

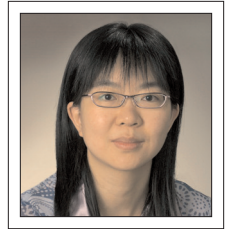
# USING CLUSTERING METHODS IN GEOSPATIAL INFORMATION SYSTEMS

Xin Wang and Jing Wang

Department of Geomatics Engineering, Schulich School of Engineering  
University of Calgary, Calgary, Alberta

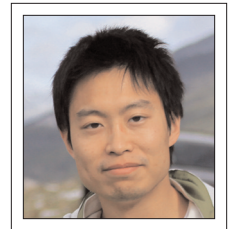
*Spatial clustering is the process of grouping similar objects based on their distance, connectivity, or relative density in space. It has been employed in the field of spatial analysis for years. In order to select the proper spatial clustering methods for geospatial information systems, we need to consider the characteristics of different clustering methods, relative to the objectives that we are trying to achieve. In this paper, we give a detailed discussion of different types of clustering methods from a data mining perspective. Analysis of the advantages and limitations of some classical clustering methods are given. Subsequently we discuss applying spatial clustering methods as part of geospatial information systems, with respect to distance functions, data models, non-spatial attributes and performance.*

*Le regroupement spatial est le processus visant à regrouper des objets similaires en fonction de leur distance, de leur connectivité ou de leur densité relative dans l'espace. Il est utilisé depuis des années dans le domaine de l'analyse spatiale. Afin de choisir les méthodes adéquates de regroupement spatial pour les systèmes d'information géospatiale, nous devons tenir compte des caractéristiques des diverses méthodes de regroupement relativement aux objectifs que nous tentons d'atteindre. Dans le présent article, nous faisons un exposé détaillé des différents types de méthodes de regroupement dans une perspective d'exploration de données. Nous présentons également une analyse des avantages et des limites de certaines méthodes classiques de regroupement. Subséquemment, nous examinons la question de l'application des méthodes de regroupement spatial dans le cadre des systèmes d'information géospatiale, en ce qui a trait aux fonctions de distance, aux modèles de données, aux attributs non spatiaux et à la performance.*



**Xin Wang**

[xcwang@ucalgary.ca](mailto:xcwang@ucalgary.ca)



**Jing Wang**

[wangjing@ucalgary.ca](mailto:wangjing@ucalgary.ca)

## 1. Introduction

A Geospatial Information System (GIS) is a computer-based information system for both managing geographical data and for using these data to solve spatial problems [Lo and Yeung 2007]. Rapid growth is occurring in the number and the size of GIS applications, including geo-marketing, traffic control, and environmental studies [Han et al. 2001]. *Spatial clustering* is the process of grouping similar objects based on their distance, connectivity, or relative density in space [Han et al. 2001]. This has been employed for spatial analysis over a number of years. Currently it is commonly used in such diverse fields as disease surveillance, spatial epidemiology, population genetics, landscape ecology, crime analysis, as well as in many other fields [Jacquez 2008]. Therefore, spatial clustering is potentially a very useful tool for spatial analysis in GIS.

Various clustering methods have been proposed in both the area of spatial data mining and the area of geospatial analysis [Agrawl et al. 1998; Ester et al. 1996; Estivill-Castro and Lee 2000a; Estivill-Castro and Lee 2000b; Gaffney et al. 2006; Gaffney and Smyth 1999; Kaymak and Setnes 2002; Klawonn and Hoppner 2003; Lee et al. 2007; Martino et al. 2008;

Nanni and Pedreschi 2006; Mu and Wang 2008; Ng and Han 1994; Sander et al. 1998; Stefanakis 2007; Tung et al. 2001a; Tung et al. 2001b; Wang and Hamilton 2003; Wang et al. 2004; Wang et al. 1997; Zaïane and Lee 2002; Zhang et al. 1996; Zhou et al. 2005]. In spatial data mining, clustering methods can be classified into different categories. In terms of the techniques adopted to define clusters, clustering algorithms have been categorized into four broad categories: hierarchical, partitional, density-based, and grid-based [Han et al. 2001]. *Hierarchical clustering methods* group objects into a tree-like structure that progressively reduces the search space. *Partitional clustering methods* partition the points into clusters, such that the points in a cluster are more similar to each other than to points in different clusters. *Density-based clustering methods* can find arbitrarily shaped clusters that are 'grown' from seeds, and established once the density in the clusters' neighborhoods exceed certain density thresholds. *Grid-based clustering methods* divide the information spaces into a finite number of grid cells and then cluster objects based on this structure.