Visualizing Computer Science Communities Using Conference Hashtags

Chaoqun Ni[1], Jiepu Jiang[2]

[1] School of Library and Information Science, Simmons College

[1] College of Information and Computer Sciences, University of Massachusetts Amherst

Abstract

This poster presents preliminary results for examining computer science communities based on Twitter conference hashtags such as #iconference2016. We construct and visualize a computer science conference network based on shared Twitter users. The network is compared with another based on shared publication authors. Results show the two networks are correlated. Both of them help explain the structure of the computer science community.

Keywords: discipline; computer science; twitter; bibliometrics; altmetrics

Citation: Editor will add citation.

Copyright: Copyright is held by the authors.

Contact: ni95@simmons.edu, jpjiang@cs.umass.edu.

1   Introduction

In recent years, it becomes increasingly popular to tweet an academic conference using hashtags, especially in the field of computer science and information science. For example, people usually include #iconference2016 in tweets if they want to address to the audience of the 2016 iConference. We examine how such conference hashtags help disclose discipline structure, with a specific example of the computer science field.

A few previous work visualized and examined discipline structures using social media, e.g., Jiang, Ni, He, and Jeng (2013) used Mendeley online groups. However, few conducted large scale analysis using twitter academic conferences. A few previous studies (Wen, Lin, Trattner, & Parra, 2014; Wen, Parra, & Trattner, 2014; Wen & Lin, 2015; Gavilanes et al., 2015) only considered a limited number of academic conferences in Twitter.

We construct and visualize a network for 100 top-ranked computer science conferences in the ACM digital library. The network is compared with another network based on publication data. Results show two networks are similar and correlated. Both display clear structures of the computer science discipline.

2   Data

We collected a Twitter dataset related to computer science conferences. The ACM digital library website[1] provides a list of conference proceedings published by ACM. We selected 1,139 conference proceedings (workshops were removed) published from 2009 to 2014 and generated 2,278 hashtag queries using conference name acronym (e.g., SIGIR) and year (four digits or only the last two digits). For example, for the WWW'15 conference, we generated two hashtag queries #www2015 and #www15.

The twitter REST API[2] only returns recent tweets for queries. In order to collect tweets for past conferences, we recruited a human worker to manually download search results for all hashtag queries. The human worker was instructed to search each hashtag query manually using a browser. After all tweets were displayed, the worker saved the webpage as an HTML file. The saved HTML file includes ids and contents of all tweets on the webpage, as well as ids and names of the users who posted these tweets. The worker downloaded the search results from October to December, 2014.

The collected dataset includes twitter search results for 2,278 hashtag queries for 1,139 conference proceedings (344 unique conferences) in total. In total 155,945 tweets posted by 43,657 unique users were retrieved. 1,334 (58.6%) hashtag queries returned results, covering 795 conference proceedings and 264 unique conferences.

---

[1] http://dl.acm.org/proceedings.cfm

[2] https://dev.twitter.com/rest/public

3   Conference-author-coupling and Conference-user-coupling Networks

We examine computer science communities by looking into connections of computer science conferences. This is because conference proceedings now become the major publication venues for the computer science field. We measure connections of computer science conferences using their twitter user communities and compare with those based on author communities. The venue-author-coupling (VAC) approach (Ni, Sugimoto, & Jiang, 2013) is adopted for analyzing connections of conferences. This approach assumes that two venues (e.g., journals or conferences) are related if they have similar groups of authors. Similarly, we assume two conferences are related if they share many users on Twitter.

We select 100 unique conferences for analysis in this study. The conferences are selected based on the total number of citations for the conferences' articles published between 2009 and 2014. Citation is counted based on a crawl of the ACM digital library database in March, 2015. These 100 conferences are from various sub-fields of computer science. We label fields of the conferences using the categorization of conferences in Microsoft academic search. Legends in Figure 1 and Figure 2 show names of these sub-fields.
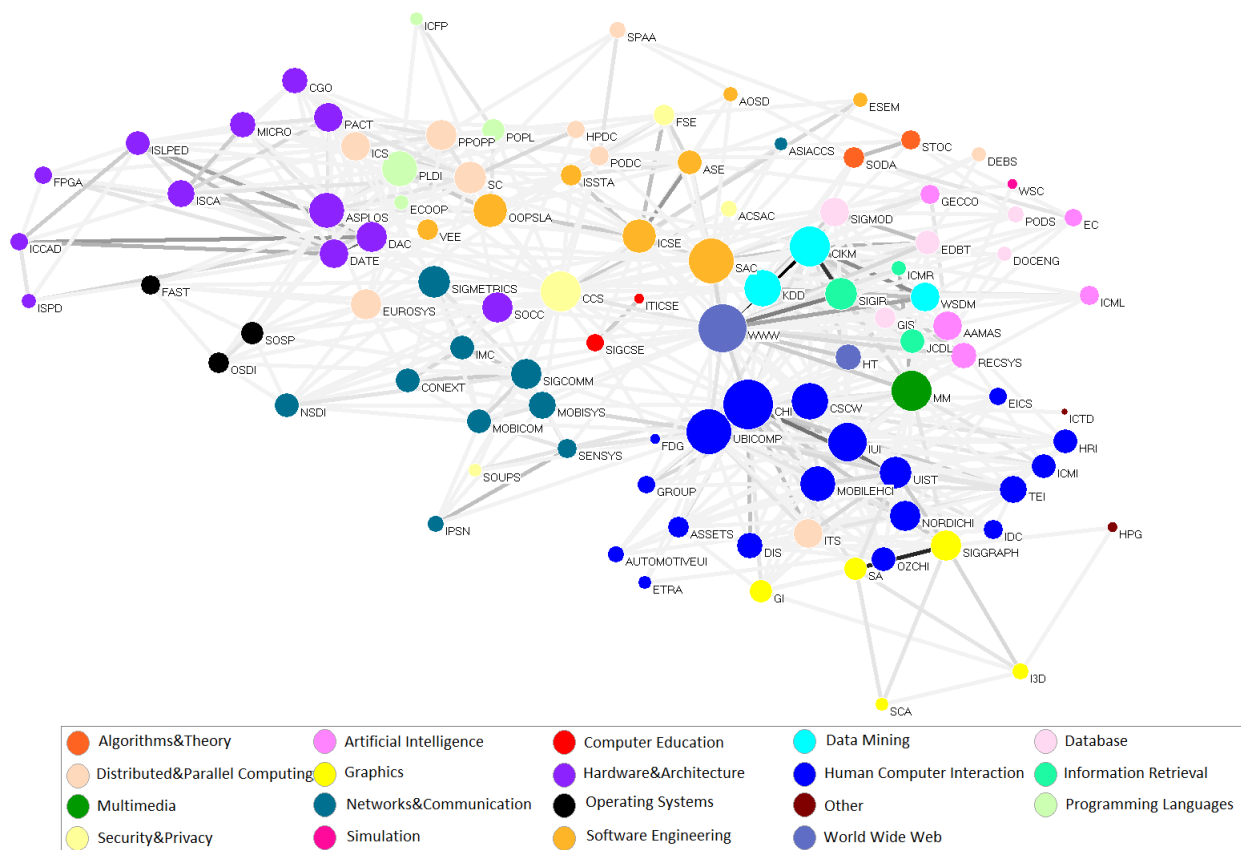
3.1   Conference-author-coupling Network

Figure 1: Conference-author-coupling network (only display edges with strength greater than 25).

We create a conference-author-coupling network using the set of shared authors who publish articles in two conferences. Each node of the network is a conference and the strength of connection between two nodes (conferences) is the number of authors who published in both conferences during 2009 and 2014 (to be comparable with the twitter dataset).

Figure 1 shows this network. We visualize the network using Pajek, and we use Kamada-Kawaii layout. Conference nodes are colored by their subject field categorization based on Microsoft Academic Search schema. The size of node is proportional to its degree centrality, e.g., the total number of shared authors with all other conferences. The grey scale of an edge is proportional to its connection strength, e.g. the number of

shared authors between two conferences: darker edges indicate two conferences have a greater number of shared authors.

The number of authors shared by two conferences varies. In total, 7,556 connections (conference pairs) were built among these conferences. About 22.8% conference pairs only shared one author with each other, and conference pairs with less than 10 common authors account for about 70% of the total. For the purpose of readability, edges between conferences with less than 25 common authors were removed (but conference nodes were kept in the figure). As Figure 1 shows, the community of conferences based on their shared authorship is very close to their subject classification from Microsoft Academic Search. Conferences in the same subject area seem to share more authors with each other in the network, such as those Human Computer Interaction conferences, Hardware & Architecture conferences, Network & Communication conferences, and Operating System conferences. Some conferences, such as WWW, also shows its central role in this network: it is closely connected with conferences in Information Retrieval, Human Computer Interaction, etc., which is not surprising considering its wide topic coverage.
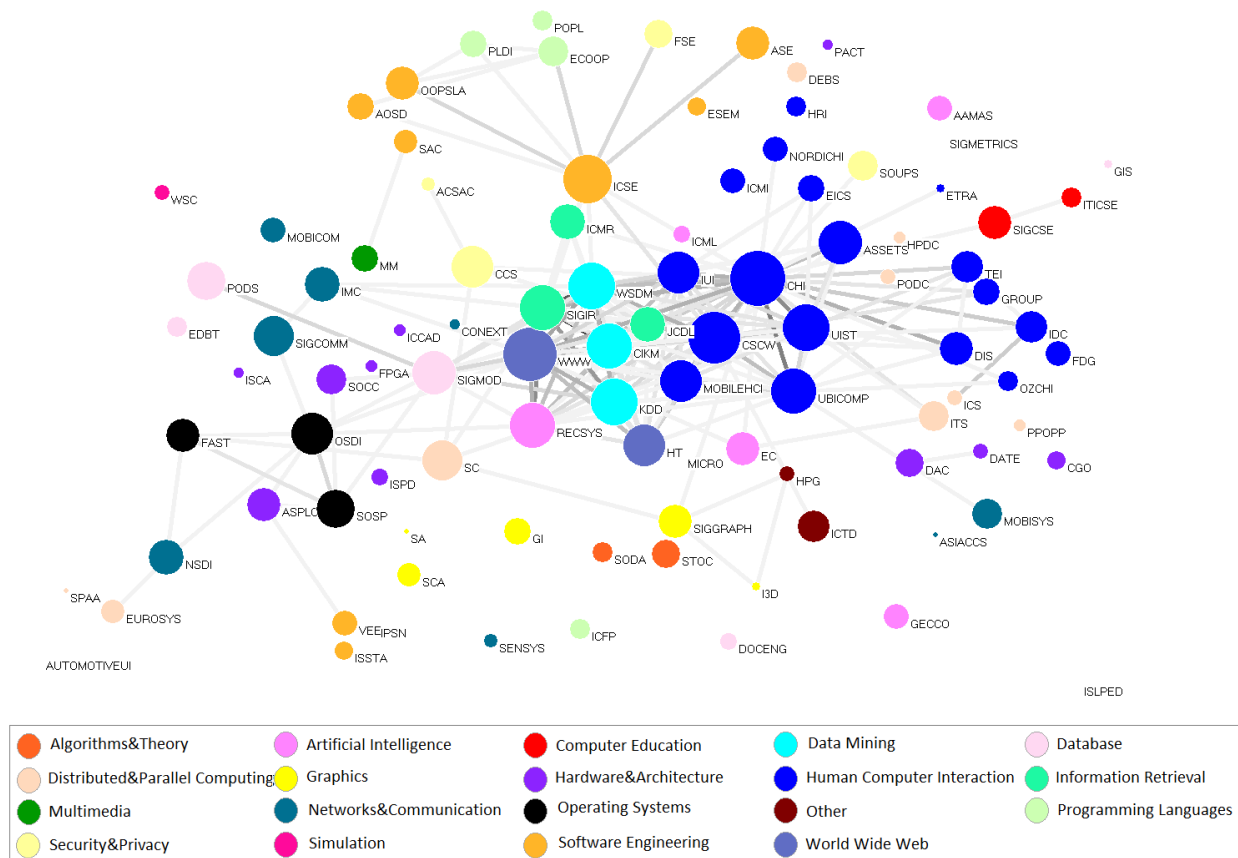
3.2   Conference-user-coupling Network

Figure 2: Conference-user-coupling network (only display edges with strength greater than 3).

We further construct a conference-user-coupling network and compare with the conference-author-coupling network. Still, we use conferences as nodes, but the strength of connection between two conferences is the number of common users posted at least one tweets to both conferences (the tweets include the conferences' hashtags). Similar to the above conference-author-coupling network, each node is a conference, and the color is coded based on Microsoft Academic Search classification schema. The network is displayed in the Kamada-Kawaii layout. The size of each node is proportional to the number of shared Twitter users, and the grey scale of each edge is proportional to the number of shared users among the conference pair connecting by the edge.

The number of shared Twitter users among each conference pairs is smaller than that of authors overall. There are about 1,920 conference pairs that share Twitter users with each other. The largest number of Twitter users shared by conference pairs is 108. About 1,000 (50.2%) of the conference pairs only have one Twitter user in common, and 83.9% of the conference pairs share less than 5 Twitter users with each other. For the purpose of a clear view of the network, edges with values lower than 3 are not displayed.

Conference communities based on shared twitter users also seem to be consistent with conferences' subject area classification. As Figure 2 shows, conferences in the same subject area seem to be closely connected with each other based on their shared Twitter users, such as those Human Computer Interaction conferences, Software Engineering conferences, Data Mining conferences, etc. Similar to the result in conference-author-coupling network, conferences in related subject areas also seem to show closeness based on their share Twitter user profile. For instance, WWW is also closely connected to Data Mining conferences (WSDM, CIKM, and SIGKDD), and Information Retrieval conferences (e.g., SIGIR).

However, comparing to the conference-author-coupling network, some of the inter-connected conferences do not show strong connections, such as the cluster of conferences for Hardware & Architecture. This cluster of conferences is separated into two groups, one of them (e.g., ASPLO, SOCC, and ISPD) are connected to Operating System conferences (which is also very related to Hardware & Architecture), while another group (e.g., DAC, DATE) show no connection with other conferences in Hardware & Architecture. This shows twitter academic conferences also show differences comparing to those based on academic publishing data. However, so far it remains unclear what causes this difference. Although some studied the motivation of online community users in websites such as Mendeley (Jeng, He, & Jiang, 2015), few looked into users motivations for tweeting academic conferences in Twitter. We leave this, and the interpretation of the differences, for future work.

3.3 QAP Correlation

Additionally, a comparison of conference proximity results based on authors and Twitter users was conducted using the Quadratic Assignment Procedure (QAP). The QAP correlation test is commonly used in social network analysis as a way of testing the correlation between different networks. The QAP approach was employed here to test the correlation between the conference-author-coupling network and conference-user-coupling network. Result shows that these two networks are significantly correlated: $r = 0.693$, $p < 0.001$, indicating moderate to high correlation between these two networks.

4 Conclusion

To conclude, the significantly high correlation between conference-author-coupling network and conference-Twitter-user network might be caused by the overlap between conference authors and Twitter users. Twitter users tweet about a certain conference probably because they are involved into the conference somehow: they might be conference participants, conference article authors, or someone who are interested in the conference. Most conference participants may also be authors of some papers, and they may tweet about the conference and their articles.

References

Gavilanes, R. O. G., Gomez, D., Parra Santander, D., Trattner, C., Kaltenbrunner, A., & Graells, E. (2015). Language, twitter and academic conferences. In *Proceedings of the 26th ACM conference on hypertext & social media* (pp. 159–163).

Jeng, W., He, D., & Jiang, J. (2015). User participation in an academic social networking service: A survey of open group users on mendeley. *Journal of the Association for Information Science and Technology*, *66*(5), 890–904.

Jiang, J., Ni, C., He, D., & Jeng, W. (2013). Mendeley group as a new source of interdisciplinarity study: How do disciplines interact on mendeley? In *Proceedings of the 13th ACM/IEEE-CS joint conference on digital libraries* (pp. 135–138).

Ni, C., Sugimoto, C. R., & Jiang, J. (2013). Venue-author-coupling: A measure for identifying disciplines through author communities. *Journal of the American Society for Information Science and Technology*, *64*(2), 265–279.

Wen, X., & Lin, Y.-R. (2015). Tweeting questions in academic conferences: Seeking or promoting information? In *iConference 2015 proceedings.*

Wen, X., Lin, Y.-R., Trattner, C., & Parra, D. (2014). Twitter in academic conferences: Usage, networking and participation over time. In *Proceedings of the 25th ACM conference on hypertext and social media* (pp. 285–290).

Wen, X., Parra, D., & Trattner, C. (2014). How groups of people interact with each other on twitter during academic conferences. In *Proceedings of the companion publication of the 17th ACM conference on computer supported cooperative work & social computing* (pp. 253–256).