Dynamics of Continuous, Discrete and Impulsive Systems
Series **B:** *Applications & Algorithms* **14 (2007)** 825-840
 http://www.watam.org

A TOPOLOGICAL FRAMEWORK FOR SPATIAL CLUSTERING ON MANIFOLD

Xiyu Liu[1], Hong Liu[2] and Hui Wang[3]

[1]School of Management & Economics
Shandong Normal University, Jinan, Shandong 250014, P. R. China
Email: xyliu@sdnu.edu.cn

[2]School of Information Science and Engineering
Shandong Normal University, Jinan, Shandong 250014, P. R. China
Email: hongliu@sdnu.edu.cn

[3]School of Information Science and Engineering
Shandong Normal University, Jinan, Shandong 250014, P. R. China
Email: bzwdwh@163.com

Abstract. Clustering analysis has become an important tool in knowledge engineering and business intelligence. There is rapidly growing number of clustering algorithms, however, one of the most important clustering algorithms is the density based spatial clustering, and its various variations. These algorithms are either based on the density connected chains, or weighted density, or the gravity based connected sets. They seem similar in idea but a unified presentation does not exist. The purpose of this paper is to give a topological view to the spatial clustering focusing on density based clustering with its variations. Moreover, we present this new framework in manifold. We propose the definition of cluster topology; define clusters as various kinds of topological connected sets. More in this paper, the topological cluster forming algorithms are presented. Then we reform some typical algorithms with the new topological view. Examples are given to show the theories.

Keywords. Spatial clustering, manifold learning, topology, connected sets, algorithms, density based clustering.

AMS (MOS) subject classification: 68T35.

1 Introduction

There have been many papers focusing on specific algorithms of data mining and clustering analysis, applications, especially in the area of business intelligence, see [3] for example. Data mining is a step in the KDD process consisting of the application of data analysis and discovery algorithms that, under acceptable computational efficiency limitations, produce a particular enumeration of patterns over the data (Fayyad et al [11]). Clustering, i.e. grouping the objects of a database into meaningful subclasses, is one of the major data mining methods (Matheus et al [1]). Cluster Analysis can be categorized as a branch of statistics. In the past three decades, it