A note on a Bonnet-Myers type diameter bound for graphs with positive entropic Ricci curvature

S. Kamtue

Department of Mathematical Sciences, Durham University

March 4, 2020

Abstract

An equivalent definition of entropic Ricci curvature on discrete spaces was given in terms of the global gradient estimate in [1, Theorem 3.1]. With a particular choice of the density function $\rho$, we obtain a localized gradient estimate, which in turns allow us to apply the same technique as in [4] to derive a Bonnet-Myers type diameter bound for graphs with positive entropic Ricci curvature. However, the case of the hypercubes indicates that the bound may be not optimal (where $\theta$ is chosen to be logarithmic mean by default). If $\theta$ is arithmetic mean, the Bakry-Émery criterion can be recovered and the diameter bound is optimal as it can be attained by the hypercubes.

1 Introduction

The notion of entropic Ricci curvature on discrete spaces (i.e., Markov chains, or graphs) was introduced by Erbar and Maas [2] inspired by work of Sturm [10] and Lott and Villani [6] which describes the lower bound of Ricci curvature via the displacement convexity of the entropy functional. Equivalent definitions of Entropic curvature are also given in terms of Bochner’s inequality [2] and in terms of gradient estimates [1]. Detailed notations and definitions are given in Section 2 and proofs are presented in Section 3.

Definition 1.1 (Entropic curvature). An irreducible and reversible Markov kernel $Q$ on a finite discrete space $X$ (with a steady state $\pi$) has entropic curvature at least $\kappa \in \mathbb{R}$, if and only if the entropy functional $H$ satisfies the following $\kappa$-convexity inequality:

$$H(\rho_t) \leq (1 - t)H(\rho_0) + tH(\rho_1) - \frac{\kappa}{2}t(1 - t)W(\rho_0, \rho_1)^2$$

for every constant speed geodesic $(\rho_t)_{t\in[0,1]}$ in the (modified) $L^2$-Wasserstein space of probability densities $(\mathcal{P}(X), W)$.

Theorem 1.1 (Equivalent definitions). For a finite set $X$ equipped with an irreducible and reversible Markov kernel $Q$ (and the steady state $\pi$), the following properties are equivalent:

1. $\text{Ric}(Q) \geq \kappa$;
Bochner’s inequality holds for all \( \rho \in \mathcal{D}(X) \) and \( f \in C(X) \):

\[
B(\rho, f) \geq \kappa A(\rho, f),
\]

(1.2)

where

\[
A(\rho, f) := \langle \hat{\rho} \cdot \nabla f, \nabla f \rangle_\pi,
\]

and

\[
B(\rho, f) := \frac{1}{2} \langle \hat{\Delta} \rho \cdot \nabla f, \nabla f \rangle_\pi - \langle \hat{\rho} \cdot \nabla f, \nabla \Delta f \rangle_\pi,
\]

where

\[
\hat{\rho}(x, y) := \theta(\rho(x), \rho(y)),
\]

\[
\hat{\Delta} \rho := \partial_1 \theta(\rho(x), \rho(y)) \Delta \rho(x) + \partial_2 \theta(\rho(x), \rho(y)) \Delta \rho(y),
\]

and \( \theta \) is a logarithmic mean: \( \theta(a, b) = \frac{a - \log a - \log b}{a - b} \).

Global gradient estimate holds for all \( \rho \in \mathcal{D}(X) \) and \( f \in C(X) \) and \( t \geq 0 \),

\[
|\nabla P_t f|_\rho^2 \leq e^{-2\kappa t} |\nabla f|_{P_t \rho}^2,
\]

or more explicitly

\[
\frac{1}{2} \sum_{u,v} (P_t f(v) - P_t f(u))^2 \hat{\rho}(u, v) Q(u, v) \pi(u) \leq e^{-2\kappa t} \frac{1}{2} \sum_{u,v} (f(v) - f(u))^2 P_t \rho(u, v) Q(u, v) \pi(u)
\]

(1.3)

where \( P_t = e^{t \Delta} \) denotes the heat semigroup.

The contribution of this paper is to follow ideas from [1] to derive a similar local gradient estimate, and to apply the same technique as in [4] to prove a Bonnet-Myers type diameter bound on the underlying graph \( X \) (when \( Q \) is a simple random walk) in the case of a positive lower bound of entropic curvature.

**Theorem 1.2** (Local gradient estimate). If \( \text{Ric}(Q) \geq \kappa \), then for all \( x \in X \) and \( f \in C(X) \) and all \( \varepsilon \geq 0 \) we have

\[
\Gamma(P_t f)(x) \pi(x) \leq \frac{e^{-2\kappa t}}{2\theta(1, \varepsilon)} [P_t \Gamma(f)(x) \pi(x) + \varepsilon \sum_{y \in N(x)} P_t \Gamma(f)(y) \pi(y)].
\]

(1.4)

where \( \Gamma(f)(x) := \sum_y (f(y) - f(x))^2 Q(x, y) \).

**Theorem 1.3** (Diameter bound). Let \( Q \) represent a simple random walk on \( X \) with strictly positive entropic Ricci curvature \( \text{Ric}(Q) \geq \kappa > 0 \). Then the diameter is bounded from above by \( \text{diam}(X) \leq \frac{2}{\kappa} \sqrt{\frac{D \log D}{D - 1}} \) where \( D \) is the maximal (vertex) degree.

**Remark 1.4.** It is known that the entropic curvature of (a simple random walk on) the discrete hypercube \( \mathbb{Q}^n \) is \( \frac{2}{n} \) (see [2, Example 5.7]). Therefore, in view of the hypercubes, the bound in Theorem 1.3 is not optimal:

\[
n = \text{diam}(\mathbb{Q}^n) \leq n \sqrt{n - 1} \log n.
\]
However, if we replace the logarithmic mean by the arithmetic mean for \( \theta \) in Bochner’s formula \( (1.2) \) and inequality \( (1.3) \), the local gradient estimate in Theorem \( 1.2 \) would imply (by taking \( \varepsilon = 0 \)) the Bakry-Émery curvature criterion \( CD(\kappa, \infty) \): \( \Gamma(P_t f)(x) \leq e^{-2\kappa t} \Gamma(f)(x) \) in the sense of \[3\] Corollary 3.3. This implication has already been mentioned in the survey \[8\]. Consequently, we obtain a diameter bound: \( \text{diam}(X) \leq 2\kappa \), which is sharp and the equality is attained if and only if \( X \) is a hypercube \( Q^n \) (for details see \[4\] and \[5\]).

2 Setup and notations

2.1 Notions associated to a discrete Markov chain

We start with a Markov chain \( (X, Q) \), where \( X \) is a finite set and \( Q : X \times X \to \mathbb{R}^+ \cup \{0\} \) is a Markov kernel \( Q \), i.e. \( \sum_{y \in X} Q(x, y) = 1 \) for all \( x \in X \). Furthermore, we assume that \( Q \) is irreducible and reversible, which implies that there exists a unique stationary probability measure \( \pi \) on \( X \) satisfying \( \sum_{x \in X} \pi(x) = 1 \) and the detail balanced equations:

\[
Q(x, y)\pi(x) = Q(y, x)\pi(y) \quad \forall x, y \in X.
\]

The set of probability densities (with respect to \( \pi \)) is defined as

\[
\mathcal{D}(X) := \left\{ \rho : X \to \mathbb{R}^+ \cup \{0\} \mid \sum_{x \in X} \pi(x)\rho(x) = 1 \right\}.
\]

The entropy functional, defined on \( \mathcal{D}(X) \), is given by

\[
\mathcal{H}(\rho) := \sum_{x \in X} \rho(x) \log \rho(x)\pi(x).
\]

The discrete gradient \( \nabla : C(X) \to C(X \times X) \), discrete divergence \( \nabla \cdot : C(X \times X) \to C(X) \), and laplacian \( \Delta := \nabla \cdot \nabla \) are defined as follows:

**Definition 2.1.** For all \( f, g \in C(X) \) and \( U, V \in C(X \times X) \),

\[
\nabla f(x, y) := f(y) - f(x),
\]

\[
\nabla \cdot V(x) := \frac{1}{2} \sum_{y \in X} (V(x, y) - V(y, x)) Q(x, y),
\]

\[
\Delta f(x) := \sum_{y \in X} (f(y) - f(x))Q(x, y).
\]

The inner products are defined as

\[
\langle f, g \rangle_\pi := \sum_{x \in X} f(x)g(x)\pi(x),
\]

\[
\langle U, V \rangle_\pi := \frac{1}{2} \sum_{x, y \in X} U(x, y)V(x, y)Q(x, y)\pi(x),
\]

3
and for all $\rho \in \mathcal{D}(X)$,

$$
\langle U, V \rangle_\rho := \langle \dot{\rho} \cdot U, V \rangle = \frac{1}{2} \sum_{x, y \in X} U(x, y)V(x, y)\dot{\rho}(x, y)Q(x, y)\pi(x),
$$

where $\dot{\rho}(x, y) := \theta(\rho(x), \rho(y))$ and $\theta$ is a suitable mean satisfying Assumption 2.1 in [2]. By default, $\theta$ is chosen to be the logarithmic mean:

$$
\theta(a, b) = \frac{a^b - b^a}{\log a - \log b}.
$$

Furthermore, we already introduced

$$
\mathcal{A}(\rho, f) := |\nabla f|^2 = \langle \dot{\rho} \cdot \nabla f, \nabla f \rangle_\pi,
$$

$$
\mathcal{B}(\rho, f) := \frac{1}{2} \langle \hat{\Delta} \rho \cdot \nabla f, \nabla f \rangle_\pi - \langle \dot{\rho} \cdot \nabla f, \nabla \Delta f \rangle_\pi,
$$

where $\hat{\Delta} \rho := \partial_1 \theta(\rho(x), \rho(y))\Delta \rho(x) + \partial_2 \theta(\rho(x), \rho(y))\Delta \rho(y)$.

**Definition 2.2** (Discrete transport metric). For $\bar{\rho}_0, \bar{\rho}_1 \in \mathcal{D}(X)$,

$$
\mathcal{W}(\bar{\rho}_0, \bar{\rho}_1) := \inf \left\{ \int_0^1 \mathcal{A}(\rho_t, f_t) \, dt \biggm| (\rho_t, f_t) \in CE(\bar{\rho}_0, \bar{\rho}_1) \right\} \frac{1}{2},
$$

where the infimum is taken over the set $CE(\rho_0, \rho_1)$ which consists of all sufficiently regular curves $(\rho_t)_{t \in [0,1]}$ on $\mathcal{D}(X)$ and $(f_t)_{t \in [0,1]}$ on $C(X)$ which satisfy the continuity equation $\partial_t \rho_t + \nabla \cdot (\hat{\rho}_t \nabla f_t) = 0$ and $\rho_0 = \bar{\rho}_0, \rho_1 = \bar{\rho}_1$. It was shown that $\mathcal{W}$ is a metric on $\mathcal{D}(X)$. We refer to [2, 7] for further details. Note that this notion of the $\mathcal{W}$-metric is relevant in Definition 1.1 of entropic curvature.

### 2.2 Graph theoretical notions

The kernel $Q$ induces a graph structure on $X$ by assigning an edge $x \sim y$ if and only if $Q(x, y) > 0$. Note that the graph is connected and undirected, due to irreducibility and reversibility of $Q$, respectively. For a vertex $x \in X$, denote $N(x) := \{ y \in X \mid x \sim y \}$ the set of neighbors of $x$, and $d_x := |N(x)|$ the degree of $x$, and $D := \max_{x \in X} d_x$ the maximal degree. The graph is equipped with the usual combinatorial distance function $d$ where $d(x, y)$ is the length of shortest path(s) between $x$ and $y$, and the diameter of the graph is defined as $diam(X) = \max_{x, y \in X} d(x, y)$. In this note, we restrict $Q$ to be a simple random walk, which is given by $Q(x, y) = \frac{1}{d_x}$ for all $y \in N(x)$, and $\pi(x) = d_x / (\sum_{v \in X} d_v)$.

### 3 Proofs

**Proof of Theorem 1.1** Equivalence (1) $\Leftrightarrow$ (2) was stated in [2, Theorem 4.5] under the assumption that $\theta$ is the logarithmic mean. Equivalence (2) $\Leftrightarrow$ (3) was stated in [1, Theorem 3.1], regardless of the choice of a ‘suitable’ mean for $\theta$. \[Q\]

**Proof of Theorem 1.2**

The proof follows ideas of [1, Corollary 3.4]. Note that the inequality (13) is homogeneous in $\rho$. Therefore, we can drop the requirement that $\rho$ is a probability density: $\sum_x \rho(x)\pi(x) =
1. We localize inequality (1.3) by choosing \( \rho = 1_x + \varepsilon \sum_{y \in N(x)} 1_y \) for a fixed \( x \in V \) and a parameter \( \varepsilon \in [0, \infty) \).

In particular we know that \( \hat{\rho}(x, y) = \hat{\rho}(y, x) = \theta(1, \varepsilon) \) for all \( y \in N(x) \). The left-hand-side of (1.3) has the following lower bound

\[
L.H.S. \geq \sum_{y \in N(x)} (P_t f(y) - P_t f(x))^2 \theta(1, \varepsilon) Q(x, y) \pi(x) = \theta(1, \varepsilon) \cdot \Gamma(P_t f)(x) \pi(x) \tag{3.1}
\]

On the other hand, we have the following bound on the right-hand-side of (1.3) by

\[
R.H.S. \leq e^{-2ct} \frac{1}{2} \sum_{u,v} (f(v) - f(u))^2 P_t \rho(u) + P_t \rho(v) Q(u, v) \pi(u)
\]

\[
= e^{-2ct} \frac{1}{2} \sum_{u,v} (f(v) - f(u))^2 P_t \rho(u) Q(u, v) \pi(u) \tag{3.2}
\]

due to \( \theta(s, t) \leq (s + t)/2 \) and symmetry from interchanging \( u \) and \( v \).

We now apply the heat kernel \( p_t(\cdot, \cdot) \) given by \( P_t g(u) = \sum_z p_t(u, z) g(z) \pi(z) \) for every function \( g \). With our chosen \( \rho \), we obtain \( P_t \rho(u) = p_t(u, x) \rho(x) \pi(x) + \varepsilon \sum_{y \in N(x)} p_t(u, y) \rho(y) \pi(y) \), which we substitute into (3.2) and use the symmetry of heat kernel: \( p_t(u, v) = p_t(v, u) \) to derive

\[
R.H.S. \leq e^{-2ct} \frac{1}{2} \left[ \pi(x) \sum_u p_t(u, x) \pi(u) \sum_v (f(v) - f(u))^2 Q(u, v) + \varepsilon \sum_{y \in N(x)} \pi(y) \sum_u p_t(u, y) \pi(u) (\sum_v (f(v) - f(u))^2 Q(u, v)) \right]
\]

\[
= e^{-2ct} \frac{1}{2} \left[ \pi(x) \Gamma(f)(x) + \varepsilon \sum_{y \in N(x)} \pi(y) P_t \Gamma(f)(y) \right]. \tag{3.3}
\]

The desired inequality then follows from (3.1) and (3.3). \( \square \)

For the underlying graph \( X \) with \( Q \) representing a simple random walk, we have the following corollary as an immediate consequence of Theorem 1.2

**Corollary 3.1.** Let \( Q \) represent a simple random walk on \( X \) with entropic Ricci curvature \( \text{Ric}(Q) \geq \kappa. \) Then we have the following gradient estimate:

\[
\Gamma(P_t f)(x) \leq c \cdot e^{-2ct} \|P_t \Gamma(f)\|_\infty \tag{3.4}
\]

where \( c := \frac{D \log D}{D - 1} \) and \( D \) is the maximal degree.

**Proof of Corollary 3.1.** Theorem 1.2 implies that \( \Gamma(P_t f)(x) \leq c_{\varepsilon, x} \cdot e^{-2ct} \|P_t \Gamma(f)\|_\infty \).

where \( c_{\varepsilon, x} := \frac{1 + \varepsilon \sum_{y \in N(x)} d_y}{2 \theta(1, \varepsilon)} \leq \frac{1 + \varepsilon D}{2 \theta(1, \varepsilon)} \). In particular when \( \varepsilon = \frac{1}{D} \), we have \( c_{\varepsilon, x} \leq c. \) \( \square \)

Finally, we present the proof of the diameter bound in the case of strictly positive entropic curvature.

**Proof of Theorem 1.3.**
The proof follows ideas of [4, Theorem 2.1]. Consider a particular choice of function \( f \in C(X) \) given by \( f(x) = d(x, x_0) \) for an arbitrary reference point \( x_0 \in X \). Since \( f \) is a Lipschitz function with constant 1, it follows that \( \Gamma(f)(x) = \sum_{y \in N(x)} \frac{1}{d_x}(f(y) - f(x))^2 \leq 1 \) for all \( x \in X \), i.e., \( \|\Gamma(f)\|_\infty \leq 1 \), which then implies \( \|P_t\Gamma(f)\|_\infty \leq \|\Gamma(f)\|_\infty \leq 1 \).

Moreover, Cauchy-Schwartz and inequality (3.4) give

\[
|\Delta P_t f(x)|^2 \leq \frac{1}{d_x} \sum_{y \in N(x)} (P_t f(y) - P_t f(x))^2 = \Gamma(P_t f)(x) \leq e^{-2\kappa t} \|P_t\Gamma(f)\|_\infty \leq ce^{-2\kappa t}.
\]

From the fundamental theorem of calculus and the definition of \( P_t \), we then obtain

\[
|f(x) - P_T f(x)| \leq \int_0^T \left| \frac{\partial}{\partial t} P_t f(x) \right| dt \leq \int_0^T \sqrt{ce^{-\kappa t}} dt \leq \frac{\sqrt{c}}{\kappa},
\]

which holds true for all \( T > 0 \).

Moreover, the chain of inequalities in (3.5) implies that \( |P_t f(y) - P_t f(x)|^2 \leq d_x \cdot ce^{-2\kappa t} \to 0 \) as \( t \to \infty \) for all neighbors \( y \sim x \). Therefore, \( |P_t f(y) - P_t f(x)| \to 0 \) as \( t \to \infty \) for an arbitrary pair of \( x, y \) (by considering a connected path from \( x \) to \( y \)).

Passing to the limit \( T \to \infty \), we can conclude from triangle inequality that

\[
d(x, x_0) = |f(x) - f(x_0)| \leq |f(x) - P_T f(x)| + |f(x_0) - P_T f(x_0)| + |P_T f(x) - P_T f(x_0)| \leq \frac{2\sqrt{c}}{\kappa}.
\]

Since \( x, x_0 \) are arbitrary, we obtain the desired diameter bound.

\[\square\]

References

[1] M. Erbar and M. Fathi, Poincaré, modified logarithmic Sobolev and isoperimetric inequalities for Markov chains with non-negative Ricci curvature, J. Funct. Anal. 274(11) (2018), 3056–3089.

[2] M. Erbar and J. Maas, Ricci Curvature of Finite Markov Chains via Convexity of the Entropy, Arch. Rational Mech. Anal. 206 (2012) 997–1038

[3] S. Liu and Y. Lin, Equivalent Properties of CD Inequality on Graph, arXiv:1512.02677

[4] S. Liu, F. Münch and N. Peyerimhoff, Bakry-Émery curvature and diameter bounds on graphs, Calc. Var. Partial Differential Equations 57(2) (2018), Art. 67, 9.

[5] S. Liu, F. Münch and N. Peyerimhoff, Rigidity properties of the hypercube via Bakry-Émery curvature, arXiv:1705.06789.

[6] L. Lott and C. Villani, Ricci curvature for metric-measure spaces via optimal transport, Ann. Math. (2) 169(3) (2009), 903–991

[7] J. Maas, Gradient flows of the entropy for finite Markov chains, J. Funct. Anal. 261(8) (2011), 2250–2292.
[8] J. Maas, *Entropic Ricci Curvature for Discrete Spaces*, Modern Approaches to Discrete Curvature, Lecture Notes in Mathematics, vol 2184 (2017), L. Najman and P. Romon (eds), Springer, Cham.

[9] S. B. Myers, *Riemannian manifolds with positive mean curvature*, Duke Math. J. 8 (1941), 401–404.

[10] K.-Th. Sturm, *On the geometry of metric measure spaces I and II*, Acta Math. 196(1) (2006), 65–177.