Strategic and Stochastic Approaches to Modeling the Structure of Multi-Layer and Interdependent Networks

by

Ebrahim Moradi Shahrivar

A thesis
presented to the University of Waterloo
in fulfillment of the
thesis requirement for the degree of
Doctor of Philosophy
in
Electrical and Computer Engineering

Waterloo, Ontario, Canada, 2016

I hereby declare that I am the sole author of this thesis. This is a true copy of the thesis, including any required final revisions, as accepted by my examiners.

I understand that my thesis may be made electronically available to the public.

Abstract

Examples of complex networks abound in both the natural world (e.g., ecological, social and economic systems), and in engineered applications (e.g., the Internet, the power grid, etc.). The topological structure of such networks plays a fundamental role in their functioning, dictating properties such as the speed of information diffusion, the influence of powerful or vulnerable nodes, and the ability of the nodes to take collective actions. There are two main schools of thought for investigating the structure of complex networks. Early research on this topic primarily adopted a stochastic perspective, postulating that the links between nodes are formed randomly. In an alternative perspective, it has been argued that optimization (rather than pure randomness) plays a key role in network formation. In such settings, edges are formed strategically (either by a designer or by the nodes themselves) in order to maximize certain utility functions. The classical literature on the structure of networks has predominantly focused on single layer networks where there is a single set of edges between nodes. However, there is an increasing realization that many real-world networks have either multi-layer or interdependent structure. While the former considers multiple layers of relationships between the same set of nodes, the latter deals with networks-of-networks consisting of interdependencies between different subnetworks. This thesis focuses on the analysis of the structure of multi-layer and interdependent networks via strategic and stochastic approaches.

In the strategic multi-layer network formation setting, each layer represents a different type of relationship between the nodes and is designed to maximize some utility that depends on its own topology and those of the other layers. By viewing the designer of each layer as a player in a multi-layer network formation game, we show that hub-and-spoke networks that are commonly observed in transportation systems arise as a Nash equilibrium. Extending this analysis to interdependent networks where there are different sets of nodes, we introduce a network design game where the objective of the players is to design the interconnections between the nodes of two different networks, G_1 and G_2. In this game, each player is associated with a node in G_1 and has functional dependencies on certain nodes in G_2. Besides showing that finding a best response of a player is NP-hard and characterizing some useful properties of the best response actions of the players, we prove existence of pure Nash equilibria in this game under certain conditions. In order to obtain further insights into the structure of interdependent networks with an arbitrary number of subnetworks, we consider a model for random interdependent networks where each edge between two different subnetworks is formed with probability p. We investigate certain spectral and structural properties of such networks, with corresponding implications for certain variants of consensus dynamics on those networks. In particular, we study a property known as r-robustness, which is a strong indicator of the ability of a network, including interdependent networks, to tolerate structural perturbations and dynamical attacks.

Acknowledgment

I would like to express my deepest appreciation to my supervisor, Shreyas Sundaram for his constant support, understanding and encouragement throughout my study and research. I am indebted to him for his patience in careful proofreading of countless drafts of my works. This thesis would not have been possible without his excellent guidance. I am also sincerely grateful of my co-supervisor Mahesh Tripunitara for his kind supports.

I would also like to thank the other members of my thesis committee (Professors Luca Scardovi, Stephen Smith, John Simpson-Porco and Chaitanya Swamy) for reading a preliminary version of this thesis and their useful comments.

I thank those with whom I collaborated during my PhD, including Mohammad Pirani and Abbas Mehrabian.

I am grateful to my friends in Waterloo and Toronto for all the hangouts, laughters, travels and campings that we have shared.

I always think that I am exceptionally blessed to have such a wonderful family. I cannot thank my parents enough for all of their love and support. This thesis is for them. Also, I would like to thank my sister for her kindness and the cheerful conversations that we had during this time.

At last but not least, I want to thank my beloved wife whose presence and supports made completing this thesis easier and smoother for me.

Table of Contents

List of Tables

List of Figures

Chapter 1

Introduction and Background

Due to rapid advances in communication technology, it is impossible to analyze behavior of individuals in society without considering the "connection" or "mutual effect" of one person to/on another. The property of being "interconnected" is not only limited to social systems. In fact, the ever-increasing sophistication of today's large-scale engineered systems requires the development of new techniques to understand the implications of their interconnections. With this paradigm shift, the concept of "networks" has started to attract even more attention over the past decades.

In the most basic sense, a network is comprised of any set of objects, which are referred to as nodes, and a set of links capturing possible connections between pairs of nodes. Depending on the context, links can define different forms of connections and relationships, such as friendship between individuals in a social network, roads in a transportation network, or data transmission paths in a computer network. Similarly, in a distributed system, objects or nodes can model different types of components, such as humans, robots, computers (agents) or in general any dynamical system.

Over the last several decades, researchers have studied various aspects of distributed systems. Some have focused on individual components, such as exploring how a robot works in a cooperative environment or how humans are influenced by society. There are also interesting studies on the nature of the connections or interactions between components, e.g., the communication protocols on the Internet or friendship dynamics among people. But there is a third aspect to the analysis of these systems, and that is their structure or pattern of interconnections.

Various studies have shown that the topological structure of interconnection between the components of a decentralized system plays a fundamental role in the system's functioning. For instance, [73] showed that the convergence rate to the rendezvous point in a multi-agent setting with consensus dynamics is directly related to the algebraic connectivity of the interconnection network among the agents. The interconnection structure in a decentralized system also crucially affects the robustness of the system against dynamical attacks or random failures [98, 97, 43, 86]. In light of this, our studies in this thesis contributes to the understanding of the structure of large-scale and complex networks, an area usually referred to as "network design" or "network formation" in the literature. In order to describe our contributions in more detail, we will first need to introduce some terminology.

1.1 Definitions

In this section, we start by providing some essential graph theory terminology and then we will discuss definitions related to the complexity of solving problems.

1.1.1 Graph Theory Terminology

An undirected network (or graph) is denoted by $G = (N, E)$ where $N = \{v_1, v_2, \ldots, v_n\}$ is the set of nodes (or vertices) and $E \subseteq \{(v_i, v_j) | v_i, v_j \in N, v_i \neq v_j\}$. The set of all possible graphs on N is denoted by G^N. Two nodes are said to be neighbors if there is an edge between them. The degree of a node $v_i \in N$ is the number of its neighbors in graph G, and is denoted by $\deg_i(G)$. A leaf node is a node that has degree one, i.e., it has only one neighbor. A path from node v_1 to v_k in graph G is a sequence of distinct nodes $v_1 v_2 \cdots v_k$ where there is an edge between each pair of consecutive nodes of the sequence. The length of a path is the number of edges in the sequence. We denote the shortest distance between nodes v_i and v_j in graph G by $d_G(i, j)$. If there is no path from v_i to v_j, we take $d_G(i, j) = \infty$. The diameter of the graph G is $\max_{v_i, v_j \in N, v_i \neq v_j} d_G(i, j)$. A cycle is a path of length two or more from a node to itself. A graph $G' = (N', E')$ is called a subgraph of $G = (N, E)$, denoted as $G' \subseteq G$, if $N' \subseteq N$ and $E' \subseteq E \cap \{N' \times N'\}$. A graph G' is said to be induced by a set of nodes $N' \subseteq N$ if $E' = E \cap \{N' \times N'\}$. A graph is connected if there is a path from every node to every other node. A subgraph $G' = (N', E')$ of G is a component if G' is connected and there are no edges in G between nodes in N' and nodes in $N \setminus N'$.

A tree is a connected acyclic graph. For a connected graph $G = (N, E)$, a connected acyclic subgraph $T = (N, E_T)$ of G is called a spanning tree of G. A spanning forest of a disconnected graph is a collection of spanning trees of each of its components.

We denote the complete graph (i.e., the graph with an edge between every pair of different nodes) by $G^c = (N, E^c)$. We use $G^e = (N, \phi)$ to denote the empty graph. Finally, $G^s = (N, E^s)$

is a star graph, which is a tree graph with one node that is connected to all other nodes. The complement of graph $G = (N, E)$ is denoted by $\sim G = (N, \sim E)$, where $\sim E \triangleq E^c \setminus E$. Two graphs on the same set of nodes are said to be disjoint if their edge sets are disjoint. For an integer $k \in \mathbb{Z}_{\geq 2}$, a graph G is *k-partite* if its vertex set can be partitioned into k sets $V_1, V_2, \ldots, V_k$ such that there are no edges between nodes within any of those sets.

The *adjacency matrix* for the graph $G = (V, E)$ is a matrix $A \in \{0, 1\}^{n \times n}$ whose (i, j) entry is 1 if $(v_i, v_j) \in E$, and zero otherwise. The *Laplacian matrix* for the graph is given by $L = D - A$, where D is the degree matrix with $D = \mathrm{diag}(d_1, d_2, \ldots, d_n)$. The eigenvalues of the Laplacian are real and nonnegative, and are denoted by $0 = \lambda_1(L) \leq \lambda_2(L) \leq \ldots \leq \lambda_n(L)$.

1.1.2 NP-hardness and Complexity

Let $\mathbb{N}_{>0}$ denote the set of positive integers. Consider two functions $f(n) : \mathbb{N}_{>0} \to \mathbb{R}$ and $g(n) : \mathbb{N}_{>0} \to \mathbb{R}$, where $\mathbb{R}$ is the set of real numbers. Then $f(n)$ is said to be in

- $O(g(n))$ if and only if there exist positive constants c and n_0 such that
$$0 \leq f(n) \leq c\, g(n) \quad \forall n \geq n_0.$$
- $\Theta(g(n))$ if and only if there exist positive constants c_1, c_2 and n_0 such that
$$0 \leq c_1 g(n) \leq f(n) \leq c_2\, g(n) \quad \forall n \geq n_0.$$
- $\Omega(g(n))$ if and only if there exist positive constants c and n_0 such that
$$0 \leq c\, g(n) \leq f(n) \quad \forall n \geq n_0.$$
- $o(g(n))$ if and only if for every positive constant c, there exists a constant $n_0 > 0$ such that
$$0 \leq f(n) < c\, g(n) \quad \forall n \geq n_0.$$

Assume that the input to a problem has size n (measured by an appropriate encoding of the input). If there is an algorithm that solves the problem in $O(n^k)$ time (for some positive constant k), the problem is said to be in the complexity class P. A decision problem is a question to which answer is either “yes” or “no”. A decision problem is said to be in the class NP if every “yes” answer has an accompanying certificate that can be verified in polynomial-time. Consider two decision problems A and B and assume that there exists a polynomial-time transformation from any instance b of problem B into some instance a of problem A such that the answer to b is "yes" if and only if answer to a is "yes". If such a transformation from B to A exists, it is called a *reduction* and problem B is said to be *polynomial-time reducible* to problem A. A problem A is NP-hard if for all problems $B \in$ NP, B is polynomial-time reducible to A; in particular, A is NP-hard if some other NP-hard problem B is polynomial-time reducible to A [19]. An NP-hard problem that is also in the class NP is said to be NP-complete.

With the above definitions and terminology in hand, we will now review previous results on network formation from stochastic and strategic points of view.

1.2 Random Network Formation

Endeavors to study the structure of large-scale networks were first initiated by formulating a stochastic setting in which links are formed randomly. A fundamental work with this perspective is the "small world" model proposed by Watts and Strogatz [93]. By randomly rewiring links in a symmetric network, they generated a network that exhibits small world characteristics such as a short average distance between any two pair of nodes and a high clustering coefficient. In [6], Albert and Barabasi showed that if nodes form links through preferential attachment (i.e., the probability of forming a link between a new node and an existing node v is proportional to $\deg(v)$), then the produced network will have power law degree distribution. Similar ideas go back to the 1950's [71, 48, 17]. Another interesting model, which illustrates certain features of social networks, was introduced by Jackson and Rogers [50]. In their model, a new node forms links in two steps: first, randomly connecting to a set of nodes, and then searching locally through the current structure of the network (e.g., friends of friends).

However, probably the most-prominent network formation model with a stochastic perspective is the Erdos-Renyi (ER) model in which, given a set of nodes, there is an edge between any two nodes with a fixed probability [26]. Although ER networks are not typically representative of real-world networks, due to the interesting properties that they demonstrate, they have become one of the most common models for studying large scale networks [74, 10, 48, 26]. We will now formally define the ER network formation model.

Definition 1. *An Erdos-Renyi (ER) random network, denoted by $G(n, p)$, is a graph on n nodes where each possible edge between two distinct vertices is present independently with probability p (which could be a function of n). Equivalently, an ER random network can be viewed as a probability space (Ω_n, F_n, P_n), where the sample space Ω_n consists of all possible graphs on n nodes, the σ-algebra F_n is the power set of Ω_n, and the probability measure P_n assigns a probability of $p^{|E|}(1-p)^{\binom{n}{2}-|E|}$ to each graph with $|E|$ edges.*

Let $\mathbb{P}(\cdot)$ denote the probability of an event. Given a property $\mathcal{P}$ and the probability of edge formation $p = p(n)$, we say that $G(n, p)$ has the property $\mathcal{P}$ asymptotically almost surely (a.a.s.) if $\mathbb{P}(G(n, p) \in \mathcal{P}) \to 1$ as $n \to \infty$. Similarly, we say that $G(n, p)$ does not have the property $\mathcal{P}$ a.a.s. if $\mathbb{P}(G(n, p) \in \mathcal{P}) \to 0$ as $n \to \infty$.

One of the fascinating features of ER networks is that they exhibit phase transition at certain thresholds for the edge formation probability p. Below we provide one definition of a threshold in random networks.

Definition 2. *Consider a function $t(n) = \frac{g(n)}{n}$ with $g(n) \to \infty$ as $n \to \infty$, and a function $x = o(g(n))$ which satisfies $x \to \infty$ as $n \to \infty$. Then $t(n)$ is said to be a (sharp) threshold function for a graph property $\mathcal{P}$ if*

1. *property $\mathcal{P}$ a.a.s. holds when $p(n) = \frac{g(n)+x}{n}$, and*
2. *property $\mathcal{P}$ a.a.s. does not hold for $p(n) = \frac{g(n)-x}{n}$.*

Loosely speaking, "$t(n)$ is a (sharp) threshold for the property $\mathcal{P}$" means that if the edge formation probability p is "larger" than threshold $t(n)$, then property $\mathcal{P}$ a.a.s. holds; if p is "smaller" than the threshold $t(n)$, then $\mathcal{P}$ a.a.s. does not hold.

1.3 Strategic Network formation

While random networks are appealing from an analytical point of view, they do not necessarily capture the driving principles beneath social, economic or engineered networks. In fact, in these kinds of networks, agents have discretion about connections that they form with other individuals and do not form their links purely by random. This led to the "strategic" approach for analyzing and modeling of the structure complex networks, driven by the economics, computer science and engineering communities, in which optimization plays the key role (rather than pure randomness). In such settings, edges are formed (either by a designer or by the nodes themselves) in order to maximize certain utility functions, resulting in networks that can be analyzed using game-theoretic notions of equilibria and efficiency [4, 9, 12, 41, 55].

However, there are two crucial challenges in modeling network formation from a strategic point of view. First an appropriate metric is needed to capture individuals' or the central designer's incentives to form or remove a link. Thus, we need to explicitly model the benefit and cost of forming edges. Another challenge is in predicting the formed networks that maximize the given measure or utility function.

The past few decades have generated a large volume of literature on the problem of network formation and design for different types of utility functions. We will now review some of these lines of work and then explain how our research fits within the context of existing literature.

1.3.1 Network Design

The selection of an optimal configuration or design of a network occurs in many different application contexts including transportation, telecommunication and electric power systems. Optimality of the designed network is measured by a utility function that depends on the nature of the problem under investigation. An example is the so-called *Facility Location* problem where there are a set of potential sites, a set of clients, and relevant profit and cost data. The goal is to find a maximum-profit plan giving the number of facilities to open, their location and an allocation of each client to an open facility. The solution to this problem can be captured as a bipartite graph. The facility location problem is NP-hard [66].

There are other types of network design problems that have been widely studied in the literature. Perhaps the most common instance is the minimum spanning tree (MST) problem which is to find a spanning tree of a weighted graph that has the least overall weight. Kruskal and Prim are two greedy algorithms that solve MST in polynomial time [19]. The Steiner tree problem is a problem in combinatorial optimization that is similar to the minimum spanning tree problem. One of the versions of the problem that is most related to our work is: given a subset of vertices $R \subseteq V$ of graph $G = (V, E)$ and a positive integer $K \leq |V| - 1$, is there a subtree of G that includes all the vertices of R and that contains no more than K edges? Steiner tree is one of Karp's 21 NP-complete problems [54].

In an important paper, Hu introduced the optimal communication spanning tree (OCST) problem defined below.

Problem 1. *Given a set of nodes N and a set of requirements r_{ij}, $i, j \in N$, the goal of the network designer is to find the spanning tree connecting the set of nodes in N such that the total cost of communication is minimized over all spanning trees. The cost of communication for a pair of nodes is r_{ij} multiplied by their geodesic distance in the tree. Summing over all $\binom{n}{2}$ pair of nodes gives the cost of the spanning tree. Thus the objective function has the form*

$$u(T) = \sum_{i,j \in N, i \neq j} r_{ij} d_T(i, j). \tag{1.1}$$

Hu in [45] showed that this problem is polynomially solvable for any set of r_{ij}.

The Network Design Problem (NDP) is another important problem in the network design literature introduced by Johnson et al. in [53], and is defined as follows.

Problem 2. *Given an undirected graph $G = (V, E)$, a weight function $L : E \to \mathbb{N}$, a budget $B \in \mathbb{N}$ and a criterion $C \in \mathbb{N}$, does there exist a subgraph $G' = (V, E')$ of G with overall weight $\sum_{(i,j) \in E'} L(i, j) \leq B$ and criterion value $F(G') \leq C$, where $F(G')$ denotes the sum of the weights of the shortest paths in G' between all vertex pairs?*

Johnson shows that this problem is NP-complete by reduction from the Subset-Sum problem which is NP-complete as well [34]. In a further investigation, Johnson showed that the special instance of this problem (known as SNDP) with $L(i, j) = 1$ for all $(i, j) \in E$ and $B = |V| - 1$ is also NP-complete.

The above problems are classical in the network design literature. An alternative formulation that explicitly incorporates the costs of forming links has been pursued by the economic community over the past two decades. We now explain this formulation.

Classical Distance-based Utility

A well-studied and natural utility function is the so-called *distance-based utility* introduced in [46, 48] by Jackson and Wolinsky, where the objective is to purchase edges to minimize the distances between all pairs of nodes in the network. In this model, there is a net benefit of $b(k)$ for each pair of nodes that are k hops away in the network, where $b(\cdot)$ is a decreasing nonnegative function (i.e., nodes that are further away from each other provide smaller benefits and $b(\infty) = 0$). There is a cost $c > 0$ for each edge in the network. The outcome of the network formation process is a graph $G = (N, E) \in G^N$. A graph is evaluated according to the utility function (or value function) $u : G^N \to \mathbb{R}$ given by

$$u(G) = \sum_{i,j \in N : i \neq j} b(d_G(i, j)) - c|E|. \tag{1.2}$$

In this formulation, there is an inherent trade-off faced by the designer: adding links to a larger number of nodes incurs a larger benefit (by reducing the distances between nodes), but also a larger cost invested in links. An *efficient* network is the network with the highest utility. In other words, if G is an efficient network, then $u(G) \geq u(G'), \forall G' \in G^N$.

The following result from [46, 48] shows that there are only a few different kinds of efficient networks, depending on the relative values of the link costs and connection benefits: the empty network (for high link costs), the star network (for medium link costs), and the fully connected network (for low link costs).

Proposition 1 ([48]). *In the distance-based utility model,*

- *if* $c < b(1) - b(2)$, *then the complete network is the unique efficient network;*
- *if* $b(1) - b(2) < c < b(1) + (n-2)b(2)/2$, *then the star is the unique efficient network;*
- *if* $b(1) + (n-2)b(2)/2 < c$, *then the empty network is the unique efficient network.*

Remark 1. *In the above proposition, whenever c is equal to one of the specified upper or lower bounds, there will be more than one efficient network: if $c = b(1) - b(2)$, then the complete network and star network are both efficient, and if $c = b(1) + (n-2)b(2)/2$, the star network and the empty network are both efficient networks with zero utility. Furthermore, for the more general case where $b(\cdot)$ is nonincreasing, the three networks given by the above result are still optimal for the corresponding ranges of costs and benefits, although they may no longer be unique.*

Proposition 1 is about finding the network that maximizes the overall societal welfare given in Equation (1.2), when there exists a central designer that decides about connections between nodes. Similarly, [48] defines the distance based utility function from the nodes' point of view as

$$u_i(G) = \sum_{j \in N: i \neq j} b(d_G(i,j)) - c\, \deg_G(i), \tag{1.3}$$

where $\deg_G(i)$ denotes the degree of node i (i.e., the number of its incident edges). Jackson in [48] assumes that consent of both of the endpoints is needed in order to form a link between them. This seems to be a reasonable assumption in social settings as a relationship between individuals must be in their mutual interest. The Nash equilibrium concept fails to capture the assumption that agents have the capability to negotiate and thus cannot be employed here. Instead the notion of pairwise stability was introduced in [48] and is fully defined in Section 2.4.

In another variation of the distance-based utility, Jackson and Rogers studied the Islands-Connection Model [49]. In this model, there are clusters of geographically close nodes (called islands) with a geographic structure to the cost. It is assumed that the price of intra-island edge construction and inter-island edge construction are c and C respectively, where $C > c > 0$. Furthermore, if the shortest path between two nodes is higher than some value D, then they do not receive any benefit from each other. In this setting, the nodes are the network designers and the overall utility to node i in network G is

$$u_i(G) = \sum_{i \neq j: d_G(i,j) \leq D} \delta^{d_G(i,j)} - \sum_{j: ij \in G} c_{ij},$$

where in the above $0 < \delta < 1$. Fabrikant et al. in [27] assumes that by expending enough effort, nodes can have edges to any other node in the network. This assumption makes it possible to use the concept of Nash equilibria, as in this configuration, nodes are assumed to be independent in their decisions. By slightly changing the distance based utility function and defining $b(k) = n - k$, [27] characterizes the price of anarchy in the network formation problem .

Comparing Different Network Design Models

As one can see, the OCST and SNDP problems are similar to the distance utility formulation of Jackson et al. To clarify the differences between these problems, consider two graphs F_1 and F_2. Assume that the solution network must be a subgraph of F_1, and F_2 denotes the set of pairs of nodes that wish to communicate. If F_1 is the complete graph, we are searching among all possible networks and if F_2 is the complete graph, then the distances between all pairs of nodes appear in the utility function. In Jackson's network formation problem, F_1 and F_2 are both the complete graph. In SNDP, F_1 is a given graph G and F_2 is the complete graph. In OCST F_1 is the complete graph, but F_2 is a given graph G (in OCST, graph G is a weighted graph with $L(i, j) = r_{ij}$).

1.4 Multi-Layer and Interdependent Networks

A common theme in the existing works on network formation is that they focus on the construction of a single set of edges between the nodes. However, many real-world networks have an inherently multi-layer or interdependent structure. While the former considers multiple layers of relationships between the same set of nodes, the latter deals with networks-of-networks consisting of interdependencies between different subnetworks. Figure 1.1 demonstrates the difference between these two structures. Examples of these networks abound in both the natural world (e.g., ecological, social and economic systems), and in engineered applications. Friendship and professional relationships in social networks, policy influence and knowledge exchange in organizational networks, and transportation networks between a group of cities (e.g., bus, train, flight, etc.) are just a few instances of multi-layer networks [94, 85, 48, 82, 57]. Similarly, examples of interdependent networks include coupled communication and energy infrastructure networks, coupled cyber and physical networks, transportation networks, and different communities of individuals in social networks joined together by 'weak ties' [84, 77, 42, 7, 15]. The objective of this thesis is to study multi-layer and interdependent networks from strategic and stochastic view points.

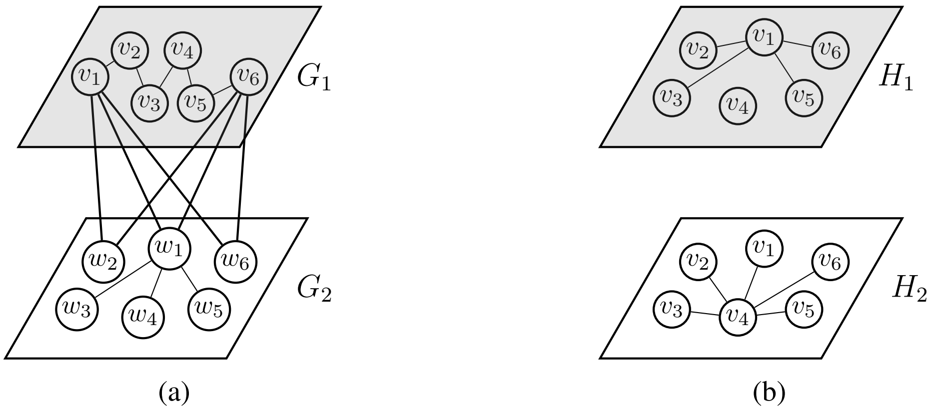


Figure 1.1: (a) An interdependent network with two subnetworks $G_1 = (V, E_1)$ and $G_2 = (W, E_2)$. As we can see there are interdependencies between nodes in V and nodes in the set W. (b) A multi-layer network with two layers $H_1 = (V, E_1)$ and $H_2 = (V, E_2)$. These layers are constructed on the same set of nodes V and have different structures.

1.4.1 Contributions and Outline of Thesis

In Chapter 2, we focus on the strategic approach where the goal is to investigate the formation of multi-layer networks among a common set of nodes. In this setting, each layer represents a different type of relationship between the nodes and is designed to maximize some utility that depends on the topology of that layer and those of the other layers. We generalize distance-based network formation to the two-layer setting, where edges are constructed in one layer (with fixed cost per edge) to minimize distances between nodes that are neighbors in another layer. We show that designing an optimal network (referred to as the best response network) in this setting is NP-hard. Despite the underlying complexity of the problem, we characterize certain properties of the optimal networks. We exploit these properties to determine best responses to networks with certain specific structures. Finally, we investigate satisfaction of the nodes (as the individuals) with the decision of the central network designer via the concept of pairwise stability. Our results illustrate that best response networks (which are socially optimal) are not necessarily pairwise stable.

Using our notion of the best response network, in Chapter 3 we formulate a multi-layer network formation game where each layer corresponds to a player that is optimally choosing its edge set in response to the edge sets of the other players. We consider utility functions that view the different layers as strategic substitutes. By applying our results about optimal networks, we show that players with low edge costs drive players with high edge costs out of the game, and that hub-and-spoke networks that are commonly observed in transportation systems arise as

Nash equilibria in this game.

In Chapter 3, we also consider a variant on the multi-layer network formation game that is based on the classical Colonel Blotto game. We study a scenario where there is a common set of nodes and each player in the game designs a network by purchasing a set of edges between these nodes. We assume that players have a limited budget with which to bid on each edge and the utility of a given set of edges to a player is a function of the resulting network. We characterize the ranges of player budgets for which the game admits pure Nash equilibria for utility functions that depend on the component sizes and diameter of the formed networks.

Extending our strategic multi-layer network formation analysis to the case that there are different sets of nodes, in Chapter 4 we introduce a network design game where the objective of the players is to design the interconnections between the nodes of two different networks G_1 and G_2 in order to maximize certain local utility functions. In this setting, each player is associated with a node in G_1 and has functional dependencies on certain nodes in G_2. We use a distance-based utility for the players in which the goal of each player is to purchase a set of edges (incident to its associated node) such that the sum of the distances between its associated node and the nodes it depends on in G_2 is minimized. We consider a heterogeneous set of players (i.e., players have their own costs and benefits for constructing edges). We show that finding a best response of a player in this game is NP-hard. Despite this, we characterize some properties of the best response actions which are helpful in determining a Nash equilibrium for certain instances of this game. In particular, we prove existence of pure Nash equilibria in this game when G_2 contains a star subgraph, and provide an algorithm that outputs such an equilibrium for any set of players.

Finally, in Chapter 5 we analyze a model for random interdependent networks which consist of a group of subnetworks where each edge between two different subnetworks is formed with probability p. We investigate certain spectral and structural properties of such networks, with corresponding implications for certain variants of consensus dynamics on those networks. We first provide a characterization of the isoperimetric constant in terms of the inter-network edge formation probability p. We then analyze the algebraic connectivity of such networks, and provide an asymptotically tight rate of growth of this quantity for a certain range of inter-network edge formation probabilities. Next, we give bounds on the smallest eigenvalue of the grounded Laplacian matrix of random interdependent networks for the case where one of the subnetworks is comprised entirely of leader nodes. We also study a property known as r-robustness, which is a strong indicator of the ability of a network to tolerate structural perturbations and dynamical attacks. Our results yield new insights into the structure and robustness properties of random interdependent networks.

Chapter 2

Strategic Multi-Layer Network Design

In this chapter, we begin a study of strategic multi-layer network formation by generalizing distance utility network formation to the case where one layer (or network) is formed by optimizing the distances between nodes that are neighbors in another layer (or network). As a motivating example, consider the problem in [60], where both the physical infrastructure network and the traffic flow network between a group of cities are studied (Figure 2.1). Interpreting traffic flow as the weight of connection between the endpoint cities, the objective is to design an optimal infrastructure network between cities with respect to the given traffic flow pattern. In the simplest

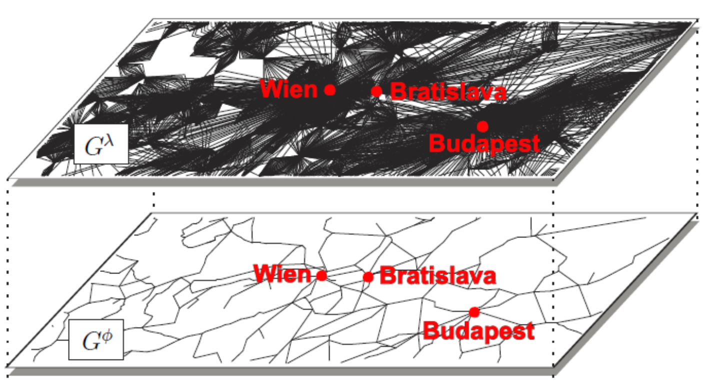


Figure 2.1: Traffic flow network G^λ and physical roads network G^ϕ among a group of cities [60]

case, this problem can be modeled as a network formation problem with a distance-based utility function where only the distances between specific pairs of nodes matter (i.e., those pairs with sufficiently high traffic flow between them). We address this class of problems by first defining a network G_1 capturing an existing set of relationships between nodes, and then studying the formation of an optimal second network G_2 based on G_1. We call the optimally designed network G_2 with respect to G_1 the *best response* network to G_1. Distance-based utilities have also been used to study social networks (where each node is an individual and the edges indicate relationships) [46, 48] and the Internet (where each node represents a router and the edges indicate communication links) [27]. Our formulation generalizes the settings presented in those papers by allowing only distances between certain pairs of nodes (e.g., individuals in the social network or routers in the Internet) to matter when evaluating the utility of the network. For instance, in the case of the Internet or other communication networks, the reference layer G_1 represents the virtual communication network indicating which pairs of nodes wish to exchange information, and the designed layer G_2 represents the physical communication network.

While the best response networks have been completely characterized in the case where G_1 is the complete network [46, 48], we show that finding a best response network with respect to an arbitrary graph G_1 is NP-hard. We characterize some useful properties of the optimal networks that arise in this setting. These properties enable us to find best response networks with respect to certain specific reference networks.

Finally, we investigate satisfaction of the individual nodes with the decisions of the central network designer in the best response network. For this purpose, we generalize the notion of "pairwise stability" to the multi-layer setting. Pairwise stability is roughly defined as the situation in which, firstly no node wants to remove any of its incident edges, and secondly it is not beneficial for any two nodes to add an edge between them (if there is no edge between them already). We characterize conditions under which best response networks are pairwise stable.

2.1 Two Layer Distance-Based Utilities: Best Response Network

In the traditional distance-based network formation problem described in Section 1.3.1, the objective is to minimize the distances between every possible pair of nodes. However, in many settings, one is only interested in minimizing distances between *certain* pairs of nodes. For example, consider a communications system where each node only wishes to exchange information with a subset of the other nodes, and the task is to design a physical network to provide short paths between those pairs of nodes. To handle these types of scenarios, in this section we generalize the study of distance-based network formation to a *multi-layer* setting. Specifically, suppose that we have a layer (or graph) $G_1 = (N, E_1)$, where the edge set E_1 specifies a type of relationship between the nodes in N. Our objective is to design another layer (or graph) $G = (N, E)$ on the same set of nodes, where the utility of the graph is given by

$$u(G|G_1) = \sum_{(v_i,v_j)\in E_1} b(d_G(i,j)) - c|E|. \tag{2.1}$$

Note that the summation is only over edges in set E_1, capturing the fact that only distances between those pairs of nodes matter in graph G; the traditional distance utility function in (1.2) is obtained as a special case when G_1 is the complete graph.

Assume $G_2 = (N, E_2)$ is a network that maximizes (2.1); we say G_2 is a *best response* (BR) network to G_1, or equivalently, an *efficient* network with respect to the utility function (2.1).

Remark 2. *The utility function* (2.1) *does not necessarily have a unique maximizer; indeed, in many cases, there are multiple best response networks with respect to a given network, as demonstrated by Example 1 below.*

When G_1 is the complete network, the best response is trivially a subgraph of G_1. However, the following example demonstrates that the best response network to a general network G_1 does not necessarily have to be a subgraph of that network.

Example 1. *Consider the ring graph G_1 with 6 nodes shown in Figure 2.2a. Suppose $b(1) = c + \epsilon$, for some small constant $\epsilon > 0$. Then,*

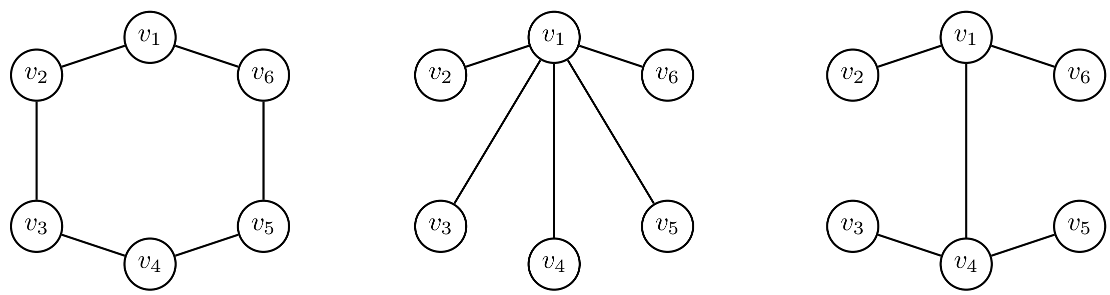
1. *The utility* (2.1) *of G_1 to itself is $u(G_1|G_1) = 6(b(1) - c) = 6\epsilon$.*
2. *Any subgraph of G_1 with 5 edges is a path graph. This has utility $5(b(1) - c) + b(5) = 5\epsilon + b(5)$.*
3. *Any subgraph of G_1 with k edges, where $k < 5$, has utility $k(b(1) - c) = k\epsilon$.*

Thus, when $b(5) > \epsilon$, the best subgraph of G_1 is the path graph with the utility given above.

Now, the star graph shown in Figure 2.2b has utility $2b(1) + 4b(2) - 5c$. This is better than the path graph if $4b(2) - 3b(1) > b(5)$, which holds, for example, when $b(2)$ is sufficiently close to $b(1)$ and $b(2) > b(5)$. Therefore, for utility functions that satisfy this property, no subgraph of G_1 can be a BR to G_1.

For certain benefit functions a star is not a BR either. The graph G_3 given in Figure 2.2c has utility $4b(1) + 2b(3) - 5c$. This is better than the path graph if $2b(3) - b(1) > b(5)$, and better than the star if $b(3) > 2b(2) - b(1)$. For instance if $c = 1$, $b(1) = 1.01, b(2) = 0.85, b(3) = 0.8, b(4) = 0.2$ and $b(5) = 0.1$, then the graph G_3 is better than the star graph or any subgraph *of G_1, i.e., $u(G_3|G_1) > u(G|G_1)$ where $G \subseteq G_1$ or $G = G_2$. In this example, one can verify (e.g., using a brute-force search) that G_3 is in fact a BR network to G_1.*

It is also instructive to consider the case where $b(1) = b(2) = b(3) > \max\{c, b(4)\}$. In this case, the graphs shown in Figure 2.2b and 2.2c are both best response networks to G_1 and have higher utility than any subgraph of G_1.



(a) G_1 (b) G_2 (c) G_3

Figure 2.2: Illustration of potential best response networks with respect to network G_1.

The above example illustrates that BR networks to an arbitrary graph G_1 are very sensitive to the relative values of the benefit function $b(\cdot)$ and the cost c. Indeed, the shape of the entire benefit function can play a role in determining the best response to general graphs, whereas only the value of $b(1)$ and $b(2)$ matter when G_1 is the complete graph (as shown in Proposition 1).

Proposition 1 showed that finding the optimal network in the classical distance-based utility framework can be done in polynomial time. We will now formally characterize the complexity of finding a best response network to a given graph with generalized distance utility function given in equation 2.1. To do this, we first cast it as a decision problem (i.e., a question to which the answer is *yes* or *no*) as follows.

Definition 3. Best Response Network (BRN) Problem.
INSTANCE: A network $G_1 = (N, E_1)$, a nonincreasing benefit function $b : \{1, 2, \cdots, n-1, \infty\} \to \mathbb{R}_{\geq 0}$, an edge cost $c \in \mathbb{R}_{>0}$ and a lower bound on utility given by $r \in \mathbb{R}_{>0}$.
QUESTION: For the utility function $u(\cdot)$ given in equation (2.1), does there exist a $G = (N, E) \in G^N$ such that

$$u(G|G_1) \geq r? \tag{2.2}$$

The following theorem is one of our main results in this chapter and shows that finding a BR with respect to an arbitrary graph with arbitrary cost and nonincreasing benefit functions does not have a polynomial-time solution, unless the answer to the long-standing open question of whether P = NP is affirmative.

Theorem 1. *BRN is NP-hard.*

We will develop the proof of Theorem 1 over the rest of this section. We will require some intermediate properties of best response networks, given by the following results.

2.1.1 Some Properties of Best Response Networks

Lemma 1. *If $G_2 = (N, E_2)$ is a BR network to $G_1 = (N, E_1)$, then the number of edges in G_2 is less than or equal to the number of edges in G_1. If $b(1) > b(2)$, then G_1 and G_2 have an equal number of edges if and only if $G_2 = G_1$.*

Proof. We use contradiction to prove the first part. Suppose that G_2 is a BR and has more edges than G_1. Then

$$\begin{aligned} u(G_2|G_1) &= \sum_{(u,v)\in E_1} b(d_{G_2}(u,v)) - c|E_2| \\ &\leq |E_1|b(1) - c|E_2| \\ &< |E_1|b(1) - c|E_1| = u(G_1|G_1), \end{aligned}$$

which contradicts our assumption that G_2 is a BR to G_1. To prove the second part, note that if $G_2 = G_1$ then the number of edges in G_2 and G_1 are equal. So we only need to show that when $b(1) > b(2)$, if the number of edges in G_2 is equal to the number of edges in G_1, then $G_2 = G_1$. If $G_2 \neq G_1$, then there exists a $(u, v) \in E_1$ such that $d_{G_2}(u, v) \geq 2$. Thus

$$\begin{aligned} u(G_2|G_1) &= \sum_{(u,v)\in E_1} b(d_{G_2}(u,v)) - c|E_2| \\ &< |E_1|b(1) - c|E_1| = u(G_1|G_1), \end{aligned}$$

contradicting the assumption that G_2 is a BR to G_1. □

The next lemma discusses the connectivity of BR networks.

Lemma 2. *Suppose that G_2 is a best response network to G_1 and $b(1) > c$. Then any two nodes that are connected by a path in G_1 will also be connected by a path in G_2. Specifically, if G_1 is connected, then G_2 must be connected.*

Proof. Let u and v be two nodes that are neighbors in G_1. By way of contradiction assume that there is no path between u and v in the BR network $G_2 = (N, E_2)$. For $G_2' = (N, E_2')$ with $E_2' = E_2 \cup \{(u, v)\}$,

$$u(G_2'|G_1) - u(G_2|G_1) \geq b(1) - c > 0,$$

contradicting the assumption that G_2 is a BR network. Now consider the case that u and v are connected through a path in G_1. Then there must be a path from u to v in G_2, since we showed that any two nodes that are directly connected in G_1 remain connected in G_2. □

Remark 3. *When $b(1) = c$, the above proof can be applied to show that there exists a best response network in which any two nodes that are connected by a path in G_1 will also be connected by a path in G_2 (although this does not have to be true of every best response network).*

For any integer $t \geq 1$, a subgraph $H = (N, E_H)$ of $G_1 = (N, E_1)$ is called a *t-spanner* if $d_H(x, y) \leq t$ for all $(x, y) \in E_1$, i.e., the distance between each pair of nodes that are neighbors in G_1 is not more than t in H [14]. A subgraph $T = (N, E_T)$ of the graph G_1 that is both a t-spanner and a tree is called a *tree t-spanner*. The following important lemma characterizes a BR to graphs that have a 2-spanner.

Lemma 3. *Suppose graph $G_1 = (N, E_1)$ has a spanning forest*[1] *$F = (N, E_F)$ that is also a 2-spanner. Assume that $b(1) - b(2) \leq c \leq b(1)$. Then F is a BR to G_1.*

Proof. Assume that G_1 has m components where $m \geq 1$. Since F is a spanning forest, $|E_F| = |N| - m$. Using the fact that $d_F(x, y) \leq 2$ for all $(x, y) \in E_1$, we have

$$u(F|G_1) = (|N| - m)(b(1) - c) + (|E_1| - (|N| - m))b(2). \tag{2.3}$$

Now assume that $H = (N, E_H)$ is a best response network to G_1 such that any two nodes that are connected in G_1 are also connected in H. The existence of such a BR network is guaranteed by Lemma 2 and Remark 3. Thus $|E_H| \geq |N| - m$. Also by Lemma 1, we have $|E_H| \leq |E_1|$. Since at most $|E_H|$ pairs of neighbors in G_1 can be directly connected in H, the remaining $|E_1| - |E_H|$ pairs of neighbors in G_1 will be at least a distance of two away from each other in H. Thus we have

$$\begin{aligned} u(H|G_1) &\leq |E_H|(b(1) - c) + (|E_1| - |E_H|)b(2) \qquad (2.4)\\ &= (|N| - m)(b(1) - c) + (|E_H| - (|N| - m))(b(1) - c) + (|E_1| - |E_H|)b(2) \\ &\leq (|N| - m)(b(1) - c) + (|E_1| - (|N| - m))b(2) \\ &= u(F|G_1). \end{aligned}$$

Thus F is a BR to the network G_1. □

[1]Whenever G_1 is a connected network, by a spanning forest of G_1 we mean a spanning tree.

The next lemma provides lower and upper bounds on the utility of BR networks when $b(1) - b(2) \leq c \leq b(1)$.

Lemma 4. *Suppose that $b(1) - b(2) \leq c \leq b(1)$ and $G_2 = (N, E_2)$ is a BR network with respect to an arbitrary connected network $G_1 = (N, E_1)$. Then*

$$|E_1|(b(1) - c) \leq u(G_2|G_1) \leq (|N| - 1)(b(1) - c) + (|E_1| - |N| + 1)b(2). \qquad (2.5)$$

Proof. The lower bound follows from the fact that $u(G_2|G_1) \geq u(G_1|G_1) = |E_1|(b(1) - c)$, by virtue of G_2 being a BR network. For the upper bound, note that since $b(1) \geq c$, G_2 can be assumed to be a connected graph (by Lemma 2 and Remark 3) and thus $|E_2| \geq |N| - 1$. The rest of the proof follows the same procedure as in the proof of Lemma 3 with $m = 1$. ☐

Remark 4. *The inequalities given in the above lemma are sharp. As we will show later, a BR to a tree is the same tree if $b(1) \geq c$. For a tree, the left and right hand sides of inequality (2.5) are equal. Also, for a graph G_1 with a tree 2-spanner T, we know that T is a BR to G_1 by Lemma 3 with utility equal to the right hand side of inequality (2.5).*

2.1.2 Proof of NP-Hardness of the BRN Problem

We now return to the BRN problem (Definition 3) and the claim of NP-hardness given in Theorem 1. To prove this theorem, we will construct a reduction from the *Tree t-spanner Problem* [14], defined below.

Definition 4. Tree t-Spanner (TtS) Problem.
INSTANCE: A connected graph $G = (N, E)$ and a positive integer t.
QUESTION: Does G have a tree t-spanner, i.e., a subgraph $T = (N, E_T)$ such that $|E_T| = |N| - 1$ and $d_T(x, y) \leq t$ for all $(x, y) \in E$?

The TtS problem is in P for $t = 2$, but NP-complete for all $t \geq 4$; the complexity of the problem for $t = 3$ is still unknown [14]. We are now in place to prove Theorem 3.

Proof of Theorem 1. We will construct a reduction from the TtS problem to the BRN problem, which will then imply that the BRN problem is NP-hard. Consider an instance of the TtS problem with graph $G = (N, E)$ and $t = 4$. Any spanning tree of G with $|N| \leq 5$ is a tree 4-spanner which is easy to find. Thus, we assume that $|N| \geq 6$. Define the corresponding instance of the

BRN problem as follows. The network $G_1 = (N, E_1)$ is the same as the graph G, i.e., $G_1 = G$. The benefit function $b(\cdot)$ and edge-cost c are chosen to satisfy

$$\begin{aligned} b(1) > b(2) = b(3) = b(4) > b(5), \\ b(1) - b(2) < c < b(1). \end{aligned} \tag{2.6}$$

For example $c = 2$, $b(1) = 3$, $b(2) = b(3) = b(4) = 2$ and $b(k) = 0\ \forall k \geq 5$ satisfies these conditions. Finally set

$$r = (|N| - 1)(b(1) - c) + (|E_1| - (|N| - 1))b(2). \tag{2.7}$$

Clearly we can construct the above BRN instance in polynomial time. Now assume that the answer to the instance of the TtS problem is "yes", i.e., graph G has a tree 4-spanner $T = (N, E_T)$. This means that T is a subtree of G_1 and $d_T(x, y) \leq 4$ for all $(x, y) \in E_1$. Thus we have that

$$\begin{aligned} u(T|G_1) &= \sum_{(x,y)\in E_1 \setminus E_T} b(d_T(x, y)) + (|N| - 1)(b(1) - c) \\ &= (|E_1| - (|N| - 1))b(2) + (|N| - 1)(b(1) - c) \\ &= r. \end{aligned}$$

Note that we used the fact that $b(2) = b(3) = b(4)$ to go from the first line to the second line in the above equation. Therefore, the answer to the defined instance of the BRN problem is also "yes".

To complete the proof, we have to show that if the answer to the constructed instance of the BRN is "yes", then the answer to the instance of the TtS is "yes". In other words, we have to show that if there exists a graph $G_2 = (N, E_2)$ such that

$$u(G_2|G_1) = \sum_{(x,y)\in E_1} b(d_{G_2}(x, y)) - c|E_2| \geq r,$$

where $b(\cdot)$ and c satisfy (2.6) and r is given by (2.7), then G_1 has a tree 4-spanner. We claim that any G_2 with utility at least r must be a tree 4-spanner of G_1.

Assume that $G_2 = (N, E_2)$ is a graph with $u(G_2|G_1) \geq r$. Since r is equal to the upper bound of the utility of the BR (by Lemma 4), G_2 must be a best response to G_1. Since $b(1) > c$, by Lemma 2 we know that G_2 is a connected graph. Therefore, $|E_2| \geq |N| - 1$. First consider the case that $|E_2| > |N| - 1$. Then similar to equation (2.4), we have that

$$\begin{aligned}
u(G_2|G_1) &\le |E_2|(b(1) - c) + (|E_1| - |E_2|)b(2) \\
&= (|N| - 1)(b(1) - c) + (|E_2| - (|N| - 1))(b(1) - c) + (|E_1| - |E_2|)b(2) \\
&< (|N| - 1)(b(1) - c) + (|E_2| - (|N| - 1))b(2) + (|E_1| - |E_2|)b(2) \\
&= (|N| - 1)(b(1) - c) + (|E_1| - (|N| - 1))b(2) = r,
\end{aligned}$$

which is a contradiction. Thus consider the case that $|E_2| = |N| - 1$, i.e., G_2 is tree. Denoting $|E_2 \cap E_1| = \gamma$, we have

$$\begin{aligned}
u(G_2|G_1) &= \gamma(b(1) - c) - (|N| - 1 - \gamma)c + \sum_{(x,y) \in E_1 \setminus E_2} b(d_{G_2}(x, y)) \qquad (2.8) \\
&\le \gamma(b(1) - c) - (|N| - 1 - \gamma)c + (|E_1| - \gamma)b(2) \\
&= \gamma(b(1) - c) + (|N| - 1 - \gamma)(b(2) - c) + (|E_1| - (|N| - 1))b(2).
\end{aligned}$$

If $\gamma < |N| - 1$, since $b(1) - c > b(2) - c$, by equation (2.8) we have that

$$u(G_2|G_1) < \gamma(b(1) - c) + (|N| - 1 - \gamma)(b(1) - c) + (|E_1| - (|N| - 1))b(2) = r,$$

which is again a contradiction. Therefore, $|E_2 \cap E_1| = \gamma = |N| - 1$. This means that G_2 is a subtree of G_1. Now if there exists $(u, v) \in E_1$ such that $d_{G_2}(u, v) > 4$, then we have

$$\begin{aligned}
u(G_2|G_1) &= (|N| - 1)(b(1) - c) + \sum_{(x,y) \in E_1 \setminus E_2} b(d_{G_2}(x, y)) \\
&< (|N| - 1)(b(1) - c) + (|E_1| - (|N| - 1))b(2) \\
&= r,
\end{aligned}$$

where the last inequality follows from the fact that $b(2) = b(3) = b(4) > b(d)$ for all $d > 4$. Therefore, for all $(u, v) \in E_1$, $d_{G_2}(u, v) \le 4$ which means that G_2 must be a tree 4-spanner for the graph G_1. Thus the answer to the instance of the TtS problem is "yes". This shows that the NP-hard problem TtS (for $t = 4$) is polynomial-time reducible to BRN, and therefore BRN is NP-hard. □

Remark 5. *Deriving approximation algorithms with provable performance guarantees is a natural approach to dealing with the inherent complexity of finding best response networks; a deeper investigation of the connections between t-spanners and the best response network design problem might lead to such algorithms. This is left as a venue for future work.*

There are certain NP-hard optimization problems (e.g., minimum vertex cover) whose solutions can be approximated to within a constant factor by simple greedy algorithms [19]. The following example considers a natural greedy algorithm where edges are added or removed one at a time, and shows that this algorithm can produce results that are arbitrarily far away from the optimal network.

Example 2. *Consider a greedy algorithm where at each step, we add or remove a link that provides the highest increase in the utility until no further improvements can be made. The following scenarios illustrate the pitfalls of such an algorithm.*

Consider a reference network G_1. Suppose we attempt to build a BR network by starting with an empty network G and repeatedly adding edges. If $b(1) < c$, then adding any single edge to G will result in negative utility, and thus the algorithm stops with the empty network. Since there can exist nonempty BR networks when $b(1) < c$ whose utility is unbounded in n (e.g., see Proposition 1), the network produced by the above algorithm can be arbitrarily bad in comparison to the true BR network.

Now suppose that we attempt to build a BR network by starting with the reference network G_1 and removing edges one at a time. Consider the graph G_1 depicted in Figure 2.3a and define $c = 1$, $b(1) = \frac{n-1}{n-2}$, $b(2) = 0.5$, $b(k) = 0$ for $3 \leq k \leq n-1$.

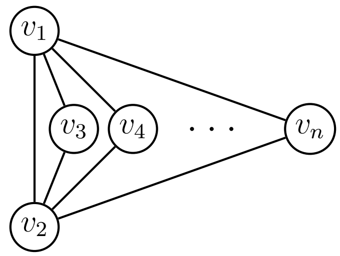
Starting with G_1, removing any of the edges increases the utility by $b(2)-(b(1)-c)$. Thus any edge is a candidate for removal. Consider removing the edge (v_1, v_2) which results in network G_2. Now no further improvements are possible by adding or removing a single edge. Next, consider network G_3 shown in the Figure 2.3c. As we will show in Proposition 3 in Section 2.3, G_3 is a best response network to G_1. We have

$$\lim_{n\to\infty} \frac{u(G_3|G_1)}{u(G_2|G_1)} = \lim_{n\to\infty} \frac{(n-1)(b(1)-c)+(n-2)b(2)}{2(n-2)(b(1)-c)+b(2)} = \infty.$$

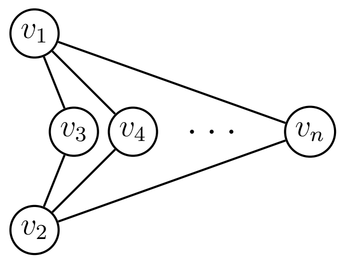
Note that same conclusion is reached even if we start with the complete graph, i.e., we can remove the edges in such a way that we end up in network G_2. Thus this greedy algorithm can perform arbitrarily poorly in comparison to the optimal solution.

2.1.3 Comparison to Other Network Design Problems

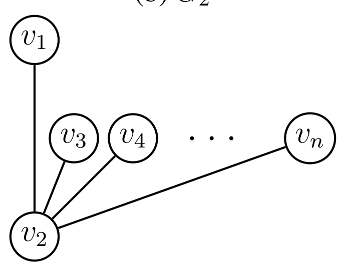
Here, we compare the BRN problem to two canonical network design problems (defined in Chapter 1) that also attempt to minimize distances between pairs of nodes: the Optimal Communication Spanning Tree (OCST) problem introduced in [45], and the Simple Network Design problem (SNDP) introduced in [53].



(a) G_1



(b) G_2



(c) G_3

Figure 2.3: Performance of a greedy algorithm. Graph G_1 in (a) is the reference network. Graph G_2 in (b) is the output of the greedy algorithm discussed above. Graph G_3 in (c) is a best response to G_1.

The relationships between the BRN, OCST and SNDP problems are as follows.

- The OCST and SNDP problems explicitly constrain the number of edges in the designed network, whereas the BRN problem includes the cost of edges in the utility function.
- The SNDP problem requires the designed network to be a subgraph of another given network, whereas the BRN and OCST problems place no such constraint.

- The objective of the SNDP problem is to minimize the sum of distances between *all* pairs of nodes, whereas the BRN and OCST problems allow the objective function to only depend on distances between selected pairs of nodes (the OCST problem does this by setting $r_{ij} = 0$ for those pairs that do not wish to communicate).

Despite the apparent similarities between the BRN problem and the OCST problem, Theorem 1 shows that the BRN problem is NP-hard, even though the OCST problem can be solved in polynomial-time. This increase in complexity is a byproduct of the additional flexibility afforded by the general nonincreasing benefit function in the BRN problem (as opposed to the scaled distances in the utility function for the OCST problem), which allows it to capture the tree-t-spanner problem as a special case.

2.1.4 Possible Extensions of the BRN Problem

There are various extensions for the BRN problem that can be considered. For instance, it is possible to consider a set of weights w_{ij} for the benefits produced from connections between different pairs of nodes $(i, j) \in E_1$. Formally, we can generalize the utility function (2.1) as

$$u_w(G|G_1) = \sum_{(v_i,v_j)\in E_1} w_{ij} b(d_G(i,j)) - c|E|. \tag{2.9}$$

We can also define a version of the problem with an edge-dependent cost, i.e., considering a utility function with the form

$$u_c(G|G_1) = \sum_{(v_i,v_j)\in E_1} b(d_G(i,j)) - c_{ij}|E|. \tag{2.10}$$

Theorem 1 immediately implies that the above versions of the BRN problem are also NP-hard.

Another interesting variant of the BRN problem is by allowing a fixed number of edges for the network $G = (N, E)$ in Definition 3, i.e., adding the condition $|E| = k$. Through a proof similar to the proof of Theorem 1, we can show that this version of the BRN problem is also NP-hard (by setting $k = |N| - 1$ in the proposed reduction from the 4-spanner problem).

In the next section, we will characterize further properties of BR networks; these will allow us to find BR networks with respect to certain specific classes of graphs, which in turn will allow us to formulate and study a multi-layer network formation setting with multiple network designers.

2.2 Further Properties of Best Response Networks

We start with the following useful result describing the relationship between the components of BR networks and the reference network.

Lemma 5. *Let G_2 be a BR network to G_1, and suppose that G_2 is not connected. Let $G_{2i} = (N_i,\ E_{2i})$, $i = 1, \dots, k$, be the components of G_2. Let $G_{1i} = (N_i,\ E_{1i})$, $i = 1, 2, \dots, k$, be the subgraphs induced by vertex sets N_i on G_1. Then network G_{2i} must be a BR network to G_{1i} for $i = 1, 2, \dots, k$.*

Proof. Consider the utility of network G_2 with respect to G_1. Since there are no edges between the components in G_2, for any $(u, v) \in E_1$ with u and v in different components of G_2, $d_{G_2}(u, v) = \infty$. Thus $\sum_{(u,v)\in E_1} b(d_{G_2}(u, v)) = \sum_{i=1}^{k} \sum_{(u,v)\in E_{1i}} b(d_{G_{2i}}(u, v))$, and the utility function can be written as

$$\begin{aligned} u(G_2|G_1) &= \sum_{(u,v)\in E_1} b(d_{G_2}(u, v)) - c|E_2| \\ &= \sum_{i=1}^{k} \left(\sum_{(u,v)\in E_{1i}} b(d_{G_{2i}}(u, v)) - c|E_{2i}| \right) \\ &= u(G_{21}|G_{11}) + \cdots + u(G_{2k}|G_{1k}). \end{aligned}$$

Now, if G_{2i} is not a BR to G_{1i} for some $i \in \{1, 2, \dots, k\}$, replace it with a BR. This will increase the utility, contradicting the fact that G_2 is a BR. □

The following lemma considers the case when there are isolated nodes in G_1.

Lemma 6. *Let $G_1 = (N, E_1)$ and suppose $v \in N$ is an isolated node. Then v is isolated in any BR to G_1.*

Proof. Let $G_2 = (N, E_2)$ be a BR network with respect to G_1, and suppose by way of contradiction that v is not isolated in G_2. If v is a leaf node in G_2 (i.e., it has a single neighbor), then the edge incident to v is not used in any of the shortest paths between nodes in $N \setminus \{v\}$. Removing that edge increases the utility of G_2 by c, contradicting the fact that it is a BR.

Now suppose that v has two or more neighbors in G_2, and denote those neighbors by the set $J = \{v_{j_1}, v_{j_2}, \dots, v_{j_l}\} \subseteq N \setminus \{v\}$ with $l \geq 2$. Construct a new network $G_3 = (N, E_3)$ with

$$E_3 = (E_2 \setminus \{(v, v_{j_1}), (v, v_{j_2}), \dots, (v, v_{j_l})\}) \cup \{(v_{j_1}, v_{j_2}), (v_{j_1}, v_{j_3}), \dots, (v_{j_1}, v_{j_l})\}, \tag{2.11}$$

i.e., we remove the l edges from v to its neighbors and add edges from $v_{j_1} \in J$ to the other nodes in J. This results in a net removal of at least one edge from the graph. Suppose that the shortest path between some pair of nodes in $N \setminus \{v\}$ passed through v in G_2; the shortest path now passes through v_{j_1} in G_3, and is at least as short as the original shortest path. Thus $u(G_3|G_1) > u(G_2|G_1)$ which contradicts the assumption that G_2 is a BR network to G_1. Therefore, v must be an isolated node in G_2. □

The properties described above are independent of the relative values of the benefit function and edge costs. The following set of results provide more details of the BR networks for certain ranges of benefits and costs.

Lemma 7. *Let $G_1 = (N, E_1)$ be an arbitrary graph.*

1. *If $b(1) - c > b(2)$, then the unique BR network to G_1 is $G_2 = G_1$.*
2. *If $b(1) < c$, then G_1 is not a BR network to G_1, unless G_1 is the empty network.*
3. *Define*
$$\alpha \triangleq \max_{2 \leq |S|, S \subseteq N} \frac{|E_{G_1}(S,S)|}{|S|-1} - 1, \tag{2.12}$$
where $E_{G_1}(S,S)$ denotes the set of edges in G_1 that have both of their endpoints in the set S, i.e., $E_{G_1}(S,S) = E_1 \cap (S \times S)$. If $c > b(1) + \alpha b(2)$, then the unique BR network with respect to G_1 is the empty network.

Proof. In order to prove the first property, assume by way of contradiction that G_2 is a BR network and $G_2 \neq G_1$. Since $b(1) > b(2)$, by Lemma 1, we know that the number of edges in G_2 is less than in G_1. So there are vertices u and v such that $(u, v) \in E_1$ and $d_{G_2}(u, v) > 1$. Adding the edge (u, v) to E_2 increases the utility by at least $b(1) - c - b(2) > 0$ which contradicts the assumption that $G_2 \neq G_1$ is a BR network. Therefore, the BR network must be equal to G_1.

For the second property note that if $G_2 = G_1 \neq \phi$, then $u(G_2|G_1) = |E_1|(b(1) - c) < 0$ due to the assumption that $b(1) < c$. Thus it must be the case that $G_2 \neq G_1$, or G_1 is the empty network.

Finally in order to prove the third property, consider an arbitrary graph $G_1 = (N, E_1)$ with n nodes. By way of contradiction assume that $G_2 \neq \phi$ is a BR network with respect to G_1. Let $G_{21} = (N_1, E_{21})$ be a component of network G_2 with $1 < |N_1| \leq n$. By Lemma 5, we know that G_{21} must be a BR to the subgraph induced by the node set N_1 on G_1, which we denote by $G_{11} = (N_1, E_{11})$. Thus

$$\begin{aligned} u(G_{21}|G_{11}) &\leq |E_{21}|(b(1) - c) + (|E_{11}| - |E_{21}|)b(2) \\ &= |E_{21}|(b(1) - c + \alpha b(2)) + (|E_{11}| - |E_{21}|(1+\alpha))b(2) \\ &= |E_{21}|(b(1) - c + \alpha b(2)) + |E_{21}| \left(\frac{|E_{11}|}{|E_{21}|} - (1+\alpha) \right) b(2). \end{aligned} \tag{2.13}$$

Due to the assumption that $c > b(1) + \alpha b(2)$, the first term in (2.13) is negative. Also, we have that

$$\frac{|E_{11}|}{|E_{21}|} \leq \frac{|E_{11}|}{|N_1| - 1} = \frac{|E_{G_1}(N_1, N_1)|}{|N_1| - 1} \leq \max_{2 \leq |S|, S \subseteq N} \frac{|E_{G_1}(S, S)|}{|S| - 1} = \alpha + 1.$$

The first inequality above follows from the fact that G_{21} is a component and thus has at least $|N_1| - 1$ edges. Thus the second term in equation (2.13) is nonpositive. Therefore, $u(G_{21}|G_{11}) < 0$ which is a contradiction. As a result G_{21} (and thereby G_2) must be the empty network. □

The parameter α is a measure of the edge density of the underlying graph, and thus the threshold to have the empty network as the best response network increases as the underlying graph becomes more dense. The following example illustrates the implication of α for various graphs.

Example 3. *In the following, we define $|N| = n$.*

- *Assume that $G_1 = (N, E_1)$ is the complete graph. Then $|E_{G_1}(S, S)| = \binom{|S|}{2}$ for any (non-singleton) $S \subseteq N$ and thus $\alpha = \frac{n-2}{2}$ in equation (2.12). This means that the BR to the complete graph is the empty graph for $c > b(1) + \frac{n-2}{2} b(2)$, yielding part (iii) of Proposition 1 (obtained in [48]) as a special case of Lemma 7.*

- *Suppose that $G_1 = (N, E_1)$ is a tree. Since any induced subgraph of a tree is a forest (it is a tree when the subgraph is connected), we have $|E_{G_1}(S, S)| \leq |S| - 1$ for any non-singleton $S \subseteq N$. Thus*

$$\frac{|E_{G_1}(S, S)|}{|S| - 1} - 1 \leq 0 \quad \forall S \subseteq N, \ |S| \geq 2.$$

This means that $\alpha = 0$ (which happens for any S that induces a connected subgraph on G_1). Therefore, we can conclude that the BR network to a tree is the empty network when $c > b(1)$.

- *Consider a cycle graph $G_1 = (N, E_1)$ with n nodes.*[2] *Any induced subgraph of G_1 on a non-singleton node set $S \subset N$ is an acyclic graph and thus $|E_{G_1}(S,S)| \leq |S| - 1$. For $S = N$, we have $|E_{G_1}(N,N)| = n$. Thus $\alpha = \frac{1}{n-1}$, and the BR network to G_1 is the empty network for $c > b(1) + \frac{1}{n-1}b(2)$.*

Remark 6. *Note that when $b(2) = 0$, Lemma 7 indicates that for $b(1) > c$, G_1 is a unique BR to itself (for any network G_1), and when $b(1) < c$, the empty network is a unique best response (both G_1 and the empty network are best responses with utility 0 when $b(1) = c$). Thus, in the rest of this section, we will assume that $b(2) > 0$.*

In the next lemma, we consider the case that we have nodes with degree one in the graph.

Lemma 8. *Let $G_1 = (N, E_1)$, and suppose $v \in N$ is a leaf node. Define the induced subgraph of G_1 under the node set $N \setminus \{v\}$ as $G_{11} = (N \setminus \{v\}, E_{11})$ (i.e., the graph obtained by removing node v and its incident edge). Then a BR to G_1 can be obtained by first finding a BR to G_{11} and then adding v as an isolated node if $b(1) \leq c$, or adding v together with a single edge to its neighbor in G_1 if $b(1) \geq c$.*

Proof. Let the neighbor of v in G_1 be denoted by u, and assume that network $H = (N, E_H)$ is a BR to network G_1. We reason as we did in the proof of Lemma 6, with a few additional details.

Consider the case that $b(1) \leq c$. Suppose that node v is not isolated in H, and let $J = \{v_{j_1}, v_{j_2}, \ldots, v_{j_l}\} \subseteq N \setminus \{v\}$ be the neighbors of v in H. If $l = 1$ (i.e., v has a single neighbor in H), the edge (v, v_{j_1}) is not used in any of the shortest paths between nodes in $N \setminus \{v\}$. Removing that edge saves a cost of c, and loses at most a benefit of $b(1)$ (due to the loss of the path from v to u in H). Since $b(1) \leq c$, the resulting graph has utility at least as large as H.

Now suppose $l > 1$. Construct the new network $H_1 = (N, E_{H_1})$ with edge set

$$E_{H_1} \triangleq (E_H \setminus \{(v, v_{j_1}), (v, v_{j_2}), \ldots, (v, v_{j_l})\}) \cup \{(v_{j_1}, v_{j_2}), (v_{j_1}, v_{j_3}), \ldots, (v_{j_1}, v_{j_l})\}. \quad (2.14)$$

In other words, we remove all of the incident edges from v in H and add edges from each node in $J \setminus \{v_{j_1}\}$ to v_{j_1}. This saves at least one edge, and $d_{H_1}(x, y) \leq d_H(x, y)$ for all $(x, y) \in E_1$. Thus, the only drop in utility in graph H_1 arises from the loss of the path from node v to u. Again, since $b(1) \leq c$, the graph H_1 has utility at least equal to the utility of the network H and thus H_1 is also a best response. The above two cases show that when $b(1) \leq c$, there exists a best response where the leaf node v is isolated.

Now consider the case where $b(1) \geq c$. Then by Lemma 2 and Remark 3, there exists a BR network $H = (N, E_H)$ containing a path from v to u. If v is a leaf node in H, it is

[2] A cycle graph with n nodes consists of only one cycle of length n.

straightforward to show that there exists a BR network H' where v is connected to u. Thus suppose v is connected to the node set $J = \{v_{j_1}, v_{j_2}, \ldots, v_{j_l}\} \subseteq N \setminus \{v\}$ in H, with $l \geq 2$. Construct a new graph $H_2 = (N, E_{H_2})$, where $E_{H_2} \triangleq E_{H_1} \cup \{(v,u)\}$ with E_{H_1} as defined in (2.14). Arguing as above, the utility of H_2 is at least as high as the utility of H, and thus H_2 is a BR to G_1. Since the edge (v,u) cannot be in the shortest path between any pair of nodes in $N \setminus \{v\}$, we see that the subgraph of H_2 induced by $N \setminus \{v\}$ must be a best response to the corresponding subgraph of G_1. This proves the result. □

The above lemma provides the following method to simplify the task of finding a best response network. Given a graph G_1, we recursively remove nodes of degree 1 until we are left with a graph where all nodes have degree two or larger (this is known as *peeling* the graph, and the resulting subgraph is known as a 2-core [68]). A best response to the 2-core can then be found using whatever means necessary, and then the removed nodes can be recursively added back as isolated nodes (if $b(1) \leq c$), or with the single edge that was removed (if $b(1) \geq c$). This process is illustrated in the following example.

Example 4. *Consider the network G_1 shown in Figure 2.4a. Suppose that*

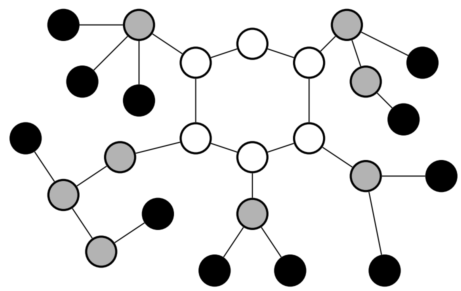
$$b(1) = 1.01,\ b(2) = 0.85,\ b(3) = 0.8,\ b(4) = 0.2,\ b(5) = 0.1$$
$$b(k) = 0\ \ \forall k \geq 6$$
$$c = 1.$$

The are 2^{300} possible candidates for best response to G_1. By Lemma 1, we can decrease the number of candidates to 33554432. However, finding a BR by brute-force search will still take hours.

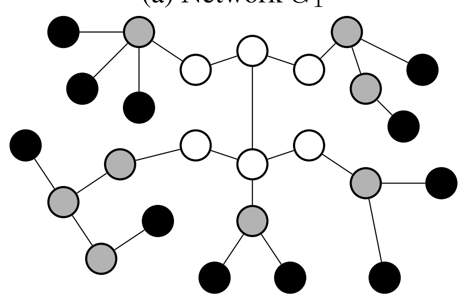
We use Lemma 8 to simplify the search for a best response network to G_1. The first step is to remove all of the leaf nodes (colored black) from G_1; this brings us to a new network containing only the gray and white nodes. Recursively removing leaf nodes, we reach the cycle network containing the white nodes; this is the 2-core of graph G_1. A best response to this cycle for the above edge cost and benefit function was found in Example 1 to be the network in Figure 2.2c. Now since $b(1) > c$, we can add the leaf nodes that we removed in each step, leading to the BR network G_2 shown in Figure 2.4b.

2.3 Best Responses to Specific Networks

We will now apply the above results to characterize best responses to acyclic networks and networks with a star subgraph. The latter models, for example, sensor or communication networks



(a) Network G_1



(b) Network G_2

Figure 2.4: Finding a best response network to G_1 using Lemma 8. The leaf nodes (starting with the black nodes) are recursively removed from G_1 until only the cycle containing the white nodes remains. A best response to the cycle is then found and the other nodes are recursively added back to obtain the best response network G_2.

where one or more base stations or fusion centers wish to communicate with all nodes, while the other nodes only need to communicate locally amongst themselves.

Proposition 2. *Let* $G_1 = (N, E_1)$ *be a forest.*

- *If* $b(1) < c$*, the empty network is the unique BR to* G_1*.*
- *If* $b(1) > c$*, then* $G_2 = G_1$ *is a BR network to* G_1*.*
- *If* $b(1) = c$*, the empty network and* $G_2 = G_1$ *are both BR networks to* G_1*.*
- *For* $b(1) > \max\{b(2), c\}$*, the unique BR to* G_1 *is* $G_2 = G_1$*.*

Proof. When $b(1) < c$, we use part 3 of Lemma 7. Following the same argument as in Example 3 for trees, we have $\alpha = 0$ for G_1. Thus the unique BR network to a forest is the empty network when $b(1) < c$.

For $b(1) - b(2) > c$, the unique best response to any network is the same network by the first part of Lemma 7. For $b(1) - b(2) \leq c \leq b(1)$, note that G_1 is a 2-spanner forest of itself, and thus G_1 is a BR to itself by Lemma 3, proving the second statement. Since this BR has a utility of zero when $b(1) = c$, the empty network is also a BR for this value of c, proving the third statement.

Finally, we prove the uniqueness of the BR when $b(1) > \max\{b(2), c\}$. If G_1 has r connected components, then $|E_1| = |N| - r$. By Lemma 2, we must have $|E_2| \geq |N| - r$. By Lemma 1, we know that $|E_2| \leq |E_1| = |N| - r$. Thus $|E_2| = |E_1|$ and since $b(1) > b(2)$, we have $G_2 = G_1$. □

Proposition 3. *Let $G_1 = (N, E_1)$ be a graph that has a star subgraph centered at node $v \in N$.*

- *If $b(1) - b(2) > c$, then G_1 is the unique BR to G_1.*
- *If $b(1) - b(2) \leq c \leq b(1)$, then the star network centered at node v is a BR network to G_1.*
- *If $b(1) \leq c$, one of the following networks is a BR to G_1:*
    1. *A star network on N with center at node v.*
    2. *A network where one component is a star and all other components are isolated nodes.*
    3. *The empty network.*

Proof. The first statement is a direct result of Lemma 7.

In order to prove the second statement we use Lemma 3. Let G^s be the star network centered at node v. Since G^s is a 2-spanner tree of G_1, it is a BR to G_1.

Next, we prove the third statement. Define G^s as the star network centered at node v. By equation (2.1), we have

$$u(G^s|G_1) = (|N| - 1)(b(1) - c) + (|E_1| - (|N| - 1))b(2). \tag{2.15}$$

Now assume that $G_2 = (N, E_2)$ is a BR network. Using the same argument as in equation (2.4), we have

$$u(G_2|G_1) \leq |E_2|(b(1) - c) + (|E_1| - |E_2|)b(2). \tag{2.16}$$

Using equations (2.15) and (2.16) we obtain

$$u(G^s|G_1) - u(G_2|G_1) \geq (|E_2| - (|N| - 1))(b(2) - b(1) + c). \tag{2.17}$$

According to the assumption of the Proposition, $c - b(1) \geq 0$ and thus the right hand side of equation (2.17) is nonnegative for all $|E_2| \geq |N| - 1$. Therefore, the utility of G^s with respect to G_1 is as high as any other connected network.

Thus assume that G_2 is a non-empty disconnected network. Suppose that it has γ components $G_{2k} = (N_k, E_{2k})$ for $k \in \{1, 2, \cdots, \gamma\}$. Denote by $G_{1k} = (N_k, E_{1k})$, $k \in \{1, 2, \cdots, \gamma\}$, the subgraphs induced by N_k on G_1. Without loss of generality, let $v \in N_1$. Then, since G_{11} contains a star subgraph (centered on v), and G_{21} is a BR to G_{11} (by Lemma 5) and connected, we can take it to be a star by the above argument. Next, we aim to show that there exists a BR (constructed based on G_2) such that all of the components are isolated nodes except G_{21}.

Suppose that some component of G_2 (not containing v) has more than one node and take this component to be G_{22} without loss of generality. We know that G_{22} is a BR to G_{12} based on Lemma 5. Arguing as in equation (2.4), we have

$$u(G_{22}|G_{12}) \leq |E_{22}|(b(1) - c) + (|E_{12}| - |E_{22}|)b(2). \tag{2.18}$$

If G_{22} has zero utility, we can replace it by the empty network and subsequently, we have the result. Thus assume by way of contradiction that it has some positive utility. Therefore, the right hand side of equation (2.18) is positive. Since G_{22} is a connected network, $|E_{22}| \geq |N_2| - 1$. Hence

$$\frac{|E_{12}| - |E_{22}|}{|E_{22}|} \leq \frac{|E_{12}| - (|N_2| - 1)}{|N_2| - 1} \leq \frac{\binom{|N_2|}{2} - (|N_2| - 1)}{|N_2| - 1} < |N_2| - 1. \tag{2.19}$$

Using the assumption that the right hand side of inequality (2.18) is positive and by inequality (2.19), we have that

$$\begin{aligned} 0 &< |E_{22}|(b(1) - c) + (|E_{12}| - |E_{22}|)b(2) \\ &= |E_{22}| \left((b(1) - c) + \frac{|E_{12}| - |E_{22}|}{|E_{22}|} b(2) \right) \\ &< |E_{22}| \left(b(1) - c + (|N_2| - 1)b(2) \right). \end{aligned}$$

Now consider a graph $\hat{G}_2$ obtained by removing all edges of G_{22} and connecting all of its nodes to node v. Since $b(1) - c + (|N_2| - 1)b(2) > 0$ we have,

$$\begin{aligned}
u(\hat{G}_2|G_1) &\geq \sum_{i\neq 2} u(G_{2i}|G_{1i}) + |N_2|(b(1) - c) + |E_{12}|b(2) \\
&> \sum_{i\neq 2} u(G_{2i}|G_{1i}) + |N_2|(b(1) - c) + |E_{12}|b(2) - (b(1) - c + (|N_2| - 1)b(2)) \\
&= \sum_{i\neq 2} u(G_{2i}|G_{1i}) + (|N_2| - 1)(b(1) - c) + (|E_{12}| - (|N_2| - 1))b(2), \qquad (2.20)
\end{aligned}$$

where the first inequality follows from the fact that the induced subgraph of $N_1 \cup N_2$ on $\hat{G}_2$ is a connected network and we neglect the benefit (if any) from indirect connections between nodes in $N_1 \setminus \{v\}$ and N_2. The second term in the first inequality captures the direct benefits and costs of the $|N_2|$ edges from nodes in N_2 to v, and the third term captures the benefits due to each pair of nodes in N_2 having a distance of 2 from each other in $\hat{G}_2$ (via v). Next, note that

$$\begin{aligned}
u(G_{22}|G_{12}) &\leq |E_{22}|(b(1) - c) + (|E_{12}| - |E_{22}|)b(2) \\
&= (|N_2| - 1)(b(1) - c) + (|E_{22}| - (|N_2| - 1))(b(1) - c) + (|E_{12}| - |E_{22}|)b(2) \\
&\leq (|N_2| - 1)(b(1) - c) + (|E_{22}| - (|N_2| - 1))b(2) + (|E_{12}| - |E_{22}|)b(2) \\
&= (|N_2| - 1)(b(1) - c) + (|E_{12}| - (|N_2| - 1))b(2). \qquad (2.21)
\end{aligned}$$

Substituting inequality (2.21) in inequality (2.20), we have that

$$\begin{aligned}
u(\hat{G}_2|G_1) &> \sum_{i\neq 2} u(G_{2i}|G_{1i}) + (|N_2| - 1)(b(1) - c) + (|E_{12}| - (|N_2| - 1))b(2) \\
&\geq \sum_{i\neq 2} u(G_{2i}|G_{1i}) + u(G_{22}|G_{12}) = u(G_2|G_1).
\end{aligned}$$

However this is a contradiction to the assumption that G_2 is a BR to G_1. Thus all of the nonempty components of G_2 (except G_{21}) must have zero utility and therefore, we can replace each of them by the empty network. $\square$

2.4 Pairwise Stability

The objective of this section is to study the satisfaction of the individual *nodes* in the network with the decision of the central designer (that chooses the best response network). Specifically, let $G_1 = (N, E_1)$ be a given graph, and let U denote the set of all possible utility functions

$u(G_2|G_1)$ for graph G_2 based on G_1. For each $v_i \in N$, define the allocation rule $u_i(G_2, G_1, u) : G^N \times G^N \times U \to \mathbb{R}$ specifying the amount of utility that we allocate to player i from the overall utility generated by the formed network G_2. For simplicity, we will use the notation $u_i(G_2)$ when G_1 and u are fixed.

For a given best response graph G_2 and individual utility functions u_i, it may be the case that a certain node can improve its own utility by removing one or more of its incident edges in G_2, or by adding additional edges from itself to other nodes. As in [48], we assume any node can remove any of its incident edges unilaterally, but that adding an edge to another node requires the consent of that node. This motivates the following definition of *pairwise stability* of a given network [48]. In this definition, when $(v_i, v_j) \notin G$, $G + ij$ denotes the graph obtained by adding an edge between v_i and v_j in G. Similarly, $G - ij$ represents the graph obtained by deleting the edge (v_i, v_j) when $(v_i, v_j) \in G$.

Definition 5 ([48]). *A graph $G = (N, E)$ is said to be pairwise stable if*

- $\forall (v_i, v_j) \in E$, $u_i(G) \geq u_i(G - ij)$ *and* $u_j(G) \geq u_j(G - ij)$, *and*
- $\forall (v_i, v_j) \notin E$, *if* $u_i(G + ij) > u_i(G)$ *then* $u_j(G + ij) < u_j(G)$.

The graph is pairwise unstable if it is not pairwise stable.

In words, pairwise stability of a network corresponds to the situation where no node has any incentive to change any (one) of its connections in the network. This is a modification of the notion of a Nash equilibrium in network formation, capturing the concept of negotiation and agreement between the endpoints prior to forming the edge. Various versions of this notion have been studied in the network formation literature [46, 9, 47, 4, 44].

We now investigate the pairwise stability properties of best response networks that we characterized in the previous section. Consider the allocation rule

$$u_i(G_2|G_1) = \frac{1}{2} \sum_{(v_i,v_j) \in E_1} b(d_{G_2}(v_i, v_j)) - \frac{c}{2} deg_i(G_2), \tag{2.22}$$

where $deg_i(G)$ is the degree of node v_i in graph G. Note that the total utility (2.1) satisfies $u(G_2|G_1) = \sum_{i \in N} u_i(G_2|G_1)$ and thus this allocation rule is *budget balanced* [39].

It is not hard to show that for any $v_i, v_j \in N$ where $G_2 = (N, E_2)$, if $(v_i, v_j) \notin E_2$, then it would not be beneficial for at least one of the nodes v_i or v_j to add the edge (v_i, v_j) to the network G_2. By way of contradiction assume that $u_i(G_2 + ij) \geq u_i(G_2)$ and $u_j(G_2 + ij) \geq u_j(G_2)$ with one of the inequalities strict. Then, $u_i(G_2+ij)+u_j(G_2+ij) > u_i(G_2)+u_j(G_2)$. However, this means that $u(G_2+ij|G_1) > u(G_2|G_1)$ which contradicts the assumption that G_2 is the optimal network. This immediately implies that if the empty network is the best response to a graph, it is pairwise stable. For a general graph, to conclude that the best response network G_2 is pairwise stable, we also need to show that removing any of the edges from network G_2 is not beneficial for any of its endpoints. However, this is *not* true in general. As an example, consider Figure 2.2c, where the edge (v_1, v_4) in G_3 is only useful for connecting nodes other than its endpoints. This network G_3 is not stable, since both nodes v_1 and v_4 could improve their utility by removing this edge.

The following result provides a condition under which the best response network obtained from solving the optimization problem (2.1) is pairwise stable with respect to the allocation rule (2.22).

Proposition 4. *If $G_2 = G_1$ is a best response network with respect to the network G_1, then G_2 is pairwise stable.*

Proof. As argued above, adding an edge is not beneficial to any node. Thus it suffices to show that removing any of the edges is unrewarding for both of its endpoints. Since $G_2 = G_1$, edge (v_i, v_j) is only useful for the connection between nodes i and j. So if it is beneficial for one of the endpoints i or j to sever the link (v_i, v_j), it is also beneficial for the other endpoint. Consequently, removing the edge (v_i, v_j) increases the utility of G_2. This contradicts the fact that $G_2 = G_1$ is a best response to G_1.

Remark 7. *Note that the above result encompasses cases where G_1 is an arbitrary network and $b(1) - c > b(2)$ (by Lemma 7), and where G_1 is a forest (by Proposition 2).*

2.5 Summary

In this chapter, we started by strictly generalizing the distance-based utility function to design a network based on another network. We called the outcome of this network design procedure the "Best Response Network". We showed that the problem of finding a best response network with respect to an arbitrary network is NP-hard. Nevertheless, we characterized certain properties of best response networks, and found the optimal networks for certain cases of the reference graphs. Through an example, we showed that a natural greedy algorithm can produce solutions that are arbitrarily far from the optimal solution. In the last section of this chapter, we studied satisfaction of the nodes as the individuals with the decision of the central designer. We used the notion of pairwise stability to investigate whether nodes wish to make any changes in their incident edges. We provided an example to show that best response networks are not necessarily pairwise stable.

Chapter 3

Multi-Layer Network Formation Game

Using the notion of the best response network that we developed in the previous chapter, here we consider a scenario with multiple network designers, each of whom is building a different layer of the network. An example of this is when multiple transportation companies build their individual service networks among a group of cities, and each company prefers to provide service between pairs of cities that are not already covered by other companies. We capture these scenarios by defining a non-cooperative multi-layer network formation game where each player corresponds to a specific layer of the network. We develop a notion of distance-based multi-layer network formation based on *strategic substitutes*, where the presence of an edge in one layer makes it less desirable to have that edge in another layer. Despite the complexity of calculating best response networks, we characterize the Nash equilibrium networks that arise in this setting. In particular, we show that players with low costs for building edges drive out players that have relatively high costs, and that our framework gives rise to the "hub-and-spoke" networks commonly seen in various transportation systems [82].

As an alternative multi-layer network formation setting, we consider the case where the players have finite budgets and directly compete with each other to purchase edges from a common set. An example of this would be when multiple telecommunications companies allocate their budgets to improve their services between pairs of cities, with the company that spends the most on a given edge winning the business on that edge. To model these situations, we introduce a network formation game based on the classical *Colonel Blotto* game [37, 90, 81]. We consider utilities for players that depend on the component sizes and diameter of the formed networks, and characterize the range of player budgets for which the game admits pure Nash equilibria. The developed game also has applications to attack and defense problems in networks [41, 43].

3.1 A Multi-Layer Network Formation Game with Strategic Substitutes

We start by defining an m-player game where each player corresponds to one of the layers.

Definition 6. *A Multi-Layer Network Formation Game has a set of m players denoted by $P = \{P_1, P_2, \ldots, P_m\}$. The strategy space for each of the players is defined to be G^N where $N = \{v_1, v_2, \ldots, v_n\}$, i.e., the set of all graphs on node set N. For each $i \in \{1, 2, \ldots, m\}$, let $G_i = (N, E_i) \in G^N$ denote the action of player P_i. The utility of player P_i is given by a function $A_i : G^N \times G^N \times \cdots \times G^N \to \mathbb{R}$, where the j^{th} argument is the action of the j^{th} player for $1 \le j \le m$.*

We will use G_{-i} to denote the vector of actions of all players except player P_i, and use $A_i(G_i, G_{-i})$ to denote the utility of player P_i with respect to the given vector $(G_1, G_2, \ldots, G_m)$. Based on the definition of the game, we say that a vector of networks $(G_1, G_2, \ldots, G_m)$ is a *Nash equilibrium* if and only if $G_i \in \arg\max_G A_i(G, G_{-i})$ for all $i \in \{1, 2, \ldots, m\}$. In this case, G_i is said to be a BR network to G_{-i} with respect to the utility function A_i.

The characteristics of the game and the optimal strategies for each player will depend on the form of the utility functions A_i. Here, as a starting point for studying such games, we will focus on distance-based utilities (thereby building on our results from Chapter 2). The reference networks for the distance-based utility function for each player will depend on the networks constructed by the other players. In the remainder of this section, we will explore functions that view different layers of the network as *strategic substitutes*, where the presence of a link in one layer makes it less desirable for that link to appear in another layer; this captures the notion that the different network layers are attempting to fill gaps in connectivity left by the other layers.[1] As a motivating example, consider competing transportation companies offering services between a common set of cities. Suppose that for economical reasons, each company would prefer to design its transportation network to provide short routes between those cities that are not directly serviced by any other company. In other words, each company designs its network with respect to the *complement* of the transportation networks provided by all other companies. If we impose further structure on such games by assuming distance-based utility functions, we obtain the game defined below. In the following definition, for a set of graphs $G_j = (N, E_j)$, $j = 1, 2, \ldots, m$, on a common set of nodes, we use the notation $\cup_{j=1}^m G_j$ to indicate the graph $G = (N, \cup_{j=1}^m E_j)$, and $\cap_{j=1}^m G_j$ to indicate the graph $G = (N, \cap_{j=1}^m E_j)$.

[1] One can also consider a *strategic complements* version of this class of games where each player wishes to provide short paths between those pairs of nodes that share an edge in each of the other layers. The analysis of such games is relatively straightforward and thus we only focus on strategic substitutes.

Definition 7. *The game in Definition 6 is said to be a Multi-Layer Network Formation Game with Strategic Substitutes and Distance-Utilities if the utility functions are of the form*

$$\begin{aligned} A_i(G_1, \ldots, G_m) &= u_i(G_i| \sim (\cup_{j=1, j\neq i}^m G_j)) \\ &= \sum_{(x,y)\notin \cup_{k=1,k\neq i}^m E_k} b_i(d_{G_i}(x, y)) - c_i|E_i|, \end{aligned} \tag{3.1}$$

where the function u_i is defined in (2.1); the benefit functions $b_i(\cdot)$ are nonnegative, nonincreasing and satisfy $b_i(\infty) = 0$, and all costs c_i are positive. The benefit functions and costs can be different for the different players.

It is clear from the definition of the game that $(G_1, G_2, \ldots, G_m)$ is a Nash equilibrium if and only if for all $1 \leq i \leq m$, G_i is a BR network with respect to $\sim (\cup_{j=1,j\neq i}^m G_j)$ for the utility function (2.1). Although we showed in Theorem 1 that finding a BR network with respect to this utility function is NP-hard in general, we now show that certain insights can nevertheless be obtained in the multiplayer setting (regardless of the number of nodes and players). To develop our results, we partition the set of players P into three sets: *high-cost players* $S_H = \{P_i \in P | c_i > b_i(1)\}$, *medium-cost players* $S_M = \{P_i \in P | b_i(1) \geq c_i \geq b_i(1) - b_i(2)\}$ and *low-cost players* $S_L = \{P_i \in P | b_i(1) - b_i(2) > c_i\}$. We start by considering the case where the game contains low-cost players.

3.1.1 Games Containing Low-Cost Players

Proposition 5. *Suppose $|S_L| \geq 1$. Then in every Nash equilibrium, every player in S_H chooses the empty network. Furthermore, any vector of disjoint networks $(G_1, G_2, \ldots, G_m)$ forms a Nash equilibrium when $\{G_k | P_k \in S_M\}$ is a set of disjoint forests and $\cup_{i\in S_L} G_i = \sim \cup_{i\in S_M} G_i$.*

Proof. Let $(G_1, G_2, \ldots, G_m)$ be any vector of networks in Nash equilibrium. Since there exists at least one player P_i whose edge cost satisfies $c_i < b_i(1) - b_i(2)$, the Nash equilibrium vector must satisfy $\cup_{j=1}^m G_j = G^c$, where G^c is the complete network. To see this, suppose that the union of the graphs is not the complete network; then there exists some edge (u, v) that does not appear in any network, and thus appears in the complement of the graph $\cup_{j=1,j\neq i}^m G_j$. By Lemma 7, the BR to $\sim \cup_{j=1,j\neq i}^m G_j$ with respect to player P_i's utility function is $\sim \cup_{j=1,j\neq i}^m G_j$, and thus the edge (u, v) appears in graph G_i, contradicting the fact that it does not appear in the union of all the graphs.

Next, note that since $\cup_{j=1}^m G_j = G^c$, for any player $P_k \in P$, the graph $G_k = (N, E_k)$ is a BR to the graph $G^c \backslash \{\cup_{j=1,j\neq k}^m G_j\} \subseteq G_k$. By Lemma 1, a BR to a graph cannot be a strict superset of that graph, and thus we have that G_k is a best response to itself with respect to the utility function of player P_k. Now if $P_k \in S_H$, we know from Lemma 7 that G_k must be the empty network, completing the first part of the proof. For the second part, note that for any vector of networks satisfying the given properties, Proposition 2 and Lemma 7 indicate that a best response to G_k is indeed G_k for $P_k \in S_M \cup S_L$, completing the proof. □

The above result shows that the presence of a player with low edge costs (relative to its own benefit function) guarantees the existence of a Nash equilibrium in the game, and furthermore, such low-cost players drive players with sufficiently high edge costs out of the game; the proposition provides the threshold for costs at which this occurs (namely $b_i(1) < c_i$). Players with medium edge costs, on the other hand, can obtain certain nonempty networks in equilibrium, and the players with low edge costs split all of the remaining edges amongst themselves.

We now study the situation where there are no low-cost players in the game (i.e., $S_L = \emptyset$). We start by considering games that contain only high-cost players.

3.1.2 Games Containing Only High-Cost Players

Suppose $P = S_H$ (i.e., $S_L = S_M = \emptyset$). For each player $P_i \in P$, define the index k_i as

$$k_i \triangleq \min\left\{t \in \mathbb{N} \mid c_i < b_i(1) + \frac{t-2}{2} b_i(2)\right\}.$$

Since $c_i > b_i(1)$, we have $k_i \geq 3$ for all $P_i \in P$. If $k_i > n$, by Lemma 7, the empty network is a BR of player P_i to any set of networks G_{-i} (since α in (2.12) satisfies $\alpha \leq \frac{n-2}{2}$ for any reference graph). Thus without loss of generality, assume that all players have $3 \leq k_i \leq n$ and players are sorted according to their k_i, i.e., $k_1 \leq k_2 \leq \cdots \leq k_m \leq n$. We will now partition the set of players P into different sets.

Define the index i_1 as

$$i_1 \triangleq \max\left\{i \in \{1, 2, \ldots, m\} \mid k_i \leq n - i + 1\right\}.$$

Next, define

$$\begin{aligned}
i_2 &\triangleq \max\left\{i \in \{1, 2, \ldots, i_1 - 1\} \mid k_i \leq i_1 - i + 1\right\}, \\
i_3 &\triangleq \max\left\{i \in \{1, 2, \ldots, i_2 - 1\} \mid k_i \leq i_2 - i + 1\right\}, \\
&\vdots \\
i_r &\triangleq \max\left\{i \in \{1, 2, \ldots, i_{r-1} - 1\} \mid k_i \leq i_{r-1} - i + 1\right\},
\end{aligned} \tag{3.2}$$

where i_r satisfies $i_r < k_1$ (so that no further sets of this form can be defined).

The above indices satisfy $1 \leq i_r < i_{r-1} < \cdots < i_1 \leq m$. Partition the set of players and nodes as follows

$$\begin{aligned} H_r &= \{P_1, \dots, P_{i_r}\}, V_r = \{v_1, \dots, v_{i_r}\} \\ H_{r-1} &= \{P_{i_r+1}, \dots, P_{i_{r-1}}\}, V_{r-1} = \{v_{i_r+1}, \dots, v_{i_{r-1}}\} \\ &\vdots \\ H_1 &= \{P_{i_2+1}, \dots, P_{i_1}\}, V_1 = \{v_{i_2+1}, \dots, v_{i_1}\}. \end{aligned} \tag{3.3}$$

Also define $H_0 = \{P_{i_1+1}, P_{i_1+2}, \dots, P_m\}$ and $V_0 = \{v_{i_1+1}, v_{i_1+2}, \dots, v_n\}$.

Proposition 6. *For each player $P_j \in H_l$ (for $1 \leq l \leq r$), define the network G_j to be the star network centered on node v_j with peripheral nodes $\cup_{t=0}^{l-1} V_t$, where H_l and V_t are defined as in equation (3.3). For each player $P_j \in H_0$, define G_j to be the empty network. Then the set of networks $(G_1, G_2, \dots, G_m)$ forms a Nash equilibrium.*

To prove Proposition 6, we will first need the following intermediate result.

Lemma 9. *Let $b(1) < c$. Consider network $G = (N, E)$ with components $G_i = (N_i, E_i)$ for $1 \leq i \leq r$ ($N = \cup_{i=1}^r N_i$ and $E = \cup_{i=1}^r E_i$). Assume that every induced subgraph of G has a 2-spanner forest. Then every BR of network G is composed of a BR to each component of G.*

Proof. Let $F = (\cup_{i=1}^r N_i, E_F)$ be a BR to G. Suppose by way of contradiction that F contains a non-empty component $F_1 = (W, R)$ with nodes from p different N_i where $p \geq 2$. Let $G_{F_1} = (W, E_{F_1})$ denote the induced subgraph of W on G and T be a 2-spanner forest of G_{F_1}. The spanner forest T has q components where $q \geq p$. Also note that $|R| \geq |W| - 1 > |W| - q$. Then we have

$$\begin{aligned} u(F_1|G_{F_1}) &\leq |R|(b(1) - c) + (|E_{F_1}| - |R|)b(2) \\ &= (|R| - (|W| - q))(b(1) - c) + (|W| - q)(b(1) - c) + (|E_{F_1}| - |R|)b(2) \\ &< (|R| - (|W| - q))b(2) + (|W| - q)(b(1) - c) + (|E_{F_1}| - |R|)b(2) \\ &= (|W| - q)(b(1) - c) + (|E_{F_1}| - (|W| - q))b(2) \\ &= u(T|G_{F_1}), \end{aligned} \tag{3.4}$$

where the first inequality comes from the fact that at most $|R|$ pairs of nodes that are neighbors in G_{F_1} have direct connections in F_1 and the remaining pairs of nodes are at a distance of at least 2 in F_1. The second inequality is due to $b(1) - c < b(2)$.

Inequality (3.4) means that by replacing F_1 with T, we can increase the utility of network F which is a contradiction to the assumption that F is a BR to G. Therefore, no component of F contains nodes from multiple components in G and thus by Lemma 6, the subgraph of F induced by N_i must be a BR to G_i for $1 \le i \le r$, yielding the result. $\square$

We are now in place to prove Proposition 6.

Proof of Proposition 6. Consider player P_j where $1 \le j \le m$. If $j > i_1$ (i.e., $P_j \in H_0$), then $G =\sim \cup_{t=1,t\neq j}^{m} G_t$ consists of disjoint complete graphs on node sets $V_r, V_{r-1}, \ldots, V_0$. Since

$$\begin{aligned}
&k_j \ge k_{i_1+1} > n - i_1\\
&k_j \ge k_{i_2+1} > i_1 - i_2\\
&\qquad\vdots\\
&k_j \ge k_{i_r+1} > i_{r-1} - i_r\\
&k_j \ge k_1 > i_r,
\end{aligned}$$

a best response of player P_j to any of these complete networks is the empty network (by Proposition 1 or Lemma 7). Every induced subgraph of G has a star network on its non-empty components (which means it has a 2-spanner forest). Thus using Lemma 9, the empty network G_j is a BR to the network of the other players.

Next, we prove that for player $P_j \in H_l$ where $1 \le l \le r-1$, the network G_j is a BR to the other players' networks. From the definition of the sets H_l in (3.3), we have that $i_{l+1} < j \le i_l$. Note that $G =\sim \cup_{t=1,t\neq j}^{m} G_t$, consists of disjoint complete graphs on node sets $V_{l+1}, \ldots, V_r$. It also has a component $C = (\cup_{t=0}^{l} V_t, E_C)$ of size $n - i_{l+1}$. The structure of the network C can be described as a set of complete networks of size $n - i_1 + 1, i_1 - i_2 + 1, \ldots, i_{l-1} - i_l + 1, i_l - i_{l+1}$ where all of them have the common node v_j. These complete networks are on node sets $V_0 \cup \{v_j\}, V_1 \cup \{v_j\}, \ldots, V_{l-1} \cup \{v_j\}, V_l$. Network G satisfies the condition of Lemma 9 and thus a BR to G can be obtained by finding a BR network to each component. Since

$$\begin{aligned}
&k_j \ge k_{i_{l+2}+1} > i_{l+1} - i_{l+2}\\
&\qquad\vdots\\
&k_j \ge k_{i_r+1} > i_{r-1} - i_r\\
&k_j \ge k_1 > i_r,
\end{aligned}$$

the best response of player P_j to each of the complete networks on node sets $V_{l+1}, \ldots, V_r$ in G is the empty network.

Network C has a star subgraph centered at node v_j and hence by Proposition 3, there exists a BR network $S = (\cup_{t=0}^{l} V_t, E_S)$ that is a star network centered at node v_j with potentially some isolated nodes. Now assume that in the network S, there are edges from v_j to a nonempty strict subset of nodes $R_q \subset V_q$ for some $0 \leq q \leq l$, and the set of nodes in $V_q \setminus R_q$ are isolated. Note that edges between node v_j and the set of nodes R_q are only useful for connections between nodes in $R_q \cup \{v_j\}$ and produces a utility of

$$|R_q| \left(b_j(1) - c_j + \frac{|R_q| - 1}{2} b_j(2) \right) \geq 0, \tag{3.5}$$

where the inequality follows from the fact that this graph has utility at least as large as that of the empty network. Now construct a new network $S' = (\cup_{t=0}^{l} V_t, E_{S'})$ by connecting a node $u \in V_q \setminus R_q$ to v_j, i.e., $E_{S'} = E_S \cup \{(v_j, u)\}$. Then we have that $u(S'|C) - u(S|C) = b_j(1) - c_j + |R_q| b_j(2)$ which must be a positive value by inequality (3.5). This contradicts the assumption that S is a BR to C. Therefore, for each $0 \leq t \leq l$, node v_j is either connected to all of the nodes in V_t or to none of them. Since

$$\begin{aligned} k_j &\leq k_{i_1} \leq n - i_1 + 1 \\ k_j &\leq k_{i_2} \leq i_1 - i_2 + 1 \\ &\vdots \\ k_j &\leq k_{i_l} \leq i_{l-1} - i_l + 1, \end{aligned}$$

a BR to all of the complete networks on nodes $V_t \cup \{v_j\}$ in C is the star network for $0 \leq t \leq l-1$. However, since $k_j \geq k_{i_{l+1}+1} > i_l - i_{l+1}$, the BR to the complete network on the set of nodes V_l is the empty network and thus all of the nodes in $V_l \setminus \{v_j\}$ must be isolated nodes.

Therefore, we can conclude that a star network centered on the node v_j with peripheral nodes $\{v_{i_l+1}, \ldots, v_n\} = \cup_{t=0}^{l-1} V_t$, and all other nodes being isolated is a BR to the network of the other players; this is precisely the network G_j given in the statement of the proposition.

Finally, we have to show that players P_j, $1 \leq j \leq i_r$ (i.e., $P_j \in H_r$) are in Nash equilibrium. Similar to the above, for player P_j, $G =\sim \cup_{t=1, t\neq j}^{m} G_t$ consists of complete networks of size $n - i_1 + 1, i_1 - i_2 + 1, \ldots, i_{r-1} - i_r + 1, i_r$ with the common node v_j. These complete networks are on node sets $V_0 \cup \{v_j\}, V_1 \cup \{v_j\}, \ldots, V_{r-1} \cup \{v_j\}, V_r$. By an argument similar to the above, since $k_j \geq k_1 > i_r$ and

$$\begin{aligned} k_j &\leq k_{i_1} \leq n - i_1 + 1 \\ k_j &\leq k_{i_2} \leq i_1 - i_2 + 1 \\ &\vdots \\ k_j &\leq k_{i_r} \leq i_{r-1} - i_r + 1, \end{aligned}$$

a star network centered on v_j with peripheral nodes $\{v_{i_r+1}, \ldots, v_n\} = \cup_{t=0}^{r-1} V_t$ (i.e., G_j) is a BR to the network of the other players.

Therefore, for each player $P_j \in P$, G_j is a BR to $G = \sim \cup_{t=1, t\neq j}^{m} G_t$ and thus the given networks are in Nash equilibrium. $\square$

The following example illustrates the structure of the Nash equilibrium specified by the above proposition.

Example 5. *Suppose that there are 11 nodes and 9 high-cost players with $k_i = 3$ for $1 \leq i \leq 5$, $k_6 = 4,\ k_7 = 5$ and $k_8, k_9 \geq 5$. From the equations in (3.2), we get $i_1 = 7, i_2 = 5, i_3 = 3$ and $i_4 = 1$. Figure 3.1 shows partition of the set of players into 5 sets based on the index values. Given these partitions, Figure 3.2 demonstrates the networks of players P_1, P_2, P_4 and P_6 in the*

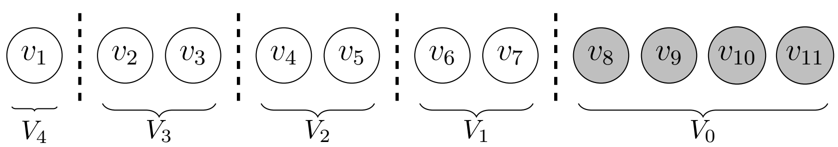
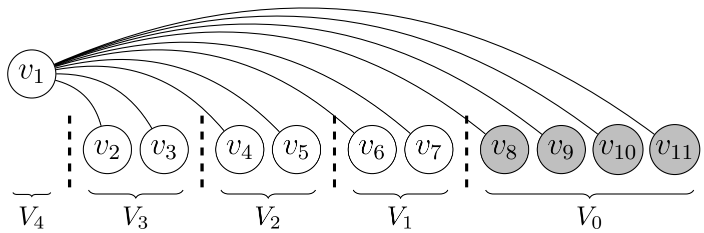


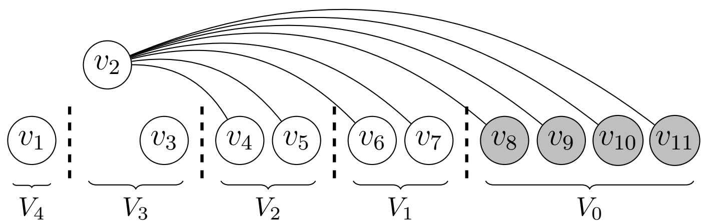
Figure 3.1: A multi-layer network formation game considered in Example 5 with 9 high-cost players and 11 nodes. Nodes are partitioned into 5 sets as shown in this figure, based on the characteristics of the players. Each node in each of the sets V_1, V_2, V_3, V_4 will be chosen by a different player as the center of a star subgraph in the Nash equilibrium.

Nash equilibrium defined in Proposition 6. Player P_3 has a similar network to player P_2 (except that the star of her network is centered on v_3). Players P_5 and P_7 have similar networks to that of P_4 and P_6, respectively (the only difference being that player P_5 has a star centered on v_5, and P_7 has a star centered on v_7). Players P_8 and P_9 each have the empty network.

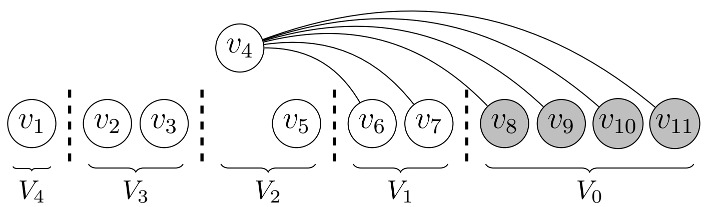
Despite the stylized nature of the multi-layer network formation game in Definition 7, it is of interest to note that the "hub-and-spoke" networks that arise in the above Nash equilibrium are predominant in real-world transportation systems (airline networks, in particular) [82, 1, 57]. While previous work has shown that such networks are optimal in the single-layer setting (e.g., Proposition 1 [48]), our analysis shows that these structures also arise when players selfishly optimize their individual networks in competitive environments. We will now consider games with a mix of medium-cost and high-cost players, and show that such structures also arise as a Nash equilibrium in that setting.



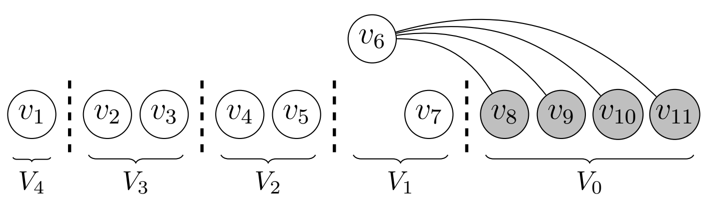
(a) G_1



(b) G_2



(c) G_4



(d) G_6

Figure 3.2: The Nash equilibrium networks of players P_1, P_2, P_4 and P_6 in Example 5 are shown in 3.2a, 3.2b, 3.2c and 3.2d, respectively. The networks of players P_3, P_5 and P_7 are not shown; they have stars centered on v_3, v_5 and v_7, respectively, with the same peripheral nodes as P_2, P_4 and P_6, respectively. Players P_8 and P_9 choose the empty network.

3.1.3 Games With Medium and High-Cost Players

Proposition 7. *Suppose that $S_L = \emptyset$, and assume without loss of generality that the first μ players in P are medium-cost players, with $1 \le \mu \le n$. For $j \in \{1, 2, \ldots, \mu\}$, define the network G_j to be the star network centered on node v_j with peripheral nodes $\{v_{j+1}, v_{j+2}, \ldots, v_n\}$. For the set of high-cost players S_H, let $(G_{\mu+1}, G_{\mu+2}, \ldots, G_m)$ be the Nash equilibrium networks on node set $\{v_{\mu+1}, v_{\mu+2}, \ldots, v_n\}$ defined in Proposition 6. Then the set of networks $(G_1, G_2, \ldots, G_m)$ is a Nash equilibrium.*

Proof. In the proof, we will use the fact that each network G_j, $1 \le j \le m$, only contains edges from node v_j to nodes with index larger than j. For each player P_j, let $G_{j,ref} \triangleq \bigcup_{i=1, i \neq j}^{m} G_j$ be the union of the networks of the other players.

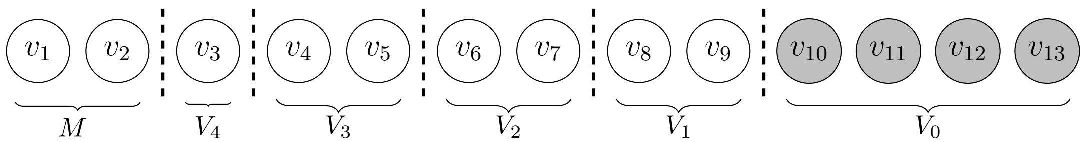
Consider a medium-cost player P_j, where $j \in \{1, 2, \ldots, \mu\}$. Since all players with index smaller than j are medium-cost players, for each node v_i with $i < j$, $G_{j,ref}$ contains an edge from node v_i to v_k for all $k > i$. Furthermore, $G_{j,ref}$ contains no edge from v_k to v_j for any $k > j$. Thus, in the network $\sim G_{j,ref}$, nodes $v_1, v_2, \ldots, v_{j-1}$ are isolated, and there is an edge from v_j to each node v_k with $k > j$. By Lemma 6, the isolated nodes in $\sim G_{j,ref}$ remain isolated in the BR; applying Proposition 3, a star network centered at v_j with edges to $\{v_{j+1}, \ldots, v_n\}$ is a BR with respect to $\sim G_{j,ref}$. Thus, G_j is a BR to $\sim G_{j,ref}$.

Now consider a high-cost player P_j, where $j \in \{\mu + 1, \mu + 2, \ldots, m\}$. Arguing as above, nodes $v_1, v_2, \ldots, v_\mu$ are isolated in the network $\sim G_{j,ref}$. Thus by Lemma 6, those nodes remain isolated in the BR to $\sim G_{j,ref}$. Since this is true for all high-cost players, we can remove the nodes $v_1, v_2, \ldots, v_\mu$ from consideration, and focus on showing that the subgraph of G_j induced by the node set $\{v_{\mu+1}, v_{\mu+2}, \ldots, v_n\}$ is a BR to the graphs $(G_{\mu+1}, G_{\mu+2}, \ldots, G_m)$ on that node set. This is true by construction, and thus the given set of networks is a Nash equilibrium. □

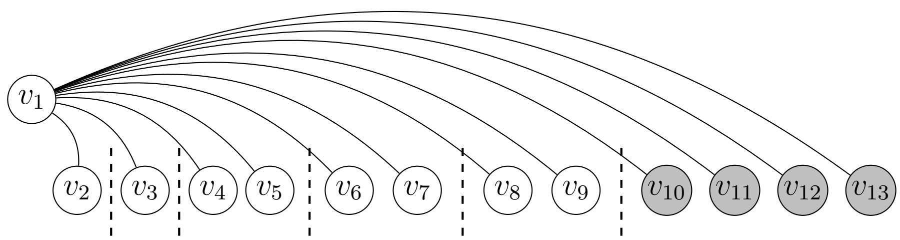
Example 6. *Consider a game with 13 nodes, 2 medium-cost players (P_1 and P_2) and 9 high-cost players ($P_3, \cdots, P_{11}$). Assume that the 9 high-cost players are the same as the high-cost players in Example 5. Based on Proposition 7, each of the medium-cost players P_1 and P_2 will have a star network centered on node v_1 and v_2, with peripheral nodes $V \setminus \{v_1\}$ and $V \setminus \{v_1, v_2\}$, respectively. These networks are shown in Figure 3.3b and 3.3c, respectively. The networks of the remaining players (which have high costs) have the same structure as in Example 5 with two extra isolated nodes, v_1 and v_2. Once again, we see that hub-and-spoke networks arise as a Nash equilibrium in this setting.*

The following corollary immediately follows from Propositions 5, 6 and 7.

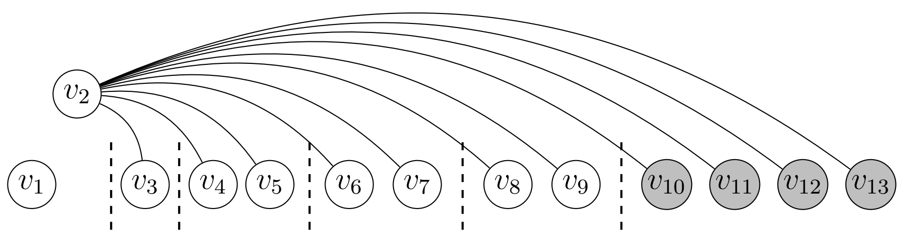
Corollary 1. *The multi-layer network formation game with strategic substitutes and distance-utilities has a pure Nash equilibrium for any set of players.*



(a) Partition of the nodes.



(b) G_1



(c) G_2

Figure 3.3: Figure 3.3a demonstrates the partition of the set of nodes into 6 sets. The first set (denoted M) contains nodes that will form the centers of the star networks chosen by the medium cost players P_1 and P_2. These star networks are depicted in Figures 3.3b and 3.3c. The networks of the remaining high-cost players have the same structure as the networks shown in Figures 3.2a to 3.2d, with v_1 and v_2 as isolated nodes.

3.2 Colonel Blotto Network Formation Game

The Multi-Layer Network Formation Game defined in the last section assumed that each player in the game constructs a separate layer of the network, based on the layers constructed by the other players. There is no hard constraint that prevents multiple players from having the same edge in their layers, although the Strategic Substitute utility function disincentivizes such behavior.

In this section, we consider an alternate setting for competitive network formation where two players are each given a budget and compete with each other to purchase edges from a common set. As an example, consider once again the example of two competing transportation companies. Each company has a fixed budget to spend on service between some or all pairs of cities. Under an idealized allocation rule, the company that spends more on a given edge wins that edge, and the utility to a company is a function of all of the edges that it wins. Similar examples can be formulated for telecommunication companies bidding on spectrum or communication links, and wireless networks where a transmitter and jammer are competing to send or disrupt information, respectively, with fixed power budgets [96].

Games of this form are traditionally known as *Colonel Blotto* games; here we will extend such games to the network formation setting and characterize the types of equilibria that occur for different utility functions and budgets of the players. We start by reviewing the classical version of this game.

3.2.1 The Colonel Blotto Game

The classical Colonel Blotto (CB) game is defined as follows [37, 90].

Definition 8. *There are two players P_1 and P_2 with S_1 and S_2 units of resources, respectively. There are n battlefields, and the players choose their actions simultaneously. The strategy space of player P_l, $l \in \{1, 2\}$ is given by*

$$X^l = \{(x_1^l, x_2^l, \dots, x_n^l) \in \mathbb{R}^n_{\geq 0} | \sum_{i=1}^{n} x_i^l = S_l\}.$$

The amount of resources allocated by player P_l to field i is x_i^l. There is a zero sum game between the players in each field i where the player with the higher amount of allocated resources to that field receives a payoff of $+1$ and the loser receives -1; both players receive 0 when they allocate equal resources to a field (although other tie-breaking rules can also be considered). The total payoff to each of the players is given by $u_1 = \sum_{i=1}^{n} sgn(x_i^1 - x_i^2)$ and $u_2 = -u_1$.

Using connections to simultaneous first-price all-pay auctions, [81] showed that this game has no pure Nash equilibrium when $\frac{1}{n} < \frac{S_2}{S_1} \leq 1$, but does have an equilibrium in mixed strategies. Other variants of this game such as having more than two players, different types of functions for scoring over battles, and fields with different values have also been addressed in the literature [37, 90]. Here, we introduce the Colonel Blotto game into the competitive network formation setting, and characterize the set of equilibria that can occur.

3.2.2 The Colonel Blotto Network Formation Game

Definition 9. *The CB Network Formation Game consists of two players P_1 and P_2, and a set of nodes $N = \{v_1, v_2, \ldots, v_n\}$. These players play a CB game where the set of battlefields are the $\binom{n}{2}$ edges between these nodes. Player P_l, $l \in \{1, 2\}$, has a positive budget S_l, and strategy space*

$$X^l = \{(x^l_{12}, x^l_{13}, \ldots, x^l_{n-1,n}) \in \mathbb{R}^{\binom{n}{2}}_{\geq 0} | \sum_{i,j \in N, i \neq j} x^l_{ij} = S_l\}, \tag{3.6}$$

where entry x^l_{ij} indicates the allocation by player P_l to edge (v_i, v_j). Each pair of strategies $(x^1, x^2) \in X^1 \times X^2$ induces graphs $G_1(x^1, x^2)$ and $G_2(x^1, x^2)$, where $G_l(x^1, x^2)$ is formed from edges in which player l has the highest allocation for $l \in \{1, 2\}$[2]. There is a utility function $u_l : G^N \times G^N \to \mathbb{R}$ for $l \in \{1, 2\}$ that determines the utility of each player based on the formed networks. The pair of strategies (x^1, x^2) is said to be a pure Nash equilibrium if and only if $x^1 \in \operatorname{argmax}_{x \in X^1} u_1(G_1(x, x^2), G_2(x, x^2))$ and $x^2 \in \operatorname{argmax}_{x \in X^2} u_2(G_1(x^1, x), G_2(x^1, x))$. For simplicity, we denote $G_l(x^1, x^2)$ by just G_l in the rest of this section.

Remark 8. *The chosen strategy $x^l \in X^l$ of player P_l in equation (3.6) can equivalently be viewed as a weighted graph[3] F_l, where each player has allocated a nonzero amount of investment on each of the links and zero investment on the links of its complement network. We refer to the strategies of the players as their chosen network or investment vector.*

Note that the above game is a version of the CB game where the battlefields exhibit *strategic complementarities* (i.e., the value of a given battlefield, or edge, depends on the set of all battlefields won by that player). While CB games with complementarities have been studied previously (e.g., [90]), the difference in our setting is that the complementarities arise from network characteristics (such as distance and connectedness), which lends additional structure to the problem.

In the rest of this section, we will assume without loss of generality that player P_1 has a budget $S_1 = 1$, and $S_2 \leq 1$. The following facts will be useful for characterizing the equilibria of the game for various utility functions.

Lemma 10. *Consider the CB Network Formation Game with budgets $S_1 = 1$ and $S_2 \leq S_1$, and node set $N = \{1, 2, \ldots, n\}$.*

[2] Similar to the classical CB game in definition 8, we assume that if the players allocate equal amount of resources to an edge, none of them wins that edge.

[3] A weighted graph associates a real number (which must be nonnegative in the CB game) to every edge in the graph.

1. *If $S_2 < \frac{2}{n(n-1)}$, there is an investment vector allowing player P_1 to win all edges, regardless of the strategy of player P_2.*

2. *If $\frac{2}{n(n-1)} \leq S_2 \leq \frac{2}{n}$, there is an investment vector for player P_1 such that player P_2 cannot win a component of size n.*

3. *If $\frac{2(m-1)}{n(n-1)} < S_2 \leq 1$ where $2 \leq m \leq n$, then for any given investment vector of player P_1, player P_2 can win a star network with at least m nodes.*

Proof. The first case is trivial as it suffices for player P_1 to allocate $\frac{2}{n(n-1)}$ on all of the edges.

For the second case, note that P_2 needs at least $n-1$ edges to win a component of size n. If player P_1 allocates $\frac{2}{n(n-1)}$ on each edge, player P_2 can never win a component of size n since that would require a total investment larger than $\frac{2}{n}$.

Finally, in the last case, let S_i^j denote a star network of size m centered at node v_i with $m-1$ peripheral nodes chosen from the set $N \setminus \{i\}$ (the index j will be used to enumerate such star networks). It is clear that for any $v_i \in N$, there exists $\binom{n-1}{m-1}$ of these star networks. Denote the sum of the investments of player 1 on the edges of the star network S_i^j by b_i^j. Then we have

$$\sum_{i=1}^{n} \sum_{j=1}^{\binom{n-1}{m-1}} b_i^j = 2\binom{n-2}{m-2}.$$

This is due to the fact that each edge (v_k, v_t) is counted $2\binom{n-2}{m-2}$ times, where $\binom{n-2}{m-2}$ is the number of possible star networks on node v_k that contains the edge (v_k, v_t) (and similarly for node v_t). Therefore, there must exist a node v_i with b_i^j such that

$$b_i^j \leq \frac{2\binom{n-2}{m-2}}{n\binom{n-1}{m-1}} = \frac{2(m-1)}{n(n-1)}.$$

By allocating his or her resources appropriately, player P_2 can win all of the edges in the star network S_i^j. $\square$

Remark 9. *Note that when $S_2 < 1$, for any given investment vector of player P_2, player P_1 can win all edges of the network by simply matching P_2's investment everywhere, and then spreading the excess budget evenly over all edges. When $S_2 = 1$, no player can win all edges because that would require a total investment larger than 1. However, for any given investment vector of player P_2, player P_1 can win $\binom{n}{2} - 1$ edges as follows: choose an edge where P_2 has a nonzero investment r, match P_2's investment on all other edges and then distribute an additional r evenly over all those edges. The same is true for P_2 by symmetry. These arguments are independent of the utility function and are standard in the study of CB games [37].*

We now study the CB network formation game for two natural utility functions.

3.2.3 Colonel Blotto Network Formation with Respect to Largest Component

In this section we will define the utility u_l of each player to be an increasing function of the size of the largest component in their formed network, capturing the notion that players wish their network to provide paths between as many nodes as possible. For instance, having a larger component is advantageous for a telecommunications company that provides service among a group of cities. This has also been a measure of robustness in the network attack and defense literature, where the defender aims at maximizing the size of his component in the presence of adversaries [9, 41]. The following proposition characterizes Nash equilibria in terms of the range of resources available to player 2.

Proposition 8. *Assume that* $\frac{2(m-1)}{n(n-1)} < S_2 \leq 1$ *where* $2 \leq m \leq n$. *Let* $G_l \in G^N$ *for* $l \in \{1, 2\}$ *denote the network formed by player* P_l *as the outcome of the game. Define the utility functions of the players in the CB network formation game as*

$$u_1(G_1, G_2) = g_1(C(G_1)), \;\; u_2(G_1, G_2) = g_2(C(G_2)),$$

where $g_1(\cdot)$ *and* $g_2(\cdot)$ *are increasing functions and* $C(G)$ *denotes the size of the largest component in graph* G. *Then for any Nash equilibrium pair of actions* $x^1 \in X^1$ *and* $x^2 \in X^2$, *we have that* $C(G_1) = n$ *and* $C(G_2) \geq m$.

Proof. By Remark 9, it is clear that the formed network by player 1 must be a connected network. Furthermore, when $\frac{2(m-1)}{n(n-1)} < S_2 \leq 1$, Lemma 10 indicates that for any given strategy chosen by player P_1, player P_2 can always choose his strategy such that his outcome network has a component of size at least m. □

Note that there are many equilibria of the form described in the above proposition, as in any complete graph on n nodes, there are $\lfloor \frac{n}{2} \rfloor$ disjoint spanning trees [75]. For any range of resources available for player P_2, any two of these trees form an equilibrium when each player allocates his budget entirely to the edges on his tree.

By Lemma 10, we know that for $S_2 \leq \frac{2}{n(n-1)}$, there exist strategies for player P_1 such that player P_2 cannot win a single edge. Thus in the case of $S_2 \leq \frac{2}{n(n-1)}$, there are instances of Nash equilibria in which player 2 forms the empty network.

3.2.4 Colonel Blotto Network Formation with Respect to Diameter

Proposition 8 considered the case where each player only cares about the size of the largest component in his or her network; however, the *diameter* of the network is often also of interest in network formation and design [46, 91]. Here, we study the CB network formation game when the utility of each player is decreasing in the diameter of the network.

Proposition 9. *Let $G_l \in G^N$ for $l \in \{1, 2\}$ denote the network formed by player P_l as the outcome of the game under strategies $x^1 \in X^1$ and $x^2 \in X^2$. Define the utility functions of the players in the CB network formation game as*

$$u_1(G_1, G_2) = h_1(D(G_1)), \;\; u_2(G_1, G_2) = h_2(D(G_2)),$$

where $h_1(\cdot)$ and $h_2(\cdot)$ are decreasing functions and $D(G)$ denotes the diameter of graph G. Then the pair of actions $x^1 \in X^1$ and $x^2 \in X^2$ are in Nash equilibrium if and only if one of the following two conditions holds.

1. *$S_2 = S_1$ and both G_1 and G_2 have a diameter of 2.*
2. *$S_2 \leq \frac{2}{n}$ and $G_1 = G^c$ and $G_2 = G^e$.*

Proof. We first argue that there is no (pure) Nash equilibrium when $\frac{2}{n} < S_2 < 1$. By Lemma 10, for any given strategy $x^1 \in X^1$, player P_2 can choose a strategy $x^2 \in X^2$ such that he obtains a star, which has diameter 2. However for any given strategy $x^2 \in X^2$, player P_1 can choose $x^1 \in X^1$ such that she wins all edges and obtains a diameter of 1 (as argued in Remark 9). Thus there is no pure Nash equilibrium for this range of S_2.

Suppose $S_2 = 1$. Note from Remark 9 that no player can win all edges, and thus no player can achieve a diameter less than 2. However, for any given investment by one player, the other has a strategy that wins all but one of the edges, thereby winning a network of diameter 2. Thus all Nash equilibria under $S_2 = S_1 = 1$ satisfy the property that both players win a network of diameter 2.

Now suppose that $S_2 \leq \frac{2}{n}$. In this case, by Lemma 10, for any given investment of player P_2, there always exists a strategy for player P_1 to win all edges and obtain a diameter of 1. Thus any Nash equilibrium must have player P_1 winning all edges and P_2 winning none. When player P_1 invests $\frac{2}{n(n-1)}$ on all edges, player P_2 can never obtain a connected network, and thus any investment vector $x^2 \in X^2$ with less than $\frac{2}{n(n-1)}$ on each edge yields a Nash equilibrium. □

Based on Proposition 9 for $S_1 = S_2$, players are in Nash equilibrium if and only if their formed networks each have a diameter of 2. By definition, the networks G_1 and G_2 formed in the CB network formation game are edge-disjoint. It is easy to check that for $n = 5$, the cycle graph 1-2-3-4-5-1 and its complement both have diameter of 2. We now show that there exist such graphs for any $n \geq 5$. We will use the following notation.

Definition 10. *Let $G_i = (N_i, E_i)$, $1 \leq i \leq T$, be a set of graphs. Then $G = (N, E) = [G_1, G_2, \ldots, G_T]$ is a graph such that $N = \cup_{i=1}^{T} N_i$ and $E = \cup_{i=1}^{T} E_i$.*

Proposition 10. *Suppose that we have $|N| = 2k$ nodes for $k \geq 3$. Let graph G_1 consist of nodes $\{v_1, v_2, \ldots, v_k\}$, G_2 consist of nodes $\{v_{k+1}, v_{k+2}, \ldots, v_{2k}\}$, and G_3 consist of all nodes $\{v_1, v_2, \ldots, v_{2k}\}$. Define the edge sets of G_1, G_2 and G_3 as follows (depicted in Figure 3.4):*

1. *G_1 is the path $v_1 v_2 v_3 \cdots v_k$.*

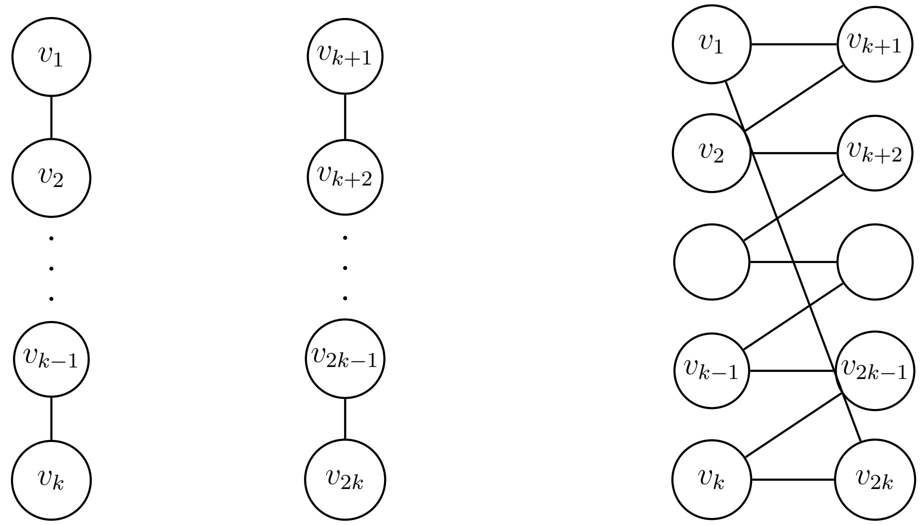
2. *G_2 is the path $v_{k+1} v_{k+2} \cdots v_{2k}$.*

3. *G_3 is the cycle $v_1 v_{k+1} v_2 v_{k+2} \cdots v_{2k-1} v_k v_{2k} v_1$.*

Define graphs $F_1 = [\bar{G}_1, \bar{G}_2, G_3]$ and $F_2 = \bar{F}_1$. Then both F_1 and F_2 have a diameter of 2. When $|N| = 2k + 1$ for some $k \geq 3$, it suffices to connect the node $(2k + 1)$ to all nodes $\{k + 1, k + 2, \ldots, 2k\}$ in graph F_1.

Proof. The proof of this proposition is straightforward but tedious, and thus we provide only a sketch of the proof. For $k = 3$ the proof can be verified by examining the graph directly, and thus suppose $k \geq 4$. In graph F_1, consider node v_1. This node is directly connected to nodes in the set $\{v_3, v_4, \ldots, v_k, v_{k+1}, v_{2k}\}$. It has a path of length 2 to nodes in the set $\{v_2, v_{k+3}, v_{k+4}, \ldots, v_{2k-1}\}$ via node v_{k+1}. Finally, it has a path of length 2 to v_{k+2} via node v_{2k}. Next consider node v_2. This node is directly connected to nodes in the set $\{v_4, v_5, \ldots, v_k, v_{k+1}, v_{k+2}\}$. It has a path of length 2 to node v_3 via node v_{k+2}. Finally, it has a path of length 2 to nodes in the set $\{v_{k+3}, \ldots, v_{2k}\}$ via v_{k+1}. This reasoning can be repeated for all nodes in the network to show that F_1 has a diameter of 2, and similarly that F_2 has a diameter of 2. □

Remark 10. *Note that the graphs provided above are not unique in having the property that both the graph and its complement have a diameter of 2. For example, consider the Erdos-Renyi graph $G(n, p)$ on n nodes, where each of the possible $\binom{n}{2}$ edges is independently present with probability $p \in [0, 1]$. When p is constant in $(0, 1)$, it is well known that asymptotically (in n) almost surely the graph $G(n, p)$ will have diameter 2 [24]. Noting that the complement of $G(n, p)$ is an Erdos-Renyi graph $G(n, 1 - p)$, we see that an Erdos-Renyi graph and its complement will both have diameter 2 with probability tending to 1 as $n \to \infty$.*



(a) G_1 (b) G_2 (c) G_3

Figure 3.4: An example of Nash equilirbium in CB network formation game with utility based on diameter.

3.3 Summary

In this chapter, we formulated a multi-layer network formation game where each player builds a different layer of the network, simultaneously. We started with the case that the layers are viewed as strategic substitutes, i.e., if there is a link between two nodes in the network of one of the players, it is less desirable for that link to appear in the network of other players. We showed that the Nash equilibria of the game exhibit certain natural characteristics. Specifically, the presence of low-cost players pushes high-cost players out of the game, and hub-and-spoke networks arise in the Nash equilibrium when there are no low-cost players.

Finally, we looked at the situation where there cannot be any overlap between the links of the network of players. We assumed that players have a fixed budget to spend on constructing edges. We modeled this situation as a Colonel Blotto game. We characterized the ranges of player budgets for which the game admits pure Nash equilibria for utility functions that depend on the component sizes and diameter of the formed networks.

Chapter 4

The Strategic Formation of Interconnections Between Networks

While we focused on multi-layer networks in the previous chapters, we now turn our attention to studying the structure of interdependent networks and considering the game-theoretic formation of edges *between* two given networks $G_1 = (V_1, E_1)$ and $G_2 = (V_2, E_2)$ on two different sets of nodes V_1 and V_2. We assume that there are dependencies between nodes in V_1 and V_2, i.e., some of the nodes in V_1 require connections to (or information from) some of the nodes in V_2 in order to function. These dependencies are captured by a bipartite network $G_I = (V_1 \cup V_2, E_I)$ where $E_I \subseteq V_1 \times V_2$, and an edge $(v_i, v_j) \in E_I$ indicates that $v_i \in V_1$ is dependent on $v_j \in V_2$. We consider a distributed network formation framework where each node in V_1 is a player and builds a set of edges between itself and nodes in V_2 in order to maximize a distance-based utility function. As a motivating abstraction, consider a cyber-physical system where G_1 is a power network (with the nodes representing substations) and G_2 is a sensor network [72, 77]. Suppose that neighboring nodes in each network are capable of exchanging information with each other. Each substation in the power network requires the information gathered by certain nodes in the sensor network; these dependencies are captured by the network G_I. The substation operators wish to construct connections to the sensor network in such a way that they minimize the number of hops required to gather data from their interdependent nodes (where the number of hops is measured with respect to the connections within G_1 and G_2 and the edges constructed between the networks). This leads to an *interconnection network design game* (INDG) with distance utilities where the utility of each player (operator) depends on its own set of edges as well as the set of edges constructed by other players. Distance-based utilities have been also used to study computer networks (where nodes represent the computers and edges are the communication links) [27, 89]. In this case, network G_I models the virtual dependencies among the computers in cluster G_1 and cluster G_2, indicating the set of pairs of nodes that wish to exchange information. The designed interconnection network represents the physical communication network between the two clusters. Another application of the INDG with distance utilities arises in studying interconnections between the transportation networks of two countries. We will elaborate on this example in Section 4.1.

We start our analysis by investigating the complexity of solving this problem and show that it is NP-hard to find a best response for each player. Despite the NP-hardness of the problem, we characterize some useful properties of the best response which consequently enable us to determine a Nash equilibrium instance for certain cases of the game. Specifically, we study the existence of Nash equilibria in an INDG with distance utilities when network G_2 has a star subgraph (similar to the "hub-and-spoke" structure seen in various transportation networks [82, 1]) and there is full interdependency between nodes in G_1 and G_2. We show that this setting possesses a Nash equilibrium for any set of players with arbitrary benefit functions and edge costs. We partition the set of players into two sets consisting of high and low edge cost players and show that in any Nash equilibrium, all of the high-cost players that have a low-cost player in their vicinity "free ride" and choose not to construct any interconnections to G_2.

Below, we have formally defined the problem.

4.1 Interconnection Network Design Game

Assume that we are given two arbitrary networks $G_1 = (V_1, E_1)$ and $G_2 = (V_2, E_2)$. In this chapter, we consider a setting in which each node in V_1 constructs a set of edges to nodes in V_2 such that some utility function is maximized. This leads to a game with the nodes of G_1 as the players.

Definition 11. *Consider two arbitrary networks $G_1 = (V_1, E_1)$ and $G_2 = (V_2, E_2)$ with $V_1 = \{x_1, \cdots, x_n\}$ and $V_2 = \{y_1, \cdots, y_m\}$. An instance of the interconnection network design game (INDG) $\mathcal{G} = (P, (S_i)_{P_i \in P}, (\Psi_i)_{P_i \in P}, G_1, G_2)$ has a set of n players $P = \{P_1, P_2, \cdots, P_n\}$ where player P_i is associated with node $x_i \in V_1$ for $1 \leq i \leq n$. The strategy space of player P_i is $S_i = 2^{\{x_i\} \times V_2}$, i.e., all possible subsets of edges from x_i to nodes in V_2. The action of player P_i is an element of S_i and is denoted by W_i, i.e., W_i is a set of edges from x_i to a certain subset of V_2. By an abuse of notation, we take $B = \cup_{j=1}^{n} W_j$ to indicate the bipartite graph $B = (V_1 \cup V_2, \cup_{j=1}^{n} W_j)$. The utility of player P_i is given by a function $\Psi_i : S_1 \times S_2 \times \cdots \times S_n \to \mathbb{R}$, where the j^{th} argument[1] is the action of the j^{th} player for $1 \leq j \leq n$.*

[1]The utility function Ψ_i is also a function of G_1 and G_2 which will be omitted from the argument list as long as it is clear from the context.

The characteristics of the game and the optimal strategies for each player will depend on the form of the utility functions Ψ_i. We consider a modified version of the distance utility function in (1.3.1) as the payoff to the players. Specifically, we assume that there are dependencies between nodes in the graphs G_1 and G_2 which is represented by a bipartite network $G_I = (V_1 \cup V_2, E_I)$ with two partitions V_1 and V_2 and $E_I \subseteq V_1 \times V_2$. Let $I_i \subseteq V_2,\ 1 \leq i \leq n$ denote the set of neighbors of $x_i \in V_1$ in the network G_I. Then the objective of player P_i is to find the optimal set of edges to construct to V_2 such that distance between its associated node x_i and the set of nodes in I_i is minimized. In addition to the technological applications that we mentioned before, the INDG can be utilized to model problems in transportation. For instance consider a modified version of the problem studied in [60] where we are given the traffic flow between cities of two different countries C_1 and C_2. Each of these countries has a domestic transportation service which connects its cities and is modeled by networks G_1 and G_2. A city in C_1 and a city in C_2 are said to be interdependent if the traffic flow between them is higher than some threshold, and this interdependency is represented by an edge between them in the network G_I. The players of the game correspond to transportation service planners at each node in C_1, who are faced with the problem of finding the optimal set of routes to establish from their associated city to cities of the country C_2 such that distance between the interdependent cities is minimized. It is clear that the structure of the interconnection between cities inside the countries C_1 and C_2 (modeled as G_1 and G_2) directly affects the optimal decisions made by the players.

Definition 12. *An instance*

$$\mathcal{G} = (P, (S_i)_{P_i \in P}, (u_i)_{P_i \in P}, G_1, G_2, G_I)$$

of the game in Definition 11 is said to be an interconnection network design game with distance utilities if the utility function of player P_i, $1 \leq i \leq n$ with action $W_i \in S_i$ has the form

$$\begin{aligned} \Psi_i(W_1, \cdots, W_n) &= u_i(\cup_{j=1}^{n} W_j | G_1, G_2, G_I) \\ &= \sum_{y \in I_i} b_i(d_G(x_i, y)) - c_i |W_i|, \end{aligned} \tag{4.1}$$

where $G = (V_1 \cup V_2, E_1 \cup E_2 \cup (\cup_{j=1}^{n} W_j))$.

As we can see in the utility function $u_i(\cdot)$, only the distances between node x_i and the set of nodes I_i matter. Furthermore, each player has to pay only for his/her constructed edges. The benefit functions $b_i(\cdot)$ are nonnegative, nonincreasing and satisfy $b_i(\infty) = 0$, and all costs c_i are positive, and can be different across players.

We will use W_{-i} to denote the vector of actions of all players except player P_i, and use $\Psi_i(W_i, W_{-i})$ to denote the utility of player P_i with respect to the given vector $(W_1, W_2, \ldots, W_n)$.

Based on the definition of the game, we say that a vector of actions $(W_1, W_2, \ldots, W_n)$ is a *Nash equilibrium* if and only if $W_i \in \arg\max_{W \in S_i} \Psi_i(W, W_{-i})$ for all $i \in \{1, 2, \ldots, n\}$. In this case, W_i is said to be a best response action to W_{-i} with respect to the utility function Ψ_i. For the rest of this chapter, whenever we say INDG, by default we mean an interconnection network design game with distance utilities.

4.2 Characteristics of the Best Responses

In this section, we characterize some important properties of the best response actions for the players. We start by determining the complexity of finding a best response action for the players in the INDG.

4.2.1 Complexity

We define the (decision) problem faced by each player in the INDG as follows.

Definition 13. Best Response Interconnection (BRI).
INSTANCE: *Given an instance*

$$\mathcal{G} = (P, (S_i)_{P_i \in P}, (u_i)_{P_i \in P}, G_1, G_2, G_I),$$

of INDG, a player $P_j \in P$, a joint strategy by all other players $W_{-j} = \cup_{i \neq j} W_i$ and a threshold $r \in \mathbb{R}_{>0}$.

QUESTION: *Does there exist an action $W_j \in S_j$ for the player P_j such that*

$$u_j(W_j \cup W_{-j} | G_1, G_2, G_I) = \sum_{y \in I_j} b_j(d_G(x_j, y)) - c_j|W_j| \geq r,$$

where $G = (V_1 \cup V_2, E_1 \cup E_2 \cup W_j \cup W_{-j})$?

We now provide the following theorem showing that finding a best response for the players, given arbitrary networks G_1, G_2, G_I, and arbitrary non-increasing benefit functions $b_i(\cdot)$ and edge costs $c_i > 0$ for the players, is impossible in polynomial-time (unless P = NP).

Theorem 2. *The Best Response Interconnection problem is NP-hard.*

To prove this theorem, we provide a reduction from the NP-complete Dominating Set Problem [19]. A dominating set of the network $G_d = (V_d, E_d)$ is a subset $S \subseteq V_d$ such that for all $u \in V_d \setminus S$, u has a neighbor in the set S.

Definition 14. Dominating Set Problem.
INSTANCE: *Network $G_d = (V_d, E_d)$ and positive integer $k \le |V_d|$.*
QUESTION: *Does the network G_d have a dominating set S with $|S| \le k$?*

Below, we have provided the proof of the Theorem 2.

Proof of Theorem 2. Given an instance of the dominating set problem with $G_d = (V_d, E_d)$ and k, define an instance of the BRI problem with $G_1 = (V_1, E_1), G_2 = (V_2, E_2)$ and $G_I = (V_1 \cup V_2, E_I)$ as follows

$$\begin{aligned} V_1 &= \{x_1\},\ E_1 = \phi \\ V_2 &= V_d,\ E_2 = E_d \\ E_I &= V_1 \times V_2 \\ b_1(3) &< b_1(1) - c_1 < b_1(2) \\ r &= k(b_1(1) - c_1) + (|V_2| - k)b_1(2). \end{aligned} \tag{4.2}$$

In the above instance of the BRI, there is only one player P_1 associated with the node x_1. Hence, the BRI problem is to determine whether P_1 has an action W_1 such that $u_1(W_1|G_1, G_2, G_I) \ge r$.

Clearly, construction of the above instance of the BRI problem can be done in polynomial time. In the rest of the proof, we show that the answer to the above instance of the BRI problem is "yes" if and only if the answer to the instance of the Dominating Set Problem is "yes".

Assume that the graph $G_2 = G_d$ has a dominating set $S \subset V_2$ with $|S| \le k$ and thus the answer to the given instance of the Dominating set problem is "yes". Then by defining $W_1 = \{(x_1, v)|v \in S\}$, the distance between node x_1 and any node in V_2 is at most 2. Since $|W_1| \le k$, we have

$$\begin{aligned} u_1(W_1|G_1, G_2, G_I) &= |W_1|(b_1(1) - c_1) + (|V_2| - |W_1|)b_1(2) \\ &= |W_1|(b_1(1) - c_1) + (|V_2| - k)b_1(2) + (k - |W_1|)b_1(2) \\ &\ge |W_1|(b_1(1) - c_1) + (|V_2| - k)b_1(2) + (k - |W_1|)(b_1(1) - c_1) \\ &= r. \end{aligned}$$

Therefore, the answer to the constructed instance of the BRI problem in (4.2) is "yes" as well.

Next suppose that the answer to the defined instance of BRI in (4.2) is "yes", i.e., there exists a $W_1 \in S_1$ such that $u_1(W_1|G_1, G_2, G_I) \geq r$. Since $b_1(1) - c_1 > b_1(3)$, if there is a node $v \in V_2$ such that $d_G(x_1, v) \geq 3$, we can add the edge (x_1, v) to W_1 and increase the utility of P_1. Thus without loss of generality we can take the distance between node x_1 and any node in V_2 to be at most 2 under the constructed edge set W_1.

Consider the set of nodes $S \subseteq V_2$ that are incident to at least one edge in W_1, i.e., $S = \{v \in V_2 | (x_1, v) \in W_1\}$. All of the nodes in $V_2 \setminus S$ are connected to at least one of the nodes in S due to the assumption that the distance between any node in V_2 and node x_1 is at most 2. Thus S is a dominating set of the network G_2. On the other hand, the assumption that $u_1(W_1|G_1, G_2, G_I) \geq r$ yields

$$\begin{aligned}
0 &\leq u_1(W_1|G_1, G_2, G_I) - r \\
&\leq |W_1|(b_1(1) - c_1) + (|V_2| - |W_1|)b_1(2) - r \\
&= (|W_1| - k)(b_1(1) - c_1) + (k - |W_1|)b_1(2) \\
&= (|W_1| - k)(b_1(1) - c_1 - b_1(2)).
\end{aligned}$$

Since $b_1(1) - c_1 < b_1(2)$, we must have that $|W_1| \leq k$. Hence, $|S| = |W_1| \leq k$. This means that network G_2 has a dominating set of size less than k and thus the answer to the given instance of the Dominating Set Problem is "yes". □

Given that BRI is a NP-hard problem, finding a Nash equilibrium instance of the INDG with distance utilities is nontrivial in general. In the next section, we provide a set of properties of best response actions which are helpful in characterizing the best responses of the players in certain cases.

4.2.2 Properties of the Best Response

Lemma 11. *Let W_j be a best response to W_{-j} in the INDG*

$$\mathcal{G} = (P, (S_i)_{P_i \in P}, (u_i)_{P_i \in P}, G_1, G_2, G_I).$$

Then we have that

1. $|W_j| \leq |I_j|$.
2. *If $b_j(1) > b_j(2)$, then $|W_j| = |I_j|$ if and only if $W_j = \{(x_j, y) | y \in I_j\}$.*

Proof. Let $G = (V_1 \cup V_2, E_1 \cup E_2 \cup W_j \cup W_{-j})$. We use contradiction to prove the first statement. Assume that $|W_j| > |I_j|$, then

$$\begin{aligned} u_j(W_j \cup W_{-j}|G_1, G_2, G_I) &= \sum_{y \in I_j} b_j(d_G(x_j, y)) - c_j|W_j| \\ &\leq |I_j|b_j(1) - c_j|W_j| \\ &< |I_j|(b_j(1) - c_j) \\ &= u_j(W_j' \cup W_{-j}|G_1, G_2, G_I), \end{aligned}$$

where in the above $W_j' = \{(x_j, y)|y \in I_j\}$. Thus W_j is not a best response to W_{-j} which is a contradiction to the assumption of the lemma.

To prove the second statement, note that if $W_j = \{(x_j, y)|y \in I_j\}$, then $|I_j| = |W_j|$. Thus we only have to show that when $b_j(1) > b_j(2)$, if $|I_j| = |W_j|$, then $W_j = \{(x_j, y)|y \in I_j\}$. Assume by way of contradiction that there exists $y^* \in I_j$ such that $(x_j, y^*) \notin W_j$. This means that $d_G(x_j, y^*) \geq 2$ and thus

$$\begin{aligned} u_j(W_j \cup W_{-j}|G_1, G_2, G_I) &= \sum_{y \in I_j} b_j(d_G(x_j, y)) - c_j|W_j| \\ &< |I_j|b_j(1) - c_j|I_j| \\ &= u_j(W_j' \cup W_{-j}|G_1, G_2, G_I), \end{aligned}$$

where in the above $W_j' = \{(x_j, y)|y \in I_j\}$. This is a contradiction and thus we must have $\{(x_j, y)|y \in I_j\} \subseteq W_j$. We also know that $|W_j| \leq |I_j|$ and therefore, have the required result. □

The next lemma characterizes a best response action of the players when the cost of constructing edges is less than a certain threshold. The proof follows the same reasoning as the proof in [48] for the formation of (single) networks under low edge costs.

Lemma 12. *Let W_j be a best response to W_{-j} in the INDG*

$$\mathcal{G} = (P, (S_i)_{P_i \in P}, (u_i)_{P_i \in P}, G_1, G_2, G_I).$$

If $c_j < b_j(1) - b_j(2)$, then $W_j = \{(x_j, y)|y \in I_j\}$. Furthermore, if $c_j = b_j(1) - b_j(2)$, then $W_j = \{(x_j, y)|y \in I_j\}$ is a best response action for player P_j.

Proof. Suppose that $y^* \in I_j$ and $(x_j, y^*) \notin W_j$. Then $b_j(d_G(x_j, y^*)) \leq b_j(2)$ where $G = (V_1 \cup V_2, E_1 \cup E_2 \cup W_j \cup W_{-j})$. Adding the edge (x_j, y^*) to W_j increases the utility of W_j by at least $b_j(1) - c_j - b_j(2) > 0$ which contradicts the assumption that W_j is a best response and thus $(x_j, y^*) \in W_j$. Hence $\{(x_j, y)|y \in I_j\} \subseteq W_j$. By Lemma 11, we know that $|W_j| \leq |I_j|$ and therefore, $W_j = \{(x_j, y)|y \in I_j\}$.

For the case that $c_j = b_j(1) - b_j(2)$, note that adding the edge (x_j, y^*) to W_j does not decrease the utility of W_j and thus as in the above argument, $W_j = \{(x_j, y)|y \in I_j\}$ is a best response action for P_j. $\square$

The next result gives an upper-bound on the maximum number of edges that a player P_j with $c_j > b_j(1) - b_j(2)$ will form in a Nash equilibrium.

Lemma 13. *Let W_j be a best response to W_{-j} in the INDG*

$$\mathcal{G} = (P, (S_i)_{P_i \in P}, (u_i)_{P_i \in P}, G_1, G_2, G_I).$$

If $b_j(1) - b_j(2) < c_j$, then $|W_j| \leq |D|$, where D denotes the smallest dominating set of the network G_2.

Proof. If $|I_j| \leq |D|$, we have the result by the first statement of Lemma 11. Thus consider the case that $|I_j| > |D|$. Assume by way of contradiction that $|W_j| > |D|$. Let $G = (V_1 \cup V_2, E_1 \cup E_2 \cup W_j \cup W_{-j})$. Then

$$\begin{aligned}
u_j(W_j \cup W_{-j}|G_1, G_2, G_I) &= \sum_{y \in I_j} b_j(d_G(x_j, y)) - c_j|W_j| \\
&\leq |W_j|(b_j(1) - c_j) + (|I_j| - |W_j|)b_j(2) \\
&= |D|(b_j(1) - c_j) + (|W_j| - |D|)(b_j(1) - c_j) + (|I_j| - |W_j|)b_j(2) \\
&< |D|(b_j(1) - c_j) + (|I_j| - |D|)b_j(2) \\
&= u_j(W'_j \cup W_{-j}|G_1, G_2, G_I),
\end{aligned}$$

where $W'_j = \{(x_j, y)|y \in D\}$. Thus W'_j produces more utility than W_j for player P_j which is a contradiction to the assumption that W_j is a best response action to W_{-j}. $\square$

We use Lemma 13 in Section 4.3 to determine a Nash equilibrium instance of the INDG when G_2 has a star subgraph.

Based on the distance utility function for the players in (4.1), increasing the number of edges decreases the utility. Therefore, there must exist a threshold for the edge costs above which it is not beneficial to construct any edges. The next lemma provides such a threshold.

Lemma 14. *Let W_j be a best response to W_{-j} in the INDG*

$$\mathcal{G} = (P, (S_i)_{P_i \in P}, (u_i)_{P_i \in P}, G_1, G_2, G_I).$$

If $c_j > b_j(1) + (|I_j| - 1)b_j(2)$, then $W_j = \phi$, i.e., it is not beneficial for the player P_j to construct any edges incident to its associated node x_j.

Proof. Assume by way of contradiction that $|W_j| \geq 1$. Given $G = (V_1 \cup V_2, E_1 \cup E_2 \cup W_j \cup W_{-j})$, we have

$$\begin{aligned} u_j(W_j \cup W_{-j} | G_1, G_2, G_I) &= \sum_{y \in I_j} b_j(d_G(x_j, y)) - c_j |W_j| \\ &\leq |W_j|(b_j(1) - c_j) + (|I_j| - |W_j|)b_j(2) \\ &= b_j(1) - c_j + (|W_j| - 1)(b_j(1) - c_j) + (|I_j| - |W_j|)b_j(2) \\ &\leq b_j(1) - c_j + (|I_j| - 1)b_j(2) < 0, \end{aligned}$$

where in the above, we are using the fact that $b_j(1) - c_j < b_j(2)$ by the assumption of the lemma. Therefore, we must have that $|W_j| = 0$ which yields the required result. □

In the next result, we propose a condition under which a player disregards the network constructed by another player when considering the best response. We define the *R-radius* of a player $P_i \in P$ with $b_i(1) - c_i > 0$ as the minimum integer $R_i > 0$ (or ∞) such that $b_i(1) - c_i > b_i(R_i + 1)$.

Lemma 15. *Consider two players $P_i, P_j \in P$ with R-radii R_i and R_j, respectively. For a given instance of INDG*

$$\mathcal{G} = (P, (S_i)_{P_i \in P}, (u_i)_{P_i \in P}, G_1, G_2, G_I),$$

assume that W_i and W_j are best response actions to W_{-i} and W_{-j}, respectively. If $d_{G_1}(x_i, x_j) \geq R_i + R_j - 1$, then the actions of the players P_i and P_j are such that shortest paths from nodes x_i and x_j to the nodes that they depend on in V_2 are node disjoint in G_1.

Proof. The idea behind the proof stems from the fact that for any two nodes $x_i, x_j \in V_1$ with $d_{G_1}(x_i, x_j) \geq R_i + R_j - 1$, there does not exist any node $x_k \in V_1$ that simultaneously has distance less than R_i to x_i and less than R_j to x_j. To formally prove the lemma, consider $\{(x_i, y_i), (x_j, y_j)\} \subseteq E_I$. By way of contradiction, assume that the shortest paths from x_i to y_i and x_j to y_j intersect at a node $x_k \in V_1$. Without loss of generality, let $d_{G_1}(x_i, x_k) \geq R_i$. This means that $d_G(x_i, y_i) \geq R_i + 1$ where $G = (V_1 \cup V_2, E_1 \cup E_2 \cup W_i \cup W_{-i})$. Now consider $W_i' = W_i \cup \{(x_i, y_i)\}$ as a modified action of player P_i. This new action will increase the utility of player P_i by at least $b_i(1) - c_i - b_i(R_i + 1) > 0$, which is a contradiction to the assumption that W_i is a best response to W_{-i}. □

The following example illustrates the application of Lemma 15 in determining a Nash equilibrium of the INDG.

Example 7. *Consider networks $G_1 = (V_1, E_1)$ and $G_2 = (V_2, E_2)$ depicted in Fig. 4.1a with the given dependency network G_I between them (shown by dashed edges). Assume that $b_i(3) > b_i(1) - c_i > b_i(4)$ for $i \in \{1, 6\}$ which yields $R_1 = R_6 = 3$. Nodes x_1 and x_6 correspond to the players P_1 and P_6, respectively. Note that since all of the other nodes $x_i \in V_1 \setminus \{x_1, x_2\}$ have $I_i = \emptyset$, their associated players do not construct any edges in any Nash equilibrium by Lemma 11. Both x_1 and x_2 are dependent on all nodes in G_2, as illustrated for x_1 in Fig. 4.1b. The*

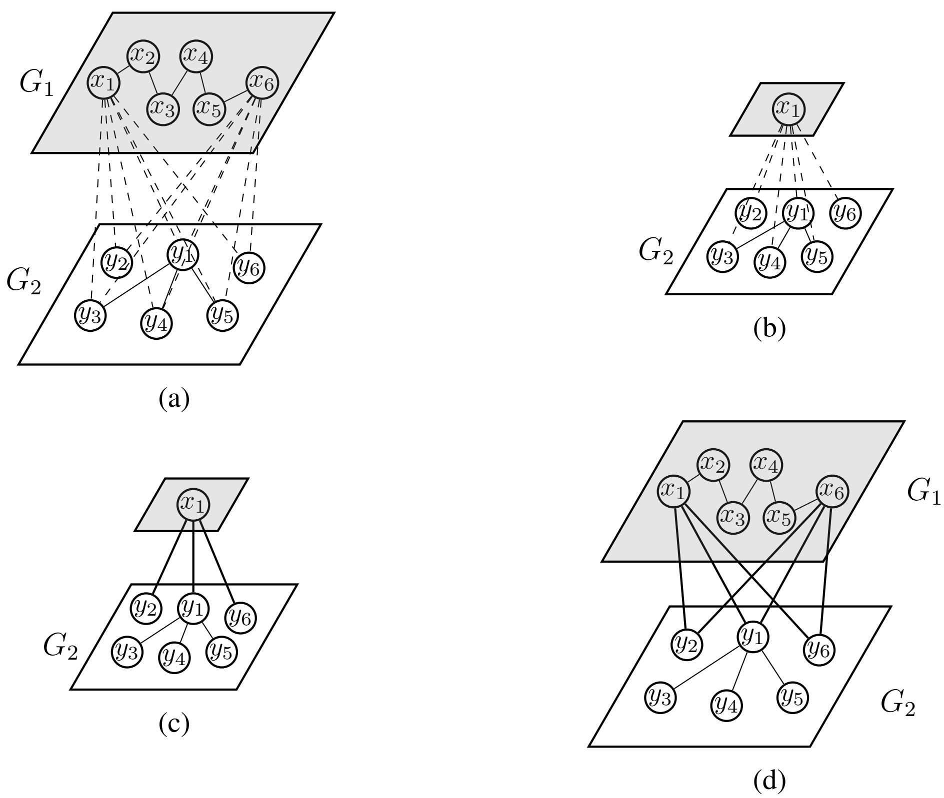


Figure 4.1: (a) Networks G_1 and G_2 with interdependencies shown by dashed edges. (b) Interdependencies of player P_1 with nodes in G_2. (c) Best response action of P_1 (d) A Nash equilibrium instance.

distance between nodes x_1 and x_6 in G_1 is 5 and thus the networks constructed by players P_1 and P_6 will be such that the shortest paths from x_1 to the nodes in G_2 are node disjoint (in G_1) from the shortest paths from x_6 to the nodes in G_2, by Lemma 15. Fig. 4.1c demonstrates a best *response for player P_1. Using the optimal action of P_1 and Lemma 15, we can determine a Nash equilibrium as shown in Fig. 4.1d.*

4.3 Nash Equilibrium of INDG for Networks Containing Star Subgraphs

With our results on best responses in hand, we now turn our attention to proving the existence of a Nash equilibrium. While it is challenging to show this for general G_1, G_2 and G_I, here we will prove that the INDG always has a Nash equilibrium when G_2 contains a star subgraph,[2] and $G_I = (V_1 \cup V_2, E_I)$ is the complete bipartite network, i.e., $E_I = V_1 \times V_2$. We allow G_1 to be arbitrary. Without loss of generality, let $y_1 \in V_2$ be the hub node in $G_2 = (V_2, E_2)$, i.e.,

$$\{(y_1, y)|y \in V_2,\ y \neq y_1\} \subseteq E_2.$$

As we illustrate later via an example, the presence of heterogeneous players (captured by individual benefit functions and edge costs) along with the arbitrary structure of G_1 leads to non-trivial interconnection networks in equilibrium, even under the above assumptions on G_2 and G_I.

To develop our results, we partition the set of players P into two sets: *high-cost players* $S_H = \{P_i \in P|b_i(1) - b_i(2) < c_i\}$ and *low-cost players* $S_L = \{P_i \in P|b_i(1) - b_i(2) \geq c_i\}$. Recall that we assumed $V_1 = \{x_1, \cdots, x_n\}$ and $V_2 = \{y_1, \cdots, y_m\}$. For the rest of this section, we denote the number of players $|P|$ by $|V_1| = n$ and the number of nodes in $|V_2|$ by m. We begin our analysis in this section with the following useful corollary of Lemma 12 which determines a best response action for the low-cost players.

Corollary 2. *Assume that $P_i \in S_L$. Then $W_i = \{(x_i, y)|y \in V_2\}$ is a best response action for player P_i regardless of the actions of the other players.*

In the remaining of this chapter, we assume that low-cost players always set their action according to the best response given by Corollary 2. In the next proposition, we discuss the best responses of high-cost players when there is a low-cost player in their neighborhood. We define the *L-radius* of a player P_i as the maximum nonnegative integer L_i such that

$$b_i(1) - c_i + (m-1)b_i(2) \leq mb_i(L_i + 1). \tag{4.3}$$

[2] Such networks can be used to represent, for example, sensor networks that have a fusion center, or transportation networks that have a "hub-and-spoke" structure [1, 82, 15].

Proposition 11. *Let P_i be a high-cost player. Suppose that there exists a low-cost player $P_j \in S_L$ such that the distance between x_j and x_i is less than $L_i + 1$ (i.e., $d_{G_1}(x_i, x_j) < L_i + 1$), where L_i is the L-radius of player P_i. Then, if P_j has constructed edges to all nodes in V_2, the empty network is a best response action for player P_i.*

Proof. Let W_i denote a best response action of the player $P_i \in S_H$ with respect to W_{-i}. Node x_i has distance $d \leq L_i$ to x_j which is associated with a low-cost player $P_j \in S_L$ that is connected to all of the nodes in V_2. Now assume that $W_i \neq \phi$. Then we have

$$\begin{aligned}
\Psi_i(W_1, \cdots, W_n) &= u_i(\cup_{j=1}^{n} W_j | G_1, G_2, G_I) \\
&= \sum_{y_j in I_i} b_i(d_G(x_i, y_j)) - c_i|W_i| \\
&\leq |W_i|(b_i(1) - c_i) + (m - |W_i|)b_i(2) \\
&\leq b_i(1) - c_i + (|W_i| - 1)b_i(2) + (m - |W_i|)b_i(2) \\
&= b_i(1) - c_i + (m-1)b_i(2) \leq mb_i(L_i + 1) \leq mb_i(d+1).
\end{aligned} \tag{4.4}$$

Therefore, player P_i can change its action to be the empty network and connect to the nodes it depends on in G_2 via edges constructed by the low-cost player P_j. □

The above result shows that the existence of a low-cost player in the proximity of a high-cost player will make the high-cost player a *free rider* in any Nash equilibrium, i.e., the high-cost player does not construct any edges and benefits from the low-cost player's edges.

Remark 11. *Note that Corollary 2 and Proposition 11 do not rely on G_2 having a star subgraph, and hold whenever the low-cost players have dependencies on all nodes in G_2.*

Corollary 3. *Assume that $P_i \in S_H$. Then for any best response action of the player P_i, node x_i is either connected to only the center of the star subgraph in G_2 (i.e., node $y_1 \in V_2$) or it does not have any edges.*

Proof. Since G_2 has a star subgraph, the size of its smallest dominating set is 1 (i.e., the center of the star, y_1). Therefore, by Lemma 13, we must have that $|W_i| \leq 1$. Furthermore, the proof of Lemma 13 shows that with a single edge, $W_i = \{(x_i, y_1)\}$ produces the highest possible utility for P_i. Proposition 11 gives an instance of the situation when $W_i = \phi$. □

Although Corollary 3 limits the set of best response actions of a high-cost player to two actions, it is not clear whether this game has a pure strategy Nash equilibrium for any set of players with arbitrary network G_1, edge cost c_i and benefit function $b_i(\cdot)$. We prove existence of

Nash equilibrium in this game by providing an algorithm that outputs such an equilibrium. In order to do this, we first need to define an index r_i for each high-cost player $P_i \in S_H$, called the r-*radius*. The r-radius[3] of player P_i with benefit function $b_i(\cdot)$ and edge cost c_i is defined as the maximum nonnegative constant r_i such that

$$b_i(1) - c_i + (m-1)b_i(2) \leq b_i(r_i+1) + (m-1)b_i(r_i+2). \tag{4.5}$$

Note that by the above definition, $L_i \geq r_i$ where L_i was defined in (4.3). For a given r-radius r_i, we define the r_i-neighborhood of node x_i as

$$N_i = \{x_j | P_j \in S_H \text{ and } d_{G_1}(x_i, x_j) \leq r_i\}. \tag{4.6}$$

If a high-cost player P_i has another high-cost player P_j with a single edge to V_2 such that $x_j \in N_i$, then player P_i is better off with no edge to V_2. This statement is also true if P_j is a low-cost player by Proposition 11 and the fact that $r_i \leq L_i$. The following proposition formally states these ideas.

Proposition 12. *Let P_i be a high-cost player with r-radius r_i. Suppose that there exists a player P_j such that $|W_j| \geq 1$ and $d_{G_1}(x_i, x_j) \leq r_i$. Then the empty network is a best response action for the player P_i with respect to W_{-i}.*

The set of propositions that we provided in this section enables us to give an algorithm that outputs a Nash equilibrium instance of the interconnection network design game with distance utilities for an arbitrary relation between the set of players (i.e., arbitrary network G_1) and arbitrary benefit function and cost of edges.

Theorem 3. *Assume that network G_2 has a star subgraph and G_I is a complete bipartite graph with partitions V_1 and V_2. Then the interconnection network design game with distance utilities in Definition 12 always has a pure strategy Nash equilibrium.*

Proof. We prove this theorem by construction. We provide an algorithm that outputs a Nash equilibrium instance of the game given by a set of actions $(W_1, W_2, \cdots, W_n)$ for the players. The steps of the algorithm are as follows:

1. Connect nodes associated to the low-cost players to all of the nodes in V_2.

2. Take S_H^∞ as the set of all high-cost players with $r_i = \infty$. Set the actions of all players $P_i \in S_H^\infty$ to be the empty network, i.e., $W_i = \emptyset$.

[3] A player P_i with $b_i(1) - c_i + (m-1)b_i(2) < 0$ is defined to have $r_i = \infty$, and their best response action is always the empty network by Lemma 14.

3. Determine the set S_H^L which consists of all high-cost players that have a low-cost player in their L_i-neighborhood where L_i denotes the L-radius, i.e.,

$$S_H^L = \{P_i \in S_H | \exists P_j \in S_L; d_{G_1}(x_i, x_j) \leq L_i\}$$

Set the actions of these players to be the empty network (by Proposition 11).

4. Consider the set of players whose actions have not been determined yet, and call it Q (we know that $Q \subseteq S_H \setminus (S_H^L \cup S_H^\infty)$). If the set Q is empty, exit the algorithm. Otherwise, let $P_i \in Q$ be the player with the lowest r-radius. Connect x_i (i.e., the node associated to P_i) via a single edge to the central node in G_2. Remove P_i from Q.

5. Set the action of all high-cost players $P_j \in Q$ with $x_i \in N_j$[4] to the empty network and remove them from the set Q.

6. Return to step 4.

We now argue that the output of the above algorithm is in fact a Nash equilibrium. Since the actions of low-cost players are in accordance with Corollary 2, they are in Nash equilibrium. The same is true for all high-cost players in S_H^∞ by Lemma 14. High cost players with a low-cost player in their L_i-neighborhood are also playing their optimal action, according to Proposition 11. Thus we only need to prove optimality of the actions of the remaining players which are determined through steps 4 to 6. Note that all of the remaining players have $r_i < \infty$.

Consider the set $S_H \setminus (S_H^L \cup S_H^\infty) = \{P_{i_1}, \cdots, P_{i_t}\}$ and assume without loss of generality that $r_{i_1} \leq r_{i_2} \leq \cdots \leq r_{i_t}$. Under the algorithm, the action of P_{i_1} is $W_{i_1} = \{(x_{i_1}, y_1)\}$. We know that there is no low-cost player in the L_{i_1}-neighborhood of P_{i_1}, since $P_{i_1} \in S_H \setminus (S_H^L \cup S_H^\infty)$. Similarly, there is no high-cost player with a nonempty action in the N_{i_1} neighborhood of P_{i_1}. Now assume that there exists a player $P_{i_j} \in S_H \setminus (S_H^L \cup S_H^\infty)$ with $j > 1$ and $|W_{i_j}| = 1$. We have to show that $x_{i_j} \notin N_{i_1}$, since otherwise the action of player P_{i_1} will not be optimal. In step 5 of the algorithm, we set the actions of all players P_{i_q} such that $x_{i_1} \in N_{i_q}$ to the empty network and remove them from the set Q. Hence, we must have that $x_{i_1} \notin N_{i_j}$, i.e.,

$$d_{G_1}(x_{i_1}, x_{i_j}) > r_{i_j} \geq r_{i_1}.$$

Therefore, $x_{i_j} \notin N_{i_1}$ and thus the action of player P_{i_1} is optimal.

The actions of all players whose actions are set to be the empty network in step 5 are optimal, by Proposition 12.

[4] N_j is the r_j-neighborhood of player P_j and was defined in (4.6).

Finally, consider any player P_{i_j} with $|W_{i_j}| = 1$ and $j > 1$. We know that $x_{i_k} \notin N_{i_j}$ for any $k < j$ with $|W_{i_k}| = 1$; otherwise the action of player P_{i_j} would have been set to the empty network in step 5 of the algorithm after assigning the action of player P_{i_k} in step 4. Moreover, by a reasoning similar to the argument for optimality of P_{i_1}'s action, we can show that for any player P_{i_t} with $t > j$ and $|W_{i_t}| = 1$, $x_{i_t} \notin N_{i_j}$. Therefore, the action of player P_{i_j} is a best response.

Thus, each player is playing their best response given the actions of the rest of the players, which implies that the given vector of actions is a Nash equilibrium. $\square$

Algorithm 1 gives the pseudo-code of the algorithm that we explained in the proof of Theorem 3. The following example illustrates the steps of the algorithm, and the corresponding Nash equilibrium.

Figure 4.2: (a) Network G_1 (b) Network G_2.

Example 8. *Consider two networks $G_1 = (V_1, E_1)$ and $G_2 = (V_2, E_2)$ depicted in Figures 4.2a and 4.2b with complete dependencies between nodes in G_1 and G_2. Assume that the cost of constructing edges is equal to 1 for all of the players, i.e., $c_i = 1, 1 \leq i \leq 9$. Suppose the benefit functions for the players take the values given in Table 4.1. Based on these values, player 7 is a low-cost player (since $c_7 < b_7(1) - b_7(2)$) and the rest of the players have high edge costs, i.e.,*

$$S_L = \{P_7\}$$
$$S_H = \{P_1, P_2, \cdots, P_6, P_8, P_9\}.$$

The corresponding values of the radii r_i and L_i (given by inequalities (4.5) and (4.3), respectively) are shown in the table.

We now follow the algorithm prescribed in the proof of Theorem 3.

1. *P_7 is the only low-cost player, and thus we connect x_7 to all of the nodes in G_2, i.e.,*

$$W_7 = \{(x_7, y_i) | 1 \leq i \leq 7\}.$$

Algorithm 1 Find Nash Equilibria

Input: Graphs G_1 and G_2, benefit functions $b_i(\cdot)$ and edge cost c_i for each player

Output: A set of actions for players that constitutes a Nash equilibrium instance of the Network Design Game with Distance Utilities

for each vertex $x_i \in V_1$ **do**
    Set[i]=0
    if $b_i(1) - c_i > b_i(2)$ **then**
        Set[i]=1
        $W_i = \{(x_i, y_j) | y_j \in V_2\}$
    end if
end for
for each vertex $x_i \in V_1$ & Set[i]=0 **do**
    $L_i = \arg\max_{L \geq 0}\{b_i(1) - c_i + (m-1)b_i(2) \leq mb_i(L+1)\}$
    if $\exists x_j \in V_1$ with set[j]=1 & $d_{G_1}(x_i, x_j) \leq L_i$ **then**
        Set[i]=2
        $W_j = \{\}$
    end if
end for
for each vertex $x_i \in V_1$ & Set[i]=0 **do**
    $r_i = \arg\max_{r \geq 0}\{b_i(1) - c_i + (m-1)b_i(2) \leq b_i(r+1) + (m-1)b_i(r+2)\}$
    $N_i = \{x_j | d_{G_1}(x_i, x_j) \leq r_i\}$
end for
while $\exists x_i \in V_1$ s.t. Set[i]=0 **do**
    $j = \arg\min_{Set[i]=0}\{r_i\}$
    Set[j]=3
    $W_j = \{(x_j, y_1)\}$
    for each $x_k \in V_1$ such that Set[k]=0 & $x_j \in N_k$ **do**
        Set[k]=2
    end for
end while

2. *For each node v_i whose distance to the low-cost player x_7 is at most L_i, we set that player's action to be empty. These nodes are given by $\{P_1, P_3, P_8, P_9\}$, and thus $W_1 = W_3 = W_8 = W_9 = \emptyset$.*

3. *The second player has the lowest r-radius among the remaining players and thus we set its action to $W_2 = \{(x_2, y_1)\}$. Since $\nexists P_j, j \in \{4, 5, 6\}$ such that $x_2 \in N_j$, we must choose*

	$b_i(1)$	$b_i(2)$	$b_i(3)$	$b_i(4)$	$b_i(5)$	L_i	r_i
P_1	1.5	1.3	1.2	1.1	0.2	2	1
P_2	1.2	0.8	0.5	0.2	0	1	0
P_3	1.1	0.9	0.1	0	0	1	0
P_4	0.9	0.8	0.7	0.5	0.2	2	1
P_5	1.2	1.1	0.9	0.2	0.1	1	0
P_6	1.3	1	0.5	0.4	0.3	1	0
P_7	3	1	0.5	0.5	0.4	NA	NA
P_8	1.2	0.8	0.7	0.5	0.4	1	1
P_9	1.2	1.1	1.1	1	0.2	3	2

Table 4.1: Benefit function, r-radius and L-radius of the players in Example 8

the next player with the lowest r_i. Recall that N_j was defined in (4.6).

4. *Player P_5 with $r_5 = 0$ has the lowest r-radius among the remaining players. Thus we set $W_5 = \{(x_5, y_1)\}$. Again since $\nexists P_j, j \in \{4, 6\}$ such that $x_5 \in N_j$, we must choose the next player with the lowest r_i.*

5. *Finally, we choose player P_6 with $r_6 = 0$ and set its action to $W_6 = \{(x_6, y_1)\}$. Due to the fact that $x_6 \in N_4$, we set the action of player P_4 to the empty network, i.e., $W_4 = \phi$.*

Fig. 4.3 demonstrates the output of the algorithm given in the proof of Theorem 3 when networks G_1 and G_2 depicted in Fig. 4.2b are given as input.

As one can see, the role of the hub nodes is crucial in the structure of the Nash equilibrium interconnection networks. While low edge cost players connect their associated nodes in G_1 to all of the nodes in the network G_2 (and thus themselves become hubs), the remaining high-cost players either choose (I) the empty network and connect via edges constructed by other players or (II) they directly connect to the hub node in network G_2.

4.3.1 Price of Anarchy

The concept of "price of anarchy" (PoA) was introduced in [58] to measure how selfish behavior of the individual players degrades the efficiency of the output in a non-cooperative game.

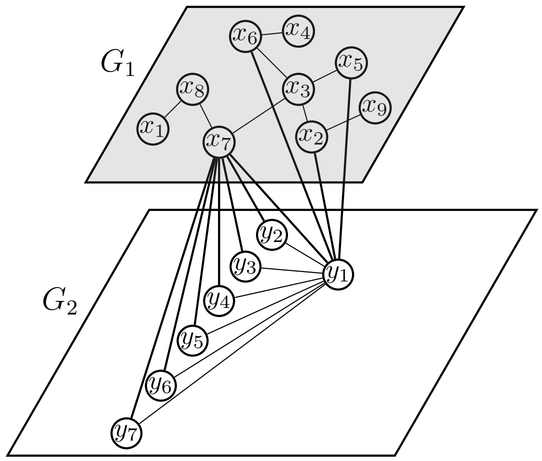


Figure 4.3: Networks G_1 and G_2 with the Nash equilibrium interconnection network G_p connecting them. Network G_p was produced by the algorithm given in the proof of Theorem 3.

Given a strategy $W = (W_1, W_2, \ldots, W_n)$ taken by the players and $T(W) = \sum_{i=1}^{n} u_i(W_i \cup W_{-i}|G_1, G_2, G_I)$ as the social welfare function, PoA is defined as

$$PoA = \frac{\max_{W \in S} T(W)}{\min_{W \in E} T(W)},$$

where in the above S denotes the joint strategy space and $E \subseteq S$ is the set of strategies in Nash equilibrium.

We show via the following example that the PoA can be arbitrarily large in the INDG.

Example 9. *Consider two networks $G_1 = (V_1, E_1)$ and $G_2 = (V_2, E_2)$, each containing star subgraphs centered on nodes $x_1 \in V_1$ and $y_1 \in V_2$, respectively, i.e.,*

$$\{(x_1, x_i)|x_i \in V_1\} \subseteq E_1$$
$$\{(y_1, y_i)|y_i \in V_2\} \subseteq E_2.$$

Suppose that we have full dependencies between nodes in V_1 and V_2, i.e., $G_I = (V_1 \cup V_2, V_1 \times V_2)$. Assume that all of the players $P_i, 1 \leq i \leq |V_1|$ in the INDG have $c_i = 2.1, b_i(1) = 1$ and $b_i(2) = b_i(3) = 1/(|V_2|-1)$, where $|V_1| > |V_2| > 1$. This means that $b_i(1) - c_i + (|V_2|-1)b_i(2) = -0.1 < 0$ for all of the players $P_i \in P$ and thus by Lemma 14, none of the players constructs any edges. Therefore, the social welfare value $T(W) = 0$ for all strategies in Nash equilibrium.

Now consider the socially optimal interconnection strategy, i.e., the strategy that maximizes $T(\cdot)$. For the strategy $W^\star = (W_1^\star, W_2^\star, \ldots, W_n^\star)$ where $W_1^\star = \{(x_1, y_1)\}$ and $W_i^\star = \phi, i \neq 1$, we have that

$$\begin{aligned} T(W^\star) &= \sum_{i=1}^{n} u_i(W_i^* \cup W_{-i}^* | G_1, G_2, G_I) \\ &= b_1(1) - c_1 + (|V_2| - 1)b_1(2) + \sum_{j=2}^{|V_1|} (b_j(2) + (|V_2| - 1)b_j(3)) \\ &= -0.1 + \frac{(|V_1| - 1)|V_2|}{|V_2| - 1} > 0. \end{aligned}$$

Therefore, the network that maximizes the social welfare function has a nonzero utility and thus PoA is trivially infinite.

Remark 12. *The network $G_{SocOpt} = \cup_{i=1}^{n} W_i$ that maximizes the social utility function $T(W)$ is called the socially optimal network. Similar to the proof of the Theorem 2, one can show that finding the socially optimal network is an NP-hard problem.*

4.4 INDG vs Islands-Connection model

The interconnection network design problem that we investigated in this chapter has similarities to the Islands-Connection (IC) model that was mentioned in the Introduction (Chapter 1). While the IC model considers a homogeneous set of players, the INDG model includes the case that players have different cost and benefit functions. Furthermore, the topologies of networks G_1 and G_2 (which correspond to islands in the IC model) in INDG are fixed, whereas in IC the structure of the islands depends on the cost of intra-island edge construction. When the cost of intra-island and inter-island edge formation are lower than certain thresholds, [49] shows that there are complete connections inside the islands. In addition, while the distance between all pairs of inter-island nodes are taken into account in the IC model, our INDG model allows the interdependency network G_I to characterize the set of important inter-island distances.

4.5 Summary

We introduced the interconnection network design game between two networks G_1 and G_2. In this game, there is a heterogeneous (in terms of utility function) set of network designers, each associated with a node in the network G_1. Each node in G_1 is dependent on certain nodes in G_2, and these dependencies are captured by a network G_I. The utility of the players is defined based on the distance-utility function where the objective of each player is to build a set of edges from its associated node to nodes in the network G_2 such that distances between its associated node and the nodes it depends on in G_2 are minimized. We showed that finding a best response action of a player is NP-hard. Nevertheless, we showed certain important properties of the best response networks, which enabled us to find a Nash equilibrium for certain instances of the game.

Chapter 5

Random Interdependent Networks

In Chapter 4, we studied the optimal allocation of interconnection edges in an interdependent network. To complement our studies, in this chapter we use random network models to investigate certain spectral and structural properties of such networks, namely *edge expansion*, r*-robustness*, *algebraic connectivity* and the *smallest eigenvalue of the grounded Laplacian matrix*. In addition to their topological implications, these properties also play a key role in certain variants of diffusion dynamics on networks. Our analysis in this chapter is applicable to interdependent networks with an arbitrary number of subnetworks.

We consider the class of *random interdependent networks* consisting of k subnetworks, where each edge between nodes in different subnetworks is present independently with a certain probability p. Our model is fairly general in that we make no assumption on the topologies within the subnetworks, and captures Erdos-Renyi graphs and random k-partite graphs as special cases. We use the graph theoretic notion of *isoperimetric constant* (denoted by $i(G)$) as the key property to derive our results. Our first result characterizes a threshold p_r for random k-partite networks to have $i(G) > r - 1$ where r is a positive integer. Furthermore, we prove that p_r is *also* the threshold for the minimum degree of the network to be r. This is potentially surprising given that $i(G) > r - 1$ is a significantly stronger graph property than r-minimum-degree. Secondly, we show that when the probability of inter-network edge formation is sufficiently high, $i(G)$ scales as $\Theta(np)$, where n is the number of nodes in each subnetwork.

We then focus on the algebraic connectivity of random interdependent networks (defined as the second smallest eigenvalue of the Laplacian matrix). We show that when the inter-network edge formation probability p satisfies a certain condition, the algebraic connectivity of the network grows as $\Theta(np)$ asymptotically almost surely, *regardless* of the topologies within the subnetworks. Given the key role of algebraic connectivity in the speed of consensus dynamics on networks [73], our analysis demonstrates the importance of the edges that connect different communities in the network in terms of facilitating information spreading, in line with classical findings in the sociology literature [42]. Our result on algebraic connectivity of random interdependent networks is also applicable to the *stochastic block model* or *planted partition model* that has been widely studied in the machine learning literature [28, 2, 20, 65]. While we consider arbitrary intra-network topologies, in the planted partition model it is assumed that the intra-network edges are also placed randomly with a certain probability. Furthermore, the lower bound that we obtain here for $\lambda_2(L)$ is tighter than the lower bounds obtained in the existing planted partition literature for the range of edge formation probabilities that we consider [28].

Next, we provide a bound for the smallest eigenvalue of the grounded Laplacian matrix (obtained by removing certain rows and columns of the Laplacian matrix) of random interdependent networks. This eigenvalue dictates the rate of convergence in consensus dynamics in a network where some nodes do not change their states; such nodes are called *stubborn* or *leaders* (depending on the context). The rate of convergence induced by certain sets of stubborn nodes in consensus dynamics has been the subject of study over the past few years [18, 35, 79], and our results add to this literature by studying such dynamics in interdependent networks. We show that in the case where all of the nodes in one of the subnetworks are leaders or stubborn agents, the smallest eigenvalue of the grounded Laplacian scales as $\Theta(np)$ asymptotically almost surely provided that the probability of edge formation between a normal node and a leader node is sufficiently high.

Finally, we analyze a metric known as r-robustness of networks. In recent years, the robustness of interdependent networks to intentional disruption or natural malfunctions has started to attract attention by a variety of researchers [32, 84, 97]. As we will describe later, r-robustness has strong connotations for the ability of networks to withstand structural and dynamical disruptions: it guarantees that the network will remain connected even if up to $r-1$ nodes are removed from the neighborhood of *every* node in the network, and facilitates certain consensus dynamics that are resilient to adversarial nodes [62, 100, 23, 98, 92]. We identify a bound p_r for the probability of inter-network edge formation p such that for $p > p_r$, random interdependent networks with arbitrary intra-network topologies are guaranteed to be r-robust asymptotically almost surely. For the special case of k-partite random graphs, we show that this p_r is tight (i.e., it forms a threshold for the property of r-robustness). In fact, the aforementioned p_r is the same as the threshold for k-partite random graphs to have isoperimetric constant higher than $r-1$ which itself is equal to the threshold to have minimum degree of r. Recent work has shown that the two properties of r-minimum degree and r-robustness share the same thresholds in Erdos-Renyi random graphs [98] and random intersection graphs [100], and our work in this chapter adds random k-partite graphs to this list.

We start with defining each property and then discussing its implications for dynamics on such networks.

5.1 Background and Application

5.1.1 Isoperimetric Constant

The *edge-boundary* of a set of nodes $S \subset V$ is given by $\partial S = \{(v_i, v_j) \in E \mid v_i \in S, v_j \in V \backslash S\}$. The *isoperimetric constant* of G is defined as [16]

$$i(G) \triangleq \min_{A \subset V, |A| \leq \frac{n}{2}} \frac{|\partial A|}{|A|}. \tag{5.1}$$

By choosing A as the vertex with the smallest degree we obtain $i(G) \leq d_{min}$. Our results about the isoperimetric constant of random interdependent networks (given in Section 5.3) are at the heart of many of the subsequent results that we provide in this chapter.

5.1.2 Algebraic Connectivity

The second smallest eigenvalue of the Laplacian matrix $\lambda_2(L)$, is called the *algebraic connectivity* of the graph and is related to the isoperimetric constant by [16]

$$\frac{i(G)^2}{2d_{max}} \leq \lambda_2(L) \leq 2i(G). \tag{5.2}$$

Algebraic connectivity has important implications in various areas [30, 21]. For instance, algebraic connectivity is a lower bound for node and edge connectivity, i.e., the minimum number of nodes or edges that have to be removed in order to make the graph disconnected [52]. Here, we focus on the application of algebraic connectivity in consensus dynamics.

Consider a multi-agent setting with n agents and interaction topology modeled by the graph $G = (V, E)$, where each node of G corresponds to an agent. There is an edge between two nodes in the graph G if their corresponding agents communicate or exchange information. Associated with each agent $v_i \in V$ is an initial state (an opinion, decision, measurement, etc.) which is represented by a real value $x_i(0) \in \mathbb{R}$ and evolves over time as a function of the states of v_i's neighbors. Assume that each agent updates its state $x_i(t)$ as

$$\dot{x}_i(t) = \sum_{v_j \in \mathcal{N}_i} (x_j(t) - x_i(t)).$$

The system-wide dynamics can then be represented by

$$\dot{X}(t) = -LX(t), \tag{5.3}$$

where $X(t) = [x_1(t), x_2(t), \ldots, x_n(t)]^T$ denotes the vector of states of all of the nodes and L denotes the Laplacian matrix. When G is a connected graph, the state of all of the agents converges to $\mathbf{1}^T X(0)/n$ (the average of the initial values) and the asymptotic convergence rate is given by $\lambda_2(L)$ [73]. We provide a tight bound for $\lambda_2(L)$ in Section 5.4 when the underlying graph belongs to the class of random interdependent networks.

5.1.3 Smallest Eigenvalue of the Grounded Laplacian

Next, consider the consensus setting with a group of agents $S \subset V$ who keep their states constant, i.e., $\forall v_s \in S, \exists x_s \in \mathbb{R}$ such that $x_s(t) = x_s,\ \forall t \geq 0$. Depending on the context, these agents are called *stubborn agents* or *leaders* [35, 79]. Let X_F and X_S denote the states of the follower and stubborn agents, respectively. Then the equation (5.3) can be written as

$$\begin{bmatrix} \dot{X}_F(t) \\ \dot{X}_S(t) \end{bmatrix} = - \begin{bmatrix} L_{11} & L_{12} \\ L_{21} & L_{22} \end{bmatrix} \begin{bmatrix} X_F(t) \\ X_S(t) \end{bmatrix}. \tag{5.4}$$

Matrices L_{21} and L_{22} are both zero by the definition of stubborn agents. Matrix L_{11} is called the *grounded Laplacian* of the system and is denoted by $L_{11} = L_g(S)$; we drop the argument S whenever it is clear from the context. It can be shown that the state of each follower agent asymptotically converges to a convex combination of the values of the stubborn agents with convergence rate given by $\lambda(L_g)$, the *smallest eigenvalue of the grounded Laplacian* [18]. Furthermore, the smallest eigenvalue of the grounded Laplacian is inversely proportional to the H_∞ coherence metric which is used to measure deviations of the non-leader agents from their steady state value when they are affected by noise [78]. In Section 5.5, we obtain a bound for $\lambda(L_g)$ when G is a random interdependent network.

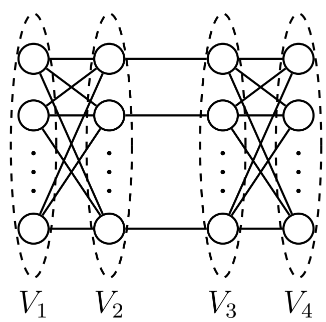
5.1.4 The Notion of r-Robustness

We use the following metric known as r-*robustness* to study robustness of networks against structural and dynamical disruptions.

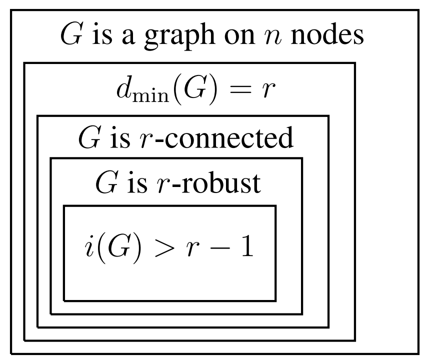
Definition 15 ([62]). *Let $r \in \mathbb{N}$. A subset S of nodes in the graph $G = (V, E)$ is said to be r-reachable if there exists a node $v_i \in S$ such that $|\mathcal{N}_i \setminus S| \geq r$. A graph $G = (V, E)$ is said to be r-robust if for every pair of nonempty, disjoint subsets $V_1, V_2 \subseteq V$, either V_1 or V_2 is r-reachable.*

Simply put, an r-reachable set contains a node that has r neighbors outside that set, and an r-robust graph has the property that no matter how one chooses two disjoint nonempty sets, at least one of those sets is r-reachable. The property of r-robustness is significantly stronger than the property of r-connectivity (or r-minimum degree); an r-robust graph will remain connected even after up to $r-1$ nodes are removed from the neighborhood of *every* remaining node, while an r-connected graph will only guarantee connectedness after the removal of $r-1$ nodes *in total* [62, 98]. Indeed, the gap between the robustness and node connectivity (and minimum degree) parameters can be arbitrarily large, as illustrated by the graph G in Fig. 5.1a. While the minimum degree and node connectivity of the graph G is $n/4$, it is only 1-robust (consider disjoint subsets $V_1 \cup V_2$ and $V_3 \cup V_4$).

The following result shows that the isoperimetric constant $i(G)$ provides a lower bound on the robustness parameter.



(a)



(b)

Figure 5.1: (a) Graph $G = (V, E)$ with $V = V_1 \cup V_2 \cup V_3 \cup V_4$ and $|V_i| = \frac{n}{4}$, $1 \leq i \leq 4$. All of the nodes in V_1 (V_3) are connected to all of the nodes in V_2 (V_4). Furthermore, there is a one to one connection between nodes in V_2 and nodes in V_3. (b) Relationships between different notions of robustness.

Lemma 16. *Let r be a positive integer. If $i(G) > r-1$, then the graph is at least r-robust.*

Proof. If $i(G) > r-1$, then every set of nodes $S \subset V$ of size up to $\frac{n}{2}$ has at least $(r-1)|S|+1$ edges leaving that set (by the definition of $i(G)$). By the pigeonhole principle, at least one node in S has at least r neighbors outside S. Now for any two disjoint non-empty sets S_1 and S_2, at least one of these sets has size at most $\frac{n}{2}$, and thus is r-reachable. Therefore, the graph is r-robust. $\square$

Note that together with (5.2), the above lemma implies that any graph is at least $\lceil \frac{\lambda_2(L)}{2} \rceil$-robust. As an example of Lemma 16, consider the graph G in Fig. 5.1a which has isoperimetric constant of at most 0.5 (since the edge boundary of $V_1 \cup V_2$ has size $n/4$), but is 1-robust. In fact, the gap between the robustness parameter and isoperimetric constant can be arbitrarily large. For instance, given any arbitrary $t \in \mathbb{N}$ and n sufficiently large, assume that each node in V_2 (V_3) is connected to exactly t nodes in V_3 (V_2)) and the rest of the graph is the same as the structure shown in Fig. 5.1a. Then the isoperimetric constant of the proposed graph is at most $t/2$. However, one can show that the constructed graph is t-robust. The relationships between these different graph-theoretic measures of robustness are summarized in Fig. 5.1b.

In order to see application of r-robustness in consensus dynamics, consider the case where agents synchronously update their states according to the following filtering dynamics [62, 98]: given $F \in \mathbb{N}$, at each time step, each node receives the values of its neighbors, disregards the largest and the smallest F values (breaking ties arbitrarily) and updates its state as

$$x_i[k+1] = w_{ii}[k]x_i[k] + \sum_{j \in \mathcal{R}_i[k]} w_{ij}[k]x_j[k], \tag{5.5}$$

where $\mathcal{R}_i[k]$ represents the set of nodes whose values were adopted by node i at time step k. In (5.5), $w_{ii}[k]$ and $w_{ij}[k]$ are the weights at time step k which satisfy $\sum_{j \in \mathcal{R}_i[k]} w_{ij}[k] = 1, \forall i \in V,\ k \in \mathbb{Z}_{\geq 0}$. Then, consensus of the agents is guaranteed if and only if the underlying graph G is at least $(F+1)$-robust [62]. Furthermore, if there are up to F adversarial (arbitrarily behaving) nodes in the neighborhood of every normal node, then under these dynamics, all regular nodes will converge to consensus in the convex hull of the initial values of the normal nodes as long as the graph is $(2F+1)$-robust [62, 98].

5.2 Random Interdependent Networks

We investigate the properties that we discussed in the last section for the class of *random interdependent networks*. The formal definition of this class of networks is given below.

Definition 16. *An interdependent network G is denoted by a tuple $G = (G_1, G_2, \dots, G_k, G_p)$ where $G_i = (V_i, E_i)$ for $i = 1, 2, \dots, k$ are called the subnetworks of the network G, and $G_p = (V_1 \cup V_2 \cup \dots \cup V_k, E_p)$ is a k-partite network with $E_p \subseteq \cup_{i \neq j} V_i \times V_j$ specifying the interconnection (or inter-network) topology.*

In the remaining of this chapter, we assume that $|V_1| = |V_2| = \cdots = |V_k| = n$ and that the number of subnetworks k is at least 2.

Definition 17. *Define the probability space* $(\Omega_n, \mathcal{F}_n, \mathbb{P}_n)$*, where the sample space* Ω_n *consists of all possible interdependent networks* $(G_1, G_2, \ldots, G_k, G_p)$ *and the index* $n \in \mathbb{N}$ *denotes the number of nodes in each subnetwork. The* σ*-algebra* $\mathcal{F}_n$ *is the power set of* Ω_n *and the probability measure* $\mathbb{P}_n$ *associates a probability* $\mathbb{P}(G_1, G_2, \ldots, G_k, G_p)$ *to each network* $G = (G_1, G_2, \ldots, G_k, G_p)$*. A random interdependent network is a network* $G = (G_1, G_2, \ldots, G_k, G_p)$ *drawn from* Ω_n *according to the given probability distribution.*

Note that deterministic structures for the subnetworks or interconnections can be obtained as a special case of the above definition where $\mathbb{P}(G_1, G_2, \ldots, G_k, G_p)$ is 0 for interdependent networks not containing those specific structures; for instance, a *random* k*-partite graph* is obtained by allocating a probability of 0 to interdependent networks where any of the G_i for $1 \leq i \leq k$ is nonempty. Through an abuse of notation, we will refer to random k-partite graphs simply by G_p. Similarly, Erdos-Renyi random graphs on kn nodes are obtained as a special case of the above definition by choosing the edges in $G_1, G_2, \ldots, G_k$ and G_p independently with a common probability p.

In this chapter, we will focus on the case where G_p is independent of G_i for $1 \leq i \leq k$. Specifically, we assume that each possible edge of the k-partite network G_p is present independently with probability p (which can be a function of n). We will characterize certain properties of such networks as n gets large. Recall that for a random interdependent network (similar to the ER networks), we say that a property $\mathcal{P}$ holds *asymptotically almost surely* (a.a.s.) if the probability measure of the set of graphs with property $\mathcal{P}$ (over the probability space $(\Omega_n, \mathcal{F}_n, \mathbb{P}_n)$) tends to 1 as $n \to \infty$.

5.2.1 Application of Consensus Dynamics on Random Interdependent Networks

Consider a society with multiple communities, where individuals have inter-community and intra-community links. Modeling the interaction among individuals by an interdependent network (where each subnetwork represents a community), the consensus (or opinion) dynamics described in the previous section can be interpreted as follows.

- The case that individuals update their decision by aggregating the opinions of all of their neighbors corresponds to the standard consensus dynamics given by equation (5.3). Then, the algebraic connectivity of the interdependent network determines how fast information spreads throughout the network.

- Equation (5.4) models the situation that one of the communities acts as a *leader community* (i.e., its members keep their opinions fixed). In this case, the smallest eigenvalue of the grounded Laplacian matrix determines the speed at which the follower nodes converge to the steady state.

- Finally, the filtering dynamics in equation (5.5) generalize DeGroot opinion dynamics [22] by allowing the nodes to discard the most extreme opinions of their neighbors before averaging the rest [98]. The notion of r-robustness enables us to understand the ability of the network to facilitate consensus under such dynamics even when some individuals behave in an adversarial or erratic manner.

5.3 Isoperimetric Constant of Random Interdependent Networks

This section provides two important results about the isoperimetric constant of random interdependent networks. We start with the following lemma, giving a threshold for the isoperimetric constant of a random k-partite network to be greater than $r-1$.

Lemma 17. *Consider a random k-partite network G_p with n nodes in each subnetwork and inter-network edge formation probability $p = p(n)$. Let $x = x(n)$ be some function satisfying $x = o(\ln \ln n)$ and $x \to \infty$ as $n \to \infty$. Then for any positive integer r and $k \geq 2$,*

1. *If $p(n) = \frac{\ln n + (r-1)\ln\ln n + x(n)}{(k-1)n}$, then $i(G_p) > r-1$ (and thus the minimum degree is at least r) a.a.s., and*

2. *If $p(n) = \frac{\ln n + (r-1)\ln\ln n - x(n)}{(k-1)n}$, then the minimum degree is at most $r-1$ (and thus $i(G_p) \leq r-1$) a.a.s.*

Proof. First consider the case that the inter-network edge formation probability is

$$p = \frac{\ln n + (r-1)\ln\ln n + x}{(k-1)n},$$

where $x = o(\ln \ln n)$ and $x \to \infty$ when $n \to \infty$. We have to show that for any set of vertices of size m, $1 \leq m \leq nk/2$, there are at least $m(r-1)+1$ edges that leave the set a.a.s. Consider a set $S \subset V_1 \cup V_2 \cup \cdots \cup V_k$ with $|S| = m$. Assume that the set S contains s_i nodes from V_i for $1 \leq i \leq k$ (i.e., $|S \cap V_i| = s_i \geq 0$). Define E_S as the event that $m(r-1)$ or fewer edges leave S.

Note that $|\partial S|$ is a binomial random variable with parameters $\sum_{l=1}^{k} s_l \left(\sum_{t=1,t\neq l}^{k}(n-s_t)\right)$ and p. We have that

$$\begin{aligned}\sum_{l=1}^{k} s_l \left(\sum_{t=1,t\neq l}^{k}(n-s_t)\right) &= \sum_{l=1}^{k} s_l(n(k-1)-m+s_l)\\ &= n(k-1)m - m^2 + \sum_{l=1}^{k} s_l^2. \end{aligned} \tag{5.6}$$

Then we have

$$\begin{aligned}\mathbf{Pr}(E_S) &= \sum_{i=0}^{m(r-1)} \binom{n(k-1)m - m^2 + \sum_{l=1}^{k} s_l^2}{i} p^i (1-p)^{n(k-1)m-m^2+\sum_{l=1}^{k} s_l^2 - i}\\ &\le \sum_{i=0}^{m(r-1)} \binom{n(k-1)m}{i} p^i (1-p)^{n(k-1)m-m^2+\sum_{l=1}^{k} s_l^2 - i}. \end{aligned} \tag{5.7}$$

For $s_i \in \mathbb{R}, 1 \le i \le k$, we have

$$\sum_{i=1}^{k} s_i^2 \ge \frac{\left(\sum_{i=1}^{k} s_i\right)^2}{k}, \tag{5.8}$$

which is a direct consequence of the Cauchy-Schwartz inequality. Applying (5.8) to the inequality (5.7), we get $0 \le m^2 - \sum_{l=1}^{k} s_l^2 \le \frac{(k-1)m^2}{k}$ and thus

$$\mathbf{Pr}(E_S) \le \sum_{i=0}^{m(r-1)} \binom{n(k-1)m}{i} p^i (1-p)^{n(k-1)m - \frac{(k-1)m^2}{k} - i}. \tag{5.9}$$

Next note that $k \ge 2$ and for $1 \le i \le m(r-1)$, we have

$$\begin{aligned}\frac{\binom{n(k-1)m}{i} p^i (1-p)^{n(k-1)m-\frac{(k-1)m^2}{k}-i}}{\binom{n(k-1)m}{i-1} p^{i-1} (1-p)^{n(k-1)m-\frac{(k-1)m^2}{k}-(i-1)}} &= \frac{n(k-1)m-i+1}{i}\frac{p}{1-p}\\ &\ge \frac{n(k-1)m-m(r-1)+1}{m(r-1)} \times \frac{p}{1-p}\\ &\ge \frac{n(k-1)-r+1}{r-1} \times \frac{p}{1-p}, \end{aligned}$$

which is lower bounded by some constant strictly larger than 1 for sufficiently large n. Thus there exists some constant $R > 0$ such that

$$\begin{aligned}
\mathbf{Pr}(E_S) &\leq \sum_{i=0}^{m(r-1)} \binom{n(k-1)m}{i} p^i (1-p)^{n(k-1)m - \frac{(k-1)m^2}{k} - i} \\
&\leq R \binom{n(k-1)m}{m(r-1)} p^{m(r-1)} \times (1-p)^{n(k-1)m - \frac{(k-1)m^2}{k} - m(r-1)}.
\end{aligned} \tag{5.10}$$

Define P_m as the probability that there exists a set of nodes T such that $|T| = m$ and $|\partial T| \leq m(r-1)$. Then using the inequality $\binom{n}{m} \leq (\frac{ne}{m})^m$ yields

$$\begin{aligned}
P_m &\leq \sum_{\substack{|S|=m, \\ S \subset \cup_{i=1}^k V_i}} \mathbf{Pr}(E_S) \\
&\leq R \binom{nk}{m} \binom{n(k-1)m}{m(r-1)} p^{m(r-1)} (1-p)^{n(k-1)m - \frac{(k-1)m^2}{k} - m(r-1)} \\
&\leq R \left(\frac{nke}{m}\right)^m \left(\frac{n(k-1)mep}{m(r-1)}\right)^{m(r-1)} (1-p)^{n(k-1)m - \frac{(k-1)m^2}{k} - m(r-1)} \qquad (5.11) \\
&= R \left(\frac{ke^r}{(r-1)^{r-1}(1-p)^{r-1}} \frac{n(1-p)^{n(k-1)}(n(k-1)p)^{r-1}}{m(1-p)^{\frac{(k-1)m}{k}}} \right)^m \\
&\leq R \left(\frac{c_1 n(1-p)^{n(k-1)}(n(k-1)p)^{r-1}}{m(1-p)^{\frac{(k-1)m}{k}}} \right)^m,
\end{aligned}$$

where c_1 is some constant satisfying $\frac{ke^r}{(r-1)^{r-1}(1-p)^{r-1}} \leq c_1 < \frac{2ke^r}{(r-1)^{r-1}}$ for sufficiently large n. Recalling the function $p(n) = \frac{\ln n + (r-1)\ln\ln n + x}{(k-1)n}$ and using the inequality $1 - p \leq e^{-p}$ yields

$$\begin{aligned}
P_m &\leq R \left(\frac{c_1 n e^{-n(k-1)p}(n(k-1)p)^{r-1}}{m(1-p)^{\frac{(k-1)m}{k}}} \right)^m \\
&= R \left(c_1 \left(\frac{\ln n + (r-1)\ln\ln n + x}{\ln n} \right)^{r-1} \frac{e^{-x}}{m(1-p)^{\frac{(k-1)m}{k}}} \right)^m \\
&\leq R \left(\frac{c_2 e^{-x}}{m(1-p)^{\frac{(k-1)m}{k}}} \right)^m.
\end{aligned}$$

Due to the fact that $\frac{\ln n + (r-1)\ln\ln n + x}{\ln n} < 2$ for sufficiently large n, c_2 is a constant upper bound for $c_1 \left(\frac{\ln n + (r-1)\ln\ln n + x}{\ln n}\right)^{r-1}$ such that $0 < c_2 < c_1 2^{r-1}$. Next, we substitute the Taylor series expansion $\ln(1-p) = -\sum_{i=1}^{\infty} \frac{p^i}{i}$ for $p \in [0,1)$ in the above inequality to obtain

$$
\begin{aligned}
P_m &\leq R\left(\frac{c_2 e^{-x} e^{-\frac{(k-1)m}{k}\ln(1-p)}}{m}\right)^m \\
&= R\left(\frac{c_2 e^{-x} e^{\frac{(k-1)mp}{k}} \exp\{\frac{(k-1)m}{k} p^2 \sum_{i=2}^{\infty} \frac{p^{i-2}}{i}\}}{m}\right)^m.
\end{aligned}
$$

Since we have $\sum_{i=2}^{\infty} \frac{p^{i-2}}{i} < \sum_{i=2}^{\infty} p^{i-2} = \frac{1}{1-p}$ and $\frac{(k-1)m}{k}p^2 \to 0$ as $n \to \infty$, there exists a constant c_3 such that $0 < \frac{(k-1)m}{k}p^2 \sum_{i=2}^{\infty} \frac{p^{i-2}}{i} < c_3 < 1$ for sufficiently large n. Therefore,

$$
P_m \leq R\left(c_2 e^{c_3} \frac{e^{-x} e^{\frac{(k-1)mp}{k}}}{m}\right)^m = R\left(c_4 \frac{e^{-x} e^{\frac{(k-1)mp}{k}}}{m}\right)^m,
$$

where $0 < c_4 = c_2 e^{c_3} < \frac{ke^{r+1}2^r}{(r-1)^{r-1}}$.

Consider the function $f(m) = \frac{e^{\frac{(k-1)mp}{k}}}{m}$. Since $\frac{df}{dm} = \frac{e^{\frac{(k-1)mp}{k}}(\frac{(k-1)mp}{k}-1)}{m^2}$, we have that $\frac{df}{dm} < 0$ for $m < \frac{k}{(k-1)p}$ and $\frac{df}{dm} > 0$ for $m > \frac{k}{(k-1)p}$. Therefore, $f(m) \leq \max\{f(1), f(nk/2)\}$ for $m \in \{1, 2, \ldots, \lfloor nk/2 \rfloor\}$. We have

$$
\begin{aligned}
f(nk/2) &= \frac{2e^{\frac{(k-1)nkp}{2k}}}{nk} \\
&= \frac{2}{k} e^{(-\ln n + (r-1)\ln\ln n + x)/2}.
\end{aligned}
$$

Since $\ln\ln n = o(\ln n)$, we have that $f(nk/2) = o(1)$. Moreover, $1 < f(1) = e^{\frac{(k-1)p}{k}} < e$ and thus for sufficiently large n we must have $f(m) \leq f(1) < e$. Therefore,

$$
P_m \leq R(c_4 e^{1-x})^m.
$$

Let P_0 be the probability that there exists a set S with size $nk/2$ or less that it is not r-reachable. Then by the union bound we have

$$
P_0 \leq \sum_{m=1}^{\lfloor nk/2 \rfloor} P_m \leq \sum_{m=1}^{\infty} R(c_4 e^{1-x})^m = \frac{Rc_4 e^{1-x}}{1 - c_4 e^{1-x}} = o(1),
$$

since $x \to \infty$ as $n \to \infty$. Thus $i(G_p) > r - 1$ a.a.s. This also implies that G_p has minimum degree at least r a.a.s. (by the relationships shown in Fig. 5.1b).

To complete the proof, we have to show that for $p = \frac{\ln n+(r-1)\ln\ln n-x}{(k-1)n}$ where $x = o(\ln\ln n)$ and $x \to \infty$ when $n \to \infty$, $i(G_p) \leq r-1$ a.a.s. In order to prove this, we show the stronger result that G_p has a node with degree less than or equal to $r-1$.

Consider the vertex set $V_1 = \{v_1, \ldots, v_n\}$, and define the random variable $X = X_1 + X_2 + \cdots + X_n$ where $X_i = 1$ if the degree of node v_i is less than r and zero otherwise. The goal is to show that if $p(n) = \frac{\ln n+(r-1)\ln\ln n-x}{(k-1)n}$, then $\mathbf{Pr}(X = 0) \to 0$ asymptotically almost surely. This means that for the specified $p(n)$, there exists a node in V_1 with degree less than r with high probability. Thus by selecting that single node as the set S, we have $i(G_p) \leq \frac{|\partial S|}{|S|} = r-1$ which is exactly what we want to show.

The random variable X is zero if and only if $X_i = 0$ for $1 \leq i \leq n$. The random variables X_i and X_j are identically distributed and independent when $i \neq j$ and thus we have

$$\begin{aligned}\mathbf{Pr}(X = 0) &= \mathbf{Pr}(X_1 = 0)^n \\ &= (1 - \mathbf{Pr}(X_1 = 1))^n \\ &\leq e^{-n\mathbf{Pr}(X_1=1)},\end{aligned} \tag{5.12}$$

where the last inequality is due to the fact that $1 - p \leq e^{-p}$ for $p \geq 0$. Now, note that

$$\begin{aligned}n\mathbf{Pr}(X_1 = 1) &= n\sum_{i=0}^{r-1}\binom{n(k-1)}{i}p^i(1-p)^{n(k-1)-i} \\ &\geq n\binom{n(k-1)}{r-1}p^{r-1}(1-p)^{n(k-1)-r+1} \\ &\geq n\binom{n(k-1)}{r-1}p^{r-1}(1-p)^{n(k-1)},\end{aligned} \tag{5.13}$$

where the last inequality is obtained by using the fact that $0 < (1-p)^{r-1} \leq 1$ for $r \geq 1$. Using the fact that $\binom{n(k-1)}{r-1} = \Omega\left(n^{r-1}\right)$ for constant r and k, and $(1-p)^{n(k-1)} = e^{n(k-1)\ln(1-p)} = \Omega(e^{-n(k-1)p})$ when $np^2 \to 0$ (which is satisfied for the function p that we are considering above), the inequality (5.13) becomes

$$n\mathbf{Pr}(X_1 = 1) = \Omega\left(n^r p^{r-1} e^{-n(k-1)p}\right).$$

Substituting $p = \frac{\ln n+(r-1)\ln\ln n-x}{(k-1)n}$ and simplifying, we obtain

$$n\mathbf{Pr}(X_1 = 1) = \Omega\left(\frac{(\ln n + (r-1)\ln\ln n - x)^{r-1}}{(\ln n)^{r-1}}e^x\right) = \Omega(e^x).$$

Thus we must have that $\lim_{n\to\infty} n\mathbf{Pr}(X_1 = 1) = \infty$, which proves that $\mathbf{Pr}(X = 0) \to 0$ as $n \to \infty$ (from (5.12)). Therefore, there exists a node with degree less than r a.a.s. and consequently, $i(G_p) \leq r - 1$ a.a.s. $\square$

The above result shows that the function $t(n) = \frac{\ln n + (r-1)\ln\ln n}{(k-1)n}$ forms a (sharp) threshold for the property $i(G_p) > r - 1$, and also for the graph to have minimum degree r (a significantly weaker property)[1]. Also note that the restriction $x(n) = o(\ln\ln n)$ in the above lemma is for technical reasons; the result will hold even if $x(n)$ grows faster than this bound due to the fact that $i(G_p) > r - 1$ and minimum degree being at least r are both *monotonic properties* (i.e., they hold if more edges are added to the graph) [31].

Lemma 17 provides the condition under which the isoperimetric constant is higher or lower than $r - 1$ (a constant value). Next, we will investigate a coarser rate of growth for the inter-network edge formation p, and show that for such probability functions, the isoperimetric constant scales as $\Theta(np)$. This will play a role in Sections 5.4 and 5.6, where we investigate algebraic connectivity and robustness of random interdependent networks.

Lemma 18. *Consider a random k-partite graph $G_p = (V_1 \cup V_2 \cup \cdots \cup V_k, E_p)$ with node sets $V_i = \{(i-1)n+1, (i-1)n+2, \ldots, in\}$ for $1 \leq i \leq k$. Assume that the inter-network edge formation probability p satisfies $\limsup_{n\to\infty} \frac{\ln n}{(k-1)np} < 1$. Fix any $\epsilon \in (0, \frac{1}{2}]$. Then there exists a constant α (that depends on p) such that the minimum degree d_{min}, maximum degree d_{max} and isoperimetric constant $i(G_p)$ a.a.s. satisfy*

$$\alpha np \leq i(G_p) \leq d_{min} \leq d_{max} \leq n(k-1)p\left(1 + \sqrt{3}\left(\frac{\ln n}{(k-1)np}\right)^{\frac{1}{2}-\epsilon}\right). \tag{5.14}$$

Proof. The inequality $i(G_p) \leq d_{min}$ follows immediately from the definition of the isoperimetric constant. We will show that $d_{max} \leq n(k-1)p\left(1 + \sqrt{3}(\frac{\ln n}{(k-1)np})^{\frac{1}{2}-\epsilon}\right)$ asymptotically almost surely. Let d_j denote the degree of vertex j, $1 \leq j \leq kn$. From the definition, d_j is a binomial random variable with parameters $n(k-1)$ and p and thus $\mathbb{E}[d_j] = n(k-1)p$. Then, for any $0 < \beta \leq \sqrt{3}$, by the Chernoff bound [67, 79]

$$\mathbf{Pr}(d_j \geq (1+\beta)\mathbb{E}[d_j]) \leq e^{-\frac{\mathbb{E}[d_j]\beta^2}{3}}. \tag{5.15}$$

[1] Loosely speaking, if $p(n)$ is "bigger" than $t(n)$ (in the sense specified by the lemma), then the stated properties a.a.s. hold and if $p(n)$ is "less" than $t(n)$, the properties a.a.s. do not hold.

Choose $\beta = \sqrt{3}(\frac{\ln n}{(k-1)np})^{\frac{1}{2}-\epsilon}$, which is at most $\sqrt{3}$ for n sufficiently large and p satisfying the conditions in the lemma. Substituting into equation (5.15), we have

$$\mathbf{Pr}(d_j \geq (1+\beta)\mathbb{E}[d_j]) \leq e^{-(\ln n)(\frac{\ln n}{(k-1)np})^{-2\epsilon}}.$$

The probability that d_{max} is higher than $(1+\beta)\mathbb{E}[d_j]$ equals the probability that at least one of the vertices has degree higher than $(1+\beta)\mathbb{E}[d_j]$, which by the union bound is upper bounded by

$$\begin{aligned}\mathbf{Pr}(d_{max} \geq (1+\beta)\mathbb{E}[d_j]) &\leq kn\mathbf{Pr}(d_j \geq (1+\beta)\mathbb{E}[d_j]) \\ &\leq ke^{(\ln n)-(\ln n)(\frac{\ln n}{(k-1)np})^{-2\epsilon}} \\ &\leq ke^{(\ln n)(1-(\frac{\ln n}{(k-1)np})^{-2\epsilon})}.\end{aligned}$$

Since the right hand-side of the above inequality goes to zero as $n \to \infty$ for p satisfying the condition in the lemma, we conclude that

$$d_{max} \leq n(k-1)p\left(1+\sqrt{3}\left(\frac{\ln n}{(k-1)np}\right)^{\frac{1}{2}-\epsilon}\right),$$

asymptotically almost surely.

Next, we prove the lower-bound for $i(G_p)$ in (5.14). We show that for any set of vertices of size m, $1 \leq m \leq nk/2$, there are at least αmnp edges that leave the set, for some constant α that we will specify later and probability p satisfying $\lim_{n\to\infty} \frac{\ln n}{(k-1)np} < 1$.

Consider a set $S \subset V_1 \cup V_2 \cup \cdots \cup V_k$ with $|S| = m$. Assume that the set S contains s_i nodes from V_i for $1 \leq i \leq k$ (i.e., $|S \cap V_i| = s_i \geq 0$). Define E_S as the event that αmnp or fewer edges leave S. Note that $|\partial S|$ is a binomial random variable with parameters $\sum_{l=1}^{k} s_l \left(\sum_{t=1,t\neq l}^{k}(n-s_t)\right)$ and p. As in the equality (5.6) and inequalities (5.7) and (5.9), we have that

$$\mathbf{Pr}(E_S) \leq \sum_{i=0}^{\lfloor \alpha mnp \rfloor} \binom{n(k-1)m}{i} p^i (1-p)^{n(k-1)m-\frac{(k-1)m^2}{k}-i}. \tag{5.16}$$

Next note that $k \geq 2$ and for $1 \leq i \leq \lfloor \alpha mnp \rfloor$,

$$\begin{aligned}\frac{\binom{n(k-1)m}{i}p^i(1-p)^{n(k-1)m-\frac{(k-1)m^2}{k}-i}}{\binom{n(k-1)m}{i-1}p^{i-1}(1-p)^{n(k-1)m-\frac{(k-1)m^2}{k}-(i-1)}} &= \frac{n(k-1)m-i+1}{i} \times \frac{p}{1-p} \\ &\geq \frac{n(k-1)m-\alpha mnp+1}{\alpha mnp} \times \frac{p}{1-p} \\ &\geq \frac{k-1-\alpha p}{\alpha} \times \frac{1}{1-p} \geq \frac{1-\alpha p}{\alpha(1-p)} \geq \frac{1}{\alpha},\end{aligned}$$

for $\alpha < 1$ which will be satisfied by our eventual choice for α.

Now let P_m denote the probability of the event that there exists a set of size m with $\lfloor \alpha mnp \rfloor$ or fewer number of edges leaving it. Then there must exist some constant $R > 0$ such that by the same procedure as in inequalities (5.10) and (5.11), we have

$$\begin{aligned} P_m &\leq R\left(\frac{nke}{m}\right)^m \left(\frac{n(k-1)mep}{\alpha mnp}\right)^{\alpha mnp} (1-p)^{n(k-1)m-\alpha mnp-\frac{(k-1)m^2}{k}} \\ &\leq R\left(\frac{(k-1)e}{\alpha}\right)^{\alpha mnp} e^{m\ln\left(\frac{nke}{m}\right)} e^{-p(n(k-1)m-\alpha mnp-\frac{(k-1)m^2}{k})} \\ &= Re^{mh(m)}, \end{aligned} \tag{5.17}$$

where

$$\begin{aligned} h(m) &= \alpha np + \alpha np\ln(k-1) - \alpha np\ln\alpha + \ln\left(\frac{nke}{m}\right) - p(n(k-1) - \alpha np - \frac{(k-1)m}{k}) \\ &= 1 + \ln k + \frac{(k-1)pm}{k} - \ln m + np\left(\underbrace{\alpha + \alpha\ln(k-1) - \alpha\ln\alpha + \frac{\ln n}{np} - (k-1) + \alpha p}_{\Gamma(\alpha)}\right). \end{aligned} \tag{5.18}$$

Since $\frac{\partial h(m)}{\partial m} = \frac{(k-1)p}{k} - \frac{1}{m}$ is negative for $m < \frac{k}{(k-1)p}$ and positive for $m > \frac{k}{(k-1)p}$, we have

$$\begin{aligned} h(m) &\leq \max\{h(1), h(nk/2)\} \\ &\leq \max\left\{1 + \ln k + \frac{(k-1)p}{k} + np\Gamma(\alpha), 1 + \ln 2 + np\left(\Gamma(\alpha) + \frac{(k-1)}{2} - \frac{\ln n}{np}\right)\right\}. \end{aligned}$$

From (5.18), $\frac{\partial\Gamma(\alpha)}{\partial\alpha} = \ln(k-1) - \ln\alpha + p > 0$ and thus $\Gamma(\alpha)$ is an increasing function in α for $\alpha < (k-1)$, with $\Gamma(0) = \frac{\ln n}{np} - (k-1)$ which is negative and bounded away from 0 for sufficiently large n (by the assumption on p in the lemma). Therefore, for sufficiently small α, there exists some positive constant $\bar{\alpha}$ such that $h(m) \leq -\bar{\alpha}np$ for sufficiently large n. Thus (5.17) becomes $P_m \leq Re^{-\bar{\alpha}mnp}$ for sufficiently large n.

The probability that $i(G_p) \leq \alpha np$ is upper bounded by the sum of the probabilities P_m for

$1 \leq m \leq \lfloor nk/2 \rfloor$. Using the above inequality, we have

$$\begin{aligned}\mathbf{Pr}(i(G_p) \leq \alpha np) \leq \sum_{m=1}^{\lfloor nk/2 \rfloor} P_m &\leq R \sum_{m=1}^{\lfloor nk/2 \rfloor} e^{-\bar{\alpha} mnp} \\ &\leq R \sum_{m=1}^{\infty} e^{-\bar{\alpha} mnp} \\ &= R \frac{e^{-\bar{\alpha} np}}{1 - e^{-\bar{\alpha} np}},\end{aligned}$$

which goes to 0 as $n \to \infty$. Therefore, we have $i(G_p) \geq \alpha np$ asymptotically almost surely. $\square$

So far in this section, we have been focused on random k-partite graphs. In the next result, we provide a bound for the isoperimetric constant of random interdependent graphs (with arbitrary topologies within the subnetworks).

Lemma 19. *Let $G = (G_1, G_2, \ldots, G_k, G_p)$ be a random interdependent network and assume that the probability of inter-network edge formation p satisfies $\limsup_{n\to\infty} \frac{\ln n}{(k-1)np} < 1$. Then $i(G) = \Theta(np)$.*

Proof. First we show that $i(G) \leq \gamma np$ for some $\gamma > 0$ a.a.s. Consider the set of nodes V_1 in the first subnetwork G_1. The number of edges between V_1 and all other V_j, $2 \leq j \leq k$ is a binomial random variable $B(n^2(k-1), p)$ and thus $\mathbb{E}[|\partial V_1|] = n^2(k-1)p$. Using the Chernoff bound [67] for the random variable $|\partial V_1|$, we have (for $0 < \delta < 1$)

$$\mathbf{Pr}(|\partial V_1| \geq (1+\delta)\mathbb{E}[|\partial V_1|]) \leq e^{\frac{-\mathbb{E}(|\partial V_1|)\delta^2}{3}}. \tag{5.19}$$

Choosing $\delta = \frac{\sqrt{3}}{\sqrt{\ln n}}$, the upper bound in the expression above becomes $\exp\left(-\frac{n^2(k-1)p}{\ln n}\right)$. Since $\ln n < n(k-1)p$ for n sufficiently large and for p satisfying the condition in the proposition, the right hand side of inequality (5.19) goes to zero as $n \to \infty$. Thus $|\partial V_1| \leq (1 + o(1))\mathbb{E}[|\partial V_1|]$ a.a.s. Therefore

$$\begin{aligned}i(G) = \min_{\substack{|A| \leq \frac{nk}{2}, \\ A \subseteq V_1 \cup V_2 \cup \cdots \cup V_k}} \frac{|\partial A|}{|A|} &\leq \frac{|\partial V_1|}{|V_1|} \\ &\leq \frac{(1+o(1))n^2(k-1)p}{n} \\ &\leq \gamma np,\end{aligned}$$

a.a.s. for some $\gamma > 0$.

Next, we have to show that $i(G) \geq \alpha np$ for some $\alpha > 0$. Consider the k-partite subgraph of network G which is denoted by G_p. By Lemma 18, we know that $i(G_p) \geq \alpha np$ a.a.s. Adding edges does not decrease the isoperimetric constant (by definition of the isoperimetric constant) and thus $i(G) \geq i(G_p) \geq \alpha np$ a.a.s. □

Remark 13. *The results that we developed in this section for isoperimetric constant of random interdependent and k-partite networks* ***only*** *depend on the inter-network edge formation probability p. This fact demonstrates the importance of inter-network edges and is in line with classical findings in the sociology literature [42]. For instance, in the setting introduced in Section 5.2.1, the set of inter-network edges work as bridges between different communities and transfer valuable information that cannot be obtained by the individuals otherwise. Moreover, since our results in the rest of this chapter are directly derived from Lemmas 17, 18 and 19, they are also* ***only*** *dependent on inter-network edge formation probability p and are independent of intra-network topologies. A deeper investigation of the role of the subnetwork topologies would potentially lead to further refinement of our results which is left as a venue for future work.*

In the following sections, we build on these results to study the spectral and structural properties of random interdependent networks (with corresponding implications for consensus dynamics that operate over these networks).

5.4 Algebraic Connectivity of Random Interdependent Networks

The algebraic connectivity of interdependent networks has started to receive attention in recent years. The authors of [80] analyzed the algebraic connectivity of deterministic interconnected networks with one-to-one weighted symmetric inter-network connections. The recent paper [64] studied the algebraic connectivity of a mean field model of interdependent networks where each subnetwork has an identical structure, and the interconnections are all-to-all with appropriately chosen weights. Spectral properties of random interdependent networks (under the moniker of *planted partition models*) have also been studied in research areas such as algorithms and machine learning [2, 28, 71]. Here, we leverage our results from the previous section to provide a bound on the algebraic connectivity for random interdependent networks that is the tightest known bound for the range of inter-network edge formation probabilities that we consider.

Theorem 4. *Consider a random interdependent network $G = (G_1, G_2, \ldots, G_k, G_p)$ and assume that the probability of inter-network edge formation p satisfies $\limsup_{n\to\infty} \frac{\ln n}{(k-1)np} < 1$. Then $\lambda_2(G) = \Theta(np)$ a.a.s.*

Proof. First note that by Lemma 19, there exists a constant $\gamma > 0$ such that $i(G) \leq \gamma np$ a.a.s. Hence from inequality (5.2), we have $\lambda_2(G) = O(np)$ a.a.s.

Next, we prove the lower bound on $\lambda_2(G)$. Consider the k-partite subgraph G_p of the network G. By Lemma 18 and the inequality (5.2), we know that $\lambda_2(G_p) \geq \alpha np$ for some constant α a.a.s. Since adding edges to a graph does not decrease the algebraic connectivity of that graph [13], we have $\lambda_2(G) \geq \lambda_2(G_p) \geq \alpha np$ a.a.s. □

Theorem 4 demonstrates the importance of inter-network edges on the algebraic connectivity of the overall network when $\limsup_{n\to\infty} \frac{\ln n}{(k-1)np} < 1$. This requirement on the growth rate of p cannot be reduced if one wishes to stay agnostic about the probability distributions over the topologies of the subnetworks. Indeed, by Lemma 17, if $\limsup_{n\to\infty} \frac{\ln n}{(k-1)np} > 1$, a random k-partite graph will have at least one isolated node a.a.s. and thus has algebraic connectivity equal to zero a.a.s. In this case the quantity $\frac{\ln n}{(k-1)n}$ forms a *coarse threshold* for the algebraic connectivity being 0, or growing as $\Theta(np)$. On the other hand, if one had further information about the probability distributions over the subnetworks, one could potentially relax the condition on p required in the above results. For instance, as mentioned in Section 5.2, when each of the k subnetworks is an Erdos-Renyi graph formed with probability p, then the entire interdependent network is an Erdos-Renyi graph on kn nodes; in this case, the algebraic connectivity is $\Omega(np)$ a.a.s. as long as $\limsup_{n\to\infty} \frac{\ln n}{knp} < 1$ [79]. This constraint on p differs by a factor of $\frac{k}{k-1}$ from the expression in Theorem 4.

5.5 Smallest Eigenvalue of the Grounded Laplacian in Random Interdependent Networks

We now turn our attention to the smallest eigenvalue of the grounded Laplacian matrix, obtained by removing certain rows and columns from the Laplacian. Specifically, we consider the case where all of the rows and columns corresponding to the nodes in one of the subnetworks are removed as described in Section 5.1.3. This represents the situation where all of the nodes in the grounded subnetwork act as leaders in consensus dynamics, while the rest of the nodes are followers.

Theorem 5. *Consider a random interdependent network $G = (G_1, G_2, \dots, G_k, G_p)$ with k subnetworks $G_i = (V_i, E_i), 1 \leq i \leq k$ where $|V_i| = n$ for $1 \leq i \leq k$ and $G_p = (\cup_{i=1}^{k} V_i, E_p)$. Suppose that one of the subnetworks consists only of leader nodes and the rest of the nodes in the network are followers. Assume that the probability of edge formation between a follower and leader node is denoted by p and satisfies $\limsup_{n\to\infty} \frac{\ln n}{np} < 1$. Then the smallest eigenvalue of the grounded Laplacian satisfies $\lambda = \Theta(np)$.*

In order to prove this theorem, we use a simplified version of Theorem 1 in [78] which is stated below.

Lemma 20 ([78])**.** *Consider a graph $G = (V, E)$ and suppose $S \subset V$ is a set of grounded nodes. For $v_i \in V \setminus S$, let β_i be the number of grounded nodes in node v_i's neighborhood and λ denote the smallest eigenvalue of the grounded Laplacian L_g. Then*

$$\min_{i\in\mathcal{V}\setminus S} \beta_i \leq \lambda \leq \max_{i\in\mathcal{V}\setminus S} \beta_i. \tag{5.20}$$

Proof of Theorem 5. Without loss of generality assume that V_1 consists entirely of leader nodes and all other nodes are followers. For any node $v_j \in V_i$ where $2 \leq i \leq k$, let β_j^i be the number of neighbors of the node v_j in the set V_1. Consider the inter-network topology $H_i = (V_i \cup V_1, E_H^i)$ between nodes in V_i and V_1, i.e, $E_H^i = E_p \cap (V_i \times V_1)$. It is clear that H_i is a bipartite network and thus β_j^i is the degree of the nodes in V_i in the network H_i. Therefore, by Lemma 18 (with $k = 2$), for the specified range of p, there exist α and γ such that

$$\alpha np \leq d_{min}(H_i) \leq \min_{1\leq j\leq n} \beta_i^j$$
$$\max_{1\leq j\leq n} \beta_i^j \leq d_{\max}(H_i) \leq \gamma np,$$

a.a.s. The above inequalities hold for all $2 \leq i \leq k$ a.a.s. Since k is a constant value, we conclude that there must exist $\alpha', \gamma' > 0$ such that $\alpha' np \leq \min_{i\in\mathcal{V}\setminus S} \beta_i$ and $\max_{i\in\mathcal{V}\setminus S} \beta_i \leq \gamma' np$ a.a.s and thus by Lemma 20, $\lambda = \Theta(np)$ a.a.s. □

Remark 14. *The analysis of grounded Laplacian matrices for scenarios where the leaders are spread across multiple subnetworks is more challenging. The existing analytical bounds in the literature for general grounded Laplacians require more information about the network topology [79], and thus further research is required to obtain bounds for random interdependent networks with arbitrary subnetwork topologies and general leader sets.*

5.6 Robustness of Random Interdependent Networks

Finally in this section, we characterize the conditions under which random interdependent networks are r-robust. We will first consider random k-partite networks, and show that they exhibit phase transitions at certain thresholds for the probability p.

Theorem 6. *For any positive integers r and $k \geq 2$,*

$$t(n) = \frac{\ln n + (r-1)\ln\ln n}{(k-1)n}$$

is a threshold for r-robustness of random k-partite graphs.

Proof. Consider a random k-partite graph G_p with edge formation probability $p(n)$ given by

$$p(n) = \frac{\ln n + (r-1)\ln\ln n + x}{(k-1)n},$$

where $r \in \mathbb{N}$ is a constant and $x = x(n)$ is some function satisfying $x = o(\ln\ln n)$ and $x \to \infty$ as $n \to \infty$. By Lemma 17, we know that $i(G_p) > r-1$. Therefore, by Lemma 16, G_p is at least r-robust a.a.s.

Next consider $p(n) = \frac{\ln n + (r-1)\ln\ln n - x}{(k-1)n}$, where $x = x(n)$ is some function satisfying $x = o(\ln\ln n)$ and $x \to \infty$ as $n \to \infty$. Lemma 17 indicates that the minimum degree of a random k-partite graph G_p is less than r a.a.s. Hence, G_p is not r-robust a.a.s (by the relationships shown in Fig. 5.1b). $\square$

Together with Lemma 17, the above result indicates that the properties of r-robustness and r-minimum-degree (and correspondingly, r-connectivity) all share the same threshold function in random k-partite graphs, despite the fact that r-robustness is a significantly stronger property than the other two properties. In particular, this indicates that above the given threshold, random k-partite networks possess stronger robustness properties than simply being r-connected: they can withstand the removal of a large number of nodes (up to $r-1$ from every neighborhood), and facilitate consensus dynamics that are resilient to a large number of malicious nodes (up to $\lfloor \frac{r-1}{2} \rfloor$ in the neighborhood of every normal node).

With the sharp threshold given by Theorem 6 for random k-partite graphs in hand, we now consider general random interdependent networks with arbitrary topologies within the subnetworks. Note that any general random interdependent network can be obtained by first drawing a random k-partite graph, and then adding additional edges to fill out the subnetworks. Using the fact that r-robustness is a *monotonic graph property* (i.e., adding edges to an r-robust graph does not decrease the robustness parameter), we obtain the following result.

Corollary 4. *Consider a random interdependent network $G = (G_1, G_2, \ldots, G_k, G_p)$. Assume that the inter-network edge formation probability satisfies $p(n) \geq \frac{\ln n + (r-1)\ln\ln n + x}{(k-1)n}$, $r \in \mathbb{Z}_{\geq 1}$ and $x = x(n)$ is some function satisfying $x = o(\ln\ln n)$ and $x \to \infty$ as $n \to \infty$. Then G is r-robust a.a.s.*

We conclude this section by characterizing the robustness of random interdependent networks under a coarser rate of growth in p.

Theorem 7. *Consider a random k-partite graph G_p with inter-network edge formation probability $p = p(n)$ that satisfies $\limsup_{n\to\infty} \frac{\ln n}{(k-1)np} < 1$. Then G_p is $\Theta(np)$-robust a.a.s.*

Proof. For $p = p(n)$ satisfying the given condition, we have $i(G_p) = \Theta(np)$ a.a.s. from Lemma 18. By Lemma 16, the robustness parameter of G_p is $\Omega(np)$ a.a.s. Furthermore, since the robustness parameter is always less than the minimum degree of the graph, the robustness parameter is $O(np)$ a.a.s. from Lemma 18. □

Once again, since adding edges to a network does not decrease the robustness parameter, the above result immediately implies that for random interdependent networks with inter-network edge formation probability satisfying $\limsup_{n\to\infty} \frac{\ln n}{(k-1)np} < 1$, the robustness is $\Omega(np)$ a.a.s.

5.7 Summary

We studied certain spectral and structural properties of random interdependent networks. We started by analyzing the isoperimetric constant of random k-partite graphs, and showed that the properties $i(G) > r - 1$ and r-minimum-degree share the same threshold function. We also provided a range for p, the probability of inter-network edge formation, for which the isoperimetric constant grows as $\Theta(np)$. We exploited these results to investigate three important characteristics of random interdependent networks, namely the algebraic connectivity, the smallest eigenvalue of the grounded Laplacian matrix and r-robustness. We determined tight asymptotic rates of growth on the algebraic connectivity of random interdependent networks for certain ranges of inter-network edge formation probabilities (regardless of the subnetwork topologies). Next, we analyzed the condition where nodes in one of the subnetworks act as leaders and all of the other nodes are followers. We specified a growth rate for the smallest eigenvalue of the grounded Laplacian matrix for a range of inter-network edge formation probabilities (again, regardless of the interaction topology among the follower nodes). Finally, we showed that r-robustness and r-minimum-degree (and r-connectivity) all share the same threshold function, despite the fact that r-robustness is a much stronger property than the others. Our results lead to insights about consensus and opinion dynamics that operate over random interdependent networks.

Chapter 6

Conclusion and Future Work

6.1 Conclusion

The objective of this thesis was to study the structure of multi-layer and interdependent networks. One approach to investigate this problem was through the framework of *random networks,* where each network is drawn from a certain probability distribution. An alternative perspective on understanding the structure of networks was to view the edges as being optimally/strategically placed (either by a central designer, or by different decision makers) in order to maximize some given utility function(s). Each of these methods has its own weaknesses and strengths, and provide different (complementary) perspectives on the structure of networks. While random models exhibit certain interesting features, they do not explain why those processes might arise. This is in contrast with the strategic models, where the reason behind a specific characteristic can be traced back to primitive elements such as the cost of constructing edges and the form of the utility function. However, strategic network formation models have two main drawbacks: the requirement to model network designer incentives in the utility function and prediction of the structure of the emerging networks. We obtained the following key results:

1. We generalized distance-based network formation to multi-layer networks, and showed that the problem of finding an optimal network in this setting is NP-hard. By characterizing some particular properties of optimal networks, we found the optimal networks for certain special cases of reference graphs. Extending the concept of pairwise stability to multi-layer networks revealed that the optimally designed networks are not necessarily stable.

2. We then focused on the case where multiple network designers are constructing their networks simultaneously. In introducing a game-theoretic framework to model this situation,

we assumed that the utility of each network designer (player) depends on the structure of its own network as well as the structure of the network designed by other players. We started the analysis of this game by considering a distance-based utility with strategic substitute for each player. This utility function has two important properties: first, the existence of a link in the network of one player makes it less desirable for that link to appear in the network of other players; and second, in this setting, each player can have its own specific cost and benefit function. By construction, we proved that this game always has a Nash equilibrium. Furthermore, our result indicates that hub-and-spoke networks commonly observed in transportation systems arise as a Nash equilibrium, and that the presence of low-cost players pushes high-cost players out of the game.

3. We also applied our multi-layer network formation game to a setting where each link can appear in the network of only one of the players. This led to a 2-players game based on the classical Colonel Blotto game where each player (as a network designer) has a limited amount of resources to invest on the edges, and the player with a higher amount of investment on one edge wins that edge. We characterized a Nash equilibrium of this game when the utility of each player is a function of the diameter or the largest connected component of the outcome network.

4. We then turned our attention to coupled interdependent networks. We defined a network design game for optimally allocating the interconnection links between nodes of two networks $G_1 = (V_1, E_1)$ and $G_2 = (V_2, E_2)$, where $V_1 \neq V_2$. There are $|V_1|$ players in this game, each associated with a node in the set V_1. Each node in V_1 has interdependencies with a subset of nodes in V_2. Then the objective of each player was defined to build a set of edges from its associated node to nodes in V_2 such that its distances to the set of nodes it depends on in G_2 are minimized. We showed that determining a best response action of a player in this game is NP-hard; however, certain insights can be gained about the structure of the optimal actions. We proved existence of a pure Nash equilibria in this game under certain conditions by providing an algorithm that outputs such an equilibrium of the game for any set of players.

5. Finally, we focused on studying structural and spectral properties of interdependent networks via random network models. In a major contribution, we showed that "$i(G) > r - 1$", r-minimum degree, r-robustness and r-connectivity all share the same threshold function, despite the fact that these properties have different degrees of strength. We also provided a condition under which algebraic connectivity of random interdependent networks scales as $\Theta(np)$, which is independent of the topology inside the subnetworks and depends only on the inter-network edge formation probability p. Our results were applicable to random interdependent networks with an arbitrary number of subnetworks and led

to insights about consensus and opinion dynamics that operate over such networks.

6.2 Future Work

It is likely that our results on multi-layer and interdependent networks can be further extended to obtain finer insights into the structure of such networks. Below we provide some immediate avenues for further studies.

1. In Chapter 2, we showed that finding a best response with respect to an arbitrary network is NP-hard. Developing approximation algorithms is an interesting and potentially challenging avenue for future research on this problem. Our analysis revealed a relationship between the tree-t-spanner problem and the best response network design problem (i.e., there is a reduction from the former to the latter), which formed the basis of our NP-hardness proof. There is a rich literature on approximation algorithms for tree-t-spanners [14, 11, 29, 25], and thus further investigations of the connections between tree-t-spanners and the best response network problem might lead to approximation algorithms for the latter.

2. Since we introduced multi-layer network formation games for the first time in this thesis, our goal was to establish the existence and structural properties of the Nash equilibria, which we could then use as a baseline to study other classes of games (such as Stackelberg or sequential games). One avenue for further research on this problem is to study sequential best response dynamics by the different players in order to capture the players changing their networks over time in response to the networks constructed by other players. Our simulations show that such best response dynamics converge to the Nash equilibria that we have identified in this thesis, but we currently have proofs of convergence only for certain cases.

3. The utility that we considered in our strategic network formation analysis was based on the distance between nodes of the network. Furthermore, the networks were assumed to be undirected unweighted graphs. Extension of this investigation to more general cases (i.e., weighted graphs and other types of utility function) would be of value. For instance, this might lead to an analysis of how economical relations between a group of countries change as the political relationships between them change.

4. There are other possible research directions on the Colonel Blotto network formation game such as investigating robustness of the pure Nash equilibrium networks against link failures, i.e., whether removing certain number of edges from the formed network of a player

makes it disconnected or increases its diameter. Another interesting subject is exploiting properties of random networks for strategic purposes. For example as we mentioned in Remark 10, an Erdos Renyi graph and its complement have diameter 2 when the number of nodes are large enough. Thus they form a pure Nash equilibria for the Colonel Blotto network formation game with respect to diameter. Other connections that can be made between random graphs and strategic formation is an interesting potential subject of research. Studying a sequential version of the Colonel Blotto network formation game and investigating convergence to Nash equilibria would also be of value.

5. In Chapter 4, we discussed strategic design of interdependent networks with two subnetworks. Extension of this work to the case where there are more than two subnetworks would be of value. We also proved the existence of a Nash equilibria when G_2 has a star subgraph and G_I was the complete bipartite graph. Proving the existence of Nash equilibria for other classes of G_2 and G_I is an interesting subject for future study.

6. The threshold function that we obtained in Chapter 5 for the isoperimetric constant of random interdependent networks depends only on the inter-network edge formation probability. A deeper investigation of the role of the subnetwork topologies would potentially lead to further refinement of our results. Also it was assumed that the inter-network edges have identical and independent Bernoulli probability distributions. Extending the analysis to other probability distributions over the inter-network edges is an important topic for further research.

7. There are various interesting directions for research on the robustness of interdependent networks, including development of a notion for structural robustness against cascading failures. Recognizing vulnerable structures of interdependent networks (which depends on the nature of failures and propagation dynamics) might be a helpful step toward this direction. Another interesting subject is to consider a defense budget against adversaries and then to investigate optimal allocation of the defense budget on the nodes to minimize the impact of random and strategic attacks on layered networks. Finally, proposing a fault diagnosis and compensation technique to reduce the spread of cascading failures is also an interesting topic for further research on interdependent networks.

References

[1] A strategy for a future mail processing and transportation network. Technical Report RARC-WP-11-006, U.S. Postal Service Office of Inspector General, 07 2011.

[2] E. Abbe, A. S. Bandeira, and G. Hall. Exact recovery in the stochastic block model. *arXiv:1405.3267v4*, 2014.

[3] Sebastian Abshoff, Andreas Cord-Landwehr, Daniel Jung, and Alexander Skopalik. Multilevel network games. In *Web and Internet Economics*, volume 8877 of *Lecture Notes in Computer Science*, pages 435–440. 2014.

[4] E. Arcaute, R. Johari, and S. Mannor. Network formation: Bilateral contracting and myopic dynamics. *IEEE Transactions on Automatic Control*, 54(8):1765–1778, 2009.

[5] Venkatesh Bala and Sanjeev Goyal. A noncooperative model of network formation. *Econometrica*, 68(5):1181–1229, 2000.

[6] Albert-Laszlo Barabasi and Reka Albert. Emergence of scaling in random networks. *Science*, 286(5439):509–512, 1999.

[7] Marco Beccuti, Giuliana Franceschinis, Mohamed Kaâniche, and Karama Kanoun. Critical information infrastructure security. chapter Multi-level Dependability Modeling of Interdependencies between the Electricity and Information Infrastructures, pages 48–59. 2009.

[8] Francis Bloch and Matthew Jackson. The formation of networks with transfer among the players. *Journal of Economic Theory*, 133:83–110, 2007.

[9] Larry Blume, David Easley, Jon Kleinberg, Robert Kleinberg, and Eva Tardos. Network formation in the presence of contagious risk. In *Proceedings of the 12th ACM Conference on Electronic Commerce*, pages 1–10, 2011.

[10] Bela Bollobas. *Random Graphs*. Cambridge University Press, 2001.

[11] Andreas Brandstdt, Feodor F. Dragan, Hong-Oanh Le, and Van Bang Le. Tree spanners on chordal graphs: complexity and algorithms. *Theoretical Computer Science*, 310(1):329 – 354, 2004.

[12] Michael Brautbar and Michael Kearns. A clustering coefficient network formation game. In *Proceedings of the 4th International Conference on Algorithmic Game Theory*, pages 224–235, 2011.

[13] A. E. Brouwer and W. H. Haemers. *Spectra of Graphs*. Springer, 2012.

[14] Leizhen Cai and Derek G. Corneil. Tree spanners. *SIAM J. Discret. Math.*, 8(3):359–387, 1995.

[15] Alessio Cardillo, Massimiliano Zanin, Jess Gmez-Gardees, Miguel Romance, Alejandro J. Garca del Amo, and Stefano Boccaletti. Modeling the multi-layer nature of the European air transport network: Resilience and passengers re-scheduling under random failures. *The European Physical Journal Special Topics*, 215(1):23–33, 2013.

[16] F. Chung. *Spectral Graph Theory*. American Mathematical Society, 1997.

[17] Fan Chung and Linyuan Lu. *Complex Graphs and Networks*. CBMS Regional Conference Series in Mathematics, vol. 107, 2006.

[18] A. Clark, B. Alomair, L. Bushnell, and R. Poovendran. Leader selection in multi-agent systems for smooth convergence via fast mixing. In *Decision and Control (CDC), 2012 IEEE 51st Annual Conference on*, pages 818–824, 2012.

[19] T. H. Cormen, C. E. Leiserson, R. L. Rivest, and C. Stein. *Introduction to Algorithms*. The MIT Press, third edition.

[20] S. Dasgupta, A. Kolla, and K. Koiliaris. Spectra of random graphs with planted partitions. 2013.

[21] N. M. Maia de Abreu. Old and new results on algebraic connectivity of graphs. *Linear Algebra and its Applications*, 423(1):53 – 73, 2007.

[22] Morris H. Degroot. Reaching a consensus. *Journal of the American Statistical Association*, 69(345):118–121, 1974.

[23] S. M. Dibaji and H. Ishii. Consensus of second-order multi-agent systems in the presence of locally bounded faults. *Systems and Control Letters*, 79:23–29, 2015.

[24] R. Diestel. *Graph Theory*. Springer-Verlag: Heidelberg, 2005.

[25] Feodor F. Dragan and Ekkehard Köhler. An approximation algorithm for the tree t-spanner problem on unweighted graphs via generalized chordal graphs. *Algorithmica*, 69(4):884–905, 2014.

[26] P. Erdos and A. Renyi. On random graphs. *Publicationes of Mathematicae Debrecen*, 6:290–297, 1959.

[27] Alex Fabrikant, Ankur Luthra, Elitza Maneva, Christos H. Papadimitriou, and Scott Shenker. On a network creation game. In *Proc. of ACM Symposium on Principle of Distributed Systems*, 2003.

[28] U. Feige and J. Kilian. Heuristics for semirandom graph problems. *Journal of Computer and System Science*, 63:639–671, 2001.

[29] Sndor P. Fekete and Jana Kremer. Tree spanners in planar graphs. *Discrete Applied Mathematics*, 108(12):85 – 103, 2001.

[30] M. Fiedler. Algebraic connectivity of graphs. *Czechoslovak Mathematical Journal*, 23(2):298–305, 1973.

[31] Ehud Friedgut and Gil Kalai. Every monotone graph property has a sharp threshold. *Proceedings of the American mathematical Society*, 124(10):2993–3002, 1996.

[32] Jianxi Gao, S. V. Buldyrev, S. Havlin, and H. E. Stanley. Robustness of a network formed by n interdependent networks with a one-to-one correspondence of dependent nodes. *Phys. Rev. E*, 85:066134, 2012.

[33] Jianxi Gao, Sergey V. Buldyrev, H. Eugene Stanley, and Shlomo Havlin. Networks formed from interdependent networks. *Nature Physics*, 8:40–48, 2012.

[34] M. R. Garey and David S. Johnson. *Computers and Intractability: A Guide to the Theory of NP-Completeness*. W. H. Freeman and Company, 1979.

[35] J. Ghaderi and R. Srikant. Opinion dynamics in social networks: A local interaction game with stubborn agents. *Automatica*, 50:3209 – 3215, 2014.

[36] A. V. Goldberg. Finding a maximum density subgraph. Technical Report UCB/CSD-84-171, EECS Department, University of California, Berkeley, 1984.

[37] Russell Golman and Scott E. Page. General Blotto game: Games of strategic allocative mismatch. *Public Choice*, 138(3-4):279–299, 2009.

[38] B. Golub and M. O. Jackson. Naive learning in social networks and the wisdom of crowds. *American Economic Journal: Microeconomics*, 2(1):112–49, 2010.

[39] Ragavendran Gopalakrishnan, Jason R. Marden, and Adam Wierman. An architectural view of game theoretic control. *SIGMETRICS Performance Evaluation Review*, 38(3):31–36, January 2011.

[40] Sanjeev Goyal. *Connections: An Introduction to the Economic of Networks*. Princeton University Press, 2009.

[41] Sanjeev Goyal and Adrien Vigier. Robust networks. *Working Paper*, 2011.

[42] M. S. Granovetter. The strength of weak ties. *American Journal of Sociology*, 6:1360–1380, 1973.

[43] Assane Gueye, Jean C. Walrand, and Venkat Anantharam. Design of network topology in an adversarial environment. In *Proceedings of the First International Conference on Decision and Game Theory for Security*, pages 1–20, 2010.

[44] Hans Haller and Sudipta Sarangi. Nash networks with heterogeneous links. *Mathematical Social Sciences*, 50(2):181 – 201, 2005.

[45] T. C. Hu. Optimum communication spanning tree. *Siam Journal of Computer*, 3(3), 1974.

[46] Mathew O. Jackson. *Social and Economic Networks*. Princeton University Press, 2008.

[47] Mathew O. Jackson and Alison Watts. The evolution of social and economic networks. *Journal of Economic Theory*, 106:265–295, 2002.

[48] Mathew O. Jackson and A. Wolinsky. A strategic model of social and economic networks. *Journal of Economic Theory*, 71(0108):44–74, 1996.

[49] Matthew O. Jackson and Brian W. Rogers. The economics of small worlds. *Journal of the European Economic Association*, 3:617–627, 2005.

[50] Matthew O. Jackson and Brian W. Rogers. Meeting strangers and friends of friends: How random are social networks? *American Economic Review*, 97(3):890–915, June 2007.

[51] A. Jadbabaie, J. Lin, and A. S. Morse. Coordination of groups of mobile autonomous agents using nearest neighbor rules. *IEEE Transactions on Automatic Control*, 48:988–1001, 2003.

[52] A. Jamakovic and S. Uhlig. On the relationship between the algebraic connectivity and graph's robustness to node and link failures. In *Next Generation Internet Networks, 3rd EuroNGI Conference on*, pages 96–102, 2007.

[53] D. S. Johnson, J. K. Lenstra, and A. R. H. Rinnooy Kan. The complexity of the network design problem. *Networks*, 8:279–285, 1978.

[54] Richard M. Karp. Reducibility among combinatorial problems. *Complexity of Computer Computations*, 1972.

[55] Michael Kearns, Stephen Judd, and Yevgeniy Vorobeychik. Behavioral experiments on a network formation game. In *Proceedings of the 13th ACM Conference on Electronic Commerce*, pages 690–704, 2012.

[56] D. B. Kingston and R. W. Beard. Discrete-time average consensus under switching network topologies. *In Proc. American Control Conference*, 2006.

[57] Mikko Kivela, Alexandre Arenas, Marc Barthelemy, James P. Gleeson, Yamir Moreno, and Mason A. Porter. Multilayer networks. *Journal of Complex Networks*, 2014.

[58] Elias Koutsoupias and Christos Papadimitriou. Worst-case equilibria. In *Proceedings of the 16th Annual Conference on Theoretical Aspects of Computer Science, STACS*, pages 404–413, 1999.

[59] Elias Koutsoupias and Christos Papadimitriou. Worst-case equilibria. In *Proceedings of the 16th Annual Symposium on Theoretical Aspects of Computer Science*, pages 404–413, 1999.

[60] Maciej Kurant and Patrick Thiran. Layered complex networks. *Phys. Rev. Lett.*, 96:138701, 2006.

[61] T. C. Kwok, L. C. Lau, Y. T. Lee, S. Oveis Gharan, and L. Trevisan. Improved Cheeger's Inequality: Analysis of spectral partitioning algorithms through higher order spectral gap. In *Proc. ACM Symp. on Theory of Computing*, pages 11–20, 2013.

[62] H.J. LeBlanc, Haotian Zhang, X. Koutsoukos, and S. Sundaram. Resilient asymptotic consensus in robust networks. *Selected Areas in Communications, IEEE Journal on*, 31(4):766–781, 2013.

[63] Matteo Magnani and Luca Rossi. Formation of multiple networks. In *Proc. 6th International Conference on Social Computing, Behavioral-Cultural Modeling and Prediction*, SBP'13, pages 257–264, 2013.

[64] J. Martin-Hernandez, H. Wang, P. Van Mieghem, and G. D'Agostino. Algebraic connectivity of interdependent networks. *Physica A: Statistical Mechanics and its Applications*, 404:92–105, 2014.

[65] F. McSherry. Spectral partitioning of random graphs. In *Proceedings of 42nd IEEE Symposium on Foundations of Computer Science*, pages 529–537, 2001.

[66] Nimrod Megiddo and Arie Tamir. On the complexity of locating linear facilities in the plane. *Operation Research Letters*, 1, 1982.

[67] M. Mitzenmacher and E. Upfal. *Probability and Computing*. Cambridge University Press, 2005.

[68] Michael Mitzenmacher and Justin Thaler. Peeling arguments and double hashing. In *50th Annual Allerton Conference on Communication, Control, and Computing*, pages 1118–1125, 2012.

[69] E. Moradi Shahrivar, M. Pirani, and S. Sundaram. Robustness and algebraic connectivity of random interdependent networks. In *5th IFAC Workshop on Distributed Estimation and Control of Networked Systems*, 2015.

[70] Ebrahim Moradi-Shahrivar and Shreyas Sundaram. Strategic multi-layer network formation. In *Proceedings of the IEEE Conference on Decision and Control*, 2013. To appear.

[71] R. R. Nadakuditi and M. E. J. Newman. Graph spectra and the detectability of community structure in networks. *Phys. Rev. Lett.*, 108:188701–188705, 2012.

[72] D. T. Nguyen, Y. Shen, and M. T. Thai. Detecting critical nodes in interdependent power networks for vulnerability assessment. *IEEE Transactions on Smart Grid*, 4(1):151–159, 2013.

[73] R. Olfati-Saber, J.A. Fax, and R.M. Murray. Consensus and cooperation in networked multi-agent systems. *Proceedings of the IEEE*, 95(1):215–233, 2007.

[74] E. M. Palmer. *Graphical Evolution: An Introduction to the Theory of Random Graphs*. John Wiley & Sons, Inc., 1985.

[75] E. M. Palmer. On the spanning tree packing number of a graph: a survey. *Discrete Math.*, 230(1-3):13–21, 2001.

[76] M. Parandehgheibi and E. Modiano. Robustness of interdependent networks: The case of communication networks and the power grid. In *IEEE Global Communications Conference*, pages 2164–2169, 2013.

[77] Marzieh Parandehgheibi and Eytan Modiano. Robustness of interdependent networks: The case of communication networks and the power grid. In *Global Communications Conference (GLOBECOM), IEEE*, pages 2164–2169, 2013.

[78] M. Pirani, E. Moradi Shahrivar, and S. Sundaram. Coherence and convergence rate in networked dynamical systems. In *Decision and Control (CDC), 2015 IEEE 54nd Annual Conference on*, 2015.

[79] M. Pirani and S. Sundaram. On the smallest eigenvalue of grounded Laplacian matrices. *IEEE Transactions on Automatic Control*, 2015. To appear.

[80] F. Radicchi and A. Arenas. Abrupt transition in the structural formation of interconnected networks. *Nature Physics*, 9:717–720, 2013.

[81] Brian Roberson. The Colonel Blotto game. *Economic Theory*, 29(1):1–24, 2006.

[82] Jean-Paul Rodrigue, Claude Comtois, and Brian Slack. *The Geography of Transport Systems*. Routledge, 3rd edition, 2013.

[83] W. Saad, Zhu Han, T. Basar, M. Debbah, and A. Hjorungnes. Network formation games among relay stations in next generation wireless networks. *IEEE Transactions on Communications*, 59(9):2528–2542, September 2011.

[84] C. M. Schneider, N. Yazdani, N. A. M. Arajo, S. Havlin, and H. J. Herrmann. Towards designing robust coupled networks. *Scientific Report*, 3(1969):1–7, 2013.

[85] Christian M. Schneider, Nuno A. M. Araujo, Shlomo Havlin, and Hans J. Herrmann. Toward designing robust coupled networks. *arXiv:1106.3234 [cond-mat.stat-mech]*, 2011.

[86] G. A. Schwartz, S. Amin, A. Gueye, and J. Walrand. Network design game with both reliability and security failures. In *Communication, Control, and Computing (Allerton), 2011 49th Annual Allerton Conference on*, pages 675–681, 2011.

[87] Ebrahim Moradi Shahrivar, Mohammad Pirani, and Shreyas Sundaram. Robustness and algebraic connectivity of random interdependent networks. In *5th IFAC Workshop on Distributed Estimation and Control in Networked Systems (NecSys)*, pages 252 – 257, 2015.

[88] Ebrahim Moradi Shahrivar and Shreyas Sundaram. Multi-layer network formation via a colonel blotto game. In *Signal and Information Processing (GlobalSIP), 2014 IEEE Global Conference on*, pages 838–841, 2014.

[89] Ebrahim Moradi Shahrivar and Shreyas Sundaram. The strategic formation of multi-layer networks. *IEEE Transactions on Network Science and Engineering*, 2(4):164–178, 2015.

[90] Martin Shubik and Robert James Weber. Systems defense games: Colonel blotto, command and control. *Naval Research Logistics Quarterly*, 28(2):281–287, 1981.

[91] Eva Tardos and Tom Wexler. Network formation games and the potential function method. In *Algorithmic Game Theory*, pages 487–516, 2007.

[92] N. Vaidya, L. Tseng, and G. Liang. Iterative approximate Byzantine consensus in arbitrary directed graphs. In *Proceedings of the ACM Symposium on Principles of Distributed Computing*, pages 365–374, 2012.

[93] D. Watts and S. Strogatz. Collective dynamics of small-world networks. *Nature*, 393(4):440–442, 1998.

[94] Kristen Weiss, Mark Hamann, Michael Kinney, and Helene Marsh. Knowledge exchange and policy influence in a marine resource governance network. *Global Environmental Change*, 22(1):178 – 188, 2012.

[95] D. B. West. *Introduction to Graph Theory*. Prentice-Hall Inc., Upper Saddle River, New Jersey, 2001.

[96] Yongle Wu, Beibei Wang, and K.J.R. Liu. Optimal power allocation strategy against jamming attacks using the Colonel Blotto game. In *IEEE Global Telecommunications Conference*, pages 1–5, 2009.

[97] O. Yagan, Dajun Qian, Junshan Zhang, and D. Cochran. Optimal allocation of interconnecting links in cyber-physical systems: Interdependence, cascading failures, and robustness. *IEEE Transactions on Parallel and Distributed Systems*, 23(9):1708–1720, 2012.

[98] H. Zhang, E. Fata, and S. Sundaram. A notion of robustness in complex networks. *IEEE Transactions on Control of Network Systems*, 2(3):310–320, 2015.

[99] Pengcheng Zhang, Srinivas Peeta, and Terry Friesz. Dynamic game theoretic model of multi-layer infrastructure networks. *Networks and Spatial Economics*, 5(2):147–178, 2005.

[100] Jun Zhao, O. Yagan, and V. Gligor. On the strengths of connectivity and robustness in general random intersection graphs. In *Decision and Control (CDC), 2014 IEEE 53rd Annual Conference on*, pages 3661–3668, 2014.