

STATE OF ART: UTILIZING SOCIAL NETWORK ANALYSIS IN DIVERSE FIELDS

ESTADO DEL ARTE: UTILIZACIÓN DEL ANÁLISIS DE REDES SOCIALES EN DIVERSOS CAMPOS

¹Ibo Luis Cerra Escobar, ²Jhon Edison Villarreal Padilla



Universidad Piloto de Colombia

¹Ibo-cerra@unipiloto.edu.co, orcid.org/0000-0003-2158-6621

²jhon-villareal@unipiloto.edu.co, orcid.org/0000-0003-3135-0866

Recibido: 24/08/2016 • Aprobado: 30/10/2016

RESUMEN

Este documento muestra una visión general de los diferentes conceptos y técnicas que se utilizan en el análisis de redes sociales; también muestra los nuevos usos del Análisis de Redes Sociales, SNA (por sus siglas en Inglés) y los desafíos que está enfrentando. En un primer momento se ilustra la forma de representar las redes sociales a través de gráficas o de matrices, resaltando la importancia de los nodos en el método gráfico y la mayor aplicación del segundo método, debido a las técnicas matemáticas allí utilizadas. Se profundiza en la importancia de los nodos, las diferentes técnicas para su análisis y se establece su influencia y se menciona la desventaja de su uso. De manera similar se presenta una comparación de las técnicas utilizadas para analizar la fortaleza de los lazos que unen estos nodos, definida como la combinación de una serie de características propias de dichas relaciones. Además, se hace un análisis de los nuevos usos del SNA y los retos que este enfrenta, orientados principalmente a la verificación de la veracidad de la información. Como conclusión se puede observar que el SNA puede aplicarse en muchos campos y que es una necesidad la investigación y desarrollo de nuevos sistemas, técnicas y herramientas de análisis en esta área. En este proceso es importante tener en cuenta el papel básico que juegan los ingenieros, matemáticos y sociólogos en la predicción de la conducta de las diferentes redes sociales y en el diseño y generación de estas nuevas técnicas.

Palabras clave: análisis de red, cercanía, autovector, grado, gráficos, interconexión social.

ABSTRACT

This paper shows an overview of the different concepts and techniques that are been used in social network analysis; also it shows new uses of SNA and different challenges that it is facing. In a first moment, it illustrates the way of representing the social networks through graphs of matrices, highlighting the importance of nodes in the graphical method and the wider application of the second method, due to the mathematical techniques used there. The importance of nodes is deepened, the different techniques for analysis; and they are defined, their influence is established and the disadvantage of their use is mentioned. Similarly, a comparison of the techniques used to analyze the strength of ties that unite these nodes, defined as the combination of a series of characteristics of these relationships, is presented. In addition, an analysis of the new uses of SNA and the challenges that it faces, oriented to the verification of the accuracy of the information, is made. As a conclusion, it can be observed that the Social Network Analysis, SNA, can be applied in many fields and that is a need the research and development of new systems, techniques and analytical tools in this áre. In this process it is important to take into account the basic role played by engineers, mathematicians and sociologists in predicting the behavior of different social networks and in the design and generation of these new techniques.

Keywords: closeness, degree, eigenvector, graphs, network analysis, social betweenness.

I. INTRODUCTION

In the last decade social networks such as Facebook, Twitter, Google and many others, have become part of daily life; it has a major impact on people, since 22% of the time spent online is dedicated to social networks (Mindjumpers 2012). But what is defined as a social network? Basically, it is a social structure between actors, individuals or organizations, which indicate the ways in which they get connected throughout a range of familiarities from simple acquaintances to people with close ties (Hanneman & Riddle 2005). Many of us based on the definition, could claim that social network analysis should be conducted by sociologists and experts in social areas and indeed, some authors define it as “an important branch of western sociology and a new way to study social science” (Kang, Duan & Wen 2011) Due to the magnitude of the amount of users and the complexity of social networks, interdisciplinary groups between sociologists, mathematicians and engineers, have been established to achieve the creation of methods of social network analysis what is known today as Social Network Analysis (SNA).

SNA is defined as: “the mapping and measuring of relationships and flows between people, groups, organizations, animals, and computers or other information / knowledge processing entities” (Jamali & Abolhassan 2006). Through the analysis of social networks, we observe in the market how social applications are created such as disease modeling (Eubank et al. 2004), marketing Carrasco et al. 2003), recommender systems (Palau et al. 2004) and propagation of influence in social network (Goyal, Bonchi & Lakshmanan 2008; Chen et al. 2009). Therefore, the analysis of social networks has become the subject of many research efforts.

The objective of this report is to analyze the use of SNA in different scenarios of people’s daily lives. As a result, this report is divided into four sections: Section II is a summary of the most common definitions and techniques for analyzing social networks, which are illustrated with examples; then in section III a series of articles are analyzed to show the different applications of SNA, in section IV the challenges of SNA are mentioned to

create a more complete and accurate analysis, and finally the conclusions of this analysis are stated.

II. ANALYSIS OF THE DEFINITIONS AND TECHNIQUES USED IN SNA

This section is a summary of the techniques used in SNA and how these are reflected in current social networks.

Social networks are commonly represented in two ways: graphically modeling social networks using graphs, as well as, being represented with matrices (Jamali & Abolhassan 2006). With graphs, each user is represented as a nodes and relationships are represented by edges. We must mention that the relationship between users can be direct. For example, on Facebook when you add a friend the two people share information, meaning that the communication is reciprocal and also indirect; Twitter allows a person to have many followers and he/she may follow different people. Not requiring this person to follow someone who is following him/her. Facebook also gives the opportunity to subscribe to these people. The second form of representation used to model social networks is the adjacent matrix. This method is applied mostly because of the use of mathematical techniques to analyze social networks (Lei & Huan 2010). Figure 1 shows an example of the two types of representation.

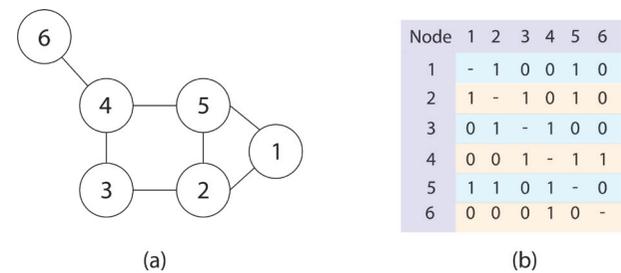


Fig. 1 Representation of Social Networks using a (a) Graph or (b) Matrix

Due to the huge number of users in social networks, these are classified as Large-Scale network, and researchers have identified a number of common properties such as size, density, degree, reachability, distance, diameter, and geodesic distance (Hanneman & Riddle 2005). Similarly, Ties and Influence have been agreed to

be the set of techniques for analyzing social networks based on the nodes and each is summarized below.

A. Importance of the Nodes

The summary of the techniques is shown in table 1.

TABLE 1

Comparison of the Techniques to Analyze the Importance of Nodes (Jamali & Abolhassan 2006; Lei & Huan 2010).

Power Aspect Name	Definition	Influence
Degree	The number of expired nodes.	Having more options to transmit information.
Closeness	Measures how close a node is to the others.	To perform a more direct approach with other nodes.
Betweenness	How many short past necessary for a node to connect to others.	Measure the shortest time it takes for a node to establish communications with a node of interest. (Broker)
Eigenvector	How important is a node depending on the importance of friends.	It is better having friends than money.

The disadvantage of these techniques in order to analysis social networks, is that there are different types of relationships between nodes. For example, If you look at the number of Facebook friends you have, you would see that this is, in most cases, a large number. In my case, I have 600 friends; yet that does not indicate the closeness I have with them. This fact takes away the support for an analysis supposing that we were based it only on these techniques.

Since we are talking about social networks, the term social comes from the Latin *soctus*, which means a group of humans. Hence, we should analyze the social networks as a group rather than an individual, and that is why other analysis techniques have been created and shall be explained later.

B. Strengths of Ties

According to Granovetter, strength of ties is defined as: “a combination of the amount of time, the emotional intensity, the intimacy, and the reciprocal services which characterize the tie” (Lei & Huan 2010).

Similar to the importance of nodes, there are three techniques to evaluate the strength of the connections which are summarized in table 2.

TABLE 2

Comparison of the Techniques to Analyze the Strength of Nodes (Lei & Huan 2010).

	Definition	Influence
Network Topology	It seeks to identify weak links between two nodes. A link means that when you remove that connection it results in the total disconnection of a group of nodes.	To prevent a node or group of people from disconnecting other connections can be suggested.
Attributes and Interactions	It seeks to analyze how users interact, and what aspects they have in common.	We see how adjacent nodes change weight as they interact.
Sequence of User Activities	It seeks to analyze the effects the activity of a node produces in the others nodes.	It identifies the network backbones. These are the ones who spread information easily.

One can see that these techniques are commonly applied within social networks. For example, the first technique is used to suggest friends, looking at Figure 2, on the right of the LinkedIn page, we find the option of suggested friends.

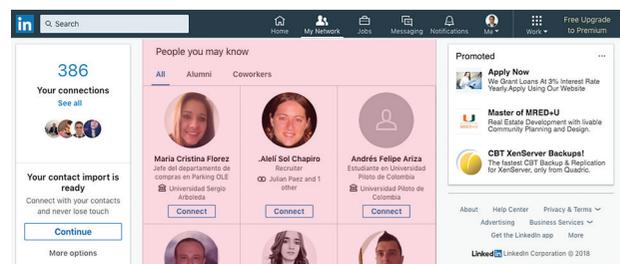


Fig. 2 Suggested friends on LinkedIn.

Note that the algorithms used in social networks like MySpace, Facebook, LinkedIn etc, are modified to achieve better results but are not revealed (Quora 2010).

The Attributes and Interactions technique can be found in Facebook when we observe the home section which shows current news from our friends. Usually, this is news from the people you interact, share photos or exchange messages, as shown in figure 3.

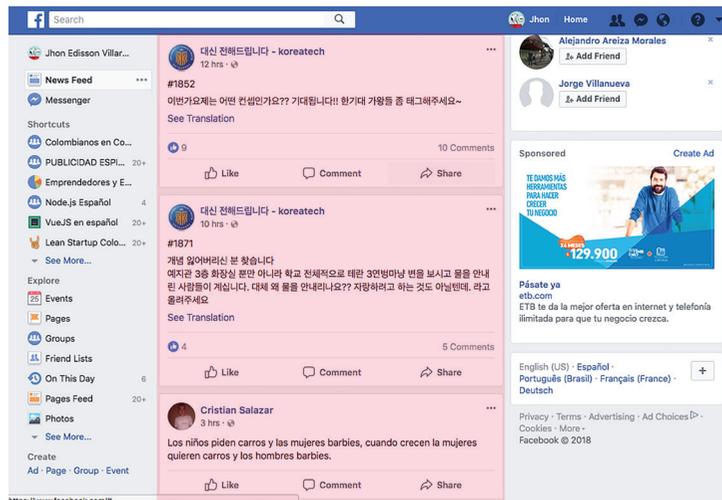


Fig. 3 Home on Facebook.

The last technique, Sequence of User Activities is used by the marketing people to analyze how nodes influence others. On Amazon, for example, when you search for a book, a series of suggested

books appears as well, which are based on books that were seen by other people who bought the same book that you're interested in. This is shown in Figure 4.

Frequently bought together

An Amazon 'Frequently bought together' section. It displays two book covers: 'Social Network Analysis' by John Scott (Paperback, \$41.80) and 'Linked: How Everything is Connected to Everything Else and What It Means for Business, Science, and...' by Albert-Laszlo Barabasi (Paperback, \$14.71). The total price for both is \$56.51. There are buttons for 'Add both to Cart' and 'Add both to List'. A note indicates that one item ships sooner than the other.

Customers who bought this item also bought

An Amazon 'Customers who bought this item also bought' section. It features a carousel of six book covers with their titles, authors, and ratings. From left to right: 'Analyzing Social Networks' by Stephen P Borgatti (4.5 stars, 13 reviews, \$45.95); 'Understanding Social Networks: Theories, Concepts, and Findings' by Charles Kadushin (4.5 stars, 17 reviews); 'Linked: How Everything is Connected to Everything Else and What It Means...' by Albert-Laszlo Barabasi (4.5 stars, 139 reviews); 'Social Network Analysis: Methods and Examples' by Song Yang (4.5 stars, 4 reviews); 'Social Network Analysis: Methods and Applications' by Stanley Wasserman (4.5 stars, 23 reviews); and 'A User's Guide to Network Analysis in R' by Douglas Luke (4.5 stars, 2 reviews).

Fig. 4 Suggested Books on Amazon.

This technique has brought a major challenge for engineers and mathematicians in order to define a model to analyze the influence of a node onto other. Table 3 shows two models that are generally used for this.

TABLE 3
Methods for Modeling the Influence of a Node (Lei & Huan 2010)

Influence Modeling	Definition
Linear Threshold Model	This consists that each node has the possibility of being influenced by another, depending on how many neighboring nodes perform a particular action. For example, If my friends get a tattoo, I should answer the question, how many friends of mine should get a tattoo before I should get one.
Independent cascade model.	It is based on the fact that each node is likely to be influenced by the action of another node. That is, if a friend shares a story on Face book, depending on the subject and my interests, I decide whether I share it or not.

In order to conclude this section, community detection and evaluation will be explained based on the events whenever groups of users form into communities and social networks are targeted to certain types of users. Some examples are: Flickr is aimed at amateur photographers, SoundCloud allows exchanges between music lovers, and LinkedIn for business contacts and others.

C. Community

We found that the nodes (users) in social networks tend to interact between different groups, which lead to the challenge of identifying these groups and make a further analysis in each group (Jamali & Abolhassan 2006). Table 4 shows a comparison of approaches to define groups and subgroups in the social networks.

TABLE 4
Comparing Different Approaches for Defining Substructures and Groups in Sociograms (Jamali & Abolhassan 2006)

Substructure	Description
Clique	Actors who have all possible ties among themselves.
N-Clique	Actors are connected to every member of the group at maximum distance of N.
N-Clans	N-Cliques that all paths among members occur, by the way of other members of N-Clique
K-Plex	Clique in which actors have ties to all but k of members of a group.
K-Core	Actors are connected to k of members of the group.
Component	Parts of sociograms that are connected within but disconnected with other components.
Cut Points	Nodes, which if removed, make the structure divide into unconnected systems.
Block	The divisions into which cut points divide a graph.
Lambda Set	Set of actors who, if disconnected, would most likely disrupt the flow among the entire actor.

The technique to find these groups within social networks is still a challenge for the SNA. Some of these proposed techniques are summarized in Table 5.

TABLE 5

Comparison of Techniques for Detecting Subgroups in Sociograms (Lei & Huan 2010).

Detection Techniques	Definition	Models and Techniques
Node Centric	It requires each node to meet certain criteria in order to belong to a group (Complete Mutuality, Reachability)	Brute Force. Pruning Strategy. Clique percolation method (CPM). Reachability.
Group Centric	Considers the connection within a group as a whole.	Density-based Groups Maximal dense subgraphs.
Network centric	This considers the entire network topology.	Vertex Similarity (automorphic equivalence, regular equivalence, Jaccard Similarity, Cosine Similarity). Latent Space (multidimensional scaling) Block model Approximation. Spectral clustering. Modularity Maximization Unified Process
Hierarchy Centric	Builds a hierarchical structure based on communities across the network.	Divisive Hierarchical Clustering (Newman Girvn Algorithm). Agglomerative Hierarchical Clustering.

As for the detection of communities and subgroups within the social networking world, is a very complex and difficult task to structure. Therefore, we can see how more websites are specialized for specific groups of people in the market of social networks, reducing its size so it can provide better analysis and a better experience for the user. Such examples are Sound Cloud for musicians and Flickr for photographers, among others.

With all this we see the complexity of social networks and how they are structured in a general way.

III. ANALYSIS OF NEW USES OF SNA

In the previous section, we saw the different SNA techniques and how they are applied to some examples. Therefore, it was observed that most of these techniques are used in common social networks like Facebook, My Space, Google and others, but the concept of social networks is broader than just social networks field and shows how SNA can be applied in other fields. Next, we will examine four different cases used.

The first one is entitled “Multi-factor analysis of Terrorist Activities based on social network” (Julei et al.2012), which presents a method for identify the nucleus of a terrorist organization, possible attacks, slipping and location of the organization. So as, to test this, the techniques proposed in section 2 are applied as: Degree Centrality, Closeness Centrality, Betweenness Centrality, and other new elements of analysis such as Cohesive Subgroup Analysis, Spatial Correlation Analysis and Invulnerability Analysis. All this was applied to an empirical test called “East Turkistan” Terrorist Activities since the founding of the People’s Republic of China and the results shown at the end are the important for some terrorists based on the importance of node techniques. In addition, the different groups have their influence and importance within the graph and the variation of such importance during different years, which can show a view of the behavior of these terrorist networks.

On the other hand, we have seen how social networks play an important role in the case of emergencies such as tornadoes, earthquakes, floods, as a means of dissemination of the events. Therefore, the second article we want to exhibit is “Research on Emergency Information Management Based on the Social Network Analysis: A Case Analysis of Panic Buying of Salt” (Kang, Duan & Wen 2011) where SNA techniques are applied to study the characteristics and effects of broadcasting emergencies, specifically in the case of “panic buying of salt” in China. Based on the collected data, they are showing the graphical representation of the network, which is directional and models the adjacent matrix in order to perform the analysis of the importance of the nodes already discussed in the previous section. After data

analysis, we concluded that when an emergency occurs, and the network is very unstable and constantly changing, the stronger nodes, in this case the government and the media, should be attentive to the information being disseminated and for it not to be altered or changed. Also, the information has to be shared within different cities to generate peace to others and for these not to spread false information.

Another outstanding example is presented in “The Application of SNA in IT-Based BRP” (Min et al. 2009), where the SNA is used for the reorganization of IT company to ensure better service to the end user. To accomplish this, first we perform the modeling of the structure of the company and then apply SNA techniques to see which entities play an important role in the company in order to propose a reorganization of the structure and new processes to reduce the time; it takes to dispatch an order by almost 160%.

Finally, the last article is: “Social Network Analysis Application in Bulletin Board Systems” (Kou & Zhang 2003), which is an analysis of a BBS (Bulletin Board Systems) an electronic message center. For the case of the example, SMTH was used as a bulletin board belonging to Tsinghua University, the largest in China (Baruah & Angelov 2012). With this information the graph is obtained and the degree of importance of the nodes are calculated; based on these nodes they can define who are the best BBS moderators.

In this section, we can conclude that the SNA is a powerful tool for engineers because of its vast scope of work, managing cases from criminal structure analysis to the reorganization of a company, since the SNA can determine the current network structure and its behavior and can improve this structure based on its results.

IV. CHALLENGES FACED BY SNA.

There are some challenges that the SNA must overcome. These are explained below:

- Social networks change in real time, means that whenever new nodes are made and connections

appear as seen in section II and III, the techniques used can be applied to the network instantaneously or at any given moment in time, and due to the large amount of data, it is difficult to maintain the information in real time. Therefore, techniques for scalability in this analysis are being proposed, as the proposed article “Dynamic Social Network Analysis using Latent Space Model and an Integrated Clustering Algorithm” (Zu & Zheng 2009).

- It is observed that most techniques focus on the elements of a network regardless of the information distributed, but as in some cases, false information can be spread just like in the case of the announcement of Justin Bieber having cancer and many of his fans shaving their heads to show support, turned out to be false (Crook 2012); or when famous bloggers who can influence thousands of people spreading rumors that can result in the loss of millions of dollars as in the case of South Korean blogger Minerva (Glionna & Park 2009). Therefore, the SNA should also adopt techniques to analyze the information or add parameters to verify the accuracy of the information.
- The SNA is applied to the personal data of people, in consequence this is an intrusive action and is often done without the consent of the user, who does not read the security policies when sharing information. In regards to this matter, laws should be created to regulate companies where this type of information is shared, in order to maintain privacy.

V. CONCLUSIONS

It is observed that the techniques that are used mostly for the analysis of social networks are the ones based on the importance of the node, whose disadvantages we indicated in Section 2, when it doesn't consider the interaction between the nodes or the closeness between them. This issue arises due to other techniques

being more complicated mathematically; therefore, the development of software tools is needed to implement the other techniques and thus provide better information to sociologists for different analysis.

The SNA can be applied in many fields as noted in Section III, and it's very important to continue the development of applications that facilitate daily life. These tools are so strong; you can analyze complex cases, such dealing with terrorism, or simplest cases.

Due to the complexity of social networks, the SNA still has a long way to go in development since each time new elements are analyzed, and these are not easily formable as we explained in Section IV. Mechanisms are needed to predict the behavior of social networks and where engineers, mathematicians and sociologists play an important role in developing such models and new techniques for analysis.

REFERENCES

- Baruah RD & Angelov, P 2012, 'Evolving Social Network Analysis: A Case Study on Mobile Phone Data.' Available in: <http://ieeexplore.ieee.org/document/6232815/>
- Carrasco, J, Fain, D, Lang, K & Zhukov, L 2003, 'Clustering of bipartite advertiser-keyword graph,' in Proc. International Conference on Data Mining (ICDM'03), Melbourne, Florida, November.
- Crook, J 2012, 'Bald Beliebers Remind Us: Just Because You Read It On Twitter, Doesn't Mean It's True.' Available in: <http://techcrunch.com/2012/10/26/bald-beliebers-remind-us-just-because-you-read-it-on-twitter-doesnt-mean-its-true/>
- Chen, D, Tang, J, Li, J & Zhou, L 2009, 'Discovering the staring people from social networks,' in WWW '09: Proceedings of the 18th international conference on World wide web. New York, NY, USA: ACM, 2009, pp. 1219–1220.
- Ding, L & Peng S 2011, 'Social Network Analysis Application in Bulletin Board Systems'. *International Conference on Intelligence Science and Information Engineering*.
- Eubank, S, Guclu, H, Kumar, V, Marathe, M, Srinivasan, A, Toroczkai, Z & Wang, N 2004, 'Modelling disease outbreaks in realistic urban social networks,' *Nature*, 429(6988), 180–184, May.
- Glionna, JM & Park J-M 2009, South Korea: Blogger 'Minerva' found not guilty. Available in: <http://articles.latimes.com/2009/apr/21/world/fg-minerva21>
- Goyal, A, Bonchi, F & Lakshmanan, LV 2008, 'Discovering leaders from community actions,' in CIKM '08: Proceeding of the 17th ACM conference on Information and knowledge management. New York, NY, USA: ACM, 2008, pp. 499–508.
- Hanneman, A & Riddle, M 2005, 'Introduction to social network methods,' Available in: <http://www.faculty.ucr.edu/hanneman/nettext/>
- Jamali, M, & Abolhassan, H 2006, 'Different Aspects of Social Network Analysis,' Proceedings of the 2006 IEEWICACM International Conference on Web Intelligence.
- Julei, F, Jian, C, Duoyong, S & Shouyang W 2012, 'Multi-factor analysis of terrorist activities based on social network', 2012 Fifth International Conference on Business Intelligence and Financial Engineering. Available in: <http://ieeexplore.ieee.org/document/6305170/>
- Kou, ZB & Zhang, Ch 2003, 'Reply networks on bulletin board system,' *Phys. Rev. E* 64(67), n32, 036117/1-036117/6.
- Lei, T & Huan, L 2010, 'Community Detection and Mining in Social Media', Synthesis Lectures on Data Mining and Knowledge Discovery, 2010, Vol. 2, No. 1, 1-137.
- Mao-Ping, W, Hsio-Yi, L, An-Pin, C & Chyan, Y 2011, 'An Integrated Home Financial Investment Learning Environment Applying Cloud Computing in Social Network Analysis', International Conference on Advances in Social Networks Analysis and Mining.
- Mindjumpers 2012, 'How People Spend Their Time Online' available in: <http://www.mindjumpers.com/blog/2012/05/time-spend-online/>
- Palau, J, Montaner, M, López B & de la Rosa JL 2004, 'Collaboration analysis in recommender systems using social networks,' in Cooperative Information Agents VIII, 8th International Workshop, CIA 2004. Berlin/Heidelberg: Springer, 2004, pp. 137–151.
- Quora 2010, How does LinkedIn's People You May Know work? Available in: <http://www.quora.com/How-does-the-LinkedIn-connection-suggestion-algorithm-work>
- Wei, K, Wen-Wu, D & Lin, W 2011, 'Research on Emergency Information Management Based on the Social Network Analysis: A Case Analysis of Panic Buying of Salt'. *International Conference on Management Science & Engineering* (18th), September 13-15, 2011 Rome, Italy.
- Xu, A, Zheng, X 2009, 'Dynamic Social Network Analysis using Latent Space Model and an Integrated Clustering Algorithm', *Eighth IEEE International Conference on Dependable, Autonomic and Secure Computing*. Available in: <http://ieeexplore.ieee.org/document/5380397/>
- Zhu, M, Liu, W, Hu, W & Fang Z 2009, 'The Application of SNA in IT-Based BRP', *Third International Symposium on Intelligent Information Technology Application*. Available in: <http://ieeexplore.ieee.org/document/5370043/>

