Indian Journal of Science and Technology, Vol 9(20), DOI: 10.17485/ijst/2016/v9i20/78483, May 2016
ISSN (Print) : 0974-6846
ISSN (Online) : 0974-5645

Finding Hubs and Outliers in Temporal Networks

Anupama Angadi[1*] and P. Suresh Varma[2]

[1]IT Department, GMR Institute of Technology, Rajam - 532127, Andhra Pradesh, India; anupamaangadi.a@gmail.com

[2]Department of CSE, University College of Engineering, Adikavi Nannaya University, Rajahmundry - 533296, Andhra Pradesh, India; vermaps@yahoo.com

Abstract

Background/Objectives: Social Network Analysis (SNA) is the analysis of a social structure that is made up of a set of social players and a pile of the interactions between these social players. An individual such as a person, or an institution such as a college, agency and a federation, can be taken to be a social player. In late years, with the extensive function of social networking such as Facebook and Twitter, a vast sum of social interaction data has established social network analysis go beyond sociology and invite analysts from many fields. **Methods/Statistical Analysis:** Analysts have offered many different metrics to assess different topological features of a social player such as degree, betweenness centrality, eigen vector centrality etc. A distinctmetric is not adequate to examine multiple features of a social player, since each indicator designate a network in a dissimilar way so it is a reasonable solution to employ collective metrics with strong correlation (Spearman's or Pearson's). **Findings:** To find out the influential nodes the framework considers three egocentric metrics replacing social centric measures in temporal networks. Previous studies applied multiple social centrality measures with a strong correlation in static (or constant) networks. But many online social networks are naturally dynamic, propagate quickly in terms of social communications. Not all social players are born identical in a network, some might be superior in the sense they interconnected with almost all others and some might not contribute at all. The framework identifies these Hubs and Outliers at every snapshot. We have done experiments on undirected and unweighted EMAIL-ENRON real-world network. **Application/Improvements:** Influenial nodes can reveal new insights such as viral marketing, epidemic control, super-spreaders of disease and more generally in information dissemination.

Keywords: Combination of Genetic and Decision Tree, Consensus of Classifiers

1. Introduction

The problem of detecting dominant propagators in networks has been attracting a substantial part of the research community. The problem can be sub-divided in two; one is the identification of individual influential node and second is the group of nodes that render the influence more efficient. Focusing on identifying single spreaders, widely used socio centric metrics include the degree, betweenness, closeness, eigen vector, PageRank centralities. A social player with a high degree centrality[1] who have the top interactions to those near to them – they might be dominant, or just deliberately significant for communication. Social players with a high betweenness centrality[2] would designate that the individual is an vital caretaker of information between different parts of an organization. Social players with a high closeness value have a minor distance to all other players in the network and would so be efficient announcers of information. Each metric characterizes network in a different way (Figure 1) so it is an accepted solution to apply collective metrics, ought to be chosen carefully based on application[5,6]. However it is better to select metrics with statistical correlation[5,7-10].

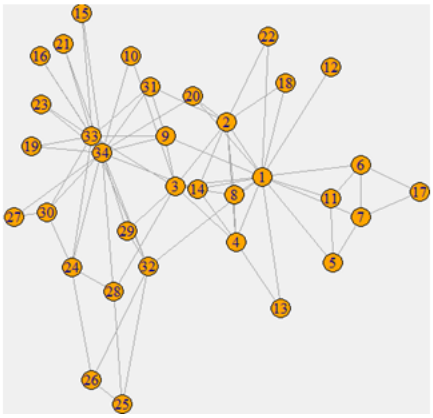
Previous studies identified only one Hub[11] for each community using multiple metrics with strong correlation[13] (Figure 2) and socio centrality measures in a static network[3,4] as shown in Figure 4. Social networks are not stable objects, where the interactions seam and vanish over a span of time, which in turn modify the metric values of the social players.When an edge adds to the Hub at a new timestamp, it increases the indicator value of the

* Author for correspondence

Hub and become more prominent and if an edge gets deleted from the Hub its indicator value decreases and the next Hub will become stronger. When all the edges of a node are with less influential nodes the framework treat it as an outlier. Inorder to track these influential nodes at every snapshot our framework identifies multiple Hubs and outliers.

Some players have interactions with almost all other players; in literature, such a player is known as Leader or Hub and some nodes might have less interactions; it is known as Outlier. Socio centrality measure involves the quantification of relationships between people in the whole network. Social networks are self-organized environments lacking centralized management operations, considering the whole network for quantifying relationships in personal networks is not an alternative at all. The alternative is to study the network of interactions neighbouring players rather than concentrating on the whole network shown in Figure 3. Thus, with its emphasis on individuals, it was anxious with making simplifications about the properties of personal networks; in literature it is called as Egocentric approach. In our study, we select collective metrics (degree centrality, ego-betweenness centrality and ego-eigenvector centrality) with strong correlation to assess the character of a node in a temporal networks.

The principal contributions of the framework are summarized as follows. firstly, we presented a spearman correlation; secondly, we present a framework to evaluate the node importance using ego-centric approach and in dynamic network; finally the investigational results on temporal data set shows how the framework can track the dynamic event changes and find influential social players (hubs, bridges and outliers).



ID	$C_{Soc}(Degree)$	$C_{Soc}(Betweenness)$	$C_{Soc}(Eigen)$
1	16	231.0714	0.657539519
2	9	28.47857	0.701590456
3	10	75.85079	0.838534476
4	6	6.288095	0.556485571
5	3	0.333333	0.213922655
6	4	15.83333	0.227545693
7	4	15.83333	0.227545693
8	4	0	0.450430215
9	5	29.52937	0.605076173
10	2	0.447619	0.271521686
11	3	0.333333	0.213922655
12	1	0	0.143296278
13	2	0	0.224862335
14	5	24.21587	0.599324757
15	2	0	0.27126871

Figure 1. Signifies Zachary's Karate Club network, a revision conducted at a US university designated by Zachary, W. (1977). Topology measures of first fifteen members on (right) showing single metric is not sufficient.

	degree	bet. Cen	eigen. cen	closen es.cen	info .cen	flow.b et	load.c en	graph .cen	stress .cen	power
Degree	---	0.9146	0.7752	0.8945	0.9691	0.868	0.908	0.603	0.8973	-0.483
bet. Cen	0.904	---	0.6928	0.8980	0.9008	0.807	0.998	0.679	0.996	-0.551
eigen. Cen	0.775	0.6928	---	0.8550	0.8572	0.429	0.694	0.502	0.7022	-0.281
closeness. Cen	0.894	0.8980	0.8550	---	0.9412	0.670	0.898	0.765	0.908	-0.506
info. Cen	0.969	0.9008	0.8576	0.9413	---	0.756	0.902	0.619	0.899	-0.464
flow. Bet	0.868	0.8070	0.4297	0.6706	0.5766	---	0.813	0.470	0.787	-0.503
load. Cen	0.908	0.9998	0.9081	0.8981	0.9025	0.8139	---	0.688	0.988	-0.547
graph. cen	0.603	0.6792	0.5022	0.7654	0.6196	0.4701	0.688	---	0.693	-0.270
stress. Cen	0.921	0.996	0.7022	0.9083	0.8990	0.7874	0.994	0.693	---	-0.538
Power	-0.489	-0.551	-0.2911	-0.506	-0.464	-0.503	-0.549	-0.291	-0.538	---

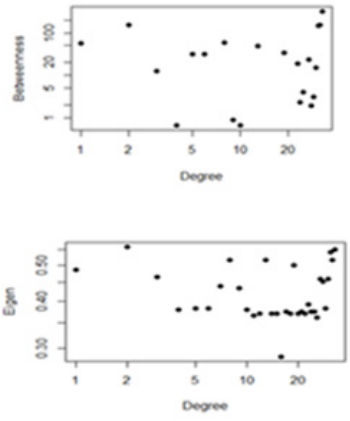


Figure 2. The Spearman correlation coefficients of player-level indicators of (Karate), shows the strong correlation between degree and other node-level indicators.

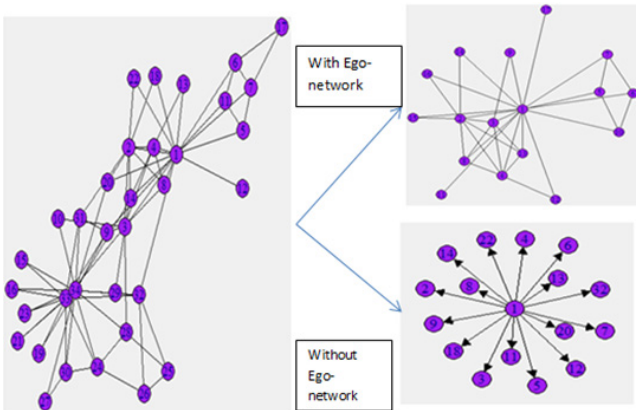


Figure 3. Karate dataset with & without Ego-network.

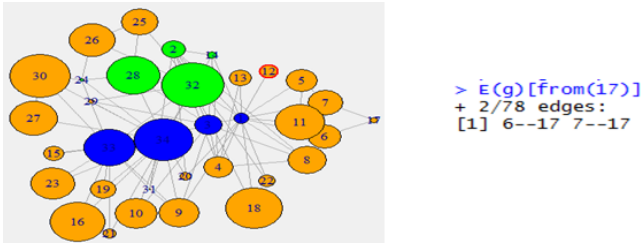


Figure 4. Influential nodes of Karate network(static), where a blue social player represents a Hub, a green player represents a Bridge. (b) Outlier in Karate data set, it has only two edges not connected to Hub node nor to a Bridge node.

2. Finding Influential Nodes In Temporal Networks

Degree, ego-betweenness, ego-eigenvector centralities are chosen for demanding no overall topology and their strong association with degree is also a slightbetter than social central measures as shown in Table 1. Correlation is a statistical technique used to measure the strength or direction of the relationship between two variables[11]. The spearman correlation coefficient is specified asfor a size n sample, the n raw scores X_i, Y_i are converted to ranks x_i, y_i and ρ is computed from:

$$\rho = 1 - \frac{6\sum d_i^2}{n(n^2-1)} \quad d_i = x_i - y_i$$

Finding influential nodes in temporal networks involves several snapshots, or timestamps, of the network, taking snapshots of the same network from different intervals and observed the changes of the network. Instead of updating the entire network with each iteration, this approach tracks those changed edges applies only those to the algorithm.

The procedure for finding influential nodes is: firstly, create a vector with three ego-centricmetrics of each player in the network; secondly, rank the node according to their indicators (Figure 5); third step calculate the average rank differences between indicators; fourth step select the hubs and bridges; fifth step get the edges of every node and find their ranking order. If all of its edges are with less ranking nodes, treat that node as outlier; finally for the remaining snapshots obtain the different in edges between the two snapshots and allow only these edges for ranking.

Algorithm 1: DIMI

Input: Graph Snapshots(0…..T)

1. run IMI for snapshot-0
2. for t in 1 thru T:
3. E = list of edges
4. for each edge in E
   if edge is an addition
      add edge to snapshot-t
   else
      remove edge from shapshot-t
   end if
5. end for
6. get the subgraph snapshot-t
7. ΔE = difference(snapshot-t , snapshot-t-1)
8. IMI(snapshot-t-1, E)
9. end for

Table 1. Comparison between Sociocentric and Egonetwork measures of Karate

	C_{SOC}(Degree)	C_{Ego}(Degree)	C_{SOC}(Betweenness)	C_{Ego}(Betweenness)	C_{SOC}(Eigen)	C_{Ego}(Eigen)
Degree	---	---	0.9146	0.92796	0.7752	0.7984
Betweenness	0.9142	0.9279	---	---	0.8032	0.8234
Eigen vector	0.9175	0.9345	0.8032	0.8212	---	---

2	8.0	9.0	11.0
9	78.0	33.0	30.0
11	56.0	50.0	61.0
14	9.0	7.0	74.0
36	2.0	18.0	7.0
42	1.0	15.0	0.0
52	80.0	43.0	66.0
57	32.0	22.0	82.0
85	46.0	10.0	94.0
86	119.0	14.0	15.0
99	4.0	3.0	96.0
125	5.0	64.0	1.0
127	42.0	34.0	54.0
136	144.0	3.0	1.5
141	3.0	18.0	0.0

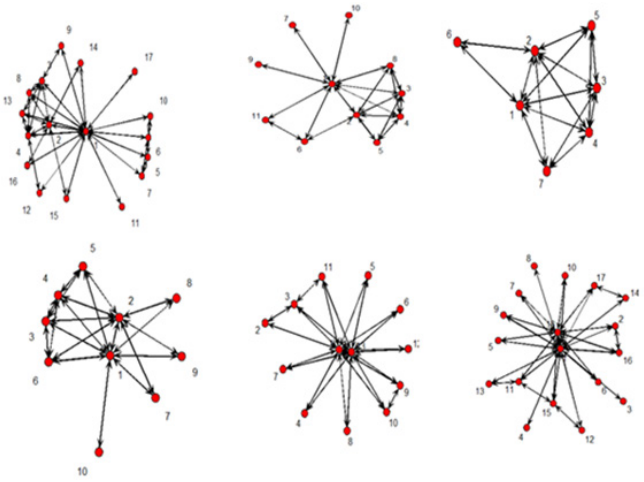


Figure 5. Ranks of ego-centric metrics of top 15 influential nodes and corresponding ego-networks in Enron.

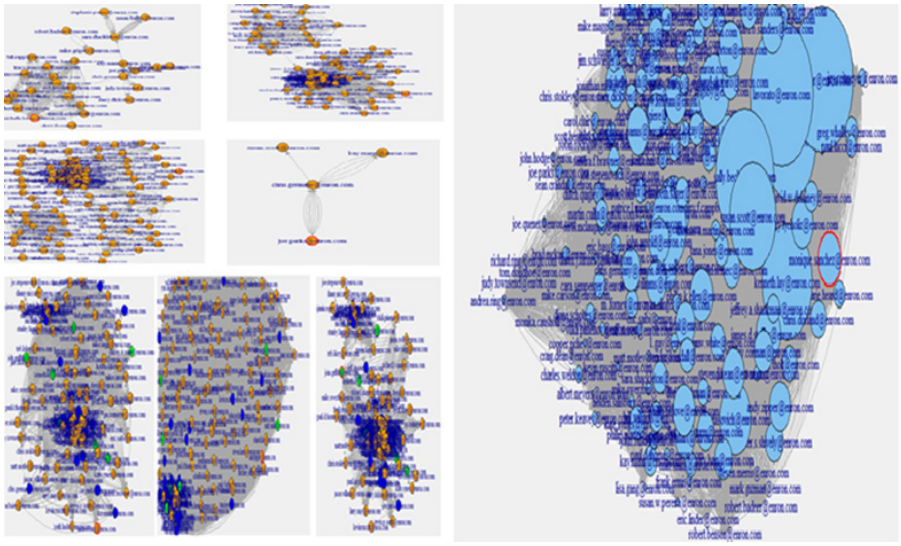


Figure 6. The Enron network formation. The size of the nodes views for the locus or prominence of the worker in the organization.

Algorithm 2: IMI(G,E)

Input: Graph G, edge changes E

1. N←VerticesCount (G)
2. M←EdgeCount (G)
3. R ← 2 *(M/N)
4. ego_diff ← 0, eig_diff ← 0
5. Hubs ← 0, Bridges ← 0, Outliers ← 0
6. Metrics < - (D(G)), egobetweenness(G), eigenvector(G))
7. Rank ←(rank(metrics[[1]])), rank(metrics[[2]]), rank(metrics[[3]])
8. tmp ← rank[[2]]
9. ego_diff ← sum(Rank[[1]]) - (Rank[[2]])/N
10. eig_diff ← sum(Rank[[1]]) - (Rank[[3]])/N
11. for each i in 1 to N
    if (Rank[[1]][i] > 0 and Rank[[1]][i] < = R and Rank[[2]][i] >= ego_diff and Rank[[3]][i] >= eig_diff)
    Hubs[num] ← i
    num ← num + 1
    end if
12. if(tmp[[i]] >= 1and tmp <= 2Rand Rank[[2]][i] < = ego_diff and Rank[[3]][i] > = eig_diff and i! = Hubs)
    Bridges [bnum] ← i

```
    bnum ← bnum + 1
    end if
13. totalrank[[i]] ← Rank[[1]][i] + Rank[[2]][i] + Rank[[3]][i]
14. Edges [[i]] ← E(g)[from(i)]
15.  if(Edges [[i]]! = Hubs) || (Edges [[i]]! = Bridges)
16. onum ← onum + 1
17. if onum = = #(Edges[[i]])
18.  Outliers[[i]] ← i
19. end if
20. end for
21. Display Hubs, Bridges, Outliers
22. Nodesize ← totalrank
23.  Plot the graph according to Order(Nodesize)
```

3. Evaluation and Results

The Enron dataset contains the 252 759 emails that 151 Enron employees exchanged during three years. It records information on the senders, receivers, and the moment they were sent. Note that by nature, the links are directed but for a fair analysis, we treated them as undirected in the present study. The above mentioned data contains temporal information. Hence, this data set has selected for evaluating the role of a node in temporal network. The Enron email dataset was made public by the Federal Energy Regulatory Commission during its investigation.

4. Conclusion and Future Works

On the source of correlation analyses of typical metrics, the framework suggests the methods to assess the prominence and the role of social players based on collective indicators with strong associations. The experimental results show the good performance of the proposed methods in analyzing the heterogeneous networks. And the previous studies have shown that the overall topology is highly steady with that based on ego networks. Therefore, the established approach would be adaptable to the huge, time-varying network whose precise global topology is always absent, such as the Internet and the social network of Facebook. In our upcoming work, we plan to extend our framework to detect communities based on prominent players.

5. References

1. Wasserman S, Faust K. Social Network Analysis (Cambridge Univ. Press, Cambridge, U.K.), 1994.
2. Freeman LC. A set of measures of centrality based on betweenness, Sociometr, 1977; 35–41.
3. Cao X et al. Identifying Overlapping Communities as Well as Hubs and Outliers via Nonnegative Matrix Factorization. Scientific Reports PMC 3 2993, 2013.
4. Haung S, Lv T, Zhang X, Yang Y, Zheng W, Wen C. Identifying node role in social network based on multiple indicators, 2014.
5. Valente TW, Coronges K, Lakon C, Costenbader E. How Correlated Are Network Centrality Measures? Connections (Toronto, Ont.). 2008; 28(1):16–26.
6. Ni C, Cassidy S. Using Social Network Knowledge For Detecting Spider Constructions In Social Security Fraud. ASONAM '13 Proceedings of The IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining. New York, NY, USA: ACM. 2013. p. 813–20. Print.
7. Bolland JM. Sorting Out Centrality: An analysis of the Performance of Four Centrality Models in real and simulated networks. Social Networks. 1998; 10:233–53.
8. Faust K. Centrality in affiliation networks. Social networks. 1997; 19(2):157–91.
9. Lee CY. Correlations among centrality measures in complex networks, 2006. No. physics/0605220.
10. Sun X-Q, Shen H-W, Cheng X-Q, Wang Z-Y. Degree-Strength Correlation Reveals Anomalous Trading Behavior, PLoS ONE. 2012; 7(10):e45598.
11. Sporns O, Honey C, Kotter R. Identification and classification of hubs in brain networks, PloSOne. 2007; 10:e1049.
12. Lu L, Zhang Y-C. Leaders in Social Networks, the DeliciousCase, Ed. Enrico Scalas. PLoS ONE, PMC. 2011; 6.6:e21202.
13. Baig, MB, Akoglu L. Correlation of Node Importance Measures: An Empirical Study through Graph Robustness. Proceedings of the 24th International Conference on World Wide Web Companion, International World Wide Web Conferences Steering Committee. 2015.