# THE EVOLUTION OF FLOW-BASED HIERARCHY IN NETWORKS

*Zizhen Chen, David W. Matula and Eli V. Olinick*

SIAM Workshop on Network Science 2017
July 13-14 · Pittsburgh

## Extended Abstract

Divisive (top-down) graph-decomposition methods based on edge centralities (e.g., Girvan and Newman 2002 [4], and Fortunato et al. [3]) have been developed to avoid deficiencies associated with agglomerative clustering methods [5]. These community detection methods iteratively identify and remove high centrality edges to produce a hierarchical decomposition of the graph into clusters (connected components).

The divisive algorithm investigated in this paper is based on "maximin" concurrent flow and its dual sparsest cut. Formally, the *maximum concurrent flow problem* (MCFP) is a maximum network flow problem in which every pair of nodes can send and receive flow concurrently. The term *throughput* is defined to be the ratio of the flow supplied between a pair of nodes to the given demand for that pair. The objective of the MCFP is to maximize the throughput, which must be the same for all pairs of nodes, subject to fixed capacity constraints on the edges [1].

A canonical MCFP solution is characterized by a maximal set of "slack" edges with residual flow capacity identifying a partition into connected components of slack edges. The complementary "critical edges" are saturated with flow by any MCFP solution, typically comprising an edge "cut set" forming a bipartition of the graph. The hierarchical MCFP (HMCFP) then further maximizes the common throughput between all node pairs connected by a path of slack edges determining a second throughput level and second set of critical edges bifurcating all the nodes. Iterating further, a series of throughput levels is determined until all edges are critical, yielding a hierarchical stratification portrayed as a *dendrogram* [6].

Real world networks such as the 15-node Florentine Families network [2] are represented as graphs. Taking edge capacities and node pair demands both as unity provides a density based hierarchy by the HMCFP. For the HMCFP at every throughput level, the slack edges at that level identify a partition into component "cluster nodes". The cluster node partition at each level identifies a graph contraction.

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 backbones for the Florentine Families graph HMCF partitions into 3 and 10 parts are shown in Figure 1. The backbones visually display relationships between clusters at each level and introduce a "distance" between clusters.

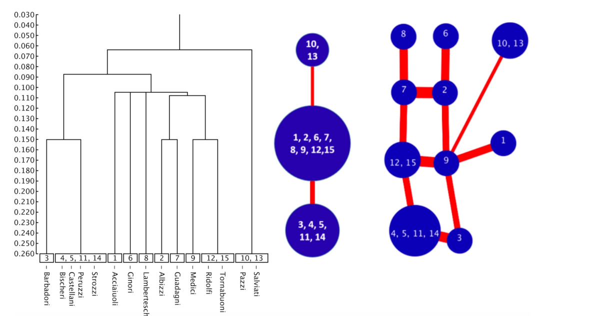


Figure 1: Florentine Dendrogram and two Backbones

## References

[1] P.-O. Bauguion, W. Ben-Ameur, and E. Gourdin. Efficient algorithms for the maximum concurrent flow problem. *Networks*, 65(1):56–67, 2015.

[2] C. L. DuBois. UCI network data repository. `http://networkdata.ics.uci.edu`, 2008.

[3] S. Fortunato, V. Latora, and M. Marchiori. Method to find community structures based on information centrality. *Phys. Rev. E*, 70:056104, Nov 2004.

[4] M. Girvan and M. E. J. Newman. Community structure in social and biological networks. *Proceedings of the National Academy of Sciences of the United States of America*, 99:7821–7826, 2002.

[5] C. F. Mann, D. W. Matula, and E. V. Olinick. The use of sparsest cuts to reveal the hierarchical community structure of social networks. *Social Networks*, 30(3):223–234, 2008.

[6] P. Sneath and R. Sokal. *Numerical Taxonomy. The Principles and Practice of Numerical Classification.* Freeman, 1973.