

Bringing together what belongs together: A recommender-system to foster academic collaboration

Fridolin Wild¹⁾, Xavier Ochoa²⁾, Nina Heinze³⁾, Raquel M. Crespo⁴⁾, Kevin Quick¹⁾

¹⁾The Open University, UK

²⁾Escuela Superior Politecnica del Litoral , Ecuador

³⁾Knowledge Media Research Center, Germany

⁴⁾Universidad Carlos III de Madrid, Spain

Introduction

The amount of information available to researchers today has increased in the last years at an unfathomable speed. Web 2.0 technology, especially social network tools and new communication platforms have added more points of access to the exploding number of informational offers. As scientists are increasingly using Web 2.0 for research, knowledge management, communication, and collaboration it is becoming more and more important for the users of these tools to be able to filter out (from their perspective and in the given situation) unwanted information. Moreover, it is even more important for researchers to filter through known and even unknown content to find *relevant* information, broaden their scope of knowledge about a certain field, remain up-to-date on the state-of-the-art and expand their reach to other institutions or domains.

One possibility to make users aware of content available in their field of interest is through the use of recommender systems. We would like to outline how we propose a new recommender system using a method-mix of data mining and social network analysis (SNA). This science proxy project is targeted towards resolving unwanted fragmentation with the help of recommendations extracted from publishing data from conferences such as EC-TEL and – on the intervention side – with the help of the networked communication instrument ‘Flashmeeting’¹. This recommender system aims at supporting scientists in the field of TEL to increase the quality of their collaboration in the wider community of peers.

Methods for the inspection of network structures

Data mining is used to transform data into meaningful information. For our purposes we used datasets of registered users from the *Flashmeeting* project as well as publication meta-data derived from EC-TEL conferences. The process known as Knowledge Discovery in Databases (KDD) (Fayyad et al. 1996), which primarily entails pre-processing raw data, mining the data, and interpreting the

¹ <http://flashmeeting.open.ac.uk/home.html>

results, was used to gather data to get a first impression of what the possibilities and the potential of our future development are.

Social network analysis (SNA) is considered a distinct research perspective within the social and behavioural sciences. Its main focus lies on the relations between actors in a defined network. This method can be used to examine a number of topics within networks including structural characteristics, linkage of actors and groups, diffusion of innovations, and transformation of network structures in parts and as a whole. In the Flashmeeting project we propose using the method of SNA to analyse the participation of the actors within the field of TEL in relation to each other as well as to analyze research topics and their diffusion or fragmentation respectively. This leads to the detection and characterisation of community structure in networks (Newman 2006), which in turn will let us identify several key aspects relevant for developing a recommender system like related or unrelated actors in similar research areas, diffusion paths of topics throughout the network, related research themes, or an overlap of certain communities in regards to topics. Data mined in the KDD process beforehand is used in this step.

Prospective science proxies

The goal of developing a recommender system for the *Flashmeeting* project is to use meta-data to support researchers by pointing out other projects, researchers, or related topics they may not be aware of yet and that are closely related to their field of interest.

Scientific community structure is revealed in a number of communication channels (events, authorship of papers, on-line meetings, etc). Data from different sources should thus be integrated and combined in order to grasp an accurate impression of the actual state of the art. The basic idea of this set of aligned recommender tools draws from the possibility to build links from comparing relationship information in two independent data sets.

The first data set contains publication meta-data from a conference series specialising on the topic of technology-enhanced learning. This set contains information on authors and their co-authorship relations. Titles, abstracts, and full texts provide means to extract representative keyword descriptors using language technology: using a combination of latent-semantic and network analysis, the medium frequent keywords at the heart of clusters in a latent-semantic term-to-term interaction space can be extracted as good descriptors for the papers by which they are activated. Additionally, the data set contains references in a structured format.

The second data set contains virtual presence information captured from meeting attendance in a virtual conferencing tool popular in the same community of authors.

By comparing these data sets, differences can be spotted which allow to influence the structure of the community (as represented in the relationship networks) by mutually proposing recommendations made possible through information gained from the complementary data set.

Recommenders made possible with this approach range from group forming to group activity support. They include but are not limited to:

- *Defrag meeting recommender*: from the co-authorship network and the co-citations therein, a recommender can identify when authors are working on the same topic (=keywords) but with different co-authors and different literature. This can be a strong indicator of unwanted fragmentation. A recommender can be trained that proposes to hold a 'getting to know each other' *Flashmeeting* that may initiate desired defragmentation.

For example, analyzing the ECTEL co-authorship and co-citation graphs, Mohamed Amine Chatti, a researcher from RWTH Aachen University, Germany is an isolated member connected only to his 2 co-authors. A textual analysis of the content of his paper reveals that he is working on the automated annotation of learning materials. If that topic is searched inside the ECTEL collection of papers, it appears that there is a strongly related group of researchers working in the same field (centred on Alexandra Cristea). The location of Mohamed and Alexandra group could be seen in Figure 1. The recommender could suggest Alexandra and Mohamed to meet.

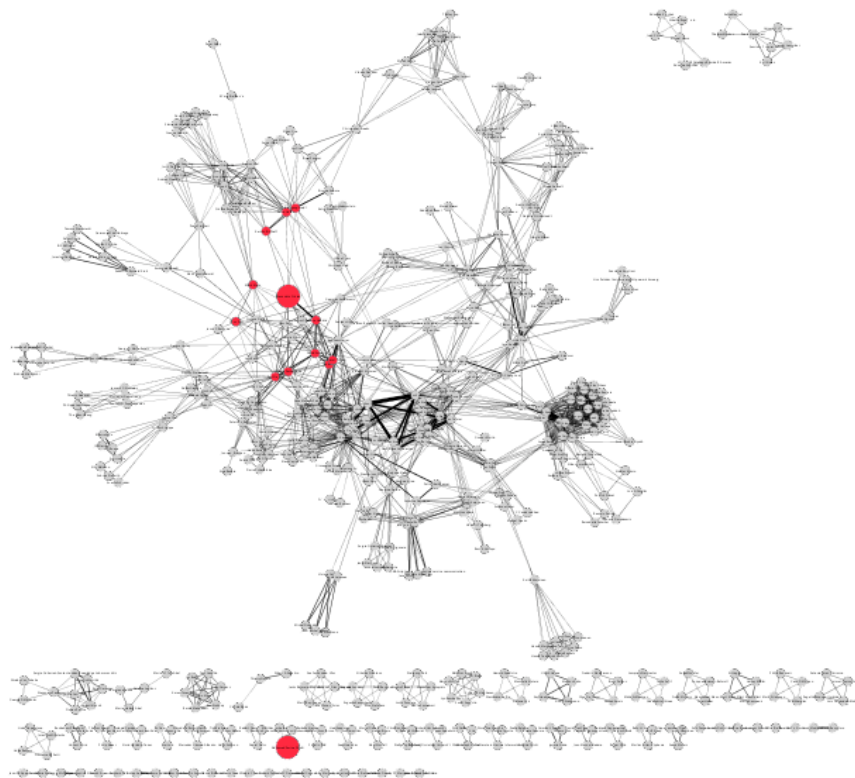


Fig. 1. ECTEL Co-citation graph with the positions of Mohamed Amine Chatti and the group of Alexandra Cristea

- *Group proposal recommendation*: existing cliques can be discovered from graph components, recommending their members to form a group for supporting the management of joint meetings.
- *Group closing recommendation*: lack of activity in a group may indicate that it no longer exists as such. Confirming group disappearance would be necessary for keeping the server tidy and an accurate map of existing active communities.
- *Group access recommender*: when raising awareness about existing groups for a given individual, the participation of his/her contacts in a certain group is a strong indicator about the interest of the group for such a person. Recommendations for joining a given group based on contacts' membership can help to avoid missing information.
- *Meeting invitation recommender*: awareness of community specific events can also be improved. Based on the known participants in the event as well as their contact relations, recommendations can be made for potential attendants.
- *Defrag group recommender*: communities are far from homogeneous. Sub-groups can emerge, particularly in big communities, which are connected by a small set (two or three) of members acting as bridge builders between otherwise disconnected components in the interaction graph. Alerts about such structural dysfunctions including the provision of solutions such as joint virtual meetings can help to mend them and improve effective collaboration inside the global community.

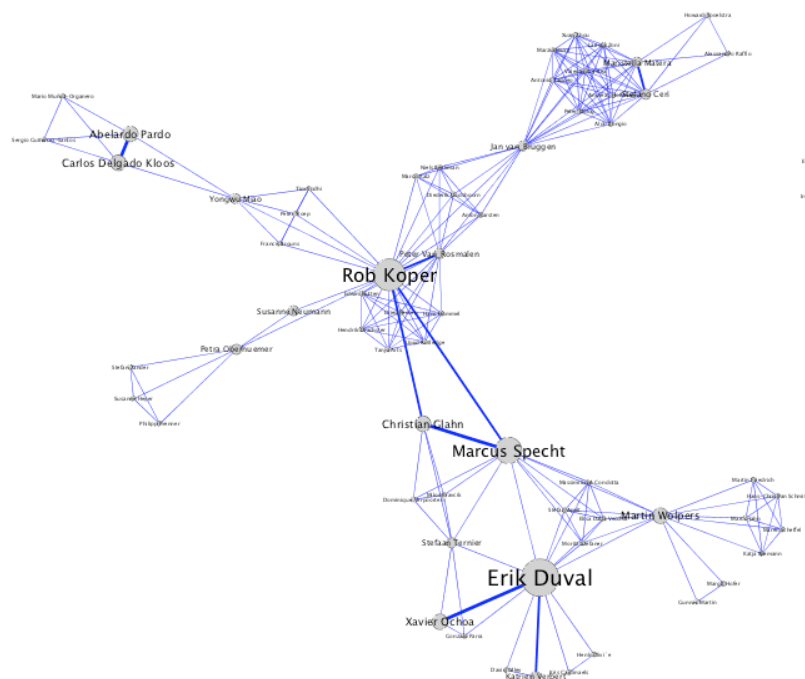


Fig. 2. ECTEL Co-authorship graph showing the linkage between OUNL and KULeuven research groups.

An example that can be extracted from the ECTEL co-authorship graph is the structure of the collaboration of the research groups lead by Rob Koper, from Open University Nederlands and Erik Duval, from KULeuven (Figure 2). The main linkage between these two groups is Marcus Specht. Encouraging joint meetings between members of these two groups could lead to stronger collaboration and a healthier structure.

Evaluation Methodology

We will conduct quasi-experiments once we have developed a closed beta-version of the planned recommender systems to analyse and evaluate our research results as well as our implementation.

Thereby, the accuracy and usefulness of as well as the satisfaction provided by the recommendations have to be carefully evaluated. There are tensions between different aims in tuning of the algorithms, not least between optimising for accuracy and optimising for serendipity.

As Herlocker et al. (2004) point out, the user tasks, the types of analysis and data sets being used, the focus of the evaluation, and the mode of evaluation all influence the outcomes of such an investigation.

Our preliminary plan for an experimental evaluation encompasses a user survey asking for individual relevance ratings of each recommendation and uses a control group provided with random recommendations to statistically test effect size and significance of the influence of the recommendation algorithms. If the test group with recommendations based on the recommender algorithms ranks their recommendations significantly more relevant than the control group that is provided with random recommendations, the algorithms shall be considered more useful than random chance.

References

Fayyad, Usama; Piatetsky-Shapiro, Gregory; Smyth, Padhraic (1996). "From Data Mining to Knowledge Discovery in Databases". In: *AI Magazine*, 17(3), pp. 37-54. <http://www.kdnuggets.com/gpspubs/aimag-kdd-overview-1996-Fayyad.pdf>. Retrieved 2008-12-17.

Herlocker, Jonathan; Konstan, Joseph; Terveen, Loren; Riedl, John (2004): Evaluating Collaborative Filtering Recommender Systems, In: *ACM Transactions on Information Systems*, Vol. 22, No. 1, pp. 5-53.

Newman, M. E. J. (2006): "Modularity and community structure in networks", In: *Proc. Natl. Acad. Sci. USA* 103, 8577-8582.