On 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, CA 92093, USA*

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

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

We establish a connection between observed scale-free topologies and hidden hyperbolic geometries of complex networks. Hidden geometries are coarse metric abstractions of the approximately hierarchical community structure of complex networks, used to estimate node similarities. Space expands exponentially in hyperbolic geometry, and scale-free topologies emerge as a consequence of this exponential expansion. Fermi-Dirac statistics connects observed topology to hidden geometry: observed edges are fermions, hidden distances are their energies; the curvature of the hidden space affects the heterogeneity of the degree distribution, while clustering is a function of temperature. Understanding the connection between topology and geometry of complex networks contributes to studying the efficiency of their functions, and may find practical applications in many disciplines, ranging from Internet routing to brain, cell signaling, or protein folding research.

PACS numbers: 89.75.Fb, 89.75.Hc, 02.40.-k, 67.85.Lm

The topology of many complex networks is scale-free. The distribution $P(k)$ of node degrees k in them often follows power laws $P(k) \sim k^{-\gamma}$ with $\gamma \in [2, 3]$ [1, 2, 3, 4]. These networks also exhibit strong clustering, i.e., high concentration of triangular subgraphs [1, 2, 3, 4]. Our previous work [5] demonstrated that the clustering peculiarities of complex networks, and in particular their self-similarity, finds a natural explanation in the existence of hidden metric spaces underlying the network and abstracting the intrinsic similarities between its nodes. Here we show that the first property—the scale-free topology of complex networks—appears 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. For example, in the hyperbolic plane, which is the two-dimensional hyperbolic space of constant curvature -1 (Fig. 1), the length of a circle and the area of a disc of radius R are $2\pi \sinh R$ and $2\pi(\cosh R - 1)$ [6], both growing as $\sim e^R$. From a purely metric perspective, the hyperbolic plane is 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.

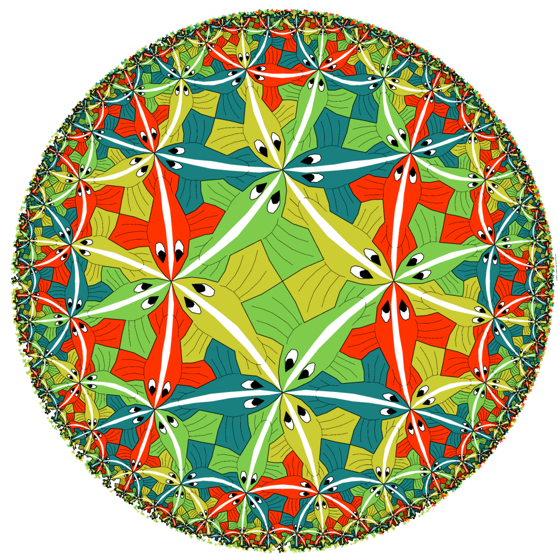


FIG. 1: Artistic visualization of the Poincaré disc model of the hyperbolic plane by Silvio Levy, based on M. C. Escher's *Circle Limit III*, with the permission from the Geometry Center, University of Minnesota. The exponential expansion of fish illustrates the exponential expansion of hyperbolic space. 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.

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. The 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, etc. The relationships between such groups and subgroups, called communities [7], can be approximated by tree-like structures, in which the distance between two nodes estimates how similar they are [8, 9]. Importantly, the node classification hierarchy need not be strictly a tree. Approximate "tree-ness," which can be formally expressed solely in terms of the metric structure of a space [10], makes the space hyper-

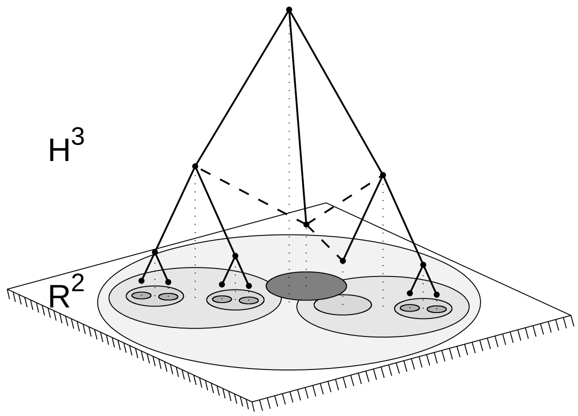


FIG. 2: Mapping between discs in the Euclidean space $\mathbb{R}^2$ and points in the hyperbolic space $\mathbb{H}^3$. The x, y-coordinates of the disc centers in $\mathbb{R}^2$ are the x, y-coordinates of the corresponding points in $\mathbb{H}^3$. The z-coordinates of these points are the radii of the corresponding discs. This mapping represents the tree-like hierarchy among the discs. Two points in $\mathbb{H}^3$ are connected by a solid link if one of the corresponding discs is the minimum-size disc that fully contains the other disc. This hierarchy is not perfect, and the tree structure is approximate. The darkest disc in the middle partially overlaps with three other discs at different levels of the hierarchy. Two points in $\mathbb{H}^3$ are connected by a dashed link if the corresponding discs partially overlap. These links add cycles to the tree.

bolic.

Fig. 2 illustrates a very general mechanism explaining why a hyperbolic, tree-like geometry naturally characterizes the community-based node similarity spaces underlying complex networks. In this illustration, communities, i.e., abstract sets of nodes, are represented by the Euclidean discs in $\mathbb{R}^2$. Each disc in $\mathbb{R}^2$ is mapped to a point in the Poincaré half-space model of the 3-dimensional hyperbolic space $\mathbb{H}^3$. Colloquially, two discs are similar if their overlap is approximately equal to each disc, i.e., if their radii are similar and centers are close in $\mathbb{R}^2$. But the shown mapping has the property that if two discs in $\mathbb{R}^2$ are similar, then the two points representing them in $\mathbb{H}^3$ are hyperbolically close, and *vice versa*. Formally, if the ratio of the discs' radii r, r' is bounded by a constant C, $1/C \leqslant r/r' \leqslant C$, and the Euclidean distance between their centers is bounded by Cr, then one can show [10] that the hyperbolic distance between the corresponding points in $\mathbb{H}^3$ is bounded by some constant C', which depends only on C, and not on the disc radii or center locations. The converse is also true. Therefore, similarity distances between sets and hyperbolic distances between their one-point representations are congruent measures.

We now put these intuitive considerations to qualitative grounds. We want to see what network topologies emerge in the simplest possible settings involving hidden hyperbolic metric spaces. Specifically, let us form a network of $N \gg 1$ nodes located in the simplest hyperbolic space of curvature -1, i.e., the hyperbolic plane. Since the number of nodes is finite, the area that nodes occupy is bounded. Let it be a disc of radius $R \gg 1$. The simplest node distribution within the disc is uniform, meaning that the node density $\rho(r)$ at distance r from the disc center is

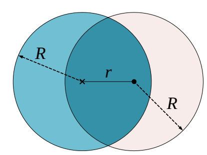
$$\rho(r) = \frac{\sinh r}{\cosh R - 1} \approx e^{r-R} \sim e^{r}. \qquad (1)$$

Next, we have to specify the connection probability $p(x)$, which is the probability that two nodes at hyperbolic distance x are connected. The only requirement to this function is that it must be integrable [11]. We first consider the simplest case, the step function

$$p(x) = \Theta(R - x), \qquad (2)$$

and justify this choice later. This $p(x)$ connects each pair of nodes if the distance between them is not larger than R.

At this point we have a network formed, and we can compute the average degree $k(r)$ of nodes at distance r from the disc center. Such nodes are connected to all nodes in the intersection area of the two discs of the same radius R, one in which all nodes reside, and the other centered at distance r from the center of the first disc:



Since the node distribution is uniform, $k(r)$ is proportional to the area of this intersection. In Euclidean geometry this area is given by a trivial expression. In hyperbolic geometry the analogous expression is far from trivial. We have computed it, it matches perfectly the simulations, but it is rather long, so that we omit it here for brevity. What matters is that $k(r)$ decreases exponentially, $k(r) \sim e^{-r/2}$. Therefore, the inverse function is logarithmic, $r(k) \sim -2 \ln k$, and the node degree distribution in the network is approximately a power law,

$$P(k) \approx \rho[r(k)]\,|r'(k)| \sim k^{-3}. \qquad (3)$$

We can generalize the node density in Eq. (1):

$$\rho(r) \approx \alpha e^{\alpha(r-R)} \sim e^{\alpha r}, \quad \alpha > 0. \qquad (4)$$

In this case we cannot compute $k(r)$ exactly, but the approximate expression reads

$$k(r) \approx N\left\{\xi e^{-\frac{1}{2}r} + (1-\xi)\,e^{-\alpha r}\right\},\ \xi = \frac{2}{\pi}\frac{\alpha}{\alpha - \frac{1}{2}}. \qquad (5)$$

The limit $\alpha \to 1/2$ is well defined, $k(r) \to N\left(1 + \frac{r}{\pi}\right)e^{-\frac{1}{2}r}$, and we see that $k(r) \sim e^{-\frac{1}{2}r}$ if $\alpha \geqslant 1/2$,

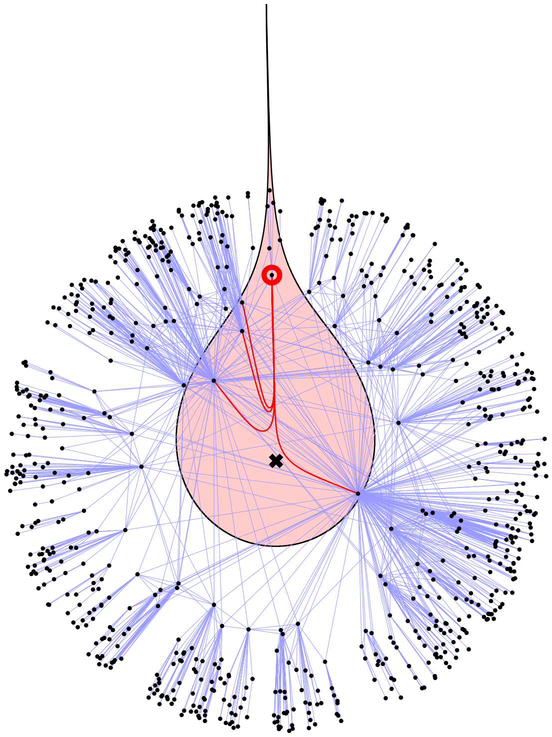


FIG. 3: Visualization of a modeled network with $N = 740$ nodes, power-law exponent $\gamma = 2.2$, and average degree $\bar{k} = 4.98$ embedded in the hyperbolic disc of radius $R = 15.47$. The Euclidean distance between a node and the origin at the disc center, shown as the cross, represents the true hyperbolic distance between the two. The Euclidean distance between any two nodes is *not* equal to the hyperbolic distance between them, as indicated by the shape of the shaded hyperbolic disc centered at the circled node located at distance $r = 10.60$ from the origin. The hyperbolic radius of this disc is also R, and according to the model, the circled node is connected to all the nodes lying in this disc. The curves show the hyperbolically straight lines, i.e., geodesics, connecting the circled node and the nodes in its disc that are closer to the origin.

and $k(r) \sim e^{-\alpha r}$ otherwise. Therefore the degree distribution in the network is

$$P(k) \sim k^{-\gamma}, \quad \text{with } \gamma = \begin{cases} 2\alpha + 1 & \text{if } \alpha \geqslant \frac{1}{2}, \\ 2 & \text{if } \alpha \leqslant \frac{1}{2}. \end{cases} \qquad (6)$$

Given Eq. (5), it is easy to see that selecting R according to $N = c\,e^{R/2}$, where c is a constant, fixes the average degree in the network. Fig. 3 visualizes one small network generated by this model. This network looks conceptually similar to well-known visualizations of real scale-free networks [12, 13].

We now pause and approach the problem from a different angle. Suppose we formally want to generate scale-free networks by assigning to N nodes two hidden variables (r, θ), with r distributed exponentially on $[0, R]$ as in Eq. (4), and θ being uniform on $[0, 2\pi]$. We want the expected degree κ of a node to depend only on r. We then see that to produce a network with the expected degree distribution $\rho(\kappa) = \kappa_0^{\gamma-1}(\gamma-1)\kappa^{-\gamma}$, we must have

$$\kappa = \kappa_0 e^{\frac{\zeta}{2}(R-r)}, \quad \frac{\zeta}{2} = \frac{\alpha}{\gamma - 1}, \quad N = c\,e^{\frac{\zeta}{2}R}, \qquad (7)$$

where ζ and c are some constants, and κ_0 is the minimum expected degree. This change of variables changes our perception of a node. Its geometric attribute r, radial coordinate, becomes its topological attribute κ, expected degree. In the κ-variables, nodes do not have any radial coordinates, they are effectively located on a Euclidean circle $\mathbb{S}^1$ of some radius, which can be set without loss of generality to $N/(2\pi)$, so that the node density on the circle is fixed to 1 [5]. Measured over this circle, the distance d between two nodes with expected degrees κ and κ' is proportional to the difference of their angular coordinates $\Delta\theta$, $d = N\Delta\theta/(2\pi)$. As shown in [5], the connection probability can be any integrable function of $d/(\mu\kappa\kappa')$, where μ is a constant that depends on the average degree. Eq. (7) translates this observation to the r-variables to yield that the connection probability can be any integrable function of $x - R$, where the new effective distance x is

$$x = r + r' + \frac{2}{\zeta} \ln \frac{\Delta\theta}{2}. \qquad (8)$$

The hyperbolic distance 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 r, r', and $\Delta\theta$ is closely approximated by

$$x = r + r' + \frac{2}{\zeta} \ln \sin \frac{\Delta\theta}{2}. \qquad (9)$$

Therefore, parameter ζ in Eq. (7) is the square root of curvature $-K$ of the hyperbolic space. The subtle difference between the truly hyperbolic distance in $\mathbb{H}^2$ with $K = -\zeta^2$ in Eq. (9), and the effective distance in $\mathbb{S}^1$ in Eq. (8) has virtually no effect on any topological property of generated networks, and it justifies *a posteriori* the choice of the connection probability as a function of $x - R$ in Eq. (2).

We thus have a different view on hyperbolic geometry. We can start with a scale-free network embedded in an asymptotically flat Euclidean space, and then naturally redefine distances in this space, Eqs. (7,8), to account for the topological, degree-induced, hierarchy among nodes. The result of this redefinition is an effective hyperbolic geometry, virtually identical, Eqs. (8,9), to the true hyperbolic geometry representing the hidden, similarity-based hierarchy. Is this equivalence “coincidental”?

To answer this question, we consider the Fermi connection probability

$$p(x) = \frac{1}{1 + e^{\frac{\zeta}{2T}(x-R)}} = \frac{1}{1 + \left(\frac{d}{\mu\kappa\kappa'}\right)^{\frac{1}{T}}}, \qquad (10)$$

where T is the temperature and $\mu = c/(\pi\kappa_0^2)$. One motivation for this connection probability is that it generates graphs belonging to the ensemble of exponential random graphs [14]. From a physical perspective, graph edges are non-interacting fermions with energies equal to their hidden hyperbolic distances, and R is the chemical potential in the grand canonical sense, i.e., it is defined by the condition that $\bar{k}N/2$, the number of edges/fermions, is fixed on average. At $T \to 0$ Eq. (10) converges to the step function in Eq. (2), and the network is in the strongly degenerate ground state of the system. 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 network losing its cold-state metric structure. Clustering remains zero for all $T > 1$.

Fermi-Dirac statistics thus provides a physical interpretation of the "coincidence" between the true hyperbolic geometry induced by hidden similarities, and the effective one, due to observable node degrees. We can freely switch between the two views on the hierarchical nature of complex networks using Eqs. (7,10). These equations also establish a formal equivalence between the $\mathbb{S}^1$ and $\mathbb{H}^2$ models we introduced in [5] and here. The two models are congruent in terms of the topology of networks that they produce, but if we are to study other, geometric properties of these networks, such as their navigability [15], 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. 3, where some hyperbolic geodesics are shown. They follow closely the radial directions between the nodes and the origin. Spherical distances d are the other extreme, as their gradient lines lie in the orthogonal tangential directions.

Our model has three parameters. The first, temperature T, controls clustering at $T < 1$. In this cold regime, the exponent of the degree distribution γ depends only on the second parameter, ratio α/ζ: $\gamma = 2\alpha/\zeta + 1$ if $\alpha/\zeta \geqslant 1/2$, and $\gamma = 2$ otherwise. Nodes are distributed uniformly in the hyperbolic space if $\alpha = \zeta$, in which case $\gamma = 3$. We can think of α as the logarithm of the average branching factor in the underlying hierarchy. Only in relation to the square root $\zeta = \sqrt{-K}$ of the hidden space curvature K does this branching factor affect the observed network topology. We can thus set $\alpha = 1/2$ without loss of generality, so that $\gamma = 1/\zeta + 1$ is fully defined by curvature $K > -1$. At $T > 1$, $\int dx/(1 + x^{1/T})$ diverges, and the chemical potential is no longer given by Eq. (7) but by $N = c\, e^{\frac{\zeta}{2T}R}$. In this hot regime, clustering is always zero, 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 regardless the hidden distance between them, and the network loses its cold-state hierarchical structure. Combining the cold and hot regimes,

$$\gamma = \begin{cases} 1/\zeta + 1 & \text{if } T < 1 \text{ and } \zeta < 1, \\ T/\zeta + 1 & \text{if } T > 1 \text{ and } \zeta < T, \\ 2 & \text{otherwise.} \end{cases} \tag{11}$$

The last parameter, c in $\mathbb{H}^2$ or $\mu = c/(\pi\kappa_0^2)$ in $\mathbb{S}^1$, fixes the average degree in the network:

$$c \approx \begin{cases} \bar{k}\,\frac{\sin \pi T}{2T}\left(1-\zeta\right)^2 \approx \kappa_0^2 \frac{\sin \pi T}{2\bar{k}T} & \text{if } T < 1, \\ \bar{k}\left(\frac{\pi}{2}\right)^{\frac{1}{T}} \frac{T-1}{T^3}\left(T-\zeta\right)^2 & \text{if } T > 1. \end{cases} \tag{12}$$

At $T \to \infty$, $c = \bar{k}$. With these parameters, the model can generate classical random graphs, and scale-free networks with any average degree, power-law exponent $\gamma > 2$, and clustering. In Fig. 4 we see that the curvature and temperature of the Internet are approximately $K = -0.83$ and $T = 0.6 \pm 0.1$.

In summary, we have shown that hyperbolic geometry naturally abstracts the two types of hierarchy in complex networks. The first hierarchy reflects a similarity-based community structure; the second is induced by node degrees. Scale-free degree distributions appear as a consequence of the exponential expansion of hyperbolic space, which is a metric space, its triangle inequality explaining strong clustering [5]. We have thus established a connection between the main properties of hyperbolic geometry and complex networks topology. We have also shown that Fermi-Dirac statistics establishes the congruency between the two hierarchies, hidden and observed. Observed edges are fermions, their energies are hidden

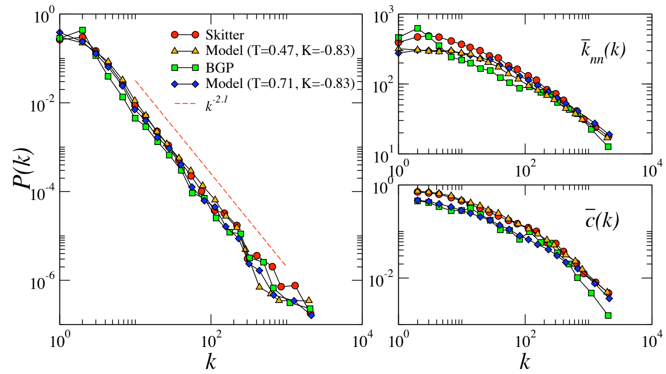


FIG. 4: Modeled networks vs. the Internet. The degree distribution $P(k)$, average nearest neighbors degree $\bar{k}_{nn}(k)$, and degree-dependent clustering coefficient $\bar{c}(k)$ are shown for the skitter (average degree $\bar{k} = 6.29$ and average clustering $\bar{C} = 0.46$) and Border Gateway Protocol (BGP) ($\bar{k} = 4.68$, $\bar{C} = 0.29$) views of the Internet from [16], 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 ($\bar{k} = 4.85$, $\bar{C} = 0.25$).

distances. The curvature of the hidden space controls the heterogeneity of the degree distribution, while clustering is a function of temperature. This analogy may lead to novel applications of the standard tools of statistical mechanics to the analysis of complex networks [14, 17], which can be informally thought of as negatively curved containers of ultracold fermions.

In this work we make a step forward specifying the common structure of hidden metric spaces underlying real networks. Discoveries of the network-specific structure of such spaces may find practical applications in many domains of science and engineering. Such potential applications include areas where the right estimate of node similarity is a key, e.g., recommender systems or other forms of data mining. Another class of applications is related to transport phenomena in networks, where hidden spaces may be used to propagate information, or other forms of media, towards specific destinations without global knowledge of network topology [15]. Examples include brain, cell signaling, protein folding processes, and our main interest, Internet routing. A question of special interest is whether the hyperbolic metric space explanation of the structure of complex networks is (implicitly) equivalent to existing models, among which preferential attachment [18] appears to be most popular?

We thank A. Goltsev, S. Dorogovtsev, A. Samukhin, R. Pastor-Satorras, A. Baronchelli, M. Newman, J. Kleinberg, Z. Toroczkai, F. Menczer, A. Clauset, D. Clark, K. Fall, kc claffy, and others for useful discussions and suggestions. This work was supported by FIS2007-66485-C02-02, Generalitat de Catalunya grant No. SGR00889, and by NSF CNS-0434996 and CNS-0722070, DHS N66001-08-C-2029, and Cisco Systems.

[1] R. Albert and A.-L. Barabási, Rev Mod Phys **74**, 47 (2002).
[2] S. N. Dorogovtsev and J. F. F. Mendes, *Evolution of Networks: From Biological Nets to the Internet and WWW* (Oxford University Press, Oxford, 2003).
[3] M. E. J. Newman, SIAM Rev **45**, 167 (2003).
[4] S. Boccaletti, V. Latora, Y. Moreno, M. Chavez, and D.-U. Hwanga, Phys Rep **424**, 175 (2006).
[5] M. Á. Serrano, D. Krioukov, and M. Boguñá, Phys Rev Lett **100**, 078701 (2008).
[6] J. W. Anderson, *Hyperbolic Geometry* (Springer-Verlag, London, 2005).
[7] M. Girvan and M. E. J. Newman, Proc Natl Acad Sci USA **99**, 7821 (2002).
[8] D. J. Watts, P. S. Dodds, and M. E. J. Newman, Science **296**, 1302 (2002).
[9] A. Clauset, C. Moore, and M. E. J. Newman, Nature **453**, 98 (2008).
[10] M. Gromov, *Metric Structures for Riemannian and Non-Riemannian Spaces* (Birkhäuser, Boston, 2007).
[11] M. Boguñá and R. Pastor-Satorras, Phys Rev E **68**, 036112 (2003).
[12] CAIDA, *Visualizing Internet topology at a macroscopic scale*, `http://www.caida.org/research/topology/as_core_network/`.
[13] J. I. Alvarez-Hamelin, L. Dall'Asta, A. Barrat, and A. Vespignani, in *Advances in Neural Information Processing Systems 18*, edited by Y. Weiss, B. Schölkopf, and J. Platt (MIT Press, Boston, 2006), pp. 41–50.
[14] J. Park and M. E. J. Newman, Phys Rev E **70**, 066117 (2004).
[15] M. Boguñá, D. Krioukov, and kc claffy, Nature Physics **5**, 74 (2009).
[16] P. Mahadevan, D. Krioukov, M. Fomenkov, B. Huffaker, X. Dimitropoulos, kc claffy, and A. Vahdat, Comput Commun Rev **36**, 17 (2006).
[17] S. N. Dorogovtsev, J. F. F. Mendes, and A. N. Samukhin, Nucl Phys B **666**, 396 (2003).
[18] A.-L. Barabási and R. Albert, Science **286**, 509 (1999).