2012 International Conference on Computer and Software Modeling (ICCSM 2012)
IPCSIT vol. 54 (2012) © (2012) IACSIT Press, Singapore
DOI: 10.7763/IPCSIT.2012.V54.09

Human-Matcher-in-the-Loop Model for Facilitating Future Collaborations

Zhe Xu[1], Rajiv Ramnath[1] and Jay Ramanathan[1 +]

Abstract. To facilitate new collaborations within social networks, most existing research focuses on "who can be connected". However, we argue that direct introduction of potential Collaborators is not as effective as an appropriate broker (human-in-the-loop) that bridges them using deeper history knowledge. We refer to the agent that helps connect potential Collaborators as a 'Matcher'. We focus on research questions: 1) which specific individual can best help connect other individuals? 2) how do we track the effectiveness of Matchers for this purpose? We propose a model for identifying Matchers based on *two connection measures* that reflect previous successful connection making. We also use a real-world data set of project collaborations to conduct an experiment. And, we show that the measures perform reasonably to identify Matchers. The practical implication is that we can identify Matchers and promote collaborations across silo organizations.

Keywords: Social Network Analysis, Computer Supported Cooperative Work, Enterprise Collaboration.

1. Introduction

One goal of any organization is to provide the management nodes that are the 'Matchers' to identify collaborators that work on tasks. Existing methods for matching are limiting and time consuming because the Matcher intermediaries are limited (by whom they know or by which social event they happen to know about and attend hoping that something will result). Generally, Matchers are defined here as individuals with tacit knowledge that help future collaborations and enhance existing social networks with new connections. They do this by facilitating other Collaborators in completing successful tasks. The Matcher role achieves this by using a form of deeper social networking knowledge resulting from their own history of interactions with individual Collaborators across organizations. In these interactions they gain knowledge about Collaborators and ways to connect them to new project needs. Because of this knowledge we claim that the 'Matcher' is an essential 'human-in-the-loop' element to successful collaborations. We contribute here by making explicit the tacit social networking knowledge and rules that identify Matchers based on the chronology of previous connections in successful collaborations. These rules can be automated and used to identify Matchers that play an important role when parties that could collaborate beneficially do not know of each other's existence, nor do they have knowledge to make a connection (via email, search queries etc.).

We know the *network knowledge* possessed by Matchers is essential in brokering connections in hierarchical organizations. For example, dedicated Matchers in management roles often conduct the connection making by applying network knowledge for purposeful organizational efficiency [5]. Identifying Collaborators by networking across organizations is very time consuming, yet increasingly important [4].

Using a corpus of project data here, we show here that 1) Matchers can be detected, 2) there also exists tacit network knowledge that can be made explicit in the form of a score that can be algorithmically applied to identify Matchers from previous collaboration history data, 3) the Matchers identified by our score are validated by additional real-world information about the project participants.

[+] Corresponding author. Tel.: + (1-614-565-4187); fax: +(1-614-292-2911).
E-mail address: jayram@cse.ohio-state.edu.

Following the related work section below, we present in the Matcher model that defines the roles and interactions within a collaboration network. In the subsequent section we present two measures to find Matchers in a social network. And, in the next section we discuss the validity of the experimental results. The last section concludes the paper with details on future research.

2. Related Research

The need for large-scale collaboration has been recently identified in [12], and architectural frameworks have been proposed [4, 5]. Searching for expertise in networks is researched by Zhang et al. [1] and compares algorithms such as breadth first search in social networks by simulating on an email dataset. A review of expertise locator systems that locate experts from the web is in [2]. A study on how external factors such as a job role could shape expertise search is conducted in [3]. Aardvark [6] exhibited the interesting human-in-the-loop aspect. This is a recommendation system that searched for the right experts to answer people's questions. Researchers have also explored how agent-based models for collaboration [7, 8].

Finding influential nodes in social networks has been studied by the SNA community [18, 19].Social matching is also related with our study [20, 21]. Social Matching finds the right experts. It focuses on the "what" not "how". "What" is concerned with the final targeted experts and "how" is concerned with the process by which the matchers play important roles. Some research shows boundary objects are important in constructing project teams [11]. This concept is related to our work. Other related work is in motivational aspects of the individual [9] and security/privacy concerns [10].

Our research hypothesis here is that we can collaborate more effectively across organizations if we can successfully identify individuals that are good ‘Matchers’ and possess ‘matching credentials’. These Matchers can help establish new future connections that *augment* an existing social network. The connections serve as a virtual organization that can better facilitate collaborations. Burt [14, 15] introduces the concept of a 'broker' which is very similar with 'Matcher'. Many have also explored collaborations in the context of academic collaborations [16, 17]. However, as far as we know there is no approach that identifies Matchers in social networks.

3. Matching Model

The conceptual Matcher model is presented next along with insights. This is followed by a discussion of how this model would be used in a collaboration environment. Our Matcher model contains three kinds of agent roles: *Collaborator, Observer and Matcher*. A Collaborator provides the resources and implements tasks in the collaboration. Collaborators are often highly motivated to do future tasks after their work products have been successfully delivered. They often become active Observers and Matchers in future projects.

An Observer observes without getting committed to becoming a Collaborator or Matcher. Observers may be interested in certain collaborations. Observers having previous successful project experiences tend to be more highly motivated to get committed. Matchers know individuals that have an interest or resources that contribute (skills, time, assets etc.) towards solving the problem. A Matcher matches Observers and assists them to become Collaborators based on intrinsic knowledge about them. A Collaborator could also be a Matcher for a given collaboration. Figure 1 shows the relationships between three agent roles.

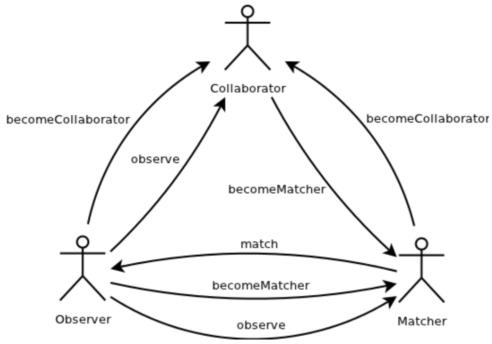

Fig. 1: The Matcher Model.

An agent could be either a human being or a machine. Any individual can also assume multiple roles and work on multiple collaborations simultaneously. But in this paper, we focus on the Matchers, especially human Matchers. Our rationale is given next.

Matcher Human-in-the-Loop: Specifically, a human Matcher interacts with Observers (those that are uncommitted to a collaboration) in ways irreplaceable by a computerized interaction. This is because machines might tell you "who" to collaborate with. But without additional explanations - they eventually fail to explain "why" as well as the human Matcher. A solid explanation of "why" could actually precipitate a collaboration, where none existed before. A machine agent does not do better than a human-in-the-loop even in the case they both have the same knowledge, let alone in situations that are dynamic and non-routine.

4. Finding Matchers

4.1. Problem Statement

Given a social network and its corresponding collaboration history: we ask:

1) Which people were matched by whom?
2) How active is a given Matcher in the network?

Before introducing our solution to the posed questions, we need to explain two simplifications we make.

The first simplification is the number of Matchers and Collaborators in a matching instance. It is possible that in one matching there could be multiple Matchers, we view each matching instance as a triple with only one Matcher and two Collaborators. Any matching that has multiple Matchers and more than two Collaborators can be converted to a standard form by generating all combination triples that have a Matcher and two Collaborators. For example, a matching instance that says M_1 and M_2 match C_1, C_2 and C_3 can be converted to 6 matching instances or triples as follows: (1) M1 matches C_1 and C_2, (2) M_1 matches C_1 and C_3, (3) M_1 matches C_2 and C_3, (4) M_2 matches C_1 and C_2, (5) M_2 matches C_1 and C_3, (6) M_2 matches C_2 and C_3.

The other simplification is Matchers are also the Collaborators. It is possible that a Matcher can match people without getting involved in the project collaboration. However, due to the data availability, it is very hard to detect such Matchers. In this paper, we only focus on finding Matchers who are also Collaborators. We will study the problem of finding Matchers who are not Collaborators in the future.

4.2. Connection Strength Related Definitions

We introduce Connection Strength (CS) to indicate how strongly people are connected. The connection may be across more than two people. If individuals are strongly connected, the connection strength between them is larger. Connection Strength is related to time and changes with time. Formally, we define Connection Strength by using the following notation:

$$\boldsymbol{CS(T_1, T_2, \{I_1, I_2, \dots I_k\})}$$

T_1 and T_2 are two timestamps, and T_1 is before T_2. $I_1, I_2 \dots I_k$ are k individuals. $CS(T_1, T_2, \{I_1, I_2, \dots I_k\})$ is the connection strength between $I_1, I_2, \dots I_k$ in the time period from T_1 to T_2. While different measures can be used to calculate connection strength, we propose and validate one where we can identify the role of individuals as Matchers. That is, here we base CS on the number of collaborations, each of which contains I_1, $I_2 \dots I_k$ as Collaborators, that happened between time T_1 and T_2. We note that this list include a Matcher. Next we use the intuition that a Matcher has a stronger connection with each individual Collaborator separately than what exists between those Collaborators, before those Collaborators are matched. Formally, this can be stated as follows. Let the Matcher be M. Let the Collaborators be C_i and C_j. At time T_m, M matched C_i with C_j. We can state this as:

$$\boldsymbol{Min\{CS(0, T_m, \{M, C_i\}), CS(0, T_m, \{M, C_j\})\} > CS(0, T_m, \{C_i, C_j\})}$$

The logic behind this is: If C_i and C_j already have a connection stronger than those between each of them and the Matcher, it will make no sense to have the Matcher to introduce them. This is because they are already connected to each other better than the Matcher can connect them. Deriving from this formulation, we introduce two key measures:

- **Matching Instance Score (MIS)** to find matching instances.

- **Accumulative Matching Score (AMS)** is the accumulative matching capabilities of Matchers.

MIS is assigned to triples of individuals. A triple can be arbitrarily composed - it does not have to be a matching instance that actually occurred between the individuals. And in each triple, one person is assumed to be the Matcher and the others are assumed to be Collaborators, arbitrarily.

MIS is a score that will predict how probable a particular triple is in reflecting a real-world matching instance. In another words, if a triple has a large MIS, the Matcher assumed in the triple is actually more likely to be the Matcher in the real world for the other individuals in the triple.

MIS is calculated in the following way. For each triple (M, C_i, C_j) M is assumed to be the Matcher while C_i and C_j are assumed to be the Collaborators. The $MIS(M, C_i, C_j)$ has two cases:

1) If the very first collaboration between C_i and C_j does not contain M as one of the Collaborators, $MIS(M, C_i, C_j) = 0$.

2) Otherwise, let the time of the first collaboration between C_i and C_j be T_m. Assume the social network 'start' and 'end' snapshots at time T_s and time T_e. $MIS(M, C_i, C_j) = \log(CS(T_s, T_m, \{M, C_i\}) + 1) * \log(CS(T_s, T_m, \{M, C_j\}) + 1) * \log2(CS(T_m, T_e, \{M, C_i, C_j\}) + 2)$

There are several considerations for designing MIS as 2) above. If $CS(T_s, T_m, \{M, C_i\})$ and $CS(T_s, T_m, \{M, C_j\})$ are small, it is possible that C_i and C_j got matched by someone else since M did not know them very well. On the other hand, if $CS(T_s, T_m, \{M, C_i\})$ and $CS(T_s, T_m, \{M, C_j\})$ are large, M is more likely to be the Matcher. However, $CS(T_s, T_m, \{M, C_i\})$ and $CS(T_s, T_m, \{M, C_j\})$ are not treated linearly with respect to the $MIS(M, C_i, C_j)$. For example, the change of $CS(T_s, T_m, \{M, C_i\})$ from 2 to 3 collaborations is of much larger significance, than the jump from 98 to 99. This reflects the fact that the Matcher plays a more important role in the earlier time interval of matching. It is difficult to have the first collaboration. As the time goes on and more collaborations happen, the Matcher's matching strength is decreased since the collaborators become familiar with each other and do not need to utilize the matcher as before. Thus the weight of $CS(T_m, T_e, \{M, C_i, C_j\})$ should be smaller than $CS(T_s, T_m, \{M, C_i\})$ and $CS(T_s, T_m, \{M, C_j\})$ indicative of this decreased significance. Once C_i and C_j are matched, the number of collaborations they are going to have in the future is not an important factor affecting the Matcher identification process.

AMS is calculated based on Matching Instance Scores. It is a score assigned to each person in the network to indicate matching capability in the future. Formally, let M be a person in the social network. Let AMS(M) be the Accumulative Matching Score of M. Assume the social network has n people $C_1, C_2 \ldots C_n$.

$$AMS(M) = \sum_{1 \le i \le n, 1 \le j \le n, C_i \ne M, C_j \ne M} MIS(M, C_i, C_j)$$

5. Experiment

The goal of the experiment is to first compute the MIS and AMS scores and then test the effectiveness in the next section. We selected Carnegie Mellon's Robotics Institute project data set for evaluation [13]. The dataset contains 5152 project collaboration records from year 1971 to 2006. Each record specifies only two kinds of information: the starting year of that project and the names of collaborators. We calculated MIS scores for all potential triple sets from all 5152 projects. The way we extracted triple set is as follows: for any three people who collaborated at least in one project, we constructed three triples. Every one of the three individuals has a chance to be the Matcher of the other two. Since there is no order between Collaborators, three people can generate only three combinations. After calculating the corresponding MIS scores, we call each triple as a predicted instance. There are 511 instances, each of which has an MIS score larger than zero. Next, to validate the effectiveness of MIS in predicting the real world, we are interested in the following question: How accurate is MIS for identifying Matchers? There exist two challenges in the validation: 1) The number of triple instances is so large that it is hard to validate each of them. 2) Thousands of individuals are involved and it is impossible to interview each of them for validation. To address this we adopted the following approach. First we sorted 511 triple instances according to their MIS prediction scores. Then we sampled by selecting one out of every twenty triples from the sorted list. This process generates 26 triple instances.

The next task was to then verify whether Matchers predicted in the 26 sets are indeed Matchers in the real world using other sources of information discussed next. For each person in the 26 triples, we separately collected their profile information from their public homepages. The information includes academic status, education background, publications, research history (labs they were associated with before and research projects), and relationship (such as students and advisors). Based on such information, we inferred whether a Matcher instance identified by our algorithm is also verifiable as the real Matcher in the real-world situation by using heuristic rules. Most of the heuristic rules come from common sense. Some of these rules applied to each triple instance are listed as follows.

1) If an academic advisor and her student both appear in a matching instance, the advisor has a high probability to be the Matcher.

2) If a person was a leader of a research project and the other two people worked in that project, then the project leader has a high probability to be the Matcher of the instance.

3) If a person co-worked with the other two people in different labs, the person has a high probability to be the Matcher.

Obviously, this kind of verification has a drawback: the inference maybe inaccurate. However, due to the challenges we mentioned earlier, this approach seems to be the best at this point. When we are not confident about the inference, we take a conservative approach and classify the MIS instances as a false positive.

We sorted 26 sample instances according to MIS scores and verified the correctness of each one by using the inference mentioned above. Each instance is classified as either true or false. There are two cases for a prediction to be false: 1) we are confident that the predicted Matcher is not the actual Matcher; 2) we are not confident that the predicted Matcher is the actual Matcher.

Out of 26 triple instances we selected for our sample, 12 of them are classified to be true and 14 are false. The way we calculate the accuracy is as follows. For each given score in the 26 instances, we use the number of true instances that have larger MIS scores than the given score, divided by the number of instances that have larger MIS scores than the given score. Figure 2 shows the relationship between accuracy and MIS score. It shows that if the score is larger than 2, we can get an accuracy as high as 80%.

We next look at the Matchers with high scores and see what roles they play in an academic organization (Table 1). The table shows the corresponding academic roles of Matcher and Collaborators in the 12 triple instances that we validated to be accurate. To understand the table, we explain some of the more obscure roles. Commercialization specialists work closely with government and industry clients to develop technologies from concept to commercialization. The role 'other' indicates those who are either outside the Robotics Institute or not found on the Internet. In conclusion, the table shows Robotics faculty members frequently match Collaborators. Thus, the predictions of individuals, with higher MIS scores as Matchers are consistent with their roles within an academic institution.

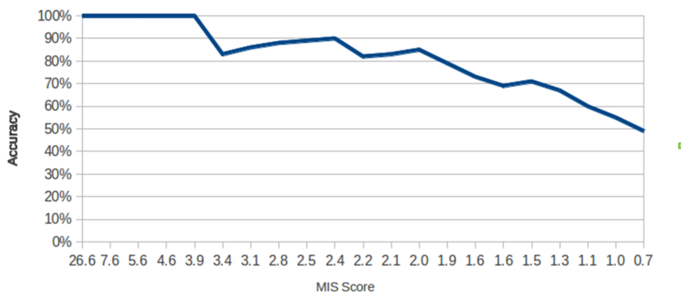

Fig. 2: The Relationship between Accuracy and MIS (Matching Instance Score).

We are interested in the most active Matchers in the network and what roles they play. We calculated AMS scores for all 3342 individuals in the dataset. We adopted a manual approach of searching their roles in their homepages or the homepage of the Robotics Institute. For the validation we extracted academic roles of 10 people who have the largest MAS scores. Their roles and AMS scores are shown in Table 2. We note that the most active predicted Matchers are 'Robotics Adjunct Faculty' and 'Robotics Faculty'.

M Role	C1 Role	C2 Role
Robotics Adjunct Faculty	Robotics Adjunct Faculty	Robotics Faculty
Commercialization Specialists	Other	Student
Commercialization Specialists	Robotics Adjunct Faculty	Student
Robotics Faculty	Student	Student
Robotics Scientist	Student	Student
Robotics Faculty	Robotics Faculty	Post-doctor
Robotics Adjunct Faculty	Other	Other
Commercialization Specialists	Other	Robotics Faculty
Robotics Faculty	Post-doctor	Other
Robotics Faculty	Robotics Faculty	Robotics Adjunct Faculty
Robotics Faculty	Robotics Scientist	Student
Other	Student	Robotics Scientist

Table 1. Role Patterns of 12 Sampled Instances with High Scores.

Ranking	Role	AMS
1	Robotics Adjunct Faculty	441.42
2	Robotics Faculty	412.60
3	Robotics Adjunct Faculty	306.40
4	Commercialization Specialists	276.23
5	Robotics Faculty	266.16
6	Robotics Faculty	263.27
7	Robotics Adjunct Faculty	201.55
8	Robotics Adjunct Faculty	184.22
9	Robotics Scientist	165.64
10	Robotics Faculty	160.07

Table 2. The Roles of Top 10 People Having largest AMS Scores.

Thus, with our limited sample size, we have nevertheless shown that 1) MIS (Matcher Instance Score) predicts the organization-specific real-world validation. And, 2) AMS (Accumulative Matcher Score) predicts the organizational roles of the top ten Matchers. The implications of this are discussed next.

6. Conclusion

Using a corpus of available project data here, we exploration here shows that 1) Matchers are detectable in existing networks; 2) there also exists tacit network knowledge about Matchers that can be captured as Matcher Instance Score (MIS) and Accumulative Matcher Score (AMS) that can eventually be used in a recommender algorithm within an agent-based ExSO Collaboration environment; and finally, 3) we show through an experiment that Matcher scores reliably correspond to independently collected real world information. Leveraging the results real-world applications is the focus of our future research. More generally we show that an environment to facilitate ExSO collaboration not only requires an understanding social network structure, but also ‘knowledge management’ processes related to the Matcher, Collaborator, Observer roles along with credibility established due to previous project successes, other interests and motivations of individuals.

7. Acknowledgement

This research is supported by the National Science Foundation under Grant No. 0753710 and the CERCS IUCRC Center for Enterprise Transformation and Innovation (CETI). We also wish to acknowledge sponsors and members of the Center.

8. References

[1] J. Zhang, and M. S. Ackerman, Searching for expertise in social networks: a simulation of potential strategies. GROUP '05.

[2] I. Becerra-Fernandez, Searching for experts on the Web: A review of contemporary expertise locator systems. ACM Transactions on Internet Technology (TOIT) 2006.

[3] K. Ehrlich and N. S. Shami, Searching for expertise. CHI'08.

[4] M. A. Schilling and C. C. Phelps, Interfirm Collaboration Networks: The Impact of Large-Scale Network Structure on Firm Innovation. Management Science, Volume: 53, Issue: 7, 2007.

[5] I.T. Hawryszkiewycz, System requirements for large scale community collaboration. Collaboration Technologies and Systems (CTS), 2011

[6] D. Horowitz and S. D. Kamvar, The anatomy of a large-scale social search engine. WWW '10.

[7] E. Bonabeau, Agent-based modeling: Methods and techniques for simulating human systems. National Acad Sciences 2002.

[8] G. Cabri, L. Ferrari, and L. Leonardi, Agent role-based collaboration and coordination: a survey about existing approaches. Systems, Man and Cybernetics, 2004.

[9] B. A. Huberman, and C. H. Loch, Collaboration, motivation, and the size of organizations. Journal of Organizational Computing and Electronic Commerce, Volume 6, Issue 2, 1996.

[10] A. F. Van Veenstra, and M. Ramilli, Exploring information security issues in public sector inter-organizational collaboration. EGOV'11.

[11] A. F. Phelps and M. Reddy, The influence of boundary objects on group collaboration in construction project teams. GROUP '09.

[12] http://cacm.acm.org/news/132915-a-made-to-measure-social-network-for-the-academic-world/fulltext

[13] http://www.ri.cmu.edu/

[14] R. Burt, Structural Holes: The Social Structure of Competition. Harvard University Press. 1992.

[15] R. Burt, Brokerage and Closure. Oxford University Press. 2005.

[16] J. N. Cummings and S. Kiesler, Collaborative research across disciplinary and organizational boundaries. Social Studies of Science, 35(5), 703-722. 2005.

[17] B. Gray, Assessing inter-organizational collaboration: Multiple conceptions and multiple methods. In Perspectives on collaboration. 2000.

[18] J. Zhou, Y. Zhang and J. Cheng Preference-based mining of top-Kinfluentialnodes in socialnetworks. Future Generation Computer Systems, June 2012. Elsevier.

[19] M. Kimura, K. Saito, R. Nakano and H. Motoda Finding Influential Nodes in a Social Network from Information Diffusion Data. Social Computing and Behavioral Modeling. 2009.

[20] L. Terveen and D. W. McDonald, Social matching: A framework and research agenda. Transactions on Computer-Human Interaction (TOCHI). 2005.

[21] S. B. Mokhtar, A. J. Mashhadi, L. Capra and L. McNamara, A self-organising directory and matching service for opportunistic social networking. SNS 2010.