Expert Recommendation based on Social Drivers, Social Network Analysis, and Semantic Data Representation

Maryam Fazel-Zarandi
University of Toronto
10 King's College Road
Toronto, Canada M5S 3G4
1-416-946-0398
mfazel@cs.toronto.edu

Hugh J. Devlin
Northwestern University
2145 Sheridan Rd
Evanston IL USA 60208
1-847-491-3669
hdevlin@northwestern.edu

Yun Huang
Northwestern University
2145 Sheridan Rd
Evanston IL USA 60208
1-847-491-3669
yun@northwestern.edu

Noshir Contractor
Northwestern University
2145 Sheridan Rd
Evanston IL USA 60208
1-847-491-3669
nosh@northwestern.edu

ABSTRACT

Knowledge networks and recommender systems are especially important for expert finding within organizations and scientific communities. Useful recommendation of experts, however, is not an easy task for many reasons: It requires reasoning about multiple complex networks from heterogeneous sources (such as collaboration networks of individuals, article citation networks, and concept networks) and depends significantly on the needs of individuals in seeking recommendations.

Although over the past decade much effort has gone into developing techniques to increase and evaluate the quality of recommendations, personalizing recommendations according to individuals' *motivations* has not received much attention. While previous work in the literature has focused primarily on *identifying* experts, our focus here is on personalizing the *selection* of an expert through a principled application of social science theories to model the user's motivation. In this paper, we present an expert recommender system capable of applying multiple theoretical mechanisms to the problem of personalized recommendations through profiling users' motivations and their relations. To this end, we use the Multi-Theoretical Multi-Level (MTML) framework which investigates social drivers for network formation in the communities with diverse goals. This framework serves as the theoretical basis for mapping motivations to the appropriate domain data, heuristic, and objective functions for the personalized expert recommendation.

As a proof of concept, we developed a prototype recommender grounded in social science theories, and utilizing computational techniques from social network analysis and representational techniques from the semantic web to facilitate combining and operating on data from heterogeneous sources. We evaluated the prototype's ability to predict collaborations for scientific research teams, using a simple off-line methodology. Preliminary results demonstrate encouraging success while offering significant personalization options and providing flexibility in customizing the recommendation heuristic based on users' motivations. In particular, recommendation heuristics based on different motivation profiles result in different recommendations, and taken as a whole better capture the diversity of observed expert collaboration.

HetRec'11, October 27 2011, Chicago, IL, USA

Categories and Subject Descriptors

H.3.3 [**Information Search and Retrieval**]: Selection process; H.1.2 [**User/Machine Systems**]: Human factors; H.5.3 [**Group and Organization Interfaces**]: Computer-supported cooperative work.

General Terms

Algorithms, Design, Human Factors, Theory.

Keywords

Scientific collaboration, expert recommendation, personalized search, social network analysis, semantic web, social drivers, multi-theoretical multi-level framework.

1. INTRODUCTION

Finding individuals with appropriate expertise is important for accomplishing knowledge intensive tasks and solving complex problems since individuals usually rely heavily on others with knowledge and information [11]. More specifically, within academic communities, to facilitate academic research and encourage interdisciplinary collaboration, it is even more crucial that individuals can identify the right expertise or resources from within a large multidimensional knowledge network and in turn interact with potential collaborators. However, finding a collaborator might not be an easy task for many reasons. Expertise is highly dynamic, difficult to qualify, and varying in level [16, 30]. Validating other people's expertise is difficult [30]. The assistance of multiple experts may be required for complex problems [16]. The difficulty of locating an expert is exacerbated in larger and more geographically distributed virtual organizations and communities such as those found on the Web.

For these reasons, the study and development of tools that recommend people who have expertise in a given area has received the attention of researchers and organizations. The resulting expert recommender systems either rely on individuals to provide accurate and comprehensive profiles of their competences and experiences [16], or use mechanisms to automatically discover up-to-date expertise information from secondary sources such as articles, email communications, and forums [45]. Given that the ultimate goal of most expert recommender systems is to emulate what is essentially a *social* process [40], in recent years there has been an increasing interest in using social networks and social network analysis (SNA) to improve the quality of recommendations. In the context of expert finding in particular, SNA offers methods for identifying important persons within social networks [22] and those individuals who are more likely to help an expert seeker.

Although over the past decade much effort has gone into developing techniques to increase and evaluate the quality of recommendations, personalizing recommendations according to individuals' *motivations* has not received much attention. Useful recommendation of experts and resources, however, depends significantly on the motivation of the user in seeking the recommendation [13]. For example, looking for a quick answer versus a potential research collaborator will require very different recommendation strategies. The effectiveness of existing expert recommenders, however, is limited in that their recommendations are either based on only one theoretical mechanism, namely, similarity between user preferences [20], or solely on network statistics without consideration of the underlying user motivations. Another important shortcoming of current expert recommender systems is related to their limitation with respect to the sources of information [39, 41] and their difficulties in extending to open data sources.

In this paper, we aim to build flexible expert recommender systems by integrating two new elements: applying individual motivations for team assembly in the recommendation process and using representational techniques from the semantic web to facilitate combining data from heterogeneous sources. We personalize the recommendation process by requiring users of the system to specify their motivations, which are used as a template for selecting the appropriate network data to draw into the recommendation heuristic. To this end, we adopt the Multi-Theoretical Multi-Level (MTML) framework [14, 35] which applies various findings in social theories for network formation in communities and constructs a catalog of diverse individual motivations such as discovering new people (novelty), collaborating with persons of similar characteristics (homophily), exploiting existing resources, and responding rapidly. Two contributions of our work are: 1) individuals can choose a recommendation profile and get personalized recommendations according to their motivations and positions in social networks; and 2) recommenders can be used as tools to investigate motivation of effective collaborations in a community. It is important to note that we do not intend to propose a novel recommendation algorithm, but rather a framework which takes into account information about users and their motivations. In other words, we suggest a better way to leverage social science insights about how and why people come together in successful teams and propose a more *prescriptive* approach to complement the existing predictive approaches for making recommendations.

The remainder of this paper is organized as follows: A brief discussion of related work is followed by a presentation of social drivers and the conceptual framework for a recommender system utilizing network analysis and semantic data representation. A demonstration prototype is described and preliminary results are reported for the testing of the system. Finally, the paper is concluded with a discussion of potential areas of future work.

2. RELATED WORK

Expert recommenders are a subset of recommender systems where experts are the "items" being recommended [19]. Comprehensive surveys of this literature are available [1, 9, 38]. We focus instead on aspects of the literature that particularly motivate and inform our model. In particular, when searching for experts, two steps can be distinguished: 1) *Expert identification*: the problem of knowing what information or skills other individuals have, 2) *Expert selection*: the problem of appropriately choosing among qualified experts [20, 32].

2.1 Expert Identification

Expert recommender systems use implicit or explicit data about individuals' expertise to identify appropriate experts. Early applications, which were also known as "corporate yellow pages" or "people finders," relied on individuals to provide accurate and comprehensive profiles of their own expertise [16, 19]. With more activities occurring in the digital environment, automated expertise extraction from sources such as user created artifacts, online activities, and reputation and authority of experts [41] has gained momentum. Examples of information retrieval techniques used to automatically identify experts include using probabilistic language models [2, 10] and considering relevance scores of documents related to a person [28]. Identifying experts solely based on social networks has also been considered in previous research [46, 47]. The idea behind these techniques is that an individual's expertise can be inferred from the expertise of other people to whom he or she is connected [27].

To match individuals to queries, both similarity-based and logic-based matchmaking approaches exist in the literature. For example, information retrieval techniques such as database querying and similarity between weighted vectors of terms [43] have been considered in previous research. Techniques for ontology-based matching have also been considered. For example, Colucci et al. [12] use description logic inferences in an ontology-supported framework for finding the best individual for a given task. In general, matchmaking strategies based on purely logic deductive facilities present high precision and recall, but are often characterized by low flexibility [4]. Flexibility refers to the ability to recognize the degree of similarity when an exact match does not exist. On the other hand, similarity-based approaches are characterized by high flexibility, but limited precision and recall [4]. To take advantage of the benefits of both approaches, hybrid strategies have also been proposed [17, 33, 42].

2.2 Expert Selection

Expert selection is in part the problem of personalizing search results. Although over the past decade or so much effort has gone into creating techniques to increase and evaluate the recommendation quality for objects such as books and movies, the personalized search for subjects such as experts in a particular field has not so far received much attention [19, 29, 36]. This is in part due to the difficulty in measuring some of the related aspects. For example, Borgatti and Cross [7] and Casciaro and Lobo [11] reveal that in addition to the expert seeker's awareness of a potential source's expertise, other factors such as timely access to the source, a degree of safety in the relationship, and willingness of an expert to cognitively engage in problem solving, all play an important part in determining whom the seeker chooses to go to. As such, social factors and contextual knowledge need to be considered when recommending experts.

Previous studies that have attempted to provide personalized search for subjects can be divided into two groups. Techniques in the first approach consider user preferences and make recommendations based on the degree of similarity between expert profiles and user preferences. Systems in the second approach, on the other hand, consider social relations and network statistics. Examples include Referral Web [21] which uses co-

authoring and co-citation relationships, Expert Recommender [32] which considers friendship and departmental relations, and the dating recommender [36] which considers network centrality when selecting individuals.

3. SOCIAL DRIVERS

Comprehending available information about the individual is recognized as a crucial aspect of generating useful, satisfying recommendations [26]. Typically user profiles may include basic demographic information or past preferences of the user [37]. But the single most important data item about the seeker of a recommendation, in terms of its significance in generating a useful recommendation, is perhaps a characterization of the user's *motivation* in seeking recommendations. We focus on this aspect for recommending experts.

Recommender implementation efforts attempting to assure that their search heuristics satisfy user expectations usually include extensive preliminary field studies including user interviews [31, 37] and, most typically, after-the-fact user satisfaction surveys. But understanding the diverse *social drivers* that influence the formation of ties in communities is one of the founding questions of social science. Recommending collaborations is proposing social ties, yet many published studies fail to acknowledge or explicitly identify the social theories and generative mechanisms that motivate their research [34]. Our work suggests how the major findings drawn from the social science literature can inform systems that recommend social relationships including scientific research collaborations.

The Multitheoretical, Multilevel (MTML) analytic framework provides a coherent, overarching framework for integrating conceptual, theoretical, and empirical work. A multitheoretical perspective helps to compare and integrate diverse theories and to increase the explanatory power of research results. MTML comprehends both individual and social relational attributes and the evolving nature of social networks over time. MTML takes a complex adaptive systems perspective on social network formation. Complex systems exhibit distinct characteristics such as multi-scale interactions, emergent behavior, pattern formation, and self-organization, and are not amenable to reductionism.

Table 1. Recommendation heuristic driven by motivation

Colloquial Term	MTML Theory	Network of Relations	Importance (centrality measure)	Node Attributes
Most qualified	Self Interest	Agent-artifact relations (e.g., citation)	In-degree within the network of concern	Measures of qualification (e.g., # of publications)
Friend-of-a-Friend	Balance	Agent-agent relations (e.g., collaborations)	Distance (e.g., # of geodesics	
Birds of a feather	Homophily			Measures of similarity (e.g., gender)
Social Exchange	Social Exchange	Agent-agent relations		Measures of tie strength
Follow the crowd	Contagion	Agent-agent relations (e.g., collaborations, citations)	In-degree within the network of concern	

Social scientists studying complex social systems who eschew the temptation to reductionism may instead choose to embrace the complexity of their domain though adopting a research framework which approaches complex social phenomenon from a multi-level, multi-theoretical perspective. Alternative social theories make differential predictions about the formation of social relationships, and the key prediction of MTML is that utilizing multiple social theories should improve our understanding of network evolution as well as *significantly increase the amount of variance accounted for* by these theoretical mechanisms. Applying the MTML framework to modeling a social phenomenon obligates researchers to consider the various social theories comprehended by the framework and encourages researchers to generate diverse hypotheses for further empirical studies or simulations [35].

3.1 Motivation Templates

The MTML-based recommender proposed in this paper is a hybrid recommender system [9], a cascade of two phases . The first phase, a knowledge-based component, aggregates data from different sources and identifies qualified experts using domain knowledge (Section 4). The second phase, a switching component, recommends the most suitable expert by choosing among different recommendation strategies based on the expert seekers' motivation. Our system implements diverse approaches to expert recommendation as user-selectable *motivation templates*. Each template implements an expert recommendation strategy as might flow from a particular, widely-accepted theory about the formation of social relationships by considering a set of relations, measures of importance, and/or individual attributes. Table 1 summarizes the motivation templates, their underlying social science theories, and the relations and attributes involved in the recommendation heuristic.

This section illustrates our method by briefly describing the underlying social theories, the recommendation heuristics, and the objective functions *f* of some of our motivation templates. We illustrate the operation of the templates with examples drawn from expert recommending in the domain of scientific collaboration. The motivation templates operate in the second, switching stage of our hybrid cascade recommender, so in each example given, a list of qualified experts has already been identified.

3.1.1 Self-interest theory: Most qualified template

Theories of self-interest are among the oldest models of human behavior, from the "rationality" of Adam Smith to the "bounded rationality" of Simon [35]. These theories postulate that people make what they believe to be a rational choice in order to acquire personal benefits, in essence conducting a personal cost-benefit analysis. From a purely self-interest point of view, the most useful collaborator is the most qualified expert. As such, expert recommendation grounded in the self-interest theory focuses on expertise measures which may be derived from individual attributes and/or position of the individual within a particular network. Individual measures of expertise include number of publications on a given topic and years of experience. Expertise networks include citation and question-answering networks, and network measures for inferring expertise include calculating in-degree and using popular webpage ranking algorithms such as PageRank [8] or HITS [23].

In the context of finding a collaborator within an academic community, for example, expertise can be estimated from historical bibliography data [28]. In this case, the "most qualified" template would take the subset of the bibliographic data qualified

by the topic of interest, and consider number of publications. An author-author citation network is then constructed based on the restricted bibliographic data. The score *f* for a qualified expert *e* on a vector of topic keywords *k* can then be defined as the number of publications *p* on the required topics by that expert, plus *c*, the *in-degree* of *e* in the author-author citation network (the number of distinct authors who have publications with those keywords which cite that expert):

$$f(e,\vec{k}) = p(e,\vec{k}) + c(e,\vec{k})$$

3.1.2 Balance theory: Friend-of-a-friend template

Early work on balance theory profoundly influenced modern social network analysis. Balance theory suggests that the likelihood of a social tie between two persons is predicated by the configuration of ties the two persons have with third individuals [35]. Our "friend-of-friend" motivation template is targeted to meet the needs of individuals seeking recommendations of experts with whom they have not previously worked. In the context of scientific collaboration, researchers may seek new collaborators when their research is stymied and in need of new methods or perspectives, or perhaps grant proposals from their usual cohort have been repeatedly rejected. Previous collaborators are specifically disallowed in the "friend-of-friend" template.

$$f(u,e) = \frac{\sqrt{g(u,e)}}{(d(u,e))^2}, \quad d > 1, \quad 0 \ \ otherwise$$

The score *f* for a qualified expert *e* is inversely proportional to the square of the distance *d* between the user *u* and the expert *e*, but also proportional to the square root of the number of distinct geodesics (shortest paths) *g* from the user *u* to the expert *e*. That is, for experts at the same distance in the collaboration network, the expert with more common intermediaries who might facilitate an introduction scores higher.

3.1.3 Homophily theory: Birds-of-a-feather template

Individuals define their social identity through self-categorization and similarity provides individuals with a basis for legitimizing their own social identity [35]. Homophily theories model the selection of others who are similar. Our "birds-of-a-feather" motivation template empowers a user to seek experts who are alike (or dissimilar) on a user-selectable menu of *k* attributes:

$$f(u,e) = \sum_k w_k \cdot a_k(u,e)$$

Attributes relevant to collaboration in an academic community include gender, organizational affiliation, tenure status, and co-citation similarity (that is, two researchers are considered similar if they have been cited together). Weights w_k indicate the relative importance of one attribute over another to the user.

3.1.4 Social exchange template

Social exchange theory seeks to explain human action by a calculus of exchange of material or information resources., and attempts to explain the likelihood of a dyadic relationship based on the supply and demand of resources that each member of the dyadic had to offer [6]. Our "exchange" template identifies discrepancies in resource allocation which may provide a basis for the realization of a social tie:

$$f(u,e) = c(u,e)$$

Each expert's sense of indebtedness to the user is estimated, for example, by the number of times *c* that expert *e* cited the user *u* in the citation network.

3.1.5 Contagion theory: Follow-the-crowd template

Contagion theories seek to explain networks as conduits for "infectious" attitudes and behaviors. Krackhardt and Brass [24] suggest practical applications of contagion theory requires that the *principle of interaction* be counterbalanced with the *principle of reflected exclusivity*, lest naïve applications of contagion theory to for example collaboration behavior entail that researchers with the most previous collaborations would always be the best choice for future collaborations. Krackhardt [25] develops the concept of *cognitive social structures* to characterize individuals perceptions of social networks, which may differ significantly from their actual, observed network. Our "follow-the-crowd" template implements a popularity measure, counterbalanced with a factor which decays the popularity with increasing distance, modeling the decrease in accuracy of the perception of the network with increasing distance:

$$f(u,e) = \frac{inDegree(e)}{d(u,e)}$$

Here, *inDegree* is an estimator of the "popularity" of expert *e*, the number of in-bound arcs in the overall network of direct and derived interactions, while distance *d* is an estimator of social distance, measured in the network of direct personal interactions.

4. EXPERT PROFILING

Many different sources of information can be used for constructing expert profiles. Skill statements can be declared by individuals about themselves or by others, and/or can be derived from 1) activities performed by the individual either online or offline including enrollment in learning activities, experiences related to the workplace, and question-answering in online forums; 2) content created by the individual such as scientific publications, Wikipedia pages, and blog posts; and 3) recommendations and the "wisdom of the crowd". We use these sources to construct employee profiles. In recent years in particular, digital media and communication networks have become an important medium for enabling new levels of interactions in organizations and communities. Many online communities and interactive collaboration spaces (such as forums and wikis) evolve into large-scale knowledge networks which are context-dependent and multi-dimensional. These networks consist of 1) knowledge entities such as keywords and research topics, 2) information entities such as articles, projects, organizations, and emails, and 3) individuals. As the structure of these networks suggests, they are ideal for creating expert profiles to be used to select experts according to the requester's perspectives and connections in the social network.

However, in order to accomplish this task, data from heterogeneous sources, some of which were not created for the purposes of expert identification or network analysis, need to be aggregated. In fact, an important shortcoming of current expert recommenders is related to their limitation with respect to the sources of information that they use [19, 39, 41] and to improve these systems, it is important to integrate data from different sources of information to better reflect individual's expertise [19]. To this end, Semantic Web technologies and ontologies provide a solution for representing, integrating, and reasoning with semantic and social data from divergent sources of information available on the Web [3, 18]. Furthermore, the emerging mass of publicly available interconnected data published on the Web according to Linked Data principles [5] and using unambiguous vocabularies, make automated data integration from diverse and distributed

sources possible for better reflecting individuals' expertise [41]. In particular, the Resource Description Framework (RDF)[1] provides a standard representation scheme for capturing domain knowledge, well-suited to federating data from diverse sources into a coherent knowledge base. RDF is a flexible knowledge representation scheme adapted at expressing network structures, and so is an ideal foundation for systems that integrate network analysis and inference technologies. In addition to providing a shared vocabulary and facilitating automatic information integration, ontologies provide the means for deduction and automated reasoning in order to generate further knowledge [18] (i.e., knowledge that is not explicitly known but that can be deducted based on the general knowledge of the domain). For example, given a taxonomy of skills in a particular domain, it is possible to infer expertise in areas which are not explicitly declared. Furthermore, by adding declarative rules, it is possible to infer additional associations or social relationships. For example, rules can be added to link individuals by co-authorship or trust relations, to infer expertise based on different evidences, and/or to determine who should seek knowledge from whom.

The heterogeneous data from divergent sources is expressed in a common representation, namely RDF. The RDF graphs can be typed using existing ontologies together with proprietary domain ontologies if needed. For example, FOAF[2] and Dublin Core[3] vocabularies can be used to represent social networks and publication metadata. Once the data is expressed in a common representation, concepts and instances of the used ontologies are mapped in order to bridge the semantic gap. We use the three-layer multi-network model presented in [20] consisting of content, artifact, and human layers, to bridge the gap by defining three types of nodes: agent, artifact, and concept. These node types can be extended as necessary by adding attributes and relations. In this model, the concept layer represents the *content* domain and consists of all knowledge entities and their semantic networks and the taxonomies in which they are organized. The *human* layer is a network of people and experts in the community. The intermediate *artifact* layer represents all information entities that can bridge the concept layer and the human layer and establish relations among them. This three-layer model is very flexible in preserving various types of relations: semantic networks in the concept layer, social networks in the human layer, and association relations in the artifact layer and between layers. The combination of different relations can be used to infer further knowledge and to better understand user behavior. This model provides a simple underlying representation for evaluating different expert identification and selection strategies.

5. EVALUATION

The intention of the motivation templates is to *personalize* search by capturing the user's *goal* in seeking a recommendation. Recognizing that the problem of *recommending* social relationships is closely related to the problem of *predicting* social relationships, we next present an evaluation of the face validity of our core idea of MTML-based recommending. We do not consider our evaluation stand-alone as a novel contribution to the considerable literature on the single best model for the motivation of the seeker of a recommendation (e.g., self-interest versus homophily). While our multi-theoretical approach to recommendation naturally invites comparisons of the various motivation theories, our purpose here is to see whether our implementation produces recommendations consistent with how previous research suggests people are choosing collaborators without assistance, and how it has the potential to identify successful collaborations. As such, we do not intend to support generalized conclusions regarding the single best motivation. Our aim is to show that the motivation of the expertise seeker is important and should be baked into recommendation strategies.

5.1 Experimental Setting

Two datasets provided data for the current study: data drawn from grant applications in the year 2009 for funding from the Northwestern University Clinical and Translational Sciences (NUCATS) Institute[4], and the publication histories of the applicants extracted from the Thomson Reuters Web of Science bibliographic database[5]. The program attracted 147 principal investigators and co-principal investigators from across the university. 98 grant proposals were submitted in 2009, of which 52 were collaborative (two 5-way collaborations, three 4-way, 16 3-way, and the rest were two-way). Two applicants collaborated on three proposals, 27 applicants collaborated on two proposals, while most applicants (129) were associated with just one proposal. The bibliographic data includes 83,213 articles, of which 15,592 are authored by at least one NUCATS applicant as an author, and the rest of which are articles cited by NUCATS authors. 38,656 distinct keywords assigned (by the author's themselves or by Thomson Reuters "Keyword Plus") to the NUCATS participants' articles were extracted.

To provide simulated topics of collaboration for our test cases, the textual descriptions of the proposed projects from the grant applications were analyzed using the Centering Resonance Analysis[6] (CRA) method [15]. Simplifying, CRA extracts a network of noun phrases from text and computes a betweenness centrality measure on each concept. The keywords extracted from the grant applications by CRA were validated by cross-referencing them with the Web of Science keywords associated with articles authored by NUCATS applicants. Keywords associated with 40 or more of the applicants, such as "cell" and "gene," were considered to be too general in discriminating between applicants, while keywords associated with fewer than 10 applicants were considered to be too specific in serving as simulated topics of collaboration, and so were not used. The remaining middle-frequency keywords served as concepts on which the applicants might seek expertise.

5.2 Evaluation Criteria

We evaluated the proposed recommender by an "off-line" methodology: "holding back" our knowledge of the actual grant proposal collaborations, and simulating the NUCATS applicants seeking each other. We use the following criteria to evaluate the recommendations made by our motivation templates:

- P@1 (P@5): The proportion of test cases for which the actual collaborator is ranked by a template in the top 1 (5) in the list of identified experts;
- ARC: The average rank of the correct recommendation (actual collaborator) [9];

[1] http://www.w3.org/RDF/
[2] http://xmlns.com/foaf/spec/
[3] http://www.dublincore.org
[4] http://www.nucats.northwestern.edu/
[5] http://apps.isiknowledge.com
[6] http://www.crawdadtech.com/

Each motivation template makes its recommendations for each test case, and the rank of an actual collaborator in the ranked list of recommended experts is recorded. For the purposes of this study, a test case consists of an ordered pair of collaborators and a concept keyword. The first individual is considered to be seeking collaboration on the specified concept keyword, and the second is considered to be the expected result (i.e., the actual collaborator). Moreover, we consider Researcher-A seeking Researcher-B to be a distinct test case from Researcher-B seeking Researcher-A. An individual is identified as an expert on one or more topics if and only if there exists a *knows-about* relation between the individual and each associated keyword. *knows-about* relations are inferred by a reasoning engine from explicit author-article and article-keyword relationships extracted from the bibliographic data. Note that for this evaluation, we mainly focus on the problem of expert selection according to requester's perspectives and connections in the social network, and do not consider other techniques for identification and expert ranking, such as topic-centric similarity measures or traversing a taxonomy of concepts.

5.3 Results

Table 2 reports results from testing the recommender against 125 test cases, running each of the motivation templates. Since in many scenarios the point of a recommender system is to help individuals find experts whom they would not have found otherwise, results for 69 cases in which the seeker and the actual collaborator had NOT collaborated (co-authored) previously are also reported separately. On average, 22 experts were identified for each case. Exchange and friend-of-a-friend templates have a limited coverage in that they were not able to make any recommendations for some test cases due to the sparsity of the underlying collaboration network (the expert at rank one had score zero). As such, ARC values for these two cases are reported within the successful subset: exchange produced recommendations for 37 (3 new collaborations) and friend-of-a-friend for 16 cases in total.

According to all evaluation criteria, birds-of-a-feather and most-qualified templates best model individuals' motivations for seeking collaboration: NUCATS applicants are more likely to collaborate with researchers who have similar status such as affiliation, gender, and tenure status (the different attributes considered for homophily) and who are most-qualified for requested topics. Considering all test cases, of the 56 cases which the most-qualified template ranked an actual collaborator in the top 5, 38 were either predicted by the birds-of-a-feather template or were previous collaborators. Consistent with previous research [14, 19], this result suggests that researchers do not tend to seek collaboration with the most qualified experts solely because of their qualifications and homophily is the most important social factor for collaboration. Exchange and follow-the-crowd templates, however, mostly recommend previous collaborators and do not perform well on recommending new ones. The structures of their social networks, i.e. reciprocal links and popular experts, have weaker impacts on collaboration decisions.

Table 2. Summary of results for test set with 10 to 40 identified experts (125 test cases). Values with * are not for all cases (exchange made recommendations for 37 (3 new collaborations) and friend-of-a-friend for 16 cases).

	All Test Cases			New Collaborators (69 of the 125 test cases)		
	P@1	P@5	ARC	P@1	P@5	ARC
Most-Qualified	0.14	0.45	8.3	**0.17**	0.45	7.9
Birds-Of-A-Feather	0.18	**0.57**	**6.6**	0.12	**0.49**	**7.5**
Follow-The-Crowd	0.11	0.29	9.9	0.06	0.16	12.6
Exchange	**0.19**	0.30	1.6*	0.04	0.04	1*
Friend-Of-A-Friend	0.03	0.05	6.2*	0.06	0.09	6.2*

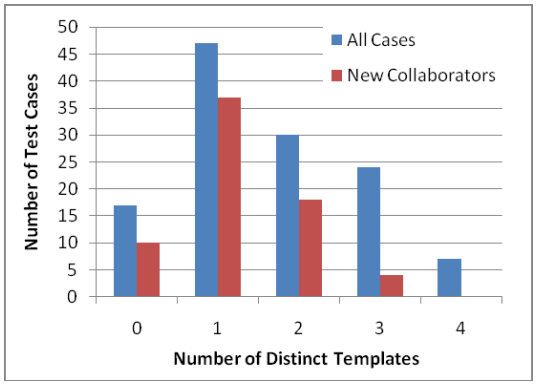

Figure 1 Number of distinct motivation templates that correctly ranked the actual collaborator for each test case in the top 5 of identified experts for all 125 and 69 new

Figure 1 illustrates that recommendation heuristics based on different theories of social drivers differ from each other and different templates are capturing different underlying dimensions. When considering the top 5 recommended experts for each test case, actual collaborators in 47 of the 125 test cases were correctly predicted by only one motivation template (22 were ranked by birds-of-a-feather, 15 by most-qualified , 5 by follow-the-crowd, 3 by friend-of-a-friend, and 2 by exchange), and 37 of which were new collaborators. The fact that many of the actual collaborators were recommended uniquely by only one of the motivation templates supports the claim that there is no one best collaboration recommendation heuristic, and reinforces the importance of considering multiple heuristics. Pure expertise matching or network relations oriented methods in isolation cannot provide accurate recommendations. Social driver based recommenders integrate the level of expertise matching and the structure of social networks and make better recommendations.

Another criterion for evaluating the utility of the system that recommends experts is to examine whether the collaborations recommended are successful, i.e. actually received the funding awards. Table 3 follows the same format as Table 2 but considers only the collaborations among funded proposals. Here the birds-of-a-feather and most-qualified templates are still the best models

Table 3. Summary of results for test set with 10 to 40 identified experts considering only the funded proposals (20 test cases). Values with * are not for all cases (exchange made recommendations for 4 (2 new), friend-of-a-friend for 3 cases).

	All Funded Test Cases			New Collaborators (11 of the 20 test cases)		
	P@1	P@5	ARC	P@1	P@5	ARC
Most-Qualified	**0.20**	0.30	**9.1**	**0.36**	**0.36**	**7.5**
Birds-Of-A-Feather	0.05	**0.35**	9.7	0.09	0.27	9.8
Follow-The-Crowd	0.00	0.20	9.6	0.00	0.18	12.0
Exchange	0.15	0.25	1.3*	0.18	0.18	1.0*
Friend-Of-A-Friend	0.10	0.15	1.3*	0.18	0.27	1.3*

for the motivation behind collaborations. However, in comparison to Table 2, in the case of new collaborations, friend-of-a-friend plays a much more important role. In addition, P@1 values for new collaborations are higher than the previous case. This reveals the very purpose of the NUCATS grant program - fostering novel, trans-disciplinary research - and illustrates the power of MTML recommendation strategies in detecting underlying user motivations and selection mechanisms.

6. CONCLUSION AND FUTURE WORK

This paper presented the design of flexible expert recommender systems capable of utilizing heterogeneous data, multiple theoretical mechanisms, and also featuring personalization of the recommendation based on user motivation. The key advantages of the proposed system are its ability to work with complex multidimensional networks, different user motivations, and heterogeneous sources of data and provide recommendations beyond the single individual's point of view.

The MTML framework helps us conduct research to better understand how context influences decision making. An MTML approach to recommender systems serves researchers as a unique tool for testing hypotheses regarding the formation of collaborations in different contexts. The diversity results reported earlier in Figure 1 support the view that, in the absence of an explicit declaration by the user of their motivation, combining the results of a suite of recommendation heuristics may result in more satisfying recommendations than any one heuristic. An important open question arises, which is devising the appropriate scheme for *combining* the heterogeneous information into a conceptually valid ranking or score. In the current system, the weights for different attributes and measures and how they are combined are specified by users. We plan to pursue ideas and techniques from successful model development tools of SNA, in particular exponential random graph models (ERGM, also known as p* models) [44], with the goal of developing a conceptual model for this optimization problem, and describe a principled and effective mechanism for estimating weights based on existing successful collaboration data in a particular domain.

In this paper, we focused on evaluating the expert selection process and used a simple method for identifying experts. The intent of the evaluation was to show that a recommender based on MTML works and makes reasonable recommendations in line with how previous research suggests individuals form teams and seek experts. As such, the presented evaluation illustrates the value of the new approach and suggests potential research directions. Further work should be done on evaluating different identification strategies and comparing the results from both stages to results obtained from applying other existing recommendation approaches. In particular, identification may be improved by techniques to facilitate users accurately selecting one or more keywords from among tens of thousands, such as semantic representations of hierarchical taxonomies. In addition, it is also important to evaluate the system against criteria like efficiency and scalability which can only be evaluated through the use of the system in diverse domains. We plan a more realistic evaluation of an MTML-based recommender by fielding a version customized to a particular domain of researchers, inviting them to try the system as an aide to team assembly, and following up with them on their successful and unsuccessful grant applications and research programs. Candidate domains under consideration include the linked open data of the VIVO project (http://vivoweb.org/) and the Elsevier SciVal community (http://scival.com/).

7. ACKNOWLEDGMENTS

The authors gratefully acknowledge the help and constructive comments of MeiKuan Huang, JinLing Li, and York Yao. This research is supported by the National Science Foundation (NSF award numbers 0838564 and 0836262) and the Natural Science and Engineering Research Council of Canada. Thompson-Reuters made available the Web of Science bibliographic data under terms of a data sharing agreement with Northwestern University.

8. REFERENCES

[1] Adomavicius, G. and Tuzhilin, A. Toward the Next Generation of Recommender Systems: A Survey of the State-of-the-Art and Possible Extensions. *IEEE Transactions on Knowledge and Data Engineering*, 17, 6 (2005), 734-749.

[2] Balog, K., Azzopardi, L. and de Rijke, M. Formal models for expert finding in enterprise corpora. In *Proceedings of the 29th annual international ACM SIGIR conference on Research and development in information retrieval* (Seattle, Washington, 2006). ACM, New York, NY, 2006.

[3] Berners-Lee, T., Hendler, J. and Lassila, O. The Semantic Web. *Scientific American* (May 2001), 29-37.

[4] Bianchini, D., De Antonellis, V. and Melchiori, M. Flexible semantic-based service matchmaking and discovery. *World Wide Web Journal*, 11, 2 (Jun 2008), 227-251.

[5] Bizer, C., Heath, T., Idehen, K. and Berners-Lee, T. Linked data on the web. In *Proceedings of the 17th International Conference on World Wide Web* (Beijing, China, 2008). ACM, New York, NY, 2008.

[6] Blau, P. M. *Exchange and power in social life : with a new introduction by the author*. Transaction Publ., New Brunswick [u.a.], 2006.

[7] Borgatti, S. P. and Cross, R. A relational view of information seeking and learning in social networks. *Management Science*, 49, 4 (Apr 2003), 432-445.

[8] Brin, S. and Page, L. The anatomy of a large-scale hypertextual Web search engine. *Computer Networks and ISDN Systems*, 30, 1-7 (1998), 107-117.

[9] Burke, R. Hybrid systems for personalized recommendations. In B. Mobasher and S. S. Anand, *Intelligent Techniques for Web Personalization* Springer-Verlag, Berlin, 2005.

[10] Cao, Y., Liu, J., Bao, S. and Li, H. Research on Expert Search at Enterprise Track of TREC 2005. In *Proceedings of the 14th Text REtrieval Conference (TREC 2005)* (2005). National Institute of Standards and Technology, Boulder, CO, 2005.

[11] Casciaro, T. and Lobo, M. S. Competent jerks, lovable fools, and the formation of social networks. *Harv Bus Rev*, 83, 6 (June 2005), 92-99, 149.

[12] Colucci, S., Di Noia, T., Di Sciascio, E., Donini, F. M., Mongiello, M. and Mottola, M. A formal approach to ontology-based semantic match of skills descriptions. *Journal of Universal Computer Science*, 9, 12 (2003), 1437-1454.

[13] Contractor, N. S. and Monge, P. R. Managing Knowledge Networks. *Management Communication Quarterly*, 16, 2 (2002), 249-258.

[14] Contractor, N. S., Wasserman, S. and Faust, K. Testing multitheoretical, multilevel hypotheses about organizational networks: An analytic framework and empirical example. *Academy of Management Review* (July 2006), 681-703.

[15] Corman, S. R., Kuhn, T., McPhee, R. D. and Dooley, K. J. Studying Complex Discursive Systems. *Human Communication Research*, 28, 2 (2002), 157-206.
[16] Earl, M. Knowledge management strategies: Toward a taxonomy. *Journal of Management Information Systems*, 18, 1 (Summer 2001), 215-233.
[17] Fazel-Zarandi, M. and Fox, M. S. Semantic Matchmaking for Job Recruitment: An Ontology-Based Hybrid Approach. In *Proceedings of the 8th International Semantic Web Conference* (2009), 2009.
[18] Fox, M. S., Barbuceanu, M. and Gruninger, M. An organisation ontology for enterprise modeling: Preliminary concepts for linking structure and behaviour. *Computers in Industry*, 29, 1-2 (1996), 123-134.
[19] Hansen, D. L., Khopkar, T. and Zhang, J. Recommender Systems and Expert Locators. *Encyclopedia of Library and Information Sciences, Third Edition* (2010), 4433 - 4441.
[20] Huang, Y., Contractor, N. and Yao, Y. CI-KNOW: recommendation based on social networks. In *Proceedings of the 2008 International Conference on Digital Government Research* (Montreal, CA, 2008). Digital Government Society of North America, 1367840, 2008.
[21] Kautz, H., Selman, B. and Shah, M. Referral web: Combining social networks and collaborative filtering. *Communications of the Acm*, 40, 3 (March 1997), 63-65.
[22] Kirchhoff, L., Stanoevska-Slabeva, K., Nicolai, T. and Fleck, M. Using social network analysis to enhance information retrieval systems. In *Proceedings of the 5th Conference on Applications of Social Network Analysis* (2008), 2008.
[23] Kleinberg, J. M. Authoritative sources in a hyperlinked environment. *J. ACM*, 46, 5 (1999), 604-632.
[24] Krackhardt, D. and Brass, D. J. *Intraorganizational networks: the micro side*, 1994.
[25] Krackhardt, D. M. *Cognitive social structures*. Johnson Graduate School of Management, [Ithaca, N.Y.], 1986.
[26] Krulwich, B. Lifestyle finder - Intelligent user profiling using large-scale demographic data. *Ai Magazine*, 18, 2 (Sum 1997), 37-45.
[27] Lappas, T., Liu, K. and Terzi, E. A Survey of Algorithms and Systems for Expert Location in Social Networks. In C. C. Aggarwal, *Social Network Data Analytics* Springer US, 2011.
[28] MacDonald, C. and Ounis, I. Voting for candidates: adapting data fusion techniques for an expert search task. In *Proceedings of the 15th ACM International Conference on Information and Knowledge Management* (Arlington, Virginia, USA, 2006). ACM, New York, NY, 2006.
[29] Malinowski, J., Keim, T., Wendt, O. and Weitzel, T. Matching People and Jobs: A Bilateral Recommendation Approach. In *Proceedings of the 39th Annual Hawaii International Conference on System Sciences (HICSS'06)* (Kauai, HI, 2006), Honolulu, HI, 2006.
[30] Maybury, M. T. *Expert finding systems*. MITRE Center for Integrated Intelligence Systems Bedford, Massachusetts, USA, 2006.
[31] McDonald, D. W. Recommending collaboration with social networks: a comparative evaluation. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (Ft. Lauderdale, FL, 2003). ACM, New York, NY, 2003.
[32] McDonald, D. W. and Ackerman, M. S. Just talk to me: a field study of expertise location. In *Proceedings of the 1998 ACM Conference on Computer Supported Cooperative Work* (Seattle, Washington, United States, 1998). ACM, New York, NY, 1998.
[33] Mochol, M., Wache, H. and Nixon, L. Improving the Accuracy of Job Search with Semantic Techniques. In *Proceedings of the 10th International Conference on Business Information Systems* (2007). Springer, New York, NY, 2007.
[34] Monge, P. R. and Contractor, N. S. Emergence of Communication Networks. In F. Jablin and L. L. Putnam, *The New Handbook of Organizational Communication* Sage, Thousand Oaks, CA, 2001.
[35] Monge, P. R. and Contractor, N. S. *Theories of communication networks*. Oxford University Press, Oxford, 2003.
[36] Pizzato, L., Rej, T., Chung, T., Koprinska, I. and Kay, J. RECON: a reciprocal recommender for online dating. In *Proceedings of the 4th ACM Conference on Recommender Systems* (Barcelona, Spain, 2010). ACM, New York, NY, 2010.
[37] Reichling, T., Veith, M. and Wulf, V. Expert Recommender: Designing for a Network Organization. *Comput. Supported Coop. Work*, 16, 4-5 (2007), 431-465.
[38] Schafer, J. B., Frankowski, D., Herlocker, J. and Sen, S. Collaborative filtering recommender systems. In B. Peter, K. Alfred and N. Wolfgang, *The adaptive web* Springer-Verlag, 2007.
[39] Serdyukov, P., Rode, H. and Hiemstra, D. Modeling multi-step relevance propagation for expert finding. In *Proceedings of the 17th ACM conference on Information and Knowledge Management* (Napa Valley, CA, 2008). ACM, New York, NY, 2008.
[40] Sinha, R. and Swearingen, K. Comparing recommendations made by online systems and friends. In *Proceedings of the DELOS Workshop: Personalisation and Recommender Systems in Digital Libraries* (Dublin, Ireland, 2001). DELOS, Trier, Germany, 2001.
[41] Stankovic, M., Wagner, C., Jovanovic, J. and Laublet, P. Looking for Experts? What can Linked Data do for You? In *Proceedings of the Linked Data on the Web (LDOW2010)* (Raleigh, NC, 2010), 2010.
[42] Tolksdorf, R., Mochol, M., Heese, R., Eckstein, R., Oldakowski, R. and Bizer, C. Semantic Web technologies in the recruitment process. In *Proceedings of the 7th International Conference Wirtschaftsinformatik* (2006), 2006.
[43] Veit, D., Müller, J. P., Schneider, M. and Fiehn, B. Matchmaking for autonomous agents in electronic marketplaces. In *Proceedings of the 5th International Conference on Autonomous Agents* (Montreal, Quebec, Canada, 2001). ACM, New York, NY, 2001.
[44] Wasserman, S. and Robins, G. An introduction to random graphs, dependence graphs, and p*. In P. J. Carrington and S. J. Wasserman, *Models and Methods in Social Network Analysis* Cambridge University Press, New York, NY, 2003.
[45] Yimam-Seid, D. and Kobsa, A. Expert-finding systems for organizations: Problem and domain analysis and the DEMOIR approach. *Journal of Organizational Computing and Electronic Commerce*, 13, 1 (2003), 1-24.
[46] Zhang, J. and Ackerman, M. S. Searching for expertise in social networks: a simulation of potential strategies. In *Proceedings of the 2005 International ACM SIGGROUP Conference on Supporting Group Work* (Sanibel Island, FL, 2005). ACM, New York, NY, 2005.
[47] Zhang, J., Ackerman, M. S. and Adamic, L. Expertise networks in online communities: structure and algorithms. In *Proceedings of the Proceedings of the 16th international conference on World Wide Web* (Banff, Alberta, CA, 2007). ACM, New York, NY, 2007.