BACKBONE STRUCTURE OF HIERARCHICAL NETWORK PARTITIONING

Zizhen Chen, David W. Matula

SIAM Workshop on Network Science 2018
July 12-13 · Portland

Extended Abstract

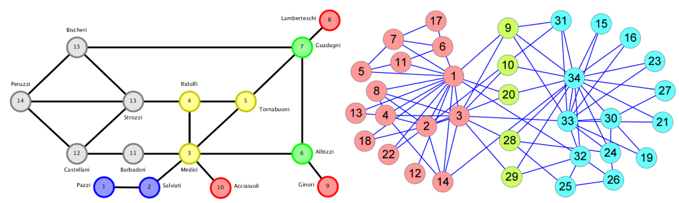
We review the foundations of a hierarchical network partitioning method based on the fundamental concept of density (i.e., sparse cuts separating dense components) and implemented via a duality to peer-to-peer network flow between all node pairs. The methodology is conceptualized by considering traffic flow in a road network with traffic demands between all cities. In periods of high flow "congestion pricing" might be imposed to lessen flow through a bottleneck set of roads, such as those into a city's downtown core. For regions of somewhat less congestion successively smaller prices might be imposed across boundary accesses. To determine this hierarchy of boundaries we employ the Maximum Concurrent Flow Problem (MCFP) which can be formulated as a Linear Program (LP) with maximin objective to maximize the minimum flow throughput guaranteed between all node pairs subject to the paths with flow sharing capacity of the edges [1, 7, 9, 10, 11, 12, 13]. Throughput is the ratio of the flow delivered between a node pair in comparison to the node pair's corresponding demand. Employing LP duality, the optimal throughput is shown to determine a critically saturated separating set of edges partitioning the network into component parts, where all edges of the components have strictly positive residual capacity.

The hierarchical MCFP is then formulated to further maximize throughput in the residual capacity components determining a second throughput level and set of critical edges. Iterating further employing a series of LP's, a series of throughput levels is determined until all edges are critical, yielding a stratification portrayed in hierarchical partitioning theory as a dendrogram.

Each LP of the HMCFP sequence of LP's determines in polynomially bounded time either a sparsest cut or a sparse grid, the latter a multipart partition proven to have at least five parts. The result is consistent with the sparsest cut problem by itself being NP-hard [8, 12, 6].

Most real world networks from many fields are sparse. Numerous examples having a power law distribution of degrees have been cited [3, 2], but the generality of this observation is problematic [5]. Importantly, we found that networks having ground truth community structure typically yield successive levels of marginally sparsest cuts before encountering a gridlock fragmentation. We utilized well-known small social networks (Figure 1) to illustrate the process and exhibit the component communities by characterizing the "backbone" concept of the network.

In Figure 1a, the 15-node Florentine Families network [4] is represented as a graph. Taking edge capacities and node pair demands both as unity provides a density based hierarchy by the HMCFP. The backbone can be defined as a "graph minor", with the components at any level formed by contraction of the node sets.



(a) Florentine Family Network (b) Karate Clube Network

Figure 1: Two Small Networks Illustrating the Procedure

An edge of the contracted graph after k cuts can be labeled by the set of j cuts that include that edge $1 \leq j \leq k$. A path of two or more edges between the end nodes of an edge that is cut by the same set of j cuts is termed a "back channel" between the nodes. The edges with no back channels between their end nodes are "backbone edges" and characterize the subgraph of the contracted graph termed a "backbone". Backbone edges provide the excess capacity to absorb the additional flow between end node pairs to maximize the concurrent flow at that level.

The minimal set of cuts crossed by a path between two components of the backbone provides a distance between the components. The concept of distance between components identifies weak and strong relations between the components. For community detection applications, this adds to the understanding of relations between communities augmenting the original data on relations between individual node pairs. Figure 2 shows the Florentine Family backbones from cut level 1 to level 5.

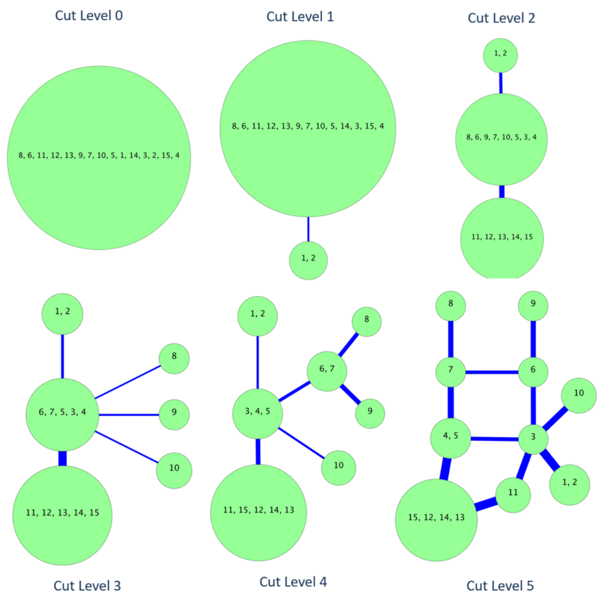


Figure 2: Florentine Backbones Level 1 to Level 5

The backbone shows the structure of the network by preserving the minimal set of critical edges that guide the flow routing. It is typically a sparse planar graph for real world networks. Figure 1b shows a denser social network: Zachary's Karate Club [14] containing 34 nodes with 78 edges. Figure 3 is the Karate Club backbone at a final cut level before gridlock occurs. The backbone provides a simplified visualization of the denser components and their interrelations. Besides the hierarchical component relational structure, the backbone may also provide a seriation of the corresponding adjacency matrix. Figure 4a shows a seriated adjacency matrix compared to Figure 4b which is a randomized adjacency matrix of the Karate Club network.

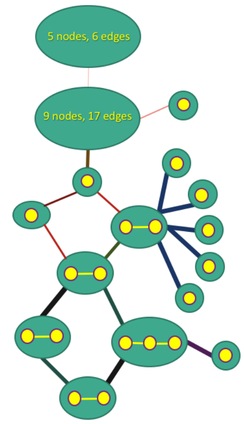
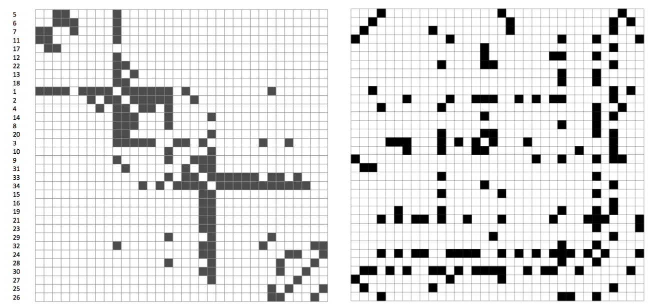


Figure 3: Karate Club Network Final Backbone



(a) Seriated Matrix (b) Random Matrix

Figure 4: Adjacency Matrices of Karate Club Network

References

[1] M. Allalouf and Y. Shavitt. Maximum flow routing with weighted max-min fairness. In *Quality of Service in the Emerging Networking Panorama*, pages 278–287. Springer, 2004.

[2] A. Barabási and M. PÃ3sfai. *Network Science*. Cambridge University Press, 2016.

[3] A.-L. Barabási and R. Albert. Emergence of scaling in random networks. *science*, 286(5439):509–512, 1999.

[4] R. L. Breiger and P. E. Pattison. Cumulated social roles: The duality of persons and their algebras. *Social networks*, 8(3):215–256, 1986.

[5] A. D. Broido and A. Clauset. Scale-free networks are rare. *arXiv preprint arXiv:1801.03400*, 2018.

[6] S.-W. Chiou. A combinatorial approximation algorithm for concurrent flow problem and its application. *Computers & operations research*, 32(4):1007–1035, 2005.

[7] Y. Dong, E. V. Olinick, T. Jason Kratz, and D. W. Matula. A compact linear programming formulation of the maximum concurrent flow problem. *Networks*, 65(1):68–87, 2015.

[8] E. Gourdin. A mixed integer model for the sparsest cut problem. *Electronic Notes in Discrete Mathematics*, 36:111–118, 2010.

[9] T. Leighton and S. Rao. Multicommodity max-flow min-cut theorems and their use in designing approximation algorithms. *Journal of the ACM (JACM)*, 46(6):787–832, 1999.

[10] D. W. Matula. Cluster validity by concurrent chaining. In *Numerical Taxonomy*, pages 156–166. Springer, 1983.

[11] D. W. Matula. Divisive vs. agglomerative vverage linkage hierarchical clustering. In W. Gaul and M. Schader, editors, *Classification as a Tool of Research*. North-Holland, 6 1985.

[12] D. W. Matula and F. Shahrokhi. Sparsest cuts and bottlenecks in graphs. *Discrete Applied Mathematics*, 27(1-2):113–123, 1990.

[13] F. Shahrokhi and D. W. Matula. The maximum concurrent flow problem. *Journal of the ACM (JACM)*, 37(2):318–334, 1990.

[14] W. W. Zachary. An information flow model for conflict and fission in small groups. *Journal of anthropological research*, 33(4):452–473, 1977.