Bavaud, F., Guex, G.: *Interpolating between Random Walks and Shortest Paths: a Path Functional Approach*. In: Aberer, K. et al. (Eds.) SocInfo 2012, LNCS 7710, pp. 68–81. Springer, Heidelberg (2012)

Interpolating between Random Walks and Shortest Paths: a Path Functional Approach

François Bavaud and Guillaume Guex *

Department of Computer Science and Mathematical Methods
Department of Geography
University of Lausanne, Switzerland

Abstract. General models of network navigation must contain a deterministic or drift component, encouraging the agent to follow routes of least cost, as well as a random or diffusive component, enabling free wandering. This paper proposes a thermodynamic formalism involving two path functionals, namely an energy functional governing the drift and an entropy functional governing the diffusion. A freely adjustable parameter, the temperature, arbitrates between the conflicting objectives of minimising travel costs and maximising spatial exploration. The theory is illustrated on various graphs and various temperatures. The resulting optimal paths, together with presumably new associated edges and nodes centrality indices, are analytically and numerically investigated.

1 Introduction

Consider a network together with an agent wishing to move (or wishing to move goods, money, information, etc.) from source node s to target node t. The agent seeks to minimise the total cost or duration of the move, but the ideal path may be difficult to realise exactly, in absence of perfect information about the network.

The above context is common to many behavorial and decision contexts, among which "small-word" social communications (Travers and Milgram 1969), spatial navigation (e.g. Farnsworth and Beecham 1999), routing strategy on internet networks (e.g. Zhou 2008, Dubois-Ferrière et al. 2011), and several others (e.g. Borgatti 2005; Newman 2005).

Trajectories can be coded, generally non-univocally, by $X = (x_{ij})$ where x_{ij} = "number of direct transitions from node i to node j". The use of the flow matrix X is central in Operational Research (e.g. Ahuja et al. 1993) and Markov Chains theory (e.g. Kemeny and Snell 1976); four optimal st-paths have in particular been extensively analysed *separately* in the litterature, namely the shortest-path, the random walk, the maximum flow (Freeman et al. 1991) and the electrical current (Kirchhoff 1850; Newman 2005; Brandes and Fleischer 2005).

* The specific remarks of two anonymous reviewers are gratefully acknowledged

This paper investigates the properties of *st*-paths resulting from the minimisation of a *free energy functional* $F(X)$, over the set $X \in \mathcal{X}$ of admissible solutions. $F(X)$ contains a resistance component privileging shortest paths, and an entropy component favouring random walks. The conflict is arbitrered by a continuous parameter $T \geq 0$, the *temperature* (or its *inverse* $\beta := 1/T$), and results in an analytically solvable unique optimum *continuously interpolating* between shortest-paths and random walks. See Yen et al. (2008) and Saerens and al. (2009) for a close proposal, yet distinct in its implementation.

Section 2 introduces the formalism, in particular the *energy functional* (based upon an edge resistance matrix R, symmetrical or not) and the *entropy functional* (based upon a Markov transition matrix W, reversible or not, related to R *or not*). Section 2.5 provides the analytic form of the unique solution minimising the free energy. Section 4 proposes the definition of edge and vertex betweenness centrality indices directly based upon the flow X. They are illustrated in sections 3 and 5 for various network geometries at various temperatures.

2 Definitions and solutions

2.1 Admissible paths

Consider a connected graph $G = (V, E)$ involving $n = |V|$ nodes together with two distinguished and distinct nodes, the source s and target t. The *st*-path or flow matrix, noted $X^{st} = (x^{st}_{ij})$ or simply $X = (x_{ij})$, counts the number of transitions from i to j along conserved unit paths starting at s, possibly visiting s again, and absorbed at t. Hence

$$x_{ij} \geq 0 \qquad \text{positivity} \tag{1}$$

$$x_{i\bullet} - x_{\bullet i} = \delta_{is} - \delta_{it} \qquad \text{unit flow conservation} \tag{2}$$

where δ_{ij} is the Kronecker delta, the components of the identity matrix. Here and in the sequel, $\bullet$ denotes the summation over the values of the replaced index, as in $x_{i\bullet} = \sum_{j=1}^{n} x_{ij}$. In particular, $x_{s\bullet} = x_{\bullet s} + 1$. Also,

$$x_{t\bullet} = 0 \qquad \text{absorbtion at } t \tag{3}$$

entailing $x_{tj} = 0$ for all j, and $x_{\bullet t} = 1$. Normalisation (2) can be extended to *valued flows*

$$x_{i\bullet} - x_{\bullet i} = v(\delta_{is} - \delta_{it}) \qquad \text{conservation for valued flow} \tag{4}$$

where $v \geq 0$, the amount sent through the network, is the *value* of the flow. Further familiar constraints consist of

$$x_{ij} \leq c_{ij} \qquad \text{capacity, where } c_{ij} \geq 0 \tag{5}$$

$$x_{ij} \geq b_{ij} \qquad \text{minimum flow requirement, } b_{ij} \geq 0 \tag{6}$$

$$x_{\bullet j_0} = 0 \qquad \text{forbidden node } j_0 \tag{7}$$

$$x_{i_0 j_0} = 0 \qquad \text{forbidden arc } (i_0 j_0)\ . \tag{8}$$

2.2 Mixtures and convexity

Any of the above constraints (1) to (8) or combinations thereof defines a *convex* set $\mathcal{X}$ of admissible *st*-paths: if X and Y are admissible, so is their *mixture* $\alpha X + (1-\alpha)Y$ for $\alpha \in [0,1]$. Mixture of paths are generally non-integer, and can be given a probabilistic interpretation, as in

- $x_{\bullet\bullet}$ = "average time (number of transitions) for transportation from s to t"
- $x_{ij}/x_{i\bullet}$ = "conditional probability to jump to j from i".

From now on, one considers by default unit flows X, generally non-integer, obeying (1), (2) and (3).

2.3 Path entropy and energy

Let $W = (w_{ij})$ denote the $(n \times n)$ transition matrix of some irreducible Markov chain. A *st*-path constitutes a random walk (as defined by W) iff $x_{ij}/x_{i\bullet} = w_{ij}$ for all visited node i, i.e. such that $x_{i\bullet} > 0$. Random walk *st*-paths X minimise the *entropy* functional

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

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 transition distributions X and W from i, taking on its minimum value zero iff $\frac{x_{ij}}{x_{i\bullet}} = w_{ij}$. Note $G(X)$ to be *homogeneous*, that is $G(vX) = v\,G(X)$ for $v > 0$, reflecting the *extensivity* of $G(X)$ in the thermodynamic sense.

By contrast, shortest-paths and other alternative optimal paths minimize *resistance* or *energy* functionals of the general form

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

where $r_{ij} > 0$ represent a cost or resistance associated to the directed arc ij, and $\varphi(x)$ is a smooth non-decreasing function with $\varphi(0) = 0$. In particular, minimizing $U(X)$ yields

- *st*-shortest paths for the choice $\varphi(x) = x$, where r_{ij} is the length of the arc ij
- *st*-electric currents from s to t for the choice $\varphi(x) = x^2/2$, where r_{ij} is the resistance of the conductor ij (see section 2.8).

As in Statistical Mechanics, we consider in this paper the class of admissible paths minimizing the *free energy*

$$F(X) := U(X) + T\,G(X) \ . \tag{9}$$

Here $T > 0$ is a free parameter, the *temperature*, controlling for the importance of the fluctuation around the trajectory of least resistance or energy (ground sate), realised in the low temperature limit $T \to 0$. In the high temperature limit $T \to \infty$ (or $\beta \to 0$, where $\beta := 1/T$ is the *inverse temperature*), the path consists of a random walk from s to t governed by W. Hence, minimising the free energy (9) generates for $T > 0$ *"heated extensions"* of classical minimum-cost problems $\min_X U(X)$, with the production of random fluctuations around the classical, "ground state" solution.

Derivating the free energy with respect to x_{ij}, and expressing the conservation constraints (2) through Lagrange multipliers $\{\lambda_i\}$ yields the optimality condition

$$T \ln \frac{x_{ij}}{x_{i\bullet} w_{ij}} + r_{ij} \varphi'(x_{ij}) = \lambda_j - \lambda_i \tag{10}$$

that is

$$x_{ij} = x_{i\bullet} w_{ij} \exp(-\beta[r_{ij} \varphi'(x_{ij}) + \lambda_i - \lambda_j]) \ . \tag{11}$$

The multipliers are defined up to an additive constant (see 15). In any case, $x_{ij} = 0$ when $w_{ij} = 0$ or $i = t$.

2.4 Minimum free energy and uniqueness

Multiplying (10) by x_{ij} and summing over all arcs yields an identity involving the entropy $G(X)$ of the optimal path X. Substitution in the free energy together with (2) demonstrates in turn the identity

$$\min_X F(X) = \sum_{ij} r_{ij} [\varphi(x_{ij}) - \varphi'(x_{ij}) x_{ij}] + \lambda_t - \lambda_s \ . \tag{12}$$

The first term is negative for $\varphi(x)$ convex, positive for $\varphi(x)$ concave, and zero for the heated shortest-path problem $\varphi(x) = x$, for which $\min_X F(X) = \lambda_t - \lambda_s$.

Also, the entropy functional is convex, that is $G(\alpha X + (1-\alpha) Y) \leq \alpha G(X) + (1-\alpha) G(Y)$ for two admissible paths X and Y and $0 \leq \alpha \leq 1$. The energy $U(X)$ is convex (resp. concave) iff $\varphi(x)$ is convex (resp. concave).

When a strictly convex functional $F(X)$ possesses a local minimum on a convex domain $\mathcal{X}$, the minimum is unique. In particular, we expect the optimal flows for $\varphi(x) = x^p$ to be unique for $p > 1$, but not anymore for $0 < p < 1$, where local minima may exist; see Alamgir and von Luxburg (2011) on "p-resistances".

In the shortest-path problem $p = 1$, the solution is unique if $T > 0$ (Section 2.5); when $T = 0$, local minima of $U(X)$ may coexist, yet all yielding the same value of $U(X)$.

2.5 Algebraic solution

Solving (11) is best done by considering separately the target node t. Define $v_{ij} := w_{ij} \exp(-\beta r_{ij} \varphi'(x_{ij}))$ as well as the $(n-1) \times (n-1)$ matrix $V = (v_{ij})_{i,j \neq t}$.

Also, define the $(n-1)$ dimensional vectors

$$a_i := x_{i\bullet} \exp(-\beta\lambda_i)|_{i\neq t} \qquad\qquad b_j := \exp(\beta\lambda_j)|_{j\neq t} \tag{13}$$
$$q_i := v_{it}|_{i\neq t} \qquad\qquad e_j := \delta_{js}|_{j\neq t}$$

Summing (11) over all i (for $j \neq t$, resp. $j = t$), then over all j for $i \neq t$ yields, using (2) and (3)

$$V'a = a - \exp(-\beta\lambda_s)\, e \qquad\qquad a'q = \exp(-\beta\lambda_t) \qquad\qquad Vb + \exp(\beta\lambda_t)\, q = b$$

Define the $(n-1) \times (n-1)$ matrix $M = (m_{ij})$ and the $(n-1)$ vector z as

$$M := (I - V)^{-1} = I + V + V^2 \ldots \qquad\qquad z := Mq \tag{14}$$

Then a and b express as

$$a_i = \exp(-\beta\lambda_s)\, m_{si} \qquad\qquad b_j = \exp(\beta\lambda_s)\, \frac{z_j}{z_s} = \exp(\beta\lambda_j)$$

implying incidentally

$$\lambda_j = T \ln z_j + C \overset{\text{(Section 2.6)}}{=} T \ln z_j + \lambda_t \quad . \tag{15}$$

Finally

$$x_{i\bullet} = m_{si}\, \frac{z_i}{z_s} \qquad\qquad x_{ij} = m_{si}\, v_{ij}\, \frac{z_j}{z_s} \quad (i \neq t) \tag{16}$$

$$x_{it} = m_{si}\, \frac{q_i}{z_s} \qquad\qquad x_{\bullet\bullet} = \frac{(Mz)_s}{z_s} = \frac{(M^2 q)_s}{(Mq)_s} \quad . \tag{17}$$

In general, V, M, q and d depend upon X. Hence (16) and (17) define a recursive system, whose fixed points may be multiple if $U(X)$ is not convex (Section 2.4), but converging to a unique solution for $p > 1$.

In the heated shortest-path case $p = 1$, the above quantities are independent of X. Hence the solution is unique, and particularly easy to compute in one single $O(n^3)$ step, involving matrix inversion, as illustrated in Sections 3 and 5.

2.6 Probabilistic interpretation

In addition to the absorbing target node t, let us introduce another "cemetery" or absorbing state 0, and define an extended Markov chain P on $n+1$ states with transition matrix

$$P = \begin{pmatrix} \begin{array}{c||c|c|c} & \mathtt{i} \neq \mathtt{t},\mathtt{0} & \mathtt{t} & \mathtt{0} \\ \hline\hline \mathtt{i} \neq \mathtt{t},\mathtt{0} & V & q & \rho \\ \hline \mathtt{t} & 0 & 1 & 0 \\ \hline \mathtt{0} & 0 & 0 & 1 \end{array} \end{pmatrix}$$

where $\rho_i = 1 - \sum_{k=1}^{n} v_{ik}$ is the probability of being absorbed at 0 from i in one step.

$M = (m_{ij})$ is the so-called *fundamental matrix* (see (14) and Kemeny and Snell 1976 p.46), whose components m_{ij} give the *expected number of visits from i to j*, before being eventually absorbed at 0 or t. Also, z_i (with $i \neq t, 0$) is the *survival probability*, that is to be, directly or indirectly, eventually absorbed at t rather than killed at 0, when starting from i. The higher the node survival probability, the higher the value of its Lagrange multiplier in view of (15).

Extending the latter to $j = t$ entails the consistency condition $z_t = 1$, making $\lambda_t \geq \lambda_i$ for all i. In particular, the free energy of the heated shortest-path case is, in view of (12),

$$F(X^{st}) = -T \ln z_s(T) \ ,$$

increasing (super-linearly in T) with the risk of being absorbed at 0 from s.

2.7 High-temperature limit

The energy term in (9) plays no role anymore in the limit $T \to \infty$ (that is $\beta \to 0$), and so does the absorbing state 0 above in view of $\rho_i = 0$. In particular, $z_i \equiv 1$ and $x_{ij}^{st} = m_{si} w_{ij}$ for $i \neq t$.

Also, $x_{\bullet\bullet}^{st}$ is the expected number of transitions needed to reach t from s. The *commute time distance* or *resistance distance* $x_{\bullet\bullet}^{st} + x_{\bullet\bullet}^{ts}$ is known to represent a *squared Euclidean distance* between states s and t: see e.g. Fouss et al. 2007, and references therein; see also Yen et al. (2008) and Chebotarev (2010) for further studies on resistance and shortest-path *distances*.

2.8 Low-temperature limit

Equations (11), (16) and (17) show the positivity condition $x_{ij} \geq 0$ to be automatically satisfied, thanks to the entropy term $G(X)$. However, the latter disappears in the limit $T \to 0$, where one faces the difficulty that the optimality condition (10) $r_{ij}\varphi'(x_{ij}) = \lambda_j - \lambda_i$ is still justified only if x_{ij} is freely adjustable, that is if $x_{ij} > 0$.

For the *st*-shortest path problem $\varphi(x) = x$, one gets, assuming the solution to be unique, the well-known characterisation (see e.g. Ahuja et al. (1993) p.107):

$$\begin{cases} r_{ij} = \lambda_j - \lambda_i & \text{if } x_{ij} > 0 \\ r_{ij} > \lambda_j - \lambda_i & \text{if } x_{ij} = 0 \end{cases}$$

occurring in the dual formulation of the *st*-shortest path problem, namely *"maximize $\lambda_t - \lambda_s$ subject to $\lambda_j - \lambda_i \leq r_{ij}$ for all i, j"*. Here λ_i is the shortest-path distance from s to i.

For the *st*-electrical circuit problem $\varphi(x) = x^2/2$, one gets $r_{ij}x_{ij} = \lambda_j - \lambda_i$ if $x_{ij} > 0$, in which case $x_{ji} > 0$ cannot hold in view of the positivity of the resistances, thus forcing $x_{ji} = 0$. Hence

$$\begin{cases} x_{ij} = \frac{\lambda_j - \lambda_i}{r_{ij}} > 0 & \text{if } \lambda_j > \lambda_i \\ x_{ij} = 0 & \text{otherwise} \end{cases}$$

expressing *Ohm's law* for the currrent intensity x_{ij} (Kirchhoff 1850), where λ_i is the electric potential at node i.

3 Illustrations and case studies: simple flow and net flow

Let us restrict on st-shortest path problems, i.e. $\varphi(x) = x$, whose free energy is homogeneous in the sense $F(vX) = vF(X)$ where $v > 0$ is the value of the flow in (4).

Graphs are defined by a $n \times n$ Markov transition matrix W together with a $n \times n$ positive resistance matrix R. Fixing in addition s, t and β, yields an unique *simple flow* x_{ij}^{st}, computable for any W (reversible or not) and any R (symmetric or not) - a fairly large set of tractable weighted networks.

An obvious class of networks consists of binary graphs, defined by a symmetric, off-diagonal adjacency matrix, with unit resistances and uniform transitions on existing edges (i.e. a simple random walk in the sense of Bollobás 1998).

Such are the graphs A (Figure 1) and B (Figure 2) below. Graph C (Figure 3) penalises in addition two edges forming short-cut from the point of view of W, but with increased values of their resistance.

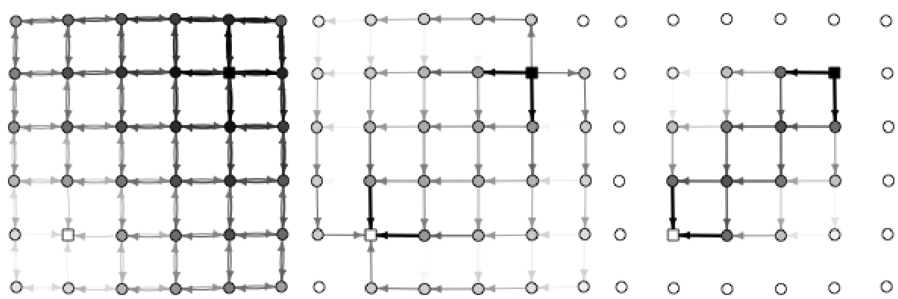

Fig. 1. Graph A is a square grid with uniform transitions and resistances. The resulting (high values in black, low values in light grey) simple flow x_{ij}^{st} and net flow ν_{ij}^{st} from s (black square) to t (white square) are depicted respectively on the left and middle picture with $\beta = 0$ (random walk) and on the right with $\beta = 50$ (shortest-path dominance). Note the simple flow and net flow to be identical at low temperatures.

Among the wide variety of graphs defined by a (W, R) pair, the plain graphs A, B and C primarily aim at illustrating the basic fact that, at high temperature, reverberation among neighbours of the source may dramatically lengthen the shortest path - an expected phenomenon (Figure 4).

Another quantity of interest is the *net flow*

$$\nu_{ij}^{st} := |x_{ij}^{st} - x_{ji}^{st}| \tag{18}$$

discounting "back and forth walks" inside the same edge, as discussed by Newman (2005): as a matter of fact, the presence of such alternate moves mechanically increases the simple flow inside an edge or node, especially near the source

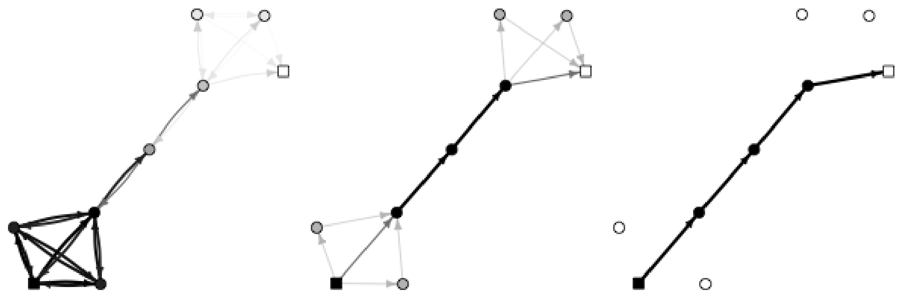

Fig. 2. Graph B consists of two cliques K_4 joined by two edges, with uniform transitions and resistances. Again, the resulting (high values in black, low values in light grey) simple flow x_{ij}^{st} and net flow ν_{ij}^{st} from s (black square) to t (white square) are depicted respectively on the left and middle picture with $\beta = 0$ (random walk) and on the right with $\beta = 50$.

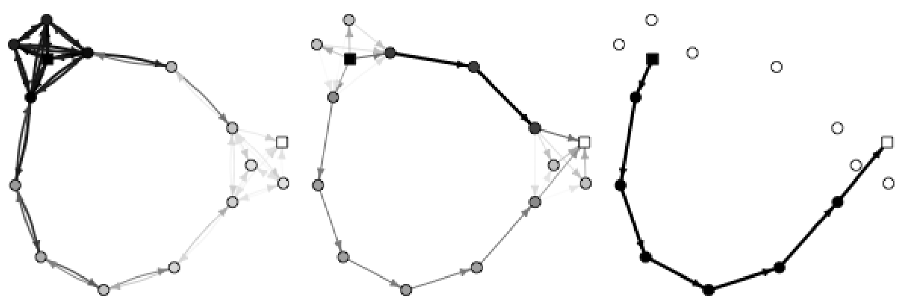

Fig. 3. Graph C consists of two cliques K_5 joined by two paths: the upper one consists of five edges, each with unit resistance, while the upper one contains two edges, each with resistance tenfold larger. The resulting (high values in black, low values in light grey) simple flow x_{ij}^{st} and net flow ν_{ij}^{st} from s (black square) to t (white square) are depicted respectively on the left and middle picture with $\beta = 0$ (random walk) and on the right with $\beta = 50$.

at high temperature (Figures 1, 2 and 3, left), giving the false impression the behaviour is more entropic (that is, random-walk dominated) around the source, which is erroneous.

The net flow "filters out" reverberations and hence captures the resulting "trend" of the agents within their random movements, who rarely go back along the edge from where they came if there is another way; cf. the circulation of "used goods" as defined in Borgatti (2005) along trails exempt of edges repetition. At low temperatures, the simple flow is directed in one way and hence converges to the simple flow (Figures 1, 2 and 3, right).

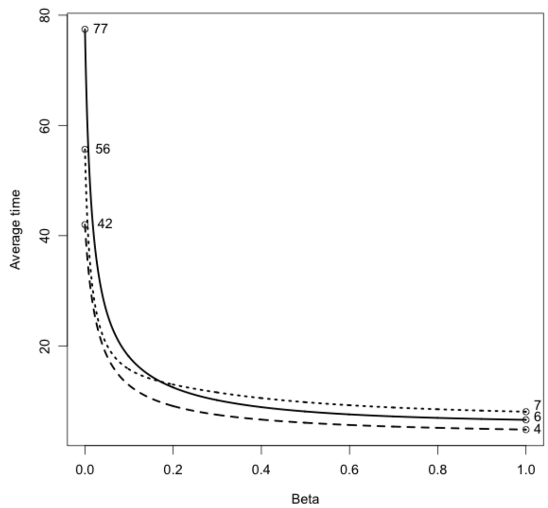

Fig. 4. The average time $x^{st}_{\bullet\bullet}$ to reach t from s is minimum for $T = 0$, and decreases with the inverse temperature β. Solid line: graph A; Dashed line: graph B; Dotted line: graph C.

4 Edge and vertex centrality betweenness

Several flow-based indices of betweenness centrality have been proposed ever since the shortest-path centrality pioneering proposal of Freeman (1977). In particular, random-walk centrality indices have been discussed by Noh and Rieger (2004) and Newman (2005). In this paper, we study the (unweighted) *mean flow betwenness*, defined for edges and vertices respectively (with complexity $O(n^5)$) as

$$\langle x_{ij} \rangle := \frac{1}{n(n-1)} \sum_{s,t|s \neq t} x^{st}_{ij} \qquad\qquad \langle x_{i\bullet} \rangle := \sum_j \langle x_{ij} \rangle = \langle x_{\bullet i} \rangle \tag{19}$$

where the latter identity results from the conservation condition (2). Definition (19) is intuitive enough: an edge is central if it carries a large amount of flow *on average*, that is by considering *all pairs of distinct source-targets couples*, thus extending the formalism to flows without specific source or target, such as monetary flows.

A more formal motivation arises from sensitivity analysis, with the result

$$\frac{\partial F(X(R))}{\partial r_{ij}} = \sum_{kl} \frac{\partial F(X(R))}{\partial x_{kl}(R)} \frac{\partial x_{kl}(R)}{\partial r_{ij}} + x_{ij}(R) = x_{ij}$$

where $F(X(R)) = \sum_{ij} r_{ij}\, x_{ij}(R) + TG(X(R))$ is the minimum free energy (9) under the constraints of Section 2.1 and r_{ij} the resistance of the edge ij.

Note that $\langle x_{\bullet\bullet} \rangle := \sum_j \langle x_{\bullet j} \rangle$ represents the average time to go from a vertex s to another vertex t and to return to s, averaged over all distinct pairs st. One can also define the *relative mean flow betwenness* as

$$c_{ij} := \frac{\langle x_{ij} \rangle}{\langle x_{\bullet\bullet} \rangle} \qquad\qquad c_i := \frac{\langle x_{i\bullet} \rangle}{\langle x_{\bullet\bullet} \rangle}$$

with the property $c_{ij} \geq 0$, $\sum_{ij} c_{ij} = 1$ and $c_i = c_{i\bullet} = c_{\bullet i}$.

Another candidate for a flow-based betweenness index is the *mean net flow*, again defined for edges and vertices as

$$\langle \nu_{ij} \rangle := \frac{1}{n(n-1)} \sum_{s,t|s \neq t} \nu_{ij}^{st} \qquad\qquad \langle \nu_{i\bullet} \rangle := \sum_j \langle \nu_{ij} \rangle = \langle \nu_{\bullet i} \rangle \tag{20}$$

Middle pictures in Figures 5, 6 and 7 below demonstrate how the mean net flow “substracts” the mechanical contribution arising from back and forth walks inside the same edge, in better accordance to a common sense notion of centrality.

Also, the sensitivity of the trip duration with respect to the edge resistance

$$\sigma_{ij} := \frac{\partial \langle x_{\bullet\bullet}(R) \rangle}{\partial r_{ij}}$$

constitutes yet another candidate, amenable to analytic treatment, whose study is beyond the size of the paper.

5 Case studies (continued): mean flow and mean net flow

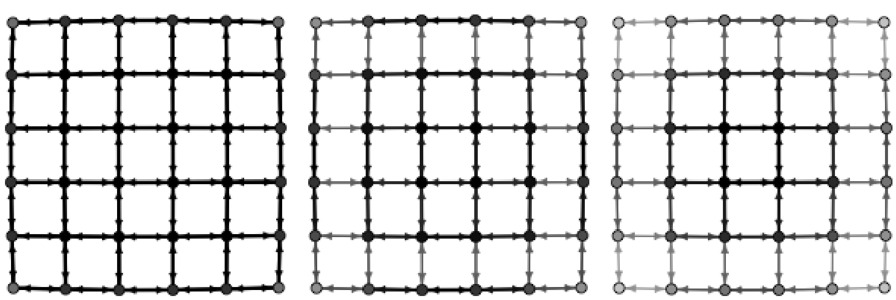

Fig. 5. Graph A: mean flow $\langle x_{ij} \rangle$ and mean net flow $\langle \nu_{ij} \rangle$, with $\beta = 0$ (left and middle) and $\beta = 50$ (right); high values in black, low values in light grey.

Figures 5, 6 and 7 depict the mean flow betweenness and the mean net flow betweenness (19) for the three graphs of Section 3, at high temperatures (left and middle) and low temperatures (right). Here $\langle x_{ij} \rangle = \langle x_{ji} \rangle$ due to the symmetry of R and the reversibility of W. Visual inspection confirms the role of the mean

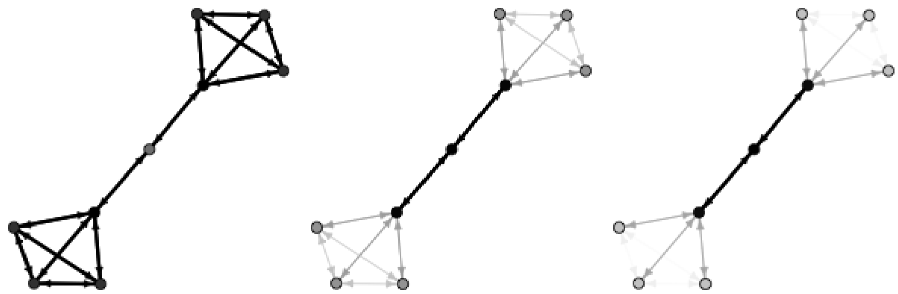

Fig. 6. Graph B: mean flow $\langle x_{ij} \rangle$ and mean net flow $\langle \nu_{ij} \rangle$, with $\beta = 0$ (left and middle) and $\beta = 50$ (right); high values in black, low values in light grey.

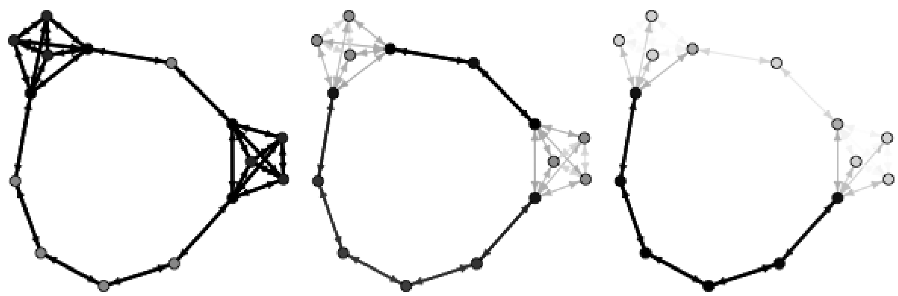

Fig. 7. Graph C: mean flow $\langle x_{ij} \rangle$ and mean net flow $\langle \nu_{ij} \rangle$, with $\beta = 0$ (left and middle) and $\beta = 50$ (right); high values in black, low values in light grey.

flow as a betweenness index, approaching the shortest-path betweenness at low temperatures.

At high temperatures, the mean flow $\langle x_{ij} \rangle$ turns out to be *constant* for all edges ij, a consistent observation for all "random-walk type" networks we have examined so far. As a consequence, the mean flow centrality of a node $\langle x_{i\bullet} \rangle$ *is proportional to its degree* for $\beta \to 0$, and identical to the shortest-path betweenness for $\beta \to \infty$. The former simply measures the local connectivity of the node, while the latter also takes into account the contributions of the remote parts of the network, in particular penalising high-resistance edges in comparison to low-resistance ones (Figure 7).

At low temperature, the net mean flow converges (together with the simple flow) to the shortest-path betweenness (Figures 5, 6 and 7, right). At high temperatures, the net mean flow betweenness is large for edges connecting clusters, but, as expected, small for edges inside clusters. Hence an original kind of centrality, the "net random walk betweenness", differing from shortest-path and degree betweeness, can be identified (Figures 5, 6 and 7, middle). As suggested

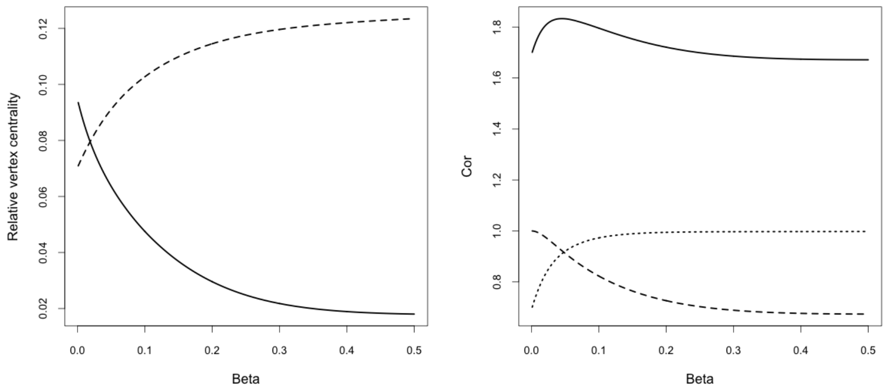

Fig. 8. Left: mean net flow centrality for the vertex in the "high-resistance path" (solid line) of network C, and for one of the nodes in the "low-resistance path" (dashed line) of network C. Right: inter-nodes correlation between the mean net flow centrality with itself at $\beta = 0$ (net random walk centrality; dashed line) and at $\beta = \infty$ (shortest-path node centrality; dotted line), in function of the inverse temperature β, for graph C. The sum of the two lines (solid line) is maximum for $\beta = 0.04$, arguably indicating a transition between an high- and a low-temperature regime.

in Figure 8 (right), contributions of both origins manifest themseves in the mean flow node centrality, for intermediate values of the temperature.

6 Conclusion

The paper proposes a coherent mechanism, easy to implement, interpolating between shortest paths and random walks. The construction is controlled by a temperature T and applies to any network endowed with a Markov transition matrix W and a resistance matrix R. The two matrices can be related, typically as (componentwise) inverses of each other (e.g. Yen et al. 2008) *or not*, in which case continuity at $T = 0$ and $T = \infty$ however requires $w_{ij} > 0$ whenever $r_{ij} < \infty$.

Modelling empirical st-paths necessitates to define W and R. The "simple symmetric model", namely unit resistances and uniform transitions on existing edges (Section 3) is, arguably, already meaningful in social phenomena and otherwise. For more elaborated applications, one can consider a possible model of tourist paths exploring Kobe (Iryio et al. 2012), consisting in choosing street directions as W with a bias towards "pleasant" street segments identified by low entries in R. Or the situation where a person at s wishes to be introduced to another person at t, by moving over an existing social network (defined by W) of friends, friends of friends, etc., where the resistance r_{ij} can express the difficulty that actor i introduces the person to actor j. One can also consider general situations where W expresses an average motion, a mass circulation, and R captures an individual specific shift, biased towards preferentially reaching a peculiar outcome t, such as a specific location, or an a-spatial goal such as fortune, power, marriage, safety, etc.

By contrast, the construction seems little adapted to the simulation of replicant agents (such as viruses, gossip or e-mails) violating in general the flow conservation condition (2).

The paper has defined and investigated a variety of centrality indices for edges and nodes. In particular, the mean flow betweenness interpolates between degree centrality and shortest-path centrality for nodes. Regarding edges, the mean net flow embodies various measures ranging from simple random-walk betweenness (as defined in Newman 2005) to shortest-path betweenness, again. The average time needed to attain another node, respectively being attained from another node

$$T_s^{\text{out}} := \frac{1}{n-1}\sum_{t\mid t\neq s} x_{\bullet\bullet}^{st} \qquad\qquad T_t^{\text{in}} := \frac{1}{n-1}\sum_{s\mid s\neq t} x_{\bullet\bullet}^{st}$$

constitute alternative centrality indices, generalising Freeman's *closeness centrality* (Freeman 1979), incorporating a drift component when $T > 0$.

Maximum-likelihood type arguments, necessitating a probabilistic framework not exposed here, suggest for W and R fixed the estimation rule for T

$$U(X^{st}) = U(X^{st}(T))$$

where U is the energy functional in Section 2.3. Here X^{st} is the observed, empirical path, and $X^{st}(T)$ is the optimal path (16, 17) at temperature T. Alternatively, T could be calibrated from the observed total time, using Figure 4 as an abacus.

References

Ahuja, R.K., Magnanti, T.L., Orlin, J.B.: Network Flows. Theory, algorithms and applications. Prentice Hall (1993)

Alamgir, M., von Luxburg, U.: Phase transition in the family of p-resistances. In: Neural Information Processing Systems (NIPS 2011), pp. 379–387 (2011)

Bollobás, B.: Modern Graph Theory. Springer (1998)

Borgatti, S.P.: Centrality and network flow. Social Networks 27, pp. 55–71 (2005)

Brandes, U., Fleischer, D.: Centrality Measures Based on Current Flow. In: Diekert, V., Durand, B. (eds.) STACS 2005. LNCS, vol. 3404, pp. 533–544. Springer (2005)

Chebotarev, P.: A class of graph-geodetic distances generalizing the shortest-path and the resistance distances. Discrete Applied Mathematics 159, pp. 295–302 (2010)

Dubois-Ferrière, H., Grossglauser, M., Vetterli, M.: Valuable Detours: Least-Cost Anypath Routing. IEEE/ACM Transactions on Networking 19, pp. 333–346 (2011)

Iryo, T., Shintaku, H., Senoo, S.: Experimental Study of Spatial Searching Behaviour of Travellers in Pedestrian Networks. In: 1st European Symposium on Quantitative Methods in Transportation Systems, EPFL Lausanne (2012) (contributed talk)

Kemeny, J.G., Snell, J.L.: Finite Markov Chains. Springer (1976)

Farnsworth, K.D., Beecham, J.A.: How Do Grazers Achieve Their Distribution? A Continuum of Models from Random Diffusion to the Ideal Free Distribution Using Biased Random Walks. The American Naturalist 153, pp. 509–526 (1999)

Fouss, F., Pirotte, A., Renders, J.-M., Saerens, M.: Random-Walk Computation of Similarities between Nodes of a Graph with Application to Collaborative Recommendation. IEEE Transactions on Knowledge and Data Engineering 19, pp. 355–369 (2007)

Freeman, L.C.: Centrality in networks: I. Conceptual clarication. Social Networks 1, pp. 215–239 (1979)

Freeman, L.C., Borgatti, S.P., White, D.R.: Centrality in valued graphs: A measure of betweenness based on network flow. Social Networks 13, pp. 141–154 (1991)

Kirchhoff, G.: On a deduction of Ohms laws, in connexion with the theory of electrostatics. Philosophical Magazine 37, p. 463 (1850)

Newman, M.E.J.: A measure of betweenness centrality based on random walks. Social Networks 27, pp. 39–54 (2005)

Noh, J.-D., Rieger, H.: Random walks on complex networks. Phys. Rev. Lett. 92, p. 118701 (2004)

Saerens, M., Achbany, Y., Fouss, F., Yen, L.: Randomized Shortest-Path Problems: Two Related Models. Neural Computation 21, pp. 2363–2404 (2009)

Travers, J., Milgram, S.: An experimental study of the small world problem. Sociometry 32, pp. 425–443 (1969)

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

Zhou, T.: Mixing navigation on networks. Physica A 387, pp. 3025–3032 (2008)