

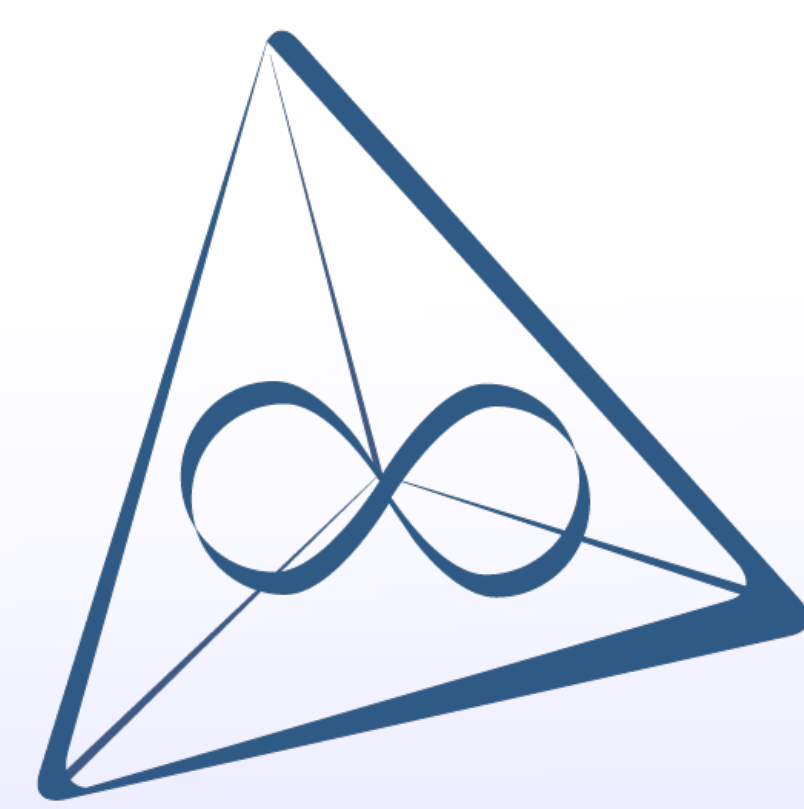


Spectrum analysis on biology networks

Jiao Gu, Peter Stadler

jiaogu@mis.mpg.de

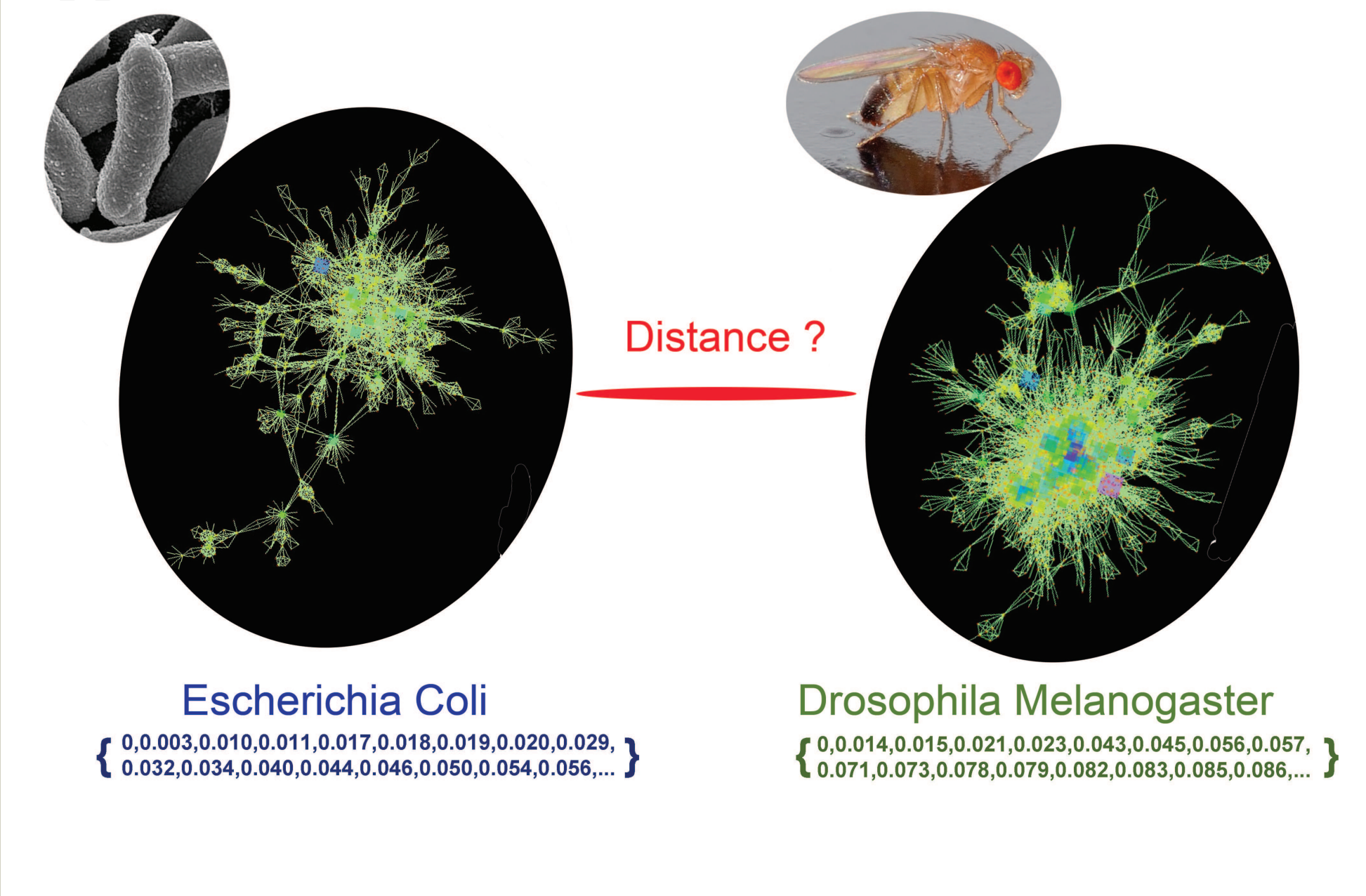
Max-Planck-Institute for Mathematics in the Sciences, Leipzig, Germany



MAX-PLANCK-GESELLSCHAFT

Problem

We want to find a proper metric on the space of graphs, with which we can measure the difference between two graphs without the identical information for each node. Since the spectrum reflects the structure information of a graph [1], we take it as the starting point of our approach.



Basic Concepts

For a graph $G = (V, E)$, the normalized Laplacian matrix $\Delta = [a_{ij}]$ has the form

$$a_{ij} = \begin{cases} 1, & \text{if } i = j \text{ and } n_i \neq 0, \\ -\frac{1}{n_j}, & \text{if } i \sim j \text{ is an edge,} \\ 0, & \text{otherwise.} \end{cases}$$

We call λ an eigenvalue of Δ if there exists some $u \neq 0$, such that

$$\Delta u = \lambda u.$$

For a network G with n nodes, we have the spectrum sequence [2] :

$$0 = \lambda_0 \leq \lambda_1 \leq \dots \leq \lambda_{n-1}.$$

Spectrum metric

We could use Gaussian kernel to smooth the eigenvalues set $\{\lambda_i\}_{i=1}^n$:

$$\rho(x) = \frac{1}{n} \sum_{i=1}^n \frac{1}{\sqrt{2\pi\sigma_G^2}} e^{-\frac{(x-\lambda_i)^2}{2\sigma_G^2}}, \quad (1)$$

where σ_G is the smooth factor (bandwidth) for this network. The algorithm to optimize bandwidth could be found in [3].

Spectrum metric for any two graphs G and G' , with eigenvalues sets $\{\lambda_i\}_{i=1}^n$ and $\{\lambda_j\}_{j=1}^m$ respectively, is defined as:

$$d_{ij} = \int |\rho_i(x) - \rho_j(x)| dx, \quad (2)$$

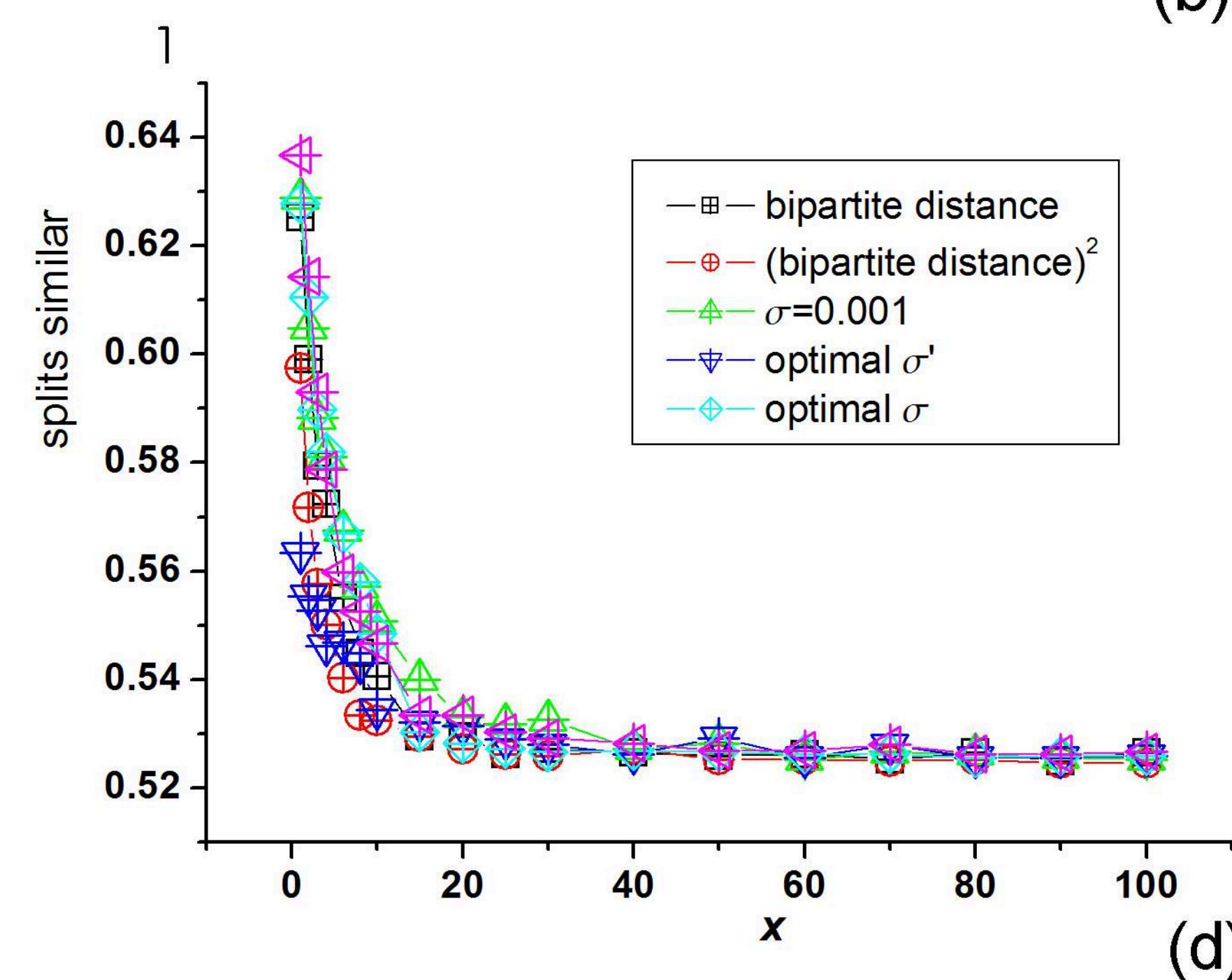
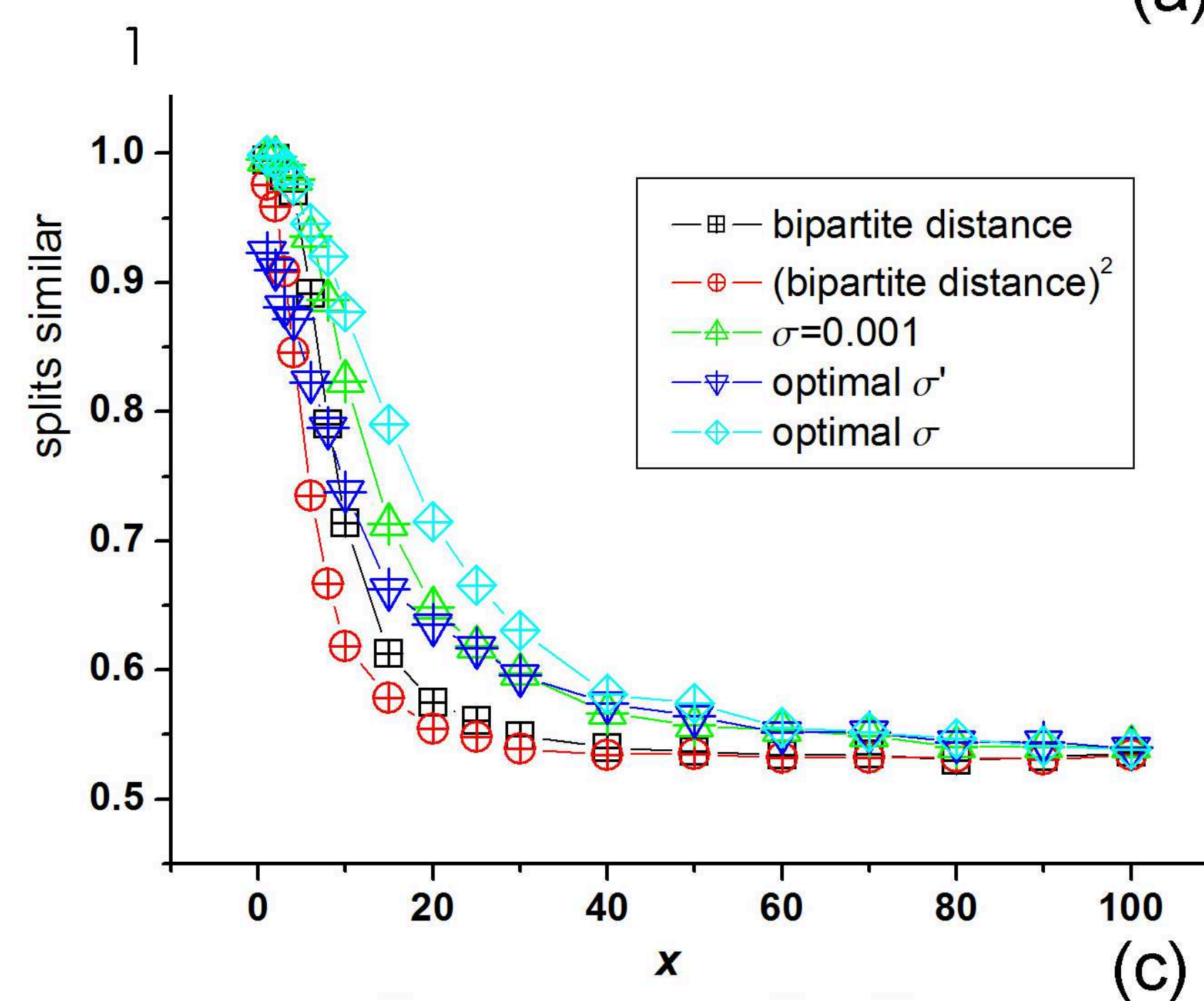
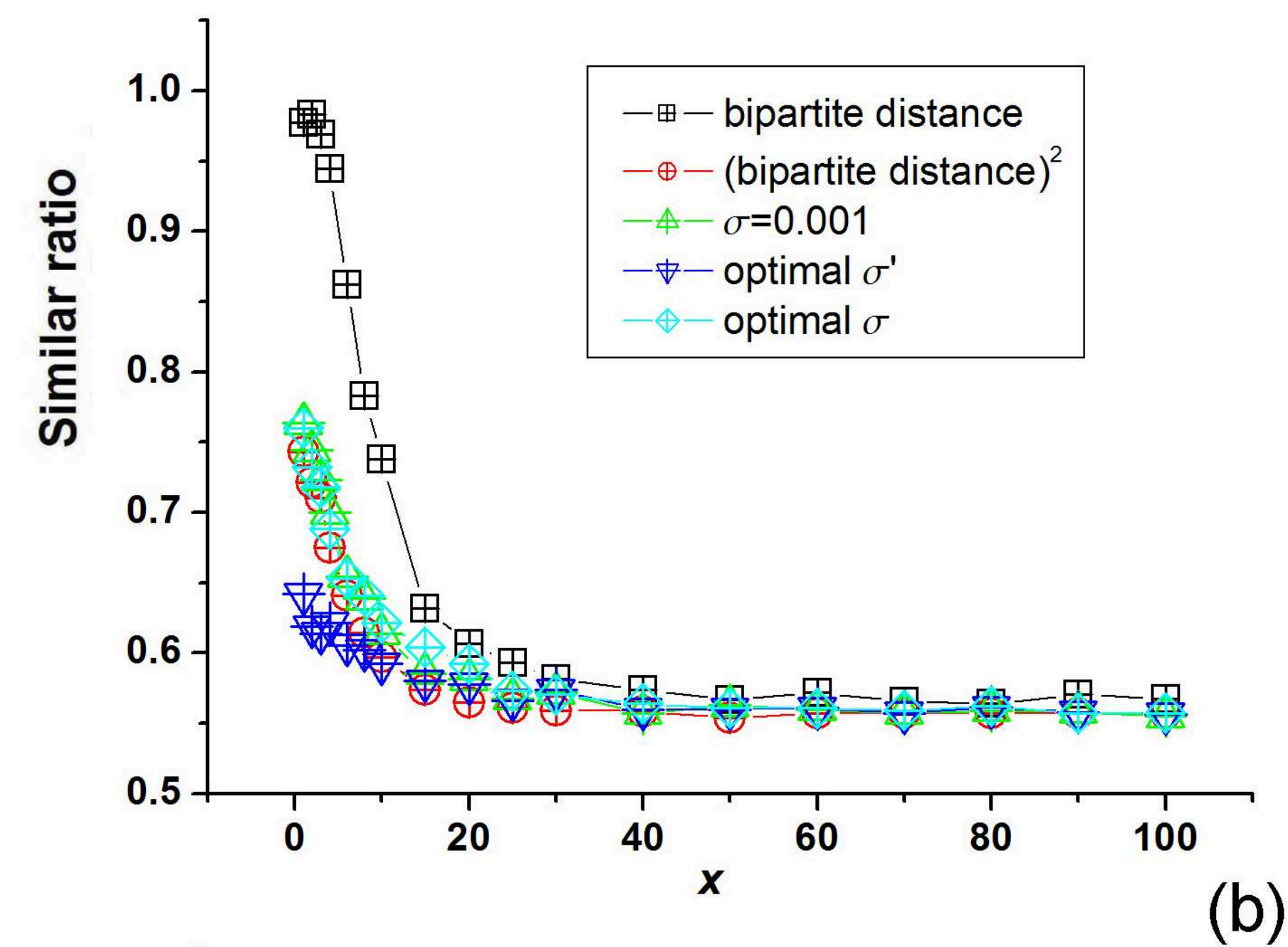
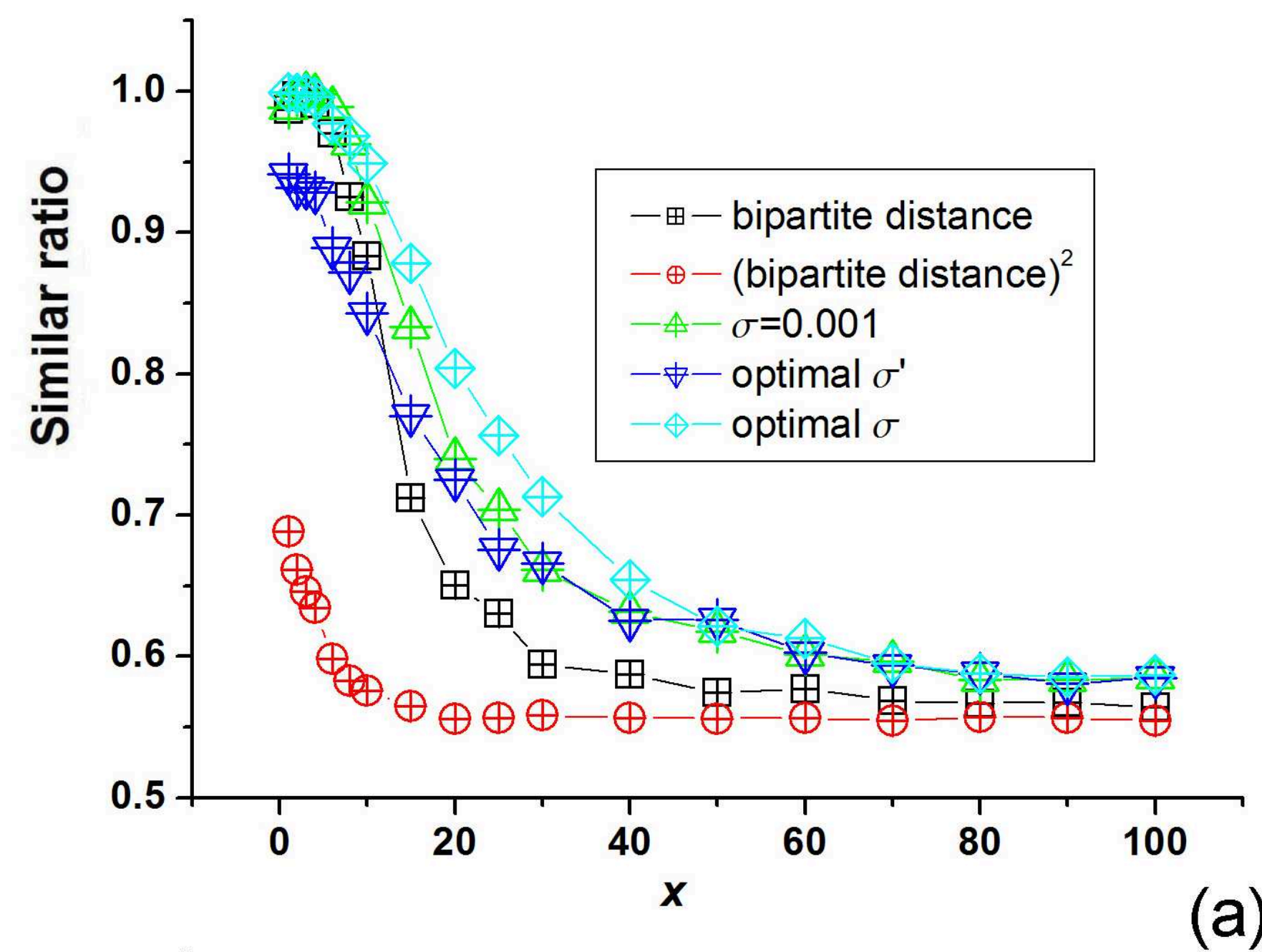
which is a value lying in $[0, 2]$.

References

- [1] Anirban Banerjee, Jürgen Jost: *The Spectrum of the Graph Laplacian as a Tool for Analyzing Structure and Evolution of Networks* (2008)
- [2] Fan Chung: *Spectral graph theory* (1992)
- [3] Z. I. Botev, J. F. Grotowski and D.P. Kroese: *Kernel density estimation via diffusion* (2010)
- [4] S. J. Sheather and M. C. Jones: *A reliable data-based bandwidth selection method for kernel density estimation* (2006)
- [5] Hans Bandelt and Andreas Dress: *Split decomposition: a new and useful approach to phylogenetic analysis of distance data* (1992)

Reconstruction with Spectrum metric and other measures

We used different metric to reconstruct balanced tree $(((((A : x, B : x) : 2x, (C : x, D : x) : 2x) : 4x, (E : x, F : x) : 2x, (G : x, H : x) : 2x) : 4x) : 8x, (((((I : x, J : x) : 2x, (K : x, L : x) : 2x) : 4x, (M : x, N : x) : 2x, (O : x, P : x) : 2x) : 4x) : 8x))$; and unbalanced tree $((((((((A : x, B : x) : x, C : 2x) : x, D : 3x) : x, E : 4x) : x, F : 5x) : x, G : 6x) : x, H : 7x) : x, I : 8x) : x, J : 9x) : x, K : 10x) : x, L : 11x) : x, M : 12x) : x, N : 13x) : x, O : 14x) : x, P : 15x)$; with 16 leaves (networks), and also the corresponding balanced and unbalanced trees with 32 leaves. The similar ratio, which measures how the reconstructed trees are closed to the original trees are shown in (a),(b), (c) and (d) respectively. The ratio using bipartite distance, spectrum metric (see Eq.2) with constant factor $\sigma = 0.001$ (in Eq.1) and optimal smooth factor σ (by [3]), σ' (by [5]) are marked in different colors. In most cases, spectrum metric with optimal σ is better than the others. However, the similar ratio falls down quickly when x becomes larger using these methods. All the methods we used here do not employ the identical nodes information.



Phylogenetic trees on biology networks

With a metric on the space of graphs in hand, we can generate phylogenetic trees for biology domain networks. For the following two trees, (b) is generated by spectrum metric we described in (2); (a) is generated by edit distance between graphs. Principally, the edit distance counts how many percentage of edges are shared by two networks. Note that the edit distance takes the node information into account whereas the spectrum metric does not. The following two trees show that spectrum metric can perform as well as edit distance does for biology domain networks.

