A Bibliometric Analysis of SBP-BRiMS

Amit Upadhyay[1] and Wen Dong[1[0000−0001−8923−2227]]

University at Buffalo, Buffalo NY 14260, USA
{amitupad,wendong}@buffalo.edu

Abstract. New technologies have brought transformative change to the social sciences. How can SBP-BRiMS, at the frontier where new technologies meet the social sciences, respond to the changing paradigm? In this paper, we apply network-science and information-science approaches to analyze the bibliography metadata of all 470 published SBP-BRiMS papers, and present our observations of the activities and interactions among the publications, authors, institutions, and fields of study. We conclude with a list of topics that show the strongest growth areas in recent years and our interpretation of what caused this growth.

Keywords: Data mining · Data visualization · Social network analysis.

1 Introduction

In recent years, cloud technologies, big data, and artificial intelligence have transformed both society as a whole and our methodologies as researchers. Topics such as disinformation, cybersecurity, data privacy, deep learning, graph neural networks, multi-agent learning, and neural natural language processing are gaining research attention. In this new environment, how have research questions and methodologies changed at the frontier where new technologies meet the social sciences, and how must the SBP-BRiMS community respond to this changing paradigm?

In this paper, we analyze the interactions among authors, institutions, and fields of study using the bibliography metadata of all 470 published SBP-BRiMS papers. We extracted this metadata from Microsoft Academic Graph (MAG) using the Project Academic Knowledge REST API, including the author, institution, field, citation, and cited-by information of each SBP-BRiMS paper. MAG is a graph database of authors, institutions, papers, publication venues, and fields of study, and each entity in MAG has a unique identifier.

Scientometrics — the field of measuring and analyzing scholarly literature — has a long history in information science and other disciplines, and our contribution is its application to the publications in this conference. A number of researchers have modeled how the citation count of a paper can be predicted by author reputation, venue impact [12], average citation count in the discipline [5], and the paper's position relative to structural holes [1, 8, 7, 9, 6] and other important works [10]. Indicators of impact such as h-index [4], g-index [2], and JIF [3] all show the importance of citation. Informed by previous research, we analyzed how authors and fields of study co-occurred in papers, how papers by salient authors were cited, and how new fields of study were introduced.

Microsoft Academic is a free public web-search engine for academic publications, and models scholarly communications with three key AI technologies: 1) natural language understanding, 2) knowledge-assisted inference and reasoning, and 3) reinforcement learning [11]. To date, it has indexed 240 million scientific papers and patents, 240 million authors, 740 thousand fields of study (FoS), 4 thousand conferences, 48 thousand journals, 25 thousand institutions, and the relationships among them. Users can access Microsoft Academic with a web browser [1], the Project Academic Knowledge REST API [2], and a subscription to Microsoft Academic Graph data [3]. The FoS are identified from Wikipedia entities automatically at the bottom four levels and manually at the top two, and are tagged to academic publications through the similarity between the textual representations of the FoS and the publications in the embedding space.

The remainder of the paper is organized as follows. In Section 2, we analyze the interactions between fields of study and authors in SBP-BRiMS publications, and the interaction of SBP-BRiMS with other publication venues. In Section 3, we analyze the dynamics of six FoS clusters individually in terms of salient authors and FoS. We make conclusions in Section 4.

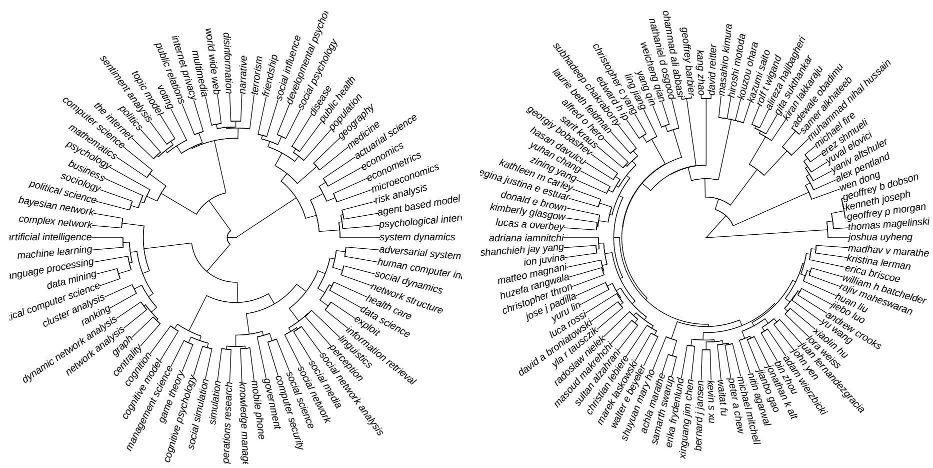
2 Fields of Study, Authors, and Related Venues

We scraped the titles and authors of all 470 SBP-BRiMS papers over the past ten years from the publisher website [4]. Then, we identified the unique IDs of the papers and their authors, fields of study, and citations in Microsoft Academic Graph with the Project Academic Knowledge REST API. With these paper metadata and further queries, we established a profile of papers, authors, institutions, fields, and venues related to the conference. In the following, we characterize SBP-BRiMS in terms of its topics, authors, citations, and publication volumes.

Fig. 1 demonstrates that SBP-BRiMS is an important forum for computer science and social science researchers to exchange ideas and establish collaborations. The two dendrograms show how the most frequent 74 fields of study and 90 authors co-appeared in papers. 95% of papers contained at least one of the 74 fields, and 55% one of the 90 authors. The dendrograms were constructed using cosine distance between how fields/authors co-occur in SBP-BRiMS papers with other fields/authors, and Ward's method to cluster the entities from the bottom up. The field dendrogram shows interdisciplinary interactions — the interactions of `artificial intelligence`, `machine learning`, and `data mining` with `complex network` and `natural language processing`; the interactions of `political science`, `sociology, business`, `psychology`,

[1] https://academic.microsoft.com/
[2] https://docs.microsoft.com/academic-services/project-academic-knowledge/
[3] https://docs.microsoft.com/academic-services/graph/
[4] https://link.springer.com/conference/sbp

`terrorism`, `public relations`, and `disinformation` with `topic modeling`, `computer science`, and `mathematics`; and the interactions of `public health`, `health`, and `medicine` with `simulation`, `agent-based modeling`, and `system dynamics`. The author dendrogram shows that collaborations in this venue are infrequent, which is typical for an interdisciplinary conference.

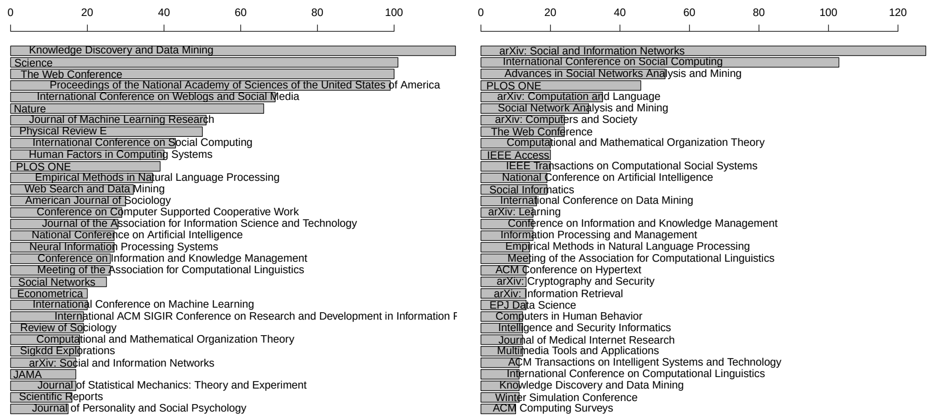


(a) Fields of study (b) Authors

Fig. 1: Interactions among fields of study and authors within SBP-BRiMS. The dendrograms were created with Ward's method and cosine distance over the numbers of co-occurrences among authors and fields.

Fig. 2 shows the times that SBP-BRiMS papers cited (left) and were cited by (right) papers from other publication venues, indicating how this venue has influenced and been influenced by others. The influence comes from (left) top scientific research journals such as Science, Nature, PLOS-One, PNAS, SBP-BRiMS, and CMOT; data-mining venues such as KDD, WWW, ICWSM, and WSDM; natural language processing venues such as EMNLP and ACL; machine learning venues such as JMLR, AAAI, and NeurIPS; and human-computer interaction venues such as HCI and CSCW. The influence also goes to (right) publication venues of scientific research, data mining, natural language processing, machine learning, and human-computer interaction, although conference publications are less likely to be cited than journal publications. The cross-citation relationships between SBP-BRiMS and other venues again show that SBP-BRiMS is an important forum for connecting computer science and social sciences researchers.

Fig. 3 shows the number of SBP-BRiMS papers published in each year and the citations they received, indicating a paradigm shift around 2014. The average number of papers published from 2010 to 2014 is about 47, but from 2016 to 2020 is 37. The average citations received by a paper are 12 in 2010-2013, but only 4 in 2014-2020. The 10 most-cited papers, their fields of study, and their



(a) Times citing

(b) Times cited by

Fig. 2: Number of times SBP-BRiMS papers have cited (left) and been cited by (right) papers from other venues. The cross-citation relationships show how this venue has influenced and been influenced by other venues.

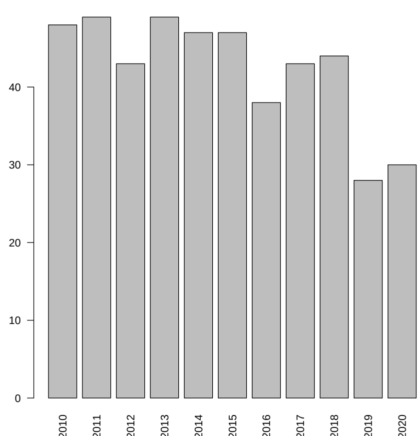
citations received are given in Table 1. We know that cloud computing, big data, and artificial intelligence are transforming technology and the social sciences. In the rest of this paper, we analyze how SBP-BRiMS has responded to this changing paradigm in terms of the activities and interactions among fields of study, publications, authors, and institutions.

3 Characterizing Dynamics by Field-of-Study Clusters

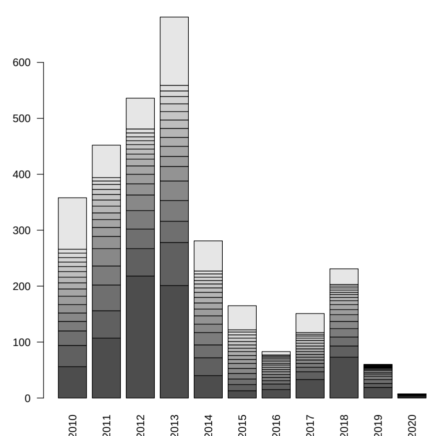
To characterize academic communications in SBP-BRiMS at the macroscopic level, we clustered the 470 papers by their fields of study. We applied cosine distance and Ward's method, and cut the resulting dendrogram into six clusters. The outcome is in Fig. 4. We show how fields of study co-occurred in papers as clusters in Fig. 4a — the darker the red the more substantial the co-occurrence. The FoS `computer science` co-occurs with almost all topics, `social media` has fewer interactions with `agent-based modeling` and `system dynamics`, and `machine learning algorithms` form a cluster. The dendrogram in Fig. 4a was constructed in the same way as in Fig. 1a. In Fig. 4b, papers were organized in clusters according to the availability of several key fields of study. We heuristically cut the paper dendrogram into six clusters so that we can analyze academic communications by area but need not deal with a prohibitive number of areas. The important topics in each cluster are shown in Fig. 4 c-f, where cluster 2 is on `politics`, cluster 3 on `sociology`, cluster 4 on `machine learning`, cluster 5 on `data mining`, and cluster 6 on `psychology`. Cluster 1 is a mega-cluster containing several smaller clusters related to general `computer science`.

Table 1: 10 most-cited SBP-BRiMS papers.

Citation	Title	Fields of Study	Year
218	automatic crime prediction using events extracted from twitter posts	latent dirichlet allocation, sentiment analysis, social media, geospatial analysis,...	2012
201	predicting personality using novel mobile phone based metrics	mobile phone, personality, phone, big data, world wide web, computer science, personality prediction	2013
107	identifying health related topics on twitter: an exploration of tobacco related tweets ...	topic modeling, social media, public health, social networks, data science, world wide web, health-related, tobacco use	2011
77	measuring user credibility in social media	credibility, social media, social movement, misinformation, internet privacy, politics, business	2013
73	aspect level sentiment classification with attention over attention neural networks	artificial neural networks, natural language processing, computer science, laptops, artificial intelligence	2018
56	social network data and practices: the case of friendfeed	social network, data science, public relations, computer science, sociological research	2010
49	lessons learned in using social media for disaster relief asu crisis response game	emergency management, social media, popularity, public relations, computer security, usability, business, crisis response, disaster response	2012
49	promoting coordination for disaster relief from crowdsourcing to coordination	crowdsourcing, emergency management, public relations, computer security, environmental disaster ...	2011
46	temporal visualization of social network dynamics prototypes for nation of neighbors	social network, social network analysis, information visualization, visualization, social dynamics, community network, testbed, user interface, human-computer interaction	2011
40	identifying users with opposing opinions in twitter debates	social media, public opinion, sentiment analysis, public policy, internet privacy, world wide web, politics, political science, label propagation, stance detection	2014

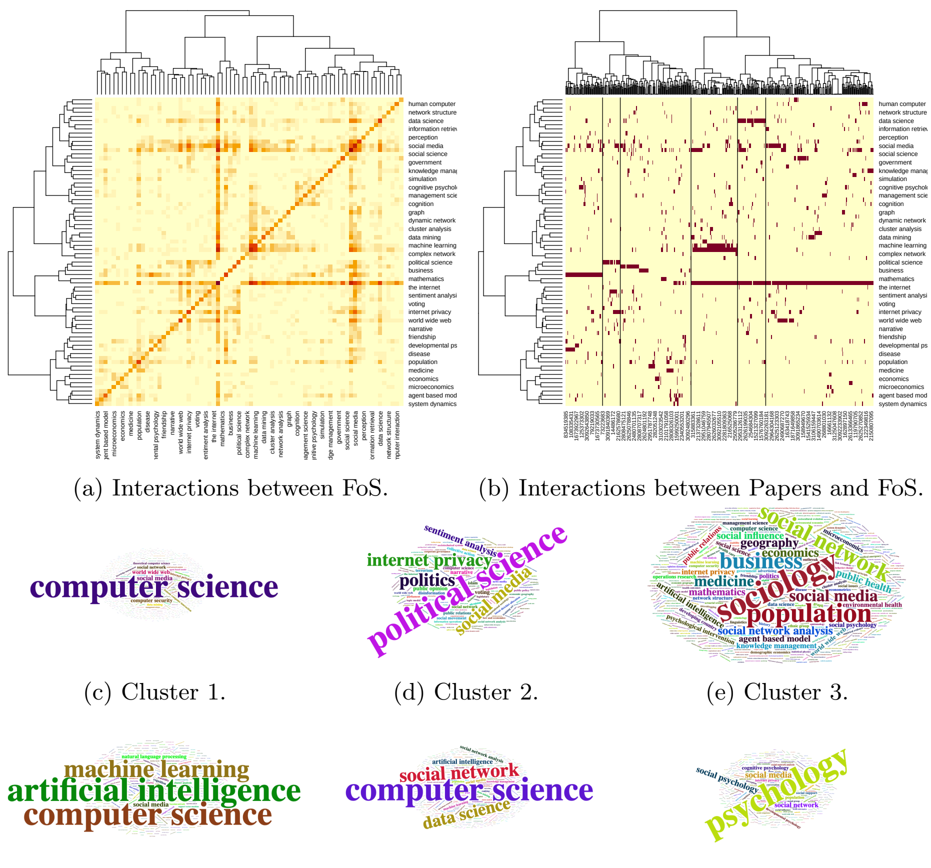


(a) Papers by year



(b) Citations by year

Fig. 3: SBP-BRiMS number of papers (left) and number of citations (right) per year. In (b), the citations of the top-16 papers in each year are shown.



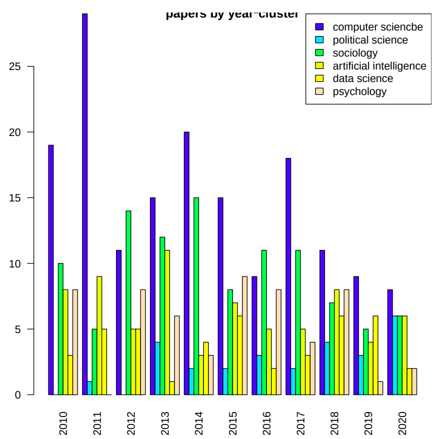
(a) Interactions between FoS. (b) Interactions between Papers and FoS.

(c) Cluster 1. (d) Cluster 2. (e) Cluster 3.

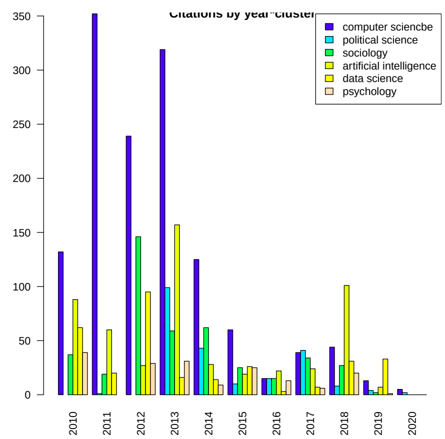
(f) Cluster 4. (g) Cluster 5. (h) Cluster 6.

Fig. 4: Interactions between fields of study and papers (a,b) in SBP-BRiMS publications, and the word clouds (c-h) showing the fields of study in each cluster identified from (b).

Fig. 5 shows the number of papers belonging to each of the six clusters published each year, and the citations attracted by these papers. In terms of paper count, the general `computer science`, `sociology`, and `artificial intelligence` clusters have higher representations, while the `political science`, `network science`, and `psychology` clusters have lower but stable representations. The publication count in the general `computer science` cluster decreases year-over-year, while the publication counts in other clusters are more stable. In terms of citation count, the general `computer science` cluster is highest, but this number is decreasing rapidly. The citations to `artificial intelligence` papers are relatively stable, demonstrating the synergy between artificial intelligence research and the social sciences. Overall, we speculate that SBP-BRiMS conference papers are more likely to be cited by computer science conferences but less likely by social sciences journals. As such, one possible approach to boosting impact is cultivating computer scientists — and especially those working with artificial intelligence — to solve problems in the social sciences.



(a) Paper counts by cluster and year



(b) Citations by cluster and year

Fig. 5: Publication counts and citations by year.

We also evaluated the global performance of the top 20 FoS in each cluster. `social media computer science` appeared in most of the clusters, and their publication volumes increased by 840% and 210% respectively from 2010 to 2020. Other FoS with the fastest growth in publication volume are `adversarial system` (1149%), and `graph` (260%) in cluster 1; `disinformation` (2765%), `sentiment analysis` (880%), `computational sociology` (682%), and `islam` (192%) in cluster 2; `psychological intervention` (249%), and `environmental health` (193%) in cluster 3; `supervised learning` (437%), `machine learning` (270%), `classifier` (265%), and `sampling` (249%) in cluster 4; `topic model`, `graph`, `centrality` (209%), and `mobile phone` (205%) in cluster 5; and `disease` (204%) in cluster 6. These fields have clearly gained research momentum, and publishing on these topics might further strengthen the impact of SBP-BRiMS.

Each cluster also has its prominent authors. Kathleen M. Carley in cluster 1 has about 26 papers, the first published in 2013 and most recent in 2020, and she actively co-authors with other prominent SBP-BRiMS authors like Nitin Agarwal and Kenneth Joseph and with new authors to SBP-BRiMS. Nitin Agarwal in cluster 2 has about 13 papers, the first in 2013 and most recent in 2020, and actively co-authors both with prominent authors like Thomas Magelinski and Carley and with new authors. Christopher C. Yang of cluster 3 has 7 papers, 2011 to 2015, and has co-authored with many new authors. Cluster 4's Hiroshi Motoda published 8 papers, 2010 to 2015, and has co-authored with prominent authors like Kazumi Saito, Kouzou Ohara, and Masahiro Kimura. Kristina Lerman in cluster 5 has published 4 papers, 2010 to 2017, and co-authored with many new authors. And Alex Pentland in cluster 6 published 12 papers from 2012 to 2017 and co-authored with both experienced authors like Yaniv Altshuler or Wen Dong and with several new authors as well.

Fig. 6 shows the co-authorship and citation relations among SBP-BRiMS authors with at least three publications. In this figure, each node represents an author, node size represents the total citations attracted by that author's SBP publications, and node color represents the publication cluster in Fig. 4 to which these publications generally belong. This figure demonstrates that good citation performance is dependent both on dense interactions within a cluster and on the influence of salient authors — such as Carley in the sociology cluster (about 26 papers), Agarwal in the social media cluster (about 13 papers), Pentland in machine learning and psychology (about 10 papers), and Huan Liu in the social media cluster (about 13 papers) — because bigger nodes often correlate with dense interactions. We note that many of the top-20 authors (based on citations) have become inactive over time: 4 published their last paper in 2011, 5 in 2013, 4 in 2015, and 2 in 2017. It is important to retain and develop influencers to keep publication clusters active.

Fig. 7 compares yearly publications of the top-10 SBP-BRiMS FoS with their corresponding year's global publication numbers. Among these 10 FoS (Fig. 7b), the yearly publication volumes of `social media` and `machine learning` rose by 840% and 270% (respectively) between 2010 and 2020; investment in these fields might be fruitful. In comparison, the research in `data science`, `internet privacy`, `social networks`, and `world wide web` appears to be saturated. The SBP-BRiMS yearly publication volume in `computer science` is shrinking, while globally growth is nonetheless strong.

We further evaluated the global performance of the top-50 SBP-BRiMS FoS (based on their publication numbers), and from them we identified the top-10 fastest growing FoS from 2010 to 2020. They are `semantic analysis` (850% change in annual publication volume), `social media` (840%), `topic modeling` (520%), `machine learning` (260%), `graphing` (270%), `psychological intervention` (250%), `centrality` (210%), `network structure` (210%), `computer science` (210%), and `exploits` (190%). These FoS have low representation at SBP-BRiMS but are gaining research interest globally. Improving the publication count of these subjects might help SBP-BRiMS to further increase its impact.

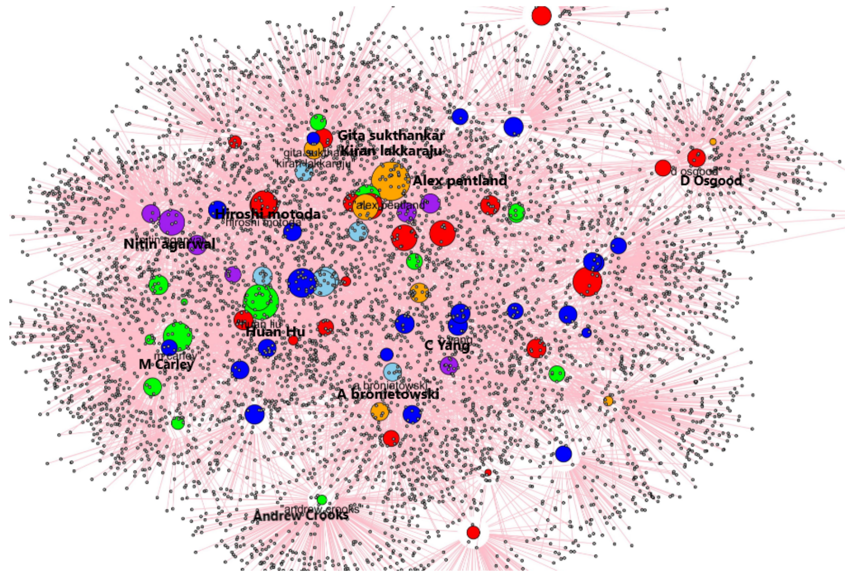
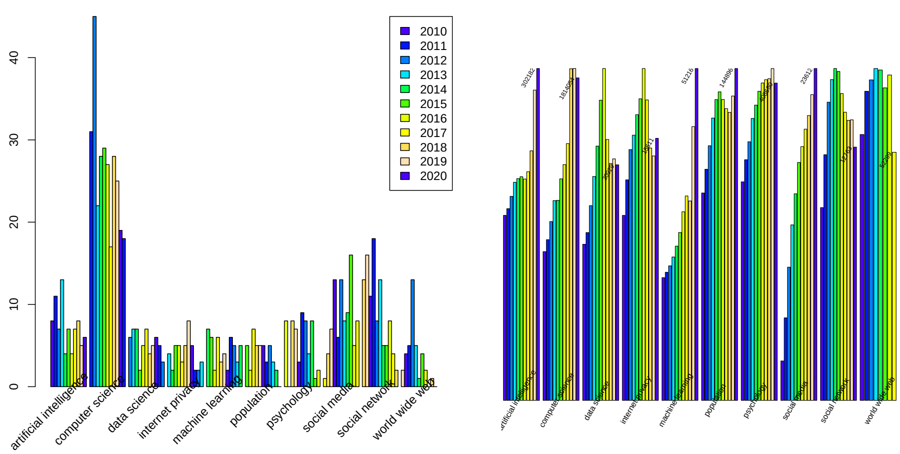


Fig. 6: Graphical representation of interaction of authors of SBP among themselves and with their references



(a) SBP-BRiMS

(b) Academia

Fig. 7: SBP-BRiMS number of papers (left) and number of citations (right) per year. In (b) the citations of the top 16 papers in each year are shown.

4 Conclusions

The transformation brought by cloud technology, big data, and artificial intelligence prompts us to rethink our research questions and methodology. In this paper, we conducted a bibliometric analysis of the 470 SBP-BRiMS publications from 2010 to 2020 — investigating how subjects and authors co-appear in a paper, how SBP-BRiMS papers cited and were cited by papers from other venues, and how publications in each year and citations to these publications changed over time. To predict future directions of SBP-BRiMS, we analyzed its publications by clusters of subjects, the interactions and influencers in each cluster, and the topics with strong growth. We conclude by noting several topics with strong potential, and the need to strength collaboration within this community.

References

1. Boyack, K.W., Klavans, R., Börner, K.: Mapping the backbone of science. Scientometrics **64**(3), 351–374 (2005)
2. Egghe, L.: Theory and practise of the g-index. Scientometrics **69**(1), 131–152 (2006)
3. Garfield, E.: The history and meaning of the journal impact factor. JAMA **295**(1), 90–93 (2006)
4. Hirsch, J.E.: An index to quantify an individual's scientific research output. Proceedings of the National Academy of Sciences of the United States of America **102**(46), 16569–16572 (2005)
5. Radicchi, F., Fortunato, S., Castellano, C.: Universality of citation distributions: Toward an objective measure of scientific impact. Proceedings of the National Academy of Sciences of the United States of America **105**(45), 17268–17272 (2008)
6. Ren, X., Liu, J., Yu, X., Khandelwal, U., Gu, Q., Wang, L., Han, J.: Cluscite: effective citation recommendation by information network-based clustering. In: Proceedings of the 20th ACM SIGKDD international conference on Knowledge discovery and data mining. pp. 821–830 (2014)
7. Shi, X., Leskovec, J., McFarland, D.A.: Citing for high impact. In: Proceedings of the 10th annual joint conference on Digital libraries. pp. 49–58 (2010)
8. Tang, J., Zhang, J., Yao, L., Li, J., Zhang, L., Su, Z.: Arnetminer: extraction and mining of academic social networks. In: Proceedings of the 14th ACM SIGKDD international conference on Knowledge discovery and data mining. pp. 990–998 (2008)
9. Uzzi, B., Mukherjee, S., Stringer, M., Jones, B.: Atypical combinations and scientific impact. Science **342**(6157), 468–472 (2013)
10. Walker, D., Xie, H., Yan, K.K., Maslov, S.: Ranking scientific publications using a model of network traffic. Journal of Statistical Mechanics: Theory and Experiment **2007**(6), 6010 (2007)
11. Wang, D., Song, C., Barabási, A.L.: Quantifying long-term scientific impact. Science **342**(6154), 127–132 (2013)
12. Yan, R., Tang, J., Liu, X., Shan, D., Li, X.: Citation count prediction: learning to estimate future citations for literature. In: Proceedings of the 20th ACM international conference on Information and knowledge management. pp. 1247–1252 (2011)