A model of geographic information system using graph clustering methods

by Tedy Setiadi

Submission date: 24-Aug-2020 06:39AM (UTC+0700)
Submission ID: 1373047169
File name: Gis_Graphcluster.pdf (445.08K)
Word count: 2793
Character count: 15423

2017 3rd International Conference on Science in Information Technology (ICSITech)

A Model of Geographic Information System using Graph Clustering Methods

Tedy Setiadi, Andri Pranolo, Muhammad Aziz
Department of Informatics
Universitas Ahmad Dahlan
Yogyakarta, Indonesia
tedy.setiadi@tif.uad.ac.id, andri.pranolo@tif.uad.ac.id,
moch.aziz@tif.uad.ac.id

Sukrisno Mardiyanto, Bayu Hendrajaya
School of Electrical Engineering and Informatics
Institut Teknologi Bandung
Bandung, Indonesia
sukrisno@informatika.org, bayu@informatika.org

Munir
Department of Computer Science
Universitas Pendidikan Indonesia
Bandung, Indonesia
munir@upi.edu

***Abstract*—This research aimed to modeling a geographic information system (GIS) using graph clustering. GIS discuss about spatial data that consists of several objects including points, lines, regions, rectangles, surfaces, volumes, and the all data representation of an object on earth. Graph-based clustering techniques are extremely useful because many real word problem domains have a natural graph representation. This research focused on case GIS for family empowerment post (POSDAYA) which is one of the programs for improving the quality of human beings, especially in achieving the Millennium Development Goals (MDGs) in Indonesia that has a priority on family-based poverty reduction. The results are the GIS framework model that use three components such as graph clustering, layering, and the view of data properties. GIS provided the spatial data and its properties, and graph clustering as a method or algorithm used to make the layers properties and related to each points based on similarities.**

Keywords—GIS; graph clustering; POSDAYA; MDGs; Indonesia

I. INTRODUCTION

Geographic Information System (GIS) is a system that has a function to provide spatial information and its properties. The data on GIS which is called as spatial data which has several objects including lines, regions, points, rectangles, volumes, surfaces, and all data representation of an object on earth [1]. GIS will provide an accurate information when spatial data are available and complete. Aside from being spatial data processing function, GIS able to provide information resources of a spatial or a particular region. The information resources according to GIS can be proceed and utilized by using data mining techniques, including methods of clustering and classification. Clustering technique related to an analysis how the dataset are organized to be groups based on similarity measures [2], [3], and features [4]. In this research, we are using clustering based graph which is including in graph partitioning topic [3]. Previous findings in spatial data clustering have been reported, such as Clustering Large Applications based upon Randomized search (CLARANS) algorithm [5], Balanced Iterative Reducing and Clustering using Hierarchies (BIRCH), DBSCAN (Density-based clustering algorithm) [6], random walks [7], Partitioning Around Medoids (PAM) [8], and CLARA (Clustering LARge Applications) [8]. Furthermore, some basic clustering algorithms and methods on graph clustering have been reported using k-Spanning Tree, Shared Nearest Neighbor, Restricted Neighborhood Search Clustering (RNSC), Betweenness Centrality Based, Highly Connected Components, Maximal Clique Enumeration, and Kernel k-means, Super Paramagnetic Clustering (SPC), Local Clique Merging Algorithm (LCMA), Geometric MST Clustering (GMC), Iterative Conductance Cutting (ICC), Markov Clustering (MCL), SideS algorithm, Harmony Search Algorithm, Molecular Complex Detection (MCODE), spectral clustering, and Genetic algorithm [3], [9]–[11].

This research aimed to propose GIS model by using graph clustering. GIS provided the spatial data and its properties, graph clustering method or algorithm given the visualization of GIS layers properties and made groups based on similarities of regions points. The discussion in this paper will be associated with the post of family empowerment (POSDAYA) as a case study. POSDAYA is a social group activities for gathering for collaboration, advocacy, communication, education, information, and also a coordination of activities to strengthen the integrated of family functions in groups [12]. Its program one of the program for especially in achieving the Millennium Development Goals (MDGs) in Indonesia.

The research findings to combine between GIS and other method were done before, among others are the finding of coupling the GIS by using artificial neural networks (ANN's) according to forecast of land use changing [13]. It's research

978-1-5090-5864-8/17/$31.00 ©2017 IEEE

resulting a land transformation model, where ANNs are used to learn the development region patterns and are tested the model forecast capacity, while the development of spatial information is using geographical information system which has function as a predictor and conclude the results of spatial analysis. Reference [14] proposed a GIS by combining cellular automata and neural networks. GIS utilized for obtaining spatial function for constructing a neural network and site attributes as a training data. In this research, the experiment of gradual land using conservation process utilize the iterative looping of neural network. Reference [21] use the generic procedure of ensemble learning and determine several weak graph clustering's.

II. Background

A. *Geographic Information System (GIS)*

Most of Geographic information system (GIS) definition can be viewed from two aspects, GIS as a technology which defined as a set of tools, and GIS as a problem solving is related to optimizing the spatial data and its attributes in the decision support activities [15]. Therefore, a GIS is representing the spatial data information system which has some relevant object or thi on the surface of the earth [16]. GIS also can defined as computer program for acquiring, storing, interpreting, and displaying spatially organized information [17]. GIS function can be divided into four components, such as input of data component, data storage and management, data analysis and manipulation, and data output [15], see Fig. 1.

Data input is the procedures to converts data from the existing form or raw into one or database which can be utilized by a GIS. It source can be form several process such a manual, scanning, digitizing, and remote sensing [15], [18]. There are two types of data to be entered in a GIS, they are spatial data (geographic location of features) and non-spatial data (descriptive or numeric information about features) [18].

Spatial data represented as a physical level in the form of raster format (grid cells/pixels) and vector format [15]. Reference [17] divided a layers in three types: 1) vector layers which has ntaining objects made up from points (e.g. town), line form (e.g. roads), or polygons form (e.g. national or state boundaries), 2) data which usually stored in the database tables contain attributes individual objects, and 3) raster layers which is containing sites data within a region.

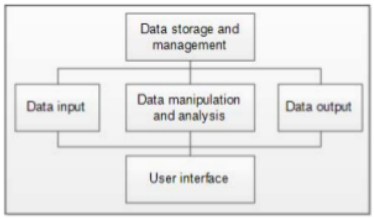

Fig. 1. Structure of GIS [15].

Type output of GIS can be classified into four categories [15], including: 1) text output such as tables, lists, number or text response to queries, 2) graphic output can be viewed such as by screen display, maps, and graphs, 3) digital data that stored in the disk or transmitted across a network, and 4) not yet commonly used, such as computer-generated sound and video clips. Data storage and management include both functions of store and retrieve the data from the database.

Two important kinds of functions within a GIS are data analysis and models [17]. Some technique can be used for comparing and analyzing data layers, such as kriging, spatial correlation of variables, nearest neighbor analysis, and fractal dimension of shapes. A GIS-based decision analysis requires representation of a real word geographic system in the digital format [15].

B. *Graph Clustering*

Graph clustering is associated with splitting and grouping vertices in a large graph based on similarity of various criteria including vertex connectivity [19]. Nowadays, graph clustering mostly used for many applications in areas such as medical field for diseases classification, grouping of compounds in the field of chemistry, and in the most popular one is in social studies, graph clustering use for classification of statistical findings [5]. Graph clustering also has been utilized for finding related people to make a group in the a networks analysis [20], give recommendations in a e-commerce based on a trend or group opinions, and in the bioinformatics field graph clustering has utilized for classification of gene expression and grouping of epidemics spreading [20].

The key issue of clustering analysis of spatial data has been defined by Ng and Han [5] as follows:

- Natural similarity aspect. Is there a natural similarity between "objects" to be clustered? Cluster analysis will be very easy if done on spatial attributes only,
- Size of objects. Classify objects number, because a traditional cluster analysis techniques or algorithms are not designing for a large data sets or objects.

The example, given Prim's minimum spanning tree algorithm as a simple clustering algorithm using graph clustering [11], shown in Fig. 2. At each iteration an edges is chosen that satisfies the following two criteria: (1) exactly one of its endpoints is in T and (2) the edge is of minimum weight among those that satisfy (1). Line 1 is initialize T with a random vertex from G. Line 3, 4, and five are executed iteratively until all the vertices in G become part of T (Line 2). When the algorithm terminates, T holds the minimum spanning tree of G.

Input: A (distance) weighted, undirected graph $G = (V, E)$
Output: A graph T that represents the minimum spanning tree
1 Initialize T with a random vertex from G
2 **while** $V(T) \neq V(G)$ **do**
3 $\quad E_T = \{(u, v) \in E(G) : u \in T \text{ or } v \in T\}$
4 $\quad min_{edge} = (u, v) : weight(u, v)$ minimum in E_T
5 $\quad$ Add min_{edge} to T
6 **end**
7 **return** T

Fig. 2. Prim's minimum spanning tree algorithm.

C. Family Empowerment Post (POSDAYA)

Family Empowerment Post (POSDAYA) is a forum gathering, advocacy ,communication, information, education and also can developed into coordination of activities strengthening family functions in an integrated manner [12]. POSDAYA is a container for all family members to their field wider. In POSDAYA wealthier families, with the support and assistance of government officials or community organizations, is expected to help families in need.

III. Method

In this preliminary study consists of three main phases such as analysis and design of spatial database systems (SDBMS) [16], [21]., data mining phase [21], and providing view on GIS.

A. Spatial Database Systems (SDBS)

Spatial Database Systems (SDBS) are database system containing of knowledge of spatial data. It used for managing data including data store and retrieving data.

B. Data mining phase

The data mining steps consisting selection, data reduction, data mining process, and evaluation [21].

- Selection. In this phase subset of all data of knowledge or attributes are selected for the next process.
- Data Reduction. Data reduction is a process of transformation technique of selected attributes to be more effective and ready to compute in the data mining phase.
- Data mining. In this phase, we applied a suitable algorithm to produce a data pattern
- Result Interpretation. This step aims to see the evaluation of the pattern to describe the information which gain from the data mining process.

C. Combining data mining and GIS visualization

For combining data mining results and GIS visualization, we use point locations on the Google Maps.

IV. Results and Discussion

A. Spatial Database Systems (SDBMS)

The spatial database systems for Posdaya shown in Fig.3 consists of Posdaya coordinate latitude and longitude.

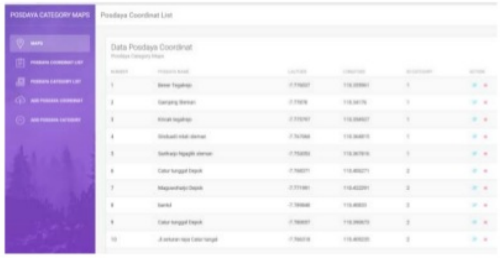

Fig. 3. Spatial database system of Posdaya.

B. Data mining and GIS visualization

The prototyping of Geographic Information System (GIS) combined with graph clustering shown in Fig 3. The sample of maps is Yogyakarta city, Indonesia. The kind graph shown is a star graph with n-vertices. A star graph is a finished bipartite graph when a single vertex according to one set and the other remaining vertices according to other set [22]. In the Fig. 4 (a), (b), (c) and (d), there are set of 'n' vertices which a single vertex has connected to all of the 'n–1' vertices. Fig. 3 (a) shows the user interface of GIS system consisting of : (1) the maps with points that connected with the a single vertex in the form of star graph, (2) the second is information of region and clustering produced, and (3) the third is information about symbol/properties. Fig. 5 shows the view of GIS with all clusters, each clustering differencing with the color of point and grouping by line with a centroid for each clusters.

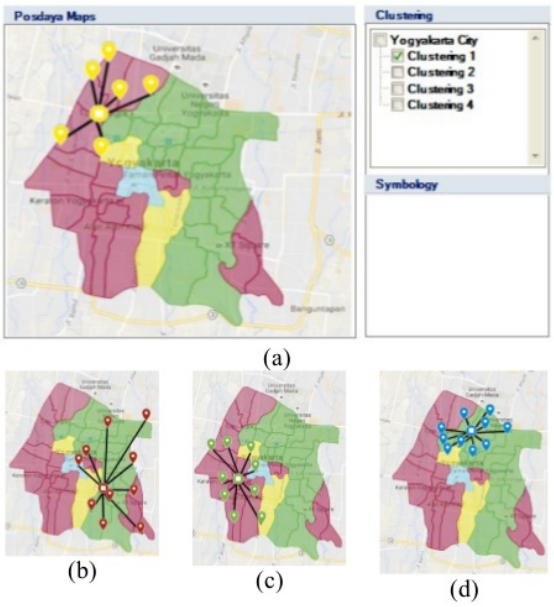

Fig. 4. View of GIS with graph clustering.

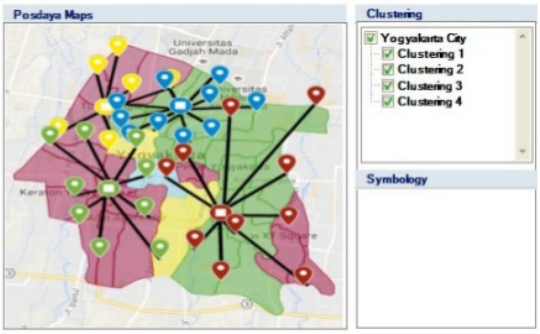

Fig. 5. View of GIS that shown all cluster.

C. Simulation results on Google Maps

In the preliminary research we have classified Posdaya data into 4 groups or clusters level such as Pre-Posdaya (Fig. 6a), semi-independent (Fig. 6b), independent (Fig. 6c), and core-independent Posdaya (Fig. 6d). The visualization on Google Maps shown in Fig 6.

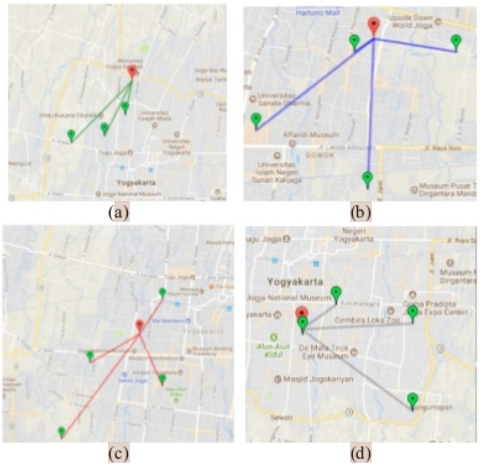

Fig. 6. Visualization of Posdaya on Google Maps.

V. CONCLUSION

The main contributor of this paper is a proposed the technique how to combine GIS with graph clustering. GIS provides spatial data and properties stored in spatial databases system (SDBS) that can be used to perform clustering. Clustering visualized in graph form. Each cluster has a centroid that is connected with the points representing a certain area or region that resembles as a star graph. Additionally, each cluster also has a calm colors.

To making the graph, the initial or first centroid is defined, and then the centroid will be looking for a similar point with the defined measurement. If the point checked is have a high similarity, then the point will be a group with the centroid and signed the point already have a group, else the point will be the next centroid. Each centroids will be looking for the other points which not yet have a group and adding the member of their group based on similarities, and the new centroid will be created if the points do not have similar meaning. It process repeated until all the dataset checked. To check the similarity we can used modified or established clustering method.

ACKNOWLEDGMENT

This work has been supported by Penelitian Kerjasama Perguruan Tinggi (PEKERTI), Directorate on research and higher education (RistekDikti) Republic of Indonesia with number grant number 011/HB-LIT/III/2016, 15 March 2016.

REFERENCES

[1] H. Samet, "Applications of spatial data structures," 1990.

[2] F.O. de França, "A hash-based co-clustering algorithm for categorical data," vol. 64, pp. 24–35, 2016.

[3] R. Mishra, S. Shukla, D. Arora, and M. Kumar, "An effective comparison of graph clustering algorithms via random graphs," Int. J. Comput. Appl., vol. 22, no. 1, pp. 22–27, 2011.

[4] D. Octaviani, M.S. Othman, N. Yusof, and A. Pranolo, "Applied Clustering Analysis for Grouping Behaviour of E-Learning Usage Based on Meaningful Learning Characteristics," J. Teknol., vol. 76, no. 1, pp. 125–138, 2015.

[5] R.T. Ng and J. Han, "CLARANS: A method for clustering objects for spatial data mining," IEEE Trans. Knowl. Data Eng., vol. 14, no. 5, pp. 1003–1016, 2002.

[6] W. Wang, J. Yang, R. Muntz, and others, "STING: A statistical information grid approach to spatial data mining," in VLDB, 1997, vol. 97, pp. 186–195.

[7] D. Harel and Y. Koren, "Clustering spatial data using random walks," in Proceedings of the seventh ACM SIGKDD international conference on Knowledge discovery and data mining, 2001, pp. 281–286.

[8] L. Kaufman and P.J. Rousseeuw, Finding groups in data: an introduction to cluster analysis, vol. 344. John Wiley & Sons, 2009.

[9] Y. Atay and H. Kodaz, "Modularity-Based Graph Clustering Using Harmony Search Algorithm," 2016, pp. 109–114.

[10] F.C. Graham and A. Tsiatas, "Finding and visualizing graph clusters using pagerank optimization," in International Workshop on Algorithms and Models for the Web-Graph, 2010, pp. 86–97.

[11] N.F. Samatova, W. Hendrix, J. Jenkins, K. Padmanabhan, and A. Chakraborty, Practical graph mining with R. CRC Press, 2013.

[12] H. Suyono and R. Haryanto, Buku pedoman pembentukan dan pengembangan pos pemberdayaan keluarga POSDAYA (The guidebook of formation and development the family empowerment post POSDAYA), 2nd ed. Jakarta Pusat: Balai Pustaka, 2011.

[13] B.C. Pijanowski, D.G. Brown, B.A. Shellito, and G.A. Manik, "Using neural networks and GIS to forecast land use changes: a land transformation model," Comput. Environ. Urban Syst., vol. 26, no. 6, pp. 553–575, 2002.

[14] X. Li and A. G.-O. Yeh, "Neural-network-based cellular automata for simulating multiple land use changes using GIS," Int. J. Geogr. Inf. Sci., vol. 16, no. 4, pp. 323–343, 2002.

[15] J. Malczewski, GIS and multicriteria decision analysis. John Wiley & Sons, 1999.

[16] M. Ester, H.-P. Kriegel, and J. Sander, "Spatial data mining: A database approach," in International Symposium on Spatial Databases, 1997, pp. 47–66.

[17] T. Bossomaier and B. A. Hope, Online GIS and spatial metadata. CRC Press, 2015.

[18] "Lecture 5-6. Data Input In GIS." [Online]. Available: http://www.columbia.edu/itc/eee/e1001/edit/lect3.html.

[19] Y. Zhou, H. Cheng, and J.X. Yu, "Clustering large attributed graphs: An efficient incremental approach," in 2010 IEEE International Conference on Data Mining, 2010, pp. 689–698.

[20] H. Almeida, D. G. Neto, W. Meira Jr, and M. J. Zaki, "Towards a better quality metric for graph cluster evaluation," J. Inf. Data Manag., vol. 3, no. 3, p. 378, 2012.

[21] M. Ester, A. Frommelt, H.-P. Kriegel, and J. Sander, "Spatial data mining: database primitives, algorithms and efficient DBMS support," Data Min. Knowl. Discov., vol. 4, no. 2–3, pp. 193–216, 2000.

[22] "Graph Theory - Types of Graphs." [Online]. Available: https://www.tutorialspoint.com/graph_theory/types_of_graphs.htm.

A model of geographic information system using graph clustering methods

ORIGINALITY REPORT

11%	4%	7%	5%
SIMILARITY INDEX	INTERNET SOURCES	PUBLICATIONS	STUDENT PAPERS

PRIMARY SOURCES

1	" 2016 2 ICSITech reviewer ", 2016 2nd International Conference on Science in Information Technology (ICSITech), 2016 Publication	1%
2	www.kemalapublisher.com Internet Source	1%
3	Submitted to National Institute of Technology Warangal Student Paper	1%
4	Submitted to University of Ghana Student Paper	1%
5	Ardiansyah Ardiansyah, Murein Miksa Mardhia, Sri Handayaningsih. "Analogy-based model for software project effort estimation", International Journal of Advances in Intelligent Informatics, 2018 Publication	1%
6	www.csc2.ncsu.edu Internet Source	1%

7	"Advances in Swarm Intelligence", Springer Science and Business Media LLC, 2013 Publication	1%
8	digitalcommons.wpi.edu Internet Source	1%
9	David G. Green, Nicholas I. Klomp, Glyn Rimmington, Suzanne Sadedin. "Chapter 3 Complexity in Landscapes", Springer Science and Business Media LLC, 2020 Publication	1%
10	Zhou, Yang, Hong Cheng, and Jeffrey Xu Yu. "Clustering Large Attributed Graphs: An Efficient Incremental Approach", 2010 IEEE International Conference on Data Mining, 2010. Publication	<1%
11	Muhamad Nursalman, Arif Sasongko, Yusuf Kurniawan, Kuspriyanto. "Improved generalizations of the Karatsuba algorithm in GF(2^n)", 2014 International Conference of Advanced Informatics: Concept, Theory and Application (ICAICTA), 2014 Publication	<1%
12	econpapers.repec.org Internet Source	<1%
13	www.cde.ca.gov Internet Source	<1%

14	Bryan C Pijanowski, Daniel G Brown, Bradley A Shellito, Gaurav A Manik. "Using neural networks and GIS to forecast land use changes: a Land Transformation Model", Computers, Environment and Urban Systems, 2002 Publication	<1%
15	Xu. "Subject Index", Clustering, 10/09/2008 Publication	<1%
16	www.ecice06.com Internet Source	<1%
17	Rashidul Hasan, Zakir Hossain, Fahim Chowdhury, Mohammad Khalad Hasan. "Extended algorithm for spatial characterization and discrimination rules", 2008 11th International Conference on Computer and Information Technology, 2008 Publication	<1%

Exclude quotes	Off	Exclude matches	Off
Exclude bibliography	On		