Utilizing Academic-Network-Based Conflict of Interests for Paper Reviewer Assignment

Sixing Yan, Jian Jin, Qian Geng, Yue Zhao, and Xirui Huang

Abstract-Peer-review is the most common practice for accessing submitted papers. Previous research mainly focused on maximizing topic relevance and avoiding direct Conflict of Interest (COI) relationship between submitted authors of papers and reviewers. However, potential relationships of Conflict of Interests were not considered which might cause unfair in reviewing process. Hence, in this research, an approach of paper-reviewer assignment was proposed considering potential Conflict of Interests between authors of submitted papers and reviewers. First, academic networks of scholar-to-scholar and institution-to-institution were extracted from their academic activities records. Second, potential conflict of interests of papers-reviewers were measured utilizing paths distance between related authors and reviewers on in academic networks, topic relevance of papers-reviewers was measured using similarity between submitted papers content and publications of reviewers. Third, the paper-reviewer assignment approach was linked to a minimum cost maximum flow problem, which maximized topic relevance and minimized Conflict of Interests under the given constraints. With a big volume of real scholar data and a real task of conference paper-reviewer assignment, the proposed approach was evaluated. The experimental results validated the effectiveness of the proposed approach.

Index Terms—Paper-reviewer assignment, expertise matching, conflict of interests, academic network.

I. INTRODUCTION

In an academic conference, assigning suitable experts to submitted papers is a crucial and challenging step in judging and making decisions of submitted papers. This assignment problem is referred to as paper-reviewer assignment (PRA), in which each paper is evaluated by the most competent reviewers among the Programme Committee (PC) members. quality of scientific knowledge is controlled and the legitimacy of decisions made is safeguarded.

In PRA, selecting reviewer for submitted paper requires relevance and fairness. Specifically, the demand of relevance based on topics, which was defined based on the where the similarity of topics shared by each paper and assigned reviewers [1]-[3]. Also, the demand of fairness requires that each assignment of papers and reviewers should not have

Manuscript received August 9, 2017; revised November 12, 2017. This work is supported by the youth fund project of ministry of education of the humanities and social sciences research (NO. 16YJC870006) and a research project funded by the ISTIC-EBSCO joint laboratory.

Sixing Yan, Jian Jin, and Qian Geng are with School of Government, Beijing Normal University, Beijing, China (e-mail: yansixing@mail.bnu.edu.cn, gengqian@bnu.edu.cn).

Yue Zhao is with the School of History, Beijing Normal University, Beijing, China (e-mail: zhaoyue@mail.bnu.edu.cn).

Xirui Huang is with the Beijing Normal University, Beijing, China (e-mail: Xhua4717@uni.sydney.edu.au).

significant Conflict-of-Interest (COI), as well as guarantee some other constrains, like *Paper Demand Constraint* and *Reviewer Workload Constraint* [4]-[6]. Some researches take the PRA as an *information retrieval* problem [7]-[9] which aims to retrieve a certain number of candidate reviewer who are most competent to the assigning papers. While this kind might not solve the constrains effectively, in more recent studies PRA was taken as a matching problem between a paper set and a reviewer set [10]-[12]. In these researches, various types of COI were introduced as the constrains instead of an optimizing goal.

The common drawback regarding COIs of these paper-reviewer assignment researches is that only obvious COI can would be included while the degree of potential COIs cannot be considered and measured. Although the information of directed COIs is required to be provided, the indirect COIs might be hardly to state. Here are some instances to illustrate which includes three types of entity as Author of paper, Reviewer, as well as Institution of author or reviewer. To begin with, for researcher-to-researcher relationships, an author and a reviewer might have COI through several intermediate coauthors. Also, an author and a reviewer might once be the students of a same doctoral supervisor, or even that reviewer is the doctoral supervisor of that author. For instances, the research institutions of authors and reviewers have a high closeness which produced by previous cooperation or co-worker relationships; Besides, COI might be existed based on the institution-to-institution relationships. The Research Institutions of authors and reviewers might have a high closeness because of a big volume of previous cooperation or co-worker relationships. A direct solution to find out such COI is to collect researcher's academic records, construct the relationship network, then retrieve COI connections and apply them into the assignment tasks. However, such relationship information is large, and retrieve this relationships in network-based data is a time-cost work.

Motivated by the above two observations, in this research, a new assigning approach based on academic network, called Minimum COIs Paper Reviewer Assignment (MinCOI-PRA), is proposed, which assigns paper to reviewers such that the basic demand of relevance is maximized while COIs is minimized, as well as some practical constraints are satisfied for requirement of fairness.

Compared with the existing matching-based methods, the proposed method in this paper achieve a broader relevance and fairness of the assigning process. It not only matches each paper-reviewer pair with high topic relevance, but also measures how a paper is unsuited to assign to a reviewer with user-defined parameters.

II. RELATED WORK

A. Paper Review Assignment

Expertise matching and assignment is matching paper and reviewer utilizing technologies of information retrieve and information filtering [13]. The similar research of reviewer assignment is research paper recommendation since they are all match or assign papers to users. while a relatively few paper will all be assigned to a relatively few users, e.g., technique committees; in paper recommendation system, a few papers will be recommended to many users by recommendation system, e.g., research readers [14].

The reviewer assignment problem mainly concerns about how to assign paper to reviewer well. The first and most important consideration is to guarantee the similarity between paper content and research area of expertise. Additionally, the own wills for research interests of reviewer were taken into consideration by [15]. Research interest areas were pointed out by reviewers first, then a deviation-based algorithm was proposed to predict the reviewing wills of the given papers. An Expertise degree matching is proposed by [16], in which the feature of time was introduced to weight the expertise authority of reviewers.

Another consideration is that, the assignment plan should meet the practice constraints, e.g., the limitation of reviewer's work load, the variety of reviewer's authorities. A nontrivial combinatorial problem involving the assignment of evaluators to grant applications was created by [17]. For each grant application, there have to be at least two evaluators in working group without conflict. Then a graph was constructed to optimize the total workload of an evaluator while assigned evaluator was the council supervisor of application but neither its reviewer nor working group supervisor. A generalized framework for fair reviewer assignment was proposed by [18]. First the topics of domain knowledge from the reviewers published paper were extracted. Then a group assignment of reviewer-to-paper where relevance of paper to topic were weighted.

While such researches mainly consider the relevance between submitted papers and assigned reviewer, the fairness demand of paper-reviewer assignment was seldom concerned.

B. Conflict of Interests in Paper Review Assignment

At the same time, the exist Conflict of Interests (COI) will affect the judgement of reviewers. Personal and group loyalty is a must at the first register, whereas it not only loses its relevance at the second but also undermines such highly valued qualities as open-mindedness and universalism. In other words, the multiple roles played by the scholar produce the situation of a conflict of interest [19]. The result indicated that when COI expertise faced increased, the weight of importance will increase while the information effectiveness will be decreased.

Some research focus on predicting potential COI relationships from other types of connections. A semantics analysis tool was proposed by [20], which detect the COI relationships and measure their possible Intensity. Semantics relation were extract by experiences of researcher, then the coverage of ontological characteristic was calculated as the weight of COI. An expertise matching problem based on a convex cost flow problem was proposed by [4], which

guarantees an optimal solution under various constraints. Then an online matching was generated to support incorporating user feedbacks in real time.

Various types of COI were taken into paper-reviewer assignment. The four aspects were considered in the proposed method by [21], including work load balancing, COI between authors and reviewers, individual preferences and multiple keywords of a proposal. Also, four rules were employed to avoid COI. Authors and reviewers were expected not to be affiliated, co-authors, a student-teacher relationship and colleagues. Different types of COI and their effectiveness on reviewer assignment were studied by [6], then a topic coverage and COI avoiding approach was proposed to achieve the goodness and fairness of paper-reviewer assignment. Also, the experiment results showed that restrict COI constrain will decrease the reviewer's topic coverage toward their assigning papers.

Most of researches studied how to avoiding the direct COI in their expertise matching framework but how to measure the potential COI in the assignment process was less studied. Even though the undirect COI relationship is common in practice, how to measure the degree of such type of COI relationship in assignment is still hard to achieve in exist methods.

C. A Brief Summary

Most research mainly study the relevance between paper and reviewer in reviewer assignment or matching with some practice constrains. However, the complex relationships between authors of paper and reviewer, e.g., potential COI, are seldom concerned, which might occur unfairness in judgement process. Accordingly, in this research, how to extract the potential COIs in academic relationship network and employ them into paper-reviewer assignment is investigated. It is believed to be benefit for achieve both goodness and fairness demands of assignment tasks.

III. PROBLEM STATEMENT

The reviewer assignment problem that is dealt in this paper can be defined as follows: given a set of papers and a set of reviewers, the aim is to maximize the total matching degree such that a certain number of reviewer is assigned to assess a proposal, and a certain number of papers are reviewed by each reviewer under some crisp constraints. The goal of this research is to implement the paper-reviewer assignment with high *Relevance* and low *Conflict of Interest (COI)*, as well as satisfy some practical constraints. Hence, the input data of assignment is denoted at the beginning, then what assigning results are generated and how to generate the assigning results are denoted correspondingly.

To begin with, for a paper-reviewer assignment task, a set of submitted papers $P = \{p_1, p_2, ..., p_n\}$ and a set of committee reviewers $R = \{r_1, r_2, ..., r_m\}$ are given. For each paper $p_i \in P$, its authors are denoted as $F(p_i)$ while their affiliate institutions are denoted as $D(F(p_i))$.Similarly, for each reviewer $r_j \in R$, his or her related institutions is denoted as $D(r_j)$. In order to consider the topic relevance of paper and reviewer, the topic of paper is denoted as $S(p_i)$, as well as topic of reviewer is denoted as $S(r_i)$. Hence, the input data of *Paper*, *Author* of paper, *Reviewer* of PC, *Institution* of paper/reviewer, and *Topic* of paper/reviewer are defined.

Given a paper $p_i \in P$ and a reviewer $r_j \in R$, $M(p_i, r_j)$ is presented as a match if p_i is assigned to r_j . Accordingly, reviewers that will review a paper p_i is denoted as $M(p_i).r$, and papers that assigned to a reviewer r_j is denoted as $M(r_j).p$. Each possible match M regarding to an assignment A between the paper set P and the reviewer R, denoted as $M \in A$. Given an assignment $A \subseteq P \times R$, the set of matches in A involving paper $p_i \in P$ is denoted by $A(p_i)$, i.e., $A(p_i) = \{M | M \in A, M.p = p_i\}$. Similar to $A(p_i), A(r_j) =$ $\{M | M \in A, M.r = r_j\}$ for each $r_j \in R$. The final result of the proposed approach is the assignment set of which performances are expected to be good.

Next, based on a paper set P and a reviewer set R, assignments are expected to generate which satisfies requirements of *Relevance* and *COI*, as well as some other constraints. For a paper-reviewer assignment A, the topic relevance of A is expected to be large, denoted as Relevance(A). Conversely, the *COI* that A involves is expected to be low, denoted as CoI(A). To generate such assignment set, several definitions and constraints are present accordingly. Given a paper p_i and a reviewer r_j , let's start by explaining the Topic relevance requirement and the *COI* avoiding requirement in paper-reviewer assignment.

Topic relevance requirement. The *Relevance* between a paper and a reviewer means that how similar topics they share in their researches are. It is measured by the similarity between the topics of that assigned paper and the research topics of that reviewer's domain. Formally, given a paper p_i and its reviewers $A(p_i)$. *r*, *Relevacen*(*A*) is denoted as,

$$Relevacen(A) = \sum_{p_i \in P} \frac{Similarity(S(p_i), S(A(p_i), r))}{|A(p_i).r|}.$$
 (1)

COI avoiding requirement. A conflict-of-interest (COI) between a paper and a reviewer indicates that the reviewer has an apriority bias for that paper. It is composed of researcher-to-researcher relationships, as well as institution-to-institution relationships of authors and reviewer. The COI among researchers is denoted as $CoI_V(F(p_i), r_j)$, and its COI of institutions is denoted as $CoI_D(D(F(p_i)), D(r_j))$. Then, CoI(A) of each assignment A is denoted as,

$$CoI(A) = \gamma \sum_{p_i \in P, r_j \in R} CoI_V(F(p_i), r_j) + (1 - \gamma) \sum_{p_i \in P, r_j \in R} CoI_D(D(F(p_i)), D(r_j)).$$
⁽²⁾

where $\gamma \in [0,1]$ is a coupling parameter to trade-off author-to-reviewer CoI_V and institution-to-institution CoI_D .

In different applications, the constraints might be different. Hence, several general constraints, named *Reviewer Workload Constraint*, *Paper Demand Constraint* and *Competitor Relationship Constraint*, are introduced in paper-reviewer assignment to take as practice constraints.

The first constraint is called as *Reviewer Workload Constraint*, indicating that each reviewer can only answer a limited number of papers. For easy to explain, here a strict constraint is added, presenting that the number of papers assigned to reviewer r_j , denoted as $|A(r_j).p|$, should be equal or larger than a minimum number Z_{lower} , while it should be equal or smaller than a maximum number Z_{upper} . Formally, this constraint can be defined as,

$$\forall r_j \in R, Z_{lower} \le |A(r_j), p| \le Z_{upper}.$$
(3)

The second constraint is called as *Paper Demand Constraint*, indicating that each paper should be assigned to exactly *H* experts. For instance, in the paper-reviewer assignment task of an academic conference, each paper is expected to be reviewed by 2 to 4 committees. This constrain can be simply added into the optimization problem. Formally, given the number of paper's assigned reviewers, this constraint can be defined as follow,

$$\forall p_i \in P, |A(p_i).r| = H. \tag{4}$$

The third constrain is called as *Competitor Relationship Constraint*, indicating that each reviewer should not be assigned with a paper of which topics are quite similar to his/her own submitted paper in the same conference. Formally, with the help of the similarity of their research topics, the *competitor relationship* between a paper Wp_i and its assigned reviewer's paper p_{r_i} can be defined as,

$$Comp(A) = Similarity(S(p_i), S(p_{r_i})).$$
(5)

Now the requirements and constraints in the paper-reviewer assignment (PRA) are defined. Continually, with the help of above definitions, the goal of PRA in this paper is expected to get the most suitable paper-reviewer matches for assigning. In general, the objective of assignment is viewed as two perspectives: maximizing the requirements and satisfying the given constraints. Therefore, a basic objective function of the PRA can be formed as an optimization problem, defined as,

$$\begin{aligned} & \operatorname{Max}(\operatorname{Relevacen}(A) - \operatorname{CoI}(A) - \operatorname{Comp}(A)) \\ & s.t. \forall r_j \in R \ Z_{lower} \leq |A(r_j).p| \leq Z_{upper} \\ & \forall p_i \in P, |A(p_i).r| = H. \end{aligned}$$

$$(6)$$

So far, the problem of how to generate the assigning results is modeled as an optimization problem. The approach of paper-reviewer assignment proposed in this research, MinCOI-PRA, is different with previous researches since that the main propose of this approach is to achieve a lower COI in assignment which further ensure the *fairness* demand of assigning papers to reviewers.

In the next section, how to measure different types of COIs from academic networks, how to measure the topic similarity and analyze the optimization problem will be discussed.

IV. METHODOLOGY

In this section, technical details about the proposed approach MinCOI-PRA will be elaborated. For the sake for reference symbols and associated definitions are listed in Table I.

TABLE I: NOTATIONS AND DEFINITIONS

Symbol	Definition
Р	The submitted paper set $P = \{p_1, p_2, \dots, p_n\}$.
F	Given a paper $p_i \in P$, its authors are $F(p_i)$.
R	The reviewer set $R = \{r_1, r_2, \dots, r_m\}.$
	Given a paper $p_i \in P$, its authors' work affiliates are
D	$D(F(p_i))$; given a reviewer $r_j \in R$, his or her work affiliates are $D(r_j)$.
S	Given a paper $p_i \in P$, its research topics are $S(p_i)$; given a reviewer $r_j \in R$, his or her research topics are $S(r_j)$.
М	For a match, reviewer $r_j \in R$ is assigned to review the submitted paper $p_i \in P$, denoted as $M(p_i, r_j)$.
A	For an assignment, $A \subseteq P \times R$, the matches of $p_i \in P$ presents as $A(p_i) = \{M M \in A, M, p = p_i\}$; the matches of $r_j \in R$ presents as $A(r_j) = \{M M \in A, M, r = r_j\}$.

A. Academic Network Extraction

The COIs in peer review are generated by previous academic experiences, which means they can be extracted and measured from the academic relationship network. In this research, the academic relationship network models the connections that built by the experience of researchers, that is, researchers might create an academic connection when they shared a same experience. Continually, five types of COIs are introduced in this research. The academic network of researcher-to-researcher is discussed in subsubsection 4.11, denoted as G_V ; the academic network of institution-to-institution is discussed in subsubsection 4.12, denoted as G_D .

In the first place, the weight of edge in a network (graph structure), called *distance*, is denoted as following: given an academic network $G = \langle V, E \rangle$ and two nodes $v_1, v_2 \in V$, the *distance* function between v_1 and v_2 is defined as $\varphi(x_1, x_2) \in [0,1]$ indicating the weight between v_1 and v_2 . It is a common sense that the closer academic relationships between two nodes is, the smaller their distance is. Thus, the *distance* of two nodes is utilized to estimate the probably degree of COI, that is, the closer academic relationships the two researcher (institutions) share, the more COI exist.

1) Academic network of researcher-to-researcher

In the previous research and common senses, three types of researcher-to-researcher COI are discussed in the PRA, including *Co-author relationship*, *Colleague relationship*, and *Advisor-advisee relationship*. The relative distance function for each relationship network are defined.

Co-author relationship. Suppose the co-author records of two members v_1 and v_2 is $C^{1,2} = \{c_i^{1,2}\}_{|C^{1,2}|}$, the distance based on co-author relationship is $\varphi_C(v_1, v_2) \in [0,1]$. For each coauthoring $c^{1,2} \in C^{1,2}$, its time before the present is $Y_{c^{1,2}} \cdot \varphi_C(v_1, v_2)$ can be measured by the sum of each coauthoring that present as *Sigmoid* function with coauthoring time Y. Generally speaking, the connection is stronger when the time of their coauthoring is recent, in other word the $Y_{c^{1,2}v_1v_2}$ is small. Formally, the distance function $\varphi_C(v_1, v_2)$ for *Co-author relationship* can be measured as the sum of all coauthor records,

$$\rho_{C}(v_{1}, v_{2}) = \frac{1}{|C^{1,2}|} \sum_{c^{1,2} \in C^{1,2}} \frac{1}{1 + e^{-Y} c^{1,2}}.$$
 (7)

where $|C^{1,2}|$ is the number of and *e* is the natural logarithm.

Colleague relationship. Given the work experiences of two members v_1 and v_2 , each institution of each member belongs to is denoted as $d^v \in D(v)$. The period time of one of work experience is present as T_{d^v} , and $T_{d^{v_1}d^{v_2}}$ present the work experience they shared. The distance based on the *Colleague relationship* is $\varphi_C(v_1, v_2) \in [0,1]$, which is measured by product of the ratio of the time they work together to the time his or her work there. The colleague relationship is closer when these two members shared more work experiences. Formally, the $\varphi_M(v_1, v_2)$ can be denoted as,

$$\varphi_{M}(v_{1}, v_{2}) = 1 - \frac{\sum_{d^{v_{1}} \in D(v_{1}), d^{v_{2}} \in D(v_{2})} \frac{|T_{d^{v_{1}}, d^{v_{2}}}|}{|T_{d^{v_{1}}}| \times |T_{d^{v_{2}}}|}}{|D(v_{1})| \times |D(v_{2})|}.$$
 (8)

where |D(v)| is the number of the institution they worked at.

Advisor-advisee relationship. Suppose two members have Advisor-advisee relationship when one member is the advisor of another member, or they are advisees of the same advisor. The distance based on the Colleague relationship is $\varphi_A(v_1, v_2) \in \{0,1\}$, which is initialized as $\varphi_A(v_1, v_2) = 1$. The $\varphi_A(v_1, v_2)$ can be formed as an indicator variable, that is if the Advisor-advisee relationship exits in two members, then φ_A is the closest, $\varphi_A(v_1, v_2) = 0$; and $\varphi_A(v_1, v_2) = 1$ otherwise. Formally, $\mathcal{A}(v)$ is utilized to present the advisor of v, then $\varphi_A(v_1, v_2)$ is defined as,

$$\varphi_A(v_1, v_2) = 0$$

s.t., $v_1 = \mathcal{A}(v_2), v_2 = \mathcal{A}(v_1), \mathcal{A}(v_1) = \mathcal{A}(v_2).$ (9)

Compared with relationship of coauthor and colleagues, the advisor-advisee relationship should be limited restrictively, which means the distance of two nodes is the closest when they have advisor-advisee relationship. In sum, given two members v_1 and v_2 , their distance $\varphi_V(v_1, v_2)$ in academic network of researcher can be measured as,

$$\varphi_V = (\alpha \varphi_C + (1 - \alpha) \varphi_M) \times \varphi_A. \tag{10}$$

where α is a coupling parameter to trade-off the distance of *co-author* and *colleague* relationships.

2) Academic network of institution-to-institution

The academic connections between institutions are built by their staffs' academic experiences, including *Cooperation relationship* and *Co-worker relationship*. Given two institutions d_1 and d_2 , different types of distance in their academic networks are discussed. Consider a member $v \in V$ and one of his or her institutions $d \in D(v)$, O(d) is denoted all the members belonged to d.

Cooperation relationship. The degree of cooperation relationship between two institutions d_1 and d_2 is considered about the members' collaboration of these two institutions. The collaboration between two members can be measure using (7). The distance based on the cooperation relationship is denoted as $\varphi_C(d_1, d_2) \in [0,1]$, which can be measured as members' collaborations of these two institutions. Formally, $\varphi_C(d_1, d_2)$ is denoted as,

$$\varphi_{I}(d_{1}, d_{2}) = 1 - \frac{\sum_{\nu_{1} \in O(d_{1}), \nu_{2} \in O(d_{2}), c^{1,2} \in C^{1,2} \frac{1}{1 + e^{-Y} c^{1,2}}}{|O(d_{1})| \times |O(d_{2})|} \quad (11)$$

where |O(d)| present the number of affiliated to institution *d*.

Co-worker relationship. The degree of co-worker relationships between two institutions is generated by staffs

they shared or once shared. In another word, if there is a staff that has work experiences at these two institutions, then a coworker relationship will be built. It can be measured by the ration of the number of co-workers to all staffs they have. With the help of O(d), the $\varphi_W(d_1, d_2)$ can be denoted as,

$$\varphi_W(d_1, d_2) = 1 - \frac{|o(d_1) \cap o(d_2)|}{|o(d_1) \cup o(d_2)|}$$
(12)

Similar to (11), given two institutions d_1 and d_2 , their distance in academic network of institution-to-institution is measured as,

$$\varphi_D = \beta \varphi_I + (1 - \beta) \varphi_W \tag{13}$$

where β is a coupling parameter to trade-off the weight between *cooperation* and *co-worker* relationships. The distance of d_1 and d_2 is small when $\varphi_D(d_1, d_2)$ is low.

After constructing the academic networks of researcher-to-researcher and institution-to-institution, how to measure the degree of COI will be discussed in Section 4.2.

B. Conflicts of Interests Mensuration

The Conflict of Interests (COI) between submitted paper and assigned reviewers come from researcher-to-researcher and institution-to-institution, where the CoI_V generated from researchers and CoI_D generated from institutions. In paper reviewer assignment, COI is mirrored by the academic relationships in which the COI are more likely higher when these distances in academic network are small.

To explain this problem clearly, some terms are clarified first. Given an academic relationship network G, and two nodes v_1 and v_2 of network, the path from v_1 to v_2 is denoted as $L_{v_1 \to v_2}$, and $|L_{v_1 \to v_2}|$ is presented the number of nodes in that path. Thus, the COI in academic relationship network is generated following three basic principles,

- 1) if two nodes are connected directly, they are more likely to have a high COI when their distance $\varphi(v_1, v_2)$ is small. Otherwise, their COI will be reflected though the path of which length is the smallest, and their COI will be measured by the sum of distances between two connected nodes in that path.
- 2) if the shortest path of two nodes contains intermediary nodes that more than a desired value *K*, denoted as $|L_{v_1 \rightarrow v_2}| > K$, then their path is thought to be very long and their academic relationships can be ignored.
- 3) if there is not any path between these two nodes, then their academic relationships can be ignored.

Noticed that the weight of edges and length of path should be concerned when the $CoI_G(v_1, v_2)$ between v_1 and v_2 is measured. For each match $M(p_i, r_j) \in A$, if v_1 presents one of the authors of a paper p_i , $v_1 \in F(p_i)$, v_2 present the assigned reviewer, then the $CoI_V(v_1, v_2)$ of v_1 and v_2 is the minimum value of all possible COI values,

$$Col_{V}(v_{1}, v_{2}) = 1 - \min_{l \in L_{v_{1} \to v_{2}}} \sum_{(v, v') \in l} \varphi_{V}(v, v')$$

s.t., min $|L_{v_{1} \to v_{2}}| \le K.$ (14)

Here a situation with two examples is illustrated. In a local academic network (Fig. 1) containing eight researchers where A1 is an author, and $\{V1, V2, V3, V4\}$ is the coauthor of A1, as well as $\{R1, R2, R3\}$ is the reviewer.



Fig. 1. A local academic network with three roles.

Example 1. As the picture shows that, there is two paths $l_{A_1 \to R_2}^1, l_{A_1 \to R_2}^2 \in L_{A_1 \to R_2}$ between A1 and R2. $l_{A_1 \to R_2}^1 = A1 \to V1 \to V3 \to V4 \to R2$, of which sum of is 1.0 and length of path $|l_{A_1 \to R_2}^2|$ is 5; of which sum of is 1.0 and length of path $|l_{A_1 \to R_2}^2|$ is 4. Then the path $L_{A_1 \to R_2}^2$ is considered as the way generating COI between A1 and R2.

Example 2. the shortest path between A1 and R3 contains 6 nodes, $|l_{A_1 \rightarrow R_3}| = 6$. If K is set as 5, then this path can be ignored which means their relationship can be ignored. Then, compared with R1 and R2, A1 might not have COI with R3.

Accordingly, the problem of measuring potential COI in academic relationship network is modeled as how to find out a path with the smallest accumulated weight in the weighted undirected graph. Generally speaking, the common and simple ways to search the shortest path in the graph are *Dijkstra* algorithm and *Floyd* algorithm. Since the academic networks (graph structure) in this research is very large, here an accelerated *Floyd* algorithm [22] is employed to search the shortest path between two nodes in the graph. The iteration time of that algorithm is set by the maximum length of the path, *K*. Note that, the COI of institution-to-institution $CoI_D(d_1, d_2)$ is measured like $CoI_V(v_1, v_2)$.

C. Topic Relevance Mensuration

In the paper-reviewer assignment, another basic requirement is the *relevance* between paper's content and reviewers' knowledge is expected to be high. In this research, the relevance of each assignment A can be simply measured by the similarity between the submitted paper and the publications of reviewer, denoted as Sim(P).

For each paper $p_i \in P$, the assigned reviewers $r_j \in A(p_i).r$ is expected to cover different distinct topics of that paper. At the same time, the related topics of that paper are expected to be covered by as many reviewers as possible. As a result, to achieve both considerations, the measure similarity is sued to combine both of them, i.e.,

$$Sim(p_i) = \frac{1}{H} \sum_{r_j \in A(p_i).r} \frac{|S(p_i) \cap S(r_j)|}{|S(p_i)|}.$$
 (15)

where $|S(p_i)|$ (resp. $|S(r_j)|$) is the number of topic involved in p_i (resp. r_j).

In the process of paper-reviewer assignment and expert recommendation, LDA model [23], Author-Topic Model and Expectation Maximization Model [18] were commonly utilized to model the topics of documents. While the traditional LDA model can be effectively applicated in topic analysis of document, in this research, it was directly employed to analysis topics of papers and reviewers. The topics of reviewer r_j can be generated from his/her previous publication set T_j which was also presented in some related works. First, submitted papers and reviewer's publications were collected as input, where each document was taken as an individual. Topic distribution of each publication t_j of reviewer and each paper p_i of submitted pool were then generated as $S(t_j)$ and $S(p_i)$, respectively. Similar methods in expert matching research can be seen in [24].

D. Optimization Problem

Generally speaking, the COIs between authors of paper and assigned reviewer are expected to be minimized; the research relevance between paper and the research of assigned reviewers are expected to be maximized; some objective demands should be satisfied. Thus, the problem of MinCOI-PRA is naturally an optimization problem with various constraints. Actually, different methods are available to analyze such optimization problem, such as the greedy-based approach [25], the fast-approximate spectral clustering [26] or transfer it to other problem [4].

In order to maximize the objective function (6), in this research, the method in [4] is adopted to solve the optimization problem, where a convex cost network with lower and upper bounds imposed on the arc flows is constructed. The constructing process as described in *Algorithm* 1 is illustrated in Fig. 2. It can be solved by transforming to an equivalent minimum cost flow problem, which is an optimal assignment with respect to (6). Accordingly, the minimum convex cost flow problem can be formulated as the following optimization problem,

$$\min \sum_{(v_1, v_2) \in E(G)} C_{v_1 v_2}(f(v_1, v_2))$$

s.t. $\forall (v_1, v_2) \in E(G), Lw_{v_1 v_2} \leq f(v_1, v_2) \leq Up_{v_1 v_2}.$ (16)

The model is defined on a directed network G = (V(G), E(G)) with lower bound $Lw_{v_1v_2}$ and upper bound $Up_{v_1v_2}$, as well as a convex cost function $C_{v_1v_2}(f(v_1, v_2))$ associated with each arc (v_1, v_2) . Accordingly, minimizing on the graph *G* constructed in *Algorithm* 1 is equivalent to maximizing (6) is proved followingly. Therefore, the optimization of MinCOI-PRA can be reduced to an equivalent minimum convex cost flow problem.



Fig. 2. The construction of convex-cost network flow according to objective function (6).

Algorithm 1: Optimization solving algorithm

Input: The set of paper *P*; the set of reviewers *Q*; the CoI matrix *COI*; the Similarity matrix *Sim*; the Competitor matrix *Comp*; Z_{lower}, Z_{upper} as described above.

Output: An assignment of reviewers to papers maximizing object function.

1.1 Create a network G with source node S and sink node T;

1.2 foreach $p_i \in P$ do

- 1.3 Create a node P_i ;
- 1 Add an arc from source node *S* to node P_i , with zero cost and flow constraint [1,1];
 - 1.5 **foreach** $r_j \in R$ **do**
 - 1.6 Create a node R_j ;
 - 1.7 Add an arc from R_j to sink node T, with square cost function ηf^2 and flow constraint $[Z_{lower}, Z_{upper}]$;
 - 1.8 **foreach** $p_i \in P, r_j \in R, s. t. COI_{ij} \neq 1$ **do**
 - 1.9 Add an arc from P_i to R_j , with linear cost function $-(Sim_{ij} COI_{ij} Comp_{ij})f$ and flow constraint [0,1];
 - 1.10 Compute the minimum cost flow on G
 - 1.11 **foreach** $p_i \in P, r_j \in R, s.t.Col_{ij} \neq 1$ **do**

```
1.12 If flow f(P_i, R_j) = 1 then Assign paper p_i to reviewer r_j;
```

V. EXPERIMENT STUDY AND DISCUSSION

A. Experiment Setup

In this experiment, one real-world dataset is used to valid the effectiveness of the proposed approach MaxCOI-PRA. For mining the academic relationship network, the dataset was built from the ACM digital library which is a publication database of Association for Computing Machinery (ACM). To build this dataset, 351 ACM Fellows from ACM during 1994 to 2016 were obtained from the official website of ACM. Next, these scholars were used as seeds and their 458,295 collaborators were found in *ACM Digital Library* with 1,305,757 papers, 529,546 affiliates records and 19,379 advisor-advisee records, as well as 10,643 institutions.

Evaluating of such proposed approach is challenging. Since there is not any dataset available for reviewer assignment task directly, a mimetic assignment task was built. all the paper published in SIGKDD from year 2013 to year 2015 were collected as the submitted papers, where 487 papers and 1229 authors in total. All program committees (PCs) of ICDM 2015 and SIGKDD 2015 are taken as the reviewers, where 905 reviewers in total. For modeling reviewers' expertise, publicized papers of reviewers were used to generated expertise profiles according to similar researches [4], [18]. Thus, a profile for each author was created in this experiment by concatenation of the abstracts of publications written by that specific author from year 2000 to year 2015.

B. Evaluation Metrics and Baselines

Include a note with your final paper indicating that you request color printing. As there are no standard answers, in order to quantitatively evaluate our method MinCOI-PRA, similar to relative research [24], four metrics are defined,

1) Average Matching Fitness (AMF)

The average matching fitness measures the topic relevance between paper-reviewer, in order to consider how the reviewers are fit for the paper.

AMF based on paper AMF_P. Evaluate the average topic relevance between each paper and its assigned reviewer $AMF_P(P)$, reflecting how paper will be evaluated well. Formally, given a paper set P, the $AMF_P(P)$ can be measured as,

$$AMF_{P}(P) = \frac{\sum_{p_{i} \in P, r_{p_{i}} \in A(p_{i}), r} Sim(S(p_{i}), S(r_{p_{i}}))}{|P|}.$$
 (17)

AMF based on reviewer AMF_R . Evaluate the average topic relevance between each reviewer and his or her

assigning papers $AMF_R(R)$, in order to reflect the how reviewer can access the relevance research they are familiar to. Formally, given a paper set *P*, the $AMF_R(R)$ can be measured as,

$$AMF_{R}(R) = \frac{\sum_{r_{j} \in R, p_{r_{j}} \in A}(r_{j}) p^{Sim(S(p_{r_{j}}), S(r_{j}))}}{|R|}.$$
 (18)

2) Average Avoiding Effect (AAE)

The average avoiding effect measures the existing COI between each paper-reviewer assigning, in order to consider how assignments avoid COI relationship between author of papers and paper's assigned reviewer.

AAE based on submitted paper (AAE_P). Evaluate the COI between authors of each paper and its assigned reviewers $AAE_P(P)$, in order to reflect how well the assignments avoid COI for each paper assigning reviewers. Formally, given a paper set *P*, $AAE_P(P)$ can be measured as,

$$AAE_{P}(P) = 1 - \frac{1}{|P|} \sum_{p_{i} \in P, r_{p_{i}} \in A(p_{i}), r} Col(p_{i}, r_{p_{i}}).$$
(19)

AAE based on reviewer AAE_R . Evaluate the COI between reviewer and authors of his or her assigning papers $AAE_R(R)$, in order to reflect how well the assignments avoid COI for each reviewer are assigned with papers. Formally, given a reviewer set R, $AAE_R(R)$ can be measured as,

$$AAE_{R}(R) = 1 - \frac{1}{|R|} \sum_{r_{j} \in R, p_{r_{j}} \in A(r_{j}), p} CoI(p_{r_{j}}, r_{j}).$$
(20)

3) Benchmark approaches

The goal of this experiment is to consider the effectiveness. To benchmark the performance in terms of four evaluation metrics, three methods are utilized. A simple approach of select reviewers for each paper, the RANDOM Assignment for PRA, is introduced as basic benchmark approach. Additionally, a greedy-based approach for PRA considering topic similarity is also employed to benchmark, which denoted as GREEDY-PRA.

RANDOM-PRA. *H* reviewers are selected randomly as the assigned reviewers to a given paper by the random algorithm, which is meant to serve as a sanity check that calibrates the results of other algorithms. In this experiment, 100 runs will be performed for the random assigning and the top *H* frequent reviewer was taken as the assigned reviewers.

GREEDY-PRA. Given the paper set *P* and the reviewer set *R*, for each paper, select reviewers of which *Relevacen* is highest and *Col* is lowest with other three constraints. Repeat this process until all the papers are assigned with reviewers.

To evaluate the performance of each approach, the *H* is set to be 2 and 4, the Z_{lower} is set to be 5 and the Z_{upper} is set to be 10. The *K* is set to be 1, 3, 5, 7. Note that, K = 1 is the situation that only consider the direct COI relationships in the academic network. Parameter α and β are set to be 0.5 which is meat that two parts in (10) and (13) are equally important.

C. On the Sensitivity of Parameter

Parameter γ of (1) will have a significant impact on the final assigning result. Thus, in this experiment, the impacts of parameter γ is conducted first. According to 5.3, $H = \{2,3\}$ and $K = \{1,3,5,7\}$. The average result of each experiment group is taken as the final results that showed in Fig. 3.



Parameter γ trade-off the COI between researcher-to-researcher and institution-to-institution. As observed from Fig. 3, higher *AMF* and *AAE* are observed if the value of γ near 0.6 and 0.7. One of potential reasons is that, for papers and reviewers, their affiliate institutions will generate more COI. In the following experiments, γ is chosen to be 0.65, in which better performance is observed in terms

D. Evaluation Results and Analysis

of all four metrics.

The proposed approach is evaluated on a basic expertise matching problem, that is paper-reviewer assignment, with two benchmark approaches. Two groups of experiments are conducted to show the effectiveness of the proposed method, and effectiveness of different COI and constraints are analyzed

1) Topic relevance analysis

In Fig. 4, the AMF metric of three approaches are compared. To begin with, when K = 1 considering direct COI only, the assignment performances of MinCOI-PRA and GREEDY-PRA are approximately identical. As observed in Fig. 4, when K > 1 where indirect COI are concerned, the AMF is declined, while GREEDY-PRA is kept. At the same time, when K is increasing, the speed of declining of AMF is more significant. This result simply indicates that considering more COI will make AMF decrease. One potential reason is that two research whose research topics are similar, they are more likely to have cooperating experience, which continually may create more COI. As a result, in order to reduce the degree of COI, MinCOI-PRA may exclude some high COI candidates who also have high topic relevance, thus reduce some topic relevance and have a relatively lower AMF.



Compare the assignment results of MinCOI-PRA in paper and reviewer, it can be observed that avoiding potential COI

decrease more $AMF_P(P)$ than $AMF_R(R)$. Perhaps the reason behind is that, the diversity of paper's research topic is smaller than the diversity of reviewer, which may cause the AMF of rest of reviewer candidates is smaller than the avoiding reviewers. Thus, the $AMF_P(P)$ of total assignment is reduced.

Also, AMF will obtain a higher value when H = 2. It may be inferred that for some rare research area, the number of well-fit candidate reviewer is small. When such candidates are excluded, the rest of candidates may decrease the degree of AMF of the assigning results.

2) Conflict of interest avoiding analysis

In Fig. 5, avoiding COI effectiveness of each approaches' assigning results are showed. Similar to AMF, MinCOI-PRA and GREEDY-PRA perform well in avoiding COI with K = 1. As observed in Fig, when potential COI is taken into consideration, the AAE of GREEDY-PRA is decreased significantly while the K is increased. On the contrary, the AAE of MinCOI-PRA is approximately lasting a high degree. According to 5.4.1, a potential reason is that MinCOI-PRA reduce some parts of topic relevance to maintain the AAE in a relatively high level.



Comparing, the *AAE* of reviewers decline faster when *K* is increasing, which indicated that *AAE* is lost when considering more potential COI of reviewers. One of possible reason behind it is that, general speaking, reviewers of technical committee are senior researchers in their area who may have more coauthorships and research experience. The more academic relationships they have, the more COI will exist. Thus, compared with authors of papers, more connections of reviewers are created in the different types of academic networks, then more COI should be avoided by MinCOI-PRA in the assignment process. Continuedly, *AAE* of assignment results is decreased for avoiding COI relationships.

Also, the results indicate that the assignment in H = 2 and H = 3. A relative better performance of MinCOI-PRA is observed when H = 2. It may release that the less reviewer are assigned, the less COI the assignments will contain.

3) A short summary

In this subsection, two groups of experiments are conducted to evaluate the effectiveness of the proposed approach MinCOI-PRA in two aspects. Compared with benchmark methods, the MinCOI-PRA is showed a better performance in avoiding COI while maintaining a relatively high topic relevance. Also, COI perform better when the number of demand reviewer per paper is small.

VI. CONCLUSION

In this research, the problem of paper-reviewer assignment is studied, in which different types and degrees of Conflict of Interest (COI) are concerned. COI in researcher-to-researcher and institution-to-institution are extracted from academic relationship network, which are expected to be minimized in the optimization problem of paper reviewer assignment. The assignment problem is then formalized as a minimum convex cost flow problem. Experimental results on a real data set demonstrate that the proposed approach can effectively and efficiently match reviewers with the submitted paper set. In the future, many new directions of this work can be study. One attractive study is to apply the such COI mensuration into Question & Answer (Q&A) system since that the Q&A match should also consider both relevance and fairness, e.g., the answer should not be provided by stakeholders regarding questions. Also, it is important to consider the influence among researchers in other social networks.

ACKNOWLEDGMENT

This research is an extent work of the graduation thesis of F. A. Sixing Yan, and he thanks the supervisor Prof. Geng Qian and Dr Jin Jian's advising.

REFERENCES

- M. Karimzadehgan, C. X. Zhai, and G. Belford, "Multi-aspect expertise matching for review assignment," in *Proc. ACM Conference* on Information and Knowledge Management, 2008, pp. 1113-1122.
- [2] D. Mimno and A. Mccallum, "Expertise modeling for matching papers with reviewers," in Proc. of ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 2007, pp. 500-509.
- [3] Y. H. Sun, J. Ma, Z. P. Fan, and J. Wang, "A hybrid knowledge and model approach for reviewer assignment," in *Proc. Hawaii International Conference on System Sciences*, vol. 34, pp. 47-47, 2007.
- [4] W. Tang, J. Tang, T. Lei, C. Tan, B. Gao, and T. Li, "On optimization of expertise matching with various constraints," *Neurocomputing*, vol. 76, no. 1, pp. 71-83, 2012.
- [5] R. N. E. Anggraini, "Conflict of interest (COI) detector in reviewer recommendation system using potential friendship finder," M.S. thesis, Dept. Computer Science and Information Engineering. College of Electrical Engineering and Computer Science. National Taiwan University of Science and Technology, Taiwan, 2012.
- [6] C. Long, C. W. Wong, Y. Peng, and L. Ye, "On good and fair paper-reviewer assignment," in *Proc. IEEE, International Conference* on *Data Mining*, 2013, pp.1145-1150.
- [7] R. Marko, A. Bollen, and Johan, "An algorithm to determine peer-reviewers," in *Proc. ACM Conf. on Information and Knowledge Management*, 2006, pp. 319-328.
- [8] P. Serdyukov, H. Rode, and D. Hiemstra. "Modeling multi-step relevance propagation for expert finding," in *Proc. ACM Conference* on Information and Knowledge Management, 2008, pp. 1133-1142.
- [9] M. Karimzadehgan, C. X. Zhai, and G. Belford, "Multi-aspect expertise matching for review assignment," in *Proc. ACM Conference* on Information and Knowledge Management, 2008, pp. 1113-1122.
- [10] N. Garg, T. Kavitha, A. Kumar, K. Mehlhorn, and M. Juli, "Assigning papers to referees," *Algorithmica*, vol. 58, no. 1, pp. 119-136, 2010.
 [11] W. Tang, J. Tang, and C. Tan, "Expertise matching via
- [11] W. Tang, J. Tang, and C. Tan, "Expertise matching via constraint-based optimization," in *Proc. of International Conference* on Web Intelligence and Intelligent Agent Technology, 2010.
- [12] L. Charlin, R. S. Zemel, and C. Boutilier, "A framework for optimizing paper matching," in *Proc. the Twenty-Seventh Conference on Uncertainty in Artificial Intelligence.*
- [13] S. T. Dumais and J. Nielsen, "Automating the assignment of submitted manuscripts to reviewers," in *Proc. of International ACM SIGIR Conference on Research and Development in Information Retrieval*, 1992, pp. 233-244.

- [14] J. Beel, B. Gipp, S. Langer, and C. Breitinger, "Research-paper recommender systems: A literature survey," *International Journal on Digital Libraries*, vol. 17, no. 4, pp. 1-34, 2015.
- [15] P. Rigaux, "An iterative rating method: Application to web-based conference management," in *Proc. of ACM Symposium on Applied Computing*, 2004, pp. 1682-1687.
- [16] X. Li and T. Watanabe, "Automatic paper-to-reviewer assignment, based on the matching degree of the reviewers," *Procedia Computer Science*, vol. 22, pp. 633-642, 2013.
- [17] K. Cechl árov á T. Fleiner, and E. Potpinkov á "Assigning evaluators to research grant applications: The case of slovak research and development agency," *Scientometrics*, vol. 99, no. 2, pp. 495-506, 2014.
- [18] N. M. Kou, H. U. Leong, N. Mamoulis, and Z. Gong, "Weighted coverage based reviewer assignment," in *Proc. of ACM SIGMOD International Conference on Management of Data*, 2015, pp. 2031-2046.
- [19] A. Oleinik, "Conflict(s) of interest in peer review: Its origins and possible solutions," *Science & Engineering Ethics*, vol. 20, no. 1, pp. 55-75, 2014.
- [20] C. Chen, "Conflict of interest detection in incomplete collaboration network via social interaction," M.S. thesis, Dept. Computer Science and Information Engineering. College of Electrical Engineering and Computer Science. National Taiwan University of Science and Technology, Taiwan, 2008.
- [21] D. K. Tayal, P. C. Saxena, A. Sharma, G. Khanna, and S. Gupta, "New method for solving reviewer assignment problem using type-2 fuzzy sets and fuzzy functions," *Applied Intelligence*, vol. 40, no. 1, pp. 54-73, 2014.
- [22] D. Q. Zhang and W. U. Guo-Lin, "Accelerated and optimized method of floyd algorithm to find out shortest path," *Computer Engineering & Applications*. vol. 45, no. 17, pp. 41-43, 2009.

- [23] X. Liu, T. Suel, and N. Memon, "A robust model for paper reviewer assignment," in *Proc. of ACM Conference on Recommender Systems*, 2014, pp. 25-32.
- [24] L. Li and N. Memon, "Mining groups of common interest: Discovering topical communities with network flows," in *Proc. of International Conference on Machine Learning and Data Mining in Pattern Recognition*, 2013, pp. 405-420.
- [25] R. Mcdonald, "A study of global inference algorithms in multi-document summarization," Advances in Information Retrieval. 2007.
- [26] W. Wei, Z. Y. Ming, L. Nie, G. Li, J. Li, and F. Zhu, "Exploring heterogeneous features for query-focused summarization of categorized community answers," *Information Sciences*, vol. 330, pp. 403-423, 2016.



Sixing Yan was borned in China. He earned his four-year bachlaor degree of information management and information system in Beijing Normal University (BNU), Beijing, China.

He worked as a one-quarter intern at Data Analyst Assistant, International Data Company (IDC) Consulting (Beijing) Ltd. Beijing, China. In this internship, he

worked in the field of end-user research statistic. Currently, he works as a project officer of Nanyang Technological University (NTU).

Yan was a project principal investigator supervised by professor and department director Liu Xiaojuan during 2014 to 2015. Also, he was an assistant undergraduate researcher of prof. Jin Jian during 2015 to 2017. His undergraduate graduation thesis is advised by prof. Geng Qian and prof. Jin Jian in 2017. He is the project officer of assistant professor Arijit Kkan, from September 2017 to present.