A State of the Art on Social Network Analysis and its Applications on a Semantic Web

Guillaume Erétéo[1], Michel Buffa[2], Fabien Gandon[3], Patrick Grohan [1], Mylène Leitzelman[4], Peter Sander[2]

[1] Orange Labs
{guillaume.ereteo, patrick.grohan}@orange-ftgroup.com
[2] KEWI, I3S, Université of Nice, France
buffa@unice.fr, sander@polytech.unice.fr
[3] EDELWEISS, INRIA Sophia-Antipolis, France
fabien.gandon@sophia.inria.fr
[4] Telecom ParisTech, Sophia Antipolis
mylene.leitzelman@telecom-paristech.fr

Abstract. The increasingly popular web 2.0 sites provide the largest social network ever analyzed - users are now considered as plain web resources. Some researchers apply classical methods of social network analysis to such networks; others provide models to leverage the semantics of their representation. We present a state of the art of these two approaches and propose an architecture to merge and exploit the best features of each.

Keywords: social network analysis, semantic web.

1 Introduction

Research conducted on large social networks has principally concerned interviews, enterprise human resources mining, or scientific publications references [17] [39] [51] [53]. However, since its birth in 1992, the web has provided many ways of interaction between people [9], revealing social network structures [54], a phenomenon amplified by the emergence of the web 2.0 [28]. Social networks have been extracted from email communications [52], hyperlink structure of home pages [1], co-occurrence of names [31] [39] [37] [30], and from web 2.0 applications [39]. Dedicated online platforms such as Facebook and Myspace now provide huge amounts of structured social network data to exploit.

In the first part of this paper we recall some classical work from Social Network Analysis (SNA), in particular we detail the popular models used by researchers for representing and visualizing social networks. Definitions of the features that characterize these networks will be presented as well as the corresponding algorithms. In a second part, we discuss the use of semantic web languages and technologies to represent social networks. Finally, we will show that these enhanced representations are a step forward to what we call the “semantic social network analysis” of online interactions.

2 Social Network Analysis

The first representations of social network were sociograms [38] where people are represented by points and relationships by lines connecting them. Much research has been conducted on SNA based on this graph-based view using graph theory [51] [53]. Among important results is the identification of sociometric features that characterize a network. The **density** indicates the cohesion of the network. The **centrality** highlights the most important actors of the network and three definitions have been proposed [19]. The **degree centrality** considers nodes with the higher degrees (number of adjacent edges). The **closeness centrality** is based on the average length of the paths (number of edges) linking a node to others and reveals the capacity of a node to be reached. The **betweenness centrality** focuses on the capacity of a node to be an intermediary between any two other nodes. A network is highly dependent on actors with high betweenness centrality and these actors have a strategic advantage due to their position as intermediaries and brokers [10] [29] [12]. Its exact computation is time consuming, several algorithms tackle this problem [20] [42] [35] [7] with a minimum time complexity of *O(n.m)* - n is the number of vertices and m the number of edges. To deal with large networks, approximating algorithms [49] [8] [5] [22] and parallel algorithms [4] [50] have been proposed.

Community detection helps understanding the global structure of a network and the distribution of actors and activities [51]. Moreover, the community structure influences the way information is shared and the way actors behave [10] [11] [12]. Scott [51] gives three graph patterns that correspond to cohesive subgroups of actors playing an important role in community detection: **components (**isolated connected subgraphs), **cliques** (complete subgraphs), and **cycles** (paths returning to their point of departure)**.** Alternative definitions have also been proposed such as **n-clique, n-clan** and **k-plex** that extend these initial concepts. Community detection algorithms are decomposed into two categories, either hierarchical or based on heuristics [44] [24] [15]. Two strategies are used in hierarchical algorithms: the divisive algorithms consider the whole network and divide it iteratively into sub communities [23] [56] [21] [49] and the agglomerative algorithms group nodes into larger and larger communities [16] [58]. Other algorithms are based on heuristics such as random walk, analogies to electrical networks or formula optimization [45] [57] [48].

Social network graphs hold specific patterns that can be used to characterize them [43] and accelerate algorithms. According to the small world effect [40], the order of the shortest path between two actors in a social network of size *n* is *log(n)*. Social networks have an important clustering tendency and a community structure**,** furthermore, the degree distribution follows a **power law** [43].

These graph-based representations are only concerned with syntax – they all lack semantics, and have an especially poor exploitation of the types of relations. We will now see how recently online social networks started to be represented with rich structured data incorporating semantics.

3 Semantic Web Representation of Online Social Networks

Semantic web frameworks provide a graph model (RDF[1]), a query language (SPARQL[1]) and type and definition systems (RDFS[1] and OWL[1]) to represent and exchange knowledge online. These frameworks provide a whole new way of capturing social networks in much richer structures than raw graphs.

Several ontologies can be used to represent social networks. The most popular is **FOAF**[2], used for describing people, their relationships and their activity. A large set of properties is dedicated to the definition of a user profile: "family name", "nick", "interest", etc. The "knows" property is used to connect people and to build a social network. Other properties are available to describe web usages: online account, weblog, memberships, etc. The properties defined in the **RELATIONSHIP**[3] ontology specialize the "knows" property of FOAF to type relationships in a social network more precisely (familial, friendship or professional relationships). For instance the relation "livesWith" specializes the relation "knows". The primitives of the **SIOC**[4] ontology specialize "OnlineAccount" and "HasOnlineAccount" from FOAF in order to model the interactions and resources manipulated by social web applications; SIOC defines concepts such as posts in forums, blogs, etc. Researchers [6] have shown that SIOC and the other ontologies presented can be used and extended for linking reuse scenarios and data from web 2.0 community sites.

In parallel, web 2.0 applications made social tagging popular: users tag resources of the web (pictures, video, blog posts etc.) The set of tags forms a folksonomy that can be seen as a shared vocabulary that is both originated by, and familiar to, its primary users [39]. Ontologies have been designed to capture and exploit the activities of **social tagging** [27] [33] [46] while researchers have attempted to bridge folksonomies and ontologies to leverage the semantics of tags (see overview in [36]). Once they are typed and structured, the relations between the tags and between the tags and the users are also a new source of social networks.

A lighter way to add semantics to the representation of persons and usages of the web is to use microformats[5] [2] [32]. Some microformats can be used for describing user profiles, including resources and social networks. For example, hCard and hResume microformats describe a person (name, email, address, personal resume etc.) and XFN (XTML Friends Network) is useful for describing relationships.

Millions of FOAF profiles [26] are now published on the web, due to the adoption of this ontology by web 2.0 platforms with large audiences (www.livejournal.net, www.tribe.net). The acquaintance and expertise networks respectively formed by the properties "foaf:knows" and "foaf:interest" reflect real social networks [18]. As a consequence, researchers have applied classical SNA methods to FOAF [47] [25] [26]. Much as today there is only one community of email users (anyone can mail anyone), the adoption of standardized ontologies for non-specialist online social networks will lead to increasing interoperability between them and to the need for uniform tools to analyse and manage them.

[1] Semantic Web, W3C, http://www.w3.org/2001/sw/
[2] http://www.foaf-project.org/
[3] http://vocab.org/relationship/
[4] http://sioc-project.org/
[5] http://microformats.org/

4 Toward a Semantic Social Network Analysis

The online availability of social network data in different formats, the availability of associated semantic models and the graph structure of the RDF language are leading to a new way of analysing social networks. Current algorithms that are applied to SNA are based on graph pattern detection and use very little semantics. The semantics of sociometric patterns that are measured are never taken into account due to the lack of semantics of the representation of the analysed networks. As an example, community detection algorithms are based on graph structure characteristics of social networks but none is based on a sociological definition of community [55] and types of relations are under-exploited. Ontologies were designed to describe particular communities [41] and can be an interesting way to extend community detection among semantically described social networks.

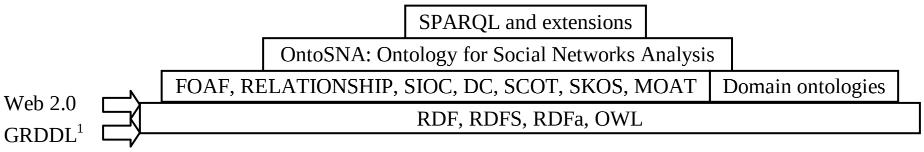

Fig 1: A semantic social network analysis architecture

We are designing an architecture (fig. 1) for a new tool to analyse online social networks. This tool explores RDF-based annotations describing profiles and interactions of users through social applications, using the conceptual vocabulary of previously mentioned ontologies and domain specific ontologies. An ontology, called **OntoSNA** (Ontology of Social Network Analysis), describes general sociometric features and their links to social RDF data. Recently, SPARQL extensions have been proposed for enhancing the RDF graph queries [3] [34] and have been implemented in the search engine CORESE [13] [14]. These extensions enable us to extract paths in RDF graphs by specifying multiple criteria such as the type of the properties involved in the path with regular expressions, or edge directions or constraints on the vertices that paths go through. We reuse these extensions and propose new ones dedicated to SNA in order to make easier the analysis of RDF-based representations of social networks. With such a tool, we can focus or parameterize the analysis specifying types of resources or properties to be considered, and extend classical algorithms with semantic features expressed in SPARQL and based on sociological definitions.

```
select count(?y) as ?cdegree
{
  {?y foaf:knows ?x}
  UNION
  {?x foaf:knows ?y}
}
group by ?x
```

```
select count(?y) as ?cdegree
{
  {?y relationship:worksWith ?x}
  UNION
  {?x relationship:worksWith ?y}
}
group by ?x
```

Table 1: SPARQL queries that extract the degree centrality of actors linked by the property foaf:knows and its specialization "relationship:worksWith".

5 Conclusion

We presented a state of the art on SNA and showed that while this research domain has been exploited for a long time, its application to the web opened new perspectives. The web is now a major medium of communication in our society and, as a consequence, an element of our socialization. The huge number of human interactions through web 2.0 platforms reveal real social networks, and understanding their life cycles is one of the challenges of knowledge sciences. Semantic models of these interactions are well developed and some are now massively integrated into online social applications. The semantic leverage of social data in a machine readable format opens a new way for SNA and the enhancement of online social experiences. We proposed an approach to go toward semantic-aware social network analysis.

References

1. Adamic, L. A., Adar, E.: Friends and Neighbors on the web. Social Networks 25: 211-230 (2003)
2. Adida, B.: hGRDDL: Bridging micorformats and RDFa. J. Web Sem. 6(1): 61-69 (2008)
3. Anyanwu, M., Maduko, A., Sheth, A.: SPARQL2L: Towards Support for Subgraph Extraction Queries in RDF Databases, Proc. WWW2007. (2007)
4. Bader, D. A., Madduri, K.: Parallel algorithms for evaluating centrality in real-world networks, ICPP2006 (2006).
5. Bader, D. A., Kintali, S., Madduri, K., Mihail, M.: Approximating betweenness centrality. WAW2007 (2007)
6. Bojars, U., Breslin, J.G, Finn, A., Decker, S.: "Using the Semantic Web for linking and reusing data across Web 2.0 communities", 2008, J. Web Sem. 6: 21-28.
7. Brandes, U.: A faster algorithm for betweenness centrality. J. Math. Socio. 25(2): 163-177 (2001).
8. Brandes, U., Pich, C.: Centrality estimation in large networks. J. Bifurcation and Chaos in Applied Sciences and Engineering 17(7), 2303–2318 (2007).
9. Buffa, M.: Du Web aux wikis : une histoire des outils collaboratifs. http://interstices.info, online journal, issue of 23/05/08 (2008)
10. Burt, RS: Structural Holes. Cambridge University Press, New York (1992)
11. Burt, R.S.: Structural Holes versus Network Closure as Social Capital. Lin, N., Cook K., Burt, R.S.: *Social Capital: Theory and Research.* Aldine de Gruyter: 31-56 (2001)
12. Burt, R.S.: Structural Holes and Good Ideas. American J. of Sociology 100(2): 339-399 (2004)
13. Corby, C., Dieng-Kuntz, R., Faron-Zucker, C.: Querying the semantic web with the corese search engine. ECAI/PAIS2004 (2004)
14. Corby, C.: http://www-sop.inria.fr/edelweiss/software/corese/v2_4_0/manual/next.php (2008)
15. Danon, L., Diaz-Guilera, A., Duch, J., Arenas,A. Comparing community structure identification. J. Stat. Mech. P09008 (2005).
16. Donetti, L., Munoz, M. A.: Detecting communities: a new systematic and efficient algorithm. J. Stat. Mech. P10012 (2004).
17. Evans, J.A.: Electronic publishing and the narrowing of science and scholarship. Science 5887 (321). (2008)
18. Finin, T., Ding, L., Zou, L.: Social networking on the semantic web. J. Learning organization 5 (12): 418-435. (2005)
19. Freeman, L.C.: Centrality in Social Networks: I. Conceptual Clarification. Social Networks 1 (1979)
20. Freeman, L. C., Borgatti, S. P.: Centrality in valued graphs: A mesure of betweenness based on network flow. Social Networks 13: 141–154 (1991)
21. Fortunato, S., Latora, V., Marchiori, M.: Method to find community structures based on information centrality. Phy. Rev. E 70(5): 056104 (2004)
22. Geisberg, R., Sanders P., Scultes, D.: Better approximation of betweenness centrality. ALENEX08 (2008).

23. Girvan, M., Newman, M. E. J.: Community structure in social and biological networks, Proc. Natl. Acad. Sci. USA 99(12) (2002)
24. Girvan, M., Newman, M. E. J.: Finding and evaluating community structure in networks. Phys. Rev. E, 69:026113 (2004)
25. Golbeck, J., Parsia, B., Hendler, J.: Trust network on the semantic web. CIA03(2003)
26. Goldbeck, J., Rothstein, M.: Linking social Networks on the web with FOAF AAA08 (2008)
27. Gruber, T.: Ontology of folksonomy: A mash-up of apples and oranges. MTSR2005 (2005)
28. Hendler, J., Goldbeck, J.: Metcalfe's law, web 2.0 and the Semantic Web. J. Web Sem. 6(1):14-20, 2008
29. Holme, P., Kim, B.J., Yoon, C.N., Han, S.K.: Attack vulnerability of complex networks. Phys. Rev. E 65: 056109 (2002).
30. Jin, Y., Matsuo, Y., Ishizuka, M.: Extracting a Social Network among Entities by Web mining. ESWC2007. (2007)
31. Kautz, H., Selman, B., Shah, M.: The hidden Web. AI magazine 18 (2) :27-35 (1997)
32. Khare, R., Celik, T.: Microformats: a pragmatic path to the Semantic Web. WWW2006 (2006)
33. Kim, H., Yang, S., Song, S., Breslin, J. G., Kim, H.: Tag Mediated Society with SCOT Ontology. ISWC2007. (2007).
34. Kochut, K.L., Janik, M.: SPARQLeR: Extended SPARQL for Semantic Association Discovery. Proc. ESWC2007 (2007)
35. Latora, V., Marchiori, M.: A measure of centrality based on the network efficiency. N. J. Phy 9 (6) : 188 (2007)
36. Limpens, F., Buffa, M., Gandon, F.: Rapprocher les ontologies et les folksonomies pour la gestion des connaissances partagées : un état de l'art. IC2008 (2008)
37. Matsuo, Y., Hamasaki, M., Takeda, H., Nishimura, T., Hasida, K., Ishizuka, M.: POLYPHONET: An advanced social network extraction system. WWW2006 (2006)
38. Moreno, J.L.: Emotions mapped by new geography. New York Times (1933)
39. Mika, P.: Social Networks and the Semantic Web. Springer (2007)
40. Milgram, S.: The Small World Problem. Psychology Today, 1(1): 61 – 67. (1967)
41. Mirbel, I.: Vers une ontologie pour les communautés de développement de logiciel libre. IC2009 (2008)
42. Newman, M. E. J.: Scientific collaboration networks. Shortests paths weighted networks, and centrality. Phys Rev E Stat Nonlin Soft Matter Phys 64: 016132 (2001)
43. Newman, M. E. J.: The structure and function of complex networks. SIAM Review 45, 167-256 (2003)
44. Newman, M. E. J.: Detecting community structure in networks. Eur. Phys. J. B 38:321-330. (2004).
45. Newman, M. E. J.: Fast algorithm for detecting community in networks. Phys. Rev. E 69, 066133 (2004)
46. Passant, A., Laublet, P.: Meaning Of A Tag: A Collaborative Approach to Bridge the Gap Between Tagging and Linked Data. LDOW2008 (2008)
47. Paolillo, J. C., Wright, E.: Social Network Analysis on the Semantic Web: Techniques and Challenges for Visualizing FOAF. Visualizing the semantic WebXml-based Internet And Information (2006).
48. Pons, P., Latapy, M.: Computing communities in large networks using random walks. ISCIS2005 (2005)
49. Radicchi, F., Castellano, C., Cecconi, F., Loreto, V., Parisi, D.: Defining and identifying communities in networks. Proc Natl Acad Sci USA 101, 2658-2663 (2004)
50. Santos, E.E., Pan, L., Arendt, D., Pittkin, M.: An Effective Anytime Anywhere Parallel Approach for Centrality Measurements in Social Network Analysis. IEEE2006 (2006)
51. Scott 2000, J.: Social Network Analysis, a handbook, second edition. Sage (2000)
52. Tyler, J. R., Wilkinson, D. M., Huberman, B. A.: Email as spectroscopy: automated discovery of community structure within organizations. Proc. C&T2003 (2003)
53. Wasserman, S., Faust, K., Iacobucci, D., Granovetter, M.: Social Network Analysis: Methods and Applications. Cambridge University Press (1994)
54. Wellman, B.: Computer Networks As Social Networks. Science 293, 2031-34 (2001)
55. Wenger, E.: Communities of Practice: Learning as a Social System Thinker. (1998)
56. Wilkinson, D. M., Huberman, B. A. A method for finding communities of related genes. Proc. Natl. Acad. Sci. (2003)
57. Wu, F., Huberman, B. A.: Finding communities in linear time: a physics approach. HP Labs (2004)
58. Zhou, H., Lipowsky, R.: Network browniam motion: A new method to measure vertex-vertex proximity and to identify communities and subcommunities. ICCS 2004 (2004).