Opportunistic Mobile Social Networks

Edited by
Jie Wu and Yunsheng Wang

CRC Press
Taylor & Francis Group

Opportunistic Mobile Social Networks

This page intentionally left blank

Opportunistic Mobile Social Networks

Edited by
Jie Wu and Yunsheng Wang

CRC Press
Taylor & Francis Group
Boca Raton London New York

CRC Press is an imprint of the
Taylor & Francis Group, an **informa** business

CRC Press
Taylor & Francis Group
6000 Broken Sound Parkway NW, Suite 300
Boca Raton, FL 33487-2742

Version Date: 20140603

International Standard Book Number-13: 978-1-4665-9495-1 (eBook - PDF)

Visit the Taylor & Francis Web site at
http://www.taylorandfrancis.com

and the CRC Press Web site at
http://www.crcpress.com

Dedication

To our parents, Zengchang Wu, Yeyi Shao, Keming Wang, and Yingyan Wang

This page intentionally left blank

Contents

This page intentionally left blank

List of Figures

This page intentionally left blank

List of Tables

Preface

Over the past few decades, social networks have attracted massive interest from scholars in fields as diverse as sociology, biology, physics, business, politics, and computer science. From these diverse fields, researchers have found that many systems can be represented as networks, and that there is much to be learned by studying those networks. With the rapid growth of the Internet and the web, large-scale social network analysis has become possible for researchers. The most important difference between the traditional and new social networks is that the traditional theories of social networks have not been very concerned with the structure of naturally occurring networks. Traditional social network analysis is deep and elegant, but it is not especially relevant to networks arising in the real world. The emergence of recent mobile devices and their applications have brought about a new landscape in studying social networks.

The recent availability of mobile devices coupled with recent advancements in networking capabilities make opportunistic networks one of the most promising technologies for next-generation mobile applications. Opportunistic networks are commonly defined as a type of network where communication is challenged by sporadic and intermittent contacts, as well as frequent disconnections and reconnections, and where the assumption of the existence of an end-to-end path between the source and the destination is relinquished. Connectivity disruptions, limited network capacity, energy and storage constraints of those participating, mobile devices, and the arbitrary movement of nodes are only a few of the challenges that must be dealt with by the protocol stack. Clearly, current Internet protocols (i.e., the TCP/IP protocol stack) suffer and can fail under such conditions. Opportunities can be useful for building both ad hoc and delay-tolerant networks for data, but they can also be mined for information about mobility and social structures. However, to do either of these, users need to be persuaded to share resources, either at the information level, which impacts privacy, or at the communications level, which impacts their own network performance.

With new challenges brought up by the aforementioned emerging mobile technology in social networks and opportunistic networks, we have recently witnessed the rise of an emerging cross-disciplinary field called *opportunistic mobile social networks*, which has started to receive much attention from practitioners, scholars, and the general public. An opportunistic mobile social network can be described as a platform that provides services via hand held and wireless devices, mainly for the purpose of fostering and maintaining social interactions and connections. Therefore, an opportunistic mobile social network can be considered a form of social network where services are provided with mobility as an added value. Thanks to mobility, many emergent applications related to social networking are now available for individuals, business enterprises, and governments.

The convergence of social networks and opportunistic networks has its own implications for theory and practice. From a theoretical perspective, new research domains have emerged to tackle opportunistic mobile social networks from technological, social, behavioral, legal, and ethical standpoints. From a practical perspective, it is a topic that can notice new forms of collaboration, such as the one between mobile network operators and social networking sites, to offer new innovative services. Moreover, new opportunities are now available to individuals, business organizations, and governments, such as location-based services, content distribution systems, early warning systems in crisis management, and business cooperation monitoring. Indeed, these implications call for urgent attention to further investigate all related and significant issues of opportunistic mobile social networks, so as to advance our understanding and knowledge in this context.

The main goal of this book is to collect the recent development on theoretical, algorithmic, and application-based aspects of opportunistic mobile social networks. This book will be of particular value to academics, researchers, practitioners, government officials, business organizations (e.g., executives, marketing professionals, and resource managers), and even customers—those working in, participating in, or even those interested in fields related to social networks. The content of the book will be especially useful for students in areas like social networks, informatics, wireless networks, data mining, and administrative sciences and management, but also applies to students of education, economy, or law, who would benefit from the information, cases, and examples therein.

This book is expected to serve as a reference book for developers in the telecommunications industry, and for a graduate course in computer science and engineering. Our focus is to expose readers the technical challenges of opportunistic mobile social networking, and to offer some ideas on how we might overcome them. This book is organized in four areas with a total of 16 chapters. Each area corresponds to an important snapshot, according to what we believe, in this fast-growing field. Although several books have emerged recently in this area, none of them address all four areas in terms of critical issues and possible solutions.

- Fundamental concepts and models in opportunistic mobile social networks (Chapters 1–5)
- Routing and forwarding schemes in opportunistic mobile social networks (Chapters 6–9)

- Privacy, security, and economics in opportunistic mobile social networks (Chapters 10–12)
- Applications and testbeds in opportunistic mobile social networks (Chapters 13–16)

Introducing fundamental concepts and models in opportunistic mobile social networks, Chapter 1 presents a systematic analytical study of the constrained information flow problem, which models a pair of networks (social and communication) as a composite graph. Chapter 2 reviews the recent literature of social influence in complex social networks. Chapter 3 provides a comprehensive overview of the fundamental characteristics of link-level connectivity in opportunistic networks, which is crucial in understanding and evaluating network performance. Chapter 4 uses WiFi interactive to discover and predict temporal networks and human population dynamics. Chapter 5 shows how mobility and dynamic network structure impact the processing capacity of opportunistic mobile networks for cloud applications.

In a discussion of routing and forwarding schemes, which spans Chapters 6 to 9, Chapter 6 provides a comprehensive overview of the routing schemes proposed in opportunistic mobile social networks, with a focus on encounter-based unicasting and social-based unicasting. A brief overview of several multicast approaches is also given. Chapter 7 takes an in-depth look into multicast protocols, which are classified based on the number of copies of the multicast message for opportunistic mobile social networks. Chapter 8 focuses on providing pervasive data access to mobile users without the support of cellular or Internet infrastructures. Chapter 9 adopts a data-driven approach, which is based on multiple mobility traces collected from conferences, university campuses, and metropolitan cities, to address four challenges: efficiency, utilization, scalability, and trust.

Issues of privacy, security, and economics in opportunistic mobile social networks are examined in chapters 10 through 12. Chapter 10 applies privacy-preserving techniques with packet forwarding to enhance communication performance and protect users' sensitive information from disclosure. Chapter 11 surveys a collection of approaches that have been recently proposed in the literature to address the need for minimizing privacy leakage during opportunistic user profile exchange. Chapter 12 introduces economics concepts to help formalize the idea of incentives for rewarding long-term participation.

In the final area of applications and testbeds, Chapter 13 deals with a P2P search framework for intelligent crowdsourcing in opportunistic mobile social networks. Chapter 14 introduces a framework for mobile peer rating using a multi-dimensional metric scheme, based on encounter and location sensing. Chapter 15 investigates Vehicular Ad hoc NETworks (VANETs), as a particular class of opportunistic mobile social networks, under the assumption of social networking for vehicular applications (i.e., safety and entertainment applications). Chapter 16 develops a network emulation testbed called QOMB, that can be used to validate the efficient operation of opportunistic network applications and protocols in scenarios that involve both node mobility and wireless communication.

We would like to express our gratitude to all authors. This book would not be possible without their generous contributions. Our special thanks are given to CRC senior editor, Richard O'Hanley, for his encouragement and guidance. Finally, we want to thank our families for their support and patience during this project. Readers are welcome to send their comments and suggestions to jiewu@temple.edu and ywang@kettering.edu.

Jie Wu
Temple University

Yunsheng Wang
Kettering University

About the Editors

Jie Wu is chair and Laura H. Carnell Professor in the Department of Computer and Information Sciences at Temple University. He is also an Intellectual Ventures endowed visiting chair professor at the National Laboratory for Information Science and Technology, Tsinghua University. His current research interests include mobile computing and wireless networks, routing protocols, cloud and green computing, network trust and security, and social network applications. Dr. Wu regularly publishes in scholarly journals, conference proceedings, and books. He serves on several editorial boards and on organization committees of ACM and IEEE conferences. Dr. Wu is the recipient of the 2011 China Computer Federation (CCF) Overseas Outstanding Achievement Award.

Yungsheng Wang is an assistant professor in the Department of Computer Science in Kettering University, Flint, Michigan. He received a B.Eng. in electronic and information engineering from Dalian University of Technology, Dalian, China, in 2007; a M.Res. in telecommunication from University College London, London, in 2008; and a Ph.D. from the Department of Computer and Information Sciences, Temple University, Philadelphia, Pennsylvania, in 2013. His research interests include various topics in the application and protocols of wireless networks. Currently, his research focuses on the efficient communication in delay tolerant networks and opportunistic mobile social networks. Dr. Wang serves as track co-chair of the International Workshop on Mobile Sensing, Computing and Communication (MSCC) conducted in conjunction with ACM Mobihoc, 2014. He also serves on technical program committees of several IEEE conferences.

This page intentionally left blank

Contributors

Amotz Bar-Noy
Department of Computer Science
The Graduate Center of the City University of New York
New York, New York

Prithwish Basu
Advanced Networking Technologies
Raytheon BBN Technologies
Cambridge, Massachusetts

Ben Baumer
Department of Mathematics and Statistics
Smith College
Northampton, Massachusetts

Razvan Beuran
Hokuriku StarBED Technology Center
National Institute of Information and Communications Technology
Ishikawa, Japan

Eyuphan Bulut
Cisco Systems
Richardson, Texas

Guohong Cao
Department of Computer Science and Engineering
The Pennsylvania State University
University Park, Pennsylvania

Chi-Kin Chau
Department of Computing and Information Science
Masdar Institute
Masdar City, United Arab Emirates

Gianpiero Costantino
Istituto di Informatica e Telematica del CNR
Pisa, Italy

Michel Diaz
LAAS-CNRS
Toulouse, France

Do Young Eun
North Carolina State University
Raleigh, North Carolina

Wei Gao
Department of Electrical Engineering and Computer Science
The University of Tennessee
Knoxville, Tennessee

Khaled A. Harras
Computer Science Department
Carnegie Mellon University
Doha, Qatar

Ahmed Helmy
Department of Computer and Information Science and Engineering
University of Florida
Gainesville, Florida

Buster O. Holzbauer
Rensselaer Polytechnic Institute
Troy, New York

Wenjie Hu
Department of Computer Science and Engineering
The Pennsylvania State University
University Park, Pennsylvania

Andreas Konstantinidis
Department of Computer Science
University of Cyprus
Nicosia, Cyprus

Udayan Kumar
Department of Computer and Information Science and Engineering
University of Florida
Gainesville, Florida

Chul-Ho Lee
North Carolina State University
Raleigh, North Carolina

Xiang Li
Adaptive Networks and Control Laboratory
Department of Electronic Engineering
Fudan University
Shanghai, China

Xiaohui Liang
Department of Electrical & Computer Engineering
University of Waterloo
Waterloo, Ontario, Canada

Thomas D.C. Little
Department of Electrical and Computer Engineering
Boston University
Boston, Massachusetts

Chengyin Liu
Sun Yat-Sen University
Guangzhou, Guangdong, P.R. China

Cong Liu
Sun Yat-Sen University
Guangzhou, Guangdong, P.R. China

Rongxing Lu
Department of Electrical & Computer Engineering
University of Waterloo
Waterloo, Ontario, Canada

Fabio Martinelli
Istituto di Informatica e Telematica del CNR
Pisa, Italy

Subhankar Mishra
Department of Computer and Information Science and Engineering
University of Florida
Gainesville, Florida

Shinsuke Miwa
Hokuriku StarBED Technology Center
National Institute of Information and Communications Technology
Ishikawa, Japan

Toshiyuki Miyachi
Hokuriku StarBED Technology Center
National Institute of Information and Communications Technology
Ishikawa, Japan

Abderrahmen Mtibaa
Electrical & Computer Engineering
Texas A & M University
Doha, Qatar

Anh-Dung Nguyen
Department of Mathematics Computer Science and Automatic Control
ISAE, University of Toulouse
Toulouse, France
LAAS-CNRS, Toulouse, France

Paolo Santi
Istituto di Informatica e Telematica del CNR
Pisa, Italy

Patrick Senac
Department of Mathematics Computer Science and Automatic Control
ISAE, University of Toulouse
Toulouse, France
LAAS-CNRS, Toulouse, France

Xuemin (Sherman) Shen
Department of Electrical & Computer Engineering
University of Waterloo
Waterloo, Ontario, Canada

Yoichi Shinoda
Research Center for Advanced Computing Infrastructure
Japan Advanced Institute of Science and Technology
Ishikawa, Japan

Boleslaw K. Szymanski
Rensselaer Polytechnic Institute
Troy, New York

My T. Thai
Department of Computer and Information Science and Engineering
University of Florida
Gainesville, Florida
and
Ton Duc Thang University
Ho Chi Minh City, Vietnam

Athanasios V. Vasilakos
Department of Electrical and Computer Engineering
National Technical University of Athens
Athens, Greece

Anna Maria Vegni
Department of Engineering
University of Roma Tre
Rome, Italy

Wei Wang
Sun Yat-Sen University
Guangzhou, Guangdong, P.R. China

Yunsheng Wang
Department of Computer Science
Kettering University
Flint, Michigan

Jie Wu
Department of Computer and Information Sciences
Temple University
Philadelphia, Pennsylvania

Demetrios Zeinalipour-Yazti
Department of Computer Science
University of Cyprus
Nicosia, Cyprus

Huiyuan Zhang
Department of Computer and Information Science and Engineering
University of Florida
Gainesville, Florida

Kuan Zhang
Department of Electrical & Computer Engineering
University of Waterloo
Waterloo, Ontario, Canada

Yi-Qing Zhang
Adaptive Networks and Control Laboratory
Department of Electronic Engineering
Fudan University
Shanghai, China

Chapter 1

Social-Communication Composite Networks

Prithwish Basu
Advanced Networking Technologies
Raytheon BBN Technologies
Cambridge, Massachusetts

Ben Baumer
Department of Mathematics and Statistics
Smith College
Northampton, Massachusetts

Amotz Bar-Noy
Department of Computer Science
The Graduate Center of the City University of New York
New York, New York

Chi-Kin Chau
Department of Computing and Information Science
Masdar Institute
Masdar City, United Arab Emirates

CONTENTS

1.1 Introduction

The recent explosive growth in online social networks has been fueled by the proliferation of high-speed and highly available communication networks such as the Internet and broadband cellular wireless networks, as well as the increasing popularity of mobile network-ready devices such as “smartphones” and tablets. People tend to share information with other people they know, who subsequently forward that information along various links in the social network—this occurs either verbatim (for example, the directives from a commander flow through the *chain of command*) or after modifications (for example, propagation of rumors, gossip, or news on Twitter). A social network’s topology thus *constrains* or *guides* the flow and spread of information through it. These constraints can force the information to traverse much longer paths in the underlying communication network between its originator and its ultimate consumers. This phenomenon, known as *stretch*, is justified because the intermediaries may play a critical role in interpreting or modifying the information or they may serve as important links in the acquaintance chain, without whom the originator and the ultimate consumers would not have known each other.

When information gets *stretched*, the total time for it to spread through the entire social network is often different from the time taken to simply *multicast* the information on the underlying communication network to the set of ultimate consumers. An additional undesirable side-effect of the “stretch” phenomenon is that an information object may traverse a communication link or a node several times during the process, thus increasing resource consumption. While this is not a major issue for lightweight content such as text (e.g., 140 character Twitter messages), it can be a significant problem for multimedia content, especially in mobile ad hoc network (MANET) or disruption-tolerant network (DTN) settings where accessing multimedia content directly from a server over a flaky network may not be feasible.

In this chapter, we present a systematic analytical study of the constrained information flow problem—in particular, we model a pair of networks (social and communication) as a *composite graph*—a structure that results from embedding or mapping the social network into the communication network using *embedding / mapping functions*. A mapping function maps a node in the social network to one or more in the communication network when the former *uses* the latter as his/her communication portal(s). We consider unicast, broadcast, and multicast versions of this scenario. We introduce several "composite graph" metrics that capture the effect of constraining the flow of information in the communication network due to the social network, for example, *composite path stretch*, *composite broadcast time*, *composite betweenness centrality*, etc. We analytically study how these metrics scale with the sizes of both networks under consideration under various random graph models and mapping functions. The above modeling / analysis can be useful in an application scenario such as the following: workers or soldiers equipped with wireless communication devices have been deployed at a disaster relief site and their group leader disseminates messages to them following a specific chain of command, which is essentially a social network. These messages trace a logical path in the social network that translates to a potentially longer physical path in the underlying communication network (which is a MANET or a DTN).

Information multicast through a chain-of-command hierarchy can also be modeled in the composite graph framework. For many operations and missions in practice, mere topological proximity to certain recipients of a message does not warrant its direct delivery to the latter. Instead, certain hierarchical policies that define different roles and ranks of network nodes may constrain the message flow through the network. For example, in military networks, communications between various nodes may need to be observed and then cleared by individuals located higher in the chain-of-command hierarchy, which is nothing but a social network. It is often the case that a subset of nodes in the hierarchy are interested in participating together in a multicast session. Therefore, we are motivated to construct multicast "routes" that *connect* these nodes while being constrained by the relationships in the hierarchy.

1.1.1 Related Work

There are three classes of related work in this area: graph embedding, network science approaches to studying composite networks, and overlay networks in the Internet.

Graph embedding has received attention in the parallel computing domain where the problem is to map a *task graph* onto a multiprocessor interconnection network (also known as *host graph*) [6, 23, 15], and in the ubiquitous computing domain where the problem is to map heterogeneous task graphs on non-regular networks such as mobile ad hoc networks [4], while attempting to determine the optimal mapping (or task to processor assignment) function such that metrics such as delay-to-task-completion, edge dilation (or stretch), node/edge congestion, etc., are minimized. Instead of the aforementioned "optimization" approaches, in this chapter, we follow the "scaling law analysis" approach where both the graphs and the mapping function are given (deterministic or stochastic), and we characterize how a different set of appropriate "constrained" metrics such as composite path stretch, composite diameter, broadcast time, and composite betweenness centrality scale as a function of composite graph attributes.

There is a large body of work pertaining to the embedding of one metric space into another—in particular, normed spaces such as d-dimensional Euclidean space $\mathbb{R}^d$)—with "low-distortion." This has been summarized well in [16]. This entails establishing the necessary and sufficient conditions on the properties of the two spaces for finding such embedding functions that yield a particular distortion, and in many cases finding the best embedding function [2]. A related idea of finding embeddings is popular in geographic routing—virtual coordinates are assigned to nodes in a hyperbolic space, and such an embedding guarantees that a greedy algorithm on the virtual coordinate space yields a route between every source and destination, if one exists [19].

Various flavors of layered or composite networks have received some attention in the network science literature. Kurant and Thiran propose the Layered Complex Network model [20] for studying load in transportation networks. They considered 2-layer graphs where the physical graph corresponds to the transportation network and the logical graph corresponds to the traffic flow between various cities—they use computational methods to determine different levels of *load* on various transportation sectors in Europe. In comparison, our approach is analytical and we study metrics that have not been studied in [20]. A recent analytical line of research considers interdependent networks such as power grid and communication networks [8]—they use percolation theory to determine the fraction of nodes whose removal is likely to generate cascading failures in such networks. Leicht and D'Souza show that percolation thresholds of composite networks is lower than the individual networks, when considered separately [21]. While these approaches are all analytical, they study a different graph metric, i.e., degree of failure tolerance.

Overlay networks have received a lot of attention in the computer networking literature in the past decade [22]. Works such as CAN [25] and CHORD [28] attempt to design good distributed hash tables for P2P applications—for storing (key, value) pairs *overlaid* on top of the Internet, so that efficient insertion and retrieval of hashed content is feasible from any part of the network. While this is a good example of a composite network, its similarities with our approach are slim. While overlay networks attempt to design good overlay graphs for the purpose of optimization of insertion/lookup overhead, in our problem space, the social network graph is given, and we are interested in a different set of information flow metrics. Moreover, unlike the Internet, which is a complete graph (or clique) for the purpose of connectivity in P2P applications, our underlying network is a multi-hop network, in general.

The focus of this chapter is not *to find* the best embedding function that yields a low distortion—rather, it is *to analyze* the distortion (or stretch) of an information flow that results from a *random* embedding of the nodes of the first graph onto the second graph, in distribution or in expectation. The material in this chapter has been derived in part from two recent publications co-authored by us [5, 3].

Our contributions in this chapter can be summarized as follows:

1. Novel models and metrics for constrained information flow in composite networks.

2. Mathematical analysis of scaling laws for constrained composite path stretch when a social network path is randomly mapped onto a general graph under both one-to-one and many-to-one mappings.

3. Scaling laws for constrained composite broadcast time of a tree social network (chain of command) randomly mapped onto different communication networks.

4. A hierarchy-compliant multicast algorithms for composite network multicast.

5. Validation of a subset of these results using two historical deployments of military chain-of-command networks as well as the FOAF (friend of a friend) data set embedded on a geometric communication graph.

We show that the composite betweenness centrality metric yields significantly better insights about the structure of a communication network compared to classic betweenness centrality computed on a single network. We also demonstrate that one has to be willing to pay a 25% overhead for adhering to the social network structure in certain realistic composite network multicast deployment scenarios.

1.2 Composite Graph Models

We define the *composite graph* $\mathcal{G}$ of two graphs G_1 and G_2 to be the 3-tuple (G_1, G_2, R), where $R \subseteq V(G_1) \times V(G_2)$ is an *embedding / mapping relation* between the vertex sets $V(G_1)$ and $V(G_2)$ of the two graphs, respectively. In general, every element of R may have multiple attributes associated with it but in this preliminary study we only consider a binary relation. This relation may be time-varying when information is replicated or moves from one communication node to another over time. Time-varying relations are outside the scope of this chapter.

1.2.1 Metrics on Composite Graphs

We first define *constrained composite path stretch*, a metric that is useful for measuring how many physical communication hops are spanned by a logical information flow under a given embedding of the logical flow on a physical network.

Throughout this chapter, let $\mathcal{G} = (G_1, G_2, R)$ be a composite graph, with $V_i = V(G_i)$ the vertex set of graph i and R an embedding relation as mentioned above. Unless otherwise noted, $P_k = P_{uv} = \{u = v_0, v_1, ..., v_k = v\}$ is a path of length k in

G_1, and $d_{G_2} : V_2 \times V_2 \to \mathbb{R}$ is a shortest path distance metric in G_2. For clarity, we introduce the notion of an *itinerary* in a graph.

Definition 1.1 (Itinerary) Given a list of vertices $v_0, v_1, \ldots, v_k$ in a graph, an *itinerary* is a not necessarily simple path passing through $v_0, v_1, \ldots, v_k$ *in order*, for which the path connecting consecutive vertices (v_i, v_{i+1}) is a shortest path, for all $0 \leq i \leq k-1$.

Intuitively, an itinerary is the shortest possible path through a sequence of not necessarily neighboring vertices.

Definition 1.2 (cstretch) Given composite graph $\mathcal{G} = (G_1, G_2, R)$, the constrained composite path stretch of $P_{uv} = \{u = v_0, v_1, ..., v_k = v\}$ in $\mathcal{G}$ is defined as:

$$cstretch_{G_2}(P_{uv}) = \sum_{i=0}^{k-1} \max_{\substack{s,t \in V_2: \\ (v_i,s) \in R \wedge (v_{i+1},t) \in R}} \{d_{G_2}(s,t)\}. \tag{1.1}$$

Equivalently, $cstretch_{G_2}(P_{uv})$ is the longest itinerary through the vertices in G_2 that are images of the vertices of P_{uv} in G_1 under the mapping R. Note that in general, R is not necessarily a bijection, and so there may be multiple vertices in G_2 that correspond to a single vertex in G_1.

CStretch characterizes the scenario with a stringent requirement that the information needs to traverse the nodes in the path P_{uv} *in order*, and in the process need to traverse the appropriately mapped nodes in G_2. This is not a far-fetched scenario—in military systems, the chain-of-command (modeled by graph G_1) often mandates a piece of information to flow through the logical chain even though the ultimate recipient of the information may be in close proximity to the origin and the intermediate nodes are farther away from them. The reason behind this is that information often needs to get refined or obfuscated at each level of the logical chain before it is passed on further. Similarly, even in non-military applications (such as online social networks such as Twitter) information such as news or gossip is often routed along logical paths of friends who may be physically located all over the globe at large "Internet distances" from each other.

In the composite graph setting, the notion of diameter[1] can be extended to that of the *constrained composite diameter*, which can be defined in terms of constrained composite path stretch.

Definition 1.3 (ccd) The constrained composite diameter of $\mathcal{G}$ is defined as

$$ccd(\mathcal{G}) = \max_{u,v \in V_1} cstretch_{G_2}(P_{uv}). \tag{1.2}$$

[1] Diameter is the maximum length of the shortest path between any pair of nodes in a graph. It is an important measure for communication networks because it gives us a sense of the amount of time required (in the worst case) to traverse a network.

The *CStretch* metric captures the extra distance in G_2 that a message has to travel in order to move through a path in G_1. We need a different metric to capture the combined stretch for a message traveling through a chain-of-command *tree* in a composite graph. In this context, it is more natural to consider the *constrained composite broadcast time* metric.

Definition 1.4 (cbtime) Let T be a tree in G_1, with root u. Then the constrained composite broadcast time of T in the composite graph $\mathcal{G}$ is defined as

$$cbtime_{G_2}(T) = \max_{v \in T} cstretch_{G_2}(P_{uv}). \tag{1.3}$$

The constrained composite broadcast time represents the stretch necessary to send a message through a chain-of-command tree that is deployed in a network topology. This may be of interest, for example, in a disaster relief situation when information needs to travel from a central director to end caregivers while relief workers are deployed in the field. In other words, it measures the time at which the last worker received the message that was broadcast through the chain of command.

We are also interested in measuring the traffic load on a particular edge in G_2 as a result of the flows along the edges in G_1.

Definition 1.5 (Load Indicator) For a specific edge $e = (x, y) \in G_2$, we say the edge bears a load from $v_i, v_{i+1} \in V(G_1)$ in the composite graph (G_1, G_2, R) if and only if e lies along a shortest path from a vertex $w_i \in G_2$ to $w_j \in G_2$, where $(v_i, w_i) \in R$ and $(v_{i+1}, w_j) \in R$. Let P_{ij} be any shortest path from $w_i \in G_2$ to $w_j \in G_2$. Then,

$$\chi_e(v_i, v_{i+1}) = \begin{cases} 1 & \text{if } e \in P_{ij} \text{ and } ((v_i, w_i), (v_{i+1}, w_j)) \in R \\ 0 & \text{otherwise.} \end{cases}$$

Definition 1.6 (cload) Let $P_{uv} = \{u = v_0 \to v_1 \to v_2 \to \cdots \to v_k = v\}$ be a path in graph G_1. Then the composite load on $e \in E(G_2)$ of P_{uv} in G_2 is defined as:

$$cload_{G_2}(P_{uv}, e) = \sum_{i=0}^{k-1} \chi_e(v_i, v_{i+1}).$$

Note that $0 \leq cload_{G_2}(P_{uv}, e) \leq k$. Naturally, we want to determine the maximum and expected measures of load upon any edge in G_2.

Finally, we extend the notion of *betweenness centrality* to composite graphs in order to measure the load on certain vertices and edges in G_2.

Definition 1.7 (cvbc) If (G_1, G_2, R) is a composite graph, let σ_{st} be the number of shortest paths in G_2 between s and $t \in V(G_2)$, and $\sigma_{st}(u)$ be the number of shortest paths in G_2 between s and t which pass through vertex u. Then the *composite vertex* *betweenness centrality* of a vertex $u \in V(G_2)$ is given by

$$cvbc(u) = \sum_{\substack{s \neq u \neq t \in V(G_2) \\ (v,w) \in E(G_1) \wedge \{(v,s),(w,t)\} \subseteq R}} \frac{\sigma_{st}(u)}{\sigma_{st}}. \tag{1.4}$$

Definition 1.8 (cebc) If (G_1, G_2, R) is a composite graph, let σ_{st} be the number of shortest paths in G_2 between s and $t \in V(G_2)$, and $\sigma_{st}(e)$ be the number of shortest paths in G_2 between s and t which pass through edge e. The *composite edge betweenness centrality* of an edge $e \in V(G_2)$ is given by

$$cebc(e) = \sum_{\substack{s \neq t \in V(G_2) \\ (v,w) \in E(G_1) \wedge \{(v,s),(w,t)\} \subseteq R}} \frac{\sigma_{st}(e)}{\sigma_{st}}. \tag{1.5}$$

1.3 Composite Stretch Analysis

In this section, we focus on analyzing *random embedding relations*, where vertices in G_1 are mapped to vertices in G_2 via some random process π. In particular, we study two cases:

1. Each vertex in G_1 is mapped to a vertex in G_2 that has been sampled uniformly at random *with replacement*. This is the many-to-one scenario, where many "social network" nodes can get mapped to the same communication network node.

2. Each vertex in G_1 is mapped to a vertex in G_2 that has been sampled uniformly at random *without replacement*. This is the one-to-one scenario, where a communication network node can host at most one social network node.

Specifically, we characterize the distribution of the constrained composite path stretch of P_k over uniform random embeddings into G_2. We first prove some general results that apply to any graph G_2, and then illustrate scaling laws for a few well-known graph families.

1.3.1 Theoretical Results

For any graph $G = (V, E)$, let D_G be the geodesic graph distance matrix between all pairs of vertices $v_i, v_j \in V$. That is, each entry d_{ij} in D_G represents the shortest path distance from v_i to v_j in G. Then we note that the sum of the geodesic distances $\Delta_G = \sum_{v_i, v_j \in V} d_{ij}$, is a constant depending only on the structure of G.

Lemma 1.1
Let G be a graph with $|V| = n$, and let X be a random variable denoting the geodesic

distance between two vertices of G chosen uniformly at random. Then:

$$\mathbb{E}[X] = \begin{cases} \frac{\Delta_G}{n(n-1)}, & \textit{when sampling without replacement} \\ \frac{\Delta_G}{n^2}, & \textit{when sampling with replacement.} \end{cases}$$

Proof 1.1 The case where sampling is done with replacement is clear: since there are n^2 pairs of vertices from which to choose, the expression given is the average distance. If sampling is done without replacement, then Δ_G double-counts the distance for each of the $\binom{n}{2}$ unique pairs of vertices. Note that the n diagonal entries in D_G contribute nothing to Δ_G.

Corollary 1.3.1 *There is no asymptotic difference in $\mathbb{E}[X]$ between sampling vertices with or without replacement.*

Proof 1.2 From the preceding lemma, it follows that the ratio of $\mathbb{E}[X]$, when sampling without replacement, to $\mathbb{E}[X]$, when sampling with replacement, is $1 + \frac{1}{n} \to 1$ as $n \to \infty$.

Next, we show that the expected stretch of a link is independent of the choices of vertices already mapped, regardless of whether sampling is done with or without replacement.

Lemma 1.2
Let $v_1, v_2, ..., v_i$ be a sequence of vertices chosen uniformly at random from V (with or without replacement), and let X_i be the random variable giving the distance between v_i and v_{i-1}. Then $\mathbb{E}[\mathbb{E}[X_{i+1}|v_1, v_2, ..., v_i]] = \mathbb{E}[X_2]$.

Proof 1.3 While the statement may be obvious for the case of sampling with replacement, we exercise more care for the case where sampling is done without replacement, and prove the statement combinatorially. For the RHS, select one vertex uniformly at random and color it red (call it v_1). Then select another from the remaining and color it blue (v_2). The RHS counts the expected distance between these two vertices. We now argue that the LHS counts the same. To see this, first color one vertex blue (call it v_{i+1}), and another vertex red (v_i). Now color $i-1$ other vertices green ($v_{i-1}, ..., v_1$). The LHS counts the expected distance between the blue vertex and the red vertex.

This leads us to a general theorem about the expected composite stretch of a path.

Theorem 1.3.1 *For a path P_k embedded uniformly at random into any graph G_2 (with the sampling performed with or without replacement),*

$$\mathbb{E}[cstretch^{\pi}_{G_2}(P_k)] = k \cdot \mathbb{E}[X], \tag{1.6}$$

where X is the random variable giving the distance between two randomly chosen vertices in G_2.

We emphasize that the expectation is being taken over the uniform random embedding R_π. But as we saw in Lemma 1.1, for a specific G_2, if the sampling method of R_π is known, then the expected distance $\mathbb{E}[X]$ is a constant.

Composite Diameter: In addition to the average case, we also want to describe the worst-case *cstretch* for a random embedding. It is easy to see that if R_π samples vertices with replacement, then each successive link in any path can simply bounce back and forth between the furthest two vertices in G_2. Thus, $ccd(\mathcal{G}) = diam(G_1) \cdot diam(G_2)$. However, when R_π samples vertices without replacement, the problem is an instance of MAX-TSP, which is MAX SNP-hard [14]. However, a greedy approximation heuristic works well in practice.

1.3.2 Composite Stretch of Some Special Graphs

Theorem 1.3.1 shows that the expected stretch of a path is equal to the length of the path times a constant depending only on the structure of G_2 and the distribution of the random embedding. In what follows, we present examples of some well-known graph families, and illustrate how their structure affects the distribution of *cstretch*.

d-dimensional Discrete Lattice: Let $D_n^d = \{0, 1, ..., n-1\}^d$ be the d-dimensional discrete lattice on n^d points, and consider a composite graph with $G_2 = D_n^d$. On this graph topology, geodesic distance is equivalent to the ℓ_1-norm (Manhattan distance) between two points in D_n^d. Thus, $d_{G_2}(v,w) = \sum_{i=1}^d |v_i - w_i|$, and summing all n^{2d} of these pairs gives

$$\Delta_{G_2} = \sum_{v,w \in V} \sum_{i=1}^{d} |v_i - w_i| = \frac{dn^{2d+1}}{3}\left(1 - \frac{1}{n^2}\right). \tag{1.7}$$

It follows from Lemma 1.1 and Theorem 1.3.1 that under a random uniform embedding with replacement into the d-dimensional discrete lattice,

$$\mathbb{E}[cstretch^{\pi}_{G_2}(P_k)] = \frac{kdn}{3}\left(1 - \frac{1}{n^2}\right). \tag{1.8}$$

Note that in this case it is also straightforward to fully explicate the distribution of X. For any $1 \le i \le d$, let $X_i = |v_i - w_i|$. Then the probability mass function for X_i is

$$p_{X_i}(\delta) = \begin{cases} \frac{1}{n} & \text{if } \delta = 0 \\ \frac{2(n-\delta)}{n^2} & \text{otherwise} \end{cases}, \tag{1.9}$$

since each coordinate can take on any of n values, and there are $n - \delta$ ways to achieve each value of δ between 0 and $n-1$. Since the X_i's are independent and identically distributed, we can extract (among other things), the second moment of X:

$$\mathrm{Var}[X] = d \cdot \frac{(n^2-1)(n^2+2)}{18n^2}. \tag{1.10}$$

We can infer from this that the expected stretch is not likely to deviate significantly from its mean.

For the discrete lattice, we have that $diam(G_2) = d(n-1)$, so as mentioned above, the *ccd* for P_k is $k(n-1)$. For the non-trivial "without replacement" scenario, we implemented a greedy approximation heuristic, and verified that *ccd* for both without and with replacement scenarios are $O(n^2)$.

Cycle: Let C_n be the cycle of length n, and consider uniform discrete mappings from P_k onto C_n. Clearly, the maximum distance between two vertices in C_n is $\lfloor \frac{n}{2} \rfloor$. But, for each possible distance x between 0 and $\frac{n}{2}$, there are exactly n such pairs for $x = 0, \frac{n}{2}$, and exactly $2n$ such pairs otherwise (we assume that in the case of a tie, only one shortest path is kept). It is thus straightforward to show that

$$\Delta_{C_n} = \begin{cases} \frac{n(n^2-1)}{4} & \text{if } n \text{ is odd} \\ \frac{n^3}{4} & \text{if } n \text{ is even} \end{cases}. \tag{1.11}$$

Application of Lemma 1.1 and Theorem 1.3.1 then reveal that for random uniform embeddings onto C_n,

$$\mathbb{E}[cstretch^{\pi}_{C_n}(P_k)] = k \cdot \left(\frac{n}{4} + o(1)\right). \tag{1.12}$$

Greedy is optimal on C_n, since if n is odd, it finds $n-1$ pairs at distance $\lfloor \frac{n}{2} \rfloor = diam(G_2)$ from each other, which is optimal by definition. On the other hand, if n is even, it picks *all* $\frac{n}{2}$ pairs at distance $\frac{n}{2} = diam(G_2)$ from each other, and another $\left(\frac{n}{2} - 1\right)$ pairs at the next greatest distance $\left(\frac{n}{2} - 1\right)$.

Balloon graph: Next, we consider a graph family with some interesting properties. Let $B_{n,m}$ be a balloon graph consisting of a string (line graph) of length m, connected to a balloon (clique) of size $n-m$, for any $0 \leq m < n$. For clarity, we specify that vertices $\{v_0, ..., v_m\}$ make up the string, while vertices $\{v_m, ..., v_{n-1}\}$ make up the balloon (see Figure 1.1). Note that for any two indices $0 \leq i < j \leq n-1$ in this graph, we have that

$$d_{B_{n,m}}(v_i, v_j) = \begin{cases} j - i & \text{if } i < j \leq m \\ m + 1 - i & \text{if } i < m \leq j \\ 1 & \text{if } m \leq i < j. \end{cases}$$

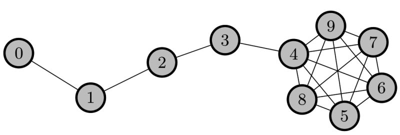

Figure 1.1: The balloon graph $B_{10,4}$.

In particular, note that $diam(B_{n,m}) = m+1$. In computing the distance matrix, we distinguish three cases based on the indices of the two vertices chosen:

1. If $i \leq j \leq m$, then both vertices lie in the string, which is D^1_{m+1}. This contributes $\Delta_{D^1_{m+1}}$ toward $\Delta_{B_{n,m}}$.
2. If $m \leq i \leq j$, then both vertices lie in the balloon, and it is clear that on the complete graph $K_n, \Delta_{K_n} = n^2 - n$, since every pair of vertices are connected by an edge, but there are n ways to choose the same vertex twice.
3. If $i < m < j$, then one vertex lies in the string, and the other lies in the balloon. Consider any vertex w_j in the balloon. Its distance from the set of vertices in the string is simply $m+1, m, m-1, ..., 2$. Thus, the contribution to $\Delta_{B_{n,m}}$ is

$$2(n-m-1)\sum_{i=2}^{m+1} i = m(m+3)(n-m-1).$$

Adding these three quantities yields

$$\Delta_{B_{n,m}} = -\frac{2}{3}m^3 + (n-2)m^2 + \left(n - \frac{4}{3}\right)m + n^2 - n. \tag{1.13}$$

The reader may verify that setting $m = 0$ corresponds to the special case where the balloon graph is itself a clique, while setting $m = n-1$ yields the special case where $B_{n,n-1} = D^1_n$.

By Theorem 1.3.1 and Lemma 1.1, the expected stretch for a path of length k onto $B_{n,m}$ is thus:

$$\mathbb{E}[cstretch^{\pi}_{B_{n,m}}(P_k)] = k \cdot \left(1 + O\left(\frac{m^2}{n}\right)\right) \tag{1.14}$$

Random Geometric Graph: Lastly, we consider the composite stretch when P_k is mapped onto a random geometric graph $G_2 = RGG(n, r(n))$, where $r(n)$ is the radius of communication. That is, G_2 consists of n vertices placed uniformly at random in $[0,1]^2$, wherein any two vertices are connected with an edge if and only if the Euclidean distance between them is at most $r(n)$. Gupta and Kumar [13] showed that a radius of connectivity of $r(n) = \sqrt{\frac{\ln n + c(n)}{\pi n}}$ ensures asymptotic connectivity in the RGG with high probability if and only if $c(n) \to +\infty$. In all of our discussions on RGG in this chapter, we assume that the radius of connectivity is at least this large, i.e., $r(n) = \Omega(\sqrt{\ln n / n})$.

As before, Theorem 1.3.1 still applies, so it remains only to characterize the distribution of the random variable X giving the geodesic distance between two vertices in $RGG(n, r(n))$ selected uniformly at random. Note that in contrast to the previous examples we have considered, we now have two sources of randomness: 1) the randomized construction of the RGG; and 2) the random uniform embedding. If the

Euclidean distance between two vertices in an RGG is δ, then recent results confirm that with high probability, the geodesic distance X differs from its minimum of δ/r by at most a constant [7].

Theorem 1.3.2 *With high probability, the expected geodesic distance in* $RGG(n,r(n))$ *satisfies*

$$\frac{\Delta(2)}{r(n)} \leq \mathbb{E}[X] \leq \kappa(n) \cdot \frac{\Delta(2)}{r(n)}, \tag{1.15}$$

where $\Delta(2) \approx 0.5214054331$ *is a known constant, and* $\kappa(n) \geq 1$ *is* $O(1)$.

Proof 1.4 Let v,w be two vertices in $RGG(n,r(n))$ selected uniformly at random, and set $\delta = ||v-w||_2$. Clearly, $X \geq \delta/r$. Conversely, if $\delta = \Omega(\log^{3.5} n/r^2)$, then by a result from [7], $X = O(\delta/r)$.

Taking expectation yields the result, since $\mathbb{E}[\delta] = \Delta(2)$ is a known constant [29].

Synthetic analysis suggests that $\kappa(n) < 1.3$ for $n > 1000$. Therefore, as before, we can easily bound (from above) the expected composite stretch.

Corollary 1.3.2 *For* $r(n)$ *sufficiently large (i.e., greater than the critical connectivity threshold), the composite stretch of a path* P_k *on a random geometric graph* $RGG(n,r(n))$ *satisfies with high probability:*

$$\mathbb{E}[cstretch^{\pi}_{RGG}(P_k)] = k \cdot \kappa(n) \cdot \frac{\Delta(2)}{r(n)} = O\left(k \cdot \sqrt{\frac{n}{\ln n}}\right). \tag{1.16}$$

1.3.3 Average vs. Worst-Case Analysis

We have so far characterized the average case (expected $cstretch$) and the worst case (ccd) for a random uniform embedding of a path onto several graph families. For both the lattice and the cycle, these quantities were of the same order of magnitude. A natural question is:

> Are there graphs for which the ratio of the maximum $cstretch$ to the average $cstretch$ of P_k is *not* $O(1)$?

Indeed, the balloon graph is one such graph. As the diameter of $B_{n,m}$ is $m+1$, the maximum stretch is $diam(G_1) \cdot (m+1)$. If we let $\phi(B_{n,m})$ be the ratio of the maximum $cstretch$ to the mean $cstretch$, we can see that:

$$\phi(B_{n,m}) = \frac{diam(G_1)(m+1)}{diam(G_1)\left(1+O\left(\frac{m^2}{n}\right)\right)} = O\left(\frac{n}{m}\right).$$

In particular then, for $m = \sqrt{n}$, the ratio of the maximum stretch to the mean stretch for the balloon graph $B_{n,m}$ is $O(\sqrt{n})$. Explicit calculations reveal that for $m = \sqrt{n}$, in fact $\mathbb{E}[X] \to 2$ as $n \to \infty$.

Table 1.1 Summary of Path Stretch Metrics for Uniform Random Embeddings of P_k

G_1	G_2	$\mathbb{E}[cstretch]$	$\max[cstretch]$
P_k	D_n^d	$\frac{kdn}{3}\left(1-n^{-2}\right)$	$kd(n-1)$
	C_n	$k\cdot\left(\frac{n}{4}+o(1)\right)$	$k\cdot\lfloor\frac{n}{2}\rfloor$
	$B_{n,m}$	$k\cdot\left(1+O\left(\frac{m^2}{n}\right)\right)$	$k(m+1)$
	$RGG(n,r(n))$	$O\left(k\sqrt{\frac{n}{\ln n}}\right)$	

More interesting is the fact that this gap appears to be mainly an artifact of the difference between sampling with and without replacement. The results of our greedy algorithm for CCD without replacement suggest that with $m=\sqrt{n}$; the CCD and expected *cstretch* are of the same order of magnitude.

Table 1.1 summarizes our theoretical results.

1.4 Composite Broadcast Time

In this section, we analytically characterize the expected composite broadcast time for tree topologies. Social networks for information dissemination commonly have tree structures (more on this in Section 1.7), hence this analysis can be useful for specific communication network deployment scenarios. Let T_k be a k-node tree of height h and maximum (out)degree δ, for some $1\le\delta<k$. We assume that T_k exists in some G_1, and examine the constrained composite broadcast time for sending a message from the root to each of the other nodes.

Star Topology: We begin with the special case where T_k is a k-star. First, we introduce a notation. Let

$$p_k=\frac{1}{\binom{n-1}{k}}\left(\underbrace{0,\ldots,0}_{k},1,\binom{k}{k-1},\ldots,\binom{n-2}{k-1}\right)\in\mathbb{R}^n$$

be a column vector, and note that $||p_k||_1=1$. The i^{th} entry in p_k represents the probability that the i^{th} largest among n values is returned, when this value is the maximum among a subset of size k chosen uniformly at random. Furthermore, let $f:\mathbb{R}^{m\times n}\to\mathbb{R}^{m\times n}$ be the function that sorts the rows of a matrix in ascending order from left to right. That is,

$$D=\begin{bmatrix}d_1\\d_2\\\vdots\\d_m\end{bmatrix}\Rightarrow f(D)=\begin{bmatrix}sort(d_1)\\sort(d_2)\\\vdots\\sort(d_m)\end{bmatrix},$$

where d_i is the i^{th} row of D. Finally, $v_m=\frac{1}{m}(1,\ldots,1)\in\mathbb{R}^m$.

Theorem 1.4.1 *For any graph G_2, the broadcast time of a star of size k satisfies*

$$\mathbb{E}[cbtime_{G_2}(S_k)] = v_n^T \cdot f(D_{G_2}) \cdot p_k. \tag{1.17}$$

Proof 1.5 Let d_i be the i^{th} row of D_{G_2}, and suppose that the root of the star S_k is mapped to node i in G_2. The broadcast time of S_k is the maximum *cstretch* from among its k children. But since the j^{th} entry of p_k is the probability that the j^{th} largest value in d_i will be returned, the inner product $\langle sort(d_i), p_k \rangle$ gives the expected value of the maximum of the k *cstretch*es. Multiplication on the left by v_n^T simply averages these n values over all n rows.

Note that this is consistent with Theorem 1.3.1 for the special case where $k = 2$. Theorem 1.4.1 allows us to compute the broadcast time of a k-star for a variety of graph families, and we later use these as building blocks for bounds on general trees. Moreover, Theorem 1.4.1 improves on the trivial upper bound of $diam(G_2)$ for the broadcast time of a star. A better bound can be derived by considering the average eccentricity of G_2. The eccentricity ε of a vertex in a graph is defined as the maximum geodesic distance between that vertex and any other.

Corollary 1.4.1 *For any graph G_2, the broadcast time of a star of size k satisfies*

$$\mathbb{E}[cbtime_{G_2}(S_k)] \leq \frac{1}{n} \sum_{v \in V_2} \varepsilon(v). \tag{1.18}$$

Proof 1.6 Substituting p_{n-1} in place of p_{k-1} returns the average eccentricity of the vertices in G_2.

Corollary 1.4.1 provides a better bound than the diameter, but is not nearly as good as when using Theorem 1.4.1 directly. To illustrate how Theorem 1.4.1 can be used for a specific G_2, we provide an upper bound on the broadcast time of a star, when G_2 is the line lattice above.

Corollary 1.4.2 *For $G_2 = D_n^1$, the line lattice, the broadcast time of a star of size k satisfies*

$$\mathbb{E}[cbtime_{G_2}(S_k)] \leq \frac{k}{k+1} \cdot n. \tag{1.19}$$

Proof 1.7 The maximum product on the right certainly occurs at $d_1 =$