How to calculate the barycenter of a weighted graph

Sébastien Gadat

Toulouse School of Economics, Université Toulouse 1 Capitole sebastien.gadat@math.univ-toulouse.fr,
http://perso.math.univ-toulouse.fr/gadat/

Ioana Gavra

Institut de Mathématiques de Toulouse, Université Toulouse 3 Paul Sabatier Ioana.Gavra@math.univ-toulouse.fr,

Laurent Risser

Institut de Mathématiques de Toulouse, CNRS lrisser@math.univ-toulouse.fr, http://laurent.risser.free.fr/

Discrete structures like graphs make it possible to naturally and flexibly model complex phenomena. Since graphs that represent various types of information are increasingly available today, their analysis has become a popular subject of research. The graphs studied in the field of data science at this time generally have a large number of nodes that are not fairly weighted and connected to each other, translating a structural specification of the data. Yet, even an algorithm for locating the *average* position in graphs is lacking although this knowledge would be of primary interest for statistical or representation problems. In this work, we develop a stochastic algorithm for finding the Fréchet mean of weighted undirected metric graphs. This method relies on a noisy simulated annealing algorithm dealt with using homogenization. We then illustrate our algorithm with two examples (subgraphs of a social network and of a collaboration and citation network).

1. Introduction.

Numerous open questions in a very wide variety of scientific domains involve complex discrete structures that are easily modeled by graphs. The nature of these graphs may be weighted or not, directed or not, observed online or by using batch processing, each time implying new problems and sometimes leading to difficult mathematical or numerical questions. Graphs are the subject of perhaps one of the most impressive growing bodies of literature dealing with potential applications in statistical or quantum physics (see, *e.g.*, [21]) economics (dynamics in economy structured as networks), biology (regulatory networks of genes, neural networks), informatics (Web understanding and representation), social sciences (dynamics in social networks, analysis of citation graphs). We refer to [37] and [43] for recent communications on the theoretical aspects of random graph models, questions in the field of statistics and graphical models, and related numerical algorithms. In [34], the authors have developed an overview of numerous possible applications using graphs and networks in the fields of industrial organization and economics. Additional applications, details and references in the field of machine learning may also be found in [28].

In the meantime, the nature of the mathematical questions raised by the models that involve networks is very extensive and may concern geometry, statistics, algorithms or dynamical evolution over the network, to name a few. For example, we can be interested in the definition of suitable random graph models that make it possible to detect specific shape phenomena observed at different scales and frameworks (the small world networks of [47], the existence of a giant connected component for a specific range of parameters in the Erdös-Rényi random graph model [20]). A complete survey may be found in [39]. Another important field of investigation is dedicated to graph visualization (see some popular methods in [45] and [35], for example). In statistics, a popular topic

deals with the estimation of a natural clustering when the graph follows a specific random graph model (see, among others, the recent contribution [36] that proposes an optimal estimator for the stochastic block model and smooth graphons). Other approaches rely on efficient algorithms that analyze the spectral properties of adjacency matrices representing the networks (see, *e.g*, [6]). A final important field of interest deals with the evolution of a dynamical system defined through a discrete graph structure: this is, for instance, the question raised by gossip models that may describe belief evolution over a social network (see [1] and the references therein).

We address a problem here that may be considered as very simple at first glance: we aim to define and estimate the barycenter of weighted graphs. Hence, this problem involves questions that straddle the area of statistics and the geometry of graphs. Surprisingly, as far as we know, this question has received very little attention, although a good assessment of the location of a weighted graph barycenter could be used for fair representation issues or for understanding the graph structure from a statistical point of view. In particular, it could be used and extended to produce "second order" moment analyses of graphs. We could therefore generalize a Principal Component Analysis by extending this framework. This intermediary step was used, for example, by [13] to extend the definition of PCA on the space of probability measures on $\mathbb{R}$, and by [14] to develop a suitable geometric PCA of a set of images. A popular strategy to define moments in complex metric spaces is to use the variational interpretation of means (or barycenters), which leads to the introduction of Fréchet (or Karcher) means (see Section 2.2 for an accurate definition). This approach has been introduced in the seminal contribution [23] that makes it possible to define p-means over any metric probability space.

The use of Fréchet means has met with great interest, especially in the field of bio-statistics and signal processing, although mathematical and statistical derivations around this notion constitute a growing field of interest.

• In continuous domains, many authors have recently proposed limit theorems on the empirical Fréchet sample mean (only a sample of size n of the probability law is observed) towards its population counterpart. These works were mostly guided by applications to continuous manifolds that describe shape spaces introduced in [19]. For example, [38] establishes the consistency of the population Fréchet mean and derives applications in the Kendall space. The study of [9] establishes the consistency of Fréchet empirical means and derives its asymptotic distribution when $n \longrightarrow +\infty$, whereas the observations live in more general Riemannian manifolds. Finer results can be obtained in some non-parametric restrictive situations (see, *e.g.*, [11, 12, 15] dedicated to the so-called *shape invariant model*). Many applications in various domains involving signal processing can also be found: ECG curve analysis [10] and image analysis [44, 2], to name a few.

• Recent works treat Fréchet means in a discrete setting, especially when dealing with phylo-genetic trees that have an important hierarchical structure property. In particular, [8] proposed a central limit theorem in this discrete case, whereas [41] used an idea of Sturm for spaces with non-positive curvature to define an algorithm for the computation of the population Fréchet mean. Other works deal with the averaging of discrete structure sequences such as diagrams using the Wasserstein metric (see, *e.g.*, [42]) or graphs [27].

Our work deviates from the above-mentioned point since we build an algorithm that recovers the Fréchet mean of a weighted graph while observing an infinite sequence of nodes of the graph, instead of finding the population Fréchet mean of a set of discrete structures, as proposed in [8, 41, 42, 27]. Our algorithm uses recent contributions on simulated annealing ([3, 4]). It relies on a continuous-time noisy simulated annealing Markov process on graphs, as well as a second process that accelerates and homogenizes the updates of the noisy transitions in the simulated annealing procedure.

The paper is organized as follows: Section 2 highlights the different difficulties raised by the computation of the Fréchet mean of weighted graphs and describes the algorithm we propose. In

Section 3.1, theoretical backgrounds to understand the behavior of our method are developed on the basis of this algorithm, and Section 3.2 states our main results. Simulations and numerical insights are then given in Section 4. The convergence of the algorithm is theoretically established in Section 5, whereas Section 6 describes functional inequalities in quantum graphs that will be introduced below.

Acknowledgments.

2. Stochastic algorithm on quantum graphs

We consider $G = (V, E)$ a finite connected and undirected graph with no loop, where $V = \{1, \ldots, N\}$ refers to the N vertices (also called nodes) of G, and E the set of edges that connect some couples of vertices in G.

2.1. Undirected weighted graphs

The structure of G may be described by the adjacency matrix W that gives a non-negative weight to each edge E (pair of connected vertices), so that $W = (w_{i,j})_{1 \leq i,j \leq N}$ while $w_{i,j} = +\infty$ if there is no direct link between node i and node j. W indicates the length of each direct link in E: a small positive value of $w_{i,j}$ represents a small length of the edge $\{i, j\}$. We assume that G is undirected (so that the adjacency matrix W is symmetric) and connected: for any couple of nodes in V, we can always find a path that connects these two nodes. Finally, we assume that G has no self loop. Hence, the matrix W satisfies:

$$\forall i \neq j \quad w_{i,j} = w_{j,i} \qquad \text{and} \qquad \forall i \in V \quad w_{i,i} = 0$$

We define $d(x, y)$ as the geodesic distance between two points $(x, y) \in V^2$, which is the length of the shortest path between them. The length of this path is given by the addition of the length of traversed edges:

$$\forall (i, j) \in V^2 \qquad d(i, j) = \min_{i = i_1 \to i_2 \to \ldots \to i_{k+1} = j} \sum_{\ell=1}^{k-1} w_{i_\ell, i_{\ell+1}}.$$

When the length of the edges is constant and equal to 1, it simply corresponds to the number of traversed edges. Since the graph is connected and finite, we introduce the definition of the diameter of G:

$$\mathcal{D}_G := \sup_{(x,y) \in V^2} d(x, y).$$

To define a barycenter of a graph, it is necessary to introduce a discrete probability distribution ν over the set of vertices V. This probability distribution is used to measure the influence of each node on the graph.

EXAMPLE 1. Let us consider a simple scientometric example illustrated in Figure 1 and consider a "toy" co-authorship relation that could be obtained in a subgraph of a collaboration network like *zbMATH*[1]. If two authors A and B share $k_{A,B}$ joint papers, it is a reasonable choice to use a weight $w_{A,B} = \phi(k_{A,B})$, where ϕ is a convex function satisfying $\phi(0) = +\infty, \phi(1) = 1$ and $\phi(+\infty) = 0$. This means that no joint paper between A and B leads to the absence of a direct link between A and B
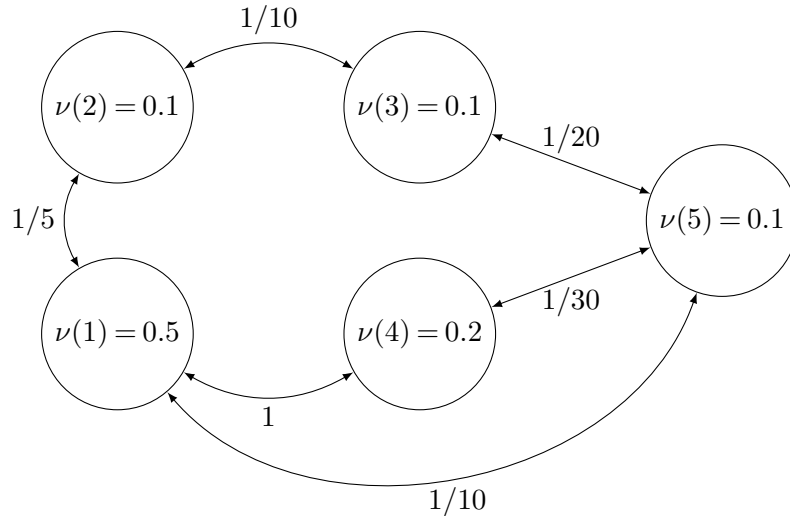
---

[1] https://zbmath.org/

Figure 1   Example of a weighted graph between five authors, where the probability mass is $\{0.5, 0.1, 0.1, 0.2, 0.1\}$. In this graph, Author 1 shares five publications with Author 2 and ten publications with Author 5, and so on.

on the graph. On the contrary, the more papers there are between A and B, the closer A and B will be on the graph. Of course, this graph may be embedded in a probability space with the additional definition of a probability distribution over the authors that can be naturally proportional to the number of citations of each author. This is a generalization of the Erdös graph. Note that this type of example can also be encountered when dealing with movies and actors, leading, for example, to the Bacon number and graph (see the website: www.oracleofbacon.org/).

2.2.   Fréchet mean of an undirected weighted graph

Following the simple remark that the p-mean of any distribution ν of $\mathbb{R}^d$ is the point that minimizes

$$x \longmapsto \mathbb{E}_{Z \sim \nu}[|x - Z|^p],$$

it is legitimate to be interested in the Fréchet mean of a graph (G, ν) where ν refers to the probability distribution over each node. The Fréchet mean is introduced in [23] to generalize this variational approach to any metric space. Although we have chosen to restrict our work to the case of $p = 2$, which corresponds to the Fréchet mean definition, we believe that our work could be generalized to any value of $p \geq 1$. If d denotes the geodesic distance w.r.t. G, we are interested in solving the following minimization problem:

$$M_\nu := \arg\min_{x \in E} U_\nu(x) \qquad \text{where} \qquad U_\nu(x) = \frac{1}{2} \int_G d^2(x, y)\nu(y). \tag{1}$$

Hence, the Fréchet mean of (G, ν), denoted M_ν, is the set of all possible minimizers of U_ν. Note that this set is not necessarily a singleton and this uniqueness property generally requires some additional topological assumptions (see [3], for example). At this point, we want to make three important remarks about the difficulty of this optimization problem:

• The problem of finding M_ν involves the minimization of U_ν, which is a non-convex function with the possibility of numerous local traps. To our knowledge, this problem cannot be efficiently solved using either a relaxed solution or using a greedy/dynamic programing algorithm (in the spirit of the Dijkstra method that makes it possible to compute geodesic paths [18]).

• It is thus natural to think about the use of a global minimization procedure, and, in particular, the simulated annealing (S.A. for short) method. S.A. is a standard strategy to minimize a function over discrete spaces and its computational cost is generally high. It relies on an inhomogeneous Markov process that evolves on the graph with a transition kernel depending on the energy estimates U_ν. In other words, it requires the computation of U_ν that depends on an integral w.r.t. ν. Since we plan to handle large graphs, this last dependency can be a very strong limitation.

• In some cases, even the global knowledge of ν may not be realistic, and the importance of each node can only be revealed through i.i.d. sequential arrivals of new observations in E that are distributed according to ν. This may be the case, for example, if we consider a probability distribution over E that is cropped while gathering interactive forms on a website.

2.3. Outline of Simulated Annealing (S.A.)

The optimization with S.A. introduces a Markov random dynamical system that evolves either in continuous time (generally for continuous spaces) or in discrete time (for discrete spaces). When dealing with a discrete setting, S.A. is based on a Markov kernel proposition $L(.,.) : V \times V \longrightarrow [0,1]$ related to the 1-neighborhoods of the Markov chain, as introduced in [30]. It is based on an inhomogeneous Metropolis-Hastings scheme, which is recalled in Algorithm 1. We can derive asymptotic guarantees of the convergence towards a minimum of U as soon as the cooling schedule is well chosen.

Algorithm 1: M.-H. Simulated Annealing

Data: Function U. Decreasing temperature sequence $(T_k)_{k \geq 0}$

1 Initialization: $X_0 \in V$;

2 for $k = 0 \ldots + \infty$ **do**

3 $\quad$ Draw $x' \sim L(X_k, .)$ and compute $p_k = 1 \wedge \left\{ e^{T_k^{-1}[U(X_k) - U(x')]} \frac{L(x', X_k)}{L(X_k, x')} \right\}$

4 $\quad$ Update X_{k+1} according to $X_{k+1} = \begin{cases} x' \text{ with probability } p_k \\ X_k \text{ with probability } 1 - p_k \end{cases}$

5 end

6 Output: $\lim_{k \longrightarrow +\infty} X_k$

When dealing with a continuous setting, S.A. uses a drifted diffusion with a vanishing variance $(\epsilon_t)_{t \geq 0}$ over V, or an increasing drift coefficient $(\beta_t)_{t \geq 0}$. We refer to [32, 40] for details and we recall its Langevin formulation in Algorithm 2.

Algorithm 2: Langevin Simulated Annealing

Data: Function U. Increasing inverse temperature $(\beta_t)_{t \geq 0}$

1 Initialization: $X_0 \in V$;

2 $\quad \forall t \geq 0 \quad dX_t = -\beta_t \nabla U(X_t) dt + dB_t$

3 Output: $\lim_{t \longrightarrow +\infty} X_t$

In both cases, we can see that S.A. with $U = U_\nu$ given by (1) involves the computation of the value of U in Line 4 of Algorithm 1, or the computation of ∇U in Line 2 of Algorithm 2. These two computations are problematic for our Fréchet mean problem: the integration over ν is intractable in the situation of large graphs and we are naturally driven to consider a noisy version of S.A. A possible alternative method for this problem is to use a homogenization technique: replacing U_ν in the definition of p_k by $U_y(\cdot) = d^2(\cdot, y)$, *where y is a value from an i.i.d. sequence $(Y_n)_{n \geq 0}$ distributed according to ν*. Such methods have been developed in [29] as a modification of Algorithm 1 with an additional Monte-Carlo step in Line 4, when $U_Y(x)$ follows a Gaussian distribution centered around the true value of $U(x) = \mathbb{E}_{Y \sim \nu} U_Y(x)$. This approach is still problematic in our case since the Gaussian assumption on the random variable $d^2(x, Y)$ is unrealistic here. Another limitation of this MC step relies on the fact that it requires a batch average of several $(U(x, Y_j))_{1 \leq j \leq n_k}$ where

n_k is the number of observations involved at iteration k, although we also plan to develop an algorithm that may be adapted to on-line arrivals of the observation $(Y_n)_{n \geq 0}$. Lastly, it is important to observe that the non-linearity of the exponential prevents the use of only *one* observation Y_k in the acceptation/reject ratio involved in the S.A. since it does not lead to an unbiased evaluation of the true transition:

$$\mathbb{E}_{Y \sim \nu} \left[e^{T_k^{-1}[U_Y(X_k) - U_Y(x')]} \right] \neq e^{T_k^{-1} \mathbb{E}_{Y \sim \nu}[U_Y(X_k) - U_Y(x')]}.$$

This difficulty does not arise in the homogenization of the simulated annealing algorithm in the continuous case since the exponential is replaced by a gradient, *i.e.*, the process we use is a Markov process of the form:

$$\mathrm{d}X_t = -\beta_t \nabla U_{Y_t}(X_t)dt + dB_t,$$

where B_t is a "Brownian motion" on the graph G, β_t^{-1} refers to the inverse of the temperature, and Y_t is a continuous time Markov process obtained from the sequence $(Y_n)_{n \in \mathbb{N}}$. This point motivates the introduction of the quantum graph induced by the initial graph. Of course, dealing with a continuous diffusion over a quantum graph G deserves special theoretical attention, which will be given in Section 3.1.

2.4. Homogenized S.A. algorithm on a quantum graph

We now present the proposed algorithm for estimating Fréchet means. To do so, we first introduce the quantum graph Γ_G derived from $G = (V, E)$ that corresponds to the set of points living inside the edges $e \in E$ of the initial graph. Once an orientation is arbitrarily fixed for each edge of E, the location of a point in Γ_G depends on the choice of an edge $e \in E$ and on a coordinate $x_e \in [0, L_e]$ where L_e is the length of edge e on the initial graph. The coordinate 0 then refers to the initial point of e and L_e refers to the other extremity.

While considering the quantum graph Γ_G, it is still possible to define the geodesic distance between any point $x \in \Gamma_G$ and any node $y \in G$. In particular, when $x \in V$, we use the initial definition of the geodesic distance over the discrete graph, although when $x \in e \in E$ with a coordinate $x_e \in [0, L_e]$, the geodesic distance between x and y is:

$$d(x, y) = \{x_e + d(e(0), y)\} \wedge \{L_e - x_e + d(e(L_e), y)\}.$$

This definition can be naturally generalized to any two points of Γ_G, enabling us to consider the metric space (Γ_G, d).

Consider a positive, continuous and increasing function $t \longmapsto \alpha_t$ such that:

$$\lim_{t \longmapsto +\infty} \alpha_t = +\infty \qquad \text{and} \qquad \forall t \geq 0 \quad \beta_t = o(\alpha_t).$$

We introduce $(N_t^\alpha)_{t \geq 0}$, an inhomogeneous Poisson process over $\mathbb{R}_+$ with intensity α. It is standard to represent N^α through a homogeneous Poisson process H of intensity 1 using the relationship:

$$\forall t \geq 0 \qquad N_t^\alpha = H_{h(t)}, \text{ where } h(t) = \int_0^t \alpha_s \mathrm{d}s.$$

Using this accelerated process N^α, our optimization algorithm over Γ_G is based on the Markov process that solves the following stochastic differential equation over Γ_G:

$$\begin{cases} X_0 \in \Gamma_G \\ dX_t = -\beta_t \nabla U_{Y_{N_t^\alpha}}(X_t)dt + dB_t. \end{cases} \tag{2}$$

Using the definition of N_t and the basic properties of a Poisson Process, it can be observed that for all $\epsilon > 0$ and all $t \geq 0$:

$$\mathbb{E}\left[\frac{N_{t+\epsilon} - N_t}{\epsilon}\right] = \frac{1}{\epsilon} \int_t^{t+\epsilon} \alpha_s \mathrm{d}s$$

Hence, α should be understood as the speed of new arrivals in the sequence $(Y_n)_{n \in \mathbb{N}}$.

Our definition (2) is slightly inaccurate: for any $y \in V(G)$, the function U_y is $\mathcal{C}^1$ except at a finite set of points of Γ_G:

- This can be the case inside an edge $x \in e$ with $x_e \in (0, L_e)$ when at least two different geodesic paths from x to y start in opposite directions. Then, we define $\nabla U_y(x) = 0$.
- This can also be the case at a node $x \in V$ when several geodesic paths start from x to y. In that case, we once again arbitrarily impose a null value for the "gradient" of U_y at node x. This prior choice will not have any influence on the behavior of the algorithm.

We will first detail below the theoretical objects involved in (2), then we will describe an efficient discretization of (2) that makes it possible to derive our practical optimization algorithm.

Algorithm 3: Homogenized Simulated Annealing over a quantum graph

---

Data: Function U. Increasing inverse temperature $(\beta_t)_{t \geq 0}$. Intensity $(\alpha_t)_{t \geq 0}$

1 Initialization: Pick $X_0 \in \Gamma_G$. ;

2 $T_0 = 0$;

3 for $k = 0 \ldots + \infty$ **do**

4   **while** $N_t^\alpha = k$ **do**

5 X_t evolves as a Brownian motion, relatively to the structure of Γ_G initialized at $X_{T_k}^-$.

6   **end**

7 $T_{k+1} := \inf\{t : N_t^\alpha = k+1\}$;

8   At time $t = T_{k+1}$, draw $Y_{k+1} = Y_{N_t^\alpha}$ according to ν.

9   The process X_t jumps from X_t^- towards Y_{k+1}:

$$X_t = X_t^- + \beta_t \alpha_t^{-1} \overrightarrow{X_t Y_{N_t^\alpha}}, \tag{3}$$

     where $\overrightarrow{X_t Y_{N_t^\alpha}}$ represents the shortest (geodesic) path from X_t to $Y_{N_t^\alpha}$ in Γ_G.

10 end

11 Output: $\lim_{t \longrightarrow +\infty} X_t$

Algorithm 3 could be studied following the road map of [5]. Nevertheless, this implies serious regularity difficulties on the densities and the Markov semi-group involved. Hence, we have chosen to consider Algorithm 3 as a natural Euler explicit discretization of our Markov evolution (2): for a large value of k, the average time needed to travel from T_k to T_{k+1} is approximately $\alpha_{T_k}^{-1} \longrightarrow 0$ as $k \longrightarrow +\infty$. On this short time interval, the drift term in (2) is the gradient of the squared geodesic distance between X_{T_k} and $Y_{N_{T_k}^\alpha}$, which is approximated by our vector $\overrightarrow{X_t Y_k}$, multiplied by β_{T_k}, leading to (3). X_t now evolves as a Brownian motion over Γ_G between two jump times and this evolution can be simulated with a Gaussian random variable using a (symmetric) random walk when the algorithm hits a node of Γ_G. Figure 2 proposes a schematic evolution of $(X_t)_{t \geq 0}$ over a simple graph Γ_G with five nodes. We will prove the following result.

THEOREM 1. *A constant $c^\star(U_\nu)$ exists such that if $\alpha_t = 1 + t$ and $\beta_t = b \log(1+t)$ with $b > \{c^\star(U_\nu)\}^{-1}$, then $(X_t)_{t \geq 0}$ defined in (2) converges a.s. to M_ν defined in (1), when t goes to infinity.*

A more rigorous form of this result and more details about the process defined in (2) are presented in Section 3.2. The proof of Theorem 3 is deferred to Section 5.

3. Inhomogeneous Markov process over Γ_G

This section presents the theoretical background needed to define the Markov evolution (2).

Figure 2 Schematic evolution of the Homogenized S.A. described in Algorithm 3 over the quantum graph. We first observe a jump at time T_1 towards $Y_{T_1} = V_1$ and a Brownian motion on Γ_G during $T_2 - T_1$. A second node is then sampled according to ν: here, $Y_{T_2} = V_5$ and a jump towards Y_{T_2} occurs at time T_2.

3.1. Diffusion processes on quantum graphs

We adopt here the convention introduced in [24] and fix for any edge $e \in E$ of length L_e an orientation (and parametrization s_e). This means that $s_e(0)$ is one of the extremities of e and $s_e(L_e)$ is the other one. By doing so, we have determined an orientation for Γ_G.

Dynamical system inside one edge Following the parametrization of each edge, we can define the second order elliptic operator $\mathcal{L}_e$ as $\forall (x, y, t) \in e \times V(G) \times \mathbb{R}_+$:

$$\mathcal{L}_e f(x, y, t) = -\beta_t \nabla_x U_y(x) + \frac{1}{2}\Delta_x f(x, y) + \alpha_t \int_G [f(x, y') - f(x, y)]d\nu(y'), \tag{4}$$

which is associated with (2) when $x \in e$: the x-component follows a standard diffusion drifted by $\nabla_x U_y(.)$ inside the edge e, although the y-component jumps over the nodes of the initial graph G with a jump distribution ν and a rate α_t.

Since the drift term $\nabla U_y(.)$ is measurable w.r.t. the Lebesgue measure over e and the second-order part of the operator is uniformly elliptic, $\mathcal{L}_e$ uniquely defines (in the weak sense) a diffusion process up to the first time the process hits one of the extremities of e (see, *e.g.*, [33]), which leads to a Feller Markov semi-group.

Dynamical system near one node We adopt the notation of [26] and write $e \sim v$ when a vertex $v \in V$ is an extremity of an edge $e \in E$. For any function f on Γ_G, at any point $v \in V$, we can define the directional derivative of f with respect to an edge $e \sim v$ according to the parametrization of e. If n_v denotes the number of edges e such that $e \sim v$, we then obtain n_v directional derivatives designated as $(d_e f(v))$:

$$d_e f(v) = \begin{cases} \lim\limits_{h \longrightarrow 0^+} \dfrac{f(s_e(h)) - f(s_e(0))}{h} & \text{if} \quad s_e(0) = v \\[2mm] \lim\limits_{h \longrightarrow 0^-} \dfrac{f(s_e(L_e + h)) - f(s_e(L_e))}{h} & \text{if} \quad s_e(L_e) = v. \end{cases}$$

It is shown in [26] that general dynamics over quantum graphs depend on a set of positive coefficients:

$$\mathcal{A} := \left\{ (a_v, (a_{e,v})_{e \sim v}) \in \mathbb{R}_+^{1+n_v} \text{ s.t. } a_v + \sum_{e \sim v} a_{e,v} > 0 : v \in V \right\}.$$

There then exists a one-to-one correspondence between $\mathcal{A}$ and the set of all possible continuous Markov Feller processes on Γ_G. More precisely, if the global generator $\mathcal{L}_t$ is defined as

$$\forall (x,y) \in \Gamma_G \times V \qquad \mathcal{L}_t f(x,y) = \mathcal{L}_{e,t}(f)(x,y) \qquad \text{when} \qquad x \in e,$$

while f belongs to the domain:

$$\mathcal{D}(\mathcal{L}) := \left\{ \forall y \in G \quad f(.,y) \in \mathcal{C}_b^\infty(\Gamma_G) : \forall v \in V \qquad \sum_{e \sim v} a_e d_e f(v,y) = 0 \right\}, \tag{5}$$

then the martingale problem is well-posed (see [26, 22]). In our setting, the *gluing conditions* are defined through the following set of coefficients $\mathcal{A}$:

$$\forall v \in V \qquad a_v = 0 \qquad \text{and} \qquad \forall e \sim v \qquad a_{e,v} = \frac{1}{n_v}.$$

The gluing conditions defined in (5) induce the following dynamics: when the x component of $(X_t, Y_t)_{t \geq 0}$ hits an extremity $v \in V$ of an edge e, it is instantaneously reflected in one of the n_v edges connected to v (with a uniform probability distribution over the connected edges) while spending no time on v. Using the uniform ellipticity of $\mathcal{L}_e$ and the measurability of the drift term, Theorem 2.1 of [26] can be adapted, providing the well-posedness of the Martingale problem associated with $(\mathcal{L}, \mathcal{D}(\mathcal{L}))$, and the next preliminary result can then be obtained.

THEOREM 2. *The operator $\mathcal{L}$ associated with the gluing conditions $\mathcal{A}$ generates a Feller Markov process on $\Gamma_G \times V \times \mathbb{R}_+$, with continuous sample paths on the x component. This process is weakly unique and follows the S.D.E. (2) on each $e \in E$.*

For simplicity, Δ_x will refer to the Laplacian operator with respect to the x-coordinate on the quantum graph Γ_G, using our gluing conditions (5) and the formalism introduced in [26].

3.2. Convergence of the homogenized S.A. over Γ_G

As mentioned earlier, we use a homogenization technique that involves an auxiliary sequence of random variables $(Y_n)_{n \geq 1}$, which are distributed according to ν. More specifically, the stochastic process $(X_t, Y_t)_{y \geq 0}$ described above is depicted by its inhomogeneous Markov generator, which can be split into three parts:

$$\mathcal{L}_t f(x,y) = \mathcal{L}_{1,t} f(x,y) + \mathcal{L}_{2,t} f(x,y).$$

In the equality above, $\mathcal{L}_{1,t}$ is the part of the generator that acts on Y_t:

$$\mathcal{L}_{1,t} f(x,y) = \alpha_t \int [f(x,y') - f(x,y)] \nu(\mathrm{d}y'), \tag{6}$$

describing the arrival of a new observation $Y \sim \nu$ with a rate α_t at time t. Concerning the action on the x component, the generator is:

$$\mathcal{L}_{2,t} f(x,y) = \frac{1}{2} \triangle_x f(x,y) - \beta_t < \nabla_x U_y, \nabla_x f(x,y) >. \tag{7}$$

Since the couple $(X_t, Y_t)_{t \geq 0}$ is Markov with a renewal of Y with ν, it can be immediately observed that the y component is distributed at any time according to ν. We introduce the notation m_t to refer to the distribution of the couple (X_t, Y_t) at time t, and we define n_t as the marginal distribution of X_t. In the following, we will also need to deal with the conditional distribution of

Y_t given the position X_t in Γ_G. We will refer to this probability distribution as $m_t(y|x)$. To sum up, we have:

$$\mathscr{L}(X_t, Y_t) = m_t \quad \text{with} \quad n_t(x)dx = \int_V m_t(x, y)dy \quad \text{and} \quad m_t(y\,|x) := \mathbb{P}[Y = y\,|X_t = x]. \qquad (8)$$

A traditional method for establishing the convergence of S.A. towards the minimum of a function U_ν consists in studying the evolution of the law of $(X_t)_{t \geq 0}$ and, in particular, its close relationship with the Gibbs field μ_{β_t} with energy U_ν and inverse temperature β_t:

$$\mu_{\beta_t} = \frac{1}{Z_{\beta_t}} \exp(-\beta_t U_\nu(x)), \qquad (9)$$

where Z_{β_t} is the normalization factor, *i.e.*, $Z_{\beta_t} = \int e^{-\beta_t U_\nu(x)} dx$.

Using a slowly decreasing temperature scheme $t \longmapsto \beta_t^{-1}$, it is expected that the process $(X_t)_{t \geq 0}$ evolves sufficiently fast (over the state space) in the ergodic sense so that its law remains close to μ_{β_t}. In addition, the Laplace method on the sequence $(\mu_{\beta_t})_{t \geq 0}$ ensures that the measure is concentrated near the global minimum of U_ν (see, for example, the large deviation principle associated with $(\mu_\beta)_{\beta \to +\infty}$ in [25]).

Hence, a natural consequence of the convergence "$\mathscr{L}(X_t) \longrightarrow \mu_{\beta_t}$" and of the weak asymptotic concentration of $(\mu_{\beta_t})_{t \geq 0}$ around M_ν would be the almost sure convergence of the algorithm towards the Fréchet mean:

$$\lim_{t \longrightarrow +\infty} P(X_t \in M_\nu) = 1.$$

We refer to [32] for further details. In particular, a strong requirement for this convergence can be considered through the relative entropy of n_t (the law of X_t) with respect to μ_{β_t}:

$$J_t := KL(n_t || \mu_{\beta_t}) = \int_{\Gamma_G} \log \left[\frac{n_t(x)}{\mu_{\beta_t}(x)} \right] dn_t(x) \qquad (10)$$

The function β_t will be chosen as a $\mathcal{C}^1$ function of $\mathbb{R}_+$, and since μ_β is a strictly positive measure over Γ_G, it implies that $t \longmapsto \mu_{\beta_t}(x)^{-1}$ is $\mathcal{C}^1(\mathbb{R}_+ \times \Gamma_G)$. Moreover, $(t, x) \longmapsto n_t(x)$ follows the backward Kolmogorov equation, which induces a $\mathcal{C}^1(\mathbb{R}_+ \times \Gamma_G)$ function. Since the semi-group is uniformly elliptic on the x-component, we have:

$$\forall t > 0 \qquad \forall x \in \Gamma_G \qquad n_t(x) > 0.$$

On the basis of these arguments, we can deduce:

PROPOSITION 1. *Assume that $t \longmapsto \beta_t$ is $\mathcal{C}^1$, then $(t, x) \longmapsto n_t(x)$ defines a positive $\mathcal{C}^1(\mathbb{R}_+ \times \Gamma_G)$ function and $t \longmapsto J_t$ is differentiable for any $t > 0$.*

If we define:

$$\alpha_t = \lambda(t + 1) \text{ and } \beta_t = b \log(t + 1),$$

with b a constant smaller than $c^\star(U_\nu)^{-1}$, where $c^\star(U_\nu)$ is the maximal depth of a well containing a local but not global minimum of U_ν, defined in (34), then our main result can be stated as follows:

THEOREM 3. *For any constant $\lambda > 0$ such that $\alpha_t = \lambda(t + 1)$ and $\beta_t = b \log(1 + t)$ with $b > \{c^\star(U_\nu)^{-1}\}^{-1}$, then:*

$$\lim J_t = 0 \qquad as \qquad t \longrightarrow +\infty.$$

This ensures that the process X_t will almost surely converge towards M_ν and, therefore, towards a global minimum of U_ν.

The idea of the proof is to obtain a differential inequality for J_t, which implies its convergence towards 0. It is well known that the Gibbs measure μ_{β_t} is the unique invariant distribution of the stochastic process that evolves only on the x component, whose Markov generator is given by:

$$\hat{\mathcal{L}}_{2,t}(f)(x) = \frac{1}{2}\Delta_x f - \beta_t \langle \nabla_x f, \nabla_x U_\nu \rangle. \tag{11}$$

Therefore, a natural step of the proof will be to control the difference between $\mathcal{L}_{2,t}$ and $\hat{\mathcal{L}}_{2,t}$ and to use this difference to study the evolution of m_t and n_t. It can be seen that $\hat{\mathcal{L}}_{2,t}$ may be written as an average action of the operator thanks to the linearity of the gradient operator:

$$\hat{\mathcal{L}}_{2,t}(f) = \frac{1}{2}\Delta_x f - \beta_t \mathbb{E}_{y\sim\nu} \langle \nabla_x f, \nabla_x U_y \rangle.$$

When $X_t = x$, we know that Y_t is distributed according to the distribution $m_t(y|x)$. Consequently, the average action of $\mathcal{L}_{2,t}$ on the x component is:

$$\tilde{\mathcal{L}}_{2,t}(f) = \frac{1}{2}\Delta_x f - \beta_t \mathbb{E}_{y\sim m_t(.|x)} \langle \nabla_x f, \nabla_x U_y \rangle = \frac{1}{2}\Delta_x f - \beta_t \int_V \langle \nabla_x f, \nabla_x U_y \rangle m_t(y|x)\mathrm{d}y, \tag{12}$$

whose expression may be close to that of $\hat{\mathcal{L}}_{2,t}$ if $m_t(.|x)$ is close to ν.

Thus, another important step is to choose appropriate values for α_t and β_t, *i.e.*, to find the balance between the increasing intensity of the Poisson process and the decreasing temperature schedule, in order to quantify the distance between ν and $m_t(.|x)$. The main core of the proof brings together these two aspects and is detailed in Section 5.

Another important step will be the use of functional inequalities (Poincaré and log-Sobolev inequalities) over Γ_G for the measure μ_β when $\beta = 0$ and $\beta \longrightarrow +\infty$. The proof of these technical results are given in Section 6.3.

COROLLARY 1. *Assume that $\beta_t = b\log(t+1)$ with $b < c^\star(U_\nu)^{-1}$ and $\alpha_t = \lambda(t+1)^\gamma$ with $\gamma \geq 1$, then for any neighborhood $\mathcal{N}$ of M_ν:*

$$\lim_{t\longrightarrow+\infty} \mathbb{P}[X_t \in \mathcal{N}] = 1.$$

Proof: The argument follows from Theorem 3. Consider any neighborhood $\mathcal{N}$ of M_ν. The continuity of U_ν shows that:

$$\exists \delta > 0 \quad \mathcal{N}^c \subset \{x \in \Gamma_G : U(x) > \min U + \delta\}.$$

Hence,

$$\begin{aligned}
\mathbb{P}[X_t \in \mathcal{N}^c] &\leq \mathbb{P}[U(X_t) > \min U + \delta] \\
&= \int_{\Gamma_G} \mathbf{1}_{U(x)>\min U+\delta}\mathrm{d}n_t(x) \\
&= \int_{\Gamma_G} \mathbf{1}_{U(x)>\min U+\delta}\mathrm{d}\mu_{\beta_t}(x) + \int_{\Gamma_G} \mathbf{1}_{U(x)>\min U+\delta}[n_t(x) - \mu_{\beta_t}(x)]\mathrm{d}x \\
&\leq \mu_{\beta_t}\{U > \min U + \delta\} + 2d_{TV}(n_t, \mu_{\beta_t}) \\
&\leq \mu_{\beta_t}\{U > \min U + \delta\} + \sqrt{2J_t},
\end{aligned}$$

where we used the variational formulation of the total variation distance and the Csiszár-Kullback inequality. As soon as $\lim_{t\longrightarrow+\infty} J_t = 0$, we can conclude the proof observing that $\mu_{\beta_t}\{U > \min U + \delta\} \longrightarrow 0$ as $\beta_t \longrightarrow +\infty$. $\qquad\square$

It can actually be proven (not shown here) that $lim_{t \longrightarrow +\infty}\mathbb{P}[X_t \in \mathcal{N}] = 1$ if and only if the constant b is chosen to be lower than $c^\star(U_\nu)^{-1}$. This means that this algorithm does not allow faster cooling schedules than the classical S.A. algorithm. Nevertheless, this is a positive result since this homogenized S.A. can be numerically computed quickly and easily on large graphs. Finally, we should consider this result to be theoretical. However, in practice, the simulation of this homogenized S.A. is performed during a finite horizon time and efficient implementations certainly deserve a specific theoretical study following the works of [17] and [46].

4. Numerical simulation results

We now present different practical aspects of our strategy. After giving numerical details about how to make it usable on large graphs, plus different insights into its calibration, we will present results obtained on social network subgraphs of *Facebook*[2] and a citation subgraph of *zbMATH*.

4.1. Algorithmic complexity

It is widely recognized that the algorithmic complexity of numerical strategies dealing with graphs can be an issue. The number of vertices and edges of real-life graphs can indeed be quite large. For instance, the *zbMATH* subgraph of Section 4.3.2 has 13000 nodes and approximately 48000 edges. In this context, it is worth justifying that the motion of $(X_t)_{t \geq 0}$ on the graph Γ_G across the iterations of Algorithm 3 is reasonably demanding in terms of computational resources. We recall that this motion is driven by a Brownian motion (lines 4 to 6 of Algorithm 3) and an attraction towards the vertex $Y_{N_t^\alpha}$ (line 9 of Algorithm 3) at random times $(T_k)_{k \geq 0}$. The number of vertices and undirected edges with non-null weights in Γ_G is also N and $|E|$, respectively. Note finally that $N - 1 \leq |E| \leq N(N-1)/2$ if Γ_G has a unique connected component.

Neighborhood structure A first computational issue that can arise when moving X_t is to find all possible neighbors of a specific vertex v. This is indeed performed every time X_t moves from one edge to another and directly depends on how Γ_G is encoded in the memory. Encoding Γ_G in a list of edges with non-null weights is common practice. In that case, the computer checks the vertex pairs linked by all edges to find those containing v, so the algorithmic cost is $2|E|$. A second classic strategy is to encode the graph in a connectivity matrix. In this case, the computer has to go through all the N indexes of the columns representing v to find the non-null weights. We instead sparsely encode the graph in a list of lists: the main list has a size N and each of its elements lists the neighbors of a specific vertex v. If the graph only contains edges with strictly positive weights, this strategy has a computational cost N, and in all of the other cases, the cost is $< N$. On average, the computational cost is $2|E|/N$ so that this strategy is particularly advantageous for sparse graphs, where $|E| << N(N-1)/2$, which are common for the targeted applications. For instance, $|E| = 4 * 10^3$ and $N(N-1)/2 = 1.24 * 10^5$ on the smallest *Facebook* subgraph of Section 4.3.1, and $|E| = 4.8 * 10^4$ and $N(N-1)/2 = 8.44 * 10^7$ on the *zbMATH* subgraph of Section 4.3.2.

Geodesic paths Another potential issue with our strategy is that it seeks at least one optimal path between the vertices X_t and $Y_{N_t^\alpha}$ at each jump time (line 9 of Algorithm 3). This may be efficiently done using a fast marching propagation algorithm (see, *e.g.*, [18] for details), where:

(i) the distance to X_t is iteratively propagated on the whole graph Γ_G until no more optimal distance to reach X_t is updated, and

(ii) considering the shortest path between X_t and $Y_{N_t^\alpha}$.

In this case, step (i) is particularly time-consuming and *cannot be reasonably performed at each step of Algorithm 3*. Fortunately, the algorithmic cost to compute the distance between all pairs of vertices is equivalent to that of computing step (i). We then compute these distances once and for

---

all at the beginning of the computations and store the result in a $N \times N$ matrix in the RAM of the computer [3]. We can therefore very quickly use these results at each iteration of the algorithm and, in particular, deduce the useful part of the geodesic paths involved to travel from X_t to $Y_{N_t^\alpha}$. The only limitation of this strategy is that the distance matrix can be memory-consuming. It can therefore be used on small to large graphs but not on huge graphs (typically when $N > 10^5$) on current desktops. An extension to deal with this scalability issue is a current subject of research, and we will therefore not describe applications of Algorithm 3 on huge graphs in this work.

4.2. Parameter tuning

Several parameters influence the behavior of the simulated process X_t. Some of them are directly introduced in the theoretical construction of the algorithm, *i.e.*, the intensity of the Poisson process or the temperature schedule. Other ones come from the practical implementation of the algorithm, *i.e.*, the maximal time up to which we generate X_t. The theoretical result given in Theorem 3 gives an upper bound for the probability of X_t to be not too distant from the set of global minima. This bound depends on β_t and α_t as well as on different characteristics of the graph such as its diameter and number of nodes. We then propose an empirical strategy for parameter tuning, which we will use later in Section 4.3.

For a graph G with N nodes and a diameter $\mathcal{D}_G$, we define:

$$T_{\max}^* = 100 + 0.1N \qquad \text{and} \qquad \beta_t^* = 2\log(t+1)/\mathcal{D}_G.$$

We then choose the intensity of the Poisson process so that it will generate a reasonably large amount of Y_n at the end of the algorithm, depending on the discrete probability distribution ν. More specifically, let S^* be the average number of jumping times between $T_{max}^* - 1$ and T_{max}^*. We then set $S^* = 1000$, which can be obtained by choosing:

$$\alpha_t^* = \lambda(2t+2) \qquad \text{with} \qquad \lambda = \frac{S^*}{2T_{\max}^* + 1}.$$

4.3. Results

This section presents results obtained on graphs of different sizes and using different parameters. We used the empirical method of Section 4.2 to define default parameters and altered them to quantify the sensitivity of our algorithm to parameter variations.

Since the process X_t lives on a quantum graph, its location at time t is between two vertices, on an edge of Γ_G. However, our primary interest is to study the properties of the initial discrete graph and therefore, the output of the algorithm will be the vertex considered as the graph barycenter. To achieve this, we associate a frequency to each vertex. For a vertex v, this frequency is the portion of time during which v was the closest vertex to the simulated process X_t. In our results, we consider frequencies computed over the last 10% of the iterations of the algorithm.

4.3.1. Facebook subgraphs

---

[3] In our experiments, we used the *all-shortest-paths* function of the Python library NetworkX.

Experimental protocol We first tested our algorithm on three subgraphs of Facebook from the Stanford Large Network Dataset Collection: **(FB500)** has 500 nodes and 4337 edges and contains two obvious clusters; **(FB2000)** has 2000 nodes and 37645 edges and fully contains (FB500); and **(FB4000)** has 4039 nodes and 88234 edges and fully contains (FB2000). For each of these subgraphs, we considered the probability measure ν as the uniform distribution over the graph's vertices and used a length of 1 for all edges. We also explicitly computed the barycenter of these graphs using an exhaustive search procedure. We are able to do that because we considered a simplified case where the distribution over the nodes is uniform. For example, this exhaustive search procedure required approximately 6 hours for the **(FB4000)** subgraph. Nevertheless, it allowed us to compare our strategy with ground-truth results.

We used the strategy of Section 4.2 to define default parameters adapted to each subgraph. We also tested different values for parameters β, S and $T_{\max}$ in order to quantify their influence. We repeated our algorithm 100 times for each parameter set to evaluate the algorithm stability.

In the tables representing quantitative results, *Error* represents the number of times, out of 100, that the algorithm converged to a node different from the ground-truth barycenter. It is a rough indicator of the ability of the algorithm to locate the barycenter of the graph, that could be replaced by a measure of the average distance between the last iterations of the algorithm and the ground-truth barycenter (not shown in this work). For each subgraph and parameter set, column *Av. time* contains the average times in seconds for the barycenter estimation, keeping in mind that the Dijkstra algorithm was performed once for all before the 100 estimations. This preliminary computation requires approximately 1, 30 and 80 seconds on a standard laptop.

Effect of β From a theoretical point of view, $(\beta_t)_{t \geq 0}$ should be chosen in relation to the constant $c^\star(U_\nu)$, which is unknown in practice. Therefore, the practical choice of $(\beta_t)_{t \geq 0}$ is a real issue to obtain a good behavior of the algorithm. Table 1 gives the results obtained on our algorithm with different values of β_t.

	FB500			FB2000			FB4000		
β	Error	Med. Freq.	Av. time	Error	Med. Freq.	Av. time	Error	Med. Freq.	Av. time
$\frac{1}{4}\beta^*$	15 %	0.6042	0.95 s	28 %	0.3519	10.12 s	27 %	0.3314	12.41 s
$\frac{1}{2}\beta^*$	2 %	0.4184	1.38 s	1 %	0.7268	4.25 s	5 %	0.6534	14.11 s
β^*	0 %	0.8008	1.48 s	0 %	0.9418	12.36 s	1 %	0.8913	13.55 s
$2\beta^*$	0 %	0.8321	11.31 s	0 %	0.9892	8.52 s	0 %	0.9647	16.79 s
$4\beta^*$	0 %	0.8233	13.19 s	0 %	0.9930	15.58 s	0 %	0.9824	31.43 s
$8\beta^*$	0 %	0.7717	2.36 s	0 %	0.9750	23.25 s	0 %	0.9445	44.46 s

Table 1 Experiments on the three Facebook subgraphs, while varying the values of $(\beta_t)_{t \geq 0}$.

We can observe that when $(\beta_t)_{t \geq 0}$ is too small, then the behavior of the algorithm is deteriorated, revealing the tendancy of the process $(X_t)_{0 \leq t \leq T^*_{\max}}$ to have an excessively slow convergence rate towards its local attractor in the graph. Roughly speaking, in such a situation, the process does not learn fast enough. When the value of β is chosen in the range $[\beta^*; 2\beta^*]$, we can observe a really good behavior of the algorithm: it almost always locates the good barycenter in a quite reasonable time of computation (less then 20 seconds for the largest graph). Finally, we can observe in the column, Med. Freq., that in most of the last iterations of the algorithm (more than 80%), the process evolves around its estimated barycenter, so that the decision to produce an estimator is quite easy when looking at an execution of the algorithm.

Effect of S and $T_{\max}$ Table 2 gives the results obtained with our algorithm while using different values of S and $T_{\max}$. As expected, we observe that increasing the ending time of simulation always improves the convergence rate (column Error in Table 2) of the algorithm towards the right node. The behavior of the algorithm is also improved by increasing the value of S, which quantifies the number of arrivals of nodes observed along the averaging procedure. Of course, the counterpart of increasing both S and T_{max} is an increasing cost of simulation (see column Av. time).

		FB500			FB2000			FB4000		
S	$T_{\max}$	Error	Med. Freq.	Av. time	Error	Med. Freq.	Av. time	Error	Med. Freq.	Av. time
$\frac{1}{2}S^*$	$T^*_{\max}$	2 %	0.7667	0.59 s	0 %	0.9344	3.22 s	0 %	0.8614	6.70 s
S^*	$2T^*_{\max}$	0 %	0.8049	2.79 s	0 %	0.9610	9.30 s	0 %	0.8970	24.42 s
S^*	$4T^*_{\max}$	0 %	0.8101	5.60 s	0 %	0.9677	22.21 s	0 %	0.9222	54.73 s
$2S^*$	$T^*_{\max}$	1 %	0.8345	2.26 s	0 %	0.9512	8.39 s	0 %	0.9062	20.95 s
$2S^*$	$2T^*_{\max}$	0 %	0.8361	4.57 s	0 %	0.9586	23.30 s	0 %	0.9121	23.30 s
$2S^*$	$4T^*_{\max}$	0 %	0.8423	11.16 s	0 %	0.9735	11.16 s	0 %	0.9366	96.84 s

Table 2 Experiments on the three Facebook subgraphs, while varying the values of S and $T_{\max}$.

As an illustration of the (small) complexity of the *Facebook* sub-graphs used for benchmarking our algorithm, we provide a representation of the FB500 graph in Figure 3. This representation has been obtained with the help of *Cytoscape* software and is not a result of our own algorithm. In Figure 3, the red node is the estimated barycenter, which is also the ground-truth barycenter located by a direct exhaustive computation. The blue nodes are the "second rank" nodes visited by our method.
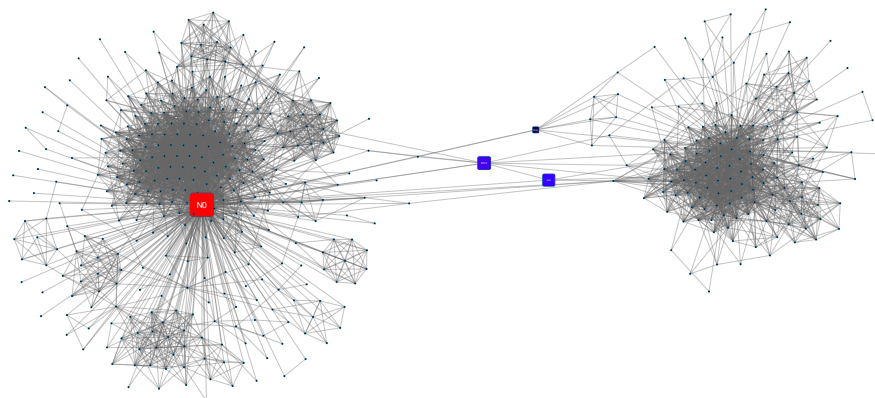

Figure 3 Region of interest presenting the results obtained on the FB500 graph.

4.3.2. zbMATH subgraph The zbMATH subgraph built from zbMATH[4] has been obtained by an iterative exploration of the co-authorship relationship in an alphabetical order. We thus naturally obtained a connected graph. This graph has been weighted by the complete number of citations obtained by each author, leading to the distribution probability $\nu \propto \sharp$(number of citations). Finally, all the edges in the graph are fixed to have a length of 1. This exploration was initialized on the entry of the first author's name (*i.e.*, S. Gadat) and we stopped the process when we obtained 13000 nodes (authors) on the graph. This stopping criterion in the exploration of the zbMATH database corresponds to a technical limitation of 40 GB memory required by the distance matrix

---

[4] https://zbmath.org/authors/

obtained with the Dijsktra algorithm. In particular, this limitation and the starting point of the exploration induce an important bias in the community of authors used to build the subgraph from the zbMATH dataset: the researchers obtained in the subgraph are generally French and applied mathematicians. The graph is more or less focused on the following research themes: probability, statistics and P.D.E. Consequently, the results provided below should be understood as an illustration of our algorithm and not as a bibliometric study!

Our experiments rely on the same choice of parameters $(\beta_t^*)_{t \geq 0}, S^*, T_{max}^*$ and α_t^* as above, indicated at the beginning of Section 4.2. Again, our algorithm produces the same outputs but of course, in this situation, we do not know the ground-truth barycenter of the zbMath subgraph. In Figure 4, we present a representation of the subgraph obtained with *Cytoscape* software, and a zoom on a region of interest (ROI for short) in Figure 4.

Again, the main nodes visited by our algorithm are represented with a red square and the size of the used square is larger when the node is frequently visited. According to the results obtained on the *Facebook* subgraphs, we then assume that the larger red square is the barycenter of the zbMath subgraph.
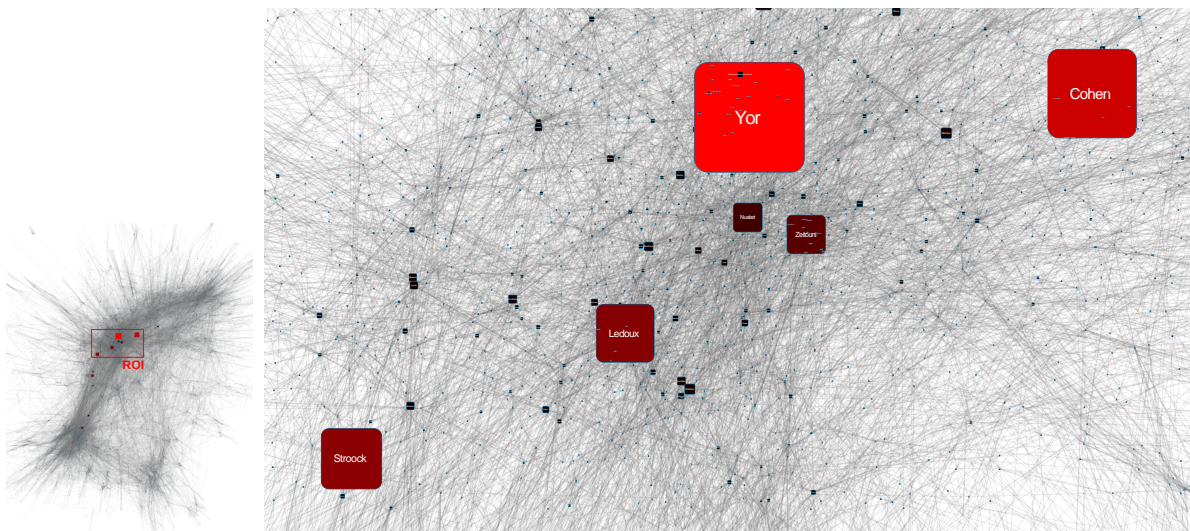

Figure 4   Left: General overview of the zbMath subgraph extracted for our experiments, containing approximately 13000 nodes. Right: Region of interest presenting the main results obtained on the zbMATH subgraph.

It is also possible to assert the robustness of our method with respect to several Monte Carlo runs of our algorithm. We have produced some boxplots for each of the main authors located in the subgraph, according to the occupation measure of the process over the last 10% of the iterations with 10 Monte-Carlo replications. These "violin" plots are represented in Figure 5. Each execution of the algorithm requires approximately 3 hours of computations. The algorithm seems to produce reliable conclusions concerning the top nodes visited all along the ending iterations. Nevertheless, it appears to be necessary to extend our investigations in order to obtain a scalable method for handling larger graphs.

5.   Proof of the main result (Theorem 3)

We establish a differential inequality that will imply the convergence of J_t. The computations actually lead us to a system of two differential inequalities. We therefore introduce another quantity
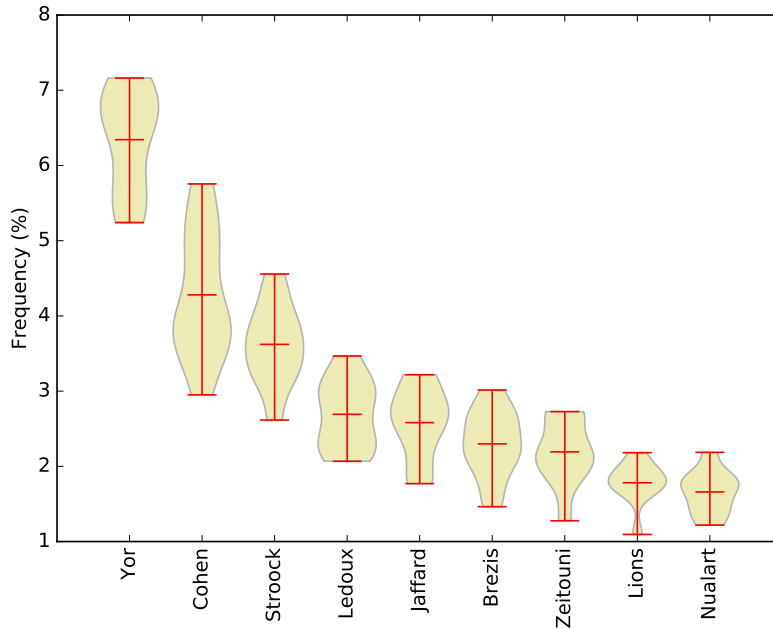
Figure 5   **"Violin" plot of the occupation measure of the algorithm on the zbMATH subgraph for 10 MC repli-
cations. The average frequency is located in the middle of the "violin" plot, although the minimal and
maximal values are shown in the extremity of the representation.**

I_t that measures the average closeness (w.r.t. x) of the conditional law of y given x at time t to ν,
defined as:

$$I_t := \int KL(m_t(.|x)||\nu)\mathrm{d}n_t(x) = \int \left[\int \log \frac{m_t(y|x)}{\nu(y)} m_t(y|x)dy \right] .\mathrm{d}n_t(x) \tag{13}$$

The next proposition links the evolution of J_t (in terms of an upper bound of $\partial_t J_t$) with the
spectral gap of μ_{β_t} over Γ_G, the diameter of the graph $\mathcal{D}_G$ and the divergence I_t.

5.1.   Study of $\partial_t J_t$

PROPOSITION 2. *Let $c^\star(U_\nu)$ be the term defined in Equation* (34) *and C_{Γ_G} be the constant given
in Proposition* 5. *We then have:*

$$\partial_t J_t \leq \mathcal{D}_G^2 \beta_t' + 8\mathcal{D}_G^2 \beta_t^2 I_t - \frac{e^{-c^\star(U)\beta_t}}{C_{\Gamma_G}(1+\beta_t)} J_t.$$

Proof: We compute the derivative of J_t and separately study each of the three terms:

$$\partial_t J_t = \underbrace{\int \partial_t \{\log(n_t(x))\}\mathrm{d}n_t(x)}_{J_{1,t}} - \underbrace{\int \partial_t \{\log(\mu_{\beta_t}(x))\}\mathrm{d}n_t(x)}_{J_{2,t}} + \underbrace{\int \log \left[\frac{n_t(x)}{\mu_{\beta_t}} \right] \partial_t n_t(x)\mathrm{d}x}_{J_{3,t}} \tag{14}$$

Study of $J_{1,t}$: This term is easy to deal with:

$$J_{1,t} = \int_{\Gamma_G} \partial_t \{\log(n_t(x))\}n_t(x)$$

$$= \int_{\Gamma_G} \frac{\partial_t\{n_t(x)\}}{n_t(x)} n_t(x) = \int_{\Gamma_G} \partial_t\{n_t(x)\}dx = \partial_t\left\{\int_{\Gamma_G} n_t(x)dx\right\} = \partial_t\{1\} = 0. \qquad (15)$$

Study of $J_{2,t}$: Using the definition of μ_{β_t} given in Equation (9), we obtain for the second term:

$$J_{2,t} = -\int_{\Gamma_G} \partial_t\{\log(\mu_{\beta_t}(x))\}dn_t(x) = -\int_{\Gamma_G} \partial_t\{-\beta_t U_\nu(x) - \log Z_{\beta_t}\}dn_t(x)$$
$$= \beta_t' \int_{\Gamma_G} U_\nu(x)dn_t(x) + \frac{\partial_t\{Z_{\beta_t}\}}{Z_{\beta_t}}.$$

According to the definition of Z_{β_t}, we have:

$$\frac{\partial_t\{Z_{\beta_t}\}}{Z_{\beta_t}} = Z_{\beta_t}^{-1}\partial_t\left\{\int_{\Gamma_G} e^{-\beta_t U_\nu(x)}dx\right\} = -\beta_t' \int_{\Gamma_G} U_\nu(x)\frac{e^{-\beta_t U_\nu(x)}}{Z_{\beta_t}}dx = -\partial_t\{\beta_t\}\int_{\Gamma_G} U_\nu(x)\mu_{\beta_t}(x)dx.$$

Hence, we obtain

$$J_{2,t} = \beta_t' \int_{\Gamma_G} U_\nu(x)[n_t(x) - \mu_{\beta_t}(x)]dx$$

The graph has a finite diameter $\mathcal{D}_G$. We therefore have: $\int_{\Gamma_G} U_\nu(x)dn_t(x) \leq \mathcal{D}_G^2$. The same inequality holds using the measure μ_{β_t} so that:

$$|J_{2,t}| \leq \mathcal{D}_G^2 \beta_t'. \qquad (16)$$

Study of $J_{3,t}$: The last term $J_{3,t}$ involves the backward Kolmogorov equation. First, since n_t is the marginal law of X_t, we have: $n_t(x) = \int m_t(x,y)dy$.

Using the backward Kolmogorov equation for the Markov process $(X_t, Y_t)_{t\geq 0}$ and the Fubini theorem, we have, for any smooth enough function $f_t : x \in \Gamma_G \longmapsto \mathbb{R}$:

$$\int_{\Gamma_G} f_t(x)\partial_t\{n_t(x)\}dx = \int_{\Gamma_G} f_t(x)\partial_t\left\{\int_V m_t(x,y)dy\right\}dx = \int_{\Gamma_G}\int_V f_t(x)\partial_t\{m_t(x,y)\}dxdy$$
$$= \int_{\Gamma_G}\int_V \mathcal{L}_t(f_t)(x)m_t(x,y)dxdy = \int_{\Gamma_G}\int_V [\mathcal{L}_{1,t} + \mathcal{L}_{2,t}](f_t)(x)m_t(x,y)dxdy,$$

where $\mathcal{L}_{1,t}$ and $\mathcal{L}_{2,t}$ are defined in Equations (6) and (7). Since the function f_t is independent of y, we have $\mathcal{L}_{1,t}(f_t) = 0$. For the part corresponding to $\mathcal{L}_{2,t}$, we have:

$$\int_{\Gamma_G}\int_V \mathcal{L}_{2,t}(f_t)(x)m_t(x,y)dxdy$$
$$= \int_{\Gamma_G}\int_V \left[\frac{1}{2}\Delta_x f_t(x) - \beta_t \nabla_x U_y(x)\nabla_x f_t(x)\right] m_t(x,y)dxdy$$
$$= \int_{\Gamma_G} \frac{1}{2}\Delta_x f_t(x)n_t(x)dx - \beta_t \int_{\Gamma_G}\int_V \nabla_x U_y(x)\nabla_x f_t(x)n_t(x)m_t(y|x)dxdy.$$

because n_t is the marginal distribution of X_t and $m_t(y|x) \times n_t(x) = m_t(x,y)$.

Thus, for any smooth enough function $f_t(x)$, using the operator introduced in (12), we have:

$$\int_{\Gamma_G} f_t(x)\partial_t n_t(x)dx = \int_{\Gamma_G} \tilde{\mathcal{L}}_{2,t}(f_t)(x)n_t(x)dx.$$

Replacing $f_t(x)$ by $\log \dfrac{n_t(x)}{\mu_{\beta_t}(x)}$, we obtain:

$$J_{3,t} = \int_{\Gamma_G} \log \frac{n_t(x)}{\mu_{\beta_t}} \partial_t n_t(x) = \int_{\Gamma_G} \tilde{\mathcal{L}}_{2,t} \left[\log \frac{n_t(x)}{\mu_{\beta_t}(x)} \right] n_t(\mathrm{d}x).$$

Since μ_{β_t} is the invariant distribution of $\hat{\mathcal{L}}_{2,t}$ (see Equation (11)), it is natural to insert $\hat{\mathcal{L}}_{2,t}$:

$$
\begin{aligned}
J_{3,t} &= \int_{\Gamma_G} \tilde{\mathcal{L}}_{2,t} \left[\log \frac{n_t(x)}{\mu_{\beta_t}(x)} \right] n_t(\mathrm{d}x) \\
&= \int_{\Gamma_G} \hat{\mathcal{L}}_{2,t} \left[\log \frac{n_t(x)}{\mu_{\beta_t}(x)} \right] n_t(\mathrm{d}x) - \underbrace{\int_{\Gamma_G} \left(\hat{\mathcal{L}}_{2,t} - \tilde{\mathcal{L}}_{2,t} \right) \left[\log \frac{n_t(x)}{\mu_{\beta_t}(x)} \right] n_t(\mathrm{d}x)}_{:=\kappa_t}.
\end{aligned}
\tag{17}
$$

Since $\hat{\mathcal{L}}_{2,t}$ is a diffusion operator and μ_{β_t} is its invariant measure, it is well known (see, *e.g.*, [7]) that the action of $\hat{\mathcal{L}}_{2,t}$ on the entropy is closely linked to the Dirichlet form in the following way:

$$\int_{\Gamma_G} \hat{\mathcal{L}}_{2,t} \left[\log \frac{n_t(x)}{\mu_{\beta_t}(x)} \right] n_t(\mathrm{d}x) = -2 \int_{\Gamma_G} \left(\nabla_x \left\{ \sqrt{\frac{n_t(x)}{\mu_{\beta_t}(x)}} \right\} \right)^2 \mu_{\beta_t}(\mathrm{d}x),
\tag{18}$$

and therefore translates a mean reversion towards μ_{β_t} in the first term of (17).

We now study the size of the difference between $\hat{\mathcal{L}}_{2,t}$ and $\tilde{\mathcal{L}}_{2,t}$ and introduce the "approximation" term of $\nabla_x U_\nu(x)$ at time t:

$$\tilde{R}_t(x) = \int_V \nabla_x (U_y)(x) m_t(y|x) \mathrm{d}y.$$

The relationship $\nabla_x \log(f) = 2 \frac{\nabla_x \sqrt{f}}{\sqrt{f}}$, the Cauchy-Schwarz inequality and $2ab \leq a^2 + b^2$ yield:

$$
\begin{aligned}
|\kappa_t| &= \beta_t \left| \int_{\Gamma_G} \left(\tilde{R}_t(x) - \nabla_x U_\nu(x) \right) \nabla_x \left\{ \log \frac{n_t(x)}{\mu_{\beta_t}(x)} \right\} n_t(\mathrm{d}x) \right| \\
&= 2\beta_t \left| \int_{\Gamma_G} \left(\tilde{R}_t(x) - \nabla_x U_\nu(x) \right) \nabla_x \left\{ \sqrt{\frac{n_t(x)}{\mu_{\beta_t}(x)}} \right\} \sqrt{\frac{\mu_{\beta_t}(x)}{n_t(x)}} n_t(\mathrm{d}x) \right| \\
&\leq 2\beta_t \sqrt{\int_{\Gamma_G} \left(\tilde{R}_t(x) - \nabla_x U_\nu(x) \right)^2 n_t(\mathrm{d}x)} \cdot \sqrt{\int_{\Gamma_G} \nabla_x \left(\sqrt{\frac{n_t(x)}{\mu_{\beta_t}(x)}} \right)^2 \mu_{\beta_t}(x) \mathrm{d}x} \\
&\leq \beta_t^2 \int \left(\tilde{R}_t(x) - \nabla_x U_\nu(x) \right)^2 n_t(x) \mathrm{d}x + \int_{\Gamma_G} \left(\nabla_x \left\{ \sqrt{\frac{n_t(x)}{\mu_{\beta_t}(x)}} \right\} \right)^2 \mu_{\beta_t}(\mathrm{d}x)
\end{aligned}
$$

If d_{TV} denotes the total variation distance, the first term of the right hand side leads to:

$$
\begin{aligned}
\left| \tilde{R}_t(x) - \nabla_x U_\nu(x) \right| &= \left| \int_V \nabla_x d^2(x,y) [m_t(y|x) - \nu(y)] \mathrm{d}y \right| \\
&\leq \| \nabla_x d^2(x,y) \|_\infty \left| \int_V [m_t(y|x) - \nu(y)] \mathrm{d}y \right| \\
&\leq 2 \| \nabla_x d^2(x,y) \|_\infty d_{TV}(m_t(.|x), \nu) \\
&\leq \sqrt{2} \| \nabla_x d^2(x,y) \|_\infty \sqrt{\int_V \log \frac{m_t(y|x)}{\nu(y)} m_t(y|x) \mathrm{d}y}
\end{aligned}
$$

where the last line comes from the Csiszár-Kullback inequality. Since $d^2(\cdot, y)$ is differentiable a.e. and its derivative is bounded for all $y \in V$ by $2\mathcal{D}_G$, we can use I_t defined in Equation (13) to obtain:

$$\int_{\Gamma_G} \left(\tilde{R}_t - \hat{R}_t \right)^2 (x) n_t(x) \mathrm{d}x \leq 8 \mathcal{D}_G^2 I_t$$

Consequently, we obtain:

$$|\kappa_t| \leq 8 \mathcal{D}_G^2 \beta_t^2 I_t + \int_{\Gamma_G} \left(\nabla_x \left\{ \sqrt{\frac{n_t(x)}{\mu_{\beta_t}(x)}} \right\} \right)^2 \mu_{\beta_t}(\mathrm{d}x) \tag{19}$$

Taking the inequalities (18) and (19), we now obtain in (17):

$$J_{3,t} \leq 8 \mathcal{D}_G^2 \beta_t^2 I_t - \int_{\Gamma_G} \left(\nabla_x \left\{ \sqrt{\frac{n_t(x)}{\mu_{\beta_t}(x)}} \right\} \right)^2 \mu_{\beta_t}(\mathrm{d}x)$$

We denote $f_t = \sqrt{\frac{n_t}{\mu_{\beta_t}}}$. Since $\|f\|_{2,\mu_{\beta_t}}^2 = 1$, one can easily see that $\mathbf{Ent}_{\mu_{\beta_t}}(f_t^2) = J_t$. Now, the Logarithmic Sobolev inequality on the (quantum) graph Γ_G for the measure μ_{β_t} stated in Proposition 5 shows that:

$$\int_{\Gamma_G} \left(\nabla_x \left\{ \sqrt{\frac{n_t(x)}{\mu_{\beta_t}(x)}} \right\} \right)^2 \mu_{\beta_t}(\mathrm{d}x) \geq \frac{e^{-c^\star(U_\nu)\beta_t}}{C_{\Gamma_G}(1 + \beta_t)} J_t,$$

where $c^\star(U_\nu)$ is defined in Equation (34) and is related to the maximal depth of a well containing a local but not global minimum. We thus obtain:

$$J_{3,t} = \int \tilde{\mathcal{L}}_{2,t} \left[\log \frac{n_t(x)}{\mu_{\beta_t}(x)} \right] n_t(\mathrm{d}x) \leq 8 \mathcal{D}_G^2 \beta_t^2 I_t - \frac{e^{-c^\star(U_\nu)\beta_t}}{C_{\Gamma_G}(1 + \beta_t)} J_t.$$

The proof is concluded by regrouping the three terms. $\qquad\square$

5.2. Study of $\partial_t I_t$

PROPOSITION 3. *Assume that $\mathcal{D}_G \geq 1$ and β_t is an increasing inverse temperature with $\beta_t \geq 1$, then:*

$$\partial_t I_t \leq -\alpha_t I_t - \partial_t J_t + \mathcal{D}_G^2 [\beta_t' + 6\beta_t^2]. \tag{20}$$

Proof: We compute the derivative of I_t. Observing that $m_t(x, y) = n_t(x) m_t(y|x)$, we have:

$$\partial_t I_t = \partial_t \left\{ \int_{\Gamma_G} n_t(x) \left(\int_V \log \frac{m_t(y|x)}{\nu(y)} m_t(y|x) \mathrm{d}y \right) \mathrm{d}x \right\}$$

$$= \partial_t \left\{ \int_{\Gamma_G} \int_V \log \frac{m_t(y|x)}{\nu(y)} m_t(x, y) \mathrm{d}x \mathrm{d}y \right\}$$

$$= \underbrace{\int_{\Gamma_G} \int_V \partial_t \{\log m_t(y|x)\} m_t(x, y) \mathrm{d}x \mathrm{d}y}_{:=I_{1,t}} - \underbrace{\int_{\Gamma_G} \int_V \partial_t \{\log \nu(y)\} m_t(x, y) \mathrm{d}x \mathrm{d}y}_{:=I_{2,t}} + \underbrace{\int \log \frac{m_t(y|x)}{\nu(y)} \partial_t m_t(x, y) \mathrm{d}x \mathrm{d}y}_{:=I_{3,t}}$$

The computation of the first two terms is straightforward. For the first one we have:

$$I_{1,t} = \int_{\Gamma_G} \int_V \partial_t \{\log m_t(y|x)\} m_t(x, y) = \int_{\Gamma_G} \int_V \frac{\partial_t m_t(y|x)}{m_t(y|x)} m_t(x, y) \mathrm{d}x \mathrm{d}y$$

$$= \int_{\Gamma_G} \int_V \partial_t m_t(y|x) n_t(x) \mathrm{d}x \mathrm{d}y = \int_{\Gamma_G} n_t(x) \left(\int_V \partial_t m_t(y|x) \mathrm{d}y \right) \mathrm{d}x$$

$$= \int_{\Gamma_G} n_t(x) \partial_t \left\{ \underbrace{\int_V m_t(y|x) \mathrm{d}y}_{:=1} \right\} \, \mathrm{d}x = 0$$

The computation of $I_{2,t}$ is easy since ν does not depend on t, implying that $I_{2,t} = 0$.

For the third term, we use the backward Kolmogorov equation and obtain:

$$I_{3,t} = \int_{\Gamma_G} \int_V \log \frac{m_t(y|x)}{\nu(y)} \partial_t m_t(x,y) = \int_{\Gamma_G} \int_V \mathcal{L}_t \left(\log \frac{m_t(y|x)}{\nu(y)} \right) m_t(x,y) \mathrm{d}x \, \mathrm{d}y$$

$$= \underbrace{\int_{\Gamma_G} \int_V \mathcal{L}_{1,t} \left(\log \frac{m_t(y|x)}{\nu(y)} \right) m_t(x,y) \mathrm{d}x \, \mathrm{d}y}_{:=I^1_{3,t}} + \underbrace{\int_{\Gamma_G} \int_V \mathcal{L}_{2,t} \left(\log \frac{m_t(y|x)}{\nu(y)} \right) m_t(x,y) \mathrm{d}x \, \mathrm{d}y}_{:=I^2_{3,t}}$$

The jump part $\mathcal{L}_{1,t}$ exhibits a mean reversion on the entropy on the conditional law: applying the Jensen inequality for the logarithmic function and the measure ν, we obtain:

$$I^1_{3,t} = \alpha_t \int_{\Gamma_G} \int_V \left[\int_V \left(\log \frac{m_t(y'|x)}{\nu(y')} - \log \frac{m_t(y|x)}{\nu(y)} \right) \nu(y') \mathrm{d}y' \right] m_t(x,y) \mathrm{d}x \mathrm{d}y$$

$$= \alpha_t \int_{\Gamma_G} \int_V \left[\int_V \left(\log \frac{m_t(y'|x)}{\nu(y')} \right) \nu(y') \mathrm{d}y' \right] m_t(x,y) \mathrm{d}x \mathrm{d}y - \alpha_t I_t$$

$$\leq \alpha_t \int_{\Gamma_G} \int_V \log \left(\int_V \frac{m_t(y'|x)}{\nu(y')} \nu(y') \mathrm{d}y' \right) m_t(x,y) \mathrm{d}x \mathrm{d}y - \alpha_t I_t$$

$$\leq \alpha_t \int_{\Gamma_G} \int_V \log \left(\int_V m_t(y'|x) \mathrm{d}y' \right) m_t(x,y) \mathrm{d}x \mathrm{d}y - \alpha_t I_t$$

$$\leq \alpha_t \int_{\Gamma_G} \int_V \log 1 \cdot m_t(x,y) \mathrm{d}x \mathrm{d}y - \alpha_t I_t \leq -\alpha_t I_t \tag{21}$$

If we consider the action of $\mathcal{L}_{2,t}$ on the entropy of the conditional law, using $m_t(x,y) = m_t(y|x) n_t(x)$ yields:

$$I^2_{3,t} = \int_{\Gamma_G} \int_V \mathcal{L}_{2,t} \left(\log \frac{m_t(y|x)}{\nu(y)} \right) m_t(x,y) \mathrm{d}x \, \mathrm{d}y$$

$$= \int_{\Gamma_G} \int_V \mathcal{L}_{2,t} \left(\log \frac{m_t(x,y)}{\nu(y) \cdot n_t(x)} \right) m_t(x,y) \mathrm{d}x \, \mathrm{d}y$$

$$= \int_{\Gamma_G} \int_V \mathcal{L}_{2,t} \log \left(m_t(x,y) \right) m_t(x,y) \mathrm{d}x \mathrm{d}y - \int_{\Gamma_G} \int_V \mathcal{L}_{2,t} \log \left(n_t(x,y) \right) m_t(x,y) \mathrm{d}x \mathrm{d}y, \tag{22}$$

because $\mathcal{L}_{2,t}(\nu)(y) = 0$ ($\mathcal{L}_{2,t}$ only involves the x component). We now study the first term of (22):

$$\int_{\Gamma_G} \int_V \mathcal{L}_{2,t} \left(\log m_t(x,y) \right) \mathrm{d}m_t(x,y) = \frac{1}{2} \int_{\Gamma_G} \int_V \Delta_x \left[\log m_t(x,y) \right] \mathrm{d}m_t(x,y)$$

$$- \beta_t \int_{\Gamma_G} \int_V \langle \nabla_x U_y, \nabla_x \log m_t(x,y) \rangle \mathrm{d}m_t(x,y)$$

$$= \frac{1}{2} \int_{\Gamma_G} \int_V \partial^2_x \left\{ \log m_t(x,y) \right\} \mathrm{d}m_t(x,y)$$

$$- \beta_t \int \partial_x U_y \partial_x \log m_t(x,y) \mathrm{d}m_t(x,y)$$

The first term of the right-hand side deserves special attention:

$$\begin{aligned}
\int_{\Gamma_G}\int_V \partial_x^2\left\{\log m_t(x,y)\right\}\mathrm{d}m_t(x,y) &= \int_{\Gamma_G}\int_V \partial_x\left\{\frac{\partial_x\{m_t(x,y)\}}{m_t(x,y)}\right\}\mathrm{d}m_t(x,y)\\
&= \int_{\Gamma_G}\int_V \frac{\partial_x^2\{m_t(x,y)\}}{m_t(x,y)}\mathrm{d}m_t(x,y) - \int_{\Gamma_G}\int_V \frac{[\partial_x\{m_t(x,y)\}]^2}{m_t(x,y)}\mathrm{d}x\mathrm{d}y\\
&= \int_{\Gamma_G}\int_V \partial_x^2\{m_t(x,y)\}\mathrm{d}x\mathrm{d}y - 4\int_{\Gamma_G}\int_V \left(\partial_x\sqrt{m_t(x,y)}\right)^2\mathrm{d}x\mathrm{d}y\\
&= \sum_{e\in E}\int_V [\partial_x m_t(e(L_e),y)] - \partial_x m_t(e(0),y)\mathrm{d}y\\
&\quad - 4\int_{\Gamma_G}\int_V \left(\partial_x\sqrt{m_t(x,y)}\right)^2\mathrm{d}x\mathrm{d}y,
\end{aligned}$$

where $\{e(s), 0\le s\le L_e\}$ is the parametrization of the edge $e\in E$ introduced in Section 3.1. The gluing conditions (5) yield:

$$\sum_{e\in E}\int_V \partial_x m_t(e(L_e),y) - \partial_x m_t(e(0),y) = \int_V \sum_{v\in V}\left[\sum_{e\sim v} -d_e m_t(v,y)\right] = 0.$$

As a consequence, we obtain:

$$\begin{aligned}
\int_{\Gamma_G}\int_V \mathcal{L}_{2,t}\left(\log m_t(x,y)\right)\mathrm{d}m_t(x,y) &= -2\int_{\Gamma_G}\int_V \left(\partial_x\sqrt{m_t(x,y)}\right)^2\mathrm{d}x\mathrm{d}y\\
&\quad - \beta_t\int_{\Gamma_G}\int_V \partial_x U_y \partial_x\log m_t(x,y)\mathrm{d}m_t(x,y)
\end{aligned} \tag{23}$$

The Cauchy-Schwarz inequality applied on the second term leads to:

$$\begin{aligned}
\left|\beta_t\int_{\Gamma_G}\int_V \partial_x U_y \partial_x\log m_t(x,y)\mathrm{d}m_t(x,y)\right| &\le \beta_t\|\partial_x d^2(.,.)\|_{2,m_t}\sqrt{\int_{\Gamma_G}\int_V (\partial_x\log m_t(x,y))^2\,\mathrm{d}m_t(x,y)}\\
&\le \beta_t\|\partial_x d^2(x,y)\|_\infty\sqrt{\int_{\Gamma_G}\int_V (\partial_x\log m_t(x,y))^2\,m_t(x,y)\mathrm{d}x\mathrm{d}y}\\
&\le 4\beta_t\mathcal{D}_G\sqrt{\int_{\Gamma_G}\int_V \left(\partial_x\sqrt{m_t(x,y)}\right)^2\mathrm{d}x\mathrm{d}y}\\
&\le 2\mathcal{D}_G^2\beta_t^2 + 2\int_{\Gamma_G}\int_V \left(\partial_x\sqrt{m_t(x,y)}\right)^2\mathrm{d}x\mathrm{d}y
\end{aligned}$$

Inserting this in Equation (23) leads to:

$$\int_{\Gamma_G}\int_V \mathcal{L}_{2,t}\log m_t(x,y)\mathrm{d}m_t(x,y) \le 2\mathcal{D}_G^2\beta_t^2. \tag{24}$$

We study the second term of (22) and use the proof of Proposition 2: the decomposition (14) with Equations (15) and (16) yield:

$$\partial_t J_t \le \beta_t'\mathcal{D}_G^2 + \int \mathcal{L}_{2,t}\left(\log\frac{n_t(x)}{\mu_{\beta_t}(x)}\right)m_t(\mathrm{d}x,\mathrm{d}y).$$

This implies:

$$-\int_{\Gamma_G}\int_V \mathcal{L}_{2,t}\left(\log n_t(x)\right)m_t(\mathrm{d}x,\mathrm{d}y) \le -\partial_t J_t + \beta_t'\mathcal{D}_G^2 - \int \mathcal{L}_{2,t}\left(\log\mu_{\beta_t}(x)\right)m_t(\mathrm{d}x,\mathrm{d}y). \tag{25}$$

Using the definition of μ_{β_t}, we obtain:

$$
\begin{aligned}
-\int_{\Gamma_G}\int_V \mathcal{L}_{2,t}\log\frac{e^{-\beta_t U_\nu(x)}}{Z_{\beta_t}} &= \int_{\Gamma_G}\int_V \left[\mathcal{L}_{2,t}\left(\beta_t U_\nu(x)\right)+\mathcal{L}_{2,t}\left(\log Z_{\beta_t}\right)\right]m_t(\mathrm{d}x,\mathrm{d}y)\\
&= \beta_t\int_{\Gamma_G}\int_V \mathcal{L}_{2,t}\left(U_\nu(x)\right)m_t(\mathrm{d}x,\mathrm{d}y)\\
&= \beta_t\int_{\Gamma_G}\int_V \frac{1}{2}\Delta_x U_\nu(x)-\beta_t\nabla_x U_y(x)\nabla_x U_\nu(x)m_t(\mathrm{d}x,\mathrm{d}y)\\
&\le \frac{\beta_t}{2}+4\beta_t^2\mathcal{D}_G^2
\end{aligned}
$$

This inequality used in (25) yields:

$$
-\int \mathcal{L}_{2,t}\left(\log n_t(x)\right)m_t(x,y) = -\partial_t J_t + \beta_t'\mathcal{D}_G^2 + \frac{\beta_t}{2}+4\beta_t^2\mathcal{D}_G^2 \tag{26}
$$

We now use (24) and (26) in (22) and our assumptions $\beta_t\ge 1$ and $\mathcal{D}_G\ge 1$ to obtain:

$$
I_{3,t}^2 \le -\partial_t J_t + \mathcal{D}_G^2[\beta_t'+6\beta_t^2] \tag{27}
$$

Combining (21) and (27) leads to the desired inequality given by Equation (20). $\qquad\square$

5.3. Convergence of the entropy

The use of Propositions 2 and 3 makes it possible to obtain a system of coupled differential inequalities.

Proof of Theorem 3: If we denote $a=\mathcal{D}_G^2$, we can write:

$$
\begin{cases}
J_t' \le -\dfrac{e^{-c^\star(U_\nu)\beta_t}}{C_{\Gamma_G}(1+\beta_t)}J_t + a\beta_t' + 8a\beta_t^2 I_t\\[2mm]
I_t' \le -\alpha_t I_t - J_t' + a(\beta_t'+6\beta_t^2)
\end{cases}
$$

We introduce an auxiliary function $K_t = J_t + k_t I_t$, where k_t is a smooth positive decreasing function for t large enough, so that $\lim_{t\to\infty} k_t = 0$. We use the system above to deduce that:

$$
\begin{aligned}
K_t' &= J_t' + k_t' I_t + k_t I_t'\\
&\le J_t' + k_t I_t'\\
&\le J_t' + k_t(-\alpha_t I_t - J_t') + ak_t(\beta_t'+6\beta_t^2)\\
&\le (1-k_t)J_t' - k_t\alpha_t I_t + ak_t(\beta_t'+6\beta_t^2).
\end{aligned}
$$

In the first inequality, we use the fact that $k_t'\le 0$ and I_t is positive. The second inequality is given by (20). The upper bound of Proposition 2 now leads to:

$$
K_t' \le -\epsilon_t(1-k_t)J_t - k_t\alpha_t I_t + 8a(1-k_t)\beta_t^2 I_t + a\beta_t' + 6ak_t\beta_t^2,
$$

where we denoted $\epsilon_t = \frac{e^{-c^\star(U_\nu)\beta_t}}{C_{\Gamma_G}(1+\beta_t)}$. We choose the function k_t to obtain a mean reversion on K_t:

$$
k_t := \frac{8a\beta_t^2}{\alpha_t + 8a\beta_t^2 - \epsilon_t/2}. \tag{28}
$$

Note that this function is decreasing for sufficiently large t as soon as $\beta_t = o(\alpha_t)$, which is the case according to the choices described in Theorem 3. Moreover, a straightforward consequence is $\lim_{t\to+\infty} k_t = 0$. This ensures that a positive T_0 exists such that:

$$
\forall t\ge T_0 \qquad 0\le k_t\le \frac{1}{2}.
$$

Consequently, we deduce that:

$$\forall t \geq T_0 \qquad K_t' \leq -\epsilon_t(1-k_t)J_t - k_t\frac{\epsilon_t}{2}I_t + a\beta_t' + 6a\beta_t^2 k_t$$
$$\leq -\frac{\epsilon_t}{2}J_t - k_t\frac{\epsilon_t}{2}I_t + a\beta_t' + 6a\beta_t^2 k_t \leq -\frac{\epsilon_t}{2}K_t + \underbrace{a\beta_t' + 6a\beta_t^2 k_t}_{:=\eta_t}$$

The next bound is an easy consequence of the Gronwall Lemma:

$$\forall t \geq T_0 \qquad K_t \leq K_{T_0}e^{-\int_{T_0}^t \frac{\epsilon_s}{2}\mathrm{d}s} + \int_{T_0}^T \eta_s e^{-\int_s^t \frac{\epsilon_u}{2}\mathrm{d}u}\mathrm{d}s,$$

which in turn implies that $K_t \longrightarrow 0$ as $t \longrightarrow +\infty$ as soon as $\eta_t = o_{t\sim+\infty}(\epsilon_t)$ with $\int_0^\infty \epsilon_u du = +\infty$.

We are now looking for a suitable choice for α_t and β_t. Let us assume that $\alpha_t \sim \lambda t^\gamma$ for any $\gamma > 0$. This choice leads to:

$$k_t \sim \frac{8ab^2 \log(t+1)^2}{\lambda t^\gamma},$$

so that:

$$\eta_t \sim \frac{ab}{t} + \frac{48a^2 b^4 \log(t+1)^4}{\lambda t^\gamma} = O(t^{-(1\wedge\gamma)}\log(t)^4).$$

At the same time, we can check that $\epsilon_t \sim \frac{t^{-c^\star(U_\nu)b}}{bC_{\Gamma_G}\log t}$. Now, our conditions on η_t and ϵ_t imply that:

$$\int_0^\infty \epsilon_u \mathrm{d}u = +\infty \qquad \Longleftrightarrow \qquad b\,c^\star(U_\nu) \leq 1.$$

At the same time:

$$\eta_t = o(\epsilon_t) \qquad \Longleftrightarrow \qquad 1 \wedge \gamma > b\,c^\star(U_\nu).$$

The optimal calibration of our parameters γ and b (minimal value of γ, maximal value of b) induces the choice $\gamma = 1$ and $b < c^\star(U_\nu)^{-1}$. Up to the choices $\alpha_t \sim \lambda t$ and $\beta_t \sim b\log(t)$, we deduce that $K_t \longrightarrow 0$ when t goes to infinity. Since I_t, k_t and J_t are positive, this also implies that $J_t \longrightarrow 0$ when t goes to infinity. □

6. Functional inequalities

This section is devoted to the proof of the Log-Sobolev inequality for the measure $(\mu_\beta)_{\beta>0}$. We dealt with the Dirichlet form using this inequality (see Equation (18)) in Proposition 2. In particular, we need to obtain an accurate estimate when $\beta \longrightarrow +\infty$. For this purpose, we introduce the generic notation for Dirichlet forms (see, *i.e.*, [7] for a more in-depth description):

$$\forall \beta \in \mathbb{R}_+^* \quad \forall f \in W^{1,2}(\mu_\beta) \qquad \mathcal{E}_\beta(f,f) = \|\nabla f\|_{\mu_\beta}^2 = \int_{\Gamma_G} |\nabla f(x)|^2 \mathrm{d}\mu_\beta(x).$$

If we denote $<f>_{\mu_\beta} = \mu_\beta(f) = \int_{\Gamma_G} f\mathrm{d}\mu_\beta$, we are interested in showing the Poincaré inequality:

$$\|f - <f>_{\mu_\beta}\|_{2,\mu_\beta}^2 = \int_{\Gamma_G} (f - <f>_{\mu_\beta})^2 \mathrm{d}\mu_\beta \leq \lambda(\beta)\mathcal{E}_\beta(f,f) \tag{29}$$

and the Log-Sobolev Inequality (referred to as LSI below):

$$\int_{\Gamma_G} f^2 \log\left(\frac{f}{\|f\|_{2,\mu_\beta}}\right)^2 d\mu_\beta \leq C(\beta)\mathcal{E}_\beta(f,f) \tag{30}$$

The specific feature of this functional inequality deals with the quantum graph settings and deserves careful adaptation of the pioneering work of [31]. This technical section is split into two parts. The first one establishes a preliminary estimate when $\beta = 0$ (*i.e.*, when dealing with the uniform measure on Γ_G). The second one then uses this estimate to derive the asymptotic behavior of the LSI when $\beta \longrightarrow +\infty$.

6.1. Preliminary control for μ_0 on Γ_G

We consider μ_0 the normalized Lebesgue measure and use the standard notation for any measure μ on Γ_G:

$$\|f\|_{W^{1,p}(\mu)} := \|f\|_{p,\mu} + \|\nabla f\|_{p,\mu}$$

Let us establish the next elementary result:

LEMMA 1. *The ordinary Sobolev inequality holds on Γ_G, i.e., for all measurable functions f we have:*

$$\|f - <f>_0\|_{p,\mu_0}^2 \leq e^{2/e}L^2\left[\frac{\mathcal{D}_G L}{2} + 1\right]\mathcal{E}_0(f,f), \tag{31}$$

where $\mathcal{D}_G$ is the diameter of the graph and L its perimeter defined by:

$$L = \sum_{e \in E} L_e.$$

Proof: First, let us remind the reader that for a given interval I in dimension 1 equipped with the Lebesgue measure λ_I, the Sobolev space $W^{1,p}(\lambda_I)$ is continuously embedded in $L^\infty(I)$ (compact injection when I is bounded). In particular, it can be shown (see, *e.g.*, [16]) that while integrating w.r.t. the unnormalized Lebesgue measure:

$$\forall p \geq 1 \qquad \|f\|_{L^\infty(I)} \leq e^{1/e}\|f\|_{W^{1,p}(\lambda_I)}.$$

We now consider $f \in W^{1,p}(\mu_0)$. Since G has a finite number of edges we can write: $\Gamma_G = \bigcup_{i=1}^n e_i$, where each e_i can be seen as an interval of length ℓ_i. We have seen that for each edge e_i:

$$\|f\|_{L^\infty(e_i)} \leq e^{1/e}\|f\|_{W^{1,p}(\lambda_{e_i})}.$$

We can use a union bound since Γ_G represents the union of edges, and deduce that:

$$\|f\|_{L^\infty(\Gamma_G)} = \max_{1 \leq i \leq n}\|f\|_{L^\infty(e_i)} \leq e^{1/e}\max_{1 \leq i \leq n}\|f\|_{W^{1,p}(\lambda_{e_i})} \leq e^{1/e}\|f\|_{W^{1,p}(\lambda_{\Gamma_G})},$$

where the inequality above holds w.r.t. the unnormalized Lebesgue measure. If L denotes the sum of the lengths of all edges in Γ_G, we obtain:

$$\|f\|_{L^\infty(\Gamma_G)} \leq e^{1/e}L\|f\|_{W^{1,p}(\mu_0)}. \tag{32}$$

Second, we establish a simple Poincaré inequality for μ_0 on Γ_G. For any function $f \in W^{1,2}(\mu_0)$, we use the equality:

$$\|f - <f>_0\|_{2,\mu_0}^2 = \frac{1}{2}\int_{\Gamma_G}\int_{\Gamma_G}[f(x) - f(y)]^2 d\mu_0(x)d\mu_0(y) = \frac{1}{2}\int_{\Gamma_G}\int_{\Gamma_G}\left[\int_{\gamma_{x,y}} f'(s)ds\right]^2 d\mu_0(x)d\mu_0(y).$$

where $\gamma_{x,y}$ is the shortest path that connects x to y, parametrized with speed 1 and $f'(s)$ refers to the derivative of f w.r.t. this parametrization at time s. It should be noted that such a path exists because the graph Γ_G is connected. The Cauchy-Schwarz inequality yields:

$$\|f - <f>_0\|_{2,\mu_0}^2 \leq \frac{1}{2} \int_{\Gamma_G} \int_{\Gamma_G} \left[\int_{\gamma_{x,y}} |\nabla f(s)|^2 ds \right] |\gamma_{x,y}| \mathrm{d}\mu_0(x) \mathrm{d}\mu_0(y) \leq \frac{\mathcal{D}_G L}{2} \|\nabla f\|_{2,\mu_0}^2. \tag{33}$$

The Sobolev inequality is now an obvious consequence of the previous inequality: consider $f \in W^{1,p}(\mu_0)$ and note that since $\mu_0(\Gamma_G) = 1$, then (32) applied with $p = 2$ leads to:

$$\forall q \geq 1 \qquad \|f - <f>_0\|_{q,\mu_0}^2 = \left(\int_{\Gamma_G} |f - <f>_0|^q \mathrm{d}\mu_0 \right)^{2/q}$$
$$\leq \|f - <f>_0\|_{L^\infty(\Gamma_G)}^2 \mu_0(\Gamma_G)$$
$$\leq e^{2/e} L^2 \|f - <f>_0\|_{W^{1,2}(\mu_0)}^2.$$

We can use the Poincaré inequality established for μ_0 on Γ_G in Equation (33) and obtain:

$$\forall q \geq 1 \qquad \|f - <f>_0\|_{q,\mu_0}^2 \leq e^{2/e} L^2 \left[\frac{\mathcal{D}_G L}{2} + 1 \right] \|\nabla f\|_{2,\mu_0}^2 = e^{2/e} L^2 \left[\frac{\mathcal{D}_G L}{2} + 1 \right] \mathcal{E}_0(f, f),$$

which concludes the proof. $\qquad\square$

REMARK 1. The constants obtained in the proof of Lemma 1 above could certainly be improved. Nevertheless, such an improvement would have little importance for the final estimates obtained in Proposition 5.

6.2. Poincaré Inequality on μ_β

In the following, we show a Poincaré Inequality for the measure μ_β for large values of β. This preliminary estimate will be useful for deriving LSI on μ_β. This functional inequality is strongly related to the classical minimal elevation of the energy function U_ν for joining any state x to any state y.

We first introduce some useful notations. For any couple of vertices (x, y) of Γ_G, and for any path $\gamma_{x,y}$ that connects them, we define $h(\gamma_{x,y})$ as the highest value of U_ν on $\gamma_{x,y}$:

$$h(\gamma_{x,y}) = \max_{s \in \gamma_{x,y}} U_\nu(s).$$

We define $H(x, y)$ as the smallest value of $h(\gamma_{x,y})$ obtained for all possible paths from x to y:

$$H(x, y) = \min_{\gamma: x \to y} h(\gamma)$$

Now, for any pair of vertices x and y, the notation $\gamma_{x,y}$ will be reserved for the path that attains the minimum in the definition of $H(x, y)$. Such a path exists for any x, y because Γ_G is connected and possesses a finite number of paths that connect any two given vertices.

Finally, we introduce the quantity that will mainly determine the size of the spectral gap involved in the Poincaré inequality and the constant in the LSI (see the seminal works of [25] for a LDP probabilistic interpretation and [32] for a functional analysis point of view):

$$c^\star(U_\nu) := \max_{(x,y) \in \Gamma_G^2} [H(x, y) - U_\nu(x) - U_\nu(y)] + \min_{x \in \Gamma_G} U_\nu(x) \tag{34}$$

In the following, every time we write $\min U_\nu$ we refer to $\min_{x \in \Gamma_G} U_\nu(x)$.

THEOREM 4 (**Poincaré inequality for μ_β**). *For all measurable functions g defined on Γ_G:*

$$Var_{\mu_\beta}(g) \leq \frac{|E| \max_{e \in E} L_e e^{2\mathcal{D}_G}}{2} e^{\beta c^\star(U_\nu)} \mathcal{E}_{\mu_\beta}(g,g).$$

Proof: Since the graph Γ_G is connected, for any two points x and y, we can find a minimal path $\gamma_{x,y}$ that links them and that minimizes H. We denote $x_0 = x, x_1, \cdots x_l, x_{l+1} = y$ where the sequence $x_1, \cdots, x_l$ refers to the nodes included in the path $\gamma_{x,y}$.

$$
\begin{aligned}
Var_{\mu_\beta}(g) &= \frac{1}{2} \int_{\Gamma_G} \int_{\Gamma_G} (g(y) - g(x))^2 \mathrm{d}\mu_\beta(x) \mathrm{d}\mu_\beta(y) \\
&= \frac{1}{2} \int_{\Gamma_G} \int_{\Gamma_G} \left(\sum_{i=0}^l (g(x_{i+1}) - g(x_i)) \right)^2 \mathrm{d}\mu_\beta(x) \mathrm{d}\mu_\beta(y) \\
&= \frac{1}{2} \int_{\Gamma_G} \int_{\Gamma_G} \left(\sum_{i=0}^l \int_{x_i}^{x_{i+1}} g'(s) \mathrm{d}s \right)^2 \mathrm{d}\mu_\beta(x) \mathrm{d}\mu_\beta(y) \\
&= \frac{1}{2} \int_{\Gamma_G} \int_{\Gamma_G} \left(\int_{\gamma_{x,y}} g'(s) \mathrm{d}s \right)^2 \mathrm{d}\mu_\beta(x) \mathrm{d}\mu_\beta(y) \\
&\leq \frac{1}{2} \int_{\Gamma_G} \int_{\Gamma_G} \int_{\gamma_{x,y}} |\nabla g(s)|^2 \mathrm{d}s |\gamma_{x,y}| \mathrm{d}\mu_\beta(x) \mathrm{d}\mu_\beta(y),
\end{aligned}
$$

The last line is implied by the Cauchy-Schwarz inequality.

We can organize the terms involved in the above upper bound in the following way: for any edge of the graph $e \in E$, we denote V_e the set of points (x,y) such that $\gamma_{x,y} \cap e \neq \emptyset$. We therefore have:

$$
\begin{aligned}
Var_{\mu_\beta}(g) &\leq \frac{1}{2} \sum_{e \in E} \int_e |\nabla g(s)|^2 \mathrm{d}s \int_{V_e} |\gamma_{x,y}| \mathrm{d}\mu_\beta(x) \mathrm{d}\mu_\beta(y) \\
&\leq \frac{1}{2} \sum_{e \in E} \int_e |\nabla g(s)|^2 \mu_\beta(s) \int_{V_e} |\gamma_{x,y}| \frac{1}{\mu_\beta(s)} \mathrm{d}\mu_\beta(x) \mathrm{d}\mu_\beta(y) \mathrm{d}s \\
&\leq \frac{1}{2} \sum_{e \in E} \left[\int_e |\nabla g(s)|^2 \mu_\beta(s) \mathrm{d}s \right] \left[\sup_{\tilde{s} \in e} \frac{1}{\mu_\beta(\tilde{s})} \int_{V_e} |\gamma_{x,y}| \mathrm{d}\mu_\beta(x) \mathrm{d}\mu_\beta(y) \right] \\
&\leq \frac{1}{2} \int_{\Gamma_G} |\nabla g(s)|^2 \mu_\beta(s) \mathrm{d}s \sup_{\tilde{s} \in \Gamma_G} \frac{1}{\mu_\beta(\tilde{s})} \int_{V_{e_{\tilde{s}}}} |\gamma_{x,y}| \mathrm{d}\mu_\beta(x) \mathrm{d}\mu_\beta(y)
\end{aligned}
$$

In this case, $e_{\tilde{s}}$ denotes the edge of the graph that contains the point $\tilde{s}$ and the set $V_{e_{\tilde{s}}}$ is still the set of couples (x,y) defined above, associated with each edge $e_{\tilde{s}}$. We introduce the quantity $\mathcal{A}$ defined as:

$$\mathcal{A} = \sup_{\tilde{s} \in \Gamma_G} \frac{1}{\mu_\beta(\tilde{s})} \int_{V_{e_{\tilde{s}}}} |\gamma_{x,y}| \mathrm{d}\mu_\beta(x) \mathrm{d}\mu_\beta(y).$$

Using this notation, we have obtained that for all functions g, we have the Poincaré inequality:

$$Var_{\mu_\beta}(g) \leq \frac{\mathcal{A}}{2} \mathcal{E}_{\mu_\beta}(g,g). \tag{35}$$

All that remains to be done is to obtain an upper bound of $\mathcal{A}$.

$$\mathcal{A} = \sup_{\tilde{s} \in \Gamma_G} \frac{1}{\mu_\beta(s)} \int_{V_{e_{\tilde{s}}}} |\gamma_{x,y}| \mathrm{d}\mu_\beta(x) \mathrm{d}\mu_\beta(y)$$

$$
\begin{aligned}
&= \sup_{\tilde{s} \in \Gamma_G} Z_\beta e^{\beta U_\nu(\tilde{s})} \int_{V_{e_{\tilde{s}}}} \frac{e^{-\beta(U_\nu(x)+U_\nu(y))}}{Z_\beta^2} |\gamma_{x,y}| \mathrm{d}x \mathrm{d}y \\
&= \frac{1}{Z_\beta} \sup_{\tilde{s} \in \Gamma_G} \int_{V_{e_{\tilde{s}}}} e^{\beta(U_\nu(\tilde{s})-U_\nu(x)-U_\nu(y))} |\gamma_{x,y}| \mathrm{d}x \mathrm{d}y
\end{aligned}
$$

Since $\gamma_{x,y}$ is the minimal path for $H(x,y)$, we have $H(x,y) = \max_{s \in \gamma_{x,y}} U_\nu(s)$. Therefore:

$$
\begin{aligned}
\mathcal{A} &\leq \frac{1}{Z_\beta} \sup_{\tilde{s} \in \Gamma_G} \int_{V_{e_{\tilde{s}}}} e^{\beta((H(x,y)-U_\nu(x)-U_\nu(y))} |\gamma_{x,y}| \mathrm{d}x \mathrm{d}y \\
&\leq \frac{1}{Z_\beta} \sup_{\tilde{s} \in \Gamma_G} \int_{V_{e_{\tilde{s}}}} e^{\beta(c^\star(U_\nu)-\min(U_\nu))} |\gamma_{x,y}| \mathrm{d}x \mathrm{d}y \\
&\leq \frac{e^{\beta(c^\star(U_\nu)-\min(U_\nu))}}{Z_\beta} \sup_{\tilde{s} \in \Gamma_G} \int_{V_{e_{\tilde{s}}}} |\gamma_{x,y}| \mathrm{d}x \mathrm{d}y \\
&\leq \frac{e^{\beta(c^\star(U_\nu)-\min(U_\nu))}}{Z_\beta} |E| \max_{e \in E} L_e.
\end{aligned}
$$

Using the definition of Z_β, we have:

$$
\mathcal{A} \leq \frac{e^{\beta c^\star(U_\nu)} |E| \max_{e \in E} L_e}{\int_{\Gamma_G} e^{-\beta(U_\nu(x)-\min(U_\nu))} \mathrm{d}x}. \tag{36}
$$

If $x^\star \in M_\nu$ is a Fréchet mean that minimizes U_ν, we designate $\mathbb{B}\left(x^\star, \frac{1}{\beta}\right)$, as the ball of center $x^\star$ and radius $\frac{1}{\beta}$, for the geodesic distance d on the graph Γ_G. It is easy to check that:

$$
|U_\nu(x) - U_\nu(x^\star)| = \left| E_{Y \sim \nu} \left[d^2(x,Y) - d^2(x^\star, Y) \right] \right| \leq E_{Y \sim \nu} \left| d^2(x,Y) - d^2(x^\star, Y) \right| \leq 2\mathcal{D}_G \times d(x,x^\star).
$$

We can then deduce a lower bound on the denominator involved in (36):

$$
\begin{aligned}
\int_{\Gamma_G} e^{-\beta(U_\nu(x)-\min(U_\nu))} \mathrm{d}x &\geq \int_{\mathbb{B}\left(x^\star, \frac{1}{\beta}\right)} e^{-\beta(U_\nu(x)-\min(U_\nu))} \mathrm{d}x \\
&\geq \int_{\mathbb{B}\left(x^\star, \frac{1}{\beta}\right)} e^{-2\mathcal{D}_G \beta d(x,x^\star)} \mathrm{d}x \\
&\geq e^{-2\mathcal{D}_G} \int_{\mathbb{B}\left(x^\star, \frac{1}{\beta}\right)} \mathrm{d}x.
\end{aligned}
$$

The Lebesgue measure of $\mathbb{B}\left(x^\star, \frac{1}{\beta}\right)$ may be lower bounded by β^{-1} since there is, at the least, one path in Γ_G passing by the point $x^\star$. Inserting this inequality in (36) gives:

$$
\mathcal{A} \leq |E| \max_{e \in E} L_e e^{2\mathcal{D}_G} \beta e^{\beta c^\star(U_\nu)}.
$$

Using this upper bound in (35) leads to the desired Poincaré inequality. $\qquad \square$

6.3.   Sobolev Inequalities on μ_β

6.3.1.   Preliminary control on Dirichlet forms The next result will be useful to derive a LSI for μ_β from a Poincaré inequality on μ_β (given in Theorem 4). It generalizes the Poincaré inequality for p norms with $p > 2$ while using the Sobolev inequality given in Lemma 1.

First, we introduce the maximal elevation of U_ν as:

$$M = \sup_{x \in \Gamma_G} U_\nu(x) - \inf_{x \in \Gamma_G} U_\nu(x). \tag{37}$$

Note that in our case, M may be upper bounded by $\mathcal{D}_G^2$.

PROPOSITION 4. *For any $p > 2$ and all measurable functions f, we have:*

$$\forall \beta \geq 0 \qquad \|f - <f>_\beta\|_{p,\mu_\beta}^2 \leq 4L^2 \left(\frac{\mathcal{D}_G L}{2} + 1\right) e^{2/e + M\beta} \mathcal{E}_\beta(f, f). \tag{38}$$

Proof: The Jensen inequality applied to the convex function $x \longmapsto x^p$ yields:

$$\begin{aligned}
\|f - <f>_\beta\|_{p,\mu_\beta}^p &= \int_{\Gamma_G} |f - <f>|_\beta^p \mathrm{d}\mu_\beta \\
&= \int_{\Gamma_G} \left| f(x) - \int_{\Gamma_G} f(y)\mathrm{d}\mu_\beta(y) \right|^p \mathrm{d}\mu_\beta(x) \\
&= \int_{\Gamma_G} \left| \int_{\Gamma_G} f(x) - f(y)\mathrm{d}\mu_\beta(y) \right|^p \mathrm{d}\mu_\beta(x) \\
&\leq \int_{\Gamma_G} \int_{\Gamma_G} |f(x) - f(y)|^p \, \mathrm{d}\mu_\beta(y)\mathrm{d}\mu_\beta(x) \\
&\leq \int_{\Gamma_G} (|f(x) - <f>_0| + |f(y) - <f>_0|)^p \, \mathrm{d}\mu_\beta(y)\mathrm{d}\mu_\beta(x)
\end{aligned}$$

Again, the Jensen inequality $|a + b|^p \leq 2^{p-1}[|a|^p + |b|^p]$ implies that:

$$\begin{aligned}
\|f - <f>_\beta\|_{p,\beta}^p &\leq 2^{p-1} \int_{\Gamma_G} |f(x) - <f>_0|^p + |f(y) - <f>_0|^p \mathrm{d}\mu_\beta(y)\mathrm{d}\mu_\beta(x) \\
&\leq 2^p \int_{\Gamma_G} |f(x) - <f>_0|^p \mathrm{d}\mu_\beta(x) = 2^p \|f - <f>_0\|_{p,\mu_\beta}^p
\end{aligned}$$

We conclude that:

$$\|f - <f>_\beta\|_{p,\mu_\beta}^2 \leq 4\|f - <f>_0\|_{p,\mu_\beta}^2. \tag{39}$$

Using the fact that:

$$\mu_\beta(x) = \frac{e^{-\beta \min U_\nu}}{Z_\beta} \leq \frac{e^{-\beta \min U_\nu}}{Z_\beta} \times Z_0 \mu_0(x),$$

we also have: $\|f - <f>_0\|_{p,\mu_\beta}^p \leq \dfrac{L e^{-\beta \min U_\nu}}{Z_\beta} \|f - <f>_0\|_{p,\mu_0}^p$, since Z_0 is the perimeter L of the graph Γ_G. Consequently, we obtain:

$$\|f - <f>_0\|_{p,\mu_\beta}^2 = \left(\|f - <f>_0\|_{p,\mu_\beta}^p\right)^{2/p} \leq \left(\frac{L e^{-\beta \min U_\nu}}{Z_\beta}\right)^{2/p} \|f - <f>_0\|_{p,\mu_0}^2$$

Using inequality (39), the fact that $Z_\beta \leq L e^{-\beta \min U_\nu}$ and the assumption $2/p < 1$, we conclude that:

$$\|f - <f>_\beta\|_{p,\mu_\beta}^2 \leq \frac{4L e^{-\beta \min U_\nu}}{Z_\beta} \|f - <f>_0\|_{p,\mu_0}^2.$$

The Sobolev inequality (1) implies that:

$$\|f- <f>_\beta\|_{p,\mu_\beta}^2 \le \frac{4L^3\left(\mathcal{D}_G L/2+1\right)e^{2/e-\beta\min U_\nu}}{Z_\beta}\mathcal{E}_0(f,f). \tag{40}$$

We now find an upper bound for the Dirichlet form $\mathcal{E}_0(f,f)$ that involves $\mathcal{E}_\beta(f,f)$:

$$
\begin{aligned}
\mathcal{E}_0(f,f) &= \int_{\Gamma_G} <\nabla f,\nabla f> \mathrm{d}\mu_0\\
&= Z_\beta \int_{\Gamma_G} <\nabla f,\nabla f> e^{\beta U_\nu(x)}\frac{e^{-\beta U_\nu(x)}}{Z_\beta}\mathrm{d}\mu_0\\
&\le \frac{Z_\beta \exp\left(\beta\sup\limits_{x\in\Gamma_G} U_\nu(x)\right)}{Z_0}\int_{\Gamma_G} <\nabla f,\nabla f>\mathrm{d}\mu_\beta\\
&\le \frac{Z_\beta \exp\left(\beta\sup\limits_{x\in\Gamma_G} U_\nu(x)\right)}{L}\mathcal{E}_\beta(f,f) \tag{41}
\end{aligned}
$$

Putting (40) and (41) together concludes the proof. $\qquad\square$

6.3.2. Log-Sobolev Inequality on μ_β For all probability measures μ and all measurable functions f, we denote:

$$\mathbf{Ent}_\mu(f^2) = \int_{\Gamma_G} f^2\log\left(\frac{f^2}{\|f\|_{2,\mu}^2}\right)\mathrm{d}\mu$$

PROPOSITION 5. *The Log-Sobolev Inequality holds on* Γ_G. *A constant* C_{Γ_G} *exists such that for all* $\beta\ge 0$ *and all* μ_β-*measurable functions* f, *we have:*

$$\mathbf{Ent}_{\mu_\beta}(f^2) \le C_{\Gamma_G}[1+\beta]e^{c^\star(U_\nu)\beta}\mathcal{E}_\beta(f,f)$$

Proof: We consider $\beta\ge 0$ and a μ_β measurable function f. We apply the Jensen inequality for the logarithmic function and the measure $f^2/\|f\|_{2,\mu_\beta}^2\mathrm{d}\mu_\beta$ to obtain:

$$
\begin{aligned}
\int_{\Gamma_G}\left(\frac{f}{\|f\|_{2,\mu_\beta}}\right)^2\log\left(\frac{f^2}{\|f\|_{2,\mu_\beta}^2}\right)\mathrm{d}\mu_\beta &= \frac{2}{p-2}\int_{\Gamma_G}\left(\frac{f}{\|f\|_{2,\mu_\beta}}\right)^2\log\left(\frac{|f|^{p-2}}{\|f\|_{2,\mu_\beta}^{p-2}}\right)\mathrm{d}\mu_\beta\\
&\le \frac{2}{p-2}\log\left(\int_{\Gamma_G}\frac{|f|^{p-2}}{\|f\|_{2,\mu_\beta}^{p-2}}\frac{f^2}{\|f\|_{2,\mu_\beta}^2}\mathrm{d}\mu_\beta\right)\\
&\le \frac{2}{p-2}\log\left(\frac{\|f\|_{p,\mu_\beta}^p}{\|f\|_{2,\mu_\beta}^p}\right) = \frac{p}{p-2}\log\left(\frac{\|f\|_{p,\mu_\beta}^2}{\|f\|_{2,\mu_\beta}^2}\right).
\end{aligned}
$$

Observing that for all $x,\delta>0$ we have $\log(x\delta)\le\delta x$, we obtain $\log(x)\le x\delta+\log(\delta^{-1})$. Therefore:

$$
\begin{aligned}
\forall\delta>0\qquad \int_{\Gamma_G} f^2\log\left(\frac{f}{\|f\|_{2,\mu_\beta}}\right)^2\mathrm{d}\mu_\beta &= \|f\|_{2,\mu_\beta}^2\int_{\Gamma_G}\left(\frac{f}{\|f\|_{2,\mu_\beta}}\right)^2\log\left(\frac{f^2}{\|f\|_{2,\mu_\beta}^2}\right)\mathrm{d}\mu_\beta\\
&\le \frac{p\|f\|_{2,\mu_\beta}^2}{p-2}\log\left(\frac{\|f\|_{p,\mu_\beta}^2}{\|f\|_{2,\mu_\beta}^2}\right)\\
&\le \frac{p\|f\|_{2,\mu_\beta}^2}{p-2}\left[\delta\frac{\|f\|_{p,\mu_\beta}^2}{\|f\|_{2,\mu_\beta}^2}+\log\frac{1}{\delta}\right].
\end{aligned}
$$

Replacing f with $f - <f>_\beta$ and choosing $\delta = e^{-M\beta}$ leads to:

$$\int_{\Gamma_G} (f - <f>_\beta)^2 \log\left(\frac{f - <f>_\beta}{\|f - <f>_\beta\|_{2,\beta}}\right)^2 \mathrm{d}\mu_\beta$$
$$\leq \frac{p}{p-2}\left[e^{-\beta M}\|f - <f>_\beta\|_{p,\mu_\beta}^2 + M\beta\|f - <f>_\beta\|_{2,\mu_\beta}^2\right].$$

Proposition 4 and the Poincaré inequality established in Theorem 4 yields:

$$\mathbf{Ent}_{\mu_\beta}\left[(f - <f>_\beta)^2\right] = \int_{\Gamma_G} (f - <f>_\beta)^2 \log\left(\frac{f - <f>_\beta}{\|f - <f>_\beta\|_{2,\mu_\beta}}\right)^2 \mathrm{d}\mu_\beta$$
$$\leq \frac{p}{p-2}\left(4L^2\frac{\mathcal{D}_G L + 2}{2}e^{2/e} + M\beta|E|\max_{e\in E} L_e \frac{e^{2\mathcal{D}_G}}{2}\beta e^{c^\star(U_\nu)\beta}\right)\mathcal{E}_\beta(f,f) \quad (42)$$

It remains to use Rothau's Lemma (see Lemma 5.1.4 of [7]) that states that for any measure μ and any constant a:

$$\mathbf{Ent}_\mu(g+a)^2 \leq \mathbf{Ent}_\mu g^2 + 2\int_{\Gamma_G} g^2 \mathrm{d}\mu.$$

Let $p = 4$ in Equation (42). Putting this together with Rothau's Lemma and Theorem 4, we obtain the following:

$$\mathbf{Ent}_{\mu_\beta}(f^2) = \int_{\Gamma_G} f^2 \log\left(\frac{f^2}{\|f\|_{2,\mu_\beta}^2}\right)\mathrm{d}\mu_\beta$$
$$\leq \mathbf{Ent}_{\mu_\beta}\left[(f - <f>_\beta)^2\right] + 2Var_{\mu_\beta}(f)$$
$$\leq \mathcal{E}_\beta(f,f)\left[\beta|E|\max_{e\in E} L_e e^{2\mathcal{D}_G}(1+M\beta)e^{c^\star(U_\nu)\beta} + 4L^2(2+\mathcal{D}_G L)e^{2/e}\right]$$
$$\leq C_{\Gamma_G}[1+\beta]e^{c^\star(U_\nu)\beta}\mathcal{E}_\beta(f,f),$$

where C_{Γ_G} is a large enough constant (independent of β) that could be made explicit in terms of constants $\mathcal{D}_G$, $|E|$ and L since we trivially have $M \leq \mathcal{D}_G^2$. $\qquad\square$

References

[1] Acemoğlu D., Fagnani F. Ozdaglar A., Como G. 2013. Opinion fluctuations and disagreement in social networks. *Math. Oper. Res.* **38**(1) 1–27.

[2] Allassonnière S., Trouvé A., Kuhn E. 2010. Construction of bayesian deformable models via stochastic approximation algorithm: A convergence study. *Bernoulli* **16** 641–678.

[3] Arnaudon M., Miclo L. 2014. Means in complete manifolds: uniqueness and approximation. *ESAIM: Probability and Statistics* **18** 185–206.

[4] Arnaudon M., Miclo L. 2014. A stochastic algorithm finding generalized means on compact manifolds. *Stochastic Processes and their Applications* **124** 34633479.

[5] Arnaudon M., Miclo L. 2016. A stochastic algorithm finding p-means on the circle. *Bernoulli, in press*.

[6] Bach F., Jordan M. 2004. Learning spectral clustering. *Advances in Neural Information Processing Systems* 305–312.

[7] Bakry D., Ledoux M., Gentil I. 2014. *Analysis and Geometry of Markov Diffusion Operators*, *Grundlehren der mathematischen Wissenschaften*, vol. 348. Springer.

[8] Barden D., Owen M., Le H. 2013. Central limit theorems for Fréchet means in the space of phylogenetic trees. *Electronic Journal of Probability* **18** 1–25.

[9] Bhattacharya R., Patrangenaru V. 2003. Large sample theory of intrinsic and extrinsic sample means on manifolds. *Ann. Statist.* **31** 1–29.

[10] Bigot, J. 2013. Fréchet means of curves for signal averaging and application to ECG data analysis. *Annals of Applied Statistics* **7** 1837–2457.

[11] Bigot J., Gadat S. 2010. A deconvolution approach to estimation of a common shape in a shifted curves model. *Annals of Statistics* **38** 224–243.

[12] Bigot J., Gendre X. 2013. Minimax properties of Fréchet means of discretely sampled curves. *Annals of Statistics* **41** 923–956.

[13] Bigot J., Klein T. Lopez A., Gouet R. 2015. Geodesic PCA in the wasserstein space by convex PCA. *Annales de l'Institut Henri Poincaré B: Probability and Statistics, to appear* .

[14] Bigot J., Lopez A., Gouet R. 2013. Geometric PCA of images. *Siam, J. Imaging Sci.* **6** 1851–1879.

[15] Bontemps D., Gadat S. 2014. Bayesian methods for the shape invariant model. *Electronic Journal of Statistics* **8** 1522–1568.

[16] Brezis, H. 1987. *Analyse fonctionelle*. Masson, Paris.

[17] Catoni, O. 1992. Rough large deviation estimates for simulated annealing : application to exponential schedules. *Ann. Probab.* **20** 1109–1146.

[18] Dijkstra, E. W. 1959. A note on two problems in connexion with graphs. *Numer. Math.* **1** 269–271.

[19] Dryden I. L., Mardia K. V. 1998. *Statistical Shape Analysis*.

[20] Erdös P., Rényi A. 1960. The evolution of random graphs. *Magyar Tud. Akad. Mat. Kutato Int. Kozl.* **5** 17–61.

[21] Estrada, E. 2015. *Introduction to Complex Networks. Structure and Dynamics, chapter of Evolutionary Equations with Applications to Natural Sciences*. Lecture Notes in Mathematics, Springer.

[22] Ethier S. N., Kurtz T. 2005. *Markov processes. Characterization and convergence.*. Wiley Series in Probability and Statistics, John Wiley & Sons.

[23] Fréchet, M. 1948. Les éléments aléatoires de nature quelconque dans un espace distancié. *Annales de l'Institut Henri Poincaré (B)* **10** 215–310.

[24] Freidlin M., Sheu S.J. 2000. Diffusion processes on graphs: stochastic differential equations, large deviation principle. *Probab. Theory Relat. Fields* **116** 181–220.

[25] Freidlin M. I., Wentzell A. D. 1979. *Random perturbations of dynamical systems*, *Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences]*, vol. 260. 2nd ed. Springer-Verlag, New York. Translated from the 1979 Russian original by Joseph Szücs.

[26] Freidlin M. I., Wentzell A. D. 1995. *Random perturbations of Hamiltonian systems*, *Memoirs of the American Mathematical Society*, vol. 523. A.M.S., New York.

[27] Ginestet, C.E. 2013. Strong consistency of set-valued fréchet sample means in metric spaces. *Preprint* .

[28] Goldenberg A., Fienberg S.E. Airoldi E. M., Zheng A. X. 2010. A survey of statistical network models. *Found. Trends Mach. Learn.* **2**(2) 129–233.

[29] Gutjahr W. J., Pflug G. 1996. Simulated annealing for noisy cost functions. *J. Global Optim.* **8**(1) 1–13.

[30] Hajeck, B. 1988. Cooling schedules for optimal annealing. *Mathematics of Operations Research* **13** 311–329.

[31] Holley, Richard, Daniel Stroock. 1988. Simulated annealing via Sobolev inequalities. *Comm. Math. Phys.* **115**(4) 553–569.

[32] Holley R., Stroock D., Kusuoka D. 1989. Asymptotics of the spectral gap with applications to the theory of simulated annealing. *J. Funct. Anal.* **83**(2) 333–347.

[33] Ikeda N., Watanabe S. 1981. *Stochastic Differential Equations and Diffusion Processes*. North-Holland.

[34] Jackson, M. O. 2008. *Social and Economic Networks*. Princeton Univ. Press, Princeton, NJ.

[35] Kaufmann M., Wagner D. 2001. *Drawing graphs: methods and models*. Lecture Notes in Computer Science, Springer.

[36] Klopp O., Verzelen N., Tsybakov A.B. 2016. Oracle inequalities for network models and sparse graphon estimation. *Annals of Statistics, in press* .

[37] Kolaczyk, E. D. 2009. *Statistical analysis of network data: methods and models*. Springer Series in Statistics, Springer New-York.

[38] Le, H. 2001. Locating Fréchet means with application to shape spaces. *Adv. Appl. Probab.* **33** 324–338.

[39] Lovasz, L. 2012. *Large networks and graph limits*, vol. 60. American Mathematical Society.

[40] Miclo, Laurent. 1992. Recuit simulé sur $\mathbf{R}^n$. Étude de l'évolution de l'énergie libre. *Ann. Inst. H. Poincaré Probab. Statist.* **28**(2) 235–266.

[41] Miller E., Provan J., Owen M. 2015. Polyhedral computational geometry for averaging metric phylogenetic trees. *Advances in Applied Mathematics* **68** 51–91.

[42] Munch E., Bendich P. Mukherjee S. Mattingly J. Harer J., Turner K. 2015. Probabilistic Fréchet means for time varying persistence diagrams. *Electronic Journal of Statistics* **9** 1173–1204.

[43] Newman, M. 2010. *Networks, An Introduction*. Oxford University Press.

[44] Pennec, X. 2006. Intrinsic statistics on Riemannian manifolds: basic tools for geometric measurements. *J. Math. Imaging Vision* **25** 127–154.

[45] Shneiderman B., Aris A. 2006. Network visualization by semantic substrates. *IEEE Transactions on Visualization and Computer Graphics* **12** 733–740.

[46] Trouvé, A. 1993. Parallélisation massive du recuit simulé. *PhD Thesis, Université d'Orsay* .

[47] Watts D. J., Strogatz S. H. 1998. Collective dynamics of 'small-world' networks. *Nature* **393**(6684) 409–10.