Presentation 1

Graphs are ubiquitous and can model complex relationships

With the proliferation of

graph applications, such as social networks, information

networks, web search, collaboration networks, E-commerce

networks, communication networks, and biology

Graph clustering

– Group vertices into clusters: dense intra connection and sparse inter

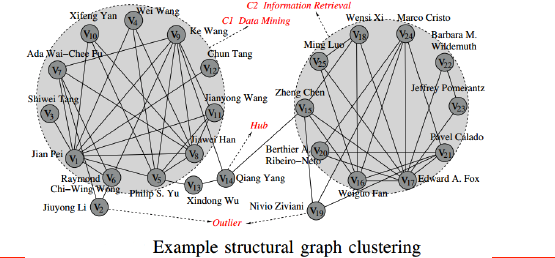
connection

Structural Graph Clustering

• SCAN [Xu+, KDD’07]

– Identifies clusters, hubs, and outliers at the same time

– Mimics DBSCAN [Ester+, KDD’96] for clustering spatial data



scan iterate through all vertices that have not been assigned to clusters.

For each such core vertex u, it creates a new cluster C initially containing

only u and iteratively adds into C all vertices that are structure-similar to a core vertex in C.

Thus, SCAN computes the structural similarity for every pair of adjacent vertices in G ( i.e., for every .u; v. 2 E); this incurs high computational cost and makes it not scalable to large graphs.

two main challenges:

1 How to reduce the number of structural similarity computations?

2 How to efficiently check whether two vertices are structure-similar to each other?

Two – step paradigm

1. cluster all cover vertices by partitioning

them into clusters, and then cluster non-core vertices by assigning each non-core vertex v to the same clusters as its neighbors that are core vertices and are structure-similar to v.

2. we propose a pruned SCAN (denoted pSCAN) approach for structural graph clustering.

We incrementally maintain an effective-degree ed.v. and a similar-degree sd.v. for every vertex v 2 V , satisfying ed.v.  sd.v..(提升分辨效率)

Then, the clusters of core vertices are computed

based on the transitive property that, two core vertices, u and

v, are in the same cluster if .u; v. 2 E and they are structuresimilar.

Moreover, we avoid computing the structural similarity

between core vertices u and v if they have already been

assigned to the same cluster. Consequently, pSCAN saves a lot

of structural similarity computations.(减少计算)