Exploring Social Network Analysis Techniques on Decision Support

Manuela Freire², Francisco Antunes¹,² and João Paulo Costa²,³
¹Department of Management and Economics, Beira Interior University, Portugal
²INESCC – Computer and Systems Engineering Institute of Coimbra, Portugal
³Faculty of Economics, Coimbra University, Portugal
maria-m-freire@telecom.pt
francisco.antunes@ubi.pt
jpaulo@fe.uc.pt

Abstract: Managers are increasingly concerned in making timely and correct decisions, using adaptable decision structures, processes and mechanisms. For this purpose, they resort to web discourse capture and analysis to obtain opinions that express different points of view, regarding organizational decision-making. Since communication and interaction among groups are very important in group decision, the exchanged and shared contents, as well as a social network analysis (SNA), can play an important role in this process. Understanding the interactions within a social network and how they can support decision-making is an important issue, representing the main motivation for this work. For the study presented in this paper, to understand how organizations can strategically use online communication as a decision support tool, we chose to use Facebook, where social actors gather around the same focus of interest or affinity (friends, co-workers, etc.) and, additionally, it is possible to create discussion groups for defining a problem, limit group members and context argument. In addition, we chose an ego-centric analysis focused on a single actor, instead of the network as a whole, to get a good picture of the “local” or his “neighbors” networks. The analyzed data were collected from Facebook and afterwards we applied SNA techniques in order to draw some conclusions.

Keywords: social networks, web discourse, SNA, decision support

1. Introduction

According to Laudon and Laudon (2011), managers routinely use Web 2.0 technologies, such as social networks, in order to make better and faster decisions, as they are able to represent, disseminate and retrieve the information circulating on the social web for decision support. The Semantic Web provides a basis for analyzing both the social actors’ communication behavior and the content of such communication, although there is modest research combining the tools to study communication’s social behavior and its contents, and fewer connect both to organizational decision-making. Connecting Web 2.0 tools within web discourse context and decision-making, shows that their effective use can help improve the extraction and use of knowledge within organizations.

The objective of this study is to understand if it is possible to structure data in a simple way, from web communication within Facebook social network, using SNA methodology, in order to extract information from the discourse produced for decision support. For that purpose, we established a debate for determining the logistics over the annual enterprise lunch (mainly the location and menu). Afterwards, we applied SNA techniques over the collected data in order to study the network. Our network was analyzed from the ego-centric perspective according to four perspectives: (1) a network with all social actors in the group; (2) a network with social actors who have had interactions with another member of the group; (3) a network with all social actors and respective posts comments; (4) a semantic network. This paper addresses the underlying complexity of web discourse analysis and, in section 2, we focus important aspects to be taken into account. A discussion on decision support within the context of social networks is also addressed in section 2, along social network analysis and most important metrics. In section 3 we present and discuss the obtained results. Final remarks are expressed in Section 4.

2. Decision support in the context of social networks

According to Choo (2006), in an ideal world the rational decision-making would require a thorough research of the available alternatives, reliable information on the consequences and preferences, in order to assess intended results. Some authors (Adam, 2008; Sueur et al., 2012; Herring, 2013) argue that network analysis is particularly suitable for the study of decision-making, because it allows recognizing the dynamic nature of networks, providing tools and techniques for measuring and evaluating the change. According to Antunes and Costa (2012b), the use of social networks decision support is still a relatively new subject, although some research has been done to determine the adequacy of social software for decision support.
Decision-making is defined by Nutt and Wilson (2010) as a process by which a course of action is chosen (by a decision maker), with a clear goal, from a set of alternatives. Literature shows several models and theories related to decision-making (Mintzberg et al., 1976; Huber, 1981; Das and Teng, 1999; Nutt and Wilson, 2010). Despite the numerous concepts, it is possible to observe that Simon (1977) is the most recognized, because it is accessible and easy to visualize. The decision model proposed by Simon (1977) is divided into four phases, with a systematic review between them and, according to Turban et al. (2005), is the most concise characterization of rational decision-making. With regard to decision process phases, and according to Turban et al. (2005), they can be described as: Intelligence phase, which involves environmental analysis, intermittently or continuously, and includes various activities seeking to identify situations, problem or opportunities; Design phase, which involves finding or developing and analyzing possible courses of action; Choice stage, is the critical time of decision-making, as it is the only one in which the current decision is made and where the commitment to follow a certain course of action is done; finally, the Implementation and Evaluation phase, where the chosen alternative is implemented.

The Web, especially the social web (online social networks), comprises interaction processes where the emerging networks are built through an established dialogue or web discourse. Studies essentially address the computer mediated communication point of view (Herring, 2010, 2013), natural language processing (NLP), web content analysis (Herring, 2010) or discourse analysis (Moser et al., 2013). These works are commonly focused on text analysis, though others are focused not only on the discursive aspects, but also on rhetorical, argumentative, semantic and pragmatic issues, to analyze the text produced in the web discourse.

For the analysis of produced texts within the web discourse, it is important to consider the following:

- The structure of the actors’ social network. It is important to consider metrics that are able to identify social actors and the key relationships between them, power metrics (which indicate prestige and influence), independence and ability to control and disseminate information;

- The communication. It encompasses the actual network organizational structure where the web discourse takes place (i.e., the so-called network architectures or topologies);

- The context. Web discourse analysis relies not only on understanding the language, but also on contextual aspects of discursive exchanges. In these exchanges, it appears that the discourse has a great deal of subjectivity and ambiguity, since it consists on sharing experiences and the construction of stakeholders’ different realities (Perelman and Olbrechts-Tyteca, 2005);

- The use of language. New technologies often require new forms of expressing emerging concepts (Bodomo, 2010). Particularly within the web discourse case, social actors tend to mix linguistic and non-linguistic resources in order to expedite communication;

- Semantics and pragmatics. Semantics are required to know the meaning of words and their relationship in the syntactic structure. In web discourse, text can be seen as a sequence of speech acts, and analyzing them is, according Dijk (1977), a central task of pragmatics;

- The content. If, on one hand, the produced web discourse content is relevant to understanding the relationship between social web, its context and communication (as they bring concepts of knowledge and social capital), on the other hand, within technological environments, the information becomes important and the social web is seen as a space that allows the interaction of heterogeneous individuals with different perceptions, where knowledge blends from the most diverse narrative sources.

Figure 1 depicts important aspects for analyzing text from a web discourse. It encompasses a network structure of social actors (social web), where players communicate (interact) with each other, within a certain context that uses its own language and must be analyzed, taking into account linguistic aspects to realize the produced content. Accordingly, at the center of the figure a social web, formed by a network of actors and text, is connected to each of the aspects. This connection is intended to express that all six aspects are interconnected and depend on each other and that any aspect of the speech must be understood in view of the inherent interaction of the actors’ social network.
According to the literature (Wasserman and Faust, 1994; D’Andrea et al., 2010; Erétéo et al., 2011), SNA methodology metrics are generally suitable for various levels of analysis and the resulting interpretation can be subdivided according to three different levels: (1) interpretation of the entire network; (2) interpretation of groups and components (socio-centric); (3) interpretation of actors’ individual positions (ego-centric).

When analyzing web discourses the main interest is to divide the structural analysis in ego-centric and socio-centric, because network members are defined by the connections they have with the ego. This approach is particularly useful to limit and define the boundaries of network analysis, given that the discourse analysis should be limited within its context. According to Hanneman and Riddle (2005) and Duijn and Vermunt (2006), the ego-centric analysis focuses on a player and not on the network as a whole, where the nodes connected to him are identified, allowing to get a good picture of the “local” network or its “neighbors” network. Regarding to socio-centric network analysis, and according to D’Andrea et al. (2010), it is used to analyze virtual social structures, in particular sets of relations among the actors considered for analytical purposes and enclosed in collective and social terms.

As stated by Wasserman and Faust (1994), there are properties and associated methods for actors alone (ego-centric) that include the way they stand out within a group, quantified by metrics such as centrality and prestige, their expansiveness level and their popularity parameters. There are also metrics for individual functions, identifying an actor’s isolation, its contacts and bridges (Marmo, 2011). The ego-centric networks give a view of the network from the perspective of a single node (ego) and its links.

The use of a smaller network allows to limit the number of elements under review since, according to Ma et al. (2010), it is a set of specific ties that bind the members of the group to the “ego”. In this way, it is possible to study the network in terms of its composition, characteristics (through metrics on the type of interaction established between the different social actors) and structure.

3. Study

We created a closed group from a Portuguese national company, geographically spread all over Portugal, within the social network Facebook where the company’s annual lunch (social meeting) topic was put to debate. The workers participated voluntarily with different levels of engagement. Some gathered information (from the web) on the topic and reasoned about it. The organization of a simple lunch has an associated logistics process, because when we gather a group the choice of location is conditioned by the number of people. Consequently, the organizer needed to know who will be present, as well as the preferences for the type of food and space in order to satisfy the attendees. Accordingly, the rational decision-making required a thorough search of the suggested alternatives, reliable information on the consequences and preferences, to get the intended results.

The data were collected between 22-Nov-2014 and 11-Jan-2015 and, afterwards, we applied SNA techniques in order to study the network, which was analyzed from the ego-centric perspective. The group participants were aware of the language in use, as well as the associated semantics and therefore, there was no need for ontologies to decipher what is implied within the speech. We used Facebook as it is a social network with some maturity, robust and relatively well documented on to how the discursive exchanges are structured (Russell, 2013). The use of a relatively small network allowed to limit the number of elements under review since,
Manuela Freire, Francisco Antunes and João Paulo Costa

according to Ma et al. (2010), it is a set of specific ties that bind the members of the group to the “ego”. In this way, it was possible to study the network in terms of its composition, characteristics (through metrics on the type of interaction established between the different social actors) and structure.

To understand the network structure of our group analysis we considered the metric Degree Centrality (Everett and Borgatti, 2005; Milka, 2007; Antunes and Costa, 2012a), which provides insights how information flows between nodes and the level of interactions between them. According to the literature (Izquierdo and Hanneman, 2006; Erétéo et al., 2011), within oriented networks, it is necessary to distinguish between incoming calls (In-degree) and outbound links (Out-degree). The Out-degree concept is usually a measure of how influential the actor can be inside the network, which substantiates the activity that the actor has to respond to others. The In-degree means how other actors send information or maintain ties with a particular actor (Wasserman and Faust, 1994), what clearly shows the importance of this actor’s posts. The fact that these posts are extensively commented or displayed might ascertain their significance.

This study also considered Betweenness Centrality and Closeness Centrality metrics. The idea underlying the Betweenness is the fact that when a particular node is between other nodes that makes it powerful (Wasserman and Faust, 1994). The more central an individual is, the better he controls the information flow, resources and communications, moving between nodes (Izquierdo and Hanneman, 2006). Regarding Closeness of an actor, this measures evidences the proximity of a node to the other network nodes (Wasserman and Faust, 1994), i.e. the length of the actor paths to all other actors.

We also considered PageRank metric, which evaluates the importance of each node within the network, based on the number and quality of the nodes that are connected to it and the Eigenvector Centrality metric that aims to measure the importance of a node, depending on the size of their neighbors. Therefore, even if a node is connected to only some other nodes of the network, having a low Degree of Centrality, these neighbors may be relevant and, consequently, the node is also important.

Data extraction was initially made by using NodeXL1, essentially because it is a free tool that simply and quickly allows importing Facebook data. However, it had some limitations as it only allowed collecting one type of network. For that reason the data were also taken from Facebook using Netvizz APP2 that allowed to extract data in three different ways and also to create four distinct networks:

- Network 1 - with all social actors who were part of the group (Group Connections Nodes). This network consisted of 27 nodes and 140 links between them (Edges). The network encompassed all social actors, regardless of any interaction between them;
- Network 2 - with all social actors with interactions (Group Interactions Nodes), i.e., exchanged “likes” or comments between them. This network consisted of 16 nodes and 32 links between them (Edges);
- Network 3 - with all social actors and interactions with all respective posts comments (Group Posts and Users-Nodes). This network consisted of 64 nodes and 142 links between them (Edges). Facebook contents were captured as plaintext, which guaranteed the distinction between discursive textual content, relevant to the analysis, and other forms of communication language;
- Network 4 - this network was created with all identified concepts and respective posts. Then, using the data collected for network 3, we selected posts containing text and built a semantic network (network 4). The resulting data set, consisting of 17 posts, was treated semi-automatically using Microsoft Excel. After cleaning the “noise words” (e.g. prepositions, conjunctions, definite articles and indefinite articles) from the 17 posts, 122 concepts were identified. For the purpose of our analysis we also considered as concepts the use of ellipsis (“...”) and smiles (“:)”) identified in the discursive exchanges. The resulting network consisted of 139 nodes and 155 links between them (edges).

Using Gephi3, we calculated the metrics for the nodes of social actors for networks 1 and 2, which included Degree, In-degree, Out-degree, Closeness, Betweenness, Eigenvector and PageRank. To facilitate interpretation of the network, Closeness, Betweenness, Eigenvector and PageRank metrics were normalized to present values between zero and one.

---

1 http://nodexl.codeplex.com/
2 https://apps.facebook.com/netvizz/
3 http://gephi.github.io/
As networks 3 and 4 had more than one type of entity, it made no sense to calculate the same metrics as in network 1 and 2. Network 3 had three types of nodes: users, photos and text only posts. In the case of network 4, it had two types of nodes (posts and concepts), for which the calculated metrics were the In-degree, Out-degree, number of “Likes” received and Modularity Class, which allowed to identify subgroups (Brandes and Erlebach, 2005). In the next subsection we describe the obtained results for all four analyzed networks.

### 3.1 Network 1 – links between social actors in the group

Network 1 contained all the group members who had at least one connection to any other element, consisting of 27 nodes and 140 edges (table 1 shows the results for this network). The most active social actor is member TM, because it had an Out-degree (14) that was higher than all others (meaning that he created links with 14 other actors). Meanwhile, actor MS was the element with the highest value in the In-degree metric (21), which meant that MS posts were viewed or commented the most. “Likes” and/or comments posted with these actors’ messages can be synonymous (or not) of agreement with an idea. Therefore, this data could provide relevant information for decision support, by revealing some level of agreement or disagreement with it. MS also have the highest values in Closeness, Eigenvector and PageRank (1; 1 and 0,17 respectively). Having a high value in Closeness represents independence and communication with other players on the network, using a minimum number of intermediates. Actors that have more links tend to be more powerful because they can affect directly other actors. Results from the Eigenvector confirm that MS is an important actor in the network, since this metric aims to measure the importance of a node, depending on the size of their neighbors. Thus, even if a node is connected to only certain other nodes, hence having a low Degree Centrality, these neighbors may be relevant and, therefore, the node is also important when it has a high Eigenvector. Concerning the Betweenness, TC is the actor showing the highest value. This social actor can be designated as a key player, with power to influence others. This is important in the web discourse analysis because the actor, through its discourse, can influence in a great deal the acts and minds of others. Figure 2 displays the importance that the above mentioned actors have within the group.

![Figure 2: Group connections nodes](image)

**Table 1:** Network 1 SNA metrics
3.2 Network 2 - social actors with interaction within the group

Network 2 is composed by all group members that had at least one interaction with any other element, constituted by 16 nodes and 31 edges. The obtained results are shown in Table 2 shows where social actors with higher values in the different metrics are MS, RA and LL.

MS had the highest value in Eigenvector and PageRank (1 and 0.35 respectively). LL was the element with the highest value regarding In-degree, reaching 13, and Betweenness recorded a value of 0.12. These results show the power of the player in control of the information flow. If, on one hand, he had a high In-degree showing the importance of its posts, the high value of Betweenness.

AR had an Out-degree of 4, meaning that he interacted with other four players within the network. AR and PD recorded the highest value of Closeness, which meant that they communicated with many actors, using a minimum number of intermediaries, quickly sending information to others. This metric is related to how long information takes before being shared by all nodes of the network.

Figure 3: Group interactions nodes

Table 2: Interactions network

3.3 Network 3 - social actors with interactions and respective content

Network 3 is composed by all group members who had at least one interaction with any other element and the published content following the interaction, consisting of 64 nodes and 142 edges. As mentioned earlier, network 2 has three types of nodes: users; photos and text posts.

“Likes” placed in the post by actors were calculated, see Table 3, because their role should not be underestimated. Their use may be interpreted as a position for a subject point of view and gives an idea in real time (or near real) of acceptance (or not) of what is being published. When social actors are part of a group they can express common interests by placing “likes” reciprocally in posts, which can be interpreted as an asynchronous conversation that can take place over a long period of time.

3.4 Network 4 - semantic network

Network 4 consisted of all the concepts and identified posts, meaning 155 edges and 139 nodes, of which 122 are concepts and 17 are text posts. Through a semi-automated textual analysis it was possible to extract a list of commonly used keywords and to rank them. Table 4 (see below) evidences the five most commonly used concepts in the web discourse, as well as the number of posts in which they appear.
Analyzing the produced discourse in social networks can be a "long shot", since we do not know in advance what type of data we will get and which will be the social actors' behavior. It is difficult to predefine the structure of any social network because, due to its dynamics, it is constantly changing and evolving. Similarly, it is not possible to predetermine its boundaries or the nature of their ties, since the links can be global or individual, compact or scattered. In this particular case, the analyzed group can be characterized by its low activity because it just...
published a few posts with text (17 in total) and posted comments with only a few words, which accounts for the low number of keywords found in the analyzed speech.

Generally, the language used in web discourse is characterized by a tendency to reduce the amount of quantity and character types to express an idea and, consequently, to speed up the communication process. By visual examining Figure 5, it is possible to state that from the 17 analyzed posts, only one has a larger number of concepts and the rest have a very low amount. We used the Modularity Class algorithm to detect the different community (post) within the network, because all concepts belonging to the same post are seen as a community. This indicates that they are more densely connected with each other than with the rest of the network being then considered as belonging to the same community. In the figure it can be easily identified which concepts belong to the different posts. In the figure it can be easily identified the different posts, and concepts that are part of each one.

4. Final remarks

Capturing and realizing everything that circulates within web discourse can be very important for organizations, because the exchanges (of opinion and information) originated in social networks are performed using computer-mediated communication. This contemporary reality requires that individuals within organizations develop new skills, particularly in terms of decision-making within the social web, by incorporating enhanced mechanisms of interaction, because its use is much more than a simple trend.

This study objective was to find out if the speech produced in a social network (Facebook), regarding a decision on where to organize a company lunch (social meeting), is likely to be used in decision-support and if the discursive exchanges can be measured through the SNA metrics. If we consider that decision support systems aim to provide decision-makers with useful mechanisms during the various stages of the decision process, then we stand that Facebook could provide information for such purpose. Through the results of Table 4 it is possible to understand how many people are going to be at the lunch, the kind of place they want and the food they prefer. These results can be useful within the intelligence phase, while gathering information on the problem and within the design phase, when analyzing the problem and seeking alternatives for its resolution. Specifically, the calculations of different SNA metrics and the analysis of multiple networks allowed a richer, more structured view of those involved in discourse, as well as most used concepts.

The analysis and visualization of concepts used in web discourse can be used as insights for follow-up decisions, as it allows efficient access to key information. The obtained result provides information on the key terms used during the discursive exchanges and can be used to analyze various alternatives of a problem and to support decision making. By identifying the number of times that the concepts appear in all posts, it is possible to determine any possible preferences and opinions of social actors. By putting all the group members and their respective interests in the same level, we get an outline of the views and group positions (in favor or against). Therefore, when a choice has to be made, argumentation gets into the “game” and network analysis can help to capture it. Therefore, network analysis is particularly suitable for the study of social network decision-making because it allows to recognize the dynamic nature of networks, providing tools and techniques for measuring and evaluating its changes.

References


