Extraction and Analysis of Tripartite Relationships from Wikipedia

Fawad Nazir[1,2], Hideaki Takeda[1]
[1]National ICT Australia, Sydney, Australia
[2]National Institute of Informatics (NII), Tokyo, Japan
E-mail: {fawad_nazir,takeda}@nii.ac.jp

Abstract

Social aspects are critical in the decision making process for social actors (human beings). Social aspects can be categorized into social interaction, social communities, social groups or any kind of behavior that emerges from interlinking, overlapping or similarities between interests of a society. These social aspects are dynamic and emergent. Therefore, interlinking them in a social structure, based on bipartite affiliation network, may result in isolated graphs. The major reason is that as these correspondences are dynamic and emergent, they should be coupled with more than a single affiliation in order to sustain the interconnections during interest evolutions. In this paper we propose to interlink actors using multiple tripartite graphs rather than a bipartite graph which was the focus of most of the previous social network building techniques. The utmost benefit of using tripartite graphs is that we can have multiple and hierarchical links between social actors. Therefore in this paper we discuss the extraction, plotting and analysis methods of tripartite relations between authors, articles and categories from Wikipedia. Furthermore, we also discuss the advantages of tripartite relationships over bipartite relationships. As a conclusion of this study we argue based on our results that to build useful, robust and dynamic social networks, actors should be interlinked in one or more tripartite networks.

1. Introduction

Social networking is an emerging field of research. A social network [1] [2] is a structured representation of social actors and their interconnections a.k.a. ties. Social networks form social groups or social communities that share interests. Social communities on the web are steadily emerging and the demand for forming an on-demand social network is immense [3][11][12]. Community members benefit from being linked to other members who share common interests, though having widely dispersed residences. Without these online social community portals on the web, people would not be able to find other people having the same interest. For example, if a person is searching for specific information he can look at his social network of people for information that interests them and get relevant references. Social tagging is also emerging. It helps create these social networks with the help of attaching keywords to information and user profile. The emerging techniques like tagging, blogging and wikis tend to create dynamic social networks. The techniques such as tagging perform very well for creating scalable social networks. Tagging basically assumes that user profile will only increase. An important aspect in tagging is completely ignored, that is how the social network should evolve if people join and leave the system or if people add as well as discard the tags.

This leads to a generic research question that is: how can we manage a dynamic social network? If we apply this dynamic nature of the social network to the existing tagging techniques it creates many isolated graphs and the robustness fails in the social network. The major problem is that the affiliation networks are always used as bipartite graphs. The problem with bipartite graphs is that just through one link, either the user is part of the group or is not. On the other hand if we look at the actual social network scenarios, people are linked with multiple similar interests. Here with the help of a simple example we will demonstrate the advantage of using tripartite social network over bipartite social networks. Suppose Alice and Bob are two students in an electrical engineering department. Alice and Bob both are interested in the semantic web. Alice is reading about tagging and Bob is reading about blogging. Here if we use conventional tagging, Alice and Bob are interrelated to each other in two communities i.e. Electrical Engineering and Semantic Web. Alice's Tag is "Semantic Web, Tagging" and Bob's tag is "Semantic Web, Blogging". Now if Alice removes this tag then her affiliation with both Semantic Web and Blogging will be gone and it will have no social connection with Bob (Fig-1a). On the other hand if we represent the association between Alice and Bob using a tripartite graph Alice will lose the relationship with the tagging community but will still be affiliated with Bob through the Semantic Web community link.

In this paper we address the research question: Can linking authors using one or more tripartite graphs help create more robust, manageable and dynamic social networks? We are inspired by the works in [5]. In this paper first of all we will analyze the Wikipedia data from different dimensions of user interactions and then we will

978-1-4244-1670-7/08/$25.00 ©2008 IEEE

make an analysis of Affiliation network [4] of Actor-instance, instance-category and derive from it actor-category affiliation network. As Wikipedia data is more reliable, it is also used in other related works [9][10]. After that we will explain our algorithm for tripartite graph generation from Wikipedia data. In the end we will discuss our results and give some concluding remarks.

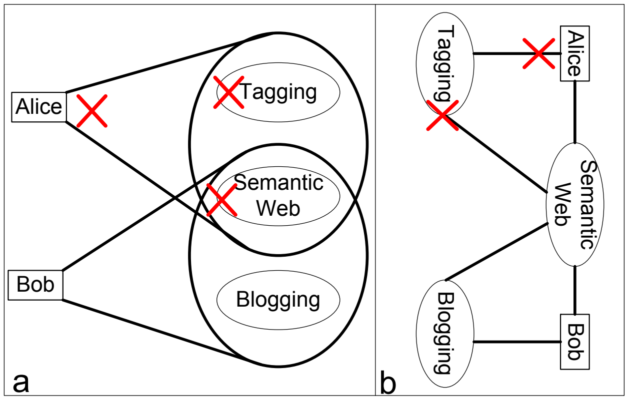

Figure 1: (a) Bipartite Social Graph (Social Tagging), (b) Tripartite Social Graph

2. Wikipedia Data Analysis

In this section we will discuss the Wikipedia analysis and some of our interesting findings. First of all we will briefly talk about the database structure of Wikipedia data to know how different tables are related to each other and what useful information we can extract from combination of different tables. Secondly we will discuss in detail the Wikipedia analysis that we performed on this data and the useful trends we discovered.

2.1. Design of Wikipedia Data

There are 41 Wikipedia tables [7]. In this paper we will only use four tables to extract the most interesting conclusions from Wikipedia. The tables we used are page table, user table, revision table and categorylink table. Page table is considered to be the core of Wikipedia. This table contains an entry for every page in Wikipedia. It does not contain the page text. It contains information about the page identity only and the reference for it in text table (this table contains the page text) and revision table (this table keeps tract of the page revision made by users). User table stores the information about the users of Wikipedia. Users are the people who are authors/editors of the Wikipedia articles. This table contains information about user identity and user privileges. Revision table is the most important table for our tripartite analysis. This table holds information about edits made by users to Wikipedia articles. It keeps track of the article that was edited, who edited it and at what time it was done. We use this table to find communities of users according to their article edit patterns. This table forms a baseline for our tripartite analysis. The last table which we used in our analysis is categorylink table. This table stores the categories to which a page is associated. This table adds the third dimension to our data i.e. Category. This dimension makes our analysis a tripartite analysis.

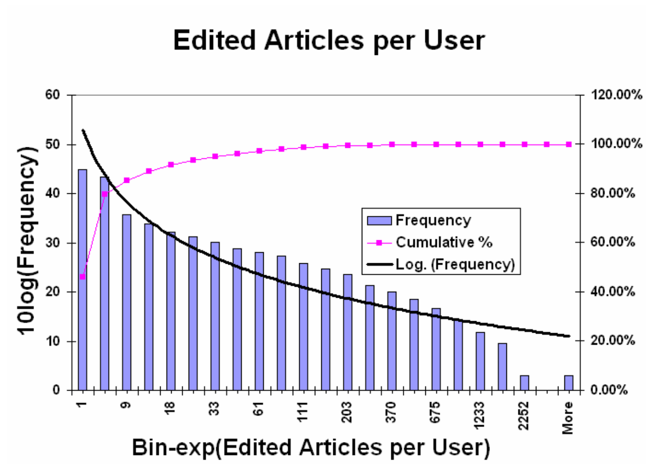

Figure 2 : Edited Articles per User: Histogram with Log Trend Line

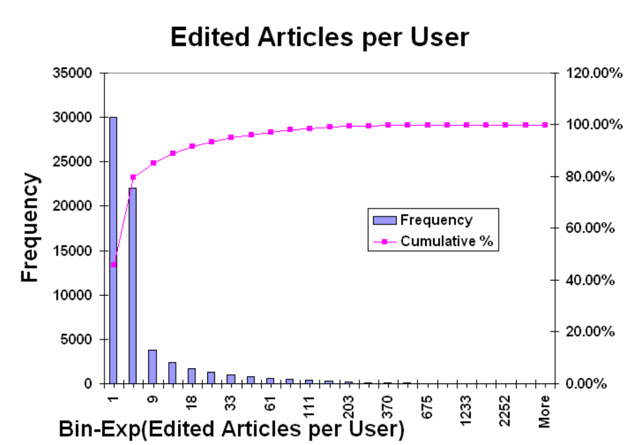

Figure 3 : Edited Articles per User: Histogram of Original Data

2.2. Aggregated Analysis and Trends

It is important to understand the general trends of Wikipedia data in order to come to some conclusion based on our analysis. Trends and aggregated analysis helps us to determine which sample of the data will best lead to conclusions from our findings. In aggregated analysis we perform analysis based on the relationship between the users and the articles they have edited. The following is the information about the data that we have used:

1. Number of articles: 10,218,632

2. Number of users: 65,678
3. Number of revised articles analyzed: 234,357
4. Number of users who made revisions: 626,413
5. Total number of article revisions studied: 31,135,556
6. Wikipedia dump date: September 08, 2007.

Table 1: Articles from Wikipedia with Highest Editors. (Not Unique)

Rank	Title	# of Edits
1	Jesus	18156
2	2006_Lebanon_war	16824
3	World_wrestling_entertainment_roster	16633
4	Runescape	15495
5	Hurricane_katrina	15091

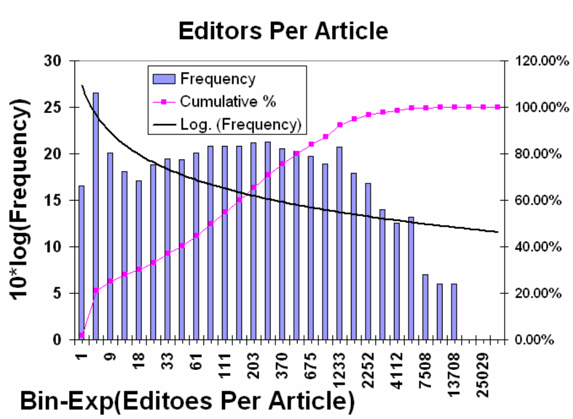

Figure 4: Editors Per Article: Histogram with Log Trend Line

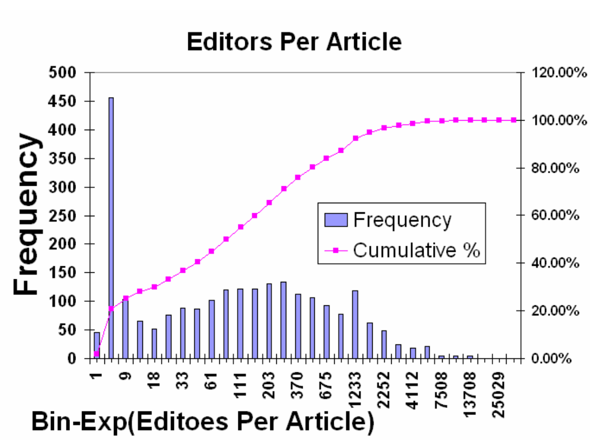

Figure 5: Editors per Article: Histogram of Original Data

Now we will discuss the aggregated analysis that we have performed with Wikipedia data. Figure-2 and Figure-3, shows the histogram of Edited Articles per User, implying thereby on average how many articles are edited by a user. Figure-2, shows the histogram with the logarithmic trend-line. This trend-line clearly shows that most authors edit fewer articles and very few users edit more than 10-15 articles. Figure-3 shows the histogram from the original data and we can clearly see that more than 80% of the users edit less than 6 articles. The maximum number of articles edited by any user is 113,872. The user name for all these contributions is: R_F (Due to privacy reasons we can not show the full user name, these are the initials only). This record is not shown in the graph in Figure 2, as per the limitation of Excel to only show 65,536 records.

Figure-4 and Figure-5 show the histogram of Edits per Article, meaning on average how many users edit an article. Figure-4 shows the editors per article in a histogram form with a logarithmic trend-line. From this trend-line we can conclude that most of the articles have few editors. If we study this trend-line it is not very steep. Therefore, this conclusion does not always hold. Interestingly this graph is a multi-modal graph. The first mode is around 1-8 users per article which covers around 50% of the articles and then the second mode is from 25-2251 which is again around 40% of the articles. In Figure-5, we plot the editors per article histogram of the original data. In this graph from the cumulative percentage we can conclude that more than 50% of the articles have less than 100 editors, and editors of more than 4000 are just 1%. We should also mention that in these two graphs we have not considered unique users. Therefore these edits per article can have multiple edits from one user. We will analyze the unique users in the next two graphs. In Table-1, we show that the top five most popular articles in Wikipedia have the highest number of editors.

The records shown in Table-1 are not included in Figure 4 and Figure-5, as per the limitation of Excel to only show 65536 records.

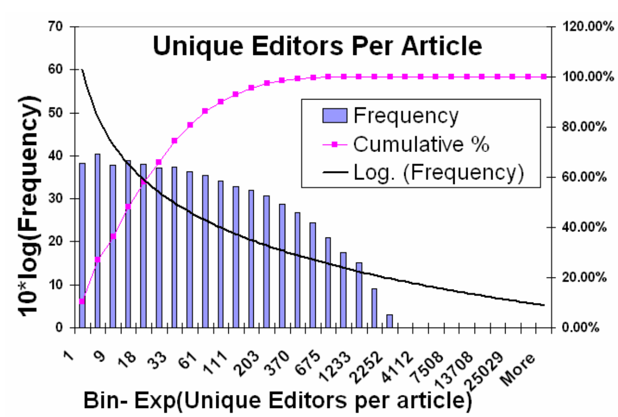

Figure 6 : Unique Editors Per Article : Histogram with Log Trend Line.

Figure-4 and Figure-5 show the histogram of Editors per Article but they include non unique users. Now in Figure 6 and Figure 7 we show the histograms of unique editors per article, meaning on an average how many unique users edit an article. In Figure 6, we show the histogram of unique editors per article with a logarithmic trend-line. This trend-line shows that the number of unique users editing an article is less and the frequency decreases if we move from fewer users to more users. In Figure 7, we show a histogram of unique users per article using original data. In this graph we can clearly see that the maximum numbers of unique editors, around 20%, is within the range of 8 per article. Another observation is that more than 50% of the unique editors per article are under 50 per article.

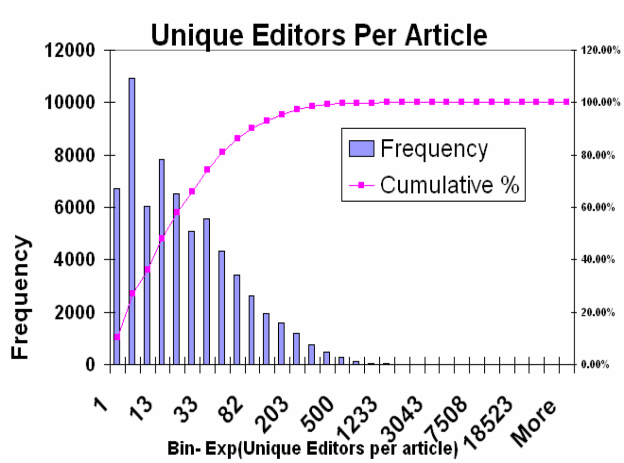

Figure 7: Unique Editors per Article: Histogram of Original Data

In Table-1, we show that the top five most popular articles in Wikipedia that have the highest unique editors. Interestingly two of the articles, Ipod and Paris_Hilton, were not in Table-1, which means in most of the articles single users edit multiple times. The records shown in Table-2 are not included in Figure 4 and Figure-5 as per the limitation of Excel to only show 65,536 records.

Table 2: Articles from Wikipedia with Highest Unique Editors.

Rank	Title	# of Unique Edits
1	Jesus	2103
2	Runescape	1958
3	Ipod	1876
4	Paris_Hilton	1846
5	Hurricane_Katrina	1756

Figure-8, is generated using GraphViz [8]. This figure represents an affiliation network of user->articles in Wikipedia. Another important trend that can be seen from the graph is that popularity as well as connectivity of the articles increases from outwards towards inwards. The articles on the boundary of the graph are least popular and mostly isolated and the articles towards the center are more popular among users. From this graph we got the following two conclusions.

The first conclusion was about the most popular article, Anarchism, with 12,385 edits. The second conclusion we got from the above discussed graph is about the biggest contributor. The biggest contributor is the person who contributed to the most articles. This is R_I (Due to privacy reason we can not show the full user name, there we only have initials) with user_id: 141644, and the total contributions made by him or her is 1,912.

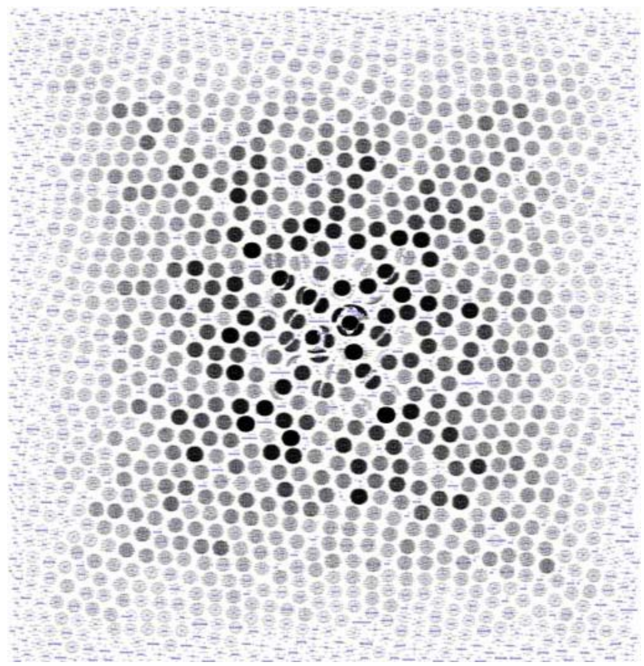

Figure 8 : Fading Graph Showing the Popularity of Articles

3. Affiliation Network Analysis (Author $\longrightarrow$ Article)

Affiliation network is a special kind of social network also known as involvement relation network. In affiliation network we have two modes i.e. set of actors and their affiliated set of social events. In this section we will analyze the affiliation network between Wikipedia users (Actors) and Wikipedia articles (Articles or instances).

$A = \{a_1, a_2, ..., a_n\}$ = Actors

$I = \{i_1, i_2, ..., i_g\}$ = Instances/articles

Where $\boldsymbol{n}$ is the total number of actors and $\boldsymbol{g}$ is the total number of articles.

3.1. Extraction of Sample Data for Analysis

In order to perform an analysis of affiliation network we extracted a subset of the data from Wikipedia containing most of the well connected users. This subset makes it easy to explain our analysis and its verification.

3.2. Data Representation for Analysis

Affiliation networks can be represented in different ways. Three of the common ways of representation are affiliation network matrix, bipartite graph and hypergraph. In our analysis we will mainly use affiliation network matrix and bipartite graphs to see the connection and distance between users. We used matrix representation to study the properties like Rate of participation, Size of events and density and bipartite graph representation to study the distance between actors and formation of one mode communities.

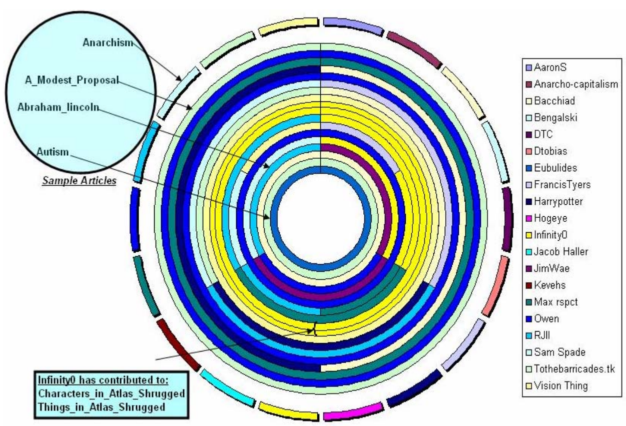

Figure 9 : Graphical Representation of Affiliation Matrix

3.2.1. Affiliation Network Matrix

Representation of an affiliation network in an affiliation network matrix is quite straight forward. We present the Affiliation Matrix as $A = \{a_{ij}\}$. Since we have ***n*** actors and ***g*** instances therefore $A = n * g$. In this matrix if we have 1 entry in (i, j), then that means i***th*** actor has edited the jth instance in Wikipedia. Formally it can be presented as:

$$a_{ij} = \begin{cases} 1 & if\, actor \;\; i \;\; has \;\; edited \;\; j \\ 0 & otherwise \end{cases}$$

Here a_{ij} is one cell in the affiliation matrix. Figure 9, shows graphical representation of the affiliation matrix between actors and instances in Wikipedia. The different colors represent different actors and each circle represents an article. As you can see the outer-most circle is Anarchism and it has multiple colors. So this means that many different actors contributed to this article. Similarly actor "infinity" who is represented in yellow has contributed to two articles: Character_in_Atlas_Shrugged and Things_in_the_Shrugged.

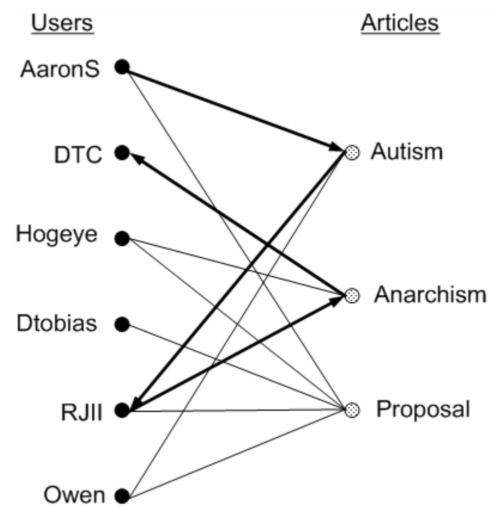

Figure 10 : Bipartite Graph of Actors -> Articles

3.2.2. Bipartite Graph

Bipartite graph [13] is another way to represent affiliation networks. In bipartite graphs we represent the relationship between two entities partitioned into two groups. In this case one group is actors and the other is articles. The connection between the actors and the articles is made if a user has contributed to that article. Bipartite graph can be used to see relationship between users from different perspectives. For example, group of users contributing to the same article, group of articles contributed by a user and distance between users. All of these above mentioned properties can be easily presented in a bipartite graph. In Figure-10, we represent a sample of bipartite graph of Actors

$$Actors \xrightarrow{Edited} Article.$$

In this bipartite graph we have 6 users and 3 articles. This bipartite graph shows properties of the relationship between the actors and articles. In this graph the out-degree of the node representing the user is equal to the number of articles that user has edited. Similarly, the degree of a node representing an article is the number of users who have contributed to that article. Another advantage of representing this relationship in a bipartite graph is that one can find indirect connections between articles and actors, and between actors and events. These relationships are much more clearly visible in a bipartite graph than in an affiliation matrix. In Figure-10, as an example we show the distance between Actors, AaronS and DTC with the help of $\longrightarrow$. AaronS and and DTC are not directly connected. They are connected through the following connections between actors and articles:

$$AaronS_A \rightarrow Autism_I \rightarrow RJII_A$$
$$\rightarrow Anarchism_I \rightarrow DTC_A$$

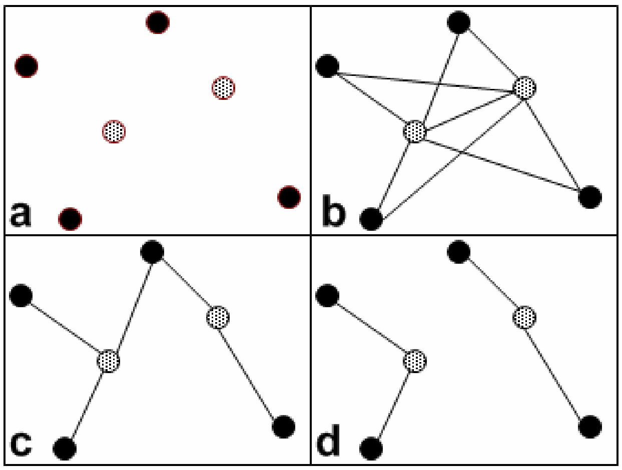

Figure 11 : Isolated Graph problem in dynamic bipartite social networks. (a) Empty Graph, (b) Complete Affiliation Network, (c) Intermediate Affiliation Network, (d) Isolated Social Network.

3.2.3. Discussion on Representation Methods

In this section we will analyze and discuss the above mentioned two representation methods and discuss the problems encountered in them. In Affiliation matrix we represented a binary relation between a user and an article. We have 1 if an actor has edited an article and 0 if the actor has not edited an article. In the case of Wikipedia this looks like an obvious case. However, in dynamic social networks we have the following case:

$$a_{ij} = \begin{cases} 1 \; if \quad in\ user \quad profile \\ 0 \qquad\qquad otherwise \end{cases}$$

So in this case this binary relation is highly dynamic. We would like to add another important point here, that when a user is affiliated with an event then it is not just one connection. This affiliation creates a whole network of users that are affiliated to that event. Moreover, it defines a reach-ability matrix of that user to other events and users. This shows that even one connection of user and event is quite important for the whole network structure. Loss of one connection to an event can result is isolation and unreach-ability of lots of other users and events. Therefore, in this paper we suggest that binary relationships are not reliable for dynamic social networks.

In Figure-9 the graphical representation of affiliation network matrix we can clearly see that if a user is taken out from one circle it will lose the network connection with all the users connected to that network. In the following example we will pictorially demonstrate our concept. In Figure-11, we have tried to explain this isolated graph problem in dynamic social networks. In this graph gray circles shows the articles, the black circles show the actors and the lines show the binary relationship between the actors and the articles. In Figure-11(a) we show an empty graph with no connection between the actors and articles. In Figure-11 (b) we show a complete affiliation network in which all the actors are connected with all the articles and articles are also connected to other articles. This case is very rare in the case of a social network. However, this is ideal for the formation of dynamic networks. In this case we have multiple paths between the actors and articles and between authors and articles. Figure 11(c) shows an intermediate affiliation social network. This type of social network is more close to the real affiliation networks and it also has an article with high between-ness centrality. In Figure 11(d) we eliminate one link between an actor and article and the connected graph in Figure 11(c) becomes a disconnected graph. This shows that bipartite affiliation networks are not robust. As we can see, in graph Figure 11(c) we can reach all the nodes from any node, but in the case of (d) just by eliminating one link between the actor and the article we end up having isolated affiliation networks. In this paper we have eliminated this problem using tripartite network instead of a bipartite network.

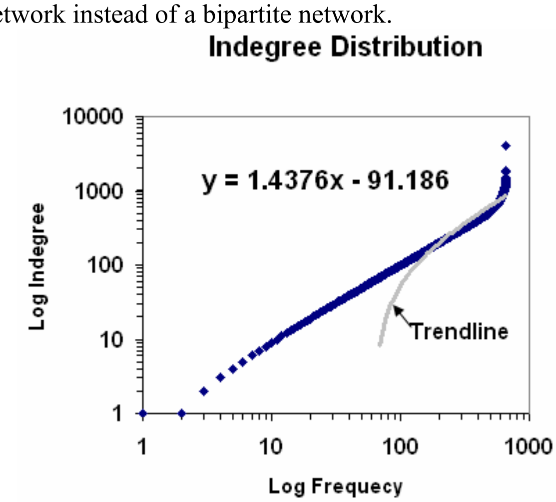

Figure 12 : In-degree Distribution Analysis for the Articles

4. Social Network Formation

In this section we will discuss how to create a social network based on a co-membership matrix using tripartite relationships. Before studying social network formation we have to make sure our data set follows the benchmark which is included in a well known cyber-communities paper [6]. Therefore we did an In-degree Distribution analysis for the articles and found that our results are close to the well known benchmark in order to proceed with our analysis. Our results are shown in Figure 12. In this graph, in-degree means the number of users editing an article. For example if article X is edited by 100 users then the in-degree of article X is 100. This analysis gives us confidence with our trimmed sample data set.

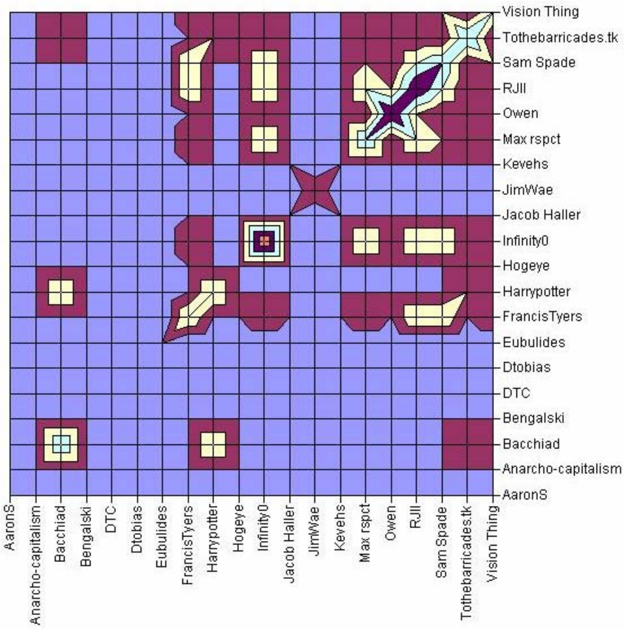

Figure 13 : Co-membership graph showing the Strength of the Relationships

4.1. Co-Membership Matrix

A Co-membership Matrix is used to check the number of articles edited by pairs of actors. If we use affiliation network matrix, we can see that two actors who have edited an article will have 1 in the same column. If actor i and j both have edited the article k, then we have: $a_{ik} = a_{jk} = 1$. From here we can deduce that the number of times two actors have 1 in the same column, means that they have edited these many articles together. Thus we define x_{ij}^{k} as the number of articles both the actors i and j have edited. Therefore:

$$x_{ij}^{k} = \sum_{k=1}^{h} a_{ik} a_{jk}$$

Where h = the total number of articles.

Based on this formula we have constructed a co-membership matrix. Figure-13 shows the visual form of this matrix. This graph shows the strength of the relationship between actors. The strength of the relationship between two actors is defined by the articles they have edited together. In this graph the dark blue color shows that those users have nothing in common and the orange color shows the strongest association.

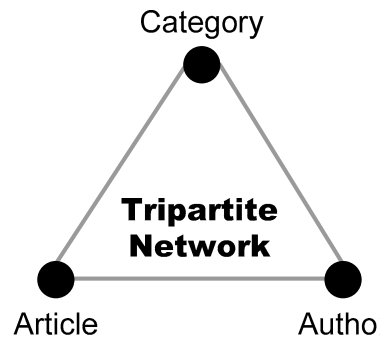

Figure 14 : Tripartite Network of Author, Article, Category

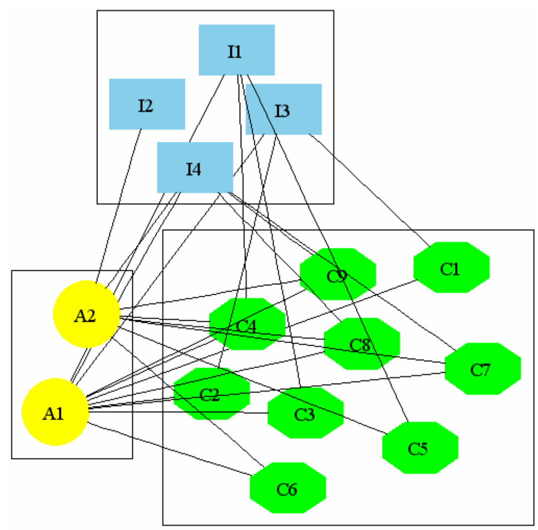

Figure 15: Proposed Tripartite Graph. Yellow Circles are actors, blue squares are articles and green octagons are categories.

4.2. Tripartite Affiliation Network

In this paper we propose a tripartite affiliation network of actors, articles and categories as shown in Figure-14. The tripartite graph is different from the bipartite graph. In a tripartite graph we divide the vertex set into three disjoint non-empty sets as opposed to two disjoint non empty sets in bipartite graphs. The tripartite graphs in this paper have the following conventions (Figure 15):

1. Actors are shown in yellow circles.
2. Article/instances are shown in blue rectangles.
3. Categories are shown in green orthogonal.

$A = \{a_1, a_2, ..., a_n\}$ => Actors.

$I = \{i_1, i_2, ..., i_g\}$ => Instances/Articles.

$C = \{c_1, c_2, ..., c_z\}$ => Categories.

Therefore, the tripartite graph becomes a graph with these three vertices:

$T \rightarrow \{A, I, C\}$ => Tripartite Affiliation Graph

A tripartite graph from Wikipedia is shown the Figure 17.

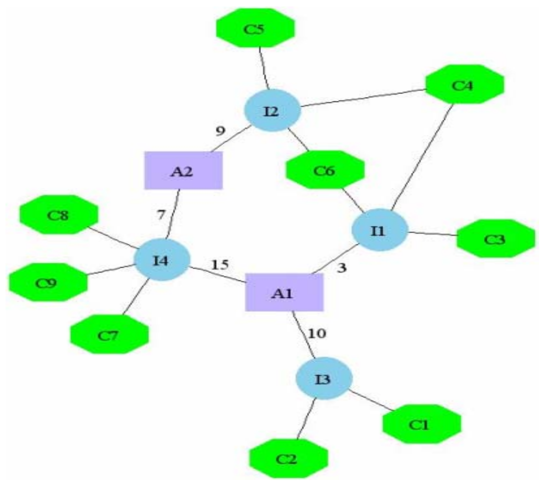

Figure 16 : Example Configuration to calculate similarity coefficient between two Wikipedia actors

5. Tripartite Data Analysis

As discussed in the previous section we perform tripartite analysis on Wikipedia data using Actors, Articles and Categories. In this section we will see how to conduct such an analysis. Furthermore, we will discuss our results and demonstrate the advantages of tripartite analysis over bipartite analysis.

In Figure 16, we show an example of our proposed tripartite social network. In this social network we have the following configuration:

Actors A_1 & A_2

$$A_1 \xrightarrow{Instances} \{I_1, I_3, I_4\}$$
$$A_2 \xrightarrow{Instances} \{I_2, I_4\}$$
$$I_1 \xrightarrow{Categories} \{C_3, C_4, C_5\}$$
$$I_3 \xrightarrow{Categories} \{C_1, C_2\}$$
$$I_4 \xrightarrow{Categories} \{C_7, C_8, C_9\}$$

From the above mentioned relationships we can derive the following indirect relationship between the actors and categories in order to complete our tripartite graph:

$$A_1 \xrightarrow{Categories} \{C_3, C_4, C_6, C_1, C_2, C_7, C_8, C_9\}$$
$$A_2 \xrightarrow{Categories} \{C_4, C_5, C_6, C_7, C_8, C_9\}$$

From the above example, if we create a social network based on a bipartite relationship between actor and instance, we will just have one mapping using I_4. On the other hand if we consider a tripartite network and use derived a relationship i.e. actors and categories, then we have five mapping C_4, C_6, C_7, C_8, C_9. This is shown here:

Bipartite network: $A_1 \xrightarrow{I_4} A_2$

Tripartite Network: $A_1 \xrightarrow{C_4, C_6, C_7, C_8, C_9} A_2$

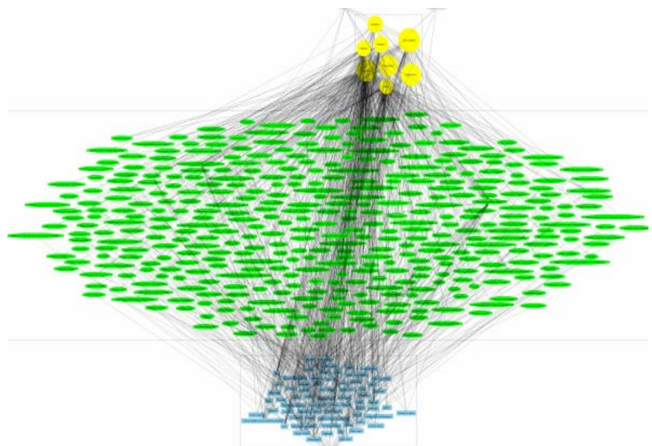

Figure 17 : Wikipedia Tripartite Relationship Visualization

This above mentioned example shows that a tripartite network builds a more robust and multi-path relationship between the actors. As we have seen in the case of a bipartite network Actor 1 was connected to Actor 2 just by one instance. On the other hand in the tripartite network Actor 1 is connected to Actor 2 by five categories. In the first case, if any one of the actors lose membership of instance 4, then they have no more connection among themselves. On the other hand, if we consider the case of a tripartite network Actor 1 has a very strong network with Actor 2. In this case, if anyone of them disjoins an instance only a few of the categories might be affected. However, using other categories both users will still have a social network. Now we will test our hypothesis on the real data from Wikipedia. In Figure

17, we show the sample data extracted from Wikipedia for a tripartite analysis. Yellow circles show Wikipedia actors, blue rectangles show the instances/articles these actors have edited and green orthogonals show the categories these instances are associated with.

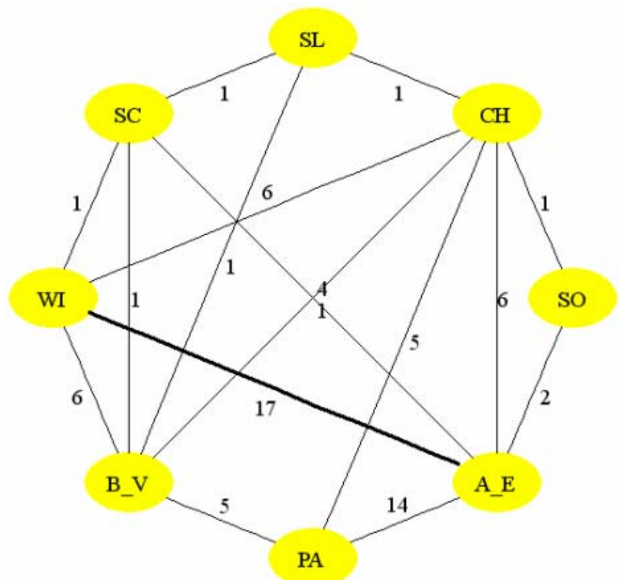

Figure 18 : Social network of Wikipedia user using the bipartite network analysis with link weights.

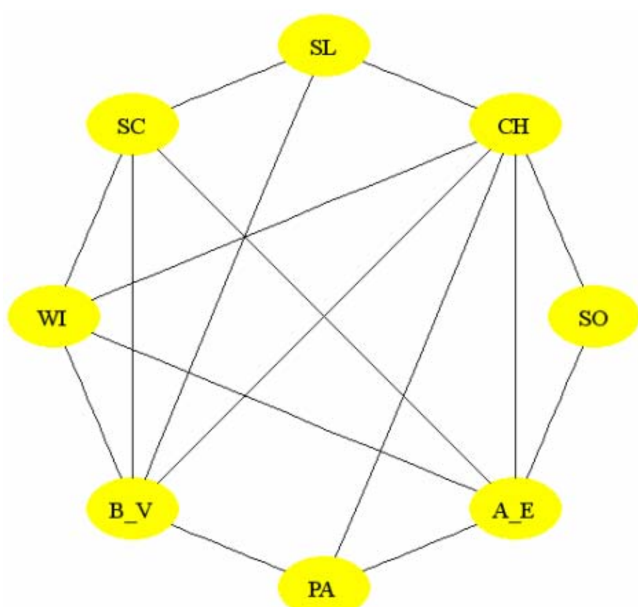

Figure 19 : Social network of Wikipedia user using the bipartite network analysis.

5.1. Density Property and Strength Analysis of Tripartite Affiliation Network

In this section we will first create a bipartite and tripartite network from the data given in Figure-17 and then we will study the density property of these networks [14][15][16]. We have taken eight Wikipedia users and due to privacy reasons we have created aliases for their names using name initials. In order to create a bipartite network of user we have taken the data of:

$$Actor \rightarrow Article$$

This is just one level of association among the users. In Figure 18, we show the social network of Wikipedia users using the bipartite network analysis. We have also created a tripartite network of users in Figure 17. In order to create a tripartite social network we used the three dimensional relationships of:

$$Actor \rightarrow Article \rightarrow Category$$

In a tripartite network we have two levels of relationships, Actor to Article and Article to Category. In Figure 20, we show a social network of Wikipedia users using the tripartite network analysis. Now we will discuss how the connections between the two actors are different in both bipartite networks and tripartite networks. In bipartite networks the actors are connected if they have edited same article. In this type of network if the user loses one level of association, he will lose the social connection to other actors. On the other hand in the tripartite network two users are connected in the following conditions:

1. If they edit the same article.
2. If they edit articles under the same category.
3. If they edit an article connected using a hierarchy of categories.
4. If they edit an article connected using a chain of categories.

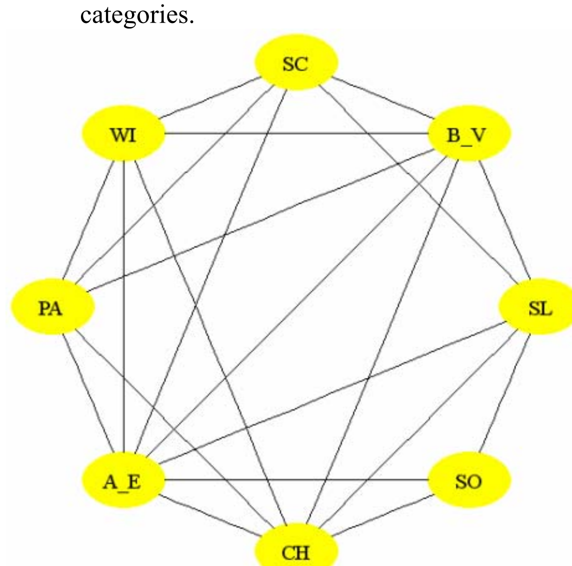

Figure 20 : Social network of Wikipedia user using the tripartite network analysis.

Now we will calculate the density of both bipartite and tripartite social networks. Density is a concept that considers the number of lines incident with each node in a graph. In a graph, $\binom{g}{2} = g(g-1)/2$ connections are possible if we eliminate the loops. In this case g is the number of nodes in the graph. In our sample data we have 8 users (nodes) in the graph. Therefore, the total number of possible connections is 28. The density of a graph is the proportion of possible connections that are actually present in the graph. It is the ratio of the number of connections present (L) to the maximum possible connections [4]. Density is denoted by Δ.

$$\Delta = 2L / g(g-1)$$

Where L is the actual number of connections in the network. We will define some of the terms before we perform the calculations:

L_b = Actual connections in bipartite network =16

L_t = Actual connections in tripartite network = 21

g_b = Total connections in bipartite network = 8

g_t = Total connections in tripartite network = 8

Δ_b = Density of the bipartite network

Δ_t = Density of the tripartite network

Therefore:

$$\Delta_b = 2L_b / g_b(g_b - 1) = 0.57$$

$$\Delta_t = 2L_t / g_t(g_t - 1) = 0.75$$

From our calculations we can clearly see that the density of the tripartite graph is better than that of a bipartite graph. This means that the social network constructed using tripartite relationships is well connected compared to the social network built using bipartite relationships. In this study we also took into consideration another important fact that is the strength of the relationships. Strength means how many different paths can be used to reach from one actor to another. Suppose we can use 10 paths to reach from A1 to A2. Then the strength of the relationship between A1 and A2 is 10. Figure 19 shows the social network of Wikipedia users using the bipartite network analysis with link weights. In this graph these links weight are actually strengths. Similarly in Figure 21, we show the social network of Wikipedia users using the tripartite network analysis with link weights. If we add all the weight of the links then that becomes the strength of the network. In the case of bipartite network the strength of the network is 72 and in the case of tripartite network the strength of the network is 219. Therefore, we can conclude that the strength of the tripartite network is greater than that of the bipartite network.

The bar graph in Figure 22 is a summary of the comparison between bipartite and tripartite networks using density and strength. In this graph, the missing blue bars show the missing link in the bipartite network. These missing links decrease the density of the social network. The height of the bars shows the strength. In the graph we can see that brown tripartite bars are higher than the bipartite bars. Therefore, we can conclude that the tripartite social network construction method is much better than the bipartite method.

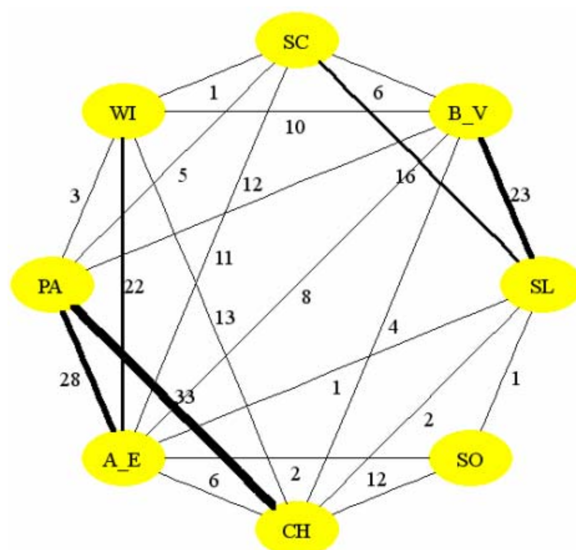

Figure 21: Social network of Wikipedia user using the tripartite network analysis with link weights.

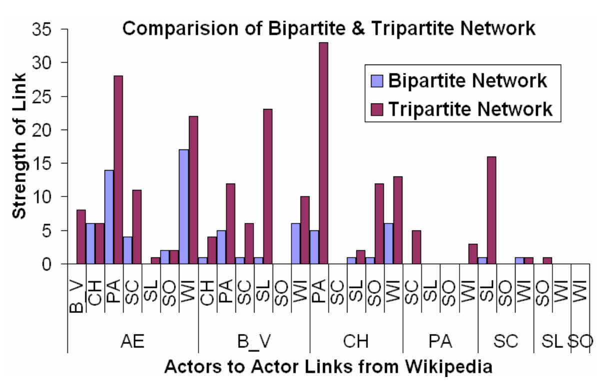

Figure 22: Comparison of Bipartite & Tripartite Network

6. Conclusions

Social networks play a vital role in creating relationships among social entities. Social networks have applications in a wide range of disciplines such as social and behavioral sciences, economics, marketing, and industrial engineering. Social network construction is the first important step before we can gauge any useful conclusions from the data. The most robust and connected social networks can give useful and meaningful results. To construct such a network, in this paper we propose to use tripartite relationships. We studied Wikipedia data thoroughly to extract tripartite relationships between actors, articles and categories. Having looked at the network representation techniques, we discussed the major problem of isolated social networks. This problem occurs when we use the existing network construction techniques for dynamic social networks. Therefore, we proposed a new approach using tripartite relationships. In the end, we performed a comparison of strength and density between bipartite and tripartite network. Our results show that a tripartite social network is better than a bipartite network in terms of strength and density. In our future work we are looking at the application of our findings and semantic analysis of our research [17]. Therefore, as a first step, we plan to build a semantic social recommendation system.

In order to put our ideas into a global perspective we would like to add that social interaction or social connection are not just one-to-one relationships between two social actors. They are actually many-to-many relationships. Furthermore, two actors can be connected through other paths that have several intermediate entities. So basically, as we discussed in this paper, two social actors are connected using multi-partite graphs. Therefore, as part of our future plan we will work on on-demand virtual social networks, on top of the complex multi-partite social network. In the end we would like to conclude that an information source combined with its social organization, derived from its origin and robust social connections, can make information more vital and authentic.

7. Acknowledgements

We would like to thank Mr. Hideyuki Tan (tanh@alpha.co.jp) from the National Institute of Informatics (NII) Japan for his help and support in setting up the database server for our Wikipedia analysis. We would also like to thanks Prof. Aruna Seneviratne and Dr. Max Ott from National ICT Australia (NICTA) for providing us with technical feedback.

8. References

[1] Adamic, L.A., Buyukkokten, O. and Adar, E. A Social Network Caught in the Web. First Monday, 8 (6). (2003). http://www.firstmonday.org/Issues/issue8_6/adamic/index.html

[2] Garton, L., Haythornthwaite, C. and Wellman, B. Studying Online Social Networks. Journal of Computer-Mediated Communication, 3 (1). Volume 3, issue 1. (1997)

[3] S¨oren Preibusch, Bettina Hoser, Seda G¨urses, Bettina Berendt, Ubiquitous social networks opportunities and challenges for privacy-aware user modeling, DM.UM'07 (2007)

[4] Wasserman S, Fausr K, Social Network Analysis, Cambridge Press,Publisher, (1994).

[5] Peter Mikka, Ontologies are us: A unified model of social networks and semantics, International Semantic Web Conference (2005).

[6] Ravi Kumar, Prabhakar Raghavan, Sridhar Rajagopalan, Andrew Tomkins, Trawling the web for emerging cyber-communities, Elsevier Science Computer Networks Amsterdam, Netherlands (1999)

[7] Wikipedia table Schema: http://www.mediawiki.org/wiki/ Category:MediaWiki_database_tables

[8] Graphviz, www.graphviz.org

[9] Gianluca Demartini, Finding Expert Using Wikipedia, 2nd International ExpertFinder Workshop, International Semantic Web Conference (ISWC) (2007).

[10] M.E.J. Newman, Who is the best connected scientist? A study of scientific co-authorship networks, Complex Networks, pp. 337–370, Springer, Berlin (2004).

[11] J. Zhang, M.S. Ackerman and L. Adamic, Expertise Networks on Online Communities: Structure and Algorithms, Proceeding of the 16th international conference on World Wide Web, pages 221-230, (2007)

[12] Han Woo Park, Hyperlink Network Analysis: A New Method for the Study of Social Structure on the Web, CONNECTIONS 25(1): 49-61, (2003)

[13] Malcolm Alexander, Using the bipartite line graph to visualize 2-mode social networks, Conferences in Computational Social and Organizational Science, (2005)

[14] M. E. J. Newman ,Mathematics of networks, The New Palgrave Encyclopedia of Economics, 2nd edition, L. E. Blume and S. N. Durlauf (eds.), Palgrave Macmillan, Basingstoke (2008).

[15] M. E. J. Newman, Analysis of weighted networks, Phys. Rev. E 70, 056131 (2004).

[16] M. E. J. Newman, Properties of highly clustered networks, Phys. Rev. E 68, 026121 (2003).

[17] Boanerges A, Meenakshi N etl. Semantic Analytics on Social Networks: Experiences in Addressing the Problem of Conflict of Interest Detection, World Wide Web (2006)