Data-driven PC-chair-in-the-loop Formation of Program Committees: An EDBT 2023 Experience

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

Katja Hose
TU Wien
Austria
katja.hose@tuwien.ac.at

ABSTRACT

The formation of a quality program committee (PC) for a conference venue is critical for a high-quality scientific program. Traditionally, PC chairs take a "manual" approach to form a PC. In practice, however, such an approach, might not create a diverse PC w.r.t. certain dimensions. Furthermore, it has been reported that the traditional manual approach may lead to dense co-authorship networks among PC members. All these aspects can easily make it challenging in practice to ensure fair and quality assignments of reviewers to submissions. In this article, we share our experiences and results of installing a novel *data-driven PC-chair-in-the-loop* PC formation framework for EDBT 2023 to mitigate some of the challenges brought by traditional PC formation methods.

1. INTRODUCTION

Program committee (PC) members of a conference venue play a pivotal role in ensuring quality and fairness of the review process. A high-quality review process does not only facilitate a high-quality scientific program but also enhances trust of a venue among community members by providing constructive feedback on scientific work that benefits authors in their scientific endeavours. PC chairs often strive to form PCs to serve this intended purpose. With the increasing number of submissions in recent years, however, sizes of PCs in major conferences can easily be in the hundreds, making quality PC formation a challenging task.

The traditional approach of PC formation for a venue is "manual" in nature. PC chairs typically invite candidate reviewers independent of other candidates primarily based on the recommendations from meta-reviewers and themselves, and lists of PC members in recent venues (*e.g.,* PC members in the last two editions of SIGMOD). Although they may invest efforts to ensure high coverage of all topics of interest (based on publication profiles of candidates) as well as diversify PC members along various dimensions (*e.g.,* gender, location, experience), these are largely ad-hoc and manual in nature. Consequently, even if the initial list of candidate reviewers is diverse with adequate coverage, the final set of PC members may not necessarily be so due to declination of invitations from many candidates during PC formation. Furthermore, a recent report [3] revealed that such a traditional approach of PC formation often leads to dense co-authorship networks between reviewers that may make fair assignments of reviewers to submissions challenging.

In this article, we report our experiences as PC chairs in installing a novel *data-driven PC chair-in-the-loop* PC formation framework for EDBT 2023 [1] to address some of the challenges of the traditional approach. Specifically, we undertook a data-driven, iterative approach to carefully select a set of PC members that is diverse w.r.t. multiple dimensions, adequately covers the topics of a venue, but forms a *sparse* co-authorship network with a low average clustering coefficient, a smaller giant component (*i.e.,* largest connected component), and low average and maximum degree. We observe that the formed PC not only exhibits features that are closer to the *PC selection criteria* of EDBT 2023 but also several quantitative measures related to the review process quality show promising results. Nevertheless, we do not claim that the our framework has a causal relationship with the review process quality. This article aims to nudge future PC chairs to adopt and explore the nexus between the proposed framework and quality review process.

The rest of this article is organized as follows. In Section 2, we introduce the criteria we have adopted for selecting PC members for EDBT 2023. We describe the novel framework for PC formation based on these criteria in Section 3. Section 4 reports the impact of the formed PC on the review process. The last section concludes the article. Note that the majority of the content reported here was presented during the *EDBT 2023 Opening Session.*

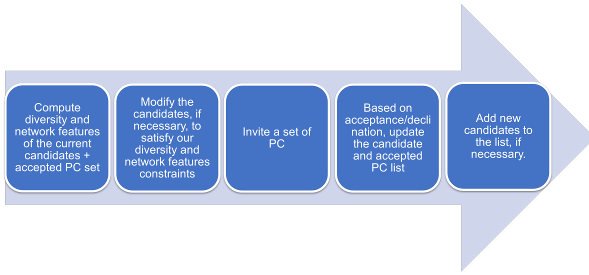

Figure 1: Framework for PC formation.

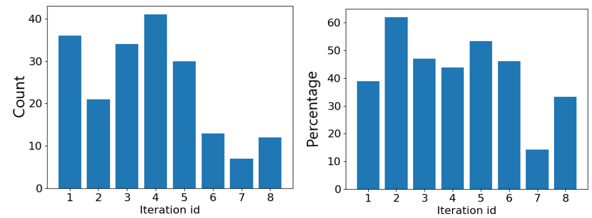

Figure 2: The number of candidates invited per iteration (left); acceptance rates of the candidates (right).

2. CRITERIA FOR PC MEMBERS

There are many ways PC chairs may form a PC. We adopted the following criteria to select high-quality and diverse PC members for EDBT 2023:

1. The PC members should adequately cover all topics of interest of EDBT 2023. The coverage of topics does not have to be uniformly distributed. Popular topics typically attract a higher number of submissions and hence a larger pool of experts needs to be recruited on these topics compared to less popular ones to balance reviewer workload and assign reviewers with the respective expertise.
2. All PC members should have prior experience in publishing their research in SIGMOD, VLDB, or EDBT. We use the past publication record of an individual in these venues as a proxy for ensuring high scientific quality of the PC.
3. The PC should have a good balance of "senior" and "junior" researchers. By "senior", we refer to researchers who are tenured faculty members or have been active in research for more than 8 years since their doctoral degree. "Junior" researchers are typically tenure-track faculty, post docs, or researchers with 8 or less years of research experience. This balance is essential as it enables us to inject sufficient experience in the review process while at the same time provide junior researchers with an opportunity to serve as reviewers.
4. Since a diverse group of individuals tends to surface different perspectives, the PC should have good geographic, institutional, and country of origin diversity. That is, no particular group should be overly represented in the PC.
5. It is desirable for the PC to form a *sparse* co-authorship network instead of a dense one. Not only do dense co-authorship networks among PC members make unbiased assignment of reviewers to submissions challenging, but they might also increase the likelihood of unethical reviewing practices such as collusion [3].
6. PC members should not be serving as PC chairs of other major venues during the review period of EDBT 2023[1]. This is to mitigate the adverse impact of competing service workload on the review process.

Observe that Criteria 1, 3, and 6 are usually considered during PC formation in major venues. However, Criteria 2 may not be adopted strictly. For example, the VLDB 2023 (resp. SIGMOD 2023) PC contained at least 9 (resp. 11) members who have not published in these two venues at the start of the review process. Although some degree of geographic and institutional diversity is considered by existing venues, country of origin (Criteria 4) is often ignored, leading to over-representation of certain groups (detailed in Section 3.3). Lastly, Criterion 5 is typically not considered by major data management venues.

3. FRAMEWORK FOR PC FORMATION

In this section, we describe our data-driven PC-chair-in-the-loop framework for PC formation.

3.1 Approach

First, we gathered recommendations from the 13 Senior Program Committee (SPC) members of EDBT 2023. This list contained 87 recommendations. Instead of inviting all of them for PC directly, we pruned the list based on Criteria 2 and 6 first. This resulted in a final list of 75 candidate PC members. Note that the check for Criterion 2 is undertaken automatically by leveraging CLOSET [2, 4]. Since the number of candidates is lower than the desired size of the PC (*i.e.,* 85-90), we iteratively added additional candidates satisfying the two criteria as discussed below. These candidates are retrieved from the reviewer database of CLOSET, which contains details of PC members in major venues in the last five years. We also exploit CLOS-

[1]Although this criterion should also exclude PC members who are serving in multiple venues concurrently, it is hard to implement it in practice as the PC lists of these venues may not be publicly available during the time of EDBT PC formation.

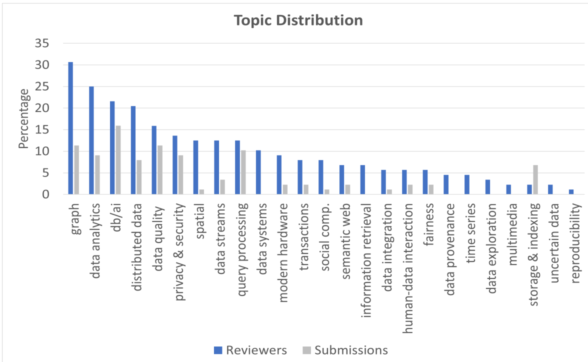

Figure 3: Distribution of expertise of reviewers and submissions. The Y-axis shows the percentage of reviewers in the PC (resp. submissions in Cycle 3) for a specific topic.

ET's capability of searching for candidates in DBLP based on topics and publication profile.

Each candidate is associated with the following attributes: *name*, *email*, *institution*, *country of residence*, *country of tertiary/secondary education (COE)*, *seniority*, *expertise area*, *DBLP*, *iteration*, and *decision*. The attribute COE represents the country where a candidate undertook his or her high school or undergraduate education. Such information is often available in a candidate's homepage or *LinkedIn* page. We use it as a proxy for country of origin since the latter may not be publicly available. *Seniority* takes one of the following two values: senior or junior (Criterion 3). The *DBLP* attribute records the DBLP URL of a candidate. The *iteration* attribute records the *iteration id* when a candidate is invited (detailed below) and *decision* captures whether a candidate accepted or declined the PC invitation or did not respond.

Next, we formed the PC *iteratively* until the desired number of PC members was attained by adopting the following steps (Figure 1). We set the desired PC size to be between 85-90 given the expected submission numbers to EDBT. In each iteration, we undertook the following steps:

1. Compute research area, level, diversity distributions, and network features (Criteria 1, 3 – 5) of the list of candidates and accepted PC members by leveraging CLOSET. We added or removed candidates to ensure that the aforementioned criteria are satisfied.
2. Send PC invitation to the refined candidate set from Step 1 and give them 10-14 days to respond. We record their decisions and update the *decision* attribute of the candidate list.
3. Add new candidates satisfying Criteria 2 and 6 to the list until we have reached the desired PC size.
4. Repeat Steps 1-3.

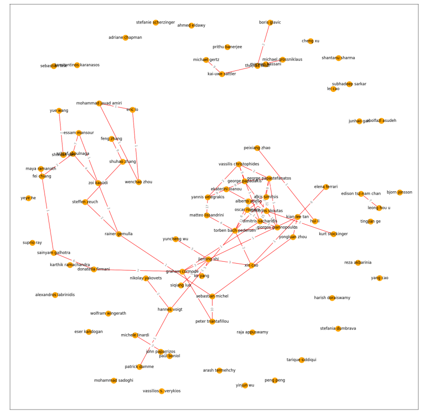

Figure 4: The EDBT 2023 reviewer network.

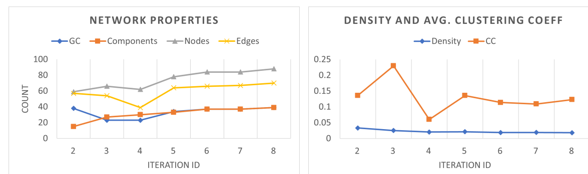

Figure 5: Network features vs iteration id.

We took 8 iterations spanning over a period of 3.5 months to recruit 88 PC members. Figure 2 depicts the number of candidates invited per iteration as well as acceptance rates of the candidates. In total, we invited 194 candidates with an overall acceptance rate of 45.4%.

3.2 Looking into the Data

In every iteration of the PC formation process, we generated various distributions and statistics (related to Criteria 1, 3–5) of the currently accepted list of PC members, invited candidates who have not yet responded, and candidates whom we intend to send invitations to. We updated the candidate list for invitations, if necessary, based on these distributions and statistics so that the aforementioned criteria are satisfactory in each iteration. This enabled us to form the PC iteratively in a data-driven manner. In this subsection, we first present data to reveal this iterative process. In the next subsection, we shall compare some of these quantitative characteristics of our PC with the PCs of recent major data management venues that are formed using the traditional approach.

Our foremost criterion was to ensure that the PC adequately covers the topics of interest of EDBT 2023 (Criterion 1). Figure 3 plots the distributions of expertise at iteration 8 (final) and the primary

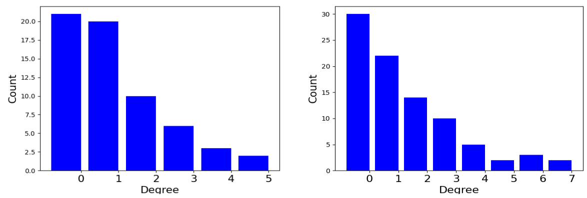

Figure 6: Degree distributions of the reviewer networks in iterations 4 and 8.

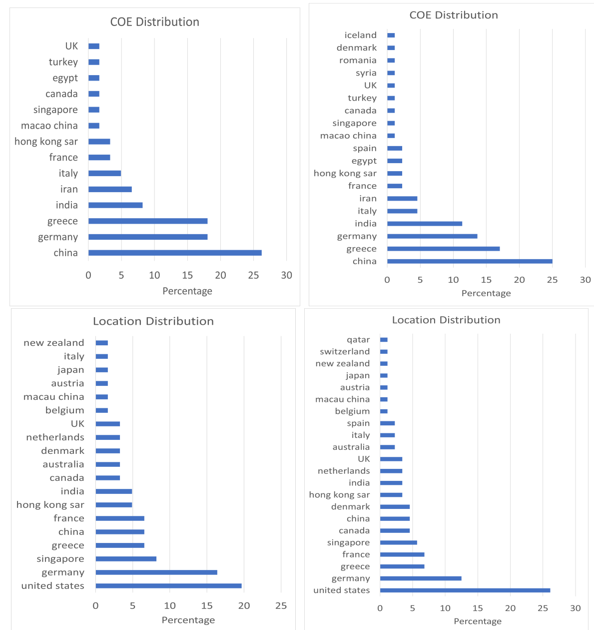

Figure 7: Distributions of COE (top) and location (bottom) in iterations 4 (left) and 8 (right).

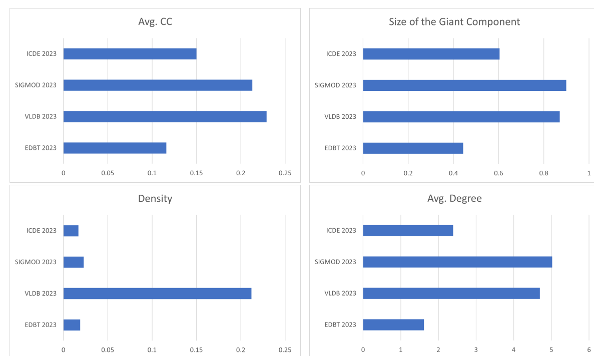

Figure 8: Comparison of reviewer networks.

subject areas of the submissions in Cycle 3 (as declared by the authors). Observe that the final PC covers the areas of submissions adequately. In particular, since graphs, DB & AI, and data analytics are popular topics, we recruited more PC members in these areas. Furthermore, at each iteration of PC invitations we maintained a healthy balance of junior and senior members (Criteria 3). Around 60% of the final PC members were seniors, which we believe is a good distribution.

Another important criterion during this process is the reviewer network of the PC members in each iteration. The *reviewer network* $G_i = (V_i, E_i)$ at iteration $0 < i \leq 8$ is an undirected, labeled, weighted graph where V_i is a set of (candidate) reviewers and E_i is a set of co-authorship edges between them. Given a pair of $u, v \in V_i$, $(u, v) \in E_i$ iff u and v have co-authored one or more articles. A node $u \in V_i$ is labeled with a unique identifier and $(u, v) \in E_i$ is labeled with a weight w representing the number of co-authored articles by u and v. We generate a reviewer network by exploiting the DBLP dataset of candidate reviewers. V_i comprises accepted PC members till iteration $i-1$ (denoted by V_a), invited candidates at iterations $0 < j \leq i-1$, denoted by V_u, who have not responded yet (*i.e.,* not marked as decline or accept), and candidates whom we intend to send invitations at iteration i (denoted by V_c). That is, $V_i = V_a \cup V_u \cup V_c$. Note that when $i = 1$, $V_a = V_u = \emptyset$. When $i = 8$, we generate G_8 *after* the desired size was attained by updating the *decision* attribute of unresponsive candidates to 'decline'. That is, $V_c = V_u = \emptyset$.

In line with Criterion 5, at each iteration, we ensured that the reviewer network was sparse and there were no significant hubs (PC members who have collaborated with many other members). Figure 4 depicts the final reviewer network. Note that the goal here is not visual clarity but to visually appreciate the sparseness of the network. Figure 5 reports the numbers of nodes, edges, and connected components, the size of the giant component (GC), density, and average clustering coefficient (CC) of the reviewer network at each iteration. The key observation here is that at each iteration of PC invitations we maintain a sparse network (low density and CC) and the size of the giant component is less than half of the number of nodes in most iterations. As we shall see later, these values are significantly superior to other major data management venues. Note that several network properties may not monotonically increase with i if there are declinations from candidates invited in the prior iterations.

Figure 6 plots the degree distributions of the network at two iterations. Observe that the maximum degree of a node is 7 in the final PC and many PC members do not have any co-authorship relationships with other members (*i.e.,* many isolates).

Figure 7 plots the distributions of COE and location of candidate reviewers at two iteration points. At each iteration we ensured that the PC is diverse

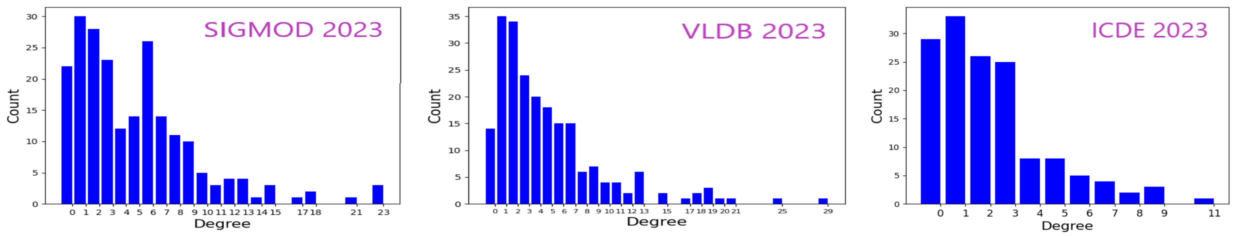

Figure 9: Degree distributions of reviewer networks.

and no single group dominates significantly. In particular, the maximum size of any group does not exceed 30% of the PC for these measures.

3.3 Comparison

In the following, we compare the reviewer network and PC properties of EDBT 2023 with the 2023 editions of SIGMOD, VLDB, and ICDE whose PC formation follows the traditional approach. Reviewer networks of all venues are generated using CLOSET using the approach described in [4]. Figure 8 plots the results for average clustering coefficient (CC), the fraction of nodes in the giant component (GC), density, and average degree of the networks. Clearly, the EDBT 2023 reviewer network is sparser than these venues.

Figure 9 reports the degree distributions. Observe that the maximum degree of nodes is lowest in EDBT 2023 (Figure 6, right). Furthermore, the number of isolates (*i.e.,* nodes with 0 degree) forms the largest group size in EDBT 2023. This is not the case for any of the other venues.

Finally, we make some observations related to the country of tertiary/secondary education (COE) and location distributions. The largest group sizes for COE (resp. location) distribution for ICDE 2023, VLDB 2023, and SIGMOD 2023 involve 53% (resp. 28%), 42% (resp. 33%), and 27% (resp. 42%), respectively, of the PC members, Hence, the COE distribution of EDBT 2023 is comparable to SIGMOD 2023 and significantly less skewed compared to ICDE 2023 and VLDB 2023. Similarly, the location distribution is comparable to ICDE 2023 and less skewed compared to the other two venues. It is evident that the manual approach of PC formation may result in overrepresentation of certain groups.

In summary, the data-driven PC formation framework deployed in EDBT 2023 facilitated the formation of a PC whose features are closer to the PC selection criteria in Section 2.

4. REVIEW PROCESS ANALYSIS

In this section, we quantitatively analyse the review process of EDBT 2023 undertaken by the PC formed by the proposed framework. There were three submission cycles for EDBT 2023. As PC chairs we were in charge of the second and third cycles. The first cycle was managed by the PC chairs of EDBT 2022. We received 29 and 88 submissions for the second and third cycles, respectively. We deployed a bidding-based automated reviewer assignment process for submissions hosted on Microsoft's CMT. We utilized CLOSET [4] to manage submissions and reviewer assignments that violate EDBT 2023 COI policy. Each reviewer was assigned a total of 5-6 submissions. Each senior PC (*i.e.,* meta-reviewer) managed 8-11 submissions. Each submission was assigned 5 reviewers and at most 4 reviewers in the second and third cycles, respectively.

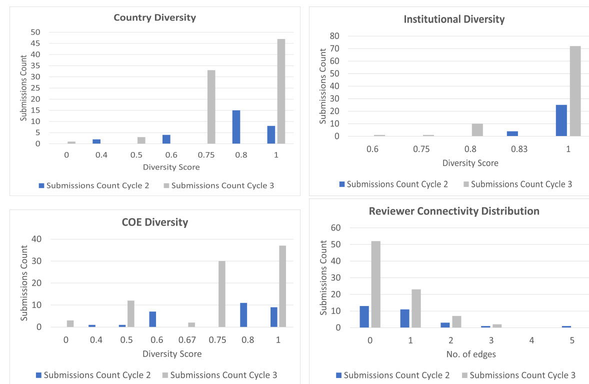

Figure 10: (top-left, top-right, bottom-left) Diversity of reviewer assignments to submissions for the two cycles. (bottom-right) Reviewer connectivity distribution for the two cycles.

Diversity of Reviewer Assignments. Since the reviewer assignment technique in CMT is opaque to end users, we could not tinker with it. Instead, we diversified the input (*i.e.,* reviewers) through our PC formation framework so that the assignments could be diverse. Figure 10 plots the results (diversity score vs number of submissions). The higher the diversity score the greater is the diversity of the reviewer assignment. For instance, COE diversity score of 1 (resp. 0) indicates all reviewers for a submission have distinct (resp. same) COE. Observe that the diversity of assignments is good across all

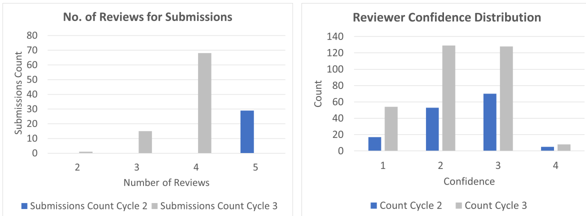

Figure 11: Quality of the review process: timeliness (left), reviewer confidence (right).

the three dimensions with very few submissions, if any, have low diversity (< 0.5).

Reviewer connectivity. Next, we report the impact of maintaining a sparse reviewer network on the reviewer assignment. We compute the number of edges between the set of reviewers in each submission. Then, we compute the distribution of the number of submissions with reviewers having coauthorship relationships. We observe that 44.8% and 62% submissions have no edges between reviewers (*i.e.,* isolated nodes) in the second and third cycles, respectively. Figure 10 (bottom-right) reports the distributions for submissions. Observe that for the majority of these cases only one or two edges exist among reviewers. That is, the sparseness of the reviewer network facilitates assignment of reviewers to submissions with very low connectivity *without* tinkering CMT's assignment algorithm.

Quality of the review process. Lastly, we report on the quality of the review process. We quantify it by measuring the following three dimensions: *review timeliness*, *author complaints*, and *reviewer confidence*[2]. We measure the timeliness by computing the number of reviews received per submission before the start of the author feedback phase. Figure 11 (left) reports the numbers. We received all five reviews for each submission on time in the second cycle. For the third cycle, except for one submission, all had at least three reviews prior to the deadline (maximum is four reviews) In particular, the missing reviews involved only 7 reviewers. That is, 92% of the PC completed their reviews before the author feedback phase. This is remarkable given that PC chairs have frequently lamented on late or missing reviews in major venues. For example, in the first three submission cycles of SIGMOD 2024, only about 20% of submissions had all three reviews by the review deadline [5].

Next, we report the number of complaints from authors regarding review quality and decision making process after the notification of results. For both the cycles, we did not receive any such complaints from the authors.

Lastly, we measure *reviewer confidence*, *i.e.,* how confident reviewers were of their reviews w.r.t. the contents of submissions. To this end, in the review form we have an item on reviewer confidence. Specifically, a reviewer must select one of the followings w.r.t. a submission he or she reviewed:

1. *"I am certain that my evaluation is correct and I have worked on this problem/main topic."*
2. *"I am willing to defend my evaluation and have a solid overview of the state of the art on the main topic of the paper."*
3. *"I am willing to defend my evaluation but am lacking a detailed knowledge of the state of the art on the main topic of the paper."*
4. *"I learnt the topic as I reviewed the paper."*

Note that *Item 4* indicates lack of confidence of a reviewer on his or her review. Figure 11 (right) plots the results. Observe that very few submissions have a review where the reviewer has selected item 4. Overall, the PC of EDBT 2023 is confident of their reviews, which is a cornerstone for any quality review process.

5. CONCLUSIONS

In this paper, we report our experiences with the novel data-driven PC-chair-in-the-loop PC formation framework that we adopted for EDBT 2023. We depart from the traditional approach of PC formation by iteratively guiding the selection of candidate reviewers using various diversity, publication profile, and collaboration network-related data. Our experiences as well as data related to the review process convince us that such a data-driven approach may contribute to a high-quality review process. We emphasize that we neither claim that the proposed framework has a causal relationship with the quality of the review process nor the review quality is superior to other venues. In particular, the latter issue demands access to the private review data of different venues which is unavailable to us.

We hope that this article will inspire future PC chairs to adopt a data-driven framework for PC formation and review management and explore its impact on the review quality. We also encourage PC chairs of existing venues to undertake a comparative analysis of the aforementioned measures related to their review processes. This will provide deeper insights on the impact of the two PC formation approaches on the review process.

[2]We acknowledge review process quality is multi-faceted with many dimensions (*e.g.,* discussion quality, engagement) at play. Here we only consider a subset of them that is easily measurable.

6. REFERENCES

[1] EDBT 2023 website. http://edbticdt2023.cs.uoi.gr/.

[2] S. S. Bhowmick. CLOSET: Conflict of Interest Detection and Management System. https://personal.ntu.edu.sg/assourav/research/DARE/closet.html, 2019.

[3] S. S. Bhowmick. How Connected Are Our Conference Review Boards? *SIGMOD Rec.*, 51(4): 74-78, Dec 2022.

[4] S. S. Bhowmick. CLOSET: Data-Driven COI Detection and Management in Peer-Review Venues. *Commun. ACM*, 66(7): 70-71, July 2023.

[5] A. Meliou, S. S. Bhowmick, K. Aberer, D. Agrawal, A. Bonifati, V. Braganholo, F. Geerts, W. Lehner, D. Srivastava. Peer-Reviewing Processes and Incentives: Data Management Community Survey Results. *SIGMOD Rec.*, 52(4), Dec 2023.