A Social Networking Model of a Web Community

Anna V. Zhdanova[1, 2], Livia Predoiu[3], Tassilo Pellegrini[4], Dieter Fensel[2]

[1] CCSR (Centre for Communication Systems Research), University of Surrey, GU2 7XH, Guildford, UK Tel: +44-1483-689492

[2] DERI (Digital Enterprise Research Institute), University of Innsbruck, Technikerstrasse 21a, A-6020, Innsbruck, Austria Tel: +43-512-507-6485

[3] Institute of Computer Science, University of Mannheim, A5, D-68159, Mannheim, Germany Tel: +49-621-181-2484

[4] Semantic Web School – Centre for Knowledge Transfer, Lerchenfeldergürtel 43, A-1160, Vienna, Austria Tel: +43-1-4021235-28

a.zhdanova@surrey.ac.uk, livia.predoiu@informatik.uni-mannheim.de, t.pellegrini@semantic-web.at, dieter.fensel@deri.org

Abstract. We propose a model to represent closeness between entities within Web communities and social networks/folksonomies conceptually and numerically, and a community dynamics notification algorithm. The proposed models add value to community environments supporting social networking, and are specifically applicable to the community-driven environments, where users create and share their vocabularies. We also show how this model can be applied within real-world use cases.

1. Introduction

Social communities on the Web are steadily emerging and the demand on virtual social community networks is immense. Community members profit from being linked to other people sharing common interests though having widely dispersed residences. Without virtual social community portals on the Web, people would not be able to find that many persons sharing the same interests and being available for discussions and collaborations.

By means of adding semantics to social community portals with ontologies yielding semantic social community portals, the content processing possibilities are enhanced. For example, searching for specific information yields more qualified results. Topic-wise processing of information, e.g. ontology-supported treatment of privacy issues and knowledge management makes discussing and sharing information easier. Most current semantic social community portals allow their users to create their own vocabularies by attaching keywords (the so-called tags) to all kinds of information they put online. Examples of community portals which allow their users to tag resources are Flickr[1] allowing their users to tag photos or the del.icio.us[2] allowing their users to tag bookmarks. Such tag collections yield community specific vocabularies and are called folksonomies.

By means of tagging, social communities can elaborate their own vocabularies. So-called "Tags" are often not considered to be ontologies, but Mika [3] indicates a way on how to evolve light-weight ontologies by means of a model of social communities which consist just of users, tags and tagged objects. The People's portal [7] is a novel community portal infrastructure which goes a step further and allows the users to create ontologies in RDF(S)[3]/OWL[4], populate and manage them. Allowing the users to evolve ontologies and not just keywords brings social community networks a big step forward in opening further interaction and content processing options.

Although current virtual social portals provide many interaction possibilities for their users, they are still limited in the way they process the community-created content. There is still a lack in dynamics tracking within a portal and in updating users with information about changes relevant for them. Users can search for information according to what the community portal developers allow. Trust and privacy issues, as well as simple identification of people with common interests remain to be challenges. Thus, it is clear that a richer model for social networking is needed, in particular the one allowing to model closeness between entities in a community.

In this paper, we propose a model which allows to represent and calculate the closeness between persons basing on the content they put online within a community portal. Our model can represent different kinds of Web communities and social networks/folksonomies conceptually and numerically. We also present a community dynamics notification algorithm which also takes advantage of the computation of the closeness value between entities within a community. The proposed model and the algorithm add much value to the users of community environments supporting social networking.

The paper is organized as follows. In Section 2, we present conceptual and numerical modelling, and an algorithm for notification of community dynamics. In Section 3, we show how the theoretical models are applied to real-world case studies, report on the first results and identify directions for further work.

[1] Flickr: http://www.flickr.com
[2] del.icio.us: http://del.icio.us
[3] Resource Description Framework: http://www.w3.org/RDF/
[4] Web Ontology Language: http://www.w3.org/2004/OWL/

2. Theoretical Model

2.1 Conceptual Modeling

A social networking/folksonomy model presented here is built on top of state of the art models for semantic social network representation [3, 5]. In this section, we provide conceptual foundations of the proposed model.

In order to model networks of folksonomies at an abstract level, a core model is represented as a tripartite graph with hyperedges. The set of vertices is partitioned into the three (possibly empty) disjoint sets $A = \{a_1, \ldots, a_k\}$, $C = \{c_1, \ldots, c_l\}$, $I = \{i_1, \ldots, i_m\}$ corresponding the set of actors (users), the set of concepts (tags, keywords, higher-order ontology constructions) and the set of objects annotated (bookmarks, photos etc.) In effect, the common bipartite model of ontologies (concepts and instances) is extended by incorporating actors in the model. In the model employed here, we follow conventions set in RDF and name representations of the set *A* members as *subjects*, and representations of the set *I* members as *objects*. The later representations can be ontology instances or literal values. *Subjects* and *objects* are also referred with a common name as *nodes*.

In a social tagging system, users tag objects with concepts, creating ternary associations between the user, the concept and the object. Thus the folksonomy is defined by a set of annotations $T \subseteq A \times C \times I$ [3]. Such a network is most naturally represented as hypergraph with ternary edges, where each edge represents the fact that a given actor associated a certain instance with a certain concept. In particular, we define the representing hypergraph of a folksonomy *T* as a (simple) tripartite hypergraph $H(T) = \langle V,E \rangle$ where $V = A \cup C \cup I$, $E = \{\{a, c, i\} \mid (a, c, i) \in T\}$. We also refer to edges connecting actors/subjects and concepts/objects as *links*. Factually in community-driven ontology construction environments as the People's portal [7], links are most often represented as arbitrary properties that connect subjects with objects.

Tripartite graphs with hyper-edges can be reduced to three bipartite graphs (also called two-mode graphs) with regular edges. These three graphs model the associations between actors and concepts (graph AC), concepts and objects (graph CO) and actors and instances (graph AI). For example, the AC valued bipartite graph is defined as follows:

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

$$w : E \to \aleph \text{ with } w(e) = |\{i \mid (a,c,i) \in E\}| \text{ for all } e = (a,c) \in E\ /$$

Therefore, the bipartite graph AC links the persons to the concepts that they have used for tagging at least one object. Each link is weighted by the number of times the person has used that concept as a tag. This kind of graph is known in the social network analysis literature as an affiliation network [5], linking people to affiliations with weights corresponding to the strength of the affiliation. An affiliation network can be used to generate two simple, weighted graphs (one-mode networks) showing the similarities between actors and events, respectively. Here, ontology construction and community analysis are mainly supported at the level of the AC graph presented above, namely involving subjects (actors), links (edges) and objects (concepts).

2.2 Numerical Modeling

In this section, we propose a numerical model to specify communities and relations within these communities on the basis of a more general conceptual model described in the previous section.

<u>Connection Strength</u>

Rewording the formalization of the previous section, a community is modelled as follows. Subjects (i.e., persons or actors) can be connected to each other only via links with the same objects (i.e., concepts). This modelling also complies with a definition of a community as a group having common interests. In the model, these interests are represented by objects.

Strictly speaking, direct links between two subjects do not exist. A subject can only be connected to another subject in the following way via an object and two or more links: "Subject1 – Link1 – Object – Link2 – Subject2". A link between a subject and an object are bi-directional. Each direction of a link has a value assigned to it. The value assignment represents the fact that a connection of one subject to an object may be stronger than a connection of another subject to the same object.

Formally, the value of the link is calculated as follows. *link_value*(*link_a*) is defined for any model where *link_a* exists between an object and a subject. The value of the function is in the range $(0,1]$.

Practically, one can determine strength/value of each link by examining subjects and objects associated with this link. Basing on the theoretical principles on language, communication and communities [2], we put forwards the following two factors as crucial in influencing the connection strength/value of a link between subjects:

- <u>Popularity of objects</u>: Growing popularity of objects (or how many subjects are linked to these objects) weakens the connection strength between subjects linked via these objects. For example, being connected with someone having an object "Community portals" as a common research topic is stronger than being linked with someone having a common concept "Female" as a "Gender" attribute.
- <u>Capacity of subjects</u>: The more objects are linked to/embraced by a subject, the weaker connections of this subject are to other subjects via these objects. In other words, the more activities a subject is involved in the less attention/time/effort is distributed to the object from the subject's side. For example, if a researcher claims to work in 10 projects, this most often means that the time invested in each of these

projects is less than it would be in case when a researcher works in just one project. Here, being involved in many projects with different partners results in weakening the connection strength between partners.

Strictly speaking, modelling connection strength between two subjects can be made more complex with taking in account additional factors and when trying to establish a very precise balance between the two main factors mentioned above. For example, in a system where a person is allowed to marry only one person, being connected to someone via an object "Marriage" is stronger than having the same connection in a community where a person may marry several persons. However, popularity of objects and capacity of subjects are in any case seen as inverse proportional to the connection strength or value of the link. Therefore and because we aim at a lightweight model that is not overloaded with a lot of factors which have a minor impact on the connection strength and increase the computation effort substantially, we chose popularity of objects and capacity of subjects as factors for our model.

Remaining generally correct and adding value from the practical point of view, it can be stated the strength of the link between a subject and an object is inversely proportional to the subject's capacity and the object's popularity. Remember that the connection strength function *link_value* is not symmetrical, i.e., subjects can be indirectly attached to one another with different strength: one subject may be linked closer to another subject, than the later to the first subject.

The value of the function *link_value* between *subject_1* and *subject_2* from the point of view of *subject_2* is calculated as follows.

$$link\_value(link\_1) = \{ \ \frac{1}{popularity(object\_1)} \frac{1}{capacity(subject\_1)}, \\ \forall link\_1 = (object\_1,\ subject\_1) \in E, \exists\, link = (object\_1,\ subject\_2) \in E \ \}$$

Here, *subject capacity* and *object popularity* are metrics signifying on the number of links connected to the node. These metrics are formally specified below. Specifically, the measures *subject_capacity* and *object_popularity* are specified via the measure *links_connected*(*node*), which returns the number of links connected with a node.

Subject Capacity

Informally, *subject capacity* reflects the number of things a person/agent is involved with, the number of activities a person/agent participates in, etc. Subject capacity is identified by the number of objects the subject is connected to.

Formally, *capacity*(*subject*) is defined for any model where *subject* exists. The value of the function is an integer in the range $[0, \infty)$ and is calculated as follows.

$$capacity(subject) = links\_connected(subject),$$

where $links\_connected(subject) = \left|\{e \mid e = (subject, object) \in E, object \in C\}\right|$

Object Popularity

Informally, *object popularity* reflects the number of persons/agents which are associated via any kind of link with the object. As it was already defined above, objects factually are represented by instances, both string values and resources connected with a subject via a property.

Formally, *popularity*(*object*) is defined for any model where *object* exists. The value of the function is an integer in the range $[0, \infty)$ and is calculated as follows.

$$popularity(object) = links\_connected(object),$$

where $links\_connected(object) = \left|\{e \mid e = (subject, object) \in E, subject \in A\}\right|$

Closeness Measure

As mentioned above, two subjects can only be connected via an intermediate object or objects and other subjects and links, but not directly to each other via a link. Therefore, in order to calculate connection strength between subjects, we look via which objects these subjects are connected and how popular or important these objects are.

The communities are dynamic and are permanently subject to changes and practically any change in closeness between two subjects is caused by a person profile change on a community portal. When a person assigns new objects to him/her, the closeness measure values towards other people connected with the person change, links to new people may appear and already existing links may disappear.

Formally, *closeness*(*node_1*, *node_2*) is defined for any model where *node_1*, *node_2* and paths between *node_1* and *node_2* exists. The value of the function is in the range $(0, \infty)$ and is calculated as follows.

$$closeness(node\_1, node\_2) = \sum_{paths(node\_1, node\_2)} \prod_{link\_1 \in paths(node\_1, node\_2)} link\_value(link\_1)$$

Here a path between *node_1* and *node_2* is defined as a chain of one or more links starting at *node_1* when following them one by one, *node_2* is reached. And *link_1* is said to belong to a path when it forms a part of the path between *node_1* and *node_2*. Function *paths*(*node_1*, *node_2*) returns all the paths leading from *node_1* to *node_2* in the given model, or $E_{node\_1 node\_2}$ in the graph notation.

As the reader may already notice, the closeness function can be used to calculate closeness between two objects, similar to the way the closeness is calculated between two subjects. Pragmatically, the function reflects how close one subject's view on the world (Weltanschauung) to the view of another subject, i.e., how many common objects they share and how strongly they are committed to these objects. When a subject can be reached via a path consisting of several links, a product of the respective link values is taken. Such modeling correlates with the fact that one direct link is stronger than several transitive links, e.g., being a friend is a stronger relation than being a friend of a friend. When a person/agent can be reached from another person/agent via several paths, the products for every path are summed up in order to receive the value reflecting all the relations connecting two persons/agents. The function is asymmetric in the same way as the function *link_value* is asymmetric, i.e., one node can be connected stronger to another than the later to the first one.

2.3 Community Dynamics Notifications

Keeping a community member up-to-date regarding the community dynamics (i.e., changes which take place in the community) is crucial for keeping the community representation correlated with reality and evolving.

Important events for the community members to be informed/notified of include the following.

Notification of a member and a community upon user profile change

A community member is to be notified upon changes in the profiles of community members who are connected to him/her via shared objects.

A notification process for community members on the profile change is as follows.

1) a community member changes his/her profile
2) community members who are to be notified of the change are identified
3) closeness degree between community members is re-calculated
4) the member who changed the profile and his/her communities are notified about the change, current closeness degrees and changes in closeness degrees (including members indirect links to whom appeared or disappeared as a result of changed relations to certain objects in profiles of other members)

Notification of a community/community member upon appearance of a new object

A community member is to be notified upon appearance of new objects in the community space, as these objects may appear relevant to a person and a community member(s) may consider assigning them to his/her profile. Selection process of a (sub-)community, which is notified upon appearance of a new object, may employ analysis of already existing links between subjects and objects and use of *closeness* value between different nodes. For instance, a community member may be notified about a new object if the closeness value between him/her and the person who introduced this object in the ontology is not smaller than a certain threshold value.

Notification of a community/community member upon popularity change of objects

A community member is to be notified upon popularity change for the objects in the community space (i.e., ontology), as these objects may appear relevant to a person and a community member(s) may consider assigning or removing links to them in his/her profile. Selection process of a (sub-)community, which is notified upon an object popularity change in an ontology, may employ analysis of already existing links between subjects and objects and use of *closeness* value between different nodes. For instance, a community member may be notified about a change in object popularity if the closeness value between him/her and this object in the ontology is not smaller than a certain threshold value.

3. Practical Experience, Results and Further Work

For reasons of acceptance a qualitative empirical approach has been chosen to test the model concept in two community settings. Following the approach of dialogue-based participatory technology assessment [1] the model concept has been discussed with the community management, two persons with a high degree of interaction and two persons with little degree of interaction following a standardized questionnaire. Based on these findings a second empirical phase is planed for autumn 2006. The community settings have been chosen by the criteria size, functional mode (community of practice or interest), administrative capacity, access and degree of integration [6].

Setting A is composed of members of the Platform Wissensmanagement[5], a Vienna-based online community dealing with applied and theoretical aspects of knowledge management active since 1998. This setting is characterised by interaction addressing organizational aspects associated with working groups, questions concerning physical meetings and events, tools and methodologies and knowledge sharing on project ideas and search for proper project partners. Setting A can be described as a loosely integrated *community of interest*,

[5] Platform Wissensmanagement: http://www.pwm.at

relatively large in size (about 1000 individual members), of heterogeneous interests, with flat hierarchical structures, of low administrative capacity and little restrictions to access.

Setting B is composed of members of the Austrian Best Bio-Gas Practice[6], a network of institutional players in Austrian bio-energy domain active since 2004. The community's purpose is to develop policies, guidelines and metrics to foster the sustainable energy market in Austria by raising awareness, networking stakeholders, lobbying and standards setting. Setting B can be described as a densely integrated *community of practice*, being of relatively little size (about 20 organisational members), higher administrative capacity, steeper hierarchical structures, more restricted access and working by a consensual mode.

Observations on the Model Applicability to Case Studies

Findings in Setting A:

Interviewees from setting A raised concerns about the managebility of the model. From their perspective, less integrated communities are characterised by a higher degree of terminological richness and ambiguity. This affects the popularity criterion of objects and undermines the expressiveness of the folksonomy. Thus questions have been raised how to produce terminological significance and how terminology management could be supported technically.

Findings in Setting B:

Interviewees from setting B have been generally more positive towards the community model. Terminological variance has been described as lower compared to less integrated communities leading to a higher expressiveness of the folksonomy. Nevertheless a phenomenon has been addressed which could be called "terminological capture" by politically motivated coalitions, overstressing certain topics and underrepresenting others. Hence questions of governance have to be seen as a crucial aspect.

General Findings:

Interviewees from both settings are facing the challenge of terminological representativity and the necessity of an editorial process that combines top down domain ontology modelling with bottom up social tagging. The representativeness of the domain ontology is a crucial criterion for the acceptance of the model in both community settings. The development of such an ontology heavily depends on the administrative capacity of the community, which is generally easier for more integrated communities, which are characterized by steeper hierarchical structures.

The capacity criterion is seen as indifferent to the setting. It is in independent variable that is not affected by any of the community criteria mentioned above. Nevertheless concerns have been raised about the importance of topicality of person related information and how this topicality can be achieved.

Beside these aspects editorial actions are a crucial factor to keep the social process going and to balance single sided views. A domain community platform solely built on self-organizing principles is a necessary but not sufficient condition to get it started. Especially in the beginning incentives have to be offered to the members to get the community process running and keep it alive over the long run. Beside the incentives already described in the paper (automatic notification mechanisms) editorial content combined with a mix of offline social activities and finally the usability of the system and the quality of the content it produces can be mentioned.

Further fields of application of the model include monitoring complex decision making processes over time for documenting policy learning [4]. This is especially relevant in multi-stakeholder policy processes at EU level at the stages of program development and program implementation. Slight methodological adaptations would be necessary like time series analysis, enhanced semi-automatic annotation functions, linguistic analysis (discourse tracking). It will be necessary to develop governance recommendations, i.e., when and how the model fits best and what should be done by community managers in charge to bring the model to its full potential.

Acknowledgements

This work has been partly supported by the FFG SemNetMan project. We would also like to thank Klaus Valentin from profactor Steyr for his involvement in this work.

References

[1] Abels, G.; Bora, A., 2004. Demokratische Technikbewertung. Bielefeld: transcript Verlag

[2] Mead, G.H., 1934. *Mind, Self, and Society, From the Standpoint of a Social Behaviourist*. University of Chicago.

[3] Mika, P., 2005. Ontologies Are Us: A Unified Model of Social Networks and Semantics. In *Proceedings of the 4th International Semantic Web Conference*, Springer-Verlag, LNCS 3729, pp. 522-536.

[4] Sabatier, P.A., 1999. Theories of the Policy Process. Oxford Westview Press.

[5] Wasserman, S., Faust, K., Iacobucci, D., Granovetter, M., 1994. *Social Network Analysis: Methods and Applications*. Cambridge University Press.

[6] Zahariadis, N., 1999. Ambuigity, Time and Multiple Streams. In Sabatier, P.A., 1999. Theories of the Policy Process. Oxford Westview Press.

[7] Zhdanova, A.V., Krummenacher, R., Henke, J., Fensel, D., 2005. Community-Driven Ontology Management: DERI Case Study. In *Proceedings of the IEEE/WIC/ACM International Conference on Web Intelligence*, IEEE Computer Society Press, pp. 73-79.

[6] Austrian Best Bio-Gas Practice: http://www.oebn.at