Towards Semantic Profiling in Social Networks

Gregorio D'Agostino, Antonio De Nicola

Computing and Technological Infrastructure Lab,
ENEA: Italian National Agency for New Technologies, Energy and Sustainable Economic Development,
Via Anguillarese 301 - 00123, Rome, Italy
gregorio.dagostino@enea.it, antonio.denicola@enea.it

Abstract. The growing available implicit knowledge about people in social networks is fostering a new generation of advanced services centered on the actual needs of persons. Precondition to them is to provide an explicit formal specification of what is known about people taking part in the social network. This specification should be in the terms of a shared conceptualization of one or more domains of interest related to the social network scope. Here we propose a method to provide semantic profiles of people taking part in a social network concerning a specific domain of interest. Our method is an inference process based on the topological structure of the relationships between people, on existing explicit interests, and on the assumption that each person has own beliefs and is partly influenced by friends. In particular, the introduced approach allows estimating a semantic profile of a newcomer, i.e., a new person joining a social network, and how it will evolve over time.

Keywords. Social Networks; Semantic Social Networks; Semantic Profiling.

1 Introduction

Social Networks (*SN*)s collect a huge amount of growing implicit knowledge about people and domains of interest. General purpose *SN*s (e.g., Facebook, Twitter) collect information on several domains (e.g., music, movies, literature, travels) whereas domain specific *SN*s, e.g., anobii and LinkedIn, collect information on specific topics (e.g., books for anoobi and job and careers for LinkedIn).

Organizing and managing such knowledge may allow definition of a new generation of services focused on specific users' needs. Examples are services devoted to a better organization of human competencies in the enterprise sector; marketing services to promote new products and services; and alert services to provide useful or necessary information to citizens, private enterprises, institutional operators, decision makers, and possibly other entities.

Precondition to these services is harnessing implicit knowledge at the basis of a *SN* (i.e., *SN knowledge*) and to provide it in an explicit form. *SN knowledge* concerns people, i.e., their relationships, their attitudes, their interests, their needs, and their activities, and, for domain specific *SN*s, the addressed *SN* sector.

In this paper we propose a first step to provide an explicit specification of *SN* knowledge. We focus on providing a semantic specification of interests (i.e., semantic profile) related to people taking part in a *SN*. In particular, we propose a method to estimate the semantic profile of a new person joining a social network and its evolution during time. Our method is based on the topological structure of people relationships, on existing explicit interests and on the assumption that each person has her/his own beliefs and is partly influenced by her/his friends.

The rest of the paper is organized as follows. First we introduce a formal definition of a social network. Then we describe an incremental approach to specify a domain of interest and we present our definition of the *interest endowed network* as a bipartite graph. Then we propose our method to estimate the evolution of a *semantic profile* and a consequent estimated welcome profile related to a *newcomer*. Finally, related work in the area and conclusions end this work.

2 Social Network

A social network consists of a community of people linked together with some kind of relationships (e.g., friendship, coauthorship, working together with). It can be represented as a directed graph $SoN = (P,F)$ (Figure 1), where the set P of nodes p_i represents people $P=\{p_1, p_2, \ldots, p_{|P|}\}$, and the set F of links $f_{j,k}$ represents friendship relationships between person j and person k as ordered pairs of people $F=\{f_{1,1}, f_{1,2}, \ldots, f_{|F|}\}$.

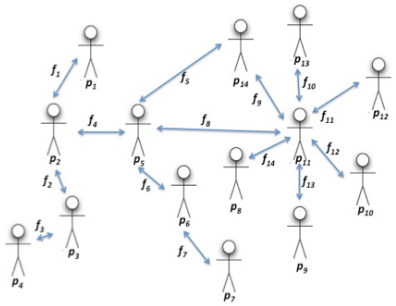

Fig. 1. An Excerpt from a Social Network where (p_i)s represent people and ($f_{j,k}$)s represent relationships

If the relationship between people is symmetric the graph becomes "undirected". There are two ways to account for the symmetric case: one consists into doubling the links always including two ordered pairs in both orders; while the other consists into associating a link to a unordered pair. To account for the most general case, the first approach has been preferred. Most of the considerations presented in the paper apply to both directed and undirected graphs (i.e., symmetric or non symmetric relationship); when results will depend on such characteristics it will be outlined in the text. To keep the problem on ground it is worth noting that while Facebook relationship is intrinsically symmetric, the same does not apply to Twitter.

3 Domain of Interest Representation

According to the modeler's purpose, a domain of interest can be conceptualized at different levels of abstraction and represented by means of incremental levels of details [1]. First, the **domain lexicon** specifies the terminology used in the domain of interest. Then the **domain glossary** allows specifying the definitions corresponding to the terms. The **semantic network** allows specifying ontological relationships. Finally, the **ontology** provides axioms and the final formalization (e.g., by using the OWL language [2]).

In the following, as a running example, we consider a community of people working in the sector of critical infrastructures protection and sharing information related to such domain.

A **domain lexicon** $\boldsymbol{DL}=\{t_1, t_2, \ldots, t_{|DL|}\}$ is defined as the set of terms (t_i) used to characterize a domain of interest.

In the example of the critical infrastructure sector (see above), an excerpt of the lexicon is reported in the Table 1.

Table 1. Excerpt from the Domain Lexicon concerning Critical Infrastructure Protection

Telecommunications, *Transportation*, *SCADA*, *Mobile Telecommunications*, *Fixed Telecommunications*, *Rail Transportation*, *Aviation*, *Maritime Transportation*, *Road Transportation*, *Water*, *Gas*, *Electricity*

A **domain glossary** $\boldsymbol{G}=\{g_1,\ldots,g_{|G|}\}$ is defined as the finite set of terms belonging to a domain lexicon *DL* paired with the corresponding definitions. The pair *term* and *definition* is defined as the *glossary entry* g_i, where

$$g_i = \{(t_i, def_i) \mid t_i \in DL \wedge def_i \in DEF\},$$

and *DEF* is the set of the definitions of the domain lexicon terms.

In the example of the critical infrastructure protection sector, an excerpt of the domain glossary is reported in the Table 2.

Table 2. An Excerpt from the Domain Glossary concerning Critical Infrastructure Protection.

Term	Definition
Aviation	The operation of aircraft to provide transportation. [From *WordNet*- http://wordnet.princeton.edu]
Maritime Transportation	Maritime transportation is a means of conveyance of passengers and goods by means of watercraft. [Inspired by *Wikipedia*- http:en.wikipedia.org]
Road Transportation	Road transportation is transport on roads of passengers or goods. [From *Wikipedia*]
SCADA	SCADA (supervisory control and data acquisition) generally refers to industrial control systems (ICS): computer systems that monitor and control industrial, infrastructure, or facility-based processes. [From *Wikipedia*]
Transportation	A facility consisting of the means and equipment necessary for the movement of passengers or goods. [From *WordNet*]

A **semantic network *SeN*** can be considered as a simplified version of an ontology since it consists of a set of concepts and a set of ontological relationships between them. Consequently, *SeN=(C,R)*, where the set *C* of nodes c_i includes concepts, $C=\{c_1,c_2, \ldots, c_{|C|}\}$, and the set *R* includes the relationships r_k between concept *i* and concept *j*, where

$$R = \left\{r_1, r_2, ..., r_{|R|}\right\} = \left\{(c_i, c_j)\right\} \subseteq C \times C .$$

An example of relationship is similarity (***sim***) [3], representing how a concept is similar to another concept.

An **ontology *O*** is a formal specification of a shared conceptualization [4] [5]. An ontology consists of a set of concepts, a set of relationships between them, and a set of axioms. Consequently, given a finite set of concepts, a finite set R of relationships established between concepts, and a finite set of semantic axioms *Ax*, an ontology is defined as a triple *O=(C,R,Ax)*, where $Ax=\{boolExp_1, boolExp_2, \ldots, boolExp_{|Ax|}\}$.

The Table 3 reports an example of axiom in the critical infrastructure protection sector using Horn Clauses notation [6].

Table 3. Example of Ontology Axiom in the Critical Infrastructure Sector

Axiom. Radioactive materials are transported only by means of special transportation.
$Good(\_x) :- RadioactiveMaterial(\_x)$
$Transportation(\_y) :- SpecialTransportation(\_y)$
$SpecialTransportation(\_y) :- RadioactiveMaterial(\_x), TransportedBy(x,y)$

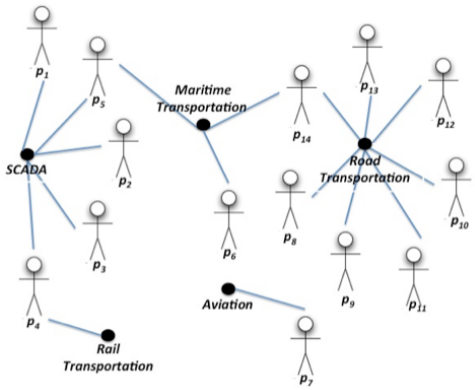

Fig. 2. A pictorial representation of an excerpt from the Interest Endowed Network concerning Critical Infrastructures

An **Interest endowed Network *IeN*** represents the interests of a community of linked people. It can be represented as a bipartite graph consisting of a set of nodes ***N*** partitioned in two groups, one representing people and the other the concepts from a domain of interest conceptualization, and a set of relationships ***I*** representing links between people and concepts only. Consequently, *IEN=(P,C,I)*, where $I=\{i_1, i_2, \ldots, i_{|I|}\}$, and

$$i_i = (p_j, c_k) \text{ with } p_j \in P \text{ and } c_k \in C .$$

An excerpt from the *interest endowed network* in the critical infrastructures protection example is reported in the Figure 2.

4 Semantic Social Network

A **semantic social network *SSN*** represents, at the same time, the domain of interest, the interests of a community of people, and the relationships among such people. It can be specified as the union of a semantic network (or an ontology), a social network, and an *interest endowed network*. Consequently, *SSN=(P,F,C,R,I)*, where *P* represents the set of people; *F* represents the relationships among people; *C* represents the set of concepts; *R* represents the relationships between concepts; and *I* represents the interest of people on concepts.

4.1 Semantic Profiling

We define *semantic profiling* as the process to associate interests (i.e., concepts of a domain ontology), related to a specific domain of interest, to a person, in other words inferring links belonging to the *i-th* person. The set of interests characterizing a person p_i is defined as her/his *semantic profile* S_{p_i}:

$$S_{p_i} = \left\{ c_k : \left(p_i, c_k \right) \in I \right\}$$

where

$$c_k \in C \text{ , } k \in (0, \left| C \right|) \text{ and } p_i \in P \text{ .}$$

In other words the *semantic profile* of a person is a subset of the *interest endowed network*.

The method we propose here allows building a semantic profile of a person p_i, i.e., *newcomer*, joining an existing *semantic social network SSN*, representing a community of people, a domain of interest, and their interest in such domain. The precondition to apply this method is to know both the topological structure of the social network and the semantic profiles of people belonging to it, that is all the *SSN*. We are looking for a kind of *welcome profiling*.

For the sake of simplicity, here, we do not take into account the linked structure of the ontology and we consider just the concepts per se.

Furthermore, for the *newcomer* p_i, we define the "*likelihood*" $L_{p_i}(c_k)$ as the probability to be interested in the concept c_k.

4.2 Semantic Profile Evolution

Here we estimate how, given a person p_i, the probability for he or she to be interested in a concept c_k evolves during time:

$$L_{p_i}(c_k \mid t + \Delta t) = (1 - x_i\left(c_k\right)) \cdot L_{p_i}(c_k \mid t) + \frac{1}{\left| N_{p_i} \right|} \cdot \sum_{p_j \in N_{p_i}} x_{ij}(c_k) \cdot L_{p_j}(c_k \mid t)$$

where $x_{ij}(c_k)$ is a positive number representing the attitude of a person p_i to be influenced by his or her neighbors (p_j) with regard to the concept c_k and x_i represents the sum over all j's that is the total influenceability:

$$x_i(c_k) \overset{def}{=} \frac{\sum_{p_j \in N_{p_i}} x_{ij}(c_k)}{\left|N_{p_i}\right|}.$$

In order to predict the probability of exhibiting an interest in a future time $L_{p_i}(c_k|t+\Delta t)$, we assumed that a person has her/his own beliefs; this assumption originates the positive term $L_{p_i}(c_k|t)$. As mentioned, we also assume that a person is partly influenced by people he or she interacts with; this originates both the negative term, $-x_i(c_k).L_{p_i}(c_k|t)$ (representing the negative influence of the friends on the term c_k), and the positive following term (representing the positive influence of the friends on the term c_k): $\frac{1}{\left|N_{p_i}\right|} \cdot \sum_{p_j \in N_{p_i}} x_{ij}(c_k) \cdot L_{p_j}(c_k \mid t)$.

If we assume that the interest in any concept experiences the same influence from friends, the x_i value does not depend on concept.

When x_i=0 we are dealing with a person that is not influenced by other people and always keeps his or her own opinions. On the other hand, when x_i=1 the person is totally bailed out by friends.

To extract a semantic profile from the L's, we may assume that a person p_i is interested in a concept c_k if $L_{p_i}(c_k|t)>L_t$, where L_t is a predefined *threshold*. Consequently:

$$S_{p_i}(t) = \left\{c_k : (p_i, c_k) \in I \wedge L_{p_i}(c_k \mid t) > L_t\right\}.$$

4.3 Welcome Profiling

Once a person joins a social network, it may happen that he or she does not have time or does not desire to express formally interests; therefore the only information available is contained in her/his links with other people. For this reason, in order to build a *welcome profile*, we omit the unknown part concerning past personal interests in the formula to estimate $L_{p_i}(c_k)$ and we just consider the part concerning the influence of the group of people in his or her neighborhood (assuming there is at least one person in the group and, consequently, $\left|N_{p_i}\right| \neq 0$):

$$L_{p_i}(c_k \mid t + \Delta t) = \frac{1}{\left|N_{p_i}\right|} \cdot \sum_{p_j \in N_{p_i}} x_{ij}\left(c_k\right) \cdot L_{p_j}(c_k \mid t).$$

In practice, we are estimating the interest of a *newcomer* as just the average of the interests of his or her friends. Alternatively, we can assume the *newcomer* to have an

"a priori" average interest $L^0_{p_i}(c_k)$ and provide a *welcome profile* based on its expected evolution:

$$L_{p_i}(c_k \mid \Delta t) = (1 - x(c_k)) \cdot L^0_{p_i}(c_k) + \frac{1}{|N_{p_i}|} \cdot \sum_{p_j \in N_{p_i}} x_{ij}(c_k) \cdot L_{p_j}(c_k \mid t = 0).$$

Please note that we consider here $t=0$ as the starting time of the observation period of the social network.

Furthermore, as a first approximation, we can assume that the friend's contribution is unitary, $L_{p_j}(c_k|t)=1$, if her/his interest on the concept c_k is declared, or null otherwise, $L_{p_j}(c_k|t)=0$. The "a priori" interest L^0 may result from analysis of some larger social network or can be provided by other modeling of the system.

5 Related Work

Business and social experts [7] recognize the growing importance of the new generation of web sites both for their implications on marketing and for the impact on the societal and political life.

A list of new generation of methodological approaches fostering the potential of social networks is presented in [8]. Examples of these methodologies are novel approaches to study viral marketing [9] and mechanism to govern social influence [10].

Merging social network analysis and semantics-based methods is a new research approach recently used with promising results [11] [12] [13] [14]. With these works we share the use of a conceptual representation of a domain of interest in the social network context.

The importance of semantic profiling, that is the main topic of our paper, is recognized by a growing number of research papers [15] [16]. Semantic profiling is mainly used to support information retrieval systems. The existing approaches are based on extracting concepts from a domain ontology and by means of a set of existing documents. While their approach employs a domain ontology, we propose a paradigm shift to estimate semantic profiles including the topology of the human relationships of the social network.

6 Conclusions

The growing implicit knowledge available in social networks opens several opportunities to develop a new generation of advanced services tailored to user characteristics. Semantic profiling is the process to add a semantics-based description of a person. The former is a precondition to foster the desired intelligent services.

In this paper we have proposed a novel approach to model the evolution of semantic profile in a social network and to provide a *newcomer* in a social network with an inferred *welcome profile*. Our approach is based on the relationships' topology of the social network and on assumption that each person carries his or her own beliefs whilst being partly influenced by friends.

Finally, the present paper is a preliminary theoretical work that is currently under ongoing validation through a large set of data related to a widespread social network. As possible applications, we envisage usages in the competencies management and in the marketing sectors.

References

1. De Nicola A., Missikoff M., Navigli R., A software engineering approach to ontology building, Information Systems, Volume 34, Issue 2, pp. 258-275, April 2009.
2. Antoniou G., van Harmelen F., Web ontology language: Owl, in: S. Staab, R. Studer (Eds.), Handbook on Ontologies, Springer, pp. 67–92, 2004.
3. Lin D., An Information-Theoretic Definition of Similarity, In Proceedings of the 15th International Conference on Machine Learning, Morgan Kaufmann, pp. 296--304, 1998.
4. Gruber T. R., A translation approach to portable ontology specification, Knowl. Acquis. 5, pp 199–220, 1993.
5. Borst W. N., Construction of engineering ontologies for knowledge sharing and reuse, University of Twente, Enschede, The Netherlands, 1997.
6. Lloyd JW. Foundations of Logic Programming (2nd edn). Springer-Verlag: Berlin, 1987.
7. D. Tapscott and A. D. Williams, Wikinomics: How Mass Collaboration Changes Everything, New York, Penguin, 2007.
8. Staab S., Domingos P., Mika P., Golbeck J., Ding L., Finin T., Joshi A., Nowak A., and Vallacher R. R., Social Networks Applied. IEEE Intelligent Systems 20, 1, January 2005.
9. Domingos P., Mining Social Networks for Viral Marketing, IEEE Intelligent Systems, Vol. 20, No. 1., pp. 80-82, 2005.
10. Novak A., Vallacher R. R., Information and Influence in the Construction of Shared Reality, IEEE Intelligent Systems, Vol. 20, No. 1., pp. 90-93, 2005.
11. Jung J., Euzenat J., Towards Semantic Social Networks. In Proc. of ESWC 2007, LNCS 4519, pp. 267–280, 2007.
12. Mika P., Social Networks and the Semantic Web: The Next Challenge, IEEE Intelligent Systems, Vol. 20, No. 1., pp. 82-84, 2005.
13. Bojrs U., Breslin J. G., Finn A., and Decker S.. Using the Semantic Web for linking and reusing data across Web 2.0 communities. Web Semant. 6, 1, February 2008.
14. Kinsella S., Passant A., et al.. The Future of Social Web Sites: Sharing Data and Trusted Applications with Semantics, Advances in Computers, Vol. 76, Is. 09, pp. 121-175, 2009.
15. Bhatt M., Rahayu W., Prakash Soni S., Wouters C., Ontology driven semantic profiling and retrieval in medical information systems, Web Semantics: Science, Services and Agents on the World Wide Web, Volume 7, Issue 4, pp. 317-331, December 2009.
16. Zhuhadar L., Nasraoui O., and Wyatt R.. Dual representation of the semantic user profile for personalized web search in an evolving domain. In Proc. of the AAAI 2009 Spring Symposium on Social Semantic Web, Where Web 2.0 meets Web 3.0, pp. 84–89, 2009.