Research Article

Journal of Algorithms & Computational Technology

Journal of Algorithms & Computational Technology
2017, Vol. 11(1) 40–57

DOI: 10.1177/1748301816665519
journals.sagepub.com/home/act
SAGE

Core–periphery models for graphs based on their δ-hyperbolicity: An example using biological networks

Hend Alrasheed and Feodor F Dragan

Abstract

Hyperbolicity is a global property of graphs that measures how close their structures are to trees in terms of their distances. It embeds multiple properties that facilitate solving several problems that found to be hard in the general graph form. In this paper, we investigate the hyperbolicity of graphs not only by considering Gromov's notion of δ-hyperbolicity but also by analyzing its relationship to other graph's parameters. This new perspective allows us to classify graphs with respect to their hyperbolicity and to show that many biological networks are hyperbolic. Then we introduce the eccentricity-based bending property which we exploit to identify the core vertices of a graph by proposing two models: the maximum-peak model and the minimum cover set model. In this extended version of the paper, we include some new theorems, as well as proofs of the theorems proposed in the conference paper. Also, we present the algorithms we used for each of the proposed core identification models, and we provide more analysis, explanations, and examples.

Keywords

Biological networks, δ-hyperbolicity, vertex eccentricity, eccentricity centrality, graph eccentricity layering, eccentricity-based bending property, shortest-path bending, core–periphery organization

Date received: 14 August 2015; accepted: 31 March 2016

Introduction

Using graph-theoretical tools for analyzing complex networks to characterize their structures has been the subject of much research. It aids identifying multiple key properties as well as explaining essential behaviors of those systems. A common structure that has been widely recognized in social networks as well as other network disciplines is the core–periphery structure. Multiple coefficients have been proposed to examine the existence of the core–periphery organization in a graph[1,2]; also, various methods have been introduced to identify the core of a graph.[3–5]

The core–periphery structure suggests partitioning the graph into two parts: the core which is dense and cohesive and the periphery which is sparse and disconnected. Vertices in the periphery part interact through a series of intermediate core vertices. This pattern of communication (where traffic tends to concentrate on a specific subset of the vertices) has been observed in trees where distant nodes communicate via the central node (or nodes) in the tree. δ-Hyperbolicity, which is a measure that shows how close a graph is to a tree, suggests that any shortest path (geodesic) between any pair of vertices bends (to some extent) toward the core of the graph. This phenomenon has been justified by the global curvature of the network which (in case of graphs) can be measured using hyperbolicity (sometimes called also the negative curvature).[6]

There are multiple equivalent definitions for Gromov's hyperbolicity. Let $G=(V,E)$ be a graph with a distance metric d on V such that the distance between two vertices x and y is the length of a shortest path between x and y. A geodesic triangle $\Delta(x,y,z)$ for three arbitrary vertices $x, y, z \in V$ is the union of three shortest paths (geodesic segments) connecting x, y, and z. In hyperbolic spaces, any vertex in any side of a geodesic triangle is contained in the d-neighborhood of the union of the two other sides. This forces the sides

Department of Computer Science, Kent State University, Kent, USA

Corresponding author:
Hend Alrasheed, Department of Computer Science, Kent State University, Kent, OH 44242, USA.
Email: halrashe@kent.edu

of the triangle to be curved toward its center as its size increases.[7]

Multiple complex networks such as the Internet,[8,9] data networks at the IP layer,[6] and social and biological networks[10,11] show low δ-hyperbolicity (low hyperbolicity suggests a structure that is close to a tree structure[12]). Also, it has been observed that networks with this property have highly connected cores.[6] Generally, the core of a graph is specified according to one or more centrality measures. For example, the degree centrality where the core of the graph is the set of vertices that have the highest number of connections, the betweenness centrality which considers the vertices that have the highest number of shortest paths passing through them as the core, and the eccentricity centrality in which the core has the subset of vertices that have the shortest distance to every other vertex.

The δ-hyperbolicity of graphs embeds multiple properties that facilitate solving several problems that found to be difficult in the general graph form. For example, diameter estimation,[13] distance and routing labeling,[14] and several routing problems.[15,16] In this paper, we investigate implications of the δ-hyperbolicity of a network and exploit them for the purpose of partitioning the graph into core and periphery parts.

Our main contributions can be summarized as follows:

1. We study the hyperbolicity of several real-world biological networks and show that the hyperbolicity of almost all the networks in our datasets is small. This confirms the results in Albert et al.[11] However, unlike previous efforts, we analyze the relationship between the hyperbolicity and other global parameters of the graph. We find in most of our networks that the hyperbolicity is bounded by the logarithm of the diameter ($\delta(G) \leq \log_2(diam(G))$) and the logarithm of the size of the graph in terms of the number of vertices and the number of edges ($\delta(G) \leq \log_2(size(G))$). Based on this analysis we classify graphs with respect to their hyperbolicity into three categories: *strongly-hyperbolic networks*, *hyperbolic networks*, and *nonhyperbolic networks*.
2. We formalize the notion of the *eccentricity layering* of a graph and employ it to introduce a new property that we find to be intrinsic to hyperbolic graphs: *the eccentricity-based bending property*. Unlike previous work, we investigate the essence of this bending in shortest paths by studying its relationship to the distance between vertex pairs.
3. We exploit the eccentricity-based bending property, and based on it, propose two core–periphery separation models: the *maximum-peak model* and the *minimum cover set model*.
4. We apply both models to our biological graph datasets. In contrast to what have been observed in Holme,[2] we find that biological networks exhibit a clear-cut core–periphery structure. Then we investigate the relationship between the hyperbolicity of a graph and the conciseness of its core.

This paper is organized as follows. The next section discusses some theoretical background on graph theory and presents the definition of δ-hyperbolic graphs. Related work on the core–periphery structure and graph centrality measures in networks in general and in biological networks in particular are also discussed. "Datasets" section describes the kinds of biological networks used in this paper and presents a summary of their parameters. In "δ-Hyperbolicity of networks" section, we measure, analyze, and classify the δ-hyperbolicity of the networks in our datasets. Our classification is based on how the *hyperbolicity* is evident in those graphs. We also discuss how other graph parameters factor in to this classification. Finally in "Core–periphery models based on δ-hyperbolicity" section, we propose our eccentricity-based bending property followed by two core–periphery separation models as one of this property's implications.

Theoretical background and related work

Preliminaries on graph theory

A simple undirected graph $G = (V, E)$ naturally defines a metric space (V, d) on its vertex set V. The distance $d(u, v)$ is defined as the number of edges in a shortest path $\rho(u, v)$ that connects two vertices u and v.

We define the *size* of the graph denoted as $size(G)$ as the sum of the number of vertices and the number of edges in G, i.e. $size(G) = |V| + |E|$.

The *diameter* of the graph $diam(G)$ is the length of the longest shortest path between any two vertices u and v in the graph, i.e. $diam(G) = \max_{u,v \in V}\{d(u, v)\}$. Obviously, when the graph is disconnected, the value of the diameter is undefined.

According to the distances between vertices in the graph, the center can be defined using the vertex eccentricities. The *eccentricity* of a vertex u is $ecc(u) = \max_{v \in V}\{d(u, v)\}$, i.e. the distance between u and any of its farthest neighbors v. The minimum value of the eccentricity represents the graph's *radius*: $rad(G) = \min_{u \in V}\{ecc(u)\}$. The set of vertices with minimum eccentricity are considered the *center* of the graph $C(G)$. In other words, $C(G) = \{u \in V : ecc(u) = rad(G)\}$.

δ-Hyperbolicity

In smooth geometry, hyperbolicity captures the notion of negative curvature which can be generalized as δ-hyperbolicity in more abstract concept of metric

spaces including graphs.[17] The presence of hyperbolic networks in a big variety of applications attracted many researchers to investigate the negative curvature of different types of graphs. It turns out that many real-world networks are hyperbolic.[6,8,9,11] Moreover, multiple applications of the hyperbolicity have been examined such as diameter estimation,[13] compact distance and routing labeling schemes and spanners,[14,16,18] and several routing and traffic flow problems.[6,15,16,19]

The δ-hyperbolicity measure of a metric space was proposed by Gromov in 1987.[20] It measures how close the metric structure is to a tree structure. A connected graph G can be viewed as a metric space with the graph distance metric d. There are multiple equivalent definitions (up to constant factors[13]) for Gromov's hyperbolicity. In this paper, we use the following four-point condition definition.

Given a graph $G=(V,E)$, x, y, u, and $v \in V$ are four distinct vertices, and the three sums: $d(x,y)+d(u,v)$, $d(x,u)+d(y,v)$, and $d(x,v)+d(y,u)$ sorted in a nonincreasing order, the hyperbolicity of the quadruple x, y, u, v denoted as $\delta(x,y,u,v)$ equals half the difference between the two larger sums. It is defined as: $\delta(x,y,u,v)=((d(x,y)+d(u,v))-(d(x,u)+d(y,v)))/2$. The δ-hyperbolicity of the graph G denoted as $\delta(G)$ (or simply δ) is $\delta(G)=\max_{x,y,u,v\in G}\delta(x,y,u,v)$.

A graph $G=(V,E)$ is considered *δ-hyperbolic* for some nonnegative real number δ if for every set of four points x, y, u, and v, the larger two of the three sums $d(x,y)+d(u,v), d(x,u)+d(y,v)$, and $d(x,v)+d(y,u)$ differ by at most 2δ. The δ-hyperbolicity of a graph is the smallest δ for which the graph G is δ-hyperbolic. For finite graphs δ-hyperbolicity or simply $\delta(G)$ is finite. Consequently, one can think of all finite graphs as hyperbolic except that the value of δ decides how *hyperbolic* the graph is. On the other hand, as mentioned in Adcock et al.,[7] when no finite δ exists (which may be the case for infinite graphs), the graph is considered nonhyperbolic.

Trees are 0-hyperbolic ($\delta(G)=0$), cliques are also 0-hyperbolic, and chordal graphs are at most 1-hyperbolic.[21] On the other hand, a cycle with n vertices is approximately $n/4$-hyperbolic and an $n\times n$ grid is $(n-1)$-hyperbolic. Generally, the smaller the value of δ the closer the graph is to a tree (metrically); this implies strong hyperbolicity. See Wu and Zhang[22] for a detailed discussion on tree-likeness and hyperbolicity.

Core–periphery and network centrality in complex networks

The notion of the core–periphery structure has a long history in social network analysis. It deals with identifying the part (or parts) of the network that represents the central part in terms of the network distance, the most congested part in terms of the network traffic, the highly connected part in terms of the vertices' degrees, or any combination of the three. Borgatti and Everett[3] formalize the core–periphery structure by developing two families of core–periphery models: the discrete model where vertices belong to one of two classes (core and periphery) and the continuous model which includes three classes (or more) of vertices (core, semi-periphery, and periphery). Then they propose algorithms for detecting each model by finding the partition which maximizes the correlation between the data matrix and the pattern matrix.

Seidman[4] proposes the k-core decomposition as a tool to study the structural properties of large networks focusing on subsets of increasing degree centrality. It partitions the graph into subsets each of which is identified by removing vertices of degree smaller than k.

Holme[2] introduces a coefficient that measures if a network has a clear core–periphery structure based on the closeness centrality (defined below) and the basic definition of clusters. He also shows that the core's neighborhood (for increasing radius) is highly dense (with respect to the number of edges). Unlike our result, he concludes that biological networks do not have a strong core–periphery organization. In Chung and Lu,[23] the authors show that for some families of random graphs with expected degrees there is a core to which almost all vertices are at distance less than or equal to $O(\log\log n)$. Leskovec et al.[24] study community structures in large networks by analyzing a big range of different real-world networks. They identify the existence of multiple (smaller) communities that are attached to the core of the network with very few connections (whiskers). They also observe that some graphs have cores with a nested core–periphery structure.

In the study of communication networks, the core is usually identified by the small dense part of the network that carries out most of the traffic under shortest-path routing.[6,25,26] Narayan and Saniee[6] show that the load scales as $O(n^2)$ with the number of vertices n at the core of the network. The asymptotic traffic flow in hyperbolic graphs has been studied in Baryshnikov and Tucci.[25] It shows that a vertex v belongs to the core if there exists a finite radius r such that the amount of the traffic that passes through the ball centered at v and with radius r behaves asymptotically as $\theta(n^2)$ as the number of vertices n grows to infinity. The existence of the core in large networks such as the Internet motivates researchers to embed the Internet distance metric in a hyperbolic space for distance estimation.[9,27]

The notion behind centrality is identifying key vertices in a graph that are considered high contributors. There are multiple centrality measures in the literature; some of which are: the betweenness centrality (similar to

Newman,[28] we will call it the shortest-path betweenness centrality) and the eccentricity centrality.[29–31]

The shortest-path betweenness measure is a widely used concept in social networks. It expresses how much effect each vertex (or actor) has in the communication of the network assuming that all traffic follows shortest paths between every two vertices (taking into account that one or more shortest paths may exist between any given pair).

Given a connected finite graph $G=(V,E)$, the shortest-path betweenness of a vertex $u \in V$ measures the total number of shortest paths between every pair of vertices x and y that pass through u such that $u \neq x$ and $u \neq y$ to exclude the cases when u is the source or the destination vertex. Let $\alpha_{xy}(u)$ be the fraction of shortest paths between x and y that pass through u, i.e. $\alpha_{xy}(u) = \sigma_{xy}(u)/\sigma_{xy}$, where $\sigma_{xy}(u)$ is the number of all shortest paths between x and y that pass through u and σ_{xy} is the number of all shortest paths between x and y. The shortest-path betweenness centrality $c_B(u)$ of a vertex u can be calculated as the following[29]: $c_B(u) = \sum_{x \in V} \sum_{y \in V} \alpha_{xy}(u)$. The higher value of this index indicates the higher importance of the role that the vertex plays in the data exchange process among distant vertices.

The eccentricity centrality suggests that the center of the graph includes the vertex (or vertices) that has the shortest distance to all other vertices (minimum eccentricity). For a given vertex u, the eccentricity centrality of u is $c_E(u) = 1/\max\{d(u,v) : v \in V\}$.[32] As discussed in ''Preliminaries on graph theory'' section, a vertex u is a central vertex if $ecc(u) = rad(G)$.

The eccentricity centrality is sometimes referred to in literature as the closeness centrality. To avoid any ambiguity, we point out that the closeness centrality considers the center as the subset of vertices with the minimum total distance to all other vertices. In other words, the closeness centrality $c_C(u)$ of a vertex u is $c_C(u) = 1/\sum_{v \in V} d(u,v)$.[29]

Biological networks and the core–periphery structure

It has been found in several fields that looking at the overall system reveals more about the functionality of its components as opposed to inspecting its individual elements. Therefore, various types of large-scale biological networks have been constructed to capture the different kinds of interactions between their components. For instance, protein–protein interaction (PPI) networks have been used to identify the function of individual proteins as well as the purpose behind some unknown interactions.[33] A lot of research efforts were directed to discover some topological properties of the biological networks. It has been shown that some protein structures and PPI networks[34,35] and a number of metabolic networks[36–38] exhibit the small-world property (a graph is small world when its diameter is bounded by the logarithm of its size ($diam(G) \leq \log_2(size(G))$). A lot of work has focused on analyzing the degree distribution of different biological networks. Power-law degree distribution was caught in protein structure networks,[34] PPI networks,[35,37] and metabolic networks.[39]

Structure analyses of some biological networks have detected the presence of hierarchal, modular, and core–periphery organization structures. The core–periphery model of several types of biological networks has been studied thoroughly in the literature. Da et al.[1] propose a parameter that detects the existence of a core–periphery structure in a metabolic network based on the closeness centrality of metabolites and the network connectivity. In Junker and Schreiber,[40] the yeast protein interaction network was analyzed based on the betweenness centrality. They conclude that the high betweenness low connectivity proteins may be working as connectors between separate modules. Using mathematical tools that have been used to analyze sociological networks, the authors in 36 study recognizing the central metabolites in a metabolic pathway network. In Luo et al.,[41] the authors demonstrate a systematic exploration of the core–periphery model in protein interaction networks that depends on the connectivity of the vertices. They also classify the peripheral vertices based on their structural relationship with the core. In Ma and Zeng,[42] the authors identify the central metabolites using degree centrality and closeness centrality. They also show the relationship between the average path length and the closeness centralization index of metabolic network.

Datasets

There are various forms of biological networks that have different characteristics according to their origins and to their construction methodologies. Generally, the vertices in a biological network represent biomolecules such as proteins, genes, or metabolites, and the edges represent a chemical, physical, or functional interaction between the connected vertices. Each of the networks used in this work belongs to one of the following types:

1. Protein interaction (PI) networks: Generally, in a PI network, the vertices represent different proteins and the edges represent the connections between the interacting proteins. PI networks have been described as small-world and scale-free networks.[40]
2. Neural networks: They contain neurons (vertices) which are connected together through synapsis (edges). Neurons have a high tendency to form clusters based on their spatial location. Neural networks are small-world networks.[40]

3. Metabolic networks: Metabolic networks are represented by metabolites (vertices) and biochemical reactions (directed edges). Usually, metabolites are small molecules (for example amino acids); however, they also can be macromolecules. Metabolic networks show the small-world property and they have a high clustering coefficient. Also, they follow the power-law degree distribution.[40]
4. Transcription networks: Networks in which vertices are genes and edges represent different interactions (interrelationships) between genes. A transcription network is one type of the gene regulatory networks.[43]

We analyze the PI networks of budding yeast,[44] *Escherichia coli*,[45] yeast,[46] *Saccharomyces cerevisiae*,[47] and *Helicobacter pylori*.[48] Also, we analyze two different brain area networks of the Macaque monkey[49,50] and the metabolic networks of the *Escherichia coli*[51] and the Caenorhabditis elegans.[52] Finally, we analyze the yeast transcription network.[43]

In this work, we consider unweighted graphs, and we only consider the largest connected component of each network. The size of this component for each network is presented in Table 1. We also ignore the directions of the edges.

δ-Hyperbolicity of networks

For the purpose of investigating the hyperbolicity of networks, it seems natural to analyze and classify the graphs based on their hyperbolicity. The classification should reflect how strong (evident) the tree-likeness is in the graph's structure.

In the following subsections not only we measure the hyperbolicity of each of the networks in our graph datasets, but also we relate this value to other graph's parameters. Upon this analysis, we provide our hyperbolicity-based classification of the graphs.

Hyperbolicity of biological networks

We measure δ-hyperbolicity on each bi-connected component of each network in the datasets presented in "Datasets" section using Gromov's four-point condition. For each network, we identify a bi-connected component with the maximum value of δ since the hyperbolicity of a graph equals the maximum hyperbolicity of its bi-connected components.[8,22,53]

Table 2 shows that almost all networks in our datasets have small hyperbolicity. Even though the definition of δ-hyperbolicity considers the difference between the largest two distance sums among any quadruples and takes into account only the maximum one, this absolute analysis is deficient. Similar to Adcock et al.[7] and Montgolfier et al.,[8] closer analysis to the distribution of the value of δ (see Figure 1) shows that only a very small percent of the quadruples (less than 1%) have the maximum value of δ while most quadruples have $\delta = 0$ (about 40–70%). Figure 2(a) to (c) presents samples of the distribution of the quadruples over the different values of δ.

This observation makes it equally important to calculate the value of the average delta $\delta'(G)$ (see Table 2). This suggests that the maximum value of δ was not expressive for the graph but was affected by some outliers. The average hyperbolicity is defined as: $\delta'(G) = \sum_{x,y,u,v \in V} \delta(x, y, u, v) / \binom{|V|}{4}$.

Analysis and discussion

Our goal is to categorize graphs with respect to their hyperbolicity into three classes: *strongly-hyperbolic graphs, hyperbolic graphs*, and *nonhyperbolic graphs*.

Taking into account that trees are 0-hyperbolic, generally, the smaller the value of $\delta(G)$, the closer the graph's structure to a tree. However, studying the tree-like

Table 1. Graph datasets and their parameters: number of vertices $|V|$; number of edges $|E|$; graph's size $size(G) = |V| + |E|$; average degree $\bar{d}$; diameter $diam(G)$; radius $rad(G)$.

Network category	Network	$\|V\|$	$\|E\|$	$\log_2(size(G))$	$\bar{d}$	$diam(G)$	$rad(G)$
PI networks	B-YEAST-PI	1465	5839	12.8	7.97	8	5
	E-COLI-PI	126	581	9.5	9.2	5	3
	YEAST-PI	1728	11003	13.6	12.7	12	7
	S-CEREVISIAE-PI	537	1002	10.5	3.7	11	7
	H-PYLORI-PI	72	112	7.5	3.1	7	5
Neural networks	MACAQUE-BRAIN-1	45	463	9	11.3	4	2
	MACAQUE-BRAIN-2	350	5198	12.4	29.7	4	3
Metabolic networks	E-COLI-METABOLIC	242	376	9.3	3.1	16	9
	C-ELEGANS-METABOLIC	453	4596	12.3	8.9	7	4
Transcription networks	YEAST-TRANSCRIPTION	321	711	10	4.4	9	5

Table 2. Graph datasets and their parameters: number of vertices $|V|$; number of edges $|E|$; diameter $diam(G)$; radius $rad(G)$; hyperbolicity $\delta(G)$; and the average hyperbolicity $\delta'(G)$.

Network	$\|V\|$	$\|E\|$	$diam(G)$	$rad(G)$	$\delta(G)$	$\delta'(G)$
B-YEAST-PI	1465	5839	8	5	2.5	0.299
E-COLI-PI	126	581	5	3	2	0.251
YEAST-PI	1728	11003	12	7	3.5	0.322
S-CEREVISIAE-PI	537	1002	11	7	4	0.419
H-PYLORI-PI	72	112	7	5	3	0.368
MACAQUE-BRAIN-1	45	463	4	2	1.5	0.231
MACAQUE-BRAIN-2	350	5198	4	3	1.5	0.203
E-COLI-METABOLIC	242	376	16	9	4	0.483
C-ELEGANS-METABOLIC	453	4596	7	4	1.5	0.133
YEAST-TRANSCRIPTION	321	711	9	5	3	0.365

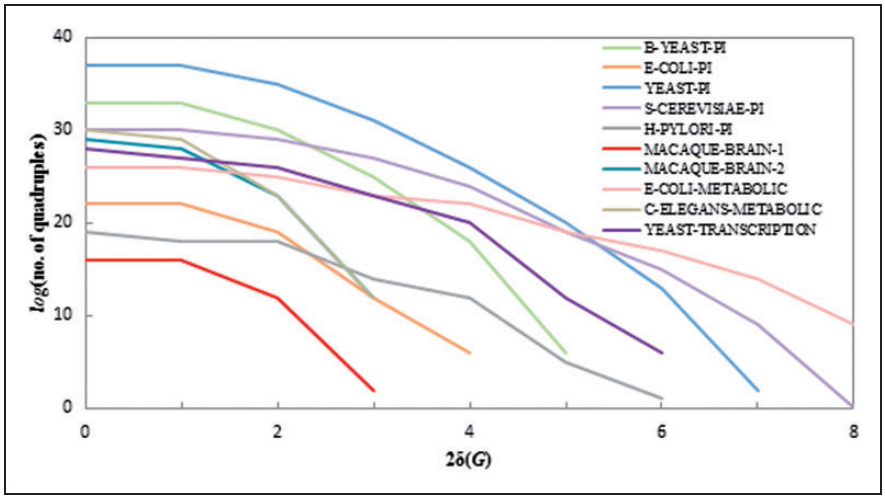

Figure 1. Distribution of quadruples over different values of δ.

(a)

2δ	Quadruples	%
0	1276345762	73%
1	448324099	26%
2	6795007	1%
3	3057	0.002%

(b)

2δ	Quadruples	%
0	5529039	55%
1	3934917	39%
2	540463	5%
3	4658	0.04%
4	48	0.0005%

(c)

2δ	Quadruples	%
0	529467	52%
1	273481	27%
2	199186	19%
3	21971	2%
4	4651	1%
5	32	0.003%
6	2	0.0002%

Figure 2. Distribution of quadruples over different values of δ (three datasets): (a) C-ELEGANS-METABOLIC, (b) E-COLI-PI, and (c) H-PYLORI-PI.

structure of graphs based solely on the value of the hyperbolicity may not be sufficient to characterize their closeness to a tree topology for two reasons. First, the hyperbolicity is a relative measure. For example, for a given graph $G=(V,E)$, a value of $\delta(G)=10$ can be seen as too large when $size(G)\simeq 10^2$; the structure of G can be fairly described as being far from a tree. However, when $size(G)\simeq 10^7$, the hyperbolicity $\delta(G)=10$ looks much smaller and indicates a tree-like structure. Second, small graph size and (or) small diameter directly yield low

hyperbolicity; our classification should be sensitive to such cases. In other words, small $\delta(G)$ does not always suggest a graph with a tree-like structure; other graph attributes that might impact the hyperbolicity must be investigated. We find that the graph's size $size(G)$ and the graph's diameter $diam(G)$ play an important role in deciding how hyperbolic a given graph is.

Since finite graphs will always have a finite value for δ such that the four-point condition is true, it is natural to think that the nonhyperbolic class includes only infinite graphs. However, in this study, we only consider finite graphs; accordingly, a nonhyperbolic graph in our sense is a graph with too large δ with respect to the logarithm of the graph's size, i.e. when it violates the following condition: $\delta(G) \leq \log_2(size(G))$.

Recall that $size(G) = |V| + |E|$. In cases where $\delta(G) \leq \log_2(size(G))$, we move on and compare δ with the diameter of the graph. To guarantee that the value of the diameter is not directly impacted by the size of the graph, first we require the diameter to be within the following bound: $diam(G) \leq \log_2(size(G))$. This is especially important for excluding the uninteresting cases where the small diameter is a result of the graph's small size.

Multiple previous works have analyzed the relationship between the hyperbolicity and the diameter. In fact, the graph's diameter represents an upper bound for δ.

Lemma 1.[54,55] *For any graph G with diameter diam(G) and hyperbolicity* $\delta(G)$, $\delta(G) \leq diam(G)/2$.

Moreover, the authors in Kennedy et al.[17] (using the slim triangles condition for hyperbolicity) and in Jonckheere et al.[56] (using the four-point condition) argue that the hyperbolicity of the graph is "actually" present when the value of δ is much smaller than the graph's diameter. They specify that for a graph to be hyperbolic the value of $\delta/diam(G)$ must asymptotically scale to zero. In this work, our goal is to ensure that the low value for the hyperbolicity is not a consequence of the graph's small diameter.

Interestingly, for most of the networks in our graph datasets, we find that $\delta(G) \leq \log_2(diam(G))$. Therefore, we say that a graph is

- *Strongly-hyperbolic* if it exhibits (1) $diam(G) \leq \log_2(size(G))$ and (2) $\delta(G) \leq \log_2(diam(G))$. Note that a graph that satisfies (2) is in fact a small-world graph;
- *Hyperbolic* when it violates either (1) or (2);
- *Nonhyperbolic* in all other cases.

Note that in the case when δ is small but greater than $\log_2(size(G))$, δ is an insufficient indication for hyperbolicity. We are not saying that the graph is far from the tree structure; still we point out that hyperbolicity is not very expressive in this case and other tree-likeness measurements may be used.

As Table 1 shows, all networks in the datasets, with the exception of networks S-CEREVISIAE-PI and E-COLI-METABOLIC, exhibit the small-world property. Also, Table 2 shows that $\delta(G) \leq \log_2(diam(G))$ in all graphs except for the S-CEREVISIAE-PI and the H-PYLORI-PI networks. As a result, those three graphs have been classified as hyperbolic graphs, and their hyperbolicity values are on the larger side ($\delta(G)$ for S-CEREVISIAE-PI is 4, $\delta(G)$ for H-PYLORI-PI is 3, and $\delta(G)$ for E-COLI-METABOLIC is 4). Also, the values of the average $\delta(G)$ ($\delta'(G)$) from Table 2 are high compared to other networks (0.419, 0.368, and 0.483, respectively). In Figure 3, we show this classification. Note that none

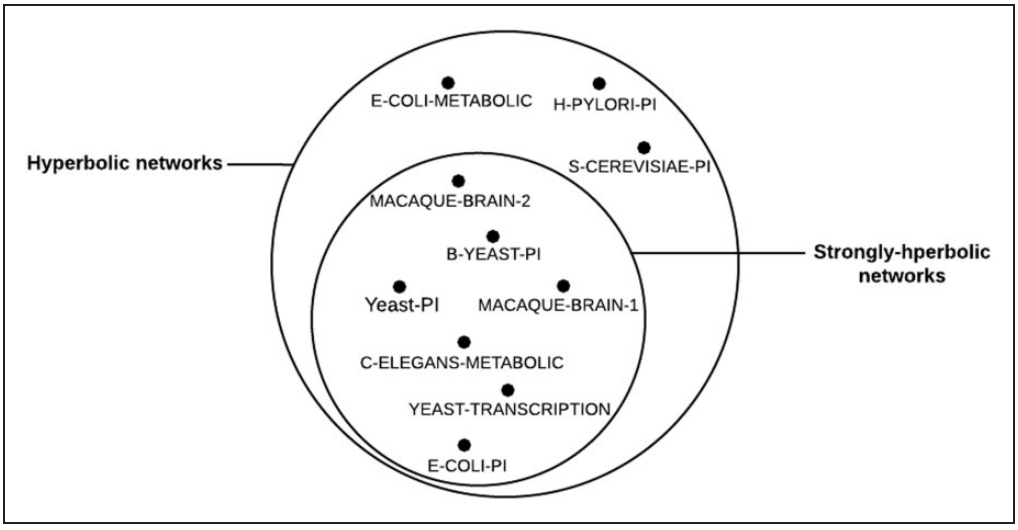

Figure 3. Classification of the graph datasets based on their hyperbolicity.

of the networks in our graph datasets is nonhyperbolic (it is always the case that $\delta(G) \leq \log_2(size(G))$).

Quantifying "small" and "large" for δ is not straightforward simply because it is relative. Therefore, we judge according to the difference between δ and $\log_2(\log_2(size(G)))$. The more substantial this difference is the closer the graph's structure to a tree structure. For example in Table 3, networks C-ELEGANS-METABOLIC and MACAQUE-BRAIN-2 are metrically closer to trees than networks YEAST-TRANSCRIPTION and YEAST-PI.

We also compared the value of δ to $\log_2(size(G))$ and to $\log_2(diam(G))$, and their average for the networks in the two classes: strongly-hyperbolic and hyperbolic. The values are listed in Table 3. Note that the three last columns of Table 3 show higher values for the hyperbolic graphs.

Core–periphery models based on δ-hyperbolicity

It was suggested by Baryshnikov and Tucci[25] and Narayan and Saniee[6] that the highly congested cores in many communication networks can be due to their hyperbolicity or negative curvature. Those cores are represented by vertices that belong to most shortest paths (shortest-path betweenness centrality) and (or) have minimum distances to all other vertices (eccentricity centrality). It was also observed that the negative curvature causes most of the shortest paths to bend making the peak of the arc formed by a shortest path to pass through a vertex in the core. In this section, we formalize this notion of bending in shortest paths by introducing an important property that is intrinsic to δ-hyperbolic graphs (the *eccentricity-based bending property*). Then we use the implication of this property to aid the partitioning of a graph into core and periphery parts by proposing two models: the *maximum-peak model* and the *minimum cover set model*. we further apply our models to the biological networks in our datasets. In contrast to what have been observed in Holme,[2] we show that biological networks do exhibit a core–periphery structure.

Eccentricity layering of a graph

The *eccentricity layering* of a graph $G = (V, E)$ denoted as $\mathcal{EL}(G)$ partitions its vertices into concentric circles or layers $\ell_r(G)$, $r = 0, 1, \ldots$. Each layer r is defined as $\ell_r(G) = \{u \in V : ecc(u) - rad(G) = r\}$. Here r represents the index of the layer. The innermost layer (layer 0) encloses the graph's center $C(G)$, i.e. all vertices with the minimum eccentricity or of eccentricity equal to $rad(G)$; this layer has index $r = 0$. Then the first layer includes all vertices who have their eccentricities equal to $rad(G) + 1$, and so on. The vertices in the last layer (outermost layer) will have eccentricities equal to the diameter of the graph. Figure 4 demonstrates an illustration for the layers.

Any vertex $v \in \ell_r(G)$ has *level* (or layer) $level(v) = r$. Figure 5 and Table 4 show the distribution of the vertices over different layers of the eccentricity layering of each graph in our datasets. Note that the vertices' population is denser in middle layers in almost all our networks.

In this work, we use layer 0 to indicate the central layer, layer 1 to indicate the layer that directly succeeds the central layer, etc. Therefore, when we say lower layers, we mean layers toward the central layer and the opposite for higher layers.

Table 3. The ratio of δ to logarithm of the size of the graph $\log_2(size(G))$ and logarithm of the diameter $\log_2(diam(G))$ for each network in our datasets. ϵ is the difference $\delta - \log_2(\log_2(size(G)))$. The average value presents the average $\left(\frac{\delta}{\log_2(size(G))} + \frac{\delta}{\log_2(diam(G))}\right)/2$.

	Network	$\log_2(size(G))$	$diam(G)$	δ	ϵ	$\frac{\delta}{\log_2(size(G))}$	$\frac{\delta}{\log_2(diam(G))}$	Average
Strongly-hyperbolic networks	C-ELEGANS-METABOLIC	12.3	7	1.5	2.1	0.121	0.536	0.329
	B-YEAST-PI	12.8	8	2.5	1.2	0.195	0.833	0.514
	MACAQUE-BRAIN-2	12.4	4	1.5	2.1	0.120	0.750	0.435
	E-COLI-PI	9.5	5	2	1.2	0.210	0.869	0.540
	YEAST-TRANSCRIPTION	10	9	3	0.3	0.300	0.937	0.636
	MACAQUE-BRAIN-1	9	4	1.5	1.5	0.166	0.750	0.458
	YEAST-PI	13.6	12	3.5	0.3	0.275	0.972	0.626
Hyperbolic networks	S-CEREVISIA-PI	10.5	11	4		0.380	1.142	0.761
	H-PYLORI-PI	7.5	7	3		0.400	1.071	0.736
	E-COLI-METABOLIC	9.3	16	4		0.430	1.000	0.715

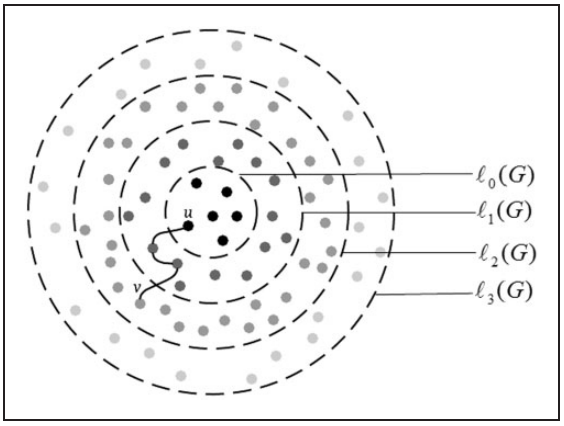

Figure 4. Eccentricity layering of a graph. Darker vertices belong to lower layers.

Eccentricity-based bending property of δ-hyperbolic networks

Let $G=(V,E)$ be a δ-hyperbolic graph, $\mathcal{EL}(G)$ be its eccentricity layering, and $C(G)$ be its center. In Chepoi et al.,[13] the following useful metric property of δ-hyperbolic graphs was proven.

Lemma 2 [14]. *Let G be a δ-hyperbolic graph and x, y, v, u be its four arbitrary vertices. If $d(v,u) \geq \max\{d(y,u), d(x,u)\}$, then $d(x,y) \leq \max\{d(v,x), d(v,y)\} + 2\delta$.*

We use this property to establish the following few interesting statements.

Proposition 1. *Let G be a δ-hyperbolic graph and x, y, s be arbitrary vertices of G. If $d(x,y) > 4\delta + 1$, then*

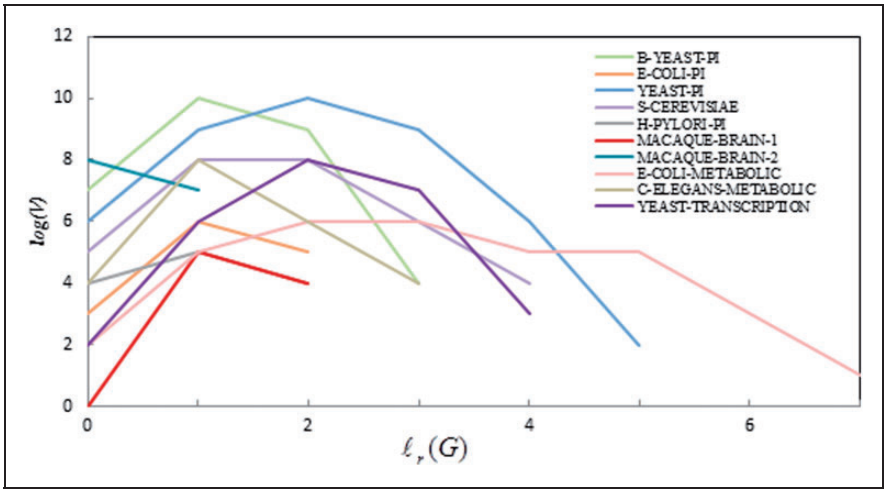

Figure 5. Distribution of vertices over different layers of the graph's eccentricity layering.

Table 4. Distribution of vertices over different layers with respect to the graph's eccentricity layering. *rad*(*G*) is the graph's radius; |*V* | is the number of vertices of each graph.

			Layer							
Network	*rad*(*G*)	|*V* |	0	1	2	3	4	5	6	7
B-YEAST-PI	5	1465	90	902	456	17				
E-COLI-PI	3	126	6	87	33					
YEAST-PI	7	1728	53	419	805	393	55	3		
S-CEREVISIAE-PI	7	537	26	198	219	79	15			
H-PYLORI-PI	5	72	14	41	17					
MACAQUE-BRAIN-1	2	45	1	30	14					
MACAQUE-BRAIN-2	3	350	194	156						
E-COLI-METABOLIC	9	242	5	43	54	50	42	36	10	2
C-ELEGANS-METABOLIC	4	453	17	353	69	14				
YEAST-TRANSCRIPTION	5	321	3	58	204	49	7			

$d(w,s) < \max\{d(x,s), d(y,s)\}$ *for any middle vertex* w *of any shortest* (x, y)*-path.*

Proof. Let w be a middle vertex of a shortest (x, y)-path and let $d(x,w) = \lfloor d(x,y)/2 \rfloor$. Assume, by way of contradiction, that $d(w,s) \geq \max\{d(x,s), d(y,s)\}$. Then, by Lemma 2, $d(x,y) \leq \max\{d(w,x), d(w,y)\} + 2\delta = d(w,y) + 2\delta$. Since $d(x,y) = d(x,w) + d(w,y)$, we obtain $d(x,w) \leq 2\delta$, giving $d(x,y) \leq 4\delta + 1$. □

Proposition 2. *Let* G *be a* δ*-hyperbolic graph and* x, y *be arbitrary vertices of* G*. If* $d(x,y) > 4\delta + 1$*, then on any shortest* (x, y)*-path there is a vertex* w *with* $ecc(w) < \max\{ecc(x), ecc(y)\}$.

Proof. Let w be a middle vertex of a shortest (x, y)-path given by Proposition 1 and s be a vertex such that $d(w,s) = ecc(w)$. Then, by Proposition 1, $ecc(w) = d(w,s) < \max\{d(x,s), d(y,s)\} \leq \max\{ecc(x), ecc(y)\}$. □

Denote by $D(y,r) := \{v \in V : d(v,y) \leq r\}$ the closed disk of G of radius r and centered at vertex y.

Proposition 3. *Let* G *be a* δ*-hyperbolic graph and* y *be its arbitrary vertex. Then,* $C(G) \subseteq D(y, 4\delta + 1)$ *or there is a vertex* $v \in D(y, 2\delta + 1)$ *such that* $ecc(v) < ecc(y)$.

Proof. Consider an arbitrary vertex $x \in C(G)$ and any shortest (x, y)-path P. Assume that $x \notin D(y, 4\delta + 1)$, i.e. P has length at least $4\delta + 2$. Consider the vertex v of this path which is at distance $2\delta + 1$ from y. We claim that $ecc(v) < ecc(y)$. Let s be a vertex such that $ecc(v) = d(v,s)$ and assume that $ecc(v) \geq ecc(y)$, i.e. $d(v,s) \geq ecc(y) \geq d(y,s)$. We know also that $d(v,s) \geq ecc(y) \geq rad(G) = ecc(x) \geq d(x,s)$. Then, Lemma 2 applied to $d(v,s) \geq \max\{d(y,s), d(x,s)\}$ gives $d(x,y) \leq \max\{d(v,x), d(v,y)\} + 2\delta = d(v,x) + 2\delta$. But, since $d(x,v) + d(v,y) = d(x,y)$, we get $2\delta \geq d(x,y) - d(x,v) = d(v,y) = 2\delta + 1$, which is impossible. □

We define the bend in shortest paths between two distinct vertices u and v with $d(u,v) \geq 2$, denoted by $bend(u, v)$, as follows:

Definition. $\forall\ u, v \in V$ $bend(u,v) = \min\{level(z) : z \in V$ and $d(u,z) + d(z,v) = d(u,v)\}$. Here $level(z) = r$ iff $z \in \ell_r(G)$.

We say that shortest paths between u and v bend if and only if a vertex z with $ecc(z) < \max\{ecc(u), ecc(v)\}$ exists in a shortest path between u and v. In this case we say also that *the pair of vertices* u *and* v *bends*. The parameter *bend* decides the extent (or the level) to which shortest paths between vertices u and v curve toward the center of the graph (since we are always looking for a z that belongs to a smaller layer according to the eccentricity layering). Note that in some cases $bend(u, v)$ will be assigned either $ecc(u)$ or $ecc(v)$, whatever is smaller. For example, see the shortest path $\rho(u, v)$ in Figure 4.

Now we analyze how vertex pairs behave in terms of their bending toward the center of the graph $C(G)$. Specifically, we investigate the effect of the distance between a pair of vertices on the bend of the shortest paths between them. Our findings in this context are summarized in the following two statements.

A. Despite their distances, most vertex pairs bend. Moreover, among those bending vertex pairs, the majority is represented by those that are sufficiently far from each other.
B. There is a direct relation between the distance among vertex pairs and how close to the center a shortest path between them bends.

Motivation and empirical evaluation of (A). In light of Proposition 2, we investigate how vertex pairs of various distances act with respect to the eccentricity-based bending property. Given two vertices u and v, we know from Proposition 2 that when $d(u,v) > 4\delta + 1$, then $\rho(u, v)$ bends. We now analyze this bending in $\rho(u, v)$ when $d(u,v) \leq 4\delta + 1$. We are motivated by the fact that the small-world property is observed in many biological networks[6] (also refer to ''Biological networks and the core–periphery structure'' and ''Datasets'' sections). Accordingly, shortest paths of lengths $4\delta + 1$ may not even exist in those networks.

Interestingly, we noticed the bend in the majority of shortest paths with lengths $\leq 4\delta + 1$. A quick look at Table 5 shows that a big percent of vertex pairs of distance at least two bend. For example, in graph C-ELEGANS-METABOLIC, about 97% of the vertex pairs with distances at least two bend. Furthermore, even though it is not always true (see the example in

Table 5. The effect of the distance k between vertex pairs on the bending property. Out of all vertex pairs with distance at least k, we show the percentage of those that bend for three of the networks in our graph datasets.

k	C-ELEGANS-METABOLIC ($diam(G) = 7$)	B-YEAST-PI ($diam(G) = 8$)	YEAST-TRANSCRIPTION ($diam(G) = 9$)
2	96.99%	93.10%	96.65%
3	99.89%	94.87%	97.77%
4	100%	98.43%	99.11%
5	100%	99.93%	99.88%
6	100%	100%	100%
7	100%	100%	100%
8		100%	100%
9			100%

Figure 6), all graphs in our graph datasets have bends for vertex pairs that are at distance equal to the graphs' diameters.

To quantify the distances at which the bend in vertex pairs happens in each of the networks in our datasets, we define two parameters: the *absolute curvity* and the *effective curvity*.

Let k be the distance between a pair of vertices ($2 \leq k \leq diam(G)$). The *absolute curvity* k^* is the minimum k such that all pairs with distance $\geq k$ bend. The *effective curvity* $\tilde{k}$ is the minimum k such that more than 90% of the pairs with distance $\geq k$ bend. Take a look at Table 6 for the results.

The sixth and the last columns in Table 6 present the absolute curvity and the effective curvity of each graph as a function of δ so we can compare it with the upper bound $4\delta + 1$ from Proposition 2. It is clear that the networks have their absolute curvity k^* either equal $2\delta + 1$ or 2δ, and $\delta - 1.5 \leq \tilde{k} \leq 2\delta$. Also, all networks (except for network MACAQUE-BRAIN-1) have their effective curvity less than their absolute curvity which means that the value of the absolute curvity may be affected by a number of outlier vertex pairs.

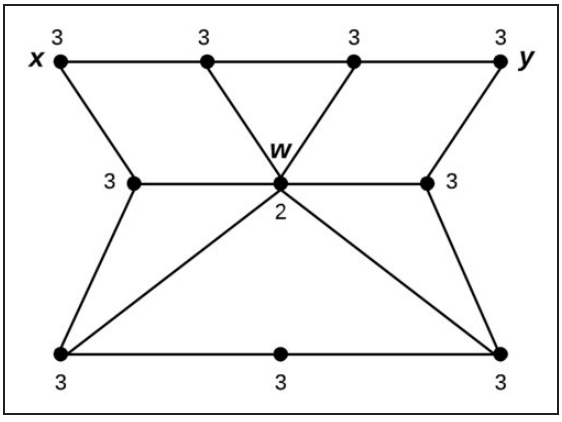

Figure 6. A case in which the shortest path between a diametral pair x, y does not bend toward the graph's center $C(G) = \{w\}$. The number next to each vertex shows the eccentricity of that vertex.

Algorithm 1. Decide the lowest layer μ_k that all vertex pairs with distance $\geq k$ bend to. $2 \leq k \leq diam(G)$.

```
1. for k ← 2 to diam(G) do
2.   max ← -1
3.   for every pair(x, y) do
4.     if (d(x, y) ≥ k) then
5.       if (bend(x,y) > max) then
6.         max ←bend(x,y)
7.       end if
8.     end if
9.   end for
10. μ_k ← max
11. print μ_k
12. end for
```

Motivation and empirical evaluation of (B). Here we examine the impact of the distance on the level to which vertex pairs bend. Let k be the distance between two vertices such that $2 \leq k \leq diam(G)$. Consider μ_k as the lowest layer that all vertex pairs of distance $\geq k$ bend to. We define it as: $\mu_k = \max\{bend(u, v) : \forall u, v \in V$ with $d(u, v) \geq k\}$.

This allows us to look at how the bends of the vertex pairs behave with respect to different distances. The algorithm used to find every μ_k is shown in Algorithm 1. A summary of the results is shown in Figure 7.

Table 6. The absolute curvity k^* and the effective curvity $\tilde{k}$ for our graph datasets. The absolute curvity k^* is the minimum k such that all pairs with distance $\geq k$ bend. The effective curvity $\tilde{k}$ is the minimum k such that more than 90% of the pairs with distance $\geq k$ bend.

Network	$diam(G)$	δ	k^*	$k^* = a\delta + b$	$\tilde{k}$	$\tilde{k} = a\delta + b$
B-YEAST-PI	8	2.5	6	$6 = 2\delta + 1$	2	$2 = \delta - 0.5$
E-COLI-PI	5	2	4	$4 = 2\delta + 0$	3	$3 = \delta + 1$
YEAST-PI	12	3.5	8	$8 = 2\delta + 1$	2	$2 = \delta - 1.5$
S-CEREVISIAE-PI	11	4	8	$8 = 2\delta + 0$	2	$2 = \delta - 2$
H-PYLORI-PI	7	3	6	$6 = 2\delta + 0$	2	$2 = \delta - 1$
MACAQUE-BRAIN-1	4	1.5	3	$3 = 2\delta + 0$	3	$3 = 2\delta + 0$
MACAQUE-BRAIN-2	4	1.5	4	$4 = 2\delta + 1$	3	$3 = 2\delta + 0$
E-COLI-METABOLIC	16	4	9	$9 = 2\delta + 1$	2	$2 = \delta - 2$
C-ELEGANS-METABOLIC	7	1.5	4	$4 = 2\delta + 1$	2	$2 = \delta + 0.5$
YEAST-TRANSCRIPTION	9	3	6	$6 = 2\delta + 0$	2	$2 = \delta - 1$

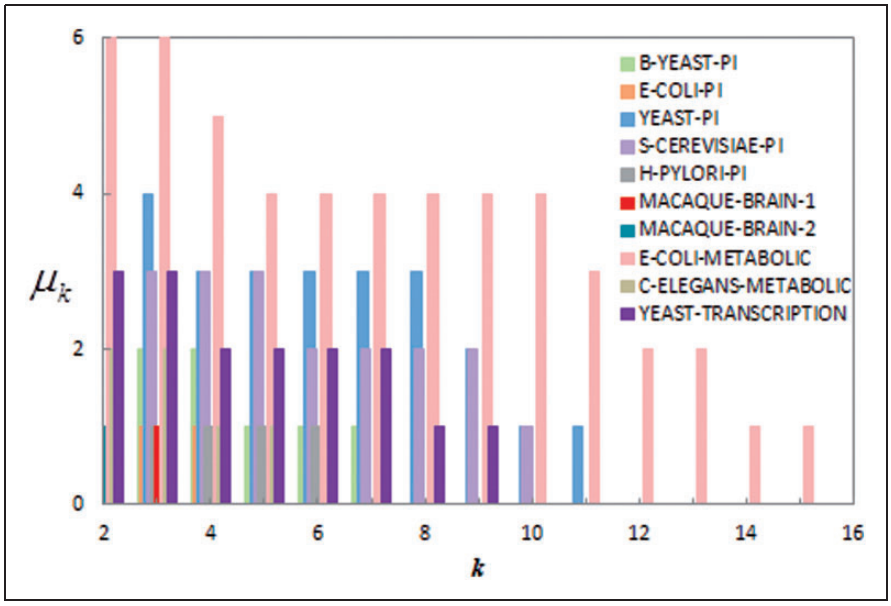

Figure 7. μ_k values for each network in the graph datasets.

As expected, we found a direct relation between the distance of vertex pairs and their bend. For example, Figure 7 shows that in network YEAST-PI, vertex pairs with distances 3, 6, and 9 from one another bend to layers 4, 3, and 2, respectively.

In fact, this observation is a direct implication of Proposition 1 and Proposition 2. For every pair of vertices u and v that are sufficiently far from each other, there is a vertex z in the middle of $\rho(u, v)$ that has less eccentricity (i.e. has a smaller level according to the eccentricity layering) than u or v. Applying this argument recursively to the new vertex pairs (u, z) and (z, v), if the distance between u and z is still sufficiently large, then there is another vertex w in the middle of a shortest path $\rho(u, z)$ such that $ecc(w) < \max\{ecc(u), ecc(z)\}$. The same applies to the vertex pair (z, v).

Also, we know from (A) that the distance among vertex pairs does not need to be very large for a vertex pair to bend. For example, most vertex pairs with distance as small as two from each other bend.

Core–periphery identification using the eccentricity-based bending property

A well-defined center of a graph is a good starting point for locating its core. According to the pattern of data exchange discussed earlier (in which a shortest path between distant vertices bends toward graph's center), we choose to identify the core of the graph using the eccentricity centrality measure (see "Core–periphery and network centrality in complex networks" section).

Even though the center $C(G)$ contains all vertices that are closer to other vertices, this subset is not sufficient to carry out the communication between every pair of vertices ($C(G) \subseteq core(G)$). More vertices should be added to the core according to their participation in routing the traffic among other vertices. We decide the participation of each vertex v based on its eccentricity and whether or not v lies on a shortest path between a pair of vertices x and y (this notion combines the classic concept of the shortest-path betweenness centrality with the eccentricity centrality). We know that a vertex z lies on a shortest path between a pair of vertices u and v if and only if $d(u, v) = d(u, z) + d(z, v)$.

Obviously, not all graphs exhibit a core–periphery structure. Also, graphs follow this structure with different extents with respect to the quality of their cores. We identify a good graph's core as the one that (1) includes a small number of layers with respect to the eccentricity layering; and (2) has size (with respect to the number of vertices) that is small compared to the total number of vertices in the graph. The core should also contain vertices that participate in the majority of interactions among other vertices.

In the following two subsections, we discuss two models that exploit the eccentricity-based bending property to partition the vertices in a graph G into two sets: the core $core(G)$ and the periphery.

Model I. The maximum-peak model. Given a δ-hyperbolic graph $G = (V, E)$ along with its eccentricity layering $\mathcal{EL}(G)$, the maximum-peak model identifies a separation layer index $p \geq 0$ and defines the core as the subset of vertices formed by layers $\ell_0(G), \ell_1(G), \ldots, \ell_p(G)$.

In light of the eccentricity-based bending property, each *bend*(x, y) for a pair of vertices x and y represents a *peak* for shortest paths connecting x and y. In this model, we are locating the index of the lowest layer p over all layers that vertex pairs bend to. Index p represents the separation point where the layers of the graph can be partitioned to a core and a periphery. See Figure 8 for an illustration.

After identifying all peaks, the core will include all vertices starting at $\ell_0(G)$ (the center) until $\ell_p(G)$. Then the periphery will include the vertices in the remaining layers ($\ell_{p+1}(G)$ to $\ell_m(G)$ where m is the index of the last layer in the eccentricity layering). The core $core(G)$ becomes: $core(G) = \bigcup_{r=0}^{p} \ell_r(G)$.

Again, to avoid the impact that outlier vertices may impose (as discussed in "Eccentricity-based bending property of δ-hyperbolic networks" section), we define two types of the separation index p: the *absolute separation index* p^* and the *effective separation index* $\tilde{p}$. The absolute separation index p^* is the lowest layer that all vertex pairs bend to; we call the core defined by this index the absolute core set C^*_{core}. The effective separation index $\tilde{p}$ is the lowest layer where 90% of the vertex pairs bend to, and the core defined by this index is called the effective core set $\tilde{C}_{core}$.

The algorithm used to decide the effective separation index $\tilde{p}$ is presented in Algorithm 2. Table 7 shows the cores for the networks in our graph datasets according to the maximum-peak model.

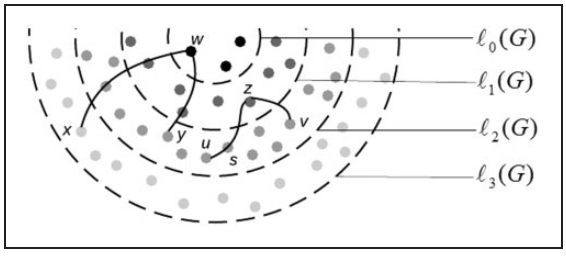

Figure 8. A simplified illustration of the eccentricity layering of a graph (with four layers) and the maximum-peak model. $\ell_r(G)$ represents each layer r, $0 \le r \le 3$. The peaks of shortest paths $\rho(x,y)$ and $\rho(u,v)$ are vertices w and z, respectively. The core contains all vertices of the layers of the peaks, i.e. $core(G) = \ell_0(G) \cup \ell_1(G)$.

Algorithm 2. Decide the effective separation index $\tilde{p}$ to partition the graph into a core and a periphery based on the maximum-peak model. Γ is the total number of vertex pairs with distance ≥ 2. m is the index of the last layer

```
1. cnt ← 0
2.  for r = 0 to m do
3.    for every not counted pair (x,y) with d(x,y) ≥ 2
      do
4.      if (bend(x,y) ≤ r) then
5.        cnt++ and mark pair(x,y) as counted
6.        if (cnt ≥ 90% of Γ) then
7.        p̃ ← r
8.        break and return p̃
9.        end if
10.     end if
11.   end for
12. end for
```

The results in Table 7 show (as expected) a big difference in the sizes (with respect to the number of vertices) of the absolute core and the effective core in the majority of the networks in the datasets. This implies that the identification of the absolute separation index was highly affected by a small percent of vertices.

Table 7. The cores of the graph datasets based on the maximum-peak model. $|V|$ is the number of vertices; $|Layers|$ is the number of layers; C^*_{core}-*lyr* and $|C^*_{core}|$ are the number of layers and number of vertices in the absolute core set; $\tilde{C}_{core}$-*lyr* and $|\tilde{C}_{core}|$ are the number of layers and number of vertices in the effective core set.

Network	$\|V\|$	$\|Layers\|$	C^*_{core}-*lyr*	$\|C^*_{core}\|$	$\|C^*_{core}\|$ to $\|V\|$	$\tilde{C}_{core}$-*lyr*	$\|\tilde{C}_{core}\|$	$\|\tilde{C}_{core}\|$ to $\|V\|$
B-YEAST-PI	1465	4	3	1448	≈ 99%	2	902	≈ 62%
E-COLI-PI	126	3	2	93	≈ 74%	2	93	≈ 74%
YEAST-PI	1728	6	5	1725	≈ 100%	2	472	≈ 27%
S-CEREVISIAE-PI	537	5	5	537	100%	2	223	≈ 42%
H-PYLORI-PI	72	3	2	56	≈ 78%	2	56	≈ 78%
MACAQUE-BRAIN-1	45	3	2	31	≈ 69%	2	31	≈ 69%
MACAQUE-BRAIN-2	350	2	2	350	100%	2	350	100%
E-COLI-METABOLIC	242	8	7	240	≈ 99%	3	102	≈ 42%
C-ELEGANS-METABOLIC	453	4	3	439	≈ 97%	1	17	≈ 4%
YEAST-TRANSCRIPTION	321	5	4	314	≈ 98%	2	62	≈ 19%

Closer analysis to the effective core set $\tilde{C}_{core}$ suggests that deciding the core according to this notion generates good cores (number of layers in the core is small and the number of vertices is about 25% of the total number of vertices) for some networks such as the YEAST-TRANSCRIPTION, C-ELEGANS-METABOLIC, and YEAST-PI. Also, networks with core sizes between 25 and 50% can be considered good as well; such as the cores of the S-CEREVISIAE-PI and E-COLI-METABOLIC. On the other hand, networks like B-YEAST-PI, E-COLI-PI, MACAQUE-BRAIN-1, MACAQUE-BRAIN-2, and H-PYLORI-PI have too large core sizes compared to the overall graph size. This model is highly affected by the distribution of vertices over the layers (see Figure 5). For example, the core of graph B-YEAST-PI has two layers (out of four). This can be considered as a balanced core–periphery separation with respect to the number of layers. However, considering the distribution of the vertices in the four layers, which is 90, 902, 465, and 17, explains the increase in the size of the core.

Recall that the core according to this model is defined by the layers and not by the vertices which may result in the addition of some vertices to the core that do not have real contribution (they were added only for their location in the graph's eccentricity layering). This issue can be resolved by the identification of the core according to the minimum cover set model presented in the following subsection.

Model II. The minimum cover set model. Consider a graph $G=(V,E)$ with the eccentricity layering $\mathcal{EL}(G)$ and with the center $C(G)$. The way this model works is to start the core set $core(G)$ as an empty set and expand it to include vertices which have smaller eccentricity, are closer to the center $C(G)$, and participate in the traffic between other vertices. This expansion should be orderly, first incorporating the vertices that are more eligible (or have higher priority) to be a part of the core, and then moving on to vertices who are less eligible. For each vertex $v \in V$, we define the following three parameters according to which we prioritize the vertices in G.

1. The *eccentricity* $ecc(v)$ (see "Eccentricity layering of a graph" section). Vertices with smaller eccentricities have higher priority to be in the graph's core.
2. The *distance-to-center* for a vertex v, denoted as $f(v)$, which expresses the distance between v and its closest vertex from the center $C(G)$, i.e. $f(v)=d(v,C(G))$. Note that $f(v)\geq 0$. Vertices with small $f(v)$ are closer to the center; therefore, they have higher priority of being in the core. For example, in Figure 8, vertex y is closer to the center than vertex u.
3. The *betweenness* $b(v)$. The definition of the betweenness $b(v)$ of v is close to the definition of the classic shortest-path betweenness centrality. It measures how many pairs of distant vertices x and y have v in one of their shortest paths (versus counting all shortest paths in the classic definition (see "Core–periphery and network centrality in complex networks" section)). It quantifies the participation of a vertex v in the traffic flow process, and we define it as: $b(v)=$ number of pairs $x,y\in V$ with $v\neq x, v\neq y, d(x,y)\geq 2$ and $d(x,v)+d(v,y)=d(x,y)$. According to the core–periphery organization, the betweenness of a vertex should increase as its eccentricity decreases. This means that a vertex that belongs to the central layer of a graph should have a higher value for its $b(v)$ parameter than a vertex that belongs to any other higher layer in the graph's eccentricity layering.

Our goal in this model is to identify the smallest subset of vertices that participate in all traffic throughout the network. The algorithm for this model comprises two stages. First, in a priority list T we lexicographically sort the vertices according to the three attributes: $ecc(v)$, $f(v)$, and $b(v)$. T now has the vertices in the order that they should be considered to become part of the core. The goal is to ensure that for each pair of vertices $x,y\in V$ there exists at least one vertex $v\in core(G)$ such that $v\in\rho(x,y)$. In such case, we say that a shortest path $\rho(x,y)$ from x to y is covered by v (a shortest path from y to x is also covered by the same vertex v since we are dealing with undirected graphs).

Algorithm 3. Decide the core of each graph based on the minimum cover set model. T is a priority list in which vertices are ordered based on: $ecc(v)$, $f(v)$, and $b(v)$. C^*_{core} is the absolute core set

1. $C^*_{core} \leftarrow$ first vertex in T
2. **while** (current core does not cover all vertex pairs) **do**
3. $v \leftarrow$ next vertex in T
4. **if** (v covers a new pair) **then**
5. $C^*_{core} \leftarrow C^*_{core} \cup v$
6. **end if**
7. **end while**
8. return C^*_{core}

The second stage starts with a vertex v at the head of T being removed from T and added to an initially empty set C^*_{core} that represents the absolute core set. This vertex must cover at least one shortest path between a pair of vertices according to the definition of the betweenness $b(v)$. After this initial step, the process continues by repeatedly removing the vertex v at the head of T and adding it to C^*_{core} if and only if v covers a shortest path between an uncovered yet pair x

and y (when there is at least one vertex $v \in C^*_{core}$ that covers a shortest path between x and y, then the pair becomes covered). This step should run until all pairs are covered. Note that we consider the core set C^*_{core} as absolute since all vertex pairs must be covered by a vertex in it.

Now the vertices in set C^*_{core} represent the core of the graph ($core(G)$) while the remaining vertices represent the periphery. Algorithm 3 presents the pseudocode, and the number of vertices in the absolute and the effective core sets of each graph of our datasets is listed in Table 8.

Close analysis of Table 8 shows that each produced absolute core C^*_{core} is of a size between 44 and 86% of the original number of vertices in the graph. It is important to note that vertices in the core are expected to have different contributions (some vertices cover more vertex pairs than others). Figure 9 shows how many vertex pairs are remained uncovered after the orderly addition of vertices to the absolute core. For example, in the network B-YEAST-PI, 80% of vertex pairs are uncovered after adding the first vertex to the absolute core set C^*_{core}. However, after adding 20 vertices to C^*_{core}, only 35% of the vertex pairs are

Table 8. The cores of the graph datasets based on the minimum cover set model. $|V|$ is the number of vertices; $\delta(G)$ is the hyperbolicity; $|C^*_{core}|$ is the number of vertices in the absolute core set; $|\tilde{C}_{core}|$ is the number of vertices in the effective core set; C^*_{MaxLyr} is the largest index layer found among vertices in C^*_{core}; and $\tilde{C}_{MaxLyr}$ is the largest index layer found among vertices in $\tilde{C}_{core}$.

Network	$\|V\|$	$\delta(G)$	$\|C^*_{core}\|$	$\|C^*_{core}\|$ to $\|V\|$	C^*_{MaxLyr}	$\|\tilde{C}_{core}\|$	$\|\tilde{C}_{core}\|$ to $\|V\|$	$\tilde{C}_{MaxLyr}$
B-YEAST-PI	1465	2.5	1117	$\approx$ 76 %	3	117	$\approx$ 8 %	1
E-COLI-PI	126	2	65	$\approx$ 52 %	2	13	$\approx$ 10 %	1
YEAST-PI	1728	3.5	902	$\approx$ 52 %	5	318	$\approx$ 18 %	2
S-CEREVISIAE-PI	537	4	438	$\approx$ 82 %	4	114	$\approx$ 21 %	1
H-PYLORI-PI	72	3	54	$\approx$ 75 %	2	15	$\approx$ 21 %	1
MACAQUE-BRAIN-1	45	1.5	20	$\approx$ 44 %	2	7	$\approx$ 16 %	1
MACAQUE-BRAIN-2	350	1.5	197	$\approx$ 56 %	1	31	$\approx$ 9 %	0
E-COLI-METABOLIC	242	4	208	$\approx$ 86 %	7	66	$\approx$ 27 %	2
C-ELEGANS-METABOLIC	453	1.5	202	$\approx$ 45 %	2	12	$\approx$ 3%	0
YEAST-TRANSCRIPTION	321	3	155	$\approx$ 48 %	4	40	$\approx$ 12%	1

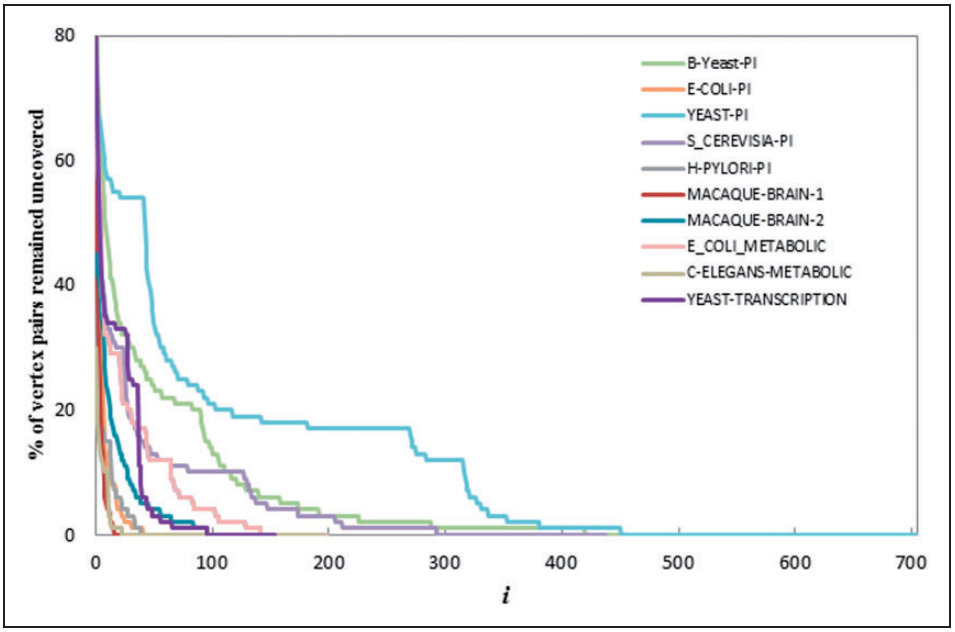

Figure 9. The percentage of the uncovered vertex pairs after the orderly addition of vertices to the core set C^*_{core}. Number i indicates the cardinality of the current core.

uncovered. It is also clear from Figure 9 that many of the vertices that have been added later to the absolute core set cover a very small percentage of the vertex pairs.

To keep only vertices that are considered higher contributors (cover a large number of vertex pairs) we define the effective core set $\tilde{C}_{core}$. The effective core set is the subset of the core that is sufficient to cover shortest paths between 90% of the vertex pairs in the graph. To obtain $\tilde{C}_{core}$, we examine the vertices of the core C^*_{core} in the same order in which they were added. A new vertex is added to current $\tilde{C}_{core}$ only if more than 10% of the vertex pairs remain uncovered. The results on the core according to both concepts in this model are presented in Table 8. Note that the index of the layer of the last vertex added to the core in each network has significantly decreased.

Because hyperbolic graphs adhere to the property of having shortest paths that bend to the core of the graph, it was natural to think that hyperbolic graphs with lower $\delta(G)$ should have even smaller number of vertices in the core. A quick comparison between the $\tilde{C}_{core}$ of each graph with its $\delta(G)$ supports this hypothesis.

Concluding remarks

The structure of several biological networks has been often described as a chain-like or tree-like topology in molecular biology.[11] This motivates investigating if those networks also admit tree-like structures based on different aspects such as their distances. In ''Hyperbolicity of biological networks'' section, we observed that most biological networks appear to have low hyperbolicity suggesting their closeness to a tree structure with respect to their distances.[10]

In the tree structure, the communication among distant vertices is carried out through the center of the tree which is one or two vertices with the smallest eccentricities. The center in this case represents the core of the network while the rest of the vertices belong to the periphery. Since strongly-hyperbolic graphs have a structure that is closer to a tree structure, this motivates the following hypothesis: does hyperbolicity indicate the existence of a sharper core–periphery dichotomy? In other words, do strongly-hyperbolic graphs have more concise cores compared to (weak) hyperbolic graphs?

Before we answer this question we will analyze the networks in our datasets. In ''δ-Hyperbolicity of networks'' section, we differentiated between two types of hyperbolicity that appear in our graphs. First, the low hyperbolicity that is caused by a small diameter or graph size; second, the low hyperbolicity with the value of $\delta(G)$ smaller than or equal to the value of $\log_2(diam(G))$ which in turn is at most $\log_2(\log_2(size(G)))$. Whereas both types are considered hyperbolic graphs, we called the latter category strongly-hyperbolic to point out the similarity in their structure to a tree despite their sizes and their diameters. Here, we noticed that graphs of strong-hyperbolicity are actually small-world networks. We are not saying that the hyperbolicity is intrinsic to small-world networks even though this notion has been suggested in Narayan and Saniee.[6] For instance, a tree with a large diameter is not small world but it is strongly-hyperbolic.

Using the eccentricity-based bending property introduced in ''Eccentricity-based bending property of δ-hyperbolic networks'' section, we defined the core vertices of each graph in our datasets. It is clear from Table 9 that hyperbolic networks have larger cores when compared to strongly-hyperbolic networks (which confirms our hypothesis). Here we only consider cores according to the minimum cover set model to eliminate any interference that may happen because

Table 9. Summary of the graph datasets' parameters and cores: size of the graph $size(G)$, diameter $diam(G)$, hyperbolicity $\delta(G)$, average hyperbolicity $\delta'(G)$, and the core $\tilde{C}_{core}$ is the effective core according to the minimum cover set model.

Network		$\log_2(size(G))$	$diam(G)$	$\delta(G)$	$\delta'(G)$	$\tilde{C}_{core}$	
Strongly-hyperbolic networks	1	C-ELEGANS-METABOLIC	12.3	7	1.5	0.133	3%
		B-YEAST-PI	12.8	8	2.5	0.299	8%
	2	MACAQUE-BRAIN-2	12.4	4	1.5	0.203	9%
		E-COLI-PI	9.5	5	2	0.251	10%
		YEAST-TRANSCRIPTION	10	9	3	0.365	12%
		MACAQUE-BRAIN-1	9	4	1.5	0.231	16%
		YEAST-PI	13.6	12	3.5	0.322	18%
Hyperbolic networks		S-CEREVISIAE-PI	10.5	11	4	0.419	21%
		H-PYLORI-PI	7.5	7	3	0.368	21%
		E-COLI-METABOLIC	9.3	16	4	0.483	27%

of the vertex distribution over different layers of the graph eccentricity layering. The sizes of the cores in strongly-hyperbolic graphs are less than 20% of the number of vertices of each network. In Saniee and Tucci,[26] the authors showed that the small-world property is not enough for a graph to have the core–periphery structure. However, it turns out that our graphs that do not exhibit small-world property also have the core–periphery organization but with larger cores. They also have higher values for $\delta(G)$ ($\delta(G) \geq 3$). Moreover, we see that the more distant the graph is from having the small-world property, (the larger the difference between its $\log_2(size(G))$ and its $diam(G)$), the larger its core is.

We also observed two patterns in strongly-hyperbolic networks named groups 1 and 2 in Table 9. The networks in the first group (the C-ELEGANS-METABOLIC and the B-YEAST-PI) have small hyperbolicity ($\delta(G) < 3$) and in the same time $\delta(G)$ is sufficiently smaller than the value of half the diameter. The cores for those networks are very small (the numbers of vertices in the cores are 3 and 8% of the total number of vertices). The second group has networks that are either with higher hyperbolicity (YEAST-TRANSCRIPTION and YEAST-PI), or low hyperbolicity with value of $\delta(G)$ very close to $diam(G)/2$. The cores for the networks in group 2 are larger than for those in group 1, yet they are small (9–18%).[57]

Acknowledgment

Results of this paper were partially presented at CompleNet 2015.

Declaration of conflicting interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The author(s) received no financial support for the research, authorship, and/or publication of this article.

References

1. Da S, Rosa M, Ma H, et al. Centrality, network capacity, and modularity as parameters to analyze the core-periphery structure in metabolic networks. *Proc IEEE* 2008; 96: 1411–1420.
2. Holme P. Core-periphery organization of complex networks. *Phys Rev E* 2005; 72: 1–4.
3. Borgatti S and Everett M. Models of core/periphery structures. *Soc Netw* 2000; 21: 375–395.
4. Seidman S. Network structure and minimum degree. *Soc Netw* 1983; 5: 269–287.
5. Shanahan M and Wildie M. Knotty-centrality: finding the connective core of a complex network. *PLoS One* 2012; 7: 5.
6. Narayan O and Saniee I. The large scale curvature of networks. *Phys Rev E* 2011; 84: 1–8.
7. Adcock A, Sullivan B and Mahoney M. Tree-like structure in large social and information networks. In: *Proceedings of the IEEE 13th International Conference on Data Mining (ICDM)*, IEEE, 2013, pp.1–10.
8. Montgolfier F, Soto M and Viennot L. Treewidth and hyperbolicity of the internet. In: *Proceedings of the 10th IEEE International Symposium on Network Computing and Applications (NCA)*, IEEE, 2011, pp.25–32.
9. Shavitt Y and Tankel T. On the curvature of the internet and its usage for overlay construction and distance estimation. In: *Proceedings of the 23rd Annual Joint Conference of the IEEE Computer and Communications Societies (INFOCOM)*, IEEE, vol 1, 2004, pp. 374–384.
10. Abu-Ata M and Dragan FF. Metric tree-like structures in real world networks: an empirical study. *Networks* 2016; 67: 49–68.
11. Albert R, DasGupta B and Mobasheri N. Topological implications of negative curvature for biological and social networks. *Phys Rev E* 2014; 89: 1–19.
12. Chen W, Fang W, Hu G, et al. On the hyperbolicity of small-world and treelike random graphs. *Internet Math* 2013; 9: 434–491.
13. Chepoi V, Dragan F, Estellon B, et al. Diameters, centers, and approximating trees of delta-hyperbolic geodesic spaces and graphs. In: *Proceedings of the 24th annual symposium on computational geometry*, ACM, 9 June 2008, pp. 59–68.
14. Chepoi V, Dragan F, Estellon B, et al. Additive spanners and distance and routing labeling schemes for hyperbolic graphs. *Algorithmica* 2012; 62: 713–732.
15. Dourisboure Y. Compact routing schemes for generalised chordal graphs. *J Graph Algorithms Appl* 2005; 9: 277–297.
16. Krioukov D, Fall K and Brady A. On compact routing for the internet. *ACM SIGCOMM Comput Commun Rev* 2007; 37: 41–52.
17. Kennedy S, Narayan O and Saniee I. On the hyperbolicity of large-scale networks. *arXiv prepr* 2013; 1–22.
18. Cvetkovski A and Crovella M. Hyperbolic embedding and routing for dynamic graphs. In: *INFOCOM 2009,* IEEE, 19 April 2009, pp. 1647–1655.
19. Baryshnikov Y. On the curvature of the internet. In: *Workshop on stochastic geometry and teletraffic*, Eindhoven, The Netherlands, 2002.
20. Gromov M. *Hyperbolic groups*. New York: Springer, 1987.
21. Brinkmann G, Koolen J and Moulton V. On the hyperbolicity of chordal graphs. *Ann Combinator* 2001; 5: 61–69.
22. Wu Y and Zhang C. Hyperbolicity and chordality of a graph. *Electr J Combinator* 2011; 18: 43.
23. Chung F and Lu L. The average distances in random graphs with given expected degrees. *Proc Natl Acad Sci* 2002; 99: 15879–15882.

24. Leskovec J, Lang K, Dasgupta A, et al. Community structure in large networks: natural cluster sizes and the absence of large well-defined clusters. *Internet Math* 2009; 6: 29–123.
25. Baryshnikov Y and Tucci G. Asymptotic traffic flow in a hyperbolic network. In: *Proceedings of the 5th International Symposium on Communications Control and Signal Processing (ISCCSP)*, IEEE, 2 May 2012, pp. 1–4.
26. Saniee I and Tucci G. Scaling of congestion in small world networks. *arXiv Prepr* 2012; 1201: 1–8.
27. Shavitt Y and Tankel T. Hyperbolic embedding of internet graph for distance estimation and overlay construction. In: *IEEE/ACM Transactions on Networking (TON)*, vol. 16, 1 February 2008, pp. 25–36.
28. Newman M. A measure of betweenness centrality based on random walks. *Soc Netw* 2005; 27: 39–54.
29. Brandes U and Erlebach T. *Network analysis: methodological foundations*. Springer Science and Business Media, 2005.
30. Freeman L. A set of measures of centrality based on betweenness. *Sociometry* 1977; 35–41.
31. Freeman L. Centrality in social networks conceptual classification. *Soc Netw* 1979; 3: 215–239.
32. Harary F and Hage P. Eccentricity and centrality in networks. *Soc Netw* 1995; 17: 57–63.
33. Raman K. Construction and analysis of protein-protein interaction networks. *Autom Exp* 2010; 2: 2.
34. Greene L and Higman V. Uncovering network systems within protein structures. *J Mol Biol* 2003; 334: 781–791.
35. Stelzl U, Worm U, Lalowski M, et al. A human protein-protein interaction network: a resource for annotating the proteome. *Cell* 2005; 122: 957–968.
36. Fell D and Wagner A. The small world of metabolism. *Nat Biotechnol* 2000; 18: 1121–1122.
37. Giot L, Bader J, Brouwer C, et al. A protein interaction map of drosophila melanogaster. *Science* 2003; 302: 1727–1736.
38. Wagner A and Fell D. The small world inside large metabolic networks. *Proc Biol Sci* 2001; 268: 1803–1810.
39. Jeong H, Tombor B, Albert R, et al. The large-scale organization of metabolic networks. *Nature* 2000; 407: 651–654.
40. Junker B and Schreiber F. *Analysis of biological networks*. John Wiley and Sons, 2008.
41. Luo F, Li B, Wan X, et al. Core and periphery structures in protein interaction networks. *BMC Bioinformat* 2009; 10: S8.
42. Ma H-W and Zeng A-P. The connectivity structure, giant strong component and centrality of metabolic networks. *Bioinformatics* 2003; 19: 1423–1430.
43. Milo R, Shen-Orr S, Itzkovitz S, et al. Network motifs: simple building blocks of complex networks. *Science* 2002; 5594: 824–827.
44. Bu D, Zhao Y, Cai L, et al. Topological structure analysis of the protein–protein interaction network in budding yeast. *Nucl Acids Res* 2003; 31: 2443–2450.
45. Butland G, Peregrin-Alvarez JM, Li J, et al. Interaction network containing conserved and essential protein complexes in *Escherichia coli*. *Nature* 2005; 433: 531–537.
46. Christian V, Krause R, Snel B, et al. Comparative assessment of large-scale data sets of protein–protein interactions. *Nature* 2002; 417: 309–403.
47. Jeong H, Mason S, Barabási A, et al. Lethality and centrality in protein networks. *Nature* 2001; 6833: 41–42.
48. Rain J, Selig L, De Reuse H, et al. The protein–protein interaction map of *Helicobacter pylori*. *Nature* 2001; 6817: 211–215.
49. Modha D and Singh R. Network architecture of the long-distance pathways in the macaque brain. *Proc Natl Acad Sci* 2010; 107: 13485–13490.
50. Négyessy L, Nepusz T, Kocsis L, et al. Prediction of the main cortical areas and connections involved in the tactile function of the visual cortex by network analysis. *Eur J Neurosci* 2006; 23: 1919–1930.
51. Ma H and Zeng A. Reconstruction of metabolic networks from genome data and analysis of their global structure for various organisms. *Bioinformatics* 2003; 19: 270–277.
52. Duch J and Arenas A. Community detection in complex networks using extremal optimization. *Phys Rev* 2005; 72: 027104-1–027104-4.
53. Bermudo S, Rodríguez JM, Sigarreta JM, et al. Gromov hyperbolic graphs. *Discrete Math* 2013; 313: 1575–1585.
54. Fang W. *On hyperbolic geometry structure of complex networks*. Technical report, Report of M1 internship in Microsoft Research Asia, 2011.
55. Rodríguez J, Sigarreta J, Marie Vilaire J, et al. On the hyperbolicity constant in graphs. *Discrete Math* 2011; 311: 211–219.
56. Jonckheere E, Lohsoonthorn P and Bonahon F. Scaled gromov hyperbolic graphs. *J Graph Theory* 2008; 57: 157–180.
57. Alrasheed H and Dragan F. Core-periphery models for graphs based on their δ-hyperbolicity: an example using biological networks. In: *Complex networks VI*, Springer International Publishing, 2015, pp. 65–77.