Guest Editorial: Tech Mining for Engineering Management: An Introduction

TECH *Mining* refers to "the application of text mining tools to science and technology information, informed by the understanding of technological innovation processes in item 1) of the Appendix." Since the 1st Global Tech Mining (GTM) Conference in 2010, the GTM community has established great connections between tech mining and a broad range of research domains, particularly bibliometrics, text mining, and technology management. Endeavours of the community touch on benefits of the big data boom, the artificial intelligence (AI) age, advanced analytical and visualization tools, novel philosophical concepts, and reshaped managerial models. Such studies provide new visions for studying science, technology & innovation (ST&I), and gaining insights in competitive technical intelligence (CTI) and R&D management in diverse levels (e.g., national strategic management, science policy, and entrepreneurship). Previous collections of GTM conferences particularly addressed the engagement of novel bibliometric approaches in tech mining in item 2) of the Appendix and concentrated on the measurement & prediction of technological emergence in item 3) of the Appendix. This special issue draws on papers presented at the *2018 Global Tech Mining Conference* in Leiden, the Netherlands, and relevant external submissions.

We highlight the following interests.

1) *Maximizing the potential of traditional and novel text data:* Tapping vital traits while minding limitations and matching analytical aims, such as ways to address new data sources (e.g., web scraping and social media cumulations) and transitioning from database mining to real-time streaming text analytics.
2) *Advancing and integrating methods:* Topics of interest include measuring and forecasting technological emergence via advanced information technologies (e.g., natural language processing, social network analysis, and topic models), ways to demonstrate multidimensional indicators of R&D and innovation activities, and means to combine quantitative and qualitative approaches in practical case applications.
3) *Translating analyses to useful intelligence, with informative indicators and compelling visualizations:* For example, case studies that offer testable projects (e.g., specific and explicit) to enable future revisiting to validate methods and capabilities of visual analytics both for analysts and for end-users.

Digital Object Identifier 10.1109/TEM.2021.3061862

We introduce 16 submissions collected in this special issue from the following two aspects: advanced tech mining methods and practical tech mining.

I. Advanced Tech Mining Methods

Inspired by the needs of analyzing ST&I sources, this collection emphasizes the development of novel approaches incorporating text mining (including natural language processing techniques and topic models) and/or social network analytics with bibliometrics.

A. Text Mining

Text mining techniques could be one of tech mining's technical backbones, and within this scope, our special issue collects papers specifically focusing on similarity measurements in diverse scenarios, in which document embedding and topic models are involved.

Topic models have been an interest of our community for decades, providing a solution of soft clustering for topic extraction and further analyses. In the paper entitled *Identify Topic Relations in Scientific Literature Using Topic Modeling*, Chen *et al.* proposed a vivid approach to measure the semantic similarities among topics generated from a latent Dirichlet allocation (LDA) model and such relationships were further validated via established collaborations between involved authors. Wang *et al.* targeted the technological complementarities between enterprises and exploited a hierarchical LDA model to measure these relationships in the paper entitled *Measuring Technology Complementarity between Enterprises with an hLDA Topic Model*.

Rising interests are also credited to embedding techniques in representing words, phrases, and documents, and to their applications in dealing with ST&I sources. The paper entitled *Bilingual Textual Similarity in Scientific Documents* aims to measure similarities between scientific documents in different languages, in which Kawamura and Egami proposed a solution of similarity measurements incorporating certain popular models in text analytics, such as document embedding, semantic graph, and BM25 (for ranking features).

B. Social Network Analytics

Despite a decade or longer history of social networks in social sciences, social network analytics are not emphasized by the GTM community until recently. However, as reflected in our special issue, it is impressive to observe ambitious integrations between social network analytics and actual needs in tech mining (e.g., discovering innovation clusters and research communities), with diverse bibliometric indicators [e.g., coauthorships, and subject-action-object (SAO) structures].

In the paper entitled *Overlapping Community Discovery for Identifying Key Research Themes*, Huang *et al.* proposed an approach of network analytics for recognizing key nodes overlapping multiple communities in a network, in which they exploited word embedding techniques for representing keywords and a cluster percolation method for detecting communities in a "soft-clustering" way. Practically, the paper entitled *Proximity and Knowledge, Business, Geographic Communities in Innovation Cluster: Evidence from China* draws attention to the proximity of innovation clusters with multidimensional indicators such as geography, business, and knowledge, and, thus, Zhou *et al.* constructed a heterogenous tri-layer network and measured such proximity within and between communities detected in the network.

As an intriguing indicator in tech mining, SAO analysis always creates challenges and opportunities from a perspective of "deep" semantic analysis. In the paper entitled *Technology Opportunity Analysis: Combing SAO Networks and Link Prediction*, Han *et al.* introduced link prediction approaches to an SAO network and predicted connections between subjects and objects for informing potential technological opportunities.

II. Practical Tech Mining

Tech mining highlights its capabilities in handling ST&I issues. This collection contains a set of contributions on analyzing technological opportunities within an integrative framework, measuring technological emergence with novel indicators, and CTI-oriented patent analysis and science–technology linkage discovery.

A. Technology Opportunity Analysis

Technology opportunity analysis, as an analytic framework, can be dated back to the 1990s in item 4) of the Appendix, and it established tech mining's fundamental bases in concepts and techniques. However, benefitting from advanced information technologies and their interactions with bibliometrics, the GTM community has started to refine this framework with new visions, methods, and indicators.

In the paper entitled *Parallel or Intersecting Lines? Intelligent Bibliometrics for Investigating the Involvement of Data Science in Policy Analysis*, Zhang *et al.* designed a framework of intelligent bibliometrics to answer not only the WHO, WHERE, WHEN, and WHAT questions, but also the question of how scientific topics evolve over time, in which they particularly exploited an approach of streaming data analytics to identify predecessor–descendant relationships in scientific topics. While Zhang *et al.* draw the evolutionary pathways in a manner of text mining, Huang *et al.* introduced citation networks and main path analysis to the same task and specifically investigated technological evolution within a framework of technology life cycle in the paper entitled *Exploring Technology Evolution Pathways to Facilitate Technology Management: From a Technology Life Cycle Perspective*. Compared to tracking technological evolution in the papers above, Park *et al.* defined technological opportunities as a set of future-oriented topics in the paper entitled *Technological Opportunities Discovery for Safety Through Topic Modeling and Opinion Mining in the Fourth Industrial Revolution: The Case of Artificial Intelligence*. In the case of artificial intelligence, they exploited topic models and sentiment analysis in this framework for identifying risk-related topics from a technological perspective, rather than traditional societal and/or political scenarios.

Ozcan *et al.* extended tech mining's framework to a novel dimension—human resource mining—in the paper entitled *Human Resource Mining for Examination of R&D Progress and Requirements*, and specifically identified R&D skillsets and capabilities, technological trajectories, and national requirements and demands by analyzing the text of job advertisements.

B. Measuring Technological Emergence

Forecasting technological emergence is the key interest of our special issue for the 2017 Global Tech Mining Conference in item 3) of the Appendix, but the community retains such an interest and, in this collection,, we focus on the measurement of technological emergence with novel indicators.

Zhou *et al.* established a hybrid approach that incorporates a set of popular text mining techniques (e.g., topic analysis, sentiment analysis, and SAO analysis) within the scope of innovation pathways in the paper entitled *Identifying and Assessing Innovation Pathways for Emerging Technologies: A Hybrid Approach based on Text Mining and Altmetrics*, and a set of altmetric indicators were proposed for measuring potential impacts of emerging technologies. Li measured technological emergence from the perspective of operations research in the paper entitled *Capturing the Risk Signals for a Specific Emerging Technology: An Integrated Framework of Text Mining*, in which he conceptually defined the basic attributes for measuring the risk of emerging technologies with bibliometric indicators, and composed the outcome via a multiattribute decision-making approach.

C. CTI-Oriented Patent Analysis

Patent data have long been one of the key ST&I sources for gaining CTI, and facilitating patent statistics, including unique attributes [e.g., international patent classification codes, claims, grant lag, and patent/nonpatent references] provide new angles to uncover insights from not only technological, but also economic and legal, perspectives.

In the paper entitled *Patent Value Analysis using Deep Learning Models – The Case of IoT Technology Mining for the Manufacturing Industry*, Trappey *et al.* constructed an indicator system with measurable patent attributes and applied deep neural networks to recognize the patterns of patent values, and thus, identified a set of valuable patents and assignees holding such valuable technologies in the global Internet of Things sector. While the above paper investigated patent value using cutting-edge AI techniques, Kronemeyer *et al.* conducted patent analysis within a framework of technology management in the paper entitled *Monitoring Competitors' Innovation Activities:*

Analyzing the Competitive Patent Landscape Based on Semantic Anchor Points, in which the authors specifically focused on the activities of patent assignees and proposed a hybrid approach to capture such activities by combining qualitative approaches (e.g., expert knowledge) and quantitative approaches (e.g., text mining).

D. Discovering Science-Technology Linkages

Discovering the linkages between science and technology is a practical task in CTI and attracts increasing attention these days. Conducting this task requires not only the integration of scientific publications and patents but also the identification of the connections between the two sides.

In the paper entitled *From Research to Industry: A Quantitative and Qualitative Analysis of Science-Technology Transferences and Emergence Patterns in Bioremediation*, Garechana *et al.* tracked the knowledge transfer between research activities and technology development via text mining techniques, including science maps and SAO analysis. Trela *et al.* handled these science-technology linkages from a new angle—i.e., seeking industrial partners for research institutions, and in the paper entitled *How to Find New Industry Partners for Public Research: A Classification Approach*, they proposed a classifier with a set of classification approaches (e.g., support vector machine, random forest, and logistic regression) and examined their performance through an integrated dataset with industry-partner data, company data, and public funding data.

III. Conclusion

This special issue collects the contribution of the GTM community on developing advanced tech mining techniques by introducing new analytic approaches and tools, and on applying tech mining to handle real-world ST&I issues, with novel concepts, frameworks, and indicators. We observe that the combination of quantitative and qualitative methodologies is still widely employed in practical use, but we also notice the rising interests of introducing artificial intelligence techniques to transfer traditional semiautomatic approaches to automatic approaches. This change reflects how the GTM community is adapting to this AI age and, in fact, it also indicates the potential "evolution" of tech mining's technical backbones has been occurring. Thus, in the venue of tech mining we anticipate multidisciplinary interactions among ST&I management, computer science, and information science could be more radical and extensive, and how to facilitate such an interaction in a rigorous and coherent framework might require further attentions.

Acknowledgment

The work of Yi Zhang was supported by the Discovery Early Career Researcher Award granted by the Australian Research Council under Grant DE190100994. The work of Ying Huang was supported from the National Natural Science Foundation of China under Grant 72004169. The work of Alan L. Porter and Denise Chiavetta was supported by the U.S. National Science Foundation Award #1759960 to Search Technology, Inc., and Georgia Tech.

YI ZHANG, *Guest Editor*
Australian Artificial Intelligence Institute
Faculty of Engineering and Information Technology
University of Technology Sydney
Ultimo, NSW 2007, Australia

YING HUANG, *Guest Editor*
School of Information Management
Wuhan University
Wuhan 430072, China

Department of Management, Strategy and Innovation, and Centre for R&D Monitoring
KU Leuven
3000 Leuven, Belgium

DENISE CHIAVETTA, *Guest Editor*
Search Technology, Inc.
Norcross, GA 30092 USA

Alan L. Porter, *Guest Editor*
Search Technology, Inc.
Norcross, GA 30092 USA

Program in Science, Technology and Innovation Policy
Georgia Institute of Technology
Atlanta, GA 30332 USA

References

[1] A. L. Porter and S. W. Cunningham, *Tech Mining: Exploiting new Technologies for Competitive Advantage*. Hoboken, NJ, USA: Wiley, 2004.

[2] Y. Zhang, A. L. Porter, and D. Chiavetta, "Scientometrics for tech mining: An introduction," *Scientometrics*, vol. 111, no. 3, pp. 1875–1878, 2017.

[3] Y. Zhang, A. Porter, D. Chiavetta, N. C. Newman, and Y. Guo, "Forecasting technical emergence: An introduction," *Technol. Forecast. Soc. Change*, vol. 146, no. C, pp. 626–627, 2019.

[4] A. L. Porter and M. J. Detampel, "Technology opportunities analysis," *Technol. Forecast. Soc. Change*, vol. 49, no. 3, pp. 237–255, 1995.