Curvature and temperature of complex networks

Dmitri Krioukov,[1] Fragkiskos Papadopoulos,[1] Amin Vahdat,[2] and Marián Boguñá[3]

[1]*Cooperative Association for Internet Data Analysis (CAIDA), University of California — San Diego (UCSD), La Jolla, California 92093, USA*

[2]*Department of Computer Science and Engineering, University of California–San Diego (UCSD), La Jolla, California 92093, USA*

[3]*Departament de Física Fonamental, Universitat de Barcelona, Martí i Franquès 1, 08028 Barcelona, Spain*

We show that heterogeneous degree distributions in observed scale-free topologies of complex networks can emerge as a consequence of the exponential expansion of hidden hyperbolic space. Fermi-Dirac statistics provides a physical interpretation of hyperbolic distances as energies of links. The hidden space curvature affects the heterogeneity of the degree distribution, while clustering is a function of temperature. We embed the internet into the hyperbolic plane and find a remarkable congruency between the embedding and our hyperbolic model. Besides proving our model realistic, this embedding may be used for routing with only local information, which holds significant promise for improving the performance of internet routing.

PACS number(s): 89.75.Hc, 02.40.−k, 67.85.Lm, 89.75.Fb

Many complex networks possess heterogeneous degree distributions. This heterogeneity is often modeled by power laws often truncated [1]. These networks also exhibit strong clustering, i.e., high concentration of triangular subgraphs. Our previous work [2] demonstrated that the clustering peculiarities of complex networks and, in particular, their self-similarity finds a natural geometric explanation in the existence of hidden metric spaces underlying the network and abstracting the intrinsic similarities between its nodes. Here we seek to provide a geometric interpretation of the first property—network heterogeneity. We show that heterogeneous or scale-free degree distributions in complex networks appear as a simple consequence of negative curvature of hidden spaces. That is, we argue that these spaces are hyperbolic.

The main metric property of hyperbolic geometry is the exponential expansion of space (see Fig. 1, left). For example, in the hyperbolic plane, i.e., the two-dimensional space of constant curvature -1, the length of a circle and the area of a disk of radius R are $2\pi \sinh R$ and $2\pi(\cosh R - 1)$, both growing as $\sim e^R$. The hyperbolic plane is thus metrically equivalent to an e-ary tree, i.e., a tree with the average branching factor equal to e. Indeed, in a b-ary tree the surface of a sphere or the volume of a ball of radius R, measured as the number of nodes lying at or within R hops from the root, grow as b^R. Informally, hyperbolic spaces can therefore be thought of as "continuous versions" of trees.

To see why this exponential expansion of hidden space is intrinsic to complex networks, observe that their topology represents the structure of connections or interactions among distinguishable heterogeneous elements abstracted as nodes. This heterogeneity implies that nodes can be somehow classified, however broadly, into a taxonomy, i.e., nodes can be split into large groups consisting of smaller subgroups, which in turn consist of even smaller subsubgroups. The relationships between such groups and subgroups can be approximated by treelike structures, sometimes called *dendrograms*, in which the distance between two nodes estimates how similar they are [3]. Importantly, the node classification hierarchy need not be strictly a tree. Approximate "treeness," which can be formally expressed solely in terms of the metric structure of a space [4], makes the space hyperbolic.

Let us see what network topologies emerge in the simplest possible settings involving hidden hyperbolic metric spaces. Let us form a network of $N \gg 1$ nodes located in the hyperbolic plane $\mathbb{H}^2$. Since the number of nodes is finite, the area that nodes occupy is bounded. Let $R \gg 1$ be the radius of a
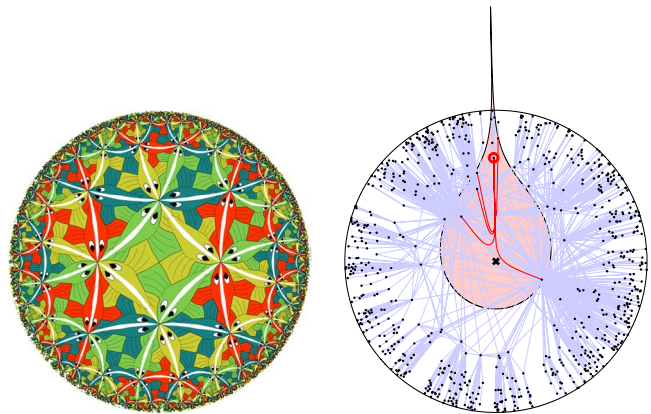

FIG. 1. (Color online) Left: artistic visualization of the Poincaré disk model of the hyperbolic plane $\mathbb{H}^2$ by Levy based on Escher's *Circle Limit III* with the permission from the Geometry Center, University of Minnesota. The exponential expansion of fish illustrates the exponential expansion of hyperbolic space. All fish are of the same hyperbolic size but their Euclidean size exponentially decreases, while their number exponentially increases with the distance from the origin. Right: a modeled network with $N=740$ nodes, power-law exponent $\gamma=2.2$, and average degree $\bar{k} \approx 5$ embedded in the hyperbolic disk of curvature $K=-1$ and radius $R \approx 15.5$. The Euclidean distance between a node and the origin at the disk center, shown as the cross, represents the true hyperbolic distance between the two. But the Euclidean distance between any two other nodes is *not* equal to the hyperbolic distance between them as indicated by the peculiar shape of the shaded hyperbolic *disk* centered at the circled node located at distance $r=10.6$ from the origin. The hyperbolic radius of this disk is also R, and according to the model, the circled node is connected to all the nodes lying in this disk. The curves show the hyperbolically straight lines, i.e., geodesics, connecting the circled node and some nodes in its disk.

disk within which nodes are uniformly distributed. In hyperbolic geometry, this means that nodes are given an angular coordinate θ randomly distributed in $[0, 2\pi]$ and a radial coordinate r following the density $\rho(r) = \sinh r / (\cosh R - 1)$ $\approx e^{r-R}$. Next, we have to specify the connection probability $p(x)$ that two nodes at hyperbolic distance x are connected. We first consider the simplest case, the step function $p(x) = \Theta(R - x)$, and justify this choice later. This $p(x)$ connects each pair of nodes if the hyperbolic distance between them is not larger than R.

The network is now formed, and we can compute the average degree $\bar{k}(r)$ of nodes at distance r from the disk center. These nodes are connected to all nodes in the intersection area of the two disks of the same radius R, one in which all nodes reside, and the other centered at distance r from the center of the first disk (see Fig. 1, right). Since the node distribution is uniform, $\bar{k}(r)$ is proportional to the area of this intersection, which decreases exponentially with r, $\bar{k}(r) \sim e^{-r/2}$. Therefore, the inverse function is logarithmic, $\bar{r}(k) \sim -2 \ln k$, and the node degree distribution in the network is approximately a power law, $P(k) \approx \rho[\bar{r}(k)]|\bar{r}'(k)| \sim k^{-3}$. If we generalize the space curvature to $K = -\zeta^2$, $\zeta > 0$, and the node density to $\rho(r) \approx \alpha e^{\alpha(r-R)}$, where we can think of $\alpha > 0$ as the logarithm of the average branching factor in the underlying hierarchy, then the average degree at radius r scales as $\bar{k}(r) \sim e^{-\zeta r/2}$ if $\alpha/\zeta \geq 1/2$ or $\bar{k}(r) \sim e^{-\alpha r}$ otherwise so that the node degree distribution becomes $P(k) \sim k^{-\gamma}$ with

$$\gamma = \begin{cases} 2\alpha/\zeta + 1 & \text{if } \alpha/\zeta \geq 1/2 \\ 2 & \text{otherwise.} \end{cases} \tag{1}$$

To fix the average degree in the network, we have to choose $N = c e^{\zeta R/2}$, where c is a constant. The result in Eq. (1) is remarkable as it shows that heterogeneous degree distributions may emerge as a simple consequence of the exponential expansion of hyperbolic space.

However, our choice of the step-function connection probability is not yet justified. To justify it and to show that scale-free networks have effective hyperbolic geometries underneath, we recall the $\mathbb{S}^1$ model introduced in [2]. In that model, networks are constructed as follows. First, distribute N nodes uniformly over the circle $\mathbb{S}^1$ of radius $N/(2\pi)$ so that the node density on the circle is fixed to 1. Second, assign to all nodes an additional hidden variable κ representing their expected degrees. To generate scale-free networks, the variable κ is power law distributed according to $\rho(\kappa) = \kappa_0^{\gamma-1}(\gamma-1)\kappa^{-\gamma}$, $\kappa \in [\kappa_0, \infty)$, where κ_0 is the minimum expected degree. Finally, let κ and κ' be the expected degrees of two nodes located at distance $d = N\Delta\theta/(2\pi)$ measured over the circle ($\Delta\theta$ is the angular distance between the nodes). We connect each pair of nodes with probability $\tilde{p}(\chi)$, where $\chi \equiv d/(\mu\kappa\kappa')$ and constant μ fixes the average degree in the network.

The key point is that the connection probability $\tilde{p}(\chi)$ can be *any integrable function*. As long as the distance over the circle is rescaled as $\chi \sim d/(\kappa\kappa')$, any integrable $\tilde{p}(\chi)$ guarantees that the expected degree of nodes with hidden variable κ

is indeed κ, $\bar{k}(\kappa) = \kappa$, so that γ, which is a model parameter, is indeed the exponent of the degree distribution in generated networks.

We now want to map the expected degree κ of each node to a radial position r within a disk of radius R, such that after the mapping, the radial distribution of nodes is $\rho(r) \approx \alpha e^{\alpha(r-R)}$, i.e., as in the hyperbolic $\mathbb{H}^2$ model introduced above. To have this $\rho(r)$, we must select the $\kappa \rightarrow r$ mapping according to

$$\kappa = \kappa_0 e^{\zeta(R-r)/2}, \quad \frac{\zeta}{2} = \frac{\alpha}{\gamma-1}, \quad N = c e^{\zeta R/2}, \quad c = \pi\mu\kappa_0^2, \tag{2}$$

where ζ is fixed by the values of γ and target α. We see that $\kappa(r)$ and consequently $\bar{k}(r)$ scale with r as in the $\mathbb{H}^2$ model, while the connection probability $\tilde{p}(\chi)$ becomes $\tilde{p}(e^{\zeta(x-R)/2})$, where the effective distance

$$x = r + r' + \frac{2}{\zeta}\ln\frac{\Delta\theta}{2} \tag{3}$$

is approximately equal to the hyperbolic distance between the two nodes in the disk. Indeed, the true hyperbolic distance x between two points with polar coordinates (r, θ) and (r', θ') in the hyperbolic space $\mathbb{H}^2$ of curvature $K = -\zeta^2$ is $\cosh \zeta x = \cosh \zeta r \cosh \zeta r' - \sinh \zeta r \sinh \zeta r' \cos \Delta\theta$, which for sufficiently large ζr, $\zeta r'$, and $\Delta\theta > 2\sqrt{e^{-2\zeta r} + e^{-2\zeta r'}}$ is closely approximated by

$$x = r + r' + \frac{2}{\zeta}\ln \sin\frac{\Delta\theta}{2}. \tag{4}$$

Since the effective and true hyperbolic distances in Eqs. (3) and (4) are approximately equal, the value of ζ in Eq. (2) is indeed the square root of curvature of the hyperbolic disk, in agreement with Eq. (1) in the $\mathbb{H}^2$ model. We also notice that since the connection probability $\tilde{p}(\chi)$ in the $\mathbb{S}^1$ model can be any integrable function, the connection probability $p(x)$ in the $\mathbb{H}^2$ model can be *any function* of the form $p(x) = \tilde{p}(e^{\zeta(x-R)/2})$.

Given this freedom of choice of the connection probability, let us consider the family of functions

$$p(x) = \frac{1}{1 + e^{\zeta(x-R)/(2T)}} \tag{5}$$

parameterized by $T > 0$. One motivation to focus on this family is that it generates exponential random graphs in the statistical mechanics sense [5]. Equation (5) is nothing but the grand canonical Fermi-Dirac distribution, and T is the system temperature. From the physical perspective, graph edges are noninteracting fermions with energies equal to their hidden hyperbolic lengths, and R is the chemical potential defined by the condition that $\bar{k}N/2$, the number of edge fermions, is fixed on average. At $T \rightarrow 0$ Eq. (5) converges to $p(x) = \Theta(R - x)$, which *a posteriori* justifies our choice of the step-function connection probability in the $\mathbb{H}^2$ model.

The dependence on temperature in the model is peculiar. At zero temperature, the network is in the strongly degener-
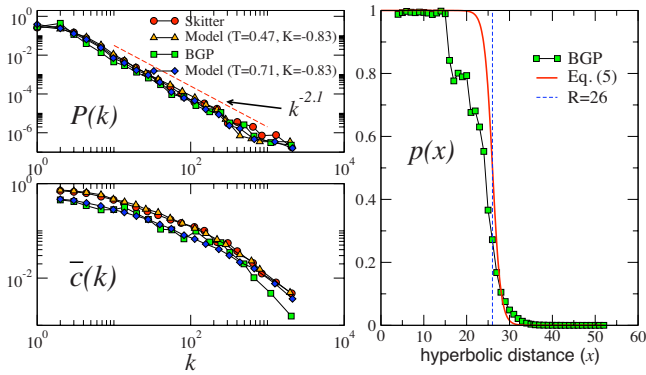
FIG. 2. (Color online) Networks in the $\mathbb{H}^2$ model vs the internet. Left: the degree distribution $P(k)$ and degree-dependent clustering coefficient $\bar{c}(k)$ are shown for the skitter ($\bar{k}=6.29$, $\bar{C}=0.46$) and border gateway protocol (BGP) ($\bar{k}=4.68$, $\bar{C}=0.29$) views of the internet from [6] and for modeled networks with curvature $K=-0.83$ and two values of temperature T, 0.47 ($\bar{k}=6.03$, $\bar{C}=0.44$) and 0.71 ($\bar{k}=4.85$, $\bar{C}=0.25$). Right: the empirical connection probability in the hyperbolically embedded internet, compared to Eq. (5).

ate ground state. As we heat it up, particles explore higher-energy states, i.e., edges connect longer distances, which affects clustering. At $T \to 0$, clustering is maximized. It monotonically decreases with T, and at $T \to 1$ we have a phase transition with clustering going to zero and the network losing its cold-state metric structure. In the cold regime with $T < 1$, the exponent of the degree distribution γ depends only on the ratio α/ζ via Eq. (1). Therefore, we can set $\alpha = 1/2$ without loss of generality so that $\gamma = 1/\zeta + 1$ is fully defined by curvature $K > -1$. In the hot regime with $T > 1$, clustering remains zero, the chemical potential is no longer given by $N = ce^{\zeta R/2}$ but by $N = ce^{\zeta R/(2T)}$, and γ also depends on temperature, $\gamma = T/\zeta + 1$. Therefore at $T \to \infty$ the graph ensemble is identical to classical random graphs, as all fermions are uniformly distributed across all energies, i.e., all pairs of nodes are connected with the same probability independent of the hidden distance between them, and the network loses its cold-state hierarchical structure. Combining the cold and hot regimes,

$$\gamma = \begin{cases} 1/\zeta + 1 & \text{if } T < 1 \text{ and } \zeta < 1, \\ T/\zeta + 1 & \text{if } T > 1 \text{ and } \zeta < T, \\ 2 & \text{otherwise.} \end{cases} \quad (6)$$

Finally, constant c fixing the average degree in the network is

$$c \approx \begin{cases} \bar{k} \dfrac{\sin \pi T}{2T}(1-\zeta)^2 \approx \kappa_0^2 \dfrac{\sin \pi T}{2\bar{k}T} & \text{if } T < 1 \\ \bar{k}\left(\dfrac{\pi}{2}\right)^{1/T} \dfrac{T-1}{T^3}(T-\zeta)^2 \underset{T\to\infty}{\to} \bar{k} & \text{if } T > 1. \end{cases} \quad (7)$$

The $\mathbb{H}^2$ model can thus generate classical random graphs and scale-free networks with any average degree power-law exponent $\gamma > 2$ and clustering. In Fig. 2, left, we see that the

curvature and temperature of the internet are approximately $K = -0.83$ and $T = 0.6 \pm 0.1$.

Equation (2) establishes a formal equivalence between the $\mathbb{S}^1$ and $\mathbb{H}^2$ models we introduced in [2] and here. The two models generate similar network topologies thanks to the similarity between the effective and true hyperbolic distances in Eqs. (3) and (4). However, if we are to study other geometric properties of these networks, such as their navigability [7], then it does matter a lot what distances, spherical d native to $\mathbb{S}^1$ or hyperbolic x native to $\mathbb{H}^2$, we use to navigate a network. The latter distances x are dominated by $r + r'$, minus some small θ-dependent corrections. This effect can be observed in Fig. 1, right, where we show some hyperbolic geodesics between nodes in a small modeled network. These geodesics follow closely the radial directions between the nodes and the origin, i.e., they follow the same pattern as the shortest paths in the embedded network. Spherical distances d are at the other extreme as their gradient lines lie in the orthogonal tangential directions.

To demonstrate how such differences in distance calculations affect the efficiency of transport processes on networks, we embed the real internet topology from [8] into $\mathbb{H}^2$ using maximum-likelihood techniques. Specifically, we first assign to nodes random angular coordinates, while their radial coordinates are fixed by Eq. (2). We then execute the Metropolis-Hastings algorithm [9] by moving random nodes to new locations with the same radial coordinate but with a randomly chosen new angular coordinate. We accept each move with probability $\min(1, \mathcal{L}_a/\mathcal{L}_b)$, where $\mathcal{L}_b$ and $\mathcal{L}_a$ are the likelihoods before and after the move, that the network is produced by our $\mathbb{H}^2$ model with parameters matching the internet in Fig. 2, left. Formally, $\mathcal{L} = \prod_{i<j} p(x_{ij})^{a_{ij}}[1-p(x_{ij})]^{1-a_{ij}}$, where $\{a_{ij}\}$ is the internet adjacency matrix and x_{ij} is the hyperbolic distance between nodes i and j.

After this process has converged, we perform greedy routing as in [7] in the resulting embedding. We randomly select a source and try to find a path to a random destination by selecting the next node on the path as the current node's neighbor closest to the destination in $\mathbb{H}^2$. This process can be unsuccessful as it can get stuck at intermediate nodes that have no neighbors closer to the destination than themselves. The percentage of successful greedy paths and their hop-length averaged over 10^5 random source-destination pairs are 94.5% and 3.95, respectively. (The average length of shortest paths is 3.46.) For comparison, the same numbers using the $\mathbb{S}^1$ distances are 75.9% and 4.29. The reason for the exceptionally high ratio of successful paths in the $\mathbb{H}^2$ case is that the shortest paths in the internet stay close to the hyperbolic geodesics followed by greedy navigation between the corresponding source and destination. In other words, the real internet topology is remarkably congruent with underlying hyperbolic geometry.

Even more striking in this regard is Fig. 2, right, where we show the empirical connection probability for the links vs their hyperbolic distances in the embedded internet, juxtaposed with the theoretical connection probability in our $\mathbb{H}^2$ model. The similarity between the two provides empirical evidence that our model reflects reality. If the real internet were not congruent with our hyperbolic model, then no

maximum-likelihood technique used for its embedding would be able to make it such.

In summary we have shown that complex network topologies are congruent with hyperbolic geometries. We can start with hyperbolic geometry and show that it naturally gives rise to scale-free topology or we can start with the latter and show that hyperbolic geometry is its effective geometry. In this geometric approach, clustering and heterogeneous degree distributions appear as simple consequences of the metric and negative-curvature properties of hyperbolic spaces. In our hyperbolic model, the space curvature controls the heterogeneity of the degree distribution, while clustering is a function of temperature. Fermi-Dirac statistics provides a physical interpretation of hidden distances as energies of the corresponding link fermions. This analogy may contribute to applications of the standard tools of statistical mechanics to the analysis of complex networks [5], which can be informally thought of as negatively curved containers of ultracold fermions. The internet embedding, besides providing empirical evidence that our model is realistic, shows that the efficiency of transport processes without global knowledge is maximized if such processes use (effective) hyperbolic distances. If networks evolve to be efficient with respect to their functions—and transport is one of such functions—then this finding further supports our hyperbolic metric space approach.

The internet embedding may also prove practically useful since routing in it is extremely efficient and requires only local information about hyperbolic coordinates of node neighbors. Global knowledge of the large-scale internet topology is a major scalability bottleneck in internet routing today [10]. Another potential application of our work is protein folding, where hidden spaces are protein conformation energy profiles, and the protein folding process is greedy routing toward the minimum-energy conformation [11]. Yet another class of applications involves cases where to have a right model for similarity distances is a key, such as recommender systems used by companies such as Amazon or Netflix. Their efficiency depends on how accurately the similarities between consumers are estimated. Our hyperbolic explanation of the structure of complex networks is by no means the only possible mechanism capable of generating scale-free topologies with strong clustering. Therefore, the question of special interest is whether our explanation is (implicitly) equivalent to existing models, among which preferential attachment [12] appears to be most popular?

[1] A. Clauset, C. R. Shalizi, and M. E. J. Newman, SIAM Rev. (to be published).

[2] M. Á. Serrano, D. Krioukov, and M. Boguñá, Phys. Rev. Lett. **100**, 078701 (2008).

[3] D. J. Watts, P. S. Dodds, and M. E. J. Newman, Science **296**, 1302 (2002); A. Clauset, C. Moore, and M. E. J. Newman, Nature (London) **453**, 98 (2008).

[4] M. Gromov, *Metric Structures for Riemannian and Non-Riemannian Spaces* (Birkhäuser, Boston, 2007).

[5] J. Park and M. E. J. Newman, Phys. Rev. E **70**, 066117 (2004); S. N. Dorogovtsev, J. F. F. Mendes, and A. N. Samukhin, Nucl. Phys. B **666**, 396 (2003).

[6] P. Mahadevan, D. Krioukov, M. Fomenkov, B. Huffaker, X. Dimitropoulos, K. C. Claffy, and A. Vahdat, Comput. Commun. Rev. **36**, 17 (2006).

[7] M. Boguñá, D. Krioukov, and K. C. Claffy, Nature Physics **5**, 74 (2009).

[8] Y. Shavitt and E. Shir, Comput. Commun. Rev. **35**, 71 (2005).

[9] M. E. J. Newman and G. T. Barkema, *Monte Carlo Methods in Statistical Physics* (Clarendon Press, Oxford, 1999).

[10] *RFC4984*, edited by D. Meyer, L. Zhang, and K. Fall (The Internet Architecture Board, Amsterdam, 2007).

[11] E. Ravasz, S. Gnanakaran, and Z. Toroczkai, e-print arXiv:0705.0912.

[12] A.-L. Barabási and R. Albert, Science **286**, 509 (1999).