ABSTRACT

Real world complex networks such as social networks, biological networks usually exhibit inhomogeneity, resulting in densely interconnected nodes, *communities*, which play an important functional role in the original system. Analyzing such communities in large networks has rapidly become one of the major topics in complex networks. In this thesis, we study four different aspects related to communities: (i) analyzing dependency of existing community detection algorithms on vertex ordering, (ii) quantifying the extent of belongingness of nodes in a community; (iii) unfolding the evolution dynamics of communities in a real-world network; (iv) designing different community-based applications.

While observing the variability in the outputs obtained from community finding algorithms, we notice that some groups of vertices always remain together despite any vertex ordering. We call these groups *constant communities*. We characterize constant communities and show that prior detection of such constant communities improves the performance of a community detection algorithm and reduces the variability of the output.

Then we quantify the membership of a vertex within a community by formulating two vertex-centric metrics: *permanence* (*Perm*) for non-overlapping communities and *overlapping permanence* (*OPerm*) for overlapping communities. We show the effectiveness of these metrics by comparing the results with the ground-truth community structure. We also design two algorithms, *MaxPerm* and *MaxOPerm*, to detect non-overlapping and overlapping communities respectively.

We crawl a massive publication dataset of computer science domain constituting more that 1.5 million scientific articles. We tag each paper by its related research field(s) that act as ground-truth communities. Then we study the temporal interactions of these communities through citations over the last fifty years and unfold the landscape of scientific paradigms. Moreover, we quantify the degree of interdisciplinarity of each field and describe the evolutionary landscape of the interdisciplinary fields over the years.

Finally, we study the citation growth of a paper after publication and discover six distinct categories of citation profile. This observation leads us to adopt *stratified learning* approach in a prediction task, whereby, we propose a two-stage model to predict the future citation count of a paper after a certain time period of its publication. We also design *FeRoSA*, a framework of faceted recommendation for scientific articles that apart from ensuring quality retrieval also efficiently arranges the recommended papers into different facets (categories) that indeed show how these recommendations are related to the query paper.

Keywords: Community analysis, Permanence, Community detection algorithms, Community evolution, Citation networks, Faceted recommendation system