

## Measuring full and partial geometrical similarity of Map objects

Mahdi JASIM

College of Computer Technologies  
Univ. of Babylon-Iraq  
[mnjisd@yahoo.com](mailto:mnjisd@yahoo.com)

cell. 009647813184069

Tawfiq Al ASADI

College of Computer technologies  
Univ. of Babylon-Iraq  
[tawfiqasadi63@yahoo.com](mailto:tawfiqasadi63@yahoo.com)

cell. 009647801463183

---

### Abstract:

Geographic Information System(GIS) are usually classified into raster, vector, and raster – vector systems. The research deals with proposing new algorithm called **GIS-Span** to measure the partial and full geometrical similarity between map objects. The algorithm is used for clustering the vector features of GIS. The vector data are usually stored in data files called shape files. These files contains the (point, lines, polygons,...,etc). The extracted data is then stored in a dataset to be processed by the proposed algorithm to discover the full and partial similarities among map objects to assist the clustering and analysis of map data. It deals with clustering the polylines and polygonal data according to their geometrical similarity.

---

### Keywords :

Graph clustering algorithms, Graph mining, GIS data analysis, Mining GIS data ,Spatial data mining, geometrical similarity.

### Introduction

During the study of related researches to Spatial clustering which deals with spatial data that is generally organized in the form of a set of *points* or *polygons*[Deep et.al 04] we have found that mining of GIS spatial data is fertile for more research to be more convenient to GIS applications especially the vector data that

deals with GIS Graph data which deals with vector map elements like points, lines, polygons,..., etc , and the huge amount of GIS-based applications make it essential to improve graph mining techniques to extract the spatial knowledge embedded in that data to support GIS applications to be installed in new data analysis fields to support geographers and GIS

users to investigate the similarity among the spaghetti of map objects, and therefore can cluster them

The proposed algorithm is intended to extract spatial knowledge based on the similarity among map objects to cluster them accordingly. The research is consisting from the following main steps. Extracting the graph features included in GIS-Shape files, design and implement dataset to store the extracted data, propose a new algorithm called **GIS-Span** to discover the fully and partially similar vector map objects, cluster the resultants similar graph features, re-visualize the discovered object on the real map giving them a distancing effects.

### **1.Related works**

Recently the GIS graph mining techniques attracts more researching efforts in both raster and vector data mining, but we are interested in the mining of graph data which concerned with knowledge discovery in GIS-vector data, the extracted data are intended to enhance the GIS-query system to be more responding to user queries.

In 2009 Paul Van Dooren Catherine Fraikin [C. FRA 02] studied the Similarity matrices for coloured graphs but only between nodes of two graphs, to the case of coloured graphs, where the colouring is either on the nodes or on the edges of both graphs. The proposed method tries to find the optimal matching between the nodes or edges of both graphs but only performs the comparison when their colours are the same.

In 2010 Akshara Pande et. al. designed an information system to work with data referenced by spatial geographical coordinates. They are detecting design patterns so that it can be used as a conceptual tool to cope with recurrent problems appearing in the GIS domain [aksh 01].

In 2011[Deep 04] used districting which is the process of dividing a geographic space or region of spatial units often represented as polygons into smaller sub-regions or districts. As such, it can be viewed as a set partitioning problem, i.e. the problem is to cluster the entire set of spatial polygons into groups according to the spatial properties involved, redistricting is like to spatial clustering which can be done using spatially-flavoured constraints such as spatial contiguity and compactness which are spatial features in GIS data.

In 2010 a study of clustering algorithms of area geographical entities based on geometric shape similarity was made and the researchers presented a similarity criterion of line segments shape and a criterion of area geographical entities comprehensively utilizing distance and geometric shape similarity. Clustering algorithms based on these criteria are more suitable for clustering analysis of area geographical entities[Chen et. al 03]. The extracted knowledge can be used to explore deeper level knowledge combined with other data mining methods and to improve the efficiency and quality of data mining [ Ilar et.al 06]. We studied clustering algorithms of area geographical entities based on geometry

similarity. They also presented a similarity criterion of line segments shape and a criterion of area entities similarity comprehensive utilizing distance and geometry similarity. Clustering algorithms based on these criterions are more suitable for clustering analysis of area geographical entities[Chen et. al 03] .

### 2.Graph Mining Algorithms

Graph mining algorithms deals with discovering the knowledge embedded GIS vector data which are mostly stored in GIS shape files. These GIS data structures are stored as one geographic feature per shape file like point feature layer, multipoint feature layer, and polygon feature layer [Kang 07] [saya 08].

We used the ESRI shape file technical report to implement a *software module* to extract the *polyline* and *polygon* features of ESRI-Based maps and designed a dataset to contain the extracted data[ESRI 05] .Fig.2.1 shows the structure of ESRI shape file,it consists of three files for each feature layer(\*.shp,\*.shx, and \*.dbf) the first one containing the geographical coordinates ,the second is an index between each feature and it's description in the database file, and the last containing a description on each polygon feature.

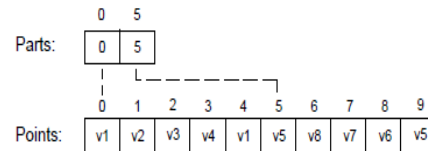


Table 1  
Polygon Record Contents

Position	Field	Value	Type	Number	Byte Order
Byte 0	Shape Type	5	Integer	1	Little
Byte 4	Box	Box	Double	4	Little
Byte 36	NumParts	NumParts	Integer	1	Little
Byte 40	NumPoints	NumPoints	Integer	1	Little
Byte 44	Parts	Parts	Integer	NumParts	Little
Byte X	Points	Points	Point	NumPoints	Little

Note: X = 44 + 4 \* NumParts

Fig.1 the shape file structure for Polygon feature.

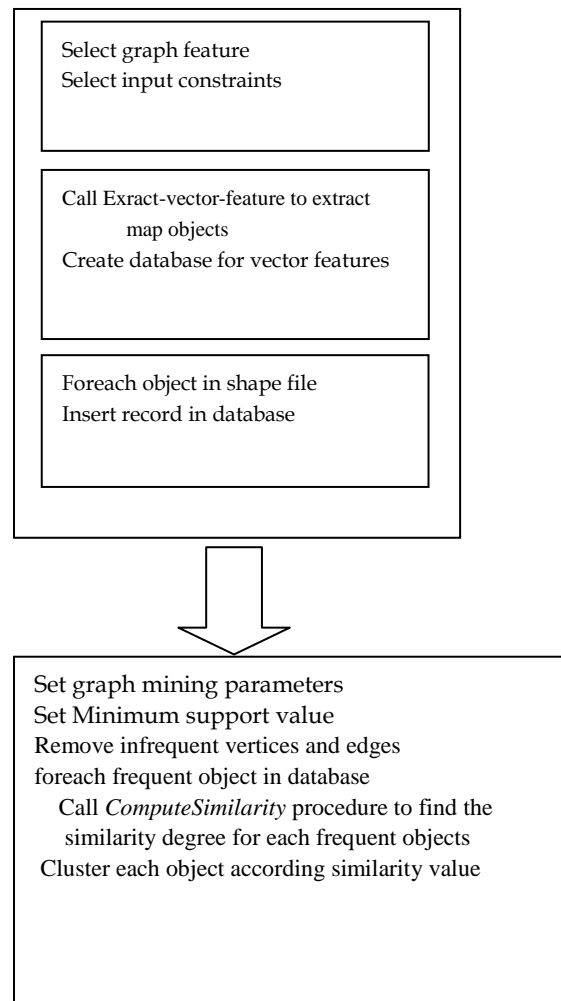


Fig2. The block diagram of the main steps of the proposal.

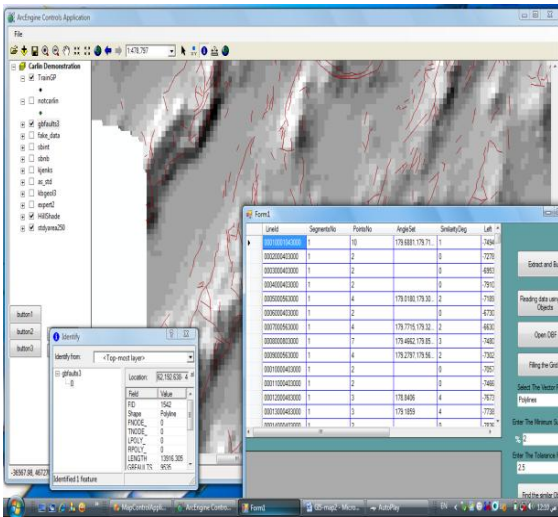


Fig.3. The dataset extracted from the shape file

Fig.3 shows part of the dataset which is built to hold the spatial-vector data that extracted from the GIS shape file by our module *Extract-vector-feature*. This data will be passed to the data mining algorithm **GIS-span** to extract the similarities between polygons and polylines map-objects.

Fig.4 shows part of the output of our Graph-mining algorithm *GIS-span* which is applied on ESRI map layer for the selected map objects features.

When we applied the proposed algorithm on the map depicted in Fig.4 we have found that it consists of 1674 objects and the is clustered to 74 clusters after excluding the objects that are not respond to the preprocessing constraints.

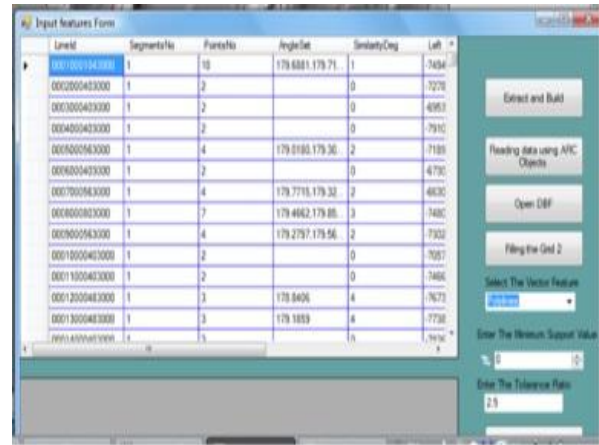


Fig.4 The dataset with similarity degrees

### Conclusions

In this paper we present a new algorithm for mining the GIS vector data (polylines, polygons, and ,etc).The algorithm clusters map objects based on the degree of geometrical similarity among map objects. The results of applying this algorithm on actual map data gives good clustering results with less computing costs because of the simple geometrical similarity measure.

The algorithm may need some enhancement to deal with mining all vector types of map objects. The algorithm is also capable to deal with vector data stored in shape files and not related to GIS system like auto CAD and other drawing tools to cluster their map objects as a type of knowledge discovery to help designers analysing their maps.

### References

1- Akshara Pande Manjari Gupta A.K. Tripathi, Design Pattern Mining for GIS Application Using Graph Matching Techniques 978-1-4244-5540-9/10,2010 IEEE.

- 2- C. Fraikin and P. Van Dooren, *Graph matching with type constraints*, ECC 07, Greece, July 2-5, 2007.
- 3- Chen Guang-xue, LI Xiao-zhou, Chen Qifeng, LI Xiao-zhou, *Clustering Algorithms for Area Geographical Entities in Spatial Data Mining*, 978-1-4244-5934-6/10/2010 IEEE, (FSKD 2010).
- 4- Deepti Joshi, Leen-Kiat Soh *Member, IEEE*, and Ashok Samal, *Member, IEEE*, Redistricting using Constrained Polygonal Clustering. Digital Object Identifier 10.1109/TKDE.2011.140.
- 5-ESRI Shapefile Technical Description An ESRI White Paper—July 1998 Copyright © 1997, 1998 Environmental Systems Research Institute, Inc.
- 6- Ilaria Bordino, Debora Donato, and Aristides Gionis Mining large networks with subgraph counting, 2008 Eighth IEEE International Conference on Data Mining
- 7- Kang-Tsung Chang, *Programming ArcObjects with VBA Second Edition*, Chapman & Hall/CRC Taylor & Francis Group 6000 Broken Sound Parkway NW, Suite 300 Boca Raton, FL 33487-2742 © 2008 by Taylor & Francis Group, LLC.
- 8-Sayan Ranu and Ambuj K. Singh. Graphsig: A scalable approach to mining significant subgraphs in large graph databases. In *ICDE*, 2009.