BY ERGIN ELMACIOGLU AND DONGWON LEE

Oracle, Where Shall I Submit My Papers?

IN APRIL 2005, A GROUP OF MIT STUDENTS PULLED PRANK[1] on the conference – *World Multi-Conference on Systemics, Cybernetics and Informatics (WMSCI)*[10] – known for sending unsolicited invitation emails to people in academia. The MIT students used software to generate bogus research papers, complete with context-free grammar, and submitted two of them to the conference. To their surprise, one of the gibberish papers was accepted without any reviews. The event received much attention, being covered in various media,[6] and has became an amusing topic for debate among scientists.

The discipline of computer science is unique in its publication practice. Its fast-moving pace of progress makes the dissemination of research findings much quicker and on a broader scale with increasing number of venues.[5] Yet the discipline demands ideas to be verified through rigorous peer-reviews before their publication in conferences. Although there are a few reputable journals that do not review papers or do only informal reviews, such cases are rare. Moreover, conferences that publish articles without peer-reviews are extremely rare in computer science – almost unprecedented before the event of WMSCI in 2005.

Why are conferences in computer science so special? More importantly, why are the qualities of some conferences questioned? The so-called "publish or perish" pressure to researchers may be one of the reasons. As researchers are pressured to be more prolific, and top-tier conferences in computer science become ever more competitive, some people undoubtedly look for easy venues solely for bean counting. Likewise, organizers of some conferences may exploit this demand for the source of revenue. Participating program committee (PC) members, authors, or readers who do not realize values of those profit-oriented conferences are in a sense victims. Furthermore, existence and prosperity of such conferences can significantly undermine the reputation and credibility of legitimate conferences in computer science.

Set-Up

What does it mean to label a conference *prestigious* or *questionable*? In general, the notion of good vs. bad conferences is rather subjective (but often agreeable). For instance, although the somewhat low-standard of WMSCI was revealed through MIT's prank, it does not directly suggest that WMSCI is a questionable conference. To simplify our task, therefore, let us not define what questionable conferences are. Instead, we assume that *magically* we are given a list of *questionable* conferences, called Q, and a list of *respectable* ones, called R. Then, to see if a conference c belongs to Q or R, our task is to measure how similar c is to Q and how dissimilar c is to R. In order words, we are only interested in finding out distinctive characteristics of Q, compared with R.

In order to measure the differences between Q and R, one may use quality metrics such as the Impact Factor,[4] acceptance rate, or other proposals studied in bibliometrics. However, not only such metrics cannot be easily obtained (particularly for questionable conferences), but they are easily affected by the age of conference (young conferences tend to have fewer citations than

well-established ones), and are domain-dependent (different research domains have different citation patterns). Therefore, in this study, we decided to utilize PC members to determine the characteristics of Q and R. Our underlying hypothesis of the study is:
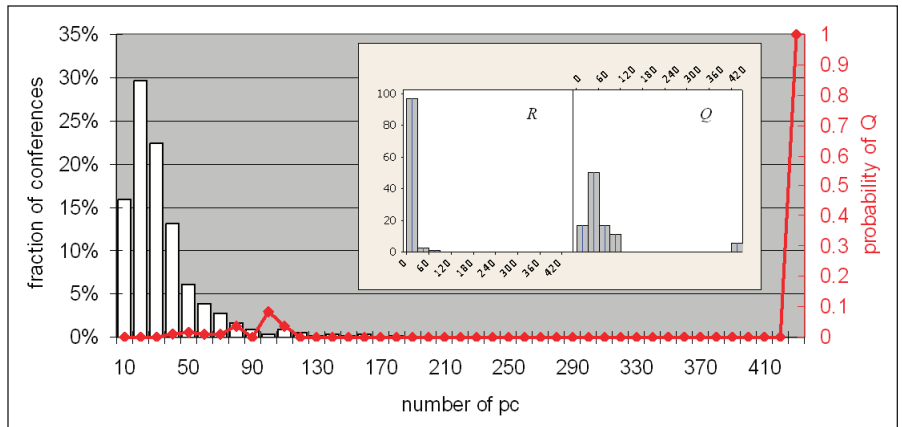
"The quality of a conference is correlated with that of its PC members."

PC information is especially useful since it can be easily extracted from "*Call for Papers* (CFPs)" data. Our test data were gathered as follows:

▸ We built Q by consulting colleagues and reading people's comments (or complaints) about certain conferences on the Web.[3] As a result, CFPs of 18 *questionable* computer science conferences of 2005 and 2006 were obtained, whose identities will not be revealed in this article. Then, between February and May of 2006, we constructed R by crawling *dbworld*,[2] an email list server to announce computer science related events including *conferences*, workshops and symposiums with more focus on Database. Note that we use the term conferences to collectively refer to all such events. At the end, 2,979 CFPs were collected for R. If a conference from Q was also included in R, it was removed from R. At the end, there were 16,136 and 930 distinct PC members in R and Q, respectively (65 common PC members in both R and Q).

▸ We used ACM Guide[9] to construct a *collaboration* graph in computer science. ACM Guide is a high quality citation digital library that has a good coverage on computing literature. We first downloaded citation data from 1950 to 2004 from ACM Guide, containing about 609,000 authors and 770,000 articles. The *collaboration graph* is a graph where nodes represent authors and edges represent co-authorship of authors.[7] ACM Guide generated about 1.2 million edges in our collaboration graph. Note that ACM Guide itself does not have a notion of "unique key" such as Digital Object Identifier (DOI). Instead, it depends on the name of authors to distinguish them. Therefore, the classical *name authority control problem* may arise (that is, same author with various spellings or different authors with the same spelling). We tried to minimize the effect of this problem by conducting two sets of experiments– one

with the full names ("Dongwon Lee") and the other with first name initial followed by last name ("D. Lee")– and used these two as the upper and lower bounds of the statistics. At the end, the difference was not significant.

▸ From the ACM collaboration graph, for each conference q in Q, we induced a sub-graph that consisted only of PC members of q. If an author in a PC list did not appear in the collaboration graph, we added a dangling node to the sub-graph representing this author. Since such nodes were isolated from the giant component of the sub-graph, they did not significantly affect the results either positively or negatively.

Results

Using these data sets, we have examined if conferences in Q and R behave differently. Among a variety of perspectives from which we have examined, due to limited space, we present three of them here: (1) the average number of PC members, (2) the average number of published articles by PC members, and (3) the average closeness centrality of PC members.

First, we measured the average number of PC members per conference. The average number of PC members for R and Q was 28.831 (std. dev. was 0.477) and 69.6 (std. dev. was 21.7), respectively (see Figure 1 Inset). The main Figure 1 has two Y-axes, adopted from:[8] *Left* Y-axis indicates the fraction of conferences (histogram) while *Right* Y-axis indicates the probability of a conference being in Q (line graph). In general, with increasing numbers of papers being submitted to respectable conferences, their PC sizes tend to increase. However, it appeared that once the number

Figure 1. The distributions of conferences by the number of PC members were different. Q tended to have more PC members in general. The probability of a conference being in Q increased as the number of PC members increased. (Inset shows different statistics of the number of PC members between R and Q, where X-axis is the number of PC while Y-axis is the fraction of conferences).

Figure 2. The average numbers of publications by PC members in R and Q appeared quite different. PC members in R generally had more publications than those in Q.
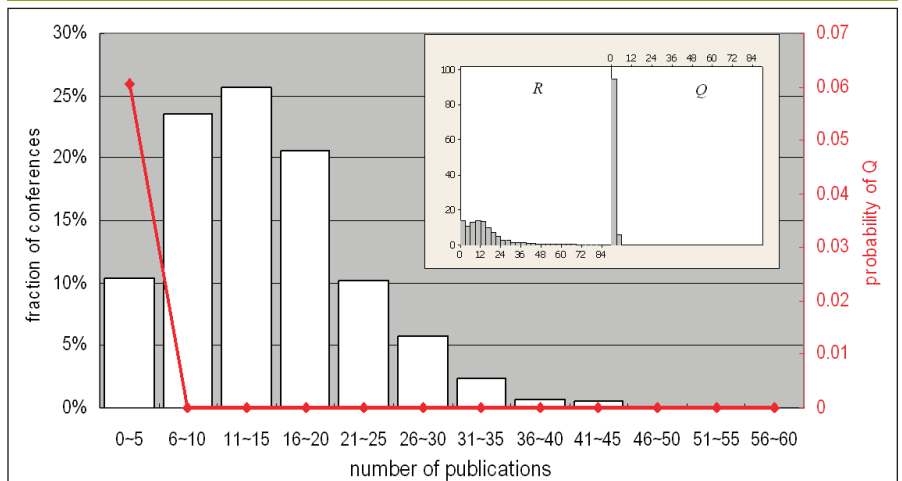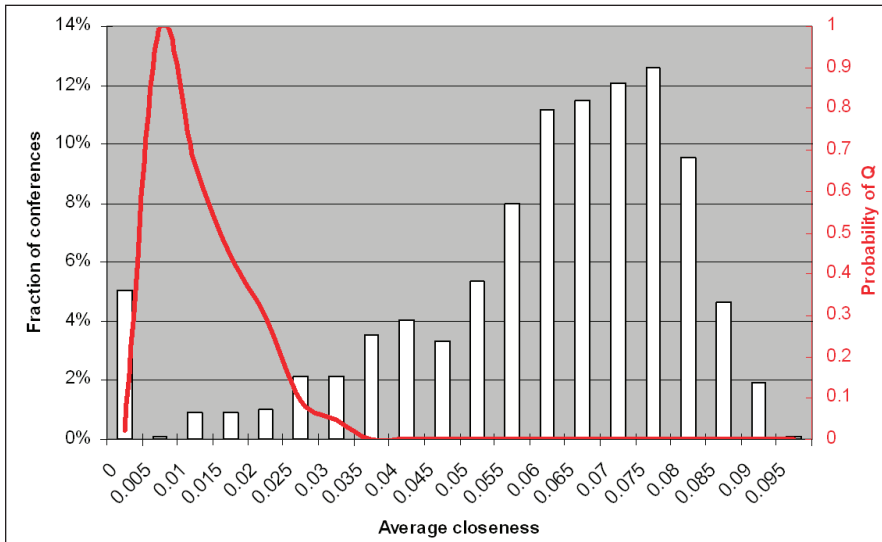
Figure 3. The distribution of average closeness and the probability of a venue belonging to Q.

of PC members surpasses a threshold of 130, the probability of a conference belonging to R becomes near 0. That is, when a conference has a substantially large number of PC members, it can be viewed as abnormal.

The second heuristics we explored was the average number of publications by PC members. Let P_c denote the set of PC members for a conference C. We mapped P_c to the ACM collaboration graph according to the person's name, and counted the number of unique publications pub_i for each PC member p_i in P_c. The average number of articles for C was then calculated as:

$$\text{pub}_{\text{avg}}(C) = \sum_{i = 1 \text{ to} |Pc|} pub_i \,/\, |P_c|$$

The result is shown in Figure 2. The mean for R was 14.7 (std. dev. of 8.2), compared with 1.5 (std. dev. of 0.9) for Q. The maximum for R and Q were 56 and 3.8, respectively. It appears that as the average number of publications by PC members increased, the probability that this conference belonging to Q decreased dramatically. Once the average number of articles exceeded a threshold of 4, it became very unlikely that the conference belonged to Q. That is, when PC members of a conference have a small number of known publications, it can be viewed as abnormal. Intuitively, this makes sense – since PC members are judges on certain areas, they usually have to have a significant number of publications on the area.

Finally, we examined the data sets using one of social network analysis (SNA) methods. The closeness centrality is to quantify individuals' locations in a community[11]. The more prominent ones are often located in the strategic locations in the social network, i.e., a collaboration graph, of the community. The closeness can be defined as how close an author is on average to all other authors. Then, authors with high closeness values could be viewed as those who can access new information quicker than others, and similarly, information originating from those authors can be disseminated to others quicker[7]. It is believed that prominent individuals in a community tend to have higher closeness values. Formally, the closeness of a node v in a connected graph G is defined as follows:

$$CC(v) = \frac{n - 1}{\sum\limits_{w \in G} d(v,w)}$$

where $d(v,w)$ is the pair-wise geodesic (shortest distance) and n is the number of nodes reachable in G. In our experiment, the average closeness of a conference C was calculated as the average closeness of its PC members, P_c, like:

$$\text{Closeness}_{\text{avg}}(C) = \sum_{i = 1 \text{ to } |Pc|} CC(i) \,/\, |P_c|$$

As shown in Figure 3, the distribution of the closeness for *respectable* conferences was right-skewed implying that the majority of the conferences had PC members with high closeness values on average. This is intuitive since PC members are usually constituted by prominent scholars of the community. However, the closeness of conferences in Q dominated the region with low average closeness. When the closeness value exceeded 0.035, the probability that a conference belonging to Q became 0. Therefore, the average closeness of a conference is another good indicator to differentiate Q from R.

Combining Heuristics

The three simple heuristics exhibited promising results at various degrees of correlation with the quality of conferences. However, each may fail to have enough distinguishing power for different extreme cases when used individually. For instance, a reputable conference such as WWW may have a large number of PC members to handle the large number of submissions, thus invalidates our first method. However, the other two methods can easily identify the quality of the conference by the quality of its PC members. Therefore, we also studied to combine the results of these heuristics in a classification framework for a better accuracy.

We used a naïve classifier, C 4.5, to classify if a given conference instance C is respectable or questionable one. We employed the ten-fold cross validation to evaluate the classification accuracy. This technique randomly divided the judged data into 10 partitions of equal size, and performed 10 training/testing phases in which nine partitions were used for training and the remaining partition was used for testing. When all three heuristics are combined, this classifier correctly judged 2,975 (99.27%) of all the 2,997 conference instances tested. The precision and recall values for filtering out questionable conferences were 0.996 and 0.997 and the false positive rate for this was very low at merely 0.003.

Conclusion

In this study, we have shown that our intuitive hypothesis, "the quality of a conference is highly correlated with that of its PC members", holds relatively well. Although we have presented only three heuristics here, other heuristics that we studied in[12] (such as, number of co-authors of PC, temporal publication pattern, and betweenness of PC, etc) also showed consistent conclusion. By measuring simple charac-

teristics of PC members extracted from CFPs, therefore, one is able to differentiate those *questionable* conferences from *respectable* ones. It is our hope that our study can raise the awareness of the issues caused by questionable conferences so that authors become more cautious about where to submit their work. Finally, it should be noted that the quality of the results of each of our heuristics is ultimately relative to the completeness and quality of the *ACM Guide to Computing Literature* and our CFP collection methodology. 🅒

References
1 Details of MIT students' prank is at: http://pdos.csail.mit.edu/scigen/
2. dbworld: http://www.cs.wisc.edu/dbworld/
3. Fake conference lists include: http://del.icio.us/tag/fakeconference, http://wiwowo.blogspot.com/2005/04/fake-conferences.html, etc.
4 Garfield, E. Citation analysis as a tool in journal evaluation. Science, Vol. 178, pp. 471-479. (1972)
5. Goodrum, A. A., McCain, K. W., Lawrence, S., Giles, C. L. Scholarly publishing in the internet age: A citation analysis of computer science literature," IPM, 37, pp 661-675, 2001.
6. Kelly-Bootle, S. Call That Gibberish? ACM Queue, 3:6 (2005)
7. Newman, M. E. Who is the best connected scientist? A study of scientific coauthorship networks. Phys. Rev. (2001).
8. Ntoulas, A., Najork, M., Manasse, M., and Fetterly, D. Detecting spam web pages through content analysis. In Proceedings of *World Wide Web Conference* (Edinburgh, Scotland) 2006.
9. The ACM Guide is at: http://portal.acm.org/guide.cfm/
10. The conference Web Site is at: http://www.iiisci.org/sci2005/
11. Wasserman, S., and Faust, K. Social networks analysis: Methods and application. Cambridge University Press, United Kingdom, 1994.
12. Zhuang, Z., Elmacioglu, E., Lee, D., Giles, C. L., Measuring conference quality by mining program committee characteristics. In *Proceedings of the ACM/IEEE Joint Conference on Digitial Libraries* (Vancouver, Canada, 2007).

Ergin Elmacioglu (ergin@yahoo-inc.com) is an engineer at Yahoo, Inc., CA.

Dongwon Lee (dongwon@psu.edu) is an associate professor of College of IST at the Pennsylvania State University, University Park, PA.