Core-Periphery Models for Graphs Based on their δ-Hyperbolicity: An Example Using Biological Networks

Hend Alrasheed and Feodor F. Dragan

Department of Computer Science, Kent State University, Kent, OH 44242, USA
`halrashe@kent.edu, dragan@cs.kent.edu`

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.*

1 Introduction

Using graph-theoretical tools for analyzing complex networks aids identifying multiple key properties as well as explaining essential behaviors of those systems. A common structure in many network disciplines is the core-periphery structure which suggests partitioning the graph into a dense core and sparse periphery. Vertices in the periphery interact through a series of core vertices. This pattern of communication (where traffic tends to concentrate on a subset of vertices) has been observed in trees where distant nodes communicate via the central nodes. δ-Hyperbolicity, which is a measure that shows how close a graph is to a tree, suggests that any shortest path between any pair of vertices bends (to some extent) towards the core. This phenomenon has been justified by the negative curvature which in case of graphs can be measured using hyperbolicity [24].

Multiple complex networks such as the Internet [28,14], data networks at the IP layer [24], and social and biological networks [4,2] show low δ-hyperbolicity (low hyperbolicity suggests a structure that is close to a tree structure [14,3]). Also, it has been observed that networks with this property have highly connected cores [24]. Generally, the core of a graph is specified according to one or more centrality measures. For example, the betweenness centrality and the eccentricity centrality. 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 [9] and compact distance and

G. Mangioni et al. (eds.), *Complex Networks VI*,
Studies in Computational Intelligence 597, DOI: 10.1007/978-3-319-16112-9_7

routing labeling schemes [10,12]. In this paper, we investigate implications of the δ-hyperbolicity of a graph and exploit them for the purpose of partitioning the graph into core and periphery parts. Our main contributions can be summarized as follows.

(a) We study the hyperbolicity of several biological networks and show that the hyperbolicity of almost all the networks in our dataset is small. This confirms the results in [4]. 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 graph's diameter and the logarithm of the graph's size. Based on this we classify graphs into: *strongly-hyperbolic*, *hyperbolic*, and *non-hyperbolic*.

(b) 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.

(c) We exploit the eccentricity-based bending property by proposing two core-periphery separation models. We apply both models to our datasets. In contrast to what was observed in [18], we find that biological networks exhibit a clear-cut core-periphery structure. Some details were omitted in this conference version of the paper. Interested readers can refer to [1].

2 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 $size(G) = |V| + |E|$. The *diameter* of the graph $diam(G)$ is the length of the longest shortest path between any two vertices u and v, i.e., $diam(G) = \max_{u,v \in V}\{d(u, v)\}$. 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. The δ-hyperbolicity measure of a metric space was proposed by Gromov [17]. 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 [9]) for Gromov's hyperbolicity. Here we use the 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 non-increasing order, the hyperbolicity of the quadruple x, y, u, v 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 $\delta(G) = \max_{x,y,u,v \in G} \delta(x, y, u, v)$.

For finite graphs δ-hyperbolicity 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, when no finite δ exists (which may be the case for infinite graphs), the graph is considered non-hyperbolic [3]. Generally, the smaller the value of δ the closer the graph is to a tree (metrically).

Core-Periphery and Network Centrality in Complex Networks. In [6], the authors formalize the core-periphery structure by developing two 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. Holme in [18] introduces a coefficient that measures if a network has a clear core-periphery structure based on the closeness centrality. Structure analyses of some biological networks have detected the presence of the core-periphery organization. [13] proposes a parameter that detects the existence of a core-periphery structure in a metabolic network based on the closeness centrality. [16] studies recognizing the central metabolites in a metabolic network. In [21], the authors identify the central metabolites using degree and closeness centrality.

In the study of communication networks, the core is usually identified by the small dense part that carries out most traffic under shortest path routing [5,24]. It is quite natural to associate the concepts of the network's core and its center. In [6], the authors argue that each central vertex is a core vertex; consequently, all coreness measures are centrality measures. The notion behind centrality is identifying vertices that are high contributors. There are multiple centrality measures in the literature. The betweenness centrality expresses how much effect each vertex has in the communication. Given a connected finite graph $G = (V, E)$, the betweenness centrality 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. The eccentricity centrality suggests that the center of the graph includes the vertex (or vertices) that has the shortest distance to all other vertices.

3 Datasets

We analyze the protein interaction networks of Budding yeast [7], Escherichia coli [8], Yeast [11], Saccharomyces cerevisiae [19], and Helicobacter Pylori [26]. Also, we analyze two brain area networks of the macaque monkey [25] [23]; and the metabolic networks of the Escherichia coli [20] and the Caenorhabditis elegans [15]. Finally, we analyze the yeast transcription network [22]. 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.

4 δ-Hyperbolicity of Networks

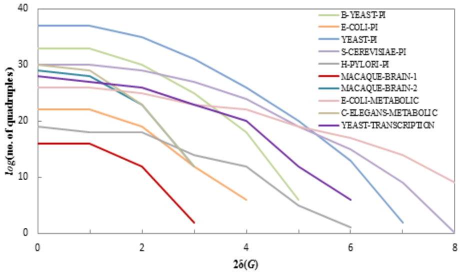
For the purpose of investigating the hyperbolicity of networks, it seems natural to analyze and classify them based on their hyperbolicity. The classification should reflect how strong (evident) the tree-likeness is in the graph's structure.

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

Network Category	Network	$\lvert V\rvert$	$\lvert E\rvert$	$\log_2(size(G))$	$\bar{d}$	$diam(G)$	$rad(G)$	$\delta(G)$	$\delta^{'}(G)$
PI Networks	B-yeast-PI	1465	5839	12.8	7.97	8	5	2.5	0.299
	E-coli-PI	126	581	9.5	9.2	5	3	2	0.251
	Yeast-PI	1728	11003	13.6	12.7	12	7	3.5	0.322
	S-cerevisiae-PI	537	1002	10.5	3.7	11	7	4	0.419
	H-pylori-PI	72	112	7.5	3.1	7	5	3	0.368
Neural Networks	Macaque-brain-1	45	463	9	11.3	4	2	1.5	0.231
	Macaque-brain-2	350	5198	12.4	29.7	4	3	1.5	0.203
Metabolic Networks	E-coli-metabolic	242	376	9.3	3.1	16	9	4	0.483
	C-elegans-metabolic	453	4596	12.3	8.9	7	4	1.5	0.133
Transcription Networks	Yeast-transcription	321	711	10	4.4	9	5	3	0.365

Hyperbolicity of Biological Networks. We measure δ-hyperbolicity 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 [14].

Table 1 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 [14,3], closer analysis to the distribution of the value of δ (see Figure 1) shows that only a very small percent of the quadruples have the maximum value of δ while most quadruples have $\delta = 0$. This observation makes it equally important to calculate the value of the average delta $\delta'(G)$ (see Table 1).



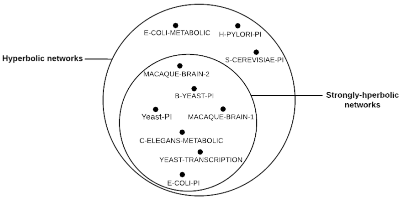


Fig. 1. The distribution of the quadruples over different values of δ

Fig. 2. Classification of the graph datasets based on their hyperbolicity

Analysis and Discussion. Our goal is to categorize graphs with respect to their hyperbolicity into three classes: *strongly-hyperbolic, hyperbolic*, and *non-hyperbolic*. Studying the tree-like structure of graphs based solely on the value of the hyperbolicity may not be sufficient for two reasons. First, the hyperbolicity is a relative measure. For example, for a given graph, a value of $\delta(G) = 10$ can be seen as too large when $size(G) \simeq 10^2$. However, when $size(G) \simeq 10^7$, the hyperbolicity $\delta(G) = 10$ looks much smaller. Second, small graph size and (or) small diameter directly yield low hyperbolicity. 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 $size(G)$ and $diam(G)$ 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 non-hyperbolic class includes only infinite graphs. However, in this study, we only consider finite graphs; accordingly, a non-hyperbolic 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 $\delta(G) \leq \log_2(size(G))$.

In cases where $\delta(G) \leq \log_2(size(G))$, we move on and compare δ with the diameter. To guarantee that the value of the diameter is not directly impacted by the graph's size, first we require that $diam(G) \leq \log_2(size(G))$. Multiple previous works have analyzed the relationship between $\delta(G)$ and the diameter.

Lemma 1 ([27]). *For any graph G with diameter $diam(G)$ and hyperbolicity $\delta(G)$, $\delta(G) \leq diam(G)/2$.*

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) $\delta(G) \leq \log_2(diam(G))$ (small-world), *hyperbolic* when it violates either (1) or (2), and *non-hyperbolic* whenever it has a large δ, i.e., $\delta > \log_2(size(G))$. As Table 1 shows, all networks in the datasets, with the exception of S-CEREVISIAE-PI and E-COLI-METABOLIC, exhibit the small-world property. Also, it 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 $\delta(G)$ and $\delta^{'}(G)$ values are on the larger side. In Figure 2, we show this classification.

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 network C-ELEGANS-METABOLIC is metrically closer to a tree than network YEAST-PI.

5 Core-Periphery Models Based on δ-Hyperbolicity

In this section, we formalize the notion of bending in shortest paths by introducing 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.

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 inner-most layer (layer 0) encloses the graph's center $C(G)$. Then the first layer includes all vertices with eccentricities equal to $rad(G) + 1$, and so on. The vertices in the last layer (outer-most) have eccentricities equal to the diameter. Any vertex $v \in \ell_r(G)$ has *level* (or layer) $level(v) = r$. Figure 3 gives an illustration. We noticed that the vertices' population is denser in the middle layers in almost all networks.

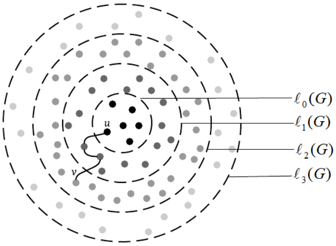


Fig. 3. Eccentricity layering of a graph. Darker vertices belong to lower layers.

Table 2. 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 networks.

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%

5.1 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 [9], the following useful metric property of δ-hyperbolic graphs was proven.

Lemma 2 ([9]). *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. The proofs are omitted in this version. We direct interested readers to [1].

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

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 $ecc(w) < \max\{ecc(x), ecc(y)\}$.*

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 $\forall\ u, v \in V\ bend(u, v) =$

$min\{level(z) : z \in V$ 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 them. In this case we say also that *pair of vertices u and v bends.* The parameter *bend* decides the extent (or the level) to which shortest paths curve towards the center. Note that in some cases $bend(u,v)$ will be assigned either $ecc(u)$ or $ecc(v)$, whatever is smaller. For example, see $\rho(u,v)$ in Figure 3.

Now we investigate the effect of the distance between a vertex pair on its bend. Our findings in this context are summarized in the following statements.

(A) Despite their distances, most vertex pairs bend. Moreover, among those bending pairs, the majority 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. Interestingly, we noticed the bend in the majority of shortest paths. A quick look at Table 2 shows that a big percent of vertex pairs of distance at least two bend.

To quantify the distances at which the bend happens, 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. When the values of k^* and $\tilde{k}$ of each graph are represented as a function of δ to compare it with the upper bound $4\delta + 1$, we find that the networks have their k^* either equal to $2\delta + 1$ or to 2δ, and $\delta - 2 \leq \tilde{k} \leq 2\delta$. Also, all networks (except for MACAQUE-BRAIN-1) have their $\tilde{k}$ less than their k^*.

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

5.2 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, we identify the core using the eccentricity centrality measure. Even though the center contains all vertices that are closer to other vertices, this subset is not sufficient. More vertices should be added to the core according to their participation in routing

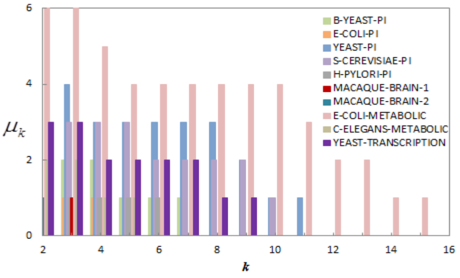


Fig. 4. μ_k values for each network in the graph datasets

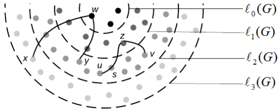


Fig. 5. Illustration of the eccentricity layering of a graph and the Maximum-Peak model. $\ell_r(G)$ represents each layer r. The peaks of $\rho(x, y)$ and $\rho(u, v)$ are w and z.

the traffic. We decide the participation of each vertex based on its eccentricity and whether or not it lies on a shortest path between a vertex pair.

Graphs follow the core-periphery 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 a 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 subsections, we discuss two core-periphery separation models.

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 $\ell_0(G), \ell_1(G), ..., \ell_p(G)$.

In light of the eccentricity-based bending property, each *bend*(x, y) for a pair of vertices x and y represents a *peak* for $\rho(x, y)$. 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 can be partitioned to a core and a periphery. See Figure 5 for an illustration. After identifying all peaks, the core will include all vertices starting at $\ell_0(G)$ until $\ell_p(G)$, i.e., $core(G) = \bigcup_{r=0}^{p} \ell_r(G)$. Then the periphery will include the vertices in the remaining layers.

Again, to avoid the impact that outlier vertices may impose, we define two types of 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 the effective core set $\tilde{C}_{core}$. Table 3 shows the cores for the networks in our datasets according to this model.

Table 3 shows a big difference in the sizes of the absolute core and the effective core in the majority of the networks. Closer analysis to $\tilde{C}_{core}$ 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-PI. Also, networks with core sizes between 25% - 50% can be considered good as well; such as the core of the S-cerevisiae-PI. On the other hand, networks like E-coli-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. 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. 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. This issue can be resolved by using the second model.

Table 3. 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%

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 as an empty set and expand it to include vertices which have smaller eccentricity, are closer to the center, and participate in the traffic. This expansion should be orderly, first incorporating the vertices that have higher priority, and then vertices who are less eligible. For each vertex $v \in V$, we define three parameters according to which we prioritize the vertices.

- The *eccentricity* $ecc(v)$. Vertices with smaller eccentricities have higher priority to be in the graph's core.
- The *distance-to-center*, denoted as $f(v)$, which expresses the distance between v and its closest vertex from the center $C(G)$, i.e., $f(v) = d(v, C(G))$. Vertices with small $f(v)$ have higher priority of being in the core. For example, in Figure 5, vertex y is closer to the center than u.
- The *betweenness* $b(v)$. The betweenness 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 of the betweennes). 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.

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 there exists at least one vertex $v \in core(G)$ such that $v \in \rho(x, y)$ for each pair of vertices $x, y \in V$. In such case, we say that a shortest path $\rho(x, y)$ is covered by v (a shortest path from y to x is also covered by v since we are dealing with undirected graphs).

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 pair. 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 an uncovered yet pair x and y (when there is at least one vertex $v \in C^*_{core}$ that covers a pair (x, y), then it 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 while the remaining vertices represent the periphery. The number of vertices in the absolute and the effective core sets of each graph in our datasets is listed in Table 4.

Table 4. 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	≈ 76 %	3	117	≈ 8 %	1
E-coli-PI	126	2	65	≈ 52 %	2	13	≈ 10 %	1
Yeast-PI	1728	3.5	902	≈ 52 %	5	318	≈ 18 %	2
S-cerevisiae-PI	537	4	438	≈ 82 %	4	114	≈ 21 %	1
H-pylori-PI	72	3	54	≈ 75 %	2	15	≈ 21 %	1
Macaque-brain-1	45	1.5	20	≈ 44 %	2	7	≈ 16 %	1
Macaque-brain-2	350	1.5	197	≈ 56 %	1	31	≈ 9 %	0
E-coli-metabolic	242	4	208	≈ 86 %	7	66	≈ 27 %	2
C-elegans-metabolic	453	1.5	202	≈ 45 %	2	12	≈ 3%	0
Yeast-transcription	321	3	155	≈ 48 %	4	40	≈ 12%	1

Close analysis of Table 4 shows that each produced absolute core C^*_{core} is of a size between 44% to 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 6 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 C^*_{core}. However, after adding 20 vertices, only 35% of the vertex pairs are uncovered. It is also clear that many of the vertices that have been added later to the absolute core set cover a very small percentage of vertex pairs.

To keep only vertices that are considered higher contributors we define the effective core set $\tilde{C}_{core}$. The effective core is the subset of the core that is sufficient to cover 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 4. 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, it was natural to think that hyperbolic graphs with lower $\delta(G)$ should have even smaller cores. A quick comparison between the $\tilde{C}_{core}$ of each graph with its $\delta(G)$ supports this idea.

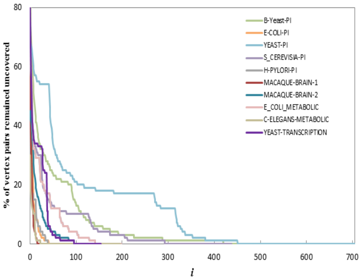


Fig. 6. The percentage of the uncovered vertex pairs after the orderly addition of vertices to the core set C^*_{core}. Number i indicates the cardinality of the current core.

Table 5. Summary of the graph datasets' parameters and cores. $\tilde{C}_{core}$ is the effective core according to the Minimum Cover Set model.

		Network	$\log_2(size)$	*diam*	δ	δ'	$\tilde{C}_{core}$
Strongly-hyperbolic Networks	1	C-elegans-metab.	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-transcr.	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-metab.	9.3	16	4	0.483	27%

6 Concluding Remarks

The structure of several biological networks has been often described as a tree-like topology in molecular biology [4]. This motivates investigating if those networks also admit tree-like structures based on their distances. In Section 4, we observed that most biological networks appear to have low hyperbolicity. Since strongly-hyperbolic graphs have a structure that is closer to a tree, this motivates the following hypothesis: do strongly-hyperbolic graphs have more concise cores compared to other hyperbolic graphs? It is clear from Tables 5 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. The sizes of the cores in strongly-hyperbolic graphs are less than 20% of the number of vertices of each network.

We also observed two patterns in strongly-hyperbolic networks named groups 1 and 2 in Table 5. The networks in group 1 have $\delta(G) < 3$ 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 second group has networks that are either with higher hyperbolicity, or low hyperbolicity with value of $\delta(G)$ very close to $diam(G)/2$. The cores for group 2 are larger than those in group 1.

References

[1] `http://www.kent.edu/~dragan/FullVersionComplexNet.pdf`

[2] Abu-Ata, M., Dragan, F.F.: Metric tree-like structures in real-life networks: an empirical study. arXiv preprint 3364(1402) (2014)

[3] Adcock, A., Sullivan, B., Mahoney, M.: Tree-like structure in large social and information networks. In: ICDM 2013 (2013)

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

[5] Baryshnikov, Y., Tucci, G.: Asymptotic traffic flow in a hyperbolic network. In: ISCCSP (2012)

[6] Borgatti, S., Everett, M.: Models of core/periphery structures. Social Networks 21(4), 375–395 (2000)

[7] Bu, D., Zhao, Y., et al.: Topological structure analysis of the protein-protein interaction network in budding yeast. Nucleic Acids Research 31(9), 2443–2450 (2003)

[8] Butland, G., Manuel, J., et al.: Interaction network containing conserved and essential protein complexes in escherichia coli. Nature 7025(433), 531–537 (2005)

[9] Chepoi, V., Dragan, F., et al.: Diameters, centers, and approximating trees of delta-hyperbolic geodesic spaces and graphs. In: SoCG 2008 (2008)

[10] Chepoi, V., Dragan, F., et al.: Additive spanners and distance and routing labeling schemes for hyperbolic graphs. Algorithmica 62(3), 713–732 (2012)

[11] Christian, V.M., Krause, R., et al.: Comparative assessment of large-scale data sets of protein-protein interactions. Nature 6887(417), 309–403 (2002)

[12] Cvetkovski, A., Crovella, M.: Hyperbolic embedding and routing for dynamic graphs. In: INFOCOM 2009 (2009)

[13] Da, S., Rosa, M., et al.: Centrality, network capacity, and modularity as parameters to analyze the core-periphery structure in metabolic networks. Proceedings of the IEEE 96, 1411–1420 (2008)

[14] de Montgolfier, F., Soto, M., Viennot, L.: Treewidth and hyperbolicity of the internet. In: 10th IEEE International Symposium, NCA, pp. 25–32. IEEE (2011)

[15] Duch, J., Arenas, A.: Community detection in complex networks using extremal optimization. Physical Review 72(2) (2005)

[16] Fell, D., Wagner, A.: The small world of metabolism. Nature Biotechnology 18(11), 1121–1122 (2000)

[17] Gromov, M.: Hyperbolic Groups. Springer, New York (1987)

[18] Holme, P.: Core-periphery organization of complex networks. Physical Review E 72(4) (2005)

[19] Jeong, H., Mason, S., et al.: Lethality and centrality in protein networks. Nature 6833(411), 41–42 (2001)

[20] Ma, H., Zeng, A.-P.: Reconstruction of metabolic networks from genome data and analysis of their global structure for various organisms. Bioinformatics 19(2), 270–277 (2003)

[21] Ma, H.-W., Zeng, A.-P.: The connectivity structure, giant strong component and centrality of metabolic networks. Bioinformatics 19(11), 1423–1430 (2003)

[22] Milo, R., Shen-Orr, S., et al.: Network motifs: simple building blocks of complex networks. Science 5594(298), 824–827 (2002)

[23] Modha, D., Singh, R.: Network architecture of the long-distance pathways in the macaque brain. Proceedings of NAS 107

[24] Narayan, O., Saniee, I.: The large scale curvature of networks. Physical Review E 84(6) (2011)

[25] Négyessy, L., Nepusz, T., et al.: Prediction of the main cortical areas and connections involved in the tactile function of the visual cortex by network analysis. European Journal of Neuroscience 23(7), 1919–1930 (2006)

[26] Rain, J.-C., Selig, L., et al.: The protein-protein interaction map of helicobacter pylori. Nature 6817(409), 211–215 (2001)

[27] Rodríguez, J., Sigarreta, J., et al.: On the hyperbolicity constant in graphs. Discrete Mathematics 311(4) (2011)

[28] Shavitt, Y., Tankel, T.: On the curvature of the internet and its usage for overlay construction and distance estimation. In: INFOCOM 2004. IEEE (2004)