Drawing a Knowledge Map of Smart City Knowledge in Academia

Feng-Tyan Lin

1239 Siping Road, Shanghai, P.R. China,
<ftlin@tongji.edu.cn>

Dr. Feng-Tyan Lin obtained his bachelor's degree in engineering (urban planning) at National Cheng Kung University, Tainan, Taiwan, in 1977. He obtained his master's and PhD degrees in computer science at Northwestern University, USA, in 1986 and 1989, respectively. Professor Lin's research has focused on information cities and computer theories which can be applied to urban planning and architectural design. He was a professor at National Taiwan University and National Cheng Kung University, Taiwan, and retired in 2015. He currently serves as an overseas co-PI in Tongji University, China.

Lin, Fen-Tyan. 2019. "Drawing a Knowledge Map of Smart City Knowledge in Academia." *Knowledge Organization* 46(6): 419-438. 27 references. DOI:10.5771/0943-7444-2019-6-419.

Abstract: This research takes the academic articles in the Web of Science's core collection database as a corpus to draw a series of knowledge maps, to explore the relationships, connectivity, distribution, and evolution among their keywords with respect to smart cities in the last decade. Beyond just drawing a text cloud or measuring their sizes, we further explore their texture by identifying the hottest keywords in academic articles, construct links between and among them that share common keywords, identify islands, rocks, reefs that are formed by connected articles—a metaphor inspired by Ong et al. (2005)—and analyze trends in their evolution. We found the following phenomena: 1) "Internet of Things" is the most frequently mentioned keyword in recent research articles; 2) the numbers of islands and reefs are increasing; 3) the evolutions of the numbers of weighted links have fractal-like structure; and, 4) the coverage of the largest rock, formed by articles that share a common keyword, in the largest island is converging into around 10% to 20%. These phenomena imply that a common interest in the technology of smart cities has been emerging among researchers. However, the administrative, social, economic, and cultural issues need more attention in academia in the future.

Received: 7 May 2019; Revised: 18 August 2019; Accepted: 23 August 2019

Keywords: articles, keywords, knowledge maps, smart cities

1.0 Motivation

In the era of the information explosion, processing information into knowledge for better management and decision making has become necessary. Many ongoing efforts explore the issues of knowledge in various fields, such as library and information science (LIS), business administration, industrial production, public health, public policy, and smart cities. Being roused by this wave, many research interests have also emerged in knowledge management (KM) as a collectively scientific discipline.

The study of KM has three interrelated aspects: methodology, ontology, and sociology. Methodologies of KM include codification, classification, tag clouds, knowledge map construction, visualization, text mining, and topological analysis. Some will be discussed further and employed later in this paper. In addition to methodology, the ontology of KM consists of organized knowledge of specific knowledge domains explored by various methods, while the sociology of KM, combining with epistemology and axiology, concerns social, cultural, organizational, and political factors associated with successful implementation of knowledge management.

Although the scope of KM ontology has been increasing, it is far from complete. As Hjørland (2008) suggests, knowledge organization (KO) should not be limited to a narrow meaning restricted to document description, indexing, classification, and organization. Rather, KO has a broader meaning related to how knowledge is socially organized and how individual sciences are organized. He claims that KO in the narrow sense cannot develop a fruitful body of knowledge without considering KO in the broader perspective. The claim also holds true in the territory of KM.

Among the various booming subjects in KM, we found that there are relatively few articles discussing the knowledge management of smart cities. Being a prevailing topic, smart cities, including smart transportation, smart public health and safety, smart education, and smart governance, etc., has attracted interest from many cities, researchers, scholars, engineers, industries, and businesses. Most of those works focus on developing frameworks, strategies, innovative technologies and devices, and application systems for constructing smart cities. Far fewer articles study the KM of smart cities, which propose conceptual visions, suggest frameworks and models, and identify key factors for building up smart city knowledge bases (e.g., Boyer 2016; Biloslavo and Zornada 2004; Brachos et al. 2007). Only a handful of articles discuss the sociological aspects of KM in smart cities (e.g., Meijer and Bolívar 2016; Jennex and Zakharova 2006). As far as we know, no research exists on keyword distribution in academic research articles.

To remedy this, we will use the academic articles in the Web of Science core collection database as a testbed to explore the relationships, connectivity, distribution, and evolution among their keywords associated with smart cities, published in the last decade. The remainder of this paper is organized as follows: In section 2 we briefly review some related work; in section 3 we introduce the methods used in this research, including definitions, the analysis process, and the database used; and in section 4 we present the results. Finally, in section 5 we draw conclusions.

2.0 Related work

2.1 The character of knowledge

Knowledge has many properties, including dispersion, evolution, reusability, and guidance. Knowledge to accomplish a job is dispersed through various organizations and staff in different disciplines. Therefore, it needs cooperation among persons from different departments or even outside experts. Knowledge sharing and communication is a key factor for completing a task successfully (e.g., Liu et al. 2019, Ahmed et al. 2019). Knowledge also evolves. In other words, knowledge has a dynamic nature and cannot be static. It continually changes with human experiences, technology advancement, knowledge explication, researchers' perspectives, and social interactions (Mclnerney 2002). Moreover, many tasks are repeated with minor variations in different contexts. The knowledge from previous tasks can be reused and adapted for new cases. New employees can learn from the experiences of similar cases completed previously by other colleagues so that the task at hand can be carried out correctly, efficiently, and effectively. If knowledge of typical experiences is recorded in an understandable format for transferal to new staff, it will largely improve the work quality of an organization. The study of knowledge has also attracted researchers' attention so that the current situation and front edge of research can be identified for guiding future work. For example, Scharnhorst et al. (2016) captured how knowledge and knowledge systems of UDC changed over time and raised some further questions for future work.

The knowledge of smart cities has all the properties mentioned above. Articles bearing knowledge are published in academic journals, conference proceedings, and magazines. Keywords provided by authors emerge, are repeated and reused, change, and evolve. Thus, studying the knowledge concerning smart city keywords should reveal the corresponding phenomena of these properties through the construction of knowledge maps with capabilities of visualization and text analytics.

2.2 Visualization

Visualization, one of the most popular approaches, acts as a collaboration catalyst to capture the big picture of dispersed knowledge for sense making and knowledge sharing (Eppler 2013). While many word cloud visualization tools deal with individual words, Heimerl et al. (2014) took it a step further to develop a prototypical system, called Word Cloud Explore, that employs linguistic knowledge about the words and their relationships for text analysis, such as multiword expression identification, term statistics, co-occurrence highlighting, and provision of linguistic information.

Many processes for classifying raw text-based materials and interpreting the visualized result are still manual. For example, Toronto 311, a non-emergency service in a smart city, maintained an online knowledge base composed of 21,000 web pages. However, these web pages were unstructured texts, and thus, not machine-readable and difficult to reuse. To recognize the knowledge requirements of the city government, Allahyari et al. (2014) manually analyzed and identified ten knowledge patterns extracting from more than 500 Toronto 311 web pages according to their importance and frequency. In another example unrelated to smart cities, Scharnhorst et al. (2016) employed a color-coding scheme to visualize complex networks of category systems of Wikipedia and Universal Decimal Classification (UDC) so their differences could be compared using human eyes.

"Map" is a geographical term, which we borrow to describe objects in knowledge maps. Ong et al. (2005) mentioned that a knowledge map had an ocean-and-island metaphor, and the size of an island provided an estimate of the number of articles contained in a category. However, they did not explore this issue further. In this article, we take a similar analogy from geographers and geologists. Not only is the metaphor of islands referenced and their sizes are measured, but also the texture of islands is further explored by studying the evolution trend of the hot keywords, the strength of connectivity among articles, and the coverage of the biggest rock formed by completely interconnected articles over the islands.

2.3 Knowledge mining and mapping

Knowledge mining is a family of methods used to reveal the structure of knowledge embedded in a mass of unorganized documents by constructing relationships of the co-citation or co-occurrence of tags, which can be words appearing in titles, abstracts, keyword lists, or full texts.

Many knowledge-mining techniques exist. Medelyan (2018) illustrates five common approaches to disclose the internal structure of unorganized materials, namely, word spotting, rules for pattern matching, text categorization, topic modelling, and thematic analysis. Cheng et al. (2018) summarize knowledge-mining techniques into two categories: statistical analysis-oriented, such as k-means and k-nn (Chemchem and Drias 2015), and knowledge discovery-oriented, such as machine learning (Ong et al. 2005).

A knowledge map results from knowledge mapping. The layout of a knowledge map may be arranged in a sequential line-by-line form, a tree, a circle, or a complex network. Concepts or words of knowledge may be organized by alphabetical order, occurrence frequency, or sematic proximity in different fonts, sizes, weights, colors, and places for readers to easily capture the whole structure and perform tasks, such as searching, browsing, impression formation, recognition, and matching (Bateman et al. 2008; Gambette and Véronis 2010; Rivadeneira et al. 2007; Heimerl et al. 2014).

Knowledge mapping is an essential subfield of knowledge management and has been applied to many fields. It assists public and private organizations and academic and research communities understand the whole picture of scattered knowledge retained in different departments or places with the purpose of making strategic plans, transferring knowledge and learning experiences, inspiring brainstorming, and stimulating new knowledge. For example, by analyzing the number of papers downloaded from the arXiv in the "artificial intelligence" (AI) section through 18 November 2018, Hao (2019) classified the research history of AI into three major trends. To cope with rapid growth and ever-changing knowledge in the field of smart production, Cheng et al. (2018) discussed and suggested the application of knowledge mapping in production management, while Su and Jiang (2007) applied it to assisting fuel pump design. Su and Jiang (2007) suggested a product design task-oriented knowledge organizing method. Liu, et al. (2009) developed a virtual collaboration platform for enterprise knowledge construction by allowing members to tag their documents, and then asked a domain expert to draw a domain knowledge map based on tags collected from members' contributions.

The research on knowledge maps of smart cities is very rare and needs to be a dedicated topic. Balaid et al. (2016) systematically reviewed the development status of knowledge mapping. He concluded that the study of knowledge mapping was still in an early stage, and a large portion of existing research only covered very limited disciplines. In the field of smart cities, this observation is also true, where Mora, Deakin and Reid (2018) remains a singular work. They mapped a network structure of publications in the field of smart cities in the period from 1992 to 2012 by combining co-citation clustering and text-based analysis. They identified five major thematic tracks in the publications concerning smart cities, namely experimental, ubiquitous, corporate, single, and holistic. Their work is closely aligned with ours, in that we also are interested in drawing a knowledge map of smart city research, analyzing its evolution in the last decade and identifying hot topics. However, some major departures differentiate the two studies, including timespan, inclusion of tags, article selection criteria, research methods and findings. We will further compare their work with ours in Section 5.1.

3.0 Method

3.1 Definition

A knowledge map is an application of graph theory that studies the topology of nodes and links. Knowledge mapping has two kinds of nodes: articles and keywords. Thus, there are also two kinds of links: single and composite links. A single link is established between two articles based on a common keyword. A composite link is composed of one or more simple links between two articles. In other words, while many simple links may exist between two articles due to sharing many common keywords, there is at most one composite link between two articles.

Articles are loosely connected, like an island, if they are directly or indirectly connected by composite links. The size of an island is the number of articles of it. In an island, one may find rocks, where articles share a common keyword. At the same time, there may be a lot of singular articles without any common keywords with other articles. The singular articles are called reefs and are not considered islands. Formal definitions of these concepts are given as follows.

Let D be a set of articles $p_1, p_2, \ldots,$ denoted as $D = \{p_1, p_2, \ldots\}$, the frequency of a keyword k in D, $freq(k, D)$, is the number of occurrences of k in P. It is noted that there are no duplicated keywords in an article. In other words, $freq(k, D)$ also indicates the number of articles in D that share a common keyword k. Furthermore, let article p_i have keywords $K_i = \{k_{i1}, k_{i2}, \ldots\}$, and p_j have $K_j = \{k_{j1}, k_{j2}, \ldots\}$. If there is a keyword k, where $k \epsilon K_i$ and $k \epsilon K_j$, ie., p_i and p_j share a common keyword k, p_i and p_j are linked with respect to k, and a simple link $L(p_i, p_j, k)$ is established. A composite link $CL(p_i, p_j)$ between articles p_i and p_j is a composition of all the simple links between them. The weight of a composite link $CL(p_i, p_j)$ is denoted as $WL(p_i, p_j)$; that is the number of simple links between p_i and p_j. $CL(p_i, p_j)$ is also called the "strength" between p_i and p_j, since it is the number of common keywords of p_i and p_j. The stronger the strength $CL(p_i, p_j)$ is, the more the common keywords. Two articles p_i and p_j are "loosely connected" if: 1) there exists a simple link $L(p_i, p_j, k_w)$ directly connecting them; or, 2) there exists simple links $L(p_i, p_v, k_w)$, and $L(p_j, p_v, k_u)$, where p_i and p_j are indirectly connected via p_v through possibly different keywords k_w and k_u. However, when $k_w = k_u$, it is said that p_i and p_j are "strongly connected." The connectivity of an article p, $con(p)$, is the number of composite links to other articles, while the weighted connectivity of an article p, $wcon(p)$, is the total number of links to other articles. An "island" is a set of loosely connected articles, while a "rock" is a set of strongly connected articles. It is noted that a rock is a complete graph, where all the elements of the rock are linked to each other. The size of an island or a rock is the number of articles belonging to it. The "coverage" of a rock on an island is defined as the size of the rock divided by that of the island. A reef is a singular article that has no link, or common keyword, to any other articles. The size of a reef is always 1. In Figure 1, article p_1 has keywords $K_1 = \{a,b,c,d,e,f,g\}$, article p_2 has keywords $K_2 = \{a,g,w\}$. Thus, there are two links $L(p_1, p_2, a)$ and $L(p_1, p_2, g)$ between p_1 and p_2. Furthermore, the composite link $CL(p_1, p_2)$ is composed of $L(p_1, p_2, a)$ and $L(p_1, p_2, g)$ with weight $WL(p_1, p_2) = 2$. Meanwhile, the composite link $CL(p_1, p_3)$ has weight $WL(p_1, p_3) = 1$, since there is only one link between them. As a result, the connectivity of p_1 is $con(p_1) = 5$ with weight $wcon(p_1) = 7$. In Figure 1, p_2, p_7, p_8, and p_9 are strongly connected as a rock with a size of four, since they share a common keyword w and form a complete graph. Meanwhile, p_2 and p_3 are loosely connected, although they do not have any common keywords, but they share different keywords, say a and b, with p_1. In this way, p_1, p_2, p_3, p_4, p_5, p_6, p_7, p_8, and p_9 are loosely connected as an island with a size of nine. The coverage of rock p_2, p_7, p_8, and p_9 on the island is 0.44%. There is another island formed by p_{10} and p_{11} with a size of two. There is a reef p_{12} that does not share any keywords with any other articles. It is noted that a reef is not taken as an island.

3.2 Analysis process

To study the evolution of knowledge associated with a topic of interest, which is represented by an exact keyword or keywords, there are three stages: construction and enumeration of knowledge maps, analyses of temporal knowledge maps, and interpretation.

3.2.1 Construction and enumeration of knowledge maps

In this stage, articles and their keywords are retrieved, and maps are constructed and enumerated based on a given keyword KW, which is the core concept on which the study focuses. We take KW as an initial keyword to retrieve all the articles D_t whose titles or keywords contain KW from a journal database in a certain time interval t, say one year.

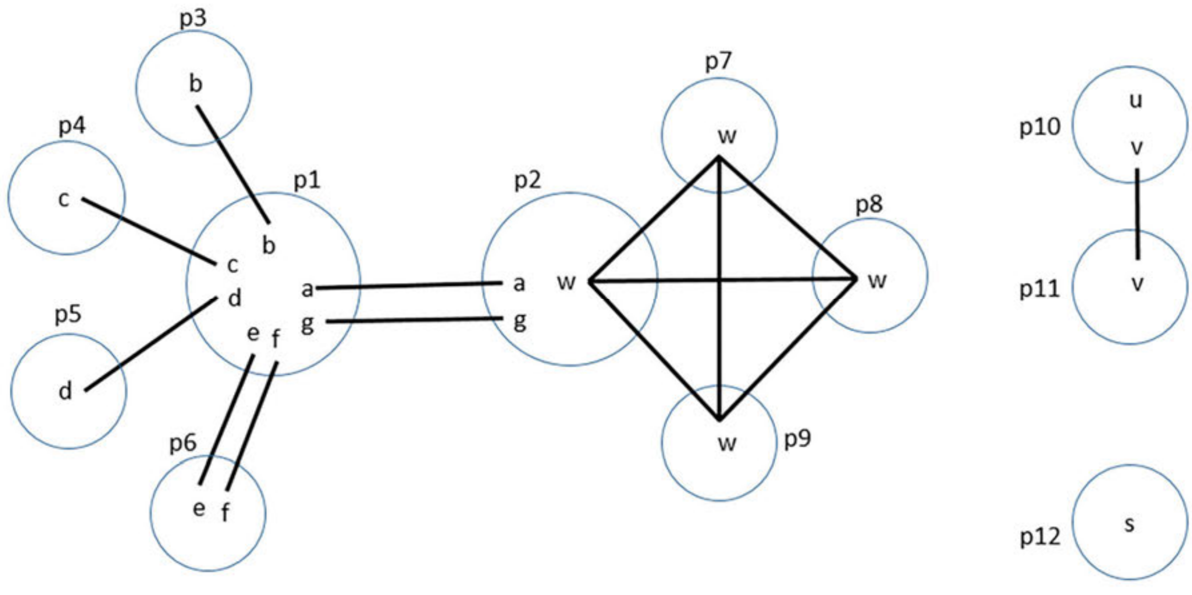


Figure 1. Explanatory diagram of a knowledge map

The journal database may cover several consecutive time intervals. Knowledge maps are constructed for each time interval, such that a temporal evolution can be analyzed.

Since KW will appear in every article under this situation, all the articles will be connected and form a single big stone island due to the common keyword KW. It turns out to be a trivial problem. Thus, the given keyword KW should be removed from data set D_t, and it can be thought of as the scope of the study. Then, the frequency for each remaining keyword, the simple and composite links between pairs of articles, the degree and their weights can be constructed and enumerated. Finally, islands, rocks, and reefs are identified, and their numbers are enumerated. The detailed procedure is illustrated below.

1. For a given keyword KW, retrieve all the articles whose titles or keywords contain KW from a journal database during a certain time period, which can be divided into several time intervals. Let D_t be the set of articles retrieved from time interval t.
2. For each time interval t
   - 2.1 Extract the set of keywords K_i for each article p_i from D_t.
   - 2.2 Transform synonyms, the original compound nouns of abbreviations, acronyms, and initials into standard keywords in lowercase letters.
   - 2.3 Remove the given keyword KW from all the K_i.
   - 2.4 Let K be the union of all the keywords in D_t; i.e., K = $K_i \cup K_j \cup \ldots = \{k_{i1}, k_{i2}, \ldots\} \cup \{k_{j1}, k_{j2}, \ldots\} \cup \ldots$
   - 2.5 For every keyword $k \in K$ associated with D_t, count its frequency $freq(k, D_t)$.
   - 2.6 For every pair of articles p_1 *and* p_2, make a simple link $L(p_i, p_j, k)$ between them if they share a common keyword k.
   - 2.7 For every pair of articles p1 and p2, make a composite link CL(pi, pj) between them if any simple link $L(p_i, p_j, k)$ exists.
   - 2.8 For every article $p \in D_t$, count its degree $deg(p)$ and weighted degree $wdeg(p)$.
   - 2.9 Identify islands, count the number of islands and the size of each island of D_t.
   - 2.10 Identify number of reefs of D_t.
   - 2.11 For every island, identify internal rocks with respect to different keywords; count the number of rocks and the size of each rock.

3.2.2 Analysis of temporal knowledge maps

After constructing and enumerating a temporal series of knowledge maps within the scope of a given keyword KW, several analyses can be performed:

– What keywords have the highest frequency? Do they change over the course of time?
– How many islands and reefs do the articles form?
– What are the sizes of the islands from the largest to the smallest? Are any trends evident?
– What is the relationship between the largest island and rock? Can one find the largest rock in the largest island?
– What is the highest strength (weighted link) between two articles?

3.2.3 Interpretation

Finally, one may interpret the results of the analysis in terms of domain knowledge. For example, if the scope of interest is "smart city" (the given keywords), some interpretations and queries may be made as follows.

– What are the hottest terms? Do they change over the course of time?
– What terms are emerging? What terms are fading out?
– Are there competing groups within the interested topic?

3.3 Datasets and software

In present research practices, researchers consult academic databases and use various tools for their research work. Many databases, such as Web of Science, Scopus, Crossref, ArXiv, etc., collect academic articles. Additionally, many tools, such as VOSViewer, CiteSpace, HistCite, SciMAT, Sci2, etc., visualize and analyze the relations among articles in databases (Chen 2017). Although they possess friendly user interfaces, convenient analysis functions, and colorful visual windows for dynamic layouts, they are general-purpose software tools, insufficient to support analyses where particular characteristics of specific disciplines need customized considerations. For instance, in this paper, instead of using a fixed selection criterion, we must choose different percentages of keywords as the hottest keywords in different time periods due to different total numbers of articles and keywords in different years.

We used the Web of Science core collection database (https://clarivate.com/products/web-of-science/web-science-form/web-science-core-collection/) as a testbed, which collects articles mainly from academic journals and conferences. Web of Science provides two methods to access their databases. The authorized users either visit their web pages or gains access through API to download retrieved articles with their titles, authors, publication names, year of published, organizations, and other auxiliary information after specifying an interested database, search words, timespan, and citation indices. In our research, articles with the keywords "smart city" and published in the period from 2009 to 2018 were selected. Keywords in academic articles are semi-structured, freely provided by the authors, and composed of an indefinite number of single or compound nouns. Individual articles and keywords are two study units, as shown in Figure 1, for further analysis using the process described in Section 3.2.

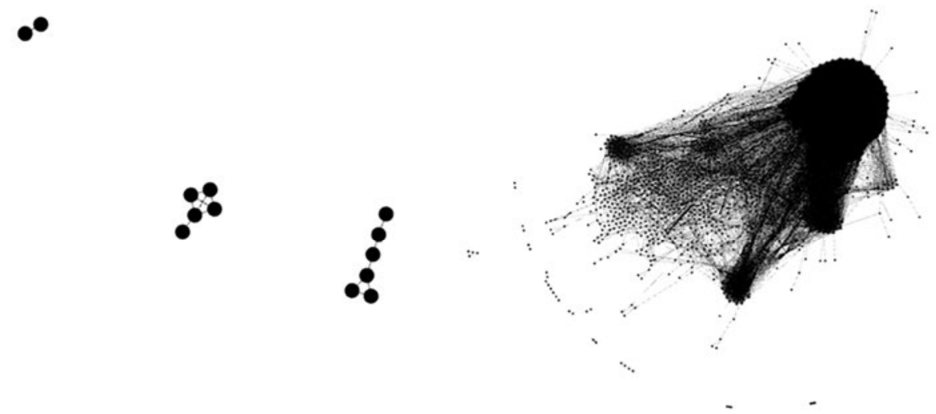
4.0 Results

4.1 Graphical knowledge maps of islands

As mentioned above, knowledge maps can be presented in graphical figures or textual lists. In this subsection, we present graphical knowledge maps of the years 2009 and 2018 in Figures 2 (a) and (b), respectively. Each dot represents an article. Each link denotes that there are common keywords between the two linked articles, while its thickness represents its strength or weight, i.e., the number of common keywords. Note that singular nodes, having no common keyword with any other articles, are not shown in the figures. Figure 2(a) is relatively simple to read and understand, as there are only thirteen articles in three islands. In this case, all the weights of the links are equal to one. However, Figure 2(b) is quite messy. There is a big black "rock" and several smaller black "rocks" in the biggest island, while some much smaller islands line the lower left. Although some thicker links can be seen in the figure, it is almost impossible to identify and count them by visualization only.

4.2 Frequency and coverage

In early years, few articles mentioned smart cities (Figure 3(a)); however, the numbers have increased dramatically in the last decade. Figure 3(b) shows that the number of articles with keywords quickly increased in the last decade



(a) (b)

Figure 2. Knowledge maps of year 2009 (a) and 2018 (b).

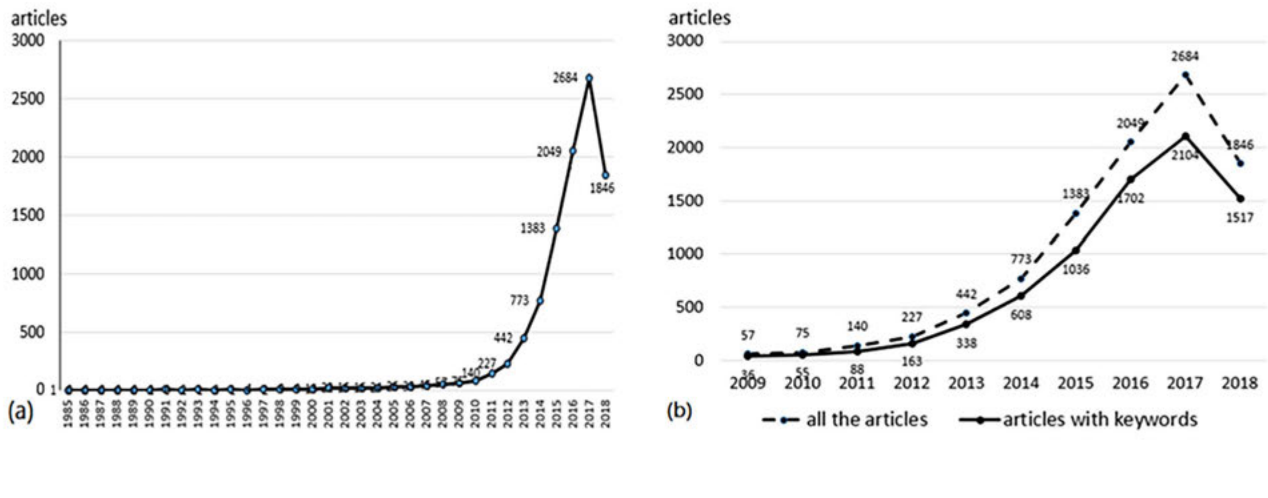


Figure 3. Number of articles since 1985 (a), and since 2009 (b).

except for a slight decrease in 2018. We will further analyze the evolution in the last decade in the remainder of this paper. It is noted that not all the articles provide their keywords. Also, the number of keywords, where duplicate keywords in different articles are counted only once, follow a similar trend; however, it drops earlier in the year 2016 (Figure 4). Although the number of both articles and keywords being used by the authors dropped in the last few years, at this moment it is hard to say whether they will continue to decrease in the near future.

It will be very interesting to know which keywords are most commonly mentioned and examine the evolution of them. For being manageable, the number of hot keywords should be limited. In this research, due to different amounts of articles in different periods, different criteria are needed to select keywords for a meaningful comparison. In this research, the number of articles with keywords in the first half of the decade is relatively smaller than that of the second half. It calls for different criteria to choose keywords from the first and last halves of the decade for identifying "hot" keywords and their associated trends. As a rule of thumb, the keywords that occurred more than once each year in the first half are chosen, while more than five times in the second half of the decade. As a result, the number and percentage of hot keywords increased in the first half of the decade from seven (4.00%) to 122 (9.8%), while the number of hot keywords in the second half varied from twenty to ninety-five, and their percentages are kept in the range between 1% to 2%. The hot keywords in the year 2013 were chosen by both of criteria of the first and second half of the decade. The gap between these two criteria is 8.27% (Figure 5).

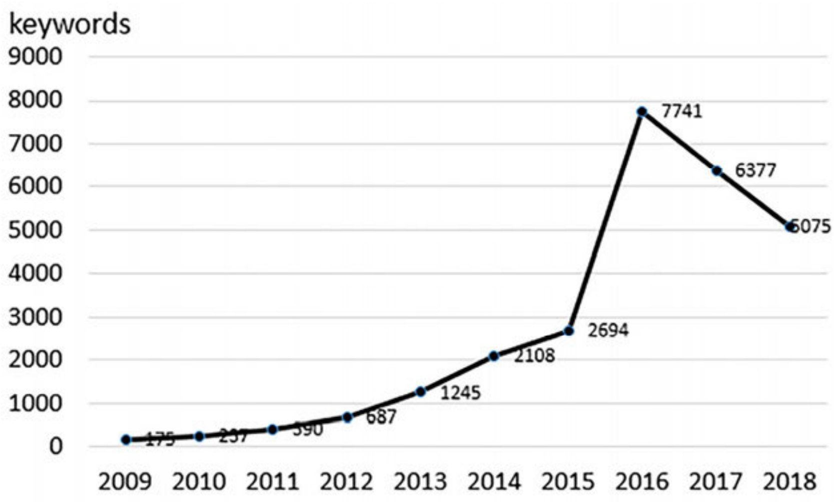


Figure 4. Numbers of keywords.

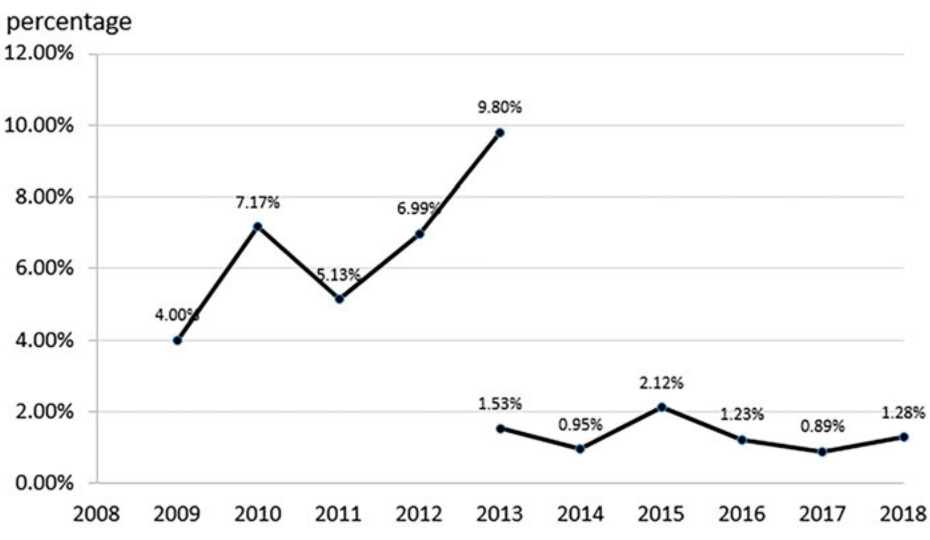


Figure 5. Percentage of hot keywords against all keywords.

As an illustration, Table 1 shows the evolution in major hot keywords, whose frequencies are in the top three highest in any year of the last decade. The percentages below the frequencies are quotients of frequencies divided by the number of articles with keywords of their corresponding years. Thus, we call the percentages "coverages." In this way, we identify the thirteen major hottest keywords, namely: IoT, big data, cloud computing, sustainability, smart grid, ICT, urban development, smart growth, GIS, tourism, ubiquitous computing, smart planet, and u-city.

The second column of Table 1 indicates the properties of these keywords, where "T" means technology that smart cities employ, while "V" means values that smart cities pursue. There are seven keywords with "T" and six with "V." Although they seem roughly equal, but as a matter of fact, keywords concerning value is overwhelmed by those concerning technology in terms of frequency. For example, in 2018, keywords with "T" cover 32.9%, while keywords with "V" only cover 4.48% of the articles with keywords.

The hottest keywords shifted yearly. In the first two years, they were "smart growth," while in the second two years they were "smart grid." The frequencies of the hottest keywords in the first four years were relatively small. During the years 2013 to 2018, "IoT" (Internet of Things) held the position of the hottest keyword with a trend of increasing frequencies and percentages against the numbers of articles with keywords.

Figures 6 and 7 illustrates the evolving trends of the thirteen hottest keywords in terms of frequencies and their coverage in the corresponding years. Figure 6 shows that the differences between the frequencies of the top and

years		**2009**	**2010**	**2011**	**2012**	**2013**	**2014**	**2015**	**2016**	**2017**	**2018**
number of articles with keywords		36	55	88	163	338	608	1036	1702	2104	1517
IoT	T				9	34	56	106	245	340	288
					5.52%	10.06%	9.21%	10.23%	14.39%	16.16%	19.98%
big data	T					4	16	46	83	96	70
						1.18%	2.63%	4.44%	4.88%	4.56%	4.61%
cloud computing	T				5	8	21	41	73	47	48
					3.07%	2.37%	3.45%	3.96%	4.29%	2.23%	3.16%
sustainability	V		3	3	3	9	18	19	31	47	48
			5.45%	3.41%	1.84%	2.66%	2.96%	1.83%	1.82%	2.23%	3.16%
smart grid	T			11	18	21	29	53	76	61	44
				12.50%	11.04%	6.21%	4.77%	5.12%	4.47%	2.90%	2.90%
ICT	T					10	12	14	29	8	28
						2.96%	1.97%	1.35%	1.70%	0.38%	1.85%
urban development	V			3					16		7
				3.41%					0.94%		0.46%
smart growth	V	4	9		10	5	6		8		7
		11.11%	16.36%		6.13%	1.48%	0.99%		0.47%		0.46%
GIS	T	2	4			6	9	8	15		6
		5.56%	7.27%			1.78%	1.48%	0.77%	0.88%		0.40%
tourism	V			3		3		6	9		6
				3.41%		0.89%		0.58%	0.53%		0.40%
ubiquitous computing	T	3			5	2		6			
		8.33%			3.07%	0.59%		0.58%			
smart planet	V			3							
				3.41%							
u-city	V		4								
			7.27%								

Table 1. Thirteen major hottest keywords.

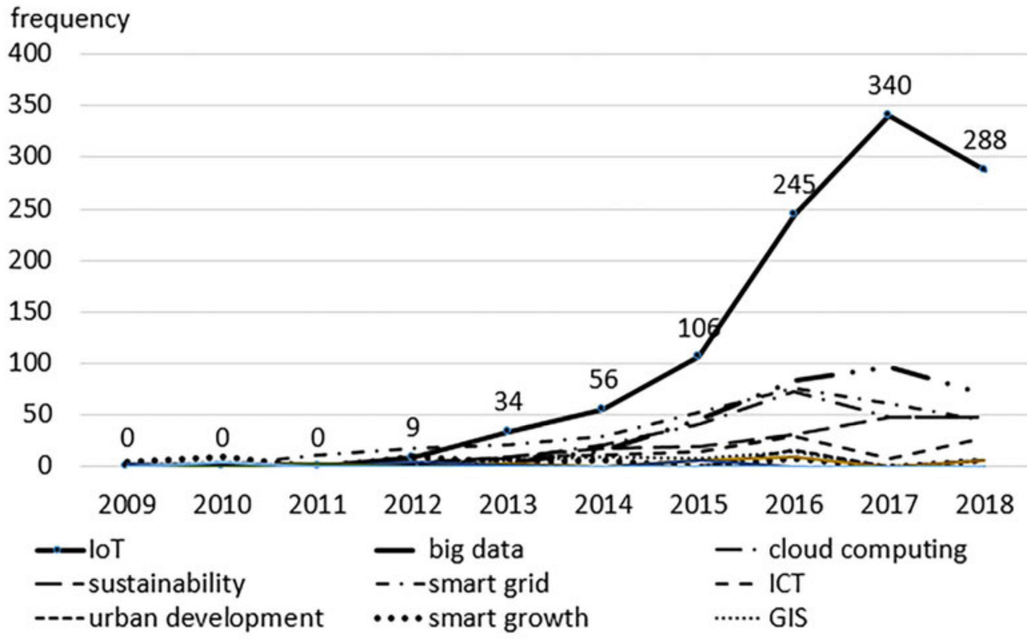


Figure 6. Frequencies of keywords.

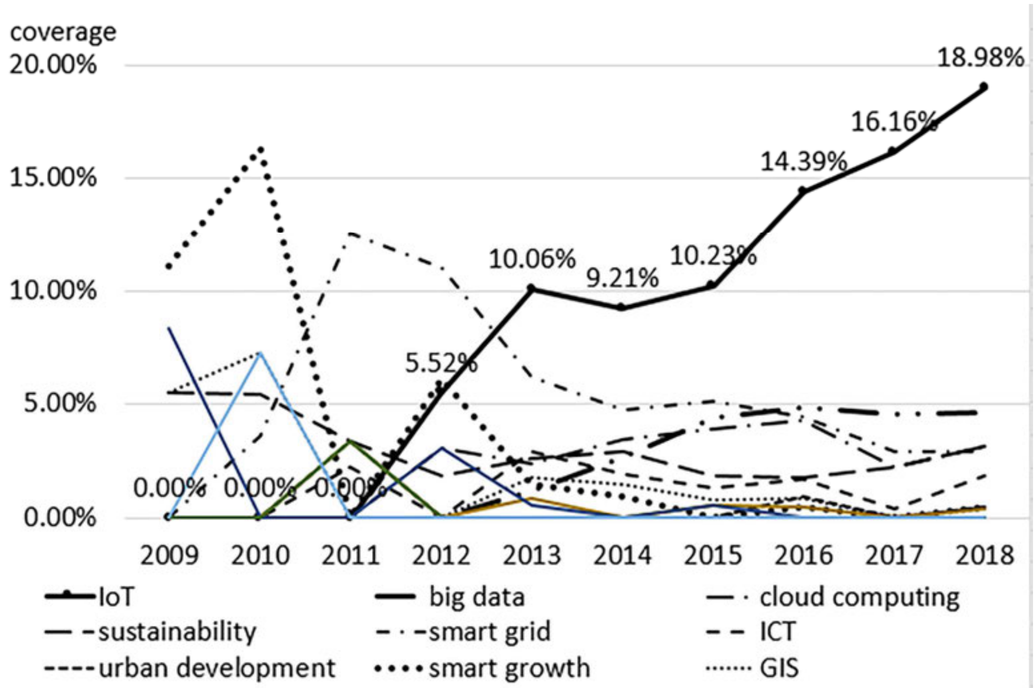


Figure 7. Coverage of keywords w.r.t all the articles.

second hottest keywords grew bigger and bigger. They can be classified into four groups. IoT, the only one member in the first group, obviously dominates the others. Some other keywords, such as "smart growth," had high coverage in the early years, but quickly shrank later. Big data, cloud computing, sustainability, smart grid, and ICT can be treated as the second group, which remain their most coverages between 1.5% and 5%. Among them, "sustainability" is the only value-oriented keyword and has had increasing coverage in recent years, while the others are techology-oriented and have flat or dropping tendencies. Urban development, smart growth, GIS, and tourism comprise the third group. Their frequencies and coverages are relatively smaller but still mentioned recently. The fourth group, consisting of ubiquitous computing, smart planet, and u-city seems to have vanished. It is noticeable that most of the keywords in groups three and four are value-oriented. In other words, in earlier years, researchers focused their efforts from exploring the meaning of smart cities to enriching the value of cities. Gradually, the focus shifted to technology-related issues for making smart cities a reality. Following the same idea, the evolution of any keyword can be explored. In the future work, researchers may further vary the definition of hot keywords and get different sets of them so that their evolution patterns can be further explored.

4.3 Knowledge islands

Connected articles form knowledge islands. As mentioned earlier, if two articles share any keywords, they are linked. Either of them can further link to another article. Thus, consecutively linked articles form an island. The size of an island is at least two articles. Any article that does not link to any another article becomes a reef. In the study concerning smart cities, the numbers of islands (Figure 8), reefs (Figure 9), and articles in islands (Figure 10) simultaneously increase. That means while the number of islands

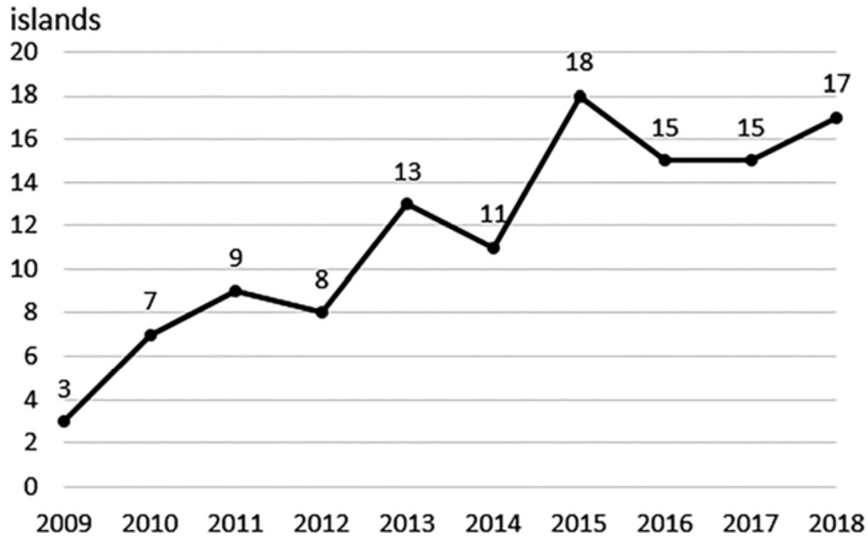


Figure 8. Numbers of islands.

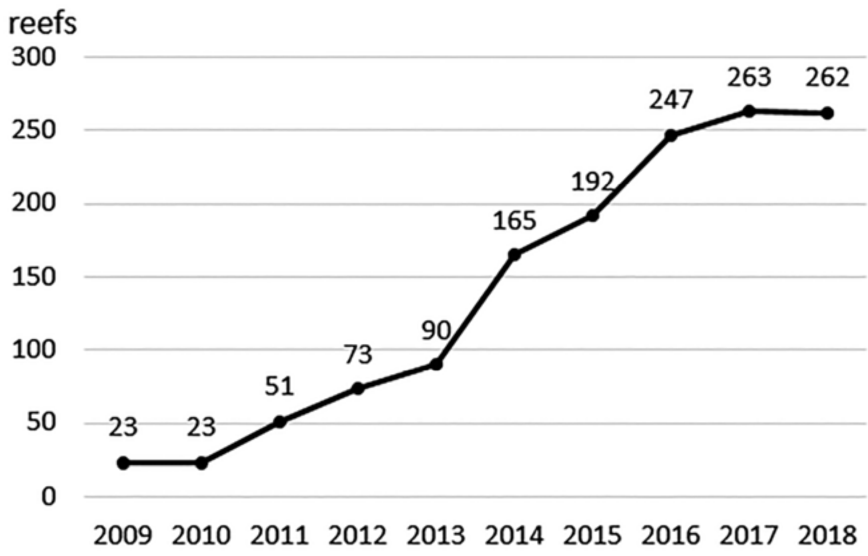


Figure 9. Numbers of reefs.

increases yearly, the largest islands are bigger and bigger, and many new standalone keywords (reefs) also emerge. Figure 10 further shows that the gap between the numbers of all the articles and articles in islands is exactly the number of reefs. Figure 11 also shows that the percentages of all the articles and articles in the largest islands against the total numbers of articles with keywords converge to around 85% gradually. In other words, there is a space of

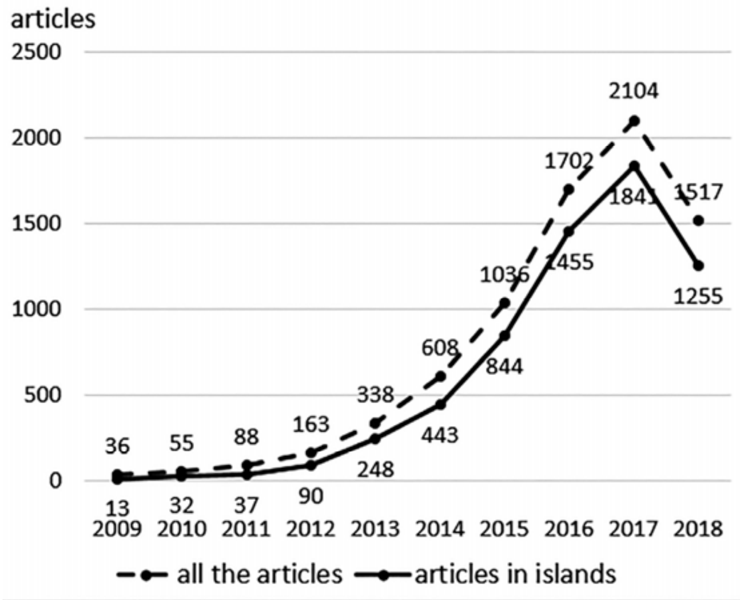


Figure 10. The numbers of all the articles and articles in islands.

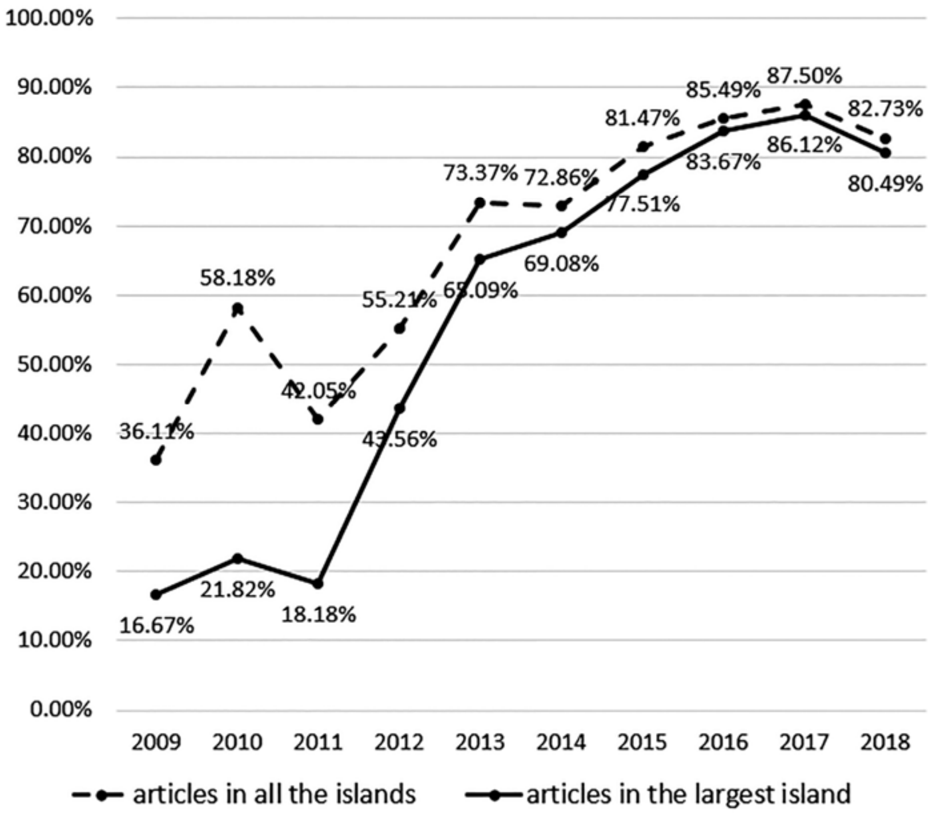


Figure 11. The percentages of all the articles and articles in the largest islands against the total numbers of articles with keywords.

around 15% for much smaller islands and reefs to keep the field of smart cities growing. Figure 12 illustrates how the percentages of number of reefs out of all the articles in corresponding years are also decreasing and converging to a range between 10% and 20%.

4.4 Island size

The size of a knowledge island is the number of articles on it. In the research field concerning smart cities, the sizes of the largest islands are much larger than those of the other islands. Thus, they are separately illustrated in two figures. In Figure 13, the sizes of the largest islands and the number of all the articles, which are shown on the polylines, dramatically increase in a similar trend except the last year. On the other hand, Figure 14 illustrates that the yearly sizes and numbers of the second-, third-, and fourth-largest islands are stable and much smaller than those of the largest islands. Furthermore, Figure 15 shows that the smaller the islands, the more numerous they are. We can imagine that the knowledge map of smart city knowledge is composed of an exceedingly large island, several much smaller islands, and a lot of reefs.

4.5 The strength of links

The weight, or strength, of a composite link is measured by the number of common keywords between two articles that the composite link connects. Table 2 depicts that numbers of links associated with weight from one to ten during years 2009 to 2018. It shows that most of the links are of weight =1, and in the second half of the decade the number of links with weight =2 are less than 3% of those of weight = 1. Furthermore, Figures 16(a)-(d) illustrate the evolutions of the numbers of links with weights from one to five. We find that they have a fractal-like structure. In other words, the relative structure of evolution curves between weight = 1 and the others (Figure 16(a)) and is similar to that between weight = 2 and weights = 3,4,5 (Figure 16(b)). This phenomenon is also held between weight = 3 and weights = 4,5 (Figure 16(c)). Their correlation coefficients between consecutive weight links are calculated in Table 3, which shows that they are highly correlated.

4.6 Connectivity

The connectivity of an article measures how many other articles have common keywords with it. Figure 17 illustrates the yearly evolution of the top three articles with the highest connectivity and the reef's connectivity. The yearly top three articles with the highest connectivity are very close, and their evolutionary trend is very similar to that of the numbers of articles. Since the connectivity of reefs is zero, their evolution line lies in the x-axis. Figure 18 illustrates the number of articles with respect to their connectivity. The number of articles with connectivity less than twenty

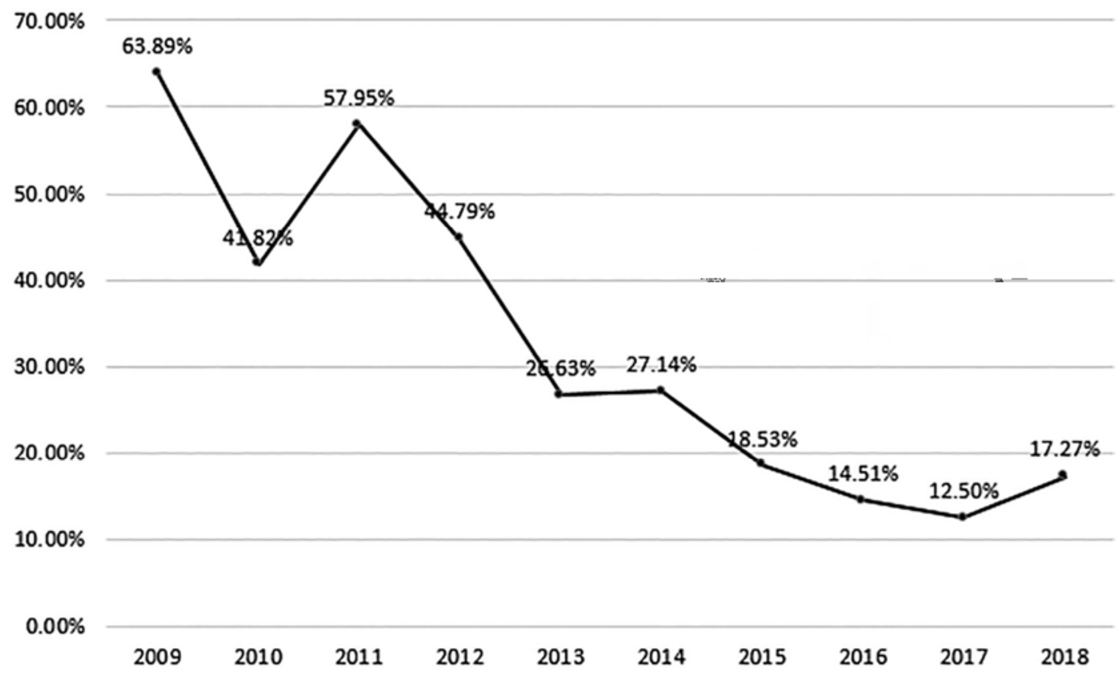


Figure 12. The percentage of number of reefs out of all the articles.

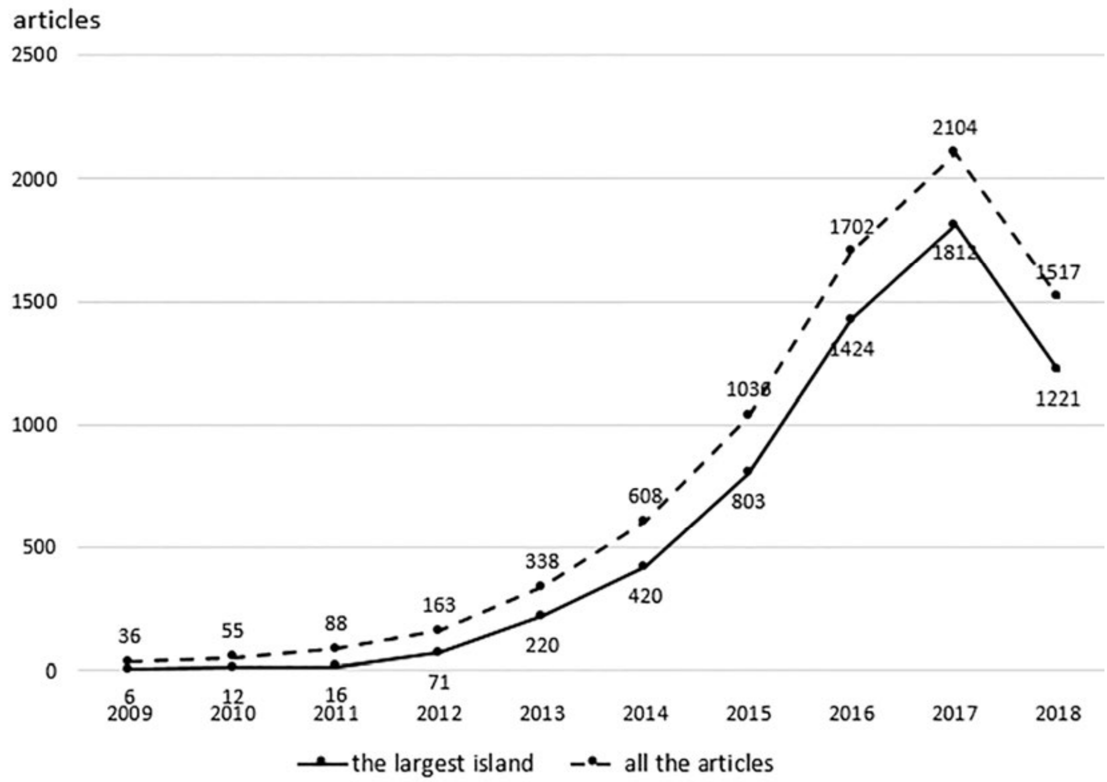


Figure 13. The yearly evolution of sizes (Y-axis) and numbers (on polylines) of the largest islands comparing to the numbers of all the articles.

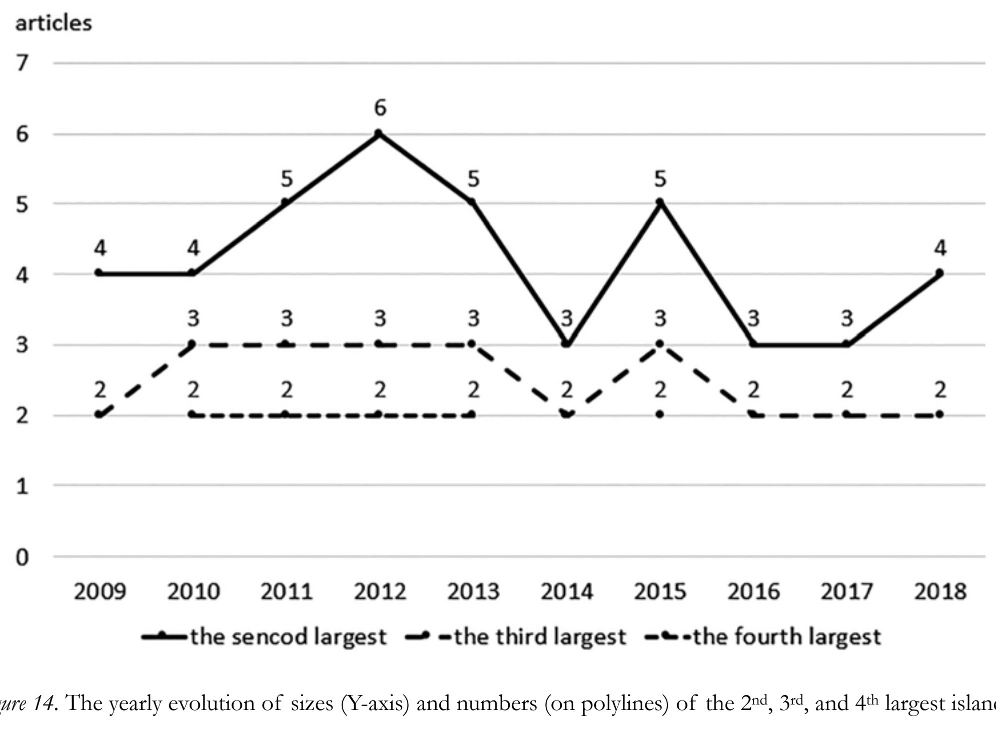


Figure 14. The yearly evolution of sizes (Y-axis) and numbers (on polylines) of the 2nd, 3rd, and 4th largest islands.

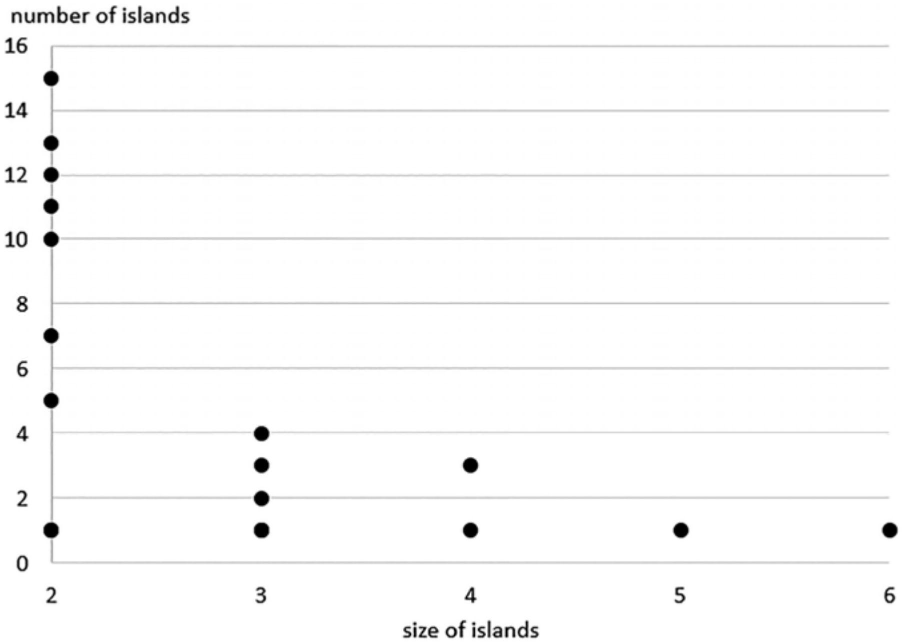


Figure 15. The relationship between numbers and sizes of islands.

	W=1	W=2	W=3	W=4	W=5	W=6	W=7	W=8	W=9	W=10
2009	14									
2010	53	5	1							
2011	43	5	1							
2012	244	4	1	1	1					
2013	911	32	1	1	3					
2014	3330	100	3	3	1					
2015	7651	199	25	13	10	3	1	3		1
2016	41056	801	27	5	3	5				
2017	63176	1470	70	7	5	4	2	1		
2018	36552	887	37	4	2					

Table 2. Numbers of links associated with different weights.

Comparison	Correlation coefficients
W=1 vs W=2	0.9953
W=2 vs W=3	0.9632
W=3 vs W=4	0.6274
W=4 vs W=5	0.9518
W=5 vs W=6	0.6307
W=6 vs W=7	0.6125
W=7 vs W=8	0.6475

Table 3. Correlation coefficients between consecutive weight links.

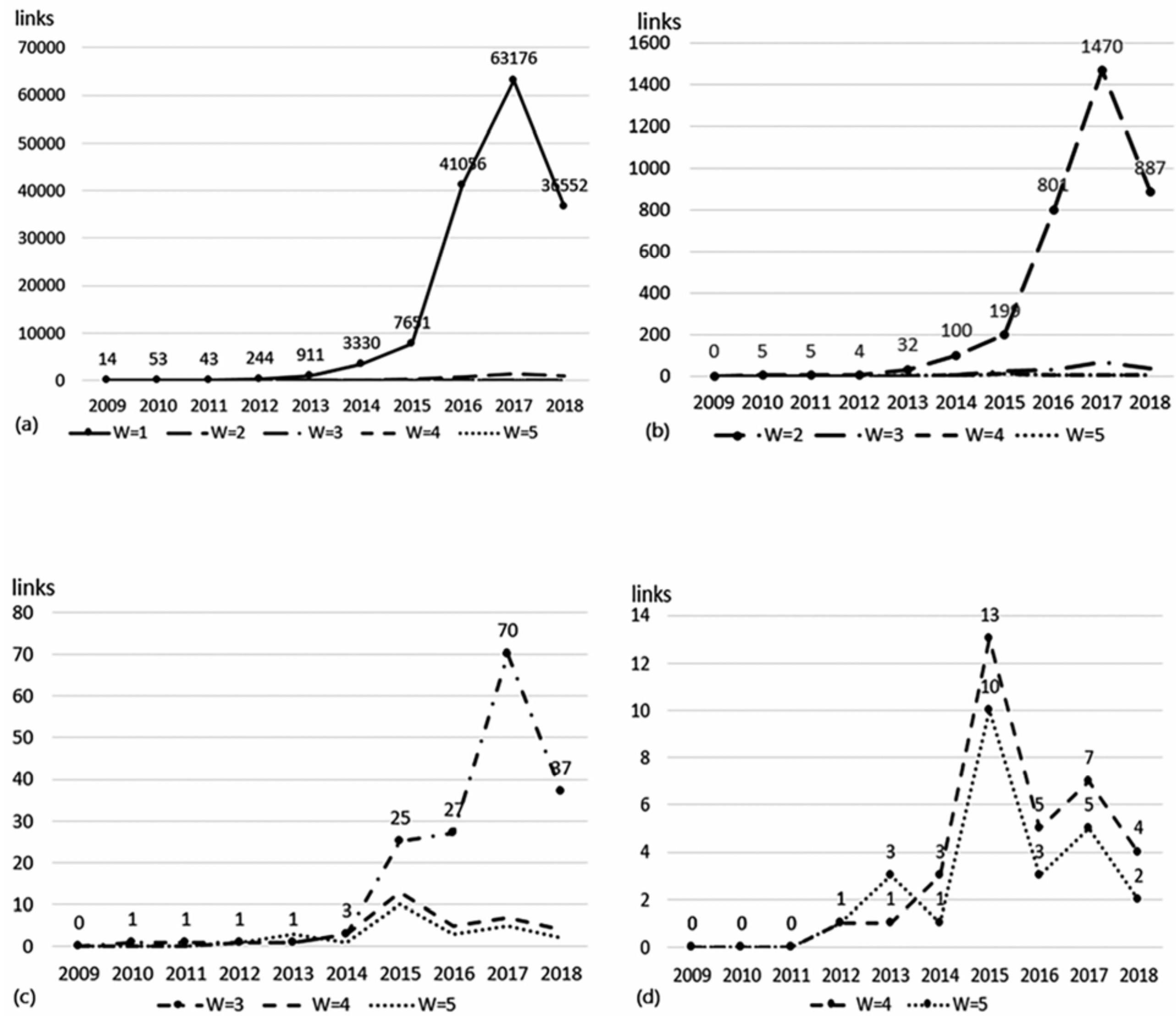


Figure 16. Numbers of links with weights 1-5 (a), 2-5 (b), 3-5(c), 4-5 (d).

is less than six. It is noted that, in this research, beyond the x-axis of Figure 18, the number of all the articles of connectivity more than twenty is one.

A rock is composed of articles sharing common keywords. These articles link to one another and form a complete graph. Different keywords will form different rocks. In other words, the number of rocks is equal to that of keywords shared by different articles. The size of a rock is measured by the number of corresponding articles. Therefore, the thirteen hottest keywords illustrated in Table 1 are the glue of the thirteen largest rocks. However, the largest rock is usually not the article with the highest connectivity. Figure 19 compares the largest rock with the highest connectivity, also shown in Figure 17. The difference between them is attributed to the other keywords that co-exist in the rock. Figure 20 shows the coverage of the largest rock in the island where it is located. It seems that the range of the coverage may be kept between 10% and 20% in the future. Figure 19 shows that the largest rocks are not in the articles with the largest degrees. The coverages of the biggest rocks in islands where they are located converge in the 10% to 20% range (Figure 20).

4.7 Summary of findings

"Smart city" is a buzzword in recent years, and the academic community is no exception. According to Clarivate Analytics' Web of Science database, the term "smart city" first appeared in 1985. However, only a few articles concerning smart cities existed until around 2009. After that, the number of articles has been dramatically increasing. By taking article keywords as pivotal tags to explore oceans of

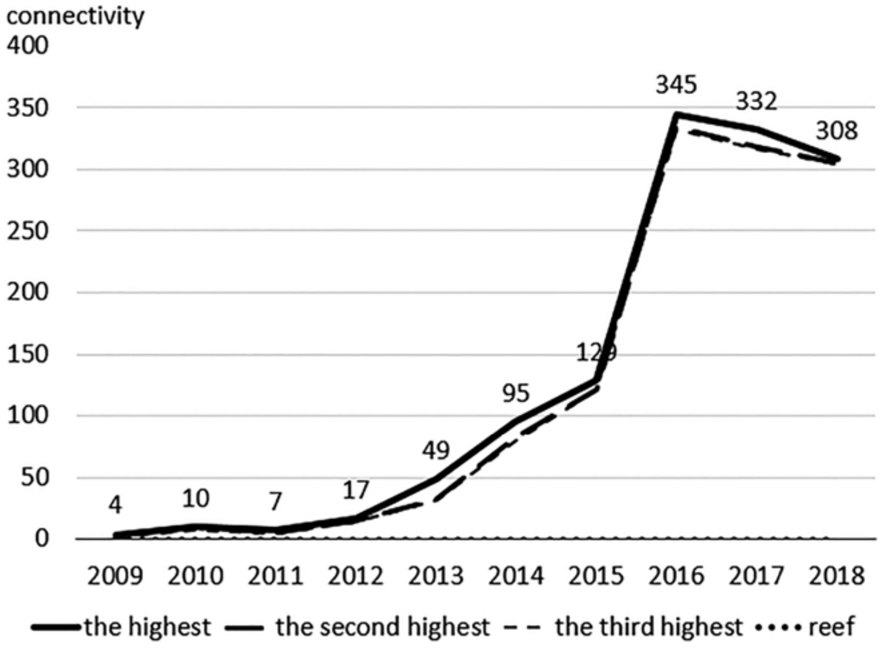


Figure 17. The yearly evolution of the top three highest and reef's connectivity.

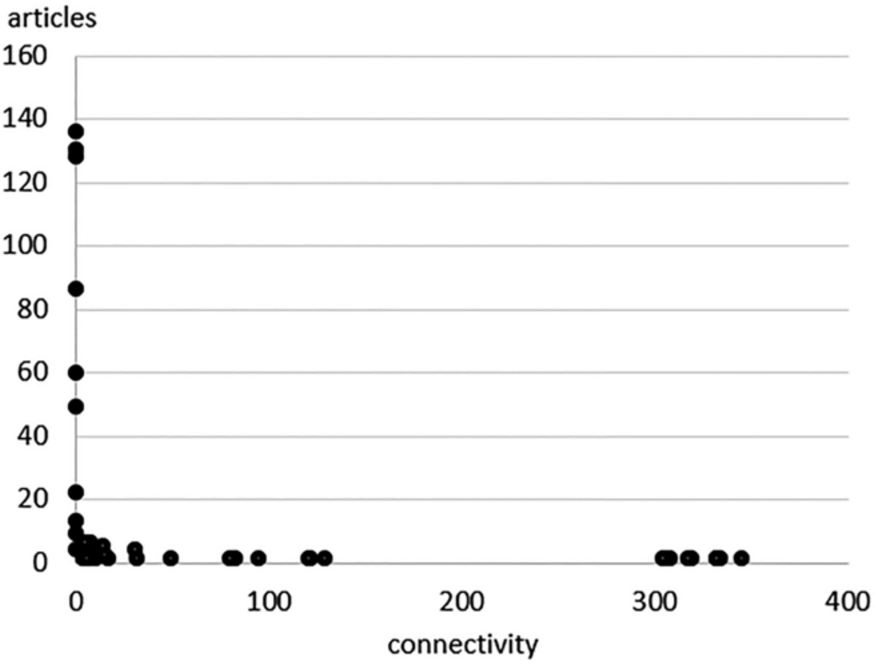


Figure 18. The relationship between the connectivity and number of articles.

academic knowledge concerning smart cities from the year 2009 to 2018, we find that the articles concerning smart cities indeed have enjoyed a booming period in the last decade, except a small drop in 2018. In other words, it is hard to judge whether the study of smart cities has matured and will decline from now on, or whether it is just a little turbulence and will keep on growing in the future.

The characteristics of hot keywords in the first and second half of the decade are quite different. They were chosen if they occurred more than two and five times in the first and second half of the decade respectively. In the first half of the decade, the percentage of the number of hot keywords against that of all the keywords increased, but it stabilized to a range of between 1% and 2%. Furthermore,

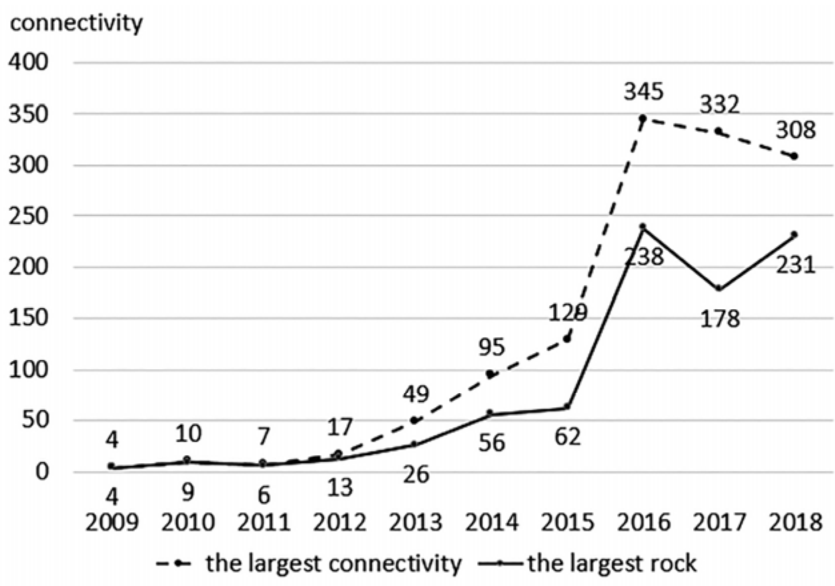


Figure 19. The comparison of connectivity between the largest rock and article.

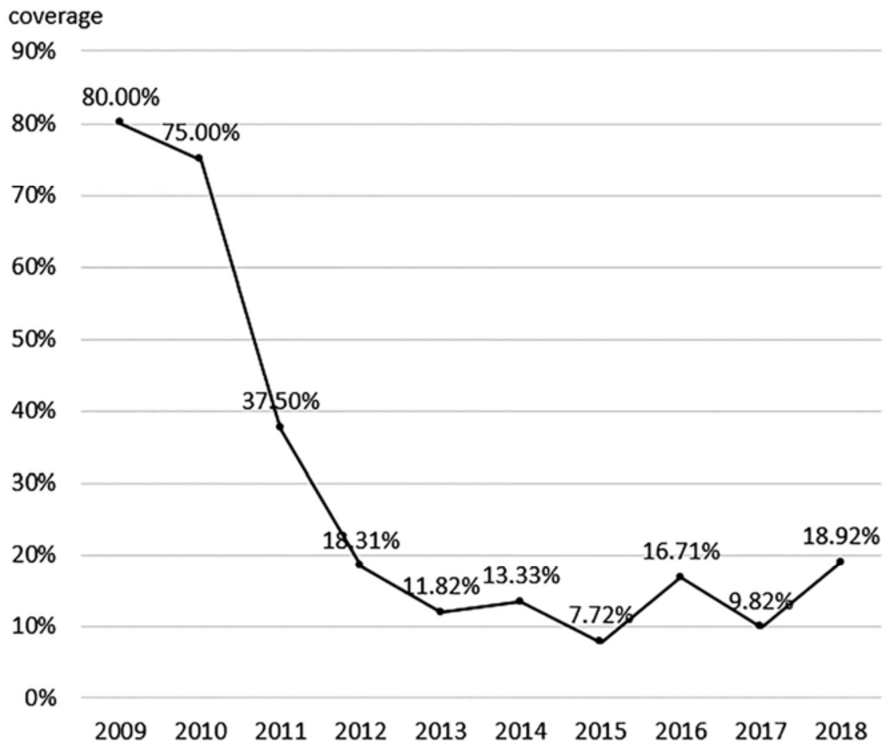


Figure 20. The coverages of the largest rocks in islands.

thirteen major hot keywords, namely IoT, big data, cloud computing, sustainability, smart grid, ICT, urban development, smart growth, GIS, tourism, ubiquitous computing, smart planet, and u-city, were selected. The numbers and percentages of keyword IoT, which was the hottest keyword in the last six years, far exceeded those of the other hot keywords.

A knowledge island contains many knowledge rocks. Since there are many keywords in an article, it can indirectly connect to other articles through different keywords, while articles sharing a common keyword are directly interconnected as a complete graph. In other words, a knowledge island is composed of loosely connected articles, while strongly connected articles form a rock. It was found that over a decade the number of islands increased from three to more than fifteen, and their sizes increased from thirteen to more than 1,200. The sizes of the largest rock covered those of islands where they are located are around 10% to 20% for the articles about smart cities. Reefs are articles that share no keywords with other articles. As a result, an exceedingly large island, several much smaller islands, and a lot of reefs are present on the knowledge map of smart city knowledge.

5.0 Discussion

5.1 Comparative study

It is worth comparing our work with similar work done by Mora, Deakin and Reid (2018), as mentioned in section 2.3. Although the two studies used different data sources, timespans, association tags, grouping approaches, and hot keyword selections to explore how the concept of a smart city is intellectually structured, both have compatible and progressive findings. Both studies took English-language literature from scholarly databases as source articles, but Mora, Deakin and Reid searched eight databases, namely Google Scholar, Web of Science, IEEE Xplore, Scopus, SpringerLink, Engineering Village, ScienceDirect, and Taylor and Francis Online, from 1992 to 2012, while we focused on Web of Science from 2009 to 2018. Both studies involved selecting articles in which the term smart city is included in the title, abstract, keyword list, while Mora, Deakin and Reid also searched the body of the text. As a result, Mora, Deakin and Reid had 2,273 source articles, and we had 7,647 in the last decade and 6,967 in the second half. The observation that research concerning smart cities has been dramatically increasing is supported by the two independent studies where the number of articles considered in the present study from one database in the last five years is much larger than that by Mora, Deakin and Reid from various databases in twenty years. To group articles, the study authors took different tags; where Mora, Deakin and Reid used a subject-oriented co-citation approach, we used a frequency-oriented co-keyword approach. Although they used different approaches, the numbers of islands (clusters) are very close, where Mora, Deakin and Reid got eighteen clusters, we found there were seventeen or eighteen islands in the last four years.

Furthermore, both studies used different criteria to select distinct or hot keywords but still achieved some agreement and implied the trend of evolution. Mora, Deakin and Reid selected the top ten keywords in eighteen clusters and made a profile of thirty-one distinct keywords, while we selected the top three keywords from each year in the last decade and came up with a list of thirteen hot keywords. There are four keywords, namely IoT, ICT, smart grid, and urban development, shown in both studies. On the other hand, hot keywords of 2018, namely big data, cloud computing, and sustainability, indicate the new trend of research interests. Furthermore, both studies agree that technology-oriented articles are overwhelming in the research community of the smart city.

5.2 Categorization vs. classification

It is worth mentioning that the characteristics of knowledge maps are closer to categorization than classification based on Jacob (2004), who identified classification and categorization by six systemic properties: process, boundaries, membership, criteria for assignment, typicality, and structure. Categorization processes entities using creative synthesis based on similarity and has a non-binding boundary. The criteria of category assignment can be context-dependent or context-independent; thus, the membership of an entity is flexible and can be associated with more than one category. There is no typical or representative member in a category since every member has its own different properties. The structure of a category may be flat or hierarchical. On the other hand, classification arranges entities in a systematic process based on their characteristics using predetermined assignment criteria; thus, classes are mutually-exclusive and non-overlapping, and boundaries are fixed where an entity either is or is not a member of a particular class. All members of a class are typically and equally representative. Classes can be hierarchically structured. In the case of knowledge maps, we group articles collected in WoS using a single criterion of assignment by connecting common keywords among them to form islands. If we change the relationship based on common keywords to other relationships, such as common authors, references, etc., articles will be grouped in different ways. Thus, we create a knowledge map in a categorization process. Since every article in an island (category) has different numbers of keywords associated with different frequencies, and different numbers of link strength associated with different other articles, no article can be a representative for other articles on the same island. Furthermore, the structure of a knowledge map is flat, non-hierarchical. Thus, the properties of process, criteria for assignment, typicality, and structure coincide with those of categorization. However, the other two properties, namely, membership and boundaries, behave like classification. When the criterion of assignment is given, and an article is once connected to an island, it will not change to any other island. Any two separated islands have no common keyword. In other words, any article either only belongs to an island or becomes a reef by itself. The boundaries of knowledge islands are fixed, and the islands are mutually exclusive.

5.3 Standardization of keywords

Keywords which were freely provided by the authors of articles need to be pre-processed in order to have a standardized analysis. A concept may be expressed in various terms or forms. For examples, "Internet of Things" may also be expressed as IoT, Internet-of-Things, Internet-of-Things (IoT), IoT(iot), etc. A reference table has to be built for integrating many synonyms into one. However, it is not necessary to define a limited set of control words so that innovative keywords are possible.

5.4 Future investigation

Our future work has two parts. While possible research directions for the academic community at large will be suggested in section 6.0, in this section we will discuss the future investigation of the construction and exploration of knowledge maps to explore the texture of a certain domain of knowledge, which can proceed based on the experience gained in this study. Some proposed approaches are as follows:

- measuring the distance between any two keywords: distance can be measured by the number of articles in the shortest path between two keywords; it would be very interesting to find how many years it takes for two keywords to become closer.
- measuring the density of an island: density can be measured by the ratio of the total degree of articles and/or keywords against that of a complete graph; it would be interesting to find the relationship between the evolution of densities and the cohesion of a community of interest.
- identifying patterns of life cycles of keywords: in this research, we find that some keywords in early years might be shrinking, disappearing or reviving, while others might suddenly appear in a great amount and increase dramatically; these phenomena might be affected by technology breakthroughs or socioeconomic issues.
- calculating the entropy of the distribution of islands: the entropy can be a measurement of the vitality of a community of interest a high entropy might imply a vital community in which there are many reefs or small islands with independent and creative ideas; on the other hand, a low entropy might imply that the community has focused on a set of specific topics.
- exploring fractal-like structure: we have revealed in the present study that there is a fractal-like structure embedded in the strengths of composite links in terms of their weights; this phenomenon deserves further exploration.

6.0 Conclusion

A Knowledge map is a powerful tool to capture the whole picture of a certain knowledge domain. However, one may get various pictures if different sources, timespans, association tags, grouping algorithms, and categorization processes are employed. In this article, we have explained how we explored and what we found. A comparison between this research and a similar but independent study was made, and the comparison shows that the collective findings are compatible and progressive.

The evolution shown by the knowledge maps not only illustrates the current situation of the academic community, but also indicates possible future research directions for the academic community interested in the field of smart cities. The results of this research imply that the academic community may have reached a common consensus about the issue of IoT recently. It may also signify the maturation of the topic of smart cities. Additionally, keywords concerning the value of smart cities for pursuing a better life and environment are overwhelmed by those concerning technology. Since the issues of smart cities have many facets, it is suggested that issues concerning values of smart technology, such as sustainability of urban development, social equity and justice, economic growth, adaption of climate change, etc., should be further explored in future research. Furthermore, although many scholarly databases collect published journal and conference papers, unpublished reports, and grey literature, many of them do not provide or only provide limited metadata for further academic research. If they can be downloaded more easily and made user-friendly, knowledge maps of different sources, viewpoints, tags and disciplines can be drawn more quickly. It would be very beneficial to accumulate holistic knowledge.

References

Ahmed, Yunis Ali, Mohammad Nazir Ahmad, Norasnita Ahmad and Nor Hidayati Zakaria. 2019. "Social Media for Knowledge-sharing: A Systematic Literature Review." *Telematics and Informatics* 37: 72-112.

Allahyari, N., Mark S. Fox and Michael Gruninger. 2014. "City Knowledge Patterns: A Standard for Smart City Knowledge Management." Paper presented at Semantic Cities: Beyond Open Data to Models, Standards and Reasoning Workshop at AAAI14, Quebec City July 28, 2014. https://www.aaai.org/ocs/index.php/WS/AAAIW14/paper/viewFile/8831/8267

Balaida, Ali, Mohd Zaidi Abd Rozana, Syed Norris Hikmi and Jamshed Memon. 2016. "Knowledge Maps: A Systematic Literature Review and Directions for Future Research." *International Journal of Information Management* 36: 451-65.

Bateman, Scott, Carl Gutwin and Miguel Nacenta. 2008. "Seeing Things in the Clouds: The Effect of Visual Features on Tag Cloud Selections." In *Proceedings of the 19th ACM Conference Hypertext and Hypermedia 16-21 June 2008 Pittsburgh, PA, USA,* ed. Brusilovsky and Hugh Davis. New York: ACM, 193-202.

Biloslavo, Roberto and Max Zornada. 2004. "Development of a Knowledge Management Framework within the Systems Context." In *The Fifth European Conference on Organizational Knowledge, Learning and Capabilities 2-3 April 2004 University of Innsbruck, Austria*. Coventry, UK: University of Warwick. https://warwick.ac.uk/fac/soc/wbs/conf/olkc/archive/oklc5/papers/h-3_bilo slavo.pdf

Boyes, Bruce. 2016. "Smart Cities and Knowledge Management." *RealKM Magazine*. https://realkm.com/2016/07/22/smart-cities-and-knowledge-management/

Brachos, Dimitris, Konstantinos Kostopoulos, Klas Eric Soderquist and Gregory Prastacos. 2007 "Knowledge Effectiveness, Social Context and Innovation." *Journal of Knowledge Management* 11, no. 5: 31-44.

Chemchem, Amine and Habiba Drias. 2015. "From Data Mining to Knowledge Mining: Application to Intelligent Agents." *Expert Systems with Applications* 42: 1436-45.

Chen, Chaomei. 2017. "Science Mapping: A Systematic Review of the Literature." *Journal of Data and Information Science* 2, no.2: 1-40. doi:10.1515/jdis-2017-0006

Cheng, Ying, Ken Chen, Hemeng Sun, Yongping Zhang and Fei Tao. 2018. "Data and Knowledge Mining with Big Data towards Smart Production." *Journal on Industrial Information Integration* 9: 1-13.

Eppler, Martin J. 2013. "What Is an Effective Knowledge Visualization? Insights from a Review of Seminal Concepts." In *Knowledge Visualization Currents*, ed. F. T. Marchese and E. Banissi. London: Springer, 3-12.

Gambette, Philippe and Jean Véronis. 2010. "Visualizing a Text with a Tree Cloud." In *Proceedings of the 11th IFCS Biennial Conference and 33rd Annual Conference of the Gesellschaft für Klassifikation e.V. March 2009 Dresden, Germany,* ed. Hermann Locarek and Claus Weihs. Berlin: Springer, 561-9.

Hao, Karen. 2019. "We Analyzed 16,625 Papers to Figure out Where AI is Headed Next." *MIT Technology Review* (blog), January 25. https://www.technologyreview.com/s/612768/we-analyzed-16625-papers-to-figure-out-where-ai-is-headed-next/

Heimerl, Florian, Steffen Lohmann, Simon Lange and Thomas Ertl. 2014. "Word Cloud Explorer: Text Analytics Based on Word Cloud." In *Proceedings of the 47th Hawaii International Conference on System Sciences January 2014 Hawaii, USA*. Washington, DC: IEEE Computer Society, 1833-42.

Hjørland, Birger. 2008. "What is Knowledge Organization (KO)?" *Knowledge Organization* 35: 86-101.

Jacob, Elin. 2004. "Classification and Categorization: A Difference that Makes a Difference." *Library Trends* 52, no. 3: 515-40.

Jennex, Murray E. and Iryna Zakharova. 2006. "Culture, Context, and Knowledge Management." *International Journal of Knowledge Management* 2: i-iv.

Liu, Jun, Zhinan Zhang, Richard Evans and Youbai Xie. 2019. "Web Services-based Knowledge Sharing, Reuse and Integration in the Design Evaluation of Mechanical Systems." *Robotics and Computer Integrated Manufacturing* 57: 271-81.

Liu, Lu, Jing Li and Chenggong Lv. 2009. "A Method for Enterprise Knowledge Map Construction based on Social Classification." *Systems Research and Behavioral Science* 26: 143-53.

Mclnerney, Claire. 2002. "Knowledge Management and the Dynamic Nature of Knowledge." *Journal of the American Society for Information Science and Technology* 53: 1009-18.

Medelyan, Alyona. 2018. "5 Text Analytics Approached - A Comprehensive Review." Thematic October 02, 2018. https://getthematic.com/insights/5-text-analytics-approaches/

Meijer, Albert and Manuel Pedro Rodríguez Bolívar. 2016. "Governing the Smart City: A Review of the Literature on Smart Urban Governance." *International Review of Administrative Sciences* 82: 392-408.

Mora, Luca, Mark Deakin and Alasdair Reid. 2018. "Combining Co-citation Clustering and Text-based Analysis to Reveal the Main Development Paths of Smart Cities." *Technological Forecasting & Social Change* 142: 56-69.

Ong, Thian-Huat, Hsinchun Chen, Wai-ki Sung and Bin Zhu. 2005. "Newsmap: A Knowledge Map for Online News." *Decision Support Systems* 39: 583-97.

Rivadeneira, Walkyria Goode, Daniel M. Gruen, Michael Muller and David Millen. 2007. "Getting Our Head in the Clouds: Toward Evaluation Studies of Tagclouds." In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems April 28-May 03, 2007 San Jose, CA, USA*. New York: ACM, 995-8. doi:10.1145/1240624.1240775

Scharnhorst, Andrea, Richard P. Smiraglia, Christophe Guéret and Alkim Almila Akdag Salah. 2016. "Knowledge Maps of the UDC: Uses and Use Cases." *Knowledge Organization* 43: 641-54. doi:10.5771/0943-7444-2016-8-641

Su, Hai and Zuhua Jiang. 2007. "Construction Method of Knowledge Map based on Design Process." *Chinese Journal of Mechanical Engineering* 20: 98-104.