# Social Network Analysis in Telecommunications

Carlos Andre Reis Pinheiro

# Contents

CHAPTER 1

# An Introduction to Social Network Analysis

This chapter introduces the origins of social network analysis as a social science. A basic approach of studying social relations comes from the 1950s, when it was an important tool in understanding relationships among people within some particular social groups. How this approach evolved during the years is also briefly described, as well as how mathematical concepts, such as graph theory, were introduced to create a formal method to analyze and provide relevant information about social groups.

Therefore, Chapter 1 presents the evolution of social network analysis, from its appearance as a social science until its development into a more formal method based on mathematical concepts.

## EVOLUTION OF SOCIAL NETWORK ANALYSIS

Social network analysis takes into consideration the fact that social life is relations among individuals. Behaviors within a network are critical to understanding social connections, and—more important—the implications that they create inside a community. The most relevant implication is how particular people inside a social network can influence others to perform similar events. In terms of business, what makes an individual a leader or follower? What is the impact when a telecommunications company loses an influential customer within a social network? Are companies able to lead other customers afterward? Are companies able to lead other customers to purchase new products and bundles or to consume new services? Understanding the consequences from the new perspective of social relations can allow companies to predict new business events and to focus the customer relationship management more accurately.

The study of social networks is formally defined as a set of *nodes*, which consist of *network members*. Those nodes are connected by different types of relations, which are formally defined as *links*. Network analysis study takes all those connections as the primary building blocks of the social world. It does not collect just unique types of data, but traditional analysis takes into consideration individual attributes. Social network analysis also considers the attributes of relations from a fundamentally different perspective than that adopted by individualist or attribute-based social science. Traditional methods of data analysis usually consider individual attributes from all observations in order to analyze the information available. What is the average characteristic from a particular population of companies, employees, customers, or markets? Particularly in telecommunications, it is quite common to analyze the individual data to understand customer behavior, such as the average billing, type of payment, frequency and amount of service usage, and so on. Besides the individual attributes, social network analysis considers all information about the relationships among the network members (nodes). As a matter of fact, the information about the relations among the individuals within a social network is usually more relevant than the individual attributes of the individuals. The relations among the individuals can tell more about customers than their individual attributes. This is the basic difference between data analysis and social network analysis. What you are is not as important as how you behave and connect with others.

Alexandra Marin and Barry Wellman published a survey describing some case studies about social relations. According to them,

> A conventional approach to understanding high-innovation regions such as Silicon Valley would focus on the high levels of education and expertise common in the local labor market. Education and expertise are characteristics of the relevant actors. In contrast, a network analytic approach to understanding the same phenomenon would draw attention to the ways in which mobility between educational institutions and multiple employers has created connections between organizations. Thus, people moving from one organization to another bring their ideas, expertise, and tacit knowledge with them. They also bring with them the connections they have made to coworkers, some of whom have moved on to new organizations themselves. This pattern of connections between organizations, in which each organization is tied through its employees to multiple other organizations, allows each to draw on diverse sources of knowledge. Since combining previously disconnected ideas is the heart of innovation and a useful problem-solving strategy, this pattern of connections—not just the human capital of individual actors—leads to accelerating rates of innovation in the sectors and regions where it occurs.[1]

The most relevant fact of analysis here is not the individual attributes or characteristics. Certainly, all the information about the individual is quite important—the level of education, papers published, conferences attended, researches conducted, and so on. However, from a perspective of innovation and how data evolves over time, the relationship among the individuals and the institutions with which they are associated are more important than the attributes of the individuals. The way individuals share their ideas, compile their works, and how their institutions collaborate their resources is more influential than individual skills, talents, and initiatives.

A social network is basically comprised of a set of nodes connected by one or more links, which represent distinct types of relationships. For instance, in telecommunications, the nodes are usually assigned to the customers; landlines, broadband, mobiles, or whatever other type of communication device. The links, especially in the mobile environment, are assigned to the calls, the texts, the messages, e-mail, and any sort of communication's connections that can be collected somehow.

The conception of nodes and links can change according to the subject of interest or the business area which is under consideration. In the financial or banking industry, the nodes can be not only the customers but the accounts, and links can be represented by transactions, policies, claims, mortgages, payments, and so on.

One of the most important stages in this sort of exploratory analysis, particularly involving social networks, is to properly define the role of the nodes and the links. This phase is certainly crucial to find applications in relation to business issues. If you do not define roles of nodes and links, you will be able to build and analyze a social network, but it will not lead to applications for business.

The task of choosing which kind of available data to use, based on different sort of problems that are particular to different kinds of businesses, constitutes the work of the network analysts. As a matter of fact, for the same industry, but considering different problems, the same data can be used differently. For example, consider the telecommunications environment. Telephone numbers are usually thought of as nodes and calls are thought of as links. However, if the subject matter is fraud, and the identification of the perpetrators of fraud is the focus, the telephone number could be viewed as a link that connects two distinct individuals. In this way, the telephone number can be used to compose a node for a particular problem and can be used to compose a link for another one.

Indeed, the first endeavor in a social network analysis approach is to properly define what the nodes are supposed to be and what the links are supposed to be. In telecommunications, it is quite straightforward to define customers as nodes and calls as links, especially in mobile operations, even though there is more than one type of link in mobile operations, such as calls, text messages, multimedia messages, e-mail, and even Internet access. In the case of Internet access, what is the destination node? Is it a Web site? Even in a straightforward scenario such as telecommunications, we can visualize more than one type of node and more than one type of link. Thinking about landlines, where the relationship between the device and customer is not one to one, nodes can be defined in different ways. For instance, the nodes can be defined as the residents in a household, splitting the links among them, or as the household itself, aggregating the links into the telephone number.

In some industries more complex than telecommunications, such as insurance and banks, it is possible to identify distinct types of nodes and links, according to the sort of business problem to be addressed.

Also, when there is more than one type of node to be considered and more than one type of link to connect nodes, it is possible to include different weights for nodes and links as a way to distinguish them based on some particular rule.

Laumann identified three distinct approaches to select nodes and links according to the specification of the edge of the social network.[2] The first approach is based on position, considering the members of a particular organization or community as the nodes within the network. The second approach is based on event, considering the nodes with their positions well defined within an organization or community as members of the network. The third approach is based on the relation, considering the members of an organization or community based on a set of nodes that clearly present strong relationships within a particular network.

All these approaches can occur simultaneously and can also be applied at the same time for some different types of networks. In this way you define not just the nodes and links but also the boundaries of the network, which then helps to define nodes and links to be included in the study of communities.

In telecommunications, for instance, customers can be considered as a position-based approach, where any type of customer can be considered as a particular node. Also, in most cases, companies affect their own customers, and, therefore, the boundary of their network analysis could be considered to be their own network. However, considering the high migration of customers among telecommunications companies, especially the number portability in mobile operations, some specific mobile numbers could be accepted as members of the inside network. These external nodes could be quite relevant for understanding churn and customer acquisition, and thus they can be included in the internal network analysis process. Finally, the consideration of the population of interest is crucial for network analysis. By considering any telephone as a customer and then as a node, some telephone numbers should be discarded from the network, such as emergency numbers, customer care numbers, some particular services based on automatic response units, and so on. Those types of numbers do not represent a node in the social perspective. In addition, when analyzing the individual customers in terms of relationships, it is possible to consider all links and all nodes, but in order to compare usage among the customers, perhaps some kinds of customers should be excluded from the entire population, such as private companies and special individuals playing professional roles.

Because the process of identifying nodes and links is not straightforward at all, considerable time and effort should be spent in this stage. This recognition process is a key factor in whether social network analysis succeeds or not.

When social networks are considered, the relations between the nodes should be considered rather than their isolated attributes. In this way, the individuals hold a set of attributes that must be analyzed when the network is built. In order to study the effects of attributes such as average revenue, billing, payment history, active products and services, and so on, they should be sorted to reveal their relevance in the network environment.

A particular method to consider the impact of some attributes in social network analysis is presented in the case study presented in Part II. The effect of business attributes in relation to the social network analysis over time can reveal the average individual behavior assigned to some business events. For instance, when analyzing product and service acquisition, it is possible to correlate influential customers, the ones who lead other customers to purchase the same type of product or service, and then highlight the average values of some particular business attributes for those leaders. By doing this, companies are able to identify the leaders for particular business events; in addition, companies can identify their average behavior in terms of usage and characteristics.

It is important to note that, when attributes are taken separately, individuals with common attributes can produce the same outcome or be considered to have the same value of corporate importance. This isolated attribute analysis can link individuals with same average values for those business variables, even if they have completely distinct behavior in terms of usage or relationship.

Conversely, when examining a social network and its behavior, individuals are analyzed by their relationships (strength, frequency, importance, etc.). This distinct approach might lead the analysis in a completely different direction: evaluating individuals with common attributes in distinctive ways, hence producing totally different outcomes. In contrast to the evaluation of individual attributes, the network analysis can put together nodes with similar behavior in terms of relationships even though they hold completely different values in their individual attributes. The key to grouping similar nodes is the relationship rather than the attributes.

The relevant message here is that individuals with common characteristics can behave in completely distinctive ways. Using traditional clustering techniques, these individuals might be considered similar because of their attributes and would probably fit in the same group. However, based on social network analysis, those same individuals would fit into distinct groups or communities. Even though individuals with common attributes can behave in similar ways, in social network analysis, the most relevant analysis is the relationship with the neighbor nodes instead of their own characteristics.

By analyzing the behavior embedded in social networks, it is possible to exploit different levels of patterns. The results are especially useful for marketing, especially in diffusing processes related to product and services acquisition. Similar behaviors can lead a business to purchase new types of products and services, or even churn.

In most industries, especially in telecommunications, churn and product and service acquisition can be understood as a chain of events. In other words, some key customers, the leaders, can lead other customers to leave after they churn or they can lead other customers to acquire similar products and services after them. For this reason, it is crucial to target the right subset of customers in a particular campaign, whether it is in relation to customer retention or to product acquisition. The right set of customers can diffuse products or services within the telecommunications network.

Also, analyzing social networks requires looking at individuals not as a group but, instead, as a *community*. A group is comprised of individuals with similar characteristics, whereas a community is grouped together by their similar behavior.

Social network analysis should also take into consideration business strategy, which indicates different weights of importance to be used in the network analysis of nodes and links. In telecommunications, for instance, you might consider different weights according to distinct types of links, such as calls, text messages, multimedia messages, e-mails, and so on. In order to increase a specific text message service, for example, social network analysis should consider the weight of the links established based on text messages as highly relevant in the network relationship, rather than the calls themselves. Similarly, in order to increase a particular roaming package, all roaming calls should be considered more important than other calls. Although all distinct types of communications or relationships should be considered in the analyses—such as local calls, national calls, international, texts, multimedia, and so on—the specific roaming calls should hold a higher weight than other calls in order to highlight the relevance of roaming in the particular study.

## BUILDING SOCIAL NETWORKS BASED ON NODES AND LINKS

Social networks analysis is not really a brand-new discipline even though it has been gaining popularity in academia. In terms of academic approach, researchers have published experiments to describe the connections among people in distinct types of environments. In addition to the experiments, several theories in relation to social network analysis, and with respect to algorithms to calculate network measures have been published. One of the oldest articles about social network analysis is the one known as the Small World Experiment.[3] It is still fairly well known, and it is about the length of the path in a particular social network.

Stanley Milgram conducted some experiments in order to examine the average length of the path for some particular networks. The main objective of these experiments was to suggest that the human communities were not too big, where one individual could reach any other one by some maximum number of steps, or, in other words, by a certain number of intermediate individuals between them. The research outcomes have, indeed, suggested that the networks were characterized by short paths among the nodes considering the number of those intermediate steps. One particular experiment was associated with the famous phrase "six degrees of separation." However, Milgram himself had never used this term to refer to his experiment.

Milgram coauthored the experiment called the Small World Problem, which created a great deal of publicity. This publicity was mostly supported by several studies conducted simultaneously and focusing on the concept that the world was becoming highly interconnected. Bear in mind that this particular experiment took place in 1967. It is not difficult to imagine now the impact this concept had, viewing the possible types of relationships people could establish, with the use of different devices, technologies, and geographies.

Milgram's experiment intended to highlight the likelihood that two randomly selected people would know each other, regardless of the length of the path or, in other words, the number of steps (nodes and links) between them. This is certainly one way to solve the problem. A straight method to solve the Small World Problem is by calculating the average length of path connecting any two distinct nodes inside the network. Milgram's experiment ultimately created an algorithm to calculate the average number of links that connected any two nodes.

The procedure in relation to the Small World Problem was established in the paper by Jeffrey Travers and Stanley Milgram called "An Experimental Study of the Small World Problem."

> The procedure may be summarized as follows: an arbitrary target person and a group of starting persons were selected, and an attempt was made to generate an acquaintance chain from each starter to the target. Each starter was provided a document and asked to begin moving it by mail toward the target. The document described the study, named the target, and asked the recipient to become a participant by sending the document on. It was stipulated that the document could be sent only to a first-name acquaintance of the sender. The sender was urged to choose the recipient in such a way as to advance the progress of the document toward the target; several times of information about the target were provided to guide each new sender in his choice of recipient. Thus, each document made its way along the

> acquaintance chain of indefinite length, a chain which would end only when it reached the target or when someone along the way declined to participate. Certain basic information, such as age, sex and occupation, was collected for each participant.
>
> PARTICIPANTS. Starting Population. The starting population for the study was comprised of 296 volunteers. Of these, 196 were residents of the state of Nebraska, solicited by mail. Within this group, 100 were systematically chosen owners of blue-chip stocks; these will be designated by Nebraska stockholders throughout this paper. The rest were chosen from the population at large; these will be termed the Nebraska random group. In addition to the two Nebraska groups, 100 volunteers were solicited through an advertisement in a Boston newspaper (the Boston random group). Each member of the starting population became the first link in a chain of acquaintances directed at the target person.
>
> INTERMEDIARIES. The remaining participants in the study, who numbered 453 in all, were in effect solicited by other participants; they were acquaintances selected by previous participants as people likely to extend the chain toward the target. Participation was voluntary. Participants were not paid, nor was money or other reward offered as incentive for completion of chains.[4]

Travers and Milgram had also defined the *rules of participation* for this particular study, which described how the participants should include personal information and forward the document ahead. The rules also defined the *target person* as a stockholder who lives in Boston, Massachusetts; the *roster*, who was the person to avoid looping; and the *tracer cards*, which were business cards that requested a reply and contained information about the participant and the person who sent the card.

## INFLUENCE

The idea behind finding out the number of degrees that separate people in a network has triggered further concepts about how people behave inside the network rather than the number of connections they have. In terms of business, it is more important to know how each individual behaves inside the network than it is to know how many connections there are among people—that is, of prime importance is their correlated nodes through their connected links.

Considering that case of six degrees of separation between any two nodes in a network, the important thing to be analyzed is the role of each one of those six nodes inside the network: how they connect the other nodes, how strong their links are, how central they are in the network, how many short paths they belong to, and so on. These answers drive the network analysis to identify the influential nodes, the leaders, the followers, and also the isolated or irrelevant nodes.

According to Simmel, social ties are primarily based on viewing components as isolated units. However, those components are better understood as being at the intersection of particular relations and as deriving their defining characteristics from the intersection of these relations. He argues that society itself is nothing more than a web of relations. He wrote:

> The significance of these interactions among men lies in the fact that it is because of them that the individuals, in whom these driving impulses and purposes are lodged, form a unity, that is, a society. For unity in the empirical sense of the word is nothing but the interaction of elements. An organic body is a unity because its organs maintain a more intimate exchange of their energies with each other than with any other organism; a state is a unity because its citizens show similar mutual effects.[5]

This statement is an argument against the premise that a society is just a bunch of individuals who react individually and independently to particular circumstances according to their personal desires. Based on his belief that the social world is found in interactions rather than in an aggregation of individuals, Simmel argued that the primary work of sociologists is to study patterns among these interactions rather than to study the individual motives.

Simmel wrote in the same study:

> A collection of human beings does not become a society because each of them has an objectively determined or subjectively impelling life-content. It becomes a society

> only when the vitality of these contents attains the form of reciprocal influence; only when one individual has an effect, immediate or mediate, upon another, is mere spatial aggregation or temporal succession transformed into society.[6]

There also are some formal theories that are focused on the study of the effects of relationships among the individuals rather than on the attributes that describe the individuals themselves. For instance, if communities within social networks are densely connected, you can expect to find short paths linking most pairs of nodes.

By using these theories, mathematical models and computer simulations can be built to analyze social networks describing the patterns inside the entire network and also in several smaller communities within the network. The study of internal communities, whether densely or sparsely connected, is quite relevant to analyze some business actions over time, such as churn, diffusion, and even fraud events. Dense communities tie their members to each other and, therefore, are often targeted in retention and sales campaigns. Sparse communities are useful in analyzing patterns of behavior of the overall network.

Community analysis is quite relevant for business purposes and can reveal distinctive patterns inside networks that can be used to target particular marketing campaigns. In telecommunications, for instance, where the entire network is huge and sparse, the identifications of smaller internal communities comprising tight connections can be used to create specific bundles of products and services directed toward particular patterns of behavior. Highly dense communities inside sparse networks can be viewed as a single unit, and, therefore, individual churn from this unit (or a product acquisition from it) can represent cascading events; that is, one churn or one purchase can trigger a chain of churns or purchases.

The average behavior in a very tight community is usually well distributed throughout the members of this community, which means that the values assigned to the individual members are similar to the average values of the entire community. Because of this, it is easy to establish bundles of products and packet of services for the entire community when the average behavior is well distributed among the members.

## STRUCTURES OF SOCIAL NETWORKS

According to the type of social network and its structure, the outcomes in terms of metrics and measurement can be quite different, indicating distinct approaches to exploit the networks.

Based on the study of Borgatti,[7] there are four main categories to classify networks according to their arguments. The *argument transmission* represents networks in which connections are widely distributed by their nodes. *Argument adaptation* represents two nodes with similar network positions that are likely to follow similar behaviors when faced with similar circumstances. *Argument binding* represents two subnetworks bound to each other so that they act as a single network. Finally, *argument exclusion* represents a particular link that, because of its presence, excludes another link within the network.

The transmission argument in telecommunications is quite relevant. As long as some types of communications, such as text messages, remain popular, the connections inside the network can spread faster and wider than before. Considering particular massive marketing campaigns and spam messages, this transmission argument can be very relevant to analyze network structure and behavior.

The adaption argument has two distinct aspects, occurring simultaneously or sequentially. When two nodes perform the same event at the same time, and they do not hold any kind of connection, we can consider that similar positions inside subnetworks or similar constraints and circumstances could take in place. However, if they do hold some sort of connection or, even more important, if they hold a strong relationship, we can consider that they have exerted some influence on each other. They may have previously decided to follow the same event at the same time. In the first case, the adaptation argument is more about the environment and scenario, but in the second case, this argument is more about influence.

The binding argument can be closely associated to a concept called articulation point. The articulation point is the node that links two connected components. Connected components are a set of nodes that can reach each other, no matter what path they take to do it. Nodes that play the articulation point role usually are central nodes, connecting small groups, subnetworks, or even communities. All nodes as articulation points can be targets for further analysis.

The exclusion argument is not so applicable in telecommunications. For example, in social networks based on employer and employees, if two persons are married, they cannot work in the same company. In this way, the married link excludes the workmate link. However, in telecommunications, it is possible to have several types of links at the same time, such as calls, texts, media, e-mail, and so on.

## ANALYSES APPROACH FOR SOCIAL NETWORKS

All data in relation to networks are critical to performing any sort of exploratory analysis. This includes information about the nodes and the links, as well as the types and weights assigned to the nodes and links.[8]

One of the most important analyses in relation to social networks is the comparison among distinct behaviors. These behaviors are assigned to the nodes as well as to the links. Most often, the links' behaviors—their frequency, relevance, regency, and so on—can change the importance of the node assigned to the links. Thinking about telecommunications networks, customers who hold several links to very important nodes also become important. It is possible to include, in the social network analysis, individual attributes, such as customer segmentation, customer value (a corporate score for each customer), average billing, and then use them in the network analysis. In the same way, it is possible to include individual attributes for the links, such as the price of the call, the duration, the time, the frequency, among others, and then use these attributes in the network analysis. Those distinct attributes can create different weights for nodes and links, and thus completely change the network analysis. Once again, nodes comprising high value links become more important, and links connecting high–value nodes also become important.

Several measures can be calculated to represent the properties or characteristics of the network, as well as the subnetworks or subgraphs within it. Metrics in relation to networks include the number of connections a particular node has. This particular metric can be split into two distinct measures, describing the number of incoming connections and the number of outgoing connections. This is quite relevant in the telecommunications industry, representing the value of the outgoing and incoming calls among customers. Other network metrics include the path through the network, such as the length of the paths connecting nodes, the distance among them, how central the nodes are within the network, and so on.

Also, the distance of these relations can show some relevant features about the network's structure. The strength of the links is very important to establish the relationship behavior among the nodes. A particular link between two nodes can be very central to the entire network. Therefore, not only can the nodes be valued in terms of closeness but also the links can be measured in terms of how they connect nodes within the network.

In addition, when a network analysis takes place, the properties of the entire network should be calculated. Properties such as the number of degrees for each node, incoming and outgoing links, centrality, closeness, betweenness, and influence over the other nodes should all be calculated as well and compared to each other. This comparison analysis enhances the pattern of behavior for the entire network but especially of the individual behavior for each node and link inside the network, revealing the high-value nodes in terms of network measures.

Networks can also be analyzed by dividing the entire population into small connected communities. Similar to the comparison analysis for nodes and links, a comparison evaluation based on the small internal communities can reveal how the network is clustered and which communities are more important in terms of network metrics. In this way, it is possible to see that the average network metrics for the communities' nodes are also the network metrics for the communities themselves.

Community analysis can describe the network behavior, but, mostly, it is useful to highlight unexpected clusters within the network in terms of behavior. When considering usage, unusual groups of nodes—in telecommunications, for instance—can represent a possible set of customers committing suspect or fraudulent events.

Even if these customers are not committing fraud but are just high-usage customers of telecommunications services, this particular group of nodes should be handled in a distinct way by the company, protecting them from competitors.

## GRAPH THEORY AND SOCIAL NETWORK ANALYSIS

Social network analysis can be explored and explained by two distinct approaches: by social sciences, as described so far in this chapter, and also by mathematical science, described briefly at the end of this chapter and later in the book (Part II).

The mathematical science that describes and explains the social network analysis is graph theory. Several concepts and measures in relation to social network analysis come from graph theory. A great advantage of graph theory is the mathematical formulas that can be applied in computing and, therefore, in business problems.

Graphs can be considered mathematical structures used to model pairs of relations between distinct objects. The study of graphs should consider a set of vertices or nodes and a set of edges or links, which connect pairs of vertices or nodes. Basically, these are the same explanations that the social sciences make, but now the explanations use mathematical formalism. Based on formal formulas, it is possible to implement algorithms and hence apply social network analysis to practical problems. Social sciences usually involve small networks, comprising a reasonable number of nodes and links. However, for business problems in some industries, such as telecommunications, banking, and credit-card companies, the social network can be huge, with a really large number of nodes and links. Considering, for instance, a medium-size telecommunications company, with 10 million customers, a regular network can reach these 10 million nodes with 1 billion links. Efficient algorithms, based on formal methods, are, therefore, completely mandatory in order to address business issues in those types of industries.

A graph may be undirected or directed. *Undirected links* mean that there is no distinction between the two vertices associated with each edge; in other words, there is no direction in relation to the link that connects two nodes. Node A relates to node B the same way as node B relates to node A. This concept is well understood when thinking about friendship: Andre is a friend of Bruno the same way Bruno is a friend of Andre. There is no differentiation between the links that connect the nodes. *Direct links* mean that there is a direction in relation to the connection between two distinct nodes. In that case, node A may relate to node B in a different way than node B relates to node A. In telecommunications, traditional graphs are always directed: Node A calls to node B or node B calls to node A. Analogously, if Andre calls to Bruno 100 times and Bruno calls back to Andre just 10 times, this differentiation should be clear in the network. The way to make it clear is by establishing a direction for the links. There are two possible approaches to represent this scenario. The first one is creating 100 links (arrows) from Andre to Bruno and an additional 10 links (arrows) from Bruno to Andre. The second approach is to create one link (arrow) from Andre to Bruno and another link (arrow) from Bruno to Andre. In this case, the link from Andre to Bruno should be 10 times thicker than the link from Bruno to Andre.

One of the first people to use graph theory was probably Leonhard Euler in 1736, when he proposed the solution for the problem about the Seven Bridges of Königsberg, presented in the book authored by Barabási.[9] Figure 1.1 shows the question of the seven bridges, a historical problem in mathematics. As many cities in Europe evolved near rivers, Königsberg was set on both sides of the Pregel River. The city included two large islands connected to the mainland by seven bridges. The problem was formulated to find a way to walk through the city crossing each bridge only once. The islands could not be reached by any route but the bridges.

Euler indicated that the only important characteristic in this particular problem was the sequence of the bridges to be crossed. Therefore, it is possible to discard any sort of characteristic except the list of bridges connecting two landmasses. This problem is a foundation of graph theory. Extrapolating this concept to social network analysis, each landmass is a node, and each bridge is a link. The mathematical structure assigned to this problem is known as a graph, and in terms of social science, components constitute a network.

In this sort of problem, only the connection information is relevant. The graph can be represented in different shapes without

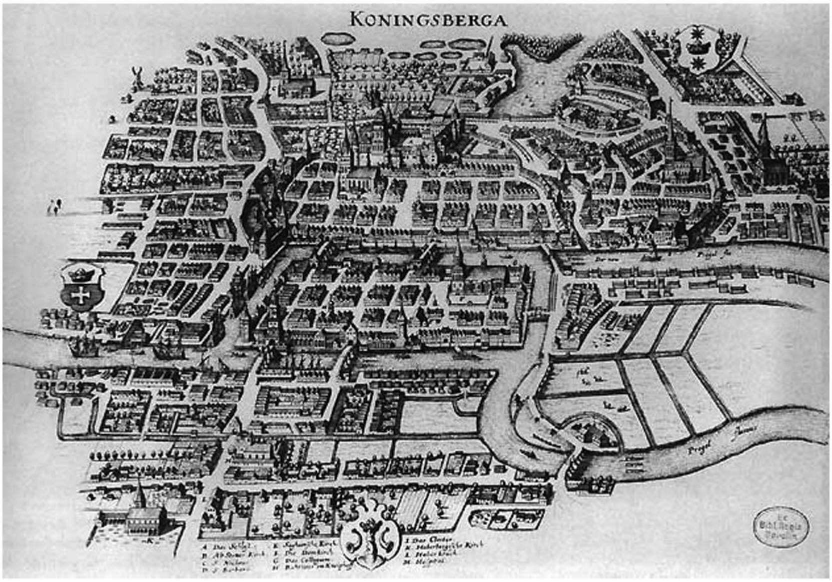


**Figure 1.1** Problem of the Seven Bridges of Königsberg

changing the graph itself. The existence or absence of a link between a pair of nodes represents everything in terms of social network analysis. Even when calculating individual network metrics for the nodes, the links are taken into consideration to perform it. For instance, the degree of a particular node is the number of related nodes it has. The related nodes are established by the link between a pair of nodes, the original one and the related one. Then the node's measure degree is computed by counting the number of links it has. Social network analysis is the study of the relationship among the nodes, and the relationship is defined in this context by the links.

Turning back to the bridge problem, Euler observed that whenever a walker gets a link and reaches a vertex by a bridge, he leaves the vertex by a bridge. In practical terms, during any walk through the city, the number of times a walker enters a bridge is equal to the number of times he leaves a bridge.

In a mathematical formalism, Euler said that the existence of a path in a particular graph depends on the nodes' degrees. The degree of a node is the number of links it has, regardless of whether the link

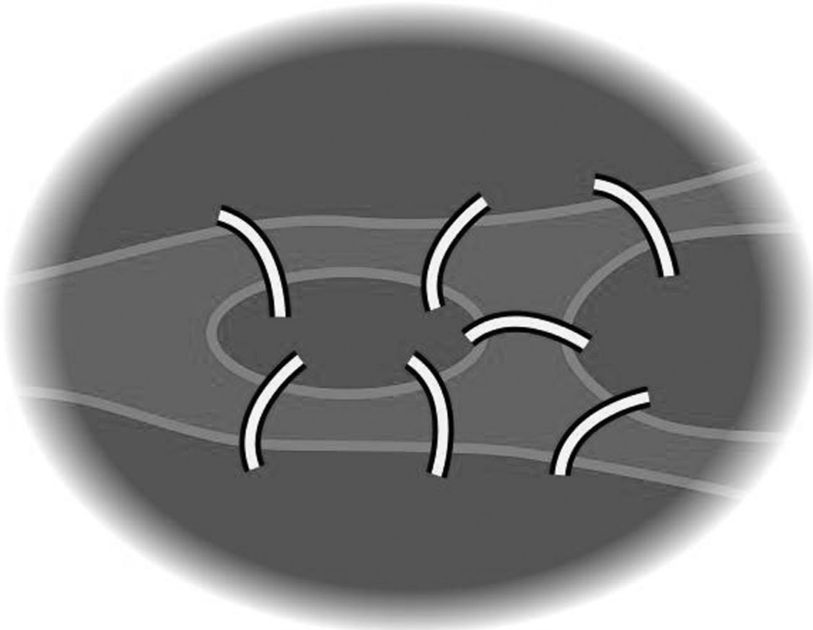

**Figure 1.2** Euler Circuit for the Seven Bridges' Problem
*Source:* Creative Commons Attribution ShareAlike 3.0.

is arriving to or departing from it. The importance is the number of links that are touching this particular node. Then, by Euler's observation, a necessary condition for a particular walk-through is that the graph be connected and have zero or two nodes with odd degrees.

An alternative formulation for the bridge problem is to discover a path by which all bridges are crossed and the starting and ending points are the same. This walk-through is known as Eulerian Circuit. This walk-through exists only if the graph is connected and there are no nodes with degrees having odd values. If this circuit crosses over all seven bridges just once, the Eulerian Circuit is called an Eulerian Path. Figure 1.2 shows the Seven Bridges' problem.

Nowadays, graph theory is regularly used for finding communities in networks. Algorithms regarding community detection have been increasing every day, making it possible for more real-world and business problems to be addressed by this sort of methodology. By detecting not just the small networks, communities, or subgraphs, but also their structures—hierarchical, circular, and others—it is possible to apply social network analysis to several distinct types of problems, such as cross-selling and up-selling in marketing and fraud and money laundering in revenue assurance. By understanding the small communities and their structures, it is possible to understand the network's behavior and also the way these networks diffuse over the time, even allowing some predictions.

The types of real applications of social network analysis have been increasing. It will be possible to address more and more kinds of problems in the future by using this methodology. Different kinds of industries, of different sizes, will be able to target their business problems, especially those related to customer behavior and relationships, by using social network analysis technique.

The case study in this book (see Part II) presents two distinct approaches for the use of social network analysis in telecommunications. The first one is the use of social network analysis for churn, which involves identifying the most influential customers who could possibly lead other customers to leave the company. The second example is the use of social network analysis for purchasing, identifying once again the influential customers who can possibly lead other customers to purchase similar products in the future. These influential customers are identified based on their relationships inside the network. Their network's measures in relation to degrees, distances, and paths will define their ability as leaders and how effective their influence can be over other customers.

## STATISTICS AND SOCIAL NETWORK ANALYSIS

Social network analysis started as a social science, and, therefore, it can be considered a branch of mathematical sociology more than it can be considered statistical or quantitative analysis.

However, the boundary between the social science approach and the statistical one is not so well defined and easy to understand. The mathematical methodology often handles the social network analysis in terms of a deterministic study, gauging the network data based on algorithms that compute the relationships among nodes and their respective strength, frequency, recency, direction, and so on. Contextualized in a mathematical model, social network analysis tends to consider all observations about a particular population, according to a specific subject of interest.

Statistical methodology often considers stochastic events about the relationship strengths in terms of distribution. Statistical analysis also considers particular subsets of the network as samples of a larger population. In the mathematical approach, a subset of a particular network is considered definite and encompasses the subject of interest of some sort of study. In the statistical approach, a subset of a particular network is considered a sample; because of this, the concern is to keep the results consistent and reproducible for similar samples from the same population.

This book, and mainly the case study presented in Part II, is more concerned with the mathematical approach. The network considered in the case study, for instance, is a subset of a particular subject of interest, considering a subset of the entire population. According to the statistical approach, the results from a network analysis about a subset of a particular population, which is a sample at the end, should be reproducible when similar analysis is applied to similar samples from the same population. This can be true in some environments or scenarios, but not in all.

In the case study, for example, the social network analysis was performed on the residential customers for a particular telecommunications company. Suppose this particular case study had taken a subset of the residential customers—for instance, a segment of them in terms of market. That social network analysis would have reached different outcomes if applied over a different sample, even though using the same residential-customer population. Thus, the mathematical methodology does not consider the subset of the network as a sample but as a definite population. The entire population could be considered all customers, including the wholesale, government, corporate companies, and so on. However, in that particular case, the subset is residential customers only.

Although the mathematical approach was put in place in order to analyze the social network of that particular telecommunications company, a statistical methodology was not ignored. Some specific statistical procedures, such as multivariate analysis, principal component analysis, clustering, and others, aimed at analyzing outcomes of social network analysis were applied to describe the results accomplished by network computation, which are depicted by several network metrics, such as degrees, centralities, closeness, betweenness, influence, and others.

Finally, a different sort of statistical approach, based on data mining techniques, can also be applied to understand some patterns of behavior for a particular social network being studied. For temporal analysis—that is, looking at the social network over time—predictive methods such as regression, decision tree, and artificial neural networks can be implemented in order to define some possible scenarios of evolution for the social network.

At the end, both mathematical and statistical approaches are put in place in order to analyze social networks, and each one is useful in some particular stage of the process. The mathematical approach is often used to compute the network measures and to establish the overall topology of the network. The statistical approach is often applied to analyze the outcomes from the network analysis, describing some relevant characteristics and particularities of the network, its distribution, outliers, communities, clusters, and so on.

## SUMMARY

This chapter introduced social network analysis from different perspectives. The first perspective was its basis as a social science, by analyzing the relationships among real social groups, such as authors, employees, students, and so on. This perspective was used to understand how people relate to one another and the impact of their actions over the social network. This perspective can be used to identify who the leaders are and who the followers are. The mathematical foundations were not so strong even though, for small networks, comprised of a reasonable number of nodes and links, it is possible to evaluate and analyze the social interactions over the network based on the social science theory.

Graph theory is a mathematical formulation that explains and depicts the social network analysis in terms of formulas and geometric concepts. These formulas can be translated into algorithms and then applied to business problems that involve huge social networks. Social networks in business occur in telecommunications, insurance, banking, and credit card operations. By creating social networks based on customers and calls, for instance, a huge network can be established, and efficient algorithms should be deployed in order to achieve useful and practical results.

Chapters 2, 3, and 4 present more concepts and foundations in relation to social network analysis. These foundations are more mathematical and, therefore, are basically built on the graph theory briefly presented here. Chapter 2 presents the theoretical foundation based on mathematical concepts in respect to social network analysis. Chapter 3 presents formal methods to create, describe, and evaluate social networks. Finally, Chapter 4 presents the measures in relation to social network analysis—metrics assigned to influence, power, and centrality. These metrics describe how influential and central the nodes are and how important they are in the social network for problem solving.

## NOTES

1. Alexandra Marin and Barry Wellman, "Social Network Analysis: An Introduction," June 2009. Available at http://www.hsph.harvard.edu/massconect/files/social_network_analysis._an_introduction.pdf (accessed October 26, 2010).

2. Edward O. Laumann, Peter V. Marsden, and David Prensky, "The Boundary Specification Problem in Network Analysis," in *Research Methods in Social Network Analysis* (Piscataway, NJ: Transaction Publishers, 1992), 61–87.

3. Jeffrey Travers and Stanley Milgram, "An Experiment Study of the Small World Problem," *Sociometry* 32 (4) (December 1969): 425–443.

4. Ibid.

5. Georg Simmel, Donald N. Levine, and Morris Janowitz, *Georg Simmel on Individuality and Social Forms: Selected Writings* (Chicago: University of Chicago Press, 1972).

6. Ibid.

7. Stephen P. Borgatti, Ajay Mehra, Daniel J. Brass, and Guiseppe Labianca, "Network Analysis in the Social Sciences," *Science* 323 (5916) (2009): 892–895.

8. For the following discussion, I am indebted to Marin and Wellman, "Social Network Analysis: An Introduction."

9. Albert-László Barabási, *Linked: The New Science of Networks* (New York: Basic Books, 2002).