RESEARCH REVIEW JOURNALS
Volume-04
Issue-05
May-2019

ISSN: 2455-3085 (Online)
RESEARCH REVIEW International Journal of Multidisciplinary
www.rrjournals.com [UGC Listed Journal no : 44945]

Detection of Same Frequency Group in Social Networking

[1]Atul Agrawal & [2]Dr. Omprakesh Singh

[1]Research Scholar, OPJS University, Churu, Rajasthan, India
[2]Associate Prof., OPJS University, Churu, Rajasthan, India

ARTICLE DETAILS

Article History
Published Online: 25 May 2019

Keywords
Online Social Networks, Frequency, Detection

ABSTRACT

All Online Social Networks (OSNs) depends on client collaboration or participation of clients. Networks in social networks are shaped when set of clients communicate with one another often. Unequivocal people group are the consequence of conscious human choices. Understood people group are rising up out of the communications and exercises of clients in the social media. At the point when an on-screen character or items share more than one community covering networks are framed. Networks are not one of a kind and they shift contingent upon the use of explicit needs or condition to be fulfilled by the community. In this way different ways to deal with recognize covering networks from social networks have been proposed. For example, the condition to be fulfilled by a community can be the presence of way between the nodes or it tends to be based on the normal thickness of a particular community. The vast majority of the past works basically centered either to investigate sentiments at the tweet level or to study the attributes of tweeters in an associated situation. Here we planned an issue to discover certain networks from the gathering of clients thinking in same example on different issues. As such Same Wavelength Communities will be networks shaped on the basis of opinions or sentiments of comparative tone towards different issues by various people. Such same wave length networks or groups essentially interface the people in a significant and intentional crew.

1. Introduction

Network systems have been generally viewed as arbitrary structures and in spite of connections being considered to happen indiscriminately between nodes, most nodes were relied upon to have nearly a similar degree. In any case, huge commitments made by specialists like uncovered that the vertex availability of vast scale real-world networks really pursues a scale-free power-law distribution. That is, for extensive estimations of k, the fraction P(k) of nodes having k associations with different nodes in the network pursues the connection as appeared in equation 1, where c is a normalization constant and γ is a parameter usually ranging between 2 < γ < 3.

$$P(k) \sim ck^{-\gamma} \qquad \textbf{(1)}$$

The development of such networks includes a rich-get-richer plot (particular connection)

wherein new nodes have a higher likelihood to connection to nodes of higher degree, or it very well may be expressed that the 3 probability of a hub getting another connection is in extent to the hub's degree.

2. Type of social networks

Multi-mode network: A social network may include heterogeneous on-screen characters where every performer in the network speaks to a substance. The substances can be clients, recordings, labels or any occasions. For example, in a 3-mode network the three on-screen characters can be clients, recordings and labels as appeared in figure 4.1(Tang & Liu 2010). Different cooperations may exist among the substances in the above network. Clients may have contacts or communications among themselves. Clients can transfer recordings and every video can be labeled. In the meantime labels can have connections based on their semantic significance. Various performing artists and heterogeneous collaborations may prompt a multidimensional network.

Two-mode network: Two mode networks includes two sort of on-screen characters normally spoken to by bipartite diagram. A bipartite diagram is a chart G(U,V,E) where U and V are two disjoint arrangements of vertices with the end goal that each edge associates a vertex in U to a vertex in V . For example, in our concern we have two sorts of on-screen characters - clients and the inclining issues in which client having positive or negative opinion. Figure 2 demonstrates a case of bipartite diagram where (u1, u2, u3, u4) are set of clients and (e1, e2, e3, e4) are set of slanting issues or occasions. Each edge from clients to occasions speaks to the connection of clients to the occasion.

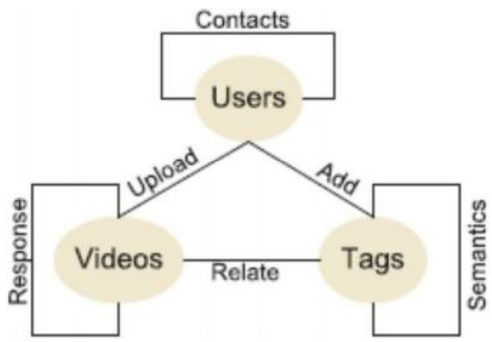

Figure 1 Three-mode network examples

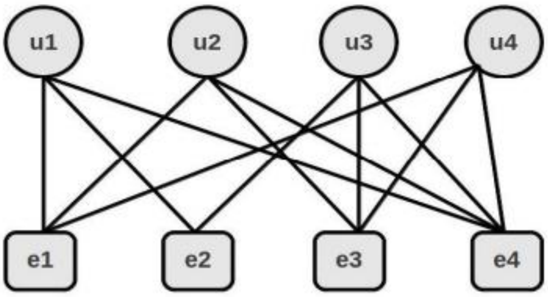

Figure 2: A bipartite graph with set of users and events example

3. Problem Statement

Let U = (u1,…..,um) indicate the arrangement of m particular clients and E = (e1, e2 ….,en) signify the arrangement of n unmistakable occasions. Consider C = (c1, c2,….,cn) indicate clubs produced based on the opinion towards n occasions. Discover all networks or a group which comprises the two clients and occasions that share same opinion on at least two occasions in the occasion set E where every community is a biclique or bipartite complete diagram. A biclique is a finished bipartite chart G (U,E) , ue is an edge in G typically spoken to by Km,n where m = |U| and n = |E|. These bicliques are called same frequency group. In the end the answer for the issue comes down to distinguishes the bicliques from a bipartite diagram produced based on the sentiments of clients on different drifting issues.

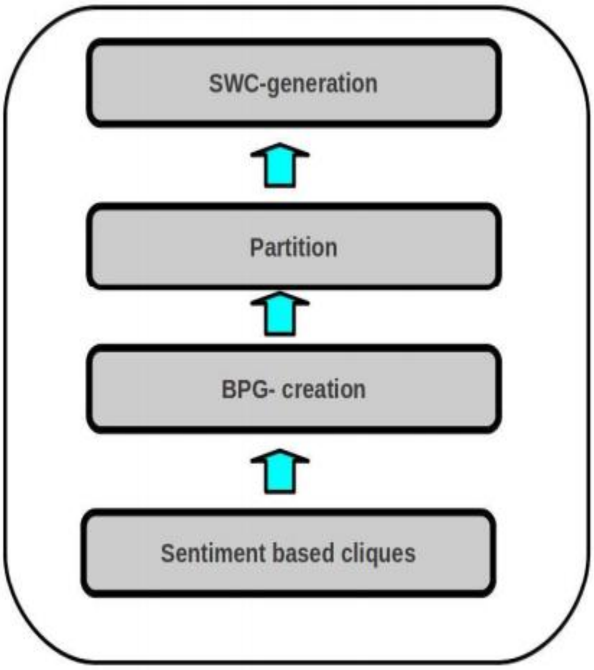

Figure 3: General frameworks for identifying SFG(SWC)

4. Research method

The SFG-FIND calculation creates a similar wavelength networks. The calculation is actualized in a dispersed situation utilizing parallel-python. The calculation acknowledges the bipartite diagram as info. Bipartite diagram is put away in a hash table as outlined in the Figure 4 C1, C2, C3… ,Cn, speaks to the name of the client groups in which clients having same opinion on the slanting issues. For example, Figure 5 shows of a toy bipartite chart and its portrayal. The quantity of edges episode on C1, C2, and C3 focuses to the rundown of clients having same sentiment on a particular issue. Henceforth applying hash work utilizing bunch name as key recovers the arrangement of clients associated with that issue. Since the quantity of clients might be vast, association of bipartite diagram as nearness grid requires enormous measure of room because of the scanty passages. Additionally hash table encourages the partitioning of bipartite diagram easily.

Parallel-python system is used. Number of partition chooses the quantity of same categories or bicliques to be produced. Speaking to bipartite chart as hash table encourages the way toward partitioning successfully by performing hash table inquiry utilizing bunch name as hash keys.

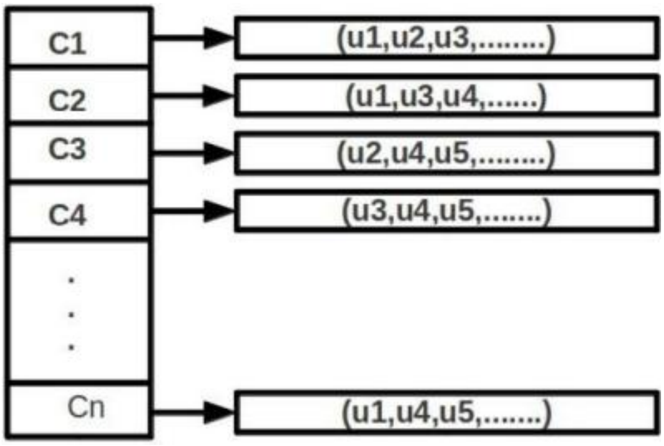

Figure 4: Hash table representation of bipartite graph

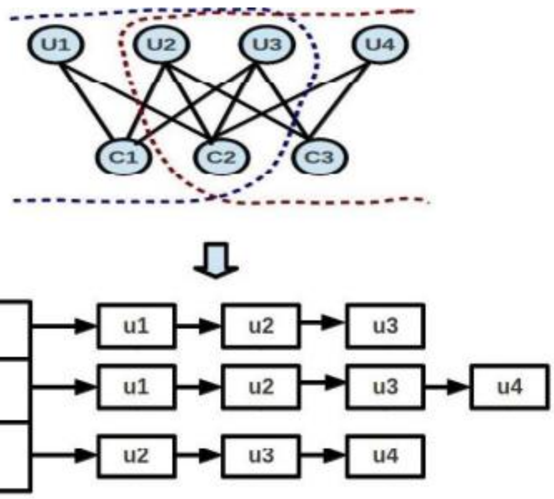

Figure 5: Hash table representation of bipartite graph example

5. Proposed algorithm

SFG-FIND algorithm identifies the same wavelength communities from the bipartite graph. PARTITION function implements the partition of bipartite graph. The number of bicliques depends on the number of partitions.

Algorithm 4 SFG-FIND algorithm

procedure SFG-FIND (G(U+C,E)) – G ← Bipartite graph E ← Occasion set having same sentiment on issues, U Set of client

S ← PARTITION(G(U+C,E))

for all partition p ε S do

for all c_i ε p do

for all Uc_i ε c_i do

$U_{sfsi} \longleftarrow$ CLIQUE-INTERSECT Uc_i , c_i
$SFG_i \longleftarrow$ CREATE-BICLIQUE , U_{sfsi} , c_i
end for
end for
end for
return SFG_i
end procedure

the total number of partitions will be $2^{|c|} - |C| - 1$. and the quantity of partitions decides the quantity of same wavelength networks where every community is a biclique. For a given a bipartite chart, testing whether it contains a biclique $K_{i,i}$, for a parameter i is a NP-complete issue (Garey & Johnson 1979). In this case additionally the quantity of bicliques to be created is exponential. So the issue is computationally unmanageable. A few methodologies are utilized to take care of computationally unmanageable issues which incorporate estimation, heuristic, parameterization and so on. Parameterization proposes the calculation time can be improved if certain parameters can be fixed. In the SFG-FIND calculation there are two parameters which choose the computational complexity - the quantity of clients and the quantity of issues in which clients make their opinions. A huge number of clients may react to an occasion or an issue. PARTITION function returns $2^{|C|} - |C| - 1$ partition where each partition is created by performing hash table lookup using the issue name in the subset of events as keys. When the partitions are created, CLIQUE-INTERSECT work finds the crossing point of partitions over all subset of occasions of length more noteworthy than one. Inner circle INTERSECT work restores the arrangement of shared clients U_{sfsi} having same opinion on various occasions. The nodes in each partition can be spoken to as set. At that point the execution of CLIQUE-INTERSECT work utilizing hash table.

6. Experiment results

Same examination set up referenced in the past parts is again utilized here. Two sorts of data sets are utilized. Engineered data set is utilized to confirm the rightness of SFG-FIND calculation. Client groups created by the sentiment analyzer from the tweets gathered on three slanting issues referenced in the past section are utilized to recognize same frequency networks. Identification of same frequency networks is an issue gotten from this. Subsequently ground truth data isn't accessible simply like numerous other social figuring tasks. So Phi-coefficient connection measure is utilized to discover the viability of the strategy.

6.1 Same Categories Based on the Real Time Tweets

Sentiment based client clubs (groups) created by the sentiment analyzer based on the three drifting issues are utilized to produce same wavelength networks. For example table 1 shows six client groups having either positive or negative opinion on the issues. (p1, p2, p3) indicates the positive and (n1, n2, n3) means the negative groups. Table 2 and figure 6 exhibit the particular categories and the quantity of clients engaged with every community produced from the sentiment based groups by the sentiment analyzer (table 4.1).

Figure 4.15 shows depiction of a SFG comprises clients having positive opinion on three occasions or issues meant by (e1, e2, e3).

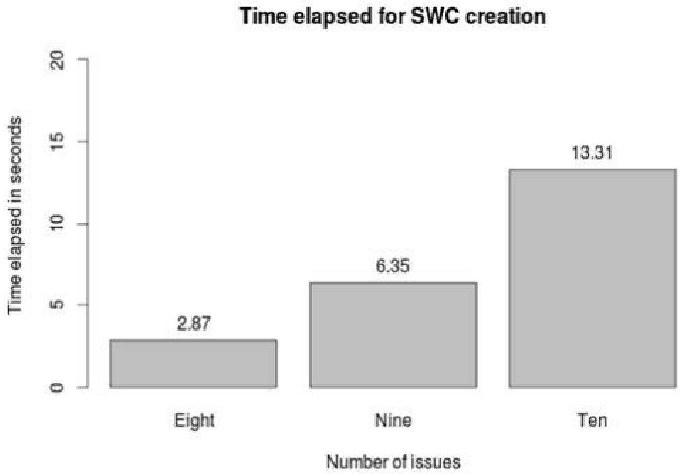

Figure 6: Running time for the SFG(SWC) calculation

Table 1: Sentiment groups based on the trending issue

Issue	Positive group(Yes)	Negative group(No)
Is Kejriwals quit from Delhi CM was good decision?	p1	n1
Is pepper-spray issue over Telgana issue deplorable?	P2	N2
Is blackout in parliament over Telgana deplorable?	P3	N3

Table 2: Same Categories based on the trending issues

S. No	SWCs	Number of users
SWCs from p1,p2,p3		
1	(p1,p2)	969
2	(p1,p3)	1528
3	(p2,p3)	736
4	(p1,p2,p3)	436
SWCs from n1,n2,n3		
1	(n1,n2)	799
2	(n1,n3)	1129
3	(n2,n3)	385
4	(n1,n2,n3)	238

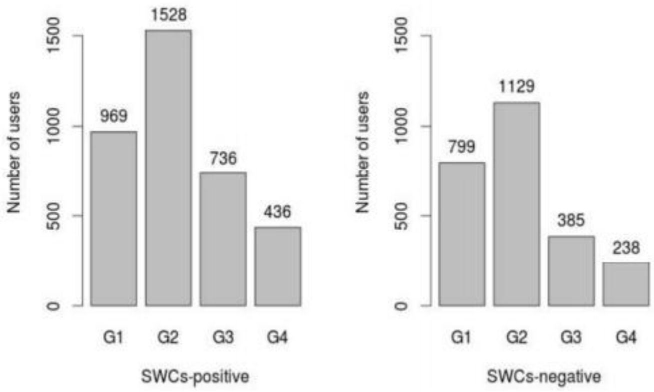

Figure 7: Distinct same categories and the number of users involved in each community

6.2 Effectiveness of the Algorithm

As referenced before identification of same wavelength networks is an issue gotten from the consequence of sentiment analysis of tweets on diverse issues and the execution relies upon the precision of the sentiment analyzer. So ground truth data isn't accessible for the constant tweets. One of the difficulties with the social registering task is the assessment without ground truth data. Thus we utilized phi-coefficient to discover the adequacy of the technique.

The phi-coefficient is really variety of Pearson's meaning of relationship coefficient r when the two conditions of every factor are given values of 0 and 1 individually. The phi-coefficient was intended for the correlation of really dichotomous conveyances, i.e., dispersions that have just two points on their scale which show some unmeasurable attribute. The phi-coefficient is particularly utilized in mental and educational testing to measure the association of two binary factors. This is otherwise called "mean square possibility coefficient" and indicated by φ. In the event that we have a 2X2 table for two attributes or factors, at that point the phi-coefficient that depicts the association between two variable is determined from the equation 1 where a, b, c and d speak to the frequencies of perception.

Table 3 Phi-coefficient 2 X 2 table

Attribute-1	Attribute-2	
	Yes	No
Yes	a	b
No	c	d

$$\frac{ad - bc}{\sqrt{(a+b)(c+d)(a+c)(c+d)}}$$

In this case we break down if there is any huge relationship exist among the people having same opinions in the initial two issues as referenced in the table 4 with the third issue by analyzing physically the sentiments of 100 example tweets from the clients reacted on three issues. Consider attribute-1, state X, is the clients reaction (positive or negative) on the initial two issues and attribute-2, state Y, is the clients reaction on the third issue. Phicoefficient table for the 100 example tweets is appeared in the table 5. The phi coefficient value determined utilizing the equation is 0.48. This demonstrates a positive connection and affirms the viability in anticipating client reaction for comparative sort of issues.

Table 4 Phi-coefficient 2 X 2 table for sample tweets

Attribute-1	Attribute-2	
	Yes	No
Yes	81	19
No	34	66

7. Conclusion

Opinions in OSNs have been distinguished as a solid measurement which initiates homophily. In this section a novel system is exhibited for distinguishing same wavelength networks or groups from online social networks like Twitter. The thought is to decide groups of individuals from the open who share same opinion on different issues or occasions. This is one unobtrusive approach to study the gathering reactions and personal conduct standards. The system is mapped to a chart hypothetical model and proposed a calculation which distinguishes the clubs based on the sentiments towards each issue and decides the covering bicliques that share similar sentiments towards a lot of issues. The calculation is actualized for a dispersed domain in parallel python and appeared to scale for a huge volume of data. The proposed calculation produces same wavelength networks in polynomial time for generally little arrangement of occasions. The analysis of such groups would be of assistance in disentangling their reaction designs and conduct highlights.

References

[1]. Tom Fawcett, Introduction to ROC Analysis, Elsevier B.V., Available online www.sciencedirct.com, 2005.

[2]. Tran Quang Hoa, Vo Ho Tien Hung, Nguyen Le Hoang, Ho Trung Thanh, Do Phuc, Finding the Cluster of Actors in Social Network based on the Topic of Messages, ACIIDS 04/2014, ThaiLan. Springer, 2014, pp. 183-190.

[3]. Wenjun Zhou, Hongxia Jin, Yan Liu, Community Discovery and Profiling with Social Messages, KDD'12, August 12–16, 2012, Beijing, China, 2012, pp. 388-396.

[4]. X. Wang, N. Mohanty, and A. McCallum. Group and topic discovery from relations and their attributes. Advances in Neural Information Processing Systems 18, 2006, pp. 1449-1456.

[5]. Zhijun Yin et. al, Latent community Topic Analysis: Integration of Community Discovery with Topic Modeling, ACM Transactions on Intelligent Systems and Technology, 2012, pp. 1-21.

[6]. Newman, M.: Fast algorithm for detecting community structure in networks. Physical Review E 69 (2003)

[7]. Raghavan, U.N., Albert, R., Kumara, S.: Near linear time algorithm to detect community structures in large-scale networks. (September 2007)

[8]. Meng, W., Feng, L., Bressan, S., Winiwarter, W., Song, W., eds.: Database Systems for Advanced Applications, 18th International Conference, DASFAA 2013, Wuhan, China, April 22-25, 2013. Proceedings, Part II. In Meng, W., Feng, L., Bressan, S., Winiwarter, W., Song, W., eds.: DASFAA (2). Lecture Notes in Computer Science, Springer (2013)