Node Expansions and Cuts in Gromov-Hyperbolic Graphs

Bhaskar DasGupta* Marek Karpinski† Nasim Mobasheri* Farzaneh Yahyanejad*

Abstract

Gromov-hyperbolic graphs (or, hyperbolic graphs for short) are a non-trivial interesting classes of non-expander graphs. Originally conceived by Gromov in 1987 in a different context while studying fundamental groups of a Riemann surface, the hyper bolicity measure for graphs has recently been a quite popular measure in the network science community in quantifying curvature and closeness to a tree topology for a given network, and ma ny real-world networks have been empirically observed to be hyperbolic.

In this paper, we provide *constructive* non-trivial bounds on node expansions and cut-sizes for hyperbolic graphs, and show that witnesses for such non- expansion or cut-size can in fact be computed efficiently in polynomial time. We also provide so me algorithmic consequences of these bounds and their related proof techniques for a few pro blems related to cuts and paths *for hyperbolic graphs*, such as the existence of a large family of s-t cuts with *small* number of cut-edges for when s and t are at least logarithmically far apart, efficient approximat ion of hitting sets of size-constrained cuts, and a polynomial-time solution for a type of small-set expansion problem originally proposed by Arora, Barak and Steurer.

*Department of Computer Science, University of Illinois at Chicago, Chicago, IL 60607. Emails: {bdas-gup,nmobash2,fyahya2}@uic.edu. Supported by NSF grant IIS-1160995.

†Department of Computer Science, University of Bonn, Bonn 53113. Research partially supported by DFG grants. Email: marek@cs.uni-bonn.de

1 Introduction

Useful insights for many complex systems such as the world-wide web, social networks, metabolic networks, and protein-protein interaction networks can often be obtained by representing them as parameterized networks and analyzing them using graph-theoretic tools. Some standard measures used for such analyses include degree based measures (*e.g.*, maximum/minimum/average degree or degree distribution) connectivity based measures (*e.g.*, clustering coefficient, largest cliques or densest sub-graphs), and geodesic based measures (*e.g.*, diameter or betweenness centrality). In this paper we consider a combinatorial measure called *Gromov-hyperbolicity* (or, hyperbolicity for short) for *finite* undirected unweighted graphs that has recently received significant attention from researchers in both the graph theory and the network science community. The hyperbolicity measure was originally conceived in a somewhat different group-theoretic context by Gromov in 1987 [18] from an observation that many results concerning the fundamental group of a Riemann surface hold true in a more general context. The measure was first defined for *infinite* continuous metric space with *bounded* local geometry via properties of geodesics [9], but was later also adopted for *finite* graphs. Off late, there has been a surge of theoretical and empirical works measuring and analyzing the hyperbolicity of networks, and many *real-world* networks have been reported to be hyperbolic. For example, *preferential attachment* scale-free networks were reported to be hyperbolic with appropriate scaling (normalization) in [19], networks of high power transceivers in a wireless sensor network were empirically observed to be hyperbolic in [2], communication networks at the IP layer and at other levels were empirically observed to be hyperbolic in [25], an assorted set of biological and social networks were empirically observed to be hyperbolic in [1], and extreme congestion at a small number of nodes in a large traffic network using the shortest-path routing was shown in [21] to be caused due to hyperbolicity of the network. On the other hand, theoretical investigations have revealed that *expanders*, *vertex-transitive* graphs and classical *Erdös-Rényi* random graphs are *not* hyperbolic [6–8, 23].

In this paper, we further expand on the non-expander properties of hyperbolic networks shown in [6, 23] and provide constructive proofs of various kinds of *witnesses* (subsets of nodes) of small expansion or small cut-size. We also provide some algorithmic consequences of these bounds and their related proof techniques for a few problems related to cuts and paths for hyperbolic graphs.

Basic Notations and Assumptions

We will use the following notations and terminologies throughout the paper. We will simply write log to refer to logarithm base 2. Our basic input is an ordered triple $\langle G, d, \delta \rangle$ denoting the given *connected undirected unweighted* graph $G = (V, E)$ having m edges and n nodes of hyperbolicity δ in which every node has a degree of at most $d > 2$. *Throughout the paper, we assume that n is always sufficiently large.* For notational convenience, **we will ignore floors and ceilings of fractional values** in our theorems and proofs, *e.g.*, we will simply write $D/3$ instead of $\lfloor D/3 \rfloor$ or $\lceil D/3 \rceil$, since this will have *no* effect on the *asymptotic* nature of the bounds. **We will also make no serious effort to optimize the constants that appear in the bounds in our theorems and proofs**. In addition, the following notations will be used throughout the paper:

- $|\mathcal{P}|$ is the *length* (number of edges) of a path $\mathcal{P}$.
- $\overline{u,v}$ is a *shortest path* between nodes u and v. In our proofs, any shortest path can be selected but, once selected, the *same* shortest path *must* be used in the remaining part of the analysis.
- $\mathsf{dist}_H(u, v)$ is the distance (number of edges in a shortest path) between u and v in a graph H (or ∞ if there is no path between u and v in H).

- $D(H) = \max_{u,v \in V'} \{\mathsf{dist}_H(u,v)\}$ is the *diameter* D of a graph $H = (V', E')$. Thus, in particular, for our input graph G there exists two nodes p and q such that $\mathsf{dist}_G(p,q) = D(G) \geq \log_d n$.
- For a subset S of nodes of the graph $H = (V', E')$, the *boundary* $\partial_H(S)$ of S is the set of nodes in $V \setminus S$ that are connected to *at least* one node in S, *i.e.*, $\partial_H(S) = \{u \in V' \setminus S \,|\, v \in S \,\&\, \{u,v\} \in E'\}$.
- $\mathcal{B}_H(u,r) = \{v \,|\, \mathsf{dist}_H(u,v) \leq r\}$ is the set of nodes contained in a *ball* of radius r centered at node u in a graph H.

Definitions of Gromov-hyperbolicity

Often the hyperbolicity measure is introduced via the thin geodesic triangles in the following manner.

Definition 1 (δ-hyperbolic graphs)[1] *A graph G has a (Gromov) hyperbolicity of δ, or simply is δ-hyperbolic, if and only if for every three ordered triple of shortest paths $(\overline{u,v}, \overline{u,w}, \overline{v,w})$, $\overline{u,v}$ lies in a δ-neighborhood of $\overline{u,w} \cup \overline{v,w}$, i.e., for every node x on $\overline{u,v}$, there exists a node y on $\overline{u,w}$ or $\overline{v,w}$ such that $\mathsf{dist}_G(x,y) \leq \delta$. A δ-hyperbolic graph is simply called a hyperbolic graph if δ is a constant.*

There is another alternate but *equivalent* ("up to a constant multiplicative factor") way of defining hyperbolicity of graphs via 4-node conditions.

Definition 2 (equivalent definition of δ-hyperbolic graphs via 4-node conditions) *For a set of four nodes u_1, u_2, u_3, u_4, let $\pi = (\pi_1, \pi_2, \pi_3, \pi_4)$ be a permutation of $\{1,2,3,4\}$ denoting a rearrangement of the indices of nodes such that $S = d_{u_{\pi_1}, u_{\pi_2}} + d_{u_{\pi_3}, u_{\pi_4}} \leq M = d_{u_{\pi_1}, u_{\pi_3}} + d_{u_{\pi_2}, u_{\pi_4}} \leq L = d_{u_{\pi_1}, u_{\pi_4}} + d_{u_{\pi_2}, u_{\pi_3}}$, and let $\rho_{u_1,u_2,u_3,u_4} = \frac{L-M}{2}$. Then, a graph G is δ-hyperbolic if and only if $\delta = \max\limits_{u_1,u_2,u_3,u_4 \in V} \left\{ \rho_{u_1,u_2,u_3,u_4} \right\}$.*

It is well-known [9] that the two above definitions of hyperbolicity are equivalent in the sense that they are related by a constant factor, *i.e.*, there is a constant $c > 0$ such that if a graph G is δ_1-hyperbolic and δ_2-hyperbolic via Definition 1 and Definition 2, respectively, then $\frac{1}{c}\delta_1 \leq \delta_2 \leq c\delta_1$. Since constant factors are not optimized in our proofs, we will use *either* of the two definitions of hyperbolicity in the sequel as deemed more convenient. Using Definition 2 and casting the resulting computation as a $(\max, \min)$ matrix multiplication problem allows one to compute δ and a 2-approximation of δ in $O\left(n^{3.69}\right)$ and in $O\left(n^{2.69}\right)$ time, respectively [15]. Several routing-related problems or the diameter estimation problem become easier if the network is hyperbolic [10–12, 17].

The hyperbolicity property enjoys many non-trivial topological characteristics. For example, adding a single node or edge can increase/decrease the value of δ abruptly (*e.g.*, a cycle of n nodes has $\delta = \lceil n/4 \rceil$ but removing a node or an edge has a value of $\delta = 0$ for the resulting graph). Examples of hyperbolic graphs include *trees*, *chordal graphs*, *cactus of cliques*, *AT-free* graphs, *link graphs of simple polygons*, and *any* class of graphs with a *fixed* diameter, whereas examples of non-hyperbolic graphs include *expanders*, *a simple cycle*, and *Erdös-Rényi random graphs* for some parameter ranges.

Note that if G is δ-hyperbolic then G is also δ'-hyperbolic for any $\delta' > \delta$ (cf. Definition 1). In this paper, to avoid division by zero in terms involving $1/\delta$, we will assume $\delta > 0$. In other words, we will treat a 0-hyperbolic graph (a tree) as a $\frac{1}{2}$-hyperbolic graph in the analysis. Alternatively, one can also replace such a division by a sufficiently large positive value.

[1] Often the definition of a hyperbolic space assumes a *bounded local geometry* (*i.e.*, the degree of nodes is uniformly bounded) [9, 20].

2 Overview of Our Results and Proof Techniques

Before proceeding with formal theorems and proofs, we first provide a informal non-technical intuitive overview of our results and proof techniques.

Overview of Results

⋆ Our three results in Section 3 provide non-trivial upper bounds for *node expansions* for the triple $\langle G, d, \delta \rangle$ as a function of n, $D(G)$, d, and δ. The first result, namely, Theorem 3, provides an absolute bound and shows that a *witness* (subset of nodes) of small expansion can in fact be found efficiently in polynomial time. The next two results, namely Theorems 5 and 6, generalize Theorem 3 (sometimes at the expense of *slightly* worse expansion bounds) to show that in fact *many* such witnesses can be found in polynomial time even satisfying criteria such as

- the witnesses (subsets) form a nested (laminar) family, or
- the witnesses (subsets) have *limited* overlap in the sense that every subset has a certain number of nodes *not* contained in any other subset.

These bounds also imply in an obvious manner corresponding upper bounds for the *edge-expansion* of G and for the smallest non-zero eigenvalue of the Laplacian of G.

To illustrate the non-trivialness of these bounds, suppose that the maximum degree d and the hyperbolicity value δ grows asymptotically very slowly[2] with respect to the number of nodes n, and the diameter D to be of the order of the minimum possible value of $\log_d n$. In Remark 1, we provide an explanation of the asymptotics of these bounds in comparison to expander-type graphs. In particular, if δ is fixed (*i.e.*, G is hyperbolic) then d has to be increased to *at least* $2^{\Omega\left(\sqrt{\log\log n \,/\, \log\log\log n}\right)}$ to get a positive non-zero Cheeger constant, whereas if d is fixed then δ need to be at least $\Omega(\log n)$ to get a positive non-zero Cheeger constant (this last implication also follows from the results in [6, 23]).

⋆ Our result in Section 4, namely Theorem 7, deals with the *absolute* size of s-t cuts in hyperbolic graphs, and shows that a large family of s-t cuts having at most $d^{O(1)}$ cut-edges can be found in polynomial time in hyperbolic graphs when every node other than s and t has a maximum degree of d and the distance between s and t is at least $\Omega(\log n)$. This result was instrumental in designing the approximation algorithm for size-constrained hitting set problem in Section 5.1 (Theorem 10).

⋆ In Section 5 we discuss some applications of these bounds in designing improved approximation algorithms for hyperbolic graphs for several combinatorial problems.

- We first show in Section 5.1 that our bounds can be used to give a *logarithmic* approximation for *size-constrained cuts* which otherwise is *provably* much more harder to approximate.
- We then show in Section 5.2 that the problem of identifying *vulnerable edges* by *minimizing shared edges* can be cast as a hitting set problem for size-constrained cuts, thereby extending the previous logarithmic approximation to this problem as well.
- Finally, in Section 5.3 we provide a *polynomial-time* solution for a type of small-set expansion problem originally proposed by Arora, Barak and Steurer [3] if the given graph is hyperbolic.

[2]As we remarked before, often the definition of a hyperbolic graph assumes d is fixed [9, 20].

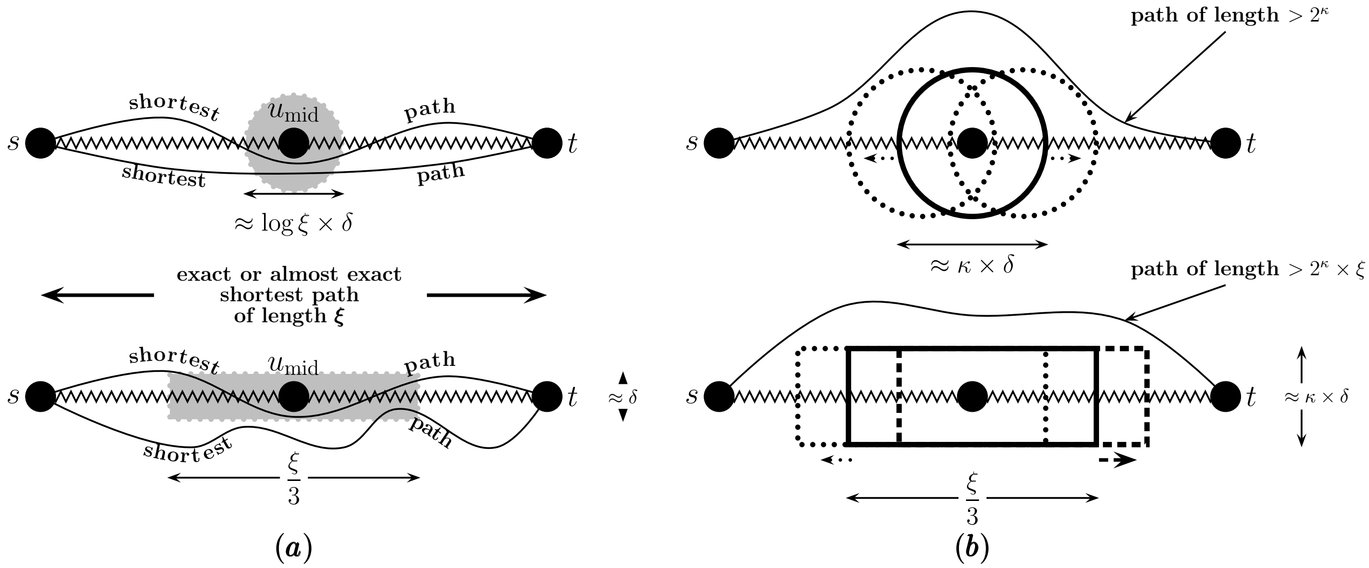

Figure 1: Effect of node removal in a hyperbolic graph. (a) All shortest paths between s and t must intersect the shaded region around $u_{\rm mid}$. (b) By growing the shaded region in (a) and removing nodes in its boundary, one can selectively extract longer paths in the graph. Translating the region somewhat does not change this property much.

Proof Techniques

⋆ One key ingredient in many of the proofs is the process *extraction* of longer paths in a hyperbolic graph in a systematic manner by removing "not too many" nodes, a slightly stronger but special version of which appears in Fact 1, and a slightly weaker but more general version of which appears in Fact 2. To illustrate the process, consider the situations depicted in Fig. 1. By removing the boundary of a specific subset of nodes (the figure shows a cylinder and a ball, but other subsets are possible) around the middle node of the shortest path between nodes s and t of the correct size, one can in effect destroy *every* shortest path between s and t (see Fig. 1 (a)). As illustrated in Fig. 1 (b), enlarging the shaded region appropriately allows us to extract longer paths in the given hyperbolic graph and small translations of the region does not significantly alter this property.

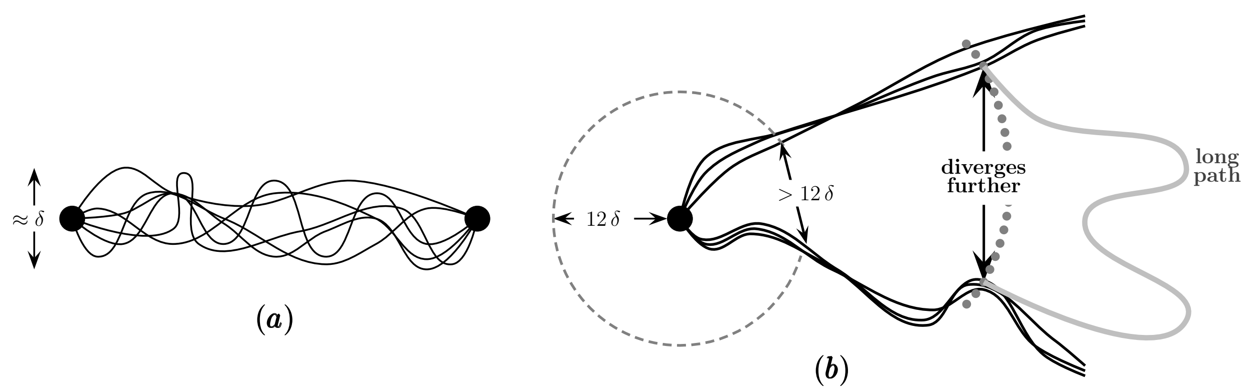

Figure 2: Divergence of geodesic rays. (***a***) Geodesic rays may continue to follow each other closely. (***b***) Once geodesic rays diverge sufficiently, they cannot connect back without using a sufficiently long path.

⋆ Another important property that is used especially in the proof of Theorem 7 on small-size cuts is the so-called "*exponential divergence property of geodesic rays emanating from the same node*"[3], namely the property that once geodesic rays emanating from the same node move sufficiently far apart they continue to keep diverging at an exponential rate and shortest paths between them tends to bend "inwards" towards the emanating node (see Fig. 2). This type of behavior of geodesics seems crucial in designing small-size cuts as in Theorem 7.

3 Node Expansion Bounds for Hyperbolic Graphs

The three results in this section is related to the node expansion ratios of a hyperbolic graph. The *node expansion ratio* $h_H(S)$ of a subset S of at most $|V'|/2$ nodes of a graph $H = (V', E')$ is defined as $h_H(S) = \frac{|\,\partial_H(S)\,|}{|\,S\,|}$. Let $\mathfrak{h}_H = \min_{S \subset V' \,:\, |S| \leq |V'|/2} \{h_H(S)\}$ be the *minimum node expansion ratio* of H. Since any subset S containing exactly $|V'|/2$ nodes has $|\,\partial_H(S)\,| \leq |V'|/2$, $\mathfrak{h}_H$ satisfies $0 < \mathfrak{h}_H \leq 1$ for any graph H.

The *edge expansion ratio* of a graph is also defined in a similar manner. For any subset S of nodes of a graph H, define $\text{cut}_H(S)$ as the set of edges that have *exactly* one end-point in S. The edge expansion ratio $\mathscr{g}_H(S)$ of a subset S of at most $|V'|/2$ nodes of a graph $H = (V', E')$ is defined as $\mathscr{g}_H(S) = \frac{|\,\text{cut}_H(S)\,|}{|\,S\,|}$. Obviously $\mathscr{g}_H(S) \leq d h_H(S)$ and thus the bounds in Theorems 3 5 and 6 for node expansions translate to some bounds of the edge expansions as well.

3.1 Bound for Minimum Node Expansion and a Polynomial-Time Witness

Theorem 3 *For any constant*[4] $0 < \mu < 1$*, the following results hold for* $\langle G, d, \delta \rangle$ *with* $D = D(G)$*:*

- $\mathfrak{h}_G \leq \Lambda = \min \left\{ \dfrac{4 \ln \left(\frac{n}{2}\right)}{D}, \max \left\{ \left(\dfrac{1}{D}\right)^{1-\mu}, \dfrac{360 \log n}{D\, 2^{\frac{D^\mu}{7\,\delta\, \log(2d)}}} \right\} \right\}$
- *A subset of nodes* S *such that* $h_G(S) \leq \Lambda$ *can be found in* $O\left(n^2 \log n + mn\right)$*.*

Corollary 4 *Since* $D > \frac{\log n}{\log d}$*, the bound in Theorem 3 implies:*

$$\mathfrak{h}_G < \max \left\{ \left(\frac{\log d}{\log n}\right)^{1-\mu}, \frac{360 \log d}{2^{\frac{\log^\mu n}{7\,\delta\,\log^{1+\mu}(2d)}}} \right\}$$

Remark 1 *The following observations may help the reader to understand the asymptotic nature of the bound in Corollary 4.*

(a) *The first component of the bound is* $O\left({}^{1}/_{\log^{1-\mu} n}\right)$ *for fixed* d*, and is* $\Omega(1)$ *only when* $d = \Omega(n)$*.*

(b) *To better understand the second component of the bound, consider the following cases (note that an expander has a constant value for* $\mathfrak{h}_G$*):*

- *Suppose that the given graph is a hyperbolic graph of constant maximum degree,* i.e., *both* δ *and* d *are constants. In that case,*

$$\frac{360 \log d}{2^{\frac{\log^\mu n}{7\,\delta\,\log^{1+\mu}(2d)}}} = O\left(\frac{1}{2^{O(1)\,\log^\mu n}}\right) = O\left(\frac{1}{\text{polylog}(n)}\right)$$

[3] Indeed, some texts *define* a hyperbolic space as one having such a property.

[4] Actually, this proof (and that of Theorems 5 and 6) works even if $o(1) \leq \mu \leq 1 - o(1)$ where $o(1)$ is a *sufficiently* slowly decreasing function of n, *i.e.*, $\frac{1}{\log^* n}$ for $o(1)$ suffices.

- *Suppose that the given graph is a hyperbolic graph but the maximum degree d is arbitrary. In that case,*

$$\frac{360\log d}{2^{\frac{\log^{\mu} n}{7\,\delta\,\log^{1+\mu}(2d)}}} = O\left(\frac{\log d}{2^{O(1)\,\frac{\log^{\mu} n}{\log^{1+\mu} d}}}\right) = O\left(\frac{\log d}{\mathrm{polylog}(n)^{\frac{1}{\log^{2} d}}}\right)$$

and thus d has to be increased to at least $2^{\Omega\left(\sqrt{\frac{\log\log n}{\log\log\log n}}\right)}$ to get a constant upper bound.

- *Suppose that the given graph has a constant maximum degree but not necessarily hyperbolic (*i.e., *δ is arbitrary). In that case,*

$$\frac{360\log d}{2^{\frac{\log^{\mu} n}{7\,\delta\,\log^{1+\mu}(2d)}}} = O\left(\frac{1}{2^{O(1)\frac{\log^{\mu} n}{\delta}}}\right)$$

and thus δ need to be at least $\Omega\left(\log^{\mu} n\right)$ to get a constant upper bound.

Proof. The bound of $\mathfrak{h}_G < \frac{4\ln\left(\frac{n}{2}\right)}{D}$ holds for any graph. Obviously, the bound is non-trivial only if $D > 4\ln\left(\frac{n}{2}\right)$ and in that case it follows easily from standard results in spectral graph theory that relates the diameter to the smallest non-zero eigenvalue λ_1 of the Laplacian of the graph (*e.g.*, Theorem 3.1 in [13]), and using the standard Cheeger inequality that states that $\mathfrak{h}_G \leq \sqrt{2\,\lambda_1}$. Here we give an elementary proof. Let p and q be two nodes such that $\mathsf{dist}_G(p,q) = D$. Assume, without loss of generality, that $\left|\mathcal{B}_G\left(p,\frac{D}{2}\right)\right| \leq \min\left\{\left|\mathcal{B}_G\left(p,\frac{D}{2}\right)\right|, \left|\mathcal{B}_G\left(q,\frac{D}{2}\right)\right|\right\} \leq {}^{n}\!/_{2}$. Consider the sequence of balls $\mathcal{B}_G(p,r)$ for $r = 0,1,2\dots, {}^{D}\!/_{2}$. Thus it follows that

$$\frac{n}{2} > \left|\mathcal{B}_G\left(p,\frac{D}{2}\right)\right| \geq \left(1+\mathfrak{h}_G\right)^{D/2}$$

$$\Rightarrow \ln\left(\frac{n}{2}\right) > \frac{D}{2}\ln\left(1+\mathfrak{h}_G\right) \geq \left(\frac{D}{2}\right)\left(\frac{\mathfrak{h}_G}{1+\mathfrak{h}_G}\right) \geq \frac{D\mathfrak{h}_G}{4} \Rightarrow \mathfrak{h}_G < \frac{4\ln\left(\frac{n}{2}\right)}{D}$$

Such a bound for $\mathfrak{h}_G$ is realized by a ball $\mathcal{B}_G(p,r)$ for some $0 \leq r \leq {}^{D}\!/_{2}$ and therefore can be found within the desired time complexity bound.

Thus, in the remaining part of the proof, we concentrate on the other two bounds only. First, assume that $D(n) = c$ for any some constant $c \geq 1$ (independent of n). Then, since $\delta \geq {}^{1}\!/_{2}$ and $d > 1$, we have

$$\frac{360\log n}{D\,2^{\frac{D^{\mu}}{7\,\delta\,\log(2d)}}} > \frac{360\log n}{c\,2^{\,c^{\mu}}} > 1 \quad \text{(since } n \text{ is sufficiently large)}$$

and the claimed bound is trivially true for *any* subset of ${}^{n}\!/_{2}$ nodes.

Otherwise, assume that $D(n) = \omega(1)$, *i.e.*, $\lim\limits_{n\to\infty} D(n) > c$ for any constant c. We reuse the same analysis as in [23] which improves upon the analysis in [6]. Let p and q be two nodes such that $\mathsf{dist}_G(p,q) = D$. Let p', q' be nodes on a shortest path between p and q such that $\mathsf{dist}_G(p,p') = \mathsf{dist}_G(p',q') = \mathsf{dist}_G(q',q) = {}^{D}\!/_{3}$. Set $0 < \alpha < {}^{1}\!/_{4}$ to be as follows:

$$\alpha = \frac{1}{7\,D^{1-\mu}\,\log(2d)} \tag{1}$$

Let $\mathcal{C}$ be set of nodes at a distance of $\lfloor\alpha D\rfloor > \alpha D - 1$ of a shortest path $\overline{p',q'}$ between p' and q'. Thus,

$$\mathcal{C} = \left\{u \,|\, \exists\, v \in \overline{p',q'} \colon \mathsf{dist}_G(u,v) = \lceil\alpha D\rceil\right\} \Rightarrow |\,\mathcal{C}\,| \leq \frac{D}{3}\,d^{\,\lfloor\alpha D\rfloor} < \frac{D}{3}\,d^{\alpha D} \tag{2}$$

Let $G_{-\mathcal{C}}$ be the graph obtained from G by removing the nodes in $\mathcal{C}$. We recall the following fact proved in [23].

Fact 1 (Cylinder removal around a geodesic) [23] *Assume that G is a δ-hyperbolic graph. Let p and q be two nodes of G such that $\mathsf{dist}_G(p,q) = \beta > 6$, and let p', q' be nodes on a shortest path between p and q such that $\mathsf{dist}_G(p,p') = \mathsf{dist}_G(p',q') = \mathsf{dist}_G(q',q) = \beta/3$. For any $0 < \alpha < 1/4$, let $\mathcal{C}$ be set of nodes at a distance of $\alpha\beta - 1$ of a shortest path $\overline{p',q'}$ between p' and q',* i.e.,

$$\mathcal{C} = \left\{ u \,|\, \exists\, v \in \overline{p',q'} \colon \mathsf{dist}_G(u,v) = \alpha\beta - 1 \right\}$$

Let $G_{-\mathcal{C}}$ be the graph obtained from G by removing the nodes in $\mathcal{C}$. Then, $\mathsf{dist}_{G_{-\mathcal{C}}}(p,q) \geq \frac{\beta}{60}\, 2^{\alpha\beta/\delta}$.

By Fact 1,

$$\mathsf{dist}_{G_{-\mathcal{C}}}(p,q) \geq \frac{D}{60}\, 2^{\alpha D/\delta} \tag{3}$$

For a ball $\mathcal{B}_G(p,r)$ of radius r centered at node p in G with $|\,\mathcal{B}_G(p,r)\,| \leq n/2$, we have :

$$|\,\mathcal{B}_G(p,0)\,| = 1 \;\text{ and }\; \frac{|\mathcal{B}_G(p,r)|}{|\mathcal{B}_G(p,r-1)|} \geq 1 + \mathfrak{h}_G \;\;\Rightarrow\;\; |\,\mathcal{B}_G(p,r)\,| \geq (1+\mathfrak{h}_G)^r \tag{4}$$

Assume without loss of generality that[5]

$$\begin{aligned}
&\left| \mathcal{B}_{G_{-\mathcal{C}}}\left(p, \frac{\mathsf{dist}_{G_{-\mathcal{C}}}(p,q)}{2}\right) \right| \leq \left| \mathcal{B}_{G_{-\mathcal{C}}}\left(q, \frac{\mathsf{dist}_{G_{-\mathcal{C}}}(p,q)}{2}\right) \right| \\
\Rightarrow &\left| \mathcal{B}_{G_{-\mathcal{C}}}\left(p, \frac{\mathsf{dist}_{G_{-\mathcal{C}}}(p,q)}{2}\right) \right| \leq \frac{\left| \mathcal{B}_{G_{-\mathcal{C}}}\left(p, \frac{\mathsf{dist}_{G_{-\mathcal{C}}}(p,q)}{2}\right) \right| + \left| \mathcal{B}_{G_{-\mathcal{C}}}\left(q, \frac{\mathsf{dist}_{G_{-\mathcal{C}}}(p,q)}{2}\right) \right|}{2} \leq \frac{n - |\mathcal{C}|}{2} < \frac{n}{2}
\end{aligned} \tag{5}$$

Define $\hbar$ as

$$\hbar = \min_{0 \leq r \leq d_{G_{-\mathcal{C}}}(p,q)/2} \left\{ h_G\left(\mathcal{B}_{G_{-\mathcal{C}}}(p,r)\right) \right\} \geq \mathfrak{h}_G \tag{6}$$

If $\hbar \leq (1/D)^{1-\mu}$ then $\mathfrak{h}_G \leq \hbar \leq (1/D)^{1-\mu}$, and there exists a subset of nodes $S = \mathcal{B}_{G_{-\mathcal{C}}}(p,r')$ for some r' (with $|S| \leq n/2$) such that $h_G(S) \leq (1/D)^{1-\mu}$.

Otherwise, assume that $\hbar > (1/D)^{1-\mu}$. Let r_p be the least integer such that $\mathcal{B}_{G_{-\mathcal{C}}}(p, r_p) = \mathcal{B}_{G_{-\mathcal{C}}}(p, r_p+1)$. Since G is a connected graph and

$$\forall\, r \leq \frac{D}{3} - \alpha D :\; \mathcal{B}_G(p,r) \cap \mathcal{C} = \emptyset \;\equiv\; \mathcal{B}_{G_{-\mathcal{C}}}(p,r) = \mathcal{B}_G(p,r) \;\Rightarrow\; \mathcal{B}_{G_{-\mathcal{C}}}(p,r) > (1+\hbar)^r \tag{7}$$

we have $r_p \geq \frac{D}{3} - \alpha D$.

We first show that our choice of α ensures that $r_p > \mathsf{dist}_{G_{-\mathcal{C}}}(p,q)/2$. Suppose for the sake of contradiction that $r_p \leq \mathsf{dist}_{G_{-\mathcal{C}}}(p,q)/2$. Then,

$$|\,\mathcal{B}_{G_{-\mathcal{C}}}(p,r_p)\,| \underset{\text{since } r_p \geq \frac{D}{3} - \alpha D}{>} \left| \mathcal{B}_{G_{-\mathcal{C}}}\left(p, \frac{D}{3} - \alpha D\right) \right| \underset{\text{by (7)}}{=} \left| \mathcal{B}_G\left(p, \frac{D}{3} - \alpha D\right) \right| \underset{\text{by (4)}}{\geq} (1+\hbar)^{\frac{D}{3} - \alpha D}$$

[5]Note that if there is no path between nodes p and q in $G_{-\mathcal{C}}$ then $d_{G_{-\mathcal{C}}}(p,q) = \infty$ and hence $\mathcal{B}_{G_{-\mathcal{C}}}\left(p, \frac{d_{G_{-\mathcal{C}}}(p,q)}{2}\right)$ and $\mathcal{B}_{G_{-\mathcal{C}}}\left(q, \frac{d_{G_{-\mathcal{C}}}(p,q)}{2}\right)$ contains all the nodes reachable from p and q, respectively, in $G_{-\mathcal{C}}$.

$$\left|\,\partial_{G_{-\mathcal{C}}}\left(\mathcal{B}_{G_{-\mathcal{C}}}(p,r_p)\right)\right| = 0 \;\Rightarrow\; \left|\,\partial_{G}\left(\mathcal{B}_{G_{-\mathcal{C}}}(p,r_p)\right)\right| \le |\,\mathcal{C}\,| \underset{\text{by (2)}}{<} \frac{D}{3}d^{\alpha D}$$

$$h_G\left(\mathcal{B}_{G_{-\mathcal{C}}}(p,r_p)\right) = \frac{\left|\,\partial_{G}\left(\mathcal{B}_{G_{-\mathcal{C}}}(p,r_p)\right)\right|}{\left|\,\mathcal{B}_{G_{-\mathcal{C}}}(p,r_p)\,\right|} < \frac{\frac{D}{3}d^{\alpha D}}{(1+\hbar)^{\frac{D}{3}-\alpha D}}$$

By our choice of α in (1):

$$d^{\alpha D} = d^{\frac{D^\mu}{7\log(2d)}} < \left(d^{1/\log d}\right)^{D^\mu/7} = 2^{D^\mu/7}$$

$$\frac{D}{3} - \alpha D = \frac{D}{3} - \frac{D^\mu}{7\log(2d)} > \frac{D}{4}, \quad \text{since } \mu < 1 \text{ and } n \text{ (and hence } D\text{) is sufficiently large}$$

$$(1+\hbar)^{\frac{D}{3}-\alpha D} > (1+\hbar)^{\frac{D}{4}} = \left[(1+\hbar)^{1/\hbar}\right]^{\frac{\hbar D}{4}} \underset{\text{since } 0<\hbar\le 1}{>} 2^{\frac{\hbar D}{4}} > 2^{\frac{\frac{1}{D^{1-\mu}}D}{4}} = 2^{D^\mu/4}$$

$$h_G\left(\mathcal{B}_{G_{-\mathcal{C}}}(p,r_p)\right) < \frac{\frac{D}{3}d^{\alpha D}}{(1+\hbar)^{\frac{D}{3}-\alpha D}} < \frac{\frac{D}{3}2^{D^\mu/7}}{2^{D^\mu/4}} < \frac{D/3}{2^{D^\mu/14}} < \frac{1}{D^{1-\mu}}, \quad \text{since } \mu>0 \text{ and } n \text{ (and hence } D\text{) is sufficiently large} \tag{8}$$

Inequality (8) contradicts the fact that $h_G\left(\mathcal{B}_{G_{-\mathcal{C}}}(p,r_p)\right) \ge \hbar > (1/D)^{1-\mu}$.

Thus, for the remaining proof, we assume that $r_p > \mathsf{dist}_{G_{-\mathcal{C}}}(p,q)/2$. We want to ensure that removal of nodes in $\mathcal{C}$ does not decrease the expansion of the balls $\mathcal{B}_{G_{-\mathcal{C}}}(p,r)$ in the new graph $G_{-\mathcal{C}}$ by more than a factor of 2, *i.e.*, the reduced expansion must be at least $\hbar/2$. If $r \le \frac{D}{3} - \alpha D$ this is not a problem since by (7)

$$\forall\, r < \frac{D}{3} - \alpha D:\; h_{G_{-\mathcal{C}}}\left(\mathcal{B}_{G_{-\mathcal{C}}}(p,r)\right) = h_G\left(\mathcal{B}_G(p,r)\right) \ge \hbar$$

For $\frac{D}{3} - \alpha D < r \le \mathsf{dist}_{G_{-\mathcal{C}}}(p,q)/2$ we will need to ensure the following:

$$\begin{aligned}
& h_{G_{-\mathcal{C}}}\left(\mathcal{B}_{G_{-\mathcal{C}}}(p,r-1)\right) \ge \frac{\hbar}{2} \\
\equiv\; & \frac{\left|\,\partial_{G_{-\mathcal{C}}}\left(\mathcal{B}_{G_{-\mathcal{C}}}(p,r-1)\right)\right|}{\left|\,\mathcal{B}_{G_{-\mathcal{C}}}(p,r-1)\,\right|} = \frac{\left|\,\partial_{G}\left(\mathcal{B}_{G_{-\mathcal{C}}}(p,r-1)\right)\right| - \left|\,\partial_{G}\left(\mathcal{B}_{G_{-\mathcal{C}}}(p,r-1)\right)\cap\mathcal{C}\,\right|}{\left|\,\mathcal{B}_{G_{-\mathcal{C}}}(p,r-1)\,\right|} \ge \frac{\hbar}{2} \\
\Leftarrow\; & \frac{\left|\,\partial_{G}\left(\mathcal{B}_{G_{-\mathcal{C}}}(p,r-1)\right)\right| - |\mathcal{C}|}{\left|\,\mathcal{B}_{G_{-\mathcal{C}}}(p,r-1)\,\right|} \ge \frac{\hbar}{2} \equiv \frac{\left|\,\partial_{G}\left(\mathcal{B}_{G_{-\mathcal{C}}}(p,r-1)\right)\right|}{\left|\,\mathcal{B}_{G_{-\mathcal{C}}}(p,r-1)\,\right|} - \frac{|\mathcal{C}|}{\left|\,\mathcal{B}_{G_{-\mathcal{C}}}(p,r-1)\,\right|} \ge \frac{\hbar}{2} \\
\equiv\; & h_G\left(\mathcal{B}_{G_{-\mathcal{C}}}(p,r-1)\right) - \frac{|\mathcal{C}|}{\left|\,\mathcal{B}_{G_{-\mathcal{C}}}(p,r-1)\,\right|} \ge \frac{\hbar}{2} \\
\Leftarrow\; & \hbar - \frac{|\mathcal{C}|}{\left|\,\mathcal{B}_{G_{-\mathcal{C}}}(p,r-1)\,\right|} \ge \frac{\hbar}{2}, \quad \text{since } h_G\left(\mathcal{B}_{G_{-\mathcal{C}}}(p,r-1)\right) \ge \hbar \text{ by definition of } \hbar \\
\equiv\; & \frac{|\mathcal{C}|}{\hbar/2} \le \left|\,\mathcal{B}_{G_{-\mathcal{C}}}(p,r-1)\,\right| \\
\Leftarrow\; & \frac{|\mathcal{C}|}{\hbar/2} \le (\hbar+1)^{\frac{D}{3}-\alpha D}, \quad \text{since } \left|\,\mathcal{B}_{G_{-\mathcal{C}}}(p,r-1)\,\right| \ge \left|\mathcal{B}_{G_{-\mathcal{C}}}\left(p, \frac{D}{3}-\alpha D\right)\right| \\
& \qquad\qquad = \left|\mathcal{B}_{G}\left(p, \frac{D}{3}-\alpha D\right)\right| \ge (\hbar+1)^{\frac{D}{3}-\alpha D} \\
\Leftarrow\; & \left(\frac{D}{3}d^{\alpha D}\right)\left(\frac{2}{\hbar}\right) \le (\hbar+1)^{\frac{D}{3}-\alpha D}, \quad \text{since } |\mathcal{C}| < \frac{D}{3}d^{\alpha D} \text{ by (2)}
\end{aligned}$$

$$
\begin{aligned}
&\equiv \quad \frac{D}{3} - \alpha D \geq \frac{\log D + \alpha\, D \log d - \log 3 + \log\left(\frac{2}{\hbar}\right)}{\log(1+\hbar)} \\
&\equiv \quad \alpha \leq \frac{\frac{D}{3}\log(1+\hbar) - \log D + \log 3 - \log\left(\frac{2}{\hbar}\right)}{D\log\big(d\,(1+\hbar)\big)} \\
&\Leftarrow \quad \alpha \leq \frac{\frac{D}{3}\log(1+\hbar) - \log D + \log 3 - \log\left(\frac{2}{\hbar}\right)}{D\log(2d)}, \quad \text{since} \quad \hbar \leq 1
\end{aligned} \tag{9}
$$

Since $\hbar > \left(\frac{1}{D}\right)^{1-\mu}$ and n (and hence D) is sufficiently large, we have:

$$
\begin{aligned}
&\frac{D}{3}\log(1+\hbar) - \log D + \log 3 - \log\left(\frac{2}{\hbar}\right) \\
&> \frac{D}{3}\log(1+\hbar) - \log D + \log \hbar, \\
&> \frac{D\hbar\,\log \mathbf{e}}{6} - \log D + \log\hbar, \quad \text{since } \log(\hbar+1) = (\log\mathbf{e})\,(\ln(\hbar+1)) \text{ and } \ln(\hbar+1) \geq \tfrac{\hbar}{1+\hbar} \geq \tfrac{\hbar}{2} \\
&> \frac{D^{\mu}\,\log\mathbf{e}}{6} - 2\log D, \quad \text{since } \hbar > {}^{1}/_{D^{1-\mu}} \\
&> \frac{D^{\mu}}{7}, \quad \text{since } D \text{ is sufficiently large}
\end{aligned}
$$

Thus, Inequality (9) is satisfied by our selection of $\alpha = \dfrac{1}{7\,D^{1-\mu}\,\log(2d)}$ in (1).

For $\frac{D}{3} - \alpha D < r \leq \frac{\mathsf{dist}_{G_{-\mathcal{C}}}(p,q)}{2}$, the number of nodes in the ball $\mathcal{B}_{G_{-\mathcal{C}}}(p,r)$ is given by:

$$
\left|\,\mathcal{B}_{G_{-\mathcal{C}}}(p,r)\,\right| \geq (\hbar+1)^{\frac{D}{3}+\alpha\,D}\left(\frac{\hbar}{2}+1\right)^{r-\frac{D}{3}-\alpha\,D} > \left(\frac{\hbar}{2}+1\right)^{r}
$$

Thus, using (3) and (5), we get

$$
\begin{aligned}
&\frac{n}{2} > \left|\,\mathcal{B}_{G_{-\mathcal{C}}}\left(p, \frac{\mathsf{dist}_{G_{-\mathcal{C}}}(p,q)}{2}\right)\right| > \left(\frac{\hbar}{2}+1\right)^{d_{G_{-\mathcal{C}}}(p,q)/2} \geq \left(\frac{\hbar}{2}+1\right)^{\frac{D}{120}\,2^{\,\alpha D/\delta}} \\
\Rightarrow\quad &\frac{n}{2} > \left(\frac{\hbar}{2}+1\right)^{\frac{D}{120}\,2^{\,\alpha D/\delta}} = \left(\frac{\hbar}{2}+1\right)^{\frac{D}{120}\,2^{\,D^{\mu}/(7\,\delta\,\log(2d)\,)}} \\
\equiv\quad &\log\left(\frac{\hbar}{2}+1\right) < \frac{120(\log n - 1)}{D\,2^{\,D^{\mu}/(7\,\delta\,\log(2d)\,)}} < \frac{120\log n}{D\,2^{\,D^{\mu}/(7\,\delta\,\log(2d)\,)}} \\
\Rightarrow\quad &\hbar < \frac{360\log n}{D\,2^{\,D^{\mu}/(7\,\delta\,\log(2d)\,)}}, \quad \text{using } \frac{\hbar}{3} \leq \frac{\frac{\hbar}{2}}{1+\frac{\hbar}{2}} = 1 - \frac{1}{1+\frac{\hbar}{2}} \leq \ln\left(\frac{\hbar}{2}+1\right)
\end{aligned}
$$

Thus, there must exist a subset of nodes $\mathcal{B}_{G_{-\mathcal{C}}}(p,r')$ for some $0 \leq r' \leq \mathsf{dist}_{G_{-\mathcal{C}}}(p,q)/2$ such that $h_G\left(\mathcal{B}_{G_{-\mathcal{C}}}(p,r')\right) \leq \max\left\{\left(\frac{1}{D}\right)^{1-\mu}, \frac{360\log n}{D\,2^{\,D^{\mu}/(7\,\delta\,\log(2d)\,)}}\right\}$. Algorithmically, such a subset of nodes can be found as follows:

- Find two nodes p and q such that $\mathsf{dist}_G(p,q) = D$ in $O\left(n^2\log n + mn\right)$ time.
- Using breadth-first-search (BFS), find the two nodes p', q' as in the proof in $O(m+n)$ time.
- Compute α and using BFS find the set of nodes $\mathcal{C}$ in $O\left(n^2 + mn\right)$ time.

- Compute $G_{-\mathcal{C}}$ in $O(m+n)$ time.
- Using BFS, compute $\mathcal{B}_{G_{-\mathcal{C}}}(p,r)$ for every $0 \leq r \leq \mathsf{dist}_{G_{-\mathcal{C}}}(p,q)/2$ in $O(m+n)$ time.
- Compute $h_G\left(\mathcal{B}_{G_{-\mathcal{C}}}(p,r)\right)$ for every $0 \leq r \leq \mathsf{dist}_{G_{-\mathcal{C}}}(p,q)/2$ in an obvious manner in $O(n^2+mn)$ time, and select the subset of node with a minimum expansion.

❑

3.2 Nested Family of Witnesses of Small Node Expansion

In previous section we were able to find one subset of nodes with a small node expansion ratio. In this section, our goal is to find *many* such subsets of nodes that form a nested family.

Let p and q be two nodes of G such that $\mathsf{dist}_G(p,q) = D(G)$. Recall that a cut S of $G=(V,E)$ that "separates p from q" is a subset S of nodes containing p but not containing q, and the set of cut edges $\text{cut}_G(S,p,q)$ corresponding to the cut S is the set of edges with exactly one end-point in S, *i.e.*,

$$\text{cut}_G(S,p,q) = \Big\{ \{u,v\} \mid p,u \in S \text{ and } q,v \in V \setminus S \Big\}$$

Our results in this section show that there exists a nested family of cuts $S_1 \subset S_2 \subset \dots$ of small node expansion for a hyperbolic graph G, and all of these cuts either contain p or all of them contain q.

Theorem 5 *Let p and q be two nodes of G such that $\mathsf{dist}_G(p,q) = D$. Then, for any constant $0 < \mu < 1$, there exists $t = \max\left\{\frac{D^\mu}{56\log d}, 1\right\}$ distinct subsets of nodes $\emptyset \subset S_1 \subset S_2 \subset \cdots \subset S_t \subset V$ such that the following properties hold:*

- $\forall\, j \in \{1,2,\dots,t\}:\ h_G(S_j) \leq \min\left\{ \frac{8\ln\left(\frac{n}{2}\right)}{D}, \max\left\{ \left(\frac{1}{D}\right)^{1-\mu}, \frac{500\ln n}{D\, 2^{\frac{D^\mu}{28\,\delta\,\log(2d)}}} \right\} \right\}.$
- *Each subset S_j has at most $n/2$ nodes, and can be found in $O\left(n^2\log n + mn\right)$ time.*
- *Either all the subsets $S_1, S_2, \dots, S_t$ contain the node p, or all of them contain the node q.*

Remark 2 *The expansion bounds in Theorem 5 are of the same asymptotic nature as those in Theorem 3. Thus, these bounds also follow the interpretation in Remark 1.*

Remark 3 *The claim in Theorem 5 actually holds for any pair of nodes u and v provided one replaced D in the claim by $\mathsf{dist}_G(u,v)$.*

Proof. The bound of $h_G(S_j) < \frac{8\ln\left(\frac{n}{2}\right)}{D}$ again holds for any graph. Let p and q be two nodes such that $\mathsf{dist}_G(p,q) = D$. Assume, without loss of generality, that $\left|\mathcal{B}_G(p,\frac{D}{2})\right| \leq \min\left\{ \left|\mathcal{B}_G(p,\frac{D}{2})\right|, \left|\mathcal{B}_G(q,\frac{D}{2})\right| \right\} \leq n/2$. Consider the sequence of balls $\mathcal{B}_G(p,r)$ for $r = 0,1,2\dots,D/2$. Thus it follows that

$$\frac{n}{2} > \left|\mathcal{B}_G\left(p,\frac{D}{2}\right)\right| \geq \prod_{\ell=0}^{\frac{D}{2}-1}\Big(1+h_G\left(\mathcal{B}_G(p,\ell)\right)\Big) \geq \prod_{\ell=0}^{\frac{D}{2}-1} \mathbf{e}^{h_G(\mathcal{B}_G(p,\ell))/2} = \mathbf{e}^{\frac{1}{2}\sum_{\ell=0}^{\frac{D}{2}-1} h_G(\mathcal{B}_G(p,\ell))}$$

$$\Rightarrow \ln\left(\frac{n}{2}\right) > \frac{1}{2}\sum_{\ell=0}^{\frac{D}{2}-1} h_G\left(\mathcal{B}_G(p,\ell)\right) \Rightarrow \frac{\sum_{\ell=0}^{\frac{D}{2}-1} h_G\left(\mathcal{B}_G(p,\ell)\right)}{D/2} < \frac{4\ln\left(\frac{n}{2}\right)}{D}$$

By a simple averaging argument, there must now exist $\frac{D}{4} > \max\left\{\frac{D^\mu}{56\log d}, 1\right\}$ distinct balls (subsets of nodes) $\mathcal{B}_G(p, r_1), \mathcal{B}_G(p, r_2), \ldots, \mathcal{B}_G\left(p, r_{D/4}\right)$ such that $|\,\mathcal{B}_G(p, r_j)\,| < \frac{8\ln\left(\frac{n}{2}\right)}{D}$ for $j = 1, 2, \ldots, D/4$, and these balls can be easily found within the desired time complexity bound.

Thus, in the remaining part of the proof, we concentrate on the other two bounds only. We will reuse the same construction as in the proof of Theorem 3 with appropriately modified calculations. If $D(n) = c$ for any some constant $c \geq 1$ (independent of n) then, since $\delta \geq 1/2$ and $d > 1$, we have

$$\frac{500\ln n}{D\,2^{\frac{D^\mu}{28\,\delta\,\log(2d)}}} > \frac{500\ln n}{D\,2^{(1/14)D^\mu}} > 1 \quad \text{(since } n \text{ is sufficiently large)}$$

Thus, any subset of $n/2$ nodes containing p satisfies the claimed bound, and the number of such subsets is $\binom{\frac{n}{2}-1}{n-2} > t$.

Otherwise, assume that $D(n) = \omega(1)$, *i.e.*, $\lim_{n\to\infty} D(n) > c$ for any constant c. To provide some intuition behind our proof, let $\overline{\boldsymbol{h}}(p, j) \stackrel{\text{def}}{=} \frac{\sum_{\ell=0}^{j-1}\mathcal{B}_G(p,\ell)}{j}$. Equation (4) regarding the size of $|\,\mathcal{B}_G(p, r)\,|$ now becomes:

$$|\,\mathcal{B}_G(p,0)\,| = 1 \;\text{ and }\; \frac{|\mathcal{B}_G(p,r)|}{|\mathcal{B}_G(p,r-1)|} = 1 + h_G\left(\mathcal{B}_G(p, r-1)\right)$$

$$\Rightarrow\; |\,\mathcal{B}_G(p,r)\,| = \prod_{j=0}^{r-1}\left(1 + h_G\left(\mathcal{B}_G(p,j)\right)\right) \geq \prod_{j=0}^{r-1} \mathbf{e}^{h_G(\mathcal{B}_G(p,j))/2} = \mathbf{e}^{\frac{1}{2}\sum_{j=0}^{r-1} h_G(\mathcal{B}_G(p,j))} = \mathbf{e}^{r\,\overline{\boldsymbol{h}}(p,r)/2} \quad (10)$$

Note that the value $(1+\mathfrak{h}_G)^r = \mathbf{e}^{r\ln(1+\mathfrak{h}_G)}$ is approximately $\mathbf{e}^{r\,\mathfrak{h}_G/2}$, and thus one would expect the moving average $\overline{\boldsymbol{h}}$ values to play the same role in the analysis as $\mathfrak{h}_G$ *provided the average was taken over at least* $\frac{D^\mu}{56\log d}$ *distinct sets*. In the remaining analysis, we will show that this can be done.

First, suppose that there exists a set of $t = \frac{D^\mu}{56\log d}$ distinct indices $\{i_1, i_2, \ldots, i_t\} \subseteq \left\{0, 1, 2, \ldots, \frac{\mathsf{dist}_{G_{-\mathcal{C}}}(p,q)}{2}\right\}$ such that:

$$\forall\, 1 \leq s \leq t:\; h_G\left(\,\mathcal{B}_G(p, i_s)\,\right) = h_G\left(\,\mathcal{B}_{G_{-\mathcal{C}}}(p, i_s)\,\right) \leq (1/D)^{1-\mu}$$

Then, our claim is obviously true and these subsets can be found in the same manner as in Theorem 3. Otherwise, we have

$$\sum_{\ell=0}^{\frac{D}{3}-\alpha D-1} h_G\left(\mathcal{B}_G(p,\ell)\right) > \left(\frac{\mathsf{dist}_{G_{-\mathcal{C}}}(p,q)}{2} - (t-1)\right)(1/D)^{1-\mu} > \left(\frac{D}{3} - \alpha D - t\right)(1/D)^{1-\mu} > \frac{D^\mu}{4} \quad (11)$$

Handling the case(s) corresponding to $r_p \leq \mathsf{dist}_{G_{-\mathcal{C}}}(p,q)/2$ is a bit more involved now. Let us write r_p as $r_{p,\alpha D}$ to show its dependence on αD and let $\alpha_1 = \frac{1}{14\,D^{1-\mu}\,\log(2d)}$ be *asymptotically similar* to choice of α shown in (1). Consider the sequence of values $r_{p,\,\alpha_1 D}, r_{p,\,\alpha_1 D-1}, \ldots, r_{p,\,\alpha_1 D/2}$ as the value of αD is decremented from $\alpha_1 D = \frac{D^\mu}{14\,\log(2d)}$ to $\frac{\alpha_1 D}{2} = \frac{D^\mu}{28\,\log(2d)}$ in steps of -1. Let $\mathcal{C}_{\alpha_1 D-\ell}$ be the set of nodes in $\mathcal{C}$ when αD is set equal to $\alpha_1 D - \ell$.

Proposition 1 *$\mathcal{C}_{\alpha_1 D-j} \neq \mathcal{C}_{\alpha_1 D-j'}$ for any $j \neq j'$.*

Proof. Assume $j < j'$ and let u be the node on a shortest path between p and q that goes through p' and q' such that $d_G(p', u) = \alpha_1 D - j' - 1$. Then u belongs to $\mathcal{C}_{\alpha_1 D-j'}$ but does not belong to $\mathcal{C}_{\alpha_1 D-j}$. ❑

Suppose first that removal of each of the $\frac{\alpha_1 D}{2}+1$ set of nodes $\mathcal{C}_{\alpha_1 D}, \mathcal{C}_{\alpha_1 D-1}, \mathcal{C}_{\alpha_1 D-2}, \dots, \mathcal{C}_{\alpha_1 D/2}$ disconnects p from q in the corresponding graphs $G_{-\mathcal{C}_{\alpha_1 D}}, G_{-\mathcal{C}_{\alpha_1 D-1}}, G_{-\mathcal{C}_{\alpha_1 D-2}}, \dots, G_{-\mathcal{C}_{\alpha_1 D/2}}$, respectively. Then, for any $0 \leq \ell \leq \alpha_1 D/2$, we have

$$r_{p,\alpha_1 D-\ell} = \alpha_1 D - \ell - 1 \geq \frac{D}{3} - \alpha_1 D + \ell \geq \frac{D}{3} - \alpha_1 D$$

$$\left| \mathcal{B}_{G_{-\mathcal{C}_{\alpha_1 D-\ell}}}\left(p, r_{p,\alpha_1 D-\ell}\right) \right| > \left| \mathcal{B}_{G_{-\mathcal{C}_{\alpha_1 D}}}\left(p, \frac{D}{3} - \alpha_1 D\right) \right| \underset{\text{by (10)}}{\geq} \mathbf{e}^{\frac{1}{2}\sum_{j=0}^{\frac{D}{3}-\alpha D-1} h_G(\mathcal{B}_G(p,j))} \underset{\text{by (11)}}{>} \mathbf{e}^{D^\mu/8}$$

$$\left| \partial_{G_{-\mathcal{C}_{\alpha_1 D-\ell}}}\left(\mathcal{B}_{G_{-\mathcal{C}_{\alpha_1 D-\ell}}}\left(p, r_{p,\alpha_1 D-\ell}\right)\right) \right| = 0 \;\Rightarrow\; \left| \partial_G\left(\mathcal{B}_{G_{-\mathcal{C}_{\alpha_1 D-\ell}}}\left(p, r_{p,\alpha_1 D-\ell}\right)\right) \right| \leq \left| \mathcal{C}_{\alpha_1 D-\ell} \right| \leq \left| \mathcal{C}_{\alpha_1 D} \right| < \frac{D}{3} d^{\alpha_1 D}$$

$$h_G\left(\mathcal{B}_{G_{-\mathcal{C}_{\alpha_1 D-\ell}}}\left(p, r_{p,\alpha_1 D-\ell}\right)\right) = \frac{\left| \partial_G\left(\mathcal{B}_{G_{-\mathcal{C}_{\alpha_1 D-\ell}}}\left(p, r_{p,\alpha_1 D-\ell}\right)\right) \right|}{\left| \mathcal{B}_{G_{-\mathcal{C}_{\alpha_1 D-\ell}}}\left(p, r_{p,\alpha_1 D-\ell}\right) \right|} < \frac{\frac{D}{3} d^{\alpha_1 D}}{\mathbf{e}^{D^\mu/8}}$$

$$d^{\alpha_1 D} = d^{\frac{D^\mu}{14\log(2d)}} < \left(d^{1/\log d} \right)^{D^\mu/14} = 2^{D^\mu/14} < \mathbf{e}^{D^\mu/14}$$

$$h_G\left(\mathcal{B}_{G_{-\mathcal{C}_{\alpha_1 D-\ell}}}\left(p, r_{p,\alpha_1 D-\ell}\right)\right) < \frac{\frac{D}{3} d^{\alpha_1 D}}{\mathbf{e}^{D^\mu/8}} < \frac{\frac{D}{3}\mathbf{e}^{D^\mu/14}}{\mathbf{e}^{D^\mu/8}} < \frac{D/3}{2^{D^\mu/20}} < \left(\frac{1}{D}\right)^{1-\mu}, \quad \text{since } \mu > 0 \text{ and } n \text{ (and hence } D\text{) is sufficiently large}$$

The last inequality implies that there exists a set of $\frac{\alpha_1 D}{2}+1 = \frac{D^\mu}{28\log(2d)}+1 > \frac{D^\mu}{56\log d}$ subsets of nodes $\mathcal{B}_{G_{-\mathcal{C}_{\alpha_1 D}}}\left(p, r_{p,\alpha_1 D}\right), \mathcal{B}_{G_{-\mathcal{C}_{\alpha_1 D-1}}}\left(p, r_{p,\alpha_1 D-1}\right), \mathcal{B}_{G_{-\mathcal{C}_{\alpha_1 D-2}}}\left(p, r_{p,\alpha_1 D-2}\right), \dots, \mathcal{B}_{G_{-\mathcal{C}_{\alpha_1 D/2}}}\left(p, r_{p,\alpha_1 D/2}\right)$ such that each such subset $\mathcal{B}_{G_{-\mathcal{C}_{\alpha_1 D-\ell}}}\left(p, r_{p,\alpha_1 D-\ell}\right)$ has $h_G\left(\mathcal{B}_{G_{-\mathcal{C}_{\alpha_1 D-\ell}}}\left(p, r_{p,\alpha_1 D-\ell}\right)\right) < (1/D)^{1-\mu}$. This proves our claim. Algorithmically, these subsets can be found in the same manner as in Theorem 3.

Otherwise, there exists an index $0 \leq t \leq \alpha_1 D/2$ such that the removal of the set of nodes in $\mathcal{C}_{\alpha_1 D-t}$ does *not* disconnect p from q in the corresponding graphs $G_{-\mathcal{C}_{\alpha_1 D-t}}$. This implies $r_{p,\alpha_1 D-t} > \mathsf{dist}_{G_{-\mathcal{C}_{\alpha_1 D-t}}}(p,q)/2$. For notational convenience, we will denote $\mathcal{C}_{\alpha_1 D-t}$ and $G_{-\mathcal{C}_{\alpha_1 D-t}}$ simply by $\mathcal{C}$ and $G_{-\mathcal{C}}$, respectively, and let $\alpha_0 = \alpha_1 - \frac{t}{D}$ such that $\alpha_1 D - t = \alpha_0 D$. Note that $\alpha_1/2 \leq \alpha_0 \leq \alpha_1$.

As in the proof of Theorem 3, we will require that removal of nodes in $\mathcal{C}$ decreases the expansions of the balls around p by no more than a factor of 2. As we will verify next, steps in the proof of Theorem 3 actually show the following result:

$$\forall \frac{D}{3} - \alpha_0 D < r \leq \mathsf{dist}_{G_{-\mathcal{C}}(p,q)}/2 \; :$$

$$h_G\left(\mathcal{B}_{G_{-\mathcal{C}}}(p, r-1)\right) > (1/D)^{1-\mu} \;\Rightarrow\; h_{G_{-\mathcal{C}}}\left(\mathcal{B}_{G_{-\mathcal{C}}}(p, r-1)\right) \geq \frac{1}{2} h_G\left(\mathcal{B}_{G_{-\mathcal{C}}}(p, r-1)\right) \quad (12)$$

The verification of (12) as shown below follow the same steps as in the proof of Theorem 3 with minor modifications:

$$\begin{aligned}
& h_{G_{-\mathcal{C}}}\left(\mathcal{B}_{G_{-\mathcal{C}}}(p, r-1)\right) \geq \frac{1}{2} h_G\left(\mathcal{B}_{G_{-\mathcal{C}}}(p, r-1)\right) \\
\equiv \quad & \frac{\left| \partial_G\left(\mathcal{B}_{G_{-\mathcal{C}}}(p, r-1)\right) \right| - \left| \partial_G\left(\mathcal{B}_{G_{-\mathcal{C}}}(p, r-1)\right) \cap \mathcal{C} \right|}{\left| \mathcal{B}_{G_{-\mathcal{C}}}(p, r-1) \right|} \geq \frac{1}{2} h_G\left(\mathcal{B}_{G_{-\mathcal{C}}}(p, r-1)\right) \\
\Leftarrow \quad & \frac{\left| \partial_G\left(\mathcal{B}_{G_{-\mathcal{C}}}(p, r-1)\right) \right| - |\mathcal{C}|}{\left| \mathcal{B}_{G_{-\mathcal{C}}}(p, r-1) \right|} \geq \frac{1}{2} h_G\left(\mathcal{B}_{G_{-\mathcal{C}}}(p, r-1)\right) \\
\equiv \quad & \frac{\left| \partial_G\left(\mathcal{B}_{G_{-\mathcal{C}}}(p, r-1)\right) \right|}{\left| \mathcal{B}_{G_{-\mathcal{C}}}(p, r-1) \right|} - \frac{|\mathcal{C}|}{\left| \mathcal{B}_{G_{-\mathcal{C}}}(p, r-1) \right|} \geq \frac{1}{2} h_G\left(\mathcal{B}_{G_{-\mathcal{C}}}(p, r-1)\right)
\end{aligned}$$

$$
\begin{aligned}
&\equiv \quad h_G\left(\mathcal{B}_{G_{-\mathcal{C}}}(p,r-1)\right) - \frac{|\mathcal{C}|}{\left|\,\mathcal{B}_{G_{-\mathcal{C}}}(p,r-1)\,\right|} \geq \frac{1}{2}\, h_G\left(\mathcal{B}_{G_{-\mathcal{C}}}(p,r-1)\right) \\
&\equiv \quad \frac{2\,|\mathcal{C}|}{h_G\left(\mathcal{B}_{G_{-\mathcal{C}}}(p,r-1)\right)} \leq \left|\,\mathcal{B}_{G_{-\mathcal{C}}}(p,r-1)\,\right| \\
&\Leftarrow \quad \frac{2\,|\mathcal{C}|}{h_G\left(\mathcal{B}_{G_{-\mathcal{C}}}(p,r-1)\right)} \leq \mathbf{e}^{D^{\mu}/8}, \quad \text{since} \quad \left|\,\mathcal{B}_{G_{-\mathcal{C}}}(p,r-1)\,\right| \geq \left|\mathcal{B}_{G_{-\mathcal{C}}}\left(p, \frac{D}{3}-\alpha_0 D\right)\right| \\
&\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad = \left|\mathcal{B}_{G}\left(p, \frac{D}{3}-\alpha_0 D\right)\right| \geq \left|\mathcal{B}_{G}\left(p, \frac{D}{3}-\alpha_1 D\right)\right| > \mathbf{e}^{D^{\mu}/8} \\
&\Leftarrow \quad \left(\frac{D}{3} d^{\alpha_0 D}\right)\left(\frac{2}{h_G\left(\mathcal{B}_{G_{-\mathcal{C}}}(p,r-1)\right)}\right) \leq \mathbf{e}^{D^{\mu}/8}, \quad \text{since} \quad |\mathcal{C}| < \frac{D}{3} d^{\alpha_0 D} \\
&\equiv \quad \frac{D^{\mu}}{8} \geq \ln D + \alpha_0\, D \ln d - \ln {}^{3}\!/_{2} - \ln\left(h_G\left(\mathcal{B}_{G_{-\mathcal{C}}}(p,r-1)\right)\right) \\
&\Leftarrow \quad \frac{D^{\mu}}{8} \geq \ln D + \alpha_1\, D \ln d - \ln {}^{3}\!/_{2} - \ln\left(h_G\left(\mathcal{B}_{G_{-\mathcal{C}}}(p,r-1)\right)\right), \quad \text{since} \quad \alpha_0 \leq \alpha_1 \\
&\equiv \quad \alpha_1 \leq \frac{\frac{D^{\mu}}{8} - \ln D + \ln {}^{3}\!/_{2} + \ln\left(h_G\left(\mathcal{B}_{G_{-\mathcal{C}}}(p,r-1)\right)\right)}{D \ln d} \\
&\Leftarrow \quad \alpha_1 \leq \frac{\frac{D^{\mu}}{8} - \ln D + \ln\left(h_G\left(\mathcal{B}_{G_{-\mathcal{C}}}(p,r-1)\right)\right)}{D \ln d} \qquad (13)
\end{aligned}
$$

If $h_G\left(\mathcal{B}_{G_{-\mathcal{C}}}(p,r-1)\right) > ({}^{1}\!/_{D})^{1-\mu}$ then since n (and hence D) is sufficiently large, we have:

$$
\frac{D^{\mu}}{8} - \ln D + \ln\left(h_G\left(\mathcal{B}_{G_{-\mathcal{C}}}(p,r-1)\right)\right) > \frac{D^{\mu}}{8} - \ln D - (1-\mu)\ln D \underbrace{>}_{\text{since } D \text{ is sufficiently large}} \frac{D^{\mu}}{7}
$$

Thus, Inequality (13) is satisfied by our selection of $\alpha_1 = \dfrac{1}{14\, D^{1-\mu}\, \log(2d)}$. This concludes the verification of (12).

Suppose now that there exists a set of $t = \frac{D^{\mu}}{56 \log d}$ distinct indices $i_1, i_2, \ldots, i_t \in \left\{ \frac{D}{3} - \alpha_0 D + 1, \frac{D}{3} - \alpha_0 D + 2, \ldots, \frac{\mathsf{dist}_{G_{-\mathcal{C}}}(p,q)}{2} \right\}$ such that

$$
\forall\, 1 \leq s \leq t: \; h_G\left(\mathcal{B}_{G_{-\mathcal{C}}}(p,i_s)\right) \leq ({}^{1}\!/_{D})^{1-\mu} \qquad (14)
$$

Then, the existence of the subsets of nodes $\mathcal{B}_{G_{-\mathcal{C}}}(p,i_1), \mathcal{B}_{G_{-\mathcal{C}}}(p,i_2), \ldots, \mathcal{B}_{G_{-\mathcal{C}}}(p,i_t)$ prove our claim. Algorithmically, these subsets can be found in the same manner as in Theorem 3. Otherwise, assume that there are no sets of t indices that satisfy (14). This implies that there exists a set of

$$
\xi = \left(\frac{\mathsf{dist}_{G_{-\mathcal{C}}}(p,q)}{2} - \frac{D}{3} + \alpha_0 D\right) - (t-1)
$$

distinct indices $j_1, j_2, \ldots, j_{\xi} \in \left\{ \frac{D}{3} - \alpha_0 D + 1, \frac{D}{3} - \alpha_0 D + 2, \ldots, \frac{\mathsf{dist}_{G_{-\mathcal{C}}}(p,q)}{2} \right\}$ such that

$$
\forall\, 1 \leq s \leq \xi: \; h_G\left(\mathcal{B}_{G_{-\mathcal{C}}}(p,j_s)\right) > ({}^{1}\!/_{D})^{1-\mu} \underbrace{\Rightarrow}_{\text{by (12)}} \forall\, 1 \leq s \leq \xi: \; h_{G_{-\mathcal{C}}}\left(\mathcal{B}_{G_{-\mathcal{C}}}(p,j_s)\right) \geq \frac{1}{2}\, h_G\left(\mathcal{B}_{G_{-\mathcal{C}}}(p,j_s)\right) \qquad (15)
$$

This implies

$$
\begin{aligned}
\left|\mathcal{B}_{G_{-\mathcal{C}}}\left(p, \frac{\mathsf{dist}_{G_{-\mathcal{C}}}(p,q)}{2}\right)\right| &> \left(\prod_{j=0}^{\frac{D}{3}-\alpha_0 D-1}\left(1+h_G\left(\mathcal{B}_{G_{-\mathcal{C}}}(p,j)\right)\right)\right)\left(\prod_{j=\frac{D}{3}-\alpha_0 D}^{\frac{D}{3}-\alpha_0 D+\xi-1}\left(1+\frac{1}{2}h_G\left(\mathcal{B}_{G_{-\mathcal{C}}}(p,j)\right)\right)\right) \\
&\qquad \text{using (15)} \\
&> \left(\prod_{j=0}^{\frac{D}{3}-\alpha_0 D-1}\mathbf{e}^{h_G\left(\mathcal{B}_{G_{-\mathcal{C}}}(p,j)\right)/2}\right)\left(\prod_{j=\frac{D}{3}-\alpha_0 D}^{\frac{D}{3}-\alpha_0 D+\xi-1}\mathbf{e}^{h_G\left(\mathcal{B}_{G_{-\mathcal{C}}}(p,j)\right)/4}\right) \\
&= \left(\mathbf{e}^{\frac{1}{2}\sum_{j=0}^{\frac{D}{3}-\alpha_0 D-1}h_G\left(\mathcal{B}_{G_{-\mathcal{C}}}(p,j)\right)}\right)\left(\mathbf{e}^{\frac{1}{4}\sum_{j=\frac{D}{3}-\alpha_0 D}^{\frac{D}{3}-\alpha_0 D+\xi-1}h_G\left(\mathcal{B}_{G_{-\mathcal{C}}}(p,j)\right)}\right) \\
&> \mathbf{e}^{\frac{1}{4}\sum_{j=0}^{\frac{D}{3}-\alpha_0 D+\xi-1}h_G\left(\mathcal{B}_{G_{-\mathcal{C}}}(p,j)\right)} \qquad (16)
\end{aligned}
$$

$$
\frac{n}{2} > \left|\mathcal{B}_{G_{-\mathcal{C}}}\left(p, \frac{\mathsf{dist}_{G_{-\mathcal{C}}}(p,q)}{2}\right)\right| > \mathbf{e}^{\frac{1}{4}\sum_{j=0}^{\frac{D}{3}-\alpha_0 D+\xi-1}h_G\left(\mathcal{B}_{G_{-\mathcal{C}}}(p,j)\right)} \quad \Rightarrow \sum_{j=0}^{\frac{D}{3}-\alpha_0 D+\xi-1} h_G\left(\mathcal{B}_{G_{-\mathcal{C}}}(p,j)\right) < 4\ln n \qquad (17)
$$

Suppose that there exists *no* set of $t = \frac{D^\mu}{56\log d}$ distinct indices $i_1, i_2, \ldots, i_t \in \left\{0, 1, \ldots, \frac{D}{3} - \alpha_0 D + \xi - 1\right\}$ such that

$$
\forall\, 1 \le s \le t: \;\; h_G\left(\mathcal{B}_{G_{-\mathcal{C}}}(p, i_s)\right) \le \frac{500\ln n}{D\, 2^{\frac{D^\mu}{28\delta\log(2d)}}} \qquad (18)
$$

Together with (17) this implies:

$$
\begin{aligned}
& 4\ln n > \sum_{j=0}^{\frac{D}{3}-\alpha_0 D+\xi-1} h_G\left(\mathcal{B}_{G_{-\mathcal{C}}}(p,j)\right) > \left(\frac{D}{3} - \alpha_0 D + \xi - \frac{D^\mu}{56\log d} + 1\right)\frac{500\ln n}{D\, 2^{\frac{D^\mu}{28\delta\log(2d)}}} \\
\Rightarrow\;& \left(\frac{\mathsf{dist}_{G_{-\mathcal{C}}}(p,q)}{2} - \frac{D^\mu}{28\log d}\right)\frac{500\ln n}{D\, 2^{\frac{D^\mu}{28\delta\log(2d)}}} < 4\ln n \\
\Rightarrow\;& \left(\frac{D}{120}2^{\frac{\alpha_1 D}{2\delta}} - \frac{D^\mu}{28\log d}\right)\frac{500\ln n}{D\, 2^{\frac{D^\mu}{28\delta\log(2d)}}} < 4\ln n, \quad \text{by (3) and since } \tfrac{\alpha_1}{2} \le \alpha_0 \\
\equiv\;& D2^{\frac{D^\mu}{28\delta\log(2d)}} < \frac{3000 D^\mu}{28\log d} \qquad (19)
\end{aligned}
$$

Inequality (19) is false since $2^{\frac{D^\mu}{28\delta\log(2d)}} > 1$, $\mu < 1$ and D is sufficiently large. Thus, there must exist a set of t distinct indices $i_1, i_2, \ldots, i_t$ such that (18) holds and the corresponding sets $\mathcal{B}_{G_{-\mathcal{C}}}(p, i_1), \mathcal{B}_{G_{-\mathcal{C}}}(p, i_2), \ldots, \mathcal{B}_G$ prove our claim. Algorithmically, these subsets again can be found in the same manner as in Theorem 3. ❑

3.3 Family of Small Node Expansion Witnesses With Limited Mutual Overlaps

The result in the previous section provided a nested family of cuts that separated node p from node q. However, pairs of subsets in this family may differ by as few as just *one* node. In some applications, one may need to generate a family of cuts that are sufficiently different from each other, *i.e.*, they are either disjoint or have limited overlap. This can be done as stated below.

Theorem 6 *For any constant $0 < \mu < 1$ and for any positive integer* $\frac{(42\,\delta\,\log(2d)\,\log(2D))^{1/\mu}}{D} \leq \tau \leq D/4$ *there exists* $\lfloor \tau/4 \rfloor$ *distinct subsets of nodes* $\emptyset \subset V_1, V_2, \ldots, V_{\lfloor \tau/4 \rfloor} \subset V$ *such that the following results hold for* $\langle G, d, \delta \rangle$ *with* $D = D(G)$*:*

- $\forall\, j \in \{1, 2, \ldots, \lfloor \tau/4 \rfloor\} \;:\; h_G(V_j) \leq \max \left\{ \left(\frac{1}{(D/\tau)} \right)^{1-\mu}, \frac{360 \log n}{(D/\tau)\, 2^{\frac{(D/\tau)^{\mu}}{7\,\delta\,\log(2d)}}} \right\}.$
- *Each subset V_j has at most $n/2$ nodes, and can be found in $O\left(n^2 \log n + mn\right)$ time.*
- (limited overlap claim) *For every distinct pair of subsets V_i and V_j, either $V_i \cap V_j = \emptyset$ or at least $\frac{D}{2\tau}$ nodes in each subset do not belong to the other subset.*

Remark 4 *Consider a bounded-degree hyperbolic graph,* i.e., *assume that δ and d are constants. Setting $\tau = \sqrt{D}$ gives $\Omega(\sqrt{D}\,)$ subsets of nodes of maximum node expansion $(1/D)^{\frac{1-\mu}{2}}$ such that every pairwise non-disjoint subsets have $\Omega(\sqrt{D}\,)$ private nodes.*

Proof. Let p and q be two nodes such that $\mathsf{dist}_G(p, q) = D(G) = D$. Let $(p = p_1, p_2, \ldots, p_{\tau+1} = q)$ be an ordered sequence of $\tau + 1$ nodes such that $\mathsf{dist}_G(p_i, p_{i+1}) = \frac{D}{\tau}$ for $i = 1, 2, \ldots, \tau$. Applying the proof of Theorem 3 and using Remark 3 for each pair (p_i, p_{i+1}), we get a subset $\emptyset \subset S_i \subset V$ of nodes such that

$$h_G(S_i) \leq \max \left\{ \left(\frac{1}{(D/\tau)} \right)^{1-\mu}, \frac{360 \log n}{(D/\tau)\, 2^{\frac{(D/\tau)^{\mu}}{7\,\delta\,\log(2d)}}} \right\}$$

Thus, in all, we have τ such subsets of nodes. The subset S_i of nodes were constructed in Theorem 3 in the following manner. Let ℓ_i and r_i be two nodes on a shortest path $\overline{p_i, p_{i+1}}$ such that $\mathsf{dist}_G(p_i, \ell_i) = \mathsf{dist}_G(\ell_i, r_i) = \mathsf{dist}_G(r_i, p_{i+1}) = \frac{1}{3}\mathsf{dist}_G(p_i, p_{i+1})$. For $\alpha = \frac{1}{7\,(D/\tau)^{1-\mu}\log(2d)} < 1/4$, we constructed the graphs $G_{-\mathcal{C}_i}$ obtained by removing the set of nodes $\mathcal{C}_i$ which are exactly at a distance of $\lceil \alpha\, \mathsf{dist}_G(p_i, p_{i+1}) \rceil$ from some node of the shortest path $\overline{\ell_i, r_i}$. The subset S_i is then the set of nodes in the ball $\mathcal{B}_{G_{-\mathcal{C}_i}}(y_i, a_i)$ for some $a_i \in \left[0, \mathsf{dist}_{G_{-\mathcal{C}_i}}(p_i, p_{i+1})/2\right]$ and for some $y_i \in \{p_i, p_{i+1}\}$. If $y_i = p_i$ then we call S_i "left handed", otherwise we call S_i "right handed". Consider the following two collection of subset of nodes:

$$\mathcal{P} = \{\, S_i \mid i \text{ is even} \quad \text{and} \quad S_i \text{ is left handed } \}$$
$$\mathcal{Q} = \{\, S_i \mid i \text{ is even} \quad \text{and} \quad S_i \text{ is right handed } \}$$

One of these collections has at least $\lfloor \tau/4 \rfloor$ sets. Suppose that $\mathcal{P}$ has at least $\lfloor \tau/4 \rfloor$ sets (the other case is similar). We show below that subsets in $\mathcal{P}$ satisfy the *limited overlap* claim. Consider two sets in $\mathcal{P}$ of the form $S_i = \mathcal{B}_{G_{-\mathcal{C}_i}}(p_i, a_i)$ and $S_j = \mathcal{B}_{G_{-\mathcal{C}_j}}(p_j, a_j)$ with $i \leq j - 2$. Let $\mathfrak{C}_i$ denote the interior of the closed cylinder of nodes in G which are at a distance of *at most* $\lceil \alpha\, \mathsf{dist}_G(p_i, p_{i+1}) \rceil$ from some node of the shortest path $\overline{\ell_i, r_i}$, *i.e.*,

$$\mathfrak{C}_i = \left\{ u \mid \exists\, v \in \overline{\ell_i, r_i} \colon \mathsf{dist}_G(u, v) \leq \lceil \alpha\, \mathsf{dist}_G(p_i, p_{i+1}) \rceil \right\}$$

Proposition 2 *If $i \neq j$ then* $\mathfrak{C}_i \cap \mathcal{B}_{G_{-\mathcal{C}_j}}\left(p_j, \frac{D}{2\tau}\right) = \emptyset$.

Proof. Assume for the sake of contradiction that $\mathfrak{C}_i \cap \mathcal{B}_{G_{-\mathcal{C}_j}}\left(p_j, \frac{D}{2\tau}\right) \neq \emptyset$, and let $u \in \mathfrak{C}_i \cap \mathcal{B}_{G_{-\mathcal{C}_j}}\left(p_j, \frac{D}{2\tau}\right)$. Since $u \in \mathfrak{C}_i$, there exists a node $v \in \overline{\ell_i, r_i}$ such that $\mathsf{dist}_G(v, u) \leq \lceil \alpha\, \mathsf{dist}_G(p_i, p_{i+1}) \rceil < \frac{1}{4}\mathsf{dist}_G(p_i, p_{i+1}) = \frac{D}{4\tau}$. Thus,

$$u \in \mathcal{B}_{G_{-\mathcal{C}_j}}\left(p_j, \frac{D}{2\,\tau}\right) \Rightarrow \mathsf{dist}_{G_{-\mathcal{C}_j}}(u, p_j) \le \frac{D}{2\,\tau} \Rightarrow \mathsf{dist}_G(u, p_j) \le \frac{D}{2\,\tau}$$

$$\Rightarrow \mathsf{dist}_G(v, p_j) \le \mathsf{dist}_G(v, u) + \mathsf{dist}_G(u, p_j) < \frac{D}{4\,\tau} + \frac{D}{2\,\tau} < \frac{D}{\tau}$$

which contradicts the fact that $\mathsf{dist}_G(v, p_j) > \mathsf{dist}_G(p_{i+1}, p_j) = \frac{D}{\tau}$. ❑

Proposition 3 *Any node $u \in S_i \cap S_j$ satisfies $\mathsf{dist}_{G_{-\mathcal{C}_j}}(u, p_j) > \frac{D}{2\tau}$.*

Proof. Assume for the sake of contradiction that $z = \mathsf{dist}_{G_{-\mathcal{C}_j}}(u, p_j) \le \frac{D}{2\tau}$. Since $u \in S_i = \mathcal{B}_{G_{-\mathcal{C}_i}}(p_i, a_i)$, this implies $\mathsf{dist}_{G_{-\mathcal{C}_i}}(p_i, u) \le a_i \le \frac{\mathsf{dist}_{G_{-\mathcal{C}_i}}(p_i, p_{i+1})}{2}$. Since $u \in S_j = \mathcal{B}_{G_{-\mathcal{C}_j}}(p_j, a_j)$, this implies $u \in \mathcal{B}_{G_{-\mathcal{C}_j}}(p_j, z)$. By Proposition 2, $\mathfrak{C}_i \cap \mathcal{B}_{G_{-\mathcal{C}_j}}(p_j, z) = \emptyset$, and therefore

$$\frac{D}{2\,\tau} \ge z = \mathsf{dist}_{G_{-\mathcal{C}_j}}(u, p_j) = \mathsf{dist}_{G_{-\mathcal{C}_i \cup \mathcal{C}_j}}(u, p_j) \ge \mathsf{dist}_{G_{-\mathcal{C}_i}}(u, p_j)$$

which in turn implies

$$\mathsf{dist}_{G_{-\mathcal{C}_i}}(p_i, p_j) \le \mathsf{dist}_{G_{-\mathcal{C}_i}}(p_i, u) + \mathsf{dist}_{G_{-\mathcal{C}_i}}(u, p_j) \le \frac{\mathsf{dist}_{G_{-\mathcal{C}_i}}(p_i, p_{i+1})}{2} + \frac{D}{2\,\tau}$$

Since the Hausdorff distance between the two shortest paths $\overline{\ell_i, r_i}$ and $\overline{p_j, p_{j+1}}$ is at least $\frac{D}{3\tau} > \lceil \alpha\, \mathsf{dist}_G(p_i, p_{i+1}) \rceil$, we have

$$\mathsf{dist}_{G_{-\mathcal{C}_i}}(p_j, p_{i+1}) = (j - i)\frac{D}{\tau} < D$$

$$\Rightarrow \mathsf{dist}_{G_{-\mathcal{C}_i}}(p_i, p_{i+1}) \le \mathsf{dist}_{G_{-\mathcal{C}_i}}(p_i, p_j) + \mathsf{dist}_{G_{-\mathcal{C}_i}}(p_j, p_{i+1}) \le \frac{\mathsf{dist}_{G_{-\mathcal{C}_i}}(p_i, p_{i+1})}{2} + \frac{D}{2\,\tau} + D$$

$$\Rightarrow \mathsf{dist}_{G_{-\mathcal{C}_i}}(p_i, p_{i+1}) \le \frac{D}{\tau} + 2D \quad (20)$$

On the other hand, by Fact 1:

$$\mathsf{dist}_{G_{-\mathcal{C}_i}}(p_i, p_{i+1}) \ge \frac{D}{60\,\tau}\, 2^{\frac{\alpha D}{\tau \delta}} = \frac{D}{60\,\tau}\, 2^{\frac{D^\mu \tau^\mu}{7\,\delta\,\log(2d)}} \quad (21)$$

Inequalities (20) and (21) together imply

$$\frac{D}{60\,\tau}\, 2^{\frac{D^\mu \tau^\mu}{7\,\delta\,\log(2d)}} < \frac{D}{\tau} + 2D$$

$$\Rightarrow D^\mu\, \tau^\mu < 7\,\delta\,\log(2d)\,\log(60 + 120\,\tau) < 42\,\delta\,\log(2d)\,\log(2D)$$
since $\tau \le D/2$

$$\tau < \frac{(42\,\delta\,\log(2d)\,\log(2D)\,)^{1/\mu}}{D} \quad (22)$$

Inequality (22) contradicts the assumption on the range of τ. ❑

To complete the proof, suppose that $S_j \cap S_j \ne \emptyset$ and let $u \in S_j \cap S_j \ne \emptyset$. Proposition 3 implies that $S_j \supset \mathcal{B}_{G_{-\mathcal{C}_j}}\left(p_j, \frac{D}{2\tau}\right)$, $u \notin \mathcal{B}_{G_{-\mathcal{C}_j}}\left(p_j, \frac{D}{2\tau}\right)$, and thus there are at least $\frac{D}{2\tau}$ node on a shortest path in $G_{-\mathcal{C}_j}$ from p_j to a node at a distance of $\frac{D}{2\tau}$ from p_j that are not part of S_i. The corresponding proof for S_i with respect to S_j is similar. ❑

4 Large Family of Small-Size Mutually Disjoint Cuts

Recall that, given two distinct nodes $s, t \in V$ of a graph $G = (V, E)$, a cut in G that separates s from t (or, simply a "s-t cut") $\text{cut}_G(S, s, t)$ is a subset of nodes S that disconnects s from t. The *cut-edges* $\mathcal{E}_G(S, s, t)$ (resp., *cut-nodes* $\mathcal{V}_G(S, s, t)$) corresponding to this cut is the set of edges with one end-point in S (resp., the end-points of these cut-edges that belong to S), *i.e.*,

$$\mathcal{E}_G(S, s, t) = \{ \{u, v\} \mid u \in S,\ v \in V \setminus S,\ \{u, v\} \in E \}$$
$$\mathcal{V}_G(S, s, t) = \{ u \mid u \in S,\ v \in V \setminus S,\ \{u, v\} \in E \}$$

Theorem 7 *Suppose that the following holds for our given graph G:*

- *s and t are two nodes of G such that $\mathsf{dist}_G(s, t) > 120\,\delta + 80\,\delta \log n$, and*
- *d is the maximum degree of any node except s and t (degrees of s and t may be arbitrary).*

Then, there exists a set of at least $\left\lfloor \dfrac{\mathsf{dist}_G(s,t) - 8\,\delta \log n}{30 \max\{\delta,\ 1/2\}} \right\rfloor = \Omega(\mathsf{dist}_G(s, t))$ (node and edge) disjoint cuts such that each such cut has at most $d^{12\delta+1}$ cut edges.

Remark 5 *Suppose that G is hyperbolic (i.e., δ is a constant), d is a constant, and s and t be two nodes such that $\mathsf{dist}_G(s, t) > 120\delta + 80\delta \log n = \Omega(\log n)$. Theorem 7 then implies that there are $\Omega(\mathsf{dist}_G(s, t))$ s-t cuts each having $O(1)$ edges.*

Proof. The result is trivially true when $\delta = 0$ (the resulting graph is either a tree or the complete graph K_n), thus we assume in the sequel that $\delta \geq 1/2$. We start by doing a BFS starting from node s. Let $\mathcal{L}_i$ be the sets of nodes at the i^{th} level (*i.e.*, $\forall\, u \in \mathcal{L}_i$: $\mathsf{dist}_G(s, u) = i$); obviously t belongs to $\mathcal{L}_{\mathsf{dist}_G(s,t)}$. To handle arbitrary paths between s and t that may not be shortest or approximately shortest, we recall the following well-known result about the exponential divergence of geodesic rays emanating from a node (*e.g.*, see [1, 9] among others).

Fact 2 (Exponential divergence of geodesic rays) [Simplified reformulation of [1, Theorem 10] in our notations] *Suppose that we are given the following:*

- *three integers $k \geq 4$, $\alpha > 0$, $r > 3k\delta$,*
- *four nodes u_1, u_2, u_3, u_4 such that:*
  - *$u_1, u_2 \in \mathcal{L}_r$ with $\mathsf{dist}_G(u_1, u_2) \geq 3k\delta$,*
  - *$u_3, u_4 \in \mathcal{L}_{r+\alpha}$ with $\mathsf{dist}_G(u_1, u_4) = \mathsf{dist}_G(u_2, u_3) = \alpha$.*

Then, the following statements hold:

(*a*) *For any shortest path $\mathcal{P}$ between u_3 and u_4, the middle node v of $\mathcal{P}$ is contained in $\mathcal{L}_j$ for some $j \leq r + \delta - \frac{3}{2}k\delta$,* i.e., $\mathsf{dist}_G(s, v) \leq r + \delta - \frac{3}{2}k\delta$ *(this obviously implies* $\mathsf{dist}_G(u_3, u_4) \geq 2 \times \left((r + \alpha) - \left(r + \delta - \frac{3}{2}k\delta\right) \right) = 2\,\alpha + (3k - 2)\delta$*).*

(*b*) *Consider any path $\mathcal{Q}$ between u_3 and u_4 that does not involve a node in $\bigcup_{0 \leq j \leq r+\alpha} \mathcal{L}_j$. Then, the length $|\mathcal{Q}|$ of the path $\mathcal{Q}$ satisfies $|\,\mathcal{Q}\,| > 2^{\frac{\alpha}{6\,\delta}+k+1}$.*

An arbitrary path between s and t visits at least one node in each of the levels $\mathcal{L}_1, \mathcal{L}_2, \ldots, \mathcal{L}_{\mathsf{dist}_G(s,t)-1}$ (and, possibly some nodes other than t in levels $\mathcal{L}_{\mathsf{dist}_G(s,t)}$ and beyond).

Lemma 8 *Assume* $\mathsf{dist}_G(s,t) > 13\,\delta + 8\,\delta\,\log n$*, and consider two arbitrary paths* $\mathcal{P}_1$ *and* $\mathcal{P}_2$ *between* s *and* t *passing through two nodes* $v_1, v_2 \in \mathcal{L}_j$ *for some* $13\,\delta \leq j \leq \mathsf{dist}_G(s,t) - 7\,\delta\,\log n$*. Then,* $\mathsf{dist}_G(v_1, v_2) < 12\,\delta$*.*

Proof. For the sake of contradiction, suppose that $\mathsf{dist}_G(v_1, v_2) \geq 12\delta$. Let v_1' and v_2' be the first node in level $\mathcal{L}_{j+6\delta\log n}$ visited by $\mathcal{P}_1$ and $\mathcal{P}_2$, respectively. Since both $\mathcal{P}_1$ and $\mathcal{P}_2$ are paths between s and t and $j + 6\,\delta\,\log n < \mathsf{dist}_G(s,t)$ implies $\mathcal{L}_{j+6\delta\log n+1} \neq \emptyset$, there must be a path $\mathcal{P}_3$ between v_1' and v_2' through t using nodes not in $\bigcup_{0\leq \ell\leq j+6\,\delta\log n} \mathcal{L}_\ell$. We show that this is impossible by Fact 2(***b***). Set the parameters in Fact 2 in the following manner:

- $k = 4$, $\alpha = 6\delta\log n$, $r = j > 12\,k\,\delta = 48\,\delta$;
- $u_1 = v_1$, $u_2 = v_2$, $u_4 = v_1'$, and $u_3 = v_2'$.

This implies that the length of $\mathcal{P}_3$ satisfies

$$|\,\mathcal{P}_3\,| > 2^{\log n+5} > n$$

which contradicts the fact that $|\,\mathcal{P}_3\,| < n$. ❑

The above lemma immediately provide one hitting set in the following manner.

Lemma 9 *Assume* $\mathsf{dist}_G(s,t) > 13\delta + 8\delta\log n$*, and let* v *be an arbitrary node in level* $\mathcal{L}_j$ *lying on a path between* s *and* t *for some* $13\delta \leq j \leq \mathsf{dist}_G(s,t) - 7\delta\log n$*. Then,* $\mathcal{B}_G(v, 12\delta)$ *provides an* s*-*t *cut* $\mathrm{cut}_G(\mathcal{B}_G(v,12\delta), s, t)$ *having at most* $\mathcal{E}_G(\mathcal{B}_G(v,12\delta), s, t) \leq d^{\,12\delta+1}$ *edges.*

Proof. Consider any path $\mathcal{P}$ between s and t and let u be the first node in $\mathcal{L}_j$ visited by the path. By Lemma 8, $\mathsf{dist}_G(u,v) \leq 12\delta$ and thus $v \in \mathcal{B}_G(v, 12\delta)$. Since d is the maximum degree of any node other than s and t and $s,t \notin \mathcal{B}_G(v,12\delta)$, it follows that $\mathcal{E}_G(\mathcal{B}_G(v,12\delta), s, t) \leq d \times \partial_G(\mathcal{B}_G(v, 12\delta - 1)) \leq d^{\,12\delta+1}$. ❑

Using Lemma 9 we can now finish the proof of our theorem in the following way. Assume that $\mathsf{dist}_G(s,t) > 120\delta + 80\delta\log n$. Consider the levels $\mathcal{L}_j$ for $j \in \left\{30\delta, 60\delta, 90\delta, \dots, \left\lfloor \frac{\mathsf{dist}_G(s,t)-8\delta\log n}{30\delta} \right\rfloor \right\}$. For each such level $\mathcal{L}_j$, select a node v_j that is on a path between s and t and consider the subset of edges $\mathrm{cut}_G(\mathcal{B}_G(v_j, 12\delta), s, t)$. By Lemma 9, $\mathrm{cut}_G(\mathcal{B}_G(v_j,12\delta), s, t)$ is a valid s-t cut. The number of such cuts is at least $\left\lfloor \frac{\mathsf{dist}_G(s,t)-8\delta\log n}{30\delta} \right\rfloor$. To see why these cuts are node and edge disjoint, note that $\mathcal{E}_G(\mathcal{B}_G(v_j,12\delta), s, t) \cap \mathcal{E}_G(\mathcal{B}_G(v_\ell,12\delta), s, t) = \emptyset$ and $\mathcal{V}_G(\mathcal{B}_G(v_j,12\delta), s, t) \cap \mathcal{V}_G(\mathcal{B}_G(v_\ell,12\delta), s, t) = \emptyset$ for any $j \neq \ell$ since $\mathsf{dist}_G(v_j, v_\ell) > 30\delta$. ❑

5 Applications in Designing Improved Algorithms

Our results in the preceding sections show that hyperbolic graphs have many subsets of nodes of small expansion or small number of cut edges. Intuitively, such subsets of nodes should be useful in problems related to cuts and paths. In this section, we consider three such problems.

5.1 Hitting Sets for Size Constrained Cuts

The edge and node hitting set problems for size constrained s-t cuts can be defined as follows:

Problem name : *edge hitting set for size constrained cuts* (Ehssc).

(resp., *node hitting set for size constrained cuts* (NHSSC)).

Input : a graph $G = (V, E)$, two nodes $s, t \in V$, and a positive integer $0 < \kappa \le |E|$.

Valid solution : a hitting set $\widetilde{E}$ of all $\mathcal{E}_G(S, s, t)$'s satisfying $|\,\mathcal{E}_G(S, s, t)\,| \le \kappa$, *i.e.*, $\widetilde{E} \subseteq E$ satisfying $\forall\, s \in S \subset V \setminus \{t\}$: $|\,\mathcal{E}_G(S, s, t)\,| \le \kappa \Rightarrow \mathcal{E}_G(S, s, t) \cap \widetilde{E} \neq \emptyset$.
(resp., a hitting set $\widetilde{V}$ of all $\mathcal{V}_G(S, s, t)$'s satisfying $|\,\mathcal{V}_G(S, s, t)\,| \le \kappa$, *i.e.*, $\widetilde{V} \subseteq V$ satisfying $\forall\, s \in S \subset V \setminus \{t\}$: $|\,\mathcal{V}_G(S, s, t)\,| \le \kappa \Rightarrow \mathcal{V}_G(S, s, t) \cap \widetilde{V} \neq \emptyset$).

Objective : *minimize* $|\,\widetilde{E}\,|$ (resp., *minimize* $|\,\widetilde{V}\,|$).

Notation : $E_{\text{EHSSC}}(G, s, t, \kappa)$ (resp., $V_{\text{NHSSC}}(G, s, t, \kappa)$) is an optimal hitting set having $\mathsf{OPT}_{\text{EHSSC}}(G, s, t, \kappa)$ edges (resp., $\mathsf{OPT}_{\text{NHSSC}}(G, s, t, \kappa)$ nodes).

The following results for the EHSSC and NHSSC problems are implied by known results:

- NHSSC is NP-hard even if every node other then s and t has a degree of at most 5 (use [24, Theorem 2] with restricted instance of set cover and note that the reduction works for NHSSC even if all edges are undirected).
- EHSSC on a graph of n nodes does not admit a $2^{\log^{1-\varepsilon} n}$-approximation for any constant $\varepsilon > 0$ unless NP $\subseteq$ DTIME$\left(n^{\log\log n}\right)$.
- EHSSC on a graph of n nodes and m edges admits a $O\left(\min\left\{\, n^{3/4},\, m^{1/2}\,\right\}\right)$-approximation. Thus, in particular, if $m = O(n)$ then HSSC admits a $O\left(n^{1/2}\right)$-approximation.

Of course, both EHSSC and NHSSC have obvious *exponential-size* LP-relaxations since they are after all hitting set problems. For example, a exponential size LP-relaxation of EHSSC is as follows:

$$
\begin{array}{lll}
\textit{minimize} & \sum_{e \in E} x_e & \\
\textit{subject to} & \forall\, s \in S \subset V \setminus \{t\} \text{ such that } |\,\mathcal{E}_G(S, s, t)\,| \le \kappa : & \sum_{e \in \mathcal{E}_G(S,s,t)} x_e \ge 1 \\
 & \forall\, e \in E : & 0 \le x_e \le 1
\end{array}
$$

Intuitively, there are at least two reasons why such a LP-relaxation may not be of sufficient interest. Firstly, known results may imply a large integrality gap. Secondly, it is even not very clear if the LP-relaxation can be solved exactly in a time efficient manner.

Our main result for EHSSC and NHSSC shows that if the given graph is a *bounded degree hyperbolic* graph then these problems admit a $O(\log n)$-approximation, improving upon the currently best approximation ratio of $O\left(n^{1/2}\right)$.

Theorem 10 *If the given graph G is hyperbolic (i.e., if δ is a constant) and every node other than s and t has a constant maximum degree d then* EHSSC *(and consequently* NHSSC*) admit a $O(\log n)$-approximation,*

Remark 6 *For the general case when d or δ are not necessarily constants, our proof provides a $O\left(\max\left\{\log n,\, d^{\,12\delta+1}\right\}\right)$-approximation for* EHSSC *and a $O\left(\max\left\{\log n,\, d^{\,12\delta+2}\right\}\right)$-approximation for* NHSSC*. This improves upon the currently best $O\left(n^{1/2}\right)$-approximation provided $\delta = o\left(\log n/\log d\right)$. Thus, for example, for fixed d we provide improved approximation as long as $\delta = o(\log n)$.*

Proof. Since d is fixed, it suffices to to prove the result for EHSSC only (for the general case when d is not fixed, the approximation factor for EHSSC gets multiplied by d for NHSSC). Since the claim is trivial when $\delta = 0$ (a tree or the complete graph K_n), we assume in the sequel that $\delta \ge 1/2$. Our algorithm for EHSSC can be summarized as follows:

1. If $\kappa \leq d^{12\delta+1} = O(1)$ then solve the problem optimally in polynomial time by enumerating all possible $\sum_{j=1}^{\kappa} \binom{nd}{j} = n^{O(1)}$ subsets of at most κ edges.

   (For arbitrary (not necessarily constant) δ and d, one can get a polynomial-time $O\left(d^{12\delta+1}\right)$-approximation by using the primal-dual algorithm in [4, 24]).

2. Otherwise, assume that $\kappa > d^{12\delta+1}$. If $\mathsf{dist}_G(s,t) \leq 120\delta + 80\delta \log n$ then return all the edges in a shortest path between s and t as the solution. Since $\mathsf{OPT}_{\text{EHSSC}}(G, s, t, \kappa) \geq 1$, this provides a $(120\delta + 80\delta \log n) = O(\log n)$-approximation.

3. Otherwise, assume that $\kappa > d^{12\delta+1}$ and $\mathsf{dist}_G(s,t) > 120\delta + 80\delta \log n$. Use Theorem 7 to find a collection $S_1, S_2, \ldots, S_\ell$ of $\ell = \left\lfloor \frac{\mathsf{dist}_G(s,t) - 8\delta \log n}{30\delta} \right\rfloor$ edge and node disjoint s-t cuts. Since $\mathcal{E}_G(S_j s, t) \leq d^{12\delta+1} < \kappa$, any valid solution of EHSSC must select *at least* one edge from $\mathcal{E}_G(S_j, s, t)$. Since the cuts are edge and node disjoint, it follows that

   $$\mathsf{OPT}_{\text{EHSSC}}(G, s, t, \kappa) \geq \left\lfloor \frac{\mathsf{dist}_G(s,t) - 8\delta \log n}{30\delta} \right\rfloor$$

   In this case also, we return all the edges in a shortest path between s and t as the solution. The approximation ratio achieved is therefore

   $$\frac{\mathsf{dist}_G(s,t)}{\left\lfloor \frac{\mathsf{dist}_G(s,t) - 8\delta \log n}{30\delta} \right\rfloor} < 60\delta$$

❑

5.2 Minimizing Bottleneck Edges

The HSSC problem can be used to characterize an optimal solution for other problems for the purpose of designing approximation algorithms for these problems. As an example, in this section we consider the following problem.

Problem name :	*unweighted uncapacitated minimum vulnerability* (UUMV) [4, 24, 31].
Input :	a graph $G = (V, E)$, two nodes $s, t \in V$, and two positive integers $0 \leq r < k$.
Definition :	an edge is called *shared* if it is in more than r paths between s and t.
Valid solution :	a set of k paths between s and t.
Objective :	*minimize* the number of shared edges.
Notation :	$\mathsf{OPT}_{\text{UUMV}}(G, s, t, r, k)$ is the number of shared edges in an optimal solution.

UUMV has applications in several communication network design problems [29–31]. When $r = 1$, the UUMV problem is known as the *minimum shared edges* (MSE) problem. It was shown in in [4, 24] that, for a graph with n nodes and m edges, MSE does not admit a $2^{\log^{1-\varepsilon} n}$-approximation for any constant $\varepsilon > 0$ unless $\text{NP} \subseteq \text{DTIME}\left(n^{\log \log n}\right)$, UUMV admits a $\lfloor k/(r+1) \rfloor$-approximation, and MSE admits a $\min\left\{ \lfloor k/2 \rfloor, n^{3/4}, m^{1/2} \right\}$-approximation. The following lemma shows that UUMV (and thus MSE) has the *same approximability properties as* EHSSC by characterizing optimal solutions of UUMV in terms of optimal solutions of EHSSC.

Lemma 11 $\mathsf{OPT}_{\text{UUMV}}(G, s, t, r, k) = \mathsf{OPT}_{\text{EHSSC}}\left(G, s, t, \lceil k/r \rceil - 1\right)$*, and thus the approximation algorithms for* EHSSC *in Theorem 10 and Remark 6 carry over to corresponding bounds for* UUMV*..*

Proof. Note that *any* feasible solution for UUMV *must* contain at least one edge from every collection of cut-edges $\mathcal{E}_G(S,s,t)$ satisfying $|\,\mathcal{E}_G(S,s,t)\,| \leq \lceil k/r \rceil - 1$, since otherwise the number of paths going from $\text{cut}_G(S,s,t)$ to $V \setminus \text{cut}_G(S,s,t)$ is at most $r \times (\,\lceil k/r \rceil - 1\,) < k$. Thus it follows that $\mathsf{OPT}_{\text{UUMV}}(G,s,t,r,k) \geq \mathsf{OPT}_{\text{EHSSC}}\,(G,s,t,\lceil k/r \rceil - 1)$.

On the other hand, $\mathsf{OPT}_{\text{UUMV}}(G,s,t,r,k) \leq \mathsf{OPT}_{\text{EHSSC}}\,(G,s,t,\lceil k/r \rceil - 1)$ can be argued as follows. Consider the set of edges $E_{\text{EHSSC}}\,(G,s,t,\lceil k/r \rceil - 1)$ in an optimal hitting set and set the capacity $c(e)$ of every edge e of G as $c(e) = \begin{cases} \infty, & \text{if } e \in E_{\text{HSSC}}\,(G,s,t,\lceil k/r \rceil - 1) \\ r, & \text{otherwise} \end{cases}$. The value of the minimum cut for G is then at least $\min\{\infty,\, r \times \lceil k/r \rceil\} \geq k$ which implies (by the max-flow-min-cut theorem) the existence of k flows each of unit value. The paths taken by these k flows provide our desired k paths for UUMV. ❑

5.3 The Small Set Expansion Problem

The *small set expansion* (SSE) problem was studied by Arora, Barak and Steurer in [3] (and also by several other researchers such as [5, 16, 26–28]) in an attempt to understand the computational difficulties surrounding the Unique Games Conjecture (UGC). of the adjacency matrix of a d-regular graph, For uniformity we also use the standard *normalized* edge-expansion of a graph which can be defined as follows [13]. For a subset of nodes S, let $\mathsf{vol}(S)$ denote the sum of degrees of the nodes in G. Then, the normalized edge expansion ratio $\Phi(S)$ of a subset of nodes of at most $|V'|/2$ nodes of a graph $H=(V',E')$ is defined as $\Phi_H(S) = \frac{\text{cut}_H(S)}{\mathsf{vol}(S)}$. Since we will deal with only d-regular graphs in this subsection, $\Phi_H(S)$ simplifies to $\Phi_H(S) = \frac{\text{cut}_H(S)}{d\,|S|}$. Note that $h_H(S) \leq \varepsilon$ implies $\Phi_H(S) \leq \frac{d\,h_H(S)}{d} \leq \varepsilon$.

Definition 12 ((SSE Problem) [a case of [3, Theorem 2.1], rewritten as a problem] *Suppose that we are given a d-regular graph $G=(V,E)$ for some fixed d, and suppose G has a subset of at most ζn nodes S (for some constant $0 < \zeta < 1/2$) such that $\Phi_G(S) \leq \varepsilon$ (or, $g_G(S) \leq \varepsilon$) for some constant $0 < \varepsilon \leq 1$. Then, find as efficiently as possible a subset S' of at most at most ζn nodes such that $\Phi_G(S) \leq \eta\,\varepsilon$ (or, $g_G(S) \leq \eta\,\varepsilon$) for some "universal constant" $\eta > 0$.*

In general, computing a very good approximation of the SSE problem seems to be quite hard; the approximation ratio of the algorithm presented in [27] roughly deteriorates proportional to $\sqrt{\log(1/\zeta)}$, and a $O(1)$-approximation described in [5] works only if the graph excludes two specific minors. The authors in [3] showed how to design a sub-exponential time (*i.e.*, $O\left(2^{c\,n}\right)$ time for some constant $c<1$) algorithm for the above problem. As they remark, expander like graphs are somewhat easier instances of SSE [3] for their algorithm, and it takes some non-trivial technical effort to handle the "non-expander" graphs. Note that *the class of hyperbolic graphs* (*i.e.*, when the hyperbolicity δ is a constant) *is a non-trivial proper subclass of non-expander graphs.* We show that SSE (as defined in Definition 12) can be solved efficiently for this proper subclass of non-expanders.

Theorem 13 (SSE for hyperbolic graphs is polynomial-time solvable) *Suppose that G is a d-regular graph that is also hyperbolic (*i.e., *δ is a constant). Then the* SSE *problem for G can be solved in polynomial time even for a wider range of parameter than Definition 12 suggests*[6], i.e., *when $d \leq c \log\log n$, $\zeta \geq \dfrac{1}{c\log n}$, and $\varepsilon \geq \dfrac{1}{c\sqrt[3]{\log n}}$ for any sufficiently large constant $c > 0$.*

Remark 7 *Our proof is quite similar to that used for Theorems 3 and 5. But, instead of looking for smallest possible non-expansion bounds, we now relax the search and allow us to consider subsets of*

[6]Better combinations of parameter ranges are possible; our intention here was to state that these parameters need *not* be constants.

nodes whose expansion is just enough to satisfy the requirement. This relaxation helps us to ensure the size requirement of the subset we need to find.

Proof. Since $d = O(\log\log n)$, one can easily verify (by setting the parameter μ to a value slightly higher than $1/2$) that it suffices to prove the result for h_G. We will use the proofs of Theorem 3 and Theorem 5 in this proof, so we urge the readers to familiarize themselves with the details of these proofs before reading the current proof. Instead of choosing two nodes p and q that are furthest apart, we now chose them such that $\mathsf{dist}_G(p, q) = \log_d n = \frac{\log n}{\log d}$ (cf. Remark 3). For convenience at the expense of slight misuse of notations, we let $D = \mathsf{dist}_G(p, q)$ instead of D being the actual diameter of G. For concreteness, set the constant μ to be $1/2$ in the proofs of Theorem 3 and Theorem 5. Note that $\frac{360\log n}{D\,2^{\frac{D^\mu}{7\,\delta\,\log(2d)}}} < \left(\frac{1}{D}\right)^{1-\mu}$ since

$$\frac{360\log n}{D\,2^{\frac{D^\mu}{7\,\delta\,\log(2d)}}} < \left(\frac{1}{D}\right)^{1-\mu} \Leftarrow \frac{360\log d}{2^{\frac{(\log n)^{1/2}}{14\,\delta\,(\log d)^{3/2}}}} < \left(\frac{\log d}{\log n}\right)^{1/2} \equiv 360\,(\log n)^{1/2} < \frac{2^{\frac{(\log n)^{1/2}}{14\,\delta\,(\log d)^{3/2}}}}{(\log d)^{1/2}}$$

$$\Leftarrow \frac{1}{2}\log\log n + 9 < \frac{(\log n)^{1/2}}{14\,\delta\,\left(\log\log\log n + \log c\right)^{3/2}} - \frac{1}{2}\log\log\log n - \frac{1}{2}\log c$$

and the last inequality holds since δ and c are constants.

First, suppose that there exists $0 \le r \le \frac{D}{3} - \alpha\,D$ such that $h_G\left(\mathcal{B}_{G_{-\mathcal{C}}}(p, r)\right) = h_G\left(\mathcal{B}_G(p, r)\right) \le \varepsilon$. We return $S' = \mathcal{B}_G(p, r)$ as our solution, To verify the size requirement, note that

$$|\,\mathcal{B}_G(p, r)\,| \le \left|\mathcal{B}_G\left(p, \frac{D}{3} - \alpha\,D\right)\right| < \left|\mathcal{B}_G\left(p, \frac{D}{3}\right)\right| < \sum_{i=0}^{D/3} d^{\,i} < d^{\frac{D}{3}+1} = d\,n^{1/3} < \zeta\,n \qquad (23)$$

where the last inequality follows since $d \le c\log\log n$ and $\zeta \ge \frac{1}{c\,\log n}$.

Otherwise, no such r exists, and this implies

$$\left|\mathcal{B}_G\left(p, \frac{D}{3} - \alpha\,D\right)\right| \ge (1+\varepsilon)^{\frac{D}{3} - \alpha\,D} > (1+\varepsilon)^{\frac{D}{4}} \ge \mathbf{e}^{\frac{\varepsilon D}{8}} = \mathbf{e}^{\frac{\varepsilon\log_d n}{8}} = n^{\frac{\varepsilon\log_d \mathbf{e}}{8}}$$

Now there are two major cases as follows.

Case 1: there exists at least one path between p and q in $G_{-\mathcal{C}}$.

We know that $\mathsf{dist}_{G_{-\mathcal{C}}}(p, q) \ge \frac{D}{60}2^{\alpha\,D/\delta}$ and (by choice of p) $\left|\mathcal{B}_{G_{-\mathcal{C}}}\left(p, \frac{\mathsf{dist}_{G_{-\mathcal{C}}}(p,q)}{2}\right)\right| < \frac{n}{2}$. Let $p = u_0, u_1, \dots, u_{t-1}, u_t = q$ be the nodes in successive order on a shortest path from p to q of length $t = \mathsf{dist}_{G_{-\mathcal{C}}}(p, q)$. Perform a BFS starting from p in $G_{-\mathcal{C}}$, and let $\mathcal{L}_i$ be the sets of nodes at the i^{th} level (*i.e.*, $\forall\, u \in \mathcal{L}_i$: $\mathsf{dist}_{G_{-\mathcal{C}}}(p, u) = i$). Note that $\left|\bigcup_{j=0}^{t/2}\mathcal{L}_j\right| \le \frac{n}{2}$. Consider the levels $\mathcal{L}_0, \mathcal{L}_1, \dots, \mathcal{L}_{t/2}$, and partition the ordered sequence of integers $0, 1, 2, \dots, t/2$ into consecutive blocks $\Delta_0, \Delta_1, \dots, \Delta_{(1+\frac{t}{2})/\kappa - 1}$ each of length $\kappa = \frac{8}{\varepsilon}\ln n$, *i.e.*,

$$\underbrace{0, 1, 2, \dots, \kappa-1}_{\Delta_0}, \underbrace{\kappa, \kappa+1, \kappa+2, \dots, 2\kappa-1}_{\Delta_1}, \dots\dots, \underbrace{\frac{t}{2} - \kappa + 1, \frac{t}{2} - \kappa + 2, \dots, \frac{t}{2}}_{\Delta_{\left(1+\frac{t}{2}\right)/\kappa - 1}}$$

We claim that for every Δ_i, there exists an index i^* within Δ_i (*i.e.*, there exists an index $i\,\kappa \le i^* \le i\,\kappa + \kappa - 1$) such that $h_G\left(\mathcal{L}_{i^*}\right) \le \varepsilon$. Suppose for the sake of contradiction that this is not true. Then, it follows that

$$\forall\, i\,\kappa \leq j \leq i\,\kappa + \kappa - 1:\; h_{G_{-\mathcal{C}}}(\mathcal{L}_j) \geq \frac{h_G(\mathcal{L}_j)}{2} > \frac{\varepsilon}{2}$$
$$\Rightarrow\; h_{G_{-\mathcal{C}}}(\mathcal{L}_{i\,\kappa+\kappa-1}) > |\,\mathcal{L}_{i\,\kappa}\,| \left(1+\frac{\varepsilon}{2}\right)^{\kappa} \geq \left(1+\frac{\varepsilon}{2}\right)^{\frac{8}{\varepsilon}\ln n} \geq \mathbf{e}^{\left(\frac{\varepsilon}{4}\right)\left(\frac{8}{\varepsilon}\ln n\right)} = n^2 > n$$

and the last inequality contradicts the fact that $\left|\bigcup_{j=0}^{t/2}\mathcal{L}_j\right| \leq \frac{n}{2}$. Since $\sum_{i=0}^{\left(1+\frac{t}{2}\right)/\kappa-1} |\,\mathcal{L}_{i^*}\,| < \frac{n}{2}$, there exists a set $\mathcal{L}_{k^*}$ such that $h_G(\mathcal{L}_{k^*}) \leq \varepsilon$ and

$$|\,\mathcal{L}_{k^*}\,| < \frac{n/2}{\frac{1+\frac{t}{2}}{\kappa}} = \frac{n\kappa}{2+t} < \frac{8\,n\,\ln n}{\varepsilon\,\frac{D}{60}\,2^{\frac{D^{1/2}}{7\,\delta\,\log(2d)}}} \leq \frac{480\,c\,n\,\ln\log n\,\sqrt[3]{\log n}}{2^{\frac{\sqrt{\log n}}{14\,\delta\,(\log\log\log n+\log c)^{3/2}}}} < \frac{n}{c\,\log n} \leq \zeta\, n$$

where we use the fact that δ and c are constants.

Case 2: there is no path between p and q in $G_{-\mathcal{C}}$.

In this case, we return $\mathcal{B}_{G_{-\mathcal{C}}}\left(p, \frac{D}{3} - \alpha\,D\right) = \mathcal{B}_G\left(p, \frac{D}{3} - \alpha\,D\right)$ as our solution. The size requirement follows since we showed in (23) that $\left|\mathcal{B}_G\left(p, \frac{D}{3} - \alpha\,D\right)\right| < \zeta\,n$. Note that nodes in $\mathcal{B}_G\left(p, \frac{D}{3} - \alpha\,D\right)$ can only be connected to nodes in $\mathcal{C}$, and thus

$$h_G\left(\mathcal{B}_G\left(p, \frac{D}{3} - \alpha\,D\right)\right) \leq \frac{|\,\mathcal{C}\,|}{\left|\mathcal{B}_G\left(p, \frac{D}{3} - \alpha\,D\right)\right|} \leq \frac{\frac{D}{3}d^{\alpha D}}{n^{\frac{\varepsilon\log_d \mathbf{e}}{8}}}$$
$$< \frac{\log n\,\left(c\,\log\log n\right)^{\frac{\sqrt{\log n}}{14\,(\log\log\log n+\log c)^{3/2}}}}{3\,\log 3\;n^{\frac{1}{8\,c\,\sqrt[3]{\log n}\,(\ln\log\log n+\ln c)}}} < \frac{1}{c\,\sqrt[3]{\log n}} \leq \varepsilon$$

where the penultimate inequality follows since c is a constant.

In all cases, the required subset of nodes can be found in $O\left(n^2\log n\right)$ time. ❑

Concluding Remarks

Computing the minimum node expansion ratio of a graph is in general NP-hard and is in fact SSE-hard to approximate within a ratio of $C\sqrt{\mathfrak{h}_G\,\log d}$ for some constant $C > 0$ [22]. Since we show that SSE is polynomial-time for hyperbolic graphs, the hardness result of [22] does not directly apply for this case, and thus additional arguments may be needed to establish similar hardness results for hyperbolic graphs.

Acknowledgements

The problem of investigating expansion properties of hyperbolic graphs was raised originally to some of the authors by Avi Wigderson.

References

[1] R. Albert, B. DasGupta and N. Mobasheri. *Topological implications of negative curvature for biological and social networks*, Physical Review E, 89 (3), 032811, 2014.

[2] F. Ariaei, M. Lou, E. Jonckeere, B. Krishnamachari, and M. Zuniga. *Curvature of sensor network: clustering coefficient*, EURASIP Journal on Wireless Communications and Networking, 213185, 2008.

[3] S. Arora, B. Barak and D. Steurer. *Subexponential Algorithms for Unique Games and Related Problems*, 51st annual IEEE Symposium on Foundations of Computer Science, 563-572, 2010.

[4] S. Assadi, E. Emamjomeh-Zadeh, A. Norouzi-Fard, S. Yazdanbod and H. Zarrabi-Zadeh. *The Minimum Vulnerability Problem*, Algorithmica, 70, 718-731, 2014.

[5] N. Bansal, U. Feige, R. Krauthgamer, K. Makarychev, V. Nagarajan, J. Naor and R. Schwartz. *Min-max Graph Partitioning and Small Set Expansion*, 52nd Annual IEEE Symposium on Foundations of Computer Science, 17-26, 2011.

[6] I. Benjamini. *Expanders are not hyperbolic*, Israel Journal of Mathematics, 108, 33-36, 1998.

[7] I. Benjamini, C. Hoppen, E. Ofek, P. Pralat and N. Wormald. *Geodesics and almost geodesic cycles in random regular graphs*, Journal of Graph Theory, 66, 115Ũ136, 2011.

[8] I. Benjamini and O. Schramm. *Finite transitive graph embedding into a hyperbolic metric space must stretch or squeeze*, Geometric Aspects of Functional Analysis, 123-126, Springer, 2012.

[9] M. R. Bridson and A. Haefliger. *Metric Spaces of Non-Positive Curvature*, Springer, 1999.

[10] V. Chepoi, F. F. Dragan, B. Estellon, M. Habib and Y. Vaxès. *Diameters, centers, and approximating trees of δ-hyperbolic geodesic spaces and graphs*, proceedings of the 24th Annual Symposium on Computational geometry, 59-68, 2008.

[11] V. Chepoi, F. F. Dragan, B. Estellon, M. Habib, Y. Vaxès and Y. Xiang. *Additive spanners and distance and routing labeling schemes for δ-hyperbolic graphs*, Algorithmica, 62(3-4), 713-732, 2012.

[12] V. Chepoi and B. Estellon. *Packing and covering δ-hyperbolic spaces by balls*, in Lecture Notes in Computer Science 4627, M. Charikar, K. Jansen, O. Reingold and J. D. P. Rolim (Eds.), 59-73, Springer, 2007.

[13] F. R. K. Chung. *Spectral Graph Theory*, CBMS Regional Conference Series in Mathematics, 92, 1997.

[14] G. Even, G. Kortsarz and W. Slany. *On network design problems: fixed cost flows and the covering steiner problem*, ACM Transaction on Algorithms, 1 (1), 74-101, 2005.

[15] H. Fournier, A. Ismail, and A. Vigneron. *Computing the Gromov hyperbolicity of a discrete metric space*, Information Processing Letters, 115, 6-8, 576Ũ579, 2015.

[16] R. Gandhi and G. Kortsarz. *On edge expansion problems and the small set expansion conjecture*, Discrete Applied Mathematics, 194, 93-101, 2015.

[17] C. Gavoille and O. Ly. *Distance labeling in hyperbolic graphs*, in Lecture Notes in Computer Science 3827, X. Deng and D.-Z. Du (Eds.), 1071-1079, Springer, 2005.

[18] M. Gromov. Hyperbolic groups, *Essays in group theory*, 8, 75-263, 1987.

[19] E. Jonckheere, P. Lohsoonthorn, and F. Bonahon. *Scaled Gromov hyperbolic graphs*, Journal of Graph Theory, 57(2), 157-180, 2007.

[20] E. A. Jonckheere, M. Lou, J. Hespanha and P. Barooah. *Effective resistance of Gromov-hyperbolic graphs: Application to asymptotic sensor network problems*, 46^{th} IEEE Conference on Decision and Control, 1453-1458, 2007.

[21] E. Jonckheerea, M. Loua, F. Bonahona and Y. Baryshnikova. *Euclidean versus hyperbolic congestion in idealized versus experimental networks*, Internet Mathematics, 7(1), 2011.

[22] A. Louis, P. Raghavendra and S. Vempala. *The Complexity of Approximating Vertex Expansion*, 54^{th} IEEE Annual Symposium on Foundations of Computer Science, 360-369, 2013.

[23] A. Malyshev. *Expanders are order diameter non-hyperbolic*, arXiv:1501.07904, 2015.

[24] M. T. Omran, J.-R. Sack and H. Zarrabi-Zadeh. *Finding paths with minimum shared edges*, Journal of Combinatorial Optimization, 26(4), 709-722, 2013.

[25] F. Papadopoulos, D. Krioukov, M. Boguna and A. Vahdat. *Greedy Forwarding in Dynamic Scale-Free Networks Embedded in Hyperbolic Metric Spaces*, IEEE INFOCOM, 1-9, 2010.

[26] P. Raghavendra and D. Steurer. *Graph Expansion and the Unique Games Conjecture*, 45^{th} ACM Symposium on Theory of Computing, 755-764, 2010.

[27] P. Raghavendra, D. Steurer and P. Tetali. *Approximations for the isoperimetric and spectral profile of graphs and related parameters*, 45^{th} ACM Symposium on Theory of Computing, 631-640, 2010.

[28] P. Raghavendra, D. Steurer and M. Tulsiani. *Reductions Between Expansion Problems*, IEEE Conference on Computational Complexity, 64-73, 2012.

[29] J. Wang, M. Yang, B. Yang and S. Q. Zheng. *Dual-homing based scalable partial multicast protection*, IEEE Transaction on Computers, 55 (9), 1130-1141, 2006.

[30] B. Yang, M. Yang, J. Wang and S. Q. Zheng. *Minimum cost paths subject to minimum vulnerability for reliable communications*, 8^{th} International Symposium on Parallel Architectures, Algorithms and Networks, 334-339, 2005.

[31] S. Q. Zheng, J. Wang, B. Yang and M. Yang. *Minimum-cost multiple paths subject to minimum link and node sharing in a network*, IEEE/ACM Transaction on Networking, 18 (5), 1436-1449, 2010.