Topological Decomposition and Heuristics for High Speed Clustering of Complex Networks

Faraz Zaidi*, Guy Melançon**

*Karachi Institute of Economics and Technology (KIET)
Korangi Creek, Karachi, 75190, Pakistan
faraz@pafkiet.edu.pk
**CNRS UMR 5800 LaBRI & INRIA Bordeaux - Sud Ouest
351, cours de la Libération, 33405 Talence cedex, FRANCE
guy.melancon@labri.fr

Abstract. With the exponential growth in the size of data and networks, development of new and fast techniques to analyze and explore these networks is becoming a necessity. Moreover the emergence of scale free and small world properties in real world networks has stimulated lots of activity in the field of network analysis and data mining. Clustering remains a fundamental technique to explore and organize these networks. A challenging problem is to find a clustering algorithm that works well in terms of clustering quality and is efficient in terms of time complexity.

In this paper, we propose a fast clustering algorithm which combines some heuristics with a Topological Decomposition to obtain a clustering. The algorithm which we call Topological Decomposition and Heuristics for Clustering (TDHC) is highly efficient in terms of asymptotic time complexity as compared to other existing algorithms in the literature. We also introduce a number of Heuristics to complement the clustering algorithm which increases the speed of the clustering process maintaining the high quality of clustering. We show the effectiveness of the proposed clustering method on different real world data sets and compare its results with well known clustering algorithms.

1 Introduction

Most real world systems take the form of networks where a set of nodes and edges might be used to represent these networks. Examples include social networks, metabolic networks, food web, transport networks (Newman (2003)). Clustering remains an important technique towards the better exploration and organization of these networks. In terms of networks representing real world data, a cluster can be defined as a group of nodes which are similar or connected in some predefined sense and dissimilar to nodes belonging to the other clusters (Schaeffer (2007)). Detection of clusters has a wide range of applications in various fields. For example, in social networks, clustering could lead us towards a better comprehension of the interactions taking place between people, or for biological networks, a useful application of clustering is in the identification of biomarkers in a protein-protein interaction network.