Optimal Leader Selection in Multi-agent Networks: Joint Centrality, Robustness and Controllability

Katherine Fitch

A Dissertation
Presented to the Faculty
of Princeton University
in Candidacy for the Degree
of Doctor of Philosophy

Recommended for Acceptance
by the Department of
Mechanical and Aerospace Engineering
Adviser: Naomi E. Leonard

November 2016

Abstract

In many engineered and natural multi-agent networked systems, a limited subset of agents, called *leaders*, have access to external information, while the remaining agents, called *followers*, obtain information through network connections. In this dissertation, we connect measures of group-level performance with properties of the underlying network. Specifically, we determine which nodes in the network should be selected as leader nodes to maximize controllability and robustness of the leader-follower consensus dynamics to noise.

Maximizing robustness of the leader-follower consensus dynamics to noise is equivalent to minimizing steady-state system error defined in terms of variance about consensus. We define a new notion of centrality of a set of nodes, called *joint centrality*, that will be maximized by the optimal leader set for robustness to noise. We demonstrate that the optimal selection of a single leader for robustness to noise is the most information central agent. We show that, in general, node sets with high joint centrality balance high individual information centralities with coverage over the network.

For unweighted path and cycle graphs, we explicitly solve the optimal leader selection problem for robustness to noise. For unweighted tree graphs, we provide a simplification of two- and three-node joint centrality and present provable bounds for computationally efficient leader selection.

Centrality of a set of nodes is significant to a variety of network applications. We explore and illustrate the use of joint centrality in an example of synthetic lethality in *Saccharomyces cerevisiae* (baker's yeast), in an example of clustering in a Facebook social network, and in a network of political books frequently purchased together on Amazon.

In leader selection for controllability, we study *average controllability*, which measures the difficulty in controlling agents to any state in finite time, and *volume of*

the subspace reachable with one unit of control input. We prove that average controllability is maximized when the leader set is composed of the least information central nodes. We demonstrate that reachable volume is dependent on the left eigenvectors of the graph Laplacian corresponding to the leader nodes. We explore the fundamental trade-off between leader selection for robustness and leader selection for controllability.

Acknowledgements

I am very grateful to have had the privilege of working with with my advisor, Naomi Leonard, for the past five years. Naomi is an incredibly talented researcher with a unique ability to connect topics from engineering to a wide range of fields spanning from art to neuroscience. As an advisor, Naomi fosters creativity and independent thought, while always ready to provide assistance when needed - even when it requires hours of discussion about research ideas and directions. I have learned a great deal and have grown as a researcher due to her guidance. I thank Naomi for her consistent encouragement and support, and for making my time as a graduate student a rewarding experience.

I thank my committee members, Clancy Rowley and Phil Holmes for their reflections and valuable insights on this work. Additionally, I am grateful to Simon Levin who, along with Clancy, took the time to read and provide feedback on this thesis. The path to this dissertation started back in Syracuse and I am greatly indebted to many of my professors there. In particular I am grateful to the late Hiroshi Higuchi, who first suggested that I work with Naomi.

Throughout the past five years years I have had the privilege of working with many talented researchers as part of the Leonard Lab. I thank Darren Pais and George Young for their help with beginning my research project. I also thank past and present lab members Paul Reverdy, Will Scott, Brendan Andrade, Peter Landgren, Liz Davidson, Bec Gray, Renato Pagliara, Tian Shen and Stephanie Goldfarb, as well as post-docs Vaibhav Srivastava, Karla Kvaternik, Biswadip Dey and Kayhan Ozcimder. I am grateful to have worked with such an intelligent and fun group of people. Will and Brendan were in my first year cohort, along with Scott Dawson, Anthony DeGennaro, A J Eldorado and Katie Riggs and Steve and Heidi Atkinson, all of whom became good friends.

The MAE department is a vibrant and collegial community that I am very grateful

to have been a part of and will certainly miss. I thank Jill Ray for all her help with administrative matters and for always welcoming me to stop by and chat. Jill's involvement in the MAE department has undoubtedly enhanced graduate student life. I thank Michael Littman for the use of his darkroom equipment, which enabled me to pursue my love for black and white film photography. I thank Daisy Fitch and the late Professor Val Fitch for their kindness and for opening up their home to me, which provided a welcoming and peaceful environment to spend my summers.

My time at Princeton was enriched by a number of friendships that enhanced life outside of work. I am grateful for all the fun times spent on climbing adventures and stadium runs with Hannah Waight; early morning BBG workouts with Bec; stellar Rockville sessions with Valerie Caplan; and trips to the Bent Spoon with Naima Hammoud. I thank Lisa DiMatteo for being the person I could always count on and a steadfast friend throughout the years.

Finally, I thank my family for their lifelong love and support. My brother, Andy Fitch, has always been someone I look up to for his creativity, talent and willingness to help others. Our wanderings through NYC have often provided me with much needed perspective and laughter. I'm so grateful for the time we have spent together, especially over the past year. My parents, Ray and Karen Fitch, nurtured my affection for math from a young age and instilled the work ethic that has helped me achieve my goals thus far. They are my role models and I owe so much of who I am to their love and sacrifice. Always remember and never forget: you're the best, Mom and Dad.

This dissertation carries the number T-3325 in the records of the Department of Mechanical and Aerospace Engineering.

Contents

List of Tables

List of Figures

Chapter 1

Introduction

Networked multi-agent systems consist of two or more agents that are interacting, sensing, or communicating, with one another and possibly the surrounding environment. The generalizability of the networked multi-agent system framework allows for its use in wide-ranging fields of research. As a result, analysis of topics such as engineered systems, collective animal behavior, and social networks can all be conducted using similar mathematical tools [41, 85, 111]. Examples include fleets of underwater vehicles moving in the ocean that may be equipped to communicate with one another and have sensors to measure ocean currents, or fish in a school that interact by sensing the relative motion of near neighbors. A common objective of these kinds of applications is to understand how the structure of the inter-agent communication, or *network topology*, affects the flow of information throughout the network and subsequently the *collective dynamics*.

Agents in a complex network have different levels of influence on the flow of information, depending on where in the network they are located and whether or not they are interacting with the external environment. We can think of an agent that is communicating with many other agents in the network as *central* in the network, and therefore important in the spread of information. This notion becomes significantly

more complex when thinking of the centrality of *sets* of agents in a network because it mandates knowledge of how members of the set *jointly* influence the network. In other words, we can not simply consider the influences or connections of each agent independently, but rather how agents influence the network relative to other members of the set.

When studying collective dynamics of a networked multi-agent system, one of the most fundamental questions we can ask is whether or not the states of agents in the group can reach *consensus*. In other words, are the agents communicating with one another in such a way that the states of all agents will eventually be in agreement with each other. Consensus of multi-agent systems is a key factor for many group level behaviors, such as decision making where agents' states may represent opinions about a decision. Furthermore, many problems that require synchronization, such as formation control [3], can be cast as consensus problems.

We can look beyond basic consensus by considering systems with agents also capable of measuring external, or environmental, information. Often in networked systems, only a limited subset of agents will have access to external information. This is because in both natural and engineered systems it is frequently more costly, i.e. requires advanced sensing capabilities, to acquire external information than it is to communicate with other agents in the network. For example in a robotic swarm, inter-agent communication can be accomplished at low-energy cost through bluetooth or similar technology, while measuring information from the environment typically requires additional sensors and more battery power. Therefore, it can be more efficient for a small subset of agents to have direct access to external information and for the remaining agents to obtain information through network connections. We call the agents accessing external information *leaders* and all remaining agents *followers*. Whenever two agents are communicating or interacting we refer to them as *neighbors* and we assume that all communication between two agents is bi-directional. We

not that not all networks of interest can be assumed as bi-directional and thus the analysis here cannot be applied in those cases. The assumption of bi-directional communication is used to simplify and improve tractability of our analysis.

In natural systems, such as herds of migrating animals, the locations of agents sensing the environment is viewed as an emergent phenomenon [37]. In [71], the authors extended a mathematical model [100] to analyze this evolutionary dynamic and to compute the location of emergent leaders as a function of the network graph and the investment cost. The model yields a distributed adaptive dynamic for taking on leadership in this context; however, the evolutionary dynamics do not guarantee a steady-state solution that is optimal for the herd.

In engineered systems we would like to determine the best locations of leaders so as to optimize the collective dynamics. Alternatively posed, we seek to determine where information should enter the network so as to optimize the performance of the system as a function of the network topology. Merging components of natural and engineered leader selection by designing networks where emergent leadership gives rise to high performing systems is an area of ongoing research.

The notion of external information discussed thus far could represent a number of different functions or signals. In this work we consider two scenarios. In the first, the external information is a static or slowly time-varying signal that the group is seeking to estimate, and reach consensus about, through measurements by the leader agents and inter-agent communication. For the second case, the external information is a (possibly time-varying) control signal. Here, each agent in the network achieves a desired state as a result of inter-agent communication and the control input which directly affects the states of the leader nodes.

All real world systems are subject to disturbances and noise. Therefore, it is critical for the leader-follower consensus dynamics to be sufficiently *robust*. Robustness of a multi-agent system can be considered with respect to a variety of parameters,

including performance under communication failures, agent failures, environmental disturbances or noise, and communication uncertainty. While all facets of robustness are important to leader-follower multi-agent systems, in this work we focus specifically on the robustness of the leader-follower consensus dynamics to environmental noise. For systems in which the leaders states are influenced by a control input, we would like the networks to be *controllable*. That is, there exists a control input such that each agent can obtain any state in finite time. It is possible that a system is controllable but takes nearly infinite input energy to achieve a target final state and because of this, we use measures of network *controllability* to quantify how controllable a system is. For many systems, one cannot simultaneously maximize both robustness to noise and controllability, therefore trade-offs must be made if both properties are desired.

This dissertation provides a bridge between performance of group level collective dynamics, such as controllability and robustness of the leader-follower consensus dynamics to noise, and properties of the underlying network topology represented by a network graph. To do so, we analytically determine network-level properties of leader sets that will give rise to optimal group behavior.

1.1 Leader Selection for Robustness

When we are confronted with the problem of choosing the locations of leaders in a graph to maximize accuracy of consensus about an external signal despite the presence of noise, we refer to this as the problem of *leader selection for robustness*. Furthermore, throughout this thesis, the use of the term robustness implies robustness of the leader-follower consensus dynamics to noise. This problem has received a significant amount attention in recent years [16, 24, 38, 49, 56, 58, 59, 74, 79]

Much of the research related to leader-follower multi-agent systems with linear stochastic dynamics has been focused on the development of off-line leader selection

algorithms that seek to find the leader set that minimizes total system error, defined as the sum of individual steady-state variances about the stationary external signal [16, 24, 58, 59, 74]. Total system error can be viewed as a measure of coherence, equivalently the H_2 norm of the system dynamics [74, 109]. In [74] a greedy algorithm was proposed to find the optimal set of m leaders. The authors argued that their greedy algorithm may be too computationally intensive for very large networks, and they derived alternative algorithms that use a bound on the total system error to improve efficiency. These and similar algorithms add a leader to the optimal set one at a time, which typically leads to a suboptimal leader set because in general the optimal set of m leaders does not include the optimal set of $m-1$ leaders. In [59] convex optimization was used to quantify bounds on performance and an efficient greedy approach was proposed. This algorithm uses a swap procedure to reduce the error associated with choosing one leader at a time. In [17, 16] the total system error was proved to be a supermodular function (see 2.6) of the leader set, and this allowed for the development of algorithms that approximate the optimal solution up to a provable bound.

While researchers have added complexity to the leader selection for robustness problem by studying convergence rate to the state of the leader(s) [93, 104], networks with time delays [79], higher order dynamics [38, 49, 56, 94], uncertain dynamics [14, 105], or contradicting leader states [64], we note that the linear and most fundamental form of the problem has not been solved explicitly even for simple network structures [75].

To this end we develop a means of quantifying the combined influence of a set of leader nodes in a network on the total system error in the leader-follower dynamic. Intuitively, this influence should correspond to some notion of centrality of a set of nodes because a leader set that gives low system error must be well connected to other nodes in the network.

We note that the problem of leader selection for robustness is very closely tied to problems in other research areas such as pinning control [94], task allocation in robotic networks [34], and opinion dynamics with stubborn agents [35, 78, 108].

1.2 Centrality of Sets of Nodes

Analysis of how influential or central an agent is in a network can be used in a wide variety of contexts to better understand the roles of individual agents in multi-agent system dynamics. Examples are numerous and include social network analysis [30, 31, 36, 89], destabilization of covert networks [11, 12], biological applications such as understanding metabolic networks [63] or lethality in protein networks [45], and analysis of self-organized vs. planned urban streets [19].

Understanding how a set of agents jointly influences the network requires a notion of centrality different from that of a single agent. Various notions of centrality of a set of agents were defined in [23], where the authors quantified degree, closeness, betweenness and flow centralities of sets of nodes by extensions of the definitions for individuals. Illustrative examples were used in [13, 23, 69] to explore the relationship to network properties and efficient computation of group betweenness centrality was studied in [40, 53, 82, 83]. Alternative group centrality measures have been derived in [9, 10, 54]. While these measure provide heuristics that are intuitively reasonable, it is difficult to ascertain whether or not the measures are rigorously representative of the centrality of a set of nodes. In contrast to the existing literature, we derive a measure of centrality of a set of nodes, called *joint centrality*, by examining the performance measure, i.e., total system error, and expressing performance in terms of graph measures.

Applications of measures of centrality of a set of nodes include a broad range of research areas, encompassing emergency response management [48] and a net-

work connectivity analysis of the quality of innovative ideas [8]. Furthermore, many combinatorial problems, such as resource allocation for network security [87] and deployment strategies for mobile ad hoc networks [50] are strongly dependent on the joint influence of a set of nodes on the rest of the network.

1.3 Leader Selection for Controllability

Broadly speaking, there are two common variations on the problem of *leader selection for controllability*. When can a network be controlled by a given number of leaders [42, 43, 61, 65, 84, 99, 114]? And where in the network should leaders be located to optimize controllability metrics [27, 66, 73, 98]? These two variations are complementary as it is important to understand whether it is feasible to reach any final state and also how difficult it is to control the system to any such final state.

The concept of structural controllability is often applied when determining if a network can be controlled by a set of leaders. If a system is structurally controllable then there exists a set of weightings on communications between agents for which the network is controllable [57]. A common assumption is that leadership is costly, i.e. requires additional hardware or energy, and therefore one objective in the literature is to determine the minimum number of agents to behave as leaders. In [61] the authors concluded that the number of leaders, defined as nodes providing control input, required for structural controllability depends largely on the degree distribution of the network. An algorithm for finding the minimum number of leaders while minimizing cost was proposed in [77].

The authors of [99], and later the authors of [84], examined the graph theoretic characteristics required for a network to be controlled by a single leader. Means for extending these characteristics to the multiple leader setting were introduced in [84] using equitable partitions, which were also studied in [65]. In [46], the authors studied

controllability as related to connectedness between leader and follower subgraphs; topological implications of the results from [46] were examined in [47]. A common goal in many areas of multi-agent systems research is to develop distributed algorithms and protocols. To this end, [76] proposed a distributed algorithm to determine the number and locations of leaders to meet specifications related to structural controllability. Leader selection in networks subject to agent or communication failures was studied in [43].

In [18] it was argued that the more important question was whether or not a system is almost uncontrollable. Arbitrarily large graphs were considered in [22], and the fraction and locations of leaders needed for controllability were found for some canonical network topologies.

A related approach is to select leaders to maximize standard measures of controllability, which are defined in terms of the controllability Gramian. In [73] a strategy was developed for selecting leader nodes with performance guarantees relative to the smallest eigenvalues of the controllability Gramian, a measure inversely related to worst case input energy. Additionally, the authors of [101] developed an algorithm for actuator placement in sensor networks that meets a bound on control energy while ensuring controllability. The largest and smallest eigenvalues of the controllability Gramian were used in [107], where the relationship between energy cost bound and control time was demonstrated.

The trace of the controllability Gramian can be applied as a measure of average controllability as in [66]. In [98] the authors established that the trace of the controllability Gramian is modular while two other energy related control measures are submodular: trace of the inverse Gramian and log product of non-zero eigenvalues of the controllability Gramian, a measure of reachable volume (see 2.6 for definitions of modular and submodular). Control energy centralities were defined for each node in a network based on these measures of controllability.

1.4 Contributions and Thesis Outline

The work in this dissertation was motivated by the desire to rigorously connect group level performance of leader-follower networks with properties of the underlying network. To this end, throughout this work we investigate the leader selection problems for robustness (1.1) and controllability (1.3) in an attempt to rigorously show as a function of network graph measures where in the network optimal leaders should be located to optimize performance. We begin in Chapter 2 by providing background material and reviewing relevant mathematical topics that will be used throughout this thesis.

Chapter 3 focuses on the leader selection problem for robustness of agents' states to external noise. We provide a new approach to solving the optimal leader selection problem in terms of network graph measures. In general, our approach reduces computational complexity significantly as compared to the brute force computation. Further, our results provide structural insights into the problem and foundations for the development of on-line strategies. Significant to our approach, we define a new notion of centrality of a set of nodes in an undirected, connected graph, that we call *joint centrality*. For the leader-follower network dynamics, we show that the total system error is inversely proportional to the joint centrality of the leader set when the leaders are noise-free. Thus, the solution to the optimal leader selection problem is the set of nodes that maximizes joint centrality. We show that joint centrality of a set of nodes is a generalization of *information centrality* for a single node, and that the optimal leader set is composed of nodes that trade off high nodal information centrality with good coverage of the graph, i.e., a well distributed set with respect to resistance and biharmonic distances among nodes in the set.

In Chapter 4 we investigate special cases of graphs, namely path graphs, tree graphs, and cycle graphs. We demonstrate how the computation of joint centrality simplifies in these cases and provide lower and upper bounds on pairwise joint cen-

trality in unweighted trees. Furthermore, we solve explicitly for optimal solutions in unweighted path and unweighted cycle graphs. Chapter 5 demonstrates the use of joint centrality as a centrality measure of a set of nodes with the use of three real world examples: synthetic lethality in *Saccharomyces cerevisiae* (baker's yeast), clustering of a facebook social network and clustering of political books sold on Amazon.

We then proceed to study leader selection for controllability in Chapter 6. We prove the dependence of average controllability on information centrality of the leader nodes. Additionally, we prove the dependence of the reachable volume on the leader nodes' entries in the eigenvectors of the graph Laplacian. We discuss the tension between optimal sets for controllability and for robustness of the network dynamics to noise, and then conclude with a summary of results and interesting open questions in Chapter 7.

Chapter 2

Background

In this chapter we establish notation and review mathematical tools that will be used throughout this thesis. In particular we present a brief introduction to graph theory, related linear algebra, and key results from control theory. We also review results that are relevant to the conclusions drawn in this work. Although we cannot provide a thoroughly comprehensive review of all topics, the referenced citations contain additional details on the expansive work in these areas.

2.1 Basic Notation and Definitions

Throughout this dissertation we will use the symbol $\mathbb{R}$ to denote the set of all real numbers, $\mathbb{C}$ to denote the set of all complex numbers, and $\mathbb{Z}$ to denote the set of all integers. Lower case Roman or Greek letters will be used to denote scalar quantities. The exceptions to this include the capital letter K_f, which represents the scalar Kirchhoff index of a graph, and the capital letter W, which we will use to denote a scalar Wiener process.

Vectors will be represented by bold lower case Roman or Greek letters, with the exception of $\mathbf{W}$, which denotes a vector-valued Wiener process. We denote scalar entries in a vector using the same, non-bold, letter with a subscript denoting the

position of the entry. For example, the i^{th} entry of vector $\mathbf{x}$ is x_i. The j^{th} standard basis vector of $\mathbb{R}^n$ will be written as $\mathbf{e}_n^{(j)}$; a vector with 1 in the j^{th} position and 0 in all other positions. The notation $\mathbf{1}_n$ denotes a vector in $\mathbb{R}^n$ with 1 in every entry.

Calligraphic capital letters will denote mathematical objects such as graphs or sets, with the exception of S and F, which will be used to represent a set of leader agents and follower agents, respectively, and $\hat{a}, \hat{b}$, and $\hat{y}$, which will be used as sets of nodes in a partitioned network in Chapter 4. We will represent matrices using capital Roman or Greek letters, with the exception of the aforementioned examples: K_f, W, S, and F. When a set S of m nodes is identified, we will assume they are the first m nodes in an ordering of the complete set of n nodes unless otherwise specified. The entry of a matrix M in the i^{th} row and j^{th} column will be denoted by $M_{i,j}$. We will denote the partition of an $n \times n$ matrix B as

$$B = \begin{bmatrix} B_S & B_{SF} \\ B_{FS} & B_F \end{bmatrix}, \tag{2.1.1}$$

where B_S is an $m \times m$ matrix corresponding to nodes in set S, and B_F is an $(n - m) \times (n - m)$ matrix corresponding to the remaining nodes. We will further let l_1 be the first node in the ordered set S, and we will use this when there is no loss of generality. The identity matrix in $\mathbb{R}^{n \times n}$ will be represented as I_n.

The i^{th} eigenvalue of a square matrix B will be written as λ_i, where λ_i are arranged in ascending order. The association of λ_i with B will typically be clear in context, but in cases where this is not clear we will write $\lambda_i(B)$.

We will use the exponent T to denote the transpose of a matrix and the exponent $*$ to denote the Hermitian transpose of a matrix. The inverse of a matrix B will be written as B^{-1}, the determinant of B as $\det(B)$, and the adjugate matrix of B as $\text{adj}(A)$, where $B^{-1} = \frac{\text{adj}(A)}{\det(A)}$. The inverse of a matrix can also be computed blockwise

via the following relationship [6]

$$B^{-1} = \begin{bmatrix} B_S^{-1} + B_S^{-1}B_{SF}(B_F - B_{FS}B_S^{-1}B_{SF})^{-1}B_3B_S^{-1} & -B_S^{-1}B_{SF}(B_F - B_{FS}B_S^{-1}B_{SF})^{-1} \\ -(B_F - B_{FS}B_S^{-1}B_{SF})^{-1}B_{FS}B_S^{-1} & (B_F - B_{FS}B_S^{-1}B_{SF})^{-1} \end{bmatrix}, \tag{2.1.2}$$

assuming B_S is invertible.

The Moore-Penrose Pseudoinverse of a rank deficient matrix C will be written as C^+, where $CC^+C = C^+$ and $C^+CC^+ = C$ [5]. The trace of a square matrix B is equivalent to the sum of diagonal entries and represented by $\mathrm{tr}(B)$. We will make use of the property that the trace of the product of square matrices of equivalent size is invariant under cyclic permutations. That is, $\mathrm{tr}(XYZ) = \mathrm{tr}(YZX) = \mathrm{tr}(ZXY)$. This property also holds for rectangular matrices of compatible dimension. For example, for $Z \in \mathbb{R}^{n\times m}$, $\mathrm{tr}(Z'Z) = \mathrm{tr}(ZZ')$.

A *normal* matrix $B \in \mathbb{R}^{n\times n}$ is a matrix that commutes with its transpose, i.e. $BB^T = B^TB$. A square matrix $U \in \mathbb{C}^{n\times n}$ is a *unitary* matrix if its inverse is equivalent to its Hermitian transpose, that is $U^*U = UU^* = I_n$.

A *tridiagonal* matrix is a matrix with entries along the main diagonal, and the first diagonals above and below the main, with zeros elsewhere. The elements of the inverse of a general tridiagonal matrix can be expressed through a recurrence relation [102]. Certain tridiagonal matrices allow for explicit expressions of the elements of

the matrix inverse [28]. A tridiagonal matrix of the form

$$T_n = \begin{bmatrix} 1+\frac{a}{a+b} & -1 & & & \\ -1 & 2 & -1 & & \\ & \ddots & \ddots & \ddots & \\ & & -1 & 2 & -1 \\ & & & -1 & 1 \end{bmatrix},$$

has a symmetric inverse such that

$$(T_n^{-1})_{i,j} = \frac{ai-b}{a} \quad i \leq j. \tag{2.1.3}$$

We can write the pseudoinverse of a real-valued, rank-deficient matrix, $B \in \mathbb{R}^{n\times n}$ plus a rank-one matrix in terms of the pseudoinverse of B plus a real valued matrix H. Formally, for $y, z \in \mathcal{R}^n$,

$$(B + \mathbf{z}\mathbf{y}^T)^+ = B^+ + H \tag{2.1.4}$$

where

$$H = -\frac{1}{\|\mathbf{w}\|^2}\mathbf{v}\mathbf{w}^T - \frac{1}{\|\mathbf{m}\|^2}\mathbf{m}\mathbf{h}^T + \frac{\beta}{\|\mathbf{m}\|^2\|\mathbf{w}\|^2}\mathbf{m}\mathbf{w}^T \tag{2.1.5}$$

and $\beta = 1+\mathbf{y}^T B^+\mathbf{z}$, $\mathbf{v} = B^+\mathbf{z}$, $\mathbf{h} = (B^+)^T\mathbf{y}$, $\mathbf{w} = (I-BB^+)\mathbf{z}$, and $\mathbf{m} = (I-B^+B)^T\mathbf{y}$ [5] . Additionally, we will make use of the *Matrix Inversion Lemma*, which states that given $X \in \mathbb{R}^{n\times n}$, $Z \in \mathbb{R}^{m\times m}$, $U \in \mathbb{R}^{n\times m}$ and $V \in \mathbb{R}^{m\times n}$ such that X, Z and $X+UZV$ are nonsingular, then, $(X+UZV)^{-1}$ can be written as [106]

$$(X+UVZ)^{-1} = X^{-1} - X^{-1}U(Z^{-1} + VX^{-1}U)^{-1}VX^{-1}. \tag{2.1.6}$$

Another important matrix property that we will make use of is the determinant of a matrix with an added row and column, otherwise called a bordered matrix. The determinant of a bordered matrix can be computed as follows,

$$\begin{vmatrix} X & \mathbf{u} \\ \mathbf{v}^T & d \end{vmatrix} = d|X| - \mathbf{v}^T(\text{adj}(X))\mathbf{u}, \tag{2.1.7}$$

where $X \in \mathbb{R}^{p\times p}$, $\mathbf{u}, \mathbf{v} \in \mathbb{R}^p$, and $d \in \mathbb{R}$ [32] .

2.2 Linear Consensus Dynamics

The objective of a multi-agent system performing consensus dynamics is for the state of every agent in the network to reach agreement, i.e. converges on the same value. Formally, we represent this by considering n agents with states denoted by $\mathbf{x} = [x_1 \;\; x_2 \;\; ... \;\; x_n]^T$, where x_i is the state of agent i. Then we can say that a system is in consensus if $x_i = x_j \;\forall\; i, j \in \{1, 2, ..., n\}$. We note that the state of the system can represent a physical state, such as location or direction of travel, or the state can represent a virtual quantity, for example, a belief or parameter estimate. A fundamental assumption of consensus dynamics is that each agent can measure its state relative to the state of its neighbors. That is, if agent i is communicating with agent j, then i has access to the difference $x_j - x_i$. The use of relative measurements rather than absolute avoids the need to establish position in global coordinates, which can be difficult to measure. Furthermore, this ensures that any agent's bias in calculating absolute measurements (e.g., imperfectly calibrated sensors) is inconsequential.

There are many variations on dynamic protocols for achieving consensus, though here we will focus on the *linear consensus protocol* [70, 86, 90, 103]. In linear consensus dynamics, each agent updates its state according to a weighted average of relative

measurements of its neighbors. For each $i \in \{1, .., n\}$ we have

$$\dot{x}_i = \sum_{j \in \mathcal{N}_i} a_{i,j}(x_j - x_i), \tag{2.2.1}$$

where $\mathcal{N}_i$ is the set of neighbors of agent i and $a_{i,j}$ is a positive weight on the relative difference in states i and j. In matrix form, equation (2.2.1) can be expressed as

$$\dot{\mathbf{x}} = -L\mathbf{x} \tag{2.2.2}$$

where L is the Laplacian matrix generated by the underlying network topology. The construction and properties of L will be discussed in the following section.

2.3 Undirected Graph Theory

The study of graph theory provides a visual and mathematical framework for analysis of networked multi-agent systems. Agents are represented by *nodes* and communication between any two agents is given by a link, otherwise known as an *edge* between the two agents.

Formally, a *graph* $\mathcal{G} = (\mathcal{V}, \mathcal{E}, A)$, where each agent corresponds to a node in the set $\mathcal{V} = \{1, 2, ..., n\}$. We will use the terms agents and nodes interchangeably. $\mathcal{E} \subseteq \mathcal{V} \times \mathcal{V}$ is the set of edges, where the edge $(i, j) \in \mathcal{E}$ if j is a neighbor of i ($j \in \mathcal{N}_i$). $A \in \mathcal{R}^{n \times n}$ denotes the adjacency matrix. If $(i, j) \in \mathcal{E}$ then element $a_{i,j}$ will be positive, otherwise $a_{i,j} = 0$. In the linear consensus dynamic the magnitude of $a_{i,j}$ is the weighting that node i puts on the relative value of information from agent j. If for every pair $(i, j) \in \mathcal{E}$ we have that $(j, i) \in \mathcal{E}$ and $a_{i,j} = a_{j,i}$ $\forall i, j, \in \{1, 2, ..., n\}$, then graph $\mathcal{G}$ is *undirected*. It follows that the adjacency matrix for an undirected graph is a symmetric matrix. For the purpose of analytic tractability, in this dissertation we limit our analysis to undirected graphs. Recent results in closely related areas of directed graph theory

[112], [113], may provide a bridge towards extending this work to directed graphs.

A visual representation of a weighted, undirected graph is shown in Figure 2.1. Nodes are given in the numbered circles, lines between nodes are edges and the weight of each edge is written adjacent to the edge. In general, if edge weights are not specified then all weights are assumed to be 1.

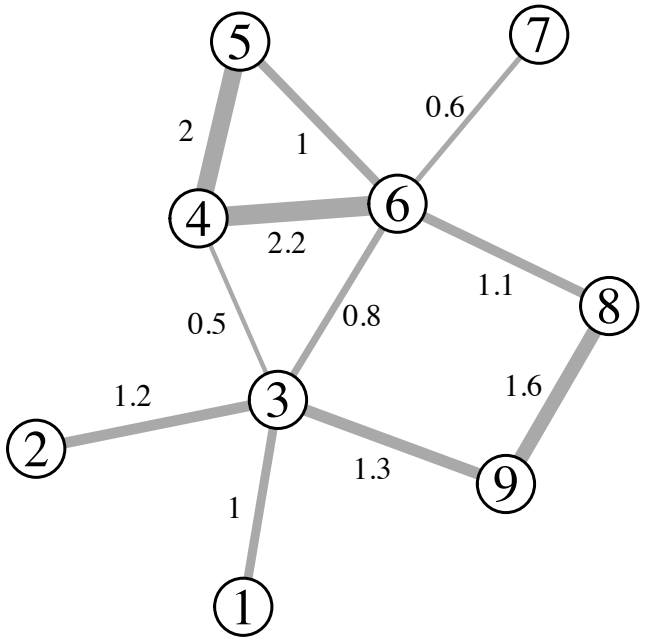

Figure 2.1: Example undirected, weighted graph with 9 nodes and 11 edges

We define the *degree* of node i, denoted by d_i, to be the sum of the weights of all edges incident to i, $d_i = \sum_{j=1}^{n} a_{i,j}$. The degree matrix associated with $\mathcal{G}$ is the diagonal matrix of node degrees, $D = \text{diag}([d_1, d_2, ..., d_n]^T) = \text{diag}(A\mathbf{1}_n)$. The *Laplacian* matrix, L, of $\mathcal{G}$ is defined as $L = D - A$. The Laplacian matrix of an undirected graph has a number of properties that will be utilized in future sections. First, the Laplacian is symmetric, and row and column sums of L are always equal to zero. The vector $\mathbf{1}_n$ lies in the null space of L, thus

$$L\mathbf{1}_n = \mathbf{0}, \tag{2.3.1}$$

and 0 is always an eigenvalue of L. All other eigenvalues will be real and non-negative, assuming the graph is connected. Finally, the following property of L^+ will be applied in proofs (see [81] for details):

$$LL^+ = L^+L = I_n - \frac{1}{n}\mathbf{1_n}\mathbf{1_n}^T. \tag{2.3.2}$$

A *path* is a sequence of nodes for which there is an edge between each consecutive pair of nodes. The number of edges between consecutive pairs of nodes is the *length* of the path. A *simple path* is a path which has no repeated nodes. A *cycle* is a path for which the initial node is equivalent to the final node, otherwise referred to as a closed path. A cycle which is a simple closed path is called a *simple cycle*. If every pair of nodes in an undirected network has a path between them then the graph $\mathcal{G}$ is *connected*. In this dissertation we will only make use of simple paths and simple cycles, therefore the use of the terms path or cycle will always indicate no repeated nodes. We will use the notation $p^{(ab)}$ to denote a path between nodes a and b, and let $|p^{(ab)}|$ be the length of the path, defined as the number of edges.

We now review relevant special cases of graphs. A *cycle graph* is a graph with n nodes and n edges where every node in the graph is contained by a simple cycle of length n. A *path graph* is a graph with n nodes and $n-1$ edges where every node in the graph is contained by a simple path of length $n-1$. We note that a path graph is equal to a cycle graph with one edge removed. A connected graph with n nodes that does not contain any cycles is a *tree*. A tree graph will always contain exactly $n-1$ edges. Any node in a tree with a degree of one is referred to as a *leaf*. Tree graphs, path graphs and cycle graphs are all examples of *planar graphs*, which implies that they can be drawn in a 2-D plane without edge crossings as shown in Figure 2.2 [2].

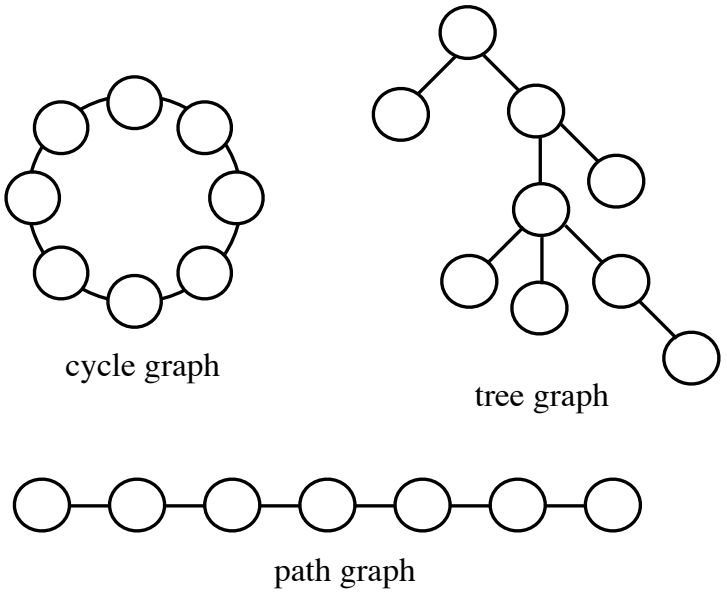

Figure 2.2: Examples of three canonical graph topologies: cycle graph (upper left), tree graph (upper right), and path graph (bottom).

2.3.1 Distance and Centrality

There are a variety of measures to characterize how central any given node in a graph is [30]. Examples include degree centrality [92], betweenness centrality [29], closeness centrality [4], eigenvector centrality [55] and information centrality [97]. Of those, only eigenvector centrality and information centrality take into account all paths between any pair of nodes in the network. Betweenness centrality and closeness centrality are both calculated using shortest paths, while degree centrality depends on the degree of each node.

We will show later (Chapter 3) that information centrality in particular is an important measure in the leader selection problem. Information centrality can be understood by first defining the information in a path between any two nodes in $\mathcal{G}$ to be the inverse of the sum of edge weights between those two nodes. Thus for an unweighted path, the longer the path the less information in that path. Total information between nodes i and j, denoted $I_{i,j}^{tot}$, is the sum of the information in all paths connecting nodes i and j. It was shown in [97] that total information can

be calculated without path enumeration by using the pseudoinverse of the Laplacian, L^+:

$$I_{i,j}^{\text{tot}} = \frac{1}{L_{i,i}^+ + L_{j,j}^+ - 2L_{i,j}^+}, \tag{2.3.3}$$

where $L_{i,j}^+$ is the i, j component of L^+. Information centrality for node i, denoted c_i, is defined as the harmonic average of total information between node i and all other nodes in $\mathcal{G}$ [97]:

$$c_i = \left(\frac{1}{n} \sum_{j=1}^{n} \frac{1}{I_{i,j}^{\text{tot}}} \right)^{-1}. \tag{2.3.4}$$

The Kirchhoff index of a graph is the sum of resistance distances between all node pairs and has been applied as a measure of robustness of a network [109, 110]. The diagonal entries of the psuedoinverse of the Laplacian are related to the Kirchhoff index through the relation [52]

$$\text{tr}(L^+) = \frac{K_f}{n}. \tag{2.3.5}$$

In [81], Poulakakis et al. evaluated the certainty of each node i in a network of decision-makers accumulating stochastic evidence towards a decision* . This certainty, denoted μ_i, is defined as the inverse of the difference between the variance of the state x_i about the reference signal and the minimum achievable variance as $t \to \infty$. The authors apply the notion of information centrality to directly interpret μ_i in terms of structural properties of the underlying communication graph. It was proven that

$$\frac{1}{\mu_i} = \frac{\sigma^2}{2} L_{i,i}^+ = \frac{\sigma^2}{2} \left(\frac{1}{c_i} - \frac{K_f}{n^2} \right). \tag{2.3.6}$$

*This paragraph through the end of Section 2.3.1 has been adapted from Fitch and Leonard [26] with some text taken verbatim.

The identity (2.3.6) implies that the ordering of nodes by certainty is equal to the ordering of nodes by information centrality.

The total information between any two nodes i and j is closely related to the resistance distance between them, denoted $r_{i,j}$. Resistance distance between nodes in the undirected graph $\mathcal{G}$ is defined as the resistance distance between the corresponding two nodes in the electrical network analog to the graph $\mathcal{G}$. By [52] for an undirected graph $\mathcal{G}$

$$r_{i,j} = L^+_{i,i} + L^+_{j,j} - 2L^+_{i,j} = {I^{\text{tot}}_{i,j}}^{-1}. \tag{2.3.7}$$

It follows that

$$\sum_{i=1}^{n} r_{i,j} = \frac{n}{c_j}. \tag{2.3.8}$$

Since there is only one path between any two nodes and all edge weights are equal to one in an unweighted tree graph, resistance distance in this case is equivalent to geodesic distance. The notation $\delta_{i,j}$ will be used to denote the resistance distance between nodes i, j in an unweighted tree. The constraint of one path between any two nodes allows us to write the information centrality of a node i for an unweighted tree graph in terms of the sum of geodesic distances from i to every other node in the network:

$$\frac{1}{c_i} = \frac{1}{n}\sum_{j=1}^{n} \delta_{ij}, \tag{2.3.9}$$

where we note that in this case $I^{i,j}_{\text{tot}} = \frac{1}{\delta_{ij}}$. This form can be simplified even further when considering paths graphs. In general for path graphs we will use the convention that node a is the node in the a^{th} position along the path, ordered from left to right.

Applying equation (2.3.9) to a path graph yields for node i,

$$\frac{1}{c_i} = \frac{2i^2 + n + n^2 - 2i(1+n)}{2n} \tag{2.3.10}$$

An additional measure with similar form to that of resistance distance is the recently derived notion of biharmonic distance, d_B [60]. This measure has been used to quantify distance between two points v_i, v_j on the surface of a discrete 3D mesh:

$$d_B(v_i, v_j)^2 = g_d(i,i) + g_d(j,j) - 2g_d(i,j), \tag{2.3.11}$$

where g_d is the discrete Green's function [15, 60] of the discretized, bilaplacian $\tilde{L}^2$, equivalent to the pseudoinverse of $\tilde{L}^2$. $\tilde{L}$ is the discretized Laplacian, obtained by the cotangent formula for discretization of the Laplacian on meshes, and normalized by the mesh area at each vertex (see [60] and references therein for details). In the context of 3D meshes, the biharmonic distance has the advantage, over diffusion and geodesic distances, of providing a balance between local and global properties of a surface, reflecting overall connectivity for faraway points [60]. We define the *biharmonic distance between two nodes i and j in the graph $\mathcal{G}$*, which we denote $\gamma_{i,j}$, analogously for the unnormalized Laplacian L as defined in the beginning of Section 2.3:

$$\begin{aligned} \gamma_{i,j} &= L^{2+}_{i,i} + L^{2+}_{j,j} - 2L^{2+}_{i,j} = \sum_{l=1}^{n} (L^{+}_{l,i} - L^{+}_{l,j})^2 \\ &= (\mathbf{e}_i - \mathbf{e}_j)^T L^{2+} (\mathbf{e}_i - \mathbf{e}_j). \end{aligned} \tag{2.3.12}$$

We observe that the definition of biharmonic distance $\gamma_{i,j}$ of (2.3.12) is very similar to the definition of resistance distance $r_{i,j}$ of (2.3.7) with the difference being the use of the pseudoinverse of L^2 in the definition of $\gamma_{i,j}$ as compared to the pseudoinverse of L in the definition of $r_{i,j}$. Since L^2 is symmetric and positive semi-definite, we

immediately have that $\gamma^{1/2}$ is a metric. In fact, it can be viewed as a Manahalobis distance, which in this case describes a dissimilarity measure between two vectors from a single distribution with covariance matrix L^2. Let Γ be the matrix with elements $\gamma_{i,j}$.

For completion, we note that both resistance distance and biharmonic distance between nodes can be written in terms of the eigenvalues λ_i and eigenvectors $\boldsymbol{\nu}_i$ of the Laplacian L [52, 26]:

$$r_{i,j} = \sum_{l=2}^{n} \frac{1}{\lambda_l}(\nu_l^i - \nu_l^j)^2, \tag{2.3.13}$$

$$\gamma_{i,j} = \sum_{l=2}^{n} \frac{1}{\lambda_l^2}(\nu_l^i - \nu_l^j)^2. \tag{2.3.14}$$

2.4 Controllability

Given a dynamical system, we often wish to rigorously understand how the behavior of that system will be modified by the presence of inputs to the system. A linear system with n states subject to m inputs, $\mathbf{u}$, can be written as

$$\dot{\mathbf{x}} = A\mathbf{x} + B\mathbf{u}, \tag{2.4.1}$$

where $A \in \mathbb{R}^{n\times n}$, $\mathbf{x} \in \mathbb{R}^{n\times 1}$, $B \in \mathbb{R}^{n\times m}$ and $\mathbf{u} \in \mathbb{R}^{m\times 1}$.

One relevant question we can ask is whether or not it is possible to design control inputs which can drive the system to a desired state. This brings us to the notion of *controllability*:

Definition 1. *[88] Controllability: The linear, time invariant system (2.4.1) is* controllable *if, for every* $\mathbf{x}_f$ *and every* $T > 0$*, there exists an input function* $\mathbf{u}(t)$*, for* $0 < t < T$*, such that the system is taken from* $\mathbf{x}(t = 0) = \mathbf{0}^{n\times 1}$ *to* $\mathbf{x}(t = T) = \mathbf{x}_f$*.*

In our discussion we are primarily concerned with steady state behavior. There-

fore we will consider *infinite horizon controllability*, that is, controllability as defined in Definition 1 with $T = \infty$. A key concept in the study of controllability is the *controllability Gramian*. The controllability Gramian associated with the dynamics (2.4.1) is defined, for stable A, as [39]

$$W_C = \int_0^\infty e^{A\tau} B B^T e^{A^T \tau} d\tau, \tag{2.4.2}$$

which is also the solution to the Lyapunov equation

$$AW_C + W_C A^T = -BB^T. \tag{2.4.3}$$

It can be shown that the following statements are equivalent [39].

1. System (2.4.1) is controllable.
2. The controllability Gramian (2.4.2) is nonsingular.
3. The $n \times n$ controllability matrix $\mathcal{C} = [B\ AB\ A^2B\ \ ...\ A^{n-1}B]$ has rank n.

In the case where W_C is not full rank we refer to the controllable subspace of the system (2.4.1) as the range of the controllability Gramian. While controllability is an important characteristic for a linear time invariant system, it is also necessary to consider how controllable a system is. This is because it is possible for a system to be controllable but take infinite time to reach the final state or require a large amount of control effort or energy. [†] The following four functions of the controllability Gramian provide four measures of controllability performance.

(a) **Average controllability**: $\mathrm{tr}(W_C)$ provides a measure of average controllability over the controllable subspace.

[†]The following text through the end of Section 2.4 has been adapted from Fitch and Leonard [27] with some text taken verbatim.

(b) **Reachable volume**: $\mathrm{ld}(W_C) = \log\left(\prod_{j=1}^{\mathrm{rank}W_C} \lambda_j(W_C)\right)$ provides a measure of the volume of the controllable subspace reachable with one unit of input. When W_C is full rank $\mathrm{ld}(W_C)$ is equal to the log determinant of W_C.

(c) **Average control energy**: $\mathrm{tr}({W_C}^{-1})$ and $\mathrm{tr}({W_C}^{+})$ provide measures of average control energy required to reach a random state in the controllable subspace.

(d) **Worst case input energy**: $\lambda_{min}(W_C)$ is inversely proportional to the input energy required to move in the least controllable direction in the controllable subspace.

We let W_{C_i} be the controllability Gramian associated with one leader node, i. W_{C_i} satisfies (2.4.3) when $B = \mathbf{e}_n^{(i)}$. In [98], the authors defined three *control energy centralities* for each node i in a network based on the value of controllability measures (a)-(c) when i is selected as a single leader node. These control energy centralities are

- *Average controllability centrality*

$$C_{AC}(i) = \mathrm{tr}(W_{c_i}) \quad i \in V$$

- *Average control energy centrality*

$$C_{ACE}(i) = -\mathrm{tr}(W_{c_i}^{+}) \quad i \in V$$

- *Volumetric control energy centrality*

$$C_{VCE}(i) = \log\left(\prod_{j=1}^{\mathrm{rank}W_{c_i}} \lambda_j(W_{c_i})\right) \quad i \in V.$$

2.5 Stochastic Differential Equations

Often it is not possibly to perfectly model the dynamics of physical systems due to the presence of *noise*, which could arise from random disturbances, imperfect measurements, or unpredictable phenomena that cannot be captured by our dynamical equations. Though we cannot precisely determine a priori the magnitude and means by which noise will effect our system at any given instant, we would still like our dynamic model to account for stochasticity. A dynamic model with noise can be represented using an Ito stochastic differential equation [33], the scalar version of which can be written as

$$dx = f(x,t)dt + g(x,t)dW, \tag{2.5.1}$$

where x is the state variable, f and g are each functions of both x and t and dW is the standard Wiener increment. Before discussing the precise meaning of dW, we first introduce Markov processes and Wiener processes. A Markov process is a process for which the future state of the process depends only on the current state of the process and not past states [72]. In this way, a Markov process is referred to as memory-less: the past history of the process provides no additional information about the future states. A Wiener process, $W(t)$, is a continuous Markov process with a zero-mean Gaussian probability density function that has a variance that grows linearly with time [1, 33]. The Wiener process is a function of time and we will often drop the t in $W(t)$ and write W for notational simplicity. Increments of W are stationary, meaning that the distribution of $W(t_2) - W(t_1)$ does not depend on the particular values of t_1 or t_2, but rather the difference $t_2 - t_1$.

The Wiener process is closely related to the concept of white noise, which is a scalar, stationary, zero-mean Gaussian process $\xi(t)$, with a constant power spectral

density. If fact, [1]

$$\int_0^t \xi(\tau)d\tau = W(t). \tag{2.5.2}$$

Therefore we can think about dW in equation (2.5.1) as a term representing the presence of white noise. Unfortunately, the Wiener process is nowhere differentiable, and we are not able to take the derivative of W in the normal way. The development of Ito calculus provides us a means with which to work with such functions. We make the definition

$$dW := \xi(t)dt,$$

and then we are able to write the integral of a white noise process as [33]

$$\int_0^t g(x,\tau)\xi(\tau)d\tau = \int_0^t g(x,\tau)dW$$

where

$$\int_0^t g(\tau)dW := \underset{n\to\infty}{\text{ms-lim}}\Big\{\sum_{i=1}^{n} g(x,\tau_{i=1})[W(\tau_i) - W(\tau_{i-1})]\Big\} \tag{2.5.3}$$

where ms-lim stands for mean square limit and $\tau_0 = t_1 < \tau_1 < \tau_2 < ... < \tau_n = t_2$. Thus (2.5.1) is a valid equation and its integral form is [33]

$$x(t) - x(0) = \int_0^t f(x,\tau)d\tau + \int_0^t g(x,\tau)dW$$

These concepts can be easily extended to higher dimensions by considering $\mathbf{W}$ to be an n-dimensional Wiener process in which each element is an independent scalar

Wiener process. The generalized version of (2.3.12) is

$$d\mathbf{x} = \mathbf{f}(\mathbf{x}, t)dt + G(\mathbf{x}, t)d\mathbf{W}, \tag{2.5.4}$$

where $\mathbf{x}$ is an n-dimensional vector, $\mathbf{f}$ is a vector-valued function of $\mathbf{x}$ and t, and G is a matrix-valued function of $\mathbf{x}$ and t.

Any process corrupted by noise will result in a different time series for each run of the process. For this reason, it is more informative to determine statistical properties of the solution rather than solve for sample solutions $\mathbf{x}(t)$. To do this we study how the probability density function of $\mathbf{x}(t)$, $p(\mathbf{x}, t)$, evolves with time. If (2.5.4) can be written as [1]

$$d\mathbf{x} = [A(t)\mathbf{x} + \mathbf{a}(t)]dt + B(t)d\mathbf{W}, \tag{2.5.5}$$

and initial condition $\mathbf{x}(t_0)$ is constant or normally distributed, then $\mathbf{x}(t)$ is a Gaussian process, and $p(\mathbf{x}, t)$ can be completely described by its mean and covariance [1]. Letting $\mathbf{E}[\cdot]$ denote the expected value operator, we define $\mu_{\mathbf{x}}(t) := \mathbf{E}[\mathbf{x}(t)]$ to be the mean of $\mathbf{x}$, and $\Sigma_{\mathbf{x}}(t) := \mathbf{E}\Big[(\mathbf{x}(t) - \mu_{\mathbf{x}}(t))(\mathbf{x}(t) - \mu_{\mathbf{x}}(t))^T\Big]$ to be the covariance of $\mathbf{x}$. It then holds that $\mu_{\mathbf{x}}(t)$ satisfies the equation [1]

$$\dot{\mu_{\mathbf{x}}}(t)) = A(t)\mu_{\mathbf{x}}(t)) + \mathbf{a}(t), \tag{2.5.6}$$

and the covariance $\Sigma_{\mathbf{x}}(t)$ satisfies

$$\dot{\Sigma_{\mathbf{x}}}(t) = A(t)\Sigma_{\mathbf{x}}(t) + \Sigma_{\mathbf{x}}(t)A(t)^T + B(t)B(t)^T. \tag{2.5.7}$$

Therefore, equations (2.5.6) and 2.5.7 completely describe the behavior of systems of the form (2.5.5) with constant or normally distributed initial conditions.

Assuming constant A and B matrices, we observe a parallel between the steady state covariance equation((2.5.7) with $\dot{\Sigma} = 0$ and the equation for the controllability Gramian (2.4.3). Indeed, one way to think of noise is as an input to our dynamical system.

2.6 Modularity and Submodularity

It is often difficult to generalize the effect of a change in the members of a set $\mathcal{S}$ on the outcome of a real valued set function $f(\mathcal{S})$. However, there are properties of $f(\mathcal{S})$, namely *modularity* and *submodularity*, that when established, give us valuable insight. Let $V = 1, ..., Z$ be a finite set and let $f : 2^V \to \mathbb{R}$ be a function from all subsets of V to real values. Then modularity and submodularity are defined as follows.

Definition 2. *[62] Modularity: A set function $f : 2^V \to \mathbb{R}$ is modular if and only if for any subset $\mathcal{S} \subseteq V$, f can be written as*

$$f(\mathcal{S}) = w(\emptyset) + \sum_{s \in \mathcal{S}} w(s) \tag{2.6.1}$$

where w is a weight function $w : V \to \mathbb{R}$.

The implication of a modular set function is that each element of a subset independently contributes to the value of the function. Moreover, $w(s) = f(\{s\})$ when $w(\emptyset) = 0$. Solving an optimization problem with a modular cost function is therefore straightforward, as the total cost is simply the sum of each element's independent contribution to the cost function.

Definition 3. *[62] Submodularity: A set function $f : 2^V \to \mathbb{R}$ is submodular if and*

only if for all subsets $\mathcal{A} \subseteq \mathcal{B} \subseteq V$ *and all elements* $s \notin \mathcal{B}$

$$f(\mathcal{A} \cup \{s\}) - f(\mathcal{A}) \geq f(\mathcal{B} \cup \{s\}) - f(\mathcal{B}). \tag{2.6.2}$$

A submodular set function has the property of diminishing returns, that is the addition of an element to a larger set has a smaller contribution than the addition of an element to a smaller set. Therefore, each element of a subset does not contribute independently and full solutions to optimization problems with nondecreasing submodular set functions are NP-hard. However, greedy algorithms can provide a solution within a provable bound from the optimal solution [68].

Definition 4. *[62, 16] Supermodularity: A set function* $f : 2^V \to \mathbb{R}$ *is supermodular if and only if for all subsets* $\mathcal{A} \subseteq \mathcal{B} \subseteq V$ *and all elements* $s \notin \mathcal{B}$

$$f(\mathcal{A}) - f(\mathcal{A} \cup \{s\}) \geq f(\mathcal{B}) - f(\mathcal{B} \cup \{s\}). \tag{2.6.3}$$

A supermodular set function has the property that adding an element s to a set $\mathcal{A}$ yields a larger decrease in f than adding s to a superset $\mathcal{B}$. A function, f, is supermodular if $-f$ is submodular.

Chapter 3

Leader Selection for Robustness and Joint Centrality*

In this chapter we study the leader selection problem for robustness. We consider a network of nodes tracking an external signal in a noisy environment, and seek to identify m leader nodes to directly measure the signal such that total system error is minimized. Total system error is defined as the sum of steady state variances of all nodes in the network and is equivalent to the $\mathcal{H}_2$-norm of the linear system (3.1.2), a measure of robustness of the leader-follower consensus dynamics to environmental noise. Therefore, high performing leader sets indicate that the system will be less influenced by the presence of noise in the state of each node. We find that total system error for m noise-free leaders can be written as proportional to the *joint centrality* of the leader set, where joint centrality is a measure associated with a set of nodes that takes into account the centrality of the nodes and coverage of the set over the network. Thus, high performing leader sets will be central, yet distributed over the graph. For one leader, we prove that the optimal leader is the most information central node.

*This chapter is adapted from Fitch and Leonard [26] with most of the text taken verbatim. Some material was first published in Fitch and Leonard [25].

3.1 Model and Problem Statement

We consider a network of n agents tasked with tracking an external signal from the environment. We denote the external signal by $\mu \in \mathbb{R}$ and suppose it to be a constant. Generalizations to vector-valued environmental signals are expected to be relatively straightforward and extensions to time-varying environmental signals are the topic of future work.

An agent $l \in \mathcal{V}$ is called a *leader* if it directly measures the external signal. Let $k_l > 0$ be the weight that agent l puts on its signal measurement. Any agent that is not a leader is called a *follower*. Let the set of leaders be denoted S with cardinality m and the set of follower nodes, denoted by F, be the complement of S with cardinality $n - m$. Summation over s denotes summation over the leader set, while summation over i denotes summation over the entire set of leaders and followers. We use the index l_1 when it is necessary to identify one leader apart from the rest of the leader set.

We assume that all leaders apply the same weight k to their measurement of the external signal, i.e., $k_i = k > 0$ for $i \in S$ and $k_i = 0$ for $i \in F$. We assume that stochastic disturbances enter the dynamics as additive noise. We model the dynamics for each agent $i \in \mathcal{V}$ by the following stochastic process:

$$dx_i = -k_i(x_i - \mu)dt - L_i\mathbf{x}dt + \sigma dW_i, \tag{3.1.1}$$

where L_i is the ith row of the Laplacian L, and σdW_i represents increments drawn from independent Wiener processes with standard deviation σ.

In the case that $k < \infty$, the dynamics of the leaders and followers are all noise corrupted. In [59], it was demonstrated that in the limit as $k \to \infty$, i.e., in the case that leaders apply an arbitrarily large weight to tracking the external signal, the dynamics (3.1.1) describe the case of noise-free leaders. Thus, our model (3.1.1)

describes both cases of noise-corrupted leaders ($k < \infty$) and noise-free leaders ($k \to \infty$).

To write (3.1.1) in vector form let $K \in \mathbb{R}^n$ be the diagonal matrix with elements k_i, let $M = L + K$ and without loss of generality let $\mu = 0$. Then (3.1.1) becomes

$$d\mathbf{x} = -M\mathbf{x}dt + \sigma d\mathbf{W}. \tag{3.1.2}$$

Since we have assumed that $\mathcal{G}$ is connected, $-M$ is Hurwitz so long as $k_i = k > 0$ for some agent i, i.e., S is nonempty.

Thus, for nonempty S, $\mathbf{x}$ will converge to a steady-state distribution about the value of the external signal, and the steady-state covariance matrix Σ of $\mathbf{x}$ is the solution to the Lyapunov equation

$$M\Sigma + \Sigma M^T = \sigma^2 I. \tag{3.1.3}$$

The steady-state variance of x_i is $\Sigma_{i,i}$, the corresponding diagonal element of Σ. Since the external signal is assumed to be constant, the system will converge to a steady-state distribution about the value of the external signal even if the nodes chosen as leaders do not guarantee system controllability.

Following [74, 16], we define *total system error* as $\mathrm{Tr}(\Sigma) = \sum_{i=1}^{n} \Sigma_{i,i}$. We define group performance as the inverse of total system error, which measures *network tracking accuracy*.

By [1] we have that the covariance matrix of (3.1.2) is

$$\mathrm{Cov}(\mathbf{x}(t), \mathbf{x}(t)) = \sigma^2 \int_0^t e^{-M(t-\tau)} e^{-M^T(t-\tau)} d\tau. \tag{3.1.4}$$

Given that $\mathcal{G}$ is undirected, the Laplacian matrix L will be symmetric and it follows that M will be symmetric and normal. Let the eigenvalues of M be λ_i, $i \in \mathcal{V}$ with

corresponding eigenvectors $\boldsymbol{\nu}_i$. Let Λ be the diagonal matrix with entries $\Lambda_{i,i} = \lambda_i$. Then there exists a unitary matrix U such that $U^*MU = \Lambda$ and (3.1.4) can be written as

$$\mathrm{Cov}(\mathbf{x}(t), \mathbf{x}(t)) = \sigma^2(UR(t)U^*), \tag{3.1.5}$$

with

$$R(t) := \int_0^t e^{-(\Lambda+\bar{\Lambda})(t-\tau)}d\tau. \tag{3.1.6}$$

From [80], this gives

$$[\mathrm{Cov}(\mathbf{x}(t), \mathbf{x}(t))]_{i,j} = \sigma^2 \sum_{p=1}^{n} \frac{1 - e^{-2\mathrm{Re}(\lambda_p)t}}{2\mathrm{Re}(\lambda_p)} \nu_i^{(p)} \bar{\nu}_j^{(p)}. \tag{3.1.7}$$

Since M is symmetric, all eigenvalues of M will be real, and the steady-state variance of each node can be written as

$$\mathrm{Var}(x_i)_{ss} = \Sigma_{i,i} = \sigma^2 \sum_{p=1}^{n} \frac{1}{2\lambda_p} |\nu_i^{(p)}|^2. \tag{3.1.8}$$

Total system error follows from summing (3.1.8) over all i,

$$\sum_{i=1}^{n} \Sigma_{i,i} = \sigma^2 \sum_{i=1}^{n} \frac{1}{2\lambda_i} = \frac{\sigma^2}{2} \sum_{i=1}^{n} M_{i,i}^{-1}. \tag{3.1.9}$$

Total system error defines the coherence of the network, and is equivalent to the H_2 norm of the system with output equation $\mathbf{y} = C\mathbf{x}$, where $C = I_n$ and I_n the $n \times n$ identity matrix [74, 109].

We define the *optimal leader selection problem* as follows.

Definition 5 (Optimal leader selection problem for robustness)**.** *Given m and undi-*

rected, connected graph $\mathcal{G}$, find a set of m leaders S^* over all possible sets S of m leaders that minimizes the total system error (3.1.9) for the leader-follower network tracking dynamics (3.1.2), i.e., find

$$S^* = \arg\min_S \sigma^2 \sum_{i=1}^{n} \frac{1}{2\lambda_i} = \arg\min_S \frac{\sigma^2}{2} \sum_{i=1}^{n} M_{i,i}^{-1}. \tag{3.1.10}$$

3.2 Joint centrality and the optimal m noise-free leaders

In this section, we prove our main result on the general solution of the optimal leader selection problem by deriving an explicit expression for total system error with m noise-free leaders in terms of properties of the underlying graph. Before stating the theorem, we first define the *joint centrality of a set of m nodes* in a network graph.

Definition 6 (Joint centrality)**.** *Let $\mathcal{G}$ be an undirected, connected graph of order n. Given integer $m < n$, let S be the set of any m nodes in $\mathcal{G}$. Choose an arbitrary element $l_1 \in S$. Let N be an $n \times n$ matrix with elements of N^{-1} given by*

$$N_{i,j}^{-1} = L_{i,j}^{+} - L_{i,l_1}^{+} - L_{j,l_1}^{+} + L_{l_1,l_1}^{+}. \tag{3.2.1}$$

Following (2.1.1), $N_{S\setminus l_1}^{-1}$ is the $(m-1)\times(m-1)$ submatrix of N^{-1} corresponding to the elements of S less the first element l_1. Let $G = \left(N_{S\setminus l_1}^{-1}\right)^{-1}$ and $\bar{G} = \begin{bmatrix} 0 & 0 \\ 0 & G \end{bmatrix} \in \mathbb{R}^{m\times m}$. Let $Q = \bar{G}\Gamma_S$, where Γ is given by (2.3.12). The joint centrality of set S in $\mathcal{G}$ *is defined as*

$$\rho_S = n\Big(\frac{K_f}{n} + n\det(G)\det(L_S^{+}) + \frac{1}{2}\mathrm{Tr}(Q) - \mathbf{1}_n^T Q \mathbf{e}_{l_1}\Big)^{-1}. \tag{3.2.2}$$

Theorem 1 (Optimal noise-free leader set)**.** *Let $\mathcal{G}$ be an undirected, connected graph*

of order n. Let S be a set of m noise-free leaders. Then, the total system error (3.1.9) for the system dynamics (3.1.2) is

$$\sum_{i=1}^{n} \Sigma_{i,i} = \frac{\sigma^2}{2}\Big(\frac{n}{\rho_S}\Big), \tag{3.2.3}$$

where ρ_S is the joint centrality of leader set S given by (3.2.2). The optimal leader set is $S^* = \arg\max_S \rho_S$, the set of leader nodes with the maximal joint centrality.

Proof. (Theorem 1). We begin by studying terms in the total system error for finite $k > 0$ and then evaluate in the limit as $k \to \infty$. From (3.1.9), the total system error is proportional to $\mathrm{Tr}(M^{-1})$ where $M = L + K$. Let K_1 be the diagonal matrix with k in the first diagonal element and zeros elsewhere and let $K_{m-1} = K - K_1$. We derive an expression for $\mathrm{Tr}(M^{-1})$ by calculating two successive updates to L^+. We first show that if we define $N = L + K_1$, and thus $M = N + K_{m-1}$, then N^{-1} satisfies (3.2.1) for $k \to \infty$.

Let $\mathbf{e} = \mathbf{d}$ be vectors of length n with $\sqrt{k}$ in the l_1 (first) entry and zeros elsewhere where l_1 is a member of the leader set. Note that the choice of l_1 will not affect the value of joint centrality for a given leader set. Then $N^{-1} = (L+K_1)^{-1} = (L+\mathbf{e}\mathbf{d}^T)^{-1}$. Applying Lemma 2.1.4 we get that $(L + \mathbf{e}\mathbf{d}^T)^{-1} = L^+ + H$, with H given by (2.1.5) such that

$$N^{-1} = L^+ - L^+_{l_1}\mathbf{1_n}^T - \mathbf{1_n}L^{+T}_{l_1} + \frac{(1 + kL^+_{l_1,l_1})}{k}\mathbf{1_n}^T\mathbf{1_n}. \tag{3.2.4}$$

Taking the limit as $k \to \infty$, the elements of N^{-1} can be written as (3.2.1).

Let $U = [-\sqrt{k}\mathbf{e}_2, \ldots, -\sqrt{k}\mathbf{e}_m] \in \mathbb{R}^{n\times(m-1)}$, let $V = U^T$ and let $\mathbb{I}_{m-1} \in \mathbb{R}^{(m-1)\times(m-1)}$ be the identity matrix. Then, $M^{-1} = (N + K_{m-1})^{-1} = (N + U\mathbb{I}V)^{-1}$. Applying the

Matrix Inversion Lemma 2.1.6 we get that

$$(N+U\mathbb{I}V)^{-1} = N^{-1} - N^{-1}U(\mathbb{I} + VN^{-1}U)^{-1}VN^{-1}. \tag{3.2.5}$$

Let $G = ({N^{-1}}_{S\setminus l_1})^{-1}$ as in Definition 6. Then if we take the limit as $k \to \infty$, sum the diagonal elements of $M^{-1} = (N + UIV)^{-1}$, and apply the identities (2.3.1) and (2.3.5) we get

$$\begin{aligned}\sum_{j=1}^{n} M^{-1}_{j,j} =& \frac{K_f}{n} + nL^+_{l_1,l_1} - \sum_{s_1,s_2\in S\setminus\{l_1\}} \sum_{i=1}^{n} G_{s_1,s_2}\Big(L^+_{l_1,l_1}(L^+_{l_1,l_1} - L^+_{l_1,s_1} - L^+_{l_1,s_2}) + \\ & L^+_{l_1,s_1}L^+_{l_1,s_2} + \frac{1}{2}\left[(L^+_{i,l_1} - L^+_{i,s_1})^2 + (L^+_{i,l_1} - L^+_{i,s_2})^2 - (L^+_{i,s_1} - L^+_{i,s_2})^2\right]\Big).\end{aligned} \tag{3.2.6}$$

Consider the square bracketed terms of (3.2.6) in which we observe the emergence of biharmonic distance, γ. Substituting (2.3.12) and defining $\bar{G}$ as in Definition 6 we get

$$\begin{aligned}&\sum_{s_1,s_2\in S\setminus\{l_1\}} \sum_{i=1}^{n} G_{s_1,s_2}\,\frac{1}{2}\left[(L^+_{i,l_1} - L^+_{i,s_1})^2 + (L^+_{i,l_1} - L^+_{i,s_2})^2 - (L^+_{i,s_1} - L^+_{i,s_2})^2\right]\Big) \\ &\qquad = -\frac{1}{2}\mathrm{Tr}(\bar{G}\Gamma_S) + \mathbf{1}_n^T[\bar{G}\Gamma_S]\mathbf{e}_{l_1}.\end{aligned} \tag{3.2.7}$$

Additional simplification is made by applying Lemma 2.1.7 to the middle terms on the right hand side of (3.2.6). We get

$$\begin{aligned}&nL^+_{l_1,l_1} - n\sum_{s_1,s_2\in S\setminus\{l_1\}} G_{s_1,s_2}\Big(L^+_{l_1,l_1}(L^+_{l_1,l_1} - L^+_{l_1,s_1} - L^+_{l_1,s_2}) + L^+_{l_1,s_1}L^+_{l_1,s_2}\Big) = \\ &\frac{n}{\det(G^{-1})}\Big(L^+_{l_1,l_1}\det(G^{-1}) - \\ &\sum_{s_1,s_2\in S\setminus\{l_1\}} C_{{N^{-1}}_{s_1,s_2}}\left[L^+_{l_1,l_1}(L^+_{l_1,l_1} - L^+_{l_1,s_1} - L^+_{l_1,s_2}) + L^+_{l_1,s_1}L^+_{l_1,s_2}\right]\Big)\end{aligned} \tag{3.2.8}$$

where $C_{N^{-1}}$ is the cofactor matrix of ${N^{-1}}_{S\setminus l_1} = G^{-1}$. We then let $\mathbf{L}^+_{l_1,s_i} = [L^+_{l_1,s_1}, ..., L^+_{l_1,s_{m-1}}]^T$ and $\mathbf{L}^+_{l_1,l_1} = [L^+_{l_1,l_1}, ..., L^+_{l_1,l_1}]^T$ to be vectors in $\mathbb{R}^{m-1}$ and apply Lemma 2.1.7 to rewrite the expression in (3.2.8) as

$$\frac{n}{\det({N^{-1}}_{S\setminus l_1})} \begin{vmatrix} {N^{-1}}_{S\setminus l_1} & \mathbf{L}^+_{l_1,l_1} - \mathbf{L}^+_{l_1,s_i} \\ \mathbf{L}^+_{l_1,l_1} - \mathbf{L}^+_{l_1,s_i} & L^+_{l_1,l_1} \end{vmatrix}. \tag{3.2.9}$$

Using (3.2.1) we expand the determinant in (3.2.9) and perform algebraic manipulation to show that (3.2.9) simplifies to

$$n \det(G) \det(L^+_S). \tag{3.2.10}$$

Thus,

$$\sum_{i=1}^{n} \Sigma_{i,i} = \frac{\sigma^2}{2}\Big(\frac{K_f}{n} + n \det(G) \det(L^+_S) + \frac{1}{2}\mathrm{Tr}(\bar{G}\Gamma_S) - \mathbf{1}_n^T[\bar{G}\Gamma_S]\mathbf{e}_{l_1}\Big) = \frac{\sigma^2}{2}\Big(\frac{n}{\rho_S}\Big) \tag{3.2.11}$$

where ρ_S is defined by (3.2.2). □

3.3 Interpretation

In this section we provide interpretation of, and intuition on, the joint centrality measure, we prove explicit solutions to the optimal leader selection problem in a few cases, and we consider noise-corrupted leaders in the case of $m = 1$ and $m = 2$. Our central insight is that joint centrality of a set of nodes is a generalization of information centrality of an individual node: the joint centrality of a set of nodes is directly related to the information centrality of each individual node in the set and a coverage of the graph by the whole set, defined in terms of distribution of the set over the graph with respect to resistance and biharmonic distances. These

components of joint centrality may be in tension, since the most information central nodes can be close to one another (e.g., in the path graph), in which case they may be insufficiently distributed over the graph to provide good coverage. The optimal leader set is composed of nodes that trade off high nodal information centrality (close to the center in the path graph example) with good coverage (close to the ends in the path graph example).

We begin in this section by examining the terms in the expression for joint centrality in the case of an arbitrary number of noise-free leaders m, and show the connection to information centralities and coverage. We solve the optimal leader selection problem in the case of a cycle graph and illustrate further with a more general example. We then specialize to the case of $m = 1$ leader, and show how joint centrality specializes to information centrality of the leader node, with or without noise corruption. Next we specialize to the case of $m = 2$ leaders, where the expression for joint centrality facilitates a close examination of the trade-off between information centralities and coverage provided by the two leaders. We prove an explicit solution for the optimal set of two leaders in the case of the path graph. We also address the problem for $m = 2$ noise-corrupted leaders and provide intuition. We finish the section with a discussion of our results in light of greedy algorithms for finding optimal leader sets, and we make connections to controllability.

3.3.1 Joint centrality and an arbitrary number of leaders m

We interpret the results of Theorem 1 in the following two remarks. We then illustrate the notion of coverage by proving the explicit solution to the optimal leader set in the case of a cycle graph. We illustrate the trade-off between centrality and coverage with an example network.

Remark 1. *Using Theorem 1 to compute the total system error in terms of joint*

centrality of the m leader nodes provides a significant reduction in computation as compared to using the definition of total system error (3.1.9). Using joint centrality one only needs to compute the inverse of two $n \times n$ matrices L^+ and L^{2+} and then for each candidate set of leaders the inverse of an $(m-1) \times (m-1)$ matrix. This is in contrast to using the definition (3.1.9), which requires computing the inverse of the $n \times n$ matrix M for each candidate set of leaders.

.

Remark 2. *Theorem 1 reveals how the solution to the optimal leader selection problem is an optimal trade-off between high information centrality of the leader nodes and high resistance distances and biharmonic distances between leader nodes. To see this we examine the terms in (3.2.2) for joint centrality ρ_S.*

First, the elements of N^{-1} given by (3.2.1) depend on resistance distances:

$$N^{-1}_{i,j} = \frac{1}{2}(r_{i,l_1} + r_{j,l_1} - r_{i,j}).$$

Thus $N^{-1}_{i,j}$ quantifies a joint resistance distance between a pair of nodes i, j and l_1, Then, $\det(G) = (\det(N^{-1}_{S\setminus l_1}))^{-1}$ depends on these joint resistance distances among leaders.

Second, by (2.3.6) each diagonal element of L^+_S corresponds to a leader node and depends directly on the inverse of its information centrality as follows:

$$L^+_{s,s} = \frac{1}{c_s} - \frac{K_f}{n^2}.$$

By (2.3.7) the off diagonal elements of L^+_S depend on information centralities and resistance distances between leaders:

$$L^+_{s,t} = \frac{1}{2}\left(\frac{1}{c_s} + \frac{1}{c_t} - r_{s,t} - 2\frac{K_f}{n^2}\right).$$

Maximizing ρ_S requires a small $\det(G)\det(L_S^+)$, which suggests a key trade-off between high information centrality of leaders and high resistance distances between leaders.

The term $\mathrm{Tr}(Q)$ in (3.2.2) is the sum of products of the biharmonic distances between pairs of leader nodes (from Γ_S), and terms in G. Since $\mathrm{Tr}(Q)$ is negative, maximizing joint centrality requires high biharmonic distances between pairs of leader nodes. Biharmonic distance between a pair of nodes depends strongly on global connectivity of the graph and together with resistance distances provides a measure of coverage of the graph by a node set. Thus, the joint centrality measure makes rigorous how the optimal leader set trades off high information centrality of each of the nodes in the set with a good coverage of the graph by the set of nodes.

To better understand the coverage term, we first consider the case of a cycle graph. Because each node in the cycle graph has the same information centrality, it is only the coverage term that matters in the optimization of joint centrality. We can use the cyclic structure of the graph Laplacian to explicitly solve for the optimal locations of m noise-free leaders. In Theorem 5, we show that the optimal leader set is a set of nodes uniformly distributed about the cycle, which corresponds to a set that maximizes coverage of the graph.

Next, to illustrate the trade-off between nodal information centrality and coverage, we consider the unweighted, undirected, connected graph shown in Figure 3.1. The optimal sets of one, two and three leaders are shown in yellow, green and blue, respectively. Visually, it is clear that the optimal choice for a single leader (node 9, in yellow) has a central position in the network. In fact, node 9 has the highest information centrality c_i (2.3.4), consistent with Corollary 1 of Section 3.3.2, where it is proved that the optimal single leader is the most information central node.

Interestingly, it is observed that the optimal single leader is not a member of the optimal set of two leaders (nodes 2 and 3, in green). This is due to the fact that the optimal two leaders need to trade off high information centrality as individuals with

a joint coverage of the graph (see also Corollary 2 in Section 3.3.3). For this reason the optimal two leaders are well connected within the graph and distanced from each other.

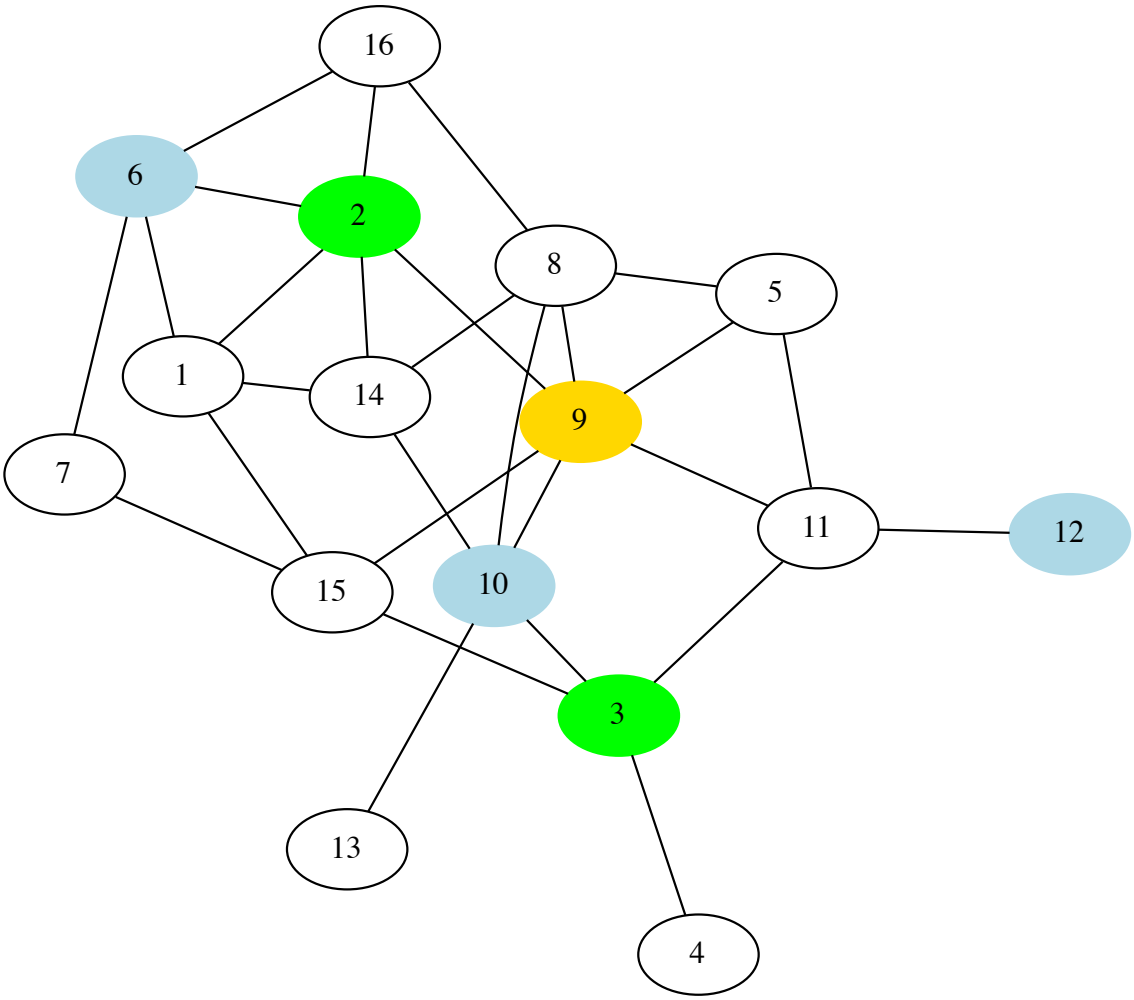

Figure 3.1: Solutions to the optimal leader set for robustness for an example graph with sixteen nodes. For $m = 1$ leader, the optimal solution is node 9, shown in yellow. For $m = 2$ leaders, the optimal solution is the set of nodes 2 and 3, shown in green. For $m = 3$ leaders, the optimal solution is the set of nodes 6, 10, and 12, shown in blue.

The optimal three leaders (nodes 6, 10, 12, in blue) further illustrate the key trade-off between leaders that are central and leaders that cover the graph. Although node 12 is not so well connected, its large resistance and biharmonic distances from nodes 6 and 10 make it part of the optimal three-leader set. That is, the three-node

leader set has optimal joint influence on the graph, as encoded by the joint centrality of the set.

The three solutions illustrate how a leader selection algorithm that first selects a leader and then iteratively adds to the set would result in a sub-optimal leader set for this example and likely in general (see also the example in [74]).

To further demonstrate how the node set with highest joint centrality is comprised of nodes that are both central and distributed over the network, we consider the highly clustered graph in Figure 3.2, where the optimal sets of m =1,2,3, 4 nodes are shown in orange. We see that as the number of leader nodes approaches the number of clusters, the optimal leaders become highly central nodes within each cluster. This characteristic will prove to be useful in Section 5.2 when we discuss the use of joint centrality for graph clustering.

3.3.2 Optimal selection of a single noise-corrupted or noise-free leader

Joint centrality reduces to information centrality in the case of a single leader ($m = 1$), with or without noise corruption. Thus, the optimal single leader is the node with the highest information centrality.

Corollary 1 (Optimal leader set, $m = 1$)**.** *Let $\mathcal{G}$ be an undirected, connected graph of order n. Let $S = \{s\}$ be a set of one noise-corrupted leader ($k < \infty$) with information centrality c_s. Then, the total system error (3.1.9) for the system dynamics (3.1.2) is*

$$\sum_{i=1}^{n} \Sigma_{i,i} = \frac{n\sigma^2}{2}\left(\frac{1}{k} + \frac{1}{c_s}\right). \tag{3.3.1}$$

If instead the leader set S is noise-free, then the total system error (3.1.9) for the

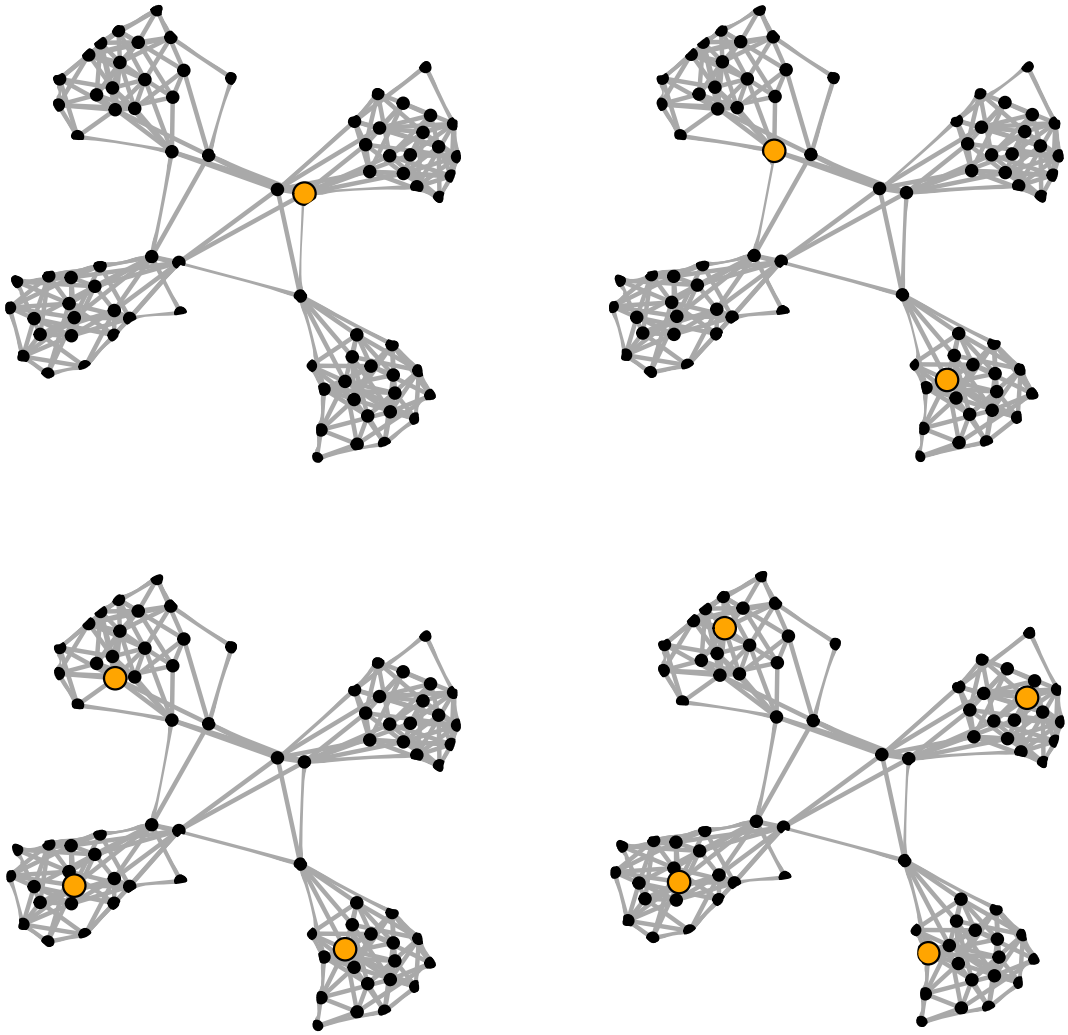

Figure 3.2: Network of $n = 80$ nodes with four distinct clusters. The orange nodes represent the optimal sets of $m = 1, 2, 3, 4$ nodes. For a single leader, the optimal node is central in the graph, with connections to all clusters. As the number of leaders increases to equal the number of clusters we see that the optimal leaders become central nodes in each cluster.

system dynamics (3.1.2) is

$$\sum_{i=1}^{n} \Sigma_{i,i} = \frac{n\sigma^2}{2} \left(\frac{1}{c_s} \right). \tag{3.3.2}$$

In both the noise-corrupted and the noise-free cases, the optimal leader set $S^* = \{s^*\} = \arg\max_s c_s$ *, the node with maximal information centrality* c_{s^*}*.*

Proof. For a single leader we only need to consider a rank-one update to the pseudoinverse of L. From (3.2.4), where $l_1 = s$, this is

$$N^{-1} = L^{-1} - L_s^+ \mathbf{1_n}^T - \mathbf{1_n} L_s^{+T} + \frac{(1 + kL_{s,s}^+)}{k} \mathbf{1_n}^T \mathbf{1_n}. \tag{3.3.3}$$

Summing the diagonal elements of (3.3.3) and applying (2.3.6), (2.3.1), (2.3.5) yields

$$\sum_{i=1}^{n} N_{i,i}^{-1} = \frac{K_f}{n} + \frac{n}{k} + n\left(\frac{1}{c_s} - \frac{K_f}{n^2}\right) = \frac{n}{k} + \frac{n}{c_s}. \tag{3.3.4}$$

Subsequently substituting into (3.1.9) gives the total system error

$$\sum_{i=1}^{n} \Sigma_{i,i} = \frac{n\sigma^2}{2}\left(\frac{1}{k} + \frac{1}{c_s}\right). \tag{3.3.5}$$

To get the total system error in the case of one noise-free leader, we take the limit of (3.3.5) as $k \to \infty$, which gives

$$\lim_{k\to\infty} \sum_{i=1}^{n} \Sigma_{i,i} = \lim_{k\to\infty} \frac{n\sigma^2}{2}\left(\frac{1}{k} + \frac{1}{c_s}\right) = \frac{n\sigma^2}{2}\left(\frac{1}{c_s}\right). \tag{3.3.6}$$

The total system error in (3.3.5) and in (3.3.6) is minimized when the leader has the highest information centrality. □

Remark 3. *Our definition of joint centrality derives from the definition of the optimal leader selection problem in terms of minimizing total system error (3.1.9). However, we have shown in Corollary 1 that joint centrality can be interpreted as a generalization of information centrality of a single node. This suggests the possibility of using joint centrality for generalizing from individual nodes to sets of nodes in problems where information centrality is a critical measure. For example, it is proved in [95] that information centrality of a node in a network performing distributed hypothe-*

sis testing determines its speed-accuracy trade-off. Joint centrality may be useful for investigating the decision-making performance of a set of nodes in this context.

3.3.3 Joint centrality and two noise-free leaders

In order to provide further intuition, we specialize Theorem 1 to the case of two noise-free leaders. In this case the expression for joint centrality simplifies as compared to the case of arbitrary m, and we can more closely examine the terms that determine the centrality versus coverage trade-off in the optimal leader set.

Corollary 2 (Optimal noise-free leader set, $m = 2$)**.** *Let $\mathcal{G}$ be an undirected, connected graph of order n. Let $S_2 = \{s_1, s_2\}$ be a set of two noise-free leaders. Then, the total system error (3.1.9) for the system dynamics (3.1.2) is*

$$\sum_{i=1}^{n} \Sigma_{ii} = \frac{\sigma^2}{2} \left(\frac{n}{\rho_{S_2}} \right), \tag{3.3.7}$$

where ρ_{S_2} is the joint centrality of S_2 given by (3.2.2), which specializes to

$$\rho_{S_2} = n \Big(\frac{K_f}{n} + \frac{n L^+_{s_1,s_1} L^+_{s_2,s_2} - n {L^+_{s_1,s_2}}^2 - \gamma_{s_1,s_2}}{r_{s_1,s_2}} \Big)^{-1}. \tag{3.3.8}$$

The optimal leader set is $S_2^* = \{s_1^*, s_2^*\} = \arg\max_{s_1,s_2} \rho_{S_2}$, the two nodes with the maximal joint centrality.

Proof. In the case of two leaders, $G = \frac{1}{r_{s_1,s_2}}$. Equation (3.3.7) follows directly from simplification of (3.2.3) and (3.2.2) from Theorem 1. □

Remark 4. *Following Remark 2, we see that in the two-leader case the term $det(G)det(L_S^+) = (L^+_{s_1,s_1} L^+_{s_2,s_2} - {L^+_{s_1,s_2}}^2)/r_{s_1,s_2}$, which is small for large leader information centrality and large resistance distance between leaders. The term $Tr(Q)$ is proportional to*

$-\gamma_{s_1,s_1}/r_{s_1,s_2}$. *For this term to be small, the biharmonic distance should be large relative to the resistance distance between leaders.*

We prove in Section 4.2 that the optimal pair of leaders on an unweighted path of length n are at locations $s_1^* = \text{rnd}(\frac{n}{5} + \frac{1}{2})$ and $s_2^* = \text{rnd}(\frac{4n}{5} + \frac{1}{2})$, where rndis rounding to the closest integer. We observe that for large n, the optimal two leader locations on the path approach 0.2 and 0.8 of the path length (starting from one end). This is in contrast with the cycle, where the optimal two leaders maintain a distance between each other equal to 0.5 of the number of nodes. Considering that the path is simply a cycle with one edge removed, it is interesting to observe that for large n, removing an edge from a cycle will cause the fraction of nodes between the optimal two leaders to increase from 0.5 to 0.6. That is, the optimal two leaders in the path are more spread out towards the two endpoints. The locations of the optimal two leaders in the path can be understood to be the optimal solution to the trade-off between high information centrality of two symmetrically distributed leaders, which increases as the two leaders get closer to midpoint and thus to each other, and good coverage, which requires the two leaders to be sufficiently distant from each other. The optimal two-leader set does not include the optimal single leader set, which is the node at the midpoint of the path, following Corollary 1 of Section 3.3.2.

3.3.4 Joint centrality and two noise-corrupted leaders

To address the case of two noise-corrupted leaders, where $k < \infty$, we define a k-dependent joint centrality of a set of two nodes. We then derive the solution to the optimal leader selection problem for two noise-corrupted leaders by calculating the total system error in terms of the k-dependent joint centrality of the two-leader set.

Theorem 2 (Optimal noise-corrupted leader set, $m = 2$)**.** *Let $\mathcal{G}$ be an undirected, connected graph of order n. Let $S_2 = \{s_1, s_2\}$ be a set of two noise-corrupted leaders*

($k < \infty$). Define ρ_{kS_2}, the k-dependent joint centrality of S_2, as

$$\rho_{kS_2} = n\Big(\frac{K_f}{n} + \frac{n[1 + k(L^+_{s_1,s_1} + L^+_{s_2,s_2})]}{k(2 + kr_{s_1,s_2})} + \frac{nk^2(L^+_{s_1,s_1}L^+_{s_2,s_2} - {L^+_{s_1,s_2}}^2) - k^2\gamma_{s_1,s_2}}{k(2 + kr_{s_1,s_2})}\Big)^{-1}. \tag{3.3.9}$$

Then, the total system error (3.1.9) for the system dynamics (3.1.2) is

$$\sum_{i=1}^{N} \Sigma_{ii} = \frac{\sigma^2}{2}\left(\frac{n}{\rho_{kS_2}}\right). \tag{3.3.10}$$

The optimal leader set is $S_2^* = \{s_1^*, s_2^*\} = \arg\max_{s_1,s_2} \rho_{kS_2}$, the two nodes with the maximal k-dependent joint centrality.

Prior to proving Theorem 2, we state a lemma from [67] that provides a simplification of the Woodbury formula in the case of a rank one update to a matrix.

Lemma 1. *[67] For rank one square matrix H and nonsingular X and $X + H$, $(X + H)^{-1}$ can be written as*

$$(X + H)^{-1} = X^{-1} - \frac{1}{1 + g}X^{-1}HX^{-1}, \tag{3.3.11}$$

where $g = \mathrm{Tr}(HX^{-1})$.

Proof. (Theorem 2). Let K_1, K_2 be rank one matrices with ${K_1}_{s_1,s_1} = k$, ${K_2}_{s_2,s_2} = k$ where $k > 0$ and all other elements of K_1, K_2 are zero. Let $K = K_1 + K_2$ and $N = L + K_1$. Then, $M = L + K = N + K_2$.

By applying Lemma 1, we compute

$$\begin{aligned} M^{-1} &= (N + K_2)^{-1} \\ &= N^{-1} - \frac{1}{1 + \mathrm{Tr}(K_2N^{-1})}N^{-1}K_2N^{-1}. \end{aligned} \tag{3.3.12}$$

By (3.2.4)

$$\begin{aligned}\mathrm{Tr}(K_2N^{-1}) &= 1 + kL^+_{s_2,s_2} - 2kL^+_{s_2,s_1} + kL^+_{s_1,s_1} \\ &= 1 + k\, r_{s_1,s_2}. \end{aligned} \tag{3.3.13}$$

Plugging (3.3.13) into (3.3.12) yields total system error (3.1.9):

$$\sum_{i=1}^{n}\Sigma_{i,i} = \frac{\sigma^2}{2}\sum_{i=1}^{n} M^{-1}_{i,i} = \frac{\sigma^2}{2}\sum_{i=1}^{n}\left(N^{-1}_{i,i} - \frac{1}{2+kr_{s_1,s_2}}(N^{-1}K_2N^{-1})_{i,i}\right). \tag{3.3.14}$$

Expanding N^{-1} in terms of L^+ and applying (2.3.1) and (2.3.5) gives

$$\begin{aligned}\sum_{i=1}^{n} M^{-1}_{i,i} =& \frac{n}{k} + \frac{K_f}{n} + nL_{s_1,s_1} - \frac{1}{2+k\, r_{s,p}}\Big(k\sum_{i=1}^{n}(L^+_{i,s_1} - L^+_{i,s_2})^2 + nk(L^+_{s_1,s_2})^2 - \\ & 2nL^+_{s_1,s_2} - 2nkL^+_{s_1,s_1}L^+_{s_1,s_2} + 2nL^+_{s_1,s_1} + nk(L^+_{s_1,s_1})^2 + \frac{n}{k}\Big). \end{aligned} \tag{3.3.15}$$

Rearranging terms and substituting from (2.3.12) results in

$$\begin{aligned}\sum_{i=1}^{n}\Sigma_{i,i} &= \frac{\sigma^2}{2}\Big(\frac{K_f}{n} + \frac{n + nk(L^+_{s_1,s_1} + L^+_{s_2,s_2})}{k(2+kr_{s_1,s_2})} + \frac{nk^2(L^+_{s_1,s_1}L^+_{s_2,s_2} - {L^+_{s_1,s_2}}^2) - k^2\gamma_{s_1,s_2}}{k(2+kr_{s_1,s_2})})\Big) \\ &= \frac{\sigma^2}{2}\Big(\frac{n}{\rho_{kS_2}}\Big). \end{aligned} \tag{3.3.16}$$

□

We observe that the k-dependent joint centrality (ρ_{kS_2} from Theorem 2) plays the same role in determining total system error with two noise-corrupted leaders (3.3.10) as joint centrality (ρ_{S_2} from Corollary 2) plays in determining total system error with two noise-free leaders (3.3.7). Further, as expected, in the limit as $k \to \infty$ we see that ρ_{kS_2} approaches ρ_{S_2}. To better understand the results in the case of finite k, we

compute the Taylor series expansion of ρ_{kS_2} (3.3.10) about $k = 0$:

$$\rho_{kS_2} = 2k + \left(r_{s_1,s_2} - (L^+_{s_1,s_1} + L^+_{s_2,s_2}) - \frac{4K_f}{n^2} \right) k^2 + O(k^3). \tag{3.3.17}$$

Thus, for $k << 1$, ρ_{kS_2} can be approximated by $2k + \left(r_{s_1,s_2} - (L^+_{s_1,s_1} + L^+_{s_2,s_2}) - \frac{4K_f}{n^2} \right) k^2$.

Remark 5. *It can be seen in (3.3.17) that $\rho_{kS_2} \to 0$ as $k \to 0$. This follows since at $k = 0$ there are no leaders and thus no centrality of leaders. For $k << 1$, the approximation $2k + \left(r_{s_1,s_2} - (L^+_{s_1,s_1} + L^+_{s_2,s_2}) - \frac{4K_f}{n^2} \right) k^2$ of ρ_{kS_2} reveals a trade-off similar to the trade-off in the noise-free case. The trade-off implies that the k-dependent joint centrality is maximized for large resistance distance r_{s_1,s_2} between the two leaders and for large information centrality of each of the two leaders. In the case of a symmetric graph where each node has the same information centrality, the optimal leader set is the one in which the pair has maximum resistance distance.*

Chapter 4

Joint Centrality and Optimal Leader Sets in Trees and Cycles

In this chapter we focus on special cases of graphs, specifically unweighted tree and cycle graphs. The constraints provided by these classes of graphs allow for increased tractability; thus, we are able to thoroughly explore the optimal leader selection problem in each case. The study of canonical graph structures provides us with valuable insight that can be applied in future work to aid leader selection in more complex networks. Section 4.3 and a preliminary version of Section 4.2 have been previously published in Fitch and Leonard [26].

4.1 Joint Centrality in Unweighted Trees

We will now investigate general unweighted tree graphs and demonstrate how the equation for joint centrality reduces for two and three nodes in an unweighted tree graph. Our motivation for studying trees comes from large scale applications where networks can be approximated by their minimum spanning tree to increase sparsity and decrease computational time.

As discussed in Section 2.3, tree graphs are in the class of planar graphs, implying

that we can embed any tree graph in a two dimensional plane without edge crossings. We can also embed a tree graph in three dimensions with a third dimension, h, which for each node is the corresponding diagonal entry of L^+. This allows us to easily view, in a single plot, how the information centrality of each node relates to its location in the network and the relative information centralities of other nodes. An example embedding is shown in Figure 4.1.

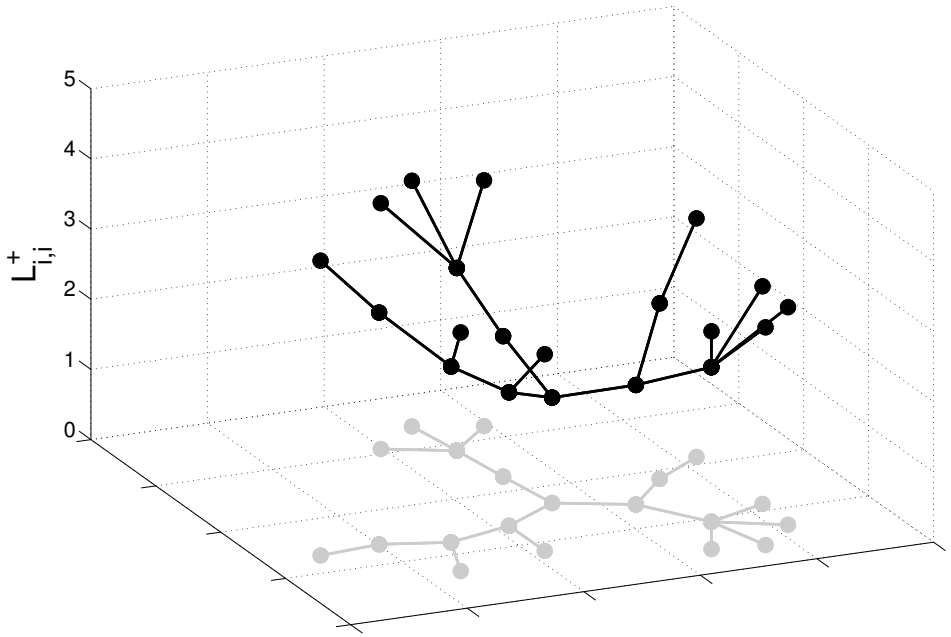

Figure 4.1: Three dimensional plot of an unweighted, undirected tree graph. The first two dimensions are the planar embedding of the tree and, for each node i, the third (vertical) dimension corresponds to $L^+_{i,i}$.

Furthermore, since there is only one unique path between any two nodes in a tree graph, we can also select a path between two nodes and represent nodes in this path on a plane where the vertical axis is h and the horizontal axis is position along the path. In other words, slice Figure 4.1 along a path. We define the function $h(z)$ as the piecewise linear function created by connecting the diagonal entries of L^+ corresponding to nodes along any path in a tree graph, ordered by position along the

path. Let a and $b = a + 1$ be indices for a sequential pair of nodes. Then, we let

$$h(z) = (L^+_{p_b,p_b} - L^+_{p_a,p_a})z + (a+1)L^+_{p_a,p_a} - aL^+_{p_b,p_b} \quad a \leq z \leq b,$$

where the subscript p_a indicates the a^{th} node along path $p^{(ij)}$. In particular, $h(a) = L^+_{p_a,p_a}$ and $h(b) = L^+_{p_b,p_b}$. We will use superscripts on $h(z)$ to denote the first and last nodes along the path. For example, $h(z)^{(i,j)}$ is generated from the path connecting nodes i, j. Figure 4.2 demonstrates the function $h(z)^{(4,18)}$ in an example tree graph.

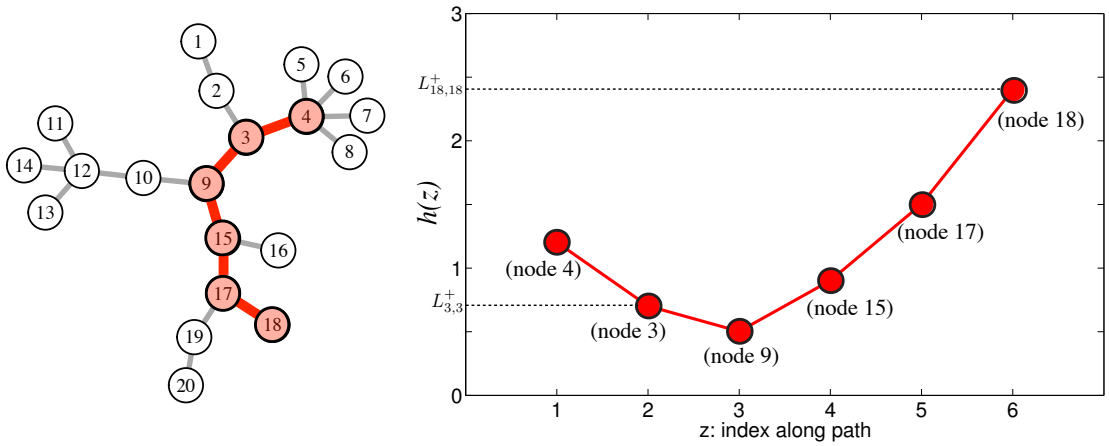

Figure 4.2: Example tree graph with path $p^{(4,18)}$ highlighted (left) and corresponding plot of $h(z)$ along path $p^{(4,18)}$

We will show in Theorem 3 that the average value of $h(z)$ along the path appears as a term in the biharmonic distance between those two nodes. Before presenting the theorem, we first introduce notation specific to this section. For a pair of nodes a and b, let $\hat{a}$ be the set of n_a nodes indexed by i for which the path $p^{(bi)}$ contains node a. Similarly let $\hat{b}$ be the set of n_b nodes indexed by j for which the path $p^{(aj)}$ contains node b. Finally, let $\hat{y}$ be the set of all nodes not in $\hat{a}$ or $\hat{b}$ and let the number of nodes in $\hat{y}$ be $n_y = n - n_a - n_b$. Figure 4.3 an example of this grouping where a is chosen as node 10 and b is node 17. We then write $L^+_{a,a}$ and $L^+_{b,b}$ in terms of these parameters:

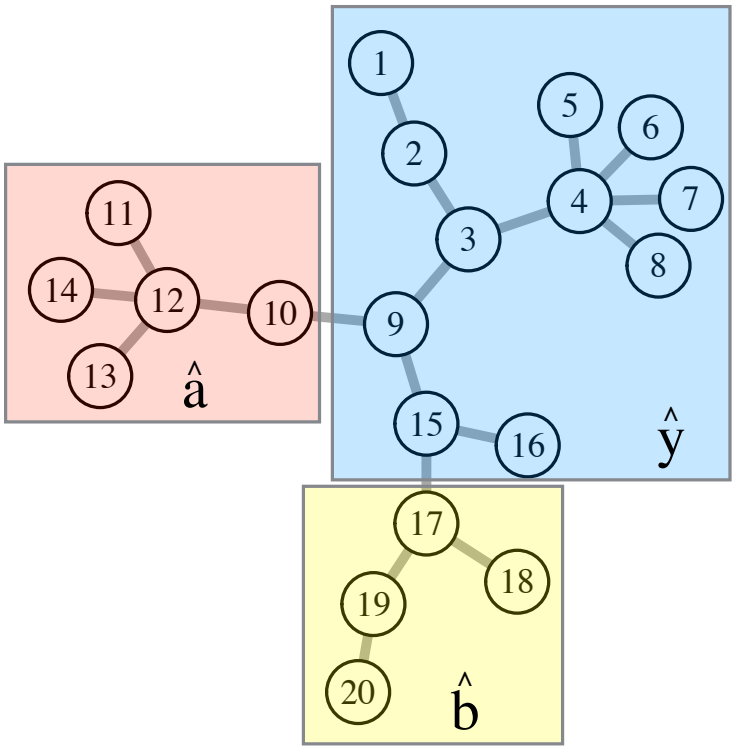

Figure 4.3: Example partitioning of a tree graph into sets $\hat{a}, \hat{b}$, and $\hat{y}$ with $a = 10$, $b = 17$. With this partitioning $n_a = 5$, $n_b = 4$ and $n_y = 11$.

$$L^+_{a,a} = -\frac{K_f}{n^2} + \frac{1}{n}\Big(n_b\delta_{a,b} + \sum_{i\in\hat{a}}\delta_{a,i} + \sum_{j\in\hat{b}}\delta_{b,j} + \sum_{h\in\hat{y}}\delta_{a,h}\Big) \tag{4.1.1}$$

$$L^+_{b,b} = -\frac{K_f}{n^2} + \frac{1}{n}\Big(n_a\delta_{a,b} + \sum_{i\in\hat{a}}\delta_{a,i} + \sum_{j\in\hat{b}}\delta_{b,j} + \sum_{h\in\hat{y}}\delta_{b,h}\Big), \tag{4.1.2}$$

where $\delta_{i,j}$ is the geodesic distance between nodes i and j. Geodesic distance is equivalent to resistance distance in trees, that is $\delta_{i,j} = r_{i,j}$.

We will index a node on the path $p^{(a,b)}$ by α, where $\alpha \neq a, b$. We define n_α to be the number of nodes, j, for which node α is incident on each path $p^{(a,b)}$, $p^{(a,j)}$and $p^{(b,j)}$. For example, consider Figure 4.3 and let $a = 10$ and $b = 17$; if $\alpha = 15$ then $n_\alpha = 2$, if $\alpha = 9$, then $n_\alpha = 9$. The notation $\sum_{\alpha\in p^{(a,b)}}$ implies that we are summing over each node on the path $p^{(a,b)}$, excluding a and b.

Theorem 3 (Biharmonic distance in unweighted tree graphs)**.** *Let $\mathcal{G}$ be a connected,*

undirected, unweighted tree graph of order n. Let a, b be two nodes in $\mathcal{G}$ and let $\gamma_{a,b}$ be the biharmonic distance between a and b. Then $\gamma_{a,b}$ is

$$\gamma_{a,b} = \frac{n}{4}\Big(- (L^+_{a,a} - L^+_{b,b})^2 - 2(L^+_{a,a} + L^+_{b,b} + 2\hat{L}^+_{a,b})\delta_{a,b} + \delta^2_{a,b}\Big), \tag{4.1.3}$$

where $\hat{L}^+_{a,b}$ is the average value of the piecewise continuous function $h(z)^{(a,b)}$.

Proof. We begin the proof by writing biharmonic distance defined in (2.3.12) as a sum in three parts,

$$\begin{aligned}\gamma_{a,b} &= \sum_{i\in\hat{a}}(L^+_{a,a} - L^+_{b,b} + \delta_{b,i} - \delta_{a,i})^2 + \sum_{j\in\hat{b}}(L^+_{a,a} - L^+_{b,b} + \delta_{b,j} - \delta_{a,j})^2 + \\ &+ \sum_{h\in\hat{y}}(L^+_{a,a} - L^+_{b,b} + \delta_{b,h} - \delta_{a,h})^2 \\ &= n_a(L^+_{a,a} - L^+_{b,b} + \delta_{a,b})^2 + n_b(L^+_{a,a} - L^+_{b,b} - \delta_{a,b})^2 + \sum_{h\in\hat{y}}(L^+_{a,a} - L^+_{b,b} + \delta_{b,h} - \delta_{a,h})^2.\end{aligned} \tag{4.1.4}$$

Expansion of the last term of (4.1.4) and application of (4.1.1), (4.1.2), and relationship $\delta_{b,\alpha} = \delta_{a,b} - \delta_{a,\alpha}$ yields

$$\gamma_{a,b} = \frac{1}{n}\Big((n-n_b)n_b\delta^2_{a,b} - \sum_{\alpha\in p^{(a,b)}} n_\alpha\delta_{a,\alpha}\Big(2n_b\delta_{a,b} + \sum_{\alpha\in p^{(a,b)}} n_\alpha\delta_{a,\alpha}\Big) + n\sum_{\alpha\in p^{(a,b)}} n_\alpha\delta^2_{a,\alpha}\Big). \tag{4.1.5}$$

Furthermore, by expressing $\delta^2_{a,\alpha}$ in terms of $L^+_{a,a}$, $L^+_{b,b}$ and associated sums it can be shown that

$$\sum_{\alpha\in p^{(a,b)}} n_\alpha\delta^2_{a,\alpha} = \frac{n}{2}(L^+_{a,a} + L^+_{b,b})(-\delta_{a,b} + 1) + n\sum_{\alpha\in p^{(a,b)}} L^+_{\alpha,\alpha} + \delta_{a,b}\sum_{\alpha\in p^{(a,b)}} n_\alpha\delta_{a,\alpha}. \tag{4.1.6}$$

We observe that $\frac{1}{2}(L^+_{a,a} + L^+_{b,b}) + \sum_{\alpha\in p^{(a,b)}} L^+_{\alpha,\alpha}$ is equivalent to the area under the

curve $h(z)^{(a,b)}$. And subsequently

$$-\frac{1}{2\delta_{a,b}}(L^+_{a,a}+L^+_{b,b})-\frac{1}{\delta_{a,b}}\sum_{\alpha\in p^{(a,b)}}L^+_{\alpha,\alpha}=-\hat{L}^+_{a,b}. \tag{4.1.7}$$

Combining (4.1.6) and (4.1.7), substituting into (4.1.5) and manipulation of terms yields

$$\gamma_{ab}=\frac{n}{4}\Big(-(L^+_{a,a}-L^+_{b,b})^2-2(L^+_{a,a}+L^+_{b,b}+2\hat{L}^+_{a,b})\delta_{a,b}+\delta^2_{a,b}\Big).$$

□

In the following subsections we apply Theorem 3 to two-node and three-node leader selection, respectively, in unweighted trees.

4.1.1 Two-node Leader Selection in Unweighted Trees

Corollary 3 (Two-node joint centrality in unweighted tree graphs)**.** *Let $\mathcal{G}$ be a connected, undirected, unweighted tree graph of order n. Let S be a set of $m=2$ nodes and let ρ_{S2_t} be the joint centrality of set $S=\{s_1,s_2\}$ given by (3.2.2). Then the two node joint centrality reduces to*

$$\rho_{S2_t}=\Big(\frac{K_f}{n^2}+L^+_{s_1,s_1}+L^+_{s_2,s_2}-\frac{1}{2}\delta_{s_1,s_2}-\hat{L}^+_{s_1,s_2}\Big)^{-1}. \tag{4.1.8}$$

Proof. Applying equations (4.1.1), (4.1.2) and (4.1.3) to (3.3.8) and simplification results in

$$\rho_{S2_t}=\Big(\frac{K_f}{n^2}+L^+_{s_1,s_1}+L^+_{s_2,s_2}-\frac{1}{2}\delta_{s_1,s_2}-\hat{L}^+_{s_1,s_2}\Big)^{-1}.$$

□

There are a number of interesting observations to be made about equation (4.1.8).

First, we see that though the leader selection problem is inherently still combinatorial, the equation for pairwise joint centrality has been greatly reduced in complexity. The constraints imposed by a tree graph allow for pairwise joint centrality to be written as a linear combination of interpertable terms. Additionally, the presence of the average of $h(z)$, rather than the average information centrality of each node in the path, signifies that the rate at which information centrality increases or decreases along a path is of importance. Furthermore, we are able to derive simple bounds for pairwise joint centrality in trees as presented in Corollary 4.

Corollary 4 (Lower bound on pairwise joint centrality in undirected, unweighted trees)**.** *Let $\mathcal{G}$ be a connected, undirected, unweighted path graph of order n. Let S be a set of $m = 2$ nodes and let ρ_{S2_t} be the joint centrality of set $S = \{s_1, s_2\}$ given by (4.1.8). Let s^* be the node with the highest information centrality. Then the pairwise joint centrality is lower bounded by*

$$\rho_{S2_t} \geq \Big(\frac{K_f}{n^2} + \Big(1 - \frac{1}{2\delta_{s_1,s_2}}\Big)(L^+_{s_1,s_1} + L^+_{s_2,s_2}) - \frac{1}{2}\delta_{s_1,s_2} - L^+_{s^*s^*} + \frac{L^+_{s^*,s^*}}{\delta_{s_1,s_2}}\Big)^{-1}. \quad (4.1.9)$$

Proof. The bound (4.1.9) follows from applying the inequality to (4.1.8)

$$\hat{L}^+_{a,b} \geq \frac{1}{2\delta_{a,b}}(L^+_{a,a} + L^+_{b,b}) + (1 - \frac{1}{\delta_{a,b}})L^+_{s^*s^*}. \quad (4.1.10)$$

□

The inequality (4.1.9) clearly demonstrates that we require optimal leaders to have high information centrality (inversely related to $L^+_{s_1,s_1}$ and $L^+_{s_2,s_2}$), yet be distant from each other in the graph (high δ_{s_1,s_2}). We also note that to compute the bound (4.1.9) we only require one inversion of an $n \times n$ matrix, whereas other approximations, such as the greedy algorithm in [74], would in general require $n(n-1)$ matrix inversions.

Corollary 5 (Upper bound on pairwise joint centrality in undirected, unweighted

trees). *Let $\mathcal{G}$ be a connected, undirected, unweighted path graph of order n. Let S be a set of $m = 2$ nodes and let ρ_{S2_t} be the joint centrality of set $S = \{s_1, s_2\}$ given by (4.1.8). Let s^* be the node with the highest information centrality. Then the pairwise joint centrality is upper bounded by*

$$\rho_{S2_t} \leq \Big(\frac{K_f}{n^2} + \frac{1}{2}(L^+_{s_1,s_1} + L^+_{s_2,s_2} - \delta_{s_1,s_2}) + \frac{1}{6}(-1 + \delta^2_{s_1,s_2}\Big)^{-1}. \tag{4.1.11}$$

Proof. The bound (4.1.11) follows from evaluating (4.1.3) with $n_\alpha = 1$ for every node α along the path between a and b. That is, evaluating (4.1.3) as though there is no branching along the path between a and b. □

While the expression for pairwise joint centrality (4.1.8) is significantly simplified from the general form (3.3.8), and provides valuable insight as to how the network structure and information centrality along the path connecting two leaders affects the joint centrality of those leaders, it may still be costly to evaluate for large networks. However, as previously mentioned, the bound (4.1.9) is computationally efficient even compared to greedy algorithms. Simulation suggests that (4.1.9) is indeed a tight bound for nodes with high joint centrality. We expect this because the difference between $L^+_{s^*,s^*}$ and $\hat{L}^+_{i,j}$ will in general be larger for i, j with very low information centralities, and smaller for nodes that are good candidates for members of the optimal set; that is i, j with high information centralities and the node with highest information centrality, L_{s^*,s^*}, along path $p^{(ij)}$.

4.1.2 Three-node Leader Selection in Unweighted Trees

In connected, undirected trees, for any three nodes i, j, l there is exactly one node, a, incident on the three paths $p^{i,j}$, $p^{i,l}$, and $p^{j,l}$, where it is possible that a is equal to i, j, or l. It follows that $\delta_{i,j} = \delta_{i,a} + \delta_{j,a}$,$\delta_{i,l} = \delta_{i,a} + \delta_{l,a}$, and $\delta_{j,l} = \delta_{j,a} + \delta_{l,a}$. We will call a the link node for set $\{i, j, l\}$.

Corollary 6 (Three-node joint centrality in tree graphs)**.** *Let $\mathcal{G}$ be a connected, undirected, unweighted tree graph of order n. Let S be a set of $m = 3$ nodes, let ρ_{S3_t} be the joint centrality of set $S = \{s_1, s_2, s_3\}$ given by (3.2.2). Let a be the link node of S. Then the three node joint centrality simplifies to*

$$\rho_{S3_t} = \Big(\frac{K_f}{n^2} + \sum_{i\in S}\Big(L^+_{i,i} - \frac{1}{2}\delta_{a,i} - \sum_{j<i}\hat{L}^+_{i,j}\Big) + \frac{\sum_{i\in S}\sum_{j<i}\hat{L}^+_{i,j}\delta_{a,i}\delta_{a,j} - \frac{1}{2}\prod_{i\in S}\delta_{a,i}}{\sum_{i\in S}\sum_{j<i}\delta_{a,i}\delta_{a,j}}\Big)^{-1}, \tag{4.1.12}$$

where $\hat{L}^+_{a,b}$ is the average value of the piecewise continuous function $h(z)^{(a,b)}$.

Proof. Representing the distance between any two nodes in the set in terms of the sum of distances to the link node, applying (4.1.1), (4.1.2) and (4.1.3) to (3.2.2), and algebraic simplification yields

$$\rho_{S3_t} = \Big(\frac{K_f}{n^2} + \sum_{i\in S}\Big(L^+_{i,i} - \frac{1}{2}\delta_{a,i} - \sum_{j<i}\hat{L}^+_{i,j}\Big) + \frac{\sum_{i\in S}\sum_{j<i}\hat{L}^+_{i,j}\delta_{a,i}\delta_{a,j} - \frac{1}{2}\prod_{i\in S}\delta_{a,i}}{\sum_{i\in S}\sum_{j<i}\delta_{a,i}\delta_{a,j}}\Big)^{-1}.$$

□

Corollary 7 (Lower bound on three-node joint centrality in undirected, unweighted trees)**.** *Let $\mathcal{G}$ be a connected, undirected, unweighted tree graph of order n. Let S be a set of $m = 3$ nodes and let ρ_{S3_t} be the joint centrality of set $S = \{s_1, s_2, s_3\}$ given by (4.1.12). Let s^* be the node with the highest information centrality. Then the three-node joint centrality is lower bounded by*

$$\rho_{S3_t} \geq \Big(\frac{K_f}{n^2} + 2L^+_{s^*,s^*} + \sum_{i\in S}\Big(L^+_{i,i} - \frac{1}{2}\delta_{a,i} - \frac{\delta_{a,i}}{6(2\delta_{a,i}+1)} - \frac{\delta_{a,i}L^+_{s^*,s^*} + L^+_{i,i}}{2\delta_{a^*}\delta_{a,i} + \delta^2_{a^*}}\Big)^{-1}. \tag{4.1.13}$$

where δ_{a^*} is the maximum distance from node a to any other node in the network.

Proof. The bound (4.1.13) follows from applying the inequality (4.1.10) and limits on

$\delta_{a,i}$, $1 \leq \delta_{a,i} \leq \delta_{a^*}$, to (4.1.12). $\square$

An important feature of the relationship (4.1.13) is that when applied to the leader selection problem we can separately evaluate the contribution of each leader node relative to the link node of the leader set. This allows us to perform computationally efficient leader selection. First, a link node, a, is selected following the steps outlines below. Then, for each branch formed by an edge incident to the link node, we search for the node that minimizes

$$q_i = L^+_{i,i} - \frac{1}{2}\delta_{a,i} - \frac{\delta_{a,i}}{6(2\delta_{a,i}+1)} - \frac{\delta_{a,i}L^+_{s^*,s^*} + L^+_{i,i}}{2\delta_{a^*}\delta_{a,i} + \delta^2_{a^*}}.$$

This results in a set of nodes with size equivalent to the degree of the link node. Of this set, the three nodes with minimal q_i are selected as the leader nodes. If the $d_a = 2$ then the link node is selected as a leader node. We note that leaf nodes cannot be link nodes.

It remains to be clarified how the link node is selected. In large networks where time complexity is a concern, an intuitive choice for a link node is the node with the highest information centrality and degree greater than or equal to three. Alternatively, one could iterate over all possible link nodes j, where $d_j \geq 2$ and select the leader set that maximizes (4.1.13).

The advantage of applying (4.1.13) for leader selection is that with an appropriate choice of link node, the resulting leader set will inherently have properties associated with high performing leader sets. That is, the nodes will have high information centralities, while also distributed over the network. By choosing leaders relative to a node common on all paths between leaders, we are able to indirectly account, in part, for the joint influence between leaders while still selecting each leader independently from the other two.

The method of first selecting a link node and then selecting leaders nodes from

branches of the link node can be extended for leader sets of size $m > 3$. At most $m-2$ link nodes are required to express joint centrality (3.2.2) in terms of information centralities of leader nodes and distances to link nodes. An area of ongoing research is bounding joint centrality of $m > 3$ leaders and developing computationally efficient heuristics for the selection of large, performing leader sets.

4.1.3 Leader Selection Examples

We provide two examples comparing the total system error for the optimal two and three node leader sets, the two and three node leader sets obtained by maximizing (4.1.9) and (4.1.13), and the two and three node leader sets obtained by the greedy algorithm in [74]. The first example, Figure 4.4, is a tree network of 20 nodes that has been used throughout this section to illustrate concepts. The second example, Figure 4.5, is the tree network of 25 nodes presented in [74]. As summarized in Table 4.1, we find that in the first example, the two-node leader set obtained by maximizing (4.1.9) is indeed the optimal leader pair. Furthermore, the three-node leader set obtained by maximizing (4.1.13) outperforms the leader set obtained by the greedy algorithm. In the second example the pair of leaders which maximizes (4.1.9) is once again an improvement over the greedy algorithm. However, total system error is lower for the 3-node leader set selected using the greedy algorithm than the set that maximizes (4.1.13). Table 4.2 summarizes leader selection results for the second example. Computational complexity for selecting leader sets through optimizing (4.1.9) and (4.1.13) is at worst $\mathcal{O}(n^3 + n^2)$, which provides a significant reduction from the complexity of the greedy algorithm in [74] that is order $\mathcal{O}(n^{m+3})$.

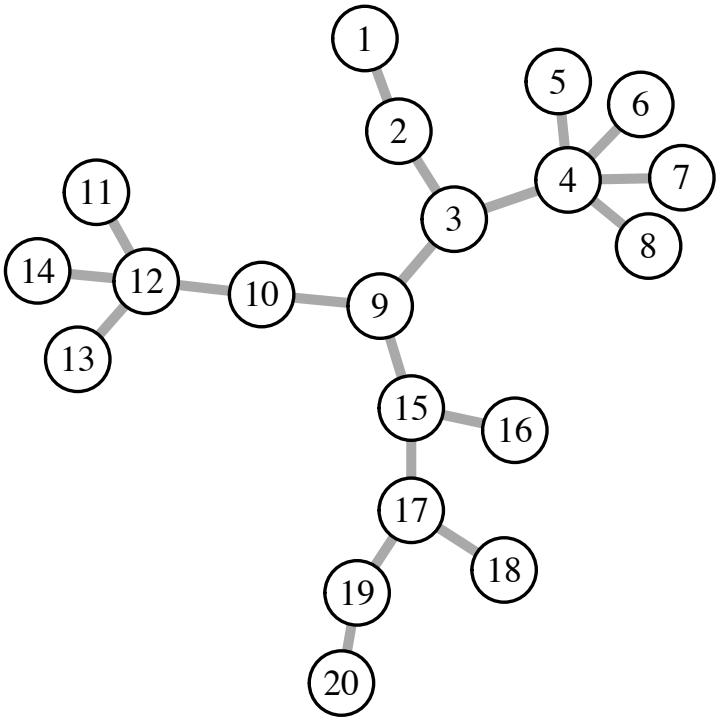

Figure 4.4: Example tree graph for leader selection results displayed in Table 4.1

	2 leader nodes	$\frac{1}{\sigma}\sum_{i=1}^{N}\Sigma_{i,i}$	ranking	3 leader nodes	$\frac{1}{\sigma}\sum_{i=1}^{N}\Sigma_{i,i}$	ranking
optimal	4, 17	16.88	1/190	4, 12, 17	19.67	1/1140
lower bound	4, 17	16.88	1/190	4, 10, 15	25.20	74/1140
greedy	4, 9	17.75	7/190	4, 9, 19	26.17	100/1140

Table 4.1: Two and three node leader selection results for tree graph shown in Figure 4.4

	2 leader nodes	$\frac{1}{\sigma}\sum_{i=1}^{N}\Sigma_{i,i}$	ranking	3 leader nodes	$\frac{1}{\sigma}\sum_{i=1}^{N}\Sigma_{i,i}$	ranking
optimal	4, 22	38.40	1/300	4, 16, 22	30.00	1/2300
lower bound	4, 19	41.67	6/300	4, 16, 19	36.20	105/2300
greedy	12, 22	44.75	10/300	5, 12, 22	33.25	12/2300

Table 4.2: Two and three node leader selection results for tree graph shown in Figure 4.5

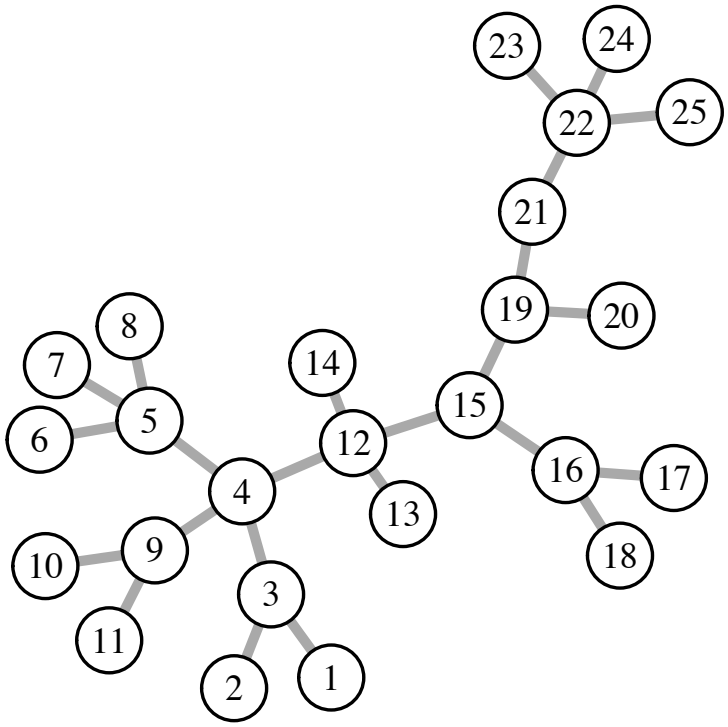

Figure 4.5: Example tree graph for leader selection results displayed in Table 4.2

4.2 Optimal Leader Selection in Unweighted Path Graphs

The unweighted path graph, a tree graph with only one path, is one of the simplest canonical graph structures. However, the optimal leader selection problem for robustness has yet to be solved for this class of graph. In this section we will prove a simple expression with conditions on rounding for the number of follower nodes between two adjacent leaders that minimizes total system error.

4.2.1 Optimal selection of m Noise-Free Leaders in an Unweighted Path Graph

We let a *block* be any set of follower nodes between two leaders or between a leader and an end of the path, including the end node. We define the positions of leaders

by defining the number of nodes in the blocks. Let the number of follower nodes in the first block, from the left end of the path to the first leader, be

$$\bar{w}_1 = \mathrm{rnd}\Big(\frac{n}{3m-1} - \frac{1}{2}\Big). \tag{4.2.1}$$

Let the number of follower nodes in the last block, from the right end of the path to the last leader be

$$\bar{w}_{m+1} = \mathrm{rnd}\Big(\frac{n}{3m-1} - \frac{1}{2}\Big). \tag{4.2.2}$$

Let the number of nodes in each middle block between two leaders along the path be

$$\bar{w}_j = \mathrm{rnd}\Big(\frac{3n}{3m-1} - 1\Big) \;\; j = 2, 3, ...m. \tag{4.2.3}$$

where rnd denotes rounding to an integer value and all blocks need not be rounded in the same direction. For a path of any length, we let the number of end blocks where (4.2.1) or (4.2.2) is rounded up to an integer value be b_e^u and the number of end blocks where (4.2.1) or (4.2.2) is rounded down to an integer value be b_e^d. Similarly, we let the number of middle blocks where (4.2.3) is rounded up to an integer value be b_m^u and the number of middle blocks where (4.2.3) rounded down to an integer value be b_e^d. Additionally, we let ξ be the remainder of $\frac{n}{3m-1}$ and let the relationship between ξ, b_e^u, b_e^d, b_m^u and b_m^d be as summarized in Table 4.3.

Theorem 4 (m noise-free leaders on a path)**.** *Let $\mathcal{G}$ be an undirected, unweighted path graph of order n. Let $S = \{s_1, ..., s_m\}$ be a set of m noise-free leaders with indices*

$$s_i = \sum_{l=1}^{i} (\bar{w}_l + 1).$$

where $\bar{w}_l$ is given by (4.2.1), (4.2.2), (4.2.3) and the number of blocks rounded up or

ξ	b_e^u	b_e^d	b_m^u	b_m^d	assumption
$\xi = 0$	1	1	n/a	n/a	
$0 < \xi < \frac{1}{3}$	2	0	$\xi(3m-1)-1$	$m - \xi(3m-1)$	
$\frac{1}{3} < \xi < \frac{1}{2}$	2	0	$\xi(3m-1)-m$	$2m-1-\xi(3m-1)$	$\xi \geq \frac{m}{3m-1}$
$\xi = \frac{1}{2}$	n/a	n/a	$\frac{1}{2}(m-1)$	$\frac{1}{2}(m-1)$	
$\frac{1}{2} < \xi < \frac{2}{3}$	0	2	$\xi(3m-1)-m$	$2m-1-\xi(3m-1)$	
$\frac{2}{3} < \xi < 1$	0	2	$\xi(3m-1)-2m+1$	$3m-2-\xi(3m-1)$	$\xi \leq \frac{3m-2}{3m-1}$

Table 4.3: Summary of relationship between ξ and the number of end and middle blocks rounded up or down to an integer.

down is in accordance with Table 4.3. Then S minimizes total system error, (3.1.9), and is the optimal set of leaders.

Proof. We begin by assuming m nodes on the path have been selected as leaders and let $M = L + K$ where K is a matrix with a value of k in the entries along the main diagonal corresponding to the leader nodes and zeros elsewhere. Rather than partition M in the usual way, we let the row and column related to each node in the graph be that node's incidence along the path, ordered from left to right. Let M_F be the matrix with leader node rows and columns deleted. Since we are assuming noise-free leaders, to compute total system error we need only to consider the sum of the diagonal elements of M_F^{-1}. M_F can be written as a block diagonal matrix where each block corresponds to a set of connected follower nodes between two leader nodes. There are $m+1$ blocks of M_F, where the i^{th} block is labeled as M_{F_i} and is represented

as follows,

$$M_{F_1} = \begin{bmatrix} 1 & -1 & & 0 \\ -1 & 2 & & \\ 0 & & \ddots & -1 \\ 0 & & -1 & 2 \end{bmatrix},$$

$$M_{F_i} = \begin{bmatrix} 2 & -1 & & 0 \\ -1 & 2 & & \\ 0 & & \ddots & -1 \\ 0 & & -1 & 2 \end{bmatrix} \quad i = 2, 3, ...m,$$

$$M_{F_{m+1}} = \begin{bmatrix} 2 & -1 & & 0 \\ -1 & 2 & & \\ 0 & & \ddots & -1 \\ 0 & & -1 & 1 \end{bmatrix}.$$

We let the length of block i be w_i. Total system error is then proportional to the sum of the trace of each $M_{F_i}^{-1}$. By applying (2.1.2) and (2.1.3) it can be shown that

$$\begin{aligned} \operatorname{tr}(M_{f_1}^{-1}) &= \sum_{\alpha=1}^{w_1} \alpha = \frac{1}{2}w_1^2 + \frac{1}{2}w_1, \\ \operatorname{tr}(M_{f_{m+1}}^{-1}) &= \sum_{\beta=1}^{w_{m+1}} \beta = \frac{1}{2}w_{m+1}^2 + \frac{1}{2}w_{m+1}. \end{aligned} \tag{4.2.4}$$

Furthermore, by [20] we have that the eigenvalues of $M_{F_i}^{-1}$ for $i = 2, .., m$ are

$$\lambda(M_{F_i}^{-1})_j = \frac{1}{2 - 2\cos\left(j\frac{\pi}{w_i+1}\right)} \quad j = 1, ..., w_i. \tag{4.2.5}$$

The sum of the eigenvalues of a block is equal to the trace of that block, which is

written as

$$\mathrm{tr}(M_{F_i}^{-1}) = \sum_{j=1}^{w_i} \lambda(M_{F_i}^{-1})_j = \frac{1}{6}w_i^2 + \frac{1}{3}w_i, \quad i = 2, ..., m. \tag{4.2.6}$$

Therefore, the optimal leader selection problem for an unweighted path becomes

$$\begin{aligned} \underset{w_1,\dots w_{m+1}}{\text{minimize}} \quad & J(w) = \mathrm{tr}(M_F^{-1}) = \frac{1}{2}\Big(w_1^2 + w_1 + w_{m+1}^2 + w_{m+1}\Big) + \sum_{j=2}^{m} \frac{1}{6}w_j^2 + \frac{1}{3}w_j \\ \text{subject to} \quad & \sum_{i=1}^{m+1} w_i = n - m, \quad w_i \in \mathbb{Z}^+. \end{aligned} \tag{4.2.7}$$

Since the optimization problem (4.2.7) is non-convex, we proceed by studying its convex relaxation in which the integer constraint, $w_i \in \mathbb{Z}^+$, is replaced with $w_i \in \mathbb{R}^+$. The relaxed problem is easily solved using Lagrange multipliers [7], yielding the following optimal (non-integer) block sizes

$$\begin{aligned} w_1^* = w_{m+1}^* &= \frac{n}{3m-1} - \frac{1}{2} \\ w_j^* &= \frac{3n}{3m-1} - 1 \quad j = 2, 3, ...m, \end{aligned} \tag{4.2.8}$$

which gives the optimal (non-integer) leader positions

$$s_i^* = \frac{1}{2} + \frac{(1 + 3(i-1))n}{3m-1} \quad i = 1, ...m. \tag{4.2.9}$$

To fully solve the optimal leader selection problem, we study the solution to the relaxed optimization problem and the increase in error induced by rounding block sizes to integer values. We let $\bar{w}_i = w_i^* + \epsilon_i$, where $\bar{w}_i \in \mathbb{Z}$ and $-1 < \epsilon_i < 1$, be w_i^* rounded to an integer value. Furthermore, $\sum_i^{m+1} \bar{w}_i = n - m$; therefore,

$$\sum_{i=1}^{m+1} \epsilon_i = 0. \tag{4.2.10}$$

The total cost associated with integer block sizes, $J(\bar{w})$ is

$$\begin{aligned} J(\bar{w}) &= \frac{1}{2}\Big((w_1^* + \epsilon_1)^2 + w_1^* + \epsilon_1 + (w_{m+1}^* + \epsilon_{m+1})^2 + w_{m+1}^* + \epsilon_{m+1}\Big) + \\ &\quad + \sum_{j=2}^{m} \frac{1}{6}(w_j^* + \epsilon_j)^2 + \frac{1}{3}(w_1^* + \epsilon_1) \\ &= J(w^*) + (w_1^* + 1)\epsilon_1 + \frac{1}{2}\epsilon_1^2 + (w_{m+1}^* + 1)\epsilon_{m+1} + \frac{1}{2}\epsilon_{m+1}^2 + \sum_{j=2}^{m} \frac{1}{3}(w_j^* + 1)\epsilon_j + \epsilon_j^2. \end{aligned} \tag{4.2.11}$$

By expressing w_1^* and w_{m+1}^* as $w_1^* = w_{m+1}^* = \frac{w_j^*}{3} - \frac{1}{6}$ and applying equation (4.2.10), the cost (4.2.11) reduces to

$$J(\bar{w}) = J(w^*) + \frac{1}{2}(\epsilon_1^2 + \epsilon_{m+1}^2) + \frac{1}{3}\sum_{j=2}^{m} \epsilon_j^2. \tag{4.2.12}$$

It is clear from equation (4.2.12) that the increase in cost induced by rounding block sizes to integer values is of order $\mathcal{O}\Big(\frac{m}{2}\epsilon^2\Big)$ where $\epsilon = \max_i \epsilon_i$. The value of each ϵ_i, as well as the number of blocks rounded up to an integer value and the number of blocks rounded down to an integer value to satisfy (4.2.10) is dependent on $\xi - \frac{1}{2}$ and 3ξ, where ξ is the remainder of $\frac{n}{3m-1}$. Table (4.4) summarizes the relationship between ξ and the total number of blocks rounded up, $T_{b^u} = b_e^u + b_m^u$.

With the exception of $\xi = 0$ and $\xi = \frac{1}{2}$, the optimal rounding of block sizes for different values of ξ cannot be discerned from Table 4.4 alone. For each case of ξ we determine how $J(\bar{w})$ varies as a function of the number of blocks rounded up or down to the nearest integer by solving for b_m^u in terms of b_e^u in Table 4.4, combining with (4.2.12) and taking the derivative of $J(\bar{w})$ with respect to b_e^u.

Subsequently, we can determine the minimum cost solution for rounding of block sizes as a function of ξ as shown in Table 4.3. Thus, the solution to the optimal leader selection problem on an undirected, unweighted path for any length and number of

ξ	T_{b^u} = total number of blocks rounded up
$\xi = 0$	$b_e^u = 1$
$0 < \xi < \frac{1}{3}$	$b_e^u + b_m^u = \xi(3m-1)+1$
$\frac{1}{3} < \xi < \frac{1}{2}$	$b_e^u + b_m^u = \xi(3m-1)-m+2$
$\xi = \frac{1}{2}$	$b_m^u = \frac{1}{2}(m-1)$
$\frac{1}{2} < \xi < \frac{2}{3}$	$b_e^u + b_m^u = \xi(3m-1)-m$
$\frac{2}{3} < \xi < 1$	$b_e^u + b_m^u = \xi(3m-1)-2m+1$

Table 4.4: Summary of the relationship between the value of ξ =remainder$\left(\frac{n}{3m-1}\right)$ and the total number of blocks rounded up to an integer value.

leaders can be solved by calculating non-integer block sizes, (4.2.8), calculating ξ, the remainder of $\frac{n}{3m-1}$, and rounding b_e^u of the end block sizes up, b_e^d of the end block sizes down, b_m^u of the middle block sizes up, and b_m^d of the end blocks down, as summarized in Table 4.3 such that integer block sizes $\bar{w}_j$ are obtained. The optimal leader locations are then

$$s_i = \sum_{j=1}^{i} (\bar{w}_j + 1). \tag{4.2.13}$$

□

We see that depending on the values of m and n, there well often be multiple optimal solutions. For example, the optimal solutions for two leaders on a 10-node path will be nodes 3 and 9 or nodes 2 and 8. A larger set of optimal solutions can occur when $m > 2$ and there are multiple middle blocks. In this case it is irrelevant which blocks are rounded up and which are rounded down, so long as Table 4.3 holds.

4.2.2 Two noise-Corrupted Leaders on a Path

To investigate the role of finite k we computed the optimal noise-corrupted leader set for the path graph of order $n = 51$. For $k = 2$ and higher values, the solution

corresponds to the optimal solution in the noise-free case given by Theorem 4, i.e., $S^* = \{11, 41\}$. In the case of $k = 0.0001$, the optimal solution is $S^* = \{13, 39\}$, i.e., the optimal noise-corrupted leaders are a little closer to the center of the path. The trend persists for larger n. For example, for a path graph of order $n = 101$, for $k = 2$ and higher values, the solution corresponds to the optimal solution in the noise-free case given by Theorem 4, i.e., $S^* = \{21, 81\}$, and in the case of $k = 0.0001$, the optimal solution is $S^* = \{25, 77\}$.

4.3 Optimal Leader Selection in Unweighted Cycle Graphs

The third canonical graph topology we study is the unweighted cycle graph. Cycle graphs have the property that information centrality is equivalent for all nodes and there are exactly two paths between any pair of nodes. With these constraints we are able to develop exact solutions to the leader selection problem for robustness. We prove the optimal locations of any m noise-free leaders and $m = 2$ noise-corrupted leaders.

4.3.1 Optimal selection of m Noise-Free Leaders in an Unweighted Cycle Graph

Theorem 5 (Optimal noise-free leader set on a cycle graph)**.** *Let $\mathcal{G}$ be an undirected, unweighted cycle graph of order n. Let $m < n$ such that $p = n/m$ is an integer. Let S be a set of m noise-free leaders. Then, an optimal leader set S^* is any set S where the leaders are uniformly distributed around the cycle, i.e., the geodesic distance between any leader and each of the other two closest leaders is $\delta_{s_a,s_b} = p$.*

Proof. We begin by assuming m nodes on the cycle have been selected as leaders and

let $M = L + K$ where K is a matrix with a value of k in the entries along the main diagonal corresponding to the leader nodes and zeros elsewhere. We partition M in the usual way. Similarly to Section 4.2, M_F can be written as a block diagonal matrix where each block corresponds to a set of connected follower nodes between two leader nodes. Each block, M_{F_i} will itself be a tridiagonal matrix of the form

$$M_{F_i} = \begin{bmatrix} 2 & -1 & & 0 \\ -1 & 2 & & \\ 0 & & \ddots & -1 \\ 0 & & -1 & 2 \end{bmatrix}. \tag{4.3.1}$$

In the case where there is one follower node in between two leader nodes the corresponding diagonal block in M_F will be one element with an entry of 2.

As before, total system error for noise-free leaders will be proportional to the trace of M_F^{-1}, which here is equivalent to the total sum of eigenvalues of each $M_{F_i}^{-1}$. The eigenvalues of each $M_{F_i}^{-1}$ and the trace of $M_{F_i}^{-1}$ are given by equations (4.2.5) and (4.2.6), respectively.

Therefore, minimizing the total sum of eigenvalues is equivalent to minimizing the sum over i of $\mathrm{tr}(M_{F_i}^{-1})$. It follows that the minimum is achieved when $w_1 = w_2 = w_3 = ...$, or in other words when the dimension of each block is the same. This corresponds to the leaders being evenly distributed around the cycle with shortest distances between leaders equal to $\delta_{s_1,s_2} = \frac{n}{2}$. □

4.3.2 Two noise-Corrupted Leaders on a Cycle

To again investigate the effect of finite k, we study the case of two noise-corrupted leaders on a cycle. We prove in the case of the cycle graph, where every node has the same information centrality, that the optimal two noise-corrupted leaders correspond to an antipodal pair of nodes, i.e., a pair with maximal resistance distance. This is

the same solution as in the case of noise-free leaders on the cycle.

Corollary 8 (Optimal noise-corrupted leader set on a cycle, $m = 2$)**.** *Let $\mathcal{G}$ be an undirected, unweighted cycle graph of order n where n is even. Let $S_2 = \{s_1, s_2\}$ be a set of two noise-corrupted leaders ($k < 0$). The optimal leader set S^* is any two nodes with maximal resistance distance $r_{s_1,s_2} = \frac{n}{4}$, which corresponds to geodesic distance $\delta_{s_1,s_2} = \frac{n}{2}$ and antipodal nodes.*

Proof. For a circulant graph, $L^+_{s_1,s_1} = L^+_{s_2,s_2} = L^+_{s,s}$ and thus $\gamma_{s_1,s_2} = \frac{1}{4}\sum_{i=1}^{n}(r_{i,s_1} - r_{i,s_2})^2$. The k-dependent joint centrality ρ_{kS_2} (3.3.9) simplifies to

$$\rho_{kS_2} = \; n\Big(\frac{K_f}{n} + \frac{n{L^+_{s,s}}^2 - n{L^+_{s_1,s_2}}^2 - \sum_{i=1}^{n}(r_{i,s_1} - r_{i,s_2})^2}{4r_{s_1,s_2}}\Big)^{-1}. \tag{4.3.2}$$

By applying (2.3.7) and re-arranging terms we have

$$\rho_{kS_2} = \frac{n^2}{4}\Big(\frac{K_f}{n^2} + \frac{2}{k} + 4L^+_{s,s} - r_{s_1,s_2} - \frac{k}{4}\frac{\sum_{i=1}^{n}(r_{i,s_1} - r_{i,s_2})^2}{2 + kr_{s_1,s_2}}\Big)^{-1}. \tag{4.3.3}$$

Using the electric circuit analog of resistance distance and applying Kirchhoff's laws, the resistance distance between any two nodes in a cycle can be written as

$$\frac{1}{r_{i,j}} = \frac{1}{\delta_{i,j}} + \frac{1}{n - \delta_{i,j}}, \tag{4.3.4}$$

where $\delta_{i,j}$ is the geodesic distance between nodes i and j. The maximum resistance distance is $r_{i,j} = \frac{n}{4}$, which is obtained between two nodes with $\delta_{i,j} = \frac{n}{2}$.

Simplifying the $\sum_{i=1}^{n}(r^+_{i,s_1} - r^+_{i,s_2})^2$ term of (4.3.3) by inserting (4.3.4) gives

$$\begin{aligned}\sum_{i=1}^{n}(r_{i,s_1} - r_{i,s_2})^2 &= \sum_{i=1}^{n}\Big(\delta_{i,s_1} - \delta_{i,s_2} + \frac{\delta^2_{i,s_2} - \delta^2_{i,s_1}}{n}\Big)^2 \\ &= \frac{\delta_{s_1,s_2}(\delta_{s_1,s_2} - n)(\delta^2_{s_1,s_2} - n\delta_{s_1,s_2} - 2)}{3n}.\end{aligned} \tag{4.3.5}$$

Substituting (4.3.5) into (4.3.3) results in

$$\rho_{kS_2} = \frac{n^2}{4}\Big(\frac{K_f}{n^2} + \frac{2}{k} + 4L^+_{s,s} - \frac{\delta_{s_1,s_2}(n-\delta_{s_1,s_2})}{n} - \frac{k\delta_{s_1,s_2}(\delta_{s_1,s_2}-n)(\delta^2_{s_1,s_2}-n\delta_{s_1,s_2}-2)}{6n(2n+k\delta_{s_1,s_2}(n-\delta_{s_1,s_2}))}\Big)^{-1}. \tag{4.3.6}$$

To determine how ρ_{kS_2} changes as a function of δ_{s_1,s_2}, we take the partial derivative of (4.3.6) with respect to δ_{s_1,s_2} to give

$$\frac{\partial \rho^{-1}_{kS_2}}{\partial \delta_{s_1,s_2}} = -\frac{1}{4}(n-2\delta_{s_1,s_2}) - \frac{nk[2(-\delta_{s_1,s_2}+\delta^3_{s_1,s_2})+(1-3\delta^2_{s_1,s_2})n+\delta_{s_1,s_2}n^2]}{3(2n+\delta_{s_1,s_2}k(-\delta_{s_1,s_2}+n))^2} - \frac{k^2[-2\delta^5_{s_1,s_2}+5\delta^4_{s_1,s_2}n-4\delta^3_{s_1,s_2}n^2+\delta^2_{s_1,s_2}n^3]}{12(2n+\delta_{s_1,s_2}k(-\delta_{s_1,s_2}+n))^2}. \tag{4.3.7}$$

Since $\delta_{s_1,s_2} \leq \frac{n}{2}$, the first term of (4.3.7) will always be nonpositive. Additionally, it can be shown algebraically that for $n > 3$ the two bracketed expressions in the second and third terms will be greater than zero. Therefore $\rho^{-1}_{kS_2}$ decreases as δ_{s_1,s_2} increases, reaching its minimum at the maximal value of $\delta_{s_1,s_2} = \frac{n}{2}$, corresponding to $r_{s_1,s_2} = \frac{n}{4}$. □

The results from Theorem 5 and 8 provide an interesting comparison with the results from noise-free and noise-corrupted pairwise leader selection on path graphs, Theorem 4 and Section 4.2.2, respectively. The removal of one edge in a cycle graph results in a path graph, and we have shown that for large n the removal of this edge causes the optimal leaders to change location from 25% and 75% about the cycle to 20% and 80% along the path. However, when we consider noise-corrupted leaders we find that the removal of an edge in the cycle causes a smaller change in the location of optimal leaders, that is the optimal noise-corrupted leaders on a path approach the optimal pair for the cycle.

Chapter 5

Applications of Joint Centrality

As mentioned in Chapter 1, the concept of centrality for a group of nodes has a wide range of applications. In this chapter we explore the use of joint centrality as a centrality measure for a set of nodes through illustrative examples. We begin by demonstrating that pairs of genes in *Saccharomyces Cerevisiae*, otherwise known as bakers yeast, which are pairwise essential to the organism, tend to have higher two-node joint centrality relative to non-essential pairs of genes. Then, we apply joint centrality to graph clustering and provide two examples. Section 5.1 was previously published in Fitch and Leonard [26].

5.1 Joint Centrality and Synthetic Lethality in Saccharomyces cerevisiae

To further investigate joint centrality of a set of nodes, we apply it in the analysis of synthetically lethal (SL) genes of the probabilistic functional gene network of *Saccharomyces Cerevisiae*, also known as baker's yeast. *S. Cerevisiae* has served as a platform for studying genetics of human diseases and is therefore an important model for biological studies [51]. A probabilistic functional gene network is one in

which nodes in the network represent genes and edges between pairs of nodes represent a measure of probability that two genes are involved in the same function. In the probabilistic functional gene network of *S. Cerevisiae* from [51], edge weights are determined by using Bayesian statistics to calculate the log likelihood of two genes being functionally coupled. If the log likelihood between two nodes is zero, in other words there is no edge between them, then it is implied that the likelihood of two genes being involved in the same function(s) is at most random expectation. Additional details can be found in [51]. Here, we focus on instances of synthetic lethality, which occur when the deletion of two genes (A and B) is lethal to the organism and the deletion of A alone or B alone is not lethal.

Using the probabilistic functional gene network of *S. Cerevisiae* from [51], (5808 genes with 362,421 edges that represent functional couplings) we calculated the two-node joint centrality for every pair of genes in the network. Then we applied experimental interaction data from the BioGrid database to identify SL pairs of nodes [96]. Figure 5.1 shows the probability distribution function of two-node joint centrality for all pairs of genes (blue) against the probability distribution function of two-node joint centrality for SL pairs of genes (red). The distributions were constructed by fitting non-parametric distributions with a normal kernel function to normalized histograms of joint centrality calculations for all node pairs and for all SL node pairs.

A clear distinction between the two distributions in Figure 5.1 is apparent. The distribution of two-node joint centralities for SL node pairs is more highly skewed towards high values of joint centrality than the distribution of two-node joint centralities for all node pairs.

We note that SL pairs of nodes are also distinguishable from all other pairs due to their having a higher average degree. This is expected, however, as there is likely a research bias towards testing high degree nodes for synthetic lethality (the set of SL pairs is not necessarily the complete set but rather the set that has been identified

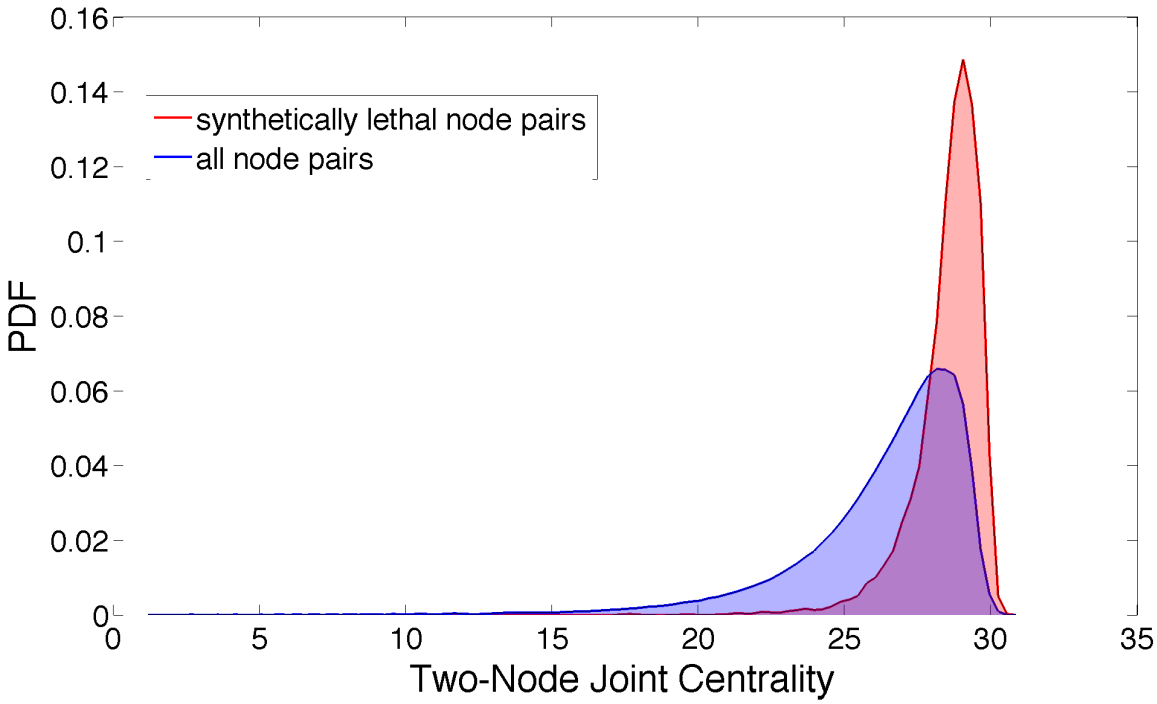

Figure 5.1: Distribution of two-node joint centrality for every node pair (blue) in the functional gene network of *S. Cerevisiae* and distribution of two-node joint centrality of synthetically lethal node pairs (red).

thus far). Accordingly, we do not suggest that joint centrality is the only way to predict possible SL pairs. Instead, we suggest that two-node joint centrality provides a natural measure for predicting SL pairs, because it takes into account the joint influence of a pair of nodes on the entire network. In contrast, a measure of pairwise average degree only considers independent, local interactions.

5.2 Joint Centrality and Graph Clustering

In [93], the authors approximated the solution to the problem of leader selection for fast convergence to consensus by applying an algorithm that approximates the solution to the metric $\hat{k}$-center problem. The solution to the metric $\hat{k}$-center problem is the set of $\hat{k}$ leaders that minimizes the maximum distance from a follower node to the closest leader. Inspired by [93], here, we apply the solution to the leader selection

problem for robustness to graph clustering, an application which uses algorithms similar to metric $\hat{k}$-center. In graph clustering, the objective is to assign each node in the graph to a group, or cluster, in a way such that all nodes in a given cluster are similar or close in terms of a graph metric. There are many variants of the clustering problem, and we refer to [91] for a more comprehensive survey on the topic.

One of the most common clustering algorithms currently used in practice is the $\hat{k}$-means algorithm. This is an iterative algorithm in which $\hat{k}$ initial "means", or *seed nodes*, are randomly selected and then $\hat{k}$ clusters are constructed by assigning each node to the cluster associated with the nearest mean in euclidean distance. Then the centroid of each cluster becomes a mean for the next iteration and the process is repeated until the clusters stabilize. Variations on $\hat{k}$-means, including $\hat{k}$-medoids, employ alternative similarity or distance metrics as opposed to strictly l_2 distances [44]. $\hat{k}$-means does not account for overlapping clusters, that is, a node cannot be assigned to more than one cluster. As a comment on notation, $\hat{k}$-means is typically written as k-means, though here we use $\hat{k}$ to denote number of clusters to distinguish from k which represents weighting on the leader's distance from external signal μ, as used in earlier chapters of this thesis.

We have previously shown that node sets with high joint centrality trade off high individual centrality with distribution over the network graph. Therefore, we suggest that instead of iteratively assigning clusters and computing centroids, one can simply *select the $\hat{k}$ nodes comprising the set with the highest $\hat{k}$-node joint centrality, $\mathcal{S}^* = \{s_1, ..., s_k\}$, as seed nodes.* Subsequently, nodes can be assigned to clusters according to a pre-decided distance measure or similarity measure, which we will clarify later, to the nodes in this set. That is, to assign a node to a non-overlapping cluster, we calculate the similarity of a node, i, to each node in $\mathcal{S}^*$ and assign i to the cluster centered at s_j where the similarity between i and s_j is greater than the similarity between i and any other node in $\mathcal{S}^*$. For overlapping cluster assignments, node i is a

partial member of many clusters, where the percentage of assignment to each cluster is determined by the relative similarity to each nodes in $\mathcal{S}^*$. For example, let $f(i, s_j)$ be a similarity measure, then if $\mathcal{S}^* = \}s_1, s_2\}$, $f(i, s_1) = 2$, and $f(i, s_2) = 6$, then node i is 25% in the cluster centered at s_1 and 75% in the cluster centered at s_2. In the examples that follow, we take the value kM^{-1}_{i,s_j} as a measure of similarity of node i to leader s_j. As demonstrated in Equation (3.1.8), M^{-1}_{i,s_j} is steady state covariance of node i and node s_j, scaled by $\frac{2}{\sigma}$. Therefore, it is an intuitive choice of similarity measure to employ. One could alternatively cluster nodes according to resistance distance or biharmonic distance from the most joint central set, though we do not explore those directions here.

In both $\hat{k}$-means and joint centrality based methods outlined above, there is little guarantee on performance. It can be shown that the clusters assigned by $\hat{k}$-means have a non-increasing measure of net distance of each node to the closest center, though it is possible the algorithm will converge to locally optimal selections of cluster centroids. Alternatively, for the joint centrality based method we have cluster centroids that are by default well distributed with respect to one another, well connected to nearby nodes, and jointly provide an optimal distribution of information over the network.

To test the performance of the proposed clustering method, I downloaded my Facebook social network using the Facebook graph API. The data was comprised of 468 entries, each with the name of a friend, their associated node ID and the node IDs of other friends to which that node is connected. The objective was to determine if the joint centrality based clustering method could appropriately cluster my friends according to geographic location. Thus far, I have primarily lived in Hamburg NY, Syracuse NY and Princeton NJ; therefore, the social network was clustered into three groups. The set with the highest three-node joint centrality (excluding myself) was calculated and found to be the set comprised of a friend from Hamburg, a friend from Syracuse and a friend from Princeton. From there, the covariance of each node in the

set with highest three-node joint centrality was calculated with each remaining node in the network. To illustrate overlapping clusters (i.e., a friend could be affiliated with both Princeton and Hamburg), nodes were colored using an rgb colormap where the percentage red of a node is the covariance of that node with the center node of the Hamburg cluster, the percentage green of a node is the covariance of that node with the center node of the Princeton cluster, and the percentage blue of a node is the covariance of that node with the center node of the Syracuse cluster. The resulting coloring can be seen in Figure 5.2.

There are a few interesting observations to make. First, there is a relative gradient in color from red (Hamburg cluster) to blue (Syracuse cluster), while the green cluster (Princeton) has clearer boundaries. This is precisely what we would expect to see because Hamburg and Syracuse are geographically close and there are a number of friends who have ties to both locations. Subsequently, the friends with ties to both locations have similar covariances to the centers of the Syracuse and Hamburg clusters. Conversely, there is very little overlap of friends from the Princeton community with friends in the Syracuse or Hamburg communities. Therefore, nodes in the Princeton cluster have high covariances with the center of the Princeton cluster and very low covariances with the center of the other two clusters. Manual verification of cluster assignments demonstrated a high level of accuracy.

As a second example, we examine a network of 105 political books on Amazon. Each book is classified as conservative (red), liberal (green), or neutral (blue), and is represented by a node in the network. There is an edge between two nodes if those two books are frequently purchased together.

For this example, we look at a clustering with non-overlapping cluster assignments, that is, if a node is in group A then it is not in group B. This allows for direct comparison with $\hat{k}$-means clustering. The political book network is shown in Figure 5.3.

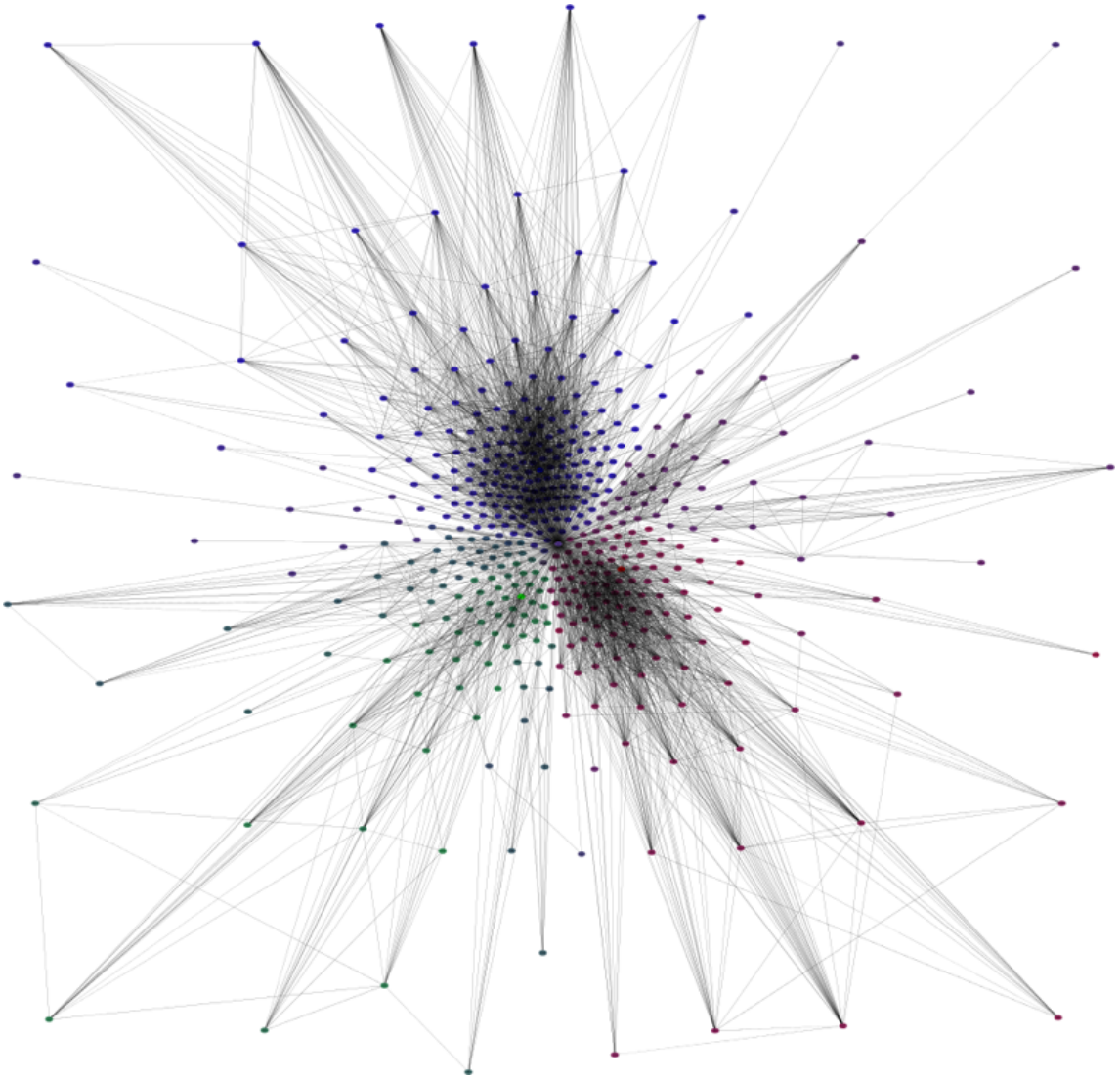

Figure 5.2: Clustering on a Facebook social network graph. Nodes colored according to covariance with the set with highest three-node joint centrality. Red signifies cluster assignment to Hamburg, NY cluster, blue signifies cluster assignment to Syracuse, NY cluster, and green signifies cluster assignment to Princeton, NJ cluster.

To test the applicability of the joint centrality based clustering method we sought to establish three clusters of political orientation solely from the network topology while assuming no knowledge of the ground truth labeling. Again, this was achieved through calculation of the set with the highest three-node joint centrality and determining the covariance of each node in the network with each node in the leader set. We assign a node to the cluster of the leader with which it has the highest covariance.

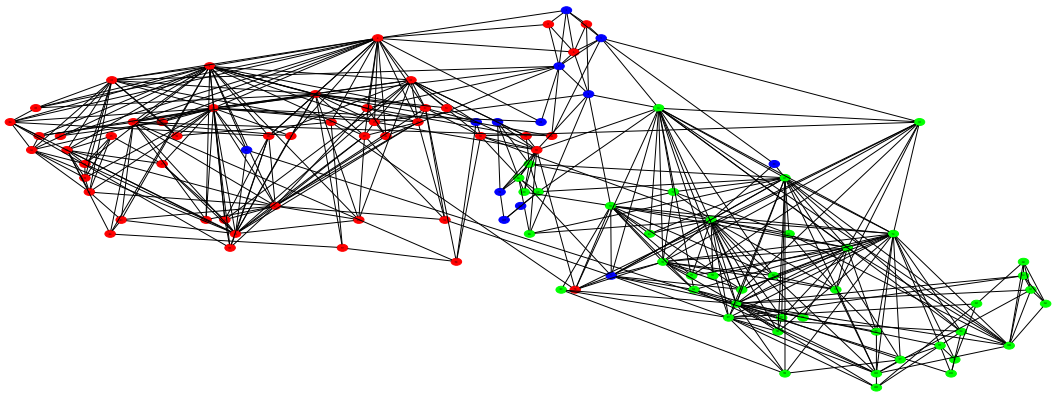

Figure 5.3: Network of political books commonly purchased together on Amazon. Red nodes are conservative books, green nodes are liberal books and blue nodes are neutral books.

The resulting cluster assignments are shown in Figure 5.4. Furthermore, Figure 5.5

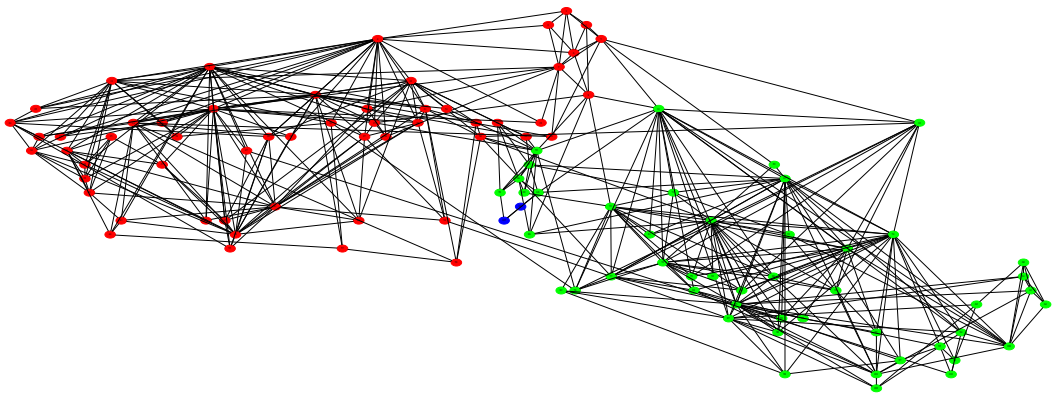

Figure 5.4: Cluster classifications of political books commonly purchased together on Amazon via the joint centrality based method.

shows the clustering assignments determined using $\hat{k}$-means.

We find that the joint centrality based clustering method and $\hat{k}$-means clustering perform similarly. Both have 13 errors and are, in general, unable to correctly identify most of the neutral political books, while having a high level of accuracy distinguishing between the conservative and liberal books. This is expected as the

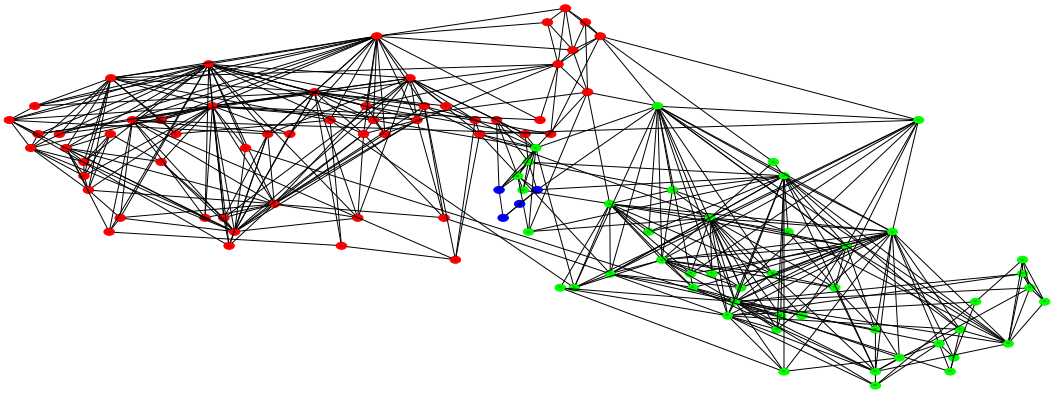

Figure 5.5: Cluster classifications of political books commonly purchased together on Amazon via $\hat{k}$-means.

neutral books are not clustered and in general cannot be isolated purely from knowledge of the graph topology. Of the 13 errors in both methods, 12 are equivalent. The difference in cluster classifications between the two methods is that the joint centrality misclassifies one neutral book as liberal, which $\hat{k}$-means correctly classifies, and $\hat{k}$-means misclassifies one liberal book as neutral, which the joint centrality method correctly classifies.

Chapter 6

Optimal Leader Selection for Controllability*

In this chapter we study the optimal leader selection problem for controllability. Specifically, we look at how controllable a system is, as a function of where in the network leaders are placed. In previous chapters, every node was receiving an input of white noise and we sought to select leaders to diminish the influence of that input in steady-state. Conversely, in this chapter only the leader node receive an input, e.g., a control signal, and we seek to select leaders to maximize average controllability or volume of the reachable subspace. Through our analysis, we prove that the optimal leaders for average controllability are the least information central nodes, revealing an interesting contrast to the optimal leader sets for robustness. We also show that the optimal leaders for reachable subspace volume are dependent on the left eigenvectors of the graph Laplacian.

*This chapter is adapted from Fitch and Leonard [27] with most of the text taken verbatim.

6.1 Optimal leader selection problem for controllability

To investigate optimal leader selection for controllability, we start by assuming that a set S of $m > 0$ nodes are leaders, which act as control inputs to the network system. The network dynamics evolve according to 2.4.1 where $\mathbf{u} \in \mathbb{R}^m$ is the control vector and B has m columns, corresponding to the standard basis vectors $\mathbf{e}_i$ for $i \in S$.

Controllability of a consensus network can be defined by restricting dynamics to the disagreement subspace (orthogonal complement of the one-dimensional agreement subspace spanned by the vector of all ones $\mathbf{1}_n \in \mathbb{R}^n$) [73]. Thus, we consider the reduced Laplacian $\bar{L} = QLQ^T$, and controllability Gramian

$$W_C = \int_{t_0}^{\infty} e^{-\bar{L}\tau} QBB^T Q^T e^{-\bar{L}\tau} d\tau \tag{6.1.1}$$

where the rows of $Q \in \mathbb{R}^{(n-1)\times n}$ form an orthonormal basis for $\mathbf{1}_n^{\perp}$, and it follows that $\bar{L}$ is stable and invertible. W_C is also the solution to the Lyapunov equation

$$\bar{L}W_C + W_C\bar{L}^T = QBB^TQ^T. \tag{6.1.2}$$

We define four cases of the *optimal leader selection problem for controllability* as follows.

Definition 7 (Optimal leader selection problem for controllability)**.** *Given $m > 0$ and undirected, connected graph $\mathcal{G}$, find a set of m leaders S_C^* over all possible sets S of m leaders that optimizes a controllability metric $\alpha(W_C)$ for the leader-follower network dynamics (2.4.1), where $\alpha(W_C)$ is determined by one of the four performance measures:*

(a) *Average controllability: $\alpha(W_C) = tr(W_C)$*

(b) *Reachable volume: $\alpha(W_C) = ld(W_C)$*

(c) *Average control energy:* $\alpha(W_C) = tr(W_C^+)$

(d) *Worst case input energy:* $\alpha(W_C) = \lambda_{min}(W_C)$.

6.1.1 Comparison of W_C and W_R

We observe that even though W_C and W_R are quite similar in formulation, the difference between choosing non-zero elements of B and non-zero diagonal elements of K is quite significant due to the fact that K influences the value of W_R as part of an element in an exponential function, whereas BB^T is simply multiplied by the state transition matrix. Thus, non-zero elements of B will have a different effect on the value of $\text{tr}(W_C)$ than non-zero diagonal elements of K will have on the value of $\text{tr}(W_R)$. Furthermore, to maximize average controllability one wants to *maximize* $\text{tr}(W_C)$; however, to maximize robustness one wants to *minimize* $\text{tr}(W_R)$. We demonstrate and discuss the implications of these observations on the resulting optimal leader sets for each problem in Sections 6.3 and 6.4.

6.2 Control energy centralities

The authors of [98] proved that the trace of the controllability Gramian, $\text{tr}(W_C)$, is a modular set function. The implication of a modular set function is that each element of a subset independently contributes to the value of the function. Solving an optimization problem with a modular cost function is straightforward, as the total cost is the sum of each element's independent contribution to the cost function.

Summers, et al. proved that the trace of the (pseudo-) inverse of the controllability Gramian, $\text{tr}({W_C}^+)$, and the log determinant of the controllability Gramian, $\text{ld}(W_C)$, are both submodular functions of the leader set. In [16], it was shown that the trace of the robustness Gramian, $\text{tr}(W_R)$, is also a submodular function of the leader set. A submodular set function has the property of diminishing returns, that

is the addition of an element to a larger set has a smaller contribution than the addition of an element to a smaller set. Therefore, each element of a subset does not contribute independently as in modular set functions. Full solutions to opimization problems with nondecreasing submodular set functions are NP-hard, although greedy algorithms can provide a solution within a provable bound from the optimal solution [68]. From this, we can predict that a closed-form solution for maximizing $\mathrm{tr}(W_C)$ is obtainable, while optimizing $\mathrm{tr}({W_C}^+)$, $\mathrm{ld}(W_C)$, and $\mathrm{tr}(W_R)$ are each combinatorially difficult problems.

The authors of [98] did not provide relationships between these centrality definitions and well defined measures of the network graph nor insight more generally on how a node's location in a network relates to the value of its three control energy centralities. For single leaders, the solutions to maximizing $\mathrm{tr}(W_C)$, and $\mathrm{ld}(W_C)$ and minimizing $\mathrm{tr}({W_C}^+)$ will align with the nodes that maximize the respective control energy centralities. In the following section, we make explicit the relationship between average controllability centrality C_{AC}, volumetric control energy centrality C_{VCE} and properties of the graph Laplacian L.

6.3 Optimal leader selection results

6.3.1 Optimal leader selection for average controllability

The following theorem provides the optimal leader set S_C^* for Problem 1(a) in terms of properties of the network graph.

Theorem 6. *Consider the dynamics (2.4.1) with the undirected, connected graph $\mathcal{G}$ of order n. Let the set S be a set of m leaders. Then average controllability depends on the inverse of the information centralities of nodes in S, and the optimal leader set S_C^* is composed of the m nodes with smallest information centrality.*

Proof. Consider the controllability Gramian W_C given by (6.1.1). We note that $\bar{L}$ has the same eigenvalues as L except for the zero eigenvalue, which we index by n. Let the diagonal matrix of eigenvalues and the matrix of right eigenvectors for $\bar{L}$ and L be $\bar{\Lambda}$, $\bar{V}$ and Λ, V, respectively. Then,

$$W_C = \bar{V}\Big(\int_{t_0}^{\infty} e^{-\bar{\Lambda}\tau}\bar{V}^T Q B B^T Q^T \bar{V} e^{-\bar{\Lambda}\tau} d\tau\Big)\bar{V}^T d\tau. \tag{6.3.1}$$

Consider the case of a single controller node, indexed by l. Then, the vector B will have a single non-zero entry, b.

Since $Q^T\bar{V}$ is equivalent to the first $n-1$ columns of V, we can represent the product in the integral of (6.3.1) as a function of the eigenvalues and eigenvectors of L. Then $W_C = b^2\bar{V}G\bar{V}^T$, where $G \in \mathbb{R}^{(n-1)\times(n-1)}$ has entries

$$g_{i,j} = \int_0^{\infty} e^{-\lambda_i\tau-\lambda_j\tau} v_{l,i}v_{l,j}d\tau = \frac{1}{\lambda_i+\lambda_j}v_{l,i}v_{l,j}. \tag{6.3.2}$$

Recall that we are interested in maximizing the trace of W_C and that trace is invariant under cyclic permutations. Thus

$$\begin{aligned}\mathrm{tr}(W_C) &= b^2\mathrm{tr}(\bar{V}G\bar{V}^T) = b^2\mathrm{tr}(G\bar{V}^T\bar{V}) = b^2\mathrm{tr}(G)\\ &= \sum_i^{n-1}\frac{b^2}{2\lambda_i}v_{l,i}^2 = \frac{b^2}{2}L^+_{l,l}\\ &= \frac{b^2}{2}\left(\frac{1}{c_l} - \frac{K_f}{n^2}\right)\end{aligned} \tag{6.3.3}$$

where c_l is the information centrality of node l. From (6.3.3), for a single controller node, $\mathrm{tr}(W_C)$ is maximized by the node with the smallest information centrality.

Due to the modularity property, $\mathrm{tr}(W_C)$ with m leader nodes will be minimized when the set of leaders S_C^* consists of the m nodes with smallest information centrality.

□

Corollary 9. *Consider the dynamics (2.4.1) with the undirected, connected graph $\mathcal{G}$ of order n. Then*

$$C_{AC}(i) = \mathrm{tr}(W_{c_i}) = \frac{1}{2}\Big(\frac{1}{c_i} - \frac{K_f}{n^2}\Big).$$

We have found that the average controllability centrality defined by [98] is in fact inversely related to a pre-existing graph measure: information centrality. Corollary 9 implies that the more information central is a leader node the lower will be the average controllability.

6.3.2 Optimal leader selection for reachable subspace volume

The following theorem provides the optimal leader set S_C^* for Problem 1(b) in terms of properties of the network graph.

Theorem 7. *Consider the dynamics (2.4.1) with the undirected, connected graph $\mathcal{G}$ of order n. Let the set S be a set of m leaders. Then reachable volume can be written as*

$$\mathrm{ld}(W_C) \propto \log\Big(\prod_{j=1}^{n-1}\Big(\sum_{i\in S} v_{i,j}^2\Big)\Big)$$

where $v_{j,i}$ is the ith entry in the jth right eigenvector of L. The optimal leader set is $S_C^* = \arg\max_S(\prod_{j=1}^{n-1}\Big(\sum_{i\in S} v_{i,j}^2\Big)$.

Proof. Using (6.3.1), the determinant of W_C is

$$\det(W_C) = b^2\det(\bar{V}G\bar{V}) = b^2\det(\bar{V})\det(G)\det(\bar{V}). \tag{6.3.4}$$

From (6.3.2), $G = \tilde{V}\Gamma\tilde{V}$ where $\tilde{V}, \Gamma \in \mathbb{R}^{(n-1)\times(n-1)}$. $\tilde{V}$ is a diagonal matrix with

$\tilde{V}_{i,i} = v_{l,i}$ and the entries of Γ are $\Gamma_{i,j} = \frac{1}{\lambda_i + \lambda_j}$. Plugging in to (6.3.4) gives

$$\det(W_C) = b^2 \det(\bar{V}) \det(\tilde{V}) \det(\Gamma) \det(\tilde{V}) \det(\bar{V}). \tag{6.3.5}$$

The only term in (6.3.5) that depends on the choice of leader node is $\det(\tilde{V})^2$. Since $\tilde{V}$ is diagonal, its determinant is the product of its diagonal entries. Thus for a single leader l

$$\det(W_C) \propto \log \prod_{j=1}^{n-1} v_{l,j}^2.$$

In the case of m leaders, $\tilde{V}_{j,j} = \sum_{i \in S_c} v_{i,j}$ and

$$\det(W_C) \propto \log \prod_{j=1}^{n-1} \Big(\sum_{i \in S_c} v_{i,j}^2 \Big).$$

It follows that the set S_C^* of m leaders that maximizes $\prod_{j=1}^{n-1} \Big(\sum_{i \in S} v_{i,j}^2 \Big)$ maximizes $\mathrm{ld}(W_C)$, the volume of the controllable subspace reachable with one unit of input. $\square$

Corollary 10. *Consider the dynamics (2.4.1) with the undirected, connected graph $\mathcal{G}$ of order n. Then*

$$C_{VCE}(i) = \log \Big(Y \prod_{j=1}^{n-1} v_{i,j}^2 \Big)$$

where Y is a constant and $Y = \det(\bar{V} \Gamma \bar{V})$.

Theorem 7 and Corollary 10 show the dependence of $\mathrm{ld}(W_C)$ and the volumetric control energy centrality on the leader nodes' entries of the eigenvectors of the graph Laplacian.

We point out that ranking nodes by volumetric control energy centrality computed as $\prod_{j=1}^{n-1} v_{i,j}^2$ for each node i is significantly less computationally intensive than ranking

through a calculation of the controllability Gramian and its determinant for each node i.

6.4 Controllability versus robustness trade-offs

Combining the results from Section 6.3.1 and Chapter 3, we see that a fundamental trade-off appears between selecting leaders for average controllability and selecting leaders for robustness. To maximize average controllability one simply selects the least information central nodes as leaders. Since the average controllability problem is modular, the solution does not depend on the relative positions of the leader nodes. Conversely, the problems of leader selection to maximize the volume reachable with one unit of energy and the problem of leader selection for robustness are both submodular; therefore the relative positions of nodes in the leader sets play a role in the optimal solutions. To maximize robustness the leader set must balance high information centrality of individual leader nodes with distribution of leader nodes over the network. Nodes with low information centralities lead to a leader set with high average controllability but often with low robustness. Therefore, both robustness and average controllability cannot be optimized by the same leader set in general graphs where all nodes do not have equivalent information centralities.

To provide intuition behind these contrasting solutions we discuss the single leader case. For the controllability problem, we have a time varying control signal modulating the state of the leader node and we would like to control the network to any configuration in the controllable subspace. Variations in the state of the leader node will have a large immediate effect on the states of neighboring nodes, while the effect on the states of distant nodes will be less significant. In other words, there is variability in the effect of the control signal depending on distance from the leader node. Since information centrality is calculated through a harmonic mean, nodes with a

larger range of distances to every other node in the network will have lower information centralities. Hence, the nodes with low information centralities will have more unique and independent paths to every other node in the network. This means that it is likely easier to independently control the states of other nodes in the network.

6.4.1 Optimal leader selection in a cycle graph

Consider a cycle graph with n nodes. For $m = 2$ leaders, the value of average controllability over the controllable subspace will be the same no matter which m nodes are chosen because all nodes have equivalent information centralities in a cycle graph. For the robustness problem, in [26] it was proven that the optimal leader selection for robustness in a cycle graph corresponds to m nodes evenly distributed about the cycle. The optimal leader set for reachable volume corresponds to any two nodes separated by a single node. Therefore the optimal leader set for reachable volume and the optimal leader set for robustness are in direct tension and cannot be simulatneously selected.

6.4.2 Optimal leader selection in a random graph

Consider a random network with $n = 100$ nodes as shown in Figure 6.1. We have highlighted the sets of $m = 3$ nodes that will optimize the leader selection Problems 1(a-c), and Problem 2. The leader set that maximizes average controllability is colored in green, the leader set that maximizes reachable volume is in orange, the leader set that minimizes average control energy is in red, and the leader set that maximizes robustness is in blue. The direct tension between optimizing the leader set for controllability metrics and optimizing the leader set for robustness is visually apparent. The optimal leader sets for maximizing average controllability, maximizing reachable volume, and minimizing average input energy are on the periphery of the network, and it is interesting to note that one node in particular is a member of all

three sets. The nodes in the optimal leader set for robustness are in central, but distributed, locations. Furthermore, nodes in the optimal leader set for robustness are less susceptible to becoming disconnected from the network through edge failures. Again, it is clear that a trade-off must be made if both controllability and robustness are important characteristics in a leader-follower multi-agent system.

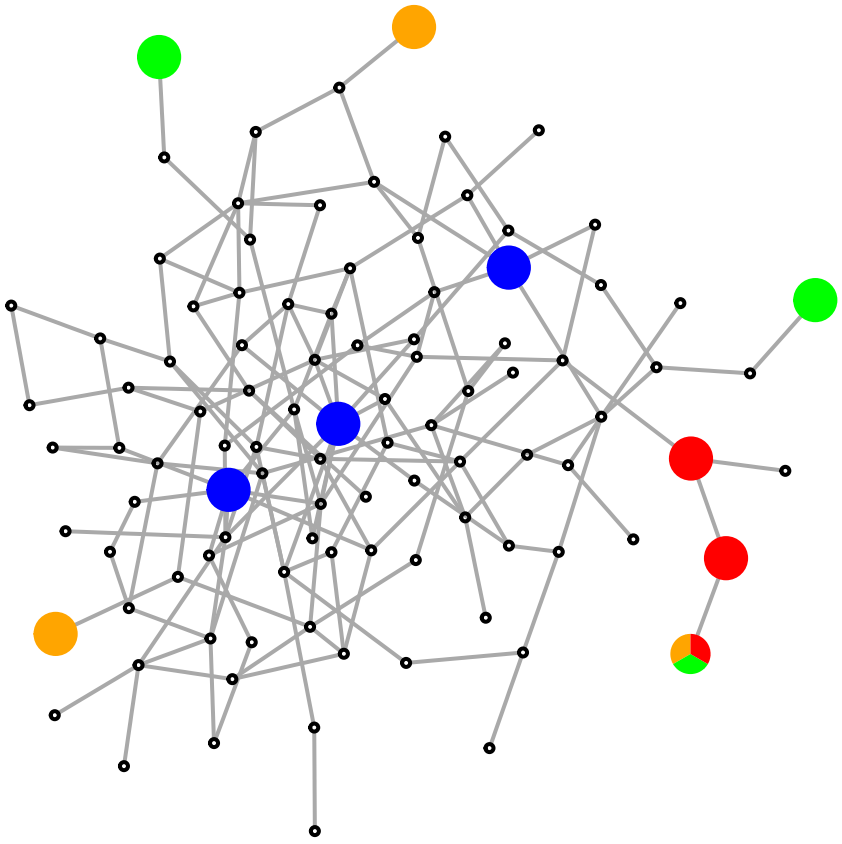

Figure 6.1: Random undirected graph with $n = 100$ nodes highlighting optimal leader sets of $m = 3$ nodes for average controllability (green), reachable volume (orange), average control energy (red) and robustness (blue).

The controllability Gramian measures are metrics over the controllable subspace and none guarantee full controllability. Therefore it is possible that a leader set that maximizes $\mathrm{tr}(W_C)$ corresponds to few highly controllable modes or many weakly controllable nodes. This may or may not be of significance depending on the application. Consideration of the rank of the controllability Gramian in addition to the measures

discussed here may be of importance.

6.5 Final Remarks

In this chapter we examined and provided new insights on the optimal leader selection problem for leader-follower multi-agent systems. We proved that the optimal leader set for average controllability consists of the least information central nodes in the network. We proved the relationship between the optimal leader set for reachable volume and entries in the eigenvectors of the network graph Laplacian. From these we derived expressions for average controllability centrality and volumetric control energy centrality in terms of properties of the underlying graph. We showed how the optimal leader sets for controllability metrics are in tension with the optimal leader set for robustness, and thus require a trade-off if both features are desirable.

Future directions include characterizing $\mathrm{tr}(W_C^+)$ and $\lambda_{min}(W_C)$ in terms of properties of the graph and expanding the decentralized leader selection algorithm to accommodate sets of more than two leaders.

Chapter 7

Final Remarks

This dissertation examined leader selection problems for robustness of the consensus dynamics to noise and controllability in multi-agent networks, and proved relationships between group level performance metrics and the network properties of leader sets. Our approach allowed us to determine, for multiple performance metrics, where optimal leaders should be located in a network graph solely as a function of the network topology. This is an important result that has not been previously established in the literature. We investigated leader selection for robustness and defined a new notion of centrality of a set of nodes, called *joint centrality*, based on our results. It was shown that joint centrality rigorously accounts for the individual centrality and distribution of nodes over a network. We compared this with our results on leader selection for controllability and explored the inherent tension between the optimal leader sets for robustness and the optimal leader sets for controllability.

Leader selection for robustness of the leader-follower consensus dynamics to noise was discussed in Chapter 3, where it was demonstrated that the total system error, equivalently the $\mathcal{H}_2$-norm of the system, could be written in terms of graph measures, dependent only on the leader submatrices of L^+ and L^{2+}. We defined joint centrality such that the node set with highest joint centrality minimizes total system error.

Additionally we demonstrated that node sets with high joint centrality are composed of nodes that trade off high individual information centralities with coverage over the network graph. We illustrated this trade off through an examination of two-node joint centrality and relevant examples. We proved that for one leader, the optimal selection for robustness is the most information central node.

We considered special cases of graphs in Chapter 4. By applying properties of unweighted tree graphs, we simplified the expression for biharmonic distance and subsequently presented reduced expressions for two- and three-node joint centrality. We also provided bounds on two- and three-node joint centrality that allow for computationally efficient leader selection and illustrated this with examples. Motivated by the results for two and three leaders, we discussed heuristic extensions for selection of $m > 3$ leaders in unweighted trees. We derived explicit solutions for the selection of m leaders in two types of one-dimensional graphs: unweighted paths and cycles. We proved that there is equal spacing, up to a rounding, between any two optimal leader nodes on a path graph and that the distance between the ends of the path and optimal first and last leaders, respectively, is less than the distance between any two leaders. Furthermore, expressions for the spacing between any two optimal leaders or a leader and the end of the path were presented along with rules on rounding to ensure optimality. We demonstrated that the optimal solution of m leaders on an unweighted cycle graph corresponds to the leaders evenly distributed around the cycle.

Chapter 5 explored the use of joint centrality in applications other than the leader selection problem. We studied the gene network of *Saccharomyces cerevisiae* (baker's yeasst) and demonstrated that synthetically lethal pairs of nodes tended to have high values of joint centrality when compared to all node pairs in the network. We showed that joint centrality can be applied to graph clustering and provided two examples in which nodes were clustered into m clusters according to covariances with nodes in

the set with highest m-node joint centrality.

We discussed leader selection to optimize controllability metrics in Chapter 6. The optimal set of m leaders for average controllability was found to be the set of m nodes with the lowest information centrality and subsequently the node with the highest average controllability centrality is the least information central node. We related leader selection for reachable volume and volumetric control energy centrality to the left eigenvectors of the graph Laplacian. Through these results, we identified an inherent trade-off between the optimal leader set for robustness and the optimal leader sets for controllability metrics, which we illustrated with examples.

7.1 Future Directions

Many compelling open directions remain related to leader selection in multi-agent networks. The combinatorial nature of the problem of leader selection for robustness makes a full solution to the general problem intractable; however, there exist possibilities to develop explicit solutions to special cases of graphs beyond those mentioned in this work. Additionally, generalizations of joint centrality and bounds on joint centrality for m leaders in unweighted trees will provide a completion of the preliminary results on two and three leaders presented in Chapter 4.

The assumption of undirected networks applied in this work implies equal communication between a pair of connected nodes. This does not always hold in real world networks, for example agent i can be sharing information with agent j but the reverse may not be true. Therefore, leader selection in directed networks is an important area of future research. For directed networks, it is currently unknown where optimal leaders for robustness of the leader-follower consensus dynamics to noise would be located as a function of the network topology. Fortunately, recent works in the literature provide a path towards defining joint centrality for directed

networks. A notion of effective resistance in directed networks was defined in [112] and [113], and the authors of [21] applied theory of positive systems to establish convexity in the problem of leader selection for robustness of the leader-follower consensus dynamics to noise in directed networks. It is possible that the combination of theory from these two works will allow for new developments and a broader understanding of leader selection in directed networks.

Another important characteristic of many real world multi-agent networks that we have not yet addressed is time variability. While the static graph approximation is valid when the graph is changing at a much slower time scale than the network dynamics, there exist many scenarios for which this assumption is not true. One approach being studied in the Leonard lab is modeling time-varying graphs as Markov jump linear systems and quantifying system error in terms of instantaneous graph properties and the transition probability matrix.

We primarily discussed joint centrality in reference to the leader selection problem for robustness of the leader-follower consensus dynamics to noise; however, there is opportunity for joint centrality to be used as a measure of centrality of a set of nodes beyond the examples presented in Chapter 6. Thus, another future direction is to establish the generalized efficacy of joint centrality in the capacity of a centrality measure for sets of nodes in a network. This would have an influence not only in the area of multi-agent systems research but the broader field of networked system analysis.

There are multiple controllability metrics, namely $\mathrm{tr}(W_C^+)$ and $\lambda_{min}(W_C)$, for which optimal leader selection has not yet been described in terms of the underlying network graph. These metrics provide valuable insight on the controllability and required control energy for the system, and so generalizations in terms of the graph Laplacian will be particularly useful. Furthermore, we would like to rigorously quantify the trade-off between leader selection for robustness and leader selection for

controllability.

The work presented here is primarily focused on 'top down' leader selection, that is, an external observer selects leaders *a-priori* based on properties of the network structure. Through this approach we have identified important nodal properties that lead to high performing leader sets. An important direction for this work is to apply these results to further develop distributed leader selection algorithms in which nodes use information obtained through message passing to determine whether or not they should act as a leader. This not only removes the need for an external observer to select leaders, but also allows for networks to be adaptive in the face of changing environment or communication structure.

Bibliography

[1] L. Arnold. *Stochastic Differential Equations: Theory and Applications*. Krieger Publishing Complany, 1992.

[2] L. Auslander and S. Parter. On imbedding graphs in the sphere. *J. Math. Mechanics*, 10:517–523, 1961.

[3] T. Balch and R. C. Arkin. Behavior-based formation control for multirobot teams. *IEEE Trans. Robot. Autom.*, 14(6):926–939, 1998.

[4] A. Bavelas. Communication patterns in task-oriented groups. *J. Acoust. Soc. Am.*, 22(6):725–730, 1950.

[5] A. Ben-Israel and T. N. E. Greville. *Generalized Inverses: Theory and Applications*. Springer-Verlag, 2nd edition, 2003.

[6] D. Bernstein. *Matrix Mathematics*. Princeton University Press, 2005.

[7] D. Bertsekas. *Nonlinear Programming*. Athena Scientific, 2nd edition, 1999.

[8] J. Björk and M. Magnusson. Where do good innovation ideas come from? *J. Prod. Innovat. Manag.*, 26(6):662–670, 2009.

[9] P. Bonacich. Simultaneous group and individual centralities simultaneous group and individual centralities. *Social Networks*, 13(2):155–168, 1991.

[10] S. Borgatti. Identifying sets of key players in a social network. *Comput. Math. Organ. Theory*, 12(1):21–34, 2006.

[11] K. M. Carley. Destabilization of covert networks. *Comput. Math. Organ. Theory*, 12:51–66, 2006.

[12] K. M. Carley, J. Reminga, and N. Kamneva. Destabilizing terrorist networks. In *Institute for software research*, page 45, 1998.

[13] P. J. Carrington, J. Scott, and S. Wasserman, editors. *Models and Methods in Social Network Analysis*. Cambridge University Press, 2005.

[14] L. Cheng, Z.-G. Hou, M. Tan, Y. Lin, and W. Zhang. Neural-network-based adaptive leader-following control for multiagent systems with uncertainties. *IEEE Trans. Neural Net*, 21(8):1351–1358, 2010.

[15] F. Chung and S. T. Yau. Discrete green's functions. *J. Combin. Theory Ser. A.*, 91:191–214, 2000.

[16] A. Clark, L. Bushnell, and R. Poovendran. A supermodular optimization framework for leader selection under link noise in linear multi-agent systems. *IEEE Trans. Autom. Control*, 59:283 – 296, 2014.

[17] A. Clark, B. Alomair, L. Bushnell, and R. Poovendran. *Submodularity in Dynamics and Control of Netowkred Systems*. Springer International Publishing, 2016.

[18] N. Cowan, E. Chastain, D. Vilhena, J. Freudenberg, and C. Bergstrom. Nodal dynamics, not degree distributions, determine the structural controllability of complex networks. *PLoS One*, 10.1371/journal.pone.0038398, 2012.

[19] P. Crucitti, V. Latora, and S. Porta. Centrality measures in spatial networks of urban streets. *Phys Review E*, 73:036125, 2006.

[20] C. M. da Fonseca and J. Petronilho. Explicit inverse of a tridiagonal k-toeplitz matrix. *Numerische Mathematik*, 100:457–482, 2005.

[21] N. K. Dhingra, M. Colombino, and M. R. Jovanovic. On the convexity of a class of structured optimal control problems for positive systems. In *Proc. ECC*, 2016.

[22] C. Enyioha, M. A. Rahimian, G. J. Pappas, and A. Jadbabaie. Controllability and fraction of leaders in infinite networks. In *Proc. IEEE CDC*, pages 1359–1364, 2014.

[23] M. G. Everett and S. P. Borgatti. The centrality of groups and classes. *J. Math. Sociol.*, 23(3):181–201, 1999.

[24] M. Fardad, F. Lin, and M. R. Jovanovic. Algorithms for leader selection in large dynamical networks: noise-free leaders. In *Proc. IEEE CDC*, pages 7188–7193, 2011.

[25] K. Fitch and N. E. Leonard. Information centrality and optimal leader selection in noisy networks. In *Proc. IEEE CDC*, pages 7510–7515, 2013.

[26] K. Fitch and N. E. Leonard. Joint centrality distinguishes optimal leaders in noisy networks. *IEEE Trans. Cont. Network Sys.*, 2016.

[27] K. Fitch and N. E. Leonard. Optimal leader selection for controllability and robustness in multi-agent networks. In *Proc. ECC*, 2016.

[28] C. M. Da Fonseca. On the eigenvalues of some tridiagonal matrices. *J. Comput. Appl. Math.*, 200:283–286, 2007.

[29] L. Freeman. A set of measures of centrality based upon betweenness. *Sociometry*, 40:35–41, 1977.

[30] L. C. Freeman. Centrality in social networks; conceptual clarification. *Social Networks*, 1:215–239, 1978.

[31] L. C. Freeman, D. Roeder, and R. R. Mulholland. Centrality in social networks: ii. experimental results. *Social Networks*, 2(2):119–141, 1979.

[32] F. G. Frobenius. Über matrizen aus positiven etementen. *S.-B. Deutsch. Akad. Wiss. Berlin*, pages 471–476, 1908.

[33] C. W. Gardiner. *Stochastic Methods: A Handbook for the Natural and Social Sciences*. Springer, London, U.K., 2009.

[34] B. P. Gerkey and M. Mataric. A formal analysis and taxonomy of task allocation in multi-robot systems. *Int. J. Robotics Res.*, 23(9):939–954, 2004.

[35] J. Ghaderi and R. Srikant. Opinion dynamics in social networks with stubborn agents: Equilibrium and convergence rate. *Automatica*, 50(12):3209–3215, 2014.

[36] K.-I. Goh, E. Oh, B. Kahng, and D Kim. Betweenness centrality correlation in social networks. *Phys Review E*, 67:017101, 2003.

[37] V. Guttal and I. D. Couzin. Social interactions, information use, and the evolution of collective migration. *PNAS*, 107:16172–16177, 2010.

[38] W. He, B. Zhang, Q.-L. Han, and F. Qian. Leader-following consensus of nonlinear multiagent systems with stochastic sampling. *IEEE Trans. Cybern.*, 2016.

[39] J. P. Hespanha. *Linear Systems Theory*. Princeton University Press, 2009.

[40] V. Ishakian, D. Erdos, E. Terzi, and A. Bestavros. A framework for the evaluation and management of network centrality* read more: http://epubs.siam.org/doi/abs/10.1137/1.9781611972825.37. In *Proc SIAM CDM*, pages 427–438, 2012.

[41] A. Jadbabaie, P. Molavi, A. Sandroni, and A. Tahbaz-Salehi. Non-bayesian social learning. *GEB*, 76:210–225, 2012.

[42] S. Jafari, A. Ajorlou, and A. G. Aghdam. Leader localization in multi-agent systems subject to failure: A graph-theoretic approach. *Automatica*, 47:1744–1750, 2011.

[43] S. Jafari, A. Ajorlou, and A. G. Aghdam. Leader selection in multi-agent systems subject to partial failure. In *Proc. ACC*, pages 5330 – 5335, 2011.

[44] A. K. Jain. Data clustering: 50 years beyond k-means. In *19th International Conference in Pattern Recognition*, volume 31, pages 651–666, 2010.

[45] H. Jeong, S. P. Mason, A.-L. Barabási, and Z. N. Oltvai. Lethality and centrality in protein networks. *Nature*, 411:41–42, 2001.

[46] Z. Ji, H. Lin, and T. H. Lee. A graph theory based characterization of controllability for multi-agent systems with fixed topology. In *Proc. IEEE CDC*, pages 5262–5267, 2008.

[47] Z. Ji, Z. Wang, H. Linc, and Z .Wang. Interconnection topologies for multi-agent coordination under leader–follower framework. *Automatica*, 45:2857–2863, 2009.

[48] N. Kapucu. Interorganizational coordination in dynamic context: Networks in emergency response management. *Connections*, 26(2):33–48, 2005.

[49] I. Katsoukis and G. A. Rovithakis. Leader-follower tracking with prescribed transient and steady state performance guarantees for a class of unknown nonlinear multi-agent systems. In *Proc IEEE MED*, pages 1027 – 1032, 2015.

[50] A. Kchiche and F. Kamoun. Access-points deployment for vehicular networks based on group centrality. In *Proc. NTMS*, pages 1–6, 2009.

[51] H. Kim, J. Shin, E. Kim, H. Kim, S. Hwang, J. E. Shim, and I. Lee. Yeastnet v3: a public database of data-specific and integrated functional gene networks for Saccharomyces Cerevisiae. *Nucl. Acids Res.*, 42((D1):):D731–D736, 2013.

[52] D. J. Klein and M. J. Randic. Resistance distance. *J. Math. Chem.*, 12:81–95, 1993.

[53] E. Kolaczyk, D. Chua, and M. Barthelemy. Group betweenness and co-betweenness: Inter-related notions of coalition centrality. *Social Networks*, 31 (3):190–203, 2009.

[54] V. Latora and M. Marchiori. A measure of centrality based on network efficiency. *New J. Phys.*, 9, 2007.

[55] W. W. Leontief. The structure of american economy, 1919-1929. *Cambridge: Harvard University Press*, page xi, 1941.

[56] H. Li, X. Liao, T. Huang, and W. Zhu. Event-triggering sampling based leader-following consensus in second-order multi-agent systems. *IEEE Trans. Autom. Control*, 60(7):1998–2003, 2015.

[57] C.-T. Lin. Structural controllability. *IEEE Trans. Autom. Control*, 19(3):201–208, 1974.

[58] F. Lin, M. Fardad, and M. R. Jovanovic. Algorithms for leader selection in large dynamical networks: noise-corrupted leaders. In *Proc. IEEE CDC*, pages 2932–2937, 2011.

[59] F. Lin, M. Fardad, and M. R. Jovanovic. Algorithms for leader selection in stochastically forced consensus networks. *IEEE Trans. Autom. Control*, 59: 1789–1802, 2014.

[60] Y. Lipman, R. M. Rustamov, and T. A. Funkhouser. Biharmonic distance. *ACM Transactions on Graphs*, 29:27:1–27:11, 2010.

[61] Y.-Y. Liu, J.-J. Slotine, and A.-L. Barabási. Controllability of complex networks. *Nature*, 473:167–173, 2011.

[62] L. Lovasz. Submodular functions and convexity. In A. Bachem, B. Korte, and M. Grötschel, editors, *Submodular functions and convexity*, pages 235–257. Springer Berlin Heidelberg, 1983.

[63] H.-W. Ma and A.-P. Zeng. The connectivity structure, giant strong component and centrality of metabolic networks. *Bioinformatics*, 19(11):1423–1430, 2003.

[64] J. Ma, Y. Zheng, and L. Wang. Topology selection for multi-agent systems with opposite leadesr. *Syst. Control Lett.*, 93:43–49, 2016.

[65] S. Martini, M. Egerstedt, and A. Bicchi. Controllability analysis of multi-agent systems using relaxed equitable partitions. *Int. J. Systems, Contr and Comm.*, 2:100–121, 2010.

[66] B. Marx, D. Koenig, and D. George. Optimal sensor and actuator location for descriptor systems using generalized gramians and balanced realizations. In *Proc. ACC*, pages 2729–2734, 2004.

[67] K. S. Miller. On the inverse of the sum of matrices. *Mathematics Magazine*, 54 (2):67–72, 1981.

[68] G. L. Nemhauser, L. A. Wolsey, and M. L. Fisher. An analysis of approximations for maximizing submodular set functions–I. *Math. Program.*, 14:265–294, 1978.

[69] C. Ni, C. Sugimoto, and J. Jiang. Degree, closeness, and betweenness: Application of group centrality measurements to explore macro-disciplinary evolution diachronically. In *Proc. ISSI*, 2011.

[70] R. Olfati-Saber and R. M. Murray. Consensus problems in networks of agents with switching topology and time-delays. *IEEE Trans. Autom. Control*, 49(9): 1520–1533, 2004.

[71] D. Pais and N. E. Leonard. Adaptive network dynamics and evolution of leadership in collective migration. *Physica D*, 267:81–93, 2014.

[72] A. Papoulis. *Probability, Random Variables, and Stochastic Processes*. McGraw-Hill, 2nd edition, 1984.

[73] F. Pasqualetti, S. Zampieri, and F. Bullo. Controllability metrics, limitations and algorithms for complex networks. *IEEE TCNS*, 1(1):3287–3292, 2014.

[74] S. Patterson and B. Bamieh. Leader selection for optimal network coherence. In *Proc. IEEE CDC*, pages 2693–2697, 2010.

[75] S. Patterson, N. McGlohon, and K. Dyagilev. Optimal k-leader selection for coherence and convergence rate in one-dimensional networks. *IEEE TCNS*, PP, 2016.

[76] S. Pequito, V. Preciado, and G. Pappas. Distributed leader selection. In *Proc. IEEE CDC*, pages 962–967, 2015.

[77] S. Pequito, J. Svacha, G. J. Pappas, and V. Kumar. Sparsest minimum multiple-cost structural leader selection. In *IFAC NecSys*, volume 48, pages 144–149, 2015.

[78] M. Pirani and S. Sundaram. Spectral properties of the grounded laplacian matrix with applications to consensus in the presence of stubborn agents. In *Proc. ACC*, pages 2160 – 2165, 2014.

[79] M. Pirani, E. M. Shahrivar, B. Fidan, and S. Sundaram. Robustness of leader-follower networked dynamical systems. *arXiv:1604.08651v1*, 2016.

[80] I. Poulakakis, L. Scardovi, and N. E. Leonard. Coupled stochastic differential equations and collective decision making in the two- alternative forced-choice task. In *Proc. ACC*, pages 69–74, 2010.

[81] I. Poulakakis, G. F. Young, L. Scardovi, and N. E. Leonard. Information centrality and ordering of nodes for accuracy in noisy decision-making dynamics. *IEEE Trans. Autom. Control*, 61(4):1040–1045, 2016.

[82] R. Puzis, Y. Elovici, and S. Dolev. Fast algorithm for successive computation of group betweenness centrality. *Phys Review E*, 76(5):056709, 2007.

[83] R. Puzis, Y. Elovici, and S. Dolev. Finding the most prominent group in complex networks. *AI Comm*, 20(4):287–296, 2007.

[84] A. Rahmani, M. Ji, M. Mesbahi, and M. Egerstedt. Controllability of multi-agent systems from a graph-theoretic perspective. *SIAM J. Control Optim.*, 48 (1):162–186, 2009.

[85] W. Ren. Multi-vehicle consensus with a time-varying reference state. *Syst. Control Lett.*, 56(7):474–483, 2007.

[86] W. Ren, R. Beard, and E. Atkins. A survey of consensus problems in multi-agent coordination. In *Proc. ACC*, pages 1859–1864, 2005.

[87] J. Riehl and M. Cao. Optimal defensive resource allocation for a centrality-based security game on multi-hop networks. In *Proc. IEEE CDC*, pages 6257–6262, 2015.

[88] C. Rowley. Automatic control systems. Lecture notes, 2015.

[89] J. Sabater and C. Sierra. Reputation and social network analysis in multi-agent systems. In *Proc AAMAS*, pages 475–482, 2002.

[90] L. Scardovi and N. E. Leonard. Robustness of aggregation in networked dynamical systems. In *Proc ROBOCOMM*, pages 1–6, 2009.

[91] S. E. Schaeffer. Graph clustering. *Comp. Sci. Rev.*, 1:27–64, 2007.

[92] M. E. Shaw. Group structure and the behavior of individuals in small groups. *J Psych.*, 38:139–149, 1954.

[93] G. Shi, K. C. Sou, H. Sandberg, and K. H. Johansson. A graph-theoretic approach on optimizing informed-node selection in multi-agent tracking control. *Physica D*, 267(15):104–111, 2014.

[94] Q. Song, J. Cao, and W. Yu. Second-order leader-following consensus of nonlinear multi-agent systems via pinning control. *Syst. Control Lett.*, 59, 2010.

[95] V. Srivastava and N. E. Leonard. Collective decision-making in ideal networks: The speed-accuracy tradeoff. *Transactions on Control of Network Systems*, 1 (1):121–131, 2014.

[96] C. Stark, B. J. Breitkreutz, T. Reguly, L. Boucher, A Breitkreutz, and M Tyers. Biogrid: A general repository for interaction datasets. *Nucleic Acids Res*, 34: D535–9, 2006.

[97] K. Stephenson and M. Zelen. Rethinking centrality: Methods and examples. *Social Networks*, 11:1–37, 1989.

[98] T. H. Summers, F. L. Cortesi, and J. Lygeros. On submodularity and controllability in complex dynamical networks. In *arXiv:1404.7665v2*, 2015.

[99] H. G. Tanner. On the controllability of nearest neighbor interconnections. In *Proc. IEEE CDC*, pages 2467–2472, 2004.

[100] C. Torney, S. levin, and I. Couzin. Specialization and evolutionary branching within migratory populations. *Proc Nat Acad Sci*, 107:20394–20399, 2010.

[101] V. Tzmoumas, M. A. Rahimian, G. J. Pappas, and A. Jadbabaie. Minimum actuator placement with optimal control constraints. In *Proc. ACC*, pages 2081 – 2086, 2015.

[102] R. A. Usmani. Inversion of a tridiagonal jacobi matrix. *Linear Algebra and its Applications*, 212-213:413–414, 1994.

[103] J. Wang, Y. Tan, and I. Mareels. Robustness analysis of leader-follower consensus. *J Syst. Sci. Complex.*, 22:186–206, 2009.

[104] Y. Wang, W. Yang, L. Wang, and X. Wang. Leader selection for fast consensus in networks. *EPL*, 112(5), 2015.

[105] G. Wen, W. Yu, Z. Li, X. Yu, and J. Cao. Neuro-adaptive consensus tracking of multiagent systems with a high-dimensional leader. *IEEE Trans. Cybern.*, PP, 2016.

[106] M. A. Woodbury. *Inverting modified matrices.* Statistical Research Group, Princeton Univ. Memorandum Rept. 42, 1950.

[107] G. Yan, J. Ren, Y.-C. Lai, C.-H. Lai, and B. Li. Controlling complex networks: How much energy is needed? *PRL*, 108:218703–1 – 218703–5, 2012.

[108] E. Yildiz, D. Acemoglu, A. Ozdaglar, A. Saberi, and A. Scaglione. Discrete opinion dynamics with stubborn agents, 2011. URL `http://ssrn.com/abstract=1744113`.

[109] G. Young, L. Scardovi, and N. E. Leonard. Robustness of noisy consensus dynamics with directed communication. In *Proc. ACC*, pages 6312–6317, 2010.

[110] G. F. Young, L. Scardovi, and N. E. Leonard. Rearranging trees for robust consensus. In *Proc. IEEE CDC*, pages 1000–1005, 2011.

[111] G. F. Young, L. Scardovi, A. Cavagna, I. Giardina, and N. E. Leonard. Starling flock networks manage uncertainty in consensus at low cost. *PLoS Comput. Biol.*, 9:1–7, 2013.

[112] G. F. Young, L. Scardovi, and N. E. Leonard. A new notion of effective resistance for directed graphs-part I: Definition and properties. *IEEE Trans. Autom. Control*, 61(7):1727 – 1736, 2016.

[113] G. F. Young, L. Scardovi, and N. E. Leonard. A new notion of effective resistance for directed graphs-part II: Computing resistances. *IEEE Trans. Autom. Control*, 61(7):1737 – 1752, 2017.

[114] B. Zhao, Y. Guan, and L. Wang. Leader selection and weight adjustment for controllability of multi-agent systems. *arXiv:1503.05913v2*, 2015.