Topological implications of negative curvature for biological and social networks

Réka Albert
Department of Physics, Pennsylvania State University, University Park, PA 16802

Bhaskar DasGupta and Nasim Mobasher
Department of Computer Science, University of Illinois at Chicago, Chicago, IL 60607
(Dated: March 6, 2014)

Network measures that reflect the most salient properties of complex large-scale networks are in high demand in the network research community. In this paper we adapt a combinatorial measure of negative curvature (also called hyperbolicity) to parameterized finite networks, and show that a variety of biological and social networks are hyperbolic. This hyperbolicity property has strong implications on the higher-order connectivity and other topological properties of these networks. Specifically, we derive and prove bounds on the distance among shortest or approximately shortest paths in hyperbolic networks. We describe two implications of these bounds to cross-talk in biological networks, and to the existence of central, influential neighborhoods in both biological and social networks.

PACS numbers: 87.18.Mp,87.18.Cf,87.18.Vf,89.75.Hc,02.10.Ox
Keywords: Hyperbolicity, Networks, Crosstalk, Influential nodes

I. INTRODUCTION

For a large variety of complex systems, ranging from the World-Wide Web to metabolic networks, representation as a parameterized network and graph theoretical analysis of this network have led to many useful insights [1, 2]. In addition to established network measures such as the average degree, clustering coefficient or diameter, complex network researchers have proposed and evaluated a number novel network measures [3–6]. In this article we consider a combinatorial measure of negative curvature (also called hyperbolicity) of parameterized finite networks and the implications of negative curvature on the higher-order connectivity and topological properties of these networks.

There are many ways in which the (positive or negative) curvature of a continuous surface or other similar spaces can be defined depending on whether the measure is to reflect the local or global properties of the underlying space. The specific notion of negative curvature that we use is an adoption of the hyperbolicity measure for a infinite metric space with bounded local geometry as originally proposed by Gromov [7]. We adopt this measure for parameterized finite discrete metric spaces induced by a network via all-pairs shortest paths and apply it to biological and social networks. Recently, there has been a surge of empirical works measuring and analyzing the hyperbolicity of networks defined in this manner, and many real-world networks were observed to be hyperbolic in this sense. For example, preferential attachment networks were shown to be scaled hyperbolic in [8, 9], networks of high power transceivers in a wireless sensor network were empirically observed to have a tendency to be hyperbolic in [10], communication networks at the IP layer and at other levels were empirically observed to be hyperbolic in [11, 12], extreme congestion at a very limited number of nodes in a very large traffic network was shown in [13] to be caused due to hyperbolicity of the network together with minimum length routing, and the authors in [14] showed how to efficiently map the topology of the Internet to a hyperbolic space.

Gromov’s hyperbolicity measure adopted on a shortest-path metric of networks can also be visualized as a measure of the “closeness” of the original network topology to a tree topology [15]. Another popular measure used in both the bioinformatics and theoretical computer science literature is the treewidth measure first introduced by Robertson and Seymour [16]. Many NP-hard problems on general networks admit efficient polynomial-time solutions if restricted to classes of networks with bounded treewidth [17], just as several routing-related problems or the diameter estimation problem become easier if the network has small hyperbolicity [18–21]. However, as observed in [14], the two measures are quite different in nature: “the treewidth is more related to the least number of nodes whose removal changes the connectivity of the graph in a significant manner whereas the hyperbolicity measure is related to comparing the geodesics of the given network with that of a tree”. Other related research works on hyperbolic networks include estimating the distortion necessary to map hyperbolic metrics to tree metrics [22] and studying the algorithmic aspects of several combinatorial problems on points in a hyperbolic space [23].

II. HYPERBOLICITY-RELATED DEFINITIONS AND MEASURES

Let $G = (V, E)$ be a connected undirected graph of $n \geq 4$ nodes. We will use the following notations:

- $u \xrightarrow{\varphi} v$ denotes a path $\varphi \equiv (u = u_0, u_1, \ldots, u_k = v)$ from node $u$ to node $v$ and $\ell(\varphi)$ denotes the length (number of edges) of such a path.
• $u_i \leftrightarrow u_j$ denotes the sub-path $(u_i, u_{i+1}, \ldots, u_j)$ of $\mathcal{P}$ from $u_i$ to $u_j$.

• $u \leftrightarrow v$ denotes a shortest path from node $u$ to node $v$ of length $d_{u,v} = \ell(u \leftrightarrow v)$.

We introduce the hyperbolicity measures via the 4-node condition as originally proposed by Gromov. Consider a quadruple of distinct nodes $u_1, u_2, u_3, u_4$, and 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

$$S_{u_1, u_2, u_3, u_4} = d_{u_1, u_2} + d_{u_3, u_4} + d_{u_1, u_3} + d_{u_2, u_4} \leq M_{u_1, u_2, u_3, u_4} = d_{u_1, u_3} + d_{u_2, u_4} + d_{u_1, u_4} + d_{u_2, u_3}$$

and let $\delta^+_{u_1, u_2, u_3, u_4} = \frac{L_{u_1, u_2, u_3, u_4} - M_{u_1, u_2, u_3, u_4}}{2}$. Considering all combinations of four nodes in a graph one can define a worst-case hyperbolicity\(^7\) as

$$\delta^\text{worst}_{\text{ave}}(G) = \max_{u_1, u_2, u_3, u_4} \{\delta^+_{u_1, u_2, u_3, u_4}\}$$

and an average hyperbolicity as

$$\delta^\text{ave}_{\text{ave}}(G) = \frac{1}{\binom{n}{4}} \sum_{u_1, u_2, u_3, u_4} \delta^+_{u_1, u_2, u_3, u_4}$$

Note that $\delta^\text{ave}_{\text{ave}}(G)$ is the expected value of $\delta^+_{u_1, u_2, u_3, u_4}$ when the four nodes $u_1, u_2, u_3, u_4$ are picked independently and uniformly at random from the set of all nodes. Both $\delta^\text{worst}_{\text{ave}}(G)$ and $\delta^\text{ave}_{\text{ave}}(G)$ can be trivially computed in $O(n^4)$ time for any graph $G$.

A graph $G$ is called $\delta$-hyperbolic if $\delta^\text{worst}_{\text{ave}}(G) \leq \delta$. If $\delta$ is a small constant independent of the parameters of the graph, a $\delta$-hyperbolic graph is simply called a hyperbolic graph. It is easy to see that if $G$ is a tree then $\delta^\text{worst}_{\text{ave}}(G) = \delta^\text{ave}_{\text{ave}}(G) = 0$. Thus all trees are hyperbolic graphs.

The hyperbolicity measure $\delta^\text{worst}_{\text{ave}}$ considered in this paper for a metric space was originally used by Gromov in the context of group theory\(^7\) by observing that many results concerning the fundamental group of a Riemann surface hold true in a more general context. $\delta^\text{worst}_{\text{ave}}$ is trivially infinite in the standard (unbounded) Euclidean space. Intuitively, a metric space has a finite value of $\delta^\text{worst}_{\text{ave}}$ if it behaves metrically in the large scale as a negatively curved Riemannian manifold, and thus the value of $\delta^\text{worst}_{\text{ave}}$ can be related to the standard scalar curvature of a Hyperbolic manifold. For example, a simply connected complete Riemannian manifold whose sectional curvature is below $\alpha < 0$ has a value of $\delta^\text{worst}_{\text{ave}}$ that is $O\left(\left(\frac{1}{\sqrt{-\alpha}}\right)^{-1}\right)$ (see [24]).

---

1 If two or more nodes among $u_1, u_2, u_3, u_4$ are identical, then $\delta^+_{u_1, u_2, u_3, u_4} = 0$ due to the metric's triangle inequality; thus it suffices to assume that the four nodes are distinct.
It can be seen that, for all scaled hyperbolicity used in this paper. The basic idea is to “scale” the values of hyperbolicity defined via the four-point condition as using theoretical or empirical calculations, the authors in [42] hyperbolicity of our networks.

Checking hyperbolicity via the scaled hyperbolicity approach

\[ \delta^*(G) = \min_{i,j \in V} \left\{ \max_{k,l \in V} \left\{ D^i_{kl} \right\} \right\} \]

An approach for testing hyperbolicity for finite graphs was introduced and used via “scaled” Gromov hyperbolicity in [6][11] for hyperbolicity defined via thin triangles and in [42] for for hyperbolicity defined via the four-point condition as used in this paper. The basic idea is to “scale” the values of \( \delta^*_i(u_1, u_2, u_3, u_4) \) by a suitable scaling factor, say \( \mu_{u_1, u_2, u_3, u_4} \), such that there exists a constant \( 0 < \varepsilon < 1 \) with the following property:

- the maximum achievable value of \( \frac{\delta^*_i(u_1, u_2, u_3, u_4)}{\mu_{u_1, u_2, u_3, u_4}} \) is \( \varepsilon \) in the standard hyperbolic space or in the Euclidean space, and
- \( \delta^*_i(u_1, u_2, u_3, u_4) \) goes beyond \( \varepsilon \) in positively curved spaces.

We use the notation \( D_{u_1, u_2, u_3, u_4} = \max_{i,j \in \{1,2,3,4\}} \{ d_{u_i, u_j} \} \) to indicate the diameter of the subset of four nodes \( u_1, u_2, u_3 \) and \( u_4 \). By using theoretical or empirical calculations, the authors in [42] provide the bounds shown in Table.I

We adapt the criterion proposed by Jonckheere, Lohsoonthorn and Ariuari [42] to designate a given finite graph as hyperbolic by requiring a significant percentage of all possible subset of four nodes to satisfy the \( \varepsilon \) bound. More formally, suppose that \( G \) has \( t \) connected components containing \( n_1, n_2, \ldots, n_t \) nodes, respectively (\( \sum_{i=1}^{t} n_i = n \)). Let \( 0 < \eta < 1 \) be a sufficiently high value indicating the confidence level in declaring the graph \( G \) to be hyperbolic. Then, we call our given graph \( G \) to be (scaled) hyperbolic if and only if

\[ \frac{\text{number of subset of four nodes } \{ u_i, u_j, u_k, u_l \} \text{ such that } \delta^*_i(u_1, u_2, u_3, u_4) > \varepsilon}{\text{number of all possible combinations of four nodes that contribute to hyperbolicity}} \]

\[ \leq \frac{\text{number of subset of four nodes } \{ u_i, u_j, u_k, u_l \} \text{ such that } \delta^*_i(u_1, u_2, u_3, u_4) > \varepsilon}{\sum_{1 \leq i < j < k < l} \binom{4}{3} \binom{n}{i,j,k,l} \} < 1 - \eta \]

The values of \( \Delta^*(G) \) for our networks are shown in Table.IV and Table.V It can be seen that, for all scaled hyperbolicity measures and for all networks, the value of \( 1 - \eta \) is very close to zero.

We next tested the statistical significance of the \( \Delta^*(G) \) values by computing the statistical significance values (commonly called \( p \)-values) of these \( \Delta^*(G) \) values for each network \( G \) with respect to a null hypothesis model of the networks. We use a standard method used in the network science literature (e.g., see [5][25]) for such purpose. For each network \( G \), we generated 100 randomized versions of the network using a Markov-chain algorithm [43] by swapping the endpoints of randomly selected pairs of edges until 20\% of the edges was changed. We computed the values of \( \Delta^*(G_{\text{rand}}) \). \( \Delta^*(G_{\text{rand}}) \), \( \ldots, \Delta^*(G_{\text{rand}}) \). We then used an (unpaired) one-sample student’s t-test to determine the probability that \( \Delta^*(G) \) belongs to the same distribution as \( \Delta^*(G_{\text{rand}}), \Delta^*(G_{\text{rand}}), \ldots, \Delta^*(G_{\text{rand}}) \). The \( p \)-values, tabulated in Table.VI and Table.VII clearly show that all social networks and all except two biological net-
TABLE IV. \( \Delta^Y(G) \) values for biological networks for \( Y \in \{D, L, L + M + S\} \).

| Network id   | \( \Delta^D(G) \) | \( \Delta^L(G) \) | \( \Delta^{L+M+S}(G) \) |
|-------------|-----------------|-----------------|-----------------|
| 1. E. coli transcriptional | 0.0014 | 0.0018 | 0.0015 |
| 2. Mammalian Signaling | 0.0021 | 0.0018 | 0.0022 |
| 3. E. Coli transcriptional | 0.0006 | 0.0006 | 0.0007 |
| 4. T LGL signaling | 0.0228 | 0.0221 | 0.0318 |
| 5. S. cerevisiae transcriptional | 0.0031 | 0.0032 | 0.0033 |
| 6. C. elegans Metabolic | 0.0020 | 0.0018 | 0.0019 |
| 7. Drosophila segment polarity | 0.0374 | 0.0558 | 0.0750 |
| 8. ABA signaling | 0.0343 | 0.0285 | 0.0425 |
| 9. Immune Response Network | 0.0461 | 0.0552 | 0.0781 |
| 10. T Cell Receptor Signalling | 0.0034 | 0.0045 | 0.0056 |
| 11. Oriented yeast PPI | 0.0013 | 0.0009 | 0.0012 |

TABLE VI. \( p \)-values for the \( \Delta^Y(G) \) values for biological networks for \( Y \in \{D, L, L + M + S\} \). In general, a \( p \)-value less than 0.05 (shown in boldface) is considered to be statistically significant, and a \( p \)-value above 0.05 is considered to be not statistically significant.

| Network id   | \( \Delta^D(G) \) | \( \Delta^L(G) \) | \( \Delta^{L+M+S}(G) \) |
|-------------|-----------------|-----------------|-----------------|
| 1. Dolphins social network | 0.0115 | 0.0120 | 0.0168 |
| 2. American College Football | 0.0435 | 0.0395 | 0.0577 |
| 3. Zachary Karate Club | 0.0195 | 0.0249 | 0.0284 |
| 4. Books about US Politics | 0.0106 | 0.0074 | 0.0116 |
| 5. Sawmill communication | 0.0069 | 0.0068 | 0.0085 |
| 6. Jazz musician | 0.0097 | 0.0117 | 0.0124 |
| 7. Visiting ties in San Juan | 0.0221 | 0.0242 | 0.0275 |
| 8. World Soccer data, 1998 | 0.0145 | 0.0155 | 0.0212 |
| 9. Les Miserable | 0.0032 | 0.0034 | 0.0049 |

B. Hyperbolicity and crosstalk in regulatory networks

Let \( C = (u_0, u_1, \ldots, u_k, u_0) \) be a cycle of \( k \geq 4 \) nodes. A path-chord of \( C \) is defined to be a path \( u_i \rightarrow u_j \) between two distinct nodes \( u_i, u_j \in C \) such that the length of \( P \) is less than \((i – j) (mod k)\) (see Fig. 1). A path-chord of length 1 is simply called a chord.

We find that large cycles without a path-chord imply large lower bounds on hyperbolicity (see Theorem 1 in Section 3 of the appendix). In particular, \( G \) does not have a cycle of more than \( 4 \delta_{\text{word}}(G) \) nodes that does not have a path-chord. Thus, for example, if \( \delta_{\text{worst}}(G) < 1 \) then \( G \) has no chordless cycle, i.e., \( G \) is a chordal graph. The intuition behind the proof of Theorem 1 is that if \( G \) contains a long cycle without a path-chord then we can select four almost equidistant nodes on the cycle and these nodes give a large hyperbolicity value. This general result has the following implications for regulatory networks:

- If a node regulates itself through a long feedback loop (e.g., of length at least 6 if \( \delta_{\text{worst}}(G) = 3/2 \)) then this loop must have a path-chord. Thus it follows that there exists a shorter feedback cycle through the same node.
- A chord or short path-chord can be interpreted as

![Path-chord of a cycle](image_url)
TABLE VII. p-values for the $\Delta^2(G)$ values for social networks for $Y \in \{D, L, L + M + S\}$. In general, a p-value less than 0.05 (shown in boldface) is considered to be statistically significant, and a p-value above 0.05 is considered to be not statistically significant.

| Network id       | 1. Dolphins social network | 2. American College Football | 3. Zachary Karate Club | 4. Books about US Politics | 5. Sawmill communication | 6. Jazz musician | 7. Visiting ties in San Juan | 8. World Soccer data, 1998 | 9. Les Miserable |
|------------------|----------------------------|------------------------------|------------------------|---------------------------|------------------------|-----------------|----------------------------|-------------------|------------------|
| p values         | $\Delta^D \leq 0.0001$   | $< 0.0001$                  | $< 0.0001$             | $< 0.0001$                | $< 0.0001$             | $< 0.0001$      | $< 0.0001$                | $< 0.0001$         | $< 0.0001$       |
|                  | $\Delta^L \leq 0.0001$   | $< 0.0001$                  | $< 0.0001$             | $< 0.0001$                | $< 0.0001$             | $< 0.0001$      | $< 0.0001$                | $< 0.0001$         | $< 0.0001$       |
|                  | $\Delta^L + M + S \leq 0.0001$ | $< 0.0001$                  | $< 0.0001$             | $< 0.0001$                | $< 0.0001$             | $< 0.0001$      | $< 0.0001$                | $< 0.0001$         | $< 0.0001$       |

C. Shortest-path triangles and crosstalk paths in regulatory networks

(a) Result related to triplets of shortest paths Originally, the hyperbolicity measure was introduced for infinite continuous metric spaces with negative curvature via the concept of the “thin” and “slim” triangles (e.g., see [46]). For finite discrete metric spaces as induced by an undirected graph, one can analogously define a shortest-path triangle (or, simply a triangle) $\Delta_{\{u_0, u_1, u_2\}}$ as a set of three distinct nodes $u_0, u_1, u_2$ with a set of three shortest paths $P_{\Delta}(u_0, u_1), P_{\Delta}(u_0, u_2), P_{\Delta}(u_1, u_2)$ between $u_0$ and $u_1$, $u_0$ and $u_2$, and $u_1$ and $u_2$, respectively. As illustrated on Fig. 2 in hyperbolic networks we are guaranteed to find shortest path between the nodes that make up $P_{\Delta}(u_0, u_1), P_{\Delta}(u_0, u_2), P_{\Delta}(u_1, u_2)$. This is formally stated in Theorem 3 in Section III C(a) of the appendix. Moreover, as Corollary 4 (in Section III C of the appendix) states, we can have a small Hausdorff distance between these shortest paths. This result is a proper generalization of our previous result on path-cords. Indeed, in the special case when $u_1$ and $u_2$ are the same node the triangle becomes a shortest-path cycle involving the shortest paths between $u_0$ and $u_1$ and the short-cord result is obtained.

(b) Results related to the distance between two exact or approximate shortest paths between the same pair of nodes It is reasonable to assume that, when up- or down-regulation

FIG. 2. An informal and simplified pictorial illustration of the claims in Section III C(a).

A proof of Theorem 3 is obtained by appropriate modification of a known similar bound for infinite continuous metric spaces.

The implications of this result for regulatory networks can be summarized as follows:

If we consider a feedback loop (cycle) or feed-forward loop formed by the shortest paths among three nodes, we can expect short cross-talk paths between these shortest paths. Consequently, the feedback or feed-forward loop will be nested with additional feed-back or feed-forward loops in which one of the paths will be slightly longer.

The above finding is empirically supported by the observation that network motifs (e.g., feed-forward or feedback loops composed of three nodes and three edges) are often nested [47].

2 By a short path here, we mean a path whose length is at most a constant times $\gamma_{\{u_0, u_1, u_2\}}$ (note that $\gamma_{\{u_0, u_1, u_2\}} \leq \gamma_{\text{word}}(G)$).
of a target node is mediated by two or more short paths starting from the same regulator node, additional very long paths between the same regulator and target node do not contribute significantly to the target node’s regulation. We refer to the short paths as relevant, and to the long paths as irrelevant. Then, our finding can be summarized by saying that:

*almost all* relevant paths between two nodes have crosstalk paths between each other.

### Formal Justifications and Intuitions

(see Theorem 5 and Corollary 6 in Section C and Theorem 7 and Corollary 8 in Section D of the appendix)

We use the following two quantifications of “approximately” short paths:

- A path $u_0 \leftrightarrow u_k = (u_0, u_1, \ldots, u_k)$ is $\mu$-approximate short provided $\ell(u_i \leftrightarrow u_j) \leq \mu d_{u_i, u_j}$ for all $0 \leq i < j \leq k$,

- A path $u_0 \leftrightarrow u_k$ is $\epsilon$-additive-approximate short provided $\ell(P) \leq d_{u_0, u_k} + \epsilon$.

A mathematical justification for the claim then is provided by two separate theorems and their corollaries:

- Let $P_1$ and $P_2$ be a shortest path and an arbitrary path, respectively, between two nodes $u_0$ and $u_1$. Then, Theorem 5 and Corollary 6 implies that, for every node $v$ on $P_1$, there exists a node $v'$ on $P_2$ such that $d_{v, v'}$ depends linearly on $\delta_{\text{worst}}(G)$, only logarithmically on the length of $P_2$ and does not depend on the size or any other parameter of the network.

To obtain this type of bound, one needs to apply Theorem 5 on $u_0$, $u_1$, and the middle node of the path $P_2$ and then use the same approach recursively on a part of the path $P_2$ containing at most $\left\lceil \frac{|P_2|}{2} \right\rceil$ edges. The depth of the level of recursion provides the logarithmic factor in the bound.

---

3 Here by short paths we mean either a shortest path or an approximately shortest path whose length is not too much above the length of a shortest path, i.e., a $\mu$-approximate short path or an $\epsilon$-additive-approximate short path, as defined in the subsequent “Formal Justifications and Intuitions” subsection, for small $\mu$ or small $\epsilon$, respectively.
Thus, only the edges inside the neighborhood are relevant to the regulation among this pair of nodes.

This result can be adapted to find the most relevant paths between the input node $u_{\text{source}}$ and output node $u_{\text{target}}$ of a signal transduction network. In many situations, for example when the signal transduction network is inferred from undirected protein-protein interaction data, a large number of paths can potentially be included in the signal transduction network as the protein-protein interaction network has a large connected component with a small average path length [47]. There is usually no prior knowledge on which of the existing paths are relevant to the signal transduction network. A hyperbolicity-based method is to first find a central node $u_{\text{central}}$ which is at equal distance between $u_{\text{source}}$ and $u_{\text{target}}$, and is on the shortest, or close to shortest, path between $u_{\text{source}}$ and $u_{\text{target}}$. Then one constructs the neighborhood around $u_{\text{central}}$ such that $u_{\text{source}}$ and $u_{\text{target}}$ are on the boundary of this neighborhood. Applying this result, the paths relevant to the signal transduction network are inside the neighborhood, and the paths that go out of the neighborhood are irrelevant. See Fig. 5 for a pictorial illustration of this implication.

(II) Finding essential nodes Again, consider an input node $u_{\text{source}}$ and output node $u_{\text{target}}$ of a signal transduction network, and let $u_{\text{central}}$ be a central node which is on the shortest path between them and at approximately equal distance between $u_{\text{source}}$ and $u_{\text{target}}$. Our results show that if one constructs a small $\xi$-neighbourhood around $u_{\text{central}}$ with $\xi = O(\delta_{\text{worst}}^{-}(G))$, then all relevant (short or approximately short) paths between $u_{\text{source}}$ and $u_{\text{target}}$ must include a node in this $\xi$-neighborhood. Therefore, “knocking out” the nodes in this $\xi$-neighborhood cuts off all relevant regulatory paths between $u_{\text{source}}$ and $u_{\text{target}}$.

See Fig. 5 for a pictorial illustration of this implication. Note that the size $\xi$ of the neighborhood depends only on $\delta_{\text{worst}}^{-}(G)$ which, as our empirical results indicate, is usually a small constant for real networks.

**Formal Justifications and Intuitions for (⋆) and (⋆⋆)** (see Theorem [10] and Corollary [11] in Section [E] of the appendix)

Suppose that we are given the following:

- three integers $κ ≥ 4$, $α > 0$, $r > (\frac{κ}{2} - 1)(6\delta_{\text{worst}}^{+}(G) + 2)$,
- five nodes $u_0, u_1, u_2, u_3, u_4$ such that
  - $u_1, u_2 \in B_r(u_0)$ with $d_{u_1, u_2} ≥ (\frac{κ}{2} - 1)(6\delta_{\text{worst}}^{+}(G) + 2)$,
  - $d_{u_1, u_4} = d_{u_2, u_3} = α$.

Then, (⋆) and (⋆⋆) are implied by following type of asymptotic bounds provided by Theorem [10] and Corollary [11].

For a suitable positive value $λ = O(\delta_{\text{worst}}^{+}(G))$, if $d_{u_1, u_4} = d_{u_2, u_3} = α > λ$ then one of the following is true for any path $Q$ between $u_1$ and $u_4$ that does not involve a node in $\cup_{r/2, r, \ell}(u_0)$:

- $Q$ does not exist (i.e., $\ell(Q) ≥ n$), or
- $Q$ is much longer than a shortest path between the two nodes, i.e., if $Q$ is a $\mu$-approximate short path or a $\varepsilon$-additive-approximate short path then $µ$ or $ε$ is large.

A pessimistic estimate shows that a value of $λ$ that is about $6\delta_{\text{worst}}^{+}(G) + 2$ suffices. As we subsequently observe, for real networks the bound is much better, about $λ ≈ \delta_{\text{worst}}^{+}(G)$.

**Empirical evaluation of (⋆)**

We empirically investigated the claim in (⋆) on relevant paths passing through a neighborhood of a central node for the following two biological networks:

**Network 1::** E. coli transcriptional, and

**Network 4::** T-LGL signaling.

For each network we selected a few biologically relevant source-target pairs. For each such pair $u_{\text{source}}$ and $u_{\text{target}}$, we found the shortest path(s) between them. For each such shortest path, a central node $u_{\text{central}}$ was identified. Then we considered the $\xi$-neighborhood of $u_{\text{central}}$ such that both both $u_{\text{source}}$ and $u_{\text{target}}$ are on the boundary of the neighborhood, and for each such neighborhood we determined what percentage of shortest or approximately short path (with one or two extra edges compared to shortest paths) between $u_{\text{source}}$ and $u_{\text{target}}$ had all edges in this neighborhood. The results, tabulated in Table [VIII] support (⋆).

**Empirical evaluation of (⋆⋆)**

We empirically investigated the size $\xi$ of the neighborhood in claim (⋆⋆) for the same two biological networks and the same combinations of source, target and central nodes as in
TABLE VIII. Effect of the prescribed neighborhood in claim (\( \star \)) on all edges in relevant paths.

| Network name | \( u_{\text{source}} \) | \( u_{\text{target}} \) | \( d_{u_{\text{source}}, u_{\text{target}}} \) | \( u_{\text{central}} \) | \( N_{\xi}^\ast (u_{\text{central}}) \) | \( n \) | \% of \( SP \) with every edge in the neighborhood of claim (\( \star \)) | \% of \( SP^\ast \) with every edge in the neighborhood of claim (\( \star \)) | \% of \( SP^{\ast +} \) with every edge in the neighborhood of claim (\( \star \)) |
|--------------|----------------|----------------|----------------|----------------|----------------|------|----------------|----------------|----------------|
| Network 1: E. coli transcriptional | fliAZY | arcA | 4 | CalF | 0.20 | 100% | 100% | 100% |
| | | | | crp | 0.27 | 100% | 100% | 70% |
| | | | | crp | 0.43 | 100% | 100% | 100% |
| | fecA | aspA | 6 | sodA | 0.28 | 100% | 100% | 62% |
| Network 4: T-LGL signaling | IL15 | Apoptosis | 4 | GZMB | 0.37 | 100% | 100% | 40% |
| | PDGF | Apoptosis | 6 | IL2, NFKB | 0.72, 0.59 | 100% | 100% | 100% |
| | | | | Ceramide | 0.60 | 80% | 64% | 36% |
| | | | | MCL1 | 0.59 | 80% | 88% | 93% |
| | Stimuli | Apoptosis | 4 | GZMB | 0.37 | 100% | 100% | 100% |

TABLE IX. The effect of the size of the neighborhood in mediating short paths.

| Network name | \( u_{\text{source}} \) | \( u_{\text{target}} \) | \( d_{u_{\text{source}}, u_{\text{target}}} \) | \( u_{\text{central}} \) | \( \% \) of \( SP \) with a node in \( \xi \)-neighborhood | \( \% \) of \( SP^\ast \) with a node in \( \xi \)-neighborhood | \( \% \) of \( SP^{\ast +} \) with a node in \( \xi \)-neighborhood |
|--------------|----------------|----------------|----------------|----------------|----------------|----------------|----------------|
| Network 1: E. coli transcriptional | fliAZY | arcA | 4 | CalF | \( \xi = 1 \) | 100% | 71% | 59% |
| | | | | crp | \( \xi = 1 \) | 100% | 100% | 100% |
| | | | | sodA | \( \xi = 1 \) | 100% | 100% | 100% |
| | fecA | aspA | 6 | crp | \( \xi = 1 \) | 100% | 100% | 100% |
| | | | | sodA | \( \xi = 1 \) | 100% | 100% | 100% |
| | IL15 | apoptosis | 4 | GZMB | \( \xi = 1 \) | 100% | 100% | 100% |
| Network 4: T-LGL signaling | PDGF | apoptosis | 6 | IL2 | \( \xi = 1 \) | 80% | 82% | 93% |
| | | | | NFKB | \( \xi = 2 \) | 100% | 100% | 100% |
| | | | | Ceramide | \( \xi = 1 \) | 40% | 23% | 40% |
| | | | | MCL1 | \( \xi = 2 \) | 100% | 100% | 100% |

E. Effect of hyperbolicity on structural holes in social networks

For a node \( u \in V \), let \( \text{Nbr}(u) = \{ v \mid (u, v) \in E \} \) be the set of neighbors of (i.e., nodes adjacent to) \( u \). To quantify the useful information in a social network, Ron Burt in [48] defined a measure of the structural holes of a network. For an undirected unweighted connected graph \( G = (V, E) \) and a node \( u \in V \) with degree larger than 1, this measure \( W_u \) of the structural hole at \( u \) is defined as [48, 49].

claim (\( \star \)). We considered the \( \xi \)-neighborhood of \( u_{\text{central}} \) for \( \xi = 1, 2, \ldots \), and for each such neighborhood we determined what percentage of shortest or approximately short path (with one or two extra edges compared to shortest paths) between \( u_{\text{source}} \) and \( u_{\text{target}} \) involved a node in this neighborhood (not counting \( u_{\text{source}} \) and \( u_{\text{target}} \)). The results, tabulated in Table [IX], show that removing the nodes in a \( \xi \leq \delta^\ast_w(G) \) neighborhood around the central nodes disrupts all the relevant paths of the selected networks. As \( \delta^\ast_w(G) \) is a small constant for all of our biological networks, this implies that the central node and its neighbors within a small distance are the essential nodes in the signal propagation between \( u_{\text{source}} \) and \( u_{\text{target}} \).
Let $M$ be the shortest path between $u$ since only one shortest path between $v$ and $y$ intersects $B_j(u)$. Thus high-degree nodes whose neighbors are not connected are the entries in the standard adjacency matrix of $G$. By observing that $a_{p,q} = a_{p,p}$ and $\max \{a_{u,x} + a_{x,u}\} = \max \{a_{u,z} + a_{z,u}\} = 2$, the above equation for $M_u$ can be simplified to

$$M_u = |\text{Nbr}(u)| - \sum_{v \in \text{Nbr}(u)} a_{v,y}$$

Thus high-degree nodes whose neighbors are not connected to each other have high $M_u$ values. For an intuitive interpretation and generalization of (1), the following definition of weak and strong dominance will prove useful (c.f. dominating set problem for graphs [50] and point domination problems in geometry [51]). A pair of distinct nodes $v, y$ is weakly $(\rho, \lambda)$-dominated (respectively, strongly $(\rho, \lambda)$-dominated) by a node $u$ provided (see Fig. 6):

(a) $\rho < d_{u,v}, d_{u,y} \leq \rho + \lambda$, and

(b) for at least one shortest path $P$ (respectively, for every shortest path $P$) between $v$ and $y$, $P$ contains a node $z$ such that $d_{u,z} \leq \rho$.

Let $(v, y) \prec_{\rho, \lambda} u$ (respectively, $(v, y) \prec_{\rho, \lambda}^\ast u$)

$$= \begin{cases} 1, & \text{if } v, y \text{ is weakly (respectively, strongly) } (\rho, \lambda)-\text{dominated by } u \\ 0, & \text{otherwise} \end{cases}$$

Since $B_1(u) = \bigcup_{0 < j \leq 1} B_j(u) = \text{Nbr}(u)$, it follows that

$$M_u = \left| \bigcup_{0 < j \leq 1} B_j(u) \right| - \frac{\sum_{v,y \in \bigcup_{0 < j \leq 1} B_j(u)} 1 - (v, y) \prec_{\rho, \lambda} u}{\left| \bigcup_{0 < j \leq 1} B_j(u) \right|}$$

A mathematical justification for the claim (\ast\ast\ast) is provided by Lemma 12 in Section D of the appendix.

An implication of (\ast\ast\ast)
We computed the number $n_1$ of all pairs of nodes from $B_{\xi}(u)$ that are weakly $(\rho, \lambda)$-dominated by $u$.

We computed the number $n_2$ of all pairs of nodes from $B_{\xi}(u)$ that are strongly $(\rho, \lambda)$-dominated by $u$.

Table X tabulates the ratio $\nu = n_2/n_1$, and shows that a large percentage of the pair of nodes that were weakly dominated were also strongly dominated by $u$.

### IV. CONCLUSION

In this paper we demonstrated a number of interesting properties of the shortest and approximately shortest paths in hyperbolic networks. We established the relevance of these results in the context of biological and social networks by empirically finding that a variety of such networks have close-to-tree-like topologies. Our results have important implications to a general class of directed networks which we refer to as regulatory networks. For example, our results imply that cross-talk edges or paths are frequent in these networks. Based on our theoretical results we proposed methodologies to determine relevant paths between a source and a target node in a signal transduction network, and to identify the most important nodes that mediate these paths. Our investigation shows that the hyperbolicity measure captures non-trivial topological properties that is not fully reflected in other network measures, and therefore the hyperbolicity measure should be more widely used.

### Appendix A: Theorem 1

**Theorem 1** Suppose that $G$ has a cycle of $k \geq 4$ nodes which has no path-chord. Then, $\delta^*_\text{wors}(G) \geq \lfloor \sqrt[k]{k} \rfloor$.

**Proof.** In our proofs we will use the consequences of the 4-node condition when the 4 nodes are chosen in a specific manner as stated below in Lemma [2].

### Empirical verification of (***)

We empirically investigated the claim in (***), for the following three social networks from Table II:

**Network 1:** Dolphin social network,

**Network 4:** Books about US politics, and

**Network 7:** Visiting ties in San Juan.

For each network we selected a (central) node $u$ such that there are sufficiently many nodes in the boundary of the $\xi$-neighborhood $B_{\xi}(u)$ of $u$ for an appropriate $\xi = \rho + \lambda$. We then set $\lambda$ to a very small value of 1, and calculated the following quantities.

![FIG. 8. For hyperbolic graphs, the further we move from the central (black) node, the more a shortest path bends inward towards the central node.](image)

![FIG. 9. Case 1 of Theorem 3 $v = u_{0,1}, v' = u_{1,2}$.](image)
Lemma 2 Let \( u_0, u_1, u_2, u_3 \) be four nodes such that \( u_3 \) is on a shortest path between \( u_1 \) and \( u_2 \). Suppose also that all the inter-node distances are strictly positive except for \( d_{u_1, u_3} \) and 
\[
\frac{d_{u_0, u_1} + d_{u_0, u_3} + d_{u_1, u_3}}{2} \leq d_{u_0, u_1} + d_{u_1, u_3}.
\]
Then,
\[
\frac{d_{u_0, u_1} + d_{u_0, u_3} + d_{u_1, u_3}}{2} \leq \frac{d_{u_0, u_1} + d_{u_0, u_3} + d_{u_1, u_3}}{2} + 2 \delta^+_{u_0, u_1, u_3, u_3}.
\]

Proof. Note that due to triangle inequality \( 0 \leq \frac{d_{u_0, u_1} + d_{u_0, u_3} - d_{u_0, u_1}}{2} \leq d_{u_1, u_2} \) and thus node \( u_3 \) always exists.

First, consider the case when \( 0 < d_{u_1, u_3} < d_{u_1, u_2} \). Consider the three quantities involved in the 4-node condition for the nodes \( u_0, u_1, u_2, u_3 \), namely the quantities \( d_{u_0, u_1} + d_{u_0, u_2} + d_{u_1, u_2} + d_{u_1, u_3} \) and \( \delta^+_{u_0, u_1, u_3, u_3} \). Note that
\[
2 (d_{u_0, u_1} + d_{u_1, u_2}) = (d_{u_0, u_1} + d_{u_0, u_2} + d_{u_1, u_2} + d_{u_1, u_3}) + d_{u_2, u_1} \\
\geq d_{u_0, u_1} + d_{u_0, u_2} + d_{u_1, u_2} + d_{u_1, u_3} \\
\Rightarrow d_{u_0, u_1} + d_{u_1, u_3} \geq \frac{d_{u_0, u_1} + d_{u_0, u_2} + d_{u_1, u_2}}{2}.
\]

Thus, \( d_{u_0, u_1} + d_{u_1, u_3} \geq \max \{ d_{u_0, u_1} + d_{u_1, u_3}, d_{u_0, u_1} + d_{u_2, u_3} \} \) and using the definition of \( \delta^+_{u_0, u_1, u_3, u_3} \) we have
\[
\frac{d_{u_0, u_1} + d_{u_0, u_3} + d_{u_1, u_3}}{2} \leq \frac{d_{u_0, u_1} + d_{u_0, u_3} + d_{u_1, u_3}}{2} + 2 \delta^+_{u_0, u_1, u_3, u_3}.
\]

Next, consider the case when \( d_{u_1, u_3} = 0 \). This implies
\[
d_{u_0, u_1} + d_{u_1, u_3} = d_{u_0, u_1} + d_{u_1, u_3} = \frac{d_{u_0, u_1} + d_{u_1, u_2}}{2} \\
\frac{d_{u_0, u_1} + d_{u_0, u_3} + d_{u_1, u_3}}{2} \leq \frac{d_{u_0, u_1} + d_{u_0, u_3} + d_{u_1, u_3}}{2}.
\]

Finally, consider the case when \( d_{u_1, u_3} = d_{u_1, u_2} \). This implies
\[
d_{u_0, u_1} - \frac{d_{u_1, u_3} + d_{u_1, u_2} - d_{u_0, u_2}}{2} < 1 \\
\Rightarrow d_{u_0, u_1} + d_{u_1, u_2} = d_{u_0, u_1} + 2 - 2 \varepsilon \text{ for some } 0 < \varepsilon \leq 1
\]

Thus, it easily follows that
\[
d_{u_0, u_1} + d_{u_1, u_2} = d_{u_0, u_2} + d_{u_1, u_3} + d_{u_0, u_1} + 2 - 2 \varepsilon \]
\[
= \frac{d_{u_0, u_1} + d_{u_1, u_2} + d_{u_0, u_1}}{2} + 1 - \varepsilon
\]
\[
\Rightarrow d_{u_0, u_1} + d_{u_1, u_2} \leq \frac{d_{u_0, u_1} + d_{u_0, u_3} + d_{u_1, u_3}}{2}.
\]

We can now prove Theorem 1 as follows. Let \( C = (u_0, u_1, \ldots, u_{k-1}, u_0) \) be the cycle of \( k = 4r + r' \) nodes for some integers \( r \) and \( 0 \leq r' < 4 \). Consider the four nodes \( u_0, u_{r' + \lceil r'/z \rceil}, u_{2r' + \lceil (r' + r')/z \rceil} \) and \( u_{3r' + r'} \). Since \( C \) has no path-chord, we have
\[
d_{u_0, u_{r' + \lceil r'/z \rceil}} = r + \lceil r'/z \rceil, \quad d_{u_0, u_{2r' + \lceil (r' + r')/z \rceil}} = 2r + \lceil (r' + r')/z \rceil, \quad d_{u_0, u_{3r' + r'}} = r \text{ and } u_{2r' + \lceil (r' + r')/z \rceil} \text{ is on a shortest path between } u_0 \text{ and } u_{3r' + r'}.
\]

Thus, applying the bound of Lemma 2 we get
\[
\delta^\ast_{\text{worst}}(G) \geq \delta^\ast_{u_0, u_{r' + \lceil r'/z \rceil}, u_{2r' + \lceil (r' + r')/z \rceil}} - 1
\]
\[
\geq \frac{d_{u_0, u_{r' + \lceil r'/z \rceil}} + d_{u_{2r' + \lceil (r' + r')/z \rceil} - d_{u_{r' + \lceil r'/z \rceil}}}}{2}
\]
\[
\geq \frac{4r + \lceil (r' + r')/z \rceil}{} - \lceil (r' + r')/z \rceil = \frac{4r + r'}{2} = \frac{r' + r'}{2} + \frac{r' + r'}{2}
\]
\[
\geq r - \frac{1}{4} \Rightarrow \delta^\ast_{\text{worst}}(G) \geq r = \lceil 4/r \rceil
\]

Appendix B: Theorem 3 and Corollary 4

The Gromov product nodes \( u_{0,1}, u_{0,2}, u_{1,2} \) of a shortest-path triangle \( \Delta_{u_{0,1}, u_{1,2}} \) are three nodes satisfying the following:

- \( u_{0,1}, u_{0,2} \) and \( u_{1,2} \) are located on the paths \( \mathcal{P}_u(u_0, u_1) \), \( \mathcal{P}_u(u_0, u_2) \) and \( \mathcal{P}_u(u_1, u_2) \), respectively, and
- the distances of these three nodes from \( u_0, u_1 \) and \( u_2 \) satisfy the following constraints:
\[
d_{u_{0,1}, u_{0,2}} + d_{u_{1,2}, u_{0,1}} = d_{u_{1,2}, u_{1,2}} + d_{u_{0,1}, u_{0,1}} = d_{u_{0,2}, u_{0,1}} + d_{u_{0,2}, u_{1,2}}
\]

To simplify exposition, we assume that \( d_{u_{0,1}, u_{1,2}} + d_{u_{0,2}, u_{1,2}} + d_{u_{0,1}, u_{0,1}} \) is an even number. Otherwise, the definition will require minor changes.
It is not difficult to see that a set of such three nodes always exists. For convenience, the nodes \( u_{1,0}, u_{2,0} \) and \( u_{2,1} \) are assumed to be the same as the nodes \( u_{0,1}, u_{0,2} \) and \( u_{1,2} \), respectively.

**Theorem 3 (see Fig. 10 for a visual illustration)** For a shortest-path triangle \( \Delta(u_{0,1}, u_{0,2}) \) and for \( 0 \leq i \leq 2 \), let \( v \) and \( v' \) be two nodes on the paths \( u_i \) and \( u_{i+2} \) (mod 3), respectively, such that \( d_{u_i,v} = d_{u_i,v'} \). Then,

\[
d_{v,v'} \leq 6 d_{u_i,v}^* + 2
\]

where \( d_{u_i,v}^* \leq \delta_{\text{worst}}(G) \) is the largest worst-case hyperbolicity among all combinations of four nodes in the three shortest paths defining the triangle.

**Corollary 4  (Hausdorff distance between shortest paths)** Suppose that \( \mathcal{P}_1 \) and \( \mathcal{P}_2 \) are two shortest paths between two nodes \( u_0 \) and \( u_1 \). Then, the Hausdorff distance \( d_H(\mathcal{P}_1, \mathcal{P}_2) \) between these two paths can be bounded as:

\[
d_H(\mathcal{P}_1, \mathcal{P}_2) \leq \max \left\{ \min_{v_1 \in \mathcal{P}_1} d_{v_1,v_2}, \min_{v_2 \in \mathcal{P}_2} d_{v_1,v_2} \right\} \leq 6 d_{u_0,v}^* + 2
\]

where \( u_2 \) is any node on the path \( \mathcal{P}_2 \).

**Proof of Theorem 3** To simplify exposition, we assume that \( d_{u_0,v} + d_{u_1,v} + d_{u_2,v} \) is even and prove a slightly improve bound of \( d_{v,v'} \leq 6 d_{u_i,v}^* + 1 \). It is easy to modify the proof to show that \( d_{v,v'} \leq 6 d_{u_i,v}^* + 2 \) if \( d_{u_0,v} + d_{u_1,v} + d_{u_2,v} \) is odd.

We will prove the result for \( i = 1 \) only; similar arguments will hold for \( i = 0 \) and \( i = 2 \). If \( d_{u_1,v} = 0 \) then \( v = v' = u_1 \) and the claim holds trivially. Thus, we assume that \( d_{u_1,v} > 0 \).

**Case 1:** \( v = u_{0,1} \) and \( v' = u_{1,2} \). In this case we need to prove that \( d_{u_{0,1},u_{1,2}} \leq 6 d_{u_1,v}^* + 1 \) (see Fig. 9). Assume that

\[
d_{u_{0,1},u_{1,2}} > 0 \text{ since otherwise the claim is trivially true. Using Lemma 2 for the four nodes } u_0, u_1, u_2, u_{1,2}, \text{ we get}
\]

\[
d_{u_{0,1},u_{1,2}} + d_{u_1,u_2} \leq \left[ \frac{d_{u_0,u_1} + d_{u_1,u_2} + d_{u_0,u_2}}{2} \right] + 2 \delta_{u_{0,1},u_{1,2}} \tag{B1}
\]

Now, we note that

\[
d_{u_{0,1},u_{1,2}} + d_{u_0,u_0} = d_{u_0,u_2} + \left[ \frac{d_{u_0,u_1} + d_{u_1,u_2} - d_{u_0,u_2}}{2} \right] = \left[ \frac{d_{u_0,u_1} + d_{u_1,u_2} + d_{u_1,u_2}}{2} \right] \tag{B2}
\]

which in turn implies

\[
\left| d_{u_{0,1},u_{1,2}} - d_{u_0,u_2} \right| = \left| \left( d_{u_0,u_1} + d_{u_1,u_2} \right) - \left( d_{u_0,u_1} + d_{u_1,u_2} \right) \right| \leq \left[ \frac{d_{u_0,u_1} + d_{u_1,u_2} + d_{u_1,u_2}}{2} \right] + 2 \delta_{u_{0,1},u_{1,2}} \tag{by inequality (B1)}
\]

\[
\left( \text{by equality (B2)} \right)
\]

\[
\leq 2 d_{u_{0,1},u_{1,2}} + 1 \tag{B3}
\]

In a similar manner, we can prove the following analog of inequality (B3):

\[
\left| d_{u_{0,1},u_{1,1}} - d_{u_0,u_2} \right| \leq 2 \delta_{u_{0,1},u_{1,2},u_{1,1}} \tag{B4}
\]

Using inequalities (B3) and (B4), it follows that

\[
\left( d_{u_{0,1},u_{1,2}} + d_{u_0,u_2} \right) \left( d_{u_{0,1},u_{1,1}} + d_{u_0,u_2} \right) \left( d_{u_{0,1},u_{1,2}} + d_{u_0,u_2} \right) \leq 2 \delta_{u_{0,1},u_{1,2}} + 2 \delta_{u_{0,1},u_{1,2},u_{1,1}} + 1 \tag{B5}
\]

Now, consider the three quantities involved in the 4-node condition for the nodes \( u_0, u_2, u_{0,1}, u_{1,2} \), namely, the quantities: \( d_{u_0,u_2} + d_{u_0,u_{0,1}} + d_{u_{0,1},u_2} \) and \( d_{u_0,u_{0,1}} + d_{u_{0,1},u_2} \). Note that

\[
d_{u_0,u_{0,1}} + d_{u_{0,1},u_2} = d_{u_0,u_{0,1}} + d_{u_{0,1},u_2} = d_{u_0,u_{0,1}} + d_{u_{0,1},u_2} \tag{B6}
\]

If \( d_{u_0,u_{0,1}} + d_{u_{0,1},u_2} \leq d_{u_0,u_{0,1}} + d_{u_{0,1},u_2} \) then by the definition of \( \delta_{u_{0,1},u_{0,1},u_{1,2}} \) we have

\[
d_{u_{0,1},u_{1,2}} = \left( d_{u_{0,1},u_{1,2}} + d_{u_{0,1},u_{1,2}} \right) - d_{u_0,u_2}
\]

\[
= \left( d_{u_{0,1},u_{1,2}} + d_{u_{0,1},u_{1,2}} \right) - \left( d_{u_0,u_{0,1}} + d_{u_{0,1},u_2} \right) \leq 2 \delta_{u_{0,1},u_{0,1},u_{1,2}} \tag{otherwise}
\]

Otherwise, \( d_{u_{0,1},u_{1,2}} > d_{u_{0,1},u_{1,2}} \) and then again by the definition of \( 2 \delta_{u_{0,1},u_{0,1},u_{1,2}} \) we have

\[
\left| d_{u_0,u_{0,1}} + d_{u_{0,1},u_2} - d_{u_0,u_{0,1}} - d_{u_{0,1},u_2} \right| \leq 2 \delta_{u_{0,1},u_{0,1},u_{1,2}}
\]
Thus, the ordered sequence of nodes without skipping over any node. Since $v \neq u_{01}, v' \neq u_{12}$.

and now using inequality (55) gives

$$d_{u_{01}, u_{12}} = (d_{u_{01}, u_{12}} + d_{u_{12}, u_{01}}) - (d_{u_{01}, u_{12}} + d_{u_{12}, u_{01}})$$

$$\leq |d_{u_{01}, u_{12}} + d_{u_{12}, u_{01}} - d_{u_{01}, u_{12}}| + |d_{u_{01}, u_{12}} - d_{u_{01}, u_{12}}| + 1 \leq 6 \delta_{u_{01}, u_{12}}^+ + 1$$

**Case 2:** $v \neq u_{01}$ and $v' \neq u_{12}$. The claim trivially holds if $d_{v, v'} \leq 1$, thus we assume that $d_{v, v'} > 1$. Let $(v_1 = u_1, v_2 = u_3, v_3, \ldots, v_h = v', \ldots, v_s = u_1, \ldots, v_r = u_2)$ be the ordered sequence of nodes in the given shortest path from $u_1$ to $u_2$ (see Fig. 11). Consider the sequence of shortest-path triangles $\Delta(u_0, u_1, v_1), \Delta(u_0, u_1, v_2), \ldots, \Delta(u_0, u_1, v_r)$ where each such triangle is obtained by taking the shortest path $P_\Delta(u_0, u_1)$, the sub-path $P_\Delta(u_1, v_j)$ of the shortest path $P_\Delta(u_1, v_j)$, from $u_1$ to $v_j$, and a shortest path $u_0 \leftrightarrow v_j$ from $u_0$ to $v_j$. Let $v_{1,j}$ be the Gromov product node on the side (shortest path) $P_\Delta(u_1, v_j)$ of the shortest-path triangle $\Delta(u_0, u_{12}, v_j)$.

We claim that if $v_{1, j} = v_{p, q}$ and $v_{1, j+1} = v_{p, q}$, then $q$ is either $p$ or $p+1$. Indeed, if $d_{v, v'} = \left\lceil \frac{d_{u_{01}, u_{12}} + d_{u_{12}, u_{01}} - d_{u_{01}, u_{12}}}{2} \right\rceil$ then

$$d_{v, v'} - d_{u_{01}, u_{12}} + d_{u_{12}, u_{01}} = \left\lceil \frac{d_{u_{01}, u_{12}} + d_{u_{12}, u_{01}} - d_{u_{01}, u_{12}}}{2} \right\rceil = \left\lceil \frac{d_{u_{01}, u_{12}} + d_{u_{12}, u_{01}} - d_{u_{01}, u_{12}}}{2} \right\rceil + 1 = 1$$

and a similar proof of $d_{v, v'} - d_{u_{01}, u_{12}} + d_{u_{12}, u_{01}} = \left\lceil \frac{d_{u_{01}, u_{12}} + d_{u_{12}, u_{01}} - d_{u_{01}, u_{12}}}{2} \right\rceil$ can be obtained if $d_{u_{01}, u_{12}} = \left\lceil \frac{d_{u_{01}, u_{12}} + d_{u_{12}, u_{01}} - d_{u_{01}, u_{12}}}{2} \right\rceil$ and $d_{u_{12}, u_{01}} = \left\lceil \frac{d_{u_{01}, u_{12}} + d_{u_{12}, u_{01}} - d_{u_{01}, u_{12}}}{2} \right\rceil$.

Thus, the ordered sequence of nodes $v_{1,1}, v_{1,2}, \ldots, v_{1, s}$ cover the ordered sequence of nodes $v_2, v_3, \ldots, v_r$ in a consecutive manner without skipping over any node. Since $v_{1,1}$ is either $v_1$ or $v_2$, and $v_{1, r} = v_r = u_{12}$, there must be an index $t$ such that $v_{1, t} = v = v_r$. Since $d_{v_1, v} = d_{v, v'}$, $v$ and $v'$ are the two Gromov product nodes for the shortest-path triangle $\Delta(u_0, u_{12}, v)$ and thus applying Case 1.1 on $\Delta(u_0, u_{12}, v)$, we have $d_{v, v'} \leq 6 \delta_{u_{01}, u_{12}}^+ + 1$.

**Appendix C: Theorem and Corollary**

**Theorem 5** (see Fig. 12 for a visual illustration) Let $P_1 \equiv u_0 \leftrightarrow u_1$ and $P_2$ be a shortest path and an arbitrary path, respectively, between two nodes $u_0$ and $u_1$. Then, for every node $v$ on $P_1$, there exists a node $v'$ on $P_2$ such that

$$d_{v, v'} \leq \min \left\{ \left( 6 \delta_{\text{worst}}^+(G) + 2 \right), \left( \left\lceil \frac{\log_2 \ell (P_2) \right\rceil - 1 \right), \left\lceil \frac{d_{u_0, u_1}}{2} \right\rceil \right\}$$

Since $\ell (P_2) \leq n$, the above bound also implies that

$$d_{v, v'} \leq \left( 6 \delta_{\text{worst}}^+(G) + 2 \right) \left( \left\lceil \frac{\log_2 n \right\rceil - 1 \right) = O \left( \delta_{\text{worst}}^+(G) \log n \right)$$

**Corollary 6** Suppose that there exists a node $v$ on the shortest path between $u_0$ and $u_1$ such that $\min_{v' \in P_2} d_{v, v'} \geq \gamma$. Then, $\ell (P_2) \geq 2^{\delta_{\text{worst}}^+(G)+2} + 1 - 1 = \Omega (2^\gamma / \delta_{\text{worst}}^+(G))$.

**Proof of Theorem 5** First, note that by selecting $v'$ to be one of $u_0$ or $u_1$ appropriately we have $d_{v, v'} \leq \left\lceil \frac{d_{u_0, u_1}}{2} \right\rceil$. Now, assume that $\ell (P_2) > 2$. Let $u_2$ be the node on the path $P_2$ such that $\ell (u_0 \leftrightarrow u_2) = \left\lceil \frac{\ell (P_2)}{2} \right\rceil$, and consider the shortest-path triangle $\Delta(u_0, u_{12}, u_2)$. By Theorem 3 there exists a node $v'$ either on a shortest path between $u_0$ and $u_2$ or on a shortest path between $u_1$ and $u_2$ such that $d_{v, v'} \leq 6 \delta_{\text{worst}}^+(G) + 2$. We move from $v$ to $v'$ and recursively solve the problem of finding a shortest path from $v'$ to a node on a part of the path $P_2$ containing at most $\left\lceil \frac{\ell (P_2)}{2} \right\rceil$ edges. Let $D(y)$ denote the minimum distance from $v$ to a node in a path of length $y$ between $u_0$ and $u_1$. Thus, the worst-case recurrence for $D(y)$ is given by

$$D(y) \leq D \left( \left\lceil \frac{y}{2} \right\rceil \right) + 6 \delta_{\text{worst}}^+(G) + 2, \quad \text{if } y > 2$$

$$D(2) = 1$$

A solution to the above recurrence satisfies $D(\ell (P_2)) \leq \left( 6 \delta_{\text{worst}}^+(G) + 2 \right) \left( \left\lceil \frac{\log_2 \ell (P_2) \right\rceil - 1 \right)$. \(\square\)
Appendix D: Theorem 7 and Corollary 8

For easy of display of long mathematical equations, we will denote $\delta^{+}_{\text{word}}(G)$ simply as $\delta^+$. 

**Theorem 7** Let $P_1$ and $P_2$ be a shortest path and another path, respectively, between two nodes. Define $\eta_{P_1, P_2}$ as

$$
\eta_{P_1, P_2} = \left(6 \delta^+ + 2 \right) \log_2 \left( \left(6 \delta^+ + 2 \right) \log_2 \left(6 \delta^+ + 2 \right) \right) + \mu
$$

Then, the following statements are true.

(a) For every node $v$ on $P_1$, there exists a node $v'$ on $P_2$ such that $d_{v, v'} \leq \left[ \eta_{P_1, P_2} \right]$. 

(b) For every node $v'$ on $P_2$, there exists a node $v$ on $P_1$ such that $d_{v, v'} \leq \left[ \eta_{P_1, P_2} \right]$.

**Corollary 8** (Hausdorff distance between approximate short paths) Suppose that $P_1$ and $P_2$ are two paths between two nodes. Then, the Hausdorff distance $d_H(P_1, P_2)$ between these two paths can be bounded as follows:

$$
d_H(P_1, P_2) \leq \max_{v \in \{v_1 \mid v_1 \in P_1 \}} \{ d_{v, v_1} \}, \quad \max_{v \in \{v_2 \mid v_2 \in P_2 \}} \{ d_{v, v_2} \}
$$

In particular, assuming real world networks have small constant values of $\delta^+$, the asymptotic dependence of $\mu$ and $\epsilon$ on $\gamma$ can be summarized as:

$$
\text{both } \mu \text{ and } \epsilon = \Omega(2^{C\gamma}) \quad \text{for some constant } 0 < C < 1
$$

**Proof of Theorem 7** Let $P_1$ and $P_2$ be a shortest path and another path, respectively, between two nodes $u_0$ and $u_1$. Note that any “sub-path” of a $\mu$-approximate short path is also a $\mu$-approximately short path, i.e., $u_i \leftrightarrow u_j$ is also a $\mu$-approximate short path, and similarly any sub-path of an $\epsilon$-additive-approximate short path is also an $\epsilon$-additive-approximate short path. $\mu$-approximate shortest paths also restrict the “span” of a path-chord of the path, i.e., if $(u_0, u_1, \ldots, u_k)$ is a $\mu$-approximate short path and $\{u_i, u_j\} \in E$ then $|j - i| \leq \mu$.

(a) Let $v$ and $v'$ be two nodes on $P_1$ and $P_2$, respectively, such that $\alpha = d_{v, v'} = \min \{ d_{v, v_0}, d_{v, v''} \}$. Let $v_0 \in u_0 \leftrightarrow v$ and $v_1 \in u_1 \leftrightarrow v$ be two nodes defined by

$$
d_{v, v_0} = \frac{2\alpha + 1}{\epsilon}, \quad \text{if } d_{v, v_0} > 2\alpha + 1
$$

$$
d_{v, v_0} = \frac{2\alpha + 1}{\epsilon}, \quad \text{if } d_{v, v_0} > 2\alpha + 1
$$

By definition of $\alpha$, there exists two nodes $v_{\ell}$ and $v_r$ on the path $P_2$ such that $d_{v_{\ell}, v_r}, d_{v_r, v_{\ell}} \leq \alpha$. Consider the path $P_3 = v_{\ell} \leftrightarrow P_2 \leftrightarrow v_r$ that is the part of path $P_2$ from $v_{\ell}$ to $v_r$. Note that

$$
d_{v_{\ell}, v_r} \leq d_{v_{\ell}, v'} + d_{v', v_r} \leq 6\alpha + 2
$$

Thus, we arrive at the following inequalities

$$
\ell(P_3) \leq \begin{cases} 
6\alpha + 2 \mu, \quad \text{if } P_2 \text{ is } \mu\text{-approximate short} \\
6\alpha + 2 + \epsilon, \quad \text{if } P_2 \text{ is } \epsilon\text{-additive-approximate short} 
\end{cases}
$$

Now consider the path $P_4 = v_{\ell} \leftrightarrow v_r \leftrightarrow P_3 \leftrightarrow v_r \leftrightarrow v_{\ell}$ obtained by taking a shortest path from $v_{\ell}$ to $v_r$ followed by the path $P_3$ followed by a shortest path from $v_r$ to $v_{\ell}$. Note that

$$
\ell(P_4) \leq \begin{cases} 
6\alpha + 2 + \epsilon + 2\alpha = 8\alpha + 2 + \epsilon, \quad \text{if } P_2 \text{ is } \epsilon\text{-additive-approximate short} 
\end{cases}
$$

We claim that $\min_{i \in P_2} \{ d_{v_i, v_i} \} = \alpha$. Indeed, if $v_i \in P_3$ then, by definition of $\alpha$, $\min_{i \in P_3} \{ d_{v_i, v_i} \} = \alpha$. Otherwise, if $v_i \in v_{\ell} \leftrightarrow v_r$, then by triangle inequality $d_{v_i, v_i} = d_{v_{\ell}, v_i} + d_{v_i, v_r} = d_{v_i, v'} + d_{v, v_r} \geq 2\alpha + 1 - d_{v, v'} > \alpha$. Similarly, if $v_i \in v_r \leftrightarrow v_{\ell}$, then by triangle inequality $d_{v_i, v_i} = d_{v_r, v_i} + d_{v_i, v_{\ell}} = d_{v, v_r} \geq 2\alpha + 1 - d_{v, v'} > \alpha$. Since $v_i \in P_2$ is a shortest path between $v_i$ and $v_i$, and $v$ is a node on this path, by Theorem 5 $\alpha \leq (6\delta^+ + 2) \{ \log_2 (\ell(P_3)) \} - 1$. Thus,
we have the following inequalities:

- If $P_2$ is a $\mu$-approximate short path then
  \[
  \ell(P_2) \leq (6\alpha + 2)\mu + 2\alpha \\
  \leq (6\mu + 2)\alpha + 2\mu \\
  \leq (6\mu + 2)\alpha + 2\mu \\
  \leq (6\mu + 2)\alpha + 2\mu + 2\alpha \\
  \Rightarrow \alpha \leq (6\delta + 2)(\log_2((3\mu + 1)\mu + \mu))
  \]

(D1)

- If $P_2$ is a $\varepsilon$-additive-approximate short path then
  \[
  \ell(P_2) \leq 8\alpha + 2 + \varepsilon \\
  \leq 8(6\delta^+ + 2)(\log_2\ell(P_2) - 1) + 2 + \varepsilon \\
  \leq 8(6\delta^+ + 2)(\log_2(8\alpha + 2 + \varepsilon)) + 1 + 2 + \varepsilon \\
  \Rightarrow \alpha \leq (6\delta + 2)(\log_2(4\alpha + 1 + \frac{\varepsilon}{2}))
  \]

(D2)

Both (D1) and (D2) are of the form $\alpha \leq a\log_2(b\alpha + c) \equiv 2^{\gamma_\alpha} \leq b\alpha + c$ where

\[
a = 6\delta^+ + 2 \geq 1 \text{ for both (D1) and (D2)}
\]

\[
b = \begin{cases} 
3\mu + 1 & \text{for (D1)} \\
4 & \text{for (D2)}
\end{cases}
\]

\[
c = \begin{cases} 
\mu & \text{for (D1)} \\
1 + \frac{\varepsilon}{4} & \text{for (D2)}
\end{cases}
\]

Thus, $\alpha$ is at most $z_0$ where $z_0$ is the largest positive integer value of $z$ that satisfies the equation:

\[
2^{\gamma_\alpha} \leq b\alpha + c
\]

In the sequel, we will use the fact that $\log_2(x + y + 1) \geq \log_2(x + y)$ for $x, y \geq 1$. This holds since

\[
x \geq 1 \text{ and } y \geq 1 \Rightarrow y(x - 1) \geq x - 1 \Rightarrow xy + 1 \geq x + y
\]

We claim that $z_0 \leq \eta = a\log_2(2a b\log_2(a b c) + c)$. This is verified by showing that $2^{\gamma_{\eta}} \geq b\eta + c$ as follows:

\[
2^{\gamma_{\eta}} = 2^{\log_2(2ab\log_2(ab c + c))} \geq 2a b\log_2(ab c + c) + c
\]

\[
b\eta + c = a b\log_2(2a b\log_2(ab c + c)) + c
\]

\[
2^{\gamma_{\eta}} > b\eta + c
\]

\[
\Rightarrow 2a b\log_2(ab c + c) \geq a b\log_2(2a b\log_2(ab c + c)) + c
\]

\[
\Rightarrow 2\log_2(ab c + c) \geq \log_2(2a b\log_2(ab c + c)) + c
\]

\[
\Rightarrow 2\log_2(ab c + c) \geq \log_2(2a b\log_2(ab c + c + 1))
\]

\[
\Rightarrow 2a b\log_2(ab c + c) \geq 2a b\log_2(ab c + c + 1)
\]

\[
\Rightarrow ab c \geq \log_2(ab c + c + 1)
\]

and the very last inequality holds since $abc \geq 4$. Thus, we arrive at the at the following bounds:

- If $P_2$ is a $\mu$-approximate short path then
  \[
  \eta = (6\delta^+ + 2)(\log_2((6\mu + 2)(6\delta^+ + 2)(3\mu + 1)\mu + \mu))
  \]

- If $P_2$ is a $\varepsilon$-additive-approximate short path then
  \[
  \eta = (6\delta^+ + 2)(\log_2((6\mu + 2)(6\delta^+ + 2)(3\mu + 1)\mu + \mu))
  \]

(b) Let the ordered sequence of nodes in the path $P_1 = v_1 \xrightarrow{\mu_1} v_2 \cdots \xrightarrow{\mu_k} v_k$ be a (length) maximal sequence of nodes such that:

\[
\forall v' \in P_3 : \min_{v \in P_1} \{d_{v,v'}\} > Z_{P_1,P_2}
\]

Consider the following set of nodes belonging to the two paths $u_0 \xrightarrow{\mu_1} v_1$ and $v_1 \xrightarrow{\mu_2} u_1$:

\[
S_1 = \bigcup \{v' \in u_0 \xrightarrow{\mu_1} v_1 | \exists v \in P_1 : d_{v,v} = \min_{v \in P_1} \{d_{v,v'}\}\}
\]

\[
S_2 = \bigcup \{v' \in v_1 \xrightarrow{\mu_2} u_1 | \exists v \in P_1 : d_{v,v} = \min_{v \in P_1} \{d_{v,v'}\}\}
\]

Since $u_0 \in S_1$ and $u_1 \in S_2$, it follows that $S_1 \neq \emptyset$ and $S_2 \neq \emptyset$. Note that

\[
\bigcup \{v \in u_0 \xrightarrow{\mu_1} u_1 | \exists v \in S_1 \cup S_2 : d_{v,v} = \min_{v \in P_1} \{d_{v,v'}\}\} = \bigcup_{v \in u_0 \xrightarrow{\mu_1} u_1} \{v\}
\]

Thus, there exists two adjacent nodes $v_4$ and $v_4'$ on $P_1$ such that both $d_{v_4,v_3}$ and $d_{v_4',v_3}$ is at most $Z_{P_1,P_2}$. Using triangle inequality it follows that

\[
d_{v_4,v_3} \leq d_{v_4,v_4'} + Z_{P_1,P_2} + 1
\]

and the very last inequality holds since $abc \geq 4$. Thus, we arrive at the at the following bounds:

- If $P_2$ is a $\mu$-approximate short path then
  \[
  \eta = (6\delta^+ + 2)(\log_2((6\mu + 2)(6\delta^+ + 2)(3\mu + 1)\mu + \mu))
  \]

- If $P_2$ is a $\varepsilon$-additive-approximate short path then
  \[
  \eta = (6\delta^+ + 2)(\log_2((6\mu + 2)(6\delta^+ + 2)(3\mu + 1)\mu + \mu))
  \]

For any node $v'$ on $P_2$, we can always use the following path to reach a node on $P_1$:

- if $d_{v',v_1} \leq d_{v_4,v_3}$ then we take the path $v' \xrightarrow{\mu_2} v_3 \xrightarrow{\mu_1} v_4$ of length at most $\left[\ell(P_2,v_3,v_4)\right] + Z_{P_1,P_2}$ to reach the node $v = v_4$ on $P_1$;

- otherwise we take the path $v' \xrightarrow{\mu_2} v_3 \xrightarrow{\mu_1} v_4$ of length at most $\left[\ell(P_2,v_3,v_4)\right] + Z_{P_1,P_2}$ to reach the node $v = v_4$ on $P_1$.

This gives the following worst-case bounds for $d_{v,v'}$:

\[
d_{v,v'} \leq \left\lfloor \frac{(\mu + 1)Z_{P_1,P_2} + \frac{\varepsilon}{4}}{2} \right\rfloor, \text{ if } P_2 \text{ is } \mu\text{-approximate short}
\]

\[
d_{v,v'} \leq \left\lfloor \frac{2Z_{P_1,P_2} + \frac{1+\varepsilon}{4}}{2} \right\rfloor, \text{ if } P_2 \text{ is } \varepsilon\text{-additive-approximate short}
\]
that we are given the following:

(iii): if $Q$ is a $\varepsilon$-additive-approximate short path then

$$\varepsilon > \frac{2\delta_{\text{worst}}^\alpha(G) + \frac{2}{\beta} - \frac{1}{\beta}}{48 \delta_{\text{worst}}(G) + \frac{12}{\beta}} - \log_2(48 \delta_{\text{worst}}(G) + 16)$$

In particular, if $\delta_{\text{worst}}(G)$ is a constant then $\varepsilon = \Omega\left(2\Theta(\varepsilon + \kappa)\right)$ and thus $\varepsilon$ increases at least exponentially with both $\alpha$ and $\kappa$.

1. Proof of Theorem 10

Consider the shortest-path triangle $\Delta_{u_0, u_3, u_4}$ and let $u_0, u_3, u_4$ be the Gromov product nodes of $\Delta_{u_0, u_3, u_4}$ on the sides (shortest paths) $u_0$ to $u_3$, $u_0$ to $u_4$ and $u_3$ to $u_4$, respectively. Thus, $d_{u_0, u_3} = d_{u_0, u_4}$, and $d_{u_0, u_4} = \left\lfloor \frac{d_{u_0, u_3} + d_{u_0, u_4} - d_{u_3, u_4}}{2} \right\rfloor$ since $d_{u_0, u_3} = d_{u_0, u_4} = r + \alpha$.

We first claim that $d_{u_0, u_3} < r = d_{u_0, u_4}$. Suppose for the sake of contradiction that $d_{u_0, u_3} = d_{u_0, u_4} \geq r$. Then, by Theorem 3 we get $d_{u_0, u_3} \leq 6 \delta_{\text{worst}}^\alpha(G) + 2$ which contradicts the assumption that $d_{u_0, u_3} = \frac{r}{2} (6 \delta_{\text{worst}}^\alpha(G) + 2)$ since $\kappa \geq 4$.

Thus, assume that $d_{u_0, u_3} = d_{u_0, u_4} = r - x$ for some integer $x > 0$. By Theorem 3 $d_{u_0, u_3, u_4} \leq 6 \delta_{\text{worst}}(G) + 2$. Let $d_{u_0, u_3, u_4} = 6 \delta_{\text{worst}}(G) + 2 - y$ for some integer $0 < y \leq 6 \delta_{\text{worst}}(G) + 2$ and $d_{u_0, u_3} = \frac{r}{2} (6 \delta_{\text{worst}}^\alpha(G) + 2) + z$ for some integer $z \geq 0$. Consider the 4-node condition for the four nodes $u_1, u_2, u_3, u_4$.

The three relevant quantities for comparison are:

- $q_\| = d_{u_1, u_3} + d_{u_0, u_3, u_4} = \left(\frac{r}{2} + 1\right) (6 \delta_{\text{worst}}^\alpha(G) + 1) + z - y$
- $q_\ast = d_{u_0, u_3, u_4} - d_{u_0, u_3} + d_{u_0, u_4, u_3} = \left(\frac{r}{2} + 1\right) (6 \delta_{\text{worst}}^\alpha(G) + 2) + 2x$
- $q_\ast' = d_{u_0, u_3, u_4} - d_{u_0, u_3} + d_{u_0, u_4, u_3} + d_{u_0, u_4, u_3} = 12 \delta_{\text{worst}}^\alpha(G) + 4 - 2y + 2x$

We now show that $x > \left(\frac{3x - 2}{12}\right) (6 \delta_{\text{worst}}^\alpha(G) + 2)$. We have the following cases.

- Assume that $q_\| \leq \min\{q_\|, q_\ast\}$. This implies

$$\left|q_\| - q_\ast\right| \leq 2 \delta_{\text{worst}}^\alpha(G)$$

$$\Rightarrow \left|\left(\frac{r}{2} + 1\right) (6 \delta_{\text{worst}}^\alpha(G) + 2) + z - y - 2x\right| \leq 2 \delta_{\text{worst}}^\alpha(G)$$

$$\Rightarrow x \geq \left(\frac{3x - 2}{12}\right) (6 \delta_{\text{worst}}^\alpha(G) + 2) + \frac{1}{6}$$

- Otherwise, assume that $q_\ast \leq \min\{q_\|, q_\ast\}$. This implies

$$\left|q_\| - q_\ast\right| \leq 2 \delta_{\text{worst}}^\alpha(G)$$

$$\Rightarrow q_\| \geq q_\| - 2 \delta_{\text{worst}}^\alpha(G)$$

$$\Rightarrow d_{u_0, u_3} + d_{u_0, u_4} \geq \left(\frac{r}{2} + 1\right) (6 \delta_{\text{worst}}^\alpha(G) + 2) + z - y - 2 \delta_{\text{worst}}^\alpha(G)$$

$$\Rightarrow \left(d_{u_0, u_3, u_4} + d_{u_0, u_3, u_4} + d_{u_0, u_3, u_4}\right) \geq d_{u_0, u_3} + d_{u_0, u_4}$$

$$\Rightarrow 2x + 2 \left(6 \delta_{\text{worst}}^\alpha(G) + 2\right) \geq \left(\frac{3x - 2}{12}\right) (6 \delta_{\text{worst}}^\alpha(G) + 2) + z - y - 2 \delta_{\text{worst}}^\alpha(G)$$

$$\Rightarrow x \geq \left(\frac{3x - 2}{12}\right) (6 \delta_{\text{worst}}^\alpha(G) + 2) + \frac{1}{6}$$
Otherwise, assume that $q = \min \{ q_\parallel, q_\perp \}$. This implies

$$|q_\parallel - q_\perp| \leq 2 \delta^*_w(G)$$

$$\Leftrightarrow 2x - (d_{u_3,u_4} + d_{u_4,u_2}) \leq 2 \delta^*_w(G)$$

$$\Rightarrow 2x \geq d_{u_3,u_4} + d_{u_4,u_2} - 2 \delta^*_w(G)$$

$$\geq (d_{u_3,u_2} - d_{u_4,u_3}) + (d_{u_3,u_2} - d_{u_4,u_3}) - 2 \delta^*_w(G)$$

$$\Rightarrow 2x \geq \kappa (6 \delta^*_w(G) + 2) + 2x - 2 \delta^*_w(G)$$

$$\Rightarrow x \geq \left( \frac{3\kappa - 2}{12} \right) (6 \delta^*_w(G) + 2) + \frac{\delta^*_w(G)}{2} + \frac{1}{6}$$

Using Theorem 3 it now follows that

$$d_{u_3,u_4} \leq d_{u_3,u_2} + d_{u_2,u_4} \leq (r - x) + (6 \delta^*_w(G) + 2)$$

$$\leq r - \left( \frac{3\kappa - 2}{12} \right) (6 \delta^*_w(G) + 2)$$

This proves part (a) with $u_{3,4}$ being the node in question. To prove part (b), note that

$$|\mathcal{P}| = 2 \beta \geq 2(r + x) - 2d_{u_3,u_4} \geq 2x + \left( \frac{3\kappa - 2}{6} \right) (6 \delta^*_w(G) + 2)$$

2. Proof of Corollary 11

Consider such a path $Q$ and consider the node $u_{3,4}$ on the shortest path between $u_3$ and $u_4$. Since every node of $Q$ is at a distance strictly larger than $r + x$ from $u_0$, by Theorem 10 the following holds for every node $v \in Q$

$$d_{u_3,v} \geq (r + x) - d_{u_0,u_3} = (r + x) - \left( r - \frac{3\kappa - 2}{12} \right) (6 \delta^*_w(G) + 2)$$

$$= \alpha + \left( \frac{3\kappa - 2}{12} \right) (6 \delta^*_w(G) + 2)$$

Thus, by Corollary 6 (with $\gamma = \alpha + \left( \frac{3\kappa - 2}{12} \right) (6 \delta^*_w(G) + 2)$), we get

$$\ell(Q) \geq 2 \delta^*_w(G) + \frac{\alpha}{2} + \frac{5}{6} - \frac{1}{6}$$

If $Q$ is a $\mu$-approximate short path, then by Corollary 9

$$\mu > \frac{2 \delta^*_w(G) + 2}{\gamma} - \frac{1}{3}$$

$$= \frac{2 \delta^*_w(G) + 2}{12 \gamma - \left( 24 + o(1) \right) \left( 6 \delta^*_w(G) + 2 \right) - \frac{1}{3}}$$

$$= \frac{2 \delta^*_w(G) + 2}{12 \alpha + (3 \kappa - 26 - o(1)) \left( 6 \delta^*_w(G) + 2 \right) - \frac{1}{3}}$$

If $Q$ is a $\varepsilon$-additive-approximate short path, then by Corollary 9

$$\varepsilon > \frac{2 \delta^*_w(G) + 2}{48 \delta^*_w(G) + \frac{\alpha}{2} + \frac{5}{6} - \frac{1}{6}} - \log_2 \left( 48 \delta^*_w(G) + 16 \right)$$

$$= \frac{2 \delta^*_w(G) + 2}{48 \delta^*_w(G) + \frac{\alpha}{2} + \frac{5}{6} - \frac{1}{6}} - \log_2 \left( 48 \delta^*_w(G) + 16 \right)$$

Appendix F: Lemma 12 (equivalence of strong and weak domination; see Fig. 7 for a visual illustration) If $\lambda \geq (6 \delta^*_w(G) + 2) \log_2 n$ then

$$\mathbb{E}_{\kappa,\lambda} \equiv \begin{array}{c|c}
\text{number of pairs of nodes } v, y \text{ such that } v, y \text{ is weakly } \\
\text{weakly } (\rho, \lambda)\text{-dominated by } u & \text{v is selected uniformly randomly} \\
\text{from } \cup_{i \geq 1} B_i(u) & \text{from } \cup_{i \geq 1} B_i(u) \end{array}$$

There exists a shortest path $v \leftrightarrow y$ between $v, y \in \mathcal{B}_{\rho,\lambda}(u)$ such that for some node $v' \neq v \leftrightarrow y$ we have $v' \notin \mathcal{B}_{\rho,\lambda}(u)$. Let $v \leftrightarrow y$ be any other path between $v$ and $y$ that does not contain a node from $\mathcal{B}_{\rho,\lambda}(u)$. Then, by Corollary 7 (with $\kappa = 4$) we have

$$\ell(Q) \geq 2 \delta^*_w(G) + \frac{\alpha}{2} + \frac{5}{6} - \frac{1}{6} \geq 2 \log_2 n + \frac{\alpha}{2} + \frac{5}{6} - \frac{1}{6} > n - 1$$

which contradicts the obvious bound $\ell(Q) < n$. Thus, no such path $Q$ exists and $v, y$ is strongly $\lambda$-dominated by $u$. $

\square$

ACKNOWLEDGMENTS

B. DasGupta and N. Mobasher were supported by NSF grants IIS-1160995. R. Albert was supported by NSF grants IIS-1161007 and PHY-1205840.

[1] M. E. J. Newman, Networks: An Introduction (Oxford University Press, 2010).

[2] R. Albert, A.-L. Barabási, Statistical mechanics of complex networks. Reviews in Modern Physics 74, 47-97 (2002).
[40] C. P. Loomis, J. O. Morales, R. A. Clifford, O. E. Leonard, Turrrialba: Social Systems and the Introduction of Change (The Free Press, Glencoe, IL, 1953), p. 45 and 78.

[41] D. E. Knuth, The Stanford GraphBase: A Platform for Combinatorial Computing (Addison-Wesley, Reading, MA, 1993).

[42] E. Jonckheere, P. Lohsoonthorn, F. Ariaei, Scaled Gromov Four-Point Condition for Network Graph Curvature Computation. Internet Mathematics 7(3), 137-177 (2011).

[43] R. Kannan, P. Tetali, S. Vempala, Markov-chain algorithms for generating bipartite graphs and tournaments. Random Structures and Algorithms 14, 293-308 (1999).

[44] B. Alberts, Molecular biology of the cell (New York: Garland Publishers, 1994).

[45] T. I. Lee, N. J. Rinaldi, F. Robert, D. T. Odom, Z. Bar-Joseph, G. K. Gerber, N. M. Hannett, C. T. Harbison, C. M. Thompson, I. Simon, J. Zeilinger, E. G. Jennings, H. L. Murray, D. B. Gordon, B. Ren, J. J. Wyrick, J.-B. Tagne, T. L. Volkert, E. Fraenkel, D. K. Gifford, R. A. Young, Transcriptional regulatory networks in Saccharomyces cerevisiae. Science 298(5594), 799-804 (2002).

[46] M. R. Bridson, A. Haefliger, Metric Spaces of Non-Positive Curvature (Springer, 1999).

[47] R. Albert, Scale-free networks in cell biology, Journal of Cell Science 118, 4947-4957 (2005).

[48] R. S. Burt, Structural Holes: The Social Structure of Competition (Harvard University Press, 1995).

[49] S. P. Borgatti, Structural Holes: Unpacking Burt’s Redundancy Measures. Connections, 20(1), 35-38 (1997).

[50] M. R. Garey, D. S. Johnson, Computers and Intractability – A Guide to the Theory of NP-Completeness (W. H. Freeman & Co., 1979).

[51] P. Gupta, R. Janardan, M. Smid, B. DasGupta, The rectangle enclosure and point-dominance problems revisited. International Journal of Computational Geometry & Applications 7(5), 437-455 (1997).
Supplemental Information
| Name                      | Brief Description                                                                                                                                                                                                 | Nodes | Edges   | Reference |
|---------------------------|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-------|---------|-----------|
| E. coli transcriptional   | Direct regulatory interactions between transcription factors and the genes or operons they regulate                                                                                                               | 311   | 451     | [25]      |
| Mammalian signaling       | Mammalian network of signaling pathways and cellular machines in the hippocampal CA1 neuron                                                                                                                      | 512   | 1047    | [26]      |
| E. coli transcriptional   | Direct regulatory interactions between transcription factors and the genes or operons they regulate                                                                                                             | 418   | 544     | #         |
| T-LGL signaling           | Signaling network inside cytotoxic T cells in the context of the disease T cell large granular lymphocyte leukemia                                                                                              | 58    | 135     | [27]      |
| S. cerevisiae transcriptional | S. cerevisiae transcriptional regulatory network showing interactions between transcription factor proteins and genes                                                                                          | 690   | 1082    | [28]      |
| C. elegans metabolic      | The network of biochemical reactions in C. elegans metabolism                                                                                                                                                | 453   | 2040    | [29]      |
| Drosophila segment polarity (6 cells) | 1-dimensional 6-cell version of the gene regulatory network among products of the segment polarity gene family that plays an important role in the embryonic development of Drosophila melanogaster | 78    | 132     | [30]      |
| ABA signaling             | Guard cell signal transduction network for abscisic acid (ABA) induced stomatal closure in plants                                                                                                                | 55    | 88      | [31]      |
| Immune response network   | Network of interactions among immune cells and pathogens in the mammalian immune response against two bacterial species                                                                                      | 18    | 42      | [32]      |
| T cell receptor signaling | Network for T cell activation mechanisms after engagement of the TCR, the CD4/CD8 co-receptors and CD28.                                                                                                      | 94    | 138     | [33]      |
| Oriented yeast PPI        | An oriented version of an unweighted PPI network constructed from S. cerevisiae interactions in the BioGRID database                                                                                           | 786   | 2445    | [34]      |

# Updated version of the network in [25]; see www.weizmann.ac.il/mcb/UriAlon/Papers/networkMotifs/coli1_1Inter_st.txt
TABLE XII. Details of 9 social networks studied

| name                        | brief description                                                                                           | type                  | # nodes | # edges | reference |
|-----------------------------|-------------------------------------------------------------------------------------------------------------|-----------------------|---------|---------|-----------|
| 1. Dolphin social network   | Social network of frequent associations between 62 dolphins in a community living off Doubtful Sound in New Zealand | undirected, unweighted | 62      | 160     | [35]      |
| 2. American College Football | Network of American football games between Division IA colleges during the regular Fall 2000 season       | undirected, unweighted | 115     | 612     | [36]      |
| 3. Zachary Karate Club      | Network of friendships between 34 members of a karate club at a US university in the 1970s                 | undirected, unweighted | 34      | 78      | [37]      |
| 4. Books about US politics  | Network of books about US politics published around the time of the 2004 presidential election and sold by the online bookseller amazon.com, edges between books represent frequent copurchasing of books by the same buyers.| undirected, unweighted | 105     | 442     | ‡         |
| 5. Sawmill communication network | A communication network within a small enterprise: a sawmill. All employees were asked to indicate the frequency with which they discussed work matters with each of their colleagues on a five-point scale ranging from less than once a week to several times a day. Two employees were linked in the network if they rated their contact as three or more. | undirected, unweighted | 36      | 62      | [38]      |
| 6. Jazz Musician network    | A social network of Jazz musicians                                                                           | undirected, unweighted | 198     | 2742    | [39]      |
| 7. Visiting ties in San Juan | Network for visiting relations between families living in farms in the neighborhood San Juan Sur, Costa Rica, 1948 | undirected, unweighted | 75      | 144     | [40]      |
| 8. World Soccer Data, Paris 1998 | Members of the 22 soccer teams which participated in the World Championship in Paris in 1998 had contracts in 35 countries. Counts of which team exports how many players to which country are used to generate this network. | directed, weighted    | 35      | 118     | †         |
| 9. Les Miserables           | Network of co-appearances of characters in Victor Hugo’s novel “Les Miserables”. Nodes represent characters as indicated by the labels and edges connect any pair of characters that appear in the same chapter of the book. The weights on the edges are the number of such coappearances. | undirected, weighted  | 77      | 251     | [41]      |

‡ V. Krebs, unpublished manuscript, found on Krebs’ website [www.orgnet.com]

† Dagstuhl seminar: *Link Analysis and Visualization*, Dagstuhl 1-6, 2001.
(see [http://vlado.fmf.uni-lj.si/pub/networks/data/sport/football.htm](http://vlado.fmf.uni-lj.si/pub/networks/data/sport/football.htm))
Biological details of source, target and central nodes ($u_{source}$, $u_{target}$ and $u_{central}$) used in Table VIII and Table IX

**NETWORK 1: E. COLI TRANSCRIPTIONAL**

| Node name | Node type | Details |
|-----------|-----------|---------|
| fliAZY    | $u_{source}$ | Contains fliA gene (sigma factor), fliZ (possible cell-density responsive regulator of sigma) and fliY (periplasmic cystine-binding protein) |
| fecA      | $u_{source}$ | Ferric citrate, outer membrane receptor |
| arcA      | $u_{target}$ | Aerobic respiration control, transcriptional dual regulator |
| aspA      | $u_{target}$ | Component of aspartate ammonia-lyase |
| crp       | $u_{central}$ | Component of CRP transcriptional dual regulator (DNA-binding transcriptional dual regulator) |
| CaiF      | $u_{central}$ | DNA-binding transcriptional activator |
| sodA      | $u_{central}$ | Component of superoxide dismutases that catalyzes the dismutation of superoxide into oxygen and hydrogen peroxide |

**NETWORK 4: T-LGL SIGNALING NETWORK**

| Node name | Node type | Details |
|-----------|-----------|---------|
| PDGF      | $u_{source}$ | Platelet-derived growth factor is one of the numerous growth factors, or proteins that regulates cell growth and division. |
| IL15      | $u_{source}$ | Interleukin 15 is a cytokine. |
| Stimuli   | $u_{source}$ | Antigen Stimulation |
| apoptosis | $u_{target}$ | process of programmed cell death |
| IL2       | $u_{central}$ | Interleukin 2 is a cytokine signaling molecule in the immune system |
| Ceramide  | $u_{central}$ | A waxy lipid molecule within the cell membrane which can participate in variety of cellular signaling like proliferation and apoptosis |
| GZMB      | $u_{central}$ | A serine proteases that is released within cytotoxic T cells and natural killer cells to induce apoptosis within virus-infected cells, thus destroying them |
| NFKB      | $u_{central}$ | nuclear factor kappa-light-chain-enhancer of activated B cells, a protein complex that controls the transcription of DNA |
| MCL1      | $u_{central}$ | Induced myeloid leukemia cell differentiation protein Mcl-1 |