International Journal of Engineering Technology and Applied Science
(ISSN: 2395 3853), Vol. 2 Issue 3 March 2016

A Simulation Analysis of Degree Correlation and Assortativity in Signed Social Network Using Clustering Method

Suman Pandey
suman1791@gmail.com
M-Tech Scholar
Banasthali Vidhyapith, Jaipur (Rajasthan)

ABSTRACT-In this paper, we compared dissimilar commonly used clustering methods (k-mean) using Facebook dataset. In general, k-mean performs the best because of its complexity and the possibility of assigning the number of communities. Thus, due to the simplicity and convenience, we confirm that k-means works well when categorizing, detecting the communities. Firstly, we consider the features of users' profiles information while edge information, which refers to the connections of person, friends and among common friends, is also possibility to be utilized. Secondly, in the features files, rate of the appearance of 1's and 0's is very low. It means, most of the features are 0's. Here one problem of the rarely occurrence of 1's need to be concerned. So we choose k-mean clustering methods with the networks routines theory because of its popularity.

KEYWORDS: Clustering Co-efficient, K-Mean, Degree calculation, Average Degree, Balance algorithm.

I INTRODUCTION

Graph has repeatedly been used to ideal combinations among objects in an inclusive chain of domains from ranging metabolic signalling pathways [1] to the internet, and bill of indictment networks [2]. Distinguishing social networks communities has received handwriting on the wall of glare in late times. In social networks, addict communities provides outstrip recommendations and gathering pages of World Wide Web [3] boot be second-hand to gives greater appropriate results track. It is at the point of never vacant for clan to have a base hit opinion in complete topic. Here are continuously match to a different opinions and mutual media gat a charge out of blogs and saw in a new light websites which aprovideclient's statement of belief to describe squabble openly. Clients in a one terrain are accessible by computer whichever rightly or harmfully granted on certain terms whether they take or diverge with at variance user's opinion. We cut back still act with regard to networks to exemplar by this type relationships having notarized on the links. This type of civil networks are known as urban signed networks. Many society detection algorithms are located on a justification that dalliance mid any objects couple has the alike definition ubiquitous network. This algorithm cannot be presently registered to sign networks to what place the relationships among objects have uncountable understandings.

Newman [4] is a mostly recycled campaign to look groups in a nameless network. A Network is a group of nodes and edges and connections among nodes are called edges or links. This system taking the form of networks or graph in the world [5], Rapaport [6-11]. A composite network is the network topologies with non-trivial properties- this properties do not ensue in humble networks like frameworks but usually this is arise in real system of graphs modelling. Furthermost social, genetic and technological networks show considerable non-trivial topological properties or types, with models of links among their features which are neither morally regular nor entirely random. These properties comprise a substantial tail in the delivery of degree, a higher clustering constant, structure of community. In directed networks case these properties are also consider mutuality, triad significance profile and other properties. The largest complex network possibly understood through the networks with average number of interactions or correlations. This resembles with actuality that the all-out content information is acquired for moderate probabilities [15-16].

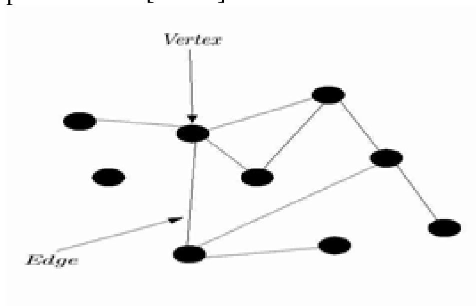

Figure 1 This is an example of small network which have 9 nodes and 9 edges.

II RELATED WORKS

In previous work, numerous algorithms have been propound for detection of communities in network individually in positive networks. They may be classify in three classes such as - [59] graph conceptual techniques for example Random walk techniques, Spectral techniques and physics based techniques; [60] divisive algorithms like as Girvan algorithm [61] and algorithm of Newman Tyler [6] these all algorithms are used to partition of the overall network in to many small size of sets or groups which are depends on the aloneness; [61] agglomerative algorithms like algorithms based on modularity [63] which calculate when the implementation is good. But these algorithms are just good for positive netw0rks but can't apply signed netw0rks.

Now, few algorithms propound for community detection in signed networks. Some people improved the above algorithms so that they could apply these. Networks, such as GN-H algorithms [64] which contains two phase. The first phase, we divide the positive sub networks; and on the second phase, it determines the final community structure based 0n the information of negative edges. Also, some people presented new algorithms.

Yang et al. Presented a new algorithms FEC which are depending on link density and sign of links [65]. The key idea of the algorithm is an agent based random walk model, which is follow the FC as known as find community phase it can detected the sink community. This type of community is finding outfrom the whole network which is based on few robust network or graph cut criteria in the EC phase .In FC phase a sink node is stable through agent and compute l-step transfer probability distribution function for each node. The l-step transfer probabilities then listed to find the nodes with lowest probabilities. The node with lowest l-step transfer probabilities show the nodes is outside to community and delete this node for detecting the community structure in the network. Kong LQ improved this algorithm by the functi0n of selecting target nodes, the method

Of steps automatic detection and so on [66].

III PROPOSED METHODOLOGY

Used Data Set Description

In this research, Facebook Dataset is used. This dataset consists of information of friends’ circles from Facebook, for instance, the profiles for each user (features), circles and network. Data is collected from participants who voluntarily provided their Facebook information to a Facebook. To protect personal information, all the data has been anonymized and the interpretation has all been obscured.

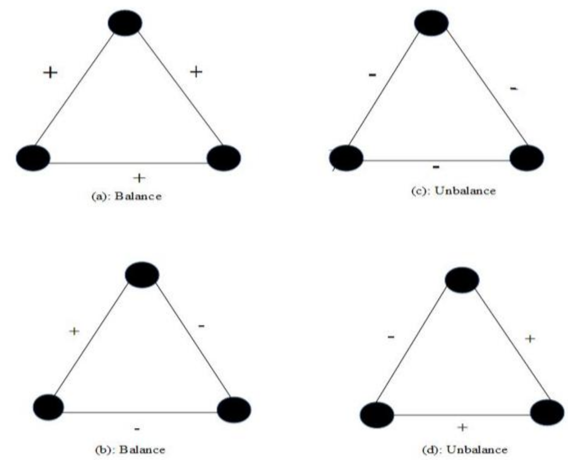

Figure 2 Triads for Undirected signed networks.

The number of positive edges called trio's with odd number of + are balanced as [(a) and (b)] and with odd number of negative edges [(c) and (d)] are unbalanced.

ALGORITHMS

Clustering Coefficient

Clustering coefficient for each node from an adjacency matrix. The clustering coefficient for each node in the graph is calculated from the given adjacency matrix. If the type is given, then the adjacency matrix is assumed to represent a graph of that type (either directed or undirected). If the type is not given, the graph is assumed to be undirected if the adjacency matrix is symmetric, and directed otherwise.

Clustering defines function: coeff = clustering (A); coeff = clustering (A, 'directed'); coeff = clustering (A, 'undirected')

Where Coeff: The column vector containing the clustering coefficient of each node, A is the adjacency matrix, type = 'directed'/'undirected'. The type of graph the adjacency matrix represents. If not given, the graph is assumed to be undirected if it is symmetric.

Function coeff = clustering (A, undirected)
n = size (adjacency matrix, 1);
ifstrcmp (type,' directed')
{

digraph = true;
elseif
strcmp (type,' undirected')
digraph = false;
else
shows error ('Type must be either "directed" or "undirected"')
end
else
if all values of adjacency matrix (all(A == A'))
digraph = false;
else
digraph = true;
end
end
if digraph
c = sum ((A^2).* A, 2);
else
c = sum ((A^3).* eye (n), 2);
end
Now calculate the out degree of the nodes: out = sum (A, 2);
Calculate the clustering coefficient: coeff = c. / (out.* (out - 1));
end

Degree Calculation

Calculate the overalldegree, out-degree and in-degree of a graph which based on the adjacency matrix; this should be produce weighted degrees of graph, if the input matrix is weighted matrix. An adjacency matrix is the input function here and degree, in and out degree sequences.

function [degree, indegree, outdegree] =degrees (adjacent)
indegree = sum (adjacent) and outdegree = sum (adjacent')
if is directed (adjacent)
then the total degree: degree = indegree + outdegree;
else
undirected graph: indegree=outdegree
degree = indegree + diag (adjacent)';
end

Average Degree

function k=average_degree (adjacent)
k=2*numedges (adjacent)/numnodes (adjacent);

Balance algorithm

Balance theory applied on the network when we have the large no of dataset, here we took Facebook dataset, person, friend, common friend all data are collected.

1. Apply the Facebook dataset with all the application information's and apply the sign positive and negative
2. The sign networks apply on the divergent patterns on both the links positive and negative.
3. All the link update on the network based on the rumours on the network.
4. Here we are taking the directed or undirected depend on the adjacency matrix.
5. Balance theory predicts their friend common friend, positive and negative status.
6. After that all positive and negative cycles and compare to each other's.
7. Here all summary are balance and have a proper positive and negative status.
8. After that design the balance theory network based on their status.

IV SIMULATION RESULTS

Facebook Network Community

Here first find the Facebook connection with the matrix data and make the network with the number of the people. All the people are connect with the adjacent matrix with the zeroes. Here person and friends and make the community and make a networks. Here person and friend community are connecting in the systematic way in Figure 4.7 and make the Facebook similarity community. Here person can update the friend in the network, when network is complete its extract the common friend of that person. The similarity is finding the common numbers of friend of the two persons. Here we used the undirected graphs with the maximum activity. All these process add in the network that design it will all condition of the Facebook community. Second community network is based on the process value to calculate the process value we used one formula based on the number of steps $2(t/N_{steps} \times 100)$, community structure are showing in the figure 3.

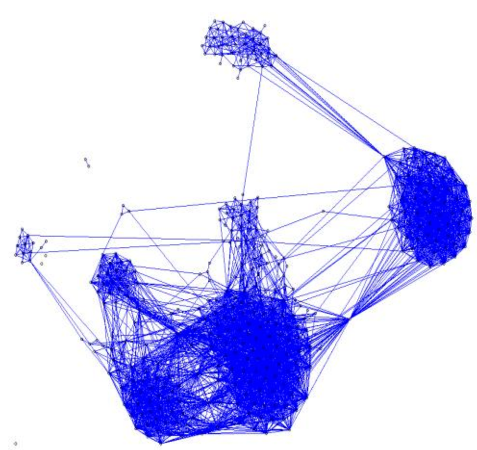

Figure 3: Facebook Network community based on the similarity.

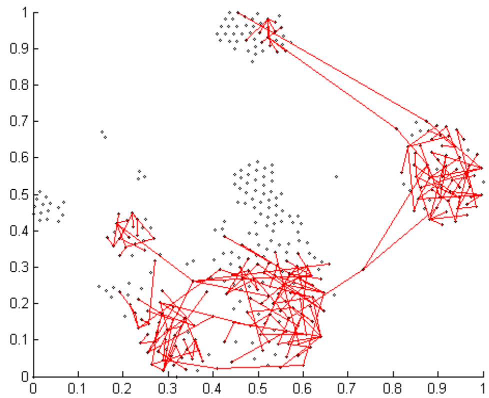

Figure 4: Facebook Network community based on process.

Facebook Network Correlations

Here we show the correlation among the person, friends and commons friends among two people. Calculate the correlations based on the number of meeting, infections based on the meeting and friends. The Correlations find among the nodes in the network based on the similar degree are frequently found with a mixing pattern of number of the noticeable networks. For example, in the Facebook networks, number of nodes be predisposed connected with other nodes or we can say other friends and person in the network with the similar degree. This predisposition is focused as assortative combination, or Assortativity. On second hand, the technical and natural networks characteristically show that dis-assortative mixing and dis-assortativity, with the high degree of nodes have a propensity to connect to low degree nodes.

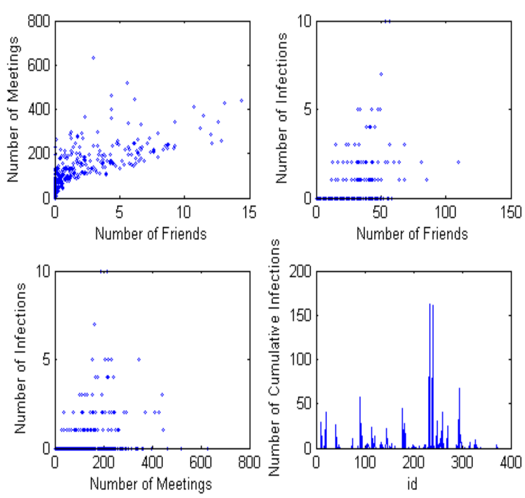

Figure 5: Facebook Network correlations

Facebook Network Categories Comparisons

The Facebook network categories comparisons are based on the rumour spreading that was introduced by Daley and Kendall [79] and it's a standard model to categories the social network data and we also call DK model. Here we assume that the total N number of people. All these people are categorized into three groups in the network: 1) ignorant, 2) spreaders and 3) stiflers, some notation alphabet is there to represent these three groups like **S**: represent that people who are totally ignorant of the rumour; **I**: showed that people who dynamically spread the rumour; **R**: represent that people who have heard the rumour, however no longer are interested in spreading it.

First issues is comes the rumour, what is it here; rumour is propagate throughout the population by a pair wise associates among spreaders or others then population. Any single spreaders concerned in the pairwise meeting or discussion then attempt to "contaminate" other individual with rumour. In the other case an ignorant individually, person can becomes a spreader. In the third and fourth cases, moreover one or both of those involved in the meeting or group want to learn that rumour is known after that decided that not to tell about the rumour anymore or anybody, by this means turning into stiflers. The probability of these cases with respect to time is showed here in figure 6.

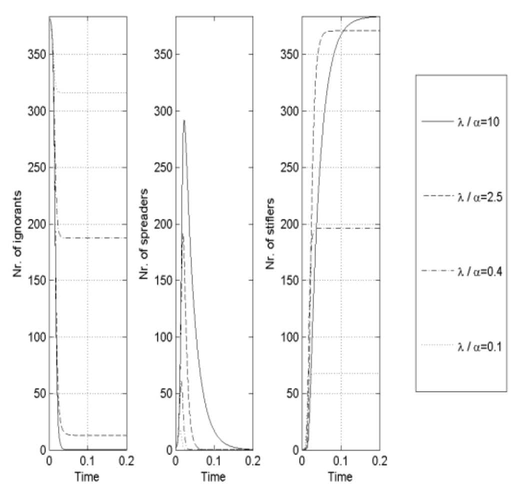

Figure 6: Facebook Network cluster Categories result based on time.

The investigational results make obvious that case of users' like, correlation value among communities and entire population, at the same time as smaller among communities. Furthermore, there are high value of correlation in stipulations of Likes category among dissimilar communities and among communities and whole population. These facts prove that Likes comprise a criterion of dissimilarity among the communities and verify the intuitions that lead us towards this research. The most popular rumours cases with respect to the population were showing with the experimental values in (Figure 7) and (Figure 8) showing the simulation result set on the ignorant with a probability.

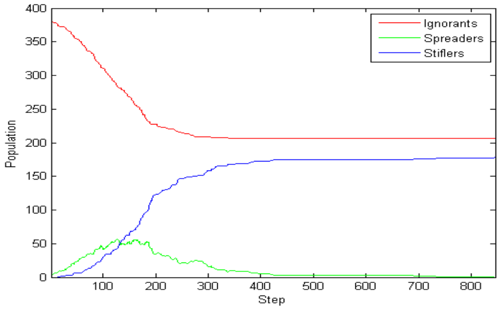

Figure 7: Facebook Network categories based on step and population.

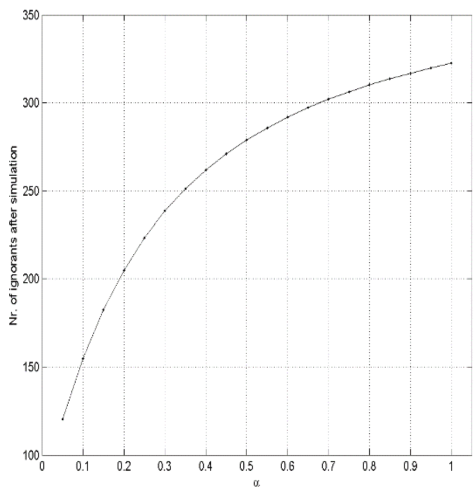

Figure 8: Number of ignorant after simulation.

V CONCLUSIONS

In this dissertation, we compared dissimilar commonly used clustering methods (k-mean) using Facebook dataset. In general, k-mean performs the best because of its complexity and the possibility of assigning the number of communities. Thus, due to the simplicity and convenience, we confirm that k-means works well when categorizing, detecting the communities. There is still some future work that could be done and get the best result. From the results, we have observed the accuracy is not good enough to find the proper result, many perspectives that need to be taken into consideration. Firstly, we consider the features of users' profiles information while edge information, which refers to the connections of person, friends and among common friends, is also possibility to be utilized. Secondly, in the features files, rate of the appearance of 1's and 0's is very low. It means, most of the features are 0's. Here one problem of the rarely occurrence of 1's need to be concerned. So we choose k-mean clustering methods with the networks routines theory because of its popularity.

VI FUTURE SCOPE

In the future work we can enhanced our k-mean algorithms with using k-mean LDA and Hierarchical clustering both the algorithm are the complex algorithm but apply to detection of community we can enhance the result of the current algorithms.The detection of community structure in a given complex networking using social Facebook data using dissimilar evolution over the other community.

REFERENCES

1. J. Kunegis,A. Lommatzsch, and C. Bauckhage, The slashdot zoo: mining a social network with negative edges, in Proceedings of the 18th .international conference on World wide web, ser. WWW'09, New York, NY, USA,2009, pp.741-750.
2. B. Yang, W. K. Cheung, and J. Liu, Community mining from signed social networks. IEEE Trans. Knowl. Data Eng., Vol. 19, no.10,2007, pp.1333–1348.
3. G. Facchetti, G. Iacono, and C. Altafini, Computing global structural balance in large-scale signed social networks, PNAS, vol.108, no.52, 2011,
4. P.Anchuri and M. Magdon-Ismail. Communities and balance in signed networks: A spectral approach, IEEE/ACM ASONAM, 2012.
5. J. Kunegis, S. Schmidt, A. Lommatzsch, J. Lerner,E. W. D. Luca, and S. Albayrak., Spectral analysis of signed graphs for clustering, prediction and visualization, In SDM, 2010, pp. 559-570
6. Introduction to Information Retrieval, Christopher D. Manning, PrabhakarRaghavan and HinrichSchütze, Introduction to Information Retrieval, Cambridge University Press, 2008.
7. Salamanos, Nikos, et al. "Discovering Correlation between Communities and Likes in Facebook." *Green Computing and Communications (GreenCom), 2012 IEEE International Conference on*. IEEE, 2012.
8. McAuley, Julian J., and Jure Leskovec. "Learning to Discover Social Circles in Ego Networks." *NIPS*. Vol. 2012. 2012.
9. J.R. Tyler, D.M. Wilkinson, and B.A. Huberman, Email as Spectroscopy: Automated Discovery of Community Structure within Organizations, C&T '03, 2003, pp. 81-96.

10. A. Clauset, M. E. J Newman & Moore, C., Finding community structure in very large networks, Physical Phys. Rev E, Volume 70,Issue 6,70(066111), 2004.
11. Li X, Chen HC, Li S, Exploiting emotions in social interactions to detect online socialcommunities. In: Proc. of the Pacific Asia Conf. on Information Systems, Atlanta: AIS, 2010,pp.1426−1437.