International Journal of Engineering Innovation & Research
Volume 1, Issue 6, ISSN : 2277 – 5668

Theoretical Concepts of Social Networks and Group Formation: A Survey

Karuna C. Gull
karuna7674@gmail.com

Padmashri Desai
padmashri@bvb.edu

Akshata Angadi
akshata_angadi@yahoo.co.in

***Abstract* — A social network is defined as a social structure of individuals, who are related (directly or indirectly to each other) based on a common relation of interest, e.g. friendship, trust, etc. Social network analysis is the study of social networks to understand their structure and behavior. The online social network field is broad, and any literature review can only focus on a selection of articles. The present article highlights recent research in the field and focuses on centrality, linkage strength, identity, trust, activity and benefits. By no means is this review comprehensive, but it should give practitioners some useful concepts to consider as they design social network based web applications. Thus the overall objective of the proposed survey is to come up with an aim to have the latest research and simultaneously discuss contributions of renowned researchers whose work created a revolution in the social networking area. The goal of the survey process is to produce new knowledge or deepen understanding of a same topic or an issue.**

***Keywords* — Behavior, Centrality, Learning, Life Cycle, Literature Review, Social Networks, Social Network Analysis and Mining (SNAM).**

I. INTRODUCTION

A. What Is Social Network?

Social networks consist of the most intelligent components – human beings. Being so, any activity involved with the social networks - be it participation, management, or optimization – becomes extremely complicated and context based. Due to the various facets of the human species, we can have multiple types of social networks in all the fields and areas. This can range from a network of physicists, to a network of doctors, to a network of soccer lovers. Each type of network has its own focus area, member size, geographical spread and societal impact. Managing such networks is not only complicated, but it requires a lot of collective effort and collaboration. Uncountable social networks have been formed, but only a few have finally achieved their true goal, which emphasizes the complexity of such a matter. A social network is usually created by a group of individuals who have a set of common interests and objectives. There are usually a set of network formulators followed by a broadcast to achieve the network membership. This advertising happens both in public and private groups depending upon the confidentiality of the network. After the minimum numbers are met, the network starts its basic operations and goes out to achieve its goal. Today's social networks also support formation of multiple groups inside the network on the basis of interests on specific areas of the network members. These groups run their own discussion threads, share viewpoints that are accessible to those group members only. Success of a social network mainly depends on contribution, interest, and motivation of its members along with technology backbone or platform support. This makes life easier to communicate and exchange information to fulfill a particular communication need. Implementing social networks and sustaining them is one of the biggest challenges, and people have formulated many mechanisms in the past to keep alive such networks. This has been largely supported by the advancements in the field of information technology.

B. Types and Behavior of Social Networks:

Social networks exist in various domains – within and outside the organizations, within and outside geographical boundaries, within and outside social boundaries, and many other areas. Such huge variations make the reach of social networks grow to all sectors of the society. Keeping these in mind, the main categories identified are given below (Fig. 1):

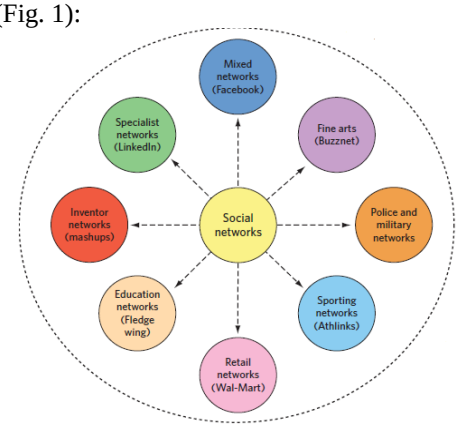

Fig.1. Different types of social networks

- Social contact networks: These types of networks are formed to keep contact with friends and family and are one of the most popular sites on the network today. They have all components of Web 2.0, such as blogging, tagging, Wikis, and forums. Examples of these include Orkut, Facebook, and Twitter.
- Study circles: These are social networks dedicated for students where they can have areas dedicated to student study topics, placement-related queries, and advanced research opportunity gathering. These have components like blogging and file sharing. Examples of these are FledgeWing and College Tonight.
- Social networks for professionals: These types of social networks are specifically designed for core fieldworkers, such as doctors, scientists, engineers, and members of the corporate industries. A very good example for this type of network is LinkedIn.

- Networks for fine arts: These types of social networks are dedicated to people linked with music, painting, and related arts and have lots of useful networking information for all aspiring people of the same line. Examples of such a network are Amie Street and Buzznet.
- Police and military networks: These types of networks, though not on a public domain, operate much like social networks on a private domain due to the confidentiality of information.
- Sporting networks: These types of social networks are dedicated to people of the sporting fraternity and have a gamut of information related to this field. An example of the same is Athlinks.
- Mixed networks: There are a number of social networks that have a subscription of people from all the above groups. It is a heterogeneous social network serving multiple types of social collaboration.
- Social networks for the 'inventors': These are the social networks for the people who have invented the concept of social networks, the very developers, and architects who have developed the social networks. Examples are Technical Forums and Mashup centers.
- Shopping and utility service networks: The present world of huge consumerism has triggered people to invest in social networks that will try to analyze the social behavior and send related information for the same to respective marts and stores.
- Others: Apart from the networks outlined above, multiple social networks serve a huge number of the Internet population in various ways. Some of these networks die out very fast due to lack of constructive sustenance thoughts while others finally migrate to a more specialist network.

C. Life Cycle of Social Networks

For any social network, there are a number of steps in its life cycle. In each step of the life cycle of an online social network, Web 2.0 concepts have a great influence. Consider the diagram below (Fig. 2). For all the steps in the life cycle, Web 2.0 has provided tools and concepts that are not only cost effective but very easy to implement. Oftentimes, online networks have a tendency to die out very fast due to lack of proper tools to communicate. Web 2.0 provides excellent communication mechanism concepts like blogging and individual email filtering to keep everyone in the network involved in its day-to-day activities.

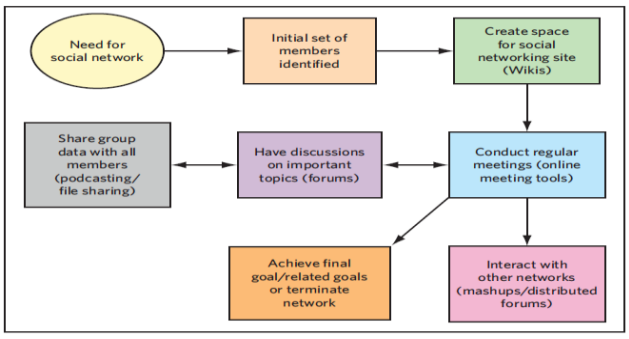

Fig.2. Life cycle of social networks

1.4 Structure and Functionalities of Social Networks

Here are some interesting facts about the social networking and its usage. Internet users spent 65% more time online in April 2010 (884M hours) than they did in April 2007 (536M hours). With the sector having grown to account for almost 23% of UK Internet time (176M hours), one in every four-and-a-half UK minutes online is accounted for by social networks and blogs. In mainland China, social networks are taking off at a tremendous speed. As of March 2010 comScore data, Facebook traffic made up 41% of all traffic on a list of popular social destinations. Social network advertisement revenue will reach $421 million, marking an overall 50% increase according to Informa. The above statistics just demonstrate how the social network scenario will be impacting the IT industry as a whole. Niche features and state-of-the-art concepts like Web 2.0 have given users a wide range of services and created a huge boom in social networking applications. This can be credited to the basic restructuring of the Web as a whole. Earlier the Web was a read-only platform modified by a select group of Webmasters (administrative users on the side of the website) with almost no opinion taken from the end users. The Web over the last decade has migrated to a platform in itself and has started a revolution where the end users are 'developing' on the Web itself with minimal intervention from the Webmasters.

Social networking applications have been able to greatly leverage from this paradigm shift of the Web, and all that a social networking application offers have been because of this evolution. The social networking applications of today have all been modeled on some common services to be provided to a user. Such services are broadly divided into the following:

- The ability to add friends/relatives/ colleagues using the same network.
- Post messages either to a contact or a group of contacts and receive messages in return.
- Upload and share files in the genre of music, video, or documents.
- Personalized home pages (such as Wikis) can be created by the users.
- Blogging capabilities
- Instant messaging capabilities with other users who are online in the same social network.
- Maintain online albums related to portraits
- Development of links in professional fields by featuring career details.
- Online gaming options among members of the same social network.

II. LITERATURE REVIEW

Humans suffer from information overload; there's much more information on any given subject than a person is able to access. As a result, people are forced to depend upon each other for knowledge. Johnson (Johnson et al., 2002) [7] mentioned about Know-who information rather

than know-what, know-how or know-why information has become most crucial. It involves knowing who has the needed information and being able to reach that person

In this context, understanding the formation, evolution and utilization of online social networks becomes important. According to Garton (Garton et al., 1997) [5] a social network is "a set of people (or organizations or other social entities) connected by a set of social relationships, such as friendship, cow or king or information exchange." While the Internet contributes to the information overload, it also provides useful tools to effectively manage one's social networks and through them gain access to the right pieces of information.

Social networks are graphs that are used to model and analyze the structure of relationships between a set of entities. The nature and scope of social network data, however, have grown well beyond traditional applications in sociology [9] and contemporary study in physics [10, 11]. Conventional social networks are now superseded by continuous streams of dynamic interaction data, or dynamic networks, which have opened the way to new techniques for analyzing the underlying populations [10-12, 13, 14, 15].

This field is of particular interest to researchers working at the intersection of information systems, sociology and mathematics. These researchers study the uses of social networks and the ways in which they are mediated in society and in the workplace through information communication technologies (ICTs) such as (but not limited to) the Internet. This literature review explores how social networks that take advantage of information communication technologies – specifically, web based technologies – begin, evolve and are utilized.

A. The Strength of Weak Ties

Social networks were first researched in the late 1940s. With the advent of the Internet, online communities and social networking websites, their significance has only increased. Any review hoping to be meaningful must begin with the normative contributions of the sociologist Mark Granovetter and the mathematician Linton C. Freeman who both wrote influential articles well before the Internet was popularized.

Granovetter (1973) [6] argued that within a social network, weak ties are more powerful than strong ties. He explained that this was because information was far more likely to be "diffused" through weaker ties. He concluded that weak ties are "indispensable to individuals' opportunities and to their incorporation into communities while strong ties breed local cohesion."

Granovetter's doctoral thesis demonstrated that most people landed jobs thanks to their weak ties and not their strong ones. It was the people that they did not know well, the ones with whom they did not have shared histories and did not see on a regular basis who were of most help. This is because people with strong ties generally share the same pieces of information and resources. Therefore they are of less help to one another.

Similarly, Granovetter identified absent ties (also called nodding ties) – those ties that lack the emotional intensity, time, intimacy and reciprocity to even qualify as weak ties. Someone living on the same street that you nod to everyday is an absent tie. An absent tie is someone that exists in your life but with whom you have no connection whatsoever. That person is not helpful in the way that a weak tie can be.

Depending upon the type of application you are building, you may want to design it so that people are encouraged to form weak ties with people that they do not know very well. They are more likely to benefit from those weak ties than from strong ones. But it is important to recognize the difference between a weak tie and an absent one. On social network sites like MySpace and Facebook, where self worth is garnered through the number of ties, the difference becomes important. Yet, the fact that you can search and connect to all kinds of ties on these networks has influenced their growth.

According to Granovetter's theory [6], there would be value in the visual depiction of weak ties. LinkedIn tells you how many ties you have at each degree of separation, but other than that you are not given much information about those ties. Are they strong, weak, or absent ties? LinkedIn has another problem too: It makes it difficult for you to connect with your weak ties. You often have to ask a common friend for permission to establish that connection. No wonder LinkedIn is being eclipsed by other social network services!

B. Centralization in a Network

An understanding of social networks needs also to include accounts of centrality and of one node's relationship to other nodes in a network. This is why Linton C. Freeman's article on centrality in social networks is important (Freeman, 1979). Freeman [4] explored how "graph centralization" was based on differences in point centralities. He also outlined three competing theories regarding the definition of centrality based on degree of a point, control and independence.

Degree of a point refers to the number of nodes connected to a given node. In simple terms, this means counting the number of friends you have in a social network. The more friends, you have, the more important you are.

Control refers to the extent to which nodes depend on one specific node to communicate with other nodes. For example, if hundreds of friends are connected to each other only when you serve as the bridge connecting them, then your centrality is high. You are the node that controls the communication flows.

And finally independence means that a node is closely related to all the nodes considered – so that it is minimally dependent on any single node and is not subject to control. This means you can reach the maximum number of people through the shortest number of links, without being dependent on a particular few nodes.

Centrality

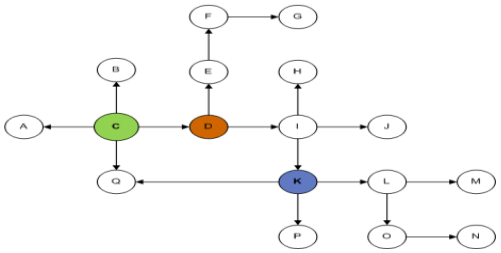

Fig.3. A depiction of centrality

- Degree point: C and K have the most nodes connected to them.
- Control: D serves as the bridge between the most nodes and controls the flow of information.
- Independence: K is most closely connected to the other nodes by multiple nodes (I and Q).

Because social networks are fundamentally social tools in which people are constantly monitoring and growing their social network, most social network media depict growth using the degree of point definition. However, control and independence can be more useful definitions. For example, a person who controls information flows is more important than one who may have more friends in the network. Centrality can also indicate which members are the most useful or well connected and therefore the best information resources.

C. Learning from Flickr & Yahoo!

The principles of node structures, tie strength and centrality have been applied to understand nodes in modern day online social networks. A good example of this is in the explanatory research conducted by Kumar, Novak and Tomkins (2006) [8]. They compared two online social networks, Flickr and Yahoo! 360, which together had more than five million users at the time. These researchers noticed that the social networks follow a standard pattern of growth, namely, rapid early growth followed a period of decline and then slow but steady growth. Kumar, Novak and Tomkins also saw that network activity is of three types:

- "Singletons," who have no connections and are least central
- The "giant component," which is the largest group of nodes tightly connected to the central nodes and to each other
- The "middle region," which represents isolated groups which interact amongst themselves but not with the rest of the network, forming isolated stars. These groups grow one user at a time. Over time they merge with the giant component.

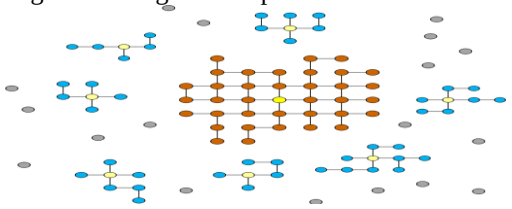

Fig.4. (a): The red section represents the giant component. The blue is the middle region comprising of isolated networks while the gray are singletons.

The node analysis of these networks showed that more than half of a social network is outside the giant component where the greatest centrality lies. They used the "control" definition of centrality to determine this. The research also highlighted a prevalence of "stars" in the middle region which are mini social networks, typically driven by one dynamic member who serves as the point of centrality with others serving as satellite nodes – connected to the dynamic member but not to each other. In Kumar, Novak and Tomkins' analysis [8] the middle region represented one-third of users on Flickr and about ten percent of users on Yahoo! 360. Also keep in mind that the most growth happens in the middle region where dynamic members influence others to join their network. These sub-networks can gradually join the giant component over time. Once they do, the importance of the dynamic member diminishes. Even if that dynamic member were to leave the network, the others would stay in the network.

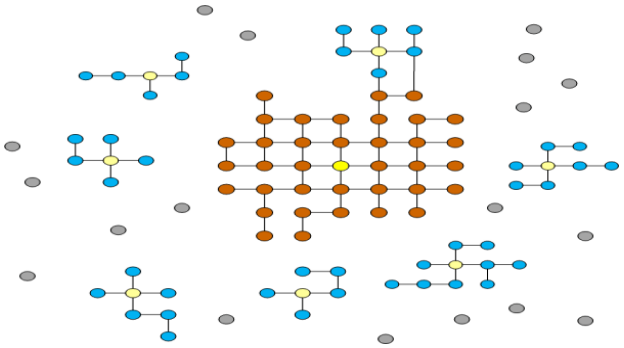

Fig.4. (b): A connection is made between one of the isolated networks from the middle region connects to the giant component.

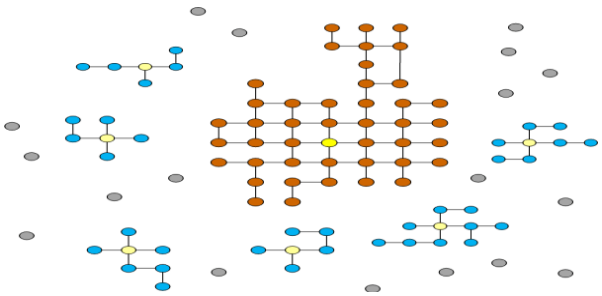

Fig.4. (c): The formerly isolated network becomes part of the giant component.

What are the implications of this? When designing your social network, be aware that most of the network will be outside the giant component. In a sense, social networks themselves are thousands of sub-networks. The more mechanisms that you provide for those sub-networks to flourish, greater the overall network growth. Social networks are fundamentally virtual ghettos. Networks like MySpace and Facebook that encourage ghettos grow the most. Ning, which lets you create your own network and join others too, cleverly understands this concept and leverages it.

D. Live Journal, DBLP & Adoption Behavior

Most online social networks grow based on the initiative of early adopters who transfer their offline networks online and serve as "stars." But it is also important to look at the evolution of social networks based on intentional activity within a network. Backstrom, Huttenlocher and Kleinberg (2006) [1] analyze group formation in large social networks. They used LiveJournal data from its ten million users and DBLP, a database of co-authorship in conference publications to study how the communities grew based on the underlying social networks. They showed that a person was more likely to join a social network if friends of the person were already closely linked together on it. Having several friends closely connected in an online social network builds trust. For those of us who are active members of social networks, this makes obvious sense.

The article conclusively showed that the most growth happened in the giant component (without using the term explicitly) where the nodes were most central. In highlighting the importance of the giant component, Backstrom, Huttenlocher and Kleinberg [1] validated the Kumar et al. (2006) [8] theory. Their article raises a critical question: Once a node becomes aware of its neighbors' behavior, under what conditions and based on what network relationships will the node adopt that behavior itself?

Another group of researchers who studied the DBLP database were Cai et al.(2005) [3]. They pointed out that each node belongs to several different social networks, with the other networks affecting the group formation patterns, evolution and information sharing on the social network. As a result, they felt that a network can't be analyzed independently but needs to be studied in the context of other networks. It may also influence whether a node leaves a network based on the activity of nodes on its other networks. This raises an important question for practitioners: Do you know how much of the activity on your social network is influenced by activity on other social networks?

This is of particular interest when examined in the context of the new Google lab efforts around Social Stream, which hopes to be a meta-social-network aggregating different networks together. Developed in partnership with Carnegie Mellon University, Social Stream is currently in private beta. The question that social network designers worry about is, once you can understand network activity on different networks via a single, consolidated interface, how will that affect your own network preferences?

It is clear that online social networks are always evolving because of both outside influences and activity within them. Butler (2001) [2] emphasized this when they showed that network size has a complex influence on the network such that more member gains results in more member losses too. They argued that it is necessary to balance the positives and negatives of size and communication activity. A final question to consider is which type of membership activity and where (giant component, middle layer or among singletons) most affects an online network?

III. NEED OF THE STUDY DATA MINING IN SOCIAL NETWORK

Social Network Analysis and Mining (SNAM) is intended to be a multidisciplinary to serve both academia and industry as a main venue for a wide range of researchers. We therefore need to solicit experimental and theoretical work on social network analysis and mining using different techniques from computer science. The main areas covered by SNAM include:

(1) Data mining advances on the discovery and analysis of communities, personalization for solitary activities (like search) and social activities (like discovery of potential friends), the analysis of user behavior in open forums (like conventional sites, blogs and forums) and in commercial platforms (like e-auctions), and the associated security and privacy-preservation challenges;

(2) Social network modeling, construction of scalable, customizable social network infrastructure, identification and discovery of dynamics, growth, and evolution patterns using machine learning approaches or multi-agent based simulation.

Research should elaborate on data mining or related methods, issues associated to data preparation and pattern interpretation, both for conventional data (usage logs, query logs, document collections) and for multimedia data (pictures and their annotations, multi-channel usage data).

IV. OBJECTIVE OF THE STUDY

The objective of the proposed study would be to make prospective clients and sources aware of the latest trends and patterns in the inclusion of the data mining tools and techniques in the dynamic social networks and fuzzy systems which effect the common man so that a better system can be developed with improved and modern techniques of data mining. The inclusion of improved and proven algorithms adds to the security level and thoughts of various researchers will be part of the research.

The overall objective of the proposed research is to come up with an aim to have the latest research and simultaneously discuss contributions of renowned researchers whose work created a revolution in the area. The contributions by eminent researchers in dynamic social network Mining and data mining in fuzzy systems will be part of research.

V. MODEL FOR THE NEW SOCIAL NETWORK

The model that we want to propose through this article is the evolution of Social Networks as a Service (SNaaS). This has been conceptualized in a different manner than the normal social networking servicer that we have. What we need are social networks focusing on a particular functionality and exposing that as a service.

The model that we have proposed has been shown in Fig. 5 We have community management as a specialized service for communities, event management, contact management .for profiles as some of the niche services that a single social networking application (shown in the horizontal plane) can provide.

The basic thought behind this model is that we will be having a host of services on the Internet that can be made available to a user. Such services have been shown under the services vertical in the model. In the service provider section, we have the social networks, where each social network will be specializing in one of the services (Fig. 5). While a social network may choose to provide multiple services, its niche area will be one of the services. As a result, the social network application itself becomes a service provider. Hence, a user availing this service can reduce multiple user profile creation and risks associated with the same.

In the proposed model, the social network would be concentrating upon one of the niche services while keeping the other common features as an additive or supporting cast to the main service.

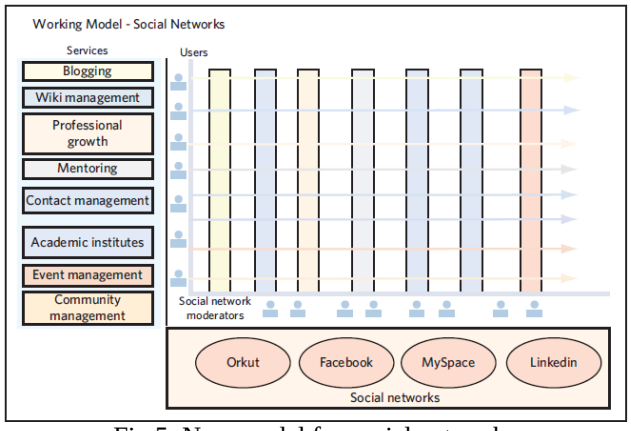

Fig.5. New model for social networks

VI. Research Methodology

The goal of the research process is to produce new knowledge or deepen understanding of a topic or issue. This process takes three main forms (although, as previously discussed, the boundaries between them may be obscure):

- Exploratory research, which helps to identify and define a problem or question.
- Constructive research, which tests theories and proposes solutions to a problem or question.
- Empirical research, which tests the feasibility of a solution using empirical evidence.

There are two ways to conduct research:

1. Secondary research

Using secondary sources, i.e., a synthesis of, interpretation of, or discussions about primary sources. There are two major research designs: qualitative research and quantitative research. Researchers choose one of these two tracks according to the nature of the research problem they want to observe and the research questions they aim to answer:

2. Qualitative research

Understanding of human behavior and the reasons that govern such behavior. Asking a broad question and collecting word-type data that is analyzed searching for themes. This type of research looks to describe a population without attempting to quantifiably measure variables or look to potential relationships between variables. It is viewed as more restrictive in testing hypotheses.

VII. Conclusion

Researchers studying group formation have incorporated the normative social network theories discussed by Granovetter and Freeman. They recognize that these are socio-technical systems that must account for human agency, meaning that the ability of human beings to make unique choices heavily influences a network's evolution. As a result, one can apply social networking theory to a web product, but one must remember that because these are human systems it is difficult to gauge the potential success of a given network. Thus, it is up to us to participate in this movement and continue to contribute towards the betterment of the technology and concept for more contribution to the society as a whole.

Acknowledgment

We are grateful to Mr. Anand S. Hanchinal and Ms. Salma M.B, who helped to complete the paper and participated in valuable discussions while writing the paper.

References

[1] Backstrom, L., Huttenlocher, D., Kleinberg, J., and Lan, X. (2006.) Group formation in large social networks: membership, growth, and evolution. In Proceedings of the *12th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining.* ACM Press: Philadelphia, PA, USA.

[2] Butler, B. (2001.) Membership size, communication activity, and sustainability: a resource-based model of online social structures. *Information Systems Research*, 12 (4), p. 26.

[3] Cai, D., Shao, Z., He, X., Yan, X., and Han, J. (2005) Mining hidden community in heterogeneous social networks. In Proceedings of the *3rd International Workshop on Link Discovery.* ACM Press: Chicago, Illinois.

[4] Freeman, L. C. (1979.) Centrality in social networks conceptual clarification. *Social Networks*, 1 pp. 215-239.

[5] Garton, L., C. Haythornthwaite and B. Wellman. (1997.) Studying online social networks. *Journal of Computer Mediated Communication*, 3 (1).

[6] Granovetter, M. S. (1973) The strength of weak ties. American *Journal of Psychology*, 78 (6), pp.1360-1380.

[7] Johnson, B., Lorenz, E. and Lundvall, B. (2002.) Why all this fuss about codified and tacit knowledge? *Industrial and Corporate Change*, 11 (2), pp. 245-262.

[8] Kumar, R., J. Novak and A. Tomkins. (2006.) Structure and evolution of online social networks. Proceedings of the *12th ACM SIGKDD International Conference on Knowledge*

Discovery and Data Mining. ACM Press: Philadelphia, PA, USA, pp. 611-617.

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

[10] A. L. Barabasi, H. Jeong, Z. Neda, E. Ravasz, A. Schubert, and T. Vicsek. *Evolution of the social network of scientific collaborations. Physica A*, 311(3–4):590–614, 2002.

[11] M. E. J. Newman. The structure of scientific collaboration networks. *PNAS*, 98:404–409, 2001.

[12] K. M. Borgwardt, H.-P. Kriegel, and P.Wackersreuther. Pattern Mining in Frequent Dynamic Subgraphs. In *Proc. of the 6th IEEE Intl. Conf. on Data Mining*, pg. 818–822, 2006.

[13] P. Desikan and J. Srivastava. Mining Temporally Evolving Graphs. In *Proc. of WebKDD 2004*, pg. 22–25, 2004.

[14] J. Leskovec, J. Kleinberg, and C. Faloutsos. Graphs over time: densification laws, shrinking diameters and possible explanations. In *Proc. of the 11th ACM SIGKDD Intl. Conf. on Knowl. Disc. in Data Mining*, pg. 177–187, 2005.

[15] D. Liben-Nowell and J. Kleinberg. The link prediction problem for social networks. In *Proc. of the 12th Intl. Conf. on Inf. and Knowl. Management*, pg. 556–559, 2003.

AUTHOR'S PROFILE

Karuna C. Gull

is from Hubli, India. She has born on 7th June 1974. She has received the B.E. degree in Electronics and Communication from Karnataka University, India in the year 1996 and the M.Tech degree in Computer science and Engineering from the Visvesvaraya Technological University, India in the year 2008. She has been working in the area of data mining and social networking since 2009. She has published 1International paper on Data Mining (2009), 2 International papers on Image Processing (2011). She has also published 3 National and 4 International papers in Conference Proceedings. She has also attended many of the workshops and conferences held in different places on High Impact Teaching Skills, Embedded System Using Microcontroller, Information Storage and Management (ISM), Data Mining, and many more. She worked as a Lecturer and Senior Lecturer for about 10 years. She is currently working as an Assistant Professor in K.L.E.IT, Hubli, India since 2011.

Padmashree Desai

is working currently as Associative Professor in the department of Computer science and Engineering, B.V.B. College of Engineering and Technology, Hubli, Karnataka. She has published many papers in international journals and International conferences in the area of Image processing, Datamining. Also published papers on outcome based education and curriculum design. She has got teaching experience of 21 years.

Akshata B. Angadi

received the BE degree in Computer Science from Visvesvaraya Technological University, India in 2011. She is currently working as a Lecturer in K.L.E.IT, Hubli since 2011.