

Enrichment of Affiliation Networks and Information Discovery in SKOS-based Datasets

Julia Anaya^{1,2}, Milan Stankovic² and Alexandre Passant¹

¹ Digital Enterprise Research Institute
National University of Ireland, Galway, Ireland
{alexandre.passant,julia.anaya}@deri.org,

² Hypios Research, 187 rue du Temple
Paris, France
milan.stankovic@hypios.com

Abstract. As a result of the Linking Open Data project, a growing amount of structured and semantically-rich data is available on the Web. Among other use-cases, it lets people link existing content (pictures, publications, etc.) to resources from well-known datasets, such as DBpedia, so that the linked information can be more easily accessed. In this paper, we focus on enhancing social networking and information discovery in such datasets, based on the hierarchical organisation of their topics. To do so, we extend the tripartite model of tagging by using SKOS relations between concepts available on the Web. Moreover, we demonstrate the validity of this approach by (1) identifying emerging social network relations and (2) enhancing people and documents identification in a network extracted from scientific publications tagged with MeSH terms.

Keywords: Semantic Web, Linked Open Data, SKOS, Social Network Analysis

1 Introduction

In a few years, the amount of information available on the Web as Linked Open Data (LOD)³ grown tremendously. Among others usages, end-users can benefit from this data by annotating existing content, by linking it to resources from well-know datasets, *e.g.* from DBpedia. Ontologies and frameworks such as MOAT [12], FLOR [1], as well as services like Faviki⁴ can be used for such semantically-tagging purposes. Similarly to what Web 2.0 practices offer with free-tagging, these systems enable users to classify their content with specific concepts, identified by URIs of resources from the LOD cloud. A main advantage of this approach is that annotation terms are structured and interlinked. For example, from any content annotated with `dbpedia:RDF`⁵, a user can browse

³ <http://linkeddata.org>

⁴ <http://faviki.com>

⁵ All the prefixes used in this paper can be resolved on <http://prefix.cc/>

content annotated `dbpedia:SPARQL`, since these resources share relationships in DBpedia.

However, there is still a need to identify potential benefits of this approach for end-users. In this paper, we discuss how our approach can be used to mine social networks from structured tagging data and to identify hidden connections, thus connecting people and bringing a Social aspect to Semantic Web technologies. In particular, we focus on SKOS-based annotations — Simple Knowledge Organization System [10] — concepts to identify such social networks, and see how different SKOS properties can be used to do so.

The remainder of this paper is organised as follows. We describe our methodology in Section 2 and illustrate our approach with a concrete example in Section 3, based on a social networking platform for researchers, the FPGG network — “Fondation Pierre-Gilles de Gennes”⁶. Subsequently, Section 4 reviews related work, and Section 5 reports both potentials and risks of this approach, which we noticed during our work. The paper is concluded by suggesting directions for future work that could further improve information discovery in this context.

2 Extending Author-Concept graph with SKOS

To identify the benefits of the Web of Data for tagging and social networking discovery, we consider the tripartite model of tagging (Tags, Users, Resources) and its extension proposed by [9]: Actors, Concepts, Instances. In our case study this model represents authors annotating scientific publications with terms from the MeSH — Medical Subject Headings⁷ — classification (Section 3). . We focus on the social and semantic dimension of this network by 1. extracting the bipartite Actors-Concepts model (built on implicit relations) from the tripartite one and 2. further enriching this network with data from the LOD cloud. Since the later is represented using ontological structures, various properties can be mined to create semantic relations between concepts. For instance, we can imagine a graph considering only `geonames:parentFeature` relationships to identify geographical hierarchy, or one using `foaf:knows` to identify people acquaintances.

However, these properties are domain specific and our motivation in this work is to provide a framework that is as generic as possible, *i.e.* domain-independent. To do so, we focus on SKOS, which “provides a standard way to represent knowledge organisation systems using the Resource Description Framework”⁸. Especially, we consider the `skos:broader` and `skos:narrower` properties that are used to hierarchically organise thesaurus and taxonomies. As a use-case, we consider MeSH data and its SKOS using proper mappings[17]. For example, consider the concept hierarchy in Fig.1, where *Phagocyte* is narrower of *Cells* and broader of *Osteoclasts*

⁶ <http://fpgg-network.org/>

⁷ <http://www.nlm.nih.gov/mesh/>

⁸ <http://www.w3.org/2004/02/skos/intro>

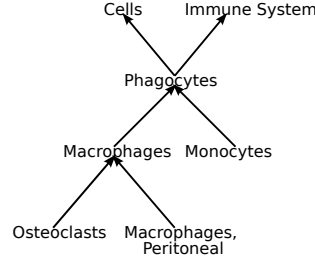


Fig. 1. Example concept hierarchy extracted from a MeSH subset

We will see how these relations between concepts can be used to generate new relations between authors that are using related concepts but that are not initially connected.

The aforementioned tripartite model of Actors-Concepts-Instances can be abstractly represented as a tripartite graph with hyper-edges that can be reduced (through a matrix transformation process) into three bipartite weighted graphs (two-mode networks), while each of these bipartite graphs could also be folded in two unipartite graphs.

Concretely, if the sets: $A = \{a_1, \dots, a_k\}$, $C = \{c_1, \dots, c_l\}$, $I = \{i_1, \dots, i_m\}$ represent actors, concepts and instances respectively, we can then define the two-mode (bipartite) AC as a triple of vertexes, edges and edge weights as in Eq. 1:

$$AC = \langle A \cup C, E_{ac}, w_{ac} \rangle, E_{ac} = \{(a, c) \mid \exists i \in I : (a, c, i) \in A \times C \times I\},$$

$$w_{ac} : E_{ac} \rightarrow \mathbb{N}, \forall e = (a, c) \in E_{ac}, w_{ac}(e) := |\{i : (a, c, i) \in A \times C \times I\}|. \quad (1)$$

The graph AC links the authors (Actors) to the keywords (Concepts) that they have used for annotating at least one publication (Instance). Each link is weighted by the number of times the author has used that keyword in a publication.

This graph is projected in two unipartite graphs (one-mode network): CC_a , where two concepts are linked if some actor used both concepts (defined in Eq.2) and AA_c where two actors are associated if they used a concept in common (defined in Eq.3).

$$AA_c = \langle A, E_{aa_c}, w_{aa_c} \rangle, E_{aa_c} = \{(a_1, a_2) \mid \exists c \in C : (a_1, c), (a_2, c) \in E_{ac}\},$$

$$w_{aa_c} : E_{aa_c} \rightarrow \mathbb{N}, \forall e = (a_1, a_2) \in E_{aa_c}, w_{aa_c}(e) := |\{c : (a_1, c), (a_2, c) \in E_{ac}\}|. \quad (2)$$

$$CC_a = \langle C, E_{cc_a}, w_{cc_a} \rangle, E_{cc_a} = \{(c_1, c_2) \mid \exists a \in A : (c_1, a), (c_2, a) \in E_{cc_a}\}, \\ w_{cc_a} : E_{cc_a} \rightarrow \mathbb{N}, \forall e = (c_1, c_2) \in E_{cc_a}, w_{cc_a}(e) := |\{a : (c_1, a), (c_2, a) \in E_{cc_a}\}|. \quad (3)$$

This association network can be semantically enriched by Social Network Analysis measures like *clustering coefficient* to distinguish general and specific terms, or extract broader and narrower relations. Especially, we focus on this extraction using relations defined in the SKOS ontology and explore two graph types, the concept graph and the authors graph, CC_a and AA_c .

2.1 Exploring the Concept Graph

Broader/narrower relations between concepts (extracted from SKOS) can also be represented as an unipartite unweighted directed graph where there is a link between two concepts if one is broader of the other.

We will denote this graph CC_b (b stands for *broader*) and define as $CC_b = \langle C, E_{cc_b} \rangle$. The vertexes set C is composed of the same concepts as in the AC graph and there is a directed edge between two concepts c_1 and c_2 if c_2 is broader of c_1 (or c_1 is narrower of c_2). The edge direction is from narrower concept to broader concept.

In this graph, we identify three significant patterns, illustrated in Fig.2:

- (a) a *parent-child* relation, where the parent is a more generic concept (or hypernym) than the child (or hyponym);
- (b) a *sibling* relation between two concepts, refers to the case where two concepts have a common broader concept;
- (c) a *co-ancestor* relation between two concepts, refers to the case where two concepts have a common narrower concept.

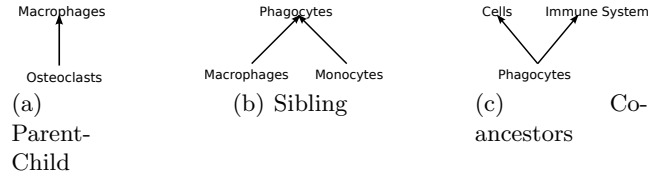


Fig. 2. Broader pattern relations in concept-concept graph

Parent-child relations are isomorphic to asymmetric dyads. A dyad is a subgraph consisting of two nodes and the possible ties among them and triads are subgraph consisting of three nodes. Sibling relations are also isomorphic to the triads denoted as *021D* ([18], page 566) and finally, co-ancestors relations are isomorphic to the triads as *021U*.

Note that the *top* concept in the case of sibling and the *down* one in the case of co-ancestors can be concepts extracted from LOD cloud that not necessary belong to the initial C set, but are responsible of the new relations between concepts that are in this set. For instance, the *top* concept *Phagocytes* in Fig. 2(b) is responsible of the sibling relation between *Macrophages* and *Monocytes* and these three concepts belongs to C set, but *Reticuloendothelial System*, extracted from LOD, is not in the C set and it is also a broader (*top*) concept of *Macrophages* and *Monocytes*.

We can then define a new set $K, K \supseteq C$ with concepts that were not in the initial C set and redefine the CC_b graph as KK_b . Now, in KK_b the edges can link concepts in C or K/C .

2.2 Exploring the Author Graph

The CC_b graph (and KK_b) aims at identifying new associations between authors, that were not defined in the original AA_c graph, and can also be used for browsing topics from the relationships they share (possibly not in the CC_a).

In the same way we identify three patterns in the CC_b graph, we can identify three relation patterns between authors linked to those concepts, illustrated in Fig.3.

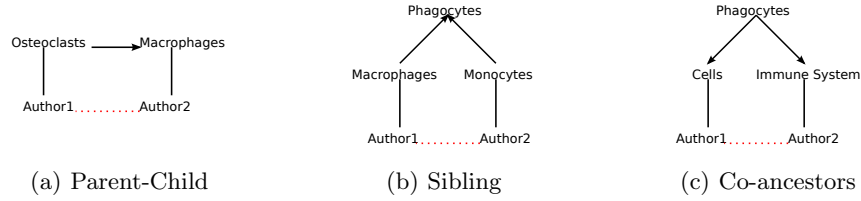


Fig. 3. author-author relations based on broader concept relations

Using sibling relations, we can infer relations between authors working in the same scientific domain, *Proteins* although part of different sub-domains, *i.e.* *Peptides* and *Carrier proteins*.

Note the concepts in the sibling and co-ancestors cases can be at the same time parent-child cases whenever the top (in sibling) or down concept (in co-ancestors) belongs to C . Note also that the down concepts in the sibling relation cannot be a parent-child relation between them at the same time, as it would mean that we are considering transitive relations. For instance, *Phagocytes* is broader of *Macrophages* and *Macrophages* is broader of *Osteoclasts*, and by transitive closure, *Phagocytes* is broader of *Osteoclasts*. If then we create a sibling relation with *Osteoclasts* and *Macrophages* as down concepts and *Phagocytes* as *top* concept, *Osteoclasts* and *Macrophages* (down concepts) are a parent-child relation too.

We will explain this in more detail in the 3 section.

We found several algorithms to compute the new relations between the authors given the initial AC graph and the KK_b graph. To determine the better one in terms of memory and time complexity is out of the scope of this paper.

One of these algorithms is to generate a new *extended* AK_b by the composition operation between AC and CC_b graphs, ie. $AK_b = AC \circ KK_b$. This graph is no more bipartite, hence we cannot project it to get the authors and concepts unipartite graphs. But we can generate the new author relations by transversing the path of our broader patterns illustrated in Fig.3 and calculate edge weights as the path distance in these traverses.

Other method would be to reduce the triads and dyads in the concept broader patterns as new unique nodes and to link the authors, that were initially linked to the concepts in the patterns, to these new nodes.

For the sake of simplicity we generate a new bipartite graph AC_{bp} (p stands for *pattern*) with the original AC and the graphs created from linking concepts by all possible broader patterns (in Fig. 2), denoted CC_{bp} (for *broader patterns*).

This graph can now be folded in a similar fashion as AC into a graph AA_{cbp} , that includes all the relations between authors by broader patterns and also by concepts they both used. To determine only the new relations between authors generated by the broader patterns, we can generate the graph AA_{bp} as the difference between AA_{cbp} and AA_c (the authors graph without broader relations), i.e.: $AA_{bp} = AA_{cbp} \setminus AA_c$

The graph transformation process is illustrated in Fig.4

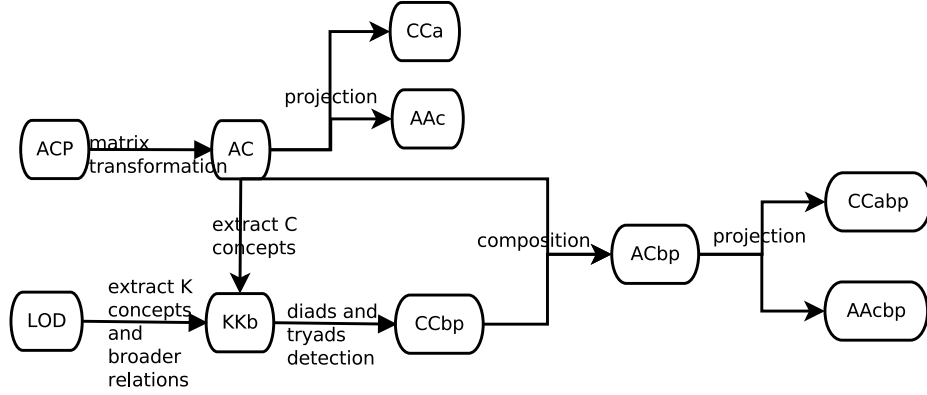


Fig. 4. Actor-Concept graph transformation

2.3 Space of possibilities with the merged graph

Many social media experts point out that the Social Web world we are living in is narrowing our view to what is shared and known by our immediate contacts.

Everything gets filtered by who we know, and information gets prioritised. Although helpful in combating information overload, this same feature is standing in the way of information discovery and meeting the unknown people, getting to the unknown knowledge and possibilities. We believe that the addition of the semantic dimension to the Social Networks might provide the missing link for the information discovery of different kinds.

In this section, we discuss how the information discovery could be achieved thanks to the semantic dimension and we explore the possibilities of gaining additional knowledge and insight based on the triangulation of the social and semantic dimensions of the graph.

Table 1 displays (1) the semantic dimension of concepts relationships on the vertical axis, (2) and the social dimension of people’s connections on the horizontal axis. Based on these two dimensions, we define a set of possible intersections that would allow us to discover interesting additional knowledge.

For instance, the lowest right corner showcases the category of people who do not work together but work in field that have a common ancestor. Such connections might indicate potential matches for some joint research. In addition, looking for persons that do not work exactly in the same domain, but rather in a domain that is closely related (by common ancestor) might introduce additional diversity in the project. Similarly, finding colleagues from sibling domains could also be interesting potential helpers for internal project, as shown on the `coworkerSiblingDomain` intersection.

Note also that these properties could be materialised in the graph as relationships between instances of `foaf:Person` (for instance by extending the `RELATIONSHIP` vocabulary⁹), so that they can be directly used when querying the graph.

	Work together	Do not work together
Broader	<code>coworkerBroaderDomain</code>	<code>shareBroaderDomain</code>
Narrower	<code>coworkerNarrowerDomain</code>	<code>shareNarrowerDomain</code>
Sibling	<code>coworkerSiblingDomain</code>	<code>shareSiblingDomain</code>
Co-ancestors	<code>coworkerCoancestorsDomain</code>	<code>shareCoancestorsDomain</code>

Table 1. Integrating relationships between concept and social connections for mining new relationships

3 A concrete case: the FPGG network

We demonstrate our research approach on the FPGG Social Network¹⁰ site for researchers. This website represents a social platform for more than 1500 researchers of the “Fondation Pierre-Gilles de Gennes”, specialised mostly in

⁹ <http://vocab.org/relationship/>
¹⁰ <http://fpgg-network.org/>

Physics, Biology and Medicine. The platform aims to provide internal social features, but also to enable external partners to identify the right experts on the platform, and build cooperations and joint ventures. In this section we show how our approach can be applied to provide meaningful visualisations of concept and user network structures for this. We are currently implementing it as a plug-in for the foundation website, and plan to test further its effectiveness in practice, including user evaluation and feedback.

Our dataset comes from users' publications, accessible via PubMed¹¹ and automatically uploaded to the site. Related meta-data (publication title, authors and keywords) is extracted from PubMed XML, and some data (authors and publication titles, but not keywords) is also available through the Neurocommons SPARQL endpoint¹².

Author names follow two different patterns: "*Lastname Firstname Initials*" and "*Lastname Firstnameinitial Initials*", for example "*Amigorena Sebastian S*" and "*Amigorena S S*". To avoid duplicate names, we assume that authors with similar last name and initials, and a first name starting with the same letter, are the same ("*Lastname Firstnameinitial(.) * Initials*"). Going further, other techniques for identity detection from publications could be considered as [16].

The concepts in our dataset are Medical Subject Headings, defined as "a comprehensive controlled vocabulary for the purpose of indexing journal articles and books in the life sciences; it can also serve as a thesaurus that facilitates searching."

A particularity of our dataset is that each author is not tagging individually a publication, as in the case of tagging systems where users tag resources, but all the authors agree on the keywords used the tagging process.

SKOS relations between concepts are extracted from the Neurocommons SPARQL endpoint. The Neurocommons project provides different graph bundles, one of these is the mesh-SKOS bundle¹³ that contains MeSH-SKOS mapping [17]. Broader/narrower concept relations could also be extracted from other LOD sources such as DBpedia. Indeed, in the case of MeSH, MeSH-DBpedia mappings exist¹⁴.

In Listing 1.1, we provide an example of SPARQL query to get broader concepts, and Listing 1.2 displays related results. Note that some results are inferred by transitivity, which normally apply with `skos:broaderTransitive` and not by using `skos:broader`. However, the endpoint we used did not make the distinction. We yet tweaked the results to limit relationships to one level, and to avoid dealing with relations between very generic and specific terms.

```
SELECT ?concept ?parent ?parentlabel
WHERE {
    ?concept skos:prefLabel "Osteoclasts" ;
    skos:broader ?parent .
```

¹¹ <http://www.ncbi.nlm.nih.gov/pubmed/>

¹² <http://sparql.neurocommons.org/>

¹³ <http://neurocommons.org/page/Bundles/mesh/mesh-skos>

¹⁴ http://esw.w3.org/HCLSIG_BioRDF_Subgroup/DERI_HCLS_KB


```

    ?parent a skos:Concept ;
      skos:prefLabel ?parentlabel
  }

```

Listing 1.1. Example SPARQL query to identify broader concepts

```

Mononuclear Phagocyte System
Phagocytes
Cells
Macrophages
Connective Tissue Cells
Immune System
Myeloid Cells
Hemic and Immune Systems

```

Listing 1.2. Example broader results

We identified these transitive relations as isomorphic to the *030T* triads [18]. If we denote T_b the subset containing these triads, we generate $KK_b \setminus T_b$ and compute all the subsequently graphs without considering transitive triads.

Using our dataset, the following Fig. 5(a) illustrate the *key* concepts in the CC_b graph (without transitive) with all C concepts in dataset. The graph is filtered by in-degree connections, represented by the node size, and colours represent the residuals from the linear regression between *eigenvector* centrality versus *betweenness* centrality. As these two metrics should be approximately linear, this could note gatekeepers to central actors (high betweenness but low eigenvector) or unique access to central actors (low betweenness and high eigenvector).

To create the initial graphs (without broader relations), we follow a similar approach to the one described by Mika [9]. We generate the AC graph with the data in our dataset and filter it on the absolute strength of associations by applying a threshold (calculated from the weights standard deviation) and generate the AA_c (and CC_a) projection. To normalise the weights, we calculate the Jaccard coefficient on the AC adjacency matrix and filter again. As a reminder, the AC graph reflects actors sharing concepts as interests (by publications), *i.e.* the associations reflect overlapping communities of interests, in the AA_c graph authors are linked if they use the same concepts and in the CC_a graph concepts are linked if authors used both concepts.

In addition:

- Table 2(a) summarises the total number of nodes in the dataset, publications (I), authors (A) and MeSH concepts (C);
- Table 2(b) summarises some averages as the average number of concepts by author in the AC graph, denoted $\bar{d}(a_i)$, where \bar{d} is the average degree and $a_i \in A$;
- Table 2(c) summarises the number of edges in the graphs before extending the graph with broader relations.

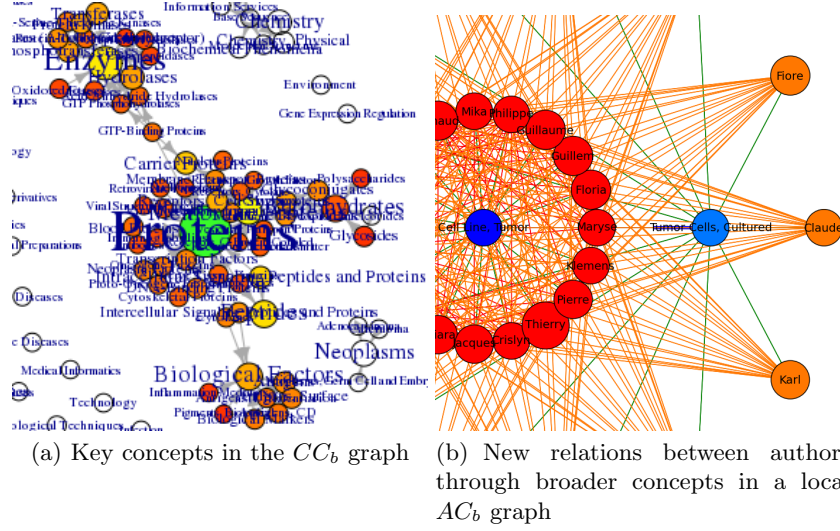


Fig. 5.

We observe that the number of concepts by author or authors by concepts is a power of 10, leading to unipartite graphs with more than a million of edges.

Table 2(d) shows the number of edges in the concept graphs generated from broader relations. CC_{bu} denotes the graph created only with sibling concepts triads from CC_b (Fig. 2(b)) and CC_{bd} the with co-ancestors triads (Fig. 2(c)).

In a similar way, in Table 2(e), AA_{bu} represents the authors graph generated only with sibling triads (Fig. 3(b)) and AA_{bd} represents the graph with co-ancestor triads (Fig. 3(c)). Here, due to computational limitations, we did not calculate the total number of edges, but the number of new edges that were not in the AA_c graph, *i.e.* $|E_G \setminus E_{aa_c}|$ included to illustrate how large are these graphs.

Fig. 5(b) displays a graph with new relations between authors and concepts, using `skos:broader` relations. In this graph, all the authors (nodes in the periphery) linked to *Tumor Cells, Cultured* are now connected to all the authors linked to *Cell Line, Tumor*.

In order to scale down the dataset (and be able to plot it), we also performed the evaluation process on a subset composed of the publications in the year 2009 from the FPGG network.

In Table 3(a) are summarised the number of nodes in this new dataset, 3(b) the degree averages and in 3(c) the number of edges in the graph without broaders, in total, after the first filtering ($|E'_G|$) and the second filtering ($|E''_G|$). Table 3(d) shows the number of edges in the graph with broaders, AA_{cbp} is the authors graph created from all possible broader relations plus the initial associations between authors that used the same concepts and CC_{abp} is created also from all possible broader relations plus the links that were in CC_a graph.

(a) Nodes

V	$ V $
I	5,230
A	13,650
C	6,412
$K \setminus C$	1673

(b) Averages

$\overline{d}(a_i)$ in AI	5.55
$\overline{d}(c_i)$ in CI	11.08
$\overline{d}(c_i)$ in AC	19.06
$\overline{d}(a_i)$ in AC	32.18

(c) Edges without
broaders

G	$ E_G $
AC	260,205
CC_a	1,309,380
AA_c	23,000,000

(d) Edges in concept graphs from
(broaders)

G	$ E_G $
CC_b	13,895
$CC_b \setminus T_b$	11,406
CC_{bu}	56,841
CC_{bd}	5,635

(e) Edges in author
graphs (from broad-
ers)

G	$ E_G \setminus E_{aa_c} $
AA_{bu}	1,809,693
AA_{bd}	16,833,594

Table 2. Number of nodes, edges and averages in the graphs from FPGG dataset

Both of them are compared with the graphs without broaders. before filtering and after filtering.

(a) Number of nodes			(b) Averages		(c) Edges (without broad- ers)			
V	$ V $	$ V' $	$\overline{d}(a_i)$ in AI	5.94	G	$ E_G $	$ E'_G $	$ E''_G $
I	481		$\overline{d}(c_i)$ in CI	9.13	AC	28,169	679	
A	2,228	87	$\overline{d}(c_i)$ in AC	12.64	AA_c	9,377	2,771	743
C	1,754	80	$\overline{d}(a_i)$ in AC	16.06	CC_a	21,980	819	230

(d) Edges with broaders		
G	$ E'_G $	$ E''_G $
CC_{abp}	1,894	525
$CC_{abp} \setminus CC_a$	1075	97
AA_{abp}	3,555	914
$AA_{cbp} \setminus AA_c$	784	21

Table 3. Number of nodes, edges and averages in the graphs from FPGG 2009 publications subset

The Fig.6 illustrate the graph AA_{cbp} after filtering. The new edges that were not in the AA_c graph are emphasized. We have also detected communities applying the spin glass model [14] and generated the layout using the Fruchterman-Reingold [5] algorithm.

Cluster Analysis; Gene Expression Profiling; Comparative Genomic Hybridization; Oligonucleotide Array Sequence Analysis; Polymorphism, Single Nucleotide; Genetic Predisposition to Disease; Risk Assessment; Proportional Hazards; Melanoma; DNA, Neoplasm; Uveal Neoplasms; Liver Neoplasms; Monosomy; Chromosomes, Human, Pair 3; Humans;
Retrospective Studies; Follow-Up Studies; Models; Humans; Female;
Algorithms; Spectrum Analysis; Signal Processing, Computer-Assisted; Image Interpretation, Computer-Assisted; Magnetic Resonance Imaging; Image Enhancement; Elasticity; Elasticity Imaging Techniques; Artifacts; Phantoms, Imaging; Computer-Aided Design; Equipment Design; Equipment Failure Analysis; Models, Theoretical; Models, Statistical; Models, Neurological; Models, Biological; Feasibility Studies; Reproducibility of Results; Sensitivity and Specificity; Sonication; Ultrasonography; Ultrasonic Therapy; Ultrasonography, Mammary; Ultrasonography, Doppler, Color; Ultrasonography, Doppler, Transcranial; Viscosity; Thrombolytic Therapy; Cerebral Hemorrhage; Stroke; Fibrinolytic Agents; Tissue Plasminogen Activator; Rheology; Muscle, Skeletal; Liver; Female; Humans;

Table 4. Concepts in communities in the AA_{cbp} graph

ones by referring to an ontology (Wordnet or Wikipedia) [15] or focus on the actor-actor relation through interlinked data (FOAF) [6]. Our work focus on both concept-concept and actor-actor relations through semantifying concepts by mapping generic relations between them to the SKOS ontology

A different approach by [4] uses semantic web frameworks to include the semantics of the graph-based representations and create SemSNA, an ontology of SNA characteristics used to annotate social networks.

5 Conclusions and perspectives

In this paper, we explored how Linked Data and specifically SKOS relationships can be used to enhancing social networking and information discovery in existing datasets, and more concretely how to connect people.

An evident result from this work is that using broader relations between concepts increases the number of associations between concepts or authors. A more accurate comparison between the initial graphs and the graphs with broader relations would be needed as future work. We also plan to validate our preliminary results through user evaluations. MeSH-SKOS mappings (and in general, well defined ontologies) can be improved by including a level in the hierarchy. (the distance or number of steps to the root concept). This will allow to easily detect broader/narrower relations that could be inferred by transitivity.

This work could be further improved by facilitating a better and faster algorithm to compute the authors and concepts graph generated from broader patterns, as in the current approach the memory and complexity time grows exponentially.

In addition, extracting broader and narrower relations using set theory could be an approach to validate data from LOD that might not be accurate enough.

Triadic study could also solve incongruity in the broader/narrower case, as the cycles detection. Clustering algorithms could also help to determine synonym sets of the more specific terms. The specificity or generality of terms could be also compared with hypernyms and hyponyms relations in Wordnet

To analyse the communities represented in the graphs, other SNA algorithms could be applied, as clustering or similarity metrics to check the sub-communities with the SKOS domains.

Other interesting study could be to obtain the relations between concept from more general contexts as DBpedia, either using existing mappings with MeSH or matching terms and checking the semantic. For example, searching *Phagocytes* as <http://dbpedia.org/page/Phagocytes>. We also could explore other kind of relations between the concepts.

Acknowledgements This work has been funded by Science Foundation Ireland under Grant No. SFI/08/CE/I1380 (Líon 2). We wish to thank Benjamin Carrillo, Václav Belák, Marcel Karnstedt and Werner Breitfuss for their valuable comments and suggestions.

References

1. Sofia Angeletou. Semantic Enrichment of Folksonomy Tagspaces. In *Proceedings of the 7th International Semantic Web Conference (ISWC2008)*, volume 5318 of *Lecture Notes in Computer Science*, pages 889–894. Springer, 2008.
2. Paul Buitelaar and Thomas Eigner. Topic Extraction from Scientific Literature for Competency Management. In *Proceedings of the PICKME workshop on Personal Identification and Collaborations: Knowledge Mediation and Extraction at the 7th International Semantic Web Conference*, volume 403. CEUR-ws.org, 2008.
3. Ali Daud, Juanzi Li, Zhou Lizhu, Lei Zhang, and Ying Ding. Modeling Ontology of Folksonomy with Latent Semantics. *IEEE/WIC/ACM International Conference on Web Intelligence (WI 2010)*, 2010.
4. Guillaume Erétéo, Michel Buffa, Fabien Gandon, and Olivier Corby. Analysis of a real online social network using semantic web frameworks. In Abraham Bernstein, David R. Karger, Tom Heath, Lee Feigenbaum, Diana Maynard, Enrico Motta, and Krishnaprasad Thirunarayan, editors, *International Semantic Web Conference*, volume 5823 of *Lecture Notes in Computer Science*, pages 180–195. Springer, 2009.
5. Thomas M. J. Fruchterman and Edward M. Reingold. Graph Drawing by Force-directed Placement. *Software: Practice and Experience*, 21(11):1129–1164, 1991.
6. Masahiro Hamasaki, Yutaka Matsuo, Takuichi Nishimura, and Hideaki Takeda. Ontology extraction using social network. In *International Workshop on Semantic Web for Collaborative Knowledge Acquisition (SWeCKa 2007)*, 2007.
7. Andreas Hotho, Robert Jäschke, Christoph Schmitz, and Gerd Stumme. Information Retrieval in Folksonomies: Search and Ranking. In *The Semantic Web: Research and Applications*, volume 4011 of *Lecture Notes in Computer Science*, pages 411–426, Heidelberg, June 2006. Springer.
8. Alexander Kreiser, Andreas Nauerz, and Fedor Bakalov. A Web 3.0 Approach for Improving Tagging Systems. In *Proceedings of the Workshop on Web 3.0: Merging*

- Semantic Web and Social Web 2009*, volume 467 of *CEUR Workshop Proceedings*, Turin, Italy, June 2009. CEUR.
9. Peter Mika. Ontologies Are Us: A unified model of social networks and semantics. In *International Semantic Web Conference*, volume 3729 of *Lecture Notes in Computer Science*, pages 522–536. Springer, November 2005.
 10. Alistair Miles and Sean Bechhofer. SKOS Simple Knowledge Organization System Reference. Technical Report W3C Recommendation, W3C, August 2009. available at <http://www.w3.org/TR/2008/WD-skos-reference-20080829/>.
 11. Alexandros Nanopoulos, Hans-Henning Gabriel, and Myra Spiliopoulou. Spectral Clustering in Social-Tagging Systems. *WISE '09: Proceedings of the 10th International Conference on Web Information Systems Engineering*, 5802:87–100, 2009.
 12. Alexandre Passant, Philippe Laublet, John G. Breslin, and Stefan Decker. A URI is Worth a Thousand Tags: From Tagging to Linked Data with MOAT. *International Journal on Semantic Web and Information Systems (IJSWIS)*, 5(3):71–94, 2009.
 13. Cartic Ramakrishnan, Krys Kochut, and Amit P. Sheth. A Framework for Schema-Driven Relationship Discovery from Unstructured Text. In *International Semantic Web Conference*, volume 4273 of *Lecture Notes in Computer Science*, pages 583–596. Springer, 2006.
 14. Joerg Reichardt and Stefan Bornholdt. Statistical Mechanics of Community Detection. *Physical Review E*, 74, 2006.
 15. Rajesh Thiagarajan, Geetha Manjunath, and Markus Stumptner. Finding Experts By Semantic Matching of User Profiles. In *Proceedings of the PICKME workshop on Personal Identification and Collaborations: Knowledge Mediation and Extraction at the 7th International Semantic Web Conference*, volume 403. CEUR-ws.org, 2008.
 16. Vette I. Torvik and Neil R. Smalheiser. Author Name Disambiguation in MEDLINE. *ACM Transactions on Knowledge Discovery from Data (TKDD)*, 3(3), July 2009.
 17. Mark van Assem, Véronique Malaisé, Alistair Miles, and Guus Schreiber. A Method to Convert Thesauri to SKOS. In *The Semantic Web: Research and Applications*, volume 4011 of *Lecture Notes in Computer Science*, pages 95–109, Heidelberg, June 2006. Springer.
 18. Stanley Wasserman and Katherine Faust. *Social Network Analysis: Methods and Applications*. Cambridge University Press, New York, USA, 1994.