Environ. Eng. Res. 2022; 27(2): 200475
https://doi.org/10.4491/eer.2020.475

pISSN 1226-1025
eISSN 2005-968X

Research

The driving force for collaboration networks in environmental engineering in South Korea

Jaebeom Park[1], Jeryang Park[2], Yongju Choi[1†]

[1]*Department of Civil and Environmental Engineering, Seoul National University, 1 Gwanak-ro, Gwanak-gu, Seoul, Republic of Korea*
[2]*Department of Civil and Environmental Engineering, Hongik University, 94 Wausan-ro, Mapo-gu, Seoul, Republic of Korea*

ABSTRACT

Here, for the first time, coauthor network and cluster analysis were utilized in the environmental engineering field to identify the driving force for scientific collaboration among individuals and the formation of clusters. Papers published in South Korean domestic environmental engineering journals from 2004 to 2018 were assessed, which enabled identification of unique network characteristics that represent not only the field of study, but also the regional boundaries of the data source. Despite being limited to a single country, the study identifies network characteristics, such as scale invariance, that are typically found in other coauthor networks. Nine clusters were identified, the identity of which could be defined by two variables: research interests and author affiliations. The clusters were divided by the sameness or geographical proximity of author affiliations and problem-oriented research topics. These also describe the inter-cluster relationships, validating the notion that the two variables are the major driving force for collaboration networks. This study substantially advances the understanding of scientific collaboration in the environmental engineering field and can guide future studies, such as the role of coauthor networks in environmental engineering within or outside of regional boundaries and the role of networks in domestic publications in other fields of study

Keywords: Cluster analysis, Coauthor network, Collaboration pattern, Complex network analysis, Environmental engineering

1. Introduction

Social network analysis (SNA) is widely used to investigate the characteristics of various social structures as well as their emerging network dynamics [1-3]. This approach simplifies a complicated real-world structure using nodes (entities) and edges (relationship among entities), which allows the structure to be represented mathematically to assess its network properties. For instance, using SNA, the current network properties may be evaluated using quantifiable indicators, the hidden network properties of a real structure may be explored, or the evolution patterns of the network structure may be investigated, thereby enabling future predictions [1-5].

A coauthor network is a type of social network that is identifiable in the domain of scientific research. Previous works have shown that coauthor network analysis can be employed to define the subdivisions of a research field [6-8], identify patterns of collaboration among researchers [9, 10], and develop strategic plans for future research [11]. The coauthor network assigns individual authors as nodes and the presence of a coauthorship as an edge, which offers an explicit means of interpreting the collaboration pattern among scientists and/or groups of scientists because coauthoring a publication typically represents a direct scientific collaboration among individuals [12]. Coauthor network analysis has been employed to assess the characteristics of scientific collaborations in several academic disciplines [13], including ecology [7], climate engineering [8], and epidemiology [11].

The field of environmental engineering emerged as a separate academic discipline in the middle of the 20th century and has rapidly grown and evolved ever since. Considering the growing need for engineering solutions to address current and future environmental problems (e.g., emerging contaminants) and improve the environmental sustainability of man-made and natural systems, the field of environmental engineering is expected to continue to grow and evolve. One can classify the scientific research in the domain of environmental engineering with respect to the environmental medium (e.g., water, soil, atmosphere, and solid waste), the fundamental science involved (e.g., physical, chemical, and biological), or other factors. However, because environmental pollu-

This is an Open Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (http://creativecommons.org/licenses/by-nc/3.0/) which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.

Copyright © 2022 Korean Society of Environmental Engineers

Received August 20, 2020 **Accepted** January 27, 2021

† Corresponding author
E-mail: ychoi81@snu.ac.kr
Tel: +82 2 880 7376 Fax: +82 2 873 2684
ORCID: 0000-0002-6187-1446

tion may involve multiple types of environmental media, and combinations of multiple approaches (e.g., physical, chemical, and biological) are required to solve most environmental problems, the boundaries between each subdivision of environmental engineering are often not clear. In addition, the boundaries of the field of environmental engineering are becoming more difficult to define because of the growing need to exchange knowledge between various academic disciplines. Among academic fields, environmental engineering is an area with particularly active collaborations, both internally (i.e., within its boundaries, if they can be defined) and externally.

Considering the evolutionary and interdisciplinary nature of environmental engineering, scientists and engineers in this field may benefit greatly from understanding past and current patterns of scientific collaboration, as doing so will aid them in developing strategic plans for future directions. The current study presents, for the first time, the results of coauthor network analysis in the environmental engineering field. The scope of this pioneering study is limited to articles published in South Korean domestic journals ("South Korea" is henceforth referred to as "Korea" for simplicity) for two principal reasons: i) As members of the Korean environmental engineering community, we are interested in identifying unique regional characteristics that contribute to scientific collaboration among our colleagues and ii) Limiting the scope in this manner ensures that the dataset is a reasonable volume, without sacrificing comprehensiveness. The current work identifies notable characteristics of scientific collaborations in environmental engineering in Korea, including some that are commonly recognized among the community members and others that are not. In addition, this work will serve as a model that can be extended to other coauthor networks in the field of environmental engineering, including the global network of coauthors of publications in international journals, which will be the topic of a follow-up paper by the current authors.

2. Methods

2.1. Scope, Data Acquisition, and Data Processing

The target social structure of the current study is defined as the environmental engineering community in Korea. The products of scientific collaborations among members of the community are limited to literature published in 13 major Korean journals dedicated to environmental engineering-related topics (listed in Table S1). These 13 Korean journals account for the majority of scientific publications in environmental engineering in Korea, as detailed below. However, a preliminary analysis of the full publication lists of several target authors suggests that publications in international journals significantly contributed to the overall number of scientific collaborations among the members. Therefore, the scientific collaborations assessed in this study should be interpreted with caution because, for the most part, the boundaries of collaboration are limited to Korea's intellectual borders. Nonetheless, this study also allows us to track how such collaborations are formed within the country.

Journal selection was based on a report issued by the Korean Citation Index (KCI), which is the most commonly used platform for indexing of scientific journals in Korea. Each year, the KCI announces a list of domestic journals for indexing and a candidate list of domestic journals for future indexing, along with the subject category of each journal. Under the "Environmental Engineering" category, 13 journals are listed or listed-as-candidate in the year 2018 [14], all of which were selected for the current study to ensure comprehensiveness (see Table S1). Raw data, including author names and affiliations, journal name, paper title, and the keywords of each paper, were collected for all scientific papers published from January 2004 to August 2018 (~15-year period); these data were available from the KCI [14]. A total of 10,418 publications were collected, with 17,138 individual authors. Note that, in essence, these authors were recognized as the members of the Korean environmental engineering community.

Raw data were processed before application to the network analysis. In a coauthor network, authors (i.e., nodes) are connected to each other (i.e., form edges) when they coauthor a publication. Therefore, a member who has only single-authored publications exists in isolation in the network (see Fig. S1(d)). A member with only one publication does not contribute to the formation of relationships with other members; in other words, he or she does not have the function of providing a new connection between other members (see Fig. S1(b)) [10]. Therefore, members with only single-authored publications and those with only one publication were screened out from the raw data. This resulted in a reduction of the number of members from 17,138 to 5,723. Afterwards, the author information data (name and affiliation) were manually inspected to correct errors that resulted in the misidentification of a single author as multiple authors (e.g., typo in the institution name, failure to comply with the data format provided by the KCI). This correction further reduced the number of members to 5,440. Preliminary inspection of the coauthor network formed by the 5,440 members revealed that 5,076 constituted one giant component while the rest formed very small, isolated groups of 2 to 12 members. A giant component is the component of a network that contains most of the connected members. A giant component is formed when collaborations among the members of a scientific community are sufficient such that a scientific collaboration network analysis is feasible [13]. In this study, the network analysis was conducted on the giant component of 5,076 members. Fig. S1 exemplifies the components of the coauthor network.

2.2. Network Properties

Network properties, including the number of nodes and edges, average degree and average weighted degree, density, clustering coefficient, average path length, and diameter, were analyzed using the open source Gephi software [15]. Appendix S1 summarizes the definitions of each of the network properties.

2.3. Clustering and Cluster Analyses

Clustering is the task of grouping the members (i.e., nodes) of a network by degree of similarity [16, 17]. In this study, clustering was employed to identify sub-components (i.e., clusters) of the giant component in which internal collaborations took place more intensively than did external collaborations. Modularity was used as a key metric for clustering because it quantifies the intensity

of internal connection versus external connection for a module of the network. After testing various network clustering algorithms available in the literature, an algorithm proposed by Sales-Pardo [18] was selected, which allowed the identification of appropriate sizes of major clusters to analyze the gross characteristics of the study network. The clusters identified by the other clustering algorithms were generally much smaller than those identified by the selected algorithm; the smaller sizes represented highly localized collaborations among authors, which was not the focus of this study.

The selected clustering algorithm divided the giant component into 43 clusters with the largest cluster containing 901 nodes. Clusters with a size equal to or larger than 2% of the giant component were used for further analysis. Smaller clusters were not likely to significantly contribute to the whole network. These minor clusters had at most 50 nodes (i.e., authors). Consequently, a total of nine major clusters were identified, which were labeled from A to I in size order. Out of 5,076 nodes of the giant component, 4,557 (89.8%) belonged to the nine-cluster set.

The inter-relationships of the nine major clusters were analyzed. There are no established indices to describe inter-relationships among clusters. Thus, a cluster inter-correlation factor, K_{ij}, was developed, which is simple to understand intuitively. K_{ij} is defined as:

$$K_{ij} = \frac{e_{ij}}{n_i \times n_j} \quad (1)$$

where e_{ij} is the total number of edges that connects a node in cluster i and a node in j, and n_i is the number of nodes in cluster i. The denominator on the right hand side of Eq. (1) represents the number of all possible edges between a node in cluster i and a node in j. Accordingly, the definition of the cluster inter-correlation factor K_{ij} is analogous to that of the density of a network (see Appendix S1).

3. Results and Discussion

3.1. Temporal Trends in Environmental Engineering Research Published in Korean Journals

To identify the temporal trends in environmental engineering research published in Korean journals, all 10,418 publications collected from 13 domestic journals from Jan 2004 – Aug 2018 were included in the analysis. An average of 719 papers (results from the year 2018 were normalized to a 1-year duration) were published annually within the collection period. Analysis of the annual average number of publications between 2004 – 2008, 2009 – 2013, and 2014 – 2018 demonstrated an increase from the first to the second time window (p = 0.011; Student's t-test) and no significant difference between the second and the third (p = 0.90). However, the increase from 2004 – 2008 to 2009 – 2013 was not remarkable (12%), indicating that the annual number of publications was fairly steady throughout the study period (Fig. 1). This result is not surprising because the data were collected from a fixed number of journals for the entire collection period. None of the 13 study journals experienced significant expansion (e.g., increased number of issues per year, increased number of publications per issue) or contraction during the collection period.

Keyword appearance frequency was analyzed for 5-year moving windows during 2004 – 2018 to observe temporal changes in popular topics in Korean journals (Table S2). Keyword appearance frequency (%) was calculated as the number of times that a given keyword appears relative to the total number of keyword appearances in a given time window. In Table S2, keywords are arranged in order of frequency during each time window. 'Recycling,' 'water quality,' 'adsorption,' and 'heavy metal' were among the six most frequent keywords in all 5-year time windows from 2004 – 2018, suggesting a consistent level of interest in these topics among the Korean environmental engineering community.

For some keywords, the frequency of appearance evidently changed over time, representing a temporal shift in the popularity of the corresponding research topics (see Fig. 2). For example, the frequency of 'anaerobic digestion' showed a consistent decrease during the study period. The high frequency of this keyword in the early 2000s may represent the urgent need at the time to advance knowledge of anaerobic processes to treat and valorize organic wastes. Scientific interest in anaerobic processes may have decreased thereafter due to maturation of the knowledge necessary for processing of high-strength organic waste [19]. The keyword frequency of 'climate change' peaked between the late 2000s and the early 2010s. Because climate change remains a long-term global issue and involves a wide range of fundamental scientific knowledge

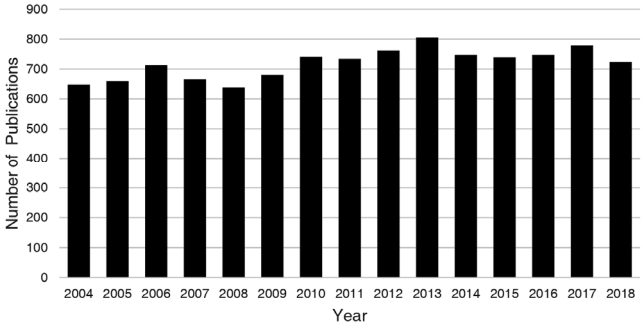


Fig. 1. Number of publications annually published in the selected Korean domestic journals during 2004-2018; the data from the year 2018 was normalized to a 1-year duration.

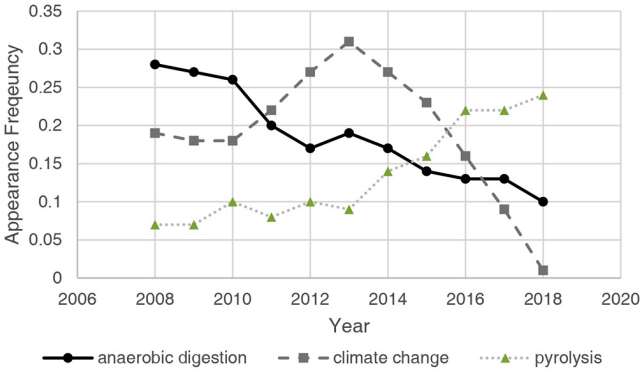


Fig. 2. Temporal trends in the frequency of appearance of three exemplary keywords: anaerobic digestion, climate change, and pyrolysis.

and technical approaches, the temporal change in the popularity of this topic cannot be explained by the maturation of relevant technologies. Government policy is presumed to be a main driver of the temporal change in the popularity of the keyword 'climate change.' The Korean administration during the late 2000s and the early 2010s declared the transition to a low-carbon society as one of its major goals. In 2010, the Framework Act on Low Carbon, Green Growth was enacted, which included provisions that support research and development of low carbon technologies [20]. 'Pyrolysis' seldom appeared as a keyword in papers published prior to the late 2000s but was highly ranked in recent years. The temporal increase in the popularity of the keyword 'pyrolysis' might represent the recent interest in the production and environmental engineering applications of biochar, which may improve the sustainability of organic waste management [21]. The appearance frequency analysis of the keywords mentioned above exemplifies the notion that the study dataset represented research interests in the Korean environmental engineering community, which change over time in response to changes in the research environment.

3.2. Network Properties

Table 1 summarizes the properties of the study network. A high clustering coefficient of a scientific collaboration network generally indicates active collaboration and close relationships among the members of the network [2].

The study network exhibited characteristics that are typically found in natural or social networks developed via self-organization. First, the study network had traits of a "small-world" network, a type of network that exhibits relatively short characteristic path lengths, as seen in random networks, but which shows a high degree of local clustering of nodes, which differentiates it from random networks [22]. To identify these characteristics, a random network was created that had the same number of nodes and edges as the study network (i.e., 5,076 nodes and 20,251 edges). While the average path lengths (L) of the study network and the random network were comparable ($L_{study} = 5.625$ and $L_{random} = 3.391$), the clustering coefficient (C) of the study network ($C_{real} = 0.667$) was more than two orders of magnitude greater than that of the random network ($C_{random} = 0.003$). Second, the study network showed scale-free characteristics in its degree distribution, i.e., the probability (P_k) that a node in the network interacts with k other nodes followed the power law ($P_k \sim k^{-\gamma}$, where γ is a power law exponent). A network exhibiting the scale-free property has nodes with extraordinarily high connectivity, namely, hubs [23]. Fig. 3 shows the power function fitted to the degree distribution of the study network with hubs. These two emergent characteristics (i.e., "small world" structure and scale-free property) are often found in various natural or social networks that are mature in terms of growth and self-organization [22, 23].

In coauthor networks, hubs are regarded as "academic stars" – individuals who have gained distinctive popularity as research

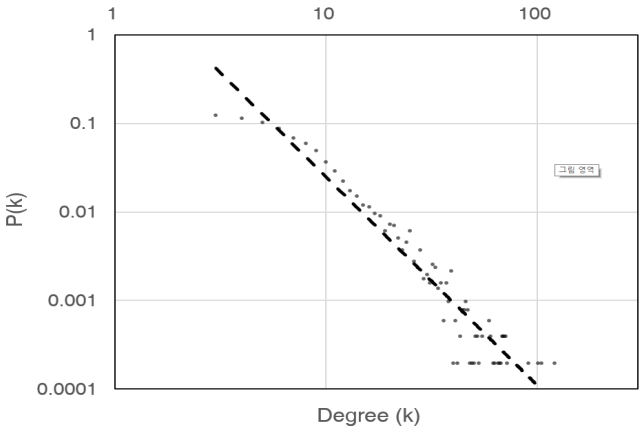


Fig. 3. A power law applies to the degree distribution of domestic environmental engineering publications in Korea. The regression model (dashed line) gives $P_k \sim k^{-2.35}$ with a correlation coefficient (R^2) of 0.9252. The exponent ($\gamma = 2.35$) is within the range (2.1 to 4) reported in the literature for other social networks [13].

collaborators among the members of a scientific community. Previous studies in various fields have shown that these "academic stars" are quite often found in international scientific collaboration networks [7, 8, 10, 24]. However, in particular cases, such as the solar radiation management and air capture subdivisions of climate engineering, coauthor networks have been shown to form from a collection of small, autonomous research groups that seldom collaborate with one another [8]. In these cases, nodes with extraordinarily high connectivity, i.e., hubs, may not be present. Only a few prior studies demonstrated the existence of hubs or "academic stars" in a coauthor network constructed using a domestic bibliography [25].

3.3. Identification of Clusters and Their Characteristics

The nine major clusters identified by the cluster analysis each consisted of 189 – 901 nodes (i.e., authors), accounting for 3.72 – 17.75% of the total node count of 5,076 (Table 2). To characterize each cluster, keyword and affiliation were used as the two major variables. The working hypothesis held that the proximity among the network members depends primarily on the presence of common research interests (as represented by keywords) and affiliate institutions. Keyword appearance frequency was calculated in each cluster following the definition given previously, and Table S3 shows the result. Table S4 shows the number of authors affiliated with each institution in each cluster. These data verify that each of the nine major clusters can be distinguished from the others using keywords and affiliation. The identity of each cluster is discussed in the following section.

Cluster A, the largest of the nine clusters, had 'recycling,' 'adsorption,' 'solvent extraction,' and 'recovery' as the most frequent

Table 1. Properties of the Study Network

Number of nodes	Number of edges	Average degree	Density	Clustering coefficient	Average path length	Diameter
5,076	20,251	7.979	0.002	0.667	5.625	17

Table 2. Number of Authors in Each Cluster

Cluster ID	A	B	C	D	E	F	G	H	I
Number of authors (nodes)	901	861	610	569	514	499	224	190	189
Proportion in giant component (%)	17.75	16.96	12.02	11.21	10.13	9.83	4.41	3.74	3.72

keywords. Inorganic chemicals, including 'copper,' 'nickel,' 'heavy metal(s),' and 'cobalt,' also appeared frequently as keywords in Cluster A. Waste management technologies dealing with inorganic constituents were the most popular area of research in this cluster, corresponding very well to the focus of the Korea Institute of Geoscience and Mineral Resources (KIGAM), the most common affiliate institution of authors in Cluster A.

Cluster B was another large cluster with waste management as the most popular research topic. 'Heavy metal' and 'recycling' were the two most frequent keywords in Cluster B. One characteristic distinguishing this cluster from Cluster A is that keywords representing waste management-related environmental media, such as 'sediment,' 'groundwater,' and 'landfill' were also at the top of the list. In addition, Cluster B showed broader interest in the technical approaches to waste management, as evidenced by the presence of 'adsorption,' 'anaerobic digestion,' 'stabilization,' 'soil washing,' and 'coagulation' in the major keyword list (Table S3). In Cluster A, physicochemical techniques such as 'adsorption' and 'solvent extraction' were much more popular than biological techniques. In terms of author affiliations, the Korea Institute of Civil Engineering and Building Technology (KICT), not KIGAM, was the most common institution in Cluster B.

Cluster C had 'VOCs,' 'particulate matter' (including 'PM10' and 'PM2.5'), 'formaldehyde,' 'odor,' 'indoor air,' 'indoor air quality,' and 'elementary school' as major keywords, clearly indicating that indoor air quality was the common research topic in this cluster. The National Institute of Environmental Research (NIER) and Yonsei University were the two most common affiliate institutions in Cluster C, suggesting the pivotal role of these institutions in indoor air quality studies in Korea.

In *Cluster D*, the major keywords collectively suggested that eutrophication of water bodies was a major research interest. 'Water quality' was the most frequent keyword, followed by 'sediment,' 'soil erosion,' and 'runoff,' which are sources of macronutrients in natural water systems. Aquatic organisms related to algal bloom such as 'phytoplankton,' 'zooplankton,' and 'cyanobacteria' were also identified as major keywords, further substantiating the conclusion. Algal bloom in bodies of fresh water has been one of the most serious issues in Korea since the early 2010s [26]. The major institution for this cluster was NIER, which may represent its responsibility as a government-funded research institution that deals with contemporary environmental issues.

Cluster E could not be sufficiently defined by keyword analysis. Although most major keywords in Cluster E were related to atmospheric pollution (e.g., 'odor,' 'PM2.5,' 'PM10,' and 'VOCs'), keywords representing environmental media other than the atmosphere were also significant (e.g., 'groundwater' and 'soil'). Keywords related to both meteorological models (e.g., 'CALPUFF' and 'CALMET') and molecular biology techniques (e.g., 'microbial community' and 'DGGE') both appeared as major keywords. On the other hand, the identity of Cluster E was evidenced by author affiliations: NIER was by far the most common. It is likely that this cluster represents collaborative works among researchers at NIER that spans various environmental engineering-related subjects.

For *Cluster F*, most of the frequently appearing keywords were related to water quality and the aquatic ecosystem. In this regard, Clusters D and F shared a common research interest. However, Cluster F is likely to more broadly address water quality and the aquatic ecosystem and be less dedicated to eutrophication than Cluster D. The frequencies of keywords related to higher-level organisms (e.g., 'fish community' and 'fish') and lower-level organisms (e.g., 'phytoplankton,' 'zooplankton,' and 'microalgae') were comparable in Cluster F. This cluster deals with a wide range of contaminants including 'arsenic,' 'phosphorus,' 'nutrient,' and 'heavy metal.' Several keywords representing treatment techniques were also identified, such as 'adsorption' and 'oxidation.' Although relatively less significant than keywords related to water quality and the aquatic ecosystem, 'soil,' 'eutrophication,' and 'wastewater treatment' also appeared in the major keyword list, further demonstrating the wide breadth of research interests among authors in Cluster F. The pattern of affiliation of authors in this cluster is analogous to that of Cluster D. The major institutions related to this cluster were NIER, Chungbuk National University, and Kangwon National University.

Cluster G's major research interest was water quality management and engineering. The major keywords in this cluster related to typical water quality indicators (e.g., 'pH,' 'T-N,' and 'COD') and typical water treatment techniques (e.g., 'coagulation,' 'activated carbon,' and 'PAC'). The most interesting feature of this cluster is that most of the authors' affiliations were located in the southeast region of Korea. These institutions include Pusan National University, Busan Metropolitan Government, Donga University, and the National Institute of Fisheries Science (NIFS), which are all located in Busan, the largest city in southeastern Korea. This finding indicates that regional proximity contributes to collaborative research among environmental scientists in Korea.

Cluster H's research focus was waste management, including management of hazardous substances. Keywords related to the valorization of wastes (e.g., 'recycling,' 'energy recovery,' and 'anaerobic digestion') and toxic and/or refractory chemicals (e.g., 'heavy metal,' 'mercury,' 'PAHs,' and 'PCBs') were popular. In terms of research interests, Cluster H was similar to Cluster B. An obvious difference between the two clusters was found in the affiliation analysis. The two major institutions in Cluster H were NIER and Yonsei University, whereas KICT and Seoul National University were the two major institutions in Cluster B.

For *Cluster I*, keywords related to organic waste valorization, including 'food waste,' 'compost,' and 'recycling,' were ranked highly on the major keyword list. However, keywords related to physicochemical treatment techniques (e.g., 'UV,' 'UF' (ultrafiltration),

'AOPs,' and 'O_3') and those related to the assessment of environmental and ecological quality (e.g., 'environmental impact assessment,' 'ecosystem,' 'species distribution model,' and '*Cryptosporidium*') also appeared frequently. Overall, the research interests of Cluster I were relatively diverse, with organic waste valorization being a major topic. Compared to the other clusters, the authors' affiliations in Cluster I were much more frequently located in southwestern Korea; these included Jeonbuk National University, the National Institute of Agricultural Sciences (NIAS), and the National Institute of Ecology (NIE). The research interests of Cluster I coincided well with the foci of NIAS and NIE.

Overall, the in-depth analysis of the characteristics of each cluster verified that the identities of the clusters could reasonably be defined by two variables, keywords and author affiliations (Table 3). Research interest and affiliation are likely to be the two key factors that promote collaboration among environmental engineers in Korea. The keyword analysis implied that the common research interests for each cluster could be defined in terms of environmental concerns (e.g., waste management, indoor air pollution, and eutrophication of water bodies) rather than technical approaches (e.g., biological treatment, physical treatment, and advanced oxidation processes). A common technology-oriented topic that represents the research interests could not be identified for any of the nine clusters. Therefore, collaboration among Korean environmental engineers is generally more likely to be problem-oriented than technology-oriented. The affiliation analysis identified universities and government-funded, non-educational research institutions (henceforth referred to as government-funded institutions) as the two major types of institutions to which the authors belonged. Intra-institutional collaborations in government-funded institutions are expected to be driven by multiple factors, such as shared research interests and tasks, close personal relationships, and geographical proximity. Because of the relatively strong driving force for intra-institutional collaboration, government-funded institutions are often found to play a key role in cluster formation. Cluster E clearly exemplifies this phenomenon, as it shows a wide diversity of research topics while being dominated by a single government institution (e.g., NIER). In some cases, government-funded institutions and universities located in close proximity to one another set up inter-institutional collaborations to address local environmental problems. Cluster G represents this type of collaboration, where several institutions located in the city of Busan formed a cluster centered on water quality management. This city, or more broadly the entire Nakdong river basin, has long suffered from water pollution problems due to the high number of industrial complexes in the area.

3.4. Correlation among Clusters

In the cluster analysis, multiple clusters shared common research interests (see Table 3). The hypothesis was extended to say that the nine clusters identified in the study network would be clustered by the degree of similarity of their research topics as the sole determinant. To verify this hypothesis, the environmental engineering research topics were divided into three categories: waste management (WM), water quality (WQ), and atmospheric pollution (AP). Based on the keyword analysis described above, each cluster could be assigned to one of the categories without ambiguity. The correlation factors (see Table 4) between each pair of the nine clusters were examined next to identify if a close relationship could be found among the clusters that belonged to the same research category.

The following groups showed relatively high correlation factors between each cluster pair: A–B, C–E, D–F, and D–E–H. The first three groups contained clusters that belong to the same research category (WM, WQ, and AP, respectively). However, the members

Table 3. The Features of the Nine Large Clusters with Respect to Research Interest (Identified by Keyword Analysis) and Author Affiliations

Cluster ID	Research interest			Major author affiliations			
				Type of institution		Regional	
	WM	WQ	AP	NI	Univ.	SE	SW
A	O			O			
B	O			O			
C			O		O		
D		O		O			
E			O	O			
F		O			O		
G		O			O	O	
H	O			O			
I	O						O

For simplicity, research interests are classified into three categories: waste management (WM), water quality (WQ), and atmospheric pollution (AP). Also, the authors' affiliate institutions are classified into two categories: government-funded, non-educational research institution (NI) and university or other higher-level educational institutions (Univ). Those that exhibited a distinct regional feature were indicated with the following abbreviations: SE, southeast region of Korea; SW, southwest region of Korea. Note that clusters belonging to the same research category may be further distinguished using a more specific research subject and also that the clusters marked with the same type of institution may be further distinguished by the name of the major affiliate institution.

Table 4. Correlation Factors among Clusters

	A	B	C	D	E	F	G	H	I
A		39	16	10	11	8	22	26	12
B	39		14	28	20	22	32	29	7
C	16	14		16	**93**	6	16	25	16
D	10	28	16		**77**	**156**	20	**58**	40
E	11	20	**93**	**77**		41	24	**89**	26
F	8	22	6	**156**	41		19	39	48
G	22	32	16	20	24	19		12	**68**
H	26	29	25	**58**	**89**	39	12		33
I	12	7	16	40	26	48	**68**	33	

The actual values are multiplied by 10^5 for convenience. Higher values indicate stronger relationships between clusters. Those with correlation factors greater than 50×10^{-5} are indicated in bold.

of the last group (D–E–H) did not fall into the same research category; rather, the research categories for the three member clusters were all different. In addition, correlation factors between a pair of clusters that belonged to the same category were not necessarily greater in value than those between pairs in different categories. For example, correlation factors between A and I (both categorized as waste management) and D and G (both categorized as water quality) were lower than the gross average of 34×10^{-5}. These results suggest that, in the Korean environmental engineering community, research topic is not a sole determinant for identification of closely related cluster groups. For the D–E–H group, NIER was commonly identified as an affiliate institution. This suggests that collaborative works across different research categories took place in this group, with NIER as the key institution that bound the authors together. Therefore, both research interest and affiliation play significant roles at not only the intra-cluster but also the inter-cluster level in the environmental engineering field in Korea.

3.5. Characteristics of Cluster Groups with the Same Research Category

The three broad research topics used to categorize the clusters, i.e., waste management (WM), water quality (WQ), and atmospheric pollution (AP), may be designated as the three major academic sub-disciplines of environmental engineering in Korea. The network properties of every cluster group that belongs to the same sub-discipline were analyzed. Table 5 shows the results of this analysis. WM was the largest sub-discipline, accounting for 42.18% of the giant component in terms of the number of authors (i.e., nodes), followed by WQ (25.45%) and AP (22.14%). The average degree of WM (3.98) was smaller than those of WQ (4.94) and AP (4.62). This suggests that, in general, the scientists in the WM sub-discipline are less likely to conduct collaborative research than the scientists in the other two sub-disciplines. The coauthor networks of the three sub-disciplines all showed relatively high clustering coefficients (0.67), as was previously shown for the whole study network ($C = 0.667$). In other words, the highly clustered nature of the whole study network was evident for any of the three sub-disciplines defined by the topic of environmental concern.

The fact that WM accounted for the largest number of authors defied expectation. This result did not coincide with the number of professionals (which could be estimated by the number of members in Korean academic societies and the number of Korean graduate school faculty members specializing in each sub-discipline) and the amount of money invested in each sub-discipline in Korea. For example, in 2019, the government budget assigned to water management (including integrated water management, water environment, and water resources) was more than 10-fold greater than that assigned to waste management (which was officially termed “resource circulation”) [27]. This disagreement is probably due to the greater preference of Korean scientists specializing in WM to publish their work in domestic journals than those specializing in WQ or AP. For a simple but clear demonstration of this statement, additional data collection and analyses were conducted as follows. From the Web of Science [28], all articles published in journals under the category “engineering: environmental” in the period from Jan 2014 – Aug 2018 that had been authored by individuals affiliated with Korean institutions were collected. This search identified 2,452 papers published in international journals. The keyword appearance frequencies of the top 30 most frequent keywords in papers published in environmental engineering-related domestic journals during the same period were calculated (Table S2, last column). Table S5 shows that the appearance frequency values for domestic and international journals differed substantially for the same keyword. One notable difference is that the keywords that are highly relevant to WM, such as recycling, food waste, and leachate, show much higher frequencies in domestic journals than international journals (2.17% versus 0.29% for recycling, 0.98% versus 0.29% for food waste, and 0.83% versus 0.12% for leachate, respectively). The tendency of Korean scientists specializing in WM to prefer publishing their works in domestic journals may reflect the highly regionalized nature of waste management practices due to their large dependence on land and resource availability and government policy, among other factors.

Table 5. Properties of Each Cluster Group

	Waste management	Water quality	Atmospheric pollution
Number of authors (nodes)	2,141	1,292	1,124
Proportion in giant component (%)	42.18	25.45	22.14
Average degree	3.98	4.94	4.62
Density	0.00372	0.00764	0.00824
Clustering coefficient	0.691	0.671	0.715

4. Conclusions

This study analyzed the coauthor network in the Korean environmental engineering field by focusing on publications in domestic journals. The coauthor network exhibited a "small-world" characteristic and its degree distribution followed a power law, indicating the growth and evolution of the coauthor network through self-organization with the presence of hubs. Investigation of the clusters revealed that two major factors played key roles in collaboration among environmental engineers in Korea: research interest and affiliation. More specifically, collaboration was principally driven by the correspondence between problem-oriented research interests and the sameness or geographic closeness of the institution(s) with which the authors were affiliated. The two factors, research interest and affiliation, were highly influential not only for individual-level relationships, i.e., the formation of clusters composed of nodes, but also for cluster-level relationships, i.e., correlation among the clusters. Government-funded, non-educational research institutions played a significant role in both individual- and cluster-level relationships.

The pattern of collaboration among environmental engineers and the driving forces identified in the current analysis are assumed to represent unique characteristics of domestic publications. Domestic journals have a relative tendency to publish works that are regionally specific, whereas international journals typically prefer papers dealing with topics of broad interest. Therefore, it seems likely that international environmental engineering collaborations are more strongly driven by similarities in technology-oriented research interests than those in problem-oriented research interests. A follow-up study analyzing a coauthor network in papers published in international environmental engineering journals will address this speculation.

Acknowledgements

This paper relied on KCI DB data supplied by the National Research Foundation of Korea(NRF) and Web of Science(WOS) database supplied by the Clarivate Analytics. This research was supported by the BK21 Four research program of the NRF. Jeryang Park was supported by the Basic Science Research Program through the NRF funded by the Ministry of Science and ICT (NRF-2019R1C1C 1008017). Yongju Choi would like to thank the Institute of Engineering Research at Seoul National University for technical assistance. The authors acknowledge the contribution of Dr. Seokheon Lee at the Korea Institute of Science and Technology for providing insights and taking part in interesting discussions.

Author Contributions

J.P. (Undergraduate Student) conducted data collection and analysis, contributed to the study design, and wrote the paper. J.P. (Associate Professor) contributed to data analysis and reviewed the manuscript. Y.C. (Associate Professor) conceptualized the idea, scrutinized the data and research, and revised the manuscript.

References

1. Newman MEJ. Networks. 2nd edition. Oxford: University Press; 2018.
2. Newman MEJ. The structure and function of complex networks. *SIAM review* 2003;45:167-256.
3. Newman MEJ, Barabási ALE, Watts DJ. The structure and dynamics of networks. Princeton: University Press; 2006.
4. Daud A, Ahmad M, Malik MSI, Che D. Using machine learning techniques for rising star prediction in co-author network. *Scientometrics* 2015;102(2):1687-1711.
5. Ductor L, Fafchamps M, Goyal S, Van der Leij MJ. Social networks and research output. *Rev. Econ. Stat.* 2014;96(5):936-948.
6. Zhao, X. A scientometric review of global BIM research: Analysis and visualization. *Automat. Constr.* 2017;80:37-47.
7. Borrett SR, Sheble L, Moody J, Anway EC. Bibliometric review of ecological network analysis: 2010–2016. *Ecol. Model.* 2018;382:63-82.
8. Belter CW, Seidel DJ. A bibliometric analysis of climate engineering research. *Wiley Interdisciplinary Reviews: Climate Change* 2013;4:417-427.
9. Lee SS. A analytical study on the properties of coauthorship network based on the co-author frequency. *J. Korean Library Info. Sci. Soc.* 2011;42:105-125.
10. Velden T, Haque AU, Lagoze CJS. A new approach to analyzing patterns of collaboration in co-authorship networks: mesoscopic analysis and interpretation. *Scientometrics* 2010;85(1):219-242.
11. Morel CM, Serruya SJ, Penna GO, Guimarães RJP. Co-authorship network analysis: a powerful tool for strategic planning of research, development and capacity building programs on neglected diseases. *PLoS Negl. Trop. Dis.* 2009;3(8).
12. Glänzel W, Schubert A. Analysing scientific networks through co-authorship. In Handbook of quantitative science and technology research. Springer: Dordrecht; 2004. p. 257-276.
13. Newman MEJ. The structure of scientific collaboration networks. *Proc. Natl. Acad. Sci. USA.* 2001;98:404-409.
14. Korean Citation Index. Domestic bibliographical data in environmental engineering [Internet]. c2018. Available from: https://www.kci.go.kr/kciportal/main.kci
15. Bastian M, Heymann S, Jacomy M. Gephi: an open source software for exploring and manipulating networks. In: 3rd international AAAI conference on weblogs and social media(ICWSM 2009); 17-20 May 2009; San Jose, California.
16. Emmons S, Kobourov S, Gallant M, Borner K. Analysis of Network Clustering Algorithms and Cluster Quality Metrics at Scale. *PLoS One* 2016;11:e0159161.
17. Girvan M and Newman MEJ. Community structure in social and biological networks. *Proc. Natl. Acad. Sci. USA.* 2002;99: 7821-7826.
18. Sales-Pardo M, Guimera R, Moreira AA, Amaral LA. Extracting the hierarchical organization of complex systems. *Proc. Natl. Acad. Sci. USA.* 2007;104:15224-15229.
19. Mao C, Feng Y, Wang X, Ren G. Review on research achievements of biogas from anaerobic digestion. *Renew. Sustain. Energy Rev.* 2015;45:540-555.
20. KOR. Framework Act on Low Carbon, Green Growth [Internet]. c2018. [cited 1 August 2020]. Available from: http://ex-

twprlegs1.fao.org/docs/pdf/kor100522.pdf
21. Cha JS, Park SH, Jung SC, et al. Production and utilization of biochar: A review. *J. Ind. Eng. Chem.* 2016;40:1-15.
22. Watts DJ, Strogatz SH. Collective dynamics of 'small-world' networks. *Nature* 1998;393:440-442.
23. Barabasi AL, Albert R. Emergence of scaling in random networks. *Science* 1999;286:509-512.
24. Ebadi A, Schiffauerova A. How to become an important player in scientific collaboration networks? *J. Informetr.* 2015;*9*(4): 809-825.
25. Leifeld P, Wankmuller S, Berger VT, Ingold K, Steiner C. Collaboration patterns in the German political science co-authorship network. *PLoS One* 2017;12:e0174671.
26. Park SB. Algal blooms hit South Korean rivers. *Nature* 2012;488:427-438.
27. Ministry of Environment, KOR. Budget Plan of Ministry of Environment [Internet]. c2020. [cited 1 August 2020]. Available from: http://www.me.go.kr/home/web/index.do?menuId=10127
28. Web of Science. Bibliographical data in environmental engineering [Internet]. c2020. Available from: www.webofknowledge.com