ARTICLE IN PRESS YJDEQ:8362

Available online at www.sciencedirect.com

ScienceDirect

J. Differential Equations ••• (••••) •••–•••

Journal of Differential Equations

www.elsevier.com/locate/jde

Convergence to global equilibrium for Fokker–Planck equations on a graph and Talagrand-type inequalities ☆

Rui Che [a], Wen Huang [a], Yao Li [b,*], Prasad Tetali [c]

[a] Wu Wen-Tsun Key Laboratory of Mathematics, USTC, Chinese Academy of Sciences, Hefei, Anhui 230026, China

[b] Department of Mathematics and Statistics, University of Massachusetts Amherst, Amherst, MA 01003, USA

[c] School of Mathematics, Georgia Institute of Technology, Atlanta, GA 30332, USA

Received 13 January 2016; revised 25 April 2016

Abstract

In 2012, Chow, Huang, Li and Zhou [7] proposed the Fokker–Planck equations for the free energy on a finite graph, in which they showed that the corresponding Fokker–Planck equation is a nonlinear ODE defined on a Riemannian manifold of probability distributions. Different choices for inner products result in different Fokker–Planck equations. The unique global equilibrium of each equation is a Gibbs distribution. In this paper we proved that the exponential rate of convergence towards the global equilibrium of these Fokker–Planck equations. The rate is measured by both the decay of the L^2 norm and that of the (relative) entropy. With the convergence result, we also prove two Talagrand-type inequalities relating relative entropy and Wasserstein metric, based on two different metrics introduced in [7]. The first one is a local inequality, while the second is a global inequality with respect to the "lower bound metric" from [7].

MSC: 37H10; 60J27; 60J60

Keywords: Fokker–Planck equation; Gibbs density; Graph; Talagrand inequality

☆ The second author is supported by NSFC (Nos. 11225105, 11371339, 11431012). The work of the last author is supported in part by NSF DMS-1101447 and DMS-1407657.

* Corresponding author.

E-mail addresses: boboke-1126@163.com (R. Che), wenh@mail.ustc.edu.cn (W. Huang), yaoli@math.umass.edu (Y. Li), tetali@math.gatech.edu (P. Tetali).

http://dx.doi.org/10.1016/j.jde.2016.05.003
0022-0396/© 2016 Published by Elsevier Inc.

Please cite this article in press as: R. Che et al., Convergence to global equilibrium for Fokker–Planck equations on a graph and Talagrand-type inequalities, J. Differential Equations (2016), http://dx.doi.org/10.1016/j.jde.2016.05.003

1. Introduction

As the stochastic differential equation becomes one of the primary and highly effective tools in many practical problems arising in diverse fields such as finance, physics, chemistry and biology [10,22,26], there are considerable efforts in understanding the properties of the classical Fokker–Planck equation that describes the time evolution of the probability distribution of a stochastic process. At the same time, the free energy functional, which is defined on the space of probability distributions, as a linear combination of terms involving a potential and an entropy, has widely been used in various subjects; it typically means different things in different contexts. For example, the notion of "free energy" in thermodynamics is related to the maximal amount of work that can be extracted from a system. The concept of free energy is also used in other fields, such as statistical mechanics, probability (particularly in the context of Markov Random Fields), biology, chemistry, and image processing; see e.g., [16,29,33].

Since the seminal work of Jordan, Kinderlehrer and Otto [12,20], it is well known that a Fokker–Planck equation is the gradient flow of the free energy functional on a Riemannian manifold that is defined by a space of probability distributions with a 2-Wasserstein metric on it. This discovery has been the starting point for many developments relating the free energy, Fokker–Planck equation, an abstract notion of a Ricci curvature and optimal transport theory in continuous spaces. We refer to the monographs [1,31,32] for an overview. Recently, a synthetic theory of Ricci curvature in length spaces has been developed by Lott–Sturm–Villani [13,27,28], which reveals the fundamental relationship between entropy and Ricci curvature. Despite the remarkable developments on this subject on a continuous space, much less is known when the underlying space is discrete, as in an (undirected) graph.

In recent work, Chow, Huang, Li and Zhou [7] considered Fokker-planck equations for a free energy function (or a certain Markov process) defined on a finite graph. For a graph on $N \geq 2$ vertices, they showed that the corresponding Fokker–Planck equation is a system of N nonlinear ordinary differential equations, defined on a Riemannian manifold of probability distributions. In fact, they point out that one could make different choices for inner products on the space of probability distributions resulting, in turn, in different Fokker–Planck equations for the same process. Furthermore, each of these systems of ordinary differential equations has a unique global equilibrium – a Gibbs distribution – and is a gradient flow for the free energy functional defined on a Riemannian manifold whose metric is closely related to certain classical Wasserstein metrics.

We recall here, more formally, the approach of Chow et al. [7]. Consider a graph $G = (V, E)$, where $V = \{a_1, \cdots, a_N\}$ is the set of vertices $|V| \geq 2$, and E denotes the set of (undirected) edges. For simplicity, assume that the graph is connected and is simple – with no self-loops or multiple edges. Let $N(i) := \{j \in \{1, 2, \cdots, N\} | \{a_i, a_j\} \in E\}$ denote the neighborhood of a vertex $a_i \in V$.

Let $\boldsymbol{\Psi} = (\Psi_i)_{i=1}^N$ be a given *potential* function on V, where Ψ_i is the potential on vertex a_i. Further denote

$$\mathcal{M} = \{\boldsymbol{\rho} = (\rho_i)_{i=1}^N \in \mathbb{R}^N | \sum_{i=1}^N \rho_i = 1 \text{ and } \rho_i > 0 \text{ for } i = 1, 2, \cdots, N\},$$

as the space of all positive probability distributions on V.

Please cite this article in press as: R. Che et al., Convergence to global equilibrium for Fokker–Planck equations on a graph and Talagrand-type inequalities, J. Differential Equations (2016), http://dx.doi.org/10.1016/j.jde.2016.05.003

Then the free energy functional has the following expression: for each $\boldsymbol{\rho} \in \mathcal{M}$, let

$$F(\boldsymbol{\rho}) := F_{\boldsymbol{\Psi},\beta}(\boldsymbol{\rho}) = \sum_{i=1}^{N} \Psi_i \rho_i + \beta \sum_{i=1}^{N} \rho_i \log \rho_i, \tag{1.1}$$

where $\beta > 0$ is the strength of "white noise" or the temperature. The free energy functional has a global minimizer, called a Gibbs density, and is given by

$$\rho_i^* = \frac{1}{K} e^{-\Psi_i/\beta}, \qquad \text{where } K = \sum_{i=1}^{N} e^{-\Psi_i/\beta}. \tag{1.2}$$

From a free energy viewpoint, Chow et al. [7] endowed the space $\mathcal{M}$ with a Riemannian metric d_Ψ, which depended on the potential Ψ as well as the structure of the graph. Then by considering the gradient flow of the free energy (1.1) on such a Riemannian manifold $(\mathcal{M}, d_\Psi)$, they obtained a Fokker–Planck equation on $\mathcal{M}$:

$$\begin{aligned} \frac{d\rho_i}{dt} = & \sum_{j \in N(i), \Psi_j > \Psi_i} ((\Psi_j + \beta \log \rho_j) - (\Psi_i + \beta \log \rho_i))\rho_j \\ & + \sum_{j \in N(i), \Psi_j < \Psi_i} ((\Psi_j + \beta \log \rho_j) - (\Psi_i + \beta \log \rho_i))\rho_i \\ & + \sum_{j \in N(i), \Psi_j = \Psi_i} \beta(\rho_j - \rho_i) \end{aligned} \tag{1.3}$$

for $i = 1, 2 \cdots, N$ (see Theorem 2 in [7]).

From a stochastic process viewpoint, the work of Chow et al. may be seen as a new interpretation of white noise perturbations to a Markov process on V. By considering the time evolution equation of its probability density function, they obtained another Fokker–Planck equation on $\mathcal{M}$:

$$\begin{aligned} \frac{d\rho_i}{dt} = & \sum_{j \in N(i), \bar{\Psi}_j > \bar{\Psi}_i} ((\Psi_j + \beta \log \rho_j) - (\Psi_i + \beta \log \rho_i))\rho_j \\ & + \sum_{j \in N(i), \bar{\Psi}_j < \bar{\Psi}_i} ((\Psi_j + \beta \log \rho_j) - (\Psi_i + \beta \log \rho_i))\rho_i \\ = & \sum_{j \in N(i)} ((\Psi_j + \beta \log \rho_j) - (\Psi_i + \beta \log \rho_i))^+ \rho_j \\ & - \sum_{j \in N(i)} ((\Psi_j + \beta \log \rho_j) - (\Psi_i + \beta \log \rho_i))^- \rho_i \end{aligned} \tag{1.4}$$

for $i = 1, 2, \cdots, N$, where $\bar{\Psi}_i = \Psi_i + \beta \log \rho_i$ for $i = 1, 2, \cdots, N$ (see Theorem 3 in [7]). For convenience, we call equations (1.3) and (1.4) Fokker–Planck equation I (1.3) and II (1.4) respectively. Both (1.3) and (1.4) share similar properties for $\beta > 0$ (see Theorem 2 and Theorem 3 in [7]):

(1) Both equations are gradient flows of the same free energy on the same probability space $\mathcal{M}$, but with different metrics.
(2) The Gibbs distribution $\boldsymbol{\rho}^* = (\rho_i^*)_{i=1}^N$, given by (1.2), is the unique stationary distribution of both equations in $\mathcal{M}$. Furthermore, the free energy F attains its global minimum at the Gibbs distribution.
(3) For both equations, given any initial condition $\boldsymbol{\rho}^0 \in \mathcal{M}$, there exists a unique solution

$$\boldsymbol{\rho}(t) : [0, \infty) \to \mathcal{M}$$

with the initial value $\boldsymbol{\rho}^0 \in \mathcal{M}$, and $\boldsymbol{\rho}(t)$ satisfying the properties:
 (a) the free energy $F(\boldsymbol{\rho}(t))$ decreases as time t increases, and
 (b) $\boldsymbol{\rho}(t) \to \boldsymbol{\rho}^*$ under the Euclidean metric of $\mathbb{R}^N$, as $t \to +\infty$.

There are differences between equations (1.3) and (1.4). Fokker–Planck equation I (1.3) is obtained from the gradient flow of the free energy F on the Riemannian metric space $(\mathcal{M}, d_{\boldsymbol{\Psi}})$. However, its connection to a Markov process on the graph is not clear. On the other hand, as explained in Section 5 of [7], Fokker–Planck equation II (1.4) is obtained from a Markov process subject to "white noise" perturbations. This equation can also be considered as a generalized gradient flow of the free energy on another metric space $(\mathcal{M}, d_{\bar{\boldsymbol{\Psi}}})$ (see Theorem 3 in [7]). However, the geometry of $(\mathcal{M}, d_{\bar{\boldsymbol{\Psi}}})$ is not smooth. In fact, Chow et al. showed that, in this case, $\mathcal{M}$ is divided into finite segments, and $d_{\bar{\boldsymbol{\Psi}}}$ is only smooth on each segment.

By the above discussion, we know that the Gibbs distribution $\boldsymbol{\rho}^* = (\rho_i^*)_{i=1}^N$ is the unique global equilibrium of both equations (1.3) and (1.4) in $\mathcal{M}$, and for any solution $\boldsymbol{\rho}(t)$ of both equations (1.3) and (1.4), $\boldsymbol{\rho}(t)$ will converge to global equilibrium $\boldsymbol{\rho}^*$, under the Euclidean metric of $\mathbb{R}^N$, as $t \to \infty$. A natural next question this raises is then that of the derivation of estimates, such as $O(e^{-ct})$ for a suitable $c > 0$, on the rate of convergence to global equilibrium for solutions of both equations (1.3) and (1.4). Answering such a question is one of the main objectives of the present paper. The rates of convergence towards global equilibrium for the solution of these Fokker–Planck equations on a graph are investigated. We will prove that the convergence is indeed exponential.

In [7], the authors introduced several metrics on the space of probability measures $\mathcal{M}$, including d_Ψ, $d_{\bar{\Psi}}$, an upper bound metric d_M and a lower bound metric d_m, where the latter two are independent of the choice of potential. These distances are obtained in a sense by discretizing Felix Otto's calculus – there is a certain similarity between these distances and the 2-Wasserstein distance on the space of probability measures on $\mathbb{R}^n$. For example, the gradient flow of free energy functional (defined using relative entropy) in these metric spaces gives rise to the discrete Fokker–Planck equation in [7]. It is worth mentioning that the geodesic of d_Ψ is a discretization of the geodesic equation in 2-Wasserstein distance on the space of probability measures on $\mathbb{R}^n$. For these reasons, sometimes we refer to these as discrete 2-Wasserstein distances.

As an important application of our convergence result, Talagrand-type inequalities are proved. We will show that the 2-Wasserstein distance is bounded from above by the relative entropy: that for all $\boldsymbol{\nu}$ absolutely continuous with respect to $\boldsymbol{\mu}$, it holds:

$$d_m^2(\boldsymbol{\nu}, \boldsymbol{\mu}) \le K H(\boldsymbol{\nu}|\boldsymbol{\mu}),$$

where K depends only on the (reference measure) $\boldsymbol{\mu}$.

In recent years, there has been considerable interest in deriving such inequalities in various spaces, with the purpose of studying geometric inequalities connected to concentration of measure and other phenomena. On a Riemannian manifold, Otto and Villani [21] showed (inter alia) that a logarithmic Sobolev inequality implied a Talagrand inequality; this work was soon generalized and simplified by Bobkov, Gentil and Ledoux [4]; the latter provided simpler proofs of several previously known results concerning log-Sobolev and transport inequalities. See also [3] for an earlier work which (along with [21]) inspired much of the research in this topic. In subsequent work, Lott and Villani [14] used the Hamilton–Jacobi semigroup approach of Bobkov et al. [4] in showing that a Talagrand inequality on a measured length space implied a global Poincaré inequality, as well as in obtaining (conversely), that spaces satisfying a certain doubling condition, a local Poincaré inequality and a log-Sobolev inequality satisfied a Talagrand inequality.

In discrete spaces, such inequalities are less understood. In part, the lack of a suitable 2-Wasserstein (W_2) distance between probability measures on a graph has slowed this progress. M. Sammer and the last author of this work observed (see [24,25] for a proof) that a derivation of Otto–Villani goes through in the context of a finite graph in yielding the implication that a (weaker) *modified* log-Sobolev inequality implies a (weaker) Talagrand-type inequality, relating a W_1-distance (rather than a W_2) and the relative entropy.

In the following, we obtain in fact two versions of a Talagrand inequality. The first one is only locally true, which means that the parameter K also depends on the range of $\boldsymbol{\nu}$ – it is true for all measures $\boldsymbol{\nu}$ in a compact neighborhood of $\boldsymbol{\mu}$, but may not be true if $\boldsymbol{\nu}$ can be arbitrarily close to $\partial\mathcal{M}$.

The "global" Talagrand inequality holds for the "lower bound" metric. If the graph G is simple and connected, then there exists a parameter K that only depends on μ and certain parameters of G. As far as we know, a Talagrand-type inequality for the W_2-distance has not been established on a discrete space, and in fact various people have independently observed that a literal translation of such an inequality, borrowed from the continuous case, need not hold even on a 2-point discrete space (see e.g., [17,11]). However, our results suggest that the metrics introduced in [7] have a further similarity with the W_2-distance on the space of probability measures of the length space.

Independently, a related class of metrics has been studied by Mielke in [18,19] and Maas in [17], which are similar to the Riemannian metrics in [7] with a constant potential. In the setting of both [18] and [17], the graphs are assumed to be associated with an irreducible and reversible Markov kernel. A finite volume discretization of the Fokker–Planck equation that preserves the gradient-flow structure is developed by [8]. After essentially finishing this paper, the authors have been informed that functional inequalities including modified Talagrand inequality and modified logarithmic Sobolev inequalities associated with the Riemannian metric studied in [17,18] are independently investigated in [9].

2. Preliminary

In this section, we recall some definitions in graph theory. A *graph* is an ordered pair $G = (V, E)$ where $V = \{a_1, \cdots, a_N\}$ is the set of vertices and E is the set of edges. We further assume that the graph G is a simple graph (that is, there are no self loops or multiple edges) with $|V| \geq 2$, and G is connected. A *weighted graph* (G, w) is a pair consisting of a graph $G = (V, E)$ and a positive real-valued function w of its edges. The function w is most conveniently described as an

$|V|$-by-$|V|$, symmetric, nonnegative matrix $w = (w_{ij})$ with the property that $w_{ij} > 0$ if and only if $(a_i, a_j) \in E$.

Given a graph $G = (V, E)$ with $V = \{a_1, a_2, \cdots, a_N\}$, we consider all *positive probability distributions* on V:

$$\mathcal{M} = \left\{ \boldsymbol{\rho} = (\rho_i)_{i=1}^N \in \mathbb{R}^N \,|\, \sum_{i=1}^N \rho_i = 1 \text{ and } \rho_i > 0 \text{ for } i \in \{1, 2, \cdots, N\} \right\}.$$

For $\boldsymbol{\mu} = (\mu_i)_{i=1}^N \in \mathcal{M}$ and any map $f : V \to \mathbb{R}$, recall the $L^2(\boldsymbol{\mu})$-*norm* of f with respect to μ, denoted by $||f||_{2,\boldsymbol{\mu}}$, and given by:

$$||f||_{2,\boldsymbol{\mu}}^2 := \sum_{i=1}^N (f(a_i))^2 \mu_i .$$

Let $\boldsymbol{\nu} = (\nu_i)_{i=1}^N \in \mathcal{M}$, then the *relative entropy* $H(\boldsymbol{\nu}|\boldsymbol{\mu})$ of $\boldsymbol{\nu}$ with respect to $\boldsymbol{\mu}$ is defined by:

$$H(\boldsymbol{\nu}|\boldsymbol{\mu}) = \sum_{i=1}^N \nu_i \log \frac{\nu_i}{\mu_i},$$

and we measure the distance between (the density of) $\boldsymbol{\nu}$ and $\boldsymbol{\mu}$ using:

$$||\frac{\boldsymbol{\nu}}{\boldsymbol{\mu}} - 1||_{2,\boldsymbol{\mu}}^2 := \sum_{i=1}^N (\frac{\nu_i}{\mu_i} - 1)^2 \mu_i .$$

Given a graph $G = (V, E)$, its *Laplacian matrix* is defined as:

$$\mathcal{L}(G) := D - A ,$$

where D is a diagonal matrix with $d_{ii} = \deg(a_i)$ (number of edges at a_i), and A is the *adjacency matrix* ($A_{ij} = 1$ if and only if $\{a_i, a_j\} \in E$). As G is a connected simple graph, it is well known that $\mathcal{L}(G)$ has one 0 eigenvalue and $|V| - 1$ positive eigenvalues.

Given a weighted graph (G, w), its *weighted Laplacian matrix* is defined as

$$\mathcal{L}(G, w) = diag(\delta_1, \delta_2, \cdots, \delta_{|V|}) - w$$

with δ_i denoting the ith row sum of w. It is well known that $\mathcal{L}(G, w)$ also has one 0 eigenvalue and $|V| - 1$ positive eigenvalues.

The second smallest eigenvalue λ_2 of $\mathcal{L}(G)$ (resp. $\mathcal{L}(G, w)$) is called the *spectral gap* of G (resp. (G, w)). We remind readers that there are various standard ways to bound the spectral gap of a graph. For example for the spectral gap λ_2 of $\mathcal{L}(G)$, see [2] for the bound

$$\lambda_2 \geq d_{max} - \sqrt{d_{max}^2 - d_{min}^2} ,$$

where d_{max} and d_{min} are the maximum and minimum degrees of vertices in G; similarly see [15], for

$$\lambda_2 \geq \frac{2N}{2 + N(N-1)d - 2Md},$$

where N is the number of vertices, M is the number of edges, and d is the diameter of G; or [30] for the bound,

$$\lambda_2 \geq 2(1 - \cos(\frac{\pi}{N})).$$

3. The trend towards equilibrium

The rate of convergence towards global equilibrium for the solution of Fokker–Planck equations (1.3) and (1.4) in weighted L^2 norm is estimated in this section. We will prove that such convergence is exponential. In addition, the relative entropy (with respect to the global equilibrium) also has exponential decay.

3.1. Convergence in weighted L^2 norm

The following is our first main result.

Theorem 3.1. *Let $G = (V, E)$ be a graph with its vertex set $V = \{a_1, a_2, \cdots, a_N\}$, edge set E, a given potential $\mathbf{\Psi} = (\Psi_i)_{i=1}^N$ on V and a constant $\beta > 0$. If $\boldsymbol{\rho}(t) = (\rho_i(t))_{i=1}^N : [0, \infty) \to \mathcal{M}$ is the solution of the Fokker–Planck equation I (1.3), with the initial value $\boldsymbol{\rho}^o = (\rho_i^o)_{i=1}^N \in \mathcal{M}$, then there exists a constant $C = C(\boldsymbol{\rho}^o; G, \Psi, \beta) > 0$ such that*

$$||\frac{\boldsymbol{\rho}(t)}{\boldsymbol{\rho}^*} - 1||^2_{2,\boldsymbol{\rho}^*} \leq ||\frac{\boldsymbol{\rho}^o}{\boldsymbol{\rho}^*} - 1||^2_{2,\boldsymbol{\rho}^*} e^{-Ct}, \tag{3.1}$$

where $\boldsymbol{\rho}^* = (\rho_i^*)_{i=1}^N$ is the Gibbs distribution given by (1.2). In particular, $\boldsymbol{\rho}(t)$ exponentially converges to global equilibrium: the Gibbs distribution $\boldsymbol{\rho}^*$ under the Euclidean metric of $\mathbb{R}^N$ as $t \to \infty$.

Proof. Given initial value $\boldsymbol{\rho}^o = (\rho_i^o)_{i=1}^N \in \mathcal{M}$. Let $\boldsymbol{\rho}(t) = (\rho_i(t))_{i=1}^N : [0, \infty) \to \mathcal{M}$ be the solution of Fokker–Planck equation I (1.3) with initial value $\boldsymbol{\rho}^o \in \mathcal{M}$. For $t \geq 0$, we define

$$L(t) = ||\frac{\boldsymbol{\rho}(t)}{\boldsymbol{\rho}^*} - 1||^2_{2,\boldsymbol{\rho}^*} = \sum_{i=1}^N \frac{(\rho_i(t) - \rho_i^*)^2}{\rho_i^*},$$

where $\boldsymbol{\rho}^* = (\rho_i^*)_{i=1}^N$ is the Gibbs distribution given by (1.2). Now for $t > 0$ by (1.3) we have

$$
\begin{aligned}
\frac{dL(t)}{dt} &= \sum_{i=1}^{N} \frac{2(\rho_i(t)-\rho_i^*)}{\rho_i^*} \frac{d\rho_i(t)}{dt} \\
&= \sum_{i=1}^{N} \frac{2(\rho_i(t)-\rho_i^*)}{\rho_i^*} \Big(\sum_{j\in N(i),\Psi_j>\Psi_i} \big((\Psi_j+\beta\log\rho_j(t)) - (\Psi_i+\beta\log\rho_i(t))\big)\rho_j(t) \\
&\quad + \sum_{j\in N(i),\Psi_j<\Psi_i} \big((\Psi_j+\beta\log\rho_j(t)) - (\Psi_i+\beta\log\rho_i(t))\big)\rho_i(t) \\
&\quad + \sum_{j\in N(i),\Psi_j=\Psi_i} \beta(\rho_j(t)-\rho_i(t))\Big).
\end{aligned}
$$

Note that $\Psi_j - \Psi_i = -\beta\log\rho_j^* + \beta\log\rho_i^*$ for $i, j \in \{1, 2, \cdots, N\}$ and $\rho_j^* = \rho_i^*$ when $\Psi_j = \Psi_i$. Combining this with the above equality, we have

$$
\begin{aligned}
&\frac{dL(t)}{dt} \\
&= \sum_{i=1}^{N} \frac{2(\rho_i(t)-\rho_i^*)}{\rho_i^*} \Big(\sum_{j\in N(i),\Psi_j>\Psi_i} \big((-\beta\log\rho_j^* + \beta\log\rho_j(t)) \\
&- (-\beta\log\rho_i^* + \beta\log\rho_i(t))\big)\rho_j(t) \\
&+ \sum_{j\in N(i),\Psi_j<\Psi_i} \big((-\beta\log\rho_j^* + \beta\log\rho_j(t)) - (-\beta\log\rho_i^* + \beta\log\rho_i(t))\big)\rho_i(t) \\
&+ \sum_{j\in N(i),\Psi_j=\Psi_i} \beta(\frac{\rho_j(t)}{\rho_j^*} - \frac{\rho_i(t)}{\rho_i^*})\frac{\rho_i^*+\rho_j^*}{2}\Big) \\
&= \sum_{i=1}^{N} \frac{2(\rho_i(t)-\rho_i^*)}{\rho_i^*} \Big(\sum_{j\in N(i),\Psi_j>\Psi_i} \beta(\log\frac{\rho_j(t)}{\rho_j^*} - \log\frac{\rho_i(t)}{\rho_i^*})\rho_j(t) \\
&+ \sum_{j\in N(i),\Psi_j<\Psi_i} \beta(\log\frac{\rho_j(t)}{\rho_j^*} - \log\frac{\rho_i(t)}{\rho_i^*})\rho_i(t) + \sum_{j\in N(i),\Psi_j=\Psi_i} \beta(\frac{\rho_j(t)}{\rho_j^*} - \frac{\rho_i(t)}{\rho_i^*})\frac{\rho_i^*+\rho_j^*}{2}\Big)
\end{aligned}
$$

We denote $\eta_i(t)$ as $\frac{\rho_i(t)-\rho_i^*}{\rho_i^*}$ for $t \geq 0$. Then the above equation can be written as

$$
\begin{aligned}
\frac{dL(t)}{dt} = \sum_{i=1}^{N} 2\eta_i(t) \Big(&\sum_{j\in N(i),\Psi_j>\Psi_i} \beta\big(\log(1+\eta_j(t)) - \log(1+\eta_i(t))\big)\rho_j(t) \\
&+ \sum_{j\in N(i),\Psi_j<\Psi_i} \beta\big(\log(1+\eta_j(t)) - \log(1+\eta_i(t))\big)\rho_i(t) \\
&+ \sum_{j\in N(i),\Psi_j=\Psi_i} \beta(\eta_j(t)-\eta_i(t))\frac{\rho_i^*+\rho_j^*}{2}\Big).
\end{aligned}
$$

For edge $\{a_i, a_j\} \in E$ with $\Psi_j > \Psi_i$, $2\eta_i\beta(\log(1+\eta_j) - \log(1+\eta_i))\rho_j$ will be in the above sum at vertex a_i; $2\eta_j\beta(\log(1+\eta_i) - \log(1+\eta_j))\rho_j$ will be in the above sum at vertex a_j. So we can write the above equality as

$$\begin{aligned}\frac{dL(t)}{dt} &= -\sum_{\{a_i,a_j\}\in E, \Psi_j>\Psi_i} 2\beta(\log(1+\eta_j(t)) - \log(1+\eta_i(t)))(\eta_j(t)-\eta_i(t))\rho_j(t) \\ &\quad - \sum_{\{a_i,a_j\}\in E, \Psi_j=\Psi_i} 2\beta(\eta_j(t)-\eta_i(t))^2 \frac{\rho_i^*+\rho_j^*}{2}. \end{aligned} \tag{3.2}$$

Using (3.2) and the following inequality

$$min\{\frac{1}{a}, \frac{1}{b}\} \le \frac{\log a - \log b}{a-b} \le max\{\frac{1}{a}, \frac{1}{b}\}$$

for $a > 0, b > 0$ with $a \ne b$, we have

$$\begin{aligned}\frac{dL(t)}{dt} &\le -\sum_{\{a_i,a_j\}\in E, \Psi_j>\Psi_i} 2\beta(\eta_j(t)-\eta_i(t))^2 \min\{\frac{1}{1+\eta_i(t)}, \frac{1}{1+\eta_j(t)}\}\rho_j(t) \\ &\quad - \sum_{\{a_i,a_j\}\in E, \Psi_j=\Psi_i} 2\beta(\eta_j(t)-\eta_i(t))^2 \frac{\rho_i^*+\rho_j^*}{2} \\ &= -\sum_{\{a_i,a_j\}\in E, \Psi_j>\Psi_i} 2\beta(\eta_j(t)-\eta_i(t))^2 \min\{\frac{\rho_i^*}{\rho_i(t)}, \frac{\rho_j^*}{\rho_j(t)}\}\rho_j(t) \\ &\quad - \sum_{\{a_i,a_j\}\in E, \Psi_j=\Psi_i} 2\beta(\eta_j(t)-\eta_i(t))^2 \frac{\rho_i^*+\rho_j^*}{2}. \end{aligned} \tag{3.3}$$

For $\mathrm{b} = (b_i)_{i=1}^N \in \mathbb{R}^N$, we let

$$m(\mathrm{b}) = \min\{b_i : 1 \le i \le N\} \text{ and } M(\mathrm{b}) = \max\{b_i : 1 \le i \le N\}.$$

Put $A(t) = 2\beta \frac{m(\rho(t))}{M(\rho(t))} m(\rho^*)$ for $t \ge 0$. Then $A(t) > 0$ and by (3.3) we have

$$\frac{dL(t)}{dt} \le -A(t)\Big(\sum_{\{a_i,a_j\}\in E} (\eta_j(t)-\eta_i(t))^2\Big). \tag{3.4}$$

Next we use the following claims (whose proofs appear *after* the proof of the present theorem), relating the above right hand side to the spectral gap of the Laplacian matrix $\mathcal{L}(G)$ of graph G.

Claim 3.2.

$$\sum_{\{a_i,a_j\}\in E} (\eta_j(t)-\eta_i(t))^2 \ge \frac{\lambda_2}{2M(\boldsymbol{\rho}^*)} L(t),$$

where $M(\boldsymbol{\rho}^*)$ is the maximal entry of $\boldsymbol{\rho}^*$ which is less than 1, and λ_2 is the second smallest eigenvalue of the Laplacian matrix $\mathcal{L}(G)$ of G, or the spectral gap of G.

We need the following definition before stating the next claim. Let us denote

$$M = \max\{e^{2|\Psi_i|} : i = 1, 2, \cdots, N\},$$
$$\epsilon_0 = 1,$$

and

$$\epsilon_1 = \frac{1}{2}\min\left\{\frac{\epsilon_0}{(1+(2M)^{\frac{1}{\beta}})}, \min\{\rho_i^0 : i = 1, \cdots, N\}\right\}.$$

For $\ell = 2, 3, \cdots, N-1$, we let

$$\epsilon_\ell = \frac{\epsilon_{\ell-1}}{1+(2M)^{\frac{1}{\beta}}}.$$

We define

$$B = \{\, \boldsymbol{q} = (q_i)_{i=1}^N \in \mathcal{M} : \sum_{r=1}^{\ell} q_{i_r} \le 1 - \epsilon_\ell \text{ where } \ell \in \{1, \cdots, N-1\},$$
$$1 \le i_1 < \cdots < i_\ell \le N\}.$$

Then B is a compact subset of $\mathcal{M}$ with respect to the Euclidean metric, with

$$\operatorname{int}(B) = \{\, \boldsymbol{q} = (q_i)_{i=1}^N \in \mathcal{M} : \sum_{r=1}^{\ell} q_{i_r} < 1 - \epsilon_\ell, \text{ where } \ell \in \{1, \cdots, N-1\},$$
$$1 \le i_1 < \cdots < i_\ell \le N\},$$

and $\boldsymbol{\rho}^0 \in \operatorname{int}(B)$. We have

Claim 3.3. *$\rho(t) \in B$ for all $t \ge 0$.*

Using Claim 3.2 and (3.4), we have

$$\frac{dL(t)}{dt} \le -\frac{\lambda_2}{2M(\boldsymbol{\rho}^*)} A(t)\, L(t). \tag{3.5}$$

We define

$$C = \beta\lambda_2 \frac{m(\boldsymbol{\rho}^*)}{M(\boldsymbol{\rho}^*)} \frac{1-\epsilon_{L-1}}{\epsilon_1}. \tag{3.6}$$

Clearly $C > 0$ is dependent on $\boldsymbol{\rho}^o$ as well as on G, Ψ, β, that is $C = C(\boldsymbol{\rho}^o; G, \Psi, \beta)$. By the definition of B and Claim 3.3, we have

Please cite this article in press as: R. Che et al., Convergence to global equilibrium for Fokker–Planck equations on a graph and Talagrand-type inequalities, J. Differential Equations (2016), http://dx.doi.org/10.1016/j.jde.2016.05.003

$$\begin{aligned} A(t) &= 2\beta \frac{m(\boldsymbol{\rho}(t))}{M(\boldsymbol{\rho}(t))} m(\boldsymbol{\rho}^*) \geq 2\beta m(\boldsymbol{\rho}^*) \min\{\frac{m(\boldsymbol{q})}{M(\boldsymbol{q})} : \boldsymbol{q} \in B\} \\ &\geq 2\beta m(\boldsymbol{\rho}^*) \frac{1-\epsilon_{L-1}}{\epsilon_1} \\ &= \frac{2M(\boldsymbol{\rho}^*)}{\lambda_2} C \end{aligned}$$

for $t \geq 0$. Combining this with (3.5), we get $\frac{dL(t)}{dt} \leq -CL(t)$ for $t > 0$. This implies that $L(t) \leq L(0)e^{-Ct}$ for $t \geq 0$. Since $L(0) = ||\frac{\rho(0)}{\rho^*} - 1||^2_{2,\boldsymbol{\rho}^*}$, we have (3.1), completing the proof of the theorem, modulo the claim (see below). □

Remark 3.4. Given a graph $G = (V, E)$, a potential $\boldsymbol{\Psi} = (\Psi_i)_{i=1}^N$ on V and a constant $\beta > 0$, the positive constant $C = C(\boldsymbol{\rho}^o; G, \Psi, \beta)$ given by (3.6) is dependent on the initial value $\boldsymbol{\rho}^o \in \mathcal{M}$. In fact $C(\boldsymbol{\rho}^o; G, \Psi, \beta) \to 0$, when the initial distribution $\boldsymbol{\rho}^o$ converges to the boundary of $\mathcal{M}$.

We now write a simple observation yielding the first claim used in the proof of the above theorem.

Proof of Claim 3.2. Indeed we have

$$\begin{aligned} \sum_{\{a_i,a_j\}\in E} (\eta_j(t) - \eta_i(t))^2 &= \frac{1}{2}\sum_{i=1}^N \sum_{j\in N(i)} (\eta_j(t) - \eta_i(t))^2 \\ \geq \frac{1}{2M(\boldsymbol{\rho}^*)} \sum_{i=1}^N \sum_{j\in N(i)} (\eta_j(t) - \eta_i(t))^2 \rho_i^* &\geq \frac{\lambda_2}{2M(\boldsymbol{\rho}^*)} Var_{\boldsymbol{\rho}^*}(\boldsymbol{\eta}(t)) \end{aligned}$$

the last inequality comes from the Poincaré-type inequality (see for example [5]). In which

$$Var_{\boldsymbol{\rho}^*}(\boldsymbol{\eta}(t)) = \sum_{i=1}^N \rho_i^* \eta_i^2(t) - \Big(\sum_{i=1}^N \rho_i^* \eta_i(t)\Big)^2$$

Note that

$$\sum_{i=1}^N \rho_i^* \eta_i(t) = \sum_{i=1}^N \rho_i(t) - \sum_{i=1}^N \rho_i^* = 0$$

and

$$\sum_{i=1}^N \rho_i^* \eta_i^2(t) = L(t)$$

Hence we have

Please cite this article in press as: R. Che et al., Convergence to global equilibrium for Fokker–Planck equations on a graph and Talagrand-type inequalities, J. Differential Equations (2016), http://dx.doi.org/10.1016/j.jde.2016.05.003

$$\sum_{\{a_i,a_j\}\in E}(\eta_j(t)-\eta_i(t))^2 \ge \frac{\lambda_2}{2M(\boldsymbol{\rho}^*)}L(t)\,. \qquad \square$$

Finally, we present the proof of the second claim.

Proof of Claim 3.3. We follow closely the argument in the proof of Theorem 4.1 in [7]. Since $\boldsymbol{\rho}^0 \in \operatorname{int}(B)$, it is sufficient to show for any $\boldsymbol{q} \in B$, the solution $\boldsymbol{q}(t)$ through $\boldsymbol{q}$ remains in $\operatorname{int}(B)$ for small $t > 0$. Let $\boldsymbol{q} = (q_i)_{i=1}^N \in B$ and

$$\boldsymbol{q}(t) : [0, c(\boldsymbol{q})) \to \mathcal{M}$$

be the solution to the equation (1.3) with initial value $\boldsymbol{q}$ on its maximal interval of existence. In order to show $\boldsymbol{q}(t) \in \operatorname{int}(B)$ for small $t > 0$, it is sufficient to show that for any $\ell \in \{1, 2, \cdots, N-1\}$ and $1 \le i_1 < i_2 < \cdots i_\ell \le N$, one has

$$\sum_{r=1}^{\ell} q_{i_r}(t) < 1-\epsilon_\ell\,,$$

for sufficiently small $t > 0$.

Given $\ell \in \{1, 2, \cdots, N-1\}$ and $1 \le i_1 < i_2 < \cdots i_\ell \le N$, since $\boldsymbol{q} \in B$, we have

$$\sum_{r=1}^{\ell} q_{i_r} \le 1-\epsilon_\ell.$$

Then there are two cases. The first one is

$$\sum_{r=1}^{\ell} q_{i_r} < 1-\epsilon_\ell.$$

It is clear that

$$\sum_{r=1}^{\ell} q_{i_r}(t) < 1-\epsilon_\ell\,,$$

for small enough $t > 0$ by continuity.

The second case is

$$\sum_{r=1}^{\ell} q_{i_r} = 1-\epsilon_\ell.$$

Let $A = \{i_1, i_2, \cdots, i_\ell\}$ and $A^c = \{1, 2, \cdots, N\} \setminus A$. Then for any $j \in A^c$,

$$q_j \le 1-\Big(\sum_{r=1}^{\ell} q_{i_r}\Big) = \epsilon_\ell. \tag{3.7}$$

Since $\boldsymbol{q} \in B$, we have

$$\sum_{j=1}^{\ell-1} q_{s_j} \le 1 - \epsilon_{\ell-1},$$

for any $1 \le s_1 < s_2 < \cdots < s_{\ell-1} \le N$. Hence for each $i \in A$,

$$q_i \ge 1 - \epsilon_\ell - (1 - \epsilon_{\ell-1}) = \epsilon_{\ell-1} - \epsilon_\ell. \tag{3.8}$$

Combining equations (3.7), (3.8) and the fact

$$\epsilon_\ell \le \frac{\epsilon_{\ell-1}}{1 + (2M)^{\frac{1}{\beta}}},$$

one has, for any $i \in A$, $j \in A^c$,

$$\Psi_j - \Psi_i + \beta(\log q_j - \log q_i) \le \Psi_j - \Psi_i + \beta(\log \epsilon_\ell - \log(\epsilon_{\ell-1} - \epsilon_\ell)) \le -\log 2. \tag{3.9}$$

For $\{a_i, a_j\} \in E$, we set

$$C(\{a_i, a_j\}) = \begin{cases} q_j & \text{if } \Psi_i < \Psi_j \\ q_i & \text{if } \Psi_i > \Psi_j \\ \frac{q_i - q_j}{\log q_i - \log q_j} & \text{if } \Psi_i = \Psi_j \end{cases}. \tag{3.10}$$

Clearly, $C(\{a_i, a_j\}) > 0$ for $\{a_i, a_j\} \in E$. Since the graph G is connected, there exists $i_* \in A$, $j_* \in A^c$ such that $\{a_{i_*}, a_{j_*}\} \in E$. Thus

$$\sum_{i \in A, j \in A^c, \{a_i, a_j\} \in E} C(\{a_i, a_j\}) \ge C(\{a_{i_*}, a_{j_*}\}) > 0. \tag{3.11}$$

Now by (3.9) and (3.11), one has

$$\begin{aligned}
\frac{\mathrm{d}}{dt} \sum_{r=1}^{\ell} q_{i_r}(t) \,|_{t=0} &= \sum_{i \in A} \Big(\sum_{j \in N(i)} C(\{a_i, a_j\}) \big(\Psi_j - \Psi_i + \beta(\log q_j - \log q_i) \big) \Big) \\
&= \sum_{i \in A} \Big(\sum_{j \in A \cap N(i)} C(\{a_i, a_j\}) \big(\Psi_j - \Psi_i + \beta(\log q_j - \log q_i) \big) + \\
&\qquad \sum_{j \in A^c \cap N(i)} C(\{a_i, a_j\}) \big(\Psi_j - \Psi_i + \beta(\log q_j - \log q_i) \big) \Big) \\
&= \sum_{i \in A} \Big(\sum_{j \in A^c \cap N(i)} C(\{a_i, a_j\}) \big(\Psi_j - \Psi_i + \beta(\log q_j - \log q_i) \big) \Big) \\
&\le \sum_{i \in A} \Big(\sum_{j \in A^c \cap N(i)} -C(\{a_i, a_j\}) \log 2 \Big)
\end{aligned}$$

$$
\begin{aligned}
&= -\log 2\Big(\sum_{i\in A, j\in A^c, \{a_i,a_j\}\in E} C(\{a_i,a_j\})\Big) \\
&\le -C(\{a_{i_*}, a_{j_*}\}) \log 2 < 0.
\end{aligned}
$$

Combining this with the fact

$$
\sum_{r=1}^{\ell} q_{i_r} = 1 - \epsilon_\ell,
$$

it is clear that

$$
\sum_{r=1}^{\ell} q_{i_r}(t) < 1 - \epsilon_\ell ,
$$

for sufficiently small $t > 0$. This finishes the proof of Claim 3.3. □

Using the same technique, we have the following second main result.

Theorem 3.5. *Let $G = (V, E)$ be a graph with its vertex set $V = \{a_1, a_2, \cdots, a_N\}$, edge set E, a given potential $\mathbf{\Psi} = (\Psi_i)_{i=1}^N$ on V and a constant $\beta > 0$. If $\boldsymbol{\rho}(t) : [0,\infty) \to \mathcal{M}$ is the solution of Fokker–Planck equation II (1.4), with the initial value $\boldsymbol{\rho}^o = (\rho_i^o)_{i=1}^N \in \mathcal{M}$, then*

$$
\|\frac{\boldsymbol{\rho}(t)}{\boldsymbol{\rho}^*} - 1\|^2_{2,\boldsymbol{\rho}^*} \le \|\frac{\boldsymbol{\rho}^o}{\boldsymbol{\rho}^*} - 1\|^2_{2,\boldsymbol{\rho}^*}\, e^{-Ct}, \tag{3.12}
$$

where $\boldsymbol{\rho}^* = (\rho_i^*)_{i=1}^N$ is the Gibbs distribution given by (1.2) and $C = \beta\lambda_2 \frac{\min\{\rho_i^*:1\le i\le N\}}{\max\{\rho_i^*:1\le i\le N\}}$, where λ_2 is the spectral gap of G. In particular, $\boldsymbol{\rho}(t)$ exponentially converges to global equilibrium: the Gibbs distribution $\boldsymbol{\rho}^*$ under the Euclidean metric of $\mathbb{R}^N$, as $t \to \infty$.

Proof. Given initial value $\boldsymbol{\rho}^0 = (\rho_i^0)_{i=1}^N \in \mathcal{M}$. Let $\boldsymbol{\rho}(t) = (\rho_i(t))_{i=1}^N : [0,\infty) \to \mathcal{M}$ be the solution of Fokker–Planck equation II (1.4) with initial value $\boldsymbol{\rho}^0 \in \mathcal{M}$. For $t \ge 0$, we define

$$
L(t) = \|\frac{\boldsymbol{\rho}(t)}{\boldsymbol{\rho}^*} - 1\|^2_{2,\boldsymbol{\rho}^*} = \sum_{i=1}^N \frac{(\rho_i(t) - \rho_i^*)^2}{\rho_i^*},
$$

where $\boldsymbol{\rho}^* = (\rho_i^*)_{i=1}^N$ is the Gibbs distribution given by (1.2). Now for $t > 0$, by (1.4), we have

$$
\begin{aligned}
\frac{dL(t)}{dt} &= \sum_{i=1}^N \frac{2(\rho_i(t) - \rho_i^*)}{\rho_i^*} \frac{d\rho_i(t)}{dt} \\
&= \sum_{i=1}^N \frac{2(\rho_i(t) - \rho_i^*)}{\rho_i^*} \Big(\sum_{j\in N(i), \bar{\Psi}_j > \bar{\Psi}_i} \big((\Psi_j + \beta\log\rho_j(t)) - (\Psi_i + \beta\log\rho_i(t))\big)\rho_j(t)
\end{aligned}
$$

$$+ \sum_{j\in N(i),\bar{\Psi}_j<\bar{\Psi}_i} \Big(\big((\Psi_j+\beta\log\rho_j(t))-(\Psi_i+\beta\log\rho_i(t))\big)\rho_i(t)\Big)$$

$$= \sum_{i=1}^{N} \frac{2(\rho_i(t)-\rho_i^*)}{\rho_i^*}\Big(\sum_{j\in N(i),\bar{\Psi}_j>\bar{\Psi}_i} \beta(\log\frac{\rho_j(t)}{\rho_j^*}-\log\frac{\rho_i(t)}{\rho_i^*})\rho_j$$

$$+ \sum_{j\in N(i),\bar{\Psi}_j<\bar{\Psi}_i} \beta(\log\frac{\rho_j(t)}{\rho_j^*}-\log\frac{\rho_i(t)}{\rho_i^*})\rho_i\Big),$$

the last equality comes from the fact $\Psi_j-\Psi_i=-\beta\log\rho_j^*+\beta\log\rho_i^*$ for $i,j\in\{1,2,\cdots,N\}$.

Denoting $\frac{\rho_i(t)-\rho_i^*}{\rho_i^*}$ by $\eta_i(t)$, for $t>0$, the equation will be written as

$$\frac{dL(t)}{dt} = \sum_{i=1}^{N} 2\beta\eta_i(t)\Big(\sum_{j\in N(i),\bar{\Psi}_j>\bar{\Psi}_i} \big(\log(1+\eta_j(t))-\log(1+\eta_i(t))\big)\rho_j(t)$$

$$+ \sum_{j\in N(i),\bar{\Psi}_j<\bar{\Psi}_i} \big(\log(1+\eta_j(t))-\log(1+\eta_i(t))\big)\rho_i(t)\Big)$$

$$= \sum_{\{a_i,a_j\}\in E,\bar{\Psi}_j>\bar{\Psi}_j} 2\beta\big(\log(1+\eta_j(t))-\log(1+\eta_i(t))\big)\big(\eta_j(t)-\eta_i(t)\big)\rho_j(t).$$

Moreover note that $\eta_i(t)=\eta_j(t)$ when $\bar{\Psi}_i=\bar{\Psi}_j$, we have

$$\frac{dL(t)}{dt} = -\sum_{\{a_i,a_j\}\in E,\bar{\Psi}_j>\bar{\Psi}_i} 2\beta\big(\log(1+\eta_j(t))-\log(1+\eta_i(t))\big)\big(\eta_j(t)-\eta_i(t)\big)\rho_j(t)$$

$$- \sum_{\{a_i,a_j\}\in E,\bar{\Psi}_j=\bar{\Psi}_i} 2\beta\big(\eta_j(t)-\eta_i(t)\big)^2\frac{\rho_i^*+\rho_j^*}{2}. \tag{3.13}$$

Using (3.13) and the following inequality

$$min\{\frac{1}{a},\frac{1}{b}\} \le \frac{\log a-\log b}{a-b} \le max\{\frac{1}{a},\frac{1}{b}\},$$

for $a>0, b>0$ with $a\neq b$, we have

$$\frac{dL(t)}{dt} \le -\sum_{\{a_i,a_j\}\in E,\bar{\Psi}_j>\bar{\Psi}_i} 2\beta\big(\eta_j(t)-\eta_i(t)\big)^2 \min\{\frac{1}{1+\eta_i(t)},\frac{1}{1+\eta_j(t)}\}\rho_j(t)$$

$$- \sum_{\{a_i,a_j\}\in E,\bar{\Psi}_j=\bar{\Psi}_i} 2\beta\big(\eta_j(t)-\eta_i(t)\big)^2\frac{\rho_i^*+\rho_j^*}{2}$$

$$
\begin{aligned}
&= - \sum_{\{a_i,a_j\}\in E, \bar{\Psi}_j > \bar{\Psi}_i} 2\beta\big(\eta_j(t)-\eta_i(t)\big)^2 \min\{\frac{\rho_i^*}{\rho_i(t)}, \frac{\rho_j^*}{\rho_j(t)}\}\rho_j(t) \\
&\quad - \sum_{\{a_i,a_j\}\in E, \bar{\Psi}_j = \bar{\Psi}_i} 2\beta\big(\eta_j(t)-\eta_i(t)\big)^2 \frac{\rho_i^*+\rho_j^*}{2}.
\end{aligned} \tag{3.14}
$$

Now note that $\frac{\rho_i^*}{\rho_i(t)} \geq \frac{\rho_j^*}{\rho_j(t)}$ when $\bar{\Psi}_j > \bar{\Psi}_i$, hence $\min\{\frac{\rho_i^*}{\rho_i(t)}, \frac{\rho_j^*}{\rho_j(t)}\}\rho_j(t) = \rho_j^*$ when $\bar{\Psi}_j > \bar{\Psi}_i$. Combining this with (3.14), we have

$$
\begin{aligned}
\frac{dL(t)}{dt} &\leq - \sum_{\{a_i,a_j\}\in E, \bar{\Psi}_j > \bar{\Psi}_i} 2\beta\big(\eta_j(t)-\eta_i(t)\big)^2 \rho_j^* \\
&\quad - \sum_{\{a_i,a_j\}\in E, \bar{\Psi}_j = \bar{\Psi}_i} 2\beta\big(\eta_j(t)-\eta_i(t)\big)^2 \frac{\rho_i^*+\rho_j^*}{2} \\
&\leq -2\beta \min\{\rho_i^* : 1 \leq i \leq N\} \sum_{\{a_i,a_j\}\in E} (\eta_j(t)-\eta_i(t))^2.
\end{aligned} \tag{3.15}
$$

By the same argument as in Claim 3.2 (used in the proof of Theorem 3.1), we have

$$
\sum_{\{a_i,a_j\}\in E} (\eta_j(t)-\eta_i(t))^2 \geq \frac{\lambda_2}{2\max\{\rho_i^* : 1 \leq i \leq N\}} L(t).
$$

Combining this with (3.15), we get $\frac{dL(t)}{dt} \leq -CL(t)$ for $t > 0$, where $C = \beta\lambda_2 \frac{\min\{\rho_i^*:1\leq i\leq N\}}{\max\{\rho_i^*:1\leq i\leq N\}}$. This implies $L(t) \leq L(0)e^{-Ct}$ for $t \geq 0$, that is (3.12) is true, completing the proof of the theorem. $\square$

Remark 3.6. We remark that the rates of exponential convergence obtained in Theorem 3.1 and Theorem 3.5 are not sharp, as we take the uniform lower bound of the coefficients of $(\eta_j(t) - \eta_i(t))^2$. These can be improved by introducing a weighted version of Claim 3.2. More precisely, one needs an optimal constant c satisfying

$$
\sum_{\{a_i,a_j\}\in E} w_{ij}(\eta_j(t)-\eta_i(t)) \geq c\lambda_2 L(t),
$$

where w_{ij} are the edge weights determined by Equation (3.3) or (3.14).

3.2. *Exponential decay of the relative entropy*

In the following, we show that the entropy decay rate of the Fokker–Planck Equation (II) on $G = (V, E)$ with its vertex set $V = \{a_1, a_2, \cdots, a_N\}$, edge set E, a given potential $\mathbf{\Psi} = (\Psi_i)_{i=1}^N$ on V and a constant $\beta > 0$ can be bounded in terms of the modified logarithmic Sobolev constant (also known as the “entropy constant”) $\gamma_0 := \gamma_0(G)$ of the underlying graph G: the optimal $\gamma_0 > 0$ such that

$$2\gamma_0 \mathrm{Ent}(f) \le \mathcal{E}(f, \log f)\,, \tag{3.16}$$

over all $f : V \to \mathbb{R}$, with $f > 0$; recall here the standard notation for the Entropy functional and the Dirichlet form:

$$\mathrm{Ent}\, f := \mathrm{Ent}_{\boldsymbol{\rho}^*} f := \mathrm{E}_{\boldsymbol{\rho}^*}(f \log f) - (\mathrm{E}_{\boldsymbol{\rho}^*} f) \log(\mathrm{E}_{\boldsymbol{\rho}^*} f)\,,$$

and

$$\mathcal{E}(f, \log f) = \sum_{i=1}^{N} \sum_{j \in N(i)} \big(\log f(a_i) - \log f(a_j)\big)\big(f(a_i) - f(a_j)\big)\rho_i^*\,,$$

where $\boldsymbol{\rho}^* = (\rho_i^*)_{i=1}^N$ is the Gibbs distribution given by (1.2). See [5] (where this constant was denoted as ρ_0) and references therein, for more information on γ_0 of a graph and that of a Markov kernel on G.

Theorem 3.7. *Let $G = (V, E)$ be a graph, with its vertex set $V = \{a_1, a_2, \cdots, a_N\}$, edge set E, a given potential $\boldsymbol{\Psi} = (\Psi_i)_{i=1}^N$ on V and a constant $\beta > 0$. If $\boldsymbol{\rho}(t) = (\rho_i(t))_{i=1}^N : [0, \infty) \to \mathcal{M}$ is the solution of Fokker–Planck equation II (1.4) with the initial value $\boldsymbol{\rho}^o = (\rho_i^o)_{i=1}^N \in \mathcal{M}$, then*

$$H(\boldsymbol{\rho}(t)|\boldsymbol{\rho}^*) \le H(\boldsymbol{\rho}^0|\boldsymbol{\rho}^*)e^{-ct} \text{ for } t \ge 0,$$

where $\boldsymbol{\rho}^* = (\rho_i^*)_{i=1}^N$ is the Gibbs distribution given by (1.2) and $c = \beta\gamma_0 \frac{\min\{\rho_i^* : 1 \le i \le N\}}{\max\{\rho_i^* : 1 \le i \le N\}}$.

Proof. Given $\boldsymbol{\rho}^o = (\rho_i^o)_{i=1}^N \in \mathcal{M}$. Let $\boldsymbol{\rho}(t) = (\rho_i(t))_{i=1}^N : [0, \infty) \to \mathcal{M}$ be the solution of Fokker–Planck equation II (1.4) with the initial value $\boldsymbol{\rho}^0$.

Recall that the relative entropy of $\boldsymbol{\rho} = (\rho_i)_{i=1}^N \in \mathcal{M}$ with respect to $\boldsymbol{\rho}^*$:

$$H(\boldsymbol{\rho}|\boldsymbol{\rho}^*) = \sum_{i=1}^{N} \rho_i \log \frac{\rho_i}{\rho_i^*}\,.$$

Since the Gibbs distribution is given by $\rho_i^* = \dfrac{1}{K} e^{-\Psi_i/\beta}$, we also have

$$\Psi_i = -\beta \log \rho_i^* - \beta \log K\,.$$

For $t \ge 0$, let $f(t) = \boldsymbol{\rho}(t)/\boldsymbol{\rho}^*$, i.e, $f(t)(a_i) = \rho_i(t)/\rho_i^*$ for $i = 1, 2, \cdots, N$, we rewrite the relative entropy as usual:

$$\begin{aligned} \mathrm{Ent}\, f(t) := \mathrm{Ent}_{\boldsymbol{\rho}^*} f(t) &= \mathrm{E}_{\boldsymbol{\rho}^*}(f \log f) - (\mathrm{E}_{\boldsymbol{\rho}^*} f) \log(\mathrm{E}_{\boldsymbol{\rho}^*} f) \\ &= H(\boldsymbol{\rho}(t)|\boldsymbol{\rho}^*). \end{aligned}$$

Please cite this article in press as: R. Che et al., Convergence to global equilibrium for Fokker–Planck equations on a graph and Talagrand-type inequalities, J. Differential Equations (2016), http://dx.doi.org/10.1016/j.jde.2016.05.003

We write simply $f_i(t) = f(t)(a_i)$. Then the Fokker–Planck equation II (1.4) becomes

$$\frac{d\rho_i(t)}{dt} = \beta\Big(\sum_{\substack{j\in N(i),\\ f_j(t)>f_i(t)}} (\log f_j(t) - \log f_i(t))\rho_j(t) + \sum_{\substack{j\in N(i),\\ f_j(t)<f_i(t)}} (\log f_j(t) - \log f_i(t))\rho_i(t)\Big).$$

Observing that $\frac{a-b}{a} \le \log a - \log b$ when $a > b > 0$, we bound the entropy decay by proceeding as follows.

$$\begin{aligned}
&\frac{\mathrm{dEnt}(f(t))}{\mathrm{d}t} = \frac{\mathrm{d}\big(\sum_{i=1}^{N} \rho_i(t)\log\frac{\rho_i(t)}{\rho_i^*}\big)}{\mathrm{d}t} \\
&= \sum_{i=1}^{N} \frac{\mathrm{d}\rho_i(t)}{\mathrm{d}t}\log f_i(t) + \sum_{i=1}^{N} \frac{\mathrm{d}\rho_i(t)}{\mathrm{d}t} = \sum_{i=1}^{N} \frac{\mathrm{d}\rho_i(t)}{\mathrm{d}t}\log f_i(t) \\
&= \beta\sum_{i=1}^{N}\log f_i\Big(\sum_{\substack{j\in N(i),\\ f_j(t)>f_i(t)}} (\log f_j(t) - \log f_i(t))\rho_j(t) + \sum_{\substack{j\in N(i),\\ f_j(t)<f_i(t)}} (\log f_j(t) - \log f_i(t))\rho_i(t)\Big) \\
&= -\beta \sum_{\substack{\{a_i,a_j\}\in E\\ f_j(t)<f_i(t)}} (\log f_i(t) - \log f_j(t))^2\rho_i(t) \\
&\le -\beta \sum_{\substack{\{a_i,a_j\}\in E\\ f_j(t)<f_i(t)}} \big(\log f_i(t) - \log f_j(t)\big)\frac{(f_i(t)-f_j(t))}{f_i(t)}\rho_i(t) \\
&= -\beta \sum_{\substack{\{a_i,a_j\}\in E\\ f_j(t)<f_i(t)}} \big(\log f_i(t) - \log f_j(t)\big)(f_i(t)-f_j(t))\rho_i^* \\
&\le -\frac{1}{2}\frac{\min\{\rho_i^*: 1\le i\le N\}}{\max\{\rho_i^*: 1\le i\le N\}}\beta \sum_{\substack{\{a_i,a_j\}\in E\\ f_j(t)<f_i(t)}} \big(\log f_i(t) - \log f_j(t)\big)(f_i(t)-f_j(t))(\rho_i^*+\rho_j^*) \\
&= -\frac{1}{2}\frac{\min\{\rho_i^*: 1\le i\le N\}}{\max\{\rho_i^*: 1\le i\le N\}}\beta \sum_{\{a_i,a_j\}\in E} \big(\log f_i(t) - \log f_j(t)\big)(f_i(t)-f_j(t))(\rho_i^*+\rho_j^*) \\
&= -\frac{1}{2}\frac{\min\{\rho_i^*: 1\le i\le N\}}{\max\{\rho_i^*: 1\le i\le N\}}\beta\mathcal{E}(f,\log f),
\end{aligned}$$

where $\mathcal{E}(\cdot,\cdot)$ is the Dirichlet form of $G = (V, E)$ with respect to the measure ρ^* on V. Now using the definition of the modified log-Sobolev constant (3.16), we conclude that

$$\frac{\mathrm{dEnt}(f)}{\mathrm{d}t} \le -c\mathrm{Ent}(f),$$

resulting in:

Please cite this article in press as: R. Che et al., Convergence to global equilibrium for Fokker–Planck equations on a graph and Talagrand-type inequalities, J. Differential Equations (2016), http://dx.doi.org/10.1016/j.jde.2016.05.003

$$\operatorname{Ent}(f(t)) \le \operatorname{Ent}(f(0))e^{-ct} ,$$

that is, $H(\boldsymbol{\rho}(t)|\boldsymbol{\rho}^*) \le H(\boldsymbol{\rho}^0|\boldsymbol{\rho}^*)e^{-ct}$ for $t \ge 0$, where $c = \beta\gamma_0 \frac{\min\{\rho_i^*:1\le i\le N\}}{\max\{\rho_i^*:1\le i\le N\}}$. This completes the proof of Theorem. □

4. Talagrand-type inequalities

4.1. Discrete Wasserstein-type metric on $\mathcal{M}$

Consider a graph $G = (V, E)$ with $V = \{a_1, a_2, \cdots, a_N\}$. As the collection of positive probability distributions on V, space $\mathcal{M}$ is defined as in the beginning of Section 2.

The tangent space $T_{\boldsymbol{\rho}}\mathcal{M}$ at $\boldsymbol{\rho} \in \mathcal{M}$ has the form

$$T_{\boldsymbol{\rho}}\mathcal{M} = \left\{ \boldsymbol{\sigma} = (\sigma_i)_{i=1}^N \in \mathbb{R}^N | \sum_{i=1}^N \sigma_i = 0 \right\} .$$

It is clear that the standard Euclidean metric on $\mathbb{R}^N$, d, is also a Riemannian metric on $\mathcal{M}$.

Let

$$\boldsymbol{\Phi} : (\mathcal{M}, d) \to (\mathbb{R}^N, d)$$

be an arbitrary smooth map given by:

$$\boldsymbol{\Phi}(\boldsymbol{\rho}) = (\Phi_i(\boldsymbol{\rho}))_{i=1}^N, \qquad \boldsymbol{\rho} \in \mathcal{M} .$$

In the following, we will endow $\mathcal{M}$ with a metric $d_{\boldsymbol{\Phi}}$, which depends on $\boldsymbol{\Phi}$ and the structure of G.

We consider the function

$$\frac{r_1 - r_2}{\log r_1 - \log r_2}$$

and extend it to the closure of the first quadrant in the plane. Denote

$$e(r_1, r_2) = \begin{cases} \frac{r_1 - r_2}{\log r_1 - \log r_2} & \text{if } r_1 \ne r_2 \text{ and } r_1 r_2 > 0 \\ 0 & \text{if } r_1 r_2 = 0 \\ r_1 & \text{if } r_1 = r_2 \end{cases} .$$

It is easy to check that $e(r_1, r_2)$ is a continuous function on the first quadrant and satisfies $\min\{r_1, r_2\} \le e(r_1, r_2) \le \max\{r_1, r_2\}$. For simplicity, we use its original form instead of the function $e(r_1, r_2)$ in the present paper.

The equivalence relation "$\sim$" on $\mathbb{R}^N$ is defined as

$$\boldsymbol{p} \sim \boldsymbol{q} \quad \text{if and only if} \quad p_1 - q_1 = p_2 - q_2 = \cdots = p_N - q_N ,$$

and let $\mathcal{W}$ be the quotient space $\mathbb{R}^N/\sim$. In other words, for $\boldsymbol{p} \in \mathbb{R}^N$ we consider its equivalent class

$$[\boldsymbol{p}] = \{(p_1 + c, p_2 + c, \cdots, p_N + c) : c \in \mathbb{R}\},$$

and all such equivalent classes form the vector space $\mathcal{W}$.

For a given Φ, and $[\boldsymbol{p}] = [(p_i)_{i=1}^N] \in \mathcal{W}$, we define an identification $\tau_{\boldsymbol{\Phi}}([\boldsymbol{p}]) = \boldsymbol{\sigma}$ from $\mathcal{W}$ to $T_{\boldsymbol{\rho}}\mathcal{M}$ by,

$$\boldsymbol{\sigma} = \boldsymbol{p}\mathcal{L}(G, w(\boldsymbol{\rho})), \tag{4.1}$$

where $w(\boldsymbol{\rho}) = \{w_{ij}(\boldsymbol{\rho})\}_{i,j=1}^N$ is the weight associated to the original graph G:

$$w_{ij}(\boldsymbol{\rho}) = \begin{cases} \rho_i & \text{if } \ \Phi_i(\boldsymbol{\rho}) > \Phi_j(\boldsymbol{\rho}), \{a_i, a_j\} \in E \\ \rho_j & \text{if } \ \Phi_i(\boldsymbol{\rho}) < \Phi_j(\boldsymbol{\rho}), \{a_i, a_j\} \in E \\ \frac{\rho_i - \rho_j}{\log \rho_i - \log \rho_j} & \text{if } \ \Phi_i(\boldsymbol{\rho}) = \Phi_j(\boldsymbol{\rho}), \{a_i, a_j\} \in E \\ 0 & \quad \text{otherwise} \end{cases}$$

and $\mathcal{L}(G, w(\boldsymbol{\rho}))$ is the weighted Laplacian matrix of weighted graph $(G, w(\boldsymbol{\rho}))$. It is not hard to check that $\tau_{\boldsymbol{\Phi}}([\boldsymbol{p}])$ is a linear isomorphism between $T_{\boldsymbol{\rho}}\mathcal{M}$ and $\mathcal{W}$ (see Lemma 2 in [7]). Hence $\boldsymbol{\sigma} \in T_{\boldsymbol{\rho}}\mathcal{M}$ can be rewritten as equivalent classes on $\mathbb{R}^N$. For simplicity, we say $\boldsymbol{\sigma} \simeq [\boldsymbol{p}] = [(p_i)_{i=1}^N]$ if $[\boldsymbol{p}] := \tau_{\boldsymbol{\Phi}}^{-1}(\boldsymbol{\sigma}) \in \mathcal{W}$.

We note that this identification (4.1) depends on $\boldsymbol{\Phi}$, the probability distribution $\boldsymbol{\rho}$ and the structure of the graph G.

Definition 4.1. By the above identification (4.1), we define an inner product on $T_{\boldsymbol{\rho}}\mathcal{M}$ by:

$$g_{\boldsymbol{\rho}}^{\boldsymbol{\Phi}}(\boldsymbol{\sigma}^1, \boldsymbol{\sigma}^2) = \sum_{i=1}^N p_i^1 \sigma_i^2 = \sum_{i=1}^N p_i^2 \sigma_i^1.$$

It is easy to check that this definition is equivalent to

$$g_{\boldsymbol{\rho}}^{\boldsymbol{\Phi}}(\boldsymbol{\sigma}^1, \boldsymbol{\sigma}^2) = \boldsymbol{p}^1 \mathcal{L}(G, w(\boldsymbol{\rho}))(\boldsymbol{p}^2)^T,$$

for $\boldsymbol{\sigma}^1 = (\sigma_i^1)_{i=1}^N$, $\boldsymbol{\sigma}^2 = (\sigma_i^2)_{i=1}^N \in T_\rho\mathcal{M}$, and $[\boldsymbol{p}^1], [\boldsymbol{p}^2] \in \mathcal{W}$ satisfying

$$\boldsymbol{\sigma}^1 \simeq [\boldsymbol{p}^1] \text{ and } \boldsymbol{\sigma}^2 \simeq [\boldsymbol{p}^2].$$

In particular,

$$g_{\boldsymbol{\rho}}^{\boldsymbol{\Phi}}(\boldsymbol{\sigma}, \boldsymbol{\sigma}) = \boldsymbol{p}\mathcal{L}(G, w(\boldsymbol{\rho}))\boldsymbol{p}^T \tag{4.2}$$

for $\boldsymbol{\sigma} \in T_\rho\mathcal{M}$, where $\boldsymbol{\sigma} \simeq [\boldsymbol{p}]$.

The associated distance $d_{\mathbf{\Phi}}(\cdot,\cdot)$ on $\mathcal{M}$ is given by

$$d_{\mathbf{\Phi}}(\boldsymbol{\rho}^1,\boldsymbol{\rho}^2)=\inf_{\gamma} L(\gamma(t))$$

where $\gamma:[0,1]\to\mathcal{M}$ ranges over all continuously differentiable curve with $\gamma(0)=\boldsymbol{\rho}^1$, $\gamma(1)=\boldsymbol{\rho}^2$. The arc length of γ is given by

$$L(\gamma(t))=\int_0^1 \sqrt{g^{\mathbf{\Phi}}_{\gamma(t)}(\dot\gamma(t),\dot\gamma(t))}dt. \tag{4.3}$$

This gives the metric space $(\mathcal{M},d_{\mathbf{\Phi}})$.

Next, we show that the metric $d_{\mathbf{\Phi}}$ is lower bounded. Given $\boldsymbol{\rho}=(\rho_i)_{i=1}^N\in\mathcal{M}$. We consider a new identifications

$$\boldsymbol{\sigma}=\boldsymbol{p}\mathcal{L}(G,w^m(\boldsymbol{\rho})), \tag{4.4}$$

where $w^m(\boldsymbol{\rho})=\{w^m_{ij}(\boldsymbol{\rho})\}_{i,j=1}^N$ is the weight associated to the original graph G:

$$w^m_{ij}(\boldsymbol{\rho})=\begin{cases}\max\{\rho_i,\rho_j\} & \text{if } \{a_i,a_j\}\in E\\ 0 & \text{otherwise}\end{cases}.$$

and $\mathcal{L}(G,w^m(\boldsymbol{\rho}))$ is the weighted Laplacian matrix of weighted graph $(G,w^m(\boldsymbol{\rho}))$. Similar to the identification (4.1), identifications (4.4) are linear isomorphisms between $T_{\boldsymbol{\rho}}\mathcal{M}$ and $\mathcal{W}$.

Furthermore, they induce inner product $g^m_{\boldsymbol{\rho}}(\cdot,\cdot)$ on $T_{\boldsymbol{\rho}}\mathcal{M}$. It is not hard to see that the map $\boldsymbol{\rho}\mapsto g^m_{\boldsymbol{\rho}}$ is smooth. By using the inner products $g^m_{\boldsymbol{\rho}}$, we can obtain distance $d_m(\cdot,\cdot)$ on $\mathcal{M}$. Then $(\mathcal{M},d_m)$ is smooth Riemannian manifold. It is shown in [7, Lemma 3.4] that for any smooth map $\mathbf{\Phi}:(\mathcal{M},d)\to(\mathbb{R}^N,d)$ and $\boldsymbol{\rho}^1,\boldsymbol{\rho}^2\in\mathcal{M}$,

$$d_m(\boldsymbol{\rho}^1,\boldsymbol{\rho}^2)\le d_{\mathbf{\Phi}}(\boldsymbol{\rho}^1,\boldsymbol{\rho}^2). \tag{4.5}$$

Now we consider two choices of the function $\mathbf{\Phi}$ which are related to Fokker–Planck equation I (1.3) and II (1.4) respectively. Let the potential $\mathbf{\Psi}=(\Psi_i)_{i=1}^N$ on V be given and $\beta>0$, where Ψ_i is the potential on vertex a_i.

It then follows from [7, Section 4] that Fokker–Planck equation I (1.3) is the gradient flow of free energy $F(\boldsymbol{\rho})$ on the Riemannian manifold $(\mathcal{M},d_{\mathbf{\Psi}})$, i.e. let $\mathbf{\Phi}(\boldsymbol{\rho})\equiv\mathbf{\Psi}$ where $\boldsymbol{\rho}\in\mathcal{M}$. Fokker–Planck equation II (1.4) is related to inner product $g^{\bar{\mathbf{\Psi}}}_{\boldsymbol{\rho}}$, where the new potential $\bar{\mathbf{\Psi}}(\boldsymbol{\rho})=(\bar\Psi_i(\boldsymbol{\rho}))_{i=1}^N$ is defined by

$$\bar\Psi_i(\boldsymbol{\rho})=\Psi_i+\beta\log\rho_i.$$

Since $g^{\bar{\mathbf{\Psi}}}_{\boldsymbol{\rho}}$ is a piecewise smooth function with respect to $\boldsymbol{\rho}$, the space $(\mathcal{M},d_{\bar{\mathbf{\Psi}}})$ is a union of finitely many smooth Riemannian manifolds. Fokker–Planck equation II (1.4) can also be seen as the generalized gradient flow of $F(\boldsymbol{\rho})$ on the metric space $(\mathcal{M},d_{\bar{\mathbf{\Psi}}})$ (see [7, Section 5]). By (4.5), $d_{\mathbf{\Psi}}$ and $d_{\bar{\mathbf{\Psi}}}$ are lower bounded by d_m.

4.2. Talagrand-type inequalities

We are now ready to prove the Talagrand-type inequalities.

Theorem 4.2. *Let $G = (V, E)$ be a graph with its vertex set $V = \{a_1, a_2, \cdots, a_N\}$ and edge set E. For each $\boldsymbol{\mu} = (\mu_i)_{i=1}^N \in \mathcal{M}$ and a compact subset B of $\mathcal{M}$ with respect to the Euclidean metric, there exist a potential function $\Psi = (\Psi_i)_{i=1}^N$ on V and a constant $K = K(B, \boldsymbol{\mu}, G) > 0$ such that for any $\boldsymbol{\nu} = (\nu_i)_{i=1}^N \in B$, we have the following Talagrand-type inequality*

$$d_\Psi^2(\boldsymbol{\mu}, \boldsymbol{\nu}) \le K H(\boldsymbol{\nu}|\boldsymbol{\mu}),$$

where $H(\boldsymbol{\nu}|\boldsymbol{\mu}) = \sum_{i=1}^N \nu_i \log \frac{\nu_i}{\mu_i}$.

Proof. Given $\boldsymbol{\mu} = (\mu_i)_{i=1}^N \in \mathcal{M}$ and a compact subset B of $\mathcal{M}$ with respect to the Euclidean metric. Let $\Psi_i = -\log \mu_i$ for $i = 1, 2, \cdots, N$ and $\beta = 1$. Then

$$F(\boldsymbol{\nu}) = H(\boldsymbol{\nu}|\boldsymbol{\mu}) \ge 0$$

for any $\boldsymbol{\nu} \in \mathcal{M}$. In the following, we are going to show that there is a constant $K = K(B, \boldsymbol{\mu}, G) > 0$ such that

$$d_\Psi^2(\boldsymbol{\nu}, \boldsymbol{\mu}) \le K H(\boldsymbol{\nu}|\boldsymbol{\mu}) = K F(\boldsymbol{\nu})$$

for any $\boldsymbol{\nu} \in B$.

Firstly using (4.1) and (4.2), for $\boldsymbol{\sigma} \in T_{\boldsymbol{\rho}}\mathcal{M}$ we have

$$||\boldsymbol{\sigma}||^2 = \boldsymbol{p}\mathcal{L}(G, w(\boldsymbol{\rho}))\mathcal{L}(G, w(\boldsymbol{\rho}))^T \boldsymbol{p}^T$$

and

$$g_{\boldsymbol{\rho}}^{\boldsymbol{\Phi}}(\boldsymbol{\sigma}, \boldsymbol{\sigma}) = \boldsymbol{p}\mathcal{L}(G, w(\boldsymbol{\rho}))\boldsymbol{p}^T,$$

where $\boldsymbol{\sigma} \simeq [\boldsymbol{p}]$ and $|| \cdot ||$ is the standard Euclidean norm on $\mathbb{R}^N$.

Since $\mathcal{L}(G, w(\boldsymbol{\rho}))$ is a real symmetric matrix, we decompose $\mathcal{L}(G, w(\boldsymbol{\rho}))$ into

$$\mathcal{L}(G, w(\boldsymbol{\rho})) = Q\Lambda Q^T$$

where Λ is a diagonal matrix whose diagonal entries are eigenvalues of $\mathcal{L}(G, w(\boldsymbol{\rho}))$, and Q is a real orthogonal matrix.

Let $\boldsymbol{w} = \boldsymbol{p}Q$, then we have

$$\begin{aligned} &g_{\boldsymbol{\rho}}^{\Psi}(\boldsymbol{\sigma}, \boldsymbol{\sigma}) = \boldsymbol{p}\mathcal{L}(G, w(\boldsymbol{\rho}))\boldsymbol{p}^T = \boldsymbol{w}\Lambda\boldsymbol{w}^T \text{ and} \\ &||\boldsymbol{\sigma}||^2 = \boldsymbol{p}\mathcal{L}(G, w(\boldsymbol{\rho}))\mathcal{L}(G, w(\boldsymbol{\rho}))^T \boldsymbol{p}^T = \boldsymbol{w}\Lambda^2\boldsymbol{w}^T. \end{aligned} \tag{4.6}$$

Please cite this article in press as: R. Che et al., Convergence to global equilibrium for Fokker–Planck equations on a graph and Talagrand-type inequalities, J. Differential Equations (2016), http://dx.doi.org/10.1016/j.jde.2016.05.003

Denote $\lambda_2(\boldsymbol{\rho})$ and $\lambda_N(\boldsymbol{\rho})$ the second smallest eigenvalue and largest eigenvalue of $\mathcal{L}(G, w(\boldsymbol{\rho}))$ respectively. Since $\mathcal{L}(G, w(\boldsymbol{\rho}))$ has one 0 eigenvalue and $N-1$ positive eigenvalues, it is not hard to see

$$\frac{1}{\lambda_N(\boldsymbol{\rho})}||\boldsymbol{\sigma}||^2 \le g^{\Psi}_{\boldsymbol{\rho}}(\boldsymbol{\sigma}, \boldsymbol{\sigma}) \le \frac{1}{\lambda_2(\boldsymbol{\rho})}||\boldsymbol{\sigma}||^2$$

by (4.6).

Let us denote

$$M = \max\{e^{2|\Psi_i|} : i = 1, 2, \cdots, N\},$$
$$\epsilon_0 = 1,$$

and

$$\epsilon_1 = \frac{1}{2}\min\left\{\frac{\epsilon_0}{(1+(2M)^{\frac{1}{\beta}})}, \min_{(\rho_i)_{i=1}^N \in B} \min\{\rho_i : i = 1, \cdots, N\}\right\},$$

where $\epsilon_1 > 0$ as B is compact. For $\ell = 2, 3, \cdots, N-1$, we let

$$\epsilon_\ell = \frac{\epsilon_{\ell-1}}{1+(2M)^{\frac{1}{\beta}}}.$$

We define

$$D = \{\, \boldsymbol{q} = (q_i)_{i=1}^N \in \mathcal{M} : \sum_{r=1}^{\ell} q_{i_r} \le 1 - \epsilon_\ell \text{ where } \ell \in \{1, \cdots, N-1\},$$
$$1 \le i_1 < \cdots < i_\ell \le N\}.$$

Then D is a compact subset of $\mathcal{M}$ with respect to the Euclidean metric and with

$$\mathrm{int}(D) = \{\, \boldsymbol{q} = (q_i)_{i=1}^N \in \mathcal{M} : \sum_{r=1}^{\ell} q_{i_r} < 1 - \epsilon_\ell, \text{ where } \ell \in \{1, \cdots, N-1\},$$
$$1 \le i_1 < \cdots < i_\ell \le N\}.$$

and $B \subset \mathrm{int}(D)$.

Let

$$C_1 = \max_{\rho \in D}\{\frac{1}{\lambda_2(\boldsymbol{\rho})}\} \text{ and } C_2 = \min_{\rho \in D}\{\frac{1}{\lambda_N(\boldsymbol{\rho})}\}.$$

Since $\lambda_2, \lambda_N : \mathcal{M} \mapsto (0, +\infty)$ are continuous and D is compact with respect to the Euclidean metric on $\mathcal{M}$, we have $0 < C_2 \le C_1 < +\infty$. It is clear that C_1, C_2 depend only on B, $\boldsymbol{\mu}$ and G, and

Please cite this article in press as: R. Che et al., Convergence to global equilibrium for Fokker–Planck equations on a graph and Talagrand-type inequalities, J. Differential Equations (2016), http://dx.doi.org/10.1016/j.jde.2016.05.003

$$C_2||\boldsymbol{\sigma}||^2 \le g^{\Psi}_{\boldsymbol{\rho}}(\boldsymbol{\sigma},\boldsymbol{\sigma}) \le C_1||\boldsymbol{\sigma}||^2 \tag{4.7}$$

for any $\boldsymbol{\rho} \in D, \boldsymbol{\sigma} \in T_{\boldsymbol{\rho}}\mathcal{M}$.

Now for $\boldsymbol{\nu} \in B$, let $\boldsymbol{\rho}(t) = (\rho_i(t))_{i=1}^N : [0, +\infty) \to \mathcal{M}$ is the solution of the Fokker–Planck Equation (1.3) for $\beta = 1$:

$$\begin{aligned}\frac{d\rho_i}{dt} = &\sum_{j\in N(i),\Psi_j>\Psi_i}((\Psi_j + \log\rho_j) - (\Psi_i + \log\rho_i))\rho_j \\ &+ \sum_{j\in N(i),\Psi_j<\Psi_i}((\Psi_j + \log\rho_j) - (\Psi_i + \log\rho_i))\rho_i \\ &+ \sum_{j\in N(i),\Psi_j=\Psi_i}(\rho_j - \rho_i)\end{aligned}$$

with initial value $\boldsymbol{\nu}$, that is, $\boldsymbol{\rho}(0) = \boldsymbol{\nu}$. Since $\boldsymbol{\nu} \in \mathrm{int}(D)$, we have $\boldsymbol{\rho}(t) \in D$ for all $t \ge 0$, which proof is similar to the proof of Claim 3.3.

Since the Gibbs distribution given by (1.2) is $\boldsymbol{\mu}$, Theorem 3.1 and (3.6) imply that there exists a constant $C = C(\boldsymbol{\mu}, B, G) > 0$ such that

$$\sum_{i=1}^N \frac{(\rho_i(t)-\mu_i)^2}{\mu_i} \le (\sum_{i=1}^N \frac{(\nu_i-\mu_i)^2}{\mu_i})e^{-Ct} \text{ for all } t \ge 0.$$

Moreover let $m = \min\{\mu_i : 1 \le i \le N\}$ and $M = \max\{\mu_i : 1 \le i \le N\}$, then

$$||\boldsymbol{\rho}(t) - \boldsymbol{\mu}||^2 \le \frac{M}{m}||\boldsymbol{\mu} - \boldsymbol{\nu}||^2 e^{-Ct} \text{ for all } t \ge 0.$$

Set $T = \frac{1}{C}\log(\frac{4M}{m})$. One obtains

$$||\boldsymbol{\rho}(T) - \boldsymbol{\mu}||^2 \le \frac{1}{4}||\boldsymbol{\mu} - \boldsymbol{\nu}||^2 \le \frac{1}{2}(||\boldsymbol{\mu} - \boldsymbol{\rho}(T)||^2 + ||\boldsymbol{\rho}(T) - \boldsymbol{\nu}||^2),$$

which implies

$$||\boldsymbol{\rho}(T) - \boldsymbol{\nu}||^2 \ge ||\boldsymbol{\rho}(T) - \boldsymbol{\mu}||^2.$$

So after time T, the solution of equation (1.3) traveled at least half of the Euclidean distance from $\boldsymbol{\nu}$ to $\boldsymbol{\mu}$.

Moreover since the Fokker–Planck equation (1.3) is the gradient flow of free energy F under the metric $d_\Psi(\cdot,\cdot)$ (see Equation (31) and Theorem 2 in [7]), we have

$$\frac{\mathrm{d}F(\boldsymbol{\rho}(t))}{\mathrm{d}t} = -g^{\Psi}_{\boldsymbol{\rho}(t)}(\dot{\boldsymbol{\rho}}(t), \dot{\boldsymbol{\rho}}(t))$$

for $t > 0$. By integrating the previous equality from 0 to T, we have

Please cite this article in press as: R. Che et al., Convergence to global equilibrium for Fokker–Planck equations on a graph and Talagrand-type inequalities, J. Differential Equations (2016), http://dx.doi.org/10.1016/j.jde.2016.05.003

$$F(\boldsymbol{\nu}) - F(\boldsymbol{\rho}(T)) = \int_0^T g^{\Psi}_{\boldsymbol{\rho}(t)}(\dot{\boldsymbol{\rho}}(t), \dot{\boldsymbol{\rho}}(t))\mathrm{d}t \geq \frac{1}{T}(\int_0^T \sqrt{g^{\Psi}_{\boldsymbol{\rho}(t)}(\dot{\boldsymbol{\rho}}(t), \dot{\boldsymbol{\rho}}(t))}\mathrm{d}t)^2$$
$$\geq \frac{1}{T}(\int_0^T \sqrt{C_2}||\dot{\boldsymbol{\rho}}(t)||\mathrm{d}t)^2 \geq \frac{C_2}{T}||\boldsymbol{\nu} - \boldsymbol{\rho}(T)||^2.$$

the last second inequality comes from (4.7) and the fact that $\boldsymbol{\rho}(t) \in D$. At the same time,

$$\begin{aligned} F(\boldsymbol{\nu}) - F(\boldsymbol{\rho}(T)) &= \int_0^T g^{\Psi}_{\boldsymbol{\rho}(t)}(\dot{\boldsymbol{\rho}}(t), \dot{\boldsymbol{\rho}}(t))\mathrm{d}t \\ &\geq \frac{1}{T}(\int_0^T \sqrt{g^{\Psi}_{\boldsymbol{\rho}(t)}(\dot{\boldsymbol{\rho}}(t), \dot{\boldsymbol{\rho}}(t))}\mathrm{d}t)^2 \\ &\geq \frac{1}{T}d^2_{\Psi}(\boldsymbol{\nu}, \boldsymbol{\rho}(T)). \end{aligned}$$

Let $\boldsymbol{s}(t) = t\boldsymbol{\rho}(T) + (1-t)\boldsymbol{\mu}$ for $t \in [0, 1]$. Since D is a convex subset of $\mathbb{R}^N$ and $\boldsymbol{\rho}(T), \boldsymbol{\mu} \in D$. We have $\boldsymbol{s}(t) \in D$ for $t \in [0, 1]$. Thus

$$\begin{aligned} d^2_{\Psi}(\boldsymbol{\rho}(T), \boldsymbol{\mu}) &\leq (\int_0^1 \sqrt{g^{\Psi}_{\boldsymbol{s}(t)}(\boldsymbol{\rho}(T) - \boldsymbol{\mu}, \boldsymbol{\rho}(T) - \boldsymbol{\mu})}\mathrm{d}t)^2 \\ &\leq (\int_0^1 \sqrt{C_1||\boldsymbol{\rho}(T) - \boldsymbol{\mu}||^2}\mathrm{d}t)^2 \\ &= C_1||\boldsymbol{\rho}(T) - \boldsymbol{\mu}||^2 \end{aligned}$$

the last second inequality comes from (4.7) and the fact that $\boldsymbol{s}(t) \in D$.

This gives us the bounds

$$\begin{aligned} d^2_{\Psi}(\boldsymbol{\rho}(T), \boldsymbol{\mu}) &\leq C_1||\boldsymbol{\rho}(T) - \boldsymbol{\mu}||^2 \leq C_1||\boldsymbol{\rho}(T) - \boldsymbol{\nu}||^2 \\ &\leq \frac{TC_1}{C_2}(F(\boldsymbol{\nu}) - F(\boldsymbol{\rho}(T))) \leq \frac{TC_1}{C_2}F(\boldsymbol{\nu}), \end{aligned}$$

and

$$d^2_{\Psi}(\boldsymbol{\nu}, \boldsymbol{\rho}(T)) \leq T(F(\boldsymbol{\nu}) - F(\boldsymbol{\rho}(T))) \leq TF(\boldsymbol{\nu}).$$

Please cite this article in press as: R. Che et al., Convergence to global equilibrium for Fokker–Planck equations on a graph and Talagrand-type inequalities, J. Differential Equations (2016), http://dx.doi.org/10.1016/j.jde.2016.05.003

In conclusion,

$$d_\Psi^2(\boldsymbol{\mu},\boldsymbol{\nu}) \le 2d_\Psi^2(\boldsymbol{\mu},\boldsymbol{\rho}(T)) + 2d_\Psi^2(\boldsymbol{\rho}(T),\boldsymbol{\nu})$$
$$\le (\frac{2TC_1}{C_2} + 2T)F(\boldsymbol{\nu}) = KH(\boldsymbol{\nu}|\boldsymbol{\mu}),$$

where $K = (\frac{2TC_1}{C_2} + 2T)$ is a parameter which only depends on B, G and $\boldsymbol{\mu}$. □

The other Talagrand-type inequality is for the "lower bound" metric $d_m(\cdot,\cdot)$.

Theorem 4.3. *Let $G=(V,E)$ be a graph with its vertex set $V=\{a_1,a_2,\cdots,a_N\}$ and edge set E. Let D be the maximal degree of G and λ_2 be the spectral gap of G. Given $\boldsymbol{\mu}=(\mu_i)_{i=1}^N \in \mathcal{M}$. Let $m=\min\{\mu_i : 1\le i\le N\}$ and $M=\max\{\mu_i : 1\le i\le N\}$. Then for any $\boldsymbol{\nu}=(\nu_i)_{i=1}^N \in \mathcal{M}$, we have the following Talagrand-type inequality*

$$d_m^2(\boldsymbol{\nu},\boldsymbol{\mu}) \le KH(\boldsymbol{\nu}|\boldsymbol{\mu})$$

where $K=\frac{M(DN^3+4)}{2\lambda_2 m}\log(\frac{18M}{m^3})$ and $d_m(\cdot,\cdot)$ is the lower bounded metric defined in Section 4.1.

Proof. Let $\beta=1$, $\Psi_i=-\log\mu_i$ and $\bar{\Psi}_i(\boldsymbol{\rho})=-\log\mu_i+\log\rho_i$ for $i=1,2,\cdots,N$. Then $F(\boldsymbol{\rho})=H(\boldsymbol{\rho}|\boldsymbol{\mu})$ for $\boldsymbol{\rho}\in\mathcal{M}$. In the following, we are going to show that there is a constant $K=K(\boldsymbol{\mu},G)>0$ such that

$$d_m^2(\boldsymbol{\nu},\boldsymbol{\mu}) \le KH(\boldsymbol{\nu}|\boldsymbol{\mu}) = KF(\boldsymbol{\nu})$$

for any $\boldsymbol{\nu}\in B$.

Now for $\boldsymbol{\nu}\in B$, let $\boldsymbol{\rho}(t)=(\rho_i(t))_{i=1}^N : [0,+\infty)\to\mathcal{M}$ is the solution of the Fokker–Planck equation (II) (1.4) with $\beta=1$,

$$\frac{d\rho_i}{dt} = \sum_{j\in N(i),\bar{\Psi}_j>\bar{\Psi}_i} ((\Psi_j+\log\rho_j)-(\Psi_i+\log\rho_i))\rho_j$$
$$+ \sum_{j\in N(i),\bar{\Psi}_j<\bar{\Psi}_i} ((\Psi_j+\log\rho_j)-(\Psi_i+\log\rho_i))\rho_i,$$

with initial value $\boldsymbol{\nu}$, that is, $\boldsymbol{\rho}(0)=\boldsymbol{\nu}$.

Let $m=\min\{\mu_i : 1\le i\le N\}$, $M=\max\{\mu_i : 1\le i\le N\}$. Since the Gibbs distribution given by (1.2) is $\boldsymbol{\mu}$, Theorem 3.5 implies

$$\sum_{i=1}^N \frac{(\rho_i(t)-\mu_i)^2}{\mu_i} \le (\sum_{i=1}^N \frac{(\nu_i-\mu_i)^2}{\mu_i})e^{-\lambda_2\frac{m}{M}t} \text{ for all } t\ge 0,$$

where λ_2 is the spectral gap of G. Moreover,

$$||\boldsymbol{\rho}(t)-\boldsymbol{\mu}||^2 \le \frac{M}{m}||\boldsymbol{\mu}-\boldsymbol{\nu}||^2 e^{-\lambda_2\frac{m}{M}t} \text{ for all } t\ge 0.$$

Set $T = \frac{M}{\lambda_2 m} \log(\frac{18M}{m^3})$. One obtains

$$||\boldsymbol{\rho}(T) - \boldsymbol{\mu}||^2 \leq \frac{m^2}{18}||\boldsymbol{\mu} - \boldsymbol{\nu}||^2 \leq \frac{m^2}{9}(||\boldsymbol{\mu} - \boldsymbol{\rho}(T)||^2 + ||\boldsymbol{\rho}(T) - \boldsymbol{\nu}||^2),$$

which implies

$$||\boldsymbol{\rho}(T) - \boldsymbol{\mu}||^2 \leq \frac{1}{8}m^2||\boldsymbol{\rho}(T) - \boldsymbol{\nu}||^2 \leq \frac{m^2}{4}$$

as $m \leq \frac{1}{N} < 1$. Thus $\boldsymbol{\rho}(T) \in N(\boldsymbol{\mu})$, where

$$N(\boldsymbol{\mu}) = \{\boldsymbol{\rho} = (\rho_i)_{i=1}^N \in \mathcal{M} : \ |\rho_i - \mu_i| \leq \frac{m}{2}, \text{ for } i = 1, 2, \cdots, N\}$$

is a compact convex subset of $\mathcal{M}$ with respect to Euclidean metric. In other words, after time T, the solution of (1.4) travels at least half of the distance from $\boldsymbol{\nu}$ to $\boldsymbol{\mu}$ and enters into the neighborhood $N(\boldsymbol{\mu})$. Then we can use the exactly same method as in Theorem 4.2 to estimate $d_m^2(\cdot, \cdot)$.

Denote $\lambda_2(\boldsymbol{\rho})$ and $\lambda_N(\boldsymbol{\rho})$ the second smallest eigenvalue and largest eigenvalue of $\mathcal{L}(G, w^m(\boldsymbol{\rho}))$ respectively. Similar to the proof of Theorem 4.2, we can prove

$$\frac{1}{\lambda_N(\boldsymbol{\rho})}||\boldsymbol{\sigma}||^2 \leq g_{\boldsymbol{\rho}}^m(\boldsymbol{\sigma}, \boldsymbol{\sigma}) \leq \frac{1}{\lambda_2(\boldsymbol{\rho})}||\boldsymbol{\sigma}||^2 \tag{4.8}$$

for any $\boldsymbol{\sigma} \in T_{\boldsymbol{\rho}}\mathcal{M}$.

For $\boldsymbol{\rho} \in \mathcal{M}$, let $\bar{\delta}(\boldsymbol{\rho})$ be the maximal of the diagonal elements in the Laplacian matrix $\mathcal{L}(G, w^m(\boldsymbol{\rho}))$ and let

$$i_{w^m(\boldsymbol{\rho})}(G) = \min_{X \subset V, |X| \leq N/2} (\sum_{i \in X, j \notin X} w_{ij}^m(\boldsymbol{\rho})/|X|),$$

where the minimum is taken over all nonempty subsets X of V satisfying $|X| \leq \frac{N}{2}$. We shall refer to $i_{w^m(\boldsymbol{\rho})}(G)$ as *the isoperimetric number* of the weighted graph $(G, w^m(\boldsymbol{\rho}))$. Since G is connected and $w_{ij}^m(\boldsymbol{\rho}) \geq \min\{\rho_i : 1 \leq i \leq N\}$ for $\{a_i, a_j\} \in E$, it is not hard to see that

$$i_{w^m(\boldsymbol{\rho})}(G) \geq \frac{2 \min\{\rho_i : 1 \leq i \leq N\}}{N}. \tag{4.9}$$

It follows from Theorem 2.2 in [2] that the spectral gap $\lambda_2(\boldsymbol{\rho})$ of the weighted graph $(G, w^m(\boldsymbol{\rho}))$ satisfies

$$\lambda_2(\boldsymbol{\rho}) \geq \bar{\delta}(\boldsymbol{\rho}) - \sqrt{\bar{\delta}(\boldsymbol{\rho})^2 - i_{w^m(\boldsymbol{\rho})}(G)^2}.$$

It then follows from inequality

$$\bar{\delta}(\boldsymbol{\rho}) - \sqrt{\bar{\delta}(\boldsymbol{\rho})^2 - i_{w^m(\boldsymbol{\rho})}(G)^2} \geq \frac{i_{w^m(\boldsymbol{\rho})}(G)^2}{2\bar{\delta}(\boldsymbol{\rho})}$$

that

$$\lambda_2(\boldsymbol{\rho}) \geq \frac{2(\min\{\rho_i : 1 \leq i \leq N\})^2}{DN^2} \tag{4.10}$$

by (4.9) and the fact that $\bar{\delta}(\boldsymbol{\rho}) \leq D$.

Let $C_1 = \max\{\frac{1}{\lambda_2(\boldsymbol{\rho})} : \boldsymbol{\rho} \in N(\boldsymbol{\mu})\}$. Note that $\min\{\rho_i : 1 \leq i \leq N\} \geq \frac{m}{2}$ for all $\boldsymbol{\rho} \in N(\boldsymbol{\mu})$. We have $C_1 \leq \frac{2DN^2}{m^2}$ by (4.10). It is well known that $\lambda_N(\boldsymbol{\rho}) \leq N$ for $\boldsymbol{\rho} \in \mathcal{M}$ (see for example [23]). Let $C_2 = \inf\{\frac{1}{\lambda_N(\boldsymbol{\rho})} : \boldsymbol{\rho} \in \mathcal{M}\}$. Then $C_2 \geq \frac{1}{N}$. Now by (4.8), we have

$$C_2||\boldsymbol{\sigma}||^2 \leq g^m_{\boldsymbol{\rho}}(\boldsymbol{\sigma}, \boldsymbol{\sigma}) \tag{4.11}$$

for all $\boldsymbol{\sigma} \in T_{\boldsymbol{\rho}}\mathcal{M}$, $\boldsymbol{\rho} \in \mathcal{M}$ and

$$g^m_{\boldsymbol{\rho}}(\boldsymbol{\sigma}, \boldsymbol{\sigma}) \leq C_1||\boldsymbol{\sigma}||^2 \tag{4.12}$$

for all $\boldsymbol{\sigma} \in T_{\boldsymbol{\rho}}\mathcal{M}$, $\boldsymbol{\rho} \in N(\boldsymbol{\mu})$.

Moreover since the Fokker–Planck equation (1.4) is the generalized gradient flow of free energy F under the metric $d_{\bar{\Psi}}(\cdot, \cdot)$ (see Equation (45) and Theorem 3 in [7]), we have

$$\frac{\mathrm{d}F(\boldsymbol{\rho}(t))}{\mathrm{d}t} = -g^{\bar{\Psi}}_{\boldsymbol{\rho}(t)}(\dot{\boldsymbol{\rho}}(t), \dot{\boldsymbol{\rho}}(t))$$

for $t > 0$. By integrating the previous equality from 0 to T, we have

$$\begin{aligned} F(\boldsymbol{\nu}) - F(\boldsymbol{\rho}(T)) &= \int_0^T g^{\bar{\Psi}}_{\boldsymbol{\rho}(t)}(\dot{\boldsymbol{\rho}}(t), \dot{\boldsymbol{\rho}}(t))\mathrm{d}t \geq \frac{1}{T}(\int_0^T \sqrt{g^{\bar{\Psi}}_{\boldsymbol{\rho}(t)}(\dot{\boldsymbol{\rho}}(t), \dot{\boldsymbol{\rho}}(t))}\mathrm{d}t)^2 \\ &\geq \frac{1}{T}(\int_0^T \sqrt{C_2}||\dot{\boldsymbol{\rho}}(t)||\mathrm{d}t)^2 \geq \frac{C_2}{T}||\boldsymbol{\nu} - \boldsymbol{\rho}(T)||^2 \end{aligned}$$

the last second inequality comes from (4.11). At the same time,

$$\begin{aligned} F(\boldsymbol{\nu}) - F(\boldsymbol{\rho}(T)) &= \int_0^T g^{\bar{\Psi}}_{\boldsymbol{\rho}(t)}(\dot{\boldsymbol{\rho}}(t), \dot{\boldsymbol{\rho}}(t))\mathrm{d}t \\ &\geq \frac{1}{T}(\int_0^T \sqrt{g^{\bar{\Psi}}_{\boldsymbol{\rho}(t)}(\dot{\boldsymbol{\rho}}(t), \dot{\boldsymbol{\rho}}(t))}\mathrm{d}t)^2 \\ &\geq \frac{1}{T}d^2_{\bar{\Psi}}(\boldsymbol{\nu}, \boldsymbol{\rho}(T)) \geq \frac{1}{T}d^2_m(\boldsymbol{\nu}, \boldsymbol{\rho}(T)) \end{aligned}$$

the last inequality comes from (4.5).

Please cite this article in press as: R. Che et al., Convergence to global equilibrium for Fokker–Planck equations on a graph and Talagrand-type inequalities, J. Differential Equations (2016), http://dx.doi.org/10.1016/j.jde.2016.05.003

Let $\boldsymbol{s}(t) = t\boldsymbol{\rho}(T) + (1-t)\boldsymbol{\mu}$ for $t \in [0, 1]$. Since $N(\boldsymbol{\mu})$ is a convex subset of $\mathbb{R}^N$ and $\boldsymbol{\rho}(T)$, $\boldsymbol{\mu} \in N(\boldsymbol{\mu})$. We have $\boldsymbol{s}(t) \in N(\boldsymbol{\mu})$ for $t \in [0, 1]$. Thus

$$\begin{aligned} d_m^2(\boldsymbol{\rho}(T), \boldsymbol{\mu}) &\leq \Big(\int_0^1 \sqrt{g^m_{\boldsymbol{s}(t)}(\boldsymbol{\rho}(T) - \boldsymbol{\mu}, \boldsymbol{\rho}(T) - \boldsymbol{\mu})}\mathrm{d}t\Big)^2 \\ &\leq \Big(\int_0^1 \sqrt{C_1||\boldsymbol{\rho}(T) - \boldsymbol{\mu}||^2}\mathrm{d}t\Big)^2 \\ &= C_1||\boldsymbol{\rho}(T) - \boldsymbol{\mu}||^2 \end{aligned}$$

the last second inequality comes from (4.12) and the fact that $\boldsymbol{s}(t) \in N(\boldsymbol{\mu})$.

This gives us the bounds

$$\begin{aligned} d_m^2(\boldsymbol{\rho}(T), \boldsymbol{\mu}) &\leq C_1||\boldsymbol{\rho}(T) - \boldsymbol{\mu}||^2 \leq \frac{m^2}{8}C_1||\boldsymbol{\rho}(T) - \boldsymbol{\nu}||^2 \\ &\leq \frac{m^2TC_1}{8C_2}(F(\boldsymbol{\nu}) - F(\boldsymbol{\rho}(T))) \leq \frac{m^2TC_1}{8C_2}F(\boldsymbol{\nu}), \end{aligned}$$

and

$$d_m^2(\boldsymbol{\nu}, \boldsymbol{\rho}(T)) \leq T(F(\boldsymbol{\nu}) - F(\boldsymbol{\rho}(T))) \leq TF(\boldsymbol{\nu}).$$

Finally, note that $C_1 \leq \frac{2DN^2}{m^2}$ and $C_2 \geq \frac{1}{N}$. We have

$$\begin{aligned} d_m^2(\boldsymbol{\nu}, \boldsymbol{\mu}) &\leq 2d_m^2(\boldsymbol{\nu}, \boldsymbol{\rho}(T)) + 2d_m^2(\boldsymbol{\rho}(T), \boldsymbol{\mu}) \\ &\leq \Big(\frac{m^2TC_1}{4C_2} + 2T\Big)F(\boldsymbol{\nu}) = T\Big(\frac{m^2C_1}{4C_2} + 2\Big)F(\boldsymbol{\nu}) \\ &\leq \frac{M}{\lambda_2 m}\log\Big(\frac{18M}{m^3}\Big)\Big(\frac{m^2\frac{2DN^2}{m^2}}{4\frac{1}{N}} + 2\Big)F(\boldsymbol{\nu}) \\ &= KH(\boldsymbol{\nu}|\boldsymbol{\mu}), \end{aligned}$$

where $K = \frac{M(DN^3+4)}{2\lambda_2 m}\log(\frac{18M}{m^3})$. This completes the proof. □

Remark 4.4. We remark that our results are for the generic setting, under which the estimates of $\lambda_2(\boldsymbol{\rho})$ and $\lambda_N(\boldsymbol{\rho})$ are too complicated to be useful. Hence uniform bounds of (4.7) and (4.8) was used in the proof of both Talagrand-type inequalities, which makes the constant K not sharp. If more information about the graph topology is known, we expect our estimates can be improved by integrating (4.7) and (4.8) over the path.

It is well known that the Talagrand inequality implies the concentration of the normalized counting measure on a graph. To establish such a result in our situation, it is important to bound the classical 1-Wasserstein distance $W_1(\boldsymbol{\mu},\boldsymbol{\nu})$ by the metric $d_{\Psi}(\boldsymbol{\mu},\boldsymbol{\nu})$ (or $d_m(\boldsymbol{\mu},\boldsymbol{\nu})$). This can be done in the following way. (Also see Section 2 of [9].)

By Kantorovich–Rubinstein Theorem ([31] or [32]),

$$W_1(\boldsymbol{\mu},\boldsymbol{\nu}) = \sup_{\phi:\mathrm{Lip}(\phi)\le 1} \Big|\sum_{i=1}^{N} \phi_i(\boldsymbol{\mu}_i - \boldsymbol{\nu}_i)\Big|, \tag{4.13}$$

where ϕ is the test function defined on vertices and $\mathrm{Lip}(\phi)$ is the Lipschitz constant of ϕ.

Let ϕ be the test function defined on vertices with $\mathrm{Lip}(\phi) \le 1$ and $\gamma : [0,1] \to \mathcal{M}$ a continuously differentiable curve with $\gamma(0) = \boldsymbol{\nu}$, $\gamma(1) = \boldsymbol{\mu}$. Then we have

$$\Big|\sum_{i=1}^{N} \phi_i(\boldsymbol{\mu}_i - \boldsymbol{\nu}_i)\Big| = \Big|\int_0^1 \boldsymbol{\phi}\cdot\dot{\gamma}(t)\,\mathrm{d}t\Big| = \Big|\int_0^1 \boldsymbol{\phi}\cdot\big(\boldsymbol{p}_t L(G, w(\gamma(t)))\big)\,\mathrm{d}t\Big|,$$

where $\boldsymbol{p}_t$ is given by the identification (4.1) for $\dot{r}(t)$. By weighted Cauchy inequality,

$$\begin{aligned}
&\Big|\int_0^1 \boldsymbol{\phi}\cdot\big(\boldsymbol{p}_t L(G, w(\gamma(t)))\big)\mathrm{d}t\Big| \\
&= \Big|\int_0^1 \sum_{\{a_i,a_j\}\in E} w_{ij}(\gamma(t))(\phi_i-\phi_j)(p_i(t)-p_j(t))\mathrm{d}t\Big| \\
&\le \int_0^1 \Big(\sum_{\{a_i,a_j\}\in E} w_{ij}(\gamma(t))(\phi_i-\phi_j)^2\Big)^{1/2}\cdot\Big(\sum_{\{a_i,a_j\}\in E} w_{ij}(\gamma(t))(p_i(t)-p_j(t))^2\Big)^{1/2}\mathrm{d}t \\
&\le \int_0^1 \Big(\sum_{\{a_i,a_j\}\in E} \mathrm{Lip}(\phi)\Big)^{1/2}\cdot\Big(\sum_{\{a_i,a_j\}\in E} w_{ij}(\gamma(t))(p_i(t)-p_j(t))^2\Big)^{1/2}\mathrm{d}t \\
&\le \sqrt{N\mathrm{Lip}(\phi)}L(\gamma).
\end{aligned}$$

Combining the above two inequalities with (4.13) and (4.3), ones has

$$W_1(\boldsymbol{\mu},\boldsymbol{\nu}) \le \sqrt{N} d_{\Psi}(\boldsymbol{\mu},\boldsymbol{\nu}),$$

since ϕ and γ are arbitrary. Similar calculations will give the bound for $d_m(\boldsymbol{\mu},\boldsymbol{\nu})$.

Once we have,

$$W_1^2(\boldsymbol{\mu},\boldsymbol{\nu}) \le c_1 d_{\Psi}^2(\boldsymbol{\mu},\boldsymbol{\nu}) \le c_2 H(\boldsymbol{\nu}\,|\,\boldsymbol{\mu}), \tag{4.14}$$

(or

$$W_1^2(\boldsymbol{\mu}, \boldsymbol{\nu}) \le c_1 d_m^2(\boldsymbol{\mu}, \boldsymbol{\nu}) \le c_2 H(\boldsymbol{\nu} \mid \boldsymbol{\mu}) \tag{4.15}$$

) for all ν absolutely continuous with respect to (the reference measure) μ, and some positive constants c_1 and c_2, there are standard results that give the subgaussian concentration of the normalized counting measure on G: For any Lipschitz function ϕ with $\mathrm{Lip}(\phi) \le 1$ and all $h > 0$, there is

$$\pi\{\phi - \mathbb{E}_\pi \phi \ge h\} \le e^{-h^2/c_2},$$

where π is the normalized counting measure on G. One chooses π to be the reference measure $\boldsymbol{\mu}$ in establishing the transport-entropy inequality, which is well-known to be dual to the subgaussian (thanks to the work of [3].) See also [6] for details on such derivations in the discrete setting of graphs. Equation (4.15) follows from the above calculation and Theorem 4.3 easily. To obtain equation (4.14), the estimates in Theorem 4.2 must be improved by making the constant K independent of the choice of $\boldsymbol{\nu}$.

References

[1] L. Ambrosio, N. Gigli, G. Savare, Gradient Flows: In Metric Spaces and in the Space of Probability Measures, Lectures Math. ETH Zürich, Birkhäuser Verlag, Basel, 2008.
[2] A. Berman, X.D. Zhang, Lower bounds for the eigenvalues of Laplacian matrices, Linear Algebra Appl. 316 (1–3) (2000) 13–20.
[3] S. Bobkov, F. Götze, Exponential integrability and transportation cost related to logarithmic Sobolev inequalities, J. Funct. Anal. 163 (1) (1999) 1–28.
[4] S. Bobkov, I. Gentil, M. Ledoux, Hypercontractivity of Hamilton–Jacobi equations, J. Math. Pures Appl. 80 (7) (2001) 669–696.
[5] S. Bobkov, P. Tetali, Modified log-Sobolev inequalities, mixing and hypercontractivity, J. Theoret. Probab. 19 (2) (2006) 289–336.
[6] S. Bobkov, C. Houdré, P. Tetali, The subgaussian constant and concentration inequalities, Israel J. Math. 156 (1) (2006) 255–283.
[7] S.N. Chow, W. Huang, Y. Li, H.M. Zhou, Fokker–Planck equations for a free energy functional or Markov process on a graph, Arch. Ration. Mech. Anal. 203 (3) (2012) 969–1008.
[8] K. Disser, M. Liero, On gradient structures for reversible Markov chains and the passage to Wasserstein gradient flows, Netw. Heterog. Media 10 (2) (2015) 233–253.
[9] M. Erbar, J. Maas, Ricci curvature of finite Markov chains via convexity of the entropy, Arch. Ration. Mech. Anal. 206 (3) (2012) 997–1038.
[10] C.W. Gardiner, Handbook of Stochastic Methods. For Physics, Chemistry and the Natural Sciences, third edition, Springer Ser. Synergetics, vol. 13, Springer-Verlag, Berlin, 2004.
[11] N. Gozlan, C. Roberto, P-M. Samson, Private Communication, November 2010.
[12] R. Jordan, D. Kinderlehrer, F. Otto, The variational formulation of the Fokker–Planck equation, SIAM J. Math. Anal. 29 (1) (1998) 1–17.
[13] J. Lott, C. Villani, Ricci curvature for metric-measure spaces via optimal transport, Ann. of Math. (2) 169 (3) (2009) 903–991.
[14] J. Lott, C. Villani, Hamilton–Jacobi semigroup on length spaces and applications, J. Math. Pures Appl. 88 (3) (2007) 219–229.
[15] M. Lu, L. Zhang, F. Tian, Lower bounds of the Laplacian spectrum of graphs based on diameter, Linear Algebra Appl. 420 (2–3) (2007) 400–406.
[16] D.H. Mathews, D.H. Turner, Prediction of RNA secondary structure by free energy minimization, Curr. Opin. Struct. Biol. 16 (3) (2006) 270–278.
[17] J. Maas, Gradient flows of the entropy for finite Markov Chains, J. Funct. Anal. 261 (8) (2011) 2250–2292.

Please cite this article in press as: R. Che et al., Convergence to global equilibrium for Fokker–Planck equations on a graph and Talagrand-type inequalities, J. Differential Equations (2016), http://dx.doi.org/10.1016/j.jde.2016.05.003

[18] A. Mielke, A gradient structure for reaction-diffusion systems and for energy-drift-diffusion systems, Nonlinearity 24 (4) (2011) 1329–1346.

[19] A. Mielke, Geodesic convexity of the relative entropy in reversible Markov chains, Calc. Var. Partial Differential Equations 48 (1–2) (2013) 1–31.

[20] F. Otto, The geometry of dissipative evolution equations: the porous medium equation, Comm. Partial Differential Equations 26 (1–2) (2001) 101–174.

[21] F. Otto, C. Villani, Generalization of an inequality by Talagrand and links with the logarithmic Sobolev inequality, J. Funct. Anal. 173 (2) (2000) 361–400.

[22] H. Risken, The Fokker–Planck Equation: Methods of Solution and Applications, second edition, Springer Ser. Synergetics, vol. 18, Springer-Verlag, Berlin, 1989.

[23] O. Rojo, A nontrivial upper bound on the largest Laplacian eigenvalue of weighted graphs, Linear Algebra Appl. 420 (2–3) (2007) 625–633.

[24] M. Sammer, Aspects of mass transportation in discrete concentration inequalities, Thesis (Ph.D.), Georgia Institute of Technology, 2005, 119 pp.

[25] M. Sammer, P. Tetali, Concentration on the discrete torus using transportation, Combin. Probab. Comput. 18 (5) (2009) 835–860.

[26] Z. Schuss, Singular perturbation methods in stochastic differential equations of mathematical physics, SIAM Rev. 22 (2) (1980) 119–155.

[27] K.T. Sturm, On the geometry of metric measure spaces I, Acta Math. 196 (1) (2006) 65–131.

[28] K.T. Sturm, On the geometry of metric measure spaces II, Acta Math. 196 (1) (2006) 133–177.

[29] B. Smolka, K.W. Wojciechowski, Contrast Enhancement of Badly Illuminated Images Based on Gibbs Distribution and Random Walk model, Lecture Notes in Comput. Sci., vol. 1296, Springer, Berlin, 1997, pp. 271–278.

[30] A. Torgasev, M. Petrovic, Lower bounds of the Laplacian graph eigenvalues, Indag. Math. (N.S.) 15 (4) (2004) 589–593.

[31] C. Villani, Optimal Transport. Old and New, Grundlehren Math. Wiss., vol. 338, Springer-Verlag, Berlin, 2009.

[32] C. Villani, Topics in Optimal Transportation, Grad. Stud. Math., vol. 58, American Mathematical Society, Providence, 2003.

[33] Y. Wu, G. Hua, T. Yu, Tracking articulated body by dynamic Markov network, in: The Ninth IEEE International Conference on Computer Vision (ICCV 2003) 2-Volume Set, 2003.

Please cite this article in press as: R. Che et al., Convergence to global equilibrium for Fokker–Planck equations on a graph and Talagrand-type inequalities, J. Differential Equations (2016), http://dx.doi.org/10.1016/j.jde.2016.05.003