SIGIR Community Survey on Preprint Services

Diane Kelly, SIGIR Chair
University of Tennessee, USA
dianek@utk.edu

Abstract

Recently, there has been a growing number of questions about the relationship between preprints and double-blind peer-review at the SIGIR conference. These questions come from authors who wish to post papers which are under review to preprint servers such as arXiv.org, as well as reviewers who become aware of author identities through such postings, and are subsequently, unsure of how to proceed in a double-blind review process. A review of current conference publication guidelines was conducted, along with a survey of SIGIR community members to gain insight about their behaviors, practices and opinions. The paper presents results of this survey, as well as recommendations about submission and review policies for the SIGIR conference.

1   Introduction

Whether or not authors should be allowed to post preprints of articles while these articles are under review at a venue that practices double-blind review has been a recent topic of discussion amongst conference leaders, paper reviewers and authors, and SIGIR leadership. It is becoming more common for reviewers to point to the existence of a preprint and ask for guidance about how to proceed, or to decidedly reject the submission for not following paper submission guidelines which specify that papers should be anonymous. One of the main purposes for anonymity, of course, is to facilitate blind-review. Double-blind review, in particular, has been shown to result in fairer evaluations, including in a recent study conducted at WSDM 2017 [11].

A *preprint* is a draft of a research paper before peer review [5]. The first preprint server and one of the most popular is arXiv (arXiv.org), which was established in 1991 by Paul Ginsparg as a way to circulate research results outside of the commercial publishing enterprise. Schöpfel and Farace [9] note that "*preprints* would not have existed without the prospect of *print,*" and that one of Ginsparg's major goals in creating arXiv was to circulate results more quickly. ArXiv was not designed to compete with commercial publishers, but was rather a way to facilitate scientific communication. Today, arXiv.org contains over 1,405,821 e-prints in physics, mathematics, computer science, and several other disciplines, and is no longer just a repository for preprints, but also for post-publication prints.

Preprints are *grey literature*, which is a type of scientific communication that occurs outside of commercial venues. Grey literature is a common part of the scholarly publishing landscape. According to Wikipedia [4]:

> *Grey literature (or gray literature) are materials and research produced by organizations outside of the traditional commercial or academic publishing and distribution channels. Common grey literature publication types include reports (annual, research, technical, project, etc.), working papers, government documents, white papers and evaluations. Organizations that produce grey literature include government departments and agencies, civil society or non-governmental organisations, academic centres and departments, and private companies and consultants.*
>
> *Grey literature may be made available to the public, or distributed privately within organizations or groups, and may lack a systematic means of distribution and collection. The standard of quality, review and production of grey literature can vary considerably. Grey literature may be difficult to discover, access, and evaluate, but this can be addressed through the formulation of sound search strategies.*

In the field of information retrieval, grey literature has always been an important part of our scholarly communication practices. Some of the most well-known ideas were initially published as reports (many of these early reports are available at the *SIGIR Museum*: http://sigir.org/resources/museum/, including reports by Cyril W. Cleverdon and Gerard Salton). Many long-standing evaluation campaign such as TREC and CLEF require participants to write reports describing their research efforts and findings, which are posted online and available for download. Institutional policies and funding agencies dictate that researchers post papers in online repositories. Thesis and dissertations are another form of grey literature that can be easily found on the Internet.

Discovering and using grey literature used to be difficult, but in today's information environments "tools for collecting, depositing, and archiving will make grey literature less ephemeral and volatile than in the past" [10, p. 2038]. The availability of grey literature is likely to only increase, as well as the community's questions about how to accommodate it within a double-blind review process. The purpose of the SIGIR community survey, which is described in this paper, is to gather feedback from community members about their behaviors, practices and opinions regarding preprints and peer-review. Rather than just hearing from those with the loudest voices, or from those with access to SIGIR leadership, this survey is intended to gather community input to inform decision-making about how SIGIR submission and reviewing policies might change in response to community needs and practices.

In the next section, the current submission guidelines related to preprints from several prominent conferences are presented, including the SIGIR submission guidelines. This is followed by a section presenting results of a survey conducted by the Association for Computational Linguistics (ACL) about preprint publishing and reviewing, as well as subsequent changes to their submission and review policies. This is followed by the results of the SIGIR community survey on preprint publishing and reviewing, as well as recommendations about submission and review policies for the SIGIR conference.

2 Overview of Current Policies

The submission policies for several major conferences, including SIGIR, are included here to provide context for the survey, results and recommendations. These policies establish starting points and also provide evidence about the extent to which other related conferences and professional societies have addressed and/or reconciled the existence of preprints and double-blind peer-review.

2.1 SIGIR Submission Guidelines

The submission guidelines for SIGIR 2018 emphasize author anonymity and also indicate that authors can upload manuscripts to preprint servers. However, the guidelines explicitly discourage this since it places anonymity at risk, which could cause a paper to be rejected,

> *Full research papers must describe work that is not previously published, not accepted for publication elsewhere, and not currently under review elsewhere (including as a short-paper submission for SIGIR 2018). Submissions should not contain any author identification and should be submitted electronically via the conference submission system. Authors are required to take all reasonable steps to preserve the anonymity of their submission. While authors can upload to institutional or other preprint repositories such as arXiv.org before reviewing is complete, we generally discourage this since it places anonymity at risk (which could result in a negative outcome of the reviewing process).* [10]

2.2 WSDM Submission Guidelines

Until recently, WSDM ran a single-blind review model, where only reviewers were anonymous, in which case preprints did not present a problem. Specifically, while WSDM 2014 used double-blind reviewing, WSDM 2015 & 2016 used single-blind. In WSDM 2017, an experiment was performed comparing single- and double-blind reviewing. The outcome of the experiment found that:

> *that single-blind reviewers ... recommended for acceptance papers from famous authors or top institutions, compared to their double-blind counterparts.* [11]

Starting in 2018, WSDM moved to a double-blind review model. The 2019 submission guidelines accommodate preprints and other grey literature by requesting that the title and abstract be sufficiently different from the WSDM submission,

> *Papers that have been published in or accepted to any peer-reviewed journal or conference/workshop with published proceedings, or that are currently under review, or that will be submitted to other meetings or publications while under review may not be submitted to WSDM 2019. However, submissions that are available online and/or have been previously presented orally or as posters in venues with no formal proceedings, are allowed. Note that if available online (e.g., via arXiv) and not anonymous, their titles and abstract must be sufficiently different from the WSDM submission in order to limit the risk that a direct search breaks the double blind reviewing procedure.* [12]

2.3 CIKM Submission Guidelines

The 2018 CIKM submission guidelines do not mention preprint servers, but do emphasize double-blind review and author anonymity,

> *CIKM 2018 full and short paper review will be double-blind, and submissions must be properly anonymized. Lack of anonymization may result in rejection without review.* [2]

2.4 KDD Submission Guidelines

KDD is structured a bit differently from SIGIR, WSDM and CIKM, in that it has two full paper tracks: *Applied Data Science (ADS) Papers* and *Research Papers*, and distinguishes between them as follows,

> *A paper for the SIGKDD research track is evaluated on its novelty and contributions in the area of data science research. So if the merit of an ADS paper does not lie in its new research insights, where does it lie? The merit of an ADS track paper lies in advancing the practice of data science, in describing lessons learning from applying data science to real problems with real users and real metrics that were moved. Thus a base requirement is that the paper describes a novel artifact that has real users with the some part of the paper describing the learnings and insights that came from users/usage of the artifact(s).* [6]

Both types of papers are peer-reviewed and have similar submission guidelines,

> *Following KDD conference tradition, reviews are not double-blind, and author names and affiliations should be listed.* [6]

Although the instructions do not explicitly say that single-blind review is used, a personal correspondence with the current KDD chair confirmed this is the case (J. Pei, personal communication, June 20, 2018).

Reference to preprints and arXiv is made in the section "Dual Submissions," where the concern is about publishing original work,

> *Submitted papers must describe work that is substantively different from work that has already been published, or accepted for publication, or submitted in parallel to other conferences or journals.*

With one exception being,

> *Submissions are permitted for papers that have previously been made available only in the form of technical report with no peer reviews, in particular on arXiv.*

This positions preprints in the context of the general concern most top-tier conference have about publishing original content, rather than compromising double-blind review since KDD does not practice this type of review. Submissions for which a preprint exists are still considered original because they have not been peer-reviewed.

2.5   RecSys Submission Guidelines

Similar to KDD, RecSys addresses preprints as they relate to original work, although they do not address the relationship between preprints and their double-blind review process,

> *The peer review process is double-blind (i.e. anonymised). This means that all submissions must not include information identifying the authors or their organisation. Specifically, do not include the authors' names and affiliations, anonymise citations to your previous work and avoid providing any other information that would allow to identify the authors, such as acknowledgments and funding.*
>
> *Submitted work should be original. Simultaneous submissions to other conferences or journals is explicitly prohibited by ACM policies. However, technical reports or ArXiv disclosure prior to or simultaneous with RecSys submission, is allowed, provided they are not peer-reviewed.* [8]

A unique feature of the RecSys submission guidelines is the invocation of the "SIGCHI Submitter Agreement" (RecSys is sponsored by SIGCHI), which is an implicit agreement made by authors when they submit papers. This agreement has four points including a declaration that (1) the submission is the work of those submitting the paper; (2) copyright is held for the content (or has been obtained); (3) submissions involving human subjects has institutional approval; and (4) if the paper is accepted, one of the authors is required to attend the conference and present the work.

3   ACL Survey

In June 2017, ACL conducted a survey to learn about preprint publishing and its impact on the review process [3]. This survey was completed by 623 respondents, most of whom were either current or previous members of ACL. Based on the results of the survey and discussion amongst association leaders and others, ACL adopted new policies on submission, review and citation [1, 7]. Nivre [7] summarized the main conclusions drawn from the survey:

1. *There is strong support for maintaining double-blind reviewing for our conferences (and for doing what we can to protect its integrity by discouraging non-anonymized preprint versions of submissions).*
2. *There is only weak support for completely banning preprint versions of conference submissions (although a majority say they are prepared to live with it).*
3. *Many people think the policies and guidelines for submitting, reviewing and citing preprint papers need to be clarified.*

The updated ACL submission, review and citation policies still emphasize the requirement that submissions be anonymous and address the issue of preprints, "A submission will not be considered anonymized if the authors post (or update) a non-anonymized preprint version within an anonymity period lasting from 1 month before the submission deadline until the time of notification (or withdrawal). Submissions will be rejected if not properly anonymized" [1]. The policy further states that anonymous preprints within the anonymity period are allowed and that non-anonymized preprints before the anonymity period starts are also allowed, but authors must declare their existence at the time of submission.

Reviewers are instructed "not to actively try to discover the identity of the authors" and postpone search for related work until "after they have read the paper completely and formed an initial impression" [1]. If the reviewer uncovers the identity of the authors, they are instructed to inform the area chair, but not other reviewers.

The updated policy also provides guidance to authors about expected citation behavior. Refereed publications take priority over preprints and while authors are expected to cite all relevant literature regardless of publication venue, "they may be excused from not knowing about all unpublished work" [1]. Furthermore, reviewers are instructed to point authors to relevant preprints and authors are expected to incorporate these into the camera-ready version of the paper.

In preparing to conduct our own community survey about these issues, we contacted ACL leadership and asked for their survey questions, as well as any suggestions for improving the questions. We are grateful to Jennifer Foster, Marti Hearst, Joakim Nivre and Shiqi Zhao for sharing the survey questions, and feedback about the design of the survey.

4   SIGIR Community Survey

4.1   Respondents

The survey opened on January 16, 2018 and closed on February 2, 2018. The survey was advertised on the SIGIR email list, as well as through social media. A total of 174 responses were received. Of these responses, 159 were complete. All responses are included in this analysis, regardless of whether the survey was completed. Because of attrition and because respondents were not required to answer all questions, the number of responses to each question varies.

Most respondents are current members of SIGIR (n=124) and the majority of respondents identified as male (Figures 1 and 2). The data published in the *Letter from the Chair* in this issue of *SIGIR Forum* show that as of March 2018, there were 881 members of SIGIR, including 551 (62%) males, 113 (13%) females and 217 (25%) who did not provide a response. Thus, about 14% of SIGIR members are represented in this dataset; 77% of member respondents are male, 16% are female and 7% did not answer, which has both more male and female representation. There was a similar number of respondents based in Europe/Africa/Middle East and North/Central/South America, with the Asia/Pacific region not far behind (Figure 3). Of the 174 respondents, only 15 indicated that they had taken the previous ACL survey about pre-prints.
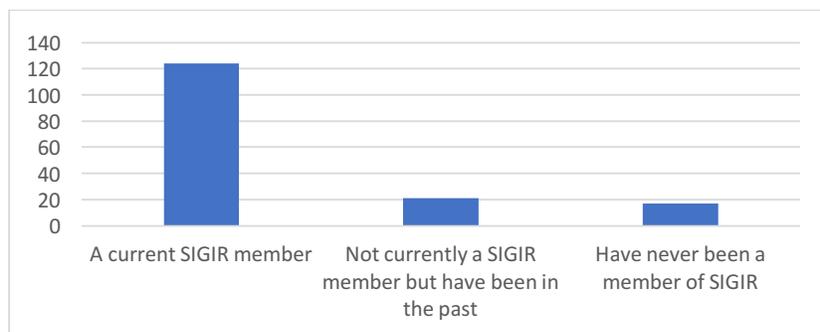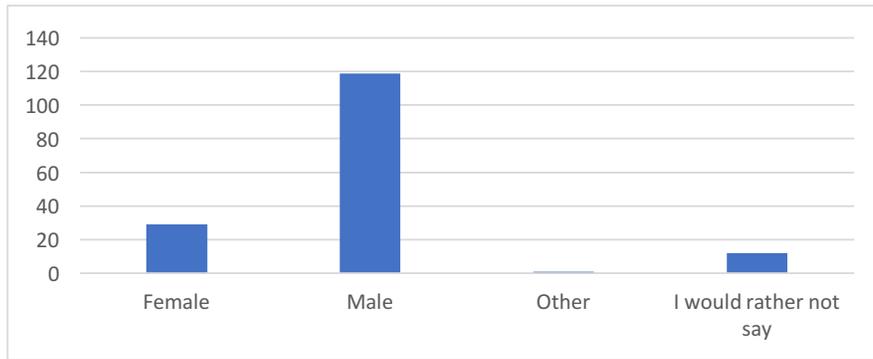
Figure 2: Respondents' descriptions of themselves when given the options female, male, other, or rather not say.
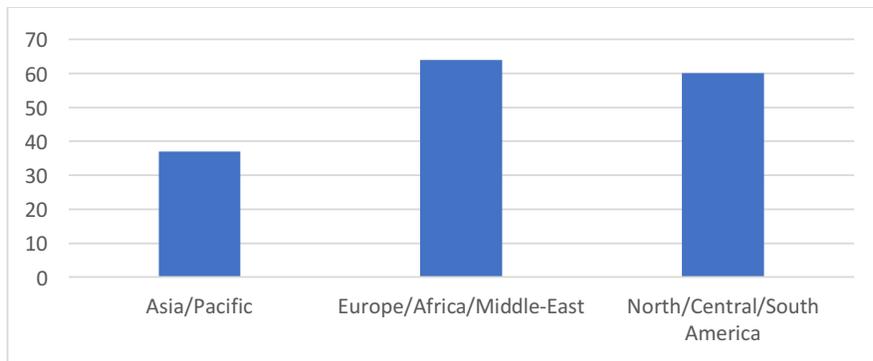

Figure 3: Geographic region in which respondents are based.

Respondents were asked to describe their work roles by selecting one or more job titles (Figure 4). Seventy-seven respondents selected professor/lecturer, 45 selected researchers in academia, 31 selected researcher in industry, 27 selected graduated student and 6 selected other. When asked how long they had been conducting research in the area of information retrieval, most respondents indicated 6-10 years, although a range of experiences was observed (Figure 5). When asked how well-known their colleagues would consider their research within their subfield, 14 respondents said extremely well-known, 41 said very well-known, 78 said somewhat well-known, 21 said not particularly well-known and 6 said unknown (Figure 6). When asked about the notoriety of their institutions, the majority of respondents indicated their institutions were very well-known (Figure 7).
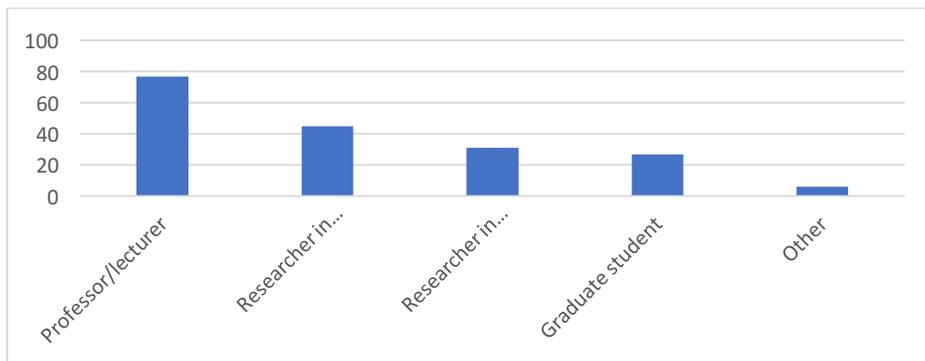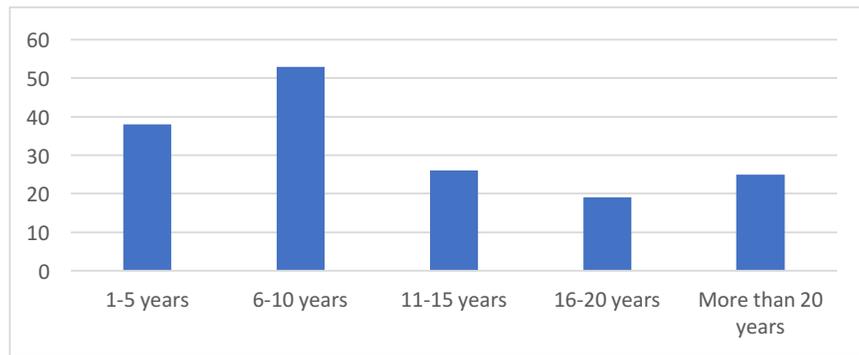
Figure 5: Length of time respondents have been conducting research in information retrieval.
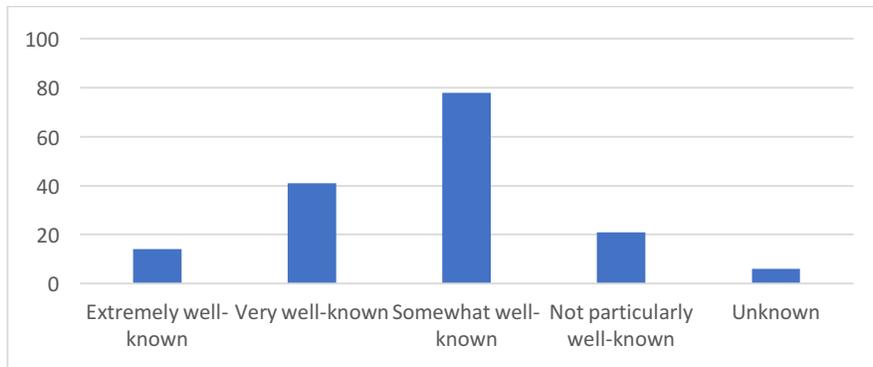

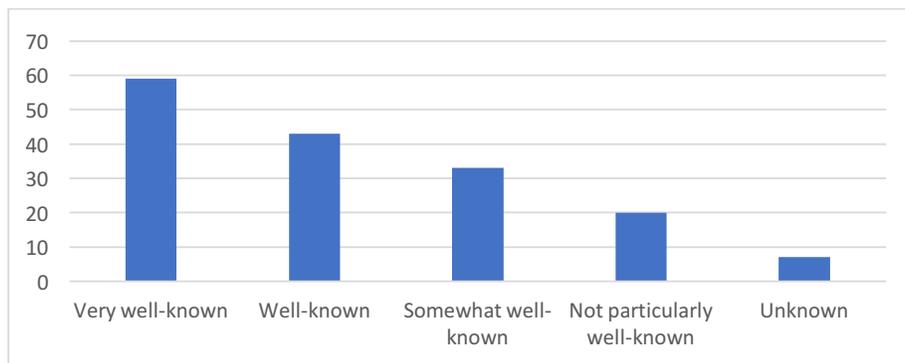**Figure 6: Respondents' perceptions of the notoriety of their research.**

Figure 7: Respondents' perceptions of notoriety of their institution.

Most respondents have published papers at several different SIGIR sponsored and co-sponsored conferences (Figure 8). Of the respondents who completed the entire survey, only 14 had not published a paper at one of these conferences. Most respondents (n=143) indicated that they had reviewed for a SIGIR sponsored or co-sponsored conference (Figure 9).
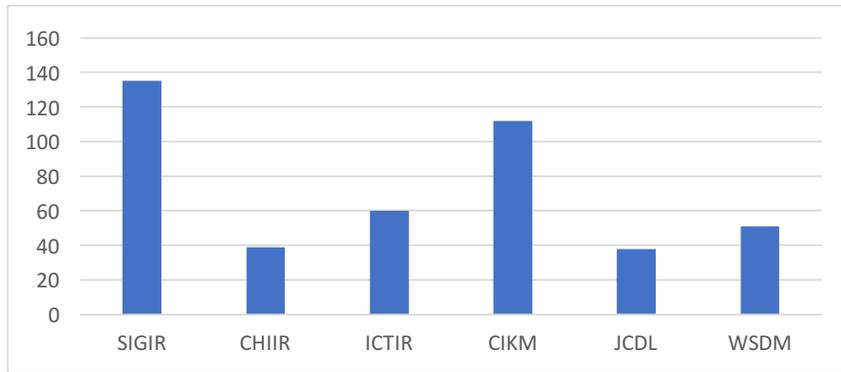
Figure 8: SIGIR sponsored or co-sponsored conferences where respondents have published.
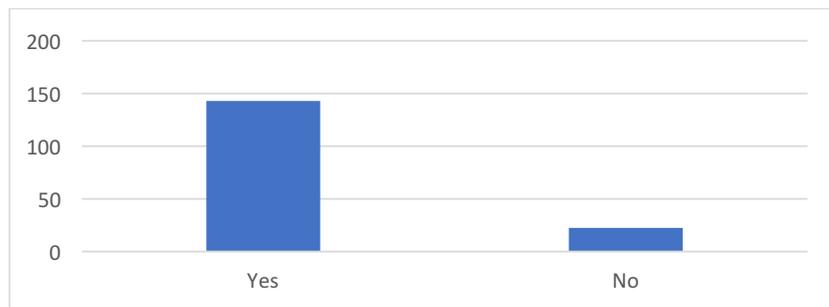

Figure 9: Number of respondents who have reviewed papers at a SIGIR sponsored or co-sponsored conference.

4.2 Preprint Services: Behaviors, Practices and Opinions

Figure 10 shows the frequency with which respondents indicated they upload their research papers to preprint servers. Most respondents indicated they "Seldom" or "Never" upload their research papers to preprint services (n=113)
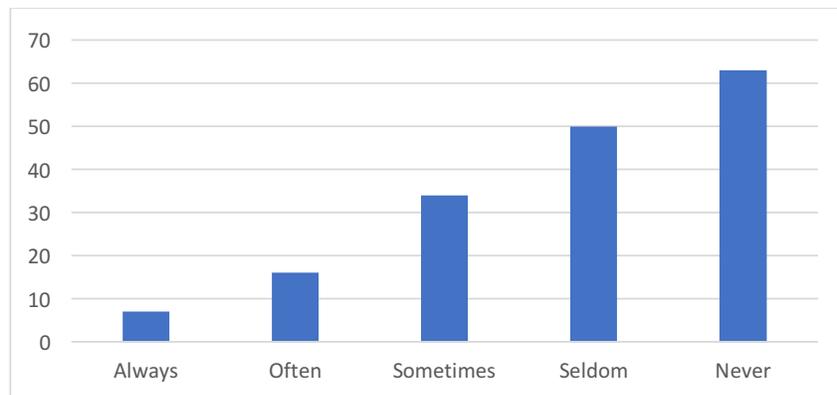

Figure 10: Frequency with which respondents upload their research papers to preprint services.

Those who indicated "Seldom" or "Never," were asked a follow-up question about why they tend not to upload their papers to preprint servers. Respondents were presented with a fixed set of options

based on the ACL survey, including an "other" option. Respondents could select as many reasons as they wanted. The two options that were selected with the greatest frequency were "I do not see the need when I intend to publish my papers at a conference or in a journal" and "I want to preserve the integrity of double-blind reviewing." Some of the more popular reasons given by those who selected "Other" included not wanting to influence the peer-review process and not wanting to publish research that might be incomplete or flawed.

#	Answer	%	Count
1	I was not aware of the possibility before now	2.88%	8
2	I haven't got a chance to try it yet	6.83%	19
3	I do not see the need when I intend to publish my papers at a conference or in a journal	25.54%	71
4	I do not want to expose my research to criticism too early in the research process	3.60%	10
5	I do not want to give away ideas too early in the research process	14.39%	40
6	I am not permitted by my employer or funding agency to make preliminary research findings publicly available	2.16%	6
7	I want to preserve the integrity of double-blind reviewing	26.26%	73
8	I do not want to split citation counts over multiple versions	12.59%	35
9	Other (please specify)	5.76%	16

Table 1: Reasons given by respondents who indicated they seldom or never upload their papers to preprint services.

Those who indicated that they "Always," "Often," or "Sometimes" upload their papers to preprint servers (n=57) were asked several follow-up questions. The first was about the stage at which they typically uploaded papers to preprint servers (Table 2). Nearly 30% of these respondents indicated, "After the paper has been accepted to publication." Ten-percent indicated, "As soon as a first draft is ready, before submitting to a conference," and 14% indicated, "After the paper has been rejected from one or more conferences."

#	Answer	%	Count
1	As soon as a first draft is ready, before submitting to a conference	10.53%	6
2	After I submit the paper to a conference, but before notification	10.53%	6
3	After the paper has been accepted to publication	29.82%	17
4	After the camera-ready version of the paper has been submitted	15.79%	9
5	After the paper has been officially published	1.75%	1
6	After the paper has been rejected from one or more conferences	14.04%	8
7	It depends on the paper whether I upload to a preprint server before or after notification (e.g., I might upload my paper when another paper on a similar topic/method appears online)	12.28%	7
8	I only upload papers to a preprint server if I do not intend to publish them elsewhere	1.75%	1
9	I only upload papers to a preprint server when my co-authors want to	0.00%	0
10	Other (please specify)	3.51%	2

Table 2: Stage at which respondents who indicated they always, often or sometimes upload their papers to preprint services.

Respondents who selected responses #1, #2, #6 or #7 were further asked why they put a paper on a preprint server before it has been accepted for publication. Table 3 shows these responses. The most popular were "To timestamp the ideas in the paper," and "To publicize my research as soon as I think it is ready."

#	Answer	%	Count
1	To publicize my research as soon as I think it is ready	19.57%	18
2	To get feedback on the research as early as possible	10.87%	10
3	To timestamp the ideas in the paper	23.91%	22
4	To maximize the paper's citation count	14.13%	13
5	To allow the evolution of the research to be tracked	9.78%	9
6	Because papers on very similar topics have been recently uploaded	7.61%	7
7	Because I did not want to wait until the next conference cycle after the paper was rejected from a conference	14.13%	13

Table 3: Reasons given by respondents who selected responses #1, #2, #6 or #7 from Table 2.

Respondents who selected responses #3, #4 or #5 from Table 2, which capture the stages after a paper has been accepted for publication, were further asked why they wait until after the paper has been accepted before putting it on a preprint server. Table 4 shows these responses; the three main choices were selected roughly equal number of times.

#	Answer	%	Count
1	I want my work to be reviewed anonymously	32.08%	17
2	I want to upload a mature version of the paper	30.19%	16
3	I want to preserve the integrity of double-blind reviewing	33.96%	18
4	Other (please specify)	3.77%	2

Table 4: Reasons given by respondents who selected responses #3, #4, or #5 from Table 2.

Respondents who indicated they upload their papers to preprint services "Always," "Often," or "Sometimes" (Figure 10) were also asked under what circumstances they update the content of the preprint. Table 4 shows the responses. The most common reasons for updating the content of a preprint paper were "If an error is discovered in the paper," and "When the camera-ready version is ready."

#	Answer	%	Count
1	If an error is discovered in the paper	25.77%	25
2	When I receive feedback that can be included in the paper	19.59%	19
3	When I get more results	14.43%	14
4	When the camera-ready version is ready	22.68%	22
5	I do not tend to update the content of a preprint paper	16.49%	16
6	Other (please specify)	1.03%	1

Table 5: Reasons given by respondents who indicated they always, often or sometimes upload their papers to preprint services (Figure 10) for updating the content of the preprint paper.

The next set of questions asked respondents about how frequently they use preprint services beyond as a repository for their work. Figure 11 shows that the majority of respondents do not subscribe to a preprint server news feed. Figures 12 and 13 show how often respondents read preprint papers and how often they cite preprint papers. One-hundred and twenty-six respondents indicated they read preprint papers "Very often," "Often," or "Sometimes," while 40 indicated "Seldom," or "Never." With respect to citing preprint papers, 66 respondents indicated they cite preprint papers "Very often," "Often," or "Sometimes," while 100 indicated "Seldom," or "Never."
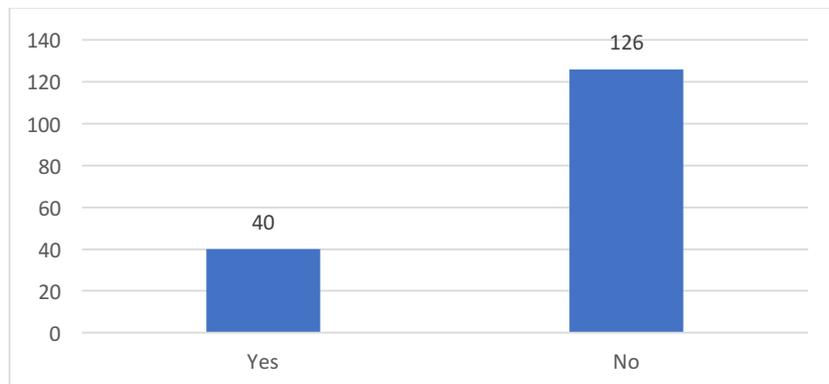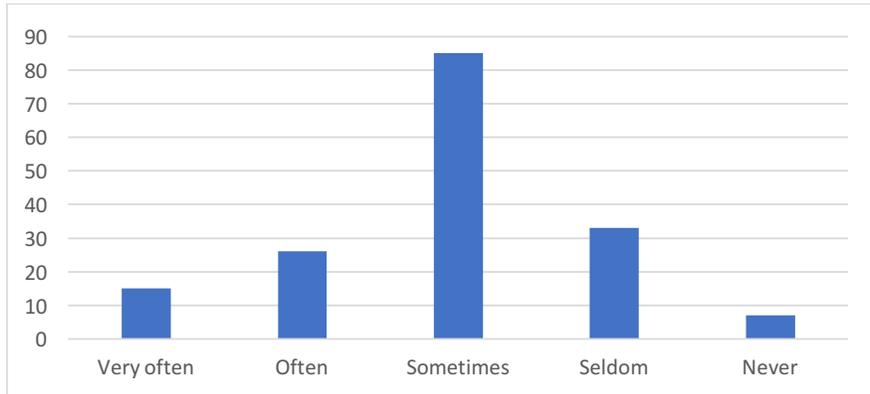
Figure 12: Frequency with which respondents read preprint papers.
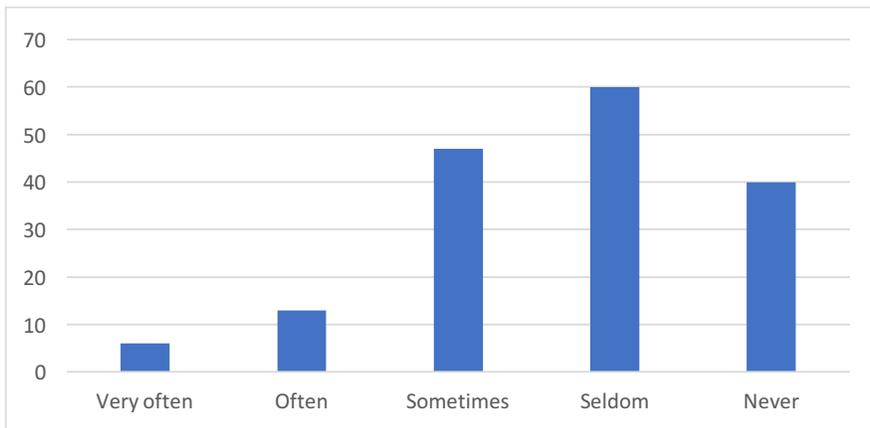

Figure 13: Frequency with which respondents cite preprint papers.

Respondents who indicated they "Seldom," or "Never" cite preprint papers were asked why not. Ninety-one of these respondents indicated "I would rather cite a published version if available," 21 indicated "I tend not to read them," and 13 selected "Other" (Table 5). The majority of respondents who selected "Other" indicated they do not cite preprints because the results are not trustworthy since they have not been peer-reviewed

#	Answer	%	Count
1	I tend not to read them	16.80%	21
2	I would rather cite a published version if available	72.80%	91
3	Other	10.40%	13

Table 5: Reasons selected by respondents who "Seldom," or "Never" cite preprint papers.

The next set of questions asked respondent about how, if at all, the existence of preprints impacts how they review papers. First, respondents were asked if they search for related work as part of the

review process when reviewing for a SIGIR or related conference. Figure 14 shows that the majority of respondents selected "Sometimes – it depends on the paper and my knowledge of the area." When asked if they have ever searched for a preprint version of a paper after an area chair (or other paper manager) has informed you that one is available, most respondents selected "Not applicable," while about 50 indicated "No" (Figure 15).
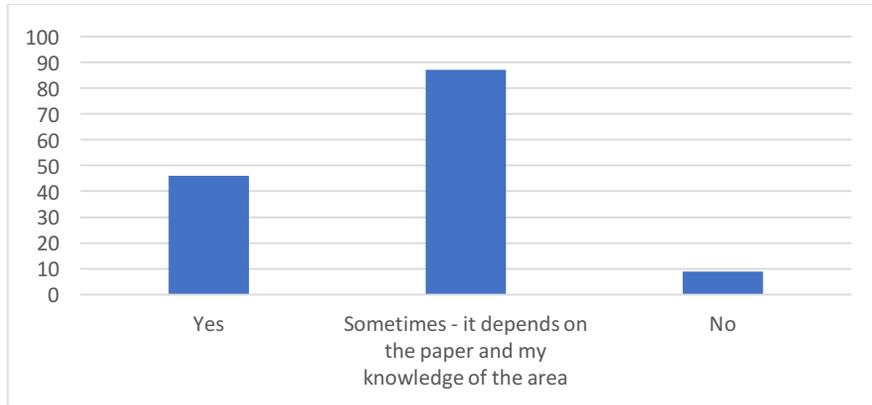

Figure 14: Responses to the question, "When you are reviewing for a SIGIR or related conference, do you search for related work as part of the review process?"
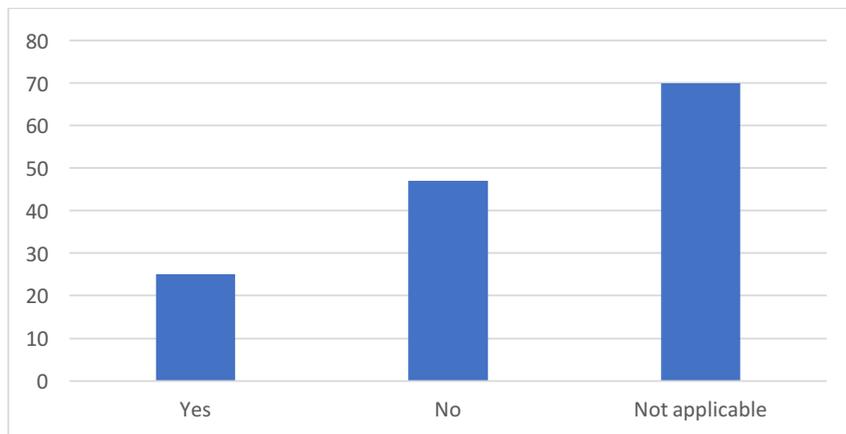

Figure 15: Responses to the question, "Have you ever searched for a preprint version of a paper after an area chair (senior reviewers, meta-reviewer) has informed you that one is available online?"

When asked if they would penalize a paper they were reviewing if they discovered it had been posted to a preprint server, most respondents indicated they would make a comment, but not adjust the scores (Figure 16). Fourteen respondents indicated that their scores would be negatively affected. The most popular reasons given by those who selected "Other" were they would notify the meta-reviewer, check the submission instructions and act accordingly. Many respondents expressed concerns about this biasing the review.
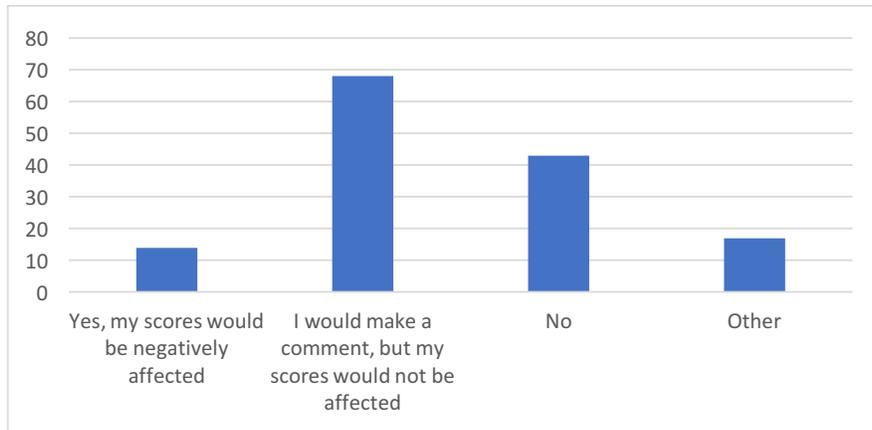
Figure 16: Responses to the question, "If you discover a preprint of a paper you are reviewing has been posted, do you (or would you) penalize the paper for this?"

When asked if they would penalize a paper they were reviewing if it did not cite a very relevant preprint of which they were aware, most respondents said either "I make a comment, but my scores are not affected" (Figure 17).
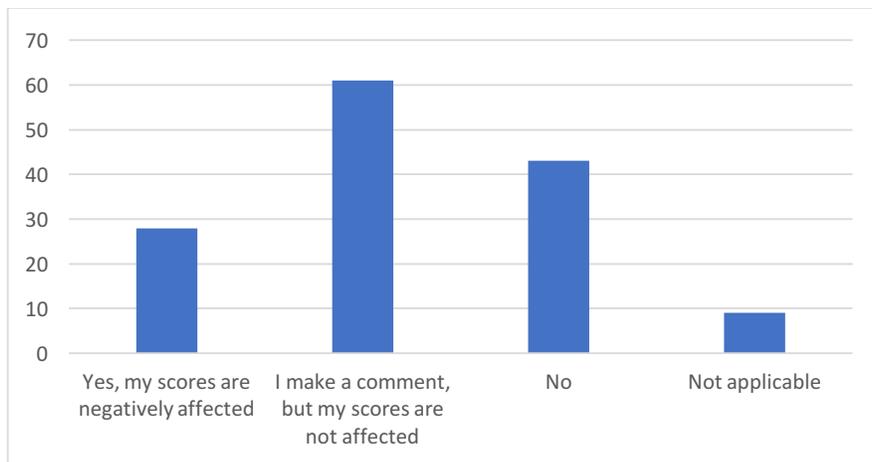

Figure 17: Responses to the question, "If a paper you are reviewing does not cite a very relevant preprint you are aware of, do you penalize the paper for this?"

When asked if they would penalize a paper they were reviewing if it did not compare with the empirical results in a very relevant preprint, most respondents selected "I make a comment, but my scores are not affected" (Figure 18). Roughly equal numbers of respondents selected "Yes, my scores are negatively affected," or "No."
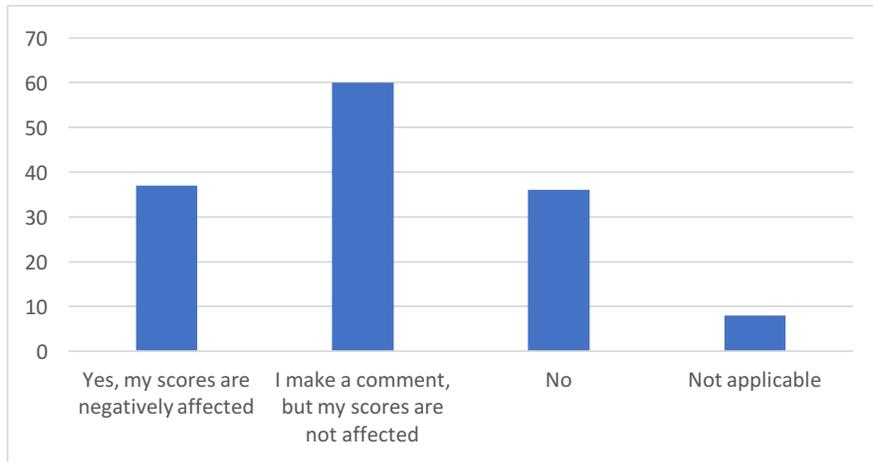
Figure 18: Responses to the question, "If a paper you are reviewing does not compare with the empirical results in a very relevant preprint paper you are aware of, do you penalize the paper for this?"

Respondents were asked if they felt that authors should be allowed to post a preprint of their paper while it is under review (Figure 19). The majority of respondents selected "Probably not," or "Definitely not."
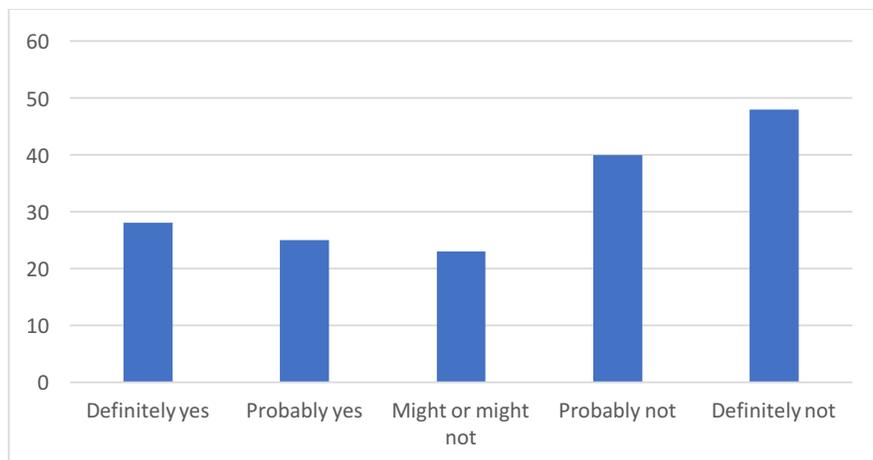

Figure 19: Responses to the question, "Do you feel that authors should be allowed to post a pre-print of their paper while it is under review?"

Respondents were asked to elaborate their selections in an open-ended follow-up question. By far, the most common explanation provided by those who selected "Probably not" or "Definitely not" was that it had a negative impact on the double-blind review process. Not only did respondents express concerns about bias, multiple respondents stated that they felt this could be exploited by well-known researchers (i.e., well-known researchers post papers to preprint servers hoping that their identify will be known by reviewers). These respondents were also concerned about the potential integrity and quality of the work being published; specifically, without peer-review, respondents expressed concerns that the work could be flawed thus generating artificial time stamp. Many of these respondents also expressed skepticism over the need to quickly get a time stamp. Several respondents made a distinction between the peer-review process and posting something to a preprint

server to get feedback on the work. These respondents saw these as two different mechanisms for getting feedback about one's work, and felt that authors should choose between these mechanisms. Several respondents felt that having something posted to a preprint server while it is under review at a venue that does double-blind review wastes reviewers' time and discounts the work they put into the review. Another common thread was related to how this would impact the review process; specifically, should reviewers now be expected to be familiar with all relevant preprints in their area? If so, should not citing a relevant preprint weigh the same as not citing a relevant article that has been subjected to peer-review? Should authors now be required to cite pre-prints, even if there is no information available about the scientific quality of the work? Many of these respondents also noted that the emphasis on speed is related to publication practices which emphasize quantity over quality.

With respect to those who selected "Definitely yes" or "Probably yes" in response to the question about whether authors should be allowed post preprints of their papers while they are under review (Figure 19), the most common explanations were to speed-up the scholarly communication cycle and to time-stamp ideas. Many of these respondents expressed concerns about having their ideas stolen or published by others in less prestigious venues. Only a few of these respondents mentioned anything about blind-review. One respondent indicated that current methods of publishing preprints has a negative impact on double-blind reviewing, but that this could be addressed by making changes to how preprints are published. Another person stated that even if preprints did not exist, the absence of an author's name on a paper that is being considered during double-blind review still does not guarantee anonymity. Another respondent advocated for open (non-blind) reviewing. Another person stated the reviewers should be discouraged from searching preprint servers.

Twenty-three respondents indicated "Might or might not." Most expressed a concern about double-blind peer-review; some of these respondents also pointed out other scholarly communication practices that are similar in nature (e.g., posting a paper to one's website). One person expressed concern about quality and work prematurely stopping in an area because a "publishable unit" was already out there.

Respondents were also asked to hypothesize about whether they think that preprint servers will become a permanent feature of research within the SIGIR community (Figure 20). A majority of respondents said "Might or might not," although when combining the two "yes" responses together ("Definitely yes" or "Probably yes") most believe preprints are not going away.
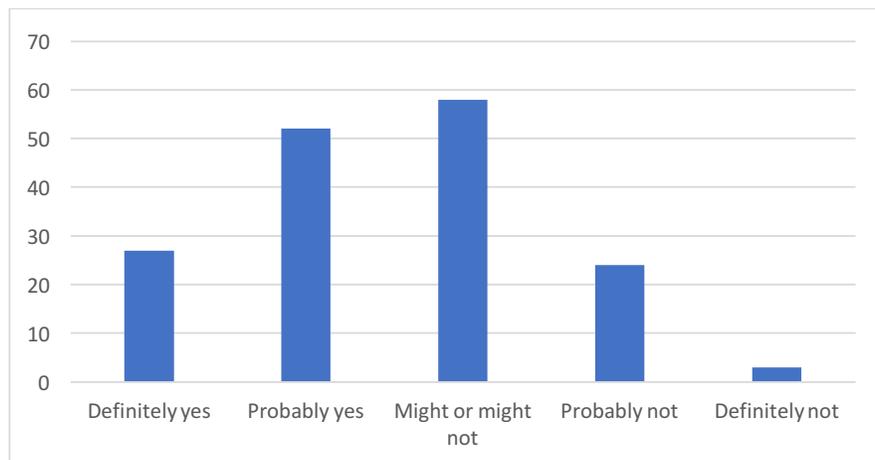
When asked if they would still submit work to SIGIR if publishing on a preprint server was banned before acceptance, the overwhelming majority of respondents selected "Definitely yes" (Figure 21).
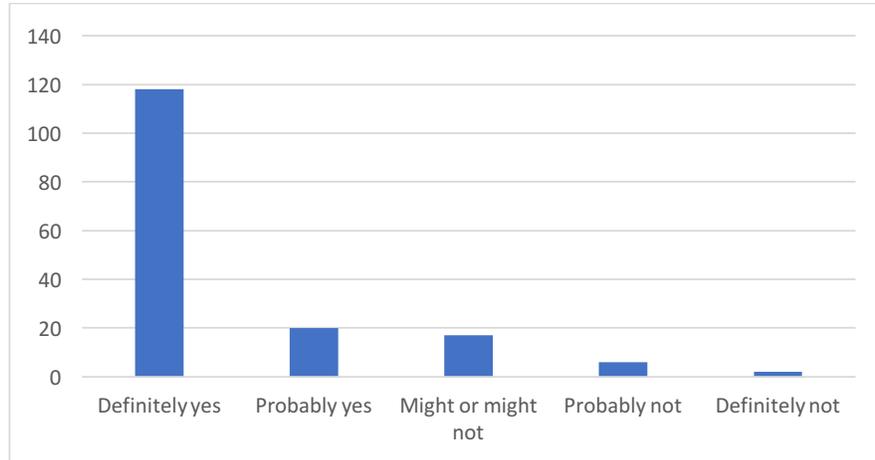
As many respondents' open-ended responses showed, one of the more challenging problems introduced by preprints is that they can compromise author anonymity and double-blind review. When asked to indicate which is more important – being able to publish your research on a preprint server or having double-blind reviewing for SIGIR conferences – the vast majority of respondents selected "Having double-blind reviewing for SIGIR conferences is more important" (Figure 22). Respondents who selected this response were asked to indicate why double-blind reviewing is important to them (Table 6). The most popular responses were "It helps to prevent unconscious bias towards authors from underrepresented groups," and "It helps to maintain the quality of SIGIR papers," although all options received a large number of selections (respondents could select all that applied).
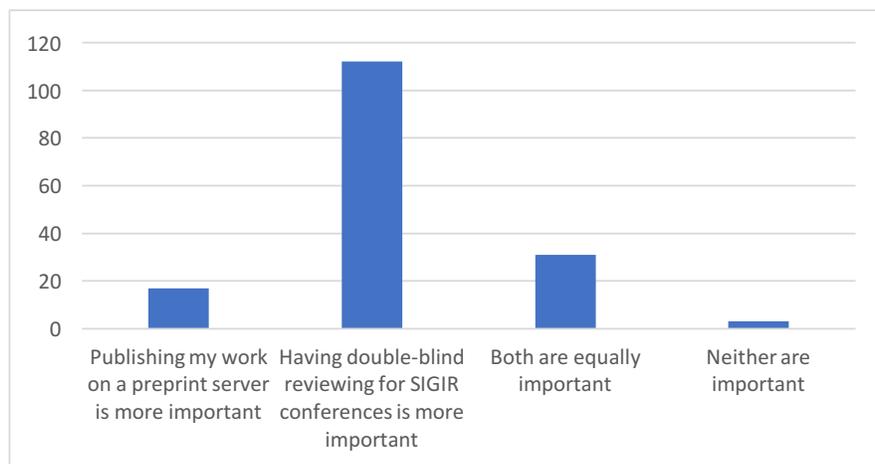
#	Answer	%	Count
1	It helps to maintain the quality of SIGIR papers	25.59%	108
2	It helps to maintain the diversity of SIGIR papers (keeps program interesting)	19.43%	82
3	It helps to prevent unconscious bias towards authors from underrepresented groups	28.91%	122
4	I do not want my reputation (or lack thereof) to affect the reviews of my work (positively or negatively)	23.22%	98
5	Other	2.84%	12

Table 6: Reasons selected by respondents who selected "Having double-blind reviewing for SIGIR conferences is more important" in Figure 22.

Finally, respondents were asked to indicate how they would like to see SIGIR's reviewing model work in the future (Table 7). Again, respondents could select all that applied. The most popular selection was "Neither reviewers nor authors are aware of each other's identities (double-blind)." Some respondents advocated for banning publishing on preprint servers before paper acceptance or allowing this with some conditions and declaration. Thirty respondents were interested in open reviewing.

#	Answer	%	Count
1	No reviewing	0.92%	3
2	Reviewers are aware of authors' identities, and vice versa (open reviewing)	9.17%	30
3	Reviewers are aware of authors' identities (single-blind)	5.50%	18
4	Neither reviewers nor authors are aware of each other's identities (double-blind)	39.76%	130
5	Publishing on preprint servers is allowed before paper acceptance but authors need to declare the existence of the preprint and reviewers are informed that a preprint exists	14.07%	46
6	Publishing on preprint servers should be discouraged but not banned before paper acceptance	9.79%	32
7	Publishing on preprint servers should be banned before paper acceptance	15.29%	50
8	Other	5.50%	18

Table 7: Potential ways that respondents would like to see the SIGIR reviewing model work.

5 Summary of Findings

Before presenting a summary of the major findings, it is important to note that these results describe the behaviors, practices and opinions of 174 people, of which 124 are SIGIR members (roughly 14% of current membership). While a larger sample size would have strengthened the findings and conclusions, these results nonetheless provide some insight that can be used to reason about how SIGIR submission and review policies might be changed.

The major findings of the survey can be summarized as follows:

- Sixty-six percent of respondents indicated that they "seldom" or "never" upload their papers to preprint servers primarily because they want to preserve the integrity of double-blind review and did not see the need since they intend to publish their papers at a conference or in a journal.
- Of those who indicated they "always," "often" or "sometimes" upload their papers to preprint servers, about 30% indicated they did so after the paper has been accepted for publication. The reasons given for doing so at this time were about a desire to have one's work reviewed anonymously, a desire to upload a mature version of the paper, and a desire to preserve the integrity of double-blind reviewing.
- The most common reasons provided by those who indicated they uploaded their papers to preprint servers at a time before the paper has been accepted for publication were to timestamp the ideas in the paper and to publicize the research.
- About 51% of respondents "sometimes" read preprint papers, while about 60% "seldom" or "never" cite preprints primarily because they would rather cite a published version if available.
- Most respondents (48%) indicated that if they discovered a preprint of a paper they were reviewing had been posted they would make a comment in their review, but not adjust their numeric scores. About 10% indicated their scores would be negatively affected.
- Most respondents (43%) indicated that if a paper they are reviewing does not site a relevant preprint they would make a comment, but not adjust their numeric scores. About 20% said their scores would be negatively affected.
- When asked if authors should be allowed to post preprints of their papers while why are under review, 54% said "probably not" or "definitely not" (24% and 30%, respectively), while 32% said "definitely yes" or "probably yes" (17% and 15%, respectively).
- The most common reason given by those who selected "probably not" or "definitely not" was related to the potentially negative impact on the double-blind review process. Another common reason was a concern about the quality and integrity of the work. These respondents also expressed skepticism over the need to quickly get a time stamp. Many of these respondents also raised questions about how allowing preprints might impact the review process.
- The most common reasons given by those who selected "probably yes" or "definitely yes" was to speed-up the scholarly communication cycle and to time-stamp ideas.
- When asked to hypothesize about whether they think that preprint servers will become a permanent feature of research within the SIGIR community, 48% of respondents said "definitely yes" or "probably yes," with another 35% saying "Might or might not."
- When asked if publishing on a preprint server was banned before acceptance, 72% indicated they would still submit their papers to SIGIR. Only 5% said "probably not" or "definitely not."

6 Discussion & Recommendations

Survey respondents were not big users of preprint servers. It is speculated that there is likely more use of preprint servers in our community than what is shown in these data. Most respondents felt preprint servers will become a permanent feature of research within the SIGIR community, which suggests we should potentially evolve our policies. On the other hand, nearly 84% of respondents said they would still submit papers to SIGIR if preprints were banned, which might lead one to conclude that preprints should be banned. However, this treats preprints as some special form of grey literature and the rationale for discriminating against this particular type is unclear. Overwhelmingly, and similar to the findings of the ACL survey, respondents expressed strong support for double-blind review.

Where does this leave us? Before making recommendations, it should first be stated that it is unclear if we should focus specifically on practices related to preprint servers or, more broadly, grey literature (at least that which is online). Perhaps papers posted on preprint servers are singled-out because they are more likely to appear in a repository and are more findable than papers posted on individual home pages, for example, or because they are elevated to the status of officially/commercially published literature as the KDD and RecSys submission guidelines suggest. Based on the review of conference submission and review policies, as well as the results of this survey, the following recommendations are made:

- **Rather than ban preprints, or any type of grey literature, the first change that might be made is to provide clearer instructions to reviewers.** A good start are the instructions ACL now provides which ask reviewers not to search for related literature until after they have read an assigned paper and formed an initial opinion. Indeed, if reviewer assignments work optimally, a reviewer should already be familiar with the related research and not need to search. If the reviewer searches and discovers the authors' names, then the reviewer should report this to paper manager, but not publicize this to other reviewers. Going a step further than the ACL guidelines, the reviewer might consider whether they can now do a fair review and if not, contact the paper manager to be released from reviewing the paper, similar to how conflicts of interest are handled once reviewing starts. This does not seem much different from reading a paper one has been assigned and realizing that one recently heard someone speak about this research as part of a seminar or at CLEF, TREC or FIRE. Reviewers also have responsibilities to uphold double-blind review and avoid conflicts of interest. Reviewers should not be allowed to use authorship discovery as a way to reject a paper. If a reviewer is uncomfortable reviewing a paper, then he/she should ask to be released from reviewing it. There are many ways that anonymity can be compromised; the key is to provide guidance to reviewers about what they should do if this happens.
- **If there is a desire to further provide instructions to authors, then we might first determine if the behavior of concern is related to posting preprints to repositories, or posting preprints online, or posting any type of grey literature.** It would seem difficult to create guidelines for the latter category. If we limit our guidelines to the first category, then we might borrow WSDM's instructions that require authors to use a different title and abstract than the submission under review, which presumably would make the preprint more difficult to find (if it were found, a comparison of the content would likely provide enough evidence for someone to determine authorship, or that a major case of plagiarism had occurred – both of which should be reported to the paper manager). The WSDM guidelines do not specify if the variations in titles and abstracts should only be maintained during the conference review period or for some other length of time. ACL introduces the notion of an *anonymity period* into their instructions

to authors, which is the period of time one month before the conference submission deadline to the time of notifications. During this period, authors are asked not post preprints. Of course, an author might have posted a preprint outside of this period; in which case, the WSDM guideline regarding titles and abstracts might be useful.

- **Provide instructions to reviewers about what to do if a relevant preprint has not been cited.** Here, ACL's guidelines seem reasonable: Refereed publications take priority over preprints. Authors are expected to cite all relevant literature regardless of publication venue, but lack of awareness/citation of a relevant preprint is not grounds to reject a paper or decrement scores. ACL instructs reviewers to point authors to relevant preprints and authors to incorporate these into the camera-ready version of the paper. However, one potential issue with this is that authors often use quality as a criterion for determining which papers to cite. Since preprints are not peer-reviewed, these instruction might conflict with other practices. It seems best to let authors determine if they find a preprint trustworthy enough to cite.

There are two other practices from the review of conference submission guidelines that might be useful to SIGIR:

- KDD has two types of full paper submission, both of which are expected to be high-quality and have impact: *Applied Data Science Papers* and *Research Papers*. The SIGIR community is very familiar with KDD's *Research Paper* – this is basically the type of research paper we already accommodate. The type of research represented by KDD's *Applied Data Science Papers*, however, is not well-recognized or incorporated into SIGIR. Creating a space for these types of papers, as well as the scientists who do this type of research, would likely strengthen and grow our community, scholarship and impact.
- RecSys asks authors to complete the SIGCHI Submitter Agreement, which is an implicit agreement by those submitting papers to a number of requirements and expectations. While the SIGIR Call for Papers typically has many of these elements listed in the submission guidelines, these details are often overlooked by authors, or forgotten, or even go unknown if one or more authors are not involved at that level. Creating an explicit agreement that all authors must agree to at the time of submission might help us make our expectations clearer. This year's SIGIR Conference submission page included several items that authors had to explicitly acknowledge about authorship, originality, conference attendance and accountability, which is a solid start in this direction.

While these last two recommendations go beyond the original purpose of this paper, they are made based on a desire to grow our community, and address misunderstandings that submitting authors sometimes have (misunderstandings which often create extra work for program committee chairs!).

7 Conclusions

Grey literature has always been an important part of the IR research landscape. Recent concerns about the relationship between preprints, a type of grey literature, and double-blind review motivated the survey described in this paper. It is clear from the survey results that respondents still consider double-blind review incredibly important. More respondents felt that authors should not be allowed to post work that is under review to preprint servers. However, rather than ban preprints, which would require first developing a very specific definition of what counts as a preprint and what does not, as well as distinguishing preprints from other grey literature, adjustments to reviewer and author instructions are first recommended. There are many ways anonymity can be broken during double-blind peer-review

beyond the existence of a preprint and it is everyone's responsibility to uphold double-blind review, not just authors.

8   Acknowledgement

9   References

[1] ACL policies for submission, review and citation. Retrieved from
https://www.aclweb.org/adminwiki/index.php?title=ACL_Policies_for_Submission,_Review_and_Citation .

[2] CIKM 2018 submission guidelines. Retrieved from
http://www.cikm2018.units.it/callforpaper.html

[3] Foster, J., Hearst, M., Nivre, J., & Zhao, S.. (2017, July). Report on ACL Survey on Preprint Publishing and Reviewing. Retrieved from
https://www.aclweb.org/portal/sites/default/files/SurveyReport2017.pdf

[4] Grey literature (n.d.). Retrieved June 19, 2018 from Wikipedia:
https://en.wikipedia.org/wiki/Grey_literature

[5] Harnad, S. (2003). Electronic preprints and postprints. In M. Drake (Ed.), *Encyclopedia of Library and Information Science (2nd edition)*. New York, NY: Marcel Dekker, Inc.

[6] KDD 2018 submission guidelines. Retrieved from http://www.kdd.org/kdd2018/calls/view/kdd-2018-call-for-research-papers

[7] Nivre. J. (2017, November 28). ACL Submission policies – whence and whither? Retrieved from
https://naacl2018.wordpress.com/2017/11/28/acl-submission-policies-whence-and-whither/

[8] RecSys 2018 submission guidelines. Retrieved from https://recsys.acm.org/recsys18/call/

[9] Schöpfel, J., & Farace, D. J. (2010). Grey literature. In M. J. Bates & M. N. Maack (Eds.), *Encyclopedia of library and information sciences (3rd edition)*. Boca Raton, FL: CRC Press.

[10] SIGIR 2018 submission guidelines. Retrieved from https://easychair.org/cfp/sigir2018

[11] Tomkins, A., Zhang, M., & Heavlin, W. D. (2017). Single versus double blind reviewing at WSDM 2017. *Computing Research Repository*, February 2017. Retrieved from **arXiv:1702.00502v6**

[12] WSDM 2019 submission guidelines. Retrieved from http://www.wsdm-conference.org/2019/