Contents

Acknowledgements ix

Preface xi

A note on software xvi

1. Introduction 1
2. From social networks to complex networks: A short history 11
3. Theoretical background 45
4. A macro level approach: The properties of complex networks 75
5. A micro level approach: Measures of centrality in the network 101
6. Centrality and marginality in the Afrikaans poetry network 111
7. Conclusion 129

Appendix A 133

Bibliography 137

Notes 149

2

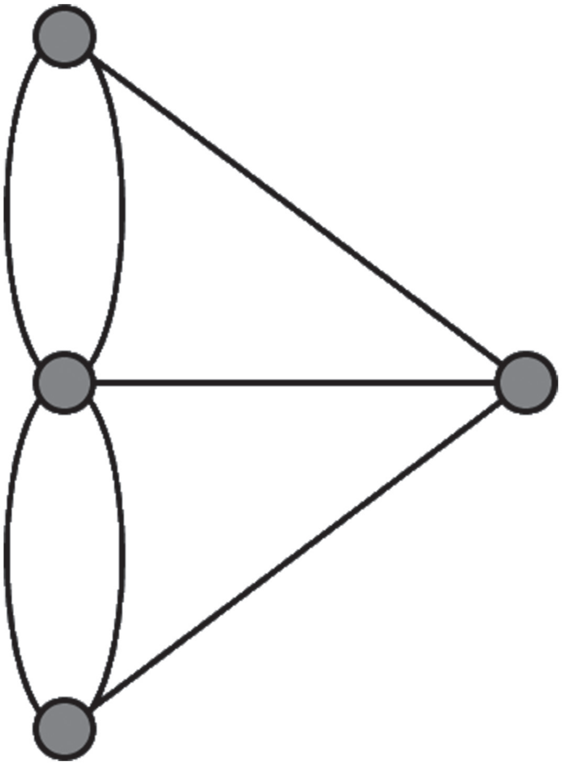
From social networks to complex networks: A short history

Network theory's interdisciplinary applications are mirrored by its diverse roots, ranging from the purely mathematical to the sociological and anthropological. Although a more detailed overview of network theory's development can for instance be found in Freeman (2004), Amaral and Ottino (2004), and Prell (2012), the current chapter provides an historical overview of how this interdisciplinary field evolved in order to contextualize some of network theory's theoretical developments that will be discussed later in the book.

Network theory is closely intertwined with graph theory, a branch of mathematics, which also constitutes network theory's oldest foundation. The development of graph theory is usually traced to Leonard Euler's famous Königsberg bridge puzzle, as formulated in 1736 (Hu 2011, 180 and Boccaletti et al. 2006, 177). Euler writes,

> In the town of Königsberg in Prussia there is an island A, called 'Kneiphoff', with the two branches of the river (Pregel) flowing around it. There are seven bridges, a, b, c, d, e, f, and g, crossing the two branches. The question is whether a person can plan a walk in such a way that he will cross each of these bridges once but not more than once. [...] On the basis of the above I formulated the following very general problem for myself: Given any configuration of the river and the branches into which it may divide, as well as any number of bridges, to determine whether or not it is possible to cross each bridge exactly once.[11]

Euler demonstrated that the puzzle could not be solved – there was no way in which all bridges could be crossed only once. In the process, however, he illustrated that physical distance had no relation to his puzzle, and was therefore the first to represent a network as a graph:



Euler's Königsberg bridge puzzle

In the above graph, the bridges are represented as links connecting nodes, which represent the different margins and islands.

Jacob Moreno's (and Helen Jennings's[12]) *Who shall survive?* (1934) was "a signal event in the history of social network analysis" (Freeman 2004, 7). Moreno and Jennings investigated why 14 girls from the Hudson School for Girls in New York had run away in just two weeks, and suggested that the phenomenon had less to do with the girls' intrinsic characteristics and more to do with their positions in a social network. Moreno used a technique he called "sociometry," which graphically represented the social network in what Moreno called a *sociogram*. Moreno (1934, 11) wrote that this approach, "enquire[s] into the evolution and organization of groups and the position of individuals within them." Similar to Euler's Königsberg bridge puzzle, Moreno emphasized structure over physical distance, where social influence would spread along connections between the girls.

Moreno's conception of social networks influencing individual behavior was of course not an isolated view, for throughout the twentieth century, numerous sociologists adopted a systems-theoretical perspective on societies. Emile Durkheim for instance argued that societies were comparable to biological systems, consisting of interrelated components, thus emphasizing the structure of the system over the intrinsic characteristics of its components. Borgatti et al. (2009, 892) write, "Moreno's sociometry provided a way of making this abstract social structure tangible." Sociometry is widely regarded as a forerunner of SNA, but Freeman (1996, 2004) notes that the roots of SNA are complex, and could also be traced to the works of among others Almack (1922), Wellman (1926), Chevaleva-Janovskaja (1927), Bott (1928), Hubbard (1929), and Hagman (1933).

Another clear antecedent of SNA was the works of Kurt Lewin. Lewin saw the social environment as a "field" (like the later Bourdieu[13]), which he defined as, "the totality of coexisting facts which are conceived of as mutually interdependent" (Lewin 1951, 240). Lewin (1939, 889) writes,

> Whether or not a certain type of behavior occurs depends not on the presence or absence of one fact or of a number of facts as viewed in isolation but upon the constellation (structure and forces) of the specific field as a whole. The 'meaning' of the single fact depends upon its position in the field; or, to say the same in more dynamical terms, the different parts of a field are mutually interdependent.

In terms of the social field, Lewin came to a similar conclusion as Euler had when attempting to solve the Königsberg puzzle, "the social field is actually an empirical space, which is as 'real' as a physical one. Euclidean space generally is not suited for adequately representing the structure of a social field – for instance, the relative position of groups, or a social locomotion" (Lewin 1939, 891). Like Moreno, then, Lewin applied a similar logic to social systems as Euler had done in his puzzle, further cementing the graphical representation of social systems as networks. Lewin also introduced the concept of the "shortest path" to sociology (Bavelas 1948, 17).

In these beginning years, SNA's roots branch out across various disciplines (e.g. graph theory and sociology). Another discipline that would eventually have an important influence on the emerging science of networks was anthropology. Borgatti et al. (2009, 893) write,

> ... building on the insights of the anthropologist Levi-Strauss, scholars began to represent kinship systems as relational algebras that consisted of a small set of generating relations (such as "parent of" and "married to") together with binary composition operations to construct derived relations such as "in-law" and "cousin." It was soon discovered that the kinship systems of such peoples as the Arunda of Australia formed elegant mathematical structures that gave hope to the idea that deep lawlike regularities might underlie the apparent chaos of human social systems.

The anthropologist Alex Bavelas (1948) was a student of Kurt Lewin, and introduced the concept of centrality in social networks. Although he set out to define what Freeman (1977) later formalized as betweenness centrality,[14] the measure he developed was nearer to closeness centrality, which measures not control over information in a network but rather independence of control (1980, 585, 593). Nevertheless, Freeman credits Bavelas for his "intuition" that contributed to the future important development of both betweenness and closeness centralities, which are discussed later in this book.

Kochen and Pool's studies in the 1950s led them to the definition of random graphs (Amaral and Ottino 2004, 152) and the identification of what is known today as the small-world phenomenon. Borgatti et al. (2009, 892) write, "On the basis of mathematical models, they speculated that in a population like the United States, at least 50% of pairs could be linked by chains with no more than two intermediaries." Although their work was only published towards the late 1970s (see e.g. Pool and Kochen 1979), it was widely circulated in preprint form, and is regarded as a direct influence on Stanley Milgram's "six degrees of separation" (1967) studies (Amaral and Ottino 2004, 152).

Like Kurt Lewin and Emile Durkheim, Siegfried Frederick Nadel (1957) saw societies not as monolithic entities, but rather as a "pattern or network (or 'system') of relationships obtaining between actors in their capacity of playing roles relative to one another" (Nadel 1957, 12). Nadel's work was one of the earliest formal treatments of the subject, and directly influenced the later work of Harrison White (Prell 2012, 34).

In the 1960s, the center of gravity of network research shifted from anthropology to sociology (Borgatti, Mehra, et al. 2009, 893). One of the foremost social network theorists was Linton Freeman, who formalized betweenness, closeness and degree centrality, with all three forms of centrality placing different emphasis on what happens in a network (see Freeman 1977, 1979, 1980). However, it was only in the late 1960s and 1970s that SNA developed into a separate field within sociology, especially at Harvard, where Harrison White institutionalized SNA.[15] White worked with many other influential researchers, including Stanley Milgram, and Mark Granovetter was one of his students. Granovetter (1973) studied the links that connect different clusters in a network, and proposed that 'weak ties' have special importance in spreading information in social networks – a key concept in contemporary network theory. Barry Wellman, another former student of White, would later form the International Network Society of Social Network Analysts (INSNA) (http://www.insna.org/), which publishes amongst others the journal *Connections*. Borgatti et al. (2006, 893) write that by the 1980s, "social network analysis had become an established field within the social sciences, with a professional organization (INSNA), an annual conference (Sunbelt), specialized software (e.g., UCINET), and its own journal (*Social Networks*)."

Within mathematics, one of the most influential advancements of graph theory was developed by Paul Erdös and Alfréd Rényi (1960). Their model laid the foundation for network models that would later develop into the scale-free and small-world models (see below).

In the late 1990s, networks became an object of interest from physicists. The first seminal publication was a paper by Duncan Watts and Steven Strogatz (1998), which appeared in *Nature* and argued that the small-world property of networks – as proposed by Stanley Milgram (1967) – was a universal attribute of complex networks, and not just of social networks. In other words, power grids, metabolic processes, neural networks, and other kinds of complex networks were comparable with social networks in terms of the average number of links that needed to be traversed to reach a node from any other node. Albert and Barabási (2002, 68) write that this paper caused an "avalanche of research on the properties of small-world networks," particularly in the physics community. Network analysis was now more than just SNA: it had become an instrument in the study of complexity in general, as Strogatz (2004[2003], 232) writes, "the 'small-world' phenomenon is much more than a curiosity of human social life: It's a unifying feature of diverse networks found in nature and technology."

In 1999, Barabási and Albert (1999) published an article in *Science* that argued that complex networks are scale-free networks that adhere to the so-called power law, which is discussed later. Boccaletti et al. (2006, 177) write that these two papers in particular "triggered" a "flurry of activity" in the physics community, directly leading to the popularity of this approach. The natural sciences in particular have focused their attentions on developing and refining models to come to a better understanding of complexity in networks. By 2004, Strogatz (2004[2003], 256) writes,

> In the past five years, the new ideas of small-world and scale-free networks have triggered an explosion of empirical studies dissecting the structure of complex networks. In case after disparate case, when the flesh is peeled back, the same skeletal structure appears from within. The Internet backbone and the primate brain – both small worlds. So are the food webs of species preying on each other, the meshwork of metabolic reactions in the cell, the interlocking boards of directors of the Fortune 1,000 companies, even the structure of the English language itself.

Along with these developments, growing computer power and the availability of large digital datasets had a profound influence on the development of the field (see Freeman 2004, 139, Barabási 2009, 413, Watts 2011, 82, and Scott 2012, 6). Network theory is heavily dependent upon computer-generated analyses, as already argued by Boissevain (1979, 392), and the development of cheaper, more user-friendly and more widely available software and hardware empowered a larger group of scientists to study networks. In the 1970s, programs such as DIP, SocPac, SOCK, COMPLT, BLOCKER and CONCOR (see Tichy, Tushman and Fombrun 1979, 513) facilitated network analysis, while by the 1990s, GRADAP, STRUCTURE, UCINET, NEGOPY and KRACKPLOT were used extensively (Haythomthwaite 1996, 331). Currently, Pajek is one of the most popular programs, along with UCINET, but a wide variety of network analysis programs have been developed – even a non-academic application to analyze Facebook contacts (TouchGraph). Anyone can now use SNA to look at his friends' connections on Facebook: SNA is no longer an approach limited to computer-savvy academics, but a tool with popular appeal outside academia.

The World Wide Web also allowed the gathering of larger datasets, which is one of the main reasons Watts and Strogatz could undertake their landmark study of small-worldedness in complex networks. Albert and Barabási (2002, 483) recognize this availability of digital data in their overview of the factors that contributed to the popularity of network theory,

> First, the computerization of data acquisition in all fields led to the emergence of large databases on the topology of various real networks. Second, the increased computing power allowed us to investigate networks containing millions of nodes, exploring questions that could not be addressed before. Third, the slow but noticeable breakdown of boundaries between disciplines offered researchers access to diverse databases, allowing them to uncover the generic properties of complex networks. Finally, there is an increasingly voiced need to move beyond reductionist approaches and try to understand the behavior of the system as a whole.

Dempwolf and Lyles (2012, 4) remind us that SNA is *both* a theoretical perspective *and* a practical set of analytical tools, as the above example of TouchGraph illustrates. The last major benefactor of network theory was the tremendous financial injection awarded to the development of software applications for SNA from a military intelligence standpoint. Unlike during the Cold War, where the US had faced a mostly monolithic enemy with a hierarchical organization and power base susceptible to attack, Al-Qaeda and its affiliates were an entirely different enemy altogether. Although Al-Qaeda had (at the time) a leadership hierarchy, small cells were capable of operating virtually independently, supplied through vast, global financial networks. Transnational terrorist networks, the Intelligence Community (IC) recognized, had to be approached *as a network*, and for that, they needed the tools to find the ties between members of terrorist organizations.[16] Already in the early 1990s, Sparrow (1991) advocated for the application of network analysis to criminal intelligence, and Glenn Henke (2009, 5) calls Arquilla, Ronfeldt and Zanini's (1999) RAND report on *Networks, Netwar, and Information-Age Terrorism*, "the first dedicated analysis of information age terrorism." Soon after the invasion of Afghanistan, studies using SNA to map terrorist networks emerged. Valdis Krebs (2002) was the first to publish a study using SNA to investigate terrorist networks, where he used open-source information to map the ties between the 9/11 hijackers, indicating that Mohamed Atta was the ring leader (Krebs 2002, 47) by using Freeman's (1979) formulas for betweenness-, closeness- and degree centralities. Rodriguez (2005) mapped the network responsible for the March 2004 Madrid bombings, Carley et al. (2003) analyzed the Al-Qaeda cell that was responsible for the bombing in Tunisia, while Koschade (2006) mapped Jemaah Islamiyah. In 2005, the Committee on Network Science for Future Army Applications (2005) published a special report on the utility of network theory. David Petraeus (2006) also includes a special section on SNA for military intelligence purposes in the new *US Army and Marine Corps Counterinsurgency Field Manual*. Numerous software platforms, including Sentinel Visualizer, Starlight VIS, and i2 Analyst's Notebook were developed from increased defense expenditure. Ressler (2006, 7) notes that government agencies, such as the Defense Advanced Research Projects

Agency (DARPA), the National Security Agency (NSA), and the Department of Homeland Security (DHS), have funded research related to SNA. Borgatti et al. (2009, 893) write that people working within security utilized SNA extensively,

> Of all the applied fields, national security is probably the area that has most embraced social network analysis. Crime-fighters, particularly those fighting organized crime, have used a network perspective for many years, covering walls with huge maps showing links between 'persons of interest.' This network approach is often credited with contributing to the capture of Saddam Hussein. In addition, terrorist groups are widely seen as networks rather than organizations, fueling research on how to disrupt functioning networks. At the same time, it is often asserted that it takes a network to fight a network, sparking military experiments with decentralized units.

The application of SNA for intelligence purposes is however not the only practical application of this approach. Zhu, Watts, and Chen (2010, 151) write,

> ...firms are using social network analysis to make hiring and transfer decisions, to optimize the flow of information among their employees, and to get the most out of talent and ideas that are embedded in the social networks of their staff.

Network theory's wide academic applications sets it apart from most other scientific theories, for it is truly interdisciplinary. However, these practical applications within military intelligence and business also sets it apart from other theoretical approaches in another way: it has been applied in the non-academic world, and found useful. Network theory's utility has therefore been demonstrated not only across academic disciplines, but also in practice, which is a relatively unique feature of the theory of complex networks as compared with other theories.