City Clustering From Road Network Graph

Wenjian Hu Jian Xu

Background

As we all know, there has been a large amount of work devoted to defining and identifying clusters or communities in social and information networks, i.e., in graphs in which the nodes represent underlying social entities and the edges represent some sort of interaction between pairs of nodes [1]. Particularly, there is a growing need to develop urban computing and analysis tools to guide the orderly development of cities. So in this project, we will be working on clustering cities from road network graphs.

Rather than using hierarchical clustering, we will use a quite different approach based from graph partitioning where a partition of the graph is determined by optimizing a merit function [1]. Defying intersections as nodes and the roads connecting these intersections as undirected edges (without edge weight information), we will deduce different clusters (cities) by calculating the conductance of different sets of nodes. The road network graph is from SNAP database [2].

However, finding cuts with exactly minimal conductance is NP-hard, thus, we will use several algorithms with good approximation performance guarantees. First, we will apply the spectral method, which uses an eigenvector of the graphs Laplacian matrix to find a cut whose conductance is no bigger than ϕ if the graph actually contains a cut with conductance $O(\phi 2)$ [3,4,5]. Also, we will use Metis+MQI, which consists of using the popular graph partitioning package Metis [6] followed by a flow-based MQI post-processing [7].

Very relevant to our work is that of Kannan, Vempala, and Vetta [8], who analyze spectral algorithms and describe a community concept in terms of a bicriterion depending on the conductance of the communities and the relative weight of inter-community edges. Flake, Tarjan, and Tsioutsiouliklis [9] introduce a similar bicriterion that is based on network flow ideas, and Flake et al. [10, 11] defined a community as a set of nodes that has more intra-edges than inter-edges. Similar edge-counting ideas were used by Radicchi et al. [12] to define and apply the notions of a strong community and a weak community.

In this project, the fundamental question we wish to answer is how to efficiently find different clusters from an undirected graph which has no edge weight information.

Reference

[1]. J. Leskovec, K. Lang, A. Dasgupta, M. Mahoney. Community Structure in Large Networks: Natural Cluster Sizes and the Absence of Large Well-Defined Clusters. Internet Mathematics 6(1) 29–123, 2009.
[2]. $https://snap.stanford.edu/data/\#road$.
[3]. J. Cheeger. A lower bound for the smallest eigenvalue of the laplacian. In Problems in Analysis, Papers dedicated to Salomon Bochner, pages 195199. Princeton University Press, 1969.
[4]. W.E. Donath and A.J. Hoffman. Algorithms for partitioning graphs and computer logic based on eigenvectors of connection matrices. IBM Technical Disclosure Bulletin, 15(3):938944, 1972.
[5]. M. Fiedler. Algebraic connectivity of graphs. Czechoslovak Mathematical Journal, 23(98):298305, 1973.
[6]. G. Karypis and V. Kumar. A fast and high quality multilevel scheme for partitioning irregular graphs. SIAM Journal on Scientific Computing, 20:359392, 1998.
[7]. K. Lang and S. Rao. A flow-based method for improving the expansion or conductance of graph cuts. In IPCO 04: Proceedings of the 10th International IPCO Conference on Integer Programming and Combinatorial Optimization, pages 325337, 2004.
[8]. R. Kannan, S. Vempala, and A. Vetta. On clusterings: Good, bad and spectral. Journal of the ACM, 51(3):497515, 2004.
[9]. G.W. Flake, R.E. Tarjan, and K.Tsioutsiouliklis. Graph clustering and minimum cut trees. Internet Mathematics, 1(4):385408, 2003.
[10]. G.W. Flake, S. Lawrence, and C.L. Giles. Efficient identification of web communities. In KDD

00: Proceedings of the 6th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, pages 150160, 2000.
[11]. G.W. Flake, S. Lawrence, C.L. Giles, and F.M. Coetzee. Self-organization and identification of web communities. Computer, 35(3):6671, 2002.
[12]. F. Radicchi, C. Castellano, F. Cecconi, V. Loreto, and D. Parisi. Defining and identifying communities in networks. Proceedings of the National Academy of Sciences of the United States of America, 101(9):26582663, 2004.