Extension of shortest path based Betweenness centrality to the multicriteria case

Keywords: Shortest path centrality; multicriteria network; multicriteria shortest path; betweenness centrality; transportation networks

Extended Abstract

Centrality measures have found wide application in social network analysis in various fields such as biology, epidemiology, management, computer science or logistics. Shortest path centrality measures provide a framework for understanding relationships between nodes in networks that represent complex systems. Betweenness centrality is defined based on connections, in the form of shortest paths, between pairs of nodes in the underlying network. Networks are usually assumed to have a single cost measure associated with traversing an edge. While there is a body of literature studying the multicriteria shortest path (MSP) problem, there is a lack of MSP based centrality concepts, unless the criteria are combined into a single one by applying a scalarisation [e.g. 1]. We introduce and apply such a MSP based centrality measure.

Background: Let $\mathscr{N} = (V, E, (c^1, \ldots, c^q))$ denote a directed weighted graph consisting of vertices V and edges $(i, j) \in E \subseteq V \times V$, where $c^k = (c^r_{ij})$ for $r \in \{1, 2, \ldots, q\}$ are vectors that represent different costs associated with the edge (i, j). A directed path P in this network consists of a sequence of consecutive edges $(i, j) \in E$ without repetition of vertices. A network is multicriteria if there are $q \geq 2$ cost vectors and single-criteria if there is only one ($q = 1$). In the following we assume that path costs are minimised. For the single-criterion case *Betweenness* centrality of a node k is determined based on the number of shortest paths in the graph that pass through node k, i.e., the number of shortest paths with node k as an intermediary node. If we denote by $\sigma(n, m)$ the number of shortest paths between the nodes n and m and by $\sigma(n, m)_k$ the number of shortest paths between n and m that go through the node k, then the Betweenness centrality of a node k is $B(k) = \sum_{k \neq n \neq m} \frac{\sigma(n,m)_k}{\sigma(n,m)}$ [2]. Betweenness centrality captures the opportunity that a node k has of being an intermediary between other nodes.

Here we propose an extension of Betweenness centrality to the multicriteria case ($q \geq 2$). While in the single-criterion case there is an optimal shortest path (or there are multiple shortest paths of the same shortest length), the MSP problem has a set of *(Pareto) efficient* solutions. We denote by $c^k(P) = \sum_{(i,j) \in P} c^k_{ij}$ the cost component k associated with path P. A path cost vector $(c^1(P), \ldots, c^q(P))$ is called *dominated* if there exists another path cost vector $(c^1(P^*), \ldots, c^q(P^*))$ such that $c^k(P^*) \leq c^k(P)$ for all $k \in \{1, \ldots, k\}$ with at least one strict inequality. Here path P^* is at least as good as path P in all criteria and strictly better in at least one, thereby making P^* an objectively better choice than P. The solution of a MSP problem is a *set of (Pareto) efficient solutions* with the property that there exists no other feasible path that dominates any of the efficient path cost vectors. We denote by $E\sigma(n, m)$ the number of efficient paths between nodes n and m and by $E\sigma(n, m)_k$ the number of efficient paths between n and m that pass through the node k. *Multicriteria Betweenness* is calculated as:

$$MB(k) = \sum_{k \neq n \neq m} \frac{E\sigma(n,m)_k}{E\sigma(n,m)}. \tag{1}$$

The computation of $MB(k)$ is based on solving MSP for all pairs of vertices [e.g. 3]. MSP can be more computationally expensive as network size increases.

While (1) is a similar formulation to the original Betweenness centrality measure there are significant differences in the multicriteria case. Figure 1 illustrates that there can be many different efficient paths in MSP that do not dominate each other, making them potential candidates for a preferred path without prescribing the actual choice being made a priori. In the context of a cycling network, for instance, one cyclist may choose a different trade-off between length of a path and its overall elevation gain depending on their fitness or perhaps availability of shower facilities when they get to their destination. The multicriteria centrality measure $MB(k)$ captures the significance of a particular node k in its ability to facilitate connections across a range of q criteria without having to prescribe a potentially unrealistic weighting between them. A high score $MB(k)$ indicates that node k is traversed by many efficient paths indicating that the network provides good connectivity with respect to the multiple criteria that are being considered. $MB(k)$ can also capture lexicographically minimal paths, and so-called non-supported paths that cannot be obtained by the popular weighted sum scalarisation [e.g. 1].

There are various transport networks that are multicriteria in nature, for instance public transport networks where the total fare, the total travel time and the number of exchanges are minimised. We will present a case study where commuter cyclist path choice is modelled assuming cyclists aim to minimise distance travelled, elevation gain and a discomfort measure associated with the quality and safety of cycling infrastructure. A comparison between distance-based single-criterion Betweenness and Multicriteria Betweenness can shed light on areas of the network where the different criteria are aligned, and, on the other hand, highlight areas where there is a need for infrastructure improvements. Based on the cycling application we will compare the two Betweenness measures and discuss insights derived from them.

References

[1] T. Opsahl, F. Agneessens, and J. Skvoretz. "Node centrality in weighted networks: Generalizing degree and shortest paths". In: *Social Networks* 32.3 (2010), pp. 245–251.

[2] L.C. Freeman. "A set of centrality measures based on betweenness". In: *Sociometry* 40.1 (1977), pp. 35–41.

[3] Y. Kergosien et al. "An Efficient Label-Correcting Algorithm for the Multiobjective Shortest Path Problem". In: *INFORMS Journal on Computing* 34 (2022), pp. 76–92.

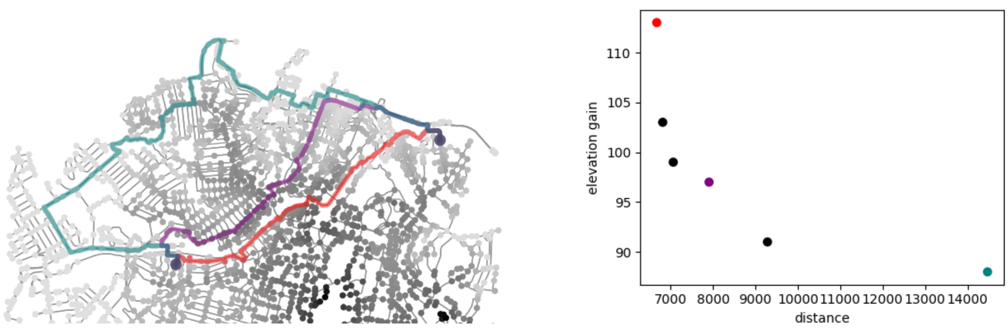

Figure 1: **Efficient paths.** Left: Auckland network with a subset of the efficient paths, and node color intensity indicating elevation; Right: Non-dominated cost vectors (distance and total elevation gain) found for a cycling network. The purple path is non-supported.