

Semantic Social Network Analysis Foresees Message Flows

Matteo Cristani, Claudio Tomazzoli and Francesco Olivieri

Dipartimento di Informatica, University of Verona, Strada le Grazie 15, Verona, Italy

Keywords: Social Network Analysis, Information Flow, Semantics of Data Flow.

Abstract: Social Network Analysis is employed widely as a means to compute the probability that a given message flows through a social network. This approach is mainly grounded upon the correct usage of three basic graph-theoretic measures: degree centrality, closeness centrality and betweenness centrality. We show that, in general, those indices are not adapted to foresee the flow of a given message, that depends upon indices based on the sharing of interests and the trust about depth in knowledge of a topic. We provide an extended model, that is a simplified version of a more general model already documented in the literature, the Semantic Social Network Analysis, and show that by means of this model it is possible to exceed the drawbacks of general indices discussed above.

1 INTRODUCTION

Social Networks are considered, on the current panorama of web applications, as the principal virtual space for online communication. Therefore, it is of strong relevance for practical applications to understand how strong a member of the network is with respect to the others.

Traditionally, sociological investigations have dealt with problems of defining properties of the users that can value their relevance (sometimes their importance, that can be considered different, the first denoting the ability to emerge, and the second the relevance perceived by the others). Scholars have developed several measures and studied how to compute them in different types of graphs, used as models for social networks. This field of research has been named *Social Network Analysis*. Sometimes the same name is attributed to a wider context, where we also mean to include analysis of the ways in which such values arise (for instance, processes able to change importance of members), or to provide methods for employing these measures in applications.

Majorly, scholars dealt with the Social Network Analysis from the viewpoint of *information flow*, namely they provide models of importance (and other aspects as well) to understand how probable would be that a piece of information passed through a given node. Mainly, the information flow has been studied for propagation of viruses (both in medical and in computer security contexts), news spread-out (and hence, studies about viral marketing as well), and

message passing in certain application contexts.

Three basic measures have been developed that belong to the family of *centrality measures*: degree centrality, closeness centrality and betweenness centrality. In this paper we criticize the models of social network analysis developed for these measures, showing that there are cases in which these measures are not adapted. The criticism arises mainly as related to the absence of semantic aspects in measures. To show what we mean with these limits, let us introduce a general example.

Example 1. Consider two users of facebook, Alice and Bob, and assume that the measures of importance is settled to coincide with the number of friends, the distance to non-friends, and the probability of being in common between two non-friends. Alice results to be much more important in the network than Bob, under all the three measures. However, to a closer observation we notice that this result is definitely true for certain topics whilst it results false for other ones. In particular, Alice is much more expert than Bob about Geography and History, equivalent with respect to Sport and weaker for Cuisine. When someone passes a message to Alice and the message regards Geography, she is much more likely to pass the message than Bob. Conversely, when a message regards Sport the opposite case holds. Cuisine information flow is better when passes through Bob.

The above described example shows that it can be the case that two members of a social network can exhibit different orders of prevalence in terms of centrality depending on the topic we refer the prevalence to.

This may produce effects that cannot be reproduced by a single index, as shown in the example below.

Example 2. Consider five individuals: Alice, Bob, John, Annie and Charlie. Alice is connected to Bob and John; John connects also to Bob and Charlie and also Annie is connected to Bob and Charlie, while Bob is connected directly to everyone and is person who loathes gossips when the others like or accept it.

If we don't consider topics we would say that dropping a gossip in the network, the right person to deliver it to have it spread is of course Bob.

Unfortunately, the message has contents of a topic which probably will see Bob cancel it, instead of forwarding it, while both John and Charlie are good choices because they are directly connected to three people each and they have a different attitude toward gossip than Bob.

The purpose of this paper is to give account to the aspects showed in the example above. We provide a model of Social Network Analysis that takes into account topics, and show that it can foresee information flow for message treating those topics in a more accurate way than classical topic-free social network analysis. We also name Semantic Social Network Analysis the techniques we studied in this investigation to cover a part of research that some previous studies did not cover satisfactorily.

The rest of the paper is organised as follows: in Section 2 we provide the actual technical part of the paper. Further we employ Section 3 to discuss related work. Finally Section 4 takes some conclusions and sketches further work.

2 SEMANTIC SOCIAL NETWORK ANALYSIS

In this section we introduce two distinct themes of graph theory: Social Network Analysis and Information Flow. The basis of both is the very general notion of a labelled graph, that we assume to be known to the reader, and specify in terms of form of the labels in Subsection 2.1. Social Network Analysis is extended in semantic terms in Subsection 2.2 and analysed in computational details in Subsection 2.3. The ability of the introduced modified measures to foresee Information Flow is discussed in Subsection 2.4.

2.1 Graph Theory Preliminaries and Social Network Analysis

A graph is a pair $G = \langle \mathcal{V}, \mathcal{E} \rangle$, where \mathcal{V} is a finite set of vertices, and \mathcal{E} is a set of edges. A graph G is

labelled when to each vertex or to each edge is associated a label, determined by Λ , a function that associated vertices and edges to the label sets (that are thus denoted by $\Delta(\mathcal{V})$ and $\Delta(\mathcal{E})$, or simply by Δ , meant to be the union of the above). We use the term node and the term vertex indifferently.

In this paper we concentrate ourselves onto undirected graphs, and delay the investigation on directed graphs to further work. We also assume that the graphs we deal with have no circular edge (although we do not assume them to be acyclic).

We name *incident edges of a node* v , the set of nodes that are on one extreme of an edge that has v as opposite extreme. We use the term *path* between two vertices v and w , to denote a sequence of nodes $p = \{v = v_1, v_2, \dots, v_n = w\}$ that connects v to w , so that every pair (v_i, v_{i+1}) is an edge of the graph. We call *length* of p the number n . We say that a vertex v is *between* two nodes v_1 and v_2 when v belongs to the shortest path from v_1 to v_2 .

The number n of vertices in a graph is named the *size* of the graph. Given a graph G of size n , if e is the number of edges in G , then the ratio $\delta = \frac{e}{n^2}$ is named the *density* of G . When a graph G_1 is denser than another graph G_2 , we also say that G_2 is sparser than G_1 .

To treat the notion of distance we employ notions derived from classic algorithmic graph theory, as widely discussed in (Eppstein, 1999). The distance between two vertices v_1 and v_2 , denoted by $\delta(v_1, v_2)$, in a graph, is the length of the shortest path connecting v_1 and v_2 .

We now introduce three basic measures of social network analysis, and discuss several flaws they exhibit.

Definition 1. A node v of a graph is said to have absolute degree centrality k when the number of edges incident to v is k .

From absolute degree centrality we can easily derive relative degree centrality, as the absolute degree centrality weighted by the size of the graph. In other terms, if a graph has n vertices, and a node v has absolute degree centrality k , the relative degree centrality of v is $\frac{k}{n}$.

The absolute degree centrality clearly depends on the size of the graph. In a graphs with n vertices absolute degree centrality can vary from 0 to $n - 1$, and therefore if a graph is larger than another, vertices may have greater degree centrality.

Relative degree centrality does not suffer of this flaw. On the other hand, if you consider two graphs of the same size, where one is denser than the other, in the denser one the relative degree centrality varies less than in the sparser one. A way to reduce this effect is

to consider a third variant of the degree centrality: the *graded degree centrality*. This variant considers the relative degree of each vertex and compares it to the corresponding values of the other vertices. A vertex v has a graded degree centrality $\delta = \frac{k}{n}$ where n is the size of the graph and k is the number of vertices in the graph with a relative degree centrality less than the relative degree centrality of v .

We now introduce the second measure of social network analysis.

Definition 2. A node v of a graph is said to have farness f such that:

$$f(v_i) = \sum_{k=1, k \neq i}^N \delta(v_i, v_k)$$

The closeness centrality of a vertex is the reciprocal of the farness of v :

$$c(v) = \frac{1}{f(v)}$$

Closeness centrality only works for connected graphs. Conceptually, however, two vertices that are not connected are as far as infinite, thus being closeness centrality of these equivalent to 0. This is clearly faulty. There have been several proposals to solve this aspect, mainly by means of techniques that are based upon weights. For the purpose of this research we assume that networks are connected.

For comparison purposes closeness centrality suffers of the same flaws of degree centrality. Larger graphs have potentially longer paths, and thus longer distances, consequently closeness centrality can vary significantly. In particular, since the distance between two vertices in our graphs is at least one (we exclude networks with self-cycles), and at most is the length of the longest possible connection path that is as long as the size (specifically $n - 1$), the farness of a vertex is obtained as the sum of the lengths $n - 1$ for each other vertex connected to it, that sums to $(n - 1) \cdot (n - 1)$. Therefore, the closeness varies from a minimum of $\frac{1}{n^2}$ to a maximum of $\frac{1}{n}$. To normalize this and make it independent from the size we cannot use fractions, as we did for degree centrality. A possible approach is to employ a logarithmic measure, and then normalize it by n . In other terms we can measure farness as the sum of the logarithmic measure of shortest paths, weighted by the size of the graph itself. Formally

$$f(v_i) = n \cdot \log\left(\sum_{k=1, k \neq i}^N \delta(v_i, v_k)\right)$$

The above introduced measure does not suffer of dependence on either size or density of the graph, and its reciprocal varies from 0 to 1. More precisely, it varies

from $\frac{1}{2}$ and 1, evidently. We can obtain a normalised measure by using

$$cc_n(v_i) = \frac{1}{f(v_i) - \frac{1}{2}} \cdot 2$$

The third measure we consider here is *betweenness centrality*. It is obtained as the number of pairs of vertices that are traversed by a path containing the measured vertex, or such that the vertex is between the elements of the pair.

Definition 3. The betweenness centrality of a vertex v is the number K of pairs in the graph for which v is between.

Analogously to the previous analysis about degree centrality we can note that large networks may exhibit wider spectrum of betweenness centrality than smaller ones, and conversely, less wide

2.2 Semantic Social Network Analysis

Social network Analysis starts from unlabelled and for the settings of this investigation indirected graphs 1.

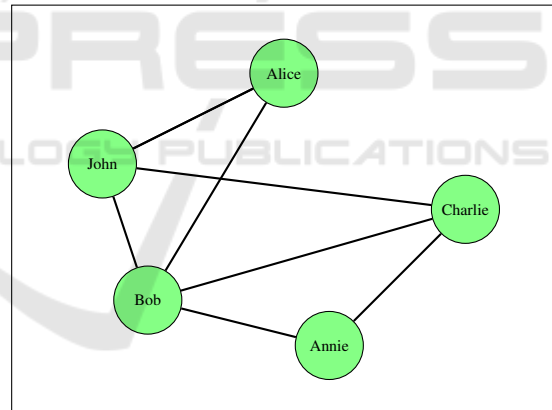


Figure 1: Representing a social network by an unlabelled indirected graph.

This, as discussed in Section 1 is unrealistic. An individual can be connected to another over a certain topic, but definitively disconnected over another topic. Moreover, the link between individuals can be independent from the shared topics. This concepts are known as *homophily* and have been discussed in Section 3.

If we measure centralities by means of unlabelled graphs, we may be rather misleading in defining the relevance (or importance, in some sense) of a vertex in the graph, since an individual can exhibit strong connections on one specific topic and weak ones on other topics, providing therefore a differentiated

degree centrality, in particular, and analogously for closeness and betweenness measures.

Consider two individuals *John* and *Alice* belonging to the same school but not sharing a hobby like music. A message regarding a class has a good probability of being forwarded from *John* to *Alice* while one regarding the hobby hasn't.

So we can say that *John* and *Alice* are connected over a topic *school* but are disconnected over topic *music*.

A method to provide this is to add a label to vertices in the graph with label corresponding to a measure of *depths* relative to a set of topics $T = \{t_1, \dots, t_k\}$. The label will be vector $D = \{d_{t_1}^1, \dots, d_{t_k}^n\}$ whose component $d_{t_j}^i$ is the *depth* over the topic t_j of the individual represented by vertex i as can be seen in Figure 2.

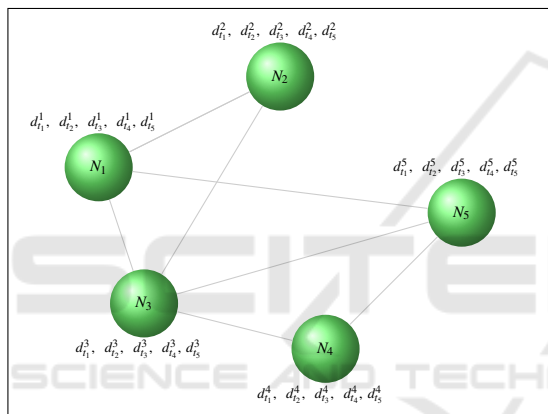


Figure 2: Introducing labels for depth in topics.

The goal of centrality measures is to provide a tool for foreseeing information flow. The basic assumption we make here is that when someone is not involved in a specific topic, it is rather unlikely that she promotes the flow of a piece of information through the vertex she occupies.

Considering the aforementioned two individuals *John* and *Alice*, their interests in the set of topics $T = \{gossip, music, sport, cooking, politics\}$ can be expressed as $D^a = (0.3, 0.5, 0.5, 0.0, 0.0)$ for *Alice* and $D^j = (0.9, 0.7, 0.0, 0.5, 0.0)$ for *John*, meaning that while both are interested in *gossip* and *music* and not interested in *politics*, *Alice* is keen to *sport* and *John* is not but he likes *cooking* while *Alice* does not, as is expressed in Figure 3.

An individual has *depth* on a certain topic measuring the degree of involvement on it; an individual has also an *activation threshold* which describes the inverse of the likelihood of that individual of becoming active when "hit" by a message. The notion introduced here is inspired by that used in scientific

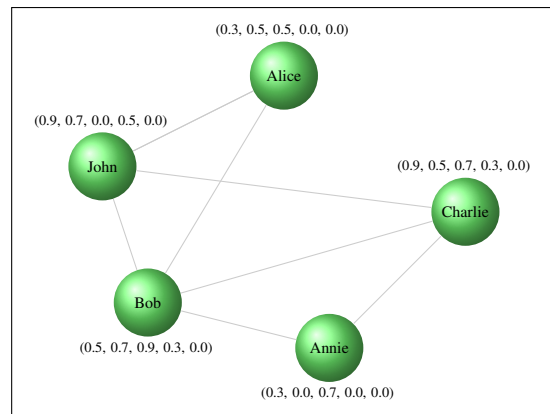


Figure 3: An Example of a social network with labels on topics *gossip, music, sport, cooking, politics*.

evaluation as proposed in (Tonta and Darvish, 2010).

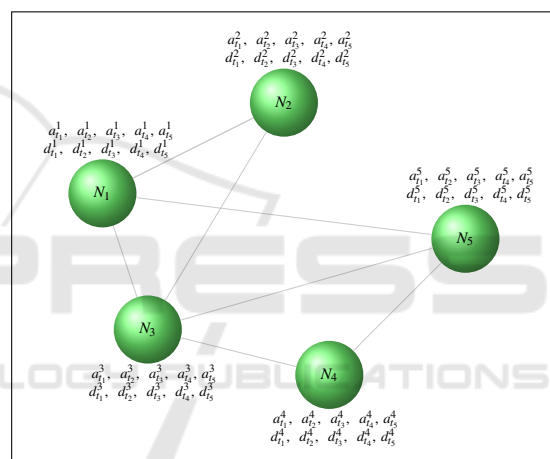


Figure 4: Activation thresholds: a means for representing sensitivity to information flows.

A method to express this is to add another label to vertices in the graph with label corresponding to a measure of *activation threshold* relative to a set of topics $T = \{t_1, \dots, t_k\}$ as can be seen in Figure 4.

The label will be vector $A = \{a_{t_1}^1, \dots, a_{t_k}^n\}$ whose component $a_{t_j}^i$ is the *activation threshold* over the topic t_j of the individual represented by vertex i , with a higher threshold meaning that it is less probable to see that individual i activated on topic t_j .

Considering the aforementioned two individuals *John* and *Alice*, their activation in the set of topics $T = \{gossip, music, sport, cooking, politics\}$ can be expressed as $A^a = (0.6, 0.3, 0.7, 0.5, 0.8)$ for *Alice* and $A^j = (0.3, 0.6, 0.3, 0.5, 0.9)$ for *John* meaning that while both are equally active in *music* and *cooking*, *Alice* is more reactive to *gossip* than *John* while he is probably forwarding more messages over topic *sport* than *Alice* and both will probably stop messages re-

garding *politics*, as is expressed in Figure 5.

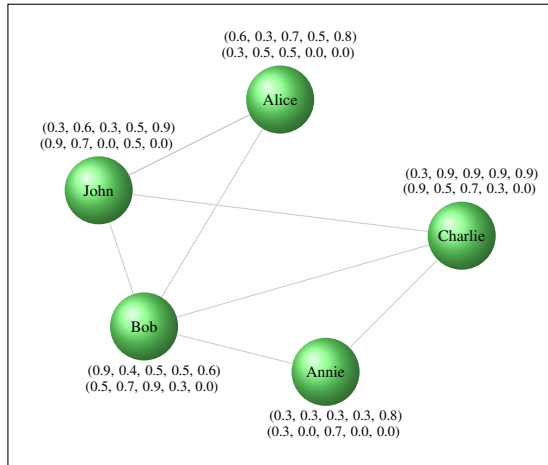


Figure 5: An example of a social network with topic depths and activation thresholds on topics *gossip*, *music*, *sport*, *cooking*, *politics*.

2.3 Computational Analysis of Semantic Social Network Analysis

Computation of degree centrality is straightforwardly obtained from the basic algorithm for Social Networks, that simply counts the number of incident edges of each vertex, and then derives the consequent computations for relative and graded variants. In Semantic Social Network Analysis, we sum the weights, instead of counting the incident edges.

The base for computing both closeness and betweenness centrality is the labelling of edges by the graph distance, meant as shortest path. We extend here the method known as *Floyd Warshall Algorithm*. In undirected unlabelled graphs, the algorithm computes just the incident edges to obtain the correct value of distances. We use the vectorial $\min()$ function both to initialize the distance matrix and to give a the value of the distance between two vertices in the core of the algorithm.

The distance of two vertices v_i, v_j with corresponding labels λ_i, λ_j will be $\delta[i, j] = \min(\lambda_i, \lambda_j)$.

The sum in *Floyd Warshall* will be substituted with the $\min()$ in our vectorial case.

As the *minimum* function is compliant with the triangle inequality the algorithm still holds.

It is easy to show that the method is polynomial and that the resulting labels provide an Euclidean distance between vertices in the graph.

Theorem 1. *FWSSNA terminates labelling of the input graph in $O(n^3)$ where n is the size of the graph.*

Algorithm 1: ALGORITHM FWSSNA. Given a graph G with vertices V , returns matrix of distances.

```

1: Input: a graph  $G$  of vertices  $V = v_1, \dots, v_n$  with
   corresponding labels  $\Lambda = \lambda_1, \dots, \lambda_n$ ;
2: Output: a matrix of distances  $\Delta$ ;
3:  $i \leftarrow 1, j \leftarrow 1$ ;
4:  $\Delta \leftarrow \emptyset$ ;
5: for  $i < |V|$  do
6:   for  $j < |V|$  do
7:      $\delta[i, j] \leftarrow \min(\lambda_i, \lambda_j)$ ;
8:   end for
9: end for
10:  $i \leftarrow 1, j \leftarrow 1, k \leftarrow 1$ ;
11: for  $k < |V|$  do
12:   for  $j < |V|$  do
13:     for  $i < |V|$  do
14:        $\delta[i, j] \leftarrow \min(\delta[i, j], \min(\delta[i, k], \delta[k, j]))$ ;
15:     end for
16:   end for
17: end for

```

Theorem 2. *FWSSNA returns a labelling that is a Euclidean distance between vertices.*

As a consequence, we can employ FWSSNA to provide semantic closeness centrality, and semantic betweenness centrality.

Closeness centrality is obtained by simply labelling vertices in the graph with the minimum of the computed distance from the node to the other nodes. Minimum is well known to respect Euclidean distance properties as well. Once we have executed the algorithm FWSSNA, every other node is connected to a given node, and thus we can consider the closeness computation as a second step. We say, in particular that an algorithm A_1 is in overhead computational cost C with respect to another algorithm A_2 , when the cost of executing A_1 is C , after A_2 has been executed as a preliminary step wrt A_1 .

Theorem 3. *Computing semantic closeness centrality is in overhead $O(n)$ wrt FWSSNA.*

Betweenness centrality can be obtained by modifying the method of FWSSNA with the introduction of a label reporting the shortest path, or better, the internal nodes of the path. If the shortest path connecting two vertices x and y is the sequence $x = x_1, x_2, \dots, x_n = y$, then the computed set of internal nodes is formed by $\{x_2, x_3, \dots, x_{n-1}\}$. To obtain, therefore, the correct betweenness it is sufficient to compute those labels (that is with no overhead wrt FWSSNA) and then, for a vertex v determine the number of shortest paths that pass through v by counting the number of appearances of v in the labels of the vertices. This can obviously

be done in a number of steps that is equivalent to the number of edges, giving the result below.

Theorem 4. *Computing semantic betweenness centrality is in overhead $O(n^2)$ wrt FWSSNA.*

2.4 Forecasting the Information Flow with Semantic Social Network Analysis

A message injected in a social network, i.e. sent to a person, has a higher probability of being forwarded if the person has a low *activation threshold* and if the content of the messages catches the attention of the person, or in other word if the information content of the message is higher or equal to the depth of the person on the topic.

In particular we assume that every information piece, that we name a *message* passing through the network, is associated with depth measures for the same topic vector employed to label the vertices of the network itself.

Therefore, we can say that the probability of a message m of topics t_1, t_2, \dots, t_k to be forwarded by an individual i can be expressed as a function of the depth on the topics, the information content over those topics of the message and the reactivity of the member. We denote by t_m the vector of topic depths of the message m , by d_i^i the vector of topic depths of the member i and by a_i^i the vector of activation threshold of the member i .

$$P(m, i) = \Psi(t_m, d_i^i, a_i^i)$$

If we consider a message of the topic *gossip* when this hits *John* or *Charlie* is less likely that it will be forwarded, as shown in Figure 6.

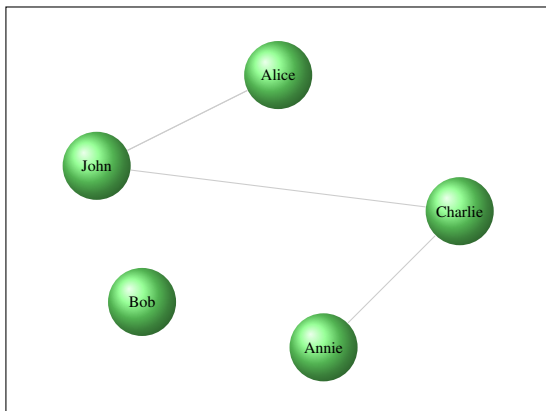


Figure 6: Our representation of a social network.

To introduce the notion we need, let us recall the idea of *probablistic complement*. When an event has

the probability p , the proablistic complement of p , namely $1 - p$ is the probability of the opposite event. When two events can occur, the probability of the occurrence of some of them is the probablistic complement of the conjunction of the opposite events. For instance, if two events have probabilities p_1 and p_2 , the probability of the occurrence of some of them is the probabilistic complement $1 - (1 - p_1)(1 - p_2)$, to be read as the probability that among all cases it would not be the case that none of the two events occur. For a set of values between 0 and 1 we $P = \{p_1, p_2, \dots, p_k\}$, we denote by P^C the probablistic complement

$$P^C = 1 - \prod_{i=1}^n 1 - p_i$$

The simplest function we can promote, when assuming that one message only deals with a single topic, is the direct product of the probabilistic complement of the activation threshold, the depth of the member in that topic and the depth of the message in the topic itself. When the topics involved are more than one, then the above is poor, and we need to work on the probabilistic complement for the depths of the involved topics.

$$P(m, i) = (t_m)^C \cdot (d_i^i)^C \cdot (a_i^i)^C$$

On a purely theoretical basis, it is clear that when two members of a network have a good activation threshold on a given topic, when a message on that topic reaches the vertex, and the message has a high depth in that topic, then the message has a good probability of being passed.

The ability of the above mentioned index to foresee correctly the information flow is under experimental evaluation and the first preliminary results are very encouraging.

3 RELATED WORK

The reference literature can be considered as articulated in three themes:

- Studies about implicit social links that exist among users of the internet (or of an internet application), or about enrichment of social web;
- Investigations of the semantics of social networks;
- Research about Social Network Analysis and relationships to semantic issues.

Regarding the first topic, we can look at methods for social link extraction, as discussed below, as one of the best structured investigations on the theme. This specific method for extracting social networks

from the web using similarity between collective contexts is proposed in (Alguliev et al., 2011). The authors construct three social networks on the same set of named entities. They use Jaccard, overlap and Normalized Google Distance (NGD) (Cilibrasi and Vitányi, 2007) coefficients to retrieve degree of closeness between entities. They show how actors may be assigned different relevance degrees and that actors having higher ranking results may be assigned lower ranks and inversely by choosing another measure to perform the ranking. In our perspective their work is solid, but lacks in one important aspect, the authors build homophily on the based of the contents.

This is a technique to *build* a network, and not an *analysis* of the network itself, as we do in this work. Suffering the same issue is the work of (Erétéo et al., 2009), where the authors present a new framework for applying Social Network Analysis to RDF representations of social data. In particular, the use of graph models underlying RDF and SPARQL extensions enables us to extract efficiently and to parameterize the classic Social Network Analysis features directly from these representations.

The main criticisms to the proposed approach lie on the fact that, as already shown in many practical cases, it makes a lot of difference, in terms of understanding of the structure of similarity between nodes, to know the relevance of the two nodes. In fact, similarity can be used, as done, for instance in (Cruz et al., 2014), for community detection, where members are related to each other based on their similarity in semantic terms. This is different in terms of relationship, with respect to measuring the relevance and study attractivity. Clearly, being interested in Football lies on liking it, but the community is formed around *authoritative* persons, for instance journalists.

A more practical research has been documented in (Yuan and Fei, 2010) where an application of semantic social networks and attraction theory to web based services is carried out.

The relation between trust and Social Network Analysis has been investigate in (Zhang et al., 2006) and specified as a means for understanding deeply the meaning of centrality and other measures as related to authority. The same concept is employed to provide a framework for the general interpretation of the logic bases of recommendation systems in (Davoodi et al., 2013).

The studies cited above all aim at discovering network links by means of mining techniques. On the other hand, the introduction of notions derived from semantic web into social networks is the core quest of many recent studies, including (Zhou et al., 2011). As a complete reference to the current literature about

meaning of social links, and relationships between social web and semantics, readers can look at (Leydesdorff, 2012).

More deeply, in (Mika, 2004) a direct and explicit comparison between social networks and the semantic web is carried out. This paper proposes a parallel between networked knowledge of members in a network and the basic notions of semantic web. The same issue is dealt with, with the specificity of a known technique, the semantic networks, in (Downes, 2005). More generally, the semantic web methods are employed for understanding the meaning of social networks as sharing platforms for common knowledge, in (Mika, 2005).

The idea of using Social Network Analysis as a means for forecasting the probability of a message to pass through a given member of the network itself is not novel at all. Base of our analysis is the criticisms to the roughness of the employed measures, criticisms that are not novel anyhow.

This has been dealt in two distinct ways: by using semantic methods for habilitating the forecast processes: in particular in (Zywica and Danowski, 2008), authors use semantic networks for foreseeing the behaviour in facebook.

On the other hand, many criticisms are applied to centrality measures ((Landherr et al., 2010), (Kang, 2007)). The main criticisms, that are met by the above mentioned investigations as well as by researches tending to correct the flaws of the general methods for centrality measures, and the measures themselves, lie on the weakness of the notion of similarity derived from the notion of centrality.

The above mentioned notion of similarity as derived from centrality measures, and its applications to the notion of reciprocity, a concept that has a crucial importance, for instance, in asymmetric social networks (Instagram, Twitter) are dealt with in (A. Ramachandra Rao, 1987). Authors show that centrality measures as used so far are unsuccessful in forecasting the information flows.

4 CONCLUSIONS

In this paper we investigated an extension to Social Network Analysis based upon the usage of a network model that includes the notion of topic. This leads to a further model that incorporates the notion of sensitivity, by means of a value, called *activation threshold* that is meant to denote the probability of a member of the network to be active in an information flow. Algorithms for computing extended notions of centrality are provided, and proved to be correct, complete and

computationally efficient. We provide examples that show that our approach can fruitfully solve few evident drawbacks of the general model, as applied to information flow forecast.

There are at least three different ways in which this investigation can be extended. First of all we aim at formalising a problem of dissemination of information pieces throughout a network. The problem can be formulated as follows: given a social network, a number k and a probability value p , select k members in such a way that the set of members reached by an information piece sent to the members in the selection and disseminated by them and the chains of members generated therefore, has a probability of being total (namely to cover the entire network) of at least p .

A second study investigates ways of providing reacher models of topics. In particular, we aim at investigating topics with sub-topics.

A third investigation will introduce the notion of *orientation*. For instance two persons can be both interested in football, but one may support Juventus F.C., whilst the other one may support A.C. Chievo Verona. These studies are taken into a track of research for viral marketing purposes, including methods to evaluate networks for advertisement, alerts, and emergencies.

REFERENCES

- A. Ramachandra Rao, S. B. (1987). Measures of reciprocity in a social network. *Sankhy: The Indian Journal of Statistics, Series A (1961-2002)*, 49(2):141–188.
- Alguliev, R., Aliguliyev, R., and Ganjaliyev, F. (2011). Investigation of the role of similarity measure and ranking algorithm in mining social networks. *Journal of Information Science*, 37(3):229–234.
- Cilibrasi, R. and Vitányi, P. M. B. (2007). The google similarity distance. *IEEE Trans. Knowl. Data Eng.*, 19(3):370–383.
- Cruz, J. D., Bothorel, C., and Poulet, F. (2014). Community detection and visualization in social networks: Integrating structural and semantic information. *ACM Trans. Intell. Syst. Technol.*, 5(1):11:1–11:26.
- Davoodi, E., Kianmehr, K., and Afsharchi, M. (2013). A semantic social network-based expert recommender system. *Applied Intelligence*, 39(1):1–13.
- Downes, S. (2005). Semantic networks and social networks. *The Learning Organization*, 12(5):411–417.
- Eppstein, D. (1999). Finding the k shortest paths. *SIAM J. Comput.*, 28(2):652–673.
- Érétéo, G., Gandon, F. L., Corby, O., and Buffa, M. (2009). Semantic social network analysis. *CoRR*, abs/0904.3701.
- Kang, S. M. (2007). A note on measures of similarity based on centrality. *Social Networks*, 29(1):137 – 142.
- Landherr, A., Friedl, B., and Heidemann, J. (2010). A critical review of centrality measures in social networks. *Business and Information Systems Engineering*, 2(6):371–385.
- Leydesdorff, L. (2012). Advances in science visualization: Social networks, semantic maps, and discursive knowledge. *CoRR*, abs/1206.3746.
- Mika, P. (2004). Social networks and the semantic web. In *Web Intelligence, 2004. WI 2004. Proceedings. IEEE/WIC/ACM International Conference on*, pages 285–291.
- Mika, P. (2005). Flink: Semantic web technology for the extraction and analysis of social networks. *Web Semantics: Science, Services and Agents on the World Wide Web*, 3(2):211 – 223. Selected Papers from the International Semantic Web Conference, 2004 ISWC, 20043rd. International Semantic Web Conference, 2004.
- Tonta, Y. and Darvish, H. R. (2010). Diffusion of latent semantic analysis as a research tool: A social network analysis approach. *Journal of Informetrics*, 4(2):166 – 174.
- Yuan, S.-T. and Fei, Y.-L. (2010). A synthesis of semantic social network and attraction theory for innovating community-based e-service. *Expert Systems with Applications*, 37(5):3588 – 3597.
- Zhang, Y., Chen, H., and Wu, Z. (2006). A social network-based trust model for the semantic web. In Yang, L., Jin, H., Ma, J., and Ungerer, T., editors, *Autonomous and Trusted Computing*, volume 4158 of *Lecture Notes in Computer Science*, pages 183–192. Springer Berlin Heidelberg.
- Zhou, L., Ding, L., and Finin, T. (2011). How is the semantic web evolving? a dynamic social network perspective. *Computers in Human Behavior*, 27(4):1294 – 1302. Social and Humanistic Computing for the Knowledge Society.
- Zywica, J. and Danowski, J. (2008). The faces of facebookers: Investigating social enhancement and social compensation hypotheses; predicting facebook and offline popularity from sociability and self-esteem, and mapping the meanings of popularity with semantic networks. *Journal of Computer-Mediated Communication*, 14(1):1–34.