Deploying Recommender Systems for Microcontent: An approach using Social Network Theory (Semantics for Microlearning)

Nikolaos Korfiatis
Department of Informatics
Copenhagen Business School (CBS) (Denmark)

Miltiades Lytras
Department of Computer Engineering and Informatics
Computer and Academic Technology Institute
University of Patras (Greece)

Abstract: The concept of microcontent poses a new set of challenges for the design of recommender systems that can assist the users to accomplish a broad set of complex informational tasks as well as to evaluate the importance of information resources such as microcontent structures. In this paper we formulate an approach by using an adaptation of the hubs and authorities model, in order to study the deployment of a recommender system for a microcontent structure such as Wikipedia.

1. Introduction

The concept of microcontent (Weinberger, 2002) apart from its implications for the field of eLearning (Hug, 2005), poses a set of challenges for the design of tools and services that can support and enhance apart from the eLearning process a broader spectrum of informational activities such as discovery of relevant pieces of information and support for social navigation (Dieberger et al., 2005).

On the other hand the enormous evolution of the World Wide Web (WWW) to the "lingua-franca" of content authoring, dissemination and accessibility has increased the volume of information available to the users. However, this also implies a cognitive load to those that want to use the web to support a set of informational activities and tasks, such as in the case of seeking learning resources (e.g. books). Studies of this cognitive load, which is addressed in the literature as information overload (Losee Jr, 1989), can be seen in several fields such as community design and in particular newsgroups (Borchers, Herlocker, Konstan, & Reidl, 1998), consumer behavior and marketing (Meyer, 1998), and to a large extent the web itself.

Modern search engines (e.g. Google – (Brin & Page, 1998)) can address cases of information overload where a resource is filtered by its popularity in a hypertextual context. However those search engines fail in cases where the hypertextual popularity is not correlated with the social context of the entities that produce this resource and are responsible for it.

Since the level of appropriateness of an informational resource to user's criteria is something that is characterized by a high level of cognitive complexity and cannot be extracted mechanically, a set of hybrid methods need to be established in order to make this kind of filtering more efficient to the eyes of the user. Such kind of methods is the family of collaborative filtering (CF) which has been developed to a set of systems that, based on collaborative filtering algorithms, provides recommendations to the users about the appropriateness of the content to their contextual needs (Shardanand & Maes, 1995).

In this paper we address the implications of microcontent to the design of recommender systems that extend the classical blackbox model and rely heavily on the socio-structural properties of the information resources in order to provide an indication of authoritativeness and trustworthiness to the user.

We consider the Hubs and Authorities model originally introduced by Kleinberg (Kleinberg, 1999b) as a departure point for the design of a recommender system capable of exploiting the advances that a microcontent structure provides. To this end our paper is organized as follows: Section (2) reviews the current state of recommender systems and the implications of microcontent to the current design of such systems. Section (3) explains the Hubs and Authorities model and how this can be incorporated in a recommender system. Finally, Section (4) conducts a case study in Wikipedia by using the hubs and authorities model, and Section (5) concludes with some remarks for future research.

2. Recommender Systems for Content and Microcontent

The use of recommender systems has been greatly advocated in electronic commerce (Shaffer et al., 1999), and several popular electronic marketplaces such as Amazon.com have incorporated recommender systems in order to be able to provide the consumers adapted interfaces, based on their preferences and needs. However, a more important aspect of recommender systems (or collaborative filtering) is the application of such techniques to online communities where filtering becomes a transposition of the word of mouth (Brown & Reingen, 1987).
Grouplens (Resnick et al., 1994) was the first implementation of a recommender system that applied collaborative filtering in the context of a community, thus delivering recommendations based on the ratings provided by community members. Since the introduction of Grouplens, several implementations of recommender systems have emerged, based on the same architecture (e.g. Movielens etc.) .

However, microcontent poses some challenges for the design of recommender systems that can be deployed on microcontent structures. In particular, we can identify the following issues when it comes to designing a recommender system for microcontent:

- **Expression of boundary and level of analysis:** Unlike traditional recommender systems, whereas recommendations are on concrete and explicitly defined objects such as movies, messages or documents, a critical issue is the definition of boundary and level of analysis to which the rating will be associated.
- **Expression of preferences and association with objects:** The subjectivity of rating scales is something that has been also discussed in the recommender systems community (Herlocker et al., 2004), however, people evaluate based on memories which can be subjective with the overall quality of the object. For instance, let us consider the evaluation of a learning module. Someone considers some of the parts positively, however, the guidance of the instructor was very poor or the learning material was of bad quality. Furthermore, the course may have been designed for learners with different learning needs. Therefore some parts were already known to the user and he/she received zero utility out of this. When it comes to the evaluation of the module, the user may consider the above in order to provide a rating based on his satisfaction/dissatisfaction. Will the rating scale be enough to capture the full range of the evaluation that the user wants to provide? Can the rating scale be associated with a part or all of the structure?

- **Aggregation of preferences and provision of the recommendation:** Shall we consider the association of the different modules when providing a recommendation? How does the social structure play a role to the recommendation? For instance, someone has taken a course on Object Oriented Design with Java and he has denoted that he has extensively taken courses on Object Oriented design with C++. Will his opinion count more than someone who has taken this course as a beginner? How can we weigh their ratings and preferences in order to provide a valuable recommendation?

Traditional Recommender Systems	Recommender Systems for MicroContent
Fixed Boundary: Recommendation can be explicit about books and courses limited to the lowest item (e.g. a book or a course and not part of a book or part of a course)	*Non Defined Boundary:* E.g. recommendations about a part of a book or a coure.
Explicit Preferences: Votes and preferences are mirrored to the book. Rating scales are the same for every item.	*Implicit Preferences:* Votes about parts of the book or the course (e.g. bad introduction, bad examples). Impact of rating scales on different parts.
Provision of Recommendation: Black box aggregation (item based filtering, nearest neighbor)	*Provision of Recommendation:* Aggregation based on hubness and authoritativeness of the evaluator. Use of the social structure and the semantic associations to influence the metrics.

Table 1: Aspects of traditional recommender systems and recommender systems for microcontent.

Aggregation is an important part of the function of a recommender system because the quality of the provided recommendation is much dependent on the way the collabora-

tive filtering algorithm calculates the similarity/dissimilarity of the user's profile with the data already gathered. However, user interaction is something that relies on both explicit and implicit data (not directly obtained) and may contain noise (Pescovitz, 2000).
The above constitutes our discussion agenda for the design of recommendation systems with emphasis to microcontent. We continue our approach with the introduction of concepts from social networks and in particular metrics of importance and prominence. We discuss an adaptation of the Hubs and Authorities model which can be used in the design of a recommender system for microstructures.

3. Hubs, Authorities and Social Networks

The importance of social networks for the study of learning communities and recommender systems has been advocated by many researchers (Downes, 2005; Rafaeli & Sudweeks, 1997) especially on an "a-posteriori" level of analysis of the social interactions that are formed through them. From its early introduction by Moreno (Moreno, 1953), Social Network Analysis (Scott, 2000) aims to unravel patterns of interactions between group members, which play a major role in the behavior of each individual, thus becoming an important indicator of the overall group activity.
Social network analysis develops the theoretical foundations for a set of measures of prominence such as centrality (Freeman, 1979; Friedkin, 1991) based on basic graph theoretic measures such as the inner and outer degree of a node.

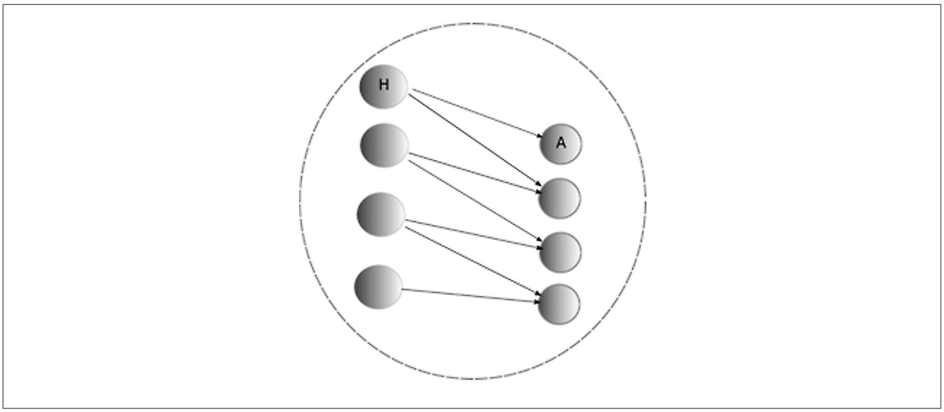

Figure 1: Hubs and Authorities on a topic boundary.

Furthermore, the concept of ranking has been studied in social networks long before the popular Pagerank algorithm was introduced (Page et al., 1998). In particular Katz (Katz, 1953) and later Hubbel (Hubbell, 1965) proposed models of ranking based on path counting and weight propagation over the network[1].

Kleinberg's work (Kleinberg, 1999a) has considered the transposition of authoritativeness over the nodes that have a relatively high inner degree – popularity. As can be seen in figure 1, hubs are characterized by a high level of outer-degree. That is the amount of links/connections departing from them is much higher than the amount of links/connections pointing to them. Taking into account the boundary of a topic, if a node is pointed by many hubs, it is considered as an authority, since flow is directed from many nodes to a single one.

In our approach we consider an adaptation of this model in order to interconnect two different graphs. In particular we consider the structural graph which contains the interactions between users and contributors of the content as well as any other kind of social interactions (e.g. ratings on raters etc). There is also a resource graph that considers the associations between the microcontent entities (e.g. topics, examples, forum posts etc.). Authoritativeness is attributed by the structural graph and hubness by the resource graph. The reason that we derive authoritativeness by the structural graph rather than the resource graph is that we start considering that all the content has the same importance factor initially set to zero. As long as someone attributes authority over the content by using it or rating it positively then this authoritativeness is transposed in the content.

Having provided the above we consider the deployment of a recommender system based on the above principles for a microcontent structure such as Wikipedia.

4. A case Study in Wikipedia

Although discussed above, the hubs and authorities model can be still quite confusing when it comes to incorporation on a recommender system. In relation with our approach in section 2 we consider an example of microcontent such as the articles that constitute a topic in the popular web based encyclopedia Wikipedia.

1 We are not going to elaborate further on graph theoretic aspects of those measures since we are only interested to the ideas that underline their development.

Wikipedia is based on Wiki software (Cunningham) and is considered to be one of the most successful collaborative editing projects on the web since it currently contains over 1 million articles[2] and has an extensive community of contributors contributing content and improving the quality of the articles.

Wikipedia is a voluntary project and since it facilitates a large amount of social interactions over a common affiliation, we consider it as an interesting example to discuss hubness and authoritativeness. In our case we consider the following social interactions:

- When a contributor edits content that has been submitted by someone else then it establishes a tie with him/her. This is depicted by an acceptance factor which represents the percentage of the content of the previous contributor that is visible afterwards.
- Every contributor that has a single or more contribution to the article establishes a relational tie with the other content contributors of the article. Evidence of participation in common projects strengthens this tie.

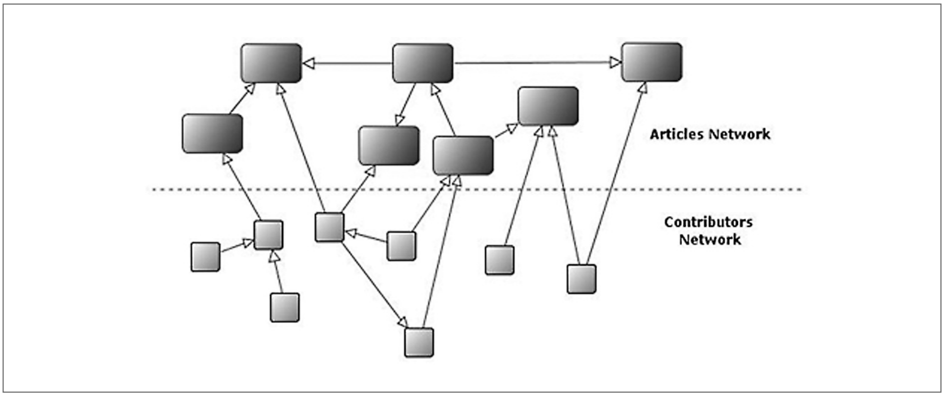

Figure 2: Network layers in the wiki publication model. Contributors are linked together by working on common projects (articles) in the same topic.

2 Statistics and data for the English Language Wikipedia. For further information about the current size of the wikipedia the reader can visit : http://en.wikipedia.org/wiki/Wikipedia:Statistics (Last access date 30th of May, 2006

As can be seen in figure 2 we define two different networks: the articles network and the contributors network.

- **The Articles Network:** Every article in Wikipedia contains references to other articles as well as external references. A set of links used for classification purposes is also available in most of the active articles of the encyclopedia. Every article represents a vertex in the article network and the internal connections between the articles the edges of the network.
- **The Contributors Network:** Wikipedia is a collaborative writing effort which means that an article has multiple contributors. We assume that a contributor establishes a relationship with another contributor if they work on the same article. In the resulted weighted network each contributor is represented by a vertex, and their social ties (positive or negative) are represented by an edge denoting the sequence of their social interaction.

Supposedly that we want to introduce a collaborative filtering system in Wikipedia that can evaluate the trustworthiness of the articles contributed. Based on the above formalization we can initialize a discussion on the following cases by using the hubs and authorities model discussed in Section 3:

- **High Article Hubness and low Contributor Authoritativeness:** In that case an article that directs to many others e.g. an index of the municipalities in Europe has been authored by contributors with low authoritativeness. This can misguide the reader so the articles that are linked from this article can also have a low degree of authoritativeness. So the article is not recommended for reading.
- **High Article Hubness and High Contributor Authoritativeness:** An important article has been written by experts on the field so the article is recommended for reading.
- **Low Article Hubness and low Contributor Authoritativeness:** A not so important article has been contributed by non experts so the article is not recommended.
- **Low Article Hubness and High Contributor Authoritativeness:** A not so popular article is written by experts on the field. This is one of the important cases in Wikipedia where a serious piece of work is not visible to others.

5. Conclusions and Future Research

We have presented the hubs and authorities model as a formal framework for the design of recommender systems for microcontent such as Wikipedia content. However this work is only based on some theoretical considerations therefore a number of research questions still stay open such as:

- Design of rating scales for different parts of the same structure
- Cases where negative authoritativeness influences negatively significant content.
- Evaluation of this class of recommender systems. Current recommender systems are evaluated using a set of guidelines (Herlocker et al, 2004). However as aforementioned, in our approach recommender systems for microcontent rely heavily on the socio-structural context whereas evaluation of those factors needs to be done very carefully.

Furthermore evaluation of that kind of model with data elicitated from a large sample from the Wikipedia is also an important matter.

6. References

Borchers, A., Herlocker, J., Konstan, J., & Reidl, J. (1998). *Ganging up on information overload. Computer, 31*(4), 106–108.

Brin, S., & Page, L. (1998). The anatomy of a large-scale hypertextual web search engine. *Proceedings of the Seventh International Conference on World Wide Web.*

Brown, J. J., & Reingen, P. H. (1987). Social ties and word-of-mouth referral behavior. *The Journal of Consumer Research, 14*(3), 350–362.

Downes, S. (2005). Semantic networks and social networks. *The Learning Organization Journal, 12*(5), 411–417.

Dieberger, A., Dourish, P., Hook, K., Resnick, P., & Wexelblat, A. (2000). Social navigation: Techniques for building more usable systems. *Interactions, 7(6)*, 36–45.

Freeman, L. C. (1979). Centrality in social networks: Conceptual clarification. *Social Networks, 1*(3), 215–239.

Friedkin, N. E. (1991). Theoretical Foundations for Centrality Measures. *The American Journal of Sociology, 96*(6), 1478–1504.

Herlocker, J. L., Konstan, J. A., Terveen, L. G., & Riedl, J. T. (2004). Evaluating collaborative filtering recommender systems. *ACM Transactions on Information Systems*

(TOIS), 22(1), 5–53.

Hubbell, C. H. (1965). An input-output approach to clique identification. *Sociometry, 28*(4), 377–399.

Hug, T. (2005). Microlearning: A new pedagogical challenge. In *Microlearning: Emerging Concepts, Practices and Technologies : Proceedings of Microlearning 2005.* Innsbruck, Austria. Innsbruck University Press.

Katz, L. (1953). A new status index derived from sociometric analysis. *Psychometrika, 18,* 39-43.

Kleinberg, J. M. (1999a). Authoritative sources in a hyperlinked environment. *Journal of the ACM (JACM), 46*(5), 604–632.

Kleinberg, J. M. (1999b). Hubs, authorities, and communities. *ACM Computing Surveys, 31*(4es)

Losee Jr, R. M. (1989). Minimizing information overload: The ranking of electronic messages. *Journal of Information Science, 15*(3), 179–189.

Meyer, J. A. (1998). Information overload in marketing management. *Marketing Intelligence & Planning, 16*(3), 200-209.

Moreno, J. L. (1953). *Who shall survive? Foundations of sociometry, group psychotherapy and Sociodrama.* Beacon House.

Page, L., Brin, S., Motwani, R., & Winograd, T. (1998). *The PageRank citation ranking: Bringing order to the Web.* Stanford Digital Library Technologies Project.

Pescovitz, D. (2000). Accounting for taste. *Scientific American, 282*(6)

Rafaeli, S., & Sudweeks, F. (1997). Networked interactivity. *Journal of Computer Mediated Communication, 2*(4), 1997.

Resnick, P., Iacovou, N., Suchak, M., Bergstrom, P., & Riedl, J. (1994). GroupLens: An open architecture for collaborative filtering of netnews. *Proceedings of the 1994 ACM Conference on Computer Supported Cooperative Work,* 175–186.

Scott, J. (2000). *Social network Analysis: A handbook* (2nd ed.). London; Thousands Oaks, Calif.: SAGE Publications.

Shardanand, U., & Maes, P. (1995). Social information filtering: Algorithms for automating ''Word of mouth''. *Proceedings of ACM CHI'95 Conference on Human Factors in Computing Systems.*

Schafer, J. B., Konstan, J., & Riedi, J. (1999). Recommender systems in e-commerce. *Proceedings of the 1st ACM Conference on Electronic Commerce,* 158–166.

Weinberger, D. (2002). *Small pieces loosely joined: A unified theory of the Web.* Perseus Books Group.