PHYSICAL REVIEW E **89**, 032811 (2014)

Topological implications of negative curvature for biological and social networks

Réka Albert*
Department of Physics, Pennsylvania State University, University Park, Pennsylvania 16802, USA

Bhaskar DasGupta† and Nasim Mobasheri‡
Department of Computer Science, University of Illinois at Chicago, Chicago, Illinois 60607, USA
(Received 11 September 2013; revised manuscript received 16 February 2014; published 24 March 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 parametrized 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 crosstalk in biological networks, and to the existence of central, influential neighborhoods in both biological and social networks.

DOI: 10.1103/PhysRevE.89.032811 PACS number(s): 89.75.Hc, 87.18.Mp, 87.18.Cf, 87.18.Vf

I. INTRODUCTION

For a large variety of complex systems, ranging from the Internet to metabolic networks, representation as a parametrized 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 of novel network measures [3–6]. In this article we consider a combinatorial measure of negative curvature (also called hyperbolicity) of parametrized 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 an infinite metric space with bounded local geometry as originally proposed by Gromov [7] using a so-called "four-point condition." We adopt this measure for parametrized 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 [15], 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 $n \geqslant 4$ nodes. We will use the following notations:

(1) $u \overset{\mathcal{P}}{\leftrightsquigarrow} v$ denotes a path $\mathcal{P} \equiv (u = u_0, u_1, \ldots, u_{k-1}, u_k = v)$ from node u to node v and $\ell(\mathcal{P})$ denotes the *length* (number of edges) of such a path.

(2) $u_i \overset{\mathcal{P}}{\leftrightsquigarrow} u_j$ denotes the subpath $(u_i, u_{i+1}, \ldots, u_j)$ of $\mathcal{P}$ from u_i to u_j.

(3) $u \overset{\mathfrak{s}}{\leftrightsquigarrow} v$ denotes a shortest path from node u to node v of length $d_{u,v} = \ell(u \overset{\mathfrak{s}}{\leftrightsquigarrow} v)$.

*ralbert@phys.psu.edu
†dasgupta@cs.uic.edu
‡nmobas2@uic.edu

TABLE I. Hyperbolicity and diameter values for biological networks.

Network id	Ref.	Average degree	$\delta^+_{\text{ave}}(G)$	$\delta^+_{\text{worst}}(G)$	$\mathcal{D}$	$\frac{\delta^+_{\text{worst}}(G)}{\frac{\mathcal{D}}{2}}$
1. *E. coli* transcriptional	[26]	1.45	0.132	2	10	0.400
2. Mammalian signaling	[27]	2.04	0.013	3	11	0.545
3. *E. coli* transcriptional	[28]	1.30	0.043	2	13	0.308
4. T LGL signaling	[29]	2.32	0.297	2	7	0.571
5. *S. cerevisiae* transcriptional	[30]	1.56	0.004	3	15	0.400
6. *C. elegans* metabolic	[31]	4.50	0.010	1.5	7	0.429
7. *Drosophila* segment polarity	[32]	1.69	0.676	4	9	0.889
8. ABA signaling	[33]	1.60	0.302	2	7	0.571
9. Immune response network	[34]	2.33	0.286	1.5	4	0.750
10. T-cell receptor signalling	[35]	1.46	0.323	3	13	0.462
11. Oriented yeast PPI	[36]	3.11	0.001	2	6	0.667

We introduce the hyperbolicity measures via the four-node condition as originally proposed by Gromov. Consider a quadruple of distinct nodes[1] 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

$$\begin{aligned} S_{u_1,u_2,u_3,u_4} &= d_{u_{\pi_1},u_{\pi_2}} + d_{u_{\pi_3},u_{\pi_4}} \\ &\leqslant M_{u_1,u_2,u_3,u_4} = d_{u_{\pi_1},u_{\pi_3}} + d_{u_{\pi_2},u_{\pi_4}} \\ &\leqslant L_{u_1,u_2,u_3,u_4} = d_{u_{\pi_1},u_{\pi_4}} + d_{u_{\pi_2},u_{\pi_3}}, \end{aligned}$$

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

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

and an average hyperbolicity as

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

Note that $\delta^+_{\text{ave}}(G)$ 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 $\delta^+_{\text{worst}}(G)$ and $\delta^+_{\text{ave}}(G)$ can be trivially computed in $O(n^4)$ time for any graph G.

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

The hyperbolicity measure δ^+_{worst} 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. δ^+_{worst} is trivially infinite in the standard (unbounded) Euclidean space. Intuitively, a metric space has a finite value of δ^+_{worst} if it behaves metrically in the large scale as a negatively curved Riemannian manifold, and thus the value of δ^+_{worst} 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 δ^+_{worst}, that is, $O((\sqrt{-\alpha})^{-1})$ (see [24]).

In this paper we first show that a variety of biological and social networks are hyperbolic. We formulate and prove bounds on the existence of path chords and on the distance among shortest or approximately shortest paths in hyperbolic networks. We determine the implications of these bounds on *regulatory* networks, i.e., directed networks whose edges correspond to regulation or influence. This category includes all the biological networks that we study in this paper. We also discuss the implications of our results on the region of influence of nodes in social networks. Some of the proofs of our theoretical results are adaptation of corresponding arguments in the continuous hyperbolic space. All the proofs are presented in the appendix for the sake of completeness.

III. RESULTS AND DISCUSSION

Section IIIA examines in detail the hyperbolicity of an assorted list of diverse biological and social networks. The remaining subsections of this section, namely Secs. IIIB–IIIE, state our findings on the implications of hyperbolicity of a network on various topological properties of the network. For Secs. IIID and IIIE, we first state our findings as applicable for biological or social networks, followed by a summary of formal mathematical results that led to such findings. Because the precise bounds on topological features of a network as a function of hyperbolicity measures are quite mathematically involved, we discuss these bounds in a somewhat simplified form in Secs. IIIB–IIIE, leaving the precise bounds as theorems and proofs in the Appendix.

A. Hyperbolicity of real networks

We analyzed 20 well-known biological and social networks (see Supplemental Material [25]). The 11 biological networks shown in Table I include three transcriptional regulatory, five signaling, one metabolic, one immune response, and

[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.

TABLE II. Hyperbolicity and diameter values for social networks.

Network id	Ref.	Average degree	$\delta^+_{\mathrm{ave}}(G)$	$\delta^+_{\mathrm{worst}}(G)$	$\mathcal{D}$	$\frac{\delta^+_{\mathrm{worst}}(G)}{\frac{\mathcal{D}}{2}}$
1. Dolphins social network	[37]	5.16	0.262	2	8	0.750
2. American College Football	[38]	10.64	0.312	2	5	0.800
3. Zachary Karate Club	[39]	4.58	0.170	1	5	0.400
4. Books about US Politics	[40]	8.41	0.247	2	7	0.571
5. Sawmill communication	[41]	3.44	0.162	1	8	0.250
6. Jazz musician	[42]	27.69	0.140	1.5	6	0.500
7. Visiting ties in San Juan	[43]	3.84	0.422	3	9	0.667
8. World Soccer data, 1998	[44]	3.37	0.270	2.5	12	0.286
9. Les Miserable	[45]	6.51	0.278	2	14	0.417

one oriented protein-protein interaction networks. Similarly, the nine social networks shown in Table II range from interactions in dolphin communities to the social network of jazz musicians. The hyperbolicity of the biological and directed social networks was computed by ignoring the direction of edges. The hyperbolicity values were calculated by writing codes in C using standard algorithmic procedures.

As shown in Tables I and II, the hyperbolicity values of almost all networks are small. If $\mathcal{D} = \max_{u,v}\{d_{u,v}\}$ is the diameter of the graph, then it is easy to see that $\delta^+_{\mathrm{worst}}(G) \leqslant \frac{\mathcal{D}}{2}$, and thus small diameter indeed implies a small value of worst-case hyperbolicity. As can be seen in Tables I and II, $\delta^+_{\mathrm{worst}}(G)$ varies with respect to its worst-case bound of $\frac{\mathcal{D}}{2}$ from 25% of $\frac{\mathcal{D}}{2}$ to no more than 89% of $\frac{\mathcal{D}}{2}$, and there does not seem to be a systematic dependence of $\delta^+_{\mathrm{worst}}(G)$ on the number of nodes (which ranges from 18 to 786), edges (from 42 to 2742), or on the value of the diameter $\mathcal{D}$.

For all the networks $\delta^+_{\mathrm{ave}}(G)$ is one or two orders of magnitude smaller than $\delta^+_{\mathrm{worst}}(G)$. Intuitively, this suggests that the value of $\delta^+_{\mathrm{worst}}(G)$ may be a rare deviation from typical values of $\delta^+_{u_1,u_2,u_3,u_4}$ that one would obtain for most combinations of nodes $\{u_1,u_2,u_3,u_4\}$.

We additionally performed the following rigorous tests for hyperbolicity of our networks.

1. Checking hyperbolicity via the scaled hyperbolicity approach

An approach for testing hyperbolicity for finite graphs was introduced and used via "scaled" Gromov hyperbolicity in [9,11] for hyperbolicity defined via thin triangles and in [46] for hyperbolicity defined via the four-point condition as used in this paper. The basic idea is to "scale" the values of $\delta^+_{u_1,u_2,u_3,u_4}$ by a suitable scaling factor, say μ_{u_1,u_2,u_3,u_4}, such that there exists a constant $0 < \varepsilon < 1$ with the following property:

(1) The maximum achievable value of $\frac{\delta^+_{u_1,u_2,u_3,u_4}}{\mu_{u_1,u_2,u_3,u_4}}$ is ε in the standard hyperbolic space or in the Euclidean space, and

(2) $\frac{\delta^+_{u_1,u_2,u_3,u_4}}{\mu_{u_1,u_2,u_3,u_4}}$ goes beyond ε in positively curved spaces.

We use the notation $\mathcal{D}_{u_1,u_2,u_3,u_4} = \max_{i,j\in\{1,2,3,4\}}\{d_{u_i,u_j}\}$ 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 [46] provide the bounds shown in Table III.

We adapt the criterion proposed by Jonckheere, Lohsoonthorn, and Ariaei [46] to designate a given finite graph as hyperbolic by requiring a *significant* percentage of all possible subsets of four nodes to satisfy the ε bound. More formally, suppose that G has t connected components containing $n_1,n_2,\ldots,n_t$ nodes, respectively ($\sum_{j=1}^{t} n_j = n$). 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

$$\Delta^{\mathrm{Y}}(G) = \frac{\text{number of subset of four nodes } \{u_i,u_j,u_k,u_\ell\} \text{ such that } \delta^{\mathrm{Y}}_{u_i,u_j,u_k,u_\ell} > \varepsilon}{\text{number of all possible combinations of four nodes that contribute to hyperbolicity}}$$

$$= \frac{\text{number of subset of four nodes } \{u_i,u_j,u_k,u_\ell\} \text{ such that } \delta^{\mathrm{Y}}_{u_i,u_j,u_k,u_\ell} > \varepsilon}{\sum_{1\leqslant j\leqslant t: n_j>3}\binom{n_j}{4}} < 1-\eta.$$

The values of $\Delta^{\mathrm{Y}}(G)$ for our networks are shown in Tables IV and 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 $\Delta^{\mathrm{Y}}(G)$ values by computing the statistical significance values (commonly called p values) of these $\Delta^{\mathrm{Y}}(G)$ 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 [5,26]) for such purpose. For each network G, we generated 100 randomized versions of the network using a Markov-chain algorithm [47] by swapping the endpoints of randomly selected pairs of edges until 20% of the edges was changed. We computed the values of $\Delta^{\mathrm{Y}}(G_{\mathrm{rand}_1})$, $\Delta^{\mathrm{Y}}(G_{\mathrm{rand}_2}),\ldots,\Delta^{\mathrm{Y}}(G_{\mathrm{rand}_{100}})$. We then used an (unpaired) one-sample student's t test to determine the probability that $\Delta^{\mathrm{Y}}(G)$ belongs to the same distribution as $\Delta^{\mathrm{Y}}(G_{\mathrm{rand}_1})$, $\Delta^{\mathrm{Y}}(G_{\mathrm{rand}_2}),\ldots,\Delta^{\mathrm{Y}}(G_{\mathrm{rand}_{100}})$.

TABLE III. [46] Various scaled Gromov hyperbolicities.

Name	Notation	μ_{u_1,u_2,u_3,u_4}	ε	Method for determining ε
Diameter-scaled hyperbolicity	$\delta^{\mathcal{D}}$	$\mathcal{D}_{u_1,u_2,u_3,u_4}$	0.2929	Empirical
L-scaled hyperbolicity	δ^{L}	L_{u_1,u_2,u_3,u_4}	$\frac{\sqrt{2}-1}{2\sqrt{2}} \approx 0.1464$	Mathematical
$(L+M+S)$-scaled hyperbolicity	$\delta^{\mathrm{L+M+S}}$	$L_{u_1,u_2,u_3,u_4} + M_{u_1,u_2,u_3,u_4} + S_{u_1,u_2,u_3,u_4}$	0.0607	Mathematical

The p values, tabulated in Tables VI and VII, clearly show that all social networks and all except two biological networks can be classified as hyperbolic in a statistically significant manner, implying that the topologies of these networks are close to a "tree topology." Indeed, for biological networks, the assumption of chainlike or treelike topology is frequently made in the traditional molecular biology literature [48]. Independent current observations also provide evidence of treelike topologies for various biological networks, e.g., the average in-out degree of transcriptional regulatory networks [26,49] and of a mammalian signal transduction network [27] is close to 1, so cycles are very rare.

B. Hyperbolicity and crosstalk in regulatory networks

Let $C = (u_0,u_1,\ldots,u_{k-1},u_0)$ be a cycle of $k \geqslant 4$ nodes. A *path chord* of C is defined to be a path $u_i \overset{\mathcal{P}}{\rightsquigarrow} u_j$ between two distinct nodes $u_i,u_j \in C$ such that the length of $\mathcal{P}$ is less than $(i-j) \pmod k$ (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 Sec. A of the Appendix). In particular, G does not have a cycle of more than $4\,\delta^{+}_{\text{worst}}(G)$ 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:

(1) If a node regulates itself through a long feedback loop (e.g., of length at least 6 if $\delta^{+}_{\text{worst}}(G) = \frac{3}{2}$) then this loop *must* have a path chord. Thus it follows that there exists a *shorter* feedback cycle through the same node.

(2) A chord or short path chord can be interpreted as *crosstalk* between two paths between a pair of nodes. With this interpretation, the following conclusion follows. If one node in a regulatory network regulates another node through two sufficiently long paths, then there must be a crosstalk path between these two paths. For example, assuming $\delta^{+}_{\text{worst}}(G) = \frac{3}{2}$, there must be a crosstalk path if the sum of lengths of the two paths is at least 6. In general, the number of crosstalk paths between two paths increases *at least* linearly with the total length of the two paths. The general conclusion that can be drawn is that independent linear pathways that connect a signal to the same output node (e.g., transcription factor) are rare, and if multiple pathways exist then they are interconnected through crosstalks.

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 [50]). 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 $\mathcal{P}_{\Delta}(u_0,u_1)$, $\mathcal{P}_{\Delta}(u_0,u_2)$, $\mathcal{P}_{\Delta}(u_1,u_2)$ between u_0 and u_1, u_0 and u_2, and u_1 and u_2, respectively. As illustrated in Fig. 2, in hyperbolic networks

TABLE IV. $\Delta^{\mathrm{Y}}(G)$ values for biological networks for $Y \in \{\mathcal{D}, L, L+M+S\}$.

Network id	$\Delta^{\mathcal{D}}(G)$	$\Delta^{\mathrm{L}}(G)$	$\Delta^{\mathrm{L+M+S}}(G)$
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 signaling	0.0034	0.0045	0.0056
11. Oriented yeast PPI	0.0013	0.0009	0.0012
Maximum	0.0461	0.0558	0.0781

TABLE V. $\Delta^{\mathrm{Y}}(G)$ values for social networks for $Y \in \{\mathcal{D}, L, L+M+S\}$.

Network id	$\Delta^{\mathcal{D}}(G)$	$\Delta^{\mathrm{L}}(G)$	$\Delta^{\mathrm{L+M+S}}(G)$
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
Maximum	0.0435	0.0395	0.0577

TABLE VI. p values for the $\Delta^{Y}(G)$ values for biological networks for $Y \in \{\mathcal{D}, L, L+M+S\}$. 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. *E. coli*	2. Mammalian signaling	3. *E. Coli* transcriptional	4. T LGL signaling	5. *S. cerevisiae* transcriptional	6. *C. elegans* Metabolic	7. *Drosophila* segment polarity	8. ABA signaling	9. Immune response network	10. T-cell receptor signalling	11. Oriented yeast PPI
p values	$\Delta^{\mathcal{D}}$	**0.0018**	**<0.0001**	**<0.0001**	**<0.0001**	0.3321	**<0.0001**	**<0.0001**	**<0.0001**	**<0.0001**	**<0.0001**	**<0.0001**
	Δ^{L}	**<0.0001**	**<0.0001**	**<0.0001**	0.011	0.3434	**<0.0001**	**<0.0001**	0.9145	**<0.0001**	**<0.0001**	**<0.0001**
	$\Delta^{\mathrm{L+M+S}}$	0.5226	**<0.0001**	**<0.0001**	**<0.0001**	0.3424	**<0.0001**	**<0.0001**	0.3342	**<0.0001**	**<0.0001**	**<0.0001**

we are guaranteed to find short paths[2] between the nodes that make up $\mathcal{P}_\Delta(u_0,u_1)$, $\mathcal{P}_\Delta(u_0,u_2)$, $\mathcal{P}_\Delta(u_1,u_2)$. This is formally stated in Theorem 3 in Sec. B of the Appendix. Moreover, as Corollary 4 (in Sec. B 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 chords. 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-chord result is obtained.

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 crosstalk paths between these shortest paths. Consequently, the feedback or feed-forward loop will be *nested* with "additional" feedback 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 [51].

(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 of a target node is mediated by two or more short paths[3] 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.

See Fig. 3 for a pictorial illustration.

Formal justifications and intuitions (see Theorem 5 and Corollary 6 in Sec. C and Theorem 7 and Corollary 8 in Sec. D of the Appendix).

We use the following two quantifications of "approximately" short paths:

(1) A path $u_0 \overset{\mathcal{P}}{\leftrightsquigarrow} u_k = (u_0,u_1,\ldots,u_k)$ is μ-approximate short provided $\ell(u_i \overset{\mathcal{P}}{\leftrightsquigarrow} u_j) \leqslant \mu\, d_{u_i,u_j}$ for all $0 \leqslant i < j \leqslant k$.

(2) A path $u_0 \overset{\mathcal{P}}{\leftrightsquigarrow} u_k$ is ε-additive-approximate short provided $\ell(\mathcal{P}) \leqslant d_{u_0,u_k} + \varepsilon$.

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

(1) Let $\mathcal{P}_1$ and $\mathcal{P}_2$ 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 $\mathcal{P}_1$, there

[2]By a short path here, we mean a path whose length is at most a constant times $\delta^+_{\Delta_{\{u_0,u_1,u_2\}}}$ [note that $\delta^+_{\Delta_{\{u_0,u_1,u_2\}}} \leqslant \delta^+_{\mathrm{worst}}(G)$].

[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 μ approximate short path or a ε-additive-approximate short path, as defined in the subsequent "Formal justifications and intuitions" subsection, for small μ or small ε, respectively.

TABLE VII. p values for the $\Delta^{Y}(G)$ values for social networks for $Y \in \{\mathcal{D}, L, L+M+S\}$. 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	$\Delta^{\mathcal{D}}$	**<0.0001**	**<0.0001**	**<0.0001**	**<0.0001**	**<0.0001**	**<0.0001**	**<0.0001**	**<0.0001**	**<0.0001**
	Δ^{L}	**<0.0001**	**<0.0001**	**<0.0001**	**<0.0001**	**<0.0001**	**<0.0001**	0.0779	**<0.0001**	**<0.0001**
	$\Delta^{\mathrm{L+M+S}}$	**<0.0001**	**<0.0001**	**<0.0001**	**<0.0001**	**<0.0001**	**<0.0001**	**<0.0001**	**<0.0001**	**<0.0001**

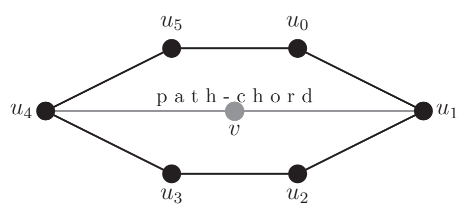


FIG. 1. Path chord of a cycle $C = (u_0.u_1,u_2,u_3,u_4,u_5,u_0)$.

exists a node v' on $\mathcal{P}_2$ such that $d_{v,v'}$ depends linearly on $\delta^{+}_{\rm worst}(G)$, only logarithmically on the length of $\mathcal{P}_2$ 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 3 on u_0, u_1 and the middle node of the path $\mathcal{P}_2$ and then use the same approach recursively on a part of the path $\mathcal{P}_2$ containing at most $\lceil \frac{\ell(\mathcal{P}_2)}{2} \rceil$ edges. The depth of the level of recursion provides the logarithmic factor in the bound.

(2) If $\mathcal{P}_1$ and $\mathcal{P}_2$ are two short paths between u_0 and u_1 then Theorem 7 and Corollary 8 imply that the Hausdorff distance between $\mathcal{P}_1$ and $\mathcal{P}_2$ depends on $\delta^{+}_{\rm worst}(G)$ only and does *not* depend on the size or any other parameter of the network. Intuitively, Theorem 7 and Corollary 8 can be thought of as generalizing and improving the bound in Theorem 5 for approximately short paths.

D. Identifying essential edges in the regulation between two nodes

For a given $\xi > 0$ and a node u, let $\mathcal{B}_\xi(u) = \{v | d_{u,v} = \xi\}$ denote the "boundary of the ξ neighborhood" of u, i.e., the set of all nodes at a distance of precisely ξ from u. Our two findings in the present context are as stated in (I) and (II) below.

(I) Identifying relevant paths between a source and a target node. Suppose that we pick a node v and consider the *strict* ξ neighborhood of v,

$$N^{+}_{\xi}(v) = \bigcup_{r \leqslant \xi} \mathcal{B}_{r'}(v) \setminus \{u |\text{ degree of } u \text{ is one}\}$$

(i.e., the set of all nodes, excluding nodes of degree 1, that are at a distance at most ξ from u) for a sufficiently large ξ. Consider two nodes u_1 and u_2 on the boundary of this neighborhood, i.e., at a distance ξ from v. Then, the following holds:

(A) The relevant (short) regulatory paths between u_1 and u_2 do *not* leave the neighborhood, i.e., *all* the edges in the relevant regulatory paths are in the neighborhood.

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 $u_{\rm source}$ and output node $u_{\rm target}$ 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 [51]. 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 $u_{\rm central}$ which is at equal distance between $u_{\rm source}$ and $u_{\rm target}$, and is on the shortest, or close to shortest, path between $u_{\rm source}$ and $u_{\rm target}$. Then one constructs the neighborhood around $u_{\rm central}$ such that $u_{\rm source}$ and $u_{\rm target}$ 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. 4 for a pictorial illustration of this implication.

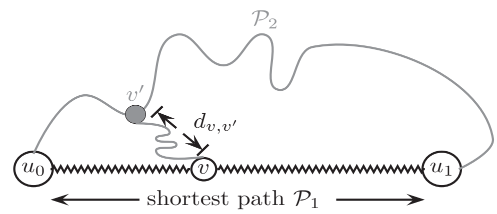


FIG. 3. An informal and simplified pictorial illustration of the claims in Sec. III C(b).

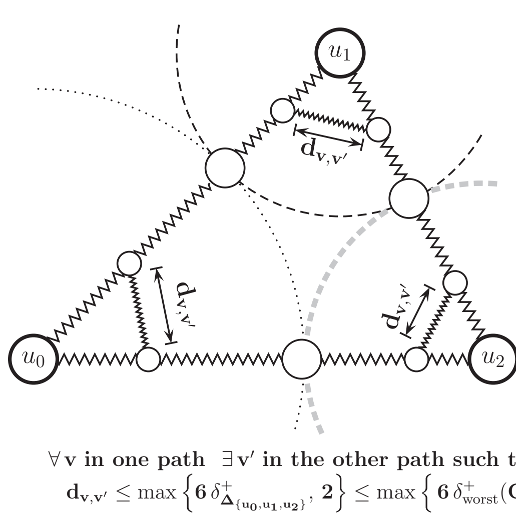


FIG. 2. An informal and simplified pictorial illustration of the claims in Sec. III C(a).

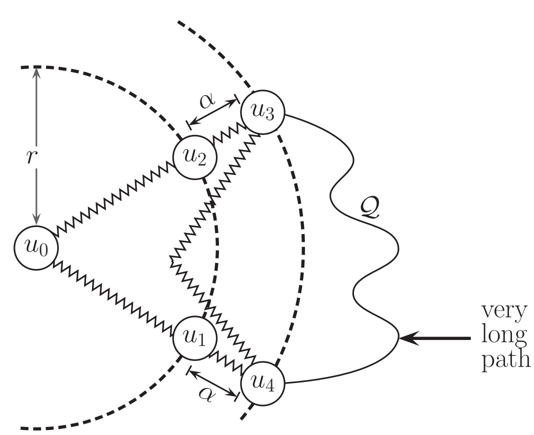


FIG. 4. An informal and simplified pictorial illustration of claim (A) in Sec. III D. As the nodes u_3 and u_4 move further away from the center node u_0, the shortest path between them bends more towards u_0 and any path between them that does not involve a node in the ball $\cup_{r' \leqslant r} B_{r'}(u_0)$ is long enough.

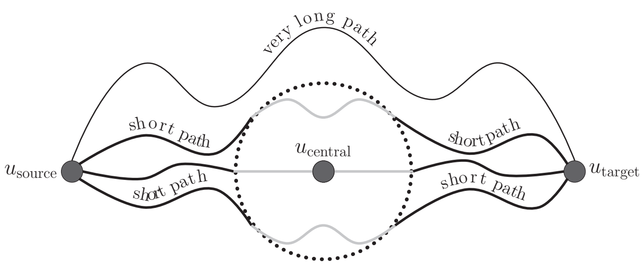


FIG. 5. An informal and simplified pictorial illustration of claim (B) in Sec. III D. Knocking out the nodes in a small neighborhood of u_{central} cuts off all relevant (short) regulation between u_{source} and u_{target}.

(II) Finding essential nodes. Again, consider an input node u_{source} and output node u_{target} of a signal transduction network, and let u_{central} be a central node which is on the shortest path between them and at approximately equal distance between u_{source} and u_{target}. Our results show that[4]

> (B) If one constructs a small ξ neighborhood around u_{central} with $\xi = \mathrm{O}\,(\delta^{+}_{\mathrm{worst}}(G))$, then *all* relevant (short or approximately short) paths between u_{source} and u_{target} must include a *node* in this ξ neighborhood. Therefore, "knocking out" the nodes in this ξ neighborhood cuts off all relevant regulatory paths between u_{source} and u_{target}.

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

Formal justifications and intuitions for ($\star$) *and* ($\star\star$) *(see Theorem 10 and Corollary 11 in Sec. E of the Appendix).*

Suppose that we are given the following:

(1) three integers $\kappa \geqslant 4, \alpha > 0, r > (\frac{\kappa}{2} - 1)(6\,\delta^{+}_{\mathrm{worst}}(G) + 2)$,

(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 $d_{u_1,u_2} \geqslant \frac{\kappa}{2}\,(6\,\delta^{+}_{\mathrm{worst}}(G) + 2)$,
- $d_{u_1,u_4} = d_{u_2,u_3} = \alpha$.

Then, (A) and (B) are implied by the following type of *asymptotic* bounds provided by Theorem 10 and Corollary 11:

> For a suitable positive value $\lambda = \mathrm{O}\big(\delta^{+}_{\mathrm{worst}}(G)\big)$, if $d_{u_1,u_4} = d_{u_2,u_3} = \alpha > \lambda$ then one of the following is true for any path Q between u_3 and u_4 that does not involve a node in $\cup_{r' \leqslant r} \mathcal{B}_{r'}\,(u_0)$:
>
> (1) Q does not exist (i.e.,, $\ell(Q) \geqslant n$), or
>
> (2) Q is much longer than a shortest path between the two nodes, i.e.,, if Q is a μ-approximate short path or a ε-additive-approximate short path then μ or ε is large.
>
> A pessimistic estimate shows that a value of λ that is about $6\,\delta^{+}_{\mathrm{worst}}(G) + 2$ suffices. As we subsequently observe, for real networks the bound is much better, about $\lambda \approx \delta^{+}_{\mathrm{worst}}(G)$.

Empirical evaluation of (A).

We empirically investigated the claim in (A) 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_{source} and u_{target}, we found the shortest path(s) between them. For each such shortest path, a central node u_{central} was identified. We then

[4] O and Ω are the standard notations used in analyzing asymptotic upper and lower bounds in the computer science literature: given two functions $f(n)$ and $g(n)$ of a variable n, $f(n) = \mathrm{O}(g(n))$ [respectively, $f(n) = \Omega(g(n))$] provided there exists two constants $n_0, c > 0$ such that $f(n) \leqslant c\,g(n)$ [respectively, $f(n) \geqslant c\,g(n)$] for $n \geqslant n_0$.

TABLE VIII. Effect of the prescribed neighborhood in claim (A) on all edges in relevant paths.
$\mathcal{SP}$, shortest path between u_{source} and u_{target}.
$\mathcal{SP}^{+1}$, paths between u_{source} and u_{target} with one extra edge than $\mathcal{SP}$(1-additive-approximate short path).
$\mathcal{SP}^{+2}$, paths between u_{source} and u_{target} with two extra edges than $\mathcal{SP}$ (2-additive-approximate short path).
$N^{+}_{\xi}(u_{\mathrm{central}})$, strict $\xi = d_{u_{\mathrm{source}}, u_{\mathrm{target}}}$ neighborhood of u_{central}.
n, size (number of nodes) of the network.
$\frac{N^{+}_{\xi}(u_{\mathrm{central}})}{n}$, fraction of strict $\xi = d_{u_{\mathrm{source}}, u_{\mathrm{target}}}$ neighborhood of u_{central} with respect to the size of the network.

Network name	u_{source}	u_{target}	$d_{u_{\mathrm{source}}, u_{\mathrm{target}}}$	u_{central}	$\frac{N^{+}_{\xi}(u_{\mathrm{central}})}{n}$	% of $\mathcal{SP}$ with every edge in the neighborhood of claim (A)	% of $\mathcal{SP}^{+1}$ with every edge in the neighborhood of claim (A)	% of $\mathcal{SP}^{+2}$ with every edge in the neighborhood of claim (A)
Network 1: *E. coli* transcriptional	fliAZY	arcA	4	CaiF	0.20	100%	100%	18%
				crp	0.27	100%	100%	70%
	fecA	aspA	6	crp	0.43	100%	100%	100%
				sodA	0.28	100%	100%	62%
Network 4: T-LGL signaling	IL15	Apoptosis	4	GZMB	0.37	100%	66%	40%
	PDGF	Apoptosis	6	IL2, NKFB	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.
$\mathcal{SP}$, shortest path between u_{source} and u_{target}.
$\mathcal{SP}^{+1}$, paths between u_{source} and u_{target} with one extra edge than $\mathcal{SP}$ (1-additive-approximate short path).
$\mathcal{SP}^{+2}$, paths between u_{source} and u_{target} with two extra edges than $\mathcal{SP}$ (2-additive-approximate short path).

Network name	u_{source}	u_{target}	$d_{u_{\text{source}},u_{\text{target}}}$	u_{central}	% of $\mathcal{SP}$ with a node in ξ neighborhood		% of $\mathcal{SP}^{+1}$ with a node in ξ neighborhood		% of $\mathcal{SP}^{+2}$ with a node in ξ neighborhood	
Network 1:	fliAZY	arcA	4	CaiF	$\xi=1$	100%	$\xi=1$	71%	$\xi=1$	59%
E. coli				crp	$\xi=1$	100%	$\xi=1$	100%	$\xi=1$	100%
transcriptional	fecA	aspA	6	crp	$\xi=1$	100%	$\xi=1$	100%	$\xi=1$	100%
$\delta^{+}_{\text{worst}}(G)=2$				sodA	$\xi=1$	100%	$\xi=1$	100%	$\xi=1$	100%
Network 4:	IL15	apoptosis	4	GZMB	$\xi=1$	100%	$\xi=1$	100%	$\xi=1$	100%
T-LGL				IL2	$\xi=1$	80%	$\xi=1$	82%	$\xi=1$	93%
signaling					$\xi=2$	100%	$\xi=2$	100%	$\xi=2$	100%
$\delta^{+}_{\text{worst}}(G)=2$				NFKB	$\xi=1$	80%	$\xi=1$	86%	$\xi=1$	76%
					$\xi=2$	100%	$\xi=2$	100%	$\xi=2$	100%
	PDGF	apoptosis	6	Ceramide	$\xi=1$	40%	$\xi=1$	23%	$\xi=1$	40%
					$\xi=2$	100%	$\xi=2$	100%	$\xi=2$	100%
				MCL1	$\xi=1$	60%	$\xi=1$	47%	$\xi=1$	73%
					$\xi=2$	100%	$\xi=2$	100%	$\xi=2$	100%
	Stimuli	apoptosis	4	GZMB	$\xi=1$	100%	$\xi=1$	100%	$\xi=1$	100%

considered the ξ neighborhood of u_{central} such that both u_{source} and u_{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_{source} and u_{target} had *all* edges in this neighborhood. The results, tabulated in Table VIII, support (A).

Empirical evaluation of (B).

We empirically investigated the size ξ of the neighborhood in claim (B) for the same two biological networks and the same combinations of source, target, and central nodes as in claim (A). We considered the ξ neighborhood of u_{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_{source} and u_{target} involved a node in this neighborhood (not counting u_{source} and u_{target}). The results, tabulated in Table IX, show that removing the nodes in a $\xi \leqslant \delta^{+}_{\text{worst}}(G)$ neighborhood around the central nodes disrupts all the relevant paths of the selected networks. As $\delta^{+}_{\text{worst}}(G)$ 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_{source} and u_{target}.

E. Effect of hyperbolicity on structural holes in social networks

For a node $u \in V$, let $\mathsf{Nbr}(u)=\{v\,|\,\{u,v\}\in E\}$ be the set of neighbors of (i.e., nodes adjacent to) u. To quantify the useful information in a social network, Ron Burt in [52] 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 $\mathfrak{M}_u$ of the structural hole at u is defined as [52,53]:

$$\mathfrak{M}_u \stackrel{\text{def}}{=} \sum_{v\in V}\left(\frac{a_{u,v}+a_{v,u}}{\max_{x\neq u}\{a_{u,x}+a_{x,u}\}}\left[1-\sum_{\substack{y\in V\\ y\neq u,v}}\left(\frac{a_{u,y}+a_{y,u}}{\sum_{x\neq u}(a_{u,x}+a_{x,u})}\right)\left(\frac{a_{v,y}+a_{y,v}}{\max_{z\neq y}\{a_{v,z}+a_{z,v}\}}\right)\right]\right),$$

where $a_{p,q}=\{\begin{smallmatrix}1, & \text{if } \{p,q\}\in E\\ 0, & \text{otherwise}\end{smallmatrix}$ are the entries in the standard adjacency matrix of G. By observing that $a_{p,q}=a_{q,p}$ and $\max_{x\neq u}\{a_{u,x}+a_{x,u}\}=\max_{z\neq y}\{a_{v,z}+a_{z,v}\}=2$, the above equation for $\mathfrak{M}_u$ can be simplified to

$$\mathfrak{M}_u = |\mathsf{Nbr}(u)| - \frac{\sum_{v,y\,\in\,\mathsf{Nbr}(u)} a_{v,y}}{|\mathsf{Nbr}(u)|}. \tag{1}$$

Thus high-degree nodes whose neighbors are not connected to each other have high $\mathfrak{M}_u$ values. For an intuitive interpretation and generalization of (1), the following definition of weak and strong dominance will prove useful (cf. dominating set problem for graphs [54] and point domination problems in geometry [55]). A pair of distinct nodes v,y is weakly (ρ,λ) dominated [respectively, strongly (ρ,λ) dominated] by a node u provided (see Fig. 6):

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

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

Let $\{\mathbf{v},\mathbf{y}\} \prec^{\rho,\lambda}_{\text{weak}} \mathbf{u}$ (respectively, $\{\mathbf{v},\mathbf{y}\} \prec^{\rho,\lambda}_{\text{strong}} \mathbf{u}$)

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

Since $\mathcal{B}_1(u) = \bigcup_{0<j\leqslant 1} \mathcal{B}_j(u) = \mathsf{Nbr}(u)$, it follows that

$$\begin{aligned}\mathfrak{M}_u &= |\cup_{0<j\leqslant 1} \mathcal{B}_j(u)| - \frac{\sum_{v,y\in\bigcup_{0<j\leqslant 1}\mathcal{B}_j(u)} \left(1 - \{\mathbf{v},\mathbf{y}\} \prec^{\mathbf{0,1}}_{\text{weak}} \mathbf{u}\right)}{|\cup_{0<j\leqslant 1} \mathcal{B}_j(u)|} \\ &= \mathbb{E}\left[\begin{array}{l|l} \text{number of pairs of nodes} & \\ v,y \text{ such that } v,y \text{ is weakly} & v \text{ is selected uniformly} \\ (0,1) \text{ dominated by } u & \text{randomly from } \bigcup_{0<j\leqslant 1}\mathcal{B}_j(u) \end{array}\right] \\ &\geqslant \mathbb{E}\left[\begin{array}{l|l} \text{number of pairs of nodes} & \\ v,y \text{ such that } v,y \text{ is strongly} & v \text{ is selected uniformly} \\ (0,1) \text{ dominated by } u & \text{randomly from } \bigcup_{0<j\leqslant 1}\mathcal{B}_j(u) \end{array}\right]\end{aligned}$$

and a generalization of $\mathfrak{M}_u$ is given by (replacing 0,1 by ρ,λ):

$$\begin{aligned}\mathfrak{M}_{u,\rho,\lambda} &= |\cup_{\rho<j\leqslant \lambda} \mathcal{B}_j(u)| - \frac{\sum_{v,y\in\bigcup_{\rho<j\leqslant \lambda}\mathcal{B}_j(u)} \left(1 - \{\mathbf{v},\mathbf{y}\} \prec^{\rho,\lambda}_{\text{weak}} \mathbf{u}\right)}{|\bigcup_{\rho<j\leqslant \lambda} \mathcal{B}_j(u)|} \\ &= \mathbb{E}\left[\begin{array}{l|l} \text{number of pairs of nodes} & \\ v,y \text{ such that } v,y \text{ is weakly} & v \text{ is selected uniformly} \\ (\rho,\lambda)\text{-dominated by } u & \text{randomly from } \cup_{\rho<j\leqslant\lambda}\mathcal{B}_j(u) \end{array}\right] \\ &\geqslant \mathbb{E}\left[\begin{array}{l|l} \text{number of pairs of nodes} & \\ v,y \text{ such that } v,y \text{ is strongly} & v \text{ is selected uniformly} \\ (\rho,\lambda)\text{-dominated by } u & \text{randomly from } \cup_{\rho<j\leqslant\lambda}\mathcal{B}_j(u) \end{array}\right].\end{aligned}$$

When the graph is hyperbolic [i.e., $\delta^+_{\text{worst}}(G)$ is a constant], for moderately large λ, weak and strong dominance are essentially identical and therefore weak domination has a much stronger implication. Recall that n denotes the number of nodes in the graph G.

Our finding can be succinctly summarized as (see Fig. 7 for a visual illustration)

(C) If $\lambda \geqslant (6\,\delta^+_{\text{worst}}(G) + 2)\log_2 n$ then, assuming v is selected uniformly randomly from $\cup_{\rho<j\leqslant\lambda}\mathcal{B}_j(u)$ for any node u, the expected number of pairs of nodes v,y that are weakly (ρ,λ) dominated by u is precisely the same as the expected number of pairs of nodes that are strongly (ρ,λ) dominated by u.

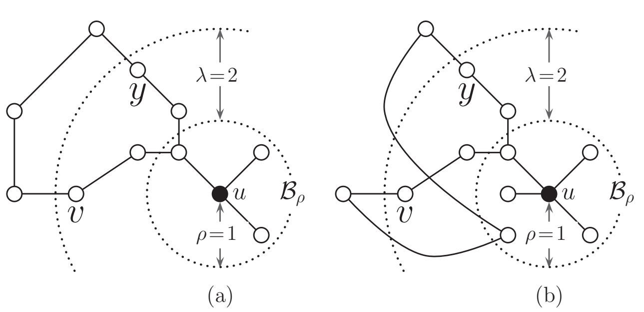


FIG. 6. Illustration of weak and strong domination. (a) v,y is weakly (ρ,λ) dominated by u since only one shortest path between v and y intersects $\mathcal{B}_\rho(u)$. (b) v,y is strongly (ρ,λ) dominated by u since all the shortest paths between v and y intersect $\mathcal{B}_\rho(u)$.

A mathematical justification for the claim (C) is provided by Lemma 12 in Sec. F of the Appendix.

An implication of (C).

If $\lambda \geqslant (6\,\delta^+_{\text{worst}}(G) + 2)\log_2 n$ and $\mathfrak{M}_{u,\rho,\lambda} \approx |\mathcal{B}_{\rho+\lambda}(u)|$, then *almost all* pairs of nodes are strongly (ρ,λ) dominated by u, i.e., for almost all pairs of nodes $v,y \in \mathcal{B}_{\rho+\lambda}(u)$, every shortest path between v and y contains a node in $\mathcal{B}_\rho(u)$.

A visual illustration of this implication is in Fig. 8 showing that as λ increases the shortest paths tend to bend more and more towards the central node u for a hyperbolic network.

Empirical verification of (C).

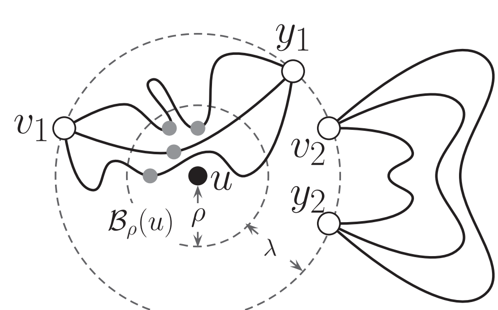


FIG. 7. Visual illustration. Either all the shortest paths are completely inside or all the shortest paths are completely outside of $\mathcal{B}_{\rho+\lambda}(u)$.

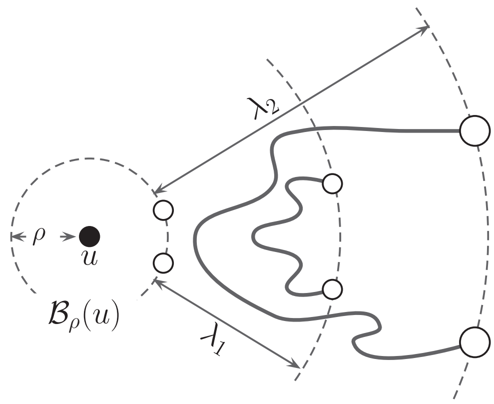


FIG. 8. For hyperbolic graphs, the further we move from the central (black) node, the more a shortest path bends inward towards the central node.

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

Network 1. Dolphin social network,
Network 4. Books about US politics,
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 ξ neighborhood $\mathcal{B}_\xi(u)$ of u for an appropriate $\xi = \rho + \lambda$. We then set λ to a very small value of 1, and calculated the following quantities.

(1) We computed the number n_1 of all pairs of nodes from $\mathcal{B}_\xi(u)$ that are weakly (ρ,λ) dominated by u.

(2) We computed the number n_2 of all pairs of nodes from $\mathcal{B}_\xi(u)$ that are strongly (ρ,λ) dominated by u.

Table X tabulates the ratio $\nu = \frac{n_2}{n_1}$, 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-treelike 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 crosstalk 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 nontrivial topological properties that are not fully reflected in other network measures, and therefore the hyperbolicity measure should be more widely used.

TABLE X. Weak domination leads to strong domination for social networks. u is the index of the central node and $\nu = \frac{n_2}{n_1} = \frac{|\{(v,y)\in\mathcal{B}_{\rho+\lambda}(u) \,|\, \{\mathbf{v},\mathbf{y}\} \prec^{\rho,\lambda}_{\text{strong}} \mathbf{u}=1\}|}{|\{(v,y)\in\mathcal{B}_{\rho+\lambda}(u) \,|\, \{\mathbf{v},\mathbf{y}\} \prec^{\rho,\lambda}_{\text{weak}} \mathbf{u}=1\}|}$.

Network name	u	ρ	λ	$\lvert\mathcal{B}_{\rho+\lambda}(u)\rvert$	ν
Network 1 Dolphin social network	14	4	1	5	80%
	37	4	1	3	100%
Network 4 Books about US politics	8	4	1	4	83%
	3	3	1	5	90%
Network 7 Visiting ties in San Juan	34	4	1	4	50%
	9	3	1	5	90%

ACKNOWLEDGMENTS

B.D. and N.M. were supported by National Science Foundation (NSF) Grant No. IIS-1160995. R.A. was supported by NSF Grants No. IIS-1161007 and No. PHY-1205840.

APPENDIX A: THEOREM 1

Theorem 1. Suppose that G has a cycle of $k \geqslant 4$ nodes which has no path chord. Then, $\delta^{+}_{\text{worst}}(G) \geqslant \lceil \frac{k}{4} \rceil$.

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

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 internode distances are *strictly* positive except for d_{u_1,u_3} and $d_{u_1,u_3} = \lceil \frac{d_{u_1,u_2}+d_{u_0,u_1}-d_{u_0,u_2}}{2} \rceil$. Then,

$$\left\lceil \frac{d_{u_0,u_1} + d_{u_0,u_2} + d_{u_1,u_2}}{2} \right\rceil \leqslant d_{u_0,u_3} + d_{u_1,u_2}$$
$$\leqslant \left\lceil \frac{d_{u_0,u_1} + d_{u_0,u_2} + d_{u_1,u_2}}{2} \right\rceil + 2\,\delta^{+}_{u_0,u_1,u_2,u_3}.$$

Proof. Note that due to triangle inequality $0 \leqslant \lceil \frac{d_{u_1,u_2}+d_{u_0,u_1}-d_{u_0,u_2}}{2} \rceil \leqslant d_{u_1,u_2}$ 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 four-node condition for the nodes u_0,u_1,u_2,u_3, namely the quantities $d_{u_0,u_3} + d_{u_1,u_2}$, $d_{u_0,u_2} + d_{u_1,u_3}$, and $d_{u_0,u_1} + d_{u_2,u_3}$. Note that

$$\begin{aligned}
2(d_{u_0,u_3} + d_{u_1,u_2}) &= (d_{u_0,u_3} + d_{u_1,u_3}) + (d_{u_0,u_3} + d_{u_2,u_3}) + d_{u_1,u_2} \\
&\geqslant d_{u_0,u_1} + d_{u_0,u_2} + d_{u_1,u_2} \\
&\Rightarrow d_{u_0,u_3} + d_{u_1,u_2} \\
&\geqslant \left\lceil \frac{d_{u_0,u_1} + d_{u_0,u_2} + d_{u_1,u_2}}{2} \right\rceil,
\end{aligned}$$

$$\begin{aligned}
d_{u_0,u_2} + d_{u_1,u_3} &= d_{u_0,u_2} + \left\lfloor \frac{d_{u_1,u_2} + d_{u_0,u_1} - d_{u_0,u_2}}{2} \right\rfloor \\
&= \left\lfloor \frac{d_{u_0,u_1} + d_{u_0,u_2} + d_{u_1,u_2}}{2} \right\rfloor,
\end{aligned}$$

$$\begin{aligned}
d_{u_0,u_1} + d_{u_2,u_3} &= d_{u_0,u_1} + \left\lceil \frac{d_{u_1,u_2} + d_{u_0,u_2} - d_{u_0,u_1}}{2} \right\rceil \\
&= \left\lceil \frac{d_{u_0,u_1} + d_{u_0,u_2} + d_{u_1,u_2}}{2} \right\rceil.
\end{aligned}$$

Thus, $d_{u_0,u_3} + d_{u_1,u_2} \geqslant \max\{d_{u_0,u_2} + d_{u_1,u_3}, d_{u_0,u_1} + d_{u_2,u_3}\}$ and using the definition of $\delta^+_{u_0,u_1,u_2,u_3}$ we have

$$\left\lceil \frac{d_{u_0,u_1} + d_{u_0,u_2} + d_{u_1,u_2}}{2} \right\rceil \leqslant d_{u_0,u_3} + d_{u_1,u_2}$$
$$\leqslant \left\lceil \frac{d_{u_0,u_1} + d_{u_0,u_2} + d_{u_1,u_2}}{2} \right\rceil + 2\,\delta^+_{u_0,u_1,u_2,u_3}.$$

Next, consider the case when $d_{u_1,u_3} = 0$. This implies

$$\begin{aligned} d_{u_0,u_1} + d_{u_1,u_3} &= d_{u_0,u_1} + d_{u_1,u_2} = d_{u_0,u_2} \\ &= \frac{d_{u_0,u_1} + d_{u_0,u_2} + d_{u_1,u_2}}{2} \\ &\leqslant \left\lceil \frac{d_{u_0,u_1} + d_{u_0,u_2} + d_{u_1,u_2}}{2} \right\rceil. \end{aligned}$$

Finally, consider the case when $d_{u_1,u_3} = d_{u_1,u_2}$. This implies

$$\begin{aligned} & d_{u_1,u_2} - \frac{d_{u_1,u_2} + d_{u_0,u_1} - d_{u_0,u_2}}{2} < 1 \\ &\equiv d_{u_0,u_2} + d_{u_1,u_2} \\ &= d_{u_0,u_1} + 2 - 2\,\varepsilon \quad \text{for some } 0 < \varepsilon \leqslant 1. \end{aligned}$$

Thus, it easily follows that

$$\begin{aligned} d_{u_0,u_3} + d_{u_1,u_2} &= d_{u_0,u_2} + d_{u_1,u_2} \\ &= \frac{d_{u_0,u_2} + d_{u_1,u_2} + d_{u_0,u_1} + 2 - 2\,\varepsilon}{2} \\ &= \frac{d_{u_0,u_2} + d_{u_1,u_2} + d_{u_0,u_1}}{2} + 1 - \varepsilon \\ &\Rightarrow d_{u_0,u_3} + d_{u_1,u_2} \\ &\leqslant \left\lceil \frac{d_{u_0,u_1} + d_{u_0,u_2} + d_{u_1,u_2}}{2} \right\rceil. \end{aligned}$$

■

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 $0 \leqslant r' < 4$. Consider the four nodes $u_0, u_{r+\lceil \frac{r'}{2} \rceil}, u_{2r+\lfloor (r'+\lceil \frac{r'}{2} \rceil)/2 \rfloor}$ and $u_{3r+r'}$. Since C has no path chord, we have $d_{u_0,u_{r+\lceil \frac{r'}{2} \rceil}} = r + \lceil \frac{r'}{2} \rceil$, $d_{u_0,u_{2r+\lfloor (r'+\lceil \frac{r'}{2} \rceil)/2 \rfloor}} = 2r + \lfloor \frac{r'+\lceil \frac{r'}{2} \rceil}{2} \rfloor d_{u_{r+\lceil r'/2 \rceil},u_{3r+r'}} = 2r + r' - \lceil \frac{r'}{2} \rceil \leqslant 2r + \lceil \frac{r'}{2} \rceil$, $d_{u_0,u_{3r+r'}} = r$, and $u_{2r+\lfloor (r'+\lceil \frac{r'}{2} \rceil)/2 \rfloor}$ is on a shortest path between u_r and $u_{3r+r'}$. Thus, applying the bound of Lemma 2, we get

$$\begin{aligned} \delta^+_{\text{worst}}(G) &\geqslant \delta^+_{u_0,\, u_{r+\lceil \frac{r'}{2} \rceil},\, u_{2r+\lfloor (r'+\lceil \frac{r'}{2} \rceil)/2 \rfloor},\, u_{3r+r'}}, \\ &\geqslant \frac{d_{u_0,u_{2r+\lfloor \frac{r'+\lceil \frac{r'}{2} \rceil}{2} \rfloor}} + d_{u_{r+\lceil \frac{r'}{2} \rceil},u_{3r+r}} - \left\lceil \frac{d_{u_0,u_{r+\lceil \frac{r'}{2} \rceil}} + d_{u_{r+\lceil \frac{r'}{2} \rceil},u_{3r+r'}} + d_{u_{3r+r},u_0}}{2} \right\rceil}{2} \\ &= \frac{4r + \left\lfloor \frac{r'+\lceil \frac{r'}{2} \rceil}{2} \right\rfloor - r' + \left\lceil \frac{r'}{2} \right\rceil - \left\lceil \frac{4r+r'}{2} \right\rceil}{2} = r + \frac{\left\lfloor \frac{r'+\lceil \frac{r'}{2} \rceil}{2} \right\rfloor - r'}{2} \\ &\geqslant r - \frac{1}{4} \;\Rightarrow\; \delta^+_{\text{worst}}(G) \geqslant r = \left\lceil \frac{k}{4} \right\rceil. \end{aligned}$$

■

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 $\Delta_{\{u_0,u_1,u_2\}}$ are three nodes satisfying the following:[5]

(1) $u_{0,1}$, $u_{0,2}$, and $u_{1,2}$ are located on the paths $\mathcal{P}_\Delta(u_0,u_1)$, $\mathcal{P}_\Delta(u_0,u_2)$, and $\mathcal{P}_\Delta(u_1,u_2)$, respectively, and

(2) the distances of these three nodes from u_0,u_1, and u_2 satisfy the following constraints:

$$\begin{aligned} & d_{u_0,u_{0,1}} + d_{u_1,u_{0,1}} = d_{u_0,u_1}, \quad d_{u_0,u_{0,2}} + d_{u_2,u_{0,2}} = d_{u_0,u_2}, \\ & d_{u_1,u_{1,2}} + d_{u_2,u_{1,2}} = d_{u_1,u_2}, \quad d_{u_1,u_{0,1}} = d_{u_1,u_{1,2}}, \\ & d_{u_0,u_{0,1}} = d_{u_0,u_{0,2}} = \left\lfloor \frac{d_{u_0,u_1} + d_{u_0,u_2} - d_{u_1,u_2}}{2} \right\rfloor. \end{aligned}$$

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.

[5]To simplify exposition, we assume that $d_{u_0,u_1} + d_{u_1,u_2} + d_{u_0,u_2}$ is an even number. Otherwise, the definition will require minor changes.

Theorem 3 (see Fig. 9 for a visual illustration). For a shortest-path triangle $\Delta_{\{u_0,u_1,u_2\}}$ and for $0 \leqslant i \leqslant 2$, let v and v' be two nodes on the paths $u_i \xleftrightarrow{\mathcal{P}_\Delta(u_i,u_{i+2 \pmod 3})} u_{i,\,i+2 \pmod 3}$ and $u_i \xleftrightarrow{\mathcal{P}_\Delta(u_i,u_{i+1 \pmod 3})} u_{i,\,i+1 \pmod 3}$, respectively, such that $d_{u_i,v} = d_{u_i,v'}$. Then,

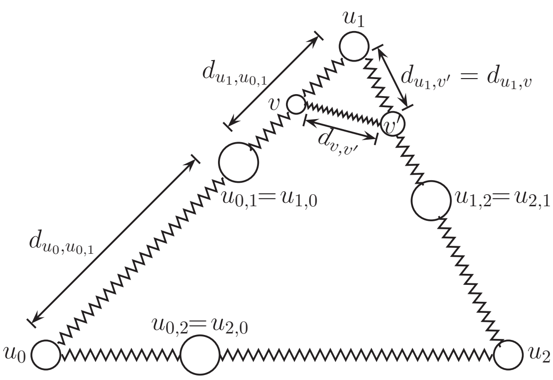
$$d_{v,v'} \leqslant 6\,\delta^+_{\Delta_{\{u_0,u_1,u_2\}}} + 2,$$

where $\delta^+_{\Delta_{\{u_0,u_1,u_2\}}} \leqslant \delta^+_{\text{worst}}(G)$ 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

$$\begin{aligned} d_H(\mathcal{P}_1,\mathcal{P}_2) &\stackrel{\text{def}}{=} \max\{\max_{v_1 \in \mathcal{P}_1} \min_{v_2 \in \mathcal{P}_2} \{d_{v_1,v_2}\}, \max_{v_2 \in \mathcal{P}_2} \min_{v_1 \in \mathcal{P}_1} \{d_{v_1,v_2}\}\} \\ &\leqslant 6\,\delta^+_{\Delta_{\{u_0,u_1,u_2\}}} + 2, \end{aligned}$$

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



$$d_{u_0,u_{0,1}} = \left\lfloor \frac{d_{u_0,u_1} + d_{u_0,u_2} - d_{u_1,u_2}}{2} \right\rfloor \quad d_{u_1,u_{0,1}} = \left\lceil \frac{d_{u_1,u_2} + d_{u_1,u_0} - d_{u_2,u_0}}{2} \right\rceil$$

$$d_{u_1,u_{0,1}} = d_{u_1,u_{1,2}} \quad d_{u_0,u_{0,1}} = d_{u_0,u_{0,2}} \quad d_{u_2,u_{0,2}} = d_{u_2,u_{1,2}}$$

$$\delta^+_{\Delta_{\{u_0,u_1,u_2\}}} \leq \delta^+_{\rm worst}(G) \quad d_{v,v'} \leq \max\left\{6\,\delta^+_{\Delta_{\{u_0,u_1,u_2\}}},\, 2\right\}$$

FIG. 9. A pictorial illustration of the claim in Theorem 3.

Proof of Theorem 3. To simplify exposition, we assume that $d_{u_0,u_1} + d_{u_1,u_2} + d_{u_0,u_2}$ is even and prove a slightly improved bound of $d_{v,v'} \leqslant 6\,\delta^+_{\Delta_{\{u_0,u_1,u_2\}}} + 1$. It is easy to modify the proof to show that $d_{v,v'} \leqslant 6\,\delta^+_{\Delta_{\{u_0,u_1,u_2\}}} + 2$ if $d_{u_0,u_1} + d_{u_1,u_2} + d_{u_0,u_2}$ is odd.

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

Case 1. $v = u_{0,1}$ and $v' = u_{1,2}$. In this case we need to prove that $d_{u_{0,1},u_{1,2}} \leqslant 6\,\delta^+_{\Delta_{\{u_0,u_1,u_2\}}} + 1$ (see Fig. 10). Assume that $d_{u_{0,1},u_{1,2}} > 0$ since otherwise the claim is trivially true. Using Lemma 2 for the four nodes $u_0,u_1,u_2,u_{1,2}$, we get

$$d_{u_0,u_{1,2}} + d_{u_1,u_2} \leqslant \left\lceil \frac{d_{u_0,u_1} + d_{u_1,u_2} + d_{u_0,u_2}}{2} \right\rceil + 2\,\delta^+_{u_0,u_1,u_2,u_{1,2}}. \tag{B1}$$

Now, we note that

$$\begin{aligned} d_{u_1,u_2} + d_{u_0,u_{0,2}} &= d_{u_1,u_2} + \left\lfloor \frac{d_{u_0,u_1} + d_{u_0,u_2} - d_{u_1,u_2}}{2} \right\rfloor \\ &= \left\lfloor \frac{d_{u_0,u_1} + d_{u_0,u_2} + d_{u_1,u_2}}{2} \right\rfloor, \end{aligned} \tag{B2}$$

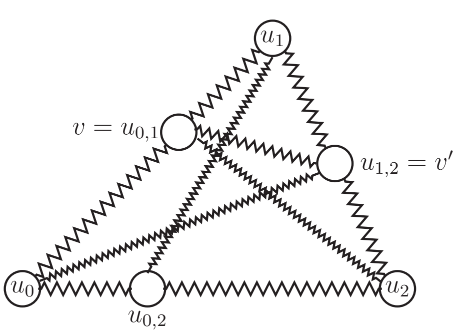


FIG. 10. Case 1 of Theorem 3. $v = u_{0,1}$, $v' = u_{1,2}$.

which in turn implies

$$\begin{aligned} \left|d_{u_0,u_{1,2}} - d_{u_0,u_{0,2}}\right| &= \left|\left(d_{u_0,u_{1,2}} + d_{u_1,u_2}\right) - \left(d_{u_1,u_2} + d_{u_0,u_{0,2}}\right)\right| \\ &\leqslant \Bigg| \underbrace{\left\lceil \frac{d_{u_0,u_1} + d_{u_1,u_2} + d_{u_0,u_2}}{2} \right\rceil + 2\,\delta^+_{u_0,u_1,u_2,u_{1,2}}}_{\text{[by inequality (B1)]}} \\ &\quad - \underbrace{\left\lfloor \frac{d_{u_0,u_1} + d_{u_0,u_2} + d_{u_1,u_2}}{2} \right\rfloor}_{\text{[by equality (B2)]}} \Bigg| \\ &\leqslant 2\,\delta^+_{u_0,u_1,u_2,u_{1,2}} + 1. \end{aligned} \tag{B3}$$

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

$$\left|d_{u_2,u_{0,1}} - d_{u_2,u_{0,2}}\right| \leqslant 2\,\delta^+_{u_0,u_1,u_2,u_{0,1}}. \tag{B4}$$

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

$$\begin{aligned} &\left|\left(d_{u_0,u_{1,2}} + d_{u_2,u_{0,1}}\right) - d_{u_0,u_2}\right| \\ &\quad = \left|\left(d_{u_0,u_{1,2}} + d_{u_2,u_{0,1}}\right) - \left(d_{u_0,u_{0,2}} + d_{u_2,u_{0,2}}\right)\right| \\ &\quad = \left|\left(d_{u_0,u_{1,2}} - d_{u_0,u_{0,2}}\right) + \left(d_{u_2,u_{0,1}} - d_{u_2,u_{0,2}}\right)\right| \\ &\quad \leqslant \left|d_{u_0,u_{1,2}} - d_{u_0,u_{0,2}}\right| + \left|d_{u_2,u_{0,1}} - d_{u_2,u_{0,2}}\right| \\ &\quad \leqslant 2\,\delta^+_{u_0,u_1,u_2,u_{1,2}} + 2\,\delta^+_{u_0,u_1,u_2,u_{0,1}} + 1. \end{aligned} \tag{B5}$$

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

$$\begin{aligned} d_{u_0,u_{0,1}} + d_{u_2,u_{1,2}} &= d_{u_0,u_{0,2}} + d_{u_2,u_{0,2}} \\ &= d_{u_0,u_2} < d_{u_0,u_2} + d_{u_{0,1},u_{1,2}}. \end{aligned} \tag{B6}$$

If $d_{u_0,u_{1,2}} + d_{u_{0,1},u_2} \leqslant d_{u_0,u_{0,1}} + d_{u_2,u_{1,2}}$ then by the definition of $\delta^+_{u_0,u_2,u_{0,1},u_{1,2}}$ we have

$$\begin{aligned} d_{u_{0,1},u_{1,2}} &= \left(d_{u_0,u_2} + d_{u_{0,1},u_{1,2}}\right) - d_{u_0,u_2} \\ &= \left(d_{u_0,u_2} + d_{u_{0,1},u_{1,2}}\right) - \left(d_{u_0,u_{0,1}} + d_{u_2,u_{1,2}}\right) \\ &\leqslant 2\,\delta^+_{u_0,u_2,u_{0,1},u_{1,2}}. \end{aligned}$$

Otherwise, $d_{u_0,u_{1,2}} + d_{u_{0,1},u_2} > d_{u_0,u_{0,1}} + d_{u_2,u_{1,2}}$ and then again by the definition of $2\,\delta^+_{u_0,u_2,u_{0,1},u_{1,2}}$ we have

$$\left|d_{u_0,u_{1,2}} + d_{u_{0,1},u_2} - d_{u_0,u_2} - d_{u_{0,1},u_{1,2}}\right| \leqslant 2\,\delta^+_{u_0,u_2,u_{0,1},u_{1,2}},$$

and now using inequality (B5) gives

$$\begin{aligned} &d_{u_{0,1},u_{1,2}} \\ &\quad = \left(d_{u_0,u_{1,2}} + d_{u_2,u_{0,1}} - d_{u_0,u_2}\right) \\ &\qquad - \left(d_{u_0,u_{1,2}} + d_{u_{0,1},u_2} - d_{u_0,u_2} - d_{u_{0,1},u_{1,2}}\right) \\ &\quad \leqslant \left|d_{u_0,u_{1,2}} + d_{u_2,u_{0,1}} - d_{u_0,u_2}\right| \\ &\qquad + \left|d_{u_0,u_{1,2}} + d_{u_{0,1},u_2} - d_{u_0,u_2} - d_{u_{0,1},u_{1,2}}\right| \\ &\quad \leqslant 2\,\delta^+_{u_0,u_1,u_2,u_{1,2}} + 2\,\delta^+_{u_0,u_1,u_2,u_{0,1}} + 2\,\delta^+_{u_0,u_2,u_{0,1},u_{1,2}} + 1 \\ &\quad \leqslant 6\,\delta^+_{\Delta_{\{u_0,u_1,u_2\}}} + 1. \end{aligned}$$

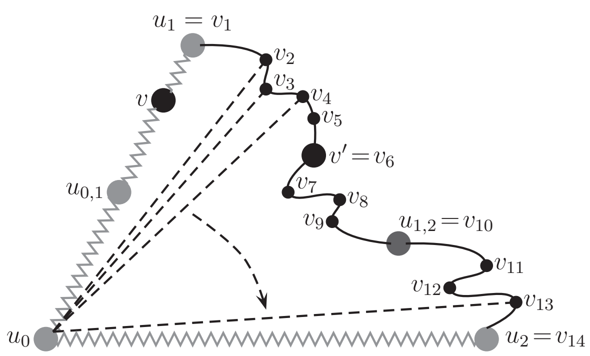


FIG. 11. Case 2 of Theorem 3: $v \neq u_{0,1}$, $v' \neq u_{1,2}$.

Case 2. $v \neq u_{0,1}$ and $v' \neq u_{1,2}$. The claim trivially holds if $d_{v,v'} \leqslant 1$, thus we assume that $d_{v,v'} > 1$. Let $(v_1 = u_1, v_2 = u_3, v_3, \ldots, v_h = v', \ldots, v_s = u_{1,2}, \ldots, v_r = u_2)$ 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 $\Delta_{\{u_0,u_1,v_2\}}, \Delta_{\{u_0,u_1,v_3\}}, \ldots, \Delta_{\{u_0,u_1,v_r\}}$, where each such triangle $\Delta_{\{u_0,u_1,v_j\}}$ is obtained by taking the shortest path $\mathcal{P}_\Delta(u_0,u_1)$, the subpath $\mathcal{P}_\Delta(u_1,v_j)$ of the shortest path $\mathcal{P}_\Delta(u_1,u_2)$, from u_1 to v_j, and a shortest path $u_0 \overset{s}{\leftrightsquigarrow} v_j$ from u_0 to v_j. Let $v_{1,j}$ be the Gromov product node on the side (shortest path) $\mathcal{P}_\Delta(u_1,v_j)$ for the shortest-path triangle $\Delta_{\{u_0,u_1,v_j\}}$.

We claim that if $v_{1,j} = v_p$ and $v_{1,j+1} = v_q$ then q is either p or $p+1$. Indeed, if $d_{u_1,v_p} = \lfloor \frac{d_{u_0,u_1} + d_{u_1,u_j} - d_{u_0,v_j}}{2} \rfloor$ and $d_{u_1,v_q} = \lfloor \frac{d_{u_0,u_1} + d_{u_1,u_{j+1}} - d_{u_0,v_{j+1}}}{2} \rfloor$ then

$$\begin{aligned} d_{u_1,v_q} - d_{u_1,v_p} &= \left\lfloor \frac{d_{u_0,u_1} + d_{u_1,v_{j+1}} - d_{u_0,v_{j+1}}}{2} \right\rfloor \\ &\quad - \left\lfloor \frac{d_{u_0,u_1} + d_{u_1,v_j} - d_{u_0,v_j}}{2} \right\rfloor \\ &\leqslant \left\lfloor \frac{d_{u_0,u_1} + (1 + d_{u_1,v_j}) - (d_{u_0,v_{j+1}} - 1)}{2} \right\rfloor \\ &\quad - \left\lfloor \frac{d_{u_0,u_1} + d_{u_1,v_j} - d_{u_0,v_j}}{2} \right\rfloor \\ &= \left\lfloor \frac{d_{u_0,u_1} + d_{u_1,v_j} - d_{u_0,v_j}}{2} + 1 \right\rfloor \\ &\quad - \left\lfloor \frac{d_{u_0,u_1} + d_{u_1,v_j} - d_{u_0,v_j}}{2} \right\rfloor \leqslant 1, \end{aligned}$$

and a similar proof of $d_{u_1,v_q} - d_{u_1,v_p} \leqslant 1$ can be obtained if $d_{u_1,v_p} = \lceil \frac{d_{u_0,u_1} + d_{u_1,u_j} - d_{u_0,v_j}}{2} \rceil$ and $d_{u_1,v_q} = \lceil \frac{d_{u_0,u_1} + d_{u_1,u_{j+1}} - d_{u_0,v_{j+1}}}{2} \rceil$. Thus, the ordered sequence of nodes $v_{1,1}, v_{1,2}, \ldots, v_{1,r}$ cover the ordered sequence of nodes $v_2, v_3, \ldots, v_s$ in a consecutive manner without skipping over any node. Since $v_{1,1}$ is either v_1 or v_2, and $v_{1,r} = v_s = u_{1,2}$, there must be an index t such that $v_{1,t} = v' = v_h$. Since $d_{u_1,v} = d_{u_1,v'}$, v, and v' are the two Gromov product nodes for the shortest-path triangle $\Delta_{\{u_0,u_1,v_t\}}$ and thus applying Case 1.1 on $\Delta_{\{u_0,u_1,v_t\}}$ we have $d_{v,v'} \leqslant 6\,\delta^+_{\Delta_{\{u_0,u_1,u_2\}}} + 1$. ■

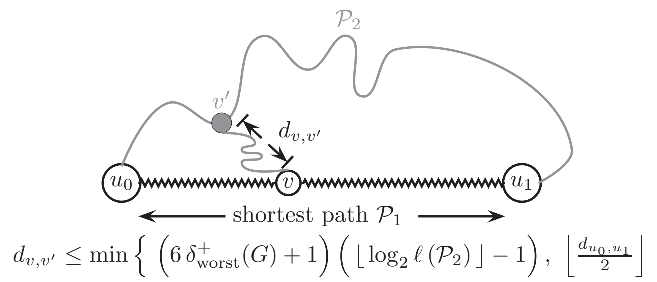


FIG. 12. Illustration of the bound in Theorem 5.

APPENDIX C: THEOREM 5 AND COROLLARY 6

Theorem 5 (see Fig. 12 for a visual illustration). Let $\mathcal{P}_1 \equiv u_0 \overset{s}{\leftrightsquigarrow} u_1$ and $\mathcal{P}_2$ be a shortest path and an arbitrary path, respectively, between two nodes u_0 and u_1. Then, for every node v on $\mathcal{P}_1$, there exists a node v' on $\mathcal{P}_2$ such that

$$\begin{aligned} d_{v,v'} &\leqslant \min \left\{ (6\,\delta^+_{\text{worst}}(G) + 2)(\lfloor \log_2 \ell(\mathcal{P}_2) \rfloor - 1), \left\lfloor \frac{d_{u_0,u_1}}{2} \right\rfloor \right\} \\ &= O(\delta^+_{\text{worst}}(G) \log \ell(\mathcal{P}_2)). \end{aligned}$$

Since $\ell(\mathcal{P}_2) \leqslant n$, the above bound also implies that

$$\begin{aligned} d_{v,v'} &\leqslant (6\,\delta^+_{\text{worst}}(G) + 2)(\lfloor \log_2 n \rfloor - 1) \\ &= O(\delta^+_{\text{worst}}(G) \log n). \end{aligned}$$

Corollary 6. Suppose that there exists a node v on the shortest path between u_0 and u_1 such that $\min_{v' \in \mathcal{P}_2} \{d_{v,v'}\} \geqslant \gamma$. Then, $\ell(\mathcal{P}_2) \geqslant 2^{\frac{\gamma}{6\,\delta^+_{\text{worst}}(G)+2}+1} - 1 = \Omega(2^{\gamma / \delta^+_{\text{worst}}(G)})$.

Proof of Theorem 5. First, note that by selecting v' to be one of u_0 or u_1 appropriately we have $d_{v,v'} \leqslant \lfloor \frac{d_{u_0,u_1}}{2} \rfloor$. Now, assume that $\ell(\mathcal{P}_2) > 2$. Let u_2 be the node on the path $\mathcal{P}_2$ such that $\ell(u_0 \overset{\mathcal{P}_2}{\leftrightsquigarrow} u_2) = \lceil \frac{\ell(\mathcal{P}_2)}{2} \rceil$. and consider the shortest-path triangle $\Delta_{\{u_0,u_1,u_2\}}$. 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 $d_{v,v'} \leqslant 6\,\delta^+_{\text{worst}}(G) + 2$. 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 $\mathcal{P}_2$ containing at most $\lceil \frac{(\mathcal{P}_2)}{2} \rceil$ 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

$$D(y) \leqslant D\left(\left\lceil \frac{y}{2} \right\rceil\right) + 6\,\delta^+_{\text{worst}}(G) + 2, \quad \text{if } y > 2, \quad D(2) = 1.$$

A solution to the above recurrence satisfies $D(\ell(\mathcal{P}_2)) \leqslant (6\,\delta^+_{\text{worst}}(G) + 2)(\lceil \log_2 \ell(\mathcal{P}_2) \rceil - 1)$. ■

APPENDIX D: THEOREM 7 AND COROLLARY 8

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

Theorem 7. Let $\mathcal{P}_1$ and $\mathcal{P}_2$ be a shortest path and another path, respectively, between two nodes. Define $\eta_{\mathcal{P}_1,\mathcal{P}_2}$ as

$$\begin{aligned}\eta_{\mathcal{P}_1,\mathcal{P}_2} &= (6\,\delta^+ + 2)\log_2((6\,\mu+2)(6\,\delta^+ + 2) \\ &\quad\times \log_2[(6\,\delta^+ + 2)(3\,\mu+1)\,\mu] + \mu) \\ &= \mathrm{O}(\delta^+ \log(\mu\,\delta^+)), \text{ if } \mathcal{P}_2 \text{ is } \mu\text{-approximate short.}\end{aligned}$$

$$\begin{aligned}\eta_{\mathcal{P}_1,\mathcal{P}_2} &= (6\,\delta^+ + 2)\log_2\left(8(6\,\delta^+ + 2)\,\log_2[(6\,\delta^+ + 2)(4+2\varepsilon)] \right. \\ &\quad \left. + 1 + \frac{\varepsilon}{2}\right) \\ &= \mathrm{O}(\delta^+ \log(\varepsilon + \delta^+\,\log\varepsilon)), \\ &\quad \text{if } \mathcal{P}_2 \text{ is } \varepsilon\text{-additive-approximate short}\end{aligned}$$

Then, the following statements are true.

(a) For every node v on $\mathcal{P}_1$, there exists a node v' on $\mathcal{P}_2$ such that $d_{v,v'} \leqslant \lfloor \eta_{\mathcal{P}_1,\mathcal{P}_2} \rfloor$.

(b) For every node v' on $\mathcal{P}_2$, there exists a node v on $\mathcal{P}_1$ such that $d_{v,v'} \leqslant \zeta_{\mathcal{P}_1,\mathcal{P}_2}$ where

$$\zeta_{\mathcal{P}_1,\mathcal{P}_2} = \begin{cases}\min\left\{\left\lfloor(\mu+1)\,\eta_{\mathcal{P}_1,\mathcal{P}_2} + \frac{\mu}{2}\right\rfloor, \left\lfloor\frac{\mu\,d_{u_0,u_1}}{2}\right\rfloor\right\} \\ = \mathrm{O}(\mu\delta^+\log(\mu\delta^+)), \text{ if } \mathcal{P}_2 \text{ is } \mu\text{-approximate short} \\ \min\left\{\left\lfloor 2\,\eta_{\mathcal{P}_1,\mathcal{P}_2} + \frac{1+\varepsilon}{2}\right\rfloor, \left\lfloor\frac{d_{u_0,u_1}+\varepsilon}{2}\right\rfloor\right\} \\ = \mathrm{O}(\varepsilon + \delta^+\log(\varepsilon + \delta^+\,\log\varepsilon)), \\ \qquad \text{if } \mathcal{P}_2 \text{ is } \varepsilon\text{-additive-approximate short.}\end{cases}$$

Corollary 8 (Hausdorff distance between approximate short paths). Suppose that $\mathcal{P}_1$ and $\mathcal{P}_2$ are two paths between two nodes. Then, the Hausdorff distance $d_H(\mathcal{P}_1,\mathcal{P}_2)$ between these two paths can be bounded as follows:

$$\begin{aligned}&d_H(\mathcal{P}_1,\mathcal{P}_2) \\ &\quad \stackrel{\text{def}}{=} \max\left\{\max_{v_1\in\mathcal{P}_1}\min_{v_2\in\mathcal{P}_2}\{d_{v_1,v_2}\},\ \max_{v_2\in\mathcal{P}_2}\min_{v_1\in\mathcal{P}_1}\{d_{v_1,v_2}\}\right\} \\ &\quad \leqslant \eta_{\mathcal{P}_1,u_0 \overset{\mathfrak{s}}{\leftrightsquigarrow} u_1} + \zeta_{\mathcal{P}_2,u_0 \overset{\mathfrak{s}}{\leftrightsquigarrow} u_1}.\end{aligned}$$

Corollary 9. Suppose that there exists a node v on the shortest path between u_0 and u_1 such that $\min_{v'\in\mathcal{P}_2}\{d_{v,v'}\} \geqslant \gamma$. Then, the following is true.

If $\mathcal{P}_2$ is a μ-approximate short path then

$$\mu > \frac{2^{\frac{\gamma}{6\,\delta^+ + 1}}}{12\,\gamma - (24 + \mathrm{o}(1))\,(6\,\delta^+ + 1)} - \frac{1}{3} \Rightarrow \mu = \Omega\left(\frac{2^{\frac{\gamma}{\delta^+}}}{\gamma}\right).$$

If $\mathcal{P}_2$ is a ε-additive-approximate short path then

$$\begin{aligned}\varepsilon &> \frac{2^{\frac{\gamma}{6\delta^+ + 1}}}{\left(48\,\delta^+ + \frac{17}{2}\right)} - \log_2(48\,\delta^+ + 8) \\ &\Rightarrow \varepsilon = \Omega\left(\frac{2^{\frac{\gamma}{\delta^+}}}{\delta^+} - \log\delta^+\right).\end{aligned}$$

In particular, assuming real world networks have small constant values of δ^+, the asymptotic dependence of μ and ε on γ can be summarized as

both μ and ε are $\Omega(2^{c\,\gamma})$ for some constant $0 < c < 1$.

Proof of Theorem 7. Let $\mathcal{P}_1$ and $\mathcal{P}_2$ be a shortest path and another path, respectively, between two nodes u_0 and u_1. Note that any "subpath" of a μ-approximate short path is also a μ-approximately short path, i.e., $u_i \overset{\mathcal{P}}{\leftrightsquigarrow} u_j$ is also a μ-approximate short path, and similarly any subpath of a ε-additive-approximate short path is also a ε-additive-approximate short path. μ-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 μ-approximate short path and $\{u_i,u_j\} \in E$ then $|j-i| \leqslant \mu$.

(a) Let v and v' be two nodes on $\mathcal{P}_1$ and $\mathcal{P}_2$, respectively, such that $\alpha = d_{v,v'} = \max_{v''\in\mathcal{P}_1}\min_{v'''\in\mathcal{P}_2}\{d_{v'',v'''}\}$. Let $v_\ell \in u_0 \overset{\mathcal{P}_1}{\leftrightsquigarrow} v$ and $v_r \in u_1 \overset{\mathcal{P}_1}{\leftrightsquigarrow} v$ be two nodes defined by

$$d_{v_\ell,v} = \begin{cases} 2\,\alpha + 1, & \text{if } d_{u_0,v} > 2\,\alpha+1, \\ d_{u_0,v}, & \text{otherwise}\end{cases}$$

$$d_{v_r,v} = \begin{cases} 2\,\alpha + 1, & \text{if } d_{u_1,v} > 2\,\alpha+1. \\ d_{u_1,v}, & \text{otherwise}\end{cases}$$

By definition of α, there exist two nodes $\widetilde{v_\ell}$ and $\widetilde{v_r}$ on the path $\mathcal{P}_2$ such that $d_{v_\ell,\widetilde{v_\ell}}, d_{v_r,\widetilde{v_r}} \leqslant \alpha$. Consider the $\mathcal{P}_3 = \widetilde{v_\ell} \overset{\mathcal{P}_2}{\leftrightsquigarrow} \widetilde{v_r}$ that is the part of path $\mathcal{P}_2$ from $\widetilde{v_\ell}$ to $\widetilde{v_r}$. Note that

$$d_{\widetilde{v_\ell},\widetilde{v_r}} \leqslant d_{\widetilde{v_\ell},v_\ell} + d_{v_\ell,v_r} + d_{v_r,\widetilde{v_r}} \leqslant 6\,\alpha + 2.$$

Thus, we arrive at the following inequalities:

$$\ell(\mathcal{P}_3) \leqslant \begin{cases}(6\,\alpha+2)\,\mu, & \text{if } \mathcal{P}_2 \text{ is } \mu\text{-approximate short} \\ 6\,\alpha + 2 + \varepsilon, & \text{if } \mathcal{P}_2 \text{ is } \varepsilon\text{-additive-approximate short.}\end{cases}$$

Now consider the path $\mathcal{P}_4 = v_\ell \overset{\mathfrak{s}}{\leftrightsquigarrow} \widetilde{v_\ell} \overset{\mathcal{P}_2}{\leftrightsquigarrow} \widetilde{v_r} \overset{\mathfrak{s}}{\leftrightsquigarrow} v_r$ obtained by taking a shortest path from v_ℓ to $\widetilde{v_\ell}$ followed by the path $\mathcal{P}_3$ followed by a shortest path from v_r to $\widetilde{v_r}$. Note that

$$\ell(\mathcal{P}_4) \leqslant \begin{cases}(6\,\alpha+2)\,\mu + 2\,\alpha, & \text{if } \mathcal{P}_2 \text{ is } \mu\text{-approximate short} \\ 6\,\alpha + 2 + \varepsilon + 2\,\alpha = 8\,\alpha + 2 + \varepsilon, \\ \qquad \text{if } \mathcal{P}_2 \text{ is } \varepsilon\text{-additive-approximate short.}\end{cases}$$

We claim that $\min_{\widetilde{v}\in\mathcal{P}_4}\{d_{v,\widetilde{v}}\} = \alpha$. Indeed, if $\widetilde{v} \in \mathcal{P}_3$ then, by definition of α, $\min_{\widetilde{v}}\{d_{v,\widetilde{v}}\} = \alpha$. Otherwise, if $\widetilde{v} \in v_\ell \overset{\mathfrak{s}}{\leftrightsquigarrow} \widetilde{v_\ell}$, then by triangle inequality $d_{v_\ell,v} \leqslant d_{v,\widetilde{v}} + d_{\widetilde{v},v_\ell} \Rightarrow d_{v,\widetilde{v}} \geqslant 2\,\alpha + 1 - d_{\widetilde{v},v_\ell} > \alpha$. Similarly, if $\widetilde{v} \in \widetilde{v_r} \overset{\mathfrak{s}}{\leftrightsquigarrow} v_r$, then by triangle inequality $d_{v_r,v} \leqslant d_{v,\widetilde{v}} + d_{\widetilde{v},v_r} \Rightarrow d_{v,\widetilde{v}} \geqslant 2\,\alpha + 1 - d_{\widetilde{v},v_r} > \alpha$. Since $v_\ell \overset{\mathcal{P}_1}{\leftrightsquigarrow} v_r$ is a shortest path between v_ℓ and v_r and v is a node on this path, by Theorem 5, $\alpha \leqslant (6\,\delta^+ + 2)(\lfloor\log_2 \ell(\mathcal{P}_4)\rfloor - 1)$. Thus, we have the following inequalities:

If $\mathcal{P}_2$ is a μ-approximate short path then

$$\begin{aligned}\ell(\mathcal{P}_4) &\leqslant (6\,\alpha+2)\,\mu + 2\,\alpha = (6\,\mu+2)\,\alpha + 2\,\mu \\ &\leqslant (6\,\mu+2)\,(6\,\delta^+ + 2)(\log_2 \ell(\mathcal{P}_4) - 1) + 2\,\mu \\ &\leqslant (6\,\mu+2)\,(6\,\delta^+ + 2)[\log_2((6\,\mu+2)\,\alpha + 2\,\mu) - 1] \\ &\quad + 2\,\mu \\ &\Rightarrow \alpha \leqslant (6\,\delta^+ + 2)[\log_2((3\,\mu+1)\,\alpha + \mu)]. \qquad \text{(D1)}\end{aligned}$$

If $\mathcal{P}_2$ is a ε-additive-approximate short path then

$$\begin{aligned}\ell(\mathcal{P}_4) &\leqslant 8\,\alpha + 2 + \varepsilon\\ &\leqslant 8\,(6\,\delta^+ + 2)(\log_2 \ell(\mathcal{P}_4) - 1) + 2 + \varepsilon\\ &\leqslant 8\,(6\,\delta^+ + 2)(\log_2(8\,\alpha + 2 + \varepsilon)\, - 1) + 2 + \varepsilon\\ &\Rightarrow 8\,\alpha + 2 + \varepsilon\\ &\leqslant 8\,(6\,\delta^+ + 2)(\log_2(8\,\alpha + 2 + \varepsilon)\, - 1) + 2 + \varepsilon\\ &\equiv \alpha \leqslant (6\,\delta^+ + 2)\left(\log_2\left(4\,\alpha + 1 + \frac{\varepsilon}{2}\right)\right).\end{aligned} \tag{D2}$$

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

$$a = 6\,\delta^+ + 2 \geqslant 1 \quad \text{for both (D1) and (D2),}$$

$$b = \begin{cases} 3\,\mu + 1 \geqslant 4 & \text{for} \quad \text{(D1)}\\ 4 & \text{for} \quad \text{(D2)}\end{cases}$$

$$c = \begin{cases} \mu \geqslant 1 & \text{for} \quad \text{(D1)}\\ 1 + \frac{\varepsilon}{2} \geqslant 1 & \text{for} \quad \text{(D2).}\end{cases}$$

Thus, α is at most z_0 where z_0 is the largest positive integer value of z that satisfies the equation,

$$2^{\frac{z}{a}} \leqslant b\,z + c.$$

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

$$x \geqslant 1 \ \&\ y \geqslant 1 \Rightarrow y\,(x - 1) \geqslant x - 1 \equiv x\,y + 1 \geqslant x + y.$$

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

$$\begin{aligned} 2^{\frac{\eta}{a}} &= 2^{\log_2(2\,a\,b\log_2(a\,b\,c)+c)} = 2\,a\,b\,\log_2(a\,b\,c) + c\\ b\,\eta + c &= a\,b\,[\log_2(2\,a\,b\log_2(a\,b\,c) + c)] + c,\\ 2^{\frac{\eta}{a}} &> b\,\eta + c\\ &\equiv 2\,a\,b\,\log_2(a\,b\,c) + c\\ &\geqslant a\,b\,[\log_2(2\,a\,b\log_2(a\,b\,c) + c)] + c\\ &\equiv 2\,\log_2(a\,b\,c)\\ &\geqslant \log_2(2\,a\,b\log_2(a\,b\,c) + c)\\ &\Leftarrow 2\,\log_2(a\,b\,c)\\ &\geqslant \log_2(2\,a\,b\,c\log_2(a\,b\,c) + 1)\\ &\qquad \text{since } 2\,a\,b\log_2(a\,b\,c)\\ &\geqslant 1 \text{ and } c \geqslant 1\\ &\equiv (a\,b\,c)^2\\ &\geqslant 2\,a\,b\,c\log_2(a\,b\,c) + 1\\ &\Leftarrow a\,b\,c \geqslant \log_2(a\,b\,c) + 1,\end{aligned}$$

and the very last inequality holds since $a\,b\,c \geqslant 4$. Thus, we arrive at the following bounds:

If $\mathcal{P}_2$ is a μ-approximate short path then

$$\eta = (6\,\delta^+ + 2)\,\log_2((6\,\mu + 2)(6\,\delta^+ + 2) \times \log_2[(6\,\delta^+ + 2)(3\,\mu + 1)\,\mu] + \mu).$$

If $\mathcal{P}_2$ is a ε-additive-approximate short path then

$$\eta = (6\,\delta^+ + 2)\,\log_2\Bigg(8\,(6\,\delta^+ + 2) \times \log_2[(6\,\delta^+ + 2)(4 + 2\varepsilon)] + 1 + \frac{\varepsilon}{2}\Bigg).$$

(b) Let the ordered sequence of nodes in the path $\mathcal{P}_3 = v_1 \overset{\mathcal{P}_2}{\leftrightsquigarrow} v_1'$ be a (length) maximal sequence of nodes such that

$$\forall\, v' \in \mathcal{P}_3 : \min_{v \in \mathcal{P}_1}\{d_{v,v'}\} > Z_{\mathcal{P}_1,\mathcal{P}_2}.$$

Consider the following set of nodes belonging to the two paths $u_0 \overset{\mathcal{P}_2}{\leftrightsquigarrow} v_1$ and $v_1' \overset{\mathcal{P}_2}{\leftrightsquigarrow} u_1$:

$$\mathcal{S}_\ell = \bigcup\{v' \in u_0 \overset{\mathcal{P}_2}{\leftrightsquigarrow} v_1 | \exists\, v \in \mathcal{P}_1 : d_{v,v'} = \min_{v'' \in \mathcal{P}_2}\{d_{v,v''}\}\},$$

$$\mathcal{S}_r = \bigcup\{v' \in v_1' \overset{\mathcal{P}_2}{\leftrightsquigarrow} u_1 | \exists\, v \in \mathcal{P}_1 : d_{v,v'} = \min_{v'' \in \mathcal{P}_2}\{d_{v,v''}\}\}.$$

Since $u_0 \in \mathcal{S}_\ell$ and $u_1 \in \mathcal{S}_r$, it follows that $\mathcal{S}_\ell \neq \emptyset$ and $\mathcal{S}_r \neq \emptyset$. Note that

$$\bigcup\{v \in u_0 \overset{\mathcal{P}_1}{\leftrightsquigarrow} u_1 | \exists\, v' \in \mathcal{S}_\ell \cup \mathcal{S}_r : d_{v,v'} = \min_{v'' \in \mathcal{P}_2}\{d_{v,v''}\}\} = \bigcup_{v \in u_0 \overset{\mathcal{P}_1}{\leftrightsquigarrow} u_1} \{v\}.$$

Thus, there exist two adjacent nodes v_4 and v_4' on $\mathcal{P}_1$ such that both d_{v_4,v_3} and $d_{v_4',v_3'}$ are at most $Z_{\mathcal{P}_1,\mathcal{P}_2}$. Using triangle inequality it follows that

$$d_{v_3,v_3'} \leqslant d_{v_3,v_4} + d_{v_4,v_4'} + d_{v_4',v_3'} = 2\,Z_{\mathcal{P}_1,\mathcal{P}_2} + 1,$$

giving the following bounds:

$$\ell(v_3 \overset{\mathcal{P}_2}{\leftrightsquigarrow} v_3') \leqslant \begin{cases} \mu\, d_{v_3,v_3'} \leqslant 2\,\mu\, Z_{\mathcal{P}_1,\mathcal{P}_2} + \mu,\\ \quad \text{if } \mathcal{P}_2 \text{ is } \mu\text{-approximate short}\\ d_{v_3,v_3'} + \varepsilon \leqslant 2\,Z_{\mathcal{P}_1,\mathcal{P}_2} + 1 + \varepsilon,\\ \quad \text{if } \mathcal{P}_2 \text{ is } \varepsilon\text{-additive-approximate short.}\end{cases}$$

For any node v' on $\mathcal{P}_3$, we can always use the following path to reach a node on $\mathcal{P}_1$:

(1) If $d_{v',v_3} \leqslant d_{v',v_3'}$ then we take the path $v' \overset{\mathcal{P}_2}{\leftrightsquigarrow} v_3 \overset{\mathfrak{s}}{\leftrightsquigarrow} v_4$ of length at most $\lfloor \frac{\ell(v_3 \overset{\mathcal{P}_2}{\leftrightsquigarrow} v_3')}{2} \rfloor + Z_{\mathcal{P}_1,\mathcal{P}_2}$ to reach the node $v = v_4$ on $\mathcal{P}_1$;

(2) otherwise we take the path $v' \overset{\mathcal{P}_2}{\leftrightsquigarrow} v_3' \overset{\mathfrak{s}}{\leftrightsquigarrow} v_4'$ of length at most $\lfloor \frac{\ell\,(v_3 \overset{\mathcal{P}_2}{\leftrightsquigarrow} v_3')}{2} \rfloor + Z_{\mathcal{P}_1,\mathcal{P}_2}$ to reach the node $v = v_4'$ on $\mathcal{P}_1$.

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

$$d_{v,v'} \leqslant \begin{cases} \left\lfloor (\mu + 1)\, Z_{\mathcal{P}_1,\mathcal{P}_2} + \frac{\mu}{2} \right\rfloor,\\ \quad \text{if } \mathcal{P}_2 \text{ is } \mu\text{-approximate short}\\ \left\lfloor 2\, Z_{\mathcal{P}_1,\mathcal{P}_2} + \frac{1+\varepsilon}{2} \right\rfloor,\\ \quad \text{if } \mathcal{P}_2 \text{ is } \varepsilon\text{-additive-approximate short.}\end{cases}$$

■

APPENDIX E: THEOREM 10 AND COROLLARY 11

Theorem 10 (see Fig. 13 for a visual illustration). Suppose that we are given the following:

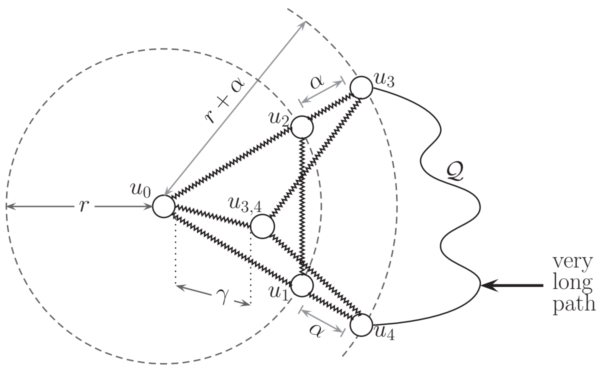


FIG. 13. Illustration of the claims in Theorem 10 and Corollary 11.

three integers $\kappa \geqslant 4$, $\alpha > 0$, $r > (\frac{\kappa}{2} - 1)(6\,\delta^{+}_{\rm worst}(G) + 2)$,
five nodes u_0,u_1,u_2,u_3,u_4 such that
$u_1,u_2 \in B_r(u_0)$ with $d_{u_1,u_2} \geqslant \frac{\kappa}{2}\,(6\,\delta^{+}_{\rm worst}(G) + 2)$, $d_{u_1,u_4} = d_{u_2,u_3} = \alpha$.

Then, the following statements are true for any shortest path $\mathcal{P}$ between u_3 and u_4:

(a) there exists a node v on $\mathcal{P}$ such that

$$\begin{aligned} d_{u_0,v} &\leqslant r - \left(\frac{3\kappa - 2}{12}\right)(6\,\delta^{+}_{\rm worst}(G) + 2) \\ &= r - \mathrm{O}(\kappa\,\delta^{+}_{\rm worst}(G)). \end{aligned}$$

(b)

$$\begin{aligned} \ell(\mathcal{P}) \geqslant \left(\frac{3\kappa - 2}{6}\right)(6\,\delta^{+}_{\rm worst}(G) + 2) \\ + 2\,\alpha = \Omega\,(\kappa\,\delta^{+}_{\rm worst}(G) + \alpha). \end{aligned}$$

Corollary 11 (see Fig. 13 for a visual illustration). Consider any path $\mathcal{Q}$ between u_3 and u_4 that does not involve a node in $\cup_{r' \leqslant r} \mathcal{B}_{r'}(u_0)$. Then, the following statements hold:

(i) $\ell(\mathcal{Q}) \geqslant 2^{\frac{\alpha}{6\,\delta^{+}_{\rm worst}(G)+2} + \frac{\kappa}{4} + \frac{5}{6}} - 1 = 2^{\Omega\,(\frac{\alpha}{\delta^{+}_{\rm worst}(G)} + \kappa)}$.
In particular, if $\delta^{+}_{\rm worst}(G)$ is a constant then $\ell(\mathcal{Q}) = 2^{\Omega(\alpha + \kappa)}$ and thus $\ell(\mathcal{Q})$ increases at least exponentially with both α and κ.

(ii) if $\mathcal{Q}$ is a μ-approximate short path then

$$\begin{aligned} \mu &\geqslant \frac{2^{\frac{\alpha}{6\,\delta^{+}_{\rm worst}(G)+2} + \frac{\kappa}{4} - \frac{1}{6}}}{12\,\alpha + (3\,\kappa - 26 - \mathrm{o}(1))(6\,\delta^{+}_{\rm worst}(G) + 2)} - \frac{1}{3} \\ &= \Omega\left(\frac{2^{\Theta\left(\frac{\alpha}{\delta^{+}_{\rm worst}(G)} + \kappa\right)}}{\alpha + \kappa\,\delta^{+}_{\rm worst}(G)}\right). \end{aligned}$$

In particular, if $\delta^{+}_{\rm worst}(G)$ is a constant then $\mu = \Omega(\frac{2^{\Theta(\alpha+\kappa)}}{\alpha+\kappa})$ and thus μ increases at least exponentially with both α and κ.

(iii) if $\mathcal{Q}$ is a ε-additive-approximate short path then

$$\varepsilon > \frac{2^{\frac{\alpha}{6\,\delta^{+}_{\rm worst}(G)+2}} + \frac{\kappa}{4} - \frac{1}{6}}{48\,\delta^{+}_{\rm worst}(G) + \frac{17}{2}} - \log_2(48\,\delta^{+}_{\rm worst}(G) + 16).$$

In particular, if $\delta^{+}_{\rm worst}(G)$ is a constant then $\varepsilon = \Omega(2^{\Theta\,(\alpha+\kappa)})$ and thus ε increases at least exponentially with both α and κ.

1. Proof of Theorem 10.

Consider the shortest-path triangle $\Delta_{\{u_0,u_3,u_4\}}$ and let $u_{0,3},u_{0,4}$, and $u_{3,4}$ 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, $d_{u_0,u_{0,3}} = d_{u_0,u_{0,4}}$, and $\beta = d_{u_3,u_{3,4}} = \lfloor \frac{d_{u_0,u_3} + d_{u_3,u_4} - d_{u_0,u_4}}{2} \rfloor = \lfloor \frac{d_{u_3,u_4}}{2} \rfloor$ since $d_{u_0,u_3} = d_{u_0,u_4} = r + \alpha$.

We first claim that $d_{u_0,u_{0,3}} < r = d_{u_0,u_2}$. Suppose for the sake of contradiction that $d_{u_0,u_{0,3}} = d_{u_0,u_{0,4}} \geqslant r$. Then, by Theorem 3 we get $d_{u_1,u_2} \leqslant 6\,\delta^{+}_{\rm worst}(G) + 2$ which contradicts the assumption that $d_{u_1,u_2} \geqslant \frac{\kappa}{2}\,(6\,\delta^{+}_{\rm worst}(G) + 2)$ since $\kappa \geqslant 4$.

Thus, assume that $d_{u_0,u_{0,3}} = d_{u_0,u_{0,4}} = r - x$ for some integer $x > 0$. By Theorem 3, $d_{u_{0,3},u_{0,4}} \leqslant 6\,\delta^{+}_{\rm worst}(G) + 2$. Let $d_{u_{0,3},u_{0,4}} = 6\,\delta^{+}_{\rm worst}(G) + 2 - y$ for some integer $0 < y \leqslant 6\,\delta^{+}_{\rm worst}(G) + 2$ and $d_{u_1,u_2} = \frac{\kappa}{2}\,(6\,\delta^{+}_{\rm worst}(G) + 2) + z$ for some integer $z \geqslant 0$. Consider the four-node condition for the four nodes $u_1,u_2,u_{0,3},u_{0,4}$. The three relevant quantities for comparison are

$$\begin{aligned} q_{\parallel} &= d_{u_1,u_2} + d_{u_{0,3},u_{0,4}} \\ &= \left(\frac{\kappa}{2} + 1\right)(6\,\delta^{+}_{\rm worst}(G) + 1) + z - y, \\ q_{=} &= d_{u_{0,3},u_2} + d_{u_{0,4},u_1} \\ &= \left(d_{u_0,u_2} - d_{u_0,u_{0,3}}\right) + \left(d_{u_0,u_1} - d_{u_0,u_{0,4}}\right) = 2x, \\ q_{\backslash\backslash} &= d_{u_{0,3},u_1} + d_{u_{0,4},u_2} \\ &\leqslant \left(d_{u_{0,3},u_{0,4}} + d_{u_{0,4},u_1}\right) + \left(d_{u_{0,3},u_{0,4}} + d_{u_{0,3},u_2}\right) \\ &= 12\,\delta^{+}_{\rm worst}(G) + 4 - 2y + 2x. \end{aligned}$$

We now show that $x > (\frac{3\kappa-2}{12})(6\,\delta^{+}_{\rm worst}(G) + 2)$. We have the following cases.

Assume that $q_{\backslash\backslash} \leqslant \min\{q_{\parallel}, q_{=}\}$. This implies

$$\begin{aligned} |q_{\parallel} - q_{=}| &\leqslant 2\,\delta^{+}_{\rm worst}(G) \\ &\equiv \left|\left(\frac{\kappa}{2} + 1\right)(6\,\delta^{+}_{\rm worst}(G) + 2) + z - y - 2x\right| \\ &\leqslant 2\,\delta^{+}_{\rm worst}(G) \\ \Rightarrow x &\geqslant \frac{(\frac{\kappa}{2} + 1)(6\,\delta^{+}_{\rm worst}(G) + 2) + z - y - 2\,\delta^{+}_{\rm worst}(G)}{2} \\ &\geqslant \left(\frac{3\kappa - 2}{12}\right)(6\,\delta^{+}_{\rm worst}(G) + 2) + \frac{1}{6}. \end{aligned}$$

Otherwise, assume that $q_{=} \leqslant \min\{q_{\parallel}, q_{\backslash\backslash}\}$. This implies

$$\begin{aligned} |q_{\parallel} - q_{\backslash\backslash}| \leqslant 2\,\delta^{+}_{\rm worst}(G) &\Rightarrow q_{\backslash\backslash} \geqslant q_{\parallel} - 2\,\delta^{+}_{\rm worst}(G) \\ &\Rightarrow d_{u_{0,3},u_1} + d_{u_{0,4},u_2} \geqslant \left(\frac{\kappa}{2} + 1\right)(6\,\delta^{+}_{\rm worst}(G) + 2) + z - y - 2\,\delta^{+}_{\rm worst}(G) \end{aligned}$$

$$\begin{aligned}
&\Rightarrow \left(d_{u_{0,3},u_{0,4}} + d_{u_{0,4},u_1}\right) + \left(d_{u_{0,3},u_{0,4}} + d_{u_{0,3},u_2}\right) \geqslant d_{u_{0,3},u_1} + d_{u_{0,4},u_2}\\
&\geqslant \left(\frac{\kappa}{2}+1\right)(6\,\delta^+_{\text{worst}}(G)+2) + z - y - 2\,\delta^+_{\text{worst}}(G)\\
&\Rightarrow 2x + 2(6\,\delta^+_{\text{worst}}(G)+2-y) \geqslant \left(\frac{\kappa}{2}+1\right)(6\,\delta^+_{\text{worst}}(G)+2) + z - y - 2\,\delta^+_{\text{worst}}(G)\\
&\Rightarrow x \geqslant \left(\frac{3\kappa-2}{12}\right)(6\,\delta^+_{\text{worst}}(G)+2) + \frac{1}{6}.
\end{aligned}$$

Otherwise, assume that $q_{\|} \leqslant \min\{q_{=}, q_{\backslash\backslash}\}$. This implies

$$\begin{aligned}
|q_{=} - q_{\backslash\backslash}| \leqslant 2\,\delta^+_{\text{worst}}(G) \;&\equiv\; \left|\,2x - \left(d_{u_{0,3},u_1} + d_{u_{0,4},u_2}\right)\right| \leqslant 2\,\delta^+_{\text{worst}}(G)\\
&\Rightarrow 2x \geqslant d_{u_{0,3},u_1} + d_{u_{0,4},u_2} - 2\,\delta^+_{\text{worst}}(G) \geqslant \left(d_{u_1,u_2} - d_{u_{0,4},u_1}\right) + \left(d_{u_1,u_2} - d_{u_{0,3},u_1}\right) - 2\,\delta^+_{\text{worst}}(G)\\
&\equiv 2x \geqslant \kappa(6\,\delta^+_{\text{worst}}(G)+2) + 2z - 2x - 2\,\delta^+_{\text{worst}}(G)\\
&\Rightarrow x \geqslant \left(\frac{3\kappa-2}{12}\right)(6\,\delta^+_{\text{worst}}(G)+2) + \frac{\delta^+_{\text{worst}}(G)}{2} + \frac{1}{6}.
\end{aligned}$$

Using Theorem 3, it now follows that

$$d_{u_0,u_{3,4}} \leqslant d_{u_0,u_{0,3}} + d_{u_{0,3},u_{0,4}} \leqslant (r-x) + (6\,\delta^+_{\text{worst}}(G)+2) < r - \left(\frac{3\kappa-2}{12}\right)(6\,\delta^+_{\text{worst}}(G)+2).$$

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

$$|\mathcal{P}| = 2\,\beta \geqslant 2(r+\alpha) - 2d_{u_0,u_{3,4}} \geqslant 2\alpha + \left(\frac{3\kappa-2}{6}\right)(6\,\delta^+_{\text{worst}}(G)+2).$$

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 $r+\alpha$ from u_0, by Theorem 10 the following holds for every node $v \in Q$:

$$d_{u_{3,4},v} \geqslant (r+\alpha) - d_{u_0,u_{3,4}} = (r+\alpha) - \left(r - \left(\frac{3\kappa-2}{12}\right)(6\,\delta^+_{\text{worst}}(G)+2)\right) = \alpha + \left(\frac{3\kappa-2}{12}\right)(6\,\delta^+_{\text{worst}}(G)+2).$$

Thus, by Corollary 6 [with $\gamma = \alpha + (\frac{3\kappa-2}{12})(6\,\delta^+_{\text{worst}}(G)+2)$], we get

$$\ell(Q) \geqslant 2^{\frac{\gamma}{6\,\delta^+_{\text{worst}}(G)+2}+1} - 1 = 2^{\frac{\alpha}{6\,\delta^+_{\text{worst}}(G)+2} + \frac{\kappa}{4} + \frac{5}{6}} - 1.$$

If Q is a μ-approximate short path, then by Corollary 9,

$$\mu > \frac{2^{\frac{\gamma}{6\,\delta^+_{\text{worst}}(G)+2}}}{12\,\gamma - (24+\mathrm{o}(1))\,(6\,\delta^+_{\text{worst}}(G)+2)} - \frac{1}{3} = \frac{2^{\frac{\alpha}{6\,\delta^+_{\text{worst}}(G)+2}+\frac{\kappa}{4}-\frac{1}{6}}}{12\,\alpha + (3\,\kappa - 26 - \mathrm{o}(1))(6\,\delta^+_{\text{worst}}(G)+2)} - \frac{1}{3}.$$

If Q is a ε-additive-approximate short path, then by Corollary 9,

$$\varepsilon > \frac{2^{\frac{\gamma}{6\delta^+_{\text{worst}}(G)+2}}}{48\,\delta^+_{\text{worst}}(G) + \frac{17}{2}} - \log_2(48\,\delta^+_{\text{worst}}(G)+16) = \frac{2^{\frac{\alpha}{6\,\delta^+_{\text{worst}}(G)+2}+\frac{\kappa}{4}-\frac{1}{6}}}{48\,\delta^+_{\text{worst}}(G) + \frac{17}{2}} - \log_2(48\,\delta^+_{\text{worst}}(G)+16).$$

APPENDIX F: LEMMA 12

Lemma 12 (equivalence of strong and weak domination; see Fig. 7 for a visual illustration). If $\lambda \geqslant (6\,\delta^+_{\text{worst}}(G)+2)\log_2 n$ then

$$\begin{aligned}
\mathfrak{M}_{u,\rho,\lambda} \stackrel{\text{def}}{=\!=} \mathbb{E}&\left[\begin{array}{l|l}\text{number of pairs of nodes} & v \text{ is selected uniformly}\\ v,y \text{ such that } v,y \text{ is weakly} & \text{randomly from}\\ (\rho,\lambda) \text{ dominated by } u & \cup_{\rho<j\leqslant\lambda}\mathcal{B}_j(u)\end{array}\right]\\
= \mathbb{E}&\left[\begin{array}{l|l}\text{number of pairs of nodes} & v \text{ is selected uniformly}\\ v,y \text{ such that } v,y \text{ is strongly} & \text{randomly from}\\ (\rho,\lambda) \text{ dominated by } u & \cup_{\rho<j\leqslant\lambda}\mathcal{B}_j(u)\end{array}\right].
\end{aligned}$$

Proof. Suppose that v,y is weakly (ρ,λ) dominated by u, i.e., there exists a shortest path $v \overset{\mathcal{P}}{\leftrightsquigarrow} y$ between $v,y \in \mathcal{B}_{\rho+\lambda}(u)$ such that for some node $v' \in v \overset{\mathcal{P}}{\leftrightsquigarrow} y$ we have $v' \in \mathcal{B}_{\rho}(u)$. Let $v \overset{\mathcal{Q}}{\leftrightsquigarrow} y$ be any other path between v and y that does not contain a node from $\mathcal{B}_{\rho}(u)$. Then, by Corollary 11 (i) (with $\kappa = 4$) we have

$$\ell(\mathcal{Q}) \geqslant 2^{\frac{\lambda}{6\,\delta^{+}_{\rm worst}(G)+2} + \frac{11}{6}} - 1 \geqslant 2^{\log_2 n + \frac{11}{6}} - 1 > n - 1,$$

which contradicts the obvious bound $\ell(\mathcal{Q}) < n$. Thus, no such path $\mathcal{Q}$ exists and v,y is strongly (ρ,λ) dominated by u. ■

[1] M. E. J. Newman, *Networks: An Introduction* (Oxford University Press, Oxford, 2010).
[2] R. Albert and A.-L. Barabási, Rev. Mod. Phys. **74**, 47 (2002).
[3] V. Colizza, A. Flammini, M. A. Serrano, and A. Vespignani, Nat. Phys. **2**, 110 (2006).
[4] V. Latora and M. Marchior, New J. Phys. **9**, 188 (2007).
[5] R. Albert, B. DasGupta, A. Gitter, G. Gürsoy, R. Hegde, P. Pal, G. S. Sivanathan, and E. Sontag, Phys. Rev. E **84**, 036117 (2011).
[6] D. S. Bassett, N. F. Wymbs, M. A. Porter, P. J. Mucha, J. M. Carlson, and S. T. Grafton, Proc. Natl. Acad. Sci. USA **108**, 7641 (2011).
[7] M. Gromov, Essays in Group Theory **8**, 75 (1987).
[8] E. A. Jonckheere and P. Lohsoonthorn, in *Proceedings of the American Control Conference* 2 (IEEE Press, Washington, DC, 2004), pp. 976–981.
[9] E. Jonckheere, P. Lohsoonthorn, and F. Bonahon, J. Graph Theor. **57**, 157 (2007).
[10] F. Ariaei, M. Lou, E. Jonckheere, B. Krishnamachari, and M. Zuniga, EURASIP J. Wireless Commun. Netw. **2008**, 213185 (2008).
[11] O. Narayan and I. Saniee, Phys. Rev. E **84**, 066108 (2011).
[12] F. Papadopoulos, D. Krioukov, M. Boguna, and A. Vahdat, in *Proceedings of the IEEE Conference on Computer Communications* (IEEE Press, Washington, DC, 2010), pp. 1–9.
[13] E. Jonckheerea, M. Loua, F. Bonahona, and Y. Baryshnikova, Internet Math. **7**, 1 (2011).
[14] M. Bogun, F. Papadopoulos, and D. Krioukov, Nat. Commun. **1**, 62 (2010).
[15] F. de Montgolfier, M. Soto, and L. Viennot, in *Proceedings of the 10th IEEE International Symposium on Networking Computing and Applications* (IEEE Press, Washington, DC, 2011), pp. 25–32.
[16] N. Robertson and P. D. Seymour, J. Comb. Theory Ser. B **35**, 39 (1983).
[17] H. L. Bodlaender, in *Lecture Notes in Computer Science 317*, edited by T. Lepistö and A. Salomaa (Springer, Berlin/Heidelberg, 1988), pp. 105–118.
[18] V. Chepoi and B. Estellon, in *Lecture Notes in Computer Science* 4627, edited by M. Charikar, K. Jansen, O. Reingold, and J. D. P. Rolim (Springer, Berlin/Heidelberg, 2007), pp. 59–73.
[19] V. Chepoi, F. F. Dragan, B. Estellon, M. Habib, and Y. Vaxès, in *Proceedings of the 24th Annual Symposium on Computational Geometry* (ACM Press, New York, 2008), pp. 59–68.
[20] V. Chepoi, F. F. Dragan, B. Estellon, M. Habib, Y. Vaxès, and Y. Xiang, Algorithmica **62**, 713 (2012).
[21] C. Gavoille and O. Ly, in *Lecture Notes in Computer Science 3827*, edited by X. Deng and D.-Z. Du (Springer, Berlin/Heidelberg, 2005), pp. 1071–1079.
[22] I. Abraham, M. Balakrishnan, F. Kuhn, D. Malkhi, V. Ramasubramanian, and K. Talwar, in *Proceedings of the 26th annual ACM Symposium on Principles of Distributed Computing* (ACM Press, New York, 2007), pp. 43–52.
[23] R. Krauthgamer and J. R. Lee, in *Proceedings of the 47th Annual IEEE Symposium on Foundations of Computer Science* (IEEE Press, Washington, DC, 2006), pp. 119–132.
[24] J. Roe, *Index Theory, Coarse Geometry, and Topology of Manifolds. Conference Board of the Mathematical Sciences Regional Conference Series 90* (American Mathematical Society, Providence, 1996).
[25] See Supplemental Material at http://link.aps.org/supplemental/10.1103/PhysRevE.89.032811 for addtional details of the networks studied.
[26] S. S. Shen-Orr, R. Milo, S. Mangan, and U. Alon, Nat. Genet. **31**, 64 (2002).
[27] A. Ma'ayan, S. L. Jenkins, S. Neves, A. Hasseldine, E. Grace, B. Dubin-Thaler, N. J. Eungdamrong, G. Weng, P. T. Ram, J. Jeremy Rice, A. Kershenbaum, G. A. Stolovitzky, R. D. Blitzer, and R. Iyengar, Science **309**, 1078 (2005).
[28] [25, updated version] see www.weizmann.ac.il/mcb/UriAlon/Papers/networkMotifs/coli1_1Inter_st.txt.
[29] R. Zhang, M. V. Shah, J. Yang, S. B. Nyland, X. Liu, J. K. Yun, R. Albert, and T. P. Loughran, Proc. Natl. Acad. Sci. USA **105**, 16308 (2008).
[30] R. Milo, S. Shen-Orr, S. Itzkovitz, N. Kashtan, and D. U. Alon, Science **298**, 824 (2002).
[31] H. Jeong, B. Tombor, R. Albert, Z. N. Oltvai, and A.-L. Barabasi, Nature (London) **407**, 651 (2000).
[32] G. von Dassow, E. Meir, E. M. Munro, and G. M. Odell, Nature (London) **406**, 188 (2000).
[33] S. Li, S. M. Assmann, and R. Albert, PLoS Biol. **4**, e312 (2006).
[34] J. Thakar, M. Pilione, G. Kirimanjeswara, E. T. Harvill, and R. Albert, PLoS Comput. Biol. **3**, e109 (2007).
[35] J. Saez-Rodriguez, L. Simeoni, J. A. Lindquist, R. Hemenway, U. Bommhardt, B. Arndt, U.-U. Haus, R. Weismantel, E. D. Gilles, S. Klamt, and B. Schraven, PLoS Comput. Biol. **3**, e163 (2007).
[36] A. Gitter, J. Klein-Seetharaman, A. Gupta, and Z. Bar-Joseph, Nucl. Acids Res. **39**, e22 (2011).
[37] D. Lusseau, K. Schneider, O. J. Boisseau, P. Haase, E. Slooten, and S. M. Dawson, Behav. Ecol. Sociobiol. **54**, 396 (2003).

[38] M. Girvan and M. E. J. Newman, Proc. Natl. Acad. Sci. USA **99**, 7821 (2002).
[39] W. W. Zachary, J. Anthropol. Res. **33**, 452 (1977).
[40] V. Krebs, www.orgnet.com.
[41] J. H. Michael and J. G. Massey, For. Prod. J. **47**, 25 (1997).
[42] P. Gleiser and L. Danon, Adv. Complex Syst. **6**, 565 (2003).
[43] C. P. Loomis, J. O. Morales, R. A. Clifford, and O. E. Leonard, *Turrialba: Social Systems and the Introduction of Change* (The Free Press, Glencoe, 1953).
[44] Dagstuhl seminar, *Link Analysis and Visualization*, Dagstuhl 1–6, 2001 http://vlado.fmf.uni-lj.si/pub/networks/data/sport/football.htm.
[45] D. E. Knuth, *The Stanford GraphBase: A Platform for Combinatorial Computing* (Addison-Wesley, Reading, 1993).
[46] E. Jonckheere, P. Lohsoonthorn, and F. Ariaei, Internet Math. **7**, 137 (2011).
[47] R. Kannan, P. Tetali, and S. Vempala, Random Struct. Alg. **14**, 293 (1999).
[48] B. Alberts, *Molecular Biology of the Cell* (Garland, New York, 1994).
[49] 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. Zeitlinger, 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, and R. A. Young, Science **298**, 799 (2002).
[50] M. R. Bridson and A. Haefliger, *Metric Spaces of Non-Positive Curvature* (Springer, New York, 1999).
[51] R. Albert, J. Cell Sci. **118**, 4947 (2005).
[52] R. S. Burt, *Structural Holes: The Social Structure of Competition* (Harvard University Press, Cambridge, 1995).
[53] S. P. Borgatti, Connections **20**, 35 (1997).
[54] M. R. Garey and D. S. Johnson, *Computers and Intractability–A Guide to the Theory of NP-Completeness* (W. H. Freeman, New York, 1979).
[55] P. Gupta, R. Janardan, M. Smid, and B. DasGupta, Int. J. Comput. Geom. Appl. **7**, 437 (1997).