Distance Distribution between Complex Network Nodes in Hyperbolic Space

Gregorio Alanis-Lobato*
Miguel A. Andrade-Navarro
Faculty of Biology, Johannes Gutenberg Universität
Institute of Molecular Biology, Ackermannweg 4, 55128 Mainz, Germany
**galanisl@uni-mainz.de*

In the emerging field of network science, a recent model proposes that a hyperbolic geometry underlies the network representation of complex systems, shaping their topology and being responsible for their signature features: scale invariance and strong clustering. Under this model of network formation, points representing system components are placed in a hyperbolic circle and connected if the distance between them is below a certain threshold. Then the aforementioned properties come out naturally, as a direct consequence of the geometric principles of the hyperbolic space containing the network. With the aim of providing insights into the stochastic processes behind the structure of complex networks constructed with this model, the probability density for the approximate hyperbolic distance between N points, distributed quasi-uniformly at random in a disk of radius $R \sim \ln N$, is determined in this paper, together with other density functions needed to derive this result.

1. Introduction

Representing the dynamic relationships between complex system components as networks of interacting nodes has found applications in biology [1], mathematics [2], technology [3], and even cosmology [4]. Despite the apparent differences between these fields, the networks arising from their systems share many structural properties: scale-free node degree distributions [5], self-similarity [6], and average graph geodesics that grow logarithmically with the number of nodes and strong clustering [7]. Moreover, these characteristics can also be present in geometric objects, like fractals or cellular automata [8–10], which has prompted the analysis of complex networks from a geometric perspective [6, 8, 9, 11, 12].

Of special interest is a recent model of network growth that advocates for the geometric principles of hyperbolic space being responsible for the emergence of the above-mentioned network signature features [13]. Thus, the formation of scale-free and strongly clustered networks is the result of an optimization process involving two variables: node popularity and similarity between nodes.

Popularity reflects the property of a node to attract connections from others over time, and it is thus associated with a node's seniority status in the system. On the other hand, nodes that are similar to each other have a high likelihood of getting connected, regardless of their rank. Dynamic generation of a network with this model can thus mimic the formation of, for example, social websites, where the concepts of (social) popularity and similarity can be intuitively understood, but also of other types of networks, such as protein interactomes. There, popularity would correspond to the tendency of a protein to establish more connections, which could correlate with a protein having appeared at an early stage in evolution (seniority), and similarity could be in protein sequence or function. This is a good example of how very simple rules—in this case, *link to the most popular and similar node*—result in complex behavior.

In the so-called $\mathbb{H}^2$ model, which can be seen as the static version of the popularity-similarity optimization discussed above [14, 15], nodes appear randomly in a hyperbolic disk with radius proportional to the number of system components. Node pairs are then linked if they are hyperbolically close. The choice of this space to place nodes is a convenient way to abstract the tradeoff between popularity and similarity via the hyperbolic distance between them: popularity is modeled by node radial coordinates and similarity by angular distances between nodes. In this way, and due to the principles of hyperbolic geometry, nodes that appear near the origin of the disk are more popular/senior and have a higher probability of connecting to other nodes in the system, thus becoming the hubs of the network. In contrast, nodes that appear on the periphery link only with nodes that are really close (similar) to them [16].

One of the advantages of this model is that with the appropriate choice of node density and hyperbolic disk radius, we can grow networks with target node degree scaling exponent γ and average node degree and clustering coefficient, which can serve as null models for testing hypotheses related to the formation of complex systems. As a result, the goal of this work is to determine an analytical expression for the probability distribution of hyperbolic distances between nodes, randomly placed in a hyperbolic disk. The resulting distribution, which is based on an approximation of the hyperbolic distance between nodes, provides insights into the stochastic processes that give rise to the structure of complex networks and serves as a reference for the type of node similarities needed to generate their scale-free and strongly clustered topologies.

2. Preliminaries

In this paper, the native representation of the hyperbolic space is used, in which the two-dimensional hyperbolic space $\mathbb{H}^2$, of constant curvature $K = -\zeta^2$, is contained in a Euclidean disk, and points are placed at polar coordinates (r, θ). The results shown in Section 3 focus on the case $\zeta = 1$. Even when the choice of different ζ changes the radius of the hyperbolic disk and scales distances between points inside it by a factor of $1/\zeta$, the network resulting from connecting them remains the same [16].

Note that while hyperbolic angles and distances from the origin of the disk (i.e., radial coordinates) are equivalent to Euclidean angles and distances from the origin, the length and area of a hyperbolic circle of radius r, $L(r) = 2\pi\sinh(r)$ and $A(r) = 2\pi(\cosh(r) - 1)$, respectively, expand exponentially with r and not polynomially like in the Euclidean scenario. Consequently, to distribute N points uniformly at random in a hyperbolic circle of radius R, angular coordinates $\theta \in [0, 2\pi]$ are sampled with the uniform density $\rho(\theta) = 1/(2\pi)$, and radial coordinates $r \in [0, R]$ are sampled with the exponential density $\rho(r) = \sinh(r)/(\cosh(R-1)) \approx e^{r-R}$ (see Figure 1(a)). Note that if a parameter $\alpha \in (0, 1]$ is introduced in this expression, we can control the density of points close to the origin of the hyperbolic circle (see equation (1) and Figure 1(b)). This has a direct impact on the node degree distribution of the network resulting from this modified node density, because, as mentioned in Section 1, nodes with small radial coordinates have a higher likelihood to be network hubs due to their proximity to all other nodes in the space:

$$\rho(r) \approx \alpha e^{\alpha(r-R)}. \tag{1}$$

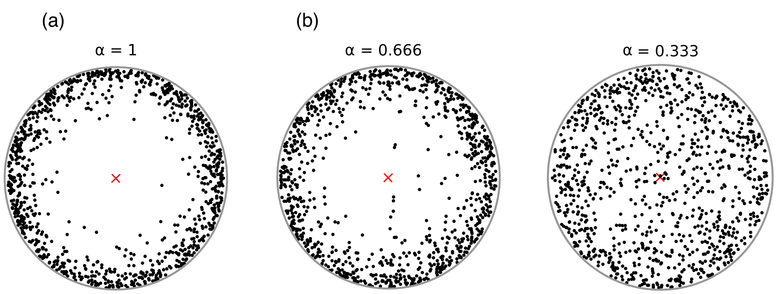

Figure 1. (a) Uniform distribution of points in the hyperbolic circle, which corresponds to setting $\alpha = 1$ in equation (1). (b) Quasi-uniform distribution of points in the hyperbolic circle for two different values of α. Disk centers are marked with a red cross and circle boundaries are depicted in gray.

Equation (1), which corresponds to a quasi-uniform distribution of points in the hyperbolic circle of radius R, is the one used in the rest of this paper.

To compute the distance between any two points (r_i, θ_i) and (r_j, θ_j) in the hyperbolic disk of constant curvature -1, we can resort to the hyperbolic law of cosines (see equation (2) and Figure 2):

$$x_{i,j} = \operatorname{arcosh}\left(\cosh(r_i)\cosh(r_j) - \sinh(r_i)\sinh(r_j)\cos(\theta_{i,j})\right). \tag{2}$$

$\theta_{i,j} = \pi - \left|\pi - \left|\theta_i - \theta_j\right|\right|$ in equation (2) is the angle between points (i.e., their angular distance).

For sufficiently large r_i and r_j, equation (2) can be closely approximated by equation (3), which is the expression for the hyperbolic distance between points that is used in the rest of this paper:

$$x_{i,j} \approx r_i + r_j + 2\ln\left(\frac{\theta_{i,j}}{2}\right). \tag{3}$$

Note that since $x_{i,j} \geq 0$, $\theta_{i,j} \geq 2\sqrt{e^{-r_i - r_j}}$.

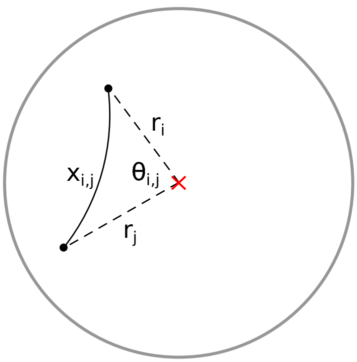

Figure 2. The hyperbolic distance $x_{i,j}$ between two points in the hyperbolic disk can be computed with the help of the hyperbolic law of cosines (see equation (2)) and closely approximated by equation (3).

Proof. Let us write equation (2) as:

$$\cosh(x_{i,j}) = \cosh(r_i)\cosh(r_j)\left(1 - \tanh(r_i)\tanh(r_j)\cos(\theta_{i,j})\right).$$

Since $\lim_{z\to\infty} \tanh z = 1$ and r_i and r_j are large, we have:

$$\cosh(x_{i,j}) \approx \cosh(r_i)\cosh(r_j)\left(1 - \cos(\theta_{i,j})\right) = \cosh(r_i)\cosh(r_j)\left(2\sin^2\frac{\theta_{i,j}}{2}\right).$$

Using the definitions of $\cosh z = \left(e^{2z} + 1\right) / 2e^{z} = (e^{z} + e^{-z}) / 2$:

$$\frac{e^{2x_{i,j}} + 1}{2e^{x_{i,j}}} \approx \left(\frac{e^{r_i} + e^{-r_i}}{2}\right)\left(\frac{e^{r_j} + e^{-r_j}}{2}\right)\left(2\sin^2\frac{\theta_{i,j}}{2}\right).$$

Since r_i and r_j are large, e^{-r_i} and e^{-r_j} are very close to 0, which gives:

$$e^{2x_{i,j}} - \left(e^{r_i+r_j}\right)\left(\sin^2\frac{\theta_{i,j}}{2}\right)e^{x_{i,j}} + 1 \approx 0.$$

The above can be solved for $e^{x_{i,j}}$ as a quadratic equation, yielding:

$$e^{x_{i,j}} \approx \frac{1}{2}\left(e^{r_i+r_j}\right)\left(\sin^2\frac{\theta_{i,j}}{2}\right)\left[1 \pm \sqrt{1 - 4\left(e^{-2r_i-2r_j}\right)\sin^{-4}\frac{\theta_{i,j}}{2}}\right].$$

Once again, since r_i and r_j are large, $e^{-2r_i-2r_j} \to 0$, making the negative term inside the squared root close to 0. As a result, considering the positive root only, the term in brackets is approximately 2, yielding:

$$e^{x_{i,j}} \approx \left(e^{r_i+r_j}\right)\left(\sin^2\frac{\theta_{i,j}}{2}\right).$$

Applying natural logarithms to both sides, the expression in equation (3) is reached:

$$x_{i,j} \approx r_i + r_j + 2\ln\left(\sin\frac{\theta_{i,j}}{2}\right) \approx r_i + r_j + 2\ln\left(\frac{\theta_{i,j}}{2}\right).$$

This completes the proof. □

In Section 3, the random variable $X_{i,j}$, the distance between two random points in a hyperbolic circle, is considered, and its probability density function (PDF) $f_{X_{i,j}}\left(x_{i,j}\right) = \rho\left(x_{i,j}\right)$ is determined. This is achieved with the help of two facts: (*i*) the PDF of a random variable is the derivative of its cumulative distribution function (CDF); and (*ii*) the PDF of the sum of two random variables is the convolution of their individual PDFs. In the following, random variables Z_q are denoted with capital letter and subscripted with the part q of equation (3) being analyzed. $F_{Z_q}\left(z_q\right)$ is the CDF of Z_q and $f_{Z_q}\left(z_q\right) = \rho(q)$ is its PDF.

3. Determination of the Probability Distribution

Let us start by determining the PDF of the rightmost part of equation (3). To do so, we first need to determine the PDF of $\theta_{i,j} = \pi - \left|\pi - |\theta_i - \theta_j|\right|$ and, at the same time, the PDF of $\theta_i - \theta_j$. It is important to emphasize that the latter expression is different from $\theta_{i,j}$. While $\theta_{i,j}$ is the angular distance between points and is always ≥ 0, $\theta_i - \theta_j$ can be negative if $\theta_j > \theta_i$. The PDF of random variable $Z_{\theta_i - \theta_j}$, the difference of two random angles in the hyperbolic circle, corresponds to the following convolution (or cross-correlation in signal-processing terminology):

$$f_{Z_{\theta_i-\theta_j}}\left(z_{\theta_i-\theta_j}\right) = \int f_{Z_{\theta_i}}\left(z_{\theta_i-\theta_j} + \theta_j\right) f_{Z_{\theta_j}}\left(\theta_j\right) d\theta_j.$$

Due to the domain of θ_j, the integration limits of this cross-correlation must ensure that $0 \leq z_{\theta_i-\theta_j} + \theta_j \leq 2\pi$. This breaks the integral in two pieces:

$$f_{Z_{\theta_i-\theta_j}}\left(z_{\theta_i-\theta_j}\right) = \begin{cases} \displaystyle\int_0^{2\pi - z_{\theta_i-\theta_j}} f_{Z_{\theta_i}}\left(z_{\theta_i-\theta_j} + \theta_j\right) f_{Z_{\theta_j}}\left(\theta_j\right) d\theta_j, & \theta_j \leq 2\pi - z_{\theta_i-\theta_j} \\ \displaystyle\int_{-z_{\theta_i-\theta_j}}^{2\pi} f_{Z_{\theta_i}}\left(z_{\theta_i-\theta_j} + \theta_j\right) f_{Z_{\theta_j}}\left(\theta_j\right) d\theta_j, & \theta_j \geq -z_{\theta_i-\theta_j}. \end{cases}$$

The solution of the integrals yields (see Figure 3(a)):

$$f_{Z_{\theta_i-\theta_j}}\left(z_{\theta_i-\theta_j}\right) = \begin{cases} \dfrac{1}{2\pi} + \dfrac{z_{\theta_i-\theta_j}}{4\pi^2}, & -2\pi \leq z_{\theta_i-\theta_j} \leq 0 \\ \dfrac{1}{2\pi} - \dfrac{z_{\theta_i-\theta_j}}{4\pi^2}, & 0 < z_{\theta_i-\theta_j} \leq 2\pi. \end{cases} \tag{4}$$

As a result, the PDF for random variable $Z_{|\theta_i-\theta_j|}$ should be:

$$f_{Z_{|\theta_i-\theta_j|}}\left(z_{|\theta_i-\theta_j|}\right) = 2 f_{Z_{\theta_i-\theta_j}}\left(z_{\theta_i-\theta_j}\right) = \frac{1}{\pi} - \frac{z_{|\theta_i-\theta_j|}}{2\pi^2}, \quad 0 \leq z_{|\theta_i-\theta_j|} \leq 2\pi.$$

To determine the PDF of $Z_{\pi-|\theta_i-\theta_j|} = \pi - Z_{|\theta_i-\theta_j|}$, let us use its CDF:

$$\begin{aligned} &F_{Z_{\pi-|\theta_i-\theta_j|}}\left(z_{\pi-|\theta_i-\theta_j|}\right) = P\left(\pi - Z_{|\theta_i-\theta_j|} \leq z_{\pi-|\theta_i-\theta_j|}\right) = \\ &\quad 1 - P\left(Z_{|\theta_i-\theta_j|} \leq \pi - z_{\pi-|\theta_i-\theta_j|}\right) = 1 - F_{Z_{|\theta_i-\theta_j|}}\left(\pi - z_{\pi-|\theta_i-\theta_j|}\right) = \\ &\quad 1 - \int_0^{\pi - z_{\pi-|\theta_i-\theta_j|}} \left(\frac{1}{\pi} - \frac{z_{|\theta_i-\theta_j|}}{2\pi^2}\right) dz_{|\theta_i-\theta_j|} = \frac{1}{2}\left(\frac{z_{\pi-|\theta_i-\theta_j|}^2}{2\pi^2} + \frac{z_{\pi-|\theta_i-\theta_j|}}{\pi} + \frac{1}{2}\right). \end{aligned}$$

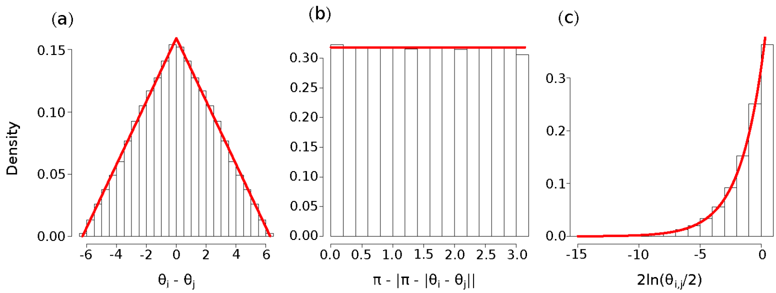

Figure 3. The shape of the distributions derived for the rightmost term in equation (3): (a) the difference between the angles of two points in the hyperbolic circle; (b) the angular distance between these points; and (c) the $\theta_{i,j}$-dependent correction applied to the sum of radial coordinates to approximate the hyperbolic distance between these points. The red line corresponds to the analytical expressions derived in this work (equations (4) through (6)), and the histograms correspond to the actual distributions of pairwise differences, distances, and corrections between 1000 points placed at random according to $\alpha = 1$.

Note that the last expression is valid for $-\pi \leq z_{\pi-|\theta_i-\theta_j|} \leq \pi$, and that differentiation with respect to $z_{\pi-|\theta_i-\theta_j|}$ finally gives:

$$f_{Z_{\pi-|\theta_i-\theta_j|}}\left(z_{\pi-|\theta_i-\theta_j|}\right) = \frac{1}{2\pi}\left(1 + \frac{z_{\pi-|\theta_i-\theta_j|}}{\pi}\right), \quad -\pi \leq z_{\pi-|\theta_i-\theta_j|} \leq \pi.$$

Now, the PDF for random variable $Z_{|\pi-|\theta_i-\theta_j||}$ is simply the positive part of $f_{Z_{\pi-|\theta_i-\theta_j|}}\left(z_{\pi-|\theta_i-\theta_j|}\right)$ plus its negative part, which yields:

$$f_{Z_{|\pi-|\theta_i-\theta_j||}}\left(z_{|\pi-|\theta_i-\theta_j||}\right) = f_{Z_{\pi-|\theta_i-\theta_j|}}\left(z_{|\pi-|\theta_i-\theta_j||}\right) + f_{Z_{\pi-|\theta_i-\theta_j|}}\left(-z_{|\pi-|\theta_i-\theta_j||}\right) = \frac{1}{\pi},$$
$$0 \leq z_{|\pi-|\theta_i-\theta_j||} \leq \pi.$$

Let $\Theta_{i,j} = \pi - Z_{|\pi-|\theta_i-\theta_j||}$ be the random variable for the angle between two random points in the hyperbolic circle (i.e., $\theta_{i,j} = \pi - \left|\pi - |\theta_i - \theta_j|\right|$). Using the CDF of $\Theta_{i,j}$, we have:

$$F_{\Theta_{i,j}}\left(\theta_{i,j}\right) = P\left(\pi - Z_{|\pi-|\theta_i-\theta_j||} \leq \theta_{i,j}\right) = 1 - P\left(Z_{|\pi-|\theta_i-\theta_j||} \leq \pi - \theta_{i,j}\right) =$$
$$1 - F_{Z_{|\pi-|\theta_i-\theta_j||}}\left(\pi - \theta_{i,j}\right) = 1 - \int_0^{\pi-\theta_{i,j}} \frac{1}{\pi} dz_{|\pi-|\theta_i-\theta_j||} = \frac{\theta_{i,j}}{\pi}.$$

Differentiating with respect to $\theta_{i,j}$, the resulting density is (see Figure 3(b)):

$$f_{\Theta_{i,j}}\left(\theta_{i,j}\right) = \frac{1}{\pi}, \quad 0 \leq \theta_{i,j} \leq \pi. \tag{5}$$

Thanks to equation (5), it is now possible to work out the PDF of term $2\ln\left(\theta_{i,j} / 2\right)$ in equation (3). Again, using the CDF method it is straightforward to determine the PDF for random variable $Z_{\theta_{ij}/2}$:

$$f_{Z_{\theta_{ij}/2}}\left(z_{\theta_{ij}/2}\right) = \frac{2}{\pi}, \quad 0 \leq z_{\theta_{ij}/2} \leq \pi.$$

With the help of the CDF method, the PDF of random variable $Z_{\ln(\theta_{ij}/2)} = \ln\left(Z_{\theta_{ij}/2}\right)$ is derived as follows:

$$F_{Z_{\ln(\theta_{ij}/2)}}\left(z_{\ln(\theta_{ij}/2)}\right) = P\left(\ln\left(Z_{\theta_{ij}/2}\right) \leq z_{\ln(\theta_{ij}/2)}\right) = P\left(Z_{\theta_{ij}/2} \leq e^{z_{\ln(\theta_{ij}/2)}}\right) =$$
$$F_{Z_{\theta_{ij}/2}}\left(e^{z_{\ln(\theta_{ij}/2)}}\right) = \int_0^{e^{z_{\ln(\theta_{ij}/2)}}} \frac{2}{\pi} dz_{\theta_{ij}/2} = \frac{2e^{z_{\ln(\theta_{ij}/2)}}}{\pi}.$$

Differentiating with respect to $z_{\ln(\theta_{ij}/2)}$:

$$f_{Z_{\ln(\theta_{ij}/2)}}\left(z_{\ln(\theta_{ij}/2)}\right) = \frac{2e^{z_{\ln(\theta_{ij}/2)}}}{\pi}, \quad z_{\ln(\theta_{ij}/2)} < \ln\left(\frac{\pi}{2}\right).$$

Finally, it is not difficult to see that for the random variable $C = 2Z_{\ln(\theta_{ij}/2)}$, the θ_{ij}-dependent correction applied to the sum of radial coordinates in equation (3) is (see Figure 3(c)):

$$f_C(c) = \frac{e^{c/2}}{\pi}, \quad c < 2\ln\left(\frac{\pi}{2}\right). \tag{6}$$

The next step of this analysis is to derive the PDF for the sum of random variables R_i and R_j, which correspond to the radial coordinates of two random points in the hyperbolic circle. The densities of these variables are $f_{R_i} = \alpha e^{\alpha(r_i - R)}$ and $f_{R_j} = \alpha e^{\alpha\left(r_j - R\right)}$, respectively, with constant R and $r_i, r_j \leq R$. The PDF for random variable $R_{i,j} = R_i + R_j$ is thus the convolution of the independent PDFs:

$$f_{R_{i,j}}\left(r_{i,j}\right) = \int f_{R_i}\left(r_{i,j} - r_j\right) f_{R_j}\left(r_j\right) dr_j.$$

To define the integration limits for the above convolution, consider that $r_j \leq R$ and thus $r_{i,j} - r_j \leq R$ too, which yields:

$$f_{R_{i,j}}\left(r_{i,j}\right) = \int_{r_{i,j}-R}^{R} \alpha e^{\alpha\left(r_{i,j}-r_j-R\right)} \alpha e^{\alpha\left(r_j-R\right)} dr_j = \int_{r_{i,j}-R}^{R} \alpha^2 e^{\alpha\left(r_{i,j}-2R\right)} dr_j = \alpha^2\left(2R - r_{i,j}\right) e^{\alpha\left(r_{i,j}-2R\right)}, \quad r_{i,j} \leq 2R. \tag{7}$$

Figure 4 shows the shape of equation (7) for different values of parameter α.

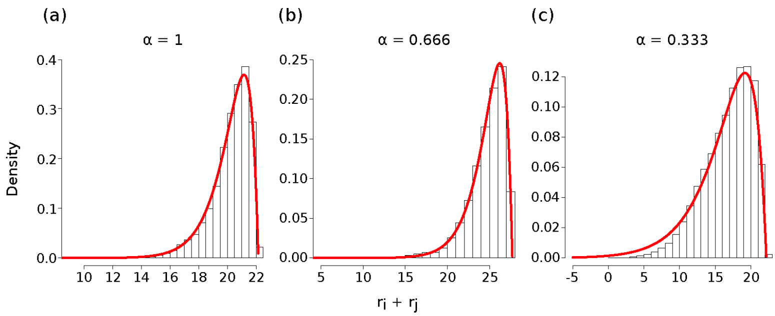

Figure 4. The shape of the distribution for the sum of the radial coordinates of two random points in the hyperbolic circle, for three different values of parameter α. The red line corresponds to the analytical expression derived in this work (equation (7)), and the histogram corresponds to the actual distribution of pairwise radial coordinate sums between 1000 points placed at random according to parameter α.

The final step of this analysis is the determination of the PDF for random variable $X_{i,j} = R_{i,j} + C$, the distance between two random points in hyperbolic space, under the important condition that variables $R_{i,j}$ and C are independent. This corresponds to the convolution of $f_C(c)$ and $f_{R_{i,j}}\left(r_{i,j}\right)$:

$$f_{X_{i,j}}\left(x_{i,j}\right) = \int f_{R_{i,j}}(h-c) f_C(c) dc.$$

The integration variable chosen is c because it has the smallest upper limit $(2\ln(\pi/2))$, which becomes the upper limit of integration. To establish the lower limit, consider that $h - c \leq 2R$, due to the domain of $f_{R_{i,j}}$, which results in:

$$f_{X_{i,j}}(x_{i,j}) = \int_{x_{i,j}-2R}^{2\ln(\pi/2)} \alpha^2(2R - x_{i,j} + c)e^{\alpha(x_{i,j}-m-2R)}\frac{e^{c/2}}{\pi}dc = \frac{1}{\pi(1-2\alpha)^2}\left(-2\alpha^2 e^{\alpha(x_{i,j}-2R-c)+c/2}\right) \left(2 + x_{i,j} - 2\alpha x_{i,j} + (4\alpha-2)R - c + 2\alpha c\right)\Big|_{x_{i,j}-2R}^{2\ln(\pi/2)}. \tag{8}$$

Changing $f_{X_{i,j}}(x_{i,j})$ to $\rho(x_{i,j})$ as defined in Section 2, the evaluation of equation (8) in the given limits is finally equal to:

$$\rho(x_{i,j}) = \frac{2\alpha^2}{\pi(1-2\alpha)^2}\Big[2e^{(x_{i,j}/2)-R} - e^{\alpha(x_{i,j}-2R-2\ln(\pi/2))+\ln(\pi/2)} \left(2 + x_{i,j}(1-2\alpha) + (4\alpha-2)\left(R + \ln\frac{\pi}{2}\right)\right)\Big], \quad x_{i,j} \leq \left(2R + 2\ln\frac{\pi}{2}\right). \tag{9}$$

Figure 5 shows the shape of equation (9) for different values of parameter α, together with the histogram of the pairwise hyperbolic distances between $N = 1000$ points, distributed quasi-uniformly at random in the hyperbolic circle of radius $R \sim \ln N$ (see [3] for the exact expression for R in the $\mathbb{H}^2$ model for complex networks). Thanks to a wxMaxima worksheet accompanying this work (see Section 4), it is possible to see that equation (9) is a valid PDF, since its integration in the domain of $x_{i,j}$ equals 1.

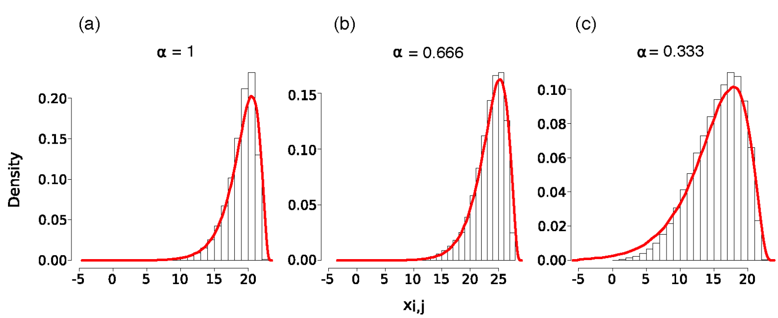

Figure 5. The shape of the distribution for the hyperbolic distance between N random points, quasi-uniformly distributed in the hyperbolic circle of radius $R \sim \ln N$ for three different values of parameter α. The red line corresponds to the analytical expression derived in this work (equation (9)), and the histogram corresponds to the actual distribution of pairwise distances between 1000 points placed at random according to parameter α.

Note that although equation (3) closely approximates equation (2), a consequence of the assumption of large r_i and r_j that leads to equation (3) is that a fraction of equation (9)'s domain is negative (see Figure 5), which is not expected from a distance distribution. This should not represent a problem for the construction of networks with scaling exponent $\gamma \in [2, 3]$, the range that is typically observed in real systems [13]. The reason why this is the case is that the choice of α impacts the network's degree exponent $\gamma = 2\alpha + 1$ [16]. When $\gamma \geq 2$, that is, $\alpha \geq 1/2$, the probability that nodes have small radial coordinates is very low, and even if their angular distances are small, the sum of their radial coordinates is big enough for equation (3) to be ≥ 0 (see Figures 1(a), 1(b) left, 4(a, b), and 5(a, b)). However, if $\alpha \in (0, 1/2)$, nodes are closer to the center of the hyperbolic disk, thus increasing the probability of observing negative distances (see Figures 1(b) right, 4(c), and 5(c)).

It is also important to note that setting $\alpha = 1/2$ in equation (9) is problematic. Nonetheless, sampling radial coordinates from equation (1) when $\alpha = 1/2$ is perfectly valid, which prompts for the determination of $\rho(x_{i,j})$ when $\alpha \to 1/2$. The resulting expression is shown in equation (10):

$$\lim_{\alpha \to 1/2} \rho(x_{i,j}) = \frac{e^{(x_{i,j}-2R)/2}}{8\pi}\left[4R^2 + 4R\left(2\ln\frac{\pi}{2} - x_{i,j}\right) + x_{i,j}^2 + 4\ln\frac{\pi}{2}\left(\ln\frac{\pi}{2} - x\right)\right], \quad x_{i,j} \leq \left(2R + 2\ln\frac{\pi}{2}\right). \tag{10}$$

4. Conclusion

In this paper, the probability density function for the approximate hyperbolic distance between points, distributed quasi-uniformly at random in a hyperbolic disk of radius R, has been determined. In the process, probability densities for the sum of the point radial coordinates, their angular coordinate difference and distance, and the natural logarithm of the latter have also been derived. These results provide insights into the processes that the components of complex systems need to optimize to give rise to the characteristic structural features of the networks they form. Equation (9) can serve as a target function or shape for geometric models of complex systems, which seek to generate scale-free and strongly clustered networks from distances between system components.

Perhaps the most important area for future work is the derivation of a general version of equation (9) for hyperbolic spaces of any dimension $D \geq 2$. In that case, nodes would be inside a hyperbolic sphere, and their position would be described by a radial coordinate r and $D-1$ angular coordinates Θ. Furthermore, equation (1) would have to change and represent a volume density rather than a planar one. Notwithstanding that such considerations are quite interesting, networks lying on $\mathbb{H}^D$ spaces ($D \gg 2$) have zero clustering in the limit of large N [15]. In contrast, real-world networks are strongly clustered, just like those generated by the $\mathbb{H}^2$ model, making them more realistic and more suitable for the study of the dynamics underlying the formation of complex systems.

Despite the fact that the most representative models of complex network formation are based on very simple rules [5, 7, 13, 16], network construction subject to optimization constraints or automata rules is a relatively fertile field in network science. Developments in the analysis of networks derived from cellular automata [17], network automata [18], trinet automata [19], and simple optimization [20] call for greater efforts aimed at understanding the relationships between simple computational systems and real-world networks.

A wxMaxima worksheet for the solution of cumbersome integrals in this paper and R code for the density, distribution, and quantile functions derived from the analytical expression determined are available at http://www.greg-al.info/code, together with a Wolfram Notebook with interactive plots of the main PDFs derived. R code for the generation of quasi-uniform point densities in hyperbolic space is available at the same link.

Acknowledgments

The authors would like to thank Ron Gordon for a post in http://math.stackexchange.com that helped with the proof associated to equation (3).

References

[1] M. Vidal, M. E. Cusick, and A.-L. Barabási, "Interactome Networks and Human Disease," *Cell*, **144**(6), 2011 pp. 986–998. doi:10.1016/j.cell.2011.02.016.

[2] G. García-Pérez, M. Á. Serrano, and M. Boguñá, "Complex Architecture of Primes and Natural Numbers," *Physical Review E*, **90**(2), 2014 022806. doi:10.1103/PhysRevE.90.022806.

[3] M. Boguñá, F. Papadopoulos, and D. Krioukov, "Sustaining the Internet with Hyperbolic Mapping," *Nature Communications*, **1**, 2010 62. doi:10.1038/ncomms1063.

[4] D. Krioukov, M. Kitsak, R. S. Sinkovits, D. Rideout, D. Meyer, and M. Boguñá, "Network Cosmology," *Scientific Reports*, **2**, 2012 793. doi:10.1038/srep00793.

[5] A.-L. Barabási and R. Albert, "Emergence of Scaling in Random Networks," *Science*, **286**(5439), 1999 pp. 509–512. doi:10.1126/science.286.5439.509.

[6] M. Á. Serrano, D. Krioukov, and M. Boguñá, "Self-Similarity of Complex Networks and Hidden Metric Spaces," *Physical Review Letters*, **100**, 2008 078701. doi:10.1103/PhysRevLett.100.078701.

[7] D. J. Watts and S. H. Strogatz, "Collective Dynamics of 'Small-World' Networks," *Nature*, **393**(6684), 1998 pp. 440–442. doi:10.1038/30918.

[8] K.-I. Goh, G. Salvi, B. Kahng, and D. Kim, "Skeleton and Fractal Scaling in Complex Networks," *Physical Review Letters*, **96**(1), 2006 018701. doi:10.1103/PhysRevLett.96.018701.

[9] C. Song, S. Havlin, and H. A. Makse, "Origins of Fractality in the Growth of Complex Networks," *Nature Physics*, **2**(4), 2006 pp. 275–281. doi:10.1038/nphys266.

[10] M. Schaller and K. Svozil, "Scale-Invariant Cellular Automata and Self-Similar Petri Nets," *The European Physical Journal B*, **69**(2), 2009 pp. 297–311. doi:10.1140/epjb/e2009-00147-x.

[11] T. Aste, T. Di Matteo, and S. T. Hyde, "Complex Networks on Hyperbolic Surfaces," *Physica A: Statistical Mechanics and Its Applications*, **346**(1–2), 2005 pp. 20–26. doi:10.1016/j.physa.2004.08.045.

[12] T. Aste, R. Gramatica, and T. Di Matteo, "Exploring Complex Networks via Topological Embedding on Surfaces," *Physical Review E*, **86**(3), 2012 036109. doi:10.1103/PhysRevE.86.036109.

[13] F. Papadopoulos, M. Kitsak, M. Á. Serrano, M. Boguñá, and D. Krioukov, "Popularity versus Similarity in Growing Networks," *Nature*, **489**(7417), 2012 pp. 537–540. doi:10.1038/nature11459.

[14] D. Krioukov and M. Ostilli, "Duality between Equilibrium and Growing Networks," *Physical Review E*, **88**(2), 2013 022808. doi:10.1103/PhysRevE.88.022808.

[15] L. Ferretti, M. Cortelezzi, and M. Mamino, "Duality between Preferential Attachment and Static Networks on Hyperbolic Spaces," *Europhysics Letters*, **105**(3), 2014 38001. http://stacks.iop.org/0295-5075/105/i=3/a=38001.

[16] D. Krioukov, F. Papadopoulos, M. Kitsak, A. Vahdat, and M. Boguñá, "Hyperbolic Geometry of Complex Networks," *Physical Review E*, **82**(3), 2010 036106. doi:10.1103/PhysRevE.82.036106.

[17] Y. Kayama, "Complex Networks Derived from Cellular Automata." arxiv.org/abs/1009.4509.

[18] D. M. D. Smith, J.-P. Onnela, C. F. Lee, M. D. Fricker, and N. F. Johnson, "Network Automata: Coupling Structure and Function in Dynamic Networks," *Advances in Complex Systems*, **14**(3), 2011 pp. 317–339. doi:10.1142/S0219525911003050.

[19] R. Southwell, H. Jianwei, and C. Cannings, "Complex Networks from Simple Rules," *Complex Systems*, **22**(2), 2013 pp. 151–173. http://www.complex-systems.com/pdf/22-2-2.pdf.

[20] B. Zheng, H. Wu, L. Kuang, J. Qin, W. Du, J. Wang, and D. Li, "A Simple Model Clarifies the Complicated Relationships of Complex Networks," *Scientific Reports*, **4**, 2014 6197. doi:10.1038/srep06197.