Spatial Scientometrics

Song Gao
Department of Geography, University of Wisconsin-Madison, Madison, WI, USA

Synonyms

Geospatial scientometrics

Definition/Introduction

The research field of scientometrics (or bibliometrics) is concerned with measuring and analyzing science, with the aim of quantifying a publication, a journal, or a discipline's structure, impact, change, and interrelations. The spatial dimension (e.g., location, place, proximity) of science has been added into account since research activities usually start from a certain region or several places in the world and then spread to other places, thus displaying spatiotemporal patterns. The analysis of spatial aspects of the science system is composed of spatial scientometrics (Frenken et al. 2009), which address the studies of geospatial distribution patterns on scientific activities, domain interactions, co-publications, citations, academic mobility, and so forth. The increasing availability of large-scale research metadata repositories in the big data age and the advancement in geospatial information technologies have enabled geospatial big data analytics for the quantitative study of science.

Main Research Topics

The earliest spatial scientometrics studies date back to 1970s. Researchers analyzed the distribution of worldwide science productivity by region and country. Later on, the availability of more detailed affiliation address information, and geographic coordinate data offers the possibility to investigate the role of physical distance in collaborative knowledge production. And the "spatial" dimension can refer to not only the "geographic space" but also the "cyberspace." The book *Atlas of Science: Visualizing What We Know* collected a series of visual maps in cyberspace for navigating the dynamic structure of science and technology (Börner 2010). According to the research framework for spatial scientometrics proposed by Frenken et al. (2009), there are at least three main topics addressed in this research domain: (1) *spatial distribution*, it studies the location arrangement of different scientific activities including research collaborations, publications, and citations across the Earth's surface. Whether geographic concentration or clustered patterns can bring advantages in scientific knowledge production is an important research issue in spatial scientometrics. (2) *Spatial bias*, it refers to those uneven spatial distributions on the scientific activities and their structure because of the limits on

L. A. Schintler, C.L. McNeely (eds.), *Encyclopedia of Big Data*,
https://doi.org/10.1007/978-3-319-32001-4_493-1

research funding, intellectual property, equipment, language, and so on. One prominent spatial bias is that researchers collaborate domestically more frequently than internationally, which might also be influenced by the number of researchers in a country. Another spatial bias is that collaborative articles from nearby research organizations are more likely to be cited than articles from research organizations further away within the same country. But there is a positive effect of international co-publications on citation impact compared with domestic co-publications. Such patterns might change with the increasing accessibility of crowdsourced or open-sourced bibliographic databases. Regarding researchers' trajectory or academic mobility patterns, they are also highly skew distributed across countries. Recent interests arise in the analysis of the origin patterns of regional or international conference participants. (3) *Citation impact*, it attracts much attention in the scientometrics studies. In academia, the number of citations is an important criterion to estimate the impact of a scientific publication, a journal, or a scientist. Spatial scientometrics researchers study and measure the geospatial distributions and impacts of citations for scientific publications and knowledge production.

Key Techniques and Analysis Methods

In order to analyze the geospatial distribution and interaction patterns of scientific activities in scientometrics studies, one important task is to get the location information of publications or research activities. There are two types of location information: (1) *place names* at different geopolitical scales (e.g., city, state, country, region) and (2) *geographic coordinates* (i.e., latitude and longitude). The place information can usually be retrieved from the affiliation information in standard bibliographic databases such as *Thomson Reuters Web of Science or Elsevier Scopus*. But the geographic coordinate information is not directly available in those databases. Additional processing techniques "georeferencing" which assigns a geographic coordinate to a place-name and "geocoding" which converts an address text into a geographic coordinate are required to generate the coordinate information for a publication, a citation, or a researcher. Popular geocoding tools include *Google Maps Geocoding API* and *ArcGIS Online Geocoding Service*.

After getting the coordinate information, a variety of statistical analysis and mapping/geovisualization techniques can be employed for spatial scientometrics analyses (Gao et al. 2013). A simplistic approach showing the spatial distribution pattern is to map the affiliation location of publications or citations or to aggregate the affiliation locations to the administrative places (e.g., city or country boundaries). Another method is to use the kernel density estimation (KDE) mapping to identify the "hotspot regions" in the geography of science (Bornmann and Waltman 2011). The KDE mapping has been widely used in spatial analysis to characterize a smooth density surface that shows the geographic clustering of point or line features. The two-dimensional KDE can identify the regions of citation clusters for each cited paper by considering both the quantity of citations and the area of geographical space, compared to the single-point representation which may neglect the multiple citations in the same location. Moreover, the concept of geographic proximity (distance) is widely used to quantify the spatial patterns of co-publications and citations. In addition, the socioeconomic factors that affect the scientific interactions have also been addressed. Boschma (2005) proposed a proximity framework of physical, cognitive, social, and institutional forms to study the scientific interaction patterns. Researchers studied the relationship between each proximity and citation impact by controlling other proximity variables. Also, the change of author affiliations over time adds complexity to the network analysis of universities. The approach with thematic, spatial, and similarity operators has been studied in the GIScience community to address this challenging issue.

When measuring the citation impact of a publication, a journal, or a scientist, traditional approaches purely counting the number of citations do not take into account the geospatial and temporal impact of the evaluating target. The spatial distribution of citations could be different even for publications with the same number of citations. Similarly, some work may be relevant and cited for decades, while other contributions only have a short-term impact. Therefore, Gao et al. (2013) proposed a theoretical and novel analytical spatial scientometrics framework which employs spatiotemporal KDE, cartograms, distance distribution curves, and spatial point patterns to evaluate the spatiotemporal citation impacts for scientific publications and researchers. Three geospatial citation impact indices ($S_{institution}$ index, S_{city} index, and $S_{country}$ index) were developed to evaluate an individual scientist's geospatial citation impact, which complement traditional nonspatial measures such as h-index and g-index.

Challenges in the Big Data Age

Considering the three V's characteristics (volume, velocity, and variety) of big data, there are many challenges in big-data-driven (spatial) scientometrics studies. These challenges require both computationally intensive processing and careful research design (Bratt et al. 2017). First, the author names, affiliation trajectory, and institution names and locations often need to be disambiguated and uniquely identified. Second, the heterogonous formats (i.e., structured, semi-structured, unstructured) of bibliographic data might be incredibly varied and cannot fit into a single spreadsheet or a database application. Moreover, the metadata standards are inconsistent across multiple sources and may change over time. All the abovementioned challenges can affect the validity and reliability of (spatial) scientometrics studies. The uncertainty or sensitivity analyses need to be included in the data processing and analytical workflows.

Conclusion

Spatial scientometrics involves the studies of spatial patterns, impacts, and trends of scientific activities (e.g., co-publication, citation, academic mobility). In the new era, because of the increasing availability of digital bibliographic databases and open data initiatives, researchers from multiple domains can contribute various qualitative, quantitative, and computational approaches and technologies into the spatial scientometrics analyses. The spatial scientometrics is still an infant interdisciplinary field with the support of spatial analysis, information science and statistic methodologies. New data sources and measurements to evaluate the excellence in the geography of science are emerging in the age of big data.

Further Readings

Börner, K. (2010). *Atlas of science: Visualizing what we know*. Cambridge: The MIT Press.

Bornmann, L., & Waltman, L. (2011). The detection of "hot regions" in the geography of science – A visualization approach by using density maps. *Journal of Informetrics, 5*(4), 547–553.

Boschma, R. (2005). Proximity and innovation: A critical assessment. *Regional Studies, 39*(1), 61–74.

Bratt, S., Hemsley, J., Qin, J., & Costa, M. (2017). Big data, big metadata and quantitative study of science: A workflow model for big scientometrics. *Proceedings of the Association for Information Science and Technology, 54*(1), 36–45.

Frenken, K., Hardeman, S., & Hoekman, J. (2009). Spatial scientometrics: Towards a cumulative research program. *Journal of Informetrics, 3*(3), 222–232.

Gao, S., Hu, Y., Janowicz, K., & McKenzie, G. (2013, November). A spatiotemporal scientometrics framework for exploring the citation impact of publications and scientists. In *Proceedings of the 21st ACM SIGSPATIAL international conference on advances in geographic information systems* (pp. 204–213). Orlando, Florida, USA: ACM.