How far are two graphs?
Exploring sets of complex networks using Gromov-Wasserstein distances

Raul Vicente[1], Dirk Oliver Theis[1], Reigo Hendrikson[1], and Victor Eguiluz[2]

[1]Institute of Computer Science, University of Tartu, Ulikooli 17, Tartu, Estonia
[2]Instituto de Fisica Interdisciplinar y Sistemas Complejos IFISC (CSIC-UIB), E07122, Palma de Mallorca, Spain

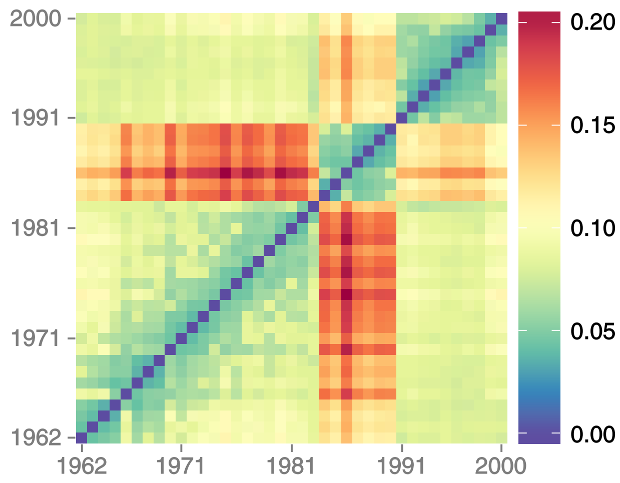
Relations between data are often presented as graphs or networks. When a complex system is naturally described as a set of networks, as for example in studying the metabolic graphs of a set of related species or following the evolution in time of a network, one is often interested in comparing the networks of that set. Such a comparative measure can provide a way to inspect the relations that the networks themselves have between each other and provide clues about their interactions and evolution.

However, it is unclear which measure is more appropriate for comparing two or more networks. For example, it is difficult to compare two networks when the number of nodes is different or more generally, when we lack a correspondence between the nodes of one network and the nodes of the other. In the past researchers have compared global statistics of some particular attribute of the networks such as their clustering coefficient, their mean degree, or their average shortest path length. However, these approaches miss important information encoded in the graphs. Ideally one would like to compare different networks with a measure that takes into account their internal structures of distance/similarity between their respective internal nodes. In another words, one would like to come with a measure of distance between the set of distances defined inside each network.

In this work we capitalize on the recent development by F. Memoli of using optimal transportation distances for the problem of object matching (comparing 3D shapes) in computer vision [1]. In particular, Memoli derived a computationally feasible bound for the Gromov distance (a metric distance between metric spaces) named Gromov-Wasserstein distance (d_{GW}) and applied it to the problem of shape matching [1, 2]. Intuitively, the Gromov-Wasserstein distance aims to solve a certain graph isomorphism problem by finding an optimal mapping (not necessarily one-to-one) between two sets of nodes such that their respective distances are best preserved by the map. The deviation from a perfect isomorphism is related to the distance between the two objects.

In our work we show analytically that Gromov-Wasserstein distance is a well behaved metric even when the two sets of distances compared do not satisfy the metric property (i.e. they do not satisfy the restrictive condition of triangular inequality). This observation dramatically expands the application of Gromov-Wasserstein type of measures to pairs of undirected graphs in which their internal (between nodes) notions of distance/dissimilarity only need to satisfy positivity and symmetry (but not triangular inequality).

After describing a numerical method to estimate Gromov-Wasserstein distance for pairs of undirected graphs we proceed to validate the measure with toy cases (see Fig. 1 for an example on yearly networks of world trade flows) and well-known families of complex networks. Finally, we apply Gromov-Wasserstein distances in combination with dimensionality reduction methods to visualize and explore the structure of several real-world sets of networks, including applications to biological networks.



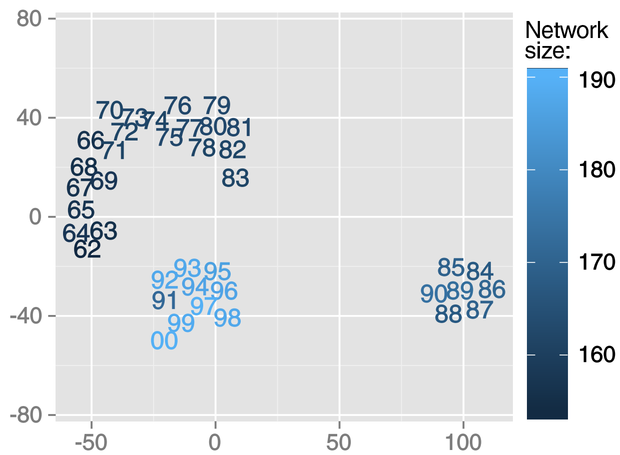


Figure 1: Matrix of Gromov-Wasserstein distances between pairs of networks (a different network is computed for every year) of world trade flows from 1962 to 2000. *Top panel*: matrix of pairwise distances ((d_{GW})). *Bottom panel*: 2d embedding of the distance matrix ((d_{GW})) using t-SNE. Numbers 62-99 represent years 1962-1999. Number 00 represents year 2000.

[1] F. Memoli, Foundations of Computational Mathematics **11**, 417-487 (2011).

[2] F. Memoli, Axioms **3**, 335-341 (2014).