

# **A Review of Existing Measures, Methods and Framework for Tracking Online Community in Social Network**

Sanjiv Sharma  
Banasthali Vidyapith  
Bansthal Rajasthan (INDIA)

G.N. Purohit  
Banasthali Vidyapith  
Bansthal Rajasthan (INDIA)

## **ABSTRACT**

Social relationships and networking are key components of human life. Social network analysis provides both a visual and a mathematical analysis of human relationships. Recently, online social networks have gained significant popularity. This popularity provides an opportunity to study the characteristics of online social network graphs at large scale. An online social network graph consists of people as nodes who interact in some way such as members of online communities sharing information using relationships among them. In this paper a state of the art survey of the works done on community tracking in social network. The main goal is to provide a road map for researchers working on different measures for tracking communities in Social Network.

## **Keywords**

Social network, community, graph, measures, analysis.

## **1. INTRODUCTION**

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 [1] focuses on the analysis of patterns of relationships among people, organizations, states and such social entities. Online social network provide social networking services over internet. The members of online social network can interact to each other in form of online communities. Communities are considered as groups of densely connected members that are only loosely connected to the rest of the network. Online community [2] is a subgroup where all members of online social network can share information about common interest. The evolution of communities over time is typically analysed by observing changes in the interaction behaviour of their members. The structure of community over time shows stable communities with a considerable amount of members who participate over a long time and a small amount of fluctuating members. Many researchers have been proposed various methods and algorithm for tracking community structure over time in online social network. This paper illustrates review of all researches for tracking online community in social network.

## **2. RELATED WORK**

Social network analysis (SNA) [1] is an interdisciplinary methodology developed mainly by sociologists and researchers in social psychology in the 1960s and 1970s, further developed in collaboration with mathematics, statistics, and computing that led to a rapid development of formal analysing techniques which made it an attractive tool for other disciplines like economics, marketing or industrial engineering . SNA is based on an assumption of the importance of relationships among interacting units or nodes.

Existing methods for tracking community is based on following review of research relevant to following problems that need to be addressed in subgroup analysis of social networks.

### **2.1 Subgroup Identification**

The problem of identifying subgroups has been a challenging issue in sociology. Wasserman and Faust [3] defined a cohesive subgroup as a set of actors (nodes) that are relatively dense and directly connected through reciprocated (bi-directional) relationships (links). They analyzed subgroups over time with in small size of network. Kumar [4] et al addressed the problem of finding emerging subgroups on the Web using heuristic style & created larger community from union of smaller community. The method proposed by him related to degree centrality as a way of selecting potential subgroups and suitable for scalable large network, but method was not fully automated and involved some filtering and interpretation by humans.

Research on finding Web communities has tended to utilize content analysis of text and tags associated with Web pages and allow the subgroup identification task to be linked to powerful search engine algorithms. Flake [5] presented a heuristic community identification algorithm for Searching and link analysis were then used to identify a community of pages relating to the topic seeds. Vein, Chau [6] et al developed a method identifying the communities associated with business Web sites by tracking back through the incoming links and carrying out data mining on the resulting network. Gruz and Haythornthwaite [7] analysed the conversation links between people and they suggest a structure of social network that represent actual conversation rather than the less informative” chain networks” and how messages connect to each other. Text in the threaded conversations can also be analyzed using natural language processing (NLP) algorithms to identify nouns and phrases, which may be used to automatically find and characterize communities. Gibson, Kumar, and Tomkins [8] presented a method for detecting densely connected groups of servers on the Web using graph theoretic research. Their analysis is focused on only static social networks. Tantipathanandh, Berger-Wolf and Kempe [9] have addressed the issue of finding subgroups in dynamic social networks using an optimization approach. Recently, Microsoft has described a supervised learning approach to identifying communities in a patent application [10]. This type of approach is based on classification of objects and their relationships is conducted through initial training to determine if an object is a member of the community using Support Vector Machines [11] and feature-based object classification algorithms such as PopRank [12].

Hence, there has been significant work on methods of automated subgroup identification, but many of them done on

small or static networks. In addition, much of the focus has been on finding communities based on links between text and web content, rather than on links between people as reflected in their online interactions. Thus, the development of a scalable and valid method for tracking subgroups or community of people based on large scale online interactions within social networks is still an open research problem.

## **2.2 Centrality**

Social network centrality is a widely used measure for identifying possible members of subgroups in networks derived from online interactions. A centrality measure can be used to identify the most important people at the center of a network or those that are well connected. Various type of centrality measures such as degree [13,14], closeness[15,16,17], betweenness[18,19,20], information [21,22], eigenvector [23,24], and dependence centrality[25,26] have been used for characterizing the social behaviour and connectedness of nodes within networks. The centrality measures find more active people from one or more subgroups. The active people have higher centrality score as compare to other people of the corresponding network. Thus, high scores of network centrality should be predictive of subgroup membership, although the relationship is by no means deterministic. Researchers have compared and contrasted centrality measures in various social networks.

Betweenness centrality has tended to more suitable to find and measure subgroup and community membership whereas degree and closeness centrality have often been used for characterizing influential members [27]. In extremely large social networks, computational efficiency may become an issue in selecting which centrality measure to use of the three measures considered here, degree centrality is the easiest to calculate, while betweenness centrality has the highest calculation complexity.

## **2.3 Clustering and Partitioning of Subgroups**

Mostly research for tracking community or finding groups within networks is based on classification methods. Methods of classification include discriminant analysis, nearest-neighbour assignment, and a number of Bayesian approaches [28]. Finding interconnected subgroups or community within social networks in the absence of prior knowledge about grouping is a problem that has attracted considerable interest. Clique analysis and related methods look directly at the links that occur in a network and identify specific patterns of connectivity. Clustering and partitioning methods are less direct (but more computationally efficient) in that they base their groupings (clusters) on proximity measures (similarities or distances) derived from the connection patterns between network nodes. Both clique analysis and clustering approaches will be considered in the following subsections.

### **2.3.1 Clique and k-plex analysis**

Cliques and k-plexes have been used to characterize groupings in social networks [29,30,31]. Cliques are fully connected subgroups [3] where each member has a direct connection to every other member in the subgroup, thus forming a completely connected graph within the subgroup. Pure cliques tend to be rare in social networks because the criterion of full connectedness is too strict to apply to most social networks. In n-cliques, the requirement that subgroup members are fully connected is relaxed. The criteria for subgroup formation can be further relaxed by allowing

subgroups to form in which members are not completely connected and where each node in the subgroup has direct ties to at least n-k members. The resulting structure is referred to as a k-plex. K-plex analysis has also been used for finding subgroup members in a network. However, similar to cliques, finding k-plexes in large networks is a computationally expensive and exhaustive process because it scales exponentially with the number of nodes in the network. A further problem is that the value of k needs to be selected and the most appropriate value of k for subgroup analysis in a particular social network may not be obvious.

### **2.3.2 Cluster analysis**

Clustering and methods initially developed for problems such as numerical taxonomy may also be applied to analysis of subgroups and communities. While clustering is a conceptually promising approach, in practice there are a huge number of clustering and other techniques that may be potentially relevant to the detection of subgroups within social hypertext networks. Hierarchical clustering can automate the process of finding subgroups. It groups nodes into a cluster if the nodes are similar and then successively merges clusters until all nodes have been merged into a single remaining cluster. Techniques based on hierarchical clustering have been used to quantify the structure of community in documents [32], web pages, blogs, [33] and discussion groups [34]. Hierarchical clustering using such algorithms as in , results in a hierarchy (tree) being formed where the leaves of the tree are the nodes that are clustered. The resulting trees can be visualized as dendrograms. There are different methods to perform hierarchical clustering depending on how distances between clusters are defined, as explained by Johnson [35]. In contrast to hierarchical cluster analysis, the groups formed in partitioning methods are not nested. The k-means algorithm [36] is a popular method for partitioning that is available in widely used statistical packages such as SPSS. K-means analysis has been used to detect clusters in blogs .

Partitioning methods are relatively efficient, but they require that the number of subgroups in the partition be defined prior to the analysis. On the other hand, hierarchical cluster analysis does not yield a partition and the hierarchy (dendrogram) that is output needs to be cut in order to identify a particular set of subgroups. In practice, for both partitioning analysis and hierarchical cluster analysis, the method needs to be supplemented with an additional selection criterion. For partition analysis, the method is run using a number of different values of k (i.e., number of groups in the partition) and the selection criterion is used to define which of the possible partitions should be chosen as the best subgrouping. For hierarchical clustering, the selection criterion is used to decide at which point the dendrogram should be cut in order to obtain a non-nested set of subgroups.

### **2.3.3 Partitioning criteria**

Orford [37] described a range of criteria for determining where to partition a dendrogram .Orford made the important point that the best criterion to use will generally vary with the problem context. In contrast to Orford's eclectic approach, recent research has tended to assess specific measures for obtaining an optimal partition. The modularity (designated as Q) discussed by Newman and Girvan[19] , has been proposed as a definitive measure of the quality of clustering. Newman [24] claimed that maximizing modularity results in a set of clusters that best represents optimum subgroup structure.

Modularity has been used for finding community structure and subgroups in networks. The computational performance of different algorithms based on modularity was evaluated in and algorithms have been revised to improve over Newman's original method. Most recently, Noack [38] showed that the set of clusters found from optimum modularity, corresponded to the clusters formed by visualizing the social network using an energy layout model of pairwise attraction and repulsion between the nodes in the network. This is an important result since many researchers often use energy layout algorithms for visualizing social network structure in social network analytic software such as Pajek and NetDraw, in addition to finding clusterings with optimal modularity. Other approaches for partitioning based on optimality include vector partitioning and normalized cut metrics.

Despite much work being done to create more efficient algorithms for modularity, there has been relatively little research on evaluating its effectiveness in finding meaningful partitions and cohesive subgroups. As noted by Radicchi [39], it is not clear whether the "optimal" partitions that are discovered using the modularity criterion are representative of real collaborations in the corresponding online communities. Van Duijn and Vermunt [40] noted that it is difficult to determine which measure is the most appropriate to use across a range of applications. Thus, Orford's [37] original insight still seems relevant, that is, the best criterion for splitting a dendrogram may depend on factors such as the type of data being collected and compared.

## **2.4 Similarity**

Most social networks are dynamic and connections between people change naturally over time for a number of reasons including recruitment to the network, attrition, and changing relationships between members of the network. Thus even if it were possible to have a definitive measure of optimality at a single point in time, the optimal subgrouping at one point in time would be unlikely to remain optimal at later times. In general, groupings that are transitory or ephemeral will be of less interest than groupings that remain cohesive over time. How can cohesion over time be measured?

### **2.4.1 Existing Models of Similarity**

Cohesive subgroups should have a core group of people that remain the same over different time periods. The situation is complicated by the fact that subgroups may split or merge, so that cohesiveness is not necessarily a property of a single subgroup, but may sometimes relate to a family of one or more related subgroups. However, in general, cohesive families of subgroups at one time period should be similar to corresponding subgroups at a different time period. Similarity is a topic that has received attention in a wide variety of scientific fields and a number of approaches are available for the measurement of similarity.

Mathematically, similarity may be viewed as a geometric property involving the scaling or transformation necessary to make objects equivalent to each other. Similarity can be defined as the inverse of distance, with a well-known distance measure being Euclidean distance [41], which itself is a special case of a family of distance measures known as Minkowski metrics [42]. However, distance measures typically require a vector (spatial) model of the entities being compared, which is often not appropriate for comparing aggregations of nodes in a network. In developing methods to assess the similarity between different species, numerical taxonomists have developed and utilized a number of

similarity measures. Many of these measures involve some sort of correlation, a construct that is conceptually related to similarity. One correlation measure is the cosine distance or dot product that measures the angle between two objects represented as vectors of numerical features. However, since features cannot always be expressed on a well-defined numerical scale, researchers (e.g., psychologists) have developed feature models of similarity that assess similarity based on a comparison of matching and mismatching features, using a set-theoretic approach.

Tversky's [43] feature contrast model expressed the degree of similarity of two stimuli to a linear combination of their common and distinctive features. A simplified version of the content similarity model is the Jaccard coefficient (first proposed in 1901 by Paul Jaccard [44]), which is defined as the size of the intersection divided by the size of the union of the objects being compared.

Johnson [35] proposed the ultrametric distance as a way of measuring distance within a hierarchy. For comparing two different clustering hierarchies, one heuristic method for estimating similarity consists of converting each hierarchy to a matrix of ones and zeros where the ones represent the parent-child links in each hierarchy.

The similarity between two hierarchies is then estimated as the correlation between the two corresponding matrices of ones and zeroes. A more formal approach is to use quadratic assignment to assess the similarity between two partitions. Quadratic assignment is a combinatorial approach, where simulation is used to create a sampling distribution of possible shuffles of a partition in terms of a correlation or regression statistic between the original partition and each shuffled version. The similarity observed between two partitions is then compared with that sampling distribution to see how extreme or notable the observed statistic actually is. Other related work by Falkowski [45] focused on finding community instances using similarity.

## **2.5 Behavioural Measures of Community**

In contrast to the fully automated methods discussed thus far, human judgment can also be used in inferring subgroupings. In some cases human judgment, as reflected in the structuring of online communities and websites, can be inferred without additional data collection effort. For instance, researchers have used corroborating events, groups and categories inherent in the structure of online communities such as LiveJournal, DBLP and IMDB [34, 46] to validate inferred community structure. User behaviour has also been studied in virtual communities. For instance, content analysis of interactions such as newsgroups, e-mail, or blogs may be used to find evidence of group or community membership. While these methods of validating subgroupings or inferred community are relatively indirect, they have the benefit of being easy to apply. Alternatively, human judgments may be collected in validating the obtained groupings, but with greater effort.

## **3. EXISTING METHODS AND FRAMEWORK FOR TRACKING COMMUNITIES**

A variety of approaches have been suggested to address this problem and the corresponding research literature on centrality, clustering, and optimization methods for finding

subgroupings are reviewed. This review will include a critical analysis of the limitations of past approaches. First, the method proposed by Chin and Chignell called SCAN [31,49] (Social Cohesion Analysis of Networks), where a combination of heuristic methods is used to identify subgroups in a manner that can potentially scale up to very large social networks. then, The DISSECT [51](Data-Intensive Socially Similar Evolving Community Tracker) method proposed by Chin and Chignell where multiple known subgroups within a social network are tracked in terms of similarity-based cohesiveness over time. The DISSECT method relies on cluster analysis of snapshots of network activity at different points in time followed by similarity analysis of subgroup evolution over successive time periods. Online social networks evolve over time, and research has looked into the temporal aspects of social networks changing over time. It has been found that groups discovered in social networks differ in their cohesiveness or bonding, which can be in time or space. Since online social networks have time inherent in their structure, cohesive subgroups can be defined as those that are similar over time based on Social Identity Theory [50] where group members feel closer if they are similar to each other. Cohesive subgroups can also be considered as optimum subgroups by calculating the optimum number of clusters, modularity, or optimizing graphs. Similarity measures are used in research to assess the cohesiveness of subgroups. Since they explicitly consider changes in subgroupings over time, similar measures take into account network membership dynamics in the social network. Different types of similar measures can be constructed depending on the particular network dynamics observed. This section addresses the problem of tracking community in social networks inferred from online interactions, by expanding on the problem of finding subgroups initially explored through the SCAN method [49] and addressing the limitations of the SCAN method.

### **3.1 Social Cohesion Analysis of Networks (SCAN)**

Social network analysis is a new research field in data mining. Finding subgroups within social networks is important for understanding and possibly influencing the formation and evolution of online communities. The SCAN (Social Cohesion Analysis of Networks) [49] methodology involves three steps: selecting the possible members (Select), collecting those members into possible subgroups (Collect) and choosing the cohesive subgroups over time (Choose). Social network analysis, clustering and partitioning, and similarity measurement are then used to implement each of the steps.

The Social Cohesion Analysis of Networks (SCAN) [49] method was developed for automatically identifying subgroups of people in social networks that are cohesive over time. The SCAN method is to be applied based on the premise that a social graph can be obtained from the online community interactions where the links are untyped (i.e., there are no associated semantics). In the social graph, each link represents an interaction between two individuals where one individual has responded to the other's post in the online community. The SCAN method has been designed to identify cohesive subgroups on the basis of social networks inferred from online interactions around common topics of interest. The SCAN method consists of the following three steps:

1. Select: Selecting potential members of cohesive subgroups from the social network.
2. Collect: Grouping these potential members into subgroups.

3. Choose: Choosing cohesive subgroups that have a similar membership over time.

### **3.2 DISSECT (Data-Intensive Socially Similar Evolving Community Tracker)**

The DISSECT method addresses the following shortcomings of the SCAN method:

1. The SCAN method only focused on betweenness centrality; other centrality measures may be useful.
2. The SCAN method only looked into two types of similarity measures (constant membership and members entering the network); there is a need to examine for other types.
3. The time periods used in the SCAN method were defined ad hoc as a matter of convenience, without any systematic evaluation.
4. The SCAN method fails if semantic properties determine subgroup membership.

### **3.3 Framework for DISSECT Method**

The Framework for DISSECT method involve following steps for tracking online community in a social network:

1. Find the initial time periods for analysis.
2. Label subgroups of people from the network dataset using content analysis and semantic properties. If possible, individuals are also labelled so as to facilitate later similarity analysis between subgroups at different time periods.
3. Select the possible members of known subgroups to be tracked using the Select step from the SCAN method.
4. Carry out hierarchical cluster analysis of interaction data taken at snapshots in time and involving known subgroups of people (using the Collect step from the SCAN method).
5. Repeat steps 3 and 4 for different values of betweenness centrality (note that the DISSECT approach is agnostic in terms of which of the many available measures of centrality should be used).
6. Calculate similarity of subgroups for the designated time periods from step 1 using the clustering results of the previous step. In this case, the similarity measure can be augmented to take into account semantic labels assigned to different people.
7. Repeat steps 2 through 6 for different time period intervals and combinations.
8. Construct a chronological view of each subgroup showing how it changes over time (as the assigned semantic labels change).

The ultimate goal in a DISSECT analysis is to trace the evolution of subgroups into communities. The DISSECT method does not stand as a theory of how communities form. However, logically it would seem that if communities do emerge out of online interactions then they are likely to evolve, initially, from smaller subgroups.

## **4. CONCLUSION**

This paper reviews the measures of social networks, formal methods and framework for tracking online community in social network. The researchers adopt different approaches for handling problem of tracking community over time. Subgroup identification, centrality, cluster analysis, clique analysis and similarity are basic measures & methods for finding structure

of community. SCAN & DISSECT are two methods or frameworks for finding cohesive group & discussed a framework for tracking community evolution in an online Community called DISSECT or Data-Intensive Socially Similar Evolving Community Tracker. This framework is an expanded and enhanced version of the SCAN method, for finding cohesive subgroups in online interactions. The framework is designed to be a step-by-step process to track the evolution of community members. This survey paper proposes future direction for improving existing methods & framework of tracking community over time.

## 5. ACKNOWLEDGMENTS

We would like to thank Department of computer science & engineering of Banasthali University for supporting our research work.

## 6. REFERENCES

- [1] Scott, Social Network Analysis. A Handbook. Sage (2000).
- [2] Shuie, Yih-Chearnng. Exploring and Mitigating Social Loafing in Online Communities. *Computers and Behavior*. v.26.4, July 2010. p. 768–777
- [3] WASSERMAN, S. & FAUST, K. Social Network Analysis: Methods and Applications, Cambridge University Press (1995).
- [4] Kumar, R., Raghavan, P., Rajagopalan, S., and Tomkins, A. Trawling the web for emerging cyber-communities. *Computer Networks* (1999).
- [5] Flake, G. W., Lawrence, S., Giles, C. L., and Coetzee, F. M. Selforganization and identification of web communities. *IEEE Computer* 35, 3 (2002), 66–71.
- [6] Chau, M., Shiu, B., Chan, I., and Chen, H. Automated identification of web communities for business intelligence analysis. In *Proceedings of the Fourth Workshop on E-Business (WEB)* (New York, NY, USA, 2003).
- [7] Gruzdz, A., and Haythornthwaite, C. Automated discovery and analysis of social networks from threaded discussions. Paper presented at the *International Network of Social Network Analysts* (2008).
- [8] Gibson, D., Kumar, R., and Tomkins, A. Discovering large dense subgraphs in massive graphs. In *VLDB '05: Proceedings of the 31st international conference on Very large data bases* (2005), VLDB Endowment, pp. 721–732.
- [9] Tantipathananandh, C., Berger-Wolf, T. Y., and Kempe, D. A framework for community identification in dynamic social networks. In *KDD '07: Proceedings of the 13th ACM SIGKDD international conference on Knowledge discovery and data mining* (New York, NY, USA, 2007), ACM, pp. 717–726.
- [10] Zhao, Q., Liu, T.-Y., and Ma, W.-Y. Predicting community members based on evolution of heterogeneous networks (patent number us 2007/0239677 a1). Microsoft Corporation, 2007.
- [11] Joachims, T. Making large-scale svm learning practical. *Advances in Kernel Methods - Support Vector Learning*, B. Scholkopf and C. Burgess and A. Smola (ed.) (1999).
- [12] Nie, Z., Zhang, Y., Wen, J.-R., and Ma, W.-Y. Object-level ranking: bringing order to web objects. In *WWW '05: Proceedings of the 14th international conference on World Wide Web* (New York, NY, USA, 2005), ACM, pp. 567–574.
- [13] Fisher, D. Using egocentric networks to understand communication. *IEEE Internet Computing* 9, 5 (2005), 20–28.
- [14] Frivolt, G., and Bielikov, M. An approach for community cutting. In *RAWS 2005 Proc. of the 1st Int. Workshop on Representation and Analysis of Web Space*, V. Svatek, V. Snasel (Eds.) (2005), pp. 49–54.
- [15] Chin, A., and Chignell, M. A social hypertext model for finding community in blogs. In *Proceedings of the 17th International ACM Conference on Hypertext and Hypermedia: Tools for Supporting Social Structures* (Odense, Denmark, 2006), ACM, pp. 11–22.
- [16] Frivolt, G., and Bielikov, M. An approach for community cutting. In *RAWS 2005 Proc. of the 1st Int. Workshop on Representation and Analysis of Web Space*, V. Svatek, V. Snasel (Eds.) (2005), pp. 49–54.
- [17] Ma, H.-W., and Zeng, A.-P. The connectivity structure, giant strong component and centrality of metabolic networks. *Bioinformatics* 19, 11 (2003), 1423–1430.
- [18] Donetti, L., and Munoz, M. A. Detecting network communities: a new systematic and efficient algorithm. *Journal of Statistical Mechanics: Theory and Experiment* 2004, 10 (2004), P10012.
- [19] Girvan, M., and Newman, M. E. Community structure in social and biological networks. *PROC.NATL.ACAD.SCI.USA* 99 (2002), 7821.
- [20] Gloor, P. A., Laubacher, R., Dynes, S. B. C., and Zhao, Y. Visualization of communication patterns in collaborative innovation networks - analysis of some w3c working groups. In *CIKM '03: Proceedings of the twelfth international conference on Information and knowledge management* (New York, NY, USA, 2003), ACM Press, pp. 56–60.
- [21] Costenbader, E., and Valente, T. W. The stability of centrality measures when networks are sampled. *Social Networks* 25 (Oct. 2003), 283–307.
- [22] Crucitti, P., Latora, V., and Porta, S. Centrality measures in spatial networks of urban streets. *Physical Review E* 73 (2006), 036125.
- [23] Estrada, E., and Rodriguez-Velazquez, J. A. Subgraph centrality in complex networks. *Physical Review E* 71 (2005), 056103.
- [24] Newman, M. E. Modularity and community structure in networks. *Proceedings of the National Academy of Sciences* 103, 23 (2006), 8577–8582.
- [25] Memon, N., Harkiolakis, N., and Hicks, D. Detecting high-value individuals in covert networks: 7/7 london bombing case study. *Computer Systems and Applications*, 2008. AICCSA 2008. IEEE/ACS International Conference on (31 2008-April 4 2008), 206–215.
- [26] Memon, N., Larsen, H. L., Hicks, D. L., and Harkiolakis, N. Detecting hidden hierarchy in terrorist networks: Some case studies. *Lecture Notes in Computer Science* 5075 (2008), 477–489.

- [27] Costenbader, E., and Valente, T. W. The stability of centrality measures when networks are sampled. *Social Networks* 25 (Oct. 2003), 283–307.
- [28] Duda, R. O., Hart, P. E., and Stork, D. G. *Unsupervised Learning and Clustering*. Wiley, New York, 2001.
- [29] Alba, R. D. A graph-theoretic definition of a sociometric clique. *Journal of Mathematical Sociology* 3 (2003), 113–126.
- [30] Balasundaram, B., Butenko, S., Hicks, I., and Sachdeva, S. Clique relaxations in social network analysis: The maximum k-plex problem. Tech. rep., Texas A and M Engineering, 2007.
- [31] Chin, A., and Chignell, M. Identifying subcommunities using cohesive subgroups in social hypertext. In *HT '07: Proceedings of the eighteenth conference on Hypertext and hypermedia* (New York, NY, USA, 2007), ACM, pp. 175–178.
- [32] Brooks, C. H., and Montanez, N. Improved annotation of the blogosphere via autotagging and hierarchical clustering. In *WWW '06: Proceedings of the 15th international conference on World Wide Web* (New York, NY, USA, 2006), ACM Press, pp. 625–632.
- [33] Li, X., Liu, B., and Yu, P. S. Mining community structure of named entities from web pages and blogs. In *AAAI Spring Symposium-2006* (2006), AAAI.
- [34] Gómez, V., Kaltenbrunner, A., and López, V. Statistical analysis of the social network and discussion threads in slashdot. In *WWW '08: Proceeding of the 17th international conference on World Wide Web* (New York, NY, USA, 2008), ACM, pp. 645–654.
- [35] Johnson, S. C. Hierarchical clustering schemes. *Psychometrika* 32.
- [36] Hartigan, J. *Clustering Algorithms*. John Wiley and Sons, New York, NY, 1975.
- [37] Orford, J. D. Implementation of criteria for partitioning a dendrogram. *Mathematical Geology* 8, 1 (1976), 75–84.
- [38] Noack, A. Modularity clustering is force-directed layout. Retrieved from <http://www.citebase.org/abstract?id=oai:arXiv.org:0807.4052> [accessed 30 September 2008], 2008.
- [39] Radicchi, F., Castellano, C., Cecconi, F., Loreto, V., and Parisi, D. Defining and identifying communities in networks. *Proceedings of the National Academy of Sciences of the United States of America* 101, 9 (2004), 2658–2663.
- [40] van Duijn, M. A. J., and Vermunt, J. K. What is special about social network analysis? *Methodology* 2 (2005), 2–6.
- [41] Elmore, K. L., and Richman, M. B. Euclidean distance as a similarity metric for principal component analysis. *Monthly Weather Review* 129, 3 (March 2001), 540–549.
- [42] Santini, S., and Jain, R. Similarity measures. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 21, 9 (September 1999), 871–883.
- [43] Tversky, A. Features of similarity. *Psychological Review* 84, 4 (1977), 327–352.
- [44] Jaccard, P. Distribution de la flore alpine dans le bassin des dranses et dans quelques rgions voisines. *Bulletin del la Socit Vaudoise des Sciences Naturelles* 37 (1901), 241–272.
- [45] Falkowski, T., Bartelheimer, J., and Spiliopoulou, M. Community dynamics mining. In *Proceedings of 14th European Conference on Information Systems (ECIS 2006)* (Gteborg, Sweden, 2006).
- [46] Leskovec, J., Lang, K. J., Dasgupta, A., and Mahoney, M. W. Statistical properties of community structure in large social and information networks. In *WWW '08: Proceeding of the 17th international conference on World Wide Web* (New York, NY, USA, 2008), ACM, pp. 695–704.
- [47] Hirsch, B. J. Psychological dimensions of social networks: A multimethod analysis. *American Journal of Community Psychology* 7, 3 (1979), 263–277.
- [48] Sarason, I. G., Levine, H. M., Basham, R. B., and Sarason, B. R. Assessing social support: The social support questionnaire. *Journal of Personality and Social Psychology* 44 (1983), 127–139.
- [49] Chin A, Chignell M Automatic detection of cohesive subgroups within social hypertext:
- [50] A heuristic approach. *New Rev Hypermed Multimed* (2008) 14(1):121–143
- [51] Tajfel, H., and Turner, J. C. The social identity theory of inter-group behavior. In S. Worchel and L. W. Austin (eds.), *Psychology of Intergroup Relations* (1986).
- [52] Chin A, Chignell M ,Wang H(2010) Tracking cohesive subgroup over time in inferred social network. In *New Review of Hypermedia and Multimedia / Hypermedia* , vol. 16, no. 1&2, (2010) pp. 113-139,