Hyperbolicity, slimness, and minsize, on average

Anna C. Gilbert[*], Joon-Hyeok Yim[†1]

[1]Yale University, New Haven, CT

December 7, 2024

Abstract

A metric space (X, d) is δ-hyperbolic if $d(x, y) + d(z, w)$ is not bigger than $\max(d(x, z) + d(y, w), d(x, w) + d(y, z))$ by 2δ. A geodesic space is δ-slim if every geodesic triangle $\Delta_{x,y,z}$ is δ-slim. It has been widely known that δ-slim, δ-hyperbolic, δ-thin(, and others) notions are all equivalent up to a constant factor. In this paper, we demonstrate that if we compute the *expectation* (or *average*) of each conditions, then the equivalent bound is no longer valid. Simply put, while $\mathbb{E}\delta$-slimness implies the $\mathbb{E}\delta$-hyperbolicity, the opposite is not true. Similarly for all of the other definitions of average versus the usual (worst-case) definitions of hyperbolicity. We exploit this discrepancy in the usual versus average definitions of hyperbolicity for the random Gaussian distribution in Eucldiean space, the random d-regular graph, and the random Erdős-Rényi graph model.

1 Introduction

There are two recent lines of work that we unite in this paper. One of them is in the realm of probability and statistical mechanics. The second is in the realm of more practical data analysis or geometric machine learning. First, Chatterjee and Sloman [6] introduce the notion of average or expected Gromov hyperbolicity in the context of the Parisi ansatz and Ising models. Their main result is that if a metric space has small hyperbolicity on average (i.e., it is tree-like on average), then it can be approximately embedded in a tree with low distortion.

Simultaneously, there has been a resurgence in computing Gromov hyperbolicity values with the burgeoning interest in geometric graph neural networks. There are many proposals to improve the efficacy of graph neural nets by discovering the appropriate geometry and then embedding a data set comprised of a graph and associated node and/or edge features in that geometry. A natural geometry to use is hyperbolic space; thus, the necessity to determine just how hyperbolic a given graph data set is.

The desire for methods to determine the "geometry of a data set" necessitate appropriate hyperbolicity measures that are adapted to finite and discrete structures. Furthermore, measures of hyperbolicity that are "worst case" are often too pessimistic for finite discrete data. As a result, Gilbert and Yim [18] adapted Chatterjee and Sloman's notion of average hyperbolicity for finite data structures and developed an algorithm for fitting tree-like on average data approximately into trees with low average distortion.

While Gromov hyperbolicity is just one measure of hyperbolicity for metric spaces, there are a number of different definitions that capture different geometric properties: slimness, thinness, and size of inscribed

---

[*]Email: anna.gilbert@yale.edu
[†]Email: joon-hyeok.yim@yale.edu

Models	(Usual) hyperbolicity	Average δ_{fp}	Average ζ_{slim}	Average $\eta_{minsize}$
G_n (Section 3)	$\Theta(\sqrt{n})$	$o(1)$	$\Theta(\sqrt{n})$	$\Theta(\sqrt{n})$
$H_{M,n}$ (Section 3)	$\Theta(M)$	$o(1)$	$\nearrow M$	$o(1)$
$\mathrm{RRG}(n,d)(d \geq 3)$ (Section 4)	$\Theta(\log n)$	$o(\omega(n))$	$\Theta(\log n)$	$\Theta(\log n)$
$\mathrm{ER}(n, \frac{\lambda}{n})(\lambda \geq 4.67)$ (Section 5)	$\Theta(\log n)$	$o(\log n)$	$\Theta(\log n)$	$\Theta(\log n)$

Table 1: The hyperbolicity constants on some graph models, and how it *breaks* when it comes to the average. Note that G_n has size $O(n^{2.5})$ and $H_{M,n}$ has size $O(n^2)$.

circles. These definitions are all equivalent up to various constants. In this paper, we adapt these geometric definitions to the finite discrete setting and define their average values. We do so in order to develop a robust set of potential algorithmic tools for analyzing the geometric properties of finite discrete structures. Surprisingly, we find that the equivalences in the *usual worst case* definition do not necessarily hold in the *average* definition and the constants are also not the same!

Another key task in adapting and using these geometric tools is interpreting their results on finite data sets. In particular, it is crucial to understand how a data set might masquerade as almost hyperbolic. For example, do points drawn from a high dimensional Gaussian distribution exhibit low average hyperbolicity (despite residing in Euclidean space)? In addition, we to use these tools on graph data sets and so it is imperative to understand how they perform on families of random graphs.
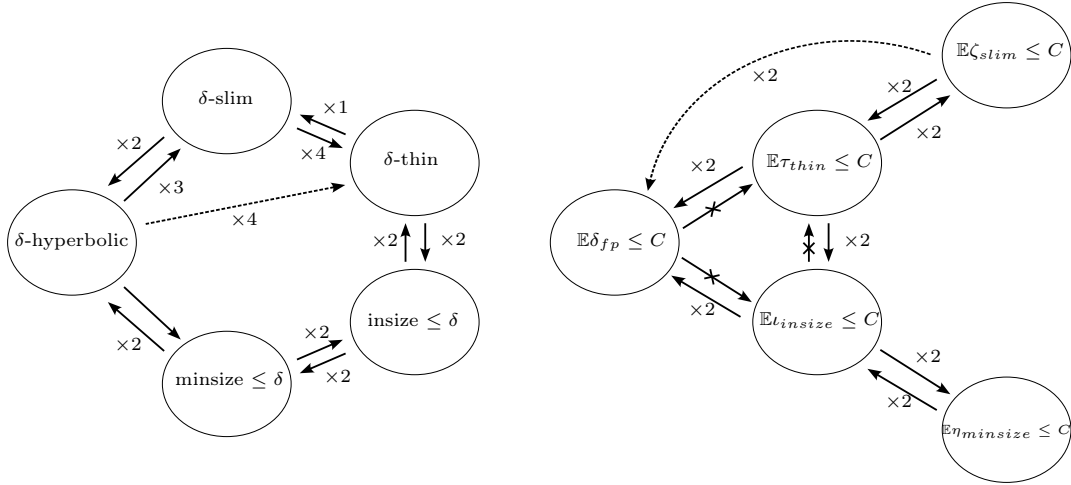

Figure 1: Overview of the results: various δ-hyperbolicity conditions are no longer equivalent when it comes to average. The figure highlights what are the (non-)equivalences.

2 Preliminaries

We begin with the usual definitions of Gromov hyperbolicity. For all of our definitions, there is an important distinction between a geodesic metric space and a discrete, combinatorial one. For graphs, this distinction is

captured by treating edges in a graph as a curve in the metric space (upon which we can place intermediate points) or as a discrete object which is not a part of the metric space and, hence, cannot be further subdivided by intermediary points. This distinction causes some slight differences in definitions which we articulate as well.

2.1 (Usual) definitions of hyperbolicity

Definition 2.1 (Gromov Hyperbolicity)**.** Given a metric space (X, d), the four-point condition of the quadruple x, y, z, w, fp is defined as

$$\text{fp}(x, y, z, w) := \frac{1}{2} \left[\max(P, Q, R) - \text{med}(P, Q, R) \right],$$

while

$$P := d(x, y) + d(z, w), Q := d(x, z) + d(y, w), R := d(x, w) + d(y, z).$$

(a) Given a metric space (X, d) and $\delta \geq 0$, if $\text{fp}(x, y, z, w) \leq \delta$ holds for all $x, y, z, w \in X$, then we call (X, d) δ-*hyperbolic*.

(b) Given a (connected) graph $G = (V, E)$ and $\delta \geq 0$, if $\text{fp}(x, y, z, w) \leq \delta$ holds for all $x, y, z, w \in V$, then we will call G as δ-*hyperbolic*. Here, we use shortest-path metric as the metric d.

Note that the above hyperbolicity condition can also be defined using the Gromov product. The *Gromov product* of two points $x, y \in X$ with respect to a base point $w \in X$ is defined as

$$\langle x, y \rangle_w := \frac{1}{2} [d(x, w) + d(y, w) - d(x, y)].$$

Then, the four point condition can be expressed as

$$\text{fp}(x, y, z, w) = \max_{\pi \text{ perm}} [\min(\langle x, z \rangle_w, \langle y, z \rangle_w) - \langle x, y \rangle_w].$$

The restriction to a base point w in this definition does not cause a structural change in the definition of Gromov hyperbolicity; we do not discuss here the necessary but trivial modification.

The more intuitive definition of hyperbolicity is that based upon the condition on a *geodesic triangle*. Given three points x, y, z in a *geodesic space* X, we denote a *geodesic triangle* $\Delta(x, y, z)$ with vertices x, y, z as the union of all three geodesic segments $[x, y], [y, z], [z, x]$ (where $[p, q]$ denotes a geodesic segment with endpoints p and q).

Definition 2.2 (δ-slim)**.** We start with, arguably the most familiar, the δ-*slim* definition.

(a) Given a geodesic metric space (X, d) and $\delta \geq 0$, a geodesic triangle $\Delta(x, y, z)$ for $x, y, z \in X$ is δ-slim if for any point $w \in [x, y]$, the distance from w to $[y, z] \cup [z, x]$ is at most δ, and similarly for $[y, z]$ and $[z, x]$. We call (X, d) δ-*slim* if every geodesic triangle in X is δ-slim.

(b) Given a (connected) graph $G = (V, E)$ and $\delta \geq 0$, a geodesic triangle $\Delta(x, y, z)$ for $x, y, z \in V$ is δ-slim if for any vertex $w \in [x, y]$, the distance from w to $[y, z] \cup [z, x]$ is at most δ, and similarly for $[y, z]$ and $[z, x]$. We refer to G as δ-*slim* if every geodesic triangle in G is δ-slim.

Note for a graph there is a slight difference on the hyperbolicity constant. For example, a cycle with length 3 C_3 is 0-slim if we realize the space as a graph, as a geodesic segment only contains two endpoints. However, it is not 0-slim if we consider the space as a geodesic metric space.

A similar but more restricted condition is the following δ-*thin* condition.

Definition 2.3 (δ-thin). (a) Given a geodesic metric space (X, d) and $\delta \geq 0$, a geodesic triangle $\Delta(x, y, z)$ for $x, y, z \in X$ is δ-thin if $d(x, v) = d(x, w) \leq \langle y, z \rangle_x$, then $d(v, w) \leq \delta$ holds for any points $v \in [x, y]$ and $w \in [x, z]$, and similar conditions hold for y and z. We say (X, d) is δ-*thin* if every geodesic triangle in X is δ-thin.

(b) Given a (connected) graph $G = (V, E)$ and $\delta \geq 0$, a geodesic triangle $\Delta(x, y, z)$ for $x, y, z \in X$ is δ-thin if $d(x, v) = d(x, w) \leq \langle y, z \rangle_x$, then $d(v, w) \leq \delta$ holds for any vertices $v \in [x, y]$ and $w \in [x, z]$, and similar conditions hold for y and z. We say G is δ-*thin* if every geodesic triangle in X is δ-thin.

Let us denote $\mathrm{slim}(\Delta(x, y, z))$ and $\mathrm{thin}(\Delta(x, y, z))$ as the infimum of the δ-slim, thin, respectively, values for all triangles; i.e.,

$$\mathrm{slim}(\Delta(x, y, z)) := \inf\{\delta \geq 0 : \Delta(x, y, z) \text{ is } \delta\text{-slim}\}$$
$$\mathrm{thin}(\Delta(x, y, z)) := \inf\{\delta \geq 0 : \Delta(x, y, z) \text{ is } \delta\text{-thin}\}.$$

Remark 2.4. By definition, a δ-thin triangle is always δ-slim. In other words, $\mathrm{slim}(\Delta(x, y, z)) \leq \mathrm{thin}(\Delta(x, y, z))$ always holds. Consequently, a δ-thin space is always δ-slim.
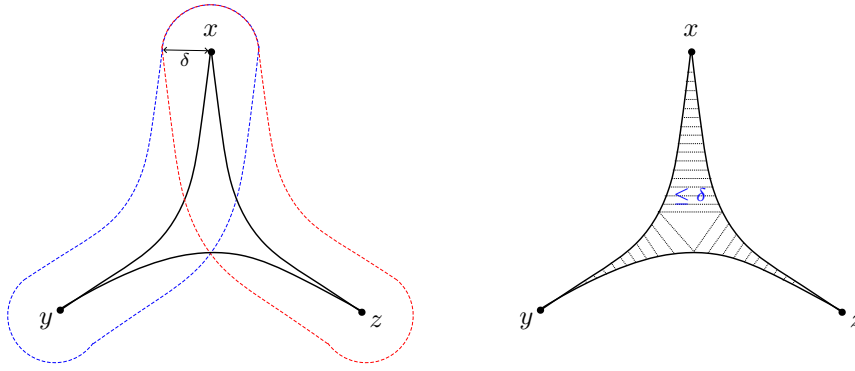

Figure 2: δ-slim and δ-thin triangle

Definition 2.5 (minsize, insize). The *minsize* of a geodesic triangle $\Delta(x, y, z)$ is defined as follows.

$$\mathrm{minsize}(\Delta(x, y, z)) = \inf_{x' \in [y, z], y' \in [z, x], z' \in [x, y]} \mathrm{diam}(\{x', y', z'\})$$

In particular, pick $m_{yz} \in [y, z]$ with $d(x, y) + d(y, m_{yz}) = d(x, z) + d(z, m_{yz})$ and $m_{zx} \in [z, x], m_{xy} \in [x, y]$ similarly. We call the diameter of $\{m_{xy}, m_{yz}, m_{zx}\}$, the *insize* of the triangle.

Remark 2.6. If given space is a graph $G = (V, E)$, by convexity, it turns out that $\inf_{x', y', z'} \mathrm{diam}(\{x', y', z'\})$ occurs when x', y', z' are all vertices (not a point on an edge). Therefore, we do not need to specifically define the minsize of triangle on a graph. Note that even if m_{xy}, m_{yz}, m_{zx} may not be vertices in a graph (instead, they are midpoint of an edge), we see that minsize $\leq$ insize.

4

2.2 Equivalences amongst definitions of (usual) hyperbolicity

The following results show that all of the above hyperbolicity definitions are equivalent up to multiplication by a small constant. In the next section, we see that this is no longer the case when we consider the *average* definitions of hyperbolicity. As many of these results have appeared in the literature (and we are simply compiling them in one place for completeness) or are straightforward consequences of the definitions, we state the results here and for the proofs, either refer to the appropriate reference or show the result in the Appendix.

Theorem 2.7. *Given a geodesic metric space* (X, d),

1. *if it is δ-hyperbolic, then it is 3δ-slim [1].*

2. *if it is δ-thin, then it is δ-hyperbolic [5].*

3. *if it is δ-slim, then it is 4δ-thin [8]. The bound is tight.*

4. *if it is δ-slim, then it is 2δ-hyperbolic [14]. The bound is asymptotically tight.*

5. *if every geodesic triangle has minsize at most δ, then it has insize at most 3δ as well [8]. We note that the original reference uses 4δ, but we can easily improve the bound.*

6. *if every geodesic triangle has insize at most δ, then it is 2δ-thin [8].*

There are similar statements for discrete graphs. Proofs are in the Appendix.

Theorem 2.8. *Given a graph $G = (V, E)$,*

1. *if it is δ-hyperbolic, then it is $3\delta + \frac{1}{2}$-slim.*

2. *if it is δ-thin, then it is $\delta + \frac{1}{2}$-hyperbolic.*

3. *if it is δ-slim, then it is $2\delta + \frac{1}{2}$-hyperbolic.*

4. *if it is δ-slim, then it is 4δ-thin.*

We conclude with the definitions of hyperbolicity of a geodesic metric space.

Definition 2.9. Given a geodesic space or graph X, denote

$$\mathrm{hyp}(X) := \inf\{\delta : X \text{ is } \delta\text{-hyperbolic}\} = \sup\{\mathrm{fp}(x, y, z, w)\}$$

$$\mathrm{slim}(X) := \inf\{\delta : X \text{ is } \delta\text{-slim}\} = \sup\{\mathrm{slim}(\Delta(x, y, z))\}$$

$$\mathrm{thin}(X) := \inf\{\delta : X \text{ is } \delta\text{-thin}\} = \sup\{\mathrm{thin}(\Delta(x, y, z))\}$$

$$\mathrm{insize}(X) := \sup\{\mathrm{insize}(\Delta) : \Delta \text{ geodesic triangle}\} = \sup\{\mathrm{insize}(\Delta(x, y, z))\}$$

$$\mathrm{minsize}(X) := \sup\{\mathrm{minsize}(\Delta) : \Delta \text{ geodesic triangle}\} = \sup\{\mathrm{minsize}(\Delta(x, y, z))\}$$

3 Average definitions of hyperbolicity

In this section, we explicate the *average* version of each of the hyperbolicity definitions from the previous section. The idea underlying the average notions is that we select some number of points at random from a space (X, d) or a graph $G = (V, E)$ and compute various geometric quantities. All of these definitions suppose that we select the points independently and identically distributed according to a distribution which we leave unspecified. We do, however, assume, that the distribution is non-trivial on all points in the space (or vertices of the graph).

5

3.1 Definitions

Definition 3.1. We define the *average* hyperbolicity on (X, d) as follows

$$\mathbb{E}\delta_{hyp}(X) := \mathbb{E}_{x,y,z,w \text{ iid}} \operatorname{fp}(x, y, z, w).$$

We define the slimness of a geodesic triangle and its average similarly.

For graphs, we must be a bit more careful. It might be the case that the geodesic path between two vertices x, y is not unique (i.e., there may be several shortest paths between two vertices), thereby making it difficult to define uniquely a geodesic triangle amongst three vertices x, y, z. In this case, we *randomly* and *uniformly* choose one segment amongst all shortest paths between x, y to constitute a segment in $\Delta(x, y, z)$. To be more precise, suppose there are $\sigma_{x,y}$ shortest paths between x and y. Then, we choose one path amongst them to form $[x, y] \in \Delta(x, y, z)$ with uniform probability $\frac{1}{\sigma_{x,y}}$, and similarly for $[y, z]$ and $[z, x]$.

Definition 3.2. We use the following notation to define the corresponding average hyperbolicity values on three vertices x, y, z:

1. $\zeta(x, y, z) := \mathbb{E} \operatorname{slim}(\Delta(x, y, z))$

2. $\tau(x, y, z) := \mathbb{E} \operatorname{thin}(\Delta(x, y, z))$

3. $\iota(x, y, z) := \mathbb{E} \operatorname{insize}(\Delta(x, y, z))$

4. $\eta(x, y, z) := \mathbb{E} \operatorname{minsize}(\Delta(x, y, z))$

For example, if the geodesic triangle on x, y, z is unique, then $\zeta(x, y, z)$ is the slimness of the triangle itself.

Definition 3.3. We define the *average* hyperbolicity values on (X, d) as:

1. $\mathbb{E}\zeta_{slim}(X) := \mathbb{E}_{x,y,z \text{iid}} \zeta(x, y, z)$

2. $\mathbb{E}\tau_{thin}(X) := \mathbb{E}_{x,y,z \text{iid}} \tau(x, y, z)$

3. $\mathbb{E}\iota_{insize}(X) := \mathbb{E}_{x,y,z \text{iid}} \iota(x, y, z)$

4. $\mathbb{E}\eta_{minsize}(X) := \mathbb{E}_{x,y,z \text{iid}} \eta(x, y, z)$

Remark 3.4. For geodetic graphs, the shortest path between two vertices x and y is always unique (See, for example, [12]). Therefore, $\zeta(x, y, z) = \operatorname{slim}(\Delta(x, y, z))$ always holds (and so on).

3.2 Equivalences: lower and upper bounds

Given the equivalences amongst the usual definitions of hyperbolicity in Section 2.2, one might expect that the average definitions of hyperbolicity also enjoy similar equivalences. Surprisingly, this is not necessarily the case. There are some definitions of average hyperbolicity that are equivalent and there are some that are not. In this subsection, we detail the equivalences, exhibiting upper and lower bounds amongst the equivalent definitions. Because the non-equivalences and the examples that witness them are more compelling and suprising, we delay their presentation until Section 4

We begin with a simple lemma that relates the slimness or minsize of a geodesic triangle with its thinness or insize, respectively.

Lemma 3.5. Given a geodesic space (X, d) or a graph $G = (V, E)$,

1. If a geodesic triangle $\Delta(x, y, z)$ is δ-slim, then it is 4δ-thin (both as a geodesic space, or as a graph).

2. If a geodesic triangle $\Delta(x, y, z)$ has minsize δ, then it has insize less than or equal to 3δ (both as a geodesic space, or as a graph).

Proof. We follow the proof of [8] and extend it slightly to the case when the geodesic triangle is in a graph. Given a geodesic triangle $\Delta(x, y, z)$ with slimness δ, suppose there exists $u \in [x, z], v \in [x, y]$ with $d(u, x) = d(v, x) \leq \langle y, z \rangle_x$ and $d(u, v) > 4\delta$. Then,

$$d(u, [x, y]) = \min(d(u, [x, v]), d(u, [v, y])) \geq \langle x, v \rangle_u, \langle v, y \rangle_u,$$

where $2\langle x, v \rangle_u = d(x, u) + d(v, u) - d(x, v) = d(v, u) > 4\delta$ and $2\langle v, y \rangle_u = d(v, u) + d(y, u) - d(v, y) = d(v, u) + d(y, u) - (d(x, y) - d(x, v)) = d(v, u) + (d(y, u) + d(x, u) - d(x, y)) \geq d(v, u) > 4\delta$. Therefore, $d(u, [x, y]) > 2\delta$ holds.

Now, pick $p_y \in [x, u]$ with $d(p_y, u) = \delta$. Even in the discrete graph case, we can choose such a point since δ is integer. Then we have $d(p_y, [x, y]) \geq d(u, [x, y]) - d(u, p_y) > \delta$ and there should be $q \in [y, z]$ such that $d(p_y, q) \leq \delta$. On the other hand,

$$d(p_y, [y, z]) \geq d(x, [y, z]) - d(x, p_y) \geq \langle y, z \rangle_x - d(x, p_y)$$
$$= d(m_y, x) - d(x, p_y) = d(p_y, m_y) = d(p_y, u) + d(u, m_y) = \delta + d(u, m_y).$$

Therefore, $u = m_y, v = m_z$ and $d(p_y, [y, z]) = \delta$. Also,

$$d(p_y, [y, z]) = \min(d(p_y, [m_x, z]), d(p_y, [y, m_x])) \geq \min(\langle m_x, z \rangle_{p_y}, \langle y, m_x \rangle_{p_y}),$$

where $2\langle m_x, z \rangle_{p_y} = d(m_x, p_y) + d(z, p_y) - d(m_x, z) = \delta + d(p_y, m_x)$ and $2\langle y, m_x \rangle_{p_y} = d(y, p_y) + d(m_x, p_y) - d(y, m_x) \geq (d(x, y) - d(x, p_y)) + d(p_y, m_x) - d(m_z, y) = d(x, m_z) + d(p_y, m_x) - d(x, p_y) = \delta + d(p_y, m_x)$. Therefore,

$$2\delta \geq 2d(p_y, [y, z]) \geq 2\min(d(p_y, [m_x, z]), d(p_y, [y, m_x])) \geq \delta + d(p_y, m_x) \Rightarrow d(p_y, m_x) \leq \delta.$$

Similarly, pick $p_z \in [x, m_z]$ with $d(p_z, m_z) = \delta$. Then we have $d(p_z, m_x) \leq \delta$. This shows

$$d(m_y, m_z) \leq d(m_y, p_y) + d(p_y, m_x) + d(m_x, p_z) + d(p_z, m_z) \leq 4\delta,$$

a contradiction. This completes the proof of part 1 of the Lemma.

For the second part of the Lemma, we modify the proof from [8] slightly. Given a geodesic triangle $\Delta(x, y, z)$, pick p_x, p_y, p_z in $[y, z], [z, x], [x, y]$ with diameter δ and let m_x, m_y, m_z be the points of the inscribed triple. We will show that $d(m_x, m_y) \leq 3\delta$. There are three cases to analyze:

1. $p_x \in [m_x, z]$ and $p_y \notin [m_y, z]$: Then we see that $d(p_y, m_y) + d(m_x, p_x) = d(p_y, z) - d(p_x, z) \leq d(p_x, p_y) \leq \delta$. Therefore,
$$d(m_x, m_y) \leq d(m_x, p_x) + d(p_x, p_y) + d(p_y, m_y) \leq \delta + \delta = 2\delta.$$

2. $p_x \notin [m_x, z]$ and $p_y \in [m_y, z]$: One can similarly show that $d(m_x, m_y) \leq 2\delta$ as well.

3. $p_x \in [m_x, z]$ and $p_y \in [m_y, z]$: We see that $p_z \in [x, m_z]$ or $p_z \in [m_z, y]$ holds. Without loss of generality, assume that the former is true. Then
$$d(p_y, m_y) + d(m_z, p_z) = d(p_y, x) - d(x, p_z) \leq d(p_y, p_z) \leq \delta,$$
and
$$d(m_x, p_x) - d(m_z, p_z) = d(y, p_x) - d(y, p_z) \leq d(p_x, p_z) \leq \delta.$$
Therefore, $d(m_x, p_x) + d(p_y, m_y) \leq 2\delta$ and $d(m_x, m_y) \leq d(m_x, p_x) + d(p_x, p_y) + d(p_y, m_y) \leq 3\delta$.

4. $p_x \notin [m_x, z]$ and $p_y \notin [m_y, z]$: Again, without loss of generality, assume $p_z \in [x, m_z]$. Then

$$d(p_x, m_x) + d(m_z, p_z) = d(p_z, y) - d(y, p_x) \leq d(p_x, p_z) \leq \delta,$$

and

$$d(m_y, p_y) - d(m_z, p_z) = d(x, p_z) - d(x, p_y) \leq d(p_y, p_z) \leq \delta.$$

Therefore, $d(m_x, p_x) + d(p_y, m_y) \leq 2\delta$ and $d(m_x, m_y) \leq d(m_x, p_x) + d(p_x, p_y) + d(p_y, m_y) \leq 3\delta$.

The above calculations cover every case and show $d(m_x, m_y), d(m_y, m_z), d(m_z, m_x) \leq 3\delta$ as desired for part 2 of the Lemma. □

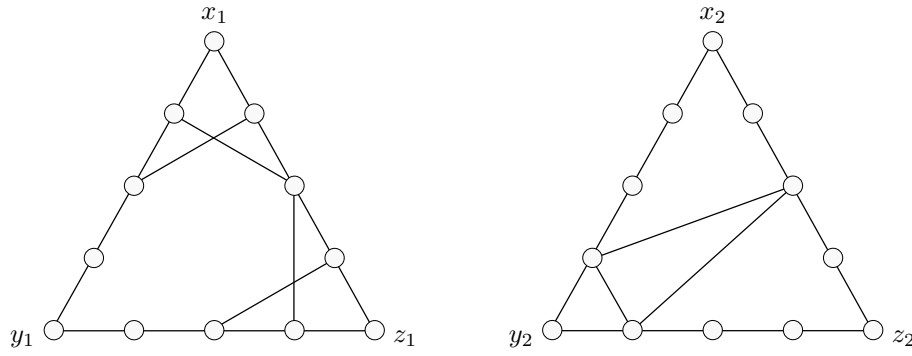The graph in Figure 3 shows that this inequality is tight.

Figure 3: The left geodesic triangle is 1-slim, but 4-thin. The right geodesic triangle has minsize 1, but its insize is 3. Therefore we see that bounds on 3.5 are tight and cannot be improved.

Corollary 3.6. For a geodesic space or a graph X,

1. $\mathbb{E}\zeta_{slim}(X) \leq \mathbb{E}\tau_{thin}(X) \leq 4\mathbb{E}\zeta_{slim}(X)$.

2. $\mathbb{E}\eta_{minsize}(X) \leq \mathbb{E}\iota_{insize}(X) \leq 3\mathbb{E}\eta_{minsize}(X)$.

3. $\mathbb{E}\eta_{minsize}(X) \leq \mathbb{E}\iota_{insize}(X) \leq \mathbb{E}\tau_{thin}(X)(+1)$ (+1 is only for the case of a discrete graph).

3.3   Slim/thin triangles, on average, imply hyperbolicity, on average

First, we begin with the proposition that the average four-point hyperbolicity condition is implied by the other (average) conditions, as we expect from the usual definitions. The main idea of the proof is to follow the proof details from [14] and [5] and simply take *expectations*. For a discrete graph, however, we must make suitable, if straightforward modifications. In all cases, we assume that we draw points at random from a non-degenerate distribution.

Theorem 3.7. *Given a geodesic space* (X, d),

1. $\mathbb{E}\delta_{hyp}(X) \leq 2\mathbb{E}\zeta_{slim}(X)$ *holds*.

8

2. $\mathbb{E}\delta_{hyp}(X) \le \mathbb{E}\iota_{insize}(X) \le \mathbb{E}\tau_{thin}(X)$ *holds.*

3. $\mathbb{E}\delta_{hyp}(X) \le 2\mathbb{E}\eta_{minsize}(X) \le 2\mathbb{E}\iota_{insize}(X)$ *holds.*

Given a (connected) graph $G = (V, E)$,

1. $\mathbb{E}\delta_{hyp}(G) \le 2\mathbb{E}\zeta_{slim}(G) + \frac{1}{2}$ *holds.*

2. $\mathbb{E}\delta_{hyp}(G) \le \mathbb{E}\tau_{thin}(G) + \frac{1}{2}$ *holds.*

3. $\mathbb{E}\delta_{hyp}(G) \le 2\mathbb{E}\eta_{minsize}(G) \le 2\mathbb{E}\iota_{insize}(G)$ *holds.*

Proof. First, assume that (X, d) is a geodesic space. The overall strategy is as follows. First, randomly sample x, y, z and w. Next, we again *randomly* sample geodesic paths between all pairs of those points, in order to constitute geodesic triangles $\Delta(x, y, z), \Delta(x, y, w), \Delta(x, z, w)$, and $\Delta(y, z, w)$. Then, we bound $\delta_{fp}(x, y, z, w)$ in terms of some constants (such as slimness, minsize, etc.) over these triangles. Finally, take the expectation to conclude the proof.

1. We will show that

$$\delta_{fp}(x, y, z, w) \le \frac{1}{2} \left[\zeta(x, y, z) + \zeta(x, y, w) + \zeta(x, z, w) + \zeta(y, z, w) \right].$$

Without loss of generality, assume $d(x, y) + d(z, w) \ge \max(d(x, z) + d(y, w), d(x, w) + d(y, z))$. Pick $v \in [x, y]$ such that $d(x, z) - d(x, v) = d(y, z) - d(y, v) = \langle x, y \rangle_z$. Then we have $d(x, y) + d(z, t) \ge d(x, z) + d(y, t) = d(x, t) + d(y, z)$. Next, by assumption, there exists $t \in [x, w] \cup [y, w]$ such that $d(w, t) \le \mathrm{slim}(\Delta(x, y, w))$. wlog assume $t \in [x, w]$. Then we have

$$\begin{aligned}
d(x, y) + d(w, v) = d(x, v) + d(v, y) + d(w, v) &\le (d(x, t) + d(t, v)) + d(y, v) + (d(t, v) + d(t, w)) \\
&\le d(x, w) + d(y, v) + 2\,\mathrm{slim}(\Delta(x, y, w)).
\end{aligned}$$

Similarly, one can show (because $d(y, v) + d(x, z) = d(x, v) + d(y, z)$,)

$$d(x, y) + d(z, v) \le d(y, v) + d(x, z) + 2\,\mathrm{slim}(\Delta(x, y, z)) = d(x, v) + d(y, z) + 2\,\mathrm{slim}(\Delta(x, y, z)).$$

This shows that

$$\begin{aligned}
d(x, y) + d(z, w) &\le (d(x, y) + d(z, v)) + (d(x, y) + d(w, v)) - d(x, y) \\
&\le (d(x, v) + d(y, z) + 2\,\mathrm{slim}(\Delta(x, y, z))) + (d(x, w) + d(y, v) + 2\,\mathrm{slim}(\Delta(x, y, w))) - d(x, y) \\
&= d(y, z) + d(x, w) + 2(\mathrm{slim}(\Delta(x, y, z)) + \mathrm{slim}(\Delta(x, y, w))),
\end{aligned}$$

which shows that $fp(x, y, z, w) \le \mathrm{slim}(\Delta(x, y, z)) + \mathrm{slim}(\Delta(x, y, w))$. Taking the expectation yields $fp(x, y, z, w) \le \zeta(x, y, z) + \zeta(x, y, w)$.

We can repeat the argument above by suitably choosing taking $v \in [z, w]$. Then we have $fp(x, y, z, w) \le \zeta(z, w, x) + \zeta(z, w, y)$ and leveraging the two inequalities proves the claim. The statement comes from taking the expectation (over the i.i.d. samples x, y, z, w).

2. Similarly, we will show

$$\delta_{fp}(x, y, z, w) \le \frac{1}{4} \left[\iota(x, y, z) + \iota(x, y, w) + \iota(x, z, w) + \iota(y, z, w) \right].$$

Without loss of generality, assume $d(x, z) + d(y, w) \geq d(x, w) + d(y, z) \geq d(x, y) + d(z, w)$. Pick $x' \in [x, w]$, $y' \in [y, w]$ such that $d(x', w) = d(y', w) = \langle x, y \rangle_w$ and $y'' \in [y, w]$, $z' \in [z, w]$ such that $d(y'', w) = d(z', w) = \langle y, z \rangle_w$. By assumption, we have $d(y', w) \geq d(y'', w)$. Then

$$
\begin{aligned}
d(x, z) &\leq d(x, x') + d(x', y') + d(y', y'') + d(y'', z') + d(z', z) \\
&\leq d(x, x') + \text{insize}(\Delta(x, y, w)) + d(y', y'') + \text{insize}(\Delta(y, z, w)) + d(z', z) \\
&= \langle y, w \rangle_x + (\langle x, y \rangle_w - \langle y, z \rangle_w) + \langle y, w \rangle_z + \text{insize}(\Delta(x, y, w)) + \text{insize}(\Delta(y, z, w)) \\
&= d(x, w) + d(y, z) - d(y, w) + \text{insize}(\Delta(x, y, w)) + \text{insize}(\Delta(y, z, w)),
\end{aligned}
$$

which immediately shows that $\delta_{fp}(x, y, z, w) \leq \frac{1}{2} \left[\text{insize}(\Delta(x, y, w)) + \text{insize}(\Delta(y, z, w)) \right]$. Taking the expectation yields $\delta_{fp}(x, y, z, w) \leq \frac{1}{2} \left[\iota(x, y, w) + \iota(y, z, w) \right]$. Again, we repeat the argument on $[x, z]$ and achieve $\delta_{fp}(x, y, z, w) \leq \frac{1}{2} \left[\iota(x, y, z) + \iota(x, z, w) \right]$, which completes the proof.

3. We will show

$$
\delta_{fp}(x, y, z, w) \leq \frac{1}{2} \left[\eta(x, y, z) + \eta(x, y, w) + \eta(x, z, w) + \eta(y, z, w) \right].
$$

Again without loss of generality, assume $d(x, z) + d(y, w) \geq \max(d(x, y) + d(z, w), d(x, w) + d(y, z))$. By assumption, pick $p_x \in [y, z]$, $p_y \in [z, x]$, $p_z \in [x, y]$ and $q_x \in [z, w]$, $q_z \in [w, x]$, $q_w \in [x, z]$ so that $\text{diam}(\{p_x, p_y, p_z\}) = \text{minsize}(\Delta(x, y, z))$ and $\text{diam}(\{q_x, q_y, q_z\}) = \text{minsize}(\Delta(x, z, w))$. WLOG assume $d(x, p_y) \leq d(x, q_w)$. Then

$$
d(y, w) \leq d(y, p_x) + d(p_x, p_y) + d(p_y, q_w) + d(q_w, q_z) + d(q_z, w)
$$

and

$$
d(x, z) + d(p_y, q_w) = d(x, q_w) + d(p_y, z) \leq d(x, q_z) + d(q_z, q_w) + d(p_y, p_x) + d(p_x, z),
$$

which shows

$$
\begin{aligned}
d(y, w) + d(x, z) &\leq d(y, p_x) + d(p_x, z) + d(x, q_z) + d(q_z, w) + 2(d(p_x, p_y) + d(q_w, q_z)) \\
&\leq d(y, z) + d(x, w) + 2(\text{minsize}(\Delta(x, y, z)) + \text{minsize}(\Delta(x, z, w))).
\end{aligned}
$$

Hence, $\delta_{fp}(x, y, z, w) \leq \eta(x, y, z) + \eta(x, z, w)$ holds, and repeating the argument completes the proof.

On a graph $G = (V, E)$, we need to modify the proof slightly as follows.

1. It may be the case that we cannot explicitly pick $v \in [x, y]$ with the desired identity, if the corresponding Gromov product is a half-integer. Instead, we will assert $|(d(x, z) - d(x, v)) - (d(y, z) - d(y, v))| \leq 1$ and the constant $1/2$ comes from the fact that two distance sum may differ by at most 1. Still, we are able to show $\delta_{fp}(x, y, z, w) \leq \text{slim}(\Delta(x, y, z)) + \text{slim}(\Delta(x, y, w)) + \frac{1}{2}$.

2. (Note that for this proof we will use the *thinness* condition instead.) Here, we will choose $d(x', w) = d(y', w) = \lfloor \langle x, y \rangle_w \rfloor$ and $d(y'', w) = d(z', w) = \lfloor \langle y, z \rangle_w \rfloor$. Then, by the thinness condition, we bound $d(x', y') \leq \text{thin}(\Delta(x, y, w))$ and $d(y'', z') \leq \text{thin}(\Delta(y, z, w))$ and show

$$
\delta_{fp}(x, y, z, w) \leq \frac{1}{2} \left[\text{insize}(\Delta(x, y, w)) + \text{insize}(\Delta(y, z, w)) + \frac{1}{2} \right].
$$

3. Finally, note that by the definition of minsize, we do not need to pick some additional nodes differently simply because we are in the discrete graph setting when it comes to the minsize. Therefore, $\mathbb{E}\delta_{hyp}(G) \leq 2\mathbb{E}\eta_{minsize}(G) \leq 2\mathbb{E}\iota_{insize}(G)$ still holds.

$\square$

Remark 3.8. The reader might wonder if we define ζ (and other constants) differently, how do our equivalences change? For example, we could define $\zeta(x,y,z)$ as the infimum (i.e., the best case) among all slimnesses over possible $\Delta(x,y,z)$. In this case, unfortunately, the desired equivalence inequalities do not hold. For example, if we consider the 4-cycle C_4, then we can always *pick* the geodesic triangle so that its slimness is always zero. The graph, however, is not a tree (far from it) and it defies intuition that its hyperbolicity is zero.

4 Non-equivalences

In this section, we detail the examples that demonstrate the non-equivalences of our definitions; that is, definition A does not imply defiition B up to multiplication by a (small) constant. . These results are suprising given the equivalences amongst all of the usual definitions of hyperbolicity. As there are a number of cases to examine, we break them up into subsections. In each subsection, we provide families of examples that illustrate the non-equivalences.

4.1 Average hyperbolicity does not imply average slimness

We answer, in the negative, whether hyperbolicity on average implies that triangles are slim/thin on average. We witness a family of graphs in which the answer to this question is no, average hyperboliticy does not necessarily imply thin or slim triangles on average.
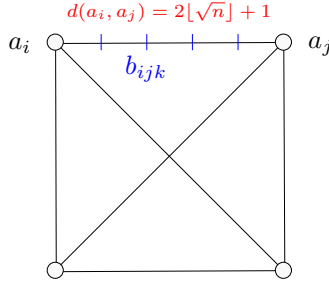

Figure 4: This figure shows the construction of the family of graphs G_n in which average hyperbolicity does not necessarily imply thin or slim triangles on average.

We construct a family of graphs $\{G_n\}$ for $n \geq 3$ as follows. Denote $m := 2\lfloor \sqrt{n} \rfloor + 1$.

- $V(G_n) = \{a_i | 1 \leq i \leq n\} \cup \{b_{ijk} | 1 \leq i < j \leq n \text{ and } 1 \leq k \leq m\}$.

- For any $1 \leq i < j \leq n$, connect $a_i - b_{ij1} - b_{ij2} - \cdots - b_{ijm} - a_j$.

Note that $|V(G_n)| = O(n^{2.5})$. We see that this graph family essentially has a *gap* between the average hyperbolicity and the average slimness.

Theorem 4.1. *The family of graphs G_n satisfies*

1. $\mathbb{E}\delta_{hyp}(G_n) \to 0$.

2. $\mathbb{E}\zeta_{slim}(G_n), \mathbb{E}\eta_{minsize}(G_n) \to \infty$.

Proof. First, we show that $\mathbb{E}\delta_{hyp}(G_n) \to 0$. We randomly choose four points x, y, z, w and consider the following *generic* configuration:

1. None of the four points are the vertices a_i for some i. This occurs with probability $1 - O(n^{-1.5})$.

2. The point x lies between $[a_{p(x)}, a_{q(x)}]$ for some $p(x), q(x) \in [n]$, and say $h_x := d(x, a_{p(x)}) < d(x, a_{q(x)})$ (therefore, $p(x) < q(x)$ may not be true). Then with probability $1 - O(n^{-1})$, all of the 8 points $p(\cdot), q(\cdot)$ for each of x, y, z, w are distinct.

Overall, the *generic* configuration occurs with probability $1 - O(n^{-1})$. In this case, the geodesic between two points is always unique and should pass a_p. Therefore,

$$d(x, y) = d(x, a_{p(x)}) + d(a_{p(x)}, a_{p(y)}) + d(a_{p(y)}, y) = m + h_x + h_y,$$

and the similar formula holds for all other pairwise distances. Hence,

$$d(x, y) + d(z, w) = d(x, w) + d(y, z) = d(x, z) + d(y, w) = 2m + h_x + h_y + h_z + h_w,$$

so that $\delta_{fp}(x, y, z, w) = 0$.

On the other hand, if the point configuration is not *generic*, then we bound $\delta_{fp}(x, y, z, w)$ by $\frac{1}{2}\operatorname{diam}(G_n)$, which is $4\lfloor\sqrt{n}\rfloor + 1 = 2m - 1$. This implies that

$$\mathbb{E}\delta_{hyp}(G_n) \leq O(\tfrac{1}{n}) \cdot (2m - 1) = O(\tfrac{1}{n} \cdot \sqrt{n}) = o(1).$$
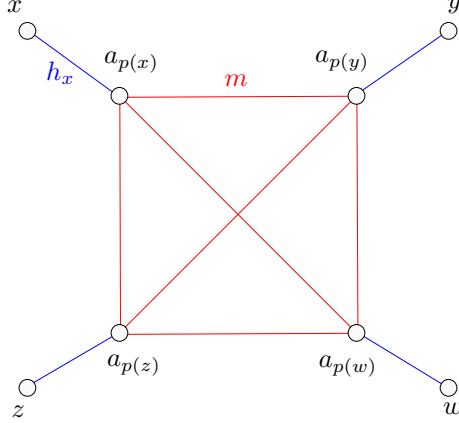

Figure 5: This figure shows the generic configuration for the four point condition in the construction of the family G_n

For the next proposition, it is enough to check that $\mathbb{E}\zeta_{slim}(G_n) \to \infty$. When we randomly sample three points x, y, z in this setting, the *generic* configuration is as follows:

1. None of the three points are the vertices a_i for some i. This occurs with probability $1 - O(n^{-1.5})$.

2. Again, the point x lies between $[a_{p(x)}, a_{q(x)}]$ for some $p(x), q(x) \in [n]$, and say $d(x, a_{p(x)}) < d(x, a_{q(x)})$. Then with probability $1 - O(n^{-1})$, all 6 indices $p(\cdot), q(\cdot)$ for each of x, y, z are distinct.

Hence, the *generic* configuration occurs with probability $1 - O(n^{-1})$. However, in this case, the geodesic triangle $\Delta(x, y, z)$ is unique and $\zeta(x, y, z) = \text{slim}(\Delta(x, y, z)) = \lfloor \sqrt{n} \rfloor = \frac{m-1}{2}$. Therefore,

$$\mathbb{E}\zeta_{slim}(G_n) \geq (1 - O(\tfrac{1}{n})) \cdot \frac{m-1}{2} = \Omega(\sqrt{n}).$$

$\square$

4.2 Average slimness and average minsize are not necessarily equivalent

In this subsection, we show that the slimness and minsize conditions are not necessarily equivalent in the average case. We have seen that slimness implies a bound on the minsize (and insize); this example indicates that the δ-slim and δ-thin conditions are the most strongest conditions amongst the equivalent hyperbolicity notions.

Fix a positive integer M. Then we construct a graph family $\{H_{M,n}\}$ for $n \geq 2$ as follows:

- $V(H_{M,n})$ consists of what we call *junction* nodes $\{c_1, \cdots, c_n\}$, nodes $\{a_{i,s}\}$ for $1 \leq i \leq n$, $1 \leq s \leq M \cdot n$ and nodes $\{b_{i,j,k}\}$ for $1 \leq i \leq n-1$, $1 \leq j \leq 2M$ and $k = 1, 2$.

- For any $1 \leq i \leq n-1$, there are two paths between c_i and c_{i+1} with length $2M$: $c_i - b_{i,1,1} - b_{i,2,1} - \cdots - b_{i,2M,1} - c_{i+1}$ (informally call these upward paths) and $c_i - b_{i,1,2} - b_{i,2,2} - \cdots - b_{i,2M,2} - c_{i+1}$ (downward paths, respectively). There is an edge between c_i and $a_{i,s}$ for any i, s.

Note that $|V(H_{M,n})| = O(n^2)$.

Theorem 4.2. *Given any positive integer M, the graph family $H_{M,n}$ satisfies*

1. $\mathbb{E}\zeta_{slim}(H_{M,n}) \to M$.

2. $\mathbb{E}\eta_{minsize}(H_{M,n}) \to 0$.

Proof. Our approach is to observe which configuration is generic and to argue that it occurs with high probability. Let us randomly sample x, y, z in $V(H_{M,n})$ and consider the following generic configuration.

1. All points x, y, z are in the "cluster" $\{a_{i,s}\}$, so that each is adjacent to $c_{i(x)}, c_{i(y)}, c_{i(z)}$, respectively.

2. $|i(x) - i(y)|, |i(y) - i(z)|, |i(z) - i(x)| \geq \log_2 n$ holds.

Note that first condition holds with probability at least $1 - O(n^{-1})$. Conditional on that, we see that the conditional distribution of $i(x)$s is uniform in $[n]$. Therefore, the second condition holds with (conditional) probability at least $1 - O(\log n/n)$. Thus, the *generic* configratuion occurs with probability at least $1 - O(\log n/n)$.

Now, consider a geodesic triangle $\Delta(x, y, z)$ in the generic configuration. Without loss of generality assume $i(x) < i(y) < i(z)$. Then we see that regardless of the choice of geodesic, $[x, y], [y, z]$ and $[z, x]$ always passes $c_{i(y)}$. In fact, $c_{i(y)}$ actually becomes the Steiner node among these points. Therefore, $\eta_{minsize}(x, y, z) = \text{minsize}(\Delta(x, y, z)) = 0$ if x, y, z are generic. On the other hand, it is easy to verify that every geodesic triangle in $H_{M,n}$ has minsize at most $2M$. Therefore,

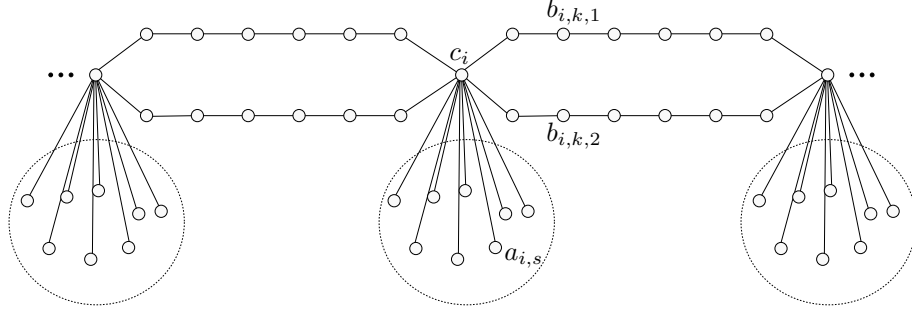$$\mathbb{E}\eta_{minsize}(H_{M,n}) \leq O(\tfrac{\log n}{n}) \cdot 2M = o(1),$$

Figure 6: This figure depicts the graph family $H_{M,n}$ in which average slimnes and average minsize are not equivalent.

as $n \to \infty$.

Now we evaluate the slimness of the generic triangle $\Delta(x, y, z)$. This depends on the choice of geodesic; let us consider $[x, y]$ and $[x, z]$ in $\Delta(x, y, z)$. When each geodesic runs from $c_{i(x)}$ to $c_{i(y)}$, if at least one of path choices between two adjacent junction nodes differs (i.e., one goes upward and the other goes downward), then the slimness of $\Delta(x, y, z)$ should be at least M, as the midpoint of each segment is distant from the other segment with distance M. As we have assumed three indices differ by at least $\log_2 n$, this phenomenon happens on at least $1 - n^{-1}$ portion of the geodesic triangles $\Delta(x, y, z)$ so that

$$\zeta_{slim}(x, y, z) \geq (1 - n^{-1}) \cdot M.$$

Therefore,

$$\mathbb{E}\zeta_{slim}(H_{M,n}) \geq (1 - O(\tfrac{\log n}{n})) \cdot (1 - \tfrac{1}{n})M \to M,$$

as $n \to \infty$. On the other hand, we can check that every geodesic triangle in $H_{M,n}$ has slimness at most M. Therefore, $\mathbb{E}\zeta_{slim}(H_{M,n}) \to M$ as desired. $\qquad\square$

4.3 Generic geodesic space examples

In this subsection, we consider generic geodesic spaces, rather than families of graphs per se, and find more extreme examples.

Example 4.3. Consider a Gaussian $N(0, \frac{I_d}{\sqrt{d}})$ random variable in $\mathbb{R}^d$. Note that Euclidean space is always non hyperbolic and we cannot bound the hyperbolicity constant δ in any sense in this space. However, if we sample points from a Gaussian distribution in $\mathbb{R}^d$ and compute the Gromov hyperbolicity constant for these points (embedded in $\mathbb{R}^d$, we see that as $d \to \infty$, $\delta_{fp} \to 0$. This is the distribution of distances between points $d(x, y)$ converges to $\sqrt{2}$. The geodesic triangle in $\mathbb{R}^d$, however, is *nowhere* slim (or thin), so that $\zeta_{slim}, \tau_{thin}, \eta_{minsize}, \iota_{insize} \to \sqrt{2}/2$.

This example highlights an important phenomenon. The first is that when one is computing hyperbolicity values numerically on finite data sets, it is important to be careful drawing conclusions. One's data set might be samples drawn from a Gaussian distribution in high dimensional Euclidean space and one might be

tempted to conclude with a small Gromov hyperbolicity value, that the points are from a hyperbolic space when, in fact, they are not.

Example 4.4. Another extreme example is a space equipped with a distribution concentrated on exactly 3 points uniformly, say $a, b, c \in X$. Then by definition, $\delta_{fp}(x, y, z, w) = 0$ always, as there would be a duplicate among four points. On the other hand, we see that $\mathbb{E}\zeta_{\text{slim}}$ is $2/9 \cdot \zeta_{\text{slim}}(a, b, c)$, and the similar thing holds on all other constants. This means that a single geodesic triangle can serve an example on the average case as well. For example, consider the graph H such that $\Delta(x, y, z)$ has minsize(insize) 1 and slimness $n/2$. Equip this graph with a distribution concentrated on x, y, z and we can see that *the bound on minsize* does not imply *the bound on slimness*.
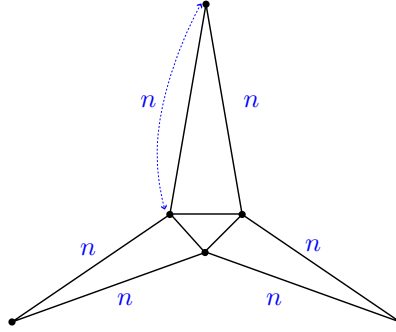

Figure 7: A geodesic triangle which is fat (as its slimness is $n/2$), but still has its insize 1

5 Random models

In the previous section, we defined the average notions of hyperbolicity and presented several examples that illustrate the *discrepancy* amongst the average definitions. This discrepancy does not exist for the usual definitions of hyperbolicity. In this section, we show that those examples are, in fact, not special and are, in fact, fairly generic. In particular, we show that on a variety of random graphs the average definitions of hyperbolicity are not equivalent. The two families of random graphs we examine are random regular graphs and Erdös-Renyi random graphs.

5.1 Random regular graph $\text{RRG}(n, d)$

In this section, we analyze the average hyperbolicity of the random regular graph model $\text{RRG}(n, d)$ with $d \geq 3$. One might think that such graphs are *not* hyperbolic on average as [3] showed that the random regular graph is *not* hyperbolic (according to the usual definitions) and, in fact, the hyperbolicity constant is at least $\frac{1}{2} \log_{d-1} n - \omega(n)$ with high probability. This value is large considering that any graph is trivially $\frac{1}{2} \text{diam}(G)$-hyperbolic and the diameter of a random regular graph is bounded by $\frac{1}{2} \log_{d-1} n + O(\log \log n)$. Furthermore, [16] investigated a technique to show these graphs are not hyperbolic using the observation that hyperbolic graphs must admit traffic congestion (or have a "core").

We demonstrate that despite these results for the usual definitions of hyperbolicity on random regular graphs, there is an asymptotic gap between the average hyperbolicity and the average slimness. It is well-

known that the random graph $\mathrm{RRG}(n, d)$ is connected with high probability (i.e., such event happens with probability $\to 1$), so that we condition on that.

First, we show that the average hyperbolicity is surprisingly bounded by $o(\omega(n))$. This is because the *typical distance* between two nodes is reasonably concentrated, so that the generic four point condition is not a large enough multiple of the number of vertices. More precisely, we consider the *distribution of shortest path lengths (DSPL)*, the distribution of distance between two fixed points. Our result builds on the recent analysis:

Lemma 5.1. [15] For fixed x_1, x_2, consider $L_n := d(x_1, x_2)$ in $G_n \in \mathrm{RRG}(n, d)$. Then the variance of L_n, $\mathrm{Var}(L_n)$, satisfies

$$\mathrm{Var}(L_n) = \frac{\pi^2}{6[\log(d-1)]^2} + \frac{1}{12} + O\left(\frac{\log n}{n}\right). \tag{1}$$

Theorem 5.2. *Given $G_n \in \mathrm{RRG}(n, d)$ for $d \geq 3$ and any growth function $\omega(n) \to \infty$,*

$$\mathbb{E}\left(\mathbb{E}\left(\delta_{fp}(G_n)\right)\right) = o(\omega(n)),$$

with high probability.

Proof. First, we observe that because our random graph model is *homogeneous*, for fixed $x_1, x_2, x_3, x_4 \in V(G_n)$,

$$\mathbb{E}\left(\mathbb{E}(\delta_{fp}(G_n))\right) = \mathbb{E}\,\mathrm{fp}(x_1, x_2, x_3, x_4)$$

(note that we condition G_n being connected). Let $P = d(x_1, x_2) + d(x_3, x_4)$ and $Q = d(x_1, x_3) + d(x_2, x_4)$. By the definition of the distribution of the distance between two fixed points (in a random regular graph), both $d(x_1, x_2)$ and $d(x_3, x_4)$ are distributed as L_n. Furthermore, even though $d(x_1, x_2)$ and $d(x_3, x_4)$ might not be independent,

$$\mathbb{E}[P] = \mathbb{E}[d(x_1, x_2) + d(x_3, x_4)] = 2\mathbb{E}(L_n) \quad \text{and} \quad \mathrm{Var}[P] \leq 4\,\mathrm{Var}(L_n),$$

and similarly for $\mathbb{E}[Q]$ and
$Var[Q]$. Therefore,

$$\mathbb{E}\,\mathrm{fp}(x_1, x_2, x_3, x_4) \leq \tfrac{1}{2}\mathbb{E}|P - Q| \leq \tfrac{1}{2}\left[\mathbb{E}|P - 2\mathbb{E}(L_n)| + \mathbb{E}|Q - 2\mathbb{E}(L_n)|\right]$$

$$\leq \tfrac{1}{2}\left[\mathrm{Var}[P]^{1/2} + \mathrm{Var}[Q]^{1/2}\right] \leq 2\,\mathrm{Var}(L_n)^{1/2}.$$

Lemma 5.1 tells us that the right hand side of the above inequality is bounded above by Equation 1. Thus, given any $\omega(n) \to \infty$,

$$\mathbb{P}\left(\mathbb{E}(\delta_{fp}(G_n)) \geq \omega(n)\right) \leq \frac{2\,\mathrm{Var}(L_n)^{1/2}}{\omega(n)} \to 0.$$

$\square$

We emphasize that we barely used the structure of quadruple of points and the four point condition to obtain the previous result; the bound on the distance distribution itself already provides a bound on the average Gromov hyperbolicity!

To prove the next theorem, we build upon the ideas from [16] and [9]to show that *most* geodesic triangles (in a random regular graph) are, in fact, extremely fat!

Theorem 5.3. *There exists a constant $C_d > 0$, which only depends on d (for $d \geq 3$), such that for $G_n \in \mathrm{RRG}(n, d)$,*

$$\mathbb{E}\left(\zeta_{\mathrm{slim}}(G_n)\right), \mathbb{E}\left(\eta_{\mathrm{minsize}}(G_n)\right) \geq \frac{1}{2}\log_{d-1} n - \frac{3}{2}\log_{d-1}\log_{d-1} n - C_d,$$

with high probability.

Corollary 5.4. For any $\epsilon > 0$, with high probability, most geodesic triangles in G_n are δ-fat with

$$\delta = \frac{1}{2}\log_{d-1} n - (\frac{3}{2} + \epsilon)\log_{d-1}\log_{d-1} n.$$

To prove Theorem 5.3, we use the following sequence of Lemmas. Denote $\Gamma_r(w) := \{x \in V : d(x, w) = r\}$ and $N_r(w) := \{x \in V : d(x, w) \leq r\}$ for $w \in V$ and any integer $r \geq 0$.

Lemma 5.5. If $G = (V, E)$ is a graph with maximal degree $d \geq 3$, then

$$|\Gamma_r(x)| \leq d(d-1)^{r-1}, \quad |N_r(x)| < \frac{d}{d-2}(d-1)^r$$

for all $x \in V, r \geq 0$.

Proof. Given $x \in V$, consider a breadth first search tree on G with root vertex x for which $|\Gamma_r(x)|$ is exactly the number of nodes with level r. As every node except the root x has at most $d - 1$ children (and at most d including x itself), $|\Gamma_{r+1}(x)| \leq (d-1)|\Gamma_r(x)| \leq d(d-1)^r$ can be shown inductively. The second part of the Lemma can be shown by the calculation

$$|N_r(x)| = \sum_{i \leq r}|\Gamma_i(x)| \leq 1 + d(1 + (d-1) + \cdots + (d-1)^{r-1})$$

$$= 1 + d\frac{(d-1)^r - 1}{d-2} < \frac{d}{d-2}(d-1)^r.$$

$\square$

In the next Lemma (from [16], we are interested in the following set of pairs of vertices, which is informally referred as a *traffic* (and is also closely related to the *betweenness centrality*),

$$T(w) := \{(x, y) \in V \times V : \text{ there exists a geodesic } [x, y] \text{ contains } w\} \quad \text{for } w \in V.$$

Lemma 5.6. [16] Let $G = (V, E)$ be a graph with diameter D. For $w \in V$ and $i \geq 0$, we have

$$T(w) \leq \sum_{k+l \leq D}|\Gamma_k(w)| \cdot |\Gamma_l(w)|.$$

Proof. Given $(x, y) \in T(w)$, it is clear that $d(x, w) + d(w, y) = d(x, y) \leq D$. The number of $(x, y) \in T(w)$ conditioned on $d(x, w) = k$ and $d(y, w) = l$ is at most $|\Gamma_k(w)| \cdot |\Gamma_l(w)|$. Therefore, the desired inequality holds. $\square$

Lemma 5.7. Suppose $G_n \in \mathrm{RRG}(n, d)$ with $d \geq 3$ has its diameter D. Then

$$|T(w)| < \frac{d^2}{(d-2)^2}(d-1)^D$$

holds for all w in $V(G_n)$.

Proof. By lemma 5.6, we see

$$
\begin{aligned}
|T(w)| &\leq \sum_{k+l\leq D} |\Gamma_k(w)| \cdot |\Gamma_l(w)| \\
&\leq \sum_{k+l\leq D} d^2(d-1)^{k+l-2} \\
&= \frac{d^2}{(d-1)^2} \sum_{0\leq t\leq D} \sum_{k+l=t} (d-1)^t \\
&\leq \frac{d^2}{(d-1)^2} \sum_{0\leq t\leq D} (D+1-t)\cdot(d-1)^t \\
&= \frac{d^2}{(d-1)^2} \left(\frac{(d-1)^{D+1}-1}{d-2} + \frac{(d-1)^D-1}{d-2} + \cdots + \frac{d-2}{d-2} \right) \\
&< \frac{d^2}{(d-1)^2} \left(\frac{(d-1)^{D+1}+(d-1)^D+\cdots+1}{d-2} \right) \\
&= \frac{d^2}{(d-1)^2} \cdot \frac{(d-1)^{D+2}-1}{(d-2)^2} < \frac{d^2}{(d-2)^2}(d-1)^D.
\end{aligned}
$$

It turns out that the bound provided by [16] is quite loose and can be improved. This improvement will be used when we compute the second leading term in the bound on the sizes of the average geodesic triangle. □

In the next Lemmas, we use the known bound on the diameter for a random regular graph.

Lemma 5.8. [4] There exists a constant C_d, only depends on $d \geq 3$, such that $\mathrm{diam}(G_n) \leq \log_{d-1} n + \log_{d-1} \log_{d-1} n + C_d$ with high probability.

The next Lemma is an improvement upon that in [9].

Lemma 5.9. Let $G = (U, V, E)$ be a bipartite graph. The edges of G are colored in such a way that so that given a node $x \in U \cup V$, the number of colors incident to x is at most $t(x)$. (u is incident to a color c if u is incident to an edge with color c). Then there exists a color that is used by at least

$$
\frac{|E|^2}{\left(\sum_{u\in U} t(u)\right)\left(\sum_{v\in V} t(v)\right)}
$$

edges in E.

Note that the above lemma immediately implies Lemma 3.2 in [9] (by letting $|U| = |V| = n$ and $t(x) \leq t$ for all nodes x). The reason we developed a new lemma is as follows. Instead of the universal bound on t (which is related to the behavior of the diameter of random graphs), we are able to use the *average* bound on t, which may be used to improve the bound later.

Proof. We provide the details of the proof in the Appendix. □

Now we are ready to prove the main theorem.

Proof. First, we will prove the statement on the *minsize*. We begin with the *sampling procedure* below.

- For every pair (v, w), *fix* a geodesic segment between v and w (by randomly choosing one). For the rest of the argument, we use only these segments when we consider a geodesic triangle $\Delta(x, y, z)$. These choices constitute a path system.

18

- Randomly choose a pivot point $z \in V(G_n)$. Then evaluate $\mathrm{minsize}(\Delta(x, y, z))$ for any $x, y \in V(G_n)$.

It is easy to see that
$$\mathbb{E} \sum_{x,y \in V} \mathrm{minsize}(\Delta(x, y, z)) = n^2 \mathbb{E} \eta_{\mathrm{minsize}}(G_n).$$

Consider a threshold $\delta \geq 0$ and construct the following bipartite graph $H = (X, Y, E)$ with $X = Y = V(G_n)$ and
$$(x, y) \in E \Leftrightarrow \mathrm{minsize}(\Delta(x, y, z)) \leq \delta.$$

We color each edge in $(x, y) \in E$ with $c(x, y) = p_z \in [x, y]$, where $\{p_x, p_y, p_z\}$ realizes the minsize of geodesic triangle. Then by assumption, $d(p_z, [x, z]), d(p_y, [y, z]) \leq \delta$. We see that

- In any $x \in X$, the number of colors incident to x is at most $|N_\delta([x, z])| \leq (d(x, z) + 1) \cdot N(\delta)$, where $N(\delta)$ is the universal bound on the size of δ-neighborhood. This is because for $p_z \in [x, y]$, there exists $p_x \in [x, z]$ such that $p_z \in N_\delta(p_x) \subset N_\delta([x, z])$.

- Given any color c, the number of edges colored with c is at most $|T(c)|$ where $T(c) = \{(x, y) \in V \times V : c \in [x, y]\}$, which is bounded by lemma.

By Lemma 5.9, we have
$$\frac{d^2}{(d-2)^2}(d-1)^D \geq |T(c)| \geq \max_{c \in \mathcal{C}} |E_c| \geq \frac{|E|^2}{(\sum_{x \in V} |N_\delta(x, z)|)^2} \geq \frac{|E|^2}{N(\delta)^2(n + \sum_{x \in V} d(x, z))^2}.$$

We have $N(\delta) \leq \frac{d}{d-2}(d-1)^\delta$ and $\sum_{x \in V} d(x, z) \leq n \, \mathrm{diam}(G_n)$. This shows that given $p := |E|/n^2$,
$$p \leq \frac{1}{n} \left[\frac{d^2}{(d-2)^2} (\mathrm{diam}(G_n) + 1)(d-1)^{t + \frac{\mathrm{diam}(G_n)}{2}} \right].$$

We note that p is the probability that our *random* geodesic triangle has the slimness at most t. Therefore, if we take expectation, we have
$$\mathbb{P}(\mathrm{minsize}(\Delta(x, y, z)) \leq t) \leq \frac{1}{n} \left[\frac{d^2}{(d-2)^2} (\mathrm{diam}(G_n) + 1)(d-1)^{t + \frac{\mathrm{diam}(G_n)}{2}} \right]$$

$$\mathbb{P}(\mathrm{minsize}(\Delta(x, y, z)) > t) \geq 1 - \frac{1}{n} \left[\frac{d^2}{(d-2)^2} (\mathrm{diam}(G_n) + 1)(d-1)^{t + \frac{\mathrm{diam}(G_n)}{2}} \right].$$

Since $\mathrm{diam}(G_n) \leq \log_{d-1} n + \log_{d-1} \log_{d-1} n + C_d$ with high probability, this suggests that w.h.p.,

$$\mathbb{P}(\mathrm{minsize}(\Delta(x, y, z)) > t) \geq 1 - \frac{1}{n} \left[\frac{d^2(d-1)^{C_d/2}}{(d-2)^2} (\log_{d-1} n + \log_{d-1} \log_{d-1} n + (C_d + 1)) \cdot \sqrt{n \cdot \log_{d-1} n} \cdot (d-1)^t \right].$$

19

Denote $C_1 := C_d + 1$ and $C_2 := \frac{d^2(d-1)^{C_d/2}}{(d-2)^2 \log(d-1)}$. Given $S > 0$ (which will be chosen later), we have

$$\mathbb{E}\zeta_{minsize}(G_n) = \int_0^\infty \mathbb{P}(\text{minsize}(\Delta(x,y,z)) > t)dt \geq \int_0^S \mathbb{P}(\text{minsize}(\Delta(x,y,z)) > t)dt$$

$$= S - \int_0^S \frac{1}{n}\left[\log(d-1)C_2(\log_{d-1} n + \log_{d-1}\log_{d-1} n + C_1) \cdot \sqrt{n \cdot \log_{d-1} n} \cdot (d-1)^t\right]$$

$$= S - \log(d-1)C_2 \frac{(\log_{d-1} n + \log_{d-1}\log_{d-1} n + C_1)(\log_{d-1} n)^{1/2}}{n^{1/2}} \int_0^S (d-1)^t dt$$

$$= S - \log(d-1)C_2 \frac{(\log_{d-1} n + \log_{d-1}\log_{d-1} n + C_1)(\log_{d-1} n)^{1/2}}{n^{1/2}} \left[\frac{(d-1)^S - 1}{\log(d-1)}\right]$$

$$> S - C_2 \frac{(\log_{d-1} n + \log_{d-1}\log_{d-1} n + C_1)(\log_{d-1} n)^{1/2}}{n^{1/2}} \cdot (d-1)^S.$$

Plugging in $S = \frac{1}{2}\log_{d-1} n - \frac{1}{2}\log_{d-1}\log_{d-1} n - \log_{d-1}(\log_{d-1} n + \log_{d-1}\log_{d-1} n + C_1)$ yields

$$\mathbb{E}\zeta_{minsize}(G_n) > S - C_2,$$

as desired. The theorem can be shown by observing that $|S - \frac{1}{2}\log_{d-1} n + \frac{3}{2}\log_{d-1}\log_{d-1} n| = o(1)$.

For the slimness case, we need to choose a different value for c, given a *slim* triangle $\Delta(x,y,z)$. In this case, we choose $c \in [x,y]$ based on the following simple proposition.

Lemma 5.10. If $\text{slim}(\Delta(x,y,z)) = \delta$, then there exists a vertex $c \in [x,y]$ such that $d(c,[x,z]), d(c,[y,z]) \leq \delta + 1$.

Proof. Pick $c \in [x,y]$ with $d(c,[x,z]) > \delta$, which is the *closest* from x. If such c does not exist, then the case is obvious. By the slimness, $d(c,[y,z]) \leq \delta$ holds. Also, for c' the adjacent vertex of c in $[x,y]$ toward x, $d(c,[x,z]) \leq d(c',[x,z]) + 1 \leq \delta + 1$ holds. $\square$

Once we have done that, we can repeat the exactly same argument up to an additional constant 1. $\square$

Remark 5.11. We improved lemmas used in [16] in order to achieve the second leading term $-\frac{3}{2}\log_{d-1}\log_{d-1} n$. If we had directly used the original lemmas, we would obtain the bound $-\frac{5}{2}\log_{d-1}\log_{d-1} n$ instead.

Intuitively, this suggests the following. It is easy to check that the slimness of any geodesic triangle is bounded by $\frac{1}{2}\text{diam}(G)$, which is about $\frac{1}{2}\log_{d-1} n + \frac{1}{2}\log_{d-1}\log_{d-1} n$. On the other hand, we have seen that *most* geodesic triangles in a random regular graph are fat so that the slimness is at least about $\frac{1}{2}\log_{d-1} n - \frac{3}{2}\log_{d-1}\log_{d-1} n$. Therefore the average behavior is somewhere between $[\frac{1}{2}\log_{d-1} n - \frac{3}{2}\log_{d-1}\log_{d-1} n, \frac{1}{2}\log_{d-1} n + \frac{1}{2}\log_{d-1}\log_{d-1} n]$.

5.2 Erdős-Rényi graph $\text{ER}(n, \lambda/n)$ with $\lambda > 1$

Finally, in this subsection, we consider the Erdős-Rényi random graph model $\text{ER}(n,p)$. In order to ensure that the graphs we study have a large connected component, we consider the supercritical case in which the probability of an edge between any two of the n vertices is $p = \lambda/n$ for some $\lambda > 1$. Following convention, we consider the giant component of these graphs, which has size roughly $\gamma(\lambda) \cdot n$. [11] showed that this sparse random graph is *not* hyperbolic by constructing a *single* geodesic triangle which is forced to be fat. (In fact, their result suggests that such triangle is $\Omega(\log n)$-fat, which is not explicitly stated in the original paper.) We want something more in this average case!

Our main result is as follows.

Theorem 5.12. *Given $G_n \in \mathrm{ER}(n, \lambda/n)$ with $\lambda \geq 4.67$, there exists $f(\lambda) > 0$ such that with high probability,*

$$\mathbb{E}\left(\zeta_{slim}(\mathscr{C}_{\max})\right), \mathbb{E}\left(\eta_{\mathrm{minsize}}(\mathscr{C}_{\max})\right) \geq f(\lambda) \cdot \log n,$$

where $\mathscr{C}_{\max}$ denotes the giant component of G_n.

Corollary 5.13. *With high probability, a geodesic triangle in (the giant component $\mathscr{C}_{\max}$ of) $G_n \in \mathrm{ER}(n, \lambda/n)$ is $\Omega(\log n)$-fat.*

The corollary answers the question raised in [11] and shows that the $\Omega(\log n)$-fat triangle in the Erdős-Rényi graph is *typical*.

Proof. The proof procedure is as follows. We use a similar method to that for the random regular graph case. While those calculations were detailed, here, we omit the details and argue simply that the leading term is $O(\log n)$.

1. First, given a point x, we want to bound the number of geodesics which contains the point x. Following the arguments for Lemmas 5.6 and 5.7, what we need to do is to bound the size of neighborhood $|\Gamma_r(x)|$ and $|N_r(x)|$ given $r > 0$. However, as the degree is not bounded, so that the upper bound is not immediate. it is not clear. Hence we adapted the lemma stated from [7].

2. We also need to bound the diameter of our giant component. We use explicit results from [13]. The condition $\lambda \geq 4.67$ is required so that the diameter $\mathrm{diam}(\mathscr{C}_{max})$ is not bigger than $2\log n/\log \lambda + O(1)$.

Lemma 5.14 (Bounds on the size of neighborhood)**.** *Given $G_n \in \mathrm{ER}(n, \lambda/n)$ with $\lambda > 1$,*

$$|\Gamma_r(x)| \leq (r+1)^2 \lambda^r \cdot \log n$$

holds for all $x \in V(G_n)$ and $0 \leq r \leq n$, with high probability.

Proof. We provide the details of the proof in the Appendix for completeness. $\qquad\square$

Lemma 5.15 (Bounds on the diameter, [13])**.** *Given $\lambda > 1$, let $\lambda_* < 1$ satisfy $\lambda_* e^{-\lambda_*} = \lambda e^{-\lambda}$ (i.e., the conjugate). Then there exists a constant C_λ that depends only on λ, such that for $G_n \in \mathrm{ER}(n, \lambda/n)$ and its giant component $\mathscr{C}_{max}$,*

$$\mathrm{diam}(\mathscr{C}_{max}) \leq \left(\frac{1}{\log \lambda} + \frac{2}{\log(1/\lambda_*)}\right) \log n + C_\lambda$$

with high probability. Note that for $\lambda \geq 4.67$, $\frac{2}{\log(1/\lambda_*)} < \frac{1}{\log \lambda}$ holds.

We are now ready to prove Theorem 5.12. Given $\lambda \geq 4.67$, pick $0 < \epsilon < 1 - \frac{2\log \lambda}{\log(1/\lambda_*)}$. With high probability, $\mathrm{diam}(\mathscr{C}_{max}) \leq (2 - \epsilon)\frac{\log n}{\log \lambda} + C_\lambda (=: D_\lambda)$ so that by Lemmas 5.6 and 5.14,

$$
\begin{aligned}
|T(w)| &\leq \sum_{k+l \leq D_\lambda} |\Gamma_k(w)| \cdot |\Gamma_l(w)| \\
&\leq \sum_{k+l \leq D_\lambda} (k+1)^2(l+1)^2 \lambda^{k+l}(\log n)^2 \\
&\leq (\log n)^2 \sum_{t \leq D_\lambda} \sum_{k+l=t} (k+1)^2(l+1)^2 \lambda^t \\
&= \frac{(\log n)^2}{30} \sum_{t \leq D_\lambda} (t-1)(t+1)(t+2)(t+3)(t+5)\lambda^t \\
&\leq \frac{(\log n)^2(D_\lambda+5)^5}{30} \sum_{t \leq D_\lambda} \lambda^t < \frac{(\log n)^2(D_\lambda+5)^5 \lambda^{D_\lambda+1}}{\lambda - 1},
\end{aligned}
$$

21

holds for all $w \in \mathscr{C}_{max}$. We see that, after substituting in the expression for D_lambda and simplifying, the right hand side becomes $O(n^{2-\epsilon} \log^7 n)$. In other words, there exists $A_\lambda > 0$ such that

$$|T(w)| \leq A_\lambda \cdot n^{2-\epsilon} \log^7 n,$$

for n sufficiently large.

Next, we use our previous proof technique again. This gives us that for $p := \mathbb{P}(\text{minsize}(\Delta(x,y,z)) \leq t)$, for a giant component $\mathscr{C}_{max}$ with size γn, there exists a color c which is used at least

$$\frac{(p(\gamma n)^2)^2}{N(t)^2(\gamma n + \sum_{x \in \mathscr{C}_{max}} d(x,z))^2} \leq |T(c)| \leq A_\lambda \cdot n^{2-\epsilon} \log^7 n$$

times.

We simply use $d(x,z) \leq \text{diam}(\mathscr{C}_{max}) \leq 2\frac{\log n}{\log \lambda} - 1$ and $N(t) \leq \sum_{r \leq t} \max |\Gamma_r(x)| \leq \sum_{r \leq t} (r+1)^2 \lambda^r \log n \leq (t+1)^3 \lambda^t \log n$, so that

$$p = \mathbb{P}(\text{minsize}(\Delta(x,y,z)) \leq t) \leq \frac{2\sqrt{A_\lambda}}{\gamma \log \lambda} n^{-\epsilon/2}(\log^{5.5} n) \cdot (t+1)^3 \lambda^t,$$

$$1 - p = \mathbb{P}(\text{minsize}(\Delta(x,y,z)) > t) \geq 1 - \frac{2\sqrt{A_\lambda}}{\gamma \log \lambda} n^{-\epsilon/2}(\log^{5.5} n) \cdot (t+1)^3 \lambda^t.$$

Now pick $S_n > 0$ such that $\int_0^{S_n} (t+1)^3 \lambda^t dt = \left(\left[\frac{\lambda^t((t+1)^3 \log^3 \lambda - 3(t+1)^2 \log^2 \lambda + 6(t+1)\log \lambda - 6)}{\log^4 \lambda} \right]_{t=0}^{t=S_n} = \right) \frac{n^{\epsilon/2}}{\log^{5.5} n}$. It turns out that

$$\mathbb{E}\zeta_{minsize}(\mathscr{C}_{max}) = \int_0^\infty \mathbb{P}(\text{minsize}(\Delta(x,y,z)) > t)dt \geq \int_0^{S_n} \mathbb{P}(\text{minsize}(\Delta(x,y,z)) > t)dt \geq S_n - \frac{2\sqrt{A_\lambda}}{\gamma \log \lambda}.$$

It is easy to check that $S_n \sim (\epsilon/2) \cdot \frac{\log n}{\log \lambda}$. Also, with high probability, γ converges to the final probability $\gamma(\lambda)$, which only depends on λ. Therefore, the right hand side of the inequality is bounded by $\Omega(\log n)$ with high probability, thus completing the proof.

The slimness case can be shown analogously. □

As we have seen in random regular graph case, we again expect a gap between the (four point) hyperbolicity and the slimness of the geodesic triangles. Although, we do not have a proof explicitly bounding the four point condition, we instead include a simple proof that the gap *exists*.

Theorem 5.16. *Given $G_n \in \text{ER}(n, \lambda/n)$ with $\lambda > 1$ and its giant component $\mathscr{C}_{\max}$, we have*

$$\mathbb{E}\left(\delta_{fp}(\mathscr{C}_{max})\right) = o(\log n)$$

with high probability.

Proof. Fix $S_\lambda > \frac{1}{\log \lambda} + \frac{2}{\log(1/\lambda_*)}$. Let B be the event such that

1. The giant component $\mathscr{C}_{max}$ has size $0.99\gamma(\lambda) \cdot n \leq |\mathscr{C}_{max}| \leq 1.01\gamma(\lambda) \cdot n$.

2. The diameter $\text{diam}(\mathscr{C}_{max}) \leq S_\lambda \cdot \log n$.

We see that $\mathbb{P}(B) \to 1$ as $n \to \infty$. Let B_{1234} be the event that B occurs and x_1, x_2, x_3, x_4 all belong to the giant component $\mathscr{C}_{\max}$ (for fixed vertices x_1, x_2, x_3, x_4). As the graph model is homogeneous, $\mathbb{P}(B_{1234}) \leq \frac{(1.01\gamma(\lambda)n)^4}{n^4}\mathbb{P}(B) = (1.01\gamma(\lambda))^4\mathbb{P}(B)$.

The expectation of average hyperbolicity *conditioned on* B is

$$
\mathbb{E}\left(\mathbb{E}\delta_{fp}(\mathscr{C}_{\max}) \mid B\right) = \mathbb{E}\left(\frac{1}{|\mathscr{C}_{max}|^4}\sum_{x,y,z,w \in \mathscr{C}_{\max}} \mathrm{fp}(x,y,z,w) \mid B\right)
$$

$$
\leq \frac{1}{(0.99\gamma(\lambda)n)^4}\mathbb{E}\left(\sum_{x,y,z,w \in \mathscr{C}_{\max}} \mathrm{fp}(x,y,z,w) \mid B\right)
$$

$$
= \frac{1}{(0.99\gamma(\lambda)n)^4}\sum_{x,y,z,w}\mathbb{E}\left(\mathbb{1}(x,y,z,w \in \mathscr{C}_{\max}) \wedge \mathrm{fp}(x,y,z,w) \mid B\right)
$$

$$
= \frac{1}{(0.99\gamma(\lambda)n)^4}\frac{n!}{(n-4)!}\mathbb{E}\left(\mathbb{1}(x_1,x_2,x_3,x_4 \in \mathscr{C}_{\max}) \wedge \mathrm{fp}(x_1,x_2,x_3,x_4) \mid B\right)
$$

$$
\leq \frac{1}{(0.99\gamma(\lambda))^4}\frac{\mathbb{P}(B_{1234})}{\mathbb{P}(B)}\mathbb{E}\left(\mathrm{fp}(x_1,x_2,x_3,x_4) \mid B_{1234}\right)
$$

$$
\leq \frac{(1.01)^4}{(0.99)^4}\mathbb{E}\left(\mathrm{fp}(x_1,x_2,x_3,x_4) \mid B_{1234}\right).
$$

Now we use the result from [17]. It states that for fixed x_1, x_2, $\frac{d(x_1,x_2)}{\log n}$ converges to $\frac{1}{\log \lambda}$ in probability, while conditioned on x_1 and x_2 belonging to the giant component $\mathscr{C}_{max}$. In other words, given any $\eta > 0$,

$$
\mathbb{P}(|d(x_1,x_2) - \log_\lambda n| > \eta\log_\lambda n \mid x_1, x_2 \in \mathscr{C}_{max}) \to 0.
$$

Fix $\eta > 0$. Let E be the event (conditional on B_{1234}) such that there exists a pairwise distance among x_1, x_2, x_3, x_4 outside of $[(1-\eta)\log_\lambda n, (1+\eta)\log_\lambda n]$. If E does not occur, then the four point condition $\mathrm{fp}(x_1,x_2,x_3,x_4)$ is bounded by $\eta\log_\lambda n$. If E does occur, then we use $\mathrm{fp}(x_1,x_2,x_3,x_4) \leq \frac{1}{2}\mathrm{diam}(\mathscr{C}_{max}) \leq \frac{1}{2}S_\lambda \log n$ so that with high probability,

$$
\mathbb{E}\,\mathrm{fp}(x_1,x_2,x_3,x_4 \mid B_{1234}) = \mathbb{E}\left(\mathrm{fp}(x_1,x_2,x_3,x_4) \mid E\right)\mathbb{P}(E \mid B_{1234}) + \mathbb{E}\left(\mathrm{fp}(x_1,x_2,x_3,x_4) \mid E^c\right)\mathbb{P}(E^c \mid B_{1234})
$$

$$
\leq \frac{S_\lambda}{2}\log n \cdot \mathbb{P}(E \mid B_{1234}) + (\eta\log_\lambda n) \cdot (1 - \mathbb{P}(E \mid B_{1234}))
$$

$$
\leq \left(\eta + \left(\frac{S_\lambda}{2} - \eta\right)\mathbb{P}(E \mid B_{1234})\right) \cdot \log_\lambda n.
$$

As $\mathbb{P}(E \mid B_{1234}) \to 0$ as $n \to \infty$, $\mathbb{E}\,\mathrm{fp}(x_1,x_2,x_3,x_4 \mid B_{1234}) < 1.01\eta\log_\lambda n$ for n sufficiently large.

Hence, for any fixed $\eta > 0$ and $M > 0$, $\mathbb{E}\left(\mathbb{E}\delta_{fp}(\mathscr{C}_{max}) \mid B\right) \leq \frac{(1.01)^5}{(0.99)^4}\frac{\eta}{2M}\log_\lambda n < \frac{\eta}{M}\log_\lambda n$ holds for n sufficiently large. Hence,

$$
\mathbb{P}(\mathbb{E}\delta_{fp}(\mathscr{C}_{max}) \geq \eta\log_\lambda n \mid B) \leq \frac{\mathbb{E}\left(\mathbb{E}\delta_{fp}(\mathscr{C}_{max}) \mid B\right)}{\eta\log_\lambda n} < \frac{1}{M} \quad \text{for } n \text{ sufficiently large.}
$$

We conclude that $\mathbb{E}\delta_{fp}(\mathscr{C}_{\max}) = o(\log n)$ with high probability (using the fact that B occurs with high probability). $\qquad\square$

6 Discussion

Our analysis establishes important distinctions between the usual and the average case definitions of hyperbolicity. We show that in contrast to the usual definitions, the average definitions are *not* all equivalent up to a constant factor. Furthermore, the examples that exhibit these differences are themselves interesting cases of random discrete structures and distributions of points in high dimensional Euclidean space that might masquerade as points from a hyperbolic space if one is not astute in one's measurements! There are a number of open questions that arise as a result of our analysis.

6.1 Improving the gap between AvgHyp and AvgSlim (or AvgMinsize)

For the Erdös-Renyi random graph $ER(n, \lambda/n)$ case, we establish that the average hyperbolicity (AvgHyp) is bounded by $o(\log n)$, using the distribution of the shortest path lengths (DSPL). Importantly, this result does not rely on any specific structural information about the quadruples of vertices, leaving room to address the gaps amongst the average case definitions. Similarly, in the case of random regular graphs, one can compute the variance of DSPL, which provides a bound on the average hyperbolicity, without using specific structural information about quadruples of points.

We did, however, assume $\lambda \geq 4.67$ to ensure that the graph's diameter remains manageable. What happens if $\lambda > 1$ but is much smaller, perhaps approaching 1? In such scenarios, can we still expect the graph and its geodesic triangles to be $\Omega(\log n)$-fat? Or does it become *truly* tree-like instead? These are questions that are valid, interesting future or follow-up work after our initial results.

6.2 "tree-likeness" of AvgSlim (or AvgMinsize)

We have seen that the *average hyperbolicity* can be used as a measure of how tree-like the space is [6, 18]. We have demonstrated that, actually, the average hyperbolicity is the weakest one amongst the average measures! In fact, the upper bound provided by [6] and [18] are both quite loose: the quantitative bound on results of [6] is hard to track since they use the weighted Szemerédi regularity lemma, which is known to have a gigantic bound. Also, it turns out that the $O(n^3)$ bound given by [18] is enormous in practice. This might not be a coincidence, as we may have assumed too weak an average hyperbolicity bound. For example, given a space or a graph where the *average slimness* is bounded, is there a good tree that fits the given distance with smaller average distortion?

6.3 Expander graphs

This is an important open question. There are negative results for the usual definitions of hyperbolicity [2, 10] but we do not know about the average definition. It's possible (given the results in [9]) that there are some families of expanders that are hyperbolic on average and possibly some that are not.

References

[1] Juan M Alonso, Thomas Brady, Daryl Cooper, Vincent Ferlini, Martin Lustig, Michael Mihalik, Michael Shapiro, and Hamish Short. Notes on word hyperbolic groups. In *Group theory from a geometrical viewpoint*. 1991.

[2] Itai Benjamini. Expanders are not hyperbolic. *Israel Journal of Mathematics*, 108:33–36, 1998.

[3] Itai Benjamini, Carlos Hoppen, Eran Ofek, Paweł Prałat, and Nick Wormald. Geodesics and almost geodesic cycles in random regular graphs. *Journal of Graph Theory*, 66(2):115–136, 2011.

[4] Béla Bollobás and W Fernandez de la Vega. The diameter of random regular graphs. *Combinatorica*, 2:125–134, 1982.

[5] Martin R Bridson and André Haefliger. *Metric spaces of non-positive curvature*, volume 319. Springer Science & Business Media, 2013.

[6] Sourav Chatterjee and Leila Sloman. Average gromov hyperbolicity and the parisi ansatz. *Advances in Mathematics*, 376:107417, 2021.

[7] Fan Chung and Linyuan Lu. The diameter of sparse random graphs. *Advances in Applied Mathematics*, 26(4):257–279, 2001.

[8] Étienne Ghys and Pierre De La Harpe. Espaces métriques hyperboliques. *Sur les groupes hyperboliques d'après Mikhael Gromov*, pages 27–45, 1990.

[9] Shi Li and Gabriel H Tucci. Traffic congestion in expanders and (p, δ)–hyperbolic spaces. *Internet Mathematics*, 11(2):134–142, 2015.

[10] Anton Malyshev. Expanders are order diameter non-hyperbolic. *arXiv preprint arXiv:1501.07904*, 2015.

[11] Onuttom Narayan, Iraj Saniee, and Gabriel H Tucci. Lack of hyperbolicity in asymptotic erdös–renyi sparse random graphs. *Internet Mathematics*, 11(3):277–288, 2015.

[12] Oystein Ore. Theory of graphs. In *Colloquium Publications*. American Mathematical Society, 1962.

[13] Oliver Riordan and Nicholas Wormald. The diameter of sparse random graphs. *Combinatorics, Probability and Computing*, 19(5-6):835–926, 2010.

[14] Mauricio Abel Gomez Soto. *Quelques propriétés topologiques des graphes et applications à Internet et aux réseaux*. PhD thesis, Paris Diderot University, 2011.

[15] Ido Tishby, Ofer Biham, Reimer Kühn, and Eytan Katzav. The mean and variance of the distribution of shortest path lengths of random regular graphs. *Journal of Physics A: Mathematical and Theoretical*, 55(26):265005, 2022.

[16] Gabriel H Tucci. Non-hyperbolicity in random regular graphs and their traffic characteristics. *Central European Journal of Mathematics*, 11:1593–1597, 2013.

[17] Remco van der Hofstad. The giant in random graphs is almost local. *arXiv preprint arXiv:2103.11733*, 2021.

[18] Joon-Hyeok Yim and Anna Gilbert. Fitting trees to ell_1-hyperbolic distances. *Advances in Neural Information Processing Systems*, 36, 2024.

7 Appendix

7.1 Proof of Lemma 5.9

We will first show the following lemma:

Lemma 7.1. Given a bipartite graph $G = (U, V, E)$,

$$\sqrt{|E|} \leq \sum_{(u,v) \in E, u \in U, v \in V} \frac{1}{\sqrt{\deg(u) \cdot \deg(v)}}.$$

Proof.

$$
\begin{aligned}
\sum_{(u,v) \in E, u \in U, v \in V} \frac{1}{\sqrt{\deg(u) \cdot \deg(v)}} &= \sum_{u \in U} \frac{1}{\sqrt{\deg(u)}} \left(\sum_{v \in N(u)} \frac{1}{\sqrt{\deg(v)}} \right) \\
&\geq \sum_{u \in U} \frac{1}{\sqrt{\deg(u)}} \left(|N(u)| \cdot \frac{1}{\sqrt{M_u/|N(u)|}} \right) \quad (M_u := \sum_{v \in N(u)} \deg(v) \leq |E|) \\
&= \sum_{u \in U} \frac{1}{\sqrt{\deg(u)}} \cdot \frac{\deg(u)^{3/2}}{\sqrt{M_u}} = \sum_{u \in U} \frac{\deg(u)}{\sqrt{M_u}} \geq \sum_{u \in U} \frac{\deg(u)}{\sqrt{|E|}} = \sqrt{|E|}.
\end{aligned}
$$

$\square$

Now we will proceed the main proof of the Lemma. Let $\mathcal{C}$ denote the collection of colors and $c(x,y)$ denote the color of $(x,y) \in E$. Let $E_c := \{(x,y) \in E : c(x,y) = c\}$, $N_c(x) := \{y : (x,y) \in E_c\}$, and $\deg_c(x) := |N_c(x)|$ for $c \in \mathcal{C}$ and $x \in U \cup V$. Then by Cauchy-Schwarz inequality,

$$
\left(\sum_{\substack{(u,v) \in E, u \in U, v \in V \\ c = c(u,v)}} \frac{1}{\sqrt{\deg_c(u) \cdot \deg_c(v)}} \right)^2 \leq \left(\sum_{\substack{(u,v) \in E, u \in U, v \in V \\ c = c(u,v)}} \frac{1}{\deg_c(u)} \right) \left(\sum_{\substack{(u,v) \in E, u \in U, v \in V \\ c = c(u,v)}} \frac{1}{\deg_c(v)} \right).
$$

We see that

$$
\sum_{\substack{(u,v) \in E, u \in U, v \in V \\ c = c(u,v)}} \frac{1}{\deg_c(u)} = \sum_{u \in U} \sum_{v \in N(u), c = c(u,v)} \frac{1}{\deg_c(u)} = \sum_{u \in U} \sum_{\substack{c \in \mathcal{C} \\ N_c(u) \neq \emptyset}} \sum_{v \in N_c(u)} \frac{1}{\deg_c(u)} = \sum_{u \in U} \sum_{\substack{c \in \mathcal{C} \\ N_c(u) \neq \emptyset}} 1 \leq \sum_{u \in U} t(u),
$$

and similarly,

$$
\sum_{\substack{(u,v) \in E, u \in U, v \in V \\ c = c(u,v)}} \frac{1}{\deg_c(v)} \leq \sum_{v \in V} t(v)
$$

holds. On the other hand,

$$
\sum_{\substack{(u,v) \in E, u \in U, v \in V \\ c = c(u,v)}} \frac{1}{\sqrt{\deg_c(u) \cdot \deg_c(v)}} = \sum_{c \in \mathcal{C}} \sum_{(u,v) \in E_c, u \in U, v \in V} \frac{1}{\sqrt{\deg_c(u) \cdot \deg_c(v)}} \geq \sum_{c \in \mathcal{C}} \sqrt{|E_c|},
$$

by the previous lemma. This shows

$$\sum_{\substack{(u,v)\in E, u\in U, v\in V \\ c=c(u,v)}} \frac{1}{\sqrt{|E_c|}} = \sum_{c\in\mathcal{C}} \sqrt{|E_c|} \leq \sqrt{\left(\sum_{u\in U} t(u)\right)\cdot\left(\sum_{v\in V} t(v)\right)},$$

so that there exists $(u,v)\in E$ with $\frac{1}{\sqrt{|E_c|}} \leq \frac{\sqrt{\left(\sum_{u\in U} t(u)\right)\cdot\left(\sum_{v\in V} t(v)\right)}}{|E|}$ holds. Thus,

$$|E_c| \geq \frac{|E|^2}{\left(\sum_{u\in U} t(u)\right)\left(\sum_{v\in V} t(v)\right)},$$

as desired.

7.2 Proof of Lemma 5.14

For fixed x, denote $a_t := |\Gamma_t(x)|$ and $b_t := \lfloor (t+1)^2 \lambda^t \log n \rfloor$. We will show the following:
Claim: Given $0 \leq t \leq n$, $a_s \leq b_s$ holds for all $0 \leq s \leq t$ with probability at least $1 - t/n^{2\lambda}$.
Proof: We induct on t. For $t = 0$, $1 = a_0 \leq b_0 = \lfloor \log n \rfloor$, which is always true (if $n \geq 3$). Now, suppose the assertion holds on $t = r$. The following Chernoff type bound is well-known.

Lemma 7.2. Given a binomial distribution $X \sim \text{Bin}(n, p)$, we have

$$\mathbb{P}(X \geq np + a) \leq \exp\left(-\frac{a^2}{2(np + a/3)}\right) \quad \forall t \geq 0.$$

We use the fact that *conditional on* $a_0, a_1, \cdots, a_r$, the distribution of a_{r+1} is dominated by the distribution of the number of boundary edges $|\partial\Gamma_r(x)|$, which is binomial with parameters $(n - |N_r(x)|)\cdot a_r \leq na_r$ and $p = \lambda/n$. Suppose $a_0 \leq b_0, \cdots, a_r \leq b_r$. Then

$$\mathbb{P}(a_{r+1} > b_{r+1}|a_0, \cdots, a_r) = \mathbb{P}(a_{r+1} > (r+2)^2 \lambda^{r+1} \log n | a_0, \cdots, a_r)$$

$$\leq \mathbb{P}\left(\text{Bin}\left(na_r, \frac{\lambda}{n}\right) > (r+2)^2 \lambda^{r+1} \log n\right)$$

$$\leq \mathbb{P}\left(\text{Bin}\left(nb_r, \frac{\lambda}{n}\right) > (r+2)^2 \lambda^{r+1} \log n\right) \quad (\because a_r \leq b_r)$$

$$\leq \exp\left[-\frac{\left((r+2)^2 \lambda^{r+1} \log n - \lambda b_r\right)^2}{2\left(\lambda b_r + \frac{(r+2)^2 \lambda^{r+1} \log n - \lambda b_r}{3}\right)}\right]$$

$$\leq \exp\left[-\frac{\left((r+2)^2 \lambda^{r+1} \log n - (r+1)^2 \lambda^{r+1} \log n\right)^2}{2\left(\frac{2(r+1)^2 \lambda^{r+1} \log n + (r+2)^2 \lambda^{r+1} \log n}{3}\right)}\right] \quad (\because \lambda b_r \leq (r+1)^2 \lambda^{r+1} \log n)$$

$$= \exp\left[-\lambda^{r+1} \log n \cdot \frac{3(4r^2 + 12r + 9)}{2(3r^2 + 8r + 6)}\right] \leq \exp\left(-2\lambda^{r+1} \log n\right) \leq \frac{1}{n^{2\lambda}}.$$

By the induction hypothesis, we see that $a_0 \leq b_0, \cdots, a_r \leq b_r$ holds with probability at least $1 - r/n^{2\lambda}$ and conditional on that, the probability of $a_{r+1} > b_{r+1}$ is bounded by $1/n^{2\lambda}$. Therefore, $a_0 \leq b_0, \cdots, a_{r+1} \leq b_{r+1}$ holds with probability at least $1 - (r+1)/n^{2\lambda}$ which concludes the proof. $\qquad\square$

Now, let $B(x)$ be the *bad* event that $|\Gamma_t(x)| > (t+1)^2 \lambda^t \log n$ for some t. By the claim, $\mathbb{P}(B(x)) \leq n^{1-2\lambda}$. Therefore, for $B := \cup_x B(x)$,

$$\mathbb{P}(B) \leq \sum_x \mathbb{P}(B(x)) \leq n^{2-2\lambda} = o(1),$$

as desired.