Social Objects Description and Recommendation in Multidimensional Social Networks: OCSO Ontology and Semantic Spreading Activation

Nicolas Marie, Fabien Gandon

Social objects description and recommendation in multidimensional social networks

OCSO Ontology and semantic spreading activation

Nicolas Marie
Social Communication / Edelweiss
Alcatel-Lucent Bell Labs / INRIA Villarceaux / Sophia-Antipolis, FR
nicolas.marie@alcatel-lucent.com

Fabien Gandon
Edelweiss
INRIA
Sophia-Antipolis, France
fabien.gandon@inria.fr

Abstract— **Two strong tendencies in social networks today are their growing multidimensionality and the key role of social objects around which they crystallize. We define OCSO as semantic web schema to describe finely these social objects and their afferent social activity. Then we defend the use of spreading activation algorithms to perform complex recommendation in multidimensional social networks. A possible approach to what we call semantic spreading activation is introduced in order to exploit, among others, the OCSO model.**

Index-Terms : social object, object centered sociality, semantic web, social semantic web, ontology, OCSO, mutidimensional social network, spreading activation, semantic spreading activation

I. INTRODUCTION

Today social networks are multidimensional networks [Contractor, 2009]. They are composed of users and others entities such as multimedia content, cultural products, places, events and so on which are linked through explicit relations such as *friend*, *like*, *recommend*, *studied at*, *work at*, *inspired by*, *is here* (eg. on facebook.com). This high level of expressivity is capturing several dimensions like thematic, social, temporal or spatial relations.

The heterogeneous content shared on these platforms generates social links between users and constitute social objects [Knorr-Cetina, 1997]. A social object is "*the reason why people affiliate with each specific other and not with anybody*" [J.Engëstrom, 2005]. It is anything that provokes, supports, maintains an interaction and creates social links: events, multimedia content, people, interests, activities, things, etc. It constitutes at the same time a social lodestone [Kinsella, 2007], a context provider and a communication anchor on social platforms. Object centered sociality [Knorr-Cetina, 1997] constitutes today a specific and widespread kind of numeric communication coexisting with others like micro-blogging, mail, forums, etc.

These multidimensionality and object centered sociality trends can be clearly observed on some generic platforms such as Facebook, plus.google.com/, vkontakte.ru, orkut.com, renren.com or linkedin.com and be forecast on social semantic networks based on ontologies such as FOAF [1] and SIOC [2] due to the intrinsically typed structures of the web of data. These phenomenons are also catalyzed by tierce buttons such as Facebook, Google, LinkedIn ones which drastically increase the amount and variety of shared content.

First this paper presents and explains the Object Centered Sociality Ontology (OCSO), which is a semantic web model for representing social objects and the corresponding social activity. Second, it focuses on advanced recommendation in multidimensional social networks. Social networks are traditionally modeled by graphs which are called sociograms. Multidimensional networks can be modeled as strongly typed graphs where nodes and links are both typed. Spreading activation is an interesting candidate to provide smart backend in this context. We share our reflection about a semantics-based version to benefit from the richness of the typed graph structure and increase recommendation quality.

II. SOCIAL OBJECTS REPRESENTATION: OCSO ONTOLOGY

Almost all of contents on the web are socially augmented through comments, recommendations or ratings consequently constituting social objects. We propose the notion of Object Centered Social Network (OCSN, figure 1) to define the union of a social object and corresponding users and interactions. An object-centered social network is a social structure formed by people and interactions centered on a single common social object. Services proposing socially augmented content like sharing platforms are supporting as many OCSN as content it contains. Thus they can be considered as OCSN platforms.

---

[1] http://xmlns.com/foaf/spec/
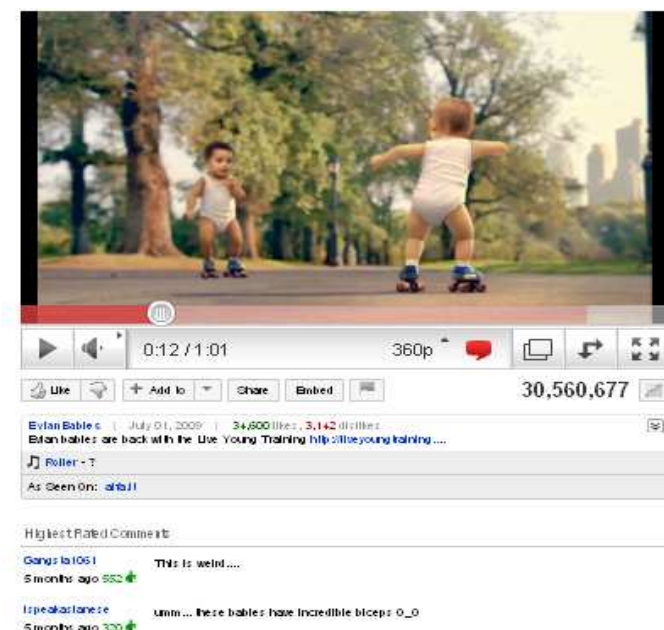
[2] http://rdfs.org/sioc/spec/

Figure 1. A typical OCSN is composed of a social object (eg. a video) plus corresponding interactions and users

[Breslin and Decker, 2007] stressed the social web need for semantics. Semantic web relies on standard knowledge representation like RDF/S [Brickley and Guha, 2000] or OWL [MacGuinness and Van Harmelen, 2004]. In terms of schemas, main social semantic web contributions include FOAF for user profiles, SIOC for online communities, OPO[3] for online presence, NiceTag[4] for tagging and SemSNA[5] for social network analysis. However, these models do not cover in details the social activity related to contents on social platform.

A. State of the art

The Open Graph Protocol[6] (OGP) proposed by Facebook uses RDFa[7] syntax to give a semantic description about any web page on the web according to OGP schema. RDFa is a syntax which allows embedding RDF in HTML pages. OGP schema is easy to appropriate and covers a set of highly popular social objects such as films, musicians or videos. These metadata are especially useful for Facebook when combined with buttons "*I like", "I send"* and "*I recommend"* which are spread all over the web by webmasters. They allow to structure information in the platform by organizing them into categories for instance. The OGP is typing web content before turning it into social objects in the platform but it does not offer any possibility to describe associated social activity all the more so as it is precisely what Facebook wants to host.

ABmeta[8] ontology is quite similar to OGP protocol. It is also RDFa based and proposes a schema which describes a finite number of content. Authors describe it themselves as "*object-centric: Focuses on everyday things that we encounter around the web*". The model itself does not capture social activity but it is mainly used on review website which are obviously social.

hreview[9] is a microformat designed for review websites which are proliferating without interoperability. Among others, the schema integrates rating, review text, tag, reviewer id. Reviewing is one of the most common social object oriented activity on the web. hreview is an interesting initiative but only covering a small part of our needs because limited to reviewing activity.

Schema.org[10] was proposed by Bing, Google and Yahoo in order to enrich their search results. The schema is an OWL ontology which is very complete, covering a wide range of common web entities from multimedia content to politic organization. The corresponding social activity is omnipresent and described in UserBlocks, UserCheckins, UserComments, UserDownloads, UserLikes, UserPageVisits, UserPlays, UserPlusOnes, UserTweets classes. Schema.org is a proof of search engine companies' interest to take into account more social information in their search results.

B. OCSO ontology

Initiatives presented in the state of the art demonstrate a growing use of semantics in the context of social web. These models are integrating more or less explicitly the concept of social object. Unfortunately they are often incomplete regarding our needs because they do not capture finely the social activity. Furthermore these models are not aligned with each other's and not coherent with social semantic models like SIOC and FOAF. OCSO (figure 2) aims to overtake these limits and precisely capture social activity around content reusing as much as possible existing models and extending them when needed. The namespace for OCSO is http://ns.inria.fr/ocso/.

The `SocialObject` class corresponds to the socially augmented content or representation, eg. a video or a place. It contains the main identification information of the social object such as its title, URL and type. The type property is aligned with the OGP and schema.org model and requires an additional computation to move from the string value of these legacy properties to a full `rdf:type` declaration. OCSO also allows to specify user who uploaded (`ocso:uploadedBy`) or created (`ocso:createdBy`) the social object.
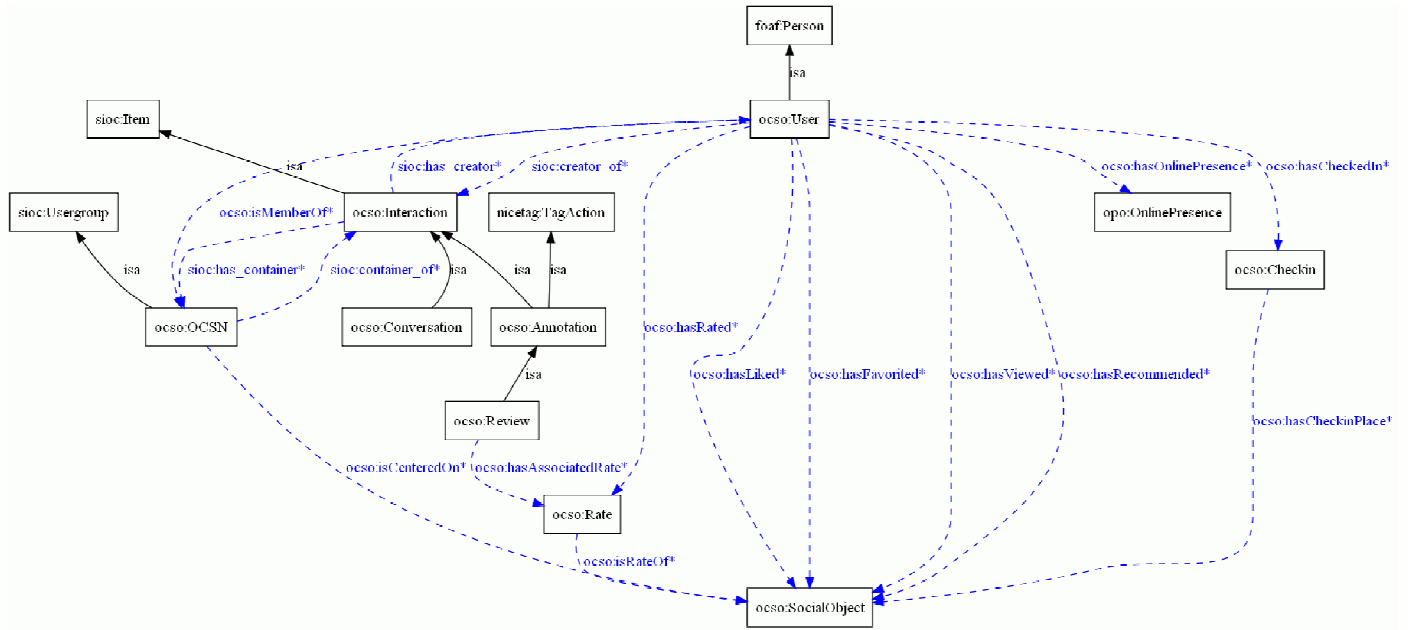
Figure 2. OCSO ontology classes

OCSN is a subclass of `sioc:Usergroup`. It represents the social activity focused on the social object and is the subject of properties that characterize the view, interaction, like, share, recommend counts. Thus it is possible to compute aggregated social activity metrics related to the social object over these properties. The OCSN class is related to the `SocialObject` class through `ocso:centeredOn` property

The `User` class is a subclass of `foaf:Person`. It contains several properties qualifying actions a user made on social object, such as *liked* or *recommend*, and which have social impacts. The class supports an online presence property offered by the Online Presence Ontology.

`Interaction` class is a subclass of `sioc:Item`. It is linked to corresponding OCSN by `sioc:has_container` and inversely by `sioc:container_of`. It has two subclasses `Annotation` and `Conversation`. An `Annotation` is a contribution which augments the meaning of a social object. It can be free text like comments or slightly more structured like tags. A `Conversation` is an exchange between two or more users in a subthread generated via "*reply to comment*" functionality for instance.

`Selection` class is capturing user selection on platforms such as favorite or subscription for contacts, content (social object) or interactions.

III. ADVANCED RECOMMANDATION IN MULTIDIMENSIONAL NETWORKS

As they are growing and becoming more complex the efficiency of human navigation in such multidimensional social networks depends on the availability of a powerful and intelligent backend providing guidance and recommendations. The heterogeneity and the amount of linked entities require more intuitive and efficient ways of navigating and accessing information overtaking classic methods we encountered today such as timelines, search and browsing.

Spreading activation, explained below, is an interesting candidate to provide smart backend in the context of multidimensional social networks because graph semantics can be integrated in the propagation process to influence the spread and increase recommendation relevance.

A. Spreading activation

Spreading activation originally comes from cognitive psychology and more specifically from works on semantic memory. It is a cognitive model which aims to describe the human remembering process. [Quillan, 1968] proposed to represent human memory and knowledge through semantic networks where concepts are represented in a network of interconnecting nodes. These concepts are linked through labeled relations. [Collins and Quillian, 1969] asserted that, in the memory, information are stored in categories which are each other logically and hierarchically related. In their paper, authors also made an experiment and validate the memory representation through semantic network structure.

Representing human memory into semantic networks helped to model some human cognitive processes. [Collin and Loftus, 1975] detail spreading activation theory which model remembering process. They explain that it is performed first through activation of concepts in the mind. Then these activations are propagated into the semantic network (*i.e.* the memory) to neighbor's nodes. During this propagation step activation power decreases inversely proportionally to

semantic relatedness which is the strength of links connecting concepts, nodes. Sufficiently activated destination nodes become the start point of another propagation round and so forth, pulse after pulse, till propagation stops.

Later these observations lead to a mathematical model [Anderson, 1983] and inspired algorithms for various purposes in mechanics, physics, chemistry, computer sciences and many others disciplines. There are plenty of spreading activation algorithm variants and huge tuning possibilities but the core is always the same. The algorithm is initialized by putting an activation value on one or several node(s) of a graph. Then this value is propagated to neighbor nodes with an increased or decreased value depending on tuning. Neighbor nodes with a sufficient value are then fired and so on till a stop condition is fulfilled (e.g. number of pulse, number of node processed). Recently some works proved that spreading activation is fast, scalable and tolerant to data incompleteness even on huge datasets [LarKC, 2008].

B. State of the art

There are some previous works about application of spreading activation in the field of multidimensional social networks. In [Kinsella and al., 2007] authors performed spreading activation through IBM galaxy tool[11] on an RDF graph composed of information about four semantic web people: Tim Berners-Lee, Dan Brickley, Andreas Harth and Tim Finin. They performed egocentric and polycentric queries to discover links between heterogeneous nodes in the graph. [Troussov and al., 2008] presents theoretically the mining method for egocentric and polycentric queries in multidimensional social network. In [Troussov and al., 2009], authors use spreading activation to perform a user similarity operation taking into account tagging information. These works does not focus very much on the semantic sensitive aspect of the algorithm. By semantic spreading activation we mean an algorithm which uses graph semantics to increase relevance of spreading activation process. Most interesting works about semantic sensitive spreading activation are not related to multidimensional social networks and are presented below.

[Crestani and Lee, 1999] is one of the earliest works about spreading activation over the web. Their spreading activation algorithm is not semantic sensitive because it is applied on hypertextual links which are unlabeled. However authors are calculating a similarity between fired nodes and their neighbors. The similarity is based on a cosine comparison between web pages search profiles. These search profiles are weighted index terms extracted from pages content. A user experiment shows that the algorithm retrieved relevant results.

[Rocha and al., 2004] presents a hybrid approach to search semantic web combining a classical search method composed with spreading activation. Authors point the fact that ontologies only give information about presence or absence of a link between two concepts. They propose to use semantics of

the ontology to calculate strength between nodes in the graph. They call this process the weight mapping. From this weight mapping operation they obtain a hybrid instances network, where each relation instance has both a semantic label and a numerical weight. This process can be assisted by a knowledge engineer indicating important parts of the ontology. Authors propose three methods to calculate the weight between two nodes: a cluster measure, an inverse domain frequency measure (called specificity measure) and a combined one. A comparative test highlights the very positive impact of weight mapping on search results. Authors stress the fact that the activation is not semantic sensitive: "*there is no semantic interpretation of the activation value flowing through the network*".

[Freitas and al., 2011] propose to simplify linked data query mechanisms for end-users through their query mechanism Treo. Authors explain that in the case of linked data end-user still have to compose with the syntax of a structured query language: SPARQL. Querying linked data with natural language queries should considerably increase their visibility and value. To reach their goal they use a combination of entity search, semantic relatedness measure and spreading activation. First, they extract concepts from the natural language query through named entity recognition. The query is then transformed into a partial order dependency structure (PODS) which is a reduced representation of the query. For instance the PODS of "*from which university did the wife of Barrack Obama graduate?*" is Barrack Obama – wife – graduate – university. The next step consists in searching a pivot entity corresponding to the first concept. In their example Barrack Obama is matched with: http://dbpedia.org/resource/Barack_Obama. Then a semantic relatedness is calculated with neighbor nodes. Authors use the Wikipedia-based semantic relatedness for this step. The activation is spread to the node with the highest similarity. This process is repeated until the algorithm retrieves the answer. Their experiment showed an average precision of 0,487, an average recall of 0,57 and 70% of answered queries.

C. Analysis and next steps

[Rocha and al., 2004] shows very encouraging results about taking into account semantics to retrieve information through spreading activation. [Kinsella, 2007] and [Troussov, 2008] present spreading activation as a very interesting candidate to retrieve information in multidimensional social networks. These previous works encouraged us to explore the trail of semantic sensitive spreading activation for multidimensional social networks. To achieve this goal, the spreading activation formula has to be adapted. A classic formula can be found in [Anderson, 1983] and [Akim and al.,2011]:

$$a_i' = \lambda u_i + \mu f \left(\sum w_{ij} * a_j \right)$$

Where a_i' is the activation at next iteration of node i, u_i is the initial activation, w_{ij} is the weight of the relation between

---

[11] http://www.alphaworks.ibm.com/tech/galaxy

nodes i and j. λ and μ are constants used for algorithm tuning. f depends on algorithm purpose, typically it can be a combination of thresholding functions, linear combinations or non-linear functions for instance.

Classic spreading activation is not semantic sensitive:
- Activation a is undifferentiated; we propose a semantic spread which requires to type activation. Possible types are provided by one or several schema(s) and associated to a dimension eg. SIOC, FOAF, OCSO for social dimension or geo[12] for spatial one.
- w_{ij} is often calculated using graph structure. For example it is common to use the called-out degree y_j (number of outgoing edges)of node a_j as $w_{ij} = 1/y_j$

Thus we propose a spreading activation formula which is taking into account typed spread:

$$a(i, n+1, t) = \frac{w_s * s(i, n+1, t) + \sum_{j,e_{i,j}}[w_p(l(e_{i,j}), t) * a(j, n, t)]}{w_s + \sum_{j,e_{i,j}} w_p(e_{i,j})}$$

and:

$$a(i, n) = \frac{\sum_t [w_a(l(i), t) * a(j, n, t)]}{\sum_t w_a(i, t)}$$

Where:
- $s(i, n, t)$ is stimulation of node i for type t at moment n.
- $a(i, n, t)$ is activation of node i for type t at moment n.
- $a(i, n)$ is aggregated activation of node i at moment n.
- w_s is the weight of the local stimulation in a node activation.
- w_p is the weight of propagation from neighbors for a specific spread type and through a specific arc type.
- w_a is the aggregation weight of typed activations for a specific activation type and a specific node type
- $l()$ gives the type of a node or an arc ; note that in this first formulation we made the assumption that only one of the most specific types was considered for each node.
- $e_{i,j}$ is the set of arcs between i and j and oriented from i to j.

Using a semantic propagation will help to increase relevance result by adding context. Application of a classic spreading activation in the context of multidimensional social network can rapidly lead to out of context results. In fact, highly heterogeneous nature of corresponding graph can easily drive propagation to irrelevant part of the graph. Moreover, undifferentiated activation in the classic formula does not allow composing activations on several dimensions eg. spatial plus social or social plus thematic (eg. music).

IV. CONCLUSION AND FUTURE WORK

We presented OCSO ontology which is capturing social objects centered interactions. These interactions constitute an important part of communication on the web nowadays. This ontology is completing previous efforts done in the field of social semantic web and connecting them to content description schemas. Information captured by OCSO ontology can be interesting for taking into account the social dimension in multidimensional social networks (during spreading activation for instance) because it clearly establish relationships social dimensions (users) and contents supporting their interactions. Integration modalities of OCSO information will be part of our future works.

Then we exposed spreading activation recent applications in the context of multidimensional social networks. We also presented a state of the art about semantic spreading activation algorithms. This kind of algorithm shows good results. Today, trends in social networks and more generally proliferation of typed graph structures through semantic web paradigm adoption renew the interest for spreading activation. Expressivity level of this data opens new perspectives. Need of advanced techniques to make sense from these massive interlinked data is increasing. Moreover, very encouraging results about spreading activation speed, scalability and tolerance to incompleteness on massive datasets are confirming that this algorithm family is highly interesting in the context of multidimensional social networks. Thus we proposed a model for the semantic spreading activation and our future work includes the implementation, experimentation and evaluation of this proposal on a dataset including at least 2 dimensions eg. social and spatial.

[1] Akim, Nazihah Md and Dix, Alan and Katifori, Akrivi and Lepouras, Giorgos and Shabir, Nadeem and Vassilakis, Costas (2011) *Spreading Activation for Web Scale Reasoning: Promise and Problems.* pp. 1-4. In: Proceedings of the ACM WebSci'11, June 14-17 2011, Koblenz, Germany.

[2] Anderson, J. R. (1983). A spreading activation theory of memory. Journal of Verbal Learning and Verbal Behavior, 22, 261-295

[3] Breslin, J. G., Decker, S. (2007) THE FUTURE OF SOCIAL NETWORKS ON THE INTERNET: THE NEED FOR SEMANTICS. *IN IEEE Internet Computing Magazine, 11 (6), 86-90*

[4] Brickley D. and Guha R.V. (2000) RDF vocabulary description language 1.0: *RDF SCHEMA. W3C proposed recommandation (http://www.w3.org/TR/rdf-schema/).*

[5] Collins, A., Loftus, E., (1975) A Spreading Activation Theory of Semantic Processing. Psychological Review, vol. 82, pp. 407-428,

[6] Collins, *A. M., &* Quillian*, M. R. (*1969)*. Retrieval Time from Semantic Memory. Journal of Verbal Learning and Verbal Behavior, 8, 240-247*

[7] N. S. Contractor, The Emergence of Multidimensional Networks. *Journal of Computer-Mediated Communication 14, 743-747 (2009).*

[8] Crestani, F. and Lee, P. L. 2000. Searching the web by constrained spreading activation. Inf. Proc. Manage. 36, 585–605.

[9] Engestrom J. (2005) *Why* some social network services work and others don't or: the case for object-centered sociality, *blog posting*

[10] Freitas, A., Oliveira, J.G., O'Riain, S., Curry, E., da Silva, P.: Querying Linked Data using Semantic Relatedness: A Vocabulary Independent Approach. In: Proceedings of the 16th International Conference on Applications of Natural Language to Information Systems, NLDB (2011)

[11] Kinsella, S., Harth, A. and Breslin, J. G. (2007). "Network Analysis of Semantic Connections in Heterogeneous Social Spaces". Proceedings of the UK Social Network Conference, London, United Kingdom.

[12] [Kinsella, 2007b] Kinsella, S., Harth, A., Troussov, A., Sogrin, M., Judge, J., Hayes, C., Breslin, J., Navigating and Annotating Semantically-Enabled Networks of People and Associated Objects, In Proc. In Proceedings of Applications of Social Network Analysis ASNA 2007, September 2007

[13] Knorr-Cetina K. (1997), Sociality with Objects: Social Relations in Postsocial Knowledge Societies, *Theory, Culture & Society, vol. 14,no. 4, 1997, pp. 1–30.*

[14] LarKC (2008) Spreading activation components (v1) A Kiryakov, D Ognyanoff, I Peikov, R Velkov… - LarKC project, deliverable D

[15] McGuinness, D.L., Van Harmelen, F.(2004) : OWL Web Ontology Language Overview. W3C Recommendation, 10 FEBRUARY 2004.

[16] Quillian, M. (1968). Semantic memory. In M. Minsky (Ed.), Semantic Information Processing, pp. 227–270. MIT Press, Cambridge, MA.

[17] Rocha, C., Schwabe, D., and de Aragão, M. P.: A Hybrid Approach for Searching in the Semantic Web. In Proc. of the 13th International World Wide Web Conference (WWW 2004), NY (2004) 374-383

[18] [Rocha u. a. 2004b] Rocha, Cristiano ; Schwabe, Daniel ; Arag̃ao, Marcus P.: Integrating Semantic Concept Similarity in Model-Based Web Applications. In: Joint Conference 10th Brazilian Symposium on Multimedia and the Web & 2nd Latin American Web Congress, (WebMedia & LA-Web 2004), 12-15 October 2004, Ribeirao Preto-SP,Brazil, 2004, S. 78–85

[19] Troussov, A., Sogrin, M., Judge, J., and Botvich, D.: Mining sociosemantic networks using spreading activation technique; In Proc. of IMEDIA' 08 and I-KNOW' 08. JUCS, 2008

[20] Troussov (2009) : Spreading activation approach to tag-aware recommenders: modeling similarity on multidimensional networks