

See discussions, stats, and author profiles for this publication at: <https://www.researchgate.net/publication/314403334>

Scientometrics for tech mining: An introduction

Article in *Scientometrics* · March 2017

DOI: 10.1007/s11192-017-2344-8

CITATIONS

0

READS

92

3 authors:



Yi Zhang

University of Technology Sydney

37 PUBLICATIONS 171 CITATIONS

SEE PROFILE



Alan L. Porter

Georgia Institute of Technology

338 PUBLICATIONS 5,453 CITATIONS

SEE PROFILE



Denise Chiavetta

Search Technology

4 PUBLICATIONS 4 CITATIONS

SEE PROFILE

Some of the authors of this publication are also working on these related projects:



Research connections [View project](#)



Advanced bibliometric methods and applications in science, technology, innovation, and policy research [View project](#)

All content following this page was uploaded by [Alan L. Porter](#) on 15 March 2017.

The user has requested enhancement of the downloaded file. All in-text references [underlined in blue](#) are added to the original document and are linked to publications on ResearchGate, letting you access and read them immediately.

Scientometrics for Tech Mining: An Introduction

Yi Zhang ¹, Alan L. Porter ^{2,3}, Denise Chiavetta ³

¹Centre for Artificial Intelligence, Faculty of Engineering and Information Technology, University of Technology Sydney, Australia

²Technology Policy and Assessment Centre, Georgia Institute of Technology, Atlanta, USA

³Search Technology, Inc., Norcross, USA

Email: yi.zhang@uts.edu.au; alan.porter@isye.gatech.edu; dchiavetta@searchtech.com;

Tech mining, a special form of “Big Data” analytics, aims to generate Competitive Technical Intelligence (CTI) using bibliometric and text mining software (e.g., VantagePoint, TDA), as well as other analytical and visualization applications, for analyses of Science, Technology & Innovation (ST&I) information resources. The big data boom challenges tech mining. How to advance tech mining techniques to devise CTI from scalable ST&I data and how to address real-world ST&I concerns using such CTI have emerged as issues. Under this circumstance, the development of tech mining has evolved in two directions:

- 1) combining scientometrics with information technologies (e.g., probabilistic methods, machine learning, network analysis, and prediction models) are increasingly used together for tech mining studies; and
- 2) the establishment of cross-disciplinary networks, with data analysts, academic researchers, policymakers, and industrial partners, to promote tech mining applications.

Such analyses offer practical insights in R&D program choices, science policy, innovation management, entrepreneurship, and national strategy.

This special issue focuses on research presented at the sixth Global Tech Mining Conference held in Valencia, Spain, in September, 2016. The twelve submissions relate to either tech mining-driven methodological innovation in scientometrics or scientometric applications for tech mining. We spotlight points of special interest.

Tech mining-driven methodological innovation

Oriented to a broad range of issues in tech mining, methodological innovation concern either the development of advanced analytic models or the recombination of existing methods. Such efforts provide novel solutions for handling new problems raised in the big data age or improving the performance of some old solutions. In this special issue methodological innovations include: micro-level scientometrics, ST&I emergence indicators, and interdisciplinarity measures.

Micro-level scientometrics emphasize the discovery of detailed interactions between/within single entities (e.g., organizations and individuals). Contrasting to emphases of large-scale data analytics, the information on individuals is usually hidden in a huge and noisy data pool. Two specific research issues are raised here: how to disambiguate author names from scientific documents with no or limited external knowledge and how to develop a lexical analytic model for a small ST&I data set.

1) Author name disambiguation (AND): the paper by Abdulhayoglu and Thijs, *Use of ResearchGate and Google CSE for author name disambiguation*, concentrates on the scholar social network platform ResearchGate. They propose a method to handle the AND issue by employing the Google Custom Search Engine as a complementary tool to access authors' web pages, in which a supervised learning classifier and a clustering technique are combined for authors' information retrieval. The paper by Han et al., *Semantic fingerprints-based author name disambiguation in Chinese documents*, integrates co-author features, institutional features, and text fingerprints to minimize the aid of external knowledge in disambiguating author names in Chinese documents.

2) Lexical analysis for nano-level scientometrics: the paper by Glänzel, Heeffer, and Thijs, *Lexical analysis of scientific publications for nano-level scientometrics*, conducts a pilot study monitoring structural changes in an 18-paper dataset, providing a novel way to grasp the dynamics of research topics at a micro level by detecting similarities between word distributions.

ST&I emergence indicators attract growing interest for tech mining, and especially for emerging technology-oriented ST&I studies. Compared with traditional expert knowledge-based approaches, the engagement of scientometrics may eliminate the bias of subjectivity and improve the ability to analyze large-scale data sets for emergence.

The paper by Carley et al., *A measure of staying power: Is the persistence of emergent concepts more significantly influenced by technical domain or scale?* develops an algorithm to calculate four specific features (i.e., novelty, persistence, community, and growth) of technical emergence by using a number of bibliometric indicators. It offers a series of case studies to evaluate the influence of technical domain or scale on the persistence of emergent research patterns.

The paper by Zhang et al., *An entropy-based indicator system for measuring the potential of patents in technological innovation: Rejecting moderation*, raises the "moderation" issue in traditional indicator systems and weighting approaches. Focusing on patents in the United States Patent and Trademark Office database, they construct an entropy-based indicator system to measure the potential of patents in technological innovation. A collaborative filtering technique iteratively ranks patents.

Interdisciplinary measures offer significant insight to ST&I management and science policy. Such interests become crucial these days as interactions between/within disciplines are increasingly dynamics and complicated.

The paper by Wang et al., *Measuring interdisciplinarity of a research system: Detecting distinction between publication categories and citation categories*, constructs a model to measure the interdisciplinarity of a research system, considering the deviations caused by exceptional categories in traditional bibliometric indicators (i.e., publication categories and citation statistics).

Scientometric applications for tech mining

Exploring specific real-world ST&I cases is vital for tech mining. The dynamics and uncertainty of modern society influence ST&I internally and externally. How to investigate cases from social, economic, political, and technological perspectives challenges existing models. In this special issue applications for tech mining include: National and industrial R&D management, and ST&I profiling.

National and industrial R&D management is a practical area to apply tech mining. Oriented to actual requirements in a given ST&I sector, CTI findings can inform relationships between science and strategy, collaboration among research communities, and shifts in innovation management systems.

The paper by Bakhtin et al., *Trend monitoring for linking science and strategy*, discusses the interactive relationships between science & technology and policy & strategy in-depth by conducting a case study of the global agriculture and food sector, with interesting text analytics and trend analyses.

The paper by Sampaio et al., *Network analysis to support public health: Evolution of collaboration among leishmaniasis researchers*, explores the benefit of tech mining in analyzing collaboration dynamics within the research communities of leishmaniasis. Interestingly, this paper integrates lexical analysis and network analysis in a coherent framework, and a link prediction technique demonstrates good capability in identifying such relationships.

The paper by Vicente-Gomila et al., *Discovering shifts in competitive strategies in probiotics, accelerated with TechMining*, applies a semantic TRIZ-based patent portfolio analysis to the global probiotic sector. They observe a shift of technological strategies in key probiotic and prebiotic ingredients. This appears to offer significant CTI for tracking technological emphases of target enterprises.

The paper by Garechana et al., *Effects of innovation management system standardization on firms: Evidence from text mining annual reports*, specifically focuses on a pioneer innovation management standard, investigating its influence on enterprises' ST&I activities. Co-word networks and emotion analyses generate CTI from annual reports of selected enterprises.

ST&I profiling can be a basic task in tech mining, seeking CTI for specific ST&I targets (e.g., data sources, research topics, and single/multiple entities) to support decision making in related R&D management and science policy. The use of scientometric techniques creates additional dimensions to achieve such targets from a wider and deeper perspective.

The paper by Fursov and Kadyrova, *How the analysis of transitional references in knowledge networks and their centrality characteristics helps in understanding the genesis of growing technology areas*, introduces a co-citation network analysis to understand the development pathways of two research arenas -- “advanced technology” and “emerging technology.”

The paper by Huang et al., *Early insights on the Emerging Sources Citation Index (ESCI): An overlay map-based bibliometric study*, sheds light on the new ESCI database. It constructs an analytic framework to integrate traditional bibliometrics and a science overlay map technique. They address specific concerns about the coverage of journals and the ability of detecting emerging fields in the new database.

In conclusion, the issue extends consideration of interactions between scientometrics and tech mining vertically and horizontally. On one hand, advanced information technologies are introduced to enhance the ability of scientometrics to support tech mining, and, on the other hand, real-world ST&I activities present new challenges for tech mining studies. In this special issue, we highlight the engagement of new techniques, concepts, and applications, such as the use of machine learning techniques for information retrieval, the application of network analysis in exploring complicated relationships, the design of emerging indicators for evaluating ST&I activities, emphases on individual-based bibliometric study, and mining of new ST&I data resources.

We welcome suggestions for the further development of tech mining, especially on how to foster interactions between scientometrics and tech mining.