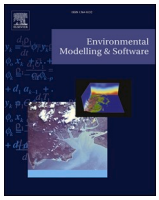
Environmental Modelling and Software 149 (2022) 105320



Contents lists available at ScienceDirect

Environmental Modelling and Software

journal homepage: www.elsevier.com/locate/envsoft



Identification of research communities of environmental engineering and their evolution using coauthor network analysis

Jaebeom Park [a], Jeryang Park [b], Yongju Choi [a,c,*]

[a] Department of Civil and Environmental Engineering, Seoul National University, Seoul, 08826, Republic of Korea
[b] Department of Civil and Environmental Engineering, Hongik University, Seoul, 04066, Republic of Korea
[c] Institute of Construction and Environmental Engineering, Seoul National University, Seoul, 08826, Republic of Korea

ARTICLE INFO

Keywords:
Bibliometric analysis
Coauthor network
Community evolution
Environmental engineering
Research pattern
Trend analysis

ABSTRACT

Networks of coauthorship for scientific publication well represent the collaboration dynamics among members of a scientific community. Here, we employed a coauthor network analysis to identify research communities and their evolutionary patterns as the first attempt to characterize the dynamics of collaboration in environmental engineering. Using bibliometric data from 2007 to 2018 in environmental engineering, 20 research communities were identified and two major drivers of scientific collaboration were revealed: research interest and country of one's affiliation. We developed a novel methodology that enables tracking cluster's evolutionary patterns. This methodology allowed systematic stereotyping of each cluster as "mature" or "emerging". The two stereotypes did not show a significant difference in the annual growth rate of the volume of publications, implying that comparable attention has been paid to advancing the knowledge of formerly-established subjects and exploring innovative subjects of environmental engineering.

1. Introduction

Environmental engineering is an academic discipline that aims to satisfy human needs by protecting the environment with resolution of environmental issues (Reible, 1999). Because problems of this discipline are not intentionally designed, they are inherently complicated and working with them requires a wide breadth of scientific knowledge in most cases. For example, domestic sewage reclamation processes require knowledge that stems from all basic disciplines of natural science including physics, chemistry, and biology due to the complexity of sewage constituents that should be removed. Hazardous chemicals released into the environment can cause multi-media pollution problems. Assessing human health risks imposed by this pollution problem requires multiple fields of applied science including analytical chemistry, fluid mechanics, meteorology, toxicology, and so forth. The wide breadth of scientific knowledge required to work with environmental engineering problems implies that advance in the academic discipline of environmental engineering necessitates collaboration among scholars with different academic backgrounds (Davis and Masten, 2020; Khan et al., 2019). Thus, tracking and characterizing collaboration among scholars in environmental engineering should provide valuable information for community members of this discipline to help understand the status quo of their current research topics, explore directions of their future research, and build a strategic plan for future collaboration.

Scientific collaboration is one of principal strategies for scholars to improve scientific productivity and quality of their research and to efficiently disseminate the knowledge (Barabás et al., 2017; McCarty et al., 2013). Analyzing scientific collaboration as a network enables disclosing hidden relationships between scientists. Among the currently available approaches of scientific collaboration network analyses, the coauthor network analysis, which represents authors and their coauthoring as nodes and edges, respectively, is regarded to be the most powerful, providing the best projection of real research collaboration patterns (Glänzel and Schubert, 2004). In addition to recording the comprehensive behavior of researchers regarding how often and with how many colleagues they coauthor, communities of authors who have common denominators can be marked by identifying densely connected groups within the network, called clusters (Porter et al., 2009).

Herein, for the first time, we applied coauthor network analysis to the international scientific literature of environmental engineering. The primary objective of the current work was to provide a scientific, systematic, and unbiased assessment of gross features of collaboration in the academic field of environmental engineering. We employed coauthor network analysis to identify subgroups of intimate intra-

* Corresponding author. Department of Civil and Environmental Engineering, Seoul National University, Seoul, 08826, Republic of Korea.
E-mail address: ychoi81@snu.ac.kr (Y. Choi).

https://doi.org/10.1016/j.envsoft.2022.105320
Received 15 December 2021; Accepted 12 January 2022
Available online 15 January 2022
1364-8152/© 2022 The Authors. Published by Elsevier Ltd. This is an open access article under the CC BY license (http://creativecommons.org/licenses/by/4.0/).

collaboration (which we referred to as research communities of environmental engineering in this study), characterized each subgroup, and investigated their evolutionary dynamics. To the best of our knowledge, except for our recent paper on domestic environmental engineering in South Korea (Park et al., 2022), any peer-reviewed article that analyzes the scientific literature of environmental engineering in the context of coauthor network has not been reported yet. Widening our search to studies that attempted such analysis in any context returned very few works. An example is a recent work that has analyzed papers published in *Environmental Science & Technology* from 2000 to 2019 (Zhu et al., 2021). This recent work focused on extracting and analyzing keywords while lacking any discussion on collaboration among researchers. Another unique feature of the current work was that in order to closely represent the gross academic field of environmental engineering, a relatively large set of data (142,129 authors with 57,534 papers in total) were collected for analysis. Coauthor network analysis has previously been employed to identify subdivisions of manifold academic disciplines (Barabás et al., 2017; Borrett et al., 2018; Huang et al., 2008; Newman, 2001, 2004; Settle et al., 2020; Velden et al., 2010) and to offer their prospects (Morel et al., 2009). Evidently, subject disciplines of these previous works, including ecological engineering (Borrett et al., 2018), epidemiology (Morel et al., 2009), and nanotechnology (Rueda et al., 2007), were much smaller in volume than environmental engineering. Relatively small volumes of data used in these previous works allowed clusters of the coauthor networks to be well evaluated for their evolution dynamics by tracking down to an individual level (Huang et al., 2008; Tomassini and Luthi, 2007). On the other hand, clusters identified in the current study were too large to be studied by this simple qualitative approach. For this reason, a novel method was developed and employed to semi-quantitatively evaluate the evolutionary pattern of clusters of coauthors.

2. Methods

2.1. Data and network

For the current analysis, we collected all 68,811 papers published from 2007 to 2019 in eight journals (Table S1 in Supplementary Material) whose impact factors were within the top 10% in the category of "Engineering, Environmental" or "Environmental Sciences" in the 2018 Journal Citation Reports issued by the Web of Science (WOS) and at the same time, whose official websites explicitly stated their scopes to be topics of environmental engineering. The WOS database was used to extract bibliographical data of papers, including title, keyword, author name, country of author's affiliation, journal name, and year of publication. Data collection was commenced in 2007 because before then, names of authors were recorded with their given names abbreviated, which imposed a significant challenge in author identification. A coauthor network was built using data from papers published between 2007 and 2018. The network had a single giant component, i.e., the one in which the majority of the volume of the whole network was found among connected subsets of a network (Newman, 2001). The presence of the giant component (accounting for 79% of all authors of the constructed network) and its scale-free degree distribution (Fig. S1) featured a typical mature social network (Barabási and Albert, 1999; Newman, 2001). The basic properties of the giant component are summarized in Table S2. Data collection and network construction were performed using Python code written by us. Network characterization was performed using Gephi software (Bastian et al., 2009). Data from papers published in 2019 were used as a validation set for the cluster growth model described below.

2.2. Cluster identification and analysis

Clustering was executed for the giant component. We adopted Leiden algorithm, one of the fastest and best-performing cluster identification algorithms (Blondel et al., 2008; Lancichinetti and Fortunato, 2009; Traag et al., 2019; Yang et al., 2016), using Gephi. We adjusted the criterion for clustering as density with a resolution set as the universal value of the giant component (5.348×10^{-5}). This enabled the identification of clusters or research communities whose members collaborated more intensively with those who belonged to the same cluster than the rest (Traag et al., 2019). Among 821 clusters identified by this means, we limited our definition of research communities to those containing over 1% of whole network members ($\approx$ 1,006). This resulted in a total of 20 research communities. They were named from ***A*** to ***T*** in descending order of the number of members.

A novel methodology was developed to explore evolutionary patterns of research communities over time. Sub-datasets were created using snowball sampling (Huang et al., 2008) [i.e., by collecting data that corresponded to cumulative time intervals of (2007–2007), (2007–2008), (2007–2009), and so forth]. The clustering algorithm described above was applied to each sub-dataset, yielding a set of clusters for each cumulative time interval. The ratio of "pre-clustered" members, $R_i(t)$, and the normalized time interval, $X(t)$, were obtained using the following Eqs. (1) and (2), respectively:

$$R_i(t) = \frac{max\{n(C_i(t) \cap c_{t,j})\}}{n(C_i(t))}; \quad i = A,\ B,\ C,\ \cdots; \quad j = A',\ B',\ C',\ \cdots \tag{1}$$

$$X(t) = \frac{t - t_0}{T - t_0} \tag{2}$$

where t was the last calendar year of the analyzed time interval; t_0 and T were the first and the last calendar year of the whole dataset, respectively; i indicated research communities identified for the whole dataset; $C_i(t)$ was the group of members of a research community i who had published at least one article by year t; $c_{t,j}$ was the group of cluster members labeled j, which was identified using a sub-dataset of publications between year t_0 and t; and n was the number of members. Among intersecting groups between multiple $c_{t,j}$ s and $C_i(t)$, members in the largest group were marked with "pre-clustered" and the rest with "complementary". $R_i(t)$ indicated the ratio of already "clustered" members among existing members in cluster i at time interval $[t_0, t]$. Of course, more $c_{t,j}$ s at each sub-dataset sharing less numbers of members with $C_i(t)$ may exist. However, some $c_{t,j}$ s did not retain a sufficient size to represent explicit characteristics in further keyword analysis. Therefore, we only focused on the maximally sharing group to allow tracking in our analysis. Considering "complementary" members would be clustered in the forthcoming period, tracking research interests of "pre-clustered" and "complementary" members could reveal the exhaustive evolutionary process of the research community. The process of determining $R_i(t)$ is illustrated in Fig. 1.

Annual growth dynamics of research communities was characterized using the following power function:

$$Y_i(t) = \mathrm{X}(t)^{b_i} \tag{3}$$

where $Y_i(t)$ was the normalized publication volume, referring to the fraction of research community i's publications published between t_0 and t relative to those between t_0 and T; and b_i was the power function exponent. It has been demonstrated that power function can well reproduce the temporal growth of a cumulative quantity with the value of the exponent representing the rapidness of growth (Manchein et al., 2020; Merrin, 2020; Zang et al., 2016; Zhu et al., 2013).

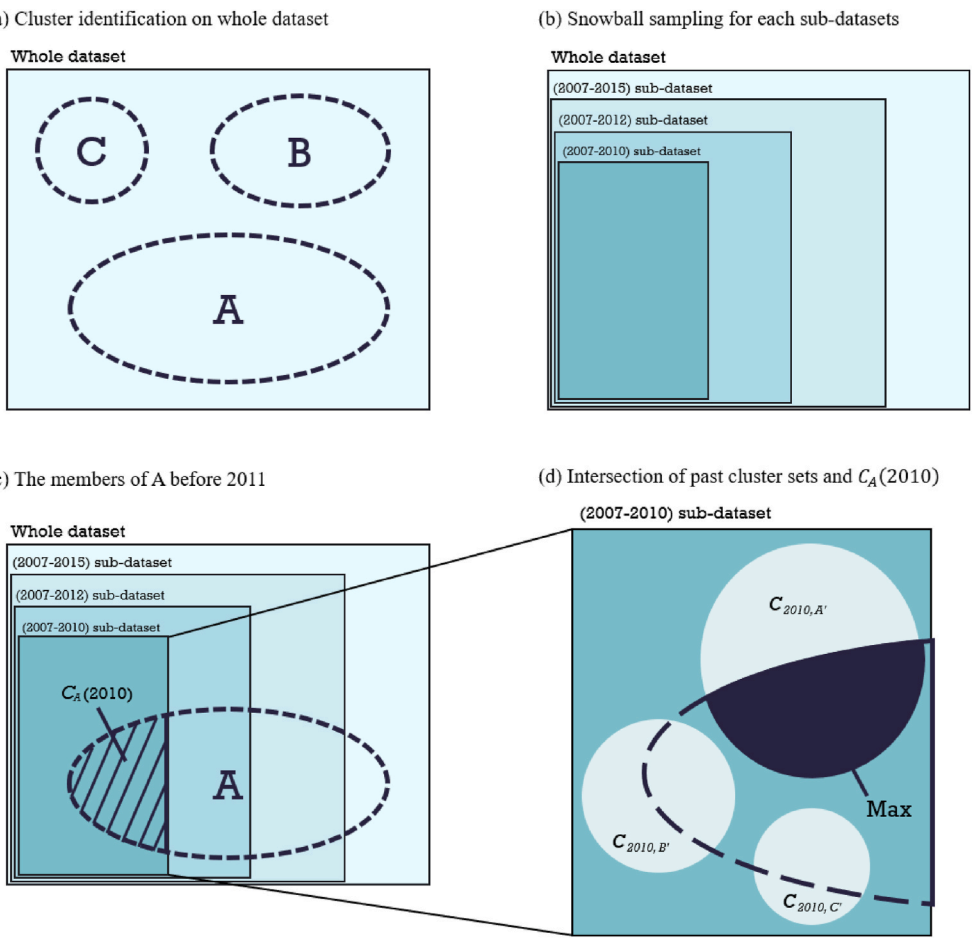


Fig. 1. Schematic of the concept of $\boldsymbol{R_i(t)}$ with the example of $\boldsymbol{R_A}(2010)$ (a) Clusters (i.e., research communities) were identified by cluster analysis in the whole-period dataset from 2007 to 2018. (b) To investigate the behavior of each cluster over time, 11 sub-datasets were formed in cumulative time intervals (2007–2007), (2007–2008), (2007–2009), and so on. (c) Among members of the research community, members who published their first papers before 2011 were captured in a sub-dataset (2007–2010) named $C_A(2010)$. (d) Past clusters identified from the same cluster analysis were $c_{2010,A'}$, $c_{2010,B'}$, $c_{2010,C'}$, and so on. Among the intersection between past clusters and $C_A(2010)$, the maximum area over $C_A(2010)$ was $R_A(2010)$. Members who belonged to the maximum area were "pre-clustered" members. The rest of them were "complementary" members.

3. Results and discussion

3.1. Research community characterization

Table 1 summarizes key features of each research community, including keywords of articles and countries of affiliations of authors with the most frequent appearance. An extended list of major keywords is given in Table S3. These major keywords were analyzed to define major research interests of each cluster, which is given in the third column of Table 1. Keywords well representing research interests are marked in bold in Table 1 and Table S3. Evaluation of research interests of the most influential researchers in each community (Table S4) justified our definition of major research interests. Countries that showed a notably (i.e., more than 2-fold) greater country occupancy ratio for a given cluster than that for the giant component are also marked in bold in Table 1. We defined country occupancy ratio as the ratio of the number of author appearances with a given country of affiliation relative to the total number of author appearances. The giant component's country occupancy ratios were in the order of USA (30.7%) > China (27.3%) > Canada (6.1%) > Spain (5.8%) > France (5.1%) and so forth.

Keywords of articles and affiliations of the authors were two determinants that defined the identity of each research community. Keyword analysis revealed that most research communities represented distinct research topics of environmental engineering. These spanned from classical topics such as persistent organic pollutants (POPs; communities ***B*** and ***I***) and biological water and wastewater treatment (***J***) to those that had recently become popular such as nanoparticles, electrochemistry, and photochemistry (***K*** and ***O***). In fact, in many cases, keywords (and in turn, research topics) alone were sufficient to differentiate a research community from others. Research interest was also observed to be the major driver of collaborative research in a domestic environmental engineering network (Park et al., 2022). Common research interests of each research community identified in the current work were scientific or technical subjects whereas those reported in the previous study would be better defined as environmental problems (e.g., inorganic waste management, indoor water quality). This is not surprising because international journals as sources of data sets for the current study highly prefer papers that discuss readily generalizable scientific and/or technical findings, whereas domestic journals would welcome submissions that address regional/local matters. Countries of authors' host institutions also made a significant contribution to the formation of research communities. For instance, communities ***A***, ***B***, ***N***, ***P***, and ***S*** showed unique country occupancy ratio profiles for China, Canada, Germany, South Korea, and USA, respectively, as primary countries. There were several pairs of communities whose research topics were analogous. These research topics could not easily be distinguished from each other without introducing country of the host institution. These included ***B*** and ***I*** pair and ***O*** and ***P*** pair. Major keywords of community ***A*** represented a much wider range of environmental engineering research topics than those of others. Community ***A*** had an extraordinarily large size of 13,510 members, which was 3.8-fold greater than that of the second-largest community. The uniqueness of community ***A*** might be attributed to an occurrence of strong driving force for intra-country collaboration in a country that a large number of authors belong to, that is, China. The return of young Chinese scientists who had studied abroad to their homeland might be one of the primary reasons for the formation of this community. Relevant studies have shown that the proportion of Chinese graduates from U.S. universities who return to China has substantially increased since the beginning of the 21st century (Kellogg, 2012) and that these "early returners" strongly prefer domestic

Table 1
Summary of characteristics of each research community. Major research interest is defined using keyword analysis and confirmed by evaluating representative authors of each community. Annual growth rate is the exponent value (i.e., the b-value) of the power function regression for normalized number of publications. Research communities are classified into two categories: 1) mature community (M); and 2) emerging community (E).

	Number of members (proportion relative to the total)	Major research interest	Top 5 keywords	Top 3 countries of affiliation (country occupancy ratio)	Category	Annual growth rate
A	13,510 (13.42%)	Water and wastewater treatment (general)	**water**, degradation, adsorption, aqueous solution, **removal**	**China** (83.7%), USA (24.8%), Australia (3.7%)	M	1.44
B	3,560 (3.54%)	Occurrence, fate, and risk of POPs	PCBs, PAHs, PBDEs, **POPs**, water	**Canada** (28.4%), USA (23.4%), China (12.2%)	M	1.11
C	3,383 (3.36%)	Occurrence and fate of atmospheric pollutants	**emissions**, united states, particulate matter, particles, **secondary organic aerosol**	**USA** (71.4%), China (19.8%), Canada (10%)	M	1.17
D	3,222 (3.2%)	Risk posed by atmospheric pollutants	**air pollution**, **exposure**, particulate matter, mortality, **health**	USA (37.5%), **Spain** (19.1%), **England** (16.7%)	M	1.51
E	2,554 (2.54%)	Micropollutant treatment in water and wastewater	**drinking water**, **wastewater**, **pharmaceuticals**, aqueous solution, **removal**	**USA** (63.1%), **Switzerland** (26.2%), **Germany** (12.8%)	M	1.13
F	2,552 (2.54%)	Fate of chemicals in the subsurface and sediment	adsorption, reduction, **sorption**, **soil**, **groundwater**	**USA** (79.6%), China (21.7%), France (5.8%)	M	1.21
G	2,253 (2.24%)	Toxicity and risk of water pollutants	**toxicity**, wastewater, PAHs, **exposure**, **water**	USA (46.6%), **England** (19.8%), **Germany** (19.4%)	E	1.24
H	2,039 (2.03%)	Public health concerns via daily chemical exposure	**exposure**, PBDEs, united states, PCBs, **house dust**	**USA** (84.8%), **Canada** (14.6%), China (9.2%)	E	1.36
I	1,947 (1.93%)	Occurrence, fate, and risk of POPs	**PBDEs**, wastewater, water, **toxicity**, **PAHs**	China (53.9%), USA (29.5%), **Canada** (21.7%)	E	1.32
J	1,894 (1.88%)	Biological water and wastewater treatment	**wastewater**, **activated sludge**, wastewater treatment, removal, system	**Australia** (23.1%), China (23.1%), USA (22.5%)	M	1.23
K	1,858 (1.85%)	Environmental nanotechnology	toxicity, water, **nanomaterial**, nanoparticles, adsorption	**USA** (75.7%), China (17.1%), France (10.0%)	M	1.25
L	1,729 (1.72%)	Pathogen monitoring and source tracking	***Escherichia coli***, drinking water, **microbial source tracking**, wastewater, water	**USA** (74.5%), **England** (9.8%), **South Korea** (5.8%)	M	1.17
M	1,678 (1.67%)	Fate and bioavailability of hazardous substances	soil, **adsorption**, **heavy metal**, water, **speciation**	China (48.3%), USA (30.1%), **Australia** (18.7%)	M	1.36
N	1,535 (1.53%)	Contaminated site remediation (bioremediation focused)	degradation, **biodegradation**, **groundwater**, **soil**, water	**Germany** (39.5%), USA (33.7%), **Canada** (12.4%)	E	1.06
O	1,534 (1.52%)	Water and wastewater treatment (electrochemistry, photocatalysis, and nanomaterial focused)	degradation, aqueous solution, wastewater, TiO_2, water	USA (31.6%), **Spain** (21.8%), China (19.7%)	E	1.15
P	1,359 (1.35%)	Water and wastewater treatment (photocatalysis and bioelectrochemistry focused)	water, degradation, **oxidation**, wastewater, performance	USA (47.5%), **South Korea** (43.2%), China (19.3%)	E	1.22
Q	1,221 (1.21%)	Occurrence, fate, and bioaccumulation of mercury	United States, atmosphere, methylmercury, emissions, deposition	**USA** (76.7%), **Canada** (19.8%), China (11.3%)	E	1.25
R	1,204 (1.2%)	Water, wastewater, and groundwater treatment (general)	aqueous solution, **sorption**, adsorption, water, **removal**	**Switzerland** (24.7%), **France** (21.5%), USA (20.0%)	E	1.13
S	1,111 (1.1%)	Occurrence and fate of contaminants in water, wastewater, and groundwater (general)	water, drinking water, wastewater, soil, **adsorption**	**USA** (90.7%), China (5.3%), **South Korea** (4.9%)	E	1.09
T	1,092 (1.08%)	Eutrophication of natural water bodies	water, sediment, soil, **phosphorus**, **eutrophication**	USA (31.0%), **England** (28.7%), China (21.2%)	E	1.26

Abbreviations: POPs, persistent organic pollutants; PCBs, polychlorinated biphenyls; PAHs, polycyclic aromatic hydrocarbons; PBDEs, brominated flame retardant; PCR, polymerase chain reaction.

collaboration over international one (Jonkers and Tijssen, 2008).

We also tested if other variables, such as published year (Table S5) and published journals (Table S6), were effective descriptors of each community's identity and thus probable drivers of research collaboration. Communities did not show any distinct fingerprints regarding the composition of published years of their articles. Unique published journal composition was observed for certain communities, which could well be explained by the research interests of the community (Table S6). This indicates that keyword appearance and published journal profile strongly depend on each other. The former should be a more effective descriptor of a community's identity than the latter because it provides much more affluent and specific information on research interest.

Overall, the current study and a previous study demonstrate that there are two major drivers of collaboration for research publication in environmental engineering: common research interests and cultural/linguistic/geographic proximity. Forming or participating in a working group sharing common research interests has been recognized as the most widespread and efficient way to improve one's scientific productivity and impact (Borrett et al., 2018; Ductor et al., 2014; Glänzel and Schubert, 2004; McCarty et al., 2013). Previous studies have demonstrated that cultural and linguistic proximities are strong drivers of scientific coauthorship (Nagtegaal and de Bruin, 1994; Narin et al., 1991). Geographical proximity has also been shown to be a key driver of coauthorship because it is one of the major motivations for initiating collaborative research in pursuit of resolving regional/local environmental problems (Park et al., 2022).

3.2. Evolution pattern

The ratio of "pre-clustered" members among existing members in each research community, $R_i(t)$, is plotted against the normalized time interval, $X(t)$, in Fig. 2. Data for periods of (2007–2007) and (2007–2008) were not shown in the figure because coauthor networks created for these periods were too small to conduct clustering. Evidently, $R_i(t)$ is a function of time. It is low at the beginning, reaching unity as the time window expands closer to $X(t) = 1$. However, its trace over $X(t)$ can vary considerably community by community. Groups whose members have a relatively long history of intensive collaboration (i.e., mature groups) are expected to consistently maintain a high value of $R_i(t)$ whereas relatively new collaboration groups should show a low $R_i(t)$ value at earlier time windows. For example, community ***A*** had $R_A(2009) = 0.8$, meaning that 80% of members of ***A*** were in the same community at the time window of (2007–2009). This indicated that this community, the one dominated by Chinese authors with broad research interests, had been constituted in or prior to 2009. It was recognized as early as in 2006 that Chinese scientists became major players in scientific publications (Zhou and Leydesdorff, 2006). An analogous pattern is seen in other mature communities, such as ***B*** and ***C*** addressing classical environmental engineering topics of POPs and atmospheric pollutants (e.g., particulate matter and organic aerosols), respectively. Conversely, the $R_i(t)$ value of ***G*** showed an abrupt increase from 0.4 to 1 between the last two years (i.e., 2017 and 2018). This dramatic change can be explained by unification of separate communities in 2018 (i.e., the emergence of this research community from convergence of different fields). Tracking R_i(t) values of all research communities allows systematically stereotyping subjects of each into two categories: mature and emerging. Examining the $R_i(t)$ value for the earliest time window, i.e., $R_i(2009)$, along with the major research interest for each cluster suggests that $R_i(2009) = 0.6$ is a reasonable threshold to categorize communities. Result of this categorization can be found in the second last column of Table 1.

3.3. Emergence of fields

We demonstrated above that the change of $R_i(t)$ with $X(t)$ might be used to reckon how a research community emerged over time. Community G exhibited a unique $X(t) - R_i(t)$ relationship, allowing this task to be intuitive. However, the pattern of change of $R_i(t)$ for emerging communities showed a wide variation. In addition, many of these communities did not show such clear $X(t) - R_i(t)$ relationship. Thus, additional information was necessary to evaluate drivers of emergence. To this end, we recalled keyword analysis. Top 5 keywords of the "pre-clustered" and "complementary" groups at different time windows were

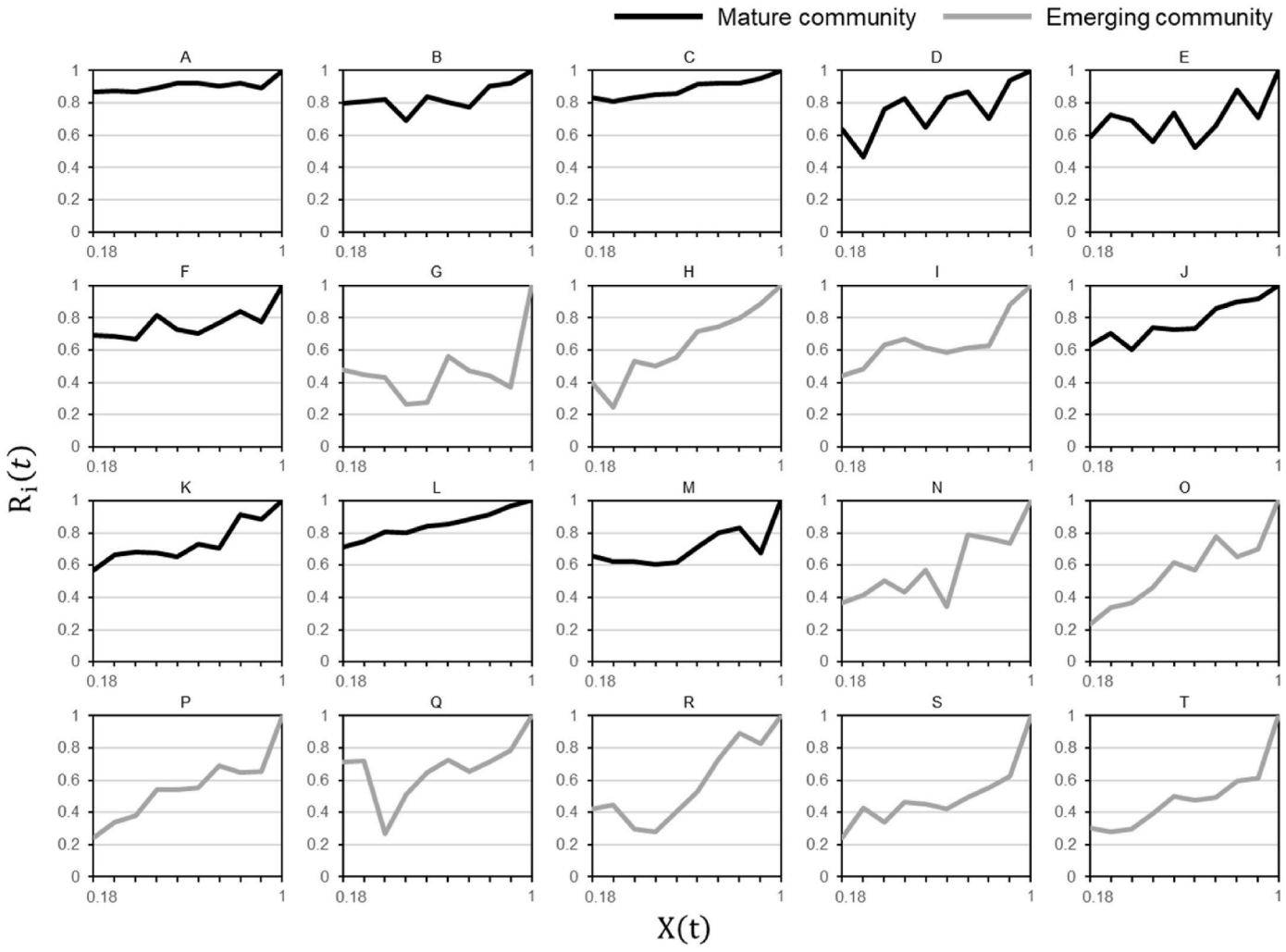


Fig. 2. Advance of clustering in each research community over time. $X(t)$ indicates the time interval, with the beginning year fixed to 2007 and the ending year of t, normalized to the whole study period of (2007–2018). $R_i(t)$ represents the ratio of "pre-clustered" members among existing members in research community i at the corresponding time interval. Data in periods (2007–2007) and (2007–2008) were excluded because coauthor networks created for these periods were too small to conduct clustering. Thus, the plot is given for an $X(t)$ ranging from 0.18 [period (2007–2009)] to 1 [period (2007–2018)].

identified for all research communities as shown in Table S7. Tracking changes in top keywords for the "pre-clustered" and "complementary" groups over time enabled us to investigate if a research community emerged by convergence of relevant fields or by congregation of researchers who worked in various fields to seek for a new field of research. *Groups of* ***G*** in 2017 had keywords of [toxicity, wastewater, plastic debris, drinking water, degradation] and [toxicity, wastewater, exposure, PAHs, United States] for the pre-clustered and complementary, respectively. Given that the $R_i(t)$ value of ***G*** remarkably increased in 2018, it could be deduced that two different groups of toxicity concerned with plastics and PAHs, respectively, converged at around the year of 2018. In a similar fashion with respect keywords variations in Table S7, the rest of emerging communities had been investigated excluding ***R***, ***S***, and ***T*** which were not large enough to arrange data over time. In ***H***, the $R_i(t)$ value increased from 2013 to 2014. Two groups with research interest in exposure to air and water had converged in that period. In ***I,*** groups of POPs working with different classes of contaminants merged in 2017. In ***N,*** groups were integrated in 2015 due to clustering of two contaminated site remediation groups of groundwater and soil. In ***O***, a group of AOPs using nanoparticles was combined with a group with a wider interest in wastewater treatment. In ***P***, researchers interested in AOPs technologies were integrated in 2018.

3.4. Relative growth rate

Annual growth dynamics of each community in publication volume, i.e., the $X(t) - Y_i(t)$ relationship, was excellently reproduced by the power function of Eq. (3) (Fig. S2). The growth rate exponent b_i for each cluster is given in the last column of Table 1. Validity of the model and exponents were further examined by an extrapolation to Year 2019. Correlation between model-predicted and real $Y_i(2019)$ values showed statistical significance at a threshold of 0.01 as determined by the Spearman correlation analysis. These results demonstrate that the growth rate exponent (i.e., b-value) can serve as an effective indicator to compare growth rates of the publication volume of a research community. All communities exhibited b-values greater than unity, indicating that they were all consistently growing in the annual volume of publications. The high growth rates of ***A*** and ***D*** can be ascribed by the remarkable growth of Chinese research output (Zhou and Leydesdorff, 2006) and increasing interest in health concerns related to air quality, respectively. Communities with POPs and bioremediation as major research interests (***B*** and ***N***) exemplify those with relatively slow growth rates. Despite the distinct difference in evolutionary patterns between mature and emerging community categories observed in cluster analysis (Fig. 2), average growth rates for these two categories were not statistically significantly different from each other ($p = 0.35$, Student's t-test; 1.26 ± 0.14 and 1.21 ± 0.10, respectively; mean $\pm$ standard deviation). This result suggests that the emergence of a research community of environmental engineering does not necessarily involve extraordinary growth in research output.

4. Implications

We conducted a coauthor network analysis for a significant portion of the international literature of environmental engineering research to identify twenty spontaneous communities with a relatively intense coauthorship. In-depth characterization of these communities revealed that major drivers of scientific collaboration through coauthoring were research interests and the country of the author's affiliation. A novel methodology to investigate the evolutionary behavior of research communities at the mesoscopic level arranged different evolutionary types of research communities such as "mature" and "emerging". All research communities have been steadily growing whether a related research topic is classic or novel, guaranteeing continuous development of environmental engineering as a promising discipline. This study is a pioneering study performing a community evolution analysis and scrutinizing collaboration in the field of environmental engineering. The analytical method developed herein allows investigation of collaborative behaviors of individuals in a research community and prediction of temporal trend of research topics in numerous complex multidisciplinary fields. As community evolution research is still in its infancy, follow-up studies are needed to further advance relevant analytical means that will help characterize various evolutionary patterns in detail and track maturation of a field after its emergence. By focusing on collaboration among individuals through scientific publications that eventually lead to a research community formation, this study allows ones to develop integrative understanding on contemporary research of environmental engineering and widen their insights on the field to strategize future contribution. It is expected that the current work will be combined with potential future reviews of environmental engineering bibliography focusing on different aspects of research activities to further assist players of the field in seeing "not only trees but also woods".

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

This paper relied on Web of Science (WOS) database supplied by the Clarivate Analytics. This research was supported by the BK21 Four research program of the National Research Foundation of Korea (NRF). Jeryang Park was supported by Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Science and ICT (NRF-2019R1C1C1008017). Jaebeom Park and Yongju Choi would like to thank the Institute of Engineering Research at Seoul National University for technical assistance.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.envsoft.2022.105320.

References

Barabás, B.l., Fülöp, O., Molontay, R., Pályi, G., 2017. Impact of the discovery of fluorous biphasic systems on chemistry: a statistical and network analysis. ACS Sustain. Chem. Eng. 5 (9), 8108–8118. https://doi.org/10.1021/acssuschemeng.7b01722.

Barabási, A.L., Albert, R., 1999. Emergence of scaling in random networks. Science 286 (5439), 509–512. https://doi.org/10.1126/science.286.5439.509.

Bastian, M., Heymann, S., Jacomy, M., 2009. Gephi: an open source software for exploring and manipulating networks. Icwsm 8, 361–362, 2009.

Blondel, V.D., Guillaume, J.L., Lambiotte, R., Lefebvre, E., 2008. Fast unfolding of communities in large networks. J. Stat. Mech. Theor. Exp. (10), P10008. https://doi.org/10.1088/1742-5468/2008/10/P10008, 2008.

Borrett, S.R., Sheble, L., Moody, J., Anway, E.C., 2018. Bibliometric review of ecological network analysis: 2010–2016. Ecol. Model. 382, 63–82. https://doi.org/10.1016/j.ecolmodel.2018.04.020.

Davis, M.L., Masten, S.J., 2020. In: Principles of Environmental Engineering & Science, fourth ed. McGraw-Hill Education, New York, pp. 2–13.

Ductor, L., Fafchamps, M., Goyal, S., van der Leij, M.J., 2014. Social networks and research output. Rev. Econ. Stat. 96 (5), 936–948. https://doi.org/10.1162/REST_a_00430.

Glänzel, W., Schubert, A., 2004. Analysing scientific networks through co-authorship. In: Handbook of Quantitative Science and Technology Research. Springer, pp. 257–276.

Huang, J., Zhuang, Z., Li, J., Giles, C.L., 2008. In: Collaboration over Time: Characterizing and Modeling Network Evolution, Proceedings of the 2008 International Conference on Web Search and Data Mining, pp. 107–116. https://doi.org/10.1145/1341531.1341548.

Jonkers, K., Tijssen, R., 2008. Chinese researchers returning home: impacts of international mobility on research collaboration and scientific productivity. Scientometrics 77 (2), 309–333. https://doi.org/10.1007/s11192-007-1971-x.

Kellogg, R.P., 2012. China's brain gain?: attitudes and future plans of overseas Chinese students in the US. J. Chin. Overseas 8 (1), 83–104. https://doi.org/10.1163/179325412X634319.

Khan, A., Choudhury, N., Uddin, S., 2019. Few research fields play major role in interdisciplinary grant success. Scientometrics 119 (1), 237–246. https://doi.org/10.1007/s11192-019-03043-w.

Lancichinetti, A., Fortunato, S., 2009. Community detection algorithms: a comparative analysis. Phys. Rev. E 80 (5), 056117. https://doi.org/10.1103/PhysRevE.80.056117.

Manchein, C., Brugnago, E.L., da Silva, R.M., Mendes, C.F., Beims, M.W., 2020. Strong correlations between power-law growth of COVID-19 in four continents and the inefficiency of soft quarantine strategies Chaos. Interdiscipl. J. Nonlinear Sci. 30 (4), 041102 https://doi.org/10.1063/5.0009454.

McCarty, C., Jawitz, J.W., Hopkins, A., Goldman, A., 2013. Predicting author h-index using characteristics of the co-author network. Scientometrics 96 (2), 467–483. https://doi.org/10.1007/s11192-012-0933-0.

Merrin, J., 2020. Differences in power law growth over time and indicators of COVID-19 pandemic progression worldwide. Phys. Biol. 17 (6), 065005 https://doi.org/10.1101/2020.03.31.20048827.

Morel, C.M., Serruya, S.J., Penna, G.O., Guimarães, R., 2009. Co-authorship network analysis: a powerful tool for strategic planning of research, development and capacity building programs on neglected diseases. PLoS Neglected Trop. Dis. 3 (8), e501. https://doi.org/10.1371/journal.pntd.0000501.

Nagtegaal, L.W., de Bruin, R.E., 1994. The French connection and other neo-colonial patterns in the global network of science. Res. Eval. 4 (2), 119–127. https://doi.org/10.1093/rev/4.2.119.

Narin, F., Stevens, K., Whitlow, E.S., 1991. Scientific co-operation in Europe and the citation of multinationally authored papers. Scientometrics 21 (3), 313–323. https://doi.org/10.1007/bf02093973.

Newman, M.E.J., 2001. The structure of scientific collaboration networks. Proc. Natl. Acad. Sci. Unit. States Am. 98 (2), 404–409. https://doi.org/10.1073/pnas.98.2.404.

Newman, M.E.J., 2004. Coauthorship networks and patterns of scientific collaboration. Proc. Natl. Acad. Sci. Unit. States Am. 101, 5200–5205. https://doi.org/10.1073/pnas.0307545100.

Park, J., Park, J., Choi, Y., 2022. The driving forces for collaboration networks of environmental engineering in Korea. Environ. Eng. Res. 27 (2), 200475. https://doi.org/10.4491/eer.2020.475.

Porter, M.A., Onnela, J.P., Mucha, P.J., 2009. Communities in networks. Notices of the AMS 56 (9), 1082–1097.

Reible, D.D., 1999. Fundamentals of Environmental Engineering. Lewis Publishers, Florida, p. 526.

Rueda, G., Gerdsri, P., Kocaoglu, D.F., 2007. In: Bibliometrics and Social Network Analysis of the Nanotechnology Field, Portland International Conference on Management of Engineering & Technology, pp. 2905–2911. https://doi.org/10.1109/PICMET.2007.4349633.

Settle, Q., Teixeira, K.L., Bechtold, R., Shoulders, C.W., Kings, A.E.H., 2020. Coauthor network analysis of journal of agricultural education articles from 2008-2017. J. Agric. Educ. 61 (2), 115–127. https://doi.org/10.5032/jae.2020.02115.

Tomassini, M., Luthi, L., 2007. Empirical analysis of the evolution of a scientific collaboration network. Phys. Stat. Mech. Appl. 385 (2), 750–764. https://doi.org/10.1016/j.physa.2007.07.028.

Traag, V.A., Waltman, L., van Eck, N.J., 2019. From Louvain to Leiden: guaranteeing well-connected communities. Sci. Rep. 9 (1), 1–12. https://doi.org/10.1038/s41598-019-41695-z.

Velden, T., Haque, A.U., Lagoze, C., 2010. A new approach to analyzing patterns of collaboration in co-authorship networks: mesoscopic analysis and interpretation. Scientometrics 85 (1), 219–242. https://doi.org/10.1007/s11192-010-0224-6.

Yang, Z., Algesheimer, R., Tessone, C.J.A., 2016. Comparative analysis of community detection algorithms on artificial networks. Sci. Rep. 6, 30750. https://doi.org/10.1038/srep30750.

Zang, C., Cui, P., Faloutsos, C., 2016. In: Beyond Sigmoids: the Nettide Model for Social Network Growth, and its Applications, Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, pp. 2015–2024. https://doi.org/10.1145/2939672.2939825.

Zhou, P., Leydesdorff, L., 2006. The emergence of China as a leading nation in science. Res. Pol. 35 (1), 83–104. https://doi.org/10.1016/j.respol.2005.08.006.

Zhu, J.J., Dressel, W., Pacion, K., Ren, Z.J., 2021. ES&T in the 21st century: a data-driven analysis of research topics, interconnections, and trends in the past 20 years. Environ. Sci. Technol. 55 (6), 3453–3464. https://doi.org/10.1021/acs.est.0c07551.

Zhu, K., Li, W., Fu, X., 2013. Modeling population growth in online social networks. Complex Adaptive Sys. Model. 1 (1), 1–16. https://doi.org/10.1186/2194-3206-1-14.