Journal of Geovisualization and Spatial Analysis (2020) 4: 24
https://doi.org/10.1007/s41651-020-00067-2

Volunteered Geographic Information Research in the First Decade: Visualizing and Analyzing the Author Connectedness of Selected Journal Articles in GIScience

Yingwei Yan [1,2,3] · Dawei Ma [4] · Wei Huang [5,6] · Chen-Chieh Feng [2] · Hongchao Fan [7] · Yingbin Deng [1,3] · Jianhui Xu [1,3]

Accepted: 20 October 2020 / Published online: 27 October 2020

Abstract

Volunteered geographic information (VGI) has been widely explored by researchers for decision support in various application domains because the data are cost-effective to collect and their richness in volume and spatiotemporal coverage is unrivaled against traditional data sources. This study visualizes and analyzes a network of the authors of selected journal articles in GIScience about the first decade of VGI research. It uses the number of citations, one local network centrality measures (i.e., degree), and three global network centrality measures (i.e., closeness centrality, betweenness centrality, and eigenvector centrality) for quantifying the author importance. A new rule-based weighting method has also been developed for taking into account author sequences when computing the global centrality measures. Results show that the connectedness of the European researchers is strong, and Europe and North America have the highest numbers of prominent VGI researchers. Closeness among researchers does not seem to contribute heavily to the increase in citations. Rather, the number of direct connections in the network, the authors' control over the network, and the quality of research connections is more important. European and North American authors as a whole play a leading role in the VGI research, but on average (per author influence) are only outstanding in terms of the citation numbers and have relatively more control over the network. Lastly, this study has revealed the relatively more diverse VGI research topics investigated over a longer time span in North America and Europe compared with other regions of the globe, highlighting the major problems that have been studied across the VGI research network.

Keywords Research network · Volunteered geographic information · Citation · Centrality · Rule-based weighting

✉ Yingbin Deng
yingbin@gdas.ac.cn

Yingwei Yan
yanyingwei@u.nus.edu

Dawei Ma
dawei.ma@whu.edu.cn

Wei Huang
huangweibuct@gmail.com

Chen-Chieh Feng
geofcc@nus.edu.sg

Hongchao Fan
hongchao.fan@ntnu.no

Jianhui Xu
xujianhui306@gdas.ac.cn

[1] Key Lab of Guangdong for Utilization of Remote Sensing and Geographical Information System, Guangdong Open Laboratory of Geospatial Information Technology and Application, Guangzhou Institute of Geography, Guangzhou, China

[2] Department of Geography, National University of Singapore, Singapore, Singapore

[3] Southern Marine Science and Engineering Guangdong Laboratory, Guangzhou, China

[4] School of Remote Sensing and Information Engineering, Wuhan University, Wuhan, China

[5] System Planning Branch, Ministry of Transportation Ontario, Toronto, Canada

[6] Department of Civil Engineering, Ryerson University, Toronto, Canada

[7] Department of Civil and Environmental Engineering, Norwegian University of Science and Technology, Trondheim, Norway

Introduction

Contemporary science in general and GIScience in particular can be described as a dynamic, complex, and constantly evolving multiple-scale network of scientists, institutions, and ideas (Fortunato et al. 2018; Sun and Manson 2011). It has been recognized that one of the dominant mechanisms of facilitating scientific advances is research collaboration (Sun and Rahwan 2017; Wuchty et al. 2007). Research collaborations increase the productivity of researchers and accelerate scientific progress. Scientific publication data has been widely used to explore the patterns and trends of research collaborations (Sun and Rahwan 2017). Through joint work, individual researchers compose research networks which are amenable to scientometric social network visualization and analysis (SNVA) (Kim and Diesner 2015; Sun and Rahwan 2017; Sun and Manson 2011). SNVA has been adopted to visually and mathematically investigate how social system structures and evolutions are defined by relationships among its elements (e.g., people and organizations) which develop and grow with the intertwined systems of social networks (Andris 2016; Rogers 1987; Sun and Manson 2011).

One of the recent studies that explored the research network of scientific publications was by Chuan et al. (2018), who proposed a new metric for edge (link) prediction in research networks, i.e., predicting potential interactions among network elements, based on content similarity. In addition, Köseoglu et al. (2018) and Sun and Rahwan (2017) visualized and examined the authorship trends and explored the structures of scientific collaborations based on network centrality metrics in lodging studies and transportation research, respectively. Xie et al. (2016) proposed a geometric graph to model research networks, the connection mechanism that expresses the effects of the homophily of authors and scholarly influences, and the collaborations at the level of research teams rather than authors. Moreover, Hu et al. (2019) visualized and analyzed the structures of cited and uncited research communities in four disciplines (i.e., chemistry organic, engineering environmental, economics, and management) and three countries (i.e., the USA, the UK, and People's Republic of China) based on co-authorship networks. One more interesting work was from Oliveira et al. (2017), where a Bayesian inferential approach was developed to measure the reliability of a research network, i.e., the probability of this network to remain connected, robust, and functionating, with emphasis on researchers (nodes).

SNVA can be considered a component of the science of science (SciSci), a field evolved from scientometrics, quantitatively examines the interactions among scientific agents (scientists, institutions, and ideas) across diverse spatial and temporal scales (Fortunato et al. 2018; Garfield 2009). For instance, some studies used large scholar datasets to explore the development of an academic field (Sun and Yin 2017), understand academic collaborations (Sun and Rahwan 2017; Sun and Manson 2011), or discover the impact of scientific work (Thelwall 2016; Wang et al. 2013). The emergence of SciSci has been driven by two main factors. The first factor is data availability (e.g., Web of Science, Scopus, and Google Scholar) pertinent to scientists from all fields and to their research output across the globe. The second factor is the collaborations among physical, social, and computational scientists, through which powerful (big) data processing, analysis, visualization tools, and models have been developed to uncover the mechanism underlying sciences and its institutions and workforce (Fortunato et al. 2018). In GIScience, there are also studies that have quantitatively analyzed certain research features such as citations, authorship, and publication patterns. For example, Biljecki (2016) analyzed 12,346 articles from 20 GIScience journals to extract patterns and trends; Duckham (2015) identified the expertise that GIScientists have in common based on keywords and citations; Wei et al. (2015) discovered and benchmarked the most important and highly cited articles published between 2003 and 2012; Sun and Manson (2011) examined the research networks and scientific collaborations. These studies have shed light on the Science of Science in the GIScience field. Moreover, there are studies using quantitative approaches to analyze research topics in GISciecne. For instance, Steiger et al. (2015) published a systematic literature review on spatiotemporal analyses of Twitter data in GIScience. Yan et al. (2020) performed a systematic review on volunteered geographic information (VGI) research topics through Latent Dirichlet allocation. An in-depth understanding of a scientific field through approaches of the SciSci can be beneficial for effective science funding allocations (Fortunato et al. 2018) as well as for high-quality education about the field.

Scientific publications in the field of VGI in particular have been booming in recent years. The term VGI was coined in 2007 by Goodchild (2007) and it has become one of the most important research topics in GIScience (Yan et al. 2020). VGI such as OpenStreetMap (OSM) and geotagged social media data can be an important source of understanding of the surface of the Earth (Goodchild 2007; Yan et al. 2017). The creators of VGI establish virtual networks to work on a common task (or subtasks) in either a synchronous or an asynchronous manner. They share their understanding of a common situation, shape contexts, and convey cognition through contextual knowledge of a place. VGI phenomenon thereby defies the traditional asymmetric power structure of geospatial information production and consumption, i.e., a minority of authorized data producers versus a majority of passive data consumers. On VGI platforms, geospatial data consumers are empowered to produce data and vice versa. The traditional division between data consumers and producers blurs (Mooney and Corcoran 2012). Indeed, the "producers" may have knowledge that is unknown to experts; local people in a sense may themselves count as experts in their own local or indigenous knowledge (Cinnamon and Schuurman 2013; Quinn and Yapa 2016).

The rapid development of VGI is attributed to Web 2.0 technologies, which favor participation and collaboration for the creation of common goods over the Internet (Goodchild 2007; Hachmann et al. 2018). The Internet in the Web 2.0 era enables the formation of a cyberspace of radical inclusion that transforms indirectly related physical communities into directly connected virtual communities. It creates platforms with techno-libertarian and egalitarian as the norms for open and pervasive collaborations of intelligence that promote digital democracy. Among the key principles of Web 2.0, utilizing collective intelligence is a key to sustaining VGI platform constructions (O'Reilly 2007). This principle encourages cyber-collectivism for the formation of Web 2.0 cyberspaces that offer opportunities for achieving higher productivity and greater innovations. In the Web 2.0 era, information technologies are more socially intertwined and new forms of social interactions within information networks are developed (Castells 2000). As such, Web 2.0 has enabled the general public to generate information and interact with one another on an unprecedented scale and in a real-time manner (Elwood et al. 2012). By contributing their collective intelligence, the general public is involved in GIS democracy in a true sense in the Web 2.0 era (Goodchild 2007).

As mentioned above, Yan et al. (2020) have recently published an article that reviews the decade-long research on VGI retrieved from 24 international refereed journals in the GIScience community. Their study has extracted 50 specific VGI research topics which have been subsequently clustered into three overarching themes including VGI contributions and contributors, main fields applying VGI, and conceptions and envisionings. The review has revealed the progress, patterns, and trends in the first decade of the VGI research. It has also proposed an agenda for future research endeavors. However, according to the review and to the best of our knowledge, no empirical research has examined the structure of the research network in the VGI research community, let alone providing insights into the collaboration mechanisms underlying the creativity and major genesis of VGI research discoveries. This study aims to build and visualize a research network of the VGI research from selected journal articles published during the first decade since the coining of VGI and investigate its patterns and structures using scientometric SNVA. This study uses four indicators including the number of citations, one local network centrality measures (i.e., degree), and three global network centrality measures (i.e., closeness centrality, betweenness centrality, and eigenvector centrality) to quantify the author importance across the research network. A new rule-based weighting method has also been developed for taking into account author sequences when computing the three global centrality measures.

Materials and Methods

VGI Research Articles

Following Yan et al. (2020), the dataset used to build the research network in this study includes the 346 articles published during the first decade of the VGI research since Goodchild (2007) coined the term (i.e., between 20 November 2007 and 20 November 2017). These articles are retrieved from 24 journals (based on the keyword "VGI") that are indexed by the Science Citation Index (SCI), Science Citation Index Expanded (SCIE), Social Sciences Citation Index (SSCI), and Emerging Sources Citation Index (ESCI) (Table 1). The four indices are among the core collections of Web of Science according to Clarivate Analytics (http://mjl.clarivate.com/). The 346 articles involve 326 research articles and 20 review articles in which VGI is the main topic of investigation and discussion or at least is used as a source of geographic data. Figure 1 illustrates the temporal distribution of the articles for each journal.

Research Network and Topics

Since scientific collaborations are not unidirectional, an undirected research network *CN* (*N*, *E*) (*N* is the set of nodes (authors) and *E* is the set of edges (connections)) has to be adopted to build the research network for the 346 articles before performing SNVA. When building the research network, this study treats the journal articles for the 10 years as a whole rather than separates them into different temporal slots. This is in part for consistency with Yan et al. (2020) and is also because the number of articles generally kept increasing during 2007 and 2017, no other special temporal variation is observed, and the dataset (i.e., the 346 articles) is not big enough for highly meaningful data split, especially for the first 5 years (Fig. 1). To build a fully connected research network of the selected journal article, this study treats that all the articles are directly or indirectly related to Goodchild (2007) and thus the network is centered with the node that represents Michael F. Goodchild who coined the term VGI. Specifically, for all authors of the 346 articles, an edge is initiated to link each of them with Michael F. Goodchild, representing an indirect connection. An edge's weight increases if there are actual collaborations (i.e., direct collaborations with co-authored articles) between the author and Michael F. Goodchild.

Gephi (https://gephi.org/) is used for network visualization. The nodes are colored based on the locations of the authors' affiliations. However, an author may have multiple affiliations. Therefore, this study employs a set of rules as follows.

Table 1 GIS journals included in this research

Journal	Number of articles
Transactions in GIS (TGIS)	59
GeoJournal (GeoJ)	39
International Journal of Geographical Information Science (IJGIS)	38
ISPRS International Journal of Geo-Information (ISPRS IJGI)	38
Cartography and Geographic Information Science (CGIS)	28
International Journal of Digital Earth (IJDE)	24
Computers, Environment and Urban Systems (CEUS)	23
Annals of the American Association of Geographers (AAG)	13
The Cartographic Journal (TCJ)	11
Cartographica: The International Journal for Geographic Information and Geovisualization (CIJGIG)	10
Journal of Location Based Services (JLBS)	10
GeoInformatica (GeoI)	8
Environment and Planning A: Economy and Space (EPA)	7
Environment and Planning B: Urban Analytics and City Science (EPB)	7
International Journal of Applied Earth Observation and Geoinformation (IJAEOG)	7
Journal of Spatial Science (JSS)	7
The Professional Geographer (TPG)	6
Computers & Geosciences (CG)	3
Applied Spatial Analysis and Policy (ASAP)	2
Geographical Analysis (GA)	2
Spatial Cognition & Computation (SCC)	1
Photogrammetrie, Fernerkundung, Geoinformation (PFG)	1
Sensors (MDPI) (S (MDPI))	1
IEEE Transactions on Geoscience and Remote Sensing (IEEE TGRS)	1

- If an author has more than one affiliation, then the one with which the author has published more articles is used for coloring;
- If an author has published an equal number of articles under each of his or her affiliations, then the latest one is used for coloring.

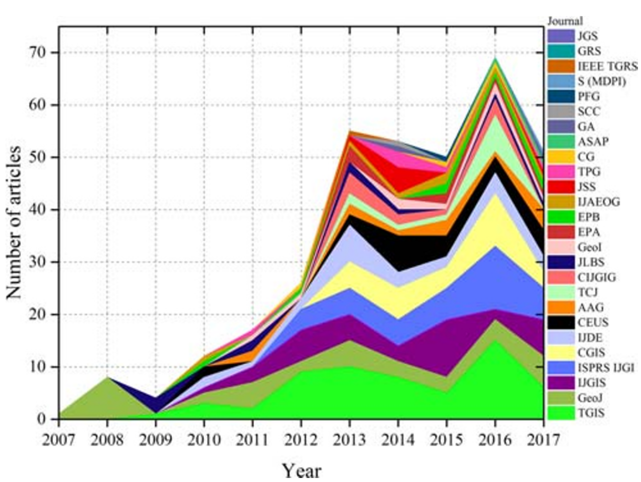

Fig. 1 Number of articles for each journal during 2007 and 2017

The network built is then analyzed using one local and three global measures to reveal the node (author) importance of this network. The notions used in these measures include the following: nodes are denoted as i, j, and an edge linking node i and node j is denoted as e_{ij}; a neighbor j of node i is denoted as Nb_{ij}; the set of authors of an article p is denoted as Au^p; the number of authors in an article p is denoted as $|\mathrm{Au}^p|$; and the sequence of author i in article p is denoted as $\mathrm{Au}S_i^p$ $(\mathrm{Au}S_i^p \geq 1)$.

The local measure employed is degree D_{i}, which is a count of how many neighbors (or co-authors) node i have:

$$D_i = \sum_{j \in N} \mathrm{Nb}_{ij}, \tag{1}$$

where Nb_{ij} equals to 1 if there is an edge directly linking i and j and 0 otherwise.

The three global measures employed, which provide the indications of author influence across the network, are closeness centrality (Bavelas 1950; Sabidussi 1966), betweenness centrality (Freeman 1977), and eigenvector centrality (Bonacich 1972).

Closeness centrality of a node is computed based on the network distance between the node and each other node in a graph. The higher the closeness centrality score is, the lower the node's total distance to all other nodes is, meaning that the node is closer to all other nodes. It can be regarded as a measure of how long it will take to spread information from a node to all other nodes sequentially. For node i, its closeness centrality (CC_i) can be calculated as the reciprocal of the sum of the length of the shortest paths between the node and all other nodes in the graph:

$$CC_i = \frac{1}{\sum_{i \neq j} d_{(ij)}}, \tag{2}$$

where $d_{(i\,j)}$ is the distance between node i and j.

Betweenness centrality of a node is the shortest path-based measure, which quantifies the number of times a node acting as a bridge along the shortest path between two other nodes. Therefore, a node with a higher betweenness centrality score has more control over the network because more information passes through that node. For node i, its betweenness centrality (BC_i) can be calculated as:

$$BC_i = \sum_{i \neq j, i \neq k, j \neq k} \frac{\delta_{jk}(i)}{\delta_{jk}}, \tag{3}$$

where δ_{jk} denotes the total number of shortest paths between node j and k, and δ_{jk} (i) denotes the number of shortest paths passing through node i $(i \neq j,\ i \neq k)$. Note that there may exist multiple shortest paths between a pair of nodes (j, k).

Eigenvector centrality is a measure related to prestige, which is a more sophisticated view of centrality. The idea is that the prestige of node i is related to the prestige of its neighbors. A node with few links may have a very high eigenvector centrality if those few links were to very well-linked others. A high eigenvector score means that a node is linked to many high score nodes. For node i, its eigenvector centrality (EC_i) is computed based on the assumption that its centrality is proportional to the sum of the centrality of its neighbors:

$$Ax = \lambda x, \tag{4}$$

where A is the adjacency matrix of the graph CN with eigenvalue λ. Based on the Perron–Frobenius theorem, there is a positive and unique solution if λ is the greatest eigenvalue associated with the eigenvector of A (Newman 2010).

For accounting for the path lengths for computing the three global centrality measures, each connection can have a fixed weight (i.e., $w = 1$) regardless of the number of authors in an article and regardless of the author sequences in an article. However, this may lead to bias in the calculation results of the centralities. To address the inflation by the number of authors, an adjusted weight parameter $\dot{W}$ introduced by Newman (2001) is adopted:

$$\dot{W}_{ij} = \sum_{p \in P} \frac{1}{|Au^p|-1} \sigma_i^p \sigma_j^p, \quad (i \in Au^p \text{ and } j \in Au^p), \tag{5}$$

where P is the article set (346 articles). σ_i^p or σ_j^p equals to 1 if i or j is an author of article p and 0 otherwise. As such, $\dot{W}_{ij}$ represents the strength of the collaboration (if any) between authors i and j. Each collaboration between two authors in an article contributes $\dot{w}_{ij} = \frac{1}{|Au^p|-1}$ units to the total weight $\dot{W}_{ij}$.

Furthermore, to account for the author sequences in an article, we introduce a readjusted weight parameter $\ddot{W}$ calculated based on the following rules:

- If a collaboration is between the first author i and a non-first author j of article p, then the collaboration contributes the following readjusted units to the total weight $\ddot{W}_{ij}$:

$$\ddot{w}_{ij} = \frac{\frac{1}{|Au^p|-1}}{AuS_j^p-1} = \frac{1}{(|Au^p|-1)\left(AuS_j^p-1\right)}, \tag{6}$$

$$\left(i \in Au^p, j \in Au^p \text{ and } AuS_i^p < AuS_j^p\right),$$

- If a collaboration is between a non-first author i and a non-first author j of article p, then the collaboration contributes the following readjusted units to the total weight $\ddot{W}_{ij}$:

$$\ddot{w}_{ij} = \frac{\frac{1}{|Au^p|-1}}{2(|Au^p|-1)} = \frac{1}{2(|Au^p|-1)^2}, \quad (i \in Au^p, j \in Au^p), \tag{7}$$

According to the above rules, for articles with two authors, the w, $\dot{w}_{ij}$, and $\ddot{w}_{ij}$ all equal to 1. Additionally, the w, $\dot{w}_{ij}$, and $\ddot{w}_{ij}$ of the edge initiated (representing an indirect connection) to link an author with Michael F. Goodchild all equal to 1. Since edges with a stronger connection have a shorter distance, following Sun and Rahwan (2017), the weighted version of the global centrality measures is computed based on the reciprocal of the weight parameters as weighted edge cost. The network analysis is performed using NetworkX (https://networkx.github.io/) which is a Python package specifically for studying complex networks.

Apart from these network-based measures, the number of citations of each paper is also used to quantify author importance. Google Scholar citations as of 12 May 2019 are used in this study. The total number of citations of author i is denoted as:

$$C_i = \sum_{p \in P} \left(ct_p \times \gamma_i^p\right), (i \in Au^p), \tag{8}$$

where ct_p is the number of citations of paper p and γ_i^p equals to 1 if author i is an author of article p and 0 otherwise.

Overall, eight measures are selected for quantifying the authors' importance, which are summarized in Table 2.

Lastly, the spatiotemporal distributions of the 50 VGI research topics derived by Yan et al. (2020) from the 346 journal articles associated with the research network are visualized. In general, this work can be considered as an extension of the review work by Yan et al. (2020).

Results

Descriptive Statistics and the Network Visualization

A total of 1106 authors are found in the 346 articles. As an author may have published more than one article, we removed duplications, which resulted in 765 unique authors that were later used to create the nodes of the research network. Between these nodes, 2651 edges are established according to the network construction rules introduced in the "Research network and topics" section. It is observed that most of the articles have two to three authors (Fig. 2a). In addition, the unique authors are found to be affiliated with 804 institutions, with the majority of these institutions located in Europe (369 or 45.9%) and North America (258 or 32.1%) (Fig. 2b).

A fully connected research network of the VGI research is shown in Fig. 3. For the six regions in Fig. 3 (i.e., Europe, North America, Asia, Oceania, South America, and Africa), it is observed that the connectedness within Europe seems stronger than the connectedness within North America. For instance, a large cluster of European connections can be seen at the bottom of the graph. In comparison, the connections within North America tend to scatter across the graph, and the size of the clusters is relatively smaller. In addition, cross-regional connections can be observed, most of which are, however, among Europe, North America, Asia, and Oceania. Authors from South America and Africa have also connected with those from the other four regions sporadically.

Research Network Metrics and the Related Topics

The rankings of authors based on the total number of citations C_i and degree D_i are shown in Table 3. Except for Michael F. Goodchild, the top three highly cited authors are Mordechai (Muki) Haklay (University College London, the United Kingdom (UK)), Andrew T. Crooks (George Mason University, the United States of America (USA)), and Alexander Zipf (Heidelberg University, Germany). Two out of the three highly cited authors are affiliated with European Universities and one is affiliated with a North American university. The three authors with the highest degrees are Alexander Zipf, Anthony Stefanidis (George Mason University, the USA), and Andrew T. Crooks. Two out of these three are affiliated with a North American university, and one is affiliated with a European university.

The rankings of authors based on adjusted closeness centrality (normalized) $\dot{CC}_i$ and readjusted closeness centrality (normalized) $\ddot{CC}_i$ are shown in Table 4. Except for Michael F. Goodchild, the three authors with the highest $\dot{CC}_i$ are Alexander Zipf, Peter Mooney (National University of Ireland, Maynooth, Ireland), and Jamal Jokar Arsanjani (Heidelberg University, Germany). All three authors are affiliated with European universities. Except for Michael F. Goodchild, the three authors with the highest $\ddot{CC}_i$ are Alexander Zipf, Peter Mooney, and Julian Hagenauer (Heidelberg University, Germany). Again, all three authors are affiliated with European universities.

The rankings of authors based on adjusted betweenness centrality (normalized) $\dot{BC}_i$ and readjusted betweenness centrality (normalized) $\ddot{BC}_i$ are shown in Table 5. Except for Michael F. Goodchild, the three authors with the highest $\dot{BC}_i$ are Alexander Zipf, Peter Mooney, and Jamal Jokar Arsanjani, and the three authors with the highest $\ddot{BC}_i$ are Alexander Zipf, Jamal Jokar Arsanjani, and Peter Mooney. For both rankings, all three authors are affiliated with European universities.

Table 2 The node (author) importance measures

Measure	Notation
Degree	D_i
Total number of citations	C_i
Adjusted closeness centrality (normalized)	$\dot{CC}_i$
Readjusted closeness centrality (normalized)	$\ddot{CC}_i$
Adjusted betweenness centrality (normalized)	$\dot{BC}_i$
Readjusted betweenness centrality (normalized)	$\ddot{BC}_i$
Adjusted eigenvector centrality (normalized)	$\dot{EC}_i$
Readjusted eigenvector centrality (normalized)	$\ddot{EC}_i$

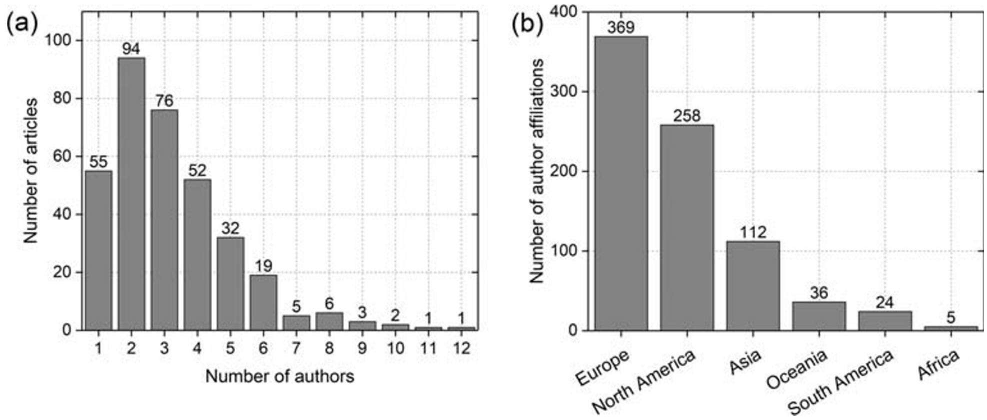

Fig. 2 **a** The number of authors across all the articles. **b** The number of affiliations associated with the articles divided by region. Note: Oceania includes only Australia and New Zealand

The ranking of authors based on adjusted eigenvector centrality (normalized) $\dot{\mathrm{EC}}_i$ and readjusted eigenvector centrality (normalized) $\ddot{EC}_i$ is shown in Table 6. Except for Michael F. Goodchild, the three authors with the highest $\dot{\mathrm{EC}}_i$ are Daniel Sui (The Ohio State University, the USA), Craig M. Dalton (Bloomsburg University of Pennsylvania and Hofstra University, the USA), and Mordechai (Muki) Haklay. Two of these three are affiliated with North American universities, and one is affiliated with a European university. Except for Michael F. Goodchild, the three authors with the highest $\ddot{\mathrm{EC}}_i$ are Daniel Sui, Craig M. Dalton, and Sterling D. Quinn (Central Washington University and The Pennsylvania State University, the USA). All three authors are affiliated with North American universities.

Table 7 shows the pairwise Pearson correlation matrix for the node importance measures that indicate the importance of authors in the VGI research network. The result suggests that D_i and C_i are not highly correlated with the closeness centrality measures ($\dot{\mathrm{CC}}_i$ and $\ddot{\mathrm{CC}}_i$). In fact, the closeness centrality measures are not highly correlated with any other centrality measures. Furthermore, Fig. 4 shows the sum of each node importance measure for each region and the sum of each node importance measure for each region divided by the number of authors from each region, i.e., the average values of the

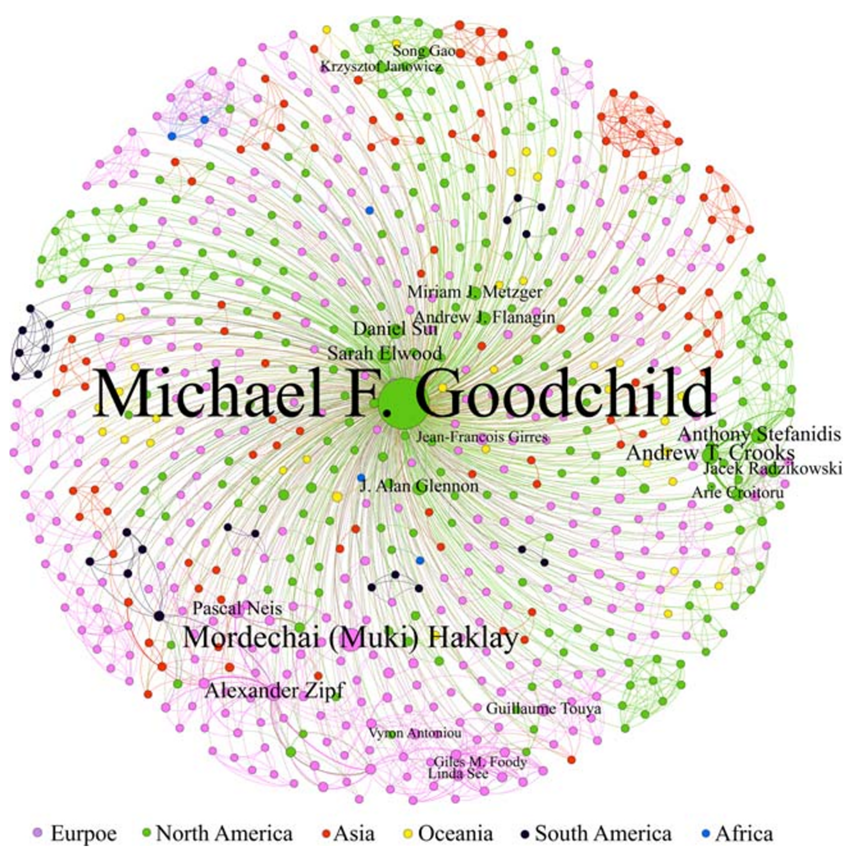

Fig. 3 A fully connected research network of the first decade of the VGI research. The node sizes are proportional to the total number of citations of the individual authors. The top 20 authors ranked by their total number of citations are labeled with their names and the name sizes are proportional to the total number of citations of the individual authors. The thickness of an edge between each pair of authors is proportional to the total number of connections between the two authors. Note: Oceania includes only Australia and New Zealand

Table 3 Ranking of authors based on the total number of citations and degree D_i. Note: Since indirect collaboration edges between Michael F. Goodchild and all the other authors are created in this study (Research network and topics), his degree is 764 (including indirect collaboration edges) and 12 (excluding indirect collaboration edges)

	Total number of citations		Degree	
	Author name	C_i	Author name	D_i
1	Michael F. Goodchild	5999	Alexander Zipf	36
2	Mordechai (Muki) Haklay	2099	Anthony Stefanidis	26
3	Andrew T. Crooks	1200	Andrew T. Crooks	25
4	Alexander Zipf	1196	Song Gao	23
5	Anthony Stefanidis	1029	Giles M. Foody	22
6	Sarah Elwood	957	Arie Croitoru	21
7	Daniel Sui	927	Krzysztof Janowicz	21
8	Pascal Neis	860	Linda See	21
9	J. Alan Glennon	764	Hongchao Fan	19
10	Andrew J. Flanagin	731	Jamal Jokar Arsanjani	19
11	Miriam J. Metzger	731	Steffen Fritz	19
11	Jacek Radzikowski	727	Dieter Pfoser	18
13	Arie Croitoru	633	Marco Minghini	17
14	Guillaume Touya	568	Vyron Antoniou	16
15	Song Gao	502	Peter Mooney	16
16	Jean-François Girres	492	Li Gong	16
17	Krzysztof Janowicz	444	Ana-Maria Olteanu-Raimond	16
18	Linda See	424	Jeroen Verplanke	15
19	Giles M. Foody	391	Michael K. McCall	15
20	Vyron Antoniou	378	Gloria Bordogna	14

Table 4 Ranking of authors based on adjusted closeness centrality (normalized) and readjusted closeness centrality (normalized)

	Adjusted closeness centrality (normalized)		Readjusted closeness centrality (normalized)	
	Author name	$\dot{CC}_i$	Author name	$\ddot{CC}_i$
1	Michael F. Goodchild	0.94321	Michael F. Goodchild	0.95023
2	Alexander Zipf	0.54395	Alexander Zipf	0.58244
3	Peter Mooney	0.54256	Peter Mooney	0.58185
4	Jamal Jokar Arsanjani	0.54141	Julian Hagenauer	0.58176
5	Vyron Antoniou	0.54076	Andriani Skopeliti	0.58153
6	Julian Hagenauer	0.53994	Jamal Jokar Arsanjani	0.58144
7	Ana-Maria Olteanu-Raimond	0.53947	Marco Minghini	0.58140
8	Steffen Fritz	0.53811	Ana-Maria Olteanu-Raimond	0.58138
9	Jacinto Estima	0.53793	Melanie Eckle	0.58136
10	Cidália Costa Fonte	0.53766	Chiao-Ling Kuo	0.58136
11	Marco Minghini	0.53749	Steffen Fritz	0.58124
11	Flavio Lupia	0.53694	Jacinto Estima	0.58115
13	Andriani Skopeliti	0.53684	Linda See	0.58115
14	Mari Laakso	0.53684	Cidália Costa Fonte	0.58113
15	Linda See	0.53660	Nuttha Sirilertworakul	0.58110
16	Yingwei Yan	0.53629	Flavio Lupia	0.58107
17	Melanie Eckle	0.53607	Vyron Antoniou	0.58106
18	Chiao-Ling Kuo	0.53607	Paul Harris	0.58097
19	Alexis Comber	0.53411	Hongchao Fan	0.58084
20	Mohamed Bakillah	0.53411	Mari Laakso	0.58084

Table 5 Ranking of authors based on adjusted betweenness centrality (normalized) and readjusted betweenness centrality (normalized)

	Adjusted betweenness centrality (normalized)		Readjusted betweenness centrality (normalized)	
	Author name	$\dot{BC}_i$	Author name	$\ddot{BC}_i$
1	Michael F. Goodchild	0.95941	Michael F. Goodchild	0.95173
2	Alexander Zipf	0.01724	Alexander Zipf	0.02503
3	Peter Mooney	0.00782	Jamal Jokar Arsanjani	0.01033
4	Jamal Jokar Arsanjani	0.00681	Peter Mooney	0.00927
5	Ana-Maria Olteanu-Raimond	0.00437	Pascal Neis	0.00799
6	Giles M. Foody	0.00383	Giles M. Foody	0.00735
7	Hansi Senaratne	0.00352	Mordechai (Muki) Haklay	0.00718
8	Vyron Antoniou	0.00345	Qing Fu	0.00709
9	John M. Davis	0.00341	Amin Mobasheri	0.00645
10	Julian Hagenauer	0.00337	Julian Hagenauer	0.00556
11	Hongchao Fan	0.00323	Ana-Maria Olteanu-Raimond	0.00532
11	Glen Hart	0.00306	Edward J. Malecki	0.00523
13	Tobias Törnros	0.00295	John M. Davis	0.00510
14	Amin Mobasheri	0.00260	Bernhard Höfle	0.00501
15	Ahmed Loai Ali	0.00239	Hongchao Fan	0.00431
16	Bisheng Yang	0.00229	Steffen Fritz	0.00403
17	João Porto de Albuquerque	0.00225	Glen Hart	0.00391
18	James Baginski	0.00217	Marco Minghini	0.00361
19	Edward J. Malecki	0.00217	Francesco Tonini	0.00348
20	Thomas Koukoletsos	0.00215	Hansi Senaratne	0.00343

Table 6 Ranking of authors based on adjusted eigenvector centrality (normalized) and readjusted eigenvector centrality (normalized)

	Adjusted eigenvector centrality (normalized)		Readjusted eigenvector centrality (normalized)	
	Author name	$\dot{EC}_i$	Author name	$\ddot{EC}_i$
1	Michael F. Goodchild	0.69953	Michael F. Goodchild	0.70373
2	Daniel Sui	0.10052	Daniel Sui	0.09351
3	Craig M. Dalton	0.08789	Craig M. Dalton	0.08977
4	Mordechai (Muki) Haklay	0.06869	Sterling D. Quinn	0.06890
5	Sterling D. Quinn	0.06743	Mordechai (Muki) Haklay	0.06849
6	Billy Tusker Haworth	0.06592	Billy Tusker Haworth	0.06733
7	Agnieszka Leszczynski	0.06592	Agnieszka Leszczynski	0.06733
8	Wen Lin	0.06592	Wen Lin	0.06733
9	Matthew James Kelley	0.06592	Matthew James Kelley	0.06733
10	Donald G. Janelle	0.06592	Donald G. Janelle	0.06733
11	David L. Tulloch	0.06592	David L. Tulloch	0.06733
11	Sarah Elwood	0.05726	Sarah Elwood	0.05762
13	Pascal Neis	0.05197	Pascal Neis	0.04852
14	Krzysztof Janowicz	0.04965	Krzysztof Janowicz	0.04669
15	Giles M. Foody	0.04959	Marcus Goetz	0.04654
16	Wenwen Li	0.04730	Wenwen Li	0.04622
17	Amin Mobasheri	0.04690	Giles M. Foody	0.04599
18	Marcus Goetz	0.04614	Peter A. Johnson	0.04588
19	Peter A. Johnson	0.04545	Amin Mobasheri	0.04553
20	Claire Ellul	0.04540	Claire Ellul	0.04535

Table 7 Pairwise Pearson correlation table for the node (author) importance measures. Values for the highly collinear pairs (i.e., $r > 0.75$) are highlighted in italics

	D_i	C_i	$\dot{C}C_i$	$\ddot{C}C_i$	$\dot{B}C_i$	$\ddot{B}C_i$	$\dot{E}C_i$	$\ddot{E}C_i$
D_i	1	*0.849***	0.295**	0.307**	*0.832***	*0.836***	*0.856***	*0.845***
C_i		1	0.411**	0.349**	*0.993***	*0.993***	*0.948***	*0.946***
$\dot{C}C_i$			1	*0.919***	0.373**	0.375**	0.144**	0.137**
$\ddot{C}C_i$				1	0.310**	0.315**	0.136**	0.125**
$\dot{B}C_i$					1	*1.000***	*0.958***	*0.958***
$\ddot{B}C_i$						1	*0.959***	*0.959***
$\dot{E}C_i$							1	*0.999***
$\ddot{E}C_i$								1

**Correlation is significant at the 0.01 level (2-tailed)

measures. It is observed that Europe and North America occupy the top two layers of the sums, each of which accounts for a percentage of the sum of each measure that is much greater than other regions do. Regarding the average values, except for the average values of the citation and the betweenness measures which are relatively greater in Europe and North America, the average values of the other measures are more or less evenly distributed across the regions.

Lastly, the spatiotemporal distributions of the VGI research topics associated with the research network are visualized in Fig. 5. It is observed that North America and Europe are associated with the highest diversity of research topics, followed by Oceania, South America, Asia, and South Africa. The USA, Canada, Germany, and the UK are among the leading countries in the VGI research that covers diverse topics such as data quality of OSM, sensor network, and disaster, crisis, emergency, and hazard management. The VGI research activities in North America and Europe also have a longer time span compared with other regions of the globe.

Discussion and Conclusions

Main Research Findings and Interpretations

This study has performed a scientometric SNVA of the first decade of the VGI research based on selected journal articles in GIScience. Recently, Yan et al. (2020) has performed a narrative review of the research articles concerning VGI published during the same period. Based on the same collection of articles, this quantitative scientometric research can be considered as a complement for the qualitative review. This study has used the number of citations, one local social network centrality measure (i.e., degree), and three global social network centrality measures (i.e., closeness centrality, betweenness centrality, and eigenvector centrality) for quantifying the node (author) importance in the network.

To take into account the number of authors in an article when computing the global centrality measures, this study has adopted an established edge weighting approach to derive an

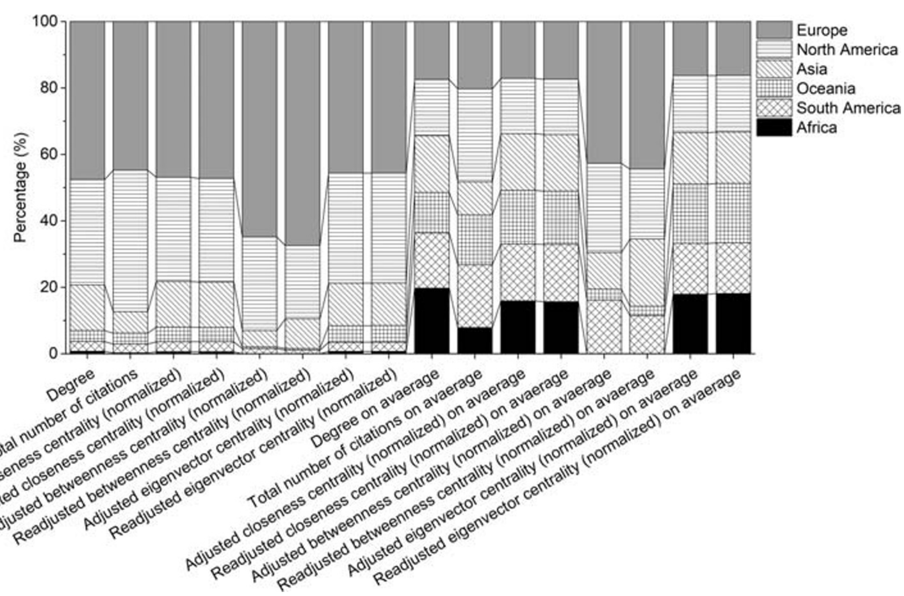

Fig. 4 The sum of each node (author) importance measure for each region and the sum of each node (author) importance measure for each region divided by the number of authors from each region. Note: Oceania includes only Australia and New Zealand

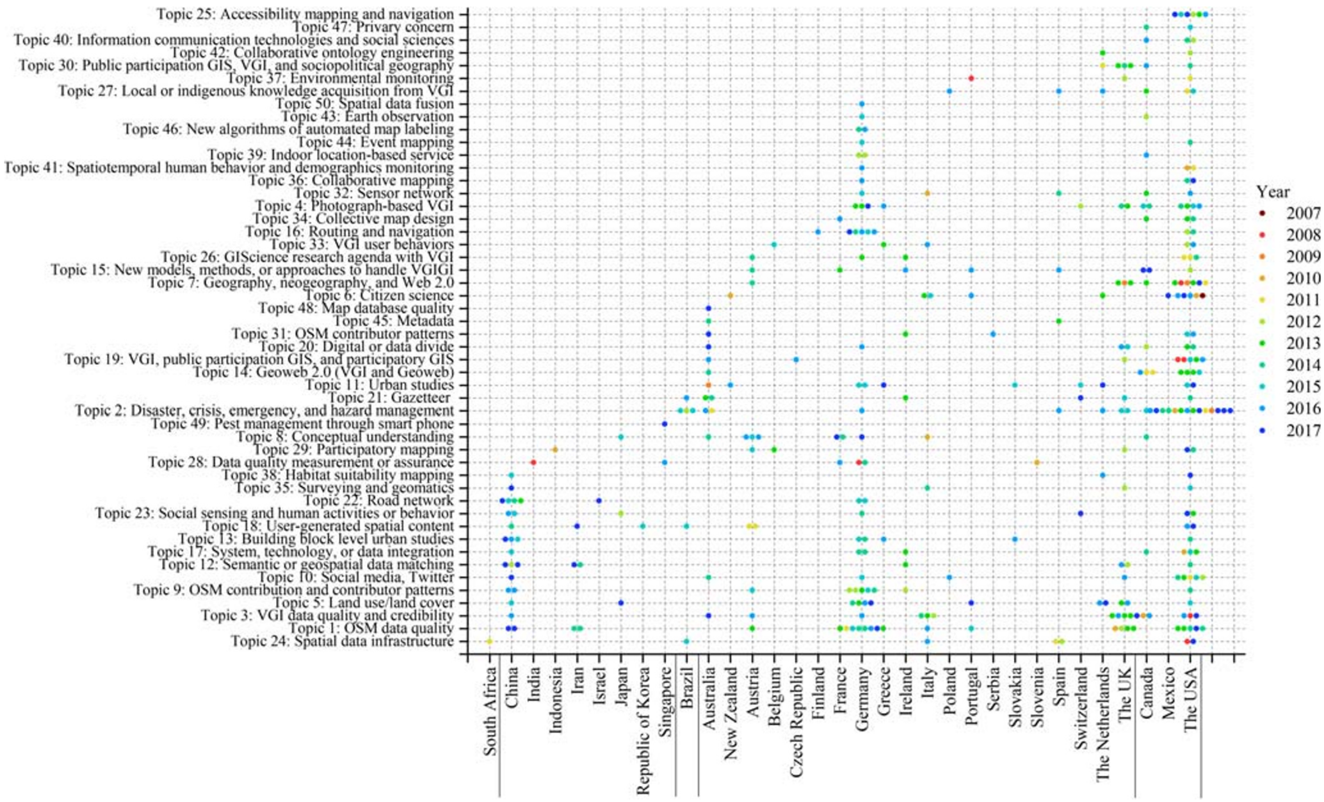

Fig. 5 Three-dimensional scatter plot showing the spatiotemporal distribution of the 50 research topics derived from the 346 journal articles. Note: Each dot in the graph represents a journal article. The countries on the *x*-axis are associated with the first authors; the years in the legend are the publication years of the articles. Multiple articles produced in a single country are arranged horizontally, extending from the central cross point (offset plotting), rather than overlapped to each other

adjusted version of the global centrality measures. In addition, to appropriately consider the author sequences in an article when computing the global centrality measures, this study has developed a new rule-based edge weighting method to derive a readjusted version of the global centrality measures. This weighting method has further reduced the bias in the calculation results of the centralities.

Regarding the main research findings, firstly, although VGI was coined by a researcher from North America (Michael F. Goodchild), European institutions seem to be more actively engaged in the VGI research than the North American counterpart (Fig. 2b). One possible reason is that OSM as the most popular VGI platform was created in the UK in Europe (Yan et al. 2020). It is also found that the connectedness within Europe to be stronger than that within North America according to the research network visualization (Fig. 3). One possible reason for this network pattern is the geographic closeness of the European institutions. In addition, this study has demonstrated that the top researchers measured using the eight-node importance measures are all affiliated with North American and European universities (Tables 3, 4, 5, and 6), confirming the leading role of the two regions in the VGI research. Apart from OSM which was developed in Europe, other diverse and most influential sources of VGI, such as Twitter, Flickr, and Geo-Wiki, were mostly developed in either the USA or Europe. The diverse VGI platforms established in these two regions and the high VGI data accessibility may explain why these two regions are the most active in the VGI research (Yan et al. 2020).

According to the pairwise correlation table (Table 7), the results suggest that the closeness among researchers does not seem to highly contribute to the increase of citations. However, the degree, betweenness centrality, and eigenvector centrality are highly correlated with the citation numbers, suggesting that the number of direct connections in the network (i.e., the number of directly linked neighbors of an author without any intermedia node), the authors' control over the network, and the quality of research connections is more important to increasing citations. Furthermore, this study discovers that the sums of the eight measures are high in Europe and North America (Fig. 4). Europe and North America also have high average values of the citation numbers and betweenness centrality measures, while the average values of the other measures are generally evenly distributed across the remaining four regions (Fig. 4). These findings suggest that European and North American authors as a whole play a leading role in the VGI research, but on average (per author influence) are only outstanding in terms of the citation numbers and have relatively more control over the network.

Lastly, this study has revealed the high diversity of the VGI research topics investigated in North America and Europe throughout the first decade of the VGI research development, highlighting the major problems that have been studied across the VGI research network (Fig. 5). The diverse VGI research topics investigated in North America and Europe and the longer time span of the research activities in North America and Europe further confirm the leading role of the two regions in the VGI research field (Fig. 5).

Research and Practical Implications

Overall, the research outcome of this work would benefit the development of policies, approaches, and tools that have the potential of accelerating the development of VGI science. This study has generated insights into the conditions and structures underlying the creativity in VGI research. It has also provided a quantitative understanding of the major genesis of scientific findings about VGI. Specifically, the authors involved in this study worked individually or collaboratively on the VGI research; therefore, this study provides indicators for person-directed funding, performance-based funding, topic-based funding, and scientist crowdfunding, contributing to the future development of the VGI research field (Fortunato et al. 2018). Additionally, the rule-based weighting method developed in this study for taking into account author sequences when computing the three global centrality measures has implications for researchers to better quantify author importance across a research network. Moreover, this study has pedagogical implications for VGI education through an in-depth understanding of the VGI science community. For example, VGI teaching can be based on the research outcomes of leading researchers from the leading regions with highly diverse VGI research topics (Figs. 3, 4, and 5 and Tables 3, 4, 5, and 6). Lastly, based on the results of this study and the relevant VGI review articles such as the one published by Yan et al. (2020), practitioners would have a direction for the enhancement of VGI platforms, e.g., seeking advice from a particular researcher or a cluster of researchers about how to motivate the user to contribute VGI and how to improve the data quality and credibility of a VGI platform.

Future Works

For future works, it will be necessary to explore the research network together with the research topics shown in Fig. 5. Yan et al. (2020) clustered the topics into three overarching themes including (1) VGI contributions and contributors, (2) main fields applying VGI, and (3) conceptions and envisionings. Further identifying these research topics that attract different degrees of regional or international collaborations and identifying the research topics that lack regional or international collaborations would be beneficial for facilitating the long-term development of this research field. Indeed, Yan et al. (2020) proposed a VGI research agenda about the identified research topics; it would be useful to discuss how researcher connectedness in the field would contribute to the fulfillment of the research agenda. Moreover, for building a fully connected network, this study assumes that every VGI article is related to Michael F. Goodchild (Research network and topics). Doing so automatically builds some relations that may not, in fact, exist. The result may be different if we do not make the assumption, and thus, an improved method is needed to build a fully connected network in order to compute the centrality measures across all the authors. Last but not least, the temporal variation of the network structure is not examined in this study. In fact, by observing Fig. 5, it can be inferred that the patterns of the research network did not vary strongly over the 10 years; i.e., basically, North America and Europe were consistently active in the VGI research throughout the 10 years. It would be interesting to keep tracking the temporal variation of the quantity of VGI articles and then investigate the structural changes of the network over the long-term run. For example, the network of the journal articles published during the first decade since the coining of VGI can be compared with that of the second decade. Indeed, there are many emerging VGI platforms that have been created in non-Western countries; VGI may attract more attention from researchers in non-Western countries, and thus, the research network structure may vary strongly over the long-term run.

Authors' Contributions All authors contributed to the research design, the algorithm design, and the manuscript writing. Yingwei Yan, Dawei Ma, and Wei Huang contributed to the data collection, processing, and analysis.

Funding This work has been supported by the GDAS' Project of Science and Technology Development (grant number: 2020GDASYL-20200103005), the National Natural Science Foundation of China (grant number: 41901330), the Key Special Project for Introduced Talents Team of Southern Marine Science and Engineering Guangdong Laboratory (Guangzhou) (grant number: GML2019ZD0301), the GDAS' Project of Science and Technology Development (grant number: 2019GDASYL-0103004), the National Natural Science Foundation of China (grant number: 41901372), and the GDAS' Project of Science and Technology Development (grant number: 2019GDASYL-0301001).

Data Availability The data used in this work is sourced from Yan et al. (2020) (cited in the manuscript).

Compliance with Ethical Standards

Conflict of Interest The authors declare that they have no conflict of interest.

Code Availability The program used in this study is NetworkX (https://networkx.github.io/) with the codes available in the website.

References

Andris C (2016) Integrating social network data into GISystems. Int J Geogr Inf Sci 30:2009–2031. https://doi.org/10.1080/13658816.2016.1153103

Bavelas A (1950) Communication patterns in task-oriented groups. J Acoust Soc Am 22:725–730. https://doi.org/10.1121/1.1906679

Biljecki F (2016) A scientometric analysis of selected GIScience journals. Int J Geogr Inf Sci 30:1302–1335. https://doi.org/10.1080/13658816.2015.1130831

Bonacich P (1972) Factoring and weighting approaches to status scores and clique identification. J Math Sociol 2:113–120. https://doi.org/10.1080/0022250X.1972.9989806

Castells M (2000) Toward a sociology of the network society. Contemp Sociol 29:693–699. https://doi.org/10.2307/2655234

Chuan PM, Son LH, Ali M, Khang TD, Huong LT, Dey N (2018) Link prediction in co-authorship networks based on hybrid content similarity metric. Appl Intell 48:2470–2486. https://doi.org/10.1007/s10489-017-1086-x

Cinnamon J, Schuurman N (2013) Confronting the data-divide in a time of spatial turns and volunteered geographic information. GeoJournal 78:657–674. https://doi.org/10.1007/s10708-012-9458-6

Duckham M (2015) GI Expertise. Trans GIS 19:499–515. https://doi.org/10.1111/tgis.12166

Elwood S, Goodchild MF, Sui DZ (2012) Researching volunteered geographic information: Spatial Data, geographic research, and new social practice. Ann Assoc Am Geogr 102:571–590. https://doi.org/10.1080/00045608.2011.595657

Fortunato S et al (2018) Science of science. Science 359:eaao0185. https://doi.org/10.1126/science.aao0185

Freeman LC (1977) A set of measures of centrality based on betweenness. Sociometry 40:35–41. https://doi.org/10.2307/3033543

Garfield E (2009) From the science of science to Scientometrics visualizing the history of science with HistCite software. J Informetrics 3:173–179. https://doi.org/10.1016/j.joi.2009.03.009

Goodchild MF (2007) Citizens as sensors: The world of volunteered geography. GeoJournal 69:211–221. https://doi.org/10.1007/s10708-007-9111-y

Hachmann S, Jokar Arsanjani J, Vaz E (2018) Spatial data for slum upgrading: volunteered geographic information and the role of citizen science. Habitat Int 72:18–26. https://doi.org/10.1016/j.habitatint.2017.04.011

Hu Z, Lin A, Willett P (2019) Identification of research communities in cited and uncited publications using a co-authorship network. Scientometrics 118:1–19. https://doi.org/10.1007/s11192-018-2954-9

Kim J, Diesner J (2015) Coauthorship networks: a directed network approach considering the order and number of coauthors. J Assoc Inf Sci Technol 66:2685–2696. https://doi.org/10.1002/asi.23361

Köseoglu MA, Okumus F, Putra ED, Yildiz M, Dogan IC (2018) Authorship trends, collaboration patterns, and co-authorship networks in lodging studies (1990–2016). J Hosp Mark Manag 27:561–582. https://doi.org/10.1080/19368623.2018.1399192

Mooney P, Corcoran P (2012) The annotation process in OpenStreetMap. Trans GIS 16:561–579. https://doi.org/10.1111/j.1467-9671.2012.01306.x

Newman MEJ (2001) Scientific collaboration networks. II. Shortest paths, weighted networks, and centrality. Phys Rev E 64:016132. https://doi.org/10.1103/PhysRevE.64.016132

Newman MEJ (2010) Networks: an introduction. Oxford University Press Inc., New York. https://doi.org/10.1093/acprof:oso/9780199206650.001.0001

O'Reilly T (2007) What is web 2.0: design patterns and business models for the next generation of software. Commun Strateg 65:17–37

Oliveira SC, Cobre J, Ferreira TP (2017) A Bayesian approach for the reliability of scientific co-authorship networks with emphasis on nodes. Soc Networks 48:110–115. https://doi.org/10.1016/j.socnet.2016.06.005

Quinn S, Yapa L (2016) OpenStreetMap and food security: A case study in the city of Philadelphia. Prof Geogr 68:271–280. https://doi.org/10.1080/00330124.2015.1065547

Rogers EM (1987) Progress, problems and prospects for network research: investigating relationships in the age of electronic communication technologies. Soc Networks 9:285–310. https://doi.org/10.1016/0378-8733(87)90001-3

Sabidussi G (1966) The centrality index of a graph. Psychometrika 31:581–603. https://doi.org/10.1007/BF02289527

Steiger E, de Albuquerque JP, Zipf A (2015) An advanced systematic literature review on spatiotemporal analyses of Twitter data. Trans GIS 19:809–834. https://doi.org/10.1111/tgis.12132

Sun S, Manson SM (2011) Social network analysis of the academic GIScience community. Prof Geogr 63:18–33. https://doi.org/10.1080/00330124.2010.533560

Sun L, Rahwan I (2017) Coauthorship network in transportation research. Transp Res A Policy Pract 100:135–151. https://doi.org/10.1016/j.tra.2017.04.011

Sun L, Yin Y (2017) Discovering themes and trends in transportation research using topic modeling. Trans Res C Emerg Technol 77:49–66. https://doi.org/10.1016/j.trc.2017.01.013

Thelwall M (2016) The discretised lognormal and hooked power law distributions for complete citation data: best options for modelling and regression. J Informetrics 10:336–346. https://doi.org/10.1016/j.joi.2015.12.007

Wang D, Song C, Barabási A-L (2013) Quantifying long-term scientific impact. Science 342:127–132. https://doi.org/10.1126/science.1237825

Wei F, Grubesic TH, Bishop BW (2015) Exploring the GIS knowledge domain using CiteSpace. Prof Geogr 67:374–384. https://doi.org/10.1080/00330124.2014.983588

Wuchty S, Jones BF, Uzzi B (2007) The increasing dominance of teams in production of knowledge. Science 316:1036–1039. https://doi.org/10.1126/science.1136099

Xie Z, Ouyang Z, Li J (2016) A geometric graph model for coauthorship networks. J Informetrics 10:299–311. https://doi.org/10.1016/j.joi.2016.02.001

Yan Y, Feng C-C, Chang KT-T (2017) Towards enhancing integrated pest management based on volunteered geographic information. ISPRS Int J Geo Inf 6:224. https://doi.org/10.3390/ijgi6070224

Yan Y, Feng C-C, Huang W, Fan H, Wang Y-C, Zipf A (2020) Volunteered geographic information research in the first decade: a narrative review of selected journal articles in GIScience. Int J Geogr Inf Sci:1–27. https://doi.org/10.1080/13658816.2020.1730848