Maximum Entropy modelling of sub-Optimal Transport

Keywords: Optimal Transport, Random Ensemble, MaxEnt, Phase Transition, dense – sparse transtition

Extended Abstract

Many real systems allocate limited resources under uncertainty. Classical Optimal Transport (OT) [1] provides the cost-minimizing benchmark for such allocations, but empirical networks—economic, ecological, infrastructural—display redundancy, noise, and multiple objectives that keep them away from strict optimality. We introduce a principled “sub-optimal transport” (sub-OT) framework that connects maximum-entropy network ensembles with discrete OT. The key idea is to replace pure-entropy null models with an information-theoretic free-energy ensemble. We start from a Shannon-entropy MaxEnt construction [2] and add an “energy” term equal to the expected transport cost (the OT objective), yielding $F=S-\beta U$ in direct analogy with the Helmholtz free energy in physics, with β playing the role of an inverse temperature. This single control parameter trades off randomness (entropy) against cost (energy), producing a continuum of bipartite weighted graphs ranging from dense, redundancy-driven configurations when $\beta\approx 0$ to sparse, cost-dominated structures that coincide with discrete OT as $\beta\to\infty$.

The model considers weighted bipartite networks with prescribed node strengths on both layers (total supply and total demand) and a unit-cost matrix. Instead of maximizing entropy subject only to strength constraints, we maximize $S - \beta U$, where S is Shannon entropy over graphs and U is the expected transport cost induced by the cost matrix. The resulting ensemble factorizes over edges and produces independent exponential weights whose expectations depend on β, the costs, and Lagrange multipliers enforcing the strengths. As β increases, expected weights shift from being strength-driven and broadly spread across many links to concentrating on low-cost links.

To quantify sub-optimality we track the mass share of the maximum spanning tree (MST) of the average network, namely the fraction of total expected weight carried by the MST. This order parameter ties topology and weights, equals one in the OT limit, and is directly measurable on data. Numerically, as β grows the MST mass share rises smoothly from near zero (dense, diffuse connectivity) to one (sparse, near-tree support). The steepness increases with system size, yet the first derivative remains bounded, indicating a sharp but non-critical crossover rather than a thermodynamic phase transition.

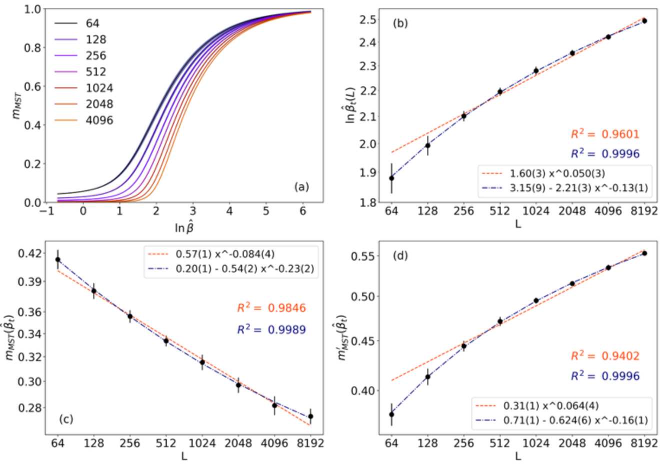
Finite-size and scaling analyses on large square bipartite instances show that the crossover location—defined by the peak slope of the MST mass share versus log β—converges to a finite limit when costs are i.i.d. uniform and strengths are Gaussian, with the same qualitative picture for log-normal or power-law strengths. A rescaled control parameter $B = \log(\tilde{\beta}/ \log L)$, where $\tilde{\beta}$ normalizes β by total mass, collapses curves at large L and is well captured by a simple limiting form: the MST mass share is approximately zero below a threshold Bc and rises like $1 - \exp[-\alpha(B - Bc)]$ above it. The mechanism is robust to cost and strength heterogeneity, including L-dependent Beta-distributed costs that preserve finite gaps among the smallest costs at large size; in all cases the system approaches the OT limit as β grows while maintaining bounded slopes.

Analytically, the partition function factorizes over edges, yielding independent exponential weights with mean 1/(β C_{iα} + t_i + θ_α), where the Lagrange multipliers (t, θ) are uniquely determined (up to a trivial gauge that adds a constant to all t's and subtracts it from all θ's) by the strength constraints via a strictly concave log-likelihood. Monotonicity in β follows directly: at fixed multipliers the expected weight on any edge decreases with β, and lower-cost edges carry larger weight. The ensemble recovers the Bipartite Weighted Configuration Model in the β→0 limit and concentrates on the discrete OT solution as β→∞, establishing continuity between redundancy-driven and cost-driven regimes. The MST mass share is bounded, continuous, and differentiable in β for finite graphs, equals one only when all expected mass lies on a spanning tree, and exhibits a bounded derivative with system size under broad cost/strength classes, confirming the non-critical nature of the dense-to-sparse crossover. A finite-size scaling with the rescaled parameter B provides an explicit limiting curve for the order parameter that is size-independent and accurately describes the onset and saturation of cost-dominated organization.

Finally, we make the theory analytically tractable. Because the ensemble factorizes edge-wise, the partition function can be computed in closed form, giving a product representation that yields explicit expressions for moments and likelihoods. In a coarse "global–constraint" approximation, where only the total transported mass is fixed, the model simplifies further and admits fully analytic formulas for expected weights and their fluctuations. Beyond this baseline, we show that modest, structured approximations for the Lagrange multipliers (e.g., homogeneous ansatz and perturbative expansions around the large system size limits) deliver closed-form results in key regimes of the full problem, providing asymptotics and scaling predictions that match our numerics.

References

[1] Villani. "Optimal transport: old and new". Springer Science \& Business Media (2001)
[2] Cimini et al. "The statistical physics of real-world networks", Nature Reviews Physics, 1 (1), (2019), 58-71



Sub-optimality transition of the model with uniformly distributed costs and Gaussian-distributed strengths. (a) Behavior of MST mass as a function of β for different network sizes L, where L is the number of nodes in each layer of the bipartite network. Each curve represents the average over at least 100 realizations. (b) Transition values ln βt as a function of L. (c) Values of the order parameter mM ST at the transition point as a function of L. (d) Maximum derivative of mM ST , evaluated at βt. In panels (b), (c), (d) orange and blue lines correspond to power-law and bounded power-law fits, respectively. Legends include the best-fit parameters along with the corresponding adjusted R-squared values.