Different Aspects of Social Network Analysis

Mohsen Jamali and Hassan Abolhassani
Web Intelligence Research Laboratory
Computer Engineering Department
Sharif University of Technology, Tehran, Iran
m_jamali@ce.sharif.edu, abolhassani@sharif.edu

***Abstract*— A social network is a set of people (or organizations or other social entities) connected by a set of social relationships, such as friendship, co-working or information exchange. Social network analysis focuses on the analysis of patterns of relationships among people, organizations, states and such social entities. Social network analysis provides both a visual and a mathematical analysis of human relationships. Web can also be considered as a social network. Social networks are formed between Web pages by hyperlinking to other Web pages. In this paper a state of the art survey of the works done on social network analysis ranging from pure mathematical analyses in graphs to analyzing the social networks in Semantic Web is given. The main goal is to provide a road map for researchers working on different aspects of Social Network Analysis.**

I. Introduction

A social network is a social structure between actors, mostly individuals or organizations. It indicates the ways in which they are connected through various social familiarities ranging from casual acquaintance to close familiar bonds[1]. Email traffic, disease transmission, and criminal activity can all be modeled as social networks.

Social network analysis is the mapping and measuring of relationships and flows between people, groups, organizations, animals, computers or other information/knowledge processing entities. The nodes in the network are the people and groups, while the links show relationships or flows between the nodes. Social network analysis provides both a visual and a mathematical analysis of human relationships. Management consultants use this methodology with their business clients and call it Organizational Network Analysis. One of the most interesting things about social structures is their substructure in terms of groupings or cliques. The number, size, and connections among the sub-groupings in a network can tell us a lot about the likely behavior of the network as a whole. How fast will things move across the actors in the network? Will conflicts most likely involve multiple groups, or two factions? To what extent do the sub-groups and social structures overlap one another?[1] All of these aspects of sub-group structure can be very relevant to predicting the behavior of the network as a whole.

Social network data consist of various elements. Following the definition by Wasserman and Faust[2], social network data can be viewed as a social relational system characterized by a set of actors and their social ties. Additional information in the form of actor attribute variables or multiple relations can be part of the social relational system. Social network data can be collected in various ways. The most common approach is by means of questionnaires, but also interviews, observations, and secondary sources are frequently used network data collection methods.[3]

In section 2, we describe different models for visualizing and analysis of social networks. Some of the most important properties of social networks is discussed in detail in section 3. In section 4 substructures and groups in social networks are illustrated. Web can be considered as a social network. A discussion on it is given in sections 5 and 6. Weblogs as special subsets of Web can also be considered as Social Networks. We discuss about Weblogs's social network in section 7. The Semantic Web (SW) is an emerging concept that launches the idea of having data on the Web defined and linked in a way that it can be used by people and processed by machines [4] [5] [6] in a "wide variety of new and exciting applications"[5]. We discuss about Semantic Web analytics on social networks and their effect on each other in section 8. Finally we have conclusions and our approaches for future works.

II. Social Network Models

A. Using formal methods to show Social Networks

One reason for using mathematical and graphical techniques in social network analysis is to represent the descriptions of networks compactly and systematically. A related reason for using (particularly mathematical) formal methods for representing social networks is that mathematical representations allow us to apply computers to the analysis of network data. The third, and final reason for using "formal" methods (mathematics and graphs) for representing social network data is that the techniques for graph processing and the rules of mathematics themselves suggest things that we might look for in our data [1].

In the analysis of complete networks, a distinction can be made between

- descriptive methods, also through graphical representations (see [7])
- analysis procedures, often based on a decomposition of the adjacency matrix
- statistical models based on probability distributions

B. Using Graphs to Represent Social Relations

Network analysis uses (primarily) one kind of graphic display that consists of points (or nodes) to represent actors and lines (or edges) to represent ties or relations. When sociologists borrowed this way of graphing things from the mathematicians, they renamed their graphs as ”sociograms”.

There are a number of variations on the theme of sociograms, but they all share the common feature of using a labeled circle for each actor in the population we are describing, and line segments between pairs of actors to represent the observation that a tie exists between the two. Visualization by displaying a sociogram as well as a summary of graph theoretical concepts provides a first description of social network data. For a small graph this may suffice, but usually the data and/or research questions are too complex for this relatively simple approach.

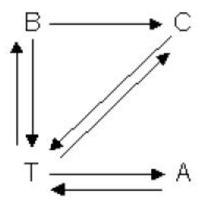

Fig. 1. Using Graphs to Represent Social Relations

C. Using Matrices to Represent Social Relations

The most common form of matrix in social network analysis is a very simple one composed of as many rows and columns as there are actors in our data set, and where the elements represent the ties between the actors. The simplest and most common matrix is binary. That is, if a tie is present, a one is entered in a cell; if there is no tie, a zero is entered. This kind of a matrix is the starting point for almost all network analysis, and is called an ”adjacency matrix” because it represents who is next to, or adjacent to whom in the ”social space” mapped by the relations that we have measured. By convention, in a directed graph, the sender of a tie is the row and the target of the tie is the column. Let’s look at a simple example. The directed graph of friendship choices among Bob, Carol, Ted, and Alice looks like figure 1. Since the ties are measured at the nominal level (that is, the data are binary choice data), we can represent the same information in a matrix that looks like figure 2 [1]:

	Bob	Carol	Ted	Alice
Bob	---	1	0	0
Carol	1	---	1	0
Ted	1	1	---	1
Alice	0	0	1	---

Fig. 2. Using Matrices to Represent Social Relations

D. Statistical Models for Social Network Analysis

Statistical analysis of social networks spans over 60 years. Since the 1970s, one of the major directions in the field was to model probabilities of relational ties between interacting units (social actors), though in the beginning only very small groups of actors were considered. Extensive introduction to earlier methods is provided by Wasserman and Faust [2]. Two of the most prominent current directions are Markov Random Fields (MRFs) introduced by Frank and Strauss [8] and Exponential Random Graphical Models (ERGMs), also known as p^* [9] [10]. The ERGM have been recently extended by Snijders et al [11] in order to achieve robustness in the estimated parameters.

The statistical literature on modeling Social Networks assumes that there are n entities called actors and information about binary relations between them. Binary relations are represented as an $n \times n$ matrix Y , where Y_{ij} is 1, if actor i is somehow related to j and 0 otherwise. For example, Y_{ij} = 1 if i considers j to be friend. The entities are usually represented as nodes and the relations as arrows between the nodes. If matrix Y is symmetric, then the relations are represented as undirected arrows. More generally Y_{ij} can be valued and not just binary, representing the strength (or value) of the relationship between actors i and j [12]. In addition, each entity can have a set of characteristics x_i such as their demographic information. Then the n dimensional vector X = x_1,...,x_n is a fully observed covariate data that is taken into account in the model (e.g. [13]).

There are several useful properties of the stochastic models. Some of them are:

- The ability to explain important properties between entities that often occur in real life such as reciprocity, if i is related to j then j is more likely to be somehow related to i; and transitivity, if i knows j and j knows k, it is likely that i knows k.
- Inference methods for handling systematic errors in the measurement of links [14].
- General approaches for parameter estimation and model comparison using Markov Chain Monte Carlo methods (e.g. [15]).
- Taking into account individual variability [16] and properties (covariates) of actors [13].
- Ability to handle groups of nodes with equivalent statistical properties [17].

There are several problems with existing models such as degeneracy analyzed by [18] and scalability mentioned by several sources [13] [19]. The new specifications for the Exponential Random Graph Models proposed in [11] attempt to find a solution for the unstable likelihood by proposing slightly different parameterization of the models than was used before.

III. Social Network Properties

There are some properties of social networks that are very important such as size, density, degree, reachability, distance, diameter, geodesic distance. Here we describe some more complicated properties which may be used in social network analysis. The following properties are taken from [1].

A. Maximum flow

One notion of how totally connected two actors are, asks how many different actors in the neighborhood of a source lead to pathways to a target. If I need to get a message to you, and there is only one other person to whom I can send this for retransmission, my connection is weak - even if the person I send it to may have many ways of reaching you. If, on the other hand, there are four people to whom I can send my message, each of whom has one or more ways of retransmitting my message to you, then my connection is stronger. This ”flow” approach suggests that the strength of my tie to you is no stronger than the weakest link in the chain of connections, where weakness means a lack of alternatives.

B. Hubbell and Katz cohesion

The maximum flow approach focuses on the vulnerability or redundancy of connection between pairs of actors - kind of a ”strength of the weakest link” argument. As an alternative approach, we might want to consider the strength of all links as defining the connection. If we are interested in how much two actors may influence on one another, or share a sense of common position, the full range of their connections should probably be considered.

Even if we want to include all connections between two actors, it may not make a great deal of sense (in most cases) to consider a path of length 10 as important as a path of length 1. The Hubbell and Katz approaches count the total connections between actors (ties for undirected data, both sending and receiving ties for directed data). Each connection, however, is given a weight, according to it’s length. The greater the length, the weaker the connection.

C. Taylor’s Influence

The Hubbell and Katz approach may make most sense when applied to symmetric data; because they pay no attention to the directions of connections (i.e. A’s ties directed to B are just as important as B’s ties to A in defining the distance or solidarity – closeness– between them). If we are more specifically interested in the influence of A on B in a directed graph, the Taylor influence approach provides an interesting alternative.

The Taylor measure, like the others, uses all connections, and applies an attenuation factor. Rather than standardizing on the whole resulting matrix, however, a different approach is adopted. The column marginals for each actor are subtracted from the row marginals, and the result is then normed. Translated into English, we look at the balance between each actor’s sending connections (row marginals) and their receiving connections (column marginals). Positive values then reflect a preponderance of sending over receiving to the other actor of the pair -or a balance of influence between the two-.

D. Centrality and Power

All sociologists would agree that power is a fundamental property of social structures. There is much less agreement about what power is, and how we can describe and analyze its causes and consequences. Table I summarizes some of the main approaches that social network analysis has developed to study power, and the closely related concept of centrality.

Power Aspect Name	Definition	Influences
Degree	Number of ties for an actor	Having more oppurtunities and alternatives
Closeness	Length of paths to other actors	Direct bargaining and exchange with other actors
Betweenness	Lying between each other pairs of actors	Brokering contacts among actors to isolate them or prevent connections

TABLE I
COMPARING THREE ASPECTS OF POWER IN SOCIOGRAMS (DEGREE, CLOSENESS, AND BETWEENNESS)

IV. GROUPS AND SUBSTRUCTURES IN SOCIAL NETWORKS

One of the most common interests of structural analysts is in the ”sub-structures” that may be present in a network. Many of the approaches to understanding the structure of a network emphasize how dense connections are compounded and extended to develop larger $cliques$ or sub-groupings. Network analysts have developed a number of useful definitions for algorithms that identify how larger structures are compounded from smaller ones.

Divisions of actors into cliques or ”sub-groups” can be a very important aspect of social structure. It can be important in understanding how the network as a whole is likely to behave. For example, suppose the actors in one network form two non-overlapping cliques; and, suppose that the actors in another network also form two cliques, but that the memberships overlap (some people are members of both cliques). Where the groups overlap, we might expect that conflict between them is less likely than when the groups don’t overlap. Where the groups overlap, mobilization and diffusion may spread rapidly across the entire network; where the groups don’t overlap, traits may occur in one group and not diffuse to the other.

The main features of a graph, in terms of its cliques or sub-graphs, may be apparent from inspection:

- How separate are the sub-graphs (do they overlap and share members, or do they divide or factionalize the network)?
- How large are the connected sub-graphs? Are there a few big groups, or a larger number of small groups?
- Are there particular actors that appear to play network roles? For example, act as nodes that connect the graph, or who are isolated from groups?

A. Cliques

The idea of a $clique$ is relatively simple. At the most general level, a clique is a sub-set of a network in which the actors are more closely and intensely tied to one another than they are to other members of the network. In terms of friendship

ties, for example, it is not unusual for people in human groups to form $cliques$ on the basis of age, gender, race, ethnicity, religion/ideology, and many other things.
The strongest possible definition of a clique is some number of actors (more than two, usually three is used) who have all possible ties present among themselves [20]. A *Maximal Complete Sub-Graph* is such a grouping, expanded to include as many actors as possible.
The strict clique definition (maximal fully connected sub-graph) may be too strong for many purposes. It insists that every member or a sub-group have a direct tie with each and every other member. You can probably think of cases of $cliques$ where at least some members are not so tightly or closely connected.
There are two major ways that the $clique$ definition has been relaxed to try to make it more helpful and general. One alternative is to define an actor as a member of a clique if they are connected to every other member of the group at a distance greater than one. Usually, the path distance two is used. This corresponds to being "a friend of a friend". This approach to defining sub-structures is called N-*clique*, where N stands for the length of the path allowed to make a connection to all other members [21].

B. N-Clans

The N-*clique* approach tends to find long and stringy groupings rather than the tight and discrete ones of the maximal approach. In some cases, N-*cliques* can be found that have a property that is probably undesirable for many purposes: it is possible for members of N-*cliques* to be connected by actors who are not, themselves, members of the $clique$. For most sociological applications, this is quite troublesome. For these reasons, some analysts have suggested restricting N-*cliques* by insisting that the total span or path distance between any two members of an N-*clique* also satisfy a condition. The kind of a restriction has the effect of forcing all paths among members of an n-*clique* to occur by way of other members of the n-*clique* [1]. This approach is the N-$Clan$.

C. K-Plexes

An alternative way of relaxing the strong assumptions of the *Maximal Complete Sub-Graph* is to allow that actors may be members of a clique even if they have tiles to all but k other members [22]. For example, if A has ties with B and C, but not D; while both B and C have ties with D, all four actors could fall in $clique$ under the K-$Plex$ approach. This approach says that a node is a member of a $clique$ of size n if it has direct ties to $n - k$ members of that clique. The k-*plex* approach would seem to have quite a bit in common with the n-*clique* approach, but k-*plex* analysis often gives quite a different picture of the substructures of a graph. Rather than the large and stringy groupings sometimes produced by n-*clique* analysis, k-*plex* analysis tends to find relatively large numbers of smaller groupings. This tends to focus attention on overlaps and co-presence (centralization) more than solidarity and reach [1].

Substructure Name	Description
Clique	Actors who have all possible ties among themselves
N-Clique	Actors are connected to every member of the group at a maximum distance of N
N-Clans	N-Cliques that all paths among members occur by the way of other members of N-Clique
K-Plex	Clique in which actors have ties to all but k of members of the group
K-Core	Actors are connected to k of members of the group
Component	Parts of sociogram that are connected within bu disconnected with other components
Cut Points	Nodes which if removed, the structure becomes divided into un-connected systems
Block	The divisions into which cutpoints divide a graph
Lambda Set	Set of actors who if disconnected, would most greatly disrupt the flow among all of the actors

TABLE II
COMPARING DIFFERENT APPROACHES FOR DEFINING SUBSTRUCTURES AND GROUPS IN SOCIOGRAMS

V. THE WEB AS A SOCIAL NETWORK

The Web is an example of a social network. Social networks are formed between Web pages by hyperlinking to other Web pages. To leverage the existence of hyperlinks, we model the Web as a graph where vertices are Web pages and hyperlinks are edges. While Web pages may be similar in terms of textual or multimedia content, a hyperlink is usually an explicit indicator that one Web page author believes that anothers page is related or relevant.

The possibility to publish and gather personal information (such as the interests, works and opinions of our friends and colleagues) has been a major factor in the success of the Web from the beginning. Remarkably, it was only in the year 2003 that the Web has become an active space of socialization for the majority of users [23]. That year has seen the rapid emergence of a new breed of Web sites, collectively referred to as social networking services (SNS). The first-mover Friendster[1] attracted over 5 million registered users in the span of a few months [24], which was followed by Google and Microsoft starting or announcing similar services.

Although these sites feature much of the same content that appears on personal Web pages, they provide a central point of access and bring structure in the process of personal information sharing and online socialization. Following registration, these sites allow users to post a profile with basic information, to invite others to register and to link to the profiles of their friends. The system also makes it possible to visualize and browse the resulting network in order to discover

[1] http://www.friendster.com/

friends in common, friends thought to be lost or potential new friendships based on shared interests. (Thematic sites cater to more specific goals, such as establishing a business contact or finding a romantic relationship [23]).

A. Applying social network analysis to the Web

Starting in 1996, a series of applications of social network analysis were made to the Web graph, with the purpose of identifying the most authoritative pages related to a user query.

1) PageRank in Google: If one wanders on the Web for infinite time, following a random link out of each page with probability $1 - p$ and jumps to a random Web page with probability p, then different pages will be visited at different rates; popular pages with many in-links will tend to be visited more often. This measure of popularity is called PageRank [25], defined recursively as

$$PageRank(v) = p/N + (1-p)\sum_{u \to v} \frac{PageRank(u)}{OutDegree(u)}$$

where '$\to$' means "links to" and N is the total number of nodes in the Web graph. (The artifice of p is needed because the Web is not connected or known to be aperiodic, therefore the simpler eigenequation is not guaranteed to have a fuxed point.) The Google search engine simulates such a random walk on the Web graph in order to estimate PageRank, which is used as a score of popularity. Given a keyword query, matching documents are ordered by this score. Note that the popularity score is precomputed independent of the query, hence Google can be potentially as fast as any relevance-ranking search engine.

2) Hyperlink induced topic search (HITS): Hyperlink induced topic search [26] is slightly different: it does not crawl or pre-process the Web, but depends on a search engine. A query to HITS is forwarded to a search engine such as Alta Vista, which retrieves a subgraph of the Web whose nodes (pages) match the query. Pages citing or cited by these pages are also included. Each node u in this expanded graph has two associated scores hu and au, initialized to 1. HITS then iteratively assigns

$$a_v = \sum_{u \to v} h_u \;\; and \;\; h_u = \sum_{u \to v} a_v$$

where $\sum_u h_u$ and $\sum_v a_v$ are normalized to 1 after each iteration. The a and h scores converge respectively to the measure of a page being an authority, and the measure of a page being a hub. Because of the query-dependent graph construction, HITS is slower than Google. A variant of this technique has been used by Dean and Henzinger to find similar pages on the Web using link-based analysis alone [27]. They improve speed by fetching the Web graph from a connectivity server which has pre-crawled substantial portions of the Web [28].

VI. Inferring Communities in Web

Community formation is one of the important activities in the Web. The Web harbors a large number of communities. A community is a group of content creators that manifests itself as a set of interlinked pages. Given a large collection of pages our aim is to find potential communities in the Web[29]. The link structure of the www represents a considerable amount of latent human annotation, and thus offers a promising starting point for structural studies of the Web. There has been a growing amount of work directed at the integration of textual content and link information for the purpose of organizing [30] [31], visualizing [28] and searching [32] [33] [27] in hypermedia such as the www. We review approaches for identification of communities from link topology in this section.
One of the key distinguishing features of the algorithms we will consider is the degree of locality used for assessing whether or not a page should be considered a community member. On the one extreme are purely local methods which consider only the properties of the local neighborhood around two vertices to decide if the two are in the same community. Global methods operate at the other extreme, and essentially demand that every edge in a Web graph be considered in order to decide if two vertices are members of the same community. In the following subsections, first we review two local algorithms (Bibliographic Metrics and Bipartite Cores), and then a global one (HITS Communities).

A. Bibliographic Metrics

Figure 3 illustrates two complementary metrics known as bibliographic coupling and co-citation coupling. In the figure, we see that the two metrics count the raw number of out-bound or in-bound references, respectively, shared by two pages u and v. Both metrics were originally formulated to capture similarity between scientific literature [34] by comparing the amount of overlap between the bibliographies or referrers for two different documents [35].

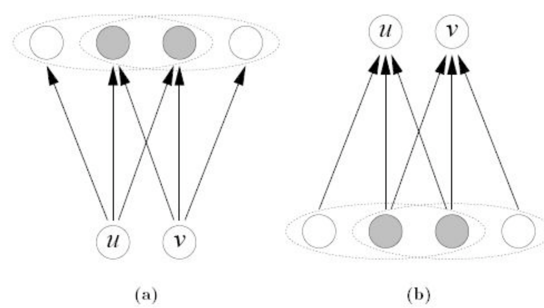

Fig. 3. Graphic portrayal of bibliographic metrics, (a) bibliographic coupling, and (b) co-citation coupling. For vertices u and v, the similarity metric is shown as the amount of overlap that vertices have in the set of outbound neighbors or in-bound neighbors.

B. Bipartite Cores

Bipartite cores are also local methods for inferring communities. Bibliographic metrics (especially when normalized) are effective for characterizing the degree of similarity between two pages in terms of what they link to and what links to them. What is missing in this framework is the notion that a collection of pages can be related to each other in an aggregate sense.

A *complete bipartite graph* is a directed graph with vertices that can be divided into two sets, L and R (for *left* and *right*) with $L \cup R = V$ and $L \cap R = \phi$, such that each vertex in L has an edge to each vertex in R. We use the notation, K_{lr} to denote a complete bipartite graph with $l = |L|$ and $r = |R|$.
Bipartite subgraphs are relevant to Web communities for at least two reasons that subtly relate to one another. First, a bipartite core, K_{lr}, has the properties that all vertices in L have a bibliographic coupling value lowerbounded by r and all vertices in R have a co-citation coupling value lowerbounded by l. Thus, bipartite subgraphs consist of vertices that have a minimal degree of similarity in terms of raw bibliographic metrics. The second reason why bipartite subgraphs are relevant to Web communities is because they empirically appear to be a signature structure of the core of a Web community [36].

C. HITS Communities

By utilization of HITS algorithm mentioned in previous section we can identify communities across documents. Note the crucial fact that the textual content of the pages involved is only considered in the initial step, when a root set is assembled from a search engine. Following this, the algorithm simply propagates weight over links without further regard to the relevance of the pages it is working with. The fact that hits can reliably identify pages that are not only authoritative but also relevant to the user's initial query implies something about the breadth of the topic: since the initial root set was sufficiently rich in relevant pages, the densest community of hubs and authorities in the surrounding base set was relevant as well. [37]

VII. Blogsphere as Social Networks

Recently, blogs (or Weblogs) have become prominent social media on the Internet that enable users to quickly and easily publish content including highly personal thoughts. A blog is typically a Web site that consists of dated entries in reverse chronological order written and maintained by a user (blogger) using a specialized tool. Since a blog entry can have hyperlinks to Web pages or other blog entries, the information structure of blogs and links (sometimes called the blogspace) can be seen as a network of multiple communities. As defined in Glossary of Internet Terms[2]:

"A blog is basically a journal that is available on the Web. The activity of updating a blog is ***blogging*** and someone who keeps a blog is a ***blogger***. Blogs are typically updated daily using software that allows people with little or no technical background to update and maintain the blog. Postings on a blog are almost always arranged in chronological order with the most recent additions featured most prominently."
Bloggers might list one another's blogs in a *blogroll* and might read, *link to a post*, or *comment* on other blogs' posts (A post is the smallest part of a blog which has some contents and readers can comment on it. A post also has a date of publish). Bloggers frequently read each other's postings, and the phenomenon of listing and commenting on information found through a user's online exploration is common. These posts and comments are intended to relay the latest interesting, humorous, or thought provoking information the user has run across. This information is added to the blog with the full realization by, or hope of, the author that it will be read by others.
Weblogs are subsets of Web and so can be considered as Social Networks. But their link structure may be somehow different if we consider comments and entry to entry links in blogs. Mohsen Jamali and Hassan Abolhassani [38] used this special link structure to extend HITS and introduce a new ranking algorithm form Weblogs. To measure Weblog popularity Gilad Mishne and Natalie Glance [39] used two indicators: the number of incoming links as reported by the Blogpulse index, and the number of page views for Weblogs that use a public visit counter such as Sitemeter. They made a great attempt to analyze the comments of blogs and the relation between the Weblog popularity and commenting patterns in it. [40] uses link structure in Weblogs to build a recommender system for Weblogs.
Cameron Marlow [41] employed social network analysis to describe the social structure of blogs. He has explored two measures of authority: popularity, as measured by Webloggers' public affiliations and influence measured by citation of each others writing. These metrics were evaluated with respect to each other and with the authority conferred by references in the popular press. Ko Fujimura et al. [42] proposed a new algorithm called "EigenRumor" that scores each blog entry by weighting the hub and authority scores of the bloggers based on eigenvector calculations. This algorithm enables a higher score to be assigned to the blog entries submitted by a good blogger but not yet linked to by any other blogs based on acceptance of the blogger's prior work. In the EigenRumor model, however, the adjacency matrix is constructed from agent-to-object links, not page-to-page (or object-to-object) links. An agent is used to represent an aspect of human being such as a blogger, and an object is used to represent any object such as a blog entity in this paper. Using the EigenRumor algorithm, the hub and authority scores are calculated as attributes of agents (bloggers) and by weighting these scores to the blog entries submitted by the blogger, the attractiveness of a blog entity that does not yet have any in-link submitted by the blogger can be estimated.

VIII. Semantic Web and Social Networks

There's a revolution occurring and it's all about making the Web meaningful, understandable, and machine-processable, whether it's based in an intranet, extranet, or Internet. This is called the Semantic Web, and it will transition us toward a knowledge-centric viewpoint of 'everything' [43]. The Semantic Web (SW) is an emerging concept that launches the idea of having data on the Web defined and linked in a way that it

[2]http://www.matisse.net/files/glossary.html

can be used by people and processed by machines [4] [44] [5] [6] in a "wide variety of new and exciting applications" [5]. It develops "languages for expressing information in a machine processable form" [4], so to enable the machine to be able to participate and help inside the information space [45].

The Semantic Web and social network models support one another. On one hand, the Semantic Web enables online and explicitly represented social information; on the other hand, social networks, especially trust networks [46], provide a new paradigm for knowledge management in which users "outsource" knowledge and beliefs via their social networks [47]. In order to turn these objectives into reality, many challenging issues need to be addressed as the following.

- **Knowledge representation.** Although various ontologies capture the rich social concepts, there is no need to have hundreds of "dialectic" ontologies defining the same concept. How can we move toward having a small number of common and comprehensive ontologies?
- **Knowledge management.** The Semantic Web is, relative the entire Web, fairly connected at the RDF graph level but poorly connected at the RDF document level. The open and distributed nature of the Semantic Web also introduces issues. How do we provide efficient and effective mechanisms for accessing knowledge, especially social networks, on the Semantic Web?
- **Social network extraction, integration and analysis.** Even with well-defined ontologies for social concepts, extracting social networks correctly from the noisy and incomplete knowledge on the (Semantic) Web is very difficult. What are the heuristics for integrating and fusing social information and the metrics for the credibility and utility of the results?
- **Provenance and trust aware distributed inference.** Provenance associates facts with social entities which are inter-connected in social network, and trust among social entities can be derived from social networks. How to manage and reduce the complexity of distributed inference by utilizing provenance of knowledge in the context of a given trust model? [48]

Despite their early popularity, users have later discovered a number of drawbacks to centralized social networking services. First, the information is under the control of the database owner who has an interest in keeping the information bound to the site. The profiles stored in these systems cannot be exported in machine processable formats, and therefore the data cannot be transferred from one system to the next. Second, centralized systems do not allow users to control the information they provide on their own terms. These problems have been addressed with the use of Semantic Web technology. The friend-of-a-friend(FOAF) project[3] is a first attempt at a formal, machine processable representation of user profiles and friendship networks.

[49],[50] show that the Friend of a Friend (FOAF) ontology is among the most used semantic Web ontologies. The Swoogle Ontology Dictionary shows that the class *foaf:Person*[4] currently has nearly one million instances spread over about 45,000 Web documents. The FOAF ontology is not the only one used to publish social information on the Web. For example, Swoogle identifies more than 360 RDFS or OWL classes defined with the local name "person".

Extracting social network from noisy, real world data is a challenging task, even if the information is already encoded in RDF using well defined ontologies. The process consists of three steps: discovering instances of *foaf:Person*, merging information about unique individuals, and linking person through various social relation properties such as *foaf:knows*. [48]

IX. Conclusions

In this paper we've reviewed social networks, formal methods to show them, and social networks' properties. Social network analysis methods provide some useful tools for addressing many aspects of social structure.

The Web itself can be considered as a social network. In the Web's social network, documens are node of the sociogram and links between documents are the edges of the sociogram. Weblogs, which are a special subset of Web could also be considered as social networks. We have described special link structure for Weblogs which contains comments other than explicit links.

The Semantic Web (SW) is an emerging concept that launches the idea of having data on the Web defined and linked in a way that it can be used by people and processed by machines. The Semantic Web and social network models support one another. Table V shows basic properties of kinds of social networks described in this paper, and shows differences in their formations.

As future works, we intend to mine the social networks of Persian Weblogs using the methods surveyed in this paper and find new interesting models. Also we're going to use semantics of those Weblogs and their link structure (their social network) to cluster the Weblogs using Semantic Web concepts.

Social Network Type	Actors	Ties	Direction
Friendship Netwotk	People in Society	Friendship relations between people	Undirected
Web's Social Network	Web Pages	Links Between Web Pages	Directed
Semantic Web Social Networks	Semantic Web Docs or Concepts in Them	Semantic Relations Between Documents or Concepts, such as $foaf:knows$.	Directed or Undirected

TABLE III

Comparing Basic Properties of Different Kinds of Social Networks Described. It should be noted that all of these networks can be analyzed by graph theory algorithms

[3]http://www.foaf-project.org/

[4]it is the Qualified name (QName) of http://xmlns.com/foaf/0.1/Person.

REFERENCES

[1] A. Hanneman and M. Riddle, "Introduction to social network methods," online at http://www.faculty.ucr.edu/ hanneman/nettext/, 2005.

[2] S. Wasserman and K. Faust, *Social Network Analysis: Methods and Applications*. Cambridge University Press, Novemver 1994.

[3] R. L. Breiger, *The Analysis of Social Networks*. London: Sage Publications Ltd, 2004, pages 505-526 in Handbook of Data Analysis, edited by Melissa Hardy and Alan Bryman.

[4] T. Berners-Lee, "Semantic web road map," online at http://www.w3.org/DesignIssues/Semantic.html, 1998.

[5] A. Swartz and J. Hendler, "The semantic web: A network of content for the digital city," in *Proceedings of Second Annual Digital Cities Workshop*, Kyoto, Japan, October 2001.

[6] J. Hendler, T. Berners-Lee, and E. Miller, "Integrating applications on the semantic web," *Journal of the Institute of Electrical Engineers of Japan*, vol. 122, no. 10, pp. 676–680, 2002.

[7] L. Freeman, *Graphic techniques for exploring social network data*. New York: Cambridge University Press, 2005, pages 248-269 in 'Models and methods in social network analysis' edited by P.J. Carrington, J. Scott, and S. Wasserman.

[8] O. Frank and D. Strauss, "Markov graphs," *Journal of the American Statistical Association*, vol. 81, pp. 832–842, 1986.

[9] S. Wasserman and P. Pattison, "Logit models and logistic regression for social networks: I. an introduction to markov graphs and p^*," *Psychometrika*, vol. 61, pp. 401–425, 1996.

[10] C. Anderson, S. Wasserman, and B. Crouch, "A p^* primer: logit models for social networks," *Social Networks*, vol. 21, pp. 37–66, 1999.

[11] T. A. Snijders, P. E. Pattison, G. L. Robins, and M. S. Handcock, "New specifications for exponential random graph models," 2004.

[12] G. Robins, P. Pattison, and S. Wasserman, "Logit models and logistic regressions for social networks," *Psychometrika*, vol. 64, pp. 371–394, November 1999.

[13] P. Hoff, A. Raftery, and M. Handcock, "Latent space approaches to social network analysis," *Journal of the American Statistical Association*, vol. 97, pp. 1090–1098, 2002.

[14] C. Butts, "Network inference, error, and informant (in)accuracy: a bayesian approach," *Social Networks*, vol. 25, no. 2, pp. 103–140, 2003.

[15] T. Snijders, "Markov chain monte carlo estimation of exponential random graph models," *Journal of Social Structure*, 2002.

[16] P. Hoff, "Random effects models for network data," Irvine, California, November 2003.

[17] Y. Wang and G. Wong, "Stochastic blockmodels for directed graphs," vol. 82, no. 8-19, 1987.

[18] M. Handcock, "Assessing degeneracy in statistical models of social networks," December 2003, working paper, University of Washington.

[19] P. Smyth, "Statistical modeling of graph and network data," in *Proceedings of IJCAI Workshop on Learning Statistical Models from Relational Data*, Acapulco, Mexico, August 2003.

[20] D. S. Johnson and M. A. Trick, *Cliques, Coloring, and Satisfiability: Second Dimacs Implementation Challenge*. American Mathematical Society, October 1996.

[21] "Connectivity and generalized cliques in sociometric group structure," *Psychometrika*, vol. 15, pp. 169–190, 1950.

[22] M. Everett, "Graph theoretic blockings, k−plexes and k−cutpoints," *Journal of Mathematical Sociology*, vol. 9, pp. 75–84, 1982.

[23] P. Mika, "Flink: Semantic web technology for the extraction and analysis of social networks," *Journal of Web Semantics*, vol. 3, pp. 211–223, October 2005.

[24] L. Kahney, "Making friendsters in high places," July 2003, the Wired News.

[25] S. Brin and L. Page, "The anatomy of a large-scale hypertextual web search engine," in *Proceedings of the 7th World-Wide Web Conference (WWW7)*, BrisBane, Australia, April 1998.

[26] J. Kleinberg, "Authoritative sources in a hyperlinked environment," in *Proceedings of Ninth Annual ACM-SIAM Symposium on Discrete Algorithms*, San Francisco, California, January 1998.

[27] J. Dean and M. R. Henzinger, "Finding related pages in the world wide web," in *In Proceedings of the 8th World Wide Web Conference*, Toronto, Canada, May 1999.

[28] K. Bharat, A. Broder, M. Henzinger, P. Kumar, and S. Venkatasubramanian, "The connectivity server: Fast access to linkage information on the web," Brisbane, Australia, April 1998.

[29] P. Reddy and M. Kitsuregawa, "Inferring web communities through relaxed cocitation and dense bipartite graphs," in *Proceedings of the Second International Conference on Web Information Systems Engineering (WISE2001)*, Kyoto, Japan, December 2001.

[30] R. A. S. Chakrabarti, B. Dom and P. Raghavan, "Using taxonomy, discriminants, and signatures to navigate in text databases," in *Proceedings of 23rd International Conference on Very Large Data Bases*, Athens, Greece, August 1997.

[31] J. Allan, "Automatic hypertext link typing," in *Proceedings of 7th ACM Conference on Hypertext, Hypertext '96*, Washington DC, March 1996, pp. 42–51.

[32] R. Agrawal, J. Gehrke, D. Gunopulos, and P. Raghavan, "Automatic subspace clustering of high dimensional data for data mining applications," in *In SIGMOD Conference on Management of Data*, Seattle, USA, June 1998.

[33] S. Chakrabarti, B. Dom, S. R. Kumar, P. Raghavan, S. Rajagopalan, A. Tomkins, D. Gibson, and J. Kleinberg, "Mining the web's link structure," *IEEE Computer*, vol. 32, no. 8, pp. 60–67, August 1999.

[34] H. G. Small, "Co-citation in the scientific literature: A new measure of the relationship between two documents," pp. 265–269, 1973.

[35] E. Garfield, *Citation Indexing: Its Theory and Application in Science*. New York: Wiley, 1979.

[36] R. Kumar, P. Raghavan, S. Rajagopalan, and A. Tomkins, "Trawling the web for emerging cyber-communities," in *Proceedings of 8th World Wide Web Conference*, Toronto, Canada, May 1999.

[37] D. Gibson, J. M. Kleinberg, and P. Raghavan, "Inferring web communities from link topology," in *UK Conference on Hypertext*, 1998, pp. 225–234.

[38] M. Jamali and H. Abolhassani, "Popularity and focality in weblogs' social network," submitted to The 2006 IEEE/WIC/ACM International Conference on Web Intelligence (WI-06), December 2006, Hong Kong.

[39] G. Mishne and N. Glance, "Leave a reply: An analysis of weblog comments," in *In WWW 2006 Workshop on the Weblogging Ecosystem: Aggregation, Analysis and Dynamics, 2006, at WWW06: the 15th World Wide Web Conference*, Edinburgh, Scotland, May 2006.

[40] K. S. Emaili, M. Neshati, M. Jamali, and H. Abolhassani, "Comparing performance of recommendation techniques in the blogsphere," in *In ECAI'06 Workshop on Recommender Systems*, Riva del Garda, Italy, August 2006.

[41] C. Marlow, "Audience, structure and authority in the weblog community." in *Proceedings of The 54th Annual Conference of the International Communication Association*, New Orleans, USA, May 2004.

[42] K. Fujimura, T. Inoue, and M. Sugisaki, "The eigenrumor algorithm for ranking blogs," in *In WWW 2005 Workshop on the Weblogging Ecosystem: Aggregation, Analysis and Dynamics, 2005, at WWW05: the 14th World Wide Web Conference*, Chiba, Japan, May 2005.

[43] M. C. Daconta, L. J. Obrst, and K. T. Smith, *The Semantic Web: A Guide to the Future of XML, Web Services, and Knowledge Management*. Wiley Publishing Inc, 2003.

[44] T. Berners-Lee, J. Hendler, and O. Lassila, "The semantic web," *Scientific American*, vol. 284, no. 5, pp. 34–43, May 2001.

[45] V. Benjamins, J. Contreras, O. Corcho, and A. Gomez-Perez, "Six challenges for the semantic web," in *Proceedings of International Semantic Web Conference (ISWC2002)*, Sardinia, Italia, 2002.

[46] J. Golbeck, B. Parsia, and J. Hendler, "Trust networks on the semantic web," in *Proceedings of Cooperative Intelligent Agents*, Helsinki, Finland, 2003.

[47] L. Ding, L. Zhou, and T. Finin, "Trust based knowledge outsourcing for semantic web agents," in *Proceedings of IEEE/WIC International Conference on Web Intelligence*, Halifax, Canada, October 2003.

[48] L. Ding, T. Finin, and A. Joshi, "Analyzing social networks on the semantic web," *IEEE Intelligent Systems*, vol. 9, no. 1, January 2005.

[49] L. Ding, L. Zhou, T. Finin, and A. Joshi, "How the semantic web is being used: An analysis of foaf," in *Proceedings of the 38th International Conference on System Sciences, Digital Documents Track (The Semantic Web: The Goal of Web Intelligence)*, Hawaii, USA, January 2005.

[50] L. Ding, T. Finin, A. Joshi, R. Pan, R. S. Cost, Y. Peng, P. Reddivari, V. C. Doshi, and J. Sachs, "Swoogle: A search and metadata engine for the semantic web," in *Proceedings of the Thirteenth ACM Conference on Information and Knowledge Management*, Washington, DC, November 2004.