On Social Network Web Sites: Definition, Features, Architectures and Analysis Tools

Vala Ali Rohani  
Department Of Software Engineering  
Faculty of Computer Science and Information Technology  
University Of Malaya  
Kuala Lumpur, Malaysia  
V.Rohani@perdana.um.edu.my

Ow Siew Hock  
Department Of Software Engineering  
Faculty of Computer Science and Information Technology  
University Of Malaya  
Kuala Lumpur, Malaysia  
Show@um.edu.my

Abstract

Development and usage of online social networking web sites are growing rapidly. Millions members of these web sites publicly articulate mutual "friendship" relations and share user-created contents, such as photos, videos, files, and blogs. The advances in web designing technology and fast growing usage of online resources prompted web designers to improve features and architectures of social web sites. This paper, with the experience of the authors in developing the Iranian Experts Social Network containing above 120,000 official members, describes the features of Social Web Sites (SWSs) and proposes a taxonomy and comprehensive definition in designing such social networking tools. We then present different SWS architectures and also introduce some useful tools for social network analysis.

Keywords: Social network web site, Social media web site, e-society, Social Network Analysis.

1. Introduction

Using IT Solutions and Web based applications in daily life of people cause the generation of wide range of social networks. The study of social networks is an almost natural approach to collaboration. In general, a network comprises of a set of nodes and a set of ties representing some relationship between the nodes. The nodes in social networks are most commonly individuals, organizations, or societies, and the ties often represent collaboration as the particular type of relationship between the nodes [6]. MySpace and Facebook are some examples of these new technology approaches and each claims more than 250 million registered users [1]. The growth of LinkedIn as a social networking website, demonstrates the impact of profile information very well. Its purpose is to help people build professional networks and find career development opportunities. Using LinkedIn, employers can look into the profile information of users to search for potential employees. Similarly, it helps employees look for potential employers [5].

The necessity of a national social network to support information interchange of Iranian experts motivated us to create the Iranian experts portal1. This project, having good infrastructure and skillful features, could attract more than 120,000 Iranian experts

to have their own official profile in it and also make their friends networks. Think tank with more than 300 daily questions and 1200 answers is another valuable outcome of this project.

There are hundreds of SNSs, with various technological affordances, supporting a wide range of interests and practices. While their key technological features are fairly consistent, the cultures that emerge around SNSs are varied. Most sites support the maintenance of pre-existing social networks, but others help strangers connect based on shared interests, political views, or activities. Some sites cater to diverse audiences, while others attract people based on common language or shared racial, religious, or nationality-based identities. Sites also vary in the extent to which they incorporate new information and communication tools, such as mobile connectivity, blogging, and photo/video-sharing [10].

In this paper, we organize the four aspects of social Web sites that we hope would serve as a comprehensive framework for understanding the features and future of social Web sites. The four aspects are definition of social Web sites, taxonomy of their essential features, some famous types of their architectures and finally, some tools for analyzing the content and relations in social web sites.

2. Social Network Sites: A Definition

In 1967, American psychologist Stanley Milgram [9] conducted the “small world experiment”, in which he sent letters to sixty volunteers in Kansas and asked them to forward the envelopes to a specific person in Massachusetts—by hand and through friends or friends of friends. The letters that reached the addressee were, on average, relayed by five to seven people. This is seen as an empirical proof that arbitrary people in our society are related to each other through friends and friends of friends.

The small world hypothesis based on Milgram’s findings states that the number of personal acquaintances needed to connect two random persons on the planet is small. The hypothesis led to the expression “the six degrees of separation”, meaning that any two random persons are associated with each other by a chain of about six individuals. The “six degrees of separation” is one of the underlying concepts of social networks on the Internet.

Social networking services offer friends a space where they can maintain their relationships, chat with each other and share information. Moreover, they offer the opportunity to build new relationships through existing friends. On the first use of the system, users are required to submit a profile containing personal information such as their name, date of birth, and a photo. The personal information is made available to other users of the system, and is used to identify friends on the network and to add them to a list of contacts. In most systems, users cannot only view their friends but also second degree friends (friends of their friends). Some networks follow an “invitation only” approach. Hence, every person in the system is automatically connected to at least one other person.

Danah M. Boyd defines social network sites as web-based services that allow individuals to (1) construct a public or semi-public profile within a bounded system, (2) articulate a list of other users with whom they share a connection, and (3) view and traverse their list of connections and those made by others within the system. The nature and nomenclature of these connections may vary from site to site [10].
Won Kim has proposed defined the social websites as those Websites that make it possible for people to form online communities, and share user-created contents (UCCs). The people may be the users of the open Internet or maybe restricted to those who belong to a particular organization (e.g., corporation, university, professional society, etc.). The community may be a network of offline friends (whose friendship is extended to online), online acquaintances, or one or more interest groups (based on school attended, hobby, interest, cause, profession, ethnicity, gender, age group, etc.). The UCC maybe photos, videos, bookmarks of web pages, user profiles, user’s activity updates, text (blog, micro blog, and comments), etc. The sharing of the UCC includes, at the minimum, the posting, viewing, and commenting of the UCC, and may also include voting on, saving, and retransmitting of the UCC. Roughly, He regards social Web sites as a union of social networking sites and social media sites.

The terms “social networking sites” and “social media sites” have already been loosely and widely used in press articles, blogs, press releases from the sites, etc., and the features of such sites are rapidly evolving. As such, we do not feel that efforts to define social Web sites (for that matter, social networking sites and social media sites as well) more precisely than above are warranted.

Roughly, social networking sites are Web sites that allow people to stay connected with other people in online communities (see [10, 11, 12] for attempts to define social networking sites rather more precisely). Some of the most widely used social networking sites in the world today include MySpace, Facebook, Windows Live Spaces, Habbo, etc. Social media sites are Web sites that allow people to share UCCs. Some of the most widely used social media sites include YouTube, Flickr, Digg, Metacafe, etc. Users of many of the popular social Web sites are dominated by teens and those in low twenties. The numbers of male and female users are roughly equal[13].


3. Social Network Sites Comparisons

Facebook now accounts for almost 60 percent of U.S. traffic to social networking sites, having increased its share of audience by 194 percent in September 2009 compared with September 2008. During the same period, the once dominant MySpace slumped to just 30 percent of overall visits, less than half of the 67 percent share it enjoyed in 2008.
Table 1. Top social networking sites by share of U.S. visitors

<table>
<thead>
<tr>
<th>Rank</th>
<th>Name</th>
<th>September 2009 (%)</th>
<th>September 2008 (%)</th>
<th>Year-over-year change (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Facebook</td>
<td>58.59</td>
<td>19.94</td>
<td>194</td>
</tr>
<tr>
<td>2</td>
<td>MySpace</td>
<td>30.26</td>
<td>66.84</td>
<td>-55</td>
</tr>
<tr>
<td>3</td>
<td>Tagged</td>
<td>2.38</td>
<td>1.62</td>
<td>47</td>
</tr>
<tr>
<td>4</td>
<td>Twitter</td>
<td>1.84</td>
<td>0.15</td>
<td>1170</td>
</tr>
<tr>
<td>5</td>
<td>myYearbook</td>
<td>1.05</td>
<td>1.76</td>
<td>-40</td>
</tr>
</tbody>
</table>

Additionally, Facebook saw its audience grow by around three percent month-on-month from August to September 2009, while MySpace experienced a drop of around three percent, suggesting its audience decline could continue. (Table 1)

Twitter's meteoric growth has resulted in its share of visits increasing by a whopping 1170 percent year-on-year, but it still accounts for just 1.8 percent of social networking traffic, thereby dwarfed by both Facebook and MySpace. It's important to note, however, that Hitwise only records traffic through the Twitter.com Web site, and does not include access through popular third-party applications, such as TweetDeck, which are likely to account for a large portion of Twitter's use [26].

Tagged received the third largest amount of visits during September 2009 with around 2 percent, and myYearbook rounded out the top five with around 1 percent.

Also, Figure 1 shows the TopTenReviews results in social networks comparison[25].

4. General Features of Social Networks

There are thousands of social Web sites, and the sites differ in details and layouts of the features they provide. Further, they continue to add new features and make changes to existing features. In this section, we will provide taxonomy of essential features that
can be extracted from today’s popular sites. To make the description of each feature concrete, we provide brief descriptions of how a few representative sites manifest the feature. We note that although all social Web sites support all these features, they manifest these in different ways and in varying degrees of sophistication.

- **Personal Profile**

Most social Web sites have members create and manage personal profiles, that is, homepages. However, they differ in the types of information included. Further, some sites have members specify privacy settings in order to control who (e.g., everyone, friends only) may access what types of information on their personal profiles. Personal profiles on social networking sites used to be more elaborate than those on social media sites. However, this distinction is becoming blurred, as social media sites are making the personal profile more elaborate [4].

- **Communicating With Online Connections**

Massage passing system is one of the friendly features of each e-community. Using this facility, the members of the same social network (also different ones) can send messages to each other and be alarmed by notification emails.

Social Web sites provide various facilities for members to use to communicate with their online connections, that is, friends and other members. These include email, instant messaging, text messaging, and public and private bulletin boards, and even Internet phone services. Such sites as MySpace and Facebook allow their members to use the messaging and phone call facilities of Internet phone services, such as Skype. Further, on behalf of the members, the sites send member updates and notices (e.g., friend request notice) using emails or text messages to members’ friends. They also send updates and notices to members of groups (e.g., notices regarding a group’s status and activities). The sites also display friends’ updates to members’ public and private boards. Some sites try to block spam by requiring the member to verify a Completely Automated Public Turing Test to tell Computers and Humans Apart (CAPTCHA) code before sending an email or text message.

For example, Twitter allows members to send messages to friends on what the members are doing. A message, which is restricted to 140 characters, may be sent to friends’ mobile devices and their Twitter accounts. A member may have one or more “followers,” and many members may be “following” a member. Facebook provides a mini bulletin board (called a “wall”) to post a member’s message for all friends to see and respond to. LinkedIn provides an “answers” function to allow members to answer questions posted by other members, and to refer the questions to their online connections [4].

- **Personal Bill Board**

Accessing the personal bill board through the profile control panel is one of the most favorite abilities of each social network. Each member can submit his (or her) daily events that want to announce in his (or her) profile.
For example, Twitter primarily allows the sharing of short text messages. The only non-text UCCs it allows are two photos of the member: one head and shoulder photo and one background photo. The member may save in his/her folder the links to selected (“favorite”) text messages. MySpace allows the sharing of blogs, photos, videos, playlists of music contents, etc. YouTube supports the posting and viewing of videos, including TV program clips, music videos, user-created video blogs, and short original videos. Members may add titles and tags when posting videos. YouTube adds the titles and tags to its search engine to support keyword-based search of the videos.

- **Thinking Room**

Most social Web sites allow members to leave comments on the UCCs. Some sites allow members to vote on them, too. The voting may take the form of ranking (e.g., checking 3 stars out of 5 stars), or marking the UCC as a “favorite,” or flagging it as spam or inappropriate. Some sites present all comments as a flat list of comments, while others organize comments as a 2-level hierarchy, that is, they allow comments on a comment. Some sites display the comments in posting timestamp order, and some in reverse timestamp order. Some sites give a few additional sorting options for displaying the comments, such as the number of comments, the number of emails that sent the UCC, the number of favorable votes received, etc.

- **e-Newsletter**

Users tend to glance at websites when they need to accomplish something or to find the answer to a specific question. In contrast, newsletters feel personal because they arrive in users’ inboxes, and users have an ongoing relationship with them. Newsletters also have a social aspect, as users often forward them to colleagues and friends.

The positive aspect of this emotional relationship is that newsletters can create much more of a bond between users and a company than a website can. The negative aspect is that newsletter usability problems have a much stronger impact on the customer relationship than website usability problems [16].

As a best practice in irexpert project, we find this fact that customers do not check their social network profile very regularly, but a majority of them check their mail box. Hence, we can encourage them to browse) our social network by sending them a user friendly newsletter.

This newsletter consists of the daily news that each member could submit in their social networks after confirming the related administrator. A web-based agent can gather daily news, generate pre-styled newsletter and then send it to the corresponding receivers.

- **Friendship Network**

Social Web sites provide various facilities for members to use to communicate with their online connections, that is, friends and other members. These include email, instant messaging, text messaging, and public and private bulletin boards, and even Internet phone services.
Each member also can invite others to make a friendship relation and develop his or friends’ network. This ability motivates members to check their profiles periodically.

- **Forums**

Many social Web sites support a small number of default groups and assign new members to one or more of them. They also allow members to explicitly form new groups, and/or join them. An explicit group is assigned a manager. Members and non-members can both view all the UCCs in all the groups. However, only members may post UCCs.

An Internet forum allows many people to chime in about their particular experiences, information, tips, tricks, etc. There are many computer forums out there which, if you have a problem with your computer, you can look up that problem and you might find 5 or 1 people who have been talking about that exact problem and how they are trying to solve it and potential solutions. A forum is an online communication between multiple users, generally through text-style communication – occasionally through voice. Internet forums really began as an off-shoot of the neighborhood bulletin board, where you would go and post your want ads or what you needed or maybe, try and get some business – basic places where communities would come together and find information about various topics. When the Internet came around, it made it much easier for people to find specific forums, specific bulletin board sites. Some of the popular ones, like Craigslist, enable you the ability to find anything you want, sell or buy things, hire people, find a job yourself, find a roommate, etc. It's a place for Internet correspondence and communication. It's a way for multiple users to give their tips, tricks, and information about how they're using something or what they've found in their day-to-day experiences. It's a way for information to be gathered and collected by every basic user about specific topics [17].

Forums prefer a premise of open and free discussion and often adopt de facto standards. Most common topics on forums include questions, comparisons and polls of opinion as well as debates. Because of their volatile and random behavior, it is not uncommon for nonsense or unsocial behavior to sprout as people lose temper. Poor understanding of argumentation theory and differences in values of the participants is a common problem of forums. Because replies to a topic are often wording aimed at someone's point of view, discussion will usually go slightly off into several directions as people question each other’s validity and sources.

5. Social Network Architecture

As huge social network site dramatically grew, the engineers had to solve three architecture challenges: performance, scalability, and availability. To understand and solve the bottlenecks on performance and scalability, the engineers have had to start using system and network monitoring tools, and Web traffic and log file analysis tools. The performance challenge is to provide real-time response to every user request. As the traffic to a site increases, the number of members increases, and/or the amount of data stored in the site increases, performance of the site is bound to degrade. The scalability challenge is to maintain the performance at the same level even as the traffic, number of members, and the amount of data stored increase. The availability challenge is to have the site provide all services and access to all the data stored to all users at all times. Fig.
4 shows the prototypical architecture of huge social networks, such as YouTube, MySpace, Facebook, Flickr, etc. Below, we describe the key components of the architecture, and techniques used in meeting the three architecture challenges.

![Architecture for huge social networks](image)

**Figure 2. An architecture for huge social networks**

### 5.1 Speeding up the Web Servers

Considering the important role of Web server, Site administrators need to ensure that the Web server does not become overloaded. The sites manage incoming requests by using various means, such as firewalls and HTTP traffic manager to block or redirect unwanted or ill-formed traffic, and a traffic shaper to smooth out peaks in requests. However, the engineers have determined that the key bottleneck for the Web servers is the time it takes waiting for responses to remote procedure calls (RPC). For higher performance, tremendous sites use Web server farms, consisting of multiple, more powerful computers with more main memory and hard disk drive capacity. Gigantic sites have also introduced a load balancer and/or reverse proxy servers in front of the Web server farm. A reverse proxy server is installed to load balance the Web server farm and to cache both static and dynamic content. It also provides an additional layer of security.

### 5.2 Use of Main-memories as Cache

Major sites, use memcaches extensively. Memcache is an in-memory hash table that is distributed across multiple servers. It has become an essential component of the architecture to meet the performance and scalability challenges. It listens to the TCP socket for requests. Hash tables are maintained in memcaches to rapidly respond to metadata keyword searches, such as the member ID. The size of memcaches is often over 16 Gigabytes. Further, some major sites, such as YouTube and MySpace, store most frequently accessed contents and less frequently accessed contents on separate server farms, serving most frequently accessed contents out of main memories of one server farm. In particular, MySpace employs more than 1200 cache servers [18].
5.3 Database Partitioning and Replication

Major sites have to deal with huge amounts of data of various types. The sites assign different server farms to manage different types of data. Major sites, partition the database based on services (e.g., member profiles, emails, news, videos, etc.). They use a different application server farm for each service. This naturally distributes the load on the application servers and database servers. Further, the sites keep multiple copies all the databases. This is to address the performance, scalability, and availability challenges.

It’s popular in huge social networks to extend database partitioning and replication to the level of relational tables. That is, they partition a huge relational into multiple smaller tables comprising of some of the columns (i.e., “vertical decomposition”) and/or rows (i.e., “horizontal decomposition”) of the original table. They also maintain multiple copies of some of the table partitions. For example, such sites as YouTube, Flickr, LinkedIn, and MySpace decompose the members table into separate partitions, and replicate each partition [18,19]. MySpace keeps one million users in each partition; and needs 300 partitions to store all 300 million registered members. Partitioning and replication can each help meet the performance, scalability, and availability challenges.

5.4 Use of Server Farms

Gigantic social network sites use server farms for each of the three key components of the site, rather than merely doubling it. For each server farm, there is a control server that manages distribution of work, load balancing, and dynamic reconfiguration of the server farm (i.e., adding a new server, taking out a server). The use of server farms helps meet the performance, scalability, and availability challenges.

5.5 Redundant Servers and Database Backup

As a major site consists of thousands or tens of thousands of commodity hardware, server failure occurs frequently. Major sites pair each server with a backup server. This helps to meet the availability challenges. When a server crashes, its backup takes over the processing load. All sites back up the database on a regular basis in order to recover from crashes of hard disk drives and servers. They back up the database in its entirety or selected parts once a day, a week, every two weeks, or a month.

6. Social Network Analysis

Social Network Analysis (SNA) is the study of social relations among a set of actors. The key difference between network analysis and other approaches to social science is the focus on relationships between actors rather than the attributes of individual actors. Network analysis takes a global view on social structures based on the belief that types and patterns of relationships emerge from individual connectivity and that the presence (or absence) of such types and patterns have substantial effects on the network and its constituents. In particular, the network structure provides opportunities and imposes constraints on the individual actors by determining the transfer or flow of resources (material or immaterial) across the network.
SNA is thus a different approach to social phenomena and therefore requires a new set of concepts and new methods for data collection and analysis. Network analysis provides a vocabulary for describing social structures, provides formal models that capture the common properties of all (social) networks and a set of methods applicable to the analysis of networks in general. The concepts and methods of network analysis are grounded in a formal description of networks as graphs. Methods of analysis primarily originate from graph theory as these are applied to the graph representation of social network data.

It is interesting to note that the formalization of network analysis has brought much of the same advantages that the formalization of knowledge on the Web (the SemanticWeb) is expected to bring to many application domains. Previously vaguely defined concepts such as social role or social group could now be defined on a formal model of networks, allowing to carry out more precise discussions in the literature and to compare results across studies.

The methods of data collection in network analysis are aimed at collecting relational data in a reliable manner. Data collection is typically carried out using standard questionnaires and observation techniques that aim to ensure the correctness and completeness of network data. Often records of social interaction (publication databases, meeting notes, newspaper articles, documents and databases of different sorts) are used to build a model of social networks [20].

This section reviews software for the statistical analysis of social networks. The programs UCINET, Pajek, NetMiner II and STRUCTURE are investigated in following. We start with three general packages, covering a wide range of analysis methods. They are presented according to age: UCINET, Pajek and NetMiner [21].

6.1. Ucinet

UCINET is a social network analysis program developed by Steve Borgatti, Martin Everett and Lin Freeman. The program is distributed by Analytic Technologies. UCINET works in tandem with freeware program called NETDRAW for visualizing networks. NETDRAW is installed automatically with UCINET.
UCINET is a comprehensive package for the analysis of social network data as well as other 1-mode and 2-mode data. It can read and write a multitude of differently formatted text files, as well as Excel files and is able to handle a maximum of 32,767 nodes (with some exceptions) although practically speaking many procedures get too slow around 5,000 - 10,000 nodes. Its social network analysis methods include centrality measures, subgroup identification, role analysis, elementary graph theory, and permutation-based statistical analysis. In addition, the package has strong matrix analysis routines, such as matrix algebra and multivariate statistics [22].

6.2 Pajek

Pajek is a windows-based program, for analysis and visualization of large networks having some thousands or even millions of vertices. The main goals in the design of Pajek are:

- to support abstraction by (recursive) decomposition of a large network into several smaller networks that can be treated further using more sophisticated methods.
- to provide the user with some powerful visualization tools.
- to implement a selection of efficient (subquadratic) algorithms for analysis of large networks.

Pajek has some features to find clusters (components, neighborhoods of ‘important’ vertices, cores, etc.) in a network, extract vertices that belong to the same clusters and show them separately, possibly with the parts of the context (detailed local view), shrink vertices in clusters and show relations among clusters (global view).

Besides ordinary (directed, undirected, mixed) networks, Pajek supports also multi-relational networks, 2-mode networks (bipartite (valued) graphs – networks between two disjoint sets of vertices), and temporal networks (dynamic graphs – networks changing over time) [23].

![Figure 4. Pajek software interface](image_url)
6.3 NetMiner

NetMiner is an innovative software tool for Exploratory Analysis and Visualization of Network Data with some abilities to explore network data visually and interactively, and also to detect underlying patterns and structures of the network.

The last version of this software is 3.4.0 which released on 23 Sep. 2009. In this new version, Layout and Drawing module that can be used for drawing large-scale 2D network map by separating the visualization process into two steps (layout and drawing) are now expanded to handle 3D network map. The coordinates of the 3D network map can be generated by Layout 3D module and drawing 3D network map using the generated coordinates can be accomplished by Drawing 3D module [24].

7. Future of Social Networks

Some reports predicts that growth in the number of people signing up to be a part of the cultural phenomenon, which has put the likes of Facebook on the map, will plateau by 2012. Growth in the membership of social-networking sites varies dramatically by region, which predicts Asia Pacific will account for 35 percent of global social networking users by the end of this year, followed by EMEA (28 percent), North America (25 percent), and the Caribbean and Latin America (12 percent).

![NetMiner software interface](image)

Figure 5. NetMiner software interface

Next generation Social Networking will present key insights not just in text, but a visually appealing manner; from transactions to transformation; from raw data to visual insights.

And finally, Enterprise Social Networking is being used to bring all the functionalities inside a company to support the key goals and strategic objectives of the organization.
References
