

# Nearest-neighbor directed random hyperbolic graphs

**Citation for published version (APA):**

Kasyanov, I. A., van der Hoorn, P., Krioukov, D., & Tamm, M. V. (2023). Nearest-neighbor directed random hyperbolic graphs. *Physical Review E*, 108(5), Article 054310. <https://doi.org/10.1103/PhysRevE.108.054310>

**DOI:**

[10.1103/PhysRevE.108.054310](https://doi.org/10.1103/PhysRevE.108.054310)

**Document status and date:**

Published: 01/11/2023

**Document Version:**

Publisher's PDF, also known as Version of Record (includes final page, issue and volume numbers)

**Please check the document version of this publication:**

- A submitted manuscript is the version of the article upon submission and before peer-review. There can be important differences between the submitted version and the official published version of record. People interested in the research are advised to contact the author for the final version of the publication, or visit the DOI to the publisher's website.
- The final author version and the galley proof are versions of the publication after peer review.
- The final published version features the final layout of the paper including the volume, issue and page numbers.

[Link to publication](#)

**General rights**

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal.

If the publication is distributed under the terms of Article 25fa of the Dutch Copyright Act, indicated by the "Taverne" license above, please follow below link for the End User Agreement:

[www.tue.nl/taverne](http://www.tue.nl/taverne)

**Take down policy**

If you believe that this document breaches copyright please contact us at:

[openaccess@tue.nl](mailto:openaccess@tue.nl)

providing details and we will investigate your claim.

## Nearest-neighbor directed random hyperbolic graphs

I. A. Kasyanov,<sup>1</sup> P. van der Hoorn,<sup>2</sup> D. Krioukov,<sup>3</sup> and M. V. Tamm<sup>4,\*</sup>

<sup>1</sup>Independent researcher, 0105 Tbilisi, Georgia

<sup>2</sup>Eindhoven University of Technology, 5612 AZ Eindhoven, Netherlands

<sup>3</sup>Northeastern University, 02115 Boston, Massachusetts, USA

<sup>4</sup>ERA Chair for Cultural Data Analytics, School of Digital Technologies, Tallinn University, 10120 Tallinn, Estonia



(Received 16 July 2023; accepted 23 October 2023; published 21 November 2023)

Undirected hyperbolic graph models have been extensively used as models of scale-free small-world networks with high clustering coefficient. Here we presented a simple directed hyperbolic model where nodes randomly distributed on a hyperbolic disk are connected to a fixed number  $m$  of their nearest spatial neighbors. We introduce also a canonical version of this network (which we call “network with varied connection radius”), where maximal length of outgoing bond is space dependent and is determined by fixing the average out-degree to  $m$ . We study local bond length, in-degree, and reciprocity in these networks as a function of spacial coordinates of the nodes and show that the network has a distinct core-periphery structure. We show that for small densities of nodes the overall in-degree has a truncated power-law distribution. We demonstrate that reciprocity of the network can be regulated by adjusting an additional temperature-like parameter without changing other global properties of the network.

DOI: [10.1103/PhysRevE.108.054310](https://doi.org/10.1103/PhysRevE.108.054310)

### I. INTRODUCTION

Reference and recommendation networks are ubiquitous [1,2]: Encyclopedia articles and scientific papers refer to each other; people recommend each other books, films, and music; online shops are full of “people who like this also like that” recommendations. These recommendations constitute directed links between objects organizing them into a directed network. The resulting networks are substantially asymmetric: The rules according to which a node becomes a source of recommendation are different from those according to which it gets recommended. In many cases the number of recommendations given *from* a node is either strictly or effectively bounded, while the number of recommendations *towards* a node is unlimited. Accordingly, out-degree and in-degree distributions in such networks are very different: Out-degree is relatively narrowly distributed, while in-degree distribution typically has a wide, often power-law, tail.

A particular example of this type of network is a network of free associations in a language [3–6]. Typically, it is constructed as follows [4,5]. Test subjects receive a set of words (stimuli) and they are to provide a first word which came to their mind as a response to each of the stimuli. The results are aggregated into a directed network of associations weighted according to the frequency with which associations appear in the dataset. The resulting networks have narrow out-degree distributions but wide in-degree distributions with power-law tails [6]. Another nice example of a directed network with asymmetry between narrow in-degree and power-law out-degree is the network of mathematical theorems studied in Ref. [7]. Yet another example where natural asymmetry between in- and out-degree distributions arises is the system

of links in encyclopedia (see, e.g., Ref. [8]). However, typically in this case both distributions have power-law tails with unequal exponents: It is less probable to have a very large out-degree than a very large in-degree.

History of network science is full of examples of how essential it is to have benchmark models of random graphs, which are able to reproduce some of the properties of the experimentally observed networks (see textbook presentation in Refs. [9–14]). Emergence of the giant cluster was understood by Flory [15] and Erdos and Renyi [16] based on a minimalistic model. Watts-Strogatz model [17] is essential for understanding the emergence of the small-world effect. Barabasi-Albert [18] and other preferential attachment models [19] shed light on the emergence of power-law degree distributions. Studying configuration [20] and exponential graph models [21] is essential to separate the effects of various topological invariants (degree, motif distributions, etc.) on the properties of networks. Hyperbolic network models, also known as random hyperbolic graphs [22–24] explain how power-law distributions, high clustering, and small-world properties coexist with each other, which is often the case in real-world networks. These models, as well as somewhat similar Apollonian networks [25,26] and their generalizations [27–30], are the first equilibrium models that unify these three properties. However, up to now hyperbolic random graph models have been confined to undirected networks (except for two very recent papers [31,32], see Discussion).

Here we develop and study a simple model of a directed network with asymmetric degree distribution: narrow distribution of the out-degree and wide distribution of the in-degree. Our model is a hyperbolic generalization of nearest-neighbor models [33–36] (see also reviews [37,38]) studied extensively for the case of Euclidean metric spaces (interestingly, nearest-neighbor graphs in high-dimensional Euclidean spaces are an important intermediate step in the construction of popular

\*Corresponding author: thumm.m@gmail.com

dimension-reduction algorithms such as diffusion maps [39], t-SNE [40], and UMAP [41]). We consider a disk in a hyperbolic space, drop a large number of points onto it uniformly at random, and then connect each point by directed links to a fixed number of its nearest neighbors. We study the limit of large networks and show that the in-degree distribution of such a network is a truncated power law, and by adjusting parameters the power-law region can be made arbitrarily wide. We also discuss how it is possible to regulate the structural parameters of the network, most importantly, the reciprocity of the network (i.e., fraction of bidirectional links in it) and the exponent of the power-law distribution.

The presentation is organized as follows. In Sec. II we define the model, discuss it qualitatively, and formulate the main results. We also define an auxiliary conjugate model, which we call “network with varied connection radius” (VCR). This model has properties similar to those of the nearest-neighbor model but is more tractable analytically. In Sec. III we turn to a more quantitative approach and provide analytical and numerical calculations for the location-dependent bond length and in-degree and show that the nearest-neighbor network has a peculiar core-periphery structure. We end Sec. III with deriving the truncated power-law behavior of the overall in-degree distribution.

In Sec. IV we recall the definition of network reciprocity and calculate it for the nearest-neighbor and VCR networks. We show how reciprocity can be regulated by introducing an additional temperature-like parameter. Finally, in Sec. V we summarize our results and discuss their possible applications and generalizations, including the control of the exponent of the in-degree distribution.

In what follows we assume some familiarity with the concept of hyperbolic spaces with constant negative curvature (see, e.g., Ref. [42] for an extended introduction). However, all the concrete formulas needed to understand the result are provided in the text to make it self-contained.

## II. DEFINITION OF THE MODEL AND QUALITATIVE DISCUSSION

### A. The $m$ -nearest-neighbor ( $m$ -NN) network model

In this section we give the definition of the nearest-neighbor network model and qualitatively discuss the bond length and the degree distribution in this network.

We start with a general definition of an  $m$ -nearest-neighbor ( $m$ -NN) network. Consider a set  $V$  of  $N$  points  $x_1, \dots, x_N$  in some normed vector space, so that distances  $\|x_i - x_j\|$  are defined for all  $i, j$ . Call an  $m$ -NN network a graph  $G$  consisting of vertices  $V$  and  $mN$  directed edges, connecting each vertex to its  $m$  nearest neighbors, i.e., to  $m$  nodes, the distance to which is smallest. Clearly, the result is a directed graph with out-degree distribution  $P_{\text{out}}(k) = \delta_{k,m}$ . The average in-degree is also  $m$ , but the in-degree distribution may be quite nontrivial.

Consider first a simple stochastic setting: Namely, let points be distributed uniformly and independently at random with a given density  $v$  per unit volume in an infinite  $d$ -dimensional space of constant curvature. This problem has been studied extensively in the case of Euclidean (zero-curvature) space [37,38]. In this case, the average network

properties are translationally invariant, and both the bond length distribution and the in-degree distribution are rather narrow. Indeed, although the exact distribution of in-degree for arbitrary  $m, d$  is unknown (see Ref. [43] where  $P_{\text{in}}(k=0)$  is calculated for  $m=1, d=2$ ), it is easy to show that the probability of both a large bond length  $r$  and a large in-degree  $k$  decays at least exponentially in both Euclidean and hyperbolic space.

Indeed, let  $A(r)$  be the volume of a ball of radius  $r$  (henceforth all distances are assumed to be hyperbolic unless mentioned otherwise; a *ball* is then defined, as usual, as a set of points at distance no more than  $r$  from the center, and a *disk* is a ball in two-dimensional space). Then define distance  $r_m$  as a solution of equation

$$vA(r_m) = m, \quad (1)$$

i.e., within the ball of radius  $r_m$  there are on average  $m$  points. The radius  $r_m$  has a meaning of “typical distance” to the  $m$ th nearest neighbor or typical bond length. Indeed, the probability that there are less than  $m$  points in the ball of radius  $r$  (which is equal to the probability that the distance to the  $m$ th neighbor is larger than  $r$ ) is

$$P_m(r) = \exp[-vA(r)] \sum_{k=0}^{m-1} \frac{[vA(r)]^k}{k!}. \quad (2)$$

Note now that in the expression

$$\exp[-vA(r)] = \sum_{k=0}^{\infty} \frac{[vA(r)]^k}{k!} \quad (3)$$

the maximal term in the right-hand side is located at  $vA(r) = k$  and thus the distribution of the  $m$ th nearest neighbor  $p_m(r) = dP_m(r)/dr$  is localized around  $r_m$ . In what follows, we use  $r_m$  as a qualitative estimate of the distance to  $m$ th nearest neighbor, since the more conventional estimate, mean distance to  $m$ th nearest neighbor,

$$r_m^{\text{mean}} = \int_0^\infty r p_m(r) dr, \quad (4)$$

is of the same order of magnitude but is a bit harder to calculate and depends on a particular form of  $A(r)$  in a more complicated way. Note also that big deviations of the bond length from  $r_m$  are extremely rare:  $A(r) \sim r^d$  in  $d$ -dimensional Euclidean space and  $A(r) \sim \exp r$  for large  $r$  in hyperbolic space, so in both cases  $P_m(r)$  given by (2) decays superexponentially, effectively limiting the possible values of  $r$ . This also means that in the translationally invariant case the in-degree of a node in the  $m$ -NN model is limited. Indeed, the superexponential decay of (2) means that there are essentially no bonds in the network longer than  $ar_m$ , where  $a$  is some numerical constant of order 1. Then the in-degree is effectively limited by

$$k_{\text{in}}^{\max} \sim vA(ar_m) = \begin{cases} ma^{d-1} & \text{for Euclidean space;} \\ me^a & \text{for hyperbolic space.} \end{cases} \quad (5)$$

The situation becomes more interesting for the case of a nearest-neighbor network in a bounded domain, where translation invariance breaks up and average properties of nodes are location dependent. Consider the simplest possible setting

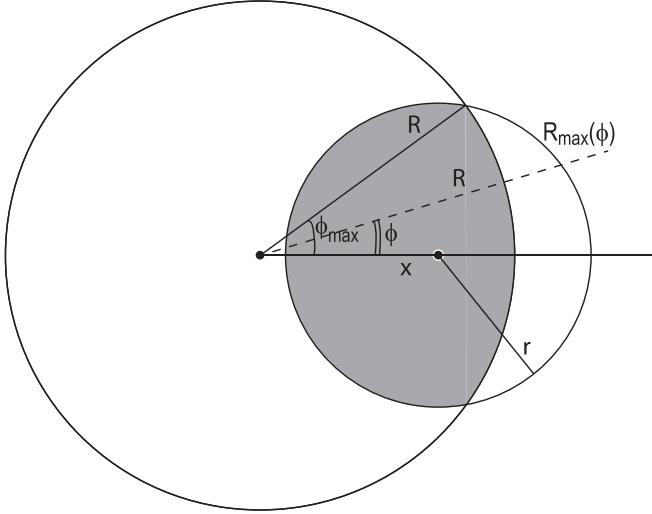


FIG. 1. Sketch of the overlap of two circles. If the circles intersect, then the shaded area  $A(r, x, R)$  is smaller than the area of the small circle  $A(r)$ . The angle  $\phi_{\max}$  and distance  $R_{\max}(\phi)$  are also shown, see (18) and (19).

when points are distributed uniformly and independently at random with a given density  $v$  inside a ball  $O(R)$  of radius  $R$  (here and in what follows whenever discussing hyperbolic space we, without loss of generality, set the space curvature  $\zeta = 1$ , i.e., distance  $R$  in the hyperbolic space is measured in the units of inverse space curvature and is therefore a dimensionless variable). In this case, the average length of a bond becomes dependent on the spatial position of its source, and the in-degree of a node depends on its radial coordinate (distance from the center of the ball). Indeed, consider the overlap of a ball of radius  $r$  with center at radial coordinate  $x$  and the ball  $O(R)$ . For  $x + r > R$  the volume of this overlap  $A(r, x, R)$  (gray area in the sketch Fig. 1) is smaller than the volume of unrestricted ball  $A(r)$  (the area of the small circle in Fig. 1). Thus, the number of points inside  $A(r_m, x, R)$  is on average smaller than  $m$ , and the typical distance to the  $m$ th nearest neighbor becomes larger than  $r_m$ . Similarly to (1) we define this typical distance  $r_m(x)$  as a solution of equation

$$vA(r_m(x), x, R) = m. \quad (6)$$

For brevity, we preserve notation  $r_m$  without explicit  $x$  dependence for the limiting value of  $r_m(x)$  in the bulk, so that  $r_m = \lim_{R \rightarrow \infty} r_m(x)$  for any fixed  $x$ .

Consider now how this spatial dependence of the bond length influences in-degree. Points in the immediate vicinity of a boundary have on average an in-degree smaller than  $m$  because part of the points, which would normally connect to them in the infinite domain case are now located outside  $O(R)$ . Conversely, points at a distance somewhat larger than  $r_m$  from the boundary have an increased average in-degree: Apart from all the bonds of the length  $r < r_m$  which they attain similarly to the points in the bulk, they also get a certain number of abnormally long incoming bonds from the points in the close vicinity of the boundary. Thus, in-degree depends on the spatial coordinate of the bond in a nonmonotonic way:

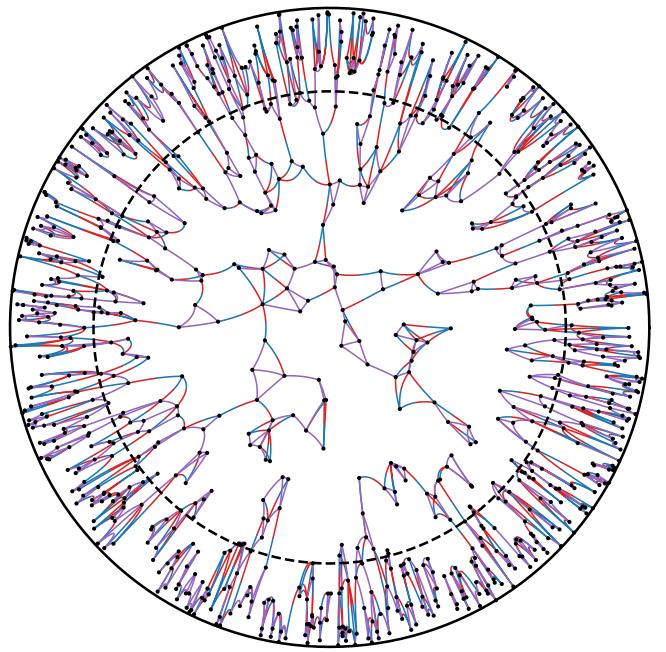


FIG. 2. Example of a directed nearest-neighbor network of  $N = 1000$  nodes on a hyperbolic disk. Each node is connected to  $m = 3$  nearest neighbors (in the colored version the source and target ends of each bond are shown with blue and red, respectively, bidirectional bonds are shown in purple). Note the core-periphery structure of the network: In the periphery the bonds are directed predominantly towards the center of the disk, and in the core there is no predominant direction. The in-degree is maximal at the crossover from core to periphery, in the vicinity of the dashed black circle.

It equals  $m$  when the distance from the boundary

$$\xi = R - x \quad (7)$$

is much larger than  $r_m$ , then increases to a larger number somewhere around  $r_m$ , and then drops to a value smaller than  $m$  in the immediate vicinity of the boundary.

This boundary effect exists both in the Euclidean and the hyperbolic space. In the Euclidean space it is relatively small (numerically, in 2D we found a spatial variation of in-degree by a factor of roughly 2), and, even more importantly, in the *thermodynamic limit*  $R \rightarrow \infty$ ,  $v = \text{const}$  it affects an infinitesimal fraction of nodes. Indeed, the fraction of volume occupied by the boundary region tends to zero for large domains in the Euclidean space. The situation is dramatically different in the hyperbolic space (see Fig. 2) where volumes of the ball  $A(r)$  and of its surface  $dA(r)/dr$  increase exponentially with  $r$ . In this case the fraction of volume taken by the boundary converges to a finite number of order

$$1 - \frac{\exp(R - r_m)}{\exp R} = 1 - \exp(-r_m). \quad (8)$$

In turn,  $r_m$  is a function of dimensionless density  $v$ , and becomes arbitrary large for small  $v$  and arbitrary small for large  $v$  [see (1)]. Thus, by changing the density of points one can cross over from the regime of bulk dominance for large

$\nu$  to the regime of boundary dominance for small  $\nu$ . As we show below, in the boundary region the in-degree distribution depends exponentially on  $\xi$  (compare to Ref. [23]), which, combined with the exponential growth of the number of nodes with a distance from the origin, leads, in full analogy with Ref. [23], to a power-law distribution of the in-degree,

$$P_{\text{in}}(k) \sim k^{-3}, \quad (9)$$

within the boundary region. Since the boundary region has a finite width, this power-law remains truncated even in the thermodynamic limit with  $k_{\text{in}}^{\max} \sim m/\nu$ .

## B. Network with VCR

Before proceeding further it is instructive to introduce an auxiliary network model, which plays the role of a canonical counterpart of the  $m$ -NN network model. Consider once again a set of points distributed uniformly and independently at random with a given density  $\nu$  inside  $O(R)$ , and for each point with radial coordinate  $x$  add outgoing bonds to all points at a distance no more than  $r_m(x)$  from it, with  $r_m(x)$  given by Eq. (6). Note that this model bears some similarity with Ref. [44], although here the rules governing connection of nodes are inhomogeneous in space, not in time. Although the out-degree of the  $m$ -NN network is  $P_{\text{out}}(k) = \delta_{k,m}$ , this auxiliary “network with varying connection radius” has a Poisson out-degree distribution with coordinate-independent mean equal to  $m$ :

$$P_{\text{out}}^{\text{MF}}(k) = e^{-m} m^k / k! \quad (10)$$

As we show below, the average in-degree of a node in this network is relatively easy to calculate: It is determined simply by the geometry of the domain and does not need averaging over simultaneous positions of several randomly located particles. On the other hand, the network with varied connection radius and the  $m$ -NN network are dual in a way similar to duality between canonical and microcanonical ensembles. Indeed, in  $m$ -NN network the out-degree is strictly fixed, while the distance to  $m$ th neighbor is fluctuating around its (position-dependent) average  $r_m(x)$  while in the network with varied connection radius the situation is inverted: The maximal length of the bond is fixed and the total number of outgoing bond is fluctuating. For large  $m$  the relative fluctuations are expected to be small

and models are expected to converge in complete analogy to the canonical and microcanonical ensembles giving the same result in thermodynamic limit.

In what follows we consider these two models in parallel, leveraging the relative analytical simplicity of the model with varied connection radius. In the next section we provide analytical and numerical calculations supporting and expanding the qualitative analysis of the two models presented above.

## III. QUANTITATIVE ANALYSIS OF THE NEAREST-NEIGHBOR MODEL

### A. Circles overlap $A(r, x, R)$

We start with finding the explicit form of the function  $A(r, x, R)$  and based on that study the asymptotic behavior of the solution of (6).

In what follows we use polar coordinates centered in the center of  $O(R)$ . Recall that the infinitesimal area element  $dA$  in hyperbolic polar coordinates is

$$dA = \sinh \rho d\rho d\theta, \quad (11)$$

and thus circumference and area of a disk of radius  $r$  are

$$\Pi(r) = dA(r)/dr = 2\pi \sinh r; \quad A(r) = 2\pi(\cosh r - 1), \quad (12)$$

respectively (recall that all the lengths are measured in the units of inverse space curvature and are therefore dimensionless). This allows us to solve equation (1) explicitly, getting the following expression for the typical length of the bond to  $m$ th nearest neighbor in the bulk

$$r_m = \cosh^{-1} \left[ \frac{m}{2\pi\nu} + 1 \right] \approx \ln \left[ \frac{m}{\pi\nu} + 2 \right], \quad (13)$$

where the last equality is valid in the small-density ( $m/\nu \gg 1$ ) limit.

Recall also the hyperbolic cosine theorem

$$\cosh a = \cosh b \cosh c - \sinh b \sinh c \cos \alpha, \quad (14)$$

which connects the length of the sides of a triangle  $a, b, c$  and the angle  $\alpha$  opposite to the side  $a$ .

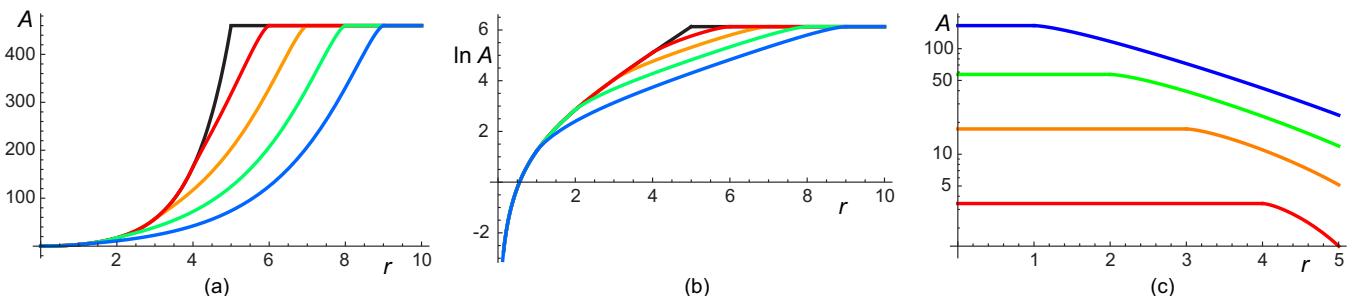


FIG. 3. The area of the overlap  $A(r, x, R)$  of two circles as a function of the radius of the inner circle  $r$  for  $R = 5$  and various coordinates of the center of smaller circle  $x$ : (from top to bottom)  $x = 0$  (black), 1 (red), 2 (orange), 3 (green), and 4 (blue) plotted in linear (a) and logarithmic (b) scale. (c) The area of the overlap  $A(r, x, R)$  as a function of the position of the center of the smaller circle  $x$  for  $R = 5$  and (from bottom to top)  $r = 1$  (red), 2 (orange), 3 (green), and 4 (blue).

These formulas are enough to explicitly calculate  $A(r, x, R)$ , the overlap of a disk with radius  $r$  and center at  $x$  and the underlying disk  $O(R)$ . Clearly, there are two simple limiting cases. If  $r \leq R - x$ , then the small circle of radius  $r$  is completely inside the disk, and thus

$$A(r, x, R) = 2\pi(\cosh r - 1). \quad (15)$$

If  $r \geq R + x$ , in turn, then the whole  $O(R)$  is inside the circle of radius  $r$  and

$$A(r, x, R) = 2\pi(\cosh R - 1). \quad (16)$$

Meanwhile, for  $R - x < r < R + x$  the two disks intersect and the area can be calculated as (see Fig. 1)

---


$$A(r, x, R) = 2\pi(\cosh r - 1) - 2 \int_0^{\phi_{\max}} d\phi \int_R^{R_{\max}(\phi)} \sinh z dz = 2\pi(\cosh r - 1) + 2\phi_{\max} \cosh R - 2 \int_0^{\phi_{\max}} \cosh R_{\max}(\phi) d\phi, \quad (17)$$

where the angle  $\phi_{\max}$ , the angular coordinate of the intersection of the circles, can be calculated from the hyperbolic cosine theorem,

$$\phi_{\max} = \arccos \left( \frac{\cosh x \cosh R - \cosh r}{\sinh x \sinh R} \right), \quad (18)$$

as well as the radial coordinate of the point at the outer part  $R_{\max}(\phi)$  of the smaller circle for any given  $\phi$ ,

$$R_{\max}(\phi) = \operatorname{arccosh} \left[ \frac{\cosh x \cosh r + \sinh x \cos \phi \sqrt{\sinh^2 r - \sin^2 \phi \sinh^2 x}}{1 + \sin^2 \phi \sinh^2 x} \right]. \quad (19)$$

The integral in (17) can be calculated explicitly:

$$\begin{aligned} \int_0^{\phi_{\max}} \cosh R_{\max}(\phi) d\phi &= \cosh r \left\{ \frac{\pi [1 - \operatorname{sgn}(\tan \phi_{\max})]}{2} + \arctan(\cosh x \tan \phi_{\max}) \right\} + \int_0^{\sinh x \sin \phi_{\max}} \frac{\sqrt{\sinh^2 r - u^2}}{1 + u^2} du \\ &= \cosh r \left\{ \frac{\pi [1 - \operatorname{sgn}(\tan \phi_{\max})]}{2} + \arctan(\cosh x \tan \phi_{\max}) \right\} - \arctan \left[ \frac{\sinh x \sinh R}{\cosh r \cosh R - \cosh x} \sin \phi_{\max} \right] \\ &\quad + \cosh r \arctan \left[ \cosh r \frac{\sinh x \sinh R}{\cosh r \cosh R - \cosh x} \sin \phi_{\max} \right]. \end{aligned} \quad (20)$$

Equations (15)–(20) collectively define the overlap area. Figure 3 shows the  $A(r, x, R)$  as a function of  $r$  and  $x$  for given  $R = 5$ . Note that in the overlap regime  $R - x < r < R + x$  the dependence of  $A$  on both  $x$  and  $r$  is roughly exponential (at least if radius  $r$  is not too small). Since this regime is the most relevant one for the dependence of the length of the bond on location, let us discuss it in more detail.

## B. Asymptotic of the bond length

It is instructive to study asymptotic behavior of the expressions (17), (18), and (20) in several steps. First, consider the thermodynamic limit  $R \rightarrow \infty$ . In this case  $\phi_{\max}$  becomes exponentially small in  $R$ :

$$\phi_{\max} \approx \tan \phi_{\max} \approx \sin \phi_{\max} \approx 2\sqrt{2} \sqrt{\cosh r - \cosh \xi} e^{\xi/2} e^{-R}, \quad \xi = R - x, \quad (21)$$

which allows us to rewrite the expression for the area  $A(r, \xi)$  in a simplified form,

$$A(r, \xi) = \begin{cases} 2\pi(\cosh r - 1) & \text{for } r \leq \xi, \\ 2\pi(\cosh r - 1) + 2\sqrt{2(\cosh r - \cosh \xi)} e^{\xi} - 2 \cosh r \arctan[\sqrt{2(\cosh r - \cosh \xi)} e^{-\xi}] \\ + 2 \arctan \left[ \frac{\sqrt{2(\cosh r - \cosh \xi)} e^{-\xi}}{\cosh r - e^{-\xi}} \right] - 2 \cosh r \arctan \left[ \cosh r \frac{\sqrt{2(\cosh r - \cosh \xi)} e^{-\xi}}{\cosh r - e^{-\xi}} \right] & \text{for } r > \xi. \\ + O(e^{-R}). \end{cases} \quad (22)$$

Introduce now  $u = r - \xi$ . If  $u$  is positive and  $\xi$  is sufficiently large (i.e., if the circle does intersect with the boundary and its center is not too close to it), then  $\cosh r \approx e^{\xi+u}/2 \gg e^{-\xi}$ , which leads to a further simplification,

$$\begin{aligned} A(r, \xi) &\approx e^{\xi} [2\sqrt{e^u - 1} + e^u (\pi - 2 \arctan \sqrt{e^u - 1})] - 2\pi \\ &\approx 4e^{\xi+u/2} - 2\pi, \end{aligned} \quad (23)$$

where the last equality is valid in the  $e^u \gg 1$  limit. Thus, we arrive at the following approximate solution of (6) giving the typical distance to the  $m$ th nearest neighbor from the point at distance  $\xi$  from the boundary of the disk,

$$r_m(\xi) = 2 \ln \left[ \frac{m}{4v} + \frac{\pi}{2} \right] - \xi = 2r_m - \xi + 2 \ln(\pi/4), \quad (24)$$

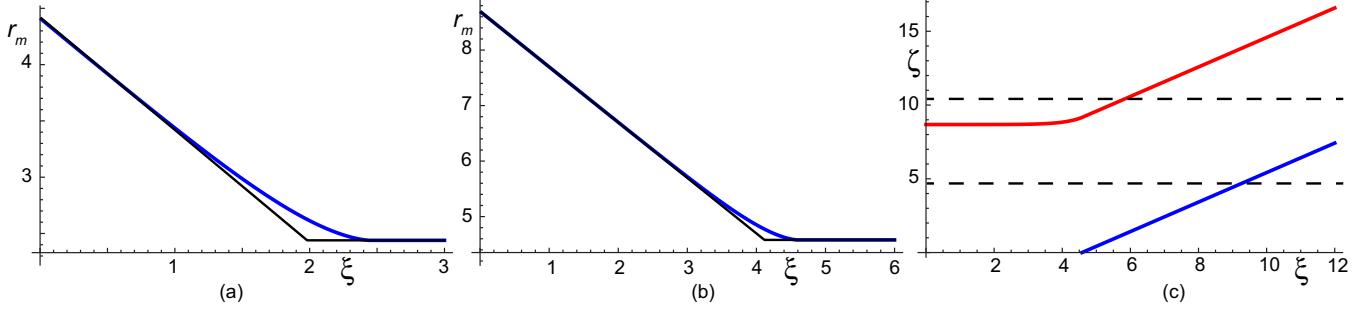


FIG. 4. The length  $r_m$  of a typical bond to  $m$ th nearest neighbor as a function of the distance  $\xi$  from the source of the bond to the boundary of the disk,  $m = 3$ ,  $v = 0.1$  (a) and  $0.01$  (b). Thick blue line corresponds to the solution of equation  $vA(r_m, \xi) = m$  in the limit of large  $R$  [i.e., for  $A(r, m)$  given by (22)]; thin black line is the approximate solution given by (24). (c) The left-hand side of Eqs. (32) for  $\xi_{\min}$  (red, top) and  $\xi_{\max}$  (blue, bottom) for  $m = 3$ ,  $v = 0.01$ . In the periphery, for  $\xi < r_m(0)$  (lower dashed line),  $\xi_{\min} = 0$ , in the core, for  $\xi > 2r_m(\infty)$  (upper dashed line),  $\xi_{\min} = \xi - r_m(\infty)$ . There is a narrow intermediate region between these two regimes.

where  $r_m$  is the limiting value of  $r_m(\xi)$  at a large distance from the boundary and is given by (13). This very simple result is valid up to exponentially small corrections if all relevant exponents are simultaneously much larger than 1:

$$e^R \gg 1, \quad e^\xi \gg 1, \quad e^{r_m(\xi)-\xi} \gg 1, \quad e^{r_m(\xi)-r_m} \gg 1, \quad (25)$$

that is, if the disk is large, then the density of points is small, and the position of point in question is not too close to the boundary and not too close to the line separating the core and periphery regions. Importantly, the solution is linear in  $\xi$ . Figure 4 shows the numerical solution of (6) for  $A$  given by (22) and the approximation (24). It is seen that while for  $m/v \sim 1$  the approximation (24) is off, it becomes better and better for smaller densities, and for  $m/v \gtrsim 10^2$  it works very well for most of the region  $0 < \xi < r_m + 2 \ln(\pi/4)$  except for small deviations when  $\xi$  approaches the upper limit. These deviations remain finite as  $v$  approaches zero and  $r_m$  diverges.

Summing up, the typical length of the bond to  $m$ th nearest neighbor depends on the position of the source in the following way: (i) If the source is more than  $r_m$  away from the boundary, then it is constant and equal to  $r_m$ ; (ii) when the source point approaches the boundary, the length of the bond grows linearly with the distance  $\xi$  to the boundary, according to (24); and (iii) there is a narrow region of finite width in the vicinity of  $\xi = r_m$ , which smoothly connects these two regimes.

In what follows we discuss the implications of this varied bond length on the *in-degree* of the nearest-neighbor network.

### C. In-degree in the network with the varied connection radius

In the previous section we discussed how the typical length of the bond depends on the position of its *source* point  $\mathbf{x} = (x, \phi)$ . In this section we are going to study the properties of the bonds with a given position of the *target* point  $\mathbf{y} = (y, \psi)$ . It is easiest to consider first the auxiliary “model with varied connection radius,” i.e., the model where a bond from  $\mathbf{x}$  to  $\mathbf{y}$  exists if and only if

$$r = \|\mathbf{x} - \mathbf{y}\| \leq r_m(x), \quad (26)$$

where  $r_m(x)$  is given by (6). Without loss of generality it is convenient to set  $\psi = 0$  and to introduce also the distance

from the target point to the boundary of the disk

$$\zeta = R - y. \quad (27)$$

Given that the points are distributed independently at random with a given density  $v$ , average in-degree of a point located at  $y$  equals

$$\bar{k}_{\text{in}}(y) = vB_m(y), \quad (28)$$

where  $B_m(y)$  is the area of the domain  $\mathcal{B}_m(y)$  defined as

$$\mathbf{x} \in \mathcal{B}_m(y) \text{ if and only if } \|\mathbf{x} - \mathbf{y}\| \leq r_m(\|\mathbf{x}\|) \text{ and } \mathbf{x} \in O(R), \quad (29)$$

and for each given  $y$  the in-degree has a Poisson distribution with average  $\bar{k}_{\text{in}}(y)$ .

In the polar coordinates associated with the center of the disk the area  $B_m(y)$  can be written formally as follows:

$$\begin{aligned} B_m(y) &= \int_{x_{\min}}^{x_{\max}} \int_{-\phi_{\max}}^{\phi_{\max}} \sinh xd\phi d\xi \\ &= 2 \int_{x_{\min}}^{x_{\max}} \phi_{\max}(x, y) \sinh xd\xi, \end{aligned} \quad (30)$$

where  $\phi_{\max}$  is the value of the angular coordinate  $\phi$  for which the distance between points  $(x, \phi)$  and  $(y, 0)$  equals  $r_m(x)$ , while  $x_{\min}$  and  $x_{\max}$  define the range of coordinates of the source points from which the target at  $y$  is accessible (provided the angular coordinates match). They are the solutions of equations

$$y - x_{\min} = r_m(x_{\min}); \quad x_{\max} - y = r_m(x_{\max}). \quad (31)$$

In terms of distances  $\xi = R - x$ ,  $\zeta = R - y$  from the boundary of the disk the equations (31) can be rewritten as

$$\zeta = \xi_{\min} + r_m(\xi_{\min}); \quad \zeta = \xi_{\max} - r_m(\xi_{\max}). \quad (32)$$

Figure 4(c) shows the behavior of the left-hand side of (32). Clearly, with respect to  $\zeta$  there are two different regimes, corresponding to the core and the periphery of the network.

In the core part, when  $\zeta \geq 2r_m$ , the accessible region is  $x \in (y - r_m, y + r_m)$ . In this case, the attachment radius is constant and equal to  $r_m$  throughout the whole accessible domain, so  $\mathcal{B}_m(y)$  is simply a circle of radius  $r_m$ , so that one immediately gets from the definition of  $r_m$  (1) that  $\bar{k}_{\text{in}}(y) = m$ .

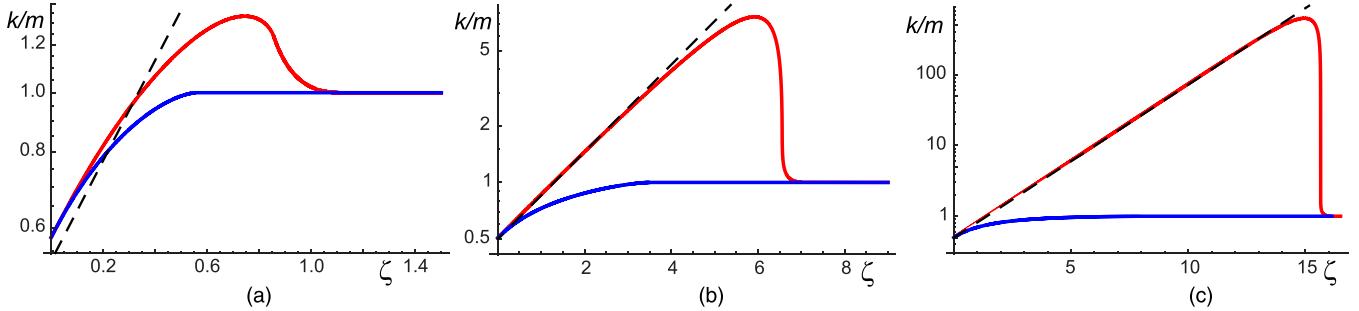


FIG. 5. The reduced average in-degree  $\bar{k}_{\text{in}}(\zeta)/m$  [(28), red, top] and bidirectional degree  $\bar{k}_{\text{periph}}(\zeta)/m$  [(52), blue, bottom] of the network with varying connection radius as a function of the distance from the boundary  $\zeta$  for (a)  $m/v = 1$ , (b)  $m/v = 10^2$ , and (c)  $m/v = 10^4$ . Dashed line indicates the approximate in-degree in the periphery region given by (36).

In the periphery, when  $\zeta < 2r_m$  the maximal length of the incoming bonds is source dependent, so  $\mathcal{B}_m(y)$  has a nontrivial shape, and its area is to be calculated using (30). The angle  $\phi_{\max}$  is obtained from the hyperbolic cosine theorem (14),

$$\cos \phi_{\max} = \frac{\cosh x \cosh y - \cosh r_m(x)}{\sinh x \sinh y};$$

$$\psi_{\max} = 2\sqrt{2}e^{-R}\sqrt{e^{\xi+\zeta}[\cosh r_m(\xi) - \cosh(\xi - \zeta)]} + O(e^{-2R}), \quad (33)$$

$$\begin{aligned} \xi_{\max} &= \zeta + r_m, \\ \xi_{\min} &= \begin{cases} 0 & \text{for } \zeta \leq r_m(0), \\ \text{solution of } A(\zeta - \xi_{\min}, \xi_{\min}) = m/v & \text{for } r_m(0) < \zeta < 2r_m. \end{cases} \end{aligned}$$

This expression is easy to calculate numerically, the results are shown in Fig. 5. In the core in-degree equals  $m$ , while in the periphery it grows exponentially with the distance from the boundary, spanning three orders of magnitude for  $m/v = 10^4$ , in between, there is a narrow crossover region where in-degree drops drastically with increasing  $\zeta$ .

The exponential asymptotic in the periphery can easily be extracted by substituting approximation (24) for  $r_m(\xi)$ . One gets

$$\begin{aligned} B_m(\zeta) &\approx 2 \int_0^{\xi_{\max}} \sqrt{e^{\xi-\zeta} \left[ \frac{\pi^2}{16} e^{2r_m-\xi} - \cosh(\xi - \zeta) \right]} d\xi \\ &\approx 2 \int_0^{\infty} \frac{\pi}{4} e^{r_m-\xi+\zeta/2} d\xi = \frac{\pi}{2} e^{\zeta+m/2} \approx \frac{1}{2} \frac{m}{v} e^{\zeta/2}, \quad (36) \end{aligned}$$

where we used the fact that the integral is controlled by its lower bound, that  $e^{2r_m-\xi} \gg \cosh(\xi - \zeta)$  in its vicinity, and that  $m/v = 2\pi(\cosh r_m - 1) \approx \pi \exp r_m$ . This asymptotic is shown with a dotted line in Fig. 5.

The in-degree of the points located exactly at the boundary is  $m/2$ . The maximal in-degree of a network is reached at  $\zeta_m = 2r_m - a$  where  $a$  numerically is close to 1. Thus, the maximal in-degree of the network scales in the leading order as

$$k_{\text{in}}^{\max} \approx m e^{\zeta_m/2} = C \times m e^{r_m} = C_2 \times \frac{m^2}{v} \quad (37)$$

with the constant  $C_2 \approx (2\pi)^{-1} \exp(-a) \approx 0.06$ .

where we switched from  $x, y$  to  $\xi, \zeta$  notation and used the expansion in powers of  $e^{-R}$ . Thus, in the leading order in  $e^{-R}$  one gets

$$\begin{aligned} B_m(\zeta) &= 2\sqrt{2} \int_{\xi_{\min}}^{\xi_{\max}} \sqrt{e^{\xi-\zeta} [\cosh r_m(\xi) - \cosh(\xi - \zeta)]} d\xi \\ &\quad + O(e^{-R}), \end{aligned} \quad (34)$$

where  $r_m(\xi)$  is a solution of  $A(r_m, \xi) = m/v$ ,  $A(r_m, \xi)$  is given by (22), and the limits of integration are

$$\begin{aligned} \xi_{\max} &= \zeta + r_m, \\ \xi_{\min} &= \begin{cases} 0 & \text{for } \zeta \leq r_m(0), \\ \text{solution of } A(\zeta - \xi_{\min}, \xi_{\min}) = m/v & \text{for } r_m(0) < \zeta < 2r_m. \end{cases} \end{aligned} \quad (35)$$

#### D. In-degree distribution of the $m$ -NN network

The network with varied connection radius described above has an advantage of being relatively easy to treat analytically. The analysis of the in-degree in the  $m$ -NN network is a bit more cumbersome. However, the result is qualitatively very similar.

The average in-degree of a point located at  $\mathbf{y} = (R - \zeta, 0)$  in the  $m$ -NN network can be formally written down as follows. For each pair of nodes define  $m_{ij}$ , the position of node  $i$  in the ordered list of neighbors of the node  $j$ ; that is,  $m_{ij} = 1$  if  $i$  is the nearest neighbor of  $j$ ,  $m_{ij} = 2$  if it is second nearest neighbor of  $j$ , etc. Then the in-degree of node  $i$

$$k_{\text{in}}^{(i)} = \sum_{j \neq i} \mathbb{I}(m_{ij} \leq m), \quad (38)$$

where  $\mathbb{I}(x)$  is the indicator function. Now fix the position of  $i$ th node at  $\mathbf{y}$  and take the average of in-degree. One gets

$$\bar{k}_{\text{in}}(\zeta) = \sum_{j \neq i} \sum_{k \leq m} \Pi(k, j | \zeta) = \int_{O(R)} v \sum_{k \leq m} \pi_k(\mathbf{x} | \zeta) d\mathbf{x}, \quad (39)$$

where  $\Pi(k, j | \zeta)$  is the probability that the node at  $\mathbf{y} = (R - \zeta, 0)$  is exactly the  $k$ th nearest neighbor of the  $j$ th node,  $\pi_k(\mathbf{x} | \zeta)$  is the similar probability for a node located at  $\mathbf{x}$ , and

we took into account that nodes are distributed uniformly at random within  $O(R)$  with density  $v d\mathbf{x}$ . The probability  $\pi_k$  is the same as the probability that there are exactly  $k - 1$  points within the area  $A(||\mathbf{x} - \mathbf{y}||, x, R)$  and is thus given by the Poisson distribution

$$\pi_k(y, \mathbf{x}, v, R) = \frac{[vA(||\mathbf{x} - \mathbf{y}||, x, R)]^{k-1}}{(k-1)!} \times \exp(-vA(||\mathbf{x} - \mathbf{y}||, x, R)), \quad (40)$$

where the function  $A(r, x, R)$  is defined by (15)–(20), and (39) can be rewritten as:

$$\begin{aligned} \bar{k}_{in}(y, m, v, R) &= \int_0^R \int_{-\pi}^{\pi} v \sinh x dx d\psi \sum_{k \leq m} \pi_k(y, \mathbf{x}, R, v) \\ &= \int_0^R \int_{-\pi}^{\pi} v \sinh x dx d\psi \\ &\times \frac{\Gamma(m, vA(||\mathbf{x} - \mathbf{y}||, x, R))}{m!}, \end{aligned} \quad (41)$$

where  $\Gamma(k, x)$  is the upper incomplete gamma function.

For large  $R$  the integrand in the right-hand side is localized in the vicinity of the position of the target node. This allows us to rewrite in the limit  $R \rightarrow \infty$ ,

$$\begin{aligned} \bar{k}_{in}(y, m, v) &= \lim_{R \rightarrow \infty} \bar{k}_{in}(y, m, v, R) \\ &= \int_0^\infty e^{-\xi} d\xi \int_{-\infty}^\infty d\omega \frac{v \Gamma(m, vA(||\mathbf{x} - \mathbf{y}||, \xi))}{m!}. \end{aligned} \quad (42)$$

where we introduced linear coordinate along the boundary,  $\omega = \psi \sinh R$ , and the function  $A(r, \xi)$  is given by (22).

Figure 6 shows the behavior of (42) for varying  $m$  and fixed  $m/v$ , as well as the results of corresponding numerical simulations (see Appendix for the details of the simulation procedure). Qualitatively the behavior of the in-degree of  $m$ -NN networks given by (42) is, indeed, similar to that of the VCR network given by (34): There are still the core region with  $\bar{k}_{in} = m$  and the periphery region with exponential dependence of the in-degree on the distance to the boundary  $\bar{k}_{in}(\xi) = (m/2) \exp(\xi/2)$ . The only difference is in the width of the crossover region, connecting these two asymptotic behaviors: It is wider for smaller  $m$  due to the fluctuations in the bond length. In the  $m/v \rightarrow \infty$  limit the behavior of the  $m$ -NN network converges, as expected, to that of the network with varying connection radius, provided that the ratio  $m/v$  is the same.

Figure 7 shows how the behavior of the in-degree changes with changing  $v$  and fixed  $m$ . Similarly to the VCR networks, the width of the periphery region and the maximal in-degree of the network increase with decreasing  $v$ , with the maximal in-degree scaling according to (37).

### E. Global in-degree distribution of the $m$ -NN network

In order to understand the global in-degree distribution of the  $m$ -NN network (i.e., summed over all nodes with various spatial positions) let us first discuss how the nodes of the network are distributed between the core and periphery

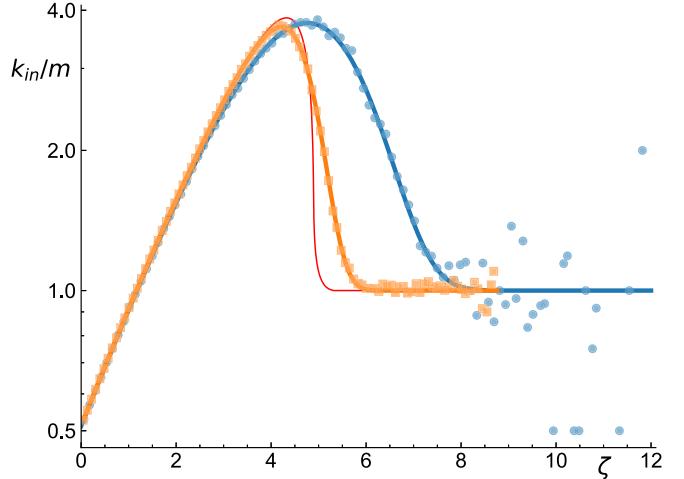


FIG. 6. Average in-degree of a node in the  $m$ -NN network as a function of the distance to the boundary  $\zeta$  for  $m/v = 40$ . Thin red line corresponds to the result (34) for the VCR network, thick blue (right) and orange (left) lines - to the  $m$ -NN networks with  $m = 2, v = 0.05$  and with  $m = 20, v = 0.5$ , respectively. The lines are obtained by numerical evaluation of (41), and the blue circles and orange squares are the results of numerical simulations averaged over 100 networks of  $10^4$  nodes (significant fluctuations for large  $\xi$  are due to the fact that fraction of nodes with large  $\xi$  decays as  $e^{-\xi}$ ).

regions. The fraction of nodes with a given distance  $\zeta$  to the boundary is

$$f(\zeta) d\zeta = \frac{\sinh(R - \zeta)}{\sinh R} d\zeta = e^{-\zeta} d\zeta + O(e^{-R}). \quad (43)$$

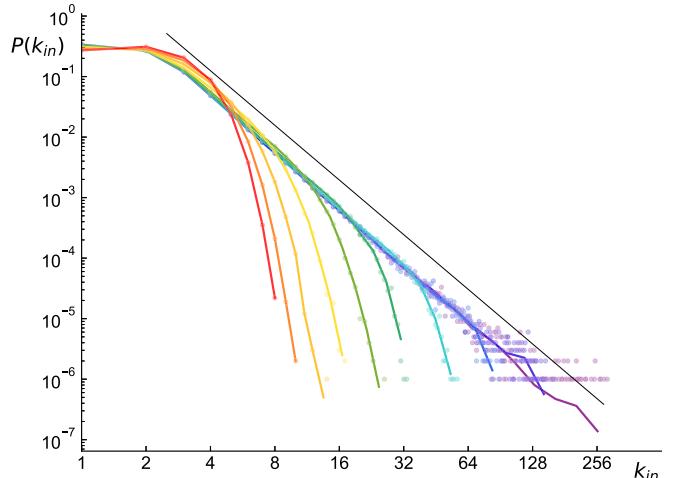


FIG. 7. In-degree distribution of the  $m$ -NN network with  $m = 2$  and  $v$  changing from 0.5 (leftmost curve) to 0.001 (rightmost curve),  $v$ 's at adjacent curves differ by a factor of 2, curves are colored from red to purple in the rainbow order. The points are obtained by averaging over 100 independent networks of  $N = 10000$  nodes for individual values of in-degree, continuous curves correspond to logarithmic binning with 25 bins per curve. The slope of the thin straight line is  $-3$ .

Thus, in the limit  $R \rightarrow \infty$  the fraction of the nodes in the core of the network is

$$\begin{aligned} f_{\text{core}} &= \int_{2r_m}^{\infty} f(\zeta) d\zeta = e^{-2r_m} \\ &= [y + 1 - \sqrt{y(y+2)}]^2, \quad y = \frac{m}{2\pi\nu}. \end{aligned} \quad (44)$$

Thus, in the large-density limit  $f_{\text{core}}$  converges to 1

$$f_{\text{core}} \approx 1 - 2\sqrt{2}\sqrt{y} + O(y) \text{ for } y \ll 1, \quad (45)$$

while in the small-density limit

$$f_{\text{core}} \approx \frac{1}{4(y+1)^2} \quad \text{for } y \gg 1 \quad (46)$$

and the core constitutes a negligibly small fraction of the nodes. The fraction of nodes belonging to the crossover region, where the in-degree increases with radial coordinate, is of the same order of magnitude,  $f_{\text{cross}} \approx f_{\text{core}}(e^a - 1)$ , where  $a$  is, once again, the width of that region.

Thus, in terms of the overall distribution of in-degree there are two regimes. For  $y = m/(2\pi\nu) \ll 1$  the network is dominated by the core where the in-degree distribution is approximately Poisson with average  $m$ . For  $y \gg 1$  the network is dominated by the peripheral region where

$$f(\zeta) = e^{-\zeta}, \quad \bar{k}_{\text{in}}(\zeta) \approx \frac{m}{2} e^{\zeta/2}, \quad (47)$$

where  $0 \leq \zeta \lesssim \ln(2(y+1) - a)$ . For each  $\zeta$  true in-degree  $k$  is narrowly distributed around  $\bar{k}_{\text{in}}(\zeta)$ . This allows us to obtain the distribution of in-degree by replacing  $k$  with  $\bar{k}_{\text{in}}$  for each  $\zeta$ :

$$\begin{aligned} P_{\text{in}}(k) dk &= f(\zeta) d\zeta \approx \left(\frac{2k}{m}\right)^{-2} d\left(2 \ln \frac{2k}{m}\right) \\ &= \frac{m^2}{2k^3} dk; \quad m < k < k_{\text{in}}^{\max}, \end{aligned} \quad (48)$$

where  $k_{\text{in}}^{\max}$  is given by (37). Thus, the in-degree distribution in the  $y \gg 1$  is a truncated power law with exponent  $-3$  spanning values from  $k_{\text{in}}^{\min} \approx m$  to  $k_{\text{in}}^{\max} \approx 0.06 m^2/\nu$ . Figure 7 shows the results of numerical simulations of the in-degree for networks with  $m = 2$  and varying  $\nu$  confirming this prediction.

#### IV. STRUCTURE BEYOND DEGREE DISTRIBUTION: RECIPROCITY

In this section we go beyond the discussion of the degree distribution and discuss the simplest nonlocal structural characteristic of the network models introduced above. In directed networks the simplest nontrivial metric characterizing link-to-link interaction is called network reciprocity. It is related to the probability that, given that there is a link  $i \rightarrow j$  there is also a link  $j \rightarrow i$  and is conventionally defined [45,46] as

$$r = \frac{N_{\leftrightarrow}}{N_b - N_{\leftrightarrow}}, \quad (49)$$

where  $N_b$  is the total number of bonds in the network ( $N_b = mN$  for the  $m$ -NN network) and  $N_{\leftrightarrow}$  is the total number of pairs of bidirectional links. Note that in the usual definition

of reciprocity [45,46] a bidirectional link is considered as a single link, while in our definition it corresponds to two links in different directions, thus the denominator of the reciprocity should be  $N_b - N_{\leftrightarrow}$  in our notation. For the network with no bidirectional links  $r = 0$ , and  $r = 1$  if all links are bidirectional ( $N_{\leftrightarrow} = N_b/2$  in this case).

For undirected networks the simplest nontrivial metric of the relative position of the edges involves not two but three edges. It is the clustering coefficient of a node  $c_i$  [11,13], defined as the fraction of node pairs adjacent to it which have a link between them, and the average clustering coefficient of a networks  $C = N^{-1} \sum c_i$ . To characterize clustering in directed networks the average clustering coefficient of the corresponding undirected network (i.e., the network obtained from the original one by replacing all directed links with undirected ones) is often used. Note, however, that such an approach leaves out the distribution of directed motifs in the original network, which is known to contain plenty of important information about its structure and function [47,48].

In geometrical random networks, where edge formation is based on distance between nodes (which is a symmetric function of node positions and obeys the triangular inequality), edges are spatially correlated. As a result both reciprocity and clustering coefficient remain of order 1 even in the limit of large networks. In what follows, we use the reciprocity as the main metric characterizing these correlations in the  $m$ -NN and VCR networks. We also provide numerical results for the average clustering coefficient of the corresponding undirected network. Study of the full distributions of directed motifs in these networks is of obvious interest but goes beyond the scope of the paper.

##### A. Reciprocity of the VCR network

We start with the simpler case of the network with varied connection radius. In this network a link  $\mathbf{x} \rightarrow \mathbf{y}$  exists if distance  $\|\mathbf{x} - \mathbf{y}\|$  is smaller than  $r_m(\xi)$  while  $\mathbf{y} \rightarrow \mathbf{x}$  exists if  $\|\mathbf{x} - \mathbf{y}\|$  is smaller than  $r_m(\zeta)$ , where  $\xi, \zeta$  are distances from points  $\mathbf{x}, \mathbf{y}$  to the boundary of  $O(R)$ . Thus the link is bidirectional if

$$\|\mathbf{x} - \mathbf{y}\| \leq \min(r_m(\xi), r_m(\zeta)) = r_m(\max(\xi, \zeta)), \quad (50)$$

where we used that  $r_m(\xi)$  is a monotonically decreasing function of  $\xi$ . Note that this means that outward-directed bonds (i.e., bonds for which  $\xi < \zeta$ ) in VCR are always reciprocated, while inward-looking ones not necessarily so. Therefore, in this case reciprocity  $r$  has also the meaning of the ratio of the numbers of outward and inward-directed bonds.

The mean number of bidirectional links from node  $\mathbf{x}$  can be written as

$$\begin{aligned} \bar{k}_{\leftrightarrow}(\mathbf{x}) &= \bar{k}_{\leftrightarrow}(\xi) = \int_{O(R)} d\mathbf{y} \nu \Theta[r_m(\max(\xi, \zeta)) - \|\mathbf{x} - \mathbf{y}\|] \\ &= \int_{\xi}^R \int_{-\pi}^{\pi} \nu \sinh(R - \zeta) d\zeta d\phi \Theta[r_m(\zeta) - \|\mathbf{x} - \mathbf{y}\|] \\ &\quad + \int_0^{\xi} \int_{-\pi}^{\pi} \nu \sinh(R - \zeta) d\zeta d\phi \Theta[r_m(\zeta) - \|\mathbf{x} - \mathbf{y}\|], \end{aligned} \quad (51)$$

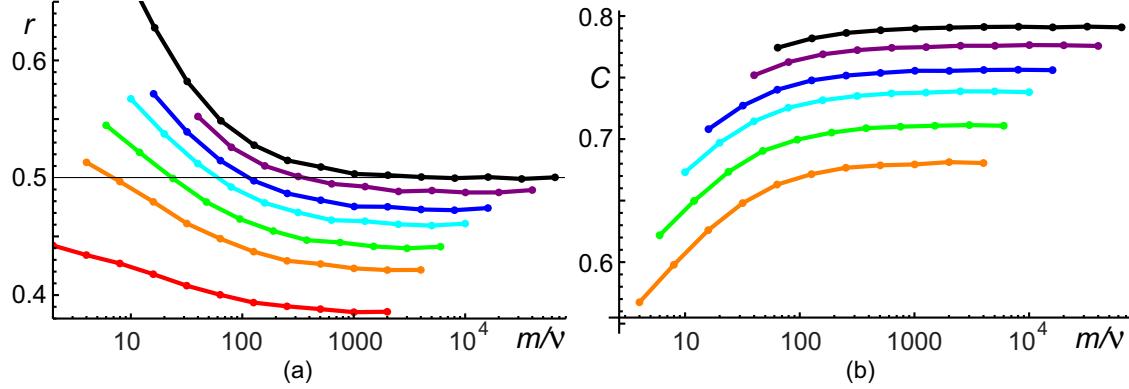


FIG. 8. (a) Network reciprocity  $r$  and (b) average clustering coefficient  $C$  of  $m$ -NN and VCR networks as functions of  $m/v$ . Numerical results for  $m$ -NN networks with (from bottom to top)  $m = 1$  (red), 2 (orange), 3 (green), 5 (cyan), 8 (blue), and 20 (purple) are shown, as well as for the VCR network (topmost, black,  $m = 32$ ). Dots are results of simulations average over 100 realizations of networks of  $10^4$  nodes. The lines are but a guide to the eye. Note that VCR reciprocity converges to  $r = 1/2$  (thin horizontal line) for large  $m/v$ . Clustering coefficient for the  $m$ -NN network with  $m = 1$  is equal to 0.

where  $\Theta(x)$  is the Heaviside theta function. Using (14) one gets in the large- $R$  limit

$$\bar{k}_{\leftrightarrow}(\xi) = 2\nu \int_{\xi}^{\zeta_{\max}} d\xi \sqrt{2[\cosh(r_m(\xi)) - 1]e^{\xi-\zeta} - (e^{\xi-\zeta} - 1)^2} + 2\nu \int_{\zeta_{\min}}^{\xi} d\xi \sqrt{2[\cosh(r_m(\xi)) - 1]e^{\xi-\zeta} - (e^{\xi-\zeta} - 1)^2}, \quad (52)$$

where

$$\begin{aligned} \zeta_{\min} &= \max[0, \xi - r_m(\xi)] \\ \zeta_{\max} &\text{ is the solution of } \zeta_{\max} = \xi + r_m(\zeta_{\max}). \end{aligned} \quad (53)$$

Given that  $r_m(\xi) \geq r_m$  for all  $\xi$  and  $r_m(\xi) = r_m$  for  $\xi \geq r_m$ , this further simplifies to

$$\begin{aligned} \zeta_{\min} &= \max[0, \xi - r_m] \\ \zeta_{\max} &= \xi + r_m. \end{aligned} \quad (54)$$

In Fig. 5 the results of numerical integration of (52) for  $m/v = 1, 100, 10^4$  are shown with blue lines. For  $\xi > r_m$   $\bar{k}_{\leftrightarrow}(\xi) \equiv m$ , i.e., all out-bonds are simultaneously in-bonds. For smaller  $\xi$  the number of bidirectional bonds is smaller than  $m$ , obviously

$$\min(\bar{k}_{\text{in}}(\xi), m) \geq \bar{k}_{\leftrightarrow}(\xi). \quad (55)$$

Reciprocity can be easily expressed in terms of mean number of bidirectional bonds  $\bar{k}_{\leftrightarrow}$ . Indeed, the expected total number of such bonds is

$$N_{\leftrightarrow} = \frac{N}{2} \int_0^{\infty} f(\xi) \bar{k}_{\leftrightarrow}(\xi) d\xi, \quad (56)$$

where  $N$  is the total number of nodes in the network,  $f(\xi) d\xi$  is the fraction of nodes with spatial coordinate  $\xi$ ,  $f(\xi) = \exp(-\xi)$  for large  $R$ , see (43). In the small-density limit  $m/v \gg 1$  one gets in the leading order in  $\mu$

$$\cosh r_m(\xi) \approx \frac{1}{32} \left( \frac{m}{v} \right)^2 e^{-\xi}. \quad (57)$$

Substituting this into (53) one gets in the leading order in  $m/v$

$$\begin{aligned} N_{\leftrightarrow} &= \frac{Nm}{4} \left( \int_0^{\infty} e^{-\xi} d\xi \int_0^{\xi} e^{-\xi/2} d\xi \right. \\ &\quad \left. + \int_0^{\infty} e^{-\xi/2} d\xi \int_{\xi}^{\infty} e^{-\xi} d\xi \right) = \frac{Nm}{3}. \end{aligned} \quad (58)$$

Thus,  $r$  converges to  $1/2$  for large densities. In turn, in the small- $m/v$  regime the network is dominated by the core where all bonds are bidirectional, and  $r$  converges to 1 for  $m/v \rightarrow 0$ . This predictions are confirmed by numerical simulations of large VCR networks as shown in Fig. 8.

## B. Reciprocity of the $m$ -NN network

Consider now the reciprocity of the  $m$ -NN networks. Most notably, it is clear that reciprocity of the  $m$ -NN network is smaller than that of the corresponding VCR network. In particular, in the VCR all bonds far from the boundary are bidirectional and all outward-directed bonds are reciprocated, while in the  $m$ -NN they are not. Indeed, for example, for VCR the condition to connect two nodes in the core is  $\|\mathbf{x} - \mathbf{y}\| < r_m$  which has a  $\mathbf{x} \leftrightarrow \mathbf{y}$  symmetry, while the relation “ $x$  is  $m$ th nearest neighbor of  $y$ ” is not. However, one expects the reciprocity of the  $m$ -NN networks to approach that of the VCR network in the limit of  $m \rightarrow \infty, m/v = \text{const}$ .

We have not been able to calculate reciprocity explicitly in the general case of an  $m$ -NN network. However, here we provide some estimates for the case of 1-NN network. Given that VCR network corresponds to the  $m \rightarrow \infty$  limit, it is natural to expect that reciprocity for arbitrary  $m$  to lie between these two limiting cases (these expectations are confirmed by numerical simulations, see Fig. 8).

Let  $\mathbf{x}$  be a position of a node and  $\mathbf{y}$  the position of its nearest neighbor at distance  $r = \|\mathbf{x} - \mathbf{y}\|$ . This means that there are no other nodes inside the light gray area in the sketch Fig. 9. In turn, for node at  $\mathbf{x}$  to be the nearest neighbor of the node at  $\mathbf{y}$  there should be no nodes closer to  $\mathbf{y}$  than  $r$ , i.e., there additionally should be no nodes in the dark gray figure in the

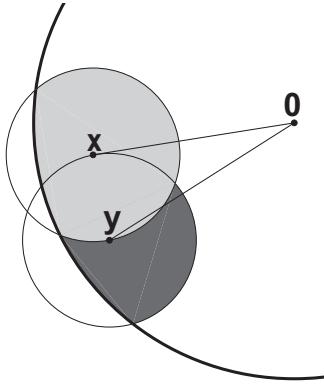


FIG. 9. A sketch illustrating the calculation of the probability of a bidirectional link in a 1-NN network. Thick line is the boundary of  $O(R)$ , the thin circles have radius  $\|\mathbf{x} - \mathbf{y}\|$ .  $\mathbf{y}$  is the nearest neighbor of  $\mathbf{x}$  if there are no nodes in the light-gray area, and  $\mathbf{y}$  is simultaneously the nearest neighbor of  $\mathbf{x}$  if additionally there are no nodes in the dark-gray area.

sketch. The area of this figure can be written formally as

$$A_{\Rightarrow}(\mathbf{x}, \mathbf{y}, R) = \int \Theta(R - z) \Theta(r - \|\mathbf{y} - \mathbf{z}\|) \Theta(\|\mathbf{x} - \mathbf{z}\| - r) d\mathbf{z}, \quad (59)$$

where we keep using the notation  $\|\mathbf{z}\| = z$ . The probability of there being no nodes in this area is  $\exp(-\nu A_{\Rightarrow})$  and the overall probability that bond from  $\mathbf{x}$  to its nearest neighbor is bidirectional is obtained by integrating over possible position of the node  $\mathbf{y}$ ,

$$\begin{aligned} p_{\Rightarrow}(\mathbf{x}, \nu, R) &= \int_{O(R)} \nu d\mathbf{y} \pi_1(\mathbf{x}, \mathbf{y}, \nu, R) \exp(-\nu A_{\Rightarrow}(\mathbf{x}, \mathbf{y}, R)) \\ &= \int_{O(R)} \nu d\mathbf{y} \exp\{-\nu[A(\mathbf{x}, \mathbf{y}, R) + A_{\Rightarrow}(\mathbf{x}, \mathbf{y}, R)]\}, \end{aligned} \quad (60)$$

where  $\pi_1$  is given by (40). Total number of bidirectional bonds  $N_{\Rightarrow}$  can be integrating (60) over the spatial distribution of the nodes. Notably,

$$\int \pi_1 d\mathbf{y} = \int_{O(R)} \nu d\mathbf{y} \exp(-\nu(A(\mathbf{x}, \mathbf{y}, R))) = 1 - O(e^{-R}) \quad (61)$$

(i.e., every node has a single nearest neighbor provided the network has at least two nodes), so  $p_{\Rightarrow} < 1$  as soon as  $A_{\Rightarrow} > 0$ . Explicit calculation of (59) is very cumbersome. Here we limit ourselves to the study of the core of the network where  $\max(x, y) \leq R - \|\mathbf{x} - \mathbf{y}\|$  and one can therefore omit the first  $\Theta$ -function in (59). In this case explicit calculation gives

$$\begin{aligned} A_{\Rightarrow}(\|\mathbf{x} - \mathbf{y}\| = r) &= 4 \cosh r \arctan\left(\frac{\cosh r}{\sqrt{1 + 2 \cosh r}}\right) \\ &\quad + 4 \arctan \sqrt{1 + 2 \cosh r} - 2\pi, \end{aligned} \quad (62)$$

which asymptotically behaves as

$$\begin{aligned} \frac{2\pi + 3\sqrt{3}}{6} r^2 + O(r^4) &\quad \text{for } r \ll 1, \\ 2\pi \cosh r - 8 \cosh(r/2) + O(e^{-r}) &\quad \text{for } r \gg 1, \end{aligned} \quad (63)$$

and thus the fraction of bidirectional bonds in the core of the network is

$$\begin{aligned} P_{\Rightarrow}^{(\text{core})} &= \pi \int_0^\infty \nu \sinh r dr \exp\{-\nu[A(r) + A_{\Rightarrow}(r)]\} \\ &= \frac{1}{2} \int_0^1 e^{-\nu A_{\Rightarrow}} d(e^{-\nu A}) \\ r^{(\text{core})} &= \left[ \frac{1}{P_{\Rightarrow}^{(\text{core})}} - 1 \right]^{-1}. \end{aligned} \quad (64)$$

This integral is easy to calculate numerically, and one gets the following limits for reciprocity in the core for large and small  $\nu$ :

$$r^{(\text{core})} = \begin{cases} \frac{3\pi}{5\pi + 3\sqrt{3}} \approx 0.4509\dots & \text{for } \nu \rightarrow \infty, \\ \frac{1}{3} & \text{for } \nu \rightarrow 0. \end{cases} \quad (65)$$

For large  $\nu$  the network is dominated by the core and one expects  $r \approx r^{(\text{core})}$ , which is confirmed by numerical results shown in Fig. 8(a) (the red line corresponds to  $m = 1$ ). In the boundary-dominated regime the reciprocity is larger than the low-density core value of  $1/3$  but still smaller than the reciprocity of the small- $\nu$  network.

### C. Regulating reciprocity with changing temperature

Reciprocity varies widely in the experimentally observed directed networks [31,45]. For example, citation networks [49] and mathematical proof networks [7] have zero or essentially zero reciprocity. Trade networks have reciprocity close to 1 [45]. However, large classes of networks have intermediate reciprocities. For example, the free association networks we mentioned in the Introduction as a proxy of a more general reference network typically have reciprocity of order 0.1 [6,45], which is, on the one hand, much larger than one would expect in a random network, but, on the other hand, significantly smaller than what is expected in the  $m$ -NN network with similar out-degree ( $m \approx 30$ ).

This gives rise to a question whether it is possible to modify the definition of  $m$ -NN and VCR models in a way that the value of reciprocity becomes adjustable. This can be done by introducing some form of stochasticity in the formation of the links in the network. Start once again with the VCR. Recall that there is a bond from a node at  $\mathbf{x}$  to a node in  $\mathbf{y}$  if and only if  $\|\mathbf{x} - \mathbf{y}\| < r_m(\mathbf{x})$ , where  $r_m(\mathbf{x})$  is the solution of (6), i.e., putting it more formally, the probability that a bond is formed between two points is

$$p_{\mathbf{x} \rightarrow \mathbf{y}} = P(\|\mathbf{x} - \mathbf{y}\|, r_m(\mathbf{x})) = \Theta[r_m(\mathbf{x}) - \|\mathbf{x} - \mathbf{y}\|], \quad (66)$$

where  $r_m(\mathbf{x})$  is determined by

$$\int_{O(r)} P(\|\mathbf{x} - \mathbf{y}\|, r_m(\mathbf{x})) d\mathbf{y} = m/\nu. \quad (67)$$

Formulated like this, it is clear that VCR model can be naturally generalized to an arbitrary non-negative kernel  $P(\|\mathbf{x} - \mathbf{y}\|, \sigma(\mathbf{x}))$  describing the probability of bond formation,  $\sigma$  being some set of governing parameters. If these parameters are chosen according to normalization rule (67),

then the resulting network has a position-independent average out-degree  $m$ . One possible natural choice is a Gaussian kernel,

$$P(||\mathbf{x} - \mathbf{y}||, \sigma(\mathbf{x})) = \exp \left[ -\frac{||\mathbf{x} - \mathbf{y}||^2}{\sigma(\mathbf{x})^2} \right], \quad (68)$$

which is used, for example, as a measure of point similarity in t-SNE [40] and to estimate the probability of bond formation in polymer physics (see, e.g., Refs. [50,51]). It is, however, not convenient for our purposes since it has only a single governing parameter  $\sigma(\mathbf{x})$ , which is to be fixed by (67). As a result, it is not possible to regulate out-degree and reciprocity independently in this case. We consider instead the Fermi-Dirac kernel

$$P(||\mathbf{x} - \mathbf{y}||, r_m, T) = \left\{ 1 + \exp \left[ \frac{||\mathbf{x} - \mathbf{y}|| - r_m(\mathbf{x}, T)}{T} \right] \right\}^{-1}, \quad (69)$$

which was suggested in a similar context in Ref. [23] (note also that it is a natural regularization of the truncated exponential kernel used for the probability of nearest-neighbor bonds in Ref. [41]). The kernel (69) has two parameters—temperature of bond formation  $T > 0$  and typical connection radius  $r_m(\mathbf{x}, T)$ —and it converges to the VCR in the limit  $T \rightarrow \infty$  [the normalization condition (67) converges to (6) in this case]. In the model with the Fermi-Dirac kernel the formation of bonds is stochastic, and one expects the reciprocity of the network to decrease with increasing  $T$  without changing of other essential properties of the networks such as the in- and out-degree distributions.

In order to understand how the structure of the network depends on  $T$ , rewrite the equation (67) in the form

$$\int dr \frac{\partial A(r, x, R)}{\partial r} \left\{ 1 + \exp \left[ \frac{r - r_m(\mathbf{x}, T)}{T} \right] \right\}^{-1} = m/v, \quad (70)$$

where  $A(r, x, R)$  is collectively defined by (15)–(20) and study the large- $r$  tail of the integrand in the left-hand side. Importantly, for large radii ( $e^r \gg 1$ ) the area  $A(r, x, R)$  grows exponentially, so that

$$\frac{dA(r, x, R)}{dr} \sim \begin{cases} \exp r & \text{for } r < R - x \\ \exp[(r + R - x)/2] & \text{for } R - x < r < R + x \end{cases} \quad (71)$$

[see (22) and (23)]. In particular, for  $x$  in the center and on the boundary of the disk we get, respectively,

$$dA(r, 0, R)/dr \sim \exp r, \quad dA(r, R, R)/dr \sim \exp(r/2). \quad (72)$$

The large- $r$  behavior of the integrand in (70) is controlled by the competition between the exponential growth of  $dA(r, x, R)/dr$  and the exponentially decaying factor  $\exp[-r/T]$ , coming from the Fermi-Dirac kernel. If the product of these two factors is decaying for large  $r$ , then the typical bond length is close to  $r_m$  as it is in the zero-temperature case (we call this case “the short bonds regime”). In turn, if the product is growing for large  $r$ , then the typical bond length decouples from  $r_m$  and is, in fact, close to the maximal possible value  $R + x$  (we call this “the long bonds regime,”

$r_m$  in this case plays a role of  $R$ -dependent normalization constant).

Thus, depending on temperature, one expects the following three regimes.

If  $T < 1$ , then all bonds are short, with their length controlled by  $r_m(\mathbf{x}, T)$ . For example, for bonds starting in the center of the disk it is easy to estimate the mean bond length by approximating the Fermi-Dirac factor (69) with

$$P(r, r_m, T) = \begin{cases} 1 & \text{for } r < r_m \\ \exp[-(r - r_m)/T] & \text{for } r > r_m, \end{cases} \quad (73)$$

which gives

$$\langle l \rangle \approx r_m - 1 + T/(1 - T). \quad (74)$$

As a result, the structure of the network in this regime is qualitatively similar to that at  $T = 0$ : The network has a core region, where properties (in-degree, reciprocity, etc.) do not “feel” the boundary of the disk, and peripheral region with roughly exponential dependence of the in-degree on the distance to the boundary.

For  $T > 2$  all bonds are long, the nodes predominantly connect to nodes which are very far from them: There are exponentially many such nodes, and Fermi-Dirac decay is not strong enough to prevent formation of the long bonds. As a result, in the thermodynamic limit bonds become completely uncorrelated and the reciprocity converges to zero.

The most interesting situation arises for  $T \in (1, 2)$ . In this case nodes close to the center of the disk predominantly are sources of long bonds (i.e., they typically connect not to their neighbors but to random nodes close to the periphery of the disk), while peripheral nodes are sources of short bonds, connecting them mostly to their neighbors. As a result, network does not have a core-periphery structure anymore. However, since the global properties of the network are mostly controlled by peripheral nodes, the overall in-degree remains a power law.

In Fig. 10 we present the results of numerical simulations of temperature-dependent VCR with the Fermi-Dirac kernel. Figure 10(a) shows the behavior of the connection radius  $r_m(\xi, T)$  for different temperatures and  $m/v$  and fixed  $N = 10000$ . It is seen that for low temperatures (upper curves in the series) the behavior is qualitatively similar to that of zero-temperature VCR: There is clear distinction between the core of the network where  $r_m$  is independent of the coordinate and periphery with linear dependence of  $r_m$  on  $\xi$ . However, the core shrinks with growing temperature and ceases to exist for  $T > 1$ . In the  $T > 1$  regime  $r_m$  actually can become negative. Note, however, that in this regime the length of bonds is close to  $R + x = 2R - \xi$  and  $r_m$  plays just a role of normalization constant in (70).

Figure 10(b) shows the behavior of the average in-degree of a node as a function of coordinate,

$$k_{\text{in}}(\xi) = \int_{O(R)} P(||\mathbf{x} - \mathbf{y}||, r_m(\mathbf{x}, T), T) v d\mathbf{x}, \quad (75)$$

where  $P(r, r_m, T)$  is given by (69) and the integral, although formally a function of the target node position  $\mathbf{y}$  in fact depends only on the distance to the boundary  $\xi = R - y$ . Interestingly, the plateau of the in-degree distribution in the core of the network exists only up to  $T = 0.5$ . However, up

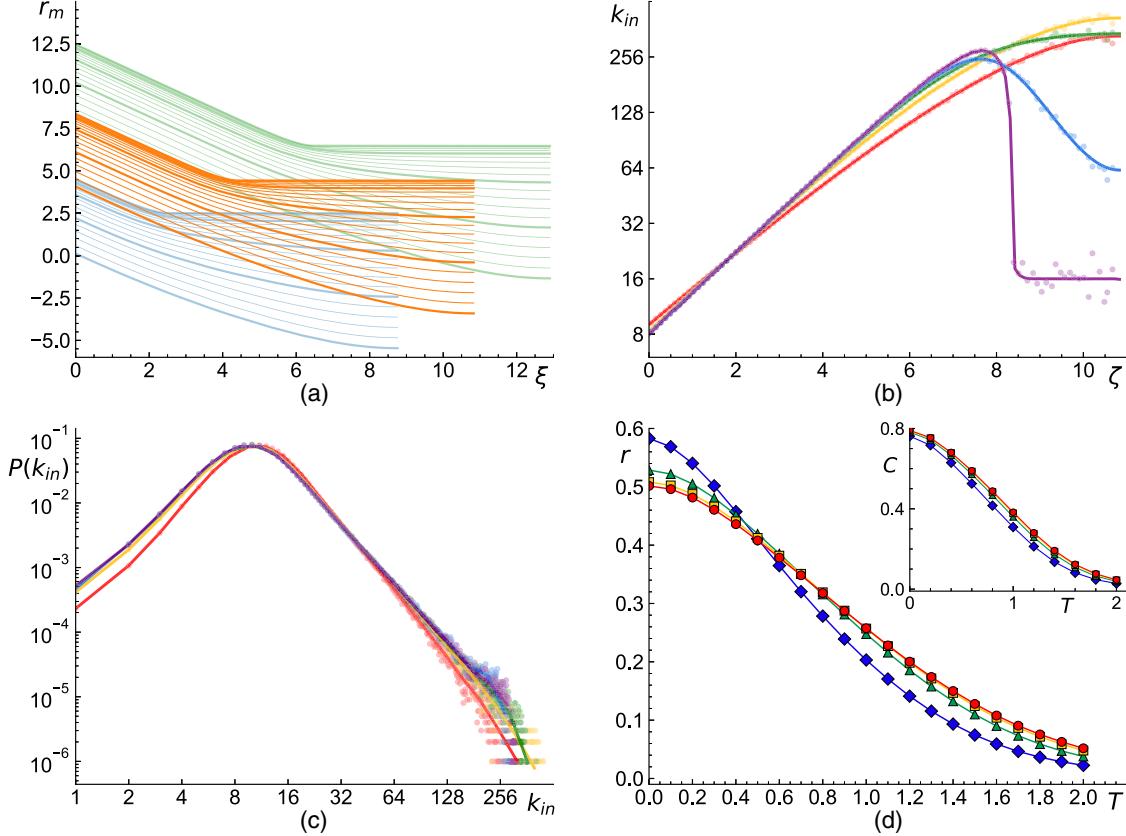


FIG. 10. Structural properties of temperature-dependent VCR networks. (a) The dependence of  $r_m(\xi, T)$  defined by (70) on the distance from the boundary  $\xi$  for networks of  $N = 10\,000$  nodes with  $m = 16$  and (from bottom to top)  $\nu = 1/2$  (blue),  $1/16$  (orange),  $1/128$  (green) (the corresponding disk radii are  $R \approx 8.76, 10.83, 12.90$ , respectively) and various temperatures. In each series of curves temperature grows from  $T = 0$  (top curve) to  $T = 2$  (bottom curve) with step 0.1. (b) The dependence of average in-degree of the node on its distance  $\xi$  from the boundary; dots are results for simulated networks of  $N = 10\,000$  nodes with  $m = 16$ ,  $\nu = 1/2$ ,  $R \approx 10.83$  and temperatures  $T = 0, 0.5, 1, 1.5$  and 2, colored from purple to red in the rainbow order; curves are results of numerical integration of (75), points correspond to average of 100 independent model networks; (c) global in-degree distribution for the same networks as in panel (b); (d) reciprocity and (in the inset) average clustering coefficient as functions of temperature for networks with  $N = 10\,000$ ,  $m = 16$  and  $\nu = 1/2$  (blue diamonds),  $1/8$  (green triangles),  $1/32$  (orange squares), and  $1/128$  (red circles).

to  $T = 1$  the in-degree dependence on  $\zeta$  is nonmonotonic, exponential increase for small  $\zeta$  is followed by decrease at larger  $\zeta$  similarly to the low-temperature limit. For  $T > 1$  the in-degree keeps slowly increasing up to the center of the disk. However, given that the majority of the nodes are located in the peripheral region where exponentially increasing  $k_{in}(\zeta)$  dependence persists for all temperatures  $T \in [0, 2]$ , it is not surprising that the global in-degree distribution remains a power law with exponent  $-3$  [see Fig. 10(c)]. In fact, erosion of the core regime with growing temperature leads to the cutoff of the power law being shifted to higher in-degrees.

Finally, Fig. 10(d) shows the numerical results for dependence of reciprocity and clustering coefficient on temperature for various  $m/\nu$  and  $N = 10\,000$ . As expected, both decrease with increasing temperature. The fact that reciprocity at  $T = 2$  remains small but finite is a finite-size effect: It is easy to show that  $r(T = 2)$  should decrease logarithmically with the size of the network.

To conclude, note that it is possible to generalize the idea of stochastic temperature-controlled bond formation to the case of  $m$ -NN networks in this case the probability of forming a

bond from node  $i$  to node  $j$  can be defined as

$$p_{i \rightarrow j} = \left\{ 1 + \exp \left[ \frac{m_{ij} - \mu(T)}{T} \right] \right\}^{-1}, \quad (76)$$

where  $m_{ij}$  is the order of  $j$  in the list of nearest neighbors of  $i$ , as defined above (38) ( $m_{ij} = 1$  for the nearest neighbor of  $i$ , 2 for the next-nearest neighbor, etc.), and  $\mu(T)$  is defined in a way to make the mean out-degree fixed:

$$\sum_{k=1}^{\infty} \left\{ 1 + \exp \left[ \frac{k - \mu(x, T)}{T} \right] \right\}^{-1} = m \quad (77)$$

or in some other similar way.

We expect the qualitative behavior of such a network to be similar to that of the T-VCR: the reciprocity of the network to decrease with growing temperature while in-degree distribution remaining a truncated power law.

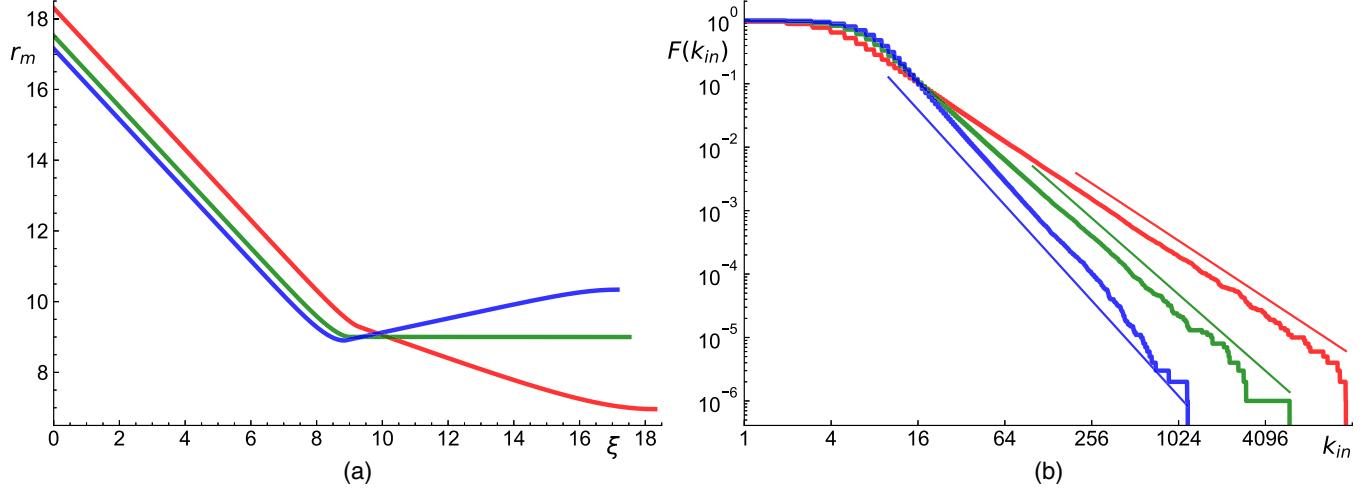


FIG. 11. (a) Connection radius  $r_m$  as a function of the distance  $\xi$  of the source node from the boundary for networks with quasiuniform distribution of nodes (78),  $\alpha = 0.75$  (red, the curve decreases inside the core, i.e., for large  $\xi$ ), 1 (green, uniform case, constant in the core), 1.25 (blue, increases in the core); numerical results for  $m = 10$ ,  $N = 50\,000$  and radius of disk  $R$  defined by equation  $r_m(\xi = 0) = R$ . It is seen that quasiuniform distribution violates the translation invariance in the core but in the periphery the dependence  $r_m(\xi) = r_m(0) - \xi$  still holds; (b) cumulative in-degree distributions of the same networks with (from top to bottom)  $\alpha = 0.75$  (red), 1 (green, uniform case), 1.25 (blue), and the corresponding straight lines indicate the corresponding power-law behavior given by (80).

## V. VARIATION OF THE EXPONENT OF THE IN-DEGREE DISTRIBUTION

The exponent of the in-degrees distribution in the power-law (boundary-dominated) regime is fixed at  $-3$  for both VCR and  $m$ -NN models. Here we briefly discuss how to relax this constraint. To do this, following Ref. [23], we introduce a “quasiuniform” distribution of nodes, so that the density of nodes at distance  $r$  from the origin behaves as

$$\nu(r) = \nu \frac{\sinh \alpha r}{\sinh r} \quad (78)$$

with some constant parameter  $\alpha$ . This distribution of nodes breaks the translation invariance of the properties in the core of the network (e.g., the mean in-degree becomes slightly coordinate dependent) but in the periphery regime the resulting behavior is quite simple: the fraction of nodes with a given distance to the boundary becomes

$$f_\alpha(\xi) d\xi = e^{-\alpha\xi} d\xi \quad (79)$$

instead of (43), which, after substitution into (48) gives

$$P_\alpha(k_{in}) \sim k_{in}^{-1-2\alpha}, \quad F_\alpha(k_{in}) \sim k_{in}^{-2\alpha}, \quad (80)$$

for the pdf and the cumulative degree distribution, respectively (compare Ref. [23]) and one needs  $\alpha \leq 2$  for the whole network to stay in the boundary-dominated regime in the thermodynamic limit.

In Fig. 11 we show the numerical results for the connection radius  $r_m(\xi)$  and for the cumulative in-degree distribution of the VCR networks with various  $\alpha$ , showing that for networks with small-enough node density, despite the violation of the translation invariance in the core, the in-degree distribution indeed does have a power-law tail (80).

## VI. GENERATING NETWORKS WITH GIVEN PROPERTIES USING THE NEAREST-NEIGHBOR MODELS

In this section we briefly discuss how to use network models introduced above to generate directed networks with given properties. Recall that out-degree in the models considered here is fixed by construction,

$$P_{out}(k) = \begin{cases} \delta_{k,m} & \text{for the } m\text{-NN model;} \\ m^k e^{-m} / k! & \text{for the VCR model,} \end{cases} \quad (81)$$

while other properties of the networks, such as the slope and width of the in-degree distribution and the reciprocity can be tuned by adjusting the parameters of the models.

As an example, consider constructing a network whose in-degree has a purely power-law tail with slope  $-3$ . We also demand that the network has a given average degree  $m$ , number of nodes  $N$ , and reciprocity  $r$ . For definiteness, let us construct it based on the VCR model.

In the large- $N$  limit the in-degree distribution of a VCR network is, for any given  $\nu$ , a truncated power law. However, for any final  $N$  it is possible to choose density of nodes in a way that only the power-law part of the distribution is observed. Indeed, the truncation of the power law is caused by the existence of the core part of the network where the expected degree does not depend exponentially on the distance to the boundary. Thus, if one chooses the density of points in a way that the width of the boundary region becomes equal to the radius of the disk, then the truncated part of the distribution ceases to exist. Using (12) for the area of hyperbolic circle and (24) for the connection radius at the boundary  $r_m(0)$

one gets

$$\begin{aligned} N &= 2\pi v(\cosh R - 1) \approx \pi v e^R \\ r_m(0) &= 2 \ln \left[ \frac{m}{4v} + \frac{\pi}{2} \right] \\ r_m(0) &= R \end{aligned} \quad \rightarrow \quad v_0 \approx \frac{\pi}{16} \frac{m^2}{N}, \quad R_0 \approx 2 \ln \left( \frac{4N}{\pi m} \right), \quad (82)$$

where  $N/m \gg 1$  is assumed. Equation (82) gives the value of  $R_0$  ( $v_0$ ) for which the in-degree distribution of the resulting networks is the closest to a true power law. If  $R < R_0$  is chosen (which means that  $v$  is larger than  $v_0$ ), then the network has a core which is unreachable from the boundary and has bulk properties, including position-independent average in-degree  $m$ , while if  $R > R_0$  is chosen, then an “inverse core” of radius  $r_m(0) - R$  is formed in the center of the network, the nodes in this inverse core get connected to all peripheral nodes, attaining in-degree of order  $N$ . Note, however, that if radius is close to  $R_0$ , then the actual number of nodes in core (reverse core) is small [it is of order  $N \exp(-r_m(0))$ ] for the core and  $N \exp[-2R + r_m(0)]$  for the reverse core).

The second step of the network construction is to distribute  $N$  points at random in the hyperbolic disk of the chosen radius  $R_0$  and calculate the distances between them. It can be done in a standard way: The coordinates  $(\rho_i, \theta_i)$  of the nodes ( $i = 1, \dots, N$ ) are chosen independently at random from probability distributions,

$$\begin{aligned} p_{\text{rad}}(\rho) &= \sinh \rho / (\cosh R_0 - 1) \quad \rho \in [0, R_0] \\ p_{\text{ang}}(\theta) &= 1/2\pi \quad \theta \in [0, 2\pi]. \end{aligned} \quad (83)$$

The distance  $d_{ij}$  between each pair of points can be calculated according to the hyperbolic cosine theorem (14) as

$$d_{ij} = \text{arccosh}[\cosh \rho_i \cosh \rho_j - \sinh \rho_i \sinh \rho_j \cos(\theta_i - \theta_j)]. \quad (84)$$

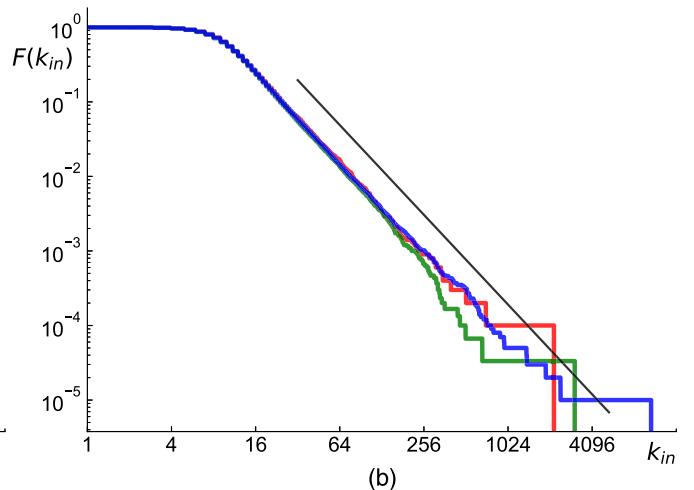
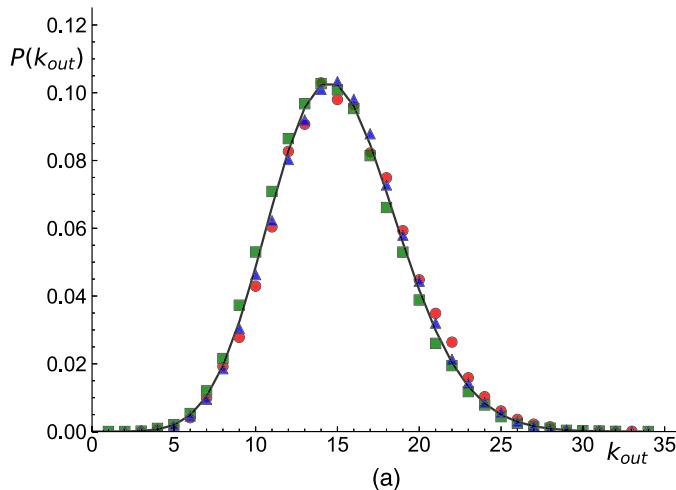


FIG. 12. (a) Out-degree distribution of models with  $m = 15$ ,  $r = 0.2$ , and  $N = 10^4$  (red circles),  $3 \times 10^4$  (green squares), and  $10^5$  (blue triangles) compared with the corresponding Poisson distribution (solid line); (b) cumulative in-degree distribution for the same networks, the straight line has a slope  $-2$  and is a guide for the eye.

is formed independently at random with probability given by (69).

We have evaluated this procedure for 27 different combinations of parameters  $N = 10^4, 3 \times 10^4, 10^5; m = 5, 15, 40$ ; and  $r = 0.1, 0.2, 0.4$ . The comparison between the seed values of the parameters and the corresponding parameters in the constructed networks are provided in the Appendix. The results are mostly satisfactory: parameters of the constructed models are within 1–2% of the desired seed values. Two minor problems are (i) notable discrepancy between constructed and seed value of reciprocity for small value of seed  $r$  and (ii) a small systematic error in the average degree. Both effects are especially notable for smaller  $N$  and seem to decline with growing  $N$ . The discrepancy in  $r$  is due to the aforementioned nonuniversal behavior of the  $r(T)$  curve for large  $T$  and can be corrected by iteratively changing the value of temperature used in simulations. The discrepancy in the average degree is due to the fact replacement of the true behavior of  $r_m(\xi)$  with a piecewise linear approximation (85) and, if needed, can be alleviated by using a more sophisticated approximation for  $r_m(\xi)$ , e.g.,

$$\begin{aligned} r_m(\xi) &= r_m(R) + [|X| - X + a \exp(-|X|/a)]/2, \\ X &= \xi - r_m(0) - r_m(R) \end{aligned} \quad (87)$$

with a properly adjusted numerical parameter  $a$  [note that for  $a \rightarrow 0$  (87) converges to (85)]. Moreover, in Fig. 12 we show the resulting in- and out-degree distributions for several of the networks proving that out- and in-degree distributions of the generated networks are indeed Poisson and power law with exponent 3, respectively.

## VII. DISCUSSION

In this paper we studied the properties of the nearest-neighbor model and a dual model with varied connection radius on a hyperbolic disk. We have shown that the properties of these networks change dramatically with the change of dimensionless (per unit of inverse curvature squared) density of nodes.

If density of nodes is large (effective space curvature is small), then the resulting networks are similar to the nearest-neighbor networks in Euclidean space: both out- and in-degree are narrowly distributed around their mean value  $m$  and the reciprocity of the network is very large, between  $r = (5/3 + \sqrt{3}/\pi)^{-1} \approx 0.4509$  for 1-NN networks and  $r = 1$  for the VCR networks, which correspond to the  $m \rightarrow \infty$  limit of  $m$ -NN.

In turn, if density of nodes is small (conversely, space curvature is large), then the networks can be thought of as directed analogs of undirected hyperbolic networks studied in Refs. [23,24]. Let us point out the important generality of this result: Although  $m$ -NN and VCR networks are 3-parametric ( $m, v, R$ ) classes of networks, in the limit of large network size their behavior is remarkably universal. Indeed, the in-degree distribution is always a power law with exponent  $-3$ , for  $k_{\text{in}}/m$  from 1 to  $m/v$ , i.e., the only relevant parameter is  $m/v$ , which controls the position of the cutoff. Reciprocities of large networks converge to finite values for  $v \rightarrow 0$ . For VCR this value is  $m$  independent, for  $m$ -NN it weakly depends on  $m$  and converges to the VCR value for  $m \rightarrow \infty$ . In Table I we summarize the properties of all the networks studied in this paper in the limit  $N \rightarrow \infty$  and for large but finite  $m/v$ .

TABLE I. Summary of the behavior of the model studied in this paper in the limit of large network size ( $N \sim ve^R \rightarrow \infty$ ).

Model	m-NN	VCR	Temperature-dependent VCR	m-NN/VCR with quasi-uniform distribution
Parameters	$m, v, R$	$m, v, R$	$m, v, R, T$	$m, v, R, \alpha$
out-degree distribution	delta-functional	Poisson	Poisson	delta/Poisson
in-degree exponent	$-3$	$-3$	$-3$	$-1-2\alpha$
in-degree cutoff	$\sim m/v$	$\sim m/v$	$\sim m/v$	?
$k_{\text{max}}/m$	of order 1 all $m/v$ ,	of order 1 for all $m/v$ ,	decreases with $T$ from a value of order	
reciprocity	converges to finite limit less than $1/2$ for $m/v \rightarrow \infty$	converges to $1/2$ for $m/v \rightarrow \infty$	$1$ for $T = 0$ to that of order $1/R$ for $T = 2$	?
Average clustering coefficient	of order 1 for all $m/v$ , converges to a finite limit for $m/v \rightarrow \infty$	of order 1 for all $m/v$ , converges to a finite limit for $m/v \rightarrow \infty$	decreases with $T$ from a value of order $1$ for $T = 0$ to that of order $1/R$ for $T = 2$	?

TABLE II. Numerically calculated values of reciprocity  $r$  for various values of temperatures  $T$  ( $N = 50\,000$ ).

$T$	$r$
0.0	0.50133
0.1	0.49449
0.2	0.48011
0.3	0.45954
0.4	0.43477
0.5	0.40725
0.6	0.37806
0.7	0.34797
0.8	0.31756
0.9	0.28723
1.0	0.25728
1.1	0.22797
1.2	0.19955
1.3	0.17234
1.4	0.14669
1.5	0.12297
1.6	0.10154
1.7	0.08262
1.8	0.06632
1.9	0.05257
2.0	0.04122

This behavior comes from the peculiar core-periphery structure of the network, where in the core network properties are translation invariant and qualitatively similar to those of Euclidean nearest-neighbor networks, while in the boundary region the degree and local reciprocity are strongly position dependent, which, after integration over the spatial distribution of nodes gives rise to the scale-free behavior. In the thermodynamic limit  $R \rightarrow \infty$ ,  $v = \text{const}$  both core and periphery regions are present for all  $m$  and  $v$ . However, distribution of the nodes between two regions depends on  $m/v$  giving rise to core-dominated behavior for large densities and boundary-dominated behavior for small densities.

We believe that the novel class of networks studied here will be useful as a reference model for scale-free directed networks. Notably, another model of directed hyperbolic networks was proposed recently in Ref. [31]. There each node attains two intrinsic popularities (akin to radial coordinates in the hyperbolic disk representation), with respect to in- and out-degree, respectively. By regulating the distributions of these popularities and the correlations between them the authors of Ref. [31] are able to generate directed networks with arbitrary in- and out-degree distributions and arbitrary correlation between them. In turn, our model is less suitable to fitting arbitrary degree distributions: indeed, out-degree distribution here is fixed. However, it is more naturally geometric, in a sense that it is a natural generalization of well-studied Euclidean  $m$ -nearest-neighbor graphs [33–38], and it might be a more natural fit to model the cases when one of the two distributions is fixed to be delta-functional or Poisson. One possible way of using these models is to study embedding of directed networks in hyperbolic space. The analogous task for the case of undirected networks is well developed [52–54] but

TABLE III. The results of generating of VCR networks with predefined properties.  $N$ ,  $m$ , and  $r$  are the target values of the number of nodes, average out-(in-)degree and reciprocity, respectively.  $R$  and  $T$  are the values of disk radius and temperature used for simulations.  $\Delta m/m$  and  $\Delta r/r$  and relative discrepancies between target values of  $m$  and  $r$  and their observed values in simulated networks.

$N$	$m$	$r$	$R$	$T$	$\Delta m/m$	$\Delta r/r$
10 000	5	0.1	15.685	1.612	-0.013	-0.011
10 000	5	0.2	15.685	1.198	-0.012	0.020
10 000	5	0.4	15.685	0.524	-0.002	0.006
10 000	15	0.1	13.488	1.612	-0.039	0.051
10 000	15	0.2	13.488	1.198	0.019	-0.013
10 000	15	0.4	13.488	0.524	-0.012	0.002
10 000	40	0.1	11.526	1.612	-0.026	0.101
10 000	40	0.2	11.526	1.198	-0.022	0.020
10 000	40	0.4	11.526	0.524	-0.017	0.022
30 000	5	0.1	17.882	1.612	0.011	-0.024
30 000	5	0.2	17.882	1.198	0.003	0.002
30 000	5	0.4	17.882	0.524	0.014	-0.024
30 000	15	0.1	15.685	1.612	-0.015	0.033
30 000	15	0.2	15.685	1.198	-0.015	0.008
30 000	15	0.4	15.685	0.524	-0.005	0.016
30 000	40	0.1	13.723	1.612	-0.014	0.046
30 000	40	0.2	13.723	1.198	0.010	-0.002
30 000	40	0.4	13.723	0.524	-0.005	0.004
100 000	5	0.1	20.290	1.612	-0.012	-0.037
100 000	5	0.2	20.290	1.198	-0.003	0.007
100 000	5	0.4	20.290	0.524	0.003	-0.003
100 000	15	0.1	18.093	1.612	-0.013	-0.008
100 000	15	0.2	18.093	1.198	0.010	-0.013
100 000	15	0.4	18.093	0.524	-0.005	0.001
100 000	40	0.1	16.131	1.612	-0.004	0.002
100 000	40	0.2	16.131	1.198	-0.010	0.009
100 000	40	0.4	16.131	0.524	-0.003	0.005

generalization to the directed case is in its infancy (although note the recent paper [32]). Using the discrepancy between embedded directed networks and  $m$ -NN networks as a loss function characterizing the quality of the embedding seems to be a very exciting perspective, although it goes beyond the scope of this paper.

## ACKNOWLEDGMENTS

The authors are grateful to M.Á. Serrano, P. Krapivsky, S. Nechaev, K. Polovnikov, and M. Schich for stimulating discussions. This work was partially supported by CUDAN ERA Chair project (Grant No. 810961 of the EU Horizon 2020 program) and NSF Grant No. IIS-1741355.

## APPENDIX

In this Appendix we provide reference data to be used for the generation of directed networks with given characteristics. In Table II we tabulate the dependence of the network reciprocity on temperature, i.e., the master curve to which curves on Fig. 10(d) converge. Table III summarizes the results of our numerical generation of 27 networks with target values of parameters  $N = 10^4, 3 \times 10^4, 10^5$ ;  $m = 5, 15, 40$ ; and

$r = 0.1, 0.2, 0.4$ . Here  $R$  and  $T$  are the values of the disk radius and temperature calculated from (82) and from Table II, respectively, while  $\Delta m/m$  and  $\Delta r/r$  are the relative discrepancies between generated and target values of average degree

and reciprocity:

$$\begin{aligned}\Delta m/m &= (m_{\text{gen}} - m_{\text{target}})/m_{\text{target}}; \\ \Delta r/r &= (r_{\text{gen}} - r_{\text{target}})/r_{\text{target}}.\end{aligned}\quad (\text{A1})$$

- 
- [1] J. Leskovec, L. A. Adamic, and B. A. Huberman, The dynamics of viral marketing, *ACM Trans. Web* **1**, 5 (2007).
- [2] G. Oestreicher-Singer and A. Sundararajan, Recommendation networks and the long tail of electronic commerce, *MIS Quart.* **36**, 65 (2012).
- [3] G. Kiss, C. Armstrong, R. Milroy, and J. Piper, An associative thesaurus of English and its computer analysis, in *The Computer and Literary Studies*, edited by A. J. Aitken, R. W. Bailey, and N. Hamilton-Smith (Edinburgh University Press, Edinburgh, UK, 1973).
- [4] D. L. Nelson, C. L. McEvoy, and T. A. Schreiber, The University of South Florida free association, rhyme, and word fragment norms, *Behav. Res. Methods Instrum. Comput.* **36**, 402 (2004).
- [5] S. De Deyne, D. J. Navarro, A. Perfors, M. Brysbaert, and G. Storms, The “Small World of Words” English word association norms for over 12,000 cue words, *Behav. Res. Methods* **51**, 987 (2019).
- [6] O. Valba, A. Gorsky, S. Nechaev, and M. Tamm, Analysis of English free association network reveals mechanisms of efficient solution of Remote Association Tests, *PLoS ONE* **16**, e0248986 (2021).
- [7] S. Viteri and S. DeDeo, Epistemic phase transitions in mathematical proofs, *Cognition* **225**, 105120 (2022).
- [8] <http://konect.cc/networks>, datasets for DBpedia, English wikipedia, zhishi.me.
- [9] S. N. Dorogovtsev and J. F. F. Mendes, *Evolution of Networks: From Biological Nets to the Internet and WWW* (Oxford University Press, Oxford, 2003).
- [10] M. O. Jackson, *Social and Economic Networks* (Princeton University Press, Princeton, NJ, 2008).
- [11] M. E. J. Newman, *Networks* (Oxford University Press, Oxford, 2018).
- [12] P. L. Krapivsky, S. Redner, and E. Ben-Naim, *A Kinetic View of Statistical Physics* (Cambridge University Press, Cambridge, UK, 2010).
- [13] A.-L. Barabási, *Network Science* (Cambridge University Press, Cambridge, UK, 2016).
- [14] M. Barthelemy, *Spatial Networks: A Complete Introduction* (Springer, Berlin, 2022).
- [15] P. J. Flory, Molecular size distribution in three dimensional polymers, *J. Am. Chem. Soc.* **63**, 3083 (1941).
- [16] P. Erdős and A. Rényi, On the evolution of random graphs, *Publ. Math. Inst. Hung. Acad. Sci.* **5**, 17 (1960).
- [17] D. J. Watts and S. H. Strogatz, Collective dynamics of “small-world” networks, *Nature (London)* **393**, 440 (1998).
- [18] A.-L. Barabási and R. Albert, Emergence of scaling in random networks, *Science* **286**, 509 (1999).
- [19] P. L. Krapivsky, S. Redner, and F. Leyvraz, Connectivity of growing random networks, *Phys. Rev. Lett.* **85**, 4629 (2000).
- [20] M. E. J. Newman, S. H. Strogatz, and D. J. Watts, Random graphs with arbitrary degree distributions and their applications, *Phys. Rev. E* **64**, 026118 (2001).
- [21] J. Park and M. E. J. Newman, The statistical mechanics of networks, *Phys. Rev. E* **70**, 066117 (2004).
- [22] M. Á. Serrano, D. Krioukov, and M. Boguná, Self-similarity of complex networks and hidden metric spaces, *Phys. Rev. Lett.* **100**, 078701 (2008).
- [23] D. Krioukov, F. Papadopoulos, M. Kitsak, A. Vahdat, and M. Boguná, Hyperbolic geometry of complex networks, *Phys. Rev. E* **82**, 036106 (2010).
- [24] F. Papadopoulos, M. Kitsak, M. Á. Serrano, M. Boguná, and D. Krioukov, Popularity versus similarity in growing networks, *Nature (London)* **489**, 537 (2012).
- [25] J. S. Andrade, H. J. Herrmann, R. F. S. Andrade, and L. da Silva, Apollonian networks: Simultaneously scale-free, small world, Euclidean, space filling, and with matching graphs, *Phys. Rev. Lett.* **94**, 018702 (2005).
- [26] T. Zhou, G. Yan, and B.-H. Wang, Maximal planar networks with large clustering coefficient and power-law degree distribution, *Phys. Rev. E* **71**, 046141 (2005).
- [27] Z. Zhang, F. Comellas, G. Fertin, and L. Rong, High-dimensional Apollonian networks, *J. Phys. A* **39**, 1811 (2006).
- [28] Z. Zhang, L. Rong, and F. Comellas, High-dimensional random Apollonian networks, *Physica A* **364**, 610 (2006).
- [29] G. Bianconi and R. M. Ziff, Topological percolation on hyperbolic simplicial complexes, *Phys. Rev. E* **98**, 052308 (2018).
- [30] M. V. Tamm, D. G. Koval, and V. I. Stadnichuk, Polygon-based hierarchical planar networks based on generalized Apollonian construction, *Physics* **3**, 998 (2021).
- [31] A. Allard, M. Á. Serrano, and M. Boguñá, Geometric description of clustering in directed networks, *Nat. Phys.* (2023), doi:10.1038/s41567-023-02246-6.
- [32] B. Kovács and G. Palla, Model-independent embedding of directed networks into Euclidean and hyperbolic spaces, *Commun. Phys.* **6**, 28 (2023).
- [33] G. L. Miller, S. H. Teng, and S. A. Vavasis, An unified geometric approach to graph separators, in *IEEE 32nd Annual Symposium on Foundations of Computer Science* (IEEE, Los Alamitos, CA, 1991), p. 538.
- [34] O. Häggström and R. Meester, Nearest neighbor and hard sphere models in continuum percolation, *Rand. Struct. Algor.* **9**, 295 (1996).
- [35] P. Balister, B. Bollobás, A. Sarkar, and M. Walters, Connectivity of random  $k$ -nearest neighbour graphs, *Adv. Appl. Probab.* **37**, 1 (2005).
- [36] S. Teng and F. Yao,  $k$ -nearest-neighbor clustering and percolation theory, *Algorithmica* **49**, 192 (2007).
- [37] R. Balister, B. Bollobás, and A. Sarkar, Percolation, Connectivity, coverage and coloring, in *Handbook of Large Scale Random Networks*, edited by B. Bollobás, R. Kozma, and

- D. Miklos (János Bolyai Mathematical Society and Springer, Budapest, 2008).
- [38] M. Walters, Random geometric graphs, in *Surveys in Combinatorics*, edited by R. Chapman (London Mathematical Society Lecture Notes Series, London, 2011), Vol. 392, p. 365.
- [39] R. R. Coifman, S. A. Lafon, B. Lee, M. Maggioni, B. Nadler, F. Warner, and F. B. Zucker, Geometric diffusions as a tool for harmonic analysis and structure definition of data: Diffusion maps, *Proc. Natl. Acad. Sci. USA* **102**, 7426 (2005); R. R. Coifman and S. Lafon, Diffusion maps, *Appl. Comput. Harmon. Anal.* **21**, 5 (2006).
- [40] L. Van der Maaten and G. Hinton, Visualizing data using t-SNE, *J. Mach. Learn. Res.* **9**, 2579 (2008).
- [41] L. McInnes, J. Healy, and J. Melville, UMAP: Uniform manifold approximation and projection for dimension reduction, [arXiv:1802.03426](https://arxiv.org/abs/1802.03426).
- [42] J. W. Cannon, W. J. Floyd, R. Kenyon, and W. R. Parry, Hyperbolic geometry, in *Flavors of geometry*, edited by S. Levy (Cambridge University Press, Cambridge, UK, 1997).
- [43] R. Tao and F. Y. Wu, The vicious neighbour problem, *J. Phys. A* **20**, L299 (1987).
- [44] P. Gracar, A. Grauer, L. Lüchtrath, and P. Mörters, The age-dependent random connection model, *Que. Syst.* **93**, 309 (2019).
- [45] D. Garlaschelli and M. I. Loffredo, Patterns of link reciprocity in directed networks, *Phys. Rev. Lett.* **93**, 268701 (2004).
- [46] M. E. J. Newman, S. Forrest, and J. Balthrop, Email networks and the spread of computer viruses, *Phys. Rev. E* **66**, 035101(R) (2002).
- [47] R. Milo, S. Shen-Orr, S. Itzkovitz, N. Kashtan, D. Chklovskii, and U. Alon, Network motifs: Simple building blocks of complex networks, *Science* **298**, 824 (2002).
- [48] R. Milo, S. Itzkovitz, N. Kashtan, R. Levitt, S. Shen-Orr *et al.*, Superfamilies of evolved and designed networks, *Science* **303**, 1538 (2004).
- [49] D. J. De Solla Price, Networks of scientific papers, *Science* **149**, 510 (1965).
- [50] I. M. Lifshitz, A. Y. Grosberg, and A. R. Khokhlov, Some problems of the statistical physics of polymer chains with volume interaction, *Rev. Mod. Phys.* **50**, 683 (1978).
- [51] I. Erkhimovich, M. V. Thamm, and A. V. Ermoshkin, Theory of the sol-gel transition in thermoreversible gels with due regard for the fundamental role of mesoscopic cyclization effects, *Macromolecules* **34**, 5653 (2001).
- [52] M. Nickel and D. Kiela, Poincaré embeddings for learning hierarchical representations, in *Advances in Neural Information Processing Systems (NeurIPS)* (2017), pp. 6338–6347.
- [53] F. Sala, C. De Sa, A. Gu, and C. Ré, Representation tradeoffs for hyperbolic embeddings, in *International Conference on Machine Learning* (2018), pp. 4457–4466.
- [54] A. Tifrea, G. Bécigneul, and O.-E. Ganea, Poincaré GloVe Hyperbolic word embeddings, in *International Conference on Learning Representations* (2019).