

# Semantic Social Network Analysis, a concrete case

**Guillaume Erétéo,**

*Orange Labs,*

*guillaume.ereteo@orange-ftgroup.com*

**Freddy Limpens, Fabien Gandon, Olivier Corby**

*EDELWEISS, INRIA Sophia-Antipolis, France*

*{firstname.lastname}@sophia.inria.fr*

**Michel Buffa, Mylène Leitzelman, Peter Sander**

*Nice Sophia Antipolis University, Kewi Team, CNRS/I3S*

*{firstname.lastname}@unice.fr*

## ABSTRACT

The World Wide Web has been evolving into a read-write medium permitting a high degree of interaction between participants, and social network analysis (SNA) seeks to understand this on-line social interaction, for example by identifying communities and sub-communities of users, important users, intermediaries between communities, etc. Semantic web techniques can explicitly model these interactions, but classical SNA methods have only been applied to these semantic representations without fully exploiting their rich expressiveness. The representation of social links can be further extended thanks to the semantic relationships found in the vocabularies (tags, folksonomies) shared by the members of these 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 chapter we present our approach to analyzing such semantic social networks and capturing collective intelligence from collaborative interactions to challenge requirements of Enterprise 2.0. Our tools and models have been tested on an anonymized dataset from Ipernity.com, one of the biggest French social web sites centered on multimedia sharing. This dataset contains over 60,000 users, around half a million declared relationships of three types, and millions of interactions (messages, comments on resources, etc.). We show that the enriched semantic web framework is particularly well-suited for representing online social networks, for identifying their key features and for predicting their evolution. Organizing huge quantity of socially produced information is necessary for a future acceptance of social applications in corporate contexts.

## INTRODUCTION

The web is now a major medium of communication in our society and, as the web is becoming more and more social, a huge amount of content is now collectively produced and widely shared online. Even early on, the social interactions on the web highlighted a social network structure (Wellman 1996), a phenomena dramatically amplified by web 2.0 which follows inexorably Metcalfe's Law<sup>1</sup> (Hendler and Golbeck 2008). Individuals and their activities are at the core of the web, along with all the easily-available social software and services, e.g., Delicious, Flickr, LinkedIn, Facebook. After the explosion of the "web of content" at the end of 90's, we are witnessing the outburst of the "web of people". Taken together, "we use people to find content whereas we use content to find people" (Morville 2004), and we need new means to investigate the relationship between people and content.

---

<sup>1</sup> Metcalfe's law states that the useful power of a network multiplies rapidly as the number of users of the network increases, "The community value of a network grows as the square of the number of its users"

**New challenges in understanding online social interactions: the case of Business Intelligence Process.**

Today every organization is forced to anticipate opportunities and threats by detecting "weak signals", to look for value-added information and knowledge, and to integrate networks of experts into its domains of activity. 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 decisions at the right moment. The Business Intelligence market is clearly bound to be seriously shaken up by the social and viral 2.0 revolution. As shown in Figure 1, it is already possible to organize (through mashups, open plugins and APIs) various free modules over the whole information cycle, i.e., identification of sources / research / analysis and treatment / creation / distribution, with an efficiency competing proprietary solutions (such as Autonomy's IDOL, Lotus Connection of IBM, and SAP BI software suite, etc.).

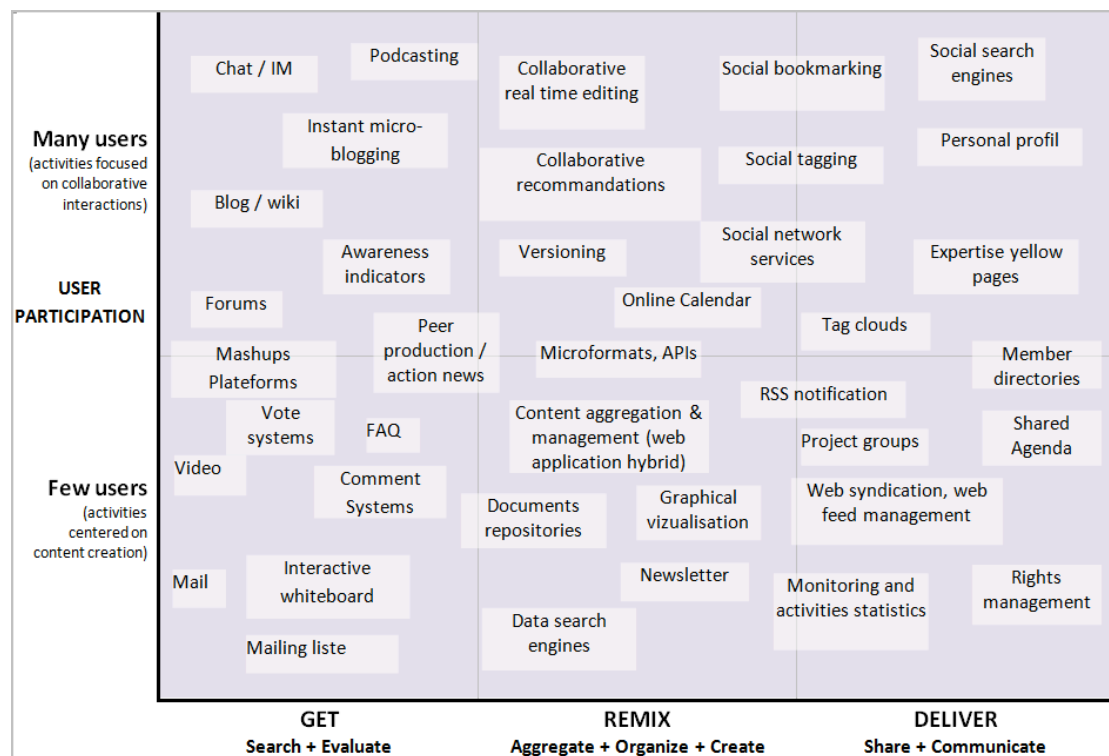


Figure 1. tools that transform Business Intelligence Process

Classical Knowledge Management and Competitive Intelligence Process inside firms are currently based on top-down business process driven approaches involving data flow analysis, subject matter expert location and Communities of Practice management. Online social data and network Software and Services (depicted in Figure 1) are reversing this whole process and empower the knowledge worker. We are witnessing the consequences on enterprises worldwide and the different generations - boomers, gen X and millennial – will have to overcome their digital divides in intra-organizational contexts (Martin 2005). Individuals inside their organization, and organizations as a whole, need tools to exploit this new wealth of knowledge to create innovation and to improve performance.

Consequently, more and more social solutions (Social Text<sup>2</sup>, Blue Kiwi<sup>3</sup>, etc.) are being deployed in corporate intranets to reproduce information sharing success stories from the web into an organizational context. This new trend is also called “Enterprise 2.0”, that Andrew

<sup>2</sup> <http://www.socialtext.com/>

<sup>3</sup> <http://www.bluekiwi-software.com/>

McAfee first coined as "the use of emergent social software platforms within companies, or between companies and their partners or customers" (McAfee 2006). These collaborative platforms allow conducting innovative strategic watch by introducing social interactions into every step of the watch cycle: search, monitoring, collecting, handling, dissemination. Information produced at different sources becomes accessible at a single entry point, is quickly shared and permanently enriched with comments and new sources. However, these platforms also augment the amount of information their users are exposed to. The benefit of information sharing is often hindered when the social network becomes so large that relevant information is lost in an overwhelming flow of activity notifications. Losing information can lead to a loss of reactivity and competitiveness in a professional context. Organizing this huge quantity of information is necessary for gaining acceptance in corporate contexts and to achieve the full potential of Enterprise 2.0. Social activities and user generated content have to be properly organized and filtered before any notification is pushed to users if we want to preserve the benefits of online collaboration. These social data are produced through different interactions between users who maintain many types of relationships.

We present here our approach to (1) capture and (2) exploit the knowledge that is contained in social interactions that emerge from the use of web 2.0 applications. The first step (capture) needs models and languages for representing the diverse knowledge that emerges from online collaboration in a machine readable and exchangeable format. The second step (exploit) requires means (languages, tools) to query such evolving and diffuse social knowledge. We answer these issues with semantic web frameworks, and will show that they address both topics efficiently. Social network analysis (SNA) is a domain that provides relevant metrics and algorithm to understand the structure of the social network that can be built from social interactions. We also show that the use of semantic web technologies is well adapted for performing SNA on online social networks, adding flexibility and simplicity to many steps of the computation of common SNA indices.

In the first part of this chapter, we recall existing works conducted by researchers from the semantic web domain - the ontologies used to represent online activities that can be combined to connect and represent online social networks. Then, we present approaches to structure and organize the shared vocabularies (folksonomies) built by users when they tag shared content on web 2.0 web sites. We will show that the tagging activities contribute to reinforcing social bonds thanks to greater involvement and freedom in publishing, organizing and sharing content and constitute a novel opportunity for analyzing social networks. In the last section, we propose a stack of tools for achieving semantic social network analysis. While existing tools discard the richness of semantic social networks, we propose a framework to handle not only their structure but also the semantics of the ontological primitives used to capture their knowledge. We present the results obtained by analyzing a real social network with over 60,000 users, connected through half a million declared relationships of three types and millions of interactions: messages, comments, visits, etc.

Finally, we present some perspectives on the exploitation of folksonomy data thanks to semantic tools and methods. We will show how the combination of Web 2.0 and semantic web approaches can help to dramatically enhance the effectiveness of bottom-up approaches to sharing and organizing resources, as well as to discover hidden social bonds within the knowledge shared among online communities.

## **REPRESENTING SOCIAL DATA WITH SEMANTIC WEB FRAMEWORKS**

### **Historical background: different graph models**

The emerging interactions between people on the internet and especially later on the World Wide Web quickly revealed social network structures (Wellman 1996) with properties that were close to those observed in the physical world. Researchers have extracted social networks from synchronous and asynchronous discussions (e.g., emails, mailing-list archives, IRC), the hyperlink structure of homepage citations, co-occurrence of names in web pages,

and from the digital traces created by web 2.0 application usages (Erétéo et al 2008). Considering this last point, turning the read web into a read/write web has led to dramatic growth in the different possibilities for interaction, producing a huge amount of heterogeneous social data. Information and content on the web are now collectively produced, socially discovered and quickly shared through mashable applications. We are witnessing the deployment of a social media landscape where "expressing tools allow users to express themselves, discuss and aggregate their social life", "sharing tools allow users to publish and share content", "networking tools allow users to search, connect and interact with each other" and "playing services integrate strong social features" (Cavazza 2009). Social platforms, like Facebook, Orkut, Hi5, etc., are at the center of this landscape as they enable us to host and aggregate these different social applications. As an example you can publish and share your Delicious bookmarks, your RSS streams or your microblog posts in the Facebook news feed, thanks to dedicated Facebook applications. This integration of various means of publishing and socializing enables us to quickly share, recommend and propagate information to our social network, trigger reactions, interact with it, and finally enrich it. Moreover web 2.0 has made social tagging popular, permitting an additional level of organization for tagged web resources (pictures, videos, blog posts etc.). A set of tags built from usage 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). This collaborative classification of web resources can be further analyzed in order to decipher implicit links between users who use similar vocabularies or tag the same content, highlighting the existence of common interests.

As more people use these social applications they expose more and more of their lives and social networks. Sociologists now have access to a valuable source of social data that captures characteristics of our societies with permanently evolving web usages and web technologies. The need for some appropriate representation to exploit them has consequently emerged. Traditionally researchers have used graph theory which proposes different graph models to represent this data (Scott 2000). People and resources are represented by nodes and relationships are represented by edges. Social networks with symmetric relationships as in Facebook, can be represented by non-oriented graphs. Inversely, oriented graphs are well suited to model social networks with non-symmetric relations like the "follows" relationships of Twitter. In weighted graphs, weights are associated to edges to specify the intensity of the relationships, useful for representing the frequency of interactions between people through messages or comments. Social networks like Ipernity.com (a French web 2.0 site for sharing pictures and videos) or Facebook propose adding labels (e.g. family, friends, favourite) on edges to represent the type of relationships that links actors. Finally, sharing sites (e.g., Flickr, YouTube, Delicious) allow interaction on shared content (e.g. photos, videos, bookmarks), connecting them through virtual artifacts. Such social networks are represented with bipartite graphs, with two types of nodes and edges that link nodes of each type. A hyperedge extends the notion of an edge by allowing more than two nodes to be connected and is often used to represent complex relationships involving at least three resources (e.g. a user, a document and a tag).

However, while human interactions in web 2.0 sites produce a huge amount of social data, capturing more and more aspects of physical social networks, this decentralized process suffers from little interoperability and little linking between diffuse data. In fact, such rich and spread-out data can't be represented using only the models of graph theory outlined above without some loss of information. These representations are poorly typed with labels on edges but with no semantic links to structure them. Moreover, they are not necessarily adapted for exchanging data and semantics across applications. We'll now see how semantic web frameworks tackle these requirements and how they can be used to represent online social networks.

## Enriching social data with semantics

Semantic web frameworks answer the problem of representing and exchanging data on such social networks with a rich typed graph model (RDF<sup>4</sup>), a query language (SPARQL<sup>4</sup>) and schema definition frameworks (RDFS<sup>4</sup> and OWL<sup>4</sup>). RDF enables us to make assertions and to describe resources with triples (domain, property, range) that can be viewed as "the subject, verb and object of an elementary sentence", "a natural way to describe the vast majority of the data processed by machines" (Berners-Lee 2001). Each element of a triple is identified by a URI (Uniform Resource Identifier), which enable every application to make its own description to identify it. These triples provide RDF with a directed labeled graph structure that is well suited to representing the social data of users that connect and interact through heterogeneous content on different web sites. First, they allow data to be spread across the internet and intranet networks, involving actors, content and relationships, and are represented with a uniform graph structure in RDF even if they are located on different sites. The URIs that are used to identify resources and properties, link distributed identities and activities. Same URIs identify the same resources so that two URIs describing the same resource can be unified with a single description stating so. Then, both nodes and relationships can be richly typed with classes and properties of ontologies that are described in RDFS and OWL adding a semantic dimension to the social graph. An ontology is "a set of representational primitives with which to model a domain of knowledge or discourse. The representational primitives are typically classes (or sets), attributes (or properties), and relationships (or relations among class members). The definitions of the representational primitives include information about their meaning and constraints on their logically consistent application" (Gruber 2009). As an example, the inheritance relation is a frequently used relation between classes and properties to define taxonomies (e.g., *web page* is a *sub class* of *document* and *parent of* is a *sub property* of *family*), but any relation between terms can be specified (e.g. *parent of* is *narrower* than *family*). Finally, SPARQL is the standard query language for querying RDF data and for performing all desired transformations on these semantic social networks (San Martin et al 2009). We will now look at ontologies for describing social activities and actors on the web.

Social data can be seen as a twofold structure: data describing the social network structure, and data describing the content produced by network members. Several ontologies exist for representing online social networks (see the chapter "Understanding Online Communities Using Semantic Web Technologies"). Currently, the most popular is FOAF<sup>5</sup>, used for describing people, their relationships and their activity. A large set of properties defines 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 for describing web usages: online accounts, weblogs, memberships, etc. The properties defined in the RELATIONSHIP<sup>6</sup> 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". *Figure 2* shows a typed graph that uses a rich model for representing the relations between nodes.

---

<sup>4</sup> Semantic Web, W3C, <http://www.w3.org/2001/sw/>

<sup>5</sup> FOAF, Friend Of A Friend <http://www.foaf-project.org/>

<sup>6</sup> RELATIONSHIP, <http://vocab.org/relationship/>

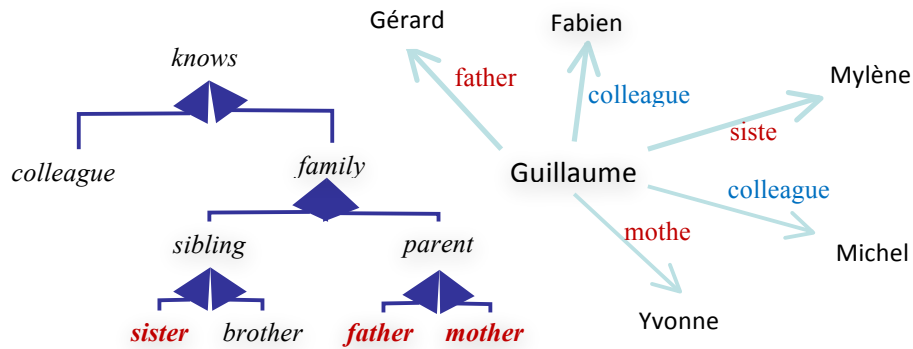


Figure 2 a typical social network represented with types relations and nodes

The primitives of the SIOC<sup>7</sup> ontology specialize “OnlineAccount” and “HasOnlineAccount” from FOAF in order to model the interactions and resources manipulated by users of social web applications (Breslin et al 2005); SIOC defines concepts such as posts in forums, blogs, etc. Researchers (Bojars et al, 2008) have shown that SIOC and the other ontologies can be used and extended for linking to and reusing scenarios and data from web 2.0 community sites. In addition, the SKOS<sup>8</sup> ontology offers a way to organize concepts with lightweight semantic properties (e.g., narrower, broader, related) and to link them to SIOC descriptions with the property “isSubjectOf” (see Figure 3).

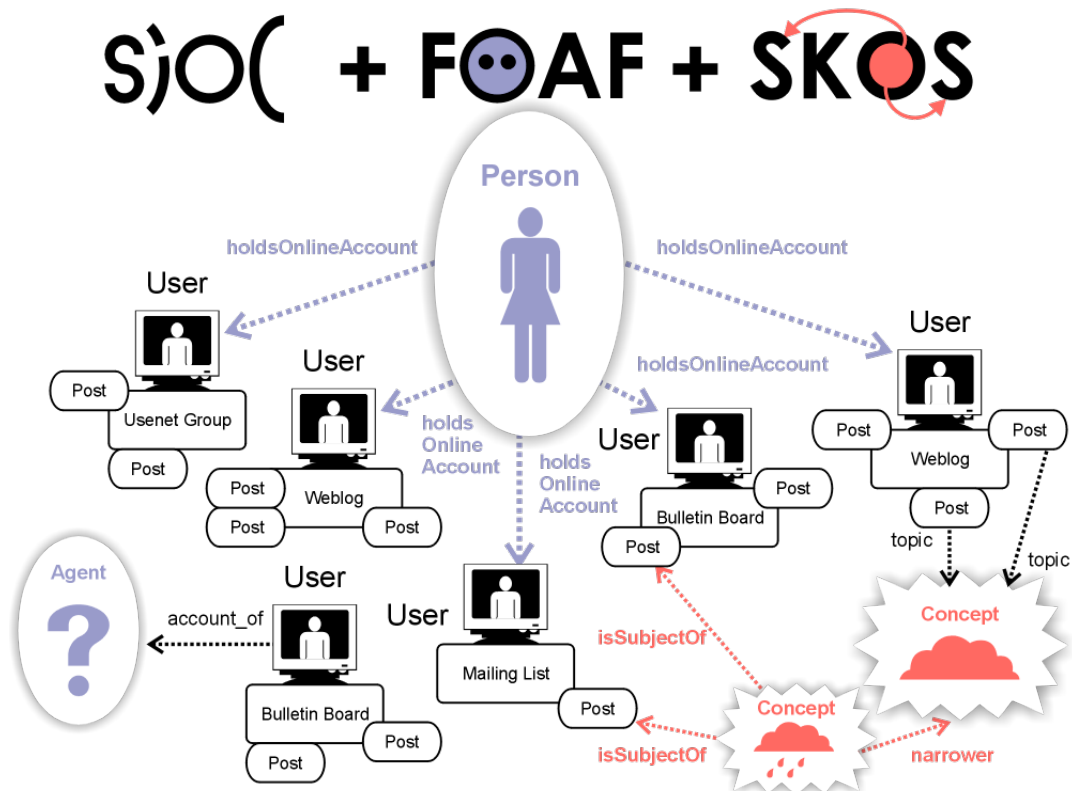


Figure 3. Alignments between SIOC, FOAF and SKOS<sup>9</sup>

Social tagging consists in allowing users to associate freely chosen key-words, called tags, with the resources they exchange such as blog posts, photos, or bookmarks (see Figure 4).

<sup>7</sup> SIOC, Semantically Interlinked Online Communities, <http://sioc-project.org/>

<sup>8</sup> SKOS, Simple Knowledge Organization System, <http://www.w3.org/2004/02/skos/>

<sup>9</sup> SIOC, Semantically Interlinked Online Communities, <http://sioc-project.org/node/158>

The result of the collection of such associations, called “taggings”, is a *folksonomy*. Social tagging and folksonomies can be improved by adding semantics that structure and link tags together. Gruber (2005) was among the first to suggest designing ontologies to capture and exploit the activities of social tagging (Newman et al. 2005) (Kim et al. 2007), These descriptions can deal with the author of the tag, or the tag itself as a character string, but also with additional properties such as the service where this tag is shared, or even a vote on the relevance of this tag. Other research work has attempted to go further by linking tags with explanations of their meaning (MOAT, Meaning Of A Tag, Passant and Laublet, 2008), or more generally, by bridging folksonomies and ontologies to leverage the semantics of tags (see an overview of this very topic in Limpens *et al.* 2008).

RDF-based descriptions of social data form a rich typed graph, exchangeable across web applications, and offer a much more powerful and significant way to represent online social networks than traditional models of SNA. However, other formalisms exist to easily attach lightweight semantics to web resources and are now widely used.

Microformats expose social data in web pages using XHTML markup. They are considered as "a pragmatic path to the semantic web" (Khare et al 2006) and solutions exist to bridge them with RDF (Adida 2008). "Microformats are a way of attaching extra meaning to the information published on a web page. This is mostly done through adding special pre-defined names to the class attribute of existing XHTML markup"<sup>10</sup>. Microformats are proposed to describe people, organizations and places (*hCard*), human relationships(*XFN - XHTML Friends Network*), events (*hCalendar*), opinions, rating and reviews (*VoteLinks, hReview*) and tags (*rel-tag*).

The following examples show some conventions of the use of XHTML attributes to add lightweight semantics with microformats.

XFN adds `rel` attributes to `<a href>` xhtml tags with all appropriate values separated by spaces to define the type of relationship(s) between the author of the page and a person represented by the URI defined in the href attribute.

```
<a href="http://jeff.example.org" rel="friend met">
```

In the same way the rel-tag microformat recommend using the value `tag` in the `rel` attribute of an `<a href>` tag to state that the link points to a tag:

```
<a href="http://technorati.com/tag/tech" rel="tag">tech</a>
```

VCard specifies values of `class` attributes to type the content of xhtml tags describing people, organization or places:

```
<div class="vcard">
  <div class="adr">
    <span class="type">Work</span>:
    <div class="street-address">169 University Avenue</div>
    <span class="locality">Palo Alto</span>,
    <abbr class="region" title="California">CA</abbr>
    <span class="postal-code">94301</span>
    <div class="country-name">USA</div>
  </div>
  <div class="tel"><span class="type">Work</span>+33651743832</div>
  <div>Email:<span class="email">ereteog@gmail.com</span></div>
</div>
```

## Adding structure and semantics to social tagging and folksonomies can help in building social graphs

Since tags are neither explicitly structured nor semantically related to each other, folksonomies have limited capacities in fully eliciting the knowledge contained in documents

---

<sup>10</sup> <http://microformats.org>

tagged by users. Tags in folksonomies remain at the stage of ad-hoc categories which serve user-centred purposes (Veres 2006). While 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.

Researchers have attempted to bridge folksonomies and ontologies to leverage the semantics of tags (Limpens and al 2008). Once semantically typed and structured, the relationships between tags and between tags and users also provide a new source of social networks. In fact social structures can be analyzed to type data produced by social actors and vice versa, data produced by social actors can be analyzed to type social networks. Consequently, tags can be used to link people, with the help of semantics (by identifying, for instance, communities that share the same interests).

Providing pivot languages to capture and exchange social data takes special importance in corporate application such as business intelligence or technology watch: these schemas and the underlying semantic web frameworks are ground foundation for data integration spanning both online sources and internal corporate applications. The network of experts, the information resources they watch, the report they produce, etc. can be integrated and articulated inside this unified graph-based set of frameworks to support transversal analysis such as identifying central experts, their interests and the sources they use regardless of where the different pieces of knowledge come from.

In the next section we will focus on the different approaches that can be used to add semantics to folksonomies.

## **BRIDGING FOLKSONOMIES AND ONTOLOGIES**

Social tagging systems have recently become very popular as a means of classifying large sets of resources shared amongst on-line communities over the social Web. The simplicity of tagging, combined with the web 2.0 culture of exchange, allow users to share their annotations on the mass of resources.

While the act of tagging is primarily for content categorization purposes, it can also be used for building social networks. For instance, we can link people who:

- used the same tag, and/or
- tagged the same resource.

The simple examples of *Figure 4* show how we can link people who share the same interest, be it symbolized by an interest on the same resource, or on the same tag. However, this approach can be greatly improved by adding semantics to the folksonomies: (1) by grouping similar or related tags; or (2) by inferring a hierarchy of tags. For instance, these semantic links can consist in stating that the tag “music” is *broader* than the tag “guitar”, or “saxophone” is *narrower* than “music” etc.

For example, if John tags a document with “saxophone” and if Freddy tags another document with “guitar”, and if “guitar” and “saxophone” are both narrower than the tag “music”, we can say that Freddy and John share the same interest for “music”, even if they share no common resources tagged with “music”. It will be now possible to state that Freddy and John are members of the community of people interested by music, *and they form an interest-based social network*.

In this section, we will first analyze folksonomy usages and limitations, and position them among the other classical ways of categorizing. Then we will present the state of the art about semantic enrichment of folksonomies and the different ways of bridging them with ontologies to be able to discover semantic links between tags. Finally we present our recent work that consists in integrating folksonomies into a collaborative construction of knowledge representations, aiming at providing additional functionalities to folksonomy-based systems and at semantically enriching folksonomies.



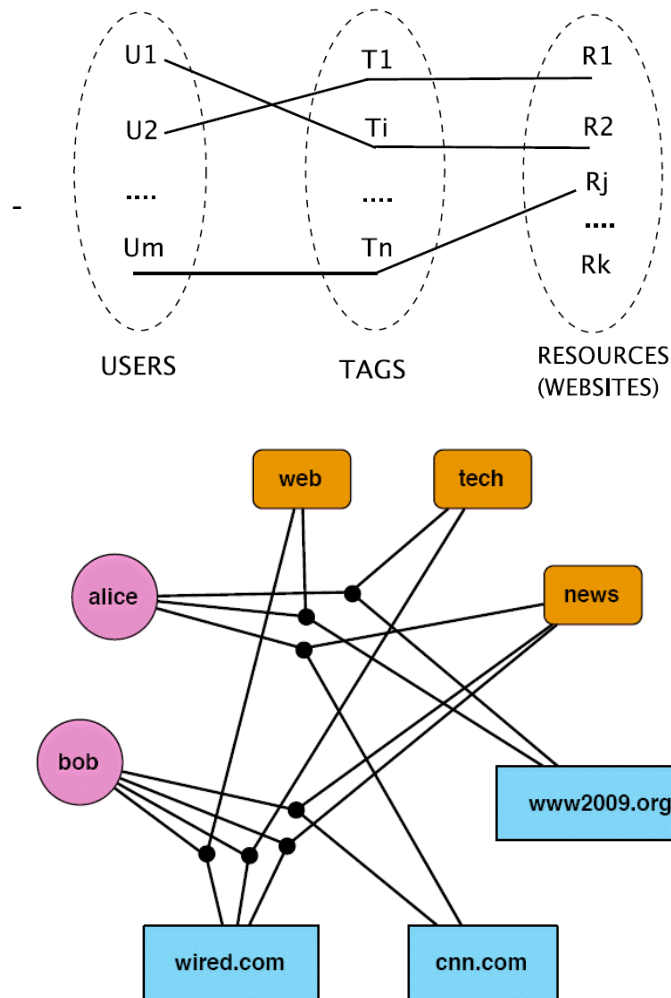


Figure 4. (at the top) Tripartite model of folksonomy (Halpin et al., 2007), and (at the bottom) illustration of the tagging of 3 web sites by 2 users using 3 tags (Markines et al., 2009) where Round = users, Rectangle = resources, Rounded rectangle = tags

## Folksonomy usages and limitations

Several qualitative studies have been conducted on folksonomies. (Golder & Huberman 2005) have analyzed the use of folksonomies and have proposed classifying the act of tagging itself into different categories in the context of a typical application of social bookmarking, such as the topic of the item tagged, or as adjectives characterizing the opinion of the author (“funny”), or such as tags oriented towards a specific task (“toread”). (Vanderwal 2004) distinguished broad folksonomies (when tags tend to be understandable by numerous users) from narrow folksonomies (when tags are more user-centered). (Veres 2006) tried to define the linguistic nature of tags and showed that some tags correspond to taxonomic categories, while other tags correspond to ad hoc categories serving users' purposes. Thus, folksonomies are a mirror of the diversity of points of view and usages of the users who share their tags. However, the exploitation of folksonomies raises several issues, as pointed out by (Mathes 2004) and (Passant 2009):

- (1) the ambiguity of tags: one tag may refer to several concepts;
- (2) the variability of the spelling: several tags may refer to the same concept;
- (3) the lack of explicit representations of the knowledge contained in folksonomies (folksonomies are “flat”, just sets of isolated keywords);
- (4) difficulties in dealing with tags from different languages.

*To overcome these limitations, the classical alternative to social tagging is the use of structured knowledge representations to classify or to index resources.*

*Formal ontologies* consist in a specification of the conceptualization of a domain of knowledge with the help of formal concepts and properties linking these concepts (Gruber 1993). *Thesauruses and taxonomies* consist in notions or concepts which are rigorously defined and hierarchically structured, but do not use formal semantics. Semi-formal and shared knowledge representations, such as *Topic Maps* (Park & Hunting 2002) have also been proposed as an intermediary representation to formal ontologies where concepts, called topics, are defined in relation to others with hierarchical relations. In comparison with these knowledge representations, folksonomies can be seen as semiotic representations of the knowledge of a community, but *they do not include any semantic structure.*

In order to overcome the limitations of folksonomies that we mentioned above, it is possible to bridge ontologies and folksonomies. The idea is to semantically enrich folksonomies in order to discover links between the tags, and in the end, *between the users behind these tags* (linking the users by the tags is a very interesting way of building social graphs for enriching the social network models described in the section dedicated to the semantic network analysis). This bridging can be done in several ways which we detail in the next subsections.

## **Extracting semantics from folksonomies**

It is possible to take into account multiple dimensions of folksonomies as they consist in a triadic structure where *tags* are associated by *people* to *resources* (“who tags what with what”). This is what (Mika 2005) does, for instance, in order to extract *broader* and *narrower* relationships between tags and to build what he calls “lightweight ontologies”. One of the advantages of this type of approach is to decipher the semantics of folksonomies and to be able to more accurately build communities of interests, for instance by considering all the persons using the tag “music” and all the tags subsumed by music (such as “guitar”, or “saxophone” in a previous example).

The first step in this task is to measure the semantic relatedness between tags. *Since usually no explicit semantic relationships are given when users tag*, this relatedness has to be first computed by analyzing the tripartite structure of folksonomies. In table 1 we compare approaches of this type.

(Cattuto et al 2008) proposed semantically grounding the measures of tag relatedness and characterizing different types of similarity measures according to the type of semantic relationships to which they correspond. Thus, their method can be used to find related tags which share a subsumption relationship with a given tag *t*, however without being sure whether these related tags may subsume or be subsumed by tag *t*.

(Mika 2005) applied social network analysis on different projections of the tripartite structure of folksonomies. Then he grouped similar communities of interest, i.e., groups of people sharing common tags, in order to derive subsumption properties between the tags thanks to the inclusion of these communities of interest.

(Hotho et al 2006) adapted the *PageRank* algorithm to the case of folksonomies in order to find not only relationships between tags, but also between users and resources. (Schmitz 2006) used conditional probability methods to induce a hierarchy from Flickr tags. (Begelman and al. 2006) looked closely at the distribution of the co-occurring tags for a given tag, and computed the threshold above which co-occurring tags are strongly related to each other. Several other approaches use distributional measures with different contexts of aggregation of the folksonomy data. The idea is to project the tri-partite model of folksonomy into bi-partite representations by aggregating the data according to a given context. For instance the tag-tag context consists in looking at the association between a tag and its co-occurring tags. (Heymann & Garcia-Molina 2006) used the tag-resource context, while (Specia & Motta 2007) used the tag-tag context, and (Schwarzkopf et al. 2007) used a composite measure mixing the tag-tag context and the tag-user context. Finally (Cattuto et al. 2008) proposed an

analysis of the different context of distributional aggregation, while (Markines et al. 2009) proposed a new type of measure based on mutual information calculus, and a framework for analyzing the different types of similarity measures between resources and tags.

	Type of similarity	Subsumption relations	Clustering
Mika (2005)	Network based	yes	no
Hotho <i>et al.</i> (2006)	FolkRank	no	no
Schmitz (2006)	conditional probability	yes	no
Begelman <i>et al.</i> (2006)	co-occurrence	no	yes
Heymann & Garcia-Molina (2006)	distributionnal (resource context)	yes	no
Specia & Motta (2007)	distributional (tag context)	no	yes
Schwarzkopf <i>et al.</i> (2007)	composite	yes	no
Cattuto <i>et al.</i> (2008)	distributional (3 contexts)	yes	no
Markines <i>et al.</i> (2009)	mutual information	yes	no

Table 1. Comparison table of approaches extracting semantic relations between tags by analyzing the structure of folksonomies

## Semantically enriching folksonomies, structure the tags!

Even if ontologies and folksonomies remain different entities, several approaches have been proposed to semantically enrich folksonomies by adding a semantic layer, or by attempting to semantically structure them with the help of other already available ontologies, or by using the tags to bootstrap an ontology.

By adding structure to the tags, we add structure to the set of users who used these tags. *Remember that by linking tags, we link people.* If we use tags to bootstrap an ontology (for example by integrating the most popular tags into the ontology), or if we link tags to a domain ontology, we help structure the tags. More generally the usefulness of these approaches for semantic social network analysis is to connect the tags to other semantic resources, such as users, shared content, or members of other social data repositories in order to *build a graph of people who share the same interests*. In addition, once the semantics of folksonomies are better known, we can use formalisms or the tools of the semantic web to support folksonomy-based social platforms.

This type of approach consists in either (1) using ontologies to represent folksonomies and properties of tags (Gruber 2005), or (2) assisting users to semantically augment tags (Tanasescu & Streibel 2007), or (3) using ontologies to automate the semantic enrichment of folksonomies (Specia & Motta 2007), or (4) involving users in the semantic organization of tags. Then semantic web formalisms can help leverage the interoperability of the exchange of this additional knowledge. In Table 2 we compare these approaches.

**The main idea consists in constructing an ontology of folksonomies** to support more advanced uses of tagging (Gruber 2005). Thus, tags can have properties and relationships, and can be grouped in tag clouds, etc. This idea has been implemented by (Newman et al. 2005), and further improved by (Kim et al. 2007) who integrated their SCOT<sup>11</sup> ontology, another ontology that models users' interaction on social Web platforms with SIOC<sup>12</sup> (Breslin et al. 2005), another ontology that models users' interaction on social Web platforms. Later, (Passant & Laublet 2008) extended these interconnected schemas with MOAT<sup>13</sup>, an ontology linking tags with online resources to define precisely the meaning of tags and to tie them with the "Web of Linked Data"<sup>14</sup>, a vision of the Web where resources are linked with each other thanks to the concepts which can be attached to them.

11 Semantic Cloud Of Tags, <http://scot-project.org/>

12 <http://sioc-project.org/>

13 Meaning Of A Tag

14 <http://esw.w3.org/topic/SweoIG/TaskForces/CommunityProjects/LinkingOpenData/>

**Research using the previous idea focused on user intervention in the process of semantically enriching folksonomies.** Huynh-Kim Bang et al. (2008) proposed the concept of structurable tags where users can add specific tags corresponding to semantic relationships between tags (such as “france” < “europe” which means “france” is narrower than “europe”). (Tanasescu & Streibel 2007) suggested letting the users tag the links existing between tags. The two latter approaches do not make direct use of semantic web formalisms, as they focus more on the flexibility of the system than on the logical consistency of the knowledge structure obtained. (Passant 2007) developed a semantically augmented corporate blog where users can attach their tags to the concepts of a centrally maintained ontology, while (Good et al. 2007) suggest terms from professional vocabularies fetched online at tagging time. Thanks to the two latter types of approaches, ambiguous tags can be associated to clearly defined concepts by the users while tagging, solving one of the limitations of folksonomies.

**Other research works proposed automating (even partially) the semantic enrichment of folksonomies. For example by applying several types of semantic processing, such as finding equivalent tags or grouping similar tags based on similarity measures.** (Specia & Motta 2007) have developed such a system; they query ontologies on the semantic web and try to match the tags from these clusters with concepts from ontologies in order to link the tags with semantic relationships. The main limitation of such an approach is the limited coverage of currently available ontologies. Similarly, (Tesconi et al. 2008) and (Ronzano et al. 2008) first built sets of terms-meaning by mining Wikipedia, and then linked each tag of a sample of delicious.com users to a unique meaning. The main difference between these two latter types of methods is that (Specia & Motta 2007) apply the mapping of tags with semantic resources on clusters of related tags, whereas (Tesconi et al. 2008) consider sets of tags belonging to the same user. The semantic enrichment of tags proposed by (Specia & Motta 2007) can be used by all the contributors of a folksonomy, and may be useful to a whole community. The tag disambiguation of (Tesconi et al. 2008) can be applied to different purposes, such as the profiling of the tagging of a user, providing for richer information when consulting the bookmark database of this user. However, if we apply the algorithm proposed by (Tesconi et al. 2008) to all the users of a community, *we can measure or detect the divergences existing among the users* and, for instance, propose discussing their points of view in the case of the collaborative construction of an ontology. (Van Damme et al. 2007), along the same lines, suggest integrating as many semantic online resources as possible, and, at the same time, integrating user intervention to build, at a reasonable cost, genuine “folksonomies”.

**The collaborative aspects of the semantic enrichment of folksonomies have been addressed by other approaches focused on ontology maturing processes.** The idea is to involve users in the semantic organization of tags so that the tags in the folksonomy will better suit the user needs than purely automatic approaches. Web 2.0 tools are used to achieve this task, such as wikis (Buffa et al. 2008), blogs (Passant 2007), e-learning platforms (Torniai et al. 2008), personal knowledge organizers (Abbattista et al. 2007), or social bookmarking sites (Braun et al. 2007). Following the distinctions brought by (Weller & Peters 2008) between the individual and the collective level at which folksonomies can be modified, we can distinguish approaches where the users merely propose new concepts to an existing ontology (Passant, 2007), with approaches where users can directly edit the whole shared ontology (Braun et al., 2007). These approaches raise also the problem of the user-friendliness of the interfaces used to edit tags and their semantic relations to other tags, as this task requires time and skills.

Another great benefit of combining ontologies and folksonomies lies in the interoperability brought by the formalism of the semantic web. The Linking Open Data project<sup>15</sup> consists in extending the Web with semantically interconnected data sources and which publish varied open data sets in RDF format following a set of ontologies describing the different types of

---

<sup>15</sup> <http://esw.w3.org/topic/SweoIG/TaskForces/CommunityProjects/LinkingOpenData/>

resources. Ontologies from the Linking Open Data initiative include ontologies like SIOC<sup>3</sup>, used to describe online communities' exchanges or SKOS<sup>4</sup>, used to describe thesauruses (see chapter "Understanding Online Communities Using Semantic Web Technologies") for more details on this aspect of the use of semantic web formalisms to empower social data repositories)

	User intervention	Ext. resources	Automatic	Sem. Web
Gruber (2005)	-	no	no	yes
Newman <i>et al.</i> (2005)	-	no	no	yes
Tanasescu & Streibel, (2007)	yes	no	no	no
Huynh-Kim Bang <i>et al.</i> (2008)	yes	no	no	no
Breslin <i>et al.</i> (2005), Kim <i>et al.</i> , (2007)	-	no	no	yes
Passant & Laublet (2008) Good <i>et al.</i> (2007)	yes	yes	no	yes
Specia & Motta (2007), Angeletou <i>et al.</i> (2008)	no	yes	yes	yes
Tesconi <i>et al.</i> (2008), Ronzano <i>et al.</i> (2008)	no	yes	yes	yes
Van Damme <i>et al.</i> (2007)	yes	yes	yes	yes
Braun <i>et al.</i>	yes	no	no	yes

Table 2. Comparison table of the approaches to enriching folksonomies which (1) exploit user intervention, and/or (2) make use of external semantic resources, and/or (3) seek the automation of the process (automatic), and/or (4) are based on semantic web formalisms.

## Concrete example: a tagging system for collaboratively building a thesaurus and for identifying a network of experts

In this section we present our approach to the semantic enrichment of folksonomies which we have applied to the evolution of a thesaurus within a French organization. It involves a social bookmarking application similar to delicious.com but adds some simple features for helping to classify the tags. We will show that a very simple application that requires little effort by users can help structure the folksonomy and build a thesaurus. *A very interesting consequence is that it also helps in building a network of experts.*

### Motivating scenario

Our scenario takes place within the French Agency for the Environment (ADEME<sup>16</sup>). In this organization, there is a distributed network of experts who publish, share and exploit resources. The goal of our collaboration with this organization is to help them improve the indexing of these resources thanks to a combination of bottom up approaches (like folksonomies) and semantic tools. In order to involve all the users in the indexing process, we designed a method based on the semantic enrichment of folksonomies. This method consists in associating the power of automatic handling of folksonomies and the expertise of users by integrating simple semantic functionality within the interface of the system. The result of this approach is a set of tags linked with semantic relationships (such as *broader*, *narrower*, or *related*) that can be connected to some nodes of the existing thesaurus thanks to ontology matching techniques (Euzenat & Shvaiko 2007). The tags which are not matched but which are semantically connected with tags that have been matched can then be proposed to the maintainer of the thesaurus as new concepts (new candidates for the integration into the ontology). In addition, our model supports confrontational views so that any user can propose semantic relationships (on the basis of automatic suggestions); divergences may arise and *can be an interesting opportunity to discover different sub-communities of interests.*

<sup>16</sup> ADEME (www.ademe.fr) has a distributed network of experts who compile data related to renewable energies (in particular for home use). They also answer questions (by email, phone) and exploit a knowledge base with simple keyword-based queries. Data is indexed using a thesaurus whose evolution is problematic.

## Semantic enrichment of folksonomies

Our approach consists in combining automatic processing of the folksonomy and semantic functionalities integrated within a navigation interface in order to assist the users in contributing to the semantic enrichment of the folksonomy.

One of the widely known limitations of folksonomies is the handling of the spelling variations between supposedly equivalent tags such as “neighbour” and “neighbor”. A simple solution to this problem consists in measuring the editing distance between these tags, such as the Levenshtein distance (Levenshtein 1966), and to identify as equivalent tags the ones whose distance is below a given threshold value. Another type of analysis consists in measuring the “similarity distance” between all the tags thanks to an analysis of the links between the tags, the users, and the tagged resources in a folksonomy. This type of handling corresponds to the solutions proposed by (Markines et al. 2009), among others. We have implemented in our system the distributional measure based on the tag-tag context.

This automatic handling is then used by functionalities such as the detection of spelling variants of tags and the possibility of related tags. These functionalities are suggested by the interface to induce users to validate, reject or correct the automatic processing.

## Implementation

The system in which we have implemented our ideas is a bookmark navigator which includes extra functionalities such as the extension of tag queries with spelling variants, and the suggestion of related tags, plus the possibility of editing these semantic relations (for a detailed presentation, see Limpens et al., 2009). Our system is composed of: (1) automatic agents applying semantic processing to folksonomies, and (2) a user interface to browse the bookmark database, and at the same time, to validate or correct the automatically suggested tags and semantic relationships.

In our model every assertion is attached to a user, recorded, and added to the database, even when it contradicts other assertions (for example the assertion “pollution” is related to “car”, has been approved by John, and rejected by Paul). This feature has the advantage of collecting all users’ contributions and letting diverging points of view coexist, each user benefiting from their own structuring of the folksonomy plus the contributions of others when they are not confrontational.

Since our model is described with semantic web formalisms (as an RDF schema), the discovery of conflicting relationships is straightforward and can make use of inference capabilities through SPARQL queries. Thus it is possible, for instance, *for a given user to know who are the other users who agreed with him on semantic relations he made on his tags.*

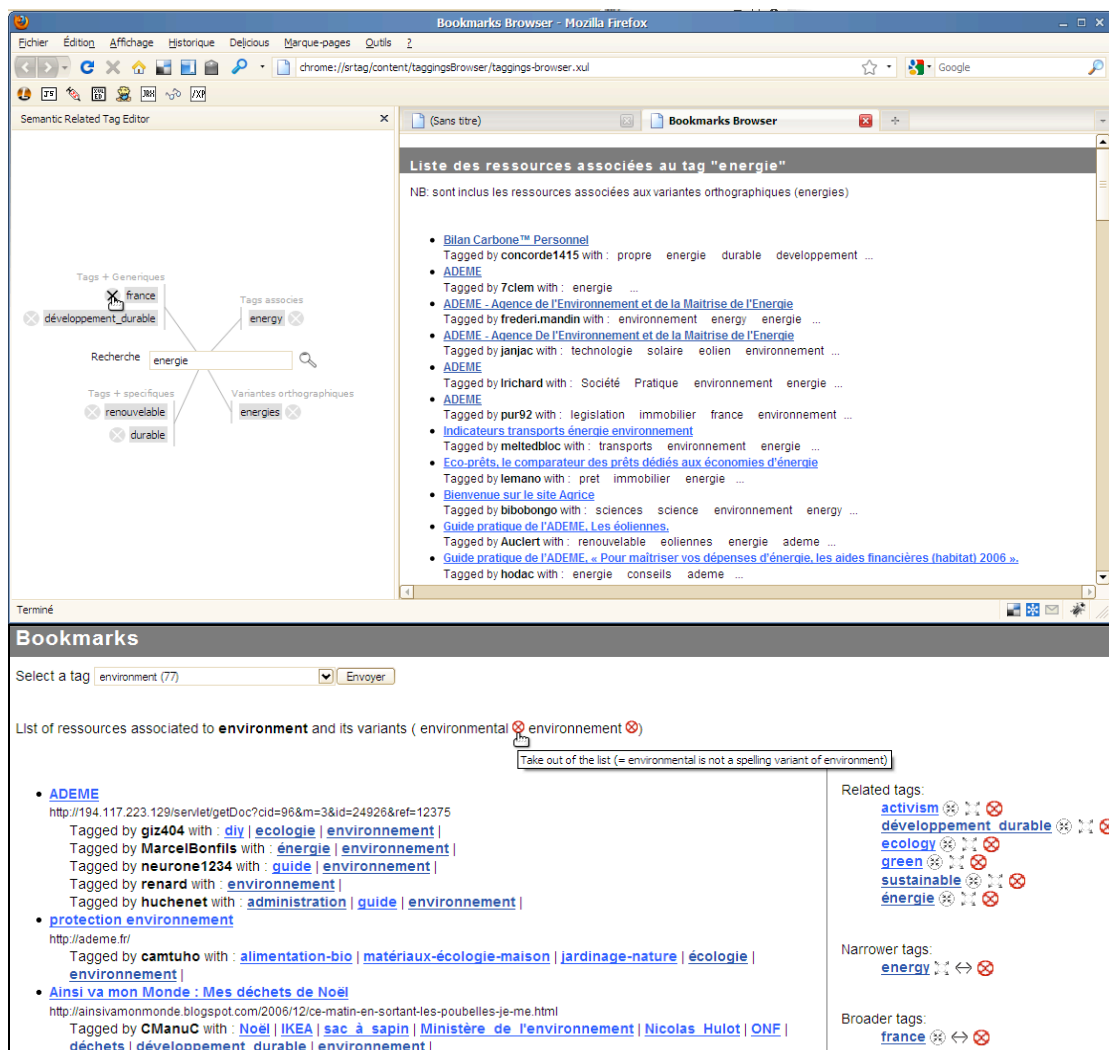
The administrators of the system can further exploit these results in different ways. The different points of views arising within the community can be highlighted thanks to the mechanism described above. For instance, the point of view of the “car’s opponents”, and the point of view of the “car’s defenders” can be highlighted if there is a conflict or an agreement in the semantic relationship that links “car” with “pollution” for example. *The hypothesis we make here is that when someone puts some effort into semantically structuring a tag, this implies a stronger commitment than mere tagging and can be a good indicator of a strong interest or an expertise in the domain described by this tag.*

## Towards novel exploitation of the semantics of social data

We have seen in this section how semantically enriching folksonomies can improve semantic social network analysis by providing additional links between tags, and thus, between people using these tags. We have presented the state of the art on bridging folksonomies and ontologies. Since folksonomies consist of the collection of the taggings by users, that is, the

association of freely chosen keywords to resources, they can connect users together through the use of the same tags or the tagging of the same resources.

Semantically enriching folksonomies can further enhance the ability to connect people via tags by discovering links between different tags which are not necessarily used for the same resources (such as “pollution” and “CO2” in the previous examples). We have also proposed a novel method to assist with automatically handling the semantic organization of folksonomies. This method consists in automatically proposing semantic relations between tags (such as “spelling variant” or “related”), and letting users validate or correct them, or even proposing new semantic relations thanks to functionalities embedded in the browsing interface (see *Figure 5*). *The results can then be exploited to highlight sub-communities of interest via the divergence or convergence between the semantic relations validated or rejected by the users. For instance if a group of users agreed on semantically connecting the tag “car” with the tag “pollution”, we can infer that they share the same view on the role of cars in pollution problems.*



*Figure 5 : Screenshot of our early interface for navigating a bookmarks database and validating or proposing semantic relations between tags.*

Adding semantics to social data such as tagging data and folksonomies can greatly enhance business intelligence processes by helping the discovery of weak signals and the deciphering of links. Indeed, mere folksonomies and the classical tag cloud visualization have the tendency to hide rarely used notions since highlighted terms are the most popular ones. In our concrete example, if a single user proposes a semantic relation between rarely used tags and

more broadly used tags, this small piece of information can benefit the whole community and render visible emergent notions more quickly.

Coming back to our initial scenario of business intelligence, a clear stake of leveraging social applications to capture and organize folksonomies is the potential of turning every user into a watcher, a contributor to business intelligence, a sensor and a categorizer, and all this, ideally, as a side effect of her day-to-day tasks such as bookmarking a resource or searching for a bookmarked resource. Now that we can capture and organize information resources and the experts who find them or who monitor them, we need to capture and analyze the networks of these experts, be they explicit or implicit.

## SEMANTIC SOCIAL NETWORK ANALYSIS

We saw in previous sections that we can represent user interaction on social web sites using several ontologies, both for representing the explicit part of the social network (network of friends, etc.) but also for building graphs of users based on other implicit markers. In particular, we focused on the semantic enrichment of the folksonomies that can be used to identify communities of interest. Once we have such graphs, we can analyse them via social network analysis (SNA).

SNA tries to understand and exploit the key features of social networks in order to manage their life cycle and predict their evolution. Much research has been conducted on SNA using graph theory (Scott 2000) (Wasserman et al 1994). Among important results is the identification of sociometric features that characterize a network. SNA metrics can be decomposed into two categories; (1) some provide information on the position of actors and how they communicate and (2) others give information on the global structure of the social network.

**Centrality highlights the most important actors and the strategic positions of the network** - three definitions have been proposed (Freeman 1979). Degree centrality considers nodes with high degrees (number of adjacent edges) as most central. It highlights the local popularity of the network, actors that influence their neighbourhood. In directed graphs the in-degree and out-degree (number of in-going and out-going adjacent edges) are alternative definitions that take into account the direction of edges, representing respectively the support and the influence of the actor. The  $n$ -degree is an alternative definition that widens the neighbourhood considered to a distance of  $n$  or less (the distance between two actors is the minimum number of relationship that link them). 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 and to join others actors. The direction of edges also modifies the interpretation of the closeness centrality by differentiating the capacity to join or to be reached. Betweenness centrality focuses on the capacity of an actor to be an intermediary between any two others. 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 (Burt 1992) (Holme 2002)(Burt 2004). Its exact computation is time consuming, several algorithms tackle this problem (Freeman et al 1991) (Newman 2001) (Latora et al 2007) (Brandes 2001) with a minimum time complexity of  $O(n.m)$  -  $n$  is the number of nodes and  $m$  the number of edges. To deal with large networks, approximating algorithms (Radicchi et al 2004) (Brandes et al 2007) (Bader et al) (Geisber et al 2008) and parallel algorithms (Bader et al 2006) (Santos et al 2006) have been proposed.

**Other metrics help understanding the global structure of the network.** The density indicates the cohesion of the network, i.e., the number of relationships expressed as a proportion of the maximum possible number of relationships ( $n*(n-1)$ , with  $n$  the number of actors). The diameter is the length of longest geodesics of the network (a geodesic is a shortest sequence of linked actors between two actors). Community detection helps understanding the distribution of actors and activities in the network (Scott 2000), by detecting groups of densely connected actors. The community structure influences the way information is shared and the way actors behave (Burt 1992) (Burt 2001) (Burt 2004)



(Coleman 1988). (Scott 2000) gives three graph patterns that correspond to cohesive subgroups of actors playing an important role in community detection: components (isolated connected sub graphs), cliques (complete sub graphs), and cycles (paths returning to their point of departure). Alternative definitions extend these initial concepts that are too restrictive for social networks. The members of an  $n$ -clique have a maximum distance of  $n$  to any other member of the group, and a member of a  $k$ -plex must be connected to all the members of the group except a maximum number of  $k$  actors. However, these extensions, still not adapted to social network structure and other criteria of cohesiveness, are proposed by community detection algorithms. Community detection algorithms are decomposed into two categories, either hierarchical or based on heuristics (Newman 2004) (Givan et al 2004) (Danon et al 2005). Two strategies are used in hierarchical algorithms: divisive algorithms consider the whole network and divide it iteratively into sub communities (Girvan et al 2002) (Wilkinson et al 2003) (Fortunato et al 2004) (Radicchi et al 2004), and the agglomerative algorithms group nodes into larger and larger communities (Donetti et al 2004) (Zhou et al 2004) (Newman 2004). Other algorithms are based on heuristics such as random walk, analogies to electrical networks (Wu et al 2004) (Pons et al 2005).

Social network graphs hold specific patterns that can be used to characterize them (Newman 2003) and accelerate algorithms. The degree distribution follows a power law, few actors have a high degree and many have a low one. According to the small world effect (Milgram 1967), the diameter in a social network with  $n$  actors is of the order of  $\log(n)$ . Social networks have an important clustering tendency forming a community structure due to a high transitivity in relationships (if Jack knows Paul and Paul knows Peter there is a good chance that Jack knows Peter or will meet him) (Newman 2003). This clustering tendency correlates with the assortativity that refers to the preference for actors of a social network to be linked to others who have similar characteristics. The size of the largest component is an indicator of the communication efficiency of the network, the more actors it contains the better the communication. In most of web 2.0 sites, the size of the largest component is of the order of the size of their social network as they are focused on user communication and centred on a viral diffusion of their content.

These algorithms are only concerned with graph structure – they all lack semantics, and have an especially poor exploitation of the types of relations. There is a need for interoperable tools and languages that could help taking into account semantics and typing. Ontologies based on semantic web standards emerged these last years to help deal with such problems. Millions of FOAF profiles (Golbeck et al 2008) are now published on the web, due to the adoption of this ontology by web 2.0 platforms with large audiences ([www.livejournal.net](http://www.livejournal.net), [www.tribe.net](http://www.tribe.net)). SIOC exporters are also proposed and available in widely deployed social applications such as blogs (e.g., SIOC plugin for Wordpress). The adoption of standardized ontologies for online social networks will lead to increasing interoperability between them and to the need for uniform tools to analyze and manage them. Consequently, some researchers have applied classical SNA methods to the graph of acquaintance and interest networks respectively formed by the properties "foaf:knows" and "foaf:interest" to identify communities of interest from the network of LiveJournal.com (Paolillo et al 2006). (Golbeck et al 2003) studied trust propagation in social networks using semantic web frameworks. (Golbeck et al 2008) worked on merging FOAF profiles and identities used on different sites. In order to perform these analyses, they chose to build *their own, untyped graphs* (each corresponding to one relationship "knows" or "interest") from the richer RDF descriptions of FOAF profiles. Too much knowledge is lost in this transformation and this knowledge could be used to parameterize social network indicators, improve their relevance and accuracy, filter their sources and customize their results. Others researchers (San Martin et al 2009) have shown that SPARQL is well suited for performing modifications on a social network but that it can't deal with global queries currently used in social network analysis (e.g., diameter, density, centrality, that require complex path computations). Consequently, researchers have extended the standard SPARQL query language in order to find paths between semantically linked resources in RDF-based graphs (Alkhateeb et al 2007) (Anyanwu 2007) (Kochut & Janik

2007) (Corby 2008) (Pérez et al 2008). In the next section, this work is used as a basis to work on graph-based and ontology-based social network representation and analysis.

## Analyzing Online Social Networks with Semantic Web Frameworks

We have designed a framework to analyse online social networks based on semantic web frameworks. Figure 6 illustrates the abstraction stack we follow. We use the RDF graphs to represent social networks, and we type those using existing ontologies together with specific domain ontologies if needed. Some social data are already readily available in a semantic format (RDF, RDFa, microformats, etc.). However, today, most of the data are still only accessible through APIs, see examples in (Rowe and Ciravegna 2008), or by crawling web pages and need to be converted. To enhance these social network representations with SNA indices, we have designed SemSNA (Figure 7), an ontology that describes the SNA notions, e.g., centrality. With this ontology, we can (1) abstract social network constructs from domain ontologies to apply our tools on existing schemas by having them extend our primitives; and we can (2) enrich the social data with new annotations (see Figure 8) such as the SNA indices that will be computed. These annotations enable us to manage more efficiently the life cycle of an analysis, by pre-calculating relevant SNA indices and updating them incrementally when the network changes over time. We propose SPARQL formal definitions of SNA operators improving the semantics of the representations. The current test uses the semantic search engine Corese (Corby et al 2004) that supports powerful SPARQL extensions particularly well suited for SNA features such as path computations (Corby et al 2008).

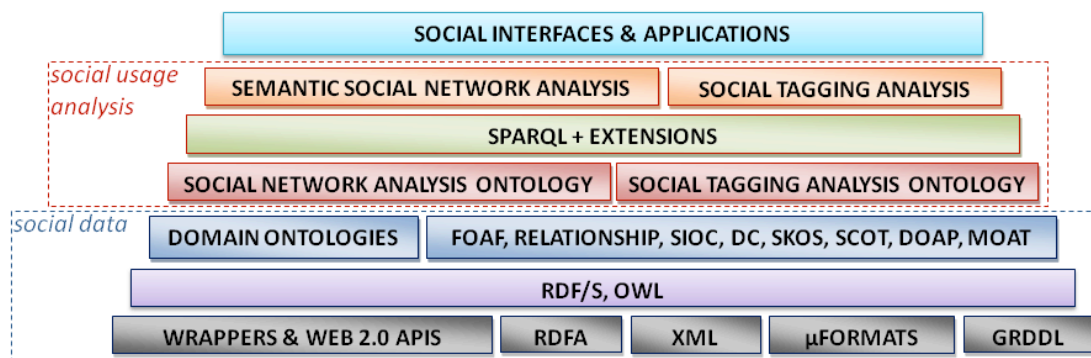


Figure 6 : Abstraction stack for social data analysis

## SemSNA: an Ontology of Social Network Analysis

SemSNA<sup>17</sup> (Figure 7) is an ontology that describes concepts of social network analysis with respect to the semantics of the analyzed relationships. First, we present the basic concepts that can be extended to integrate any SNA features and then we present different primitives that extend this basis to annotate social networks with popular SNA metrics.

The main class `SNAConcept` is used as super class for all SNA concepts. The property `isDefinedForProperty` indicates for which relationship (i.e. sub-network) an instance of SNA concept is defined. An SNA concept is attached to a social resource with the property `hasSNAConcept`. The class `SNAIndice` describes valued concepts such as centrality, and the associated value is set with the property `hasValue`. As an example, with this basis a general declaration of a valued concept will be:

```
<http://www.inria.fr/John> hasSNAConcept _:a
_:a hasValue 12
_:a isDefinedForProperty "foaf:knows"
```

<sup>17</sup> <http://ns.inria.fr/semsna/2009/06/21/voc>

A set of primitives can be used to annotate positions in the network based on centrality. The class `Centrality` is used as a super class for all centralities defined by the classes `Degree`, `InDegree`, `OutDegree`, `Betweenness`, `BetweennessCentrality` and `ClosenessCentrality`. The property `hasCentralityDistance` defines the distance of the neighbourhood taken into account for a centrality measure.

Next a set of primitives are proposed for metrics on the global structure of the social network. Primitives are defined to annotate groups of resources linked by particular properties. The class `Group` is a super class for all classes representing any definition of groups of resources. The class `Component` represents a set of connected resources. The class `StrongComponent` defines a component of a directed graph where the paths connecting its resources don't contain any change of direction. The `Diameter` subclass of `Indice` defines the length of the longest of the shortest paths of a component. The property `maximumDistance` enables us to restrict the membership to components with a maximum path length between members. A clique is a complete sub graph for a given property according to our model. An  $n$ -clique extends this definition with a maximum path length ( $n$ ) between members of the clique; the class `Clique` integrates this definition, and the maximum path length is set by the property `maximumDistance`. Resources in a clique can be linked by shortest paths going through non clique members. An `NClan` is a restriction of a clique that excludes this particular case. As `KPlex` relaxes the clique definition to allow connecting to  $k$  members with a path longer than the clique distance,  $k$  is determined by the property `nbExcept`. Finally the concept `Community` supports different community definitions:

`InterestCommunity`, `LearningCommunity`,  
`GoalOrientedCommunity`, `PraticeCommunity` and `EpistemicCommunity`  
 (Conein 2004) (Henri et al 2003). These community classes are linked to more detailed ontologies, such as used by (Vidou et al 2006) to represent communities of practice.

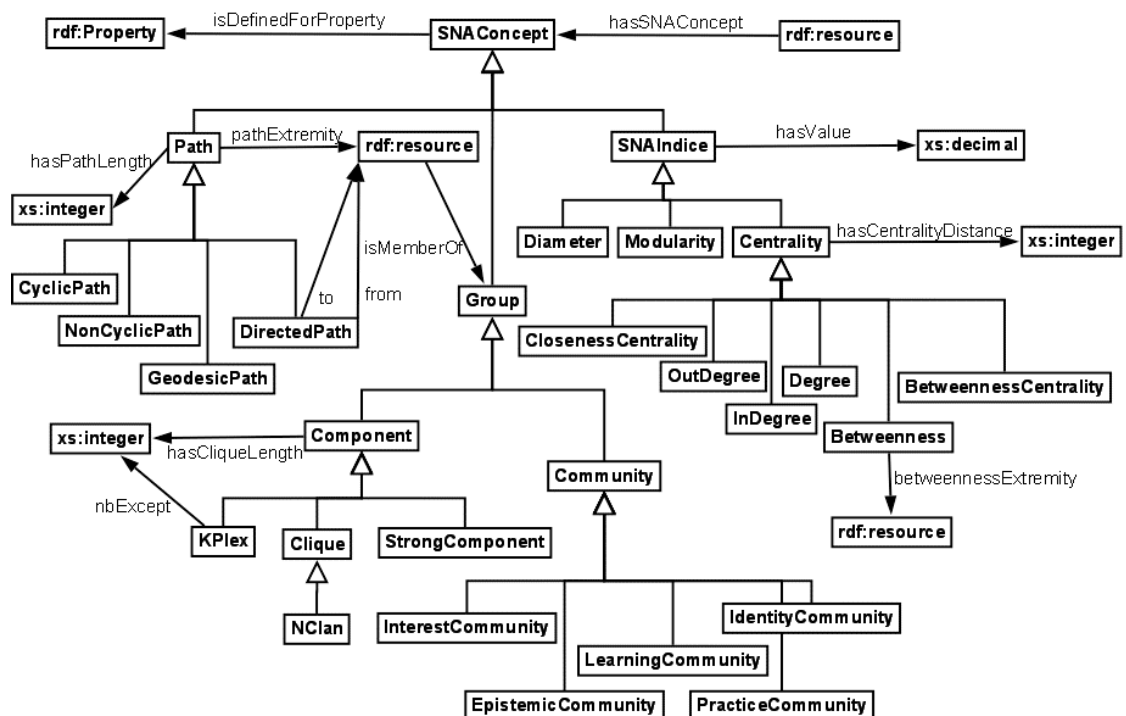


Figure 7 : Schema of SemSNA: the ontology of social network analysis

With this ontology we can enrich the RDF description of social data with SNA metrics that are semantically parametrized (Figure 8). These annotations are useful 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. Moreover, using a schema to add the results of our queries (rules) to the network also allows us to decompose complex processing into two or more stages and to factorize some computation among different

operators, e.g., we can augment the network with in-degree calculation and betweenness calculation and then run a query on both criteria to identify some nodes (e.g., what are the nodes with  $\text{indegree} > y$  and  $\text{betweenness} > x$  ?).

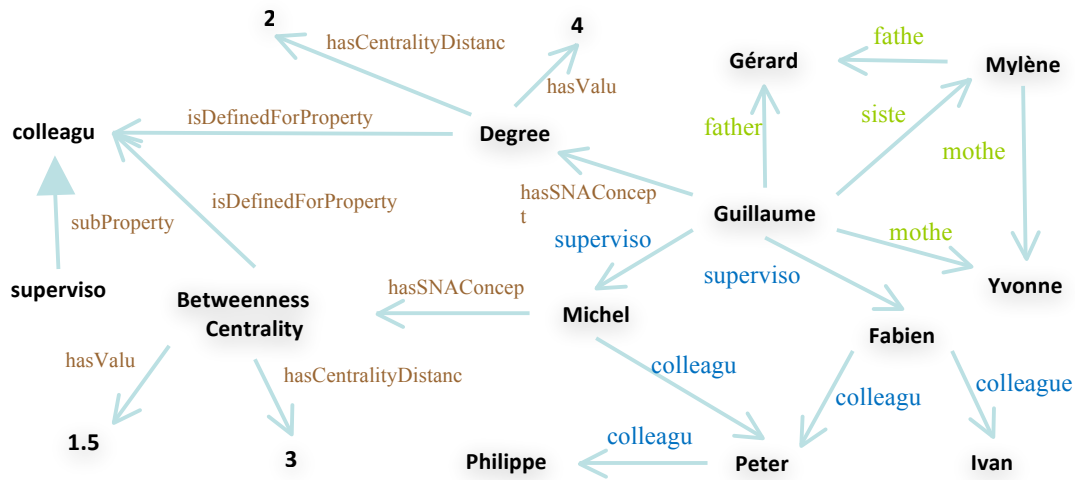


Figure 8 : social network enhanced with SemSNA indices (Degree, Betweenness).

### Querying and transforming the social network with SPARQL

Based on our model, we propose SPARQL formal definitions to compute semantically parametrized SNA features and to annotate the graph nodes, 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 require path computations (Corby 2008). In (San martin et al 2009), researchers have shown that SPARQL "is expressive enough to make all sensible transformations of networks". However, this work also shows that SPARQL is not expressive enough to meet SNA requirements for global metric querying, e.g., density, of social networks. Such global queries are mostly based on result aggregation and path computation which are missing from the standard SPARQL definition. The Corese search engine provides such features with result grouping, aggregating functions like  $\text{sum}()$  or  $\text{avg}()$  and path retrieving (Corby et al 2008) (Erétéo et al 2009). Moreover, inheritance relations are natively taken into account when querying the RDF graph in SPARQL with CORESE. Thus parametrized operators formally defined in SPARQL allow adjusting the granularity of the analysis of interactions/relations while classical SNA ignores the semantics of richly typed graphs like RDF. The Figure 9 illustrates the calculation of a parametrized degree where only family relations are considered by exploiting the hierarchy of relationships).

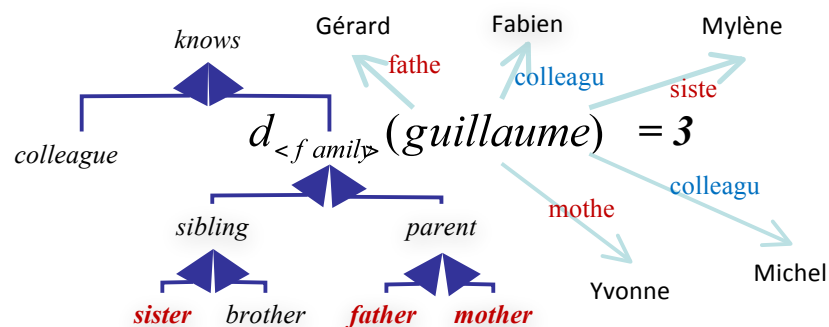


Figure 9 : A Parametrized degree that considers a hierarchy of relations.

Different SPARQL queries, exploiting Corese features, are presented in (Erétéo et al 2009) to perform social network analysis combining structural and semantic characteristics of the network. This approach is easily extensible as other queries can be defined at anytime, to compute new operators. As a simple example, the parametrized degree described in Figure 8 is computed with the following query in Corese:

```
select ?y count(?x) as ?degree where {
  { ?x rel:family <http://inria.fr/guillaume>::?y }
  UNION
  { <http://inria.fr/guillaume>::?y rel:family ?x }
} group by ?y
```

In order to be exploited in web services to leverage the social experience, these queries must be applied in batch on a large number of stored RDF triples. Consequently the social data are enhanced with the results of these parametrized SNA metrics using the SemSNA ontology to provide services based on this analysis (e.g., filter social activity notifications), to use them in the computation of more complex indices or to support iterative or parallel approaches in the computation.

Corese is a freeware that can handle millions of nodes but other engines with the same extensions could be used just as well. The W3C SPARQL Working Group<sup>18</sup> is currently investing some of the extensions that are presented in (Erétéo et al 2009), such as project expression, aggregation, group by and property paths. ARQ<sup>19</sup>, PSPARQL<sup>20</sup> and SPARQLeR (Kochut and Janik 2007) also implement property paths. However, some necessary extensions are unique to Corese, like the `group by any` statement that groups results that share a value through any variable, computing connected results.

Inside companies, these operators can analyze in real time or in batch the expert networks of the organization and its projects, providing a directory of the relevant persons to contact for every field of interest it is involved in. Leveraging both graphs (structured folksonomies and social networks) and the semantics of the schema, parametrized operators can produce reports and snapshots of the current assets and trends of the activity of the company, its markets and its competitors. But all this formalized knowledge can also be used in production rules to automatically produce new knowledge with potentially high added value as we will see in the next section.

## TRANSFORM, ENRICH AND WRAP SOCIAL DATA

Semantic web frameworks offer different ways to enrich RDF data with reasoning mechanisms. We first investigate how to infer new knowledge from an ontology by defining rules and schema properties. Then we'll see how SPARQL enables us to generate RDF by performing queries with a CONSTRUCT clause and its extension in Corese to leverage such features.

The OWL schema "specifies property characteristics, which provides a powerful mechanism for enhanced reasoning about a property"<sup>21</sup>. New properties can be defined automatically and inconsistencies among data can be easily inferred. For example, a property *family* can be defined as symmetric and transitive, and inferring on social data containing *Paul family Jack* and *Jack family Peter* will produce the knowledge *Jack family Paul*, *Peter family Jack*, *Paul family Peter* and *Peter family Paul*. The Figure 10 : Owl in One picture summarizes the characteristics that can be defined on properties with OWL.

---

<sup>18</sup> [http://www.w3.org/2009/sparql/wiki/Main\\_Page](http://www.w3.org/2009/sparql/wiki/Main_Page)

<sup>19</sup> [http://jena.sourceforge.net/ARQ/property\\_paths.html](http://jena.sourceforge.net/ARQ/property_paths.html)

<sup>20</sup> <http://exmo.inrialpes.fr/software/psparql/>

<sup>21</sup> <http://www.w3.org/TR/2004/REC-owl-guide-20040210/#SimpleProperties>

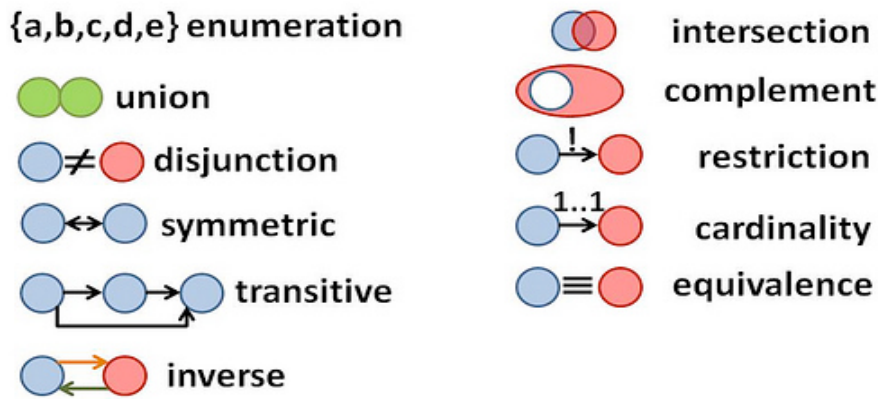


Figure 10 : Owl in One picture <sup>22</sup>

Other pre-processing can also enrich the semantics, such as rules crawling the network to add types or relations whenever they detect a pattern, e.g., every actor frequently commenting resources or posts by another actor is linked to him by a relation “monitors”. Corese can automate some transformations with inference rules (Corby et al 2002). As an example we can infer a property `SemSNI:hasInteraction` (SemSNI<sup>23</sup> is an ontology of Social Network Interaction, see (Erétéo et al 2009)) between two actors when one has commented on the other's resource using the following rule

```
<cos:if>
  { ?doc sioc:has_creator ?person1 .
    ?doc sioc:has_reply ?comment .
    ?comment sioc:has_creator ?person2 }
</cos:if>
<cos:then> {?person1 semsni:hasInteraction ?person2 } </cos:then>
```

The preceding syntax is specific to Corese but the Rule Interchange Format<sup>24</sup> (RIF) proposes XML dialects for exchanging rules on the semantic web and providing interoperability between the different inference engines. These dialects include in particular Basic Logic Dialect (BLD) and Production Rule Dialect (PRD).

Another tool to leverage the social network representation is to process it with a SPARQL query using a construct block to generate RDF and enrich the social data with it (San martin et al 2009) (Erétéo et al 2009). The following query produces the same result as the preceding example with a Corese rule:

```
CONSTRUCT {?person1 semsni:hasInteraction ?person2}
WHERE {
  ?doc sioc:has_creator ?person1 .
  ?doc sioc:has_reply ?comment .
  ?comment sioc:has_creator ?person2 }
```

Such queries produce RDF triples in respect with the construct block, which can be stored next. Corese enables us to re-inject the knowledge produced directly into the knowledge base with an add clause. The following example highlights the enrichment of a social network, using SemSNA, with degrees computed in the select clause:

```
ADD { ?y semsna:hasSNAConcept :_b .
      :_b rdf:type semsna:Degree .
      :_b semsna:isDefinedForProperty rel:family .
      :_b semsna :hasValue ?degree}
```

<sup>22</sup> Slide, Owl in one, by F. Gandon, <http://twitpic.com/60pdy>

<sup>23</sup> SemSNA is an ontology that describes concepts of Social Network Analysis, while SemSNI is used for representing interactions in a social network. For example, SemSNA can be used to compute centrality of nodes in a social network, nodes linked together using relations inferred from interactions form SemSNI (i.e network of people who commented a same resource).

<sup>24</sup> RIF Working Group [http://www.w3.org/2005/rules/wiki/RIF\\_Working\\_Group](http://www.w3.org/2005/rules/wiki/RIF_Working_Group)

```

SELECT ?y count(?x) as ?degree where {
  { ?x rel:family ?y } UNION { ?y rel:family ?x }
}group by ?y

```

## Wrap XML and SQL social data with RDF

We used Corese to query social data stored in a relational database or in XML (most of web 2.0 social data are exposed in XML through restful APIs) and to turn it into RDF/XML. While some researchers, like (Waseem et al. 2008) proposed a solution with the XSPARQL language for turning XML data into RDF, without the need for costly XSLT transformations, Corese proposes a different approach: an extension that enables us to nest an SQL query or an XQuery within SPARQL (Corby et al 2009). This is done by means of the `sql()` (respectively `XPath`) function that returns a sequence of results for each variable in the SQL select clause (respectively result of the node-set). Corese proposes an extension to the standard SPARQL select clause that enables binding these results to a list of variables. In the following example, we show how we retrieve the friend relationships from the relational database, using this `sql()` function:

```

construct { ?id1 foaf:knows ?id2 }
select  sql(<server>, <driver>, <user>, <pwd>, 'SELECT
user1_id, user2_id from relations')
as (?id1, ?id2)
where { }

```

## Experiment on a Real Online Social Network

We conducted an experiment on an anonymized dataset of Ipernity.com<sup>25</sup>, one of the largest French social networks centered on multimedia sharing. This dataset contains 61,937 actors, 494,510 declared relationships of three types and millions of interactions (messages, comments on resource, etc.). Ipernity.com, proposes to its users several options for building their social network and sharing multimedia content. Every user can share pictures, videos, music files, create a blog, a personal profile page, and comment on other's shared resources. Every resource can be tagged and shared. For building the social network, users can specify the type of relationship they have with others: friend, family, or simple contact (like a favourite you follow). Relationships are not symmetric, *Fabien* can declare a relationship with *Michel* but *Michel* can declare a different type of relationship with *Fabien* or not have him in his contact list at all; thus we have a directed labelled graph. Users have a homepage containing their profile information and pointers to the resources they share. Users can post on their profile and their contacts' profiles depending on the access rights. All these resources can be tagged including the homepage. A publisher can configure the access to a resource to make it public, private or accessible only for a subset of its contacts, depending on the type of relationship (family, friend or simple contact), and can monitor who visited it. Groups can also be created for topics of discussion with three kinds of visibility, public (all users can see it and join), protected (visible to all users, invitation required to join) or private (invitation required to join and consult).

We analyzed the three types of relations separately (*favourite*, *friend* and *family*) and also used polymorphic queries to analyze them as a whole using their super property: `foaf:knows`. We also analyzed the interactions produced by exchanges of private messages between users, as well as those produced by someone commenting someone else's documents.

We first applied quantitative metrics to get relevant information on the structure of the network and activities: the number of links and actors, the *components* and the *diameters*. 61,937 actors are involved in a total of 494,510 relationships. These relationships are decomposed into 18,771 *family* relationships between 8,047 actors, 136,311 *friend* relationships involving 17,441 actors and 339,428 *favourite* relationships for 61,425 actors.

<sup>25</sup> <http://www.ipernity.com>

These first metrics show that the semantics of relations are globally respected, as family relations are less used than friendship and favourite. 7,627 actors have interacted through 2,874,170 comments and 22,500 have communicated through 795,949 messages. All these networks are composed of a largest *component* containing most of the actors (fig 5) and a few very small components (less than 100 actors) that show "the effectiveness of the social network at doing its job" (Newman 2003), i.e., at connecting people. The interaction sub networks have a very small diameter (3 for comments and 2 for messages) due to their high density. The *family* network has a high diameter (19), consistent with its low density. However the *friend* and *favourite* networks have a low density and a low diameter revealing the presence of highly intermediary actors.

The betweenness and degree centralities confirm this last remark. The *favourite* network is highly centralized, with five actors having a betweenness centrality higher than 0, with a dramatically higher value for one actor: one who has a betweenness centrality of 1,999,858 while the other 4 have a value comprised between 2.5 and 35. This highest value is attributed to the official animator of the social network who has a *favourite* relationship<sup>26</sup> with most actors of the network, giving him the highest degree: 59,301. In the *friend* network 1,126 actors have a betweenness centrality going from 0 to 96,104 forming a long tail, with only 12 with a value higher than 10,000. These actors don't include the animator, showing that the *friend* network has been well adopted by users. The *family* network has 862 actors with a betweenness centrality from 0 to 162,881 with 5 values higher than 10,000. Only one actor is highly intermediary in both *friend* and *family* networks. The centralization of these three networks present significant differences showing that the semantics of relations have an impact on the structure of the social network. The betweenness centralities of all the relations, computed using the polymorphism in SPARQL queries with the "knows" property, highlight both the importance of the animator that has again the significantly highest centrality and the appropriation of users with 186 actors playing a role of intermediary. The employees of Ipernity.com have validated these interpretations of the metrics that we computed, showing the effectiveness of a social network analysis that exploits the semantic structure of relationships.

The Corese engine works in main memory and such an amount of data is memory consuming. The 494,510 relations declared between 61,937 actors use a space of 4.9 Go. The annotations of all messages use 14.7 Go and the representation of documents with their comments use 27.2 Go. On the other hand working in main memory allows us to process the network very rapidly. The path computation is also time and space consuming and some queries had to be limited to a maximum number of graph projections when too many paths could be retrieved. However, in that case, approximations are sufficient to obtain relevant metrics on a social network, i.e., for centralities (Brandes & Pich 2007). Moreover, we can limit the distance of the paths we are looking for by using other metrics. For example, we limit the depth of paths to be smaller or equal to the diameter of the components when computing shortest paths.

## **Toward an efficient navigation of the social capital**

The framework we presented enables analyzing the rich typed representations of semantic social networks and managing the diversity of interactions and relationships with parametrized SNA metrics. The exploitation of these semantic based SNA metrics permits structuring overwhelming flows of corporate social activities. The amount of metadata used to organize content will continue to increase as the success of social-tagging based system shows. Current methods are still limited at structuring this data and exploiting it for the analysis of social networks. As we have shown, combining semantic tools and methods with a collaborative approach is a promising track which needs to be further explored. Several challenges have to be tackled to provide efficient exploitation of the social capital (Lin 2008) (Krebs 2008) built through online collaboration, and to foster social interactions.

---

<sup>26</sup> This animator is an employee of the company that animates the social network, he declares as *favourite* every user who just created an account and sends him welcome messages.



First, computation is time consuming and even if Corese runs in main memory, experiments reported in the chapter show that handling a network with millions of actors is out of our reach today. Different approaches can be investigated to address that problem: (1) identifying computation techniques that are iterative, parallelizable, etc.; (2) identifying approximations that can be used and under which conditions they provide good quality results; (3) identifying graph characteristics (small worlds, diameters, etc.) that can help us cut the calculation space and time for the different operators.

Social web applications permit publishing, sharing and connecting so easily that a huge amount of social data is permanently produced, with a potential impact on the structure of the social network and the importance of its actors. Even if Corese enables loading data to the graph of a running engine, the computing cost and the volume of the data suggest only measuring relevant impacted metrics which change significantly. Consequently methods need to be developed to handle and quantify the impact of new social data on a semantic social graph.

Furthermore, community detection is one of the main focuses of social network analysis. Existing algorithms are based on heuristics to detect densely connected and cohesive groups of actors. But these algorithms are once again only based on the structure of the social network and they discard the semantic primitives used to type both relationships and actors. This lost knowledge could be used to determine semantically the cohesiveness of a community, to propose algorithms based on sociological definitions and to focus on relevant elements of the social graph for more efficient computation.

## **PERSPECTIVES: BUILDING “SHARED KNOWLEDGE GRAPHS”**

As discussed above, 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 the ontologies of the semantic web - they were thought of as knowledge representations capable of grasping the multi-dimensionality of the information we exchange (see Baget and 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 features and the ability for both machines and humans to exploit them at different levels of functionality. Folksonomies are a recent example of the “shared knowledge structures” which have emerged from web 2.0 applications as an affordable way to massively categorize resources.

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 many ways of describing content on the Web. The “Web of Linked Data<sup>27</sup>” 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 “representing social data with semantic web frameworks”, which describe content exchanged by and within on-line communities. These types of approaches are a good start as they already assist users in identifying, for instance, all the content 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 consists in “shared knowledge hubs”. The DBpedia project (Auer et al., 2007) is an example of such a hub, as it proposes

---

<sup>27</sup> [www.linkeddata.org](http://www.linkeddata.org)

expressing the knowledge structure of 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 elaborate “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 led 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 taxonomy categories among the tags, and to 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 of 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.

## **CONCLUSION**

Through the example of Business Intelligence Process we highlighted that the systematic exploitation of information to foster economic performance and facilitate decision making is one of the keys to success for all organizations worldwide. The progressive integration of successful web 2.0 applications into intranets to foster collaboration and knowledge sharing offers new perspectives for the competitiveness of innovative enterprises. Every user of the intranet becomes an actor of a collective watch by organizing, sharing, producing and enriching information as a side effect of using social applications. Semantic web frameworks provide models to connect and exchange the social data and the knowledge embedded in the social network, spread in a collaborative intranet. The semantic enrichment of social data such as folksonomies in intranets involves all the collaborators in an efficient elaboration of a shared and structured corporate vocabulary. The semantic SNA stack provides a way to fully exploit the RDF representations of online interactions and to enhance the social data with contextualized SNA features. These semantic intranets of people, combined with semantic descriptions of the knowledge they exchange, will allow for the construction of shared knowledge graphs. This will help to efficiently manipulate the overwhelming flow of data of a semantic intranet of people. An effective approach to building these shared knowledge graphs and to turning on-line social experiences into collective intelligence will permit efficiently capturing and managing the social capital embedded in the network structure of "knowledge workers" collaboration.

## **REFERENCES**

- Abbattista F., Gendarmi D. & Lanubile F. (2007). Fostering knowledge evolution through community-based participation. In Workshop on Social and Collaborative Construction of Structured Knowledge (CKC 2007) at WWW 2007, Banff, Canada.
- Adida, B.: hGRDDL: Bridging micorformats and RDFa. *Journal of Web Semantics*. 6(1): 61-69 (2008)
- Alkhateeb F., Baget J.-F. and Euzenat J., RDF with Regular Expressions INRIA RR-6191, <http://hal.inria.fr/inria-00144922/en> (2007)
- Angeletou S., Sabou M. & Motta E. (2008). Semantically enriching folksonomies with FLOR. In CISWeb Workshop at Europ. Semantic Web Conf.

Anyanwu, M., Maduko, A., Sheth, A.: SPARQL2L: Towards Support for Subgraphs Extraction Queries in RDF Databases, Proc. WWW2007, 2007.

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.

Bader, D. A., Madduri, K.: Parallel algorithms for evaluating centrality in real-world networks, ICPP2006 (2006).

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).

Begelman G., Keller P. & Smadja F. (2006). Automated tag clustering: Improving search and exploration in the tag space.

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.

Berners-Lee, T., Hendler, J., Lassila, O. (2001): The Semantic Web . Scientific American.

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.

Brandes, U. (2001): A faster algorithm for betweenness centrality. J. Math. Socio. 25(2): 163-177.

Brandes, U., Pich, C.: Centrality estimation in large networks. J. Bifurcation and Chaos in Applied Sciences and Engineering 17(7), 2303–2318 (2007).

Braun S., Schmidt A., Walter A., Nagypál G. & Zacharias V. (2007). Ontology maturing: a collaborative web 2.0 approach to ontology engineering. In CKC, volume 273 of CEUR Workshop Proceedings: CEUR-WS.org.

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

Buffà M., Erétéo G., Faron-Zucker C., Gandon F. & Sander P. (2008). SweetWiki: A Semantic Wiki. Journal of Web Semantics, special issue on Web 2.0 and the Semantic Web, 6(1).

Burt, RS: Structural Holes. Cambridge University Press, New York (1992)

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)

Burt, R.S.: Structural Holes and Good Ideas. American J. of Sociology 100(2): 339-399 (2004)

Cattuto C., Benz D., Hotho A. & Stumme G. (2008). Semantic grounding of tag relatedness in social bookmarking systems. 7th International Semantic Web Conference, Karlsruhe, Germany..

Coleman, J.S. (1988): Social Capital in the creation of the human capital. The American journal of sociology, Vol 94, Supplement: Organisations and institutions: Sociological and Economic Approaches to the Analysis of Social Structure.

Conein, B.: Communautés épistémiques et réseaux cognitifs: coopération et cognition distribuée. Revue D'Economie Politique 113, 141-159. (2004)

Corby, O., Faron-Zucker, F. (2002): Corese: A Corporate Semantic Web Engine. Workshop on Real World RDF and Semantic Web Applications.

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), Toulouse, France.

Corby, O., Kefi-Khelif, L., Cherfi Hacène, Gandon Fabien, Khelif Khaled (2009): Querying the Semantic Web of Data using SPARQL, RDF and XML. INRIA Research Report n°6847.

Donetti, L., Munoz, M. A.: Detecting communities: a new systematic and efficient algorithm. J. Stat. Mech. P10012 (2004)

Erétéo, G., Buffà, 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.

Erétéo, G., Buffà, M., Gandon., Corby, O. (2009): Analysis of a Real Online Social Network Using Semantic Web Frameworks, International Semantic Web Conference 2009.

EUZENAT J. & SHVAIKO P. (2007). *Ontology Matching*. Berlin, Heidelberg: Springer.

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

Fortunato, S., Latora, V., Marchiori, M.: Method to find community structures based on information centrality. Phy. Rev. E 70(5): 056104 (2004)

Freeman, L. C., Borgatti, S. P.: Centrality in valued graphs: A mesure of betweenness based on network flow. Social Networks 13: 141–154 (1991)

- Geisberg, R., Sanders P., Scultes, D.: Better approximation of betweenness centrality. ALENEX08 (2008).
- Girvan, M., Newman, M. E. J.: Community structure in social and biological networks, Proc. Natl. Acad. Sci. USA 99(12) (2002)
- Girvan, M., Newman, M. E. J.: Finding and evaluating community structure in networks. Phys. Rev. E, 69:026113 (2004)
- Golbeck, J., Parsia, B., Hendler, J.(2003): Trust network on the semantic web. CIA03.
- Goldbeck, J., Rothstein, M.(2008): Linking social networks on the web with FOAF. AAA08.
- Golder S. & Huberman B. A. (2005). The structure of collaborative tagging systems.
- Good B., Kawas E. & Wilkinson M. (2007). Bridging the gap between Social Tagging and Semantic Annotation: E.D. the Entity Descriptor. Available from Nature Precedings.
- 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
- Gruber, T. (2009): Ontology. Encyclopedia of Database Systems, Ling Liu and M. Tamer Özsu (Eds.), Springer-Verlag.
- Hendler, J., Goldbeck, J. (2008): Metcalfe's law, web 2.0 and the Semantic Web. J. Web Sem. 6(1):14-20.
- Henri, F., Pudelko, B.: Understanding and analyzing activity and learning in virtual communities. Journal of Computer Assisted Learning 19, 474-487. (2003)
- Heymann P. & Garcia-Molina H. (2006). Collaborative Creation of Communal Hierarchical Taxonomies in Social Tagging Systems. Rapport interne, Stanford InfoLab.
- Holme, P., Kim, B.J., Yoon, C.N., Han, S.K.: Attack vulnerability of complex networks. Phys. Rev. E 65: 056109 (2002).
- Hotho A., Jäschke R., Schmitz C. & Stumme G. (2006). Information Retrieval in Folksonomies: Search and Ranking.
- Huynh-Kim Bang B., Dané E. & Grandbastien M. (2008). Merging semantic and participative approaches for organising teachers' documents. In Proceedings of ED-Media 08 ED-MEDIA 08 - World Conference on Educational Multimedia, Hypermedia & Telecommunications, p. p. 4959–4966, Vienne France.
- Khare, R., Celik, T. (2006): Microformats: a pragmatic path to the Semantic Web. WWW2006
- Kim, H., Yang, S., Song, S., Breslin, J. G., Kim, H. (2007): Tag Mediated Society with SCOT Ontology. ISWC2007.
- Kochut, K.L., Janik, M. (2007): SPARQLer: Extended SPARQL for Semantic Association Discovery. Proc. ESWC2007,
- Krebs, V. (2008): Social Capital: the Key to Success for the 21<sup>st</sup> Century Organization. IHIRM Journal, Volume 12, N°5, p 38-42.
- Latora, V., Marchiori, M.: A measure of centrality based on the network efficiency. N. J. Phy 9 (6) : 188 (2007)
- Levenshtein V. I. (1966). Binary codes capable of correcting deletions, insertions and reversals. Soviet Physics Doklady., 10(8), 707–710.
- 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.
- Lin, N. (2008): A network theory of social capital. In D.Castiglione, J.W. van Deth, and G. Wolleb; editors, Handbook on Social Capital. Oxford University Press.
- McAfee, A.-P. (2006): Enterprise 2.0: The Dawn of Emergent Collaboration, MIT Sloan Management Review, Management of Technology and Innovation, April 1
- Markines B., Cattuto C., Menczer F., Benz D., Hotho A. & Stumme G. (2009). Evaluating similarity measures for emergent semantics of social tagging. In 18th International World Wide Web Conference, p. 641–641.
- 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
- Mathes A. (2004). Folksonomies - Cooperative Classification and Communication Through Shared Metadata. Rapport interne, GSLIS, Univ. Illinois Urbana-Champaign.
- Mika, P. (2005): Ontologies are us: A unified model of social networks and semantics. In Gil, Y., Motta, E., Milgram, S.: The Small World Problem. Psychology Today, 1(1): 61 – 67. (1967)
- 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

- Newman, M. E. J.: Scientific collaboration networks. Shortests paths weighted networks, and centrality. *Phys Rev E Stat Nonlin Soft Matter Phys* 64: 016132 (2001)
- Newman, M. E. J.: The structure and function of complex networks. *SIAM Review* 45, 167-256 (2003)
- Newman, M. E. J.: Detecting community structure in networks. *Eur. Phys. J. B* 38:321-330. (2004).
- Newman R., Ayers D. & Russell S. (2005). Tag Ontology Design. <http://www.holygoat.co.uk/owl/redwood/0.1/tags/>.
- Paolillo, J. C., Wright, E. (2006): Social Network Analysis on the Semantic Web: Techniques and Challenges for Visualizing FOAF. *Visualizing the semantic Web Xml-based Internet And Information*.
- PARK J. & HUNTING S. (2002). XML Topic Maps: Creating and Using Topic Maps for the Web. Addison-Wesley Professional.
- Passant A. (2007). Using Ontologies to Strengthen Folksonomies and Enrich Information Retrieval in Weblogs. In *International Conference on Weblogs and Social Media*.
- Passant A. (2009). Technologies du Web Sémantique pour l'Entreprise 2.0. PhD thesis, Université Paris IV - Sorbonne.
- Passant, A., Laublet, P. (2008): Meaning Of A Tag: A Collaborative Approach to Bridge the Gap Between Tagging and Linked Data. LDOW2008.
- Pérez, J., Arenas, M., Gutierrez, C. (2008): nSPARQL: A Navigational Language for SPARQL. In the proceedings of ISWC 2008.
- Pons, P., Latapy, M.: Computing communities in large networks using random walks. *ISCIS2005* (2005)
- 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)
- Ronzano F., Marchetti A. & Tesconi M. (2008). Tagpedia: a semantic reference to describe and search for web resources. In *WWW 2008 Workshop on Social Web and Knowledge Management*, Beijing, China.
- Rowe, M., Ciravegna, F. (2008): Getting to Me – Exporting Semantic Social Network Information from Facebook. In *proceedings of Social Data on the Web Workshop. ISWC 2008*. Karlsruhe, Germany.
- San Martin, M., Gutierrez, C.: Representing, Querying and Transforming Social Networks with RDF / SPARQL. *ESWC09*. (2009)
- Santos, E.E., Pan, L., Arendt, D., Pittkin, M.: An Effective Anytime Anywhere Parallel Approach for Centrality Measurements in Social Network Analysis. *IEEE2006* (2006)
- Schmitz P. (2006). Inducing ontology from flickr tags. In *Proc. of the Collaborative Web Tagging Workshop (WWW06)*.
- Schwarzkopf E., Heckmann D., Dengler D. & Kroner A. (2007). Mining the structure of tag spaces for user modeling. *Proceedings of the Workshop on Data Mining for User Modeling at the 11th International Conference on User Modeling*, p. 63–75.
- Scott J. (2000): *Social Network Analysis, a handbook*, second edition. Sage
- Specia L. & Motta E. (2007). Integrating folksonomies with the semantic web. *4th European Semantic Web Conference*.
- Tanasescu V. & Streibel O. (2007). ExtremeTagging: Emergent Semantics through the Tagging of Tags. In *ESOE at ISWC*.
- Tesconi M., Ronzano F., Marchetti A. & Minutoli S. (2008). Semantify del.icio.us: Automatically turn your tags into senses. In *Proceedings of the First Social Data on the Web Workshop (SDoW2008)*.
- Torniai C., Jovanovic J., Bateman S., Gašević D. & Hatala M. (2008). Leveraging folksonomies for ontology evolution in e-learning environments. In *ICSC '08: Proceedings of the 2008 IEEE International Conference on Semantic Computing*, p. 206–213, Washington, DC, USA: IEEE Computer Society.
- Van Damme C., Hepp M. & Siorpaes K. (2007). Folksonology: An integrated approach for turning folksonomies into ontologies. In *Bridging the Gap between Semantic Web and Web 2.0 (SemNet 2007)*, p. 57–70.
- Vanderwal T. (2004). Folksonomy Coinage and Definition. <http://www.vanderwal.net/folksonomy.html>.
- 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.
- Vidou, G., Dieng-Kuntz, R., El Ghali, A., Evangelou, C., Giboin, A., Tifous, A., Jacquemart, S. (2006): Towards an Ontology for Knowledge Management in Communities of Practice.
- Wasserman, S., Faust, K., Iacobucci, D., Granovetter, M.: *Social Network Analysis: Methods and Applications*. Cambridge University Press (1994)
- Waseem Akhtar, Jacek Kopecky, Thomas Krennwallner, and Axel Polleres. XSPARQL: Traveling between the XML and RDF worlds - and avoiding the XSLT pilgrimage. In *Proceedings of the 5th European Semantic Web Conference (ESWC2008)*, pages 432-447, Tenerife, Spain, June (2008). Springer.

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.

Weller K. & Peters I. (2008). Seeding, weeding, fertilizing. different tag gardening activities for folksonomy maintenance and enrichment. In S. Auer, S. Schaffert & T. Pellegrini, Eds., Proceedings of I-Semantics08, International Conference on Semantic Systems. Graz, Austria, September 3-5, p. 10–117.

Wilkinson, D. M., Huberman, B. A. A method for finding communities of related genes. Proc. Natl. Acad. Sci. (2003)

Wu, F., Huberman, B. A.: Finding communities in linear time: a physics approach. HP Labs (2004)

Zhou, H., Lipowsky, R.: Network brownian motion: A new method to measure vertexvertex proximity and to identify communities and subcommunities.

## **ADDITIONAL READING SECTION**

Berners-Lee, T., Hendler, J., Lassila, O. (2001): The Semantic Web . Scientific American.

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.

Erétéo, G., Buffa, M., Gandon., Corby, O. (2009): Analysis of a Real Online Social Network Using Semantic Web Frameworks, International Semantic Web Conference 2009.

Golbeck, J., Parsia, B., Hendler, J.(2003): Trust network on the semantic web. CIA03.

Goldbeck, J., Rothstein, M.(2008): Linking social Networks on the web with FOAF. AAA08.

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

Hendler, J., Shadbolt, N., Hall, W., Berners-Lee, T., Weitzner, D. (2008): Web science: an interdisciplinary approach to understanding the web. Communication of the ACM. Vol. 51, Issue 7, p60-69.

Krebs, V. (2008): Social Capital: the Key to Success for the 21<sup>st</sup> Century Organization. IHRM Journal, Volume 12, N°5, p 38-42.

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.

McAfee, A.-P. (2006): Enterprise 2.0: The Dawn of Emergent Collaboration, MIT Sloan Management Review, Management of Technology and Innovation, April 1

Markines B., Cattuto C., Menczer F., Benz D., Hotho A. & Stumme G. (2009). Evaluating similarity measures for emergent semantics of social tagging. In 18th International World Wide Web Conference, p. 641–641.

Mathes A. (2004). Folksonomies - Cooperative Classification and Communication Through Shared Metadata. Rapport interne, GSLIS, Univ. Illinois Urbana-Champaign.

Mika, P. (2007): Social Networks and the Semantic Web. Semantic web and beyond, Vol. 7. Springer.

Paolillo, J. C., Wright, E. (2006): Social Network Analysis on the Semantic Web: Techniques and Challenges for Visualizing FOAF. Visualizing the semantic Web Xml-based Internet And Information.,

San Martin, M., Gutierrez, C.: Representing, Querying and Transforming Social Networks with RDF / SPARQL. ESWC09. (2009)

Scott J. (2000): Social Network Analysis, a handbook, second edition. Sage

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.

Zhou, H., Lipowsky, R. (2004): Network brownian motion: A new method to measure vertexvertex proximity and to identify communities and subcommunities. ICCS 2004