# Power and Network Integration: Structural and Algorithmic Analysis of Organizational Networks

Anastasia Moskvina

PhD

2016

# Power and Network Integration: Structural and Algorithmic Analysis of Organizational Networks

Anastasia Moskvina

A thesis submitted to
Auckland University of Technology
in fulfillment of the requirements for the degree of
Doctor of Philosophy(PhD)

2016

School of Engineering, Computer & Mathematical Sciences

Auckland University of Technology

# Abstract

This thesis constitutes the results of our research towards structural and algorithmic analysis of organizational networks. We study how interpersonal ties become crucial empowerment channels that shape organizational structure. We develop an organizational network model that is consistent with management studies; moreover, by incorporating both formal and informal ties into the model, we build a promising theory that is capable to explain several organizational phenomena including flattening, workplace homophily, and loss of control. Through rigorous analysis, we demonstrate that our theoretical framework can be used to reflect general properties of organizations.

Understanding how different departments and employees of an organization interact with one another leads to comprehension of how well the organization operates. Studying an organizational structure often reveals critical positions that may require additional attention. It is the organizational structure from which one may extract hidden clues about concealed communication obstacles. In this thesis, we consider organizational structures from the network perspective. We see the following problems:

(1) There is a lack of mathematical analysis on the dual-structure of formal and informal organizations.

(2) Existing formal definitions of power only deal with networks whose edges have a single interpretation of social links, while not incorporating formal roles and levels.

(3) Network evolution represents a substantial direction of the structural analysis of social networks but yet there is a lack of models suitable for joining two networks as an outcome of strategic calculations.

The aim of this thesis, therefore, is to challenge the problems by developing a mathematical model that sits at the confluence of algorithmic and structural analyses. Our

investigation unfolds in two main directions: the first covers individual power in organizations; the second lies in integrating two disjoint organizational networks.

The first focal point of this thesis is our centrality-based definition of power which is accompanied with comprehensive and deep analysis, case studies and experiments. Our power based model provides novel insights into a range of organizational properties: 1) Organizations have limited hierarchy height. 2) Flattening is closely related to changes in the power of employees. 3) Informal relations significantly impact power of individuals. 4) Leadership styles could be reflected and analyzed through understanding weights on the ties in an organizational network. 5) The model endorses a natural interpretation of the loss of managerial control.

Our second research direction concerns computational and algorithmic aspects of network integration. The integration process amounts to the fundamental question that arises in numerous social, political, and physical domains. We study the algorithmic nature of network integration, analyze the corresponding computational problems, apply a formal framework to tackle the problems and employ various heuristics that reflect natural intuition. To compare the methods, we perform thorough experimental analysis on both synthesized and real-world data.

The significance of this thesis lies in theoretical models, simulations and analysis. Our novel, structural approach to organizational analysis provides new insights, explanations and potentially predictive guidelines for organizational decision making.

# Acknowledgments

I would like to express my gratitude to Dr. Jiamou Liu, my supervisor and co-author, for encouraging my research and being a great mentor. Thank you for your guidance and profound advice.

I am very grateful to Prof. Ajit Narayanan, my supervisor, for kindness, valuable support and helpful discussions.

I would also like to thank Michael Ouředník for collaboration, interesting suggestions and assistance with implementing software tools. Working with you was a great pleasure.

This research would not have been possible without the support of the Marsden Fund of the Royal Society of New Zealand.

Finally, to my dear soulmate and loving husband, Mikhail Kokho, for pushing me forward and for constructive criticism, for inspiration and for lighting up my path during this research journey.

# Contents

# List of Figures

# List of Tables

# Attestation of Authorship

*I hereby declare that this submission is my own work and that, to the best of my knowledge and belief, it contains no material previously published or written by another person (except where explicitly defined in the acknowledgements), nor material which to a substantial extent has been submitted for the award of any other degree or diploma of a university or other institution of higher learning.*

Anastasia Moskvina

December 2016

# Chapter 1

# Introduction

The rapid progress of information technology in the last half-century greatly improved communication and productivity in corporations, facilitating them to grow into giant enterprises. Clearly, the bigger the company gets, the more incentive there is to identify hidden information inside its structure – the more sense it makes to study how decisions pass from the top levels to the bottom, how individuals interact with each other, and which are the most important positions. Chaotic growth can lead to inefficient management, and, hence, loss of money. This is why corporations are willing to pay large sums of money not only on hiring talented managers who define the firm's direction, but also on costly business intelligence that guarantees all the layers are on the same wavelength [30].

In behavioral management theory, one usually considers how different tenuous aspects – such as motivation, personality, expectation, and conflicts – define the productivity of an individual. Alternatively, a structural theory moves away from these personal traits and rather views an organization as patterned and repeated interactions among social actors within the organization [112]. In this thesis, we focus on the structural theory. More precisely, we represent an organizational structure as a network where agents are connected with each other by interpersonal relations. Using this network, one can analyze how the flow of information circulates within the organism of the firm and uncover properties invisible at first glance. Moreover, centrality measures are capable of detecting the most powerful positions in the network, and hence, provide a rigorous analysis of *power* in organizations.

Thus, the **first main goal** of this thesis is to analyze organizational structures from a network perspective and to develop a power based model that reflects general properties of organizations. This novel, structural approach to organizational analysis provides us new insights, explanations and potentially predictive guidelines for organizational decision making.

The **second main goal** of this thesis is to study computational and algorithmic aspects of *network integration* through the prism of social and behavior bases for integration of two networks, and, consequently, to propose several heuristics that reflect natural intuition.

Indeed, every social network evolves through time. So do organizational networks. When interpersonal links disappear, a network may be split into several components; this corresponds to a process of *disintegration*, which closely relates to the problem of graph partitioning and the identification of *strong ties* [84]. On the contrary, when new relations are forged, several disjoint networks may be linked together; this corresponds to a process of *integration*, which relates to the establishment of *weak ties*. Since the seminal work of Granovetter [49], there has been a major interest in social network research on the establishment of weak ties [87].

This thesis consists of theoretical, algorithmic, and experimental results that could largely be split into three main parts:

- Structural analysis and modeling of organizational networks;
- Defining individual power in organizational networks;
- Studying network integration as a form of network evolution.

This work has found its inspiration in several sources grounded in management, economics, social theory, mathematics, and computer science. The primary focus is laid on the area of social network analysis, and the main results lie within the theoretical scope. However, we build our theory in line with management and social behavior studies; moreover, we perform series of experiments to validate our models.

## 1.1 Social Network Analysis and Power in Organizations

To better understand the way a large company functions, one should learn not only how decisions are made on the top of the iceberg, but also how the flow of information circulates within the organism of the firm [94]. Studies of the inner structure of a firm, the hierarchy, can often uncover problems invisible at first glance. The reporting relation is the strongest indicator of power in a company; ranks, titles, and uniform clearly define privilege of individuals. On the other hand, a network perspective of power posits that informal social ties and communication also grant power. For example, Brass in his work [22] suggested that individual power in organizations comes from a structural perspective, which includes both formal and informal communication.

Power has been a key notion in sociology, politics, and management theory. The term is often interpreted as the capability of a person to affect the behaviors of others [73], and is closely related to notions such as influence and authority. Organizational power may come from multiple bases such as personal traits, ranks, skills as well as interpersonal ties [45]. While we acknowledge the importance of behavioral factors in power, the focus in this thesis is solely on power that arises from interpersonal ties.

The concepts introduced in the current work are rooted in the classical and notable work by Chester Barnard [11], who laid most of the foundation of the structural theory of organizations. According to Barnard, formal organizations coexist with informal organizations within the same entity. Barnard defines formal organizations as dictated by a set of rules and policies. An informal organization by Barnard is the personal contacts and interactions between workers that form into small groups; these informal groups of workers form their own organization in the larger organization. Power, thus, arises out of the amalgamation of the formal and informal organizations.

Barnard's dual-structural approach to organization studies has been revisited many times. For example, Emerson claims in [41] that power is a property of the social relations, and resides largely in the dependence between social actors in a network. Brass in [22] focused on an informal network that unifies workflow, communication and friendship relations, and displayed correlations between powerful nodes and central positions in this network. In a similar vein, Krackhardt and Hanson in [61] drew an analogy between a company and a human body: the formal structure of a company is the skeleton, while the informal structure is the central nervous system. Informal networks are more flexible and adaptive, formal structure is static. Krackhardt then related power to em cognitive accuracy of an individual surrounding network in [62]. Cross et al. in [32] adopted a computational approach and argued that even though informal networks are invisible, they are more reflective than the formal organizations.

The authors defined scenarios where social network analysis is useful to assess informal networks and facilitate effective collaboration. We also mention a number of works that show how informal networks can be used to uncover the reporting hierarchy, revealing interesting insight of an organization [42, 108].

There are two major technical ingredients in defining power in this thesis. Firstly, we follow Barnard's view that any model of an organization must contain the dual-structure of formal and informal relations. Secondly, we define power from a structural perspective based on interpersonal ties.

In [40], the authors advocates that social network analysis (SNA) helps business in knowledge management in collaboration (communities and cross-functional connections), team-building (finding individuals which are likely to be exposed to new ideas), human resources (monitor formal relationships), sales and marketing (track the adoption of new products) and strategy (interacting firms analysis). We apply SNA techniques to study organizational structures, and build our theory around network characteristics; power may be regarded as one of the instruments to reveal hidden insights into organizational functioning.

## 1.2 Integration as a Form of Network Evolution

Network evolution represents a substantial direction of the structural analysis of social networks, and the creation of interpersonal ties is one of the components of this process. While strong ties emerge between individuals with similar social circles, forming a basis of trust and hence community structure, weak ties link two members who share few common contacts. The influential work of Granovetter reveals the vital roles of weak ties: it is weak ties that enable information transfer between communities and provide individuals positional advantage and hence influence and power [49].

Natural questions arise regarding the establishment of weak ties between communities: How to merge two departments in an organization into one? How does a company establish trade with an existing market? How to create a transport map from existing routes? We refer to such questions as *network integration*. The basic setup involves two networks; the goal is to establish ties between them to achieve certain desirable properties in the combined network. A real-life example of network integration is the inter-marriages between members of the Medici, the leading family of Renaissance

Florence, and numerous other noble Florentine families, towards gaining power and control over the city [54]. Another example is by Paul Revere, a prominent Patriot during the American Revolution, who strategically created social ties to raise a militia [109].

In this thesis, we study network integration, the process of establishing new links between two networks. The problem of how and why new ties appear in a network has been a fundamental question in the study of complex networks. While links often emerge due to natural network evolution, there are many scenarios where links are created "by design", i.e. connections are set up in order to meet certain targets. As opposed to the field of *dynamic network models*, which focuses on network evolution, we assume that in the course of network integration ties are established not as a physical process but an outcome of strategic calculations. This is similar to the assumption of *agent-based models* which employs game theory to explain how a network arrives at a particular structure. The vital difference is that the network integration process involves two already established networks and the outcome of this process guarantees preserving topology within each network.

An example of network integration could be a merger between two companies. As discussed by the authors of [1], the success of mergers and acquisitions of companies often hinges on whether firms can socialize employees effectively into the merged new entity. Therefore a big challenge faced by the top managers of both companies is how to establish links between the two companies to ensure coherence and efficient communication.

Integrating two arbitrary organizational networks can also be motivated from two perspectives: On the one hand, the resulting network should provide both organizations with efficient channels for *collaboration*: bringing all individuals (or a particular set of persons) in one organization closer to individuals in the other would be a beneficial outcome of the collaboration. On the other hand, acquisition suggests that in the result of integration, individuals in the acquiring company should be within reach of the top manager. Such form of integration is called *dominant integration*. In this thesis, we consider integration of arbitrary homogeneous networks, and then extend the results to collaborative and dominant integration of organizational networks.

Inspired by the examples of the Medici and Paul Revere, we consider also a more restricted scenario of network integration: indeed, one if the networks may contain only a single node, and the goal is to establish this node in the other network. We motivate this setup from two directions: 1) This setup amounts to the problem of *socialization*: the situation when a newcomer joins a network as an organizational member. A natural question for the newcomer is the following: How should I forge new relationships in order to take an advantageous position in the organization? As indicated in [79], socialization is greatly influenced by the social relations formed by the newcomer with "insiders" of the network. 2) This setup also amounts to the problem of *network expansion*. For example, an airline expands its existing route map with a new destination, while trying to ensure a small number of legs between any cities.

## 1.3 Measures to Evaluate the Effect of Integration

Forging new links between two disjoint networks brings these networks together. Several questions naturally arise concerning such processes: How should two departments in an organization merge into a single unit? How to establish effective collaboration between two research teams? How does socialization between two groups of people occur? How to bridge existing bus routes to create a unified public transport map? How to create hyperlinks connecting two web domains to allow convenient browsing?

Motivated by the questions above, we address the algorithmic nature of network integration. The problem asks to build links between members of two networks so that the combined network becomes a unified whole. It is then a major question how "together" the unified networks should be as an outcome of this process. Naturally, the more links there are that connect the two networks, the closer they become. On the other hand, there is normally a cost associated with establishing and maintaining links. Therefore, it is important to strike a balance between the amount of *togetherness* and the number of new links created between the networks.

In recent years there has been a surge of the use of "togetherness" in sociology [75], communication studies [6], politics [101] and biology [31]. In its most original form, togetherness is a concept rooted in Kantian philosophy, meaning the confluence of intuition and concepts [51]. In mathematics, togetherness is regarded as a "mark of being integrated into a single unity" and influences the creations of notions such as continuity and connectedness [72]. The notion is first discussed in information science by cybernetic pioneer Gordon Pask in his 1980 essay [89]; Pask refers togetherness as an "index of human proximity" that is "determined by a communication/computation medium". He goes on to discuss how togetherness can be "engineered" through a process of "conversation", which is abstractly represented as the integration of two concept networks. We rediscover and follow this seminal work, and provide a formal interpretation of togetherness in the context of network science.

In this thesis, we consider different levels of togetherness. Already in the work of Pask, it is mentioned that an appropriate measure of togetherness comes from the notion of *distance*. In a network, the distance between two nodes is the smallest number of "hops" needed to move from one node to the other. It is natural to adopt distance as an indication of togetherness. In particular, *diameter* refers to the largest distance between nodes in the network. It is well-known that most real-world networks enjoy small diameters – this is the so-called *small world* property. We hold the view that all nodes of a network have certain resources; and when a network has a small diameter, the resource on each node can be reached out from everyone else within a few steps, and each member is able to influence others. Hence, the diameter of the integrated network forms the strongest form of togetherness.

When expressing togetherness between two networks in their integration, diameter may be too strong. We define further two weaker notions of togetherness. Firstly, *existential togetherness* considers distances between every node in one network to some node in the other network. Secondly, *universal togetherness* considers distances between every node in one network to all nodes in the other network. The former measure of togetherness may be reasonable if we assume all nodes in any network hold the same resource, and it is enough to reach any node in a network. The latter measure of togetherness may be reasonable if the distances to all nodes in the other network are important. In this thesis, we relate and compare these three notions of togetherness.

Finally, we note that formal ties reflect the reporting structure of an organization, building the managerial hierarchy. As most of the social networks have bounded diameter (small-world properties), any organizational network has bounded hierarchical height. Moreover, the diameter may not be important at all in organizational networks: indeed, a manager does not want to be too far from his or her subordinates, but subordinates may not necessarily be close to each other. Some of the departments (subnetworks) may require more control from the top manager (the root). Thus, to commend the traits of organizational networks and to define another level of togetherness, we consider the distances from any node in the subnetwork to the root, calling it *hierarchical togetherness*.

## 1.4 Related Works

This thesis is predated by *organizational behavioral studies* [99, 52, 79], which look at how social ties affect a newcomer's integration and assimilation to the organization. The authors in [34, 109] argue that brokers – those who bridge and connect to diverse groups of individuals – enable good network building; creating ties with and even becoming a broker oneself allows a person to gain private information, wide skill set and hence power. *Network building theory* has also been applied to various other contexts such as economics (strategic alliance of companies) [102], governance (forming inter-government contracts) [5], and politics (individuals' joining of political movements) [90]. Compared to these works, the novelty of this thesis lies in proposing a formal framework of network building, which employs techniques from complexity theory and algorithmics.

This work is also related to two forms of network formation: *dynamic models* and *agent-based models*, both aim to capture the natural emergence of social structures [54]. The former originates from random graphs, viewing the emergence of ties as a stochastic process which may or may not lead to an optimal structure [37]. The latter comes from economics, treating a network as a multiagent system where utility-maximizing nodes establish ties in a competitive setting [60, 53]. Our work differs from network formation as the focus here is on calculated strategies that achieve desirable goals in the combined network.

*Organizational network analysis* (ONA) amounts to a collection of tools in business management [33, 40]. Existing works apply SNA to study organizational processes and problems; they include identifying important individuals [16], improving awareness about informal networks [24], improving collaboration [32], building a new business [35].

*The importance of non-reporting links* within a business hierarchy has also been intensively studied. Firstly, in management studies, Krackhardt and Hanson [61] noted that much of the work in a company happens despite the formal organization. They draw an analogy between a company and a human body: the formal structure of a company is the skeleton, while the informal structure is the central nervous system. Informal networks are more flexible and adaptive; the formal structure is static. Kilduff and Krackhardt [58] develop this idea further by studying interpersonal networks in organizations; more specifically, they focused on the cognitive and personality distinctiveness of individuals and the ways in which such distinctiveness affects relationships in organizations. They argued that business organizations are built by incorporating formal relations of authority and informal links that connect people across departmental and hierarchical boundaries. Secondly, Cross et al. in [32] adopted a computational approach and argued that even though informal networks are invisible, they are more reflective than the formal organizations. The authors defined scenarios where SNA is useful to assess informal networks and facilitate effective collaboration. Finally, a number of works show how informal networks can be used to reveal the reporting hierarchy [42, 108].

Social networks play a central role in studies on important problems such as structural holes [2], group cohesion [21] and resource allocation [19]. The first mentioned work discusses a dual-structure within an organization, which consists of direct ties and indirect ties. However, different to our notions of formal and informal ties, direct ties refer to edges between nodes, and indirect ties correspond to paths in the network. The second mentioned work used a mathematical model to simulate the interactions between newly hired employees and relates cohesion with managerial autonomy. The third paper applies a mathematical, centrality-based approach to study two strategies in designing status-based competitions.

The network aspect of *power* and *influence* flourished in the last 5-10 years [23]. Power is a multiplex concept that is affected by many factors. There are indeed various approaches to measure influence and importance of individuals in a social networks. Our definition of power is based on Bonacich centrality [14], a widely-adopted eigenvector centrality measure in social networks. In [20] Bothner et. al. also used Bonacich centrality to social network analysis. However, in contrast to our work, they emphasized individuals' "statues" in a social network but did not take into account hierarchical levels. More recently, Franceschet and Bozzo [44] provided a definition of power that is motivated from negative exchange networks: a node gains power by connecting to nodes with low power.

Moving the reader's attention to network integration, we would like to mention the following interrelated research areas:

Firstly, *strategic network formation* considers how new links emerge due to rational and self-centred decisions of members of the network [53]. Strategic network formation models aim to explain how a network evolves in time [60]. A well known example along this line is on the rise of the Medici Family in the XV century [88], which explains how inter-family ties shape political structures. In a certain sense, the network integration problem can be regarded as network formation between two established networks. The network formation models are suitable for vast analysis of a single network transformation, but they lack tools to operate with several networks.

Secondly, *interdependent networks* discusses the complex structures formed through an integration of networks of different types (e.g. a transportation network with an electrical network) [36]; the focus here is mainly on interdependence among the nodes and robustness of network, i.e. whether node failures leads to a cascading failure throughout the overall infrastructure.

The interdependent networks aim to model a complex environment where multiple networks interact and form a type of *network of networks*. The networks in such a complex environment are non-homogeneous, i.e., the networks are of different types. For example, one may be interested in the interdependence between a telecommunication network and a transportation networks and how such interaction affects robustness of the entire infrastructure.

The interdependent networks indeed cover many multidisciplinary problems in economics, sociology, biology and so on. However, the models do not explain what happens when two homogeneous networks 'dissolve in each other' and become whole. An example could be two merging companies: not only new relations are being formed, but the entire structure of the new company should satisfy some certain criteria. Compared to interdependent networks, the integration problem considered in this thesis involves homogeneous networks and concerns a type of dynamic that 'dissolves' the two networks into one.

A third related area is *link prediction*, which aims to infer potential ties between nodes of a network [67]. Here, most approaches take into account surrounding contexts such as homophily and maximum likelihood.

Another work to mention is *network weaving* [63], a process where several disjoint components of a network become a single unit by establishing a hub.

Lastly, we mention management studies on *collaborative team building*. When two companies merge, the success of the new entity largely hinges on whether the firms can effectively socialize employees from both sides to a unified direction [1]. The challenge lies in how venues could be set up (e.g. meetings, group assignments, etc.) that nurture collaboration and allow efficient communication [102, 104]. The framework proposed in this thesis addresses this challenge through an algorithmic perspective.

## 1.5 Summary of Results

This thesis constitutes the results of our research towards structural and algorithmic analysis of organizational networks. We study how interpersonal ties become crucial empowerment channels that shape organizational structure. We develop an organizational network model that is consistent with management studies; moreover, by incorporating both formal and informal ties into the model, we build a promising theory that is capable to explain several organizational phenomena including flattening, workplace homophily, and loss of control. Through rigorous analysis, we demonstrate that our theoretical framework can be used to reflect general properties of organizations.

Our investigation unfolds in two main directions: the first covers individual power in organizations; the second lies in integrating two disjoint organizational networks.

The first focal point of this thesis is our centrality-based definition of *power* which is accompanied with comprehensive and deep analysis, case studies and experiments. Our power based model provides novel insights into a range of organizational properties: 1) Organizations have limited hierarchy height. 2) Flattening, the process when a business changes from a tall hierarchy to a flat structure by delayering, is closely related to changes in the power of employees. 3) Informal relations significantly impact power of individuals. 4) Leadership styles could be reflected and analyzed through understanding weights on the ties in an organizational network. Moreover, the model endorses a natural interpretation of the loss of managerial control: the more connections managers maintain, the less their power depends on each of their neighbors power.

Not only can interpersonal ties facilitate individual communication but also they serve as instruments of bringing entire organizations together. We define *network integration* as a process of building links between two networks so that the networks form a single unified network. To evaluate the effect of integration, we introduce the notion of *togetherness*, which measures the proximity of two networks. We first investigate the integration process for two homogeneous networks. We study the algorithmic nature of network integration, analyze the corresponding computational problems of network integration, and propose methods that generate solutions. To compare the methods, we perform thorough experimental analysis on both synthesized and real-world data.

Finally, we expand the integration concept to organizational networks. Motivated by different scenarios, we define two main approaches: *collaborative and dominant integration*. The collaborative integration simulates project communication and collaboration between two organizations, while the dominant integration represents an organizational acquisition. The dominant integration requires a wider perspective on togetherness, which we extend by introducing a new level of applicability.

The rest of the manuscript is structured as follows: Chapter 2 contains main definitions and a necessary background. In Chapter 3, we introduce our power based model, analyze organizational structures from the network perspective and perform experimental analysis. In Chapter 4, we consider integration of two arbitrary homogeneous networks: we define togetherness as a measure of proximity of two networks, study relevant network integration problems and propose several heuristics for solving these problems. The network integration concepts are then extended and modified to suit organizational networks with formal and informal ties in Chapter 5. Finally in Chapter 6, we mention directions of further research and state several open questions.

**Chapter 2. Preliminaries**

This chapter contains main definitions and a necessary background into social network analysis as well as our model of organizational structures and eight heuristics for finding small distance dominating sets. The model was first introduced in our work in [68], and the algorithms appeared in [80].

In the first section, we define a network as a connected graph $G = (V, E)$ with set of nodes $V$ and set of edges $E$. We list main characteristics of a network and define most common properties. We also mention some of the properties that a typical social network has. We then list and describe a number of centrality measures, which would later be used in this thesis. Centrality measures are most often applied to define advantageous positions in a network and/or to quantify importance of a certain node.

In the second section, we describe our model of organizational networks. An organizational structure consists of a network where employees are connected by working and social ties. Analyzing this network, one can discover valuable insights into information flow within the organization.

By integrating different interpersonal relations in the same network model, we suggest a uniform approach to perform organizational network analysis [32, 33, 40, 24]. Our model is consistent with management theory, and captures main traits of large corporations. More specifically, we define the structure of a firm as a network where employees are connected to their managers and each other by working ties. The carcass of the model is an organizational hierarchy. We extend it by allowing additional types of connections between two employees (e.g. collaboration, friendship, family relations and others), and introduce the notion of an *organizational network*. Having both reporting and non-reporting relationships, our model supports a multiplex approach to organization structures.

We define two main types of relationships: reporting (formal ties) and non-reporting (informal ties). Two principles, unity of direction and maximal relative capacity, enable us to define "well-built" organizational networks. The first principle refers to the idea of having a single source of instructions, and the second principle reflects a natural proposition that a person can maintain only a limited number of ties.

Finally, we mention the problem of finding distance $k$ dominating sets on graphs. We describe four greedy algorithms Max, Min, MinLeaf, and Btw. The first three algorithms have been introduced in [38] for regular graphs; Btw is a novel algorithm although it uses the same intuition as the three algorithms above. We also present our modifications of these algorithms, S-Max, S-Min, S-Btw, S-MinLeaf, that are also suitable for finding small distance $k$ dominating sets [80].

### Chapter 3. Power in Organizational Networks

The aim of this chapter is to analyze organizational structures from a network perspective using our notion of *power*. The results presented in this chapter have been published in [68, 69, 70].

We consider organizational networks as defined in Chapter 2. Then, we define a notion of power based on the Bonaich centrality [14]. This notion not only enriches the mathematical management theory [15] but also enables formal analysis of concepts specific to organizations such as stability and flattening. Comparing to existing centrality notions, our definition of power is novel in the following aspects:

(1) the model takes into account three types of interactions: the interaction between a manager and her subordinates, the mutual interaction effect between two employees connected by a non-reporting relation, and the *backflow* effect from a subordinate to her manager;

(2) the model enables a natural interpretation of the "loss of control" of a manager: the more connections a manager maintains, the less her power depends on each of her neighbors' power [77].

We introduce a novel business intelligence software tool, CORPNET, which is designed to provide automated and accurate decision support. The prototype implements statistical and stability analysis, community detection, synthesizing networks, and visualization. Using a range of parameters, the software not only allows identification of personal power in a company but also reasoning about leadership styles and strategies.

Based on our network model and definition of power, we are able to formally study multiple important phenomena relevant to organizational management. In particular, we build our theory around the following issues:

1. *Bounded height*: A management hierarchy typically involves a bounded number of levels, regardless of the individual capabilities. A common belief is that a tall hierarchy reduces the effectiveness of communication. Using a natural measure on the stability of a network, we provide an alternative explanation: as a company creates more and more levels in its hierarchy, it will eventually become unstable, i.e. employees at lower levels possess more power than those at the higher levels. See Section 3.2.

2. *Flattening*: Flattening is a well-known phenomenon of organizational change when a company acquires a new structure with fewer hierarchical levels. The alleged benefits of flattening include empowering employees, increasing flexibility, pushing down decision making, and improving information flow [64]. We provide a somewhat paradoxical view on flattening through computation: Although

flattening reduces average power in the company, the majority of employees gain more power. See Section 3.3.

3. *Workplace homophily*: Homophily refers to the tendency of individual to be associated and linked to others who are similar to themselves. In the workplace, this principle translates to the fact that employees tend to associate with people in the same unit (i.e. department, office, etc.) as well as the same level [76, 28]. Clustering of the formal tie hierarchy alone does not reveal this tendency in an organizational network. Hence, we provide a benchmark for informal ties that is in line with the observed homophily principle. See Section 3.4.

4. *Importance of informal ties*: As argued by numerous studies, informal ties significantly impact on organizations [32]. We analyze this phenomenon from the point of view of power consistency: A network is more likely to be destabilized by social links in taller hierarchies than in flattened hierarchies. On the other hand, the gap between the power of upper and lower levels can be diminished with the presence of informal ties. See Section 3.4

5. *Leadership styles*: Leadership styles refer to ways in which a manager leads by setting directions, carrying out plans and communicating with subordinates. It is generally agreed in management studies that leadership styles play a decisive role in shaping the working atmosphere and effectiveness of an organization [105]. Here we deviate from the traditional, behavioral approach to analyzing leadership styles, but provide network-oriented angle using parameters in the definition of power. See Section 3.5.

Through these analyses above, we demonstrate that our theoretical framework can be used to reflect general properties of organizations. This novel, structural approach to organization analysis provides us new insights, explanations and potentially predictive guidelines for organizational decision making.

This chapter has the following structure: Section 3.1 introduces our definition of power in the organization. Section 3.2 focuses on formal hierarchies of organizations and discusses the relation between stability and height. Section 3.3 discusses the phenomenon of flattening and tries to explain it from a network point of view by considering power distribution. In Section 3.4, we look at informal ties and study how they affect power. Section 3.5 applies our model to the analysis of leadership styles. In Section 3.6, we consider a real-world case study (Krackhardt and Hanson's network [1993]) and apply our power based approach. Section 3.7 discusses the software CORPNET.

### Chapter 4. Integrating Homogeneous Networks

This chapter approaches the problem of integrating two homogeneous networks. We define network integration as a process of building links between two networks so that the networks form a single unified network. Creating new ties in a network (particularly, in a social network) facilitates knowledge exchange and affects positional advantage. We study the process of establishing ties between two existing networks in order to reach certain structural goals.

We consider networks represented by connected undirected unweighted graphs, i.e. there is only one type of relations between two nodes. To evaluate the effect of integration, we introduce the notion of *togetherness*, which measures the proximity of two networks. As we mentioned above, this notion is fundamental to social networks and is relevant to important concepts such as trust, coherence and solidarity. We study the algorithmic nature of network integration and formally introduce three notions of togetherness.

We analyze the corresponding computational problems of network integration:

1. *Network Integration under Togetherness constraint.* Given two networks and a desired level of togetherness, build links between members of these networks so that the overall network meets the togetherness criterion. An optimal solution of this problem is a set of edges that has the smallest cardinality. We analyze optimal solutions to this problem, describe several heuristics and compare their performance through experimental analysis.

2. *Network Integration under Edge constraint.* Given a fixed number of edges, the problem asks which nodes to connect in order to maximize the togetherness. An optimal solution of this problem is a set of edges that leads to the largest togetherness.

We study computational complexity of the problems and propose methods that generate solutions. Broadly, these methods could be split into two parts:

1. The first type of methods are heuristics that are based on the equi-privilege properties of the networks. We propose heuristics that search for small sets of edges that integrate two networks. We assume that the networks enjoy *equi-privilege property*, i.e. any pair of nodes between networks can be freely connected. The goal of these heuristics is to gain maximal togetherness in the integrated networks. To motivate this formal framework, we make three assumptions: 1) the integration takes place assuming *equipotency* of nodes; 2) creating weak ties between the networks can be encouraged and forced; and 3) structural properties such as distance provide a measure of effective communication and resource accessibility.

2. The second type are simulations based on certain priorities given to nodes of the network. We propose four scenarios where nodes in one network preferentially establish links with nodes in the other network. We assume that every node is given a *priority* which is determined by the network structure. The difference between these priority-based methods and the heuristics in the first category is that their aim is to simulate the preferential attachments of links during integration, rather than explicitly searching for good solutions.

To compare the methods, we perform experimental analysis on both synthesized and real-world data.

As a special case, we focus on the scenario when one of the two networks consists only of a single member and motivate this case from two perspectives. The first perspective is *socialization*: we ask how a newcomer can forge relationships with an existing network to place herself at the center. We prove that obtaining optimal solutions to this problem is NP-complete, and present several efficient algorithms to solve this problem and compare them with each other. The second perspective is *network expansion*: we investigate how a network may preserve or reduce its diameter through linking with a new node, hence ensuring small distance between its members. For both perspectives the experiment demonstrates that a small number of new links is usually sufficient to reach the respective goal.

This chapter has the following structure: In Section 4.1, we introduce integrated networks, define three levels of togetherness and bounds on their values. We propose two network integration problems, show that they are, in fact, closely related to each other, argue that finding optimal solutions is in general computationally hard. We separately approach the network integration problems under the assumption of equi-privilege property and then by assigning priorities to all nodes. Section 4.2 contains algorithms for integrating networks with equi-privilege property (optimizing togetherness). In Section 4.3, we define four priority based heuristics and introduce a mechanism for integrating two networks following these heuristics. In Section 4.4, we consider the special case when one of the networks is represented by a single node. Finally, in Section 4.5, we present experimental results obtained by implementing proposed algorithms in 14 experiments. To compare our heuristics, we use both generated and real-world datasets.

Results presented in this chapter have appeared in [80, 81, 82].

**Chapter 5. Dominant and Collaborative Integration of Organizational Networks**

This chapter extends the results presented in Chapter 4, knitting them together with the concepts from Chapter 3. We consider organizational networks with two types of ties that represent formal hierarchical relations and informal non-hierarchical relations such as collaboration or friendship. Motivated by different scenarios, we define two approaches: *collaborative* and *dominant* integration. The first one represents collaboration of two companies; the second one illustrates an acquisition when one of the companies becomes the other company's part. The vital difference between the approaches is that collaborative integration is established by informal ties only. As a result, the networks are still independent. On the other hand, in the process of dominant integration, one of the networks becomes the other network's subnetwork.

To evaluate the effect of integration, we revisit our notion of togetherness. We argue that all edges, regardless of their type, serve as channels of communication between people [94]. Formal organizational structure is designed to perform a function of delivering commands; commands can also go through informal networks taking the form of advice [61]. However, a directive would unlikely pass from a subordinate to his or her manager: indeed, information may travel in any direction, while the presence of authority suggests that orders may not go up the hierarchical structure.

Universal, existential, and diametric togetherness, as defined in Chapter 4, are designed to capture the proximity of two networks; in the context of organizations these measures could be used to indicate how fast information gets from one network to the other. To capture how fast a command can reach individuals in a certain department, we introduce a new level of togetherness, *hierarchical* togetherness.

We extend the applicability of togetherness measures to reveal how well a certain department is "integrated" in the entire organization. Thus, togetherness in this chapter is considered as a measure of interaction between a certain unit with the rest of the network, or more generally, between two organizational networks after they establish some new relationships. Finally, we consider separately dominant and collaborative network integration problems and propose several ideas and heuristics for solving these problems.

This chapter has the following structure: In Section 5.1, we explain and motivate network integration for two organizations. Section 5.2 considers togetherness and introduces two approaches how this measure could be used: one is a local measure of proximity within a single network, and the other one is a measure to evaluate integration of two organizational networks. In Section 5.3, we study dominant integration: we split the problem into two steps and investigate what should be taken into account in order to solve the problem. Section 5.4 considers collaborative integration of organizational networks; in this section we consider integration of two subnetworks as a special case.

**Chapter 6. Conclusion**

In the last chapter, we suggest several directions in which the results presented in this thesis could be expanded.

# Chapter 2

# Preliminaries

## 2.1 Social Network Analysis

### 2.1.1 Defining a Network

Network analysis helps to understand and predict the behavior of complex networked systems [86]. Examples of these systems include the World Wide Web, transportation networks, distribution networks, neural networks, social networks, networks of business relations between companies, collaboration networks, and many others. *Social network analysis* (SNA) brings together social and mathematical sciences by applying network and graph theories to structures, which represent interactions and relations of individuals within the networks.

We view a *network* as a connected graph $G = (V, E)$ where $V$ is a set of nodes and $E$ is a set of edges on $V$. For any vertex $u \in V$, the *degree* is the number of edges incident to the vertex.

Some of networks we consider have *weighted* edges, meaning that edges have different importance or represent stronger/weaker connections than others. An edge is *directed*, denoted $\overrightarrow{(u, v)} \in E$ if it runs in only one direction, and undirected if it runs in both directions, denoted simply $(u, v) \in E$.

A *path* (of *length* $l$) is a sequence of nodes $u_0, u_1, \ldots, u_l$ where $u_i u_{i+1} \in E$ for any $0 \leq i < l$. A *geodesic path* is the shortest path, in terms of number of edges traversed, between a specified pair of vertices.

For any nodes $u, v \in V$, the distance between $u$ and $v$, denoted by $\mathsf{dist}(u, v)$, is the length of a shortest (geodesic) path from $u$ to $v$. The *eccentricity* of $v$, denoted $\mathsf{ecc}(v)$

is the largest path from $v$ to any other node in the graph. The largest eccentricity is the *diameter* of the graph, and the smallest is its *radius* denoted $\mathsf{diam}(G)$ and $\mathsf{rad}(G)$, respectively.

The *center* of $G$ is the set of all nodes that have the least eccentricity, i.e., $C(G) = \{u \in V \mid \mathsf{ecc}(u) = \mathsf{rad}(G)\}$. The *periphery* of $G$ is the set $P(G) = \{u \in V \mid \mathsf{ecc}(u) = \mathsf{diam}(G)\}$. The network $G$ is called *diametrically uniform* if $C(G) = V$.

### 2.1.2 Properties of Social Networks

Social networks have been intensively studied in the last few decades. In this chapter, we mention briefly some properties of social networks that are relevant to our research. More details could be found for example in [111]; for more complex insights one may consult in [86].

*Homophily* is the extent to which a person forms connections to similar people. Similarity can be defined by gender, race, age, occupation, educational achievement, status, values or any other salient characteristic [76].

A *small world property*, or six degree of separation, was first empirically shown by the social psychologist Stanley Milgram in the 1960s. This phenomenon demonstrates that any two individuals are, in fact, quite close to each other. We will use Newman-Watts-Strogatzs small-world network model [85], which produces graphs with small average path lengths and high clustering coefficient.

Another essential property of numerous real-world networks is that degree distribution of nodes follows a power law. Such networks are called *scale-free networks* [86] We will use Barabasi-Alberts preferential attachment model that generates scale-free graphs [3].

We say that two networks are *homogeneous* if they are of the same nature: For example, two social networks, representing collaboration among two groups of people, are homogeneous.

*Interdependent networks* is a system of coupled networks where nodes of one or more networks depend on nodes in other networks. Interdependent networks, clearly, are often non-homogeneous [36].

### 2.1.3 Centrality Measures

As pointed out in [84], centrality measures address the question, "Who is the most important or central person in this network?" There are several approaches to answer this question; following these approaches different centrality measures could be defined. However, one needs to be careful when choosing the most suitable centrality measure [18].

Let $R$ be the adjacency matrix of the network. We implicitly mean there is an indexing of all nodes in the matrix as natural numbers $1, \ldots, n$, and $R_{i,j}$ denotes the $(i, j)$-entry of $R$. If we consider unweighted networks, all $R_{i,j}$ are either 1, or 0, meaning there is or there is not an edge between $i$ and $j$.

The simplest measure is *degree centrality*, or simply degree, defined as $x_i = \sum_j R_{ij}$. This measure is too weak to be suitable for defining important positions in a social network. For example, the recent work [25] demonstrates the negative effect of the number of contacts in terms of job search. This rather contra-intuitive conclusion, however, confirms that not only the number of connections matter.

The approach taken whereby the centrality of a node is recursively related to the centralities of the node's neighbors, seems to be the one that we need. Eigenvector centrality is based on the same idea, but it takes into account that different connections have different significance (or quality):

$$\lambda p_i = \sum_j R_{ij} p_j,$$

or in matrix notation:

$$\lambda p = Rp,$$

where $p$ is an eigenvector of $R$, which is associated with the largest eigenvalue $\lambda$ (as can be proved using the Perron-Frobenius theorem).

While degree centrality gives a simple count of the number of connections a vertex has, eigenvector centrality acknowledges the fact that connections may be not equal. The intuition behind is that connections to people who are more influential would contribute to one's own influence more than connections to less influential individuals.

However as mentioned in [15], the *eigenvector centrality* can be used effectively only for symmetric adjacency matrices. In our model, the adjacency matrix is clearly asymmetric.

This measure is often called "approximate importance" - the centrality of a node depends on the centralities of its neighbors. The number of edges is still important, but nodes with fewer but more important connections can outrank nodes with higher degrees.

In 1987, Philip Bonacich [14] expanded the eigenvector by introducing parameter $\beta$: using positive and negative values of the parameter, one can apply it for "traditional" networks (i.e. when connections to more powerful nodes give more power), as well as for negative exchange networks (i.e. connections to less central nodes make a node more central.)

Let $R$ be an adjacency matrix. Then for any node $i$, Bonacich power is defined as following:

$$p_i(\alpha, \beta) = \sum_j (\alpha + \beta p_j) R_{ij} \tag{2.1}$$

In matrix notation, Bonacich power can be defined as following:

$$p(\alpha, \beta) = \alpha (I - \beta R)^{-1} Re, \tag{2.2}$$

where $I$ is an identity matrix, and $e$ is a column vector of ones.

The parameter $\alpha$ is a scalar that affects only the length of the power vector $p$. It means that we use it only to normalize the powers. In this thesis, $\alpha$ is selected such that the squared length of $p$ equals the number of nodes in the network. Then, $p_i(\alpha, \beta) = 1$ means (approximately) that position $i$ does not have an unusually large or small degree of centrality.

The parameter $\beta$ can be any value on the interval $[-\frac{1}{\lambda}, \frac{1}{\lambda}]$ where $\lambda$ is the largest eigenvalue. When $\beta = 0$, Bonacich power is the same as the degree centrality. When $\beta$ is positive, nodes that are connected to more central nodes are more central. In contrast, when $\beta$ is negative, more scores get nodes that are connected to less central.Note that when $\beta$ approaches $\frac{1}{\lambda}$ the centrality vector approaches the eigenvector, which is associated with $\lambda$.

Finally, another useful centrality measure is *betweenness centrality*, which is based on the network paths. The betweenness centrality of vertex $v$ is the fraction of geodesic paths between other vertices that pass through $v$. To compute the betweenness centrality, one needs to find the shortest paths between every pair of vertices, and thus, it is in general computationally costly.

## 2.2 Organizational Networks: Model and Key Properties

An organizational structure is often defined as a set of positions, groups of positions, reporting relationships, and interaction patterns [12]. We use the network approach and propose a model that captures main traits of a company. On the one hand, our model delineates the organizational hierarchy of a firm by featuring reporting relationship. On the other hand, we enrich the model by also including non-reporting relations. Indeed, as we will show later, these non-reporting relations can significantly affect a company as a whole.

In our model of *organizational networks*, the structure is represented as a network, where nodes stand for work positions (or departments, or people) and they are connected to each other by reporting relationships or some other social interactions:

**Definition 2.2.1** (Organizational network)**.** *An* organizational network *is a structure* $\mathcal{G} = (V, r, E_{\mathsf{fml}}, E_{\mathsf{inf}})$*, where* $V$ *is a set of* nodes*,* $E_{\mathsf{fml}}, E_{\mathsf{inf}} \subseteq V^2$ *are edge relations such that*

1. $r \in V$ *is called the* root *and* $(r, r) \in E_{\mathsf{fml}}$*;*
2. *the pair* $(V, E_{\mathsf{fml}})$ *forms a directed acyclic graph (ignoring the edge* $(r, r)$*), where every node apart from* $r$ *has an incoming edge from another node;*
3. *the pair* $(V, E_{\mathsf{inf}})$ *forms an undirected graph.*

Informally, the set $V$ denotes the individuals (or work positions) in the network. The root $r$ is the top manager, i.e. $r$ does not report to anyone else. The edge set $E_{\mathsf{fml}}$ represents the *reporting relation* on members of the network; if $(u, v) \in E_{\mathsf{fml}}$ then $v$ reports to $u$ and is called a *subordinate* of $u$. By the definition above, any nodes in the network may play the roles of *managers* and *subordinates*. Clearly, any node $v \neq r$ reports to its manager $u$ and thus is a subordinate of $u$; at the same time, $u$ may also have its subordinates. A node that has no subordinates is called an *operative*.

The edge set $E_{\mathsf{inf}}$ represents the undirected dyadic *non-reporting relation*. This could be collaborations, advice relations, or friendship between employees, etc. We will refer to edges in $E_{\mathsf{fml}}$ as *formal ties* since reporting relations are usually more important. We will call undirected edges in $E_{\mathsf{inf}}$ *informal ties*. For simplicity, we assume that any two nodes $(u, v)$ can be connected either by a formal tie or a informal tie, but not both. In fact, this can be justified intuitively: any reporting relation presumes some social interaction between a manager and her subordinates.

The definition of organizational networks covers most essential characteristics of a firm's structure, however, it is somewhat vague and is far from representing a well-built structure. To define a "well-built" structure, we accompany the definition above with two principles:

Firstly, *unity of direction* refers to the principle that there are one leader and one plan for business activities. It has been a fundamental criterion for an effective organization [114]. Having several sources of instructions, which in real life happens sometimes, often causes bungling decision-making. Translating this principle to our model, we assert that each person should have exactly one manager in the formal tie hierarchy.

**Principle 1: Unity of Direction.** Each node has exactly one incoming directed edge, which represents relationship with its manager, i.e., for all $u \in V$ there is a unique $v \in V$ with $(v, u) \in E_{\mathsf{fml}}$.

Principle 1 requires the directed graph $(V, E_{\mathsf{fml}})$ to form a tree structure, which we call the *formal tie hierarchy* (or *reporting hierarchy*) of $\mathcal{G}$. The managers of the hierarchy are all the internal nodes of the tree $(V, E_{\mathsf{fml}})$ and the operatives are the leaves. The top (level 0) of the hierarchy contains only the root $r$. We will use the following terminology.

**Definition 2.2.2.** *The* level *of any node $v$ in $\mathcal{G}$ is the length of the path from $r$ to $v$ in the reporting hierarchy. The* height *of the hierarchy is the number of levels.*

Secondly, one may notice that a person can maintain only a limited number of interpersonal relations, due to limited time and effort. In fact, all social networks emerge under the constraint of limited resources. For example, in the context of online social networks, the number of formal ties (mutual communication during some period) for networks of more than 500 nodes on Facebook varies from 10 to 20 [39].

In defining the notion of *capacity* of individuals, we distinguish the formal and informal ties in regarding how much resource each of them consumes. Let $\Delta$ be an abstract quantity that defines the maximum amount of resources (working hours, for instance) that a person can distribute between his or her ties. For simplicity, we assume that each individual in the network has the same amount of resources $\Delta$. We also assume that a person needs $f$ resources and $i$ resources to maintain a formal and an informal tie, respectively. The root node also spends $f$ resource on some exogenous factors, which are represented by the loop $(r, r)$. Therefore, for any node $v$, if $|E_{\mathsf{fml}}(v)|$ is the number of directed edges (including self-loop), and $|E_{\mathsf{inf}}(v)|$ is the number of undirected edges, then $A$ spends $|E_{\mathsf{fml}}(v)| \times f + |E_{\mathsf{inf}}(v)| \times i \leq \Delta$ resources to maintain all his connections. Let $\delta := \frac{i}{f}$ be called the *correlation coefficient*.

**Definition 2.2.3.** *The* relative degree *of a node $v \in V$ is defined as* $\mathsf{deg}(v) = |E_{\mathsf{fml}}(v)| + |E_{\mathsf{inf}}(v)| \times \delta$, *where* $|E_{\mathsf{fml}}(v)|$, $|E_{\mathsf{inf}}(v)|$ *are the numbers of formal ties (including both incoming and outgoing edges) and informal ties $v$ maintains, respectively.*

Clearly, if $\delta = 1$, then we assume that maintaining a informal tie requires the same amount of resources as maintaining a formal tie; in this case, the relative degree $\mathsf{deg}(v)$ is the conventional degree notion in graph theory. Such assumption may be reasonable when the organization contains equipotent members who have respective expertise (e.g., a research team). The lower $\delta$ is, the greater distinction there is between formal and informal ties.

The *relative capacity* of a node $v$ is a given number that defines the upper bound on its relative degree $\mathsf{deg}(v)$. In other words, it defines the total available resources for a person to maintain all ties.

**Principle 2: Maximal Relative Capacity.** There is a constant relative capacity $c$ for any node $v \in V$ .

Management theory defines the *span of control* of a manager as the number of her direct subordinates. If we only consider the reporting relation, Principle 2 guarantees that the span of control of every individual is limited[57], and, thus, refers to the "limited managerial attention", a phenomenon in hierarchy theory [47]. The loop $(r, r)$ guarantees that the root must not have more direct subordinates than all the other managers and, hence, make our approach uniform.

**Definition 2.2.4.** *An organizational network is called* well-built *if it satisfies the principles 1 and 2.*

In this thesis we assume that all organizational networks are well-built without explicit mention. Given its simplicity, the model of organizational networks above has natural limitations, which we explain in the remarks below:

**Remark 1.** We remark that the requirement that there is a single root of the network may seem too restrictive. Indeed, large corporations tend to have a board of directors. Nevertheless, we argue that this simplified model be still reasonable as the board of directors normally perform as a whole by hiring a CEO. The loop $(r, r) \in E_{\mathsf{fml}}$ indicates that the root makes decisions by herself. Another reason why we need this loop is technical – as we will show later, it makes the capacity of nodes uniform.

**Remark 2.** In this model, we eliminated the functional differences between individuals in the organization. Such a restriction again may seem like a departure from real life; Indeed, people in a corporation perform vastly different tasks, and it is these tasks that give their positions real "meaning". Nevertheless, we argue that the model still encapsulate meaningful interpretation: Firstly, the interpersonal ties in the model capture in a certain sense the channels of information/resource flow within an organization, irrespective of the duties of individuals. Secondly, regardless of the tasks of individuals, any person who has at least one subordinate will need to act in the role of a manager, which involves a type of decision-making process or directive actions that are common to all managers. Thirdly, our goal is to use a general model that captures any types of organizational structure which may perform very different functions (e.g. a university department or a bank). It is very difficult to fix tasks for all kinds of roles in such a general setting.

## 2.3 Finding a Small Distance-$k$ Dominating Set in the Network

Many problems on social networks involve searching for the shortest paths. Solving such problems on large datasets is often very costly, and it requires incorporating different heuristics. We will consider computational problems that are related to the problem of finding dominating sets on graphs.

**Definition 2.3.1.** *Let $G = (V, E)$ be a graph. The subset $D \subset V$ is called a* dominating set *for $G$ if every vertex not in $D$ is adjacent to at least one member of $D$. The* domination number *$\gamma(G)$ is the number of vertices in a smallest dominating set for $G$.*

The dominating set problem concerns testing whether $\gamma(G) \leq K$ for a given graph $G$ and input $K$. The (connected) dominating set problem is one of the fundamental mathematical problem underlying routing. It is a classical NP-complete decision problem in computational complexity theory.

**Definition 2.3.2.** *We say that $D_k \subset V$ is a* distance $k$ dominating set *if for every vertex $v \in V$, either $v \in D_k$ or there exists a vertex $u \in D_k$ that is at distance at most $k$ from $v$.*

Distance $k$ dominating sets may be used, for example, to solve resource allocation problems [9].

We let $\gamma_k(G)$ denote the number of nodes in a smallest distance $k$-dominating set. Chang and Nemhauser in [29] showed that the problem of finding the minimum distance $k$ dominating set is also NP-hard.

We will now present several heuristics for finding small distance $k$ dominating sets

### 2.3.1 Four Greedy Algorithms

We present four greedy algorithms that are suitable for finding distance-$k$ dominating sets. Each algorithm below applies a heuristic that iteratively adds new nodes to the solution set $S$.

The starting configuration is $S = \varnothing$ and $U = V$. During its computation, the algorithm maintains a subgraph $F = (U, E \restriction U)$, which is induced by the set $U$ of all "uncovered" nodes, i.e., nodes that have distance $> (r-1)$ from any current nodes in $S$. It repeatedly performs the following operations until $U = \varnothing$, at which point it outputs $S$:

1. Select a node $v \in U$ based on the corresponding heuristic and add $v$ to $S$.

2. Compute all nodes at distance at most $(r-1)$ from $v$. Remove these nodes and all attached edges from $F$.

**Algorithm Max (Max-Degree).** The first heuristic is based on the idea that one should connect to the person with the highest number of social ties. The intuition here is that a newcomer would be benefited by linking to someone with maximal connections.

At each iteration, it adds to $S$ a node with maximum degree in the graph $F$. Using a priority queue (e.g., a Fibonacci heap) to keep track of maximal degrees and assuming the radius rad is given, the algorithm can be implemented in time $O(n \log n + m)$ where $n = |V|$ and $M = |E|$.

**Algorithm Min (Min-Degree).** The second heuristic first picks a node $v$ with minimum degree in $F$, then adds to $S$ a node with maximum degree that is adjacent to $v$. The intuition is as follows: A node is called a *leaf* if it has minimum degree in the graph; leaves correspond to least connected members in the network, and may become outliers once nodes with higher degrees are removed from the network. Hence this heuristic gives first priority to "covering" leaves; Namely, it always picks a node with maximum degree that is adjacent to a leaf.

To implement this algorithm, one may use a min-priority queue to keep track of degrees of all nodes; this is used to extract a leaf at each iteration. Then with each node $v$, one may associate a max-priority queue to keep track of degrees of all adjacent nodes of $v$. The overall running time of min-priority queue operations is $O(m+n \log n)$. The overall running time of max-priority queue operation is $O(m \log d)$ where $d$ is the maximum degree in $G$. Therefore assuming rad is given, the algorithm can be implemented in time $O(n \log n + m \log d)$.

**Algorithm Btw (Betweenness).** The third heuristic is based on *betweenness*, an important centrality measure in networks [13]. More precisely, the *betweenness* of a node $v$ is the number of shortest paths from all nodes to all others that pass through $v$. Hence, high betweenness of $v$ implies, in some sense, that $v$ is more likely to have short distance with others. This heuristic works in the same manner as Max but picks nodes with maximum betweenness in $F$.

Using Brandes' algorithm, calculating betweenness centrality of all nodes requires time $O(mn)$. However since the algorithm requires updating the graph and re-computing betweenness after each of the $O(n)$ iterations, the algorithm can be implemented in time $O(mn^2)$.

**Algorithm** MinLeaf **(Min-Leaf).** The fourth heuristic is an improvement of Min. To remove a leaf $v$ from $F$, one does not need to always add an adjacent node to $v$; it is enough to add to $S$ a node that has distance at most $r-1$ from $v$. The heuristic first picks a leaf $v$ in $F$, then applies a sub-procedure to find the next node $w$ to be added to $S$. The sub-procedure determines a path $v = u_1, u_2, \ldots$ in $F$ iteratively as follows:

1. Suppose $u_i$ is picked. If $i = r$ or $u_i$ has no adjacent node in $F$, set $u_i$ as $w$ and terminate the process.

2. Otherwise select a node $u_{i+1}$ (which is different from $u_{i-1}$) among adjacent nodes of $u_i$ with maximum degree.

After the process above terminates, the algorithm adds $w$ to $S$. Note that the distance between $w$ and $v$ is at most $r-1$.

The algorithm can be implemented using the same data structures as Min. The sub-procedure used to compute $w$ at each iteration does not take extra computational time as edges of each $u_i$ are removed during the procedure. Hence assuming $r$ is given, the algorithm takes time $O(n \log n + m \log d)$.

We mention that Algorithms Max, Min, and MinLeaf have been applied in [38] to *regular graphs*, i.e., graphs where all nodes have the same degree. In particular, MinLeaf has been shown to produce small $k$-dominating sets for a given $k$ in the average case for regular graphs: : e.g. the size of the output $S$ is roughly 10% of the number of nodes when all nodes have degree 5. The Btw-algorithm was introduced in [80]

### 2.3.2 Simplified Greedy Algorithms

One significant shortcoming of the four algorithms above lies in the fact that by deleting nodes from the network $G$, the network may become disconnected, and nodes that could have been connected via short paths are no longer reachable from each other. This process may produce *isolated* nodes in $F$, i.e., nodes having degree 0, which are subsequently all added to the output set $S$. Moreover, maintaining the graph $F$ at each iteration also makes implementations more complex. Therefore we next propose *simplified* versions of Algorithms Max, Min, Btw, and MinLeaf.

**Algorithms S-Max, S-Min, S-Btw, S-MinLeaf.** The simplified versions of Algorithms Max, Min, Btw, and MinLeaf act in a similar way as their "non-simplified" counterparts; the difference is that here the heuristic works over the original network $G$ as opposed to the updated network $F$. Hence, the graph $F$ is no longer computed. Instead we only need to maintain a set $U$ of "undominated" nodes.

The simplified algorithms have the following general structure: Start from $S = \varnothing$ and $U = V$, and repeatedly perform the following until $U = \varnothing$:

1. Select a node $v$ from $U$ based on the corresponding heuristic and add $v$ to $S$.
2. Compute all nodes in the sub-radius ball $B(v)$ of $v$, and remove any node in $B(v) \cap U$ from $U$.

The set $S$ is the output. We stress that we apply the same heuristic as described above in Algorithms Max, Min, Btw, and MinLeaf, but replace any mention of "$F$" in the description with "$U$", while all notions of degrees, distances, and betweenness are calculated based on the original network $G$.

# Chapter 3

# Power in Organizational Networks

In this chapter, we present a centrality-based definition of power. This measure enables us to identify important individuals in the network. Our model provides novel insights into a range of organizational properties: 1) Organizations have limited hierarchy height. 2) Flattening, the process when a business changes from a tall hierarchy to a flat one by delayering, is closely related to changes in the power of employees. 3) Informal relations significantly impact power of individuals. 4) Leadership styles could be reflected and analyzed through understanding weights on the ties in an organizational network. We implement our model and tools in a stand-alone application CORPNET, which provides functions for generating synthesized organizational networks, analyzing and visualizing interpersonal relations, and computing network measures.

The aim of this chapter is to analyze organizational structures from a network perspective. More specifically, the main contribution of this chapter is three-fold.

**Firstly,** by integrating different interpersonal relations in the same network model, we suggest a uniform approach to perform *organizational network analysis* (ONA) [32, 33, 40, 24]. Our model is consistent with management theory, and captures main traits of large corporations. More specifically, we define the structure of a firm as a network where employees are connected to their managers and each other by working ties. The carcass of the model is an organizational hierarchy. We extend it by allowing additional types of connections between two employees (e.g. collaboration, friendship, family relations and others), and introduce the notion of an *organizational network*. Having both reporting and non-reporting relationships, our model supports a multiplex approach to organization structures.

Throughout this chapter we consider well-build organizational networks as defined in Chapter 2.

**Secondly,** we define a notion of power based on a centrality measure for individuals in an organization. This notion not only enriches the mathematical management theory [14, 15] but also enables formal analysis of concepts specific to organizations such as stability and flattening. Comparing to existing centrality notions, our definition of power is novel in the following aspects: 1) the model takes into account three types of interactions: the interaction between a manager and her subordinates, the mutual interaction effect between two employees connected by a non-reporting relation, and the *backflow* effect from a subordinate to her manager. 2) the model enables a natural interpretation of the "loss of control" of a manager: the more connections a manager maintains, the less her power depends on each of her neighbors' power [77].

**Thirdly,** based on our model, we design and implement a novel business intelligence software tool, CORPNET, to provide automated and accurate decision support. The prototype implements statistical and stability analysis, community detection, synthesizing networks, and visualization. Using a range of parameters, the software not only allows identification of personal power in a company but also reasoning about leadership styles and strategies.

## 3.1 Measure of Power

### 3.1.1 What Defines Power: a Network Perspective

Power is a multiplex concept that is affected by behavioral, cognitive as well as social factors. Among these factors, the social network has been identified as the most significant one; as pointed out in [91], "power is first and foremost a structural phenomenon, and should be understood as such". Following a social network approach to organization analysis [23], we focus on power that emerges from the organizational network.

**Example 1.** Consider the organizational network as described in Figure 3.1. The directed edges form the set $E_{\mathsf{fml}}$ of formal reporting relations, and undirected edges form the set $E_{\mathsf{inf}}$ of informal ties. It is natural to believe that $r$ would enjoy a high power in this organization as $r$ is on the top level of the formal tie hierarchy. Comparing the managers $A$, $B$ and $C$ on level 1, we see the following differences: $A$ has two direct subordinates, but he does not maintain any informal tie; $C$ has three subordinates, but only one of them is a child, while $B$ does not have any outgoing directed edges; nevertheless, she has collaboration with all nodes except $A$. Several natural questions arise: which position is the 'best' among $A$, $B$, and $C$? How much power does each node has? Does the link between $B$ and $C$ affects $B$'s power the same way as the link between $B$ and $D$ does? All of these questions originate from a social network perspective of power, which we elaborate below:

- The network structure defines the formal tie hierarchy of the network, and hence expresses certain *legitimate power* in the organization [93]. For example, $r$ naturally has power since as top manager, $r$ has a responsibility to make decisions.
- The network structure also implies a type of *referent power* [45]. For example, having extensive and broad interpersonal ties (e.g. manager $B$) also means that the individual is capable of developing statues and building loyalty.
- Viewing power as a product of the competition for resources, one may regard interpersonal ties as sources for resources, i.e., they serve as access points for resources such as information and skills (human resources) [92]. For example, $A$ has access to information that may be passed from his subordinates, while $B$'s informal ties provide her with information across diverse departments. Both of these cases empower the particular individuals.
- From a social exchange theory perspective, interpersonal ties provides people alternatives during negotiations and hence enhance one's power [8]. For example, the fact that $A$ has two subordinates means that $A$ is at a more advantageous position when he assigns tasks to the subordinates, i.e., the competition between the two subordinates may allow $A$ to exercise more control. On the other hand, $B$ is also in an advantageous position when seeking advice from his peers as he has informal connections across wide parts of the network.

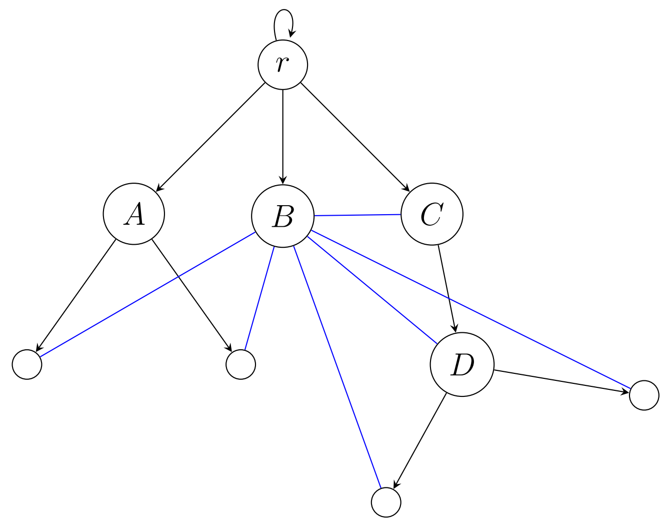



Figure 3.1: Defining power of $A$, $B$ and $C$

### 3.1.2 Power, Influence, and Authority

Marketing and management studies customarily compare power with other notions such as authority and influence. Indeed, it is tempting to use these terms interchangeably as they all imply the ability to affect others and infer leadership. Before we proceed with a formal definition of power, it is necessary to clarify the differences between these notions.

***Influence*** is a measure of the ability of one person to affect another person's perception, attitude, and thought. In management studies, influence commonly relies on skillful tactics to alter the other person's point of view [59]. In marketing, influence is often associated with the word of mouth in consumer decisions [55]. A major line of research concerns the use of social networks to analyze the individuals' connections to harness *influence* word of mouth by identifying influencers and predicting adoption probabilities. More recently, efforts have been focused on the spread of influence through a physical diffusion model, and influential individuals in a complex network that maximally spread influence in this model [78][56][4].

***Authority*** refers to the right given to a person to achieve the objectives of the organization. In other words, it is an entitlement of the individual and thus is predominantly a positional concept [50]. For example, the top manager of an organization has the authority to make decisions about the future directions of the company largely thanks to her position. Hence authority refers to a certain form of privilege.

***Power*** of an individual is defined by three intuitive factors: The first is the person's proximity to the root of the reporting hierarchy. The second is the number of ties the individual maintains – more connections provide more sources of information. Finally, the span of control indicates how many subordinates a person has, and, hence, how much involved he or she is in making decisions over the network.

The notion of power is distinct from both influence and authority in the following aspects. Firstly, power refers to the overall ability of the person to define the entire course of the organization. Thus power is considerably distinct with influence, which in principle relates to the capacity to affect the behavior of one's neighbors. In this sense, there is an overlap between power and the *spread* of influence. However, the spread of influence is the outcome of a physical process [55], while power is a fixed attribute that is defined by the network structure. Secondly, authority denotes the type of power that is accepted within an organization and is derived from the formal roles. While formal ties affect power, informal ties also play a crucial role, which is not captured by authority. Hence power is also inherently different from authority.

### 3.1.3 Definition of Power

To capture the difference between formal and informal ties, we assign a weight of 1 to all formal ties and a weight of $k$ between the values 0 and 1 to all informal ties to represent the strength of the tie. We call the parameter $k$ *interaction effect*. In some sense, interaction effect measures the capability to affect the neighbors and is therefore similar to the notion of influence. However, here we keep a uniform weight to for all informal ties as the focus is not on the influence of individual ties, but rather the overall power of nodes. Furthermore, we introduce a weight $\rho$ to the self-loop on the root, which measures the "self-assertiveness" of the root. A weight $\rho$ of 1 suggests that it has the same effect as all the other formal ties, while 0 means that the loop only affects the capacity of the root node but not the power. We will show later that the weight $\rho$ is useful for defining *leadership styles*: A larger $\rho$ indicates a more "autocratic" style of management. For simplicity, we will assume that $\rho = 1$ if without explicit mention.

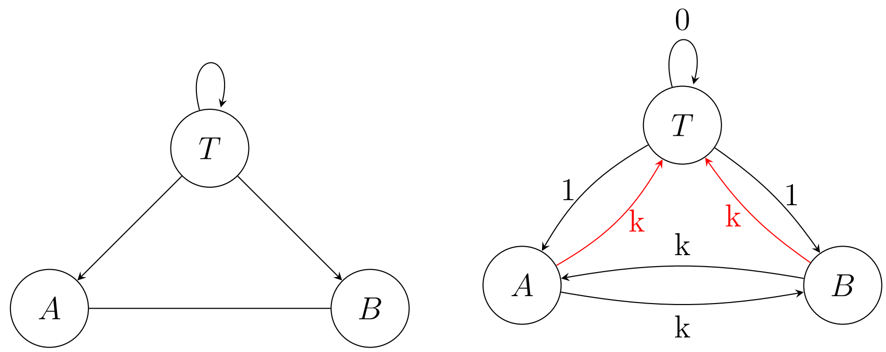



Figure 3.2: An organizational network (on the left) and its weighted interaction graph (on the right)

More formally, we define the *weight function* $\mu$ which depends on the two parameters $k$ and $\rho$:

- *Formal ties*: If $e \in E_{\mathsf{fml}}$, we define the following:

$$\mu_{\rho,k}(e) = \begin{cases} 1 & \text{if any end point } e \text{ is a non-root node} \\ \rho & \text{if } e = (r, r) \end{cases}$$

- *Informal ties*: If $e \in E_{\mathsf{inf}}$, we assign $\mu_{\rho,k}(e) = k$. The range of $k$ guarantees that directed edges are more important than undirected. An edge from $A$ to $B$ can be interpreted as the interaction effect between $A$ and $B$, ranked as the weight of this edge.

- *Backflow*: It is natural to assume that the interaction effect between an employee and her manager is not one-way: While the manager is empowered by having subordinates, the subordinate also gains power from her manager through social interaction. On one hand, as a manager acquires subordinates, the manager has increased her span of control and, hence, becomes more powerful. On the other hand, the subordinate also increases power through support and patronage of the manager. Hence, this effect could be captured by an informal tie from the subordinate back to the manager, i.e., *a backflow*; and we assume that such tie exists. Therefore, we set $\mu_{\rho,k}(e) = k$ where $e = (u, v)$ whenever $u$ reports to $v$.

Based on the definition above, we regard any organizational network as a *weighted interaction graph*; an example of this is shown in Figure 3.2.

**Definition 3.1.1** (Weighted interaction graph). *Let $\mathcal{G} = (V, r, E_{\mathsf{fml}}, E_{\mathsf{inf}})$ be an organizational network. The weighted interaction graph of $\mathcal{G}$ is*

$$W(\mathcal{G}) = (V, r, E_{\mathsf{fml}}, E_{\mathsf{inf}}, k, \rho, \mu_{\rho,k})$$

*where the parameters $k, \rho \in [0, 1]$ and the weight function $\mu_{\rho,k}$ are defined as above.*

Bonacich power, introduced in [14], is a widely-adopted eigenvector centrality measure in social networks. The basic idea is that the power of any individual depends on the power of those it is connected to; the difference between Bonacich power and the usual eigenvector centrality is the inclusion of a parameter $\beta$, which affects the meaning of centrality.

**Definition 3.1.2** (Bonacich power). *Let $R$ be the adjacency matrix of the network (here we implicitly mean there is an indexing of all nodes in the matrix as natural numbers $1, \dots, n$), and $R_{i,j}$ denotes the $(i, j)$-entry of $R$. The* Bonacich power *of $i = 1, \dots, n$ is*

$$p_i = \sum_{j=1}^{n} (\alpha + \beta p_j) R_{i,j} \tag{3.1}$$

*where $\alpha, \beta$ are scalar constants. In matrix form, the vector of Bonacich power $\vec{p} = (p_1, \dots, p_n)$ is*

$$\vec{p} = \alpha(I_n - \beta R)^{-1} R \vec{e}_n \tag{3.2}$$

*where $I$ is the $n \times n$ identity matrix, and $\vec{e}_n$ is the column vector of ones with length $n$.*

It is clear that different values of $\alpha$ and $\beta$ result in different centrality measures. Here $\alpha$ only serves as a normalizing factor; It is selected such that the norm $||\vec{p}||$ equals $\sqrt{n}$. Thus, the most "evenly distributed" case is when $p_i = 1$ for every $i = 1, \dots, n$.

For the matrix $I_n - \beta R$ to be invertible, the parameter $\beta$ can be any value on the interval $[-\frac{1}{\lambda}, \frac{1}{\lambda}]$ where $\lambda$ is the dominating eigenvalue of $R$. In some sense, it captures the contribution of ties of a node to its power. When $\beta = 0$, Bonacich power coincides with degree centrality. When $\beta > 0$, a node becomes more powerful as its neighbors become more powerful. In contrast, when $\beta$ is negative, nodes become more powerful as their neighbors become less powerful[1].

[1] A negative value of $\beta$ implies a negative exchange power where connections to nodes with smaller power results in a bigger power. See e.g. [15] and [44]

Intuitively, when $\beta > 0$, it specifies how much the power of a person depends on the power of her neighbors. Thus the parameter $\beta$ also corresponds to a managerial reality. The principle of *loss of control* states that as an individual acquires more social ties, the less her power depends on each of her neighbor's power [77]. We can reflect this principle by setting a range for $\beta$. In particular, since the capacity $c$ indirectly indicates how much effect a person spends with each of their subordinates, friends or collaborator, we require $\beta$ to be inversely proportional to the capacity minus one (the "minus one" is for the relation with its manager):

$$\beta < \begin{cases} \frac{1}{\lambda} & \text{if } \lambda > 1, \\ \frac{1}{c-1} & \text{otherwise} \end{cases} \tag{3.3}$$

To derive a measure of power in an organizational network, we adopt Bonacich power on the interaction graph of the network.

**Definition 3.1.3.** *Let $i = 1, \ldots, n$ be a node in $\mathcal{G}$. Let $\mathsf{Fml}_i$ denote the set $\{j \mid 1 \leq j \leq n, (i,j) \in E_{\mathsf{fml}}\}$ of all subordinate of $i$, let $\mathsf{Inf}_i$ denote the set $\{j \mid 1 \leq j \leq n, (i,j) \in E_{\mathsf{inf}}\}$ of nodes connected from $i$ by informal ties, and let $\mu_i$ be the node such that $(\mu_i, i) \in E_{\mathsf{fml}}$. We define the* power $p_i$ *of $i$ as discussed above, i.e., by (3.1) it is*

$$p_i = \sum_{s \in \mathsf{Fml}_i} (\alpha + \beta p_s) + k \sum_{w \in \mathsf{Inf}_i \cup \{\mu_i\}} (\alpha + \beta p_w) \tag{3.4}$$

Now we can answer the questions stated in Example 1. Let the correlation coefficient $\delta = 0.5$. Assume that capacity of each node is 4, and $\beta = 0.3 < \frac{1}{3}$. Figure 3.3 shows the resulting power of each node when $k = 0.5$ (left) and $k = 0.1$ (right).

When $k = 0.5$, even though $B$ does not have a single subordinate, she is almost as powerful as the top manager while $A$ and $C$ possess similar power. However, when $k = 0.1$, $A$ and $C$ are much more powerful than $B$. Hence $k$ captures in some sense the "importance" of informal ties.

Note also that the power of $D$, who has two subordinates (through formal ties) and an informal tie with $B$, exceeds her manager $C$ in both cases above. We interpret this situation as follows: Since our notion of power aims to capture a node's ability to promote the node's ideas and decisions to others, it denotes in a sense a level of "real power". In the case of $D$, the real power is higher than her "nominal power",

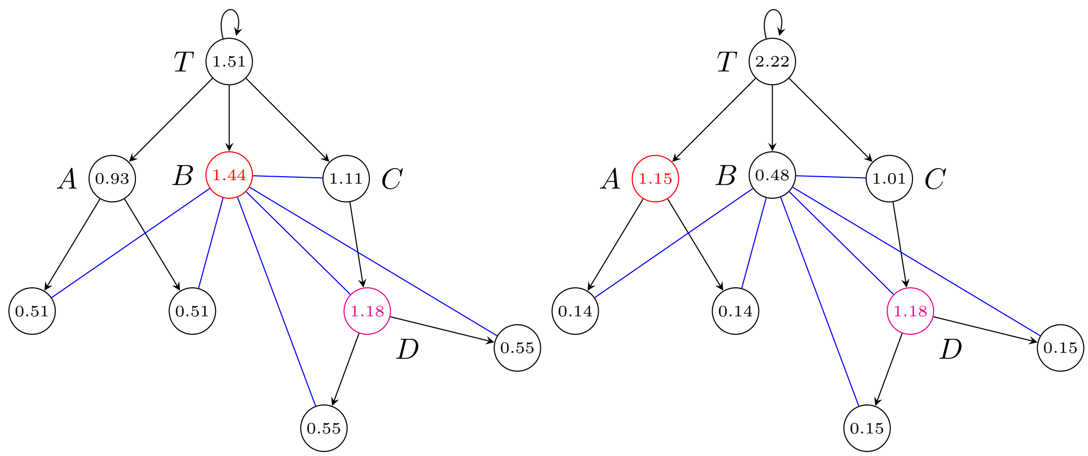



Figure 3.3: Individual power with $k = 0.5$ (left) and $k = 0.1$ (right)

which is indicated by her level formal position. This may imply a form of "inconsistency" within the structure, as $D$ may seek more formal recognition (say, in the form of promotion). Furthermore, $C$ may experience certain loss of control over $D$'s subordinates, as communication may not effectively pass down from $C$ to these nodes. Such inconsistency gives the network a potential to change. Thus, we say that in this case the organizational network is *unstable*. We stipulate that in a *stable* network, the levels in the formal tie hierarchy should truthfully reflect the actual power of individuals. In other words, the power of nodes is consistent with their respective levels in the reporting hierarchy.

**Definition 3.1.4.** *An organizational network $\mathcal{G}$ is* stable *if for any nodes $i, j \in V$, $\mathsf{lev}(i) < \mathsf{lev}(j)$ implies that $p_i > p_j$ where $\mathsf{lev} : V \to \mathbb{N}$ maps every node to its level in the hierarchy of $E_{\mathsf{fml}}$. We say that $\mathcal{G}$ is* unstable *if it is not stable.*

This definition allows us to formally analyze several phenomena, which we elaborate in the subsequent sections.

## 3.2 Stability and Height

Some scholars argue that not only the span of control is limited in organizations, but there is an "optimal" number of employees for any company [113] resulting in the limited size of its network. In this section, we continue to study the relation between the notion of stability introduced above and the height of an organizational network.

The goal is to explore the formal tie hierarchy throughout this and the next section, thus, we assume the set of informal ties $E_{\mathsf{inf}} = \varnothing$.

### 3.2.1 Chain Networks

Consider a network $\mathcal{C}_n$ consisting of $n$ nodes $1, \ldots, n$ such that $E_{\mathsf{fml}} = \{(i, i+1) \mid 1 \leq i \leq n-1\}$; this is a *chain* of $n$ nodes. A chain network does not appear as a typical management structure; clearly, a large number of nodes connected in a chain structure leads to ineffective communication as the top node will find it difficult to pass down her power to the bottom of the chain. In the following, we show that the notion of stability provides us a formal evidence for this ineffectiveness of the chain network. By (3.4), the power of node $i$ is

$$p_i = \begin{cases} \alpha + \beta p_2 & \text{if } i = 1 \\ k(\alpha + \beta p_{n-1}) & \text{if } i = n \\ \alpha + \beta p_{i+1} + (\alpha + \beta p_{i-1})k & \text{if } 2 \leq i \leq n-1 \end{cases} \tag{3.5}$$

**Lemma 3.2.1.** *If $0 < \beta \leq k < 1$, $\mathcal{C}_n$ is unstable for any $n > 2$.*

*Proof.* By (3.5) we get the following derivation

$$\begin{aligned} p_1 &= \alpha + \beta p_2 \\ &= \alpha + \beta(\alpha + \beta p_3 + (\alpha + \beta p_1)k) \\ &= \alpha + \alpha\beta + \alpha\beta k + \beta^2 p_3 + \beta^2 k p_1 \end{aligned}$$

In other words, $p_1 = \dfrac{\alpha + \alpha\beta + \alpha\beta k + \beta^2 p_3}{1 - \beta^2 k}$.

Similarly, by (3.5) we get

$$\begin{aligned} p_2 &= \alpha + \beta p_3 + (\alpha + \beta p_1)k \\ &= \alpha + \alpha\beta k + \alpha k + \beta p_3 + \beta^2 k p_2 \end{aligned}$$

In other words, $p_2 = \dfrac{\alpha + \alpha k + \alpha\beta k + \beta p_3}{1 - \beta^2 k}$.

Combining the above, we get

$$p_1 - p_2 = \frac{\alpha\beta + \beta^2 p_3 - \alpha k - \beta p_3}{1 - \beta^2 k}$$

Since $1 - \beta^2 k > 0$ for any positive $\beta, k < 1$, $p_1 - p_2$ is negative whenever $\alpha\beta + \beta^2 p_3 < \alpha k + \beta p_3$. Clearly, since $\alpha$ is positive and $\beta^2 p_3 < \beta p_3$, $\beta \leq k$ implies $p_1 < p_2$ for any $n > 2$ □ □

**Lemma 3.2.2.** *The chain $\mathcal{C}_n$ is stable if and only if $p_1 > p_2$.*

*Proof.* We only need to prove the "only if" direction. Suppose $p_1 > p_2$. Then by (3.5), $\beta p_2 > \beta p_3 + (\alpha + \beta p_2)k$. Since $(\alpha + \beta p_2)k > 0$, $p_2 > p_3$. Consequently, we have $\beta p_2 > \beta p_4 + (\alpha + \beta p_2)k$, and hence $p_2 > p_4$. Inductively, we may show that $p_2 > p_i$ for any $i = 3, \ldots, n$.

We now prove that $p_i > p_{i+1}$ for any $i = 3, \ldots, n-1$. Suppose on the contrary that $i > 2$ is the smallest such that $p_{i+1} \geq p_i$. Then by (3.5), we have $p_{i+2} + kp_i \geq p_{i+1} + kp_{i-1}$. Since $p_{i-1} > p_i$, it must be that $p_{i+2} \geq p_{i+1} \geq p_i$. Iterate the same argument we conclude $p_n \geq p_i$. However, by (3.5) again this would mean that

$$\alpha + k\beta p_{n-1} \geq \alpha + \beta p_{i+1} + k(\alpha + \beta p_{i-1}) > \alpha + k\beta p_{n-1}$$

A clear contradiction. Hence, such an $i$ does not exist and we conclude $p_1 > p_2 > \cdots > p_n$. □ □

Combining the two lemmas above, we get the following theorem.

**Theorem 3.2.1.** *Fix $k$ and $\beta$ such that $0 < k < 1$ and $0 \leq \beta < 1$. There is some $n \geq 1$ such that $\mathcal{C}_m$ is unstable for any $m \geq n$.*

*Proof.* Lemma 3.2.1 shows the statement holds when $\beta \leq k$ (where $n = 3$). Suppose $\beta > k$, by Lemma 3.2.2 we need to find $n$ such that $p_1 < p_2$ holds in $\mathcal{C}_n$. Iteratively applying (3.5), we get that

$$\begin{aligned} p_1 &= \alpha + \alpha\beta + \cdots + \alpha\beta^{n-2} + k(\beta(\alpha + \beta p_1) + \beta^2(\alpha + \beta p_2) + \\ &\quad \cdots \beta^{n-1}(\alpha + \beta p_{n-1})) \\ p_2 &= \alpha + \alpha\beta + \cdots + \alpha\beta^{n-3} + k((\alpha + \beta p_1) + \beta(\alpha + \beta p_2) + \\ &\quad \cdots \beta^{n-2}(\alpha + \beta p_{n-1})) \end{aligned}$$

Subtracting the first equation with the second, we get

$$\begin{aligned} p_1 - p_2 &= \alpha\beta^{n-2} - \alpha k \left(1+\beta+\beta^2+\cdots+\beta^{n-2}\right)(1-\beta) - \\ &\quad \beta k \left(p_1 + \beta p_2 + \beta^2 p_3 + \cdots + \beta^{n-2} p_{n-1}\right)(1-\beta) \\ &= \alpha\beta^{n-2} - \alpha k \frac{1-\beta^{n-1}}{1-\beta}(1-\beta) - (1-\beta)\beta k \sum_{i=0}^{n-2} \beta^i p_{i+1} \\ &= \alpha\beta^{n-2} - \alpha k (1-\beta^{n-1}) - (1-\beta)\beta k \sum_{i=0}^{n-2} \beta^i p_{i+1} \end{aligned}$$

Since $0 \leq \beta < 1$, $p_1 < p_2$ if $\alpha\beta^{n-2} \leq \alpha k(1-\beta^{n-1})$. We solve this inequality and get

$$n \geq \left\lceil \log_\beta \frac{k}{1+k\beta} \right\rceil + 2$$

Thus, the theorem is proved. □ □

Theorem 3.2.1 justifies that the chain networks are not suitable for organizations from the point of view of stability: The network will become unstable as the number of people (and thus levels) increases.

**Remark 3.** The above example also provides a mathematical explanation for the use of backflows in the model. Recall that a backflow represents the reciprocal interaction effect from a subordinate to the supervisor in a formal relation, which means that the subordinate is empowered by the supervisor through support and privilege, and it is given a weight of $k$ in the weighted interaction graph. If such weight is not given, then the adjacency matrix $R$ of the weighted interaction graph of a chain network will be $R_{ij} = 1$ if $j = i-1$ and 0 otherwise. The corresponding Bonacich power vector will be:

$$\vec{p} = \begin{pmatrix} \alpha + \alpha\beta + \cdots + \alpha\beta^{n-3} + \alpha\beta^{n-2} \\ \alpha + \alpha\beta + \cdots + \alpha\beta^{n-3} \\ \cdots \\ \alpha + \alpha\beta \\ \alpha \\ 0 \end{pmatrix}$$

In this case, the power is strictly decreasing from the top of the chain to the bottom, and the structure will remain stable regardless of the size of the chain, which does not meet with our intuition. Thus, the weight added to backflows in our model is necessary.

### 3.2.2 Perfect Tree Networks

With a similar but more involved technical analysis, we can generalize Theorem 3.2.1 to perfect tree networks.

**Definition 3.2.1.** *Fix $d > 1$. A* perfect $d$-ary tree network *is an organizational network where the formal ties $E_{\mathsf{fml}}$ form a tree in which every non-leaf node has exactly $d$ children and all leaves are at the same level in the tree. We use $\mathcal{D}_h^d$ to denote a perfect $d$-ary tree network of height $h$. The number $d$ is called the* arity *of the tree.*

Note that a unary perfect tree is simply a chain network. The arity $d$ in the perfect tree network equals to the capacity $c$ minus one, and therefore we get $d\beta < 1$ by our earlier assumption (3.3) that $\beta < \frac{1}{c-1}$. Similarly to (3.5), the power of node $i$ is computed by

$$p_i = \begin{cases} d(\alpha + \beta p_2) & \text{if } i = 1 \\ k(\alpha + \beta p_{h-1}) & \text{if } i = h \\ d(\alpha + \beta p_{i+1}) + (\alpha + \beta p_{i-1})k & \text{if } 2 \leq i \leq h-1 \end{cases} \tag{3.6}$$

The following is a lemma that generalizes Lemma 3.2.1.

**Lemma 3.2.3.** *If $\beta \leq \frac{k}{d^2}$, then any perfect $d$-ary tree network $\mathcal{D}_h^d$, with $d \geq 1$ and height $h > 2$, is unstable.*

*Proof.* By (3.6) we derive the following equations:

$$p_1 = \frac{d\alpha + d^2\alpha\beta + d\alpha\beta k + d^2\beta^2 p_3}{1 - d\beta^2 k}$$

and

$$p_2 = \frac{d\alpha + \alpha k + d\alpha\beta k + d\beta p_3}{1 - d\beta^2 k}$$

Combining the above we get

$$p_1 - p_2 = \frac{d^2\alpha\beta + d^2\beta^2 p_3 - \alpha k - d\beta p_3}{1 - d\beta^2 k}$$

By assumption we get $d\beta < 1$. Thus $1 - d\beta^2 k > 0$ for any positive $k < 1$. Therefore, $p_1 - p_2$ is negative whenever

$$d^2\beta - k < \frac{d\beta p_3(1 - d\beta)}{\alpha}$$

This clearly holds for any $\beta \leq \frac{k}{d^2}$ □ □

The next theorem generalizes Theorem 3.2.1 to $d$-ary perfect trees. Lemma 3.2.3 handles the case when $\beta \leq \frac{k}{d^2}$. The case when $\beta > \frac{k}{d^2}$ can be proved similarly to Theorem 3.2.1.

**Theorem 3.2.2.** *For any arity $d \geq 1$, there is a constant $c_d \in \mathbb{R}$ such that any perfect tree network $\mathcal{D}_h^d$ is unstable if*

$$n \geq c_d + \log_{d\beta}(1/d), \tag{3.7}$$

*where an upper bound for the constant $c_d$ is defined as $c_d \leq \log_{d\beta} k/(1 + k\beta) + 2$.*

*Proof.* Lemma 3.2.3 shows that the statement holds when $\beta \leq k/d^2$ (where $h \geq 3$). Suppose $\beta > k/d^2$. Iteratively applying (3.6), we get that:

$$\begin{aligned}
p_1 &= d\alpha + d\alpha(d\beta) + d\alpha(d\beta)^2 + \cdots + d\alpha(d\beta)^{h-2} + \\
&\qquad d\beta k(\alpha + \beta p_1) + (d\beta)^2 k(\alpha + \beta p_2) + \cdots + (d\beta)^{n-1} k(\alpha + \beta p_{h-1})) \\
&= d\alpha \sum_{j=0}^{h-2} (d\beta)^j + k \sum_{r=1}^{h-1} (d\beta)^r (\alpha + \beta p_r) \\
p_2 &= d\alpha + d\alpha(d\beta) + d\alpha(d\beta)^2 + \cdots + d\alpha(d\beta)^{h-3} + \\
&\qquad k(\alpha + \beta p_1) + (d\beta)k(\alpha + \beta p_2) + \cdots + (d\beta)^{h-2} k(\alpha + \beta p_{h-1})) \\
&= d\alpha \sum_{j=0}^{h-3} (d\beta)^j + k \sum_{r=1}^{h-1} (d\beta)^{r-1} (\alpha + \beta p_r)
\end{aligned}$$

Subtracting the first equation by the second, we obtain:

$$\begin{aligned} p_1 - p_2 &= d\alpha(d\beta)^{h-2} - \alpha k \left(1 - (d\beta)^{h-1}\right) - \\ &\qquad \beta k(1 - d\beta) \left(p_1 + d\beta p_2 + (d\beta)^2 p_3 + \cdots + (d\beta)^{h-2} p_{h-1}\right) \\ &= d\alpha(d\beta)^{h-2} - \alpha k \left(1 - (d\beta)^{h-1}\right) - (1 - d\beta)\beta k \sum_{i=0}^{h-2} (d\beta)^i p_{i+1} \end{aligned}$$

Since $0 \leq \beta < 1/d$ by our assumption, $p_1 < p_2$ if $d\alpha(d\beta)^{h-2} \leq \alpha k(1 - (d\beta)^{h-1})$. Solving this inequality we get:

$$h \geq \left\lceil \log_{d\beta} \frac{k}{1 + k\beta} \right\rceil + 2$$

Thus, the theorem is proved. □ □

**Remark 4.** The proof of Theorem3.2.2 gives us an upper bound for the constant $c_d$ which only depends on $d$:

$$c_d \leq \log_{d\beta} k/(1 + k\beta) + 2 \tag{3.8}$$

The inequality (3.8) provides a theoretical upper bound on the number of levels for a perfect $d$-ary tree to stay stable. Note that this bound may be much larger than the minimum value for such $c_d$. For example, using UCINET [17], we computed the actual limits on numbers of hierarchy levels with $k = 0.5$: for $d = 2$, it is 5; for $d = 3$, it is 8 (the theoretical bounds are 18 and 21, respectively.) Furthermore, by increasing the span of control (i.e., $d$) of nodes, the theorem implies a logarithmic growth on the bounds on the number of levels. Theoretical bounds for small values of $d$ can be found in Table 3.1.

We now interpret the main result (Theorem 3.2.2) of the section.

A general and significant organizational change trend in the last 50 years is the shift from *tall hierarchies* with many levels to *flat hierarchies*, where the number of levels is kept bounded. Research has found that most large companies changed their structures to the flattened ones in the past 3-4 decades [115], e.g., back in 1950s companies had up to twenty layers in their hierarchies while by the end of the twentieth century they had been trimmed to five or six. We conjecture that this delayering process implies

| span of control | $\beta < \frac{1}{d}$ | k =0.1 | k =0.5 | k =0.75 |
|---|---|---|---|---|
| 1 | 0.9 | 25 | 13 | 10 |
| 2 | 0.45 | 31 | 18 | 15 |
| 3 | 0.3 | 35 | 21 | 18 |
| 4 | 0.225 | 38 | 23 | 20 |
| 5 | 0.18 | 40 | 25 | 22 |
| 6 | 0.15 | 42 | 27 | 23 |
| 7 | 0.128571 | 43 | 28 | 25 |
| 8 | 0.1125 | 44 | 29 | 26 |
| 9 | 0.1 | 45 | 30 | 27 |
| 10 | 0.09 | 46 | 31 | 28 |

Table 3.1: Stable $d$-ary tree networks: theoretical bound on the number of layers computed as $n = \left\lfloor \log_{d\beta} \frac{k}{d(1+k\beta)} \right\rfloor + 2$

some fundamental truth regarding organizational networks. The well-known theory of "six degrees of separation" has been extensively studied and verified in the social network analysis community[84]. This theory states that six is a natural bound in the acquaintance relation on the distance between two people in the world.

Analogously, it seems that *for organizational networks, a bound on the number of levels of the hierarchy also exists.* Moreover, this upper bound is natural as it allows the top manager to maintain control over the hierarchy.

Theorem 3.2.2 provides an evidence of the existence of such a bound: *As the arity d is bounded (by capacity of individuals), the maximum height for a perfect tree network to maintain power consistency is bounded.*

## 3.3 Flattening – Workplace Democratization or Power Concentration?

*Flattening* (or delayering) is the phenomenon when an organization acquires a new structure by decreasing the number of hierarchy levels. It reflects a notable trend in organizational structure in the last half a century. Researches show that in last few decades most large companies changed their structures to the flattened ones. For instance, as it is shown in [115], the average number of those who reported directly to the CEOs in large companies was 4.7 in 1980, and 9.8 in 1999.

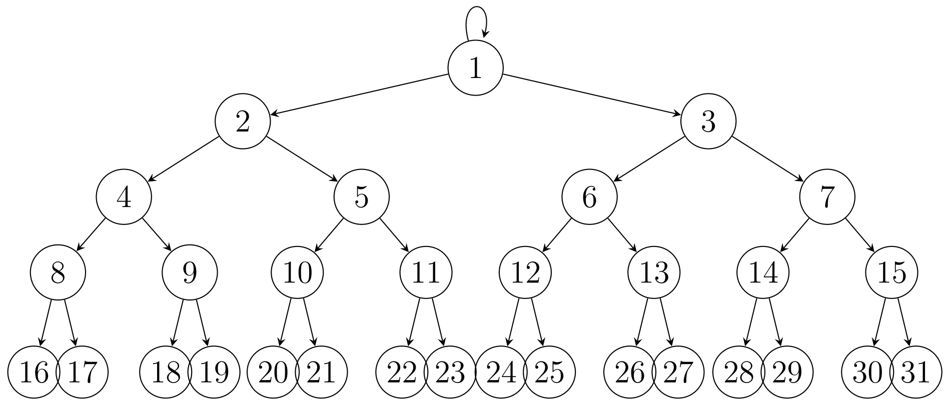



Figure 3.4: Network $\mathcal{A}$ with 31 nodes

The alleged reasons for flattening include empowering staff with decision making entitlement, increasing flexibility of employees, pushing down decision making, improving information flow and consequently enabling faster decision. In general, flattening is viewed as a strategy for democratization of the work place. However, some researchers also argue that flattening leads to the opposite effect – more control and decision making is concentrated on the top in the flattened organization, and hence it is also a strategy of strengthening controls by the top managers [106, 115].

In this section, we analyze the flattening process from the point of view of power. Based on our organizational network model, we argue that most employees indeed obtain more power through flattening, although the average power decreases. Moreover, the upper level managers are the ones whose power improves considerably.

**Example 2.** Consider a network $\mathcal{A}$ representing a perfect tree network without informal ties, such that there are 31 nodes and the capacity of each node is 3. Suppose that the span of control is exactly two for all nodes. This network can be represented as a perfect binary tree of height 5 as in Figure 3.4.

Suppose we increased the span of control of each node to 5, and the total number of nodes is kept the same, the network becomes a 5-ary perfect tree with height 3 (See Figure 3.5.)

We ran several tests computing the individual power with different parameters and list results in the Table 3.2, where $n$ is the number of nodes and $\ell$ is a hierarchy level. One can see that the average power in $\mathcal{A}$ is strictly greater than the average power in $\mathcal{B}$. Similarly, the figures for the case $k = 0.1$ shows that flattening negatively impact power of individuals in the network: only *four* nodes increase their power while 11

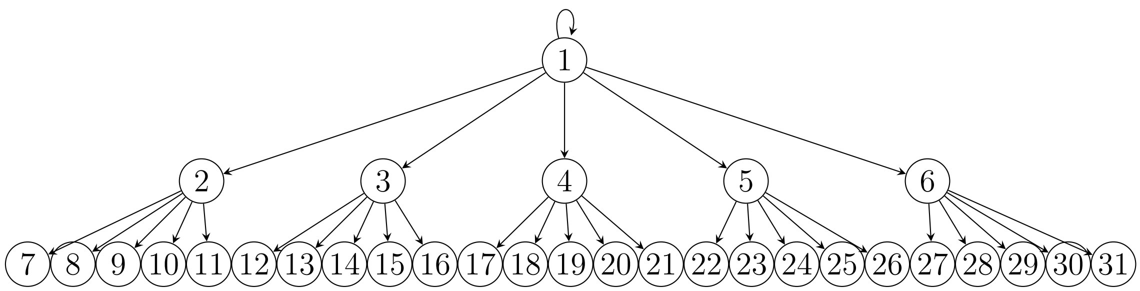



Figure 3.5: Network $\mathcal{B}$ with 31 nodes

others become less powerful and 16 stay the same. However, when we even slightly increase $k$ to 0.15, the majority of nodes increase their power. Moreover, when $k = 0.8$, the network $\mathcal{A}$ becomes unstable while $\mathcal{B}$ is still stable.

| | $\mathcal{A}$: $d = 2$, $\beta = 0.45$ | | | | |
|---|---|---|---|---|---|
| | $n$ | k=0.1 | k=0.15 | k=0.5 | k=0.8 |
| $l = 0$ | 1 | 2.613 | 2.553 | 2.193 | 1.949 |
| $l = 1$ | 2 | 2.175 | 2.181 | 2.181 | 2.149 |
| $l = 2$ | 4 | 1.522 | 1.527 | 1.556 | 1.587 |
| $l = 3$ | 8 | 0.819 | 0.828 | 0.864 | 0.882 |
| $l = 4$ | 16 | 0.070 | 0.100 | 0.252 | 0.323 |
| max | | 2.613 | 2.533 | 2.193 | 2.149 |
| min | | 0.070 | 0.100 | 0.252 | 0.323 |
| average | | 0.668 | 0.685 | 0.765 | 0.800 |
| | $\mathcal{B}$: $d = 5$, $\beta = 0.18$ | | | | |
| $l = 0$ | 1 | 3.503 | 3.450 | 3.056 | 2.733 |
| $l = 1$ | 5 | 1.929 | 1.941 | 1.966 | 1.936 |
| $l = 2$ | 25 | 0.070 | 0.103 | 0.306 | 0.437 |
| max | | 3.503 | 3.450 | 3.056 | 2.733 |
| min | | 0.070 | 0.103 | 0.306 | 0.437 |
| average | | 0.481 | 0.507 | 0.662 | 0.753 |

Table 3.2: Comparing individual power in networks $\mathcal{A}$ and $\mathcal{B}$ (tests performed using UCINET [17])

In [64], the authors carried out a survey in a company after introducing a new flat structure. The survey showed that 65.9% of employees were very happy, 26.3% were not happy, and 7.8% were not concerned about the change. This correlates very well with the results we obtain: the computation reveals that 64.5% (20 out of 31) of nodes when $k = 0.15$ become more powerful.

Through this example, we argue the following rather paradoxical aspect of flattening in an organizational network *(flattening paradox): Flattening decreases the average power in the company, but empowers most employees.*

The above displays a complicated relation between organizational power and structural properties of the formal tie hierarchy. To develop a better understanding of this relation, we perform a series of experiments.

**Experiment 1. Perfect tree hierarchies**

We first focus on perfect tree hierarchies. Here the goal is to investigate the distribution of power in perfect trees of different heights and arities.

Using CORPNET, we generate perfect tree hierarchies of various heights and set the parameter $k = 0.1$ and $\rho = 0$. The arity of the trees is set between 2 and 9, while the height $h$ is between 2 and 7. When the tree has more than three levels and the sum of the arity and height exceeds 9, the tree becomes too large for the software to handle. Thus, we only explore the results for the remaining cases. In particular, Table 3.3, Table 3.4 and Table 3.5 list the power of the roots of the trees, the average power of all nodes, and its variance, respectively.

| arity\height | 8 | 7 | 6 | 5 | 4 | 3 |
|---|---|---|---|---|---|---|
| 2 | 3.5684 | 3.2424 | 2.922 | 2.6126 | 2.317 | 2.031 |
| 3 | | 5.0269 | 4.4021 | 3.7703 | 3.1529 | 2.5671 |
| 4 | | | 5.7099 | 4.8204 | 3.9219 | 3.0557 |
| 5 | | | | 5.7602 | 4.6223 | 3.5034 |
| 6 | | | | | 5.2635 | 3.9172 |
| 7 | | | | | 5.8549 | 4.3025 |
| 8 | | | | | 6.4049 | 4.6639 |
| 9 | | | | | 6.9198 | 5.0049 |

Table 3.3: The Power of Top-level managers (roots) in perfect trees of varying heights and arities

Expectedly, when the arity is fixed, as the tree becomes taller, the root node gets more powerful; at the same time, the average power among all nodes drops, which means that the distribution of power becomes more uneven. This observation gives the impression that when we fix individuals' capability, power is distributed more evenly in flattened hierarchies than in taller hierarchies.

However, when taking into account possible changes in capability, the situation is rather different. As shown in the tables, if the height of the tree is fixed, as the arity

| arity\height | 8 | 7 | 6 | 5 | 4 | 3 |
|---|---|---|---|---|---|---|
| 2 | 0.5865 | 0.5663 | 0.5625 | 0.6503 | 0.6795 | 0.6883 |
| 3 | | 0.5620 | 0.5652 | 0.5710 | 0.5806 | 0.5923 |
| 4 | | | 0.5080 | 0.5109 | 0.5172 | 0.5277 |
| 5 | | | | 0.4679 | 0.4722 | 0.4807 |
| 6 | | | | | 0.4376 | 0.4445 |
| 7 | | | | | 0.4100 | 0.4156 |
| 8 | | | | | 0.3873 | 0.3918 |
| 9 | | | | | 0.3680 | 0.3726 |

Table 3.4: The average power of nodes in perfect trees of varying heights and arities

| arity\height | 8 | 7 | 6 | 5 | 4 | 3 |
|---|---|---|---|---|---|---|
| 2 | 0.5876 | 0.5840 | 0.5759 | 0.5538 | 0.5383 | 0.5263 |
| 3 | | 0.6841 | 0.6810 | 0.6740 | 0.6629 | 0.6492 |
| 4 | | | 0.7277 | 0.7052 | 0.6686 | 0.6561 |
| 5 | | | | 0.7811 | 0.7771 | 0.7690 |
| 6 | | | | | 0.8085 | 0.8024 |
| 7 | | | | | 0.8319 | 0.8273 |
| 8 | | | | | 0.8500 | 0.8465 |
| 9 | | | | | 0.8646 | 0.8613 |

Table 3.5: Variance of power of nodes in perfect trees of varying heights and arities

increases (that is, as people's span of control increases), the root becomes significantly more powerful, while the average power drops and variance increases.

We then plot the distribution of power across all levels of perfect tree hierarchies with arity ranging in 2,3,4 and heights $3 \leq h \leq 8$ (See Figure 3.6. )

In each plot, as the hierarchy flattens, the power of any level strictly decreases. At the mean time, when we consider the distance of nodes from the leaves, for almost any $k$, the $k$-th last level of the trees gains power slightly. When the arity increases, the difference between upper levels and lower levels becomes increasingly visible. One can interpret this intuitively: when managers get more subordinates, there is a wider power gap between an upper and a lower level.

**Experiment 2. Random tree hierarchies** Flattening may refer to two types of structural changes in an organization: The first type reduces the height of an organization by removing nodes, while not changing the capacity of its managers. An effect of this process may be conceptually revealed in the results of Experiment 1. The second type reduces the height of a hierarchy by improving the capacity of nodes, while

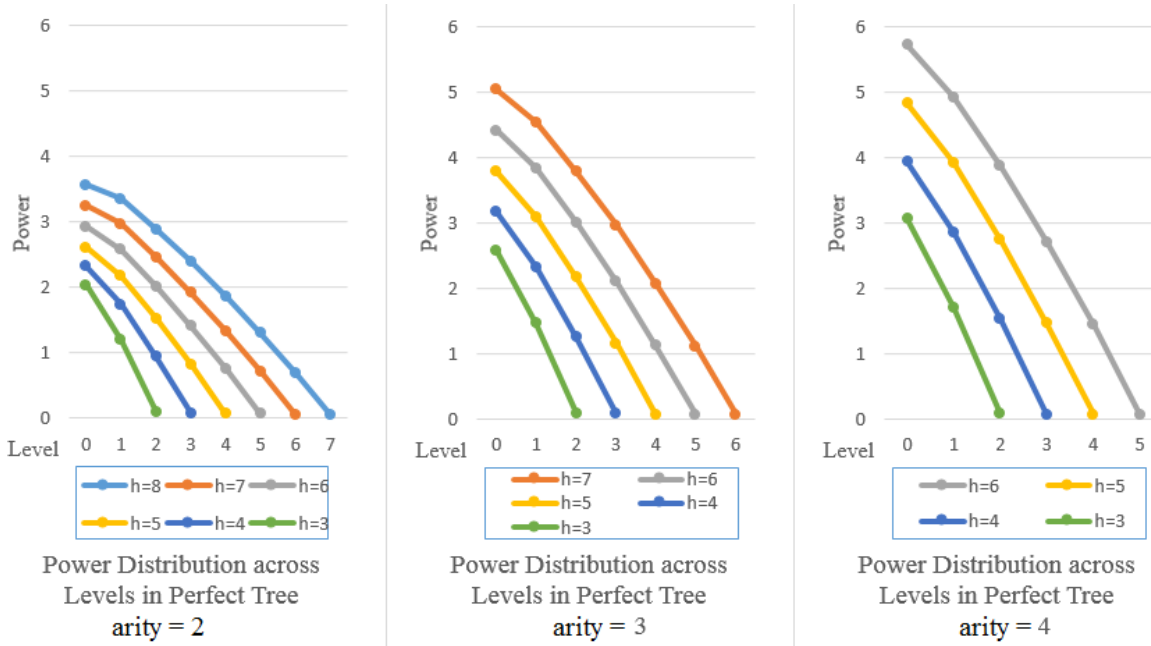



Figure 3.6: The power of three perfect tree networks with arities 2,3, and 4. The plots show the power across all levels of the hierarchies.

not changing too much the number of members of the organization. We now focus on this type of changes. Here perfect tree hierarchies no longer apply as they tend to have very different sizes when arities and heights differ.

For this experiment, we simulate formal tie hierarchies using a random tree model. Procedure 1 is a simple procedure that produces trees with a given height $h$, where the degrees of all internal nodes are taken from a normal distribution with mean $d$ and standard deviation.

---

**Procedure 1** RandomTree$(d, h, s)$ where $h \in \mathbb{Z}$, $d, s \in \mathbb{R}$

---

Initialize an empty tree and add a root
$\mathsf{lev} = 0$
**while** $\mathsf{lev} < h$ **do**
  **for** every node $u$ at level $\mathsf{lev}$ of the current tree **do**
    Randomly generate a number $m$ in a normal distribution with mean $d$ and standard deviation $s$
    Create $m$ children for the node $u$
  **end for**
**end while**
Return the constructed tree

---

We generate 10 random trees for each height between 3 and 7 while setting $s = 1$ and $k = 0.1$. We then compute the average power of nodes in each level and take the average over all trees of the same height. The details of the generated trees are listed in Table 3.6.

Starting from random trees with height 7 in the top row, we can imagine a flattening process that reduces the height of the hierarchy iteratively, while expanding the expected arity. Hence the first row corresponds to the tallest hierarchy where the last row corresponds to the flattest. The third column lists the average number of nodes of trees for each height; as the hierarchy flattens, the number of nodes reduces slightly.

| $h$ | $d$ | Avg size | Level 0 | 1 | 2 | 3 | 4 | 5 | 6 | Avg |
|---|---|---|---|---|---|---|---|---|---|---|
| 7 | 1.7 | 84 | 1.804 | 1.785 | 1.232 | 1.293 | 1.355 | 0.878 | 0.071 | 0.647 |
| 6 | 2 | 76.4 | 1.873 | 1.997 | 1.738 | 1.358 | 0.953 | 0.070 | | 0.647 |
| 5 | 2.5 | 67.2 | 2.279 | 2.106 | 1.649 | 1.247 | 0.071 | | | 0.611 |
| 4 | 3.5 | 65.4 | 2.930 | 2.400 | 1.567 | 0.067 | | | | 0.536 |
| 3 | 7.5 | 65.2 | 4.314 | 2.445 | 0.056 | | | | | 0.399 |

Table 3.6: Random trees generated by Procedure 1 with height $h$ and expected arity $d$. The third column shows the average number of nodes of the generated trees. The subsequent columns show the average power of nodes across all levels. The last column shows the average power of all nodes for height $h$.

We then plot the average power of nodes across all levels for each height. Each curve shows changes to the average power of nodes at a particular level as the organization flattens. We also illustrate the change of power in the last level, and in the second-to-last level (See Figure 3.7.)

From these plots, one can identify the following pattern: As the hierarchy is flattened, the root (level 0) gains the most power, while the other levels tend to lose power the closer we get to the last level. In particular, the leaves (operatives) lose power as a result of flattening, while levels that are above the last level tend to gain power. Overall, this flattening process reduces the average power in the hierarchy.

These results are consistent with our observation in Example 2, where flattening provides upper levels of the hierarchy with more power. In general, flattening empowers managers in the organization, and therefore a large number of individuals would prefer a flat hierarchy to a taller one. However, the process also reduces the power of the operatives.

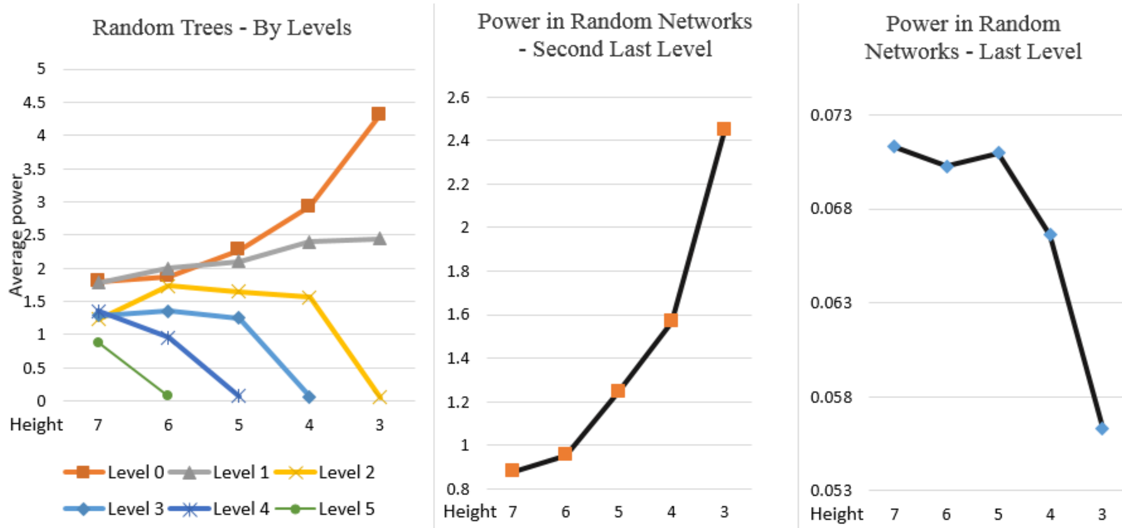



Figure 3.7: The average power of nodes across all levels in different random trees as the hierarchy flattens (left). Each curve corresponds to a particular level, and the horizontal axis refers to the different trees. Average power of nodes in the last level of these trees (right). Average power of nodes in the second-to-last level (center).

## 3.4 Understanding Informal Ties

Our analysis so far discusses only formal tie hierarchies, where no undirected informal tie is maintained. A long argument in management studies addresses the importance of informal social relations among members of organizations. Such relations, such as collaboration, advice or friendship, play important roles in the cohesion and effectiveness of the organizational structure [61, 32]. It is, therefore, crucial to incorporate informal ties into our analysis. In this section, we no longer assume that $E_{\mathsf{inf}} = \varnothing$ and aim to find how these informal ties affect the power of individuals.

### 3.4.1 A Benchmark for Organizational Networks

To correctly predict the impacts of informal ties on the network, it is imperative to adopt a reasonable benchmark for generating random social links. Popular benchmarks models such as planted $\ell$-partition, relaxed caveman graphs, and the LFR graphs [43] are not suitable for organizational networks as the informal ties generated by these models will be independent of the formal tie hierarchy of the organization. Naturally, the establishment of informal ties in an organizational network is significantly affected by the position of individuals in the reporting hierarchy.

*Homophily* is a recurring theme in social network studies which means that individuals have a natural tendency to bond with others that are similar to themselves. Numerous management studies also observe a kind of homophily in the workplace: Employees in an organization are more likely to establish social connections within a certain "circle", such as departments, offices. Moreover, individuals tend to establish personal ties with those who are at the same or similar levels in the hierarchy [76, 28]. No social network model so far has been defined taking into account this multiplex view of an organization. Hence, we provide a new benchmark graph; our view is that the new benchmark graph shall exhibit community structure that reflects the observed homophily phenomenon in an organization.

We adopt a distributed approach where each node randomly chooses to set up informal ties with other, in such a way that closer nodes (in distance) enjoy a higher "probability" of an informal tie. The procedure is described in Procedure 2.

**Procedure 2** RandomInformalTies($T, \gamma, p$) where $T = (V, E_{\mathsf{fml}})$ is a formal tie hierarchy, $\gamma \in \mathbb{N}$, $p \in [0, 1]$

Initial a set of undirected edges $E_{\mathsf{inf}} := \varnothing$
**for** Every node $u \in V$ **do**
  Compute the level $\ell(u)$ of $u$ in $T$
**end for**
**for** Every node $u \in V$ **do** ▷ Generate a probability distribution of all nodes in $V \setminus \{u\}$ by setting for any node $v \neq u$ a probability $\mathsf{Pr}_u(v)$
  **for** $v \in S_u$ **do**
    Compute the *lowest common ancestor* of $u$ and $v$, that is, a node $w$ such that $w$ is an ancestor of both $u$ and $v$, and $\ell(w)$ is maximal.
    Set $\Delta := \max\{\ell(u), \ell(v)\} - \ell(w)$
    Set $\mathsf{Pr}_u(v) := p^{\Delta}$
  **end for**
  Randomly select $\gamma$ nodes that are not $u$ based on the probability $\mathsf{Pr}_u$
  For each selected node $x$, add to $E_{\mathsf{inf}}$ an edge $(u, x)$ if it is not in $E_{\mathsf{inf}}$ already
**end for**
Return the constructed set $E_{\mathsf{inf}}$ of undirected edges

To generate a random organizational network with both formal and informal ties, we first apply a procedure that constructs a formal tie hierarchy $T = (V, E_{\mathsf{fml}})$, and then apply Procedure 2 with the parameters $T$, $\gamma$, $p$ to derive the set of informal ties. The resulting network not only captures main characteristics of social networks (such as community structure), but also entails reporting hierarchy of the network; see Figure 3.8 for a generated network visualized using a force-directed method. The community structure clearly resembles departments and reflect hierarchical levels in an organization.

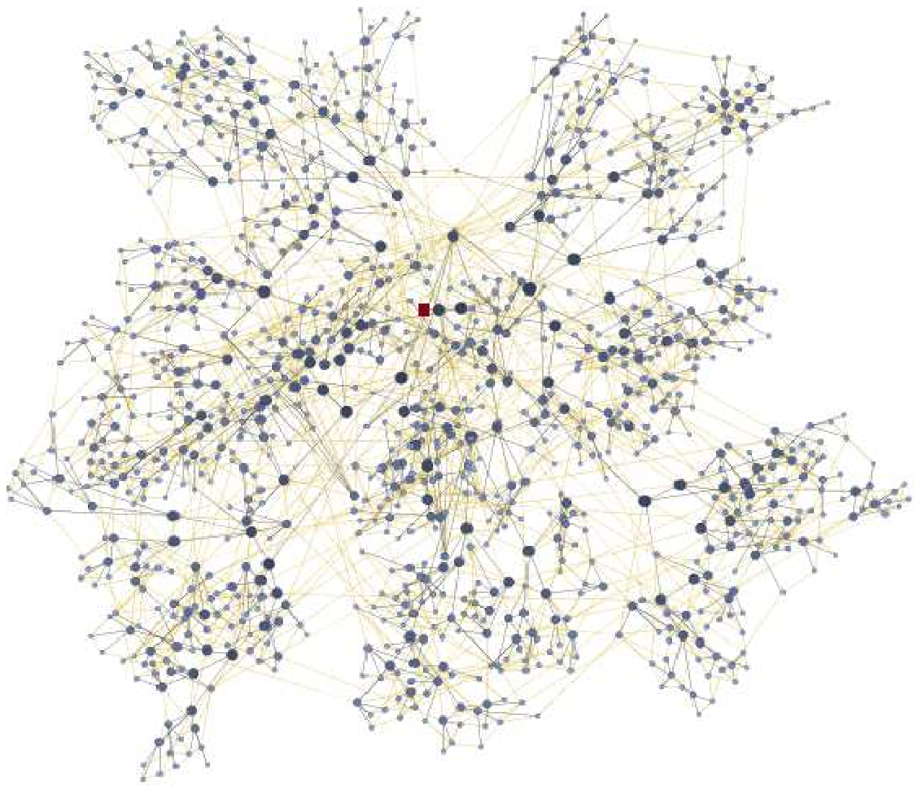


Figure 3.8: A randomly generated network for $d = 3$ and 7 levels. Blue and yellow lines are formal and informal ties, resp. The root is the brown square. Sizes of nodes indicates their power. The graph is generated and visualized by CORPNET.

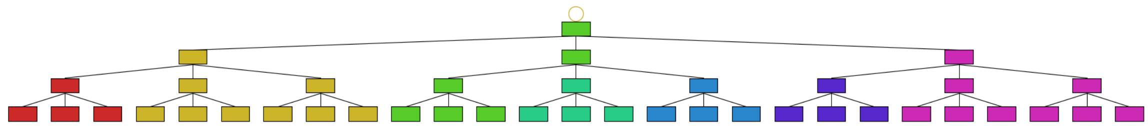
**Example 3.** To further validate our benchmark, in Figure 3.9(a), we consider a perfect 3-ary tree hierarchy with no social ties on it. We perform Newman's spectral graph clustering algorithm on this network and show the identified cluster (i.e. community) of each node, as indicated by its color [43]. In this case, we can see that clusters reflect departments, but not the levels of hierarchy. However, when we enrich this hierarchy with a generated social network, we get a quite different picture in Figure 3.9(b).

The resulting clustering clearly indicates the following pattern:

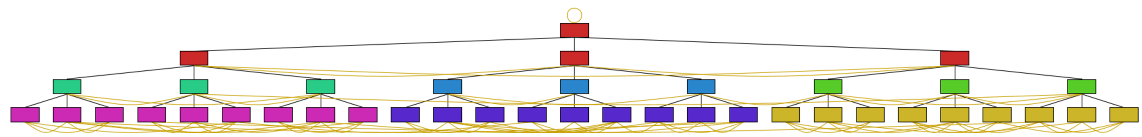
1. Clusters typically reflect departments: people in the same department tend to form a cluster.

2. Clusters also reveal levels: managers in the same level tend to form a cluster.

Clustering of the same nature is observed for another network, with randomly generated formal tie hierarchy using Procedure 1; See Figure 3.10.



(a) Clustering without informal ties



(b) Clustering with informal ties

Figure 3.9: Power of informal ties on community formations in organizations. Clusters are indicated by different colors. The graphs and their clustering are computed by CORPNET

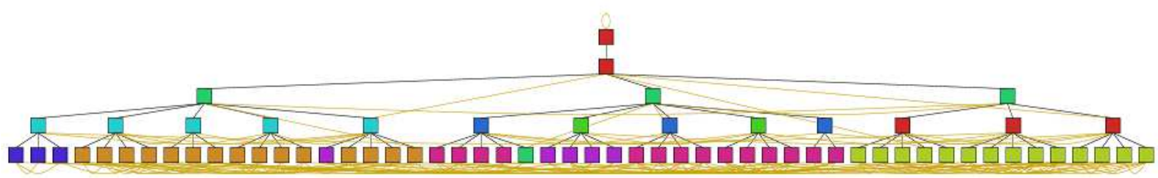


Figure 3.10: Random tree and random social network. The graph is generated and visualized by CORPNET

### 3.4.2 Importance of Informal Ties

We use two experiments to test how informal ties affect power in an organizational network. The first goal is to compare the power distribution before and after informal ties are introduced to a formal tie hierarchy. The second goal is to see how the formal tie hierarchy impacts power with the presence of informal ties. The third goal is to see how changing *density* of informal ties in the network affects structural properties and power. All experiments are carried out using CORPNET.

**Experiment 3. Stability and Informal Ties** We consider two perfect trees: one has the span of control $d = 3$ and 7 levels, the other one has the span of control $d = 10$ and 4 levels. The resulting values of individual power in both networks are listed in Table 3.7. Note that both hierarchies are stable. We then generate random informal ties with $\gamma = 8$ and different parameter $p$ over the formal tie hierarchies.

In Figure 3.11, we plot the distribution of average values of power at each hi-

| $\beta$ | 0.3 | | 0.07 | |
|---|---|---|---|---|
| $d$ | 3 | | 10 | |
| $n$ | 1093 | | 1111 | |
| Level: | k = 0.1 | k =0.5 | k =0.1 | k=0.5 |
| 0 | 5.03 | 4.8 | 6.16 | 5.66 |
| 1 | 4.53 | 4.79 | 4.86 | 4.7 |
| 2 | 3.77 | 3.9 | 2.89 | 2.85 |
| 3 | 2.95 | 2.91 | 0.05 | 0.22 |
| 4 | 2.06 | 1.96 | - | - |
| 5 | 1.1 | 1.07 | - | - |
| 6 | 0.06 | 0.24 | - | - |

Table 3.7: Power Distribution in Two Perfect Tree Hierarchies

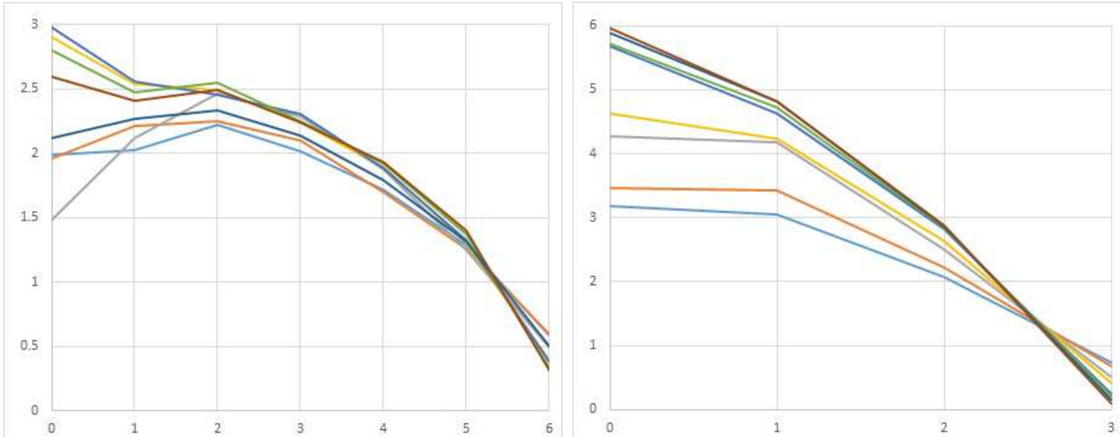
erarchy level in eight randomly generated social networks over the tall and the flat organizations. In Figure 3.11(a), one may see that only two out of eight generated networks are stable. However, as shown in Figure 3.11(b), in the flat organizations the non-reporting relations does not change the power distribution: all the networks stay stable.

As the result shows, the taller hierarchy's power consistency is very fragile – adding informal ties in all experiments makes the network unstable. On the other hand, the flattened hierarchy stays stable in most of our experiments with $k = 0.1$ and the probability $p = 0.5$ of existing friendship between two nodes which have the same direct manager. When the probability is small, corporate networks stay stable even with $k = 0.5$. Thus, this experiment justifies following: *As an organizational hierarchy has more levels, it is much more likely to be destabilized by non-reporting connections.*

**Experiment 4. Perfect Tree Networks with Social Ties**

We generate three perfect tree formal tie hierarchies. The first is a tall hierarchy with arity 2 and height 8; there are 255 nodes in the tree. The second is a flat hierarchy with arity 6 and height 4; there are 259 nodes in the tree. The third one is a hierarchy in between the previous two, with arity 4 and height 5; there are 341 nodes in the tree. We then generate informal ties by fixing $p = 0.5$ and varying parameter $\gamma \in \{0, 1, \ldots, 10\}$.

In Figure 3.12 we plot the average power of nodes and its variance in each network. It is clear that as more informal ties are introduced to the network the average power in all network increases, although this increase is more evident in the flattened



(a) Distribution of power in the tall organization (b) Distribution of power in the flat organization

Figure 3.11: Average values of power at each hierarchy level in randomly generated social networks: (a) in the tall organization, and (b) in the flat organization. The different lines indicate differences in "density" of the informal ties; in general, a denser informal relation causes a more even distribution of power across levels, hence a "flattened" (less-steep) curve.

.

networks and when there are fewer informal ties. On the other hand, the variance drops significantly as more informal ties are introduced, showing that power tends to be more evenly distributed with informal ties.

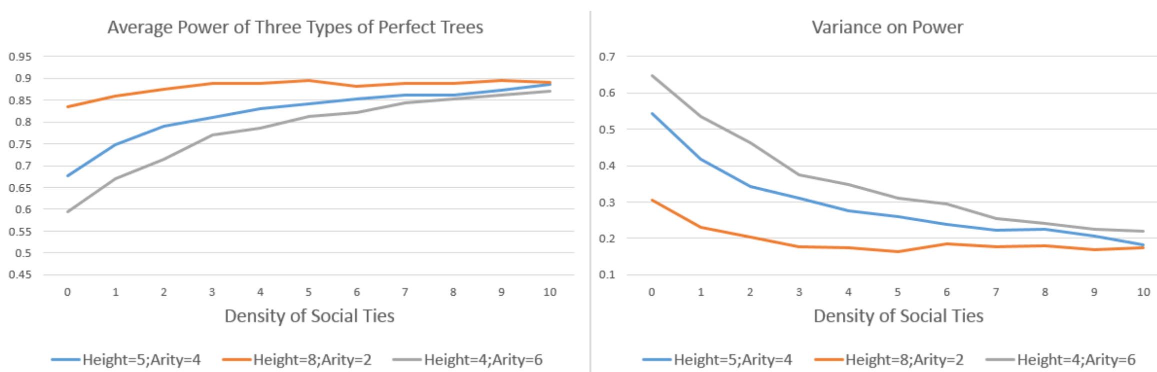



Figure 3.12: Power (left) and variance (right) for three types of perfect trees with increasing density of informal ties.

We next measure the level of stability in the network as more informal ties are added. To do that, we recall Definition 3.1.4 where a network is called *stable* if power in upper levels is consistently higher than power in lower levels. We thus call an internal node (a manager) $u$ *stable* if there is no node $v$ with a lower level than $u$ and whose power exceeds the power of $u$; otherwise, the internal node is called *unstable*.

Note that this definition only applies to managers (people with subordinates) of the network. In the following definition, we introduce a measure for the level of *instability* within a network.

**Definition 3.4.1.** *The* instability index $\iota$ *of the network is the proportion of unstable internal nodes among the set of all internal nodes in the formal tie hierarchy.*

Based on the definition above, in Figure 3.13 (left), we plot the instability index for all three networks as more informal ties are added. In all three formal tie hierarchies, instability increases as more informal ties are added. However, the tall hierarchy is especially unstable only after a small set of informal ties ($\gamma = 1$) is added, while in the flat hierarchy, the instability index is kept low (lower than 0.1) even when a large number of informal ties are added. This further confirms the finding in Experiment 3 about the differences between tall and flat hierarchies.

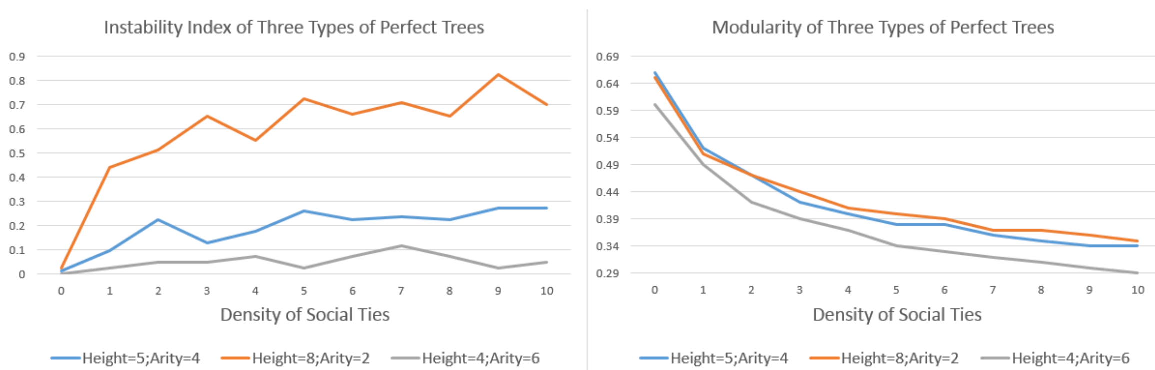



Figure 3.13: (left) The instability index for three types of perfect trees with informal ties. (right) The modularity for these networks.

We then apply Newman's spectral community detection algorithm to the networks with informal ties and compute *modularity* in each case. Modularity is a standard measure for a network and indicates how "clustered" a network is, i.e., how strong the network exhibits community structure with the detected communities [83]. Intuitively, the higher the modularity is, the more evident the clusters in the network become.

The experiments show that Newman's spectral algorithm produces roughly the same number of communities in each case (between 18 to 22). However, there has been a considerable variation regarding modularity as more informal ties are added. As the density increases, modularity significantly drops from above 0.6 to around 0.4 (for tall networks) and 0.3 (for flat networks). In a certain sense, this result captures the fact that all members of the organization form a more cohesive and unified team with more informal ties.

We further perform analysis on the distribution of power and show the results in Figure 3.14:

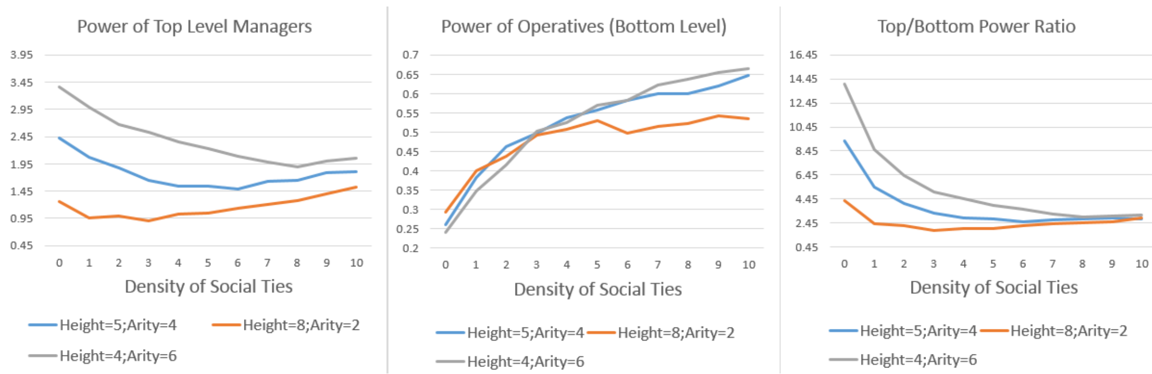



Figure 3.14: The results include three types of perfect trees with informal ties. The horizontal axis for all plots is the density of informal ties in the network. (left) The average power of the root. (center) The average power of the leaves. (right) The ratio between the average power of the root against the leaves.

- We plot the power of roots in the networks. The plot reveals that as the density of informal ties increases, the power of roots initially drops but then lifts up when more and more informal ties are present.

- We plot the average power of the leaves in the networks. The plot shows that in general more informal ties brings higher power to the operatives in the network, and the change can be dramatic, the power of operatives almost triples when the parameter $\gamma$ is changed from 0 to 10.

- We compute the ratio between the power of the root and the average power of leaves in the networks. In some sense, this ratio reflects the level of inequality of the network. In general, more informal ties makes the power among members more equal and the ratio drops. The most significant change occurs for flat networks, where without informal ties, this ratio reaches above 14, much higher than the other taller hierarchies. However, as more informal ties are added to the network, this ratio converges to about 3, which is very similar to the other hierarchies.

This experiment also gives us more insights towards the general phenomenon of flattening. Recall from Section 3.3, the experimental results show that flattening increases the power of the majority, but widens the gap between upper-level managers to lower levels. Here, our experiments show the importance of informal ties under this context: *One can significantly reduce the gap between power across upper and lower levels by enabling more informal ties in the organization.*

## 3.5 Leadership Styles: a Network Perspective

A leadership style is usually defined as a general approach acquired by the organization's leader into directing and performance management, i.e. how to set goals, implement plans, and motivate team members. A good leader makes sure the whole team is working towards a common goal, delivers outcomes and develop a healthy working atmosphere within the organization. Thus, the cohesion and productivity of an organization often hinge on the adoption of effective leadership styles by its top managers [97].

Traditionally, management studies focus on a behavioral perspective of leadership styles and analyze important traits of good leaders. Based on traits of cognitive, social and psychological factors, management theorists and practitioners typically classify leadership styles into several well-established categories, such as autocratic, paternalistic/consultive and democratic styles [48]. In this section, our goal is to provide an alternative, structural angle to the categorization and analysis of management styles. We start with a detailed discussion of the factors affecting the distribution of power.

**Experiment 5.** We carry out this experiment based on the 4-ary perfect tree hierarchy with height 5. The hierarchy is "standard" in the sense that it is neither tall (with a low span of control) nor flat (with a large span of control) and therefore should capture an idealized typically formal tie hierarchy. We generate random informal ties in the hierarchy using the parameters $\gamma = 5$ and $p = 0.5$. Two important factors influence power in this network: the value of the parameter $k$, which measures the weight of informal ties as compared to formal ties, and the value of $\rho$, which measures the *self-assertiveness* of the root of the hierarchy. Our goal for this experiment is to see how the combination of $k$ and $\rho$ affects the distribution of power.

We take $k \in \{0.1, \ldots, 1\}$ and $\rho \in \{0.1, \ldots, 1\}$ and generate 10 networks for each combination of $k$ and $\rho$ values. We then calculate using CORPNET the average power of nodes for each combination of $k$ and $\rho$ and plot them in Figure 3.15. As $k$ increases, the average power in the network also increases; on the other hand, a higher value of $\gamma$ leads to a decrease in average power, although the difference is minor.

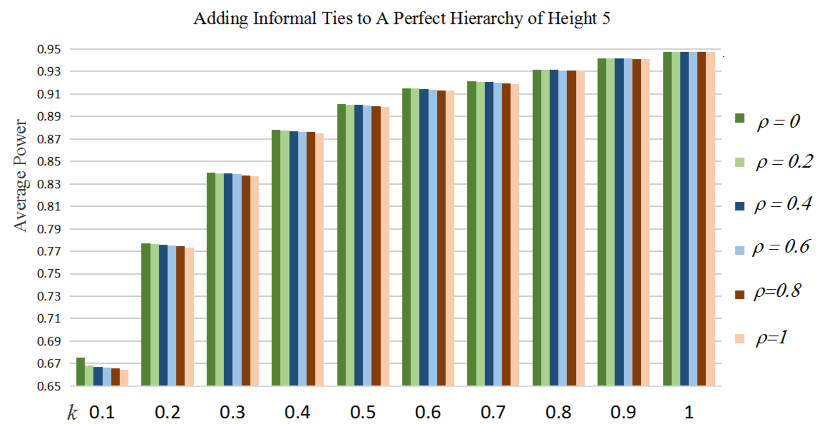



Figure 3.15: Average power of perfect tree of arity of 4 and height 5 with random informal ties

We then plot more statistics in Figure 3.16. Firstly, as shown in the plot on the left, the power of the root of the network decreases linearly on the increase in $k$, and increases with respective to the value of $\rho$. Secondly, the variance of the distribution of power drops exponentially with increasing $k$. Thirdly, the ratio of power between the root and leaves also drops exponentially with increasing $k$. Combined, these facts indicate that the power is distributed more evenly as $k$ increases.

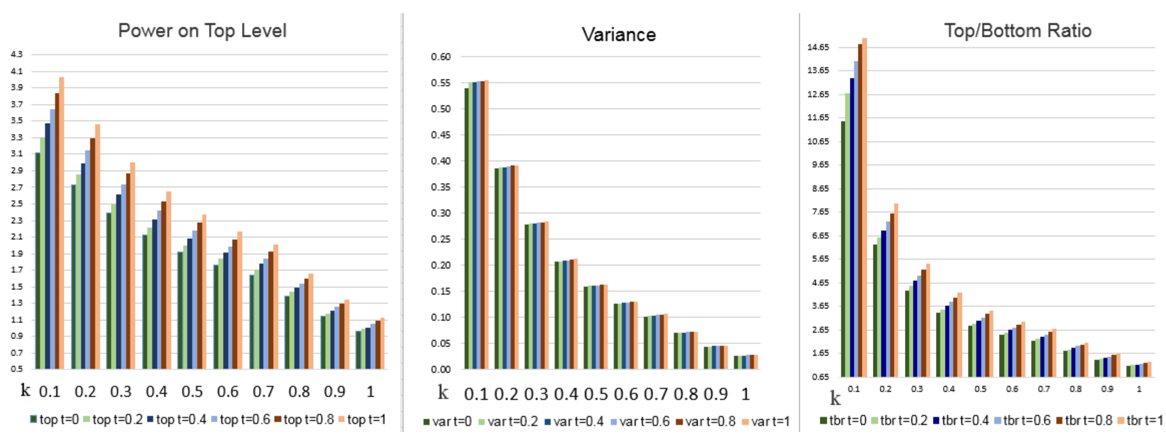



Figure 3.16: Perfect Tree of height 5 with informal ties. (left) The power of the roots. (center) The variance of power among all nodes. (right) The ratio between power of roots against leaves.

Finally, we plot the modularity and instability index of the community structure

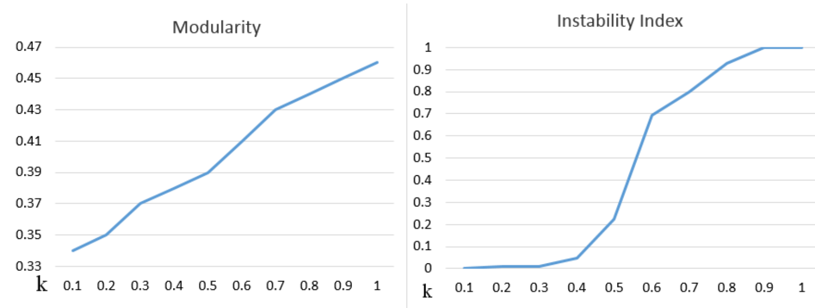



Figure 3.17: Perfect Tree of height 5 with informal ties. (left) Modularity of the identified clusters by Newman's spectral algorithm. (right) Instability index of the networks.

identified by Newman's spectral algorithm with increasing $k$ in Figure 3.17. Just like in Experiment 4, the algorithm identified 18-22 communities in all cases, and as $k$ increases from 0.1 to 1, the modularity increases between 0.34 and 0.46 in a linear fashion. This observation shows that the network tends to be more clustered as $k$ increases. On the other hand, as $k$ increases, the instability also gets larger. More interestingly, the instability index grows in different speed as $k$ grows and roughly separate into three stages: It grows very slowly (below 0.1) when $k$ is small (between 0.1 and 0.4), and then grows very fast when $k$ is between 0.4 and 0.75 before slowing down again for large k (above 0.8) but at a much higher value. We also obtain the same pattern regardless of the value of $\rho$. Thus in the plots, we do not modify the values of $\rho$.

The results of Experiment 5 demonstrate that the distribution of power is greatly influenced by the value $k$ and, at a much smaller scale, the value of $\rho$. These results suggest that a network-based approach to define and assess leadership styles is possible. In general, leadership styles in organizations are classified by the level of control exercised by the top managers. For example, managers in an autocratic organization make decisions unilaterally with no initiatives from the bottom while in a democratic organization, decisions are made by majority rather than by the top managers. The value of $k$ characterizes the amount of influence to the power of informal ties compared to formal ties; in this sense, $k$ can be regarded as an indicator of a level of control from higher to lower level of the reporting relation. Moreover, $\rho$ also intuitively characterizes the top manager's sense of self-determination when it comes to decision making and hence also affects leadership style.

We describe below some major classified leadership styles defined in management science [105]. We also interpret each style using a combination of parameters $k$ and $\rho$:

**Autocratic style.** This style assumes that the decisions are made from the top management unilaterally. There are few or no initiatives from the bottom of the hierarchy. The number of connections of any individual employee is relatively low because maintaining a reporting relation requires a lot of resources. Hence, we say that an organizational network (with a fixed weighted interaction graph) is *autocratic* if $\rho > 1$, $k$ is very small (i.e. within the range $[0, 0.1]$). The benefit of this style is high stability, while the negative effect is the lack of motivation of employees.

**Democratic style.** Here the decisions are made by the majority of the employees. There are many initiatives from the bottom of the hierarchy and collaboration requires as many resources to maintain as the reporting relations. Thus, we say that an organizational network is *democratic* if $\rho = 0$, $k \in [0.5, 1)$. The benefit of this style lies in job satisfaction and quality. However, it does mean a higher level of instability and inefficiency.

**Paternalistic (consultative) style.** This leadership style sits somewhere between autocratic and democratic styles. While the decisions are made mainly by the top managers, they take into account the best interests of the employees. The interaction is mainly one-directional (downwards), but feedbacks are encouraged. Hence, we define an organizational network to be *paternalistic* if $\rho > 0$, $k \in (0.1, 0.5)$.

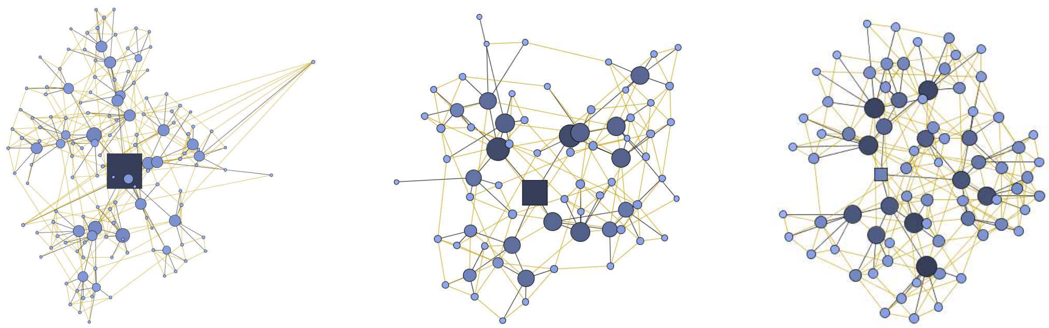
**Chaotic style.** This is a more recent style of management, which gives employees total control over decision making. Here, informal ties become the dominant personal links and thus require a larger weight as their effect and a large amount of resources to maintain. We define an organizational network to be *chaotic* if $\rho = 0$, $k = 1$. One would expect any chaotic organizational structure to be unstable.

**Example 4.** Based on the description above, we use CORPNET to generate organizational networks that capture each management styles above, by setting the parameters correspondingly. Figure 3.18 shows three typical networks with different management styles: autocratic, paternalistic, and democratic. The sizes of the nodes (drawn in

| **Leadership style** | $\rho$ | $k$ | Instability index |
|---|---|---|---|
| Autocratic | $\rho = 1$ | $k \in [0, 0.1]$ | low |
| Paternalistic | $\rho > 0$ | $k \in [0.1, 0.5)$ | low |
| Democratic | $\rho = 0$ | $k \in [0.5, 1)$ | high |
| Chaotic | $\rho = 0$ | $k \in 1$ | high |

Table 3.8: Organizational networks with different leadership styles

force-based layout) represent their power. One can clearly identify that the distribution of power in such networks is quite different, and reflect the reasoning above.



(a) A network with autocratic management styles (b) A network with paternalistic management styles (c) A network with democratic management styles

Figure 3.18: Distribution of power in networks with different management styles

**Experiment 6. Leadership styles**

We elaborate the discussion above by carrying out systematic experiments on random networks. Here once again we consider both a tall hierarchy (Height 7) and a flat hierarchy (Height 4). We generate random trees using Procedure 1 setting the mean arity of the trees to be 3 (for tall hierarchy) and 6 (for flat hierarchy). For each leadership style in Table 3.8, we generate ten networks of each type and compute average power at each level. We then plot the average power of each level in Figure 3.19.

In both organizations, autocratic style results in the largest variation of power across levels. The difference between a tall organization and a flat one is that the flat organization is stable under autocratic style whereas middle levels of the tall organization display much fluctuation. The other three styles, on the other hand, gives a much more even distribution of power across the levels. In the tall organization, the network is stable under the paternalistic style and becomes highly unstable under chaotic style.

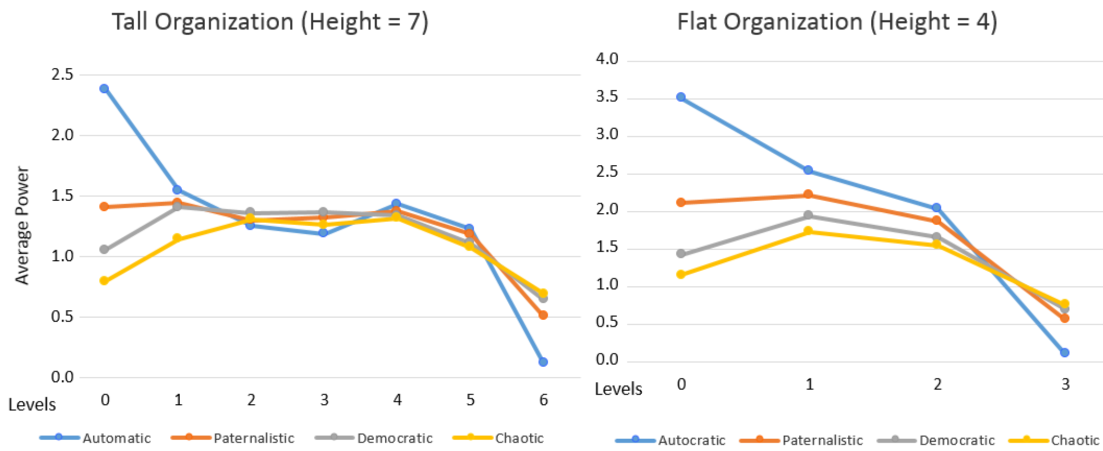



Figure 3.19: The distribution of average power across all levels in randomly generated networks. The networks consist of random formal tie hierarchy and random informal ties. Power in random tall organizations (left) where the formal tie hierarchy has height 7 and mean arity 3. Power in random flat organizations (right) where the formal tie hierarchy has height 4 and mean arity 6.

## 3.6 Case Study: Krackhardt and Hanson's Network

Krackhardt and Hanson in [61] studied a high-tech company with 21 managers. They analyzed the reporting hierarchy in the company, as well as reconstructed two types of social links on the same group of employees through a series of interviews – one type of social link is the advice relation (based on the interview question "*To whom do you go for advice?*") and the other is friendship (based on the question "*Who are your friends?*"). This data provides a real world case study for testing our model. In [61], the friendship links are directed; to fit our model we make them undirected by keeping only mutual friendship connections.

The reporting hierarchy of the network is depicted in Figure 3.20: there is one top manager (7), four departments, managed by 2, 14, 17, and 21, respectively.

We considered separately a reporting hierarchy and a "hybrid" organizational network that contains both formal and informal ties. We applied our power based approach to the both cases, and obtained the results that are listed in Table 3.9.

From the results, we draw two main conclusions:

(1) There is no correlation between power and the age, nor years of service of employees.

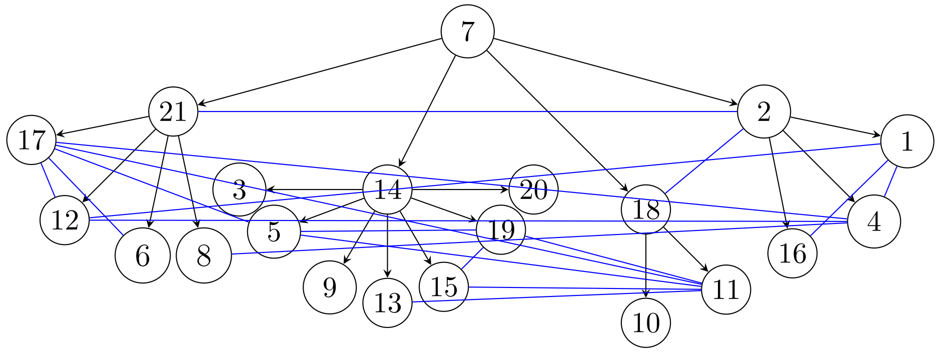



Figure 3.20: Krackhardt and Hanson's hierarchy with 21 nodes.

| | Attribute | | | Hierarchy | | | Hybrid | | |
|---|---|---|---|---|---|---|---|---|---|
| ID | Dept | Age | YoS | $k = 0.1$ | 0.5 | 0.75 | $k = 0.1$ | 0.5 | 0.75 |
| 1 | 4 | 33 | 9 | 0.06 | 0.25 | 0.34 | 0.18 | 0.68 | 0.84 |
| 2 | 4 | 42 | 20 | 1.35 | 1.41 | 1.42 | 1.43 | 1.55 | 1.48 |
| 3 | 2 | 40 | 13 | 0.07 | 0.32 | 0.45 | 0.11 | 0.41 | 0.5 |
| 4 | 4 | 33 | 8 | 0.06 | 0.25 | 0.34 | 0.22 | 0.85 | 1.06 |
| 5 | 2 | 32 | 3 | 0.07 | 0.32 | 0.45 | 0.2 | 0.77 | 0.97 |
| 6 | 1 | 59 | 28 | 0.06 | 0.27 | 0.37 | 0.1 | 0.39 | 0.49 |
| 7 | - | 55 | 30 | 2.39 | 2.13 | 1.96 | 2.33 | 1.67 | 1.33 |
| 8 | 1 | 34 | 11 | 0.06 | 0.27 | 0.37 | 0.1 | 0.38 | 0.47 |
| 9 | 2 | 62 | 5 | 0.07 | 0.32 | 0.45 | 0.07 | 0.24 | 0.28 |
| 10 | 3 | 37 | 9 | 0.05 | 0.23 | 0.31 | 0.05 | 0.18 | 0.21 |
| 11 | 3 | 46 | 27 | 0.05 | 0.23 | 0.31 | 0.26 | 1 | 1.24 |
| 12 | 1 | 34 | 9 | 0.06 | 0.27 | 0.37 | 0.19 | 0.72 | 0.9 |
| 13 | 2 | 48 | 0 | 0.07 | 0.32 | 0.45 | 0.11 | 0.42 | 0.51 |
| 14 | 2 | 43 | 10 | 3.06 | 2.98 | 2.88 | 2.99 | 2.34 | 1.96 |
| 15 | 2 | 40 | 8 | 0.07 | 0.32 | 0.45 | 0.16 | 0.59 | 0.73 |
| 16 | 4 | 27 | 5 | 0.06 | 0.25 | 0.34 | 0.1 | 0.36 | 0.45 |
| 17 | 1 | 30 | 12 | 0.06 | 0.27 | 0.37 | 0.27 | 1.03 | 1.29 |
| 18 | 3 | 33 | 9 | 0.92 | 1.03 | 1.07 | 0.96 | 1.04 | 1 |
| 19 | 2 | 32 | 5 | 0.07 | 0.32 | 0.45 | 0.24 | 0.89 | 1.09 |
| 20 | 2 | 38 | 12 | 0.07 | 0.32 | 0.45 | 0.07 | 0.24 | 0.28 |
| 21 | 1 | 36 | 13 | 1.77 | 1.8 | 1.78 | 1.9 | 1.68 | 1.53 |

Table 3.9: Power in Krackhardt and Hanson's network, $\beta = 0.1$

(2) By taking into consideration the informal ties, the power of individuals on the bottom (leaves) increases while those on higher levels lose some of their power.

Note further that this network is unstable by our definition as 14 has more power than 7 in all the cases. We suggest two possible ways to interpret this fact:

- A high power of a manager may suggest high capability and performance, as well as a high workload. This could be used as a rigorous basis for certain rewards to the particular employee in the form of, for instance, bonuses or promotion. Such bonuses would increase the loyalty of the employee, and, as a result, decrease possible risks.

- The node 14 is overwhelmed, as it has too many direct subordinates. To reduce this number and, therefore, to "stabilize" the structure, certain structural changes can be done. One of the possible solutions is to promote two of 14's most powerful direct subordinates (5 and 19) and distribute the rest (3, 9, 13, 15, 20) between them.

## 3.7 CORPNET: an ONA Tool

An ONA tool is a software that provides analytics the network structures within a company. It should reveal information flows, identify potential structural holes, gaining insights into properties that are invisible at first sight. Such insights can then be used to derive beneficial business strategies such as restructuring or promotion/demotion of staffs.

There are various existing ONA tools, examples of which include InFlow[2], SYNAPP[3], and SYNDIO[4]. These software tools usually perform data visualization tasks, as well as extracting standard network measures. However, two significant limitations exist:

1) Such products are mostly commercially available which made them difficult to be adopted for management science research, and,

2) Most importantly, they do not study correlations between formal and informal structures.

The aim of CORPNET is a stand-alone software application created to perform ONA functions based on our model above. CORPNET provides interactive simulation, analysis, and visualization functionalities. It is developed using the Scala programming language and runs on the Java Virtual Machine[5].

[2]Retrieved from `http://orgnet.com/inflow3.html`
[3]Retrieved from `http://www.seeyournetwork.com/`
[4]Retrieved from `https://synd.io/`
[5]A prototype of CORPNET and its source code can be downloaded from `https://github.com/mourednik/corpnet`

The main features of CORPNET include instruments of statistical and stability analysis, community detection, and functions to generate random formal tie hierarchies and social networks:

***Network creation and visualization*** The network graph can be visualized as either a *tree layout* or a *force directed layout*. The tree layout is hierarchical with respect to the directed edges only. The color of a node can represent either its power or its membership within a detected community; See Figure 3.21.

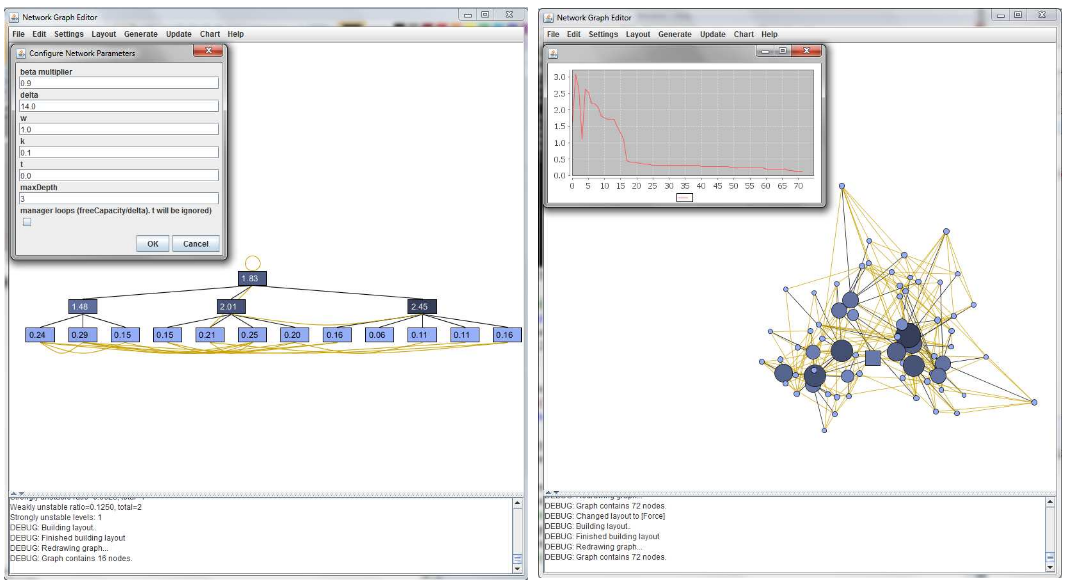



Figure 3.21: CORPNET user interface: a tree layout (left) and a force directed layout with a power distribution (right)

***Power analysis.*** CORPNET computes Bonacich power as defined above and visualizes power in multiple ways. For both layout styles, a darker shade of blue indicates a higher power within the network. The force layout draws nodes with varying sizes, such that more powerful nodes are relatively larger. Two plots are available providing statistical information regarding individual power:

- *Descending power grouped by level*: This is a scatter plot of individual node powers. The nodes are arranged from left to right in descending order of power, grouped by level such that the nodes on higher levels are to the left of nodes on lower levels.
- *Power histogram*: This is a histogram of node powers with a configurable number of bins.

See Figure 3.22 for examples of both types of plots.

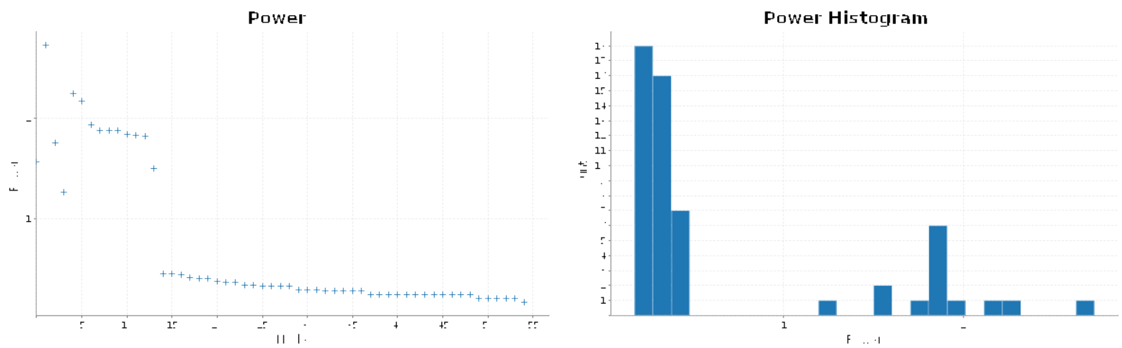



Figure 3.22: Power distribution plots: Descending power grouped by level (left) and power histogram (right)

***Community detection.*** CORPNET incorporates a module which computes clusters in the network based on formal and informal ties using Newman's spectral algorithm [43]. This is visualized by node colors which represent the detected clusters. See Figure 3.9 and Figure 3.10 for examples of clusterings in the tree layout.

***Random network generation.*** To facilitate experiments on organizational networks, CORPNET incorporates a network simulator which is able to generate synthetic benchmark organizational networks.

1. *Random tree generator.* This module generates tree using Procedure 1. The number of children is sampled from a normal distribution with a specified mean and standard deviation.

2. *Social network generator.* This module generates a benchmark social network of undirected edges over the existing hierarchical network, as described in Procedure 2.

# Chapter 4

# Integrating Homogeneous Networks

This chapter approaches the challenge of bringing two homogeneous networks together. To motivate our formal framework, we make two main assumptions. The first one is that creating weak ties between the networks can be encouraged and forced; the second is that structural properties, such as distance, provide a measure of effective communication and resource accessibility.

The first condition arises from the nature of interpersonal relations. Social networks are usually the result of complex interactions among autonomous individuals whose relationships cannot be simply controlled and forced. For example, in business networks, although working relations might be clearly prescribed, a firm is seldom in control of informal relationships (especially strong ties) among its employees [95]. Nevertheless, a company can prepare the ground for future weak ties: conferences and meetings, group assignments, special promotions etc. can be instruments of bringing people together.

The second condition discusses how the integrated network provides members with appropriate access to resource and information. When building new links between members of two networks, the crucial question is how to make the the combined network a unified whole. It is then a major question how "together" the unified networks should be as an outcome of the integrating process. Naturally, the more links there are that connect two networks, the closer they become. On the other hand, there is normally a cost associated with establishing and maintaining links. Therefore, it is important to strike a balance between the amount of *togetherness* and the number of new links created between the networks.

*Distance* is an important factor of information dissemination in a network [65]: a network with a small diameter means that members are, in general, close to each other and information could be passed from one person to any others within a small number of steps [110]. This argument has been used to explain how small-world property – the property that any node is reachable from others via only a few hops – becomes a common feature of most real-world social networks [3]. We hold the view that all nodes of a network have certain resources; and when a network has a small diameter, the resource on each node can be reached out from everyone else within a few steps, and each member is able to influence others. Hence, the diameter of the integrated network forms the strongest form of togetherness.

When expressing togetherness between two networks in their integration, diameter may be too strong. We define further two weaker notions of togetherness. Firstly, *existential togetherness* considers distances between every node in one network to some node in the other network. This measure is reasonable if we assume all nodes in any network hold the same resource, and it is enough to reach any node in a network. Secondly, *universal togetherness* considers distances between every node in one network to all nodes in the other network, which implies that the distances to all nodes in the other network are important.

Bearing the concepts of togetherness and the conditions in mind, we define *network integration* as the process when one or more edges are established across two existing networks in such a way that the integrated network has a bounded value of togetherness. Furthermore, a new edge always costs effort and time to establish and maintain. To minimize the number of new edges during the integration process, one needs to carefully choose which nodes to connect. Hence, the integration problem is comprised of two parameters, the value of togetherness and the number of edges. An additional challenge could be posed by assigning priorities to nodes that define which individuals should be connected first. We formally define corresponding *network integration problems*, study their complexity, propose several algorithms to solve them and perform numerous experiments to test these heuristics.

## 4.1 Togetherness and Network Integration

In this chapter, we view a *network* as a connected undirected unweighted graph $G = (V, E)$ where $V$ is a set of nodes and $E$ is a set of undirected edges on $V$.

For two disjoint sets of nodes $V_1, V_2$, we use $V_1 \otimes V_2$ to denote the set of all edges $\{uv \mid u \in V_1, v \in V_2\}$; these edges will be our instruments for integration two networks with nodes $V_1$ and $V_2$, respectively.

**Definition 4.1.1.** *Let $G_1 = (V_1, E_1)$ and $G_2 = (V_2, E_2)$ be two networks. Fix a non-empty set of edges $E \subseteq V_1 \otimes V_2$. The* integrated network *of $G_1$ and $G_2$ by $E$ is $G_1 \oplus_E G_2 := (V_1 \cup V_2, E_1 \cup E_2 \cup E)$.*

### 4.1.1 Three Levels of Togetherness on Integrated Networks

Consider the integration $G_1 \oplus_E G_2$. Any edge $uv \in E$ represents a channel for the flow of certain resources (information, traffic, knowledge, etc.) between $G_1$ and $G_2$. Hence, the set $E$ should provide nodes in each network with appropriate access to resources in the other network.

*Togetherness* is an index for proximity of $G_1$ and $G_2$, and thus measures the effectiveness of $E$. As argued above, distances between nodes play a significant role. Further, we introduce three levels of togetherness and motivate each notion with an example scenario in organizational management:

**(a)** Imagine two groups of specialists who provide information and advices to each other (e.g. the accounting and the procurement teams in a company). A member of one group needs to access some but not necessarily all members of the other group. In this case, it is sufficient to measure togetherness based on the distance from a node in a network to any node in the other network.

In particular, the $\exists$-*span* $\sigma_E^{\exists}(u)$ of $u \in V_i$ refers to $\min\{dist(u,v) \mid v \in V_{3-i}\}$ where $i \in \{1,2\}$. Let $\sigma_E^{\exists}(G_1, G_2) := \max\{\sigma_{\exists}(u) \mid u \in V_1 \cup V_2\}$. This distance forms the base for *existential togetherness*.

**(b)** Imagine two groups of people with varying skills who collaborate on a joint project. To fully utilize skills and incorporate knowledge, a person in one group should access everyone in the other group. Hence, we measure togetherness based on the distance from a node in a network to all members of the other network.

In particular, the $\forall$-*span* $\sigma_E^{\forall}(u)$ of $u \in V_i$ refers to $\max\{\mathsf{dist}(u,v) \mid v \in V_{3-i}\}$ where $i \in \{1,2\}$. Let $\sigma_E^{\forall}(G_1, G_2) := \max\{\sigma_{\forall}(u) \mid u \in V_1 \cup V_2\}$, this is to be used to define *universal togetherness*.

**(c)** Imagine two groups of people who merge into a single group. To ensure the resulting group is a cohesive, tightly-knit unit, we measure togetherness based on the diameter of the combined group. The diameter is to be used to define *diametrical togetherness*.

**Definition 4.1.2.** *Let $G_1 \oplus_E G_2$ be an integration of two networks $G_1 = (V_1, E_1)$ and $G_2 = (V_2, E_2)$. We define three notions of togetherness of $G_1$ and $G_2$ as follows:*

1. *The* $\exists$-togetherness *(or* existential togetherness*) is defined as* $\tau_E^{\exists}(G_1, G_2) \coloneqq \left(\sigma_E^{\exists}(G_1, G_2)\right)^{-1}$.

2. *The* $\forall$-togetherness *(or* universal togetherness*) is defined as* $\tau_E^{\forall}(G_1, G_2) \coloneqq \left(\sigma_E^{\forall}(G_1, G_2)\right)^{-1}$.

3. *The* $\Delta$-togetherness *(or* diametrical togetherness*) is defined as* $\tau_E^{\Delta}(G_1, G_2) \coloneqq \left(\mathsf{diam}\,(G_1 \oplus_E G_2)\right)^{-1}$.

When $G_1, G_2$ and $E$ are clear from context, we abuse the notation writing $\chi^{\natural}$ for $\chi_E^{\natural}(G_1, G_2)$ for all $\chi \in \{\sigma, \tau\}$ and $\natural \in \{\exists, \forall, \Delta\}$.

In the following proposition, we use $\mathsf{diam}_{\max}$ and $\mathsf{diam}_{\min}$ to denote $\max\{\mathsf{diam}(G_1), \mathsf{diam}(G_2)\}$ and $\min\{\mathsf{diam}(G_1), \mathsf{diam}(G_2)\}$, respectively.

**Proposition 4.1.1.** *The following properties hold for all networks $G_1 = (V_1, E_1), G_2 = (V_2, E_2)$ and $E \subseteq V_1 \otimes V_2$:*

*(a)* $\left(\sigma^{\exists} + \mathsf{diam}_{\min}\right)^{-1} \leq \tau^{\forall} \leq \tau^{\exists}$

*(b)* $\tau^{\Delta} \leq \tau^{\forall}$*; and* $\tau^{\forall} = \tau^{\Delta}$ *whenever* $\sigma^{\forall} \geq \mathsf{diam}_{\max}$

*Proof.* For (a), it is clear that $\tau^{\forall} \leq \tau^{\exists}$ as $\sigma^{\exists}(u) \leq \sigma^{\forall}(u)$ for every node $u$. Without loss of generality assume $\mathsf{diam}(G_2) \leq \mathsf{diam}(G_1)$. From any node $u \in V_1$, there is $v \in V_2$ where $\mathsf{dist}(u, v) \leq \sigma^{\exists}$, and for all $w \in V_2$, $\mathsf{dist}(v, w) \leq \widetilde{d}$. Thus $\mathsf{dist}(u, w) \leq \tau^{\exists} + \mathsf{diam}_{\min}$. This means that $\sigma^{\forall} \leq \sigma^{\exists} + \mathsf{diam}_{\min}$ and hence $\tau^{\forall} \geq \left(\sigma^{\exists} + \mathsf{diam}_{\min}\right)^{-1}$.

For (b), it is clear that $\tau^{\Delta} \leq \tau^{\forall}$ as $\mathsf{diam}\,(G_1 \oplus_E G_2) \leq \sigma^{\forall}(u)$ for any node $u$. When $\sigma^{\forall} \geq \mathsf{diam}_{\max}$, $\mathsf{dist}(u, v) \geq \max\{\mathsf{diam}(G_1), \mathsf{diam}(G_2)\}$ for any $u \in V_1$ and $v \in V_2$. Thus, $\mathsf{diam}(G_1 \oplus_E G_2) = \sigma^{\forall}$, which means $\tau^{\forall} = \tau^{\Delta}$. □

**Example 5.** As an example, we integrate two networks in three ways in Figure 4.1. One may see that the choice of edges affects togetherness in resulting integrated networks.

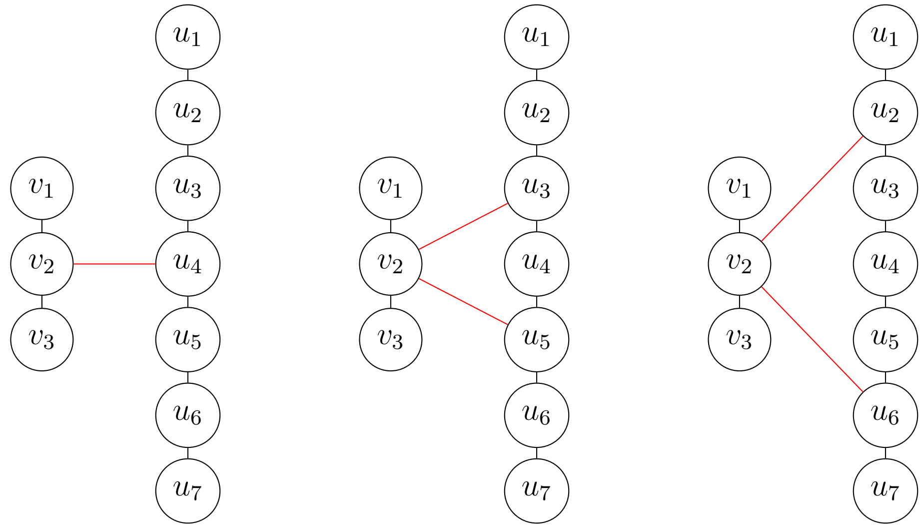



Figure 4.1: Integrating two line networks: $E_1 = \{v_2u_4\}$ with $\tau^\exists = 1/4$, $\tau^\forall = 1/5$, and $\tau^\Delta = 1/6$ (on the left); $E_2 = \{v_2u_3, v_2u_5\}$ with $\tau^\exists = 1/3$,$\tau^\forall = 1/4$, and $\tau^\Delta = 1/6$ (in the middle); and $E_3 = \{v_2u_2, v_2u_6\}$ with $\tau^\exists = 1/3$, $\tau^\forall = \tau^\Delta = 1/4$ (on the right).

### 4.1.2 The Network Integration Problems

When integrating two networks $G_1$ and $G_2$, we have two constraints:

(a) The first constraint is the togetherness of $G_1$ and $G_2$ in the integrated network.

(b) The second constraint is the number of new edges established during the process.

To ensure high togetherness, one needs to create sufficiently many edges between $G_1$ and $G_2$. As each edge requires certain resources to set up and maintain, the challenge is to obtain maximal togetherness while creating minimal number of new edges.

Formally, fix $\natural \in \{\exists, \forall, \Delta\}$. We define the following problems:

1. *Network Integration under Togetherness constraint* $\mathsf{NIT}_t^\natural(G_1, G_2)$ (where $t \in (0, 1]$): This problem asks for, given $G_1 = (V_1, E_1)$ and $G_2 = (V_2, E_2)$, a set of edges $E \subseteq V_1 \otimes V_2$ such that the togetherness $\tau_E^\natural(G_1, G_2) \geq t$. An *optimal solution* $E$ of this problem is one that has the smallest cardinality.

2. *Network Integration under Edge constraint* $\mathsf{NIE}_e^\natural(G_1, G_2)$ (where $e \geq 1$ is an integer): This problem asks for, given $G_1 = (V_2, E_1)$ and $G_2 = (V_2, E_2)$, a set $E \subseteq V_1 \otimes V_2$ that has cardinality $e$. An *optimal solution* $E$ of this problem is one that leads to the largest togetherness $\tau_E^\natural(G_1, G_2)$.

We first will show that the two network integration problems are closely related.

**Theorem 4.1.1** (Duality)**.** *For any $\natural \in \{\exists, \forall, \Delta\}$, there is a solution of $\mathsf{NIT}_t^\natural(G_1, G_2)$ containing at most $e$ edges iff there is a solution $E$ of $\mathsf{NIE}_e^\natural(G_1, G_2)$ that leads to $\tau^\natural \geq t$.*

The next result discusses the size of any optimal solution $E$ of $\mathsf{NIT}_1^\natural(G_1, G_2)$ (i.e., when we desire maximal togetherness).

**Lemma 4.1.1.** *For any networks $G_1, G_2$ and $E \subseteq V_1 \otimes V_2$:*

*(a)* $\tau^\exists$=1 *iff* $\forall u \in V_i \exists v \in V_{3-i} \colon uv \in E$ *where* $i \in \{1, 2\}$.

*(b)* $\tau^\forall$=1 *iff* $E = V_1 \otimes V_2$.

*(c)* $\tau^\Delta$=1 *iff both* $G_1, G_2$ *are complete and* $E = V_1 \otimes V_2$.

*Proof.* For (a), let $E \subseteq V_1 \otimes V_2$ be a set of edges with cardinality $|E| = \max\{|V_1|, |V_2|\}$. If $E$ connects every node in $V_1$ with some node in $V_2$ and vice versa, then the $\exists$-span $\sigma^\exists(u) = 1$ for every $u \in V_1 \cup V_2$. On the other hand, if $\exists u \in V_1 \forall v \in V_2 \colon uv \notin E$, then $\sigma^\forall(u) \geq 2$.

For (b), it is sufficient to note that if there is some $u \in V_1, v \in V_2$ with $uv \notin E$, then $\mathsf{dist}(u, v) \geq 2$; (c) is straightforward as $\tau^\Delta = 1$ iff $G_1 \oplus_E G_2$ is a complete graph. □

The next result discusses togetherness achieved by optimal solutions of $\mathsf{NIE}_1^\natural(G_1, G_2)$ (i.e., adding one new edge $\{u, v\}$).

**Lemma 4.1.2.** *For any networks $G_1, G_2$ and $uv \in V_1 \otimes V_2$, the maximum value of $\tau_{\{uv\}}^\natural(G_1, G_2)$ is*

*(a)* $(\max\{\mathsf{rad}(G_1), \mathsf{rad}(G_2)\}+1)^{-1}$ *if* $\natural = \exists$.

*(b)* $(\mathsf{rad}(G_1)+\mathsf{rad}(G_2) + 1)^{-1}$ *if* $\natural = \forall$.

*(c)* $(\max\{\mathsf{rad}(G_1)+\mathsf{rad}(G_2)+1, d\})^{-1}$ *if* $\natural = \Delta$.

*Proof.* The optimal solution $\{uv\}$ connects a center $u$ in $G_1$ with a center $v$ in $G_2$. The properties (a)(b)(c) can then be easily checked. □

The problem of finding optimal solutions for network integration is in general computationally hard.

**Theorem 4.1.2.** *For any $\natural \in \{\exists, \forall, \Delta\}$, the following problems are hard for $\mathsf{W}[2]$, the second level of the $\mathsf{W}$-hierarchy:*

1. *Fix* $t \in (0, 1/2]$. *Decide if* $\mathsf{NIT}^\natural_t(G_1, G_2)$ *has a solution with* $\leq e$ *edges for given* $G_1, G_2$ *and integer* $e > 0$

2. *Fix* $e > 1$. *Decide if* $\mathsf{NIE}^\natural_e(G_1, G_2)$ *has a solution* $E$ *that leads to* $\tau^\natural \geq t$ *for given* $G_1, G_2$ *and* $t \in (0, 1]$.

*Proof.* Due to the duality (Theorem 4.1.1) between the two problems $\mathsf{NIT}^\natural_t(G_1, G_2)$ and $\mathsf{NIE}^\natural_e(G_1, G_2)$, it is sufficient to prove one of (1) and (2). As shown in [71], finding the smallest distance-$r$ dominating set in $G$ with diameter $r + 1$ is complete for $\mathsf{W}[2]$ (for any fixed $r$). We now show a reduction from this problem to $\mathsf{NIT}^\natural_t(G_1, G_2)$ for $t \in (0, 1/2]$.

Suppose, without loss of generality, that $t = k^{-1}$ for some integer $k \geq 2$. Now let $G_1 = (V_1, E_1)$ be a graph with diameter $k$ and let $G_2$ be a graph that contains only a single node $\{u\}$. For any distance-$(k-1)$ dominating set $S \subseteq V_1$, the set of edges $S \otimes \{u\}$ is a solution of $\mathsf{NIT}^\natural_t(G_1, G_2)$. Conversely, suppose $S \subseteq V_1$ is not distance-$(k-1)$ dominating. Then there is a node $w \in V_1$ that is at distance at least $k$ away from any node $v \in S$. This means that $\mathsf{dist}(w, u)$ in the integrated network is at least $k+1$ and $S$ is not a solution of $\mathsf{NIT}^\natural_t(G_1, G_2)$. Thus $\mathsf{NIT}^\natural_t(G_1, G_2)$ has a size-$\ell$ solution if and only if $G_1$ has a size-$\ell$ distance-$(k-1)$ dominating set. □

### 4.1.3 Privilege and Priorities

We note that to find optimal solutions for $\mathsf{NIT}^\natural_t(G_1, G_2)$ and $\mathsf{NIE}^\natural_e(G_1, G_2)$ problems, one needs to make an assumption that nodes in the integrated networks are *equipotent*. This condition assumes the networks follow *peer-to-peer* relational dynamic, which refers to social structures where information and resources are distributed.

In such a social structure, as discussed by Baker in [7], members have no formal authority over each other, and have equal privileges regardless their roles [27]. Examples of such social groups include volunteer organizations, teams of scientists, and companies that embrace a holacracy management style [96]. Baker claims that in order for such a peer-to-peer network to operate efficiently, there must be clear and open communication; moreover each individual should be aware of the resources available from other nodes. The main challenge of merging two such organizations is to establish channels that allow exchanges of intellectual ideas and trusted transactions.

However, when integrating two organizations, each person may have constraints over who he or she may connect to; this is determined largely by *privilege*, i.e., the type of social inequality created from difference in positions, titles, ranks, etc [27]. We, thus, consider the special case when certain priorities are assigned to all nodes in the networks.

Summing up these ideas, we present several algorithms for integrating two networks. The mechanisms are broadly divided into two categories:

1. We propose heuristics that search for small sets $E$ that integrate two networks $G_1, G_2$. We assume that the networks enjoy *equi-privilege property*, i.e., any pair of nodes between networks can be freely connected. The goal of these heuristics is to gain maximal togetherness in the integrated networks.

2. We propose four scenarios where nodes in one network preferentially establish links with nodes in the other network. Here every node is given a *priority* which is determined by the network structure. The difference between these priority based methods and the heuristics in the first category is that their aim is to simulate the preferential attachments of links during integration, rather than explicitly searching for good solutions.

## 4.2 Algorithms for Integrating Networks with Equi-Privilege Property

In this section we focus on social networks with equipotent nodes, and therefore assume all nodes have unbounded and equal privilege. Discussions on *equi-privilege property* originates from organizational behavioral studies of social networks. In [7], Mila Baker describes peer-to-peer organizations as social structures where members have equal privileges regardless of their roles (such as volunteer groups, research teams, etc.); these organizational structures are said to have equi-privilege property.

In subsequent subsections we focus on heuristics for solving the $\mathsf{NIT}_t^\natural(G_1, G_2)$ and $\mathsf{NIE}_e^\natural(G_1, G_2)$ problems. We consider each level of togetherness and present several heuristics for solving the network integration problems. Indeed, the mechanism for network integration depends on the level of togetherness one desires to achieve. If the goal is to optimize $\exists$-togetherness, by Theorem 4.2.1, the key is to identify dominating sets in the networks $G_1$ and $G_2$. If the goal is to optimize $\forall$- or $\Delta$- togetherness, then it is desirable to establish links that minimize diameter.

### 4.2.1 Optimizing $\exists$-togetherness

We first focus on $\exists$-togetherness and characterize the optimal solutions of $\mathsf{NIT}_t^{\exists}(G_1, G_2)$.

**Theorem 4.2.1** ($\exists$-Togetherness Theorem)**.** *Suppose $E$ is an optimal solution of $\mathsf{NIT}_t^{\exists}(G_1, G_2)$. Then*

*(1) If $t = 1$, then $|E| = \max\{|V_1|, |V_2|\}$*

*(2) If $t < (\max\{\mathsf{rad}(G_1), \mathsf{rad}(G_2)\})^{-1}$, then $|E| = 1$*

*(3) If $1 > t \geq (\max\{\mathsf{rad}(G_1), \mathsf{rad}(G_2)\})^{-1}$, $|E| = \max\{\gamma_1, \gamma_2\}$, where $\gamma_i$ is the $(t^{-1}-1)$-dominating number of the network $G_i$ for each $i \in \{1, 2\}$.*

*Proof.* (1) and (2) directly follow from Lemma 4.1.1(a) and Lemma 4.1.2(a), resp. We now prove (3).

Let $k = t^{-1}$ and $D_1 \subseteq V_1$ and $D_2 \subseteq V_2$ be minimum distance-$(k-1)$ dominating sets for $G_1$ and $G_2$, resp. In other words, $|D_1| = \gamma_1$ and $|D_2| = \gamma_2$. Without loss of generality, assume $\gamma_1 \geq \gamma_2$. Then there is a set $E \subseteq V_1 \otimes V_2$ that contains for every $u \in D_i$, some edge $uv$ where $v \in D_{3-i}$ where $i \in \{1, 2\}$, and $|E| = \gamma_1$. Our goal is to show that $E$ is an optimal solution of $\mathsf{NIT}_t^{\exists}(G_1, G_2)$.

Note that any node $w$ in $V_i$ is at most $k - 1$ steps away from some node $u \in D_i$, which means that the $\exists$-span $\sigma^{\exists}(w) \leq k$. Thus $E$ is a solution of $\mathsf{NIT}_t^{\exists}(G_1, G_2)$. Now take a set $E' \subseteq V_1 \otimes V_2$ has $|E'| < \Gamma_1$. Let $S = \{u \in V_1 \mid \exists v \in V_2 \colon uv \in E'\}$. Then there is some node $w \in V_1$ that is at least $k$ steps away from any node in $S$. Thus the $\exists$-span $\sigma^{\exists}(w) > k$ and $E'$ is not a solution. This means that $E$ is an optimal solution. $\square$

Finding optimal solution $E$ for the $\mathsf{NIT}_t^{\exists}(G_1, G_2)$ (where $t \in (0, 1]$) is equivalent to finding two distance $k$ dominating sets for each $G_1$ and $G_2$. As soon as the dominating sets are found, create $\max\{|D_1|, |D_2|\}$ new edges: connect each node in a smaller dominating set with a node in the larger set (one-to-one) until the set is empty. Then connect the remaining nodes with any nodes in the other graph.

In Chapter 2, we presented eight greedy algorithms for finding distance $k$-dominating sets: Max, Min, Btw, MinLeaf, S-Max, S-Min, S-Btw, S-MinLeaf. Any of these algorithms could be applied for finding small sets of edges $E$. We compare the performance of these algorithms below in Section 4.5.

**Example 6.** To illustrate the idea, consider an example as in Figure 4.2. Here, we have two random Newman-Watts-Strogatz connected graphs with 50 nodes each. One of the graphs has radius 6 and diameter 9. The second graph has radius 7 and diameter 10. Thus, using the S-Max algorithm, we get that if $\tau^{\exists} = \frac{1}{6}$, 3 new edges should be created; for $\tau^{\exists} = \frac{1}{5}$, the number of edges if 4, for $\tau^{\exists} = \frac{1}{4}$, it is 7; and finally if $\tau^{\exists} = \frac{1}{3}$, the total number of edges is 11. One may see that the larger the value of $\exists$-togetherness is, the less the initial structure of the graph is visible.

To integrate networks with $\exists$-togetherness constraint we apply S-MinLeaf, which proves to be the most efficient for finding small distance $k$ dominating sets (see Experiment 16 in Section 4.5):

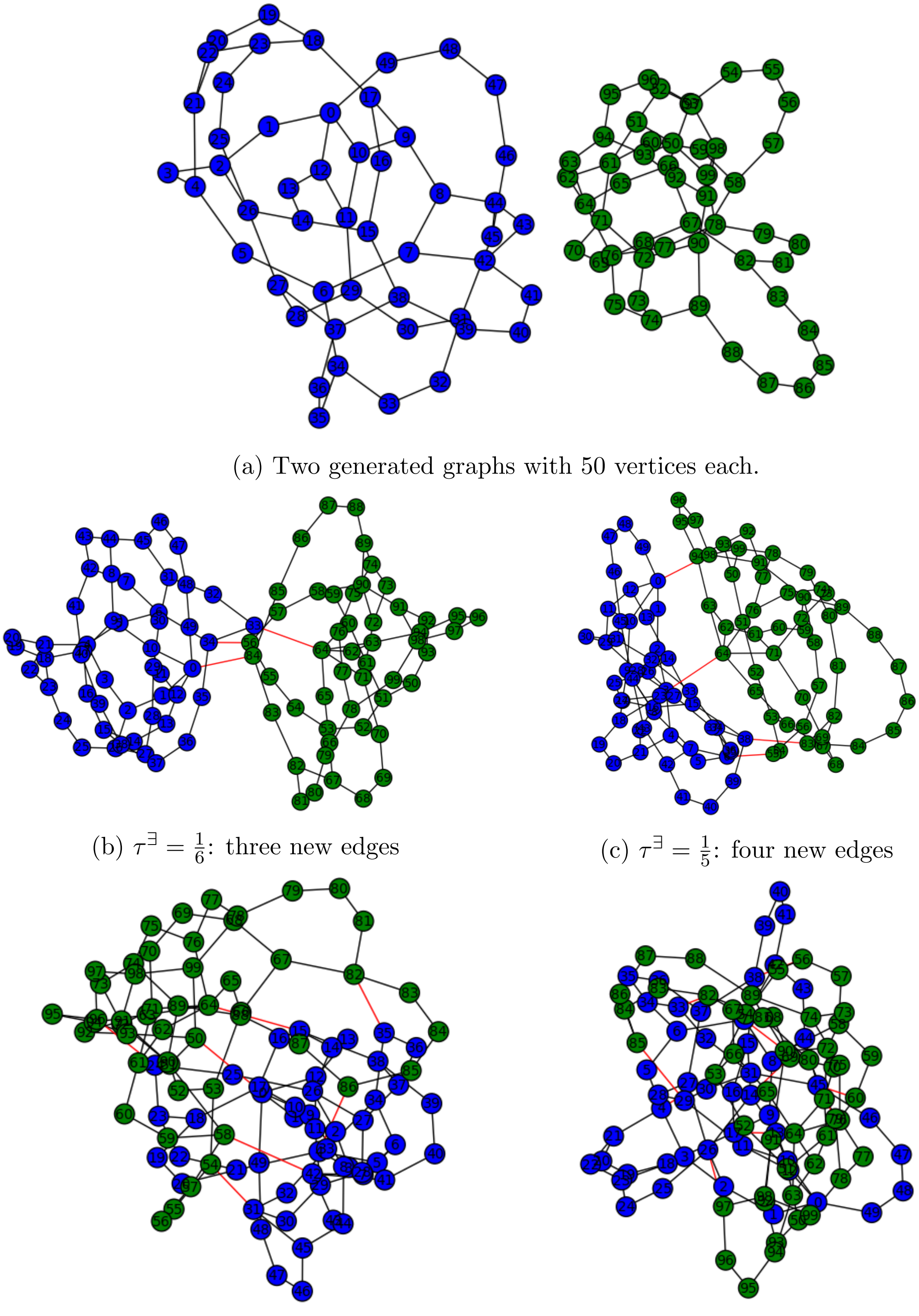
Our algorithm takes two networks $G_1 = (V_1, E_1)$ and $G_2 = (V_2, E_2)$ and $t \in (0, 1]$ as input and outputs a set $E = D_1 \oplus D_2$, where $D_1$ and $D_2$ are distance $(t^{-1}-1)$ dominating sets in $G_1$ and $G_2$, respectively. By Theorem 4.2.1, the set $E$ is a solution of $\mathsf{NIT}_t^{\exists}(G_1, G_2)$.

For $i \in \{1, 2\}$, the algorithm iteratively builds $D_i \subseteq V_i$ by maintaining a set $U_i \subseteq V_i$ of "uncovered" nodes, i.e., nodes that have distance $\geq k$ from any current node in $D_i$. The initial configuration is when $U_i = V_i$. It repeatedly performs the following operations until $U = \varnothing$:

1. Select a node $u \in U$ and add $u$ to $D_i$ (see below).

2. Compute all nodes at distance at most $(k-1)$ from $v$ and remove these nodes from $U$.

We now describe the heuristic for finding a node $u$ in each iteration. A node is called a *leaf* if it has minimum degree in the graph; leaves correspond to least connected members in the network, and may become outliers once nodes with higher degrees are removed from the network. Therefore, the heuristic first picks a leaf $v$ in $U_i$, then applies a sub-procedure to find the next node $u$ to be added to $U_i$. The sub-procedure determines a path $v = w_1, w_2, \ldots$ iteratively as follows:



(a) Two generated graphs with 50 vertices each.

(b) $\tau^{\exists} = \frac{1}{6}$: three new edges

(c) $\tau^{\exists} = \frac{1}{5}$: four new edges

(d) $\tau^{\exists} = \frac{1}{4}$: seven new edges

(e) $\tau^{\exists} = \frac{1}{3}$: eleven new edges

Figure 4.2: Integrating two networks with different values of $\tau^{\exists}$

1. Suppose $w_i$ is picked. If $i = r$ or $w_i$ has no adjacent node in $U_i$, set $w_i$ as $u$ and

terminate the process.

2. Otherwise select a $w_{i+1}$ (which is different from $w_{i-1}$) among adjacent nodes of $w_i$ with maximum degree.

When this process terminates, the algorithm adds $u$ to $D_i$. Note that the distance between $u$ and $v$ is at most $k-1$.

**Theorem 4.2.2.** *For any networks $G_1, G_2$ and $t \in (0,1]$, the* S-MinLeaf *algorithm outputs a solution of the* $\mathsf{NIT}_t^{\exists}(G_1, G_2)$ *problem.*

### 4.2.2 Optimizing $\forall$-/$\Delta$-togetherness

In this subsection we consider $\mathsf{NIT}_t^{\natural}(G_1, G_2)$ problem for $\natural \in \{\forall, \Delta\}$. Clearly, to optimize $\forall$- or $\Delta$- togetherness, one needs to establish links that minimize diameter.

The next theorem bounds the size of optimal solutions of $\mathsf{NIT}_t^{\natural}(G_1, G_2)$ for large $t$ (i.e., $t \geq 1/3$). Recall that $\gamma(G)$ denotes the dominating number of $G$.

**Theorem 4.2.3.** *Suppose $E$ be an optimal solution of* $\mathsf{NIT}_t^{\natural}(G_1, G_2)$ *where* $\natural \in \{\forall, \Delta\}$.

*(1) If $t = 1$, then $|E| = |V_1| \cdot |V_2|$*

*(2) If $t = 1/2$, then $|E| \leq \min\{\gamma(G_1) \cdot |V_2|, \gamma(G_2) \cdot |V_1|\}$*

*(3) If $t = 1/3$, then $|E| \leq |V_1| + |V_2| - 1$*

*Proof.* (1) directly follows from Lemma 4.1.1 (b)(c).

For (2), let $D_1, D_2$ be a dominating set in $G_1$ and $G_2$, respectively. Let $E_1 = \{uv \mid u \in D_1, v \in V_2\}$ (so $|E_1| = \gamma(G_1) \cdot |V_2|$) and $E_2 = \{uv \mid u \in V_1, v \in D_2\}$ (so $|E_2| = \gamma(G_2) \cdot |V_1|$). Then both $G_1 \oplus_{E_1} G_2$ and $G_1 \oplus_{E_2} G_2$ have diameter 2, and thus $\tau^{\forall} = \tau^{\Delta} = 1/2$.

For (3), pick any node $u \in V_1$ and $v \in V_2$ and let $E' = \{uy \mid y \in V_2\} \cup \{xv \mid x \in V_1\}$. Then $|E'| = |V_1| + |V_2| - 1$. The diameter of the integrated network $G_1 \oplus_{E'} G_2$ is 3, and thus $\tau^{\forall} = \tau^{\Delta} = 1/3$. $\square$

**Diameter oriented approach: $\mathsf{NI}_\Delta(G_1, G_2)$ integration problem**

Clearly, the problem $\mathsf{NIT}_t^\Delta(G_1, G_2)$ is equivalent to the problem of integrating two networks such that $\mathsf{diam}(G_1 \oplus_E G_2) \leq t^{-1}$ for some given $t \in (0,1]$ (as $\tau_E^\Delta(G_1, G_2) = (\mathsf{diam}\,(G_1 \oplus_E G_2))^{-1}$ ). Thus, we reformulate $\mathsf{NIT}_t^\Delta(G_1, G_2)$ as following:

Take an integer $\mathsf{dm} \geq 1$ (that is $\mathsf{dm} = \tau_E^\Delta(G_1, G_2)^{-1}$). We propose the following modified network integration problem, $\mathsf{NI}_\Delta(G_1, G_2)$:

**INPUT** Two networks $G_1 = (V_1, E_1), G_2 = (V_2, E_2)$ where $V_1 \cap V_2 = \varnothing$.

**OUTPUT** A set $E \subseteq V_1 \otimes V_2$ such that $\mathsf{diam}(G_1 \oplus_E G_2) \leq \mathsf{dm}$.

In the rest of the section we investigate $\mathsf{NI}_\Delta(G_1, G_2)$ on two networks $G_1 = (V_1, E_1)$ and $G_2 = (V_2, E_2)$ where $V_1 \cap V_2 = \varnothing$. The problem naturally depends on the value of $\mathsf{dm}$. When $\mathsf{dm} = 1$, it is easy to see that $\mathsf{NI}_\Delta(G_1, G_2)$ has a solution if and only if both networks $G_1, G_2$ are complete. When $\mathsf{dm} \geq 2$, since $G_1 \oplus_{V_1 \otimes V_2} G_2$ has diameter 2, $\mathsf{NI}_\Delta(G_1, G_2)$ guarantees to have a solution.

Throughout, we assume $\mathsf{dm} \geq 2$. We are interested in a solution $E$ to the problem $\mathsf{NI}_\Delta(G_1, G_2)$ that contains the least number of edges; such an $E$ is called an *optimal solution* of $\mathsf{NI}_\Delta(G_1, G_2)$.

The brute-force way of finding optimal solutions for $\mathsf{NI}_\Delta(G_1, G_2)$ examines all possible sets of edges until it finds a required solution set $E$. This will take time $2^{O(|V_1|\cdot|V_2|)}$. In fact, obtaining optimal solutions is computationally-hard. Theorem 4.1.2 implies that this problem is unlikely to be polynomial-time solvable.

We propose two heuristics to perform network integration. The first is a naive greedy method that iteratively creates edges to minimize the diameter of the resulting network. The second method separately discusses two cases: 1) When $\mathsf{dm}$ is at least the diameter of the original networks, we create edges by considering center and peripheral nodes in the networks. 2) When $\mathsf{dm}$ is smaller than the original diameter of the original networks, we first reduce the distance between nodes in the respective networks and then apply the procedure in case 1).

We, therefore, turn to heuristics for finding small solution sets of $\mathsf{NI}_\Delta(G_1, G_2)$. In the next subsection, we present several efficient approaches.

### 4.2.3 Algorithms for Solving $\mathsf{NI}_\Delta(G_1, G_2)$

We first introduce a greedy heuristic that approximates a solution of $\mathsf{NI}_\Delta(G_1, G_2)$; this heuristic will also be used as a benchmark in our experiments in Section 4.5 (see Experiment 10, Experiment 11).

**Definition 4.2.1.** *A set of edges $E \subseteq V_1 \otimes V_2$ is called a* naive greedy set *if we can write it as $\{e_1, \dots, e_\ell\}$ such that for all $1 \le i \le \ell$, and $e' \in (V_1 \otimes V_2) \setminus \{e_1, \dots, e_{i-1}\}$, $\mathsf{diam}(G_1 \oplus_{\{e_1,\dots,e_i\}} G_2) \le \mathsf{diam}(G_1 \oplus_{\{e_1,\dots,e_{i-1},e'\}} G_2)$. A* naive greedy solution *to $\mathsf{NI}_\Delta(G_1, G_2)$ is a solution that is a naive greedy set.*

As its name suggests, a naive greedy set can be constructed incrementally using a greedy strategy that locally optimizes diameter in the integrated network. Naive greedy solutions to $\mathsf{NI}_\Delta(G_1, G_2)$ are not necessarily optimal, and vice versa:

**Example 7.** Let both $G_1$ and $G_2$ be paths of length 5, i.e., $G_1$ contains nodes $a_1, \dots, a_5$ while $G_2$ contains nodes $b_1, \dots, b_5$ with edges $a_i a_{i+1}$, $b_i b_{i+1}$ for any $1 \le i < 5$.

Suppose $\mathsf{dm} = 3$. The only optimal solution $E$ contains four edges, i.e., $E = \{a_1b_3, a_3b_1, a_3b_5, a_5b_3\}$. However, for any edge $e \in E$, $\mathsf{diam}\left(G_1 \oplus_{\{e\}} G_2\right) = 7$, while $\mathsf{diam}\left(G_1 \oplus_{\{a_3b_3\}} G_2\right) = 5$. Thus $E$ is not a naive greedy solution, nor will any naive greedy solution be optimal.

**Theorem 4.2.4.** *There exists an algorithm $\mathsf{Naive}_\Delta(G_1, G_2)$ that runs in time $O(n^6)$ and computes a naive greedy solution for $\mathsf{NI}_\Delta(G_1, G_2)$ where $n = |V_1 \cup V_2|$.*

*Proof.* The algorithm $\mathsf{Naive}_\Delta(G_1, G_2)$ iteratively adds edges $e_1, e_2, \dots$ to the solution set $E$. It also computes a matrix $D : (V_1 \cup V_2)^2 \to \mathbb{N}$ that represents the distance between nodes in the current integrated graph. See Procedure 3
Since $\mathsf{dm} \ge 2$, the algorithm will terminate. Furthermore, the set of edges created by the algorithm is a naive greedy solution.

At each iteration, computing each matrix $D_e$ takes time $O(n^2)$; computing $F$ takes $O(n^2)$. Since there are $O(n^2)$ edges in $(V_1 \otimes V_2) \setminus E_i$, this iteration runs in $O(n^4)$. Since there are at most $n^2$ iterations, the algorithm takes times $O(n^6)$. □ □

We remark that when $\mathsf{dm} > 2$, the maximum number of edges required is $O(n)$, and hence $\mathsf{Naive}_\Delta(G_1, G_2)$ will take $O(n^5)$. The algorithm $\mathsf{Naive}_\Delta(G_1, G_2)$ is still too inefficient in most practical cases and hence in subsequent subsections we discuss more efficient heuristics for $\mathsf{NI}_\Delta(G_1, G_2)$.

**Procedure 3** $\mathsf{Naive}_\Delta(G_1, G_2)$; Output $E$

```
Initialize D (for the disjoint union of G1, G2); Set E := ∅
while diam (G1 ⊕_E G2) > dm do
    for e := xy ∈ (V1 ⊗ V2) \ E do
        for (u, v) ∈ (V1 ∪ V2)^2 do          ▷ define a temporary De : (V1 ∪ V2)^2 → ℕ
            De(u, v) := min{Di(u, v), Di(u, x) + Di(y, v) + 1}
        end for
        Set diam_e := max{De(u, v): (u, v) ∈ (V1 ∪ V2)^2}.
    end for
    Set F := {e ∈ (V1 ⊗ V2) \ Ei | diam_e ≤ diam_e' for all e' ∈ (V1 ⊗ V2) \ Ei}.
    Pick a random edge ei ∈ F and set D := D_ei, E := E ∪ {ei}
end while
```

**Efficient Algorithms for** $\mathsf{NI}_\Delta(G_1, G_2)$ We separately discuss two cases:

(a) when the integrated network's diameter is at least the diameters of the given networks, i.e. $\mathsf{dm} \geq \max\{\mathsf{diam}(G_1), \mathsf{diam}(G_2)\}$;

(b) when we 'improve' the diameter, i.e. $\mathsf{dm} < \max\{\mathsf{diam}(G_1), \mathsf{diam}(G_2)\}$.

**(a) The case when** $\mathsf{dm} \geq \max\{\mathsf{diam}(G_1), \mathsf{diam}(G_2)\}$

When integrating two networks, it makes sense first to establish a link between the most central persons in the networks, as they have the closest proximity to other nodes.

Furthermore, if $x, y$ are nodes that are furthest apart in the integrated network, they are unlikely to communicate effectively thanks to their shear distance; this, in a certain sense, represents a form of *structural hole* [26]. Hence, it makes sense to connect $x, y$ by an edge.

Formally, the *center* $C(G)$ of a graph $G = (V, E)$ is the set of all nodes that have the least eccentricity, i.e., $C(G) = \{v \in V \mid \mathsf{ecc}(v) = \mathsf{rad}(G)\}$. A pair of nodes $(x, y)$ in $G$ forms a *peripheral pair*, denoted by $(x, y) \in \mathsf{P}(G)$, if $\mathsf{dist}(x, y) = \mathsf{diam}(G)$.

Our heuristic first creates an edge between two nodes that are in $C(G_1)$ and $C(G_2)$ respectively, and then iteratively "bridges" peripheral pairs.

**Definition 4.2.2.** *A set* $E \subseteq V_1 \otimes V_2$ *is called a* center-periphery set *if we can write it as* $\{e_0, \ldots, e_\ell\}$ *such that:*

1. $e_0 \in C(G_1) \otimes C(G_2)$;

2. *for all* $1 \le i \le \ell$, $e_i \in \mathsf{P}\left(G_1 \oplus_{\{e_0,e_1,\ldots,e_{i-1}\}} G_2\right)$.

*A* center-periphery solution *is a solution to* $\mathsf{NI}_\Delta(G_1, G_2)$ *that is also a center-periphery set.*

Clearly, if $\mathsf{dm} > \mathsf{rad}(G_1) + \mathsf{rad}(G_2)$, then for any $uv \in C(G_1) \otimes C(G_2)$, we have $\mathsf{diam}\left(G_1 \oplus_{\{uv\}} G_2\right) \le \max\{\mathsf{diam}(G_1), \mathsf{diam}(G_2), \mathsf{rad}(G_1) + \mathsf{rad}(G_2) + 1\} \le \mathsf{dm}$. Thus $\{uv\}$ forms a solution of $\mathsf{NI}_\Delta(G_1, G_2)$. In this case, center-periphery solutions coincide with optimal solutions.

**Theorem 4.2.5.** *There exists an algorithm* CtrPer *that has* $O(n^4)$ *running time* ($n = |V_1 \cup V_2|$) *and computes a center-periphery solution for* $\mathsf{NI}_\Delta(G_1, G_2)$, *where* $\mathsf{dm} \ge \max\{\mathsf{diam}(G_1), \mathsf{diam}(G_2)\}$.

*Proof.* The CtrPer algorithm also maintains a matrix $D : (V_1 \cup V_2)^2 \to \mathbb{N}$ such that $D(u, v)$ is the distance between $u, v$.

The eccentricity of each node can be easily extracted from $D$ allowing the algorithm to identify the centers $C(G_1)$ and $C(G_2)$, respectively. The algorithm then iteratively adds edges that connect peripheral pairs in the integrated graph until its diameter becomes at most $\mathsf{dm}$. See Procedure 4.

Suppose the algorithm creates a set $E \subseteq V_1 \otimes V_2$ and $\mathsf{diam}\left(G_1 \oplus_E G_2\right) > \mathsf{dm}$.

The algorithm will update the matrix $D$ and then picks $(u, v)$ with the largest $D(u, v)$. By definition of $D$, $(u, v)$ forms a peripheral pair in $G_1 \oplus_E G_2$. We need to show that $uv$ is a valid edge to add, that is, $u, v$ cannot both lie in one of $V_1$ and $V_2$. Indeed, if $\{u, v\} \subseteq V_1$ or $\{u, v\} \subseteq V_2$, then $\mathsf{dist}(u, v) \le \max\{\mathsf{diam}(G_1), \mathsf{diam}(G_2)\} \le \mathsf{dm} < \mathsf{diam}\left(G_1 \oplus_E G_2\right)$. Thus, $uv \in V_1 \otimes V_2$.

Now either $E \cup \{uv\}$ is a solution, or $\mathsf{diam}\left(G_1 \oplus_{E\cup\{uv\}} G_2\right) > \mathsf{dm}$. In the latter case the algorithm repeats the iteration to find another peripheral pair. Thus, the algorithm will terminate and produce a center-periphery solution to $\mathsf{NI}_\Delta(G_1, G_2)$.
It takes $O(n^3)$ to initialize the matrix $D$ using Floyd-Warshall algorithm. At each iteration, the algorithm takes $O(n^2)$ to update $D$ and finds a peripheral pair. Since there are at most $n^2$ iterations, the algorithm takes time $O(n^4)$. □ □

It is easy to see that, consequently, the following theorem also holds:

**Theorem 4.2.6.** *For any networks* $G_1, G_2$ *and* $t \in (0, 1]$, CtrPer *algorithm outputs a solution of the* $\mathsf{NIT}_t^\forall(G_1, G_2)$ *problem.*

**Procedure 4** CtrPer: $\mathsf{dm} \geq \max\{\mathsf{diam}(G_1), \mathsf{diam}(G_2)\}$; Output $E$

```
Initialize the matrix D so that D(u,v) = dist(u,v) in the un-integrated graphs
Take a node u ∈ C(G_1) and a node v ∈ C(G_2)
Set e := uv, E := {e}
while (diam(G_1 ⊕_E G_2) > Δ) do
    for (x,y) ∈ (V_1 ∪ V_2)^2 do                          ▷ define D' : (V_1 ∪ V_2)^2 → ℕ
        D'(x,y) := min{D(x,y), D(x,u) + D(v,y) + 1} where e = uv
    end for
    D := D'                                                ▷ update matrix D
    Pick (u,v) with the largest D(u,v). Set e := uv and E := E ∪ {e}
end while
```

**(b) The case when $\Delta < \max\{\mathsf{diam}(G_1), \mathsf{diam}(G_2)\}$**

When the diameter bound $\mathsf{dm}$ is less than the diameters of the two component networks $G_1, G_2$, the goal is to *improve* the connectivity of each original network through integration. In other words, the integration should "bring people closer". In this case CtrPer no longer applies as it is possible for both nodes in a peripheral pair to lie in the same component graph $G_1$ or $G_2$, forbidding us to create the edge $xy$. We, therefore, need to first decrease the distance between nodes in each $G_1$ and $G_2$.

Suppose $a, b$ are two people in an organization with large distance. When their organization merges with another organization, $a$ and $b$ can be brought closer if they both know a 'third person' $c$ in the other organization, i.e., the ties $ac$ and $bc$ allows $a, b$ to be only 2 steps away.

**Definition 4.2.3.** *Let $E \subseteq V_1 \otimes V_2$ be a set of edges and $i \in \{1, 2\}$. The* diameter of $G_i$ relative to $E$ *is the maximum distance between any two nodes in $V_i$ in the integrated network $G_1 \oplus_E G_2$; we denote this by $\mathsf{diam}_E(G_i)$. A set of edges $E \subseteq V_1 \otimes V_2$ is a* $\mathsf{dm}$-bridge *if $\mathsf{diam}_E(G_i) \leq \mathsf{dm}$ for both $i \in \{1, 2\}$.*

**Theorem 4.2.7.** *For any $\mathsf{dm} \geq 2$, there exists an algorithm $\mathsf{Bridge}_{\mathsf{dm}}(G_1, G_2)$ that runs in time $O(n^4)$ and computes a $\mathsf{dm}$-bridge $E$, where $n = |V_1 \cup V_2|$.*

*Proof.* The algorithm has two phases. In phase $i \in \{1, 2\}$, it makes $\mathsf{diam}_E(G_i) \leq \mathsf{dm}$. Phase $i$ consists of several iterations; at each iteration, the algorithm takes a pair $(u, v) \in V_i$ with maximum distance and a node $w \in V_{3-i}$, and builds two edges $uw$ and $vw$. See Procedure 5.

Throughout, the algorithm computes and maintains a matrix $D : (V_1 \cup V_2)^2 \to \mathbb{N}$ such that $D(u, v)$ is the current distance between nodes $u, v$. When a pair of new edges

$uw, vw$ are added, the new distance $D'(x,y)$ between any pair of nodes $(x,y) \in V_i^2$ is calculated as follows:

$$D'(x,y) = \min\{D(x,y), D(x,u) + D(w,y) + 1, D(x,u) + D(v,y) + 2, \\ D(x,v) + D(w,y) + 1, D(x,v) + D(u,y) + 2\} \quad (4.1)$$

In the worst case, the algorithm adds edges $uw, vw$ for any pair $(u,v) \in V_i^2$ where $i \in \{1,2\}$. Thus, the algorithm terminates in at most $n^2$ iterations. Finding nodes $u,v,w$ and updating the matrix $D$ at each iteration takes time $O(n^2)$. Therefore, the total running time is $O(n^4)$. □ □

**Procedure 5** $\mathsf{Bridge}_{\mathsf{dm}}(G_1,G_2)$: $\mathsf{dm} < \max\{\mathsf{diam}(G_1), \mathsf{diam}(G_2)\}$; Output $E$

```
Initialize the matrix D so that D(u,v) = dist(u,v) in the un-integrated graphs
Initialize E := ∅
for i = 1,2 do                                        ▷ The two phases
    while diam_E(G_i) > dm do
        Take a pair of nodes u,v ∈ V_i with maximum D(u,v)
        Take a node w in V_{3−i}
        E := E ∪ {uw, wv}
        for (x,y) ∈ (V_1 ∪ V_2)^2 do                  ▷ define D' : (V_1 ∪ V_2)^2 → ℕ
            Compute D'(x,y) as in (4.1)
        end for
        D := D'                                       ▷ update matrix D
    end while
end for
```

**Remark.** Suppose the $\mathsf{Bridge}_{\mathsf{dm}}(G_1,G_2)$ algorithm adds edges $uw, vw$. Here $w$ plays the role as a *bridging node* that links $u$ and $v$. Naturally, the choice of $w$ affects the performance of the algorithm: by carefully choosing the bridging node $w$, we may reduce the number of new ties that need to be created. Imagine that $G_1, G_2$ represent two organizations.

1. To allow smooth flow of information between the two organizations and avoid *information gate keepers*, we should have many bridging nodes in $G_2$.

2. A node with a higher degree means it has better access to resource and information, and thus is a more appropriate bridging nodes.

Therefore, we introduce the following heuristics to $\mathsf{Bridge}_{\mathsf{dm}}(G_1,G_2)$:

Suppose the algorithm has selected a set $E$ of edges and picked $(u, v) \in V_i$ where $i \in \{1, 2\}$ with the largest $D(u, v)$. To pick a bridging node $w$:

**Heuristic 1** For any node $w \in V_{3-i}$, let $b(w) = |\{v \mid wv \in E\}|$. The chosen bridging node $w$ is taken from $B_i = \{w \in V_{3-i} \mid b(w) \leq b(w') \text{ for all } w' \in V_{3-i}\}$.

**Heuristic 2** The chosen bridging node $w$ has the highest degree in $B_i$.

We now extend $\mathsf{Bridge}_{\mathsf{dm}}(G_1, G_2)$ to an algorithm that solves $\mathsf{NI}_\Delta(G_1, G_2)$.

Suppose $E$ is a set of edges output by $\mathsf{Bridge}_{\mathsf{dm}}(G_1, G_2)$. For any set $E' \supseteq E$ with $\mathsf{diam}\,(G_1 \oplus_{E'} G_2) > \mathsf{dm}$, if $(u, v)$ is a peripheral pair in $G_1 \oplus_{E'} G_2$, then:

$$\begin{aligned}\mathsf{dist}(u, v) = \mathsf{diam}\,(G_1 \oplus_{E'} G_2) &> \mathsf{dm} \\ &\geq \max\{\mathsf{diam}_E(G_1), \mathsf{diam}_E(G_2)\} \\ &\geq \max\{\mathsf{diam}_{E'}(G_1), \mathsf{diam}_{E'}(G_2)\}\end{aligned}$$

This implies that $uv$ must be a pair in $V_1 \otimes V_2$. Therefore, we can apply the same procedures as in $\mathsf{CtrPer}$ to link peripheral pairs in the integrated network to obtain a solution to $\mathsf{NI}_\Delta(G_1, G_2)$; See Procedure 6.

Note that when $\mathsf{dm} \geq \max\{\mathsf{diam}(G_1), \mathsf{diam}(G_2)\}$, $\mathsf{Bridge}_{\mathsf{dm}}(G_1, G_2)$ will output $E = \varnothing$.

Hence, we have:

**Theorem 4.2.8.** *The* $\mathsf{Integrate}_{\mathsf{dm}}(G_1, G_2)$ *algorithm runs in time* $O(n^4)$ *and computes a solution to* $\mathsf{NI}_\Delta(G_1, G_2)$ *for any networks* $G_1, G_2$ *and* $\Delta \geq 2$*, where* $n = |V_1 \cup V_2|$*.*

**Algorithm 6** $\mathsf{Integrate}_{\mathsf{dm}}(G_1, G_2)$; Output $E$

Run $\mathsf{Bridge}_{\mathsf{dm}}(G_1, G_2)$ to obtain a set $E \subseteq V_1 \otimes V_2$
Run $\mathsf{CtrPer}$ to add edges to $E$ (instead of building $E$ from scratch)

Consequently, we can see the following:

**Theorem 4.2.9.** *For any networks* $G_1, G_2$ *and* $t \in (0, 1]$, $\mathsf{Integrate}_{\mathsf{dm}}(G_1, G_2)$ *algorithm outputs a solution of the* $\mathsf{NIT}_t^\natural(G_1, G_2)$ *problem, where* $\natural \in \{\forall, \Delta\}$ *and* $t = \mathsf{dm}^{-1}$*.*

## 4.3 Integrating Networks with Priority Based Approaches

In real-world networks nodes differ in various ways which may affect their integration. For example, in an organization, people sometimes connect to each other according to their abilities or roles. We therefore consider the network integration problem under the assumption that all nodes have certain priorities. We investigate cases when higher priorities are assigned to nodes with different structural properties:

1. MaxDegree: A high degree indicates the possession of certain advantage such as capability or resources. Hence, we give higher priorities to nodes with higher degrees.

2. MinDegree: A low degree indicates a certain disadvantage such as isolation and lack of resources. To ensure togetherness, it also may be reasonable to give higher priorities to nodes with lower degree.

3. MaxBtw: *Betweenness* indicates the centrality of a node, i.e., how much the node serve as a "gatekeeper" and connects diverse parts of the network [13]. Hence in this scenario, we give higher priorities to nodes with higher betweenness.

4. Random: Lastly, we consider the case when the priorities are assigned randomly. This corresponds to a case when the priorities are assigned according to some extraneous factors.

For each of the four scenarios above, we implement a mechanism that integrates networks $G_1$ and $G_2$ to achieve $\natural$-togetherness $t \in (0, 1]$. The procedure iteratively builds a set $E \subseteq V_1 \otimes V_2$ that is a solution to $\mathsf{NIT}_t^\natural(G_1, G_2)$. If $uv$ is created, the nodes $u, v$ become *bridging nodes* that link $G_1$ with $G_2$. We have the following intuition:

1. To allow smooth flow of resources between the two networks and avoid *information gate keepers*, we should have many different bridging nodes.

2. Nodes with higher priorities should serve more as bridging nodes and be linked.

Suppose a set of edges $E'$ has already been created. We adopt the following mechanism to find two nodes $u \in V_1$ and $v \in V_2$:

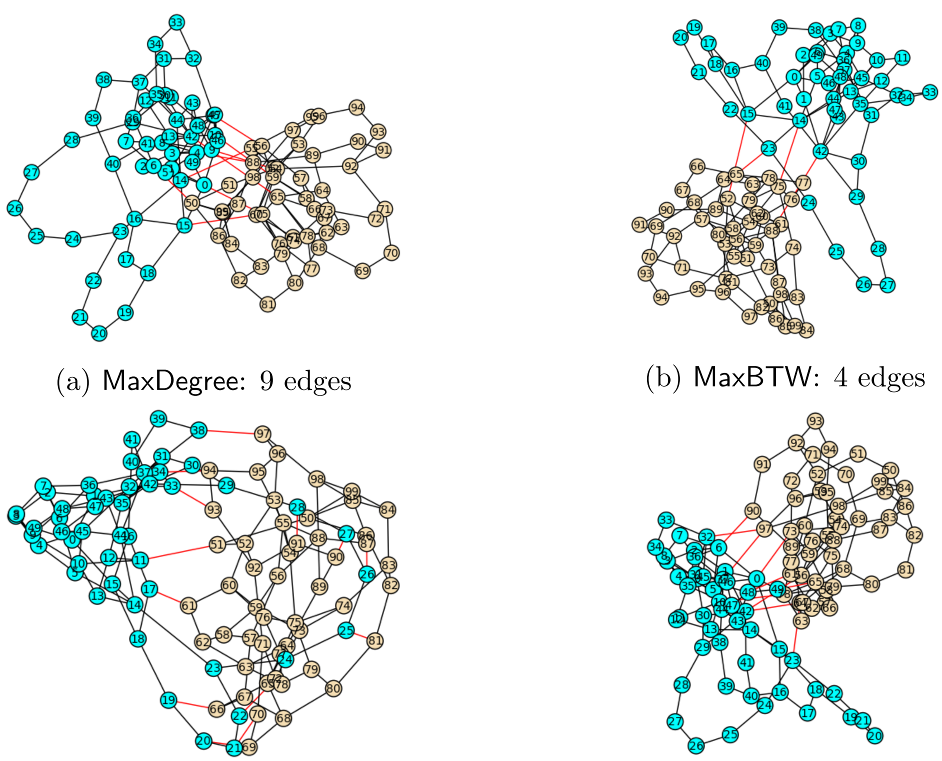
(a) For any node $w \in V_i$, let $b(w) = |\{v \in V_{3-i} \mid wv \in E'\}|$.

(b) Choose the node $u$ from $B_1 = \{w \in V_1 \mid b(w) \leq b(w') \text{ for all } w' \in V_1\}$ that has the highest priority.

(c) Choose the node $v$ from $B_2 = \{w \in V_2 \mid b(w) \leq b(w') \text{ for all } w' \in V_2\}$.

**Example 8.** To illustrate the ideas above, consider Figure 4.3. It shows the results of integrating two Newman-Watt-Strogatz random networks (see Section 4.5) with 50 nodes each using various approaches. Each of the integrated networks has diameter 9.

According to this example, while MaxBtw requires the smallest number of edges, there is a big variation in terms of the number of edges created using different priorities. It is therefore interesting to compare the results of the different heuristics in more detail.



(a) MaxDegree: 9 edges

(b) MaxBTW: 4 edges

(c) MinDegree: 14 edges

(d) Random: 10 edges

Figure 4.3: Integrating two Newman-Watts-Strogatts networks with 50 nodes to achieve diameter of 9 in the integrated network.

## 4.4 Socialization as a Special Case of Integration

When an individual joins a social network, the position he or she is going to acquire clearly depends on the quality of the established connections. The process of the new links formation is often called *socialization* [79], and has still many open questions in social science, for example, is it better to have many connections, or only few but with powerful and central individuals?, should the relationships be diverse or homogeneous?, and so on [34].

Clearly, edges play an important role: some of them can be defined as (local) bridges, others provide nodes with better connectivity. For instance, calculating betweenness, the amount of flow of information that goes through a node, is one of the most common approaches in evaluating importance of edges and nodes [39].

In this section, we consider integration of two networks from a different angle: suppose one of the networks consists of a single node. This can be interpreted as a situation when a person joins a social network.

Clearly, a social network inside this company is already formed, and the newcomer needs to establish his own connections in order to gain a certain status. In [109], the authors argue about importance of the social network formation for an individual. They mention that any network provides an individual with three advantages: access to private information, access to diverse skill sets, and power. All information is either public (for instance, available on the Internet), or private. Private information can only be acquired via network connections, and thus is more valuable.

Stuart in his works [102, 103, 104] study how the environment affects formation of strategic alliances and interorganizational collaborations. Particularly, he noticed that the focus of research has been changed from outside-the-network explanations for relationship formation (e.g., strategic alliances arise among pairs of firms with complementary resource profiles) to within-the-network explanations.

We, thus, focus on the integrated network from the perspective of a "newcomer", study the complexity of the problem, and suggest possible solutions. Personal network should be indeed carefully constructed: the number of connections is always limited, but it should provide the individual with access to information and ability to affect the network.

### 4.4.1 Network Building: the Problem Setup

We first reformulate the definition of an integrated network in terms of one network and a 'newcomer' to define network building:

**Definition 4.4.1.** *Let $G = (V, E)$ be a network and $u$ be a node not in $V$. For $S \subseteq V$, denote by $E_S$ the set of edges $\{uv \mid v \in S\}$. Define $G \oplus_S u$ as the graph $(V \cup \{u\}, E \cup E_S)$.*

We require that $S \neq \varnothing$ and thus $G \oplus_S u$ is a network built by incorporating $u$ into $G$. By [109], for a newcomer $u$ to establish herself in $G$ it is essential to identify *information brokers* who connect to diverse parts of the network. Following this intuition, we define a *broker set* as any $S \subseteq V$ such that $\mathsf{ecc}(u) = \mathsf{rad}(G \oplus_S u)$; namely, linking with $S$ enables $u$ to get in the center of the network.

Formally, given a network $G = (V, E)$, the problem of *network building for $u$* means selecting a set $S \subseteq V$ so that the combined network $G \oplus_S u$ satisfies certain conditions. Moreover, the desired set $S$ should contain as few nodes as possible. We focus on the following two key problems:

1. $\mathsf{BROKER}$: $S$ is a broker set.

2. $\mathsf{DIAM}_{\mathsf{dm}}$: $\mathsf{diam}(G \oplus_S u) \leq \mathsf{dm}$ for a given $\mathsf{dm} \leq \mathsf{diam}(G)$.

In the network $G \oplus_V u$, $\mathsf{ecc}(u) = 1 = \mathsf{rad}(G \oplus_V u)$ and $\mathsf{diam}(G \oplus_V u) = 2$. Hence, a desired $S$ must exist for $\mathsf{BROKER}$ and $\mathsf{DIAM}_{\mathsf{dm}}$ where $\mathsf{dm} \geq 2$. In the subsequent sections we systematically investigate these two problems.

### 4.4.2 Complexity and Algorithms for $\mathsf{BROKER}$

$\mathsf{BROKER}$ can be easily illustrated with an example. Suppose we have a person that joins a social network and wants to know how many new connections he or she needs to create in order to become central. We now show that this problem is in fact NP-complete.

#### Complexity

We investigate complexity of the problem $\mathsf{BROKER}(G, k)$: Given $G, k$, does $G$ have a broker set of size $k$?

The problem is trivial if $G$ has radius 1, as then $V$ is the only broker set. When $\mathsf{rad}(G) > 1$, we recall the notion of *dominating set*, which refers to a set $S$ of nodes where every node not in $S$ is adjacent to at least one member of $S$. The *domination number* $\gamma(G)$ is the size of a smallest dominating set for $G$. The $\mathsf{DOM}(G, k)$ problem concerns testing whether $\gamma(G) \leq k$ for a given graph $G$ and input $k$; it is a classical NP-complete decision problem [46].

More generally, a subset $S \subseteq V$ is called *distance $k$-dominating* if every node has distance at most $k$ with some node in $S$. We let $\gamma_k(G)$ denote the size of a smallest distance $k$-dominating set; when $k = 1$, we simply write it as $\gamma(G)$.

**Theorem 4.4.1.** $\mathsf{BROKER}(G, k)$ *is NP-complete.*

*Proof.* $\mathsf{BROKER}(G, k)$ is clearly in NP. Therefore we only show NP-hardness. We present a reduction from $\mathsf{DOM}(G, k)$ to $\mathsf{BROKER}(G, k)$. Note that when $\mathsf{rad}(G) = 1$, $\gamma(G) = 1$. Hence $\mathsf{DOM}(G, k)$ remains NP-complete if we assume $\mathsf{rad}(G) > 1$. Given a graph $G = (V, E)$ where $\mathsf{rad}(G) > 1$, we construct a graph $H$. The set of nodes in $H$ is $\{v_i \mid v \in V, 1 \leq i \leq 3\}$. The edges of $H$ are as follows:

- Add an edge $v_i v_{i+1}$ for every $v \in V$, $1 \leq i < 3$
- Add an edge $v_1 w_1$ for every $v, w \in V$
- Add an edge $v_2 w_2$ for every edge $vw \in E$

Namely, for each node $v \in V$ we create three nodes $v_1, v_2, v_3$ which form a path. We link the nodes in $\{v_1 \mid v \in V\}$ to form a complete graph, and nodes in $\{v_2 \mid v \in V\}$ to form a copy of $G$. Since $\mathsf{rad}(G) \geq 2$, for each node $v \in V$ there is $w \in V$ with $\mathsf{dist}(v, w) \geq 2$. Hence in $H$, $\mathsf{dist}(v_3, w_3) \geq 4$, and $\mathsf{dist}(v_2, w_3) \geq 3$. As the longest distance from any $v_1$ to any other node is 3, we have $\mathsf{rad}(H) = 3$.

Suppose $S$ is a dominating set of $G$. If we add all edges $uv$ where $v \in D = \{v_2 \mid v \in S\}$, $\mathsf{ecc}(u) = 3 = \mathsf{rad}(H \oplus_D u)$. Hence $D$ is a broker set for $H$. Thus the size of a minimal broker set of $H$ is at most the size of a minimal dominating set of $G$.

Conversely, for any set $D$ of nodes in $H$, define the *projection* $p(D) = \{v \mid v_i \in D \text{ for some } 1 \leq i \leq 3\}$.

Suppose $p(D)$ is not a dominating set of $G$. Then there is some $v \in V$ such that for all $w \in p(D)$, $\mathsf{dist}(v_2, w_2) \geq 2$. Thus if we add all edges in $\{ux \mid x \in D\}$, $\mathsf{dist}(u, v_3) \geq 4$.

But then $\mathsf{ecc}(w_1) = 3$ for any $w \in p(D)$. So $D$ is not a broker set. This shows that the size of a minimal dominating set of $G$ is at most the size of a minimal broker set.

Then for any $v \in V$ there is $w \in p(D)$ with $\mathsf{dist}(v, w) \leq 1$, and thus $\mathsf{dist}(v_3, w_2) \leq 2$ and $\mathsf{ecc}(u) = \mathsf{rad}(H \oplus_D u)$. Therefore the set $D$ is a broker set for $H$. On the contrary, suppose $p(D)$ is not a dominating set of $G$. Then there exists some $v \in V$ with distance at least 2 from any node in $p(D)$. This means that $\mathsf{dist}(v_3, w) \geq 3$ for all $w \in D$. Hence $\mathsf{ecc}(u) \geq 4 > \mathsf{rad}(H \oplus_D u)$. Therefore $D$ is not a broker set for $H$.

The above argument implies that the size of a minimal broker set for $H$ coincides with the size of a minimal dominating set for $G$. This finishes the reduction and hence the proof. □

#### Efficient Algorithms

Theorem 4.4.1 implies that computing optimal solution of BROKER is computationally hard. Nevertheless, we present a number of efficient algorithms that take as input a network $G = (V, E)$ with radius $\mathsf{rad}(G)$ and output a small broker set $S$ for $G$. A set $S \subseteq V$ is called *sub-radius dominating* if for all $v \in V$ not in $S$, there exists some $w \in S$ with $\mathsf{dist}(v, w) < \mathsf{rad}(G)$. Our algorithms are based on the following fact, which is clear from definition:

**Lemma 4.4.1.** *Any sub-radius dominating set is also a broker set.*

We now are ready to present efficient algorithms for solving BROKER:

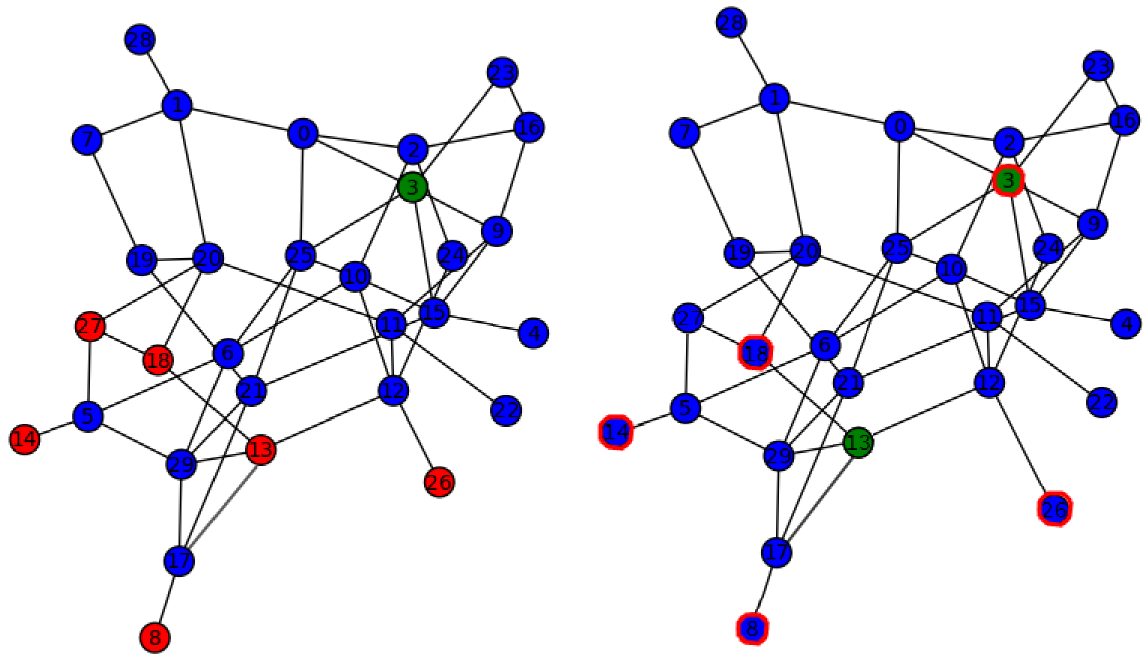
**(a) Four greedy algorithms.** In Chapter 2, we presented four greedy algorithms, Max, Min, Btw, and MinLeaf for finding small distance $k$ dominating sets, i.e., sub-radius dominating set with $\mathsf{rad}(G) = k$. Therefore, they are also suitable for solving BROKER.

**(b) Simplified greedy algorithms.** Recall that algorithms S-Max, S-Btw, S-MinLeaf act in a similar way as their "non-simplified" counterparts; the difference is that here the heuristic works over the original network $G$ as opposed to the updated network.

As an example, in Figure 4.4 we run Max and S-Max on the same network and show how S-Max may output a smaller sub-radius dominating set. The network $G$ has $\mathsf{rad}(G) = 4$. Iteration 1: Both Max and S-Max add the same green node into $S$,

$U$ contains the red nodes. S-Max outputs the green nodes $\{3, 13\}$; Max outputs the red-circled nodes $\{3, 18, 14, 8, 26\}$.

We further verify via experiments below (see Section 4.5) that the simplified algorithms lead to much smaller output $S$ in almost all cases.



(a) Pick a node with max degree (green node 3), red nodes are 'not covered'

(b) S-Max outputs the green nodes $\{3, 13\}$; Max outputs the red-circled nodes $\{3, 18, 14, 8, 26\}$

Figure 4.4: Two iterations of Max and S-Max algorithms

**(c) Center-based algorithms.** The algorithms above can all be applied to find $k$-dominating set for arbitrary $k \geq 1$. Since our focus is in finding sub-radius dominating set to answer the BROKER problem, we describe two algorithms that are specifically designed for this task. When building the network for a newcomer, it is natural to consider nodes that are already in the center of the network $G$. Hence our two algorithms are based on utilizing the center of $G$:

**Algorithm Center.** The algorithm finds a center $v$ in $G$ with minimum degree, then output all nodes that are adjacent to $v$. Since $v$ belongs to the center, for all $w \in V$, we have $\mathsf{dist}(v, w) \leq \mathsf{rad}(G)$ and thus there is $v'$ adjacent to $v$ such that $\mathsf{dist}(w, v') = \mathsf{dist}(w, v) - 1 < \mathsf{rad}(G)$. Hence, the algorithm returns a sub-radius dominating set.

Despite its apparent simplicity, Center returns surprisingly good results in many cases, as will be shown in the experiments below.

**Algorithm Imp-Center.** We present a modified version of Center, which we call Imp-Center.

**Procedure 7** Imp-Center: Given $G = (V, E)$ (with radius $\mathsf{rad}(G) = r$)

```
Pick a center node v in G with minimum degree d
Sort all adjacent nodes of v to a list u_1, u_2, ..., u_d in decreasing order of degrees
Set S ← ∅ and i ← 1
while U ≠ ∅ do
    Set C as the largest connected component in F
    if rad(C) < rad(G) − 1 then
        Pick a center node w of C. Set S ← S ∪ {w}
        Set U ← U \ {w' ∈ U | dist(w, w') < r}
    else
        Set S ← S ∪ {u_i}
        Set U ← U \ {w' ∈ U | dist(u_i, w') < r}
        Set i ← i + 1
    end if
    Set F as the subgraph induced by the current U
end while
return S
```

The algorithm first picks a center with minimum degree, and then orders all its neighbors in decreasing degree. It adds the first neighbor to $S$ and remove all nodes $\leq (r-1)$-steps from it. This may disconnect the graph into a few connected components. Take the largest component $C$. If $C$ has a smaller radius than $r$, we add the center of this component to $S$; otherwise we add the next neighbor to $S$. We then remove from $F$ all nodes at distance $\leq (r-1)$ from the newly added node. This procedure is repeated until $F$ is empty. See Procedure 7.

Figure 4.5 shows an example where Imp-Center out-performs Center.

The next theorem follows from Lemma 4.4.1.

**Theorem 4.4.2.** *Any of Algorithms in (a)–(c) outputs a broker set $S$ for $G$.*

### 4.4.3 Complexity and Algorithms for $\mathsf{DIAM}_{\mathsf{dm}}$

Let $G = (V, E)$ be a network and $u \notin V$. The $\mathsf{DIAM}_{\mathsf{dm}}$ problem asks for a set $S \subseteq V$ such that the network $G \oplus_S u$ has diameter $\leq \mathsf{dm}$; we refer to any such $S$ as dm-

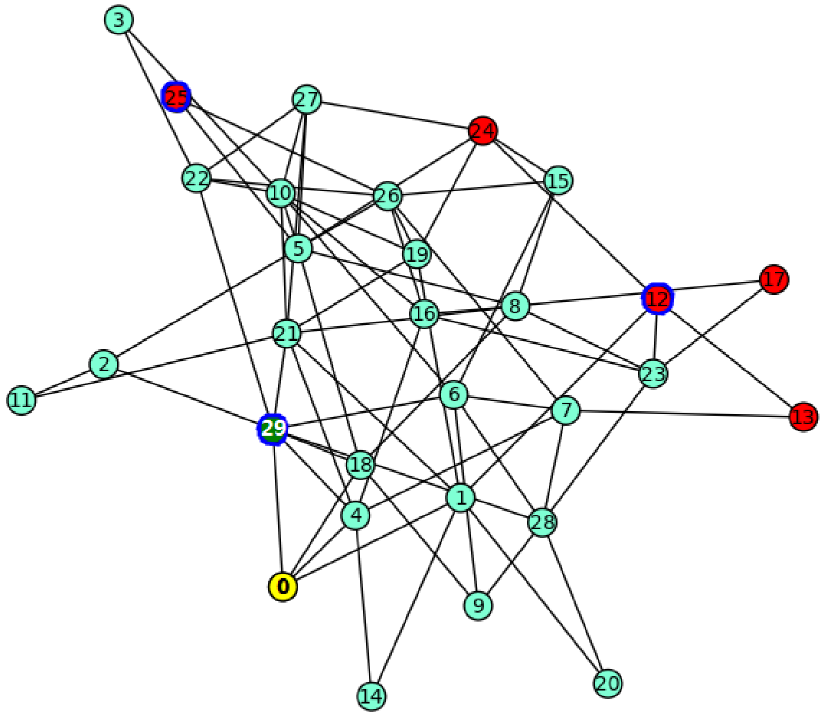



Figure 4.5: $\mathsf{rad}(G) = 3$. The yellow node 0 is a center with min degree 4. Thus, Center outputs 4 nodes. The dark green node 29 adjacent to 0 has max degree; Red nodes are “uncovered” by 29. Thus Imp-Center outputs the 3 blue circled nodes.

*enabling.*

**Preserving the diameter**

We first look at a special case when $\mathsf{dm} = \mathsf{diam}(G)$, which has a natural motivation: How can an airline expand its existing route map with an additional destination while ensuring the maximum number of hops between any two destinations is not increased? We are interested in creating as few new connections as possible to reach this goal. Let $\delta(G)$ denote the size of the smallest dm-enabling set for $G$. We say a graph is *diametrically uniform* if all nodes have the same eccentricity.

**Theorem 4.4.3.** *(a) If $G$ is not diametrically uniform, $\delta(G)=1$.*

*(b) If $G$ is complete, then $\delta(G) = |V|$.*

*(c) If $G$ is diametrically uniform and incomplete, then $1 < \delta(G) \leq d$ where $d$ is the minimum degree of any node in $G$, and the upper bound $d$ is sharp.*

*Proof.* For **(a)**, suppose $G$ is not diametrically uniform. Take any $v$ where $\mathsf{ecc}(v) <$

$\mathsf{diam}(G)$. Then in the expanded network $G \oplus_{\{v\}} u$, we have $\mathsf{ecc}(u) = \mathsf{ecc}(v) + 1 \leq \mathsf{diam}(G)$.

**(b)** is clear.

For **(c)**, Suppose $G$ is diametrically uniform and incomplete. For the lower bound, suppose $\gamma_{\mathsf{diam}(G)-1}(G) = 1$. Then there is some $v \in V$ with the following property: In the network $G \oplus_{\{v\}} u$ we have $\mathsf{ecc}(u) \leq \mathsf{diam}(G)$, which means that $\mathsf{ecc}(v) < \mathsf{diam}(G)$.

This contradicts the fact that $G$ is diametrically uniform. For the upper bound, take a node $v \in V$ with the minimum degree $d$. Let $N$ be the set of nodes adjacent to $v$. From any node $w \neq v$, there is a shortest path of length $\leq \mathsf{diam}(G)$ to $v$. This path contains a node in $N$. Hence $w$ is at distance $\leq \mathsf{diam}(G) - 1$ from some node in $N$. Furthermore as $G$ is not complete, $\mathsf{diam}(G) \geq 2$ and $v$ is at distance $1 \leq \mathsf{diam}(G) - 1$ from nodes in $N$. □

In [71] calculating the exact value of $\delta(G)$ is shown to be complete for $\mathsf{W}[2]$, second level of the $\mathsf{W}$-hierarchy. Hence $\mathsf{DIAM}_{\mathsf{dm}}$ is unlikely to be in $\mathsf{P}$. On the other hand, we argue that real-life networks are rarely diametrically uniform. Hence, by Theorem 4.4.3(a), the smallest number of new connections needed to preserve the diameter is 1.

### Reducing the diameter

We now explore the question $\mathsf{DIAM}_{\mathsf{dm}}$ where $2 \leq \mathsf{dm} < \mathsf{diam}(G)$.

This refers to the goal of placing a new member in the network and creating ties to allow a closer distance between all pairs of members. We suggest two heuristics to solve this problem: one is based on connecting two most distant vertices via the 'external' node, the other one uses a center of the graph.

The first one is very intuitive: indeed, if we start to connect the most distant vertices we eventually decrease the diameter of the graph. The second algorithm instead of taking a pair of the most distant nodes, takes only one node. Since it also uses the center of the graph, it guarantees that the new path would be $r(G)+1$ to the second node.

**Algorithm** Periphery**.** The *periphery* $P(G)$ of $G$ consists of all nodes $v$ with $\mathsf{ecc}(v) = \mathsf{diam}(G)$.

Suppose $\mathsf{diam}(G) > 2$. Then the combined network $G \oplus_{P(G)} u$ has diameter smaller than $\mathsf{diam}(G)$. Hence, we apply the following heuristic: The algorithm first adds the new node $u$ to $G$ and repeats the following procedure until the current graph has diameter $\leq \mathsf{dm}$:

1. Randomly pick a peripheral pair $v, w$ in the current graph
2. Adds the edges $uv, uw$ if they have not been added already
3. Compute the diameter of the updated graph

Note that once $v, w$ are chosen as a peripheral pair and the corresponding edges $uv, uw$ added, $v$ and $w$ will have distance 2 and they will not be chosen as a peripheral pair again. Hence, the algorithm eventually terminates and produces a graph with diameter at most $\mathsf{dm}$.

**Algorithm CP (Center-Periphery).** This algorithm applies a similar heuristic as Periphery, but instead of picking peripheral pairs at each iteration, it first picks a node $v$ in the center and adds the edge $uv$.

Then, it repeats the following procedure until the current graph has diameter $\leq \mathsf{dm}$:

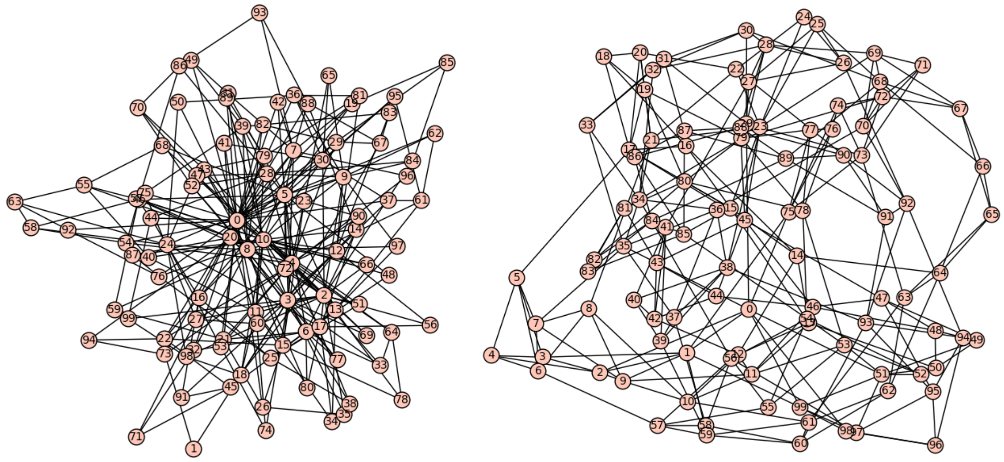
1. Randomly pick a node $w$ in the periphery of the graph
2. Add the edge $uw$ if it has not been added already
3. Compute the diameter of the updated graph

Suppose at one iteration the algorithm picks $w$ in the periphery. Then, after this iteration the eccentricity of $w$ is at most $r + 2$ where $r$ is the radius of the graph.

## 4.5 Experimental Analysis

In this section, we present experimental results obtained by implementing the proposed heuristics and algorithms for network integration.

We implemented the algorithms using Sage [100], which provides a variety of tools for working with graphs, and measured the performance of each algorithm on both artificial and real networks.



(a) BA graph $G_1$: $\mathsf{rad}(G_1) = 3$ (b) NWS graph $G_2$: $\mathsf{rad}(G_2) = 4$

Figure 4.6: Two examples of generated networks with 100 nodes

**Generated networks.** We apply two models of random graphs: The first (BA) is Barabasi-Albert's preferential attachment model which generates scale-free graphs whose degree distribution of nodes follows a power law; this is an essential property of numerous real-world networks [10]. The second (NWS) is Newman-Watts-Strogatz's small-world network [85], which produces graphs with small average path lengths and high clustering coefficient.

Examples of an NWS graph and a BA graph with 100 each nodes are in Figure 4.6.

**Real world datasets.** We test the algorithms on several real-world datasets: The Facebook dataset, collected from survey participants of Facebook App, consists of friendship relation on Facebook [74].

Enron is an email network of the company made public by the FERC [66]. Nodes of the network are email addresses and if an address $i$ sent at least one email to address $j$, the graph contains an undirected edge from $i$ to $j$.

We present a short summary of the datasets in Table 4.1

| | Facebook | Enron |
|---|---|---|
| Number of nodes | 4,039 | 33,969 |
| Number of edges | 88,234 | 180,811 |
| diameter | 8 | 13 |
| radius | 4 | 7 |

Table 4.1: Facebook and Enron datasets

| | Collaboration 1 | Collaboration 2 |
|---|---|---|
| total number of nodes | 5,242 | 9,877 |
| total number of edges | 14,496 | 25,998 |
| number of nodes in the subgraph | 4,158 | 8,638 |
| number of edges in the subgraph | 13,422 | 24,806 |
| diameter | 17 | 18 |
| radius | 9 | 10 |

Table 4.2: Collaboration 1 and Collaboration 2 datasets

Col1 and Col2 are collaboration networks that represent scientific collaborations between authors of papers submitted to General Relativity and Quantum Cosmology category (Col1), and to High Energy Physics Theory category (Col2) [65]. If an author $i$ co-authored a paper with author $j$, the graph contains a undirected edge from $i$ to $j$. Table 4.2 lists details of these networks. As both networks are initially unconnected, we considered the giant component in each graph.

### 4.5.1 Solving Network Integration Problems

In this subsection, we consider algorithms for solving the network integration problems, namely, $\mathsf{NIT}_t^\natural(G_1, G_2)$, $\mathsf{NIE}_e^\natural(G_1, G_2)$, and $\mathsf{NI}_\Delta(G_1, G_2)$. The main goals of the experiments are:

(1) to compare heuristics proposed in Section 4.2;

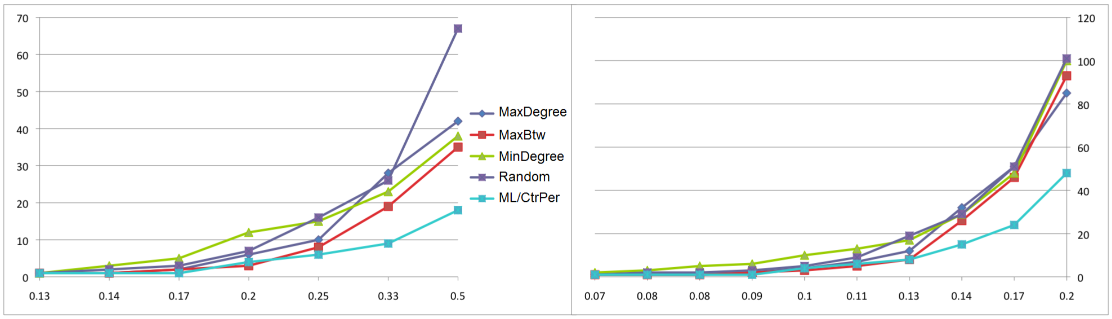
(2) to get insights into the structural properties of the integrated networks. In particular, we want to analyze how density of the input networks affects the solutions to the integration problem;

(3) to apply the heuristics onto real-world datasets and to see whether the results are consistent with our assumptions.

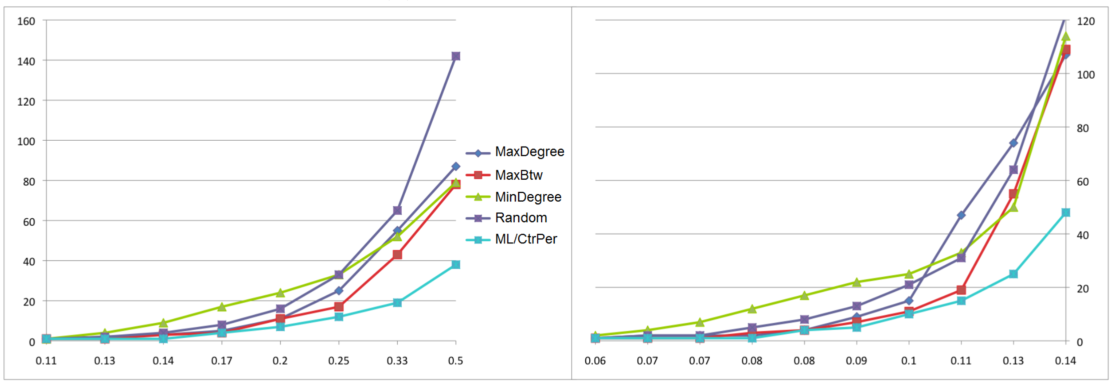
**Experiment 7. Comparing heuristics on networks with fixed $\tau^\exists$ and $\tau^\forall$.**

In the series of runs, we generated 20 pairs of the Newman-Watts-Strogattz (NWS) and the Barabasi-Albert (BA) networks with 50, 100 and 200 vertices each. For each pair we compute a solution for the $\mathsf{NIT}_t^\natural(G_1, G_2)$ problem (where $\natural \in \{\exists, \forall\}$) using MaxDegree, MinDegree, MaxBtw, Random as well as MinLeaf (when $\natural = \exists$) and CtrPer (when $\natural = \forall$). Using each of these heuristics, we add edges to the set $E$ until the integrating network $G_1 \oplus_E G_2$ satisfies a certain given parameter.

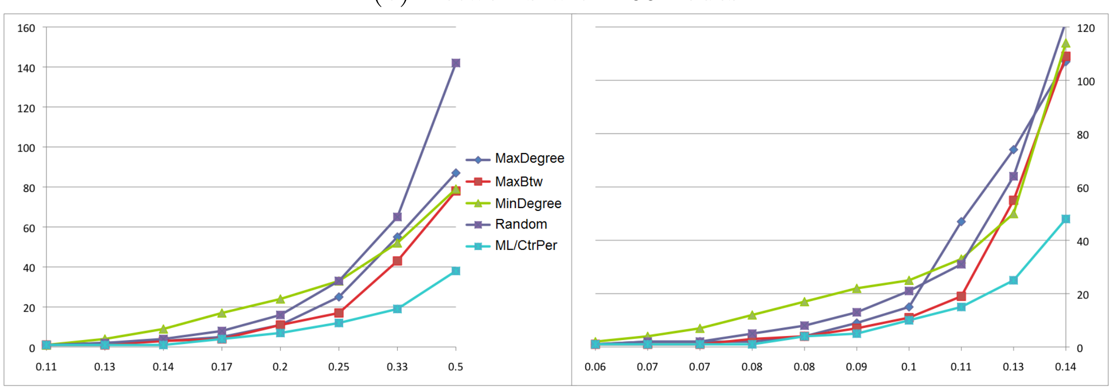
Figure 4.7 and Figure 4.8 display the average number of new edges in the solution sets for the NWS and BA networks, respectively. Results show how the number of edges increases with increasing $\tau^{\exists}$ and $\tau^{\forall}$.




(a) Networks with 50 nodes



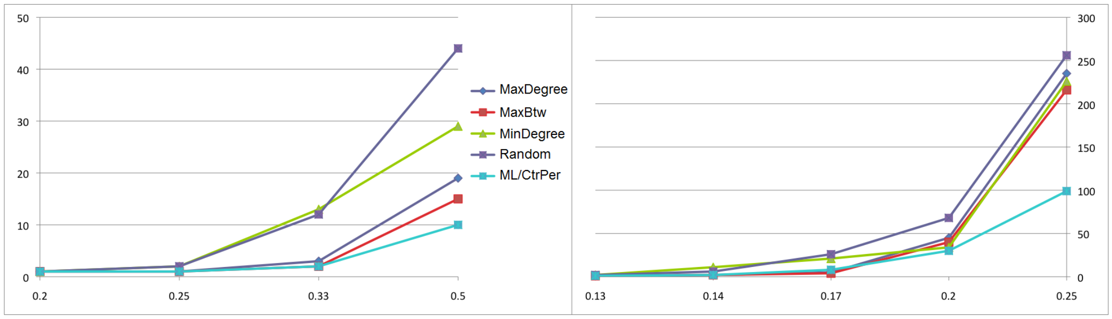

(b) Networks with 100 nodes



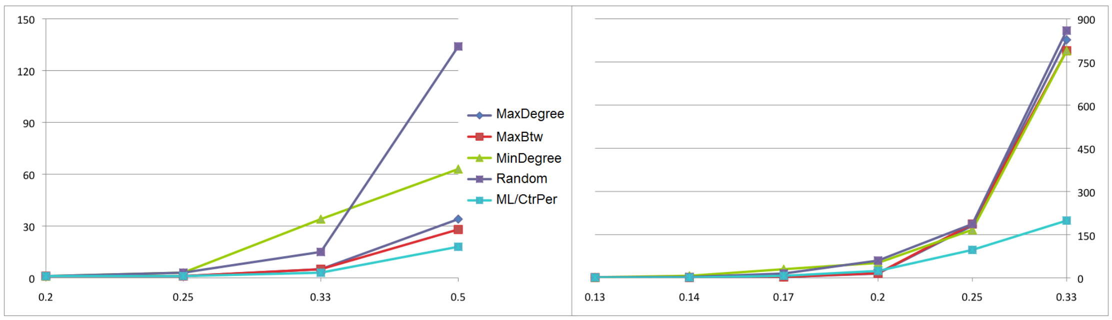

(c) Networks with 200 nodes

Figure 4.7: Comparing heuristics: average numbers of edges required to integrate two NWS networks with fixed $\tau^{\exists}$ (on the left) and fixed $\tau^{\forall}$ (on the right)

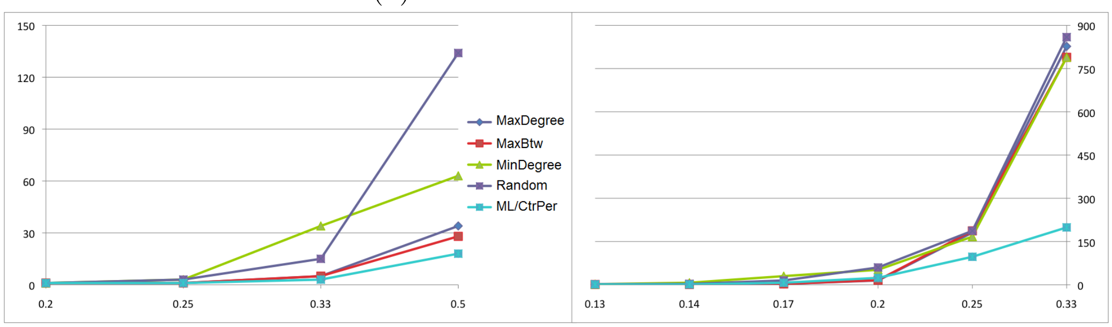
Furthermore, for small togetherness ($\tau^{\exists} \leq 0.17$ for NWS and $\tau^{\exists} \leq 0.25$ for BA, and $\tau^{\forall} > \max\{d(G_1), d(G_2)\}^{-1}$), different types priorities do not significantly affect the size of the resulting sets. However, the difference increases as togetherness increases. We conclude also that, in general, the MinLeaf and CtrPer algorithms output much smaller edge sets.




(a) Networks with 50 nodes




(b) Networks with 100 nodes




(c) Networks with 200 nodes

Figure 4.8: Comparing heuristics: average numbers of edges required to integrate two BA networks with fixed $\tau^{\exists}$ (on the left) and fixed $\tau^{\forall}$ (on the right)

**Experiment 8. Comparing heuristics for solving $\mathsf{NIE}_e^{\natural}(G_1, G_2)$.**

To perform this experiment, we generate BA and NWS networks with 50 nodes each . We set the numbers of edges $e$ to values $1, 10, 20, 50$, and then we compute the average togetherness in the integrated networks by applying different heuristics.

Figure 4.9 and Figure4.10 plot the results for NWS and BA networks, respectively.

The best performance is given by MaxBtw. In general, MinDegree gives the worst performance when $e$ is small. However, its performance catches up with other heuristics when $e$ becomes larger. On the contrary, Random has an opposite behavior: togetherness grows slower as more edges are randomly added.

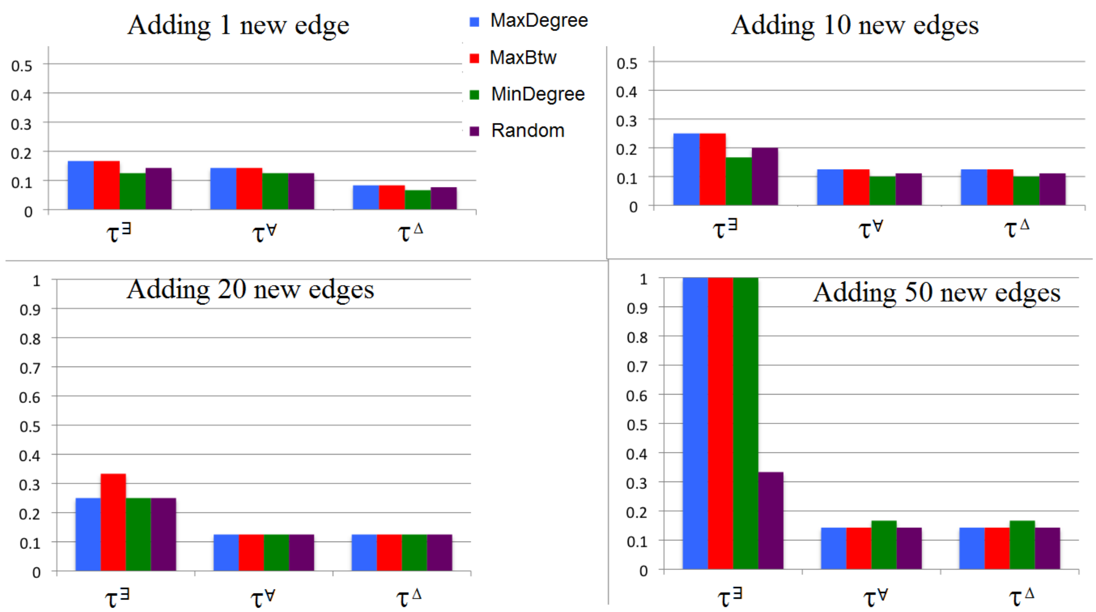



Figure 4.9: Integrating NWS networks by establishing 1, 10, 20 and 50 edges

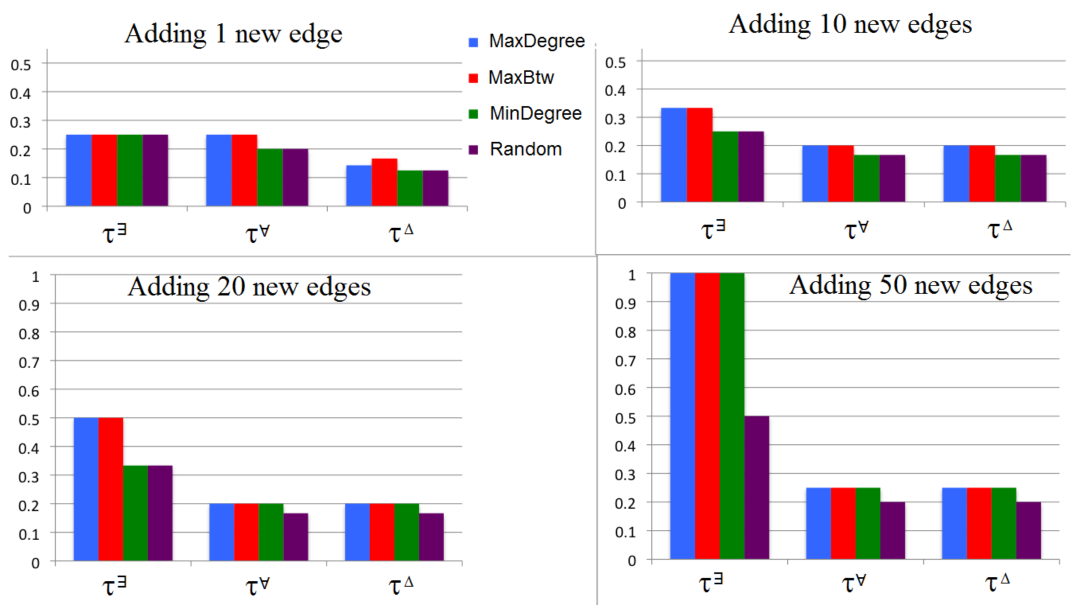



Figure 4.10: Integrating BA networks by establishing 1, 10, 20 and 50 edges

**Experiment 9. Integrating real networks with equally privileged nodes.**

We use two real datasets to reconfirm the results obtained for the synthesized datasets. Col1 and Col2 are networks that represent scientific collaborations in General Relativity, Quantum Cosmology (Col1), and in High Energy Physics Theory (Col2) [65] (See Table 4.2).

We first apply the MinLeaf and CtrPer algorithms. Structurally Col1 and Col2 resembles the small-world networks of NWS models. Thus, we expect that when integrating these two networks, despite the large number of nodes and edges, the number of new edges would be relatively small.

In Figure 4.11, we show how the togetherness measures $\tau^{\exists}$ and $\tau^{\forall}$ change with the number of added edges. The weaker togetherness notion, $\tau^{\exists}$, grows faster than $\tau^{\forall}$: decreasing the $\forall$-togetherness is harder. Consistent with our expectation, we notice that very few edges are required when integrating these large collaboration networks.

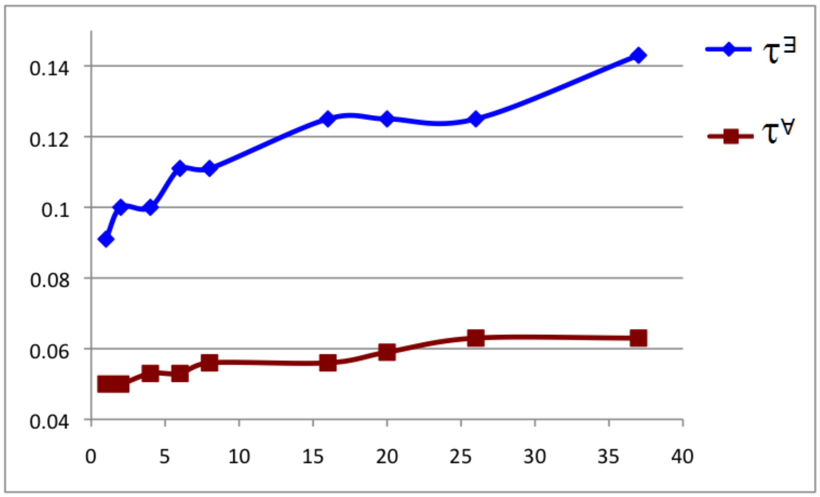



Figure 4.11: Integrating two collaboration networks with $\tau^{\exists}$ and $\tau^{\forall}$ constraints

**Decreasing diameter: solving $\mathsf{NI}_{\Delta}(G_1, G_2)$**

We now compare the algorithms for solving $\mathsf{NI}_{\Delta}(G_1, G_2)$ network integration problem: we consider $\mathsf{Integrate}_{\mathsf{dm}}(G_1, G_2)$ algorithm and the $\mathsf{Naive}_{\Delta}(G_1, G_2)$ algorithm and test their performance in terms of running time and size of the output solutions. We analyze our algorithms on synthesized as well as real-world datasets.

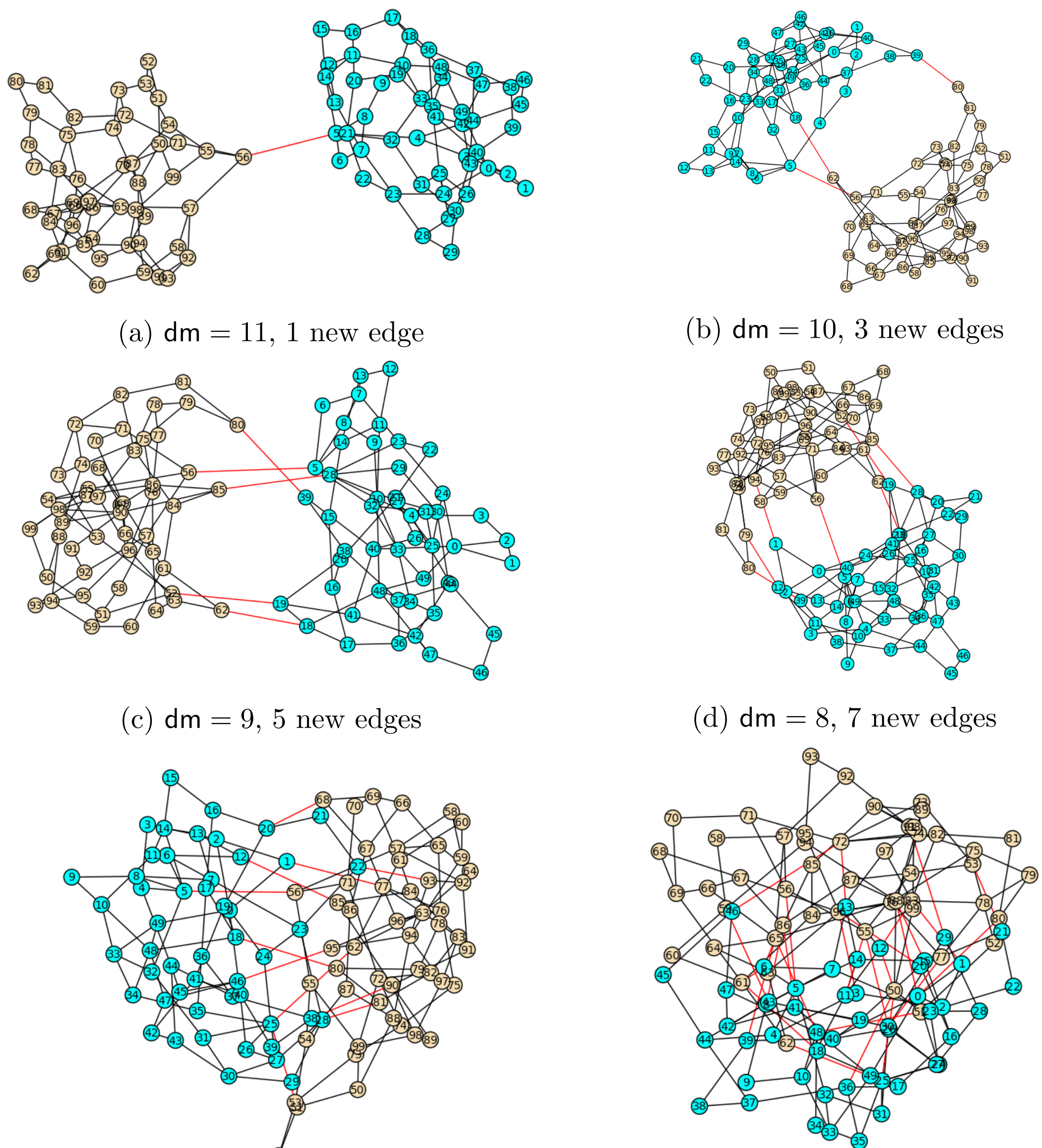
We also analyze how density of the input networks affects the solutions to the integration problem. Indeed, consider the following example:

**Example 9.** As an example, we integrate two NWS graphs as well as two BA graphs using the $\mathsf{Integrate}_{\mathsf{dm}}(G_1, G_2)$ algorithm. The statistics for each graph is shown in Table 4.3.

In Figure 4.12 and Figure 4.13, we integrate two NWS and two BA graphs, respectively, using the $\mathsf{Integrate}_{\mathsf{dm}}(G_1, G_2)$ algorithm. For the NWS graphs, we set parameter $\mathsf{dm}$ ranges from 6 to 11, while for the BA graphs, $\mathsf{dm}$ ranges from 4 to 7.

| | NWS Graph 1 | NWS Graph 2 | BA Graph 1 | BA Graph 2 |
|---|---|---|---|---|
| Number of nodes | 50 | 50 | 50 | 50 |
| Number of edges | 77 | 78 | 141 | 141 |
| Diameter | 7 | 8 | 4 | 4 |
| Radius | 5 | 5 | 3 | 3 |

Table 4.3: Two pairs of generated BA and NWS random networks



(a) dm = 11, 1 new edge

(b) dm = 10, 3 new edges

(c) dm = 9, 5 new edges

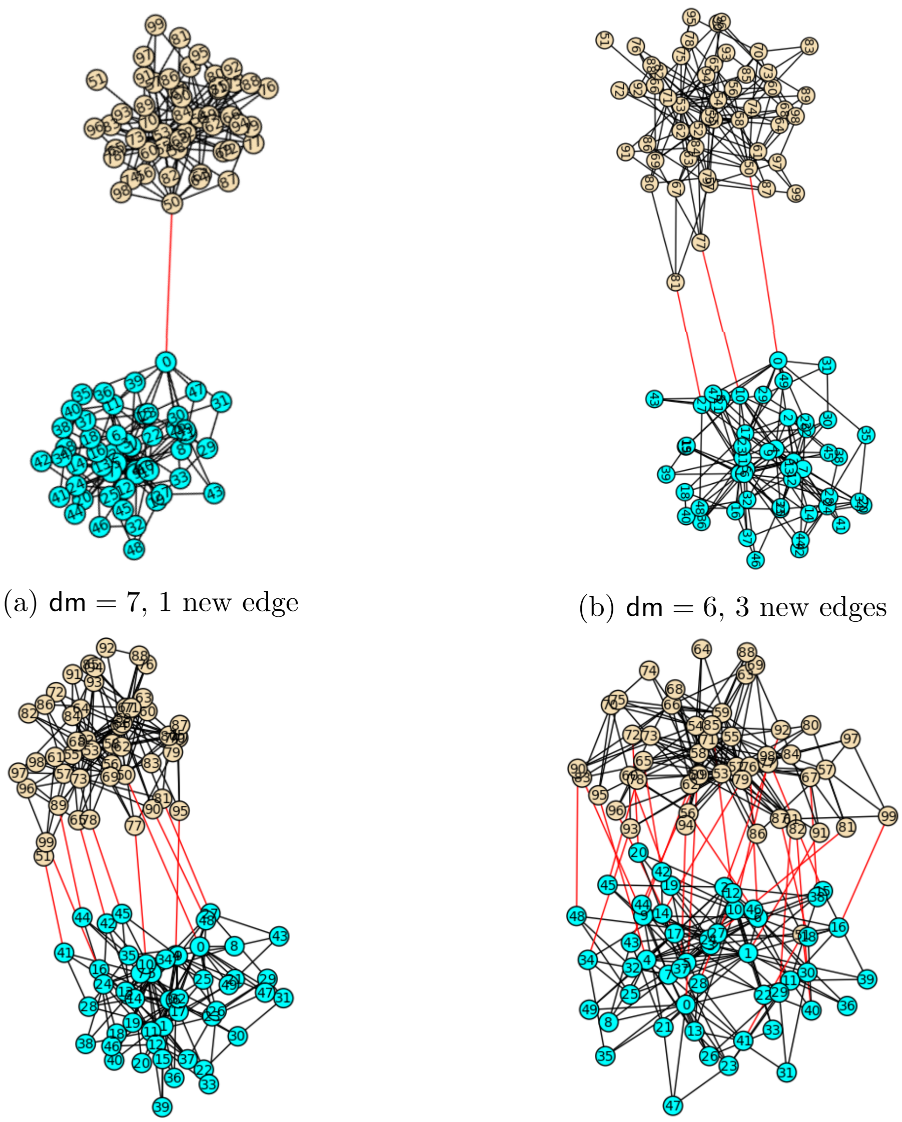
(d) dm = 8, 7 new edges

(e) dm = 7, 11 new edges

(f) improved diameter, dm = 6, 29 new edges

Figure 4.12: Integrating two NWS networks with different diameter dm

Both figures show how with decreasing parameter dm the networks gradually dissolve in each other: when very few edges exist between the networks, the initial struc-



(a) $\mathsf{dm} = 7$, 1 new edge

(b) $\mathsf{dm} = 6$, 3 new edges

(c) $\mathsf{dm} = 5$, 10 new edges

(d) $\mathsf{dm} = 4$, 30 new edges

Figure 4.13: Integrating two BA networks with different parameter $\mathsf{dm}$

ture of the graphs is clearly visible, however, the more edges we create the more united the networks become.

One may also notice from these examples that the number of edges to decrease the diameter by one grows faster for Barabasi-Albert networks. We indeed will show this tendency in the experiments below.

In Section 4.2, we computed bounds on the running time of $\mathsf{Naive}_\Delta(G_1, G_2)$ and $\mathsf{Integrate}_{\mathsf{dm}}(G_1, G_2)$ algorithms. We now show experimental results by executing and applying these heuristics on generated networks.

**Experiment 10.** $\mathsf{Naive}_\Delta(G_1, G_2)$ **and** $\mathsf{Integrate}_{\mathsf{dm}}(G_1, G_2)$**: comparing running times.**

We implement both algorithms $\mathsf{Naive}_\Delta(G_1, G_2)$ and $\mathsf{Integrate}_{\mathsf{dm}}(G_1, G_2)$ and record their running times on 300 generated NWS and BA networks. The results indicate that $\mathsf{Integrate}_{\mathsf{dm}}(G_1, G_2)$ outperforms $\mathsf{Naive}_\Delta(G_1, G_2)$ significantly, with the former runs more than 3000 times faster on networks with 1000 nodes to add a single edge in the solution set.

Figure 4.14 plots how much longer (on average) $\mathsf{Naive}_\Delta(G_1, G_2)$ takes to add a single edge to the solution set compared to $\mathsf{Integrate}_{\mathsf{dm}}(G_1, G_2)$, against the number of nodes in the networks. As one may see, on graphs with 1000 nodes, this difference becomes significant.

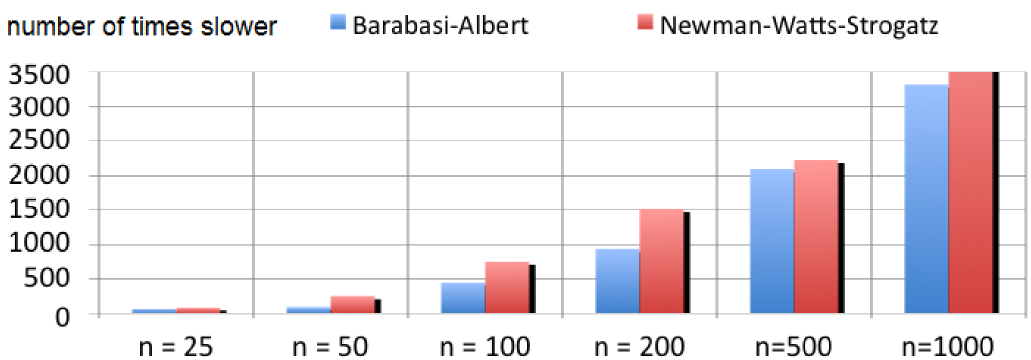



Figure 4.14: The number of times $\mathsf{Naive}_\Delta(G_1, G_2)$ runs slower than $\mathsf{Integrate}_{\mathsf{dm}}(G_1, G_2)$

**Experiment 11.** $\mathsf{Naive}_\Delta(G_1, G_2)$ **and** $\mathsf{Integrate}_{\mathsf{dm}}(G_1, G_2)$**: comparing solution size.**

We compare the output of $\mathsf{Integrate}_{\mathsf{dm}}(G_1, G_2)$ against the $\mathsf{Naive}_\Delta(G_1, G_2)$ algorithm. While $\mathsf{Naive}_\Delta(G_1, G_2)$ may output smaller solutions when $\mathsf{dm}$ is large, $\mathsf{Integrate}_{\mathsf{dm}}(G_1, G_2)$ is more likely to produce smaller solutions as $\mathsf{dm}$ decreases.

Figure 4.15 plots the percentage of the cases where $\mathsf{Integrate}_{\mathsf{dm}}(G_1, G_2)$ returns smaller sets.

Note that $\mathsf{Integrate}_{\mathsf{dm}}(G_1, G_2)$ almost always returns smaller sets whenever $\mathsf{dm} < \max\{\mathsf{diam}(G_1), \mathsf{diam}(G_2)\}$. Figure 4.16 plots the average output size of $\mathsf{Integrate}_{\mathsf{dm}}(G_1, G_2)$ and $\mathsf{Naive}_\Delta(G_1, G_2)$, against absolute and relative values of $\mathsf{dm}$. Here, each graph consists of 100 nodes. Even though $\mathsf{Naive}_\Delta(G_1, G_2)$ may outperform $\mathsf{Integrate}_{\mathsf{dm}}(G_1, G_2)$ when $\mathsf{dm}$ is large, the difference is not very significant; as $\mathsf{dm}$ decreases, the advantage of $\mathsf{Integrate}_{\mathsf{dm}}(G_1, G_2)$ becomes increasingly significant.

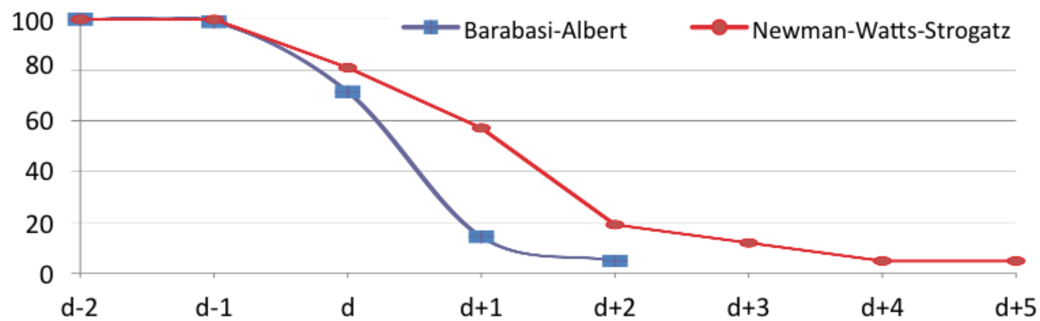
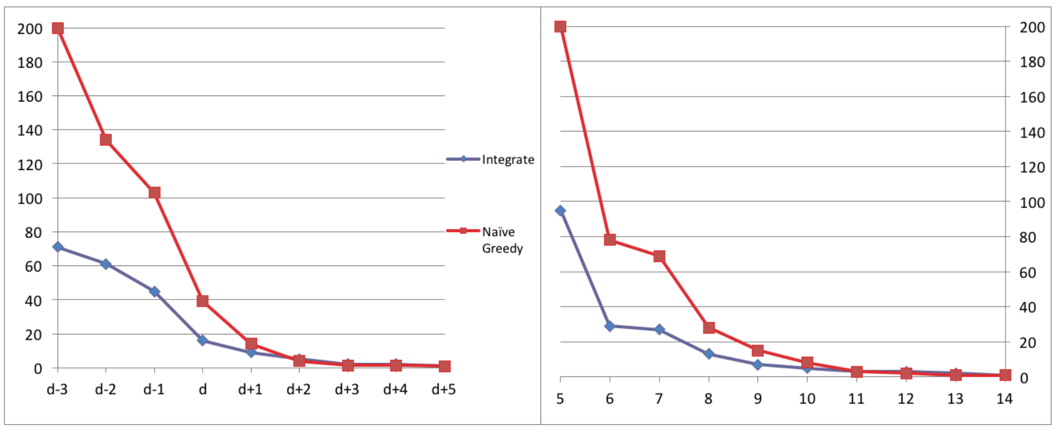
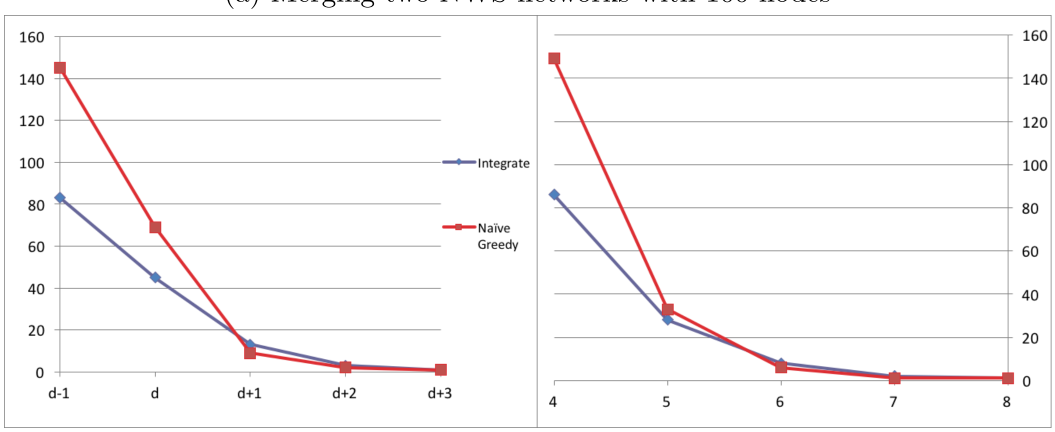



Figure 4.15: The probability that $\mathsf{Integrate}_{\mathsf{dm}}(G_1, G_2)$ outputs smaller sets with varying $\mathsf{dm} \in \{d-2, \ldots, d+5\}$ where $d = \max\{\mathsf{diam}(G_1), \mathsf{diam}(G_2)\}$




(a) Merging two NWS networks with 100 nodes




(b) Merging two BA networks with 100 nodes

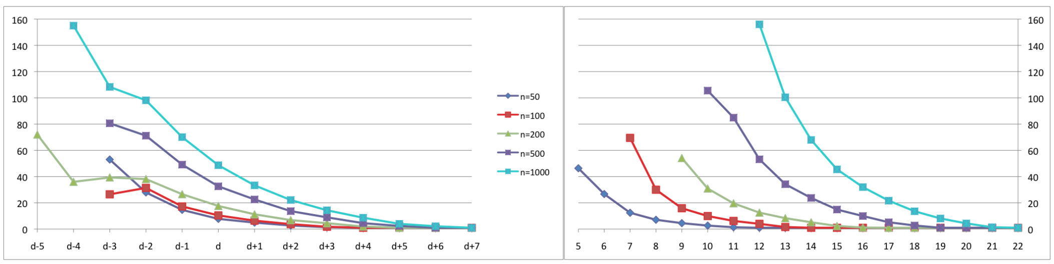
Figure 4.16: Comparing the $\mathsf{Integrate}_{\mathsf{dm}}(G_1, G_2)$ algorithm and the $\mathsf{Naive}_{\Delta}(G_1, G_2)$ algorithm: average number of edges with different parameter $\mathsf{dm}$

.

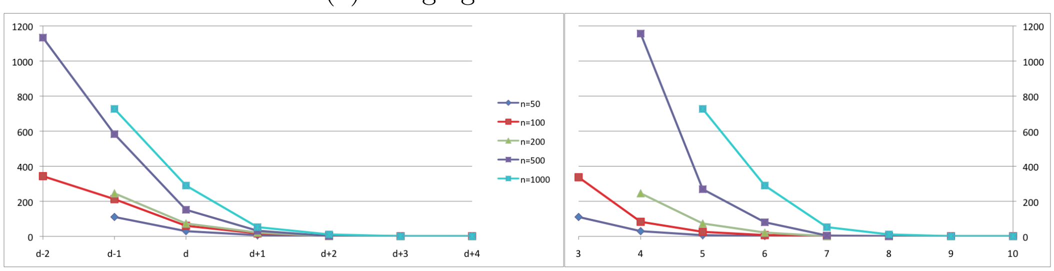
**Experiment 12. Solving $\mathsf{NI}_{\Delta}(G_1, G_2)$: structural analysis.**

The BA networks are usually dense and have smaller diameter compared to the NWS networks with the same number of nodes. We would like to understand how this difference affects network integration.

Figure 4.17 plots the average output size of $\mathsf{Integrate}_{\mathsf{dm}}(G_1, G_2)$ on both types of random networks with different $\mathsf{dm}$. It shows that less edges are needed to reduce the diameter of the integrated network to a certain $\mathsf{dm} \geq \max\{\mathsf{diam}(G_1), \mathsf{diam}(G_2)\}$ for BA graphs than for NWS graphs. On the other hand, if $\mathsf{dm} < \max\{\mathsf{diam}(G_1), \mathsf{diam}(G_2)\}$, the number of edges required for NWS graphs becomes significantly less than that for BA graphs. e.g., when $\mathsf{dm} = \max\{\mathsf{diam}(G_1), \mathsf{diam}(G_2)\} - 1$, the BA graphs requires about ten times more edges than NWS graphs.




(a) Merging two NWS networks




(b) Merging two BA networks

Figure 4.17: The average number of edges for integrating networks with $n$ nodes: applying $\mathsf{Integrate}_{\mathsf{dm}}(G_1, G_2)$

**Experiment 13. Applying $\mathsf{Integrate}_{\mathsf{dm}}(G_1, G_2)$ on real world datasets.**

To reconfirm results obtained for random networks, we integrate two real world collaboration networks, which represent paper co-authorship between scientists in physical science [65]. Both networks are initially unconnected. For the sake of experiments, we considered the largest connected components (see Table 4.2).

Note that the networks are rather sparse: their diameters are 17 and 18, and the radiuses are 9 and 10. The networks that satisfy the small-world principles, satisfy also the six degrees of separation principle, that is every person is six or fewer steps away from any other person in the network.

We assume that Col1 and Col2 are more similar to the Newman-Strogatz-Watts

| | | $\mathsf{rad}(G_1)+\mathsf{rad}(G_2)$ | $\max\{\mathsf{diam}(G_1), \mathsf{diam}(G_2)\}$ | | | |
|---|---|---|---|---|---|---|
| Diameter $\mathsf{dm}$ | 20 | 19 | 18 | 17 | 16 | 15 |
| new edges | 1 | 4 | 8 | 20 | 26 | 84 |

Table 4.4: Integrating two collaboration networks: applying $\mathsf{Integrate}_{\mathsf{dm}}(G_1, G_2)$

graphs. Thus we expect that when merging these two networks, despite the large number of nodes and edges, the number of resulting edges would be relatively small.

We set the bounds for $\mathsf{dm}$ as following: $\min\{\mathsf{diam}(G_1), \mathsf{diam}(G_2) - 2\} \leq \mathsf{dm} \leq \mathsf{rad}(G_1) + r(G_2) + 1$, or that is $15 \leq \mathsf{dm} \leq 20$. As proved earlier, when $\mathsf{dm} \geq 20$, i.e. $\mathsf{dm} > \mathsf{rad}(G_1) + \mathsf{rad}(G_2)$, a unique edge between two centers allows to create a desired network.

Then, we run the $\mathsf{Integrate}_{\mathsf{dm}}(G_1, G_2)$ algorithm, and got the following results: when $\mathsf{dm} = 19$, we create only 4 new edges, when $\mathsf{dm} = \max\{\mathsf{diam}(G_1), \mathsf{diam}(G_2)\} = 18$, the algorithm returns 8 new edges. When $\mathsf{dm} = \max\{\mathsf{diam}(G_1), \mathsf{diam}(G_2)\} - 1 = 17$, i.e. we improve the diameter by one, only 20 edges are required. For $\mathsf{dm} = 16$, one needs 26 edges, and to make the diameter $\mathsf{dm} = 15$, one needs 84 edges. In Table 4.4, we collected the results on connecting two collaboration networks with different diameters.

Surprisingly, the results show that very few edges are required when merging these large collaboration networks. Even when the diameter is improved by three, i.e. $\mathsf{dm} = \max\{\mathsf{diam}(G_1), \mathsf{diam}(G_2)\} - 3$, we still need relatively few edges. Thus, we conclude that results support our proposition that on sparse graphs one needs relatively small number of edges to improve the diameter.

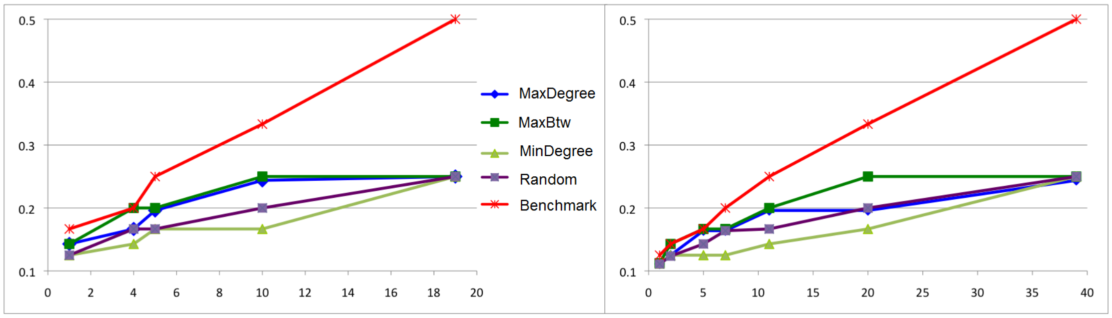
### 4.5.2 Priority Based Methods

In this subsection, we compare performance of priority based heuristics for integrating two homogeneous networks. We consider both, generated and real-world datasets.

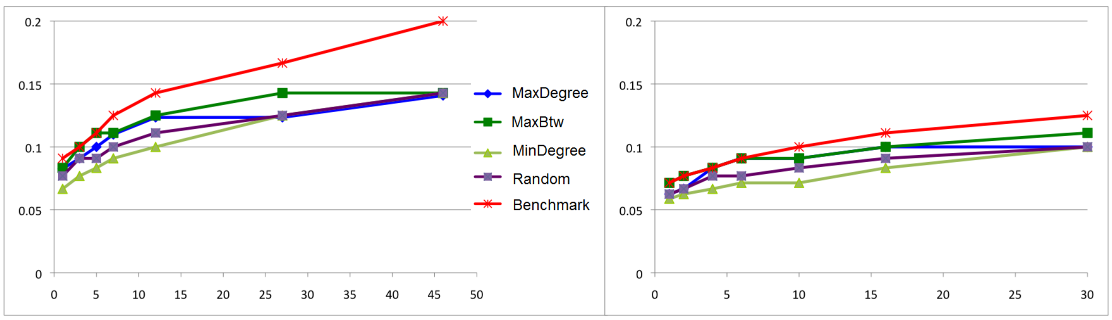
**Experiment 14. Comparing the performance of the priority based methods.**

As MinLeaf and CtrPer algorithms in general give small solution sets for the integration problems, we first apply them and use the resulting solution size as *benchmarks* to test the performance of the priority based methods.

For each value of $\exists$- and $\forall$-togetherness, we calculate the average number $e$ of edges in the output solution sets. Then, we apply the priority based heuristics to compare the result of these methods against the benchmarks. The resulting togetherness (as well as the benchmarks) are plotted in Figure 4.18 and Figure 4.19.




(a) Average values of $\tau^{\exists}$ for NWS networks




(b) Average values of $\tau^{\forall}$ for NWS networks

Figure 4.18: Integrating NWS networks with 50 nodes (on the left) and 100 nodes (on the right)

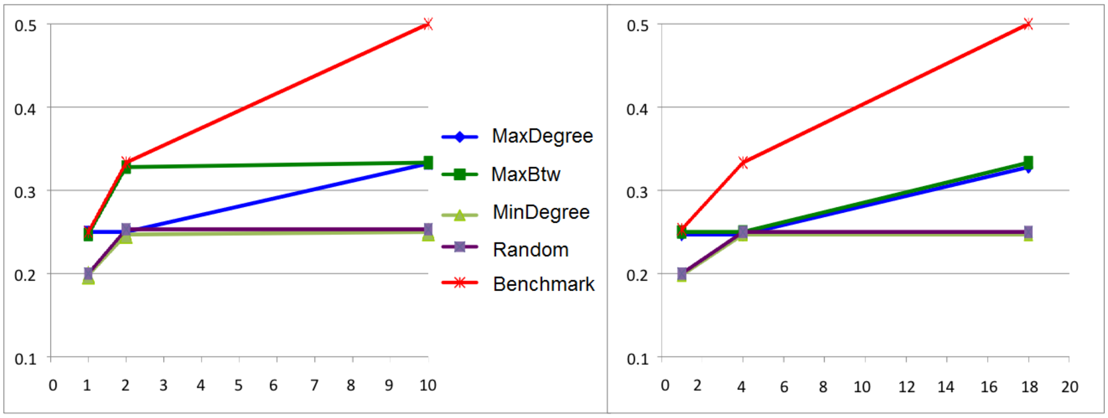
The results show that, when we add a small number of edges, the priority based heuristics perform well: the MaxBtw method results in the same togetherness as the benchmark. The MinDegree method, as in Experiment 3, proves to be the worst for small number of edges, however performs better when more edges are added.

Rather surprisingly, integrating networks with the random strategy often produce solutions that are comparable with the other strategy.

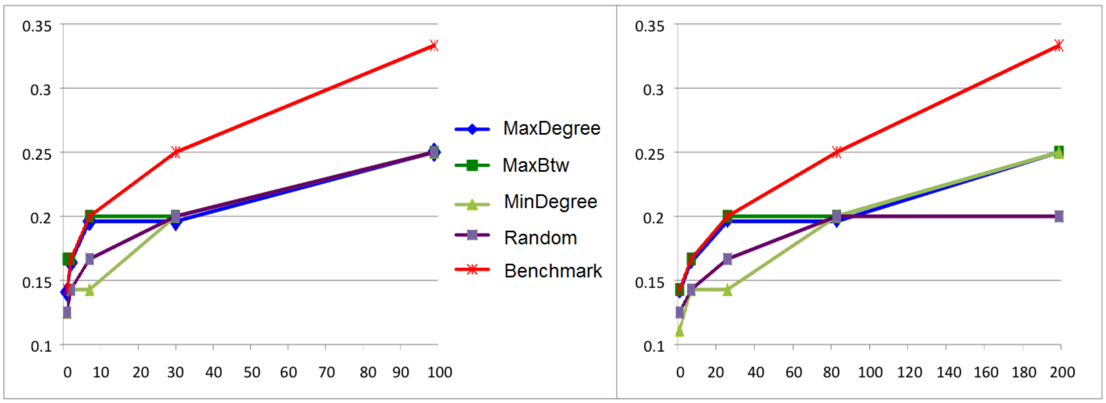
**Experiment 15. Integrating real networks: priority based heuristics**

We then apply the priority based methods to the networks Col1 and Col2. Similarly to Experiment 9, we fix the number of added edges according the benchmarks provided by MinLeaf and CtrPer and then apply the different priority based methods. The results are plotted in Figure 4.20.

Here again, we conclude that in general, the Random method gives comparable performance against other priority based strategies.




(a) Average values of $\tau^{\exists}$ for BA networks




(b) Average values of $\tau^{\forall}$ for BA networks

Figure 4.19: Integrating BA networks with 50 nodes (on the left) and 100 nodes (on the right)

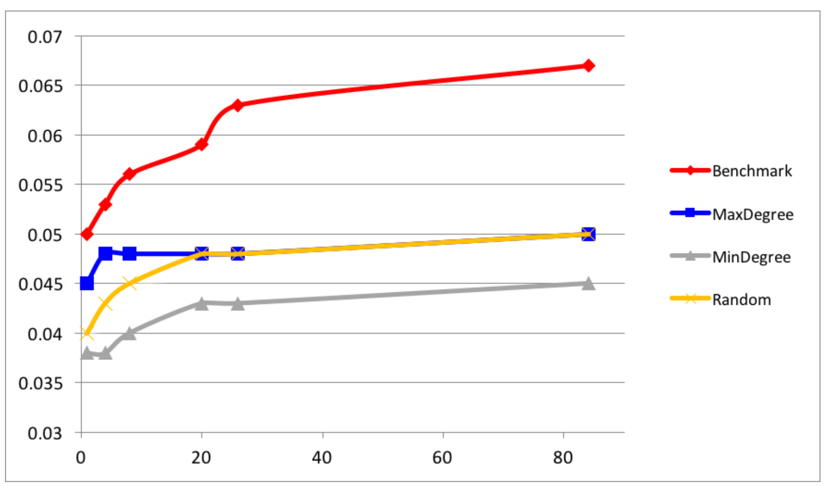



Figure 4.20: Integrating two collaboration networks: comparing different strategies

## 4.5.3 Solving BROKER and DIAM

### (a) Experiments for BROKER

For each algorithm we are interested in two indicators of its performance:

1. *Output size*: The average size of the output broker set (for a specific class of random graphs).

2. *Optimality rate*: The probability that the algorithm gives optimal broker set for a random graph. To compute this we need to first compute the size of an optimal broker set (by brute force) and count the number of times the algorithm produces optimal solution for the generated graphs.

**Experiment 16. Comparing greedy algorithms: output sizes.**

We generate 300 graphs whose numbers of nodes vary between 100 and 1000 using each random graph model. We compute averaged output sizes of generated graphs by their number of nodes $n$ and radius $rad$. The results are shown in Figure 4.21.

We make two main conclusions following this experiment:

(a) The simplified algorithms produce significantly smaller broker sets compared to their unsimplified counterparts. This shows superiority of the simplified algorithms. For instance, applied to the Barabasi-Albert graphs with 1000 nodes and radius 3, Max results in 168 new ties on average while S-Max only results in 56; Min results in 199 new ties while S-Min improves this to 41.

(b) BA graphs in general allow smaller output set than NWS graphs. This may be due to the scale-free property which results in high skewness of the degree distribution.

**Experiment 17. Comparing algorithms for solving BROKER: optimality rates.**

For the second goal, we compute the optimality rates of algorithms when applied to random graphs.

The results are shown in Figure 4.22. For BA graphs, the simplified algorithm S-MinLeaf has significantly higher optimality rate ($\geq 85\%$) than other algorithms. On the contrary, its unsimplified counterpart MinLeaf has the worst optimality rate. This is somewhat contrary to Duckworth and Man's work showing MinLeaf gives very small solution set for regular graphs [38]. The second efficient algorithm for Barabasi-Albert graph is Imp-Center.

For NWS graphs, several algorithms have almost equal optimality rate. The three best algorithms are S-Max, S-Btw and S-MinLeaf which have varying performance for graphs with different sizes (See Figure 4.23).

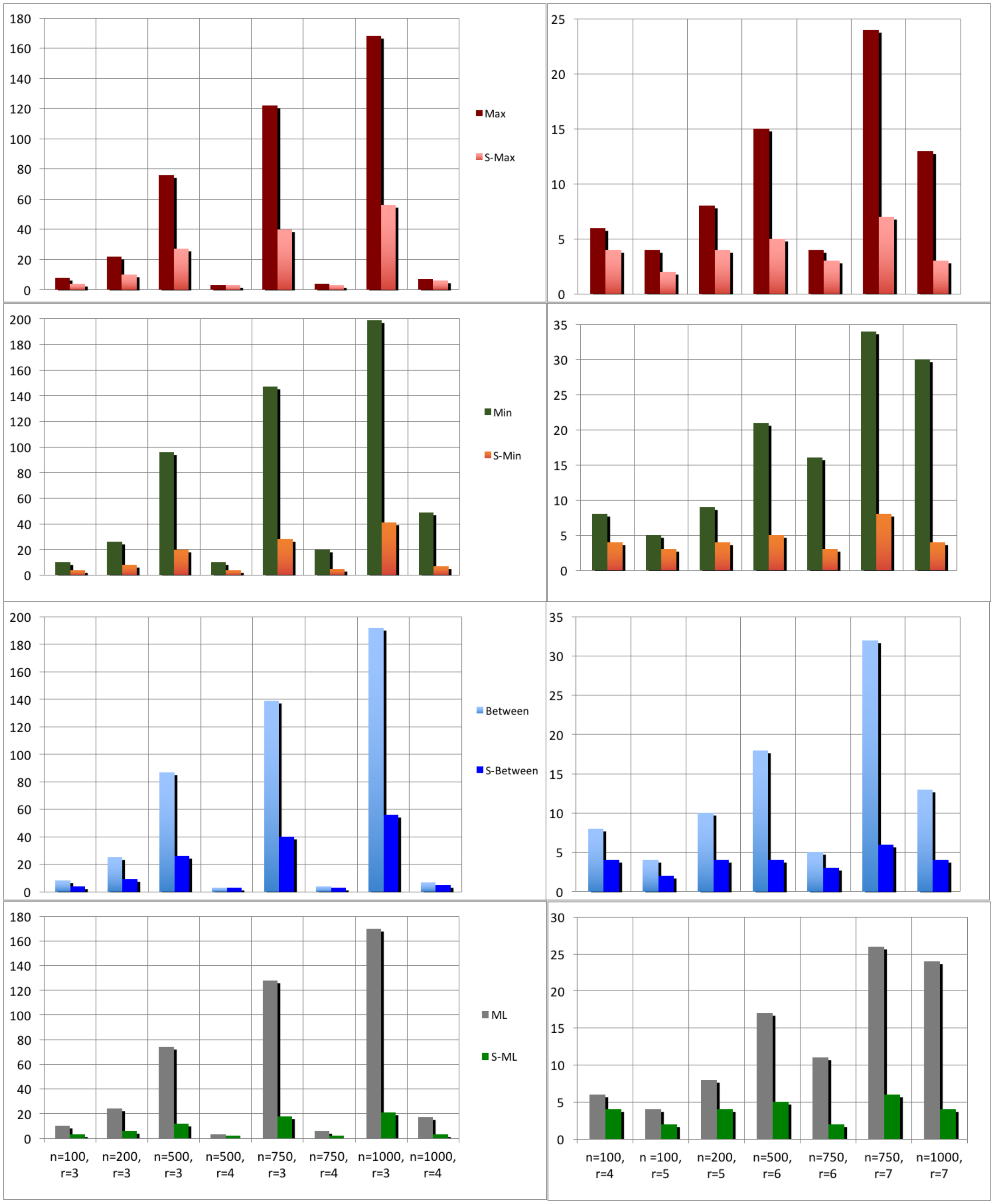



Figure 4.21: Comparing results: average performance of the Max, Min, Btw and MinLeaf algorithms versus their simplified versions on randomly generated graphs (BA graphs on the left; NWS on the right)

**Experiment 18. Experiment 12. Solving BROKER on real-world datasets.**

Results on the datasets are shown in Figure 4.24. Btw and S-Btw become too inefficient as it requires computing shortest paths between all pairs in each iteration.

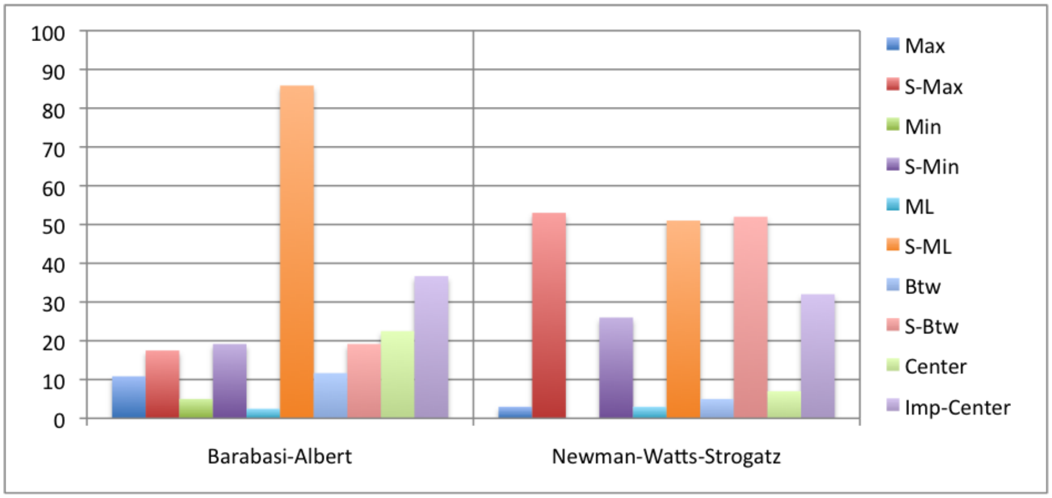



Figure 4.22: Optimality rates for different types of random graphs

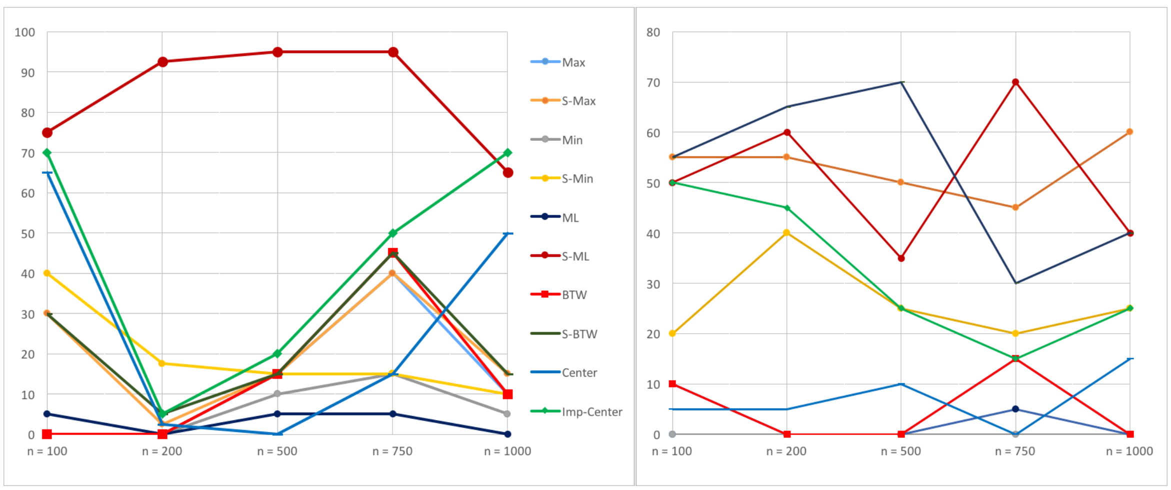



Figure 4.23: Optimality rates when graphs are classified by sizes: BA on the left and NWS on the right

Moreover, S-Max also did not terminate within reasonable time for the Enron dataset. Even though the datasets have many nodes, the output sizes are in fact very small (within 10). For instance, the smallest output sets of the Enron, Col1 and Col2 contain just two nodes.

In some sense, it means that to become in the center even in a large social network, it is often enough to establish only very few connections.

Among all algorithm Imp-Center has the best performance, producing the smallest output set for all networks. Moreover, for Enron, Col1 and Col2, Imp-Center returns the optimal broker set with cardinality 2. A rather surprising fact is, despite straightforward seemingly-naive logic, Center also produces small outputs in three networks.

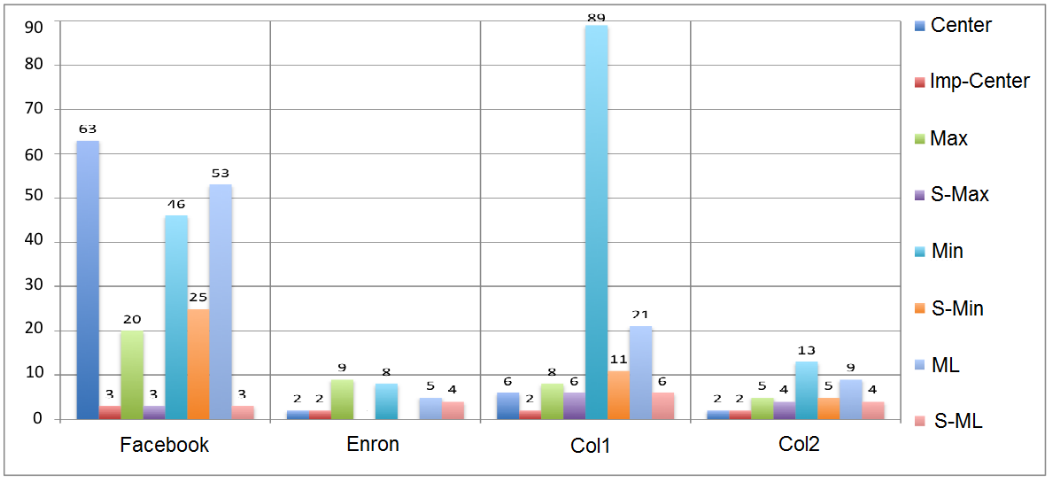



Figure 4.24: The number of new ties for the real-world networks

This reflects the fact that in order to become central it is often a good strategy to create ties with the friends of a central person.

### (b) Experiments for $\mathsf{DIAM}_{\mathsf{dm}}$

We implement and test the performance of the algorithms for solving $\mathsf{DIAM}_{\mathsf{dm}}$.Their performance are measured by the number of new ties created.

**Experiment 19. Solving $\mathsf{DIAM}_{\mathsf{dm}}$ for generated graphs.**

We generated 350 graphs and considered the case when $\mathsf{dm} = \mathsf{diam}(G) - 1$, i.e. the aim was to improve the diameter by one.

For both types of random graphs (fixing size and radius), the average number of new ties are shown in Figure 4.25.

The experiments show that Periphery performs better when the radius of the graph is close to the diameter (when radius is $> 2/3$ of diameter), whilst CP is slightly better when the radius is significantly smaller than the diameter.

**Experiment 20. $\mathsf{DIAM}_{\mathsf{dm}}$ on real-world datasets.**

We run both Periphery and CP on the networks Col1 and Col2 introduced above, setting $\mathsf{dm} = \mathsf{diam}(G) - i$ for $1 \leq i \leq 4$. The numbers of new edges obtained by Periphery and CP are shown in Figure 4.26; naturally for increasing $i$, more ties need to be created.

We point out that, despite the large total number of nodes, one needs less than 19 new edges to improve the diameter even by four. This reveals an interesting phe-

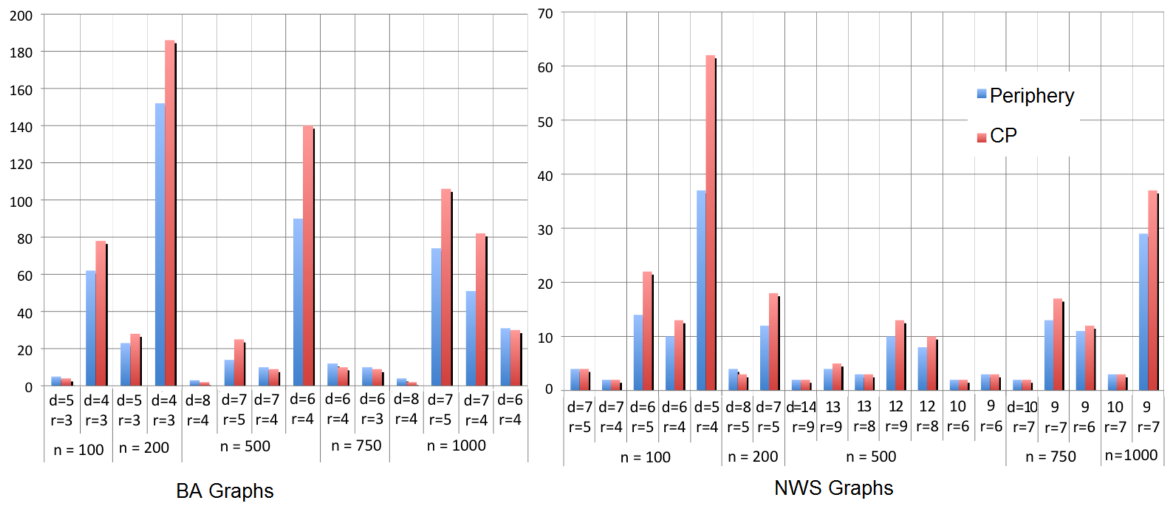



Figure 4.25: Comparing two methods for improving diameter applied to BA (left) and NWS (right) graphs

nomenon: While a collaboration network may be large, a few more collaborations are sufficient to reduce the diameter of the network.

On the Facebook dataset, Periphery is significantly better than CP: To reduce the diameter of this network from 8 to 7, Periphery requires 2 edges while CP requires 47. When one wants to reach the diameter 6, the numbers of new edges increase to 6 for Periphery and 208 for CP.

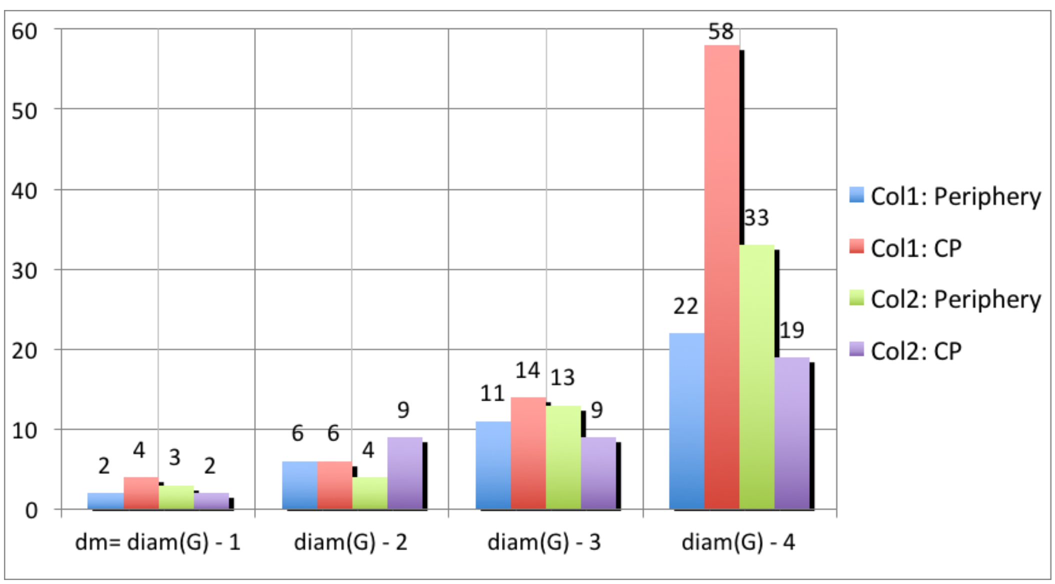



Figure 4.26: Applying algorithms for improving diameter to Collaboration 1 and Collaboration 2 datasets

# Chapter 5

# Integration of Two Organizational Networks

When two social groups merge, new relations need to be set up, which are often weak ties between these groups [49]. Integrating two companies is a process of the similar nature: take as an example a merger between two companies. As discussed by the authors of [1], the success of mergers and acquisitions of companies often hinges on whether firms can socialize employees effectively into the merged new entity. Therefore, a big challenge faced by the top managers of both companies is how to establish links between the two companies to ensure coherence and efficient communication.

In this chapter, we consider organizational networks with two types of ties that represent formal hierarchical relations and informal non-hierarchical relations such as collaboration or friendship. Motivated by different scenarios, we define two approaches: *collaborative* and *dominant* integration. The vital difference between the approaches is that the collaborative integration is established by informal ties only. As a result, the networks are still independent. On the other hand, in the process of the dominant integration, one of the networks becomes the other network's subnetwork. Such integration is established by one formal and several informal ties.

To evaluate the effect of integration, we revisit our notion of togetherness. We argue that all edges, regardless of their type, serve as channels of communication between people [94]. Formal organizational structure is designed to perform a function of delivering commands; commands can also go through informal networks taking the form of advice [61]. However, a directive would unlikely pass from a subordinate to his or her manager: indeed, information may travel in any direction, while the presence of authority suggests that orders may not go up the hierarchical structure.

Universal, existential, and diametric togetherness, as defined in Chapter 4, are designed to capture the proximity of two networks; in the context of organizations these measures could be used to indicate how fast information gets from one network to the other. To capture how fast a command can reach individuals in a certain department, we introduce a new level of togetherness, *hierarchical* togetherness.

Finally, we extend the applicability of togetherness measures to reveal how well a certain department is "integrated" in the entire organization. Thus, togetherness in this chapter is considered as a measure of interaction between a certain unit with the rest of the network, or more generally, between two organizational networks after they establish some new relationships.

## 5.1 Understanding of Network Integration for Organizations

In Chapter 4, we considered interaction and integration of two homogeneous 'flat' networks – networks that are represented by connected undirected graphs with a single type of relationships. In general, given two networks $G_1 = (V_1, E_1)$ and $G_2 = (V_2, E_2)$, an *integrated network* $G_1 \oplus_E G_2 = (V_1 \cup V_2, E_1 \cup E_2 \cup E)$ is built by establishing edges $E \subseteq V_1 \otimes V_2$ between the two networks. To extend this idea to organizational networks, one needs to define the set of edges $E$, bearing in mind the traits of organizations.

Our model of organizational structures takes into account formal and informal relations between individuals. Formal ties reflect organizational hierarchy and serve as channels to transfer commands and directives. If two individuals are connected by a formal link, it means that one of them has authority over the other one. Informal ties, on the other hand, represent mutual relationships such as collaboration or friendship; these relations are symmetrical. Through informal channels individuals communicate as equal and exchange knowledge, ideas, and experience. The presence of formal hierarchical structure is what makes organizational networks different from homogeneous networks considered in Chapter 4. Indeed, if we ignore the formal hierarchy, the integration problem would be exactly the same as in the previous chapter.

All edges, regardless of their type, serve as channels of communication between people [94]. If vertices $u$ and $v$ are connected by a formal edge $\overrightarrow{(u,v)}$, not only it means that $u$ has authority over $v$, but also that these two nodes communicate with each other. On the other hand, formal organizational structure is designed to perform a function of delivering commands; moreover, commands can also go through informal networks taking the form of advice [61]. Thus, we keep in mind these two functions: communication and transmission of directives.

Similarly, two organizations may be connected by interorganizational ties that serve as channels of communications, or signify that both organizations become two parts of a single entity with a smooth distribution of orders. We, thus, explore two scenarios.

Imagine we have two companies. The reasons to initiate an integration may be broadly split into two parts:

(1) **Mergers and acquisitions**: one of the companies becomes the other one's unit. As it is very hard to model an actual merger, we stick to the rule that both organizations preserve their initial structure: all existing relations are kept, only new one can be added.

(2) **Collaboration**: the companies want to establish new working relations but no actual consolidation occurs.

Following these scenarios, we define *dominant* and *collaborative* integration.

The main difference between these two types of integration is the resulting integrated network. Informally, collaborative integration may be regarded as establishing temporary project connections (informal edges), while dominant integration involves more serious structural changes. Indeed, in the course of dominant integration, one of the networks (dominated network) becomes the other network's (dominating network) subnetwork. To preserve the property of having only one manager, we link the root of the dominated network with a node in the dominating network by a formal edge. Then, we create as many informal edges as required to satisfy certain conditions.

We now give formal definitions of collaborative and dominant integration. Suppose we have two organizational networks $\mathcal{G}_1 = (V_1, r_1, E_1)$ and $\mathcal{G}_2 = (V_2, r_2, E_2)$ such that the sets of edges $E_1$, $E_2$ contain both formal and informal edges.

**Definition 5.1.1.** *Let $\mathcal{G}_1 \oplus_E \mathcal{G}_2 = (V_1 \cup V_2, E_1 \cup E_2 \cup E)$ be an integrated network. We say that the integration is* dominant *if there exists exactly one formal edge $\overrightarrow{(u,v)} \in E$*

*where $u \in V_1$, $v \in V_2$, and the set of edges $E \subseteq V_1 \otimes V_2$. Moreover, we say that $\mathcal{G}_1$ is the* dominating *subnetwork, and $\mathcal{G}_2$ is the* dominated *subnetwork in $\mathcal{G}_1 \oplus \mathcal{G}_2$.*

Clearly, a dominant integrated network $\mathcal{G}_1 \oplus_E \mathcal{G}_2$ is a new organizational network with root $r_1$.

On the other hand, a collaborative integration signifies establishing new links in order to encourage interaction and cooperation between two organizational networks. However, the networks are still independent, and their formal and informal structures are preserved. Thus, we do not create formal edges; however, we create as many informal edges as required to satisfy certain conditions.

**Definition 5.1.2.** *Let $\mathcal{G}_1 \oplus_E \mathcal{G}_2 = (V_1 \cup V_2, E_1 \cup E_2 \cup E)$ be an integrated network. We say that the integration is* collaborative *if all edges in set $E \subseteq V_1 \otimes V_2$ are informal.*

As the result of collaborative integrating, none of the given networks stands out in the hierarchy. Moreover, a collaborative integrated network $\mathcal{G}_1 \oplus_E \mathcal{G}_2$ is not an organizational network according to our definition as it has two roots (See Remark 1 in Chapter 2.)

Thus, we reason that the collaborative integration should be treated the same as integration of homogeneous networks with a single type of relationships. In this case, formal ties should be regarded as ordinary channels of communication rather than sources of directions.

**Remark 1.** We acknowledge that mergers and acquisitions usually result in rather more significant changes, such as restructuring or downsizing, than just creating new interorganizational ties [98]. However, in this chapter our primarily goal is to develop a model that considers a simpler case with preserving initial structures including formal and informal relationships.

## 5.2 Togetherness in Organizational Networks

In this section, we revisit the notion of togetherness, which we defined in Chapter 4.

Universal, existential, and diametric togetherness, as defined in Chapter 4, are designed to capture the proximity of two network; in the context of organizations these measures could be used to indicate how fast information gets from one network to the other. To capture how fast a command can reach individuals in a certain department, we introduce a new level of togetherness, *hierarchical* togetherness.

The first level, *universal togetherness*, reflects the principle of six degrees of separation, also known as the 'small-world' phenomenon. We argue that the same reasoning could be applied with respect to two departments (or two companies) that desire to establish collaboration: this measure captures how close all members of one network are to all members in the other network.

There also could be a different scenario: suppose workers in one of the departments possess a certain resource, which all (or some) other employees want to be able to easily acquire. Then, the first network contains the 'target' subset of vertices: we want to make sure that each node in the second network is able to reach one of the vertices in the subset within small number of steps. Then, *existential togetherness* should be applied.

These two measures could be used for any organizational network as well as any other connected (homogeneous) network. Moreover, to define a collaborative integration, all edges including formal may be regarded as undirected. Indeed, any communication is a mutual relationship, regardless how much authority one individual has over the other one.

On the other hand, we also need a measure that captures how fast a directive from a manager can reach his or her subordinates. We thus introduce a new level of togetherness, which is suitable for organizational networks only.

We note that directed edges reflect the formal structure of an organization, building the managerial hierarchy. As most of the social networks have bounded diameter (small-world properties), any organizational network has a bounded hierarchical height [113]. Moreover, the diameter may not be important at all in organizational networks: indeed, managers does not want to be too far from their subordinates, but subordinates may not necessarily be close to each other. Some of the departments (subnetworks) may require more control from the top manager (the root). Thus, to commend the traits of organizational networks and to define another level of togetherness, we consider the distances from any node in the subnetwork to the root.

In contrast to Chapter 4, we propose a more universal approach to togetherness: it is suitable not only to measure the proximity of two networks, but also to evaluate the effect of interaction between a connected subnetwork (for example, a single department) and the entire organization.

### 5.2.1 Togetherness as a Local Measure of Proximity

Let $\mathcal{G} = (V, r, E_{\mathsf{fml}} \cup E_{\mathsf{inf}})$ be an organizational network with root node $r \in V$. We call connected graph $\mathcal{G}' = (V', r', E'_{\mathsf{fml}} \cup E'_{\mathsf{inf}})$ a *subnetwork* of $\mathcal{G}$ if $V' \subset V$, $E'_{\mathsf{fml}} \subset E_{\mathsf{fml}}$, and $E'_{\mathsf{inf}} \subset E_{\mathsf{inf}}$.

As we discussed above, we need measures to evaluate 1) the proximity of a subnetwork to other members of the network, and 2) the proximity of a manager to the subnetwork. The first approach is based on measuring communication, while the second one approximates how quick commands reach this subnetwork.

Suppose we have an organizational network $\mathcal{G} = (V, r, E_{\mathsf{fml}} \cup E_{\mathsf{inf}})$. We build an undirected copy of this network by creating the set of *undirected* edges $E$ such that $(u, v) \in E$ if either $(u, v) \in E_{\mathsf{inf}}$, or $(u, v) \in E_{\mathsf{fml}}$, or $(v, u) \in E_{\mathsf{fml}}$. We say that the graph $\mathcal{G}_{\mathsf{com}} = (V, E)$ is a *communication network* of $\mathcal{G}$. Indeed, this undirected graph reflects all channels of communication between individuals, based on formal and informal relationships.

**Definition 5.2.1** (Togetherness as a measure of communication)**.** *Let $\mathcal{G} = (V, E_{\mathsf{fml}} \cup E_{\mathsf{inf}})$ be an organizational network, and let $V' \subset V$ be a subset of vertices that form the* subnetwork *$\mathcal{G}'$. Then, we define following measures of **togetherness** $\tau^{\natural}$:*

1. Existential togetherness $\exists$ *of $\mathcal{G}'$ is the smallest positive number $0 < \tau^{\exists} \leq 1$ such that for any $u \in V \setminus V'$ there exists $v \in V'$: $(\mathsf{dist}(u, v))^{-1} \leq \tau^{\exists}$ in the communication network $\mathcal{G}_{\mathsf{com}}$.*

2. Universal togetherness $\forall$ *of $\mathcal{G}'$ is the smallest positive number $0 < \tau^{\forall} \leq 1$ such that for any $v \in V'$, for any $u \in V \setminus V'$ : $(\mathsf{dist}(v, u))^{-1} \leq \tau^{\forall}$ in the communication network $\mathcal{G}_{\mathsf{com}}$.*

On the other hand, taking into account the direction of formal ties, may reveal insights into how a directive from the root node reaches individuals in the subnetwork.

**Definition 5.2.2** (Togetherness as a measure of control)**.** *Let $\mathcal{G} = (V, r, E_{\mathsf{fml}} \cup E_{\mathsf{inf}})$ be an organizational network where $E_{\mathsf{fml}}$ are directed edges and $E_{\mathsf{inf}}$ are undirected. Let $\mathcal{G}'$ be the* subnetwork *of $\mathcal{G}$ with set of vertices $V' \subset V$.*

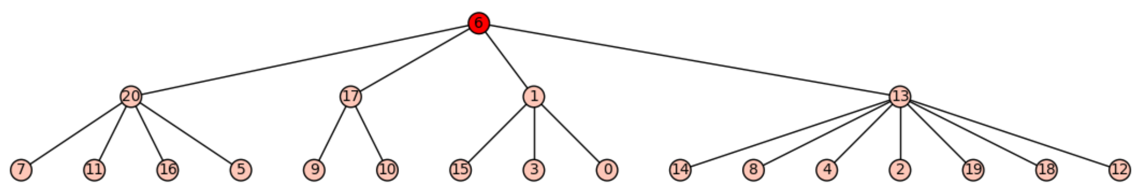
Hierarchical togetherness $h$ *of* $\mathcal{G}'$ *is the smallest positive number* $\tau^h$ *such that for any* $v \in V'$ : $(\mathsf{dist}(r, v))^{-1} \leq \tau^h$.

We illustrate these measures with a real-world example.

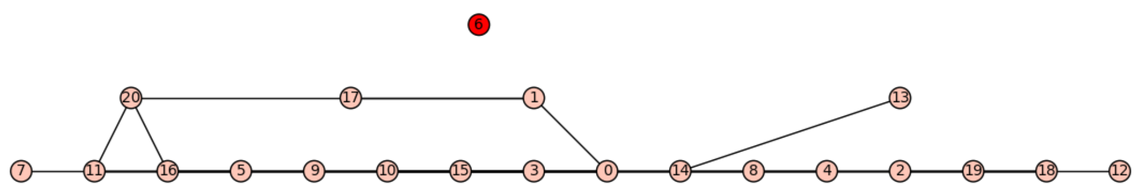
**Example 10.** We consider a small real world dataset, Krackhardt and Hansons hierarchy [61]. Krackhardt and Hansons studied a high-tech company with 21 managers. They used a formal hierarchy as well as reconstructed two social networks by interviewing the managers - one is the advice network (based on criterion "To whom do you go for advice?"), and the other one is the friendship network ("Who is your friend?").

We combine the formal hierarchy and the friendship network to obtain Krackhard organizational network. Then, we compute togetherness with respect to different subnetworks on this organizational network.

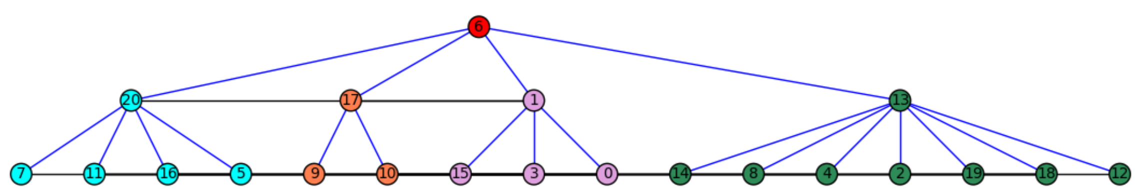
In Figure 5.1, we show the formal hierarchy, the friendship network, and the resulting organizational network.




(a) Reporting hierarchy




(b) Friendship network




(c) (Communication) organizational network: combining formal and informal relations

Figure 5.1: Different levels of Krackhardt and Hansons dataset

Consider subnetwork $\mathcal{G}_1 \subset \mathcal{G}$ such that $V_1 = \{20, 7, 16, 5, 11\}$. Then, $\tau^{\exists} = \frac{1}{3}$, $\tau^{\forall} = \frac{1}{4}$, $\tau^h = \frac{1}{2}$. Comparing these results with togetherness of the other departments (depicted in different colors in Figure 5.1), we found that all other departments are equally located and have the same values of togetherness as $\mathcal{G}_1$.

However, a single edge could change this for the department led by 17: it is easy to see that for $\mathcal{G}_2$ with $V_2 = \{17, 9, 10\}$, the universal togetherness could be improved by establishing a new edge $(17, 13)$: then, $\tau^{\forall} = \frac{1}{3}$ for $\mathcal{G}_2$ with respect to $\mathcal{G}$. This suggests an approach of improving togetherness of a certain department by creating additional informal links.

A well-known phenomenon in sociology, called *homophily*, explains that people tend to make connections with those who are similar to them. Workplace homophily may be interpreted as the fact that employees are more likely to establish social connections with people in the same unit (i.e. department, office, etc.) as well as at the same level [76]. However there might exist some informal links that connect people from different levels. To capture this phenomenon and to define some of the bounds on togetherness, we introduce the notion of *shortcuts*.

**Definition 5.2.3.** *Let $\mathcal{G} = (V, r, E_{\mathsf{fml}} \cup E_{\mathsf{inf}})$ be an organizational network. For any two nodes $v$, $u$ such that $u$ is a non-direct subordinate of $v$, we call a geodesic path from $u$ to $v$ a* shortcut *if its length smaller than the path from $v$ to $u$ using edges from $E_{\mathsf{fml}}$ only.*

**Example 11.** In Figure 5.2, consider a simple organizational network. The subnetwork $\mathcal{G}' \subset \mathcal{G}$ (depicted in red) has $\tau^h = \frac{1}{3}$. The universal togetherness of $\mathcal{G}'$ is $\tau^{\forall} = \frac{1}{4}$ since the distances $\mathsf{dist}(4, 9)$; and the existential togetherness is $\tau^{\exists} = \frac{1}{2}$. Moreover, there is a single shortcut, the edge $(1, 10) \in E_{\mathsf{inf}}$.

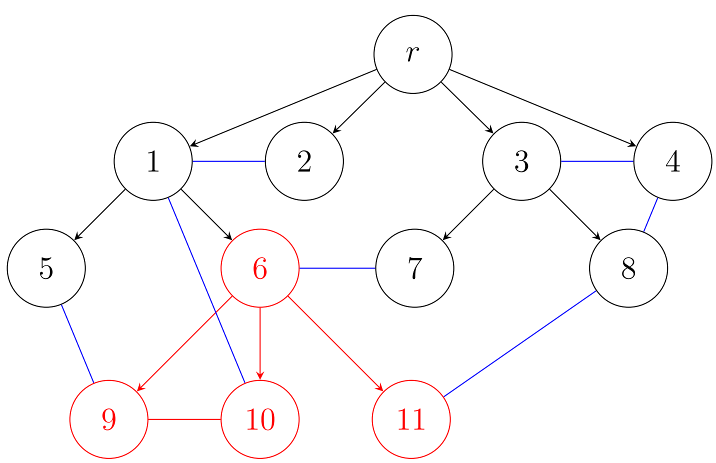



Figure 5.2: Togetherness of the subnetwork $\mathcal{G}'$ (red): $\tau^h = \frac{1}{3}$, $\tau^{\forall} = \frac{1}{4}$, $\tau^{\exists} = \frac{1}{2}$

Clearly, if an organizational network contains only formal ties, there are no shortcuts.

**Theorem 5.2.1** (Bounds on togetherness for subnetworks)**.** *Let* $\mathcal{G} = (V, r, E_{\mathsf{fml}} \cup E_{\mathsf{inf}})$ *be an organizational network. Then for any subnetwork* $\mathcal{G}'$ *of* $\mathcal{G}$*:*

*(a)* $\mathsf{height}(\mathcal{G})^{-1} \leq \tau^h \leq \infty$;

*(b)* $\mathsf{diam}(\mathcal{G})^{-1} \leq \tau^\exists \leq 1$;

*(c)* $\mathsf{diam}(\mathcal{G})^{-1} \leq \tau^\forall \leq \mathsf{rad}(\mathcal{G})^{-1}$

*Proof.* (a) For $h$, the lower bound is the hierarchical height of the organizations, i.e. the largest path from the root $r$ to any vertex using only edges from $E_{\mathsf{fml}}$. Indeed, for any node $u \in V$, there exists a path from $r$ such that $\mathsf{dist}(r, u) \leq \mathsf{height}(\mathcal{G})$.

The case when $\mathcal{G}'$ contains only one node, the root $r$, is meaningless, as the distance $dist(r, r) = 0$. Then, we say that $\tau^h = \infty$.

(b) If subnetwork $\mathcal{G}'$ contains all but one node from $\mathcal{G}'$, then $\tau^\exists = 1$; however, if $\mathsf{dist}(u, v) = \mathsf{diam}(\mathcal{G})$, and $u \in V'$ and $v \notin V'$, then $\mathcal{G}'$ has $\tau^\exists = \mathsf{diam}(\mathcal{G})^{-1}$

Similarly for (c), the lower bound is defined as $\mathsf{diam}(\mathcal{G})^{-1}$. The upper bound $\tau^\forall = \mathsf{rad}(\mathcal{G})^{-1}$ when $\mathcal{G}'$ contains a single node $u$, which is a center of $\mathcal{G}$, and thus $\mathsf{ecc}(u) = \mathsf{rad}(\mathcal{G})$.

□

### 5.2.2 Using Togetherness to Evaluate Integration

A measure of togetherness could be applied to evaluate the effect of integration.

**Dominant integration.** A dominant integration comprises a new organizational network such that the dominated and the dominating networks are part of its structure. Thus, one may apply existential, universal, and hierarchical levels of togetherness of the dominated network $\mathcal{G}_2$ with respect to the whole network $\mathcal{G}_1 \oplus_E \mathcal{G}_2$.

**Collaborative integration.** To evaluate the proximity of two organizational networks in this case, we regard all edges as undirected, that is we consider their communication networks. Then, the problem is equivalent to the problem of integrating two communication homogeneous networks as in Chapter 4.

**Definition 5.2.4.** *Let $\mathcal{G} = \mathcal{G}_1 \oplus_E \mathcal{G}_2$ be a collaborative integration of two organizational networks $\mathcal{G}_1 = (V_1, r_1, E_1)$ and $\mathcal{G}_2 = (V_2, r_2, E_2)$. We define togetherness of $\mathcal{G}_1$ and $\mathcal{G}_2$ in $\mathcal{G}$ as follows:*

1. *The* $\exists$-togetherness *(or* existential togetherness*) is the smallest positive number $0 < \tau^{\exists} \leq 1$: for any $u_1 \in V_1$ there exists $v_1 \in V_2$ such that $(\mathsf{dist}(v_1, u_1))^{-1} \leq \tau^{\exists}$, and for any $v_2 \in V_2$ there exists $u_2 \in V_1$ such that $(\mathsf{dist}(v_2, u_2))^{-1} \leq \tau^{\exists}$ in the communication network $\mathcal{G}_{\mathsf{com}}$.*

2. *The* $\forall$-togetherness *(or* universal togetherness*) is the smallest positive number $0 < \tau^{\forall} \leq 1$: for any $v \in V_1$, for any $u \in V_2$, the $(\mathsf{dist}(v, u))^{-1} \leq \tau^{\forall}$ in the communication network $\mathcal{G}_{\mathsf{com}}$.*

Collaborative integration could also be considered through the prism of bringing two certain departments together (not the entire organizations). We will consider this problem separately.

## 5.3 Dominant Integration of Two Organizational Networks

Let $\mathcal{G}_1, \mathcal{G}_2$ be two organizational networks. We say $\mathcal{G}_1 \oplus_E \mathcal{G}_2$ is the *dominant integrated* network if there exists exactly one directed edge $\overrightarrow{(u, v)} \in E$ such that $u \in V_1$, $v \in V_2$. To evaluate the effect of integration, we define optimal set of edges:

**Definition 5.3.1.** *Let $G = \mathcal{G}_1 \oplus_E \mathcal{G}_2$ be an integrated organizational network. We say that the set of edges $E$ is* optimal for $\natural \in \{\exists, \forall, h\}$ *if $\tau^{\natural}$ is greater or equal than $\tau'^{\natural}$ of $\mathcal{G}_1 \oplus_{E'} \mathcal{G}_2$ for any set of edges $E' \neq E$ of the same cardinality.*

In other words, an optimal set is the smallest set of edges to satisfy given parameter $\natural \in \{\exists, \forall, h\}$.

The **dominant network integration problem** could be stated as following: Given two networks $\mathcal{G}_1$, $\mathcal{G}_2$, with capacities $c_1$, $c_2$, respectively, build a dominant integrating network such that $\mathcal{G}_1$ is dominating and $\mathcal{G}_2$ is dominated, and the togetherness of $\mathcal{G}_2$ with respect to $\mathcal{G}_1 \oplus_E \mathcal{G}_2$ satisfies certain requirements. Moreover, the number of edges is as small as possible. By our assumption, each node has exactly one incoming edge. Thus, the only way to integrate networks is to link the root of the dominated network with one of the nodes in the other network.

**INPUT** Two organizational networks $\mathcal{G}_1$, $\mathcal{G}_2$ with capacities $c_1$, $c_2$, respectively, and $t \in (0, 1]$ is the desired value of togetherness.

**OUTPUT** A set $E \subseteq V_1 \otimes V_2$ such that the togetherness $\tau^\natural(\mathcal{G}_2, \mathcal{G}_1 \oplus_E G_2) \geq t$. An *optimal solution* $E$ of this problem is one that has the smallest cardinality.

Clearly, one needs to carefully choose which vertices to connect, as this choice would affect the togetherness. We illustrate this with an example:

**Example 12.** Suppose we have two networks $\mathcal{G}_1$ and $\mathcal{G}_2$ with $V_1 = \{0, 1, 2, 3, 4, 5, 6\}$ and $V_2 = \{a, b, c\}$. Both networks contain only directed edges, and, thus, form trees. In Figure 5.3, we show two possible ways to build a dominant integrating network using single directed edge. In Figure 5.3a, the resulting network has the same value of $h$ as in Figure 5.3b, however, $\tau^\forall(\mathcal{G}_2, \mathcal{G}_1 \oplus_E \mathcal{G}_2) > \tau^\forall(\mathcal{G}_2, \mathcal{G}_1 \oplus_{E'} \mathcal{G}_2)$ and $\tau^\exists(\mathcal{G}_2, \mathcal{G}_1 \oplus_E \mathcal{G}_2) > \tau^\exists(\mathcal{G}_2, \mathcal{G}_1 \oplus_{E'} \mathcal{G}_2)$. To make the values of universal and existential togetherness equal, one needs to create an additional undirected edge, for example $e = (5, a)$, i.e. $E'' = \{(3, a), (5, a)\}$.

In this example, $|E| = |E'|$; however, set $E$ is optimal for any togetherness $\exists$, $\forall$ and $h$ assuming that the capacity $c = 4$ – which means that the root $r = 0$ does not have free capacity: three units are spent on communicating with direct subordinates, and one unit is used by the hidden self-loop (See Chapter 2); while $E'$ is optimal only for $h$.

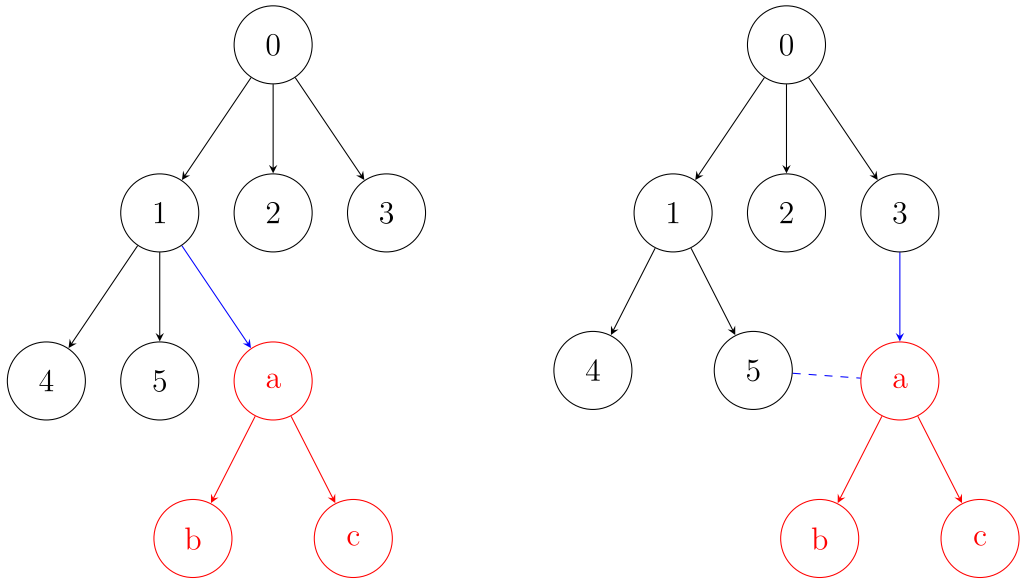
Suppose we have two organizational networks $\mathcal{G}_1$, $\mathcal{G}_2$.

**Theorem 5.3.1.** *Let $\mathcal{G}_1 \oplus_E \mathcal{G}_2$ and $\mathcal{G}_1 \oplus_{E'} \mathcal{G}_2$ be dominant integrated networks. If $\tau^\exists(\mathcal{G}_2, \mathcal{G}_1 \oplus_E \mathcal{G}_2) > \tau^\exists(\mathcal{G}_2, \mathcal{G}_1 \oplus_{E'} \mathcal{G}_2)$, then $\tau^\forall(\mathcal{G}_2, \mathcal{G}_1 \oplus_E \mathcal{G}_2) > \tau^\forall(\mathcal{G}_2, \mathcal{G}_1 \oplus_{E'} \mathcal{G}_2)$.*

*Proof.* Let $u_1 \in V_1$, $v_1 \in V_2$ be two nodes such that $(\mathsf{dist}_{\mathcal{G}}(u_1, v_1))^{-1} = \tau^\exists(\mathcal{G}_2, \mathcal{G}_1 \oplus_E \mathcal{G}_2)$; then for any $u \in V_1$, for any $v \in V_2$: $\mathsf{dist}_{\mathcal{G}}(u_1, v_1) \geq \mathsf{dist}_{\mathcal{G}}(u, v)$.

On the other hand, let $u_2 \in V_1$, $v_2 \in V_2$ be two nodes such that $(\mathsf{dist}_{\mathcal{G}'}(u_2, v_2))^{-1} = \tau^\exists(\mathcal{G}_2, \mathcal{G}_1 \oplus_{E'} \mathcal{G}_2)$; then for any $u \in V_1$, for any $v \in V_2$: $\mathsf{dist}_{\mathcal{G}'}(u_2, v_2) \geq \mathsf{dist}_{\mathcal{G}'}(u, v)$.

By Definition 5.2.1, since $\tau^\exists(\mathcal{G}_2, \mathcal{G}_1 \oplus_E \mathcal{G}_2) > \tau^\exists(\mathcal{G}_2, \mathcal{G}_1 \oplus_{E'} \mathcal{G}_2)$, we imply that $\mathsf{dist}_{\mathcal{G}}(u, v) \leq \mathsf{dist}_{\mathcal{G}}(u_1, v_1) < \mathsf{dist}_{\mathcal{G}'}(u_2, v_2)$. Thus, for any $u \in V_1$, for any $v \in V_2$, $\mathsf{dist}_{\mathcal{G}}(u, v) < \mathsf{dist}_{\mathcal{G}'}(u_2, v_2)$, and hence $\tau^\forall(\mathcal{G}_2, \mathcal{G}_1 \oplus_E \mathcal{G}_2) > \tau^\forall(\mathcal{G}_2, \mathcal{G}_1 \oplus_{E'} \mathcal{G}_2)$ by definition.




(a) Integrated network $\mathcal{G}_1 \oplus_{\{(1,a)\}} \mathcal{G}_2$:
$\tau^h(\mathcal{G}_2, \mathcal{G}_1 \oplus_E \mathcal{G}_2) = \frac{1}{3}$,
$\tau^\forall(\mathcal{G}_2, \mathcal{G}_1 \oplus_E \mathcal{G}_2) = \frac{1}{4}$,
$\tau^\exists(\mathcal{G}_2, \mathcal{G}_1 \oplus_E \mathcal{G}_2) = \frac{1}{3}$

(b) Integrated network $\mathcal{G}_1 \oplus_{\{(3,a)\}} \mathcal{G}_2$:
$\tau^h(\mathcal{G}_2, \mathcal{G}_1 \oplus_{E'} \mathcal{G}_2) = \frac{1}{3}$,
$\tau^\forall(\mathcal{G}_2, \mathcal{G}_1 \oplus_{E'} \mathcal{G}_2) = \frac{1}{5}$,
$\tau^\exists(\mathcal{G}_2, \mathcal{G}_1 \oplus_{E'} \mathcal{G}_2) = \frac{1}{4}$

Figure 5.3: Dominant integration of two networks

$\square$

This theorem shows that universal and existential togetherness are closely related: increasing one of the parameters, the other value would be eventually increased as well.

Since the capacity is limited, there are some undecidable problems: no matter how many edges one creates, it is not possible to reach a certain togetherness. Thus, given any two organizational network, we can compute the limits on each parameter of togetherness.

**Claim 5.3.1** (Condition of existence.)**.** *Let $\mathcal{G}_1$, $\mathcal{G}_2$ be two organizational networks. A dominant integrating network $\mathcal{G}_1 \oplus_E \mathcal{G}_2$ exists only if the relative degree of the root in $\mathcal{G}_2$ is smaller than the relative capacity:* $\deg(r_2) < c$.

Note that in the claim above we have no restrictions over the values of togetherness nor the final number of edges.

### 5.3.1 Dominant Integration with a Single Node

We first mention a simple case when, given a dominating organizational network $\mathcal{G}_1 = (V, r_1, E_{\mathsf{fml}}, E_{\mathsf{inf}})$ we integrate it with a dominated network that contains a single node $v$: $\mathcal{G}_2 = (\{v\}, v, \emptyset, \emptyset)$. The resulting dominant integration network would be $\mathcal{G} = (V \cup \{v\}, r_1, E_{\mathsf{fml}} \cup \{(u, v)\}, E_{\mathsf{inf}} \cup E)$. $E = \emptyset$ implies that we only create a formal edge. Clearly in this case, the closer $v$ is placed to the root $r_1$, the larger the hierarchical togetherness $\tau^h$ of $\mathcal{G}_2$.

If we can establish several undirected edges as well, some other strategies could be applied in order to gain a better position. In Section 4.4, we consider the case when a single node is added to a social network. The same ideas can be applied for organizational networks: in order to make $v$ central (that is to gain higher values of existential and universal togetherness), we need to place it closer to the center of $\mathcal{G}_1$.

### 5.3.2 Dominant Integration of Two Hierarchies

In this subsection, we consider another special case ("pure hierarchies") when organizational networks are represented by tree structures, i.e. there known only formal relationships between employees. This case helps better understand the nature of dominant integration.

Dominant integrating consists of two steps: we first create a directed edge and then add undirected edges.

**Step 1. Creating a single directed edge.** Suppose we have two trees $T_1 = (V_1, E_1)$ and $T_2 = (V_2, E_2)$ with roots $r_1 \in V_1$ and $r_2 \in V_2$, and heights $\mathsf{height}(T_1)$, $\mathsf{height}(T_2)$. Suppose that we integrate the trees such that $T_1$ is dominating.

Technically, when we integrate two networks that contain only directed edges (i.e. directed rooted trees), we obtain a new directed rooted tree where the root is the root of the dominating network. Note that by Claim 5.3, root $r_2$ must have free capacity.

Given two trees $T_1 = (V_1, E_1)$ and $T_2 = (V_2, E_2)$, create an edge $e$ such that $T_1 \oplus_E T_2$ with $E = \{e\}$ is a dominant integrating network. In fact, we do not have choice, which node to pick from $V_2$ as it must be the root $r_2$; we only choose the most suitable node $v \in V_1$, i.e. $e = \overrightarrow{(v, r_2)}$.

To evaluate the effect of integration, we compute a togetherness of the dominated tree $T_2$ with respect to $T_1 \oplus_E T_2$, that is we need to compute $\tau^\natural(T_2, T_1 \oplus_E T_2)$.

**Existential and Universal togetherness.** As we noted, we do not have freedom to choose a vertex in $V_2$. Thus, to maximize both, universal and existential togetherness, one needs to pick node $v \in V_1$ that would be as close to a center of the tree $T_1$ as possible (or be a center itself). If $v$ is a center, than for any $u \in V_1$, $\mathsf{dist}(v, u) \leq \mathsf{rad}(T_1)$; however, center nodes may not be available due to capacity restrictions.

By creating edge $\overrightarrow{(v, r_2)}$, we make the single path available from nodes in $V_1$ to $V_2$ and back, which always goes through $v$. Existential togetherness $\tau^\exists$ is the inverse of the largest distance from any $u \in V_1$ to $v$ plus one, that is $\tau^\exists = (\mathsf{ecc}(v) + 1)^{-1}$. Similarly, $\tau^\forall = (\mathsf{ecc}(v) + \mathsf{height}(T_2) + 1)^{-1}$

Thus, we make the following observation:

**Proposition 5.3.1.** *In a dominant integrating network $T_1 \oplus_E T_2$ with $E = \{\overrightarrow{(v, r_2)}\}$, if $v$ is a center of $T_1$ then the set $E$ is optimal for $\exists$ and $\forall$. Moreover, $\tau^\exists = (\mathsf{rad}(T_1)+1)^{-1}$ and $\tau^\forall = (\mathsf{rad}(T_1) + \mathsf{height}(T_2) + 1)^{-1}$.*

This proposition follows directly from the reasoning above.

**Hierarchical togetherness.** Suppose we pick the vertex $v \in V_1$ to be connected with $r_2 \in V_2$. Then, hierarchical togetherness of the resulting integrated network $T_1 \oplus_E T_2$ where $E = \{\overrightarrow{(v, r_2)}\}$ is $\tau^h = (\mathsf{dist}(r_1, v) + \mathsf{height}(T_2) + 1)^{-1}$.

Moreover, if $\mathsf{height}(T_1) < \mathsf{dist}(r_1, v) + \mathsf{height}(T_2) + 1$, then $\tau^h$ is the inverse of the height of the resulting rooted tree, i.e. $\tau^h(T_2, T_1 \oplus_E T_2) = (\mathsf{height}(T_1 \oplus_E T_2))^{-1}$.

Thus, to maximize the hierarchical togetherness $h$, one needs to place the dominated subnetwork $T_2$ as close to the root $r_1$ as possible. If there is no other requirements and constrains, we then make the following simple conclusion:

**Proposition 5.3.2.** *Let $T_1 \oplus_E T_2$ be a dominant integrating network with $E = \{\overrightarrow{(v, r_2)}\}$, and $S \subseteq V_1$ be the set of nodes with free capacity. The set $E$ is optimal for $h$ iff for any $u \in S$, $\mathsf{dist}(v, r_1) \leq \mathsf{dist}(u, r_1)$. Moreover, $\tau^h = (\mathsf{dist}(r_1, v) + \mathsf{height}(T_2) + 1)^{-1}$.*

Clearly, if root $r_1$ has free capacity, the lemma guarantees that the only optimal set for $h$ is set $E = \{\overrightarrow{(r_1, r_2)}\}$.

We conclude that the key difference is that we want to be as close to the top manager as possible to maximize $h$, and to be as close to the center of the network as possible to maximize $\forall$ ($\exists$).

**Step 2. Creating undirected edges.** Suppose we have created a directed edge $\overrightarrow{(v, r_2)}$, however, the togetherness $\natural = \{h, \forall, \exists\}$ in the resulting dominant integrated network is not as desired.

Note that $T_1 \oplus_E T_2$ at this stage is $T = (V_1 \cup V_2, E_1 \cup E_2 \cup \{\overrightarrow{(v, r_2)}\})$.

**Existential togetherness.** The second step of dominant integration with the aim of reaching a certain existential togetherness, is the same as for any homogeneous networks as in Chapter 4. However, tree structure makes computations of distance $k$ dominating sets significantly easier.

If the desired value of $\exists$-togetherness $\epsilon$ is smaller than current value $\tau^{\exists}$ of $T_1 \oplus_E T_2$ with $E = \{\overrightarrow{(v, r_2)}\}$, one needs to find a distance $k$ dominating set $D_{V_1} \subset V_1$ such that $v \in D_{V_1}$ and $k = \frac{1}{\epsilon} - 1$.

This can be done using the following procedure: Fix $k = \frac{1}{\epsilon} - 1$, and let $D_{V_1} = \{v\}$. Let set $S \subset V_1$ be the set of vertices such that for any $w \in S$ there does not exist $u \in D_{V_1}$ such that $\mathsf{dist}(w, u) \leq k$. Take a vertex $l \in S$ of the largest depth, and find $n \in V_1$ such that $\mathsf{dist}(n, l) = k - 1$ (up the tree). Add $n$ to the set $D_{V_1}$, and check if it is a distance $k$ dominating set for $V_1$. See Procedure 8.

**Procedure 8** Computing distance $k$ dominating set for tree $T = (V, E)$; Output $D_V$

```
Given v ∈ V, D_V := {v}
Set S := {u | dist(u, v) ≤ k for all v ∈ D_V}
while S ≠ V do
    l = leaf(T, S)                ▷ Pick l of the largest depth such that l ∉ S
    n = parent(T, l, k)           ▷ Find n ∈ V such that dist(n, l) = k − 1
    D_V = D_V ∪ {n}
    Update S = {u | dist(u, v) ≤ k for all v ∈ D_V}
end while
```

When the distance $k$ dominating set $D_{V_1}$ is built, connect each node in this set with some nodes in $V_2$; the number of edges in $E$ would coincide with the cardinality of the set $D_{V_1}$. Then, the existential togetherness in the dominated integrated network $T_1 \oplus_E T_2$ is $\tau^{\exists} = (k + 1)^{-1}$. Note that for existential and universal togetherness we consider communications networks, i.e. the trees are treated as undirected.

**Universal togetherness.** Let $\sigma$ be the desired $\forall$-togetherness, $\tau^{\forall}$ is the current togetherness of $T_1 \oplus_E T_2$. We need to make sure that all nodes in $V_1$ are at distance smaller than given $\sigma$ from all nodes in $V_2$. Below we present sketches for developing an approach to solve the problem:

- **Connect periphery vertices**: Suppose $x, y \in V_1 \cup V_2$ are nodes that are furthest apart in the integrated network, they share the "weakest channel". Hence, it makes sense to connect $x, y$ by an edge. Thus, a simple approach is to find two most distant vertices and connect them by an edge, then recompute the distances, and repeat the operation until the desired $\tau^{\forall} \geq \sigma$. This approach does not take into account the tree structure of the networks, moreover, as soon as we create a single undirected edge, the network does not have a tree structure anymore. Thus, as the number of paths between vertices increases, recomputing distances becomes more time consuming. For more details, see Chapter 4.

- Find a distance $l$ dominating set $D_{V_1} \subset V_1$, and a distance $m$ dominating set $D_{V_2} \subset V_2$ such that $v \in D_{V_1}$, $r_1 \in D_{V_2}$, and $m + l = \frac{1}{\sigma} - 1$. Then, create $m \times l$ edges by connecting each vertex in $D_{V_1}$ with each vertex in $D_{V_2}$. The problem here is that with limited capacity, vertices in the dominating sets may not be able to maintain all required connections. Moreover, there is no clear knowledge how to properly choose $m$ and $l$.

- Mix these two approaches: find distance dominating sets; then, connect each vertex in the set with at least one in the other set (we assume that distance dominating sets contain vertices with some free capacity) until no edges between the sets can be established. Finally, recompute the $\forall$- togetherness, and, if needed, create more edges by linking two most distant vertices. This approach is to be used with limited capacity.

**Hierarchical togetherness.** If the desired value of the hierarchical togetherness $h$ is larger than the current value $\tau^h = (\mathsf{dist}(r_1, v) + \mathsf{height}(T_2))^{-1}$, then we need to create a number of undirected edges to decrease $\mathsf{dist}(r_1, u)$ for some $u \in V_2$. This could be done by creating *shortcuts*.

**Proposition 5.3.3.** *Let $T_1 \oplus_E T_2$ be a dominant integrating network with $|E| > 1$. If the set $E$ is optimal for $h$ then there is at least one edge in $E$ that is a shortcut.*

The proof follows from the definition of shortcuts and reasoning above. Indeed, shortcuts provide vertices with alternative paths that are shorter than a path that uses formal edges only.

### 5.3.3 Solving the Dominant Integration Problem with Fixed Hierarchical Togetherness

Let $h$ be the desired value of togetherness,and $\tau^h$ is the current togetherness of the network (we assume that the directed edge is already established). The idea is following: to make $\tau^h \geq h$, connect closest to the root $r_1$ vertices to the vertices in $V_2$ such that the new edges provide the root $r_1$ with short access to as many nodes in $V_2$ as possible.

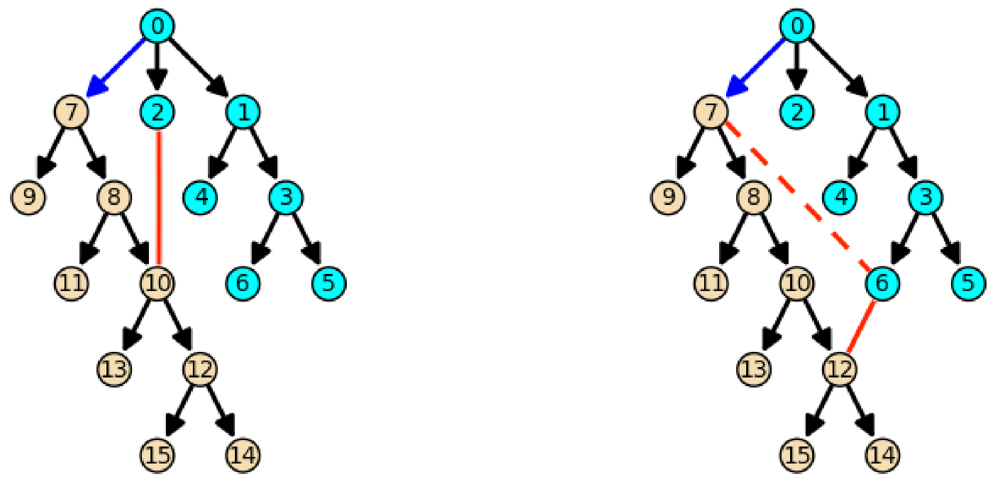
The problem may seem very similar to the problem of finding a distance $k$ dominating set, which is soluble in polynomial time on trees. However, restricting capacity is what makes it hard.

**Example 13.** Consider dominant integration of two trees as in Figure 5.4. Suppose we want to reach hierarchical togetherness $\tau^h = \frac{1}{4}$. Let $\overrightarrow{(0,7)}$ be the directed edge created at the first stage of integration. Without any undirected edge, the integrated network $T_1 \oplus_{\{\overrightarrow{(0,7)}\}} T_2$ has $\tau^h = \frac{1}{5}$ because $\mathsf{dist}(0,14) = \mathsf{dist}(0,15) = 5$. Creating, for example, an edge $(2,10)$ could improve the hierarchical togetherness (See Figure 5.4a), but this may not be possible if we assume that only vertices at the bottom $\{5,6\}$ in $T_1$ have free capacity. Then creating two shortcuts through one of them, for example, edges $(6,7)$ and $(6,12)$ would also increase $\tau^h$ (See Figure 5.4b)

Creating undirected edges between hierarchies quickly deprives us of having single paths in the network. For example in Figure 5.4b, creating an edge that connects 6 and 12 leads to creating a new path from the root 0 to 12: $(0,7,8,10,12)$ and $(0,1,3,6,12)$. $T_1 \oplus_E T_2$ is no longer a tree, thus, more options should be considered when creating a next undirected edge.

#### The LeafToRoot Paradigm

Reaching a desired value of togetherness $h$ implies that all leaf nodes in $V_2$ are at most at distance $\frac{1}{h}$ from $r_1$ in the resulting network. We propose an approach, called the




(a) one undirected edge if at least one of $\{1, 2, 3, 4\}$ has free capacity

(b) two undirected edges if only 5 and 6 have free capacity

Figure 5.4: Integrating two hierarchies $T_1$ and $T_2$ with $\tau^h = \frac{1}{4}$: capacity affects the cardinality of an optimal set of edges

LeafToRoot **paradigm**, that is focused on "covering" nodes in $V_2$ by following up the tree (and therefore, covering leaf-nodes first). This approach has always three main steps:

Step 1 Find a node $n \in V_1$ with free capacity that is as close to the root $r_1$ as possible.

Step 2 Find a node $w \in V_2$ with free capacity that is as far from leaves as possible and/or covers as many leaves as possible.

Step 3 Create undirected edge $(n, w)$. Recompute togetherness $\tau^h$; if needed, repeat the procedure.

The pseudocode is presented in Procedure 9.

**Example 14.** To illustrate how this paradigm works, suppose we have two balanced trees: $T_1$ with branching factor 3 and height 2 (thus, it has 12 nodes in total), and $T_2$ with branching factor 2 and height 4 (30 nodes in total). We want to integrate these networks such that $T_1$ is dominating and $T_2$ is dominated. For simplicity, suppose also that each vertex has exactly one unit of free capacity. Fix the desired value of togetherness $h = \frac{1}{4}$.

First, we create the formal tie that connects both roots $\overrightarrow{(0, 13)}$ (that is to minimize current $\tau^h$ value). In the resulting integrated network at this stage we have

**Procedure 9** Algorithm LeafToRoot for dominant integration with fixed hierarchical togetherness $h$; Output $E$

$E = \{\overrightarrow{(v, r_2)}\}$ ▷ Create a directed edge between $T_1$ and $T_2$
$T = T_1 \oplus_E T_2$
$\tau^h = computeHierarchicalTogetherness(T)$
**while** $\tau^h < h$ **do**
  $n = closestAvailable(V_1, r_1)$ ▷ $n$ should have free capacity
  $distToRoot = \mathsf{dist}(r_1, n)$ ▷ Step 1 ends
  $L = getLeaves(T, V_2)$ ▷ Step 2
  Find $W \subset V_2$: $\mathsf{dist}(w, l) \leq \frac{1}{h} - m$, and $\mathsf{dist}(r_2, w)$ is minimal; $l \in L$, $w \in W$
  Choose $w$ according to chosen heuristics ▷ $w$ should have free capacity
  Create edge $e = (w, n)$
  $E = E \cup \{e\}$; update $T = T_1 \oplus T_2$
  Recompute $\tau^h = computeHierarchicalTogetherness(T)$
**end while**

$\tau^h = (\mathsf{dist}(r_1, r_2) + \mathsf{height}(T_2))^{-1} = \frac{1}{5}$. Since the desired value $h > \tau^h$, we apply the LeafToRoot paradigm.

First, we pick vertex 1 as the closest to root 0 vertex with free capacity (the root itself if not available as we created edge $(0, 13)$), and fix $m = \mathsf{dist}(0, 1) = 1$. Then, we found leaf 28 that is at distance 5 from the root. Node 16 is at distance $\frac{1}{h} - m - 1 = 2$ from 28, thus, we create an edge $(1, 16)$.

Repeating this procedure, we create four more edges as shown in Figure 5.5.

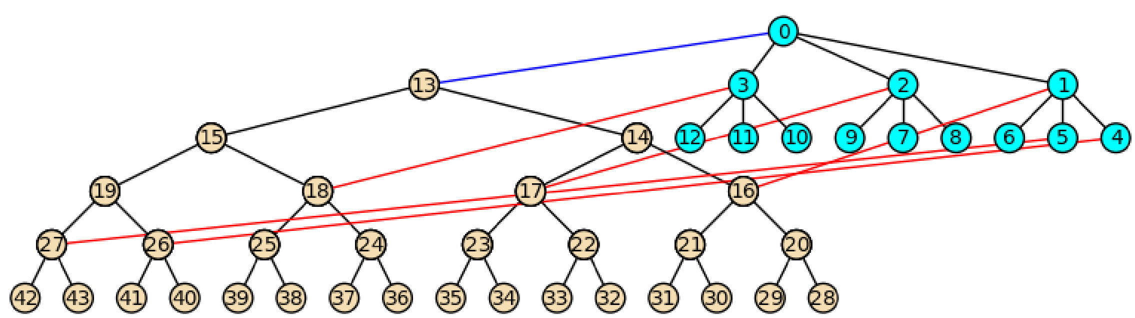
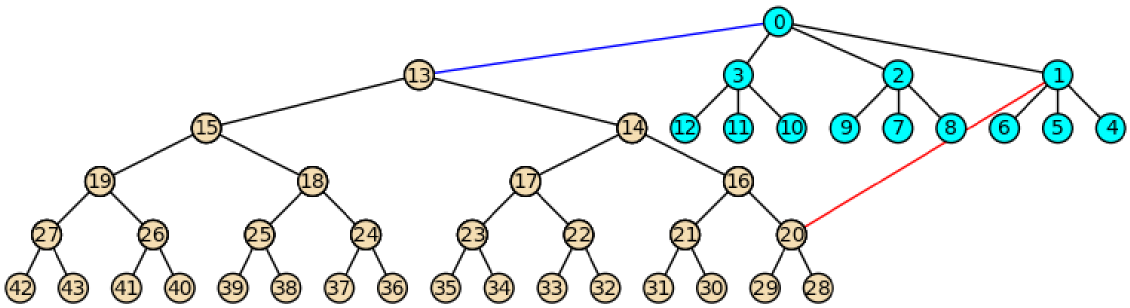


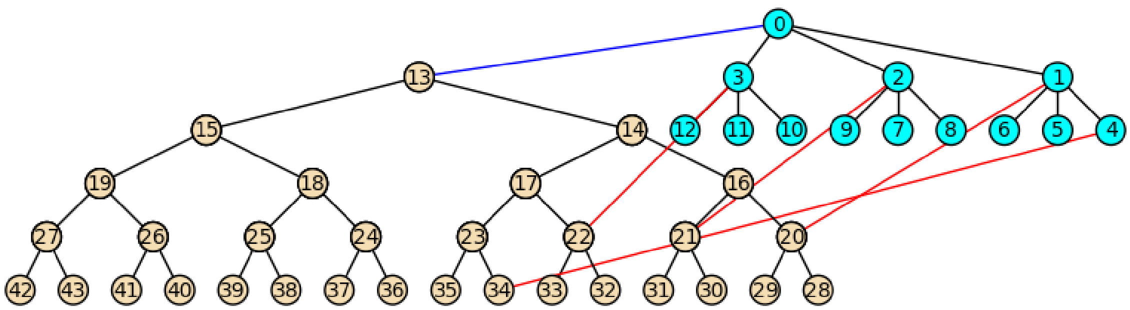

Figure 5.5: Dominant integrating of two balanced trees with $\tau^h = \frac{1}{4}$

However, if we fix $h = \frac{1}{3}$ for this example, the solution would not exist: see Figure 5.6. We again create edges one by one until we run out of vertices with free capacity in $V_1$. However, node 43 is still at distance 5 from the root 0, and thus, $\tau^h = \frac{1}{5}$.

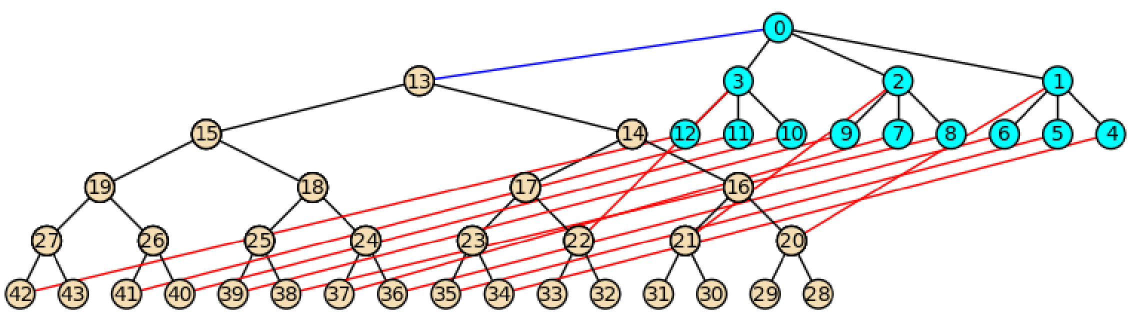
This example allows us to make the following observation. Suppose we have two



(a) Creating one edge



(b) Creating four edges



(c) Creating twelve edges

Figure 5.6: Integrating two balanced trees: no solution exists with $\tau^h = \frac{1}{3}$

balanced perfect trees: one has root degree $a$ and height $\mathsf{height}(T_1)$, and the other one has root degree $b$ and height $\mathsf{height}(T_2)$. To reach a desired value of togetherness, we need to make sure that all leaves in $V_2$ are at distance not larger than $\frac{1}{h}$ from the root. Clearly, there are $b^{\mathsf{height}(T_2)}$ leaves in this tree.

Let $h_2 = \mathsf{height}(T_2)$ and $h_1 = \min\{\mathsf{height}(T_1), \frac{1}{h}\}$. Root $r_1$ can be attached to a vertex at the level $h_1 - 1$, and thus would 'cover' $b^{h_1}$ leaves. (Assuming it has free capacity). Nodes in $V_1$ that are at distance $k$ from $r_1$ ($k < h_1$) can 'cover' $b^{h_1-k}$ leaves each. Under assumption that each vertex has exactly one unit of free capacity, we get the following condition:

$$[b^{h_1} * 0] + b^{h_1-1} * a^1 + b^{h_1-2} * a^2 + \cdots + b^0 * a^{h_1} \geq b^{h_2}$$

$[b^{h_1} * 0]$ means that we may need to exclude the root $r_1$ if its unit of free capacity was used on maintaining the formal tie between $T_1$ and $T_2$.

In our example above, we have $b = 2$, $a = 3$, $h_1 = 2$, $h_2 = 4$, which implies $2^1 * 3^1 + 2^0 * 3^2 = 15 < 2^4 = 16$. Thus, the solution does not exist – as it was earlier also shown experimentally.

This observation suggests that *capacity limitations cause limits on the number of edges that can be established between the networks*, and thus on the smallest hierarchical togetherness that could possible be reached.

Let $T_1 = (V_1, r_1, E_1)$, $T_2 = (V_2, r_2, E_2)$ be two trees.

**Theorem 5.3.2.** *Assume that any $v \in V_1 \cup V_2$ has unlimited capacity and consider $T_1 \oplus_E T_2$. The set $E = \bigcup_{i=1}^{m} \{(r_1, w_i)\}$ is optimal for given $h$ iff $W = \{w_1, w_2, \ldots w_m\}$ is a minimum distance $k$ dominating set with $k = \frac{1}{h} - 1$.*

*Proof.* Suppose $E$ is optimal for $h$. Since each edge to the root $r_1$ goes through $r_1$ itself, the path from each $w_i$ to any node in $V_2$ is not larger than $\frac{1}{h} - 1$, i.e. for any $u \in V_2$, there exists $w \in W$ such that $\mathsf{dist}(w, u) \leq \frac{1}{h} - 1$. Thus, $W$ is a distance $k$ dominating set where $k = \frac{1}{h} - 1$. Suppose there exists another distance dominating set $W'$ of cardinality $r$ such that $|W'| < |W|$. Then the set $E' = \bigcup_{i=1}^{r} \{(r_1, w'_i)\}$ with each $w'_i \in W'$ would also be a solution for $h$; moreover $r < m$ implies that $E$ is not optimal. Hence, $W$ must be a minimum distance $k$ dominating set.

Suppose $W \subset V_2$ is a minimum distance $k$ dominating set where $k = \frac{1}{h} - 1$. By definition of distance dominating sets, for any $v \in V_2$ there is $u \in V_2$ such that $\mathsf{dist}(v, u) \leq k$. Thus, if we connect each $w \in W$ to the root $r_1$, for any $v \in V_2$ $\mathsf{dist}(r_1, v) \leq k + 1$. Hence $\tau^h = \frac{1}{k+1} = h$. $E$ is optimal as there does not exists a dominating set $W' \subset V_2$ such that $|W'| < |W|$. $\square$

The theorem suggests also that to solve the integration problem for $h$ without capacity restrictions, one should find a minimum distance dominating set for the dominated tree $T_2$. For trees, this could be done using Algorithm 8.

**Remark 2.** Theorem 5.3.2 also holds with weaker restrictions of capacity: if $r_1 \in V_1$ has unlimited capacity, and all $v \in V_2$ have at least one unit of free capacity. Indeed, any edge in $E$ connects the root $r_1$ and some node in $V_2$.

Moreover, the theorem can naturally be extended into the following proposition. If $T_1$ have nodes with free capacity only at a specific hierarchical level $lev < \frac{1}{h}$, then the problem of dominant network integration for $h$ is equivalent to the problem of finding distance $k$ dominating set $D_k$ for $V_2$ where $k = \frac{1}{h} - lev - 1$ and $r_2 \in D_k$.

Using the LeafToRoot paradigm, we develop three algorithms. Step 1 stays the same: let a node $n \in V_1$ with free capacity such that for any node $n' \in V_1$ that also has free capacity $\mathsf{dist}(r_1, n) \leq \mathsf{dist}(r_1, n')$. Let $m = \mathsf{dist}(r_1, n) + 1$.

At Step 2, we choose a node from $V_2$ to be connected with $n \in V_1$ such that all its subordinates are no further than at a desired distance from the root $r_1$. Limited capacity is what makes solving the integration problem complicated. Under different assumptions, we need to apply different strategies to choose the proper node. There are several ways to find this node, and the choice of heuristics shapes several (greedy) algorithms:

**ANY LEAF.** Take a leaf $u \in V_2$ that is the furthest vertex from $r_1$, i.e. $\mathsf{dist}(u, r_1) = (\tau^h)^{-1}$. Find a node $w \in V_2$ with free capacity that is as close to the node $r_2$ as possible, and the distance $\mathsf{dist}(w, u) \leq \frac{1}{h} - m$, i.e. such that for any other $w'$: $\mathsf{dist}(w', u) \leq \frac{1}{h} - m$ with free capacity, $\mathsf{dist}(w, r_2) \leq \mathsf{dist}(w', r_2)$ . Create undirected edge $(n, w)$, and add it to the set $E$.

**Proposition 5.3.4.** *Given two perfect balanced trees $T_1$ and $T_2$ with capacities $c_1$, $c_2$, respectively, the algorithm* AnyLeaf *returns an optimal set of edges $E$.*

The proof follows from the structural properties of balanced trees: all nodes at each hierarchical level are identical to each other in terms of units of free capacity, and the number of predecessors/ancestors. Moreover, in a balanced tree, the number of nodes at each level grows exponentially.

Regardless from which leaf to start, the AnyLeaf paradigm would return sets of the same size. Moreover, at each step we cover the largest possible number of vertices (the greedy approach), thus the resulting set is optimal. See Example 14.

**MAX Degree.** Find set of leaves $L \subset V_2$ that are at largest distance from $r_1$, i.e. $\mathsf{dist}(u, r_1) = (\tau^h)^{-1}$. Find a set $W \subset V_2$ such that for any $l \in L$ there exists $w \in W$:

1) $w$ has free capacity;

2) $\mathsf{dist}(w, l) \leq \frac{1}{h} - m$;

3) there does not exists $w' \in V_2$ with free capacity such that $\mathsf{dist}(w', l) \leq \frac{1}{h} - m$ and $\mathsf{dist}(r_2, w') < \mathsf{dist}(r_2, w)$.

Choose node $w \in W$ that has a largest degree.

This greedy algorithm is suitable when we assume that at least one node at each hierarchical level has at least one unit of free capacity, and thus at each cycle of the algorithm, $m_i \leq m_{i-1} + 1$.

**Example 15.** Suppose we are integrating two trees as in Figure 5.7a; the desired hierarchical togetherness is $h = \frac{1}{3}$ and each node has exactly one unit of free capacity. The "any Leaf" strategy may lead to creating edge $(1, 14)$ at the first step, then $(2, 15)$ at the second, and $(3, 16)$ at the third. Nodes $(26, 27, 28, 29)$ are still at distance 4 from the root 0. Thus, in total the algorithm would output 7 undirected edges plus one directed.

However, a better choice of nodes to be connected could result in a smaller set $E$: In Figure 5.7b, there is an alternative way: creating edges $(2, 17)$ and $(1, 15)$ first would lead to the smaller set of edges $E$ at the end of the run – there would only be created 5 undirected and one directed edges.

**Keep best, remove worst.** Consider $T_1 = (V_1, E_1)$. Let $C_0, C_1 \dots C_{h^{-1}-1}$ be the sum of free capacities at each hierarchical level, i.e. $C_0$ represents how many edges can go directly to the root $r_1$, $C_1$ indicates how many edges can go the nodes that are at distance one from the root $r_1$, etc.

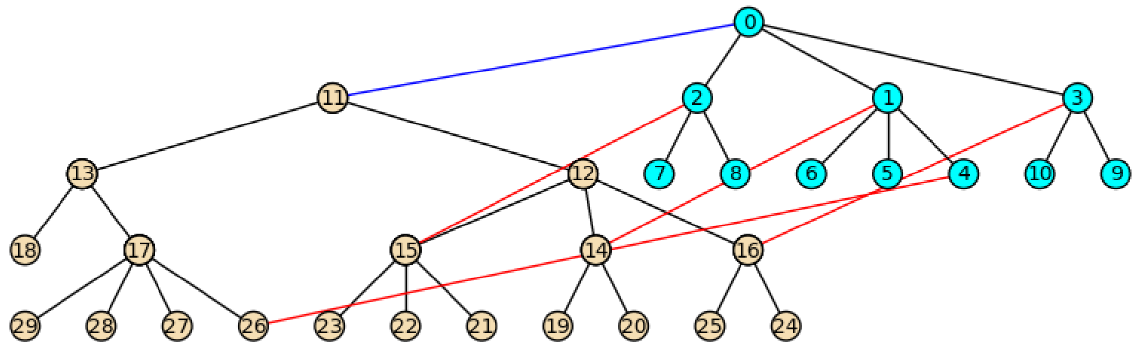
Let $j = 0, \dots, h^{-1} - 1$, $m = h^{-1} - m$. Consider $T_2 = (V_2, E_2)$. Find set of leaves $L \subset V_2$. Find a set $W \subset V_2$ such that for any $l \in L$ there exists $w \in W$:

1) $w$ has free capacity;

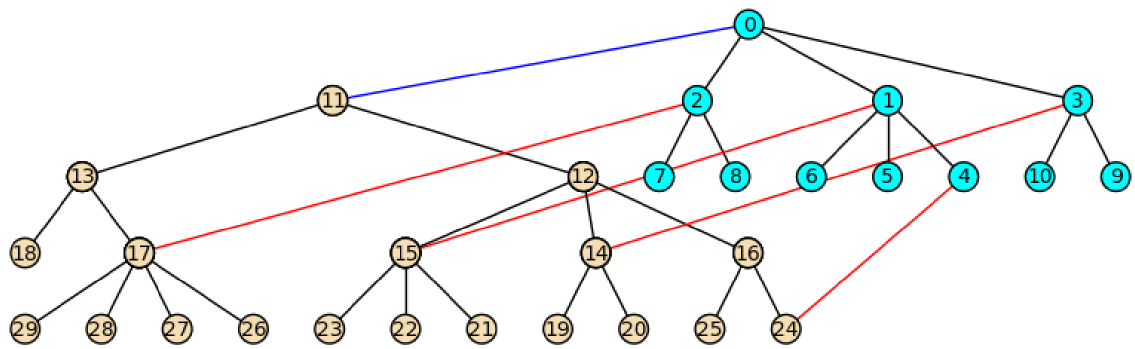
2) $\mathsf{dist}(w, l) \leq j$;

3) there does not exists $w' \in V_2$ with free capacity such that $\mathsf{dist}(w', l) \leq j$ and $\mathsf{dist}(r_2, w') < \mathsf{dist}(r_2, w)$.

Among all nodes in $W$, choose the subset $W_+ \subseteq W$ such that $w_1 \in W_+$ but $w_2 \notin W_+$ implies that $\deg(w_1) \geq \deg(w_2)$ for any $w_1, w_2 \in W$.




(a) $T_1 \oplus_E T_2$ after four steps of the LeafToRoot algorithm (ANY LEAF): three more edges will be created




(b) $T_1 \oplus_E T_2$ after four steps of the Greedy LeafToRoot algorithm (MAX Degree): only one more edge will be created

Figure 5.7: LeafToRoot: Choosing between ANY LEAF and MAX Degree

Then, create edges that link each node in $W_+$ with a node at distance $j$ from the root in $T_2$.

### 5.3.4 Generalization

In this subsection, we summarize our findings about dominant integration and consider a general case when both organizational networks have formal and informal ties. The main difference with the previous section is that now we operate with networks that are not trees, thus, all computations are more complex and time consuming.

**Existential and universal togetherness.** The first step, when we create a formal edge between the networks, is the same as for trees: we want to be as close to a center in $\mathcal{G}_1$ as possible. The second step should guarantee that we bring the networks together by establishing additional ties. Existence of informal ties does not change the main ideas, however it makes the algorithms for finding distance $k$ dominating sets computationally harder.

Thus, when adding undirected edges, with focus on existential togetherness $\exists$, we can use one of the algorithms to find a small distance $k$ dominating set for $\mathcal{G}_1$ (See Chapter 2). Then, connect each vertex in this set to any vertex in $V_2$.

For $\forall$ togetherness, we also combine three methods:

(1) Periphery: Connect two most distant vertices from $V_1$ and $V_2$ (See Chapter 4 for details.)

(2) CtrPer: Connect two most distant vertices in $V_1$ to a center node in $V_2$ (See Chapter 4 for details.)

(3) Find distant dominating sets using MinLeaf for $\mathcal{G}_1$ and $\mathcal{G}_2$, connect vertices from these sets as long as capacity allows it. Then use (1) or (2), to reach desired togetherness.

**Hierarchical togetherness.** Again, the first step is to create a directed edge $e = \overrightarrow{(v, r_2)}$ with $v \in V_1$ and the root node $r_2 \in V_2$. This step is motivated as for the case with trees: we want to place $G_2$ as close to the root node $r_1 \in V_1$ as possible in order to minimize togetherness $\tau^h$.

The resulting value is $\tau^h = (\mathsf{dist}(r_1, v) + \mathsf{dist}(r_2, u) + 1)^{-1}$ for any $u \in V_2$. Note that undirected edges may lead to the case when $\mathsf{dist}(r_2, u) < \mathsf{height}(T_2)$ when there are shortcuts.

We need to create additional undirected edges when there exists $u \in V_2$ such that $\mathsf{dist}(r_1, u)^{-1} < h$. Unlike the case with the trees, $u$ may not be at the bottom hierarchical level.

To integrate two organizational networks, we propose an algorithm that has a very similar idea to the algorithm S-MinLeaf [80], and the paradigm LeafToRoot presented above:

Step 1 Let set $S \subset V_2$ be the subset of vertices such that for any $s \in S$, $\mathsf{dist}(s, r_1) > \frac{1}{h}$. If $S = \emptyset$, $\mathcal{G}$ has already hierarchical togetherness greater or equal $h$.

Step 2 Take a node $n \in V_1$ with free capacity such that for any node $n' \in V_1$ that also has free capacity $\mathsf{dist}(r_1, n) \leq \mathsf{dist}(r_1, n')$. Let $m = \mathsf{dist}(r_1, n)$.

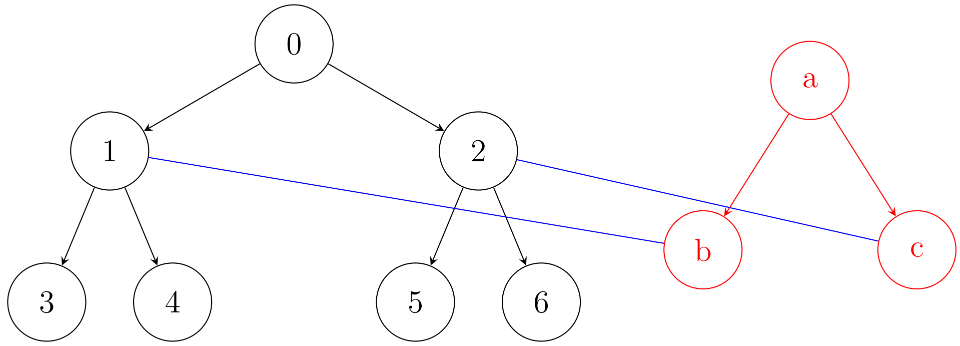



Figure 5.8: Collaborative integration of two organizational networks: $\tau^{\forall} = \frac{1}{3}$, $\tau^{\exists} = \frac{1}{2}$

Step 3 Among all nodes in $S$, find a node $s$ with a smallest degree. Then find a node $w \in V_2$ such that $\mathsf{dist}(w, s_1) = m - 1$ and it has free capacity (at each iteration we choose the node with the largest current degree).

Step 4 Create edge $(n, w)$, recompute set $S$, and repeat the procedure if needed.

## 5.4 Collaborative Integration of Two Organizational Networks

In this section, we consider collaborative integration of organizational networks. We are no longer interested in hierarchical togetherness as directions of ties are ignored for collaboration. The **collaborative network integration problem** could be stated as following: Given two networks $\mathcal{G}_1$, $\mathcal{G}_2$, with capacities $c_1$, $c_2$, respectively, build a collaborative integrated network such that togetherness of $\mathcal{G}_1$ and $\mathcal{G}_2$ in $\mathcal{G}_1 \oplus_E \mathcal{G}_2$ satisfies certain requirements.

**INPUT** Two organizational networks $\mathcal{G}_1$, $\mathcal{G}_2$ with capacities $c_1$, $c_2$, respectively, and $t \in (0, 1]$ is the desired value of togetherness.

**OUTPUT** A set $E \subseteq V_1 \otimes V_2$ such that the togetherness $\tau^{\natural}(\mathcal{G}_1, \mathcal{G}_2) \geq t$. An *optimal solution* $E$ of this problem is one that has the smallest cardinality.

In Figure 5.8, we present a simple collaborative integrated network with $\tau^h = \frac{1}{3}$, $\tau^{\forall} = \frac{1}{3}$, $\tau^{\exists} = \frac{1}{2}$.

The problem of finding small sets of edges in organizational networks with fixed parameters of existential and universal togetherness, $\tau^{\exists}$ and $\tau^{\forall}$, is equivalent to the same problem on networks with undirected edges; this problem has been considered in great detail in Chapter 4. Indeed, direction does not matter in this case (we assume that there is a backflow from a subordinate to her manager).

We, however, mention two special cases for organizational networks:

(1) there is only a formal hierarchy: organizational networks in question do not have informal undirected ties; they are represented by tree structures.

(2) collaborative integration of subnetworks: we are only interested in bringing together members of certain departments

Both these cases are considered below.

### 5.4.1 Collaborative Integration of Two Hierarchies

Suppose we have two organizational networks $T_1 = (V_1, E_1)$, $T_2 = (V_2, E_2)$ such that they both have only directed edges (formal hierarchies).

The problem of finding a minimal set of edges with a given existential togetherness $\epsilon$ is equivalent to the problem of finding distance $k$- dominating sets on trees. In [107], the authors propose a distributed algorithm to determine a minimum (connected) distance-k dominating set of a tree T. It terminates in $O(\mathsf{height}(T))$ rounds and uses $O(log\ k)$ space.

**Existential togetherness.** Let $\epsilon$ be the desired value of $\exists$-togetherness. One needs to find two distance $k$ dominating sets: $D_{V_1}$ and $D_{V_2}$ for $T_1$ and $T_2$, respectively, where $k = \frac{1}{\epsilon} - 1$. Then create $\max\{|D_{V_1}|, |D_{V_2}|\}$ edges by connecting each vertex in $D_{V_1}$ with at least one vertex $D_{V_2}$ and vice verse.

**Universal togetherness.** Let $\sigma$ be the desired value of $\forall$-togetherness.

This case is, in fact, very similar to dominant integration, except we do not need to create a directed edge, and thus simply skip Step 1. To solve collaborative integration problem for $\forall$-togetherness, one may use approaches for homogeneous networks in Chapter 4.

### 5.4.2 Collaborative Integration of Subnetworks

When dealing with two large organizational networks, defining their togetherness on the entire collaborative integrated network may be too restrictive. Indeed, collaboration between two organization often means establishing new connections only between certain departments.

Let $\mathcal{G}_{sub_1}$ and $\mathcal{G}_{sub_2}$ be subnetworks of $\mathcal{G}_1$ and $\mathcal{G}_2$ with $S_1 \subseteq V_1$ and $S_2 \subseteq V_2$, respectively. We say that:

- $\exists$-*togetherness* of $\mathcal{G}_{sub_1}$ and $\mathcal{G}_{sub_2}$ in $\mathcal{G}_1 \oplus_E \mathcal{G}_2$ is the smallest positive number $0 < \tau^{\exists} \leq 1$: for any $u_1 \in S_1$ there exists $v_1 \in S_2$ such that $(\mathsf{dist}(v_1, u_1))^{-1} \leq \tau^{\exists}$, and for any $v_2 \in S_2$ there exists $u_2 \in S_1$ such that $(\mathsf{dist}(v_2, u_2))^{-1} \leq \tau^{\exists}$ in the communication network $\mathcal{G}_{\mathsf{com}}$.

- $\forall$-*togetherness* of $\mathcal{G}_{sub_1}$ and $\mathcal{G}_{sub_2}$ in $\mathcal{G}_1 \oplus_E \mathcal{G}_2$is the smallest positive number $0 < \tau^{\forall} \leq 1$: for any $v \in S_1$, for any $u \in S_2$, the $(\mathsf{dist}(v, u))^{-1} \leq \tau^{\forall}$ in the communication network $\mathcal{G}_{\mathsf{com}}$.

**Example 16.** Consider a simple example. Suppose we have two organizational networks $\mathcal{G}_1 = (V_1, r_1, E_1)$ and $\mathcal{G}_2 = (V_2, r_2, E_2)$ with $V_1 = \{0, 1, 2, 3, 4, 5, 6, 7, 8\}$ and $V_2 = \{a, b, c, d, e\}$. Let $\mathcal{G} = \mathcal{G}_1 \oplus_E \mathcal{G}_2$ be a collaborative integrated network where $E = \{5, b\}$. See Figure 5.9.

As the result of integration, $\mathcal{G}_1$ and $\mathcal{G}_2$ have $\tau^{\forall}(\mathcal{G}_1 \oplus_E \mathcal{G}_2) = \frac{1}{6}$ because the length of a largest shortest path is $\mathsf{dist}(3, c) = 6$; $\tau^{\exists}(\mathcal{G}_1 \oplus_E \mathcal{G}_2) = \frac{1}{4}$ because the distance $\mathsf{dist}(3, b) = 4$. However, suppose that $\mathsf{dist}(3, c)$ is not important; we want to make sure instead that two certain departments are close to each other. Let $S_1 \subset V_1$, and $S_2 \subset V_2$ such that $S_1 = \{5, 6, 7, 8\}$ and $S_2 = \{b, d, e\}$. In this case, $\tau^{\forall}_{sub} = \frac{1}{3}$, $\tau^{\exists}_{sub} = \frac{1}{2}$.

The **collaborative integration problem of subnetworks** could be stated as following:

**INPUT** Two organizational networks $\mathcal{G}_1$, $\mathcal{G}_2$ with capacities $c_1$, $c_2$, and subsets $S_1 \subseteq V_1$, $S_2 \subseteq V_2$ that form the subnetworks $\mathcal{G}_{sub_1}$, $\mathcal{G}_{sub_2}$, respectively; $t \in (0, 1]$ is the desired value of togetherness.

**OUTPUT** A set $E \subseteq V_1 \otimes V_2$ such that the togetherness $\tau^{\natural}(G_{sub_1}, G_{sub_2}) \geq t$ in $\mathcal{G}_1 \oplus_E \mathcal{G}_2$.

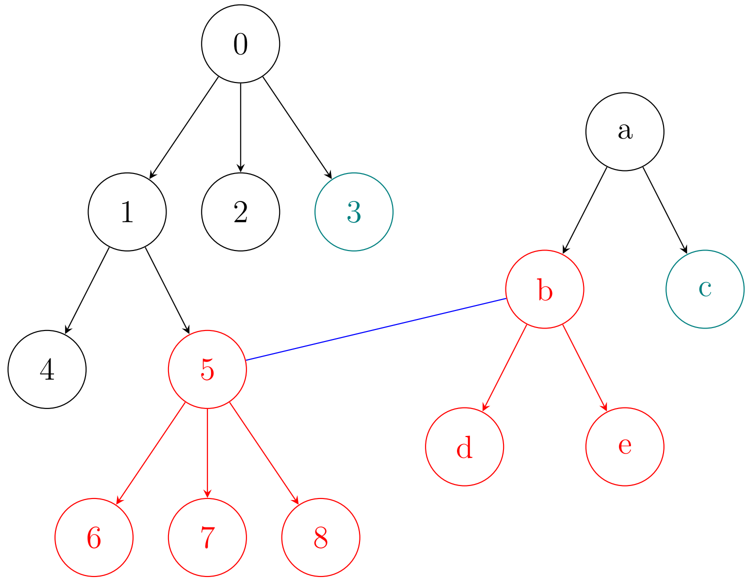



Figure 5.9: Collaborative integration of two organizational networks $\mathcal{G}_1 \oplus_{\{5,b\}} G_2$. Togetherness of the networks: $\tau^{\forall}(\mathcal{G}_1 \oplus_E \mathcal{G}_2) = \frac{1}{6}$, $\tau^{\exists}(\mathcal{G}_1 \oplus_E \mathcal{G}_2) = \frac{1}{4}$; togetherness of the subnetworks: $\tau^{\forall}_{sub}(\mathcal{G}_1 \oplus_E \mathcal{G}_2) = \frac{1}{3}$, $\tau^{\exists}_{sub}(\mathcal{G}_1 \oplus_E \mathcal{G}_2) = \frac{1}{2}$

From the algorithmic point of view, there are two possible ways to approach this problem:

1. Suppose subgraphs $\mathcal{G}_{sub_1} \subseteq \mathcal{G}_1$, $\mathcal{G}_{sub_2} \subseteq \mathcal{G}_2$ represent the desired subnetworks. Then, the set of edges $E$ that form integrated network $\mathcal{G}_{sub_1} \oplus_E \mathcal{G}_{sub_2}$ would also form collaborative integrated networks $\mathcal{G}_1 \oplus_E \mathcal{G}_2$ with desired parameters of togetherness.

2. Edges in the set $E$ that form the integrated network $\mathcal{G}_1 \oplus_E \mathcal{G}_2$ may connect some vertices that are not in the subnetworks $\mathcal{G}_{sub_1}, \mathcal{G}_{sub_2}$. This may be especially useful when integrating networks under capacity restrictions.

The first approach would result in the same algorithms as to solve the collaborative integration problem; while the second approach build a new problem setup. We leave this question for future works (See Chapter 6).

# Chapter 6

# Conclusion and Future Works

Understanding how different departments and employees of an organization interact with one another leads to comprehension of how well the organization operates. Studying an organizational structure often reveals critical positions – both most influential and impotent – that may require additional attention. Moreover, it is the organizational structure from which one may extract hidden clues about concealed communication obstacles.

In this thesis, we studied organizational structures from the network perspective. To define an organizational network we considered both formal and informal relations between employees. We noticed that:

1. there is a lack of mathematical analysis on the dual-structure of formal and informal organizations; and

2. existing formal definitions of power only deal with networks whose edges have a single interpretation of social links, while not incorporating formal roles and levels.

Our aim, therefore, was to develop a mathematical model that sits at the confluence of the two directions above. Our organizational network model is simple, elegant and novel in the sense that it unifies existing studies. Our experimental results are consistent with the knowledge that informal relations significantly affect individual power in an organization. Moreover, we demonstrate that our mathematical model has the potential to provide an explanation to complex phenomena such as flattening and leadership style.

The significance of this thesis lies mainly in theoretical models, simulations and analysis. Nevertheless, we would like to mention that the model is ready to be applied to field works where the hypotheses are verified in an empirical setting. Carefully designed experimentations are required on a real organization to collecting data about formal and informal relations and analyze the internal structures. This would be a natural and crucial next step of our research.

There are several obvious ways in which the model can be enriched:

(1) As argued in Chapter 2, a company may be led by a board of directors rather than a single person. Hence, one may allow several nodes in the network making the reporting hierarchy a forest rather than a single tree.

(2) Our model can be extended by allowing more flexible assignment of weights. It is an interesting direction of future works to explore how enabling different types of informal ties affects the distribution of power in an organization. Indeed, different types of informal ties (such as friendship,advice, or collaborations) may result in different impacts on power. Hence, one may allow several informal ties (undirected or directed) with different correlation coefficients and $k$.

(3) The capacity of individuals are different, and therefore, one may assign different capacities to different individuals.

(4) The current model only applies to functional or divisional structural type of organization, while other types, such as team-based or matrix-organizations are not captured. A model that incorporate the above considerations will provide more realistic analysis. Hence the current work is a first step towards building a generally applicable organizational network analytical tool set.

Another interesting future direction is to add an extra layer of complexity to the model by incorporating different roles which specify tasks an individual perform in the organization. We note that different roles may rely on each other (with common interests) or be in conflict with each other (with conflicting interests). The type of effects that arises due to such interactions is different and hence requires a more complex definition of power.

Network integration amounts to the fundamental question that arises in numerous social, political and physical domains. In this thesis, we applied a formal framework and employed various heuristics to tackle the problem. The key conceptual contribution is in proposing four measures of togetherness (three measures for informal networks in Chapter 4, and one for capturing authority in Chapter 5), which are useful indication of proximity between sub-networks. We believe that togetherness will be helpful not only in this context, but in any problem domains where solidarity and distances are of concern.

Contrary to intuition, our experiments demonstrate that the random strategy for building links between two social networks performs comparable to other heuristics in a few situations; It would be an interesting future work to explore the mathematical reason behind this phenomenon, e.g., what is the expected togetherness if we connect two random graphs using $k$ random edges.

It would also be interesting to incorporate node characteristics in surrounding contexts and apply other principles, e.g., homophily, to guide the establishment of links. Another future work is to incorporate directed or weighted edges in the networks. A potential application is to develop technology that advise potential links or collaborations (say, in an online social platform) to members of two social groups.

Finally, we acknowledge that Chapter 5 could be significantly extended by performing a number of experiments on the collaboration and the dominant integration. This would be interesting to pin down what happens with individual power across organizations when they establish new links between their agents. Power could also be used as a base for priority based heuristics when edges are established between individuals with higher priorities.

Overall, we believe that our approaches to power and the network integration problems represent promising areas for future research. There are still many open questions to answer and interesting problems to solve, and we hope that the ideas posed in this thesis would be inspiring and exciting for future investigators.

# Bibliography


[1] Aguilera, R., Dencker, J., and Yalabik, Z. (2006). Institutions and organizational socialization: integrating employees in cross-border mergers and acquisitions. *International Journal of Human Resource Management*, 15(8):1357–1372.

[2] Ahuja, G. (2000). Collaboration networks, structural holes, and innovation: a longitudinal study. *Administrative Science Quarterly*, 45:425–455.

[3] Albert, R. and Barbási, A.-L. (2002). Statistical mechanics of complex networks. *Reviews of Modern Physics*, 74(47).

[4] Anagnostopoulos, A., Kumar, R., and M, M. (2008). Influence and correlation in social networks. In *Proceedings of the 14th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pages 7–15. ACM.

[5] Andrew, S. A. (2010). Adaptive versus restrictive contracts: Can they resulve different risk problems? In Feiock, R. and Scholz, J., editors, *Self-Organizing Federalism: Collaborative Mechanisms to Mitigate Institutional Collective Action Dilemmas*. Cambridge University Press.

[6] Bakardjieva, M. (2003). Virtual togetherness: an everyday-life perspective. *Media Culture Society*, 25(3):291–313.

[7] Baker, M. (2014). *Peer-to-Peer Leadership: Why the Network Is the Leader*. Berrett-Koehler Publishers.

[8] Baldwin, D. (1978). Power and social exchange. *The American Political Science Review*, 72(4):1229–1242.

[9] Bar-Ilan, J., Kortzars, G., and Peleg, D. (1993). How to allocate network centers. *Journal of Algorithms*, 15(3):385–415.

[10] Barbási, A.-L. and Albert, R. (1999). Emergence of scaling in random networks. *Science*, 286(5439):509–512.

[11] Barnard, C. (1938). *The Functions of the Executive*. Harvard University Press.

[12] Barney, J. and Griffin, R. (1992). *The management of organizations: Strategy, structure, behavior*. Houghton Mifflin, Boston.

[13] Barthèlemy, M. (2004). Betweenness centrality in large complex networks. *Eur. Phys. J. B*, 38:163–168.

[14] Bonacich, P. (1987). Power and centrality: A family of measures. *American Journal of Sociology*, 92:1170–1182.

[15] Bonacich, P. and Lloyd, P. (2001). Eigenvector-like measures of centrality for asymmetric relations. *Social networks*, 23(3):191–201.

[16] Borgatti, S. (2006). Identifying sets of key players in a social network. *Computational & Mathematical Organization Theory*, 12(1):21–34.

[17] Borgatti, S., Everett, M., and Freeman, L. (2002). *Ucinet for Windows: Software for Social Network Analysis*. Harvard, MA: Analytic Technologies.

[18] Borgatti, S. P. (2005). Centrality and network flow. *Social networks*, 27(1):55–71.

[19] Bothner, M., Podolny, J., and Smith, E. (2011). Organizing contests for status: The matthew effect v.s. the mark effect. *Management Science*, 57:439–457.

[20] Bothner, M., Smith, E., and White, H. (2010). A model of robust positions in social networks. *American Journal of Sociology*, 116:943–992.

[21] Bothner, M., Stuart, T., and White, H. (2004). Status differentiation and the cohesion of social networks. *Journal of Mathematical Sociology*, 28:261–295.

[22] Brass, D. (1984). Being in the right place: A structural analysis of individual influence in an organization. *Administrative Science Quarterly*, (4):518–539.

[23] Brass, D. and Krackhardts, D. (2011). *Power, Politics, and Social Networks*. Routledge.

[24] Bryan, L., Matson, E., and Weiss, L. (2007). Harnessing the power of informal employee networks. *The McKinsey Quarterly*, 3:13–19.

[25] Buettner, R. (2016). Getting a job via career-oriented social networking sites: The weakness of ties. In *Proc. of 49th Annual Hawaii International Conference on System Sciences*. IEEE.

[26] Burt, R. (2004). Structural holes and good ideas. *American Journal of Sociology*, 110.

[27] Casella, E. and Fowler, C. (2005). *The Archaeology of Plural and Changing Identities: Beyond Identification*. Springer.

[28] Castilla, E. (2011). Bringing managers back in: Managerial influences on workplace inequality. *American Sociology Review*, 76(5):667–694.

[29] Chang, G. J. and Nemhauser, G. L. (1984). The $k$-domination and $k$-stability problems on sun-free chordal graphs. *SIAM J. Alg. Disc. Meth.*, 5:332–345.

[30] Chaudhuri, S., Dayal, U., and Narasayya, V. (2011). An overview of business intelligence technology. *Communications of the ACM*, 54(8):88–98.

[31] Cox, E. and Bonner, J. (2001). The advantages of togetherness. *Science*, 292(5516):448–449.

[32] Cross, R., Borgatti, S., and Parker, A. (2002). Making invisible work visible: Using social network analysis to support strategic collaboration. *California Management Review*, 44(2):25–46.

[33] Cross, R. and Parker, A. (2004). *The Hidden Power of Social Networks*. Harvard University Press.

[34] Cross, R. and Thomas, R. (2011). Managing yourself: a smarter way to network. *Harvard Business Review*, 89(7–8):149–153.

[35] Cross, R., Thomas, R., Dutra, A., and Newberry, C. (2007). Using network analysis to build a new business. *Organizational Dynamics*, 36:345–362.

[36] Danziger, M. M., Bashan, A., Berezin, Y., Shekhtman, L. M., and Havlin, S. (2014). *An Introduction to Interdependent Networks*, pages 189–202. Springer International Publishing.

[37] Donetti, L., Hurtado, P. I., and Munoz, M. A. (2005). Entangled networks, synchronization and optimal network topology. *Phys. Rev. Lett.*, 95(188701).

[38] Duckworth, W. and Mans, B. (2005). Randomized greedy algorithms for finding small $k$-dominating sets of regular graphs. *Random Structures and Algorithms*, 27(3):401–412.

[39] Easley, D. and Kleinberg, J. (2010). *Networks, Crowds, and Markets: Reasoning About a Highly Connected World*. Cambridge University Press.

[40] Ehrlich, K. and Carboni, I. (2005). Inside social network analysis. Technical report, IBM Corporation.

[41] Emerson, R. (1962). Power-dependance relations. *American Sociological Review*, 27(1):31–41.

[42] Fire, M., Puzis, R., and Elovici, Y. (2015). Organization mining using online social networks. *Network and Spatial Economics*, pages 1–34.

[43] Fortunato, S. (2010). Community detection in graphs. `CoRR abs/0906.0612`.

[44] Franceschet, M. and Bozzo, E. (2015). A theory on power in networks. *CoRR abs/1510.08332*.

[45] French, J. and Raven, B. (1959). *The bases of social power*. University of Michigan Press, Ann Arbor, MI.

[46] Garey, M. R. and Johnson, D. S. (1979). *Computers and Intractability: A Guide to the Theory of NP-Completeness*. W.H.Freeman.

[47] Geanakoplos, J. and Milgrom, P. (1991). A theory of hierarchies based on limited managerial attention. *Journal of the Japanese and International Economies*, 5(3):205–225. Elsevier.

[48] Goleman, D. (2000). Leadership that gets results. *Harvard Business Review*, 78(2):80–90.

[49] Granovetter, M. S. (1973). The strength of weak ties. *The American Journal of Sociology*, 78(6):1360–1380.

[50] Grimes, A. (1978). Authority, power, influence and social control: A theoretical synthesis. *The Academy of Management Review*, 3(4):724–735.

[51] Hanna, R. (2005). Kant and nonconceptual content. *European Journal of Philophy*, 13:247–290.

[52] Jablin, F. M. and Krone, K. J. (1987). Organizational assimilation. In Berger, C. and Chaffee, S., editors, *Handbook of communication science*. Sage.

[53] Jackson, M. O. (2004). A survey of models of network formation: Stability and efficiency. In Demange, G. and Wooders, M., editors, *Group Formation in Economics; Networks, Clubs and Coalitions*. Cambridge University Press.

[54] Jackson, M. O. (2006). The economics of social networks. In Blundell, R., Newey, W., and Persson, T., editors, *Proceedings of the 9th World Congress of the Econometric Society*. Cambridge University Press.

[55] Katona, Z., Zubcsek, P., and Sarvary, M. (2011). Network effects and personal influences: The diffusion of an online social network. *Journal of marketing research*, 48(3):425–443.

[56] Kempe, D., Kleinberg, J., and Tardos, E. (2015). Maximizing the spread of influence through a social network. *Theory of Computing*, 11:105–147.

[57] Keren, M. and Levhari, D. (1979). The optimal span of control in a pure hierarchy. *Management Science*, 25:1162–1172.

[58] Kilduff, M. and Krackhardt, D. (2008). *Interpersonal Networks in Organizations: Cognition, Personality, Dynamics, and Culture*. Cambridge University Press.

[59] Kipnis, D., Schmidt, S., and Wilkinson, I. (1980). Intraorganizational influence tactics: Explorations in getting one's way. *Journal of Applied Psychology*, 65(4):440–452.

[60] Kleinberg, J., Suri, S., Tardos, E., and Wexler, T. (2008). Strategic network formation with structural holes. *ACM SIGecom Exchanges*, 7(3).

[61] Krackhardt, D. and Hanson, J. (1993). Informal networks: The company behind the chart. *Harward Business Review*, 71(4):104–113.

[62] Krackhardts, D. (1990). Assessing the political landscape: Structure, cognition, and power organzation. *Administrative Science Quarterly*, 35:342–369.

[63] Krebs, V. and Holley, J. (2006). Building smart communities through network weaving. http://www.orgnet.com/BuildingNetworks.pdf.

[64] Kubheka, I., Kholopane, P., and Mbohwa, C. (2013). The effects of flattening hierarchies on employee performance in organizations: A study of a south african retail group. In *International Conference on Law, Entrepreneurship and Industrial Engineering (ICLEIE'2013)*, pages 217–222.

[65] Leskovec, J., Kleinberg, J., and Faloutsos, C. (2007). Graph evolution: Densification and shrinking diameters. *ACM Transactions on Knowledge Discovery from Data (ACM TKDD)*, 1(1).

[66] Leskovec, J., Lang, K. J., Dasgupta, A., and Mahoney, M. (2009). Community structure in large networks: Natural cluster sizes and the absence of large well-defined clusters. *Internet Mathematics*, 6(1):29–123.

[67] Liben-Novell, D. and Kleinberg, J. (2003). The link prediction problem for social networks. In *CIKM '03 Proceedings of the twelfth international conference on Information and knowledge management*, pages 556–559.

[68] Liu, J. and Moskvina, A. (2015). Hierarchies, ties and power in organizational networks: Model and analysis. In *ASONAM '15 Proceedings of the 2015 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining*, pages 202–209.

[69] Liu, J. and Moskvina, A. (2016). Hierarchies, ties and power in organizational networks: Model and analysis. *Social Network Analysis and Mining (SNAM)*, 6(106):1–26.

[70] Liu, J., Moskvina, A., and Ourednik, M. (2017). Corpnet: Towards a decision support system for organizational network analysis using multiplex interpersonal relations. In *Multi-agent and Complex Systems*, volume 670, pages 33–47. Springer.

[71] Lokshtanov, D., Misra, N., Philip, G., Ramanujan, M. S., and Saurabh, S. (2013). Hardness of r-dominating set on graphs of diameter $(r-1)$. In *Proceeds of the 8th International Symposium Parameterized and Exact Computation (IPEC 2013)*, pages 255–267, Sophia Antipolis, France.

[72] Lucas, J. (2000). *Conceptual Roots of Mathematics*. Routledge.

[73] Martinez, A., Kane, R., Ferris, G., and Brooks, C. (2012). Power in leader-follower work relationships. *Journal of Leadership & Organizational Studies*, 19(2):142–151.

[74] McAuley, J. and Leskovec, J. (2012). Learning to discover social circles in ego networks. In *The Twenty-sixth Annual Conference on Neural Information Processing Systems*.

[75] McCarth, J. R. (2012). The powerful relational language of 'family': togetherness, belonging and personhood. *The Sociological Review*, 60(1):68–90.

[76] McPherson, M., Smith-Lovin, L., and Cook, J. (2001). Birds of a feather: Homophily in social networks. *Annual Review of Sociology*, 27:415–444.

[77] Meagher, K. (2003). Generalizing incentives and loss of control in an optimal hierarchy: the role of information technology. *Economics Letters*, 78(2):273–280.

[78] Morone, F. and Makse, H. (2015). Influence maximization in complex networks through optimal percolation. *Nature*, Nature:65–68.

[79] Morrison, E. W. (2002). Newcomers' relationships: The role of social network ties during socialization. *The Academy of Management Journal*, 45(6):1149–1160.

[80] Moskvina, A. and Liu, J. (2016a). How to build your network: A structural analysis. In *Proceedings of the 25th International Joint Conference on Artificial Intelligence (IJCAI 2016)*, pages 2597–2603. AAAI Press.

[81] Moskvina, A. and Liu, J. (2016b). Integrating networks of equipotent nodes. In *Proceedings of the 5th International Conference on Computational Social Networks (CSoNet'16)*, pages 39–50.

[82] Moskvina, A. and Liu, J. (2016c). Togetherness: An algorithmic approach to network integration. In *ASONAM '16 Proceedings of the 2016 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining*, pages 223–230.

[83] Newman, M. (2006). Modularity and community structure in networks. *Proceedings of the National Academy of Science of USA*, 103(23):8577–8696.

[84] Newman, M. (2008). *Mathematics of networks*. Palgrave Macmillan, Basingstoke.

[85] Newman, M. E., Watts, D. J., and Strogatz, S. H. (2002). Random graph models of social networks. *Proc. Nat. Acad. Sci. USA*, 99:2566–2572.

[86] Newman, M. E. J. (2003). The structure and function of complex networks. *SIAM Review*, 45(2):167–256.

[87] Onnela, J.-P., Saramäki, J., Hyvönen, J., Szabó, G., Lazer, D., Kaski, K., Kertész, J., and Barabási, A.-L. (2007). Structure and tie strengths in mobile communication networks. *Proceedings of the National Academy of Science*, 104(18):7332–7336.

[88] Padgett, J. F. and Ansell, C. K. (1993). Robust action and the rise of the medici. *American Journal of Sociology*, 98(6):1259–1319.

[89] Pask, G. (1980). The limits of togetherness. In *In S.H. Lavington (ed.) Proc. Information Processing'80*. North Rolland Publishing Co.

[90] Passy, F. (2003). Social networks matter. but how? In Diani, M. and McAdam, D., editors, *Social movement and networks: relational approaches to collective action*. Oxford University Press.

[91] Pfeffer, J. (1981). *Power in Organizations*. Pitman, Marshfield, MA.

[92] Raven, B. (1965). *Social Influence and Power*, pages 371–382. Holt, Rinehart, Winston, New York.

[93] Raven, B. (1992). A power interaction model on interpersonal influence: French and raven thirty years later. *Journal of Social Behavior and Personality*, 7(2):217–244.

[94] Redding, W. C. (1972). *Communications Within the Organization: An Interpretive Review of Theory and Research*. Industrial Communication Council.

[95] Ritter, T., Wilkinson, I. F., and Johnston, W. J. (2004). Managing in complex business networks. *Industrial Marketing Management*, 33(3):175–183.

[96] Robertson, B. (2007). Organization at the leading edge: Introducing holacracy. *Integral Leadership Review*, 7(3).

[97] Sales, S. (1966). Supervisory style and productivity: Review and theory. *Personnel Psychology*, 19:275–286.

[98] Sherman, A. J. (2011). *Mergers and Acquisitions from A to Z*. New York: American Management Association.

[99] Sherman, J., Smith, H. L., and Mansfield, E. R. (1986). The impact of emergent network structure on organizational socialization. *Journal of Applied Behavioral Science*, 22.

[100] Stein, W. A. (2012). Sage – a computer system for algebra and geometry experimentation. Technical report. http://wstein.org/sage.html.

[101] Strøm, K. and Müller, W. C. (1999). The keys to togetherness: Coalition agreements in parliamentary democracies. *The Journal of Legislative Studies*, 5(3-4):255–282.

[102] Stuart, T. E. (1998). Network positions and propensities to collaborate: An investigation of strategic alliance formation in a high-technology industry. *Administrative science quarterly*, 43(3):668–698.

[103] Stuart, T. E. (2000). Interorganizational alliances and the performance of firms: A study of growth and innovation rates in a high-technology industry. *Strategic management journal*, 8:791–811.

[104] Stuart, T. E. and Sorenson, O. (2007). Strategic networks and entrepreneurial ventures. *Strategic Entrepreneurship Journal*, 1(3–4):211–227.

[105] Tannenbaum, R. and Schmidt, W. (1973). How to choose a leadership pattern. *Harvard Business Review*, pages 162–180. May – June.

[106] Teubner, T. (2001). Flattening the organizational structure: Encouraging empowerment or reinforcing control? In *Research in Organizational Change and Development (Research in Organizational Change and Development, Volume 13)*, pages 147–168.

[107] Turan, V. and Köhler, S. (2015). A distributed algorithm for minimum distance-k domination in trees. *Journal of Graph Algorithms and Applications*, 19(1):223–242.

[108] Tyler, J., Wilkinson, D., and Huberman, B. (2003). Email as spectroscopy: Automated discovery of community structure within organizations. In *Proceedings of the First International Conference on Communities and Technologies*, pages 81–96, Amsterdam, Netherlands.

[109] Uzzi, B. and Dunlap, S. (2005). How to build your network. *Harvard Business Review*, 83(12):53–60.

[110] Wang, X. and Chen, G. (2003). Complex networks: Small-world, scale-free and beyond. *IEEE circuits and systems magazine*, pages 6–20.

[111] Wasserman, S. and Faust, K. (1994). *Social Networks Analysis: Methods and Applications*. Cambridge University Press.

[112] Weick, K. (1979). *The Social Psychology of Organizing, 2nd ed.* McGraw-Hill.

[113] Williamson, O. (1967). Hierarchical control and optimal firm size. *Journal of Political Economy*, 75:123–138.

[114] Wren, D. A., Bedeian, A. G., and Breeze, J. D. (2002). The foundations of henri fayol's administrative theory. *Management Decision*, 40(9):906–918.

[115] Wulf, J. (2012). The flattened firm - not as advertised. *California Management Review*, 55(1):5–23.