



## Leveraging Social data with Semantics

Guillaume Erétéo, Michel Buffa, Fabien Gandon, Mylène Leitzelman, Freddy Limpens

► **To cite this version:**

Guillaume Erétéo, Michel Buffa, Fabien Gandon, Mylène Leitzelman, Freddy Limpens. Leveraging Social data with Semantics. W3C Workshop on the Future of Social Networking, Jan 2009, Barcelone, Spain. <<http://www.w3.org/2008/09/msnws/>>. <hal-01154486>

**HAL Id: hal-01154486**

**<https://hal.inria.fr/hal-01154486>**

Submitted on 22 May 2015

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# Leveraging Social data with Semantics

Guillaume Erétéo<sup>1</sup>, Michel Buffa<sup>4</sup>, Fabien Gandon<sup>2</sup>, Mylène Leitzelman<sup>3</sup>, Freddy Limpens<sup>2</sup>

1. Orange Labs

guillaume.ereteo@orange-ftgroup.com

2. EDELWEISS, INRIA Sophia-Antipolis, France

{fabien.gandon, freddy.limpens}@sophia.inria.fr

3. Telecom ParisTech, Sophia Antipolis

mylene.leitzelman@telecom-paristech.fr

4. Université Nice Sophia Antipolis, Kewi Team, CNRS/I3S

micbuffa@gmail.com

**Abstract:** One of the challenges of social network analysis (SNA) is to understand and exploit on-line social interactions. Research in Semantic Web has provided models to leverage the richness of these interactions that we use to represent these social networks. Classical social network analysis methods have been applied to these semantic representations without fully exploiting their rich expressiveness. Furthermore, we can extend the representation of social links thanks to the semantic relationships found in the vocabularies shared by the members of the social networks. These “enriched” representations of social networks, combined with a similar enrichment of the semantics of the meta-data attached to the shared resources, will allow the elaboration of “shared knowledge graphs”. In this paper we present our approach to analyse such semantic social networks and capture collective intelligence from collaborative interactions.

## 1. Introduction

The web is now a major medium of communication in our society and, as a consequence, an element of our socialization. As the web is becoming more and more social, we are now collecting huge amount of knowledge on-line. Semantic web researchers provide models to capture such activities that have to be fully exploited in order to be turned into collective intelligence (Gruber, 2008). However current methods of analysis of social networks tend to discard the expressiveness of these representations, and we are still lacking efficient tools to connect the communities scattered over the web via semantic links. We present here our approach to tackle this problem, and future perspectives about the collaborative construction of graph of knowledge, assisted by smart and semantically powered tools.

First, we present some of the challenges in understanding the social interactions that emerged from the use of web platforms. Then, we recall the ontologies used to represent on-line activities that can be combined to connect and represent on-line social networks. In section 4, we propose a stack of semantic social network analysis to exploit the best of semantic representations and social network analysis. Finally, we present some perspectives to exploit such approach and extract the knowledge produced by on-line social interactions by building “shared knowledge graphs”.

## 2. New challenges in understanding online social interactions: The case of Intelligence Process

Before trying to develop what are the challenges of understanding on-line social interactions, we want to start with some definitions to delimit the subject. Let’s consider on one hand that a Social Network is a representation of social interactions between individuals or organizations. It indicates in fact the ways they are connected through different social bounds (i.e., acquaintance, family, professional...) (Wellman, 1996). On the other hand, Social Network Analysis (SNA) forms a family of methodologies to map and evaluate the

relationships and flows of data between people, groups, communities or any type of social structures. While Internet follows inexorably Metcalfe's Law [1](#), on-line social interactions have never been so high on the internet, almost with the web 2.0 advent (Hendler and Golbeck 2008), which places the individuals and their activities at the core of the Web, with all the social softwares and services massively available on the web (like delicious, flickr, linkedin, Facebook...). After the explosion of the "Web of contents" in the end of 90's, we witness the outburst of the "Web of people" and we have to find new means to exploit the reciprocal relationship between people and content, as Peter Morville stated earlier in 2004 "we use people to find content whereas we use content to find people" (Morville 2004). The main challenge to tackle with Internet lies in the fact that any on-line information is networked information and its relevance depends on the networks it is linked to. Google was indeed the first to detect with its Pagerank algorithm that the importance of a page relies more on the links to its environment than to its proper content.

So, among the different human activities impacted by Social Network Analysis, we foresee specific high challenges concerning the Business Intelligence Process. The big question is how individuals inside their organization and the organizations as a whole will tap into this new knowledge wealth to create innovation and improve their competitiveness? Today every organization is forced to anticipate opportunities and threats concerning its profession by detecting "weak signals", to look for value-added information and knowledge and to integrate networks of experts in its fields of excellence. In this context, structured and unstructured information from the web has become a key factor of economic development and innovation. The competitiveness of firms is related to the adequacy of their decisions, which depends heavily on the quality of available information and their ability to capitalize, enrich and distribute this relevant information to people who will make the right decision at the right moment. One of the big challenges we foresee is that the Business Intelligence market is clearly bound to be seriously shaken by the social and viral 2.0 "revolution", which gives the power to the "knowledge worker". As the following matrix states, it is already possible to organize (through mashups, open plugins and APIs) various open-source modules on the whole information cycle (i.e. identification of sources / research / analysis & Treatment / creation / distribution) with an equivalent performance or better than the existing solutions (such as Autonomy's IDOL, Lotus Connection, Hummingbird, Connectivity, etc.).

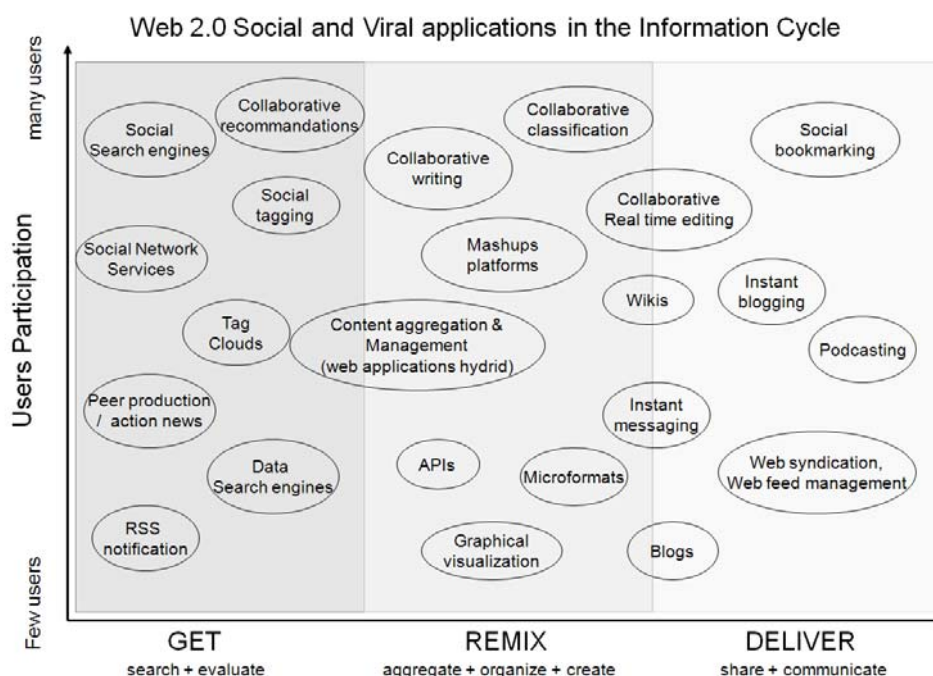


Figure 1 APIs atomizing Business Intelligence Process

We assume that these new 2.0 technological and social trends are transforming totally the classical Knowledge Management and Competitive Intelligence Process inside the firms. While the latter are actually based on data flow analysis, top-down approaches, business process driven, "subject matter experts" location, Communities Of Practice management, the social data and network Softwares and Services (depicted in Fig 1) are reversing the whole process. We are just at the beginning of discovering what will be

the consequences on enterprises worldwide and how it will take for the different generations between boomers, gen X and millennial to overcome their digital divides in an intra-organizational context (Martin 2005).

So the means to better understand the chaos emerging from on-line social data is probably to mix classical Information Retrieval techniques centred on content and Social Information Retrieval Systems centred on social data and human relationships (see (Vuorikari et al 2007)) for more details). We present in the following sections our approach based on Semantic social network analysis.

### 3. Representing social data with semantic web

Web frameworks provide a graph model (RDF [2](#)), a query language (SPARQL [2](#)) and schema definition frameworks (RDFS [2](#) and OWL [2](#)) to represent and exchange knowledge on-line. These frameworks provide a whole new way of capturing social networks in much richer structures than raw graphs.

Social data can be seen as a twofold structure: data that describe the social network, and data that describes what is produced by their members. Several ontologies can be used to represent social networks. The most popular is FOAF [3](#), 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: on-line accounts, weblog, memberships, etc. The properties defined in the RELATIONSHIP [4](#) 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 [5](#) ontology specialize "OnlineAccount" and "HasOnlineAccount" from FOAF in order to model the interactions and resources manipulated by users of social web applications; SIOC defines concepts such as posts in forums, blogs, etc. Researchers (Bojars et al, 2008) have shown that SIOC and the other presented ontologies can be used and extended for linking and reuse scenarios and data from web 2.0 community sites. In addition, the SKOS [6](#) ontology offers a way to organize manipulated concepts and to link them to SIOC descriptions with the property "isSubjectOf". RDF based description of social data forms a rich typed graph and offers a much more powerful and significant way to represent online social networks than traditional models of SNA.

In parallel, web 2.0 applications made social tagging popular: users tag resources of the web (pictures, video, blog posts etc.). A set of tags built from the use of such applications forms a folksonomy that can be seen as a shared vocabulary that is both originated by, and familiar to, its primary users (Mika 2005). Ontologies have been designed to capture and exploit the activities of social tagging, as Gruber (2005) suggested it. These descriptions can deal with ... (Passant and Laublet, 2008) and in parallel researchers have attempted to bridge folksonomies and ontologies to leverage the semantics of tags (see overview in (Limpens et al 2008)). SCOT (Kim 2007) has been designed to be used with well adopted models like FOAF and SKOS. Once they are typed and structured, the relations between the tags and between the tags and the users also provide a new source of social network. In fact social structures can be analysed to type data produced by social actors and vice versa, data produced by social actors can be analysed to type social networks.

### 4. Semantic social network analysis

We build on enhanced RDF-based representations to carry out fully "semantic social network analysis" of online interactions. We leverage semantic web technologies to merge and exploit the best features of each of these approaches. To achieve this, we designed a framework using the graph models underlying RDF and SPARQL extensions enabling us to extract efficiently and semantically parameterize the SNA features such as strategic positions, roles, community detection etc. directly using these representations.

Figure 2 illustrates the abstraction stack we follow. We use the RDF graphs to represent social data, using existing ontologies (FOAF, SIOC, SCOT, RELATIONSHIP, DOAP) together with specific domain ontologies if needed. Some social data are already readily available in a semantic format (RDF, RDFa, hCard  $\mu$ format, etc.) and can be exploited straightforward (e.g. FOAF profiles from LiveJournal.com, RDF data from openguides.com, SIOC metadata from blogs like Wordpress, GRDDL extraction of  $\mu$ formats, etc.). However, today, most of the data are still only accessible through APIs (Youtube, flickr, Open Social, Facebook, etc.) or by crawling web pages and need to be converted.

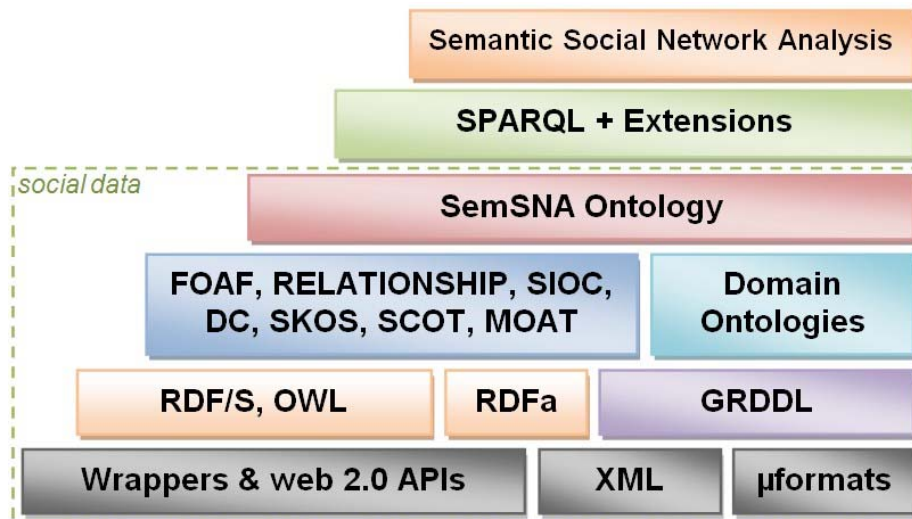


Figure 2 Abstraction stack for semantic social network analysis

We also designed SemSNA, an ontology that describes the SNA notions (e.g. centrality) and allows us to manage the life cycle of an analysis. With this ontology, we can (1) abstract social network construct from domain ontologies to apply our tools on existing schemas by making them extend our primitives ; and we can (2) enrich the social data (the RDF graph nodes) with new annotations such as the SNA indices that will be computed (degree centrality, betweenness centrality, closeness centrality, etc.). These annotations enable us to manage more efficiently the life cycle of an analysis, by calculating the SNA indices only once and updating them incrementally when the network changes over time.

Based on this model, we propose SPARQL formal definitions to compute semantically parameterized SNA features and annotate the graph nodes consequently, caching the results. The current test uses the semantic search engine CORESE (Corby et al 2004) based on graph representations and processing that supports powerful SPARQL extensions particularly well suited for the computation of the SNA features that, in particular, require path computations (Corby 2008). Among important results in SNA is the identification of sociometric features that characterize a network (Er t e o et al 2008). For example, the centrality highlights the most important actors of the network. We have adapted with our approach the computation of the three kinds of centrality that have been proposed by Freeman (Freeman 1979). Our perspectives include the adaptation of other algorithms in particular for community detection and new semantic algorithms based on adaptation of classical SNA definitions.

## 5. Building “shared knowledge graphs”

As we have seen it before, we need to enrich with semantics the simple representations of social networks and the content their users share, in order to fully exploit the wealth of data and interactions on the web. Doing so could consist in building “shared knowledge graphs” which help users find relevant resources or persons. In the field of knowledge management, this was the idea behind Topic Maps and ontologies of the Semantic Web: they were thought of as knowledge representations capable of grasping the multi-dimensions of the information we exchange (see Bajet et al (2008) for an overview of the different knowledge representations based on graphs). These shared knowledge graphs can be seen as a generalization of these two types of knowledge representation, with a focus on the shareable feature and the ability for both machines and humans to exploit them at different levels of functionalities.

Folksonomies are a recent example of “shared knowledge structures” which have emerged from Web 2.0 applications as an affordable way to massively categorize resources. But folksonomies do not truly consist in an elicitation of the knowledge contained in the documents which are tagged by users, since tags are not structured nor semantically related to each other. Tags in folksonomies remain at the stage of ad-hoc categories which serve user-centred purposes (Veres, 2006). If tags can be interpreted by humans, we still lack effective tools to integrate them with richer semantic representations shared by other members of their web communities, or by other web communities. (also miss opportunities)

In order to map the knowledge exchanged by Web communities, several challenges have to be addressed. First, for interoperability purposes we need to find a good balance in the standardization of the manifold

ways of describing content on the Web. The “Web of Linked Data” initiative proposes weaving a web of scattered sources of knowledge thanks to a combination of “good practices” and conceptual schemes describing them. Examples of such conceptual schemes can be seen in the formal ontologies presented in section 3, which describe content exchanged by and within on-line communities. This type of approaches are a good start as they already assist users in identifying, for instance, all the contents posted by a user across multiple sites, but we are still missing tools and methods to connect communities at a semantically richer level (Newell, 1982)

The next step lies in enriching the semantics by which we intend to map contents from multiple platforms. A possible means to achieve this consist in “shared knowledge hubs”. The DBpedia project (Auer et al., 2007) is an example of such hubs, as it proposes expressing the knowledge structure of the Wikipedia pages in machine processable data. By doing so, they provide a sort of common reference (the hierarchically organized Wikipedia Categories for instance) to which we can start connecting more elaborated “knowledge graphs”.

Of course, these common references are not sufficient to describe each community’s field of knowledge, but they provide common terminologies, which need not be exclusive, and to which it is possible to hook more specific terms. The “Web of linked Data” is made of multiple webs of tacit bits of knowledge that are still today rarely explicitly expressed in both machine and human understandable representations.

Web 2.0 applications and folksonomies have lead to novel user experiences and yielded rich materials which are still missing appropriate representations to be efficiently browsed. This goal can be achieved by developing tools to assist community members to connect their own knowledge categories to common references. For instance, current terminology extractors can be exploited in the context of folksonomies in order to detect common taxonomies categories among the tags, and propose to the contributors of these folksonomies to map their tags with these categories, or to create new ones when needed (Passant & Laublet, 2008). The semantic structure of the folksonomies could also combine automatic inferences with the expertise of the users by integrating the validation these inferences within the “natural” use of the systems. This aspect opens up new perspectives to create novel interfaces to knowledge repositories that exploit the best of semantic technologies and the dynamism of the social web.

## 6. Conclusion

Through the example of Business Intelligence Process we highlighted that the systematic exploitation of information to foster economic performances and facilitate decision making is one of the keys to success for all organizations worldwide. Our semantic SNA stack provides a way to fully exploit the RDF representations of on-line social networks and enrich social data with contextualized SNA features. These “enriched” social networks, combined with semantic descriptions of the knowledge they exchange, will allow for the construction of shared knowledge graphs that will help efficiently use the overwhelming flow of data on the Web. We believe that an effective approach to building these shared knowledge graphs and turning on-line social experiences into collective intelligence will have to exploit the power of both computational methods and the expertise of humans in a synergistic way.

## End notes

1. Metcalfe’s law states that the useful power of a network multiplies rapidly as the number of peers in the network increases, exactly “The community value of a network grows as the square of the number of its users increase”
2. Semantic Web, W3C, <http://www.w3.org/2001/sw/>
3. <http://www.foaf-project.org/>
4. <http://vocab.org/relationship/>
5. <http://sioc-project.org/>
6. <http://www.w3.org/2004/02/skos/>

## References

Auer S., Bizer C., Kobilarov G., Lehmann J., Cyganiak R. and Ives Z. G. (2007). Dbpedia: A nucleus for a web of open data. In ISWC/ASWC, volume 4825 of Lecture Notes in Computer Science, p. 722–735:

Springer.

Breslin J., Harth A., Bojars U. and Decker S. (2005). Towards Semantically-Interlinked Online Communities. In ESWC.

Baget, J.; Corby, O.; Dieng-Kuntz, R.; Faron-Zucker, C.; Gandon, F.; Giboin, A.; Gutierrez, A.; Leclère, M.; Mugnier, M. and Thomopoulos (2008), R. Griwes: Generic Model and Preliminary Specifications for a Graph-Based Knowledge Representation Toolkit Proc. of the 16th International Conference on Conceptual Structures (ICCS'2008).

Bojars, U., Breslin, J.G, Finn, A., Decker, S. (2008): "Using the Semantic Web for linking and reusing data across Web 2.0 communities", J. Web Sem. 6: 21-28.

Corby, O., Dieng-Kuntz, R., and Faron-Zucker, C. (2004), Querying the semantic web with the corese search engine. ECAI/PAIS2004.

Corby, O (2008): Web, Graphs & Semantics, Proc. of the 16th International Conference on Conceptual Structures (ICCS'2008)

Erétéo, G., Buffa, M., Gandon, F., Grohan, P., Leitzelman, M., Sander, P. (2008): A State of the Art on Social Network Analysis and its Applications on a Semantic Web, SDoW2008 (Social Data on the Web), workshop at the 7th International Semantic Web Conference.

Freeman, L.C.: Centrality in social Networks: Conceptual Clarification. Social Networks 1, 215-239, 1979.

Gruber T. (1993). A Translation Approach to Portable Ontology Specifications. Knowledge Acquisition, 5(2), 199–220.

Gruber, T. (2005): Ontology of folksonomy: A mash-up of apples and oranges. MTSR2005.

Gruber, T. (2008): Collective knowledge systems: Where the Social Web meets the Semantic Web J. Web Sem, 6, 4-13

Kim, H., Yang, S., Song, S., Breslin, J. G., Kim, H. (2007): Tag Mediated Society with SCOT Ontology. ISWC2007.

Hendler, J., Goldbeck, J. (2008): Metcalfe's law, web 2.0 and the Semantic Web. J. Web Sem. 6(1):14-20.

Limpens F., Gandon F. and Buffa M. (2008). Bridging Ontologies and Folksonomies to Leverage Knowledge Sharing on the Social Web: a Brief Survey. In Proc. 1st International Workshop on Social Software Engineering and Applications (SoSEA), L'Aquila, Italy.

Martin A.C. (2005). From high maintenance to high productivity: What managers need to know about Generation Y, Industrial and Commercial Training, Vol 37, n°1 pp 39-44

Mika, P. (2005): Ontologies are us: A unified model of social networks and semantics. In Gil, Y., Motta, E., Benjamins, V. R. and Musen, M. A., (2005): Eds., The Semantic Web. Proceedings of the 4th International Semantic Web Conference, ISWC2005, volume 3729 of Lecture Notes in Computer Science, p. 522–536: Springer.

Morville, P. (2004): "Ambient findability", Digital Web Magazine, [http://www.digital-web.com/articles/ambient\\_findability/](http://www.digital-web.com/articles/ambient_findability/)

Newell, A. (1982): The Knowledge Level Artificial Intelligence, 18, 87-127

PARK J. & HUNTING S. (2002). XML Topic Maps: Creating and Using Topic Maps for the Web. Addison-Wesley Professional.

Passant, A., Laublet, P. (2008): Meaning Of A Tag: A Collaborative Approach to Bridge the Gap Between Tagging and Linked Data. LDOW2008.

Veres C. (2006). The language of folksonomies: What tags reveal about user classification. In Natural

Language Processing and Information Systems, volume 3999/2006 of Lecture Notes in Computer Science, p. 58–69, Berlin / Heidelberg: Springer.

Vuorikari, R., Manouselis, N., Duval, E. (2007): Metadata for social recommendations: storing, sharing and reusing evaluations of learning resources, *Social Information Retrieval Systems: Emerging Technologies and Applications for Searching the Web Effectively*, (Goh, D.H. and Foo, S., eds.), Idea Group Inc., pp.87-107.

Wellman, B. (1996): "For a social network analysis of computer networks: a sociological perspective on collaborative work and virtual community", *Proceedings of the ACM SIGCPR/SIGMIS conference on Computer personnel research*, Denver, Colorado, United States p 1 – 11.