PISTIS: A Conflict of Interest Declaration and Detection System for Peer Review Management

Siyuan Wu
University of Macau
Macau SAR
mb55408@umac.mo

Leong Hou U
University of Macau
Macau SAR
ryanlhu@umac.mo

Sourav S Bhowmick
Nanyang Technological University
Singapore, Singapore
assourav@ntu.edu.sg

Wolfgang Gatterbauer
Northeastern University
Boston, USA
wolfgang@ccis.neu.edu

ABSTRACT

Detecting conflicts of interest (COIs) is key for guaranteeing the fairness of a peer-review process. In many conference management systems, the COIs of authors and reviewers are self-declared, and the declaration process is time consuming and potentially incomplete. To address this problem, we demonstrate a novel interactive system called PISTIS that assists the declaration process *in a semi-automatic manner*. Apart from keyword search and simple filtering, our system provides an interactive graphical interface that helps users explore potential COIs based on the heterogenous data sources. To simply the process of declaration, we also recommend latent COIs using a supervised ranking model that can be iteratively refined from the data collected from past declarations. We believe that PISTIS can be useful as an assistant tool in many real world conference management systems.

ACM Reference Format:
Siyuan Wu, Leong Hou U, Sourav S Bhowmick, and Wolfgang Gatterbauer. 2018. PISTIS: A Conflict of Interest Declaration and Detection System for Peer Review Management. In *Proceedings of 2018 International Conference on Management of Data (SIGMOD'18).* ACM, New York, NY, USA, 4 pages. https://doi.org/10.1145/3183713.3193552

1 INTRODUCTION

A fair peer-review process is a key ingredient for running a successful academic event. From an author's point of view, fairness of a review process is paramount to her research endeavor. From the academic event's point of view, fairness has direct impact on its reputation. Fairness is affected by many factors, such as the expertise of reviewers, the quality of review comments, the design of the review form, etc. However, the most important factor is the relationships between authors and reviewers. In this demonstration, we present a novel reviewer suggestion system that focuses on declaration and detection of *conflicts of interest* (COIs) in the peer-review process [6], an issue that has received scant attention despite its significance in upholding quality and fairness of an academic event.

SIGMOD'18, June 10–15, 2018, Houston, TX, USA

ACM ISBN 978-1-4503-4703-7/18/06…$15.00
https://doi.org/10.1145/3183713.3193552

In academic peer-review, we can categorize COIs into two types, namely *definite COIs* and *latent COIs*. Some examples are given in Table 1. As a common practice in existing conference management systems (e.g., Conference Management Toolkit[1] and EasyChair[2]), the definite COIs can be collected by a set of *declaration rules*. Unfortunately, these rules cannot cover all COIs. For instance, an author and her academic siblings (i.e., two researchers with the same advisor but have never published together) may have conflicts of interest but this is not required to be declared according to the rules. One possible reason is that not every academic sibling relationship has conflicts of interest. These COIs can be determined subjectively but they could potentially influence the quality of a peer review process [3].

Definite COIs	1. Collaborator in the past two years 2. Advisor - Advisee 3. Colleague in the same university 4. ...
Latent COIs	1. Close friend 2. Academic Sibling 3. Academic Sibling's colleague 4. ...

Table 1: Latent and definite COIs

There is also a lack of effective and efficient tools to facilitate the self-declaration process. Existing conference management systems request the program committee members and authors to declare their potential COIs by displaying a potentially lengthy list of reviewers. In certain venues, this list may contain several hundreds of reviewers. We argue that this declaration process is too time-consuming and incomplete (e.g., authors may intentionally or unintentionally overlook some potential COI cases). Even though program committee chairs (in conferences) or editors (in journals) may check suspicious COIs on their own [3], this approach is very time consuming, incomplete and error-prone. Hence there is a need for a framework that can help us detect COIs automatically.

[1] https://cmt3.research.microsoft.com/
[2] http://easychair.org/

Prior work [1] proposes to detect latent COIs automatically using some weighted relationships. However, it is difficult to obtain the ground truth COI cases since there is lack of a universal definition of COI. The lack of high quality ground truth datasets significantly restricts the development of the supervised COI learning.

In this demonstration, we present a novel interactive system called PISTIS[3] (**P**latform for Confl**I**ct of Intere**ST**-aware Rev**I**ewer **S**uggestion) to address the COI declaration and detection problem. It consists of three main modules for (1) data extraction, (2) graphical declaration, and (3) COI learning. The *data extraction* module is used to extract some latent COIs from public sources. For instance, the advisor-advisee information can be extracted from DBLP[4] using a factor graph model [5]. The *graphical declaration* module provides a graphical interface that helps us to explore potential COIs of authors based on a meta-path graph [4]. For instance, a user can explore her academic siblings using a meta-path "advisor-advisee". The *COI learning* module uses a supervised learning model based on the declaration data [6]. Based on the learning result, we rank the latent COIs and display a recommended COI list to the end user. Note that the feedback of the recommended list will also be considered in the subsequent round of supervised learning.

We also demonstrate how our system can be used in a user-friendly manner in two common scenarios: (1) reviewer and author COI declaration, and (2) program chair COI verification. PISTIS not only provides an interactive graphical interface and a recommended list of COIs but also shows the detail of each meta-path instance. The detail information will help reviewers and authors to better comprehend the author-reviewer relationship and make their decision. Once all COI information has been collected, PISTIS will return the information back to the conference management system and then the peer-review assignment can be conducted subject to these constraints. Subsequently, it will collect the assignment result and rank suspicious assignment pairs based on their latent COI scores. This enables the program chairs to verify the assignment result and manually drop some suspicious assignment pairs. Note that the inputs from program chairs will also be considered in the next round of supervised learning. Consequently, we envisage that the quality of COI detection in PISTIS will be continuously improved using this mutually reinforced model.

2 SYSTEM OVERVIEW

Figure 1 depicts the architecture of PISTIS. As mentioned in Section 1, we have three main components, including the data extraction module (backend), the graphical declaration panel (frontend), and the COI learning module (backend). Besides these main modules, our system provides the COI ranking list panel (frontend), the meta-path information panel (frontend), and the COI data management (backend). In addition, the interface layout is changed according to the role of users (reviewer, author, and program chair).

[3]In Greek mythology, PISTIS was the personified spirit (*daimona*) of trust, honesty, and good faith.
[4]http://dblp.uni-trier.de/db/

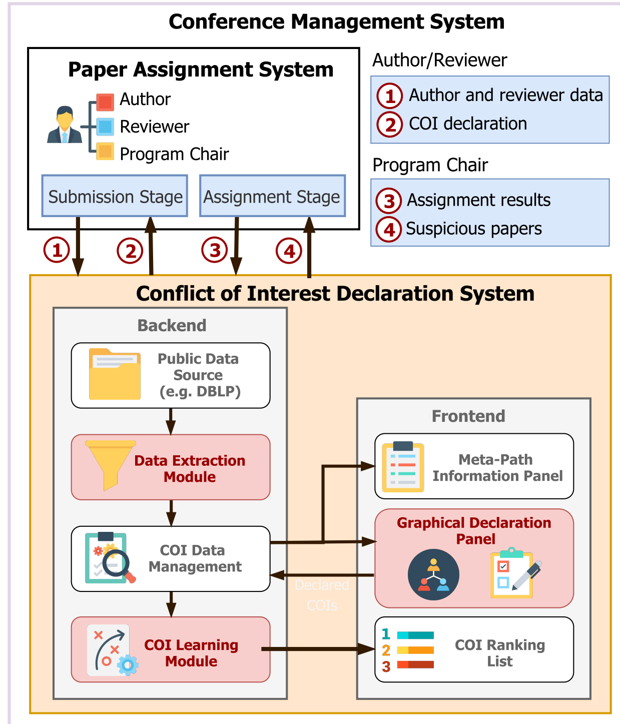

Figure 1: Architecture of PISTIS

2.1 Backend Modules

We first discuss the backend modules that include all data preprocessing, management, extraction and mining processes.

Data extraction module. Our system extracts the heterogenous relationships from some public sources, e.g., co-authorship (from DBLP) and co-working period (extracted from the affiliations). In addition, we extract the advisor-advisee relationship using a factor graph model [5]. This heterogenous information is then used to build a heterogeneous publication network that forms the main knowledge base of the system.

COI learning module. To automatically detect COIs, our prior work [6] studies a *logistic regression* model to compute that can learn the latent COI scores between users. Specifically, we define the proximity of two authors based on their collaboration information. To infer the latent COI score, we attempt to train the weight of the meta-paths, e.g., advisor-advisee and collaborator-collaborator, based on the positive and negative COI cases. Given the trained weights and the proximity value, we calculate the latent COI score of two users u and v as $\frac{e^{W^T L+b}}{e^{W^T L+b}+1}$, where W and L are the weight and the proximity vectors of the meta-paths between u and v, respectively, and b is a constant in the logistic regression model. However, our learning module suffers from the lack of a publicly-available high quality ground truth data for COI cases. This is one main reason for developing an interactive system to collect the COI information in a semi-automatic manner.

COI data management. The heterogeneous publication network is stored in standard relational tables, which include co-authorship information, co-working information, and advisor-advisee information. This information then supports the learning and the frontend modules. To enhance our learning module quality, we also store newly discovered positive cases (e.g., user self-declared COIs) and negative cases (e.g,. unselected COI cases) in the database for the subsequent round of training. We shall discuss how to declare the positive and negative cases in the next section.

2.2 Frontend Panels

We next discuss our frontend panels and show how they can help during COI declaration. To simplify our discussion, we assume the user to be an *author* and use the user interface prototype in Figure 2.

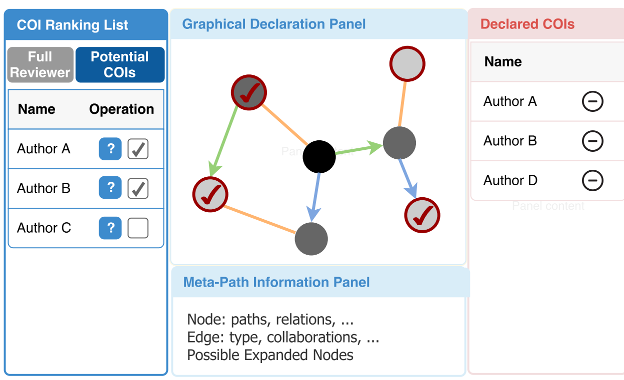

Figure 2: Prototype of the declaration system interface

COI ranking list panel. When an author enters into the COI declaration system, this panel lists her potential COI cases in descending order of the COI scores, where each case must be in the reviewer set. In addition, we automatically mark some reviewers as initial COI cases if they can be detected objectively by rules, e.g., co-author in the past 2 years. When an author clicks on the name of a latent COI case, their relationships, e.g., the collaboration records, will be shown in the meta-path information panel. After subjectively considering the detailed information, the author can report this case as a COI by simply clicking on the button beside the name. From our point of view, this panel is not only an assistant tool for helping the COI declaration but also a reminder for the suspicious cases of the author. This panel thus not only assists the authors during COI declaration, but it also brings their attention to the suspicious cases.

Graphical declaration panel. This graphical panel is the key module in the declaration system (see Figure 3). At the beginning, the panel draws a subgraph centered around the author, where the subgraph includes all reachable reviewers of 2-hops[5] based on all relationships. We use different graph patterns to represent the heterogenous relationships and types of nodes. For instance, we use the gray level to indicate the hop distance from the author. In this example, the node color of 'Alice' is darker than that of 'James' since 'Alice' is a closer node to the author. We use red border nodes to indicate the reviewer nodes of the venue, e.g., 'Mike', 'James',

[5]We only show the reviewers of 2-hops initially since some prestigious scholars may have a huge collaboration network.

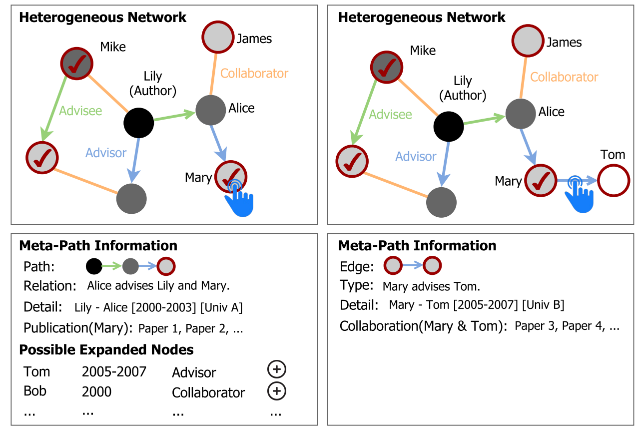

Figure 3: Graphical declaration panel

and 'Mary' are in the reviewer set but 'Alice' is not. In addition, all detected and self-declared COIs are highlighted by ticks.

When clicking on an edge or a node in the graph, the corresponding edge and path(s) (from the author node to the selected node) will be highlighted. The detailed information of the selected edge or the selected path(s) will be shown in the meta-path information panel. Moreover, the author may expand the subgraph using the meta-path information panel, that shows all possible expanded candidates from the selected node. As an example in Figure 3, 'Tom' and 'Bob' are two possible candidates to be expanded from 'Mary'. 'Lily' picks 'Tom' into the subgraph for further investigation since 'Tom' is an advisee of her close academic sibling 'Mary'. The expansion feature is particularly helpful when the author is looking for the academic siblings of her advisor(s) or advisee(s).

Meta-path information panel. When clicking on a node, the corresponding path(s) from the author node to the selected node are extracted from the subgraph. The system retrieves the heterogenous information of the selected path(s) from the COI data management module. According to the information, the panel will show a summary of the selected path(s). For instance, we say 'Alice' advises 'Lily' and 'Mary' in Figure 3. The detail of these relationships is also listed in the panel, which includes the type of each relation, the period of each relation, and the affiliation, etc. We also show the recent 3 years publication records of the selected node. When clicking on an edge, the panel will show the information in the way but the information are only related to this edge.

Initially our subgraph only contains 2-hop information, which can be expanded using this panel. When clicking on a node, some candidates will be shown in a list at the bottom of the panel. The author can add any candidate into the subgraph by pressing the corresponding button. This expansion procedure (which is a heterogenous network traversal) helps the author find more COI cases.

3 DEMONSTRATION SCENARIOS

In this section, we show the screenshots of PISTIS under two scenarios: (1) COI declaration (for reviewers and authors) and (2) COI verification (for program chairs).

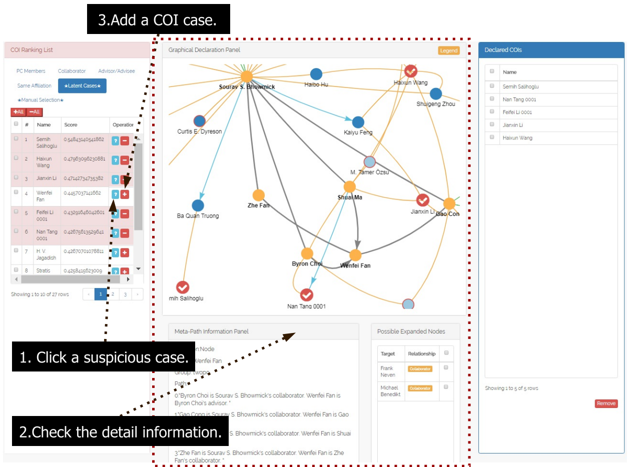

Figure 4: Screenshot of COI declaration

COI declaration (Figure 4). Initially, the latent COIs are extracted automatically and listed in the declared COI list (at right hand side of the system). To self-declare definite COI cases, our system offers two options, (1) searching from the ranking list and (2) exploring from the graphical interface. For example, the user can click on the question mark of a suspicious case, then the graphical interface will highlight this case and show the detail information in the meta-path panel. A suspicious case can be added into the declared COI list by clicking the ⊕ button beside the case.

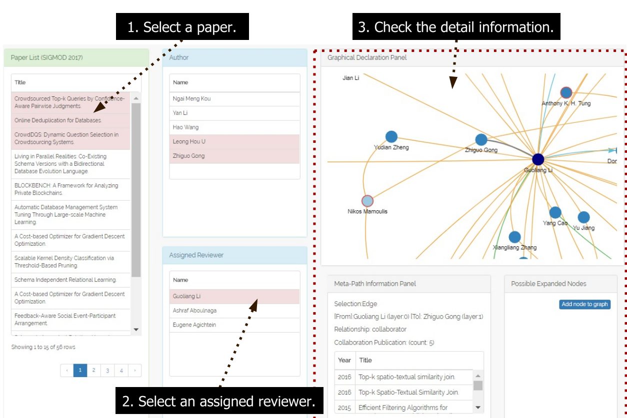

Figure 5: Screenshot of COI verification

COI verification (Figure 5). This sub-system is specifically designed for program chair(s) to verify the assignment result. The verification system has a panel to navigate the assignment result, where the papers are sorted based on the latent COI scores. When the program chair(s) investigates on a suspicious case, the relationships between the assigned reviewers and the authors are highlighted in the graphical declaration panel.

4 RELATED WORK

Automatic COI detection is a very new arena in data management and analytics. Existing studies on COIs mainly focus on its concept and importance but not on methods for detection. Cheng et al. classified some COI types by their features (co-author, colleague, advisor-advisee and competitor relationship) and emphasized the effects of COIs in paper assignment process [3]. A semantic COI detection application introduced in [1] assesses COI level in collaboration and social hybrid network where edges are directly assigned weights based on rules. Note that this approach is inflexible and ignores the topological structure of the whole graph. Our work is also orthogonal to topic-based reviewer suggestion [2, 3] and PISTIS can be easily integrated with such framework. More importantly, *to the best of our knowledge, a COI detection and declaration platform has not been demonstrated in a major venue.*

5 DEMONSTRATION OBJECTIVES

PISTIS (http://degroup.cis.umac.mo/coi/) is implemented in Node.js and HTML5. Our demonstration will be loaded with a few real datasets (e.g., DBLP, AMiner, and ResearchGate). These datasets generate a graph of 670k edges, which provides different types of relationships such as advisor-advisee, collaborations, follower-followee, spatial closeness, colleague, conference organizing committee, etc. The reviewer set R and the author set A are extracted from some prestigious conferences, e.g., SIGMOD'17, KDD'17, VLDB'17 and SIGIR'17. The audience can choose one conference and play with the role of an author ($\in A$), a reviewer ($\in R$), or a program chair, through our GUI.

The key goal of the demonstration is to experience our interactive declaration system. Through the *graphical declaration panel*, the audience can explore the latent COI cases using the graph expansion feature of our system. The *meta-path information panel* will show details of the relationships. In addition, audience members can input their name as the user and check the *COI ranking list*. During the demonstration session, we will invite the audience to declare their COIs as our ground truth dataset. In the second demonstration session, we will compare the untrained and trained COI ranking lists to demonstrate the effectiveness of the learning module.

Acknowledgements This work was supported in part by MYRG-2016-00182-FST from UMAC RC, 61502548 from NSFC and NSF IIS-1762268.

REFERENCES

[1] Boanerges Aleman-Meza, Meenakshi Nagarajan, Cartic Ramakrishnan, Li Ding, Pranam Kolari, Amit P. Sheth, Ismailcem Budak Arpinar, Anupam Joshi, and Tim Finin. 2006. Semantic analytics on social networks: experiences in addressing the problem of conflict of interest detection. In *WWW*. 407–416.

[2] Ngai Meng Kou, Leong Hou U, Nikos Mamoulis, and Zhiguo Gong. 2015. Weighted Coverage based Reviewer Assignment. In *SIGMOD*. 2031–2046.

[3] Cheng Long, Raymond Chi-Wing Wong, Yu Peng, and Liangliang Ye. 2013. On Good and Fair Paper-Reviewer Assignment. In *ICDM*. 1145–1150.

[4] Yizhou Sun, Jiawei Han, Xifeng Yan, Philip S. Yu, and Tianyi Wu. 2011. PathSim: Meta Path-Based Top-K Similarity Search in Heterogeneous Information Networks. *PVLDB* 4, 11 (2011), 992–1003.

[5] Chi Wang, Jiawei Han, Yuntao Jia, Jie Tang, Duo Zhang, Yintao Yu, and Jingyi Guo. 2010. Mining advisor-advisee relationships from research publication networks. In *KDD*. 203–212.

[6] Siyuan Wu, Leong Hou U, Sourav S. Bhowmick, and Wolfgang Gatterbauer. 2017. Conflict of Interest Declaration and Detection System in Heterogeneous Networks. In *CIKM*. 2383–2386.