Guex, G. , Bavaud, F. : *Flow-Based Dissimilarities: Shortest Path, Commute Time, Max-Flow and Free Energy*. In: Lausen, B. et al. (Eds.) Data Science, Learning by Latent Structures, and Knowledge Discovery, pp. 101–111. (Series: Studies in Classification, Data Analysis, and Knowledge Organization). Springer, Heidelberg (2015)

Flow-based dissimilarities: shortest path, commute time, max-flow and free energy

Guillaume Guex[1] and François Bavaud[2]

[1] Department of Geography, University of Lausanne guillaume.guex@unil.ch
[2] Department of Geography, University of Lausanne francois.bavaud@unil.ch

Abstract. Random-walk based dissimilarities on weighted networks have demonstrated their efficiency in clustering algorithms. This contribution considers a few alternative network dissimilarities, among which a new max-flow dissimilarity, and more general *flow-based dissimilarities*, freely mixing shortest paths and random walks in function of a free parameter - the temperature. Their geometrical properties, and in particular their squared Euclidean nature are investigated through their power indices and multidimensional scaling properties. In particular, formal and numerical studies demonstrate the existence of critical temperatures, where flow-based dissimilarities cease to be squared Euclidean. The clustering potential of medium range temperatures is emphasised.

1 Introduction

The last decade has witnessed an increasing interest in centrality indices, community detection algorithms or network clustering assisted by random walks. Most approaches imply network dissimilarities, among which the shortest path and the commute time, closely linked to the minimisation of path functionals, namely a resistance or energy functional, respectively a relative entropy functional. The maximum flow and p-resistances (Alamgir and Von Luxburg (2011)) constitute alternative path functionals. Optimal flows minimising mixtures of path functionals characterise the global properties of the network, beyond the limited local view provided by binary or weighted adjacency matrices. Optimal flows also generate network dissimilarities, such as the presumably original maximum-flow dissimilarity (Section 2.2).

In particular, the free energy path functional (Saerens et al. (2009)) generates optimal flows interpolating between shortest paths and random walks (Section 2.4), where the edge resistances and transition matrix can be fixed independently (Bavaud and Guex (2012)). After reviewing the main definitions involved in the taxonomy of dissimilarities (Section 2.1), the geometric properties of the energy and free energy path functional dissimilarities (Section 2.5)

are investigated. In particular, the question of their squared Euclidean character is studied through the well-known Torgeson criterium, as well as through the less-known power index criterium of Joly and Le Calvé (1986). Numerical examples and applications demonstrate the existence of phase transitions, where path functional dissimilarities become non-Euclidean below some critical temperatures (Section 3.1). Section 2.5 addresses the issue of multidimensional scaling network reconstruction by path functional dissimilarities. Their better efficiency in clustering and classification of categorical data, as compared to chi-square dissimilarities, is illustrated in Section 3.3 for intermediate temperature ranges.

2 Dissimilarities

2.1 A few definitions and properties

Let us recall a few standard definitions (e.g. Joly and Le Calvé (1986); Citchley and Fichet (1994)): a dissimilarity on a set S of n objects is a $n \times n$ symmetric non-negative matrix $D = (d_{ij})$ with a null diagonal. The dissimilarity is *separable* if $d_{ij} = 0$ iff $i = j$, and *metric* if $d_{ij} \leq d_{ik} + d_{kj}$ (for all triples of S). A *distance* is a metric dissimilarity. One further distinguish between

- **ultrametric** distances ($D \in \mathcal{D}_U$) for which $d_{ij} \leq \max(d_{ik}, d_{jk})$.
- **Minkowski(q)** distances ($D \in \mathcal{D}_q$, where $q \geq 1$) if one can find n vectors $x_{ik} \in \mathbb{R}^p$ such that $d_{ij} = (\sum_{l=1}^{p} |x_{il} - x_{jl}|^q)^{\frac{1}{q}}$. $\mathcal{D}_2$ corresponds to the **Euclidean distance**.
- **squared Euclidean** dissimilarities ($D \in \mathcal{D}_2^2$) if there exists an embedding of the form $d_{ij} = \sum_{l=1}^{p} (x_{il} - x_{jl})^2$.
- **Chebychev** or **Frechet** distances ($D \in \mathcal{D}_\infty$) if there exists an embedding of the form $d_{ij} = \max_{l=1}^{p} |x_{il} - x_{jl}|$.

$\mathcal{D}_\infty$ is the set of all distances, and $\mathcal{D}_U \subset \mathcal{D}_2 \subset \mathcal{D}_1 \subset (\mathcal{D}_\infty \cap \mathcal{D}_2^2)$ holds.

Our study of the squared Euclidean character of the graph dissimilarities (Sections 2.5 and 3.2) mainly relies upon the following results, the first often attributed to Torgeson (1958) (with many precursors e.g. mentioned by Lew (1978)), and the second due to Joly and Le Calvé (1986). Here $f_i > 0$ is the relative weight of object i, normalized to unity, and δ_{ij} is Kronecker's delta:

Proposition 1. *A dissimilarity D on finite set S is $\mathcal{D}_2^2$ iff the* matrix of scalar products $B := -\frac{1}{2}HDH'$ *is positive semi-definite. Here, $H = (h_{ij})$ is the* centering matrix *where $h_{ij} = \delta_{ij} - f_j$, for any fixed normalised distribution f.*

Proposition 2. *For any dissimilarity D, there is a number $a \geq 0$ such that the elementwise power D^a is a squared Euclidean dissimilarity. Also, D^b is $\mathcal{D}_2^2$ as well for $0 \leq b \leq a$.*

Define $\mathrm{pow}(D)$*, the* ***power*** *of D, as the maximum value of a making D^a squared Euclidean. Then* $\mathrm{pow}(D) \geq 1$ *iff* $D \in \mathcal{D}_2^2$, $\mathrm{pow}(D) \geq 2$ *iff* $d \in \mathcal{D}_2$ *and* $\mathrm{pow}(D) = \infty$ *iff* $d \in \mathcal{D}_U$.

2.2 Commute-time, max-flow graph and chi-square distances

A binary graph $G = (V, E)$ on $|V| = n$ nodes is specified by a $n \times n$ symmetric adjacency matrix $A = (a_{ij})$ taking on values 0 or 1. The associated random walk is defined by the Markov transition matrix $W = (w_{ij})$ with $w_{ij} = a_{ij}/a_{i\bullet}$ (here "$\bullet$" denotes the summation over the replaced index).

Conversely, any regular Markov chain $W = (w_{ij})$ with stationary distribution f defines a *weighted graph* with associated node weights f_i, and *edge weights* or *exchange matrix* (Berger and Snell 1957) $e_{ij} := f_i w_{ij}$, giving the probability to select the pair of nodes ij. For unoriented graphs, $e_{ij} = e_{ji}$, that is W is *reversible*. By construction, $e_{i\bullet} = e_{\bullet i} = f_i$ and $e_{\bullet\bullet} = 1$.

Let $\mathcal{X}_{st}$ denote the set of *unit st-flows* from the source node $s \in V$ to the target node $t \in V$, as specified by the edge transitions counts $X = (x_{ij})$ of the trajectories or paths. By construction

$$x_{ij} \geq 0 \qquad , \qquad x_{i\bullet} - x_{\bullet i} = \delta_{is} - \delta_{it} \qquad \text{and} \qquad x_{t\bullet} = 0 \ . \tag{1}$$

The *commute-time distance* d^{ct}_{st} associated with W is the average time to go from s to t and back to i. That is $d^{\mathrm{ct}}_{st} = x^{st}_{\bullet\bullet} + x^{ts}_{\bullet\bullet}$, where x^{st}_{ij} is the random walk flow, obeying (1) and $x^{st}_{ij} = x^{st}_{i\bullet} w_{ij}$. One knows that $D^{\mathrm{ct}} \in (\mathcal{D}_\infty \cap \mathcal{D}_2^2)$, even in the oriented case (e.g. Boley et al. (2011)). Also, D^{ct} is *graph-geodetic* (Klein and Zhu (1998); Chebotarev (2010)), that is obeys $d^{\mathrm{ct}}_{ik} + d^{\mathrm{ct}}_{kj} = d^{\mathrm{ct}}_{jk}$ whenever all ij-paths and ji-paths moving over edges with non-zero weights pass through k.

Let us introduce a presumably new distance on unoriented weighted graphs, the ultrametric *max-flow distance* D^{mf}. In this setup, e_{ij} represents the *edge capacity*, controlling for the flow of maximum value v_{st} between s and t, solution of the problem

$$v_{st} := \max v \quad \text{such that} \quad 0 \leq x_{ij} \leq e_{ij}, \quad x_{i\bullet} - x_{\bullet i} = v(\delta_{is} - \delta_{it}), \quad x_{t\bullet} = 0.$$

By construction, $v_{ij} \geq 0$, $v_{ij} = v_{ji}$, $v_{ii} = \infty$ and $v_{ij} \geq \min(v_{ik}, v_{kj})$ for all triples in V^3. For $e_{ij} = e_{ji}$, define the max-flow distance as $d^{\mathrm{mf}}_{ij} := 1/v_{ij}$. Then d^{mf}_{ij} is a dissimilarity obeying $d^{\mathrm{mf}}_{ij} \leq \max(d^{\mathrm{mf}}_{ik}, d^{\mathrm{mf}}_{jk})$, that is $D^{\mathrm{mf}} \in \mathcal{D}_U$.

Categorial data analysis can also be cast in the above setup: let $N = (n_{il})$ be a $n \times m$ contingency table. Define a pair selection scheme by first choosing a row i, then a category l present in i, then another row j containing l. The resulting edge weight, node weight and transition matrix read (e.g. Bavaud and Xanthos 2005)

$$e_{ij} = \sum_{l=1}^{m} \frac{n_{il} n_{jl}}{n_{\bullet\bullet} n_{\bullet l}} \qquad f_i = \frac{n_{i\bullet}}{n_{\bullet\bullet}} \qquad w_{ij} = \sum_{l=1}^{m} \frac{n_{il} n_{jl}}{n_{i\bullet} n_{\bullet l}} \tag{2}$$

On the other hand, the chi-square dissimilarity $D^\chi = (d^\chi_{ij})$ between rows reads

$$d^\chi_{ij} := n_{\bullet\bullet} \sum_l \frac{1}{n_{\bullet l}} \left(\frac{n_{il}}{n_{i\bullet}} - \frac{n_{jl}}{n_{j\bullet}}\right)^2 \tag{3}$$

$D^\chi \in \mathcal{D}_2^2$, but $D^\chi \notin \mathcal{D}_\infty$. Neither D^χ nor D^{mf} are graph-geodetic.

2.3 Shortest-path distances

Let $r_{ij} \geq 0$ denote the length, cost, travel time or *resistance* of edge ij. The *shortest-path* length from s to t is

$$d^{\mathrm{sp}}_{st} := \min_{X \in \mathcal{X}_{st}} U(X) \qquad \text{where} \quad U(X) := \sum_{ij} r_{ij}\, x_{ij}$$

Then $d^{\mathrm{sp}}_{ii} = 0$, $d^{\mathrm{sp}}_{ik} + d^{\mathrm{sp}}_{jk} \geq d^{\mathrm{sp}}_{ij}$ and $d^{\mathrm{sp}}_{ij} = d^{\mathrm{sp}}_{ji}$ for symmetric $R = (R_{ij})$. In general, $D^{\mathrm{sp}} \notin \mathcal{D}_2^2$ (e.g. Deza and Laurent (1997) or Bavaud (2010)).

An important body of literature considers the *plain setup* $r_{ij} = c/a_{ij}$, where $c > 0$ is a normalisation constant, as e.g. in Yen (2008) and references therein, or as in the celebrated Doyle and Snell monograph (1984); see also the illustrations of Section 3. For $c = 1$, the seemingly counter-intuitive inequality $d^{\mathrm{ct}}_{ij} \leq d^{\mathrm{sp}}_{ij}$ holds, with equality iff the graph is a tree (e.g. Chandra et al. (1989); Deza and Deza (2009); Bavaud (2010)). Also, D^{sp} is graph-geodetic.

2.4 Interpolating random walks and shortest paths

Measuring the navigation effort within a weighted network depends on the nature of the moving agents (people, goods, money, information, etc), either knowledgeable of all networks characteristics, or only aware of their immediate neighborhood: d^{sp} models agents moving directly to their target, while d^{ct} models agents just wandering randomly until the target is reached. The former is more sensitive to short-cuts and to the length of paths in the network, while the latter is more sensitive to the degree and the number of paths between two nodes. Both capture information on the network structure, although of different kind.

This section presents a flow formalism aimed at continuously interpolating between the shortest path and the random walk, already detailed in Bavaud and Guex (2012)); see also Yen et al. (2008) and Saerens and al. (2009) for a close yet independent proposal, distinct in its implementation.

First, consider the general *twofold setup* endowed with two distinct edges valuations, namely the transition matrix $W = (w_{ij})$ of Section 2.2, and the resistances $R = (r_{ij})$ of Section 2.3. Both can be chosen independently, except for the consistency condition $r_{ij} = \infty$ iff $w_{ij} = 0$.

Secondly, define for each path $X = (x_{ij})$ the **flow energy** as

$$U(X) := \sum_{ij} r_{ij}\, \varphi(x_{ij})$$

where $\varphi(x)$ is a smooth non-decreasing function with $\varphi(0) = 0$. Flows of $\mathcal{X}_{st}$ minimizing $U(X)$ yield st-shortest paths for the choice $\varphi(x) = x$ and st-electic currents for the choice $\varphi(x) = x^2/2$ (Alamgir and Von Luxburg (2011); Li et al. (2011)). Define also the **flow entropy**:

$$G(X) := \sum_{ij} x_{ij} \ln \frac{x_{ij}}{x_{i\bullet} w_{ij}} = \sum_i x_{i\bullet} K_i(X||W) = x_{\bullet\bullet} \sum_i \frac{x_{i\bullet}}{x_{\bullet\bullet}} K_i(X||W)$$

where $K_i(X||W) := \sum_j \frac{x_{ij}}{x_{i\bullet}} \ln \frac{x_{ij}}{x_{i\bullet} w_{ij}} \geq 0$ is the Kullback-Leibler divergence between the empirical transitions $x_{ij}/x_{i\bullet}$ and the theoretical transitions w_{ij}. The entropy $G(X)$ takes on its minimum value zero iff $x_{ij}/x_{i\bullet} = w_{ij}$. Note that the multiplicative factor $x_{\bullet\bullet}$ aims at making $G(X)$ *homogeneous*, that is $G(vX) = vG(X)$ for $v > 0$.

Third, the aforementioned interpolation is implemented by considering the minimizing solution, noted $\tilde{X}^{st}$ or simply $\tilde{X}$, of the **free energy**

$$\tilde{X}^{st} := \arg\min_{X \in \mathcal{X}_{st}} F(X) \qquad\qquad F(X) := U(X) + T\, G(X)$$

where $T > 0$ is a free parameter, the **temperature**, which arbiters between the conflicting objectives: st-flows $\tilde{X}^{\mathrm{sp}}$ minimizing $F(X)$ realise shortest paths when $T \to 0$ (and $\varphi(x) = x$), and random walks $\tilde{X}^{\mathrm{rw}}$ when $T \to \infty$. We will also use $\beta = 1/T$, the **inverse temperature**.

On one hand, the feasible set $\mathcal{X}_{st}$ defined by (1) is convex; on the other hand, $F(X)$ is convex iff $\varphi(x)$ is convex, in which case the solution $\tilde{X}^{st}$ is unique and given by (Bavaud and Guex 2012)

$$\tilde{x}_{ij} = \tilde{x}_{i\bullet} w_{ij} \exp(-\beta[r_{ij}\varphi'(\tilde{x}_{ij}) + \lambda_i - \lambda_j]) \tag{4}$$

where λ_i are the Lagrange multipliers associated to (1). Equivalently, defining $v_{ij} := w_{ij} \exp(-\beta r_{ij} \varphi'(\tilde{x}_{ij}))$ as well as $V := (v_{ij})$ (where $i \neq t$ and $j \neq t$), $M := (I - V)^{-1}$, $q := (v_{it})_{i \neq t}$ and $z := Mq$, the solution reads:

$$\tilde{x}_{ij} = m_{si}\, v_{ij}\, \frac{z_j}{z_s} \quad (j \neq t) \qquad\qquad \tilde{x}_{it} = m_{si}\, \frac{q_i}{z_s} \ .$$

The optimal flow $\tilde{X} = (\tilde{x}_{ij})$ can be interpreted as the expected number of passages on edge ij when starting from s until eventually reaching t. The minimum free energy simply express as $F(\tilde{X}) = -T \ln z_s$. In what follows, we assume $\varphi(x) = x$. Then (4) is solved in a single step, instead of iteratively.

2.5 Energy and Free Energy Dissimilarities

Define, for any pair st of nodes, the **energy dissimilarity** D^{U} and the **free energy dissimilarity** D^{F} as

$$d^{\mathrm{U}}_{st} := \frac{1}{2}(U(\tilde{X}^{st}) + U(\tilde{X}^{ts})) \qquad\qquad d^{\mathrm{F}}_{st} := \frac{1}{2}(F(\tilde{X}^{st}) + F(\tilde{X}^{ts})) \ . \tag{5}$$

d^{U}_{st} is the expected resistance for going from s to t and coming back. In general, $D^{\mathrm{U}} \notin \mathcal{D}_\infty$; however, $D^{\mathrm{F}} \in \mathcal{D}_\infty$, and is graph-geodetic as well (see Kivimäki et al. (2012) and references therein). Furthemore, one can prove that

$$\lim_{\beta\to\infty} d^{\mathrm{F}}_{st} = \lim_{\beta\to\infty} d^{\mathrm{U}}_{st} = \frac{1}{2}(d^{\mathrm{sp}}_{st} + d^{\mathrm{sp}}_{ts})$$

$$\lim_{\beta\to 0} d^{\mathrm{F}}_{st} = \lim_{\beta\to 0} d^{\mathrm{U}}_{st} = \frac{1}{2}\sum_{ij} r_{ij}\,(\tilde{x}_{ij}^{st\ \mathrm{rw}} + \tilde{x}_{ij}^{ts\ \mathrm{rw}}) := d^{\mathrm{wct}}_{st}$$

where d^{wct}_{st} is the commute cost or *weighted commute time*, proportional to d^{ct}_{st} (Françoisse et al. (2013); Kivimäki et al. (2012)). The above suggests the possibility of an *Euclidean phase transition*, with dissimilarities in $\mathcal{D}_2^2$ for $T \geq T_c$, but not anymore for $T < T_c$ whenever $D^{\mathrm{sp}} \notin \mathcal{D}_2^2$.

3 Numerical examples and applications

3.1 Experiments with small graphs

Both D^{U} and D^{F} capture network information related to D^{sp} as well as to D^{ct}. Let us investigate their squared Euclidean nature by using the two criteria of Section (2.1). Figure 1 depicts the behaviour of the smallest eigenvalue of B and the power index, in the plain setup $r_{ij} := 1/a_{ij}$ and $w_{ij} := a_{ij}/a_{i\bullet}$.

The first example, K_{23}, demonstrates the existence of *critical temperatures* T_c in the sense of Section 2.5. The second example, C_{15}, shows D^{F} to be $\mathcal{D}_2^2$ in the whole temperature range, contrarily to D^{U} which is not $\mathcal{D}_2^2$ for intermediate values of β. In the third example, both D^{F} and D^{U} are $\mathcal{D}_2^2$ over the whole temperature range, with a contrasted behaviour between the last eigenvalue of B, monotonously decreasing, and the power, maximum around $\beta = 0.5$. On all examples, D^F stays in $\mathcal{D}_2^2$ longer than D^U when β is raised, a potentially interesting property since squared Euclidean dissimilarities are often needed for statistical applications. Lacunary as they are, those results underline the wide behavioral range of simple graphs, as expected from statistical mechanical entities.

3.2 Multidimensional scaling

A planar graph with $n = 50$ nodes, aimed at imitating a realistic road network, is generated following a variant of an algorithm due to Gastner and Newman (2006) (Figure 2). We define r_{ij} as the Euclidean distances between the pairs of nodes, and apply the simple setup $e_{ij} = c/r_{ij}$. After computing D^{U} or D^{F} for various β, we extract the MDS eigen-coordinates in the two first dimensions (regardless of the possible negativity of the last eigenvalue of B^{U}).

In addition, we evaluate the similarity between the original dissimilarities and the path functional dissimilarities by means of a presumably original *configuration similarity* index $\mathtt{CS}_{ab} := \frac{\mathrm{Tr}(B^a B^b)}{\sqrt{\mathrm{Tr}((B^a)^2)\mathrm{Tr}((B^b)^2)}} \in [0,1]$, where B^a and B^b are the scalar products corresponding to configurations D^a and D^b. The maximum similarity of $\mathtt{CS}_{\mathrm{planar,\ U}} = 0.86$ for which D^{U} is still squared Euclidean obtains for $\beta = 0.3$ (Figure 2).

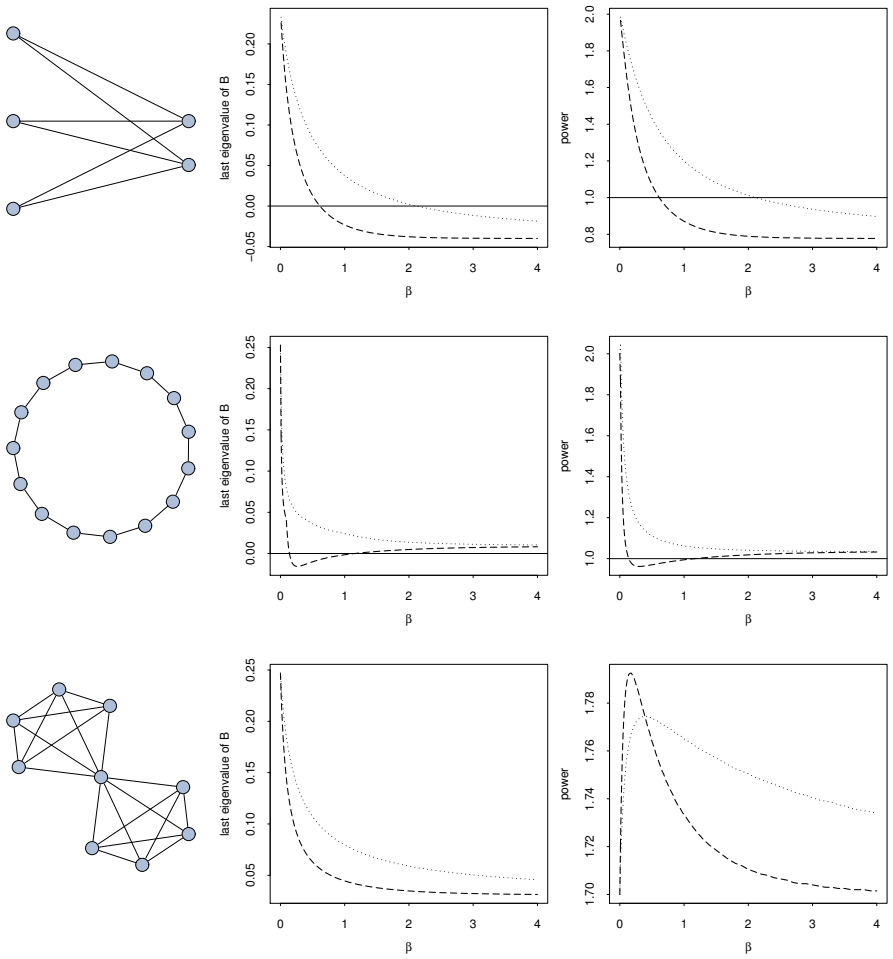

Fig. 1. Three toy graph examples, scanning the squared Euclidean character of D^{U} (dashed line) and D^{F} (dotted line). The first plot depicts the evolution of the last eigenvalue of B versus β. The second plot exhibits the power index versus β.

3.3 Clustering

A standard approach to clustering categorical data consists in computing chi-square dissimilarities D^{χ} (3) between objects, and then applying a k-means procedure. Alternatives based upon D^{U} and D^{F} might reveal more efficient, as demonstrated here in a supervised context with groups known a priori.

Specifically, one considers the document-terms contingency table $N = (n_{il})$ of the $n = 160$ documents of the `Reuters21578` corpus, belonging to $m = 8$ different groups (20 documents in each group). Exchanges e_{ij} obtain as in

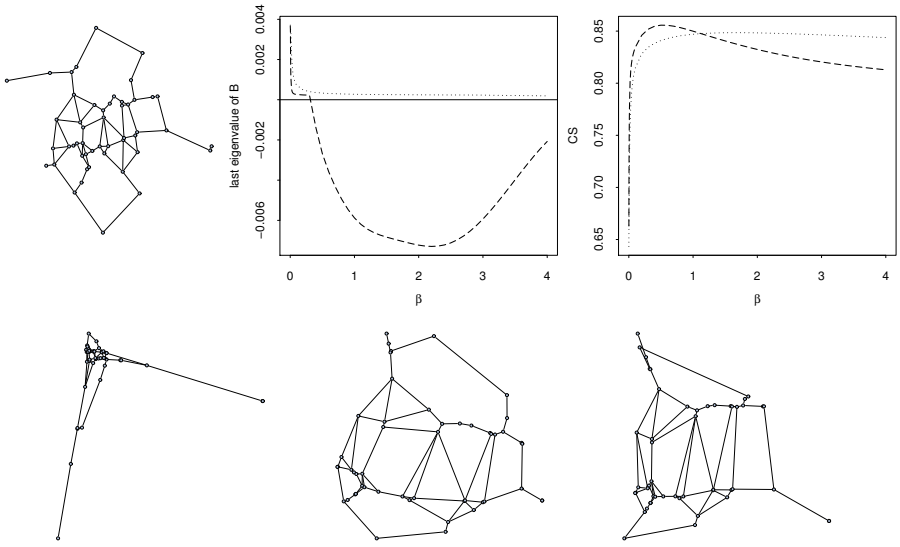

Fig. 2. Top: the original planar graph (left) and the behaviour, regarding β, of the last eigenvalue of B (middle) and the configuration similarity `CS` (right) between the original configuration and the graph dissimilarities D^U (dashed line) and D^{F} (dotted line). Bottom: MDS graph reconstruction from D^{U}, whose first two dimensions respectively explain 42.6% of the inertia (left, for $\beta = 0.001$), 53.8% (middle, for $\beta = 0.3$) and 48.1% (right, for $\beta = 4$). In the last case, the dissimilarity is not squared Euclidean and negative eigenvalues have been removed from the inertia.

(2), and resistances as $r_{ij} = 1/e_{ij}$ (plain setup). Document eigen-coordinates are obtained from *weighted MDS* on D^{U} and D^{F}, that is by considering the spectral decomposition of $K = (k_{ij})$ with $k_{ij} = \sqrt{f_i f_j} b_{ij}$ instead of $B = (b_{ij})$ (e.g. Bavaud (2010)); also, negative eigenvalues of K^{U} or K^{F} are set to zero. A k-means procedure with 8 clusters is then applied on the eigen-coordinates, and the resulting partition C is compared to the true partition C^{true} by means of the *variation of information dissimilarity* $d_{\mathrm{VI}}(C, C^{\mathrm{true}}) := H(C) + H(C^{\mathrm{true}}) - 2I(C, C^{\mathrm{true}})$, where H is the entropy of the partition and I the mutual information (Meila (2003)). Figure 3 shows that both D^{U} and D^{F} yield noticeably better results than D^{χ} for intermediate value of β, but quickly cease to be squared Euclidean. D^{U} gives for $\beta = 5 \cdot 10^{-6}$ the clustering most similar to the true classification, with a rate of correct classification (under optimal group permutation) of 65.63%. See Kivimäki et al. (2102) for further clustering experiments involving D^{U} and D^{F}, and e.g. Liu et al. (2013) for random walk clustering.

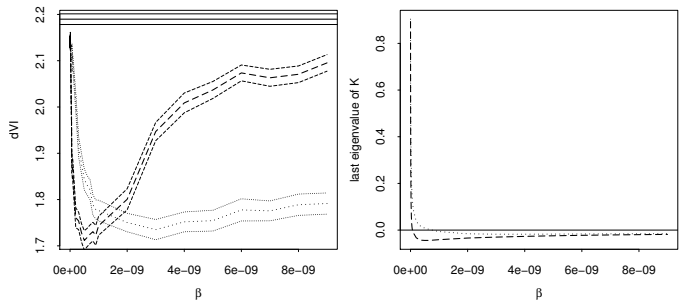

Fig. 3. Left: average information dissimilarity and 95% CI, comparing the true classification and the clustering obtained from D^{U} (dashed line), D^{F} (dotted line) and D^{χ} (baseline, top). Right: last eigenvalue of K^{U} (dashed line) and K^{F} (dotted line) as a function of β.

4 Conclusion

Transforming a graph into a dissimilarity matrix most facilitates the discussion of clustering and visualisation issues. It permits to open graph problems to a large body of statistical and mathematical methods: classical data analysis, machine learning, spectral graph theory and operations research. To that extent, enriching the family of flow-based dissimilarities by considering further sensible yet tractable functionals appears as a research priority. In particular, squared Euclidean dissimilarities immediately allow for MDS visualisation and Ward hierarchical clustering.

Various dissimilarities capture (and hide) various aspects of the graph. Results show that D^U generally gives a better representation of the graph structure than D^F for a precise value of the temperature T^{opt}, while the latter is more stable over temperature changes. This suggest that D^U should be used when T^{opt} is known, whereas D^F is preferable otherwise.

Despite the above results, the questions of knowing how to determine, even approximatively, T^{opt}, which aspect of the graph structure is best enlightened by which dissimilarity, and which dissimilarity is most efficient for a specific clustering or visualisation task, remain largely open.

References

ALAMGIR, M. and VON LUXBURG, U. (2011): Phase transition in the family of p-resistances. *Neural Information Processing Systems (NIPS 2011), 379–38.*

BAVAUD, F. (2010): Euclidean Distances, Soft and Spectral Clustering on Weighted Graphs. In: *Proceedings of ECML-PKDD 2010. LNCS 6321, 103–118.*

BAVAUD, F. and GUEX, G. (2012): Interpolating between random walks and shortest paths: a path functional approach. In: *Proceedings of SocInfo 2012. LNCS 7710, 68–81.*

BAVAUD, F. and XANTHOS, A. (2005): Markov Associativities. *Journal of Quantitative Linguistics 12, 123–137.*

BERGER, J. and SNELL, J.L. (1957): On the concept of equal exchange. *Behavioral Science 2, 111–118.*

BOLEY, D., RANJAN, G. and ZHANG, Z.-L. (2011): Commute Times for a Directed Graph using an Asymmetric Laplacian. *Linear Algebra and its Applications 435, 224–242.*

CHANDRA, A.K., RAGHAVAN, P., RUZZO, W.L., SMOLENSKY, R. and TIWARI, P. (1989): The Electrical Resistance Of A Graph Captures Its Commute And Cover Times. *Proceedings of the twenty-first annual ACM symposium on Theory of computing (STOC '89), 574–586.*

CHEBOTAREV, P. (2010): A class of graph-geodetic distances generalizing the shortest-path and the resistance distances. *Discr. Appl. Math. 159, 295–302.*

CRITCHLEY, F. and FICHET (1994) The partial order by inclusion of the principal classes of dissimilarity on a finite set, and some of their basic properties. In: B. van Cutsem (Ed.): *Classification and dissimilarity analysis. LNS 93, 5–65.*

DEZA, M. and DEZA, E. (2009): *Encyclopedia of Distances.* Springer.

DEZA, M. and LAURENT, M. (1997): *Geometry of cuts and metrics.* Springer.

DOYLE, P. and SNELL, J. (1984): *Random walks and electric networks.* Mathematical Association of America.

FRANÇOISSE, K., KIVIMÄKI, I., MANTRACH, A. ROSSI, F. and SAERENS, M. (2013): A bag-of-paths framework for network data analysis. *arXiv:1302.6766*

GASTNER, M.T. and NEWMAN, M.E.J. (2006): The spatial structure of networks. *Eur. Phys. J. B 49, 247–252.*

JOLY, S. and LE CALVÉ, G. (1986): Etude des puissances d'une distance. *Statistique et analyse des donn?es, 11, 30–50.*

KIVIMÄKI, I., SHIMBO, M. and SAERENS, M. (2012): Developments in the theory of randomized shortest paths with a comparison of graph node distances. *arXiv:1212.1666*

KLEIN, D.J. and ZHU, H.Y. (1998): Distances and volumina for graphs. *J. Math. Chem. 23, 179–195.*

LEW, J.S. (1978): Some counterexamples in multidimensional scaling. *Journal of Mathematical Psychology 17, 247–254.*

LI, Y., ZHANG; Z-L. and BOLEY, D. (2011): The Routing Continuum from Shortest-Path to All-Path: A Unifying Theory. *31st International Conference on Distributed Computing Systems (ICDCS), 847–856.*

LIU, S., MATZAVINOS, A. and SETHURAMAN, S. (2013): Random walk distances in data clustering and applications. *Adv. Data Analysis and Classif. 7, 83–108.*

MEILA, M. (2003): Comparing clusterings by the variation of information. *Proceedings of the Sixteenth Annual Conference of Computational Learning Theory (COLT).* Springer.

SAERENS, M., ACHBANY, Y., FOUSS, F. and YEN, L. (2009): Randomized Shortest-Path Problems: Two Related Models. *Neural Computation 21, pp. 2363–2404.*

TORGESON, W.S. (1958) : *Theory and methods of scaling.* Wiley, New York.

YEN, L., SAERENS, M., MANTRACH, A. and SHIMBO, M. (2008): A Family of Dissimilarity Measures between Nodes Generalizing both the Shortest-Path and the Commute-time Distances. *Proceedings of the 14th SIGKDD International Conference on Knowledge Discovery and Data Mining, pp. 785–793.*