International Journal of Geospatial and Environmental Research

Volume 7 | Number 3 Article 1

July 2020

A Comparison of Network Clustering Algorithms in Keyword Network Analysis: A Case Study with Geography Conference Presentations

Youngho Lee
Kyung Hee University, emfo0124@khu.ac.kr

Yubin Lee
Kyung Hee University, leeyubin@khu.ac.kr

Jeong Seong
University of West Georgia, jseong@westga.edu

Ana Stanescu
University of West Georgia, astanesc@westga.edu

Chul Sue Hwang
Kyung Hee University, hcs@khu.ac.kr

A Comparison of Network Clustering Algorithms in Keyword Network Analysis: A Case Study with Geography Conference Presentations

Abstract

The keyword network analysis has been used for summarizing research trends, and network clustering algorithms play important roles in identifying major research themes. In this paper, we performed a comparative analysis of network clustering algorithms to find out their performances, effectiveness, and impact on cluster themes. The AAG (American Association for Geographers) conference datasets were used in this research. We evaluated seven algorithms with modularity, processing time, and cluster members. The Louvain algorithm showed the best performance in terms of modularity and processing time, followed by the Fast Greedy algorithm. Examining cluster members also showed very coherent connections among cluster members. This study may help researchers to choose a suitable network clustering algorithm and understand geography research trends and topical fields.

Keywords

Network Clustering Algorithm, Keyword Network Analysis, Geography, Research Trends

Acknowledgements

This research was supported by the MSIT (Ministry of Science, ICT), Republic of Korea, under the High-Potential Individuals Global Training Program (IITP-2019-0-01603) supervised by the IITP (Institute for Information and Communications Technology Planning and Evaluation). We also thank the AAG Headquarter and Council Members for providing the abstracts dataset for this research.

1 INTRODUCTION

Geography is a complex and comprehensive academic field that analyzes geographical space where human and natural environmental factors are interconnected (Tuan 1971). Geographic research can contribute to help solve social problems or determine urban policies by considering the space in which people are active (Cho et al. 2013; Seong et al. 2011; Rogalsky 2010; Hanchette 1999). Geographers tackled such problems from many different perspectives, and sometimes it is very difficult to figure out evolving research trends among, for example, physical geography, cultural geography, economic geography, political geography, climatology, biogeography, GIS, and remote sensing (Gorraiz et al. 2016). Summarizing papers published in academic journals is useful for understanding research trends (Ke et al. 2009), and it also helps researchers choose future research topic (Tsai and Lydia Wen 2005).

Bibliometric methods have been used successfully for summarizing articles and analyzing research trends (Nederhof 2006). The traditional bibliometric methods used mostly quantitative indicators such as the frequency of papers or citations. The traditional approaches have limitations in identifying influential keywords, relationships between keywords and clustering research fields. Recently, network analysis has been used to overcome the limitations of the traditional bibliometric methodologies (Tijssen 1992). Network analysis is a way of quantitatively analyzing the structure of a network by building relationships between entities like people and objects (Zhang et al. 2015). Keyword network analysis finds important keywords from frequency or centrality values. It identifies relationships among keywords from their co-occurrence (Su and Lee 2010). Its visual output also intuitively presents important keywords and their relationships (Cheng et al. 2018).

Keyword network analysis involves frequency analysis, co-occurring frequency analysis, centrality analysis, network clustering analysis, etc. (Lee et al. 2019; Kang et al. 2017). Frequency analysis, co-occurring frequency analysis and centrality analysis have been used in various research projects to identify important keywords and relationships among them (De Rezende et al. 2018; Zhuang et al. 2013). However, network clustering analysis has been overlooked frequently considering its critical roles in determining overall network structures and cluster membership (Sathik et al. 2011). Network clustering analysis, also known as community detection, is a way of identifying clusters of nodes that are densely connected to each other on the network (Bu et al. 2013). It reveals underlying relationships among nodes that are not easily identifiable (Lancichinetti et al. 2008). This approach has received significant attention in recent years as an important topic in network science (Emmons et al. 2016; Yang et al. 2016; Sathik et al. 2011; Lancichinetti and Fortunato 2009). Many network clustering algorithms have been developed, but their performances and effectiveness were not tested in summarizing geographic research trends.

The goal of this paper is to compare multiple network clustering algorithms with keyword datasets. Specific objectives are (1) identifying the best algorithm in terms of modularity and processing time, and (2) identifying cluster effectiveness by examining cluster members. This paper has five sections including Introduction. Section 2 reviews network clustering algorithms. Section 3 introduces research data and methodologies.

Section 4 evaluates each algorithm's performance using modularity, processing time, and cluster effectiveness. Section 5 is a summary.

2 LITERATURE REVIEW

2.1 Network Clustering Algorithms

Network clustering analysis has been widely used to identify the structure of a network in various research fields. For this reason, more than eleven network clustering algorithms have been developed (Waltman and Van Eck 2013; Blondel et al. 2008; Van Dongen 2008; Raghavan et al. 2007; Rosvall and Bergstrom 2007; Newman 2006; Reichardt and Bornholdt 2006; Duch and Arenas 2005; Pons and Latapy 2005; Clauset et al. 2004; Girvan and Newman 2002). In this study, we compared seven algorithms: Edge Betweenness, Fast Greedy, Walktrap, Leading Eigenvector, Infomap, Label Propagation and Louvain, because they have been widely used in network clustering analysis (Zhao et al. 2018; Wang and Koopman 2017; Emmons et al. 2016; Yang et al. 2016; Liu et al. 2012; Orman et al. 2011; Sathik et al. 2011; Lancichinetti and Fortunato 2009).

The first algorithm is the Edge Betweenness (EB) algorithm (Girvan and Newman 2002). It is based on a hierarchical divisive algorithm that is a top-down clustering process. The basic idea is to iteratively remove the edges that have high edge betweenness centrality values. The edge betweenness centrality is defined as the number of shortest paths between pairs of nodes that go through the edge in the network. The procedure of edge removal is looped until the modularity of the result reaches a maximum. The modularity (Newman and Girvan 2004) is a well-known function that evaluates the quality of a division of clusters. The Edge Betweenness algorithm is suitable for small networks because of its slow performance.

The Fast Greedy (FG) algorithm (Clauset et al. 2004) is a fast implementation of an algorithm developed by Newman (2004). It is based on a hierarchical agglomerative algorithm that is a bottom-up clustering process. This algorithm starts with each node that forms a singleton cluster. Then the expected improvement of modularity for each pair of clusters is calculated. After that, clusters are merged into new clusters by choosing the pairs that give the maximum improvement of modularity. The above procedure stops when there are no cluster pairs whose merger results in an increase in modularity.

The Walktrap algorithm (Pons and Latapy 2005) is based on a hierarchical agglomerative clustering process. The general idea of this algorithm is that short distance random walks are more likely to stay within the same cluster. The distance is computed from the probabilities given by random walks in the graph. If the two nodes are in same cluster, the distance must be small. The basic procedure is as follows: each node is regarded as a cluster, and the distances between all adjacent nodes are calculated. Then, this algorithm selects two adjacent clusters based on the distance, and merges these two clusters into a new cluster. After that, the distances between clusters are recalculated. The Walktrap algorithm has several advantages like it can be computed efficiently and it captures many characteristics on the structure of clusters.

The Leading Eigenvector (LE) algorithm (Newman 2006) is based on the process that maximizes network modularity by calculating the eigenvalues and eigenvectors of a modularity matrix. A modularity matrix is a remarkable attribute of a network and is independent of any partitioning of a network to clusters. This algorithm starts with calculating the leading eigenvector of a modularity matrix. Then, it divides a graph into two sub-graphs in a way that the improvement of modularity is maximized by using the leading eigenvector. This procedure is performed repeatedly until the value of modularity contribution becomes negative.

The Infomap algorithm (Rosvall and Bergstrom 2007) is based on the principles of information theory. It optimizes a quality function, namely the minimum description length (Rissanen 1978, Grünwald et al. 2005). The basic idea of the minimum description length is that the best explanation about the data is the one that permits the greatest compression of the data. The Infomap algorithm identifies clusters by optimally compressing description of information flows on a network. This algorithm can find an acceptable approximation to the optimal solution in large-scale networks. However, it has a limitation in that an increase in network size results in a rapid increase of running time.

The Label Propagation (LP) algorithm (Raghavan et al. 2007) is based on an iterative process to find stable clusters in a network. The basic idea of this algorithm is that each node in a network belongs to the cluster most common amongst its neighbor. It begins by assigning each node in the network to a distinct label (cluster). Then, through a random sequence, each node selects the label of the majority of its neighbors. This process is iteratively simulated, and stops once the nodes with the same label are grouped together into one cluster. An advantage of this algorithm is time efficiency. This algorithm's processing time is nearly linear in the number of nodes.

The Louvain algorithm (Blondel et al. 2008) is a hierarchical agglomerative method. It is based on a greedy approach to modularity optimization. It starts with assigning each node in a network to a unique cluster. Then, each node is placed into another cluster in order to improve network modularity. The procedure is repeated for all nodes until modularity does not increase any further. After that, each cluster is regarded as a single node, and the same procedure is repeated until no further improvement of modularity is achieved. This algorithm provides a fair compromise between computational complexity and the accuracy of the estimate of the modularity maximum.

2.2 Comparative Studies on Network Clustering Algorithms

Lancichinetti and Fortunato (2009) conducted a comparative analysis of the performances of several network clustering algorithms using benchmarks and random graphs. They concluded that the Infomap algorithm performed better than other algorithms. Also, the Louvain algorithm and the Multiresolution community detection algorithm proposed by Ronhovde and Nussinov (2009) had excellent performance showing low computational complexity in large networks. Sathik et al. (2011) compared the algorithm that they introduced with the Extremal Optimization algorithm (Duch and Arenas 2005) and the algorithm proposed by Newman (2004). They used various datasets including small networks and large networks, and they concluded that their algorithm provided enhancement in discovering clusters. Orman et al. (2011) assessed the performance of the five network clustering algorithms: Fast Greedy, Infomap, Louvain, Markov Cluster (Van Dongen 2008) and Walktrap. When measured with the normalized mutual information (NMI), the Infomap algorithm ranked first. Liu et al. (2012) compared a new hybrid clustering method with the Leading eigenvector and Louvain algorithms. They performed quantitative comparisons and also investigated whether the outcome provides an optimal representation.

Emmons et al. (2016) evaluated four widely used network clustering algorithms: Louvain, Infomap, Label Propagation, and Smart Local Moving (Waltman and Van Eck 2013). Overall, the Smart Local Moving algorithm was the best, followed by Louvain and Infomap. The Label Propagation showed the widest variability in the evaluation of algorithms. Yang et al. (2016) performed a comparative study on the accuracy and computing time of eight network clustering algorithms: Edge Betweenness, Fast Greedy, Infomap, Label Propagation, Leading Eigenvector, Louvain, Spinglass (Reichardt and Bornholdt 2006), and Walktrap. The Louvain algorithm outperformed over the other algorithms. They summarized algorithm recommendations based on the parameter and network size. Wang and Koopman (2017) compared the Louvain algorithm with the K-Means algorithm which is one of the simplest unsupervised learning algorithms. These two algorithms had similar performances but the K-Means algorithm was highly scalable for a bigger dataset. Zhao et al. (2018) compared six network clustering algorithms using the modularity that evaluates the quality of the network clustering and time complexity. The algorithms included Edge Betweenness, Fast Greedy, Louvain, Label Propagation, Infomap and Leading Eigenvector. The Louvain algorithm achieved the best performance on four different datasets.

Considering previous comparative studies on network clustering algorithms, the Infomap algorithm and the Louvain algorithm were superior to other algorithms. However, it is unclear whether these algorithms are suitable for keyword network analysis in geography, because most of the existing studies used benchmark networks, random networks, and social networks. The applicability and performances of those algorithms is worth being examined with a large keyword-based network dataset from the field of geography.

3 DATA AND METHODS

In this study, we used four network datasets to evaluate the impact of different network sizes. The network datasets were extracted from the abstracts of AAG (American Association of Geographers) annual conferences. AAG is one of the largest geography conferences in the United States, and thousands of abstracts are been submitted to the conference every year. Via web crawling and with help from the AAG Headquarter office, we were able to obtain conference abstracts. Each record contains title, authors, author affiliations, abstract, keywords, and topics. Among them, we used only the keywords of each abstract. The first dataset was retrieved from the 2019 conference papers that contain “GIS” as a keyword. It consists of 1,091 nodes and 2,790 edges. The second dataset was retrieved from the 1999 AAG conference, and consists of 3,219 nodes and 7,180 edges. The third dataset was retrieved from the 2009 AAG conference, and consists of 8,560 nodes and 35,364 edges. The fourth dataset was retrieved from the 2015-2019 AAG conferences, and consists of 40,392 nodes and 251,410 edges. Four sample datasets were chosen to account for changes in geographic research topics and potential effects of sample size.

To evaluate the performances of the network clustering algorithms, we used the modularity proposed by Newman and Girvan (2004), because it is broadly used for quantifying the goodness of network clusters (Emmons et al. 2016; Zhao et al. 2018). The modularity measures the density of edges inside clusters as compared to edges between clusters. It is a scalar value between -1 and 1, where a positive value closer to one indicates that clusters in the network are strongly connected internally and weakly connected externally, a negative value indicates the opposite case, and the value of zero means that edges simply exhibit a random distribution. The modularity is defined as follows.

$$Q = \frac{1}{2m}\sum_{i,j}\left[A_{ij} - \frac{k_i k_j}{2m}\right]\delta(c_i, c_j) \tag{1}$$

where A_{ij} represents the number of connections between node i and j, c_i indicates the cluster to which node i is assigned, $\delta(x, y)$ is 1 if $x = y$ and 0 otherwise, $k_i = \sum_j A_{ij}$ and $m = \frac{1}{2}\sum_{i,j} A_{ij}$. In addition to the modularity, we calculated the processing time of each algorithm to consider time efficiency. We conducted network clustering analysis in the igraph package (Csardi and Nepusz 2006) in R. The igraph package supports multiple data types and provides various functions for graph construction, graph visualization, network clustering and centrality calculations. It supports multiple export file formats to be used in other programs like Cytoscape. The analyses were performed on a computer equipped with 6 Core Intel Xeon Gold 6128 CPU, 1TB SSD storage and 32GB memory.

4 RESULTS

4.1 Comparative Analysis of Network Clustering Algorithms with Modularity and Processing Time

Table 1 shows the modularity values of different datasets and algorithms. First of all, modularity values decrease as the network size increases. No matter what algorithm is used, modularity is dependent on network size. The Louvain algorithm outperforms all the other algorithms in all datasets. Even though we used geography keywords, an excellent performance of the Louvain algorithm was observed, which aligns with previous studies. The Edge Betweenness algorithm ranks second in some datasets. However, it could not complete the network clustering analysis in a large size network (e.g. 2015-2019 AAG paper keywords). It works well with small networks. The Fast Greedy algorithm that uses a greedy approach like the Louvain algorithm shows excellent performance in large size networks. The modularity of the Label Propagation algorithm is not relatively stable compared with other algorithms due to its stochasticity. In particular, it shows a very low modularity in the 2015-2019 AAG dataset. Occupying the middle ranks, the Infomap, Leading Eigenvector, and Walktrap algorithms do not achieve good performances in either of the datasets.

Table 1. The comparison of modularity from different algorithms on four datsets

Dataset	EB	FG	Infomap	LP	LE	Louvain	Walktrap
2019 AAG related to GIS	0.8475	0.8066	0.8052	0.7626	0.8091	0.8499	0.8183
1999 AAG	0.6702	0.6666	0.6194	0.6279	0.5842	0.6859	0.5940
2009 AAG	0.5140	0.5122	0.4692	0.4389	0.4381	0.5447	0.4430
2015-2019 AAG	-	0.4127	0.3377	0.0581	0.3045	0.4239	0.3320

Table 2 shows the results of processing time. Overall, the Label Propagation and Louvain algorithms are faster than other algorithms. These two algorithms consistently rank first or second in both small sized-networks and large sized-networks because they hold a computational advantage. While the Fast Greedy and Walktrap algorithms are very fast with small networks, they show an increase in time in larger networks. The Edge Betweenness algorithm showed the longest processing times with all networks. The Infomap algorithm did not perform well. The Leading Eigenvector algorithm does not show remarkable results, but its time efficiency is good for large size networks.

Table 2. The comparison of processing time (in seconds) from different algorithms on the four datasets

Dataset	EB	FG	Infomap	LP	LE	Louvain	Walktrap
2019 AAG related to GIS	39	0.005	0.313	0.004	0.215	0.006	0.043
1999 AAG	2437	0.042	2.421	0.049	0.938	0.021	0.283
2009 AAG	181326	0.742	12.751	0.078	2.713	0.097	3.186
2015-2019 AAG	∞	29.609	137.501	0.424	8.523	0.565	119.938

Considering both modularity and processing time on four datasets, we conclude that the Louvain algorithm shows the best performance in keyword network analysis of geography. While the Label Propagation is superior to other algorithms in some datasets in terms of the processing time, the modularity of this algorithm is low compared with the Louvain algorithm. The Fast Greedy algorithm also shows relatively high modularity value and fast processing time.

4.2 Comparative Analysis with Visualization

The quantitative results described above need to be validated further by examining cluster members. With the Cytoscape program (Shannon et al. 2003), we visualized the results of network clustering analysis to assess how well the results reflect clusters of actual research trends. The Cytoscape program is an open source software platform for visualizing complex and interaction networks. It allows users to customize their own network graphs dynamically using graphical user interfaces. The 2015-2019 AAG dataset was used, and only the top 100 most frequent keywords were visualized due to space restriction.

Table 3 shows the number of clusters from different algorithms, except Edge Betweenness, using all keywords and using the top 100 keywords. In the case of all keywords, the Leading Eigenvector algorithm has the least number of clusters, and the Infomap and Walktrap algorithms kept the most. In the top 100 keywords, the Label Propagation algorithm has only one cluster, and the Infomap algorithm is the most. Having one cluster means that all the top 100 keywords are in the same cluster, which is not appropriate.

Table 3. The number of clusters from different algorithms in dataset 2015-2019 AAG

Dataset	FG	Infomap	LP	LE	Louvain	Walktrap
All keywords	1574	4117	1511	689	1011	4531
Top 100 keywords	4	16	1	6	7	3

Figure 1 shows the network graph of the Fast Greedy algorithm. It has three major groups and one minor group. The major groups (A, B, C) represent climate change/environment, urban/people, and methodologies/transportation. Although China, Urbanization, and Innovation appear in the small group (D), the Fast Greedy algorithm classifies the top 100 keywords into three large groups, similar to the traditional classification of geography as human geography, physical geography, and GIS.

Figure 2 shows the network graph of the Infomap algorithm. It consists of one large group, three mid-sized groups, and twelve small groups. The Infomap algorithm created the highest number of groups. The largest group (A) comprises human geography, and the three mid-sized groups (B, C, D) cover GIS, remote sensing, and climate change, respectively. It is interesting that ‘GIS’ and ‘Remote Sensing’, which have strongly connected edges, are separated into different clusters.

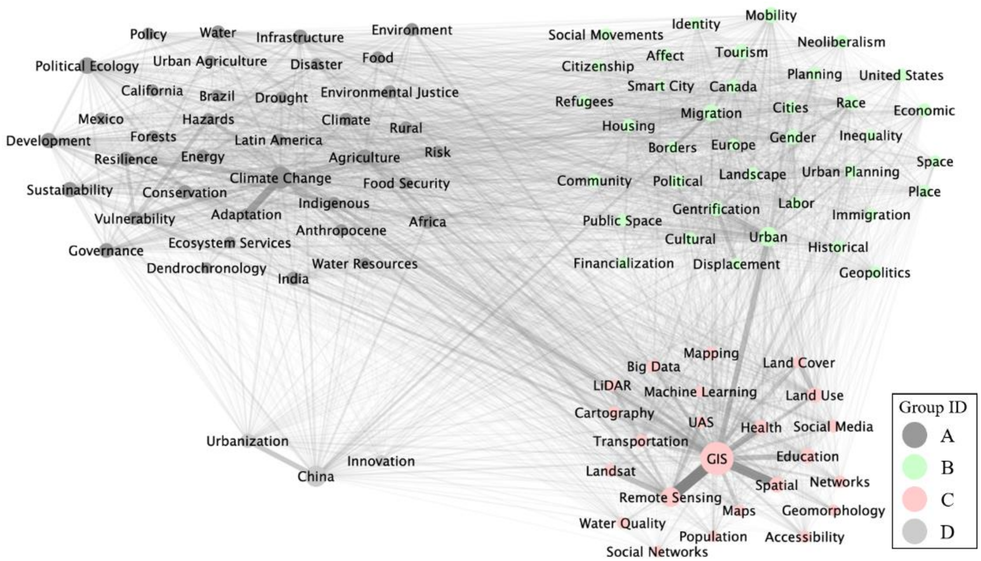

Figure 1. Network clustering with Fast greedy algorithm

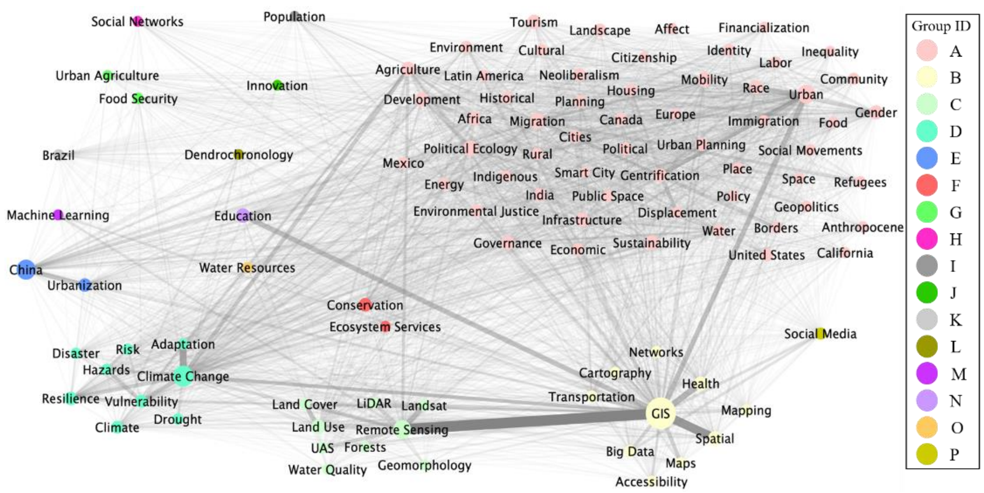

Figure 2. Network clustering with Infomap algorithm

Figure 3 shows the network graph produced by the Leading Eigenvector algorithm. It has four major groups (A, B, C, D) and two small groups (E, F), and each major group includes the third world and natural environment, people, developed countries, and methodologies, thus tying together physical geography and climate. It is quite different from the results of other algorithms and the traditional classification of geography.

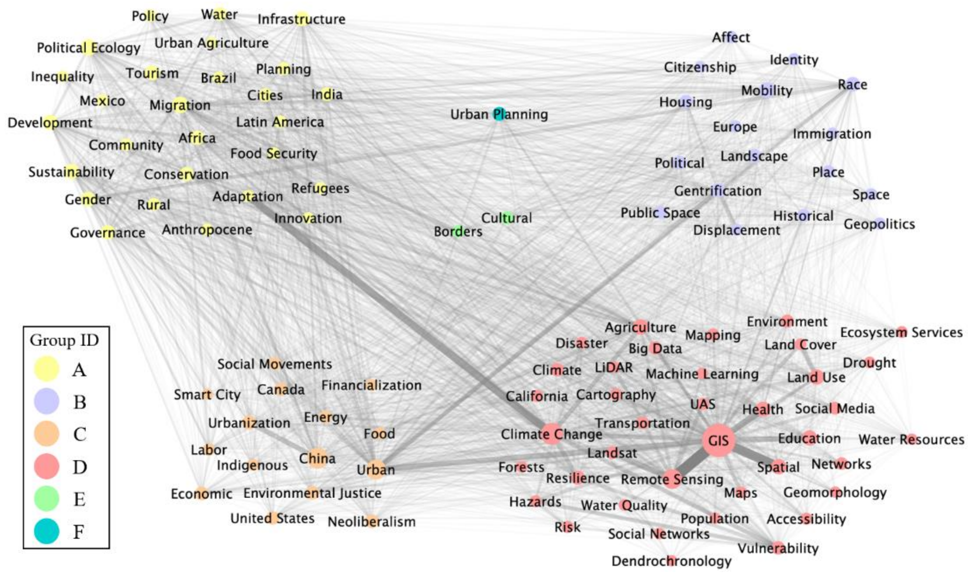

Figure 3. Network clustering with Leading eigenvector algorithm

Figure 4 shows the network graph of the Louvain algorithm. It consists of four major groups (A, B, C, D) and three small groups (E, F, G), and each major group covers environment, climate change, urban/people, and methodologies/transportation. The results of this algorithm are similar to the Fast Greedy algorithm except that environment and climate change are separated into different groups.

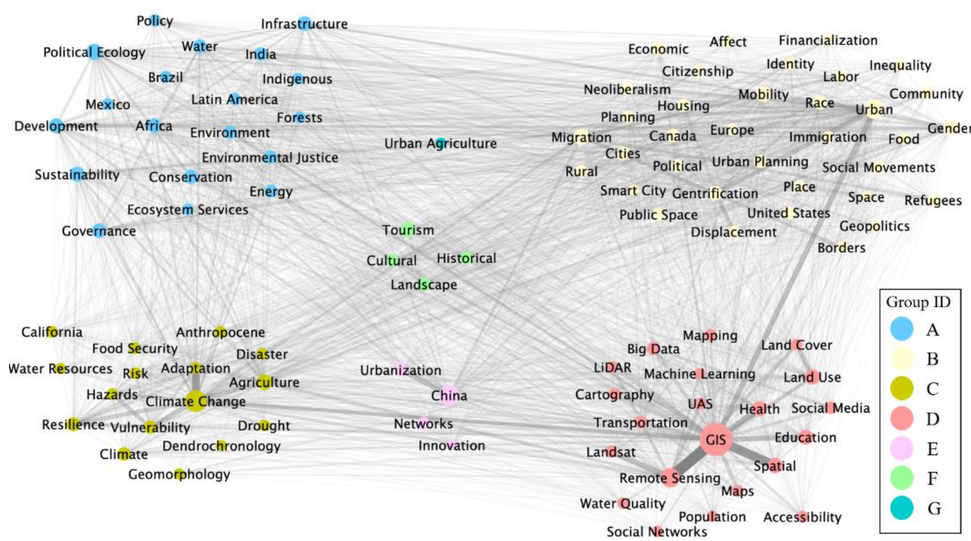

Figure 4. Network clustering with Louvain algorithm

Figure 5 shows the network graph of the Walktrap algorithm. It has one large group (A) and two smaller-sized groups (B, C) relative to the large group, but does not have any minor groups. The largest group embraces human geography like the Infomap algorithm, and each of the smaller groups comprise climate change and GIS/transportation, respectively. The keywords that are related to remote sensing are separated from GIS, and appear with climate change.

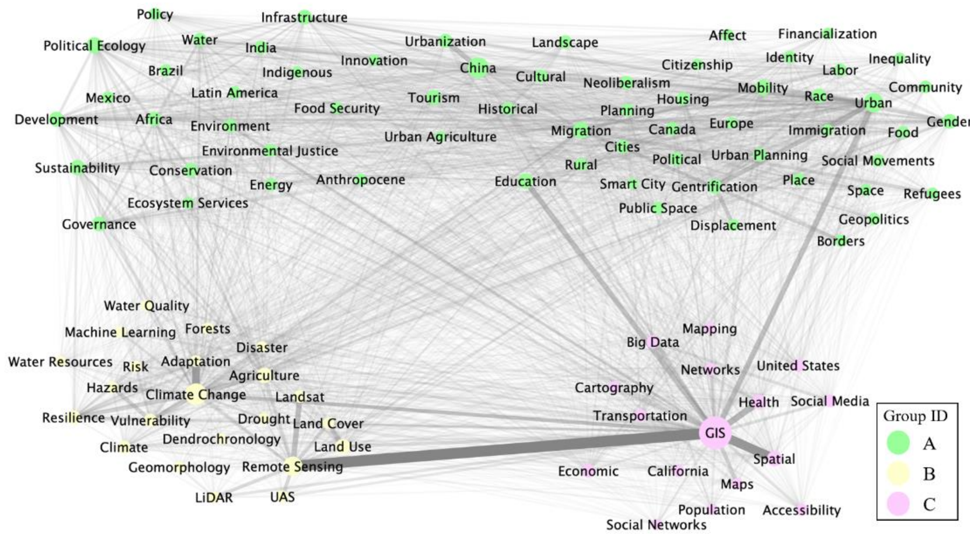

Figure 5. Network clustering with Walktrap algorithm

The Label Propagation algorithm was not visualized because it yielded only one cluster in the top 100 keywords. Considering the results of all keywords, 86% of all keywords were assigned to only one cluster with the rest of them divided into 1,510 clusters. It is expected to be due to the large network size which results in a relatively long process for the algorithm. As the procedure repeated a lot, the keywords related to the main keywords continued to assign to one cluster by the property of the algorithm that each node selects the cluster of the majority of its neighbors, and the unrelated keywords continued to be isolated.

Even if it is difficult to determine which algorithm is superior to other algorithms by examining clusters and their members, we see that the Fast Greedy and Louvain algorithms have portrayed geography research fields suitably when the AAG Specialty Groups and the AAG Conference topical fields are considered. While the Fast Greedy algorithm generated more comprehensive clusters in general, the Louvain algorithm generated more detail clusters with the top 100 frequent keywords.

5 CONCLUSIONS

Geography is a comprehensive field covering various research branches such as biology, politics, and economy. The variety makes it difficult for professionals to identify overall research trends. Recently, keyword network analysis has been used for summarizing research trends, and many network clustering algorithms were used to determine overall network cluster structures and cluster memberships. In this study, our goal was to conduct a comparative analysis of network clustering algorithms to identify an optimal algorithm to work with keywords of conference presentations. We prepared and used four datasets built from AAG conference abstracts, and evaluated the applicability and performance of seven algorithms in terms of modularity and runtime. The results of this study show that the Louvain algorithm outperforms other algorithms in terms of modularity and processing time. The Fast Greedy algorithm also showed high modularity and fast processing time in all four datasets. The Label Propagation algorithm performed better than other algorithms except for the Louvain algorithm in the processing time but it showed low modularity values. When network graphs were visualized and analyzed, we found that the Fast Greedy and Louvain algorithms were superior in portraying geography research trends. Additional network characteristics such as the small world effect, scale free networks and inherent scaling hierarchy are valuable in analyzing complex networks, but this study is focusing on discovering the optimal network clustering algorithm in summarizing the geography research trends. In addition, user's discretion is needed when the Louvain algorithm is applied to other domain datasets. This study demonstrated that a network clustering analysis may be used to summarize research trends in geography. This research can also help researchers select a suitable algorithm when performing keyword network analysis.

REFERENCES

Blondel, V. D., Guillaume, J. L., Lambiotte, R. and Lefebvre, E. (2008) Fast unfolding of communities in large networks. *Journal of Statistical Mechanics: Theory and Experiment*, P10008.

Bu, Z., Zhang, C., Xia, Z. and Wang, J. (2013) A fast parallel modularity optimization algorithm (FPMQA) for community detection in online social network. *Knowledge-Based Systems*, 50, 246-259.

Cheng, F. F., Huang, Y. W., Yu, H. C. and Wu, C. S. (2018) Mapping knowledge structure by keyword co-occurrence and social network analysis. *Library Hi Tech*. 36(4), 636-650.

Cho, S., Kang, J. Y., Yasar, A. U. H., Knapen, L., Bellemans, T., Janssens, D., Wets, G. and Hwang, C. S. (2013) An activity-based carpooling microsimulation using ontology. *Procedia Computer Science*, 19, 48-55.

Clauset, A., Newman, M. E. and Moore, C. (2004) Finding community structure in very large networks. *Physical Review E*, 70(6), 066111.

Csardi, G. and Nepusz, T. (2006) The igraph software package for complex network research. *InterJournal, Complex Systems*, 1695(5), 1-9.

De Rezende, L. B., Blackwell, P. and Pessanha Goncalves, M. D. (2018) Research focuses, trends, and major findings on project complexity: A bibliometric network analysis of 50 years of project complexity research. *Project Management Journal*, 49(1), 42-56.

Duch, J. and Arenas, A. (2005) Community detection in complex networks using extremal optimization. *Physical Review E*, 72(2), 027104.

Emmons, S., Kobourov, S., Gallant, M. and Börner, K. (2016) Analysis of network clustering algorithms and cluster quality metrics at scale. *PLOS ONE*, 11(7), e0159161.

Girvan, M. and Newman, M. E. (2002) Community structure in social and biological networks. *Proceedings of the National Academy of Sciences*, 99, 7821-7826.

Gorraiz, J., Gumpenberger, C. and Glade, T. (2016) On the bibliometric coordinates of four different research fields in Geography. *Scientometrics*, 107(2), 873-897.

Grünwald, P. D., Myung, I. J. and Pitt, M. A. (Eds.). (2005) *Advances in Minimum Description Length: Theory and Applications*. MIT Press.

Hanchette, C. L. (1999) GIS and decision making for public health agencies: childhood lead poisoning and welfare reform. *Journal of Public Health Management and Practice*, 5(4), 41-46.

Kang, G. J., Ewing-Nelson, S. R., Mackey, L., Schlitt, J. T., Marathe, A., Abbas, K. M. and Swarup, S. (2017) Semantic network analysis of vaccine sentiment in online social media. *Vaccine*, 35(29), 3621-3638.

Ke, Y., Wang, S., Chan, A. P. and Cheung, E. (2009) Research trend of public-private partnership in construction journals. *Journal of Construction Engineering and Management*, 135(10), 1076-1086.

Lancichinetti, A. and Fortunato, S. (2009) Community detection algorithms: a comparative analysis. *Physical Review E*, 22, 80(5), 056117.

Lancichinetti, A., Fortunato, S. and Radicchi, F. (2008) Benchmark graphs for testing community detection algorithms. *Physical Review E*, 78(4), 046110.

Lee, Y., Lee, Y., Seong, J. Stanescu, A., Hwang, T. and Hwang, C. S. (2019) Analysis of GIS Research Trends in the U.S. with AAG Conference Presentations from 2000-2019. *The Geographical Journal of Korea*, 53(4), 495-508.

Liu, X., Glänzel, W. and De Moor, B. (2012) Optimal and hierarchical clustering of large-scale hybrid networks for scientific mapping. *Scientometrics*, 91(2), 473-493.

Nederhof, A. J. (2006) Bibliometric monitoring of research performance in the social sciences and the humanities: A review. *Scientometrics*, 66(1): 81-100.

Newman, M. E. (2004) Fast algorithm for detecting community structure in networks. *Physical Review E*, 69(6), 066133.

Newman, M. E. (2006) Finding community structure in networks using the eigenvectors of matrices. *Physical Review E*, 74(3), 036104.

Newman, M. E. and Girvan, M. (2004) Finding and evaluating community structure in networks. *Physical Review E*, 69(2), 026113.

Orman, G. K., Labatut, V. and Cherifi, H. (2011) Qualitative comparison of community detection algorithms. In *International Conference on Digital Information and Communication Technology and its Applications*. Springer, Berlin, Heidelberg, 265-279.

Pons, P. and Latapy, M. (2005) Computing communities in large networks using random walks. In *International Symposium on Computer and Information Sciences*, Springer, Berlin, Heidelberg, 284-293.

Raghavan, U. N., Albert, R. and Kumara, S. (2007) Near linear time algorithm to detect community structures in large-scale networks. *Physical Review E*, 76(3), 036106.

Reichardt, J. and Bornholdt, S. (2006) Statistical mechanics of community detection. *Physical Review E*, 74(1), 016110.

Rissanen, J. (1978) Modelling by shortest data description. *Automatica*, 14(5), 465–471.

Rogalsky, J. (2010) The working poor and what GIS reveals about the possibilities of public transit. *Journal of Transport Geography*, 18(2), 226-237.

Ronhovde, P. and Nussinov, Z. (2009) Multiresolution community detection for megascale networks by information-based replica correlations. *Physical Review E*, 80(1), 016109.

Rosvall, M. and Bergstrom, C. T. (2007) An information-theoretic framework for resolving community structure in complex networks. *Proceedings of the National Academy of Sciences*, 104(18), 7327-7331.

Sathik, M. M., Senthamarai, K. and Rasheed, A. A. (2011) Comparative analysis of community discovery methods in social networks. *International Journal of Computer Applications*, 14(8), 0975-8887.

Seong, J. C., Park, T. H., Ko, J. H., Chang, S. I., Kim, M., Holt, J. B. and Mehdi, M. R. (2011) Modeling of road traffic noise and estimated human exposure in Fulton County, Georgia, USA. *Environment International*, 37(8), 1336-1341.

Shannon, P., Markiel, A., Ozier, O., Baliga, N. S., Wang, J. T., Ramage, D., Amin, N., Schwikowski, B. and Ideker, T. (2003) Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Research*, 13(11), 2498-2504.

Su, H. N. and Lee, P. C. (2010) Mapping knowledge structure by keyword co-occurrence: a first look at journal papers in Technology Foresight. *Scientometrics*, 85(1), 65-79.

Tijssen, R. J. (1992) A quantitative assessment of interdisciplinary structures in science and technology: co-classification analysis of energy research. *Research Policy*, 21(1), 27-44.

Tsai, C. C. and Lydia Wen, M. (2005) Research and trends in science education from 1998 to 2002: A content analysis of publication in selected journals. *International Journal of Science Education*, 27(1), 3-14.

Tuan, Y. F. (1971) Geography, phenomenology, and the study of human nature. *Canadian Geographer/Le Géographe Canadien*, 15(3), 181-192.

Van Dongen, S. (2008) Graph clustering via a discrete uncoupling process. *SIAM Journal on Matrix Analysis and Applications*, 30(1), 121-141.

Waltman, L. and Van Eck, N. J. (2013) A smart local moving algorithm for large-scale modularity-based community detection. *The European Physical Journal B*, 86(11), 471.

Wang, S. and Koopman, R. (2017) Clustering articles based on semantic similarity. *Scientometrics*, 111(2), 1017-1031.

Yang, Z., Algesheimer, R. and Tessone, C. J. (2016) A comparative analysis of community detection algorithms on artificial networks. *Scientific Reports*, 6, 30750.

Zhang, W., Zhang, Q., Yu, B. and Zhao, L. (2015) Knowledge map of creativity research based on keywords network and co-word analysis, 1992–2011. *Quality and Quantity*, 49(3), 1023-1038.

Zhao, Z., Zheng, S., Li, C., Sun, J., Chang, L. and Chiclana, F. (2018) A comparative study on community detection methods in complex networks. *Journal of Intelligent and Fuzzy Systems*, 35(1), 1077-1086.

Zhuang, Y., Liu, X., Nguyen, T., He, Q. and Hong, S. (2013) Global remote sensing research trends during 1991–2010: a bibliometric analysis. *Scientometrics*, 96(1), 203-219.