Publications mathématiques de Besançon
Algèbre et théorie des nombres

Nicolas Wicker, Canh Hao Nguyen, and Hiroshi Mamitsuka

Some properties of a dissimilarity measure for labeled graphs

2016, p. 85-94.

<http://pmb.cedram.org/item?id=PMB_2016_____85_0>

Publication éditée par le laboratoire de mathématiques de Besançon, UMR 6623 CNRS/UFC

Publications mathématiques de Besançon – 2016, 85-94

SOME PROPERTIES OF A DISSIMILARITY MEASURE FOR LABELED GRAPHS

by

Nicolas Wicker, Canh Hao Nguyen and Hiroshi Mamitsuka

Abstract. — We investigate the problem of comparing different graphs on the same set of vertices. It is a problem arising when using different biological networks to elucidate cellular processes. We wish to see their similarity and difference via connectivity-aware graph dissimilarity for graphs with the same node set. We extend a previous result and present some results concerning the orders of magnitude of the dissimilarity as the graphs' sizes grow to infinity. We find that removing an edge playing a very important role in graph connectivity, such as a bridge between two fully connected subgraphs, can have a dramatic effect on the dissimilarity compared to the removal of any "ordinary" edge.

Résumé. — Nous nous intéressons au problème de la comparaison de graphes sur un même ensemble de sommets. C'est un problème apparaissant dans l'étude de réseaux biologiques lorsqu'on veut comprendre le fonctionnement de processus cellulaires. L'objectif est de lier leur similarité ou différence, par une mesure consciente de la connectivité du graphe sur un même ensemble de sommets. Nous étendons un résultat antérieur et présentons de nouveaux résultats sur l'ordre de grandeur de la dissimilarité lorsque la taille des graphes tend vers l'infini. En particulier, nous montrons que la suppression d'une arête qui joue une grande importance dans la connectivité d'un graphe, comme un pont, peut avoir un effet dramatique sur la dissimilarité par rapport à la suppression d'une arête « ordinaire ».

1. Introduction

Biological networks are a major source of information for understanding complex biological processes [8]. One of the ways to elucidate the cellular machinery and to predict interaction and function is to study the similarity and difference in networks of different species or on different conditions. Many statistical models and computational methods have been developed to compare graphs [3, 4, 6]. However, the key idea is that properties of networks are determined by its motifs such as paths, subgraphs and graphlets. It is not possible to use these methods to compare networks for their global property such as network connectivity and robustness.

Mathematical subject classification (2010). — 05C50,15A18.
Key words and phrases. — Labeled graphs, Graph spectrum, Dissimilarity, Bridge effect.

We consider the problem of comparing networks taking into account global connectivity [9, 5]. This would be useful for various biological tasks. For example, the two networks would be considered similar in robustness if they are both not robust in the sense that there exist small changes (such as the removal of one edge) that can result in large network topological changes (such as disconnectivity). Two networks would be considered similar in modularity if they share many common well-connected subnetworks and bottlenecks. This is the case of biological networks sharing many modules. These kinds of information reflect global connectivity of the networks. To the best of our knowledge, the method in [9] is the first attempt in this direction.
The problem setting is as follows. Given graphs $G_i = (X, E_i)$ that all share the vertex set X with different edge set E_i, we want to compare them using the graph normalized Laplacians L_i, given for G_i by:

$$L_i(u, v) = \begin{cases} 1 & \text{if } u = v \text{ and } d_v = 0, \\ -\frac{1}{\sqrt{d_u d_v}} & \text{if } u \text{ and } v \text{ are adjacent,} \\ 0 & \text{otherwise} \end{cases}$$

where d_v stands for the degree of vertex v. Then, the dissimilarity measure we study has the following form:

$$d(G_1, G_2) = \sum_{i,j} \frac{(\lambda_i - \mu_j)^2}{(\lambda_i + \mu_j)^\alpha} \langle u_i, v_j \rangle^2 \text{ with } \alpha \in [0, 2),$$

where $L_1 = \sum_i \lambda_i u_i u_i^T$ and $L_2 = \sum_i \mu_i v_i v_i^T$ are eigendecompositions of the graph normalized Laplacians. The dissimilarity measure in [9] is a special case with $\alpha = 1$. In the special case of $\alpha = 1$, d is the dissimilarity measure of graphs taking into account global information of graphs such as clusters and bottlenecks [9, 5].
The main purpose is to see the change of d on large graphs in order to quantify the significance of the differences among large graphs, as the case of biological networks. While deriving formulas for d for all pair of graphs would be difficult, we selected here some canonical cases to show formally how d change, to roughly estimate the differences between graphs under this measure.
The setting of our simulation is as follows. We study two cases. Limiting ourselves to comparing graphs with minimum difference of only one edge, the *worst case* is that the edge makes the most topological change in the graph. It is the case of a bridge between two fully connected subgraphs as follows. Define two graphs U and U_b where U is the union of two complete graphs K_n and K_n and U_b is equal to U with an additional edge (*bridge*) connecting the two complete subgraphs. In the literature, U_b is sometimes called a barbell graph [2]. Without loss of generality, we suppose that this edge is between the n^{th} node and the $n + 1^{\text{th}}$ node, where $2n$ is the size of the graphs. To compare with the worst case, we design another case of graph U_r which is obtained by removing one edge (not the bridge) from U.
In this paper, we generalize the dissimilarity measure found in [9] and motivate this generalization by theorem 1, then we compare the magnitudes of $d(U, U_b)$ and $d(U, U_r)$ as a function of n in Theorems 2 and 3.

2. Results

First, let us show that adding a parameter α makes sense and essentially can help avoiding all graphs without isolated vertices to be equidistant to the graph with only isolated vertices. The theorem stands as follows:

Theorem 1. — *Let G_1 be the graph with only isolated vertices, then the most distant graph to it is the bipartite graph with $n/2$ disconnected edges if $\alpha < 1$, the set of all graphs without any isolated vertices if $\alpha = 1$ and the complete graph is $\alpha > 1$.*

Proof. — If G_2 is any graph, then the dissimilarity can be rewritten:

$$d(G_1, G_2) = \sum_{i,j} \frac{\mu_j^2}{\mu_j^\alpha} \; u_i, v_j \;\; ^2 = \sum_{i=1}^{n} \mu_i^{2-\alpha}$$

as all eigenvalues of G_1 are equal to 0. If we use $\alpha = 1$ as in [9], then as $\sum_{i=1}^n \mu_i = n$ is constant all graphs have the same distance to G_1. We can thus consider the problem of maximising the distance $d(G_1, G_2)$ with a given α, rephrasing it in terms of $\beta = 2 - \alpha$ we obtain the following maximisation problem:

$$\max \sum_{i=1}^{n} \mu_i^\beta \text{ subject to } \sum_{i=1}^{n} \mu_i = n, 2 \geq \mu_i \geq 0 \text{ and } \mu_1 = 0.$$

The constraint $2 \geq \mu_i \geq 0$ is necessary as $\mu_1, \ldots, \mu_n$ are eigenvalues of a normalized Laplacian [1]. The Lagrangian is then

$$L(\mu, \lambda, \eta, \xi) = \sum_{i=1}^{n} \mu_i^\beta + \lambda \left(\sum_{i=1}^{n} \mu_i - n \right) + \sum_{i=1}^{n} \eta_i(\mu_i - 2) - \xi_i \mu_i \text{ where } \lambda, \eta_i \text{ and } \xi_i$$

are Lagrange multipliers.

Deriving leads to

$$\frac{\partial L}{\partial \mu_i} = \beta \mu_i^{\beta-1} + \lambda + \eta_i - \xi_i.$$

Let us consider an eigenvalue $\mu_i \quad \{0, 2\}$ then $\beta \mu_i^{\beta-1} + \lambda = 0$. This shows that eigenvalues different from 0 and 2 can only take one unique value which will be denoted z.
Then, $n = 2x + yz$ where x is the number of eigenvalues equal to 2 and y the number of eigenvalues equal to an eigenvalue different from 0 and 2. The function to optimize becomes then

$$\sum_{i=1}^{n} \mu_i^\beta = x 2^\beta + y z^\beta$$

with constraints $2x + yz = n$ and $x + y \leq n - 1$ as one eigenvalue is equal to 0. We can then consider two cases, either $\beta > 1$ or $\beta < 1$.
If $\beta > 1$, $2^\beta > z^\beta$ as $z < 2$. The optimum is then obtained for $x = n/2$ and $y = 0$. This corresponds to the simple bipartite graph containing $n/2$ disconnected edges.

If $\beta < 1$, by concavity of function $f(w) = w^\beta$, if we consider that: $x + y = k$ with $k \leq n - 1$:

$$\begin{aligned} x2^\beta + yz^\beta &= k\left(\frac{x}{k}2^\beta + \frac{y}{k}z^\beta\right) \\ &\leq k\left(\frac{2x + yz}{k}\right)^\beta \\ &\leq k\left(\frac{n}{k}\right)^\beta \text{ as } 2x + yz = n. \end{aligned}$$

This upper bound can be obtained by taking $x = 0$, $y = k$ and $z = \frac{n}{k}$. Besides, the bound is maximized for $k = n - 1$. This corresponds to the spectrum of K_n the complete graph of size n. If we summarize this in terms of α, if $\alpha = 1$ all graphs are equally distant to G_1. If $\alpha > 1$ the most distant graph to G_1 is the complete graph, and if $\alpha < 1$ the most distant graph is the bipartite graph with $n/2$ disconnected edges. Interestingly, this shows that the farthest graph from G_1 can be very different depending upon the value of α with a kind of transition phase at $\alpha = 1$.

Now, we can notice that the dissimilarity behaves nicely when two graphs are concatenated i.e. when we keep all the vertices and edges of the two graphs. Namely, we have the following lemma.

Lemma 1. — *When two graphs G_1 and H_1 of equal size and G_2 and H_2 two other graphs of equal size are concatenated, then $d(G_1 \quad G_2, H_1 \quad H_2) = d(G_1, H_1) + d(G_2, H_2)$*

Proof. — Let us denote by $u_1^1, \ldots, u_n^1$ and $\lambda_1^1, \ldots, \lambda_n^1$ the eigenvectors and eigenvalues of G_1 by $u_1^2, \ldots, u_m^2$ and $\lambda_1^2, \ldots, \lambda_m^2$ the eigenvectors and eigenvalues of G_2. Similarly, the eigenvectors and eigenvalues of H_1 and H_2 are given respectively by: $v_1^1, \ldots, v_m^1$ and $\mu_1^1, \ldots, \mu_m^1$, and $v_1^2, \ldots, v_n^2$ and $\mu_1^2, \ldots, \mu_n^2$.
Then, the eigenvalues of $L(G_1 \quad G_2)$ are given by $0, 0, \lambda_2^1, \ldots, \lambda_n^1, \lambda_2^2, \ldots, \lambda_m^2$, as 0 is always the eigenvalue of a Laplacian and as G_1 and G_2 are disconnected. Similarly, the eigenvalues of $L(H_1 \quad H_2)$ are given by $0, 0, \mu_2^1, \ldots, \mu_n^1, \mu_2^2, \ldots, \mu_m^2$. The eigenvectors of $L(G_1 \quad G_2)$ are denoted $x_1, \ldots, x_{n+m}$ and those of $L(H_1 \quad H_2)$ by $y_1, \ldots, y_{n+m}$. The eigenvectors after the first two ones are the eigenvectors of $L(G_1)$, $L(G_2)$, $L(H_1)$ and $L(H_2)$, completed with m or n 0 respectively. For example, $x_3 = (u_2, 0, \ldots, 0)$.

$$\begin{aligned} d(G_1 \quad G_2, H_1 \quad H_2) = &\sum_{1 \leq i,j \leq n} \frac{(\lambda_i^1 - \mu_j^1)^2}{(\lambda_i^1 + \mu_j^1)^\alpha} \; x_i, y_j \;{}^2 + \sum_{\substack{1 \leq i \leq n \\ n+1 \leq j \leq n+m}} \frac{(\lambda_i^1 - \mu_j^2)^2}{(\lambda_i^1 + \mu_j^2)^\alpha} \; x_i, y_j \;{}^2 + \\ &\sum_{n+1 \leq i,j \leq n+m} \frac{(\lambda_i^2 - \mu_j^2)^2}{(\lambda_i^2 + \mu_j^2)^\alpha} \; x_i, y_j \;{}^2 + \sum_{\substack{n+1 \leq i \leq n+m \\ 1 \leq j \leq n}} \frac{(\lambda_i^2 - \mu_j^1)^2}{(\lambda_i^2 + \mu_j^1)^\alpha} \; x_i, y_j \;{}^2 \\ = &\sum_{1 \leq i,j \leq n} \frac{(\lambda_i^1 - \mu_j^1)^2}{(\lambda_i^1 + \mu_j^1)^\alpha} \; u_i^1, v_j^1 \;{}^2 + \sum_{n+1 \leq i,j \leq n+m} \frac{(\lambda_i^2 - \mu_j^2)^2}{(\lambda_i^2 + \mu_j^2)^\alpha} \; u_i^2, v_j^2 \;{}^2 \\ = &\; d(G_1, H_1) + d(G_2, H_2). \end{aligned}$$

This result is shared by the edit distance [7] defined by:

$$\mathrm{ed}(G_1, G_2) = |E_1 \setminus E_2 \quad E_2 \setminus E_1|$$

where $G_1 = (V, E_1)$ and $G_2 = (V, E_2)$. However, if one considers the normalized edit distance, i.e the edit distance divided by the maximum number of edges, this is no more true. Indeed, if d_1 and d_2 are the edit distances between G_1 and H_1 and G_2 and H_2 respectively. Then, the normalized distances are $\mathrm{ned}(G_1, H_1) = \frac{2d_1}{n(n-1)}$, $\mathrm{ned}(G_2, H_2) = \frac{2d_2}{m(m-1)}$ and $\mathrm{ned}(G_1 \quad G_2, H_1 \quad H_2) = \frac{2d_1+2d_2}{(n+m)(n+m-1)}$ which in general is not equal to: $\mathrm{ned}(G_1, H_1) + \mathrm{ned}(G_2, H_2) = \frac{2d_1}{n(n-1)} + \frac{2d_2}{m(m-1)}$.

Theorem 2. — *The dissimilarity* $D(U, U_b) = \left(\frac{2}{n^2}\right)^{2-\alpha} + \frac{2^{1-\alpha}}{n^2} + o(n^{-2})$.

The normalized Laplacians L_1 and L_2 of U and U_b respectively are given by:

$$L_1 = \begin{pmatrix} A_n^n & 0 \\ 0 & A_n^n \end{pmatrix} \text{ and } L_2 = \begin{pmatrix} A_n^{n-1} & B & 0 & 0 \\ B^T & 1 & -n^{-1} & 0 \\ 0 & -n^{-1} & 1 & B \\ 0 & 0 & B^T & A_n^{n-1} \end{pmatrix}$$

$$\text{with } A_n^m = \begin{pmatrix} 1 & -(n-1)^{-1} & \dots & -(n-1)^{-1} \\ -(n-1)^{-1} & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & -(n-1)^{-1} \\ -(n-1)^{-1} & \dots & -(n-1)^{-1} & 1 \end{pmatrix},$$

$$\text{and } B^T = \left(\frac{-1}{\sqrt{n(n-1)}}, \dots, \frac{-1}{\sqrt{n(n-1)}}\right).$$

with m is the matrix size.

Spectral analysis of L_1. — Eigenvalues are $\lambda_1 = \lambda_2 = 0$ and $\lambda_3 = \cdots = \lambda_{2n} = n/(n-1)$. The first two eigenvectors are: $u_1 = \frac{1}{\sqrt{2n}}(1, \dots, 1)$ and $u_2 = \frac{1}{\sqrt{2n}}(1, \dots, 1, -1, \dots, -1)$. Eigenvalue $n/(n-1)$ has multiplicity $2n-2$ as the following vectors are linearly independent eigenvectors for it: $u_3 = \left(1, \frac{-1}{n-1}, \dots, \frac{-1}{n-1}, 0, \dots, 0\right)$, $u_4 = \left(\frac{-1}{n-1}, 1, \frac{-1}{n-1}, \dots, \frac{-1}{n-1}, 0, \dots, 0\right)$, $u_{n+1} = \left(\frac{-1}{n-1}, \dots, \frac{-1}{n-1}, 1, -\frac{1}{n-1}, 0, \dots, 0\right)$, $u_{n+2} = \left(0, \dots, 0, 1, \frac{-1}{n-1}, \dots, \frac{-1}{n-1}\right), \dots,$ $u_{2n} = \left(0, \dots, 0, \frac{-1}{n-1}, \dots, \frac{-1}{n-1}, 1, \frac{-1}{n-1}\right)$. The orthonormal vectors $u_3, \dots, u_{2n}$ are not needed explicitly in the proof.

Spectral analysis of L_2. — (computation of μ_1, μ_2, v_1 and v_2) The smallest eigenvalue is 0 with eigenvector:

$$v_1 = \frac{1}{\sqrt{2(n-1)^2 + 2n}}(\sqrt{n-1}, \dots, \sqrt{n-1}, \sqrt{n}, \sqrt{n}, \sqrt{n-1}, \dots, \sqrt{n-1}).$$

Looking for a similar eigenvector, we find

$$v_2 = \frac{1}{\sqrt{2(n-1)}\sqrt{n-1+\frac{1}{n}}}\left(1, \dots, 1, \frac{-(n-1)^{\frac{3}{2}}}{\sqrt{n}}, \frac{-(n-1)^{\frac{3}{2}}}{\sqrt{n}}, 1, \dots, 1, \right)$$

for eigenvalue $1 + \frac{1}{n(n-1)}$. Eigenvalue $n/(n-1)$ has multiplicity $2n-4$ as the following vectors are linearly independent eigenvectors for it: $v_5 = \left(1, \frac{-1}{n-2}, \ldots, \frac{-1}{n-2}, 0 \ldots, 0\right)$, $v_6 = \left(\frac{-1}{n-2}, 1, \frac{-1}{n-2}, \ldots, \frac{-1}{n-2}\right)$, $v_{n+2} = \left(\frac{-1}{n-2}, \ldots, 1, \frac{-1}{n-2}, 0, \ldots, 0\right)$, $v_{n+3} = \left(0, \ldots, 0, 1, \frac{-1}{n-2}, \ldots, \frac{-1}{n-2}\right)$, $v_{n+4} = \left(0, \ldots, 0, \frac{-1}{n-2}, 1, \frac{-1}{n-2}, \ldots, \frac{-1}{n-2}\right)$, $v_{2n} = \left(0, \ldots, 0, \frac{-1}{n-2}, \ldots, \frac{-1}{n-2}, 1, \frac{-1}{n-2}\right)$ where each time $n+1$ values are equal to 0. Concerning the rest of the spectrum, lemma 2 tells us what v_3, v_4, $\tilde{\mu}_3$ and $\tilde{\mu}_4$ are equal to. Before going to the proof of theorem 2, the following lemma is needed.

Lemma 2. — *There are two eigenvectors v_3 and v_4 of L_2 having the form $v_3 = (1 + 4x_3(2n)^{-1/2} + 2x_3^2)^{-1/2}(u_2 + x_3e_n - x_3e_{n+1})$ and $v_4 = (1 + 4x_4(2n)^{-1/2} + 2x_4^2)^{-1/2}(u_2 + x_4e_n - x_4e_{n+1})$ where $e_1, \ldots, e_{2n}$ is the canonical basis of $\mathbb{R}^{2n}$, $x_3 = -\frac{3n^{-\frac{3}{2}}}{2\sqrt{2}} + o(n^{-3/2})$ and $x_4 = -\frac{\sqrt{n}}{\sqrt{2}} - \frac{3}{2\sqrt{2n}} + o(n^{-1/2})$. The corresponding eigenvalues are $\mu_3 = \frac{2}{n^2} + o(n^{-2})$ and $\mu_4 = 1 + \frac{2n-1}{n(n-1)} - \mu_3$.*

Proof. — We will show that there exists x_3 that makes v_3 an eigenvector satisfying the lemma.
Solving the equation $L_2\, v_3 = \mu v_3$ gives respectively for the first $n-1$ lines and the n^{th} line:

$$\frac{\frac{1}{\sqrt{2n}} - \frac{n-2}{\sqrt{2n}(n-1)} - \frac{(2n)^{-1/2} + x_3}{\sqrt{(n-1)n}}}{\sqrt{1 + 4x_3(2n)^{-1/2} + 2x_3^2}} = \frac{\mu_2(2n)^{-1/2}}{\sqrt{1 + 4x_3(2n)^{-1/2} + 2x_3^2}} \text{ and}$$

$$\frac{-\frac{n-1}{\sqrt{(n-1)n}\sqrt{2n}} + \left((2n)^{-1/2} + x_3\right)\left(1 + \frac{1}{n}\right)}{\sqrt{1 + 4x_3(2n)^{-1/2} + 2x_3^2}} = \frac{\mu_2\left((2n)^{-1/2} + x_3\right)}{\sqrt{1 + 4x_3(2n)^{-1/2} + 2x_3^2}}.$$

The last n lines are not considered for symmetry reasons. These equations can be rewritten:

(1) $$\frac{1}{n-1} - \frac{1 + x_3\sqrt{2n}}{\sqrt{(n-1)n}} = \mu;$$

(2) $$-\frac{\sqrt{n-1}}{\sqrt{n}} + (1 + x_3\sqrt{2n})(1 + 1/n) = \mu(1 + x_3\sqrt{2n}).$$

The two above equations are equivalent to:

$$x_3\sqrt{2n} = \frac{\sqrt{n}}{\sqrt{n-1}} - \sqrt{(n-1)n}\mu - 1 \text{ and } \mu = -\frac{\sqrt{n-1}}{\sqrt{n}}\frac{1}{1 + x_3\sqrt{2n}} + 1 + \frac{1}{n}.$$

Then, $\mu = -\frac{n-1}{n-(n-1)n\mu} + 1 + \frac{1}{n} \Leftrightarrow n(n-1)\mu^2 - (n^2 + n - 1)\mu + 2 = 0$ and taking the smallest solution yields:

$$\mu_3 = \frac{n^2 + n - 1 - \sqrt{(n^2 + n - 1)^2 - 8n(n-1)}}{2n(n-1)} = \frac{2}{n^2} + o(n^{-2})$$

and $x_3 = \frac{1}{\sqrt{2(n-1)}} - \frac{\sqrt{n-1}}{\sqrt{2}}\mu_3 - \frac{1}{\sqrt{2n}} = -\frac{3n^{-\frac{3}{2}}}{2\sqrt{2}} + o(n^{-\frac{3}{2}})$. Consequently, $\mu_4 = 1 + \frac{2n-1}{n(n-1)} - \mu_3$ and $x_4 = -\frac{\sqrt{n}}{\sqrt{2}} - \frac{3}{2\sqrt{2n}} + o(n^{-1/2})$.

Finally, we can conclude with the proof of Theorem 2.

Proof of Theorem 2. —

$$\begin{aligned}
D(U,U_b) &= \sum_{i\geq 3}^{2n}\sum_{j\geq 5}^{2n} 0\ \ u_i, v_j\ {}^2 + \sum_{i=1}^{2}\sum_{j=1}^{4} \frac{(\lambda_i-\mu_j)^2}{(\lambda_i+\mu_j)^\alpha}\ \ u_i, v_j\ {}^2 + \\
&\sum_{i=1}^{2}\sum_{j\geq 5}^{2n} \mu_j^{2-\alpha}\ \ u_i, v_j\ {}^2 + \sum_{i\geq 3}^{2n}\sum_{j=1}^{4} \frac{(\lambda_i-\mu_j)^2}{(\lambda_i+\mu_j)^\alpha}\ \ u_i, v_j\ {}^2 \\
&= \sum_{j=2}^{4} \mu_j^{2-\alpha}\ \ u_2, v_j\ {}^2 + \mu_2^{2-\alpha}\ \ u_1, v_2\ {}^2 + \sum_{i\geq 3}^{2n}\left(\frac{n}{n-1}\right)^{2-\alpha}\ \ u_i, v_1\ {}^2 + \\
&\sum_{i\geq 3}^{2n} \frac{\left(\frac{n}{n-1}-\mu_2\right)^2}{\left(\frac{n}{n-1}+\mu_2\right)^\alpha}\ \ u_i, v_2\ {}^2 + \sum_{i\geq 3}^{2n} \frac{\left(\frac{n}{n-1}-\mu_3\right)^2}{\left(\frac{n}{n-1}+\mu_3\right)^\alpha}\ \ u_i, v_3\ {}^2 + \sum_{i\geq 3}^{2n} \frac{\left(\frac{n}{n-1}-\mu_4\right)^2}{\left(\frac{n}{n-1}+\mu_4\right)^\alpha}\ \ u_i, v_4\ {}^2 \\
&= \mu_2^{2-\alpha}\ \ u_2, v_2\ {}^2 + \mu_3^{2-\alpha}\ \ u_2, v_3\ {}^2 + \mu_4^{2-\alpha}\ \ u_2, v_4\ {}^2 + \mu_2^{2-\alpha}\ \ u_1, v_2\ {}^2 + \\
&\left(\frac{n}{n-1}\right)^{2-\alpha}(1 -\ \ u_1, v_1\ {}^2 -\ \ u_2, v_1\ {}^2) + \frac{\left(\frac{n}{n-1}-\mu_2\right)^2}{\left(\frac{n}{n-1}+\mu_2\right)^\alpha}\left(1 -\ \ u_1, v_2\ {}^2 -\ \ u_2, v_2\ {}^2\right) + \\
&\frac{\left(\frac{n}{n-1}-\mu_3\right)^2}{\left(\frac{n}{n-1}+\mu_3\right)^\alpha}\left(1 -\ \ u_1, v_3\ {}^2 -\ \ u_2, v_3\ {}^2\right) + \frac{\left(\frac{n}{n-1}-\mu_4\right)^2}{\left(\frac{n}{n-1}+\mu_4\right)^\alpha}\left(1 -\ \ u_1, v_4\ {}^2 -\ \ u_2, v_4\ {}^2\right)
\end{aligned}$$

Computing the scalar products gives: $u_1, v_1\ {}^2 = \frac{[2(n-1)\sqrt{n-1}+2\sqrt{n}]^2}{4n[(n-1)^2+n]} = 1 + O(n^{-3})$, $u_1, v_2\ {}^2 = \frac{1}{4n(n-1)(n-1+1/n)}\left(2n-2-2\frac{(n-1)^{\frac{3}{2}}}{\sqrt{n}}\right)^2 = o(n^{-2})$, $u_1, v_3\ {}^2 = 0$, $u_1, v_4\ {}^2 = 0$, $u_2, v_1\ {}^2 = 0$, $u_2, v_2\ {}^2 = 0$, $u_2, v_3\ {}^2 = \frac{\left(1+\frac{\sqrt{2}x_3}{\sqrt{n}}\right)^2}{1+4x_3(2n)^{-1/2}+2x_3^2} = 1 + O(n^{-3})$ and $u_2, v_4\ {}^2 = \frac{\left(1+\frac{\sqrt{2}x_4}{\sqrt{n}}\right)^2}{1+4x_4(2n)^{-1/2}+2x_4^2} = O(n^{-3})$. Besides, $\frac{(\frac{n}{n-1}-\mu_2)^2}{\left(\frac{n}{n-1}+\mu_2\right)^\alpha} = \frac{1}{2^\alpha n^2} + o(n^{-2})$, $\frac{(\frac{n}{n-1}-\mu_3)^2}{\left(\frac{n}{n-1}+\mu_3\right)^\alpha} = 1 + o(1)$ and $\frac{(\frac{n}{n-1}-\mu_4)^2}{\left(\frac{n}{n-1}+\mu_4\right)^\alpha} = \frac{1}{2^\alpha n^2} + o(n^{-2})$.

Thus, gathering the above results, we obtain that:

$$\begin{aligned}
D(U,U_b) &= \mu_3^{2-\alpha} + \frac{2^{1-\alpha}}{n^2} + o(n^{-2}) \\
&= \left(\frac{2}{n^2}\right)^{2-\alpha} + \frac{2^{1-\alpha}}{n^2} + o(n^{-2}).
\end{aligned}$$

Now let us consider what happens if an edge is taken away from U, this gives U_r. Let us denote by K_n^{-1} the complete graph of size n with edge between vertices 1 and 2 taken away. Then, $U_r = K_n^{-1} \quad K_n$.

Theorem 3. — *The dissimilarity* $D(U,U_r) = \frac{2^{1-\alpha}}{n^2} + o(n^{-2})$.

Proof. — The eigenvectors of K_n are similar to those found in Theorem 2, that is: $\lambda_1 = 0, \lambda_2 = \cdots = \lambda_n = \frac{n}{n-1}, u_1 = \frac{1}{\sqrt{n}}(1,\ldots,1), u_2 = (1, \frac{-1}{n-1}, \ldots, \frac{-1}{n-1}), u_2 = (\frac{-1}{n-1}, 1, \frac{-1}{n-1}, \ldots, \frac{-1}{n-1})$ and $u_n = (\frac{-1}{n-1}, \ldots, \frac{-1}{n-1}, 1, \frac{-1}{n-1})$. Then,

$$L(K_n^-) = \begin{pmatrix} 1 & 0 & C^T & \\ 0 & 1 & & \\ C & & A_n^{n-2} & \end{pmatrix}$$

with

$$C^T = \begin{pmatrix} -\sqrt{(n-2)(n-1)}^{-1} & \ldots & -\sqrt{(n-2)(n-1)}^{-1} \\ -\sqrt{(n-2)(n-1)}^{-1} & \ldots & -\sqrt{(n-2)(n-1)}^{-1} \end{pmatrix}.$$

We remark for $L(K_n^{-1})$ that 0 is the eigenvalue associated to the eigenvector

$$v_1 = \frac{1}{\sqrt{(n+1)(n-2)}}(\sqrt{n-2}, \sqrt{n-2}, \sqrt{n-1}, \ldots, \sqrt{n-1}),$$

that 1 is the eigenvalue for the eigenvector

$$v_2 = \frac{1}{\sqrt{2}}(1, -1, 0, \ldots, 0),$$

that $\frac{n+1}{n-1}$ is the eigenvalue for the eigenvector

$$v_3 = \frac{1}{\sqrt{2\frac{n+1}{n-1}}}(1, 1, \frac{-2}{\sqrt{(n-1)(n-2)}}, \ldots, \frac{-2}{\sqrt{(n-1)(n-2)}})$$

and that $n/(n-1)$ for eigenvectors

$$\begin{aligned} v_4 &= (0, 0, 1, \frac{-1}{n-3}, \ldots, \frac{-1}{n-3}), \\ v_5 &= (0, 0, \frac{-1}{n-3}, 1, \frac{-1}{n-3}, \ldots, \frac{-1}{n-3}), \\ &\vdots \\ v_n &= (0, 0, \frac{-1}{n-3}, \ldots, \frac{-1}{n-3}, 1, \frac{-1}{n-3}). \end{aligned}$$

Then

$$\begin{aligned}
d(K_n, K_n^-) &= \sum_{1\le i,j\le n} \frac{(\lambda_i-\mu_j)^2}{(\lambda_i+\mu_j)^\alpha}\ u_i, v_j\ ^2\\
&= \sum_{j=2}^{n} \frac{(\lambda_1-\mu_j)^2}{(\lambda_1+\mu_j)^\alpha}\ u_1, v_j\ ^2 + \sum_{i=2}^{n} \frac{(\lambda_i-\mu_1)^2}{(\lambda_i+\mu_1)^\alpha}\ u_i, v_1\ ^2 + \sum_{i=2}^{n} \frac{(\lambda_i-\mu_2)^2}{(\lambda_i+\mu_2)^\alpha}\ u_i, v_2\ ^2 +\\
&\quad \sum_{i=2}^{n} \frac{(\lambda_i-\mu_3)^2}{(\lambda_i+\mu_3)^\alpha}\ u_i, v_3\ ^2 + \sum_{2\le i,4\le j} \frac{(\lambda_i-\mu_j)^2}{(\lambda_i+\mu_j)^\alpha}\ u_i, v_j\ ^2\\
&= \mu_3^{2-\alpha}\ u_1, v_3\ ^2 + \left(\frac{n}{n-1}\right)^{2-\alpha}(1-\ u_1, v_1\ ^2) + \frac{\left(\frac{n}{n-1}-1\right)^2}{\left(\frac{n}{n-1}+1\right)^\alpha}(1-\ u_1, v_2\ ^2)+\\
&\quad \frac{\left(\frac{n}{n-1}-\frac{n+1}{n-1}\right)^2}{\left(\frac{n}{n-1}+\frac{n+1}{n-1}\right)^\alpha}(1-\ u_1, v_3\ ^2).
\end{aligned}$$

Computing the scalar products gives: $u_1, v_1\ ^2 = 1 + O(n^{-3})$, $u_1, v_2\ ^2 = 0$ and finally $u_1, v_3\ ^2 = O(n^{-3})$. Besides, $\frac{\left(\frac{n}{n-1}-1\right)^2}{\left(\frac{n}{n-1}+1\right)^\alpha} = \frac{1}{2^\alpha n^2} + o(n^{-2})$ and $\frac{\left(\frac{n}{n-1}-\frac{n+1}{n-1}\right)^2}{\left(\frac{n}{n-1}+\frac{n+1}{n-1}\right)^\alpha} = \frac{1}{2^\alpha n^2} + o(n^{-2})$ so that:

$$d(K_n, K_n^-) = \frac{2^{1-\alpha}}{n^2} + o(n^{-2}). \tag{3}$$

Finally,

$$\begin{aligned}
d(U, U_r) &= d(K_n \quad K_n, K_n^{-1} \quad K_n)\\
&= d(K_n, K_n^{-1}) + d(K_n, K_n) \text{ (by Lemma 1)}\\
&= \frac{2^{1-\alpha}}{n^2} + o(n^{-2}) \text{ (using Equation 3).}
\end{aligned} \tag{4}$$

3. Conclusion

Taken together, the three theorems of this article show that it is very different to add an edge between two well connected graphs (complete graphs), and to delete one edge inside one of the complete graphs if α is greater than 1. Indeed, adding an edge that way builds a bridge and thus modifies the first eigencomponents which have the more impact on the dissimilarity as was shown empirically in our first work [9] and here formally. This property is interesting when one looks at graphs from the points of view of flows. If one keeps in mind that for $\alpha = 0$ we get the Bregman divergence, this means that $\alpha = 1$ is the smallest value to have our dissimilarity behave differently from Bregman divergence. A greater value of α would enhance this difference. This is confirmed by Theorem 1 showing that the farthest graph becomes the complete graph which is perfectly connected. Nevertheless, $\alpha = 2$ is to be avoided as in that case there is a continuity problem in expression $\frac{(\lambda-\mu)^2}{(\lambda+\mu)^2}$ when both eigenvalues tend to 0. Further work on dissimilarities between graphs would involve working on unlabeled graphs, coloured graph and multigraphs.

References

[1] F.R.K. Chung. *Spectral Graph Theory*. American Mathematical Society, 1997.

[2] A. Ghosh, S. Boyd, and A. Saberi. Minimizing effective resistance of a graph. In *Proceedings of the 17th International Symposium on Mathematical Theory of Networks and Systems, Kyoto, Japan*, pages 1185–1196, July 2006.

[3] H. Kashima, K.Tsuda, and A. Inokuchi. Marginalized kernels between labeled graphs. In *ICML*, pages 321–328, 2003.

[4] M. Koyuturk, A. Grama, and W. Szpankowski. An efficient algorithm for detecting frequent subgraphs in biological networks. In *ISMB/ECCB (Supplement of Bioinformatics)*, pages 200–207, 2004.

[5] C. H. Nguyen, N. Wicker, and H. Mamitsuka. Selecting graph cut solutions via global graph similarity. *IEEE Transactions on Neural Networks and Learning Systems*, 25:1407–1412, 2014.

[6] N. Pržulj. Biological network comparison using graphlet degree distribution. *Bioinformatics*, 26(6):853–854, march 2010.

[7] A. Sanfeliu and K. S. Fu. A distance measure between attributed relational graphs for pattern recognition. *IEEE Transactions on Systems, Man and Cybernetics*, 13:353–362, 1983.

[8] R. Sharan and T. Ideker. Modeling cellular machinery through biological network comparison. *Nature Biotechnology*, 24(4):427–433, april 2006.

[9] N. Wicker, C. H. Nguyen, and H. Mamitsuka. A new dissimilarity measure for labeled graphs. *Linear Algebra and its Applications*, 483:2331–2338, 2013.

December 4, 2015

Nicolas Wicker • *E-mail :* nicolas.wicker@math.univ-lille1.fr, Laboratoire Painlevé, Université de Lille 1, 59655 Villeneuve d'Ascq, France, +0(33)320434226

Canh Hao Nguyen • *E-mail :* canhhao@kuicr.kyoto-u.ac.jp, Bioinformatics Center, ICR, Kyoto University, Gokasho, Uji, Kyoto, 611-0011, Japan

Hiroshi Mamitsuka • *E-mail :* mami@kuicr.kyoto-u.ac.jp, Bioinformatics Center, ICR, Kyoto University, Gokasho, Uji, Kyoto, 611-0011, Japan