Characterizing the Potential of Being Emerging Generic Technologies: A Bi-Layer Network Analytics-based Prediction Method

Yi Zhang[1], Yihe Zhu[2], Lu Huang[2], Guangquan Zhang[1] and Jie Lu[1]

[1] *yi.zhang@uts.edu.au; guangquan.zhang@uts.edu.au; jie.lu@uts.edu.au*
Centre for Artificial Intelligence, Faculty of Engineering and Information Technology, University of Technology Sydney (Australia)

[2] *zhuyihe12@foxmail.com; huanglu628@163.com*
School of Management and Economics, Beijing Institute of Technology (China)

Abstract

Despite tremendous involvement of bibliometrics in profiling technological landscapes and identifying emerging topics, how to predict potential technological change is still unclear. This paper proposes a bi-layer network analytics-based prediction method to characterize the potential of being emerging generic technologies. Initially, based on the innovation literature, three technological characteristics are defined, and quantified by topological indicators in network analytics; a link prediction approach is applied for reconstructing the network with weighted missing links, and such reconstruction will also result in the change of related technological characteristics; the comparison between the two ranking lists of terms can help identify potential emerging generic technologies. A case study on predicting emerging generic technologies in information science demonstrates the feasibility and reliability of the proposed method.

Introduction

An early definition of emerging generic technologies can be traced back to the early 1990s, highlighting technologies that enable revolutionary impacts on the economy and society (Martin 1995), and Maine and Garnsey (2006) moved down on the line and specified the 'generic' nature as benefits on a wide range of sectors and the 'emerging' nature as the potential for innovation. Even though emerging generic technologies conceptually contain overlaps with emerging technologies (Rotolo et al. 2015), it is clear that, compared to technologies that hold disruptive powers to a given technological area – e.g., dye sensitized solar cells (Zhang et al. 2014b), the exploitations and applications of emerging generic technologies would create values for fostering innovation in broad disciplines (Coccia 2017) – e.g., nanotechnology (Maine & Thomas 2017). During the past decades, investigations on how to measure the impacts of emerging generic technologies on accelerating the economic growth (Bresnahan & Trajtenberg 1995; Crafts 2004; Qiu & Cantwell 2018) and how to transfer technological breakthroughs into impactful innovations (Sinfield & Solis 2016; Appio et al. 2017) have been conducted in the innovation literature.

The engagement of bibliometrics on assisting in the management of technology has been well observed, e.g., profiling a given technological area (Guo et al. 2010; Chakraborty et al. 2015), identifying emerging topics in science and technology (Glänzel & Thijs 2012; Small et al. 2014), and tracking the pathways of technological change (Zhou et al. 2014; Hou et al. 2018). The use of advanced information technologies, e.g., topic models, streaming data analytics, and machine learning techniques, greatly strengthens the capability of traditional bibliometrics in handling large-scale data analytics (Ding & Chen 2014; Klavans & Boyack 2017), discovering hidden relationships (Zhang et al. 2017; Zhang et al. 2018), and visualizing complicated landscapes and structures (Börner et al. 2012; Suominen & Toivanen 2016).

Even though it has been a long time since the use of network analytics in social science (Borgatti et al. 2009), network analytics was introduced to bibliometric studies in the late 2000s, which were initially used to investigate research collaborations and disciplinary interactions through analyzing bibliographic couplings (Yan et al. 2009; Yang et al. 2010). Its effective combination with citation networks has attracted great attentions on identifying emerging topics and evaluating research impacts (Takeda & Kajikawa 2009; Yan 2015). Such advantages have been applied for predicting emerging technologies (Érdi et al. 2013) and discovering technological opportunities (Park & Yoon 2018). However, despite recognitions from the both communities, concerns are still raised, e.g., bibliometrics is insufficient on 'characterizing the potential of what is detected to be emerging' (Rotolo et al. 2015). Additionally, with the rapid development of natural language processing (NLP) techniques, co-word statistics provide a new angle for bibliometrics, but how to explore insights based on semantics retrieved from co-word-based networks is still elusive. Apparently, such insights are complementary with citation networks.

Aiming to address these concerns, this paper is to propose a bi-layer network analytics-based prediction method for characterizing the potential of being emerging generic technologies. Initially, we refer studies conducted by Maine and Garnsey (2006) and Rotolo et al. (2015), and consider emerging generic technologies as novel and fast-growing technologies with prominent impacts on a relative broad range of disciplines. A co-authorship network and a co-term network are constructed and integrated as a bi-layer network to represent the content of involved disciplines/technologies, and indicators for profiling the topological structure of networks are introduced to identify technological characteristics from three aspects – i.e., fundamentality, connectivity, and externality. Further, a link prediction approach is incorporated to calculate weights of all links (including missing links) in the both networks, and such reconstruction would be the key to capture a potential characteristic change of involved technologies. Thus, investigating such change could be the way of characterizing the potential of being emerging generic technologies. We then demonstrate the feasibility and reliability of the proposed method through a case study, which predicts emerging generic technologies in information science disciplines by analyzing 17,882 articles published in 15 selected journals in the field between Jan 1, 2000 and Dec 31, 2016.

The rest of this paper is organized as follows: The Methodology section describes the details of the proposed bi-layer network analytics-based prediction method, and the Case Study section follows, presenting the data, results, and empirical insights derived from the case. We then conclude our study and outline potential future directions.

Methodology

The research framework of the proposed bi-layer network analytics-based prediction method for characterizing the potential of being emerging generic technologies is given in Fig. 1.

Technological characteristics

Following the studies conducted by Maine and Garnsey (2006) and Rotolo et al. (2015), we identify new technologies that can be fundamentally applied to a broad range disciplines, with capabilities of connecting diverse technological areas and adaptively transferring among enterprises, as emerging generic technologies, and the characteristics of emerging generic technologies are specifically defined from the following three perspectives:

- Fundamentality is to measure *whether this technology can be applied to a broad range of sectors, disciplines, or research areas*.
- Connectivity is to measure *whether this technology is sharing close relationships with other technologies in the same or different technological areas*.

- Externality is to measure *whether this technology is involved and can be transferred among diverse enterprises and research groups*.

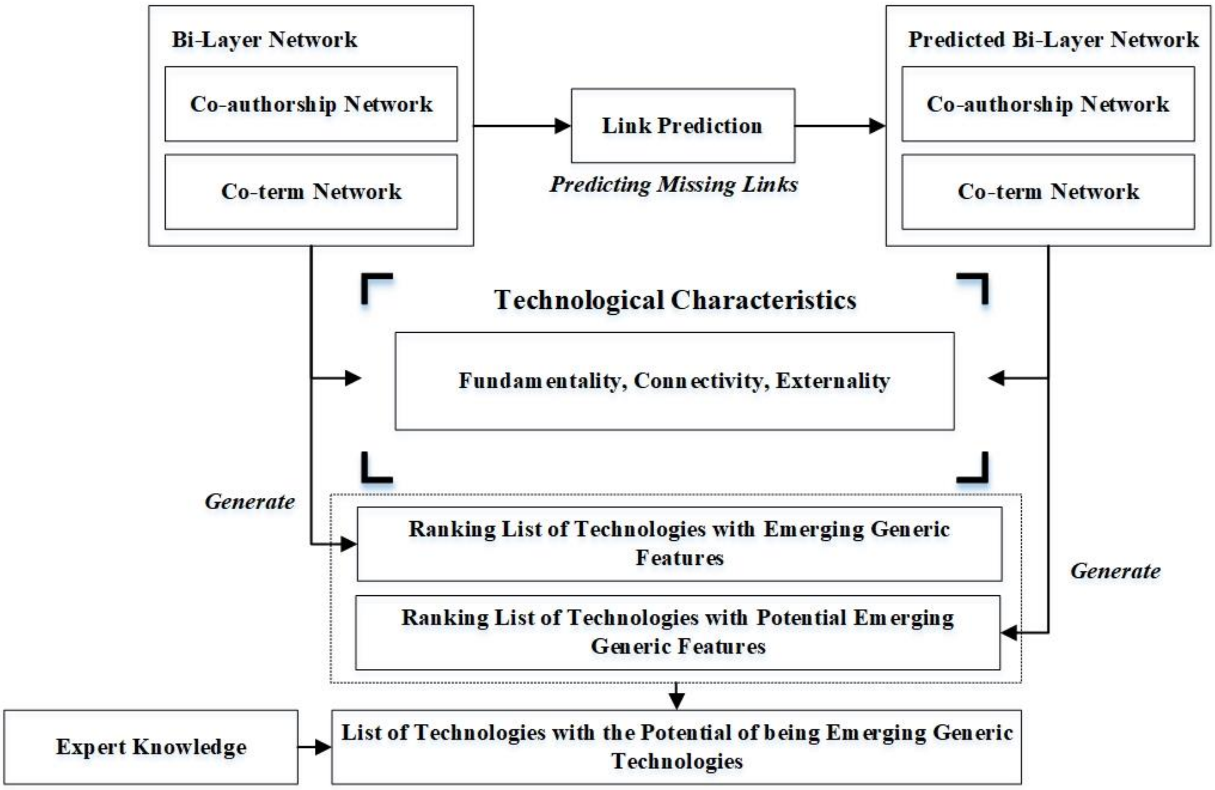

Figure 1. Research framework of the bi-layer network analytics-based prediction method.

Bi-layer network analytics

A bi-layer network includes a co-term network and a co-authorship network. We denote $N = \{(V^t, E^t), (V^a, E^a), E^{at}\}$ as a bi-layer network, in which (V^t, E^t) and (V^a, E^a) are the sets of nodes and links in the co-term network and the co-authorship network respectively and E^{at} is the set of links between the two networks. A sample of a bi-layer network is given in Fig. 2.

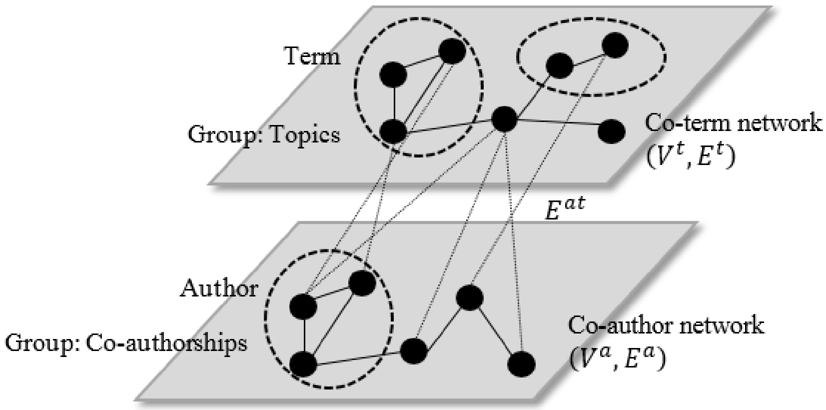

Figure 2. Sample of a bi-layer network.

Specifically, the co-term network is generated based on the co-occurrence statistics of terms derived from the title and abstract fields of collected records, and the co-authorship network is based on the statistics of co-authorship behaviors. The both networks are non-direct graphs, in which 1) each node represents either a term or an author and 2) each link represents the co-occurrence/co-authorship relationships between connected nodes and is weighted by the frequency of such co-occurrence/co-authorship. Significantly, the authorships of terms are used to be the links between the two layers – i.e., the co-term network and the co-authorship network.

When considering each term represents a technological component (e.g., materials, functions, manufacturing processes, and applications), we apply network analytics for investigating the topological structures of the bi-layer network to quantify the three characteristics of emerging generic technologies.

- Fundamentality

The fundamentality of a technology is to measure the breadth and depth of a technology's influence in given technological areas. In the co-term network, centrality, a traditional indicator of measuring network topological structures (Freeman 1977; Freeman 1978), is exploited to quantify the value of the fundamentality. Since a number of centrality-based indicators have been developed for different emphases, three forms of centrality are involved in this study:

1) Degree Centrality – the degree of a node, reflecting the breadth of its potential influence. The degree centrality of node v_i^t in the co-term network can be calculated as follows:

$$DC(v_i^t) = \frac{\sum_{j=1}^{|V^t|} w_{v_i^t, v_j^t}}{|V^t| - 1}$$

where $|V^t|$ is the number of nodes in the co-term network and $w_{v_i^t, v_j^t}$ is the weight of the link between node v_i^t and node v_j^t.

2) Closeness Centrality – the closeness between a node and other nodes in the same network, reflecting its professionalism in the given area. The closeness centrality of node v_i^t can be calculated as follows:

$$CC(v_i^t) = \frac{|V^t| - 1}{\sum_{j=1}^{|V^t|} d_{v_i^t, v_j^t}}$$

where $d_{v_i^t, v_j^t}$ is the shortest distance between node v_i^t and node v_j^t.

3) Between Centrality – the number of the shortest paths crossing a node, reflecting its role in a cross area. The between centrality of node v_i^t can be calculated as follows:

$$BC(v_i^t) = \frac{2 \sum \frac{\sigma(v_i^t)_{v_s^t, v_p^t}}{\sigma_{v_s^t, v_p^t}}}{(|V^t| - 1)(|V^t| - 2)}, v_i^t \neq v_s^t \neq v_p^t$$

where v_s^t and v_p^t are two different nodes in the network, $\sigma_{v_s^t, v_p^t}$ represents the number of the shortest paths between nodes v_s^t and v_p^t, and $\sigma(v_i^t)_{v_s^t, v_p^t}$ is the number of the shortest paths between nodes v_s^t and v_p^t, crossing node v_i^t.

The three forms of centrality exploit different topological structures – e.g., degree centrality concentrates on the number of neighbor nodes, closeness centrality highlights the capability of connecting other nodes, and between centrality emphasizes the importance of a node in the communication of a network. Thus, aiming to involve all these benefits, we calculate the fundamentality of a node $F(v_i^t)$ as the average value of the three indicators.

- Connectivity

The connectivity of a technology is considered as its relationships with other technologies and technological groups, indicating its capability of involving diverse sectors, disciplines, and technological areas. In the co-term network, 1) initially, a smart local moving algorithm (Waltman & Van Eck 2013) is applied for community detection – i.e., identifying technological groups G^t; and then, 2) we calculate the connectivity $C(v_i^t)$ between node v_i^t and its community as follows:

$$C(v_i^t) = \frac{\sum_{j=1}^{|G^t(v_i^t)|} w_{v_i^t, v_j^t}}{|G^t(v_i^t)|}$$

where $|G^t(v_i^t)|$ represents the number of nodes in the community to which node v_i^t belong.

- Externality

The externality of a technology takes both technologies and their owners into considerations – i.e., if a technology is owned by more than one owner (e.g., enterprises and research institutions) and can be easily transferred between those owners, or even between different sectors, we consider this technology is generic in the related fields. Thus, both the co-term network and the co-authorship network in a bi-layer network will be exploited for measuring the externality of a node $E(v_i^t)$ as follows:

$$L(v_m^a) = \sum_{n=1}^{|V^a|} w_{v_m^a, v_n^a}$$

$$E(v_i^t) = \sum_{n=1}^{|V^a|} w_{v_i^t, v_n^a} \times L(v_n^a)$$

where v_m^a is a node in the co-authorship network and $|V^a|$ is the number of nodes in the co-authorship network.

Despite some weighting approaches, e.g., entropy-based and standard deviation-based weights, we decide to use a 3D map to visualize values of the three technological characteristics, highlighting distinctive values based on diverse requirements and preferences.

Link prediction

A common neighbors (CN)-based link prediction approach (Newman 2001) is exploited to weight all links (including missing links) in the bi-layer network, and such reconstruction of the network would represent possible connections between terms and potential collaborations between authors in future. The basic assumption of the CN-based approach is that if two unlinked nodes have many common neighbors, it is highly possible that a link will appear between the two nodes. Thus, the CN value of each link can be calculated as follows:

$$CN(v_x, v_y) = \sum_{z=1}^{|V(v_x, v_y)|} \left(w_{v_x, v_z} + w_{v_y, v_z}\right)$$

where v_x and v_y are two different and unlinked nodes in a bi-layer network (either the co-term network or the co-authorship network) and $|V(v_x, v_y)|$ is the set of nodes in the bi-layer network, which connect v_x and v_y.

The output of this link prediction approach is a ranking list of all links in the bi-layer network, including missing links in the current network. Thus, a predicted bi-layer network will be generated, reflecting potential technological change in the near future.

Identification of emerging generic technologies

According to the technological characteristics, a ranking list (List A) of technologies with emerging generic features will be generated based on a bi-layer network. With the exploitation of link prediction approaches, missing links in the bi-layer network will be created and existing links will be re-weighted, i.e., a predicted bi-layer network is constructed. Apparently, the change of the topological structure of the existing network will result in the change of the technological characteristics of related technologies, and thus, a new ranking list (List B) will be generated. Therefore, comparing the two lists respectively generated by the two bi-layer networks will help characterize the potential of being emerging generic technologies. Several selection criteria will be highlighted, including:

- A technology only appears in List B and with a high rank;
- Compared to List A, the rank of a technology in List B dramatically increase;
- A technology appears in the top rank of the both lists;

Case Study: What are emerging generic technologies in information science?

It would likely be arguable that information science can only represent an individual discipline and it is critical to identify emerging generic technologies from such one discipline rather than a broad range of disciplines. Our consideration here is that information science has been spearheading a cross-disciplinary direction that bridges fundamental studies (e.g., mathematics, physics, and computer science) with real-world needs raised in disciplines of social science. Therefore, it would be interesting to identify emerging generic technologies from such a cross-disciplinary area, which would originate from other disciplines but build up the foundations of information science and create extensive impacts on and out of the discipline. We followed the search strategy proposed by Hou et al. (2018) and selected 15 journals and conference proceedings, covering 17,445 records between January 1, 1996 and December 31, 2016.

Table 1. List of selected journals

Journal Name	*Journal Name*
Annual Review of Information Science and Technology	Library Resources & Technical Services
Information Processing & Management	Program: Automated Library and Information Systems
Journal of the Association for Information Science and Technology	Information Research
Journal of Documentation	Journal of Informetrics
Journal of Information Science	Research Evaluation
Library & Information Science Research	The Electronic Library
ASIS&T Annual Meeting Proceedings	Information Technology and Libraries
Scientometrics	

Note that the table only lists the current names of selected journals, but we fully considered their previous names when collecting data.

We combined the title and abstract fields of the 17,445 records and retrieved 213,031 terms by a natural language processing (NLP) function integrated in the VantagePoint[1]. A term clumping process (Zhang et al. 2014a) was applied for data cleaning by removing noise and consolidating synonyms, and the stepwise results are given in Table 2. The 25,359 terms were used for constructing the co-term network.

Table 2. Stepwise results of term clumping

Step	*Description*	*#Terms*
0	Raw terms retrieved by the NLP technique;	213,031
1	Remove single-word terms, e.g., "information";	189,111

[1] VantagePoint is a software platform for bibliometrics-based text analytics and knowledge management, owned by Search Technology Inc. More details can be found at the website: www.vantagepoint.com.

2	Remove terms starting/ending with non-alphabetic characters, e.g., “step 1” and “1.5 m/s”;	180,209
3	Remove meaningless terms, e.g., pronouns, prepositions, and conjunctions;	175,488
4	Remove common terms in scientific articles, e.g., “research framework”;	157,041
5	Consolidate synonyms based on expert knowledge, e.g., “co-word analysis” and “word co-occurrence analysis”;	135,967
6	Consolidate terms with the same stem, e.g., “information system” and “information systems”	109,115
7	Remove terms appearing less than 3 times;	25,359

Note that: 1) Expert knowledge in Step 5 were mostly based on previous experiments and experiences; and 2) we usually remove terms appearing once in the dataset, but we decided to increase the threshold to keep the scale of terms at a relatively small level in Step 7.

Regarding to author names, we collected 5349 distinctive authors from a raw list of 18,882 authors, and the cleaning process includes: 1) a light author name disambiguation function integrated in the VantagePoint was applied to consolidate potential variations – e.g., “Eugene Garfield”, “Garfield, Eugene”, and “E Garfield”; and 2) authors who only published one paper in our dataset were removed. The co-authorship network was then constructed.

Thus, a bi-layer network was built up by connecting the co-term network and the co-authorship network with links, representing the authorships of terms and weighted by the frequency. a demonstration of the bi-layer network in VOSViewer (Waltman et al. 2010) is given in Fig. 3. Note that links between the co-term network and the co-authorship network are not given.

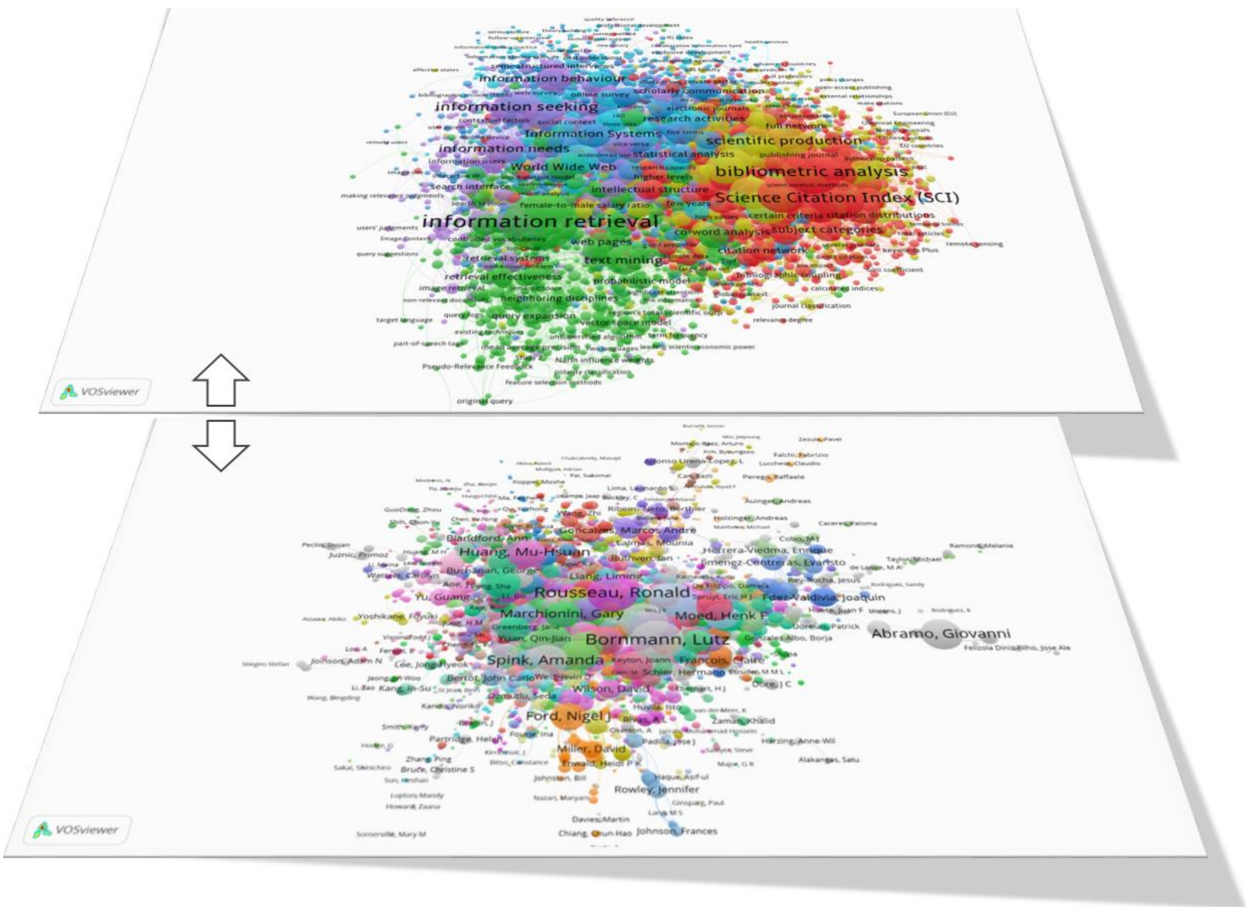

Figure 3. A bi-layer network for information science.

Network analytics were applied to quantify the three technological characteristics of the 25,359 terms, and the descriptive statistics of the results are given in Table 3. Based on the mean of the three characteristics, we selected 1000 terms and generated one 3D map in Fig. 4 (Left), visualizing and locating distinctive terms in a 3D solution.

Table 3. Descriptive statistics for technological characteristics.

No.	*Characteristics*	*Sub-characteristics*	*Max*	*Min*	*Mean*	*S.D.*
1	Fundamentality	Degree Centrality	1	0	0.014	0.023
		Closeness Centrality	1	0	0.690	0.068
		Between Centrality	1	0	0.001	0.008
		Average	1	0	0.235	0.028
2	Connectivity	N/A	7.47	0	0.050	0.220
3	Externality	N/A	20853	0	154.3	437.5

Note that regarding to fundamentality, we used the average of the three sub-characteristics as the value of fundamentality in further analytics.

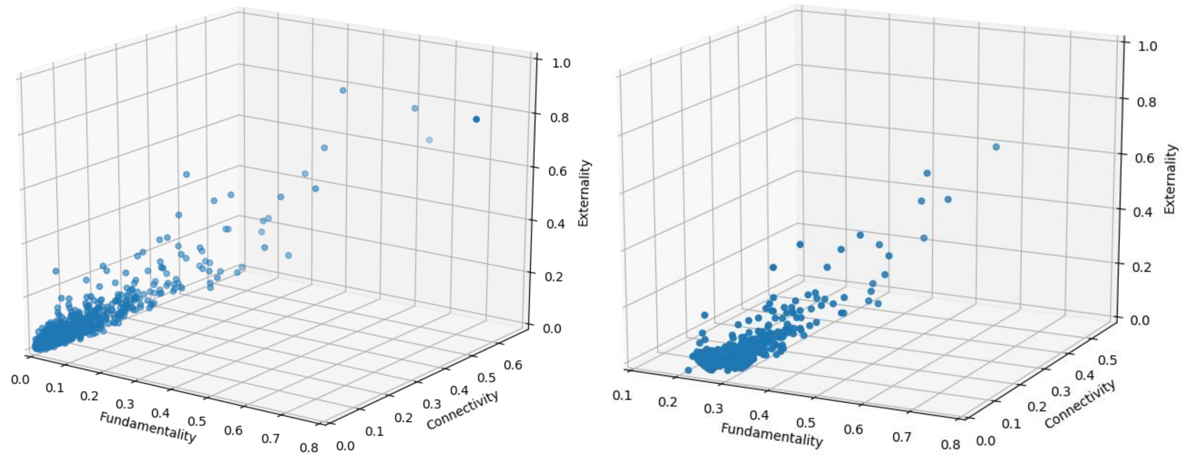

Figure 4. 3D map for 1000 terms with technological characteristics – Left for the current bi-layer network and Right for the predicted bi-layer network.

The common neighbor-based link prediction approach was then applied to calculate the CN values of all links, including missing links. With such values, the structure of the bi-layer network was changed and the technological characteristics of all nodes could be re-calculated. The descriptive statistics for technological characteristics in the predicted bi-layer network are given in Table 4, and a 3D map for visualizing selected 1000 terms with technological characteristics is given in Fig. 4 (Right).

Table 4. Descriptive statistics for technological characteristics in the predicted bi-layer network.

No.	*Characteristics*	*Sub-characteristics*	*Max*	*Min*	*Mean*	*S.D.*
1	Fundamentality	Degree Centrality	1	0	0.010	0.036
		Closeness Centrality	1	0	0.001	0.014
		Between Centrality	1	0	0.693	0.063
		Average	1	0	0.235	0.031
2	Connectivity	N/A	1	0	0.001	0.015
3	Externality	N/A	1	0	0.027	0.023

We exploited the receiver operating characteristic (ROC) analysis and the value of area under the curve (AUC) to validate the performance of the link prediction approach (Fawcett 2006). Briefly, in the ROC analysis the applied dataset was randomly divided into a training set and a test set, then, the ranking list generated by the link prediction approach in the training set would be compared with the true ranking list in the test set, and an AUC value can be calculated. The AUC values for links in the co-term network, in the co-authorship network, and between the two networks are given in Fig. 5.

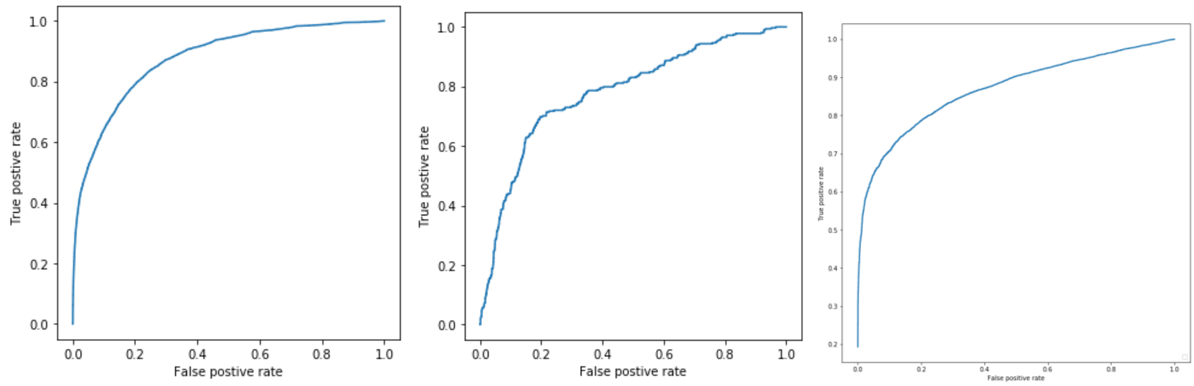

Figure 5. AUC values for validating the link prediction approach – Left for links in the co-term network, Middle for links in the co-authorship network, and Right for links between the both.

As shown in Fig. 5, AUC values for links in the co-term network, in the co-authorship network, and between the two networks are 0.89, 0.79, and 0.95 respectively, indicating an acceptable result of the link prediction approach.

We then compared the difference between the two ranking lists (i.e., the ones generated by the bi-layer network and the predicted bi-layer network respectively) and picked up a list of terms (given in Table 5) whose ranking is within the Top 50 in the one generated by the predicted bi-layer network but largely different from the previous one, indicating its potential of being emerging generic technologies in information science.

Table 5. Selected terms indicating the potential of being emerging generic technologies

No.	*Terms*	*Rank Change*	*No.*	*Terms*	*Rank Change*
1	Information retrieval	63 - 1	6	Text mining	383 - 24
2	Information seeking	15 - 4	7	Social network analysis	52 - 28
3	Digital libraries	23 - 9	8	Science policy	54 - 36
4	Information systems	45 - 17	9	Co-authorship network	79 - 43
5	H index	169 - 22			

Terms appearing in Table 5 are coherent with the study conducted by Zhang et al. (2018), where several key topics in bibliometrics were identified. Several insights are summarized below:

- As a fundamental toolkit, the involvement of *information retrieval* (e.g., *text mining*) and *information systems* techniques has significantly changed the information science discipline, but with the rapid development of information technologies, especially artificial intelligence, the involvement would be further enhanced and become an emergent direction in information science.
- *Information seeking* and *digital libraries* would be considered as two mainstream tasks of information science and library science, and the boom of social media would become a key to dramatically extend its current research areas and generate new topics.
- *Social network analysis* and *co-authorship network* are the applications of complex network analytics for analyzing science maps, which could be a cross-disciplinary direction and have attracted great attention in the past decades.

- *H index* is a traditional indicator for research evaluation in bibliometrics and could be considered as the application of complex network analytics as well. How to modify h index to evaluate researchers and research institutions from comprehensive aspects is still a hot topic in bibliometrics.
- *Science policy* could be a practical area of information science (e.g., bibliometrics). Even though such applications have appeared in the literature for decades, new problems in the area of science, technology, and innovation policy (STIP), and new solutions for existing STIP problems are still challenging researchers in information science.

Conclusions and Future Studies

This paper provides a bi-layer network analytics-based prediction method for characterizing the potential of being emerging generic technologies, in which 1) three technological characteristics are identified and then quantified by topological indicator, and 2) a common neighbor-based link prediction approach is applied for reconstructing networks with weighted missing links. Comparison between the ranking lists of terms indicating the potential of being emerging generic technologies, which are respectively generated by the current and the reconstructed networks, is used to identify potential emerging generic technologies. A case study on predicting emerging generic technologies in information science demonstrates the feasibility and reliability of the proposed method.

Future directions can be conducted to address limitations of this study from the following aspects: 1) a modified link prediction approach can be developed to better adapt to a bi-layer network, and comparisons with baselines can be applied as well; 2) it is more convincing to quantitatively or qualitatively validate the results based on different indicators and with diverse practical needs; and 3) examining the proposed method in cases with relatively broad disciplines would further help demonstrate its reliability.

Acknowledgments

This work is supported by the Australian Research Council under Discovery Early Career Researcher Award DE190100994.

References

Appio, F. P., Martini, A., & Fantoni, G. (2017). The light and shade of knowledge recombination: Insights from a general-purpose technology. *Technological Forecasting and Social Change, 125*, 154-165.

Borgatti, S. P., Mehra, A., Brass, D. J., & Labianca, G. (2009). Network analysis in the social sciences. *Science, 323*(5916), 892-895.

Börner, K., Klavans, R., Patek, M., Zoss, A. M., Biberstine, J. R., Light, R. P., . . . Boyack, K. W. (2012). Design and update of a classification system: The UCSD map of science. *PLoS One, 7*(7), e39464.

Bresnahan, T. F., & Trajtenberg, M. (1995). General purpose technologies ‘Engines of growth’? *Journal of econometrics, 65*(1), 83-108.

Chakraborty, T., Kumar, S., Goyal, P., Ganguly, N., & Mukherjee, A. (2015). On the categorization of scientific citation profiles in computer science. *Communications of the ACM, 58*(9), 82-90.

Coccia, M. (2017). The source and nature of general purpose technologies for supporting next K-waves: Global leadership and the case study of the US Navy's Mobile User Objective System. *Technological Forecasting and Social Change, 116*, 331-339.

Crafts, N. (2004). Steam as a general purpose technology: A growth accounting perspective. *The Economic Journal, 114*(495), 338-351.

Ding, W., & Chen, C. (2014). Dynamic topic detection and tracking: A comparison of HDP, C-word, and cocitation methods. *Journal of the Association for Information Science and Technology, 65*(10), 2084-2097.

Érdi, P., Makovi, K., Somogyvári, Z., Strandburg, K., Tobochnik, J., Volf, P., & Zalányi, L. (2013). Prediction of emerging technologies based on analysis of the US patent citation network. *Scientometrics, 95*(1), 225-242. doi: 10.1007/s11192-012-0796-4

Fawcett, T. (2006). An introduction to ROC analysis. *Pattern recognition letters, 27*(8), 861-874.

Freeman, L. C. (1977). A set of measures of centrality based on betweenness. *Sociometry*, 35-41.

Freeman, L. C. (1978). Centrality in social networks conceptual clarification. *Social networks, 1*(3), 215-239.

Glänzel, W., & Thijs, B. (2012). Using "core documents" for detecting and labelling new emerging topics. *Scientometrics, 91*(2), 399-416. doi: 10.1007/s11192-011-0591-7

Guo, Y., Huang, L., & Porter, A. L. (2010). The research profiling method applied to nano-enhanced, thin-film solar cells. *R&D Management, 40*(2), 195-208.

Hou, J., Yang, X., & Chen, C. (2018). Emerging trends and new developments in information science: a document co-citation analysis (2009–2016). *Scientometrics, 115*(2), 869-892.

Klavans, R., & Boyack, K. W. (2017). Which type of citation analysis generates the most accurate taxonomy of scientific and technical knowledge? *Journal of the Association for Information Science and Technology, 68*(4), 984-998.

Maine, E., & Garnsey, E. (2006). Commercializing generic technology: The case of advanced materials ventures. *Research Policy, 35*(3), 375-393.

Maine, E., & Thomas, V. (2017). Raising financing through strategic timing. *Nature nanotechnology, 12*(2), 93.

Martin, B. R. (1995). Foresight in science and technology. *Technology Analysis & Strategic Management, 7*(2), 139-168.

Newman, M. E. (2001). Clustering and preferential attachment in growing networks. *Physical Review E, 64*(2), 025102.

Park, I., & Yoon, B. (2018). Technological opportunity discovery for technological convergence based on the prediction of technology knowledge flow in a citation network. *Journal of Informetrics, 12*(4), 1199-1222.

Qiu, R., & Cantwell, J. (2018). General purpose technologies and local knowledge accumulation—A study on MNC subunits and local innovation centers. *International Business Review, 27*(4), 826-837.

Rotolo, D., Hicks, D., & Martin, B. R. (2015). What is an emerging technology? *Research Policy, 44*(10), 1827-1843.

Sinfield, J., & Solis, F. (2016). Finding a lower-risk path to high-impact innovations. *MIT Sloan management review, 57*(4), 79.

Small, H., Boyack, K. W., & Klavans, R. (2014). Identifying emerging topics in science and technology. *Research Policy, 43*(8), 1450-1467.

Suominen, A., & Toivanen, H. (2016). Map of science with topic modeling: Comparison of unsupervised learning and human-assigned subject classification. *Journal of the Association for Information Science and Technology, 67*(19), 2464–2476.

Takeda, Y., & Kajikawa, Y. (2009). Optics: A bibliometric approach to detect emerging research domains and intellectual bases. *Scientometrics, 78*(3), 543-558.

Waltman, L., van Eck, N. J., & Noyons, E. C. (2010). A unified approach to mapping and clustering of bibliometric networks. *Journal of Informetrics, 4*(4), 629-635.

Waltman, L., & Van Eck, N. J. (2013). A smart local moving algorithm for large-scale modularity-based community detection. *The European Physical Journal B, 86*(11), 471.

Yan, E., Ding, Y., & Zhu, Q. (2009). Mapping library and information science in China: A coauthorship network analysis. *Scientometrics, 83*(1), 115-131.

Yan, E. (2015). Research dynamics, impact, and dissemination: A topic-level analysis. *Journal of the Association for Information Science and Technology, 66*(11), 2357-2372.

Yang, C., Park, H., & Heo, J. (2010). A network analysis of interdisciplinary research relationships: The Korean government's R&D grant program. *Scientometrics, 83*(1), 77-92.

Zhang, Y., Porter, A. L., Hu, Z., Guo, Y., & Newman, N. C. (2014a). "Term clumping" for technical intelligence: A case study on dye-sensitized solar cells. *Technological Forecasting and Social Change, 85*, 26-39.

Zhang, Y., Zhou, X., Porter, A. L., Gomila, J. M. V., & Yan, A. (2014b). Triple Helix innovation in China's dye-sensitized solar cell industry: Hybrid methods with semantic TRIZ and technology roadmapping. *Scientometrics*, *99*(1), 55-75.

Zhang, Y., Zhang, G., Zhu, D., & Lu, J. (2017). Science evolutionary pathways: Identifying and visualizing relationships for scientific topics. *Journal of the Association for Information Science and Technology*, *68*(8), 1925-1939.

Zhang, Y., Lu, J., Liu, F., Liu, Q., Porter, A., Chen, H., & Zhang, G. (2018). Does deep learning help topic extraction? A kernel k-means clustering method with word embedding. *Journal of Informetrics*, *12*(4), 1099-1117.

Zhou, X., Zhang, Y., Porter, A. L., Guo, Y., & Zhu, D. (2014). A patent analysis method to trace technology evolutionary pathways. *Scientometrics*, *100*(3), 705-721.