On the complexity of degree-constrained network design problems

Daniele Catanzaro[1], Gwenaël Joret[2], Brieuc Pierre[1], and Francesco Pisanu[1]

[1]Center for Operations Research and Econometrics, Université Catholique de Louvain, Ottignies-Louvain-la-Neuve, Belgium, ✉ {daniele.catanzaro, brieuc.pierre, francesco.pisanu}@uclouvain.be

[2]Computer Science Department, Universisé Libre de Bruxelles, Brussels, Belgium, ✉ gwenael.joret@ulb.be

Several network design problems require finding a spanning tree that minimizes or maximizes a prescribed cost function, often under capacity or budget constraints [5, 6]. In a number of relevant variants, the objective function is not defined in terms of local edge costs, but rather depends on global properties of the topology induced by the solution, such as distances, hop-counts, or aggregate path lengths between selected pairs of terminals. Problems of this type naturally arise in inverse or structure-driven optimization settings, where the underlying network is inferred from global interaction data rather than being directly specified.
Prominent examples arise in computational biology and hierarchical clustering. In these contexts, tree structures are used to encode nested similarity relations among a given set of items, and the quality of a solution is evaluated through objective functions that depend on pairwise dissimilarities and topological distances. A representative example is the *Balanced Minimum Evolution Problem (BMEP)* [4]. Given a set of items $\Gamma = \{1, \ldots, n\}$ and a dissimilarity matrix $\mathbf{D} = \{d_{ij}\}$, the BMEP asks for an *unrooted binary tree*, that is, a tree whose leaves are exactly the elements of Γ and whose internal vertices all have degree exactly 3, minimizing the objective function

$$\sum_{\substack{i,j \in \Gamma \\ i \neq j}} d_{ij}\, 2^{-\tau_{ij}^{T}},$$

where τ_{ij}^{T} denotes the number of edges on the unique path between leaves i and j in the tree T.
The BMEP is $\mathcal{APX}$-hard and impossible to approximate within a factor c^n, for some $c > 1$, unless $\mathcal{P} = \mathcal{NP}$ [3]. Moreover, the problem is extremely challenging in practice: current state-of-the-art exact formulations can barely solve instances of size 32 within one hour of computation on modern hardware [2], and the exponential decay in the objective function inherently leads to numerical instability. As a consequence, several existing approaches rely on heuristic strategies that either attempt to infer suitable edge weights on an auxiliary graph or simplify the objective function through approximations. These include learning-based and machine-learning-driven methods [8, 1], as well as stochastic and local-search approaches [7].

In this talk, we study the computational complexity of a family of structure-driven network design problems inspired by these models, focusing on spanning trees subject to uniform internal degree constraints, as in the case of the BMEP. Our goal is to understand whether commonly adopted strategies aimed at improving tractability fundamentally change the computational nature of the problem.

Problem 1 (Path-based objective under degree constraints). Given a connected undirected graph $G = (V, E)$ and a set of terminals $\Gamma \subseteq V$, consider the problem of finding a spanning tree T whose internal vertices satisfy strict degree constraints and whose leaf-set is exactly Γ, minimizing or maximizing the objective function

$$\sum_{\substack{i,j \in \Gamma \\ i \neq j}} d_{ij}\, (\tau_{ij}^{T})^{\alpha},$$

where d_{ij} denotes the dissimilarity between terminals i and j, τ_{ij}^{T} is the number of edges on the path between i and j in T, and $\alpha \geq 1$ is a fixed scalar. We show that this problem is $\mathcal{NP}$-hard, in both its minimization and maximization variants, even under strong structural restrictions on the input graph.

Problem 2 (Degree-constrained minimum spanning tree on an auxiliary graph). Assume that the pairwise dissimilarities d_{ij} have been encoded into edge weights of a suitably constructed auxiliary graph. We then consider the problem of finding a minimum spanning tree of this graph subject to strict internal degree constraints. We show that this problem remains $\mathcal{NP}$-hard, indicating that retrieving or learning edge weights does not, by itself, disrupt the inherent computational difficulty.

These results show that most approaches that aim to enhance tractability or numerical stability—such as replacing exponential terms with polynomial or linear approximations of the objective function—still yield optimization problems that remain computationally hard. This provides a complexity-theoretic explanation for the persistent difficulty encountered by both exact and heuristic methods in this class of network design problems.

References

[1] Dana Azouri, Shiran Abadi, Yishay Mansour, Itay Mayrose, and Tal Pupko. Harnessing machine learning to guide phylogenetic-tree search algorithms. *Nature communications*, 12(1):1983, 2021.

[2] Daniele Catanzaro, Raffaele Pesenti, Allan Sapucaia, and Laurence Wolsey. Optimizing over path-length matrices of unrooted binary trees. *Mathematical Programming, in press*, 2025.

[3] S. Fiorini and G. Joret. Approximating the balanced minimum evolution problem. *Operations Research Letters*, 40(1):31–35, January 2012.

[4] O. Gascuel and M. A. Steel. *Reconstructing evolution.* Oxford University Press, New York, NY, 2007.

[5] D. S. Johnson, J. K. Lenstra, and A. H. G. Rinnooy Kan. The complexity of the network design problem. *Networks*, 8:279–285, 1978.

[6] P. C. Pop. *Generalized network design problems: Modeling and optimization.* De Gruyter, Berlin, Germany, 2012.

[7] Ling Qin, Yixin Chen, Yi Pan, and Ling Chen. A novel approach to phylogenetic tree construction using stochastic optimization and clustering. *Bmc Bioinformatics*, 7(Suppl 4):S24, 2006.

[8] Tianjian Qin, Koen J van Benthem, Luis Valente, and Rampal S Etienne. Parameter estimation from phylogenetic trees using neural networks and ensemble learning. *Systematic biology*, page syaf060, 2025.