Pure and Applied Mathematics Quarterly
Volume 18, Number 6, 2569–2598, 2022

# Concentration inequalities in spaces of random configurations with positive Ricci curvatures

Linyuan Lu* and Zhiyu Wang

**Abstract:** In this paper, we prove an Azuma-Hoeffding-type inequality in several classical models of random configurations by a Ricci curvature approach. Adapting Ollivier's work on the Ricci curvature of Markov chains on metric spaces, we prove a cleaner form of the corresponding concentration inequality in graphs. Namely, we show that for any Lipschitz function $f$ on any graph (equipped with an ergodic random walk and thus an invariant distribution $\nu$) with Ricci curvature at least $\kappa > 0$, we have

$$\nu\left(|f - E_\nu f| \geq t\right) \leq 2\exp\left(-\frac{t^2\kappa}{7}\right).$$

**Keywords:** Ricci curvature, concentration inequality, random graphs.

## 1. Introduction

One of the main tools in probabilistic analysis and random graph theory are concentration inequalities, which are meant to bound the probability that a random variable deviates from its expectation. Many of the classical concentration inequalities (such as those for binomial distributions) provide best possible deviation results with exponentially small probabilistic bounds. Such concentration inequalities usually require certain independence assumptions (e.g., the random variable is a sum of independent random variables). For concentration inequalities without the independence assumptions, one popular approach is the martingale method. A martingale is a sequence of random variables $X_0, X_1, \ldots, X_n$ with finite means such that $E[X_{i+1}|X_i, X_{i-1}, \ldots, X_0] = X_i$ for all $0 \leq i < n$. For $\mathbf{c} = (c_1, c_2, \ldots, c_n)$ with positive entries, a martingale $X$ is said to be $\mathbf{c}$-Lipschitz if $|X_i - X_{i-1}| \leq c_i$ for

Received May 27, 2021.
**2010 Mathematics Subject Classification:** Primary 05C81, 60F10, 53C44.
*This author was supported in part by NSF grant DMS-1600811 and NSF DMS-2038080.

$i \in [n]$. A powerful tool for controlling martingales is the Azuma-Hoeffding inequality [3, 27]: if a martingale is $\mathbf{c}$-Lipschitz, then

$$\Pr\left(|X - E[X]| \geq t\right) \leq 2\exp\left(-\frac{t^2}{2\sum_{i=1}^{n} c_i^2}\right).$$

For more general versions of martingale inequalities as well as applications of martingale inequalities, we refer the readers to [2, 12].

A graph $G = (V(G), E(G))$ consists of a vertex set $V(G)$ and an edge set $E(G)$ of $G$ where each edge is an unordered pair of two vertices. Given a vertex $v \in V(G)$, we use $\Gamma_G(v)$ to denote the set of *open neighbors* of $v$ in $G$, i.e., $\Gamma_G(v) = \{u \in V(G) : vu \in E(G)\}$. We use $d_G(v)$ to denote $|\Gamma_G(v)|$. Moreover, let $N_G(v) = \Gamma_G(v) \cup \{v\}$ be the *closed neighbors* of $v$ in $G$. When $G$ is clear from the context, we may ignore the subscript. For a graph $G$ and $S \subseteq V(G)$, we use $G[S]$ to denote the subgraph of $G$ induced by $S$. Given two vertices $x, y \in V(G)$, let $d(x, y)$ be the graph distance between $x$ and $y$, i.e., the length of a shortest path between $x$ and $y$. The *diameter* of $G$, denoted by $\mathrm{diam}(G)$, is the maximum $d(x, y)$ over all pairs of vertices $x, y$ in $G$. A graph parameter/function $X$ is called *vertex-Lipschitz* (or *edge-Lipschitz*) if $|X(G_1) - X(G_2)| \leq 1$ whenever $G_1$ and $G_2$ can be made isomorphic by deleting one vertex (or at most one edge) from each graph. Many graph parameters are vertex(edge)-Lipschitz, e.g., the independence number $\alpha(G)$, the chromatic number $\chi(G)$, the clique number $\omega(G)$, the domination number $\gamma(G)$, the matching number $\beta(G)$, etc.

Concentration inequalities are among the most important tools in the probabilistic analysis of random graphs. The classical binomial random graph model, which is now more commonly referred to as the the Erdős-Rényi random graph model, denoted by $G(n, p)$, is a random graph model in which a graph on $n$ vertices is constructed by connecting the vertices randomly such that each vertex pair appears as an edge with probability $p$ independently from every other edge. Another Erdős-Rényi random graph model, denoted by $G(n, M)$, is the model in which a graph is chosen uniformly at random from the collection of all graphs with $n$ vertices and $M$ edges. A standard application of the Azuma-Hoeffding inequality gives us that for any vertex-Lipschitz function $X$ defined on a vertex-exposure martingale (see e.g. [2] for definition), we have

$$\Pr(|X - \mathrm{E}(X)| \geq t) \leq 2\exp\left(-\frac{t^2}{2n}\right). \tag{1}$$

Similar concentration results can be obtained for edge-exposure martingale as well.

In this paper, we will take an alternative approach for such an inequality. The main idea is using Ollivier's work [38] on the Ricci curvature of Markov chains on metric spaces. Although the Ricci curvature of graphs has been introduced since 2009, it has not been widely used by the communities of combinatorists and graph theorists. In this paper, we prove a clean concentration result (Theorem 1) on graphs with positive Ricci curvature. Then we show that it can be applied to some classical models of random configurations including the Erdős-Rényi random graph model $G(n, p)$ and $G(n, M)$, the random $d$-out(in)-regular directed graphs, and the space of random permutations, through a geometrization process.

Consider a graph (loops allowed) $G$ equipped with a random walk whose transition probabilities are given by $m := \{m_v \colon v \in V(G)\}$. Here for each vertex $v$, $m_v \colon N(v) \to [0, 1]$ is a distribution, i.e., $\sum_{x \in N(v)} m_v(x) = 1$. A random walk is *ergodic* if starting from any initial distribution, the random walk converges to some unique invariant distribution $\nu$. In the context of random walks on graphs, in order for the random walk to be ergodic, it is sufficient that the underlying graph $G$ is connected and non-bipartite. Note that $\nu$ is a probability measure on $V$. It turns $V$ into a probability space. A function $f \colon V \to \mathbb{R}$ is called $c$-Lipschitz on $G$ if

$$|f(u) - f(v)| \leq c \quad \text{for any } uv \in E(G). \tag{2}$$

We have the following theorem on the concentration result of $f$. All we need is that the graph $G$ (equipped with a random walk) has positive Ricci curvature at least $\kappa > 0$. (See the definition of Ricci curvature (in Ollivier's notion) in next section.) We remark that Theorem 1 is a graph-theoretical version of the corresponding inequality in Theorem 33 of [38].

**Theorem 1.** *Suppose that a graph $G = (V, E)$ equipped with an ergodic random walk $m$ (and invariant distribution $\nu$) has positive Ricci curvature at least $\kappa > 0$. Then for any 1-Lipschitz function $f$ and any $t \geq 1$, we have*

$$\nu\left(f - E_\nu[f] > t\right) \leq \exp\left(\frac{-t^2\kappa}{7}\right), \tag{3}$$

$$\nu\left(f - E_\nu[f] < -t\right) \leq \exp\left(\frac{-t^2\kappa}{7}\right). \tag{4}$$

**Remark 1.** *The constant 7 can be improved to 5 if $\kappa \to 0$ as $|V(G)| \to \infty$. It can be improved to $1 + o(1)$ if we further assume that $t\kappa \to 0$ as $|V(G)| \to \infty$.*

**Remark 2.** *Ollivier [38] proved a concentration inequality for any random walk on a metric space with positive Ricci curvature at least $\kappa > 0$ and unique invariant distribution $\nu$. His result is more general but more technical to apply in the context of graphs. In particular, he defined two quantities related to the local behavior of the random walk: the* diffusion constant $\sigma(x)$ *and the* local dimension $n_x$ *at vertex $x$. Moreover, define $D_x^2 = \frac{\sigma(x)^2}{n_x \kappa}$, $D^2 = E_\nu[D_x^2]$, $t_{max} = \frac{D^2}{\max(\sigma_\infty, 2C/3)}$ where $C$ satisfies that the function $x \to D_x^2$ is $C$-Lipschitz. He proved ([38] Theorem 33, on page 834) for any 1-Lipschitz function $f$ and for any $t \leq t_{max}$, we have*

$$\nu\left(f - E_\nu[f] > t\right) \leq \exp\left(\frac{-t^2}{6D^2}\right). \tag{5}$$

*and for $t \geq t_{max}$,*

$$\nu\left(f - E_\nu[f] > t\right) \leq \exp\left(\frac{-t^2}{6D^2} - \frac{t - t_{max}}{\max(3\sigma_\infty, 2C)}\right). \tag{6}$$

*Note in Ollivier's result for graphs, we have $D^2 = O(\kappa^{-1})$ and $\sigma_\infty \approx 1$. Inequality (3) has about the same power as Inequalities (5) and (6), but simpler to apply in the context of graphs. We also give a more graph-theoretical proof for Theorem 1.*

Besides Ollivier's definition of Ricci curvature, another notion of Ricci curvature on discrete spaces, via geodesic convexity of the entropy (in the spirit of Sturm [40], Lott and Villani [35]), was proposed in [36] and systematically studied in [23] and [37]. Similar Gaussian-type concentration inequalities (as the ones in Theorem 1) in this notion of Ricci curvature are proven in [23]. Erbar, Maas, and Tetali [24] recently calculated the Ricci curvature lower bound of some classical random walks, e.g., the Bernoulli-Laplace model and the random transposition model of permutations.

In this paper, we adopt Ollivier's notion of coarse Ricci curvature as it does not require the reversibility of the random walk on graphs. The paper is organized as follows. In Section 2, we will give the history and definitions of Ricci curvature. The proof of Theorem 1 will be given in Section 3. In the last section, we will give applications of Theorem 1 in four classical models of random configurations, including the Erdős-Rényi random graph model $G(n, p)$ and $G(n, M)$, the random $d$-out(in)-regular directed graphs, and the space of random permutations.

## 2. Ricci curvatures of graphs

In Riemannian geometry, spaces with positive Ricci curvature enjoy nice properties, some of them with probabilistic interpretations. Many interesting properties are found on manifolds with non-negative Ricci curvature or on manifolds with Ricci curvature bounded below. The definition of the Ricci curvature on metric spaces first came from the Bakry and Emery notation [4] who defined the "lower Ricci curvature bound" through the heat semi-group $(P_t)_{t\geq 0}$ on a metric measure space. Ollivier [38] defined the coarse Ricci curvature of metric spaces in terms of how much small balls are closer (in Wasserstein transportation distance) than their centers are. This notion of coarse Ricci curvature on discrete spaces was also made explicit in the Ph.D. thesis of Sammer [39]. Under the assumption of positive curvature in a metric space, Gaussian-like or Poisson-like concentration inequalities can be obtained. Such concentration inequalities have been investigated in [31] for time-continuous Markov jump processes and in [38, 32] in metric spaces.

Graphs and manifolds share some similar properties through Laplace operators, heat kernels and random walks, etc. A series of work in this area were done by Chung, Yau and their coauthors [9, 13, 14, 15, 16, 10, 17, 8, 18, 11, 19]. The first definition of Ricci curvature on graphs was introduced by Chung and Yau in [14]. For a more general definition of Ricci curvature, Lin and Yau [34] gave a generalization of the lower Ricci curvature bound in the framework of graphs. Lin, Lu, and Yau [33] defined a new kind of Ricci curvature on graphs, which is based on Ollivier's work [38].

In this paper, we will use the same notation as in [33]. Given a graph $G = (V, E)$, a probability distribution (over the vertex set $V$) is a mapping $m : V \to [0, 1]$ satisfying $\sum_{x\in V} m(x) = 1$. The *support* of $m$, denoted by $\mathrm{Supp}(m)$, is defined as the set of vertices $x \in V(G)$ such that $m(x) > 0$. Suppose two probability distributions $m_1$ and $m_2$ have finite support. A coupling between $m_1$ and $m_2$ is a mapping $A : V \times V \to [0, 1]$ with finite support so that

$$\sum_{y\in V} A(x, y) = m_1(x) \text{ and } \sum_{x\in V} A(x, y) = m_2(y).$$

Recall that $d(x, y)$ is the graph distance between two vertices $x$ and $y$. The *transportation distance* between two probability distributions $m_1$ and $m_2$ is defined as follows:

$$W(m_1, m_2) = \inf_A \sum_{x,y\in V} A(x, y)d(x, y),$$

where the infimum is taken over all couplings $A$ between $m_1$ and $m_2$. By the duality theorem of a linear optimization problem, the transportation distance can also be written as follows:

$$W(m_1, m_2) = \sup_f \sum_{x \in V} f(x) \left( m_1(x) - m_2(x) \right),$$

where the supremum is taken over all 1-Lipschitz functions $f$.

A random walk $m$ on $G = (V, E)$ is defined as a family of probability measures $\{m_v(\cdot)\}_{v \in V}$ such that $m_v(u) = 0$ for all $\{v, u\} \notin E$. It follows that $m_v(u) \geq 0$ for all $v, u \in V$ and $\sum_{u \in N(v)} m_v(u) = 1$. The Ricci curvature $\kappa$ of $G$ can then be defined as follows:

**Definition 1.** *Given $G = (V, E)$, a random walk $m = \{m_v(\cdot)\}_{v \in V}$ on $G$ and two vertices $x, y \in V$,*

$$\kappa(x, y) = 1 - \frac{W(m_x, m_y)}{d(x, y)}.$$

**Remark 3.** *We say a graph $G$ equipped with a random walk $m$ has Ricci curvature at least $\kappa_0$ if $\kappa(x, y) \geq \kappa_0$ for all $x, y \in V$.*

For $0 \leq \alpha < 1$, the *$\alpha$-lazy random walk $m_x^\alpha$* (for any vertex $x$), is defined as

$$m_x^\alpha(v) = \begin{cases} \alpha & \text{if } v = x, \\ (1-\alpha)/d(x) & \text{if } v \in \Gamma(x), \\ 0 & \text{otherwise.} \end{cases}$$

In [33], Lin, Lu and Yao defined the Ricci curvature of graphs based on the $\alpha$-lazy random walk as $\alpha$ goes to 1. More precisely, for any $x, y \in V$, they defined the $\alpha$-Ricci-curvature $\kappa_\alpha(x, y)$ to be

$$\kappa_\alpha(x, y) = 1 - \frac{W(m_x^\alpha, m_y^\alpha)}{d(x, y)}$$

and the Ricci curvature $\kappa_{\mathrm{LLY}}$ of $G$ to be

$$\kappa_{\mathrm{LLY}}(x, y) = \lim_{\alpha \to 1} \frac{\kappa_\alpha(x, y)}{(1-\alpha)}.$$

They showed [33] that $\kappa_\alpha$ is concave in $\alpha \in [0, 1]$ for any two vertices $x, y$. Moreover,

$$\kappa_\alpha(x, y) \leq (1-\alpha) \frac{2}{d(x, y)}$$

for any $\alpha \in [0, 1]$ and any two vertices $x$ and $y$.

In the context of graphs, the following lemma shows that it is enough to consider only $\kappa(x, y)$ for $xy \in E(G)$.

**Lemma 1.** *[38, 33] If $\kappa(x, y) \geq \kappa_0$ for any edge $xy \in E(G)$, then $\kappa(x, y) \geq \kappa_0$ for any pair of vertices $(x, y)$.*

## 3. Proof of Theorem 1

We first define an averaging operator associated to the random walk.

**Definition 2** (Discrete averaging operator)**.** *Given a function $f : X \to \mathbb{R}$, let the averaging operator $M$ be defined as*

$$Mf(x) := \sum_{y \in V} f(y) \cdot m_x(y).$$

The following proposition shows a Lipschitz contraction property in the metric measure space. We include its proof here for the sake of completeness.

**Proposition 1** (Lipschitz contraction)**.** *[38, 22] Let $(G, d, m)$ be a random walk on a simple graph $G$. Let $\kappa \in \mathbb{R}$. Then the Ricci curvature of $G$ is at least $\kappa$, if and only if, for every $k$-Lipschitz function $f : X \to \mathbb{R}$, the function $Mf$ is $k(1-\kappa)$-Lipschitz.*

*Proof.* Suppose that the Ricci curvature of $G = (V, E)$ is at least $\kappa$. For $x, y \in V$, let $A : V \times V \to [0, 1]$ be the optimal coupling measure of $m_x$ and $m_y$. Then

$$\begin{aligned}
Mf(y) - Mf(x) &= \sum_{u \in V} f(u) m_y(u) - \sum_{u \in V} f(u) m_x(u) \\
&= \sum_{u \in V} f(u) \sum_{v \in V} A(v, u) - \sum_{u \in V} f(u) \sum_{v \in V} A(u, v) \\
&= \sum_{u,v} (f(v) - f(u)) A(u, v) \\
&\leq k \sum_{u,v} d(u, v) A(u, v) \\
&= kW(m_x, m_y) \\
&= k(1 - \kappa(x, y)) d(x, y).
\end{aligned}$$

Conversely, suppose that whenever $f$ is 1-Lipschitz, $Mf$ is $(1 - \kappa)$-Lipschitz. Then by the duality theorem for the transportation distance, we have that for all $x, y \in V(G)$,

$$\begin{aligned}
W(m_x, m_y) &= \sup_{f \text{ 1-Lipschitz}} \sum_{z \in V} f(z)\left(m_x(z) - m_y(z)\right) \\
&= \sup_{f \text{ 1-Lipschitz}} Mf(x) - Mf(y) \\
&\leq (1-\kappa)d(x,y).
\end{aligned}$$

It follows that

$$\kappa(x,y) = 1 - \frac{W(m_x, m_y)}{d(x,y)} \geq \kappa. \qquad \square$$

**Remark 4.** *Note that for any constant $c$,*

$$\operatorname{Var}(f) = E[(f-c)^2] - \left(E[f]-c\right)^2. \tag{7}$$

*Thus for any $x \in V$ and an $\alpha$-Lipschitz function $f : \operatorname{Supp}(m_x) \to \mathbb{R}$,*

$$\begin{aligned}
\operatorname{Var}_{m_x} f &\leq E_{m_x}[(f - f(x))^2] \\
&\leq \sum_{y \in \operatorname{Supp}(m_x)} (f(y) - f(x))^2 m_x(y) \\
&\leq \alpha^2.
\end{aligned}$$

**Lemma 2.** *[33, 38] Let $G$ be a finite graph with Ricci curvature at least $\kappa > 0$. Then*

$$\kappa \leq \frac{2}{\operatorname{diam}(G)}.$$

*Moreover, if $m_x(x) = \alpha$ for all $x \in V(G)$, then $\kappa \leq (1-\alpha)\frac{2}{\operatorname{diam}(G)}$.*

The following lemma is similar to Lemma 38 in [38].

**Lemma 3.** *Let $\phi : V(G) \to \mathbb{R}$ be an $\alpha$-Lipschitz function with $\alpha \leq 1$. Then for $x \in V(G)$, we have*

$$\left(Me^{\lambda\phi}\right)(x) \leq e^{\lambda M\phi(x) + \frac{1}{2}\lambda^2 e^{2\lambda}\alpha^2}.$$

*Proof.* For any smooth function $g$ and any real-valued random variable $Y$, a Taylor expansion with Lagrange remainder gives

$$Eg(Y) \leq g(EY) + \frac{1}{2}(\sup g'')\operatorname{Var} Y.$$

Applying this with $g(Y) = e^{\lambda Y}$, we get

$$(Me^{\lambda\phi})(x) = E_{m_x} e^{\lambda\phi} \le e^{\lambda M\phi(x)} + \frac{\lambda^2}{2}\left(\sup_{\mathrm{Supp}(m_x)} e^{\lambda\phi}\right) \mathrm{Var}_{m_x}\phi.$$

Note that $\mathrm{diam}(G[\mathrm{Supp}(m_x)]) \le 2$ and $\phi$ is $\alpha$-Lipschitz, it follows that

$$\sup_{\mathrm{Supp}(m_x)} \phi \le E_{m_x}\phi + \alpha \cdot \mathrm{diam}(G[\mathrm{Supp}(m_x)]) \le E_{m_x}\phi + 2\alpha.$$

Moreover, by Remark 4, $\mathrm{Var}_{m_x}\phi \le \alpha^2$. Hence we have that

$$\begin{aligned}\left(Me^{\lambda\phi}\right)(x) &\le e^{\lambda M\phi(x)} + \frac{\lambda^2}{2}(\alpha^2)e^{\lambda M\phi(x)+2\lambda\alpha} \\ &\le e^{\lambda M\phi(x)}\left(1 + \frac{\lambda^2}{2}\alpha^2 e^{2\lambda\alpha}\right) \\ &\le \exp\left(\lambda M\phi(x) + \frac{1}{2}\lambda^2\alpha^2 e^{2\lambda\alpha}\right). \qquad \square\end{aligned}$$

*Proof of Theorem 1.* In the proof, the probability measure $\Pr(\cdot)$ is with respect to a vertex sampling according to the stationary distribution $\nu$.

First, note that since $f$ is 1-Lipschitz, it follows that $|f(x) - f(y)| \le \mathrm{diam}(G)$ for any $x, y \in V(G)$. It follows that for any fixed vertex $x$,

$$\begin{aligned}|f(x) - E_\nu[f]| = \left|\sum_{u\in V(G)} \nu(u)\left(f(x) - f(u)\right)\right| &\le \sum_{u\in V(G)} \nu(u)|f(x) - f(u)| \\ &\le diam(G).\end{aligned}$$

Hence if $t > \frac{2}{\kappa}$, then by Lemma 2,

$$\Pr\left(|f - E_\nu[f]| \ge t\right) \le \Pr\left(|f - E_\nu[f]| > \frac{2}{\kappa}\right) \le \Pr\left(\mathrm{diam}(G) > \frac{2}{\kappa}\right) = 0,$$

in which case we are done. So from now on, assume $t \le 2/\kappa$.

Apply Lemma 3 iteratively and use Proposition 1, we obtain that for any $i \ge 1$,

$$\begin{aligned}M^i(e^{\lambda f}) &\le e^{\lambda M^i f} \cdot \prod_{j=0}^{i-1} \exp\left(\frac{1}{2}\lambda^2(1-\kappa)^{2j}e^{2\lambda}\right) \\ &\le \exp\left(\lambda M^i f + \frac{1}{2}\lambda^2 e^{2\lambda}\sum_{j=0}^{i-1}(1-\kappa)^{2j}\right).\end{aligned}$$

Meanwhile, observe that $(M^i e^{\lambda f})(x)$ converges to $E_\nu e^{\lambda f}$ as $i \to \infty$. Hence

$$\begin{aligned} E_\nu e^{\lambda f} &\leq \lim_{i\to\infty} \exp\left(\lambda M^i f + \frac{1}{2}\lambda^2 e^{2\lambda}\sum_{j=0}^{i-1}(1-\kappa)^{2j}\right) \\ &\leq \exp\left(\lambda E_\nu f + \frac{\lambda^2 e^{2\lambda}}{2\kappa(2-\kappa)}\right). \end{aligned}$$

Let $\lambda_0$ be the root of the equation $x \cdot e^{2x} = 2(2-\kappa)$ and set $\lambda = \frac{t\kappa\lambda_0}{2}$. Note that since $t \leq \frac{2}{\kappa}$, we have $\lambda \leq \lambda_0$. Now, we have

$$\begin{aligned} \Pr\left(f - E_\nu f \geq t\right) &\leq \Pr\left(e^{\lambda f} \geq e^{t\lambda + \lambda E_\nu f}\right) \\ &\leq E_\nu e^{\lambda f} \cdot e^{-t\lambda - \lambda E_\nu f} \\ &\leq \exp\left(-t\lambda + \frac{\lambda^2 e^{2\lambda}}{2\kappa(2-\kappa)}\right) \\ &\leq \exp\left(-t\lambda + \frac{\lambda t \lambda_0 e^{2\lambda}}{4(2-\kappa)}\right) \\ &\leq \exp\left(-t\lambda + \frac{\lambda t \lambda_0 e^{2\lambda_0}}{4(2-\kappa)}\right) \\ &= \exp\left(-\frac{1}{2}t\lambda\right) \\ &\leq \exp\left(-\frac{t^2\kappa\lambda_0}{4}\right). \end{aligned} \tag{8}$$

If $G$ is the complete graph, then $|f - E_\nu(f)| \leq 1$ holds for all vertices. Inequality 3 holds. If $G$ is not the complete graph, then we must have $\kappa \leq 1$ (otherwise, contradiction to $diam(G) \leq \frac{2}{\kappa}$). Thus $\lambda_0 \leq 0.60108...$, which is the root of $x \cdot e^{2x} = 2$. We have $\frac{\lambda_0}{4} > \frac{1}{7}$. Hence we obtain that

$$\Pr\left(f - E_\nu f \geq t\right) \leq \exp\left(-\frac{t^2\kappa}{7}\right).$$

If $\kappa \to 0$ as $|V(G)| \to \infty$ (which is true in all the examples in Section 4), then we have $\lambda_0 \to 0.80290...$ which is the root of $x \cdot e^{2x} = 4$. We have $\frac{\lambda_0}{4} > \frac{1}{5}$. Hence

$$\Pr\left(f - E_\nu f \geq t\right) \leq \exp\left(-\frac{t^2\kappa}{5}\right).$$

Furthermore, if $\kappa \to 0$ and $t\kappa \to 0$ as $|V(G)| \to \infty$, then continuing from inequality (8), we have that $e^{2\lambda} \to 1$ and $(2-\kappa) \to 2$ (as $|V(G)| \to \infty$). By setting $\lambda_0 = 4$, we have

$$\begin{aligned}
\Pr\left(f - E_\nu f \geq t\right) &\leq \exp\left(-t\lambda + \frac{\lambda t \lambda_0 e^{2\lambda}}{4(2-\kappa)}\right) \\
&\leq \exp\left(-\left(\frac{1}{2} + o(1)\right) t\lambda\right) \\
&\leq \exp\left(-\left(\frac{1}{4} + o(1)\right) t^2 \kappa \lambda_0\right) \\
&\leq \exp\left((1+o(1))t^2\kappa\right).
\end{aligned}$$

The lower tail can be obtained from the upper tail by changing $f$ to $-f$ since $-f$ is also 1-Lipschitz. □

## 4. Applications to random models of configurations

In order to apply Theorem 1 to a finite probability space $(\Omega, \mu)$, we will construct a graph $H$ with the vertex set $\Omega$ such that $\mu$ is the invariant distribution over some random walk $m$ on $H$. We call the pair $(H, m)$ a *geometrization* of $(\Omega, \mu)$. For example, consider the classical Erdős-Rényi random graph model $G(n, p)$. In this model, $\Omega$ is the space of all labelled graphs on $n$ vertices, and for any fixed graph $G \in \Omega$, $\mu(G) = p^{e(G)}(1-p)^{\binom{n}{2} - e(G)}$. To geometrize $(\Omega, \mu)$, we construct an auxiliary graph $H$ such that $V(H)$ is the set of all labeled graphs on $n$ vertices, and two graphs $G_1, G_2 \in V(H)$ are adjacent in $H$ if and only if there exists some $v \in V(G_1)(= V(G_2))$ such that $G_1 - v = G_2 - v$. Given $H$, we can then define a random walk $m = \{m_G : G \in V(H)\}$ on $H$, where for each fixed $G$, $m_G$ depends on $p$ and $G$ (see the definition of $m_G$ in Section 4.1) below. It is then not hard to verify that $\mu$, which is the distribution of $\Omega$ in the $G(n, p)$ model, is the invariant distribution of the random walk $m$ on $H$. Given such geometrization $(H, m)$, we can then construct a coupling to compute a lower bound on the graph curvature of $(H, m)$, which then enables us to apply Theorem 1 to obtain a concentration inequality on vertex/edge-Lipschitz functions in the $G(n, p)$ model.

In this section, we will present geometrizations and concentration inequalities in four popular random models of configurations.

### 4.1. Vertex-Lipschitz functions on $G(n, p)$

Let $H$ be the graph such that $V(H)$ is the set of all labeled graphs with $n$ vertices. Moreover, two graphs $G_1, G_2 \in V(H)$ are adjacent in $H$ if and only if there exists some $v$ such that $G_1 - v = G_2 - v$, i.e., $G_2$ can be obtained from $G_1$ by 're-connecting' the edges incident to a single vertex. Now define a random walk $m$ on $H$ as follows: Let $G \in V(H)$. Define

$$m_G(G') = \begin{cases} \frac{1}{n} \sum\limits_{\substack{v \in V(G) \\ G-v=G'-v}} p^{d_{G'}(v)}(1-p)^{n-1-d_{G'}(v)} & \text{if } G' \in N_H(G), \\ 0 & \text{otherwise.} \end{cases}$$

For a fixed $G \in V(H)$, the random walk $m_G$ can be thought of as follows: first pick a vertex $v$ uniformly at random among all vertices of $G$, then re-connect $v$ with the rest of the vertices such that each vertex pair $\{v, u\}$ appears with probability $p$ (for each vertex $u \neq v$).

Observe that

$$\sum_{G' \in V(H)} m_G(G') = \frac{1}{n} \sum_{v \in V(G)} \sum_{k=0}^{n-1} \binom{n-1}{k} p^k (1-p)^{n-1-k} = 1.$$

**Proposition 2.** *Let $\nu$ be the unique invariant distribution of the random walk defined above. A random graph $G$ picked according to $\nu$, satisfies that $\nu(G) = p^{e(G)}(1-p)^{\binom{n}{2}-e(G)}$.*

*Proof.* Observe that $H$ is not bipartite thus the random walk is ergodic. It suffices to show that the distribution $\nu'(G) = p^{e(G)}(1-p)^{\binom{n}{2}-e(G)}$ for every $G$ is an invariant distribution for the random walk. Indeed, for every fixed $G \in V(H)$,

$$\begin{aligned}
&\sum_{G' \in V(H)} \nu'(G') m_{G'}(G) \\
&= \sum_{v \in V(G)} \sum_{G'-v=G-v} \nu'(G') \frac{1}{n} p^{d_G(v)} (1-p)^{n-1-d_G(v)} \\
&= \sum_{v \in V(G)} \frac{1}{n} p^{d_G(v)} (1-p)^{n-1-d_G(v)} \sum_{G'-v=G-v} \nu'(G') \\
&= \sum_{v \in V(G)} \frac{1}{n} p^{d_G(v)} (1-p)^{n-1-d_G(v)} \cdot \\
&\quad \left( p^{e(G)-d_G(v)} (1-p)^{\binom{n-1}{2}-(e(G)-d_G(v))} \sum_{i=0}^{n-1} \binom{n-1}{i} p^i (1-p)^{n-1-i} \right) \\
&= \frac{1}{n} p^{e(G)} (1-p)^{\binom{n}{2}-e(G)} \sum_{v \in V(G)} \sum_{i=0}^{n-1} \binom{n-1}{i} p^i (1-p)^{n-1-i}
\end{aligned}$$

$$=p^{e(G)}(1-p)^{\binom{n}{2}-e(G)}$$
$$=\nu'(G).$$

Therefore, $\nu'$ is equal to $\nu$, the unique invariant distribution for the random walk. ☐

**Lemma 4.** *Let $H$ and the random walk $m$ be defined as above. Then*

$$\kappa(G_1, G_2) \geq \frac{1}{n}$$

*for all $G_1, G_2 \in V(H)$.*

*Proof.* Again, by Lemma 1, we can assume that $G_1, G_2$ are neighbors in $H$. It then follows from definition that

$$\kappa(G_1, G_2) = 1 - W(m_{G_1}, m_{G_2}).$$

Assume that $v$ is the unique vertex such that $G_1 - v = G_2 - v$. When $G_1$ and $G_2$ differ by an edge, it is possible that there are two vertices $v$ satisfying $G_1 - v = G_2 - v$. We remark that the analysis is similar. Consider the support of $m_{G_1}$. For each $G_1' \in \Gamma(G_1)\backslash\{G_2\}$, we will match $G_1'$ with a distinct graph $\phi(G_1') \in N(G_2)$. There are two possible types of neighbors $G_1' \in \Gamma(G_1)\backslash\{G_2\}$:

Type I: $G_1' \in \Gamma(G_1) \cap \Gamma(G_2)$, i.e., $G_1 - v = G_1' - v$. Then it follows that $G_1' - v = G_2 - v$ and we let $\phi(G_1') = G_1'$.

Type II: $G_1' \in \Gamma(G_1)\backslash(\Gamma(G_2) \cup \{G_2\})$, which implies that $G_1 - u = G_1' - u$ for some $u \neq v$. In this case, we claim that for each $G_1'$ of Type II, there exists a unique $G_2' = \phi(G_1') \in \Gamma(G_2)\backslash\{\Gamma(G_1) \cup \{G_1\}\}$ such that $G_2' - u = G_2 - u$ and $G_1' - v = G_2' - v$. Indeed, let $G_2'$ be obtained from $G_2$ by replacing the neighbors of $u$ in $G_2$ by the neighbors of $u$ in $G_1'$. It's not hard to see that $G_2' - u = G_2 - u$ and $G_1' - v = G_2' - v$.

It is not hard to see that $\phi$ is a bijection between $\Gamma(G_1)\backslash\{G_2\}$ and $\Gamma(G_2)\backslash\{G_1\}$. Moreover, for every $G_1'$ of Type II, we have that $m_{G_1}(G_1') = m_{G_2}(\phi(G_1'))$ since $d_{G_1'}(u) = d_{\phi(G_1')}(u)$. Note that for $G_1'$ of type I (in which case $G_1' = \phi(G_1')$), it is not necessarily true that $m_{G_1}(G_1') = m_{G_2}(G_1')$. This happens when $G_1'$ differs by an edge $vu$ from $G_1$ or $G_2$ for some $u \neq v$; hence if $G_1' = G_1 \pm vu$, then $m_{G_1}(G_1') \geq m_{G_2}(G_1')$ and if $G_1' = G_2 \pm vu$, then $m_{G_1}(G_1') \leq m_{G_2}(G_1')$. Let us call a Type I graph $G_1'$ with $m_{G_1}(G_1') > m_{G_2}(G_1')$ a *Type I-A* graph, and call a type I graph $G_1'$ with $m_{G_1}(G_1') < m_{G_2}(G_1')$ a *Type I-B* graph.

Let us now define a coupling $A$ (not necessarily optimal) between $m_{G_1}$ and $m_{G_2}$. Define first $A : V(H) \times V(H) \to \mathbb{R}$ for the following pairs of $G_1', G_2'$.

(9)

$$A(G_1', G_2') = \begin{cases} m_{G_2}(G_1) & \text{if } G_1' = G_2' = G_1, \\ m_{G_1}(G_2) & \text{if } G_1' = G_2' = G_2, \\ \min(m_{G_1}(G_1'), m_{G_2}(G_2')) & \text{if } G_1' \in \Gamma(G_1)\backslash\{G_2\} \\ & \quad \text{and } G_2' = \phi(G_1'), \\ m_{G_1}(G_1') - m_{G_2}(G_1') & \text{if } G_1' \text{ is Type I-A and } G_2' = G_2, \\ m_{G_2}(G_2') - m_{G_1}(G_2') & \text{if } G_2' \text{ is Type I-B and } G_1' = G_1, \\ 0 & \text{otherwise if } (G_1', G_2') \neq (G_1, G_2). \end{cases}$$

The only pair $(G_1', G_2')$ for which $A(G_1', G_2')$ is undefined is when $(G_1', G_2') = (G_1, G_2)$. In this case, define

$$A(G_1, G_2) = m_{G_1}(G_1) - m_{G_2}(G_1) - \sum_{G_1' \text{ is type I-B}} (m_{G_2}(G_1') - m_{G_1}(G_1'))\,.$$

Moreover, since $\sum_{G'} m_{G_1}(G') = \sum_{G'} m_{G_2}(G') = 1$, we have that

$$A(G_1, G_2) = m_{G_2}(G_2) - m_{G_1}(G_2) + \sum_{G_1' \text{ is type I-A}} (m_{G_1}(G_1') - m_{G_2}(G_1'))\,.$$

We claim that $A$ is a coupling between $m_{G_1}$ and $m_{G_2}$. We verify that for each $G \in N_H(G_1)$, $\sum_{G' \in V(H)} A(G, G') = m_{G_1}(G)$ and remark that the other side is similar. First, $\sum_{G' \in V(H)} A(G_1, G') = A(G_1, G_1) + A(G_1, G_2) + \sum_{G_1' \text{ is type I-B}} (m_{G_2}(G_1') - m_{G_1}(G_1')) = m_{G_1}(G_1)$. For each $G$ of Type II, $\sum_{G' \in V(H)} A(G, G') = A(G, \phi(G)) = m_{G_1}(G)$. For each $G$ of Type I-A, we have $\sum_{G' \in V(H)} A(G, G') = A(G, \phi(G)) + m_{G_1}(G) - m_{G_2}(G) = m_{G_1}(G)$. For each $G$ not of Type I-A, $\sum_{G' \in V(H)} A(G, G') = A(G, \phi(G)) = m_{G_1}(G)$. Finally, $\sum_{G' \in V(H)} A(G_2, G') = A(G_2, G_2) = m_{G_1}(G_2)$.

Recall that $W(m_{G_1}, m_{G_2}) \leq \sum_{G_1', G_2'} A(G_1', G_2') d(G_1', G_2')$. Observe that in our coupling defined above, for every $G_1', G_2$ such that $A(G_1', G_2) \neq 0$, $d(G_1', G_2') \leq 1$. Moreover, $\sum_{G_1', G_2'} A(G_1', G_2') = 1$. It follows that

$$\begin{aligned} \kappa(G_1, G_2) &\geq 1 - W(m_{G_1}, m_{G_2}) \\ &\geq \sum_{\substack{G_1' \in N_H(G_1), G_2' \in N_H(G_2) \\ d(G_1', G_2') = 0}} A(G_1', G_2'). \end{aligned}$$

Now observe that the $(G_1', G_2')$ pairs for which $d(G_1', G_2') = 0$ are precisely the pairs $(G_1', G_2')$ such that $G_1' = G_2'$ and $G_1' - v = G_1 - v = G_2 - v$. Hence we have that

$$\begin{aligned}\kappa(G_1, G_2) &\geq \sum_{\substack{G_1' \in N_H(G_1), G_2' \in N_H(G_2) \\ d(G_1', G_2') = 0}} A(G_1', G_2') \\ &\geq \frac{1}{n} \sum_{i=0}^{n-1} \binom{n-1}{i} p^i (1-p)^{n-1-i} \\ &\geq \frac{1}{n}.\end{aligned}$$

□

We remark that the lower bound $\frac{1}{n}$ in Lemma 4 is tight up to a constant factor of 2. Observe that $\mathrm{diam}(H) = n - 1$ since the distance between the empty graph and the complete graph is $n-1$ in $H$. By Lemma 2, the curvature $\kappa$ of $(H, m)$ is at most $\frac{2}{n-1}$.

Note that any 1-Lipschitz function $f$ on $(H, m)$ is vertex Lipschitz. Hence, it follows by Theorem 1 that for any vertex-Lipschitz function $f$ on graphs, if we sample an $n$-vertex graph according to the $G(n, p)$ model, then

$$\Pr\left(|f - E[f]| \geq t\right) \leq 2 \exp\left(-\frac{t^2}{5n}\right),$$

which in this context has the same strength as the Azuma-Hoeffding inequality on vertex-exposure martingale.

### 4.2. Edge-Lipschitz functions on $G(n, M)$

Let $G \sim G(n, M)$ be a random graph with $n$ vertices and $M$ edges. Let $H$ be the graph such that $V(H)$ is the set of all labeled graphs with $n$ vertices and $M$ edges. Moreover, two graphs $G_1, G_2 \in V(H)$ are adjacent in $H$ if and only if there exist two distinct vertex pairs $e_1, e_2$ such that $e_1 \in E(G_1) \backslash E(G_2)$, $e_2 \in E(G_2) \backslash E(G_1)$ and $G_1 - e_1 = G_2 - e_2$. In other words, $G_1, G_2$ are adjacent in $H$ if one can be obtained from the other by swapping an edge with a non-edge. It is easy to see that $H$ is a connected regular graph. Moreover, for every $G \in V(H)$, $d_H(G) = M\left(\binom{n}{2} - M\right)$.

The following proposition is clear from the definition of $H$.

**Proposition 3.** *If $G_1, G_2$ are adjacent in $H$, then there exists a unique pair of distinct vertex pairs $e_1, e_2$ such that $e_1 \in E(G_1) \backslash E(G_2)$, $e_2 \in E(G_2) \backslash E(G_1)$ and $G_1 - e_1 = G_2 - e_2$.*

Now define a random walk $m$ on $H$ as follows: Let $G \in V(H)$. Define

$$m_G(G') = \begin{cases} \frac{1}{M(\binom{n}{2}-M)+1} & \text{if } G' \in N_H(G), \\ 0 & \text{otherwise.} \end{cases}$$

It's easy to see that for any fixed $G$, $\sum_{G'} m_G(G') = 1$. Moreover, $m$ is simply a random walk such that the transition probability from a graph $G$ to each graph in $N_H(G)$ is equal.

**Proposition 4.** *Let $\nu$ be the unique invariant distribution of the random walk defined above. A random graph $G$ picked according to $\nu$, is equally likely to be one of the $\binom{\binom{n}{2}}{M}$ graphs that have $M$ edges.*

*Proof.* Observe that $H$ is not bipartite thus the random walk is ergodic. It suffices to show that $\nu'(G) = \binom{\binom{n}{2}}{M}^{-1}$ for every $G$ is an invariant distribution for the random walk. Indeed, for every fixed $G \in V(H)$,

$$\begin{aligned} \sum_{G' \in V(H)} \nu'(G') m_{G'}(G) &= \binom{\binom{n}{2}}{M}^{-1} \sum_{G' \in N(G)} m_{G'}(G) \\ &= \binom{\binom{n}{2}}{M}^{-1} \sum_{G' \in N(G)} m_G(G') \\ &= \binom{\binom{n}{2}}{M}^{-1} \\ &= \nu'(G). \end{aligned}$$

Since $\nu$ is the unique invariant distribution, it follows then that $\nu = \nu'$. □

**Lemma 5.** *Let $H$ and the random walk $m$ be defined as above. Then*

$$\kappa(G_1, G_2) \geq \frac{\binom{n}{2}}{M\left(\binom{n}{2} - M\right) + 1}$$

*for all $G_1, G_2 \in H$.*

*Proof.* By Lemma 1, we can assume that $G_1, G_2$ are neighbors in $H$. It then follows from definition that

$$\kappa(G_1, G_2) = 1 - W(m_{G_1}, m_{G_2}).$$

Suppose $e_1, e_2$ are the unique vertex pairs with $e_1 \in E(G_1), e_2 \notin E(G_1)$ such that $G_2 = G_1 - e_1 + e_2$. Consider the support of $m_{G_1}$, i.e., $N(G_1)$. For each $G_1' \in N(G_1)$, we will match $G_1'$ with a distinct graph $\phi(G_1') \in N(G_2)$. First, let $\phi(G_1) = G_1$ and $\phi(G_2) = G_2$. For other neighbors $G_1' \in N(G_1)$, there are three types:

Type 1: $G_1 - e_1 = G_1' - e_3$ for some $e_3 \neq e_2$. Then it follows that $G_1' - e_3 = G_2 - e_2$ and we let $\phi(G_1') = G_1'$.

Type 2: $G_1 - e_3 = G_1' - e_2$ for some $e_3 \neq e_1$. Then it follows that $G_1' - e_1 = G_2 - e_3$ and we let $\phi(G_1') = G_1'$.

Type 3: $G_1 - e_3 = G_1' - e_4$ for some $e_3, e_4 \notin \{e_1, e_2\}$. In this case, we claim that there exists a unique $G_2' = \phi(G_1') \in N(G_2)$ such that $G_1' - e_1 = G_2' - e_2$. Indeed, $G_2' = G_2 - e_3 + e_4$ will satisfy the aforementioned property.

Let us now define a coupling $A$ (not necessarily optimal) between $m_{G_1}$ and $m_{G_2}$. Define $A : V(H) \times V(H) \to \mathbb{R}$ as follows:

$$A(G_1', G_2') = \begin{cases} \frac{1}{M(\binom{n}{2} - M) + 1} & \text{if } G_1' \in N(G_1) \text{ and } G_2' = \phi(G_1'), \\ 0 & \text{otherwise.} \end{cases} \tag{10}$$

Let us verify that $A$ is a coupling of $m_{G_1}$ and $m_{G_2}$. Indeed, for each fixed $G_1'$, if $G_1' = G_1$, then $\sum_{G_2'} A(G_1', G_2') = A(G_1, G_1) = m_{G_1}(G_1)$; if $G_1' \neq G_1$, then $\sum_{G_2'} A(G_1', G_2') = A(G_1', \phi(G_1')) = m_{G_1}(G_1')$. Similarly, $\sum_{G_1'} A(G_1', G_2') = m_{G_2}(G_2')$. Now by definition,

$$\begin{aligned} W(m_{G_1}, m_{G_2}) &\leq \sum_{G_1', G_2'} A(G_1', G_2') d(G_1', G_2') \\ &\leq \sum_{G_1' \in N(G_1)} A(G_1', \phi(G_1')) d(G_1', \phi(G_1')) \\ &= \sum_{\substack{G_1' \in N(G_1) \\ G_1' \text{ is Type 3}}} A(G_1', \phi(G_1')) \\ &\leq \left( (M-1) \left( \binom{n}{2} - M - 1 \right) \right) \cdot \frac{1}{M\left(\binom{n}{2} - M\right) + 1}. \end{aligned}$$

It follows that

$$\begin{aligned} \kappa(G_1, G_2) &= 1 - W(m_{G_1}, m_{G_2}) \\ &\geq \frac{\binom{n}{2}}{M\left(\binom{n}{2} - M\right) + 1}. \end{aligned}$$

□

We remark that the lower bound in Lemma 5 is asymptotically tight. Observe that $\mathrm{diam}(H) = \min\{M, \binom{n}{2} - M\}$. By Lemma 2, the curvature $\kappa$ of $(H, m)$ is at most $\frac{2}{\min\{M, \binom{n}{2} - M\}}$.

Let $G(n, M)$ be an Erdős-Rényi random graph with $M$ edges. Let $F$ be a fixed graph and $X_F$ be the number of copies of $F$ in the random graph $G(n, M)$. Denote the number of vertices and edges of $F$ by $v(F)$ and $e(F)$ respectively. Let $p = M/\binom{n}{2}$ and $\mathrm{Aut}(F)$ denote the set of automorphisms of $F$. Then

$$E[X_F] = (1 + o(1)) \frac{v(F)!}{|\mathrm{Aut}(F)|} \binom{n}{v(F)} p^{e(F)} = \Theta\left(n^{v(F)} p^{e(F)}\right).$$

For a series of results on the upper tail of $X_F$ using different techniques, see, for example, the survey [30] and the papers [1, 7, 20, 21, 26, 29]. For $G(n, M)$ in particular, Janson, Oleszkiewicz, Ruciński [29] showed the following theorem:

**Theorem 2.** *[29] For every graph $F$ and for every $t > 1$, there exist constants $c(t, F) > 0$ such that for all $n \ge v(F)$ and $e(F) \le M \le \binom{n}{2}$, with $\mathrm{p} := M/\binom{n}{2}$,*

$$\Pr\left(X_F \ge tE[X_F]\right) \le \exp\left(-c(t, F) M_F^*(n, p)\right),$$

*where $M_F^*(n, p) \le n^2 p = O(M), M_{C_k}^*(n, p) = \Theta(n^2 p^2)$ and $M_{K_k}^*(n, p) = \Theta(n^2 p^{k-1})$.*

Let us now apply Theorem 1 to obtain the concentration results from the perspective of the Ricci curvature. Recall that $H$ is defined as the graph such that $V(H)$ is the set of all labeled graphs with $n$ vertices and $M$ edges. Moreover, two graphs $G_1, G_2 \in V(H)$ are adjacent in $H$ if and only if there exist two distinct vertex pairs $e_1, e_2$ such that $e_1 \in E(G_1) \backslash E(G_2)$, $e_2 \in E(G_2) \backslash E(G_1)$ such that $G_1 - e_1 = G_2 - e_2$.

Again let $X_F$ be the random variable denoting the number of copies of $F$ in $G(n, M)$. For ease of reference, let $k = v(F)$. Observe that $X_F$ is $\binom{n}{k-2}$-Lipschitz on $H$, i.e., if $G_1, G_2$ are adjacent in $H$, then $|X_F(G_1) - X_F(G_2)| \le \binom{n}{k-2}$. Thus by Theorem 1,

$$\Pr\left(\frac{X_F}{\binom{n}{k-2}} > \frac{E[X_F]}{\binom{n}{k-2}} + \frac{t}{\binom{n}{k-2}}\right) \le \exp\left(-\frac{t^2 \kappa}{5\binom{n}{k-2}^2}\right).$$

It follows that

$$\Pr\left(X_F > E[X_F] + t\right) \le \exp\left(-\frac{t^2\kappa}{5\binom{n}{k-2}^2}\right).$$

Let $p = M/\binom{n}{2}$. We then obtain that

$$\Pr\left(X_F \ge tE[X_F]\right) \le \exp\left(-\frac{\left((t-1)E[X_F]\right)^2\kappa}{5\binom{n}{k-2}^2}\right) \tag{11}$$

$$\le \exp\left(-C_k(t-1)^2 n^2 p^{2e(F)-1}\right). \tag{12}$$

Note that when $p = \Theta(1)$, i.e., $M = \Theta\left(\binom{n}{2}\right)$, the concentration inequalities obtained from Theorem 1 has the same asymptotic exponent as Theorem 2. For other ranges of $p$ with $n^2p \to \infty$, the asymptotic exponent in (12) is worse than the bound in Theorem 2. Nonetheless, let us compare the bounds obtained from the Ricci curvature method with those obtained from other concentration inequalities. Janson and Ruciński [30] surveyed the existing techniques on estimating the exponents for upper tails in the small subgraphs problem in $G(n,p)$ (ignoring logarithmic factors). Please see Table 2 of [30] for a detailed summary and exposition.

Although we are mainly dealing with $G(n,M)$ in this section, it is well known that $G(n,M)$ and $G(n,p)$ with $p = M/\binom{n}{2}$ behaves similarly when $n^2p \to \infty$. Applying the inequalities in (12) to $K_3, K_4, C_4$ respectively, we have that the exponents (ignoring constant) obtained from the Ricci curvature method are $n^2p^5$, $n^2p^{11}$ and $n^2p^7$ respectively. In this context, the concentration we obtained from Theorem 1 has the same strength as Talagrand inequality [41] and slightly stronger than Azuma's inequality (see Table 1).

Table 1: Comparison between (12), Azuma's and Talagrand's Inequality

| Exponents for Upper Tails in the Small Subgraphs Problem | | | |
|---|---|---|---|
| Inequalities | $K_3$ | $K_4$ | $C_4$ |
| Azuma [3, 27] | $n^2p^6$ | $n^2p^{12}$ | $n^2p^8$ |
| Talagrand [41] | $n^2p^5$ | $n^2p^{11}$ | $n^2p^7$ |
| Ricci curvature | $n^2p^5$ | $n^2p^{11}$ | $n^2p^7$ |

### 4.3. Edge-Lipschitz functions on random hypergraphs

Let $\mathcal{H} \sim \mathcal{H}^k(n,M)$ be a random $k$-uniform hypergraph with $n$ vertices and $M$ edges. Let $H$ be a graph such that $V(H)$ is the set of all labeled $k$-uniform hypergraphs with $n$ vertices and $M$ edges. Moreover, two hypergraphs $\mathcal{H}_1, \mathcal{H}_2 \in V(H)$ are adjacent in $H$ if and only if there exist two distinct $k$-sets

$h_1, h_2$ such that $h_1 \in E(\mathcal{H}_1)\backslash E(\mathcal{H}_2)$, $h_2 \in E(\mathcal{H}_2)\backslash E(\mathcal{H}_1)$ and $\mathcal{H}_1 - h_1 = \mathcal{H}_2 - h_2$. In other words, $\mathcal{H}_1, \mathcal{H}_2$ are adjacent in $H$ if one can be obtained from the other by swapping a hyperedge with a non-hyperedge. It is easy to see that $H$ is a connected regular graph. Moreover, for every $\mathcal{H} \in V(H)$, $d_H(\mathcal{H}) = M\left(\binom{n}{k} - M\right)$. Now define a random walk $m$ on $H$ as follows: Let $\mathcal{H} \in V(H)$. Define

$$m_{\mathcal{H}}(\mathcal{H}') = \begin{cases} \frac{1}{M\left(\binom{n}{k}-M\right)+1} & \text{if } \mathcal{H}' \in \Gamma(\mathcal{H}), \\ 0 & \text{otherwise.} \end{cases}$$

As before, for any fixed $\mathcal{H}$, $\sum_{\mathcal{H}'} m_{\mathcal{H}}(\mathcal{H}') = 1$. Moreover, $m$ is simply a random walk such that the transition probability from a hypergraph $\mathcal{H}$ to each hypergraph in $N_H(\mathcal{H})$ is equal.

By the same logic in Section 4.2, we can obtain a lower bound for the Ricci curvature of $H$, i.e., for all $\mathcal{H}_1, \mathcal{H}_2 \in V(H)$,

$$\kappa(\mathcal{H}_1, \mathcal{H}_2) \geq \frac{\binom{n}{k}}{M\left(\binom{n}{k} - M\right) + 1}.$$

Similar to before, we can also apply Theorem 1 to obtain concentration results for the number of copies of fixed sub-hypergraphs in a uniformly random hypergraph on $n$ vertices and $M$ edges. The idea is similar to Section 4.2 and we leave the details to the readers.

### 4.4. Vertex-Lipschitz functions on random $d$-out(in)-regular graphs

Given a directed graph $G$ and a vertex $v$, we use $\delta^+(v)$ and $\delta^-(v)$ to denote the outdegree and indegree, respectively, of a vertex $v$. A *$d$-out-regular graph* $G$ is a directed graph in which $\delta^+(v) = d$ for every $v \in V(G)$. Similarly, a *$d$-in-regular graph* $G$ is a directed graph in which $\delta^-(v) = d$ for every $v \in V(G)$. Moreover, let $\Gamma^+(v) = \{u \in V(G) : vu \in E(G)\}$, $\Gamma^-(v) = \{u \in V(G) : uv \in E(G)\}$, $N^+(v) = \Gamma^+(v) \cup \{v\}$ and $N^-(v) = \Gamma^-(v) \cup \{v\}$.

Let $H$ be a graph such that $V(H)$ is the set of all labeled $d$-out-regular graphs on $n$ vertices. Two graphs $G_1, G_2 \in V(H)$ are adjacent in $H$ if and only if there exists some vertex $v \in V(G_1) = V(G_2)$ such that one can be obtained from the other by changing $\Gamma^+(v)$. It is not hard to see that $H$ is a connected graph with $\mathrm{diam}(H) \leq n$. Moreover, it is also clear that if $G_1, G_2$ are adjacent in $H$, there is a unique vertex $v$ such that one can be obtained from the other by changing $\Gamma^+(v)$.

Now define a random walk $m$ on $H$ as follows: let $G \in V(H)$ and define

$$m_G(G') = \begin{cases} \frac{1}{n(\binom{n-1}{d}-1)+1} & \text{if } G' \in N^+(G), \\ 0 & \text{otherwise.} \end{cases}$$

It's easy to see that for a fixed $G$, $\sum_{G'} m_G(G') = 1$. Moreover, $m$ is simply a random walk such that the transition probability from a graph $G$ to each graph in $N_H(G)$ is equal.

**Proposition 5.** *Let $\nu$ be the unique invariant distribution of the random walk defined above. A random graph $G$ picked according to $\nu$, is equally likely to be one of the $d$-out-regular graphs on $n$ vertices.*

*Proof.* Observe that $H$ is not bipartite thus the random walk is ergodic. There are $\binom{n-1}{d}^n$ many $d$-out-regular graphs in total. Hence, it suffices to show that $\nu'(G) = \binom{n-1}{d}^{-n}$ for every $G$ is an invariant distribution for the random walk. Indeed, for every fixed $G \in V(H)$,

$$\begin{aligned}\sum_{G' \in H} \nu'(G') m_{G'}(G) &= \binom{n-1}{d}^{-n} \sum_{G' \in H} m_{G'}(G) \\ &= \binom{n-1}{d}^{-n} \sum_{G' \in H} m_G(G') \\ &= \binom{n-1}{d}^{-n} \\ &= \nu'(G).\end{aligned}$$

Since $\nu$ is the unique invariant distribution, it follows then that $\nu = \nu'$. $\square$

**Lemma 6.** *Let $H$ and the random walk $m$ be defined as above. Then*

$$\kappa(G_1, G_2) \geq \frac{1}{n}$$

*for all $G_1, G_2 \in V(H)$.*

*Proof.* Again, by Lemma 1, we can assume that $G_1, G_2$ are neighbors in $H$. It then follows from definition that

$$\kappa(G_1, G_2) = 1 - W(m_{G_1}, m_{G_2}).$$

Suppose $v$ is the unique vertex such that $G_2$ can be obtained from $G_1$ by changing $\Gamma^+(v)$. Consider the support of $m_{G_1}$. For each $G_1' \in N(G_1)$, we will match $G_1'$ with a distinct graph $\phi(G_1') \in N(G_2)$. Again, let $\phi(G_1) = G_1$ and $\phi(G_2) = G_2$. For other neighbors $G_1'$ of $G_1$, there are two possible cases:

Case 1: $G_1 - v = G_1' - v$. Then it follows that $G_1' - v = G_2 - v$ and we let $\phi(G_1') = G_1'$.

Case 2: $G_1 - u = G_1' - u$ for some $u \neq v$. In this case, we claim that for each $G_1'$ such that $G_1 - u = G_1' - u$, there exists a unique $G_2' = \phi(G_1')$ such that $G_2' - u = G_2 - u$ and $G_1' - v = G_2' - v$. Indeed, let $G_2'$ be obtained from $G_2$ by replacing the out-neighbors of $u$ in $G_2$ by the out-neighbors of $u$ in $G_1'$. It's not hard to see that $G_2' - u = G_2 - u$ and $G_1' - v = G_2' - v$.

Let us now define a coupling $A$ (not necessarily optimal) between $m_{G_1}$ and $m_{G_2}$. Define $A : V(H) \times V(H) \to \mathbb{R}$ as follows:

$$A(G_1', G_2') = \begin{cases} \frac{1}{n(\binom{n-1}{d} - 1) + 1} & \text{if } G_1' \in N(G_1) \text{ and } G_2' = \phi(G_1'), \\ 0 & \text{otherwise.} \end{cases} \tag{13}$$

It is not hard to verify that $A$ is a coupling of $m_{G_1}$ and $m_{G_2}$. Now by definition,

$$\begin{aligned} W(m_{G_1}, m_{G_2}) &\leq \sum_{G_1', G_2'} A(G_1', G_2') d(G_1', G_2') \\ &\leq \sum_{u \neq v} \sum_{\substack{G_1' \in N(G_1) \\ G_1' - u = G_1 - u}} A(G_1', \phi(G_1')) d(G_1', \phi(G_1')) \\ &\leq (n-1) \left( \binom{n-1}{d} - 1 \right) \frac{1}{n \left( \binom{n-1}{d} - 1 \right) + 1}. \end{aligned}$$

It follows that

$$\kappa(G_1, G_2) = 1 - W(m_{G_1}, m_{G_2}) \geq \frac{\binom{n-1}{d}}{n \left( \binom{n-1}{d} - 1 \right) + 1} \geq \frac{1}{n}.$$

This completes the proof of the lemma. $\square$

We remark that the lower bound in Lemma 6 is tight up to a constant of 2. Observe that $\operatorname{diam}(H) = n$. By Lemma 2, the curvature $\kappa$ of $(H, m)$ is at most $\frac{2}{n}$.

Let $G$ be a uniformly random $d$-out-regular graph. A *directed triangle* is a cycle of length 3 with vertices $u, v, w$ such that $uv, vw$ and $wu$ are all directed edges. Let $X_{n,d} := X(G)$ be the random variable denoting the number of directed triangle in $G$. It is not hard to see that

$$E[X_{n,d}] = (2 + o(1))\binom{n}{3}\left(\frac{d}{n-1}\right)^3,$$

since there are $\binom{n}{3}$ vertex triples, and the probability that a fixed vertex triple forms a directed triangle is $(2 + o(1))\left(\frac{d}{n-1}\right)^3$. We will now use Theorem 1 to derive the concentration behavior of $X_{n,d}$. Note that $X_{n,d}$ is $(d^2)$-Lipschitz. Hence by Theorem 1, we have that

$$\Pr\left(\left|\frac{X_{n,d}}{d^2} - \frac{E[X_{n,d}]}{d^2}\right| > \frac{t}{d^2}\right) \leq 2\exp\left(-\frac{t^2\kappa}{5d^4}\right).$$

It follows that

$$\Pr\left(|X_{n,d} - E[X_{n,d}]| > t\right) \leq 2\exp\left(-\frac{t^2\kappa}{5d^4}\right) \leq 2\exp\left(-\frac{t^2}{5nd^4}\right).$$

This concentration result is useful when $d = \Omega(\sqrt{n})$.

### 4.5. Lipschitz functions on random linear permutations

We will denote a linear permutation $\sigma$ by $\sigma = [a_1a_2\dots a_n]$ such that $a_i \in [n]$ for all $i$ and $\sigma(i) = a_i$. A linear permutation on $[n]$ can be viewed as a sequence of $n$ distinct numbers from $[n]$. Thus, without loss of generality suppose $\{a_1, a_2, \dots, a_n\} = [n]$. Given two permutations $\sigma_1, \sigma_2$ where $\sigma_1 = [a_1a_2\dots a_n]$, we say $\sigma_1$ is $(i, j)$-*alike* to $\sigma_2$ if $\sigma_2$ can be obtained from $\sigma_1$ by moving the number $i$ to the position after the number $j$ in $\sigma_1$; moreover, $\sigma_1$ is $(i, 0)$-alike to $\sigma_2$ if $\sigma_2$ can be obtained from $\sigma_1$ by moving the number $i$ to the first position of $\sigma_1$. For example, $\sigma_1 = [12345]$ is $(2, 4)$-alike to $\sigma_2 = [13425]$ and is $(4, 0)$-alike to $\sigma_3 = [41235]$. Two *distinct* linear permutations $\sigma_1, \sigma_2$ are *insertion-alike* if one is $(i, j)$-alike to the other for some $i \neq j$. Such insertion operations are related to sorting algorithms, e.g., insertion/shell sortings.

Let $H$ be the graph such that $V(H)$ is the set of all linear permutations of $[n]$ and two linear permutation $\sigma_1, \sigma_2$ are adjacent in $H$ if and only if they are insertion-alike. Clearly $H$ is a connected graph with diameter at most $n$. Moreover, every vertex (which is a linear permutation) in $H$ has $(n-1)^2$ neighbors in $H$.

Now define a random walk $m_\alpha$ on $H$ as follows: let $\sigma \in V(H)$ and define

$$m_\sigma(\sigma') = \begin{cases} \frac{1}{(n-1)^2+1} & \text{if } \sigma = \sigma' \text{ or } \sigma \text{ is insertion-alike to } \sigma', \\ 0 & \text{otherwise.} \end{cases}$$

It's not hard to see that for a fixed $\sigma$, $\sum_{\sigma'} m_\sigma(\sigma') = 1$. Moreover, $m_\sigma(\sigma') = m_{\sigma'}(\sigma)$ for every pair of $\sigma, \sigma'$. Observe that $m$ is simply a random walk such that the transition probability from a permutation $\sigma$ to each permutation in $N_H(\sigma)$ is equal.

**Proposition 6.** *Let $\nu$ be the unique invariant distribution of the random walk defined above. A random permutations $\sigma$ picked according to $\nu$, is equally likely to be one of the $n!$ permutations.*

*Proof.* Observe that $H$ is not bipartite thus the random walk is ergodic. There are $n!$ permutations in total. Hence, it suffices to show that $\nu'(\sigma) = (n!)^{-1}$ for every $\sigma$ is an invariant distribution for the random walk.

$$\begin{aligned} \sum_{\sigma' \in H} \nu'(\sigma') m_{\sigma'}(\sigma) &= \frac{1}{n!} \sum_{\sigma' \in V(H)} m_{\sigma'}(\sigma) \\ &= \frac{1}{n!} \sum_{\sigma' \in V(H)} m_\sigma(\sigma') \\ &= \frac{1}{n!} \\ &= \nu'(\sigma). \end{aligned}$$

Since $\nu$ is the unique invariant distribution, it follows then that $\nu = \nu'$. □

**Lemma 7.** *Let $H$ and the random walk $m$ be defined as above. If $\sigma_1, \sigma_2 \in V(H)$ are neighbors in $H$, then $\kappa(\sigma_1, \sigma_2) \geq \frac{1}{n}$.*

*Proof.* Suppose that $\sigma_1$ is $(i,j)$-alike to $\sigma_2$ (with $\sigma_2 \neq \sigma_1$). Consider the support of $m_{\sigma_1}$. For each $\sigma_1' \in N(\sigma_1)$, we will match $\sigma_1'$ with a distinct permutation $\phi(\sigma_1') \in N(\sigma_2)$. First let $\phi(\sigma_1) = \sigma_1$ and $\phi(\sigma_2) = \sigma_2$. For other neighbors $\sigma_1'$ of $\sigma_1$, there are two cases:

Case 1: $\sigma_1$ is $(i,k)$-alike to $\sigma_1'$ where $k \neq j$. Then it follows that $\sigma_1'$ is also $(i,j)$-alike to $\sigma_2$ and we let $\phi(\sigma_1') = \sigma_1'$.

Case 2: $\sigma_1$ is $(i',j')$-alike to $\sigma_1'$ where $i' \neq i$ and $\sigma_1$ is not $(i,k)$-alike to $\sigma_1'$ for any $k$. In this case, let $\sigma_2'$ be the permutation such that $\sigma_2$ is $(i',j')$-alike to $\sigma_2'$. It follows easily that $\sigma_1'$ is also $(i,j)$-alike to $\sigma_2'$. We then define $\phi(\sigma_1') = \sigma_2'$.

Let us now define a coupling $A$ (not necessarily optimal) between $m_{\sigma_1}$ and $m_{\sigma_2}$. Define $A : V(H) \times V(H) \to \mathbb{R}$ as follows:

$$A(\sigma_1', \sigma_2') = \begin{cases} \frac{1}{(n-1)^2+1} & \text{if } \sigma_1' \in N(\sigma_1) \text{ and } \sigma_2' = \phi(\sigma_1'), \\ 0 & \text{otherwise.} \end{cases} \tag{14}$$

It is not hard to verify that $A$ is a coupling of $m_{\sigma_1}$ and $m_{\sigma_2}$. Now by definition,

$$\begin{aligned} W(m_{\sigma_1}, m_{\sigma_2}) &\leq \sum_{\sigma_1', \sigma_2'} A(\sigma_1', \sigma_2') d(\sigma_1', \sigma_2') \\ &\leq \sum_{\sigma' \in N(\sigma_1)} A(\sigma_1', \phi(\sigma_1')) d(\sigma_1', \phi(\sigma_1')) \\ &\leq 1 - \frac{n}{(n-1)^2+1}. \end{aligned}$$

It follows that

$$\kappa(\sigma_1, \sigma_2) = 1 - W(m_{\sigma_1}, m_{\sigma_2}) \geq \frac{n}{(n-1)^2+1} \geq \frac{1}{n}.$$

This completes the proof of the lemma. ☐

We remark that the lower bound in Lemma 7 is tight up to a constant of 2. Observe that $\mathrm{diam}(H) = n-1$. By Lemma 2, The curvature $\kappa$ of $(H, m)$ is at most $\frac{2}{n}$.

Now we give an example of concentration results on the space of random linear permutations. In particular, we discuss the number of occurrences of certain patterns in random permutations. Denote the set of length $n$ linear permutations by $\mathcal{S}_n$. Given a permutation pattern $\tau \in \mathcal{S}_k$, we say that a permutation $\pi = [\pi_1 \ldots \pi_n] \in \mathcal{S}_n$ contains the pattern $\tau$ if there exists $1 \leq i_1 < i_2 < \ldots < i_k \leq n$ such that the $\pi_{i_s} < \pi_{i_t}$ if and only if $\tau_s < \tau_t$ for every pair $s, t$. Each such subsequence in $\pi$ is called an *occurrence* of the pattern $\tau$. Let $\tau$ be a random permutation in $\mathcal{S}_n$ and let the random variable $X_{\tau,n} := X_\tau(\pi)$ be the number of copies of $\tau$ in $\pi$. We consider asymptotics as $n \to \infty$ for (one or several) fixed $\tau$.

The (joint) distribution of the $X_{\tau,n}$ has been investigated in a series of paper [5, 6, 28]. In particular, Bona [5] showed that for every $\tau \in \mathcal{S}_k$, as $n \to \infty$,

$$\frac{X_{\tau,n} - E[X_{\tau,n}]}{n^{k-\frac{1}{2}}} \xrightarrow{d} N(0, Z_\tau) \tag{15}$$

for some $Z_\tau > 0$. Janson, Nakamura and Zeilberger [28] showed that the above holds jointly for any finite family of patterns $\tau$.

Note that as a consequence of the convergence (in distribution) in (15), we obtain the following concentration inequality:

$$\Pr\left(|X_{\tau,n} - E[X_{\tau,n}]| > t\right) \leq 2\exp\left(-\frac{t^2}{2n^{2k-1}Z_\tau}\right) \tag{16}$$

which is sharp up to a polynomial factor.

On the other hand, consider the graph $H$ defined at the beginning of this subsection, where $V(H)$ is the set of all linear permutations of $[n]$. It is not hard to see that the function $X_{\tau,n} : V(H) \to \mathbb{Z}$ is $\binom{n-1}{k-1}$-Lipschitz. It follows by Theorem 1 that

$$\begin{aligned}\Pr\left(\left|\frac{X_{\tau,n}}{\binom{n-1}{k-1}} - \frac{E[X_{\tau,n}]}{\binom{n-1}{k-1}}\right| > \frac{t}{\binom{n-1}{k-1}}\right) &\leq 2\exp\left(-\frac{t^2\kappa}{5\binom{n-1}{k-1}^2}\right)\\ &\leq 2\exp\left(-\frac{t^2}{C_k n^{2k-1}}\right)\end{aligned}$$

for some $C_k > 0$. Hence the concentration result in Theorem 1 is in fact asymptotically optimal in the case of counting occurrences of patterns in random permutations.

**Remark 5.** *Similar Ricci curvature and concentration results can be obtained for the space of cyclic permutations as well.*

**Remark 6.** *Another possible way to geometrize the space of linear permutations is the random transposition model (see, e.g., [24]) as follows: let $V(H) = \mathcal{S}_n$ and two permutations $\sigma_1, \sigma_2 \in V(H)$ are adjacent in $H$ if $\sigma_2 = \tau \circ \sigma_1$ for some transposition $\tau$. Define a random walk $m$ on $H$ by*

$$m_\sigma(\sigma') = \begin{cases} \frac{2}{n(n-1)} & \text{if } \sigma \text{ and } \sigma' \textit{ are adjacent in } H,\\ 0 & \text{otherwise.}\end{cases}$$

The invariant distribution is the uniform measure on $\mathcal{S}_n$. The Ricci curvature of this graph is $\Theta(n^{-2})$, as observed by Gozlan et al. [25].

## Acknowledgement

We would like to sincerely thank the anonymous referees for their valuable comments and suggestions that greatly improved the manuscript.

## References


[1] R. Adamczak and P. Wolff, Concentration inequalities for non-Lipschitz functions with bounded derivatives of higher order, *Probab. Theory Related Fields* **162(3)** (2015), 531–586. MR3383337

[2] N. Alon and J. H. Spencer, *The Probabilistic Method*, Wiley-Interscience, third edition, 2008. MR2437651

[3] K. Azuma, Weighted sums of certain dependent random variables, *Tohoku Math. J.* **19(3)** (1967), 357–367. MR0221571

[4] D. Bakry and M. Emery, Diffusions hypercontractives, Séminaire de probabilités, XIX, 1983/84, 177–206, Lecture Notes in Math., 1123, Springer, Berlin, 1985. MR0889476

[5] M. Bóna, The copies of any permutation pattern are asymptotically normal, arXiv:0712.2792 (2007).

[6] M. Bóna, On three different notions of monotone subsequences, *Permutation Patterns*, 89–114, London Math. Soc. Lecture Note Ser., 376, Cambridge Univ. Press, Cambridge, 2010. MR2732825

[7] S. Chatterjee, The missing log in large deviations for triangle counts, *Random Structures Algorithms* **40(4)** (2012), 437–451. MR2925306

[8] F. Chung, Spanning trees in subgraphs of lattices, *Contemp. Math.* **245**, 201–119, Amer. Math. Soc., Providence, R.I., 1999. MR1732103

[9] F. Chung, A. Grigor'yan and S.-T. Yau, Upper bounds for eigenvalues of the discrete and continuous Laplace operators, *Adv. Math.* **117** (1996), 165–178. MR1371647

[10] F. Chung, A. Grigor'yan and S.-T. Yau, Eigenvalues and diameters for manifolds and graphs, *Tsing Hua lectures on geometry & analysis* (Hsinchu, 1990–1991), 79–105, Int. Press, Cambridge, MA, 1997. MR1482032

[11] F. Chung, A. Grigor'yan and S.-T. Yau, Higher eigenvalues and isoperimetric inequalities on Riemannian manifolds and graphs, *Comm. Anal. Geom.* **8** (2000), 969–9026. MR1846124

[12] F. Chung and L. Lu, Concentration inequalities and martingale inequalities: a survey, *Internet Math.* **3(1)** (2006), 79–127. MR2283885

[13] F. Chung and S.-T. Yau, A Harnack inequality for homogeneous graphs and subgraphs, *Comm. Anal. Geom.* **2** (1994), 627–740, also in *Turkish J. Math.* **19(2)** (1995), 119–129. MR1349565

[14] F. Chung and S.-T. Yau, Eigenvalues of graphs and Sobolev inequalities, *Combin. Probab. Comput.* **4** (1995), 11–25. MR1336652

[15] F. Chung and S.-T. Yau, Logarithmic Harnack inequalities, *Math. Res. Lett.* **3** (1996), 793–812. MR1426537

[16] F. Chung and S.-T. Yau, A combinatorial trace formula, *Tsing Hua lectures on geometry & analysis* (Hsinchu, 1990–1991), 107–716, Int. Press, Cambridge, MA, 1997. MR1482033

[17] F. Chung and S.-T. Yau, Coverings, heat kernels and spanning trees, *Electron. J. Combin.* **6** (1999), Research Paper 12, 21pp. MR1667452

[18] F. Chung and S.-T. Yau, A Harnack inequality for Dirichlet eigenvalues, *J. Graph Theory* **34** (2000), 247–757. MR1771459

[19] F. Chung and S.-T. Yau, Discrete Green's functions, *J. Combin. Theory Ser. A* **91** (2000), 191–214. MR1779780

[20] B. DeMarco and J. Kahn, Upper tails for triangles, *Random Structures Algorithms* **40(4)** (2012), 452–459. MR2925307

[21] B. DeMarco and J. Kahn, Tight upper tail bounds for cliques, *Random Structures Algorithms* **41(4)** (2012), 469–487. MR2993131

[22] H. Djellout, A. Guillin, L. Wu, Transportation cost-information inequalities and applications to random dynamical systems and diffusions, *Ann. Probab.* **32(3B)** (2004), 2702–2732. MR2078555

[23] M. Erbar, J. Maas, Ricci curvature of finite Markov chains via convexity of the entropy, *Arch. Ration. Mech. Anal.* **206** (2012), 997–1038. MR2989449

[24] M. Erbar, J. Maas, and P. Tetali, Discrete Ricci Curvature bounds for Bernoulli-Laplace and Random Transposition models, *Annales de la faculté des sciences de Toulouse Mathématiques* **24(4)** (2016), 781–800. MR3434256

[25] N. Gozlan, J. Melbourne, W. Perkins, C. Roberto, P-M. Samson, and P. Tetali, Working Group in New directions in mass transport: discrete versus continuous. *AIM SQuaRE report, October*, 2013.

[26] H. Guo and L. Warnke, Packing nearly optimal Ramsey $R(3,t)$ graphs, *Combinatorica* **40(1)** (2020), 63–103. MR4078812

[27] W. Hoeffding, Probability inequalities for sums of bounded random variables, *J. Amer. Statist. Assoc.* **58** (1963), 13–30. MR0144363

[28] S. Janson, B. Nakamura, and D. Zeilberger, On the asymptotic statistics of the number of occurrences of multiple permutation patterns, *Journal of Combinatorics* **6** (2015), 117–143. MR3338847

[29] S. Janson, K. Oleszkiewicz, and A. Ruciński, Upper tails for subgraph counts in random graphs, *Israel J. Math.* **142** (2004), 61–92. MR2085711

[30] S. Janson and A. Ruciński, The infamous upper tail, *Random Structures and Algorithms* **20** (2002), 317–342. MR1900611

[31] A. Joulin, A new Poisson-type deviation inequality for Markov jump processes with positive Wasserstein curvature, *Bernoulli* **15** (2009), 532–249. MR2543873

[32] A. Joulin and Y. Ollivier, Curvature, concentration and error estimates for Markov chain Monte Carlo, *Ann. Probab.* **38(6)** (2010), 2418–2442. MR2683634

[33] Y. Lin, L. Lu, S. T. Yau, Ricci curvature of graphs, *Tohoku Math. J.* **63** (2011), 605–627. MR2872958

[34] Y. Lin and S.-T. Yau, Ricci curvature and eigenvalue estimate on locally finite graphs, *Mathematical Research Letters* **17** (2010), 345–358. MR2644381

[35] J. Lott and C. Villani, Ricci curvature for metric-measure spaces via optimal transport, *Ann. of Math. (2)* **169(3)** (2009), 903–991. MR2480619

[36] J. Maas, Gradient flows of the entropy for finite Markov chains, *J. Funct. Anal.* **261** (2011), 2250–2292. MR2824578

[37] A. Mielke, Geodesic convexity of the relative entropy in reversible Markov chains, *Calc. Var. Partial Differential Equations* **48** (2013), 1–31. MR3090532

[38] Y. Ollivier, Ricci curvature of Markov chains on metric spaces, *J. Funct. Anal.* **256(3)** (2009), 810–864. MR2484937

[39] M.D. Sammer, Aspects of mass transportation in discrete concentration inequalities. PhD thesis, Georgia Institute of Technology, 2005. MR2707115

[40] K.-Th. Sturm, On the geometry of metric measure spaces. I and II, *Acta Math.* **196(1)** (2006), 65–177. MR2237207

[41] M. Talagrand, Concentration of measure and isoperimetric inequalities in product spaces, *Inst. Hautes Études Sci. Publ. Math.* **81** (1995), 73–205. MR1361756

Linyuan Lu
University of South Carolina
Columbia, SC 29208
United States
E-mail: lu@math.sc.edu

Zhiyu Wang
Georgia Institute of Technology
Atlanta, GA 30332
United States
E-mail: zwang672@gatech.edu