Proceedings of the Twenty-Second International FLAIRS Conference (2009)

Discovering Patterns of Collaboration for Recommendation

Sidath Gunawardena, Rosina Weber
The iSchool at Drexel, College of Information Science & Technology, Drexel University
{sidath.gunawardena, rweber}@ischool.drexel.edu

Abstract

Collaboration between research scientists, particularly those with diverse backgrounds, is a driver of scientific innovation. However, finding the right collaborator is often an unscientific process that is subject to chance. This paper explores recommending collaborators based on repeating patterns of previous successful collaboration experiences, what we term prototypical collaborations. We investigate a method for discovering such prototypes to use them as a basis to guide the recommendation of new collaborations. To this end, we also examine two methods for matching collaboration seekers to these prototypical collaborations. Our initial studies reveal that though promising, improving collaborations through recommendation is a complex goal.

Introduction

Collaboration in science and engineering is one of the main drivers of scientific progress. Today, science faces grand challenges that require teams of scientists and engineers with complementary expertise. These include environmental science challenges such as how to address local and regional climate variability, health challenges such as how to create new vaccines, and global challenges such as eradicating poverty and hunger (Omenn 2006). Boundary spanning collaboration is believed to be so essential to groundbreaking scientific discoveries that US federal agencies mandated to promote science, such as the National Science Foundation (NSF), put a strong emphasis on interdisciplinary and inter-institution collaboration when awarding grants (NSF 2006).

Despite the importance of collaboration, researchers typically rely on chance meetings at conferences and casual conversations that bring to light knowledge about a potential collaborator. One of the roles of artificial intelligence (AI) is to assist humans in solving problems by revealing new solutions that could not be discovered or that would take great efforts to achieve without the help of the tool. The problem of discovering collaborators is not solely one of cognitive overload on the part of a human user; simply manipulating data is not sufficient. Thus, a reasoning task needs to be performed. An AI-based solution is desirable to recommend collaborators. There are existing technological options that can provide some assistance, ranging from search engines, to online communities geared towards scientific research such as the Community of Science (www.cos.com), to expert locator systems which can recommend individuals with a particular pre-specified expertise. However, these are makeshift solutions, not built to solve this particular problem. The burden is placed on the user, or collaboration seeker, to laboriously search for collaboration partners. Social networking theories can be useful for finding collaborators in the same or very closely aligned domains (Wohlfarth and Ichise 2008). When collaborators are required from disparate disciplines, these solutions becomes less effective. Finding such collaborators requires increased effort (Kreiner and Schultz 1993). Factors such as multiple domain vocabularies and the lack of past experiences to draw on, make finding multidisciplinary collaborators a daunting task. Lacking a satisfactory technological solution, or scientific methodology, humans use their best judgment and gut instincts when going about finding collaborators. Alternatively, collaborations may arise from individuals who act as bridges by maintaining ties in different domains, giving them the ability to connect collaboration partners (Crane 1972; Granovetter 1973). This paper's intended contribution is to further our understanding of recommendation of collaborations. Our ultimate goal is to create a system that can automatically unearth collaborators who can enrich the quality of a researcher's collaboration experience

In the next section we formulate our problem and discuss the idea of prototypical collaborations; repeating patterns of collaboration that can be leveraged to make recommendations. In the following section we explore this concept using an approach that extracts repeating patterns of collaboration from two datasets of grant information that are used as proxies for these past experiences of collaboration. We mine association rules to discover patterns of collaboration. Finally, we discuss some background work and then conclude in by listing some future work.

Problem Formulation

We perceive collaborations as experiences; where past successful collaborations can provide a blueprint for future ones. However, each individual only has a handful of such experiences to draw upon. Thus, we propose identifying collaboration prototypes; recurring patterns of collaboration that have been successful in the past. As we model collaborations as experiences and contend that past examples of successful collaborations can be used to recommend new collaborations, we adopt case-based reasoning (CBR) (Aamodt and Plaza 1994) as our methodology.

Collaborator recommendation is closely related to expert locator systems (ELS) (Becerra-Fernandez 2006; Maybury 2002). The task of an ELS is to recommend qualified experts to a user who has a need for a particular expertise. The primary contrast between the two approaches is that in the case of the ELS, the user species the criteria of the expert they seek. However, with a collaboration recommender, the user specifies details about themselves, which the system then uses to generate the recommendations. An ELS typically builds profiles on experts either by extracting expertise related data such as publication information from websites, from document collections located on intranets (if they are in house systems), or by using self-reported proficiency information (Becerra-Fernandez 2006). When seeking a collaborator, factors additional to expertise need to be included. The research interests of the collaboration seeker are the most obvious. In addition, seniority, collaboration style, geographic location, and experience in the field all play a part.

In the case where a user seeks an expert, the user has a clear idea of the expertise required, and relies on the system to indicate individuals with the desired expertise. What the user seeks in a collaborator is more vague and ill-defined. Furthermore, the collaboration seeker likely does not know all the domains where suitable collaboration partners reside. We see the location of the expert as the last step of the process of recommending a collaborator.

Collaboration is an idiosyncratic process, and when it happens across disciplinary boundaries it can create or exacerbate issues such as trust, the need for negotiation, and the need for a common vocabulary (Jeffrey 2003). Thus, factors that can mitigate such problems need to be taken into account. We propose an approach that takes into account the commonalities of research interests of all parties involved. Such a commonality increases the likelihood of both mutual respect between the parties involved and common mental models and vocabularies.

When recommending collaborators, there are several options as to the nature of the recommendation. The recommendation could be in the form of suggesting potential collaborators that are a good fit or it could be a recommendation of a prototype of collaboration best suited to the collaboration seeker, or it could be a combination, where both the type of collaboration and potential collaborators are recommended. We begin by exploring the prototypes of collaboration. Although we acknowledge multiple formulations to the problem of improving collaborations through recommendations, this paper focuses on one approach. We break down recommendation, first recommending potential collaboration prototypes and second submitting descriptions of individuals from recommended prototypes into an expert locator to identify instances of researchers who match the description in the prototypes.

Prototypical Collaborations

Our ultimate goal (and main hypothesis) is to enhance the quality of the collaboration experience of researchers. We also do not consider in this paper many factors that lead to a successful collaboration such as compatibility, the socio-technical environment etc (Hara et al. 2003).

There are myriad dimensions to consider when formulating this problem. To make this initial investigation more manageable, we narrow our scope to consider only research interests, and study them at two different levels of granularity. As we prepare further studies to better understand collaborative success, we now adopt obtaining grant funding as our definition of a discovered successful collaboration opportunity.

A collaboration can be a multi-party endeavor with participants from several domains. Thus, the collaboration in its entirety may constitute a prototype, or the prototype may only be a subset of the members involved in the collaboration. The goal is to identify the pattern that consistently repeats over time, indicating a successful collaboration. Thus, our cases in the case base would not be comprised of all past collaborations, but only the wholes and subsets of collaborations that occur repeatedly. Table 1 shows an example of how a simple prototype is generated. In the example, the repeating pattern is of an anthropologist collaborating with a geologist. Our first goal then, is to determine what constitutes a collaboration prototype.

Collaboration 1	Sociologist + *Anthropologist* + *Geologist*
Collaboration 2	Archeologist + *Anthropologist* + *Geologist*
Prototype	*Anthropologist + Geologist*

Table 1: Example of a collaboration prototype

Recommending Collaborations

Once we have established our collection of prototypes, the next step is to use them to provide recommendations. We envision a collaboration seeker providing their research interest or interests (in our simplified study that is all that we require), but now we need a metric to determine which collaboration prototype is the most useful to this particular collaboration seeker. The standard methodology in CBR would be to assess similarity between each member of a prototype and the collaboration seeker. However, similarity is a proxy measure for usefulness. In this problem, similarity assessments based solely on the similarity between the collaboration seeker and individual members of a prototype may not be sufficient to indicate usefulness. Opposed to the distance measure usually adopted in CBR systems as a reference to usefulness we propose an alternate measure of usefulness. In this recommender system, we match seekers to prototypes through a measure of fit.

Such a measure matches a seeker's research interests to prototypes by considering how similar the collaboration seeker is to the entire collaboration prototype. Taxonomies of the various domains allow for the resolution of research interests stated at different levels of specificity, e.g., geology is a type of earth science. The taxonomies distinguish the similarities between fields e.g., a geologist is more closely related to a seismologist than to a biochemist.

The output is a recommendation of the type of collaboration prototype that the user should seek out. Using the prototype in Table 1as an example, if the collaboration seeker is a geologist, then the system's output is to seek collaborators in the domain of Anthropology. Then, existing solutions such as an expert locator can determine the best experts for a suggested collaboration prototype. While we consider no additional factors than expertise when making our recommendation, the utility over simple expert location comes from the fact that the collaboration seeker only has to specify their own interests. Future iterations of our research will consider additional factors when making this recommendation.

A new problem is characterized by a collaboration seeker who is matched against prototypical cases. The goal is to produce a measure where the highest ranked prototypes indicate the collaborations that are most likely to result in success. We now proceed to explain in greater detail our methodology, starting with how to discover prototypical collaborations.

Discovering Prototypical Collaborations

In this section we discover prototypical collaborations via learning association rules. The discovery of patterns through the use of Association Rules is a standard procedure in data mining. We use the classic algorithm for discovering association rules, Apriori (Agrawal and Srikant 1994), to uncover prototypical collaborations. Our analysis was performed using the WEKA data mining package (Witten and Frank 2005).

Due to requirements of the analysis, we use two different data sets to perform our experiments. The prototypes are based on success in obtaining competitive funding. The datasets used for the two experiments are comprised of grants awarded by the National Institute of Health (NIH) and by the NSF, respectively. The NIH dataset comprised of 201 grants, each grant is associated with a set of research activities. The number of research activities associated with a grant ranged from 4 to 53, with an average of 14.6 research activities per grant. A weakness of the NIH data set is that the research activities are associated with the grant and not the participants.

The dataset comprising grant information from the NSF was from the period 2006-2008. Additionally, investigators' research interest data was obtained from departmental websites and online resumes. Our dataset comprises 80 grant proposals, with 220 participating PIs and Co-PIs, encompassing 34 different fields of research interests. All collaborations are multidisciplinary in nature with at least 2 members coming from different domains. The average number of domains in a collaboration is 2.75. The number of participants in a collaboration ranges from 2 to 6.

We perform this analysis at two levels of granularity: at a research interest level (e.g., cancer prevention) using the NIH dataset, and at a higher level of granularity based on the collaborator's domain (e.g., mechanical engineering) using the NSF dataset.

Association Rules from Research Activities

For the purposes of generating association rules we use a minimum support parameter of 5%. This requires that there be at least 5% of the total cases that contain the antecedent of a rule present in the data for it to be considered valid, i.e. a minimum of 10 instances in our NIH data set. We used a minimum confidence threshold of 0.5. The confidence threshold allows the selection of rules that are correct at least 50% of the time. Table 2 displays the resulting rules and confidence values.

Association Rule	Conf.
human subject + human therapy evaluation → clinical research	1.00
opioid receptor → peptide analog	1.00
neuropeptide → peptide analog	0.91
peptide analog → opioid receptor	0.83
peptide analog → neuropeptide	0.83
peptide analog → drug design/synthesis/production	0.83
chemoprevention → cancer prevention	0.79
drug design/synthesis/production → peptide analog	0.56
cancer prevention → chemoprevention	0.55
antineoplastic → drug screening /evaluation	0.50

Table 2: Research Interest prototype association rules

The first association rule shows that collaborations with *human subject research* and *human therapy evaluation* also always included *clinical research*. This overly general result highlights a characteristic of the data set that the research activities are associated with the collaboration and not the individual, for clarity, 16 other of these overly general prototypes have been removed from the list and also from the analyses that are described later in this paper. While the average number of research activities is 14.6, the the rules generated show that there are very few large scale repeating patterns within the dataset. However, within these results, interesting rules can be found. The rules found here largely describe subsets of collaborations, but are the starting point that we seek. We now describe our second attempt at discovering prototypes.

Association Rules from Domain Information

We repeated the process of generating association rules for the NIH data set of high level domain information. Here, we ignore the number of members in each domain within a collaboration. For the purposes of this experiment, we do not care how many of a particular type of researcher was involved in a collaboration, we only care about the presence or absence of that type in the collaboration. We use a minimum support parameter of 5% (4 collaborations in our NSF set). In this experiment, the minimum confidence threshold used was 0.25, i.e. a selected rule must be correct at least 25% of the time. These parameters set thresholds that are lower than the previous experiment as each collaboration has far fewer attributes in common with other collaborations, when compared to the NIH dataset used in the previous experiment. Thus, nature of this dataset is such that it becomes necessary to use a low threshold to discover rules. Table 3 shows the association rules that were generated.

Association Rule	Confidence
Ecologist → Mathematician	0.56
Ecologist → Biologist	0.44
Mech. Engineer → Mathematician	0.42
Physicist → Chemist	0.42
Electrical Engineer → Computer Scientist	0.40
Computer Scientist → Biologist	0.38
Communication → Mathematician	0.36
Chemist → Physicist	0.36
Physicist → Mathematician	0.33
Mech. Engineer → Computer Scientist	0.33
Biologist → Computer Scientist	0.30
Chemist → Biologist	0.29
Mathematician → Mech. Engineer	0.25
Mathematician → Ecologist	0.25
Computer Scientist → Mech. Engineer	0.25
Computer Scientist → Elec. Engineer	0.25
Computer Scientist → Mathematician	0.25

Table 3: Domain prototypes association rules

The first association rule shows that when an ecologist is engaged in a collaboration, 56% of the time it was with a mathematician (note that in reverse, mathematicians only collaborated with ecologists 25% of the time).

Recommending Prototypes

The resulting prototypes discovered in the previous section describe collaborations at different levels of granularity. We propose to examine these levels to determine how well each performs when used to recommend collaborations. The goal is to determine the effectiveness of the resulting recommendations.

A collaboration is a special instance of a case in CBR, where the problem and solution are defined after similarity assessment. When assessing similarity between a new problem (a collaboration seeker) and a candidate case (a collaboration), the seeker must be compared to all members in the collaboration. Thus, the collaborator that the seeker best matches becomes the problem part of the candidate case, and the solution part becomes the rest of the collaboration; the problem-solution pair only becomes defined after similarity assessment. To take into account this aspect of the problem we test two ways of measuring similarity. The first takes the best match, in terms of similarity, between the features of the new problem (the collaboration seeker) and the features of each element of the candidate case (members of the collaboration). In our experiment the features are the research domain or research interest of the collaboration seeker and those of the members of the collaboration. We term this the *individual method;* it is the equivalent of a traditional CBR approach. Instead of assessing similarity, the second method considers the fit between the new problem (collaboration seeker) and the candidate case (the collaboration as a whole). Fit is determined by how similar the features of the collaboration seeker (research interest or domain) are to those that comprise the collaboration as a whole. We term this the *aggregate method*. This is to cater to the possibility there is knowledge in a case as whole that is lost when it is decomposed into its components parts.

Experimental Design

Based on the prototypes discovered, in this study, we model the collaboration seeker as having only one associated attribute. In the case if of the NIH data this is one research interest, in the case of the NSF data it is the domain of the collaboration seeker. To assess similarity and best fit, for the NIH data, we used text matching to assess similarity and for the NSF data we in addition incorporated domain taxonomies to determine the relation of one domain is to another.

The taxonomy for NIH is based on the keyword taxonomy used by the Community of Science (www.cos.com). The taxonomies for the NSF data are obtained from National Academies' Board of Higher Education and Workforce (http://www.nationalac ademies.org/bhew). Here, taxonomies spanning the domains of Life Sciences; Physical Sciences, Mathematics, and Engineering; Social and Behavior Sciences; Arts and Humanities are used, each consisting of three levels. Similarity between nodes is determined as follows:

- 1.00 if both domains share the same node
- 0.85 if they are sibling nodes
- 0.40 if they have parent-child relationship
- 0.00 otherwise

To evaluate the two methods, we define as a measure of accuracy the percentage of the top 5% recommended collaborations (10 in the case of the NIH data and 4 in the case of the NSF data) that include a prototypical collaboration learned through the association rules. The measure of accuracy is limited to the top 5% of recommendations because the 5% support parameter chosen for learning the association rules ensures that there will be at least that many occurrences of each prototype. Thus, 100% accuracy occurs in the case where the top 5% of recommendations are comprised fully of prototypical collaborations. Using this as a basis we performed a leave-one-out cross-validation whose results are described in the following section. Our hypotheses:

H1: the more specific research interest based prototypes will have a higher accuracy than the domain based ones.

H2: the aggregate method will be more accurate than the individual method.

Results & Discussion

Table 4 and Table 5 show the accuracy of the two methods using the prototypes generated by the association rules as a basis for recommending collaborations. It also groups the results by the confidence of the rules and shows the cumulative number of rules that fall within that range.

Confidence	**>=0.95**	**>=0.80**	**>=0.65**	**>=0.5**
# of prototypes	**1**	**5**	**6**	**9**
Individual Method	100%	90%	80%	80%
Aggregate Method	90%	82%	70%	65%

Table 4: Accuracy of rules based on res. interest (NIH data)

Confidence	**>=0.50**	**>=0.40**	**>=0.30**	**>=0.25**
# of prototypes	**1**	**5**	**11**	**16**
Individual Method	50%	35%	30%	25%
Aggregate Method	75%	70%	50%	44%

Table 5: Accuracy of rules based on domain (NSF data)

The results show that the rules generated from the more granular research activities perform at a higher level of accuracy than the domain data. This suggests that prototypes will be more effective when created from low granularity data (H1). Unsurprisingly, both measures show decreasing accuracy as confidence decreases.

When comparing methods, the aggregate method is superior only in the case of the domain-based association rules. Thus, it is inconclusive whether similarity assessment should be based on individual elements of a prototype or whether it should consider the prototype as a whole (H2).

This result may be attributable to the nature of the data. The NSF data set is a lot more diverse and is at a higher level of granularity than the NIH dataset. Thus, when the potential collaborators have less degrees of obvious commonality the aggregate method may prove to be superior in providing recommendations. The next step of this research will be to test both methods on a single dataset that has multiple levels of granularity.

Background

The first step in designing a system to facilitate collaboration is to understand the reasons behind why individuals collaborate. Table 6 combines motivations for collaboration gathered from (Bozeman and Corley 2004) and (Katz and Martin 1997). Our methodology does not cater to all the motivations, but it does cover collaborators who are motivated by the need to access expertise, who wish to work with members of other disciplines, who seek to pool knowledge, and who need to collaborate with others due to their own specialization.

1. access to expertise
2. access to equipment and resources
3. access to funds
4. cross-fertilization of ideas across disciplines
5. increased visibility and prestige
6. to learn tacit knowledge about a technique
7. to pool knowledge
8. to enhance productivity
9. to educate a student
10. the increased specialization of science
11. for fun and pleasure
12. the escalating demands on scientists

Table 6: Motivation for collaboration

Recommender systems have been designed based on different paradigms. Some well known categories include collaborative, content-based, demographic, utility-based, and knowledge-based (Burke 2002). Burke (2002) explains that one hurdle recommender systems used for e-commerce must overcome is that they must recommend new options and not ones previously encountered by the user. In the case of recommending collaborators, this problem is exacerbated as the goal of the system is not to recommend potential collaborators with exactly the same research interests. Nonetheless, the system has to be designed to recognize that there must be some common ground between the two parties. Without some minimal commonality to create a shared vision and purpose, it is unlikely that a successful collaboration can take place (Mattessich 2001). Balancing these two conflicting demands is a challenge in developing such a recommender system.

Many successful recommender systems rely on collaborative filtering (Cotter and Smyth 2000; Miller et. al. 2003). These systems rely on the assumption that users can be fit in groups whose members are eligible to similar recommendations. This is an alternative approach we plan to test. The goal of identifying multiple collaborators at once may pose an additional challenge.

The initial tests we are describing in this paper evaluate a form of reactive recommendation. Nevertheless, as users interact with the expert locator module, recommendations of collaborations can be proactive, where a query is not necessary (Smyth and Cotter 1999).

Another aspect to explore in this type of recommendation is a dialogue where critiquing may be included (Bridge et al. 2005). This becomes applicable when multiple dimensions are used to describe a collaborator and the collaboration seeker may wish to change the emphasis given to a particular dimension. These preliminary single-shot tests are meant to explore the problem and identify particularities we can later utilize for enriched recommendations.

Conclusions and Future Work

In this paper, we have introduced the problem of recommending collaboration prototypes to researchers seeking collaborators to do research. We proposed to address the complexity of the problem by focusing on the selection of prototypical collaborations that were successful in obtaining government funding. In this single-shot approach we explored, once a collaboration seeker is matched to a prototype, she will have a description of one or more target collaborators. This description can then be fed into an expert locator that identifies real instances of individuals who fit the description and can become collaborators.

For the step of recommending collaboration prototypes, we adopted a case-based approach. From a case base of collaboration prototypes, a new collaboration seeker is submitted and one or more useful collaboration prototypes are retrieved. For retrieval, we adopted as proxy for usefulness a measure that matches a collaboration seeker to the entire prototype. This matching relies on a concept that a seeker could be one of the collaborators in the prototype.

Although we have just scratched the surface of a very deep problem, our initial evaluation showed promise in this approach to the problem. We plan to validate usefulness of prototypes with expanded data. The use of richer characterization of the problem will provide recommendations to enhance the users' collaboration experience.

Acknowledgments

This work is supported in part by the U.S. EPA Science to Achieve Results (STAR) Program and the U.S. Department of Homeland Security Programs, Grant # R83236201.

References

Aamodt, A. and Plaza, E. 1994. "Case-Based Reasoning: Foundational Issues, Methodological Variations, and System Approaches," *AI Communications*, vol. 7, no. 1, pp. 39-59.

Agrawal, R. and Srikant, R. 1994. "Fast Algorithms for Mining Association Rules in Large Databases," *in proceedings of the 20th International Conference on Very Large Data Bases*, pp 487-499.

Aha, D. 1998. "The Omnipresence of Case-Based Reasoning in Science and Application," *Knowledge-Based Systems*, vol. 11, no.5-6, pp. 261-273.

Becerra-Fernandez, I. 2006. "Searching for Experts on the Web: A Review of Contemporary Expertise Locator Systems," *ACM Transactions on Internet Technology*, vol. 6, no. 4, pp. 333-355.

Bozeman, B. and Corley,E. 2004. "Scientists' collaboration strategies: implications for scientific and technical human capital," *Research Policy*, vol. 33,no. 4, pp. 599-616.

Bridge, D., Göker, M. H., McGinty, L., and Smyth, B. 2005 "Case-based recommender systems," *Knowledge Engineering Review*, vol. 20, no. 3, pp. 315-320.

Cotter, P. and Smyth, B. 2000. "PTV: Intelligent Personalized TV Guides," *in proceedings of the Seventeenth National Conference on Artificial intelligence and Twelfth Conference on innovative Applications of Artificial intelligence* (July 30 - August 03, 2000), AAAI Press / The MIT Press, pp. 957-964.

Crane, D. 1972. *Invisible Colleges*. University of Chicago Press.

Granovetter, M. S. 1973. "The Strength of Weak Ties," *American Journal of Sociology,* vol. 78, no. 6, pp. 1360-1380.

Hara, N., Solomon, P., Kim, S.-L., & Sonnenwald, D. H. 2003. An emerging view of scientific collaboration: Scientists' perspectives on collaboration and factors that impact collaboration. Journal of the American Society for Information Science and Technology, 54(10), 952-965.

Jeffrey, P. 2003. "Smoothing the Waters: Observations on the Process of Cross-Disciplinary Research Collaboration," *Social Studies of Science,* vol. 33*,* no. 4, pp. 539-562.

Katz, J. and Martin, B. 1997. "What is research collaboration?" *Research Policy*, vol. 26, pp 1-18, 1997.

Kreiner, K., & Schultz, M. 1993. Informal Collaboration in R&D. The formation of Networks Across Organizations. Organization Studies, 14(2), 189-209.

Maybury, M. T. 2002. "Knowledge on demand: Knowledge and expert discovery," *Journal of Universal Computer Science* vol. 8, no.5, pp. 491-505,.

Mattessich, P., Murray-Close, M.and Monsey, B. 2001. *Collaboration-what makes it work*, 2nd ed., Amherst H. Wilder Foundation: Saint Paul, Minnesota.

Miller, B., Albert, I., Lam, S., Konstan, J., and Riedl, J. 2003. "Movielens unplugged: Experiences with a recommender system on four mobile devices," *in proceedings of the 17th Annual Human-Computer Interaction Conference*, September 8–12 2003.

National Science Foundation, Strategic Plan FY 2006-2011. *Investing in America's Future* (NSF 06-48, September 2006. Retrieved November 14, 2007, from Government of United States, National Science Foundation Website: http://www.nsf.gov/publications/pub_summ.jsp?ods_key=nsf0648#.)

Omenn, G. (2006). Grand Challenges and Great Opportunities in Science, Technology, and Public Policy. Science, 314.

Smyth, B. and Cotter, P. 1999. "Surfing the digital wave: Generating personalized TV listings using collaborative, casebased recommendation," in Althoff, KD, Bergmann, R & Branting, LK (eds.), Proceedings of the 3rd International Conference on Case-Based Reasoning, Berlin: Springer, 561–571.

Wohlfarth, T. & Ichise, R. 2008. Semantic and Event-based Approach for Link Prediction, in Yamaguchi, Takahira (ed.), Proceedings of the 7th International Conference on Practical Aspects of Knowledge Management, Berlin: Springer.

Witten, I. and Frank, E. 2005. Data Mining: Practical machine learning tools and techniques, 2nd Ed, Morgan Kaufmann, San Francisco.