PART I

GETTING STARTED WITH ANALYZING SOCIAL MEDIA NETWORKS

This volume is organized in the form of a tree with roots, a trunk, and branches. The roots (Part I: Chapters 1 through 3) provide grounding in the history and core concepts of social media and social network analysis. The trunk (Part II: Chapters 4 through 7) focuses on the practical details of operating the free and open source NodeXL extension of the familiar Microsoft Excel spreadsheet application used for all exercises in this volume. And the branches (Part III: Chapters 8 through 15) each focus on one form of social media by describing each system, the nature of the networks that are created when people interact through it, and the kinds of analysis that can be performed to identify key people, documents, groups, and events. The results are actionable insights that can guide community managers, marketers, organizations, and members as they try to improve the quality and value of their social media initiatives.

Part I discusses the novel ways people connect through social media tools and the formal network analysis techniques that can elucidate those connections. Chapter 1 provides a high-level overview of social media initiatives and network science, explaining why the two can be so effectively combined. Chapter 2 describes the design space of social media tools, providing examples of different types of connections created by popular tools. Finally, Chapter 3 introduces the core concepts of social network analysis and visualization, which will be put to work in the remainder of the book to understand social media networks. The chapters assume no prior knowledge of these topics.

This page intentionally left blank

CHAPTER

1

Introduction to Social Media and Social Networks

OUTLINE

1.1 INTRODUCTION

Billions of people create trillions of connections through social media each day, but few of us consider how each click and key press builds relationships that, in aggregate, form a vast social network. Passionate users of social media tools such as email, blogs, microblogs, and wikis eagerly send personal or public messages, post strongly felt opinions, or contribute to community knowledge to develop partnerships, promote cultural heritage, and advance development. Devoted social networkers create and share digital media and rate or recommend resources to pool their experiences, provide help for neighbors and colleagues, and express their creativity. The results are vast, complex networks of connections that link people to other people, documents, locations, concepts, and other objects. New tools are now available to collect, analyze, visualize, and generate insights from the collections of connections formed from billions of messages, links, posts, edits, uploaded photos and videos, reviews, and recommendations. As social media have emerged as a widespread platform for human interaction, the invisible ties that link each of us to others have become more visible and machine readable. The result is a new opportunity to map social networks in detail and scale never before seen. The complex structures that emerge from webs of social relationships can now be studied with computer programs and graphical maps that leverage the science of social network analysis to capture the shape and key locations within a landscape of ties and links. These maps can guide new journeys through social landscapes that were previously uncharted.

1.2 A HISTORICAL PERSPECTIVE

Network science focuses on the study of patterns of connection in a wide range of physical and social phenomena. Network researchers have explored foundational physical systems created by chemical and genetic connections, webs of consumption of which animals eat which others, and profound distributed human social phenomena such as collective action, empathy, social cohesion, privacy, responsibility, markets, motivation, and trust. In the past few decades, network researchers have developed new data collection methods, innovative mathematical techniques, and surprising predictive theories. Just as Lord Kelvin (1824–1907) encouraged careful measurement as the method of advancing science, the new sciences of collective action, collaboration, and productive communities require new forms of measurement. Similarly, where Newton (1643–1727) and Leibniz (1646–1716) created the mathematical methods of calculus to grasp the physical world of objects in motion, social scientists are developing advanced mathematical methods for capturing social network evolution, diffusion, and decay. Like Galileo's telescope (1564–1642), Hooke's microscope (1635–1703), or Roentgen's (1845–1923) x-rays, new information analysis tools are creating visualizations of never before seen structures. Jupiter's moon, plant cells, and the skeletons of living creatures were all revealed by previous technologies. Today, new network science concepts and analysis tools are making isolated groups, influential participants, and community structures visible in ways never before possible.

Social network analysis is the application of the broader field of network science to the study of human relationships and connections. Social networks are primordial; they have a history that long predates systems like Facebook and Friendster, and even the first email message. Ever since anyone exchanged help with anyone else, social networks have existed, even if they were mostly invisible. Social networks are created from any collection of connections among a group of people and things. Social network science is itself relatively new, with roots in the early twentieth century that built on two centuries of work in the mathematics of graphs and topology. In the twenty-first century, network science has blossomed alongside a new global culture of commonplace networked communications. With widespread network connectivity, within just the past few decades, billions of people have changed their lives by creatively using social media. We use social media to bring our families and friends closer together, reach out to neighbors and colleagues, and invigorate markets for products and services. Social media are used to create connections that can bind local regions and span continents. These connections range from the trivial to the most valued, potent collaborations, relationships, and communities. Social media tools have been used successfully to create large-scale successful collaborative public projects like Wikipedia, open source software used by millions, new forms of political participation, and scientific collaboratories that accelerate research. Unheard of just a few years ago, today systems such as blogs, wikis, Twitter, and Facebook are now headline news with social and political implications that stretch around the globe. Despite the very different shapes, sizes, and goals of the institutions involved in social media, the common structure that unifies all social media spaces is a social network. All of these systems create connections that leave traces and collectively create networks.

1.3 THE RISE OF SOCIAL MEDIA AS CONSUMER APPLICATIONS

Social media are visible in the form of consumer applications such as Facebook and Twitter, but significant use of social media tools takes place behind the firewalls that surround most corporations, institutions, and organizations. Inside these enterprises employees share documents, post messages and engage in extensive discussions, document annotation, and create extensive patterns of connections with other employees and other resources. Networked communication has become an indispensable link to customers and partners and a critical internal nervous system required for every aspect of commerce. Social media tools cultivate the internal discussions that improve quality, lower costs, and enable the creation of customer and partner communities that offer new opportunities for coordination, marketing, advertising, and customer support.

As enterprises adopt tools like email, message boards, blogs, wikis, document sharing, and activity streams, they generate a number of social network data structures. These networks contain information that has significant business value by exposing participants in the business network who play critical and unique roles. Some employees act as bridges or brokers between otherwise separated segments of the company. Others have patterns of connection that indicate that they serve as sources of information for many others. Social network analysis of organizations offers a form of MRI or x-ray image of the organizational structure of the company. These images illuminate the ways the members of the organization are actually structured in contrast to the formal hierarchies of traditional "org-charts."

Technology consulting firms have recently started to highlight the value of analyzing patterns of connection within an organization. The Gartner Group reported that social network analysis would prove to be a strategic advantage for a corporation, calling it an "untapped information asset."[1] They recommend the analysis of "business intelligence on the ties, information flows and value exchanges" within a corporation. Network analysis can be focused, they argue, on three separate regions of commerce: organizational network analysis, value network analysis, and influence analysis, which map loosely to internal, vendor, and consumer populations. In each segment, network analysis is a useful method for identifying choke points and positions of leverage, locating expertise, and enhancing innovation.

[1]www.gartner.com/it/page.jsp?id=1239913

1.4 INDIVIDUAL CONTRIBUTIONS GENERATE PUBLIC WEALTH

Social media collective goods are a remarkable story of bottom-up individual initiative that leads to the creation of public value and wealth. Collections of individual social media contributions can create vast, often beneficial, yet complex social institutions. The intriguing challenge for the authors of this book and for a growing circle of social media analysts is to focus on individual behaviors while recognizing the emergent, collective properties of social media contributions. Seeing the social media forest, and not just the trees, branches, and leaves, requires tools that can assemble, organize, and present an integrated view of large volumes of records of interactions. Building a better view of the social media landscape of connection can lead to improved user interfaces and policies that increase individual contributions and their quality. It can lead to better management tools and strategies that help individuals, organizations, and governments to more effectively apply social media to their priorities.

Many utopian commentators have reported and proclaimed the benefits of social media. However, dangerous criminals, malicious vandals, promoters of racial hatred, and oppressive governments can also use social media tools to enable destructive activities. Critics of social media warn of the dangers of lost responsibility and respect for creative contributions, when vital resources are assembled from many small pieces [1]. These dangers heighten interest in understanding how social media phenomena can be studied, improved, and protected. Why do some groups of people succeed in using these tools while many others fail? Community managers and participants can learn to use social network maps of their social media spaces to cultivate their best features and limit negative outcomes. Social network measures and maps can be used to gain insights into collective activity and guide optimization of their productive capacity while limiting the destructive forces that plague most efforts at computer-mediated communications. People interested in cultivating these communities can measure and map social media activity in order to compare and contrast social media efforts to one another.

Around the world, community stakeholders, managers, leaders, and members have found that they can all benefit from learning how to apply social network analysis methods to study, track, and compare the dynamics of their communities and the influence of individual contributions. Business leaders and analysts can study enterprise social networks to improve the performance of organizations by identifying key contributors, locating gaps or disconnections across the organization, and discovering important documents and other digital objects. Marketing and service directors can use social media network analysis to guide the promotion of their products and services, track compliments and complaints, and respond to priority customer requests. Community managers can apply these techniques to public-facing systems that gather people around a common interest and ensure that socially productive relationships are established. Social media tools have become central to national priorities requiring government agency leaders to become skillful in building and managing their communities and connections. Governments at all levels must learn to optimize and sustain social media tools for public health information dissemination, disaster response, energy conservation, environmental protection, community safety, and more.

In this book we explore the social structure and organization of these social media phenomena through the application of the methods and concepts of social network analysis. Network analysis is a relatively recent scientific method for describing and analyzing a web of links among entities, including people. Network analysis provides powerful ways to summarize networks and identify key people or other objects that occupy strategic locations and positions within the matrix of links. Visualizations that map these structures complement numerical measures to provide intuitions and insights into the shape, size, density, sub-regions, and key locations within a connected population. Over decades, dedicated scientists, technologists, and entrepreneurs have dramatically improved the tools, analytic methods, visualization approaches, and sources of data for social network analysis. Increasingly potent software applications are available to study these phenomena and uncover useful, actionable insights. Until recently, these tools demanded significant programming and data management skills that excluded many interested users. We focus on a social network analysis tool designed for ease of use and customized for application to social media, the free and open Network Overview Discovery and Exploration add-in for Excel 2007 (NodeXL).

1.5 WHO SHOULD READ THIS BOOK

Practitioners, researchers, and students interested in the study of social media can benefit from this book. Using this volume, business leaders, instructors, and students can shift their attention from a focus on mastering social network tools and concepts to their application in measurement, analysis, and interpretation. Readers will learn how to extract insights from networks to reveal internal business activity, external customer communities, and the competitive landscape.

Professors and instructors in a range of disciplines may find this volume useful in semester-long courses as well as shorter units related to computing, business, and social sciences. Technical classes in computer science/engineering, information science, human computer interaction, information visualization, and even social physics have been increasingly focused on the topic of “social computing.” In business and management schools, enterprise collaboration and customer communities have become hot topics. In the social sciences, sociologists, anthropologists, historians, economists, demographers, political scientists, and other students of collective intelligence, collective action, and communities of practice have a historical opportunity to do data analysis at a scale never before possible. Digital humanities scholars and instructors have increasingly used social media to understand and enable novel connections between people and cultural artifacts.

1.6 APPLYING SOCIAL MEDIA TO NATIONAL PRIORITIES

Government agencies around the world are attracted to the possibilities of improved delivery of services at reduced costs but challenged by the loss of content control, liability for libel, pornography, or terrorist use. Open access to vast stores of government data expands their value while potentially calling attention to unfavorable information that could be used by political adversaries. Government professionals are excited by the prospects of increased citizen engagement but concerned by what that engagement may mean for their control over the flow of information and their obligations to protect privacy, avoid censorship while preventing libel, and other inappropriate uses of government information technology resources. Across the planet, developed and developing countries are attracted to the potential to use social media to change their societies, from promoting energy conservation and smoking cessation to new levels of political engagement with citizens.

Early efforts by innovative citizen/residents gave encouragement that existing social media platforms could be harnessed for national priorities, such as disaster response. In the aftermath of Hurricane Katrina in New Orleans in August 2005, volunteers created web sites to coordinate assistance, offer food, provide housing, and eventually help rebuild. When 32 Virginia Tech campus members were killed by a lone shooter on April 16, 2007, 1500 people self-organized and composed a Wikipedia page to inform outsiders what was happening. Each summer’s California wildfires have stimulated innovative uses of Flickr to collect photos and Twitter for communication among firefighters and residents. But these and other uses could have been far more effective if designers had anticipated these needs and made modest design changes so that photos or tweets could be conveniently organized by groups, location, or time.

Disaster response is typically seen as a federal government responsibility, but other services such as community safety, are seen as a local government responsibility. Here too innovative individuals have created web sites and services to enable resident-to-resident assistance, such as the Nation of Neighbors project,[2] which describes its effort as follows:

> Real-Time Collaborative Neighborhood Watch and Reporting: Nation of Neighbors’ web-based community networks enable neighbors to share local crime, suspicious activity and other community concerns in real time. Our mission is to enable citizens and law enforcement to work together to fight and deter crime and build strong communities. You can participate at the individual level or as a member of a Community Group or Agency.

Another successful community safety effort has been Amber Alert,[3] named after a child who was abducted and murdered in 1996. This alerting system, now coordinated by the U.S. Department of Justice, claims to have directly assisted in almost 500 safe returns of abducted children. It may also have raised awareness enough to have prevented many other abductions.

Resident reports are being solicited for tornadoes, earthquakes, floods, or other natural disasters as well as for reporting on fraud, abuse, and waste of government funds. Positive contributions such as fixes to the Library of Congress Card Catalog or the National Library of Medicine’s PubMed service lead the way in suggesting further possibilities. The innovative Peer-to-Patent system (see Noveck’s 2009 book on *WikiGovernment* [2]) invited specialists in certain technical areas to contribute information on prior art related to patent applications. Then a group discussion ranked the 10 items for submission to the patent examiners so as to speed up and improve their work at the U.S. Patent and Trademark Office. Noveck summarizes her case this way: “Ordinary citizens have more to offer than voting or talking. They can contribute their expertise and, in so doing, realize the opportunity now to be powerful… Collaborative governance is an idea whose time has come.” [2, p. 190].

Volunteers to museums, parks, hospitals, or schools could also improve public services at national and local sites. Government-run web sites such as http://serve.gov and http://nationalservice.gov now facilitate these service efforts, while data at www.volunteeringinamerica.gov indicate that more than 38 percent of residents

[2] www.nationofneighbors.com
[3] www.amberalert.gov, America’s Missing: Broadcast Emergency Response

volunteer in the top three states of Utah, Nebraska, and Minnesota, while the bottom ranking states of Florida, Nevada, and New York have volunteer rates of under 20 percent. Could increased visibility of volunteer efforts and greater public recognition increase participation in volunteer programs?

Web services can also be used to distribute tasks among volunteers who prefer to stay at home and contribute their technology and skills, as in the Search for Extraterrestrial Intelligence:

> SETI@home is a scientific experiment that uses Internet-connected computers in the Search for Extraterrestrial Intelligence (SETI). You can participate by running a free program that downloads and analyzes radio telescope data.
>
> *http://setiathome.berkeley.edu/sah_about.php*

The early success of the National Aeronautics and Space Administration (NASA) in getting "clickworkers" to help with the image analysis of Martian craters led to the recent "Be a Martian" effort to help map features and count craters.[4] Similarly the Smithsonian Institution provides a home for the ambitious Encyclopedia of Life[5] to achieve Edward O. Wilson's goal of a web page for each of the 1.8 million species on earth. The payoffs in scientific knowledge to support biodiversity and environmental preservation are potentially large, but the challenges of getting professional scientists, citizen scientists, and nature enthusiasts to work together are substantial. Designers and community managers have yet to find the motivational structures and recognition strategies to gain broader participation.

Another powerful use of social networks is to support rapid dissemination of public information (e.g., on flu vaccinations, weather alerts, or community safety threats). Many people are more likely to trust and act on cell phone calls or email messages from friends and family than from pronouncements by public officials on television programs or newspaper reports. Public officials could also disseminate less time-sensitive information on energy conservation strategies, environmental protection initiatives, or health alerts about obesity reduction or smoking cessation. Citizen-generated YouTube videos often have more impact than carefully scripted appeals by professionals at press conferences. Counts of retweets and video downloads can show how effective various strategies are in reaching different demographic segments.

Leaders of many nongovernmental community groups have come to appreciate the growing power of social media and ubiquitous cell phones with their increasingly rich services. Communities can become energized by modern technology-mediated versions of parent-teacher associations, neighborhood watches, and disaster planning teams. Even smaller groups such as book clubs, high-school orchestras, or local birdwatchers benefit from use of communications tools such as Twitter feeds, Facebook pages, or Google groups.

1.7 WORLDWIDE EFFORTS

Although this book emphasizes examples that we are more familiar with in the United States, there are worldwide efforts to apply social media strategies and encourage further research and development. The European-based Institute for Prospective Technological Studies produced two thoughtful reports on The Impact of Social Computing on the EU Information Society and Economy[6] and Public Services 2.0: The Impact of Social Computing on Public Services.[7]

The first report suggests that "social computing could play an increasingly important role in re-engaging citizens in political debate, in securing social cohesion and harmony, and it could provide a platform for dialogue on the grand challenges of the EU [European Union] and the rest of the world." The second report encourages "policy makers to seize the opportunities of social computing but also to mitigate any undesirable effects" and laments the "limited provision of citizen-centered public services by governments." The report further warns "that the empowerment and transparency characteristics of social computing initiatives seem to disrupt existing power balances."

There have been other efforts as well, such as that of the European Society of Socially Embedded Technologies (www.eusset.eu), which produced a manifesto seeking to shift the attention of researchers and policy makers from technology to social issues:

> Many people have been involved in the attempt to shift the focus of computing—and informatics more generally—away from a purely technical approach concerned with hardware and software only, to one that considers the human activities of design and use of information systems as being of central concern … how can we design in support of cultural diversity, social inclusion, as well as social and environmental sustainability; how can we design for new forms of engagement and participation?
>
> *http://www.eusset.eu/index.php?id=5*

We are also aware of strong efforts by the Chinese Academy of Sciences, which launched a social computing laboratory focusing on (1) testbed and simulation

[4]http://beamartian.jpl.nasa.gov
[5]www.eol.org
[6]http://ftp.jrc.es/EURdoc/JRC54327.pdf
[7]http://ftp.jrc.es/EURdoc/JRC54203.pdf

tool development; (2) foundations of social computing; and (3) social, cyber, and physical systems.

The international Web Science Trust (www.webscience.org) takes as a central tenet that the web is socially embedded, and among the key research questions is "How can we develop interdisciplinary epistemologies that will enable us to understand the web as a complex socio-technical phenomenon?" Its position papers [3,4] affirm the need to develop a strong scientific foundation that integrates sociotechnical systems thinking. Hendler et al. [5] point out how difficult this challenge is:

> The social model enabled by humans interacting in ways allowed by that technology is more difficult to explain … the success or failure of the sites hinges on the rules, policies, and user communities they support. Given that the success or failure of Web technologies often seems to rely on these social features, the ability to engineer successful applications requires a better understanding of the features and functions of the social aspects of the systems.

The Web Science effort is in harmony with the long-standing efforts to create a science of sociotechnical systems (STSs), which began at London's Tavistock Institute as a way of generating fresh thinking about workplace experiences. It has broadened to address the experience of individuals, teams, and communities as they apply technology in organizations and beyond. The integrative view of STS thinking addresses the whole task and clarifies responsibility for decisions in ways that increase motivation and enrich the experience for all participants. While some see the STS more narrowly as a management and work design topic, the term has come to reflect a balanced view of how technology design can be successfully attuned to human needs and values.

Many social scientists have taken an interest in sociotechnical systems design. Applied anthropologists use ethnographic methods for mobile phone companies to better understand how these devices are changing family, business, and social structures. Similarly, some sociologists and political scientists are studying how information and communications technologies are changing societies around the world.

The authors of this book see many opportunities for diverse academics and professionals to contribute to our understanding of how social media are already changing our world. We also see an active role for them in the design of future technologies, as well as future social, economic, and political systems. Technology promoters succeed more often when they address usability and sociability, are alert to human needs and values, and are sensitive to balancing policies and norms.

1.8 PRACTITIONER'S SUMMARY

Existing social institutions, educational curricula, business plans, and government policies are shifting as a result of social media tools and their application. Forward-looking universities are adding courses on the study of social computing, social informatics, new media, digital society, and other terms. Journal editors, conference organizers, and national science funding agencies are working to take advantage of the opportunities of using these new tools and techniques.

Individuals, organizations, and government agencies are devoting resources to using social media for their benefit while avoiding the dangers. Understanding how these social media networks thrive, change, or fail is a substantial challenge to researchers and professionals. Researchers in social network analysis have provided a set of concepts and metrics to systematically study these dynamic processes. Innovators in information visualization have also contributed to helping users to discover patterns, trends, clusters, gaps, and outliers, even in complex social networks. Each day solutions are being found that bring competitive advantages to business product developers, opportunities for government agency staffers, and new possibilities for nongovernmental social entrepreneurs.

1.9 RESEARCHER'S AGENDA

Now is an exciting time for those involved in the emerging discipline of social media network analysis. Researchers are designing novel collaborative technologies and social strategies that enable new forms of working and playing. They are also analyzing existing communities to find out what strategies and design decisions lead to success. Although social media success stories abound, there are countless examples of failed attempts to effectively apply social media to achieve desired goals. Social network analysis offers a systematic method to evaluate social media efforts, replacing anecdotes with scientifically based evidence. Unfortunately, many observers see no urgency in changing business strategies, marketing plans, research directions, curricula, or government programs. We hope this book will change their minds by showing the compelling business opportunities, attractive research challenges, strong educational needs, and important national priorities that social media can address and network analysis can elucidate.

Just as the physicists of the 1940s were challenged and troubled by the awesome forces they unleashed, researchers studying these social phenomena may yet create technologies that release human chain reactions, harnessing vast amounts of human energy to overcome the social problems that challenge our world. In the past 400 years, scientists have focused on fundamental physical phenomena, such as gravity, magnetism, nuclear forces, and genetic information. Their work has produced profound changes in human life as we gain insights and control over core physical forces. Cellular communication networks, material sciences, and nuclear power are leading examples of the accomplishments of this vast intellectual endeavor. Similarly, biologists have revealed the core processes of all life, exposing the structure of DNA and opening the door to powerful techniques and practices that are only just unfolding. Because powerful technologies are eagerly sought by those who might put them to destructive purposes, urgent efforts are needed to ensure constructive outcomes. We believe that open discussion, broad participation, and respect for individual rights can help lead the way to more beneficial results.

References

[1] J. Lanier, You Are Not a Gadget: A Manifesto, Knopf Publishers, New York, 2010.

[2] B. Noveck, Wiki Government: How Technology Can Make Government Better, Democracy Stronger, and Citizens More Powerful, Brookings Institution Press, Washington, DC, 2009.

[3] T. Berners-Lee, W.T. Hall, J.W. Hendler, N. Shadbolt, D. Weitzner, Creating a science of the web, Science 313 (5788) (2006) 769–771.

[4] N. Shadbolt, T. Berners-Lee, Web science emerges, Scientific American (October 2008) 32–37.

[5] J.W. Hendler, N. Shadbolt, W.T. Hall, T. Berners-Lee, D. Weitzner, Web Science: An Interdisciplinary Approach to Understanding the World Wide Web, Communications of the ACM 51 (7) (2008).

Additional Resources

Benkler, Y. (2005). *The wealth of networks: How social production transforms markets and freedom*. New Haven, CT: Yale University Press.

Castells, M. (1996). *The rise of the network society*. Malden, MA: Blackwell.

Christakis, N., & Fowler, J. (2009). *Connected: The surprising power of our social networks and how they shape our lives*. New York: Little, Brown.

Cross, R.L., Parker, A., & Cross, R. (2004). *The hidden power of social networks: Understanding how work really gets done in organizations*. Boston, MA: Harvard Business Press.

Heath, C., & Luff, P. (2000). *Technology in action*. Cambridge, UK: Cambridge University.

Kleinberg, J. (2008). The convergence of social and technological networks. *Communications of the ACM*, *51*(11), 66–72.

Preece, J., & Shneiderman, B. (2009). The reader-to-leader framework: Motivating technology-mediated social participation. *AIS Transactions on Human-Computer Interaction*, *1*(1), 13–32. Available at: http://aisel.aisnet.org/thci/vol1/iss1/5.

Putnam, R.D. (2000). *Bowling alone: Collapse and revival of the American Community*. New York, NY: Simon and Schuster.

Shneiderman, B., & Plaisant, C. (2009). *Designing the User Interface: Strategies for effective Human-Computer Interaction* (5th ed.). Boston, MA: Addison-Wesley.

Surowiecki, J. (2004). *The wisdom of crowds*. New York, NY: Anchor Books.

Vieweg, S., Palen, L., Liu, S.B., Hughes, A.L., & Sutton, J. (2008). Collective intelligence in disaster: Examination of the phenomenon in the aftermath of the 2007 Virginia Tech shootings (pp. 44–54). Proc. 5th International ISCRAM Conference, May 5–7, Washington, DC.

Whitworth, B., & de Moor, A. (Eds.), (2009). *Handbook of research on socio-technical design and social networking*. Hershey, PA: IGI Global.

This page intentionally left blank

CHAPTER

2

Social Media

New Technologies of Collaboration

OUTLINE

2.1 INTRODUCTION

One of the marvels of our time is the unprecedented development and use of technologies that support social interaction. Social mediating technologies have engendered radically new ways of working, playing, and creating meaning, leaving an indelible mark on nearly every domain imaginable. Billions of people now weave a complex collection of email, Twitter, mobile short text messages, shared photos, podcasts, audio and video streams, blogs, wikis, discussion groups, virtual reality game environments, and social networking sites like Facebook and MySpace to connect them to the world and the people they care about. Increasingly, people access these tools using mobile devices that can tie content to locations in real time. Behind organizational firewalls, a host of "Enterprise 2.0" social media tools echo the "Web 2.0" tools in the public Internet. The novel ways that people have adopted and adapted these technologies to their particular needs is a testament to human ingenuity and sociability.

Despite the growing ubiquity of social technologies, their potential has hardly been tapped. Effectively using and improving social technologies is far from trivial. Within them a complex interplay between social practices and technological infrastructures takes place. Architects will tell you that the physical design of a building or city can dramatically influence the ways in which people interact with one another. Teaching a course in a room with seats arranged in a circle versus seats arranged in rows facing the front invites a different form of participation from students. Although the physical layout does not wholly determine the forms of interaction, it does make certain interactions easier and others more challenging.

Similarly, the sociotechnical infrastructure, or platform, that underlies online activity influences social interaction. Admitting so is not technological determinism. Rather it is a solid materialism that recognizes that technologies change the fabric of the material world, which in turn changes the social world. For example, microblogging sites like Twitter enable short exchanges ideal for efficiently pointing out resources or knowing what conferences people are attending, while discouraging in-depth discussion and analysis on the platform itself. In contrast, traditional blogs without length limitations and with their support for sharing multimedia content and comments are better suited for more in-depth presentations and conversations. Other media including books, newspapers, wikis, email, social networking sites, and so forth each have a set of properties that create a unique terrain of interaction. Learning to effectively meet your objectives using social media requires an understanding of that terrain and the social practices that have grown up around its use.

One of the most exciting aspects of online social media tools is that they produce an enormous amount of social data that can be used to better understand the people, organizations, and communities that inhabit them. More specifically, they create relational data: information about who knows or is friends with whom, who talks to whom, who hangs out in the same places, and who enjoys the same things. These relational data provide a wealth of new opportunities to understand and improve the social worlds we inhabit, as discussed throughout this book.

The purpose of this chapter is to introduce some of the important social media systems and provide a language and framework to talk about their key properties. It is also intended to informally introduce the relationship between social media systems and the networks they implicitly and explicitly create. This chapter begins with a definition of social media, followed by a framework for characterizing types of social media, and then takes you on a whirlwind tour of several important social media technologies that have emerged in the recent past.

2.2 SOCIAL MEDIA DEFINED

Social media refers to a set of online tools that supports social interaction between users. The term is often used to contrast with more traditional media such as television and books that deliver content to mass populations but do not facilitate the creation or sharing of content by users. Social media is about "transforming monologue (one-to-many) into dialog (many-to-many)."[1] In practice, it is a catchall phrase intended to describe the many novel online sociotechnical systems that have emerged in recent years, including services like email, discussion forums, blogs, microblogs, texting, chat, social networking sites, wikis, photo and video sharing sites, review sites, and multiplayer gaming communities. Related terms that describe many of these systems include Web 2.0, the read/write web, social computing, social software, collective action tools, sociotechnical systems, computer-mediated communication, groupware, computer supported cooperative work (CSCW), virtual or online communities, user-generated content, and consumer-generated media.

Pioneers of the information age such as Vannevar Bush who envisioned a hypertext-like device called the "memex" [1] and Douglas Engelbart who saw a future of graphical interfaces (i.e., windows), computer mice, and multiple-authored digital content [2] decades before it was realized, were interested in augmenting human intellect. In other words, they wanted to develop systems that "increase the capacity of man to approach a complex problem situation, to gain comprehension to suit his particular needs, and to derive solutions to problems" [2]. These goals have slowly been realized through remarkable developments in hypertext, human-computer interaction, the World Wide Web, and mobile technologies [3].

As the world has become increasingly connected, the focus has shifted to augmenting social experience and collective intelligence. Social media tools allow users to collaboratively create, find, share, evaluate, and make sense of the mass of information available online. They also allow users to connect, inform, inspire, and track other people. The new blend of social action and technological infrastructure allows entirely new ways of collaborating. Users can receive personalized recommendations based on the prior purchasing habits of thousands of other "similar" people, identify high-quality news stories based on real-time voting by the crowd, collaboratively author the world's largest and most-read encyclopedia, and instantly notify hundreds of followers about an online video presentation they found insightful.

2.3 SOCIAL MEDIA DESIGN FRAMEWORK

Social media systems come in a variety of forms and support numerous genres of interaction. Although they all connect individuals, they do so in dramatically different ways depending in part on the technical design choices that determine questions like these: Who can see what? Who can reply to whom? How long is content visible? What can link to what? Who can link to whom? As discussed in the introduction, these design choices can influence the social interactions that they enable and mediate. In addition, social practices, personalities, and history heavily influence how social media systems are used. If designers have learned anything

[1] www.webpronews.com/blogtalk/2007/06/29/the-definition-of-social-media

from successful social media systems like email and discussion forums, it is that they can be adapted to meet a surprisingly wide array of individual and community needs. Despite the adaptability of many social media systems, it is important to distinguish among systems as different as email, wikis, and massively multiplayer video games while recognizing their similarities.

One way to make sense of the bewildering proliferation of systems and services is to consider a set of key dimensions along which many social media services can be located. This approach provides a language and framework for comparing social media tools. This section considers six key dimensions:

- Size of producer and consumer population
- Pace of interaction
- Genre of basic elements
- Control of basic elements
- Types of connections
- Retention of content

These are not the only dimensions of possible interest, but they capture many of the important differences between social media tools. They also help lay the groundwork for the remainder of the book, which will use more formal methods to analyze the networks that are implicitly or explicitly created by the various social media platforms.

2.3.1 Size of Producer and Consumer Population

In most social media systems, producers and consumers are drawn from the same set of users. Users are producers one moment and consumers the next. However, differentiating between those who produce and consume content can be useful in comparing social media systems, even if the set of producers and consumers are not mutually exclusive.

Social media services vary in terms of their intended number of producers and consumers. An email is usually authored by just one person, whereas a wiki document is likely to be authored by several or even hundreds of people. An individually authored email might be sent to just one other person or be broadcasted to thousands. More generally, social media tools support different scales of production and consumption of digital objects. Table 2.1 provides some examples of social media systems, as well as some traditional media systems, and where various actions related to them fall within the producer and consumer size dimensions. You may notice that some systems show up in different places based on their usage scenario or the features that are being discussed.

Many social media tools help individuals or small groups interact. Instant messaging (IM), video chat, and

TABLE 2.1 Examples of Social Media and Pre-digital Media Systems Organized by the Size of Producer and Consumer Populations

Size of Consumer Population	Size of Producer Population		
	Small	Medium	Large
Small	Instant messaging Personal messaging (e.g., within Facebook) Video chat Phone call Face-to-face office meeting	Committee report to a decision maker Online survey Social networking friend feed Twitter homepage showing tweets of people you follow	Professional services reports for decision makers Personalized suggestions based on recommender systems
Medium	"Social" or family blog Profile page on community site or social network Departmental email list Tweet sent to followers Wall post on Facebook	Group blog on niche topic Internet relay chat room Internal department wiki Facebook group Niche YouTube channels Local markets (e.g., Craigslist)	Professional report for specialty group NASA clickworkers[1] Idea-generation sites (e.g., IdeaConnection[2])
Large	Popular blog, podcast, or webcast Message to large forum or email list Popular Twitter user's tweet Popular YouTube video Company web site Novel	News rating site (e.g., Digg) Wikipedia page Television program Popular discussion forum Online user-generated databases (e.g., IMDB) or marketplace (e.g., Threadless)	Large online marketplaces (e.g., eBay) Wikipedia encyclopedia YouTube video sharing Flickr photo sharing Popular massively multiplayer game

[1] *www.scienceofcollaboratories.org/resources/collab.php?317*
[2] *www.ideaconnection.com/crowdsourcing/procter-gamble-00007.html*

personal messaging within general-purpose social networking sites provide intimate communication channels comparable to phone calls and face-to-face office meetings. Social media can help individuals reach out to medium-sized groups of friends or acquaintances by broadcasting a personal message (e.g., a tweet sent to a user's followers on Twitter; a post sent to a departmental email list) or allowing others to overhear a comment (a post to someone's Facebook wall). They can also allow individuals to reach large groups through popular blog posts, podcasts, videos posted on sites like YouTube, or updates on Twitter by companies or celebrities with numerous followers.

Other social media tools help medium-sized groups reach out. Friend feeds on social networking sites and Twitter homepages allow a group (i.e., your friends or those you follow) to implicitly "create" a personalized stream of information just for you. Other tools like online surveys help aggregate information from many people for a small group of people analyzing the data. Group blogs or collections of related blogs about niche topics within the blogosphere allow a medium-sized number of bloggers and commenters to interact with one another. A host of tools facilitate interaction between medium-sized groups whether they are part of a Facebook group, YouTube channel, Internet Relay Chat (IRC) room, or combinations of tools as in a Ning social networking site on a niche topic. In some cases, a department or workgroup wiki can allow members or co-workers to co-author materials that are of interest to their group. Finally, some tools enable medium-sized producer groups to reach large consumer groups in a way similar to TV programs that take considerable effort to produce, but often reach the masses. Some of these include online databases such as the Internet Movie Database (IMDB), where user-generated movie content is shared with the world; news rating sites like Digg, where dozens to hundreds of people recommend a given news article that is consumed by much larger populations of read-only users; discussion forums, where posts by a few dozen active members may be seen by thousands of readers; and Wikipedia pages that are edited by dozens of people and read by thousands.

Some of the most interesting social media tools are those that help harness the power of the masses. For example, some recommender systems (e.g., Last.fm; MovieLens) provide personalized suggestions of books, movies, or songs by comparing your ratings with ratings' of other users. Other large groups help generate ideas that are used by medium-sized groups such as small businesses, corporate departments, or government agencies. For example, Procter & Gamble's IdeaConnection allows anyone to contribute ideas to solve their problems, whereas NASA's ClickWorkers help identify craters on Mars. Many of the most well-known social media sites allow large producer groups to interact with large consumer groups. Although there are many more Wikipedia readers than contributors, both groups are enormous. Online marketplaces like eBay allow the masses to sell and purchase goods. Meanwhile, social sharing sites like YouTube and Flickr make it easy for large numbers of producers and even larger numbers of consumers to interact. While these sites often facilitate smaller group interaction, they also aggregate those interactions so that people can search and navigate through large corpora of user-generated content. Massively multiplayer games rely on having large numbers to produce content and social experiences that make use of the entertaining environment.

2.3.2 Pace of Interaction

The pace at which interaction occurs is another important dimension along which social media systems can be organized. Traditionally, a distinction has been made between asynchronous and synchronous communication. Asynchronous systems like email, discussion forums, and voicemail presume a staccato pattern of interaction spread out over hours or days or weeks. Though less immediate, these systems have the advantage of allowing each participant to schedule their participation without much coordination with other people who may be in a wide range of time zones. They also encourage more careful contributions. In contrast, synchronous systems, like chat, instant messaging, videoconferencing, and graphical worlds, require that partners interact at the same time, as in face-to-face interactions and telephone calls. Although they require temporal coordination, they create a richer environment for interaction as participants quickly react and adjust to one another's signals in near real time. The pace of interaction has implications for the kinds of groups that form using each kind of tool. Global collaborations are often easier using asynchronous tools that don't require some people to change their sleeping habits. But some interactions need more rapid turn-taking to accomplish their goals.

More recently, the distinction has become blurred. For example, Twitter users often reply within minutes to another's tweet, but it is completely acceptable to reply a day later as well. Replying to a wall post or status update on a social networking site is similar in this regard. Tools like Google Buzz and Google Chat are now integrated with the widely used Gmail web email system, again blurring the distinction between synchronous and asynchronous modes of communication. However, users' varied expectations about the pace of interaction within these tools remains important for understanding social media environments.

2.3.3 Genre of Basic Elements

Digital objects, the basic elements of social media systems, vary in size and type. Twitter posts (i.e., tweets) are limited to 140 characters, whereas email messages are typically a few lines to a few paragraphs in length. This difference in size produces dramatically different patterns of interaction. Although size limits of instant messaging are not typically enforced, design choices such as the size of the text box and messaging window promote brevity. Meanwhile, MediaWiki (the wiki platform used by Wikipedia) supports six levels of headers and automatically generates a table of contents, making it relatively easy to create large pages.

The various types of digital objects supported by social media tools are another way to understand the similarities and differences among them. Social media systems have often evolved around a distinct type of digital object: videos at YouTube, photos at Flickr, bookmarks (i.e., web site URLs) at Delicious, books at Amazon, music or podcasts at iTunes, TV shows at Hulu, people at Facebook, tweets at Twitter, messages at discussion forums or email lists, pages at Wikipedia, products at eBay, presentations at SlideShare, 3D objects in Second Life, and career professionals at LinkedIn. Each of these provides different levels and mechanisms of engagement. For example, virtual worlds more closely model embodied physical interactions, where avatars can convey meaning through proximity and orientation [4]. They also introduce many of the burdens of face-to-face interaction, demanding attention to successfully puppet the avatar in interactions with other partners [5]. Although these differences may relate to the type of media (e.g., video, audio, text, 3D model), there are further distinctions within each type. Wikis support structured text elements like tables and bullets, whereas email does not. Some virtual worlds like Webkinz use cartoon characters, while multiplayer games like World of Warcraft include realistic-looking creatures. Of course some social media systems like Facebook include many basic elements: profile pages, wall posts, personal messages, applications, instant messages, notes, groups, photos, tags, status updates, and so on. Wikipedia has user pages, talk pages, articles, edits, categories, and so forth. Even in these systems, identifying the basic elements of the system is important because they are the building blocks of the interactions. They are also the building blocks of networks when they are connected together or exchanged, as you will learn throughout this book.

2.3.4 Control of Basic Elements

Social media systems provide different levels of control over their basic elements. They can restrict who can create, edit, read, invite, respond to, subscribe to, and share content of various types. Some systems differentiate between anonymous users, registered users, and those with special privileges such as administrators. For example, some discussion forums require that users log in before they post, but they allow anyone to read the messages created by the community. This helps reduce spam by creating a higher barrier to entry, while still allowing anyone access to the content. It also allows users to exclude participants they define as social deviants. In other discussion communities of a more sensitive nature (e.g., patient support groups), access to content can be limited along with contribution until a person is registered, a process that may require some type of approval process by current administrators. Other systems like eBay require users to provide validated credit card information before they can sell items. The more open a community, the more potential there is for deviant behavior as evidenced by the frequent spam sent to wikis and Usenet. However, closing a community off too much may reduce the number of contributors, whereas openness may attract high-quality contributions that include combating the effects of spam and abuse. This is evidenced in the many high-value pages in Wikipedia, where poor edits left by non-registered users are quickly reverted by other registered and non-registered users.

Deciding the right types of barriers to entry can be an important part of online community building and one that deserves careful consideration [6]. For example, in studying real-world communities, Ostrom [7] found that successful communities had clearly defined boundaries, largely to overcome problems associated with outsiders taking advantage of internally produced or maintained resources. Boundaries are also important in that they encourage frequent, ongoing interaction among group members. This is critical because repeated interaction is perhaps the single most important factor in encouraging cooperation [8]. If individuals are not likely to interact in the future, there is a huge temptation to behave selfishly and free ride. On the other hand, knowing that one will be interacting with others on a continual basis can lead to the creation of reputations and serve as a powerful deterrent to short-run, selfish behavior. Boundaries can have an impact on the kinds of interactions people are willing to engage in because of the ways they shape the expected audience. Some media, like telephones and post offices, encourage the expectation that only a specific group of selected others will be the audience to your message. Other media, like public notices or messages on web discussion boards, are likely to be seen by any number of unknown people. Further, some media prevent the identity of message creators to be known with certainty, if at all (e.g., pay phones or anonymous letters).

Control structures can heavily influence governance structures and the distributed or centralized nature of the environment. Although email lists, message boards, and Usenet newsgroups are all examples of asynchronous threaded conversation (see Section 2.4.1), their control structures are different. In centralized systems like email lists and many web boards, all communication flows through a single point, which is controllable by a single person (administrator) or small group who can wield dictatorial control over resources (messages) and access (who can subscribe). As a result, mail list owners can remove people for inappropriate conduct or prevent spam or other inappropriate messages from being sent out or stored in the archive. They often serve as "benevolent dictators." In contrast, distributed systems like Usenet are composed of hundreds of thousands of interlinked systems all interacting with a set of neighbors. The lack of a central point of control makes it impossible for a member to exclude others or remove content.

The granularity of control is another important factor. Users of a wiki can edit individual characters of a shared document, whereas other systems limit users to authoring entire messages that cannot then be edited. Twitter users can follow another user but then must receive all tweets from that user. In contrast, Delicious users can follow another user or can follow a user/tag combination so they only receive messages from the user that use a tag of interest. In many systems, users can only edit their own content, whereas in other systems, such as wikis and Google Documents, users can edit others' messages, documents, or objects. The level of granularity may differ for different user groups within the same system. For example, discussion boards on many web sites create a preset number of containers for interaction or leave control over the creation of new spaces in the hands of a small number of administrators. In contrast, email lists allow anyone to start new threads. Of course, systems in which users can create spaces with little restraint often contain many more of such spaces.

The pace of interaction can be crossed with the granularity of user control to characterize systems (see Table 2.2).

2.3.5 Types of Connections

There are many ways that the basic elements of social media systems can be connected. It is important to understand these connections or ties in order to construct and understand networks from each kind of social media system. We will describe the theory and language of networks in the next chapter. This section describes the many types of connections that exist in social media systems and explores the ways those collections of connections create a larger social system that we can now analyze with the math, tools, and insights of social network analysis.

The basic elements of many social media systems can be connected to one another explicitly or implicitly. Users intentionally and knowingly create explicit connections, whereas implicit connections are inferred from online behaviors. Perhaps the most common type of explicit social media connection is friending on social networking sites, where both people must approve the connection before it is realized. Other examples of explicit connections are following another user on Twitter, hyperlinking a wiki page to another page, tagging two photos or videos with the same tag, and adding someone to an IM buddy list.

Implicit connections can be inferred when a person sends another person an email message, "favorites" content (and by extension its author), replies to a discussion post, or "pokes," "waves," or "throws sheep" at another user as some sites allow. Although these actions were intentional, they were not performed with the explicit intention of creating a connection with the person. Other more subtle implicit connections can be identified, such as connecting people who "hang out" in the same discussion forums or Facebook groups or who edit the same wiki pages. These individuals may not know one another, but they are connected by their shared interests

TABLE 2.2 Examples of Social Media Categorized by the Pace of Interaction and the Granularity of Control over Content

Pace of Interaction	Granularity of Control		
	Fine: Users can directly control smallest units of content (characters, pixels, bytes)	Medium: Users control medium-sized blocks of content (objects, attributes, tracks, players) that they can only indirectly alter or that can be altered by other users	Coarse: Users control large block of content (documents, messages, blog posts, photos), rarely edited or modified by others
Synchronous	Real-time shared canvas	Virtual worlds, multiplayer games, real-time networked musical jamming	Chat, instant messaging, texting, Twitter
Asynchronous	Shared documents (e.g., Google Docs), source code, Wikipedia	Contribution to collected works like an album, anthology, report section, discussion group, or photosets	Email; blog posts and comments; sharing of links, photos, videos, and documents; turn-based games

and activities. Other connections can be inferred from data that are often not public but are available to the hosts or owners of social media systems such as reading patterns of discussion forums, music downloads, patterns of telephone calls, and location information. In innumerable ways, users now leave behind traces that form an intricate web connecting those users with the people, locations, and digital objects around them.

Another important distinction is between directed and undirected connections. If you and another person become friends on Facebook, the connection is a mutual one. In other words, it is undirected. Likewise, if you both are tagged as an "expert," then you are connected by an association that is mutual and thus undirected. In contrast, some systems like Twitter allow people to follow other users without first gaining those users' approval. This creates a different type of tie, where the directionality of the tie is important (i.e., who is following whom). Directed ties are also created when a person invites another person, favorites content, and creates a hyperlink from one page pointing to another page. In all of these cases, connections flow from one person or object to another and may not necessarily be reciprocated. Therefore, the relationships that are reciprocated are different and perhaps stronger than those that are not.

Finally, connections mean different things and can have different weights and values. For example, two people on Facebook can either be friends or not be friends; it is a binary connection that is either on or off. In contrast, two Facebook friends may send each other personal messages. The strength of their messaging connection could be measured based on the number of messages or the number of different days they each sent one another messages. This is an example of a weighted connection that varies in intensity. These weights often contain important information about the strength of a tie. For example, if Marc sent 10 messages to Ben last week and only 1 to Derek, it is probably safe to say that last week Marc was more strongly connected to Ben than to Derek (at least via that messaging medium).

The examples shared so far primarily connect people to each other, objects to each other, or people to objects. Recently, location has become an expanding part of social media services, allowing connections to be created between people, objects, and places. Smart mobile devices are opening a new era of social media that integrates information about location and activity in novel and powerful ways. New kinds of ties are being formed by just being in the same place as someone else, even at different times. Just having a laptop nearby someone else's can create new kinds of ties called "hyper-ties" that are like hyperlinks but applied to people.

2.3.6 Retention of Content

Social media systems also vary in how long content is retained. On one end of the spectrum are systems like wikis that typically create a permanent history of all actions that occurred in the system. Not only is each action recorded and stored, it is made available on article history pages and user contribution pages. At the other end of the spectrum, some instant messaging or voice-over Internet Protocol (IP) systems do not centrally record the interactions at all, allowing for fleeting exchanges more reminiscent of most face-to-face conversations. End points to these conversations can, of course, record them but extra effort must be taken. Many social media systems fall somewhere in the middle. For example, as of this writing, searches of the Twitter network only pull from the most recent 1.5 weeks. The horizon of the past in these systems is in flux as data volumes grow along with information processing capacity. The desire to add social media data to our long-term cultural memory has prompted interesting partnerships, as evidenced by the recent agreement between Twitter and the Library of Congress.

Some types of social media systems vary in their retention policies depending on the specific product or user settings. For example, some instant messaging clients do not archive conversations, whereas other clients retain them by default. Likewise, some email lists create a searchable archive of prior messages sent to the list, whereas others do not. However, it is important to realize that even if there is no centralized archive, individuals at the end points of these services may archive content and make it public at a later date. People can collect email messages, record Skype calls, log chat sessions, and collect most digital content fairly easily. We are now living in a world of easy data collection, retention, analysis, and publication suggesting prudence in using social media systems. Choose your words carefully; they may outlive you.

2.4 SOCIAL MEDIA EXAMPLES

This section provides a brief description of some popular types of social media as of this writing. Table 2.3 lists the social media systems and categorizes examples of each. It also serves as an index to this section. For a much more comprehensive list of social media tools, see Go2Web20[2] and Wikipedia. Because of the focus of this book, we highlight the types of networks that these social media tools create by discussing their basic elements and types of connections.

[2]www.go2web20.net

TABLE 2.3 Types of Social Media Listed with Example Services

Social Media Type	Examples
ASYNCHRONOUS THREADED CONVERSATION	
Email	Gmail, Hotmail, AIM Mail, Yahoo! Mail, MS Outlook
BBS, discussion forums, Usenet newsgroups, email lists	Slashdot, Google groups, Yahoo! Groups, Yahoo! Answers, Listserv
SYNCHRONOUS CONVERSATIONS	
Chat, instant messaging, texting	UNIX Talk, IRC, Yahoo! Messenger, MSN Messenger, AIM, Google Talk, ChaCha
Audio and videoconferencing	Skype, Gizmo, iChat, Window's Live
WORLD WIDE WEB	**TRADITIONAL WEB SITES, HOMEPAGES, AND DOCUMENTS**
Corporate, organizational, and government websites and documents	Ford.com, UMD.edu, Prevent.org, Serve.gov; Data.gov
Homepages	Faculty member websites, artists' portfolio websites, family history websites
COLLABORATIVE AUTHORING	
Wiki	Wikipedia, Wikia (Lostpedia), pbwiki, wetpaint
Shared documents	Google Docs, Zoho, Etherpad
BLOGS AND PODCASTS	
Blogs	LiveJournal, Blogger, WordPress
Microblogs and activity streams	Twitter, Yammer, Buzz, Activity Streams
Multimedia blogs and podcasts	Vlogs (video blogs such as Qik), photo blogs (Fotolog, FAILblog.org), moblog (mobile blogging such as moblog.net), podcasts (iTunes, NPR)
SOCIAL SHARING	
Video and TV	YouTube, Hulu, Netflix, Vimeo, Chatroulette
Photo and art	Flickr, Picasso, deviantART
Music	Last.Fm; imeem; Sonic Garden
Bookmarks, news, and books	Delicious, Digg, Reddit, StumbleUpon, Goodreads, LibraryThing, citeulike
SOCIAL NETWORKING SERVICES	
Social and dating	Facebook, MySpace, BlackPlanet, Tagged, eHarmony, Match
Professional	LinkedIn, Plaxo, XING
Niche networks	Ning (e.g., classroom 2.0), Ravelry, Grou.ps
ONLINE MARKETS AND PRODUCTION	
Financial transaction	eBay, Amazon, craigslist, Kiva
User-generated products	Instructables, Threadless, TopCoder, Sourceforge, Codeplex
Review sites	ePinions, Amazon, Angie's List, Yelp
IDEA GENERATION	
Idea generation, selection, and challenge sites	IdeaConnection, Chaordix, IdeaScale, Imaginatik
VIRTUAL WORLDS	
Virtual reality worlds	Second Life, Club Penguin, Webkinz, Habbo
Massively multiplayer games	World of Warcraft, Lord of the Rings Online, Aion
MOBILE-BASED SERVICES	
Location sharing, annotation, and games	Foursquare, Gowalla, Loopt, MapMyRun, Geocaching, Letterboxing, SCVNGR

2.4.1 Asynchronous Threaded Conversation

Asynchronous threaded conversations take on many forms such as email, email lists, Usenet newsgroups, discussion forums, and web boards. They have been the backbone of online communities since before the Internet and continue to play essential roles in a variety of settings both within organizations and on the public web. Even newer forms of social media such as social networking sites, social sharing sites, and virtual worlds often have asynchronous threaded conversations embedded within them to facilitate discussions. Although there are several flavors of threaded conversation, they share some key properties in that they are asynchronous, messages are sent with an address associated, typically, with a single author and authors can reply to messages and reply to others' replies creating conversation "threads." Authors do not typically edit one another's messages or even their own messages after (and some say before) it has been initially contributed. This basic structure has proven to be extremely versatile, supporting a wide range of individual, organizational, and community needs.

Threaded conversations, in all their forms, create implicit ties that connect senders and receivers of content in what is often called a "reply network" or "reply graph." These reply networks can be analyzed to identify important relationships, distinctive patterns of connections that reflect social roles, subgroups or clusters of people, interdepartmental connections, and many other important relationships. Next are brief descriptions of some of the most important asynchronous threaded conversation systems. See Chapters 8 and 9 for a more complete history and description.

Email

Email messaging was introduced in the late 1960s and the familiar "@" was introduced by Ray Tomlinson in 1971. Email lists were first introduced in 1972, the same year as PONG and the end of the moon landings. Today, email is almost universal with hundreds of millions of daily users exchanging tens of billions of messages a day. Because of the ubiquity of email, the implicit reply network created by its exchange is often an authentic representation of real-world connections. In Chapters 8 and 9 we discuss how to analyze three types of email collections: personal email collections (e.g., your own email archive), organizational email collections (e.g., your company's email traffic), and community collections (e.g., email lists).

Email Lists

Email lists turn email into a community experience by allowing people to send a message to a single email list address, which is then forwarded to everyone who has subscribed to the list. This allows large groups of individuals to contribute and, more often, listen in on the conversations of others. These collective email discussions are widely used in enterprise discussion lists or Internet Listservs covering a nearly unlimited array of topics. Email lists facilitate discussions on a topic of interest, technical support, neighborhood gatherings and advocacy, workgroup interactions, internal communities of practice, and even the exchange of goods (e.g., FreeCycle). They are particularly good for reaching less tech savvy users such as older adults who are familiar with email but not more advanced technologies. They differ from web boards and forums in that they are a push technology that shows up in your inbox rather than requiring you to visit a site to get the latest information. Although it is hard to measure, they are likely among the most common forms of group interaction on the Internet, with some of them including thousands or tens of thousands of members.

Usenet Newsgroups

Usenet newsgroups were first created in 1979 at the University of North Carolina. The Usenet grew over decades to contain tens of thousands of newsgroups, each devoted to a variety of topics and containing chains of messages in reply to one another in structures called threads. Like a collection of email lists, Usenet newsgroups can foster the collective construction of discussion threads through the individual creation of messages in reply to one another. From politics to cats, from cooking to computers, Usenet organized billions of messages into millions of conversations sorted into tens of thousands of newsgroups. Usenet newsgroups are distinguished from email lists largely in terms of their comparative lack of centralized control and weak boundaries. Anyone can post a message to any newsgroup without regard to membership or the desires of others receiving the messages. Like email and email lists, newsgroups contain a core social network structure called a "reply graph" created when authors are connected with those they reply to in a thread. Although Usenet newsgroups are not the cutting-edge social platform they once were, they did help inspire innovations that continue to live on in other forms. For example, the Microsoft Research Netscan project demonstrated the value of visualizing social interaction on Usenet, a promise only now being realized to its full potential [9]. Usenet newsgroups also helped inspire one of the first collaborative filtering systems called GroupLens, which made personalized recommendations of content you are likely to enjoy based on the preferences of like-minded people [10].

From Bulletin Board Systems (BBS) to Discussion Forums

In the late 1970s,[3] dialup bulletin board systems (BBSs) hosted a wide range of message boards that

[3]www.bbsdocumentary.com

allowed people to interact in various ways. Using telephone connections and acoustic modems, thousands of individual BBSs run on early desktop personal computers exchanged digital data including early forms of networked threaded message discussions. Messages traveled more slowly through the BBS network than through current social media services: messages traveled from one computer to another over days not seconds. But like later generations of discussion technologies on the World Wide Web, these systems allowed for threaded conversations and some levels of centralized control over boundaries. BBS managers selected who could access their services and what content would be retained, exchanged, and copied from other systems.

Decades after dialup BBS systems, modern World Wide Web discussion board applications transplant many aspects of email, email lists, newsgroup, and discussion threads into a web-based environment. A web forum or discussion board is now a common feature of many sites that allow their users to create messages in multiple categories (called topics or threads) and link them to one another in chains of reply. Often a conversation thread can be found hanging beneath articles in newspaper and blog sites along with video services like YouTube, expanding further the scope of implicit conversational social networks. Now that message boards and other interaction media have integrated with the interface tools made available through the World Wide Web, conversation interfaces and databases have been refined into elaborate discussion systems like Slashdot.org, Plastic.com, and Fool.com where discussions about technology, popular culture, and financial markets, respectively, are held within a system that awards points and cleverly filters content. These sites provide web interfaces that offer new ways of organizing conversational content beyond the long-standing defaults of time and alphabetization. Through the use of back-end databases that allow readership data and reviews of content to be tracked, these systems can dynamically filter content based on personal preferences to help overcome information overload.

2.4.2 Synchronous Conversation

Synchronous conversations such as text messaging, chat, instant messaging, and audio and videoconferencing differ from asynchronous conversations in that they occur in real time. Precursors to these Internet-based conversations occurred via telegraph, phone, two-way radios, and similar technologies. UNIX talk messaging, first used in the early 1970s, was among the first instantiations of text-based synchronous conversations based on computer networks. This simple system is the early precursor to chat and instant messaging, allowing two people to initially share a text stream, both users typing characters that would appear intermingled in the same space. Later innovations and refinements separated the text streams and evolved into the many forms of short messaging and texting services available today. Synchronous conversations today differ in that some such as instant messaging or videoconferencing support one-on-one, or small group conversations, whereas others such as chat enable more open conversations with anyone "tuned" to a specific "channel."

Chat

Chat was one of the most popular forms of interaction on the early Internet and accounted for up to a third of the revenue of the original commercial online providers such as America Online and CompuServe. In the late 1970s, "Citizen's band" (CB) radio in the United States was an analog technology that opened up new space for social interaction that was later mirrored by Chat services. CB radios, though restricted to a limited geographical space, organized discussions into a series of a few dozen "channels" located at frequency intervals allowing for open discussions and debates among people who might never have otherwise met. Most chat systems support a great number of "channels" dedicated to a vast array of subjects and interests. The IRC network of "Internet Relay Chat" servers remains a thriving and teeming space filled with chat from many people on numerous topics streaming nearly continuously around the world.

To the outsider, text chats can be confusing and chaotic places. But experienced users can accomplish a great deal of communication in these media. In contrast to Usenet, most text chat systems are organized around a centralized server that grants the server owner a great deal of power over access to the system and over individual channels. In commercial chat services, chat channels are frequently policed by the provider's staff or by appointed volunteers. In the largest noncommercial system, IRC, each channel has an owner who can eject people from the channel, control who enters the channel, and decide how many people can enter. Because of a lack of explicit links tying specific comments to one another, it can be hard for humans and computers to know who is talking to whom. This means that the reply networks created by chat may be error prone or probabilistic in nature.

Instant Messaging

Instant message text chat clients offer a private, often one-on-one, chat environment. Sometimes referred to as "buddy lists," these tools allow people to keep a list of their friends and contacts who also use the same or compatible tool. The messenger software indicates which if any of a person's "buddies" are available at that moment for possible conversation. Selecting a person on your buddy list opens up a private window for exchanging short lines of text in real time. More recent tools like Skype merge the buddy list and text chat with full voice telephony, blurring the distinctions between these modes of communication further. Two primary networks connect users of these services. One network is a friendship network that connects users to those in their buddy list. Another network is a conversation network that connects people based on how often they talk with one another. Organizations that provide instant messaging services can use these networks to capture latent and active internal connections.

Texting

Text chat has a mobile form called Mobile Short Messaging Service (SMS), which has taken the world by storm, becoming the most widely used form of electronic communication. Billions of texts are exchanged each day among almost 3 billion users of mobile phones and other devices. Text messaging lacks many of the features of richer forms of message exchange. Where a desktop, laptop, or "smart phone" device routinely now handles audio, video, and document attachments, text messaging is limited to 160 characters of text. The simplicity and low bandwidth requirements make it almost universally available, relatively cheap, and accessible from mobile devices. Texting has become a basic entry point for Internet services and the means of weaving together family, work, and civic relationships. SMS generates communication networks connecting phones to each other based on the number of messages exchanged.

Audio Conferencing

More than even the most engaged chat session, audio and videoconferencing are highly synchronous forms of social interaction. People interact in near real time, speaking and replying and gesturing in a nearly continuous loop of symbolic exchange similar to face-to-face encounters. Audio conferencing using standard phones has grown steadily to become a widely used service for small teams, as well as for training or marketing sessions for hundreds of users. The simplicity of use, low cost, and emphasis on human voice has turned phone conferences into widely used and productive applications. In an audio conference, neither speakers nor listeners need to worry about their dress, facial expressions, or eye contact, and therefore can engage in other tasks simultaneously without offending others. Audio conferencing is now increasingly taking place through Voice over Internet Protocol (VoIP) connections with dramatic reductions in cost and expansion of use around the planet. Tools like Skype now make an audio conference among a dozen or more people in as many time zones as simple and a text chat.

Videoconferencing

Videophones have been the promised "vision of the future" since the 1940s but have failed to reach mass market production. High costs of early system hardware and connections were multiplied by the social awkwardness that video connections impose. Unlike audio only connections, video requires that people comb their hair, straighten their tie, or put on a dress and organize at least the area the camera can see. Video requires a continuous display of engagement, more like a face-to-face encounter, without the flexibility of a phone call that allows wandering attention and multitasking. Despite these hurdles, as hardware and connection costs have plummeted to near zero, the vision of widely used videoconferencing is now starting to be realized. Early Internet-based phone systems such as CUSeeMe from Cornell University developed an enthusiastic community of users starting in 1992. Inclusion of built-in video cameras and the iChat program on Apple computers and video services in Skype and other instant messaging clients have triggered a much wider community of users, including grandparents seeing distant grandchildren and distributed musical bands and project work teams. Corporate use of videoconferencing has increased steadily and has been integrated with screen and document sharing services using dedicated communication lines to ensure higher-quality sound and images and improved reliability. Video cameras on video game consoles and smart phones have further expanded the scope of videoconference users.

Social networks based on who converses with whom via audio or videoconferencing can be created relatively easily if the data are available. However, more fine-grained analysis of the conversations themselves (e.g., who replies to whom) is challenging because the data are difficult to capture automatically. Advances in automated speech to text conversion and video facial recognition may make it possible to efficiently extract these network exchanges in the years to come.

2.4.3 World Wide Web

The largest public, machine-readable network is the World Wide Web where web content, such as web pages and images (identified by their Uniform Resource Locator or URL), is connected together by hyperlinks. The World Wide Web, WWW, or just the web, was initially conceived by Tim Berners-Lee in the 1980s, but it was not realized until the 1990s. For the first time it married the concept of hypertext, developed in the 1960s by Ted Nelson (Xanadu) and Douglas Engelbart (oN-Line System [NLS]), with the Internet [11]. The result was a highly flexible platform that allowed people to view web content hosted on servers throughout the world with the use of a web browser. Today the web is the primary platform upon which most of the social media tools are built. In this section, and in Chapter 12, we focus primarily on traditional web sites such as corporate, organizational, and government web sites, homepages, and documents (e.g., images and pdf files).

Although you may not realize it, network analysis plays a role in millions of people's everyday lives when they search the web via Google. As the amount of content on the web increased, the role of search engines became more essential for content on the web to be accessible. Early search engines looked only within the text of each web page (and its associated metadata fields) to determine its relevance. The first generation of search engines, like Alta Vista, built an index of all the words on millions of pages and matched them to search queries. Google made a breakthrough in the quality of its results by developing its PageRank algorithm, which determines how important a page on the web is based on its position in the web's network. At its core, the PageRank algorithm views a link to a page as a "vote" for that page's importance, so that pages with many incoming links score well. In addition, it considers the importance of each of the incoming links: receiving a link from a highly linked-to site counts more than receiving a link from an unknown site. This and related concepts are captured in various measures of the "centrality" of a point within a network as discussed in Chapter 5. A detailed analysis of several networks found within the World Wide Web is found in Chapter 12, which also introduces a "web crawler" tool that can generate networks of interconnected web pages.

2.4.4 Collaborative Authoring

Word processing is an innovation for those who grew up with typewriters. Although many believe creating and editing documents on a computer is a clear improvement over the use of typewriters, it often remains a solitary process. A further refinement of electronic writing is not just to improve the layout and fonts used but to change the way that groups of people collectively author and consume documents. Several social media tools facilitate the collaborative authoring of documents and repositories, enabling small groups and even communities of thousands to effectively create, maintain, and organize documents and document repositories.

Wikis

The most widely known example of collective document creation is Wikipedia, which is only one example of thousands of smaller wikis scattered throughout the web. Created by Ward Cunningham in the early 1990s, wikis are tools that allow a group of people, potentially any Internet user, to access and edit a shared collection of documents in the form of web pages. There are many "wiki engines" (i.e., platforms) including open source products like MediaWiki that are used by Wikipedia, as well as corporate products like Wetpaint, PBWiki, and SocialText. Wikis are used to create encyclopedias (e.g., Wikipedia), fan or game sites, corporate intranet content, and information repositories on topics ranging from educational resources to technical documentation to patient support information. Despite many differences among implementations, all wiki engines track each edit of each wiki page, creating a detailed page history. These edits can be reversed, creating a social version of the "undo" function of a word processor. People can follow updates to content on the RecentChanges page, get notified of changes to specific pages after subscribing to (i.e., "Watching") them, or view people's "user contribution pages" that chronicle all of their edits.

Wikis include many implicit networks. Some networks connect pages to other pages through hyperlinks or when pages are grouped together into categories. Other networks connect people to people when, for example, a person posts on another person's user page or when two people co-edit the same page. These networks can be mined to better understand content relationships or social roles as described in Chapter 15.

Shared Documents

The idea of collaboration through shared documents such as word processing and spreadsheet files is well established and goes beyond the community approach provided by wikis. Collaborative document creation continues to grow in popularity as corporations, governments, and community organizations discover that they can conveniently share and edit documents through Google Docs, Windows LiveSync, or DropBox. Users who edit the same documents form a network that may reveal patterns of cooperation, shared interest, or opportunities for new collaborations, although these networks are not yet explicit.

2.4.5 Blogs and Podcasts

Blogging is a special form of web page publication. Deceptively simple, a blog is essentially a low-cost mechanism for publication of rich digital content. Early blogs presented a series of text messages or "posts" in reverse chronological order so the most recent post was always at the top of the page. Today, a blog is a rich platform for content presentation and commentary. Blogs may contain text messages that are now complemented with pictures, video, and audio. Blogs present this content along with search, feeds such as Really Simple Syndication (RSS) for subscribing to the content, functionality for readers to comment on each blog post, tags to categorize posts, pointers to related blogs (i.e., blogrolls), and a range of applications and widgets. Popular tools like LiveJournal, Blogger, and WordPress make blogging essentially free and widely available. They are also commonly embedded within larger sites.

Blogs and by extension bloggers are now often able to build audiences that rival predigital media and challenge many more established information providers, particularly in the news arena where current information is paramount. Blogs and bloggers are seen as potentially powerful makers and breakers of brands, political candidates, and news stories. They also serve as micropublishing fronts for families sharing stories and photos of their children and niche interest groups exploring an obscure topic. A single blog may be authored by an individual or a handpicked set of authors. Others participate via comments or by linking their own blog to other blogs, creating a monolithic collection of interconnected blogs often referred to as the blogosphere. Specialized search engines such as Technorati scour the blogosphere for recent blogs, using the unique properties of blog links (called trackbacks) and number of comments to assign credibility scores to blogs. Because the connections can be automatically captured using web crawlers, researchers have analyzed the blogosphere to better understand issues like the nature of political discourse in the highly bifurcated United States arena [12], as well as other areas such as Iran [13].

Microblogs and Activity Streams

The Twitter microblogging system burst into public view only a few years ago and has now become a worldwide phenomenon. It is similar to traditional blogs in its focus on recent posts, but differs in that its posts, called "tweets," are restricted to 140 characters of text. Twitter takes advantage of the idea of blog feeds by allowing you to subscribe to, that is, "follow" any other Twitter user. A user's personalized feed shows the most recent tweets of all individuals he or she is following, creating a live stream of bite-sized information nuggets. A number of competing services exist, some of which overlay the service on top of other services. For example, Facebook and LinkedIn have status messages that serve as microblogs that are broadcast to friends. More recently, Google Buzz and the open source identi. ca provide similar services.

Microblogging sites create several interesting social network structures. The most obvious network is the one created by the "follows" and "is followed by" relationships. Unlike Facebook, these "follow" relationships are potentially directed: you can follow people who don't follow you and visa versa. This is in contrast to the undirected ties present in Facebook and LinkedIn. Other networks are created that connect users together based on the number of times they reply to others' microblog posts or repost messages they come across (i.e., "retweet" or RT). A detailed analysis of several networks found within Twitter is found in Chapter 10.

Multimedia Blogs and Podcasts

As bandwidth and multimedia support has increased, a variety of services related to blogs have appeared including video blogs (vlogs), photo blogs, mobile blogs (moblogs), and audio blogs (typically called podcasts). These multimedia blogs may focus on a specific topic or may focus on the everyday experiences of a specific individual as a lifelogging/lifeblogging process. Typically, people can comment in text to the initial posts, and occasionally systems allow you to reply using multimedia as with vlogs that encourage video replies to other videos. Some multimedia sites, such as the humorous photo blog Failblog, encourage submissions of content from readers but are vetted by those in charge of the site before being posted. Other multimedia blogs are authored by an individual or small group, which are read by small to large sized groups. Moblogs are unique in that mobile devices such as smart phones are used to upload photos, videos, and text and often automatically tagged with location information (e.g., see moblog.net and Qik). Podcasts may include audio or video content and like traditional blogs can be subscribed to using tools like iTunes so that new content is automatically downloaded. They differ in that the facilities for commenting on podcasts are not as common, although they may be provided as part of a web site. Technological improvements will likely make search tools and annotation of multimedia content such as videos, images, and audio more common in the future.

The networks created by multimedia blogs and podcasts are similar to those created by blogs and microblogs. They connect content to content and, by extension, content authors. People can also connect to others who use similar tags to describe multimedia posts. In addition, moblogs can connect individuals based on shared locations.

2.4.6 Social Sharing

Social sharing sites are designed to allow individuals to share content, typically of a certain type (e.g., videos, photos, web sites). They provide an alternative to purposeful searching for content on search engines by allowing a community of peers to collaboratively identify and share interesting content. They are a modern incarnation of browsing, where the masses decide what is on the shelf. Although social sharing occurs in other social media tools such as social networking sites (e.g., photo sharing and tagging on Facebook), social sharing sites are typically more content-centric and stand-alone tools.

Video and TV

Since the widespread use of digital video cameras, people have been uploading their videos to share with others. Sites like YouTube and Vimeo allow the masses to freely upload and share video content and link to it or embed it within other web sites such as blogs. Corporations, universities, and media outlets post content on their own YouTube "channels" and occasionally participate in YouTube's partnership program. Sites like Google Video, Metacafe, Hulu, and Netflix streaming service allow users to search for and view TV shows, movies, and shorter video clips. These sites are home to a number of social network structures. For example, YouTube allows users to become "friends" or contacts with one another. In addition, relationships can be created between users when they comment on one another's videos, make a video a "favorite," or subscribe to a user's stream of uploaded videos. Networks of videos that relate to one another are also created based on having shared tags or shared viewers. A detailed analysis of several networks found within YouTube is found in Chapter 14.

Photo and Art

Since the invention of cameras, people have shared photos via albums, scrapbooks, and fridge magnets. Likewise, art is largely created so it can be shared. Social media tools enable the sharing of photos and digital copies of artwork with a select group of other people or the world at large. One of the most popular services is Yahoo's Flickr, which hosts a vast collection of digital photographs that are attached to individuals (e.g., photographers who upload them), groups, and tags that describe them. Like most social sharing sites, Flickr allows users to create networks of contacts (like friends in social networks sites) and limit the distribution of photos to just those individuals or to the world at large. Many similar services exist, such as Picasa, SmugMug, PhotoBucket, and DeviantART, and some sites like Slide that allow you to arrange photos in a slideshow with accompanying music. There are also a number of stock photo and vector art sites like stock.xchng and iStockphoto that allow individuals to purchase content one image at a time. Additionally, sites like Facebook include photo sharing and tagging elements.

The richly annotated content enables the construction of many types of networks. Some networks connect people who appear in photos together, whereas other networks connect people who follow others' art or are in the same group. Implicit networks can be generated that connect people who use similar tags, comment on others' photos, favorite others' photos, take pictures in similar locations, or use the same type of camera. A detailed analysis of several networks found within Flickr is found in Chapter 13.

Music

A number of social sharing sites revolve around music, including sites like Last.Fm, imeem, and Sonic Garden. These sites share many properties with video and photo sharing sites such as the ability to friend others, post comments, and navigate the site via various metadata fields such as tags and artist. One way that some sites, such as imeem, differ is their capacity to create explicit playlists, a modern incarnation of the mix tape, recognizing the value that comes from just the right collection of songs. Some sites like Last.Fm utilize collaborative filtering technologies to help recommend music that you are likely to enjoy based on music you currently enjoy. Collaborative filtering tools, a subset of recommender systems, use data from other users to create personalized recommendations. If you like songs that a certain group of users also likes, then you will probably also like songs they like that you have not yet heard. The networks created by music sharing sites are similar to those of other sharing sites.

Bookmarks, News, and Books

As users review content on the World Wide Web, it is common to want to save a web page or a pointer to it for later use. These collections of bookmarks can be valuable for other people as well. Several services have emerged to allow users to save a bookmark pointer to a web site and share that link with others. Services like deli.cio.us, digg, reddit, and stumbledupon provide a range of tools for users to collect pointers to useful and interesting material on the web, annotate it in various ways, and publish it to select others or the public. A related set of tools allows users of these sites to filter, search, and sort the accumulated links from many other users. Many users are eager to recall useful material on the web and are often willing to signal their interest or appreciation for certain web sites to others. Sites such as digg refined the idea of sharing bookmarks by adding a voting mechanism to the process to further sort quality content. These services have developed a strong following of users who signal their interest in web sites or blog postings, which are then aggregated and ranked for others to see. Bookmark sharing services provide a rapid way to identify novel and interesting information, build a historical trace, and form communities of shared interest. Similar sites exist for non-web-based content such as books (e.g., Goodreads, LibraryThing) and journal articles (citeulike, connotea), although they often include more extensive user-contributed reviews. These systems include similar networks to other content sharing sites, as well as networks based on voting behavior.

2.4.7 Social Networking Services

In 1971, Les Earnest wrote the "finger" program that allowed users of a system to check on another user's status. When a user requested it through the finger program, a file named ".plan" would be displayed to other users.[4] This file soon became the business card and office door for many early users of the Internet and the networks that preceded it. Some users even updated the file regularly to note their current location, activity, and state of mind. This simple status and profile system evolved over time and inspired the creation of systems that allow people to present themselves to others. Modern incarnations, called social networking services, allow people to share contact information, text, images, and videos about themselves with their self-identified friends or followers. Earlier examples like Friendster established the basic outlines of the social network feature set in which users "friend" other users, enabling them to share content and receive updates about each others' activity.

Social and Dating

Services such as Facebook and MySpace have come to dominate social networking services, even though they are just a few years old. With 400+ million users in 2010, many of whom are regularly active, Facebook contains one of the largest machine readable "social graphs" on earth. There are many ways people connect to one another in Facebook, from the obvious "friending" that starts a Facebook relationship, to the many ways people can subsequently interact by writing on one another's "wall," indicating that they "like" other people's content, sending messages, tagging photos, and joining common fan clubs or groups. Facebook and related systems are rich sources of social network data as a result.

Other services have come to dominate specific geographical regions like the prominence of Orkut in Brazil and India and Bebo in the UK and StudioVZ in Germany. Some social sites have a focus on dating such as Tagged, eHarmony, and Match, or a specific demographic group like BlackPlanet. These often include detailed profile information not available on more general-purpose social networks so that members can find "compatible" companions.

Many of these social networking services, such as Facebook, impose restrictive terms of use for their data. In contrast to the mostly open and free model that surrounds Twitter, Facebook constrains what data can be accessed and the duration that the data may be used. As a result, analysis of Facebook social networks can be challenging. Individuals may extract some Facebook data related to their own interactions and social network, but even that data may only be used for short periods and for specific purposes. See Chapter 15 for a discussion of how to analyze your personal Facebook network.

Professional

Services like LinkedIn, Plaxo, and XING provide a social network feature set tuned to the self-presentation of career professionals engaged in business networking. Users can post their resume, receive and send targeted job invitations, recommend co-workers, introduce a colleague to another colleague, exchange private messages, and join groups such as university alumni associations or special interest groups. These networks are becoming a vital part of the job search process in many industries.

Niche Networks

More recently, sites like Ning and Grou.ps make it possible for users to easily create their own niche social networking site based around particular topics. For example, educators interested in using new technologies meet at the Classroom 2.0 Ning site, and people who cope with diabetes meet at TuDiabetes. These sites can be open to anyone to join or closed so that only invited members can join. In addition, a host of other online communities incorporate the core social networking features such as profile pages, while integrating them with community content. For example, Ravelry, a community of knitters, shows users current yarn stashes, favorite plans, and completed knitting projects. In Chapter 9 we explore Ravelry.

2.4.8 Online Markets and Production

Many social media sites facilitate the creation, evaluation, and exchange of goods and services.

Financial Transaction

Networks of exchange have always been at the core of marketplaces where buyers and sellers meet, exchange news, and make trades, purchase goods, or form plans for future activities. There has been enormous demand for online marketplaces in the form of auction sites such as eBay and Amazon or advertising sites like craigslist that facilitate location-specific ads for products, services, apartments, and jobs. These services generate communities of buyers and sellers who share an interest in the same products. Many small businesses and professionals, such as artists, craftspeople, or photographers, routinely advertise their products through personal or collective web sites, along with service providers, consultants, and personal trainers. These independent small businesspeople can reach a wide audience and develop credibility through reputation system tools like eBay's feedback mechanism. These marketplaces create networks that connect sellers and buyers through transactions, creating a trade network.

[4] www.rajivshah.com/Case_Studies/Finger/Finger.htm

Related services such as Kiva allow users to donate money to entrepreneurs in developing countries, facilitating microloans, and then follow their progress via blog posts and public repayment statistics. Finally, a growing number of services called "prediction markets" allow people to buy and sell assets whose cash value is tied to a future event (e.g., a Republican will be the next U.S. president). The market prices are interpreted as the probability of the event occurring. Examples include Intrade, Betfair, and more specific markets such as the Hollywood Stock Exchange. These financial transaction services create a wealth of network data on purchasing patterns that can be used to understand market dynamics, or as the basis for recommender systems like those found at Amazon.

User-Generated Products

A host of social media sites focus on collaboratively developing, sharing, or selling products. The open source software movement is an excellent example, where users contribute code to develop software tools that are then made freely available. Sites like Sourceforge and Codeplex provide tools to support developer communities by tracking changes to the software, monitoring the number of downloads, and providing basic discussion capabilities. Other sites like TopCoder allow users to compete against each other in programming tasks, where the winner gets paid by the organization or individual who posted the challenge. These sites create networks that connect people based on shared projects and challenges. Amazon's Mechanical Turk provides a platform for supporting a host of "human intelligence tasks" such as classifying items and identifying images. This "marketplace for work" allows people from around the world to perform these tasks for micropayments that can add up over time. Other communities focus on a specific product. For example, at Threadless, users can purchase T-shirts from a host of user-created designs. These sites generate trade networks, as well as networks that connect people who work on similar projects or use similar tags to describe their products.

Review Sites

Many social media sites allow people to post reviews of products or services. Some sites like ePinions and Amazon, support written reviews as well as ratings of almost every conceivable product. Local versions of review sites, such as Angie's List and Yelp, focus on location-based services such as restaurants, shopping, and nightlife, or service providers such as doctors, contractors, and service professionals. These sites create explicit networks when people friend each other (e.g., on Yelp), as well as implicit networks when users favorite places or services. They also create networks that connect people who provide similar ratings as others.

2.4.9 Idea Generation

Organizations are increasingly looking for ways to benefit from the collective intelligence of the masses. Several social media sites use "idea generation" tools to help solicit and evaluate new ideas. Companies like IdeaConnection allow organizations to post proprietary challenges to a community of problem solvers. If someone solves the problem, that person is awarded a specified dollar amount. If nobody solves the problem, no money is exchanged. Other tools by companies like Chaordix and IdeaScale allow users to post ideas and vote on others' ideas, helping the best ones bubble to the top. These services create networks that connect people based on who voted on whose ideas. They also create networks that connect ideas to other ideas based on the number of people who liked both ideas.

2.4.10 Virtual Worlds

Virtual worlds, graphical worlds, and massively multiplayer games attempt to model physical places as well as face-to-face interaction. Modern virtual worlds allow users to build new spaces, create objects, and use powerful programming languages to automate their behavior. These sophisticated forms of social media create remarkably rich collections of networks. Even services offering relatively simple game experiences like card games and backgammon offer sophisticated ways of creating friend networks, teams, and rankings. Game systems commonly allow users to create affiliation networks when players join clubs, guilds, tribes, or teams. Within the game play are other processes that create networks as records are created when users shoot or kill one another or trade less lethal materials.

Virtual Reality Worlds

Although many multiplayer games continue to focus on combat role playing, many "social" virtual worlds have become a means for widely dispersed groups to maintain personal contact. These include systems like Second Life and The Sims, designed for adults, and far more popular systems designed for children and youth such as Club Penguin, Webkinz, and Habbo. Virtual worlds typically offer a range of traditional communication channels, as well as the ability to manipulate "avatar" bodies that pose near one another. Like text chat, these systems support synchronous communication. Virtual worlds allow a number of people who occupy the same "room" to meet and "talk" by speaking, posing, gesturing, and sending lines of text or shared spatialized audio conferencing with one another. Because this interaction happens in real time, all the participants must be active at the same time. But in return, virtual worlds provide a powerful sense of social and physical presence that is absent in asynchronous media. Virtual worlds often support simulations of the multichannel quality and nuances of face-to-face interaction by integrating lines of text with gesture, pose, and voice. For children and youth, they provide an engaging environment where users can earn virtual cash by playing games or completing virtual jobs and use their virtual cash to decorate their virtual home or feed their virtual pets.

Virtual worlds and the social data created in them are typically owned by the company that provides them. Thus, owners of virtual world servers have had monopoly control over their systems making it hard to access data for analysis purposes. However, when data are available, it is a rich source of network connections, which are created in virtual worlds whenever users exchange text or virtual items, are near one another, or interact with the same objects.

Massively Multiplayer Games

Massively multiplayer online games (MMOs) are video games that include hundreds or thousands of players who interact simultaneously in a persistent virtual world. They utilize the Internet, although they may be played on a computer or game console such as Xbox 360 and PlayStation 3. There are many types of MMOs including role-playing games (e.g., World of Warcraft, Everquest), strategy games (Mankind, War of Legends), first-person shooter games, racing games, and many more. They may take place in fantasy worlds or virtual worlds that correspond with specific locations such as World War II battlefields or cities. Many MMOs include complex social arrangements such as guilds, tribes, or teams; sophisticated collaboration tools including live audio-feeds; and virtual currencies that allow players to purchase items needed to complete quests or build their empires.

2.4.11 Mobile-Based Services

The next step from virtual reality is into physical reality. Recent waves of social media tools integrate hardware and software tools to enable users to annotate physical locations. Mobile computer devices like the Apple iPhone and Motorola Droid phones are now integrated with location detection features based on Global Positioning System (GPS) and cell tower location services. Combined with compasses, these devices can reliably locate and orient a user in space. Once the device has an accurate sense of its location, that information can be associated with all the digital objects created with the device. Photos can be easily linked to the place on earth where they were taken. Restaurant reviews can be associated with a map of the location of the restaurant. A comment can be linked to where it was made. In addition, mobile devices lead to the creation of more social media content because it can be created and captured in spare moments and when notable events unfold.

Location is key to social media services that want to provide contextual information about the world immediately around users. When a device is in a known location, simply asking for content associated with "here" can quickly deliver relevant material. Furthermore, a host of additional sensors can be used to provide functionality that can be called upon by new social media services. These include still and video cameras, audio and motion sensors, accelerometers that detect motion, and tools that enable device-to-device connections. This combined functionality and the proliferation of smart devices promises a bright future for mobile social media tools. Already we see companies making their services easy to access and digest using small screens and relatively slower network and computing resources. As more users access social media services from mobile devices, new opportunities will emerge. Although we are still in the early stages of mobile-based social media services, some interesting examples exist.

Location Sharing, Annotation, and Games

Services such as Google Latitude allow users to sense and plot their location on a map and then share that location with other authorized users. The position of other users who have selected to share their location can be seen on the map. Social media services like Gowalla, Loopt, and Foursquare allow users to share their current location with friends (often via other sites like Facebook), leave virtual messages or gifts for others, and write reviews of locations such as restaurants, bars, and parks. Foursquare even adds a game layer with score keeping based on patterns of location sharing. For example, the person who "checks in" from the same location more than anyone else is declared the "mayor" of that place. Mobile, social games like Geocaching and Letterboxing encourage people to hide "caches" that often include small awards or stamps that other players can find, often using GPS tools. SCVNGR provides a more general geogaming platform that lets anyone create location-based mobile games and tours that those with a mobile device can play. Finally, sites like MapMyRun allow users to create and share jogging trails on maps, as well as log training miles. Similar sites exist for bikers, hikers, and even golfers, some of which use data from GPS devices such as Garmin products that allow users to upload data to a computer about their heart rate, elevation, and speed.

These location sharing and annotation services often contain network structures similar to those found in other social media services, but with geographical location as an additional dimension. Place joins the set of other entities found in many social media services like people, tags, dates, and connections. This allows you to create networks that connect people to each other based on who is within a certain distance or who frequents the same locations.

2.5 PRACTITIONER'S SUMMARY

Social media tools have become ubiquitous, despite their recent development. Business success strategies that make good use of the power of social networks have been reported [14–16], whereas others that have shown positive effects when applied to government projects [17, 18] are evolving rapidly. However, for every successful example, there are numerous failed attempts suggesting the need for more systematic methods for analyzing and understanding social media environments.

There is no agreed-upon taxonomy of social media tools or characteristics. Yet it is essential that practitioners spend time understanding which services match their personal and professional needs. Corporate and government decision makers who are seeking to use social media for advertising and promoting their products and services will be more successful if they learn which mechanism best reaches their desired audience, and what constitutes acceptable etiquette in those communities. As a starting point, we discussed the following six key dimensions that help characterize social media systems:

- Size of producer and consumer population
- Pace of interaction
- Genre of basic elements
- Control of basic elements
- Types of connections
- Retention of content

In addition we briefly introduced some of the more popular social media systems, which are outlined in Table 2.3. In doing so, we highlighted the types of networks each of them create, laying the groundwork for the rest of the book, which will discuss how to gain actionable insights from the analysis and visualization of those networks.

2.6 RESEARCHER'S AGENDA

The widespread adoption of social media tools has begun to usher in a golden age of social science research. Social media systems provide a wealth of data about communication patterns, location information, friendships, and other social arrangements. Mining this data is bound to provide numerous insights into human nature. There are also many important questions that need answering to help us effectively utilize social media tools to achieve our goals. For example, we need to understand how to motivate voluntary participation [19], develop persuasive systems [20], govern social media communities [21], organize activities to meet specific goals, find the limits of scalability, and develop tools to better visualize and understand social activity. Related issues of trust, empathy, responsibility, and privacy have strong research foundations, which can be helpful to a wide range of practitioners. Addressing these issues will help designers and community managers make well-informed decisions rather than simply relying on intuition and anecdotes.

The rapid pace of commercial development offers new challenges to the research community to evaluate the impact of design changes, novel policies, and evolving norms. What forms of recognition or reward are appropriate for different domains? How can communities that involve participants with different expectations, skills, and experience be accommodated? How can malicious behavior be reduced? Can envisioned benefits to health, education, energy, or international development become a reality?

References

[1] V. Bush, "As We May Think," Atlantic Monthly, 1945.
[2] D. Engelbart, A conceptual framework for augmentation of man's intellect, in: P.W. Howerton, D.C. Weeks (Eds.), Vistas in Information Handling, vol. I, Spartan Books, Washington, 1963, pp. 1–29.
[3] B. Shneiderman, C. Plaisant, Designing the User Interface: Strategies for Effective Human-Computer Interaction: Fifth ed., Addison-Wesley Publ. Co., Reading, MA, 2010.
[4] E. Hall, The Silent Language, Doubleday Press, New York, 1990.
[5] M. Smith, S. Farnham, S. Drucker, The Social Life of Small Graphical Chat Spaces, Proceeding ACM CHI 2000 Conference, The Hague, Netherlands, March 2000, New York: ACM Press, 2000.
[6] D. Powazek, Chapter 8: Barriers to Entry … Making Them Work for It, In Design for Community, Waite Group Press, 2001.
[7] E. Ostrom, Governing the Commons: The Evolution of Institutions for Collective Action, Cambridge University Press, New York, 1990.
[8] R. Axelrod, The Evolution of Cooperation, Basic Books, New York, 1984.
[9] M. Smith, Tools for navigating large social cyberspaces, Commun. ACM 45 (4) (April 2002) 51–55.
[10] P. Resnick, N. Iacovou, M. Sushak, P. Bergstrom, J. Riedl, GroupLens: An Open Architecture for Collaborative Filtering of Netnews, in: ACM Conference on Computer Supported Collaborative Work Conference, 10/1994, Chapel Hill, NC, 1994, pp. 175–186.
[11] T. Berners-Lee, Weaving the Web: The Original Design and Ultimate Destiny of the World Wide Web, Harper Paperbacks, 2000.

[12] L.A. Adamic, E. Adar, Friends and neighbors on the web, Soc. Networks 25 (3) (2003) 211–230.
[13] J. Kelly, B. Etling, Mapping Iran's Online Public: Politics and Culture in the Persian Blogosphere, Berkman Center Research Publication No. 2008-01, 2008, Available at: http://cyber.law.harvard.edu/publications/2008/Mapping_Irans_Online_Public.
[14] C. Li, J. Bernoff, Groundswell: Winning in a World Transformed by Social Technologies, Harvard Business Review, Boston, MA, 2008.
[15] J. Porter, Designing for the Social Web, New Riders, Berkeley, CA, 2008.
[16] D. Tapscott, A.D. Williams, Wikinomics: How Mass Collaboration Changes Everything, Portfolio, Penguin Books, London, UK, 2007.
[17] T. Kalil, Leveraging cyberspace, IEEE Commun. Mag. 34 (7) (1996) 82–86.
[18] B. Noveck, Wiki Government: How Technology Can Make Government Better, Democracy Stronger, and Citizens More Powerful, Brookings Institution Press, Washington, DC, 2009.
[19] K. Ling, G. Beenen, P. Ludford, X. Wang, K. Chang, X. Li, et al., Using social psychology to motivate contributions to online communities, J. of Computer-Mediated Commun. 10 (4) (2005) article 10.
[20] B.J. Fogg, Persuasive Technology: Using Computers to Change What We Think and Do, Morgan Kaufmann, San Francisco, CA, 2002.
[21] J. Preece, Online Communities: Designing Usability, Supporting Sociability, John Wiley & Sons, Chichester, UK, 2000.

Additional Resources

Easley, D., & Kleinberg, J. (2010). *Networks, crowds, and markets: Reasoning about a highly connected world*. New York: Cambridge University Press.
Rheingold, H. (2002). *Smart mobs: The next social revolution*. Cambridge, MA: Basic Books.
Shirky, C. (2008). *Here comes everybody: The power of organizing without organizations*. New York, NY: The Penguin Press.
Smith, M., & Kollock, P. (Eds.), (1999). *Communities in Cyberspace*. London, UK: Routeledge.
Wenger, E., White, N., & Smith, J.D. (2009). *Digital Habitats; stewarding technology for communities*. Portland, OR: CPsquare.

This page intentionally left blank

CHAPTER

3

Social Network Analysis

Measuring, Mapping, and Modeling Collections of Connections

OUTLINE

3.1 INTRODUCTION

Human beings have been part of social networks since our earliest days. We are born and live in a world of connections. People connect with others through social networks formed by kinship, language, trade, exchange, conflict, citation, and collaboration. Computer technologies used to create social networks are relatively new, but networks of social interactions and exchanges are primordial. Simply stated, a network is a collection of things and their relationships to one another. The "things" that are connected are called nodes, vertices, entities, and in some contexts people. The connections between the vertices are called edges, ties, and links. Many natural and artificial systems form networks, which exist in systems from the atomic level to the planetary level. *Social* networks are created whenever people interact, directly or indirectly, with other people, institutions, and artifacts. Social network theory and analysis is a relatively recent set of ideas and methods largely developed over the past 80 years. It builds on and uses concepts from the mathematics of graph theory, which has a longer history. Using network analysis, you can visualize complex sets of relationships as maps (i.e., graphs or sociograms) of connected symbols and calculate precise measures of the size, shape, and density of the network as a whole and the positions of each element within it.

The recent proliferation of Internet social media applications and mobile devices has made social connections more visible than ever before (Chapter 2). The idea of networks, whether they are composed of friends, ideas, or web pages, is increasingly an important way to think about the modern world. Social network analysis helps you explore and visualize patterns found within collections of linked entities that include people. From the perspective of social network analysis, the treelike “org-chart” that commonly represents the hierarchical structure of an organization or enterprise is too simple and lacks important information about the cross connections that exist between and across departments and divisions. In contrast with the simplified tree structure of an org-chart, a social network view of an organization or population leads to the creation of visualizations that resemble maps of highway systems, airline routes, or rail networks (See Chapter 8). Social network maps can similarly guide journeys through social landscapes and tell a story about how some points are at the center or periphery of the network. Transportation networks where distance is measured in number of flights or roads from one city to another city are familiar. They inspire application to less familiar networks of electrical connections, protein expression, and webs of information, conversation, and human connection.

Social network analysis and metrics are described in several excellent books and journals [1–6]. This chapter touches on the key historical developments, ideas, and concepts in social network analysis and applies them to social media network examples. We have left details of advanced topics and mathematical definitions of various concepts to the many fine technical works. The following is intended as an introductory survey of the core network concepts and methods used in subsequent chapters, which focus on the networks that can be extracted from social media sources like Twitter, Facebook, email, discussion forums, YouTube, Flickr, wikis, and the web.

3.2 THE NETWORK PERSPECTIVE

Network analysts see the world as a collection of interconnected pieces. Those studying social networks see *relationships* as the building blocks of the social world, each set of relationships combining to create emergent patterns of connections among people, groups, and things. The focus of social network analysis is between, not within people. Whereas traditional social science research methods such as surveys focus on individuals and their attributes (e.g., gender, age, income), network scientists focus on the connections that bind individuals together, not exclusively on their internal qualities or abilities. This change in focus from attribute data to relational data dramatically affects how data are collected, represented, and analyzed. Social network analysis complements methods that focus more narrowly on individuals, adding a critical dimension that captures the connective tissue of societies and other complex interdependencies.

Network analysis shares some core ideas with the real estate profession. In contrast to approaches that look at internal attributes of each individual, network analysis shares the real estate focus on location, location, location! The interior of a house may be a liability, but where a property is located matters far more when trying to get a good sale price. The network perspective looks at a collection of ties among a population and creates measurements that describe the location of each person or entity within the structure of all relationships in the network. The position or location of a person or vertex in relation to all the others is a primary concern of social network analysis. Many network explanations look for causes of outcomes in the patterns of connections around an individual instead of their personal characteristics. “Know who” is often more important in network explanations than “know how.” Network approaches observe that different people in similar circumstances and social positions often act in similar ways. Positions within networks may be as significant a factor as any aspect of the people who occupy them. Network analysis argues that explanations about the success or failures of organizations are often to be found in the structure of relationships that limit and provide opportunities for interaction [7].

Many network concepts are intuitive and echo familiar phrases like “friend of a friend,” “word of mouth,” and “six degrees of separation.” Other network terms like “transitivity,” “triadic closure,” and “centrality” (see Section 3.5) may be unfamiliar terms for familiar social arrangements. Many of us recognize social network differences among people: we know some people who are “popular” and have connections to many others. We may also know some people who may be less “popular” but are still “influential,” connecting to a smaller number of people who have “better” connections. Network analysis recognizes these and other less intuitively sensed patterns in social relationships, like measuring the number of your friends who know each other and how much a person occupies a gatekeeper or bridge role between two groups. The network analysis approach makes the web of interconnections that bind people to one another visible, creating a mathematical and graphical language that can highlight important people, events, and subgroups.

3.2.1 A Simple Twitter Network Example

To better understand the network perspective, consider the social network of Twitter users shown in Figure 3.1 (see Chapters 2 and 10 for a description of Twitter). It is an example of a sociogram, also called a network graph, which is a common way of visualizing networks. Like all networks, it consists of two primary building blocks: vertices (also called nodes or agents) and edges (also called ties or connections). The vertices are represented by images of the Twitter users, and the edges are represented by the lines that point from one vertex to another.

This simple graph paints a picture of the social relationship of the Twitter users who tweeted about a 2009 workshop on information in networks at New York University[1] by including the text string "#WIN09." The size of each Twitter user's profile image is determined by the user's total number of tweets as reported by the Twitter Application Programmer Interface (API), which gives sophisticated users access to powerful services. This is one example of how attribute data (e.g., data that describe a person) can be overlaid onto a network. A line, or edge, exists between two people when one "follows" the other or if one user "mentions" or "replies" to the other. All of these connections in aggregate reveal the emergent structure of two distinct groups with few connecting links. This accurately represents the way the workshop brought together previously separate clusters of people from different disciplines. It also helps identify individuals who fill important positions in the network, such as those who many people follow and those who are connected to both clusters. This and following

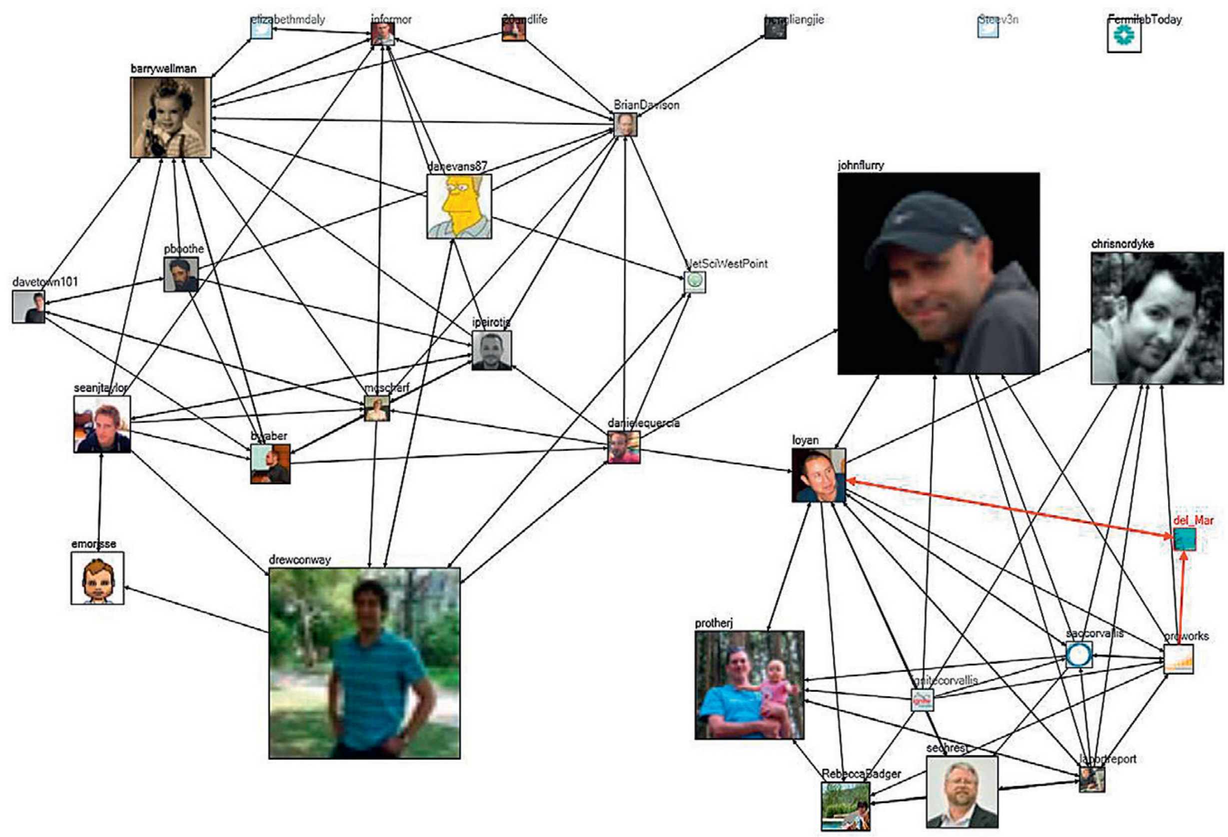

FIGURE 3.1 A NodeXL social media network diagram of relationships among Twitter users mentioning the hashtag "#WIN09" used by attendees of a conference on network science at New York University in September 2009. The size or each user's vertex is proportional to the number of tweets that user has ever made.

[1]http://winworkshop.net

chapters will provide a guide to creating maps like these from Twitter and other social media platforms and data sources. For now, let's consider the major components of a network in a bit more detail.

3.2.2 Vertices

Vertices, also called nodes, agents, entities, or items, can represent many things. Often they represent people or social structures such as workgroups, teams, organizations, institutions, states, or even countries. At other times they represent content such as web pages, keyword tags, or videos. They can even represent physical or virtual locations or events. They often correspond with the primary building blocks of social media platforms as described in Chapter 2: pages in wikis, friends in social networking sites, and posts or authors in blogs.

Although not necessary for network analysis, having attribute data that describe each of the vertices can add insights to the analysis and visualizations. For example, Figure 3.1 used descriptive attribute data about the total number of posts to convey a sense of who is most active on Twitter. Other attribute data from Twitter, such as the number of followers, people they follow, and their join date, can also be mapped to visual attributes (see Chapter 10). More generally, attribute data may describe demographic characteristics of a person (age, gender, race), data that describe the person's use of a system (number of logins, messages posted, edits made) or other characteristics such as income or location. In network visualization tools such as NodeXL, attribute data can be mapped to visual properties such as the size, color, or opacity of the vertices (see Chapter 4).

3.2.3 Edges

Edges, also known as links, ties, connections, and relationships, are the building blocks of networks. An edge connects two vertices together. Edges can represent many different types of relationships like proximity, collaborations, kinship, friendship, trade partnerships, citations, investments, hyperlinking, transactions, and shared attributes. A tie can be said to exist if it has some official status, is recognized by the participants, or is observed by exchange or interaction between them. A tie is any form of relationship or connection between two entities.

Network scientists have developed a language to describe different types of edges. In Section 2.3.5 of Chapter 2, we introduced the core types of connections that occur in social media networks. Here we describe how those concepts map to network and graph theory concepts more generally.

Undirected or *directed* edges are the two major types of connections. Directed edges (also known as asymmetric edges) have a clear origin and destination: money is lent from one person to another, a Twitter user follows another user, an email is sent to a recipient, or a web page links to another web page. They are represented on a graph as a line with an arrow pointing from the source vertex to the recipient vertex (see Figure 3.1). Directed edges may be reciprocated or not. If I sent you a message you may send one back in return, or not. An undirected edge (also known as a symmetric edge) simply exists between two people or things: a couple is married, two Facebook users are friends, or two people are members of the same organization. No origin or destination is clear in these mutual relationships. They cannot exist unless they are reciprocated. Undirected edges are represented on a graph as a line connecting two vertices with no arrows.

Edges can be represented by different types of data. The simplest type of edge, an *unweighted* edge or binary edge, only indicates if an edge exists or not. For example, a friendship tie between Facebook users either exists or it does not. In contrast, a *weighted* edge includes values associated with each edge that indicate the strength or frequency of a tie. For example, a weighted edge between two Facebook users may indicate the number of photo comments exchanged or the duration of a friendship. Weighted edges are often represented visually as thicker or darker lines or as more or less opaque lines. Including weighted data is preferable because they provide additional information about each tie. However, many social network analysis metrics (discussed later) are designed for unweighted networks. Fortunately, any weighted network can be converted to an unweighted one by choosing a cutoff point. For example, an unweighted edge could be shown between individuals who exchanged at least 10 email messages, with no edge between people who exchanged fewer than 10 messages.

3.2.4 Network Data Representations

Because network data differ from attribute data, there are different ways of representing it. With attribute data, it is common to create a data matrix where each row represents an individual and each column represents individuals' characteristics, behaviors, or answers to survey questions. A related approach can be used to represent relational data. Like attribute matrices, each row represents an individual in the network. However, unlike attribute matrices, each column also represents an individual as shown in Table 3.1.

Different types of edges can be represented in network matrices. Table 3.1 describes a directed network because not all connections are reciprocated. For example, Ann "points to" Bob as shown in row 1, but Bob

ADVANCED TOPIC

The Foundations of Graph Theory

Network analysis is rooted in the work of the mathematician Leonhard Euler who in 1736 studied whether a single path could be walked over the Seven Bridges of Königsberg that connected islands in the river Pregel (which flows through what was then Prussia and is now Kaliningrad in Russia) without crossing any bridge more than once. By reimagining the problem in terms of vertices and edges, he showed it is impossible to cross each bridge just once. Although the problem seems abstract, its solution led to the development of the mathematics of graph theory and, notably, hundreds of years later, the mathematical work of Paul Erdös and Alfréd Rényi on random graphs in the 1950s, an important theoretical development that allows for the generation of a graph from random processes. Social network analysis builds on these concepts and extends them to capture the nonrandom connections that occur among groups of people.

TABLE 3.1* A Network Represented as a Matrix

	Ann	Bob	Carol
Ann	0	1	1
Bob	0	0	0
Carol	1	0	0

**This network is a directed network, as it is not symmetrical (i.e., Ann points to Bob in row 1, but Bob doesn't point to Ann in row 2). It is a simple binary network: either a tie exists (value = 1) or not (value = 0).*

TABLE 3.2* A Network Represented as an Edge List

Vertex1	Vertex2
Ann	Bob
Ann	Carol
Carol	Ann

**Individuals in the Vertex1 column "point to" those in the Vertex2 column in this directed network. The network is implied to be a binary network. Additional columns could be used to describe each edge. For example, an Edge Weight column could be added with values representing the strength of various ties.*

does not "point to" Ann as shown in row 2. If it were an undirected network it would be a symmetric matrix; if Ann points to Bob then Bob must necessarily point to Ann. This network is a binary network because it only includes 1s and 0s, where a 1 indicates that there is a connection and a 0 indicates that there is no connection. Allowing additional values would create a weighted network. For example, the 1s could be replaced with the number of email messages sent to the other person. Finally, notice that the diagonal of the matrix connects each person with himself or herself. In this network, like most networks, the diagonal values are 0 indicating that a person does not "point to" herself. However, in some networks a "self-loop" connecting a person to herself can exist. For example, a person may send herself an email message as a reminder. Network matrices are powerful forms of representation that lend themselves to efficient mathematical manipulation for those inclined. However, they can also become quite large and challenging to navigate, particularly when networks have relatively few connections.

An alternate network representation is called an "edge list." Like its name suggests, it is simply a list of all edges in the network as shown in Table 3.2. This is the same network as shown in Table 3.1. Individuals in the Vertex1 column "point to" those in the Vertex2 column. Unless data describing the value of each edge are provided in additional columns, the network is implied to be a binary one. Self-loops are possible to

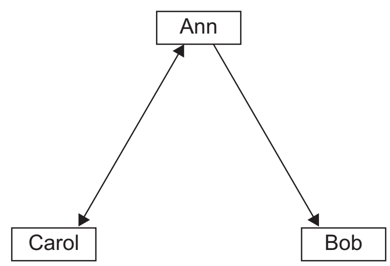

FIGURE 3.2 The directed, binary network described in Tables 3.1 and 3.2 represented as a network graph. Arrows indicate the direction of the connection (e.g., from Ann to Bob).

represent in edge lists by having a row with the person's name repeated in both columns. Throughout this book, we will use edge lists instead of matrices.

The final method for representing networks is through network graphs. Figure 3.2 is a network graph based on the data in Table 3.2. It makes immediately clear that the relationship between Ann and Carol is reciprocated (i.e., there are arrows on both sides of the line connecting them) and that there is no connection between Bob and Carol. Our earlier analysis of Figure 3.1, another network graph, demonstrates how network graphs can lead to insights that are hard to identify in tabular data, particularly when large networks are presented. However, many network graphs require significant preparation to assure that they are readable as described in Section 3.9.

3.3 TYPES OF NETWORKS

Social networks range in size from a handful of people to national and planetary populations. They also differ in the types of vertices they include, the nature of the edges that connect them, and the ways in which they are formed. In this section we introduce some of the distinctions that network scientists have identified to describe different types of networks. These distinctions affect the metrics and maps generated from them, as well as their interpretation.

3.3.1 Full, Partial, and Egocentric Networks

It is often useful to consider social networks from an individual member's point of view. Network analysts call the individual that is the focus of attention "ego" and the people he or she is connected to "alters." Some networks, called egocentric networks, only include individuals who are connected to a specified ego. For example, a network of your personal Facebook friends would be an egocentric network because you are, by definition, connected to all other vertices. Other egocentric networks and their associated "subgraphs" (see Chapter 6) may extend out from an ego, reaching not only friends, but also friends of friends. More generally, egocentric networks can extend out any number of "degrees" from ego. The basic "1-degree" ego network consists of the ego and their alters. The "1.5-degree" ego network extends the 1-degree network by including connections between all of the alters. For example, a Facebook 1.5 degree ego network would characterize which of your friends know each other (see Chapter 11). The "2-degree" ego network extends the 1.5-degree network by including all of the alters' own alters (i.e., friends of friends), some of whom may not be connected to ego. These three ego networks allow you to look at increasingly larger, but still "local" neighborhoods around a particular individual in a social network. Higher-degree networks (e.g., 2.5, 3) are feasible to create but not used as often in practice because they can quickly become intractable.

A full or complete network contains all the people or entities of interest and the connections among them. All egos are treated equally. A full network is often created and available when a single system, such as a social media platform, acts as a hub among a group of connected people or groups. For example, the Twitter network includes all users of the service and the connections between them. In practice, it is not always feasible (or particularly insightful) to analyze a full network. Instead, analysts create a partial network by selecting a sample or slice of the full network. For example, Figure 3.1 showed the slice of the Twitter network that included people who used the hashtag "#WIN09." This partial network was not egocentric. Rather it was topic centric. Other partial networks may be created to include a subgroup of users (e.g., all conference attendees), only people and connections that occurred within a specified time frame, or people who have certain characteristics (e.g., CEOs of Fortune 500 companies).

3.3.2 Unimodal, Multimodal, and Affiliation Networks

Up until this point we have only considered networks that connect the same type of entity. These standard networks are called *unimodal networks* because they include one type (i.e., mode) of vertex. They connect users to users or they connect documents to documents, but they don't include both users and documents. However, networks can include different types of vertices creating *multimodal networks*. For example, a network may connect users to discussion forums and blog posts they have commended on. Each vertex on the graph would represent a user, a forum, or a blog post, which could be visually distinguished by different colors or shapes. The SeriousEats Network discussed in Chapter 6 is an example. The rich sets of intersecting networks that form in social media environments include connections between people, photos, videos, messages, documents, groups, organizations, locations, and services. In many cases, these multimodal networks have to be transformed into simpler unimode networks to perform meaningful network analysis, as most network metrics (see Section 3.5) are designed for unimodal networks.

A common type of multimodal network is a *bimodal network* with exactly two types of vertices. Data for these networks often include individuals and some event, activity, or content with which they are affiliated, creating an *affiliation network*. For example, an affiliation network may connect users with wiki pages they edit. People are affiliated with pages. In this network, no two users would directly connect to each other. Likewise, no two wiki pages would directly connect to each other in this type of network.

Bimodal affiliation networks can be transformed into two separate unimodal networks: a user-to-user network and an affiliation-to-affiliation network (e.g., article–to–article network in a wiki) (see Chapter 6, Advanced Topic, Transforming Multimodal Affiliation Networks into Unimodal Networks for details). The user-to-user network connects people based on their links to one another. For example, in a wiki co-edit affiliation network Derek and Marc would be strongly connected because they both edit many of the same wiki pages. The affiliation-to-affiliation network connects the affiliations based on the number of shared users. For example, a pair of wiki pages would be closely connected if many people edited both of the pages (see Chapter 15). More generally, this approach can be used to relate objects of all types (e.g., books, photos, and audio recordings) based on users' behaviors (e.g., purchasing or reading habits) and preferences (e.g., ratings). Affiliation networks are the raw material of many recommender systems that recommend items of interest, such as Amazon's "Customers Who Bought This Item Also Bought" feature. A network data structure can return results to queries like "people who linked to this document also linked to these documents" or "if you link to this document, you may want to link to these people."

3.3.3 Multiplex Networks

Although it is common for two people to be connected in many different ways (for example, by exchanging phone calls, emails, sharing group membership, and being married), most networks only include one type of connection or edge. However, it is possible to consider networks with multiple types of connections, called *multiplex networks*. For example, the Twitter network shown in Figure 3.1 includes three types of directed edges: following relationships, "reply to" relationships, and "mention" relationships. The graph could have uniquely represented each type of edge by using color, different edge types (e.g., dotted lines, solid lines), or edge labels (see Chapter 4). In the case of Figure 3.1, the type of edge was not deemed important, so the multiplex network data were condensed into a standard network that showed a single directed edge if one or more of the three types of connections were present. This strategy of combining multiple types of edges is a common one that allows for the use of network metrics, which are mostly based on standard networks.

3.4 THE NETWORK ANALYSIS RESEARCH AND PRACTITIONER LANDSCAPE

You can find network scientists in nearly every academic discipline and an increasing number of practitioner communities. Network concepts and techniques are now widely found throughout a range of disciplines including sociology, anthropology, communications, computer science, education, economics, physics, management, information science, medicine, political science, public health, psychology, biology, and the humanities. In the past several decades, social scientists have shown that network structures have an influence on health, work, and community. Getting a job, being promoted, catching an illness, adopting an innovation, and many more activities and processes have been explained in terms of social networks. Network structures are important in the biological sciences where research is focused on connections between metabolic and genetic processes. The shape and function of networks can have great consequences as ideas, genes, innovations, or pathogens diffuse through populations. Researchers now apply network theory and methods to understanding how Supreme Court decisions relate to previous cases, how the United States Senate votes (see Chapter 7), how epidemics spread within cities, and how characters in a novel relate to one another (see Chapter 7). Networks are formed from many physical processes and are echoed in a number of structures created inside information systems such as the collection of linked documents within the World Wide Web or an enterprise's collections of files. Information scientists use these links to identify high-quality web pages (e.g., Google's PageRank algorithm), or use the citations from research articles to identify high-impact articles and authors.

Network methods are diffusing beyond academic research, becoming an important tool for managing organizations, markets, and movements. Entrepreneurs apply network analysis techniques to understand how to leverage the powerful effects of word-of-mouth marketing as their customers spread news about their new products to one another. Many politicians recognize the potential power of a connected network of supporters who can be turned into contributors, volunteers, and voters. Engineers use network analysis to build more effective power grids, computer networks, and transportation systems. Law enforcement officers and lawyers analyze email networks to identify and defend potential criminals. And the intelligence community hunts down terrorists by looking at networks created by money trails and kinship. Having at least a basic understanding of network thinking and concepts is a core literacy of our time. Like statistics, network analysis has countless applications to a number of fields.

This book primarily focuses on social network analysis, a subfield of network sciences that focuses on networks that connect people or social units (i.e., organizations, teams) to one another (see Advanced Topic: Early Social Network Analysis). We are also interested in networks that connect human-generated content or artifacts together, such as web sites or cell phones.

ADVANCED TOPIC
Early Social Network Analysis

The social science roots of social network analysis can be found in the early 1800s in the work of the person credited with being the first sociologist, Auguste Comte, and later in the early 1900s in the work of the sociologist Georg Simmel. Both saw patterns of social ties as the main focus of sociology in contrast to the study of individuals and their attributes. Early in the nineteenth century, Comte defined society as more than simply a group of people. He argued that a population became a society only when people had influence on one another and considered the choices and interests of others as part of their own choices. Simmel echoed these ideas at the turn of the twentieth century, focusing social science on the study of how people come together and form groups and associations. These sociologists imagined society as composed of a web of relationships—more than a mass of individuals; they saw societies as networks of reciprocal influence.

The idea of connected actions linking people to one another has remained at the core of the social sciences, but efforts to create a systematic language to record social relationships started only in the twentieth century. Anthropologists studying the range of kinship systems they documented in fieldwork from around the world created symbol systems that are related to social network analysis. Their maps of who is related to whom were early forms of social networks focused on just the subset of social ties that are considered to be "family." The core concepts and methods of modern social network analysis date from the 1930s and the pioneering work of Jacob Moreno and his many collaborators. Researchers at New York University, Columbia, and Harvard created the first scholarly works featuring the distinctive core components of modern social network theory: measures, maps, and models. Moreno and his research partners created the first pictures of patterns of people and their partners, using visual maps with symbols that represented individuals with different types of lines connecting them to others that represented different kinds of relationships.

Moreno documented relationships among schoolchildren and the way an innovative behavior, running away, moved through chains of student connections. In 1934, Moreno [8] published "Who shall survive," which catalyzed work among a group of scholars who refined his approach and added critical mathematical elements that today are a standard part of network analysis. These approaches were applied to various settings, and revealed the key roles a few people played in their networks and often the presence of subgroups of distinct people. For example, in the 1930s, Davis et al. collected detailed records of observed attendance at 14 social events by 18 southern women, and the graph of that data revealed two distinct groups with minimal overlap [9]. Moreno, who developed sociometry and is often considered the founder of the sociogram, studied relationships among members of a football team and found patterns of friendship and animosity (see Fig 3.3) (as produced in Freeman [10]).

At Harvard in the 1930s, a group formed around W. Lloyd Warner and Elton Mayo to explore interpersonal relationship in workplaces. Early social network analysis work focused on connections in small work groups in industrial factory settings. For example, Roethlisberger and Dickson [11] studied the Western Electric Wiring room, documenting the ways individuals within a group worked with one another. As seen in Figure 3.4, some workers in the study emerged as the most connected, whereas others appeared as peripheral or isolated. Another data set was created that represented the relationships among 14 employees of the Western Electric Hawthorne Plant. Employees and two inspectors were observed, and each contact among them was coded. When employees played games with one another, argued, were openly friendly, confrontational, or helpful a note and tie was recorded. The result were six networks, which led to a seminal work by the Harvard sociologist George Homans [12] and later more mathematical work that focused on automatically finding clusters or groups within these data

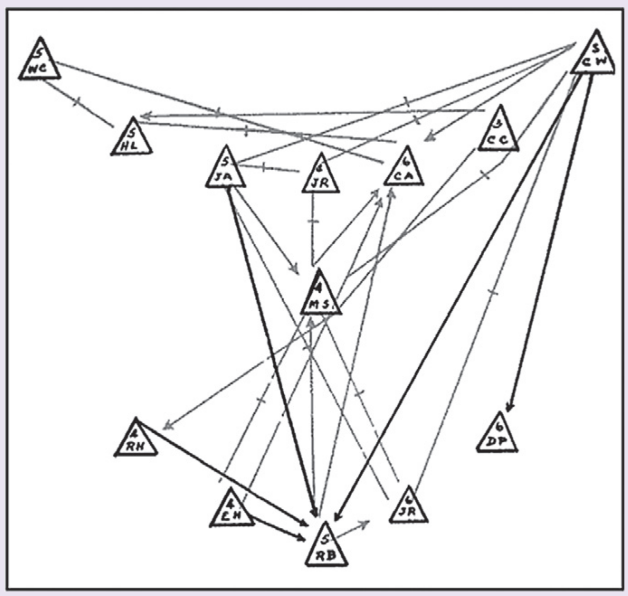

FIGURE 3.3 Jacob Moreno's early social network diagram of positive and negative relationships among members of a football team. Originally published in Moreno, J. L. (1934). *Who shall survive?* Washington, DC: Nervous and Mental Disease Publishing Company.

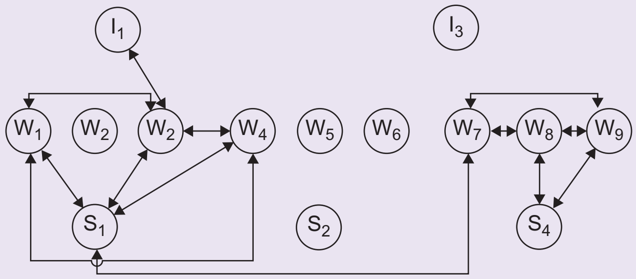

FIGURE 3.4 An early social network diagram of relationships among workers in a factory illustrates the positions different workers occupy within the workgroup. Originally published in Roethlisberger, F., and Dickson, W. (1939). *Management and the worker.* Cambridge, UK: Cambridge University Press.

sets [13]. In the 1950s, Nadel wrote about social roles and the social structures that define them [14]. He saw that the patterns of connections people had might be similar, even if they were connected to different people. These patterns, Nadel suggested, could be studied systematically, but in the 1950s the data and computational resources made that ambition a challenge.

Over time, Moreno's colleagues, including Paul Lazarsfeld, added key ingredients of the modern form of social network analysis: metrics and algorithms for calculating important network properties of the graph as a whole and for each individual in the graph (see Freeman [10] for details).

3.5 NETWORK ANALYSIS METRICS

Social scientists, physicists, computer scientists, and mathematicians have collaborated to create theories and algorithms for calculating novel measurements of social networks and the people and things that populate them. These quantitative *network metrics* allow analysts to systematically dissect the social world, creating a basis on which to compare networks, track changes in a network over time, and determine the relative position of individuals and clusters within a network.

Social network measures initially focused on simple counts of connections and became more sophisticated as concepts of density, centrality, structural holes, balance, and transitivity developed. Some metrics describe a network as a whole. For example, network density captures how highly connected vertices are by calculating the percentage of all possible connections that are realized. Other metrics are calculated for each vertex in a network. For example, centrality measures, of which there are many, capture how "important" (central) a vertex is within the network based on some objective criteria. Some people sit at the edge or periphery of their networks, whereas others are firmly at the center, connected to all the other most connected people. Even among a highly connected network, some pairs are not directly connected. When a third person bridges their connection, we can think of that person as a broker or connector. When that person is missing, we can think of a structural hole, a gap in which there is a missing connector. The following sections describe some of these metrics in more detail. Chapter 5 introduces some of the core metrics found in NodeXL through hands-on exercises.

ADVANCED TOPIC
Historical Obstacles to the Development of Network Analysis

Following the rapid development of the major elements of social network analysis in the 1930s there was a period of stagnation and neglect. For a variety of reasons, from Moreno's own personal and professional conflicts to the cost and lack of available network data sets and computing resources, social network analysis languished for decades.

The early social network literature was built on manually collected and processed data about social ties. Researchers would typically observe or survey population members, asking each to list those they came in contact with regularly for a variety of tasks and purposes. The prohibitive cost of this approach was a major limiting factor in the widespread application of social network analysis in enterprises and organizations. The recent explosion of computer-mediated social relationships and the associated drop in the costs of creating network data sets have made network approaches increasingly practical. As more details about our interactions and associations are tracked and captured by mobile devices and social media services, network analysis becomes increasingly useful.

Network analysis is computationally intensive: to generate many network metrics can require millions of calculations even when managing modest sized data sets. The recent explosion of computing power and the associated drop in costs have made network approaches increasingly practical, even if network methods remain among the most computationally intensive in use.

3.5.1 Aggregate Networks Metrics

A number of metrics describe entire networks. In some cases, a single network is broken into several disconnected pieces, called *components*. Some aggregate network metrics only work on networks where all of the vertices are connected in a single component, whereas others can be applied to entire networks even if they are split up. Here we describe just a few aggregate network metrics to give a flavor for what is possible, leaving a fuller discussion for Chapter 5.

Density is an aggregate network metric used to describe the level of interconnectedness of the vertices. Density is a count of the number of relationships observed to be present in a network divided by the total number of possible relationships that could be present. It is a quantitative way to capture important sociological ideas like cohesion, solidarity, and membership.

Centralization is an aggregate metric that characterizes the amount to which the network is centered on one or a few important nodes. Centralized networks have many edges that emanate from a few important vertices, whereas decentralized networks have little variation between the numbers of edges each vertex possesses.

Other metrics integrate attribute data with network data. For example, metrics that measure homophily look at the similarity of people who are connected. Studies typically show that people are connected to others who are similar to themselves on core attributes like income, education level, religious affiliation, and age.

3.5.2 Vertex-Specific Networks Metrics

Another set of metrics identifies individuals' positions within a network. Paramount among these is the set of centrality measures, which describe how a particular vertex can be said to be in the "middle" of a network. In the 1970s and 1980s, the sociologist Philip Bonacich developed a refined measure of centrality that took into consideration the different value a well-connected person can have in contrast to people with few connections. Network theorists noted that simply having many connections, called "degree centrality," was only one way to be "at the center" of things. A person with fewer connections might have more "important" connections than someone with more connections. One connection can be more important than another in different ways. Some are better because they bridge across otherwise separated portions of the network, whereas others are important because they connect to well-connected people. The following centrality metrics provide quantifiable measures for these concepts (see Chapter 5 for more details).

Degree Centrality

Degree centrality is a simple count of the total number of connections linked to a vertex. It can be thought of as a kind of popularity measure, but a crude one that does not recognize a difference between quantity and quality. Degree centrality does not differentiate between a link to the president of the United States and a link to a high school dropout. Degree is the measure of the total number of edges connected to a particular vertex. For directed networks, there are two measures of degree. In-degree is the number of connections that point inward at a vertex. Out-degree is the number of connections that originate at a vertex and point outward to other vertices.

Betweenness Centralities: Bridge Scores for Boundary Spanners

The notion of paths is central to the study of networks. Perhaps one of the most natural questions to ask about any two people in a network it is "How far apart are they?" This distance is measured simply: the distance between people who are not neighbors is measured by the smallest number of neighbor-to-neighbor hops from one to the other. For instance, people who are not your neighbors, but are your neighbors' neighbors, are a distance 2 from you, and so on. The shortest path between two people is called the "geodesic distance" and is used in many centrality metrics. For example, betweenness centrality is a measure of how often a given vertex lies on the shortest path between two other vertices. This can be thought of as a kind of "bridge" score, a measure of how much removing a person would disrupt the connections between other people in the network. The idea of brokering is often captured in the measure of betweenness centrality.

A structural hole is a missing bridge. Wherever two or more groups fail to connect, one can argue that there is a structural hole, a missing gap waiting to be filled. Burt provides compelling evidence that individuals who bridge structural holes are promoted faster than others [15]. Social network analysis has many strategic applications when people in an organization can analyze their position and the position of others. Managers and leaders can recognize gaps or disconnections within organizations and devote resources to traversing the divide. People may be able to apply social network analysis to identify locations in which a gap exists and elect to fill them, recognizing the value they can generate as broker between two otherwise separate groups.

Closeness Centrality: Distance Scores for Broadly Connected People

Closeness centrality takes a different perspective from the other network metrics, capturing the average distance between a vertex and every other vertex in the network. Assuming that vertices can only pass messages to or influence their existing connections, a low closeness centrality means that a person is directly connected or "just a hop away" from most others in the network. In contrast, vertices in very peripheral locations may have high closeness centrality scores, indicating the number of hops or connections they need to take to connect to distant others in the network. Think of closeness, paradoxically, as a "distance" score. Some people are just a few miles from the big city, others must drive for hours: similarly, people with high "closeness" centrality scores have many miles or rather personal connections that they must travel to reach many other people in the network. Note that in some cases the inverse of the average distance to others in the network is used as a measure of closeness centrality. In that case, higher values indicate a more central position.

Eigenvector Centrality: Influence Scores for Strategically Connected People

Eigenvector centrality is a more sophisticated view of centrality: a person with few connections could have a very high eigenvector centrality if those few connections were themselves very well connected. Eigenvector centrality allows for connections to have a variable value, so that connecting to some vertices has more benefit than connecting to others. The PageRank algorithm used by Google's search engine is a variant of Eigenvector Centrality.

CLUSTERING COEFFICIENT: HOW CONNECTED ARE MY FRIENDS?

The clustering coefficient differs from measures of centrality. It is more akin to the aggregate density metric, but focused on egocentric networks. Specifically, the clustering coefficient is a measure of the density of a 1.5-degree egocentric network. When these connections are dense, the clustering coefficient is high. If your "friends" (alters) all know each other, you have a high clustering coefficient. If your "friends" (alters) don't know each other, then you have a low clustering coefficient. People have different measures for their clustering coefficient depending on the ways they cultivate connections to others and the environments they are in.

3.5.3 Clustering and Community Detection Algorithms

A network approach contrasts with those that presume the existence and boundaries of groups. In a network perspective, people occupy many relationships and are potentially members in many groups and less defined clusters. Defining exact boundaries in networks may be difficult, reflecting the reality of multiple and shifting memberships. From a network perspective, a group is a collection of vertices that are more connected to one another than they are to others. Relatively more cohesive or densely connected sets of vertices form regions, also called clusters, that may reflect the existence of groups without regard to whether they are officially recognized or even if members recognize their connections to one another. A rapidly growing body of research describes clustering algorithms, also called community detection algorithms, that automatically identify these clusters based on networks structures. We discuss these in more detail in Chapter 7.

3.5.4 Structures, Network Motifs, and Social Roles

Two people within a network may sometimes share a pattern of connection to other people, even if they do not connect to the same people. Certain professions have distinct patterns of connections, either linking with many others (real estate agents, and other retail professionals) or few (reclusive authors and artists, peripheral workers, and other people focused on things rather than people). In addition to the number of connections, some people share the pattern of connections among the people they connect. In some cases people are connected to people who are strangers to one another, in other cases a group may be densely connected to one another. These secondary patterns of connection are a distinctive feature of network analysis approaches: networks are as much about the attributes and patterns of connection among neighbors as they are about the attributes and connections of any individual.

Social roles are complex cultural and structural features of social life. An example social role like "father" is explicitly recognized in society, has a wide set of culturally shared meanings and expectations, is associated with particular goals and interests, and is partly defined by the content and structure of actions directed toward other distinctive role holders. Although social roles may not be as clearly defined or explicitly recognized by all the actors in a given social setting, they have identifiable content, behavioral, and structural features.

Studies of social media have illustrated the ways contributors create distinctive network patterns that reflect their role or status within the community (e.g., Welser, Gleave, and Smith [16]). These patterns are evidence of specialization of behavior in these social spaces. An example of a role in a social media space is the "answer person" who disproportionately provides the answers to questions asked in message board environments, "discussion people" who engage in extended exchanges of messages in large and populous threaded discussions, "discussion starters" who demonstrate influence over the topics discussed by the "discussion people," "influential" people who are well connected to others who are more highly connected than they are, and boundary spanners who bridge between unconnected subgroups. These roles are described in greater detail in Chapter 9, devoted to email lists and discussion groups.

ADVANCED TOPIC

A Renaissance of Network Research and Data

Since the 1960s, network analysis has blossomed. New research and methods have flourished and social networking has developed a new prominence in mainstream culture. Despite early challenges, in the past several decades a healthy and growing subfield has reemerged around social network analysis. New network tools and concepts have been created and applied to a wide and growing range of domains. Mathematical sociology has developed as a major subdiscipline in the social sciences, dedicated to finding elegant descriptions of complex social phenomena. Starting 80 years ago with simple hand-drawn charts and diagrams that described small groups of people and their connections, network science concepts, methods, and tools are used today to calculate a range of measures that describe the shape, structure, and dynamics of potentially multimillion or billion vertex networks. New methods have been developed for automatically organizing and displaying visualizations of the links among large populations. This combination of structural models, visualizations, and metrics forms the key features of modern social network analysis.

In the late 1960s, Stanley Milgram explored the idea of small world networks in a study that came to be referred to as "Six Degrees of Separation" [17], which later inspired the 1990 John Guare play and 1993 movie of the same name. The study explored the question of how connected any two people selected at random might be. Milgram sent a collection of letters to people around the country asking them to send the message to someone they knew who could move their letter closer to the target, a stock broker in Massachusetts. On average, the letters took six steps to arrive at their destination. The "six degrees" or steps suggested that even in large networks where most people are not directly connected, people can be reached from every other person through a small number of steps.

Sampson's study in the late 1960s of relationships among members of a residential monastery captured social network data during an event in which several members were expelled or chose to leave [18]. A series of social network data sets were collected by asking participants about who they liked and spent time with. Social network analysis of this data allowed Sampson to identify the future lines of division among the members of the network. The idea that members of a network can be grouped based on how densely they are connected is an important concept in network analysis. These groups can be important divisions with consequences for the future of the network. For example, a notable study by Zachary in the 1970s mapped the structure of a Karate club based on affinities and connections between students and teachers. These maps predicted the ways the club eventually split when a new teacher, in conflict with the owner, left the studio and took many students with him [19].

The sociologist Barry Wellman demonstrated in the 1970s that real-world communities are composed of interlocking social networks of specialized relationships that changed dramatically in composition over a period of years.[2] He proposed that society was now characterized by networked individualism in contrast to the group memberships and identities of prior periods. Rather than defining oneself in professional or political terms, people create personal networks in which they occupy distinct locations and roles. He later applied these techniques to study online networks [20]. In 1977, Wellman founded a social network analysis professional association, the International Network for Social Network Analysis (INSNA). INSNA now has more than a thousand members, many of whom have gathered for more than 20 years for an annual conference ("Sunbelt") on social network analysis research.[3] Journals and publications devoted to social network analysis include *Social Networks*, *Connections*, and the *Journal of Social Structure*. Social network data, methods, and visualizations appear across a much wider spectrum of journals and conference publications.

[2] www.chass.utoronto.ca/~wellman/vita/index.html
[3] www.insna.org

In the early 1970s, the sociologist Mark Granovetter did research on the employment market, looking at how people discovered new job opportunities. He observed that, in contrast to the view held by classical economics, people were not freely floating independent actors in the labor market. They were embedded in a set of different relationships with particular people. Granovetter found that job news passed through connections that were not the closest and most intense relationships [21]. A person's "weak ties" brought news from distant parts of the social network to which "strong ties" did not have access because they occupied such a similar place in the network as the job seeker. Thus weak ties proved particularly useful for finding novel information, such as information about job prospects. Because weak ties were less intense, they were also less costly to maintain in terms of time and attention. As a result, it is possible to have many weak ties but only a few strong ties.

Armed with new network metrics and the means to calculate them, network analysts have focused on a variety of data sources and questions. Social networks have been applied to historical studies using records of investments, marriages, and memberships in elected positions. In the 1400s in the city of Florence, the Medici and Strozzi families struggled for domination. These families, like many others, were locked in political struggles. In the 1970s, John Padgett collected records of the social relations among Renaissance Florentine families that he extracted from historical documents. Families were often connected through a variety of ties, relations, and business connections. A data set was created that represented the financial loans, credits and joint partnerships, and marriages that bound families to one another. The resulting data set included information about each family as well as their links to others. Each family had a value representing its net wealth in the year 1427, the number of seats it held in the local government between the years 1282 and 1344, and the number of business or marriage ties among the population of 116 families. Analyzing these data, Padget found that the Medici held great power because, he argued, they sat at the center of business and family networks, brokering connections that no other family could equal [22, 23].

A more modern version of the study of historical Florentine politics can be found in the study of interlocking directorships in modern corporations. Many corporations and other institutions have a board of directors, some of whom serve on more than one board. When board members serve on two or more boards, they link those corporations and, in aggregate, create interlocking directorships that combine to form even larger meta-institutions. By building on research on interlocking directorships in U.S. corporations [24, 25], modern web sites like "They Rule" provide an interactive map that displays the common links between major corporations.[4]

In 1992, Robin Dunbar famously argued that people have an innate ability to handle a number of social relationships but not an endless number of them. Remembering people's names may have a biological limit as our brains evolved over long periods in which there were rarely more than a few hundred people within any region, group, or tribe. The number 150 has been loosely associated with the idea of a "Dunbar" number, an upper limit on the number of relationships a person can normally manage.[5] This number can be expanded with augmentation, through analog technologies like diaries, address books, and the "filo-fax." More recently, social media tools like Facebook and email contact lists extend our ability to maintain more relationships. These additional relationships can be said to be "weaker" than the core 150 "organic" relationships, but as Granovetter has shown, weak ties can collectively be of enormous value.

Business Applications of Social Network Analysis

Social network analysis has historically been an academic endeavor, but as network analysis tools and data sets become more available, pioneering businesses are applying it to help manage business challenges, gain insight into markets and communities, and build more robust industry relationships. For example, the work of Rob Cross and the Network Roundtable focuses on several practical applications of social network analysis for corporations and other large organizations, highlighting differences between healthy and underperforming divisions and the value of organization spanning connections [26, 27]. Others apply network analysis to the improvement of corporate structures and processes [28]. In the early 1990s, Monge and Contractor [29] documented the many forms of social network patterns that emerge inside of organizations and institutions.

Social networks have been shown to have a significant influence on the adoption of new technologies or social practices. The sociologist Everett Rogers described the concept of the "diffusion of innovations," arguing that

[4] www.theyrule.net
[5] www.lifewithalacrity.com/2004/03/the_dunbar_numb.html

(*Continued*)

ADVANCED TOPIC (*Continued*)

people with particular patterns of connections to others played pivotal roles in the success or failure of a new idea or message being adopted or distributed through the network [30]. Networks with different patterns of connection have different properties in terms of how they propagate a new message, rumor, or product and how they resist being dissolved when vertices are removed from the graph. These observations have significant implications for interventions into disease and rumor propagation and the cultivation of innovation [31].

Networks play an important role in e-commerce where collaborative filtering powers the familiar list of "books that people who liked this book also liked." Businesses are also interested in learning the requirements of viral marketing. We will discuss diffusion and marketing in more detail in the discussion of Twitter in Chapter 10, but for now the important thing to know is that diffusion can often lead to "cascades" where an unknown, even marginal idea can spread rapidly throughout the entire network and become the norm. Memes are a commonly-cited example of contagion, as are viral products, such as viral videos on YouTube that go from dozens to millions of viewers in a few months or even weeks.

3.6 SOCIAL NETWORKS IN THE ERA OF ABUNDANT COMPUTATION

The widespread adoption of networked communication technologies has significantly expanded the population of people who are both aware of network concepts and interested in network data. Although the idea of networks of connections of people spanning societies and nations was once esoteric, today many people actively manage an explicit social network of Internet friends, contacts, buddies, associates, and addresses that compose their family, social, professional, and civic lives. Email messages forwarded from person to person have become a common and visible example of the ways information passes through networks of connected people. The notion of "friends of friends" is now easy to illustrate in the features of Internet social media applications like Facebook, MySpace, and LinkedIn that provide explicitly named "social networking" services. Viral videos and chain emails illustrate the way word of mouth has moved into computer-mediated communication channels. The idea of "six degrees of separation" has moved from the offices of Harvard sociologists to become the dramatic premise of a Broadway play to now appear as an expected feature of services that allow people to browse and connect to their friend's friends.

As network concepts have entered everyday life, the previously less visible ties and connections that have always woven people together into relationships, cliques, clusters, groups, teams, partnerships, clans, tribes, coalitions, companies, institutions, organizations, nations, and populations have become more apparent. Patterns of sharing information, investments, personal time, and attention have always generated network structures, but only recently have these linkages been made plainly visible to a broad population. In the past few decades, the network approach to thinking about the world has expanded beyond the core population of researchers, analysts, and practitioners who have applied social network methods and perspectives to understand their businesses, communities, markets, and disciplines. Today, because many of us manage many aspects of our social relationships through a visibly computer-networked social world, it is useful for many more people to develop a language and literacy in the ways networks can be described, analyzed, and visualized. Visualizing and analyzing a social network is an increasingly common personal or business interest. The science of networks is a growing topic of interest and attention, with a growing number of courses for graduates and undergraduates and even becoming the topic of a television documentary.[6]

The availability of cheaper computing resources and network data sets has enabled a new generation of researchers access to studies of the structures of social relationships at vastly larger scale and detail. Since the late 1960s, as computing resources and network data sets have grown in availability and dropped in cost, researchers began developing tools and concepts that enabled a wider and more sophisticated application of social network analysis.

[6] "Connected: The Power of Six Degrees," http://ivl.slis.indiana.edu/km/movies/2008-talas-connected.mov

ADVANCED TOPIC

Social Network Analysis Research Meets the Web

As access to electronic networks grew in the 1970s, academic and professional discussions and collaborations began to take place through them. Systems to support the exchange of messages and the growth of discussions and even decisions became a major focus of systems development and the focus of study itself. Freeman and Freeman [32] collected data from the records of the Electronic Information Exchange System (EIES) that itself hosted a discussion among social network researchers. Two relations were recorded: the number of messages sent and acquaintanceship. These systems became the focus of the first systematic research into naturally occurring social media. Even before the Internet, early computer network applications supported the creation of exchanges, discussions, and therefore social networks, built by reply connections among authors.

Early proprietary systems evolved into the public World Wide Web. In the 1990s, the computer scientist Jon Kleinberg identified the patterns of links between high-quality web pages, an algorithm that went on to inspire Stanford graduate students who founded the Google corporation. Kleinberg described different locations in a population of linked documents on the World Wide Web. On the World Wide Web, a document or page can link to another page, forming a complex network of related documents. Some documents contain many pointers to other documents, whereas others have many documents point at them. These hubs and authorities defined two broad classes of web pages that offered a path to identifying high-quality content. Links from one page to another are considered to be indicators of value. Refinements of the HITS algorithm made use of eigenvector centralities to implement the page rank algorithm that is the core of the Google web search ranking method [33].

Network researchers studying social networks and the Internet found that empirical networks often exhibit "small-world properties": most nodes are not neighbors with each other, but nodes can be reached from every other node in a small number of hops. In the late 1990s, the physicist/sociologist Duncan Watts, working with the mathematician Steven Strogatz, created mathematical models of "small world" networks and contrasted them with purely random networks such as those proposed by Erdos and Renyi [34]. Their model captured the natural properties of social networks far better than those that assumed a purely random or normal distribution of links. Although most connections are to others who are local, a few connections importantly can jump far from the individual. Many of our friends are likely to live or work near us, but a few may be very far away. These relatively rare far-reaching links can dramatically change the properties of the network, making the widespread transmission of messages much easier. This model significantly improved on earlier models of network growth and structure, better approximating the observed structure of naturally occurring social networks. Later researchers have built upon their work to devise models that generate "small world" networks that more closely match empirical networks, helping us to understand how networks may have become the way they are. For example, Barabasi and Albert have developed a family of models of preferential attachment that can generate "scale-free" networks, which are a common feature of social networks [35]. Scale free networks have a power law degree distribution, meaning that there are a few key hubs in a network and many poorly connected vertices. While none of these models perfectly predict social networks, they provide a method for systematically comparing networks and focus attention on the processes that may have led to the characteristics that we see in networks.

In the past few years, researchers have begun to study large web-based networks. For example, Leskovec and Horvitz calculated metrics for a graph that includes more than 300 million users of the Microsoft Messenger service [36]. Each user typically had one or more "buddies" with whom he or she might send one or more messages and receive some in return. Buddies often listed their locations, allowing these linkages to be aggregated into a complex map of the world and the flow of conversation around it. Others have reported on the hyperlink network created by web pages hyperlinking to other web pages (e.g., Park and Thelwall [37]). A number of studies have examined the blog network. For example, Adamic and Adar [38] showed how political blogs are divided into two clear clusters with minimal overlap that represent the left and right political populations in the United States. More recently, Kelly and Etling mapped Iran's blogosphere, identifying more than 20 subcommunities of bloggers who wrote in Farsi for an Iranian audience.[7]

Another line of research has focused on visualizing social networks. A representative paper by Heer and Boyd [39] described a tool called Vizster that allowed users to navigate through their friends from a social networking site to explore social connections.

[7]http://cyber.law.harvard.edu/publications/2008/Mapping_Irans_Online_Public

3.7 THE ERA OF ABUNDANT SOCIAL NETWORKS: FROM THE DESKTOP TO YOUR POCKET

We now live in a new era of network data abundance. Network data collection was once a time-consuming and laborious process that yielded small data sets at great cost. Observations, surveys and interviews took many days or weeks to perform, could not be repeated frequently, required many people to produce, and often yielded low rates of participation with inherent biases and errors. Asking people about their relationships with others continues to have benefits and offers unique sources of insight, but people have been shown to be a poor source of accurate information as bias and faulty memory warp what people report about who they know and with whom they interact. The challenge of creating a data set that spanned long periods or large numbers of people or contained records of many events proved insurmountable using traditional methods.

Today, interactions between people increasingly take place through computing systems. Users create many types of networks in a machine-readable form each day as our interactions are documented in a computer. When we use these communication tools, databases are created and maintained with records and log files that document the details of the time, place, and participants of each interaction, whether via computers or telephones and even televisions. These event logs describe many different kinds of connection but share a common structure in which one person or entity is linked to another by some relationship.

The creation of these machine-readable network data sets mean that long periods of time or large populations connected by many events can now be studied using widely available computing equipment and data sources.

Like a jump from Galileo's handmade telescope to the orbiting *Hubble*, network science has made a vast leap in scale and scope as we create a digitally networked world around ourselves.

As the historical drought of social network data has ended with a flood of sources of network data, the challenge has been to rapidly develop the tools and concepts needed to process and analyze them. Technical methods for building multiterabyte databases have shifted to the even vaster task of managing petabytes of data. New methods of harnessing thousands and even millions of computers in parallel have been driven by the growing need to manage vast data stores growing from the web. The challenge is likely to grow steeper as new sources of network data come pouring out off an emerging class of sensor-rich devices that record vast streams of data from millions to billions of people, devices, and locations. The early wave of this surge of data can be seen in new sources of data from everyday life that are being captured and recorded with mobile devices, creating a new stream of archival material that is richer than all but the most obsessively observed biographies. It has become common in recent years that the most timely and well-placed photographs and video recordings have come from everyday individuals with phones and computers rather than from news photographers and reporters.

The coming wave of mobile technologies is likely to deepen this trend, with new ways for phones or other devices to capture information about their users and the world around them. Research projects like SenseCam from Microsoft Research, which captures a continuous stream of photographs and temperature and motion data, are now becoming products,[8] and services like nTag, Spotme, Loopt, Foursquare, and Google Latitude using devices like iPhones and Droids are weaving location into every application (see Chapter 2).

As phones are increasingly able to notice each other, a new set of mobile social software applications are becoming possible, as evidenced by new services such as Bump.[9]

A service like SenseNetworks[10] is a good example of a mobile data collection, analysis, and presentation service. Other services and products like CureTogether[11] and FitBit[12] are examples of social movements that are enabled by web applications integrated with devices that provide self-monitoring medical tracking. These communities overlap with the trail-based exercise communities of runners, bikers, skaters, hikers, and skiers, some with artistic inclinations (they hike in paths that resemble drawings when seen on a map). A new wave of devices is emerging that extensively quantifies your "self" and "others," allowing you to perhaps swap sensor data with other people. The result could be an aggregated map of the heath and environmental conditions of the planet, not unlike early examples of collectively authored road maps of whole nations accomplished by the Open Street Map project.[13]

[8] www.nytimes.com/2010/03/09/health/09memory.html
[9] http://bu.mp
[10] www.sensenetworks.com
[11] www.curetogether.com
[12] www.fitbit.com
[13] www.openstreetmap.org/

3.8 TOOLS FOR NETWORK ANALYSIS

The growth of interest in network analysis has been dramatic, but until recently the development of social network analysis tools has lagged, and they remained a challenge to use for many people. Applying network approaches has been traditionally a challenge that involved much more than simply mastering a new set of concepts and ideas that focus on relationships and patterns. Network data have traditionally been difficult to create and collect, and the tools for analyzing and visualizing networks have demanded significant technical skill and often mastery of programming languages. Many tools that exist to support network analysis demand significant commitment to learn and master. Existing network tools that are relatively easier to use have typically lacked support for easily importing social media network data. In the past few years, many network analysis projects and research papers have focused on computer-mediated networks of people, documents, and systems. Only recently have new tools made it simpler for people to extract data from major social media network sources and to perform a basic network analysis workflow without requiring programming skills or using a command line interface.

Social media network data collection, scrubbing, analysis, and display tasks have historically required a remarkable collection of tools and skills. A great example of the variety of tools that can be used in concert to extract, analyze, and display social media networks can be found on Drew Conway's blog.[14] This is a powerful set of tools for those who can master the demands of python or other programming languages and the application programmer interfaces (API) that give sophisticated users special access to resources. In contrast, this book focuses on a single tool designed for nonprogrammers, NodeXL, because of its relative ease of use, support for rich visuals and analytics, and integration with the ubiquitous Excel spreadsheet software. The python path is certainly the high road for experts and those with demanding volumes of data or esoteric data analysis requirements. But for the noncoding user, NodeXL may be one of the easiest ways to both manipulate network graphs and get graphs from a variety of social media sources. A detailed step-by-step guide to the core features of NodeXL can be found in Part II of the book.

3.9 NODE-LINK DIAGRAMS: VISUALLY MAPPING SOCIAL NETWORKS

One of the key elements that characterizes modern social network analysis is the use of visualizations of complex networks. Compared to staring at edge lists or network matrices (see Section 3.2.4), looking at a network graph can provide an overview of the structure of the network, calling out cliques, clusters, communities, and key participants. It could be said that a graph is worth a thousand ties. Not only can network visualizations inspire understanding and insights, they can also be appealing and even beautiful. They can serve as persuasive tools that demonstrate important points about networks. The ability to map attribute data and network metric scores to visual properties of the vertices and edges (see Chapters 4 and 5) makes them particularly powerful.

However, network visualizations are often as frustrating as they are appealing. Network graphs can rapidly get too dense and large to make out any meaningful patterns as illustrated in Figure 3.5. Many obstacles like vertex occlusions and edge crossings make creating well-organized and readable network graphs challenging. There is an upper limit on the numbers of vertices and edges that can be displayed in a bounded set of pixels; typically only a few hundred or thousand vertices can be meaningfully and distinctly represented on average-sized computer screens. In his appeal for better-quality network visualization, Shneiderman [40] has suggested that we aspire to reach the worthy but not always attainable goal of "netviz nirvana" in which the following goals are proposed:

- Every vertex is visible.
- Every vertex's degree is countable (i.e., the number of connections that start or end at that vertex).
- Every edge can be followed from source to destination.
- Clusters and outliers are identifiable.

To approach netviz nirvana, careful preparation, layout, and filtering techniques must be used. In practice, network visualizations often fall far from the mark. However, the graphs shown throughout this book illustrate the value of carefully crafting network graphs. We hope they will inspire network analysts to take the care needed to create substantive, understandable, and aesthetically pleasing graphs.

3.10 COMMON NETWORK ANALYSIS QUESTIONS APPLIED TO SOCIAL MEDIA

Once a set of social media networks has been constructed and social network measurements have been calculated, the resulting data set can be used for many applications. For example, network data sets can be used to create reports about community health, comparisons of subgroups, and identification of important

[14] www.drewconway.com/zia/?p=204

individuals, as well as in applications that rank, sort, compare, and search for content and experts.

The value of a social network approach is the ability to ask and answer questions that are not available to other methods. This means focusing on relationships. Although analysts, marketers, and administrators often track social media participation statistics, they rarely consider relationships. Traditional participation statistics can provide important insights about the engagement of a community, but can say little about the connections between community members. Network analysis can help explain important social phenomena such as group formation, group cohesion, social roles, personal influence, and overall community health. Combining traditional participation metrics with network metrics provides the best of both worlds and allows you to answer important questions such as the following:

- What kinds of social roles are being performed within a social media repository? Does a community have enough people filling the important roles?
- Which individuals play important social roles within a group or collection? Who would make a good administrator based on that person's network position?
- What subgroups exist? Do connections between subgroups exist? Who plays the bridge roles that connect otherwise unconnected groups?
- How do new ideas propagate through a network? Who are the influencers?
- How do the overall structures of a social network change after a particular event (e.g., a company social, a round of new hires or layoffs)?

3.11 PRACTITIONER'S SUMMARY

The opportunities for practitioners to apply network analysis to contemporary business, community management, political influence, and team collaboration have dramatically increased in recent years. The once esoteric

FIGURE 3.5 A medium-sized node-link network diagram visualization of Twitter users linked by patterns of following. This sized graph illustrates many issues with a network graph containing more than a few dozen vertices. Many vertices sit on or overlap with other vertices. The number of edges associated with some vertices is impossible to count, whereas other edges cannot be traced from source to destination. Improvements to network layout are an active area of research.

concepts and metrics of network analysis have become talk show and airport lounge topics. The difficulties in collecting and analyzing network data have been dramatically reduced by powerful database methods and well-designed network analysis and visualization tools. There is still a lot of work to be done, but practitioners now have the potential to make more effective decisions based on network analyses of their own data conducted in a few hours, rather than a few months.

Learning the concepts and tools is a necessary first step, but the payoffs are large. The growing industry of social media and networking consultants complemented by a vast array of books plus informative web sites, online seminars, and Wikipedia pages, makes the necessary training widely available. At the same time, network analysis methods are spreading through university curricula rapidly and filtering into high school courses.

Attending public seminars and professional conferences provides other means to acquire skills and make valuable connections. Your first steps may be a struggle, but we hope that with each step the processes become smoother and the professional benefits larger.

3.12 RESEARCHER'S AGENDA

The research progress on network analysis has been dramatic in the past few decades, transforming an exotic research topic into a thriving research community in academia, government, and industry. The existing metrics, clustering, and layout algorithms are stabilizing, but innovative approaches are still emerging to trigger bursts of new research. As practitioner pressure builds to apply network analysis to ever larger data sets, researchers have developed remarkably more efficient algorithms, while hardware developers have produced powerful graphics processors (based on gaming computers), huge arrays of computers, and scalable cloud computing services. Meanwhile, new social media services generate more relational data than ever before, ushering in a golden era of social science research on human relationships and collaboration.

The algorithms and hardware provide the platforms, but the concomitant development of vastly improved user interfaces for network analysis has begun to enlarge the community of users from the dedicated sociologists who are also programmers to the broad segment of business analysts who use spreadsheets or simplified web-based tools. Packaging the complex processes of frequently applied network analyses into a few clicks is the next challenge in many fields, thereby inspiring other researchers and developers to simplify the processes even further, while increasing the power offered to users. The best is yet to come.

References and Resources

[1] J.P. Scott, Social Network Analysis: A Handbook, Sage, Thousand Oaks, CA, 2000.
[2] D. Easley, J. Kleinberg, Networks, Crowds, and Markets: Reasoning about a Highly Connected World, Cambridge University Press, Cambridge, UK, 2010.
[3] M. Newman, A.-L. Barabasi, D.J. Watts (Eds.), The Structure and Dynamics of Networks, Princeton University Press, Princeton, NJ, 2006.
[4] P. Carrington, J. Scott, S. Wasserman (Eds.), Models and Methods in Social Network Analysis (Structural Analysis in the Social Sciences), Cambridge University Press, Cambridge, UK, 2005.
[5] W. de Nooy, A. Mrvar, V. Batageli, Exploratory Social Network Analysis with Pajek (Structural Analysis in the Social Sciences), Cambridge University Press, Cambridge, UK, 2005.
[6] S. Wasserman, K. Faust, Social Network Analysis: Methods and Applications (Structural Analysis in the Social Sciences), Cambridge University Press, Cambridge, UK, 1994.
[7] B. Wellman, Structural analysis, in: B. Wellman, S.D. Berkowitz (Eds.), Social Structures, Cambridge University Press, Cambridge, UK, 1988, pp. 19–61.
[8] J.L. Moreno, Who shall survive? A new approach to the problem of human interrelations, Nervous and Mental Disease Publishing Co., Washington, 1934.
[9] A. Davis, B.B. Gardner, M.R. Gardner, Deep South: A social Anthropological Study of Caste and Class, University of Chicago Press, Chicago, Ill., 1941.
[10] L.C. Freeman, The Development of Social Network Analysis: A Study in the Sociology of Science, BookSurge, LLC, North Charleston, SC, 2004.
[11] F. Roethlisberger, W. Dickson, Management and the Worker, Cambridge University Press, Cambridge, UK, 1939.
[12] G. Homans, The Human Group, Harcourt-Brace, New York, 1950.
[13] R. Breiger, S. Boorman, P. Arabie, An algorithm for clustering relational data with applications to social network analysis and comparison with multidimensional scaling, J Math Psychol 12 (1975) 328–383.
[14] S.F. Nadel, The Theory of Social Structure, Cohen & West, London, 1957.
[15] R. Burt, Brokerage and Closure: An Introduction to Social Capital, Oxford University Press, Oxford, 2005.
[16] H. Welser, E. Gleave, M. Smith, Visualizing the signatures of social roles in online discussion groups, J. Soc. Struct. 8 (2) (2007).
[17] S. Milgram, The small world problem, Psychology Today 2 (1967) 60–67.
[18] S. Sampson, Crisis in a cloister. Unpublished doctoral dissertation, Cornell University, 1969.
[19] W. Zachary, An information flow model for conflict and fission in small groups, J Anthropol Res 33 (1977) 452–473.
[20] B. Wellman, An electronic group is virtually a social network, in: Kiesler Sara (Ed.), Culture of the Internet, Lawrence Erlbaum, Mahwah, NJ, 1997.
[21] M. Granovetter, The strength of weak ties, Am J Sociol 78 (6) (1973) 1360–1380.
[22] J. Padgett, C. Ansell, Robust action and the rise of the medici, 1400-1434, Am. J. Sociol. 98 (6) (1993) 1259–1319.
[23] D. Kent, The Rise of the Medici: Faction in Florence, 1426-1434, Oxford University Press, Oxford, 1978.
[24] M. Mizruchi, L.B. Stearns, A longitudinal study of the formation of interlocking directorates, Adm. Sci. Q 33 (2) (1988) 194–210.
[25] B. Mintz, M. Schwartz, The Power Structure of American Business, University of Chicago Press, Chicago, 1985.

[26] R. Cross, R.J. Thomas, Driving Results through Social Networks: How Top Organizations Leverage Networks for Performance and Growth, Jossey-Bass, San Francisco, CA, 2009.

[27] R. Cross, R.J. Thomas, Driving Results through Social Networks: How Top Organizations Leverage Networks for Performance and Growth, John Wiley & Sons, San Francisco, CA, 2009.

[28] M. Kilduff, W. Tsai, Social Networks and Organizations, Sage, Thousand Oaks, CA, 2003.

[29] P.R. Monge, N. Contractor, Theories of Communication Networks, Oxford University Press, New York, 2003.

[30] D.E.M. Rogers, Diffusion of Innovations, fifth ed., Simon and Schuster, New York, 2003.

[31] M.E.J. Newman, The structure and function of complex networks, SIAM Review 45 (2003) 167–256.

[32] L.C. Freeman, Centrality in social networks: conceptual clarification, Social Networks 1 (1979) 35–41.

[33] S. Brin, L. Page, The anatomy of a large-scale hypertextual web search engine, in: Proc. 7th World-Wide Web Conference (WWW7), BrisBane, Australia, April 1998.

[34] D.J. Watts, S.H. Strogatz, Collective dynamics of 'small-world' networks, Nature 393 (6684) (4 June 1998) 440–442.

[35] A.L. Barabasi, R. Albert, Emergence of scaling in random networks, Science 286 (15 October 1999) 509–512.

[36] J. Leskovec, E. Horvitz, Planetary-scale views on a large instant-messaging network, in: Proc. 17th international Conference on World Wide Web (Beijing, China, April 21–25, 2008). WWW '08. ACM, New York, NY, 2008, 915–924.

[37] H.W. Park, M. Thelwall, Hyperlink analyses of the world wide web: a review, Journal of Computer Mediated Communication 8 (4) (July 2003).

[38] L.A. Adamic, E. Adar, Friends and neighbors on the web, Social Networks 25 (3) (July) (2003) 211–230.

[39] J. Heer, D. Boyd, Vizster: Visualizing Online Social Networks, in: Proc. 2005 IEEE Symposium on Information Visualization (October 23-25, 2005), INFOVIS. IEEE Computer Society, Washington, DC, 2005.

[40] B. Shneiderman, A. Aris, Network visualization with semantic substrates, IEEE Symposium on Information Visualization and IEEE Trans, Visualization and Computer Graphics 12 (5) (2006) 733–740.

Additional Resources

Barabasi, A.L. (2003). *Linked: How everything is connected to everything else and what it means*. New York: Penguin Group.

Bonacich, P. (1987). Power and centrality: A family of measures. *The American Journal of Sociology*, *92*(5), 1170–1182.

Borgatti, S., Mehra, A., Brass, D., & Labianca, G. (2009). Network analysis in the social sciences. *Science*, *323*(5916), 892–895.

Brandes, U., & Erlebach, T. (Eds.), (2005). *Network analysis: Methodological foundations*. Berlin, Heidelberg: Springer-Verlag.

Buchanan, M. (2002). *Nexus: Small worlds and the groundbreaking theory of networks*. New York, NY: Norton.

Burt, R. (1995). *Structural holes: The social structure of competition*. Cambridge, MA: Harvard University Press.

Burt, R.S. (1992). *Structural holes*. Cambridge, MA: Harvard University Press.

Haythornthwaite, C. (1996). Social network analysis: An approach and technique for the study of information exchange. *Library and Information Science Research*, *18*(4), 323–342.

Johnson, S. (2002). *Emergence: The connected lives of ants, brains, cities, and software*. London, UK: Penguin.

Knoke, D., & Yang, S. (2007). *Social network analysis (Quantitative Applications in the Social Sciences)*. Thousand Oaks, CA: Sage.

Nooy, W., De, Mrvar, A., & Batagelj, V. (2005). *Exploratory social network analysis with pajek*. Cambridge, UK: Cambridge University Press.

Watts, D. (1999). *Small worlds*. Princeton, NJ: Princeton University Press.

Watts, D. (2003). *Six degrees*. New York: Norton.

Wellman, B., & Berkowitz, S.D. (1988). *Social structures: A network approach*. Cambridge, UK: Cambridge University Press.