© 2013 FIZ Karlsruhe

io-port 06055452

Peter, S.J.

Structural similarity clustering for detection of outliers and hubs using minimum spanning tree.

J. Discrete Math. Sci. Cryptography 14, No. 2, 167-183 (2011).

Summary: Clustering is a process of discovering groups of objects such that the objects of the same group are similar, and objects belonging to diff erent groups are dissimilar. A number of clustering algorithms exist that can solve the problem of clustering. Minimum spanning tree based clustering algorithm is capable of detecting clusters with irregular boundaries. MST based Clustering is an important task for the discovery of underlying structures in graph. In data mining detection of anomalous pattern in data is more interesting than detecting inliers. Many algorithms find clusters by maximizing the number of intra-cluster edges. While such algorithms find useful and interesting structures, they tend to fail to identify and isolate two kinds of vertices that play special roles - vertices that bridge clusters (hubs) and vertices that are marginally connected to clusters (outliers). Identifying hubs is useful for applications such as viral marketing and epidemiology since hubs are responsible for spreading ideas and disease. In contrast, outliers have little or no infl uence, and many may be isolated as noise in the data. In this paper we propose a novel algorithm called Structural Similarity Clustering Algorithm using Minimum Spanning Tree (SSCAMST), which detect clusters, outliers and hubs in graph. The key feature of our algorithm is it finds noise-free/error-free clusters.

Keywords: Euclidean minimum spanning tree; clustering; structural similarity; outliers; hub
http://www.connectjournals.com/file_html_pdf/1080602H_05_JDMSC_T14_14-2_pp167-183A.pdf