

## APPLICATION OF SOCIAL NETWORK ANALYSIS [SNA] METRICS FOR SYSTEMATIC LITERATURE REVIEW [SLR]

Ramakrishnan Venkatesakumar\* and Murugaiyan Pachayappan\*\*

---

**Abstract:** Social Network Analysis (SNA) is the pattern identification study of people (actors) connected in a social network. SNA seeks to uncover two different types of patterns; the first one to collect the actors who are closely associated to one another and the other one is to identify the sets of actors who are associated in the total social network in similar behavior. SNA adapts certain metrics from *Graph Theory* to identify the patterns and influential information in the huge social network. Various metrics from *Graph Theory* are used in many domains like Psychology, Sociology, and Social Science etc. The objective of the current research work is to adapt the SNA metrics and apply them in literature review context and draw meaningful inferences from the literatures reviewed.

Literature review is a pioneer activity for any academic research. The traditional way of reviews narrating the literature in a comprehensive manner and describe or summarize the findings from the research works, but often failed to express insights and hypothesis/relationships. With the massive growth in internet resources, availability of research works online for any research topic is plenty; this creates more difficulty in spotting a research gap for future research activities. To facilitate research works and create quantitative parameters from the existing research works, the paper elaborates a SNA based metrics for doing systematic literature review.

This paper discusses a procedure for literature data collection, constructing the literature data into structured data and using this input for SNA based analysis, and finally, introduces SNA metrics like degree centrality, closeness centrality, betweenness centrality, clustering coefficient, eccentricity centrality, eigen vector centrality, average shortest path, connected component, graph density and modularity and their correspondence with review of literatures. These metrics are visualized by Gephi 0.9, a graph visualization and manipulation tool.

**Keyword:** Systematic Literature Review, Gephi, Social Network Analysis (SNA), Graph Theory, Centrality Measures.

### INTRODUCTION

The necessity of reviewing existing literature is more important for any research work. The researcher needs to know the emerging research interests and should regularly update their research knowledge based on their research topic (Barry, 1997). The preliminary stage of any research is to perform a literature review. A review used to elaborate, describe, summarize, evaluate and clarify the existing research works. Review of literature are often found useful to define the theoretical basis of the research topics and narrow down the scope of research (Boote and Beile 2005). Creswell (1994) classified the literature review into three criteria, "to present results of similar studies, to relate the present study to the ongoing

---

\* Associate Professor, Department of Management Studies, School of Management, Pondicherry University. Email: [venkatesh.dms@pondiuni.edu.in](mailto:venkatesh.dms@pondiuni.edu.in)

\*\* Assistant Professor, Department of International Business, School of Management, Pondicherry University

dialogue in the literature, and to provide a framework for comparing the results of a study with other studies”. Further Creswell (2002) recommends five steps for reviewing the literature; identifying terms to typically use in your literature search, locating literature, reading and checking the relevance of the literature, organizing the literature you have selected, and writing a literature review”. Reviewing the literature is categorized in eight different aspects (Petticrew and Roberts 2006) viz. “narrative, conceptual, systematic or evidence synthesis, rapid, realistic, critical, expert and state-of-the-art”.

Systematic Literature Review (SLR) is an iterative procedure of defining proper search keywords, searching in the literature, identifying the relevant literature and analysis the collected literature and structuring the literature data for further analysis (Rowley and Slack 2004). The systematic literature review is considered as evidence enriched method and making more sense for huge bodies of information; it answers to the questions about what works and what does not work; this method maps out the areas of certainty, uncertainty and defines where less or no relevant research work has been done or sometime, to identify the spurious faith in the literature.

In the beginning stage of any research work, the researchers often use to collect huge amount of literature data; many online databases like Google Scholar, Scopus, Web of Science, EBSCO etc. are accessible to the researchers and provide huge amount of previous studies in the research domain; a manual review of them often fail to provide finer points in the research works and consolidated information.

To manage this complexity and to identify the influential information [variables] from literature, various methods are used by the researchers. In this paper, a Social Network Analysis (SNA) metrics is introduced to examine the huge literature network. *The literature network is visualization of a network diagram of relationships studied in the existing literature in terms of the variables used by the researchers.*

Importance of the literature review can be classified in different aspects. It helps the researchers to discern the old theories with proposed ones or update the existing ones or to justify lack of evidence in testing relationships. It is also used to visualize the reason why different research studies addressing the same set of questions with different conclusions. *In this paper, a framework is proposed to relate various SNA metrics to literature review context.*

Open source software such as Gephi 0.9, could be useful to measure various metrics for the literature network data and social network data. For explanation purpose of the proposed methodology, the paper adapted two sample cases; one is Facebook friendship network context and the other from a literature variable network.

### PURPOSE OF THE PROPOSED RESEARCH WORK

A decade back the researches are in-between certain boundaries. The accessibility levels of the research outputs are really not easy; presence of online databases were skewed. Only certain groups of research communities know the existing status and update of the research problems (topics) like government agencies, scientific collaboration networks and higher education institutions. High degree of internet penetration and developments in high-tech communication and information technologies, availability of the electronic repositories, brought flexibility in accessing many research works across continents.

If a researcher needs literature data of specific topic, the keyword of the specific research topic is selected (Rowley and Slack 2004) and searched in the online databases like Google scholar, Scopus etc. The result obtained by the online database is vast (more than thousands for certain research topics) and to read and analyze each and every document is not humanly possible. So select the limited number of research works which are centrally positioned in your research topic rather than collecting enormous number of research (Boote and Beile 2005).

In the next step researcher will filter the collected literature data. For example filter the literature data based on certain periods of time, or filter the data based on top-tier journals, or filter based on any single journal who is pioneer in that particular research field. By doing this, the specific research topic will come under a countable literature boundaries. In the next step the researchers will review each and every topic, and then gather the information in a sample literature data table format as shown in Table-1. It consists of certain details and these details are useful to describe the flow of literature of selected topic for selected period of time.

For instance, let us consider a review of a researcher, summarizes a collection of 250 research articles; each and every author examined the same problem (topic) with their own theories, variables, methodology etc. As a researcher, we need to know which theories, variables, methodology are majorly contributed or less contributed to the selected research topic by him/her. This question can easily answered by summing each column of Table 1, but this will not give an in-depth insight of the specific research problem. The literature review should clearly define the similarities (relationship) present in-between the articles. To identify the influential similarity among the 250 articles, tabling the data will not provide sufficient summary.

For example, a researcher who is reviewing 250 articles, considers reviewing 60<sup>th</sup> article; variables used in the 60<sup>th</sup> articles are attitude towards advertisement, brand loyalty and purchase Intention, which is a partial set of variables used by the researcher of 4<sup>th</sup> article (some time before reviewed). Thus, the similarities between these two articles based on the variables are, attitude toward advertisement and purchase intention. These two variables are commonly used in both the articles (authors). The same situation might be prevailing in all the 250 cases. This type



of similarity identification is not possible to do manually for all the 250 articles. Hence, by identifying similarities among the existing research works, a researcher may define a set of highly influencing variables, another set which is less influential and variables acting as intermediate [moderators/mediators]. This research work proposes a conceptual framework to review huge amount of literatures and detecting various relationships frequently revisited by the researchers. To analyze the literature articles and to identify the variable based influential (similar) information of selected topics, the paper introduces a Social Network Analysis (SNA) metrics and their correspondence with literature review context.

### APPLICABILITY OF SNA METRICS FOR LITERATURE REVIEW

In this section the applicability of the metrics for literature review is elaborated via case example. For the discussion purpose, we have taken Facebook friendship network and literature variable network. Both the networks are directed graph (digraph) and the direction (connection) of the networks are given Table-2 for friendship network and Tabl-3 for literature variable network.

The network represented by the list in table-2 is a kind of network in social network analysis (SNA), which is also called as ego network. Ego means a node/person, which is having a direct connection between other nodes/person. The node/person to which ego is directly connected is called as alters in SNA. In this case, consider Rias as an ego and alters are Panneer and Umma. Now let we take this logic to literature variable network analysis as given in Table 3.

**TABLE 2: FRIENDSHIP NETWORK CONNECTIONS**

<i>S. NO</i>	<i>Node Name</i>	<i>Connecting node</i>
1	PANNEER	VENKAT
2	VENKAT	RIAS
3	RIAS	PANNEER UMMA
4	KASI	PANNEER
5	CHITRA	VENKAT
6	UMMA	CHITRA
7	LAVANYA	PANNEER

The Table 3 represents the relationships between the variables in research papers. The variables are classified into independent and dependent variables. So as per the SNA concept, the ego is a node which is directly connected to other nodes (alters); in the literature variable network, the independent variables are considered as ego and dependent variables are considered as alters.

In this case, sample of three research papers are selected to demonstrate the relationships studied in the papers; and the papers are based on ‘Celebrity

Endorsement’ in marketing domain. Now consider, in one of research work, the independent variable ‘Celebrity consumer congruence’ [*ego*] related with ‘Attitude towards advertisement’ ‘Attitude towards brand’ and ‘Purchase intention’ [*alters*].

**TABLE 3: LITERATURE VARIABLE DATA**

S. No	Research Paper	Connection/Relationships	
		Independent Variables	Dependent Variables
1	Paper 1	Celebrity consumer congruence	Attitude towards advertisement; Attitude towards brand; Purchase intention
2		Celebrity product congruence	Attitude towards advertisement; Attitude towards brand; Purchase intention
3		Attitude towards advertisement	Attitude towards brand
4		Attitude towards brand	Purchase intention
5	Paper 2	Attitude towards advertisement	Purchase intention
6		Attitude towards advertisement	Brand loyalty
7	Paper 3	Attitude towards celebrity	Brand loyalty
8		Celebrity product congruence	Brand loyalty
9		Attitude towards brand	Brand loyalty

## A NETWORK REPRESENTATION OF FRIENDSHIP NETWORK CONNECTIONS VIS-À-VIS

### Literature Variable Data

Let’s move on from this initial table representation to visual representation. Consider the two digraphs given in Figure 3 and Figure 4, Figure 3 is a friendship network where the nodes represents friends and edges are the connections between them. These network representation may be obtained by using ‘Network Visualization Software’; here the authors used Gephi 0.9, open source network visualization software. Since we are dealing with digraph the edges plays a vital role because according to SNA the edge defines both sending and receiving.

The pin point direction of the arrowhead of the edges represents that the node/person receiving the message from other nodes/person and fairly opposite if the node sends the message. For instance select a node PANNEER from Figure 1, PANNEER is having three pinpoint arrowheads, which means PANNEER is receiving messages from three of his friends KASI, LAVANYA and RIAS. As well as PANNEER sends the message to VENKAT.

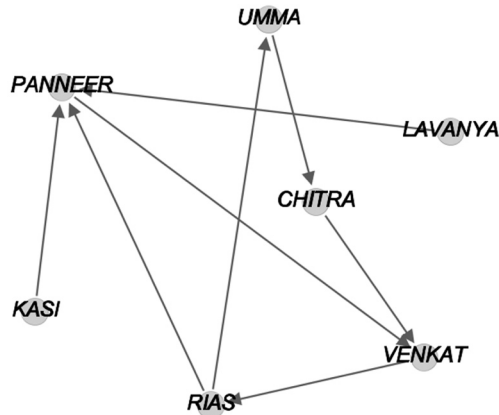


Figure 1: Friendship Network

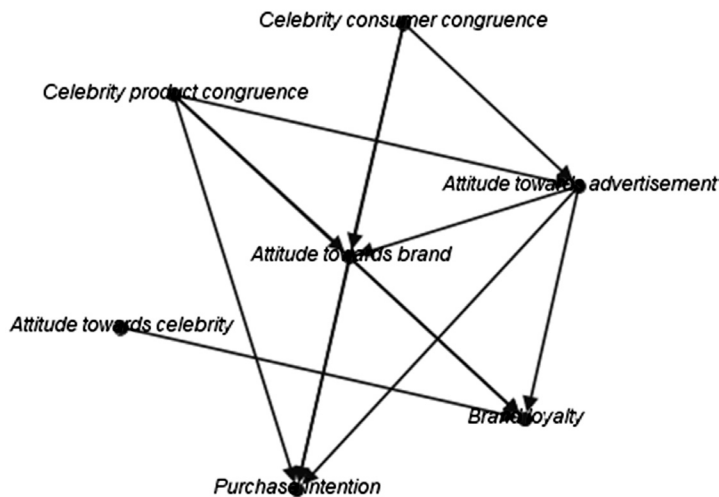


Figure 2: Literature Variables Network

Map the same SNA concept for literature variable network as shown in the Figure 2; in this network the nodes are the variables (independent or dependent) and edges are the connections (directions) between any pair of variables. Since we are dealing with independent and dependent variables in our literature review context, independent variables are considered as equivalent to a 'sender' and dependent variables are considered as 'receiver' in a friendship network. In a graph, the edges may be given weight based on several criteria; however, we do not consider those cases at this moment.

Based on Table 2 & 3 the digraphs are constructed and it is shown in Figure 1 & 2. Figure 2 has 7 nodes and 13 edges in the network; 7 nodes are nothing but

the total number variables which are used in the research papers listed in Table 3. The 13 edges define the relationships studied in the research papers. Based on the arrowheads, we can easily identify whether a variable is an independent or dependent one.

For example, consider a one node [variable] in the network [Figure 2], *Attitude towards Advertisement*; total number of connections (edges) connecting '*Attitude towards Advertisement*' with other nodes in the network is five. Out of five connections, two arrowheads is pointing towards the variable '*Attitude towards Advertisement*'; it represents '*Attitude towards Advertisement*' is studied as a dependent variable (receiving) with two other independent variables [nodes] namely, '*Celebrity Consumer Congruence*' and '*Celebrity Product Congruence*'. At the same time '*Attitude towards Advertisement*' is acting as an independent variable in relationship with other three variables namely, '*Attitude towards brand*' '*Brand loyalty*' and '*Purchase intention*'.

By computing various Social Network Analysis metrics, a researcher can quickly summarize the information available among the literature/articles collected and importance of each variable and critical relationships which are often studied or rarely studied.

### **Key Network Characteristics & Metrics**

Before proceeding with the details of the metrics, which are useful to draw insights from SNA, let us review few basic terms in graph theory. A graph is a connection between two or more entities. The entities might be anything viz. human being, machines, animals, buildings, characters in the movie, or keywords. According to graph theory, these entities are considered as nodes (vertices/actors) and the relationship between these entities are named as edges (ties/arcs) (Wasserman & Faust, 1994). The edge is a line, and the arrowheads in the lines indicate the direction of the edges. The graph with directed line is referred to as a direct graph (di-graph). The direction of an arrowhead has a specific meaning; it connects an origin node with a destination node; it is strongly connected if there is any direct relationship/path from any node to any other node.

The direct graph may be a friendship network, family network or a transportation network. The undirected graph is called as a simple graph; in that lines are connected to the nodes in the network without any direction from any node to any other nodes. This type of graph represents only the strength of the node which has more connections. The mixed graph represents both direction and un-direction between the nodes in the network.

The intuitive sense of Social Network Analysis is to know the central position of a network of person/node; and explore more visible actors in the network (Linton Freeman 1979). It means a centrality measures give a rough indication of social power of a person/node based on how well they are connected in the network



(Scott 2000). SNA adapted metrics from graph theory and the metrics are listed and detailed in the table below.

**TABLE 4: SNA METRICS**

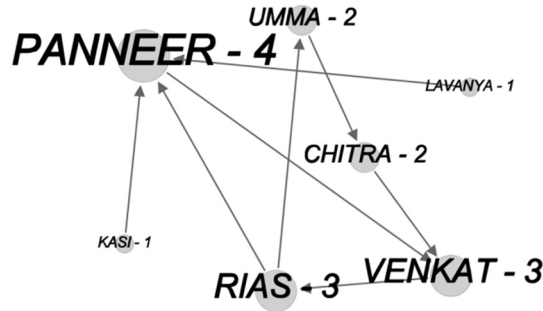
<i>S. No</i>	<i>Metrics Name</i>	<i>Description</i>
1	Degree	Degree centrality is simply the number of connections or number of immediate contacts a node has in a network
2	In-Degree	In-degree (receiver) centrality is the number of edges received by a node from others (or) in-degree measures provide information about the number of people who respond to a message from a certain participant (Laat et. al., 2007)
3	Out-Degree	Out-degree (sender) centrality is the number of edges given (sent) by that node to others (or) Out-degree gives an indication of the number of messages a person has sent to other individual members of network (Laat et. al., 2007)
4	Betweenness	The extent to which a node lies between other nodes in the network. This measure takes into account the connectivity of the node's neighbors, giving a higher value for nodes which bridge clusters. The measure reflects the number of people who a person is connecting indirectly through their direct links (Curran & Curran, 2014).
5	Closeness	Closeness refers to the degree with which an individual is nearer to all others in a network either directly or indirectly (Laat et. al., 2007) (or) "The degree an individual is near all other individuals in a network (directly or indirectly). It reflects the ability to access information through the "grapevine" of network members. Thus, closeness is the inverse of the sum of the shortest distances between each individual and every other person in the network. The shortest path may also be known as the "geodesic distance (Curran & Curran, 2014).
6	Eccentricity	The maximum of the shortest paths to other nodes in the network; indicates how far given node is from the furthest one in the network (Matos et 2009)
7	Graph Density	The degree a respondent's ties know one another/proportion of ties among an individual's nominees. Network or global-level density is the proportion of ties in a network relative to the total number possible (sparse vs. dense networks). Curran & Curran (2014).
8	Modularity	Modularity measures the structure and strength of division of graphs modules (groups/clusters/communities) (or) Modularity measures the density of links inside communities as compared to links between communities.
9	Connected Components	Determine the number of connected components in the network.
10	Clustering Coefficient	A measure of the likelihood that two associates of a node are associates. A higher clustering coefficient indicates a greater 'cliquishness'. Curran & Curran (2014).
11	Eigenvector Centrality	A measure of the importance of a node in a network. It assigns relative scores to all nodes in the network based on the principle that connections to nodes having a high score contribute more to the score of the node in question (Curran & Curran 2014).
12	Average Path Length	The average of distance between all path node (Matos et. al., 2009)

### Comparison of Metrics for a 'Facebook Friendship Network' and 'Literature Variable Network'

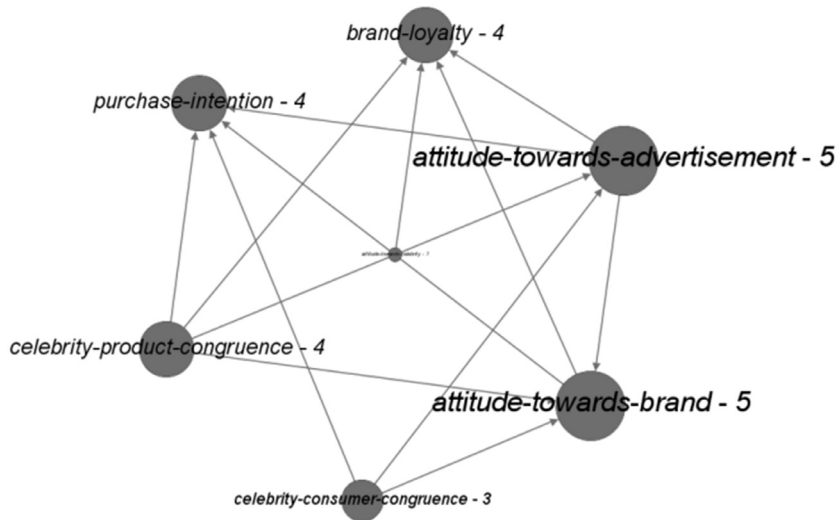
In this section, we illustrate how few metrics for SNA in a friendship network context might be useful to draw insights in a literature variable network context.

#### Degree

*Degree* is the number of connections or number of immediate contacts a node has in a network; in a friendship network context, for example, the node 'PANNEER' is having more connections than other nodes/persons, symbolizes the number of 'friends' connected with a node/person. Higher the degree coefficient signifies 'highly connected' node/person in the network.



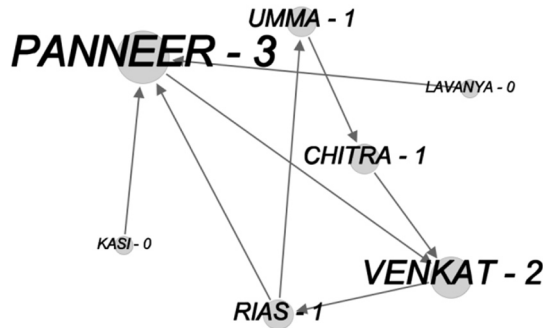
In similar analogy, we define the *degree* as number of times the variables studied with other variables in the literature collection. It is intuitively clear that higher the degree coefficient for a node in the 'literature variable network' signifies the



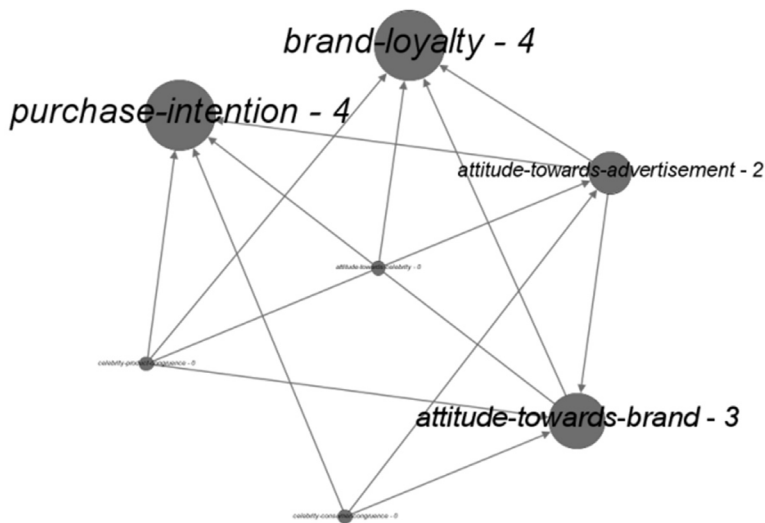
variable (s) frequently studied in conjunction with other variables [more significant/ more popular] in the pool of literature collection by the researcher.

**In-Degree**

In-degree (receiver) centrality is the number of edges received by a node from others; for example, the node ‘PANNEER’ is having more arrowheads pointing towards it than [inward connections] than other nodes/persons. It signifies the ‘logical’ close of communication in a SNA.

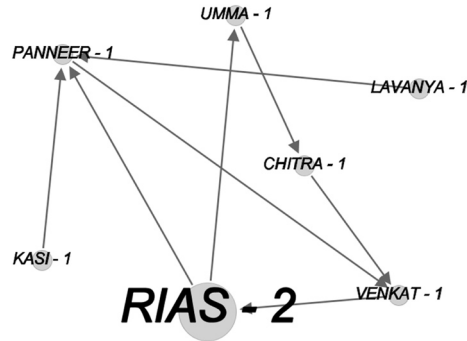


In similar analogy, we define the in-degree as number of times a variable studied as dependent variable in the literature collection. If we extend the similar argument for a literature variable network, ‘brand loyalty’ and ‘purchase intention’ studied as dependent variable by many researches than other variables and the variables could be logical conclusion for future studies. Higher the in-degree coefficient signifies more number of times the variable used as dependent variable.

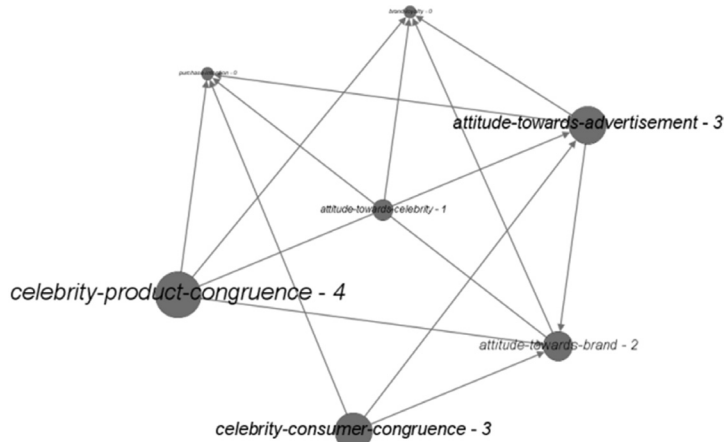


### Out-Degree

Out-degree (*sender*) centrality is the number of edges starting from a node to other nodes in the network; for example, the node ‘RIAS’ is having more arrowheads pointing towards others than [outward connections] than other nodes/persons. It signifies the ‘logical’ starting stage of communication in a SNA.



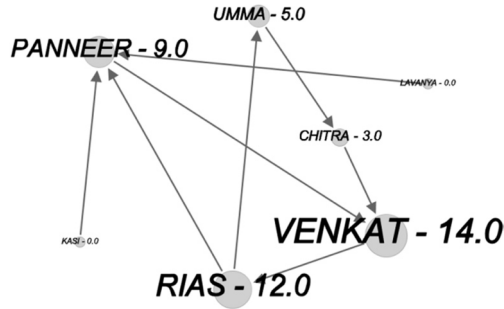
In similar analogy, we define the *out-degree* as number of times a variables studied as independent variable in the literature collection. If we extend the similar argument for a literature variable network, ‘Celebrity-product congruence’ studied as independent variable by many researches than other variables and the variable could be considered as a logical independent variable in the proposed relationships.



### Betweenness

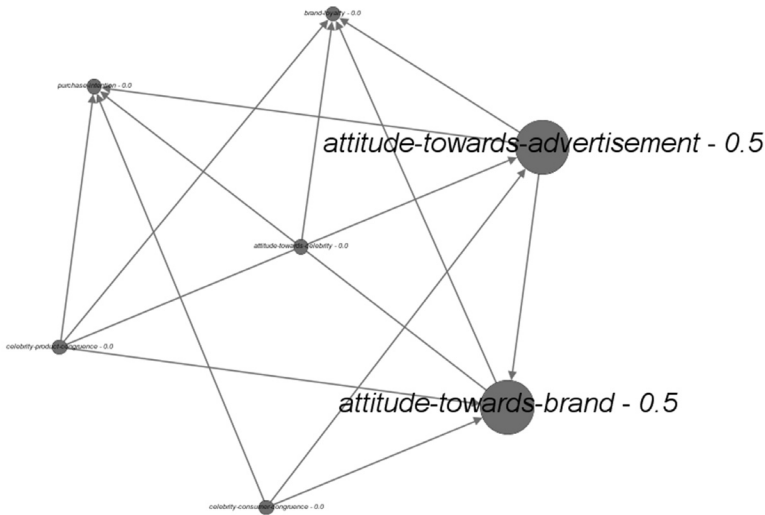
*Betweenness centrality* measures the extent to which a node lies between other nodes in the network; this measure reflects the number of people who a person is connecting indirectly through their direct links. Higher the ‘betweenness’

coefficient indicates the higher importance attached to the node/person in passing the communication.



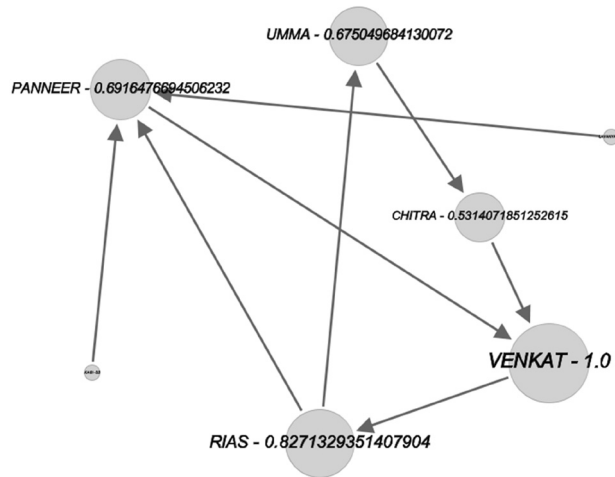
In similar analogy, we define betweenness centrality as the degree *a variables studied both independent and as well dependent variable in the literature collection*. Higher the coefficient signifies the variable could be used as a logical start as well end variable in the relationships to be studied.

If we extent the similar argument for a literature variable network, ‘attitude-towards advertisement’ and attitude towards brand studied as both independent and dependent variable by many researches than other variables.

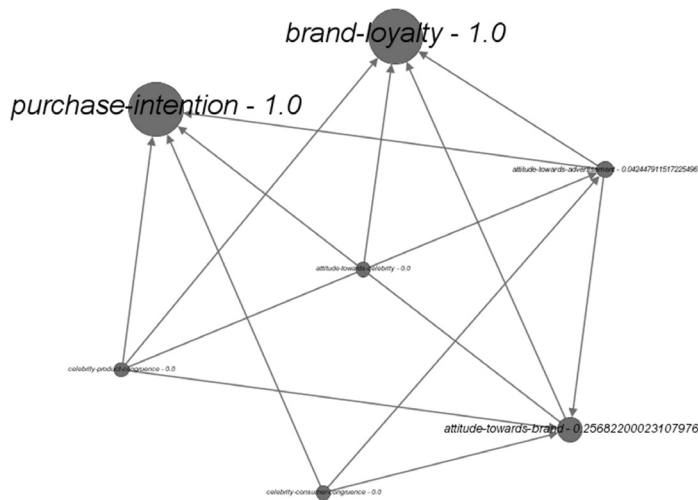


**Eigenvector**

Eigen vector measure of the importance of a node in a network. It assigns relative scores to all nodes in the network based on the principle that connections to nodes having a high score contribute more to the score of the node.



In similar line, in the literature variable network, the eigen vector scores signifies the importance of the variables with due consideration to its relationships with other variables in the network. Hence, higher the score reflects higher the importance of the variable among the variables considered.



## CONCLUSION

This paper aims to develop a conceptual discussion to bring out how the social network analysis analogies could be used to draw inferences for huge literature variable network and consolidate the information available in the literature and quantitatively conclude the importance of variables in the studies. Few important

and simple metrics are introduced and their definition and its fitment with literature variable network context is discussed.

This methodology may quickly eliminate the tediousness in understanding the numerous relationships among variables in a huge collection of earlier studies. Also, other metrics such as clustering coefficient, eccentricity centrality or average shortest path have the ability to summarize the information from SNA to SLR. However, to develop input for such analysis, the researcher has to prepare a detailed working file containing the information of variables studied by earlier studies.

### *References*

- Boote, D.N., & Beile, P. (2005). Scholars before researchers: On the centrality of the dissertation literature review in research preparation. *Educational researcher*, 34(6), 3-15.
- Barry, C.A. (1997). Information skills for an electronic world: training doctoral research students. *Journal of information science*, 23(3), 225-238.
- Creswell, J.W. (1994). *Research design: Qualitative and quantitative approaches*. Thousand Oaks, CA: SAGE.
- Creswell, J.W. (2002). *Educational research: Planning, conducting, and evaluating quantitative and qualitative research*. Upper Saddle River, NJ: Merrill Prentice Hall.
- Rowley, J., & Slack, F. (2004). Conducting a literature review. *Management Research News*, 27(6), 31-39.
- Scott, J.: *Social Network Analysis—A Handbook*, 2nd edn. SAGE Publications, Beverly Hills (2000).
- Freeman, L.C., Roeder, D., & Mulholland, R.R. (1979). Centrality in social networks: II. Experimental results. *Social networks*, 2(2), 119-141.
- Analysis of Mathematical attack on RSA algorithm using factoring a prime number by Division algorithm method” AICTE sponsored National Conference Recent Research Trends in Network Security NCRTNS-2012 organized by Srinivas Institute of Technology, Valachil, Farangipet Post, Mangalore, 19th and 20th April 2012.
- ”Wireless sensor network Attacks and Key Management Authentication” Presented in National conference on Networking and computing Technologies Organized by Dept of MCA, CSE & IT, Vel Tech MultiTech Dr. Rangarajan Dr. Sakunthala Engg college, Chennai on 10.08.2012.
- Camarinha-Matos, L.M., Paraskakis, I., & Afsarmanesh, H. (Eds.). (2009). *Leveraging Knowledge for Innovation in Collaborative Networks: 10th IFIP WG 5.5 Working Conference on Virtual Enterprises, PRO-VE 2009*, Thessaloniki, Greece, October 7-9, 2009, Proceedings (Vol. 307). Springer.
- Blondel, V.D., Guillaume, J.L., Lambiotte, R., & Lefebvre, E. (2008). Fast unfolding of communities in large networks. *Journal of statistical mechanics: theory and experiment*, 2008(10), P10008.

