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: 02 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

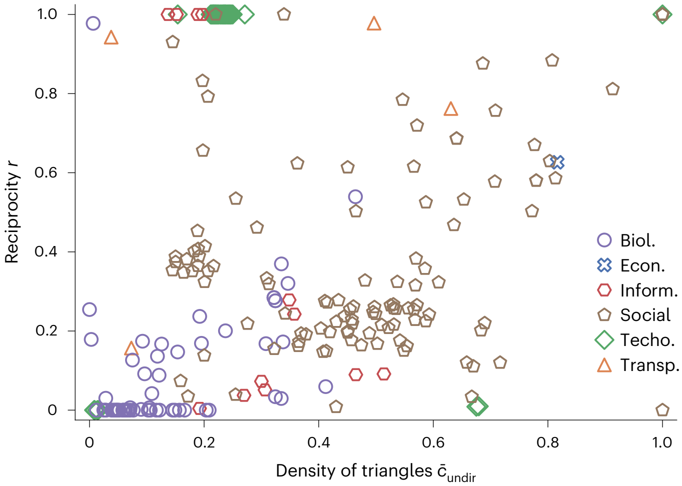

Fig. 1 | Reciprocity in real directed networks. Reciprocity versus density of triangles in 292 real directed biological (Biol.), economical (Econ.), information (Inform.), social, technological (Techno.) and transportation (Transp.) networks. The reciprocity is defined as $r = L^{\leftrightarrow}/L$,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

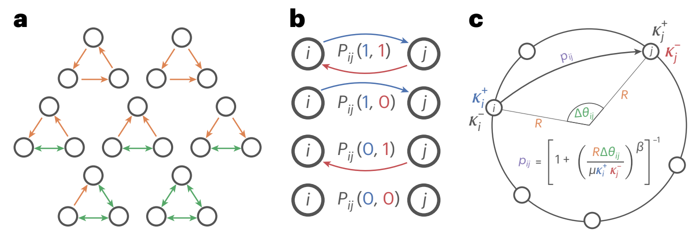

Fig. 2 | Illustrations of the concepts behind the modelling framework. **a**, The seven configurations of triangles in directed networks[45,47]. **b**, The joint probabilities $P_{ij}(a_{ij}, a_{ji})$ used in the general framework controlling reciprocity in random directed networks. **c**, The geometric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

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

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

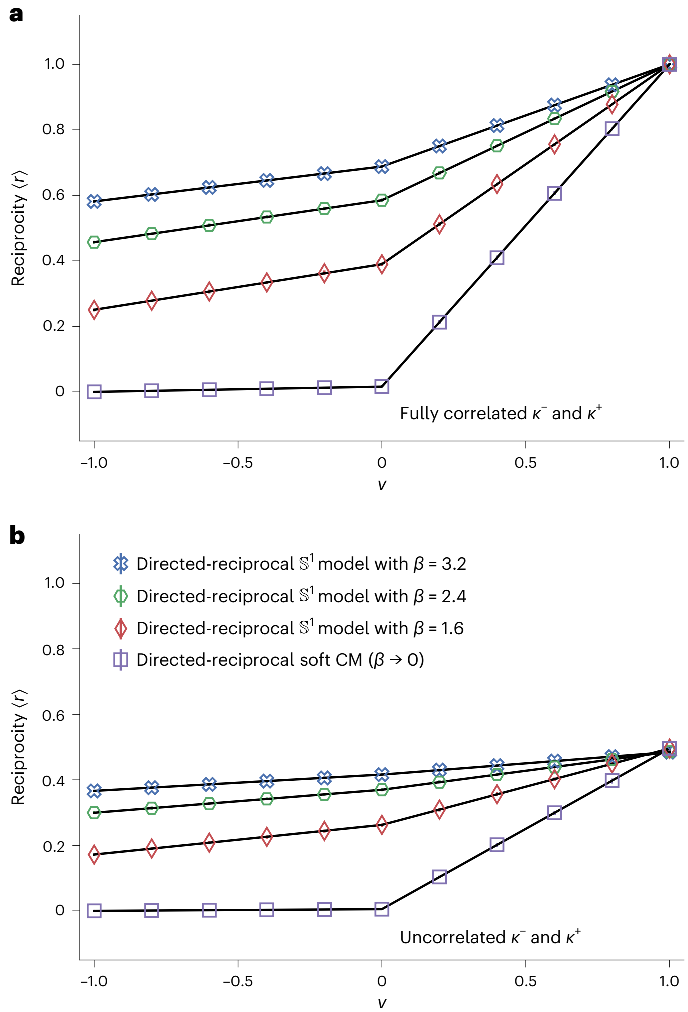

Fig. 3 | Validation of the general framework controlling reciprocity. **a**, Reciprocity versus its control parameter ν, setting $k_i^- = k_i^+$ for $i = 1, \ldots, N$ to fully correlate κ^- and κ^+. **b**, Reciprocity versus its control parameter ν, shuffling the sequence $\{k_i^+\}_{i=1,\ldots,N}$ 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

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

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

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

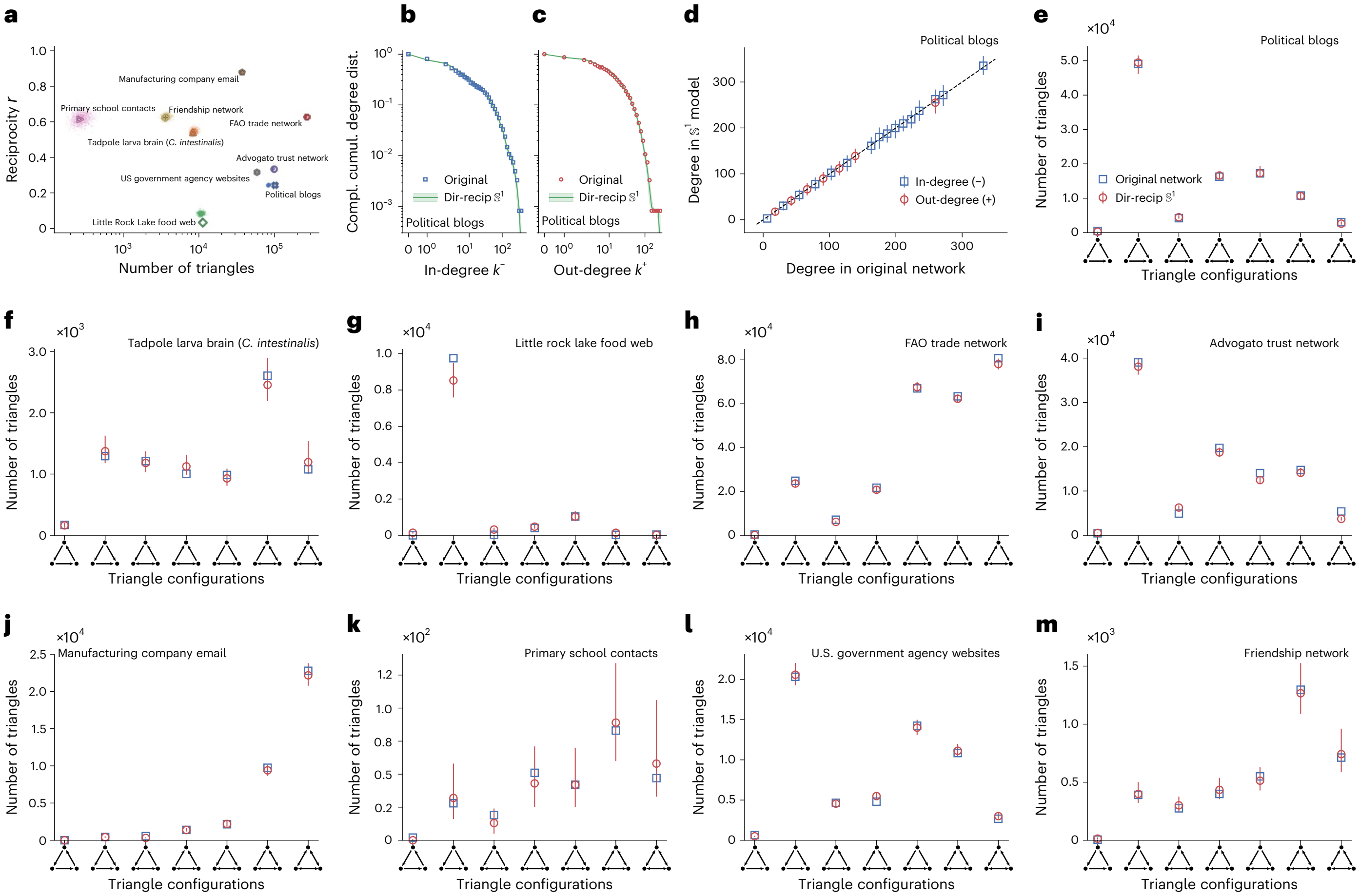

Fig. 4 | Reproducing topological features of real directed networks with the directed-reciprocal $\mathbb{S}^1$ model (Dir-recip). a, Reciprocity and the number of triangles measured on the real networks (symbols) compared to those measured on synthetic networks (1000 network instances; small translucid 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, $\bar{c}_{\mathrm{undir}}$, 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

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

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

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

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

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

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

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

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.