www.ijecs.in
International Journal Of Engineering And Computer Science
Volume 9 Issue 03 March 2020, Page No. 24978-24983
ISSN: 2319-7242 DOI: 10.18535/ijecs/v9i03.4449

A Note on Detection of Communities in Social Networks

P.Sridevi

Dept. of Information Technology Gayatri Vidya Parishad College of Engineering for Women
Visakhapatnam, India

Abstract:
The modern Science of Social Networks has brought significant advances to our understanding of the Structure, dynamics and evolution of the Network. One of the important features of graphs representing the Social Networks is community structure. The communities can be considered as fairly independent components of the social graph that helps identify groups of users with similar interests, locations, friends, or occupations. The community structure is closely tied to triangles and their count forms the basis of community detection algorithms. The present work takes into consideration, a triangle instead of the edge as the basic indicator of a strong relation in the social graph. A simple triangle counting algorithm is then used to evaluate different metrics that are employed to detect strong communities. The methodology is applied to Zachary Social network and discussed. The results bring out the usefulness of counting triangles in a network to detect strong communities in a Social Network.

Keywords: Strong, communities, Triangle counting, Social Networks, Clustering Coefficient.

1. Introduction

In recent years, there is a growing interest in understanding the structure, dynamics and evolution of complex Networks such as World Wide Web (www), Biological Networks, Technological Networks, Social Networks etc., [1]. A network is basically a set of items called the vertices or nodes with connections between them called edges. As graphs are a ubiquitous data representation that can be used to model complex relations in a wide variety of applications ranging from Social Sciences to Information Systems [2], graph theory can be used to study the complex networks modeled as graphs. The social network provides a record of global human interactions at a scale that is hitherto unprecedented and these are an invaluable resource for analyzing social allegiances, discovering entities with shared interests and identifying the key players in the social media [3]. It is observed that the size of social networks such as Facebook, Twitter, Instagram etc. with hundreds of millions of users and billions of social connections are growing day by day and an analysis of such networks is highly difficult. However, Graph theory provides techniques for fruitful analysis of these networks. Social network analysis can be used to identify important social actors, highly or sparsely connected communities and interactions among the various entities in the underlying network [4]. The Social networks differ from most other types of networks in two important ways namely network transitivity and assortative mixing or positive correlations [5]. Social networks are often seen as emerging from various social processes or mechanisms and the pattern of network ties in them tend to reveal the processes that have given rise to them [6]. Furthermore, in a social network, the distribution of edges is not only globally, but also locally inhomogeneous

With high concentration of edges within special groups of vertices and low concentration between these groups leading to the concept of community structure [7]. The community structure plays a significant role in the analysis of social networks and intense studies on this, is bound to reveal important patterns in the network aiding the analysis of the dynamics and structure of the system. The community structure is closely related

to triangles and the degree behavior of the triangles is an integral part of the structure [8]. The present work is basically aimed at utilizing the concept of triangle counting together with the associated metrics in finding strong communities in a social network.

A triangle is basically a transitive relation between three vertices and is an important building block and distinguishable feature of a network [9]. Further, a triangle is the shortest non-trivial cycle (i.e. a cycle of length 3) and the smallest non-trivial clique (i.e. a clique of size 3) that is at the heart of the definition of many important measures of network analysis such as the clustering coefficient, transitivity, triangular connectivity etc. [10].The concept of triangles have been used successfully in the detection of spamming activity, uncovering the hidden thematic structure of the web and link recommendation in online social networks [11]. A large presence of triangles in social networks is a consequence of the homophily principle which suggests that similar entities in a network tend to establish connections, such as people with similar interests, members of the same family or work mates [12].

There exist two categories of triangle-counting algorithms, the exact and the approximate. A brief review of the exact and approximate triangle counting algorithms is presented in [13]. In the exact counting method, triangle counting can be achieved by using the concept of eigen values or trace of the adjacency matrix of a graph. For triangle counting using trace, matrix multiplication is need to find the cube of the adjacency matrix. It is to be noted that both the matrix multiplication process and the eigen value computation process are time consuming when the matrix is large as in a social network. Parallelization and new matrix approaches have been developed to overcome some of the deficiencies in these methods. The present work utilizes both approaches for finding the number of triangles in a network and then using the associated metrics to find the strong communities in the social network.

2. Basic Concepts

Let us assume that a social network can be captured by an undirected sample graph $G = (V, E)$ where V is the nonempty set of vertices or nodes and E is a set of edges or connections. The vertices or edges may have a variety of properties associated with them in a social Network. For example, the vertices may represent men or women of different nationalities, religion, location, ages etc., while edges may represent friendship, animosity, professional acquaintance or location proximity etc. It is to be noted that if a vertex p is connected to vertex q and vertex q to vertex r then there is a high chance that vertex p is connected to vertex r. In terms of social networks, it translates to friend of your friend is likely to be your friend [14]. This is basically the property of network transitivity or clustering that leads to the presence of heightened number of triangles in the network. We present the definitions, metrics along with the corresponding results that have been put into use in the present work.

2.1 Adjacency Matrix

For a graph $G = (V, E)$ with vertex set $V = \{v_1, v_2, \ldots v_n\}$, the adjacency matrix of G is the $n \times n$ matrix $A = [a_{ij}]$ where $a_{ij} = \begin{cases} 1 & \text{if } \langle v_i, v_j \rangle \in E \\ 0 & \text{otherwise} \end{cases}$ i.e., a_{ij} is 1 if there is an edge from v_i to v_j. Some results on adjacency matrix that will be of use in the present work is outlined here for completeness.

1. For an undirected Graph G, the adjacency matrix is symmetric.
2. The (i,j) entry of A^n is the number of $v_i - v_j$ distinct walks of length n in G.
3. Cycles of length n are diagonal elements in A^n and a cycle of length 3, the global count of triangles $(\Delta(G))$ in the graph G is given by $\Delta(G) = \frac{1}{6}\mathrm{Tr}(A^3)$, where Tr refers to the trace operator of a matrix.
4. The sum of the eigen values of a square matrix is equal to its trace.
5. If λ is an eigen value of A then λ^n is an eigen value of A^n.
6. The total number of triangles $\Delta(G)$ in G is also given by $\Delta(G) = \frac{1}{6}\sum_{i=1}^{n} \lambda_i^3$

Where $\lambda_1, \ldots, \lambda_n$ are eigen values of A.

2.2 Degree of a Vertex

The Degree of a vertex is the number of edges incident with that vertex. It is a measure of the connectedness of a person (node) with other

persons (nodes) in the network. A vertex with maximum degree can highlight the influence of the community around that node. The degree of vertex V is denoted by $deg(V)$.

2.3 Clustering Coefficient & Transitivity of a Graph

In the seminal paper [15], the authors propose a model that explains several properties of social networks such as the abundance of triangles and the shortest paths among any pair of nodes. They also introduced such as clustering coefficient which is a measure of the frequency of triangles. The clustering coefficient $C(v)$of a vertex $v \in V$ with degree $deg(v)$ defined by

$$C(v) = \frac{\Delta(v)}{\binom{deg(v)}{2}} = \frac{\text{no of triangles incident to vertex v}}{\text{no of wedges centered at v}}$$

The global clustering coefficient also known as the transitivity denoted by $C(G)$ is defined as

$$C(G) = \frac{3\times\text{total number of triangles}}{\text{total number of wedges}} = \frac{3\times\Delta(G)}{\sum_{v\in V}\binom{deg(v)}{2}}$$

It is to be noted that a wedge is a path of length 2 and large clustering coefficients are considered a manifestation of the community structure. The local clustering coefficient $C(v)$ has been used quite widely in the sociological literature wherein it is referred to as the network density [14]

3. Methodology

In this section, we present two algorithms for triangle counting; one is based on computing the trace of the cube of the adjacency matrix while the other is based on finding the sum of the cubes of the eigenvalues of the adjacency method.

3.1 Algorithm 1- Trace Algorithm

Step1: Read the $n \times n$ adjacency matrix A of the graph G consisting of n vertices.

Step2: Calculate A^3 by matrix multiplication.

Step3: Find the trace of A^3 i.e., Tr (A^3).

Step4: Calculate$\Delta(G) = \frac{Tr(A^3)}{6}$.

Observations:

1. If the adjacency matrix consists of $N = 2m$ non-zeroes then the number of operations for computing $Tr(A^3)$ is $4Nn + N - 1$.

If using Hadamard product and sparse matrix-vector products [16], then number of operations is $2N(n+1) - 1$ which is approximately half the number of operations observed made in observation 1.

3.2 Algorithm 2- Eigen value Algorithm

Step 1: Read $n \times n$ Adjacency matrix A of Graph G.

Step 2: Find the eigen values $\lambda_1, \ldots, \lambda_n$ of A.

Step 3: Calculate $\lambda_1^3, \ldots, \lambda_n^3$.

Step4: Compute $\Delta(G) = \frac{1}{6}\sum_{i=1}^{n}\lambda_i^3$

Observations:

1. The eigen values can be computed using the LancZos method and is based on matrix-vector products that is easy to parallelize [17].
2. The QR algorithm can be employed to find the eigen values of A and is available as a direct command in Matlab.

4. Numerical Experiment

The calculations were carried out on a computer with Intel Core2 Duo E 4500 processor @2.20Ghz and 4GB of RAM. The total number of triangles in a random G were calculated using both the Trace Algorithm that employs matrix multiplication and the Eigen value Algorithm that calculates the eigen values of the adjacency matrix of the Graph G. A random symmetric matrix having a specified number of vertices/rows (columns) with 0,1, entries and 0 on the diagonal is generated with a probability parameter controlling the density (number) of edges. The two algorithms were run for the randomly generated adjacency matrix of the social graph having vertices 500, 750, 1000, 250, 1500, 1750 and 2000 and having 70% of the edges of a corresponding complete graph. The system time for running both the algorithms were calculated and compared in Fig 1.

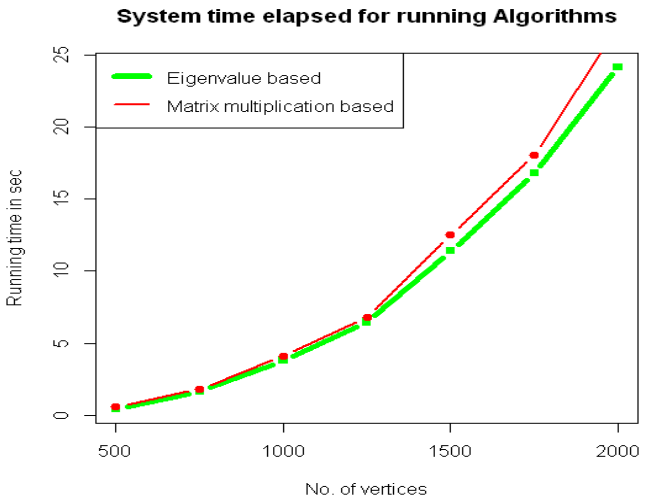

Fig. 1: System time elapsed for running Algorithms

From the presented Figure, it is noticed that the Eigenvalue method is faster than the matrix multiplication method in calculating the total number of triangles in the graph. Using Hadamard products for matrix multiplication and Arnoldi - Lanczos method for Eigen value calculations significantly lowers the running time for both the algorithms.

5. Case Study

As a case study we consider the well-known and much used Zachary Karate club network, The data was collected from members of a university Karate club by Wagne Zachary in 1977 [18]. Each node in the corresponding social graph represents a member of the club and each edge represents a tie between two members of the club. The graph of the network consists of 34 vertices and 78 edges. The matrix norm of the adjacent matrix is 6.7257 and the condition number of the matrix is 1.327e+18. The matrix has a rank of 24 and null space dimension of 10. A plot of the graph is given in Fig. 2. The local clustering coefficient is dependent on the number of triangles incident to a vertex. The local friendship structure or connectivity of each entity can be obtained by counting how many triangles a vertex is part of. The corresponding connectivity of the network is shown in Fig. 3. The maximum number of triangles a vertex is part is 18.

The average clustering coefficient of the graph is 0.5879 which is an average of the local clustering coefficients. This metric places more weights on the low degree nodes. Around 11 vertices of the graph have a local clustering coefficient 1 which quantifies how close its neighbors are to being a clique. The network has a triangle count of 45 which is validated from the two algorithms and has a wedge count of 528. Pons and Latapy [19] developed an algorithm to find communities in a graph via random walks basing on the idea that short random walks tend to stay in the same community. On using its implementation from igraph R package, we find that the network is divided into 5 groups based on the random walks and a pictorial representation is given in Fig. 4. The basic interest lies in finding groups of vertices within which connections are dense, but between which connections are sparser. A number of algorithms that appear to work with real-network data is finding the community structure is outlined in [20].

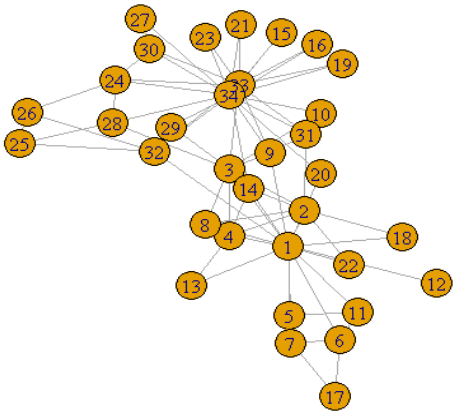

Fig. 2: Graph of Zachary Network

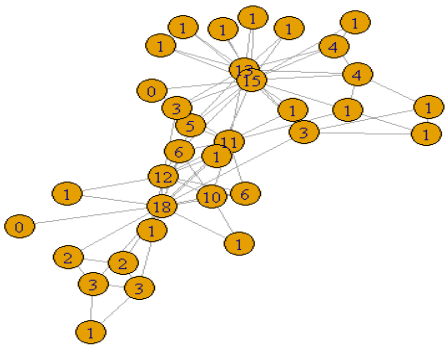

Fig. 3: Triangle count at each vertex of Zachary network

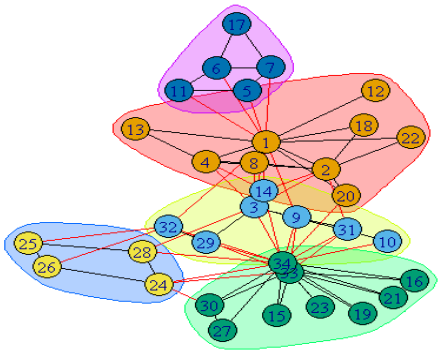

Fig. 4: Community structure using Random Walks

On experimentation it is observed that most of the methods appear to be ideal in detecting community structure but the metrics such as the transitivity (local and global) as well as the average clustering coefficient appear to give an approximate idea to these algorithms in detecting the community structure.

6. Conclusion

The calculation of the modularity and detection of the community structure of a social network depicted by its adjacency matrix is studied through the metrics associated with triangle counting. The clustering coefficient defines the quality of the communities and is an important metric. Methods at reducing the computational time of the algorithms for triangle counting are necessary for large social networks.

7. References

[1.] Thomas Schank and Dorothea Wagner, "Approximating Clustering Coefficient and Transitivity", J. Graph Algorithms and Applications, vol. 9, no. 2, pp. 265-275, 2005.

[2.] Luca Becchetti Paolo and Boldi Carlos Castillo, "Efficient Algorithms for Large Scale Local Triangle counting", ACM Trans. in Knowledge Discovery from Data, vol. 4, no. 3, pp. 13;1-27, 2010

[3.] Benjamin Paul Chamberlain, "Real-time community detection in full social networks on a laoptop", PLoS ONE, vol. 13, no. 1:eo188702,2018

[4.] S.Wasserman and K.Faust, 'Social Network analysis: methods and applications, Structural analysis in social Sciences, Cambridge University Press, NewYork USA,1994

[5.] M.E.J Newman and Juyong Park. 'Why social networks are different from other types of networks' Physical Review E 68,036122, 2003.

[6.] D.Lusher and G.Robins, 'Formation of Social Network Structure in Exponential Random Graph Models for Social Networks, Cambridge University Press, 2013

[7.] Santo Fortunato, Community detection in graphs, Physics Reports, 486, 75-174, 2010

[8.] C.Seshadri, T.G.Kolda and A.Pinar, 'Community structure and scale free collections of Erdos-Renyi graphs,, Physical Review E.85,056109,2012.

[9.] N.Durak, A.Pinar,, T.G.Kolda and c.Seshadri, 'Degree relations of trianglesin Real World Networks and Graph models ', CIKM'12,Oct-29-Nov2,HI,USA,2012.

[10.] Shumo Chu, James Cheng.Triangle Listing in Massive Networks and its Applications, KDD'11, Aug 21-24, California, USA, 2011

[11.] C.E.Tsourakakis, ' Counting of triangles in Real Networks, using projections, know,Inf.Syst, 26:501-520,2011

[12.] A.P-Perez, D.D.SAL,J-m.Brunat, J-L.L-Pey., ' put three and three together' : Triangle-Driven Community Detection, ACM Trans.on.Know. D.S.C from Data, 10(3), 22, 2016.

[13.] M.N.Kolountzakis, G.L.Miller,R.Peng and C.E.Tsourakakis, 'Efficient Triangle Counting in Large graphs via degree based vertex Partitioning', Internet Mathematics,8(1-2),161-185,2011

[14.] M.E.J. Newman, 'The structure and Function of complex Networks' SIAM Review, 45(2),

[15.] D.Watts and S.Strogatz,'Collective dynamics of small world networks' , Nature 393, 440-442,1998.

[16.] P.Burkhardt, 'Graphing.E.Tsourakakis, ' Fast counting of triangles in Large Real Networks:Algorithms and Laws', Eighth IEEE. Intl. conference on DataMining, Pisa, Italy,2008.

[17.] W.W.Zachary, 'An information flow model for conflict and fission in small groups', J.Anthropological Research 33, 452-473.1977.

[18.] P.Pons and M.Latapy, 'Computing communities in large networks using random walks, J.Graph Algorithms and Applications 10(2), 191 – 218, 2006

[19.] M.E.J.Newman, ' Detecting community structure in Networks', Eur. Phys. J.B,38,321-330,2004