Scientometrics (2020) 122:607–624
https://doi.org/10.1007/s11192-019-03288-5

Review on emerging research topics with key-route main path analysis

Shuo Xu[1] · Liyuan Hao[1] · Xin An[2] · Hongshen Pang[3] · Ting Li[4]

Received: 24 June 2019 / Published online: 9 November 2019

Abstract

The fast development of the emerging research topics field results in hundreds of theoretical and empirical publications. However, to our knowledge, there is no comprehensive and objective literature review on this field until now. To this end, a citation network consisting of 1607 papers between 1965 and early 2019 is explored to discover the knowledge diffusion trajectory of the emerging research topics field by the key-route main path analysis approach, armed with the traversal weight of search path link count. From the convergence–divergence patterns in the local and global main paths, the development of emerging research topics field can be divided into three different stages: the emergence, exploration and development stages. In the meanwhile, several research drifts can also be observed: (1) from citation-based approaches to machine learning based ones, (2) from the measurement to the identification, and (3) from the papers to the patents. Finally, the directions of future research are suggested.

Keywords Emerging research topics · Literature review · Key-route main path analysis · Knowledge diffusion trajectory

✉ Xin An
anxin@bjfu.edu.cn

Shuo Xu
xushuo@bjut.edu.cn

Liyuan Hao
haoliyuan@emails.bjut.edu.cn

Hongshen Pang
phs@szu.edu.cn

Ting Li
li_ting08@gibh.ac.cn

1 Research Base of Beijing Modern Manufacturing Development, College of Economics and Management, Beijing University of Technology, Beijing 100124, People's Republic of China

2 School of Economics and Management, Beijing Forestry University, Beijing 100083, People's Republic of China

3 Library, Shenzhen University, Shenzhen 518060, People's Republic of China

4 Guangzhou Institutes of Biomedicine and Health, Chinese Academy of Sciences, Guangzhou 510530, People's Republic of China

Introduction

There is considerable and growing interest in emerging research topic (Mogoutov and Kahane 2007; Upham and Small 2010; Halaweh 2013; Zhao and Strotmann 2014; Kyebambe et al. 2017; Soriano et al. 2018; Calabrese et al. 2018; Kuhlmann et al. 2019, and among others). While elaborating this concept in the literature, many different phrases have been used, such as *emerging research topics* (Wang 2018; Xu et al. 2019), *emerging topics* (Ohniwa et al. 2010; Glänzel and Thijs 2012; Tu and Seng 2012; Small et al. 2014), *research fronts* (Persson 1994; Åström 2007; Chen et al. 2007; Aris et al. 2009; Toivanen 2014; Ma and Liu 2016), *emerging trends* (Glassey 2009; Xie 2015; Liu and Gui 2016; Sangam 2000; Naaman et al. 2011; Liu et al. 2018), *emerging technologies* (Smalheiser 2001; Adner and Levinthal 2002; Roche et al. 2010), *emerging research fields* (Scalise et al. 2007; Lee 2008; Bettencourt et al. 2008; Chang et al. 2010; Jang et al. 2011; Jarić et al. 2014; Rohrbeck et al. 2015; Weismayer and Pezenka 2017) and so on. In our opinion, main reason is that there is no consensus one a universal concept and accurate properties attached to it. Several examples related to the definition include: (1) the clusters of highly interactive literatures in science (Persson 1994; Morris et al. 2003; Åström 2007; Aris et al. 2009; Ma and Liu 2016), (2) the themes that appear frequently during a specified time period (Chen et al. 2007; Tu and Seng 2012), and (3) the topics that have properties of radical novelty and prominent impact (Rotolo et al. 2015; Wang 2018), fast relatively growth (Ohniwa et al. 2010), coherence (Rotolo et al. 2015; Wang 2018), persistence (Porter et al. 2018; Wang et al. 2018) and so on. On closer examination, we argue that the meanings of these terms are very close, so the emerging research topic is chosen to collectively refer to above mentioned concepts.

Last five decades witnessed significant progress in the field of emerging research topics ever since de Solla Price (1965). Several recent research activities, such as Emerging Research Areas and their Coverage (Reiss et al. 2013) supported by the European Research Council (ERC), Foresight and Understanding from Scientific Exposition (FUSE) (Small et al. 2014) funded by the Intelligence Advanced Research Projects Activity (IARPA) and 2018–2019 Contest of Measuring Tech Emergence,[1] promote further the development of emerging research topics. These studies can benefit the research foundations and policymakers, whose main purpose is to promote and enhance the development of potentially promising research topics.

Until now, more than 1500 closely-related scholarly articles are published in the Web of Science database (see "Dataset" section for more details). To the best of our knowledge, there is no comprehensive literature review on emerging research topics until now. This makes it very difficult for beginning researchers to gain a proper understanding of the constantly updating field in the limited time. Moreover, it is very possible to miss important development in the areas beyond a researcher's specialty. Therefore, this work devotes to systematically reviewing the main knowledge flows that reflect crucial development paths in this field and pointing out the directions of future research. It is well known that most existing literature surveys are conducted from the perspective of the resulting researchers' experience. Thus, these reviews may be constrained by the time and the cognitive ability of the corresponding scholars on the interested field (Raghuram et al. 2010). Hence, it is very possible that some important articles in the field are excluded, which may result in several

[1] https://vpinstitute.org/academic-portal/tech-emergence-contest/.

missing research lines. As an objective methodology, main path analysis (MPA) (Hummon and Doreain 1989) is a very promising alternative to review thousands of articles in this field. Furthermore, it has been successfully applied to survey the development of *data envelopment analysis* (Liu et al. 2013a, b), *Hirsch index* (Liu and Lu 2012), *peer review* (Batagelj et al. 2017), *graphene* (Yeo et al. 2014), *IT outsourcing* (Liang et al. 2016), and *fuel cell* (Verspagen 2007) fields.

It is well-documented that the citation-link network bears rich valuable information about how the knowledge diffuses (Bhupatiraju et al. 2012; Zhu et al. 2016). Therefore, it should be rational and feasible for MPA to scrutinize the development of the interested domain from the large citation-link network with efficient algorithms (Batagelj 2003) in terms of single main path (Hummon and Doreain 1989), multiple main paths (Liu and Lu 2012), key-route main path (Liu and Lu 2012) and so on. Since the most significant links (i.e., the links with the highest traversal weight) are not guaranteed to be included in the single main path and multiple main paths (Liu and Lu 2012), the key-route main path is preferred in this study. For purpose of calculating traversal weights, four methods are put forward in the literature (Hummon and Doreain 1989; Batagelj 2003): (a) search path link count (SPLC), (b) search path nodes pair (SPNP), (c) node pair projection count (NPPC), and (d) search path count (SPC), among which NPPC is not suitable for large networks due to high computation complexity (Batagelj 2003). Liu et al. (2019) recommended SPLC for tracing knowledge diffusion trajectory, since it imitates most closely the knowledge diffusion scenario in the scientific development where the intermediate nodes (i.e., individual publications) not only pass knowledge, but also are knowledge sources per se. Hence, SPLC is also adopted here. In summary, the key-route main path methodology, armed with the traversal weight of SPLC, is utilized in this work to review the literature in the emerging research topics field.

The rest of the article is organized as follows. After "Methodology" section introduces briefly the traversal weight calculation and key-route main path analysis, "Dataset" section shows how to define the emerging research topics fields as precisely as possible. Then, a knowledge diffusion trajectory with two convergence-divergence patterns is discovered in "Analysis and Results" section, from which one can identify a series of papers that play important roles in the emerging research topics field. The last section concludes this study with the directions of future research.

Methodology

In this study, the key-route main path analysis is utilized to explore the knowledge diffusion trajectories, and then to further understand the development track of emerging research topics area. From the perspective of evolution, the main path of a citation-link network can be seen as a time sequence diagram that reflects the knowledge diffusion of an interested area.

Traversal weights

In a citation-link network, a citation link's SPC value is the total times that a citation link is traveled if one runs through all the possible citation paths from all the sources (i.e., the nodes that are cited while referring to no other nodes) to all the sinks (i.e., the nodes that are referred to while citing to no nodes). A citation link's SPLC value is the total counts

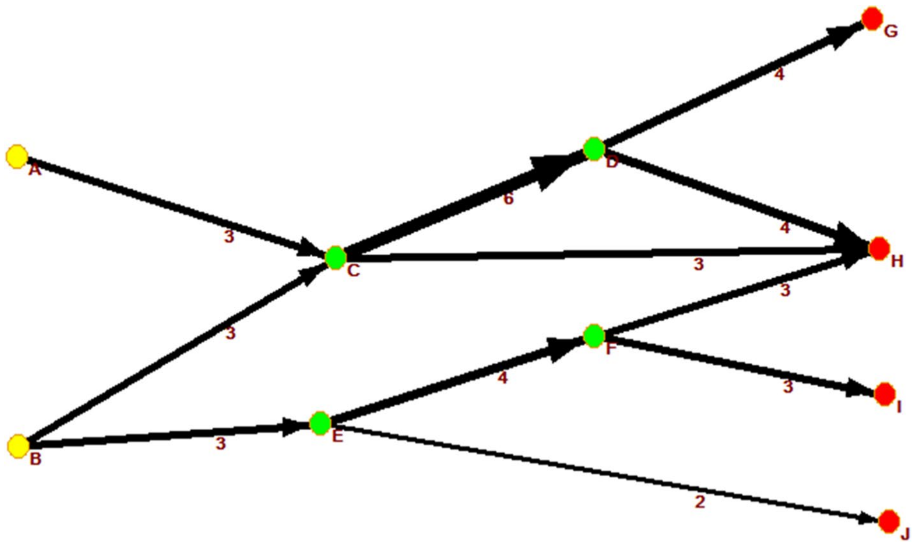

Fig. 1 A citation-link network with SPLC traversal weights. The arrow indicates that the tail node is cited by the head one, and the thickness of each citation link is drawn in proportion to its SPLC values

that a citation link is traveled if one runs through all the possible citation paths from all the ancestors (including themselves) of the tail nodes to all the sinks. The SPNP weight is more sophisticated, and it is the total number that a citation link is traveled if one runs through all the possible citation paths from all the ancestors (including themselves) of the tail nodes to all the descendants (including themselves) of the head nodes. The NPPC method calculates the number of times each link involved in connecting all node pairs in all subgraphs derived from the network.

Among these traversal weights, only SPC follows Kirchhoff's node law, so a preference on SPC is suggested by Batagelj (2003). However, it is well known that in the diffusion process of the scientific knowledge, intermediate nodes not only pass knowledge, but also generate knowledge (Liu et al. 2019). In view of this, SPLC is closer in line with the actual knowledge diffusion situation, therefore this work decides to apply SPLC method for computing citation links' weights.

Key-route main path analysis

The key-route search procedure (Liu and Lu 2012) begins with the citation-links with the top traversal weight (i.e., key-route), and then searches forward from the head node of the key-route until a sink is hit and searches backward from the tail node of the key-route until a source is hit. The resulting path-fragments are merged together to form the key-route main path. The more key-routes one selects, the more details of the main path will be revealed. Thus, prior knowledge can be readily incorporated into main path analysis, so that one can guarantee that the most significant links are included in the main path (Liu and Lu 2012). Hence, the key-route main path analysis method is preferred to in this study.

For convenient understanding, let's take Fig. 1 as an example. It is easy to see that *A* and *B* are *source* nodes, and *G*, *H*, *I* and *J* are *sink* nodes. In all the citation links, <*C*, *D*>

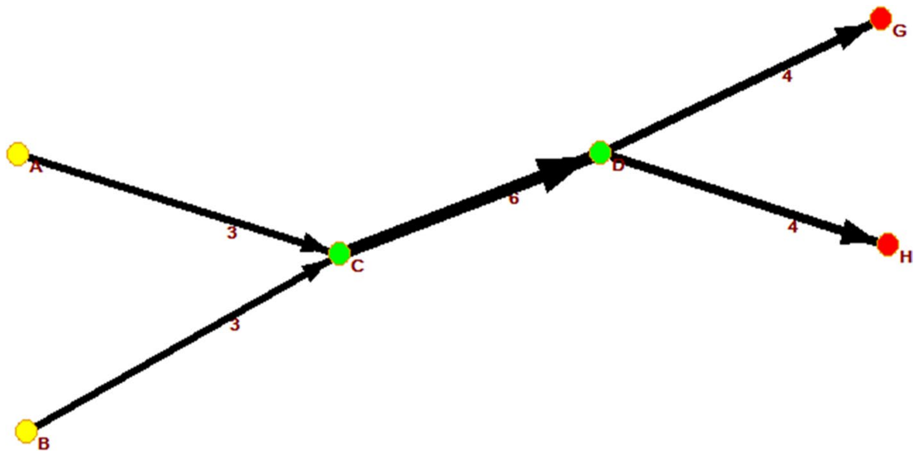

Fig. 2 The key-route main path from the citation-link network in Fig. 1. The arrow indicates that the tail node is cited by the head one, and the thickness of each citation link is drawn in proportion to its SPLC values

has the largest SPLC value. This means that this citation link plays a significant role in the knowledge diffusion process, so the link $<C, D>$ is chosen as the key-route. When searching forward and backward from the link $<C, D>$ from global or local viewpoint, the same key-route main path can be obtained in this case, as shown in Fig. 2. It is worth noting that the global main paths are usually not identical to the local counterparts in the real-world applications.

Dataset

The scientific publications in the emerging research topics field are retrieved from Web of Science (WoS) core collection on March 14, 2019 from the library of Beijing University of Technology. The data time span is set to range from 1965 to 2019. Given that the research front was first defined by de Solla Price in 1965, de Solla Price (1965) is often regarded as a pioneering work in this domain. So, the starting year of the dataset is fixed to 1965.

The following search strategy from Lu et al. (2019) is utilized in this study: TS = (“emerg* topic*” OR “emerg* research* topic*” OR “emerg* scien* topic*” OR “emerging* academic* topic*” OR “emergence of topic*” OR “emergence of scien* topic*”OR “emergence of research* topic*” OR “emergence of academic* topic*” OR “topic* emergence” OR “research* topic* emergence” OR “scien* topic* emergence” OR “academic* topic* emergence” OR “research* front*” OR “scien* front*” OR “academic* front*” OR “emerg* field*” OR “emerg* scien* field*” OR”emerg* research field*” OR “emerg* academic* field*” OR “emergence of field*”OR “emergence of scien* field*”OR “emergence of research field*” OR “emergence of academic* field*” OR “field* emergence” OR “research* field* emergence” OR “scien* field* emergence” OR “academic* field* emergence” OR “emerg* trend*” OR “emerg* research* trend*” OR “emerg* scien* trend*” OR “emerg* academic* trend*” OR “emergence of trend*” OR “emergence of research* trend*”OR “emergence of scien* trend*” OR “emergence of academic* trend*” OR “trend* emergence” OR “research* trend* emergence” OR “scien* trend* emergence”

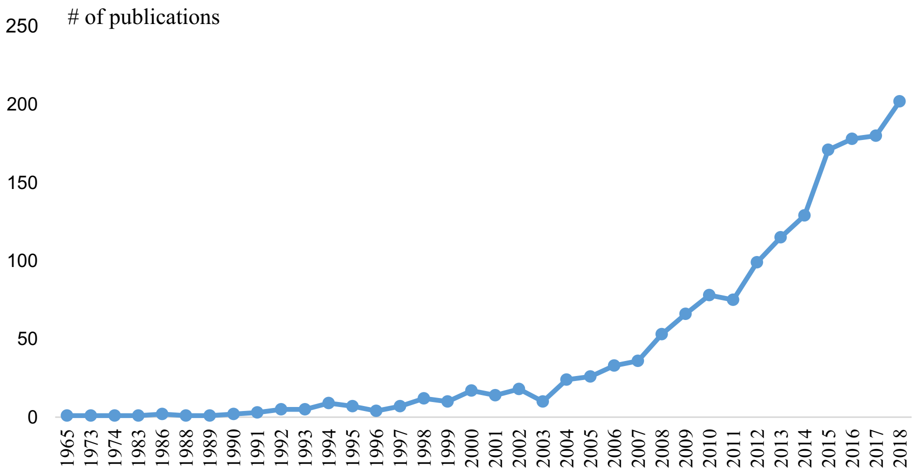

Fig. 3 Distribution of number of publications over year for *emerging research topics* dataset

OR "academic* trend* emergence"). The document type is limited to *article*, *review*, and *proceedings paper*, and the research area includes *science technology other topics*, *information science library science*, and *social sciences other topics*. In total, the number of publications is 1567.

To make sure that all papers are pertaining to the emerging research topics, scholarly articles in the dataset is checked carefully one by one by reading the resulting titles and abstracts. In the meanwhile, several important papers from the reference list in Lu et al. (2019) but missed from the WoS core collection are added manually to the dataset. Finally, the dataset includes 1607 scientific publications. We assume that these articles will enable researchers to have a full picture on the development of emerging research topics field. The distribution of number of publications over year is illustrated in Fig. 3. From Fig. 3, one can see that the number of articles increases at a very slow rate before 2003, but after 2003, the number of publications increases rapidly, especially after 2011. This demonstrates that this field has been receiving increasing attention. In addition, a downward trend can be observed after 2018, since the latest publications are unavailable in the WoS core collection when the search is executed.

Analysis and results

According to our understanding of emerging research topics field, the following publications have contributed greatly to the development of the field and promoted largely the follow-up studies: de Solla Price (1965), Small (1973), Chen (2006), Rotolo et al. (2015) and so on. Intuitively, several convergence-divergence patterns should appear in the discovered main path with the above articles as the convergent nodes. Hence, the number of key-routes with top SPLC counts is chosen from 10 to 40 with a step size 5 (i.e., 10, 15, 20, …, 40), and then the exploratory main path analysis is conducted with Pajek software (Batagelj and Mrvar 1998). The expected convergence-divergence patterns emerge when the number of key-routes is 35, as shown Figs. 4 (local main path) and 5 (global main path),

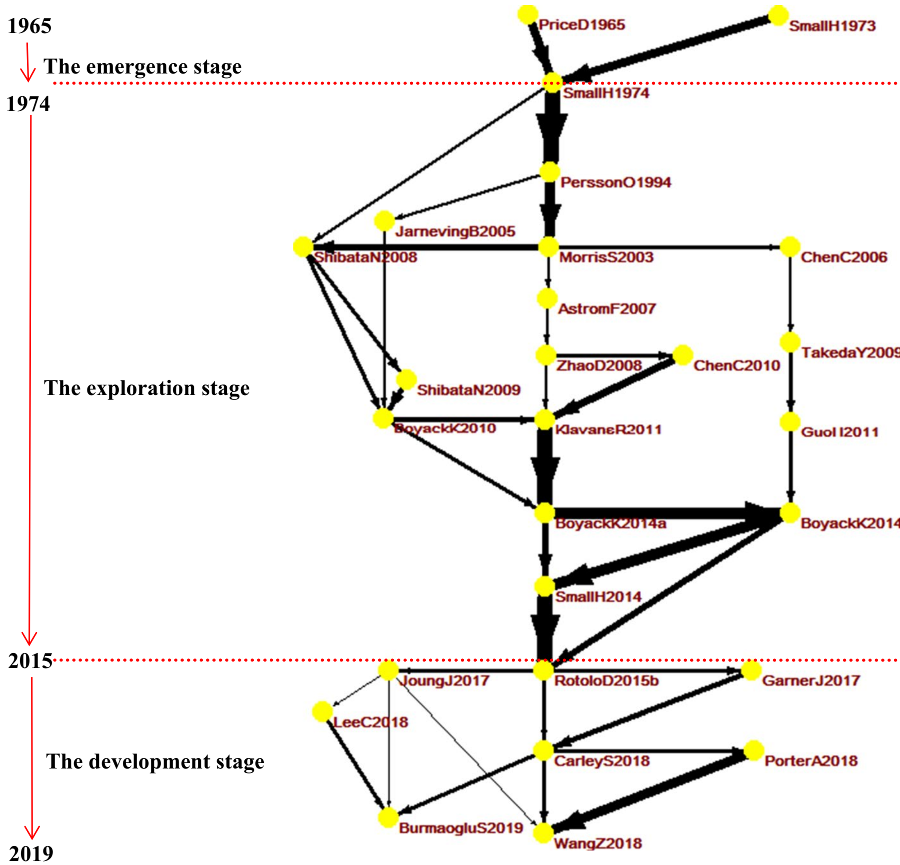

Fig. 4 The local main paths with top 35 key-routes for the emerging research topics field. The citation links thicken by their SPLC values

but becomes blurred if the number of key-routes is further increased or decreased. Hereinafter, our analysis will be limited to the case with the number of key-routes 35.

Before delving into more specifies, all scientific publications appearing in the key-route main paths are listed in Table 1, in which the first column (ID) corresponds to the resulting node's label in Figs. 4 and 5. The nodes are labeled with the last name of the first author, the initial capitals of the corresponding first name, and then followed by the publication year. In the meanwhile, for the convenience of referring to specific scientific publications, scholarly articles are also given in the second column in Table 1. The last column in Table 1 highlights explicitly which main path the resulting nodes come from.

Three stages of emerging research topics field development

From Figs. 4 and 5, the development process of emerging research topics field can be divided roughly into three stages. The *emergence stage* begins from de Solla Price (1965) (ID: PriceD1965) and Small (1973) (ID: SmallH1973), which lay the theoretical and methodological foundation of the field, respectively. The *exploration stage* is rather long, from

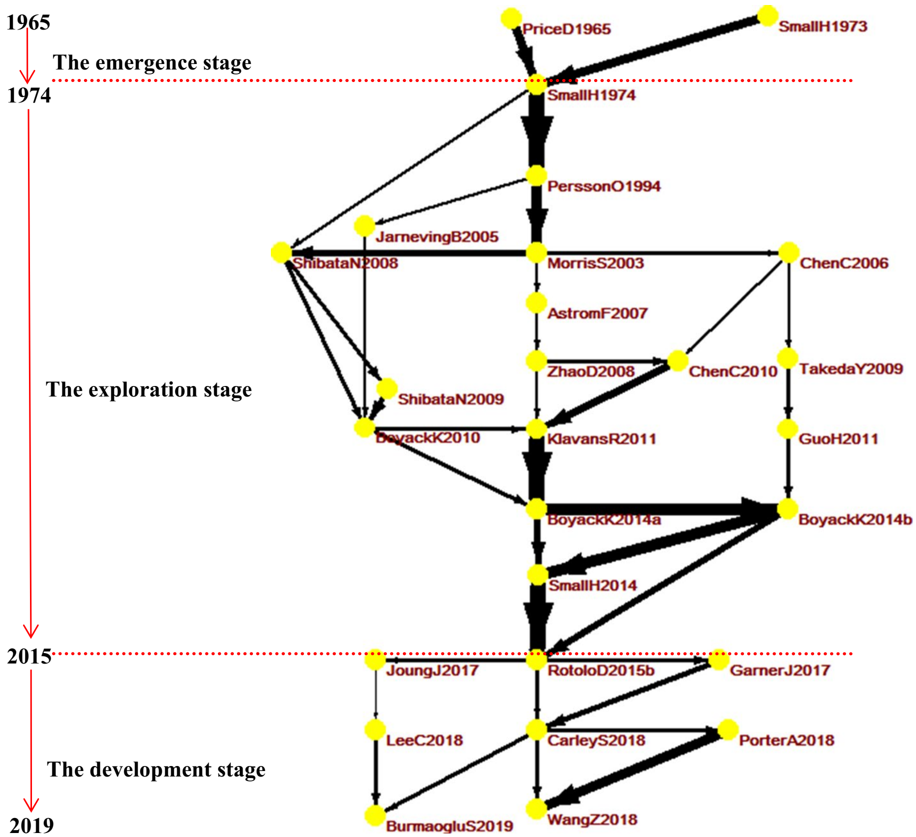

Fig. 5 The global main paths with top 35 key-routes for the emerging research topics field. The citation links thicken by their SPLC values

Table 1 A list of the scientific publications appearing in the key-route global or local main paths

ID	Scientific publication	ID	Scientific publication
PriceD1965	de Solla Price (1965)	KlavansR2011	Klavans and Boyack (2011)
SmallH1973	Small (1973)	GuoH2011	Guo et al. (2011)
SmallH1974	Small and Griffith (1974)	BoyackK2014a	Boyack and Klavans (2014)
PerssonO1994	Persson (1994)	BoyackK2014b	Boyack et al. (2014)
MorrisS2003	Morris et al. (2003)	SmallH2014	Small et al. (2014)
JarnevingB2005	Jarneving (2005)	RotoloD2015b	Rotolo et al. (2015)
ChenC2006	Chen (2006)	GarnerJ2017	Garner et al. (2017)
AstromF2007	Åström (2007)	JoungJ2017	Joung and Kim (2017)
ShibataN2008	Shibata (2008)	LeeC2018	Lee et al. (2018)
ZhaoD2008	Zhao and Strotmann (2008)	PorterA2018	Porter et al. (2018)
ShibataN2009	Shibata et al. (2009)	CarleyS2018	Carley et al. (2018)
TakedaY2009	Takeda and Kajikawa (2009)	WangZ2018	Wang et al. (2018)
ChenC2010	Chen et al. (2010)	BurmaogluS2019	Burmaoğlu et al. (2019)
BoyackK2010	Boyack and Klavans (2010)		

Small and Griffith (1974) (ID: SmallH1974) to Rotolo et al. (2015) (ID: RotoloD2015), and it constantly separates, extends, occasionally weakens or disappears. Rotolo et al. (2015) (ID: RotoloD2015), which gives an elaborate overview of the emerging technology, bridges the *exploration* and *development* stages. From 2015 till now, two research lines can be observed from Figs. 4 and 5, where some citation-links gradually cross and merge.

It is noted that many triangle structures can be observed from Figs. 4 and 5, such as Shibata et al. (2008) (ID: ShibataN2008), Shibata et al. (2009) (ID: ShibataN2009) and Boyack and Klavans (2010) (ID: BoyackK2010). This phenomenon can be interpreted by a citation-based index of *disruptiveness* (Funk and Owen-Smith 2017, Wu et al. 2019). The intuition behind the knowledge paths containing the triangle structures is straightforward: if the papers cite not only a given article, but also a substantial portion of that article's references, then this article can be seen as consolidating its scientific fields. If the opposite is true, that is to say, if the future citations to the article do not acknowledge the article's own intellectual forebears, then the article can be seen as disrupting its field (Azoulay 2019; Bornmann and Tekles 2019). In this context, the articles (e.g. Shibata et al. 2009; Chen et al. 2010; Klavans and Boyack 2011; Small et al. 2014) tend to develop existing studies in emerging research topics field, whereas some other articles (e.g. Åström 2007; Takeda and Kajikawa 2009; Chen 2006) tend to disrupt this scientific field with new insights, ideas and methods. All in all, as Azoulay (2019) points out, sustained development of the given scientific fields requires both radical and incremental contributions. Consequently, both the consolidated and disruptive works are essential to emerging research topics fields.

To corroborate each other, we search forward and backward from the top 35 key-routes locally and globally, as seen in Figs. 4 and 5. Many scientific publications are shared by the global and local main paths (cf. Table 1). This is no doubt to admit that these works have great impact on emerging research topics development, and it is feasible for the key-route main path analysis to discover the knowledge diffusion trajectory.

The emergence stage: from 1965 to 1974

The groundbreaking paper (de Solla Price 1965) firstly defined the concept of research front and vividly expressed an active research front as a sort of growing tip or epidermal layer. It is easy to see that this concept emphasizes the *novelty* attribute. Several years later, the co-citation analysis method was proposed by Small (1973) to identify emerging research topics for future research. These two publications act as the root nodes of key-route main paths, and then converge to Small and Griffith (1974). On the basis of previous studies, Small and Griffith (1974) defined the research fronts as the clusters of highly interactive literature in science, and co-citation analysis approach was used to identify research fronts. The top articles in the exploration stage extend, modify, and explore further the ideas of these studies.

The exploration stage: from 1974 to 2015

This stage mainly deals with the research of emerging research topics from the perspective of the citation network analysis, such as co-citation, bibliographic coupling, direct citation and their variants. Furthermore, from Figs. 4 and 5, three streams can be roughly found: (1) the left stream emphasizes the performance comparison among several similar methods, (2) the middle stream mainly discusses citation-based methodological improvement, and (3) the right stream embodies citation-based and lexical-based methodological application.

These three streams are described in more detail in the following paragraphs. It is worth mentioning that the Morris et al. (2003) redefined the research fronts as the clusters of documents that tend to cite a fixed and time invariant set of base documents (i.e., intellectual base).

The left stream consists of 4 articles: Jarneving (2007), Shibata et al. (2008, 2009), and Boyack and Klavans (2010). These publications are arguing which citation-link analysis approach should be preferred to detect the emerging research topics. Jarneving (2005) compared the results of research front portrayed by two different methods based on co-citation analysis and bibliographic coupling, and later on, Jarneving (2007) combined bibliographic coupling with a cluster method and found that this combined method could effectively identify coherent research themes. Shibata et al. (2008) performed a comparative study in two research domains [gallium nitride (GaN) and complex networks (CN)]. Shibata et al. (2009) and Boyack and Klavans (2010) both further investigated the performance of the various citation approaches through extensive experiments. Shibata et al. (2009) concluded that direct citation approach performs the best and the co-citation method shows the worst performance. However, Boyack and Klavans (2010) had a different view that bibliographic coupling approach slightly outperforms the co-citation method and the direct citation has the worst performance.

The papers in the middle stream discuss the citation-based methodological improvement and innovation regarding emerging research topics on the basis of the previous studies. Åström (2007) redefined the research fronts as current and influential co-cited articles on the basis of the Morris et al. (2003) and studied changes in research fronts of library and information science (LIS) domain using document co-citation analysis (DCA). The author bibliographic-coupling analysis (ABCA) was introduced by Zhao and Strotmann (2008) to map the research activities of active authors themselves for a more realistic picture of the current state of research in the information science (IS) field, and founded that research fronts can be detected based on weak signals in author bibliographic coupling analysis. Then, Chen et al. (2010) performed a comprehensive, multiple-perspective method for both author co-citation analysis (ACA) and document co-citation analysis (DCA) studies in the IS field. In addition, Klavans and Boyack (2011) recreated the document co-citation map of the IS field published by Chen et al. (2010) to compare the accuracy of local and global maps of science generated by intellectual base and research front. In the end, Boyack and Klavans (2014) absorbed and improved their previous work (Boyack and Klavans 2010; Klavans and Boyack 2011).

The right stream slightly differentiates local and global main paths, as presented in Figs. 4 and 5. More specifically, global main path includes the citation-link <ChenC2006, ChenC2010>. The topological clustering method was performed by Takeda and Kajikawa (2009) to detect emerging research domains. Chen (2006) considered that the highly bursting words identified by the burst algorithm (Kleinberg 2003) can be viewed as the indicators of emerging research front, and improved their influential visualization tool—CiteSpace and then was used in Chen et al. (2010). This tool makes substantial theoretical and methodological contributions to progressive emerging research topics domain visualization (Kim and Chen 2015; Li 2017). A mixed model, put forward by Guo et al. (2011), combines three different indicators: sudden increases in the frequency of specific words, the number and speed by which new authors are attracted to an interested area. The empirical studies on four emerging research areas show that these indicators are very indicative.

Later on, the right stream merges to the node Boyack et al. (2014), which challenges the fixed notion that the topic emerges once a certain number of articles have published on that topic. A methodology is proposed by Boyack et al. (2014) to characterize known emerging research topics and found two different patterns of emergence: one where the topic is not focused but then grows explosively, and one where the topic quickly becomes an area of focus and then grows steadily. In fact, as Cozzens et al. (2010) pointed out, most previous studies concentrate on how to measure emerging research topics, rather than how to identify them. Since Cozzens et al. (2010), several researchers begin to explore approaches of identification. For instance, Small et al. (2014) combined direct citation and co-citation analysis to recognize emerging topics in terms of the novelty (or newness) and growth indicators.

In addition, an interesting phenomenon can be observed that the citation-based methods dominate the field, and only several lexical-based studies (such as Chen 2006 and Guo et al. 2011) appear in the global or local key-route paths. Since the movement of open access (OA) to all research literature and the construction of comprehensive bibliographic databases, more and more text resources can be easily accessed. Furthermore, due to many known problems (such as time-lagging problem) with the citation-based methods, intuitively the lexical-based approaches should be able to serve as alternatives to the citation-based ones. To check if the dominance of the citation-based methods alienates the lexical-based studies, *branch path* procedure (Ho et al. 2014) is used here to trace the development trajectory of the lexical-based approaches. After the top 3 high-cited papers (cf. Table 2) on lexical-based studies are firstly selected from our dataset, the resulting branch paths are shown in Fig. 6 on the basis of the global main path. The branch paths for the local main path are very similar to those for the global main path, so they are not illustrated here. Table 2 describes the detailed information of these papers appearing in the branch paths. From Fig. 6, it is not difficult to find the citation links of branch paths are very thin. This indicates that our intuitive assumption does not hold. As a matter of fact, we also tried the top 4, 5, 6 and 7 high-cited papers on lexical-based studies. Though the resulting branch paths for these cases are not exactly the same, the same conclusion can be drawn. A reasonable explanation is given by Small et al. (2014) that the lexical-based methods can list emerging topics within some specific field, but they are not suitable for identifying emerging research topics across a wide swath of science.

The development stage: from 2015 to now

A step forward in 2015 triggered a new development in the emerging research topics field. Rotolo et al. (2015) reviewed the previous works and gave a clear and transparent concept of the emerging technologies that can be characterized by five attributes: (1) radical novelty, (2) relatively fast growth, (3) coherence, (4) prominent impact, and (5) uncertainty and ambiguity. This should be the first time to develop a comprehensive definition of an emerging technology, so its viewpoint is popular and widespread till now. After the node, this stage could be classified into two subgroups.

One subgroup includes three nodes (ID: JoungJ2017, LeeC2018 and BurmaogluS2019). They follow the viewpoint of Rotolo et al. (2015) and explore new approaches of identification and novel definition. Joung and Kim (2017) proposed a keyword-based patent analysis model to monitor emerging technologies, where the relatedness between the keywords was quantified by TF-IDF function. Lee et al. (2018) presented a machine learning approach by combining multiple patent indicators to identify emerging technologies at early stage of patents issued. Burmaoğlu et al. (2019) argued that the concept of emergence in technology was still ambiguous, so they traced emergence discussions to find the evolution of related concepts, and explored further usage in the technological context.

Table 2 A list of the scientific publications appearing in the branch paths

ID	Authors	Title	Journal
MorrisS2004	Morris, SA; Yen, GG	Crossmaps: Visualization of overlapping relationships in collections of journal papers	PNAS
ChenC2006	Chen, CM	CiteSpace II: Detecting and visualizing emerging trends and transient patterns in scientific literature	JASIST
LeeW2008	**Lee, WH**	**How to identify emerging research fields using scientometrics: An example in the field of Information Security**	**Scientometrics**
TakedaY2009	Takeda, Y; Kajikawa, Y	Optics: a bibliometric approach to detect emerging research domains and intellectual bases	Scientometrics
GuoH2011	**Guo, HN; Weingart, S; Borner, K**	**Mixed-indicators model for identifying emerging research areas**	**Scientometrics**
ZittM2011	Zitt, M; Lelu, A; Bassecoulard, E	Hybrid Citation-Word Representations in Science Mapping: Portolan Charts of Research Fields?	JASIST
AroraS2013	**Arora, SK; Porter, AL; Youtie, J; Shapira, P**	**Capturing new developments in an emerging technology: an updated search strategy for identifying nanotechnology research outputs**	**Scientometrics**

The top 3 high-cited papers on lexical-based studies are highlighted in bold fonts

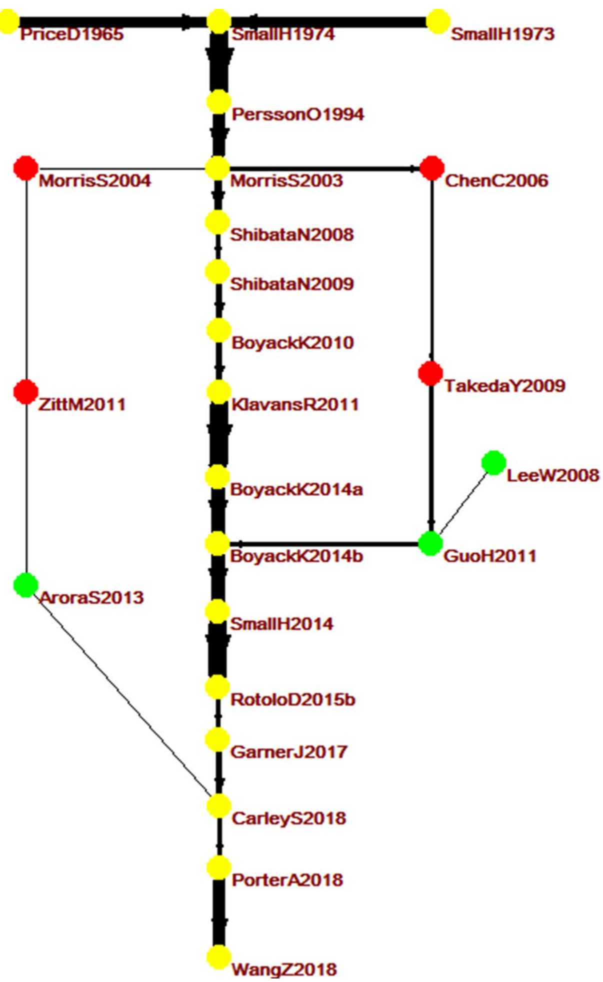

Fig. 6 The branch paths based on the global main path for the emerging research topics field. The citation links thicken by their SPLC values. The papers appearing in the global main path are represented by the yellow nodes, the top 3 papers are represented by the green nodes. (Color figure online)

The other subgroup is mainly composed by the works from Professor Alan Porter and his research team. They have been working in the field of technology foresight and have developed several emergence indicators. Unlike Rotolo et al. (2015), four attributes of emerging research topics are taken into consideration: novelty, persistence, community, and growth. Garner et al. (2017), Carley et al. (2018), and Porter et al. (2018) developed a script to calculate emergence scores and set a family of viable technical emergence indicator based on the described emergence scoring. The primary emergence indicators identify "hot topics" and can be used to generate secondary indicators that reflect countries, organizations or authors actively engaging these hot topics. Wang et al. (2018) chose 3D printing as a case study and employed methodological steps in Porter et al. (2018) to identify emerging research topics of technological convergence using patent information.

Discussion and conclusion

Knowledge diffusion trajectory of emerging research topics field is characterized by the local and global key-route main paths. According to convergence-divergence patterns in the main paths, three different development stages can be observed: the emergence, exploration and development stages. After the research fronts are defined in the first stage, more and more attentions are paid to emerging research topics field, which result directly in hundreds of scientific publications in the second stage. It promotes in turn the rapid development of emerging research topics, including methodologies of identification and detection, analysis of characterization, description and visualization. Nevertheless, bibliographic methods are still widely used and developed by a lot of researchers, and citation analysis is considered as a useful tool for identifying emerging research topics.

Intuitively, due to many known problems (such as time-lagging problem) with citation-based methods and large-scale accessible text resources, the lexical-based approaches may be considered as alternatives or supplementations to recognize emerging research topics. For example, since burst-detection algorithm (Kleinberg 2003) can be adapted for detecting sharp increases of interest in target domains, it is general enough to be applied to a time series of multi-word terms or citations of articles (Chen 2006; Guo et al. 2011). However, as one can see from the branch paths in Fig. 6 that the citation-based method has been the mainstream approach of studying emerging research topics in the exploration stage. After 2015 (i.e., the development stage), with the development of large-scale text-processing technologies, machine learning methods are gradually utilized to detect emerging research topics due to their potential power, such as the feed-forward multilayer neural network (Lee et al. 2018).

In addition, from the key-route main paths in the emerging research topics field, several drifts can be observed as follows: (1) the identification methods are drifting from citation-based approaches to machine learning based ones, such as JoungJ2017 and leeC2018 in the Figs. 4 and 5. (2) The research direction transfers from the measurement (such as ShibataN2008 and ShibataN2009) to the identification (e.g. PorterA2018, WangZ2018 and so on). (3) Information resource is drifting from the papers to the patents (e.g. JoungJ2017, LeeC2018, PorterA2018 and WangZ2018 in Figs. 4, 5). The above drifts indicate that the scholars have being realized the complexity of the problem on detecting emerging research topics, and tried to solve this problem from different perspectives.

In the near future, there is still room to improve the identification method and to expand the types of information resources (such as scientific reports, projects and so on). What's more, to our knowledge, there are no benchmark datasets public available with known emerging research topics until now. Currently, the judgment on the detected emerging topics still relies largely on expert-centric approaches such as Delphi and large-scale survey methods. Therefore, evaluation of emerging research topics also deserves to be explored.

Acknowledgements This work was supported partially by the Social Science Foundation of Beijing Municipality (Grant Number 17GLB074), and Natural Science Foundation of Guangdong Province (Grant Number 2018A030313695). Our gratitude also goes to the anonymous reviewers and the editor for their valuable comments.

References

Adner, R., & Levinthal, D. (2002). The emergence of emerging technologies. *California Management Review, 45*(1), 50–66.

Aris, A., Shneiderman, B., Qazvinian, V., & Radev, D. (2009). Visual overviews for discovering key papers and influences across research fronts. *Journal of the American Society for Information Science and Technology, 60*(11), 2219–2228.

Åström, F. (2007). Changes in the LIS research front: Time-sliced cocitation analyses of LIS journal articles, 1990–2004. *Journal of the American Society for Information Science and Technology, 58*(7), 947–957.

Azoulay, P. (2019). Small research teams 'disrupt' science more radically than large ones. *Nature, 566,* 330–332.

Batagelj, V. (2003). *Efficient algorithms for citation network analysis*. University of Ljubljana, Institute of Mathematics, Physics and Mechanics, Department of Theoretical Computer Science.

Batagelj, V., Ferligoj, A., & Squazzoni, F. (2017). The emergence of a field: A network analysis of research on peer review. *Scientometrics, 113*(1), 503–532.

Batagelj, V., & Mrvar, A. (1998). Pajek—Program for large network analysis. *Connections, 21*(2), 47–57.

Bettencourt, L., Kaiser, D., Kaur, J., Castillo-Chávez, C., & Wojick, D. (2008). Population modeling of the emergence and development of scientific fields. *Scientometrics, 75*(3), 495–518.

Bhupatiraju, S., Nomaler, Ö., Triulzi, G., & Verspagen, B. (2012). Knowledge flows—Analyzing the core literature of innovation, entrepreneurship and science and technology studies. *Research Policy, 41*(7), 1205–1218.

Bornmann, L., & Tekles, A. (2019). Disruptive papers published in scientometrics. *Scientometrics, 120*(1), 331–336.

Boyack, K., & Klavans, R. (2010). Co-citation analysis, bibliographic coupling, and direct citation: Which citation approach represents the research front most accurately? *Journal of the American Society for Information Science and Technology, 61*(12), 2389–2404.

Boyack, K., & Klavans, R. (2014). Creation of a highly detailed, dynamic, global model and map of science. *Journal of the Association for Information Science and Technology, 65*(4), 670–685.

Boyack, K., Klavans, R., Small, H., & Ungar, L. (2014). Characterizing the emergence of two nanotechnology topics using a contemporaneous global micro-model of science. *Journal of Engineering and Technology Management, 32,* 147–159.

Burmaoğlu, S., Sartenaer, O., Porter, A., & Li, M. (2019). Analysing the theoretical roots of technology emergence: An evolutionary perspective. *Scientometrics, 119*(1), 97–118.

Calabrese, A., Castaldi, C., Forte, G., & Levialdi, N. (2018). Sustainability-oriented service innovation: An emerging research field. *Journal of Cleaner Production, 193,* 533–548.

Carley, S., Newman, N., Porter, A., & Garner, J. (2018). An indicator of technical emergence. *Scientometrics, 115*(1), 35–49.

Chang, P., Wu, C., & Hoang-Jyh, L. (2010). Evolution of research subjects in library and information science based on keyword, bibliographical coupling, and co-citation analyses. *Scientometrics, 82*(1), 5–19.

Chen, C. (2006). CiteSpace II: Detecting and visualizing emerging trends and transient patterns in scientific literature. *Journal of the American Society for Information Science and Technology, 57*(3), 359–377.

Chen, C., Ibekwe-SanJuan, F., & Hou, J. (2010). The structure and dynamics of cocitation clusters: A multiple-perspective cocitation analysis. *Journal of the American Society for Information Science and Technology, 61*(7), 1386–1409.

Chen, K., Luesukprasert, L., & Chou, S. (2007). Hot topic extraction based on timeline analysis and multidimensional sentence modeling. *IEEE Transactions on Knowledge and Data Engineering, 19*(8), 1016–1025.

Cozzens, S., Gatchair, S., Kang, J., Kim, K., Lee, H., Ordóñez, G., & Porter, A. (2010). Emerging technologies: Quantitative identification and measurement. *Technology Analysis & Strategic Management, 22*(3), 361–376.

de Solla Price, D. (1965). Networks of scientific papers. *Science, 149*(3683), 510–515.

Funk, R., & Owen-Smith, J. (2017). A dynamic network measure of technological change. *Management Science, 63*(3), 791–817.

Garner, J., Carley, S., Porter, A., & Newman, N. (2017). Technological emergence indicators using emergence scoring. In *2017 Portland international conference on management of engineering and technology (PICMET).*

Glänzel, W., & Thijs, B. (2012). Using 'core documents' for detecting and labelling new emerging topics. *Scientometrics, 91*(2), 399–416.

Glassey, O. (2009). Exploring the weak signals of starts-ups as a folksonomic system. *Technology Analysis & Strategic Management, 21*(3), 321–332.

Guo, H., Weingart, S., & Börner, K. (2011). Mixed-indicators model for identifying emerging research areas. *Scientometrics, 89*(1), 421–435.

Halaweh, M. (2013). Emerging technology: What is it? *Journal of Technology Management and Innovation, 8*(3), 19–20.

Ho, J., Saw, E., Lu, L., & Liu, J. (2014). Technological barriers and research trends in fuel cell technologies: A citation network analysis. *Technological Forecasting and Social Change, 82,* 66–79.

Hummon, N., & Doreain, P. (1989). Connectivity in a citation network: The development of DNA theory. *Social Networks, 11*(1), 39–63.

Jang, S., Yu, Y., & Wang, T. (2011). Emerging firms in an emerging field: an analysis of patent citations in electronic-paper display technology. *Scientometrics, 89*(1), 259–272.

Jarić, I., Knežević-Jarić, J., & Lenhardt, M. (2014). Relative age of references as a tool to identify emerging research fields with an application to the field of ecology and environmental sciences. *Scientometrics, 100*(2), 519–529.

Jarneving, B. (2005). A comparison of two bibliometric methods for mapping of the research front. *Scientometrics, 65*(2), 245–263.

Jarneving, B. (2007). Bibliographic coupling and its application to research-front and other core documents. *Journal of Informetrics, 1*(4), 287–307.

Joung, J., & Kim, K. (2017). Monitoring emerging technologies for technology planning using technical keyword based analysis from patent data. *Technological Forecasting and Social Change, 114,* 281–292.

Kim, M., & Chen, C. (2015). A scientometric review of emerging trends and new developments in recommendation systems. *Scientometrics, 104*(1), 239–263.

Klavans, R., & Boyack, K. (2011). Using global mapping to create more accurate document-level maps of research fields. *Journal of the American Society for Information Science and Technology, 62*(1), 1–18.

Kleinberg, J. (2003). Bursty and hierarchical structure in streams. *Data Mining and Knowledge Discovery, 7*(4), 373–397.

Kuhlmann, S., Stegmaier, P., & Konrad, K. (2019). The tentative governance of emerging science and technology—A conceptual introduction. *Research Policy, 48*(5), 1091–1097.

Kyebambe, M., Cheng, G., Huang, Y., He, C., & Zhang, Z. (2017). Forecasting emerging technologies: A supervised learning approach through patent analysis. *Technological Forecasting and Social Change, 125,* 236–244.

Lee, W. (2008). How to identify emerging research fields using scientometrics: An example in the field of Information Security. *Scientometrics, 76*(3), 503–525.

Lee, C., Kwon, O., Kim, M., & Kwon, D. (2018). Early identification of emerging technologies: A machine learning approach using multiple patent indicators. *Technological Forecasting and Social Change, 127,* 291–303.

Li, M. (2017). An exploration to visualise the emerging trends of technology foresight based on an improved technique of co-word analysis and relevant literature data of WOS. *Technology Analysis & Strategic Management, 29*(6), 655–671.

Liang, H., Wang, J., Xue, Y., & Cui, X. (2016). IT outsourcing research from 1992 to 2013: A literature review based on main path analysis. *Information & Management, 53*(2), 227–251.

Liu, C., & Gui, Q. (2016). Mapping intellectual structures and dynamics of transport geography research: A scientometric overview from 1982 to 2014. *Scientometrics, 109*(1), 159–184.

Liu, J., Lu, L., & Ho, M. (2019). A few notes on main path analysis. *Scientometrics, 119*(1), 379–391.

Liu, J., & Lu, L. (2012). An integrated approach for main path analysis: Development of the Hirsch index as an example. *Journal of the Association for Information Science and Technology, 63*(3), 528–542.

Liu, J., Lu, L., Lu, W., & Lin, B. (2013b). Data envelopment analysis 1978–2010: A citation-based literature survey. *Omega, 41*(1), 3–15.

Liu, X., Jiang, T., & Ma, F. (2013a). Collective dynamics in knowledge networks: Emerging trends analysis. *Journal of Informetrics, 7*(2), 425–438.

Liu, Y., Lin, D., Xu, X., Shan, S., & Sheng, Q. (2018). Multi-views on nature index of Chinese academic institutions. *Scientometrics, 114*(3), 823–837.

Lu, C., Hou, H., Ding, Y., & Zhang, C. (2019). Review of international studies on discovering emerging topics. *Journal of the China Society for Scientific and Technical Information, 38*(1), 97–110.

Ma, V., & Liu, J. (2016). Exploring the research fronts and main paths of literature: A case study of shareholder activism research. *Scientometrics, 109*(1), 33–52.

Mogoutov, A., & Kahane, B. (2007). Data search strategy for science and technology emergence: A scalable and evolutionary query for nanotechnology tracking. *Research Policy, 36*(6), 893–903.

Morris, S., Yen, G., Wu, Z., & Asnake, B. (2003). Time line visualization of research fronts. *Journal of the American Society for Information Science and Technology, 54*(5), 413–422.

Naaman, M., Becker, H., & Gravano, L. (2011). Hip and trendy: Characterizing emerging trends on twitter. *Journal of the American Society for Information Science and Technology, 62*(5), 902–918.

Ohniwa, R., Hibino, A., & Takeyasu, K. (2010). Trends in research foci in life science fields over the last 30 years monitored by emerging topics. *Scientometrics, 85*(1), 111–127.

Persson, O. (1994). The intellectual base and research fronts of JASIS 1986–1990. *Journal of the American Society for Information Science, 45*(1), 31–38.

Porter, A. L., Garner, J., Carley, S. F., & Newman, N. C. (2018). Emergence scoring to identify frontier R&D topics and key players. *Technological Forecasting and Social Change, 146*, 628–643.

Raghuram, S., Tuertscher, P., & Garud, R. (2010). Mapping the field of virtual work: A cocitation analysis. *Information Systems Research, 21*(4), 983–999.

Reiss, T., Vignola-Gagné, E., Kukk, P., Glänzel, W., & Thijs, B. (2013). ERACEP – Emerging Research Areas and their Coverage by ERC-supported Projects. *Technical Report European Research Council.*

Roche, I., Besagni, D., Francois, C., Horlesberger, M., & Schiebel, E. (2010). Identification and characterisation of technological topics in the field of molecular biology. *Scientometrics, 82*(3), 663–676.

Rohrbeck, R., Battistella, C., & Huizingh, E. (2015). Corporate foresight: An emerging field with a rich tradition. *Technological Forecasting and Social Change, 101,* 1–9.

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

Sangam, S. (2000). Emerging trends in scientometrics: Essays in honour of Dr. Ashok Jain. *Scientometrics, 47*(1), 165–166.

Scalise, K., Bernbaurn, D., & Timms, M. (2007). Adaptive technology for e-learning: Principles and case studies of an emerging field. *Journal of the American Society for Information Science and Technology, 58*(14), 2295–2309.

Shibata, N., Kajikawa, Y., Takeda, Y., & Matsushima, K. (2008). Detecting emerging research fronts based on topological measures in citation networks of scientific publications. *Technovation, 28*(11), 758–775.

Shibata, N., Kajikawa, Y., Takeda, Y., & Matsushima, K. (2009). Comparative study on methods of detecting research fronts using different types of citation. *Journal of the American Society for Information Science and Technology, 60*(3), 571–580.

Smalheiser, N. (2001). Predicting emerging technologies with the aid of text-based data mining: The micro approach. *Technovation, 21*(10), 689–693.

Small, H. (1973). Co-citation in the scientific literature: a new measure of the relationship between two documents. *Journal of the American Society for Information Science, 24,* 265–269.

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

Small, H., & Griffith, B. (1974). The structure of scientific literatures I: Identifying and graphing specialties. *Science Studies, 4*(1), 17–40.

Soriano, A., Alvarez, C., & Valdes, R. (2018). Bibliometric analysis to identify an emerging research area: Public relations intelligence—A challenge to strengthen technological observatories in the network society. *Scientometrics, 115*(3), 1591–1614.

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

Toivanen, H. (2014). The shift from theory to innovation: The evolution of Brazilian research frontiers 2005–2011. *Technology Analysis & Strategic Management, 26*(1), 105–119.

Tu, Y., & Seng, J. (2012). Indices of novelty for emerging topic detection. *Information Processing and Management, 48*(2), 303–325.

Upham, S., & Small, H. (2010). Emerging research fronts in science and technology: Patterns of new knowledge development. *Scientometrics, 83*(1), 15–38.

Verspagen, B. (2007). Mapping technological trajectories as patent citation networks: A study on the history of fuel cell research. *Advances in Complex Systems, 10*(1), 93–115.

Wang, Q. (2018). A bibliometric model for identifying emerging research topics. *Journal of the Association for Information Science and Technology, 69*(2), 290–304.

Wang, Z., Porter, A., Wang, X., & Carley, S. (2018). An approach to identify emergent topics of technological convergence: A case study for 3D printing. *Technological Forecasting and Social Change, 146*, 723–732.

Weismayer, C., & Pezenka, I. (2017). Identifying emerging research fields: A longitudinal latent semantic keyword analysis. *Scientometrics, 113*(3), 1757–1785.

Wu, L., Wang, D., & Evans, J. (2019). Large teams develop and small teams disrupt science and technology. *Nature, 566,* 378–382.

Xie, P. (2015). Study of international *anticancer* research trends via co-word and document co-citation visualization analysis. *Scientometrics, 105*(1), 611–622.

Xu, S., Hao, L., An, X., Yang, G., & Wang, F. (2019). Emerging research topics detection with multiple machine learning models. *Journal of Informetrics* (accepted).

Yeo, W., Kim, S., Lee, J., & Kang, J. (2014). Aggregative and stochastic model of main path identification: A case study on graphene. *Scientometrics, 98*(1), 633–655.

Zhao, D., & Strotmann, A. (2008). Evolution of research activities and intellectual influences in information science 1996–2005: Introducing author bibliographic-coupling analysis. *Journal of the American Society for Information Science and Technology, 59*(13), 2070–2086.

Zhao, D., & Strotmann, A. (2014). The knowledge base and research front of information science 2006–2010: An author cocitation and bibliographic coupling analysis. *Journal of the Association for Information Science and Technology, 65*(5), 995–1006.

Zhu, H., Yin, X., Ma, J., & Hu, W. (2016). Identifying the main paths of information diffusion in online social networks. *Physica A, 452*(15), 320–328.