International Journal of Computer Applications Technology and Research
Volume 4– Issue 12, 889 - 893, 2015, ISSN: 2319–8656

The Mathematics of Social Network Analysis: Metrics for Academic Social Networks

Tasleem Arif
Department of Information Technology
Baba Ghulam Shah Badshah University
Rajouri, J&K, India

Abstract: Social network analysis plays an important role in analyzing social relations and patterns of interaction among actors in a social network. Such networks can be casual, like those on social media sites, or formal, like academic social networks. Each of these networks is characterised by underlying data which defines various features of the network. Keeping in view the size and diversity of these networks it may not be possible to dissect entire network with conventional means. Social network visualization can be used to graphically represent these networks in a concise and easy to understand manner. Social network visualization tools rely heavily on quantitative features to numerically define various attributes of the network. These features also referred to as social network metrics used everyday mathematics as their foundations. In this paper we provide an overview of various social network analysis metrics that are commonly used to analyse social networks. Explanation of these metrics and their relevance for academic social networks is also outlined

Keywords: Social network analysis, mathematics, metrics, academic networks

1. INTRODUCTION

Social networks have been around us since time immemorial. They can be found around us in a variety of shapes and forms. They can be real or virtual but their basic properties still remain the same. Social network has been defined in different ways but their reliance on the fundamentals of mathematics has remained unchanged in almost all of the definitions. Social networks have been and are increasingly being represented through mathematical representations like graphs, matrices and relations. A standard definition of social network can be found in [1] as "a structured representation of the social actors (nodes) and their interconnections (ties)". These networks can be represented as a graph G = (V, E). The set V denotes entities (people, places, organizations, Webpages, etc.) joined in pairs by edges in E denoting acquaintances or relationships (friends, siblings, co-authors, hyperlinks, etc.) [2].

Social networks are ubiquitous characterized by underlying social groups that share common interests. These networks and the underlying groups have emerged on the Web at a rapid pace and have become one of the widely used online activity [3]. These networks are an aggregation of groups or virtual communities with each of these communities different from the other in composition, purpose and intent. Members of these virtual communities profit from being linked to other people sharing common interests despite their geographically dispersed affiliations. Social networks can be constructed for business entities like a company or firm, for educational entities like a school or University, or for any other set of entities [4].

Social networks have got a lot of attention from the research community long before the advent of the Web [5]. Between 1950 and 1980, when Vannevar Bush's proposed hypertext medium 'Memex' was gaining acceptance, Social Sciences also contributed a lot in measuring and analyzing social networks [6]. There are numerous examples of social networks formed by social interactions like co-authoring, advising, supervising, and serving on committees between academics; directing, acting, and producing between movie personnel; composing and singing between musicians; trading and diplomatic relations between countries; sharing interests, connections, and transmitting infections between people; hyper linking between Web pages; and citations between papers.

There have been a number of efforts to study these networks with the first formal attempt of its kind undertaken around eight decades ago. Manual methods can be used to analyze these networks but with new actor and relationship the complexity of the network increases many folds thus rendering manual methods ineffective. With the advent of Internet and developments in information and communication technologies the size, reach and diversity of these networks has become immense. One can't even think of analyzing these networks with a simple computer leave aside manual techniques. One can't even think of keeping aside these networks because they engage enormous number of users thus providing a large customer base to businesses on the one hand and the most common medium of interaction among geographically separated users on the other. To meet this demand a specialized science called Social Network Analysis (SNA) with its roots in social sciences particularly Sociology has emerged. With each passing day the dependence of these networks on mathematics and mathematical tools has been increasing.

The focus of Social Network Analysis (SNA) is relationships, their patterns, implications, etc. Using it, one can study these patterns in a structural manner [5]. SNA can be used to identify important social actors, central nodes, highly or sparsely connected communities and interactions among actors and communities in the underlying network [5]. SNA has been used to study social interaction in a wide range of domains, e.g. collaboration networks [7], directors of companies [8], inter-organizational relations [9], etc.

The study of social networks for behaviour analysis of actors involves two aspects: (a) the use of formal theory organized on the basis of mathematical conventions and (b) the empirical analysis of network data as quantified by various social network analysis metrics. So it can be understood that

social network metrics play an important role in SNA. This paper identifies various social network metrics and the mathematics behind them. These metrics have different meanings in different types of networks. In addition this paper also examines the use and relevance of these metrics in academic social networks.

2. SOCIAL NETWROK ANALYSIS: LEVELS AND METRICS

Like in other fields metrics help define certain attributes in quantitative terms. This section illustrates different levels of social network analysis along with the metrics that are used to draw inferences about the network.

There are five different levels of social network analysis, each of them characterised by the structure of the underlying network. It may be at actor level, dyadic level, triadic level, subset level, or network level. Metrics like centrality, prestige and roles such as isolates, liaisons, bridges, etc. are used to analyse the social network at actor level, whereas distance and reachability, structural and other notions of equivalence, and tendencies toward reciprocity are important at dyadic level. At triadic level one is interested in balance and transitivity. At subset level one is interested in finding cliques, cohesive subgroups, components whereas metrics like connectedness, diameter, centralization, density, prestige, etc. are used for analysis at network level[1].

Some of the commonly used SNA metrics are:

Centrality: As said earlier 'relationships' is the focus of SNA and the 'actors' are central to all types of relationships. Thus attribute description or profiling of actors is an important aspect of any social network analysis. In this context Chelmis and Prasanna [10] proposed several social network analysis measures (metrics) that can be used to indentify influential nodes in a social networks. Centrality is a measure of the information about the relative importance of nodes and edges in a graph. Centrality measures like Degree Centrality, Closeness Centrality, Betweenness Centrality, Eigenvector Centrality, Katz Centrality and Alpha Centrality play an important role in graph theory and network analysis to measure the importance or prestige of actors or nodes in a network2. Several centrality measures like betweenness centrality, closeness centrality, and degree centrality have been proposed in [10] to identify the most important actors (leaders) in a social network.

- ***Degree Centrality:*** It is the simplest of all the centrality measures and its value for a given node in the network is the number of links incident on it and is used to identify nodes that have highest number of connections in the network. However it does not takes into account the centrality or prestige of the incident nodes. For a graph $G = (V, E)$, the degree of a node or vertex $v, (v \in V)$ can be expressed using Equation (1).

$$C_D(v) = deg(v) \tag{1}$$

where $deg(v)$ is the number of edges incident on the vertex *v*.

For entire graph *G* the *Degree Centrality* can be expressed using Equation. 2.

$$C_D(G) = \frac{\sum_{i=1}^{|V|}[C_D(v*) - C_D(v_i)]}{H} \tag{2}$$

Where $v*$ is the node in G with highest degree centrality and $H = \sum_{j=1}^{|y|} C_D(y*) - C_D(y_j)$, where $y*$ be the node with the highest degree centrality in a graph X of G with Y nodes. The value of H is maximum when a graph has a star like structure.

- ***Eigenvector Centrality:*** A more sophisticated version of degree centrality is eigenvector centrality. It not only depends on the number of incident links but also the quality of those links. This means that having connections with high prestige nodes contributes to the centrality value of the node in question. Google's *PageRank* and *Katz Centrality* is a variation of eigenvector centrality and closely related to eigenvector centrality respectively.
  Let $A = (a_{v,u})$ be the adjacency matrix of a graph G with V vertices and E edges. Then A can be defined as:

$$A_{v,u} = \begin{cases} a_{v,u} = 1, if\ vetrex\ 'v'\ is\ linked\ to\ vertex\ 'u' \\ a_{v,u} = 0, otherwise \end{cases}$$

The eigenvector centrality of a vertex v can be defined using Equation (3).

$$C_E(v) = \frac{1}{\lambda}\sum_{u \in N(v)} x_u = \frac{1}{\lambda}\sum_{u \in G} a_{v,u} x_u \tag{3}$$

where $N(v)$ represents the set of neighbours of the vertex *v* and λ is a constant.

- ***Closeness Centrality:*** The degree of nearness (direct or indirect) between any node and rest of the nodes in the network is represented by "*closeness centrality*". It is the inverse of sum of the shortest distance (also called geodesic distance) between a node and rest of all in the network. For a graph *G* with '*n*' nodes the closeness centrality of a node '*v*' can be expressed using Equation (4).

$$C_C(v) = \frac{n-1}{\sum_{k=i,}^{n} d(u_i, v)} \tag{4}$$

where $d(u_i, v)$ denotes the geodesic distance between u_i and *v*.

- ***Betweenness Centrality:*** In order to identify the leaders in the network, the quantity of interest in many social network studies is the "*betweenness centrality*" of an actor '*i*'. Betweenness centrality measures the fraction of all shortest paths that pass through a given node or in simple terms it quantifies the number of times a node acts as a bridge along the shortest path between two other nodes. Nodes with high betweenness centrality play a crucial role in the information flow and cohesiveness of the network and are considered central and indispensable to the network due to their role in the flow of information in the network. Nodes with the high betweenness act as

[1]http://lrs.ed.uiuc.edu/tse-portal/analysis/social-network-analysis/#analysis

[2] http://en.wikipedia.org/wiki/Centrality

gate keeper. The betweenness centrality of vertex *v* can be expressed using Equation (5).

$$C_B(v) = \sum_{s \neq v \neq t \in V} \frac{\sigma_{st}(v)}{\sigma_{st}} \quad (5)$$

where σ_{st} is the total number of shortest paths from node s to t and $\sigma_{st}(v)$ is the number of paths that pass through v.

- ***Clustering coefficient:*** It signifies how well a node's neighbourhood is connected[3]. Clustering coefficient is a measure of the ability of a node's neighbour to form a complete graph, also called a clique. The value of clustering coefficient is directly proportional to the degree of connectedness of the neighbours of that node: more the connections among the neighbours, the higher the clustering coefficient. The clustering co-efficient of a network as given in [11] is the average of the clustering co-efficient of all the nodes in the network. It is therefore considered to be a good measure if a network demonstrates "*small world*" behaviour [11]. Stanley Milgram's [12] theory of the "*6 Degree of Separation*" utilises the average path length metric. A graph is considered small world if its average clustering coefficient is significantly higher than a random graph constructed from the same set of vertices.
  The average clustering coefficient can be expressed using Equation (6) as follows:

$$\overline{C} = \frac{1}{n}\sum_{i=1}^{n} C_i \quad (6)$$

  where $C_i = \frac{\lambda_G(v)}{\tau_G(v)}$, $\lambda_G(v)$ is the number of subgraphs of G having 3 edges and 3 vertices including the vertex v. $\tau_G(v)$ is the number of subgraphs of G having 2 edges and 3 vertices including v such that v is incident on both the edges.

- ***Average Degree:*** The number of vertices adjacent to a vertex *v* is called as the degree of *v* or *deg*(*v*). Based on this measure one can get maximum degree, minimum degree or average degree. The average degree of a graph is a network level measure and it is calculated from the value of degree or all the nodes in the network. For a graph *G* with *V* vertices and *E* edges the average degree of *G* can be expressed using Equation (7).

$$D_A(G) = \frac{2 \times |E|}{|V|} \quad (7)$$

- ***Density:*** The Density of a graph quantifies the number of connections between various actors in the network. The graph is considered dense if the number of edges in the graph approaches the maximal number of edges which one can have in that graph and sparse otherwise. For an undirected graph G with V vertices and E edges, the density of G can be expressed using Equation (8) as follows:

$$D_G = \frac{2|E|}{|V|(|V|-1)} \quad (8)$$

3. SNA METRICS FOR ACADEMIC SOCIAL NETWORKS

Universities, research laboratories and other institutions of higher learning are known for providing solutions to various problems confronting the society [13]. Research has been providing answers to many such problems. Modern day research is faced with both extraordinary opportunities and challenges. A fast paced modern society turns to academics for immediate answers to an array of practical problems created by its own increasing needs and desires. Society is willing to invest in research as the basis of a knowledge economy as long as research proves to be responsive to its needs, is productive and effective. Knowledge sharing and interactions are at the heart of research practice and collaboration. Collaboration is defined as "working jointly with others or together especially in an intellectual endeavor"4. Research interactions and collaborations include working on a research project jointly and publishing the results of the research undertaken. These collaborations help promote and proliferate research [13, 14], therefore, they should be encouraged, supported and monitored. Studies [13, 15, 16, 17] indicate that there is a direct relationship between scientific collaboration and creation of new knowledge. Co-authorship is one of these collaborations. In order to understand and analyse the social networks formed by any form of academic collaborations, they need to be viewed from a network perspective. SNA metrics can be used to seek answer, inter alia, to the following questions:

- Who are the hubs/leaders?
- Who has more connections?
- How strong are the collaboration ties?
- How collaborative the authors are?
- How connected the network is?

Value of various SNA metrics discussed above can be used to answer these and many other questions that help us understand the structure of network, flow of information in the network, strategic positions occupied by the authors in the network, important individuals, prestige of important authors in the network, etc. In the following we discuss the applicability of the above listed SNA metrics in academic social networks.

- ***Degree Centrality:*** In case of academic social networks degree centrality means the centrality of an actor in terms of frequency of the considered activity. The more the activity the better the degree centrality. For example in co-authorship networks it is a measure of how often an author collaborates with other authors in the network. However it does not takes into account the quality of collaborators. Having connections with such nodes (authors) may not necessarily rate you

[3] http://en.wikipedia.org/wiki/Clustering_coefficient

4 Merriam-Webster's Collegiate Dictionary (1999). Tenth Edition. Springfield, MA: Merriam-Webster, Incorporated.

higher in terms of your prestige in the academic social network.

- ***Eigenvector Centrality:*** Since the value of eigenvector centrality of a node depends upon the quality of connections nodes with higher eigenvector centrality lie at the centre of flow of ideas and information in the network. In co-authorship networks it is a representation of an author's ability to receive new research ideas that spread across the network [18].
- ***Betweenness Centrality:*** Nodes with high betweenness centrality occupy strategic positions in the network. Removal of such nodes result in breakdown of the information flow and the nature of connectivity in the network may change altogether. Analytical results obtained [19] testify that in academic social networks actors (scientists in this case) having high value of betweenness centrality in a network play a positive role in advancing scientific cooperation.
- ***Closeness Centrality:*** It is measure of the proximity of an academic with others in the network. Here the diversity is link is important than the quality of links. If a node is connected with majority of other nodes in the network, either directly or indirectly, the closeness centrality of that node will be more than of those have connections with other high profile nodes.
- ***Clustering Co-efficient:*** Measure of connectivity in the network. In academic social networks clustering coefficient means is a way of predicting future collaborations between any two academics that are indirectly collaborating with each other i.e. collaborating through a mutual collaborator [20].
- ***Average Degree:*** Each of the nodes (academics) may have different potential of connectivity with other nodes in the network. It is a network metric and in academic social networks it is considered as a measure of how collaborative the academics are.
- ***Density:*** The density refers to the potential of connectivity in the network. In academic social networks it represents the degree of collaboration that takes place in the network [18].

4. CONCLUSIONS

Mathematics has been called as mother of all the sciences and SNA is no exception. Fundamentals of mathematics play an important role in the formulation of SNA. A social network can have any shape and form but the basic considerations remain almost same. In this paper we explained various social network analysis metrics and their dependence on mathematical concepts. After elaborating these metrics we discussed their use and relevance in analysis of academic social networks.

5. REFERENCES

[1] Jin, Y., Matsuo, Y. and Ishizuka, M. (2006) Extracting a social network among entities by web mining. In Proceedings of ISWC'06 Workshop on Web Content Mining with Human Language Technologies, Athens, GA, USA.

[2] Lee, J. (2007) A study of collaborative product commerce by co-citation analysis and social network analysis. In Proceedings of 2007 IEEE International Conference on Industrial Engineering and Engineering Management, Singapore, pp. 209-213.

[3] Mislove, A., Marcon, M., Gummadi, K. P. Drushel, P., and Bhattacharjee, B. (2007) Measurement and analysis of online social networks. In Proceedings of the 5th ACM/USENIX Internet Measurement Conference, San Diego, CA, USA, pp. 29-42.

[4] Arif, T., Ali, R. and Asger, M. (2012) Scientific co-authorship social networks: A case study of computer science scenario in India. International Journal of Computer Applications, 52(12), pp. 38-45.

[5] Wasserman, S. and Faust, K. (1994) Social network analysis: methods and applications, structural analysis in social sciences. Cambridge University Press, New York City, New York, U.S.A.

[6] Chakrabarti, S. (2003) Mining the Web: Discovering knowledge from hypertext data."Morgan Kaufmann Publishers, USA.

[7] Newman, MEJ. (2001) Co-authorship networks and patterns of scientific collaboration. In Proceedings of the National Academy of Sciences, 101(1), pp. 5200-5205

[8] Davis, G.F. and Greve, H.R. (1997) Corporate elite networks and governance changes in the 1980s. The American Journal of Sociology, 103(1), pp. 1-37.

[9] Stuart, T.E. (1998) Network positions and propensities to collaborate: An investigation of strategic alliance formation in a high-technology industry. Administrative Science Quarterly, 43(3), pp. 668-98.

[10] Chelmis, C. and Prasanna, V.K. (2011). Social networking analysis: A state of the art and the effect of semantics. In Proceedings of 3rd IEEE Conference on Social Computing (SocialCom), Boston, MA, pp. 531-536.

[11] Watts, D.J. and Strogatz, S.H. (1998). Collective dynamics of 'small-world' networks. Nature, 393(6684), pp. 440–442.

[12] Milgram, S. (1967). The Small World Problem. Psychology Today, 2(1), pp. 60-67.

[13] Ahn, J., Oh, D. and Lee, J. (2014) The scientific impact and partner selection in collaborative research at Korean universities. Scientometrics, 100 (1), pp. 173-188.

[14] Patel, N. (1973) Collaboration in the professional growth of American sociology. Social Science Information, 6, pp. 77-92.

[15] Wray, K. B. (2002) The epistemic significance of collaborative research. Philosophy of Science, 69, pp. 150-168.

[16] Beaver, D. D. (2004) Does collaborative research have greater epistemic authority? Scientometrics, 60(3), pp. 399-408.

[17] Guerrero-Bote, V. P., Olmeda-Gomez, C., and Moya-Anegon, F. (2013) Quantifying the benefits of international scientific collaboration. Journal of the American Society for Information Science and Technology, 64(2), pp. 392-404.

[18] Zervas, P., Tsitmidelli, A., Sampson, D.G., Chen, N.-S., and Kinshuk. (2014). Studying research collaboration patterns via coauthorship analysis in the field of TeL: The case of Educational Technology & Society journal. Educational Technology & Society, 17 (4), pp. 1–16.

[19] Newman, M. E. (2001). The structure of scientific collaboration networks. Proceedings of the National Academy of Sciences, 98(2), pp. 404–409.

[20] Farashbandi, F. Z., Geraei, E. and Siamaki, S. (2014). Study of co-authorship network of papers in the Journal of Research in Medical Sciences using social network analysis. Journal of Research in Medical Sciences, 19(1), pp. 41–46.