE OF
Doctor of Philosophy

Prof. Zhi-Li Zhang

May, 2013

Acknowledgements

Completing my PhD degree is the most exciting and challenging activity during the first 32 years of my life, which would not have been possible without the help of many people. First and foremost, I would like to thank my adviser, Zhi-Li Zhang, for his great help and support. Zhi-Li guided me to *know*, *understand*, and *get interested in* doing research. I would like to thank Zhi-Li for so many hours of discussion, insightful feedback, comments, and guidance on my work. The six years (2007-2013) experience of research and life in Minnesota is the most invaluable treasure for my academic life.

I would like to thank professors Daniel Boley, Volkan Isler, and Shuzhong Zhang for serving on my thesis committee, and Daniel Boley for chairing the committee. Their feedback and guidance have been absolutely invaluable.

I am grateful to Professor Zhi-Li Zhang and Daniel Boley for contributing to my work reported in this dissertation and for their recommendations and help. I would like also to thank Professor John Carlis for guiding me on how to conduct research, presentation skills, and technical writing.

I would like to thank my current and fellow graduate students at the University of Minnesota (in alphabetic order), Vijay Kumar Adhikari, Jie Bao, Yingying Chen, Jin Cheng, Qiang Fu, Yu Gu, Shuo Guo, Ranjan Gyan, Jaehoon Jeong, Zhe Jiang, Dong Jiao, Pengkui Luo, Guanlin Lv, Hesham Z. Mekky, Dongchul Park, Esam Sharafuddin, Ziguo Zhong, Jia Zhou, Xun Zhou, Ting Zhu.

Moreover, I would like to acknowledge those industry partners that supported this work, Dr. Wei Chen and Yajun Wang (Microsoft Research), Moritz Steiner, Limin Wang, and T.V. Lakshman (Bell Labs), and Haiyong Xie (HUAWEI Research).

The work reported in this dissertation was supported in part by US National Science Foundation grants CNS-0626808, 0831734, CNS-0905037, CNS-1017092, CNS-1017647, the DTRA

grant HDTRA1-09-1-0050, a DoD ARO MURI Award W911NF-12-1-0385, and a University of Minnesota DTC DTI grant.

Dedication

To *my paraents*, *my wife*, *Xurong*, and *my sons*, *Aaron* and *Lucas*.

Abstract

Complex networks, including the Internet, wireless and cellular networks, and on-line social networks, are becoming indispensable parts of our daily lives. These networks arising from a wide range of applications can be represented and studied as graphs, and the underlying link patterns play an important role in understanding and solving problems in such applications. For example, due to the unreliable and asymmetric wireless channel, ad hoc wireless networks can be viewed as directed graphs, and the link directions contain crucial information about the possibility and efficiency of routing over such networks. Moreover, many online social networks, such as Twitter and Google+, can be viewed as directed graphs with uni-directional "following" relations among users, and the link directions contain crucial information about how users form social communities. In another application, online social networks such as Slashdot and Epinions represent relationships between users as links with positive or negative weights, which correspond to friend and foe relations. These networks are referred to as signed networks, where those signed links generate new challenges in understanding and studying the underlying network properties.

In this dissertation, I present my work on developing theories for studying and characterizing various crucial properties and application challenges in undirected, directed, and signed complex networks. First of all, we develop and extend random walk theory and the intrinsically related spectral graph theory for undirected graphs to directed graphs. Then, we explore application challenges raised by various link patterns. To be precise, we introduce a novel social community detection algorithm for social networks with both uni- and bi-directional links. In undirected communication networks, we establish the routing continuum theory that spans from short path routing to "potential" based all-path routing, based on the connection between routing and network flow optimization problems. Moreover, we investigate the social influence diffusion dynamics and influence maximization in social networks with both friend and foe relations.

Contents

Acknowledgements i

Dedication iii

Abstract iv

List of Tables ix

List of Figures x

1 Introduction 1

1.1 Directed graph model on wireless and social networks. 2

1.2 Undirected graph model: Towards traffic intelligence on complex networks. . . 2

1.3 Signed graph model: Influence diffusion dynamics and influence maximization on social networks. 3

1.4 Roadmap 3

2 Digraph Laplacian and the Degree of Asymmetry 4

2.1 Introduction 4

2.2 Preliminaries: Random Walks on Undirected Graphs 6

2.3 Random Walk Theory on Digraphs 7

2.3.1 Random Walks on Directed Graphs and Fundamental Matrix 7

2.3.2 Diplacian and Green's Function for Digraphs 9

2.3.3 Computing Hitting and Commute Times for Digraphs using Diplacian . 12

2.4 Degree of Asymmetry, Generalized Cheeger Constant and Bounds on Mixing Rate . . . 13
2.4.1 The Degree of Asymmetry, and Relations to Symmetrized Laplacian . . 13
2.4.2 Bounding the Mixing Rate of Random Walks on Digraphs 18
2.4.3 Numerical Analysis 20

3 Identifying Stable Clusters in Social Networks with Uni- and Bi-directional Links 22
3.1 Introduction 22
3.2 Preliminaries, Related Work and Problem definition 24
3.2.1 Dyadic Analysis and Mutuality Tendency 25
3.2.2 Spectral Clustering Theory and Extensions to Digraphs via Symmetrization 28
3.3 Cluster-based Mutuality Tendency Theory 31
3.3.1 Cluster based mutuality tendency 31
3.4 Mutuality-tendency-aware spectral clustering algorithm 34
3.4.1 Mutuality-tendency-aware spectral clustering: K=2 34
3.4.2 Mutuality-tendency-aware spectral clustering: $K > 2$ 38
3.5 Evaluations 39
3.5.1 Synthetic datasets 39
3.5.2 Real Social Networks 41
3.6 Summary 44

4 Influence Diffusion Dynamics and Influence Maximization in Social Networks with Friend and Foe Relationships 45
4.1 Introduction 45
4.1.1 Our contributions 47
4.1.2 Related work 48
4.2 Voter model for signed networks 49
4.3 Analysis of voter model dynamics on signed digraphs 51
4.3.1 Short-term dynamics 51
4.3.2 Convergence of signed transition matrix with relation to structural balance of signed digraphs 52
4.3.3 Long-term dynamics 60

4.4 Influence maximization . . . 65
4.4.1 Influence maximization problem . . . 66
4.4.2 Short-term influence maximization . . . 68
4.4.3 Long-term influence maximization . . . 68
4.5 Evaluation . . . 73
4.5.1 Performance comparison with baseline heuristics . . . 74
4.5.2 The impacts of signed information . . . 80
4.6 Summary . . . 82

5 From Shortest-path to All-path: The Routing Continuum Theory and its applications 83
5.1 Introduction . . . 83
5.2 Shortest Path and "All-Path" Routing as Network Flow Optimization . . . 85
5.2.1 Network and Flows: Basic Notations . . . 85
5.2.2 Shortest-Path Routing & L_1-norm Flow Optimization . . . 87
5.2.3 Potential-based ("All-path") Routing and L_2-norm Flow Optimization . 89
5.3 Mixed L_1 and L_2-norm Network Flow Optimization and the Routing Continuum 90
5.4 Computing the Routing Continuum . . . 94
5.4.1 Routing Continuum Algorithm (RCA) . . . 95
5.4.2 Numerical Illustration . . . 98
5.5 Generalizations and Applications . . . 101
5.5.1 Multiple Flows, Link/node Capacity Constraints and Traffic Engineering 102
5.5.2 Flow Optimization with Heterogeneous L_1/L_2 Link Weights . . . 104
5.5.3 Network robustness analysis via generalized centrality measure . . . 106
5.6 Related Work and Discussion . . . 111
5.7 Summary . . . 114

6 Conclusion, Discussion, and Future Research Directions 115
6.1 Future directions . . . 116

Bibliography 118

Appendix A. Omitted Proofs **133**

A.1 Properties of ergodic digraphs 133

A.2 Special matrix power series 134

A.3 Illustration of exponential convergence time of P^t on ergodic digraph. 135

List of Tables

3.1 Ave. mutuality tendency comparison on synthetic dataset 40
3.2 Statistics of Slashdot social network Dataset 43
3.3 Ave. mutuality tendency comparison on Slashdot dataset 44
4.1 Notations and terminologies 50
4.2 Statistics of Epinions and Slashdot datasets 79
5.1 Edge ranking in mixed-flow betweenness ($\theta \geq 0$). 110
5.2 Edge ranking in mixed-flow betweenness in Abilene network ($\theta \geq 0$). 111
5.3 Node betweenness ranking in the Abilene network for all θ's. 112

List of Figures

2.1 Knoke's data [1,2] on information exchanges among organizations operating in the social welfare field. The degree of asymmetry of this digraph is 0.3096. . . 20
2.2 Degree of asymmetry . . . 21
2.3 Comparing mixing rate bounds . . . 21
3.1 An example network . . . 23
3.2 Simulation results on synthetic dataset with $K = 2$ clusters. (a)(i)-(a)(iii) show the clusters detected by traditional spectral clustering algorithm, and (b)(i)-(b)(iii) show the clusters extracted using our mutuality-tendency-aware spectral clustering algorithm. . . . 39
3.3 This synthetic dataset is generated in $K = 3$ clusters, with 500, 400 and 300 nodes, respectively. There are 54675 directional edges, among which 27336 edges are bidirectional and 27339 edges are unidirectional. We are randomly placed 90.02% of the bidirectional edges in clusters, and 89.6% of the unidirectional edges across clusters. (a)(i)-(a)(iii) show that traditional spectral clustering algorithm detects clusters with 661, 538 and 1 entities, respectively, while our method identify correct clusters (See (b)(i)-(b)(iii)). . . . 41
3.4 Simulation results on Slashdot social network dataset. (a)(i)-(a)(iii) show the clusters detected by traditional spectral clustering algorithm, and (b)(i)-(b)(iii) show the clusters extracted using our mutuality-tendency-aware spectral clustering algorithm. . . . 42
4.1 G is balanced . . . 76
4.2 G is anti-balanced . . . 76
4.3 G is strictly unbalanced . . . 76
4.4 G is weakly connected . . . 76

4.5 G is disconnected . . . 76
4.6 G is disconnected with WCC . . . 76
4.7 Instant influence in Epinions data with $k = 6$k . . . 77
4.8 Instant influence in Epinions data with $k = 500$ . . . 77
4.9 Instant influence in SCC with $k = 6$k . . . 77
4.10 Instant influence in SCC with $k = 500$ . . . 77
4.11 Average influence in Epinions data with $k = 6$k . . . 78
4.12 Average influence in Epinions data with $k = 500$ . . . 78
4.13 Average influence in SCC with $k = 6$k . . . 78
4.14 Average influence in SCC with $k = 500$ . . . 78
4.15 Instant influence in Slashdot data with $k = 6$k . . . 80
4.16 Instant influence in Slashdot SCC with $k = 6$k . . . 80
4.17 Average influence in Slashdot data with $k = 6$k . . . 80
4.18 Average influence in Slashdot SCC with $k = 6$k . . . 80
4.19 Synthetic balanced digraph . . . 81
4.20 Synthetic weakly connected digraph . . . 81
4.21 Epinions (the entire dataset) . . . 81
4.22 Epinions (the largest SCC) . . . 81
5.1 Example 1 with uniform weight $w_{ij} = 1$ . . . 98
5.2 Flow distribution evolution of graph Fig. 5.1 . . . 98
5.3 Example 2: Weight Graph . . . 99
5.4 Flow distribution, $\theta_0 = 0$ . . . 99
5.5 Flow distribution with $\theta_1 = 0.0914$ . . . 99
5.6 Flow distribution with $\theta_2 = 0.2850$ . . . 99
5.7 Flow distribution with $\theta_3 = 0.5700$ . . . 99
5.8 Flow distribution with $\theta_4 = 2$ . . . 99
5.9 Weights on the Abilene network . . . 100
5.10 Flow distribution $\theta_0 = 0$ . . . 100
5.11 Flow distribution $\theta_1 = 0.1082$ . . . 100
5.12 Flow distribution $\theta_1 = 0.2498$ . . . 100
5.13 Flow distribution $\theta_3 = 0.4943$ . . . 100
5.14 Flow distribution $\theta_4 = 3.2108$ . . . 100

5.15 Abilene network topology 101
A.1 An example digraph with exponential convergence time. All edges are with unit weights. 135

Chapter 1

Introduction

Complex networks are becoming indispensable parts of our daily lives. The Internet, wireless (cellular) networks, on-line social networks, and transportation networks are examples of some well-known complex networks around us. These networks generate an immense range of big data: weblogs, social media, the Internet traffic, which have increasingly drawn attentions from the computer science research community to explore and investigate the fundamental properties of, and improve the user experiences on, these complex networks.

Complex networks arising from many applications can be represented and studied as graphs. For example, in an ad hoc wireless network, nodes represent wireless devices, whereas edges characterize the available wireless links among those devices. On an online social network (OSN), users and their social interactions can be viewed as nodes and edges in a graph. The topological properties of the underlying complex networks play an important role in understanding and solving the application problems. In the following, I elaborate on my works on developing theories [3, 4] and studying the crucial and unique properties of various types of graphs, such as directed graphs (digraphs), undirected graphs, and signed graphs, with applications to wireless networking [5–8], the Internet traffic engineering [9], and online social network analysis [10, 11].

1.1 Directed graph model on wireless and social networks.

Graphs arising from many applications are directed, where entity connections in a direct graph can be categorized into two types, namely, bi-directional links (mutual connections) and uni-directional links (one-way connections), for example, the users' following relations on Twitter, and the hyperlinking relations on the World Wide Web (WWW) network. Differing from undirected graphs, the direction of links contains crucial information, which makes it challenging to model and characterize such directed complex networks. One main contribution of my work is to extend and generalize the standard random walk theory and the intrinsically related spectral graph theory on undirected graphs to digraphs, which is further applied to solve various practical problems in wireless networks and social networks. For example, we model and estimate various transmission costs in wireless networks, design optimal opportunistic forwarding strategy, and develop a clustering algorithm to identify more stable social communities in directed OSNs, e.g., Twitter, Slashdot, etc. Those works have been reported in [3–6, 8, 11].

1.2 Undirected graph model: Towards traffic intelligence on complex networks.

Routing is a critical operation in networks. In the context of data and sensor networks, routing strategies such as shortest-path, multi-path and potential-based ("all-path") routing have been developed, which intrinsically represent the tradeoff between the latency and energy dissipation of paths used for routing, namely, shorter paths lead to better routing with low latency, while diffusing traffic along more paths generally reduces energy dissipation. Based on the connection between routing and flow optimization in a network, in [9], we consider networks as undirected graphs and develop a unifying theoretical framework by considering flow optimization with mixed (weighted) L_1/L_2-norms. We obtain a surprising result: as we vary the trade-off parameter θ, the routing graphs induced by the optimal flow solutions span from shortest-path to multi-path to all-path routing – this entire sequence of routing graphs is referred to as the routing continuum. We also develop an efficient iterative algorithm for computing the entire routing continuum. Several generalizations are also considered, with applications to traffic engineering and wireless sensor networks.

1.3 Signed graph model: Influence diffusion dynamics and influence maximization on social networks.

In many OSNs, such as Slashdot and Epinions, the edges (i.e., relations between users) carry heterogeneous weights, which can be either positive or negative, representing trust (or friend) and distrust (or foe) relations. These networks are referred to as signed networks, where those signed weights generate new challenges in understanding and studying the underlying network properties. For example, the matrix representation of a signed network, namely, the adjacency matrix is no longer non-negative, and it is not clear how the information is propagated on such graphs. In [10], we study the influence diffusion and influence maximization in social networks with both positive and negative relationships, where we extend the classic voter diffusion model to signed networks and design efficient seed selection algorithms for online viral marketing applications, that maximize the influence coverage in short term or long term, respectively.

1.4 Roadmap

- Chapter 2 develops and extends random walk theory and the intrinsically related spectral graph theory for undirected graphs to directed graphs.
- Chapter 3 introduces a novel social community detection algorithm for social networks with both uni- and bi-directional links.
- Chapter 4 establishes the routing continuum theory that spans from short path routing to "potential" based all-path routing, based on the connection between routing and network flow optimization problems.
- In Chapter 5, we investigate the social influence diffusion dynamics and influence maximization in social networks with both friend and foe relations.
- Chapter 6 concludes this dissertation and discusses future research directions.

Chapter 2

Digraph Laplacian and the Degree of Asymmetry

2.1 Introduction

Graphs arising from many applications such as web and online social networks are directed, where direction of links contains crucial information. Random walks are frequently used to model certain dynamic processes on directed or undirected graphs, for example, to reveal important network structural information, such as the importance of nodes using Page-Rank algorithm [12] and the community structures using spectral clustering algorithm [**?**], or to study ways to efficiently explore complex networks.

Random walks on undirected graphs have been extensively studied and are well-understood (see [13]). They are closely related to *spectral graph theory* [14], which has produced powerful tools to study many important properties of undirected graphs that are of both theoretical and practical significance. Well-known results include bounds on Cheeger constant and mixing rate in terms of the second smallest eigenvalue of the graph Laplacian. On the other hand, there are relatively few similar studies on directed graphs, see for example [15, 16], where the authors circumvent the "directedness" of digraphs by converting them into undirected graphs through *symmetrization*. Moreover, some initial attempts [5, 8] are made to extend the random walk theory to directed graphs with applications in wireless networking.

In this chapter we develop a spectral digraph theory. We introduce the notion of *Diplacian*, a direct generalization of the graph Laplacian for undirected graphs, denoted by Γ. Instead of

using the node degrees as in the case of undirected graphs, the Diplacian is defined using stationary probabilities of the Markov chain governing random walks on digraphs. Furthermore, instead of relying on the positive semi-definiteness of the graph Laplacian matrix for undirected graphs, we establish a key connection between the Diplacian Γ and its *Moore-Penrose pseudo-inverse* [17], denoted by Γ^+, and use the properties of this connection to prove several parallel results for random walks on digraphs. In particular, we show that 1) the Moore-Penrose pseudo-inverse Γ^+ of the Diplacian is exactly the discrete Green's function of the Diplacian matrix Γ, acting as an operator on digraphs [18], and 2) Γ^+ is the normalized *fundamental matrix* [19] of the Markov chain governing random walks on digraphs.

Based on the connection between Γ^+ and the fundamental matrix, we show how hitting and commute times can be directly computed in terms of the singular values and vectors of the Diplacian – this yields a more direct and efficient way to compute hitting and commute times than existing methods based on the fundamental matrix. More generally, our results suggest a "spectral digraph theory", where graph properties can be studied using the singular values of the Diplacian in place of the eigenvalues of the Laplacian. In particular, our theory of random walks on digraphs subsumes the existing theory of random walks on undirected graphs as a special case.

Furthermore, we show that the well-known Cheeger constant – generalized by Chung to digraphs in [15] – is fundamentally a quantity intrinsic to undirected graphs, as there are infinitely many digraphs with the same symmetrized undirected graph. Hence, bounds based on the eigenvalues of the symmetrized graph Laplacian do not uniquely capture the properties of digraphs. This leads us to introduce the *degree of asymmetry* to capture the overall "directedness" of digraphs, formally defined as follows: we express a Diplacian Γ in terms of a *symmetric part*, $\bar{\mathcal{L}} = (\Gamma + \Gamma^T)/2$, and a *skew-symmetric* part, $\nabla = (\Gamma - \Gamma^T)/2$, where $\bar{\mathcal{L}}$ is the symmetrized graph Laplacian for digraphs introduced by Chung in [15] and ∇ is referred to as skewed Laplacian for digraphs. The largest singular value of ∇, δ_{max}, is referred to as the *degree of asymmetry*, which provides a quantitative measure of the asymmetry in digraphs. Many key properties of digraphs can then be bounded by the eigenvalues of $\bar{\mathcal{L}}$ and the degree of asymmetry. For instance, by accounting for the asymmetry of digraphs, we are able to obtain a tighter bound than that of Chung's in [15] on non-reversible Markov chain mixing rate.

2.2 Preliminaries: Random Walks on Undirected Graphs

We use a triple $G = (V, E, A)$ to denote an undirected and weighted graph on the node set $V = \{1, 2, \ldots, n\}$. The $n \times n$ nonnegative weight matrix $A = [a_{ij}]$ is symmetric, and is defined in such a way that $a_{ij} = a_{ji} > 0$, if $(i, j) \in E$, and $a_{ij} = a_{ji} = 0$ otherwise. For $1 \leq i \leq n$, the degree of node i is $d_i = \sum_{j=1}^{n} a_{ij}$. The volume of G, denoted by $\mathrm{vol}(G)$, is defined as the sum of all node degrees, $d = \sum_{i=1}^{n} d_i$, that is, $\mathrm{vol}(G) = d$.

A random walk on G is a Markov chain defined on G with the transition probability matrix $P = [p_{ij}]$, where $p_{ij} = a_{ij}/d_i$. Let $D = \mathrm{diag}[d_i]$ be a diagonal matrix of node degrees, then $P = D^{-1}A$. Without loss of generality, we assume that the undirected graph G is connected, namely, any node can reach any other node in G. Then it can be shown (see [19]) that the Markov chain is irreducible, and there exists a unique stationary distribution, $\{\phi_1, \phi_2, \ldots, \phi_n\}$. Let $\phi = [\phi_i]_{1 \leq i \leq n}$ be the column vector of the stationary probabilities. Then $\phi^T P = \phi^T$, where the superscript T represents vector or matrix transpose. Furthermore, this Markov chain on G is *reversible*, namely

$$\phi_i p_{ij} = \phi_j p_{ji}, \text{ for any } i, j, \tag{2.1}$$

and

$$\phi_i = \frac{d_i}{\sum_k d_k} = \frac{d_i}{d}, \; i = 1, 2, \ldots, n. \tag{2.2}$$

Following [14], we will use the normalized graph Laplacian instead of the unnormalized version $L = D - A$. Given an undirected G, the normalized graph Laplacian of G, which is also called normalized Laplacian matrix of G, is defined as follows:

$$\mathcal{L} = D^{-\frac{1}{2}}(D - A)D^{-\frac{1}{2}} = D^{\frac{1}{2}}(I - P)D^{-\frac{1}{2}}. \tag{2.3}$$

A key property of the graph Laplacian for an undirected graph is that $\mathcal{L}$ is symmetric and positive semi-definite [17]. Hence, all eigenvalues of $\mathcal{L}$ are nonnegative real numbers. In particular, for a connected undirected graph G, $\mathcal{L}$ has rank $n - 1$ and has exactly one zero eigenvalue as its smallest eigenvalue. Let $\lambda_1 = 0 < \lambda_2 \leq \cdots \leq \lambda_n$ be the n eigenvalues of $\mathcal{L}$ arranged in an increasing order, and μ_i, $1 \leq i \leq n$, be the corresponding eigenvectors of unit norm. In particular, one can show that the column eigenvector, μ_1, of $\mathcal{L}$ associated with the eigenvalue $\lambda_1 = 0$, is given by

$$\mu_1 = \phi^{\frac{1}{2}} = [\sqrt{\phi_i}] = [\frac{\sqrt{d_i}}{\sqrt{d}}]. \tag{2.4}$$

Define Λ=diag$[\lambda_1, \ldots, \lambda_n]$, the diagonal matrix formed by the eigenvalues, and $U = [\mu_1, \ldots, \mu_n]$, an orthonormal matrix formed by the eigenvectors of $\mathcal{L}$, where $UU^T = U^TU = I$. It is easy to see that the graph Laplacian $\mathcal{L}$ admits an eigen-decomposition [17], namely, $\mathcal{L} = U\Lambda U^T$. Using the eigenvalues and eigenvectors of $\mathcal{L}$, we can compute the hitting times and commute times using the following formula [13]:

$$H_{ij} = \sum_{k>1} \frac{d}{\lambda_k} \left(\frac{\mu_{kj}^2}{d_j} - \frac{\mu_{ki}\mu_{kj}}{\sqrt{d_i d_j}} \right), \tag{2.5}$$

and

$$C_{ij} = \sum_{k>1} \frac{d}{\lambda_k} \left(\frac{\mu_{ki}}{\sqrt{d_i}} - \frac{\mu_{kj}}{\sqrt{d_j}} \right)^2, \tag{2.6}$$

where μ_{kj} is the jth entry of the column vector μ_k.

2.3 Random Walk Theory on Digraphs

In this section, we develop the random walk theory for digraphs. In particular, we generalize the graph Laplacian defined for undirected graphs, and introduce the *Diplacian* matrix. We prove that the Moore-Penrose pseudo-inverse of this Diplacian is exactly equal to a normalized version of the fundamental matrix of the Markov chain governing random walks on digraphs, and show that it is also the Green's function of the Diplacian. Using these connections, we illustrate that how hitting and commute times of random walks on digraphs can be directly computed using the singular values and vectors of the Diplacian. We also show that when the underlying graph is undirected, our results reduce to the well-known results for undirected graphs. Hence, our theory includes undirected graphs as a special case.

2.3.1 Random Walks on Directed Graphs and Fundamental Matrix

As alluded earlier, random walks can be defined not only on undirected graphs, but also on digraphs. Let $G = (V, E, A)$ be a weighted digraph defined on the vertex set $V = \{1, 2, \ldots, n\}$, where A is a nonnegative, but generally asymmetric weight matrix such that $a_{ij} > 0$ if and only if the directed edge (or arc) $(i, j) \in E$. As before, we will simply refer to A as the adjacency matrix of G. For $i = 1, 2, \ldots, n$, we define the *out-degree* of vertex i, $d_i^+ = \sum_{j=1}^n a_{ij}$, and the *in-degree* of vertex i, $d_i^- = \sum_{j=1}^n a_{ji}$. In general, $d^+ \neq d^-$. However, we have

$d = \sum_{i=1}^{n} d_i^+ = \sum_{i=1}^{n} d_i^- = \sum_{i=1}^{n} \sum_{j=1}^{n} a_{ij}$. As before, we refer to d as the volume of the directed graph G, that is, $\mathrm{vol}(G) = d$. For conciseness, in the following unless otherwise stated, we refer to the out-degree of a vertex simply as its degree, and use d_i for d_i^+.

Let $D = \mathrm{diag}[d_i]$ be a diagonal matrix of the vertex out-degrees, and define $P = D^{-1}A$. Then $P = [p_{ij}]$ is the transition probability matrix of the Markov chain associated with random walks on G, where at each vertex i, a random walk has the probability $p_{ij} = a_{ij}/d_i$ to transit from vertex i to vertex j, if $(i, j) \in E$. We assume that G is *strongly connected*, namely, there is a directed path from any vertex i to any other vertex j. Then the Markov chain P is irreducible, and has a unique stationary probability distribution, $\{\phi_i\}$, where $\phi_i > 0, 1 \leq i \leq n$, namely, $\phi^T P = \phi^T$, where $\phi = [\phi_1, \ldots, \phi_n]^T$ be the column vector of stationary probabilities. Unlike undirected graphs, the Markov chain associated with random walks on directed graphs is generally non-reversible, and (2.1) and (2.2) for undirected graphs do not hold.

For random walks on directed graphs, quantities such as hitting times and commute times can be defined exactly as in the case of undirected graphs. However, since the normalized Laplacian matrix $\mathcal{L}$ is so far defined only for undirected graphs, we cannot use the relations (2.5) and (2.6) to compute hitting times and commute times for random walks on directed graphs. On the other hand, using results from the standard Markov chain theory, we can express the hitting times and commute times in terms of the fundamental matrix. In [19], Aldous and Fill define the fundamental matrix $Z = [z_{ij}]$ for an irreducible Markov chain with the transition probability matrix P:

$$z_{ij} = \sum_{t=0}^{\infty} (p_{ij}^{(t)} - \phi_j), 1 \leq i, j \leq n, \tag{2.7}$$

where $p_{ij}^{(t)}$ is the (i, j)-th entry in the t-step transition probability matrix $P^t = \underbrace{P \cdots P}_{t}$.

Let $\Phi = \mathrm{diag}[\phi_i]$ be the diagonal matrix containing the stationary probabilities ϕ_i's on the diagonal, and $J = [J_{ij}]$ the all-one matrix, that is, $J_{ij} = 1$ for $1 \leq i, j \leq n$. We can express Z alternatively as the sum of an infinite matrix series:

$$Z = \sum_{t=0}^{\infty} (P^t - J\Phi) = \sum_{t=0}^{\infty} (P^t - \mathbf{1}\phi^T), \tag{2.8}$$

where $\mathbf{1} = [1, \ldots, 1]^T$ is the all-one column vector. Hence, $J = \mathbf{1} \cdot \mathbf{1}^T$ and $\mathbf{1}^T\Phi = \phi^T$ hold.

While the physical meaning of the fundamental matrix Z may not be obvious from its definition (2.7) or (2.8), it plays a crucial role in computing various quantities related to random

walks, or more generally, various stopping time properties of Markov chains [19]. For instance, the hitting times and commute times of random walks on a directed graph can be expressed in terms of Z as follows (see [19]):

$$H_{ij} = \frac{z_{jj} - z_{ij}}{\phi_j} \tag{2.9}$$

and

$$C_{ij} = \frac{z_{jj} - z_{ij}}{\phi_j} + \frac{z_{ii} - z_{ji}}{\phi_i}. \tag{2.10}$$

In (2.7) and (2.8), the fundamental matrix Z is defined as an infinite sum. We show that Z in fact satisfies a simple relation (2.11), thus can be computed directly using the standard matrix inverse.

Theorem 1 *Let P be the transition probability matrix for an irreducible Markov chain. Then its corresponding fundamental matrix Z as defined in (2.7) satisfies the following relation*

$$Z + J\Phi = (I - P + J\Phi)^{-1}. \tag{2.11}$$

Proof: Note that $J\Phi = \mathbf{1}\phi^T$. From $\phi^T P = \phi^T$ and $P\mathbf{1} = \mathbf{1}$, we have $J\Phi P = J\Phi$ and $PJ\Phi = J\Phi$. Using these two relations, it is easy to prove the following equation by induction.

$$P^m - J\Phi = (P - J\Phi)^m, \text{ for any integer } m > 0. \tag{2.12}$$

Substituting this in (2.8) yields Theorem 1. ∎

As undirected graphs are a special case of directed graphs, (2.9) and (2.10) provide an alternative way to compute hitting times and commute times for random walks on connected undirected graphs. In this chapter we will show that (2.5) and (2.6) are in fact equivalent to (2.9) and (2.10).

2.3.2 Diplacian and Green's Function for Digraphs

We now generalize the existing spectral graph theory defined for undirected graphs to digraphs by introducing the Diplacian – an appropriately generalized Laplacian matrix – for strongly connected diagraphs. Let $G = (V, E, A)$ be a strongly connected weighted digraph defined on the vertex set $V = \{1, 2, \ldots, n\}$, where in general the weight (or adjacency) matrix A is asymmetric. A major technical difficulty in dealing with digraphs is that if one naively extends

the normalized Laplacian matrix, $\mathcal{L} = D^{-1/2}(D - A)D^{-1/2}$, or its un-normalized version, $L = D - A$, defined for undirected graphs to digraphs, $\mathcal{L}$ is in general asymmetric; thus the nice properties such as positive semi-definiteness of $\mathcal{L}$ no longer hold. Past attempts in generalizing the spectral graph theory to digraphs have been simply symmetrized $\mathcal{L}$, for example, by introducing a symmetric matrix, $\bar{\mathcal{L}} = I - (\Phi^{\frac{1}{2}} P \Phi^{-\frac{1}{2}} + \Phi^{-\frac{1}{2}} P^T \Phi^{\frac{1}{2}})/2$ [15, 16]. Unfortunately, as will be shown in the Section 2.4, such symmetrized $\bar{\mathcal{L}}$ does not directly capture the unique characteristic of the random walk on the digraph as defined earlier, since a set of diagraphs can have the same $\bar{\mathcal{L}}$.

For a strongly connected digraph G, let $\Phi^{\frac{1}{2}} = \mathrm{diag}[\sqrt{\phi_i}]$. We define the normalized digraph Laplacian matrix[1] (in short, *Diplacian*) for G, $\Gamma = [\Gamma_{ij}]$ as follows:

Definition 1 (Diplacian Γ)

$$\Gamma = \Phi^{\frac{1}{2}}(I - P)\Phi^{-\frac{1}{2}}, \tag{2.13}$$

namely, for $1 \leq i, j \leq n$,

$$\Gamma_{ij} = \begin{cases} 1 - p_{ii} & \text{if } i = j, \\ -\phi_i^{\frac{1}{2}} p_{ij} \phi_j^{-\frac{1}{2}} & \text{if } (i,j) \in E, \\ 0 & \text{otherwise.} \end{cases} \tag{2.14}$$

Treating this Diplacian matrix Γ as an asymmetric operator on a digraph G, we now define the discrete Green's function $\mathcal{G}$ without boundary conditions for digraphs in exactly the same manner as for undirected graphs [18], namely, $\mathcal{G}$ is a matrix with its entries, indexed by vertices i and j, that satisfies the following conditions:

$$[\mathcal{G}\Gamma]_{i,j} = I_{i,j} - \sqrt{\phi_i \phi_j}, 1 \leq i, j \leq n, \tag{2.15}$$

and expressed in the matrix form,

$$\mathcal{G}\Gamma = I - \phi^{\frac{1}{2}} \phi^{\frac{1}{2}^T}. \tag{2.16}$$

In the following we will show that $\mathcal{G}$ is precisely Γ^+, the pseudo-inverse of the Laplacian operator Γ on the digraph G. Furthermore, we will relate Γ^+ directly to the fundamental matrix Z of the Markov chain associated with random walks on the digraph G. Before we establish a main result of this chapter, we first introduce a few more notations and then prove the following useful lemma.

[1] An un-normalized digraph Laplacian is defined as $L = \Phi(I - P)$ in [20].

Lemma 1 *Define $\mathcal{Z} = \Phi^{\frac{1}{2}} Z \Phi^{-\frac{1}{2}}$ as the normalized fundamental matrix, and $\mathcal{J} = \Phi^{\frac{1}{2}} J \Phi^{\frac{1}{2}} = \phi^{\frac{1}{2}} \phi^{\frac{1}{2}^T}$. The following relations regarding $\mathcal{Z}$ and $\mathcal{J}$ hold: (i) $\mathcal{J} = \mathcal{J}^2$, (ii) $\mathcal{J}\Gamma = \Gamma\mathcal{J} = \mathcal{J}\mathcal{Z} = \mathcal{Z}\mathcal{J} = \boldsymbol{0}_{n\times n}$, and (iii) $\Gamma\phi^{\frac{1}{2}} = \mathcal{Z}\phi^{\frac{1}{2}} = \boldsymbol{0}$, $\phi^{\frac{1}{2}^T}\Gamma = \phi^{\frac{1}{2}^T}\mathcal{Z} = \boldsymbol{0}^T$.*

Proof : These relations can be established using the facts that $J = \mathbf{1}\mathbf{1}^T$, $\mathbf{1}^T\Phi = \phi^T$, $\phi^T J = \mathbf{1}^T$, $\Phi J = \phi\mathbf{1}^T$, $J\Phi J = J$, $\phi^T(I-P) = \mathbf{0}^T$, $(I-P)\mathbf{1} = \mathbf{0}$, $\phi^T Z = \mathbf{0}^T$, and $Z\mathbf{1} = \mathbf{0}$. The last four equalities imply that the matrices $I - P$ and Z have the same left and right eigenvectors, ϕ and $\mathbf{1}$, corresponding to the eigenvalue 0. ∎

We are now in a position to prove a main theorem of the chapter, which states the Green's function for the Diplacian is exactly its Moore-Penrose pseudo-inverse, and it is equal to the normalized fundamental matrix, namely, $\mathcal{G} = \Gamma^+ = \mathcal{Z}$.

Theorem 2 (Diplacian matrix and Green's function for digraphs) *Let $G = (V, E, A)$ be a strongly connected digraph with the normalized fundamental matrix $\mathcal{Z}$ as defined in Lemma 1, and the Diplacian matrix Γ as defined in (2.13). Then $\mathcal{Z} = \Gamma^+$, is the pseudo-inverse of the Diplacian matrix Γ. Furthermore, $\mathcal{Z}$ is the discrete Green's function for Γ, namely,*

$$\mathcal{Z}\Gamma = I - \Phi^{\frac{1}{2}} J \Phi^{\frac{1}{2}} = I - \phi^{\frac{1}{2}} \phi^{\frac{1}{2}^T}, \tag{2.17}$$

where J is the all-one matrix and $\phi^{\frac{1}{2}} = [\phi_1^{\frac{1}{2}}, \ldots, \phi_n^{\frac{1}{2}}]^T$ is a column vector.

Proof : From equation (2.11) in Theorem 1, we have

$$\mathcal{Z} + \mathcal{J} = (\Gamma + \mathcal{J})^{-1}. \tag{2.18}$$

Multiplying (2.18) from the right by $\Gamma + \mathcal{J}$, and using Lemma 1, it is easy to see that

$$\mathcal{Z}\Gamma = I - \mathcal{J}, \tag{2.19}$$

which establishes that $\mathcal{Z}$ is the Green's function of the Diplacian Γ. Similarly, by multiplying (2.18) from the left by $\Gamma + \mathcal{J}$, we can likewise prove $\Gamma\mathcal{Z} = I - \mathcal{J}$. Hence, $\mathcal{Z}\Gamma = \Gamma\mathcal{Z} = I - \mathcal{J}$ is a real symmetric matrix, which infers that $(\Gamma\mathcal{Z})^T = \Gamma\mathcal{Z}$ and $(\mathcal{Z}\Gamma)^T = \mathcal{Z}\Gamma$ hold. Furthermore, as $\mathcal{J}\mathcal{Z} = 0$, equation (2.19) yields $\mathcal{Z}\Gamma\mathcal{Z} = \mathcal{Z}$. Similarly, as $\Gamma\mathcal{J} = \Phi^{\frac{1}{2}}(I-P)J\Phi^{\frac{1}{2}} = 0$, equation (2.19) yields $\Gamma\mathcal{Z}\Gamma = \Gamma$. These establish that $\mathcal{Z}$ satisfies the four conditions of matrix pseudo-inverse. Hence, $\mathcal{Z}$ is also the Moore-Penrose pseudo-inverse of Γ. Therefore, $\mathcal{G} = \mathcal{Z} = \Gamma^+$. ∎

2.3.3 Computing Hitting and Commute Times for Digraphs using Diplacian

Using the relationship between the Diplacian Γ, its pseudo-inverse Γ^+, and the normalized fundamental matrix $\mathcal{Z}$, we can now express the hitting times and commute times of random walks on digraphs in terms of Γ^+, or alternatively in terms of the singular values and singular vectors of the Diplacian matrix Γ.

From $Z = \Phi^{-\frac{1}{2}}\mathcal{Z}\Phi^{\frac{1}{2}} = \Phi^{-\frac{1}{2}}\Gamma^+\Phi^{\frac{1}{2}}$, and using (2.9) and (2.10), we can compute the hitting times and commute times for random walks on digraphs directly in terms of the entries of Γ^+.

Theorem 3 (Computing hitting and commute times using Γ^+) *The hitting times and commute times of random walks on a strongly connected digraphs can be computed using the pseudo-inverse of the Diplacian matrix Γ^+ as follows:*

$$H_{ij} = \frac{\Gamma^+_{jj}}{\phi_j} - \frac{\Gamma^+_{ij}}{\sqrt{\phi_i\phi_j}}, \tag{2.20}$$

and

$$C_{ij} = H_{ij} + H_{ji} = \frac{\Gamma^+_{jj}}{\phi_j} + \frac{\Gamma^+_{ii}}{\phi_i} - \frac{\Gamma^+_{ij}}{\sqrt{\phi_i\phi_j}} - \frac{\Gamma^+_{ji}}{\sqrt{\phi_i\phi_j}}, \tag{2.21}$$

where Γ^+_{ij} is the (i,j)-th entry of Γ^+, and ϕ_i is the stationary probability of vertex i.

For undirected graphs, we show that (2.20) and (2.21) reduce to (2.5) and (2.6) in Section 3.2. This can be seen from the fact that for undirected graphs, $\Gamma = \mathcal{L}$ is symmetric and positive semi-definite. Hence, the singular value decomposition of $\mathcal{L}$ is the same as the eigen-decomposition of $\mathcal{L}$.

For digraphs we can express Γ^+_{ij} directly in terms of the singular values and left/right singular vectors of the generally asymmetric Diplacian matrix Γ. Let σ_i, u_i and v_i be the i-th singular value and the corresponding left and right singular vectors of Γ arranged in an increasing order, where $\|u_i\|_2 = 1$ and $\|v_i\|_2 = 1$, $i = 1, 2, \ldots, n$. In particular, $0 = \sigma_1 < \sigma_2 \leq \cdots \leq \sigma_n$. Hence, $\Gamma = U\Sigma V^T$, where $\Sigma = \mathrm{diag}[\sigma_i]$, $U = [u_1, \ldots, u_n]$, $V = [v_1, \ldots, v_n]$, $UU^T = I$ and $VV^T = I$, and $\Gamma^+ = V\Sigma^+U^T$, where $\Sigma^+ = \mathrm{diag}[\sigma_i^+]$. Therefore, $\Gamma^+_{ij} = \sum_{k>1}\frac{1}{\sigma_k}v_{ki}u_{kj}$. Plugging these equations into (2.20) and (2.21), we can express the hitting times and commute times for random walks in a digraph in terms of the singular values, left and right singular vectors of the Diplacian matrix Γ as below:

$$H_{ij} = \sum_{k>1}\frac{1}{\sigma_k}\left(\frac{v_{kj}u_{kj}}{\phi_j} - \frac{v_{ki}u_{kj}}{\sqrt{\phi_i\phi_j}}\right), \tag{2.22}$$

and

$$C_{ij} = \sum_{k>1} \frac{1}{\sigma_k} \left(\frac{v_{kj} u_{kj}}{\phi_j} + \frac{v_{ki} u_{ki}}{\phi_i} - \frac{v_{ki} u_{kj}}{\sqrt{\phi_i \phi_j}} - \frac{v_{kj} u_{ki}}{\sqrt{\phi_j \phi_i}} \right). \tag{2.23}$$

2.4 Degree of Asymmetry, Generalized Cheeger Constant and Bounds on Mixing Rate

In this section we explore the relation between Diplacian Γ and the symmetrized Laplacian $\bar{\mathcal{L}}$. We first show that the symmetrized Laplacian matrix $\bar{\mathcal{L}}$, and the Cheeger constant $h(G)$ as defined in [15] are in a sense primarily determined by an undirected graph associated with the random walks with the transition probability matrix $\bar{P} = (P + \Phi^{-1} P^T \Phi)/2$, thus cannot capture the unique characteristics of each individual diagraph. As a result, we investigate two questions: 1) how can the "degree of asymmetry" of a digraph be quantified and measured? and 2) how does the degree of asymmetry affect crucial properties of a digraph such as the mixing rate? In the following we propose one metric – the largest singular value of $\nabla = (\Gamma - \Gamma^T)/2$ – as a measure of the degree of asymmetry in a digraph. We show that by explicitly accounting for the degree of asymmetry, we can obtain generally tighter bounds on quantities (for example, mixing rate) associated with random walks (or Markov chains) on digraphs.

2.4.1 The Degree of Asymmetry, and Relations to Symmetrized Laplacian

In [15], Chung introduces the symmetrized Laplacian matrix for digraphs, $\bar{\mathcal{L}} = \frac{\Gamma + \Gamma^T}{2}$, generalizes the Cheeger constant to digraphs and bounds it in terms of the second smallest eigenvalue of $\bar{\mathcal{L}}$. In the following we show that the symmetrized Laplacian $\bar{\mathcal{L}}$ and the Cheeger constant introduced by Chung are in fact two quantities intrinsic to undirected graphs.

Theorem 4 *Given a digraph G, with transition probability matrix P, there exist infinite digraphs which have the same stationary distribution matrix Φ and the same symmetrized transition probability matrix $\bar{P} = (P + \Phi^{-1} P^T \Phi)/2$. As a result, all these graphs have the same symmetrized Laplacian matrix and Cheeger constant.*

Proof : We prove it by construction. Given a digraph $G = (V, E, A)$, with transition probability matrix P, all the digraphs G''s with the transition probability P' as

$$P'(\alpha) = \alpha P + (1 - \alpha) \Phi^{-1} P^T \Phi, \tag{2.24}$$

form an infinite digraph set, denoted by $\mathcal{G}_G$, where $\alpha \in [0,1]$.
It is easy to check that any $P'(\alpha)$ defined in (2.24) is non-negative, and satisfies $\phi^T P'(\alpha) = \phi^T$, and $P'(\alpha)\mathbf{1} = \mathbf{1}$, thus $P'(\alpha)$ represents a transition probability matrix of a random walk with stationary distribution ϕ.

For any $G' \in \mathcal{G}_G$, the Diplacian matrix is given by $\Gamma' = \Phi^{\frac{1}{2}}(I - P')\Phi^{-\frac{1}{2}}$, and the symmetrized Laplacian is only determined by $\bar{P}$, since we have

$$\bar{\mathcal{L}} = \frac{\Gamma' + \Gamma'^T}{2} = \Phi^{\frac{1}{2}}(I - \frac{P + \Phi^{-1}P^T\Phi}{2})\Phi^{-\frac{1}{2}} = \Phi^{\frac{1}{2}}(I - \bar{P})\Phi^{-\frac{1}{2}}. \tag{2.25}$$

In particular, when $\alpha = \frac{1}{2}$, $P'(\frac{1}{2}) = \bar{P}$ represents the undirected graph $\bar{G}$.
For any $S \subset N=\{1,\ldots,n\}$, define an n-element vector f_S, as $f_S(i) = 1/F_\phi(S)$, $i \in S$ and $f_S(i) = -1/F_\phi(\bar{S})$, $i \in \bar{S}$, where $F_\phi(S) := \sum_{i\in S}\phi_i$ and $F_\phi(\bar{S}) := \sum_{i\in\bar{S}}\phi_i$ are the circulation function [15]. Define $x_S = \Phi^{-\frac{1}{2}} f_S$. Then, we have

$$\begin{aligned} x_S^T\Gamma x_S &= f_S^T\Phi(I-P)f_S = \sum_{i\in S, j\in\bar{S}} \phi_i p_{ij} = F_\phi(\partial S) \\ x_S^T\Gamma x_S &= x_S^T\Gamma^T x_S = \frac{1}{2}(x_S^T\Gamma x_S + x_S^T\Gamma^T x_S) = x_S^T\bar{\mathcal{L}}x_S. \end{aligned} \tag{2.26}$$

Hence, we know for all the graphs $G' \in \mathcal{G}_G$, the circulation $F_\phi(\bar{S})$, $F_\phi(S)$ and $F_\phi(\partial S)$ hinge only on the partition $S \vee \bar{S} = V$. Therefore, any graph $G' \in \mathcal{G}_G$ has the same Cheeger constant, that is,

$$\min_S \frac{F_\phi(\partial S)}{\min\{F_\phi(S), F_\phi(\bar{S})\}} = h(G') = h(\bar{G}). \tag{2.27}$$

■

To capture the "degree of asymmetry" in a digraph, we express Γ as a sum of a symmetric part and a skew-symmetric part:

$$\Gamma = \bar{\mathcal{L}} + \nabla, \tag{2.28}$$

where we define the skew symmetric part $\nabla = (\Gamma - \Gamma^T)/2$ as skewed Laplacian matrix of a digraph. Note that $\Gamma^T = \bar{\mathcal{L}} + \nabla^T = \bar{\mathcal{L}} - \nabla$. Hence, ∇ captures the difference between Γ and its transpose, which induces a reserved Markov chain or random walk. When Γ is symmetric, then $\nabla = 0$. Let $0 = \sigma_1 \leq \sigma_2 \leq \ldots \leq \sigma_n$ denote the singular values in an increasing order of Γ. Likewise, let $0 = \bar{\lambda}_1 \leq \bar{\lambda}_2 \leq \ldots \leq \bar{\lambda}_n$ denote the eigenvalues of $\bar{\mathcal{L}}$, and

$0 = \delta_1 \leq \delta_2 \leq \ldots \leq \delta_n = \delta_{max}$ the singular values of ∇. The following relations among them hold (See [21]):

$$\bar{\lambda}_i \leq \sigma_i \leq \bar{\lambda}_i + \delta_n, \; i = 1, 2, \ldots, n. \tag{2.29}$$

From equation (2.29), we see that $\sigma_i - \bar{\lambda}_i \leq \delta_n$, $i = 1, \ldots, n$. When the graph is undirected, we have $\Gamma = \bar{\mathcal{L}}$, thus $\delta_n = 0$ and $\sigma_i = \lambda_i$, $i = 1, \ldots, n$. We therefore propose the largest singular value of ∇, $\delta_n = \delta_{max}$ as a measure of the degree of asymmetry in the underlying digraph. Note that $\delta_n = \|\nabla\|$, where $\|\cdot\|$ is the operator bound norm of a matrix: $\|\nabla\| = \sup_{\|x\|=1} \|\nabla x\|_2 = \sup_{\|y\|=\|x\|=1} |\langle y, \Gamma x\rangle - \langle y, \Gamma^T x\rangle| = \sup_{\|y\|=\|x\|=1} |\langle y, \Gamma x\rangle - \langle x, \Gamma y\rangle|$ (see [21], p.6 and p.91). On the other hand, $\langle x, \Gamma x\rangle - \langle x, \Gamma^T x\rangle = 0$ for any x.

In the following, we relate and bound δ_n – the degree of asymmetry – to two other important quantities associated with the Markov chain on a digraph: the *digraph gap* $g(G)$ defined below and the second largest singular value of the transmission probability matrix P.

Given a digraph G, the circulation function $F_\phi(\cdot)$, where $F_\phi(i, j) = \phi_i p_{ij}$, obeys the flow conservation law at every node of a digraph: $\sum_k F_\phi(k, i) = \sum_j F_\phi(i, j)$ for all i's. Now, define the *digraph gap* $g(G) = \max_S \sum_{i \in S} |\sum_{j \in \bar{S}} (F_\phi(i, j) - F_\phi(j, i))|$, which quantifies the maximum difference between two bipartite subgraphs S and $\bar{S}$ among all partitions. We have the following theorem relating the degree of asymmetry with $g(G)$ and $\sigma_{n-1}(\mathcal{P})$, the second largest singular value of $\mathcal{P} = \Phi^{\frac{1}{2}} P \Phi^{-\frac{1}{2}}$.

Theorem 5 (Bounds on the degree of asymmetry)

$$2g(G) \leq \delta_n \leq \lambda_{n-1}^{\frac{1}{2}}(\mathcal{P}^T \mathcal{P}) = \sigma_{n-1}(\mathcal{P}), \tag{2.30}$$

where $\mathcal{P} = \Phi^{\frac{1}{2}} P \Phi^{-\frac{1}{2}}$.

Proof : For a subset of vertices $S \subset N = \{1, \ldots, n\}$, define two n-element vectors y_S and z_S as

$$y_S(i) = \begin{cases} \phi_i^{\frac{1}{2}} & \text{if } i \in \bar{S} \text{ or } g(i, \bar{S}) > 0 \\ -\phi_i^{\frac{1}{2}} & \text{otherwise} \end{cases}, z_S(i) = \begin{cases} \phi_i^{\frac{1}{2}} & \text{if } i \in \bar{S} \text{ or } g(i, \bar{S}) < 0 \\ -\phi_i^{\frac{1}{2}} & \text{otherwise} \end{cases} \tag{2.31}$$

where $g(i, \bar{S}) = \sum_{j \in \bar{S}} (p_{ij}\phi_i - p_{ji}\phi_j)$. Then, we have $\|y_S\| = 1$, $\|z_S\| = 1$, and the lower bound of δ_n is obtained as

$$2g(G) = \max_S y_S^T \nabla z_S \leq \delta_n. \tag{2.32}$$

Moreover, we have the following equation hold true,

$$\nabla^T\nabla = \frac{1}{2}(\mathcal{P}^T\mathcal{P} + \mathcal{P}\mathcal{P}^T) - \frac{1}{4}(\mathcal{P}^T + \mathcal{P})^2. \tag{2.33}$$

It is easy to check that $\phi^{\frac{1}{2}}$ is the left and right eigenvector of ∇ corresponding to the smallest eigenvalue 0, and the eigenvector of $\mathcal{P}$ corresponding to the largest eigenvalue 1. We also know that $(\mathcal{P}^T + \mathcal{P})^2$ is positive semi-definite. Therefore, for any n by 1 column vector $x \perp \phi^{\frac{1}{2}}$, we have

$$\delta_n^2 = \max_{x\perp\phi^{\frac{1}{2}}} x^T\nabla^T\nabla x \le \frac{1}{2}\max_{x\perp\phi^{\frac{1}{2}}} x^T(\mathcal{P}^T\mathcal{P} + \mathcal{P}\mathcal{P}^T)x \le \max_{x\perp\phi^{\frac{1}{2}}} x^T\mathcal{P}^T\mathcal{P}x = \lambda_{n-1}(\mathcal{P}^T\mathcal{P}),$$

$$\delta_n \le \lambda_{n-1}^{\frac{1}{2}}(\mathcal{P}^T\mathcal{P}) = \sigma_{n-1}(\mathcal{P}). \tag{2.34}$$

(2.32) and (2.34) yield Theorem 5. ∎

Theorem 6 below relates and bounds the second smallest singular value σ_2 of Γ in terms of the degree of asymmetry δ_n, the Cheeger constant, and the second smallest eigenvalue $\bar{\lambda}_2$ of $\bar{\mathcal{L}}$.

Theorem 6 (Relations among σ_2, $\bar{\lambda}_2$, δ_n and the Cheeger constant) *Given a strongly connected graph $G = (V, E, A)$, and its Laplacian matrix $\Gamma = \Phi^{\frac{1}{2}}(I - P)\Phi^{-\frac{1}{2}}$, we have the bounds for the second smallest singular value of Γ as*

$$\frac{h^2(G)}{2} \le \sigma_2 \le \left(1 + \frac{\delta_n}{\bar{\lambda}_2}\right)\cdot 2h(G). \tag{2.35}$$

When the graph is undirected, we have $\frac{h^2(G)}{2} \le \sigma_2 = \bar{\lambda}_2 \le 2h(G)$, which is exactly the same as the bounds obtained in [15].

Proof : *(i) Proof for the upper bound:* Let σ_2 be the second smallest (or the smallest non-zero) singular value of the Diplacian matrix Γ. and denote the singular decomposition of the Moore-Penrose pseudo-inverse of the Diplacian as $\Gamma^+ = V\Sigma^+U^T$, where each diagonal entry of Σ^+ is $\sigma_i^+ = \frac{1}{\sigma_i}$, if $\sigma_i \neq 0$; $\sigma_i^+ = 0$ otherwise. Then, the largest singular value of Γ^+ is $\max_i \sigma_i^+ = \frac{1}{\sigma_2}$.

Let S denote a subset of vertices $S \subset V$, and $\bar{S}$ denotes the complement of S, that is, $\bar{S} = V - S$. Then we define $x_0 = \Phi^{\frac{1}{2}} f_S$, $y_0 = \Phi^{\frac{1}{2}}(I - P)f_S = \Gamma x_0$, where $f : V \to \mathbb{R}$ is defined as follows:

$$f_S(i) = \begin{cases} \frac{1}{F_\phi(S)} & \text{if } i \in S \\ -\frac{1}{F_\phi(\bar{S})} & \text{if } i \in \bar{S} \end{cases}, \tag{2.36}$$

with $F_\phi(S) = \sum_{i\in S}\phi_i$ and $F_\phi(\bar{S}) = \sum_{i\in\bar{S}}\phi_i$. Then, for each i ($1 \le i \le n$), we have $\langle v_i, x_0\rangle = \sigma_i^+\langle u_i, y_0\rangle$, which indicates that $\langle v_i, x_0\rangle$ and $\langle u_i, y_0\rangle$ have the same sign. Then we have

$$\sigma_2 = \frac{1}{\max_i \sigma_i^+} \le \frac{1}{\frac{\sum_{i=1}^n \langle v_i,x_0\rangle\sigma_i^+\langle u_i,y_0\rangle}{\sum_{i=1}^n \langle v_i,x_0\rangle\langle u_i,y_0\rangle}} = \frac{x_0^T V U^T y_0}{x_0^T\Gamma^+ y_0} \tag{2.37}$$

The denominator of (2.37) can be rewritten as

$$x_0^T\Gamma^+ y_0 = f_S^T\Phi^{\frac{1}{2}}\Gamma^+\Phi^{\frac{1}{2}}(I-P)f_S = f_S^T\Phi Z(I-P)f_S = f_S^T\Phi f_S = F_\phi(S)F_\phi(\bar{S}) \ge 0, \tag{2.38}$$

where Z is the fundamental matrix. The third equality holds true, because we have $Z(I-P) = I - J\Phi$ by Theorem 1, and f_S is orthogonal to ϕ, that is, $J\Phi f_S = \phi\phi^T f_S = 0$, where J is all-1 matrix, thus we have $(I-J\Phi)f_S = f_S - J\Phi f_S = f_S$. From (2.37) and (2.38), we have

$$\begin{aligned}\sigma_2 &\le \frac{x_0^T V U^T y_0}{F_\phi(S)F_\phi(\bar{S})} = \frac{x_0^T V U^T y_0}{x_0^T y_0}\cdot\frac{x_0^T y_0}{F_\phi(S)F_\phi(\bar{S})} \\ &= \frac{x_0^T V U^T y_0}{x_0^T y_0}\cdot\frac{f_S^T\Phi(I-P)f_S}{F_\phi(S)F_\phi(\bar{S})} \le \frac{x_0^T V U^T y_0}{x_0^T y_0}\cdot 2h(G) = c_0\cdot 2h(G),\end{aligned} \tag{2.39}$$

where $h(G)$ is the Cheeger constant of digraph $G = (V, E, A)$. Now, we are in a position to bound $c_0 = \frac{x_0^T V U^T y_0}{x_0^T y_0} \le 1 + \frac{\delta_n}{\lambda_2}$.

Let $R = VU^T$. Since $x_0^T\Gamma x_0 = x_0^T\Gamma^T x_0$, we can rewrite c_0 as

$$c_0 = \frac{x_0^T R y_0}{x_0^T y_0} = \frac{x_0^T R\Gamma x_0}{x_0^T\Gamma x_0} = \frac{x_0^T V\Sigma V^T x_0}{x_0^T\bar{\mathcal{L}}x_0},$$

where both $R\Gamma$ and $\bar{\mathcal{L}}$ are symmetric positive semi-definite matrices. Denote the eigen-decomposition of $\bar{\mathcal{L}}$ as $\bar{\mathcal{L}} = W\Lambda W^T = C^T C$, where $C = \Lambda^{\frac{1}{2}}W^T$ and define $z_0 = Cx_0$. Then, we have $x_0^T C^T (C^+)^T = x_0^T(I-\mathcal{J}) = x_0^T$, with $x_0 \perp \phi^{\frac{1}{2}}$ and $\mathcal{J} = \phi^{\frac{1}{2}}{\phi^{\frac{1}{2}}}^T$. As a result, we have

$$c_0 = \frac{z_0^T {C^+}^T R\Gamma C^+ z_0}{z_0^T z_0} = \frac{z_0^T T z_0}{z_0^T z_0}. \tag{2.40}$$

Since $T = {C^+}^T R\Gamma C^+$ is symmetric, it has real eigenvalues. By substituting $z = Cx$ in $Tz = \lambda z$, we prove that T and $\bar{\mathcal{L}}^+ R\Gamma$ have the same eigenvalues:

$$\begin{aligned}&Tz - \lambda z = \bar{\mathcal{L}}^+ R\Gamma x - \lambda x = \mathbf{0}, \text{ if } x \perp \phi^{\frac{1}{2}}, \\ &Tz - \lambda z = \bar{\mathcal{L}}^+ R\Gamma(I-\mathcal{J})x - \lambda(I-\mathcal{J})x = \bar{\mathcal{L}}^+ R\Gamma\mathbf{0} - \lambda\mathbf{0} = \mathbf{0}, \text{ if } x = \alpha\phi^{\frac{1}{2}},\end{aligned} \tag{2.41}$$

where $\alpha \in \mathbb{R}$. From (2.41) and (2.40), we have

$$c_0 \leq \max_z \frac{z^T T z}{z^T z} = \lambda_n(T) = \lambda_n(\bar{\mathcal{L}}^+ R\Gamma) = \lambda_n(R\Gamma\bar{\mathcal{L}}^+) \leq \sigma_n(R\Gamma\bar{\mathcal{L}}^+)$$
$$= \sigma_n(\Gamma\bar{\mathcal{L}}^+) = \sigma_n((\bar{\mathcal{L}} + \nabla)\bar{\mathcal{L}}^+) \leq 1 + \frac{\delta_n}{\bar{\lambda}_2}, \tag{2.42}$$

where the last two inequalities hold true because of Theorem 6.72(b)(iii) on page 118 and Theorem 6.80 on page 120 [22].

From (2.39) and (2.42), we have $\sigma_2 \leq (1 + \frac{\delta_n}{\bar{\lambda}_2}) \cdot 2h(G)$. In particular, when the graph is undirected, we have $\bar{\mathcal{L}} = \Gamma$ and $1 + \frac{\delta_n}{\bar{\lambda}_2} = 1$, thus, the above upper bound of σ_2 reduces to $\bar{\lambda}_2 \leq 2h(G)$, which subsumes the results obtained in [15].

(ii) Proof for lower bound: The authors in [14, 15] prove that the second smallest eigenvalue $\bar{\lambda}_2$ of the symmetrized Laplacian matrix $\bar{\mathcal{L}}$ can be bounded by Cheeger constant as $\bar{\lambda}_2 \geq \frac{h^2(G)}{2}$. Since for any directed graph $G = (V, E, A)$, $\sigma_i \geq \bar{\lambda}_i$ $(1 \leq i \leq n)$ holds true, based on Theorem 5.11.25 on page 355 [23], we have $\sigma_2 \geq \bar{\lambda}_2 \geq \frac{h^2(G)}{2}$. ∎

Finally, we introduce a generalized Cheeger constant, $\tilde{h}(G)$, defined as

$$\tilde{h}(G) = \min_S \frac{\|\Gamma x_S\|}{\|x_S\|} = \min_{x_S \perp \phi^{\frac{1}{2}}} \frac{(x_S^T \Gamma^T \Gamma x_S)^{\frac{1}{2}}}{(x_S^T x_S)^{\frac{1}{2}}}, \tag{2.43}$$

where for any $S \subset N = \{1, 2, \ldots, n\}$, $x_S = \Phi^{-\frac{1}{2}} f_S$ is defined above. We see that the generalized Cheeger constant thus defined minimizes the 2-norm of the circulations across bipartite subgraphs S and $\bar{S}$, whereas $h(G)$ minimizes the 1-norm (the sum of absolute values) of the circulations across S and $\bar{S}$. Clearly, $\sigma_2 \leq \tilde{h}(G)$.

2.4.2 Bounding the Mixing Rate of Random Walks on Digraphs

In this section, using mixing rating bounds as an example, we show that by considering the degree of asymmetry, we obtain a better bound for the mixing rate of random walks on digraphs.

The mixing rate is a measure of how fast a random walk converges to its stationary distribution. Many papers have studied the problem of bounding the mixing rate of random walks (or reversible Markov chains) on undirected graphs, such as [24, 25]. Relatively few papers [15, 26, 27] have addressed the problem of bounding the mixing rate of Markov chains (or random walks) on digraphs. In bounding the convergence rate from an initial distribution

to the stationary distribution of a Markov Chain with the transition probability matrix P, the χ-*square* distance [15, 26] is commonly used, and is defined as follows:

$$\chi(t) = \max_{i \in V(G)} \left(\sum_{j \in V(G)} \frac{(P^t(i,j) - \phi_j)^2}{\phi_j} \right)^{\frac{1}{2}}. \tag{2.44}$$

Since not all random walks on strongly connected digraphs are convergent, Chung defines a lazy random walk on G with transition probability matrix $P_L = \frac{I+P}{2}$, and derives the following bound on the mixing rate of this lazy random walk using the χ-square distance [15], where a bound using a closely related total variance distance is also derived: define $M = \Phi^{\frac{1}{2}} P_L \Phi^{-\frac{1}{2}}$, then $\chi(t)^2 \leq \varepsilon^t \max_i \phi_i^{-1}$ where

$$\varepsilon = \max_{f \perp \phi^{\frac{1}{2}}} \frac{f^T M^T M f}{f^T f} = \max_{f \perp \phi^{\frac{1}{2}}} \frac{\|Mf\|^2}{\|f\|^2} \leq 1 - \frac{\bar{\lambda}_2}{2}. \tag{2.45}$$

From Theorem 4, we know that this bound leads to the same upper bound for all digraphs with the same $\bar{\mathcal{L}}$. By accounting for the degree of asymmetry, we obtain a generally tighter upper bound on $\frac{\|Mf\|^2}{\|f\|^2}$ as follows, which in turn yields a tighter bound on $\chi(t)$:

Theorem 7 *For an irreducible Markov chain P and a lazy random walk $P_L = \frac{I+P}{2}$,*

$$\varepsilon = \max_{f \perp \phi^{\frac{1}{2}}} \frac{\|Mf\|^2}{\|f\|^2} \leq 1 - \frac{\bar{\lambda}_2}{2} - \frac{1 - \sigma_{n-1}^2(\mathcal{P})}{4} \tag{2.46}$$

$$\leq \frac{3}{4} - \frac{\bar{\lambda}_2}{2} + \frac{(\sigma_{n-1}(\bar{\mathcal{P}}) + \delta_n)^2}{4} \tag{2.47}$$

holds true, where $M = \Phi^{\frac{1}{2}} P_L \Phi^{-\frac{1}{2}}$ and $\bar{\mathcal{P}} = \frac{\mathcal{P}+\mathcal{P}^T}{2} = I - \bar{\mathcal{L}}$.

Proof : $M^T M$ can be rewritten as

$$M^T M = (I - \frac{\Gamma^T}{2})(I - \frac{\Gamma}{2}) = \frac{3}{4} I - \frac{\Gamma + \Gamma^T}{4} + \frac{\mathcal{P}^T \mathcal{P}}{4} = (I - \frac{\bar{\mathcal{L}}}{2}) - \frac{I - \mathcal{P}^T \mathcal{P}}{4}.$$

Since $f \perp \phi^{\frac{1}{2}}$ is the left and right singular vectors of Γ corresponding to the largest singular value 1, and is also the eigenvector of $\bar{\mathcal{L}}$ corresponding to the smallest eigenvalue 0, by applying the Min-max theorem (see Theorem 6.58(e) on p.108 [21]) we prove (2.46). Then $\sigma_{n-1}^2(\mathcal{P})$ in (2.46) can be further bounded as follows:

$$\sigma_{n-1}^2(\mathcal{P}) \leq \sigma_{n-1}^2(\bar{\mathcal{P}} + \nabla) \leq (\sigma_{n-1}(\bar{\mathcal{P}}) + \delta_n)^2, \tag{2.48}$$

which yields (2.47). ∎

We remark that since $\sigma_{n-1}(\mathcal{P}) \leq \sigma_n(\mathcal{P}) = 1$, $1 - \sigma_{n-1}^2(\mathcal{P}) \geq 0$ holds, the bound in (2.46) is tighter than Chung's bound in (2.45). In (2.47) we further bound $\sigma_{n-1}(\mathcal{P})$ in terms of the degree of asymmetry and the singular value of $\bar{\mathcal{P}}$. In particular, when $\mathcal{P}$ is symmetric, the underlying graph is undirected and the bound in (2.47) boils down to (2.46).

2.4.3 Numerical Analysis

In this section, we study a real directed network in terms of its degree of asymmetry δ_n and the mixing rate bound. Fig. 2.1 presents the information flow network with 10 formal organizations concerned with social welfare issues in one mid-western U.S. city, where the data were collected by Knoke and Kuklinski [1, 2].

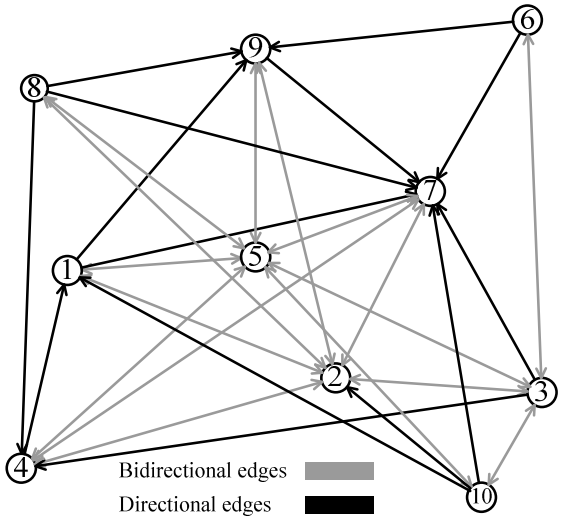


Figure 2.1: Knoke's data [1, 2] on information exchanges among organizations operating in the social welfare field. The degree of asymmetry of this digraph is 0.3096.

It is easy check that Knoke information exchange digraph is strongly connected, where any node can reach any other node in the graph. It has degree of asymmetry $\delta_n = 0.3096$. Using (2.24), we construct a set of new digraphs with α ranging from 0 to 1. As we can see from Fig.2.2, the new graph $G(\alpha)$ becomes more symmetric with smaller δ_n, when the parameter α gets closer to $1/2$. On the other hand, when $|\alpha - \frac{1}{2}|$ is larger, or equivalently α is closer to 0 or 1, the graph becomes more asymmetric, with larger δ_n.

Moreover, as we change α, we compute various mixing rate bounds on ϵ, such as our bounds

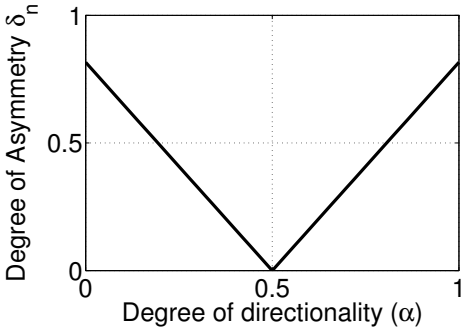


Figure 2.2: Degree of asymmetry

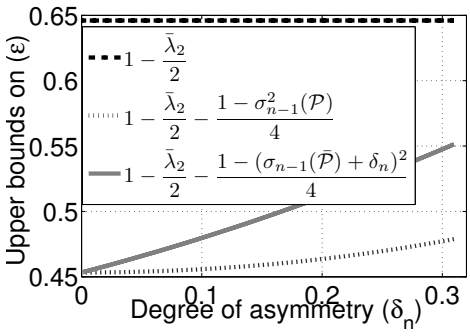


Figure 2.3: Comparing mixing rate bounds

(2.46) and (2.47), and the bound (2.45) obtained in [15]. Fig. 2.3 shows the comparisons between mixing rate bounds over the degree of asymmetry in Knoke information exchange digraph, where (2.45) provides an invariant bound for all digraphs constructed by (2.24), because the construction preserves the stationary distribution and the symmetrized Laplacian matrix $\bar{\mathcal{L}}$. On the other hand, (2.46) and (2.47) bound the random walk mixing rates on those digraphs more precisely, where digraphs that are more symmetric with lower δ_n have lower mixing rate bounds.

Chapter 3

Identifying Stable Clusters in Social Networks with Uni- and Bi-directional Links

3.1 Introduction

Graph models are widely utilized to represent relations among entities in social networks. Especially, many online social networks, e.g., Slashdot and Twitter, where the users' social relationships are represented as directed edges in directed graphs (or in short, *digraphs*). Entity connections in a digraph can be categorized into two types, namely, bi-directional links (mutual connections) and uni-directional links (*one-way* connections). Social theories [28] and online social network analysis [28–30] have revealed that various types of connections exhibit different stabilities, where mutual connections are more stable than one-way connections. In other words, mutual connections are the source of social cohesion [31, 32] that, if two individuals mutually attend to one another, then the bond is reinforced in each direction.

Studying the social network structure and properties of social ties have been an active area of research. Clustering and identifying social structures in social networks is an especially important problem [33–35] that has wide applications, for instance, community detection and friend recommendation in social networks. Existing clustering methods [36, 37] are originally developed for *undirected* graphs, based on the classical *spectral clustering theory*. Several recent

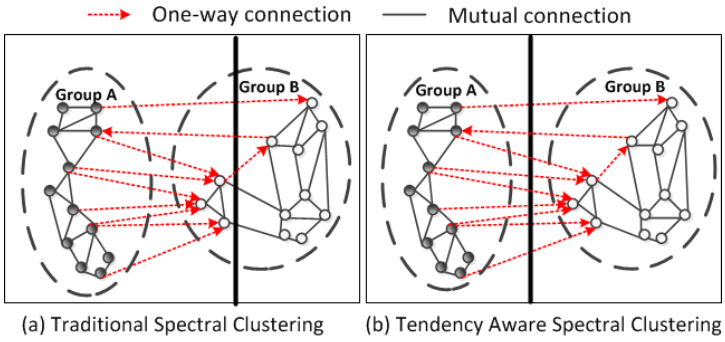


Figure 3.1: An example network

studies (see, e.g., [36–39]) extend the spectral clustering method to digraphs, by first converting the underlying digraphs to undirected graphs via some form of *symmetrization*, and then apply spectral clustering to the resulting symmetrized (undirected) graphs. However, all these methods have two common drawbacks, which prevent them from obtaining *stable* clusters with *more mutual connections*. First, these methods do not explicitly distinguish between *mutual* and *one-way* connections commonly occurring in many social networks, treating them essentially as the same and therefore ignoring the different social relations and interpretations these two types of connections represent (see Section 3.2 for more in-depth discussion). Second, by simply minimizing the total cross-cluster links (that are symmetrized in some fashion), these methods do not explicitly account for the potential tendencies of node pairs to become mutually connected. As a simple example, Fig. 3.1 shows two groups of people in a network, where people in the same group tend to have more mutual (stable) connections, and people across two groups have more one-way (unstable) connections. When using the traditional spectral clustering method, as shown in Fig. 3.1(a), group B will be partitioned into two clusters, due to its strict rule of minimizing the total number of across cluster edges. On the other hand, the correct partition should be done as shown in Fig. 3.1(b), where the majority of mutual (stable) connections are placed within clusters, and one-way (unstable) connections are placed across clusters.

In this chapter, we propose and develop a stable social cluster detection algorithm that takes into account the tendencies of node pairs whether to form mutual (thus stable) connections or not, which can result in more *stable* cluster structures. To tackle this clustering problem, we need to answer the following questions: 1) how to track and evaluate the tendencies of node pairs to become mutual (stable) relations? and 2) how to cluster the entities in social networks by accounting for their mutuality tendencies so as to extract more stable clustering structures?

To address these questions, we utilize dyadic methods to analyze social networks, and develop a generalized mutuality tendency theory which better captures the tendencies of node pairs that tend to establish mutual connections more frequently than those occur by chance. Using these results, we develop a *mutuality-tendency-aware* spectral clustering algorithm to detect more stable clusters by maximizing the *within-cluster* mutuality tendency and minimizing the *cross-cluster* mutuality tendency. Our contributions are summarized as follows.

◇ Motivated by the social science mutuality tendency theory, we establish a new *cluster-based* mutuality tendency theory which yields a symmetrized mutuality tendency for each node pair, and provides a measure of strength of social ties among nodes in a cluster.

◇ Based on our theory, we develop a *mutuality-tendency-aware* spectral clustering algorithm that can partition the social graphs into stable clusters, by maximizing the within-cluster mutuality tendencies and minimizing the across cluster mutuality tendencies.

◇ The experimental results – based on both social network structures of synthetical and real social network datasets – confirm that our clustering algorithm is able to generate more stable clusters than the traditional spectral clustering algorithms.

To our best knowledge, this is the first work studying the impact of tendencies of node pairs to become mutual connections on the stability of cluster structure of social networks. The remainder of the chapter is organized as follows. In section 3.2 we briefly discuss the existing dyadic analysis methods, the traditional spectral clustering algorithms and other related work. In section 3.3 we introduce a cluster-base mutuality tendency theory, and based on this theory, we develop a mutuality-tendency-aware spectral clustering algorithm in section 3.4. In section 3.5, we evaluate the performances of our method using synthetic and real social network (e.g., Slashdot) datasets. We summarize the chapter in section 3.6.

3.2 Preliminaries, Related Work and Problem definition

In this section, we first introduce the existing dyadic analysis methods in the social theory literature for analyzing and characterizing social network mutual connections and one-way connections. We then present the classic spectral clustering theory which was developed for *undirected* graphs, and briefly survey some related works which apply this theory to *digraphs* through *symmetrization*. We argue that these existing methods for clustering digraphs via symmetrization

are inadequate in solving social network clustering problems, as they ignore different social ties (and mutuality tendencies) represented by mutual and one-way connections in social networks. We end the section with the problem definition, namely, how to identify *stable* clusters in social networks by taking into account mutuality tendencies of mutual and one-way connections.

3.2.1 Dyadic Analysis and Mutuality Tendency

Given a social network with both uni- and bi-directional links, such a network can be represented as a (simple) digraph $G = (V, E)$ with $|V| = n$ nodes. If the links also have weights (say, representing the strength of connections or social ties), such a network can be more generally represented as a *weighted* digraph, $G = (V, E, A)$ where A_{ij} represents the strength of connection or "affinity" from node i to node j. When A is a 0-1 matrix, G reduces to a simple digraph, and A is the standard adjacency matrix of the digraph, where $A_{ij} = 1$ if the directed edge $i \to j$ is present, and $A_{ij} = 0$ otherwise. In this chapter for simplicity we focus primarily on simple (unweighted) digraphs with no selfloops, namely, social networks with unweighted directional links. Most online social networks are of unweighted variety.

Social scientists commonly view the social network G as a collection of dyads [28], where *a dyad is an* unordered *pair of nodes and* directed *edges between two nodes in the pair.* Denote a dyad as $Dy_{ij} = (A_{ij}, A_{ji})$, for $i < j$. Since dyad is an unordered notion, we have in total $N_d = n(n-1)/2$ dyads in G. Hence, there are only three possible isomorphism dyads. The first type of dyads is *mutual* relationship, where both directional edges $i \to j$ and $j \to i$ are present. The second type of dyads is *one-way* relationship, where either $i \to j$ or $j \to i$ is present, but not both. The last type of dyads is *null* relationship, where no edges show up between i and j.

Interpretations of dyads. Social scientists have observed that mutual social relations and one-way relations in social networks typically exhibit different stabilities, namely, mutual relations are more stable than one-way relations [28]. Hence in the social science literature, one prevalent interpretation of dyadic relations in social networks are the following: mutual dyads are considered as stable connections between two nodes and null relation dyads represent no relations; the one-way dyads [40–44] are viewed as an *intermediate* state of relations, which are in transition to more stable equilibrium states of reciprocity (mutual or no relation). Several recent empirical studies [30,45] of online social networks have further revealed and confirmed that mutual social relations are more stable relations than one-way connections.

Computing dyad census. Given a (simple) digraph $G = (V, E)$, with $n = |V|$ nodes. Let

m, b, and u denote the number of mutual, one-way, and null dyads in the network. Clearly, $m + b + u = n(n-1)/2$. The triple $\langle m, b, u\rangle$ is referred to as the *dyad census*, since it is derived from an examination of all (possible) dyads in the network. The dyad census triple can be computed in terms of the adjacency matrix A of G as follows (in both scalar and matrix forms):

$$\begin{aligned} m &= \sum_{i<j} A_{ij}A_{ji} = \frac{1}{2}\mathbf{tr}(AA), \\ b &= |E| - 2m = \mathbf{tr}(AA^T) - \mathbf{tr}(AA), \\ u &= N_d - b - m = N_d - \mathbf{tr}(AA^T) + \mathbf{tr}(AA). \end{aligned}$$

Measuring mutuality tendency. The notion of mutuality tendency has been introduced in the social science literature (see, e.g., [28,46]) to measure the tendency for a node pair to establish mutual connections. For any dyad between i and j in a digraph G, if i places a link to j, ρ_{ij} represents the tendency that j will reciprocate to i more frequently than would occur by chance.

Let $\mathbf{X}_{ij}$ denote the random variable that represents whether or not node i places a directed edge to node j. There are only two possible events (i.e., $\mathbf{X}_{ij}$ takes two possible values): $\mathbf{X}_{ij} = 1$, representing the edge is present; or $\mathbf{X}_{ij} = 0$, the edge is not present. Let X_{ij} (resp. $\bar{X}_{ij}$) denote the event $\{\mathbf{X}_{ij} = 1\}$ (resp. $\{\mathbf{X}_{ij} = 0\}$). Then the probability of the event X_{ij} occurring is $P(X_{ij})$. The probability that i places a directed edge to j and j reciprocates back (i.e., node i and node j are mutually connected) is thus given by

$$P(X_{ij}, X_{ji}) = P(X_{ij})P(X_{ji}|X_{ij}), \tag{3.1}$$

Wofle [28] introduces the following measure of mutuality tendency in terms of the conditional probability $P(X_{ji}|X_{ij})$ as follows:

$$\begin{aligned} P(X_{ji}|X_{ij}) &= P(X_{ji}) + \rho_{ij}P(\bar{X}_{ji}), \\ \rho_{ij} &= \frac{P(X_{ij}, X_{ji}) - P(X_{ij})P(X_{ji})}{P(X_{ij})P(\bar{X}_{ji})}, \end{aligned} \tag{3.2}$$

where $-\infty < \rho_{ij} \leq 1$ ensures $0 \leq P(X_{ji}) + \rho P(\bar{X}_{ji}) \leq 1$ to hold. Like many indices used in statistics, ρ_{ij} is dimensionless and easy to interpret, since it uses 0 and 1 as benchmarks. If $\rho = 1$, the mutuality tendency is maximum, meaning that given that node i places a link to node j, node j will for sure reciprocate. If $\rho_{ij} = 0$ (i.e., $P(X_{ji}|X_{ij}) = P(X_{ji})$), then node

j reciprocates and places a link to node i purely by chance, namely, it is independent of the event that node i places a link to node j. Hence when $0 < \rho_{ij} \leq 1$, it suggests more than a chance tendency for node j to reciprocate back. Furthermore, if $\rho_{ij} < 0$ (i.e., $P(X_{ji}|X_{ij}) < P(X_{ji})$), there is less than chance tendency for node j to reciprocate; in other words, it suggests a tendency away from mutual dyads, toward one-way and null dyads. Hence, $-\infty < \rho_{ij} \leq 1$ provides a measure of the strength of tendency for reciprocation.

From eq.(3.2), the joint distribution $P(X_{ij}, X_{ji})$ in eq.(3.2) can be measured by the observed graph, namely, either $P(X_{ij}, X_{ji}) = P^{(\omega)}(X_{ij}, X_{ji}) = 1$, when i and j have mutual connection, or $P(X_{ij}, X_{ji}) = P^{(\omega)}(X_{ij}, X_{ji}) = 0$, otherwise, where the superscript ω indicates that the probability is obtained from the observed graph. On the other hand, the distribution for each individual edge is measured by $P(X_{ij}) = P^{(\mu)}(X_{ij}) = \frac{d_i}{|V|-1}$, where d_i is the out-going degree of node i. $P^{(\mu)}(X_{ij})$ represents the probability of edge $i \to j$ being generated under a random graph model, denoted by the superscript μ, with edges randomly generated while preserving the out-degrees. Hence, the tendency ρ_{ij} is obtained by implicitly comparing the observed graph with a reference random digraph model.

Limitations of Wolfe's mutuality tendency measure for stable social structure clustering. Although the node pair in a dyad is unordered (i.e., the two nodes are treated "symmetrically" in terms of dyadic relations), Wolfe's measure of mutual tendency is in fact *asymmetric*. This can be easily seen through the following derivation. By definition,

$$P(X_{ji}|X_{ij}) = P(X_{ji}) + \rho_{ji}P(\bar{X}_{ji}),$$
$$P(X_{ij}|X_{ji}) = P(X_{ij}) + \rho_{ij}P(\bar{X}_{ij}).$$

Multiplying the above two equations with $P(X_{ij})$ and $P(X_{ji})$ respectively and from eq.(3.1), we have

$$\frac{\rho_{ji}}{\rho_{ij}} = \frac{P(X_{ji})P(\bar{X}_{ij})}{P(X_{ij})P(\bar{X}_{ji})} = \frac{P(X_{ji}) - P(X_{ij})P(X_{ji})}{P(X_{ij}) - P(X_{ij})P(X_{ji})}$$

We see that $\rho_{ij} = \rho_{ji}$ if and only if $P(X_{ij}) = P(X_{ji})$ holds. Hence, given an arbitrary dyad in a social network Wolfe's measure of mutuality tendency of the node pair is asymmetric – in a sense that it is a *node-specific* measure of mutuality tendency. It does not provide a measure of mutuality tendency of the (unordered) *node pair* viewed together. While such asymmetric (*node-specific*) measure of mutuality tendency can be useful in some social network analysis,

as will be clear later, such an asymmetric measure poses difficulty in identifying and extracting *stable* cluster structures in social networks. For instance, given a partition $V = (S, \bar{S})$ of a digraph, generalizing Wolfe's measure to clusters, the mutuality tendencies *across* the two clusters, denoted by $\rho(S, \bar{S}) = \sum_{i \in S, j \in \bar{S}} \rho_{ij}$ and $\rho(\bar{S}, S) = \sum_{i \in \bar{S}, j \in S} \rho_{ij}$, are generally not symmetric, namely, $\rho(S, \bar{S}) \neq \rho(\bar{S}, S)$. In Section 3.3, we will introduce a new measure of mutuality tendency that is *symmetric* and captures the tendency of a node pair in a dyadic relation to establish mutual connection. This measure of mutuality tendency can be applied to clusters and a whole network in a straightforward fashion, and leads us to develop a *mutuality-tendency-aware* spectral clustering algorithm.

3.2.2 Spectral Clustering Theory and Extensions to Digraphs via Symmetrization

Spectral clustering methods (see, e.g., [36, 38, 47–49]) are originally developed for clustering data with symmetric relations, namely, data that can be represented as *undirected* graphs, where each relation (edge) between two entities, $A_{ij} = A_{ji}$, represents their similarity. The goal is to partition the graph such that entities within each cluster are more similar to each other than those across clusters. This is done by minimizing the total weight of cross-cluster edges (possibly weighted by the total weight of edges within clusters). In the following we present the basics of spectral clustering theory (see [50] for more details).

Given the (non-negative) similarity matrix A, the cut function is defined to quantatively measure the quality of a partition $V = (S_1, \cdots, S_K)$, and is defined as follows:

$$Cut(S_l, \bar{S}_l) := \sum_{i \in S_l, j \in \bar{S}_l} A_{ij},$$

$$Cut(S_1, \cdots, S_K) := \sum_{i=1}^{K} Cut(S_i, \bar{S}_i).$$

To account for cluster sizes – especially to obtain relatively balanced clusters (in terms of sizes), the ratio cut function $RCut$ [51] and the normalized cut function $NCut$ [48] have also been

defined:

$$RCut(S_1, \cdots, S_K) := \sum_{i=1}^{K} \frac{Cut(S_i, \bar{S}_i)}{|S_i|},$$
$$NCut(S_1, \cdots, S_K) := \sum_{i=1}^{K} \frac{Cut(S_i, \bar{S}_i)}{Vol(S_i)},$$

where $vol(S_i) = \sum_{j \in S_i} d_j$ is the volume of the cluster S_i.

In the following (and the remainder of the chapter), we will use the ratio cut function as the objective function. All the results also hold true for the normalized cut. Using the ratio cut, the clustering problem formulated as a graph mincut optimization problem can be rewritten in the following form:

$$\min_{S_1, \cdots, S_k} RCut(S_1, \cdots, S_K), \tag{3.3}$$

The (unnormalized) Laplacian matrix $L = D - A$ is used to solve the above mincut problem, where $D = diag[d_i]$ with $d_i = \sum_j A_{ij}$ is the diagonal degree matrix. Given a (nonnegative) symmetric A, L is symmetric and positive semi-definite. If we take K eigenvectors corresponding to the smallest eigenvalues of L, the optimal solution to the problem eq.(3.3), namely, the optimal partition into K clusters, can be well approximated by applying the K-means algorithm to clustering the data points projected to the subspace formed by these K eigenvectors [50]. Moreover, [**?**] provides a systematic study on comparing a wide range of undirected graph based clustering algorithms using real large datasets, which gives a nice guideline of how to select clustering algorithms based on the underlying networks and the targeting objectives.

Extensions to digraphs via symmetrization. When relations between entities are *asymmetric*, or the underlying graph is *directed*, spectral clustering cannot be directly applied, as the notion of (semi-)definiteness is only defined for *symmetric* matrices. Several recent studies (see, e.g., [36–39]) all attempt to circumvent this difficulty by first converting the underlying digraphs to undirected graphs via some form of *symmetrization*, and then apply spectral clustering to the resulting symmetrized (undirected) graphs. For example, the authors in [37] discuss several symmetrization methods, including the symmetrized adjacency matrix $\bar{A} = (A + A^T)/2$, the bibliographic coupling matrix AA^T and the co-citation strength matrix $A^T A$, and so forth. Symmetrization can also be done through a random walk on the underlying graph, where $P = D^{-1}A$ is the probability transition matrix and $D = diag[d_i^{out}]$ is a diagonal matrix of

node out-degrees. For example, taking the objective function as the random walk flow circulation matrix $F_\pi = \Pi P$, where Π is the diagonal stationary distribution matrix, we have the symmetrized Laplacian of the circulation matrix as

$$\bar{\mathcal{L}} = \frac{\tilde{\mathcal{L}} + \tilde{\mathcal{L}}^T}{2} = I - \frac{\Pi^{\frac{1}{2}} P \Pi^{-\frac{1}{2}} + \Pi^{-\frac{1}{2}} P^T \Pi^{\frac{1}{2}}}{2}.$$

where $\tilde{\mathcal{L}}$ is the (asymmetric) digraph Laplacian matrix [3]. Then the classical spectral clustering algorithm can then be applied using $\bar{\mathcal{L}}$ which is symmetric and semi-definite. Zhou and et al [36, 38] use this type of symmetrization to perform clustering on digraphs. Moreover, Leicht and Newman [39] propose the digraph modularity matrix $Q = [Q_{ij}]$, which captures the difference between the observed digraph and the hypothetical random graph with edges randomly generated by preserving the in- and out-degrees of nodes, namely, $Q_{ij} = A_{ij} - d_i^{out} d_j^{in}/m$. Then, if the sum of edge modularities in a cluster S is large, nodes in S are well connected, since the edges in S tend to appear with higher probabilities than occur by chance. However, Q by definition is asymmetric, where [39] uses the symmetrized $\bar{Q} = (Q + Q^T)/2$ as objective to perform spectral clustering method. Essentially, the edge modularity captures how an individual edge appears more frequently than that happens by chance, thus the modularity based clustering method tends to group those nodes with more connections than expected together, which like all other clustering methods presented above completely ignores the distinction between mutual and one-way connections.

Problem definition: Clustering and identifying stable clusters in social networks with mutual and one-way connections. As discussed earlier, one-way and mutual dyadic connections in social networks often represent different states or types of social ties and exhibit various stabilities over time. Hence when performing clustering to extract community structures in social networks, one-way and mutual connections should be distinguished and treated differently. Existing digraph clustering methods via symmetrization, e.g., those mentioned above, on the other hand, ignore these different types of connections and treat them as the same: the process of symmetrization essentially weighs one-way connections as a fraction of mutual connections, and then attempt to minimize the total weight of the (symmetrized) cross-cluster links. Moreover, different from Leicht and Newman's [39] reference random graph model, as presented in earlier section, the mutuality tendency compares the observed the digraph with a random graph model where edges are randomly generated by preserving only the out-degrees, which better reflects the underlying model of how social network users establish social ties.

In this chapter we want to solve the following clustering problem in social networks with bi- and uni-directional links: Given a directed (social) graph where mutual connections represent more stable relations and one-way connections represent intermediate transferring states, *how can we account for mutual tendencies of dyadic relations and cluster the entities in such a way that nodes within each cluster have maximized mutuality tendencies to establish mutual connections, while across clusters, nodes have minimized tendencies to establish mutual connections?* The clusters (representing social structures or communities) identified and extracted thereof will hence likely be more stable.

3.3 Cluster-based Mutuality Tendency Theory

Inspired by Wolfe's study in [28], we propose a new measure of mutuality tendency for dyads that can be generalized to groups of nodes (clusters), and develop a *mutuality tendency theory* for characterizing the strength of social ties within a cluster (network structure) as well as across clusters in an asymmetric social graph. This theory lays the theoretical foundation for the network structure classification and community detection algorithms we will develop in section 3.4.

3.3.1 Cluster based mutuality tendency

Let $\mathbf{X}_{ij}$ denote the random variable that represents whether or not node i places a directed edge to node j. There are only two possible events (i.e., $\mathbf{X}_{ij}$ takes two possible values): $\mathbf{X}_{ij} = 1$, representing the edge is present; or $\mathbf{X}_{ij} = 0$, the edge is not present. Let X_{ij} (resp. $\bar{X}_{ij}$) denote the event $\{\mathbf{X}_{ij} = 1\}$ (resp. $\{\mathbf{X}_{ij} = 0\}$). Given an *observed* (asymmetric) social graph G, to capture the *mutuality tendency* of dyads in this graph, we compare it with a *hypothetical*, *random* (social) graph, denoted as $G^{(\mu)}$, where links (dyadic relations) are generated randomly (i.e., by chance) in such a manner that the (out-)degree d_i of each node i in $G^{(\mu)}$ is the same as that in the observed social graph G. Under this random social graph model, the probability of the event X_{ij} occurring is $P^{(\mu)}(X_{ij}) = \frac{d_i}{|V|-1}$; namely, i places a (directed) link to node j randomly or by chance (the superscript μ indicates the probability distribution of link generations under the random social graph model). The probability that i places a directed edge to j and j reciprocates back (i.e., node i and node j are mutually connected) is thus given by $P^{(\mu)}(X_{ij}, X_{ji}) = P^{(\mu)}(X_{ij})P^{(\mu)}(X_{ji}|X_{ij}) = P^{(\mu)}(X_{ij})P^{(\mu)}(X_{ji})$, since $\mathbf{X}_{ij}$

and $\mathbf{X}_{ji}$ are independent under the random social graph model. On the observed social graph, denote $P^{(\omega)}(X_{ij}, X_{ji})$ to represent the event whether there is a mutual connection (symmetric link) between node i and node j, i.e., $P^{(\omega)}(X_{ij}, X_{ji}) = 1$, if the dyad Dy_{ij} is a mutual dyad in the *observed* social graph, and $P^{(\omega)}(X_{ij}, X_{ji}) = 0$, otherwise. We define the *mutuality tendency* of dyad Dy_{ij} as follows:

$$\begin{aligned}\theta_{ij} :&= P^{(\omega)}(X_{ij}, X_{ji}) - P^{(\mu)}(X_{ij}, X_{ji}) \\ &= P^{(\omega)}(X_{ij}, X_{ji}) - P^{(\mu)}(X_{ij})P^{(\mu)}(X_{ji}),\end{aligned} \tag{3.4}$$

which captures how the node pair i and j establish a mutual dyad more frequently than would occur by chance.

This definition of mutuality tendency is a symmetric measure for dyad Dy_{ij}, i.e., $\theta_{ij} = \theta_{ji}$. In addition, it is shown that $\theta_{ij} \in [-1, 1]$. We remark that $\theta_{ij} = 0$ indicates that if node i places a directed link to node j, the tendency that node j will reciprocate back to node i is no more likely than would occur by chance; the same holds true if node j places a directed link to node i instead. On the other hand, $\theta_{ij} > 0$ indicates that if node i (resp. node j) places a directed link to node j (resp. node i), node j (resp. node i) will more likely than by chance to reciprocate. In particular, with $\theta_{ij} = 1$, node j (resp. node i) will almost surely reciprocate. In contrast, $\theta_{ij} < 0$ indicates that if node i (resp. node j) places a directed link to node j (resp. node i), node j (resp. node i) will tend not to reciprocate back to node i (resp. node j). In particular, with $\theta_{ij} = -1$, node j (resp. node i) will almost surely not reciprocate back. Hence θ_{ij} provides a measure of strength of social ties between node i and j: $\theta_{ij} > 0$ suggests that the dyadic relation between node i and j is stronger, having a higher tendency (than by chance) to become mutual; whereas $\theta_{ij} < 0$ suggests that node i and j have weaker social ties, and their dyadic relation is likely to remain asymmetric or eventually disappear.

Mutuality tendency of clusters. The mutuality tendency measure for dyads defined in eq.(3.4) can be easily generalized for an arbitrary cluster (a subgraph) in an observed social graph, $S \subseteq G$. We define the mutuality tendency of a cluster S, Θ_S, as follows:

$$\begin{aligned}\Theta_S :&= \sum_{i\sim j; i,j\in S} P^{(\omega)}(X_{ij}, X_{ji}) - \sum_{i\sim j; i,j\in S} P^{(\mu)}(X_{ij}, X_{ji}) \\ &= \sum_{i\sim j; i,j\in S} P^{(\omega)}(X_{ij}, X_{ji}) - \sum_{i\sim j; i,j\in S} P^{(\mu)}(X_{ij})P^{(\mu)}(X_{ji}),\end{aligned} \tag{3.5}$$

where the subscript $i \sim j : i, j \in S$ means that the summation accounts for all (unordered) dyads, and i, j are both in S. Denote the second term in eq.(3.5) as $m_S^{(\mu)}$, and the (out-degree) volume of the cluster S as $d_S := \sum_{i \in S} d_i$. As $P^{(\mu)}(X_{ij}) = d_i/(|V| - 1)$ and $P^{(\mu)}(X_{ji}) = d_j/(|V| - 1)$,

$$m_S^{(\mu)} = \sum_{i \sim j; i,j \in S} \frac{d_i d_j}{(|V| - 1)^2} = \frac{d_S^2 - \sum_{i \in S} d_i^2}{2(|V| - 1)^2}, \tag{3.6}$$

which represents the expected number of mutual connections among nodes in S *under the random social graph model*. Given the cluster S in the observed social graph G, define $m_S^{(\omega)} := \sum_{i \sim j; i,j \in S} P^{(\omega)}(X_{ij}, X_{ji})$, namely, $m_S^{(\omega)}$ represents the number of (observed) mutual connections among nodes in the cluster S in the observed social graph G. The mutual tendency of cluster S defined in eq.(3.5) is therefore exactly $\Theta_S = m_S^{(\omega)} - m_S^{(\mu)}$.

Hence Θ_S provides a measure of strength of (likely mutual) social ties among nodes in a cluster: $\Theta_S > 0$ suggests that there are more mutual connections among nodes in S than would occur by chance; whereas $\Theta_S < 0$ suggests that there are fewer mutual connections among nodes in S than would occur by chance. Using Θ_S, we can therefore quantify and detect clusters of nodes (network structures or communities) that have strong social ties.

In particular, when $S = G$, Θ_G characterizes the mutuality tendency for the entire digraph G, i.e.,

$$\Theta_G = m_G^{(\omega)} - m_G^{(\mu)} = \sum_{i \sim j} \theta_{ij}, \tag{3.7}$$

where $m_G^{(\omega)} := \sum_{i \sim j} P^{(\omega)}(X_{ij}, X_{ji})$ represents the number of (observed) mutual dyads among nodes in the observed social graph G, and

$$m_G^{(\mu)} = \sum_{i \sim j} \frac{d_i d_j}{(|V| - 1)^2} = \frac{d^2 - \sum_{i \in V} d_i^2}{2(|V| - 1)^2}, \tag{3.8}$$

represents the expected number of mutual dyads among nodes in G *under the random social graph model*. Likewise, given a bipartition $(S, \bar{S})$ of G, we define the cross-cluster mutuality tendency as

$$\Theta_{\partial S} := \sum_{i \in S \sim j \in \bar{S}} (P^{(\omega)}(X_{ij} X_{ji}) - P^{(\mu)}(X_{ij}) P^{(\mu)}(X_{ji})) \tag{3.9}$$

Denote the second quantity in eq.(3.9) as $m_S^{(\mu)}$,

$$m_{\partial S}^{(\mu)} = \sum_{i \in S \sim j \in \bar{S}} \frac{d_i d_j}{(|V|-1)^2} = \frac{d_S d_{\bar{S}}}{(|V|-1)^2} \tag{3.10}$$

which represents the expected number of mutual connections among nodes across S and $\bar{S}$ under the random social graph model. Define $m_{\partial S}^{(\omega)} := \sum_{i \in S \sim j \in \bar{S}} P^{(\omega)}(X_{ij}, X_{ji})$ representing the number of (observed) mutual connections among nodes across clusters S and $\bar{S}$ in the observed social graph G. The mutuality tendency across cluster S and $\bar{S}$ defined in eq.(3.9) is therefore exactly $\Theta_{\partial S} = m_{\partial S}^{(\omega)} - m_{\partial S}^{(\mu)}$.

The mutuality tendency theory outlined above accounts for different interpretations and roles mutual and one-way connections represent and play in asymmetric social graphs, with the emphasis in particular on the importance of mutual connections in forming and developing stable social structures/communities with strong social ties. In the next section, we will show how we can apply this mutuality tendency theory for detecting and clustering stable network structures and communities in asymmetric social graphs.

3.4 Mutuality-tendency-aware spectral clustering algorithm

In this section, we first consider the simpler case of mutuality-tendency-aware clustering problem with $K = 2$ and establish the basic theory and algorithm. We then extend it to the general case with $K > 2$.

3.4.1 Mutuality-tendency-aware spectral clustering: K=2

Without loss of generality, we consider only simple (unweighted) digraphs $G = (V, E)$ (i.e., the adjacency matrix A is a 0-1 matrix). Define the mutual connection matrix $M := \min(A, A^T)$, which expresses all the mutual connections with unit weight 1. In other words, if node i and node j are mutually connected (with bidirectional links), $M_{ij} = M_{ji} = 1$, otherwise, $M_{ij} = M_{ji} = 0$. Hence, we have $M_{ij} = P^{(\omega)}(X_{ij}, X_{ji})$, representing the event whether there is a mutual connection (symmetric link) between node i and node j, i.e., in the dyad Dy_{ij} in the observed social graph. In addition, let δ_{ij} be the Kronecker delta symbol, i.e., $\delta_{ij} = 1$ if $i = j$,

and $\delta_{ij} = 0$ otherwise. Then, we define matrix

$$\bar{M} = \frac{dd^T - diag[d^2]}{(|V|-1)^2}$$

with d as the out-going degree vector, where each entry

$$\bar{M}_{ij} = \frac{d_i d_j - \delta_{ij} d_i^2}{(|V|-1)^2} = \begin{cases} \frac{d_i d_j}{(|V|-1)^2} & \text{if } i \neq j \\ 0 & \text{if } i = j \end{cases} \tag{3.11}$$

represents the probability that two nodes i and j independently place two unidirectional links to each other to form a mutual dyad. Hence, $\bar{M}_{ij} = P^{(\mu)}(X_{ij})P^{(\mu)}(X_{ji})$ represents the probability of node pair i and j to establish a mutual connection under random graph model with edges randomly generated by preserving the node out-degrees. We denote $T = M - \bar{M}$ as the mutuality tendency matrix, with each entry

$$T_{ij} = P^{(\omega)}(X_{ij}, X_{ji}) - P^{(\mu)}(X_{ij})P^{(\mu)}(X_{ji}) = \theta_{ij} \tag{3.12}$$

as the individual dyad mutuality tendency.

Mutuality Tendency Lapacian. T is symmetric and those entries associated with non-mutual dyads are negative, representing less mutuality tendencies to establish mutual connections than those occur by chance. Define the mutuality tendency Laplacian matrix as

$$L_T = D_T - T \tag{3.13}$$

where $D_T = diag[d_T(i)]$ is the diagonal degree matrix of T, with $d_T(i) = \sum_j T_{ij}$. We have the following theorem presenting several properties of L_T.

Theorem 8 *The mutuality tendency Laplacian matrix L_T as defined in eq.(3.13) has the following properties*

- *Given a column vector $x \in \mathbb{R}^{|V|}$, the bilinear form $x^T L_T x$ satisfies*

$$x^T L_T x = \sum_{i \sim j} T_{ij}(x_i - x_j)^2. \tag{3.14}$$

- *L_T is symmetric and in general indefinite. In addition, L_T has one eigenvalue equal to 0, with corresponding eigenvector as $\mathbf{1} = [1, \cdots, 1]^T$.*

Proof : (1) By expanding the bilinear form $x^T L_T x$,

$$x^T L_T x = \sum_{i,j \in V} T_{ij}(x_i^2 - x_i x_j) = \sum_{i \sim j} T_{ij}(x_i - x_j)^2$$

(2) The symmetry of both M and $\bar{M}$ in eq.(3.12) insures the symmetry of L_T, thus L_T has all real eigenvalues. However, L_T is in general indefinite, because T_{ij} in eq.(3.14) could be either positive or negative. On the other hand, since $\mathbf{1}^T L_T = \mathbf{0}^T$ and $L_T \mathbf{1}^T = \mathbf{0}$ hold true, L_T has an eigenvalue equal to 0 with corresponding eigenvectors as identity vector $\mathbf{1} = [1, \cdots, 1]^T$. ∎

Mutuality tendency ratio cut function. For a digraph $G = (V, E)$, and a partition $V = (S, \bar{S})$ on G, we define the *mutuality tendency ratio cut function* as follows.

$$TRCut(S, \bar{S}) = \Theta_{\partial S} \left(\frac{1}{|S|} + \frac{1}{|\bar{S}|} \right), \tag{3.15}$$

which represents the overall mutuality tendency across clusters balanced by the "sizes" of the clusters. Then, the clustering problem is formulated as a minimization problem with $K = 2$ clusters. (More general cases with $|V| \geq K > 2$ will be discussed in the next subsection.)

$$\min_S TRCut(S, \bar{S}) \tag{3.16}$$

Since $\Theta_{\partial S} = \Theta_G - (\Theta_S + \Theta_{\bar{S}})$ holds true, we have

$$TRCut(S, \bar{S}) = (\Theta_G - (\Theta_S + \Theta_{\bar{S}})) \left(\frac{1}{|S|} + \frac{1}{|\bar{S}|} \right).$$

For a given graph G, the graph mutuality tendency Θ_G is a constant, the minimization problem in eq.(5.6) is equivalent to the following maximization problem:

$$\max_S \left\{ (\Theta_S + \Theta_{\bar{S}} - \Theta_G) \left(\frac{1}{|S|} + \frac{1}{|\bar{S}|} \right) \right\}. \tag{3.17}$$

Hence, minimizing the cross-cluster mutuality tendency is equivalent to maximize the within-cluster mutuality tendency. Using the results presented in Theorem 8, we prove the following theorem which provides the solution to the above mutuality tendency optimization problem.

Theorem 9 *Given the tendency Laplacian matrix $L_T = D_T - T$, the signs of the eigenvector of L_T corresponding to the smallest non-zero eigenvalue indicate the optimal solution $(S, \bar{S})$ to the optimization problem eq.(5.6).*

Proof : Define the column vector $f_S = [f_S(1), \cdots, f_S(n)]^T$ with respect to a partition $S \cup \bar{S} = V$ as follows:

$$f_S(i) = \begin{cases} \sqrt{|\bar{S}|/|S|} & \text{if } i \in S \\ -\sqrt{|S|/|\bar{S}|} & \text{if } i \in \bar{S} \end{cases} . \tag{3.18}$$

Then, by applying Theorem 8, we have

$$\begin{aligned} f_S^T L_T f_S &= \sum_{i \sim j} T_{ij} (f_S(i) - f_S(j))^2 \\ &= (\frac{|S|}{|\bar{S}|} + \frac{|\bar{S}|}{|S|} + 2) \sum_{i \in S \sim j \in \bar{S}} T_{ij} \\ &= |V|(m_{\partial S}^{(\omega)} - m_{\partial S}^{(\mu)})(\frac{1}{|\bar{S}|} + \frac{1}{|S|}), \\ &= |V| \Theta_{\partial S} (\frac{1}{|\bar{S}|} + \frac{1}{|S|}). \end{aligned} \tag{3.19}$$

In addition, we have $f_S^T f_S = \|f_S\|^2 = |V|$. Hence, Rayleigh-quoient for L_T is

$$\frac{f_S^T L_T f_S}{f_S^T f_S} = TRCut(S, \bar{S}) \geq \lambda(L_T),$$

where $\lambda(L_T)$ is the smallest non-zero eigenvalue of L_T. Here $\lambda(L_T)$ cannot be 0, because we have the constraint $f_S \perp \mathbf{1}$. From Theorem 8, $\mathbf{1}$ is an eigenvector associated with eigenvalue 0. Hence, the problem of minimizing eq.(5.6) can be equivalently rewritten as

$$\min_S f_S^T L_T f_S, \text{ s.t.: } f_S \perp \mathbf{1} \text{ in form of eq. (3.18)}, \|f_S\|^2 = |V|.$$

Since the entries of the solution vector f_S are only allowed to take values in form of eq.(3.18), this is a discrete optimization problem, which is known to be NP hard [50]. By relaxing the discreteness condition and allowing $f_S(i)$ to take arbitrary values in $\mathbb{R}$, we have the following relaxed optimization problem.

$$\min_{f_S \in \mathbb{R}^n} f_S^T L_T f_S, \qquad \text{s.t.: } f_S \perp \mathbf{1}, \text{ and } \|f_S\|^2 = |V|.$$

The solution to this problem, i.e., the vector f_S, is the eigenvector corresponding to the smallest non-zero eigenvalue $\lambda(L_T)$. Hence, we can approximate the minimizer of $TRCut(S, \bar{S})$ using the eigenvector corresponding to $\lambda(L_T)$. To obtain a partition of the graph, we need to convert the real-valued solution vector f_S of the relaxed problem to an indicator vector. One way to do this [50] is to use the signs of f_S as indicator function, where node $v_i \in S$, if $f_S(i) \geq 0$, and $v_i \in \bar{S}$, otherwise. ∎

3.4.2 Mutuality-tendency-aware spectral clustering: $K > 2$

For the case of finding $K > 2$ clusters $S_1 \cup \cdots \cup S_K = V$, we define the indicator vectors $h_k = (h_{1k}, \cdots, h_{nk})$,

$$h_{ik} = \begin{cases} \frac{1}{\sqrt{|S_k|}} & \text{if } v_i \in S_k \\ 0 & \text{otherwise} \end{cases} \tag{3.20}$$

where $i = 1, \cdots, n$ and $k = 1, \cdots, K$. Let H denote the indicator matrix containing those K indicator vectors as columns. Observe that $H^T H = I$, $h_k^T h_k = 1$, and

$$h_k^T L_T h_k = \frac{\Theta_{\partial S_k}}{|S_k|}. \tag{3.21}$$

Define the mutuality tendency ratio cut $TRCut(S_1, \cdots, S_K)$ for $K > 2$ clusters as follows:

$$TRCut(S_1, \cdots, S_K) = \sum_{k=1}^{K} \frac{\Theta_{\partial S_k}}{|S_k|}, \tag{3.22}$$

where the ratio cut reduces to eq.(3.15) when $K = 2$. The problem of minimizing $TRCut$ can be formulated as

$$\min_{S_1, \cdots, S_K} TRCut(S_1, \cdots, S_K) = \min_{S_1, \cdots, S_K} tr(H^T L_T H)$$
$$\text{s.t.: } H^T H = I, \text{ where } H \text{ is defined in eq. (3.21).}$$

One way of solving this problem is utilizing the method used in [50] by relaxing the discreteness condition to have a standard trace minimization problem as

$$\min_{H \in \mathbb{R}^{|V| \times K}} tr(H^T L_T H), \qquad \text{s.t.: } H^T H = I$$

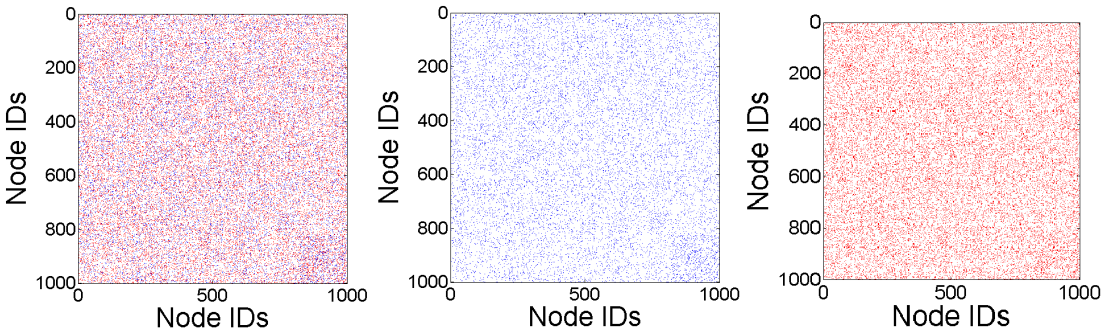
The optimal solution H contains the first K eigenvectors of L_T as columns. The clusters can be then obtained by applying the K-means algorithm on those K eigenvectors. The solution obtained minimizes the mutuality tendency across clusters (which is equivalent to maximizing the within-cluster mutuality tendency).

Choice of K. We choose K, i.e., the total number of clusters, using the eigengap heuristic [50]. Theorem 8 shows that L_T has all real eigenvalues. Denote the eigenvalues of L_T in an increasing order, i.e., $\lambda_1 \leq \cdots \leq \lambda_n$, The index of the largest eigengap, namely, $K := \text{argmax}_{2 \leq K \leq n}(g(K))$, where $g(K) = \lambda_K - \lambda_{K-1}$, $K = 2, \cdots, n$, indicates how many clusters there are in the network.

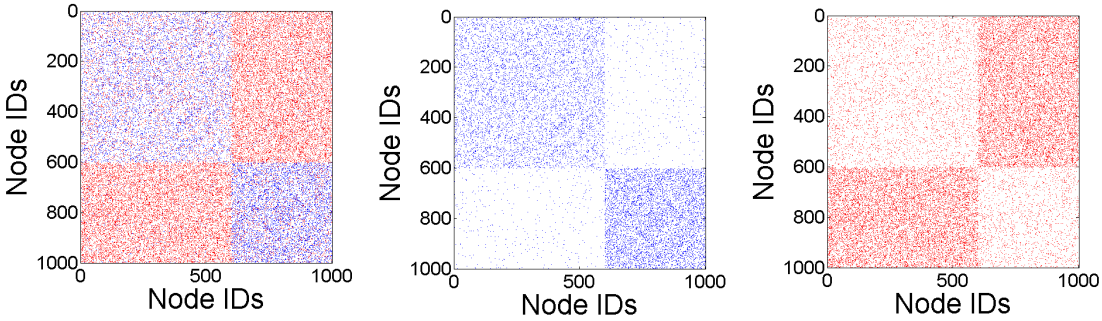
3.5 Evaluations

In this section, we evaluate the performance of the *mutuality-tendency-aware* spectral clustering method by comparing it with various symmetrization methods based digraph spectral clustering algorithms. We only present the comparison results for the adjacency matrix symmetrization method, with objective matrix as $\bar{A} = (A + A^T)/2$. For other settings, we obtained similar results and omit them here for brevity. We will 1) first test the performances using synthetic datasets, and then 2) apply our method to real online network datasets, e.g., Slashdot social network, and discover stable clusters with respect to mutual and one-way connections.

3.5.1 Synthetic datasets



(a) (i)All edges, (ii)Bidirectional edges, (iii)Unidirectional edges



(b) (i)All edges, (ii)Bidirectional edges, (iii)Unidirectional edges

Figure 3.2: Simulation results on synthetic dataset with $K = 2$ clusters. (a)(i)-(a)(iii) show the clusters detected by traditional spectral clustering algorithm, and (b)(i)-(b)(iii) show the clusters extracted using our mutuality-tendency-aware spectral clustering algorithm.

Table 3.1: Ave. mutuality tendency comparison on synthetic dataset

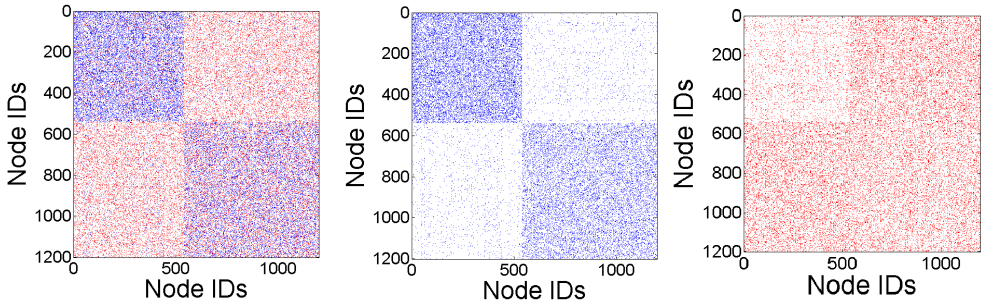
.	θ_G	θ_S	$\theta_{\bar{S}}$	$\theta_{\partial S}$
Mutuality tendency aware clustering	0.0112	**0.0172**	**0.0314**	**8.25e-5**
Traditional clustering	0.0112	0.0115	0.0202	0.0096

We first consider synthetic datasets designed specifically to test the performance of our mutuality-tendency-aware spectral clustering method. We randomly generate a network, with 1000 nodes. There are 38000 directed edges (around 3.8% of all directed node pairs[1]) in total, among which one third of them around 12666 edges are bidirectional, and two third of them around 25334 edges are unidirectional. Those nodes fall into 2 clusters, with 600 and 400 nodes respectively, where around 93.5% of the bidirectional edges are randomly placed *within* clusters, and around 80.8% of the unidirectional edges are randomly placed *across* clusters.

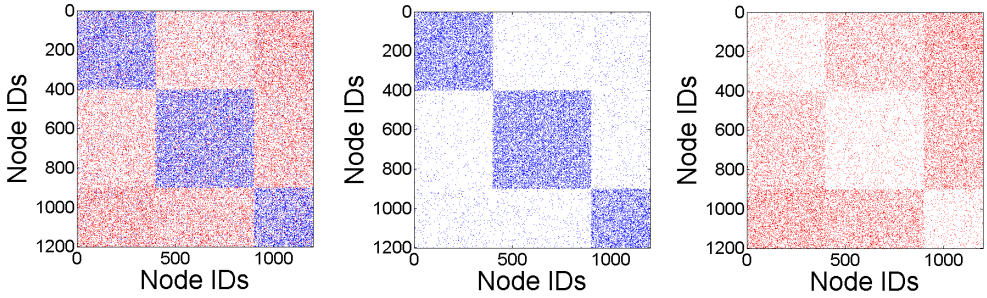
We show in Fig. 3.2(a)(i)-Fig. 3.2(a)(iii) that the traditional spectral clustering algorithm with $\bar{A} = (A + A^T)/2$ as the objective results in clusters with 180 and 820 nodes respectively, which does not reflect the underlying structure (See Fig. 3.2(a)(i)-Fig. 3.2(a)(iii), because it clusters nodes without considering the stability difference between mutual connections and one-way connections. On the other hand, our proposed mutuality-tendency-aware spectral clustering method can cluster the nodes into groups with exactly 600 and 400 nodes (See Fig. 3.2(b)(i)-Fig. 3.2(b)(iii)), which clearly group nodes with more mutual (stable) connections together and separate nodes connected via one-way connections.

Furthermore, given the cluster mutuality tendency Θ_S, we denote the average mutuality tendency of S as $\theta_S = \Theta_S/N_S$, with $N_S = |S|(|S|-1)/2$ as the total number of dyads in S. Similarly, we have the average mutuality tendency of G, $\bar{S}$, and ∂S as $\theta_G = \Theta_G/N_d$, $\theta_{\bar{S}} = \Theta_{\bar{S}}/N_{\bar{S}}$, and $\theta_{\partial S} = \Theta_{\partial S}/(|S||\bar{S}|)$, respectively. Table 3.1 shows the average mutuality tendencies of the cluster results obtained by two methods, where we can see that the *mutuality-tendency-aware* spectral clustering algorithm can group nodes together with higher within-cluster tendencies than that by traditional spectral clustering. On the other hand, the cross-cluster tendency obtained using our method is very close to 0, which means that the dyads across the clusters establish the mutual connections without any tendency (or purely independently). In addition, we generated synthetic dataset with $K > 2$ clusters, and similar results are obtained shown in

[1] As we observed from real social networks, e.g., Slashdot.com [52], an online commenting network dataset, which will be discussed in the next section, the sparsity of its "core" network is around 0.19%. Here, we choose 3.8%, that is 20 times large of the real network sparsity, just for the ease of visualization of the clustering structure.



(a) (i)All edges, (ii)Bidirectional edges, (iii)Unidirectional edges



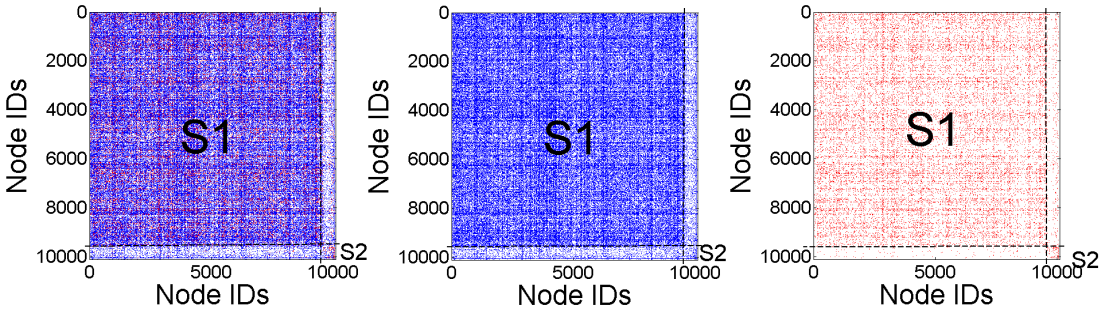
(b) (i)All edges, (ii)Bidirectional edges, (iii)Unidirectional edges

Figure 3.3: This synthetic dataset is generated in $K = 3$ clusters, with 500, 400 and 300 nodes, respectively. There are 54675 directional edges, among which 27336 edges are bidirectional and 27339 edges are unidirectional. We are randomly placed 90.02% of the bidirectional edges in clusters, and 89.6% of the unidirectional edges across clusters. (a)(i)-(a)(iii) show that traditional spectral clustering algorithm detects clusters with 661, 538 and 1 entities, respectively, while our method identify correct clusters (See (b)(i)-(b)(iii)).

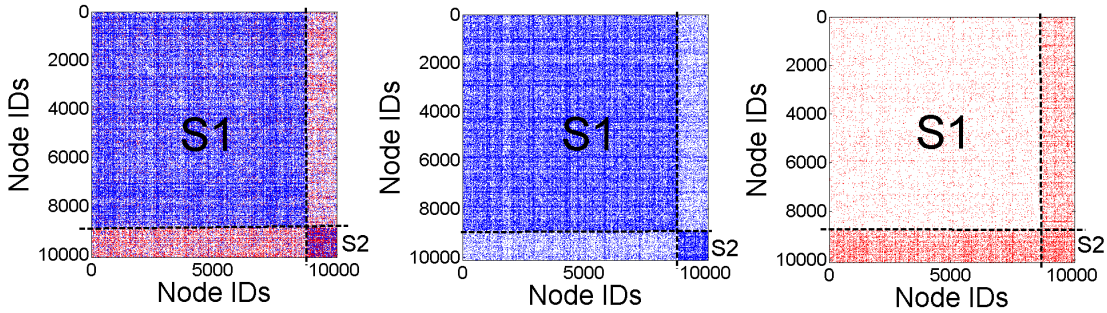
Fig. 3.3.

3.5.2 Real Social Networks

In the second set of simulations, we applied our *mutuality-tendency-aware* spectral clustering algorithm to several real social network datasets, e.g., Slashdot [52], Epinions [53], and email communication network [54] datasets, and compare with various symmetrization methods based digraph clustering algorithms, such as $A = (A + A^T)/2$, AA^T and $F_\pi = \Pi P$. Here we only show the comparison results with adjacency matrix symmetrization based digraph spectral clustering on Slashdot dataset. All other settings lead to similar results and we omit them



(a) (i)All edges, (ii)Bidirectional edges, (iii)Unidirectional edges



(b) (i)All edges, (ii)Bidirectional edges, (iii)Unidirectional edges

Figure 3.4: Simulation results on Slashdot social network dataset. (a)(i)-(a)(iii) show the clusters detected by traditional spectral clustering algorithm, and (b)(i)-(b)(iii) show the clusters extracted using our mutuality-tendency-aware spectral clustering algorithm.

here for brevity.

Slashdot is a technology-related news website founded in 1997. Users can submit stories and it allows other users to comment on them. In 2002, Slashdot introduced the Slashdot Zoo feature which allows users to tag each other as friends or foes. The network data we used is the Slashdot social relation network, where a directed edge from i to j indicates an interest from i to j's stories (or topics). Hence, two people with mutual connections thus share some common interests, while one-way connections infer that one is interested in the other's posts, but the interests are not reciprocated back. The Slashdot social network data was collected and released by Leskovec [52] in November 2008.

Table 3.2: Statistics of Slashdot social network Dataset

Nodes	77360
Edges	828161
Unidirectional edges	110199
Bidirectional edges	717962
Nodes in largest SCC	70355
Edges in largest SCC	818310
Unidirectional edges in largest SCC	100930
Bidirectional edges in largest SCC	717380
Nodes in the "core" component	10131
Edges in the "core" component	197378
Unidirectional edges in the "core" component	21404
Bidirectional edges in the "core" component	175974

The statistics[2] are shown in Table 3.2. It shows that the largest strongly connected component (SCC) include about 70355 nodes. Then, we remove those nodes with very low in-degrees and out-degrees, say no more than or equal to 2. By finding the largest strongly connected component of the remaining graph, we extract a "core" of the network with 10131 nodes and 197378 edges, among which there are 21404 unidirectional edges and 175974 bidirectional edges, respectively.

In our evaluations, we observe that there is a large "core" of the network, and all other users are attached to this core network. In our study, we are interested in extracting the community structure from the "core" network.

When applying our spectral clustering algorithm to the "core" network, two clusters with 8892 and 1239 nodes are detected (shown in Fig.3.4(b)(i)-Fig.3.4(b)(iii)). In our result, a large portion (about 35.04%) of cross-cluster edges are unidirectional edges which in turn yield lower mutuality tendency across clusters. On the other hand, when using the traditional symmetrized $\bar{A} = (A + A^T)/2$, two clusters with 9640 and 491 nodes are extracted instead (shown in Fig.3.4(a)(i)-Fig.3.4(a)(iii)). We can see that the clustering result obtained using the traditional spectral clustering method has only around 5.75% of the total edges across clusters as unidirectional edges, which boost up the mutuality tendency across clusters. However, in our clustering result, we have more unidirectional edges placed across clusters, which decreases the mutuality tendency across clusters. From Fig. 3.4(b)(i), we can clearly see that we have unidirectional

[2] Here, the total number of edges is smaller than that is shown on the website [52], because we do not count for those selfloops.

(red) edges dominating the cross-cluster parts.

Table 3.3 shows the average mutuality tendency comparison between different clustering methods, where we can see that the mutuality-tendency-aware spectral clustering algorithm can group nodes together with higher within-cluster tendencies than that of traditional spectral clustering.

Table 3.3: Ave. mutuality tendency comparison on Slashdot dataset

.	θ_G	θ_{S1}	θ_{S2}	$\theta_{\partial S}$
Mutuality tendency aware clustering	0.0017	**0.0049**	**0.0028**	**0.00033**
Traditional clustering	0.0017	0.0018	0.0021	0.00070

3.6 Summary

In this chapter, we establish a generalized mutuality tendency theory to capture the tendencies of clustered node pairs to establish mutual connections more frequently than those occur by chance. Based on our mutuality tendency theory, we develop a mutuality-tendency-aware spectral clustering algorithm that can detect stable clusters, by maximizing the within-cluster mutuality tendency and minimizing the cross-cluster mutual tendency. Extensive simulation results on synthetic, and real online social network datasets, such as Slashdot, demonstrate that our proposed *mutuality-tendency-aware* spectral clustering method resolves more stable social community structures than traditional spectral clustering methods.

Chapter 4

Influence Diffusion Dynamics and Influence Maximization in Social Networks with Friend and Foe Relationships

4.1 Introduction

As the popularity of online social networks (OSNs) such as Facebook and Twitter continuously increases, OSNs have become an important platform for the dissemination of news, ideas, opinions, etc. The openness of the OSN platforms and the richness of contents and user interaction information enable intelligent online recommendation systems and viral marketing techniques. For example, if a company wants to promote a new product, it may identify a set of influential users in the online social network and provide them with free sample products. They hope that these influential users could influence their friends, and friends of friends in the network and so on, generating a large influence cascade so that many users adopt their product as a result of such word-of-mouth effect. The question is how to select the initial users given a limited budget on free samples, so as to influence the largest number of people to purchase the product through this "word-of-mouth" process. Similar situations could apply to the promotion of ideas and opinions, such as political candidates trying to find early supporters for their political proposals

and agendas, government authorities or companies trying to win public support by finding and convincing an initial set of early adopters to their ideas.

The above problem is referred to as the *influence maximization* problem in the literature, which has been extensively studied in recent years [55–65]. In these studies, several influence diffusion models are proposed to formulate the underlying influence propagation processes, including linear threshold (LT) model, independent cascade (IC) model, voter model, etc. A number of approximation algorithms and scalable heuristics are designed under these models to solve the influence maximization problem.

However, all existing studies only look at networks with positive (i.e., friend, altruism, or trust) relationships, where in reality, relationships also include negative ones, such as foe, spite or distrust relationships. In Ebay, users develop trust and distrust in agents in the network; In online review and news forums, such as Epinions and Slashdot, readers approve or denounce reviews and articles of each other. Some recent studies [66–68] already look into the network structures with both positive and negative relationships. As a common sense exploited in many existing social influence studies [55, 58, 60–62], positive relationships carry the influence in a positive manner, i.e., you would *more likely* trust and adopt your friends' opinions. In contrast, we consider that negative relationships often carry influence in a reverse direction — if your foe chooses one opinion or votes for one candidate, you would *more likely* be influenced to do the opposite. This echoes the principles that "the friend of my enemy is my enemy" and "the enemy of my enemy is my friend". Structural balance theory has been developed based on these assumptions in social science (see chapter 5 of [69] and the references therein). We acknowledge that in real social networks, people's reactions to the influence from their friends or foes could be complicated, i.e., one could take the opposite opinion of what her foe suggests for one situation or topic, but may adopt the suggestion from the same person for a different topic, because she trusts her foe's expertise in that particular topic. In this study, we consider the influence diffusion for a single topic, where one always takes the opposite opinion of what her foe suggests. This is our first attempt to model influence diffusion in signed networks, and such topic-dependent simplification is commonly employed in prior influence diffusion studies on unsigned networks [55, 58, 60–63]. Our work aims at providing a mathematical analysis on the influence diffusion dynamic incorporated with negative relationship and applying our analysis to the algorithmic problem of influence maximization.

4.1.1 Our contributions

In this chapter, we extend the classic voter model [70, 71] to incorporate negative relationships for modeling the diffusion of opinions in a social network. Given an unsigned directed graph (digraph), the basic voter model works as follows. At each step, every node in the graph randomly picks one of its *outgoing* neighbors and adopts the opinion of this neighbor. Thus, the voter model is suitable to interpret and model opinion diffusions where people's opinions may switch back and forth based on their interactions with other people in the network. To incorporate negative relationships, we consider signed digraphs in which every directed edge is either positive or negative, and we consider the diffusion of two opposite opinions, e.g., black and white colors. We extend the voter model to signed digraphs, such that at each step, every node randomly picks one of its outgoing neighbors, and if the edge to this neighbor is positive, the node adopts the neighbor's opinion, but if the edge is negative, the node adopts the opposite of the neighbor's opinion (Section 4.2).

We provide detailed mathematical analysis on the voter model dynamics for signed networks (Section 4.3). For short-term dynamics, we derive the exact formula for opinion distribution at each step. For long-term dynamics, we provide closed-form formulas for the steady state distribution of opinions. We show that the steady state distribution depends on the graph structure: we divide signed digraphs into three classes of graph structures — balanced graphs, anti-balanced graphs, and strictly unbalanced graphs, each of which leads to a different type of steady state distributions of opinions. While balanced and unbalanced graphs have been extensively studied by structural balance theory in social science [69], the anti-balanced graphs form a new class that has not been covered before, to the best of our knowledge. Moreover, our long-term dynamics not only cover strongly connected and aperiodic digraphs that most of such studies focus on, but also weakly connected and disconnected digraphs, making our study more comprehensive.

We then study the influence maximization problem under the voter model for signed digraphs (Section 4.4). The problem here is to select at most k initial white nodes while all others are black, so that either in short term or long term the expected number of white nodes is maximized. This corresponds to the scenario where one opinion is dominating the public and an alternative opinion (e.g. a competing political agenda, or a new innovation) tries to win over supporters as much as possible by selecting some initial seeds to influence on. We provide

efficient algorithms that find optimal solutions for both short-term and long-term cases. In particular, for long-term influence maximization, our algorithm provides a comprehensive solution covering weakly connected and disconnected signed digraphs, with nontrivial computations on influence coverage of seed nodes.

Finally, we conduct extensive simulations on both real-world and synthetic networks to verify our analysis and to show the effectiveness of our influence maximization algorithm (Section 4.5). The simulation results demonstrate that our influence maximization algorithms perform much better than other heuristic algorithms.

To the best of our knowledge, we are the first to study influence diffusion and influence maximization in signed networks, and the first to apply the voter model to this case and provide efficient algorithms for influence maximization under voter model for signed networks.

4.1.2 Related work

In this subsection, we discuss the topics that are closely related to our problem, such as: (1) influence maximization and voter model, (2) signed networks, and (3) competitive influence diffusion.

Influence maximization and voter model. Influence maximization has been extensively studied in the literature. The initial work [55] proposes several influence diffusion models and provides the greedy approximation algorithm for influence maximization. More recent works [56–63] study efficient optimizations and scalable heuristics for the influence maximization problem. In particular, the voter model is proposed in [70, 71], and is suitable for modeling opinion diffusions in which people may switch opinions back and forth from time to time due to the interactions with other people in the network. Even-Dar and Shapira [58] study the influence maximization problem in the voter model on simple unsigned and undirected graphs, and they show that the best seeds for long-term influence maximization are simply the highest degree nodes. As a contrast, we show in this chapter that seed selection for signed digraphs are more sophisticated, especially for weakly connected or disconnected signed digraphs. Chung and Tsiatas [72] consider the voter models on hypergraphs as random walks on the associated weighted directed state graphs with various restrictions, such as “memoryless” or “partial memoryless” random walks, and provide bounds of the convergence time in various voter model settings. More voter model related research is conducted in physics domain, where the voter model, the zero-temperature Glauber dynamics for the Ising model, invasion process, and other related

models of population dynamics belong to the class of models with two absorbing states and epidemic spreading dynamics [73–75]. However, none of these works study the influence diffusion and influence maximization of voter model under signed networks.

Signed networks. The signed networks with both positive and negative links have gained attentions recently [66, 67, 76, 77]. In [66, 67], the authors empirically study the structure of real-world social networks with negative relationships based on two social science theories, i.e., balance theory and status theory. Kunegis et al. [76] study the spectral properties of the signed undirected graphs, with applications in link predictions, spectral clustering, etc. Borgs et al. [77] proposes a generalized PageRank algorithm for signed networks with application to online recommendations, where the distrust relations are considered as adversarial or arbitrary user behaviors, thus the outgoing relations of distrusted users are ignored while ranking nodes. Our algorithm can also be considered as an influence ranking algorithm that generalizes the PageRank algorithm, but we treat distrust links as generating negative influence rather than ignoring distrusted users' opinions, and thus our ranking method is different from [77]. None of the above work studies influence diffusion and influence maximization in signed networks.

Competitive influence diffusion. A number of recent studies focus on competitive influence diffusion and maximization [78–83], in which two or more competitive opinions or innovations are diffusing in the network. Although they consider two or more competitive or opposing influence diffusions, they are all on unsigned networks, different from our study here on diffusion with both positive and negative relationships.

4.2 Voter model for signed networks

We consider a weighted directed graph (digraph) $G = (V, E, A)$, where V is the set of vertices, E is the set of directed edges, and A is the weighted adjacency matrix with $A_{ij} \neq 0$ if and only if $(i, j) \in E$, with A_{ij} as the weight of edge (i, j). The voter model was first introduced for unsigned graphs, with nonnegative adjacency matrices A's. In this model, each node holds one of two opposite opinions, represented by black and white colors. Initially each node has either black or white color. At each step $t \geq 1$, every node i randomly picks one outgoing neighbor j with the probability proportional to the weight of (i, j), namely $A_{ij}/\sum_\ell A_{i\ell}$, and changes its color to j's color. The voter model also has a random walk interpretation. If a random walk starts from i and stops at node j at step t, then i's color at step t is j's color at step 0.

Table 4.1: Notations and terminologies

$G = (V, E, A)$, $\bar{G} = (V, E, \bar{A})$	G is a signed digraph, with signed adjacency matrix A and $\bar{G}$ is the unsigned version of G, with adjacency matrix $\bar{A}$
A^+, A^-	A^+ (resp. A^-) is the non-negative adjacency matrix representing positive (resp. negative) edges of G, with $A = A^+ - A^-$ and $\bar{A} = A^+ + A^-$.
$\mathbf{1}$, π, x_0, x_t, x, x_e, x_o	Vector forms. All vectors are $\|V\|$-dimensional column vectors by default; $\mathbf{1}$ is all one vector, π is the stationary distribution of ergodic digraph $\bar{G}$; x_0 (resp. x_t) is the white color distribution at the beginning (resp. at step t); x is the steady state white color distribution; x_e (resp. x_o) is the steady state white color distribution for even (resp. odd) steps.
d, d^+, d^-, D	d, d^+, and d^- are weighted out-degree vectors of G, where $d = A\mathbf{1}$, $d^+ = A^+\mathbf{1}$, and $d^- = A^-\mathbf{1}$; $D = diag[d]$ is the diagonal degree matrix filled with entries of d.
P, $\bar{P}$	$P = D^{-1}A$ is the signed transition matrix of G and $\bar{P} = D^{-1}A$ is the transition probability matrix of $\bar{G}$.
v_Z, $\hat{v}_S$, $\hat{v}_{Z,S_Z}$	Given a vector v, a node set $Z \subseteq V$, v_Z is the projection of v on Z. Given a partition $S, \bar{S}$ of V, $\hat{v}_S$ is signed such that $\hat{v}_S(i) = v(i)$ if $i \in S$, and $\hat{v}_S(i) = -v(i)$ if $i \notin S$. Given a partition $S_Z, \bar{S}_Z$ of Z, $\hat{v}_{Z,S_Z}$ is taking the projection of v on Z first, then negating the signs for entries in $\bar{S}_Z$.
I, $\hat{I}_S$, B_Z	I is the identity matrix. $\hat{I}_S = diag[\hat{\mathbf{1}}_S]$ is the signed identity matrix. B_Z is the projection of a matrix B to $Z \subseteq V$.

In this chapter, we extend the voter model to signed digraphs, in which the adjacency matrix A may contain negative entries. A positive entry A_{ij} represents that i considers j as a friend or i trusts j, and a negative A_{ij} means that i considers j as a foe or i distrusts j. The absolute value $|A_{ij}|$ represents the strength of this trust or distrust relationship. The voter model is thus extended naturally such that one always takes the same opinion from his/her friend, and the opposite opinion of his/her foe. Technically, at each step $t \geq 1$, i randomly picks one outgoing neighbor j with probability $|A_{ij}|/\sum_\ell |A_{i\ell}|$, and if $A_{ij} > 0$ (or edge (i, j) is positive) then i changes its color to j's color, but if $A_{ij} < 0$ (or edge (i, j) is negative) then i changes its color to the opposite of j's color. The random walk interpretation can also be extended for signed networks: if the t-step random walk from i to j passes an even number of negative edges, then i's color at step t is the same as j's color at step 0; while if it passes an odd number of negative edges, then i's color at step t is the opposite of j's color at step 0.

Given a signed digraph $G = (V, E, A)$, let $G^+ = (V, E^+, A^+)$ and $G^- = (V, E^-, A^-)$ denote the unsigned subgraphs consisting of all positive edges E^+ and all negative edges E^-, respectively, where A^+ and A^- are the corresponding non-negative adjacency matrices. Thus

we have $A = A^+ - A^-$. Similar to unsigned digraphs, G is *aperiodic* if the greatest common divisor of the lengths of all cycles in G is 1, and G is *ergodic* if it is strongly connected and aperiodic. A *sink component* of a signed digraph is a strongly connected component that has no outgoing edges to any nodes outside the component. When studying the long-term dynamics of the voter model, we assume that all signed strongly connected components are ergodic. We first study the case of ergodic graphs, and then extend it to the more general case of weakly connected or disconnected graphs with ergodic sink components. Table 4.1 provides notations and terminologies used in the chapter. Note that one basic fact we often use in studying long-term convergence behavior is: If matrix P satisfies $\lim_{t\to\infty} P^t = \mathbf{0}$, then $I - P$ is invertible and $(I - P)^{-1} = \lim_{t\to\infty} \sum_{i=0}^{t} P^i$.

4.3 Analysis of voter model dynamics on signed digraphs

In this section, we study the short-term and long-term dynamics of the voter model on signed digraphs. In particular, we answer the following two questions.
(i) Short-term dynamics: Given an initial distribution of black and white nodes, what is the distribution of black and white nodes at step $t > 0$?
(ii) Convergence of voter model: Given an initial distribution of black and white nodes, would the distribution converge, and what is the steady state distribution of black and white nodes?

4.3.1 Short-term dynamics

To study voter model dynamics on signed digraphs, we first define the *signed transition matrix* as follows.

Definition 2 (Signed transition matrix) *Given a signed digraph $G = (V, E, A)$, we define the* signed transition matrix *of G as $P = D^{-1}A$, where $D = diag[d_i]$ is the diagonal matrix and $d_i = \sum_{j\in V} |A_{ij}|$ is the weighted out-degree of node i.*

Next proposition characterizes the dynamics of the voter model at each step using the *signed transition matrix*.

Proposition 1 *Let $G = (V, E, A)$ be a signed digraph and denote the initial white color distribution vector as x_0, i.e., $x_0(i)$ represents the probability that node i is white initially. Then, the*

white color distribution at step t, denoted by x_t can be computed as

$$x_t = P^t x_0 + (\sum_{i=0}^{t-1} P^i) g^-, \tag{4.1}$$

where $g^- = D^{-1}A^-\mathbf{1}$, i.e. $g^-(i)$ is the weighted fraction of outgoing negative edges of node i.

Proof : Based on the signed digraph voter model defined in Section 4.2, x_t can be iteratively computed as

$$x_t(i) = \sum_{j \in V} \frac{A^+_{ij}}{d_i} x_{t-1}(j) + \sum_{j \in V} \frac{A^-_{ij}}{d_i} (1 - x_{t-1}(j)). \tag{4.2}$$

In matrix form, we have

$$x_t = D^{-1}Ax_{t-1} + D^{-1}A^-\mathbf{1} = Px_{t-1} + g^-, \tag{4.3}$$

which yields Eq.(4.1) by repeatedly applying Eq.(4.3). ∎

4.3.2 Convergence of signed transition matrix with relation to structural balance of signed digraphs

Eq.(4.1) infers that the long-term dynamics, i.e., the vector x_t when t goes to infinity, depends critically on the limit of P^t and $\sum_{i=0}^{t-1} P^i$. We show below that the limiting behavior of the two matrix sequences is fundamentally determined by the structural balance of signed digraph G, which connects to the social balance theory well studied in the social science literature (cf. [69]). We now define three types of signed digraphs based on their balance structures.

Definition 3 (Structural balance of signed digraphs) *Let $G = (V, E, A)$ be a signed digraph.*

1. **Balanced digraph**. *G is* balanced *if there exists a partition $S, \bar{S}$ of nodes in V, such that all edges within S and $\bar{S}$ are positive and all edges across S and $\bar{S}$ are negative.*
2. **Anti-balanced digraph**. *G is* anti-balanced *if there exists a partition $S, \bar{S}$ of nodes in V, such that all edges within S and $\bar{S}$ are negative and all edges across S and $\bar{S}$ are positive.*
3. **Strictly unbalanced digraph**. *G is* strictly unbalanced *if G is neither balanced nor anti-balanced.*

The balanced digraphs defined above correspond to the balanced graphs originally defined in social balance theory. It is known that a balanced graph can be equivalently defined by the condition that all circles in G without considering edge directions contain an even number of negative edges [69]. On the other hand, the concept of anti-balanced digraphs seems not appearing in the social balance theory. Note that balanced digraphs and anti-balanced digraphs are not mutually exclusive. For example, a four node circle with one pair of non-adjacent edges being positive and the other pair being negative is both balanced and anti-balanced. However, for studying long-term dynamics, we only need the above categorization for aperiodic digraphs, for which we show below that balanced digraphs and anti-balanced digraphs are mutually exclusive.

Proposition 2 *An aperiodic digraph G cannot be both balanced and anti-balanced.*

Proof :

Suppose, for a contradiction, that an aperiodic digraph G is both balanced and anti-balanced. By the equivalent condition of balanced graphs, we know that all cycles of G have an even number of negative edges. Since an anti-balanced graph will become balanced if we negate the signs of all its edges, we know that all cycles of G also have an even number of positive edges. Therefore, all cycles of G must have an even number of edges, which means their lengths have a common divisor 2, contradicting to the assumption that G is aperiodic. ∎

With the above proposition, we know that balanced graphs, anti-balanced graphs, and strictly unbalanced graphs indeed form a classification of aperiodic digraphs, where anti-balanced graphs and strictly unbalanced graphs together correspond to unbalanced graphs in the social balance theory. We identify anti-balanced graphs as a special category because it has a unique long-term dynamic behavior different from other graphs. An example of anti-balanced graphs is a graph with only negative edges. In general, anti-balanced graphs could be viewed as an extreme in which many hostility exist among individuals, e.g., networks formed by bidders in auctions [84, 85].

Case of ergodic signed digraphs. Now, we discuss the limiting behavior of P^t of ergodic signed digraphs with three balance structures. A signed digraph $G = (V, E, A)$ is ergodic if and only if for any node i, there always exists a signed path to any other node in G and the common divisor of all cycle path lengths of i is 1. Here, a signed path R in a signed graph G is a sequence of nodes with the edges being directed from each node to the following one, where

the length of the path, denoted as $|R|$, is the total number of directed edges in R. The sign of a path is positive, if there is an even number of negative edges along the path; otherwise the sign of a path is negative. Below, we first introduce Proposition 3 presenting that the balance structures of ergodic signed digraphs can be interpreted and distinguished in terms of the path lengths and path signs in G. As a result, Lemma 2 introduces the various limiting behaviors of P^t of ergodic signed digraphs with respect to three balance structures.

Proposition 3 *Let $G = (V, E, A)$ be an ergodic strictly unbalanced digraph. There exist two nodes i and j, and two directed paths from i to j with the same length but different signs.*

Proof : Given the following three statements, we prove `Statement 1` $\Rightarrow$ `Statement 2` $\Rightarrow$ `Statement 3`, which in turn proves this proposition, i.e., $\neg$`Statement 3` $\Rightarrow$ $\neg$`Statement 1`.
We assume that G is a signed ergodic digraph.
Statement 1: For any two nodes i and j, all paths from i to j with the same length have same signs.
Statement 2: For any two nodes i and j, all paths from i to j with even length have same signs.
Statement 3: G is either balanced or anti-balanced.
(1) Proof by contradiction for `Statement 1` $\Rightarrow$ `Statement 2`. We assume that in G, there exist two even length paths R_{e1} and R_{e2} from i to j with different signs. Since G is ergodic, by Proposition 4 in Appendix A.1, there must exist a path, denoted by R_o, from j to i with odd length (no matter what sign it carries). Denote the length of these three paths as $|R_{e1}|$, $|R_{e2}|$ and $|R_o|$, respectively.

Then, $R_{c1} = R_{e1} + R_o$ forms a cycle at node i with odd length $|R_{e1}| + |R_o|$ and $R_{c2} = R_{e2} + R_o$ forms another cycle at i with odd length $|R_{e2}| + |R_o|$. Clearly, two cycles R_{c1} and R_{c2} carry different signs. Then, let $R'_{c1} = R_{c1}^{|R_{c2}|}$ denote a cycle of node i, by continuing R_{c1} for $|R_{c2}|$ times, which has the same sign with R_{c1} since $|R_{c2}|$ is odd. Similarly, we construct a cycle $R'_{c2} = R_{c2}^{|R_{c1}|}$ by continuing R_{c2} for $|R_{c1}|$ times, which has the same sign as R_{c2}. Thus R'_{c1} and R'_{c2} have the same length of $|R_{c1}||R_{c2}|$ but different signs, which contradicts to Statement 1.
(2) Proof for `Statement 2` $\Rightarrow$ `Statement 3`. By Proposition 4 in Appendix A.1, we know that between any two nodes there must exist even-length paths. By Statement 2, we partition V into S and $\bar{S}$, based on the signs of even length paths originated from a particular node $i \in V$. More specifically, S contains the nodes to which all even length paths from i have positive signs, and $\bar{S}$ contains the other set of nodes (note that i may not be in S).

We argue that (a) within S and $\bar{S}$, all edges have same signs; and (b) all edges between S and $\bar{S}$ have same signs. Since G contains both negative and positive edges, it must be either balanced or anti-balanced.

For (a), assume to the contrary that there exist two directed edges $R_{ab} = a \rightarrow b$ and $R_{cd} = c \rightarrow d$, which both reside in the same set, e.g., S with different signs. (The case for $\bar{S}$ is similar.)

We construct two even length paths from i to c and i to d as follows.

$$R_e(i,c) = R_e(i,b) + R_e(b,c),$$
$$R_e(i,d) = R_e(i,a) + R_{ab} + R_e(b,c) + R_{cd}$$

where $R_e(x,y)$ represents the constructed even length path from node x to node y.

Since both $c, d \in S$, by construction, then $R_e(i,c)$ and $R_e(i,d)$ have same signs

$$sgn(R_e(i,c)) = sgn(R_e(i,d)). \tag{4.4}$$

On the other hand, since a and b are in the same group as c and d, $sgn(R_e(i,a)) = sgn(R_e(i,b))$. Then, we have

$$sgn(R_e(i,c)) = sgn(R_e(i,b))sgn(R_e(b,c)), \tag{4.5}$$
$$\begin{aligned} sgn(R_e(i,d)) &= sgn(R_e(i,a))sgn(R_{ab})sgn(R_e(b,c))sgn(R_{cd}) \\ &= -sgn(R_e(i,b))sgn(R_e(b,c)). \end{aligned} \tag{4.6}$$

Eq.(4.6) comes from the assumption that R_{ab} and R_{cd} have different signs. Eq.(4.4) contradicts with Eq.(4.5) and Eq.(4.6).

For (b), assume that there exist two edges R_{ab} and R_{cd} with different signs between S and $\bar{S}$. Still consider the two even length paths $R_e(i,c)$ and $R_e(i,d)$ constructed before. Since c and d are not in the same side, $R_e(i,c)$ and $R_e(i,d)$ have opposite signs by the construction, i.e.,

$$sgn(R_e(i,c)) = -sgn(R_e(i,d)). \tag{4.7}$$

On the other hand, since a and b are in the different groups as well, $sgn(R_e(i,a)) = -sgn(R_e(i,b))$. Then, we have

$$sgn(R_e(i,c)) = sgn(R_e(i,b)) \cdot sgn(R_e(b,c)), \tag{4.8}$$
$$\begin{aligned} sgn(R_e(i,d)) &= sgn(R_e(i,a))sgn(R_{ab})sgn(R_e(b,c))sgn(R_{cd}) \\ &= sgn(R_e(i,b)) \cdot sgn(R_e(b,c)). \end{aligned} \tag{4.9}$$

However, Eq.(4.7) contradicts with Eq.(4.8) and Eq.(4.9). This completes the proof. ∎

The next lemma characterizes the limiting behavior of P^t of ergodic signed digraphs with all three balance structures. Given a signed digraph $G = (V, E, A)$, let $\bar{G} = (V, E, \bar{A})$ corresponds to its unsigned version ($\bar{A}_{ij} = |A_{ij}|$ for all $i, j \in V$). When $\bar{G}$ is ergodic, a random walk on $\bar{G}$ has a unique stationary distribution, denoted as π. That is, $\pi^T = \pi^T \bar{P}$, where $\bar{P} = D^{-1}\bar{A}$ is the transition probability matrix for $\bar{G}$. Henceforth, we always use $S, \bar{S}$ to denote the corresponding partition for either balanced graphs or anti-balanced graphs. We define the infinity norm of matrix $M \in \mathbb{R}^{m\times m}$ as: $\|M\|_\infty := \max_{1\le i\le m} \sum_{j=1}^m |M_{ij}|$.

Lemma 2 *Given an ergodic signed digraph $G = (V, E, A)$, let $\bar{G} = (V, E, \bar{A})$ be the* unsigned *digraph. When G is balanced or strictly unbalanced, P^t converges, and when G is anti-balanced, the odd and even subsequences of P^t converge to opposite matrices.*

$$\begin{aligned}
&\textit{Balanced } G: && \lim_{t\to\infty} P^t = \hat{\boldsymbol{1}}_S \hat{\pi}_S^T;\\
&\textit{Strictly unbalanced } G: && \lim_{t\to\infty} P^t = \boldsymbol{0};\\
&\textit{Anti-balanced } G: && \lim_{t\to\infty} P^{2t} = \hat{\boldsymbol{1}}_S \hat{\pi}_S^T,\ \lim_{t\to\infty} P^{2t+1} = -\hat{\boldsymbol{1}}_S \hat{\pi}_S^T.
\end{aligned}$$

Proof : (1) When G is balanced, the signed transition matrix P can be written as $P = \hat{I}_S \bar{P} \hat{I}_S$. Since $\bar{G}$ is ergodic, we have $\lim_{t\to\infty} \bar{P}^t = \mathbf{1}\pi^T$. Thus,

$$\lim_{t\to\infty} P^t = \lim_{t\to\infty} (\hat{I}_S \bar{P} \hat{I}_S)^t = \hat{\mathbf{1}}_S \hat{\pi}_S^T,$$

where we use simple facts $\hat{I}_S^2 = I$, $\hat{I}_S \mathbf{1} = \hat{\mathbf{1}}_S$, and $\pi^T \hat{I}_S = \hat{\pi}_S^T$.
(2) When G is anti-balanced, we have $P = -\hat{I}_S \bar{P} \hat{I}_S$. Thus,

$$\begin{aligned}
\lim_{t\to\infty} P^{2t} &= \lim_{t\to\infty} (-\hat{I}_S \bar{P} \hat{I}_S)^{2t} = \hat{\mathbf{1}}_S \hat{\pi}_S^T\\
\lim_{t\to\infty} P^{2t+1} &= \lim_{t\to\infty} (-\hat{I}_S \bar{P} \hat{I}_S)^{2t+1} = -\hat{\mathbf{1}}_S \hat{\pi}_S^T.
\end{aligned}$$

(3) By Proposition 3, given a signed strictly unbalanced digraph G, there exist a pair of nodes i and j, such that two paths R_1 and R_2 from i to j have the same length $\ell(i)$ and opposite signs. Consider a random walk from i. Let p_1 (resp. p_2) be the probability that the walk exactly follows R_1 (resp. R_2) in the first $\ell(i)$ steps. Let $R_{i,k}^{\ell(i)}$ be the set of all paths from i to k with

length $\ell(i)$. Then, for a unit vector e_i with i-th entry equal to 1 and other entries as 0, we have

$$||e_i^T P^{\ell(i)}||_1 = \sum_{k\in V} \left| \sum_{R\in R_{i,k}^{\ell(i)}} \text{Prob}[R] sgn(R) \right| \leq 1 - \min(p_1, p_2) = \rho_i.$$

For any other node i', there must exist a path R' from $i' \to i$, due to the ergodicity of G, thus two paths $R'_1 = R' + R_1$ and $R'_2 = R' + R_2$ from i' to j have the same length, but opposite signs. With similar arguments as that for node i, $||e_{i'}^T P^{\ell(i')}||_1 \leq \rho_{i'}$ holds for any $i' \in V$. Let $\rho = \max_i \rho_i < 1$ and $\ell = \max_i \ell(i)$, we conclude for any $i \in V$, $||e_i^T P^{\ell}||_1 \leq \rho$ holds. Hence, when $t \geq T = 2\ell$, the following inequality holds

$$||e_i^T P^t||_1 = ||e_i^T P^{\frac{t}{\ell}\ell}||_1 \leq \rho^{\lfloor \frac{t}{\ell} \rfloor} \leq \rho^{\frac{t}{T}}.$$

Hence $\lim_{t\to\infty} \|P^t\|_\infty = 0$, i.e., $\lim_{t\to\infty} P^t = \mathbf{0}$. ∎

The above lemma clearly shows different convergence behaviors of P^t for three types of graphs. In particular, P^t of anti-balanced graphs exhibits a bounded oscillating behavior in the long term.

Case of weakly connected signed digraphs. Now, we consider a weakly connected signed digraph $G = (V, E, A)$ with one ergodic sink component G_Z with node set Z, which only has incoming edges from the rest of the signed digraph G_X with node set $X = V \setminus Z$. Then, the signed transition matrix P has the following block form.

$$P = \left[\begin{array}{c|c} P_X & P_Y \\ \hline \mathbf{0} & P_Z \end{array}\right], \tag{4.10}$$

where P_X and P_Z are the block matrices for component G_X and G_Z, and P_Y represent the one-way connections from G_X to G_Z. Then, the t-step transition matrix P^t can be expressed as

$$P^t = \left[\begin{array}{c|c} P_X^{(t)} & P_Y^{(t)} \\ \hline \mathbf{0} & P_Z^{(t)} \end{array}\right], \tag{4.11}$$

where $P_X^{(t)} = P_X^t$, $P_Z^{(t)} = P_Z^t$ and $P_Y^{(t)} = \sum_{i=0}^{t-1} P_X^i P_Y P_Z^{t-1-i}$. When G_Z is balanced or anti-balanced, we use $S_Z, \bar{S}_Z$ to denote the partition of Z defining its balance or anti-balance structure. Then, we denote column vectors

$$u_b = (I_X - P_X)^{-1} P_Y \hat{\mathbf{1}}_{Z,S_Z}, \tag{4.12}$$

$$\text{and } u_u = (I_X + P_X)^{-1} P_Y \hat{\mathbf{1}}_{Z,S_Z}. \tag{4.13}$$

The reason that $I_X - P_X$ is invertible is because $\lim_{t\to\infty} P_X^t = \mathbf{0}$, which is in turn because there is a path from any node i in G_X to nodes in Z (since Z is the single sink), and thus informally a random walk from i eventually reaches and then stays in G_Z. The same reason applies to $I_X + P_X$. Lemma 3 provides the formal proof of the fact $\lim_{t\to\infty} P_X^t = \mathbf{0}$.

Let π_Z denote the stationary distribution of nodes in G_Z, and $\hat{\pi}_{Z,S_Z}$ is signed, with $\hat{\pi}_{Z,S_Z}(i) = \pi_Z(i)$ for $i \in S_Z$, and $\hat{\pi}_{Z,S_Z}(i) = -\pi_Z(i)$ for $i \in Z \setminus S_Z$. Lemma 3 discloses the convergence of P^t given various balance structures of G_Z.

Lemma 3 *For weakly connected signed digraph $G = (V, E, A)$ with one ergodic sink components, with signed transition matrix given in Eq.(4.11), we have*

$$\begin{aligned}
\textit{Balanced } G_Z\textit{:} \qquad & \lim_{t\to\infty} P^t = \left[\begin{array}{c|c} \boldsymbol{0} & u_b \hat{\pi}_{Z,S_Z}^T \\ \hline \boldsymbol{0} & \hat{\boldsymbol{I}}_{Z,S_Z} \hat{\pi}_{Z,S_Z}^T \end{array}\right] \\
\textit{Strictly unbalanced } G_Z\textit{:} \qquad & \lim_{t\to\infty} P^t = \boldsymbol{0} \\
\textit{Anti-balanced } G_Z\textit{:} \qquad & \lim_{t\to\infty} P^{2t} = \left[\begin{array}{c|c} \boldsymbol{0} & -u_u \hat{\pi}_{Z,S_Z}^T \\ \hline \boldsymbol{0} & \hat{\boldsymbol{I}}_{Z,S_Z} \hat{\pi}_{Z,S_Z}^T \end{array}\right], \\
& \lim_{t\to\infty} P^{2t+1} = \left[\begin{array}{c|c} \boldsymbol{0} & u_u \hat{\pi}_{Z,S_Z}^T \\ \hline \boldsymbol{0} & -\hat{\boldsymbol{I}}_{Z,S_Z} \hat{\pi}_{Z,S_Z}^T \end{array}\right]
\end{aligned}$$

Proof : We discuss the convergence of P_X^t, P_Z^t, and $P_Y^{(t)}$ in Eq.(4.11), respectively.
(1) We first prove that P_X^t converges to $\mathbf{0}$, i.e., $\lim_{t\to\infty} P_X^t = \mathbf{0}$.

Since G_X does not contain sink components, any node $i \in X$ has a path to component G_Z. Let R_{iZ} be the shortest path from i to some node in Z, and $Prob[R_{iZ}]$ denote the probability that a random walk starting from i takes the path R_{iZ}. Hence we denote

$$p = \min_{i\in X} Prob[R_{iZ}], \text{ and } m = \max_{i\in X} |R_{iZ}|,$$

which implies that starting from any node $i \in X$, after m steps of random walk, there is at least probability p that it reaches component G_Z. Hence, we have $\|P_X^m\|_\infty \leq (1-p) < 1$. Let $T = 2m$, then for any $t > T$, we have

$$\|P_X^t\|_\infty = \|P_X^{\frac{t}{m}m}\|_\infty \leq (1-p)^{\lfloor \frac{t}{m} \rfloor} \leq (1-p)^{\frac{t}{T}},$$

which implies $\lim_{t\to\infty} \|P_X^t\|_\infty = 0$, i.e., $\lim_{t\to\infty} P_X^t = \mathbf{0}$.

(2) For subgraph G_Z, Lemma 2 directly yields

$$\lim_{t\to\infty} P_Z^t = \begin{cases} \mathbf{0}, & \text{Strictly unbalanced } G_Z; \\ \mathbf{1}_{Z,S_Z}\pi_{Z,S_Z}^T, & \text{Balanced } G_Z; \\ \mathbf{1}_{Z,S_Z}\pi_{Z,S_Z}^T, & \text{Anti-balanced } G_Z, \text{ even } t; \\ -\mathbf{1}_{Z,S_Z}\pi_{Z,S_Z}^T, & \text{Anti-balanced } G_Z, \text{ odd } t. \end{cases} \tag{4.14}$$

(3) Below, we focus on proving the results on $\lim_{t\to\infty} P_Y^{(t)}$ using Proposition 6 in Appendix A.2. When G_Z *is strictly unbalanced,* from Lemma 2 and (1) in this proof, $\lim_{t\to\infty} P_X^t = \mathbf{0}$ and $\lim_{t\to\infty} P_Z^t = \mathbf{0}$ hold, thus by Proposition 6 in Appendix A.2 $\lim_{t\to\infty} P_Y^{(t)} = \mathbf{0}$.
When G_Z *is balanced,* Lemma 2 and Proposition 5 in Appendix A.1 directly yield $(P_Z - \mathbf{1}_{Z,S_Z}\pi_{Z,S_Z}^T)^t = P_Z^t - \mathbf{1}_{Z,S_Z}\pi_{Z,S_Z}^T$ for any integer $t > 0$, and $\lim_{t\to\infty}(P_Z - \mathbf{1}_{Z,S_Z}\pi_{Z,S_Z}^T)^t = \mathbf{0}$, thus

$$\begin{aligned} \lim_{t\to\infty} P_Y^{(t)} &= \lim_{t\to\infty} \sum_{i=0}^{t-1} P_X^i P_Y (P_Z^{t-1-i} - \mathbf{1}_{Z,S_Z}\pi_{Z,S_Z}^T + \mathbf{1}_{Z,S_Z}\pi_{Z,S_Z}^T) \\ &= \lim_{t\to\infty} \sum_{i=0}^{t-1} P_X^i P_Y (P_Z - \mathbf{1}_{Z,S_Z}\pi_{Z,S_Z}^T)^{t-1-i} + \lim_{t\to\infty} \sum_{i=0}^{t-2} P_X^i P_Y \mathbf{1}_{Z,S_Z}\pi_{Z,S_Z}^T \\ &= (I_X - P_X)^{-1} P_Y \mathbf{1}_{Z,S_Z}\pi_{Z,S_Z}^T = u_b \pi_{Z,S_Z}^T, \end{aligned}$$

where the first term in the second line being $\mathbf{0}$ is due to Proposition 6 (ii) in Appendix A.2.
When G_Z *is anti-balanced,* applying Lemma 2 and Proposition 5 in Appendix A.1, we have for any integer $t > 0$, $(P_Z + \mathbf{1}_{Z,S_Z}\pi_{Z,S_Z}^T)^t = P_Z^t - (-1)^t \mathbf{1}_{Z,S_Z}\pi_{Z,S_Z}^T$, and $\lim_{t\to\infty}(P_Z + \mathbf{1}_{Z,S_Z}\pi_{Z,S_Z}^T)^t = \mathbf{0}$ hold true, thus

$$\begin{aligned} \lim_{t\to\infty} P_Y^{(t)} &= \lim_{t\to\infty} \sum_{i=0}^{t-1} P_X^i P_Y (P_Z^{t-1-i} - (-1)^{t-1-i}(\mathbf{1}_{Z,S_Z}\pi_{Z,S_Z}^T - \mathbf{1}_{Z,S_Z}\pi_{Z,S_Z}^T)) \\ &= \lim_{t\to\infty} \sum_{i=0}^{t-1} P_X^i P_Y (P_Z + \mathbf{1}_{Z,S_Z}\pi_{Z,S_Z}^T)^{t-1-i} + \lim_{t\to\infty} \sum_{i=0}^{t-2} (-1)^{t-1-i} P_X^i P_Y \mathbf{1}_{Z,S_Z}\pi_{Z,S_Z}^T \\ &= (-1)^{t-1} \lim_{t\to\infty} \sum_{i=0}^{t-2} (-P_X)^i P_Y \mathbf{1}_{Z,S_Z}\pi_{Z,S_Z}^T = (-1)^{t-1}(I_X + P_X)^{-1} P_Y \mathbf{1}_{Z,S_Z}\pi_{Z,S_Z}^T \\ &= (-1)^{t-1} u_u \pi_{Z,S_Z}^T. \end{aligned}$$

Hence, we have for anti-balanced G_Z: $\lim_{t\to\infty} P_Y^{(2t)} = -u_u \hat{\pi}_{Z,S_Z}^T$, and $\lim_{t\to\infty} P_Y^{(2t+1)} = u_u \hat{\pi}_{Z,S_Z}^T$. ∎

Multiple sink components and disconnected signed digraphs. When there exist $m > 1$ ergodic sink components, i.e., $G_{Z1}, G_{Z2}, \cdots, G_{Zm}$, the rest of the graph G is considered as G_X. Then the signed transition matrix P and P^t can be written as

$$P = \left[\begin{array}{c|c|c|c} P_X & P_{Y1} & \cdots & P_{Ym} \\ \hline \mathbf{0} & P_{Z1} & \mathbf{0} & \mathbf{0} \\ \hline \mathbf{0} & \mathbf{0} & \ddots & \mathbf{0} \\ \hline \mathbf{0} & \mathbf{0} & \mathbf{0} & P_{Zm} \end{array}\right], P^t = \left[\begin{array}{c|c|c|c} P_X^t & P_{Y1}^{(t)} & \cdots & P_{Ym}^{(t)} \\ \hline \mathbf{0} & P_{Z1}^t & \mathbf{0} & \mathbf{0} \\ \hline \mathbf{0} & \mathbf{0} & \ddots & \mathbf{0} \\ \hline \mathbf{0} & \mathbf{0} & \mathbf{0} & P_{Zm}^t \end{array}\right], \tag{4.15}$$

where $P_{Yi}^{(t)} = \sum_{j=0}^{t-1} P_X^j P_{Yi} P_{Zi}^{t-1-j}$, $1 \leq i \leq m$. Hence, each sink ergodic component P_{Zi} along with P_X and P_{Yi} independently follows Lemma 3. For disconnected signed digraph, with $m \geq 1$ ergodic or weakly connected components, each of which satisfies Lemma 2 or Lemma 3, respectively. For brevity, we omit the details here.

4.3.3 Long-term dynamics

Based on the structural balance classification and the convergence of signed transition matrix discussed above, we are ready now to analyze the long-term dynamics of the voter model on signed digraphs. Formally, we are interested in characterizing x_t with $t \to \infty$, i.e.,

$$x = \lim_{t\to\infty} x_t = \lim_{t\to\infty} (P^t x_0 + (\sum_{i=0}^{t-1} P^i) g^-). \tag{4.16}$$

If the even and odd subsequences of x_t converge separately, we denote $x_e = \lim_{t\to\infty} x_{2t}$, $x_o = \lim_{t\to\infty} x_{2t+1}$.

Before presenting the results on long-term dynamics of voter model, we first introduce the following useful lemma connecting a signed digraph G with another graph G' where all edge signs in G are negated.

Lemma 4 *Given a signed digraph $G = (V, E, A)$, let $G' = (V, E, -A)$ be a signed digraph with all edge signs negated from G. Then, for any initial color distribution x_0, at any $2t$ steps ($t > 0$), the color distributions $x_{2t}(G)$ on G and $x_{2t}(G')$ on G' are identical.*

Proof : Let $P' = -P$ denote the signed transition matrix of G', and denote the vector $g^- = D^{-1}A^-\mathbf{1}$ and $g'^- = D^{-1}(-A)^-\mathbf{1} = D^{-1}A^+\mathbf{1}$. Thus $g'^- = \mathbf{1} - g^-$. By Eq.(4.1), after two

steps, we have

$$x_2(G') = P'^2 x_0 + P'g'^- + g'^- = P^2 x_0 - P(\mathbf{1} - g^-) + \mathbf{1} - g^- \\ = P^2 x_0 + Pg^- + g^- = x_2(G),$$

where the last equality uses facts $\mathbf{1} = D^{-1}\bar{A}\mathbf{1}$ and $P = D^{-1}A$. Since the lemma holds for two steps, then clearly it holds for all even steps. ∎

Next theorem discusses the case of ergodic signed digraphs.

Theorem 10 *Let $G = (V, E, A)$ be an ergodic signed digraph, we have*

$$\textit{Balanced } G: \quad x = \hat{\boldsymbol{1}}_S \hat{\pi}_S^T (x_0 - \tfrac{1}{2}\boldsymbol{1}) + \tfrac{1}{2}\boldsymbol{1} \tag{4.17}$$

$$\textit{Strictly unbalanced } G: \quad x = \tfrac{1}{2}\boldsymbol{1} \tag{4.18}$$

$$\textit{Anti-balanced } G: \quad x_e = \hat{\boldsymbol{1}}_S \hat{\pi}_S^T (x_0 - \tfrac{1}{2}\boldsymbol{1}) + \tfrac{1}{2}\boldsymbol{1} \tag{4.19}$$

$$x_o = -\hat{\boldsymbol{1}}_S \hat{\pi}_S^T (x_0 - \tfrac{1}{2}\boldsymbol{1}) + \tfrac{1}{2}\boldsymbol{1} \tag{4.20}$$

Proof :

We discuss the limit in Eq. (4.16) for three possible balance structures of G.

Balanced digraphs. From Lemma 2 and Proposition 5 in Appendix A.1, it is easy to prove $P^m - \hat{\mathbf{1}}_S \hat{\pi}_S^T = (P - \hat{\mathbf{1}}_S \hat{\pi}_S^T)^m$ for any integer $m > 0$, which yields the following result on the second part in Eq. (4.16).

$$\lim_{t\to\infty} \sum_{i=0}^{t-1} P^i g^- = (I - P + \hat{\mathbf{1}}_S \hat{\pi}_S^T)^{-1} g^- + \lim_{t\to\infty} \sum_{i=1}^{t-1} \mathbf{1}_S \hat{\pi}_S^T g^- \tag{4.21}$$

$$= (I - P + \hat{\mathbf{1}}_S \hat{\pi}_S^T)^{-1} g^- = \frac{1}{2}\mathbf{1} - \frac{1}{2}\hat{\mathbf{1}}_S \hat{\pi}_S^T \mathbf{1}, \tag{4.22}$$

where the last term of Eq.(4.21) is canceled out due to the digraph flow circulation law [3, 15], i.e.,

$$\hat{\pi}_S^T g^- = \hat{\pi}_S^T D^{-1} A^- \mathbf{1} = \sum_{i\in S} \pi(i) \sum_{j\in \bar{S}} \bar{P}_{ij} - \sum_{i\in \bar{S}} \pi(i) \sum_{j\in S} \bar{P}_{ij} = 0.$$

The last equality in Eq.(4.22) holds because

$$\frac{1}{2}(I - P + \hat{\mathbf{1}}_S \hat{\pi}_S^T)(\mathbf{1} - \hat{\mathbf{1}}_S \hat{\pi}_S^T \mathbf{1}) - g^- = 0.$$

Eq.(4.17) is obtained by combining Eq.(4.22) with Lemma 2.

Anti-balanced Digraphs. Lemma 4 directly yields Eq.(4.19). The odd step influence distribution sequence is obtained by

$$x_o = Px_e + g^- = -\mathbf{1}_S \hat{\pi}_S^T (x_0 - \frac{1}{2}\mathbf{1}) + \frac{1}{2}\mathbf{1}.$$

Strictly unbalanced digraphs. From Theorem 2, $\lim_{t\to\infty} P^t = \mathbf{0}$ holds and thus we have

$$\lim_{t\to\infty} \sum_{i=0}^{t-1} P^i g^- = (I-P)^{-1} g^- = (D-A)^{-1} A^- \mathbf{1} = \frac{1}{2}\mathbf{1}. \tag{4.23}$$

The last equality comes from the facts $(D-A)\mathbf{1} = 2A^-\mathbf{1}$. ∎

Theorem 10 has several implications. First of all, for strictly unbalanced digraphs, each node has equal steady state probability of being black or white, and it is not determined by the initial distribution x_0. Secondly, anti-balanced digraphs has the same steady state distribution as the corresponding balanced graph for even steps, and for odd steps, the distribution oscillates to the opposite ($x_o = \mathbf{1} - x_e$). Moreover, Eq.(4.17) can also be intuitively explained from the random walk interpretation of the voter model. In particular, starting from node i, if we perform a random walk for an infinite number of steps, the probability that the random walk stops at j is given by the stationary distribution $\pi(j)$. For balanced graphs, if i and j are from the same component (either S or $\bar{S}$), then the random walk must pass an even number of negative edges, so i takes the same color as j; if i and j are from opposite components, then the walk passes an odd number of negative edges and i takes the opposite of j's color. Thus, the steady distribution of $i \in S$ being white is given by $\pi_S^T x_{0S} + \pi_{\bar{S}}^T(\mathbf{1}_{\bar{S}} - x_{0\bar{S}})$, and the case of $i \in \bar{S}$ is symmetric. Some algebra manipulations can lead us to Eq.(4.17).

For a balanced ergodic digraph G with partition $S, \bar{S}$, it is easy to check that it has the following two equilibrium states: in one state all nodes in S are white while all nodes in $\bar{S}$ are black; and in the other state all nodes in S are black while all nodes in $\bar{S}$ are white. We call these two states the *polarized states*. Using random walk interpretation, we show in the following theorem that with probability 1, the voter model dynamic converges to one of the above two equilibrium states.

Theorem 11 *Given an ergodic signed digraph $G = (V, E, A)$, if G is balanced with partition $S, \bar{S}$, the voter model dynamic converges to one of the polarized states with probability* 1*, and*

the probability of nodes in S being white is $\hat{\pi}_S^T(x_0 - \frac{1}{2}\mathbf{1}) + \frac{1}{2}$. Similarly, if G is anti-balanced, with probability 1 the voter model dynamic oscillates between the two polarized states eventually, and the probability of nodes in S being white at even steps is $\hat{\pi}_S^T(x_0 - \frac{1}{2}\mathbf{1}) + \frac{1}{2}$.

Proof :

Consider a balanced ergodic digraph G with partition $S, \bar{S}$. By ergodicity, given any two nodes i and j, with probability 1 the random walks starting from i and j will meet eventually. If i and j are both in S, when the two walks meet at some node u, they both pass either an even number of negative edges (if $u \in S$) or an odd number of negative edges (if $u \in \bar{S}$). Therefore, i and j must be in the same color with probability 1. If i and j are from different components S and $\bar{S}$, a similar argument shows that they will have the opposite color with probability 1. Therefore the final state is one of the two polarized states. The probability of nodes in S being white is simply given by Theorem 10, Eq.(4.17). The case of anti-balanced ergodic digraphs can be argued in a similar way. ∎

Theorem 12 below introduces the long-term dynamics of the weakly connected signed digraphs. We consider weakly connected G with a single sink ergodic component G_Z, and use the same notations as in Section 4.3.2.

Theorem 12 *Let $G = (V, E, A)$ be a weakly connected signed digraph with a single sink component G_Z and a non-sink component G_X. The long-term white color distribution vector x is expressed in two parts:*

$$x^T = \lim_{t\to\infty} x_t^T = [x_{XY}^T, x_Z^T].$$

where x_Z is the limit of x_{tZ} on G_Z with initial distribution x_{0Z} and is given as in Theorem 10, and vector x_{XY} is given below with respect to the balance structure of G_Z:

$$\begin{aligned}
\textit{Balanced } G_Z\textit{:} \quad & x_{XY} = \tfrac{1}{2}\mathbf{1}_X + u_b\hat{\pi}_{Z,S_Z}^T(x_{0Z} - \tfrac{1}{2}\mathbf{1}_Z) \\
\textit{Strictly unbalanced } G_Z\textit{:} \quad & x_{XY} = \tfrac{1}{2}\mathbf{1}_X \\
\textit{Anti-balanced } G_Z\textit{, even } t\textit{:} \quad & x_{XY,e} = \tfrac{1}{2}\mathbf{1}_X - u_u\hat{\pi}_{Z,S_Z}^T(x_{0Z} - \tfrac{1}{2}\mathbf{1}_Z) \\
\textit{Anti-balanced } G_Z\textit{, odd } t\textit{:} \quad & x_{XY,o} = \tfrac{1}{2}\mathbf{1}_X + u_u\hat{\pi}_{Z,S_Z}^T(x_{0Z} - \tfrac{1}{2}\mathbf{1}_Z)\ ,
\end{aligned}$$

where u_b and u_u are defined in Eq.(4.12) and Eq.(4.13).

Proof : Let initial distribution $x_0^T = [x_{0X}^T, x_{0Z}^T]$ and $g^{-T} = [g_X^{-T}, g_Z^{-T}]$. When $t \to \infty$, Eq. (4.1) can be written as

$$x^T = \lim_{t\to\infty}(P^t x_0)^T = [x_{XY}^T, x_Z^T] = [x_X^T + x_Y^T, x_Z^T],$$

where $x_X = \lim_{t\to\infty}(P_X^t x_{0X} + \sum_{i=0}^{t-1} P_X^i g_X^-)$, $x_Y = \lim_{t\to\infty}(P_Y^{(t)} x_{0Z} + \sum_{i=0}^{t-1} P_Y^{(i)} g_Z^-)$, and $x_Z = \lim_{t\to\infty}(P_Z^t x_{0Z} + \sum_{i=0}^{t-1} P_Z^i g_Z^-)$.

From Lemma 3, $\lim_{t\to\infty} P_X^t = \mathbf{0}$, thus $x_X = (I_X - P_X)^{-1} g_X^-$ holds for any ergodic G_Z. Since G_Z is ergodic, x_Z follows Theorem 10. Below we will focus on deriving x_Y, where the first part of x_Y satisfies Lemma 3, i.e.,

$$\lim_{t\to\infty} P_Y^{(t)} x_{0Z} = \begin{cases} \mathbf{0} & G_Z \text{ is strictly unbalanced} \\ u_b \hat{\pi}_{Z,S_Z}^T x_{0Z} & G_Z \text{ is balanced} \\ -u_u \hat{\pi}_{Z,S_Z}^T x_{0Z} & G_Z \text{ is anti-balanced, even } t \\ u_u \hat{\pi}_{Z,S_Z}^T x_{0Z} & G_Z \text{ is anti-balanced, odd } t. \end{cases}$$

The second part of x_Y can be further written down as

$$\begin{aligned} \lim_{m\to\infty} \sum_{t=1}^{m} P_Y^{(t)} g_Z^- &= \lim_{m\to\infty} \sum_{t=0}^{m-1} \sum_{i=0}^{t} (P_X^{t-i} P_Y P_Z^i) g_Z^- \\ &= \lim_{m\to\infty} \sum_{t=0}^{m-1} \sum_{i=0}^{m-t} (P_X^t P_Y P_Z^i) g_Z^- = \sum_{t=0}^{\infty} (P_X^t P_Y \sum_{i=0}^{\infty} P_Z^i) g_Z^- \end{aligned} \tag{4.24}$$

Now we discuss Eq.(4.24) under different balance structures of G_Z.

(1) G_Z **is strictly unbalanced.** From Lemma 3, $\lim_{t\to\infty} P^t = \mathbf{0}$. Then by Eq.(4.23) we directly obtain that $x_{XY} = \frac{1}{2}\mathbf{1}_X$. Applying Eq.(4.23) to $\sum_{i=0}^{\infty} P_Z^i g_Z^-$ in Eq.(4.24), we have

$$\lim_{m\to\infty} \sum_{t=1}^{m} P_Y^{(t)} g_Z^- = \frac{1}{2}(I_X - P_X)^{-1} P_Y \mathbf{1}_Z.$$

Thus, we obtain the following equation:

$$x_{XY} = x_X + x_Y = (I_X - P_X)^{-1}(g_X^- + \frac{1}{2} P_Y \mathbf{1}_Z) = \frac{1}{2}\mathbf{1}_X.$$

(2) G_Z **is balanced.** Using Eq.(4.22), we have

$$\lim_{m\to\infty} \sum_{t=1}^{m} P_Y^{(t)} g_Z^- = \frac{1}{2}(I_X - P_X)^{-1} P_Y (\mathbf{1}_Z - \hat{\mathbf{1}}_{Z,S_Z} \hat{\pi}_{Z,S_Z}^T \mathbf{1}_Z).$$

Hence, we have

$$x_{XY} = (I_X - P_X)^{-1}(g_X^- + \frac{1}{2}P_Y\mathbf{1}_Z) + u_b\hat{\pi}_{Z,S_Z}^T(x_{0Z} - \frac{1}{2}\mathbf{1}_Z)$$
$$= \frac{1}{2}\mathbf{1}_X + u_b\hat{\pi}_{Z,S_Z}^T(x_{0Z} - \frac{1}{2}\mathbf{1}_Z). \tag{4.25}$$

(3) G_Z **is anti-balanced.** Using Lemma 4, we can negate the signs of all edges in G so that the sink becomes balanced. Hence, we know that at even steps in long term,

$$x_{XY,e} = \frac{1}{2}\mathbf{1}_X - u_u\hat{\pi}_{Z,S_Z}^T(x_{0Z} - \frac{1}{2}\mathbf{1}_Z), \tag{4.26}$$

where Eq.(4.26) and Eq.(4.25) are identical in the sense that P_X's and P_Y's in Eq.(4.26) and Eq.(4.25) have opposite signs. Moreover, the odd step influence distribution sequence is obtained

$$x_{XY,o} = P_X x_{XY,e} + P_Y x_{Z,e} + g_X^- = \frac{1}{2}\mathbf{1}_X + u_u\hat{\pi}_{Z,S_Z}^T(x_{0Z} - \frac{1}{2}\mathbf{1}_Z). \tag{4.27}$$

■

Theorem 12 characterizes the long-term dynamics when the underlying graph is a weakly connected signed digraph with one ergodic sink component. We can see that the results for balanced and anti-balanced sink components are more complicated than the ergodic digraph case, since how non-sink components are connected to the sink subtly affects the final outcome of the steady state behavior. In steady state, while the sink component is still in one of the two polarized states as stated in Theorem 11, the non-sink components exhibit more complicated color distribution, for which we provide probability characterizations in Theorem 12. Using Eq.(4.15), Theorem 10 and Theorem 12 can be readily extended to the case with more than one ergodic sink components and disconnected digraphs. When the network only contains positive directed edges, the voter model dynamics can be interpreted using digraph random walk theory [3–5, 8].

4.4 Influence maximization

With the detailed analysis on voter model dynamics for signed digraphs, we are ready now to solve the influence maximization problem. Intuitively, we want to address the following question: *If only at most k nodes could be selected initially and be turned white while all other nodes are black, how should we choose seed nodes so as to maximize the expected number of white nodes in short term and in long term, respectively?*

4.4.1 Influence maximization problem

Influence maximization objectives. We consider two types of short-term influence objectives, one is the *instant influence*, which counts the total number of influenced nodes at a step $t > 0$; the other is the *average influence*, which takes the average number of influenced nodes within the first t steps. These two objectives have different implications and applications. For example, political campaigns try to convince voters who may change their minds back and forth, but only the voters' opinions on the voting day are counted, which matches the *instant influence*. On the other hand, a credit card company would like to have customers keep using its credit card service as much as possible, which is better interpreted by the *average influence*. When t is sufficiently large, it becomes the long-term objective, and long-term average influence coincides with long-term instant influence when the dynamic converges.

Formally, we define the *short-term instant influence* $f_t(x_0)$ and the *short-term average influence* $\bar{f}_t(x_0)$ as follows:

$$f_t(x_0) := \mathbf{1}^T x_t(x_0) \text{ and } \bar{f}_t(x_0) := \frac{\sum_{i=0}^{t} f_i(x_0)}{t+1}. \tag{4.28}$$

Moreover, we define *l*ong term influence as

$$f(x_0) := \lim_{t\to\infty} \frac{\sum_{i=0}^{t} f_i(x_0)}{t+1}. \tag{4.29}$$

Note that when the dynamic converges (e.g. ergodic balanced or ergodic strictly unbalanced graphs), $f(x_0) = \lim_{t\to\infty} f_t(x_0)$. For ergodic anti-balanced graphs (or sink components), it is essentially the average of even- and odd-step limit influence.

Given a set $W \subseteq V$, Let e_W be the vector in which $e_W(j) = 1$ if $j \in W$ and $e_W(j) = 0$ if $j \notin W$, which represents the initial seed distribution with only nodes in W as white seeds. Let e_i be the shorthand of $e_{\{i\}}$. Unlike unsigned graphs, if initially no white seeds are selected on a signed digraph G, i.e., $x_0 = \mathbf{0}$, the instant influence $f_t(\mathbf{0})$ at step t is in general non-zero, which is referred to as the *ground influence* of the graph G at t. The influence contribution of a seed set W does not count such ground influence, as shown in definition 4.

Definition 4 (Influence contribution) *The* instant influence contribution *of a seed set W to the t-th step instant influence objective, denoted by $c_t(W)$, is the difference between the instant influence at step t with only nodes in W selected as seeds and the ground influence at step t: $c_t(W) = f_t(e_W) - f_t(\boldsymbol{0})$. The* average influence contribution $\bar{c}_t(W)$ *and* long-term influence

contribution $c(W)$ *are defined in the same way:* $\bar{c}_t(W) = \bar{f}_t(e_W) - \bar{f}_t(\boldsymbol{0})$ *and* $c(W) = f(e_W) - f(\boldsymbol{0})$.

We are now ready to formally define the influence maximization problem.

Definition 5 (Influence maximization) *The* influence maximization *problem for short-term instant influence is finding a seed set W of at most k seeds that maximizes W's instance influence contribution at step t, i.e., finding $W_t^* = \arg\max_{|W|\leq k} c_t(W)$. Similarly, the problem for average influence and long-term influence is finding $\bar{W}_t^* = \arg\max_{|W|\leq k} \bar{c}_t(W)$ and $W^* = \arg\max_{|W|\leq k} c(W)$, respectively.*

We now provide some properties of influence contribution, which lead to the optimal seed selection rule. By Eq.(4.1), we have

$$c_t(W) = f_t(e_W) - f_t(\mathbf{0}) = \mathbf{1}^T x_t(e_W) - \mathbf{1}^T x_t(\mathbf{0}) = \mathbf{1}^T P^t e_W. \tag{4.30}$$

Let $c_t(i)$ be the shorthand of $c_t(\{i\})$, and let $c_t = [c_t(i)]$ denote the vector of influence contribution of individual nodes. Then $c_t^T = [c_t(i)]^T = \mathbf{1}^T P^t$. When $t \to \infty$, the long term influence contributions of individual nodes are obtained as a vector c:

$$c^T = \lim_{t\to\infty} \frac{\sum_{i=0}^t c_i^T}{t+1} = \lim_{t\to\infty} \frac{\mathbf{1}^T \sum_{i=0}^t P^i}{t+1}. \tag{4.31}$$

When P^t converges, we simply have

$$c^T = \mathbf{1}^T \lim_{t\to\infty} P^t. \tag{4.32}$$

Lemma 5 below discloses the important property that the influence contribution is a linear set function.

Lemma 5 *Given a white seed set W, $c_t(W) = \sum_{i\in W} c_t(i)$, $\bar{c}_t(W) = \sum_{i\in W} \bar{c}_t(i)$, and $c(W) = \sum_{i\in W} c(i)$.*

Proof : From Eq.(4.30), we have

$$c_t(W) = \mathbf{1}^T P^t e_W = \mathbf{1}^T P^t \sum_{i\in W} e_i = \sum_{i\in W} \mathbf{1}^T P^t e_i = \sum_{i\in W} c_t(i).$$

The linearity of $\bar{c}_t$ and c can be derived from that of c_t. ∎

Given a vector v, let $n^+(v)$ denote the number of positive entries in v. By applying Lemma 5, we have the optimal seed selection rule for instant influence maximization as follows.

Optimal seed selection rule for instant influence maximization. *Given a signed digraph and a limited budget k, selecting top $\min\{k, n^+(c_t)\}$ seeds with the highest $c_t(i)$'s, $i \in V$, leads to the maximized instant influence at step $t > 0$.*

Note that the influence contributions of some nodes may be negative and these nodes should not be selected as white seeds, and thus the optimal solution may have less than k seeds. The rules for average influence maximization and long-term influence maximization are patterned in the same way. Therefore, the central task now becomes the computation of the influence contributions of individual nodes. Below, we will introduce our SVIM algorithm, for Signed Voter model Influence Maximization.

4.4.2 Short-term influence maximization

By applying Definition 4 and Lemma 5, we develop SVIM-S algorithm to solve the short-term instant and average influence maximization problem, as shown in Algorithm 1.

Algorithm 1 Short-term influence maximization SVIM-S

1: **INPUT:** Signed transition matrix P, short-term period t, budget k;
2: **OUTPUT:** White seed set W.
3: $c_t = \mathbf{1}$; $\bar{c}_t = \mathbf{1}$;
4: **for** $i = 1 : t$ **do**
5: $\quad c_t^T = c_t^T P$;(for instant influence maximization.)
6: $\quad \bar{c}_t = \bar{c}_t + c_t$; (for average influence maximization.)
7: $W =$ top $\min\{k, n^+(c_t)\}$ (resp. $\min\{k, n^+(\bar{c}_t)\}$) nodes with the highest $c_t(i)$ (resp. $\bar{c}_t(i)$) values, for instant (resp. average) influence maximization.

SVIM-S algorithm requires t vector-matrix multiplications, each of which takes $|E|$ times entry-wise multiplication operations. Hence the total time complexity of SVIM-S is $O(t \cdot |E|)$.

4.4.3 Long-term influence maximization

We now study the long-term influence contribution c and introduce the corresponding influence maximization algorithm SVIM-L. We will see that the computation of influence contribution c and seed selection schemes depends on the structural balance and connectedness of the graph.

While seed selection for balanced ergodic digraphs still has intuitive explanations, the computation for weakly connected and disconnected digraphs is more involved and less intuitive.

Case of ergodic signed digraphs

When the signed digraph $G = (V, E, A)$ is ergodic, Lemma 6 below characterizes the long-term influence contributions of nodes, with respect to various balance structures.

Lemma 6 *Consider an ergodic signed digraph $G = (V, E, A)$. If G is* balanced*, with bipartition S and $\bar{S}$, the influence contribution vector $c = (|S| - |\bar{S}|)\hat{\pi}_S$. If G is* anti-balanced *or* strictly unbalanced*, $c = \mathbf{0}$.*

Proof : (1) When G is balanced, by Lemma 2 and Eq.(4.32),

$$c^T = \mathbf{1}^T \lim_{t\to\infty} P^t = \mathbf{1}^T \hat{\mathbf{1}}_S \hat{\pi}_S^T = (|S| - |\bar{S}|)\hat{\pi}_S.$$

(2) When G is strictly unbalanced, again by Lemma 2 and Eq.(4.32), we have $c^T = \mathbf{1}^T \lim_{t\to\infty} P^t = \mathbf{0}$.

(3) When G is anti-balanced, by Lemma 2 and Eq.(4.31), we have

$$c^T = \mathbf{1}^T \frac{\lim_{t\to\infty} P^{2t} + \lim_{t\to\infty} P^{2t+1}}{2} = \mathbf{0}.$$

■

Based on Lemma 6, Algorithm 2 summarizes how to compute the long-term influence contribution c on ergodic signed digraphs.

Algorithm 2 $c = ergodic(G)$

1: **INPUT:** Signed transition matrix P.
2: **OUTPUT:** Long term influence contribution vector c
3: Detect the structure of ergodic signed digraph G;
4: **if** G is balanced, with bipartition S and $\bar{S}$ **then**
5: Compute stationary distribution π of $\bar{P}$;
6: $c = (|S| - |\bar{S}|)\hat{\pi}_S$;
7: **else**
8: $c = \mathbf{0}$;

Lemma 6 suggests that for ergodic balanced digraphs, we should pick the larger component, e.g., S, if $|S| > |\bar{S}|$, and select the top $\min\{k, |S|\}$ nodes from S with the largest stationary

distributions as white seeds. Selecting these nodes will make the probability of the larger component being white the largest.

Theorem 10 indicates that given an anti-balanced digraph G, with bipartition S and $\bar{S}$, the long-term dynamic x_t oscillates on odd and even steps, and their long-term influence contribution is 0. However, we can still maximize the strength of the oscillation of the voter model on an anti-balanced ergodic digraph by properly choosing the initial white seeds (See Remark 1.)

Remark 1 *In an anti-balanced ergodic digraph $G = (V, E, A)$ with the bipartition S and $\bar{S}$ and a budget k. Let W' (resp. W'') denote two initial seed sets, where $\min\{k, |S|\}$ (resp. $\min\{k, |\bar{S}|\}$) nodes, with highest stationary distribution $\pi(i)$'s in S (resp. $\bar{S}$), are selected. Then, the optimal W^* that maximizes the strength of oscillation is*

$$W^* := \text{argmax}_{W \in \{W', W''\}} |\hat{\pi}_S^T (e_W - \frac{1}{2}\boldsymbol{I})|. \tag{4.33}$$

Proof : From Theorem 10, when t becomes sufficiently large, the vector x oscillates at two vectors on odd and even steps, respectively. The strength of the oscillation is

$$\begin{aligned}\frac{|f_o(x_0) - f_e(x_0)|}{2} = |\mathbf{1}^T \frac{x_o(x_0) - x_e(x_0)}{2}| &= |\mathbf{1}^T \hat{\mathbf{1}}_S \hat{\pi}_S^T (x_0 - \frac{1}{2}\mathbf{1})| \\ &= ||S| - |\bar{S}|| \cdot |\hat{\pi}_S^T (x_0 - \frac{1}{2}\mathbf{1})|.\end{aligned}$$

Let W be the initial seed set, then the oscillation strength maximization is formulated as

$$\begin{aligned}&\max_{|W| \le k} ||S| - |\bar{S}|| \cdot |\hat{\pi}_S^T (e_W - \frac{1}{2}\mathbf{1})| \\ &= ||S| - |\bar{S}|| \cdot \max\{\max_{|W| \le k} \{\hat{\pi}_S^T e_W\} - \frac{1}{2}\hat{\pi}_S^T \mathbf{1}, \max_{|W| \le k} \{-\hat{\pi}_S^T e_W\} + \frac{1}{2}\hat{\pi}_S^T \mathbf{1}\},\end{aligned} \tag{4.34}$$

which contains two sub-problems, i.e., $\max_{|W| \le k}\{\hat{\pi}_S^T e_W\}$ and $\max_{|W| \le k}\{-\hat{\pi}_S^T e_W\}$. The first maximization problem can be rewritten as

$$\max_{|W| \le k} \{\hat{\pi}_S^T e_W\} = \max_{|W| \le k} \big(\sum_{i \in S} \pi(i) e_W(i) - \sum_{j \in \bar{S}} \pi(i) e_W(j) \big). \tag{4.35}$$

Thus, let W' denote the optimal solution to the problem in Eq.(4.35), which is obtained by choosing $\min\{k, |S|\}$ seeds with highest $\pi(i)$'s from S. Similarly, choosing $\min\{k, |\bar{S}|\}$ nodes

with the highest $\pi(i)$'s from $\bar{S}$ yields the optimal solution, denoted by W'', to the second maximization problem $\max_{|W|\leq k}\{-\hat{\pi}_S^T e_W\}$. The optimal W to the problem in eq.(4.34) that maximizes the oscillation strength is the one in $\{W', W''\}$, with higher $|\hat{\pi}_S^T(e_W - \frac{1}{2}\mathbf{1})|$, which completes the proof of eq.(4.33).

■

Case of weakly connected signed digraphs

We first consider a weakly connected signed G which has a single ergodic sink component G_Z with only incoming edges from the remaining nodes $X = V \setminus Z$.

Lemma 7 *Consider a weakly connected digraph $G = (V, E, A)$ with a single ergodic sink component G_Z. If G_Z is* balanced*, with partition S_Z and $\bar{S_Z}$, the long term influence contribution vector $c^T = [c_X^T, c_Z^T]$, where $c_X = \mathbf{0}_X$ and $c_Z = (\mathbf{1}_X^T u_b + |S_Z| - |\bar{S}_Z|)\hat{\pi}_{Z,S_Z}$. If G is* anti-balanced *or* strictly unbalanced*, $c = \mathbf{0}$.*

Proof : (1)When G_Z is balanced, by Lemma 3, $c_X = \mathbf{0}_X$, and

$$c_Z^T = (\mathbf{1}_X^T u_b + \mathbf{1}_Z^T \hat{\mathbf{1}}_{Z,S_Z})\hat{\pi}_{Z,S_Z}^T = (\mathbf{1}_X^T u_b + |S_Z| - |\bar{S}_Z|)\hat{\pi}_{Z,S_Z}^T.$$

(2) When G_Z is strictly unbalanced, $c^T = \mathbf{1}^T \lim_{t\to\infty} P^t = \mathbf{0}$

(3) When G_Z is anti-balanced, by Lemma 3 the limits of odd and even subsequences of P^t cancel out, thus $c = \mathbf{0}$. ■

Lemma 7 indicates that influence contribution of the balanced ergodic sink component is more complicated than that of the balanced ergodic digraph. This is because the sink component affects the colors of the non-sink component in a complicated way depending on how non-sink and sink components are connected. Therefore, the optimal seed selection depends on the calculation of the influence contributions of each sink node, and is not as intuitive as that for the ergodic digraph case.

Theorem 12 shows that in a weakly connected signed digraph G, with single anti-balanced sink component G_Z, the long term influence $f(x_0)$ oscillates on odd and even steps, and the average is $|V|/2$, which is invariant to the initial seed selection. Similar to Remark 1, we can

maximize the oscillation strength by properly selecting initial seeds, i.e.,

$$\begin{aligned} W^* &= \text{argmax}_{|W|\leq k} \, |f_e(e_W) - f_o(e_W)|/2 \\ &= \text{argmax}_{|W|\leq k} \, |(\mathbf{1}_X^T u_u \hat{\pi}_{Z,S_Z}^T + \mathbf{1}_Z^T \hat{\mathbf{1}}_{Z,S_Z} \hat{\pi}_{Z,S_Z}^T)(e_{WZ} - \frac{1}{2}\mathbf{1}_Z)| \\ &= |\mathbf{1}_X^T u_u + |S_Z| - |\bar{S}_Z|| \cdot \text{argmax}_{|W|\leq k} \, |\hat{\pi}_{Z,S_Z}^T (e_{WZ} - \frac{1}{2}\mathbf{1}_Z)| \end{aligned} \tag{4.36}$$

where the maximization objective is independent from x_{0X}, thus oscillation strength maximization problem objective in Eq.(4.36) for G is identical to that in Remark 1. Hence, Remark 1 also applies here.

Using Eq.(4.15), Lemma 6 and Lemma 7 can be readily extended to the case with more than one ergodic sink components and disconnected digraphs. Algorithm 3 below summarizes how to compute the node influence contributions of weakly connected signed digraphs. Note that by our assumption, we consider all sink components to be ergodic.

Algorithm 3 $c = weakly(G)$

INPUT: Signed transition matrix P.
OUTPUT: Influence contribution vector c.
Detect the structure of the weakly connected signed digraph G, and find its $m \geq 1$ signed ergodic sink components $G_{Z1}, \cdots, G_{Zm}$;
for $i = 1 : m$ **do**
  if G_{Zi} is balanced with partition $S_{Zi}, \bar{S}_{Zi}$ **then**
    Compute stationary distribution π_{Zi} of $\bar{P}_{Zi}$;
    $u_{bi} = (I_X - P_X)^{-1} P_{Yi} \hat{\mathbf{1}}_{Zi,S_{Zi}}$;
    $c_{Zi} = (\mathbf{1}_X^T u_{bi} + |S_{Zi}| - |\bar{S}_{Zi}|) \hat{\pi}_{Zi,S_{Zi}}^T$;
$c = [\mathbf{0}_X; c_{Z1}; \cdots; c_{Zm}]$

General case and SVIM-L algorithm

Given the above systematic analysis, we are now in a position to summarize and introduce our SVIM-L algorithm which solves the long-term voter model influence maximization problem for general aperiodic signed digraphs.

In general, a signed digraph consists $m \geq 1$ disconnected components, within each of which the node influence contribution follows Lemma 7. The long-term signed voter model influence maximization (SVIM-L) algorithm is constructed in Algorithm 4.

Complexity analysis. We consider $G = (V, E, A)$ to be weakly connected, since disconnected graph case can be treated independently for each connected component for the time complexity.

Algorithm 4 Long-term influence maximization SVIM-L

INPUT: Signed transition matrix P, budget k.
OUTPUT: White seed set W.
Detect the structure of a general aperiodic signed digraph G, and find the $m \geq 1$ disconnected components $G_1, \cdots, G_m$;
for $i = 1 : m$ **do**
    $c_{G_i} = weakly(G_i)$;
$c = [c_{G_1}; \cdots; c_{G_m}]$;
$W =$ top $\min\{k, n^+(c)\}$ nodes with the highest $c(i)$ values.

SVIM-L algorithm consists of two parts. The first part extracts the connectivity and balance structure of the graph, which can be done using depth-first search with complexity $O(|E|)$. The second part uses Algorithm 3 to compute influence contributions of balanced ergodic sink components. The dominant computations are on the stationary distribution π_{Zi}'s and $(I_X - P_X)^{-1}$, which can be done by solving a linear equation system [86] and matrix inverse in $O(|Z_i|^3)$ and $O(n_X^3)$, respectively, where $n_X = |X|$. Let b be the number of balanced sink components in G, n_Z be the number of nodes in the largest balanced sink component. Thus SVIM-L can be done in $O(bn_Z^3 + n_X^3)$ time. Alternatively, we can use iterative method for computing both π_{Zi}'s and $\mathbf{1}_X^T(I_X - P_X)^{-1}$, if the largest convergence time t_C of P_{Zi}^t's and P_X^t is small. (Note that the convergence time of ergodic digraphs could be exponentially large in general, as illustrated by an example in Appendix A.3). In this case, each iteration step involves vector-matrix multiplication and can be done in $O(m_B)$ time, where m_B is the number of edges of the induced subgraph G_B consisting of all nodes in the balanced sink components and X. Note that m_B and t_C are only related to subgraph G_B, which could be significantly smaller than G, and thus $O(t_C m_B)$ could be much smaller than the time of naive iterations on the entire graph. Overall SVIM-L can be done in $O(|E| + \min(bn_Z^3 + n_X^3, t_C m_B))$ time.

4.5 Evaluation

In this section, we first use both synthetic datasets and real social network datasets to demonstrate the efficacy of our short-term and long-term seed selection schemes by comparing the performances with four baseline heuristics. Then, we evaluate how much the short-term and long-term influence can be improved by taking the edge signs into consideration.

4.5.1 Performance comparison with baseline heuristics

For different scenarios, we compare our SVIM-L and SVIM-S algorithms with four heuristics, i.e., (1) selecting seed nodes with the highest weighted outgoing degrees (denoted by $d^+ + d^-$ in the figures), (2) highest weighted outgoing positive degrees (denoted by d^+), (3) highest differences between weighted outgoing positive and negative degrees (denoted by $d^+ - d^-$), and (4) randomly selecting seed nodes (denoted by "Rand"), where in our evaluations, we run random seed selection 1000 times, and compare the average number of white nodes between our algorithm and other heuristics. Our evaluation results demonstrate that our seed selection scheme can increase up to 72% long-term influence, and 145% short-term influence over other heuristics.

Synthetic datasets

In this part, we generate synthetic datasets with different structures to validate our theoretical results.

Dataset generation model. We generate six types of signed digraphs, including balanced ergodic digraphs, anti-balanced ergodic digraphs, strictly unbalanced ergodic digraphs, weakly connected signed digraphs, disconnected signed digraphs with ergodic components, and disconnected signed digraph with weakly connected components (WCCs). All edges have unit weights. The following are graph configuration details.

We first create an unsigned ergodic digraph $\bar{G}$ with 9500 nodes, which has two ergodic components $\bar{G}_A$ and $\bar{G}_B$, with $[3000, 6500]$ nodes and $[3000, 6500] \times 8$ random directed edges, respectively. Moreover, there are 3000×8 random directed edges across $\bar{G}_A$ and $\bar{G}_B$. Ergodicity is checked through a simple connectivity and aperiodicity check. Given $\bar{G}$, a *balanced digraph* is obtained by assigning all edges within $\bar{G}_A$ and $\bar{G}_B$ with positive signs, and those across them with negative signs. Then, an *anti-balanced digraph* is generated by negating all edge signs of the balanced ergodic digraph. To generate a *strictly unbalanced digraph*, we randomly assign edge signs to all edges in $\bar{G}$ and make sure that there does not exist a balanced or anti-balanced bipartition.

Moreover, we generated a *disconnected signed digraph* and a weakly connected signed digraph for our study. We first generate 5 ergodic unsigned digraphs, $\bar{G}_1, \cdots, \bar{G}_5$ with $[500, 200, 800, 300, 2700]$ nodes and $[500, 200, 800, 300, 2700] \times 8$ edges, respectively. Then, we group $G_{23} = (G_2, G_3)$

and $G_{45} = (G_4, G_5)$ to form two ergodic balanced digraphs, and generate a strictly unbalanced ergodic digraph G_1 by randomly assigning signs to edges in $\bar{G}_1$. Three disconnected components G_1, G_{23}, G_{45} together form a disconnected signed digraph. To form a *weakly connected signed digraph*, we place in total 3000 random direct edges from G_1 to the balanced ergodic components G_{23} and G_{45}, where the nodes in subgraph G_1 only have outgoing edges to G_{23} and G_{45}. Moreover, we combine the above generated balanced ergodic digraph and the weakly connected signed digraph together forming a larger *disconnected signed digraph, with the weakly connected signed digraph as a component*.

Fig. 4.1-Fig. 4.6 present the evaluation results for one set of digraphs, where we observe that all digraphs we randomly generated exhibit consistent results. Our tests are conducted using Matlab on a standard PC server.

Long-term influence maximization. In the evaluations, we set the influence budget as $k = 500$, and compare the average numbers of white nodes over steps between our algorithm and other heuristics. Fig. 4.1 shows that in the balanced ergodic digraph, SVIM-L algorithm achieves the highest long-term influence over other heuristics. When applying a heuristic seed selection scheme, denoted by $\mathtt{H}$, $f_t^{\mathtt{H}}$ represents the number of white nodes at step $t(\geq 1)$. Similarly, denote $f_t^{\mathtt{SVIM}}$ as the number of white nodes at step $t(\geq 1)$ for SVIM algorithm. We consider $\Delta f_t(\mathtt{SVIM}, \mathtt{H}) = (f_t^{\mathtt{SVIM}} - f_t^{\mathtt{H}})/f_t^{\mathtt{H}}$ as the influence increase of SVIM over the heuristic algorithm $\mathtt{H}$ at step t. The maximum influence increase is the maximum $\Delta f_t(\mathtt{SVIM}, \cdot)$ among all steps ($t \geq 1$) and all heuristics. Hence, in Fig. 4.1, we see that our SVIM-L algorithm outperforms all other heuristics. Especially, a maximum of 14% influence increase is observed for $t \geq 4$ with $4.68k$ and $4.1k$ white nodes for SVIM-L and random selection scheme, respectively. In the rest of this section, we will use the maximum influence increase as a metric to illustrate the efficacy of our SVIM algorithm. Fig. 4.2 shows the clear oscillating behavior on the anti-balanced ergodic digraph, and the average influence is the same for all algorithms. The inset shows that our algorithm (denoted as "Max. Osc.") indeed provides the largest oscillation. Fig. 4.3 shows the results in strictly unbalanced graph case, where the long-term influences of all algorithms converge to $4750 = |V|/2$, which matches Theorem 10. Fig. 4.4 and Fig. 4.5 show that SVIM-L algorithm performs the best, and it generates $5.6\% - 72\%$ long-term influence increases after the sixth step over other heuristics in the weakly connected signed digraph and the disconnected signed digraph. Fig. 4.6 shows that in a more general signed digraph, which consists of a weakly connected signed component and a balanced ergodic component,

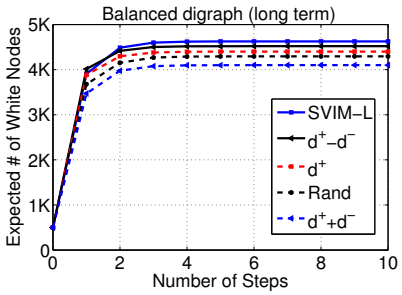


Figure 4.1: G is balanced

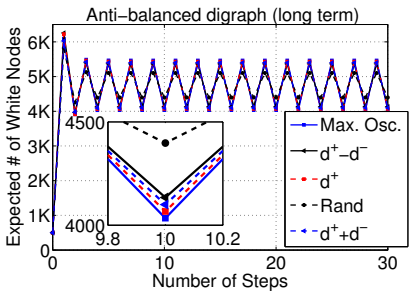


Figure 4.2: G is anti-balanced

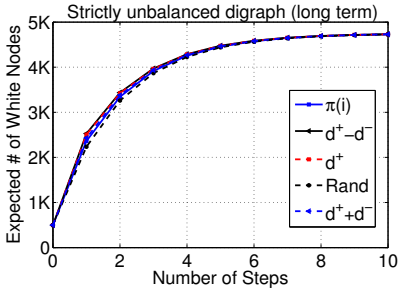


Figure 4.3: G is strictly unbalanced

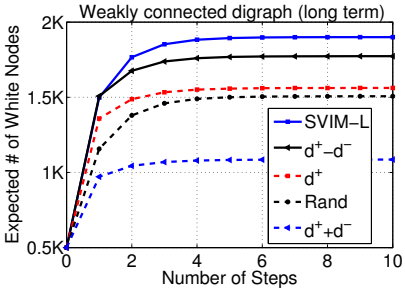


Figure 4.4: G is weakly connected

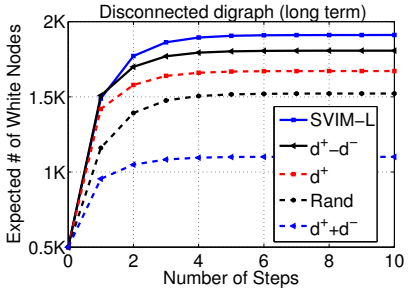


Figure 4.5: G is disconnected

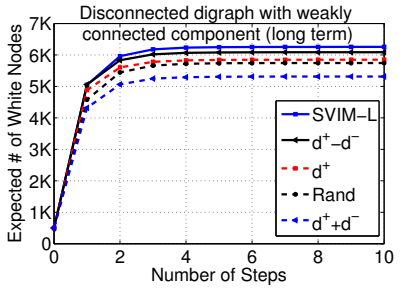


Figure 4.6: G is disconnected with WCC

SVIM-L algorithm outperforms all other heuristics with up to 17% more long term influence, which occurs for $t \geq 4$. In general, we see that for weakly connected and disconnected digraphs, SVIM-L has larger winning margins over all other heuristics than the case of balanced ergodic digraphs (Fig. 4.4–4.6 vs. Fig.4.1). We attribute this to our accurate computation of influence contribution in the more involved weakly connected and disconnected digraph cases. Moreover, in all cases, the dynamics converge very fast, i.e., in only a few steps, which indicates that the convergence time of voter model on these random graphs are very small.

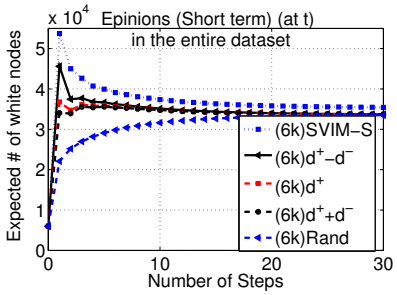


Figure 4.7: Instant influence in Epinions data with $k = 6$k

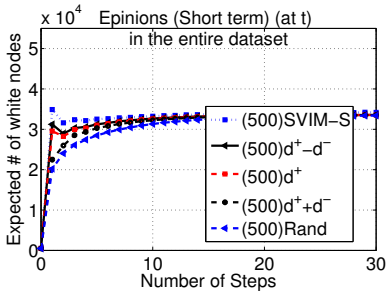


Figure 4.8: Instant influence in Epinions data with $k = 500$

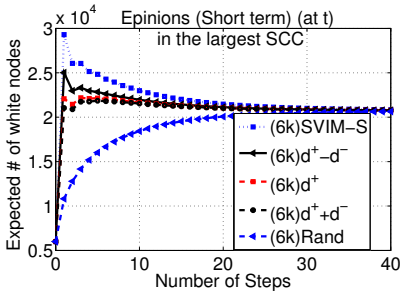


Figure 4.9: Instant influence in SCC with $k = 6$k

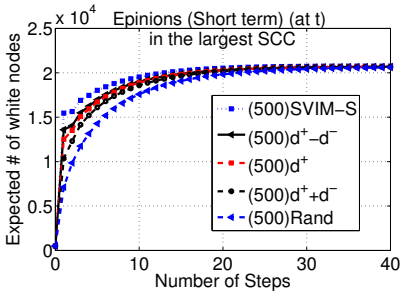


Figure 4.10: Instant influence in SCC with $k = 500$

Real datasets

We conduct extensive simulations using real datasets, such as Epinions and Slashdot datasets, to validate our theoretical results and evaluate the performance of our SVIM algorithm.

Epinions Dataset. Epinions.com [87] is a consumer review online social site, where users can write reviews to various items and vote for or against other users. The signed digraph is formed with positive or negative directed edge (u, v) meaning that u trusts or distrusts v. The statistics are shown in Table 4.2. We compare our short-term SVIM-S algorithm with four heuristics, i.e., $d^+ + d^-$, d^+, $d^+ - d^-$ and random seed selection, on the entire Epinions digraph as well as the largest strongly connected component (SCC).

Our tests are conducted on both Epinions dataset and its largest strongly connected component (SCC), where the largest SCC is ergodic and strictly unbalanced. We first look at the comparison of instant influence maximization (*at* step t) among various seed selection schemes. Fig. 4.7-4.10 shows the expected maximum instant influence at each step by different methods. Note that since the initial seeds selected by SVIM-S algorithm hinge on t, the values on the curve of our selection scheme are associated with different optimal initial seed sets. On the

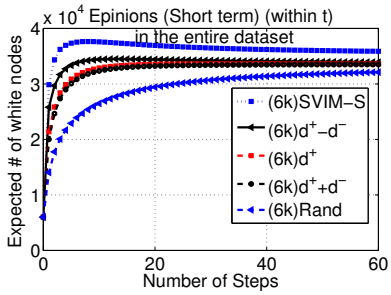


Figure 4.11: Average influence in Epinions data with $k = 6$k

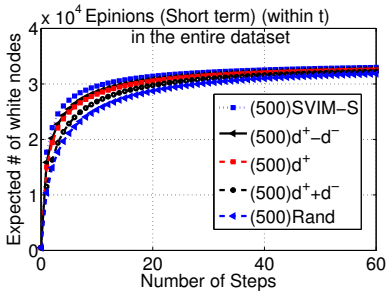


Figure 4.12: Average influence in Epinions data with $k = 500$

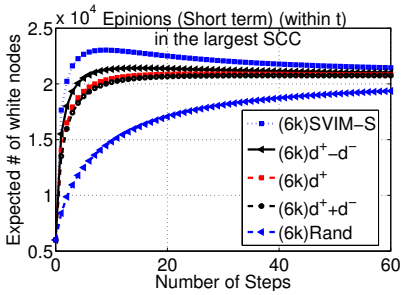


Figure 4.13: Average influence in SCC with $k = 6$k

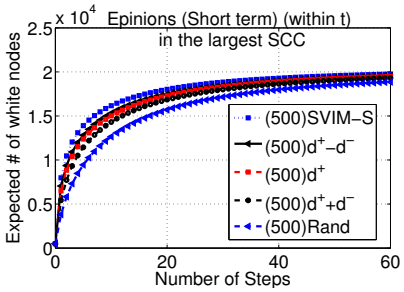


Figure 4.14: Average influence in SCC with $k = 500$

other hand, the seed selections of other heuristics are independent to t, thus the corresponding curves represent the same initial seed sets. We choose the budget as 500 and 6000 in our evaluations, i.e., selecting at maximum 500 or 6000 initial white seeds. From Fig. 4.7-4.10, SVIM-S algorithm consistently performs better, and in some cases, e.g., Fig. 4.9, it generates $16\% - 145\%$ more influence than other heuristics at step 1.

Next we compare the seed selection schemes for maximizing the average influence *within* the first t steps. Fig. 4.11-4.14 show the expected maximum average influence within the first t steps by different methods. Again, the values on the curve of SVIM-S algorithm are associated with different initial seed sets. Fig. 4.11-4.14 show that with different budgets, i.e., 500 and 6000 seeds, SVIM-S algorithm performs better than all other heuristics, where in Fig. 4.13 a maximum of 64% more influence is achieved at $t = 8$. Moreover, in all these figures, we observe that our seed selection scheme results in the highest long-term influence over other heuristics.

Moreover, from Fig. 4.7-4.14, we observe that as t increases, the influences (i.e., the expected number of white nodes), for SVIM-S and all heuristics except for random seed selection

Table 4.2: Statistics of Epinions and Slashdot datasets

Statistics	Epinions	Slashdot
# of nodes	131580	77350
# of edges	840799	516575
# of positive edges	717129	396378
# of negative edges	123670	120197
# of nodes in largest SCC	41441	26996
# of edges in largest SCC	693507	337351
# of positive edges in largest SCC	614314	259891
# of negative edges in largest SCC	79193	77460
# of strongly connected components	88361	49209

schedule, increase for small t's, and then decrease and converge to the stationary state. In contrast, from Fig. 4.1-4.6, the influence increases monotonically with t. This happens because Epinions dataset (as well as many real network datasets) has large portion (around 80%) of nodes in the non-sink components, where to maximize the long-term influence, only nodes in sink components should be selected, which governs the long-term influence dynamics of the whole graph, namely, sink nodes have higher long-term influence contributions. However, for short-term influence maximization, nodes with higher chances to influence more nodes in a few steps generally have large number of incoming links, which are able to influence a large number of nodes in either sink or non-sink components in a short period of time. Hence, in signed digraphs with large non-sink component, given a sufficiently large budget, the short-term influence can definitely outnumber the long-term influence. Our evaluations confirm this explanation. This interesting observation also leads to a problem that given a budget k, how to find the optimal time step t that generates the largest influence among all possible t's. We leaves this problem as our future work.

Slashdot Dataset. Slashdot.org [88] provides a discussion forum on various technology-related topics, where members can submit their stories, and comment on other members' stories. Its Slashdot Zoo feature allows members to tag each other as friends or foes, which in turn forms a signed online social network. The network was collected on 6-th November 2008 [67] and the statistics are shown in Table 4.2.

We evaluate instant influence and average influence of our SVIM-S algorithm on the entire slashdot dataset and its largest strongly connected component, respectively. Our results for $k = 6000$ are presented in Fig. 4.15-Fig. 4.18, which show that our SVIM-S algorithm performs

the best among all methods tested, especially in the early steps. When changing the budget k, similar results were obtained, where we omitted them here for brevity.

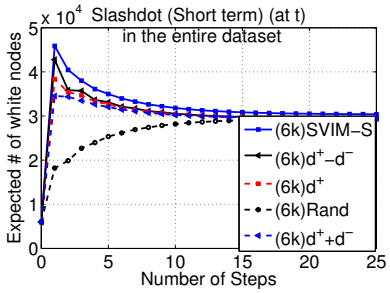


Figure 4.15: Instant influence in Slashdot data with $k = 6$k

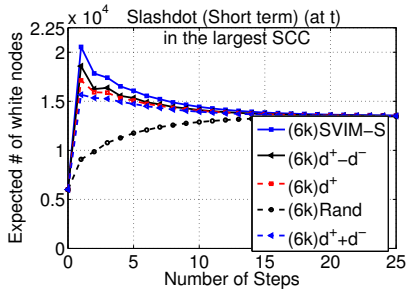


Figure 4.16: Instant influence in Slashdot SCC with $k = 6$k

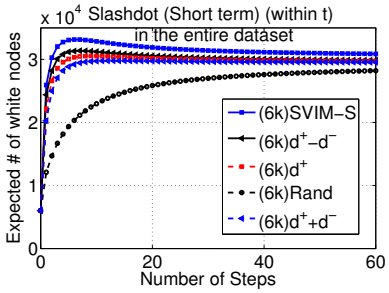


Figure 4.17: Average influence in Slashdot data with $k = 6$k

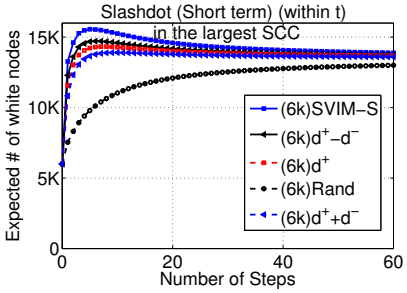


Figure 4.18: Average influence in Slashdot SCC with $k = 6$k

Moreover, the convergence times for both real-world datasets are fast, in a few tens of steps, indicating good connectivity and fast mixing property of real-world networks. In summary, our evaluation results on both synthetic and real-world networks validate our theoretical results and demonstrate that our SVIM algorithms for both short term and long term are indeed the best, and often have significant winning margins.

4.5.2 The impacts of signed information

Unlike Epinions and Slashdot, many online social networks such as Twitter are simply represented by unsigned directed graphs, where friends and foe relationships are not explicitly represented on edges. Without edge signs, two types of information may be mis-represented or under-represented: (1) one may follow his foes for tracking purpose, but this link may be mis-interpreted as friend or trust relationship; and (2) one may not follow his foes publicly to avoid being noticed, but his foes may still generate negative influence to him. In this section,

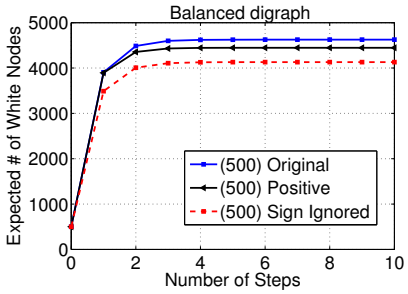


Figure 4.19: Synthetic balanced digraph

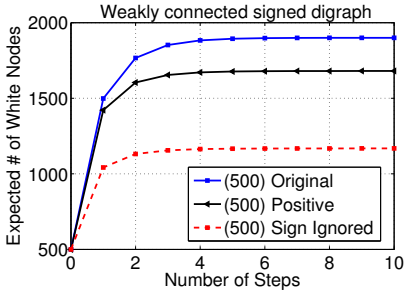


Figure 4.20: Synthetic weakly connected digraph

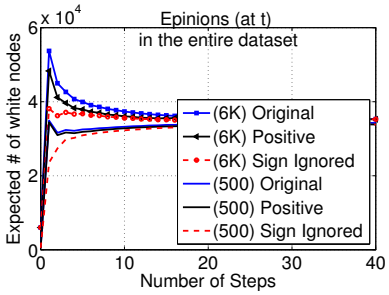


Figure 4.21: Epinions (the entire dataset)

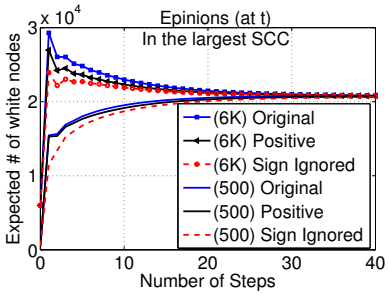


Figure 4.22: Epinions (the largest SCC)

we investigate how much influence gain can be obtained by taking the edge signs into consideration, thus illustrate the significance of utilizing both friend and foe relationships in influence maximization.

Taking the synthetic networks and Epinions dataset (used in Sec 4.5.1) as examples, we apply our SVIM algorithm to compute the optimal initial seed sets in the original signed digraphs, and two types of "sign-missing" scenarios, i.e., the unsigned digraphs with only original positive edges (denoted by "Positive" graphs) and with all edges labeled by the same signs (denoted by "Sign ignored" graphs). Then, we examine the performances of those three initial seed sets in original signed digraphs.

Fig. 4.19-4.22 show the evaluation results, where the seed sets obtained by considering edge signs perform consistently better than those using unsigned graphs. In synthetic networks, we observed $5\% - 16\%$ more influence in balanced digraph for $t \geq 6$ (See Fig. 4.19), and $11.7\% - 58\%$ more influence in weakly connected digraph for $t \geq 6$ (See Fig. 4.20). Moreover, in Epinions dataset from Fig. 4.21-4.22, there is no impact on the long-term influence, since the underlying graphs are strictly unbalanced. However, in short term, the results demonstrate that taking edge signs into consideration always performs better, which generates at maximum of 38% and 21% more influence for the entire dataset (See Fig. 4.21) and the largest SCC (See

Fig. 4.22), respectively. Both maximums occur at step 1. These results clearly demonstrate the necessity of utilizing sign information in influence maximization.

4.6 Summary

In this chapter, we propose and study voter model dynamics on signed digraphs, and apply it to solve the influence maximization problem. We provide rigorous mathematical analysis to completely characterize the short-term and long-term dynamics, and provide efficient algorithms to solve both short-term and long-term influence maximization problems. Extensive simulation results on both synthetic and real-world graphs demonstrate the efficacy of our signed voter model influence maximization (SVIM) algorithms. We also identify a class of anti-balanced digraphs, which is not covered in the social balance theory before, and exhibits oscillating steady state behavior.

There exist several open problems and future directions. One open problem is the convergence time of voter model dynamics on signed digraphs. For balanced and anti-balanced ergodic digraphs, our results show that their convergence times are the same as the corresponding unsigned digraphs. For strictly unbalanced ergodic digraphs and more general weakly connected signed digraphs, the problem is quite open. A future direction is to study influence diffusion in signed networks under other models, such as the voter model with a background color, the independent cascade model, and the linear threshold model.

Chapter 5

From Shortest-path to All-path: The Routing Continuum Theory and its applications

5.1 Introduction

Routing is a crucial operation in many types of networks from communication networks to transportation networks. For instance, in modern IP-based data networks, shortest path routing is most commonly used. In traditional telecommunication networks, dynamic alternative routing strategies that employ paths that are longer than shortest paths have been also proposed to reduce call blocking probabilities (see, e.g., [89, 90]). In wireless networks, due to the unstable channel characteristics, using a single "shortest" path (e.g., with link quality as link weights) for routing is often not the best choice; routing strategies that go beyond shortest path routing (see, e.g., [5, 91–93] and references therein) using multiple paths are often more effective. In the other extreme, in wireless sensor networks – due to their power and other resource constraints – potential-based routing [94] has been proposed, where the source essentially utilizes all (eligible) paths to transmit data to the destination. In [94], it is shown that such "all-path" routing minimizes the total energy dissipation of routing and thus maximizes the network lifetime. Clearly, what routing strategies to employ in a network hinges on what objectives are important *in practice*, therefore should be optimized. However, from a *theoretical* perspective,

when using multi-path routing that goes beyond a single shortest path, two questions arise: i) what set of paths should be used for routing? and ii) how traffic should be split (and merged) at any node along the multiple paths, especially when the paths are not all disjoint?

In addressing these questions, in this chapter we consider routing as flow optimization in a network with *mixed (weighted) L_1/L_2-norms* as the objective. This formulation is inspired by the earlier results where it has been shown that shortest path routing can be derived from network flow optimization with L_1 norm [95], whereas potential-based, "all-path" routing can be derived from network flow optimization with L_2-norm objective [89, 94]. Intuitively, the mixed L_1/L_2-norm objective can be interpreted as a trade-off between the latency and energy dissipation of paths used for routing (collectively, the paths form a routing graph): shorter paths lead to better routing with low latency, while diffusing traffic along more paths generally reduces energy dissipation. Using this formulation, we obtain a *surprising result*: as we vary the trade-off parameter θ, the routing graphs induced by the optimal flow solutions span from the shortest-path routing to multi-path routing with increasing path lengths to the potential-based ("all-path") routing – this entire (*finite*) sequence of routing graphs is referred to as the *routing continuum*. Our theory therefore subsumes the earlier L_1 and L_2 network flow optimization results [89, 95] as two extreme points in the entire routing continuum.

Furthermore, by considering the dual of the mixed L_1/L_2-norm network flow optimization problem, we develop an efficient iterative algebraic process as well as algorithms for identifying precisely the boundary conditions separating the finite sequence of routing graphs, and for computing the entire routing continuum and optimal flow solutions $X^*(\theta)$ for any $\theta \geq 0$. In particular, $X^*(\theta)$ specifies how traffic should be split and merged in the induced routing graph. We also generalize the theory to account for multiple flows (traffic demands), link capacity constraints and heterogeneous L_1/L_2 link weights, with applications to traffic engineering and wireless sensor networks. For instance, given a set of link weights and traffic demands on a network, our theory can be used to find the "best" routing graph (i.e., the best mix of shorter and longer paths) that minimizes the overall maximum link utilization.

In summary, our contributions are i) we develop a *unifying* theory using mixed L_1/L_2-norm network flow optimization and show that it can generate the entire routing continuum from shortest-path to "all-path" routing; ii) we develop an efficient iterative process for computing the entire routing continuum and optimal flow solutions $X^*(\theta)$ for any $\theta \geq 0$; iii) the basic theory is further generalized to account for multiple flows (traffic demands), link capacity

constraints and heterogeneous L_1/L_2 link weights, with applications to traffic engineering and wireless sensor networks. iv) Moreover, by applying the routing continuum theory, we generalize the betweenness centrality measure using mixed network flow, with applications in network robustness analysis. Last but not the least, while we focus on network routing in this chapter, we believe that our results can be applied to many other applications where the problems can be cast in terms of flows in a network.

The remainder of the chapter is organized as follows. In Section 5.2 we introduce the basic notations and re-state the known L_1 and L_2 flow optimizations using our notations. In Section 5.3 the general theory and results using the mixed L_1/L_2-norm flow optimization are established, and the iterative computation process and algorithms are described in Section 5.4. In Section 5.5, we consider several generalizations, with applications to traffic engineering, wireless sensor networks, and network robustness analysis. Section 5.6 discusses the related work, and the chapter is summarized in Section 5.7.

5.2 Shortest Path and "All-Path" Routing as Network Flow Optimization

In this section, we first introduce the basic notations that will be used throughout the chapter. Then, we illustrate how shortest path routing and potential-based "all-path" routing can be formulated as the flow optimization problems in a network using metric norms (on the flow space). More specifically, the shortest path routing results from minimizing the (weighted) L_1-norm of flows between a given source-destination pair in a network, whereas the potential-based, "all-path" routing results from minimizing the corresponding L_2-norm.

5.2.1 Network and Flows: Basic Notations

We represent a n-node network as an undirected, weighted graph, $G = (V, E, W)$, where $V = \{1, 2, \ldots, n\}$ is the set of vertices, E is the set of edges, and each edge $(i, j) \in E$ is assigned a positive weight w_{ij}. As G is undirected, (i, j) and (j, i) represent the same edge in E, and $w_{ij} = w_{ji} > 0$. Define $w_{ij} = 0$ if $(i, j) \notin E$, then the weight matrix $W = [w_{ij}]$ is symmetric. In particular, if all edges have a unit weight, i.e., W is a 0-1 matrix, then G represents a simple graph, and W is the corresponding adjacency matrix.

Let $d = [s,t]$, $s,t \in V, s \neq t$, denote a source-destination (or source-sink) pair in the network G. A flow of $I^{(d)}$-unit amount that flows from source s to destination t is mathematically defined as a function, $X^{(d)} : V \times V \rightarrow \mathbb{R}^+$ ($\mathbb{R}^+$ is the set of non-negative real numbers), satisfying the following constraints:

$$\text{along one direction:} \quad \text{if } X_{ij}^{(d)} > 0 \text{ then } X_{ji}^{(d)} = 0, \tag{5.1}$$

$$\text{along network edges:} \quad \text{if } (i,j) \notin E \text{ then } X_{ij}^{(d)} = 0, \tag{5.2}$$

$$\text{flow conservation at } s: \quad I^{(d)} + \sum_{k=1}^{n} X_{ks}^{(d)} = \sum_{j=1}^{n} X_{sj}^{(d)}, \tag{5.3}$$

$$\text{intermediate node } i \neq s,t: \quad \sum_{k=1}^{n} X_{ki}^{(d)} = \sum_{j=1}^{n} X_{ij}^{(d)}, \tag{5.4}$$

$$\text{at destination } t: \quad \sum_{k=1}^{n} X_{kt}^{(d)} = \sum_{j=1}^{n} X_{tj}^{(d)} + I^{(d)}. \tag{5.5}$$

Note that in this flow definition, for each (undirected) edge $(i,j) \in E$, both $X_{ij}^{(d)}$ and $X_{ji}^{(d)}$ are defined, and the constraint in eq.(5.1) states that if $X_{ij}^{(d)} > 0$, then $X_{ji}^{(d)} = 0$; or if $X_{ji}^{(d)} > 0$, then $X_{ij}^{(d)} = 0$. It is possible that for $(i,j) \in E$, both $X_{ij}^{(d)} = X_{ji}^{(d)} = 0$. In particular, by the constraint in eq.(5.2), $X_{ij}^{(d)} = X_{ji}^{(d)} = 0$ for $(i,j) \notin E$. The flow constraints in eqs.(5.3)-(5.5) state that a amount of $I^{(d)}$ units of flow is injected at source s, and the same amount is removed from destination t, while the amount of flow entering any intermediate node i is the same as the amount leaving the node.

Given a flow $X^{(d)}$ between a source-destination pair $d = [s,t]$, it induces an *oriented* (or *directed*) sub-graph of G, $G_{X^{(d)}} = (V_{X^{(d)}}, E_{X^{(d)}})$, where an arc $\langle i,j \rangle \in E_{X^{(d)}}$ and $i,j \in V_{X^{(d)}}$ if and only if $X_{ij}^{(d)} > 0$. As a directed acyclic graph (DAG) between s and t, $G_{X^{(d)}}$ represents the routes used to route the flow $X^{(d)}$ (of $I^{(d)}$ units) from source s to destination t, and we refer to it as the *routing graph* for the flow $X^{(d)}$. When $G_{X^{(d)}}$ consists of more than a single path between s and t, then $X_{ij}^{(d)}$ indicates how much flow is routed along the edge (arc) $\langle i,j \rangle$. In general, the flow may be split or merged[1] at nodes in $G_{X^{(d)}}$, and routed along different paths between s and d. We will use $\mathcal{F}^{(d)}$ to denote the collection of flows, i.e., all functions that satisfy eqs.(5.2)-(5.5).

In the next two subsections we will use two well-known results [89, 95] to illustrate that certain common routing strategies, namely, shortest path routing and potential-based, "all-path"

[1] The flow definition implicitly assumes that flows are "infinitely divisible fluid" – they can be split and merged arbitrarily at any node of the network, as long as the above flow conservation constraints are met. This mathematical definition of network flows thus provides an *idealized* (fluid) abstraction of, e.g., traffic demands routed from a source to a destination in a communication network, or commodities transported from a source to a destination in a transportation network, and so forth.

routing, can be derived by minimizing the (weighted) L_1-norm and L_2-norm, respectively, of flows between a given source-destination pair in a network. In Section 5.3 we will generalize these results and establish that by minimizing flows using mixed L_1-norm and L_2-norm, we can generate a *continuum* of routing strategies, resulting in a sequence of routing graphs with varying numbers of paths of differing costs selected, from the shortest paths to all paths (between a source-destination pair).

5.2.2 Shortest-Path Routing & L_1-norm Flow Optimization

Without loss of generality, unless otherwise specified, we assume that $s = 1$ and $t = n$, and $I^{(d)} = 1$. For clarity of notation, we drop the superscript d from $X^{(d)}$. In other words, the flow X (as a function) is equivalently specified by a set of n^2 variables, X_{ij}'s, $1 \leq i, j \leq n$.

Consider the following L_1-norm network flow optimization problem, which can be solved using linear programming (LP).

L_1-norm Network Flow Optimization (L_1 Primal):

$$\min_X \sum_{i=1}^{n} \sum_{j=1}^{n} w_{ij} X_{ij} \tag{5.6}$$

subject to the flow conservation constraints eqs.(5.2)-(5.5), which are more compactly represented below using X_{ij}'s:

$$\sum_{j:(i,j)\in E} X_{ij} - \sum_{k:(k,i)\in E} X_{ki} = \begin{cases} 1 & \text{if } i = 1 \\ 0 & \text{if } i = 2, \ldots, n-1 \\ -1 & \text{if } i = n, \end{cases} \tag{5.7}$$

$$\text{and } X_{ij} \geq 0,\ 1 \leq i, j \leq n. \tag{5.8}$$

Note that the feasible solutions to eq.(5.6) subject to eqs.(5.7) and (5.8) satisfy constraints eqs.(5.2)-(5.5), and an optimal solution to this must also satisfy eq.(5.1) automatically. Hence without loss of generality, when considering the optimization in eq.(5.6), we can restrict ourselves to X's that are flows, i.e., $X \in \mathcal{F}$. Thus we can re-state the optimization in eq.(5.6) as

$$\min_{X\in\mathcal{F}} \sum_{i}^{n} \sum_{j}^{n} w_{ij} X_{ij}.$$

In other words, the optimization solution to eq.(5.6) is the flow that minimizes the weighted L_1-norm.

To show that the optimal solution to this L_1-norm network flow optimization gives rise to the shortest-path routing, we consider its dual, stated below in terms of the Lagrange multipliers $-U_i$'s (corresponding to the flow conservation constraints eq.(5.7)[2]):

Dual of L_1-norm Network Flow Optimization (L_1 Dual):

$$\max_{U} U_1 \tag{5.9}$$

$$\text{subject to } U_n = 0 \text{ and } U_i - U_j \leq w_{ij}, \forall (i,j) \in E. \tag{5.10}$$

Let X^* denote the optimal flow solution to the primal problem eq.(5.6), and U^* the optimal solution to the dual problem. The duality and complementary slackness give us the following relations between X^*_{ij}'s and U^*_i's (*cf.* Lemma 1 in [95] and the transportation and network flow problems in chapter 5 in [96]).

$$\text{if } X^*_{ij} > 0, \text{then } U^*_i - U^*_j = w_{ij}; \tag{5.11}$$

$$\text{and if } X^*_{ij} = 0, \text{then } U^*_i - U^*_j < w_{ij}. \tag{5.12}$$

Using these relations, the authors in [95], show that the optimal solution to the dual problem, U^*_i's, have the following properties (*cf.* Theorem 1 and its proof in [95]):

Lemma 8 *Let P be a path from node 1 to node n. If for each edge (arc) $\langle i,j \rangle \in P$, $U^*_i - U^*_j = w_{ij}$, then P is a shortest path from node 1 to node n (with respect to the weights w_{ij}'s), and $U^*_1 = \sum_{\langle i,j \rangle \in P} w_{ij}$. Alternatively, if Q is a path from node 1 to node n that is* not *a shortest path, then $U^*_1 < \sum_{\langle i,j \rangle \in Q} w_{ij}$.*

The above lemma implies that for any node i on a shortest path, U^*_i is the *shortest-path* distance from node i to node n (the destination). Furthermore, the optimal flow X^* is only routed along the shortest paths between source 1 and destination n. In other words, the resulting routing graph G_{X^*} is the DAG formed by the shortest paths from 1 to n only. When there are multiple shortest paths between 1 and n, X^*_{ij} specifies the amount of flow carried on the edges of node i that are on the shortest paths, thus how the flow should be split among multiple shortest path at node i.

[2] Note that our Lagrange multipliers are *negatives* of those used in the "Dual Shortest Path Formulation (D-SP)" in [95], p. 3.

5.2.3 Potential-based ("All-path") Routing and L_2-norm Flow Optimization

We now consider the following (weighted) L_2-norm network flow optimization problem:

L_2-norm Network Flow Optimization (L_2 Primal):

$$\min_{X \in \mathcal{F}} \sum_{i=1}^{n} \sum_{j=1}^{n} w_{ij} X_{ij}^2. \tag{5.13}$$

To show that the optimal solution to this L_2-norm network flow optimization gives rise to the potential-based, "all-path" routing, we again consider its dual, stated below in terms of the Lagrange multipliers U_i's (where for convenience we have used $-2U_i$'s as the multipliers for the flow conservation constraints eq.(5.7)):

Dual of L_2-norm Network Flow Optimization (L_2 Dual):

$$\max_{U} U_1 - \frac{1}{2} \sum_{i=1}^{n} \sum_{j: U_i > U_j} \frac{(U_i - U_j)^2}{w_{ij}}. \tag{5.14}$$

subject to $U_n = 0$.

Let X^* denote the optimal flow solution to the primal problem eq.(5.13), and U^* the optimal solution to the dual problem. The duality and complementary slackness give us the following relations between X_{ij}^*'s and U_i^*'s: for any edge $(i,j) \in E$,

$$\text{if } U_i^* > U_j^*, \text{ then } X_{ij}^* = \tfrac{U_i^* - U_j^*}{w_{ij}} > 0; \tag{5.15}$$

$$\text{and if } U_i^* \leq U_j^*, \text{ then } X_{ij}^* = 0. \tag{5.16}$$

If we treat w_{ij} as the resistance on edge $(i,j) \in E$, then the relation eq.(5.15) gives us precisely Ohm's law [89], and U_i^* is the voltage (potential) at node i when a unit of current is injected at source node 1 and removed at sink node n (and grounded with $U_n^* = 0$). For any $(i,j) \in E$, if $U_i^* > U_j^*$, then the current I_{ij} flowing from node i to node j along edge (i,j) is exactly X_{ij}^*, as $I_{ij} = (U_i^* - U_j^*)/w_{ij} = X_{ij}^* > 0$. (In a electrical network, the reverse current flow, i.e., the current from node j to node i is defined as $I_{ji} := -I_{ij} = -X_{ij}^* < 0$.) Hence the optimal solution to the dual problem eq.(5.14), U^*, is a potential function (the voltage potential in the electrical network G): U_i^* is the voltage potentialfrom node i to destination node n (ground).

For $(i,j) \in E$, define $a_{ij} := 1/w_{ij}$, the conductance on edge (i,j), and for $(i,j) \notin E$, $a_{ij} = 0$. From the flow conservation constraints (or directly by solving the dual optimization

problem eq.(5.14)), we see that

$$U_i^* = \begin{cases} \sum_{j=1}^n \frac{a_{ij}}{\sum_k a_{ik}} U_j^* + \frac{1}{\sum_k a_{ik}} & \text{if } i = 1 \\ \sum_{j=1}^n \frac{a_{ij}}{\sum_k a_{ik}} U_j^* & \text{if } i = 2, \ldots, n-1 \\ 0 & \text{if } i = n \end{cases} , \tag{5.17}$$

which gives the Kirchhoff's law for voltage in an electrical network. The dual problem eq.(5.14) gives us the *Dirichlet principle* [89]: the voltage potentials, U^*, taken within the electrical network G minimizes the total energy dissipation. Likewise, the L_2-norm flow optimization problem also has a physical interpretation (Thompson's Principle [89]): among all flows $X \in \mathcal{F}$, the optimal (current) flow, X^*, minimizes the energy dissipation in the (electrical) network.

This connection between currents (and voltage) in electrical networks and L_2-norm network flow optimization is well known in the literature (see, e.g., [89, 97–100]), where the expected round-trip commute times between two nodes in a random walk over a network, whose link weights are conductances (reciprocals of resistances), is the same as the effective resistance between the those two nodes treating the graph as an electrical network. These connections give rise to potential-based ("all-path") routing (or "stochastic routing") in communication and wireless sensor networks [89, 94]. Using the relations eq.(5.15) and eq.(5.16), it is easy to see that for any path P from node 1(source) to node n (destination) in the network G, the (current) flow along P is nonzero (i.e., $X_{ij}^* > 0, \forall \langle i, j \rangle \in P$) *if and only if* the potential (voltage) at any node i along the path from node 1 to node n is strictly decreasing (i.e., $\forall \langle i, j \rangle \in P$, $U_i^* > U_j^*$). Hence the routing graph G_{X^*} induced by the optimal flow to the L_2-norm flow minimization problem is a DAG consisting of any path from source node 1 to destination node n with strictly decreasing potentials – that is what we also refer to the potential-based routing as "all-path" routing. Moreover, Ohm's law specifies how flows along the paths are split – proportional to the potential difference along an edge and inverse to the resistance of the edge, namely, $X_{ij}^* = (U_i^* - U_j^*)/w_{ij}$.

5.3 Mixed L_1 and L_2-norm Network Flow Optimization and the Routing Continuum

The results in the previous section show that the optimal flows that minimize the (weighted) L_1-norm and L_2-norm in a network yield the shortest path and (potential-based) "all-path"

routing, respectively. Intuitively, if we treat w_{ij} as "delay"[3] on each link (i,j), then the L_1-norm minimization produces an optimal flow routing that minimizes the total delay; whereas the L_2-norm minimization produces an optimal flow routing that minimizes the total energy dissipation (treating w_{ij} as the resistance of link (i,j)). This gives rise to a natural question: can we generate other routing strategies between these two extremes, e.g., routing using shortest paths as well as second-shortest paths, via network flow optimization with respect to some other forms of cost metrics? In particular, can these routing strategies be derived by trading off the total delay (the L_1-norm) and the total energy (the L_2-norm)? This leads us to posing the following mixed L_1- and L_2-norm network flow optimization problem with $\theta \geq 0$, subject to flow conservation law eqs.(5.7) and (5.8), denoted as $X \in \mathcal{F}$.

Mixed L_1- and L_2-norm Network Flow Optimization (Primal):

$$\min_{X \in \mathcal{F}} \sum_{i=1}^{n} \sum_{j=1}^{n} w_{ij} X_{ij}^2 + 2\theta \sum_{i=1}^{n} \sum_{j=1}^{n} w_{ij} X_{ij}. \tag{5.18}$$

Theorem 13 below presents the dual and optimal solution to this flow optimization problem, by introducing Lagrange multipliers $-2U_i$.

Theorem 13 ***Mixed L_1- and L_2-norm Network Flow Optimization (Dual):***

$$\max_{U} U_1 - \frac{1}{2} \sum_{i} \sum_{j: U_i - U_j > \theta w_{ij}} \frac{(U_i - U_j - \theta w_{ij})^2}{w_{ij}} \tag{5.19}$$

$$s.t. U_n = 0. \tag{5.20}$$

Let $X^*(\theta)$ be the optimal solution to the primal problem eq.(5.18), and $U^*(\theta)$ the optimal solution to the dual problem eq.(5.19). $X^*(\theta)$ and $U^*(\theta)$ follow the following relations.

$$X_{ij}^*(\theta) = \begin{cases} \frac{U_i^*(\theta) - U_j^*(\theta)}{w_{ij}} - \theta & \text{if } U_i^*(\theta) - U_j^*(\theta) > \theta w_{ij} \\ 0 & \text{if } U_i^*(\theta) - U_j^*(\theta) \leq \theta w_{ij}. \end{cases} \tag{5.21}$$

Proof : By introducing Lagrangian multiplier $2U_i$ $(1 \leq i \leq n)$ for each equality constraint in eq.(5.7), and Lagrangian multiplier $2t_{ij}$ $(1 \leq i, j \leq n)$ for each inequality constraint in eq.(5.8), the Lagrangian function of the problem eq.(5.18) can be written as

[3] Or in a physical meaning that is more consistent with the L_2-norm interpretation, the "(resistive or fricative) force" needed to move a unit of flow across link (i,j).

$$\mathcal{L}(X,U,t) = \sum_i \sum_j \left(X_{ij}^2 w_{ij} - 2(U_i - U_j - \theta w_{ij} + t_{ij})X_{ij}\right) + 2(U_1 - U_n). \tag{5.22}$$

Then, we take the partial derivative of $\mathcal{L}(X,U,t)$ (eq.(5.22)) with respect to X_{ij}, and solve the equation that the partial derivative equals 0 for each $i,j = 1,\ldots,n$.

$$\frac{\partial \mathcal{L}(X,U,t)}{\partial X_{ij}} = 2\left(X_{ij}w_{ij} - (U_i - U_j - \theta w_{ij} + t_{ij})\right) = 0, \tag{5.23}$$

$$X_{ij} = \frac{U_i - U_j + t_{ij}}{w_{ij}} - \theta. \tag{5.24}$$

Plugging the eq.(5.24) into eq.(5.22) yields the following dual problem.

$$\max_U U_1 - \frac{1}{2}\sum_i \sum_j \frac{(U_i - U_j - \theta w_{ij} + t_{ij})^2}{w_{ij}} \tag{5.25}$$

$$\text{s.t. } U_n = 0 \text{ and } \mathrm{t}_{ij} \geq 0, \text{ for } i,j = 1,\ldots,n, \tag{5.26}$$

Since the primal problem is convex, the strong duality and complementary slackness hold, thus the Karush-Kuhn-Tucker (KKT) conditions [101] are sufficient and necessary to be the optimal solution to both of the primal and dual problems. The KKT conditions include the primal constraints eq.(5.7)–(5.8) and the following three conditions.

$$t_{ij} \geq 0, \tag{5.27}$$

$$t_{ij}X_{ij} = 0, \tag{5.28}$$

$$X_{ij}w_{ij} - (U_i - U_j - \theta w_{ij} + t_{ij}) = 0. \tag{5.29}$$

From the eq.(5.28), X_{ij} or t_{ij} cannot both be zero. By setting one of them to be zero, we can solve the other. Then by checking the positivity of the solution, we get the optimal solution.

$$t_{ij}^* = \begin{cases} 0 & \text{if } U_i^* - U_j^* > \theta w_{ij}, \\ -(U_i^* - U_j^* - \theta w_{ij}) & \text{if } U_i^* - U_j^* \leq \theta w_{ij}, \end{cases} \tag{5.30}$$

$$X_{ij}^* = \begin{cases} \frac{U_i^* - U_j^*}{w_{ij}} - \theta & \text{if } U_i^* - U_j^* > \theta w_{ij}, \\ 0 & \text{if } U_i^* - U_j^* \leq \theta w_{ij}. \end{cases} \tag{5.31}$$

Since the optimal t_{ij}^* is a function of U_i^*'s, we can plug it in eqs.(5.25)-(5.26) to simply the dual problem, and eliminate the variable t_{ij}, which yields eq.(5.21). ∎

Clearly, $\theta = 0$ gives us the L_2-norm network flow optimization. In the following we will show that for sufficiently large θ, the routing graph induced by the optimal solution to eq.(5.18) gives the same shortest path DAG as the L_1-norm flow optimization. In other words, for sufficiently large θ, the optimal solution to eq.(5.18) yields the shortest path routing. Furthermore, for θ in between, the optimal solution to eq.(5.18) yields a *continuum* of routing graphs with the "all-path" and shortest-path DAGs as two extremes in the continuum.

Fix $\theta \geq 0$, and let $G_{X^*(\theta)}$ denote the routing graph (DAG) induced by the optimal flow solution $X^*(\theta)$ to eq.(5.18), i.e., for any edge $(i,j) \in E$, the arc $\langle i,j \rangle$ is included in $G_{X^*(\theta)}$ if and only $X^*_{ij}(\theta) > 0$. We use $P \in G_{X^*(\theta)}$ to denote a path P from node 1 (source) to node n (destination) where the flow along this path is nonzero, i.e., for any $\langle i,j \rangle \in P$, $X^*_{ij}(\theta) > 0$. We have the following lemma:

Lemma 9 *Consider any path $P \in G_{X^*(\theta)}$, and Q be* any *path from node 1 to node n. The following holds:*

$$\theta \sum_{\langle i,j \rangle \in P} w_{ij} < U^*_1(\theta) \leq (\theta + 1) \sum_{\langle i,j \rangle \in Q} w_{ij}. \tag{5.32}$$

Proof : For any $\langle i,j \rangle \in P$, since $X^*_{ij}(\theta) > 0$, from eq.(5.21) we have $U^*_i(\theta) - U^*_j(\theta) = \theta w_{ij} + w_{ij} X^*_{ij}(\theta)$. Therefore $\sum_{\langle i,j \rangle \in P} (U^*_i(\theta) - U^*_j(\theta)) = \sum_{\langle i,j \rangle \in P} (\theta w_{ij} + w_{ij} X^*_{ij}(\theta))$. Hence

$$U^*_1(\theta) = \sum_{\langle i,j \rangle \in P} (\theta w_{ij} + w_{ij} X^*_{ij}(\theta)) > \theta \sum_{\langle i,j \rangle \in P} w_{ij}, \tag{5.33}$$

as $X^*_{ij} > 0$ for any $\langle i,j \rangle \in P$. On the other hand, for any $\langle i,j \rangle \in Q$, from eq.(5.21) we have $U^*_i(\theta) - U^*_j(\theta) \leq \theta w_{ij} + w_{ij} X^*_{ij}(\theta)$, where the inequality holds when $X^*_{ij}(\theta) = 0$. Summing up along all edges $\langle i,j \rangle \in Q$, we have

$$U^*_1(\theta) \leq \sum_{\langle i,j \rangle \in Q} (\theta w_{ij} + w_{ij} X^*_{ij}(\theta)) \leq (\theta + 1) \sum_{\langle i,j \rangle \in Q} w_{ij},$$

where the last inequality follows from the fact that $X^*_{ij} \leq 1$. Combining this and the inequality in eq.(5.33) proves the lemma. ∎

From Lemma 9, the following holds for any $\theta > 0$,

$$\sum_{\langle i,j \rangle \in P} w_{ij} < (1 + \theta^{-1}) \min_{Q} \sum_{\langle i,j \rangle \in Q} w_{ij}. \tag{5.34}$$

Using this Lemma, we establish the following theorem.

Theorem 14 (Routing Continuum) *Let $\mathcal{R}(0)$ denote the (potential-based) "all-path" routing graph in Section 5.2.3, namely, the routing graph induced by the optimal L_2-norm flow $X^*(0)$, the optimal solution to eq.(5.18) with $\theta = 0$. Let $\mathcal{P}$ denote the collection of all paths (with* nonzero *flow), $P \in \mathcal{R}(0)$, from source node 1 to destination n. Sort and group the paths based on their length, i.e., $|P| := \sum_{\langle i,j\rangle \in P} w_{ij}$, which yields a partition (equivalent classes) of $\mathcal{P}$: $\mathcal{P}_1, \ldots, \mathcal{P}_M$, where $\mathcal{P}_m = \{P \in \mathcal{R}(0) : |P| = L_m\}$, $m = 1, \ldots, M$, and $L_1 < \ldots < L_M$. Clearly L_1 is the length of the shortest paths.*

For $\theta > 0$, let $\mathcal{R}(\theta)$ denote the routing graph induced by the optimal flow $X^*(\theta)$, the solution to the mixed L_1- and L_2-norm flow optimization problem. Then for $(L_m - L_1)/L_1 < \theta^{-1} \leq (L_{m+1} - L_1)/L_1$, $m = 1, \ldots, M$ (here we define $L_{M+1} = \infty$), we have

$$\mathcal{R}(\theta) \subseteq \cup_{k=1}^{m} \mathcal{P}_k. \tag{5.35}$$

In other words, paths in $\mathcal{R}(\theta)$ have length at most L_m.

Proof : We prove by contradiction. Given any m, $m = 1, \ldots, M$, and $\theta > 0$ where $(L_m - L_1)/L_1 < \theta^{-1} \leq (L_{m+1} - L_1)/L_1$, suppose there exists $P \in \mathcal{R}(\theta)$ such that $|P| > L_m$ (thus $|P| \geq L_{m+1}$). From Lemma 9, the length of any path in the routing graph $\mathcal{R}(\theta)$ used to route the optimal flow $X^*(\theta)$ is less than $(1 + \theta^{-1})L_1 \leq L_{m+1}$. This leads to a contradiction. ∎

Theorem 14 states as θ increases from 0 to ∞, or equivalently θ^{-1} decreases to 0, longer paths in $\mathcal{R}(0)$ are pruned, yielding a "sparser" routing graph $\mathcal{R}(\theta)$ that contains only paths of length less than $(1 + \theta^{-1})L_1$. In fact, there are a finite sequence of routing graphs $\mathcal{R}_m$, $1 \leq m \leq M$, where $\mathcal{R}_m$ only contains paths of length at most L_m. We refer to this sequence of routing graphs as the *routing continuum*. In the next section we will present an algorithm for explicitly constructing the routing continuum, and in particular, for computing the optimal flow solution, $X^*(\theta)$, which specifies how the optimal flow is routed among the paths in $\mathcal{R}_m$.

5.4 Computing the Routing Continuum

In this section we describe an efficient algorithm for computing the routing continuum and the associated optimal flow $X^*(\theta)$ for all θ's, and use two simple examples to illustrate the algorithm and results obtained thereof.

5.4.1 Routing Continuum Algorithm (RCA)

We introduce an iterative process for computing the routing continuum and the optimal flow $X^*(\theta)$, starting with $\theta = 0$, where each step involves solving a set of linear equations in $U_i^*(\theta)$'s.

For any θ, let $\mathcal{R}(\theta) = (V(\theta), E(\theta))$ denote the routing graph induced by $X^*(\theta)$, a subgraph of $G = (V, E)$, where $(i, j) \in \mathcal{R}(\theta)$ if and only if $X_{ij}^*(\theta) > 0$. In the following, we will treat $\mathcal{R}(\theta)$ as an *undirected* graph. Hence an edge $(i, j) \in \mathcal{R}(\theta)$ if and only if either $X_{ij}^*(\theta) > 0$ or $X_{ji}^*(\theta) > 0$, or equivalently, $(i, j) \in \mathcal{R}(\theta)$ if and only if $|U_i^*(\theta) - U_j^*(\theta)| > \theta w_{ij}$. For $i \in V(\theta)$, let $\Delta_i^{(+)}(\theta)$ denote the number of edges (k, i) with *incoming* flow (i.e., $X_{ki}^*(\theta) > 0$); or formally, $\Delta_i^{(+)}(\theta) := \sum_k \mathbf{1}\{U_k^*(\theta) - U_i^*(\theta) > \theta w_{ki}\}$. Likewise, let $\Delta_i^{(-)}(\theta)$ denote the number of edges (i, j) with *outgoing* flow (i.e., $X_{ij}^*(\theta) > 0$); thus $\Delta_i^{(-)}(\theta) := \sum_j \mathbf{1}\{U_i^*(\theta) - U_j^*(\theta) > \theta w_{ij}\}$. Define $\Delta_i(\theta) := \Delta_i^{(-)}(\theta) - \Delta_i^{(+)}(\theta)$, and $d_i(\theta) = \sum_{j:|U_i - U_j| > \theta w_{ij}} a_{ij} = \sum_{j:(i,j)\in E(\theta)} a_{ij}$, where $a_{ij} := w_{ij}^{-1}$ if $w_{ij} > 0$, and $a_{ij} := 0$ if otherwise. Then from eq.(5.19), we see that the optimal $U_i^*(\theta)$'s satisfy the following conditions:

$$d_i(\theta)U_i^*(\theta) - \sum_{j:(i,j)\in E(\theta)} a_{ij}U_j^*(\theta) - \theta\Delta_i(\theta) = \begin{cases} 1 & i = 1 \\ 0 & 1 < i < n, \end{cases} \tag{5.36}$$

$$\text{and } U_n^*(\theta) = 0. \tag{5.37}$$

We can rewrite eq.(5.36) more compactly in the matrix form:

$$L(\theta)U^*(\theta) - \theta\Delta(\theta) = \mathrm{b}. \tag{5.38}$$

Here, $L(\theta) := [L_{ij}(\theta)]$ is the $n-1$ by $n-1$ submatrix of the standard graph Laplacian [102] of $\mathcal{R}(\theta)$ (with the adjacency matrix $A(\theta) := [a_{ij}], i, j \in V(\theta)$), *restricted to* $V(\theta) - \{n\}$, namely, $L_{ii}(\theta) := d_i(\theta)$, and $L_{ij}(\theta) := -a_{ij}, i, j \in V(\theta) - \{n\}$. $\Delta(\theta) := [\Delta_i(\theta)]$, the vector consisting of $\Delta_i(\theta)$, and $\mathrm{b} = [1, 0, \ldots, 0]^T$ is the vector corresponding to the right hand side of eq.(5.36). Since $\mathcal{R}(\theta)$ is connected, $L(\theta)$ is non-singular and thus

$$U^*(\theta) = L^{-1}(\theta)(\theta\Delta(\theta) + \mathrm{b}). \tag{5.39}$$

Hence given θ, we can explicitly solve for $U^*(\theta)$ using eq.(5.39). However, the definitions of both $L(\theta)$ and $\Delta(\theta)$ hinge on the routing graph $\mathcal{R}(\theta) = (V(\theta), E(\theta))$, which is itself defined *assuming we know* $X_{ij}^*(\theta)$*!*

This *circular dependency* fortunately can be broken. From Theorem 14, we know that there exist only a *finite* sequence of routing graphs, $\mathcal{R}(\theta_m)$, $0 \le m \le M$, where $0 = \theta_0 < \theta_1 <$

$\ldots < \theta_M$. In other words, for $\theta_m \leq \theta < \theta_{m+1}$, $0 \leq m \leq M$ (and define $\theta_{M+1} = \infty$), $\mathcal{R}(\theta) = \mathcal{R}(\theta_m)$. Hence if we know $\mathcal{R}(\theta_m)$, we can solve $U^*(\theta)$ for any $\theta_m \leq \theta < \theta_{m+1}$ and thus $X^*(\theta)$. This leads to the following recursive process for computing the routing continuum and $X^*(\theta)$ for all $\theta \geq 0$.

Phase 1: from $\theta_0 = 0$ to θ_1:

When $\theta = 0 (= \theta_0)$, $L(0)$ is $n-1$ dimensional square submatrix of the graph Laplacian on the original network G (restricted to $V - \{n\}$). Then $U^*(0) = L^{-1}(0)\mathrm{b}$ is the optimal solution to the L_2-norm flow optimization, and $\mathcal{R}(0)$ is the "all-path" routing graph induced by the optimal L_2-norm flow $X^*(0)$.

Now consider any sufficient small $\theta > 0$ (any $\theta < \theta_1$ would suffice) such that $\mathcal{R}(\theta) = \mathcal{R}(0)$ (thus $X^*_{ij}(\theta) > 0$ for any $(i,j) \in \mathcal{R}(0)$. Hence $\Delta(\theta) = \Delta(0)$, $L(\theta) = L(0)$, and $U^*(\theta)$ is given by

$$U^*(\theta) = L^{-1}(0)(\theta\Delta(0) + \mathrm{b}) = U^*(0) + \theta L^{-1}(0)\Delta(0). \tag{5.40}$$

From eq.(5.21) and eq.(5.40), if $U_i(\theta) - U_j(\theta) > \theta w_{ij}$,

$$X^*_{ij}(\theta) = \frac{U^*_i(\theta) - U^*_j(\theta)}{w_{ij}} - \theta = X^*_{ij}(0) - \theta\alpha_{ij}(0), \tag{5.41}$$

where $\alpha_{ij}(0) = 1 - (\beta_i(0) - \beta_j(0))/w_{ij}$ is a constant, with $\beta_i(0) = [L^{-1}(0)\Delta(0)]_i$, if $i \in V - \{n\}$; and $\beta_i(0) = 0$, if $i = n$. Eq.(5.40) shows that $U^*(\theta)$ is linear function of θ, and for any edge (i,j) where $X^*_{ij}(0) > 0$, $X^*_{ij}(\theta)$ is also linear in θ. Clearly, on edge (i,j) with $\alpha_{ij}(0) > 0$, the optimal flow $X^*_{ij}(\theta)$ decreases when θ increases; whereas on those with $\alpha_{ij}(0) < 0$, the optimal flow $X^*_{ij}(\theta)$ increases (θ has no impact on those edges with $\alpha_{ij}(0) = 0$). Hence we know precisely the (first) boundary condition, namely, the smallest *positive* θ, when the first set of edges are to be truncated from $\mathcal{R}(0)$, namely, those where $X^*_{ij}(\theta)$ becomes 0:

$$\theta_1 := \min_{\langle i,j\rangle : \alpha_{ij}(0) > 0} \{X^*_{ij}(0)/\alpha_{ij}(0)\}.$$

Removing these edges yields the next routing graph $\mathcal{R}(\theta_1)$, for which $L(\theta_1)$ and $\Delta(\theta_1)$ can now be defined. Using eqs.(5.40) and (5.41), we can solve for the optimal solution, $U^*(\theta_1)$, and consequently, $X^*(\theta_1)$.

Phase 2: from θ_k to θ_{k+1}:

More generally, given $\mathcal{R}(\theta_k)$, and the corresponding optimal solutions, $U^*(\theta_k)$ and $X^*(\theta_k)$, we can solve for $U^*(\theta)$ and $X^*(\theta)$ for any $\theta_k \leq \theta < \theta_{k+1}$, using a similar argument. Again

from eq.(5.39), with $L(\theta) = L(\theta_k)$ and $\Delta(\theta) = \Delta(\theta_k)$, we have

$$\begin{aligned} U^*(\theta) &= L^{-1}(\theta_k)(\theta\Delta(\theta_k) + \mathrm{b}) \\ &= U^*(\theta_k) + (\theta - \theta_k)L^{-1}(\theta_k)\Delta(\theta_k), \end{aligned} \tag{5.42}$$

and if $U_i(\theta) - U_j(\theta) > \theta w_{ij}$,

$$X^*_{ij}(\theta) = \frac{U^*_i(\theta) - U^*_j(\theta)}{w_{ij}} - \theta = X^*_{ij}(\theta_k) - (\theta - \theta_k)\alpha_{ij}(\theta_k), \tag{5.43}$$

where $\alpha_{ij}(\theta_k) = 1 - (\beta_i(\theta_k) - \beta_j(\theta_k))/w_{ij}$ is a constant, with $\beta_i(\theta_k) = [L^{-1}(\theta_k)\Delta(\theta_k)]_i$, if $i \in V(\theta_k) - \{n\}$; and $\beta_i(\theta_k) = 0$, if $i = n$. This gives us the next boundary, θ_{k+1}, for the next set of links to be truncated (from $\mathcal{R}(\theta_k)$), where

$$\theta_{k+1} = \min_{\langle i,j\rangle : \alpha_{ij}(\theta_k) > 0} \{X^*_{ij}(\theta_k)/\alpha_{ij}(\theta_k)\} + \theta_k. \tag{5.44}$$

Removing these edges from $\mathcal{R}(\theta_k)$ yields $\mathcal{R}(\theta_{k+1})$, using which we can then solve for the optimal solutions, $U^*(\theta_{k+1})$ and $X^*(\theta_{k+1})$.

A pseudo-code algorithm for computing the boundary conditions θ_m's, $0 \le m \le M$, is given in Algorithm 5, and for computing the optimal flow solution, $X^*(\theta)$, is given in Algorithm 6. The inputs of the Algorithm 5 include the weight matrix W, the source and destination node ids, i.e. node 1 and node n. The outputs are the boundary condition sequence $[\theta_0 = 0, \theta_1, \ldots, \theta_M]$, and the correspond optimal flow distribution $[X^*(\theta_0), \ldots, X^*(\theta_M)]$. Initially, the optimal flow distributions of all-path potential based routing and the shortest path routing, denoted by $X^*(\theta_0 = 0)$ and $X^*(\infty)$, are computed at Lines 4-5. Then, the loop (Lines 6-10) generates each subsequent boundary condition θ_k, $(1 \le k)$, until the $X^*(\theta_k)$ converges to the shortest path solution. Given the outputs of Algorithm 5, for a certain $\theta \in [\theta_k, \theta_{k+1}]$, we know exactly which link has positive flow in the selected route, i.e. $X^*_{ij}(\theta_k) > 0$, based on the routing graph $\mathcal{R}(\theta_k)$. Algorithm 6 computes the optimal flow distribution $X^*(\theta)$ for any $\theta \ge 0$. The inputs of Algorithm 6 include the system parameters, W, the source and destination node ids, and the outputs of Algorithm 5, i.e. the bound condition sequence $[\theta_1, \ldots, \theta_M]$, and corresponding $[X^*(\theta_1), \ldots, X^*(\theta_M)]$. For a given θ, the output of Algorithm 6 is the corresponding optimal flow distribution matrix $X(\theta)$.

Complexity analysis. Since each step of the recursive process involves solving a set of linear equations, the worse case complexity of which is $O(n^3)$, and M is at most $|E|$ (the number of edges), the worst-case complexity of computing the *entire* routing continuum is $O(n^3|E|)$, or $O(n^5)$ in the worst case.

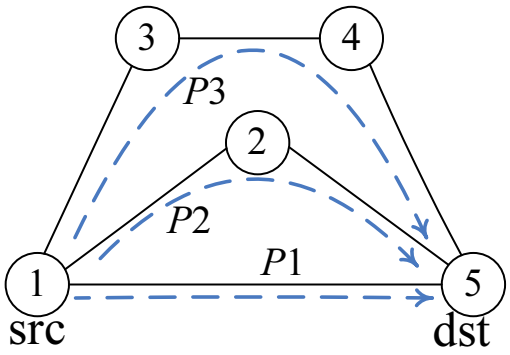


Figure 5.1: Example 1 with uniform weight $w_{ij} = 1$

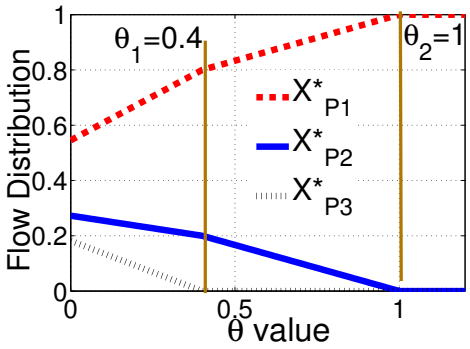


Figure 5.2: Flow distribution evolution of graph Fig. 5.1

Algorithm 5 Routing Continuum Algorithm

1: **INPUT:** Weight matrix W, source-id 1, destination-id n;
2: **OUTPUT:** Boundary vector $[\theta_0 = 0, \ldots, \theta_M]$, the corresponding $[X^*(\theta_0), \ldots, X^*(\theta_M)]$;
3: $k = 0$;
4: Compute All-path flow distribution $X^*(\theta_0 = 0)$ from eq.(5.15)(5.16);
5: Compute the shortest path flow distribution $X^*(\infty)$ from eq.(5.11)(5.12);
6: **while** $(X^*(\theta_k) \neq X^*(\infty))$ **do**
7: $k = k + 1$;
8: Compute θ_k and $X^*_{ij}(\theta_k)$ using eq.(5.44) and eq.(5.43) respectively;

5.4.2 Numerical Illustration

We now use two synthetic networks and a real network to show how the routing continuum grows as the parameter θ changes. Fig. 5.1 shows an example topology, with three disjoint paths between the source 1 and the destination 5, and each link is with 1-unit weight. As the parameter $\theta \geq 0$ increases, the longer paths $P3 = \{1 \to 3 \to 4 \to 5\}$ and $P2 = \{1 \to 2 \to 5\}$ are truncated gradually, and the shortest path $P1 = \{1 \to 5\}$ is obtained when θ increases to 1. Fig. 5.2 shows the routing continuum, i.e. the optimal flow distributions at

Algorithm 6 Solving the Optimal Flow Distribution $X^*(\theta)$, for a $\theta \geq 0$.

1: **INPUT:** $[\theta_0 = 0, \ldots, \theta_M]$, $[X^*(\theta_0), \ldots, X^*(\theta_M)]$,$\theta$, W, source-id 1, destination-id n;
2: **OUTPUT:** Optimal flow distribution $X^*(\theta)$;
3: Search for the interval, such that $\theta \in [\theta_k, \theta_{k+1})$;
4: Compute the $X^*(\theta)$ from eq.(5.43), for θ.

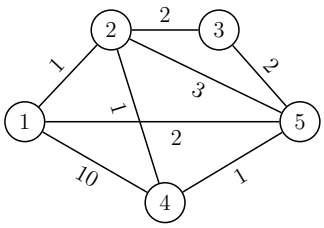


Figure 5.3: Example 2: Weight Graph

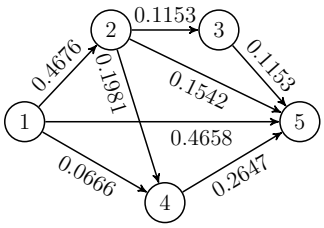


Figure 5.4: Flow distribution, $\theta_0 = 0$

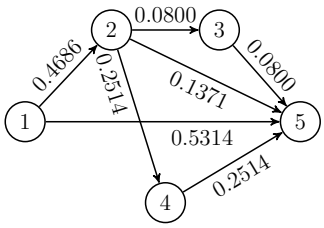


Figure 5.5: Flow distribution with $\theta_1 = 0.0914$

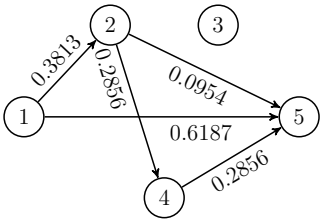


Figure 5.6: Flow distribution with $\theta_2 = 0.2850$

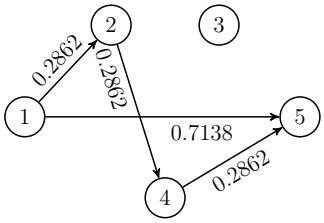


Figure 5.7: Flow distribution with $\theta_3 = 0.5700$

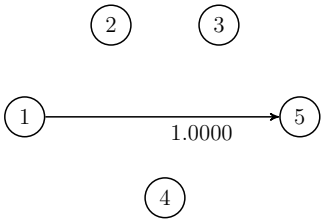


Figure 5.8: Flow distribution with $\theta_4 = 2$

each θ. We see that within the interval $\theta \in [0, 0.4]$, the flows on the longer paths $P2$ and $P3$ get linearly redistributed to the shortest path $P1$, and the longest path $P3$ gets truncated when $\theta = 0.4$. Then the flow of the second longest path $P2$ keeps decreasing as θ increases, until the second boundary condition $\theta = 1$ holds, where $P2$ is truncated. During the routing evolution process, the network flows are always redistributed from longer paths to the shorter path, while increasing θ. When $\theta > 1$, namely, the largest boundary condition, the routing solution is stabilized to the shortest path, i.e. $P1$.

Fig. 5.3 shows another example with five connected nodes in the topology. Weights w_{ij}'s are marked on the links. The flow initiates at source 1 and is removed from destination 5. Fig. 5.4~Fig. 5.8 show the optimal flow distributions (marked on individual links) under five boundary conditions, $[\theta_0 = 0, \theta_1 = 0.0914, \theta_2 = 0.2850, \theta_3 = 0.5700, \theta_4 = 2]$, In Fig. 5.4($\theta_0 = 0$), every link is active and follows the potential based "all-path" routing. Then, as θ increases to $\theta_1 = 0.0914$ (in Fig.5.5), link $(1, 4)$ is truncated, and within the interval $\theta \in [0, \theta_1]$, only the flow on path $\{1 \to 4 \to 5\}$ decreases, and gets redistributed to other paths, because this path with total length 11 is the longest path in $\mathcal{P}(0)$, i.e. the "all-path" routing graph. Then, when θ increases to $\theta_2 = 0.2850$, the flows on links $(2, 3)$ and $(3, 5)$ are truncated, because these two

links are on the second longest path $\{1 \to 2 \to 3 \to 5\}$, with path length 5. Similarly, when θ keeps increasing to θ_3 and θ_4, the rest two longer paths $\{1 \to 2 \to 5\}$ and $\{1 \to 2 \to 4 \to 5\}$ get removed, respectively, and only the shortest path $\{1 \to 5\}$ is left at last.

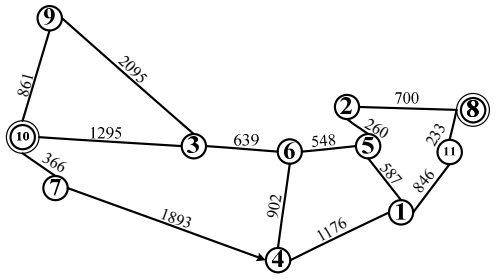


Figure 5.9: Weights on the Abilene network

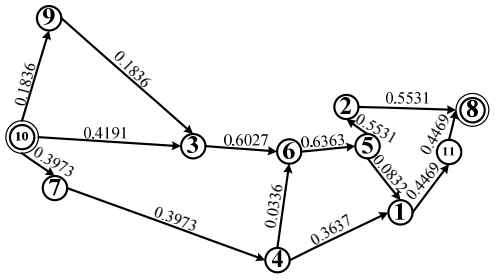


Figure 5.10: Flow distribution $\theta_0 = 0$

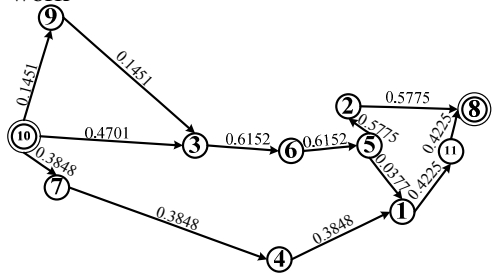


Figure 5.11: Flow distribution $\theta_1 = 0.1082$

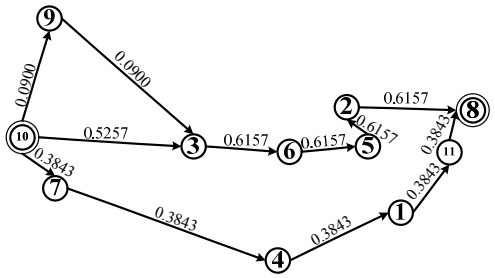


Figure 5.12: Flow distribution $\theta_1 = 0.2498$

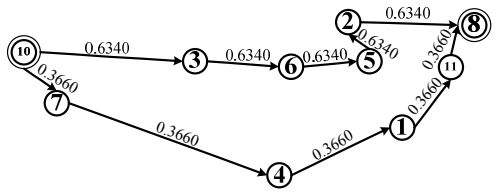


Figure 5.13: Flow distribution $\theta_3 = 0.4943$

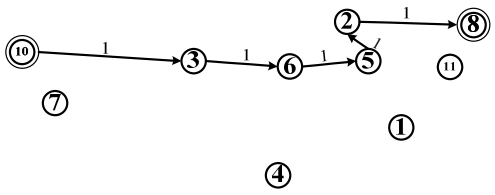


Figure 5.14: Flow distribution $\theta_4 = 3.2108$

Now, we apply the routing continuum theory to Internet2 Abilene Network [103]. The Abilene network was a high-performance backbone network established by the Internet2 community in the late 1990s. The Abilene Network was retired and became the "Internet2 Network"

in 2007. Fig. 5.15 shows its 11 regional network aggregation points and backbone connections across them (primarily OC192 or OC48 backbone). We consider the transmission cost between two end points roughly proportional to their actual geographic distance, because the velocity of light in an optical fiber becomes 60-70% compared to it in vacuum [104–106]. Hence, in the numerical analysis, we simply use the geographical distance as the link weight for the transmission cost as marked in Fig. 5.9. We choose the flow demand from `Sunnyvale` to `New York`. As we increase θ from 0, we observe a sequence of five boundary θ's, i.e., $[\theta_0 = 0, \theta_1 = 0.1082, \theta_2 = 0.2498, \theta_3 = 0.4943, \theta_4 = 3.2108]$, in which order links $(4 \to 6)$, $(5 \to 1)$, $(10 \to 9 \to 3)$ and $(10 \to 7 \to 4 \to 1 \to 11 \to 8)$ get truncated in sequence, and the optimal flow distribution evolves from the "all-path" routing to "shortest-path" routing. When $\theta_0 = 0$, all paths are present in delivering the contents, whereas only the shortest path $\{10 \to 3 \to 6 \to 5 \to 2 \to 8\}$ is active for $\theta \geq 3.2108$.

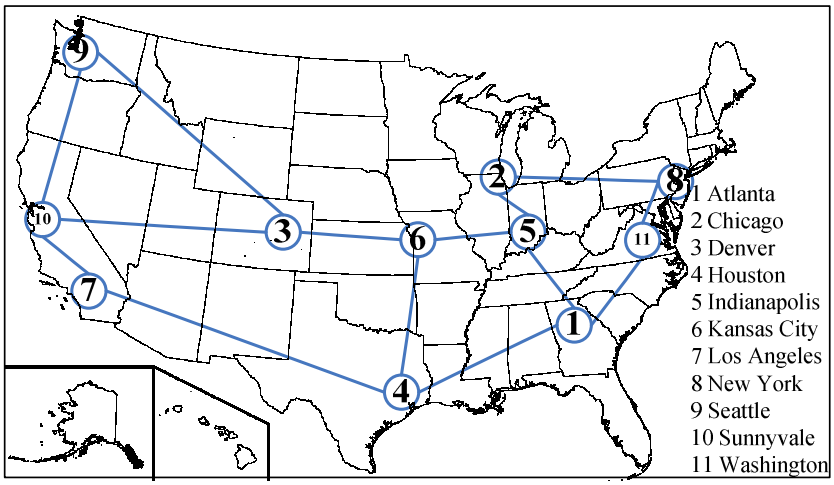


Figure 5.15: Abilene network topology

5.5 Generalizations and Applications

In this section, we present some extensions to the mixed L_1/L_2-norm network flow optimization, and briefly touch on their potential applications to traffic engineering and wireless sensor networks.

5.5.1 Multiple Flows, Link/node Capacity Constraints and Traffic Engineering

In the previous sections, for simplicity we have assumed a single flow of unit 1 from source node 1 to destination n. The formulation can be easily extended to accommodate multiple flows [95, 107, 108] between different source-destination pairs and with different units, as flows are additive on (links of) the network. Consider K flows, where the k-th flow $X^{(k)}$ of $I^{(k)}$ units is routed from source node s_k to destination node t_k, $1 \leq k \leq K$. Thus each flow $X^{(k)}$ satisfies the following conservation constraints:

$$\sum_{j:(i,j)\in E} X_{ij}^{(k)} - \sum_{l:(l,i)\in E} X_{li}^{(k)} = \begin{cases} I^{(k)} & \text{if } i = s_k \text{ (Src)} \\ -I^{(k)} & \text{if } i = t_k \text{ (Dst)} \\ 0 & \text{if } i \neq s_k, t_k, \end{cases} \tag{5.45}$$

We use $\mathcal{F}_k$ to denote the collection of flows satisfying eq.(5.45). Then the mixed L_1/L_2-norm *multi-flow* optimization is given in eq.(5.46). It is not too hard to see that this problem can be decomposed into K subproblems, each of which forms a single-flow mixed L_1 and L_2-norm optimization problem, and thus can be solved using the method presented before.

$$\min_{\substack{X^{(k)}\in\mathcal{F}_k \\ 1\leq k\leq K}} \sum_{k=1}^{K}\sum_{i=1}^{n}\sum_{j=1}^{n} (w_{ij}{X_{ij}^{(k)}}^2 + 2\theta w_{ij} X_{ij}^{(k)}) \tag{5.46}$$

subject to $X_{ij}^{(k)} \geq 0$, $1 \leq i, j \leq n$, $1 \leq k \leq K$.

In addition to having multiple flows (demands), many practical network flow problems, e.g., traffic engineering in a data network, also impose the *link capacity* constraints [109, 110]. Namely, given a network $G = (V, E)$, for each edge $(i, j) \in E$, let $C_{ij} (= C_{ji})$ denote the link capacity. Then the total amount of flows on link (i, j) cannot exceed C_{ij}. Given any set of K flows, $X^{(k)} \in \mathcal{F}_k$, $1 \leq k \leq K$, let α be a variable representing the maximum link utilization in the network, i.e., $\sum_k X_{ij}^{(k)} \leq \alpha C_{ij}$. Similar to [95], we consider the following maximum link utilization optimization and mixed L_1/L_2-norm flow optimization with link capacity constraints (where $\epsilon = \theta^{-1}$):

Capacity Constrained Mixed Flow Optimization (Prime):

$$\min_{\substack{X^{(k)}\in\mathcal{F}_k \\ 1\le k\le K}} \alpha + \sum_{k=1}^{K}\sum_{i=1}^{n}\sum_{j=1}^{n}\left(\frac{\epsilon}{2}w_{ij}{X_{ij}^{(k)}}^2 + w_{ij}X_{ij}^{(k)}\right) \tag{5.47}$$

$$\textit{subject to } X_{ij}^{(k)} \ge 0,\ 1\le k\le K\textit{; and} \tag{5.48}$$

$$\sum_{k=1}^{K} X_{ij}^{(k)} \le C_{ij}\alpha, 1\le i,j\le n. \tag{5.49}$$

Let $U_i^{(k)}$ be the Lagrange multipliers for the flow conservation constraints eq.(5.45), and s_{ij} the Lagrange multipliers for the inequality constraints $\sum_{k=1}^{K} X_{ij}^{(k)} - C_{ij}\alpha \le 0$. Then the dual problem is given by

Capacity Constrained Mixed Flow Optimization (Dual):

$$\max_{U,\mathrm{s}} \sum_{k=1}^{K} I^{(k)}U_1^{(k)}$$
$$-\frac{1}{2}\sum_{k=1}^{K}\sum_{i=1}^{n}\sum_{j:U_i^{(k)}-U_j^{(k)}>w_{ij}+s_{ij}} \frac{(U_i^{(k)}-U_j^{(k)}-(w_{ij}+s_{ij}))^2}{\epsilon w_{ij}}$$
$$\textit{subject to } s_{ij}\ge 0, \sum_{i=1}^{n}\sum_{j=1}^{n} s_{ij}C_{ij} = 1, \text{ and } U_n^{(k)} = 0.$$

Let α^* and ${X^{(k)}}^*_{ij}$'s be the optimal solution to the primal problem and ${U^{(k)}}^*_i$'s and s_{ij}^*'s the optimal solution to the dual problem. Then by the complementary slackness, we have ${X^{(k)}}^*_{ij} > 0$ if and only if ${U^{(k)}}^*_i - {U^{(k)}}^*_j \ge w_{ij} + s_{ij}^*$; and furthermore, if $s_{ij}^* > 0$, then $\sum_{k=1}^{K} {X^{(k)}}^*_{ij} = C_{ij}\alpha^*$. The latter implies that any link $(i,j)\in E$ with $s_{ij}^* > 0$ is a "bottleneck" link where the (optimal) maximum link utilization is attained. We see that on a bottleneck link (i,j), if ${X^{(k)}}^*_{ij} > 0$, then ${X^{(k)}}^*_{ij} = ({U^{(k)}}^*_i - {U^{(k)}}^*_j - (w_{ij}+s_{ij}^*))/(\epsilon w_{ij})$; whereas on a non-bottleneck link (i.e., $s_{ij}^* = 0$), if ${X^{(k)}}^*_{ij} > 0$, then ${X^{(k)}}^*_{ij} = ({U^{(k)}}^*_i - {U^{(k)}}^*_j - w_{ij})/(\epsilon w_{ij})$.

Comparing this with the optimal flow solutions to the mixed L_1/L_2-norm without the capacity constraints, an additional $s_{ij}^*/(\epsilon w_{ij})$ amount is reduced from each flow ${X^{(k)}}^*_{ij}$ on the bottleneck links (i,j). Intuitively, it is as if the weights on the bottleneck links were replaced with $w'_{ij} = w_{ij} + s_{ij}^*$ to discourage and shift away flows on the bottleneck links. In fact, suppose s_{ij}^*'s are known *a priori*. We can convert the network flow optimization eq.(5.47) *with link capacity constraints* to one (*without link capacity constraints*) as eq.(5.46), where w_{ij}'s in the

L_1-norm term are replaced by $w'_{ij} := w_{ij} + s^*_{ij}$'s, *but not* those in the L_2-norm term. This yields an example of network flow optimization with *heterogenous L_1/L_2 costs* to be discussed in the next subsection. Intuitively, this implies that the optimal flow with link capacity constraints that minimizes overall maximum link utilization is the one that discourages the usage of bottleneck links by increasing the (L_1) link costs on these links and thus shifting flows away from them.

Finally, for each $\epsilon(= \theta^{-1})$, we can use the optimal solution $\alpha^*(\theta)$ to eq.(5.47) to determine the best trade-offs between using shorter paths and longer paths, namely, the best routing graph $\mathcal{R}(\theta)$ which minimizes the overall network link utilization $\alpha(\theta)$ among any choice of $\theta \geq 0$:

$$\theta^* := \text{argmin}_{\theta \geq 0}\, \alpha^*(\theta). \tag{5.50}$$

In general, with the link capacity constraints, finding the optimal θ^* requires search in the solution space, $\{\theta : \alpha^*(\theta)\}$. On the other hand, assuming that w_{ij}'s are fixed, we can find the optimal θ^* in polynomial time by first computing the entire routing continuum using Algorithms 5 and 6, and then calculating the corresponding maximum link utilization $\alpha^*(\theta) := \max_{(i,j)}\{\sum_k {X^{(k)}}^*_{ij}/C_{ij}\}$ for each $\theta > 0$. Thus with respect to a fixed set of link weights w_{ij}'s, the routing graph $\mathcal{R}(\theta^*)$ yields the best trade-offs in usage of shorter and long paths: *it minimizes the overall network utilization among all routing graphs*.

Moreover, practical network flow problems, e.g., routing in bandwidth constrained wireless networks, may involve *node capacity* constraints [111–113], where for node $i \in V$, with node capacity C_i, the total amount of flows going through node i cannot exceed C_i. For a set of K flows, $X^{(k)} \in \mathcal{F}_k$, $1 \leq k \leq K$, $\sum_k \sum_j X^{(k)}_{ij} \leq \xi C_i$ holds true, where ξ is the maximum node capacity utilization in the network. While considering node capacity constrained mixed L_1/L_2-norm flow optimization problem, similar results can be obtained as the link capacity constrained L_1/L_2-norm optimization problem. We omit the details here for brevity.

5.5.2 Flow Optimization with Heterogeneous L_1/L_2 Link Weights

We consider the following generalization where L_1-norm and L_2-norm have different sets of link weights, w_{ij}'s and r_{ij}'s:

Flow Optimization with Heterogeneous L_1/L_2 Weights (Prime):

$$\min_{\substack{X^{(k)}\in\mathcal{F}_k \\ 1\le k\le K}} \sum_{k=1}^{K}\sum_{i=1}^{n}\sum_{j=1}^{n}\left(r_{ij}{X_{ij}^{(k)}}^2 + 2\theta w_{ij}X_{ij}^{(k)}\right), \tag{5.51}$$

$$\textit{subject to } X_{ij}^{(k)} \ge 0,\ 1 \le k \le K.$$

We have already seen one instance of such generalization in the application of traffic engineering with link/node capacity constraints. Another application arises more naturally in wireless sensor networks, where deciding on the best strategies hinge on trading off different cost considerations [5], e.g., transmission latency as well as energy consumption – the latter is important, for example, to maximize the sensor network life time, where it is shown in [94] that potential-based routing using L_2-norm maximizes the network life time. Let w_{ij}'s denote the per-hop transmission latency, and r_{ij}'s be the transmission energy costs. Then, eq.(5.51) represents the mixed L_1/L_2-norm network flow optimization problem with *heterogeneous* L_1/L_2 link weights. The dual problem can be formulated as follows:

Flow Optimization with Heterogeneous L_1/L_2 Weights (Dual):

$$\max_{U,\mathrm{s}} \sum_{k=1}^{K} I^{(k)}U_1^{(k)} - \frac{1}{2}\sum_{k=1}^{K}\sum_{i=1}^{n}\sum_{j:(i,j)\in E^{(k)}(\theta)} \frac{(U_i^{(k)}-U_j^{(k)}-\theta w_{ij})^2}{r_{ij}}$$

$$\textit{subject to } U_n^{(k)} = 0, 1 \le k \le K,$$

where $E^{(k)}(\theta)$ is the edge set, link $(i,j) \in E^{(k)}(\theta)$ if and only if $U_i^{(k)} - U_j^{(k)} > \theta w_{ij}$. Let ${X^{(k)}}_{ij}^*$'s and ${U^{(k)}}_i^*$'s be the optimal solution to the primal and dual problems, respectively. By complementary slackness, we have ${X^{(k)}}_{ij}^* = ({U^{(k)}}_i^* - {U^{(k)}}_j^* - \theta w_{ij})/r_{ij} > 0$ if and only if ${U^{(k)}}_i^* - {U^{(k)}}_j^* > \theta w_{ij}$. Using this relation, we can generalize Lemma 9 as below:

$$\theta \sum_{\langle i,j\rangle\in P^{(k)}} w_{ij} < {U^{(k)}}_1^*(\theta) \le \theta \sum_{\langle i,j\rangle\in Q^{(k)}} w_{ij} + \sum_{\langle i,j\rangle\in Q^{(k)}} r_{ij}, \tag{5.52}$$

where $P^{(k)}$ is a routing path with nonzero flow ${X^{(k)}}^*$ from source s_k to destination t_k (i.e., $P^{(k)} \in G_{{X^{(k)}}^*}$), whereas $Q^{(k)}$ is an arbitrary (simple) path in the network G from source s_k to destination t_k. For any given $\theta \ge 0$ and $1 \le k \le K$, using eq.(5.52) we can again characterize all paths in the routing graph $\mathcal{R}^{(k)}(\theta)$ – the routing graph induced by ${X^{(k)}}^*(\theta)$: for any $P \in \mathcal{R}^{(k)}(\theta)$, its path length, $|P| < L_{\min,1}^{(k)} + \theta^{-1}L_{\max,2}^{(k)}$, where $L_{\min,1}^{(k)} := \min_{Q^{(k)}} \sum_{\langle i,j\rangle\in Q^{(k)}} w_{ij}$ is the path length of the shortest (in terms of L_1 link weights) paths from s_k to t_k, and

$L^{(k)}_{\max,2} := \max_{Q^{(k)} \in \mathcal{R}^{(k)}(0)} \sum_{\langle i,j \rangle \in Q^{(k)}} r_{ij}$ is the the path length of the *longest* (in terms of L_2 link weights) paths in the routing graph $\mathcal{R}^{(k)}(0)$, the routing graph induced by the optimal L_2-norm flow $X^{(k)}(0)$. Therefore, we can establish a generalized *routing continuum* theorem analogous to Theorem 14, yielding a finite sequence of routing graphs, $\mathcal{R}^{(k)}(\theta^{(k)}_m)$'s. Furthermore, the boundary conditions for $\theta^{(k)}_m$'s can be precisely characterized using a similar iterative process as presented in Section 5.4.1, and Algorithms 5 and 6 can be analogously generalized to compute the entire routing continuum and $\{X^{(k)^*}(\theta), 1 \leq k \leq K\}$ for all $\theta > 0$. We omit the details here for brevity.

5.5.3 Network robustness analysis via generalized centrality measure

The optimal flow distribution $X^*(\theta)$ to the mixed L_1- and L_2-norm network flow optimization problem indicates exactly the loads on each link (resp. node) for certain flow demands. When considering flow demands from all source destination pairs, the average network flow on each link (resp. node) infers the "importance" of the link (resp. node), namely, the influence of the link (resp. node) in case of failure or being attacked, which in turn reveals the robustness structure of networks, i.e., which area of the network is more vulnerable to attacks. The robustness centrality measure of links and nodes in the network has been extensively studied, and has been applied to design topology control algorithm and routing protocol in wireless sensor networks and delay tolerant networks [114, 115]. Below, we show how our routing continuum theory can be used to generalize various robustness centrality measures of links/nodes in networks, where the ranking of links/nodes in terms of their betweenness infer the network robustness structure, i.e., those areas with high betweenness links/nodes expose more risks to attacks or failures, as when removing these links/nodes, more flows have to be rerouted or failed.

Centrality measures for mixed network flow

Centrality measures were first developed in social network analysis [116,117], for example, how influential a user is in a social network, with applications in robust community detection [118, 119], mobility prediction [120], and etc. There are four widely used centrality measures [117], that capture the relative importance of a vertex or an edge within a network from various aspects:

degree [4] , *eigenvector centrality* [5] , *betweenness* [123, 124], and *closeness* [125]. Betweenness and closeness centrality measures are directly interpretable in terms of the shortest path and all-path routing, thus can be generalized using our routing continuum theory to account for mixed network flows. In the following, we will introduce the mixed-flow betweenness for nodes or edges, a natural generalization of the existing betweenness centrality measures. The mixed-flow betweenness measures indicate the importance of nodes or edges in terms of the degree to which a node or an edge is participating in the communication between node pairs in the network, which has implications in network resource relocations and detecting robust subgraphs that are resilient to attacks and failures. Note that closeness centrality can similarly be generalized, and we omit these results here for brevity.

Node Betweenness centrality

Node betweenness has been studied in the past as a measure of the centrality and influence of nodes in networks.

Shortest-path betweenness. A simple example of such a betweenness measure initially proposed by Freeman [124, 126, 127] is shortest-path betweenness. Given a node i, its shortest-path betweenness is defined as the number of shortest (geodesic) paths between pairs of all other nodes that run through i. To be precise, given a graph $G = (V, E)$, node i's betweenness centrality [124, 126] C_i^S is defined as[6]

$$C_i^S = \frac{2\sum_{s<t\in V} g_i^{(st)}}{n(n-1)}, \tag{5.53}$$

where $g_i^{(st)}$ is the number of shortest paths from node s to node t that pass through i. Since the graph is undirected, $g_i^{(st)} = g_i^{(ts)}$ always holds, thus computing $g_i^{(st)}$ for only half of all node pairs (i.e., for $s < t$) is sufficient. If there is more than one shortest path between a node pair,

[4] The node degree centrality is simply defined as the number of links associated with a node, which reflects locally (i.e., within one hop,) how well the node is connected to other nodes.

[5] Eigenvector centrality takes the leading Eigenvector, i.e., the Eigenvector corresponding to the largest Eigenvalue, of the adjacent matrix A as relative scores to all nodes in the network, which follows the concept that connections to nodes with higher scores contribute more to the score of the node than connections to nodes with lower scores. PageRank [121] and Katz centrality [122] can be viewed as two variations of the Eigenvector centrality measure.

[6] Here, the normalizing constant is $n(n-1)$, where i may also be a start or end node of a source destination pair. Some definitions only count for those node pairs without i as a start or end node, where the normalizing constant becomes $(n-1)(n-2)$ instead.

each path is given equal weight such that the total weight of all of the paths is unity. Since when θ is large enough, the optimal flow distribution denoted by $X^*(\infty)$ represents the shortest path solution, the shortest path betweenness centrality C_i^S can be written as

$$C_i^S = \frac{2\sum_{s<t\in V}\sum_{k\in V} {X^{(st)}}^*_{ki}(\infty)}{n(n-1)}. \tag{5.54}$$

Current-flow betweenness[7] **.** Considering that the circuit created by placing a resister on each edge of the network and unit current source and destination at a particular node pair. The resulting current flow in the network will follow Kirchhoff's and Ohm's laws, going from source to destination along a multitude of paths. Hence, The current-flow betweenness [128] for a node i is defined as the absolute value of the currents summed over all node pairs that run through i. The optimal optimal flow distribution ${X^{(st)}}^*(0)$ of the L_2 norm flow optimization problem represents exactly the current flow for source destination pair (s,t) with $\theta = 0$. The current-flow betweenness C_i^C of node i can be written in terms of ${X^{(st)}}^*(0)$ as

$$C_i^C = \frac{2\sum_{s<t\in V}\sum_{k\in V} {X^{(st)}}^*_{ki}(0)}{n(n-1)}. \tag{5.55}$$

Mixed-flow betweenness. Shortest-path betweenness and current-flow betweenness present two extremes. One uses only shortest paths, and the other favors all-path to deliver network flow. Our routing continuum theory naturally leads to a generalized mix-flow betweenness, $C_i(\theta)$, which captures how much mixed flow $X^*(\theta)$ runs through a node given a flow combination parameter θ.

$$C_i(\theta) = \frac{2\sum_{s<t\in V}\sum_{k\in V} {X^{(st)}}^*_{ki}(\theta)}{n(n-1)}, \tag{5.56}$$

with $\theta \geq 0$. Note that the shortest-path betweenness (eq.(5.54)) and the current-flow betweenness (eq.(5.55)) are two special cases of mixed-flow betweenness, as $C_i^S = C_i(\infty)$ and $C_i^C = C_i(0)$, respectively. Given a specific $\theta \geq 0$, $C_i(\theta)$ captures the importance of node i, in terms of the average optimal flow going through node i over all source destination pairs.

Edge betweenness centrality

Analogically, the betweenness centrality can be defined for *edges*, capturing how much network flow going through a particular edge, summed over all node pairs in the network.

[7] **Current-flow betweenness is proven to be equivalent to random walk (RW) betweenness [128]. For a node i, we calculate the expected number of times that a random walk between a particular node pair will pass through i, and RW betweenness is the summation over all node pairs.**

The *shortest-path betweenness* of an edge (i, j) is the total number of shortest paths running along (i, j), which was first introduced by Anthonisse in [129], and Newman formally defined it in [118]. It can be written in terms of the optimal shortest path flow distribution denoted by $X^*(\infty)$ as

$$C_{ij}^S = \frac{2\sum_{s<t\in V} {X^{(st)}}_{ij}^*(\infty)}{n(n-1)} \tag{5.57}$$

Similarly, the *current-flow betweenness* of an edge (i, j) is the current flow running along (i, j) [118, 123], which can be computed using the following eq.(5.58) in terms of the optimal L_2 network flow distribution denoted by $X^*(0)$ as

$$C_{ij}^C = \frac{2\sum_{s<t\in V} {X^{(st)}}_{ij}^*(0)}{n(n-1)} \tag{5.58}$$

The mixed-flow betweenness of an edge (i, j) is then a natural generalization of eq.(5.57) and eq.(5.58) for $\theta \geq 0$.

$$C_{ij}(\theta) = \frac{2\sum_{s<t\in V} {X^{(st)}}_{ij}^*(\theta)}{n(n-1)} \tag{5.59}$$

As discussed earlier, the trade-off parameter $\theta \geq 0$ governs how much shortest path flow vs current flow is considered in the mixed flow optimization problem. The mixed-flow betweenness centrality measure for a link/node captures how crucial the link/node is in carrying the network flow for all possible node pairs. In communication networks, the link/node betweenness measures in fact indicate how much (mixed) network flow has to go through a particular link/node for all source-destination pairs. An attack or failure to the links/nodes with high betweenness leads to more influential impacts to the network traffic. Hence the ranking of the links/nodes in terms of their betweenness infer the network robustness structure, namely, areas with high betweenness links/nodes are more vulnerable to attacks or failures, since more flows have to be rerouted or failed if these links/nodes fail.

Numerical examples

Next, we use the topology in Fig. 5.3 and a real network topology, i.e., Internet2 Abilene Network [103], as examples, to show how the ranking of node/link in terms of betweenness changes over θ.

Table 5.1: Edge ranking in mixed-flow betweenness ($\theta \geq 0$).

Edge ranking	**when** $\theta \in [0, 0.002)$	**when** $\theta \in [0.002, 0.06)$	**when** $\theta \geq 0.06$
# 1	**(1,2)**	**(1,2)**	**(2,4)**
# 2	**(4,5)**	**(2,4)**	**(1,2)**
# 3	**(2,4)**	**(4,5)**	**(4,5)**
# 4	(2,3)	(2,3)	(2,3)
# 5	(3,5)	(3,5)	(3,5)
# 6	(1,5)	(1,5)	(1,5)
# 7	(2,5)	(2,5)	(2,5)
# 8	(1,4)	(1,4)	(1,4)

When computing the betweenness centrality measures for the five node topology in Fig. 5.3, we observe that as increasing $\theta \geq 0$, the ranking of links in terms of their mixed-flow betweenness keeps relatively robust, namely, there are only three different link ranking orders (See Tab 5.1). The highest betweenness links are $(1, 2)$, $(4, 5)$ and $(2, 4)$, which all have the smallest link weights. The link $(2, 4)$ steps up to the highest ranking, when $\theta \geq 0.06$. The node betweenness ranking is more stable, which is unchanged over θ's for topology in Fig. 5.3 with nodes ranked as $\{2, 5, 4, 1, 3\}$ in a decreasing order. Nodes with more links and lower link weights are ranked higher, since they are more likely to serve as hubs to carry more network flows.

Now we investigate how the link and node betweenness ranking vary over θ in Abilene network. As we increase $\theta \geq 0$, there are twelve boundary θ's, governing the different ranking of link betweenness in Abilene network as shown in Table 5.2. We observe that when θ changes, namely, the network flow evolves from "all-path" flow to "shortest-path" flow, the ranks of links with the highest betweeness keep high ranking over θ, i.e., links at rank # 1 to # 5 are unchanged. On the other hand, the betweenness centrality link $(3, 10)$ increases from the rank #11 to #6 gradually (as highlighted in Table 5.2), which happens because the high link weight of $(3, 10)$ suppress the "all-path (current)" flow going through it, but it resides on more shortest paths among node pairs, thus generates higher shortest-path flow when θ is large. Moreover, the ranks of links such as $(1, 11)$, $(2, 8)$, and $(8, 11)$ decrease as θ increases. The ranks of some other links, including $(1, 5)$ and $(4, 7)$, keep stable at rank #7-#9. When looking at the node betweenness, the ranking is more stable than links, which is unchanged for all θ's as shown in Table 5.3. The nodes placed in central US, such as `Kansas City` and `Indianapolis` posses highest node betweenness centrality, namely, being the busiest nodes in carrying network

Table 5.2: Edge ranking in mixed-flow betweenness in Abilene network ($\theta \geq 0$).

θ	0	0.0535	0.0617	0.0647	0.0839	0.0918
# 1	(5,6)	(5,6)	(5,6)	(5,6)	(5,6)	(5,6)
# 2	(3,6)	(3,6)	(3,6)	(3,6)	(3,6)	(3,6)
# 3	(2,5)	(2,5)	(2,5)	(2,5)	(2,5)	(2,5)
# 4	(7,10)	(7,10)	(7,10)	(7,10)	(7,10)	(7,10)
# 5	(1,4)	(1,4)	(1,4)	(1,4)	(1,4)	(1,4)
# 6	(2,8)	(2,8)	(2,8)	(2,8)	(2,8)	(2,8)
# 7	(8,11)	(8,11)	(8,11)	(8,11)	(4,7)	(4,7)
# 8	(4,7)	(4,7)	(4,7)	(4,7)	(8,11)	(1,5)
# 9	(1,11)	(1,5)	(1,5)	(1,5)	(1,5)	(8,11)
# 10	(1,5)	(1,11)	(1,11)	**(3,10)**	**(3,10)**	**(3,10)**
# 11	**(3,10)**	**(3,10)**	**(3,10)**	(1,11)	(1,11)	(1,11)

θ	0.0964	0.1079	0.1335	0.2650	0.2874	0.3176
# 1	(5,6)	(5,6)	(5,6)	(5,6)	(5,6)	(5,6)
# 2	(3,6)	(3,6)	(3,6)	(3,6)	(3,6)	(3,6)
# 3	(2,5)	(2,5)	(2,5)	(2,5)	(2,5)	(2,5)
# 4	(7,10)	(7,10)	(7,10)	(7,10)	(7,10)	(7,10)
# 5	(1,4)	(1,4)	(1,4)	(1,4)	(1,4)	(1,4)
# 6	(2,8)	(2,8)	(2,8)	(2,8)	**(3,10)**	**(3,10)**
# 7	(1,5)	(1,5)	(1,5)	**(3,10)**	(2,8)	(1,5)
# 8	(4,7)	(4,7)	**(3,10)**	(1,5)	(1,5)	(2,8)
# 9	(8,11)	**(3,10)**	(4,7)	(4,7)	(4,7)	(4,7)
# 10	**(3,10)**	(8,11)	(8,11)	(8,11)	(8,11)	(8,11)
# 11	(1,11)	(1,11)	(1,11)	(1,11)	(1,11)	(1,11)

flow.

We also computed the ranking of link/node betweenness in other real networks, such as Roofnet [130] (with 38 nodes), CERNET [131] (with 36 nodes), GEANT [132] (with 23 nodes), where similar results are obtained and we omit them here for brevity. From all these results, the node betweenness centrality ranking is overall more stable than link betweenness centrality ranking, through the entire routing continuum, i.e., all $\theta \geq 0$.

5.6 Related Work and Discussion

Routing in networks has been extensively studied under practical settings, with a literature too vast to cite completely. Here we will mention a few that are most relevant. For example, the

Table 5.3: Node betweenness ranking in the Abilene network for all θ's.

Rank	**# 1**	**# 2**	**# 3**	**# 4**	**# 5**
City	**Kansas City**	**Indianapolis**	**Denver**	**Atlanta**	**Sunnyvale**
# 6	**# 7**	**# 8**	**# 9**	**# 10**	**# 11**
Houston	**Chicago**	**Los Angeles**	**New York**	**Washington**	**Seattle**

authors in [133] propose an optimization model for QoS routing protocol design with multiple L_1-norm performance objectives, where the objectives are linearly combined with tunable parameter. In the context of traffic engineering in IP data networks, the authors in [134] show that given a set of traffic demands, optimizing the link weights in a network (assuming shortest-path routing) is NP-hard, and develop heuristics. The authors in [135] propose a new link-state routing protocol *PEFT* that goes beyond shortest paths by allowing longer paths and splitting traffic over multiple paths with an exponential penalty on longer paths. Via convex optimization, the authors show PEFT achieves optimal traffic engineering. The studies in [136–139] analyze the trade-offs between shortest path routing and multi-path routing in both wired and wireless network settings.

Different from earlier works, which focus on routing protocol designs for specific (wired/wireless) network scenarios, our work studies routing from a more general and theoretical perspective. It is partly inspired by the finding in [95], where motivated by traffic engineering in IP networks, the authors show that shortest path routing results from the optimal flow minimizing the L_1-norm in a network. In contrast, the optimal flow minimizing the L_2-norm in a network and its connection to currents in resistive electrical networks (and random walks on a graph) are well-known (see [89] and references thereof); it leads to the potential-based, "all-path" (or stochastic) routing that has been applied in wireless sensor networks, e.g., to maximize network life time [94], or to minimize state maintenance [140]. Our work generalizes these earlier results to show that using the mixed L_1/L_2-norm flow optimization, we can construct the entire routing continuum from the shortest-path to all-path, with routing graphs consisting of paths of increasing path lengths.

In a broader context, the mixed L_1/L_2 optimization formulation has been widely used, e.g., in the classical LASSO problems [141], namely, the least square optimization problems with a L_1-norm penalty term, and more recently, in compressive sensing [142, 143]. It is therefore

well-known that the L_1-norm penalty forces the least-square solution, X^*, to meet certain sparsity constraints, i.e., $||X^*||_1 \leq \epsilon$. Compared with LASSO and compressive sensing settings, our setting has a set of additional flow conservation constraints — these are what makes the problem *unique* and leads to solutions that have interesting interpretations and consequences, where the solutions to the more general LASSO and compressive sensing settings may not have, *apart from the sparsity of the solutions*.

Indeed, Lemma 9 shows that our results can be interpreted in terms of the "sparsity" of the solutions also: the optimal flow solution $X^*(\theta)$ to the mixed L_1/L_2-norm flow optimization leads to a *sparser* routing graph, where the *path length* of routes used for routing the optimal flow from a source to a destination can not be $(1 + \theta^{-1})$ longer than the shortest paths. More surprising and interesting is that we can generate the *entire routing continuum* from the mixed L_1/L_2-norm flow optimization. The flow conservation constraints in fact play a key role here: it leads to the duality of the optimal flows, $X^*(\theta)$, a function defined on the *edges* of a network, and the optimal (*generalized*) potential functions, $U^*(\theta)$, a function defined on the *nodes* of a network. This allows us to solve $U^*(\theta)$ through a set of linear equations, and yields an efficient process to compute the entire routing continuum and the optimal flow $X^*(\theta)$ for any $\theta \geq 0$. Last but not the least, we remark that although we only focus on network routing in this chapter, we believe that our results can be applied to many other applications where the problems can be cast in terms of flows in a network.

Another line of works that is related to our study is parametrized dissimilarity measure (or distance) between nodes. Yen *et al.* [144] develop a family of link-based dissimilarity measures, namely, the randomized shortest-path (RSP) dissimilarity, which generalizes both the weighted shortest path distance and the resistance distance. It is interpreted as the path probability distribution that minimizes the expected energy for transiting from a source node to a destination node, constrained by a fixed relative entropy (Kullback-Leibler divergence) with respect to the reference probability. Chebotarev [145] introduces a similar parametric family of node distance to [144] by matrix forest theorem and the transition inequality, which possess a unique graph-geodetic property: $d(i, j) + d(j, k) = d(i, k)$ if and only if every path from i to k passes through j. Different from our work, the constraints exploited in these works are no longer flow conservation law, thus the solutions obtained have different interpretations of the underlying "flow".

5.7 Summary

In this chapter, we have formulated the network routing problem as flow optimization problem in a network with mixed L_1/L_2-norms. Using this formulation, we established a surprising result: the routing graphs induced by the optimal flow solutions span the entire routing continuum from the shortest-path to all-path routing. Using the duality theory, we also developed an efficient iterative process for computing the entire routing continuum and optimal flow solutions $X^*(\theta)$ for any $\theta \geq 0$. The basic theory is further extended to account for multiple flows (traffic demands), link capacity constraints and heterogeneous L_1/L_2 link weights, with applications to traffic engineering and wireless sensor networks, and network robustness analysis.

Chapter 6

Conclusion, Discussion, and Future Research Directions

This dissertation presents my work on analyzing and characterizing diverse link patterns in complex networks in theory and practice, with various applications to opportunistic wireless networks, the Internet, and multiplex online social networks. The presented works focus on developing theories and models for understanding and employing versatile and oblivious network information – asymmetrical characteristics of the wireless transmission channels, multiplex social relations, e.g., trust and distrust relations, etc – in solving various application problems. In Chapter 2, we discuss how to generalize the random walk theory and the intrinsically related spectral graph theory from undirected graphs to directed graphs. Chapter 3 investigates and develops a novel social community detection algorithm for social networks with both uni- and bi-directional links. Chapter 4 establishes routing continuum theory for undirected communication networks, that spans from shortest path routing to the "potential" based all path routing, based on the connection between routing and network flow optimization problems.

Another line of my PhD research work is empirically analyzing, understanding, and making sense of the big data from real complex networks [146–150]. Many real world networks, such as the Internet, online social networks, video sharing websites, are large scale in natural, which generate a gigantic set of big data. For example, based on its own counting [151], YouTube serves a total of more than 2 billion views a day. According to a recent study [152], YouTube traffic contributes to a significant portion of inter-domain network traffic. Foursquare [153],

one of the most popular location based social networks (LBSNs), had more than 10 million registered users with 1 billion check-ins in September 2011 [154]. Characterizing such large scale networks, e.g., revealing and estimating the network statistics, including the network size, degree distributions, is of great interest, which may shed light on the network capacity needed, network traffic dynamics, and users' personal preferences and mobility patterns, with various applications to content pre-fetching, personalized recommendation, and online advertising. However, in most cases, complete network datasets are not available, or very costly to obtain. Thus it is important to develop efficient sampling algorithms to characterize these networks. In [146, 147, 149], we design efficient algorithms to sample and characterize YouTube, the largest video sharing website, and Foursquare, a leading location based social network (LBSN). Moreover, in [148], we evaluate how various sampling techniques perform in estimating the graph mixing time and demonstrate that bias in sampling algorithms accepted in the literature is rather metric-dependent.

6.1 Future directions

My future research plan on analyzing and characterizing versatile complex networks include carrying out active research in the following areas.

Analyzing the dynamics of complex networks. In reality, complex networks are dynamic in natural, i.e., nodes and edges join and leave over time. As one of my future research directions, I am interested in understanding, modeling, and interpreting the dynamics of various complex networks, such as online social and media networks, the Internet, wireless (cellular) networks, and investigating how the dynamics of these networks affect and be affected by the local and global network topological structure. Firstly, I intend to further explore some unaddressed problems from my existing work, including studying the dynamics and evolution of YouTube video statistics and users behaviors to reveal how (fast) YouTube is growing in size and what motivate, drive and prompt the YouTube uploaders to upload videos. These studies will not only imply how much the storage and bandwidth are needed, but also provide a basis to study the socio-economic factors in online video sharing and distribution. Moreover, in location based social networks, I intend to investigate various characteristics of venues during their life spans, to uncover factors that drive venues becoming popular, i.e., attracting a large number of visits from users over time, which are directly beneficial to applications, including venue recommendation

and targeted advertising in LBSNs.

Investigating multivariate phenomena in multi-attributed complex networks. Entities in complex networks usually possess multiple attributes that reflect their different characteristics, for example, the numbers of users, who follow and be followed by a Twitter user, indicate the popularity and the activeness of the Twitter user, respectively. The emphasis in complex networks has been on analyzing and characterizing individual attributes separately, e.g., degree distributions as univariate distributions, which have implications on issues such as social influence, community formation, resilience, and epidemics. However, multivariate phenomena that arise as a result of repeated non-linear interactions among its participants are mostly ignored, where valuable information about the correlations across different network attributes is lost. Thus, bringing multivariate probabilistic models and analytics to bear in complex network analysis will provide a more rigorous theoretical underpinning, and will lead to deeper insights into many network phenomena. I am interested in developing efficient algorithms to characterize the multivariate statistics and structures of complex networks through sampling, e.g., joint in- and out-degree distributions. The results by analyzing the sampled data will infer correlations and causality in complex networks, which in turn have implications in revealing more precise network evolution process with respect to multiple attributes, and uncovering hidden social communities in the underlying networks.

Social-aware communication networks. In recent years, online social networks (OSNs) are becoming increasingly popular, which have generated a great interest in studying and understanding the users' behavior patterns and preferences, with applications to personalized friendship, item, venue recommendations and targeted online advertising. While such social information is useful and beneficial to directly improve the user experience and impression of the ads on online social networks, it is also profitable to improve the performances of the underlying data forwarding networks. I am interested in incorporating those social information, e.g., the video preferences of YouTube users, mobility patterns of mobile phone users, and the social community structure, to design intelligent social-aware content pre-fetching and caching mechanisms for communication networks, including content delivery networks (CDNs), content-centric networks (CCNs), and wireless (cellular) networks. These designs will facilitate and advance the transmission layer network performances, such as reducing the data transmission delay and promoting the data traffic congestion control.

Bibliography

[1] David Knoke and James H. Kuklinski. *Network analysis*. Beverly Hills: Sage Publications (ISBN: 0-8039-1914-X), 1982.

[2] Robert A. Hanneman and Mark Riddle. *Introduction to social network methods*. Riverside, CA: University of California, Riverside (published in digital form at http://faculty.ucr.edu/~hanneman/), 2005.

[3] Yanhua Li and Zhi-Li Zhang. Random walks on digraphs, the generalized digraph laplacian and the degree of asymmetry. *WAW'10: 7th Workshop on Algorithms and Models for the Web Graph*, pages 74–85, 2010.

[4] Yanhua Li and Zhi-Li Zhang. Digraph laplacian and the degree of asymmetry. *Internet Mathematics*, 8(4), 2012.

[5] Yanhua Li and Zhi-Li Zhang. Random walks on digraphs: A theoretical framework for estimating transmission costs in wireless routing. In *INFOCOM'10: The 29th IEEE International Conference on Computer Communications*, pages 1–9. IEEE, 2010.

[6] Yanhua Li, Wei Chen, and Zhi-Li Zhang. Optimal forwarder list selection in opportunistic routing. In *MeshTech'09: Third IEEE International Workshop on Enabling Technologies and Standards for Wireless Mesh Networking*, pages 670–675. IEEE, 2009.

[7] Yanhua Li, Wei Chen, and Zhi-Li Zhang. Design of forwarder list selection scheme in opportunistic routing protocol. Technical report, Technical Report TR-08-033, CSE Department of University of Minnesota, 2008.

[8] Yanhua Li and Zhi-Li Zhang. Random walks and green's function on digraphs: A framework for estimating wireless transmission costs. *IEEE/ACM Transactions on Networking*, 21(1):135–148, 2013.

[9] Yanhua Li, Zhi-Li Zhang, and Daniel Boley. The routing continuum from shortest-path to all-path: A unifying theory. In *ICDCS'11: 31st International Conference on Distributed Computing Systems*, pages 847–856. IEEE, 2011.

[10] Yanhua Li, Wei Chen, Yajun Wang, and Zhi-Li Zhang. Influence diffusion dynamics and influence maximization in social networks with friend and foe relationships. In *WSDM'13: The 6th ACM International Conference on Web Search and Data Mining*, pages 1–10. ACM, 2013.

[11] Yanhua Li, Zhi-Li Zhang, and Jie Bao. Mutual or unrequited love: Identifying stable clusters in social networks with uni- and bi-directional links. In *WAW'12: The 9th Workshop on Algorithms and Models for the Web Graph*, pages 113–125, 2012.

[12] Sergey Brin and Lawrence Page. The anatomy of a large-scale hypertextual web search engine. *Computer Networks and ISDN Systems*, 30(1–7):107–117, 1998.

[13] Làszlò Lovàsz. Random walks on graphs: A survey. *Combinatorics*, 2:1–46, 1993.

[14] Fan R. K. Chung. *Spectral Graph Theory (CBMS Regional Conference Series in Mathematics, No. 92)*. Cbms Regional Conference Series in Mathematics, 2006.

[15] Fan R. K. Chung. Laplacians and the cheeger inequality for directed graphs. *Annals of Combinatorics*, 9:1–19, sep 2005.

[16] Dengyong Zhou, Jiayuan Huang, and Bernhard Schölkopf. Learning from labeled and unlabeled data on a directed graph. In *ICML '05: Proceedings of the 22nd international conference on Machine learning*, pages 1036–1043, New York, NY, USA, 2005. ACM.

[17] Roger Horn and Charles R. Johnson. *Matrix Analysis*. Cambridge University Press, 1st edition, 1985.

[18] Fan R. K. Chung and S. T. Yau. Discrete green's functions. *Journal of Combinatorial Theory, Series A*, pages 191–214, July 2000.

[19] David Aldous and Jim Fill. Reversible markov chains and random walks on graphs. *http://www.stat.berkeley.edu/~aldous/RWG/book.html*.

[20] Daniel Boley, Gyan Ranjan, and Zhi-Li Zhang. Commute times for a directed graph using an asymmetric laplacian. *Linear Algebra and its Applications*, 435(2):224–242, 2011.

[21] Rajendra Bhatia. *Matrix Analysis*. Springer, 1997.

[22] George A. F. Seber. *A Matrix Handbook for Statisticians*. Wiley-Interscience, New York, NY, USA, 2007.

[23] Dennis S. Bernstein. *Matrix Mathematics: Theory, Facts, and Formulas with Application to Linear Systems Theory*. Princeton University Press, New York, NY, USA, 2005.

[24] Stephen Boyd, Arpita Ghosh, Balaji Prabhakar, and Devavrat Shah. Mixing times for random walks on geometric random graphs. In *SIAM Workshop on Analytic Algorithmics and Combinatorics (ANALCO)*, pages 240–249, Vancouver, 2005. SIAM.

[25] Mark Jerrum and Jung-Bae Son. Spectral gap and log-sobolev constant for balanced matroids. *the 43rd Annual IEEE Symposium on Foundations of Computer Science (FOCS 2002)*, 2002.

[26] James Allen Fill. Eigenvalue bounds on convergence to stationarity for nonreversible Markov chains, with an application to the exclusion process. *Ann. Appl. Probab.*, 1(1):62–87, 1991.

[27] Milena Mihail. Conductance and convergence of markov chains-a combinatorial treatment of expanders. *Proceedings of FOCS'1989*, pages 526–531, 1989.

[28] Alvin W Wolfe. Social network analysis: Methods and applications. *American Ethnologist*, 24(1):219–220, 1997.

[29] Peter DeScioli, Robert Kurzban, Elizabeth N Koch, and David Liben-Nowell. Best friends alliances, friend ranking, and the myspace social network. *Perspectives on Psychological Science*, 6(1):6–8, 2011.

[30] Mohsen Jamali, Gholamreza Haffari, and Martin Ester. Modeling the temporal dynamics of social rating networks using bidirectional effects of social relations and rating patterns. In *WWW'11: Proceedings of the 20th international conference on World wide web*, pages 527–536. ACM, 2011.

[31] Alvin W Gouldner. The norm of reciprocity: A preliminary statement. *American sociological review*, pages 161–178, 1960.

[32] Scott A Golder, Sarita Yardi, Alice Marwick, and Danah Boyd. A structural approach to contact recommendations in online social networks. In *Workshop on Search in Social Media, SSM*, 2009.

[33] Miklós Kurucz, Andras Benczur, Károly Csalogány, and László Lukács. Spectral clustering in telephone call graphs. In *Proceedings of the 9th WebKDD and 1st SNA-KDD 2007 workshop on Web mining and social network analysis*, pages 82–91. ACM, 2007.

[34] Scott White and Padhraic Smyth. A spectral clustering approach to finding communities in graphs. In *Proceedings of the fifth SIAM international conference on data mining*, volume 119, page 274. Society for Industrial Mathematics, 2005.

[35] Nina Mishra, Robert Schreiber, Isabelle Stanton, and Robert E Tarjan. Clustering social networks. In *Algorithms and Models for the Web-Graph*, pages 56–67. Springer, 2007.

[36] Dengyong Zhou, Jiayuan Huang, and Bernhard Schölkopf. Learning from labeled and unlabeled data on a directed graph. In *ICML*, pages 1036–1043, 2005.

[37] Venu Satuluri and Srinivasan Parthasarathy. Symmetrizations for clustering directed graphs. In *Proceedings of the 14th International Conference on Extending Database Technology*, EDBT/ICDT '11, pages 343–354, New York, NY, USA, 2011. ACM.

[38] Jason Weston, Christina S. Leslie, Eugene Ie, Dengyong Zhou, André Elisseeff, and William Stafford Noble. Semi-supervised protein classification using cluster kernels. *Bioinformatics*, 21(15):3241–3247, 2005.

[39] Elizabeth A Leicht and Mark EJ Newman. Community structure in directed networks. *Physical review letters*, 100(11):118703, 2008.

[40] Fritz Heider. Attitudes and cognitive organization. *The Journal of psychology*, 21(1):107–112, 1946.

[41] Ellen Berscheid and Pamela C Regan. *The psychology of interpersonal relationships*. Pearson Prentice Hall Upper Saddle River, NJ, 2005.

[42] Harold Miller and Dennis Geller. Structural balance in dyads. *Journal of Personality and Social Psychology*, 21(2):135–38, 1972.

[43] Kendall O Price, Ernest Harburg, and Theodore M Newcomb. Psychological balance in situations of negative interpersonal attitudes. *Journal of Personality and Social Psychology*, 3(3):265, 1966.

[44] Aroldo Rodrigues. Effects of balance, positivity, and agreement in triadic social relations. *Journal of Personality and Social Psychology*, 5(4):472, 1967.

[45] Haewoon Kwak, Hyunwoo Chun, and Sue Moon. Fragile online relationship: a first look at unfollow dynamics in twitter. In *Proceedings of the 2011 annual conference on Human factors in computing systems*, pages 1091–1100. ACM, 2011.

[46] Leo Katz and James H Powell. Measurement of the tendency toward reciprocation of choice. *Sociometry*, 18(4):403–409, 1955.

[47] Ulrike Luxburg. A tutorial on spectral clustering. *Statistics and Computing*, 17:395–416, December 2007.

[48] Jianbo Shi and Jitendra Malik. Normalized cuts and image segmentation. *IEEE Trans. Pattern Anal. Mach. Intell.*, 22(8):888–905, 2000.

[49] Xiang Wang and Ian Davidson. Flexible constrained spectral clustering. In *KDD*, 2010.

[50] Ulrike von Luxburg. A tutorial on spectral clustering. Technical Report No.TR-149, Max Planck Institute for Biological Cybernetics, 2006.

[51] Lars Hagen and Andrew B Kahng. New spectral methods for ratio cut partitioning and clustering. *Computer-aided design of integrated circuits and systems, ieee transactions on*, 11(9):1074–1085, 1992.

[52] Slashdot. dataset. *http://snap.stanford.edu/data/soc-Slashdot0811.html*.

[53] Matthew Richardson, Rakesh Agrawal, and Pedro Domingos. Trust management for the semantic web. In *The Semantic Web-ISWC 2003*, pages 351–368. Springer, 2003.

[54] Jure Leskovec, Jon M. Kleinberg, and Christos Faloutsos. Graph evolution: Densification and shrinking diameters. *TKDD*, 1(1), 2007.

[55] David Kempe, Jon Kleinberg, and Éva Tardos. Maximizing the spread of influence through a social network. In *Proceedings of the ninth ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 137–146. ACM, 2003.

[56] Masahiro Kimura and Kazumi Saito. Tractable models for information diffusion in social networks. In *PKDD*, 2006.

[57] Jure Leskovec, Andreas Krause, Carlos Guestrin, Christos Faloutsos, Jeanne M. VanBriesen, and Natalie S. Glance. Cost-effective outbreak detection in networks. In *KDD*, 2007.

[58] Eyal Even-Dar and Asaf Shapira. A note on maximizing the spread of influence in social networks. In *WINE*, pages 281–286. Springer, 2007.

[59] Ramasuri Narayanam and Y. Narahari. Determining the top-k nodes in social networks using the shapley value. In *AAMAS*, 2008.

[60] Wei Chen, Yajun Wang, and Siyu Yang. Efficient influence maximization in social networks. In *KDD*, 2009.

[61] Wei. Chen, Chi. Wang, and Yajun. Wang. Scalable influence maximization for prevalent viral marketing in large-scale social networks. In *KDD*, 2010.

[62] Wei. Chen, Yifei. Yuan, and Li. Zhang. Scalable influence maximization in social networks under the linear threshold model. In *ICDM*, 2010.

[63] Amit Goyal, Wei Lu, and Laks V. S. Lakshmanan. Simpath: An efficient algorithm for influence maximization under the linear threshold model. In *ICDM*, 2011.

[64] Amit Goyal, Francesco Bonchi, and Laks V. S. Lakshmanan. A data-based approach to social influence maximization. *PVLDB*, 5(1):73–84, 2008.

[65] Nishith Pathak, Arindam Banerjee, and Jaideep Srivastava. A generalized linear threshold model for multiple cascades. In *ICDM*, 2010.

[66] Jure Leskovec, Daniel Huttenlocher, and Jon Kleinberg. Predicting positive and negative links in online social networks. In *WWW*, 2010.

[67] Jure Leskovec, Daniel Huttenlocher, and Jon Kleinberg. Signed networks in social media. In *CHI*. ACM, 2010.

[68] Kai-Yang Chiang, Nagarajan Natarajan, Ambuj Tewari, and Inderjit S Dhillon. Exploiting longer cycles for link prediction in signed networks. In *CIKM*, pages 1157–1162. ACM, 2011.

[69] David Easley and Jon Kleinberg. *Networks, Crowds, and Markets: Reasoning About a Highly Connected World*. Cambridge, 2010.

[70] Peter Clifford and Aidan Sudbury. A model for spatial conflict. *Biometrika*, 60(3):581–588, 1973.

[71] Richard A Holley and Thomas M Liggett. Ergodic theorems for weakly interacting infinite systems and the voter model. *The annals of probability*, pages 643–663, 1975.

[72] Fan Chung and Alexander Tsiatas. Hypergraph coloring games and voter models. *WAW '12: Proceedings of the 9th Workshop on Algorithms and Models for the Web Graph*, pages 1–16, 2012.

[73] Vishal Sood, Tibor Antal, and Sidney Redner. Voter models on heterogeneous networks. *Physical Review E*, 77(4):041121, 2008.

[74] M Angeles Serrano, Konstantin Klemm, Federico Vazquez, Víctor M Eguíluz, and Maxi San Miguel. Conservation laws for voter-like models on random directed networks. *Journal of Statistical Mechanics: Theory and Experiment*, 2009(10):P10024, 2009.

[75] Naoki Masuda and Hisashi Ohtsuki. Evolutionary dynamics and fixation probabilities in directed networks. *New Journal of Physics*, 11(3):033012, 2009.

[76] Jerome Kunegis, Stephan Schmidt, Andreas Lommatzsch, Jurgen Lerner, Ernesto W. De Luca, and Sahin Albayrak. Spectral analysis of signed graphs for clustering, prediction and visualization. In *SDM*, 2010.

[77] Christian Borgs, Jennifer Chayes, Adam Tauman Kalai, Azarakhsh Malekian, and Moshe Tennenholtz. A novel approach to propagating distrust. In *WINE*, pages 87–105. Springer, 2010.

[78] Shishir Bharathi, David Kempe, and Mahyar Salek. Competitive influence maximization in social networks. In *WINE*, 2007.

[79] Allan Borodin, Yuval Filmus, and Joel Oren. Threshold models for competitive influence in social networks. In *WINE*, 2010.

[80] Ceren Budak, Divyakant Agrawal, and Amr El Abbadi. Limiting the spread of misinformation in social networks. In *WWW*, 2011.

[81] Wei Chen, Alex Collins, Rachel Cummings, Te Ke, Zhenming Liu, David Rincón, Xiaorui Sun, Yajun Wang, Wei Wei, and Yifei Yuan. Influence maximization in social networks when negative opinions may emerge and propagate. In *SDM*, 2011.

[82] Hao Ma, Haixuan Yang, Michael R. Lyu, and Irwin King. Mining social networks using heat diffusion processes for marketing candidates selection. In *CIKM*, 2008.

[83] Xinran He, Guojie Song, Wei Chen, and Qingye Jiang. Influence blocking maximization in social networks under the competitive linear threshold model. In *SDM*, 2012.

[84] John Morgan, Ken Steiglitz, and George Reis. The spite motive and equilibrium behavior in auctions. *Contributions in Economic Analysis & Policy*, 2(1), 2003.

[85] Felix Brandt, Tuomas Sandholm, and Yoav Shoham. Spiteful bidding in sealed-bid auctions. In *IJCAI*, pages 1207–1214, 2007.

[86] William J Stewart. Numerical methods for computing stationary distributions of finite irreducible markov chains. In *Computational Probability*, pages 81–111. Springer, 2000.

[87] Epinions. Dataset. *http://www.epinions.com/*.

[88] Slashdot. Dataset. *http://slashdot.org/*.

[89] Frank P. Kelly. Network routing. *Philosophical Transactions of the Royal Society*, A(337):343–367, 1991.

[90] Gerald R. Ash. *Dynamic Routing in Telecommunications Networks*. McGraw-Hill Professional, 1st edition, 1997.

[91] Sanjit Biswas and Robert Morris. Exor: Opportunistic routing in multi-hop wireless networks. In *Proceedings of the ACM SIGCOMM '05 Conference*, Philadelphia, Pennsylvania, August 2005.

[92] Deepak Ganesan, Ramesh Govindan, Scott Shenker, and Deborah Estrin. Highly-resilient, energy-efficient multipath routing in wireless sensor networks. *SIGMOBILE Mobile Computing and Communications Review*, 5(4):11–25, October 2001.

[93] Lucian Popa, Costin Raiciu, Ion Stoica, and David Rosenblum. Reducing congestion effects in wireless networks by multipath routing. In *ICNP'06: IEEE International Conference on Network Protocols*, pages 96–105, Los Alamitos, CA, USA, 2006. IEEE.

[94] Haiyang Liu, Zhi-Li Zhang, Jaideep Srivastava, and Victor Firoiu. Pwave: A multi-source multi-sink anycast routing framework for wireless sensor networks. In *NETWORKING '07: Proceedings of the 6th International IFIP-TC6 Networking*, Atlanta, GA, USA, 2007. IFIP.

[95] Yufei Wang, Zheng Wang, and Leah Zhang. Internet traffic engineering without full mesh overlaying. In *Infocom '01: Proceedings of the 20th Annual Joint Conference of the IEEE Computer and Communications Societies*, Anchorage, Alaska, USA, 2001. IEEE.

[96] David G. Luenberger. *Linear and nonlinear programming (Second edition)*. Addison-Wesley, 1984.

[97] Peter G. Doyle and Laurie J. Snell. Random walks and electric networks. *Mathematical Association of America*, Jan 1984.

[98] A.K. Chandra, P. Raghavan, W.L. Ruzzo, R. Smolensky, and P. Tiwari. The electrical resistance of a graph captures its commute and cover times. In *STOC'89: Proceedings of Annual ACM Symposium on Theory of Computing*, pages 574–586, 1989.

[99] Douglas J Klein and M Randić. Resistance distance. *Journal of Mathematical Chemistry*, 12(1):81–95, 1993.

[100] Prisad Tetali. Random walks and the effective resistance of networks. *Journal of Theoretical Probability*, 4(10):1572–9230, 2005.

[101] Stephen Boyd and Lieven Vandenberghe. *Convex optimization*. Cambridge university press, 2004.

[102] Fan R. K. Chung and S. T. Yau. Discrete green's functions. *Journal of Combinatorial Theory, Series A*, pages 191–214, 2000.

[103] Abilene Project by Internet2 community. *http://www.internet2.edu/abilene/*.

[104] Stuart Cheshire. Latency survey results (for "it's the latency, stupid"), 1996.

[105] Kaoru Yoshida, Yutaka Kikuchi, Masateru Yamamoto, Yoriko Fujii, Kenichi Nagami, Ikuo Nakagawa, and Hiroshi Esaki. Inferring pop-level isp topology through end-to-end delay measurement. In *Passive and Active Network Measurement*, pages 35–44. Springer, 2009.

[106] Katsunari Okamoto. *Fundamentals of optical waveguides*. Academic press, 2010.

[107] Punyaslok Purkayastha and John S. Baras. An optimal distributed routing algorithm using dual decomposition techniques. *Communications in Information and Systems*, 8(3):277–302, 2008.

[108] Michael J Neely, Eytan Modiano, and Charles E Rohrs. Dynamic power allocation and routing for time-varying wireless networks. *Selected Areas in Communications, IEEE Journal on*, 23(1):89–103, 2005.

[109] Li Li, Marina Thottan, Bin Yao, and Sanjoy Paul. Distributed network monitoring with bounded link utilization in ip networks. In *INFOCOM '03: Proceedings of the 22th IEEE Conference on Computer Communications*, pages 1189–1198, Hong Kong, China, 2003. IEEE.

[110] Will Fisher, Martin Suchara, and Jennifer Rexford. Greening backbone networks: reducing energy consumption by shutting off cables in bundled links. In *Proceedings of the first ACM SIGCOMM workshop on Green networking*, Green Networking '10, pages 29–34, New Delhi, India, 2010. ACM.

[111] Alberto V Donati, Roberto Montemanni, Luca M Gambardella, and Andrea E Rizzoli. Integration of a robust shortest path algorithm with a time dependent vehicle routing model and applications. In *Computational Intelligence for Measurement Systems and Applications, 2003. CIMSA'03. 2003 IEEE International Symposium on*, pages 26–31. IEEE, 2003.

[112] Gil Zussman and Adrian Segall. Energy efficient routing in ad hoc disaster recovery networks. In *INFOCOM '03: Proceedings of the 22th IEEE Conference on Computer Communications*, Hong Kong, China, 2003. IEEE.

[113] Thomas Bonald, Matthieu Jonckheere, and Alexandre Proutière. Insensitive load balancing. In *ACM Sigmetrics Performance Evaluation Review*, volume 32, pages 367–377. ACM, 2004.

[114] Alfredo Cuzzocrea, Alexis Papadimitriou, Dimitrios Katsaros, and Yannis Manolopoulos. Edge betweenness centrality: A novel algorithm for qos-based topology control over wireless sensor networks. *Journal of Network and Computer Applications*, 35(4):1210–1217, 2012.

[115] Giuliano Grossi and Federico Pedersini. Hub-betweenness analysis in delay tolerant networks inferred by real traces. In *Modeling and Optimization in Mobile, Ad Hoc and Wireless Networks (WiOpt), 2010 Proceedings of the 8th International Symposium on*, pages 318–323. IEEE, 2010.

[116] Ulrik Brandes and Thomas Erlebach. *Network analysis: methodological foundations*, volume 3418. Springer, 2005.

[117] Mark EJ Newman. *Networks: An Introduction*. Oxford University Press, 1st edition, 2010.

[118] Mark EJ Newman and Michelle Girvan. Finding and evaluating community structure in networks. *Physical review E*, 69(2):026113, 2004.

[119] Michelle Girvan and Mark EJ Newman. Community structure in social and biological networks. *Proceedings of the National Academy of Sciences*, 99(12):7821–7826, 2002.

[120] Yaniv Altshuler, Rami Puzis, Yuval Elovici, Shlomo Bekhor, and Alex (Sandy) Pentland. Augmented betweenness centrality for mobility prediction in transportation networks. In *NEMO' 11: Proceedings of the international workship on finding patterns of human behaviors in network and mobility data*, 2011.

[121] Lawrence Page, Sergey Brin, Rajeev Motwani, and Terry Winograd. The pagerank citation ranking: bringing order to the web. 1999.

[122] Leo Katz. A new status index derived from sociometric analysis. *Psychometrika*, 18(1):39–43, 1953.

[123] Ulrik Brandes and Daniel Fleischer. Centrality measures based on current flow. 2005.

[124] Linton C. Freeman. A set of measures of centrality based on betweenness. *Sociometry*, 40(1):35–41, March 1977.

[125] Gert Sabidussi. The centrality index of a graph. *Psychometrika*, 31(4):581–603, 1966.

[126] Linton C. Freeman. Centrality in social networks conceptual clarification. *Social Networks*, 1(3):215–239, 1979.

[127] Stephen P Borgatti and Martin G Everett. A graph-theoretic perspective on centrality. *Social networks*, 28(4):466–484, 2006.

[128] Mark EJ Newman. A measure of betweenness centrality based on random walks. *Social Networks*, 27(1):39–54, January 2005.

[129] Jacob M Anthonisse. The rush in a directed graph. *Stichting Mathematisch Centrum. Mathematische Besliskunde*, (BN 9/71):1–10, 1971.

[130] MIT roofnet. *http://pdos.csail.mit.edu/roofnet/*.

[131] The China Education and Research Network (CERNET). *http://www.edu.cn/cernet_1377/index.shtml*.

[132] (GN3) GEANT Project. *http://www.geant.net/pages/home.aspx*.

[133] Piet Van Mieghem and Lieven Vandenberghe. Trade-off curves for qos routing. In *Infocom '06: Proceedings of the 25th Annual Joint Conference of the IEEE Computer and Communications Societies*, Barcelona, Catalunya, Spain, 2006. IEEE.

[134] Bernard Fortz and Mikkel Thorup. Increasing internet capacity using local search. *Computational Optimization and Applications*, 29(1):13–48, 2004.

[135] Dahai Xu, Mung Chiang, and Jennifer Rexford. Link-state routing with hop-by-hop forwarding can achieve optimal traffic engineering. In *INFOCOM 2008. 27th IEEE International Conference on Computer Communications*, pages 466–474, Phoenix, AZ, USA, 2008. IEEE.

[136] Peter Pham. Performance analysis of reactive shortest single-path and multi-path routing mechanism with load balance. In *INFOCOM 2003. 22nd IEEE International Conference on Computer Communications*, 2003.

[137] Israel Cidon, Raphael Rom, and Yuval Shavitt. Analysis of multi-path routing. *IEEE/ACM Transactions on Networking.*, 7:885–896, December 1999.

[138] Christopher L. Barrett, Stephan J. Eidenbenz, Lukas Kroc, Madhav V. Marathe, and James P. Smith. Probabilistic multi-path vs. deterministic single-path protocols for dynamic ad-hoc network scenarios. In *Proceedings of the 2005 ACM symposium on Applied computing*, SAC '05, pages 1166–1173, New York, NY, USA, 2005. ACM.

[139] Anindya Basu, Alvin Lin, and Sharad Ramanathan. Routing using potentials: a dynamic traffic-aware routing algorithm. In *SIGCOMM '03: Proceedings of the 2003 conference on Applications, technologies, architectures, and protocols for computer communications*, pages 37–48, New York, NY, USA, 2003. ACM Press.

[140] Chi-Kin Chau and Prithwish Basu. Exact analysis of latency of stateless opportunistic forwarding. In *INFOCOM '09: Proceedings of the 28th IEEE Conference on Computer Communications*, Rio de Janeiro, Brazil, 2009. IEEE.

[141] Robert Tibshirani. Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society. Series B (Methodological)*, pages 267–288, 1996.

[142] Yaakov Tsaig and David L. Donoho. Compressed sensing. *IEEE Transactions on Information Theory*, 52:1289–1306, 2006.

[143] Emmanuel J Candès and Michael B Wakin. An introduction to compressive sampling. *Signal Processing Magazine, IEEE*, 25(2):21–30, 2008.

[144] Luh Yen, Marco Saerens, Amin Mantrach, and Masashi Shimbo. A family of dissimilarity measures between nodes generalizing both the shortest-path and the commute-time distances. In *Proceedings of the 14th ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 785–793. ACM, 2008.

[145] Pavel Chebotarev. A class of graph-geodetic distances generalizing the shortest-path and the resistance distances. *Discrete Applied Mathematics*, 159(5):295–302, 2011.

[146] Jia Zhou, Yanhua Li, Vajay Kumar Adhikari, and Zhi-Li Zhang. Counting youtube videos via random prefix sampling. In *IMC'11: ACM SIGCOMM conference on Internet measurement conference*, pages 371–380. ACM, 2011.

[147] Yanhua Li, Moritz Steiner, Limin Wang, Zhi-Li Zhang, and Jie Bao. Dissecting foursquare venue popularity via random region sampling. In *CoNEXT'12 student workshop: The 8th International Conference on emerging Networking EXperiments and Technologies*, pages 1–2, 2012.

[148] Abedelaziz Mohaisen, Pengkui Luo, Yanhua Li, Yongdae Kim, and Zhi-Li Zhang. Measuring bias in the mixing time of social graphs due to graph sampling. In *MILCOM'12: The 2012 Military Communications Conference*, pages 1–6, 2012.

[149] Yanhua Li, Moritz Steiner, Limin Wang, Zhi-Li Zhang, and Jie Bao. Exploring venue popularity in foursquare. In *NetSciCom'13: Fifth IEEE International Workshop on Network Science for Communication Networks, in conjunction with INFOCOM'13*, 2013.

[150] Yanhua Li, Haiyong Xie, Yonggang Wen, and Zhi-Li Zhang. Coordinating in-network caching in content-centric networks: Model and analysis. In *ICDCS'13: 33st International Conference on Distributed Computing Systems*, pages 1–11. IEEE, 2013.

[151] Youtube statistics. *www.youtube.com/t/press_statistics*.

[152] C. Labovitz, S. Iekel-Johnson, D. McPherson, J. Oberheide, and F. Jahanian. Internet inter-domain traffic. In *ACM SIGCOMM Computer Communication Review*, volume 40, pages 75–86. ACM, 2010.

[153] Foursquare. *https://foursquare.com/*.

[154] A million check-ins isn't cool. you know what's cool?
http://blog.foursquare.com/2011/09/20/billion/.

Appendix A

Omitted Proofs

A.1 Properties of ergodic digraphs

Proposition 4 *Let $G = (V, E, A)$ be an ergodic digraph. For any nodes $i, j \in V$, there exist two paths from i to j with even and odd length, respectively.*

Proof : Suppose, for a contradiction, that all paths from i to j have even lengths. This implies that all cycles passing through i must be even length, since otherwise we could follow node i's odd-length cycle followed by the even length path from i to j, making the entire path from i to j odd. Now we can consider any cycle C_r in G, not necessarily passing i. We claim that C_r must have even length. In fact, we can pick any node u on C_r, and construct a path from i to j with the following segments: R_1 from i to u, C_r, R_2 from u back to i, and R_3 from i to j. Since we know that $R_1 + R_2$ has even length and R_3 has even length, it must be the case that C_r has even length by our assumption. However, this means that all cycles in C has even lengths, contradicting to the aperiodicity of G.

The case of odd length paths can be proved in the same way. ∎

Proposition 5 *Let $\bar{G} = (V, E, \bar{A})$ be an ergodic unsigned digraph, with transition probability matrix $\bar{P}$ and stationary distribution vector π. $\bar{P}^t - \mathbf{1}\pi^T = (\bar{P} - \mathbf{1}\pi^T)^t$ holds for any integer $t > 0$.*

Proof :

Using the facts that $\bar{P}\mathbf{1} = \mathbf{1}$ and $\pi^T\bar{P} = \pi^T$, it is easy to prove by induction that for any integer $t > 0$ $\bar{P}^t - \mathbf{1}\pi^T = (\bar{P} - \mathbf{1}\pi^T)^t$ holds. ∎

A.2 Special matrix power series

Proposition 6 *Let $X \in \mathbb{R}^{m\times m}$, $Y \in \mathbb{R}^{m\times n}$ and $Z \in \mathbb{R}^{n\times n}$. If $\lim_{t\to\infty} X^t = \lim_{t\to\infty} Z^t = \mathbf{0}$, the following equalities hold:*

$$\textit{(i)}\ \lim_{t\to\infty}\sum_{i=0}^{t-1} X^i = (I-X)^{-1}, \tag{A.1}$$

$$\textit{(ii)}\ \lim_{t\to\infty}\sum_{i=0}^{t-1} X^i Y Z^{t-1-i} = 0, \tag{A.2}$$

Proof : (i) Let $\rho(X)$ be the spectral radius of matrix X, i.e., the largest absolute value of the eigenvalues of X. Notice that $\lim_{t\to\infty} X^t = \mathbf{0}$ if and only if $\rho(X) < 1$.

We first claim that, $I-X$ and $I-Z$ are invertible. Suppose $I-X$ is not invertible, there is a non-zero vector p such that $(I-X)p = \mathbf{0}$. Therefore, p is the eigenvector of X with eigenvalue 1, which contradicts $\lim_{t\to\infty} X^t = \mathbf{0}$. Same argument can be applied to $I - Z$. Hence, the left hand side of Eq.(A.1) equals to

$$\lim_{t\to\infty}\sum_{i=0}^{t} X^i = \lim_{t\to\infty}(I-X)^{-1}(I-X^{t+1}) = (I-X)^{-1}.$$

(ii) The max-norm of X is given by $\|X\|_{max} = \max_{i,j\le m}\{X_{ij}\}$. Let $X = Q_X\mathcal{J}Q_X^{-1}$ be the standard Jordan form of X, where Q_X is an invertible matrix. Denote $J = \mathbf{1}\mathbf{1}^T$ as the all-one matrix. Hence, we have

$$\begin{aligned}\|X^i\|_{max} = \|Q_X\mathcal{J}^iQ_X^{-1}\|_{max} &\le \|Q_X\|_{max}\|Q_X^{-1}\|_{max}\|J\mathcal{J}^iJ\|_{max}\\ &\le \|Q_X\|_{max}\|Q_X^{-1}\|_{max}m^2\|\mathcal{J}^i\|_{max}\end{aligned}$$

$\mathcal{J}^i$ is in form as

$$\mathcal{J}^i = \begin{bmatrix} \lambda_1^i & C_i^1\lambda_1^{i-1} & C_i^2\lambda_1^{i-2} & 0 & 0\\ 0 & \lambda_1^i & C_i^1\lambda_1^{i-1} & 0 & 0\\ 0 & 0 & \lambda_1^i & 0 & 0\\ 0 & 0 & 0 & \lambda_{m_0}^i & C_i^1\lambda_{m_0}^{i-1}\\ 0 & 0 & 0 & 0 & \lambda_{m_0}^i \end{bmatrix}, \tag{A.3}$$

where $C_i^\ell = \frac{i!\ell!}{(i-\ell)!} \leq i^m$ and each non-zero entry in $\mathcal{J}^i$ can be expressed as $C_\ell^i \lambda_k^{i-\ell}$, $1 \leq k \leq m_0$, $1 \leq \ell \leq \ell_0(k)$, with m_0 as the number of different eigenvalues of X and $\ell_0(k)$ as the multiplicity of the k-th eigenvalue of X. Hence, the absolute value of each non-zero entry in $\mathcal{J}^i$ is upper bounded as $|C_i^\ell \lambda_k^{i-\ell}| \leq i^m \rho(X)^{i-m}$, which implies that

$$\|X^i\|_{max} \leq \|Q_X\|_{max} \|Q_X^{-1}\|_{max} m^2 i^m \rho(X)^{i-m}$$

Let $\rho = \max(\rho(X), \rho(Z))$, we have

$$\lim_{t\to\infty} \|\sum_{i=0}^{t-1} X^i Y Z^{t-1-i}\|_{max} \leq \lim_{t\to\infty} tmn \|X^i\|_{max} \|Y\|_{max} \|Z^{t-1-i}\|_{max}$$

$$\leq \lim_{t\to\infty} tmn T_{max} (m^2 t^m \rho^{i-m})(n^2 t^n \rho^{t-i-1-n}) \leq \lim_{t\to\infty} m^3 n^3 T_{max} t^{m+n+1} \rho^{t-1-n-m} = 0$$

where $T_{max} = \|Y\|_{max} \|Q_X\|_{max} \|Q_X^{-1}\|_{max} \|Q_Z\|_{max} \|Q_Z^{-1}\|_{max}$.

■

A.3 Illustration of exponential convergence time of P^t on ergodic digraph.

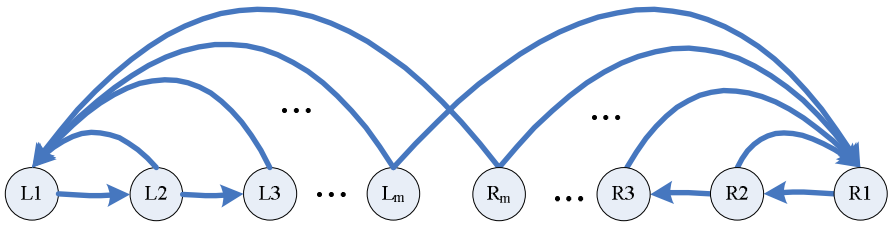


Figure A.1: An example digraph with exponential convergence time. All edges are with unit weights.

Given an unsigned ergodic digraph $\bar{G} = (V, E, \bar{A})$, with transition probability matrix $\bar{P}$, it has fixed stationary distribution π, i.e., $\pi^T = \pi^T \bar{P}$.

The convergence time (or mixing time) of a random walk Markov chain on G is the time until the Markov chain is "close" to its stationary distribution π. To be precise, for an initial distribution x_0, let $x_t^T = x_0^T \bar{P}^t$ be the distribution at step t. The variation distance mixing time is defined as the smallest t such that for any subset $W \subseteq V$,

$$|(x_t^T - \pi^T) e_W| \leq \frac{1}{4},$$

where e_W is the vector such that $e_W(i) = 1$ if $i \in W$, and $e_W(i) = 0$ if $i \in V \setminus W$.

The convergence time is said to be exponentially large if there exists x_0 such that the convergence time of the random walk starting from x_0 is $2^{\Omega(n)}$, where $n = |V|$. Lemma 10 below illustrates that the convergence time of random walk on ergodic digraphs could be exponentially large.

Lemma 10 *There exist ergodic digraphs, such that the convergence time of the random walks on these digraphs are exponentially large.*

Proof :

We prove this by construction. Fig. A.1 shows an example digraph G, with $|V| = 2m$ nodes. On the left hand side, there are $m \geq 3$ nodes $L_1, L_2, \cdots, L_m$ connected by $m-1$ directed edges from L_1 to L_m, and every node L_i with $i > 1$ has a directed connection to the leftmost node L_1. The right hand side nodes have symmetric connections as the left hand side. Moreover, node L_m and R_m also have one more connection to R_1 and L_1, respectively, which connect two components together. It is clear that the graph is strongly connected and aperiodic (there exist cycles of length 2 and 3), and thus ergodic.

Let $x_t(L_i)$ denote the probability that the random walk is at node L_i at step t, and $x(L_i)$ be its stationary distribution. Similarly define $x_t(R_i)$ and $x(R_i)$ for node R_i. The graph is symmetric, thus we have $x(L_i) = x(R_i)$ for $1 \leq i \leq m$. Let $x(L_1) = x(R_1) = \rho/4$, we have $x(L_i) = x(R_i) = \rho/2^i$ for $i = 2, 3, \ldots, m$. Then, by solving $\sum_{i=1}^{m}(x(L_i) + x(R_i)) = 1$, we obtain $\rho = \frac{2^{m-1}}{3 \cdot 2^{m-2}-1}$. It is easy to verify that indeed the obtained x is the stationary distribution of the random walks on the digraph.

Then, we consider the initial distribution as $x_0 = [1, 0, 0, \ldots, 0]$, and the subset $W = \{R_1, \cdots, R_m\}$ including all m nodes on the right-hand side. Let $x_t(W) = x_t^T \cdot e_W$ denote the total probability that the random walk is in some node in W at step t. The only edge from the left half to the right half is the edge from L_m to R_1. Thus all additions to $x_{t+1}(W)$ from $x_t(W)$ comes from this edge, namely $x_{t+1}(W) - x_t(W) \leq x_t(L_m)/2$. We now bound $x_t(L_m)$. For $t \leq m-1$, we know that $x_t(L_m) = 0$. For $t \geq m$, we have

$$x_t(L_m) = x_{t-1}(L_{m-1})/2 = x_{t-2}(L_{m-2})/2^2 = \cdots = x_{t-m+2}(L_2)/2^{m-2} \leq 1/2^{m-2}.$$

Hence, we have

$$x_t(W) = \sum_{i=1}^{t} (x_i(W) - x_{i-1}(W)) \leq t \cdot x_t(L_m)/2 \leq t/2^{m-1}.$$

Therefore, the smallest t that satisfies $|(x_t^T - \pi^T)e_W| = |x_t(W) - 1/2| \leq 1/4$ is such that $x_t(W) \geq 1/4$, which implies that $t/2^{m-1} \geq 1/4$ and $t \geq 2^{m-3}$. This completes the proof. ∎