The ultrametric backbone is the union of all minimum spanning forests

Keywords: Shortest paths, network backbone, minimum spanning trees, sparsification

Extended Abstract

Many problems in network science and graph theory, such as predicting links [1], optimizing traversal [2], locating central (or redundant) nodes and edges [3], identifying primary transmission modes in spreading dynamics and community structure [4], or predicting the size of cascades when nodes or edges are attacked, depend strongly on the structure of shortest paths [5]. Often the length of a path is computed as the sum of its edge weights, but the underlying system or process may suggest other choices, such as multiplying the edge weights or taking only the largest edge weight. The method of aggregating edge weights determines the distances between nodes, and which paths are shortest in the context of a specific optimization problem. The aggregation operation encodes the cost of indirect associations or interactions. Moreover, other methods beyond shortest paths, such as diffusion and resistance distances, are possible to aggregate indirect associations in networks [6, 7]. A unifying framework to study families of algebraically consistent edge weighting and path aggregation that quantifies node-to-node distance in weighted graphs is provided by the distance closure [8]. In this framework, applying triangular metric space operations [9, 10] leads to general algebraic definitions of network distances including shortest path distance, diffusion distance, and resistance as the closure of an algebraic structure [8].

Minimum spanning trees (MST) and forests are powerful sparsification techniques that remove cycles from weighted graphs to minimize total edge weight while preserving node reachability, with applications in computer science, network science, and graph theory. Despite their utility and ubiquity, they have several limitations, including that they are only defined for undirected networks, they significantly alter dynamics on networks, and they do not generally preserve important network features such as shortest distances, shortest path distribution, and community structure. In contrast, *distance backbones*, which are subgraphs formed by all edges that obey a generalized triangle inequality, are well defined in directed and undirected graphs (via the distance closure framework) and preserve those and other important network features [3,4].

The distance backbone of a graph is defined with respect to a specified path-length operator that aggregates weights along a path to define its length, thereby associating a cost to indirect connections. The backbone is the union of all shortest paths between each pair of nodes according to the specified operator. One such operator, the **max** function, computes the length of a path as the largest weight of the edges that compose it (a weakest link criterion). It is the only operator that yields an algebraic structure for computing shortest paths that is consistent with De Morgan's laws. Applying this operator yields the *ultrametric backbone* of a graph in that (semi-triangular) edges whose weights are larger than the length of an indirect path connecting the same nodes (i.e. those that break the generalized triangle inequality based on max as a path-length operator) are removed. We demonstrate that the ultrametric backbone is the union of minimum spanning forests in undirected graphs and provides a new generalization of *minimum spanning trees* to directed graphs that, unlike *minimum equivalent graphs* and *minimum spanning arborescences*, preserves all **max-min** shortest paths and De Morgan's law consistency.

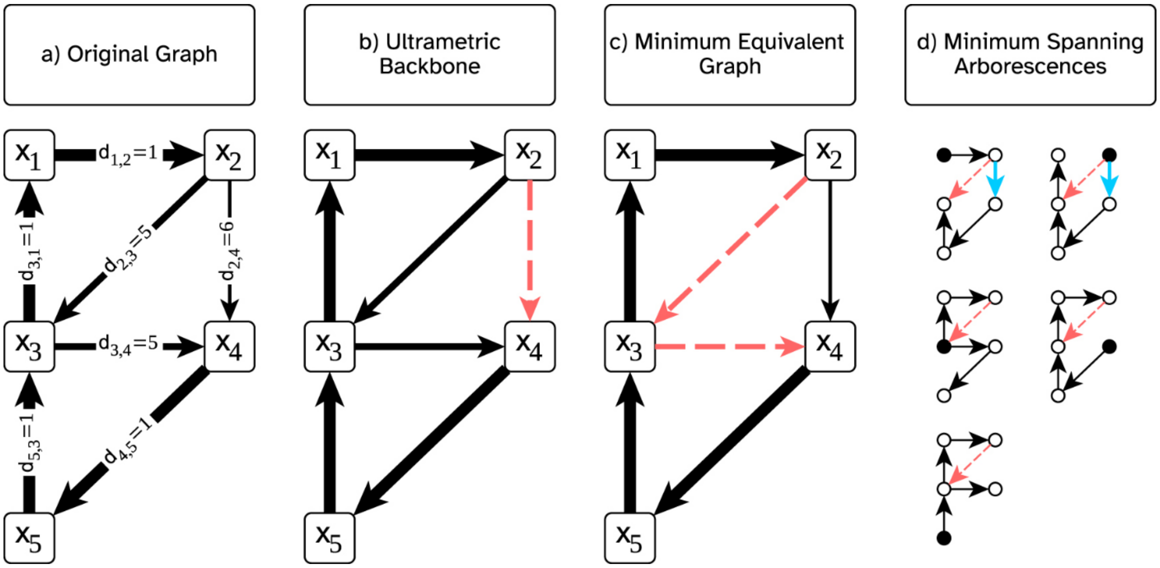

Figure 1: The ultrametric backbone is distinct from unions of MST analogs in directed graphs. (a) An example distance graph with thicker edges corresponding to smaller distance weights. (b) The ultrametric backbone is shown with edge weights omitted for visual clarity. Edge $d_{2,4}$ is removed, as indicated by the red dashes; it is redundant for **max-min** *shortest paths because it breaks the* **max-min** *transitivity. (c) A minimum equivalent graph is shown, which in this example is unique. Note that it is distinct from the ultrametric backbone and does not preserve the shortest* **max-min** *path from x_2 to x_4. (d) Five (in this case, unique) minimum spanning arborescences with the root node filled in with black are shown. The red dashed line indicates an edge, $d_{2,3}$, that is not in any minimum spanning arborescence, but is in the ultrametric backbone and required for* **max-min** *shortest paths (its weight increases from 5 to 6). The blue edge, $d_{2,4}$, is present in the union of these five graphs, but is redundant for* **max-min** *shortest paths and therefore is not in the ultrametric backbone.*

References

[1] Rocha L [2003] “Semi-metric behavior in document networks and its application to recommendation systems”. *Soft Computing Agents: A New Perspective for Dynamic Information Systems* 137–63.
[2] Van Mieghem P and van Langen S [2005]. “Influence of the link weight structure on the shortest path”. *Phys. Rev. E* **71** 056113.
[3] Simas, T., R. Brattig Correia, and L. M. Rocha. "The distance backbone of complex networks." *Journal of Complex Networks* **9**.6 (2021): cnab021.
[4] Brattig Correia R, Barrat A, Rocha L M [2023]. “Contact networks have small metric backbones that maintain community structure and are primary transmission subgraphs.” *PLOS Comput. Biol*. **19** 1–23.
[5] Van Mieghem P and Magdalena S M [2005] “Phase transition in the link weight structure of networks.” *Phys. Rev. E* **72** 056138.
[6] Estrada E and Hatano N [2010] “Resistance distance, information centrality, node vulnerability and vibrations in complex networks.” *Network Science*. 13–29
[7] Silver G, Akbarzadeh M and Estrada E [2018]. “Tuned communicability metrics in networks. The case of alternative routes for urban traffic.” *Chaos Solitons Fractals* **116** 402–13
[8] Simas T and L.M. Rocha [2015]. “Distance closures on complex networks.” *Network Science*. **3**:227–68.
[9] Menger K [1942] “Statistical metrics.” *PNAS* **28** 535–7.
[10] Klir GJ and Yuan B [1995]. *Fuzzy Sets and Fuzzy Logic, Theory and Applications* (Prentice Hall)