Supercharging Enterprise 2.0 Analytics modules with SNA metrics

Lamia Benhiba, Mohammed Abdou Janati-Idrissi

Equipe TIME, ENSIAS, Mohammed V Souissi University
Rabat, Morocco

Abstract

The increasing virtualization of collaboration has led to the use of enterprise 2.0 platforms to support work processes and consequently triggered the need for new tools to monitor collaborative initiatives. While many Enterprise 2.0 platforms offer analytics modules where they provide web-based metrics, these metrics do little to emphasize the social network aspect of users' interactions with each other. And seen that the social structure affects greatly the communication and information dissemination flows, which are critical for virtual collaboration, we believe that Social Network Analysis (SNA) metrics are a viable addition to the analytics toolset available in Enterprise 2.0 platforms today. In this article we present how supercharging the Analytics modules in enterprise 2.0 platforms with Social Network Analysis provides a more complete toolset for monitoring virtual collaboration.

Keywords: *Social network analysis; Business Intelligence; Analytics; enterprise 2.0; social collaborative platforms; virtual collaboration*

1. Introduction

In the face of the growing virtualization of organizational structures, organizations are resorting to Social and collaborative platforms to ensure collaboration and information dissemination among collaborators. These platforms, often dubbed Enterprise 2.0 tools [1], aim to facilitate communication and bypass pre-existing physical distances by offering a collaboration space that is common to all team members. Closely monitoring this virtual workplace is critical to the success of the team. Many available Enterprise 2.0 tools offer metrics inspired from web-analytics that help track users' behavior. Although these metrics reflect the generic state of activity on the platform, they fail to convey the social dynamics reigning the team's interactions. We thus propose using, in conjunction with classical web-based indicators, SNA-based metrics [2] in order to offer managers a more complete toolkit to gauge the use of the Enterprise 2.0 within virtual collaboration initiatives.

In the next section, we introduce "Enterprise 2.0" and the use of Social and collaborative platforms in organizations. Section 3 overviews the Analytics features available within Enterprise 2.0 platforms. To enrich these existent features, we propose, in section 4, a set of SNA-based metrics that assess the structural aspect of the virtual collaboration network. Section 5 presents the opportunity of supercharging Enterprise 2.0 Analytics module with SNA and hence rendering a more comprehensive monitoring tool for assessing virtual collaboration. In the last section, we provide concluding remarks and specify some future research directions.

2. The use of Enterprise 2.0 in organizations

Organizations are increasingly using social computing capabilities within their firewall to leverage the network effect and foster collaboration and knowledge management. Based on distributed technologies that collectively transform mass participation into valuable emergent outcomes, social collaboration platforms have become the new imperative for organizations to thrive. Applying some of the web 2.0 principles in organizations is what McAfee refers to as "Enterprise 2.0" and defines as the use of emergent social collaborative software platforms within companies, or between companies and their partners or customers in order to make visible the practices and outputs of their knowledge workers [1]. Enterprise 2.0 sprung from the thesis that openness, peering, sharing and acting global can harness external and internal resources and talent and achieve unparalleled growth and success as a result [3].

The socialization of collaboration platforms aims to translate implicit knowledge into explicit knowledge and combine it to render new knowledge. In fact, Enterprise 2.0 platforms aim to help knowledge workers generate, share and refine information through six technologies that McAfee refers to with the acronym SLATES: "Search" capabilities for discoverability of information, URLs to forge "Links" between enterprise content, ensuring easy access for "Authoring", allowing organization of data through "Tags", "Extensions" through applying recommendations by mining patterns and user activity, and notification via "Signals".

Based on these concepts, Enterprise 2.0 tools (such as wikis, weblogs, microblogs, social tagging tools and RSS etc.) focus on improving business processes by enhancing internal collaboration, internal knowledge management and knowledge retrievals within organizations [4]. And as organizations are evolving into new forms of team based, geography independent structures, collaboration related aspects are becoming increasingly challenging. In fact, geographic, cultural and functional lines in virtual teams generally cause breakdowns of communication flows. And since good, timely and interactive communication is a key factor for the success of virtual teams, managers are seeking more effective tools to ensure the dissemination of the right level of information, to the right people and in the right time. Hence, Enterprise 2.0 tools are gaining traction as they offer functionalities guaranteeing easy, flexible and global access, configurable security levels and subscription and search options.

While the use of Enterprise 2.0 tools for virtual collaboration has become the new imperative to overcoming communication issues, it causes many concerns regarding content quality and potential information overload. Therein lies the importance of monitoring users' activity on such platforms as it ensures an optimized level of interaction: Enough to spark and sustain collaboration but not too much to drown users in a sea of irrelevant information. Therefore, many Enterprise 2.0 platforms offer Analytics as part of their business intelligence module, to help make sense of the generated content and provide feedback that can enhance the navigability and user-experience on the platform [5]. The next section will present some of these BI features.

3. Business Intelligence in Enterprise 2.0 platforms

Business intelligence (BI) software is a collection of decision support technologies for the enterprise aimed at enabling knowledge workers to make better and faster decisions by providing the right data, at the right time to the right people. BI software aims to consolidate, analyze large quantities of operational data and make them available to decision makers. Originally based on a data-centric approach, BI is increasingly focusing on text and web analytics for unstructured contents due to the proliferation of web tools [6]. The use of web analytics to track the behavior of users on web-based platforms has thus made Analytics a viable toolset to assess employees' activities on enterprise social collaboration software. Most enterprise 2.0 platforms are hence offering a set of web-based metrics within their BI features.

The principle of web analytics is to rely on logfiles in order to measure the performance of a website by relating two sets of data: User data and context data. User data relate to the number and type of visitors and their technical environment while context data present the context of the website. The table 1 overviews some of the most common metrics provided by web analytics tools [5].

Table 1: Web-Analytics based metrics

User data based metrics		
Visits	***Loyalty***	***Technical***
Total visits	Length of visit	Browser type
Unique visitors	Depth of visit	Connection speed
Total page views	Return rate	Operating system
Average page views	Recency of visits	Screen settings
Bounce rate	Time on site	Flash settings
Location		Java settings
Context data based metrics		
Traffic sources	***Content***	***Internal search***
Search engines	Top content	Search terms
Referring sites	Top landing pages	Start pages
Direct traffic	Top exit pages	Destination pages

Many Enterprise 2.0 platforms offer an Analytics module where they provide KPIs inspired from web metrics such as illustrated in figure 1. We rely on the descriptions and the feature list of four popular enterprise 2.0 solutions, namely Jive, SocialText, Telligent and IBM Connections[1], to categorize the offered web-based metrics. These metrics relate to the various social objects available on the platform.

1 Sources: Jivesoftware.com; socialtext.com; telligent.com; www-03.ibm.com/software/products/us/en/conn/

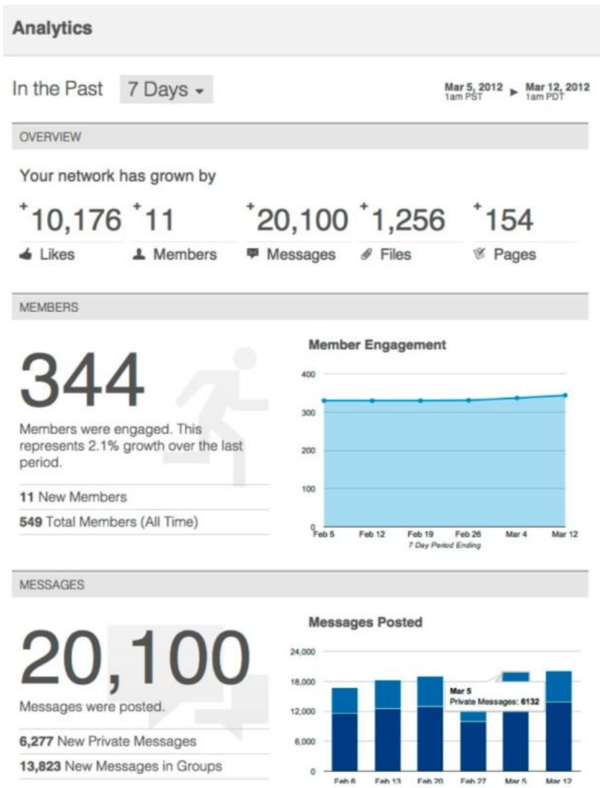

Fig. 1 Web-based Analytics on a Jive platform[1]

We consider all social objects and applications available on the collaborative platform as "Social Artifacts". This includes social containers such as communities, pages etc. and Social objects such as blog posts, threads, signals (regular posts, replies, or private). Artifacts, containers (components that host the artifacts) and activities available on the platforms can be listed as follows:

Table 2: Social Artifacts and activities in Enterprise 2.0 platforms

Containers	*Social Objects*	*Activities*
Blog Forum Community (Group) Page Conversation Project Calendar	Task Document Message Thread Status update Links User profile page Tag Notation Comment	Creation Consultation Modification Suppression Mentioning Following Unfollowing

An overview of the offered metrics shows that they could be classified in 2 categories:

[1] Source: Jivesoftware.com

- **Content centric metrics:** focus on assessing how users create content and social artifacts, share them and interact with other users through these objects.

Table 3: Content centric metrics

Activity count
Activity count by container
Count of end-user activity by container
Number of views of Social Artifact/container
The number of unique members who viewed/interacted with Social Artifact
Number of unique contributions (creation and modification of content)
Number of new Artifacts/containers
Number of unique people who created Artifact
Most followed/active Artifacts/containers

- **User centric metrics:** focus on the users and the explicit connections among them.

Table 4: User centric metrics

Number of unique authenticated visitors
Active Members count
Creator engagement
Consumer engagement
Active Members count by group
Count of connections made
Count of distinct members making a connection
Member profile completion
People with the most new followers
People with the most followers overall
Collaborativeness
Dashboard Personalizations
People Views
People Contributions
Total visitors
Total unique visitors
Time on site
Depth of visit
Loyalty
Visitor frequency
Visitor recency

Where some of these metrics can be defined as follows :

- Collaborativeness: contributors as a percentage of page viewers.
- Dashboard Personalizations: number of times a user customized his dashboard by adding, removing, moving or configuring widgets
- People Views: number of times a user viewed a People list, viewed a person's profile, viewed a People tag or the People tag list
- People Contributions: number of times a user edited his own profile, tagged or untagged a person, followed or 'unfollowed' a person

- Depth of visit: measurement of how many people are viewing how many pages
- Loyalty: measurement of how many times a visitor returns to the platform
- Visitor recency: The period between the last and current visit

While applying web-based analytics in the context of Enterprise social software offers various metrics to assess activity within the platform, it does little to emphasize the social network aspect of users' interactions with each other. Apart from the number (current and new) of connections and the number of views of a user's profile, no other metric assesses the interconnections among the platform's users. Moreover these metrics do not help recognize interaction patterns that harm the collaborative process. Whereas, a network perspective can give a formal definition of the social structure and patterns of relationships, helping thus pinpoint the sources of collaboration inhibitors. Therefore, we believe that Social Network Analysis (SNA) metrics are a viable addition to the social collaboration software analytics' feature.
In the next section, we introduce Social Network Analysis and present the SNA metrics that will complement web-based analytics in assessing virtual collaboration.

4. Social Network Analysis metrics

Social Network Analysis is a descriptive, empirical research method for mapping and measuring relationships and flows between people, groups, organizations and other connected information/knowledge entities. SNA has four features [7]: 1) It is motivated by a structural intuition based on ties linking social actors, 2) It is grounded in systematic empirical data, 3) It draws heavily on graphic imagery, and 4) It relies on the use of mathematical and/or computational models.
SNA is used to model complex systems that present interactive characteristics. Which is the case of collaboration as it is a dynamic network phenomenon. In fact, as a team collaborates, ties among stakeholders are built to allow knowledge and information dissemination. Thus, techniques from SNA can provide a viable approach to help make sense of communication flow patterns and examine the ties between team members.
SNA models the collaboration network as a set of nodes interconnected whenever they communicate, share information or interact on the social collaborative platform. The collaboration network is an undirected unweighted network that can be defined G := (V,E) as follows: Let V be the set of all stakeholders collaborating. The communication network is an ordered pair (V,E) where

$$G = \{\{p, q\} \in G_2(V) \mid p \ interacted \ with \ q\} \quad (1)$$

G2 (V) is the set of all subsets of V with size 2. $|V|$ are the nodes (vertices) and $|E|$ the links (edges) of the network. Interactions are defined as the activities operated by users on social artifacts within the social and collaborative platform as defined in Section 3.
Research identified a very diverse set of metrics that describe the structure of networks on the team level [8]. In a previous work [2], we classified the metrics pertaining to team network structure in three dimensions: Density, centrality and disconnected cliques and bridges. The density dimension offers an insight on the level of connectivity among the collaborators. The centrality dimension highlights the critical positions of certain actors within the team. The third dimension provides a view on the nature of the relationships among collaborators and their connections with cliques of the network (subgraphs of the network where every two nodes are connected).
In order to choose pertinent metrics to illustrate each dimension of the framework, we resorted to the SNA tool Networkx. The choice of this tool also stems from being the socle of the system we aim to build [9][10]. "Networkx" is a Python language package for exploration and analysis of networks. It is widely used within the SNA community as it has the most permissive license which allows integrating it within proprietary software. An adapted version of our framework can hence be presented as follows:

Table 5: Measures of network structure

	Dimensions	***Metrics***
Network Structure	Density	Density
		Average neighbor degree
	Centrality	Degree centrality
		Closeness centrality
		Centralization
	Disconnected cliques and bridges	Clustering coefficient
		Brokerage score (Betweeness)
		Number of cliques
		Heterogeneity

The metrics cited above are calculated in research in various ways. The underlying algorithms of "Networkx" define them as follows:

4.1 The density dimension

According to [11], density is the actual number of edges in the graph as a proportion of the total number of possible edges. It can be formulated as follows:

$$Density = \frac{2L}{n(n-1)} \quad (2)$$

where: L is the number of edges and n is the number of nodes

The average neighbor degree is a measurement that returns the average degree of the neighborhood of each node [12]. We define the neighborhood Ni of a node i as the group of nodes directly connected to that node.

$$k_{nn,i} = \frac{1}{|N(i)|} \sum_{j \in N(i)} k_j \quad (3)$$

where: where N(i) are the neighbors of node i and k_j is the degree of node j which belongs to N(i).

4.2 The centrality dimension

Centrality can be calculated on the node level and on a general level that is the network. The centrality on the node level can be quantified using various metrics. The most used measures are degree centrality and closeness centrality.

The degree centrality of a node is the number of links it has in the network [11].

$$C_D(v) = \frac{deg(v)}{n-1} \quad (4)$$

Where: v is a node (vertex) and n is the number of nodes in the network.

The degree centrality helps identify the central actors respectively in their local neighborhoods as it takes into account the immediate ties [13]. The closeness centrality, on the other hand, focuses on the distance of an actor to all other nodes in the network and is defined as follows [14].

$$C(u) = \frac{n-1}{\sum_{v=1}^{n} d(v,u)}, \quad (5)$$

where $d(v,u)$ is the shortest-path distance between v and u, and n is the number of nodes in the graph.

Centralization is calculated in order to contrast the gap between the largest actor centrality in a network and the other values [14].

$$C_G = \frac{\sum_{i=1}^{n}(C_{max} - C(v_i))}{(n-1)(n-2)} \quad (6)$$

Where Cmax is the maximum value possible for degree/closeness centrality and $C(v_i)$ is the degree/closeness centrality of node n_i.

4.3 The Disconnected cliques and bridges dimension

The clustering coefficient [15] is a way to measure how close a node (or vertex) and its neighbors are from being a clique. It is measured on the node level and on the network level. The local Clustering coefficient of a node is the proportion of existing connections among its neighbors compared to the number of all possible connections. For the undirected unweighted graph G:= (V,E) the clustering of a node u is defined as follows [16]:

$$c_u = \frac{2T(u)}{deg(u)(deg(u)-1)}, \quad (7)$$

where T(u) is the number of triangles through node u and deg(u) is the degree of u.

The Network Clustering coefficient [17] is the average of the local clustering coefficients of all the nodes.

$$\bar{C} = \frac{1}{n} \sum_{1}^{n} C_i \quad (8)$$

The Brokerage score is often calculated using the Betweenness measure which computes, for a node v, the extent to which all-pairs shortest paths pass through v [17].

$$C_b(v) = \sum_{s \neq v \neq t} \sigma_{st}(v)/\sigma_{st} \quad (9)$$

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

The number of cliques returns the number of maximal cliques for each node. A clique that cannot be extended by including one more adjacent vertex is referred to as a maximal clique. A maximal clique can't thus be a subset of a larger clique [18]. Networkx uses the Bron-Kerbosch algorithm as adapted by Tomita et al. [19] to render the number of cliques within the network.

Heterogeneity is calculated as the symmetric of assortatitivity because heterogeneous networks are disassortative. Assortativity, also called Homophily, describes the tendency of nodes to attach to similar nodes. Assortativity can be calculated related to the degree of a node and thus measures the similarity of connections in the graph with respect to the node degree. This coefficient is calculated using Newman's formula [20]:

$$r = \frac{n^{-1}\sum_{i,j} k_i k_j - (n^{-1}\sum_{i,j} \frac{1}{2}(k_i + k_j))^2}{n^{-1}\sum_{i,j} \frac{1}{2}(k_i^2 + k_j^2) - (n^{-1}\sum_{i,j} \frac{1}{2}(k_i + k_j))^2} \quad (10)$$

Where k_i ,k_j are respectively the degree of node i and j, and n the number of nodes in the network.

We thus define the heterogeneity of a network as:

$$h = 1 - r \quad (11)$$

Now that we listed the most pertinent SNA metrics in our context, the next section will present the opportunity of using SNA metrics and integrating them within the Enterprise 2.0's Analytics module in order to assess virtual collaboration.

5. Monitoring virtual collaboration by supercharging Analytics in Enterprise 2.0 platforms using SNA

The underlying social dynamics existing among collaborators are a critical factor to the success of any collaborative endeavour. Because SNA explicitly captures the interactions that would otherwise remain unseen, it provides a perspective that helps diagnose issues that could escape managers' attention. These collaboration issues can be identified through the metrics as well as through the graphical representation of the network. In fact, network visualization - as illustrated in figure 3- is important as it helps understand the network data, provides basis for analysis [21] and improve the way problems are solved. Many tools enable network visualization including the powerful open source package named Gephi [22]. The graph representation in the figure below illustrates for example a network dominated by few actors. When the network presents signs of bottlenecks (central nodes), shifting responsibilities to less burdened team members can help unload the bottlenecks and thus enhance the team's effectiveness.

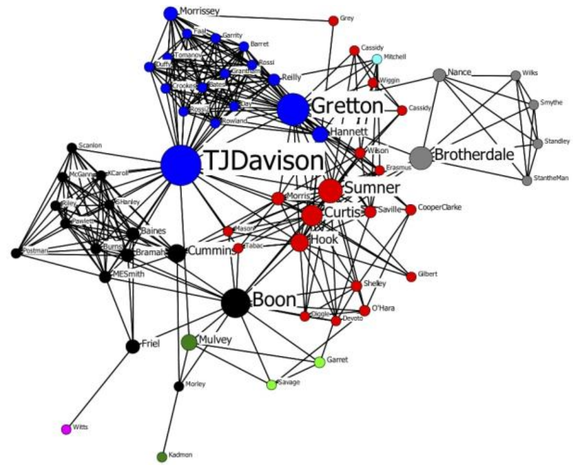

Fig. 2 An example of Social Network visualization of a team[1]

The observations that are drawn from the network visualization can be complemented by conclusions drawn from SNA metrics. These metrics are either overall network metrics, which convey the general structure of the network, or node-level metrics that represent the role collaborators play within the network. Combining the SNA-based metrics with the metrics provided within the Analytics module of Enterprise 2.0 platforms would hence offer a more comprehensive toolkit for assessing virtual collaboration.

Based on the rendered metrics and visualization, the decision maker can diagnose the network and then take the appropriate decisions to improve the team's structure. The table below presents possible interpretations of the metrics as listed in section 4.

Table 6: Interpretation of SNA measures

Metrics	***Interpretation***
Density	When density approaches 1, the majority of collaborators are connected to each other. This can be good for coordination but is often inefficient as the effort to sustain these interactions drains the collaborators' energy.
Average neighbor degree	A high average neighbor degree hints to a dense structure which can present redundant connections and thus inefficient collaboration practices.
Degree Centrality	High node centrality means that some collaborators are bottlenecks and can hinder the collaboration effort.

[1] Source: http://www.methods.manchester.ac.uk/methods/sna/

Closeness Centrality	A high closeness centrality reflects a high ability of a node to make connections across the network and it being central to relay information throughout the team. However, it can also hint to information hoarding practices.
Centralization	A high network centralization shows that its most central collaborator is dominating the whole network which means that information flows from this central source outwards. This can impact collaboration negatively.
Clustering coefficient	A high clustering coefficient on the node level means that the collaborators with whom he interacts are likely to form a cohesive group. A network with high clustering coefficient means the tendency of people to collaborate in subgroups. Too many subgroups represent a fragmented network that has few collaboration standards which can be harmful to the whole process.
Brokerage score	A node with high brokerage score has a high potential of becoming a bridge within the network through which communication flows between two subgroups otherwise disconnected. Although such nodes are exposed to a variety of information and ideas, their sensitive position presents a risk to the sustainability of the collaboration network.
Number of cliques	Too many cliques present a fragmented network that needs bridges for information to flow effectively.
Heterogeneity	A high heterogeneity means that non similar nodes frequently interact which often has a good impact on problem solving and innovation.

Although SNA may not pinpoint all the problems related to virtual collaboration, when combined with available Analytics it provides a comprehensive decision support tool that can help shed the light on many collaboration issues. It is worth pointing out that human intervention is necessary for the interpretation as the interventions that need to be performed depend greatly on each manager's objectives and his organizational context.

6. Conclusion

Managers have long used the web-based analytics of Enterprise 2.0 tools to help them supervise usage patterns. However, integrating a network perspective into the social collaboration software analytics' offers an additional toolkit that can bring more clarity into the team's social dynamics. Activity metrics, based on web analytics, help monitor users' behavior on the platform. SNA metrics, on the other hand, convey the interaction patterns among team members. In [10], we propose a system that supercharges the Enterprise 2.0 analytics with SNA in order to provide managers with a toolkit to diagnose collaboration issues. Further research will need to examine in details the interpretation aspect of the metrics in the context of virtual collaboration. An experimental study will also have to be conducted to attempt standardizing and refining the decision-making process that translates the observed metrics into business interventions aiming to enhance collaboration within virtual teams.

References

[1] A.P. McAfee, "Enterprise 2.0: The Dawn of Emergent Collaboration", MIT Sloan Management Review, vol. 47, pp. 21-28, 2006.

[2] L. Benhiba, M.A. Janati Idrissi, "An SNA-based evaluation framework for virtual teams". in OTM 2012 Workshops, LNCS, Vol. 7567, pp. 597–607, 2012.

[3] D. Tapscott and A. D. Williams, Wikinomics: How mass collaboration changes everything, Portfolio Penguin, USA, 2010.

[4] I. Gordeyeva, "Enterprise 2.0: theoretical foundations of social media tools influence on knowledge sharing practices in organizations", M.S. thesis, University of Twente, Enschede, Netherlands, 2010.

[5] A. Schneider, A. Steinhoff, "Applying Web Analytics Tools in the Context of Enterprise Social Software". In ECKM'12, 2011,

[6] H. Chen, R. Chiang, V. Storey, "Business Intelligence and Analytics: From Big Data to Big Impact", MIS Quarterly, Vol 36, N. 4, 2012, pp.1165-1188,.

[7] L. C. Freeman, The development of social network analysis: A study in the sociology of science. ΣP Empirical Press, Vancouver, Canada, 2004.

[8] K. Henttonen, "Exploring social networks on the team level - A review of empirical literature". Journal Eng. Tech. Mgt., Vol. 27, 2010, pp. 74–109.

[9] L. Benhiba, M. A. J. Idrissi.: "Towards an SNA system for evaluating project virtual teams based on social collaborative platforms". In NGNs'12, IEEE. (in press)

[10] L. Ben Hiba, and M. A. J. Idrissi. "Towards an SNA-BI based support system for the structural evaluation of project virtual teams." In Intelligent Systems: Theories and Applications (SITA), 2013 8th International Conference IEEE, 2013, pp. 1-7.

[11] S. Wasserman, C. Faust, Social network analysis methods and applications. Cambridge university press, Cambridge, NY, 1994

[12] P. Satorras, Romualdo, A. Vázquez, and A. Vespignani. "Dynamical and Correlation Properties of the Internet." Phys. Rev. Lett., Vol. 87, No. 25, 2001.

[13] R. A Hanneman, and M. Riddle: "Introduction to social network methods". Riverside, CA: University of California, Riverside (published in digital form at http://faculty.ucr.edu/~hanneman/). 2005.

[14] L.C. Freeman, "Centrality in networks: I. Conceptual clarification", Social Networks, Vol. 1, pp 215–239, 1979.

[15] D. J. Watts and Steven Strogatz "Collective dynamics of 'small-world' networks". Nature, Vol. 393, N. 6684, pp. 440–442, 1998.

[16] Saramäki, Jari, M. Kivelä, J.P. Onnela, K. Kaski, and J. Kertesz. "Generalizations of the clustering coefficient to weighted complex networks." Physical Review E, Vol. 75, No. 2, 2007.

[17] U. Brandes: "A Faster Algorithm for Betweenness Centrality"., Journal of Mathematical Sociology, Vol. 25 No.2, 2001, pp 163-177,

[18] C. Bron, and J. Kerbosch:. "Algorithm 457: finding all cliques of an undirected graph". Commun. ACM, Vol. 16, No. 9, 1973, pp 575-577.

[19] E. Tomita, A. Tanaka, H. Takahashi, "The worst-case time complexity for generating all maximal cliques and computational experiments", Theoretical Computer Science, Vol. 363, No. 1, 2006, pp 28-42

[20] M. E. J. Newman, "Assortative mixing in networks", Phys. Rev. Lett., Vol 89, No. 20, 2002.

[21] F. Carbone, J. Contreras, J.Z. Hernández, J.M. Gómez-Pérez, “Open Innovation in an Enterprise 3.0 framework: Three case studies”, Expert Systems Application, Vol. 39, No. 10, 2012, pp. 8929-8939.

[22] M. Bastian, S. Heymann, and M. Jacomy: "Gephi: an open source software for exploring and manipulating networks". In ICWSM. 2009