
Detecting Localized Categorical Attributes on Graphs

Abstract:

Do users from Carnegie Mellon University form social communities on Facebook? Do signal processing researchers from tightly collaborate with each other? Do Chinese restaurants in Manhattan cluster together? These seemingly different problems share a common structure: an attribute that may be localized on a graph. In other words, nodes activated by an attribute form a subgraph that can be easily separated from other nodes. In this paper, we thus focus on the task of detecting localized attributes on a graph. We are particularly interested in categorical attributes such as attributes in online social networks, ratings in recommender systems and viruses in cyber-physical systems because they are widely used in numerous data mining applications. To solve the task, we formulate a statistical hypothesis testing problem to decide whether a given attribute is localized or not. We propose two statistics: graph wavelet statistic and graph scan statistic, both of which are provably effective in detecting localized attributes. We validate the robustness of the proposed statistics on both simulated data and two real-world applications: high air-pollution detection and keyword ranking in a co-authorship network collected from IEEE Xplore. Experimental results show that the proposed graph wavelet statistic and graph scan statistic are effective and efficient.