Nonparametric search algorithm for outliers and hubs using minimum spanning tree

S. John Peter*

Department of Computer Science and Research Center
St. Xavier’s College
Palayamkottai
Tamil Nadu
India

Abstract

In data mining detection of anomalous pattern in data is more interesting than detecting inliers. A novel breadth-first search based algorithm for detection of outliers and hubs is proposed. Minimum Spanning Tree based Clustering is an important task for the discovery of underlying structures in graph. 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). In this paper, we model hubs as high-degree nodes having marginal distance (sparseness) with neighboring nodes. 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 influence, and many may be isolated as noise in the data. In this paper we propose an algorithm called Minimum Spanning Tree based Search Algorithm for Outliers and Hubs (MSTSOH), which detect outliers and hubs in graph.

Keywords: Euclidean minimum spanning tree; Degree; Outliers; Hub, Sparseness;