nature physics

Article

https://doi.org/10.1038/s41567-023-02246-6

Geometric description of clustering in directed networks

Received: 27 March 2023

Accepted: 13 September 2023

Published online: 2 November 2023

Check for updates

Antoine Allard [1,2] ✉, M. Ángeles Serrano [3,4,5] & Marián Boguñá [3,4]

First-principle network models are crucial to understanding the intricate topology of real complex networks. Although modelling efforts have been quite successful in undirected networks, generative models for networks with asymmetric interactions are still not well developed and unable to reproduce several basic topological properties. Progress in this direction is of particular interest, as real directed networks are the norm rather than the exception in many natural and human-made complex systems. Here we show how the network geometry paradigm can be extended to the case of directed networks. We define a maximum entropy ensemble of random geometric directed graphs with a given sequence of in-degrees and out-degrees. Beyond these local properties, the ensemble requires only two additional parameters to fix the levels of reciprocity and the frequency of the seven possible types of three-node cycles in directed networks. A systematic comparison with several representative empirical datasets shows that fixing the level of reciprocity alongside the coupling with an underlying geometry is able to reproduce the wide diversity of clustering patterns observed in real directed complex networks.

The network geometry paradigm is a comprehensive framework that successfully explains the topology, the multiscale organization and the navigability of real complex networks[1]. This framework consists of a handful of simple models and has been shown to accurately model several features observed in static, growing, weighted or multilayer networks[2–8]. The hallmark of network geometry is how it naturally reproduces the clustering patterns observed in real complex networks as one of their most fundamental properties[9]. Clustering is indeed notoriously difficult to model because triangles imply three-node interactions, and most existing approaches must rely on approximations (such as an underlying tree-like organization[10–15]), give up sparsity[16] or turn to numerical simulations[17–19].

Network geometry overcomes this difficulty by assuming that nodes are embedded in a metric space and that the probability p_{ij} that a link exists between nodes i and j is a decreasing function of the distance between them. Non-fortuitous clustering—clustering that does not occur by sheer luck—can therefore be seen as the topological counterpart of the triangle inequality of the metric space: if nodes j and l are both close to node i, then they must also be close to each other. Hence, a triangle composed of nodes i, j and l is likely, even in the limit of very large networks. In fact, network geometry interprets the clustering coefficient as a measure of the coupling between the topology of the network and an underlying latent metric space.

To date, however, network geometry has only been fully developed for complex networks with symmetric interactions, weighted or unweighted. Yet, a large number of real complex networked systems contain a mixture of symmetric and asymmetric interactions (for example, connectomes, food webs and communication networks)[9,20,21]. In addition to the ubiquity of asymmetry, such systems are relevant because they represent processes out of equilibrium where detailed balance is not fulfilled. These systems are also typically non-normal[20] and display trophic coherence[22] (or lack thereof); these features have a drastic impact on the systems' dynamics that cannot be foreseen

[1]Département de physique, de génie physique et d'optique, Université Laval, Québec, Québec, Canada. [2]Centre interdisciplinaire en modélisation mathématique, Université Laval, Québec, Québec, Canada. [3]Departament de Física de la Matèria Condensada, Universitat de Barcelona, Barcelona, Spain. [4]Universitat de Barcelona Institute of Complex Systems (UBICS), Universitat de Barcelona, Barcelona, Spain. [5]Institució Catalana de Recerca i Estudis Avançats (ICREA), Barcelona, Spain. ✉e-mail: antoine.allard@phy.ulaval.ca

https://doi.org/10.1038/s41567-023-02246-6

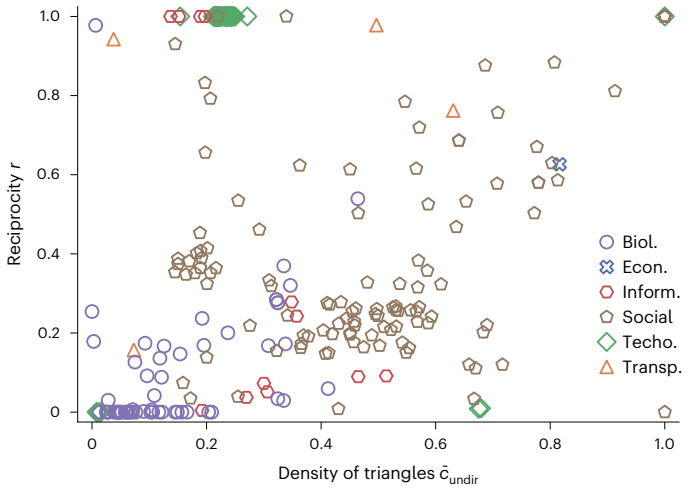

Fig. 1 | Reciprocity in real directed networks. Reciprocity versus density of triangles in 292 real directed biological (Biol.), economical (Econ.), information (Inform.), social, technological (Techno.) and transportation (Transp.) networks. The reciprocity is defined as $r = L^{\leftrightarrow}/L$, where $L^{\leftrightarrow}$ is the number of reciprocal links and L is the number of links. The density of triangles is computed as the average local clustering coefficient of the undirected projection of the original directed network $\bar{c}_{undir}$ (Methods). Details about the network datasets are provided in the Methods.

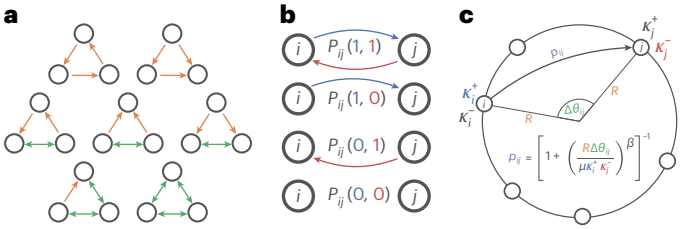

Fig. 2 | Illustrations of the concepts behind the modelling framework. **a**, The seven configurations of triangles in directed networks[45,47]. **b**, The joint probabilities $P_{ij}(a_{ij}, a_{ji})$ used in the general framework controlling reciprocity in random directed networks. **c**, The geometric directed soft configuration model where p_{ij} denotes the probability of connection $P(a_{ij} = 1 \,|\, \kappa_i^+, \kappa_j^-, \Delta\theta_{ij})$ of equation (1a).

if the directionality of the interactions is simply neglected[20,21,23–29]. Although extensions have recently been explored[30–34], the apparent contradiction between the symmetry of metric distances and asymmetric interactions has kept this important class of systems out of the reach of the network geometry framework.

In this Article we propose a simple solution to this impasse. By rethinking the relationship between distance and connection, we introduce the directed-reciprocal $\mathbb{S}^1$ model, a general and versatile adaptation of the framework of network geometry that reconciles the intrinsic symmetry of metric distances with asymmetric interactions between nodes in directed networks. Our model is able to reproduce the joint distribution of both in-degrees and out-degrees, and the model has an additional parameter that tunes the level of reciprocity, that is, the propensity for the two different directed links to exist between the same pair of nodes, a fundamental property of real directed networks[35,36] (Fig. 1). Our model is also amenable to several analytical and semi-analytical calculations. We also provide a more general probabilistic formulation of our framework that can be adapted to control the level of reciprocity in any non-geometric model as long as it defines pairwise connection probabilities.

Most importantly, we use the directed-reciprocal $\mathbb{S}^1$ model to show that even the more complex patterns of clustering in directed networks—quantified by the relative occurrences of the seven triangle configurations possible with directed links or triangle spectrum (Fig. 2a)—are in fact a byproduct of the joint distribution of in-degree and out-degree, of reciprocity and of the triangle inequality in the underlying metric space. Our contribution offers a rigorous path to extend network geometry to directed networks, thus allowing this powerful approach to be used to study real complex systems where asymmetric interactions are crucial, like the brain, food webs, information networks and human interactions.

The directed-reciprocal $\mathbb{S}^1$ model consists in combining two original frameworks: a geometric framework that controls clustering and a probabilistic non-geometric framework that controls reciprocity. In what follows, we first introduce these two frameworks individually, we then combine them to form the directed-reciprocal $\mathbb{S}^1$ model, which we finally use to model real directed complex networks.

Results

The directed $\mathbb{S}^1$ model

We introduce a generalization of the $\mathbb{S}^1$ model[2] to directed networks; this directed $\mathbb{S}^1$ model generates networks with nontrivial levels of clustering, even in the limit $N \to \infty$. However, note that this extension to directed networks generates reciprocal links only by chance. In subsequent sections, we introduce the general framework that allows for control of the level of reciprocity and then combine the two approaches to propose the definitive formulation of the directed-reciprocal $\mathbb{S}^1$ model.

The ensemble of random directed networks defined by the directed $\mathbb{S}^1$ model consists of N nodes positioned on a circle of radius $R = N/2\pi$ (thus setting the density of nodes to 1 without loss of generality). Each node i is independently and identically assigned an angular position θ_i and a pair of hidden degrees κ_i^- and κ_i^+ which, as shown below, are related to each node's in-degree and out-degree, respectively. The angular positions are random variables distributed according to the uniform probability density function (pdf) $\varphi(\theta) = \frac{1}{2\pi}$, although other densities—for instance, to include community structure[37–39]—could be considered. The hidden degrees are random variables distributed according to the joint pdf $\rho(\kappa^-, \kappa^+)$, whose exact form is a free parameter of the model. The only constraint imposed on $\rho(\kappa^-, \kappa^+)$ is that its two first moments coincide, that is, $\langle\kappa^-\rangle = \langle\kappa^+\rangle \equiv \langle\kappa\rangle$. This constraint ensures that $\langle k^-\rangle = \langle k^+\rangle$, which must be true for any directed network. Note that we can also consider another formulation of the model in which the angular positions and hidden degrees are fixed—they become parameters of the model—instead of being random variables with a specified pdf. This formulation is convenient when adjusting the model to real network datasets (see 'Modelling real directed complex networks') and facilitates various analytical calculations (Supplementary Information Section II).

A directed link exists from node i to node j with probability

$$P\left(a_{ij} = 1 \,|\, \kappa_i^+, \kappa_j^-, \Delta\theta_{ij}\right) = \frac{1}{1 + \chi_{ij}^{\beta}} \tag{1a}$$

with

$$\chi_{ij} = \frac{R\Delta\theta_{ij}}{\mu\kappa_i^+\kappa_j^-} = \frac{N\Delta\theta_{ij}}{2\pi\mu\kappa_i^+\kappa_j^-}, \tag{1b}$$

where $\Delta\theta_{ij} = \Delta\theta_{ji} = \pi - |\pi - |\theta_i - \theta_j||$ is the minimal angular distance between nodes i and j and $\mu = \frac{\beta}{2\pi\langle\kappa\rangle}\sin(\frac{\pi}{\beta})$ with $\beta > 1$ is a parameter of the model that controls clustering, as we explain below. Two links therefore exist independently from one another, or their existence may be conditionally independent if they have a node in

common (Supplementary Information Section II.A gives a complete discussion). Figure 2c provides an illustration of the model.

The choice of equation (1a) has two advantages. First, fixing the hidden degrees κ^- and κ^+ allows specifying the expected in-degree and out-degree of each node and thus the expected joint in- and out-degree distribution. As shown in Supplementary Information Sections II.B and II.C, the expected in-degrees and out-degrees of nodes with hidden variables κ_i^- and κ_i^+ are simply given by

$$\left\langle k_i^- \mid \kappa_i^- \right\rangle \simeq \kappa_i^- \quad \text{and} \quad \left\langle k_i^+ \mid \kappa_i^+ \right\rangle \simeq \kappa_i^+. \tag{2}$$

Second, it casts the ensemble of random networks generated by the model into a hyper-grand-canonical ensemble, which is a prime candidate to be the unbiased maximum entropy spatial network model for sparse heterogeneous small worlds with nonzero clustering[5]. The generalization of the $\mathbb{S}^1$ model presented here recovers the directed soft configuration model in the limit $\beta \to 0$ (Methods) but unlike its non-geometric counterpart, it has a nonvanishing clustering in the limit $N \to \infty$ (due to the triangle inequality of its embedding space). As in the undirected $\mathbb{S}^1$ model, clustering in this generalization is tuned using the parameter β; the limit $\beta \to \infty$ yielding the highest density of triangles, while clustering goes to zero when $\beta = 1$. The detailed derivation of these results as well as their validation using numerical simulations are provided in Supplementary Information Section II.

We set clustering aside to introduce a second framework that generates directed networks with a given level of reciprocity.

Reciprocity in random directed networks

We introduce a general framework to control the level of reciprocity in any random directed network models with pairwise connection probabilities. Let p_{ij} be the probability for a directed link to exist from node i to node j and N be the number of nodes. The assumption that interactions are pairwise implies that the existence of links between two different pairs of nodes, i, j and k, l, are statistically independent events. If this condition also applies to the two possible links between the same pair of nodes i, j, then the probability to have a reciprocal link is simply $p_{ij}p_{ji}$. Defining these pairwise probabilities generates a certain level of reciprocity in the network, although it is not possible to tune it.

To gain control over reciprocity, we must relax the assumption of independence of p_{ij} and p_{ji} in a pair of nodes. Thus, similarly to the seminal dyad independence model[40], our framework focuses on the four ways two nodes may or may not be connected (Fig. 2b). We define the joint probabilities $P_{ij}(a_{ij}, a_{ji})$ with $1 \le i < j \le N$ and where a_{ij} is 1 if there is a directed link from node i to node j and 0 otherwise. For our framework to be coherent with the model defining the pairwise connection probabilities, we impose that the joint probability $P_{ij}(a_{ij}, a_{ji})$ preserves the marginal connection probabilities so that

$$P_{ij}(1,0) + P_{ij}(1,1) = p_{ij}, \tag{3a}$$

$$P_{ij}(0,1) + P_{ij}(1,1) = p_{ji} \tag{3b}$$

and we assume that they are normalized, that is,

$$\sum_{a_{ij}=0}^{1} \sum_{a_{ji}=0}^{1} P_{ij}(a_{ij}, a_{ji}) = 1 \tag{4}$$

for every pair (i, j). Equations (3a), (3b) and (4) leave one of the four probabilities $P_{ij}(a_{ij}, a_{ji})$ undefined, giving the model an extra degree of freedom to fix the reciprocity of the network. This can be done by considering the correlation coefficient

$$\rho_{ij} = \frac{\langle a_{ij}a_{ji}\rangle - \langle a_{ij}\rangle\langle a_{ji}\rangle}{\sqrt{\left(\left\langle a_{ij}^2\right\rangle - \langle a_{ij}\rangle^2\right)\left(\left\langle a_{ji}^2\right\rangle - \langle a_{ji}\rangle^2\right)}} \tag{5a}$$

$$= \frac{P_{ij}(1,1) - p_{ij}p_{ji}}{\sqrt{p_{ij}(1-p_{ij})p_{ji}(1-p_{ji})}} \tag{5b}$$

where $\langle\cdot\rangle$ corresponds to an average over the network ensemble defined by the joint probabilities. Note that because $P_{ij}(1,1) \in [0,1]$, equation (5) is not guaranteed to be bounded between −1 and 1. Enforcing these bounds yields an expression for $P_{ij}(1,1)$ in terms of p_{ij}, p_{ji} and a parameter $\nu \in [-1,1]$ controlling the level of reciprocity between nodes i and j

$$P_{ij}(1,1) = \begin{cases} (1+\nu)p_{ij}p_{ji} + \nu(1-p_{ij}-p_{ji})H(p_{ij}+p_{ji}-1) & \text{for } -1 \le \nu \le 0 \\ (1-\nu)p_{ij}p_{ji} + \nu \min\{p_{ij}, p_{ji}\} & \text{for } 0 \le \nu \le 1, \end{cases} \tag{6}$$

where $H(\cdot)$ is the Heaviside step function (a detailed derivation is provided in Supplementary Information Section I). For instance, the cases $\nu = 1$, 0, and −1 correspond, respectively, to the highest level of reciprocity that is structurally possible, random reciprocity (that is, directed links exist in both directions with probability $p_{ij}p_{ji}$) and *anti*-reciprocity meaning the minimum level of reciprocity achievable given the joint probabilities. Note that fully reciprocal networks ($r = 1$) are only possible when $\nu = 1$ and $p_{ij} = p_{ji}$ for every pair of nodes i and j.

Alongside equations (3a), (3b) and (4), equation (6) fully defines the four joint probabilities $P_{ij}(a_{ij}, a_{ji})$ prescribing how nodes i and j are connected and thus the level of reciprocity in the network ensemble. The latter can be made explicit by computing the expected reciprocity[36]

$$\langle r\rangle = \left\langle \frac{L^{\leftrightarrow}}{L} \right\rangle \approx \frac{\langle L^{\leftrightarrow}\rangle}{\langle L\rangle} = \frac{\langle k^{\leftrightarrow}\rangle}{\langle k^+\rangle}, \tag{7}$$

where L is the number of links, $L^{\leftrightarrow}$ is the number of links that are reciprocated (that is, a directed link that has another link in the opposite direction),

$$\langle k^+\rangle = \langle k^-\rangle = \frac{1}{N}\sum_{i=1}^{N}\sum_{\substack{j=1 \\ j\neq i}}^{N} p_{ij} \tag{8}$$

is the expected degree (in or out) and

$$\langle k^{\leftrightarrow}\rangle = \frac{2}{N}\sum_{i=1}^{N}\sum_{j=i+1}^{N} P_{ij}(1,1) \tag{9}$$

is the expected reciprocated degree.

Having introduced these two frameworks, we now combine them into the definitive formulation of the directed-reciprocal $\mathbb{S}^1$ model.

The directed-reciprocal $\mathbb{S}^1$ model

As mentioned previously, the directed $\mathbb{S}^1$ model generates reciprocal links by chance, that is, when two directed links happen to exist in opposite directions between a given pair of nodes. We found, however, that although Fig. 1 shows that reciprocity and the density of triangles are somewhat correlated in real directed complex networks, relying on luck does not allow the accurate reproduction of the levels of reciprocity found in most real network datasets. In other words, once $\{\kappa_i^-\}_{i=1,\ldots,N}$ and $\{\kappa_i^+\}_{i=1,\ldots,N}$ have been set to reproduce the joint degree sequence and β has been chosen to reproduce the density of triangles, an additional parameter is required to accurately tune the level of reciprocity to march the level in a real directed complex network targeted for study.

The directed-reciprocal $\mathbb{S}^1$ model is the combination of the two modelling frameworks defined in the previous sections. Combining

https://doi.org/10.1038/s41567-023-02246-6

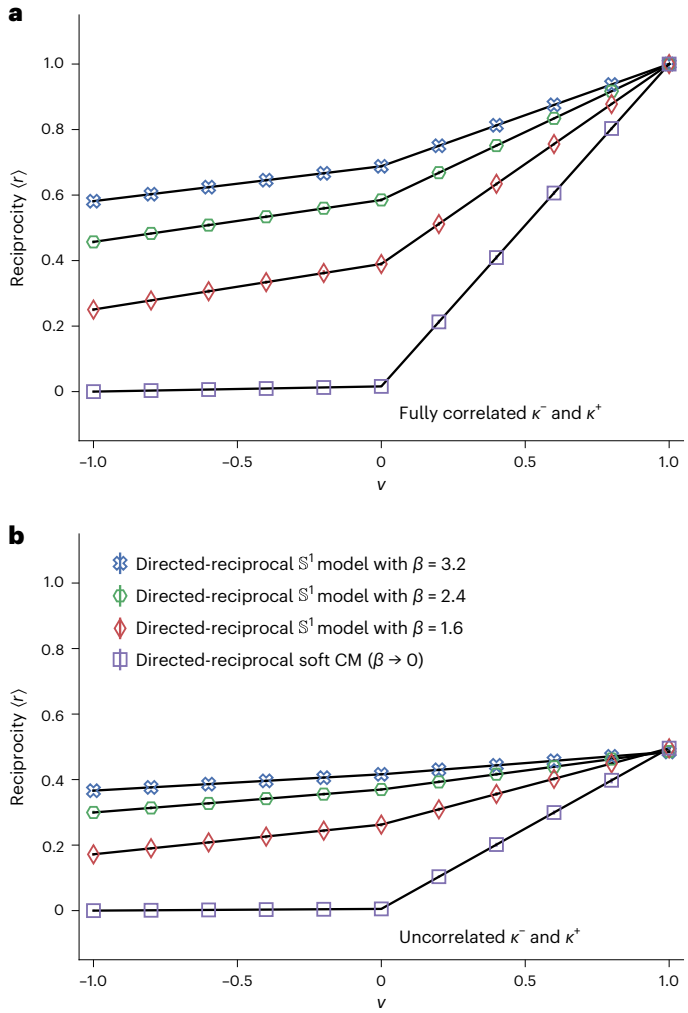

Fig. 3 | Validation of the general framework controlling reciprocity. **a**, Reciprocity versus its control parameter ν, setting $k_i^- = k_i^+$ for $i = 1, \ldots, N$ to fully correlate κ^- and κ^+. **b**, Reciprocity versus its control parameter ν, shuffling the sequence $\{k_i^+\}_{i=1,\ldots,N}$ used in **a** to decorrelate κ^- and κ^+. We consider both the directed-reciprocal soft configuration model (Methods) and the directed-reciprocal $\mathbb{S}^1$ model (main text). Each symbol shows $\langle r\rangle$ estimated from 100 random synthetic networks composed of $N = 2{,}500$ nodes. Solid lines show the predictions of equations (6)–(9). Error bars show the estimated 95% confidence interval (almost always smaller than the width of the solid lines). To highlight the dependency of $\langle r\rangle$ on β and on the correlation between κ^- and κ^+, we drew a sequence $\{\kappa_i^-\}_{i=1,\ldots,N}$ from the pdf $\rho(\kappa) \propto \kappa^{-2.5}$ with $5 < \kappa < 100$ and a sequence $\{\theta_i\}_{i=1,\ldots,N}$ from the pdf $\varphi(\theta) = \frac{1}{2\pi}$. All symbols and lines were obtained using these two sequences.

equations (1a) and (6) fixes $P_{ij}(1,1)$, which in turn fixes $P_{ij}(1,0)$ and $P_{ij}(0,1)$ via equations (3a) and (3b). Finally, asking for normalization sets $P_{ij}(0,0)$. The parameter ν can therefore serve as the extra parameter required to control the level of reciprocity.

Figure 3 illustrates the range of reciprocity that can be obtained with the directed-reciprocal $\mathbb{S}^1$ model as well as with the directed soft configuration model, which corresponds to the limit $\beta \to 0$ (ref. 41). In both Fig. 3a and Fig. 3b, nodes were distributed homogeneously at random on the circle and assigned hidden degrees. In Fig. 3a, the in-degrees and out-degrees are fully correlated—so that $\kappa_i^+ = \kappa_i^- \; \forall i$—while in Fig. 3b they are uncorrelated. Links were then added randomly according to the joint probabilities $P_{ij}(a_{ij}, a_{ji})$ defined by equations (1a), (3a), (3b), (4) and (6). Figure 3 illustrates the effect that both the parameter β and the correlation between κ^- and κ^+ have on reciprocity. Indeed, we note that stronger correlations between κ^- and κ^+ and larger values of β both yield networks with a higher reciprocity. To understand this interplay, we introduce $\kappa_{ij} = \kappa_i^+ \kappa_j^-$ and use equation (1a) to rewrite equations (6)–(9) as

$$\langle r\rangle \approx \frac{\langle k^{\leftrightarrow}\rangle}{\langle k^{+}\rangle} = \begin{cases} (1+\nu)\langle r \,|\, \nu = 0\rangle - \nu\langle r \,|\, \nu = -1\rangle & \text{for } -1 \le \nu \le 0 \\ (1-\nu)\langle r \,|\, \nu = 0\rangle + \nu\langle r \,|\, \nu = +1\rangle & \text{for } 0 \le \nu \le 1 \end{cases} \tag{10a}$$

with

$$\langle r \,|\, \nu = +1\rangle \simeq \frac{1}{\langle\kappa\rangle^2}\left\langle \min\{\kappa_{ij}, \kappa_{ji}\}\right\rangle, \tag{10b}$$

$$\langle r \,|\, \nu = 0\rangle \simeq \frac{1}{\langle\kappa\rangle^2}\left\langle \kappa_{ij}\kappa_{ji}\frac{\kappa_{ij}^{\beta-1} - \kappa_{ji}^{\beta-1}}{\kappa_{ij}^{\beta} - \kappa_{ji}^{\beta}}\right\rangle, \tag{10c}$$

and

$$\langle r \,|\, \nu = -1\rangle \simeq \frac{\sin(\pi/\beta)}{\langle\kappa\rangle^2(\pi/\beta)}\langle f(\kappa_{ij}, \kappa_{ji}, \beta)\rangle, \tag{10d}$$

where $f(\kappa_{ij}, \kappa_{ji}, \beta)$ is a symmetric function with respect to κ_{ij} and κ_{ji} and an increasing function with respect to β. A detailed derivation of these equations is provided in Supplementary Information Section II.F. Equation (10a) already explains the observed linear behaviour with parameter ν, although with two different slopes for positive or negative values.

Regarding the dependence on parameter β and in- and out-degree correlations, first, we observe that equation (10b) does not depend on β and therefore that maximal reciprocity—reached at $\nu = 1$—only depends on the correlation between κ^- and κ^+. This observation is confirmed in Fig. 3. Equation (10b) also confirms our previous observation that fully reciprocal networks (that is, $r = 1$) can only be expected when $P(a_{ij} = 1|\kappa_i^+, \kappa_j^-, \Delta\theta_{ij}) = P(a_{ji} = 1|\kappa_j^+, \kappa_i^-, \Delta\theta_{ij})$ which implies that κ^- and κ^+ are fully correlated (that is, $\kappa_i^- = \kappa_i^+$ for $i = 1, \ldots, N$). Any weaker correlation will imply a lower reciprocity since the step functions will oversample $\min\{\kappa_{ij}, \kappa_{ji}\}$ leading to the right-hand side of equation (10b) being less than 1.

Second, we observe in Fig. 3 that larger values of β allow for higher levels of reciprocity. This can be understood by noting that equation (1a) becomes a step function as $\beta \to \infty$. In this limit, any pair of nodes i and j for which $\max\{\chi_{ij}, \chi_{ji}\} < 1$ will be connected by a reciprocal link with probability 1. As β decreases, this probabilities for these same pairs of nodes will also decrease, and this drop in likelihood will not be compensated by the fact that reciprocal links between pairs of nodes with larger χ_{ij} or χ_{ji} are becoming likelier (equation (1a) decreases too quickly). As a consequence, the reciprocity increases with β. This relationship becomes explicit when κ^- and κ^+ are fully correlated (that is, $\kappa_{ij} = \kappa_{ji}$) as equation (10c) becomes $\langle r \,|\, \nu = 0\rangle \simeq 1 - 1/\beta$.

Modelling real directed complex networks

We now explore the capacity of the directed-reciprocal $\mathbb{S}^1$ model to reproduce the structure of real directed complex networks, most notably their level of reciprocity and their clustering patterns (Fig. 2a). Inspired by the parameter inference procedure of ref. 42, we designed an inference algorithm for the $2N + 2$ parameters—$\{\kappa_i^-, \kappa_i^+\}_{i=1,\ldots,N}$, β and ν—so that the directed-reciprocal $\mathbb{S}^1$ model reproduces, on average, the joint in- and out-degree sequence, the reciprocity and the density of triangles (regardless of their configuration) of an original real directed complex network ($2N + 2$ constraints). These $2N + 2$ parameters are inferred when averaging over all possible angular positions (assuming a uniform pdf), meaning that angular positions $\{\theta_i\}_{i=1,\ldots,N}$ are not inferred. A detailed description of the inference algorithm is provided

https://doi.org/10.1038/s41567-023-02246-6

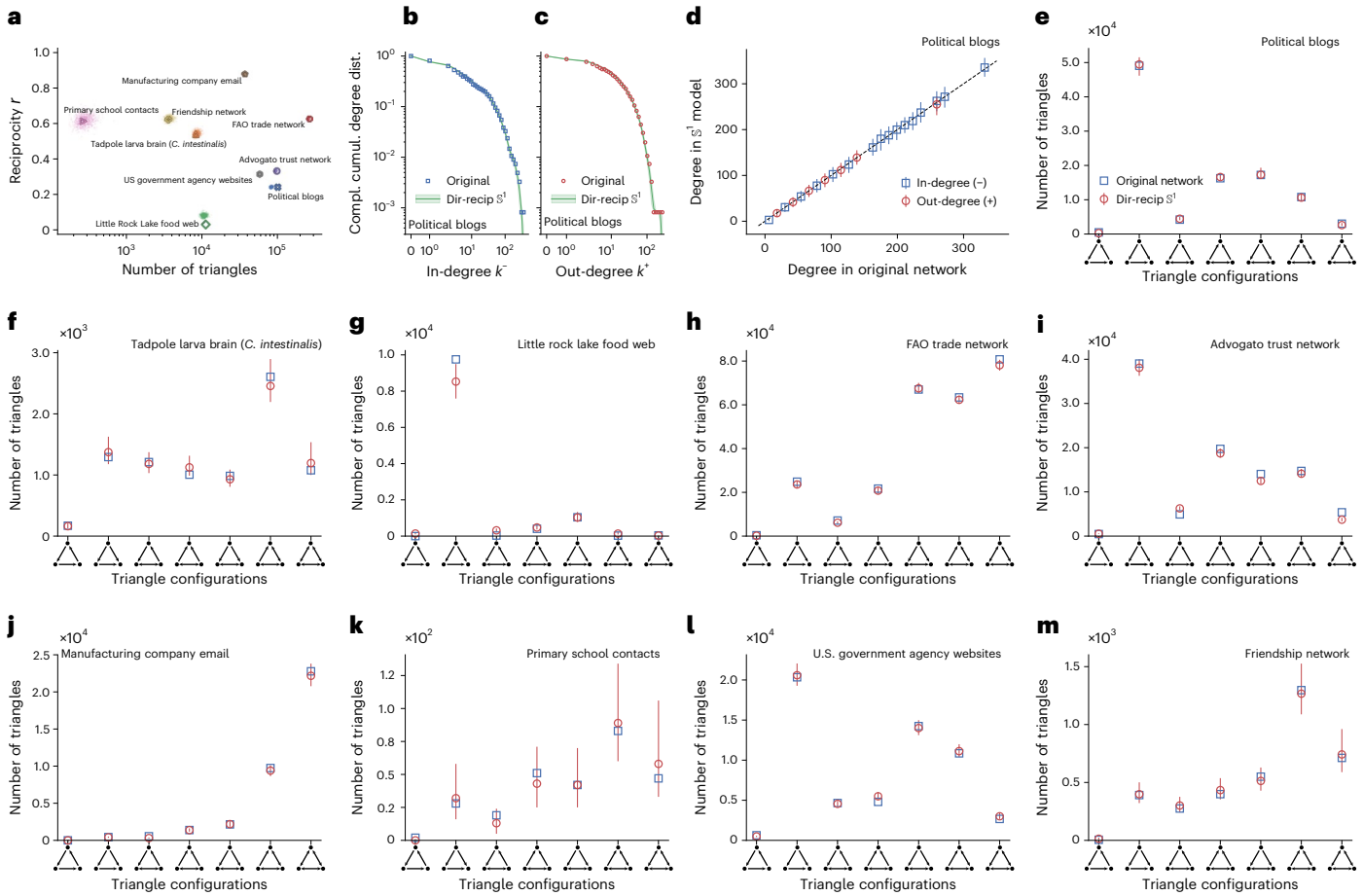

Fig. 4 | Reproducing topological features of real directed networks with the directed-reciprocal $\mathbb{S}^1$ model (Dir-recip). a, Reciprocity and the number of triangles measured on the real networks (symbols) compared to those measured on synthetic networks (1000 network instances; small translucid regions around corresponding symbols). **b**, Complementary cumulative in-degree distribution (Comp. cumul. degree dist.) for the political blogs (polblogs) dataset. **c**, Complementary cumulative out-degree distribution for the polblogs dataset. **d**, In-degree and out-degree of individual nodes for the polblogs dataset, plotting the degrees measured in the real dataset versus the values calculated by the model. Only a fraction of the symbols are shown to avoid cluttering the plot. **e**, Number of triangles of each possible configuration as shown in Fig. 2a (that is, the triangle spectrum) for the polblogs dataset. The key in **e** applies to **f**–**m** as well. **f**, Same as **e**,but for the connectome of a tadpole larva of *Ciona intestinalis* using the cintestinalis dataset. **g**, Same as **e**, but for the food web of Little Rock Lake using the foodweb_little_rock dataset. **h**, Same as **e**, but for trade relationships between countries using the fao_trade dataset. **i**, Same as **e**, but for trust relationships among users in an online community of software developers using the advogato dataset. **j**, Same as **e**, but for emails among employees of a manufacturing company usign the email_company dataset. **k**, Same as **e**, but for friendships between high school students using the sp_high_school_diaries dataset. **l**, Same as **e**, but for links between Washington State's government agencies' websites using the us_agencies_washington dataset. **m**, Same as **e**, but for friendships among students living in a residence hall using the residence_hall dataset. For each dataset, the parameters of the directed-reciprocal $\mathbb{S}^1$ model were adjusted using the inference procedure described in Supplementary Information Section IV. Green shaded areas in **b** and **c**, and vertical lines in **d**–**m** show the estimated 95% confidence interval (the 2.5 and 97.5 percentiles) obtained from 1,000 random network instances. The excellent congruence of measured versus modelled category counts for these nine real complex networks shows the quality and utility of the proposed method.

in Supplementary Information Section IV, and its implementation in C++ is publicly available (Methods).

We ran our algorithm on more than two dozen representative datasets from the Netzschleuder network catalogue and repository (https://networks.skewed.de). The results for nine of the datasets are shown in Fig. 4. Figure 4a shows the ability of our model to reproduce the reciprocity and the number of triangles. Figure 4b–d provides a representative illustration of the excellent agreement between the local properties of networks generated by our model, the in- and out-degree sequence and those of the real counterpart. Beyond the degree sequences, Fig. 4d shows that the model reproduces the observed correlations between in-degrees and out-degrees. This agreement is somewhat expected given that $2N$ parameters are dedicated to fixing the expected in-degrees and out-degree of each node. The most striking result, however, consists in the accuracy with which the directed-reciprocal $\mathbb{S}^1$ model can reproduce the variety of clustering patterns observed in a wide range of real directed complex networks (that is, their triangle spectrum). Indeed, Fig. 4e–m, as well as Supplementary Information Fig. 4, show that only two parameters are necessary to match the observed reciprocity and nontrivial clustering patterns. Therefore, our results imply that clustering in directed networks arises as a consequence of geometry and of the tendency to generate reciprocated interactions.

Discussion

Asymmetric interactions within complex systems are the norm rather than the exception[20]. Yet, for lack of sufficiently adequate modelling frameworks, it is common to see directionality neglected and treated somewhat as an afterthought[21], the underlying assumption being that the undirected representation of many complex systems encodes most

of the relationship between the structure of these systems and their behaviour. However, mounting evidence argues that this is not the case, and that directionality drastically impacts the global organization and the behaviour of these systems[20,21,23–29,43]. Hence, overlooking directionality provides an incomplete if not misleading picture.

Extending the framework of network geometry to directed networks has therefore been an urgent matter for many years, but progress was impeded by the fundamental incompatibility between asymmetric interactions and the symmetry of distances in any metric space. In this paper, we showed that this incompatibility can be bypassed by rethinking the relationship between connections and distances. This approach has resulted in a powerful and versatile framework amenable to analytical calculations that is easily adjusted to reproduce properties observed in a large variety of real network datasets.

We showed that our framework reproduces the intricate patterns of reciprocity and clustering observed in real directed complex networks. Albeit local, these features have a major impact on the global behaviour of these networks. For instance, they affect the outcome of spreading dynamics[26], impact the stability of food webs[24,25] and play a central role for flexible navigation and context-dependent action selection in connectomes[44]. Also, the information encoded in the patterns of reciprocity and of clustering is rich enough for them to act as a signature of the nature of real complex networks (social, technological, physical, biological and so on)[36,45,46]. The method fulfils the paramount need that any realistic modelling approach be able to reproduce the intricate patterns of reciprocity and clustering of real complex networks under study. Now that the gap between asymmetric interactions and symmetric metric distances has been bridged, accurate modelling of a wide and diverse range of complex systems is within reach.

Online content

Any methods, additional references, Nature Portfolio reporting summaries, source data, extended data, supplementary information, acknowledgements, peer review information, details of author contributions and competing interests, and statements of data and code availability are available at https://doi.org/10.1038/s41567-023-02246-6.

References

1. Boguñá, M. et al. Network geometry. *Nat. Rev. Phys.* **3**, 114–135 (2021).
2. Serrano, M. Á., Krioukov, D. & Boguñá, M. Self-similarity of complex networks and hidden metric spaces. *Phys. Rev. Lett.* **100**, 078701 (2008).
3. García-Pérez, G., Boguñá, M. & Serrano, M. Á. Multiscale unfolding of real networks by geometric renormalization. *Nat. Phys.* **14**, 583–589 (2018).
4. Zheng, M., García-Pérez, G., Boguñá, M. & Serrano, M. Á. Scaling up real networks by geometric branching growth. *Proc. Natl Acad. Sci. USA* **118**, e2018994118 (2021).
5. Boguñá, M., Krioukov, D., Almagro, P. & Serrano, M. Á. Small worlds and clustering in spatial networks. *Phys. Rev. Res.* **2**, 023040 (2020).
6. Papadopoulos, F., Kitsak, M., Serrano, M. Á., Boguñá, M. & Krioukov, D. Popularity versus similarity in growing networks. *Nature* **489**, 537–540 (2012).
7. Allard, A., Serrano, M. Á., García-Pérez, G. & Boguñá, M. The geometric nature of weights in real complex networks. *Nat. Commun.* **8**, 14103 (2017).
8. Kleineberg, K.-K., Boguñá, M., Serrano, M. Á. & Papadopoulos, F. Hidden geometric correlations in real multiplex networks. *Nat. Phys.* **12**, 1076–1081 (2016).
9. Newman, M. E. J. *Networks* (Oxford Univ. Press, 2018).
10. Allard, A., Hébert-Dufresne, L., Young, J.-G. & Dubé, L. J. General and exact approach to percolation on random graphs. *Phys. Rev. E* **92**, 062807 (2015).
11. Gleeson, J. P. & Melnik, S. Analytical results for bond percolation and k-core sizes on clustered networks. *Phys. Rev. E* **80**, 046121 (2009).
12. Newman, M. E. J. Properties of highly clustered networks. *Phys. Rev. E* **68**, 026121 (2003).
13. Karrer, B. & Newman, M. E. J. Random graphs containing arbitrary distributions of subgraphs. *Phys. Rev. E* **82**, 066118 (2010).
14. Miller, J. C. Percolation and epidemics in random clustered networks. *Phys. Rev. E* **80**, 020901 (2009).
15. Battiston, F. et al. Networks beyond pairwise interactions: structure and dynamics. *Phys. Rep.* **874**, 1–92 (2020).
16. Lee, C. & Wilkinson, D. J. A review of stochastic block models and extensions for graph clustering. *Appl. Netw. Sci.* **4**, 122 (2019).
17. Orsini, C. et al. Quantifying randomness in real networks. *Nat. Commun.* **6**, 8627 (2015).
18. Serrano, M. Á. & Boguñá, M. Tuning clustering in random networks with arbitrary degree distributions. *Phys. Rev. E* **72**, 036133 (2005).
19. Volz, E. Random networks with tunable degree distribution and clustering. *Phys. Rev. E* **70**, 056115 (2004).
20. Asllani, M., Lambiotte, R. & Carletti, T. Structure and dynamical behavior of non-normal networks. *Sci. Adv.* **4**, eaau9403 (2018).
21. Johnson, S. Digraphs are different: why directionality matters in complex systems. *J. Phys. Complex* **1**, 015003 (2020).
22. Levine, S. Several measures of trophic structure applicable to complex food webs. *J. Theor. Biol.* **83**, 195–207 (1980).
23. Duan, C., Nishikawa, T., Eroglu, D. & Motter, A. E. Network structural origin of instabilities in large complex systems. *Sci. Adv.* **8**, eabm8310 (2022).
24. Johnson, S., Domínguez-García, V., Donetti, L. & Muñoz, M. A. Trophic coherence determines food-web stability. *Proc. Natl Acad. Sci. USA* **111**, 17923–17928 (2014).
25. Johnson, S. & Jones, N. S. Looplessness in networks is linked to trophic coherence. *Proc. Natl Acad. Sci. USA* **114**, 5618–5623 (2017).
26. Klaise, J. & Johnson, S. From neurons to epidemics: How trophic coherence affects spreading processes. *Chaos* **26**, 065310 (2016).
27. Nicolaou, Z. G., Nishikawa, T., Nicholson, S. B., Green, J. R. & Motter, A. E. Non-normality and non-monotonic dynamics in complex reaction networks. *Phys. Rev. Res.* **2**, 043059 (2020).
28. Qu, B., Li, Q., Havlin, S., Stanley, H. E. & Wang, H. Nonconsensus opinion model on directed networks. *Phys. Rev. E* **90**, 052811 (2014).
29. Shao, J., Havlin, S. & Stanley, H. E. Dynamic opinion model and invasion percolation. *Phys. Rev. Lett.* **103**, 018701 (2009).
30. Michel, J., Reddy, S., Shah, R., Silwal, S. & Movassagh, R. Directed random geometric graphs. *J. Complex Netw.* **7**, 792–816 (2019).
31. Wolf, F., Kirsch, C. & Donner, R. V. Edge directionality properties in complex spherical networks. *Phys. Rev. E* **99**, 012301 (2019).
32. Wu, Z., Di, Z. & Fan, Y. An asymmetric popularity-similarity optimization method for embedding directed networks into hyperbolic space. *Complexity* **2020**, 8372928 (2020).
33. Kovács, B. & Palla, G. Model-independent embedding of directed networks into Euclidean and hyperbolic spaces. *Commun. Phys.* **6**, 28 (2023).
34. Peralta-Martinez, K. & Méndez-Bermúdez, J. A. Directed random geometric graphs: structural and spectral properties. *J. Phys. Complex.* **4**, 015002 (2022).
35. Wasserman, S. & Faust, K. *Social Network Analysis: Methods and Applications* (Cambridge Univ. Press, 1994).
36. Garlaschelli, D. & Loffredo, M. I. Patterns of link reciprocity in directed networks. *Phys. Rev. Lett.* **93**, 268701 (2004).

37. García-Pérez, G., Serrano, M. Á. & Boguñá, M. Soft communities in similarity space. *J. Stat. Phys.* **173**, 775–782 (2018).
38. Muscoloni, A. & Cannistraci, C. V. A nonuniform popularity-similarity optimization (nPSO) model to efficiently generate realistic complex networks with communities. *New J. Phys.* **20**, 052002 (2018).
39. Désy, B., Desrosiers, P. & Allard, A. Dimension matters when modeling network communities in hyperbolic spaces. *PNAS Nexus* **2**, pgad136 (2023).
40. Holland, P. W. & Leinhardt, S. An exponential family of probability distributions for directed graphs. *J. Am. Stat. Assoc.* **76**, 33–50 (1981).
41. van der Kolk, J., Serrano, M. Á. & Boguñá, M. An anomalous topological phase transition in spatial random graphs. *Commun. Phys.* **5**, 245 (2022).
42. García-Pérez, G., Allard, A., Serrano, M. Á. & Boguñá, M. Mercator: uncovering faithful hyperbolic embeddings of complex networks. *New J. Phys.* **21**, 123033 (2019).
43. Coletta, L. et al. Network structure of the mouse brain connectome with voxel resolution. *Sci. Adv.* **6**, eabb7187 (2020).
44. Hulse, B. K. et al. A connectome of the Drosophila central complex reveals network motifs suitable for flexible navigation and context-dependent action selection. *eLife* **10**, e66039 (2021).
45. Ahnert, S. E. & Fink, T. M. A. Clustering signatures classify directed networks. *Phys. Rev. E* **78**, 036112 (2008).
46. Jia, M., Gabrys, B. & Musial, K. Directed closure coefficient and its patterns. *PLoS ONE* **16**, e0253822 (2021).
47. Holland, P. W. & Leinhardt, S. Local structure in social networks. *Sociol. Methodol.* **7**, 1–45 (1976).

Methods

Density of triangles in directed networks

We quantify the density of triangles in a directed network with the average local clustering coefficient, $\bar{c}_{\text{undir}}$, computed using the undirected version of the original directed network. From the adjacency matrix of the directed network $A = \{a_{ij}\}$, we define the undirected adjacency matrix $\tilde{A}$ whose elements are $\tilde{a}_{ij} = \max(a_{ij}, a_{ji})$. The density of triangles is then

$$\bar{c}_{\text{undir}} = \frac{1}{N}\sum_{i=1}^{N} \frac{2T_i}{k_i(k_i-1)} \mathbb{1}_{\{k_i>1\}}, \tag{11}$$

where $T_i = \frac{1}{2}[\tilde{A}^3]_{ii}$ is the number of triangles to which node i participates, $k_i = \sum_{j=1}^{N}[\tilde{A}]_{ij}$ is the degree of node i and $\mathbb{1}_{\{\cdot\}}$ is the indicator function.

Correspondence with the directed soft configuration model

The directed soft configuration model is the unique ensemble of unbiased sparse random graphs whose entropy is maximized across all graphs with a given expected joint in- and out-degree distribution[48,49]. It consists of N nodes, each of which is assigned a pair of hidden degrees κ^- and κ^+ according to $\rho(\kappa^-, \kappa^+)$. In this model, a directed link from node i to node j exists with probability

$$P\left(a_{ij}=1 \,|\, \kappa_i^+, \kappa_j^-\right) = \frac{1}{1+\frac{N\langle\kappa\rangle}{\kappa_i^+\kappa_j^-}} \simeq \frac{\kappa_i^+\kappa_j^-}{N\langle\kappa\rangle}, \tag{12}$$

where the approximation holds in the sparse limit. In this limit the directed soft configuration model falls back to a directed version of the Chung–Lu model[50]. Note that the directed $\mathbb{S}^1$ model falls back to the directed soft configuration model in the limit $\beta \to 0$ (ref. 5).

To see how the directed $\mathbb{S}^1$ model falls back on the directed soft configuration model, we first average equation (1a) over the angular distance $\Delta\theta_{ij}$ to obtain the expected probability for a link to exist from node i to node j in the network ensemble

$$\left\langle a_{ij} \,|\, \kappa_i^+, \kappa_j^- \right\rangle = {}_2F_1\left(1, \frac{1}{\beta}, 1+\frac{1}{\beta}, -\left(\frac{N}{2\mu\kappa_i^+\kappa_j^-}\right)^{\beta}\right), \tag{13}$$

where ${}_2F_1$ is the hypergeometric function. From this expression, we show in Supplementary Information equation (23b) that in the limit $N/(\kappa_i^+\kappa_j^-) \to \infty$ the average connection probability becomes

$$\left\langle a_{ij} \,|\, \kappa_i^+, \kappa_j^- \right\rangle \simeq \frac{\kappa_i^+\kappa_j^-}{N\langle\kappa\rangle}, \tag{14}$$

which we identify as the connection probability of the sparse directed soft configuration model, equation (12). The generalization of the $\mathbb{S}^1$ model (main text) can therefore be seen as the geometric extension of the directed soft configuration model which, unlike its non-geometric counterpart, has a nonvanishing clustering in the limit $N \to \infty$ (due to the triangle inequality of its embedding space).

Directed-reciprocal soft configuration model

Akin to the directed-reciprocal $\mathbb{S}^1$ model, we introduce the directed-reciprocal soft configuration model, a combination of the framework controlling reciprocity of Supplementary Information Section II.B and of the directed soft configuration model presented above (which provides the marginal probabilities).

Network datasets

The list of all datasets is provided in Supplementary Information Section III. The datasets used in Fig. 4 were originally published as follows: polblogs in ref. 51, cintestinalis in ref. 52, foodweb_little_rock in ref. 53, fao_trade in ref. 54, advogato in ref. 55, email_company in ref. 56, sp_high_school_diaries in ref. 57, us_agencies_washington in ref. 58 and residence_hall in ref. 59.

Data availability

The network datasets used in the article have been made publicly available by the original authors and were downloaded from the Netzschleuder network catalogue and repository (https://networks.skewed.de).

Code availability

The scripts and the source code of the programs used to produce the figures are publicly available on Zenodo (https://doi.org/10.5281/zenodo.8264693).

References

48. Bianconi, G. Entropy of network ensembles. *Phys. Rev. E* **79**, 036114 (2009).
49. van der Hoorn, P., Lippner, G. & Krioukov, D. Sparse maximum-entropy random graphs with a given power-law degree distribution. *J. Stat. Phys.* **173**, 806–844 (2018).
50. Chung, F. & Lu, L. Connected components in random graphs with given expected degree sequences. *Ann. Comb.* **6**, 125–145 (2002).
51. Adamic, L. A. & Glance, N. The political blogosphere and the 2004 U.S. election: divided they blog. In *Proceedings of the 3rd International Workshop on Link Discovery* 36–43 (2005).
52. Ryan, K., Lu, Z. & Meinertzhagen, I. A. The CNS connectome of a tadpole larva of Ciona intestinalis (L.) highlights sidedness in the brain of a chordate sibling. *eLife* **5**, e16962 (2016).
53. Martinez, N. D. Artifacts or attributes? Effects of resolution on the Little Rock Lake food web. *Ecol. Monogr.* **61**, 367–392 (1991).
54. De Domenico, M., Nicosia, V., Arenas, A. & Latora, V. Structural reducibility of multilayer networks. *Nat. Commun.* **6**, 6864 (2015).
55. Massa, P., Salvetti, M. & Tomasoni, D. Bowling alone and trust decline in social network sites. In *2009 Eighth IEEE International Conference on Dependable, Autonomic and Secure Computing* 658–663 (2009).
56. Michalski, R., Palus, S. & Kazienko, P. Matching organizational structure and social network extracted from email communication. In *Business Information Systems* (ed. Abramowicz, W.) 197–206 (Springer, 2011).
57. Mastrandrea, R., Fournet, J. & Barrat, A. Contact patterns in a high school: a comparison between data collected using wearable sensors, contact diaries and friendship surveys. *PLoS ONE* **10**, e0136497 (2015).
58. Kosack, S. et al. Functional structures of US state governments. *Proc. Natl Acad. Sci. USA* **115**, 11748–11753 (2018).
59. Freeman, L. C., Webster, C. M. & Kirke, D. M. Exploring social structure using dynamic three-dimensional color images. *Soc. Networks* **20**, 109–118 (1998).

Acknowledgements

We are grateful to L. J. Dubé for comments and to the nursing staff at the Centre de recherche clinique et évaluative en oncologie (CRCEO) where part of this work was done. A.A. acknowledges financial support from the Sentinelle Nord initiative of the Canada First Research Excellence Fund and from the Natural Sciences and Engineering Research Council of Canada (project 2019-05183). M.A.S. and M.B. acknowledge support from Grant TED2021-129791B-I00 funded by MCIN/AEI/10.13039/501100011033 and the European Union

NextGenerationEU/PRTR, Grant PID2022-137505NB-C22 funded by MCIN/AEI/10.13039/501100011033, Grant PID2019-106290GB-C22 funded by MCIN/AEI/10.13039/501100011033 and Generalitat de Catalunya grant number 2021SGR00856. M.B. acknowledges the ICREA Academia award funded by the Generalitat de Catalunya.

Author contributions

All authors designed the research. A.A. and M.B. did the analytical calculations. A.A. performed the numerical simulations. All authors discussed the results and implications and wrote the manuscript.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41567-023-02246-6.

Correspondence and requests for materials should be addressed to Antoine Allard.

Peer review information *Nature Physics* thanks the anonymous reviewers for their contribution to the peer review of this work.

nature physics

Article

https://doi.org/10.1038/s41567-023-02246-6

Geometric description of clustering in directed networks

In the format provided by the authors and unedited

CONTENTS

S.I. CONTROLLING THE RECIPROCITY IN RANDOM DIRECTED NETWORKS

We consider a general random directed networks model in which p_{ij} is the probability for a directed link to exist from node i to node j. We denote the number of nodes with N. To control the level of reciprocity, our approach focuses on each *pair* of directed links between two nodes and defines the four following symmetrical probabilities

$$\begin{aligned}
&P_{ij}(a_{ij}=0,a_{ji}=0) && \text{(none of the two possible directed links exist)}\\
&P_{ij}(a_{ij}=1,a_{ji}=0) && \text{(the link from node } i \text{ to node } j \text{ exists but the other does not)}\\
&P_{ij}(a_{ij}=0,a_{ji}=1) && \text{(the link from node } j \text{ to node } i \text{ exists but the other does not)}\\
&P_{ij}(a_{ij}=1,a_{ji}=1) && \text{(both directed links exists)}
\end{aligned}$$

with $1 \leq i < j \leq N$ and where a_{ij} is 1 if there is a directed link from node i to node j, and 0 otherwise. These four joint probabilities are normalized

$$P_{ij}(a_{ij}=0,a_{ji}=0)+P_{ij}(a_{ij}=1,a_{ji}=0)+P_{ij}(a_{ij}=0,a_{ji}=1)+P_{ij}(a_{ij}=1,a_{ji}=1)=1\ , \tag{S1}$$

and their marginals must be coherent with the random directed network model

$$P_{ij}(a_{ij}=1,a_{ji}=0)+P_{ij}(a_{ij}=1,a_{ji}=1)=p_{ij}\ , \tag{S2a}$$
$$P_{ij}(a_{ij}=0,a_{ji}=1)+P_{ij}(a_{ij}=1,a_{ji}=1)=p_{ij}\ . \tag{S2b}$$

To connect the probabiilities $P_{ij}(a_{ij},a_{ji})$ with the reciprocity in the network ensemble defined by the model, we look at the following correlation coefficient

$$\rho_{ij}=\frac{\langle a_{ij}a_{ji}\rangle-\langle a_{ij}\rangle\langle a_{ji}\rangle}{\sqrt{\left(\langle a_{ij}^2\rangle-\langle a_{ij}\rangle^2\right)\left(\langle a_{ji}^2\rangle-\langle a_{ji}\rangle^2\right)}}=\frac{P_{ij}(1,1)-p_{ij}p_{ji}}{\sqrt{p_{ij}(1-p_{ij})p_{ji}(1-p_{ji})}}\ . \tag{S3}$$

for each pair (i,j) with $1 \leq i < j \leq N$, and where $\langle\cdot\rangle$ corresponds to an average over the network ensemble. A closed form for $P_{ij}(1,1)$ in terms of p_{ij} and p_{ji} can be obtained by combining Eqs. (S2)–(S3) alongside the requirement that each of the four joint probabilities $P_{ij}(a_{ij},a_{ji})$ is bounded in $[0,1]$:

1. From Eq. (S3), we can isolate

$$P_{ij}(a_{ij}=1,a_{ji}=1)=p_{ij}p_{ji}+\rho_{ij}\sqrt{p_{ij}(1-p_{ij})p_{ji}(1-p_{ji})}\ , \tag{S4}$$

which will be bounded in $[0,1]$ if

$$-\frac{p_{ij}p_{ji}}{\sqrt{p_{ij}(1-p_{ij})p_{ji}(1-p_{ji})}}\leq\rho_{ij}\leq\frac{1-p_{ij}p_{ji}}{\sqrt{p_{ij}(1-p_{ij})p_{ji}(1-p_{ji})}}\ . \tag{S5}$$

2. Combining Eqs. (S2a) and (S4), we can isolate

$$P_{ij}(a_{ij}=1,a_{ji}=0)=p_{ij}-P(a_{ij}=1,a_{ji}=1)=p_{ij}(1-p_{ji})-\rho_{ij}\sqrt{p_{ij}(1-p_{ij})p_{ji}(1-p_{ji})} \tag{S6}$$

which will be bounded in $[0,1]$ if

$$\frac{p_{ij}(1-p_{ji})-1}{\sqrt{p_{ij}(1-p_{ij})p_{ji}(1-p_{ji})}}\leq\rho_{ij}\leq\frac{p_{ij}(1-p_{ji})}{\sqrt{p_{ij}(1-p_{ij})p_{ji}(1-p_{ji})}}\ . \tag{S7}$$

3. Combining Eqs. (S2b) and (S4), we can isolate

$$P_{ij}(a_{ij}=0,a_{ji}=1)=p_{ji}-P_{ij}(a_{ij}=1,a_{ji}=1)=p_{ji}(1-p_{ij})-\rho_{ij}\sqrt{p_{ij}(1-p_{ij})p_{ji}(1-p_{ji})} \tag{S8}$$

which will be bounded in $[0,1]$ if

$$\frac{p_{ji}(1-p_{ij})-1}{\sqrt{p_{ij}(1-p_{ij})p_{ji}(1-p_{ji})}}\leq\rho_{ij}\leq\frac{p_{ji}(1-p_{ij})}{\sqrt{p_{ij}(1-p_{ij})p_{ji}(1-p_{ji})}}\ . \tag{S9}$$

4. Combining Eqs. (S1), (S4), (S6) and (S8), we can isolate

$$\begin{aligned} P_{ij}(a_{ij}=0,a_{ji}=0) &= 1 - P_{ij}(a_{ij}=1,a_{ji}=0) - P_{ij}(a_{ij}=0,a_{ji}=1) - P_{ij}(a_{ij}=1,a_{ji}=1) \\ &= (1-p_{ji})(1-p_{ij}) + \rho_{ij}\sqrt{p_{ij}(1-p_{ij})p_{ji}(1-p_{ji})} \end{aligned} \tag{S10}$$

which will be bounded in $[0,1]$ if

$$-\frac{(1-p_{ji})(1-p_{ij})}{\sqrt{p_{ij}(1-p_{ij})p_{ji}(1-p_{ji})}} \leq \rho_{ij} \leq \frac{1-(1-p_{ji})(1-p_{ij})}{\sqrt{p_{ij}(1-p_{ij})p_{ji}(1-p_{ji})}} \,. \tag{S11}$$

The lower and upper bounds for ρ_{ij} are therefore

$$\begin{aligned} \rho_{ij}^{\min} &= \frac{1}{\sqrt{p_{ij}(1-p_{ij})p_{ji}(1-p_{ji})}} \max\Big\{ -p_{ij}p_{ji},\ p_{ij}(1-p_{ji})-1,\ p_{ji}(1-p_{ij})-1,\ -(1-p_{ji})(1-p_{ij}) \Big\} \\ &= \begin{cases} -\dfrac{p_{ij}p_{ji}}{\sqrt{p_{ij}(1-p_{ij})p_{ji}(1-p_{ji})}} & \text{if } p_{ij}+p_{ji}<1 \\ -\dfrac{(1-p_{ji})(1-p_{ij})}{\sqrt{p_{ij}(1-p_{ij})p_{ji}(1-p_{ji})}} & \text{if } p_{ij}+p_{ji}>1 \end{cases} \end{aligned} \tag{S12}$$

and

$$\begin{aligned} \rho_{ij}^{\max} &= \frac{1}{\sqrt{p_{ij}(1-p_{ij})p_{ji}(1-p_{ji})}} \min\Big\{ 1-p_{ij}p_{ji},\ p_{ij}(1-p_{ji}),\ p_{ji}(1-p_{ij}),\ 1-(1-p_{ji})(1-p_{ij}) \Big\} \\ &= \begin{cases} \dfrac{p_{ij}(1-p_{ji})}{\sqrt{p_{ij}(1-p_{ij})p_{ji}(1-p_{ji})}} & \text{if } p_{ij}<p_{ji} \\ \dfrac{p_{ji}(1-p_{ij})}{\sqrt{p_{ij}(1-p_{ij})p_{ji}(1-p_{ji})}} & \text{if } p_{ij}>p_{ji} \,. \end{cases} \end{aligned} \tag{S13}$$

To control the level of reciprocity, we introduce a parameter $\nu \in [-1,1]$ controlling ρ_{ij} such that $\rho_{ij}^{\min} \leq \rho_{ij} \leq \rho_{ij}^{\max}$

$$\rho_{ij} = \begin{cases} |\nu|\rho_{ij}^{\min} & \text{if } -1 \leq \nu \leq 0 \\ |\nu|\rho_{ij}^{\max} & \text{if } \ \ 0 \leq \nu \leq 1 \end{cases} \,. \tag{S14}$$

Substituting Eq. (S14) into Eq. (S4) allows us to isolate

$$P_{ij}(a_{ij}=1,a_{ji}=1) = \begin{cases} (1+\nu)p_{ij}p_{ji} - \nu(p_{ij}+p_{ji}-1)H(p_{ij}+p_{ji}-1) & -1 \leq \nu \leq 0 \\ (1-\nu)p_{ij}p_{ji} + \nu \min\Big\{p_{ij},p_{ji}\Big\} & 0 \leq \nu \leq 1 \end{cases} \,, \tag{S15}$$

where $H(\cdot)$ is the Heaviside step function. This last equation alongside Eqs. (S1) and (S2) complete the approach for controlling reciprocity, whose level is tuned by the parameter ν.

S.II. ANALYSIS OF THE DIRECTED-RECIPROCAL $\mathbb{S}^1$ MODEL

A. Description of the model

We consider N nodes positioned on a circle of radius $R = N/2\pi$, thus setting the density of nodes to 1 without loss of generality. Each node i is independently and identically assigned an angular position θ_i and a pair of *hidden* degrees, κ_i^- and κ_i^+ which, as shown below, are related to their in- and out-degree, respectively. The angular positions are scattered on the circle according to the uniform probability density function (pdf)

$$\varphi(\theta) = \frac{1}{2\pi} \ . \tag{S16}$$

The hidden degrees are also assigned randomly according to the joint pdf $\rho(\kappa^-, \kappa^+)$, whose only constraint is

$$\iint \kappa^- \rho(\kappa^-, \kappa^+) d\kappa^- d\kappa^+ = \langle \kappa^- \rangle \equiv \langle \kappa \rangle \tag{S17a}$$

$$\iint \kappa^+ \rho(\kappa^-, \kappa^+) d\kappa^- d\kappa^+ = \langle \kappa^+ \rangle \equiv \langle \kappa \rangle \ . \tag{S17b}$$

A directed link exists from node i to node j with probability

$$P(a_{ij} = 1 | \kappa_i^+, \kappa_j^-, \Delta\theta_{ij}) = \frac{1}{1 + \chi_{ij}^\beta} \qquad \text{with} \quad \chi_{ij} = \frac{R\Delta\theta_{ij}}{\mu \kappa_i^+ \kappa_j^-} = \frac{N\Delta\theta_{ij}}{2\pi\mu\kappa_i^+\kappa_j^-} \tag{S18}$$

where $\Delta\theta_{ij} = \Delta\theta_{ji} = \pi - |\pi - |\theta_i - \theta_j||$ is the minimal angular distance between nodes i and j, and where $\mu > 0$ and $\beta > 1$ are parameters of the model. Note that we will omit writing explicitly the dependency over β, μ and N for brevity. Note also that $\varphi(\theta) = \frac{1}{2\pi}$ implies that the pdf for $\Delta\theta_{ij}$ is simply $1/\pi$.

Two connection events may either be independent or *conditionally* independent. To see this, let us consider the following two entries of the adjacency matrix a_{ij} and a_{lk} with the associated probabilities of connection

$$P(a_{ij} = 1 | \kappa_i^+, \kappa_j^-, \Delta\theta_{ij}) = \frac{1}{1 + \left[\frac{N\Delta\theta_{ij}}{2\pi\mu\kappa_i^+\kappa_j^-}\right]^\beta} \ , \qquad \text{and} \qquad P(a_{lk} = 1 | \kappa_l^+, \kappa_k^-, \Delta\theta_{lk}) = \frac{1}{1 + \left[\frac{N\Delta\theta_{lk}}{2\pi\mu\kappa_l^+\kappa_k^-}\right]^\beta} \ ,$$

where we assume that $i \neq j$ and $l \neq k$ (i.e., no self loops). We distinguish several scenarios:

1. If $i = l$ and $j = k$, then the two connection events are trivially the same.

2. If $i = l$ and $j \neq k$ ($j = k$ and $i \neq l$), then the two links are outgoing (incoming) links from (on) the same node i (j) and both depend on the parameter κ_i^+ (κ_j^-). They are therefore conditionally independent.

3. If $i = k$ and $j \neq l$ ($j = l$ and $i \neq k$) then one link is outgoing from node i (j) and the other is incoming to node i (j). The two connection events will conditionally independent if and only if κ_i^{out} and κ_i^{in} (κ_j^{out} and κ_j^{in}) are correlated. Otherwise, the two connection events are independent.

4. If $i = k$ and $j = l$, then the two links are in opposite direction between the same two nodes. Both connection events depend on the angular separation $\Delta\theta_{ij}$ between the two nodes, and are therefore conditionally independent. The two connection events could be even further correlated if κ_i^{out} and κ_i^{in} (or equivalently κ_j^{out} and κ_j^{in}) are also correlated.

5. If i, j, l and k all take distinct values, then the two connection events are independent.

B. Out-degree of nodes

Let us first consider N nodes, each of which has been assigned an angular position θ, a hidden in-degree κ^- and a hidden out-degree κ^+. The sequence of angular positions, noted $\boldsymbol{\theta} \equiv \{\theta_1, \ldots, \theta_N\}$ is distributed according to the pdf $\prod_{i=1}^N \varphi(\theta_i) = (2\pi)^{-N}$, and the hidden degrees sequence, noted $\kappa \equiv \{\kappa_1^-, \kappa_1^+, \ldots, \kappa_N^-, \kappa_N^+\}$ is distributed according

to the pdf $\prod_{i=1}^{N}\rho(\kappa_i^-,\kappa_i^+)$. The connection probability given by Eq. (S18) alongside the sequences $\boldsymbol{\theta}$ and $\boldsymbol{\kappa}$ define a random network ensemble in which node i has a out-degree equal to k_i^+ with probability $P_i^+(k_i^+|\boldsymbol{\theta},\boldsymbol{\kappa})$. The associated probability generating function (pgf) is defined as

$$H_i^+(z|\boldsymbol{\kappa},\boldsymbol{\theta}) = \sum_{k_i^+=0}^{N-1} P_i^+(k_i^+|\boldsymbol{\theta},\boldsymbol{\kappa}) z^{k_i^+} = \prod_{\substack{j=1\\ j\neq i}}^{N}\left[1 - P(a_{ij}=1|\kappa_i^+,\kappa_j^-,\Delta\theta_{ij}) + zP(a_{ij}=1|\kappa_i^+,\kappa_j^-,\Delta\theta_{ij})\right] , \tag{S19}$$

where we used the fact that the existence of each link is *conditionally* independent from the existence of the others (i.e. they are independent events given the hidden variables θ_i and κ_i^+). General expressions for the expected out-degree of node i, its variance and for the ensemble average out-degree are respectively

$$\left\langle k_i^+\middle|\boldsymbol{\kappa},\boldsymbol{\theta}\right\rangle = \left.\frac{\partial H_i^+(z|\boldsymbol{\kappa},\boldsymbol{\theta})}{\partial z}\right|_{z=1} = \sum_{\substack{j=1\\ j\neq i}}^{N} P(a_{ij}=1|\kappa_i^+,\kappa_j^-,\Delta\theta_{ij}) , \tag{S20}$$

$$\begin{aligned}\mathrm{Var}\left[k_i^+\middle|\boldsymbol{\kappa},\boldsymbol{\theta}\right] &= \left.\frac{\partial^2 H_i^+(z|\boldsymbol{\kappa},\boldsymbol{\theta})}{\partial z^2}\right|_{z=1} + \left.\frac{\partial H_i^+(z|\boldsymbol{\kappa},\boldsymbol{\theta})}{\partial z}\right|_{z=1} - \left[\left.\frac{\partial H_i^+(z|\boldsymbol{\kappa},\boldsymbol{\theta})}{\partial z}\right|_{z=1}\right]^2 \\ &= \sum_{\substack{j=1\\ j\neq i}}^{N} P(a_{ij}=1|\kappa_i^+,\kappa_j^-,\Delta\theta_{ij})[1 - P(a_{ij}=1|\kappa_i^+,\kappa_j^-,\Delta\theta_{ij})] ,\end{aligned} \tag{S21}$$

and

$$\left\langle k^+\middle|\boldsymbol{\kappa},\boldsymbol{\theta}\right\rangle = \frac{1}{N}\sum_{i=1}^{N}\left\langle k_i^+\middle|\boldsymbol{\kappa},\boldsymbol{\theta}\right\rangle = \frac{1}{N}\sum_{i=1}^{N}\sum_{\substack{j=1\\ j\neq i}}^{N} P(a_{ij}=1|\kappa_i^+,\kappa_j^-,\Delta\theta_{ij}) . \tag{S22}$$

Let us now zoom out of the random network ensemble defined by specific sequences $\boldsymbol{\theta}$ and $\boldsymbol{\kappa}$ to focus instead on the random network ensemble defined by the pdfs $\varphi(\theta)$ and $\rho(\kappa^-,\kappa^+)$ (i.e. any sequences $\boldsymbol{\theta}$ of length N and $\boldsymbol{\kappa}$ of length $2N$ drawn from their respective pdf). Averaging over all angular positions (or, equivalently, over all angular distances), the expected probability for a link to exist from node i to node j in the network ensemble becomes

$$\begin{aligned}\left\langle a_{ij}\middle|\kappa_i^+,\kappa_j^-\right\rangle &= \frac{1}{\pi}\int_0^\pi P(a_{ij}=1|\kappa_i^+,\kappa_j^-,\Delta\theta_{ij})d\Delta\theta_{ij} \\ &= \frac{1}{\pi}\int_0^\pi \frac{1}{1+\chi_{ij}^\beta}d\Delta\theta_{ij} \\ &= \frac{2\mu\kappa_i^+\kappa_j^-}{N}\int_0^{\frac{N}{2\mu\kappa_i^+\kappa_j^-}} \frac{1}{1+\chi_{ij}^\beta}d\chi_{ij} \\ &= {}_2F_1\left(1,\frac{1}{\beta};1+\frac{1}{\beta};-\left[\frac{N}{2\mu\kappa_i^+\kappa_j^-}\right]^\beta\right) \qquad \text{(S23a)} \\ &\simeq \frac{2\pi\mu\kappa_i^+\kappa_j^-}{\beta N\sin(\pi/\beta)} \\ &= \frac{\kappa_i^+\kappa_j^-}{N\langle\kappa\rangle} , \qquad \text{(S23b)}\end{aligned}$$

where $\simeq$ denotes an approximation that becomes exact in the limit $N/(\kappa_i^+\kappa_j^-)\to\infty$ [see Eqs. (S95) and (S104)], and where we set $\mu = \frac{\beta\sin(\pi/\beta)}{2\pi\langle\kappa\rangle}$ in the last equality. Averaging Eq. (S19) over every possible sequence $\boldsymbol{\theta}$ and $\boldsymbol{\kappa}\backslash\{\kappa_i^+\}$ then

yields [note that $H_i^+(z|\boldsymbol{\kappa},\boldsymbol{\theta})$ does not depend on κ_i^-]

$$\begin{aligned}
H_i^+(z|\kappa_i^+) &= \int\cdots\int H_i^+(z|\boldsymbol{\kappa},\boldsymbol{\theta})\prod_{\substack{j=1\\j\neq i}}^{N}\frac{1}{\pi}d\Delta\theta_{ij}\rho(\kappa_j^-,\kappa_j^+)d\kappa_j^-d\kappa_j^+ \\
&= \prod_{\substack{j=1\\j\neq i}}^{N}\Big[\iiint\Big[1-P(a_{ij}=1|\kappa_i^+,\kappa_j^-,\Delta\theta_{ij}) \\
&\qquad\qquad + zP(a_{ij}=1|\kappa_i^+,\kappa_j^-,\Delta\theta_{ij})\Big]\frac{1}{\pi}d\Delta\theta_{ij}\rho(\kappa_j^-,\kappa_j^+)d\kappa_j^-d\kappa_j^+\Big] \\
&= \prod_{\substack{j=1\\j\neq i}}^{N}\Big[1-\langle a_{i\bullet}|\kappa_i^+\rangle + z\langle a_{i\bullet}|\kappa_i^+\rangle\Big] \\
&= \Big[1-\langle a_{i\bullet}|\kappa_i^+\rangle + z\langle a_{i\bullet}|\kappa_i^+\rangle\Big]^{N-1}\ ,
\end{aligned} \tag{S24}$$

where

$$\langle a_{i\bullet}|\kappa_i^+\rangle = \iint\langle a_{ij}|\kappa_i^+,\kappa_j^-\rangle\,\rho(\kappa_j^-,\kappa_j^+)d\kappa_j^-d\kappa_j^+ \simeq \frac{2\pi\mu\langle\kappa\rangle\kappa_i^+}{\beta N\sin(\pi/\beta)} = \frac{\kappa_i^+}{N} \tag{S25}$$

is the average probability for the existence of any outgoing link from node i. From Eq. (S24), we conclude that the out-degree of node i in the random networks ensemble will be distributed according to a binomial distribution with average

$$\langle k_i^+|\kappa_i^+\rangle = (N-1)\langle a_{i\bullet}|\kappa_i^+\rangle \simeq \frac{2\pi\mu\langle\kappa\rangle\kappa_i^+}{\beta\sin(\pi/\beta)} = \kappa_i^+\ , \tag{S26}$$

and variance

$$\mathrm{Var}\left[k_i^+|\kappa_i^+\right] = (N-1)\langle a_{i\bullet}|\kappa_i^+\rangle\left(1-\langle a_{i\bullet}|\kappa_i^+\rangle\right) \simeq \frac{2\pi\mu\langle\kappa\rangle\kappa_i^+}{\beta\sin(\pi/\beta)} = \kappa_i^+\ . \tag{S27}$$

Finally, the average out-degree in the ensemble of random networks is

$$\langle k^+\rangle = \iint\langle k_i^+|\kappa_i^+\rangle\,\rho(\kappa_i^-,\kappa_i^+)d\kappa_i^-d\kappa_i^+ \simeq \frac{2\pi\mu\langle\kappa\rangle^2}{\beta\sin(\pi/\beta)} = \langle\kappa\rangle\ . \tag{S28}$$

As $N/(\kappa^+\kappa^-)\to\infty$, the relative fluctuations around the expected out-degree, $\sqrt{\mathrm{Var}\left[k_i^+|\kappa_i^+\right]}/\langle k_i^+|\kappa_i^+\rangle$, will fall as $1/\sqrt{\kappa_i^+}$ and will become negligible for high out-degree nodes. The binomial distribution obtained in Eq. (S24) can therefore be approximated by a Poisson distribution in this limit

$$H_i^+(z|\kappa_i^+) = \sum_{k_i^+=0}^{N-1}P_i^+(k_i^+|\kappa_i^+)z^{k_i^+} \simeq \Big[1+(z-1)\langle a_{i\bullet}|\kappa_i^+\rangle\Big]^{N-1} = \Big[1+(z-1)\frac{\kappa_i^+}{N}\Big]^{N-1} \simeq \sum_{k_i^+=0}^{\infty}\frac{[\kappa_i^+]^{k_i^+}\mathrm{e}^{-\kappa_i^+}}{k_i^+!}z^{k_i^+}\ , \tag{S29}$$

where we identify

$$P_i^+(k_i^+|\kappa_i^+) \simeq \frac{[\kappa_i^+]^{k_i^+}\mathrm{e}^{-\kappa_i^+}}{k_i^+!} \tag{S30}$$

as the probability for node i with hidden out-degree κ_i^+ to have a degree equal to k_i^+.

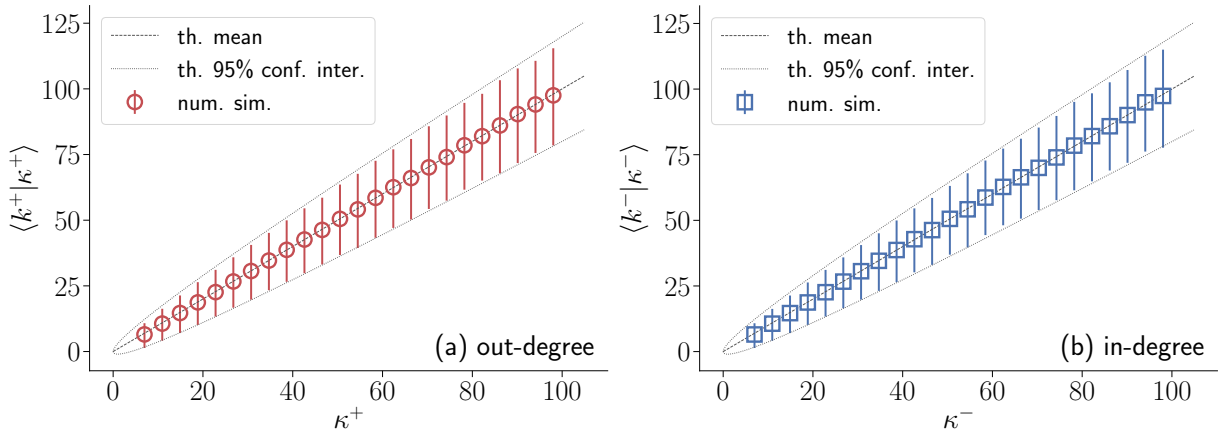

FIG. S1. **Validation of Eqs. (S26), (S27), (S32) and (S33) using numerical simulations.** Both κ^- and κ^+ were independently and identically drawn from the pdf $\rho(\kappa) \propto \kappa^{-2.5}$ with $5 < \kappa < 100$. Symbols show $\langle k^-|\kappa^-\rangle$ and $\langle k^+|\kappa^+\rangle$ estimated from 100 random synthetic networks with $N = 25000$. Only a fraction of the symbols are shown to avoid cluttering the plot. Error bars show the estimated 95% confidence interval.

C. In-degree of nodes

Repeating the same steps from the previous section yields an expression similar to Eq. (S25) for the average probability for the existence of any incoming link into node i

$$\langle a_{\bullet i}|\kappa_i^-\rangle = \iint \langle a_{ji}|\kappa_j^+,\kappa_i^-\rangle\, \rho(\kappa_j^-,\kappa_j^+) d\kappa_j^- d\kappa_j^+ \simeq \frac{2\pi\mu\langle\kappa\rangle\kappa_i^-}{\beta N \sin(\pi/\beta)} = \frac{\kappa_i^-}{N} \tag{S31}$$

an expression similar to Eq. (S26) for the expected in-degree of nodes

$$\langle k_i^-|\kappa_i^-\rangle \simeq \frac{2\pi\mu\langle\kappa\rangle\kappa_i^-}{\beta\sin(\pi/\beta)} = \kappa_i^- \ , \tag{S32}$$

and variance

$$\mathrm{Var}\left[k_i^-|\kappa_i^-\right] \simeq \frac{2\pi\mu\langle\kappa\rangle\kappa_i^-}{\beta\sin(\pi/\beta)} = \kappa_i^- \ , \tag{S33}$$

as well as an expression similar to Eq. (S28) for the ensemble average in-degree

$$\langle k^-\rangle \simeq \frac{2\pi\mu\langle\kappa\rangle^2}{\beta\sin(\pi/\beta)} = \langle\kappa\rangle \ . \tag{S34}$$

We also find that the probability for node i with hidden in-degree κ_i^- to have a degree equal to k_i^- to be

$$P_i^-(k_i^-|\kappa_i^-) \simeq \frac{[\kappa_i^-]^{k_i^-}\,\mathrm{e}^{-\kappa_i^-}}{k_i^-!} \ , \tag{S35}$$

similarly to Eq. (S30).

D. Joint in-/out-degree distribution

Since the existence of links is conditionally independent given the values of the hidden in- and out-degrees, the joint in-/out-degree distribution is

$$\begin{aligned} P(k^-,k^+) &= \iint P^-(k^-|\kappa^-)P^+(k^+|\kappa^+)\rho(\kappa^-,\kappa^+)d\kappa^- d\kappa^+ \\ &\simeq \iint \frac{[\kappa^-]^{k^-}\mathrm{e}^{-\kappa^-}}{k^-!}\frac{[\kappa^+]^{k^+}\mathrm{e}^{-\kappa^+}}{k^+!}\rho(\kappa^-,\kappa^+)d\kappa^- d\kappa^+ \ . \end{aligned} \tag{S36}$$

Hence, the in-degree and out-degree distributions are prescribed by their corresponding marginal pdf of $\rho(\kappa^-,\kappa^+)$ as

$$P^-(k^-)=\sum_{k^+=0}^{\infty}P(k^-,k^+)\simeq\int\frac{\mathrm{e}^{-\kappa^-}[\kappa^-]^{k^-}}{k^-!}\left[\int\rho(\kappa^-,\kappa^+)d\kappa^+\right]d\kappa^- \tag{S37a}$$

$$P^+(k^+)=\sum_{k^-=0}^{\infty}P(k^-,k^+)\simeq\int\frac{\mathrm{e}^{-\kappa^+}[\kappa^+]^{k^+}}{k^+!}\left[\int\rho(\kappa^-,\kappa^+)d\kappa^-\right]d\kappa^+\ , \tag{S37b}$$

and the correlations between k^- and k^+ are governed by the correlations between κ^- and κ^+ encoded in $\rho(k^-,\kappa^+)$.

E. Reciprocal degree of nodes

Let us denote the probability for a reciprocal link to exist between nodes i and j by

$$P(a_{ij}=1,a_{ji}=1|\kappa_i^-,\kappa_i^+,\kappa_j^-,\kappa_j^+,\Delta\theta_{ij})\ , \tag{S38}$$

with the additional assumption that this connection probability is symmetrical, i.e.

$$P(a_{ij}=1,a_{ji}=1|\kappa_i^-,\kappa_i^+,\kappa_j^-,\kappa_j^+,\Delta\theta_{ij})=P(a_{ji}=1,a_{ij}=1|\kappa_j^-,\kappa_j^+,\kappa_i^-,\kappa_i^+,\Delta\theta_{ji})\ . \tag{S39}$$

Following similar steps to that in Sec. S.II B, we define the pgf associated with the reciprocal degree of node i given the sequences $\boldsymbol{\theta}$ and $\boldsymbol{\kappa}$ as

$$\begin{aligned}H_i^{\leftrightarrow}(z|\boldsymbol{\kappa},\boldsymbol{\theta})&=\sum_{k_i^{\leftrightarrow}=0}^{N-1}P_i^{\leftrightarrow}(k_i^{\leftrightarrow}|\boldsymbol{\theta},\boldsymbol{\kappa})z^{k_i^{\leftrightarrow}}\\&=\prod_{\substack{j=1\\j\neq i}}^{N}\Big[1-P(a_{ij}=1,a_{ji}=1|\kappa_i^-,\kappa_i^+,\kappa_j^-,\kappa_j^+,\Delta\theta_{ij})+zP(a_{ij}=1,a_{ji}=1|\kappa_i^-,\kappa_i^+,\kappa_j^-,\kappa_j^+,\Delta\theta_{ij})\Big]\ .\end{aligned} \tag{S40}$$

Hence, general expressions for the expected reciprocal degree of node i and for the ensemble average reciprocal degree are respectively

$$\langle k_i^{\leftrightarrow}|\boldsymbol{\kappa},\boldsymbol{\theta}\rangle=\left.\frac{\partial H_i^{\leftrightarrow}(z|\boldsymbol{\kappa},\boldsymbol{\theta})}{\partial z}\right|_{z=1}=\sum_{\substack{j=1\\j\neq i}}^{N}P(a_{ij}=1,a_{ji}=1|\kappa_i^-,\kappa_i^+,\kappa_j^-,\kappa_j^+,\Delta\theta_{ij})\ , \tag{S41}$$

and

$$\begin{aligned}\langle k^{\leftrightarrow}|\boldsymbol{\kappa},\boldsymbol{\theta}\rangle&=\frac{1}{N}\sum_{i=1}^{N}\langle k_i^{\leftrightarrow}|\boldsymbol{\kappa},\boldsymbol{\theta}\rangle\\&=\frac{1}{N}\sum_{i=1}^{N}\sum_{\substack{j=1\\j\neq i}}^{N}P(a_{ij}=1,a_{ji}=1|\kappa_i^-,\kappa_i^+,\kappa_j^-,\kappa_j^+,\Delta\theta_{ij})\\&=\frac{2}{N}\sum_{i=1}^{N}\sum_{j=i+1}^{N}P(a_{ij}=1,a_{ji}=1|\kappa_i^-,\kappa_i^+,\kappa_j^-,\kappa_j^+,\Delta\theta_{ij})\ .\end{aligned} \tag{S42}$$

Averaging $H_i^{\leftrightarrow}(z|\boldsymbol{\kappa},\boldsymbol{\theta})$ over every possible sequence $\boldsymbol{\theta}$ and $\boldsymbol{\kappa}\backslash\{\kappa_i^-,\kappa_i^+\}$ yields

$$\begin{aligned}
H_i^{\leftrightarrow}(z|\kappa_i^-,\kappa_i^+) &= \int\cdots\int_0^\pi H_i^{\leftrightarrow}(z|\boldsymbol{\kappa},\boldsymbol{\theta})\prod_{\substack{j=1\\j\neq i}}^{N}\frac{1}{\pi}d\Delta\theta_{ij}\rho(\kappa_j^-,\kappa_j^+)d\kappa_j^-d\kappa_j^+\\
&= \prod_{\substack{j=1\\j\neq i}}^{N}\Bigg[\iiint_0^\pi\Big[1-P(a_{ij}=1,a_{ji}=1|\kappa_i^-,\kappa_i^+,\kappa_j^-,\kappa_j^+,\Delta\theta_{ij})\\
&\qquad\qquad + zP(a_{ij}=1,a_{ji}=1|\kappa_i^-,\kappa_i^+,\kappa_j^-,\kappa_j^+,\Delta\theta_{ij})\Big]\frac{1}{\pi}d\Delta\theta_{ij}\rho(\kappa_j^-,\kappa_j^+)d\kappa_j^-d\kappa_j^+\Bigg]\\
&= \prod_{\substack{j=1\\j\neq i}}^{N}\Bigg[\iint\Big[1-\left\langle a_{ij}a_{ji}\middle|\kappa_i^-,\kappa_i^+,\kappa_j^-,\kappa_j^+\right\rangle\\
&\qquad\qquad + z\left\langle a_{ij}a_{ji}\middle|\kappa_i^-,\kappa_i^+,\kappa_j^-,\kappa_j^+\right\rangle\Big]\rho(\kappa_j^-,\kappa_j^+)d\kappa_j^-d\kappa_j^+\Bigg]\\
&= \prod_{\substack{j=1\\j\neq i}}^{N}\left[1-\left\langle a_{i\bullet}a_{\bullet i}\middle|\kappa_i^-,\kappa_i^+\right\rangle + z\left\langle a_{i\bullet}a_{\bullet i}\middle|\kappa_i^-,\kappa_i^+\right\rangle\right]\\
&= \left[1-\left\langle a_{i\bullet}a_{\bullet i}\middle|\kappa_i^-,\kappa_i^+\right\rangle + z\left\langle a_{i\bullet}a_{\bullet i}\middle|\kappa_i^-,\kappa_i^+\right\rangle\right]^{N-1}\,,
\end{aligned}\tag{S43}$$

where

$$\left\langle a_{ij}a_{ji}\middle|\kappa_i^-,\kappa_i^+,\kappa_j^-,\kappa_j^+\right\rangle = \int_0^\pi P(a_{ij}=1,a_{ji}=1|\kappa_i^-,\kappa_i^+,\kappa_j^-,\kappa_j^+,\Delta\theta_{ij})\frac{1}{\pi}d\Delta\theta_{ij}\tag{S44}$$

and

$$\left\langle a_{i\bullet}a_{\bullet i}\middle|\kappa_i^-,\kappa_i^+\right\rangle = \iint\left\langle a_{ij}a_{ji}\middle|\kappa_i^-,\kappa_i^+,\kappa_j^-,\kappa_j^+\right\rangle\rho(\kappa_j^-,\kappa_j^+)d\kappa_j^-d\kappa_j^+\,.\tag{S45}$$

The expected reciprocal degree of node i is then

$$\left\langle k_i^{\leftrightarrow}\middle|\kappa_i^-,\kappa_i^+\right\rangle = (N-1)\left\langle a_{i\bullet}a_{\bullet i}\middle|\kappa_i^-,\kappa_i^+\right\rangle\,.\tag{S46}$$

The ensemble average expected reciprocal degree is

$$\langle k^{\leftrightarrow}\rangle = \iint\left\langle k_i^{\leftrightarrow}\middle|\kappa_i^-,\kappa_i^+\right\rangle\rho(\kappa_i^-,\kappa_i^+)d\kappa_i^-d\kappa_i^+\,.\tag{S47}$$

F. Reciprocity

We are now in a position to combine the results from Sec. S.I with those from the previous subsections to study the reciprocity in the networks generated by the directed-reciprocal $\mathbb{S}^1$ model. The reciprocity is defined as

$$r = \frac{L^{\leftrightarrow}}{L}\,,\tag{S48}$$

where L is the number of links, and $L^{\leftrightarrow}$ is the number of reciprocal links. Note that for r to be such that $0\leq r\leq 1$, each reciprocal connection (e.g. when two nodes are connected by two links in the opposite direction) must contribute 2 to $L^{\leftrightarrow}$. Hence $L^{\leftrightarrow}$ is an even number. Averaging Eq. (S48) over all possible angular positions $\boldsymbol{\theta}$ and hidden in/out-degrees $\boldsymbol{\kappa}$, we get

$$\langle r\rangle = \left\langle\frac{L^{\leftrightarrow}}{L}\right\rangle \approx \frac{\langle L^{\leftrightarrow}\rangle}{\langle L\rangle} = \frac{N\langle k^{\leftrightarrow}\rangle}{N\langle k^{+}\rangle} = \begin{cases}(1+\nu)\langle r|\nu{=}0\rangle - \nu\langle r|\nu{=}{-}1\rangle & -1\leq\nu\leq 0\\ (1-\nu)\langle r|\nu{=}0\rangle + \nu\langle r|\nu{=}1\rangle & 0\leq\nu\leq 1\end{cases}\,,\tag{S49}$$

where we used Eqs. (S44)–(S47), and where we defined the following quantities.

1. $\langle r|\nu{=}1\rangle$ is the expected reciprocity when $\nu = 1$

$$\begin{aligned}\langle r|\nu{=}1\rangle &= \frac{\langle k^{\leftrightarrow}|\nu{=}1\rangle}{\langle k^{+}\rangle} \\ &= \frac{N-1}{\langle k^{+}\rangle}\iiiint \langle a_{ij}a_{ji}\big|\kappa_i^-,\kappa_i^+,\kappa_j^-,\kappa_j^+,\nu{=}1\rangle \\ &\qquad\qquad \times \rho(\kappa_i^-,\kappa_i^+)\rho(\kappa_j^-,\kappa_j^+)d\kappa_i^- d\kappa_i^+ d\kappa_j^- d\kappa_j^+ \end{aligned} \tag{S50}$$

with $\langle a_{ij}a_{ji}\big|\kappa_i^-,\kappa_i^+,\kappa_j^-,\kappa_j^+,\nu{=}1\rangle$ being the expected reciprocal connection probability, Eq. (S44), when $\nu = 1$

$$\begin{aligned}\langle a_{ij}a_{ji}\big|\kappa_i^-,\kappa_i^+,\kappa_j^-,\kappa_j^+,\nu{=}1\rangle &= \frac{1}{\pi}\int_0^\pi \min\left\{P(a_{ij}=1|\kappa_i^+,\kappa_j^-,\Delta\theta_{ij}),P(a_{ji}=1|\kappa_j^+,\kappa_i^-,\Delta\theta_{ij})\right\}d\Delta\theta_{ij} \\ &= \frac{1}{\pi}\int_0^\pi \min\left\{\frac{1}{1+\chi_{ij}^\beta},\frac{1}{1+\chi_{ji}^\beta}\right\}d\Delta\theta_{ij} \\ &= H(1-\xi_{ij})\frac{1}{\pi}\int_0^\pi \frac{1}{1+\chi_{ij}^\beta}d\Delta\theta_{ij} + H(\xi_{ij}-1)\frac{1}{\pi}\int_0^\pi \frac{1}{1+\chi_{ji}^\beta}d\Delta\theta_{ij} \\ &= H(1-\xi_{ij})\frac{2\mu\kappa_i^+\kappa_j^-}{N}\int_0^{\frac{N}{2\mu\kappa_i^+\kappa_j^-}} \frac{1}{1+\chi_{ij}^\beta}d\chi_{ij} \\ &\qquad + H(\xi_{ij}-1)\frac{2\mu\kappa_j^+\kappa_i^-}{N}\int_0^{\frac{N}{2\mu\kappa_j^+\kappa_i^-}} \frac{1}{1+\chi_{ji}^\beta}d\chi_{ij} \\ &= H(1-\xi_{ij})_2F_1\left(1,\frac{1}{\beta};1+\frac{1}{\beta};-\left[\frac{N}{2\mu\kappa_i^+\kappa_j^-}\right]^\beta\right) \\ &\qquad + H(\xi_{ij}-1)_2F_1\left(1,\frac{1}{\beta};1+\frac{1}{\beta};-\left[\frac{N}{2\mu\kappa_j^+\kappa_i^-}\right]^\beta\right) \end{aligned} \tag{S51}$$

$$\begin{aligned} &\simeq H(1-\xi_{ij})\frac{2\pi\mu\kappa_i^+\kappa_j^-}{\beta N\sin(\pi/\beta)} + H(\xi_{ij}-1)\frac{2\pi\mu\kappa_j^+\kappa_i^-}{\beta N\sin(\pi/\beta)} \\ &= H(1-\xi_{ij})\frac{\kappa_i^+\kappa_j^-}{N\langle\kappa\rangle} + H(\xi_{ij}-1)\frac{\kappa_j^+\kappa_i^-}{N\langle\kappa\rangle}\ , \end{aligned} \tag{S52}$$

where we used Eqs. (S95) and (S104), and where we set $\mu = \dfrac{\beta\sin(\pi/\beta)}{2\pi\langle\kappa\rangle}$ and defined $\xi_{ij} = \frac{\kappa_i^+}{\kappa_i^-}\frac{\kappa_j^-}{\kappa_j^+}$.

2. $\langle r|\nu{=}0\rangle$ is the expected reciprocity when $\nu = 0$

$$\begin{aligned}\langle r|\nu{=}0\rangle &= \frac{\langle k^{\leftrightarrow}|\nu{=}0\rangle}{\langle k^{+}\rangle} \\ &= \frac{N-1}{\langle k^{+}\rangle}\iiiint \langle a_{ij}a_{ji}\big|\kappa_i^-,\kappa_i^+,\kappa_j^-,\kappa_j^+,\nu{=}0\rangle \\ &\qquad\qquad \times \rho(\kappa_i^-,\kappa_i^+)\rho(\kappa_j^-,\kappa_j^+)d\kappa_i^- d\kappa_i^+ d\kappa_j^- d\kappa_j^+ \end{aligned} \tag{S53}$$

with

$$\begin{aligned}\left\langle a_{ij}a_{ji}\middle|\kappa_i^-,\kappa_i^+,\kappa_j^-,\kappa_j^+,\nu=0,\xi_{ij}=1\right\rangle &= \frac{1}{\pi}\int_0^\pi P(a_{ij}=1|\kappa_i^+,\kappa_j^-,\Delta\theta_{ij})P(a_{ji}=1|\kappa_j^+,\kappa_i^-,\Delta\theta_{ij})d\Delta\theta_{ij}\\ &= \frac{1}{\pi}\int_0^\pi \frac{1}{1+\chi_{ij}^\beta}\frac{1}{1+\chi_{ji}^\beta}d\Delta\theta_{ij}\\ &= \frac{1}{\pi}\int_0^\pi \frac{1}{\left(1+\chi_{ij}^\beta\right)^2}d\Delta\theta_{ij}\\ &= \frac{2\mu\kappa_i^+\kappa_j^-}{N}\int_0^{\frac{N}{2\mu\kappa_i^+\kappa_j^-}} \frac{1}{\left(1+\chi_{ij}^\beta\right)^2}d\chi_{ij}\\ &= {}_2F_1\left(2,\frac{1}{\beta};1+\frac{1}{\beta};-\left[\frac{N}{2\mu\kappa_i^+\kappa_j^-}\right]^\beta\right) \qquad \text{(S54)}\\ &\simeq \frac{2\pi\mu\kappa_i^+\kappa_j^-}{\beta N\sin(\pi/\beta)}\left(1-\frac{1}{\beta}\right)\\ &= \frac{\kappa_i^+\kappa_j^-}{N\langle\kappa\rangle}\left(1-\frac{1}{\beta}\right) \qquad \text{(S55)}\end{aligned}$$

when $\xi_{ij}=\frac{\kappa_i^+}{\kappa_i^-}\frac{\kappa_j^-}{\kappa_j^+}=1$, and

$$\begin{aligned}\left\langle a_{ij}a_{ji}\middle|\kappa_i^-,\kappa_i^+,\kappa_j^-,\kappa_j^+,\nu=0,\xi_{ij}\neq 1\right\rangle &= \frac{1}{\pi}\int_0^\pi P(a_{ij}=1|\kappa_i^+,\kappa_j^-,\Delta\theta_{ij})P(a_{ji}=1|\kappa_j^+,\kappa_i^-,\Delta\theta_{ij})d\Delta\theta_{ij}\\ &= \frac{1}{\pi}\int_0^\pi \frac{1}{1+\chi_{ij}^\beta}\frac{1}{1+\chi_{ji}^\beta}d\Delta\theta_{ij}\\ &= \frac{1}{\pi}\int_0^\pi \frac{1}{1+\chi_{ij}^\beta}\frac{1}{1+\xi_{ij}^\beta\chi_{ij}^\beta}d\Delta\theta_{ij}\\ &= \frac{2\mu\kappa_i^+\kappa_j^-}{N}\int_0^{\frac{N}{2\mu\kappa_i^+\kappa_j^-}} \frac{1}{1+\chi_{ij}^\beta}\frac{1}{1+\xi_{ij}^\beta\chi_{ij}^\beta}d\chi_{ij}\\ &= \frac{1}{1-\xi_{ij}^\beta}{}_2F_1\left(1,\frac{1}{\beta};1+\frac{1}{\beta};-\left[\frac{N}{2\mu\kappa_i^+\kappa_j^-}\right]^\beta\right)\\ &\qquad -\frac{\xi_{ij}^\beta}{1-\xi_{ij}^\beta}{}_2F_1\left(1,\frac{1}{\beta};1+\frac{1}{\beta};-\left[\frac{N}{2\mu\kappa_j^+\kappa_i^-}\right]^\beta\right) \qquad \text{(S56)}\\ &\simeq \frac{2\pi\mu\kappa_i^+\kappa_j^-}{\beta N\sin(\pi/\beta)}\frac{1-\xi_{ij}^{\beta-1}}{1-\xi_{ij}^\beta}\\ &= \frac{\kappa_i^+\kappa_j^-}{N\langle\kappa\rangle}\frac{1-\xi_{ij}^{\beta-1}}{1-\xi_{ij}^\beta} \qquad \text{(S57)}\end{aligned}$$

otherwise. In the last two equations, we again set $\mu=\frac{\beta\sin(\pi/\beta)}{2\pi\langle\kappa\rangle}$, and used Eqs. (S101), (S104) and (S105).

3. $\langle r|\nu=-1\rangle$ is the expected reciprocity when $\nu=-1$

$$\begin{aligned}\langle r|\nu=-1\rangle &= \frac{\langle k^{\leftrightarrow}|\nu=-1\rangle}{\langle k^+\rangle}\\ &= \frac{N-1}{\langle k^+\rangle}\iiiint\left\langle a_{ij}a_{ji}\middle|\kappa_i^-,\kappa_i^+,\kappa_j^-,\kappa_j^+,\nu=-1\right\rangle\\ &\qquad\times\rho(\kappa_i^-,\kappa_i^+)\rho(\kappa_j^-,\kappa_j^+)d\kappa_i^-d\kappa_i^+d\kappa_j^-d\kappa_j^+ \qquad \text{(S58)}\end{aligned}$$

with

$$\begin{aligned}\left\langle a_{ij}a_{ji}\middle|\kappa_i^-,\kappa_i^+,\kappa_j^-,\kappa_j^+,\nu=-1\right\rangle &= \frac{1}{\pi}\int_0^{\pi}\Big[P(a_{ij}=1|\kappa_i^+,\kappa_j^-,\Delta\theta_{ij})+P(a_{ji}=1|\kappa_j^+,\kappa_i^-,\Delta\theta_{ij})-1\Big]\\ &\qquad\times H\Big(P(a_{ij}=1|\kappa_i^+,\kappa_j^-,\Delta\theta_{ij})+P(a_{ji}=1|\kappa_j^+,\kappa_i^-,\Delta\theta_{ij})-1\Big)d\Delta\theta_{ij}\\ &= \frac{1}{\pi}\int_0^{\Delta\theta_{ij}^{\rm c}}\Big[P(a_{ij}=1|\kappa_i^+,\kappa_j^-,\Delta\theta_{ij})+P(a_{ji}=1|\kappa_j^+,\kappa_i^-,\Delta\theta_{ij})-1\Big]d\Delta\theta_{ij}\\ &= \frac{1}{\pi}\int_0^{\Delta\theta_{ij}^{\rm c}}\frac{d\Delta\theta_{ij}}{1+\chi_{ij}^{\beta}}+\frac{1}{\pi}\int_0^{\Delta\theta_{ij}^{\rm c}}\frac{d\Delta\theta_{ij}}{1+\chi_{ji}^{\beta}}-\frac{1}{\pi}\int_0^{\Delta\theta_{ij}^{\rm c}}d\Delta\theta_{ij}\\ &= \frac{2\mu\kappa_i^+\kappa_j^-}{N}\int_0^{\frac{N\Delta\theta_{ij}^{\rm c}}{2\pi\mu\kappa_i^+\kappa_j^-}}\frac{d\chi_{ij}}{1+\chi_{ij}^{\beta}}+\frac{2\mu\kappa_j^+\kappa_i^-}{N}\int_0^{\frac{N\Delta\theta_{ij}^{\rm c}}{2\pi\mu\kappa_j^+\kappa_i^-}}\frac{d\chi_{ji}}{1+\chi_{ji}^{\beta}}-\frac{\Delta\theta_{ij}^{\rm c}}{\pi}\\ &= \frac{\Delta\theta_{ij}^{\rm c}}{\pi}\left[{}_2F_1\left(1,\frac{1}{\beta};1+\frac{1}{\beta};-\left[\frac{N\Delta\theta_{ij}^{\rm c}}{2\pi\mu\kappa_i^+\kappa_j^-}\right]^{\beta}\right)\right.\\ &\qquad\left.+{}_2F_1\left(1,\frac{1}{\beta};1+\frac{1}{\beta};-\left[\frac{N\Delta\theta_{ij}^{\rm c}}{2\pi\mu\kappa_j^+\kappa_i^-}\right]^{\beta}\right)-1\right]\end{aligned}\tag{S59}$$

where we used Eq. (S95), and where $\Delta\theta_{ij}^{\rm c}$ is the solution of

$$P(a_{ij}=1|\kappa_i^+,\kappa_j^-,\Delta\theta_{ij}^{\rm c})+P(a_{ji}=1|\kappa_j^+,\kappa_i^-,\Delta\theta_{ij}^{\rm c})=1\ .\tag{S60}$$

To explore the limit $N\to\infty$ such that $N/(\kappa_i^+\kappa_j^-)\to\infty$ and $N/(\kappa_j^+\kappa_i^-)\to\infty$, we note that

$$\begin{aligned}1 &= P(a_{ij}=1|\kappa_i^+,\kappa_j^-,\Delta\theta_{ij}^{\rm c})+P(a_{ji}=1|\kappa_j^+,\kappa_i^-,\Delta\theta_{ij}^{\rm c})\\ &= \frac{1}{1+\left[\frac{N\Delta\theta_{ij}^{\rm c}}{2\pi\mu\kappa_i^+\kappa_j^-}\right]^{\beta}}+\frac{1}{1+\left[\frac{N\Delta\theta_{ij}^{\rm c}}{2\pi\mu\kappa_j^+\kappa_i^-}\right]^{\beta}}\\ &\simeq \frac{[2\pi\mu\kappa_i^+\kappa_j^-]^{\beta}}{[N\Delta\theta_{ij}^{\rm c}]^{\beta}}+\frac{[2\pi\mu\kappa_j^+\kappa_i^-]^{\beta}}{[N\Delta\theta_{ij}^{\rm c}]^{\beta}}\ ,\end{aligned}\tag{S61}$$

and thus

$$\Delta\theta_{ij}^{\rm c}\simeq\frac{2\pi\mu}{N}\left[[\kappa_i^+\kappa_j^-]^{\beta}+[\kappa_j^+\kappa_i^-]^{\beta}\right]^{\frac{1}{\beta}}=\frac{2\pi\mu\kappa_i^+\kappa_j^-}{N}\left[1+\xi_{ij}^{-\beta}\right]^{\frac{1}{\beta}}\ .\tag{S62}$$

Equation (S59) then becomes

$$\begin{aligned}\left\langle a_{ij}a_{ji}\middle|\kappa_i^-,\kappa_i^+,\kappa_j^-,\kappa_j^+,\nu=-1\right\rangle &\simeq \frac{2\mu\kappa_i^+\kappa_j^-}{N}\left[1+\xi_{ij}^{-\beta}\right]^{\frac{1}{\beta}}\left[{}_2F_1\left(1,\frac{1}{\beta};1+\frac{1}{\beta};-1-\xi_{ij}^{-\beta}\right)\right.\\ &\qquad\left.+{}_2F_1\left(1,\frac{1}{\beta};1+\frac{1}{\beta};-1-\xi_{ij}^{\beta}\right)-1\right]\\ &= \frac{\kappa_i^+\kappa_j^-}{N\langle\kappa\rangle}\frac{\sin(\pi/\beta)}{\pi/\beta}\left[1+\xi_{ij}^{-\beta}\right]^{\frac{1}{\beta}}\left[{}_2F_1\left(1,\frac{1}{\beta};1+\frac{1}{\beta};-1-\xi_{ij}^{-\beta}\right)\right.\\ &\qquad\left.+{}_2F_1\left(1,\frac{1}{\beta};1+\frac{1}{\beta};-1-\xi_{ij}^{\beta}\right)-1\right]\ ,\end{aligned}\tag{S63}$$

where we set $\mu=\dfrac{\beta\sin(\pi/\beta)}{2\pi\langle\kappa\rangle}$.

S.III. NETWORK DATASETS

The network datasets used in the article have been made publicly available by the original authors and were downloaded from The Netzschleuder network catalogue and repository (https://networks.skewed.de). The name of each dataset is listed below:

7th_graders
add_health_comm15
add_health_comm27
add_health_comm28
add_health_comm33
add_health_comm40
add_health_comm41
add_health_comm50
add_health_comm61
add_health_comm62
add_health_comm68
add_health_comm73
add_health_comm75
add_health_comm79
add_health_comm81
add_health_comm83
add_health_comm84
advogato
anybeat
bison
bitcoin_alpha
bitcoin_trust
caida_as_20040105
caida_as_20040202
caida_as_20040301
caida_as_20040405
caida_as_20040503
caida_as_20040607
caida_as_20040705
caida_as_20040802
caida_as_20040906
caida_as_20041004
caida_as_20041101
caida_as_20041206
caida_as_20050103
caida_as_20050207
caida_as_20050307
caida_as_20050404
caida_as_20050502
caida_as_20050606
caida_as_20050704
caida_as_20050801
caida_as_20050905
caida_as_20051003
caida_as_20051107
caida_as_20051205
caida_as_20060102
caida_as_20060109
caida_as_20060116
caida_as_20060123
caida_as_20060130
caida_as_20060206
caida_as_20060213
caida_as_20060220
caida_as_20060227
caida_as_20060306
caida_as_20060313
caida_as_20060320
caida_as_20060327
caida_as_20060403
caida_as_20060410
caida_as_20060417
caida_as_20060424
caida_as_20060501
caida_as_20060508
caida_as_20060515
caida_as_20060522
caida_as_20060529
caida_as_20060605
caida_as_20060612
caida_as_20060619
caida_as_20060626
caida_as_20060703
caida_as_20060710
caida_as_20060717
caida_as_20060724
caida_as_20060731
caida_as_20060807
caida_as_20060814
caida_as_20060821
caida_as_20060828
caida_as_20060904
caida_as_20060911
caida_as_20060918
caida_as_20060925
caida_as_20061002
caida_as_20061009
caida_as_20061016
caida_as_20061023
caida_as_20061030
caida_as_20061106
caida_as_20061113
caida_as_20061120
caida_as_20061127
caida_as_20061204
caida_as_20061211
caida_as_20061218
caida_as_20061225
caida_as_20070101
caida_as_20070108
caida_as_20070115
caida_as_20070122
caida_as_20070129
caida_as_20070205
caida_as_20070212
caida_as_20070219
caida_as_20070226
caida_as_20070305
caida_as_20070312
caida_as_20070423
caida_as_20070917
cattle
celegans_2019_hermaphrodite_chemical
celegans_2019_hermaphrodite_chemical_corrected
celegans_2019_hermaphrodite_chemical_synapse
celegans_2019_male_chemical
celegans_2019_male_chemical_corrected
celegans_2019_male_chemical_synapse
celegansneural
chess
chicago_road
cintestinalis
college_freshmen
copenhagen_calls
copenhagen_sms
cora
dblp_cite
dutch_school_klas12b-net-1
dutch_school_klas12b-net-2
dutch_school_klas12b-net-3
dutch_school_klas12b-net-3m
dutch_school_klas12b-net-4
dutch_school_klas12b-net-4m
dutch_school_klas12b-primary
ecoli_transcription_v1_0
ecoli_transcription_v1_1
email_company
faa_routes
fao_trade
foodweb_baywet
foodweb_little_rock
fresh_webs_AkatoreA
fresh_webs_AkatoreB
fresh_webs_Berwick
fresh_webs_Blackrock
fresh_webs_Broad
fresh_webs_Canton
fresh_webs_Catlins
fresh_webs_Coweeta1
fresh_webs_Coweeta17
fresh_webs_DempstersAu
fresh_webs_DempstersSp
fresh_webs_DempstersSu
fresh_webs_German
fresh_webs_Healy
fresh_webs_Kyeburn
fresh_webs_LilKyeburn
fresh_webs_Martins
fresh_webs_Narrowdale
fresh_webs_NorthCol
fresh_webs_Powder
fresh_webs_Stony
fresh_webs_SuttonAu
fresh_webs_SuttonSp
fresh_webs_SuttonSu
fresh_webs_Troy
fresh_webs_Venlaw
freshman_t0
freshman_t2
freshman_t3
freshman_t5

freshman_t6
freshmen_t0
freshmen_t2
freshmen_t3
freshmen_t5
freshmen_t6
genetic_multiplex_Arabidopsis
genetic_multiplex_Bos_Multiplex_Genetic
genetic_multiplex_Candida
genetic_multiplex_Celegans
genetic_multiplex_DanioRerio
genetic_multiplex_Drosophila
genetic_multiplex_Gallus
genetic_multiplex_HepatitusCVirus
genetic_multiplex_HumanHIV1
genetic_multiplex_HumanHerpes4
genetic_multiplex_Mus
genetic_multiplex_Oryctolagus
genetic_multiplex_Plasmodium
genetic_multiplex_Rattus
genetic_multiplex_Sacchpomb
genetic_multiplex_Xenopus
gnutella_04
gnutella_06
gnutella_08
gnutella_09
gnutella_25
hens
high_tech_company
highschool
inploid
interactome_figeys
interactome_stelzl
jdk
jung
law_firm
macaques
messal_shale
moreno_sheep
moreno_taro
openflights
packet_delays
physician_trust
polblogs
qa_user_mathoverflow_a2q
qa_user_mathoverflow_c2a
qa_user_mathoverflow_c2q
residence_hall
rhesus_monkey
sp_high_school_diaries
sp_high_school_survey
un_migrations
uni_email
us_agencies_alabama
us_agencies_alaska
us_agencies_arizona
us_agencies_arkansas
us_agencies_california
us_agencies_colorado
us_agencies_connecticut
us_agencies_delaware
us_agencies_florida
us_agencies_georgia
us_agencies_hawaii
us_agencies_idaho
us_agencies_illinois
us_agencies_indiana
us_agencies_iowa
us_agencies_kansas
us_agencies_kentucky
us_agencies_louisiana
us_agencies_maine
us_agencies_maryland
us_agencies_massachusetts
us_agencies_michigan
us_agencies_minnesota
us_agencies_mississippi
us_agencies_missouri
us_agencies_montana
us_agencies_nebraska
us_agencies_nevada
us_agencies_newhampshire
us_agencies_newjersey
us_agencies_newmexico
us_agencies_newyork
us_agencies_northcarolina
us_agencies_northdakota
us_agencies_ohio
us_agencies_oklahoma
us_agencies_oregon
us_agencies_pennsylvania
us_agencies_rhodeisland
us_agencies_southcarolina
us_agencies_southdakota
us_agencies_tennessee
us_agencies_texas
us_agencies_utah
us_agencies_vermont
us_agencies_virginia
us_agencies_washington
us_agencies_westvirginia
us_agencies_wisconsin
us_agencies_wyoming
us_air_traffic
webkb_webkb_cornell_link1
webkb_webkb_texas_link1
webkb_webkb_washington_link1
webkb_webkb_wisconsin_link1
wiki_talk_br
wiki_talk_cy
wiki_talk_eo
wiki_talk_gl
wiki_talk_ht
wiki_talk_nds
wiki_talk_oc
wikipedia_link_si
word_adjacency_darwin
word_adjacency_french
word_adjacency_japanese
word_adjacency_spanish
yeast_transcription

S.IV. INFERENCE ALGORITHM

The inference algorithm used in the main text is an adaptation of the parameter inference procedure of the embedding algorithm introduced in Ref. [1]. Its objective is to infer the $2N+2$ parameters $\boldsymbol{\kappa} = \kappa_1^-, \kappa_1^+, \ldots, \kappa_N^-, \kappa_N^+$, β and ν so that the directed-reciprocal $\mathbb{S}^1$ model will reproduce, on average, the joint in/out-degree sequence, the reciprocity and the density of triangles of an original real directed network ($2N+2$ constraints).

Note that, contrary to the embedding algorithm introduced in Ref. [1], the inference algorithm does not aim to infer the angular positions, $\boldsymbol{\theta}$; the aforementioned $2N+2$ parameters are therefore inferred when averaging over all possible angular positions.

A. Inputs

The following $2N+2$ constraints are measured on an original real directed network and used as inputs for the inference algorithm.

1. The joint in/out-degree sequence $\boldsymbol{k} = \{k_1^-, k_1^+, \ldots, k_N^-, k_N^+\}$, where

$$k_i^- = |\partial_i^-| \tag{S64a}$$
$$k_i^+ = |\partial_i^+| \ , \tag{S64b}$$

   and where ∂_i^- (∂_i^+) is the set of in-neighbors (out-neighbors) of node i in the original real directed network.

2. The reciprocity r^{obs} computed as

$$r^{\text{obs}} = \frac{L^{\leftrightarrow}}{L} = \frac{\sum_{i=1}^N |\partial_i^- \cap \partial_i^+|}{\sum_{i=1}^N |\partial_i^+|} \ , \tag{S65}$$

   where $|\partial_i^- \cap \partial_i^+|$ counts the number of neighbors with which node i shares both possible directed links (i.e. reciprocal connection).

3. The density of triangles, $\bar{c}_{\text{obs}}$, as measured by the average undirected local clustering coefficient

$$\bar{c}_{\text{undir}}^{\text{obs}} = \frac{1}{N_{>1}} \sum_{i=1}^N c_i = \frac{1}{N_{>1}} \sum_{i=1}^N \frac{2T_i}{|\partial_i^- \cup \partial_i^+|\big(|\partial_i^- \cup \partial_i^+| - 1\big)} \tag{S66}$$

   where T_i is the number of triangles to which node i participates, where the quantity $|\partial_i^- \cup \partial_i^+|$ corresponds to the degree of node i in the undirected version of the network, and where $N_{>1}$ is the number of nodes for which $|\partial_i^- \cup \partial_i^+| > 1$. Note that we set $c_i = 0$ for the $N - N_{>1}$ nodes for which $|\partial_i^- \cup \partial_i^+| < 2$.

B. Inferring the hidden in/out-degrees

This subroutine assumes that a maximal deviation tolerance $\varepsilon_{\text{tol}}^{\max}$ and the parameter β have both been assigned some value (e.g. $\varepsilon_{\text{tol}}^{\max} = 0.01$), and uses

$$\mu = \frac{\beta \sin\left(\frac{\pi}{\beta}\right)}{2\pi \langle k^+ \rangle} \tag{S67}$$

where $\langle k^+ \rangle$ is the average out-degree (or equivalently average in-degree) in the original real directed network

$$\langle k^+ \rangle = \frac{1}{N} \sum_{i=1}^N k_i^+ = \frac{1}{N} \sum_{i=1}^N k_i^- \ . \tag{S68}$$

1. *Initialize* the hidden in/out-degrees by setting $\kappa_i^- = k_i^-$ and $\kappa_i^+ = k_i^+$ for all $i = 1, \ldots, N$.

2. *Compute expected in/out-degrees* as

$$\langle k_i^- | \boldsymbol{\kappa} \rangle = \sum_{\substack{j=1 \\ j\neq i}}^{N} \langle a_{ji} | \kappa_j^+, \kappa_i^- \rangle = \sum_{\substack{j=1 \\ j\neq i}}^{N} {}_2F_1\left(1, \frac{1}{\beta}; 1+\frac{1}{\beta}; -\left[\frac{N}{2\mu\kappa_j^+\kappa_i^-}\right]^\beta\right) , \tag{S69a}$$

$$\langle k_i^+ | \boldsymbol{\kappa} \rangle = \sum_{\substack{j=1 \\ j\neq i}}^{N} \langle a_{ij} | \kappa_i^+, \kappa_j^- \rangle = \sum_{\substack{j=1 \\ j\neq i}}^{N} {}_2F_1\left(1, \frac{1}{\beta}; 1+\frac{1}{\beta}; -\left[\frac{N}{2\mu\kappa_i^+\kappa_j^-}\right]^\beta\right) , \tag{S69b}$$

for all $i = 1, \ldots, N$. Equations (S69a) and (S69b) are obtained by averaging Eq. (S20) (and its equivalent for in-degrees) over all angular positions, combined with Eq. (S23a).

3. *Compute the largest deviation*, $\varepsilon^{\max}$, between the expected in/out-degrees and the in/out-degrees in the original network as

$$\varepsilon^{\max} = \max\left\{ \max\left\{ \left| \langle k_i^- | \boldsymbol{\kappa} \rangle - k_i^- \right|, \left| \langle k_i^+ | \boldsymbol{\kappa} \rangle - k_i^+ \right| \right\} : i = 1, \ldots, N \right\} . \tag{S70}$$

The hidden in/out-degrees have converged to acceptable values if $\varepsilon^{\max} < \varepsilon_{\text{tol}}^{\max}$ and we proceed to step 6. Otherwise, they require more refinement and we proceed to step 4.

4. *Update the hidden in/out-degrees* according to

$$\kappa_i^- \leftarrow \left| \kappa_i^- + \left[k_i^- - \langle k_i^- | \boldsymbol{\kappa} \rangle \right] u^- \right| , \tag{S71a}$$

$$\kappa_i^+ \leftarrow \left| \kappa_i^+ + \left[k_i^+ - \langle k_i^+ | \boldsymbol{\kappa} \rangle \right] u^+ \right| , \tag{S71b}$$

for all $i = 1, \ldots, N$, and where $u^- \sim \text{Uniform}(0, 1)$ and $u^+ \sim \text{Uniform}(0, 1)$. The random variables prevent the subroutine from getting trapped in a local minimum.

5. *Proceed to step 2* using the updated values for the hidden in/out-degrees.

6. *Compute the expected in-degree and out-degree* as

$$\langle k^- | \boldsymbol{\kappa} \rangle = \frac{1}{N} \sum_{i=1}^{N} \langle k_i^- | \boldsymbol{\kappa} \rangle \tag{S72a}$$

$$\langle k^+ | \boldsymbol{\kappa} \rangle = \frac{1}{N} \sum_{i=1}^{N} \langle k_i^+ | \boldsymbol{\kappa} \rangle \ . \tag{S72b}$$

Equations (S72a) and (S72b) are obtained by averaging Eq. (S22) (and its equivalent for in-degrees) over all angular positions. Note that $\langle k^- | \boldsymbol{\kappa} \rangle$ and $\langle k^+ | \boldsymbol{\kappa} \rangle$ will be equal up to the numerical error induced by $\varepsilon_{\text{tol}}^{\max}$.

C. Inferring parameter ν

This subroutine assumes that the parameter β has been assigned some value, and uses the parameter μ, the hidden in/out-degrees, $\boldsymbol{\kappa} = \kappa_1^-, \kappa_1^+, \ldots, \kappa_N^-, \kappa_N^+$, as well as the expected out-degree, $\langle k^+ | \boldsymbol{\kappa} \rangle$, computed in Sec. S.IV B.

The expected reciprocity in the directed-reciprocal $\mathbb{S}^1$ model is computed as

$$\langle r | \boldsymbol{\kappa} \rangle = \left\langle \frac{L^{\leftrightarrow}}{L} \middle| \boldsymbol{\kappa} \right\rangle \approx \frac{\langle L^{\leftrightarrow} | \boldsymbol{\kappa} \rangle}{\langle L | \boldsymbol{\kappa} \rangle} = \frac{N \langle k^{\leftrightarrow} | \boldsymbol{\kappa} \rangle}{N \langle k^+ | \boldsymbol{\kappa} \rangle} , \tag{S73}$$

where $\langle k^+ | \boldsymbol{\kappa} \rangle$ is taken from Eq. (S72b) and $\langle k^{\leftrightarrow} | \boldsymbol{\kappa} \rangle$ is computed by averaging Eq. (S42) over all angular positions. Equation (S73) then takes a similar form as Eq. (S49) and becomes

$$\langle r | \boldsymbol{\kappa} \rangle \approx \begin{cases} (1+\nu) \langle r | \boldsymbol{\kappa}, \nu = 0 \rangle - \nu \langle r | \boldsymbol{\kappa}, \nu = -1 \rangle & \text{if } -1 \leq \nu \leq 0 \\ \\ (1-\nu) \langle r | \boldsymbol{\kappa}, \nu = 0 \rangle + \nu \langle r | \boldsymbol{\kappa}, \nu = 1 \rangle & \text{if } \ \ 0 \leq \nu \leq 1 \end{cases} , \tag{S74}$$

and the inferred value of ν is obtained such that $\langle r|\boldsymbol{\kappa}\rangle = r^{\rm obs}$. This subroutine computes $\langle r|\boldsymbol{\kappa}\rangle$ and ν via the following steps.

1. *Compute the expected reciprocity when* $\nu = 1$ using

$$\langle r|\boldsymbol{\kappa},\nu{=}1\rangle = \frac{\langle k^{\leftrightarrow}|\boldsymbol{\kappa},\nu{=}1\rangle}{\langle k^{+}|\boldsymbol{\kappa}\rangle} = \frac{2}{N\langle k^{+}|\boldsymbol{\kappa}\rangle}\sum_{i=1}^{N}\sum_{j=i+1}^{N}\left\langle a_{ij}a_{ji}\middle|\kappa_i^-,\kappa_i^+,\kappa_j^-,\kappa_j^+,\nu{=}1\right\rangle \tag{S75a}$$

where

$$\left\langle a_{ij}a_{ji}\middle|\kappa_i^-,\kappa_i^+,\kappa_j^-,\kappa_j^+,\nu{=}1\right\rangle = \begin{cases} {}_2F_1\left(1,\frac{1}{\beta};1+\frac{1}{\beta};-\left[\frac{N}{2\mu\kappa_i^+\kappa_j^-}\right]^{\beta}\right) & \text{if } \xi_{ij} = \frac{\kappa_i^+}{\kappa_i^-}\frac{\kappa_j^-}{\kappa_j^+} < 1 \\ {}_2F_1\left(1,\frac{1}{\beta};1+\frac{1}{\beta};-\left[\frac{N}{2\mu\kappa_j^+\kappa_i^-}\right]^{\beta}\right) & \text{if } \xi_{ij} = \frac{\kappa_i^+}{\kappa_i^-}\frac{\kappa_j^-}{\kappa_j^+} > 1 \end{cases} . \tag{S75b}$$

Equations (S75a) and (S75b) are obtained by averaging Eq. (S42) over all angular positions, combined with Eq. (S51).

2. *Compute the expected reciprocity when* $\nu = 0$ using

$$\langle r|\boldsymbol{\kappa},\nu{=}0\rangle = \frac{\langle k^{\leftrightarrow}|\boldsymbol{\kappa},\nu{=}0\rangle}{\langle k^{+}|\boldsymbol{\kappa}\rangle} = \frac{2}{N\langle k^{+}|\boldsymbol{\kappa}\rangle}\sum_{i=1}^{N}\sum_{j=i+1}^{N}\left\langle a_{ij}a_{ji}\middle|\kappa_i^-,\kappa_i^+,\kappa_j^-,\kappa_j^+,\nu{=}0\right\rangle \tag{S76a}$$

where

$$\left\langle a_{ij}a_{ji}\middle|\kappa_i^-,\kappa_i^+,\kappa_j^-,\kappa_j^+,\nu{=}0\right\rangle = \begin{cases} \frac{1}{1-\xi_{ij}^{\beta}}{}_2F_1\left(1,\frac{1}{\beta};1+\frac{1}{\beta};-\left[\frac{N}{2\mu\kappa_i^+\kappa_j^-}\right]^{\beta}\right) & \\ \quad -\frac{\xi_{ij}^{\beta}}{1-\xi_{ij}^{\beta}}{}_2F_1\left(1,\frac{1}{\beta};1+\frac{1}{\beta};-\left[\frac{N}{2\mu\kappa_j^+\kappa_i^-}\right]^{\beta}\right) & \text{if } \xi_{ij} = \frac{\kappa_i^+}{\kappa_i^-}\frac{\kappa_j^-}{\kappa_j^+} \neq 1 \\ {}_2F_1\left(2,\frac{1}{\beta};1+\frac{1}{\beta};-\left[\frac{N}{2\mu\kappa_i^+\kappa_j^-}\right]^{\beta}\right) & \text{if } \xi_{ij} = \frac{\kappa_i^+}{\kappa_i^-}\frac{\kappa_j^-}{\kappa_j^+} = 1 \end{cases} . \tag{S76b}$$

Equations (S76a) and (S76b) are obtained by averaging Eq. (S42) over all angular positions, combined with Eqs. (S54) and (S56).

3. *Compute the expected reciprocity when* $\nu = -1$ using

$$\langle r|\boldsymbol{\kappa},\nu{=}0\rangle = \frac{\langle k^{\leftrightarrow}|\boldsymbol{\kappa},\nu{=}-1\rangle}{\langle k^{+}|\boldsymbol{\kappa}\rangle} = \frac{2}{N\langle k^{+}|\boldsymbol{\kappa}\rangle}\sum_{i=1}^{N}\sum_{j=i+1}^{N}\left\langle a_{ij}a_{ji}\middle|\kappa_i^-,\kappa_i^+,\kappa_j^-,\kappa_j^+,\nu{=}-1\right\rangle \tag{S77a}$$

where

$$\left\langle a_{ij}a_{ji}\middle|\kappa_i^-,\kappa_i^+,\kappa_j^-,\kappa_j^+,\nu{=}-1\right\rangle = \frac{\Delta\theta_{ij}^{\rm c}}{\pi}\left[{}_2F_1\left(1,\frac{1}{\beta};1+\frac{1}{\beta};-\left[\frac{N\Delta\theta_{ij}^{\rm c}}{2\pi\mu\kappa_i^+\kappa_j^-}\right]^{\beta}\right)\right. \\ \left. + {}_2F_1\left(1,\frac{1}{\beta};1+\frac{1}{\beta};-\left[\frac{N\Delta\theta_{ij}^{\rm c}}{2\pi\mu\kappa_j^+\kappa_i^-}\right]^{\beta}\right) - 1\right] \tag{S77b}$$

and where $\Delta\theta_{ij}^{\rm c} \in [0,\pi]$ is the solution of

$$P(a_{ij}=1|\kappa_i^+,\kappa_j^-,\Delta\theta_{ij}^{\rm c}) + P(a_{ji}=1|\kappa_j^+,\kappa_i^-,\Delta\theta_{ij}^{\rm c}) = 1 . \tag{S77c}$$

Equations (S77a)–(S77c) are obtained by averaging Eq. (S42) over all angular positions, combined with Eq. (S59).

4. *Compute the inferred value of* ν according to

$$\nu = \begin{cases} \dfrac{r^{\rm obs} - \langle r|\boldsymbol{\kappa}, \nu{=}0\rangle}{\langle r|\boldsymbol{\kappa}, \nu{=}{-}1\rangle + \langle r|\boldsymbol{\kappa}, \nu{=}0\rangle} & \text{if } r^{\rm obs} < \langle r|\boldsymbol{\kappa}, \nu{=}0\rangle \\ \dfrac{r^{\rm obs} - \langle r|\boldsymbol{\kappa}, \nu{=}0\rangle}{\langle r|\boldsymbol{\kappa}, \nu{=}1\rangle - \langle r|\boldsymbol{\kappa}, \nu{=}0\rangle} & \text{if } r^{\rm obs} > \langle r|\boldsymbol{\kappa}, \nu{=}0\rangle \end{cases} . \tag{S78}$$

D. Estimating the expected density of triangles

This subroutine assumes that the parameter β has been assigned some value, uses the parameter ν computed in Sec. S.IV C, and uses the parameter μ as well as the hidden in/out-degrees, $\boldsymbol{\kappa} = \kappa_1^-, \kappa_1^+, \ldots, \kappa_N^-, \kappa_N^+$ computed in Sec. S.IV B.

The density of triangles is quantified using the average *undirected* local clustering coefficient, that is the average local clustering coefficient measured on the *undirected projection* of the directed network. This projection is specified via its adjacency matrix, $\tilde{\mathbf{A}}$, whose elements are $\tilde{a}_{ij} = \max(a_{ij}, a_{ji})$. In other words, two nodes are connected in the projection if they are connected by at least one directed link, which occurs with probability

$$\begin{aligned} P(\tilde{a}_{ij} = 1|\kappa_i^-, \kappa_i^+, \kappa_j^-, \kappa_j^+, \Delta\theta_{ij}) = {} & P_{ij}(a_{ij} = 1, a_{ji} = 0|\kappa_i^-, \kappa_i^+, \kappa_j^-, \kappa_j^+, \Delta\theta_{ij}) \\ & + P_{ij}(a_{ij} = 0, a_{ji} = 1|\kappa_i^-, \kappa_i^+, \kappa_j^-, \kappa_j^+, \Delta\theta_{ij}) \\ & + P_{ij}(a_{ij} = 1, a_{ji} = 1|\kappa_i^-, \kappa_i^+, \kappa_j^-, \kappa_j^+, \Delta\theta_{ij}) . \end{aligned} \tag{S79}$$

This last expression can be rewritten as

$$\begin{aligned} P(\tilde{a}_{ij} = 1|\kappa_i^-, \kappa_i^+, \kappa_j^-, \kappa_j^+, \Delta\theta_{ij}) = {} & P_{ij}(a_{ij} = 1|\kappa_i^+, \kappa_j^-, \Delta\theta_{ij}) \\ & + P_{ij}(a_{ji} = 1|\kappa_i^-, \kappa_j^+, \Delta\theta_{ij}) \\ & - P_{ij}(a_{ij} = 1, a_{ji} = 1|\kappa_i^-, \kappa_i^+, \kappa_j^-, \kappa_j^+, \Delta\theta_{ij}) , \end{aligned} \tag{S80}$$

where the three probabilities on the right-hand side are obtained using Eqs. (S15) and (S18). Let us also introduce $P(\tilde{a}_{ij} = 1|\kappa_i^-, \kappa_i^+, \kappa_j^-, \kappa_j^+)$ which corresponds to $P(\tilde{a}_{ij} = 1|\kappa_i^-, \kappa_i^+, \kappa_j^-, \kappa_j^+, \Delta\theta_{ij})$ averaged over all possible angular positions

$$\begin{aligned} P(\tilde{a}_{ij} = 1|\kappa_i^-, \kappa_i^+, \kappa_j^-, \kappa_j^+) &= \int P(\tilde{a}_{ij} = 1|\kappa_i^-, \kappa_i^+, \kappa_j^-, \kappa_j^+) P(\Delta\theta_{ij}) d\Delta\theta_{ij} \\ &= \left\langle a_{ij}\middle|\kappa_i^+, \kappa_j^-\right\rangle + \left\langle a_{ji}\middle|\kappa_j^+, \kappa_i^-\right\rangle - \left\langle a_{ij}a_{ji}\middle|\kappa_i^-, \kappa_i^+, \kappa_j^-, \kappa_j^+\right\rangle , \end{aligned} \tag{S81}$$

where $\left\langle a_{ij}\middle|\kappa_i^+, \kappa_j^-\right\rangle$ and $\left\langle a_{ji}\middle|\kappa_j^+, \kappa_i^-\right\rangle$ are computed using Eq. (S23a), and where

$$\left\langle a_{ij}a_{ji}\middle|\kappa_i^-, \kappa_i^+, \kappa_j^-, \kappa_j^+\right\rangle = \begin{cases} (1+\nu)\left\langle a_{ij}a_{ji}\middle|\kappa_i^-, \kappa_i^+, \kappa_j^-, \kappa_j^+, \nu{=}0\right\rangle & \\ \qquad -\nu\left\langle a_{ij}a_{ji}\middle|\kappa_i^-, \kappa_i^+, \kappa_j^-, \kappa_j^+, \nu{=}{-}1\right\rangle & -1 \le \nu \le 0 \\ (1-\nu)\left\langle a_{ij}a_{ji}\middle|\kappa_i^-, \kappa_i^+, \kappa_j^-, \kappa_j^+, \nu{=}0\right\rangle & \\ \qquad +\nu\left\langle a_{ij}a_{ji}\middle|\kappa_i^-, \kappa_i^+, \kappa_j^-, \kappa_j^+, \nu{=}1\right\rangle & 0 \le \nu \le 1 \end{cases} \tag{S82}$$

is computed using Eqs. (S51), (S54), (S56) and (S59).

From these quantities, we use Bayes theorem to define two probability distributions with which we estimate the expected density of triangles. The first one corresponds to the probability that neighbor j of node i has hidden degrees κ_j^-, κ_j^+ regardless of the angular distance

$$P(\kappa_j^-, \kappa_j^+|\tilde{a}_{ij} = 1, \kappa_i^-, \kappa_i^+) = \frac{P(\tilde{a}_{ij} = 1|\kappa_i^-, \kappa_i^+, \kappa_j^-, \kappa_j^+) P(\kappa_j^-, \kappa_j^+)}{P(\tilde{a}_{i\bullet} = 1|\kappa_i^-, \kappa_i^+)} \tag{S83}$$

where $P(\tilde{a}_{i\bullet} = 1|\kappa_i^-, \kappa_i^+)$ is a normalization constant. The second distribution provides the probability that neighboring nodes i and j are at angular distance $\Delta\theta_{ij}$

$$P(\Delta\theta_{ij}|\tilde{a}_{ij} = 1, \kappa_i^-, \kappa_i^+, \kappa_j^-, \kappa_j^+) = \frac{P(\tilde{a}_{ij} = 1|\kappa_i^-, \kappa_i^+, \kappa_j^-, \kappa_j^+, \Delta\theta_{ij}) P(\Delta\theta_{ij})}{P(\tilde{a}_{ij} = 1|\kappa_i^-, \kappa_i^+, \kappa_j^-, \kappa_j^+)} . \tag{S84}$$

Recall that $P(\Delta\theta_{ij}) = 1/\pi$ in the directed-reciprocal $\mathbb{S}^1$ model.

With these quantities in hand, the expected density of triangles is estimated by computing

$$\bar{c}_{\text{undir}} \approx \frac{1}{MN_{>1}} \sum_{i=1}^{N} \sum_{m=1}^{M} c_i^{(m)} \mathbb{1}_{\{k_i^- + k_i^+ > 1\}} \ , \tag{S85}$$

where $\mathbb{1}_{\{\cdot\}}$ is the indicator function, M is the number of samples to be drawn for each node i, and where the m-th sample, $c_i^{(m)}$, is obtained with the following procedure.

1. *Pick the hidden degrees of two neighbors*, (κ_1^-, κ_1^+) and (κ_2^-, κ_2^+), by sampling Eq. (S83) twice.
2. *Pick the angular distance* between node i and nodes 1 and 2, $\Delta\theta_{i1}$ and $\Delta\theta_{i2}$, by sampling Eq. (S84) twice.
3. *Set the angular distance between nodes 1 and 2.* Since it is equally likely for nodes 1 and 2 to be "on the same side" or "on opposite sides" from node i, we set

$$\Delta\theta_{12} = \begin{cases} \min\left\{|\Delta\theta_{i1} + \Delta\theta_{i2}|, 2\pi - |\Delta\theta_{i1} + \Delta\theta_{i2}|\right\} & \text{with probability } 1/2 \\ \\ |\Delta\theta_{i1} - \Delta\theta_{i2}| & \text{with probability } 1/2 \end{cases} . \tag{S86}$$

4. *Compute the probability for nodes 1 and 2 to be connected* and set $c_i^{(m)} = P(\tilde{a}_{12} = 1 | \kappa_1^-, \kappa_1^+, \kappa_2^-, \kappa_2^+, \Delta\theta_{12})$.

Note: The average undirected local clustering coefficient is a convenient measure to estimate the value of β. However, because it does not fully embrace the direction of links, leading to an ambiguous definition of the degree used at the denominator (see Methods), the estimated value for β may require some manual adjustments for the model to accurately reproduce the number of triangles observed in the original network. See Fig. S2 for an illustration.

E. The algorithm

The algorithm assumes that a maximal deviation tolerance, η^{tol}, has been assigned, as well as defines $\bar{c}_{\text{undir}}^{\text{min}} = 0$, $\bar{c}_{\text{undir}}^{\text{max}} = 1$, $\beta^{\text{min}} = 1$ and $\beta^{\text{max}} = 25$.

1. *Set the initial value for the parameter* β as $\beta = 1 + u$ where $u \sim \text{Uniform}(0, 1)$.
2. *Infer the hidden in/out-degrees* $\boldsymbol{\kappa} = \kappa_1^-, \kappa_1^+, \ldots, \kappa_N^-, \kappa_N^+$ by following the procedure explained in Sec. S.IV B.
3. *Infer the parameter* ν by following the procedure explained in Sec. S.IV C.
4. *Estimate the triangle density* $\bar{c}_{\text{undir}}$ by following the procedure explained in Sec. S.IV D.
5. *Check for convergence* by checking if $|\bar{c}_{\text{undir}} - \bar{c}_{\text{undir}}^{\text{obs}}| < \eta^{\text{tol}}$, then all $2N + 2$ parameters have been estimated within the tolerance parameters. Otherwise, proceed to step 6.
6. *Update the value of the parameter* β (bisection method):
   (a) If $\bar{c}_{\text{undir}} > \bar{c}_{\text{undir}}^{\text{obs}}$, then set $\beta^{\text{max}} = \beta$, set $\bar{c}_{\text{undir}}^{\text{max}} = \bar{c}_{\text{undir}}$ and proceed to step 6c.
   (b) If $\bar{c}_{\text{undir}} < \bar{c}_{\text{undir}}^{\text{obs}}$, then set $\beta^{\text{min}} = \beta$, set $\bar{c}_{\text{undir}}^{\text{min}} = \bar{c}_{\text{undir}}$ and proceed to step 6c.
   (c) Update β to its new value according to

$$\beta = \beta^{\text{min}} + (\beta^{\text{max}} - \beta^{\text{min}}) \frac{\bar{c}_{\text{undir}}^{\text{obs}} - \bar{c}_{\text{undir}}^{\text{min}}}{\bar{c}_{\text{undir}}^{\text{max}} - \bar{c}_{\text{undir}}^{\text{min}}} \ .$$

   (d) Proceed to step 2.

S.V. MODELING REAL DIRECTED COMPLEX NETWORKS

A. Reproducing clustering and reciprocity

Figures S2 and S3 compare the accuracy with which the directed-reciprocal $\mathbb{S}^1$ model reproduces the average undirected local clustering coefficient, the number of triangles, as well as the reciprocity.

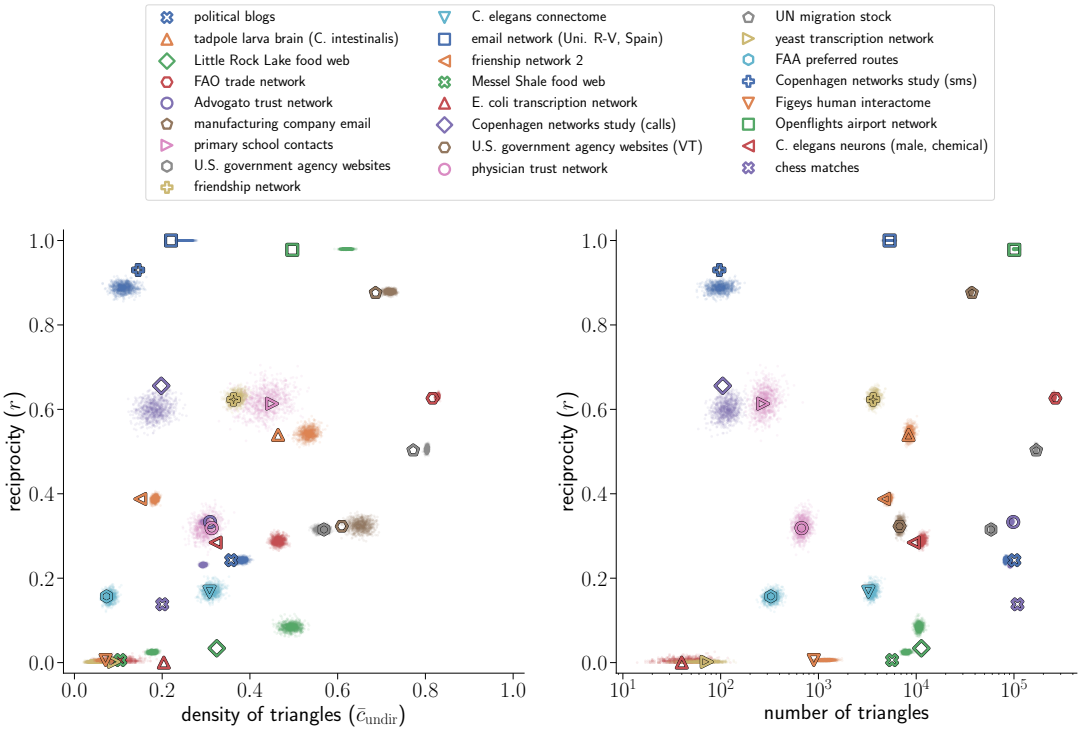

FIG. S2. Symbols represent the values measured on the original networks, and the small translucid circles show the same values measured on synthetic networks generated using the parameters inferred by the algorithm presented in Sec. S.IV (1000 network instances).

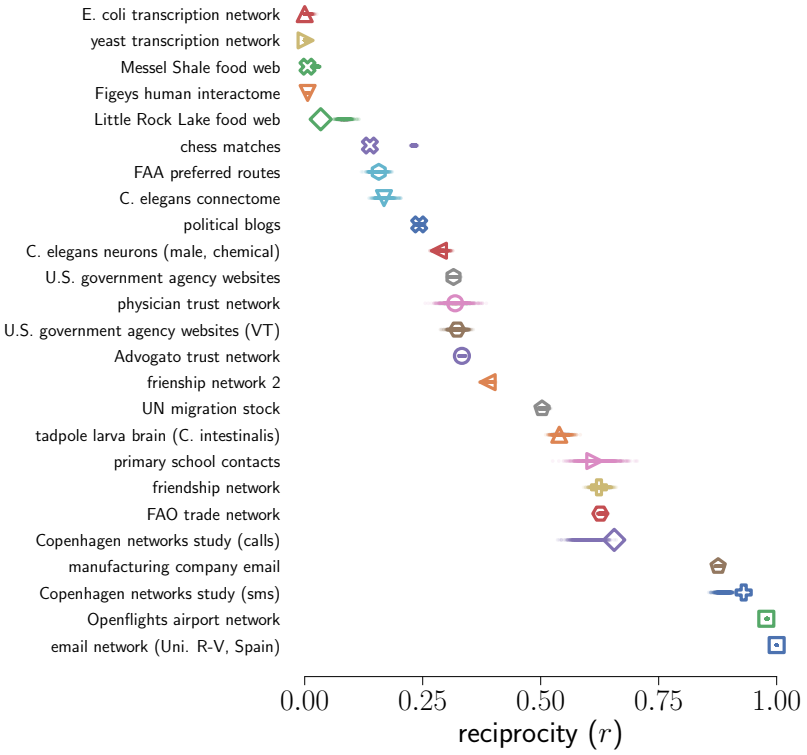

FIG. S3. Symbols represent the values measured on the original networks, and the small translucid circles show the same values measured on synthetic networks generated using the parameters inferred by the algorithm presented in Sec. S.IV (1000 network instances).

B. Additional triangle spectra

Figure S4 provides further examples of the capacity of the directed-reciprocal $\mathbb{S}^1$ model to reproduce the triangle spectra observed in various real directed networks. Table SI contains the parameters inferred for each network dataset.

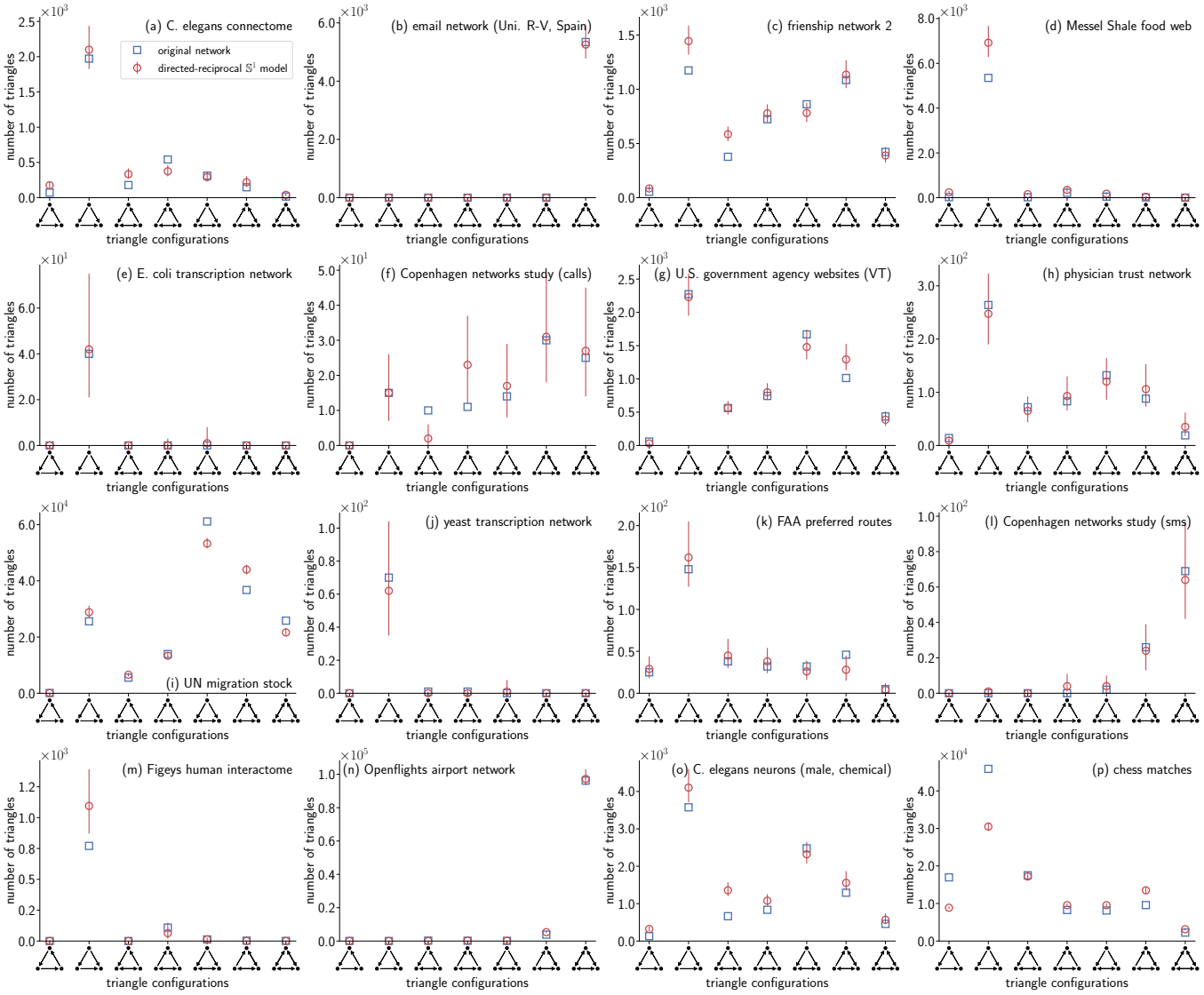

FIG. S4. **Reproducing triangle spectrum of real directed networks with the directed-reciprocal $\mathbb{S}^1$ model.** (a) Neural connections of the C. elegans nematode (dataset `celegansneural` [2, 3]). (b) Emails among members of a university (dataset `uni_email` [4]). (c) Friendships among high school students(dataset `add_health_comm50` [5]). (d) Messel Shale food web (dataset `messal_shale` [6]). (e) E. coli transcription network (dataset `ecoli_transcription_v1_0` [7]). (f) Social interactions among university students (dataset `copenhagen_calls` [8]). (g) Links between Vermont's government agencies websites (dataset: `us_agencies_vermont`) [9]). (h) Trust relationships among physicians (dataset `physician_trust` [10]). (i) Migration between countries (dataset `un_migrations` [11]). (j) Yeast transcription network (dataset `yeast_transcription` [12]). (k) Air traffic routes (dataset `faa_routes` [13]). (l) Social interactions among university students (dataset `copenhagen_sms` [8]). (m) Binding interactions between human proteins (dataset `interactome_figeys` [14]). (n) Regularly occurring flights among airports worldwide (dataset `openflights` [15]). (o) Networks among neurons of both the adult male and adult hermaphrodite worms C. elegans (dataset `celegans_2019_male_chemical` [16]). (p) Match outcomes between chess players (dataset `chess` [17]). Network datasets were downloaded from The Netzschleuder network catalogue and repository (`https://networks.skewed.de`). For each dataset, the parameters of the directed-reciprocal $\mathbb{S}^1$ model were adjusted using the inference procedure described in Sec. S.IV. Vertical lines show the estimated 95% confidence interval (2.5 and 97.5 percentiles).

TABLE SI. Parameters inferred for the network datasets used in Fig. 4 and Fig. S4.

Dataset	Figure	β	ν
political blogs	4(e)	1.75	0.43
tadpole larva brain (C. intestinalis)	4(f)	2.00	0.23
Little Rock Lake food web	4(g)	1.50	-1.00
FAO trade network	4(h)	1.75	0.34
Advogato trust network	4(i)	1.24	0.66
manufacturing company email	4(j)	1.50	1.00
primary school contacts	4(k)	2.43	0.88
U.S. government agency websites	4(l)	1.26	0.32
friendship network	4(m)	1.82	0.73
C. elegans connectome	S4(a)	1.28	-0.41
email network (Uni. R-V, Spain)	S4(b)	1.45	1.00
frienship network 2	S4(c)	1.40	0.48
Messel Shale food web	S4(d)	1.01	-0.91
E. coli transcription network	S4(e)	1.50	-0.02
Copenhagen networks study (calls)	S4(f)	1.50	1.00
U.S. government agency websites (VT)	S4(g)	2.00	0.11
physician trust network	S4(h)	1.67	0.27
UN migration stock	S4(i)	4.50	-0.73
yeast transcription network	S4(j)	1.20	0.02
FAA preferred routes	S4(k)	1.01	0.08
Copenhagen networks study (sms)	S4(l)	1.30	1.00
Figeys human interactome	S4(m)	1.01	1.00
Openflights airport network	S4(n)	3.25	1.00
C. elegans neurons (male, chemical)	S4(o)	2.00	-0.42
chess matches	S4(p)	1.30	-0.17

S.VI. USEFUL RESULTS INVOLVING THE HYPERGEOMETRIC FUNCTION

Letting $a, b \in \mathbb{C}$, $c \in \mathbb{C}\backslash\{0, -1, -2, -3, \ldots\}$ and $z \in \mathbb{Z}$, the hypergeometric function is defined by the Gauss series as [18]

$$ {}_2F_1(a, b; c; z) = \sum_{n=0}^{\infty} \frac{\Gamma(a+n)}{\Gamma(a)} \frac{\Gamma(b+n)}{\Gamma(b)} \frac{\Gamma(c)}{\Gamma(c+n)} \frac{z^n}{n!} \tag{S87} $$

for $|z| < 1$ and elsewhere by analytic continuation. In what follows, we will use the following identity [19]

$$ \begin{aligned} {}_2F_1(a, b; c; z) = \frac{\pi\Gamma(c)}{\sin\pi(b-a)} \Bigg[& \frac{(-z)^{-a}}{\Gamma(c-a)\Gamma(b)\Gamma(a-b+1)} \, {}_2F_1\left(a, a-c+1, a-b+1; \frac{1}{z}\right) \\ & - \frac{(-z)^{-b}}{\Gamma(c-b)\Gamma(a)\Gamma(b-a+1)} \, {}_2F_1\left(b, b-c+1, b-a+1; \frac{1}{z}\right) \Bigg] \end{aligned} \tag{S88} $$

valid for $\arg(1-z) < \pi$, as well as [20]

$$ z \, {}_2F_1(a, b+1; c+1; z) = \frac{c}{b} \, {}_2F_1(a, b; c; z) - \frac{c}{b} \, {}_2F_1(a-1, b; c; z) \, . \tag{S89} $$

We will also need Euler's reflection formula [21]

$$ \Gamma(z)\Gamma(1-z) = \frac{\pi}{\sin(\pi z)} \tag{S90} $$

valid for $z \neq 0, \pm 1, \pm 2, \ldots$

We seek to evaluate the integral $\int \frac{dx}{1+x^\beta}$ for $x > 0$ and $\beta > 1$. To do so, we split the open interval $x > 0$ into two parts. First, we find for $0 < x < 1$

$$ \begin{aligned} \int \frac{dx}{1+x^\beta} &= \int \frac{1}{1-(-x^\beta)} dx \\ &= \int \sum_{n=0}^{\infty} \left(-x^\beta\right)^n dx \\ &= x \sum_{n=0}^{\infty} \frac{\left(-x^\beta\right)^n}{\beta n + 1} + C \\ &= x \sum_{n=0}^{\infty} \frac{\Gamma(1+n)}{n!\Gamma(1)} \frac{\Gamma(\frac{1}{\beta})}{\Gamma(\frac{1}{\beta})} \frac{\Gamma(\frac{1}{\beta}+n)}{\Gamma(\frac{1}{\beta}+n)} \frac{\frac{1}{\beta}}{\frac{1}{\beta}+n} \left(-x^\beta\right)^n + C_1 \\ &= x \sum_{n=0}^{\infty} \frac{\Gamma(1+n)}{\Gamma(1)} \frac{\Gamma(\frac{1}{\beta}+n)}{\Gamma(\frac{1}{\beta})} \frac{\Gamma(1+\frac{1}{\beta})}{\Gamma(1+\frac{1}{\beta}+n)} \frac{\left(-x^\beta\right)^n}{n!} + C_1 \\ &= x \, {}_2F_1\left(1, \frac{1}{\beta}; 1+\frac{1}{\beta}; -x^\beta\right) + C_1 \end{aligned} \tag{S91} $$

where $C_1 \in \mathbb{R}$. Second, we find for $x > 1$

$$\begin{aligned}
\int \frac{dx}{1+x^\beta} &= \int \frac{1}{x^\beta} \frac{1}{1-(-x^{-\beta})} dx \\
&= -\int \sum_{n=0}^{\infty} \left(-x^{-\beta}\right)^{n+1} dx \\
&= -x \sum_{n=0}^{\infty} \frac{\left(-x^{-\beta}\right)^{n+1}}{-\beta(n+1)+1} + C_2 \\
&= -x \sum_{m=1}^{\infty} \frac{\left(-x^{-\beta}\right)^{m}}{-\beta m+1} + C_2 \\
&= -x \sum_{m=1}^{\infty} \frac{\Gamma(1+m)}{m!\Gamma(1)} \frac{\Gamma(-\frac{1}{\beta})}{\Gamma(-\frac{1}{\beta})} \frac{\Gamma(-\frac{1}{\beta}+m)}{\Gamma(-\frac{1}{\beta}+m)} \frac{-\frac{1}{\beta}}{-\frac{1}{\beta}+m} \left(-x^{-\beta}\right)^{m} + C_2 \\
&= -x \sum_{m=1}^{\infty} \frac{\Gamma(1+m)}{\Gamma(1)} \frac{\Gamma(-\frac{1}{\beta}+m)}{\Gamma(-\frac{1}{\beta})} \frac{\Gamma(1-\frac{1}{\beta})}{\Gamma(1-\frac{1}{\beta}+m)} \frac{\left(-x^{-\beta}\right)^{m}}{m!} + C_2 \\
&= -x \left[{}_2F_1\left(1, -\frac{1}{\beta}; 1-\frac{1}{\beta}; -x^{-\beta}\right) - 1 \right] + C_2
\end{aligned} \tag{S92}$$

where $C_2 \in \mathbb{R}$. Combining Eqs. (S88) and (S90), we find

$${}_2F_1\left(1, -\frac{1}{\beta}; 1-\frac{1}{\beta}; -x^{-\beta}\right) = \frac{\frac{1}{\beta}}{1+\frac{1}{\beta}} x^\beta \, {}_2F_1\left(1, 1+\frac{1}{\beta}, 2+\frac{1}{\beta}; -x^\beta\right) + \frac{1}{\beta} \frac{\Gamma(-\frac{1}{\beta}-1)\Gamma(2+\frac{1}{\beta})}{x} \, . \tag{S93}$$

Using Eq. (S89), we find

$$x^\beta \, {}_2F_1\left(1, 1+\frac{1}{\beta}, 2+\frac{1}{\beta}; -x^\beta\right) = -\frac{1+\frac{1}{\beta}}{\frac{1}{\beta}} \, {}_2F_1\left(1, \frac{1}{\beta}; 1+\frac{1}{\beta}; -x^\beta\right) + \frac{1+\frac{1}{\beta}}{\frac{1}{\beta}} \tag{S94}$$

Combining Eqs. (S91)–(S94), we finally get

$$\int \frac{dx}{1+x^\beta} = x \, {}_2F_1\left(1, \frac{1}{\beta}; 1+\frac{1}{\beta}; -x^\beta\right) + C_3 \tag{S95}$$

for $x > 0$ and $\beta > 1$, and where $C_3 \in \mathbb{R}$.

We also seek to evaluate the integral $\int \frac{dx}{(1+x^\beta)^2}$ for $x > 0$ and $\beta > 1$. Again, we split the open interval $x > 0$ into two parts. First, we find for $0 < x < 1$

$$\begin{aligned}
\int \frac{dx}{(1+x^\beta)^2} &= \int \frac{d}{d(-x^\beta)} \frac{1}{1-(-x^\beta)} dx \\
&= \int \frac{d}{d(-x^\beta)} \sum_{n=0}^{\infty} \left(-x^\beta\right)^{n} dx \\
&= \int \sum_{n=1}^{\infty} n \left(-x^\beta\right)^{n-1} dx \\
&= x \sum_{m=0}^{\infty} \frac{(m+1)\left(-x^\beta\right)^{m}}{\beta m+1} + C_4 \\
&= x \sum_{m=0}^{\infty} (m+1) \frac{\Gamma(1+m)}{m!\Gamma(1)} \frac{\Gamma(\frac{1}{\beta})}{\Gamma(\frac{1}{\beta})} \frac{\Gamma(\frac{1}{\beta}+m)}{\Gamma(\frac{1}{\beta}+m)} \frac{\frac{1}{\beta}}{\frac{1}{\beta}+m} \left(-x^\beta\right)^{m} + C_4 \\
&= x \sum_{m=0}^{\infty} \frac{\Gamma(2+m)}{\Gamma(2)} \frac{\Gamma(\frac{1}{\beta}+m)}{\Gamma(\frac{1}{\beta})} \frac{\Gamma(1+\frac{1}{\beta})}{\Gamma(1+\frac{1}{\beta}+m)} \frac{\left(-x^\beta\right)^{m}}{m!} + C_4 \\
&= x \, {}_2F_1\left(2, \frac{1}{\beta}; 1+\frac{1}{\beta}; -x^\beta\right) + C_4
\end{aligned} \tag{S96}$$

where $C_4 \in \mathbb{R}$. Second, we find for $x > 1$

$$\begin{aligned}
\int \frac{dx}{(1+x^\beta)^2} &= \int \frac{1}{x^{2\beta}} \frac{dx}{(1+x^{-\beta})^2} \\
&= \int x^{-2\beta} \frac{d}{d(-x^\beta)} \frac{1}{1-(-x^{-\beta})} dx \\
&= \int x^{-2\beta} \frac{d}{d(-x^{-\beta})} \sum_{n=0}^{\infty} \left(-x^{-\beta}\right)^n dx \\
&= \int \sum_{n=1}^{\infty} n \left(-x^{-\beta}\right)^{n+1} dx \\
&= x(-x^{-\beta}) \sum_{n=1}^{\infty} (-x^{-\beta}) \frac{d}{d(-x^{-\beta})} \frac{\left(-x^{-\beta}\right)^n}{-\beta(n+1)+1} + C_5 \\
&= x(-x^{-\beta})^2 \frac{d}{d(-x^{-\beta})} \sum_{n=1}^{\infty} \frac{\left(-x^{-\beta}\right)^n}{-\beta(n+1)+1} + C_5 \\
&= x(-x^{-\beta})^2 \frac{d}{d(-x^{-\beta})} \sum_{n=1}^{\infty} \frac{\Gamma(1+n)}{\Gamma(1)n!} \frac{-\frac{1}{\beta}}{1-\frac{1}{\beta}+n} \frac{1-\frac{1}{\beta}}{1-\frac{1}{\beta}} \frac{\Gamma(1-\frac{1}{\beta})}{\Gamma(1-\frac{1}{\beta})} \frac{\Gamma(1-\frac{1}{\beta}+n)}{\Gamma(1-\frac{1}{\beta}+n)} \left(-x^{-\beta}\right)^n + C_5 \\
&= x(-x^{-\beta})^2 \frac{-\frac{1}{\beta}}{1-\frac{1}{\beta}} \frac{d}{d(-x^{-\beta})} \sum_{n=1}^{\infty} \frac{\Gamma(1+n)}{\Gamma(1)} \frac{\Gamma(1-\frac{1}{\beta}+n)}{\Gamma(1-\frac{1}{\beta})} \frac{\Gamma(2-\frac{1}{\beta})}{\Gamma(2-\frac{1}{\beta}+n)} \frac{\left(-x^{-\beta}\right)^n}{n!} + C_5 \\
&= x(-x^{-\beta})^2 \frac{-\frac{1}{\beta}}{1-\frac{1}{\beta}} \frac{d}{d(-x^{-\beta})} \left[{}_2F_1\left(1, 1-\frac{1}{\beta}; 2-\frac{1}{\beta}; -x^{-\beta}\right) - 1 \right] + C_5 \\
&= \frac{-\frac{1}{\beta}}{2-\frac{1}{\beta}} x(-x^{-\beta})^2 \, {}_2F_1\left(2, 2-\frac{1}{\beta}; 3-\frac{1}{\beta}; -x^{-\beta}\right) + C_5
\end{aligned} \tag{S97}$$

where $C_5 \in \mathbb{R}$ and where we used the following identity [22] to obtain the last equality

$$\frac{d}{dz} \, {}_2F_1(a, b; c; z) = \frac{ab}{c} \, {}_2F_1(a+1, b+1; c+1; z) \, . \tag{S98}$$

Using Eqs. (S88) and (S90), Eq. (S97) becomes

$$\int \frac{dx}{(1+x^\beta)^2} = x \, {}_2F_1\left(2, \frac{1}{\beta}; 1+\frac{1}{\beta}; -x^\beta\right) + (1-\tfrac{1}{\beta})\Gamma(1-\tfrac{1}{\beta})\Gamma(1+\tfrac{1}{\beta}) + C_5 \, , \tag{S99}$$

which, combined with Eq. (S96), yields

$$\int \frac{dx}{(1+x^\beta)^2} = x \, {}_2F_1\left(2, \frac{1}{\beta}; 1+\frac{1}{\beta}; -x^\beta\right) + C_6 \, , \tag{S100}$$

for $x > 0$ and $\beta > 1$, and where $C_6 \in \mathbb{R}$.

We additionally seek to evaluate the following integral, which can be solved using Eqs. (S95) and (S100)

$$\begin{aligned}
\int \frac{1}{1+x^\beta} \frac{1}{1+(\kappa x)^\beta} dx &= \frac{1}{1-\kappa^\beta} \int \frac{dx}{1+x^\beta} - \frac{\kappa^\beta}{1-\kappa^\beta} \int \frac{dx}{1+(\kappa x)^\beta} \\
&= \begin{cases}
x \, {}_2F_1\left(2, \frac{1}{\beta}; 1+\frac{1}{\beta}; -x^\beta\right) + C_7 & \text{for } \kappa = 1 \\
\frac{x}{1-\kappa^\beta} \, {}_2F_1\left(1, \frac{1}{\beta}; 1+\frac{1}{\beta}; -x^\beta\right) & \\
\qquad - \frac{x\kappa^\beta}{1-\kappa^\beta} \, {}_2F_1\left(1, \frac{1}{\beta}; 1+\frac{1}{\beta}; -(\kappa x)^\beta\right) + C_8 & \text{for } \kappa \neq 1
\end{cases}
\end{aligned} \tag{S101}$$

with $\kappa > 0$ and $\beta > 1$, $C_7, C_8 \in \mathbb{R}$ and $x > 0$.

Letting $d \in \{1, 2\}$, we use Eq. (S88) to write

$$z\ {}_2F_1\left(d, \frac{1}{\beta}; 1 + \frac{1}{\beta}; -z^\beta\right) = \frac{(-1)^d \pi}{\sin(\frac{\pi}{\beta})} \frac{\Gamma(1 + \frac{1}{\beta})}{\Gamma(1 + \frac{1}{\beta} - d)} \left[\sum_{n=0}^{\infty} \frac{(-1)^n \Gamma(d+n)\Gamma(d - \frac{1}{\beta} + n)}{\Gamma(d - \frac{1}{\beta})\Gamma(\frac{1}{\beta}) n! \Gamma(d - \frac{1}{\beta} + 1 + s)} z^{1-(n+d)\beta} - 1\right], \quad \text{(S102)}$$

which yields

$$\lim_{z\to\infty} z\ {}_2F_1\left(d, \frac{1}{\beta}; 1 + \frac{1}{\beta}; -z^\beta\right) = \frac{(-1)^{d+1}\pi}{\sin(\frac{\pi}{\beta})} \frac{\Gamma(1 + \frac{1}{\beta})}{\Gamma(1 + \frac{1}{\beta} - d)}\ , \quad \text{(S103)}$$

and more specifically

$$\lim_{z\to\infty} z\ {}_2F_1\left(1, \frac{1}{\beta}; 1 + \frac{1}{\beta}; -z^\beta\right) = \frac{\pi}{\beta} \frac{1}{\sin(\frac{\pi}{\beta})} \quad \text{(S104)}$$

and

$$\lim_{z\to\infty} z\ {}_2F_1\left(2, \frac{1}{\beta}; 1 + \frac{1}{\beta}; -z^\beta\right) = \frac{\pi(\beta - 1)}{\beta^2} \frac{1}{\sin(\frac{\pi}{\beta})}\ . \quad \text{(S105)}$$

REFERENCES

[1] G. García-Pérez, A. Allard, M. Á. Serrano, and M. Boguñá, Mercator: uncovering faithful hyperbolic embeddings of complex networks, New J. Phys. **21**, 123033 (2019).
[2] D. J. Watts and S. H. Strogatz, Collective dynamics of small-world networks, Nature **393**, 440 (1998).
[3] J. G. White, E. Southgate, J. N. Thomson, and S. Brenner, The structure of the nervous system of the nematode Caenorhabditis elegans, Philos. Trans. Royal Soc. B **314**, 1 (1986).
[4] R. Guimerà, L. Danon, A. Díaz-Guilera, F. Giralt, and A. Arenas, Self-similar community structure in a network of human interactions, Phys. Rev. E **68**, 065103 (2003).
[5] J. Moody, Peer influence groups: identifying dense clusters in large networks, Soc. Networks **23**, 261 (2001).
[6] J. A. Dunne, C. C. Labandeira, and R. J. Williams, Highly resolved early Eocene food webs show development of modern trophic structure after the end-Cretaceous extinction, Proc. R. Soc. B **281**, 20133280 (2014).
[7] S. S. Shen-Orr, R. Milo, S. Mangan, and U. Alon, Network motifs in the transcriptional regulation network of Escherichia coli, Nat. Genet. **31**, 64 (2002).
[8] P. Sapiezynski, A. Stopczynski, D. D. Lassen, and S. Lehmann, Interaction data from the Copenhagen Networks Study, Sci. Data **6**, 315 (2019).
[9] S. Kosack, M. Coscia, E. Smith, K. Albrecht, A.-L. Barabási, and R. Hausmann, Functional structures of US state governments, Proc. Natl. Acad. Sci. U.S.A. **115**, 11748 (2018).
[10] J. Coleman, E. Katz, and H. Menzel, The Diffusion of an Innovation Among Physicians, Sociometry **20**, 253 (1957).
[11] United Nations, Department of Economic and Social Affairs, Population Division, *Trends in International Migrant Stock: The 2015 Revision*, United Nations database POP/DB/MIG/Stock/Rev.2015 (2015).
[12] R. Milo, S. Shen-Orr, S. Itzkovitz, N. Kashtan, D. B. Chklovskii, and U. Alon, Network Motifs: Simple Building Blocks of Complex Networks, Science **298**, 824 (2002).
[13] J. Kunegis, KONECT: the Koblenz network collection, in *Proceedings of the 22nd International Conference on World Wide Web* (2013) pp. 1343–1350.
[14] R. M. Ewing, P. Chu, F. Elisma, H. Li, P. Taylor, S. Climie, L. McBroom-Cerajewski, M. D. Robinson, L. O'Connor, M. Li, R. Taylor, M. Dharsee, Y. Ho, A. Heilbut, L. Moore, S. Zhang, O. Ornatsky, Y. V. Bukhman, M. Ethier, Y. Sheng, J. Vasilescu, M. Abu-Farha, J.-P. Lambert, H. S. Duewel, I. I. Stewart, B. Kuehl, K. Hogue, K. Colwill, K. Gladwish, B. Muskat, R. Kinach, S.-L. Adams, M. F. Moran, G. B. Morin, T. Topaloglou, and D. Figeys, Large-scale mapping of human protein–protein interactions by mass spectrometry, Mol. Syst. Biol. **3**, 89 (2007).
[15] The openflights.org website, https://openflights.org/data.html.
[16] S. J. Cook, T. A. Jarrell, C. A. Brittin, Y. Wang, A. E. Bloniarz, M. A. Yakovlev, K. C. Q. Nguyen, L. T.-H. Tang, E. A. Bayer, J. S. Duerr, H. E. Bülow, O. Hobert, D. H. Hall, and S. W. Emmons, Whole-animal connectomes of both Caenorhabditis elegans sexes, Nature **571**, 63 (2019).
[17] Kaggle, Chess ratings - Elo versus the Rest of the World, https://www.kaggle.com/c/chess/data.
[18] NIST Digital Library of Mathematical Functions: https://dlmf.nist.gov/15.2.E1.
[19] NIST Digital Library of Mathematical Functions: https://dlmf.nist.gov/15.8.E2.
[20] NIST Digital Library of Mathematical Functions: https://dlmf.nist.gov/15.5.E16_5.
[21] NIST Digital Library of Mathematical Functions: https://dlmf.nist.gov/5.5.E3.
[22] NIST Digital Library of Mathematical Functions: https://dlmf.nist.gov/15.5.E1.