RESEARCH REVIEW JOURNALS
Volume-04
Issue-04
April -2019

ISSN: 2455-3085 (Online)
RESEARCH REVIEW International Journal of Multidisciplinary
www.rrjournals.com [UGC Listed Journal]

Evaluation of Similarity Measures for Existing Frameworks for Tracking Online Social Network Community Structure

[1]Deepak Adhikari & [2]Dr. Jitendra Sheetlani

[1]*Research Scholar, Department of Computer Science, Sri Satya Sai University of Technology & Medical Sciences, Sehore, M.P.*
[2]*Research Guide, Department of Computer Science, Sri Satya Sai University of Technology & Medical Sciences, Sehore, M.P.*

ARTICLE DETAILS

Article History
Published Online: 15 April 2019

Keywords
social network, community, similarity, groups

ABSTRACT

Many genuine-world social networks are intimately systematized correspondingly to a community construction. Most social network are energetic and associations between individuals alter actually extra minutes. Analysts have started to consider the issue of following the arrangement of bunches of clients in social network. The circumstance is involved by the truth that subgroups may part or consolidate, so that cohesiveness isn't fundamentally a property of a single subgroup, may often refer to a family of one or more similar subgroups. In any case, cohesive subgroup families at one time should be comparable to comparing subgroups at a distinguished period of time. Similarity may be a point that has gotten consideration in a wide range of logical areas and a number of approaches are accessible for the estimation of likeness. This paper describes an efficient way to determine similarities in dynamic networks in subgroups or clusters and track community that persists over time.

1. Introduction

A social network is essentially a system made up of people or other entities embedded in a social context, with a relationship between people who had interaction, collaboration, or influence between entities. Strong success and rapid development of these online social networks indicate an isolated opportunity to study, understand and leverage their properties. A community, Tantipathananandh et al. [2007], is understood intuitively as a collection of entities wherever every entity is nearer, within the network sense, to the opposite entities between the community as well as outside entities. Therefore, communities area unit teams of entities that presumptively share certain important properties and produce similar roles inside the interacting development that's being depicted.

This chapter describes a similarity measure, Gregson et al. [1975], for tracking community development and structure in multiple snapshots of a dynamic network, Greene et al. [2010], where the life-cycle of every community is described by a number of important events. Based on this model, a simple but effective method was proposed to effectively identify and track these dynamic communities by Radicchi et al. [2004], which involved matching of communities found at follow-up steps in the individual snapshot graphs. Unlike other approaches, the approach is independent of the option of the underlying community finding algorithm applied to each phase graph. According to McDaid et al. [2010], it can also approximate data from disjunct or overlapping node groupings. To assess the similarity method for observing community, a procedure is introduced in this chapter.

2. Related Work

Cohesive subgroups should have a core group of people who remain unchanged over various periods of time. The problem is compounded by the fact that separating or combining subgroups are not always the responsibility of a single subgroup, but can sometimes be linked to one or more similar subgroups themselves. In general, however, during one point, cohesive subgroup families should be identical to the related subgroups at a different period of time. Similarity is a subject that has been examined in a wide variety of scientific fields and there are a variety of methods for measuring similarity. Chin et al [2009], prior two steps (Select and Collect) may be replicated at any time to detect cohesive subgroups of candidates. If subgroups are cohesive, over time they should be less likely to change. This should therefore be applied to subgroups that retain unity over time over those that do not.

The second perspective is that high levels of cohesion are unlikely to happen by chance. According to the social identity theory, group members feel connected if they are the same, Clauset et al. [2004]. It is therefore unlikely that capitalizing on the best betweenness cutoff to create the most coherent subgroups would distort the trends that occur in the results, Wang et al.[2002]. Subdiagrams are chosen and measured. The unified subgrouping selected is then centered on the centrality of the cutoff betweenness, which maximizes the similarity of the subgroups obtained in adjacent times.

Memon et al.[2008] Use a set theoretical approach to describe similarity as the ratio of the set subgroup intersection to the respective set union. In practice, this is corresponding to the proportion of the number of specific pairs in all clusters in both time periods relative to the total number of possible combinations of pairings. In developing a similarity measure, three issues need to be considered. First, the selection of the time periods or windows to be used for comparison, second, the comparison of time windows and third, the network dynamics.

For the selection of time periods, the approach used in this chapter is to select a similarity measure on the basis of static time windows for the entire time analyzed to simplify the research problem, in order to see how it works. The approach used in this chapter could then be extended to varying time windows and the similarity measure would be modified

RRIJM 2015, All Rights Reserved

accordingly after relaxing the view of persistence over time. More research could be conducted to use a multi-window method to compare participants for candidate subgroups over a specific time period with other time periods. To manage multiple windows, this would involve changing the Select criteria. As far as network dynamics are concerned, the present theory only considers networks where membership is either fixed or new members join elsewhere, but existing members do not leave the network (and subgroups) at different times. Consequently, the fulfilment of the SCAN approach considered here does not recognize members entering and leaving the network or both. These cases are left for future work.

Two similarity-based approaches have been developed to find cohesive subgroups based on the assumed network dynamics noted above. In the first method, Cohesion was investigated across all subgroups where there was continuous membership in subgroups over the corresponding time periods. Cohesion for the main subgroup (only) has been examined in the first of the two time periods measured in the second approach. Both similarity measures may be useful, but these are several ways of measuring cohesion where the former takes into account the constant membership of adjacent subgroups and the latter considers new participants in the subgroup.

The continuity between two consecutive time periods T1 and T2 can be determined according to the formula for the first similarity method 5.1:

$$Sim T_1, T_2 = \frac{2 * N\left(T_1 \cap T_2\right)}{N\left(T_1 \cup T_2\right)} \ldots\ldots\ldots\ldots\ldots\ldots (1.1)$$

Where $N\left(T_1 \cap T_2\right)$ is the number of pairs in both T1 and T2 in the same cluster and $N\left(T_1 \cup T_2\right)$ is the total number of pairs in either (or both) T1 or T2 in the same cluster. A factor of 2 is added to this expression's numerator as a multiplier in order to normalize the resulting similarity measure among 0 and 1.

The second approach to similarity tests the cohesion of the largest single subgroup. It analyses all possible pair relationships between subgroup members and decide how many of the pairs exist in the second span of time (i.e. within a subgroup). The similarity can be determined as per the equation using the following formula. 5.2.

$$Sim\ T_1 T_2 = \frac{N\left(S_1 \cap T_2\right)}{N\left(S_1\right)} \ldots\ldots\ldots\ldots\ldots\ldots\ldots (1.2)$$

where S1 is the largest subgroup in T1, N(S1 ∩ T2) is the number of common pairs from the largest subgroup S1 that still exist in T2 and N(S1) is the number of pairs in the largest subgroup S1.

3. Limitation:

A 3rd similarity measure that takes both new members and members leaving the subgroups would need to be defined, as the two currently defined similarity measures only take into consideration constant membership and new membership. This assumption that participants in the same subgroups can exist over different periods of time can then be relaxed and the impact of the time window varies. Subsequent similarity calculations should also take into account the time window variance, which offers ample potential for evaluating more complex similarity evaluation schemes over time that could integrate a variety of techniques, including time series data analysis, temporal algorithms and sliding window algorithms.

4. Proposed Work

The chapter indicates a Social network as a set of t time graphs {G_1, . . . , G_t}, provide overall network snapshots of nodes and edges at successive intervals. Then the problem gets the identification of a set of k communities D = {D_1, . . . , D_k} that are presenting the network across multiple time steps. Step communities are identified at individual time steps, which represent specific observations of dynamic communities at a given point in time. These need not necessarily consist of cliques, unlike the approach mentioned by Palla et al.[2007]. Alternatively, the results can be drawn from any disjoint or overlap grouping that includes roles for some or all of the nodes throughout the network. and refers to the set of communities k_t step or clusters found at the time t as C_t = {C_{t1}, . . . ,C_{tkt}}.

In the context of the model described above, a key question concerned is how best to map step communities at each time t the existing set of dynamic communities D. Further question may arise regarding the feasibility of performing this correspondence process in an efficient manner for graphs containing a large number of nodes and communities.

The first step of grouping C1 is to apply a selected algorithm to the Graph G1 and use this graph to bootstrap the process. For each static community, a definite dynamic community is formed. The next C_2 category will be produced on the Graph G_2

To perform the actual matching between C_t and the fronts {F_1, . . . , F_k}, Jaccard coefficient is employed for binary sets. The Jaccard distance which calculates the gap between sample sets, is analogous to the Jaccard coefficient and is calculated by subtracting the Jaccard coefficient from 1 and, equivalently, by dividing the union size difference between two sets and the union size intersection: The similarity between the pair is calculated based on the static community C_{ta} and a recent community R_i as:

$$sim\left(C_{ta}, R_i\right) = \frac{C_{ta} \cap R_i}{C_{ta} \cup R_i} \ldots\ldots\ldots\ldots\ldots\ldots (1.3)$$

The second approach to similarity tests the unity of the largest single subgroup. It analyses all existing pair interactions between members of the subgroup and calculates how many of the pairs still remain in the second period of time (i.e. within a subgroup). The similarity can be determined as per the equation using the following formula 5.4:

$$sim\left(S_i, R_i\right) = \frac{C_{ta} \cap R_i}{S_i} \ldots\ldots\ldots\ldots\ldots\ldots (1.4)$$

Where S_i is number of pairs in the largest subgroup.

It is necessary to define the proposed similarity measure which considers both new members and members leaving the subgroups, as the two similarity measures currently defined only take constant membership into account and new membership. The Jaccard distance which calculates the gap between sample sets, is analogous to the Jaccard coefficient and is calculated by subtracting the Jaccard coefficient from 1 and, equivalently, by dividing the union size difference between two sets and the union size intersection:

$$sim\left(C_{ta}, R_i\right) = \frac{\left(C_{ta} \cup R_i\right) - \left(C_{ta} \cap R_i\right)}{C_{ta} \cup R_i} \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots (1.5)$$

If the similarity crosses a corresponding threshold T C[0, 1], match the pair and add C_{ta} to the dynamic group D_i timeline . The intersection calculations needed for Eqn are for practical purposes. 5.3 Optimizations based on sorted sets, Baeza-Yates[2004] or bit array operations, Asur et al.[2007] can be effectively implemented using a variety of strategies. In the implementation used in this section, dynamic communities are represented as a node-community graph comparing incoming communities. This move leads to significant improvements in performance compared to a naive implementation focused on pairs of fixed structures. If there is no suitable match for C_{ta} above the T threshold, C_{ta} will create a new dynamic community. A framework of the entire process is provided as per follows:

1. Calculate Eigen value centrality and generate important nodes.
2. Extract cluster or static community $C_1$of social network G_1using spectral clustering.
3. Initialize D by creating a new dynamic community for each static community $C_{1i}CC_1$.
4. For each subsequent step t > 1, extract C_t from G_t.
5. Process every $C_{ta}C$ C_t as follows:
1. Match all dynamic communities Di for which sim(C_{ta}, R_i) >T or sim(S_i, R_i)>T.
2. If there are no matches, create new dynamic community containing C_{ta}.
3. Otherwise, add C_{ta} to each matching dynamic community.
6. Update the current R_i community array to be the new matched static community for each dynamic community. Create a split dynamic community for each case where an existing dynamic community has balanced 2 or more static communities.
7. Repeat from step 2 until all time graphs have been processed.

5. Experiment

The goal here was to determine whether applying a step-based dynamic community finding process could improve ability to detect dynamic communities, when compared with traditional static community finding methods which treat dynamic networks as a single graph without regard to temporal information. Considering a social network data set, Sampson recorded a group of monks ' social interactions while residing as a vision experimenter and collected numerous socio-metric rankings. A political "crisis in the cloister" during his stay resulted in the expulsion of four monks (Nos. 2, 3, 17 and 18) and the voluntary departure of various others, most immediately Nos. 1, 7,14, 15 and 16. (In the end, there were only 5, 6, 9 and 11 left).

Most of this information were obtained retrospectively after the separation occurred. The period in question during which a new cohort entered the monastery near the end of the study but before the beginning of the major conflict. The examples were three-fold "liking" data collected: SAMPLK1 to SAMPLK3-representing shifts in group opinion over time (SAMPLK3 was collected in the same wave as the following data). Senior monk details has not been provided.

There are four coded relationships, with different matrices for positive and negative relationship bonds. Only three of his top choices were ranked by each member on that tie. Relationships include confidence (SAMPES) and disesteem (SAMPDES), like (SAMPLK), positive influence (SAMPIN), disliking (SAMPDLK), negative influence (SAMPNIN), affection (SAMPPR) and fault (SAMPNPR). 3 specify the greatest or first choice in all ratings and 1 specify the last choice. (Some top four subjects offered tied ranks).

Data:

(TOOFAN)	(GARVESH)	2
(TOOFAN)	(DARSH)	3
(TOOFAN)	(NUGAH)	1
(DEEPAK)	(TOOFAN)	3
(DEEPAK)	(SEJUN)	2
(DEEPAK)	(NUGAH)	1
(GARVESH)	(TOOFAN)	2
(GARVESH)	(DEEPAK)	3
(GARVESH)	(DHONU)	1
(DHARMA)	(DARSH)	3
(DHARMA)	(BIRAJ) 1	
(DHARMA)	(UJESH)	2
(DARSH)	(DHARMA)	3
(DARSH)	(HARTAJ)	2
(DARSH)	(BIBHU) 1	
(BIRAJ)	(TOOFAN)	1
(BIRAJ)	(DHARMA)	3
(BIRAJ)	(YAMA) 2	
(SEJUN)	(DEEPAK)	2
(SEJUN)	(YAGGA)	1
(SEJUN)	(DEVAJ)	3
(YAGYA)	(TOOFAN)	3
(YAGYA)	(DEEPAK)	2
(YAGYA)	(YAMA) 1	
(YAMA)	(DARSH)	2
(YAMA)	(YAGYA)	3
(YAMA)	(DEVAJ)	1
(UJESH)	(DHARMA)	3
(UJESH)	(YAGGA)	1
(UJESH)	(NUGAH)	2
(HARTAJ)	(DARSH)	3
(HARTAJ)	(YAGYA)	1
(HARTAJ)	(NUGAH)	2
(UNNAT)	(TOOFAN)	3
(UNNAT)	(DEEPAK)	2
(UNNAT)	(NUGAH)	1
(BIBHU)	(DARSH)	2
(BIBHU)	(SEJUN)	1
(BIBHU)	(TEJ)	3
(NUGAH)	(TOOFAN)	3
(NUGAH)	(HARTAJ)	1
(NUGAH)	(UNNAT)	2
(NUGAH)	(BATSA)	2
(BATSA)	(TOOFAN)	3
(BATSA)	(DEEPAK)	2
(BATSA)	(NUGAH)	1
(DEVAJ)	(TOOFAN)	1
(DEVAJ)	(DEEPAK)	2

RRIJM 2015, All Rights Reserved

(DEVAJ)	(SEJUN)	3
(DHONU)	(GARVESH)	3
(DHONU)	(BIBHU)	2
(DHONU)	(TEJ)	1
(TEJ)	(TOOFAN)	2
(TEJ)	(DEEPAK)	3
(TEJ)	(SEJUN)	1

Above dataset a analyzed by UCINET in form of graph G, in which all actor represented by vertex set V is communicate to each other by edges E is analyzed in time T_1 , time T_2 and find the community structure as per the following steps:

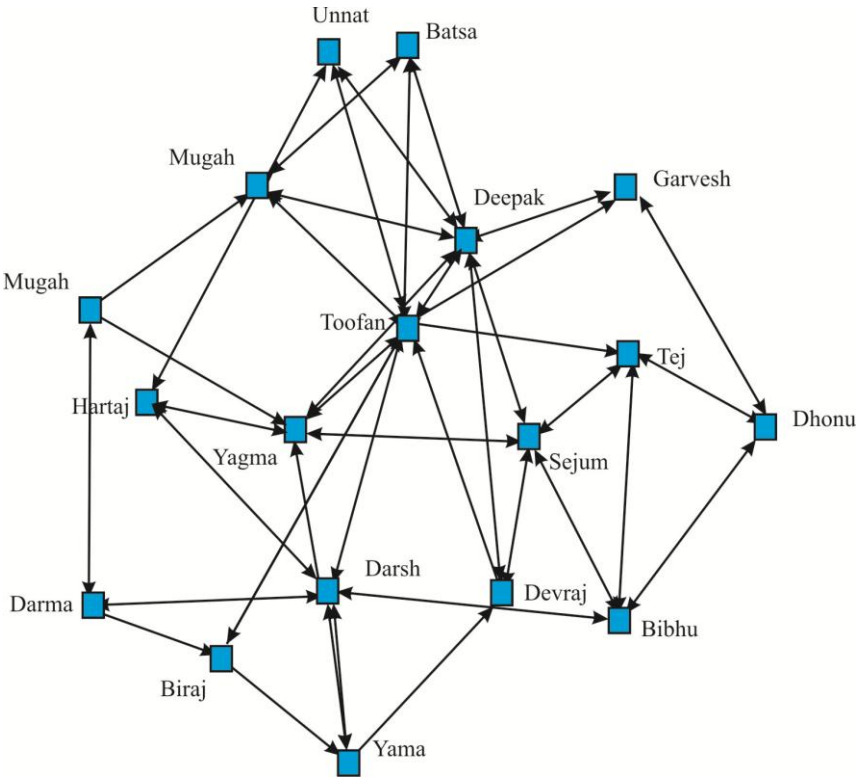

Fig 1.1: Social Network Graph G1 for representing real social interaction dataset collected by Sampson

Step 1: Calculate Eigen value centrality and generate important nodes subsequently Extract cluster or static community C of social network G1 using spectral clustering. In Graph G2 red vertices shows static community and blue vertices are not a part of community.

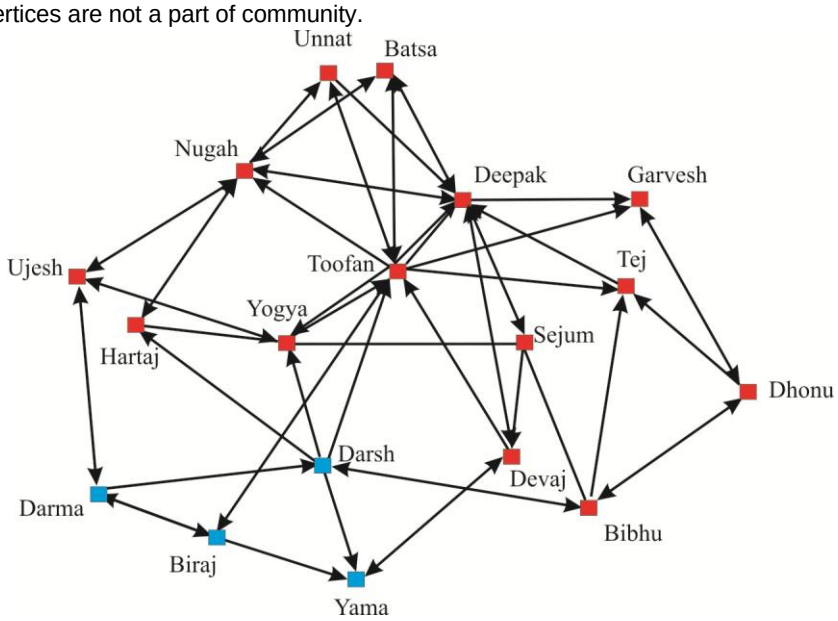

Fig 1.2: Graph G2 for representing static community

Step 2: In time T1 static community C1 is created and generate Graph G3. Initialize D by creating a new dynamic community for each static community C1iאC1

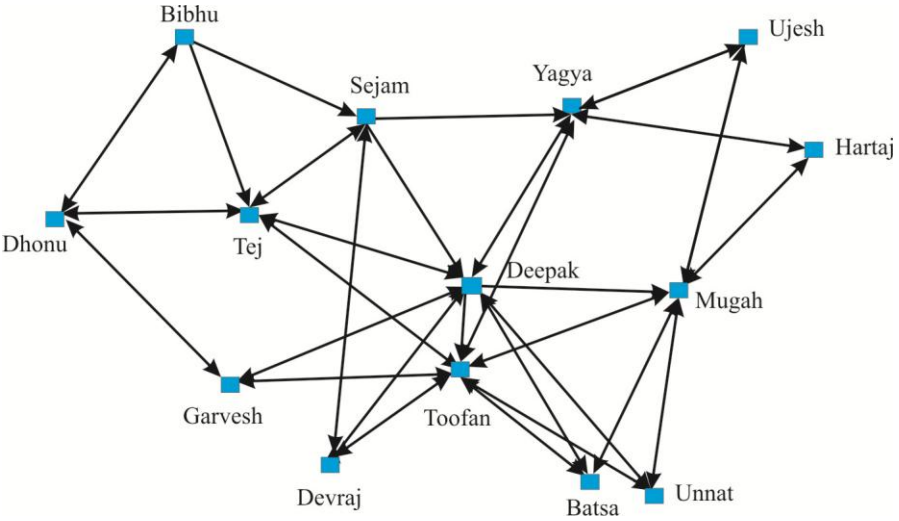

Fig 1.3 : Graph G3 for representing static community C1in time T1

Step 3: In time T2 static community C2 is created and generate Graph G4.

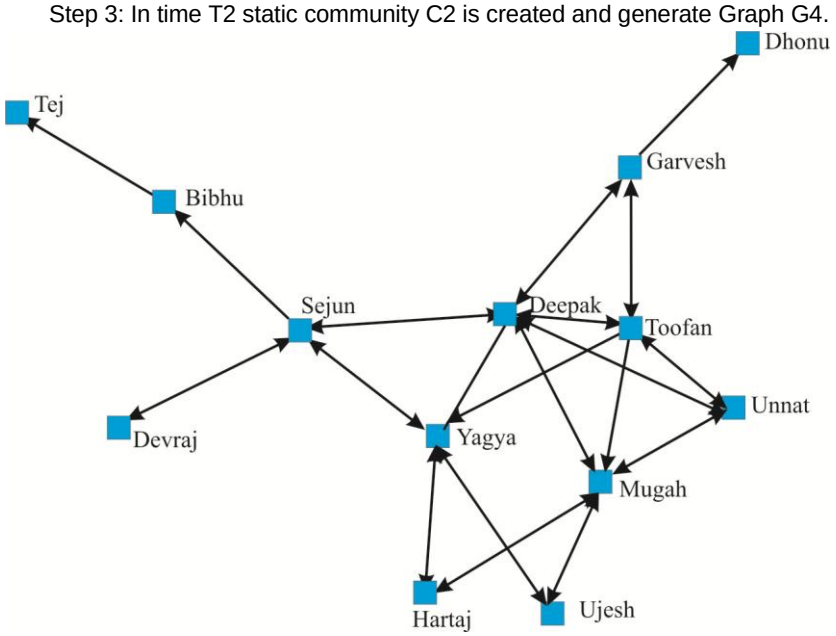

Fig 1.4: Graph G4 for representing static community C2 in time T2

Step 4: Update the current R_i community set to be the latest matched static community for each dynamic community. For each event, where an existing dynamic community is matched to two or more static communities, a split dynamic community is formed. Graph G5 shows the resulting tracked dynamic community.

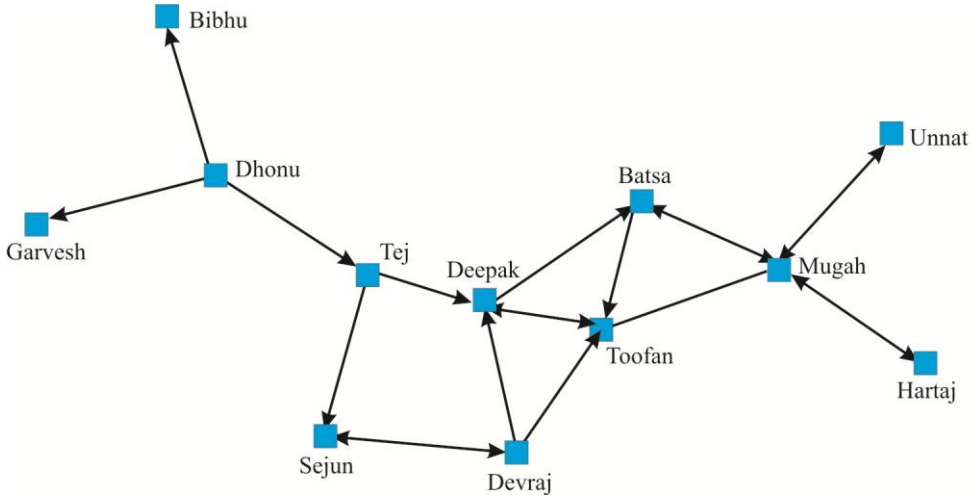

Fig 1.5: Graph G5 shows Resultant tracked dynamic community

6. Summary

This chapter includes an effective method for finding similarity and tracking communities in dynamic networks. Social network database review reveals that the approach proposed performs better than traditional static communities finding strategies which do not take in to account the temporal information into account. Furthermore, a third measure of similarity was proposed to include both new members and members leaving the subgroups, as the two currently defined similarity measures only take into consideration constant membership and new membership, which was a weakness of previous methods.

Reference

1. Memon, N., Harkiolakis, N., Hicks, D. [2008] 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, 206–215.
2. Memon, N., Larsen, H.L., Hicks, D.L., Harkiolakis, N. [2008] Detecting hidden hierarchy in terrorist networks: Some case studies. Lecture Notes in Computer Science 5075 , 477–489.
3. Clauset, A. [2005] Finding local community structure in networks. Phys Rev E 72:026132.
4. Chin, A., Chignell, M. [2008] Automatic detection of cohesive subgroups within social hypertext: A heuristic approach. New Rev Hypermed Multimed 14(1):121–143.
5. Chin, A. [2009] Social cohesion analysis of networks: a method for finding cohesive subgroups in social hypertext. PhD thesis, University of Toronto.
6. 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, pp. 113-139.
7. Radicchi, F., Castellano, C., Cecconi, F., Loreto, V., Parisi, D. [2004] Defining and identifying communities in networks. Proceedings of the National Academy of Sciences of the United States of America 101, 9, 2658– 2663.

RRIJM 2015, All Rights Reserved