

Introduction to Social Networks: Analysis and Case Studies

Nagehan İlhan, Şule Gündüz-Öğüdücü, and A. Şima Etaner-Uyar

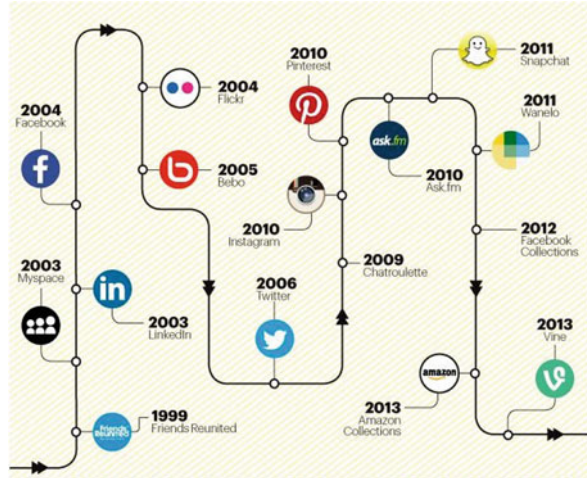
Abstract Social networks are platforms to share media, ideas, news, links or any kind of content between users and their neighbors, thus providing a perfect reflection of the structure and dynamics of the society. The recent advances in social media and the growing use of social networking tools have led to explosive growth of information available over the Internet and created a need to better understand the underlying structure of the knowledge flow. Social Network Analysis focuses on analyzing the relationships within and between users/groups in order to model the interactions and includes assumptions about how best to describe and explain the social network. Social Network Analysis and understanding the dynamics of social networks have become popular research topics and a vast number of studies have been performed. This chapter provides definitions of the basic concepts of Social Network Analysis and briefly introduces the topics covered in the book.

1 Introduction

This chapter will provide an introduction to social networks and give a description of resources and principal topics covered by Social Network Analysis (SNA). A social network is a social structure made up of actors called nodes, which are connected by various types of relationships. SNA is used to analyze and measure these relationships between people, groups and other information/knowledge processing entities and provides both a structural and a mathematical analysis. Therefore, the objects under observation are not actors and their attributes, but the relationships between actors and their structure. Relationships show what kinds of information are exchanged between which actors. SNA techniques are used to identify the

N. İlhan (✉) • Ş. Gündüz-Öğüdücü • A.Ş. Etaner-Uyar
Computer Engineering Department, Istanbul Technical University, Istanbul, Turkey
e-mail: nilhan@itu.edu.tr; sunduz@itu.edu.tr; etaner@itu.edu.tr

Fig. 1 History of online social networks [43]



characteristics of positions held by the actors in a social graph and the characteristics of the graph structure. With the advent of social network based applications, there has been more interest in computational techniques to discover the properties of networks. SNA has also attracted a significant interest in many fields such as sociology, epidemiology, anthropology, social psychology, communication studies, information sciences, etc.

Social networking has the ability to connect geographically dispersed users and provides social contact using the Internet. Social networks have become very popular in recent years due to the increasing usage of Internet enabled devices such as personal computers and mobile devices. The ever increasing popularity of many online social networks such as e.g. Facebook,¹ Twitter,² MySpace³ etc. is a good indicator for this observation. Figure 1 illustrates the history of online social networking sites in terms of when they were created. However, the social network concept is not restricted to internet based social networks.

Previous studies on SNA generally do not focus on online social interactions. In the last century, researchers in the behavioral sciences have stated studying social networks of offline interactions, such as person to person interactions, letters, telephone conversations, and so on. According to John Scott [42], there are three main research lines in SNA:

- Sociometric Analysts used graph theory and methods developed by Jacob Moreno who first introduced the concept of a sociogram [28]. A sociogram is

¹<http://www.facebook.com/>.

²<http://www.twitter.com/>.

³<http://www.myspace.com/>.

a visual diagram of relationship networks in which individuals are represented as points and their links to others as lines.

- Harvard University researchers first studied clique formations in social groups to identify cohesive subgroups in social systems in 1930s [24, 46].
- A group of anthropologists in Manchester University studied relational structures characterizing a tribal community. They directed their attention to people's informal social relationships rather than those associated with institutions and associations. Their work focused on the effective configuration of relationships deriving from conflicts between individuals and changes in these networks. John Barnes introduced the term 'Social Networks' and provided a remarkable advancement in the analysis of social structures [5].

Inspired by these studies, Harrison C. White and his colleagues focused on exploring the mathematical basis of social structures. They introduced important algebraic characteristics through the use of algebraic models of groups based on set theory, aiming to formalize different structural relations inside a group. The main idea of this study is that the search of structures in a network should be based on real interactions between the nodes and on how these interactions affect it, instead of on categories defined in advance. Later on, Mark Granovetter proposed a study on the importance of weak ties called "the strength of weak ties". He claimed that weak ties can be important in seeking information and members of a clique should look beyond the clique to their other friends and acquaintances in order to find new information [17]. In addition, a novel theory known as the small world phenomenon was proposed by Stanley Milgram [26]. In his famous small world experiment, a sample of individuals were asked to reach a particular target person by passing a message along a chain of acquaintances. The median length of the successful chains turned out to be five intermediaries or six separation steps. These studies constituted the foundation of the methods in SNA today.

The rest of the chapter is organized as follows. In the next section, we provide general definitions in SNA. In Sect. 3, tools commonly used to manipulate social networks are given. In Sect. 4, we discuss the remaining chapters in this book. Section 5 concludes the chapter.

2 Definitions in Social Network Analysis

2.1 Graphs

Social networks are usually represented as graphs. A graph $G(V, E)$ consists of a set of nodes V and a set of edges E .

- A graph may be directed or undirected: for instance, an e-mail may be from one person to another and will have a directed edge, or a mutual e-mailing event may be represented as an undirected edge.

- Graphs may be weighted; there may be multiple edges between two nodes. In a weighted graph G , let $e_{i,j}$ be the edge between node i and node j and $w_{i,j}$ is the weight of $e_{i,j}$. The total weight w_i of node i is the sum of weights of all its neighboring edges:

$$w_i = \sum_{n=1}^{d_i} w_{i,n} \quad (1)$$

where d_i represents its degree.

- Graphs may be unipartite, bipartite or multipartite.
 - A unipartite graph is a normal graph whose nodes are individuals and links. Nodes belong to the same class.
 - Bipartite graphs are graphs where nodes belong to two classes with no edges between nodes of the same class.
 - Multipartite graphs are graphs where nodes belong to more than one class, with no edges between the nodes of the same class.

A graph is generally described by an adjacency matrix A which is a square matrix with as many rows and columns as the nodes in the graph. The $[i, j]$ element of the adjacency matrix corresponds to the information about the ties between nodes i and j . An adjacency matrix may be symmetric (undirected graphs) or asymmetric (directed graphs). All entries of the adjacency matrix of an unweighted graph are 0 or 1 where a zero indicates that there is no tie and a one indicates that a tie is present between the nodes. For a weighted graph, the value of the $[i, j]$ entry in the matrix is the weight that labels the edge from node i to node j .

2.2 *Fundamental Metrics*

The study of SNA involves the measurement of particular structural metrics in order to understand the fundamental concepts of social graphs. Metrics are used to characterize and analyze connections within a given social network. Some of these metrics represent the characteristics of individual nodes whereas others infer a pattern that belongs to the network as a whole. Here, we describe the fundamental metrics that are used in SNA.

Centrality: Centrality measures the relative importance of a node and gives an indication about how influential a node is within the network. Betweenness Centrality, Closeness Centrality, Degree Centrality and Eigenvector Centrality are all measures of centrality.

Betweenness Centrality: Betweenness counts the number of shortest paths in a network that passes through a node. It takes into account the connectivity of the node's neighbors by giving a higher value for nodes which bridge clusters.

The minimum number of edges that must be traversed to travel from a node to another node of a graph is called the shortest path length. Nodes having a high betweenness value play an important role in communication within the network [12]. Betweenness centrality C_B for a node i is calculated as:

$$C_B(i) = \sum_{j < k} g_{jk}(i) / g_{jk} \quad (2)$$

where $g_{jk}(i)$ is the number of shortest paths between j and k that pass through i and g_{jk} represents all paths between j and k .

Closeness Centrality: Closeness centrality measures how close a node is to all the other nodes. A node is considered important if it is relatively close to all other nodes. Closeness is based on the inverse of the distance of each node to other nodes in the network [12]. Closeness centrality C_C for a node i is calculated as:

$$C_C(i) = \left[\sum_{j=1}^N d(i, j) \right]^{-1} \quad (3)$$

where $d(i, j)$ is the geodesic distance between node i and j . The geodesic distance is the length of the shortest path between two connected nodes.

Degree Centrality: Degree centrality of a node is the number of links to other nodes in the network. A node's in or out degree is the number of links that lead into or out of the node, respectively. In an undirected graph they are identical. This measure can be used for evaluating which nodes are central with respect to transferring information and influencing others in their immediate neighborhood [31]. It can be calculated by using the adjacency matrix:

$$C_D(i) = \sum_{j=1}^n a_{ji} \quad (4)$$

where a_{ji} is the $[j, i]$ entry of the matrix.

Eigenvector Centrality: A node's eigenvector centrality is proportional to the sum of the eigenvector centralities of all nodes directly connected to it [6]. It is a useful measure in determining which node is connected to the most connected nodes. For a node i , the Eigenvector Centrality is defined as:

$$C_E(i) = v_i = 1/\lambda_{\max}(A) \sum_{j=1}^N a_{ji} v_j \quad (5)$$

where $v = (v_1, \dots, v_n)^T$ is the eigenvector for the maximum eigenvalue $\lambda_{\max}(A)$ of the adjacency matrix A .

Clustering Coefficient: Clustering coefficient is a measure of the degree to which the nodes in a graph tend to cluster together. The clustering coefficient of a node i is the fraction of pairs of i 's neighbor nodes that are connected to each other by edges [48]. Clustering coefficient for node i with degree $d(i) \geq 2$ is defined as:

$$C_{cc}(i) = \delta(i)/\tau(i) \quad (6)$$

where $\delta(i)$ is the number of three connected cliques defined as $\delta(i) = \{\{i, j\} \in E : \{j, k\} \in E : \{i, k\} \in E\}$ and $\tau(i)$ is the number of triples of node i . Triple of a node i is a path of length two for which i is the center node.

Density: Density is the ratio of the number of edges in the network over the total number of possible edges between all pairs of nodes. It is a useful measure in comparing networks against each other. Density of a graph is calculated as follows:

$$G_G = 2 * |E| / (|V| * (|V| - 1)) \quad (7)$$

where $|V|$ is the number of nodes and $|E|$ is the number of edges in the network.

Path Length: Path length is a measure of the distances between pairs of nodes in the network. Average path length in a network is the average of these distances between all pairs of nodes. A shorter average path length means that the information will spread faster within the network.

Radiality: Radiality shows how far a node reaches into the network. It also measures the amount of novel information provided by the node and the influence it induces on the network. Nodes that have high radiality values usually have convenient positions to be innovators, thus they can relay the ideas in their neighborhood into other parts of the network.

Structural Cohesion: Structural cohesion is defined as the minimal number of nodes in a social network that need to be removed to disconnect the group [27]. It is used to identify cohesive subgroups in a network and reveals how such groups relate to one another.

3 Social Network Analysis Tools

SNA tools are used to represent, analyze and simulate a network by describing the features of the network either through a numerical or a visual representation. Network analysis tools enable researchers to examine different sizes of networks, from small to very large. Representation, visualization, characterization and community detection are expected functionalities of an analysis software. A software should be able to represent both the directed and the undirected structure of a network. Visualization of social networks is also important to understand the

network data and examine the result of the analysis. It allows to display nodes and edges in various layouts and distinct colors, size and other advanced properties of the network representation. Many quantitative measures have been defined to characterize networks. Computation of various measures provided at the node level or on the whole graph should be done by a software. Communities are groups of nodes, where nodes within the same community tend to be highly similar, sharing common features, while nodes of different communities have low similarity. Detecting and evaluating the community structure of graphs constitutes an essential task in SNA. Thus, community detection is one of the expected functionalities of an analysis software.

A wide variety of tools, each specialized on one or more of the expected functionalities, exist. Most prominent tools can be listed as follows:

1. **Gephi** Gephi provides an open source software package and a Java library for graph and network analysis [14]. It is a graph visualization tool that uses a 3D render engine to display large networks and works with huge datasets and graphs. Gephi is very convenient to explore dynamic networks. It supports graphs whose content changes over time, and has a timeline component where a timestamp of the network can be retrieved. From the time range of the timestamp, the software retrieves all nodes and edges that match and update the visualization module. Therefore, it allows to display a dynamic graph as a movie sequences. Gephi's framework offers the most common metrics like betweenness, closeness, diameter, clustering coefficient, average shortest path, PageRank, HITS, community detection using modularity, random generators for SNA.
2. **Pajek** Pajek is a network analysis package that performs analysis and visualization of large networks. The main goals in the design of Pajek are decomposition of a large network into several smaller networks, providing powerful visualization tools and implementing subquadratic algorithms to analyze large networks. Pajek analysis and visualization are performed using different data types like graph, partition, vector, cluster, permutation and general tree structures.
3. **NetworkX** NetworkX is a Python library for network analysis [18]. It is used for the creation, manipulation, and study of the structure, dynamics, and functions of complex networks. The library includes support for reading and writing graphs in various file formats. It also includes functions to generate graphs according to a variety of well known graph generation models. NetworkX is well documented, but the clustering algorithms are missing. NetworkX is not primarily a graph drawing package but basic drawing with Matplotlib as well as an interface to use the open source Graphviz software package are included. For advanced visualization, other tools should be preferred.
4. **Igraph** Igraph is a library for network analysis which uses Python and the R environment. It is one of the most essential libraries and is used in large graphs, similar to NetworkX [19]. It also provides properties for graph statistics such as computing degree centrality, closeness centrality and betweenness centrality etc. Dyad and triad census are available in Igraph and Pajek. Igraph offers a

few community detection algorithms like Walktrap [36], Fast Greedy [16], Label Propagation [37], etc.

5. **JUNG** JUNG(the Java Universal Network/Graph Framework) is an open-source software JAVA library which is mainly developed for creating interactive graphs [32]. Main functionality of JUNG is network and graph manipulation, modeling, analysis, and visualization. It has built-in support for GraphML, Pajek, and some text formats. It presents customizable visualization, graph types and includes graph theory, data mining and SNA algorithms (random graph generation, clustering, decomposition, optimization, statistical analysis, distances, flows, and centrality measures). It supports a native sparse matrix format and a graphical user interface, which makes JUNG's representations and algorithms both space and time efficient.

4 Topics in Social Network Analysis

The SNA field encapsulates a wide range of research topics and new methods and approaches are continuously being developed. Therefore, it is very hard to cover the entire network analysis literature. In this section, the topics are covered within the scope of the chapters in the book.

4.1 Node Analysis

The emergence of social networks has resulted in an exponential increase in the amount of the information about individuals, their activities, connections and features representing the characteristics of the individuals. These features may be of different types: demographic features like age, gender; features which represent political and religious beliefs; features representing hobbies, interests, affiliations etc. These features appear on the user's profile within the network, or attached to other objects like photos or videos. There are many applications that can make use of these features and connections like suggesting new connections to individuals based on finding individuals with similar interests and relationships, recommendation systems to suggest movie, music or other products, advertising systems which show advertisements to individuals in which they will most likely be interested. Thus, analyzing the nodes from different perspectives such as how closely related individuals are, who is the connector or hub in the network, who has best visibility of what is happening in the network, what are the distances and similarities of individuals from each other is crucial.

Widely used measures to identify influential nodes within the network are the degree centrality, closeness centrality and betweenness centrality [12]. However, some other methods like PageRank is adopted to find effective nodes. PageRank

is a popular Google patented algorithm to examine the entire link structure of the Web and determine which pages are most important by viewing the hyperlinks as recommendations [33]. A page with more inlinks (recommendations) must be important than a page with a few inlinks. A Web page is important if it is hyperlinked by an important recommender. In chapter “Ranking Authors on the Web: A Semantic AuthorRank”, a model is proposed to rank authors on the Web like ranking Web pages by considering their co-author links. They have adopted FOAF (friend of friend ontology), the so-called CO-AUTHORONTO ontology, in order to represent authors and also their co-author links on the Web. CO-AUTHORONTO extended with PageRank and AuthorRank metrics for ranking authors is based on the co-author links of the authors. Their framework builds on top of several known ranking metrics and side parameters like the number of authors, co-authorship exclusivity, PageRank, and co-author weight.

In chapter *Detecting Neutral Nodes in a Network of Heterogeneous Agent Based System*, a method is proposed to detect neutral agents in a social network using multi agent systems. They aim to reduce the complexity of analysis in a network consisting of heterogeneous software agents as nodes. Their method suggests detecting neutral agents (the agents that behave with similar frequencies and have similar behaviour) with respect to inter-type and intra-type communications. Thus, the complexity of the network and the algorithms to analyze them will be decreased by identifying and eliminating these agents.

4.2 Edge Analysis

Social networks are usually modeled using graphs where an edge between two nodes represents a relationship between them. Every node and each of the corresponding edges belonging to the nodes carry certain characteristics. Each node represents an entity, while each edge carries attributes that describe the nature of the relationship. There exist two types of social networks: homogeneous and heterogeneous [7]. In homogeneous social networks, there is only one kind of relationship between nodes and the knowledge flow is through this relationship. Heterogeneous social networks have several kinds of relationships between nodes and are also known as multi-relational social networks. Thus, the knowledge flow is through different kinds of relationships and network elements exchange different types of knowledge according to the type of the relationship. In the real world, social networks are usually multi-relational and users establish a large number of relationships with varying edge strengths and types: friends, family, colleagues and so on. Each relation defines a single relational network. A multi-relational network can be defined as a merger of multiple single relational networks. An edge between two nodes consists of all relations and interactions between the two individuals [9]. The weight of an edge in a multi-relational social network should consider the weight of all relationship types between the two nodes. Analyzing multi-relational

networks is much more complex than a single network analysis. In single relational networks, nodes between different groups are assumed to be loosely connected but in multi-relational networks the edges between different groups may be as dense as the edges within the groups. The authors in chapter *Global Structure in Social Networks with Directed Typed Edges*, present a spectral approach to understanding and analyzing graphs with a finite number of edge types. Their contribution is to extend conventional spectral graph analysis to networks where edges have different types. In this manner, it is possible to combine features of different types of social networks into a single network framework by taking into account the qualitative differences in the functionality of edges. They construct a multi-layer graph and embed this multi-layer graph using a directed-graph spectral technique. Their technique enables to answer several edge prediction questions such as if there should be an edge between given two nodes and what type of edge is it likely to be.

In a social graph, the presence of an edge between any two nodes indicates the efficiency of communication between them. It also means that they may belong to the same social group and work together. Understanding the structure and dynamics of social groups is crucial for network analysis. From the organizational perspective, groups are core organizational work units. The effectiveness of a group can provide a large contribution to the organizational success. Social edges in work groups are informal links between group members. Group members have different skills and capabilities which are essential for the effectiveness of the group and thus for the organization as a whole. Interaction within a group or interaction between groups are also important factors in a group's processes and outcomes. In literature, many studies exist that analyze the factors contributing to team effectiveness [20, 21, 41]. Unresolved empirical questions exist about the correlation between group density and group effectiveness. Studies show that social interactions and the communication frequency between members of a group are positively related with team effectiveness [4, 39]. They state that teams with densely configured interpersonal edges reach their goals better. In chapter *Social Networks and Group Effectiveness: The Role of External Network Ties*, the relationship between group effectiveness and social networks is examined. A communication network is formed using a 5-month ethnographic observation within three work groups employed in an Italian clothing company by recording all interactions occurring within the groups and between the groups. They show how qualitative information on the nature and dynamics of the ties between group members and other organizational actors can enhance comprehension of the impact of network relationships on organizational behaviors. They claim that the prolonged observation of group members' interactions offers researchers a privileged, thorough perspective into a group's social network. They emphasize that a high centrality degree in the request for information/advice network as opposed to the reporting of a problem or the communication of information/advice network can generate different effects on members' behaviors and on the evaluation of groups' effectiveness. In addition, they highlight the positive outcomes of team leaders who have also a high external prestige, in addition to internal prestige.

4.3 *Community Detection and Classification*

Extraction of social communities is one of the most important problems in today's SNA. Community detection attempts to solve the identification of groups of vertices that are more densely connected to each other than to the rest of the network. Communities correspond to groups of nodes in a graph that share common properties or have a common role in the organization of the system [16]. The ability to find and analyze communities has proved invaluable in understanding the underlying structure of the network. A number of methods to address this problem have been proposed. They vary in the type of network they can handle (unpartite vs. bipartite, weighted vs. unweighted, etc.) and the type of community structure they can detect (disjoint, overlapping, hierarchical, etc.). A comprehensive recent survey of community detection algorithms is proposed in [11]. The vast majority of the community detection algorithms find disjoint communities. However, in real networks, communities are usually overlapping which means that some nodes may belong to more than one community. Thinking about a person's personal social network, it is naturally considerable that a person may belong to several communities: for example family, co-workers, college friends, and so on. These kinds of networks are usually defined as overlapping networks [34]. However, most of the community detection algorithms find discrete communities which do not capture the overlapping community structure. Thus, for a correct representation of real network communities, it is crucial to find overlapping community structures. In chapter *Overlapping Community Discovery Methods: A Survey*, a review of the most recent proposals in the topic of overlapping community detection is introduced. Methods are classified by taking into account the underlying principles guiding them to obtain division of a network into groups sharing part of their nodes.

With the emergence of social networks, a large amount of information about individuals, their activities, connections has become available. A large part of these individuals which are represented as nodes in a graph structure may be labeled. This leads to a problem of providing correct and high quality labeling for every node, in other words, the node classification problem. In literature, there are a variety of node classification techniques. The simplest one is to use data about the labeled nodes and use a simple classifier in order to classify the unlabeled ones based on only those attributes that are local to these nodes. The techniques that work in this manner are called local classifiers. As in classical machine learning techniques, first, the features of the nodes should be identified. These features may be properties common to all nodes like age, gender, homeland, etc. But the existence of the connections and relationships makes the graph labeling problem different from the traditional machine learning classification techniques, where the classified nodes are assumed to be independent. The techniques that use the link information of the graph are called relational classifiers. Additional features such as the degree, centrality and so on based on adjacency in the graph can be defined in order to achieve a higher classification accuracy. Also, the labels of the neighbors constitute a useful feature. In social networks, the edges indicate some degree of similarity between the

connected nodes and constitute a useful input for the learning algorithm. Homophily and co-citation regularity are the two important phenomena used in the labeling process of the nodes. The labeling process can be iterative. Iterative algorithms use local neighborhood information to generate features that are used to learn local classifiers [30]. An iterative algorithm assumes that all of the neighbors' attributes and labels of that node are already known in determining the label of a node. Then, it calculates the most likely label with a local classifier which uses node content and neighbors' labels. In chapter *Classification in Social Networks*, the details of a method which uses content, link and label information on social network data for classification are given. Important properties of social network data which may be used to characterize a social network dataset and a list of aggregation operators which are used to aggregate the labels of neighbors are defined. A number of label aggregation methods are also experimented with. Different classifier accuracies with usage of only the content, only the link or both the content and the link information are evaluated. It is shown that homophily plays an important role in evaluating whether the network information would help in classification accuracy or not.

4.4 Graph Crawling

With the emergence and popularity of social networking sites, such as Facebook, LinkedIn, MySpace, Twitter, etc., the number of users joining these sites has dramatically increased. Alexa [2], a well-known traffic analytics website, reported that Facebook is the second most visited website on the Internet, LinkedIn ranked as the eighth and Twitter follows them with the tenth rank. Thus, online social networks have become an important phenomenon on the Internet. This global phenomenon has generated lots of interest in many disciplines to analyze human social behavior depending on observations of these networks. However, it is usually not possible to obtain datasets from the Online Social Network services due to privacy issues and it is hard to get such data directly from the service providers. Moreover, the huge size and access limitations of most of the services make it hard to completely cover the whole social graph. A widespread approach is crawling and sampling the network. It is desirable to crawl a small but representative sample of the network. In crawling, a user is randomly chosen and the friend list of the user is retrieved. Again one user is selected from out of the friend list and a new list of friends retrieved. In principle, the process repeats until every user in the network has been visited.

Several crawling strategies for single social networks have been proposed. These methods differ in the selection of the next friend. Breadth First Search (BFS) [50], Depth First Search (DFS) [47], Simple Random Walk (SRW) [23], Simple Random Walk with re-weighting (SRW-rw) [38, 40] and Metropolis-Hastings Random Walk (MHRW) [25] are the most popular ones among them. In BFS and DFS techniques, the graph is crawled node per node adding all discovered nodes to a list of nodes to visit. But, the difference between BFS and DFS techniques for graph crawling is the order in which the next node in the graph is selected. BFS selects the first node

of the list as the next node to visit and removes it from the list, while DFS selects the last node and marks it as visited. In Simple Random Walk technique, the next node is chosen uniformly at random among the neighbors of the current node. This algorithm is biased towards the high degree nodes. SWR-rw operates based upon a sequence of random nodes obtained by a Simple Random Walk with a proper re-weighting process to provide the unbiased sampling. MHRW is a crawling technique which applies the Markov Chain Monte Carlo (MCMC) method [15] for sampling from a probability distribution that is difficult to sample from directly. SWR-rw and MHRW ensure unbiased graph sampling. These crawling techniques are good for single social networks but they are not suitable for Social Internetworking Systems, where users are members of multiple social networks.

Recent studies show that users are often affiliated with different social network sites and in each one of them they exhibit different site-specific interaction patterns. This knowledge provides better understanding of the users' tastes and improves the quality of service they can get. The authors in chapter *Experiences Using BDS: A Crawler for Social Internetworking Scenarios*, confirm that a crawling strategy which is good for single social networks should not be expected to be appropriate for SIS, due to their specific topological features. They also propose a new crawling strategy Bridge-Driven Search (BDS), specifically designed for SISs, which overcomes the drawbacks of other crawling strategies. BDS is based on the bridge concept, which represents the structural element that interconnects different social networks. Bridge nodes are described as the users who joined more than one social network and explicitly declared their different accounts. They conduct several experiments and show that BDS outperforms state-of-the-art techniques. In addition they perform a large number of experiments to derive detailed information about the bridge nodes and argue that most of the required information on the structural properties of SISs can be obtained through studying bridges in detail.

4.5 Privacy and Social Networking Ethics

Online social networks such as Facebook, MySpace, Twitter, Orkut, LinkedIn and etc. are Web sites which are used widely to build connections and relationships. In principle, they allow their users to communicate with other people around the world, contact their closest friends, share experiences, photos, videos with them in real time. Users publish detailed personal information and information about their preferences and daily life. Social Web sites are also collecting a variety of data about their users, both to personalize the services for the users and to sell them to advertisers. However, some of the data revealed in these networks should remain private and not be published at all. Besides, scammers, identity thieves, stalkers, and companies looking for a market advantage are using social networks to gather information about customers.

In literature, some key concepts of privacy on online social networks are identified [35]. These concepts are network anonymization [22, 29], privacy preservation

[1, 45] and access control [8, 10]. Privacy issues of social networks include the disclosure of nodes' identity information, relationship information, data information related to nodes, etc. Network anonymity graphs are obtained by removing the nodes' identity information in social networks in order to preserve privacy. This makes node identification difficult for attackers. However, anonymization is not sufficient to protect privacy and the attackers still identify nodes' identity from the features and the structure of the network. Privacy preservation focuses on protecting sensitive information of users through techniques based on hiding sensitive attributes, identities and modifying data. But hidden attributes of the users can still be inferred. For example, it is possible to predict the home address of the user by analyzing the geographical place of the most frequent updates posted or it is possible to predict the work address by analyzing the relationships, namely if the majority of the users' friends are in the same institution, the user is most likely to work there. Access control mechanisms are used to reinforce access to users' sensitive information without explicit authorization by performing appropriate access control mechanisms. Many existing social network owners offer access control mechanisms that are primitive, permitting coarse-grained visibility control to users to place restrictions on who may view their personal information. Indeed, even with current access control mechanisms, users lose their control over data after its very first publication in the network. From the service owner perspective, it is crucial to protect users' privacy while providing useful data. In chapter *Privacy and Ethical Issues in Social Network Analysis*, privacy issues in social network data have been discussed. Different aspects of graph publication issues in graph publishing models have been introduced.

4.6 Cloud Computing with Social Media

Cloud computing is the delivery of computing services over a network such as the Internet. It allows individuals immediate access to a large number of supercomputers and their corresponding processing power that exist at remote locations. The cloud computing model allows users to access information and computer resources from anywhere with an available Internet connection. Social networking sites, online file storage, webmail, online software applications are examples of cloud services. Cloud computing service models can be divided into three main categories: Software as a Service, Platform as a Service and Infrastructure as a Service. Software as a Service (SaaS) provides running applications without installing on your hard drive and without any configuration requirements. Applications are hosted by a service provider and are made available to users over Internet. Platform as a Service (PaaS) model provides a development and production environment which consists of the operating system, the hardware and the network. Users install or develop their own software and applications.

The Infrastructure as a Service (IaaS) model provides the use of resources such as virtual machines, storage etc. It supplies the hardware and network resources; the

user installs or develops its own operating systems, software and applications. Cloud computing services reduce the cost and complexity of owning hardware, installing and configuring infrastructure. The other benefits to users are scalability and reliability. Cloud computing services are scalable because they scale up and down as needed without any cost and they offer processing and storage capacity. They are reliable since applications, documents and data are accessible anywhere in the world via the Internet [13].

Cloud computing services are integrated within online social networking sites in a variety of forms. Typically, cloud platforms are used to host social networks or scalable applications being created and hosted in the cloud. Cloud computing drives many social networking sites that has been accessed virtually by millions of users every day. When the user stores photographs to Flickr, posts them to Facebook or uploads a video to YouTube, the corresponding media are stored in the cloud. For instance, Facebook which is the most popular social networking site, provides scalable cloud based applications hosted by Amazon Web Services (AWS) [3]. The data related to a user is kept in a backend database and the user accesses the social cloud by the access control mechanism [49]. Social network forms a dynamic social cloud by enabling friends to share resources within the context of a social network. Users share their opinions, experiences in various topics and also are able to learn about other users' thoughts and experiences on a particular topic such as health. Social networks are substantial tools for healthcare. It helps patients to receive information and social support, ask advice from other patients with similar disease or medical experts. "e-health" term is used to describe the healthcare services and information delivered through the Internet. The main role of an e-health social network is to find other patients in similar situations and share information about treatments, symptoms and conditions. Some e-health services provide emotional support and some of them provide ability to ask questions to a medical expert [44]. The authors in chapter *Social Media: The Evolution of E-health Services*, analyze the e-health services provided by different Social Media, give an overview of different studies on Social Media in the healthcare sector and a description about the different activities and relationships on Social Media among physicians and patients. They introduce a Hybrid CLOUD E-health Services architecture (HCLES) which is capable of providing open, interoperable, scalable, and extensible services for all the e-health activities. Their proposed architecture integrates the use of tele-consulting service of Skype for a direct communication and synchronous data transmission and the cloud platform. The cloud platform consists of Infrastructure as a Service (IaaS) by providing a Web interface that allows patients and physicians accessing virtual machines; Software as a Service (SaaS) by providing all the e-health services offered by Social Media to patients and physicians; Platform as a Service (PaaS) by offering an integrated set of Social Media that provides an online environment for quick communication and collaboration between patients and physicians. The proposed architecture provides the need of supporting existing communities and facilitating their connection by Skype, and creating communities of interests of patients, physicians or hybrid communities that provide/receive emotional and

psychological support, medical support, medical information, health care education, and tele-consulting.

5 Conclusion

Recently, SNA has attracted a significant interest in many fields such as sociology, epidemiology, anthropology, social psychology, communication studies, information sciences, etc. This chapter provides definitions of the basic concepts of SNA and briefly introduces the topics in the book. A vast number of topics exist in the SNA field, therefore it is not possible to cover all of them comprehensively. However, this book includes most of the significant works and achieves the following:

- presents background on social networks and SNA.
- reviews the related works and their outcomes obtained on the addressed topics.
- demonstrates various important applications and studies in the areas of social networks, social community mining, social behavior and network analysis.

Through these, the book aims to introduce readers to the area of SNA and to become a reference book for academicians and industrial practitioners.

References

1. Aggarwal CC, Yu PS (2008) An introduction to privacy-preserving data mining. In: Privacy-preserving data mining, vol 34. Springer, Berlin, pp 1–9 [ISBN: 978-0-387-70991-8]
2. Alexa traffic statistics for Facebook (2013) <http://www.alexa.com/siteinfo/facebook.com>
3. Amazon, (2013) Building facebook applications on aws website. <http://aws.amazon.com/solutions/global-solution-providers/facebook/>
4. Balkundi P, Harrison DA (2006) Ties, leaders, and time in teams: strong inference about network structures effects on team viability and performance. *Acad Manage J* 49(1):49–68
5. Barnes JA (1954) Class and committee in a Norwegian island parish. *Hum Relat* 7:39–58
6. Bonacich P, Lloyd P (2001) Eigenvector-like measures of centrality for asymmetric relations. *Soc Netw* 23:191–201. doi:10.1016/S0378-8733(01)00038-7
7. Cai D, Shao Z, He X, Yan X, Han J (2005) Community mining from multi-relational networks. In: Jorge A, Torgo L, Brazdil P, Camacho R, Gama J (eds) PKDD. Springer, Berlin, pp 445–452 [ISBN: 3-540-29244-6]
8. Cheek GP, Shehab M (2012) Policy-by-example for online social networks. In: SACMAT. ACM, New York, S. 23–32 [ISBN 978-1-4503-1295-0]
9. Dai BT, Chua FCT, Lim E-P (2012) Structural analysis in multi-relational social networks. In: SDM. SIAM/Omnipress, California, USA, pp 451–462 [ISBN: 978-1-61197-232-0]
10. Fong PWL, Siahaan I (2011) Relationship-based access control policies and their policy languages. In: SACMAT. ACM, New York, S. 51–60 [ISBN 978-1-4503-0688-1]
11. Fortunato S (2010) Community detection in graphs. *Phys Rep* 486:75–174
12. Freeman LC (1979) Centrality in social networks: conceptual clarification. *Soc Netw* 1:215–239. doi:10.1016/0378-8733(78)90021-7

13. Motta G, Sfondrini N, Sacco D (2012) Cloud computing: an architectural and technological overview. In: 2012 international joint conference on service sciences (IJCSS). IEEE, New York, pp 23–27
14. Bastian M, Heymann S, Jacomy M (2009) Gephi: an open source software for exploring and manipulating networks. In: Proceedings of the AAAI international conference on weblogs and social media (ICWSM'09)
15. Gilks W, Spiegelhalter D (1996) Markov chain Monte Carlo in practice. Chapman & Hall/CRC, London/Boca Raton
16. Girvan M, Newman MEJ (2002) Community structure in social and biological networks. *Proc Natl Acad Sci* 99:7821–7826
17. Granovetter M (1973) The strength of weak ties. *Am J Sociol* 78:1360–1380
18. Hagberg AA, Schult DA, Swart PJ (2008) Exploring network structure, dynamics, and function using networkX. In: Proceedings of the 7th python in science conference, Pasadena, CA, USA, pp 11–15. <http://networkx.lanl.gov/>
19. Csardi G, Nepusz T (2006) The igraph software package for complex network research. *Inter-Journal, Complex Systems*. http://www.interjournal.org/manuscript_abstract.php?361100992
20. Kozlowski SWJ, Bell BS (2003) Work groups and teams in organizations. In: Borman WC, Ilgen DR, Klimoski R (eds) *Handbook of psychology: industrial and organizational psychology*, vol 12. Wiley, New York, pp 333–375
21. Kozlowski SWJ, Ilgen DR (2006) Enhancing the effectiveness of work groups and teams. *Psychol Sci Public Interest* 7(3):77–124
22. Liu K, Terzi E (2008) Towards identity anonymization on graphs. In: Wang JT-L (ed) *SIGMOD Conference*, pp 93–106. ACM, New York [ISBN: 978-1-60558-102-6]
23. Lovasz L (1993) Random walks on graphs: a survey, *Bolyai Soc. Math. Stud. Keszthely, Hungary*, 2:1–46
24. Mayo E (1945) *The social problems of an industrial civilization*. Routledge and Kegan Paul, London
25. Metropolis N, Rosenbluth AW, Rosenbluth MN, Teller AH, Teller E (1953) Equation of state calculations by fast computing machines. *J Chem Phys* 21:1087–1092. doi:10.1063/1.1699114
26. Milgram S (1967) The small world problem. *Psychol Today* 22:61–67
27. Moody J, White D (2003) Structural cohesion and embeddedness: a hierarchical concept of social groups. *Am Sociol Rev* 68(1):103–127
28. Moreno J (1934) *Who shall survive?* Beacon Press, New York
29. Narayanan A, Shmatikov V (2009) De-anonymizing social networks. In: *IEEE symposium on security and privacy*, IEEE Computer Society, pp 173–187
30. Neville J, Jensen D (2000) *Iterative classification in relational data*. AAAI Press, Texas, USA, pp 13–20
31. Nieminen J (1974) On centrality in a graph. *Scand J Psychol* 15:322–336
32. O'Madadhain J, Fisher D, White S, Boey Y (2003) The JUNG (java universal Network/Graph) framework. Technical report, UCI-ICS
33. Page L, Brin S, Motwani R, Winograd T (1998) The PageRank citation ranking: bringing order to the web. Stanford Digital Library Technologies Project, California, USA
34. Palla G, Barabasi A-L, Vicsek T (2007) Quantifying social group evolution. *Nature* 446:664–667. doi:10.1038/nature05670
35. Pfitzmann A, Hansen M (2010) A terminology for talking about privacy by data minimization: anonymity, unlinkability, undetectability, unobservability, pseudonymity, and identity management (version 0.34 vom 10. August 2010)
36. Pons P, Latapy M (2005) Computing communities in large networks using random walks (long version). In: *Computer and Information Sciences-ISCIS 2005*, pp 284–293
37. Raghavan UN, Albert R, Kumara S (2007) Near linear time algorithm to detect community structures in large-scale networks. *Phys Rev E* 76:036106
38. Rasti AH, Torkjazi M, Rejaie R, Duffield NG, Willinger W, Stutzbach D (2009) Respondent-driven sampling for characterizing unstructured overlays. In: *INFOCOM*. IEEE, New York, pp 2701–2705

39. Reagans R, Zuckerman EW (2001) Networks, diversity, and productivity: the social capital of corporate R&D. *Teams Organ Sci* 12(4):502–517
40. Salganik MJ, Heckathorn DD (2004) Sampling and estimation in hidden populations using respondent-driven sampling. *Sociol Methodol* 34:193–239. doi:10.2307/3649374
41. Sanna L, Parks C (1997) Group research trends in social and organizational psychology: whatever happened to intragroup research? *Psychol Sci* 8:261–267
42. Scott J (2005) *Social network analysis: a handbook*. 2nd edn. Sage Publications, London
43. Stocker K (2013) Battle of the social network climbers. <http://www.drapersonline.com/news/ecommerce/webwatch/battle-of-the-social-network-climbers/5052344.article>
44. Swan M (2009) Emerging patient-driven health care models: an examination of health social networks, consumer personalized medicine and quantified self-tracking. *Int J Environ Res Public Health* 6(2):492–525
45. Verykios VS, Bertino E, Fovino IN, Provenza LP, Saygin Y, Theodoridis Y (2004) State-of-the-art in privacy preserving data mining. *SIGMOD Rec* 33:50–57
46. Warner WL, Lunt PS (1941) *The social life of a modern community*. Yale University Press, New Haven
47. Wasserman S, Faust K, Iacobucci D, Granovetter M (1994) *Social network analysis: methods and applications*. Cambridge University Press, Cambridge
48. Watts D, Strogatz S (1998) Collective dynamics of ‘small-world’ networks. *Nature* 393(6684):440–442
49. Wooten R, Klink R, Sinek F, Bai Y, Sharma M (2012) Design and implementation of a secure healthcare social cloud system. In: *CCGRID. IEEE*, New York, pp 805–810 [ISBN: 978-1-4673-1395-7]
50. Ye S, Lang J, Wu F (2010) Crawling online social graphs. In: *2010 12th International Asia-Pacific Web Conference (APWEB)*, pp 236–242