Efficient computation of the weighted clustering coefficient

Stefano Leonardi, Sapienza University of Rome

The clustering coefficient of an unweighted network has been extensively used to quantify how tightly connected is the neighbor around a node and it has been widely adopted as an important measure for assessing the quality of nodes in a social network. The computation of the clustering coefficient is a challenging computational task that requires to count the number of triangles in the graph. Several recent works proposed efficient sampling, streaming and MapReduce algorithms that allow to overcome this computational bottleneck.

As a matter of fact, the intensity of the interaction between nodes, that is usually represented with weights on the edges of the graph, is also an important measure of the statistical cohesiveness of a network.
Recently various notions of weighted clustering coefficient have been proposed but so far all those techniques are hard to implement on large-scale graphs.

In this work we first show how standard sampling techniques can be used to obtain efficient estimator for the most commonly used measures of weighted clustering coefficient. Furthermore we also propose a novel graph-theoretic notion of clustering coefficient in weighted networks. Based on the observation that edges with large weights are more likely to play a role in the social network, we give an interpretation of the weight of an edge as the probability of existence of the edge in the graph. We therefore define the weighted clustering coefficient as the expected clustering coefficient on a family of random graphs. We show that our notion of weighted clustering coefficient can be computed in polynomial time and that can be efficiently approximated with sampling algorithms. We finally show experimentally interesting properties of the weighted clustering coefficient and we prove the accuracy and efficiency of our estimators.

Joint work with Silvio Lattanzi (Google NYC)