BLUEPRINTS FOR SUCCESS

Guidelines for Building Multidisciplinary Collaboration Teams

Sidath Gunawardena and Rosina O. Weber

The iSchool, Drexel University, Philadelphia, Pennsylvania, U.S.A.

Keywords: Recommender Systems, Case-Based Reasoning, Multidisciplinary Collaboration.

Abstract: Finding collaborators to engage in academic research is a challenging task, especially when the collaboration is multidisciplinary in nature and collaborators are needed from different disciplines. This paper uses evidence of successful multidisciplinary collaborations, funded proposals, in a novel way: as an input for a method of recommendation of multidisciplinary collaboration teams. We attempt to answer two questions posed by a collaboration seeker: what disciplines provide collaboration opportunities and what combinations of characteristics of collaborators have been successful in the past? We describe a two-step recommendation framework where the first step recommends potential disciplines with collaboration potential based on current trends in funding. The second step recommends characteristics for a collaboration team that are consistent with past instances of successful collaborations. We examine how this information source can be used in a case-based recommender system and present a preliminary validation of the system using statistical methods.

1 INTRODUCTION

Multidisciplinary collaboration brings together groups of researchers from different fields to solve a common problem, one that cannot be solved using the theories and methods of a single field (National Academies, 2005). US federal agencies encourage multidisciplinary research through increased funding initiatives (National Academies, 2005; National Science Foundation, 2006). Obtaining such funding is one way that academics, particularly tenure-track junior faculty, can advance their careers (Higgins and Walsh, 2009). Thus, academic researchers may need to find collaborators in fields very different from their own.

The traditional methods for finding a collaborator, such as leveraging one's professional ties, attending conferences, joining learned societies, and participating in on-line discussion groups (Clegg, 2003), by their nature, tend to focus inwards, towards one's own discipline (Kogan, 2000). Thus, such methods are much more likely to be successful when employed to find a collaborator in one's own discipline than when used to find a partner in a different discipline. Junior faculty members are at even greater disadvantage as they lack both experience and personal ties.

Currently available technological means provide little assistance in solving this problem. Technologies that leverage social networks to identify collaborators are limited to single disciplines (Ayanegui-Santiago et al., 2009; Liben-Nowell & Kleinberg, 2003; Newman, 2001). Expert locator systems focus on either finding an individual with pre-specified expertise or an expert able to answer to a pre-specified question (Serdyukov et al., 2008). They solve a very narrow problem of locating an expert to meet a pre-specified short term knowledge need. Hence, there is scope for a systematic, technological method for recommending synergistic disciplines and the desired characteristics of potential collaborators.

In order to find data that can help provide useful guidance, we look to existing successful multidisciplinary collaborations. In the context of competitive grant funding, we find repositories of experiences of successful multidisciplinary collaborations in the form of funded grant proposals. In order to make proper use of those experiences, we adopt a Case-Based Reasoning (CBR) methodology, a reasoning methodology that enables the reuse of experiences in multiple forms (Bride et al., 2005). While recommender systems are found in myriad contexts, we have yet to find any that attempt the task of recommending collaborators for

Gunawardena S. and O. Weber R..
BLUEPRINTS FOR SUCCESS - Guidelines for Building Multidisciplinary Collaboration Teams.
DOI: 10.5220/0003753503870393
In *Proceedings of the 4th International Conference on Agents and Artificial Intelligence* (ICAART-2012), pages 387-393
ISBN: 978-989-8425-95-9

multidisciplinary research.

In the next section we present some background literature, we then detail our data sources in Section 3. In Section 4 we present our methodology and in Section 5 our experiments and a discussion of our results. We close with our conclusions, and some thoughts about future work.

2 BACKGROUND

Recommending multidisciplinary collaborations has not been explored before, so the background of this work comes from recommending collaborators within the same discipline and also at the work on locating experts.

2.1 Social Networks

The links between researchers created by co-authorship, co-publication, or citation, can be leveraged to create social networks (Barabási et al, 2002; Tang et al., 2008), with co-authorship being the strongest link. In the case of co-authorship, the ‘distance’ between two authors is represented by the number of links that have to be traversed to make the connection between them. The number of co-authorships between two authors can be used as a measure of the strength of such linkages (Newman, 2001). Social networks can also be combined with other approaches as expert location systems to improve their usefulness to users by taking into account social dynamics in addition to expertise (McDonald, 2003). Work in social networking shows some promise for discovering collaborators who have the potential to work together, but the work is limited to researchers in the same field (Ayanegui-Santiago et al., 2009; Newman, 2003).

2.2 Expert Locator Systems

Collaborator recommendation is related to expert locator systems (ELS) (Becerra-Fernandez, 2003); where the system can recommend qualified experts to a user who has a need for a particular expertise. The level of expertise must be narrowly defined either as a question that needs an expert answer (Serdyukov, 2008) or limited to one organization (Maybury, 2002; McDonald 2003). When the user needs a particular type of expertise, the system selects the candidate that best matches the user’s expertise criteria. Additional factors such as availability can also be taken into account (McDonald & Ackerman, 2000).

When seeking a collaborator, the criteria to be satisfied are vague and ill-defined. We define researchers seeking to engage in multidisciplinary collaboration as *collaboration seekers*. The collaboration seeker likely does not know all the domains where suitable collaboration partners reside. Furthermore, factors additional to expertise need to be included. Hence, we perceive the potential usefulness of recommender systems.

We see collaboration recommendation and expert location as two separate parts of the process of finding a collaborator. The recommendation identifies the disciplines and the characteristics of the collaborators, and subsequently, expert location is used to identify the specific individuals who meet those characteristics.

2.3 Collaboration

A summary of some of the literature on collaboration can be found in Gunawardena et al. (2010). Collaboration is an idiosyncratic process, and when it occurs 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, when recommending collaboration teams, factors that can mitigate such problems need to be taken into account. Collaborators who are nearby and can facilitate face to face communications (Kat, 1994), senior colleagues can act as mediators (Bozeman & Corley, 2004; Wood & Gray, 1991), and collaborating with those at institutions with high research productivity can be beneficial (Jones et al., 2008). We examine data sources to find reasonable proxies for these factors. An initial experiment on this problem used funded grants but was limited to only area of expertise (Gunawardena & Weber 2009) showed that even with limited information it was possible to provide a basic recommendation. This work broadens the scope to include additional features of researchers known in the literature to have an impact on collaborative behaviors: the researchers’ location, their title, which is used as a proxy for their seniority, and the type of institution they belong. We take the most literal definition of multidisciplinary; in the collaborations we study are required to contain at least two members who have different departmental affiliations.

2.4 Case-based Reasoning

In CBR, the cases are typically composed of a problem context and a lesson that can be learned about it (Kolodner, 1993). The lesson can be thought of as the as the solution applicable to that particular problem context. In a case-based recommender system this takes the form of collection (case-base) of problems and associated solutions. A new problem is solved by reusing the solution of the most similar old problem (Bridge et al., 2005). We approach the problem of recommending collaborators by looking at what lessons we can learn from past successful collaborations.

In collaboration recommendation the problem to be solved is finding suitable collaboration partners for a faculty member, who is described by a set of characteristics (title, research area, institution, etc). The solution is described by the characteristics of the faculty with the best potential for collaborative success. Here the solution is presented by the same features that are used to describe the problem. Thus, the process of recommendation for a new collaboration seeker involves searching the case base for the collaboration with a member most similar to the collaboration seeker and then recommending the remaining collaborators in that collaboration, that is, the complementary portion of the collaboration, as the recommended collaboration team.

3 DATA SOURCES

We use funded grant proposals as experiences of successful multidisciplinary collaborations. The grant proposals contain the name and affiliated institution of the principal investigator and the names of the co-investigators. Thus, the information pertaining is incomplete with respect to what is required for solving the recommendation problem. To obtain a fuller picture of the collaborations we use additional sources of information.

3.1 Grant Data

For our experiments, we use grants funded by the Office of Multidisciplinary Activities (OMA), a directorate of the National Science Foundation (NSF), whose goal is to fund research in the mathematical and physical sciences that crosses disciplinary boundaries. We also utilize two additional sources to obtain the data required for these experiments. COS Scholar Universe, is a database of 2 million profiles of full time faculty supported by ProQuest LLC[1]. We obtain our data on researchers' departmental affiliations and titles from this source. Our third source of data is Academic Analytics LLC[2], a private company that provides the ranking of doctoral programs. We obtain our information on institution type and location from this source.

3.2 The Data Set

The dataset includes NSF grants from the period 2005-2010 that are composed of two to five members, with at least two members from different departments. The dataset contains 173 collaborations, involving 530 total faculty members from US academic institutions.

We aggregated the data, limiting the collaborations chosen to those comprised only of researchers with the titles of Assistant Professor, Associate Professor and Full Professors. Table 1 presents a summary of the data, and how it is coded. The departmental names have non-relevant terms removed to assign values to the feature Discipline (e.g. Department of Physics would be reduced to Physics).

Table 1: Summary of data.

Feature	Description
Title	Full, Associate, or Assistant Professor
Discipline	143 possible values (Chemistry, Astrophysics, Civil Engineering, …)
Institution Type	Large Research Inst, Small Research Inst, Specialized Inst.
Institution Location	Region (Northeast, Midwest, South, West)

We use the definition employed by Academic Analytics to categorize institutions by type. A university is considered a Large Research University (LRU) if it has at least fifteen PhD programs each with at least ten faculty members. A Small Research University (SRU) has between one and fourteen PhD programs. A Specialized University is one that awards a majority of their degrees in one field.

4 METHODOLOGY

In this section we describe the evolution of our research process, as we sequentially developed our method, with each step of the process informing the design of the subsequent experiments.

[1] www.proquest.com

[2] www.academicanalytics.com

4.1 Similarity Functions

We begin by explaining similarity in CBR and go onto describe the similarity functions we employ.

In CBR, the similarity function determines which cases in the case-base are selected, and thus which solutions are reused. The similarity function compares the characteristics of the new problem to the problems in the case-base and gives each case a score based on how similar it is to the new problem, with the higher scores assigned to the candidates to have their solutions reused.

Our initial analyses employed standard similarity methods: weighted and unweighted feature counting. We compared these to a baseline method of random recommendation and also to a modified random recommendation based on location. The purpose of the experiments is to demonstrate that the data does contain knowledge to make recommendations and then build on that to determine how to make more accurate recommendations.

4.1.1 Baseline Method: Random Recommendation

A collaborator is selected from the dataset and then n collaboration teams are randomly selected, with no team being selected twice, where *n* has the set of values {1, 3, 5, 10}.

4.1.2 Random Recommendation by Located Region

A collaborator is selected from the dataset and then randomly *n* collaboration teams are selected from the same region as the original collaborator, with no team being selected twice, *n* {1, 3, 5, 10}.

4.1.3 Feature Counting

As a first step, this method considers the selected features to have equal importance for similarity assessment. In a feature counting method, the similarity between the target artificial case *t* and candidate case *c* is given by Equation (1):

$$\text{Similarity} = \frac{1}{n} . \sum_{i=1}^{n} \text{Sim}(t_i, c_i) \qquad (1)$$

Where n is the number of features and $\text{Sim}(t_i,c_i) = 1$ if $t_i = c_i$, and 0 otherwise. Each collaboration has as many candidates as members. The similarity score used is the highest score obtained from all members. The remaining collaborators in that collaboration will be the team that is recommended.

4.1.4 Weighted

The weighted similarity method takes into consideration the relative importance of the features. Here the similarity between the target artificial case *t* and candidate case *c* is given by Equation (2):

$$\text{Similarity} = \frac{1}{n} . \sum_{i=1}^{n} w_i \text{Sim}(t_i, c_i) \qquad (2)$$

Where *n* is the number of features, w_i is the weight associated with feature *i*, and $\text{Sim}(t_i,c_i) = 1$ if $t_i = c_i$, and 0 otherwise.

To determine weights, we employ a genetic algorithm, a machine learning method used for optimization. It is based around the evolutionary principle of survival of the fittest, that is, in a population, the strongest genetic chromosomes survive and are passed on to future generations (Kelly & Davis, 1991). Genetic algorithms are a common method to derive weights for use in CBR systems (Beddoe & Petrovic, 2006; Dogan et al., 2006; Fu & Shen, 2004; Jarmulak et al., 2000). In this experiment, each characteristic of a collaborator (title, research interest, etc) is a chromosome. A genetic algorithm can be broken down into the following steps: initial weight generation, fitness evaluation, reproduction (including possible mutation). It also requires a predefined stopping criterion to terminate the process. For this experiment we apply a genetic algorithm with the following parameters: a crossover of 0.5 where each parent has an equal chance of providing the chromosome, a 1% chance of mutation where a gene is replaced by a new, random chromosome. The fitness function which determines which genes go to the next generation is determined based on accuracy at the top1 threshold. The algorithm will stop after 100 iterations. The execution of the genetic algorithm produced the following weights:

Table 2: Genetic algorithm derived weights.

Title	Discipline	Region	Inst. Type
0.24	0.34	0.34	0.08

4.2 Two Step Recommendation

There are two broad dimensions required to be considered when making this particular recommendation: a collaborator's research interest and their personal characteristics. The derived weights suggest that, combined, the personal characteristics (title, region, institution type) combined have a greater importance than that of research interest. This does not make intuitive sense as if a mathematician is seeking to engage in collaboration, then the previous collaborations of, for example, biologists have little value for *the purposes of identifying potential domains*. Thus, we take into account the practical aspects of a useful recommendation, similar to Baccigalupo & Plaza (2007) who in their work on song recommendation ignore songs that are irrelevant based on the user's specifications. Here the discipline is the primary determining factor, and the other factors secondary. To reflect this, in this experiment, we break the recommendation process into two steps.

Step 1: determine all the cases in the case-base that could provide useful recommendations. This is done by limiting the cases used to those that have at least one member from the same discipline or a discipline that is a sibling on a disciplinary taxonomy as the collaboration seeker. For our experiments we use the taxonomy used by the National Academies to classify doctoral programs[3].

Step 2: recommend the secondary characteristics of collaborators based on the characteristics of the collaborations seeker. We use the remaining features, title, location, and institution type to then recommend a potential team: the complementary portion of that collaboration.

The recommendation of the disciplines is decoupled from the recommendation of the characteristics of collaborators. Thus, with the two step approach the system is no longer limited to recommendations that exist as collaborations within the case-base. It can recommend the disciplines from one collaboration with the collaborator characteristics of another if it determines that that is the best recommendation for a particular collaboration seeker.

4.2.1 Feature Counting with Two-step

In the first step we limit the cases to those that have at least one member from the same discipline or a discipline that is a sibling on the disciplinary taxonomy as the collaboration seeker. Then we perform the feature counting similarity assessment as before, but only using title, location and institution type as features.

4.2.2 Weighted with Two-step

Here too we apply the two-step approach, using the first step to reduce the case-base and then run the GA to determine the weights of the remaining three features. We execute the GA using the same parameters as before. The execution of the genetic algorithm produced the following weights:

Table 3: Genetic algorithm derived weights.

Title	Region	Inst. Type
0.26	0.51	0.23

Thus we have the following hypotheses:

H1: Randomly selecting teams by region is more accurate than random selection.

H2: The feature counting method is more accurate than randomly selecting teams by region.

H3: The weighted method is more accurate than the feature counting method.

H4: The 2 step feature counting method is more accurate than the feature counting method.

H5: The 2 step weighted is more accurate than the weighted method.

H6: The 2 step weighted is more accurate than the 2 step feature counting method.

5 EXPERIMENTS

In this section we present the experiments we conducted on the grant dataset to demonstrate the effectiveness of this approach. These experiments are used to increase our understanding of the data, to allow us to determine whether it can be utilized to make useful recommendations.

5.1 Evaluation

A leave-one-out cross-validation (LOOCV) is a standard method to evaluate recommender systems. To apply LOOCV, a collaboration is removed from the collection and its members used as target cases. Accuracy is measured by whether the system retrieves the most similar case to the complementary portion of the removed case. However, we do not have the ability to determine similarity between collaborations to determine second best solution. To overcome this hurdle, we use what we term 'artificial collaboration seekers' who we can artificially create as being very similar to the original collaborators in the system. We describe this process in the following section.

[3] http://www.nationalacademies.org/

5.2 Generating Artificial Collaboration Seekers

From a collaboration we select each collaborator in turn and randomly select one of the features (discipline, title, institution type, or location) and modify it. The modification is such that when a feature value is modified, it is changed to an adjacent value, that is, a collaborator's title may change from assistant to associate professor, but not to full, where as an associate professor may be changed to either a full professor or an assistant professor. If the feature to be modified is discipline, then we use the taxonomy and modify the discipline and replace it with one that is a sibling.

5.3 Accuracy

In our experiments we measure accuracy as follows: when an artificial collaboration seeker is submitted to the system as a new target problem the retrieval set contains the complementary members of the original collaboration that generated the artificial collaboration seeker is retrieved within the top n cases. We examine results for the top n cases, considering n = {1, 3, 5, 10}. Tied values are considered to be equivalent in rank when determining whether a particular retrieval was successful or not. An artificial collaboration seeker is created for every collaborator in the dataset and accuracy is measured by whether collaboration team of the original collaborator is one of the top *n* recommended teams, *n* = {1, 3, 5, 10}. Each experiment is repeated ten times with an average accuracy calculated.

A one-way ANOVA test is used to determine if there is a significant difference between the means of the various methods ($\alpha = 0.05$), post hoc analyses of Scheffe, Tukey's Honest Significant Differences, Bonferroni Adjustment, and Least Significant Differences are then used to perform multiple comparisons between the means. A difference is reported as significant only if all four tests concur. The random methods are outperformed at all levels of accuracy, but the other methods only show a significant difference only when the top result is considered.

5.4 Results and Discussion

Based on the post hoc analysis at the 0.05 confidence level, we are able to reject the null hypothesis that there is no difference between the random methods and the feature counting and weighted methods at all levels of accuracy. In addition, at the top level of accuracy, the weighted methods outperform their feature counting counterparts and the two-step method shows an improvement in accuracy in both weighted and feature counting methods (Table 4). No significant difference was observed between these 4 methods at other levels of accuracy.

Table 4: Average accuracy (standard deviation) top1 results.

Feature Counting	0.492 (0.012)
2Step Feature Counting	0.521 (0.012)
Weighted	0.526 (0.017)
2Step Weighted	0.551 (0.011)

Our results versus a random baseline show that this data does possess knowledge and can be used as the basis for the recommendation of multidisciplinary collaboration teams. The subsequent results are mixed, showing statistically significant improvement only at the top level of accuracy. This is less improvement than expected of the two step method. However, the two step method recommends the best potential collaboration, which may not be one that exists in the case-base, penalizing its accuracy.

6 CONCLUSIONS AND FUTURE WORK

In this paper we show how funded grants may be used as a basis for solving a novel problem: recommending multidisciplinary collaboration teams. Using the grant dataset, we demonstrated that the proposed approach can provide recommendations that are superior to random, and showed further improvements to increase their quality. These results suggest this is a viable approach to using this data on this problem. This approach has room for improvement but it is unique in its use of the data and in providing a solution to this problem. Out of many possible improvements, we name a few. Instead of *discipline* the use of publication keywords can provide a more detailed recommendation. Additionally, these experiments focus solely on analogical reasoning, incorporating analytical knowledge from the literature on collaboration may add to the quality of the recommendation.

REFERENCES

Aha, D. (1997) Editorial on Lazy Learning. *Artificial*

Intelligence Review, 11, 7-10

Ayanegui-Santiago, H., Reyes-Galaviz, O., Chávez-Aragón, A., Ramírez-Cruz, F., Portilla, A., & García-Bañuelos, L. (2009). Mining Social Networks on the Mexican Computer Science Community. *In MICAI 2009: Advances in Artificial Intelligence* (pp. 213-224).

Barabási, A. L., Jeong, H., Néda, Z., Ravasz, E., Schubert, A., & Vicsek, T. (2002). Evolution of the social network of scientific collaborations. *Physica A: Statistical Mechanics and its Applications*, 311(3-4), 590-614.

Becerra-Fernandez, I. (2006). Searching for Experts on the Web: A Review of Contemporary Expertise Locator Systems, *ACM Transactions on Internet Technology*, 6(4), 333-355.

Beddoe, G. R., & Petrovic, S. (2006). Selecting and weighting features using a genetic algorithm in a case-based reasoning approach to personnel rostering. *European Journal of Operational Research*, 175(2), 649-671.

Baccigalupo, C. and Plaza, E. (2007), A Case-Based Song Scheduler for Group Customised Radio. Proc. ICCBR 2007, *Lecture Notes in Computer Science* Vol. 4626, p. 433-448. Springer Verlag.

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., Goker, M. H., McGinty, L, Smyth, B. (2005) On Case-Based Recommender Systems. *Knowledge Engineering Review*, 20 (3):315-320

Clegg, S. 2003. Problematising ourselves: Continuing professional development in higher education. *International Journal for Academic Development*, 8, 37-50.

Dogan, S. Z., Arditi, D., Gunayd, X, & N, H. M. (2006). Determining Attribute Weights in a CBR Model for Early Cost Prediction of Structural Systems. *Journal of Construction Engineering and Management*, 132(10), 1092-1098.

Fu, Y., & Shen, R. (2004). GA based CBR approach in Q&A system. *Expert Systems with Applications*, 26(2), 167-170.

Gunawardena, S., & Weber, R. (2009). Discovering Patterns of Collaboration for Recommendation. *In Proceedings 22th International FLAIRS Conference*, FLA IRS'09. AAAI Press, Menlo Park, California, 2009.

Gunawardena, S., Weber, R., & Agosto, D. E. (2010). Finding that Special Someone: Modeling Collabo-ration in an Academic Context. *Journal of Education for Library and Information Science*, 51.

Higgins, S. E., & Welsh, T. S. (2009). The Tenure Process in LIS: A Survey of LIS/IS Program Directors. *Journal of Education for Library and Information Science*, 50(3), 176-189.

Jarmulak, J., Craw, S., & Rowe, R. (2000). Genetic Algorithms to Optimise CBR Retrieval. *In Advances in Case-Based Reasoning* (Vol. 1898, pp. 159-194): Springer Berlin / Heidelberg.

Jeffrey, P. (2003). Smoothing the Waters: Observations on the Process of Cross-Disciplinary Research Collaboration. *Social Studies of Science*, 33(4), 539-562.

Jones, B. F., Wuchty, S., & Uzzi, B. (2008). Multi-University Research Teams: Shifting Impact, Geography, and Stratification in Science. *Science*, 322(5905), 1259-1262.

Katz, J. (1994). Geographical proximity and scientific collaboration. *Scientometrics*, 31(1), 31-43.

Kelly, J. and Davis, L. (1991). A hybrid genetic algorithm for classification. *In Proceedings of the 12th IJCAI*, Sidney, Australia, 645–650.

Kogan, M. (2000). Higher Education Communities and Academic Identity. *Higher Education Quarterly*, 54, 207-216.

Kolodner, J. (1993). *Case-Based Reasoning*. Morgan Kaufmann, San Francisco

Liben-Nowell, D., & Kleinberg, J. (2003). The link prediction problem for social networks. *Paper presented at the Proceedings of the twelfth international conference on Information and knowledge management*.

McDonald, D. W., & Ackerman, M. S. (2000). Expertise recommender: a flexible recommendation system and architecture. *Paper presented at the Proceedings of the 2000 ACM conference on Computer supported cooperative work*.

Maybury, M. T. 2002. Knowledge on demand: Knowledge and expert discovery, *Journal of Universal Computer Science* (8)5, pp. 491-505.

McDonald, D. (2003). Recommending collaboration with social networks: a comparative evaluation. *Paper presented at the CHI '03: Proceedings of the SIGCHI conference on Human factors in computing systems*, Ft. Lauderdale, Florida, USA.

National Academy of Sciences, National Academy of Engineering, Institute of Medicine. (2005). Facilitating Interdisciplinary Research. Washington, *DC.: National Academies Press*.

National Science Foundation, Strategic Plan FY 2006-2011. Investing in America's Future (NSF 06-48), September 2006.

National Academy of Sciences, National Academy of Engineering, Institute of Medicine. (2005). Facilitating Interdisciplinary Research. Washington, *DC.: National Academies Press.*

Newman, M. E. J. (2001). From the Cover: The structure of scientific collaboration networks (Publication no. 10.1073/pnas.021544898).

Serdyukov, P., Feng, L., van Bunningen, A., Evers, S., van Heerde, H., Apers, P., Fokkinga, M., and Hiemstra, D. (2008). *The Right Expert at the Right Time and Place*. PAKM: 38-49.

Tang, J., Zhang, J., Yao, L., Li, J., Zhang, L., & Su, Z. (2008). ArnetMiner: Extraction and Mining of Academic Social Networks. *Paper presented at the Fourteenth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (SIGKDD'2009)*, Las Vegas, Nevada, USA.

Wood, D., & Gray, B. (1991). Towards a Comprehensive Theory of Collaboration. *Journal of Applied Behavioral Science*, 27(2), 139-162.