scientific reports

OPEN

The Fitness-Corrected Block Model, or how to create maximum-entropy data-driven spatial social networks

Massimo Bernaschi[1], Alessandro Celestini[1], Stefano Guarino [1]✉, Enrico Mastrostefano[1] & Fabio Saracco[1,2]

Models of networks play a major role in explaining and reproducing empirically observed patterns. Suitable models can be used to randomize an observed network while preserving some of its features, or to generate synthetic graphs whose properties may be tuned upon the characteristics of a given population. In the present paper, we introduce the Fitness-Corrected Block Model, an adjustable-density variation of the well-known Degree-Corrected Block Model, and we show that the proposed construction yields a maximum entropy model. When the network is sparse, we derive an analytical expression for the degree distribution of the model that depends on just the constraints and the chosen fitness-distribution. Our model is perfectly suited to define maximum-entropy data-driven spatial social networks, where each block identifies vertices having similar position (e.g., residence) and age, and where the expected block-to-block adjacency matrix can be inferred from the available data. In this case, the sparse-regime approximation coincides with a phenomenological model where the probability of a link binding two individuals is directly proportional to their sociability and to the typical cohesion of their age-groups, whereas it decays as an inverse-power of their geographic distance. We support our analytical findings through simulations of a stylized urban area.

The definition of a suitable data-driven spatial social network model is a widely studied problem in computational social sciences. Various dynamical processes (e.g., diseases spread) can be represented on such networks, and the topology of the network has a direct impact on the evolution of the process. Having general models for social interactions, based on available data and capable of reconstructing stylized facts known from the literature, is of utmost importance to prevent from reaching conclusions biased by incorrect or ill-defined assumptions. With widely available survey and census data, it is now possible to generate synthetic geo-localized populations, stratified by age and organized into households. However, there is no equally direct way to accurately model interpersonal relationships.

Many real social networks exhibit some form of group mixing, driven by the tendency of individuals to socialize with their peers[1]. This property can be reproduced using the so-called Stochastic Block Model (SBM), risen to prominence as a way to generate networks with a known community structure[2,3]. In the SBM, the vertex set is partitioned into disjoint *blocks* and the probability of an edge between two nodes depends on the blocks to which the two nodes belong. In the original formulation of the SBM, all vertices belonging to the same block are indistinguishable, so that the degree distribution within each block tends to be Poisson-like for large graphs[2]. To produce a more realistic network, a few extensions to the model have been proposed in the literature, including the Degree-Corrected Block-Model (DCBM)[2] and its maximum-entropy version[4]. These models are, in some sense, the SBM-equivalent of the well-known configuration model. They are based on enforcing both the desired group mixing and a target degree-sequence, either exactly or in expectation. If L_{IJ} is the number of links between blocks I and J – or twice that number, if $I = J$ – and $\deg_i$ is the degree of node v_i, the maximum-entropy DCBM works by imposing that $\langle L_{IJ} \rangle = K_{IJ}$ and $\langle \deg_i \rangle = k_i$ for suitable constants K_{IJ} and k_i. The internal consistency of the model requires $\sum_J K_{IJ} = \sum_{i \in I} k_i$ for all I, and the density of the network is fixed equal to $\frac{\sum_{IJ} K_{IJ}}{N(N-1)}$, where N is the size of the network.

In this paper, we define and analyze the Fitness-Corrected Block Model (FCBM), a variation of the DCBM where the network density p is a configuration parameter. In the FCBM the block-level mixing is specified in terms of a matrix of edge-densities Δ—as in the original SBM—whereas a sequence $\vec{f}$ of vertex intrinsic fitness

---

[1]Institute for Applied Computing "Mauro Picone", National Research Council of Italy, Via dei Taurini 19, 00185 Rome, Italy. [2]"Enrico Fermi" Research Center (CREF), Via Panisperna 89A, 00184 Rome, Italy. ✉email: s.guarino@iac.cnr.it

values[5,6] measures the propensity of each vertex to establish links and can be used to enforce the desired intra-block heterogeneity. In the DCBM the constants K_{IJ} and k_i, that bound, respectively, $\langle L_{IJ} \rangle$ and $\langle \deg_i \rangle$, need to be known explicitly. The DCBM was in fact conceived as an instrument to randomize an observed graph while preserving some of its features. In the FCBM, instead, K_{IJ} and k_i are determined based on p, Δ and $\bar{f}$, making the FCBM a suitable model for generating random graphs whose properties may be tuned upon the characteristics of a given population or set of entities.

To the purpose of having a model that is maximally random, the FCBM is defined following the approach first presented in Ref.[4] for the DCBM, and later clarified and generalized in Ref.[7]. In a nutshell, the approach consists in: (1) the definition of an ensemble of networks, each with the same number of nodes, but with all possible configurations of links; (2) the constrained maximization of the entropy associated to the network ensemble, via the method of Lagrangian multipliers. By imposing the conditions $\langle L_{IJ} \rangle = K_{IJ}$ and $\langle \deg_i \rangle = k_i$, for all I, J, i, the probability per graph in the ensemble factorises in terms of probabilities per link. The numerical value of the Lagrangian multipliers has to be calculated solving the system of nonlinear equations given by the model constraints. The resolution of this system might become expensive when the network is large and, to the best of our knowledge, it is not implemented in any publicly available software library. For the present paper, we implemented a parallel version of the solver for the maximum-entropy DCBM, written in C making use of the Intel MKL scientific library and the OpenMP API, which can also be used to solve our FCBM. The solver is released as open-source software at https://gitlab.com/cranic-group/dcbm_solver.

A known sparse-regime approximation for the edge probability of the DCBM implies that, in the sparse FCBM, the probability of a link binding two individuals i and j is directly proportional to their sociability and to the cohesion of their blocks, i.e., $p_{ij} \propto p f_i f_j \Delta_{I_i J_j}$ for all $i \in I$ and $j \in J$. Under the sparse-regime approximation, the maximum entropy condition leads to a system of equations that admits a closed-form solution. We make use of this approximate solution to find two closed-form estimates for the degree distribution p_k of the FCBM—one more accurate, the other neater. These two estimates put in direct relation p_k with the fitness distribution p_f, showing that the degree distribution of the FCBM, albeit not known a priori, can be essentially controlled through the model's parameters. In particular, if p_f follows a power-law, lognormal or exponential distribution, then the same holds, approximately, for p_k.

Among the many possible applications, the FCBM is especially well suited for generating a data-driven social network of geo-referenced and age-stratified individuals. The partition of the population into blocks can be obtained by grouping the individuals having similar position (e.g., residence) and age, whereas the expected block-to-block edge-density matrix Δ can be calibrated based on survey data that quantify the dependence of contact frequencies upon geographic and socio-demographic factors. Finally, the fitness vector $\bar{f}$ may be drawn from a suitable probability distribution, modelled upon measurable features such as wealth, employment, or mobility. In this context, the sparse FCBM is well approximated by the phenomenological model presented in Refs.[8,9].

To show how the FCBM can be used in practice, and to provide empirical support to our analytical findings, we generate a set of synthetic social networks for a stylized urban population distributed on a disk of radius 2.5 Km. We use the SOCRATES[10,11] tool to extract age-based social mixing patterns, and we embrace the widely-accepted assumption that an inverse-power-law relation binds the distance between two individuals and the frequency of their social interactions[12,13]. We generate instances of our FCBM using either the exact model or its sparse-regime approximation, varying both the spatial density of the individuals in the territory and the fitness distribution. We show that the empirical degree distribution obtained with all considered configurations is in sharp agreement with the analytical estimates. The implementation of the FCBM is publicly available as part of the Urban Social Networks (USN) framework at https://gitlab.com/cranic-group/usn.

Contributions and results

In the following, we summarize the main contributions and present the main analytical and experimental results of this paper. For all methodological details, we refer the reader to the "Methods" section.

The Fitness Corrected Block Model.   We propose the Fitness-Corrected Block Model (FCBM), a new maximum-entropy model for modular networks, parameterized by the network density $p \in (0, 1)$, the block-wise mixing structure Δ—a symmetric matrix such that $\sum_{I,J} \Delta_{IJ} = 2$—and the vertex-intrinsic fitness $\bar{f}$—a vector that controls the tendency of each vertex to establish links with other vertices.

Formally, let V be a vertex set of size N partitioned into n blocks $\{B_I\}_{I=0}^{n-1}$. The FCBM is defined as the maximum entropy probability distribution P, over all networks having vertex set V, fulfilling the following two conditions:

$$\langle L_{IJ} \rangle_P = p \binom{N}{2} \Delta_{IJ} \quad \text{for all } I, J \tag{1}$$

$$\langle \deg_i \rangle_P = p \binom{N}{2} \frac{f_i}{\sum_{u \in I_i} f_u} \sum_J \Delta_{I_i J} \quad \text{for all } i \tag{2}$$

where L_{IJ} is the number of edges between B_I and B_J; $\deg_i$, f_i and I_i are the degree, fitness and pertaining block of vertex v_i; $\langle \cdot \rangle_P$ denotes the expected value with respect to P.

The FCBM can be seen as a generalization of the Degree-Corrected Block Model (DCBM). However, contrarily to the DCBM, we show that the FCBM is consistent for any choice of the configuration parameters, making it suitable for generating random graphs with tunable topological properties.

Leveraging on the framework of entropy-based null-models, we prove that the probability $P(G)$ of generating a specific graph G with the FCBM can be factorized as the product of independent edge probabilities, namely $P(G) = \prod_{i,j} p_{ij}$, where p_{ij} is the probability of an edge between v_i and v_j. The edge probabilities can be recovered solving a system of non-linear equations obtained from (1) and (2).

Efficient solver for FCBM/DCBM system of equations. The system of nonlinear equations, needed to explicitly calculate edge probabilities, becomes computationally expensive as the size of the network increases. Already with a few thousand nodes, a straightforward implementation may be too slow to be used in practical applications. We implemented a parallel C-program that efficiently solves this system, as well as the analogous system arising from the DCBM. The solver follows the Sequential Quadratic Programming approach presented in Ref.[14], using Newton's method for the Hessian approximation. To the best of our knowledge, this is the first publicly available solver for the DCBM. The source code is publicly released as open-source software at https://gitlab.com/cranic-group/dcbm_solver.

Properties of the FCBM. The fitness sequence $\vec{f}$ guarantees intra-block heterogeneity. In fact, (2) can be rewritten as

$$\left\langle \deg_i \right\rangle_P = \frac{f_i}{\sum_{u \in I_i} f_u} \left\langle \deg_I \right\rangle_P$$

where $\left\langle \deg_I \right\rangle_P$ is the expected total degree of block B_I, i.e., the total number of edges incident to B_I.

When the network is sparse, we show that the system from which all p_{ij}'s must be derived admits a closed-form approximate solution. The sparse-regime approximation allows to estimate the expected degree distribution of the sampled graph based on the fitness distribution p_f. We derive two estimates for the degree distribution p_k of the network. The first estimate reads

$$p_k(k) \approx \frac{\langle f \rangle_{p_f}}{N} \sum_I p_f \left(k \frac{\langle f \rangle_{p_f}}{\mu_I} \right) \frac{N_I}{\mu_I} \tag{3}$$

where μ_I is the expected average degree of the vertices in B_I. The second estimate, less accurate but easier to interpret and use in practice, reads

$$p_k(k) \approx p_f \left(k \frac{\langle f \rangle_{p_f}}{\mu} \right) \frac{\langle f \rangle_{p_f}}{\mu} \tag{4}$$

where μ is the average degree of the network. The obtained analytical expressions for p_k have a very desirable property: for many choices of p_f—including power-law, lognormal or exponential distributions—p_k essentially has the same "shape" of p_f.

Data-driven FCBM for spatial social networks. We propose an application for our FCBM as a maximum-entropy model for data-driven spatial social networks. In particular, we envision its application to generate geographic networks informed by census data, contact surveys, and geospatial data. Indeed, the phenomenological model described at the end of this section has already been employed to develop a realistic social network at the urban scale[9] and to study the spread of an epidemic process[15–17], on it.

Let the vertex set V describe an age-stratified population of N individuals living in a territory tessellated into square tiles of side l. Each v_i is thus characterized by two data-driven discrete attributes: its tile of residence $t_i \in T$, that is, the discretized position of v_i in the territory, and its age-group $g_i \in \Gamma$. These two attributes induce a partition of the population into $n = |T| \cdot |\Gamma|$ blocks $\{B_I\}_{i=0}^{n-1}$, with $v_i \in B_I = (t_I, g_I)$ if and only if $t_i = t_I$ and $g_i = g_I$.

To define a data-driven block-wise mixing matrix Δ, we observe that:

- Social mixing patterns, derived from heterogeneous data sources such as surveys, cell phones or wearable sensors[18–20], can be used to reconstruct a data-driven age-based mixing matrix S[9], whose s_{IJ} element measures the tendency of age groups g_I and g_J to socialize with each other.
- The frequency of social relations between individuals living in t_I and t_J is generally assumed to decay as $d_{IJ}^{-\beta}$, where d_{IJ} is the normalized (geographic or euclidean) distance between tiles t_I and t_J, and the exponent $\beta > 0$ depends on the type of relation and on the extension of the territory[12,13].

This leads to:

$$\Delta_{IJ} = \frac{d_{IJ}^{-\beta} s_{IJ}}{\sum_{O \leq Q} d_{OQ}^{-\beta} s_{OQ}} \tag{5}$$

Finally, we extract the fitness vector $\vec{f}$ from a suitable distribution p_f, set the density parameter $p \in (0, 1)$ and, for all pairs i, j, compute the edge probability p_{ij} as prescribed by the FCBM. As a result, the graph sampling probability P guarantees that: (1) the expected number of links between B_I and B_J is proportional to $s_{I,J}$ and decays as $d_{I,J}^{-\beta}$; (2) the expected degree of v_i is proportional to f_i and to the expected total degree of block I_i.
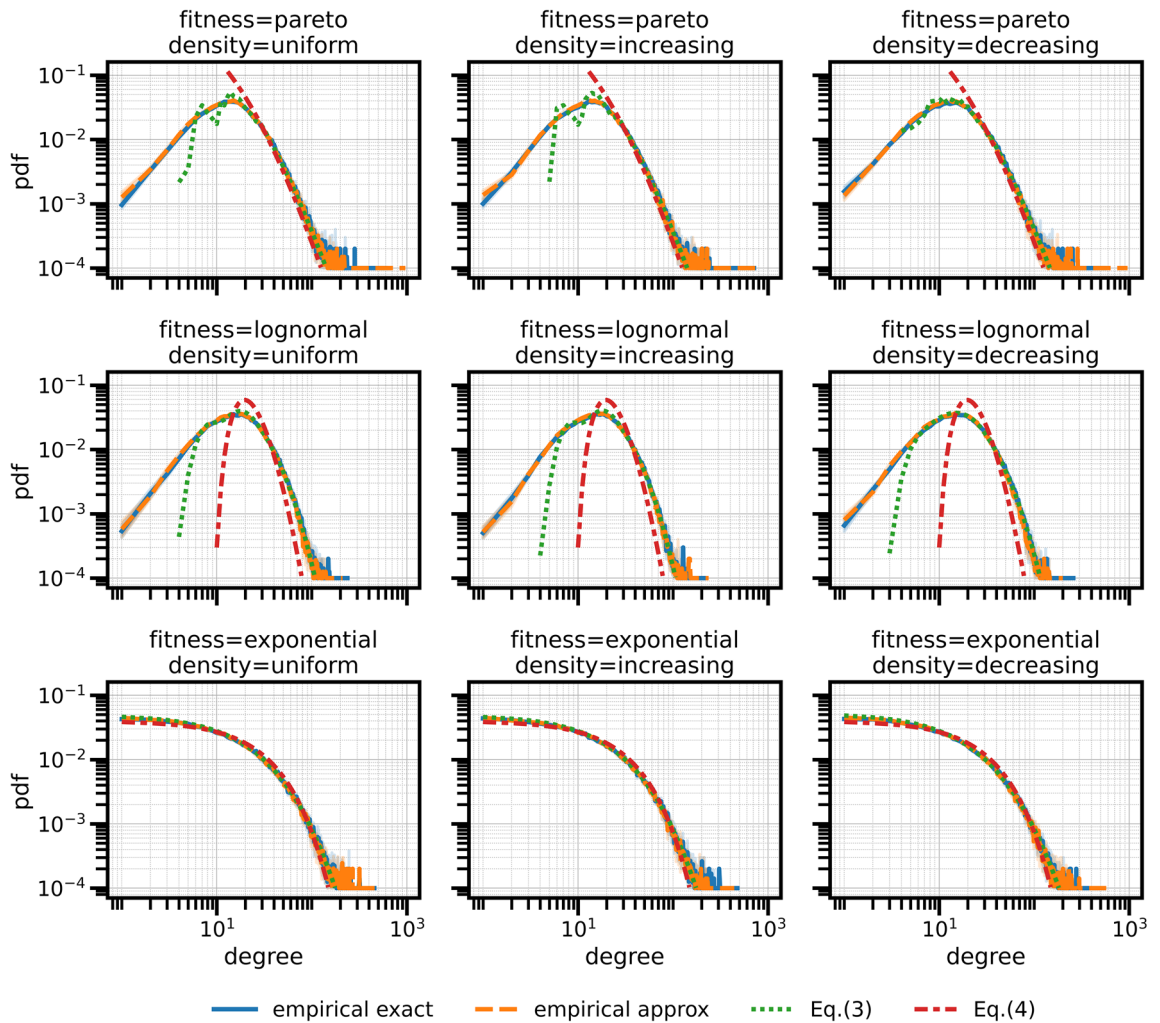
Figure 1. The degree distribution for an ideal city (circular, 10K inhabitants) for 3 different densities (uniform, radial increasing, radial decreasing) and 3 different fitness distributions (powerlaw, lognormal, exponential).

In this case, the sparse-regime approximation yields

$$p_{ij} \approx p \binom{N}{2} \frac{f_i}{\sum_{u \in I_i} f_u} \frac{f_j}{\sum_{w \in J_j} f_w} \frac{d_{I_i J_j}^{-\beta} s_{I_i J_j}}{\sum_{I \leq J} d_{IJ}^{-\beta} s_{IJ}} \tag{6}$$

Expression (6) defines a phenomenological model, analogous to the one presented in Ref.[9], where the probability of two individuals being connected is proportional to their sociability and to the cohesion of their age-groups, while decaying as a power of their distance. Clearly, the estimates obtained in (3) and (4) for the degree distribution stay valid in the data-driven model. If the network is sufficiently sparse and the population of all tiles/groups is sufficiently large, the degree distribution of the sampled graph G is controlled by the available data and by the chosen fitness distribution p_f.

Experimental analysis. We implemented both the exact FCBM and its sparse-regime approximation. The code is released, as open-source software, as part of the Urban Social Networks (USN) framework at https://gitlab.com/cranic-group/usn.

We used our data-driven FCBM to generate instances of a spatial social network for a stylized city of 10K inhabitants living in disk of radius 2.5 Km. We set p so that the average degree of the network is $\mu = 25$, we set $\beta = 1$, we used age-density data for Italy as released by the Italian National Institute of Statistics (ISTAT), and we extracted the matrix S from data released by the POLYMOD project[21]. We considered three possible spatial densities—uniform, and increasing or decreasing with the distance from the disk's center—and three possible fitness distributions—pareto, lognormal and exponential. For all nine combinations, we generated 10 independent graph instances.

In Fig. 1 we show the empirical degree distribution, averaged over the 10 instances of the FCBM for each configuration, considering both the exact model and the sparse-regime approximation. We also show the two estimates (3) and (4). The plots confirm that the approximation is sound and that the two estimates can be safely

used in practice—at least, for sparse networks—with (3) working especially well in all cases—except, possibly, for very small degrees.

Discussion

Developing realistic network models for social interactions is of paramount importance to understand the underlying mechanisms that lead to the observed features of such networks and to study all dynamical processes, such as disease spreading, that are strongly influenced by the network topology. Extreme care must be taken to avoid that any bias is unintentionally injected in the model. For this reason, maximum-entropy models are extensively used in network analysis, either as null models, or to generate synthetic networks that have specific characteristics, but are otherwise maximally random.

In this paper, we introduced the maximum-entropy Fitness-Corrected Block Model (FCBM), an adjustable-density model for modular and heterogeneous networks, whose block-wise mixing pattern is known in expectation. The model has a general and flexible formulation, but it was designed with a key application in mind: providing a working tool for building synthetic social networks informed with census, survey and geospatial data. Publicly available spatial density and demographic data are in fact necessarily discrete, thus inducing a partition of the population into blocks of agents who belong to the same age-group and live in the same area. The expected block-to-block edge-density can be estimated based on empirical findings that quantify the dependence of contact frequencies upon geographic and demographic features. However, both the density and the degree of heterogeneity of real-world social networks depend on the considered type of interpersonal relations and are rarely explicitly known beforehand. Contrarily to other block-models in the literature, the FCBM makes both these network features adjustable. In particular, the desired intra-block heterogeneity can be enforced through a vertex-intrinsic social fitness, possibly modelled upon observable population-level variables, such as wealth, employment or mobility.

We implemented the FCBM and made it publicly available as open-source software. The released software includes a parallel code that speeds up the computation of the most expensive part of the required maximization procedure, a step of the algorithm that is also needed in the well-known Degree-Corrected Block Model. We tested our implementation of the FCBM by reconstructing instances of a social network connecting the individuals of a stylized city of 10K inhabitants. The experiments allowed to verify that the exact model and its efficient sparse-regime approximation yield networks with almost identical degree distributions. We also showed and experimentally verified that, in the sparse-regime, the expected degree distribution of the output network can be estimated by two closed-form expressions. Thanks to these two estimates, the shape of the degree distribution can be predicted based on the chosen fitness distribution.

In the next future, we plan to use the FCBM—and, possibly, a temporal extension of the model—to simulate dynamical processes in real-world territories and understand how socio-demographic features and social habits affect the outcomes of these processes. To this end, we will work towards gaining a better understanding of how the topological properties of the FCBM depend on the configuration parameters, with special attention paid to a set of network properties, such as the clustering coefficient and the excess degree distribution, that can be used to study percolation and diffusion dynamics on the network.

Related work

Under the pressure of the COVID-19 pandemic, several simulation frameworks have been developed to provide realistic descriptions of the disease spread process on different spatial and temporal scales, from a single building to complex urban areas up to a global scale[22–29]. World-scale meta-population models and agent-based systems describing small and large areas can be informed by a variety of data sources. Census data and/or grid based population counts can be integrated to reconstruct populations that are statistically indistinguishable from real ones, including age, geographic distribution, education and wealth. The number and intensity of contacts in specific settings, such as workplaces, schools, or households, can be collected through surveys, questionnaires, diaries, and, if possible, supplemented with data obtained from digital technologies such as cell phones or wearable sensors[18–20]. These data are then used to reconstruct contact matrices, individual or group schedules, and are widely used to reproduce synthetic interactions[30,31].

In comparison, generative network models that can reproduce the characteristics of real-world social networks by incorporating information from data, have received much less attention. The task of inferring a realistic distribution of social ties is quite challenging, since friendship ties can only be measured for a small subset of real-world networks, and the mechanism underlying tie formation, while thoroughly studied, is still far from being fully understood. Ideally, the models should reproduce the main features of real-world social networks, well summarized in Ref.[32]. These networks show a heavy-tailed (e.g, lognormal) degree distribution, often with a finite cutoff in agreement with Dunbar's number. The transitivity of the networks is high, compared to a random graph model, as a consequence of the well-established principle that "friends of my friends are my friends". Moreover, they show positive assortativity by degree and *type*. By quantitatively looking at ego networks, mobile phone networks, and online social networks we now have a better understanding of some of their peculiar features and underline mechanisms. Education, wellness, age and spatial proximity are regarded as critical elements in the formation of friendship bonds[33,34]. Among these, age is probably the most studied, possibly due to the fact that age-related data is actually available at different spatial scales[35]. While there is wide evidence that geographical factors alone cannot explain the structure of real-world spatial social networks[12,36,37], the dependence of friendship on distance is widely assumed to follow an inverse power-law with exponent $\beta \in [0.5, 2]$[12,13,36–41]—and this surprisingly holds even for online relationships[42]. In particular, $\beta < 1$ seems to work better for short range contacts (< 20 km)[13] and for urban networks[40], in line with sociological studies[43]. However, contrary to other real-world networks[44,45], such networks do not present very large hubs[12,37,38,46]. Their degree distribution is right

skewed and relatively long-tailed[12,37], and it has been, at times, approximated by a power-law with a large (5–8) exponent[38,46] or by a lognormal distribution[13]. Within cities, population density impacts on the frequency of close-range contacts, but usually not on the overall size of each person's network[41]. While geographical proximity and community structure appear to be related[37,40,41], some authors argue that only small clusters (< 30 members) are geographically bounded[39], whereas the large ones may span across very large areas of a city[37].

Defining simple models that capture all of these features is not an easy task. Models designed to mimic the scale-free degree distribution emerging in many real networks, for instance, may fail to yield the expected clustering structure[47,48]. Exponential random graphs have been shown to overcome some of these limitations[49,50]. Stochastic Block Models (SBMs)[2,3], on the other hand, have been specifically developed to reproduce networks with a community structure, a typical characteristic of real-world social networks, which follows from the presence of some kind of group-level homophily. In this type of network models, the nodes are partitioned into disjoint sets named *blocks* and the probability of an edge existing between two nodes depends on the blocks to which the two nodes belong. The SBM and its generalization have gained their success in the last decades as they can be used to discover and understand the structure of a network, as well as for clustering purposes[51,52].

Spatial network models are often obtained by incorporating vertices into a metric space and induced constraints can determine some of the network properties[53]. Introducing a penalty on "long" edges, which mimics a penalty in maintaining long-distance relationships, has an impact on clusters, path lengths, degree distributions, and more[54]. Recently, network instances having suitable features have been generated by means of the so-called *random geometric* models[55–57], where the popularity and similarity of the nodes depend on their position in some *latent* metric space[58]. Embedding the vertices into a hyperbolic disk[55] has proved a way to obtain both high clustering and heavy-tailed degree distribution.

In the present paper, we leverage on the framework of entropy-based null-models for real complex networks, revised in Ref.[7]. Among the very first fundamental papers, the work of Park and Newman has a particular relevance[59]: based on Jaynes' derivation of Statistical Physics from Information Theory[60], they proposed a general maximum entropy approach for the randomization of complex networks. Among the extension to a different context, the main innovations of Ref.[59] are the introduction of local constraints, as the degree sequence, and the interpretation of the general framework of Exponential Random Graphs (ERGs) in terms of maximum entropy models. The present construction was later extended to the analysis of real networks[61,62], tailoring the entropy-based model on the observed network. As a matter of fact, the various Lagrangian multipliers, introduced for the entropy maximisation in Ref.[59], can be numerically calculated by maximising the (Log-)Likelihood associated with the real network. Such a construction represents a perfect benchmark for the analysis of real systems, since it is maximally random (due to the entropy maximisation) and tailored on the observed system (due to the Likelihood maximisation). It is not surprising that it has been extensively applied to the study of non trivial structural patterns of different systems, as financial and trade networks, biological systems and online social networks[7,63]. Moreover, the general framework can be easily extended to tackle different kinds of networks, as undirected, directed, weighted, directed and weighted[64–66], bipartite[67], bipartite weighted[68] and degree corrected block models[4]. Another relevant branch of research focuses on the reconstruction of networks from limited information[69]. This application is of particular interest for risk assessment of financial networks, and limited information is available due to privacy concerns[70]. Reconstruction approaches based on entropy-based null models have proven particularly effective in this context, and the maximally random nature of the framework described here is critical to avoid the introduction of bias into the predictions[69–73].

Methods

Formalism.
Let $\mathcal{G}$ be the ensemble of all simple graphs of N vertices. If P is a probability distribution over $\mathcal{G}$, $P(G)$ is the probability of graph $G \in \mathcal{G}$, and $\langle \cdot \rangle_P$ denotes the expectation with respect to P. The vertex set $V = \{v_i\}_{i=0}^{N-1}$ is partitioned into n blocks $\{B_I\}_{I=0}^{n-1}$. The size of block I is $N_I = |B_I|$ and, for each $v_i \in V$, I_i denotes the index of the block to which v_i belongs. For all pairs I, J, N_{IJ} denotes the number of possible pairs (i, j) with $v_i \in I$, $v_j \in J$ and $i \neq j$, i.e., $N_{II} = N_I(N_I - 1)$ and $N_{IJ} = N_I N_J$ if $I \neq J$—notice that in the case of N_{II} we are counting twice the number of couples, as we will do for the counts of edges in the following.

Each $G \in \mathcal{G}$ is uniquely determined by its adjacency matrix $A(G) = \{a_{ij}(G)\}_{i,j=0}^{N-1}$, where $a_{ij}(G) = 1$ if edge $(i, j) \in E(G)$ and $a_{ij}(G) = 0$ otherwise. The degree of vertex v_i in G is $\deg_i(G) = \sum_j a_{ij}(G)$. The total degree of block I is $\deg_I(G) = \sum_{i \in I} \deg_i(G)$ and, for all I, J, $L_{IJ}(G) = \sum_{i \in I} \sum_{j \in J} a_{ij}(G)$ is the number of edges between B_I and B_J in G, or, if $I = J$, twice that number. Therefore, using the definitions, $\deg_I(G) = \sum_J L_{IJ}(G)$. For the sake of simplicity, the dependence of these quantities on the specific graph G will be often omitted in the following.

Maximum entropy Degree-Corrected Block Model[4].
The maximum entropy Degree-Corrected Block Model (DCBM) is defined as the maximum entropy probability distribution P over $\mathcal{G}$ in which the number of links per block and the degree sequence are constrained on average, i.e.

$$\langle L_{IJ} \rangle_P = K_{IJ} \quad \text{for all } I, J \tag{7}$$

$$\langle \deg_i \rangle_P = k_i \quad \text{for all } i, \tag{8}$$

with $\sum_J K_{IJ} = \sum_{i \in I} k_i$, for all I. If $H(P)$ denotes the Shannon entropy of P, the sought P can be obtained by finding the stationary points of

$$H'(P, \vec{\eta}, \vec{\theta}) = H(P) - C(P, \vec{\eta}, \vec{\theta}) = -\langle \ln P(G) \rangle_P - \left(\sum_{I \leq J} \eta_{IJ} (\langle L_{IJ} \rangle_P - K_{IJ}) + \sum_i \theta_i (\langle \deg_i \rangle_P - k_i) + \alpha (\sum_G P(G) - 1) \right) \quad (9)$$

where η_{IJ}, θ_i and α are Lagrange multipliers: while, η_{IJ} and θ_i control the conditions (7) and (8), respectively, α is necessary for the normalization of the probability $P(G)$. Since the functional derivatives with respect to $P(G)$ are $\frac{d\langle L_{IJ} \rangle_P}{dP(G)} = \sum_{i \in I} \sum_{j \in J} a_{ij}(G)$ and $\frac{d\langle \deg_i \rangle_P}{dP(G)} = \sum_j a_{ij}(G)$, then

$$\frac{dC(P, \vec{\eta}, \vec{\theta})}{dP(G)} = \sum_{I \leq J} \eta_{IJ} \sum_{i \in I} \sum_{j \in J} a_{ij}(G) + \sum_i \theta_i \sum_j a_{ij}(G) + \alpha = \sum_{i<j} \eta_{I_iJ_j} a_{ij}(G) + \sum_{i<j} (\theta_i + \theta_j) a_{ij}(G) + \alpha$$

This results in

$$P(G) \propto \exp \left[-\sum_{i<j} (\theta_i + \theta_j + \eta_{I_iJ_j}) a_{ij}(G) \right]$$

and the probability per graph factorises in terms of probabilities per link as

$$p_{ij} = \begin{cases} \dfrac{x_i x_j y_{I_iJ_j}}{1 + x_i x_j y_{I_iJ_j}} & \text{if } i \neq j \\ 0 & \text{otherwise} \end{cases} \quad (10)$$

where $x_i = e^{-\theta_i}$ and $y_{IJ} = e^{-\eta_{IJ}}$. For all i and all $I \leq J$, x_i and y_{IJ} can be found by solving the system of equations

$$\sum_{i \in I} \sum_{j \in J} p_{ij} = K_{IJ} \quad \text{for all } I \leq J \quad (11)$$

$$\sum_j p_{ij} = k_i \quad \text{for all } i \quad (12)$$

Sparse DCBM. When the average degree $\langle k \rangle$ is small, i.e., when the network is sparse, the edge probability of the DCBM can be approximated as $p_{ij} \approx x_i x_j y_{I_iJ_j}$. This allows to rewrite (11) and (12) as

$$y_{IJ} \sum_{i \in I} x_i \sum_{\substack{j \in J \\ j \neq i}} x_j = K_{IJ} \quad \text{for all } I \leq J \quad (13)$$

$$x_i \sum_J y_{I_iJ} \sum_{\substack{j \in J \\ j \neq i}} x_j = k_i \quad \text{for all } i \quad (14)$$

Since the constants K and $\vec{k}$ are, by construction, bound by the relation $\sum_J K_{IJ} = \langle \deg_I \rangle = \sum_{i \in I} k_i$, (13) and (14) admit the following solution

$$x_i = k_i \quad \text{for all } i$$

$$y_{IJ} = \frac{K_{IJ}}{\left(\sum_{i \in I} k_i \right) \left(\sum_{\substack{j \in J \\ j \neq i}} k_j \right)} \quad \text{for all } I \leq J$$

Maximum entropy Fitness-Corrected Block Model. Given a scalar $p \in (0, 1)$, a symmetric matrix Δ such that $\sum_{I,J} \Delta_{I,J} = 2$, and a fitness sequence $\vec{f}$, we define the Fitness-Corrected Block Model (FCBM) as the maximum entropy model fulfilling the following two conditions:

$$\langle L_{IJ} \rangle_P = p \binom{N}{2} \Delta_{IJ} \quad \text{for all } I, J \quad (15)$$

$$\langle \deg_i \rangle_P = p \binom{N}{2} \frac{f_i}{\sum_{u \in I_i} f_u} \sum_J \Delta_{I_iJ} \quad \text{for all } i \quad (16)$$

The FCBM is a variation of the DCBM where the network density p is a configuration parameter. As in the original stochastic block model, the block-level mixing is specified in terms of a set of edge-densities Δ_{IJ}, rather

than a set of edge-counts K_{IJ}. Similarly, the degree sequence $\vec{k}$ is replaced by a vertex intrinsic fitness sequence $\vec{f}$, in line with previous models available in the literature[5,6]. f_i measures v_i's propensity to establish links and $\langle \deg_i \rangle_P$ is set proportional to f_i by a constant that depends on I_i, other than p. By design, (15) and (16) imply $\sum_J \langle L_{IJ} \rangle_P = \langle \deg_I \rangle_P = \sum_{i \in I} \langle \deg_i \rangle_P$, and (16) can be rewritten as

$$\langle \deg_i \rangle_P = \frac{f_i}{\sum_{u \in I_i} f_u} \langle \deg_I \rangle_P$$

which clarifies the role of $\vec{f}$ as an element of intra-block heterogeneity.

For fixed Δ and $\vec{f}$, the derivation of the maximum entropy FCBM is identical to that of the DCBM: the maximum entropy probability per graph factorizes into the probability per edge given by (10) and the vectors of constants $\vec{x}$ and $\vec{y}$ can be obtained by solving the analogous of (11) and (12), i.e.

$$\sum_{i \in I} \sum_{j \in J} p_{ij} = p \binom{N}{2} \Delta_{IJ} \quad \text{for all } I \leq J \tag{17}$$

$$\sum_j p_{ij} = p \binom{N}{2} \frac{f_i}{\sum_{u \in I_i} f_u} \sum_J \Delta_{I_i J} \quad \text{for all } i \tag{18}$$

Sparse FCBM. In the sparse-regime, i.e, when $p \ll 1$, using $p_{ij} \approx x_i x_j y_{I_i J_j}$ yields the following approximate solution to (17) and (18):

$$x_i = p \binom{N}{2} \frac{f_i}{\sum_{u \in I_i} f_u} \sum_J \Delta_{I_i J} = \frac{f_i}{\sum_{u \in I_i} f_u} \langle \deg_I \rangle_P \quad \text{for all } i$$

$$y_{IJ} = \frac{p \binom{N}{2} \Delta_{IJ}}{p \binom{N}{2} (\sum_O \Delta_{IO}) p \binom{N}{2} (\sum_O \Delta_{OJ})} = \frac{p \binom{N}{2} \Delta_{IJ}}{\langle \deg_I \rangle_P \langle \deg_J \rangle_P} \quad \text{for all } I \leq J$$

In this regime, the probability per edge can thus be rewritten as

$$p_{ij} \approx p \binom{N}{2} \frac{f_i}{\sum_{u \in I_i} f_u} \frac{f_j}{\sum_{w \in J_j} f_w} \Delta_{I_i J_j}. \tag{19}$$

Degree distribution for the sparse FCBM. Let us assume that each fitness value f_i is drawn from a suitable distribution p_f. The sparse-regime approximation allows to estimate the expected degree distribution of the sampled graph G with respect to both p_f and the graph sampling probability P. If N_{I_i} is large enough, we have $\sum_{u \in I_i} f_u \approx \langle f \rangle_{p_f} N_{I_i}$. Now, following the approach used in Ref.[5], we have

$$k(f, I) = \langle \deg_i \mid f_i = f, I_i = I \rangle_{p_f, P} \approx \frac{f}{\langle f \rangle_{p_f} N_I} \langle \deg_I \rangle_P = \frac{f}{\langle f \rangle_{p_f}} \mu_I \tag{20}$$

where $\mu_I = \frac{\langle \deg_I \rangle_P}{N_I} = \langle \deg_i \mid I_i = I \rangle_P$ is the expected average degree of the vertices in I. Equation (20) can be inverted leading to estimate

$$f(k, I) \approx k \frac{\langle f \rangle_{p_f}}{\mu_I}$$

so that

$$p_k(k, I) = \Pr[k(f) = k \mid I] = \Pr[f(k) = f \mid I] \frac{df(k)}{dk} \approx p_f \left(k \frac{\langle f \rangle_{p_f}}{\mu_I} \right) \frac{\langle f \rangle_{p_f}}{\mu_I} \tag{21}$$

and, hence,

$$p_k(k) = \Pr[k(f) = k] = \sum_I p_k(k, I) \frac{N_I}{N} \approx \sum_I p_f \left(k \frac{\langle f \rangle_{p_f}}{\mu_I} \right) \frac{\langle f \rangle_{p_f}}{\mu_I} \frac{N_I}{N} = \frac{\langle f \rangle_{p_f}}{N} \sum_I p_f \left(k \frac{\langle f \rangle_{p_f}}{\mu_I} \right) \frac{N_I}{\mu_I} \tag{22}$$

A slightly less accurate, yet much simpler, approximation can be obtained computing first

$$k(f) = \langle \deg_i \mid f_i = f \rangle_{p_f,P} = \sum_I k(f,I) \frac{N_I}{N} \approx \frac{f}{\langle f \rangle_{p_f}} \sum_I \mu_I \frac{N_I}{N} = \frac{f}{\langle f \rangle_{p_f}} \mu \tag{23}$$

where $\mu = \sum_I \mu_I \frac{N_I}{N}$ is the average degree of the network. Then, (23) can be inverted as

$$f(k) \approx k \frac{\langle f \rangle_{p_f}}{\mu}$$

yielding

$$p_k(k) = \Pr[k(f) = k] = \Pr[f(k) = f] \frac{\mathrm{d}f(k)}{\mathrm{d}k} \approx p_f \left(k \frac{\langle f \rangle_{p_f}}{\mu} \right) \frac{\langle f \rangle_{p_f}}{\mu} \tag{24}$$

In many cases, p_k belongs to the same family of probability distributions of p_f: e.g., if p_f follows a power-law, lognormal or exponential distribution, then the same holds for p_k.

Data-driven FCBM for spatial social networks. Our FCBM can be easily tuned upon real data and empirical findings to produce instances of a maximum entropy spatial social network. On one hand, the local density and demographic profile of the population are generally available in the form of discrete, geographically located (e.g., residents in 500 m × 500 m tiles) and/or age-stratified (e.g., 0–5 years old) population segments. These data naturally induce a partition of the population into blocks. On the other hand, intra-block population heterogeneity can be controlled by a vertex-related social fitness, possibly modelled upon measurable features such as wealth, employment or mobility.

Formally, let the vertex set V describe an age-stratified population of N individuals living in a territory tessellated into square tiles of side l. Each v_i is characterized by two data-driven discrete attributes: its tile of residence $t_i \in T$, that is, the discretized position of v_i in the territory, and its age-group $g_i \in \Gamma$. t_i and g_i may either be directly available—in the case of a real population—or be drawn, respectively, from given spatial density p_t and age-distribution p_g—in the case of a synthetic population. These two attributes induce a partition of the population into $n = |T| \cdot |\Gamma|$ blocks $\{B_I\}_{i=0}^{n-1}$, with $v_i \in B_I = (t_I, g_I)$ if and only if $t_i = t_I$ and $g_i = g_I$. We embrace the widely-acknowledged assumption that an inverse-power-law relation binds the distance d_{IJ} and the frequency of social relations between individuals living in t_I and t_J[12,13]. For all pairs of blocks I, J, we thus define the edge-density

$$\Delta_{IJ} = \frac{d_{IJ}^{-\beta} s_{IJ}}{\sum_{O \leq Q} d_{OQ}^{-\beta} s_{OQ}}$$

where d_{IJ} is the normalized (geographic or euclidean) distance between tiles t_I and t_J; the normalization is obtained through a division by $\frac{l}{2}$. We set $d_{II} = 1$, so that the distance between individuals in the same tile is half the distance of individuals living in neighboring tiles. $\beta > 0$ is a configuration parameter. s_{IJ} measures the tendency of age groups g_I and g_J to socialize with each other; such a $|\Gamma| \times |\Gamma|$ symmetric age-based social mixing matrix S can be obtained, by imposing reciprocity and normalizing, from a suitable data-driven contact matrix[9].

Finally, we extract the fitness vector $\vec{f}$, set the density parameter $p \in (0,1)$ and, for all pairs i, j, compute the edge probability p_{ij} as described for the FCBM. As a result, the graph sampling probability P guarantees that: (1) the expected number of links between B_I and B_J is proportional to $s_{I,J}$ and decays as $d_{I,J}^{-\beta}$; (2) the expected degree of v_i is proportional to f_i and to the expected total degree of block I_i.

In this case, the sparse-regime approximation yields

$$p_{ij} \approx p \binom{N}{2} \frac{f_i}{\sum_{u \in I_i} f_u} \frac{f_j}{\sum_{w \in J_j} f_w} \frac{d_{I_i J_j}^{-\beta} s_{I_i J_j}}{\sum_{I \leq J} d_{IJ}^{-\beta} s_{IJ}} \tag{25}$$

Expression (25) defines a phenomenological model, where the probability of two individuals being connected is proportional to their sociability and to the cohesion of their age-groups, while decaying as a power of their distance. Clearly, the estimates obtained in (22) and (24) for the degree distribution stay valid in the data-driven model. If the network is sufficiently sparse and the population of all tiles/groups is sufficiently large, the degree distribution of the sampled graph G is controlled by the available data and by the chosen fitness distribution p_f.

Data availability

References
1. McPherson, M., Smith-Lovin, L. & Cook, J. M. Birds of a feather: Homophily in social networks. Annual review of sociology 27, 415–444 (2001).

2. Karrer, B. & Newman, M. E. Stochastic blockmodels and community structure in networks. *Physical review E***83**, 016107 (2011).
3. Peixoto, T. P. Hierarchical block structures and high-resolution model selection in large networks. *Phys. Rev. X***4**, 011047. https://doi.org/10.1103/PhysRevX.4.011047 (2014).
4. Fronczak, P., Fronczak, A. & Bujok, M. Exponential random graph models for networks with community structure. *Physical Review E***88**, 032810. https://doi.org/10.1103/PhysRevE.88.032810 (2013).
5. Caldarelli, G., Capocci, A., De Los Rios, P. & Munoz, M. A. Scale-free networks from varying vertex intrinsic fitness. *Physical review letters***89**, 258702 (2002).
6. Servedio, V. D., Caldarelli, G. & Buttà, P. Vertex intrinsic fitness: How to produce arbitrary scale-free networks. *Physical Review E***70**, 056126 (2004).
7. Cimini, G. *et al.* The statistical physics of real-world networks. *Nat. Rev. Phys.* **1**, 58–71. https://doi.org/10.1038/s42254-018-0002-6. arxiv:1810.05095 (2019).
8. Guarino, S. *et al.* A model for urban social networks. In *International Conference on Computational Science*, 281–294 (Springer, Cham, 2021).
9. Guarino, S. et al. Inferring urban social networks from publicly available data. Future Internet. https://doi.org/10.3390/fi13050108 (2021).
10. Willem, L. et al. Socrates: An online tool leveraging a social contact data sharing initiative to assess mitigation strategies for Covid-19. BMC Research Notes. https://doi.org/10.1186/S13104-020-05136-9 (2020).
11. Verelst, F. *et al.* Socrates-comix: a platform for timely and open-source contact mixing data during and in between covid-19 surges and interventions in over 20 european countries. *BMC Med.* **19**, 1–7 (2021).
12. Liben-Nowell, D., Novak, J., Kumar, R., Raghavan, P. & Tomkins, A. Geographic routing in social networks. Proceedings of the National Academy of Sciences 102, 11623–11628 (2005).
13. Illenberger, J., Nagel, K. & Flötteröd, G. The role of spatial interaction in social networks. Networks and Spatial Economics 13, 255–282 (2013).
14. Vallarano, N. et al. Fast and scalable likelihood maximization for exponential random graph models with local constraints. Sci. Rep. 2021, 11: 1–33, 2021, doi: 10.1038/s41598-021-93830-4.
15. Guarino, S. *et al.* Data-driven simulation of contagions in public venues. In *2021 Annual Modeling and Simulation Conference (ANNSIM)*, 1–12 (IEEE, 2021).
16. Celestini, A., Colaiori, F., Guarino, S., Mastrostefano, E. & Zastrow, L. R. Epidemics in a synthetic urban population with multiple levels of mixing. In *International Conference on Complex Networks and Their Applications*, 315–326 (Springer, Cham, 2021).
17. Celestini, A., Colaiori, F., Guarino, S., Mastrostefano, E. & Zastrow, L. R. Epidemic risk assessment from geographic population density. *Applied Network Science* **7**, 39. https://doi.org/10.1007/s41109-022-00480-0 (2022).
18. Mossong, J. et al. Social contacts and mixing patterns relevant to the spread of infectious diseases. PLOS Medicine 5, 1–1 (2008).
19. Eagle, N., Pentland, A. S. & Lazer, D. Inferring friendship network structure by using mobile phone data. Proceedings of the national academy of sciences 106, 15274–15278 (2009).
20. Klepac, P. *et al.* Contacts in context: large-scale setting-specific social mixing matrices from the BBC pandemic project. *medRxiv*https://doi.org/10.1101/2020.02.16.20023754 *(2020).*
21. Mossong, J. *et al.* Social contacts and mixing patterns relevant to the spread of infectious diseases. *PLOS Medicine***5**, e74. https://doi.org/10.1371/JOURNAL.PMED.0050074 (2008).
22. Kerr, C. C. *et al.* Covasim: an agent-based model of covid-19 dynamics and interventions. *PLoS Comput. Biol.* **17**, e1009149 (2021).
23. Mahmood, I. et al. Facs: A geospatial agent-based simulator for analysing Covid-19 spread and public health measures on local regions. J. Simul. 16:1–19 (2020).
24. Liu, P., McQuarrie, L., Song, Y. & Colijn, C. Modelling the impact of household size distribution on the transmission dynamics of covid-19. Journal of the Royal Society Interface 18, 20210036 (2021).
25. Coletti, P. *et al.* A data-driven metapopulation model for the belgian covid-19 epidemic: assessing the impact of lockdown and exit strategies. *BMC Infect. Dis.* **21**, 1–12 (2021).
26. Bouchnita, A. & Jebrane, A. A hybrid multi-scale model of covid-19 transmission dynamics to assess the potential of non-pharmaceutical interventions. *Chaos, Solitons & Fractals* **138**, 109941 (2020).
27. Aleta, A. et al. Modelling the impact of testing, contact tracing and household quarantine on second waves of covid-19. Nature Human Behaviour 4, 964–971 (2020).
28. Aleta, A. *et al.* Quantifying the importance and location of sars-cov-2 transmission events in large metropolitan areas. *Proc. Natl. Acad. Sci.* **119**, e2112182119 (2022).
29. Chang, S. et al. Mobility network models of covid-19 explain inequities and inform reopening. Nature 589, 82–87 (2021).
30. Del Valle, S. Y., Hyman, J. M., Hethcote, H. W. & Eubank, S. G. Mixing patterns between age groups in social networks. *Social Networks* **29**, 539–554 (2007).
31. Barrett, C. L. *et al.* Generation and analysis of large synthetic social contact networks. In *Proceedings of the 2009 Winter Simulation Conference (WSC)*, 1003–1014 (IEEE, 2009).
32. Kertész, J., Török, J., Murase, Y., Jo, H.-H. & Kaski, K. Modeling the complex network of social interactions. In *Pathways Between Social Science and Computational Social Science*, 3–19 (Springer, 2021).
33. Palla, G., Barabási, A.-L. & Vicsek, T. Quantifying social group evolution. Nature 446, 664–667 (2007).
34. Huang, Y., Shen, C. & Contractor, N. S. Distance matters: Exploring proximity and homophily in virtual world networks. Decis. Support Syst. 55, 969–977 (2013).
35. Worldpop. https://www.worldpop.org/ (2020).
36. Scellato, S., Noulas, A., Lambiotte, R. & Mascolo, C. Socio-spatial properties of online location-based social networks. In *Proceedings of the International AAAI Conference on Web and Social Media*, Vol. 5 (2011).
37. Herrera-Yagüe, C. et al. The anatomy of urban social networks and its implications in the searchability problem. Scientific reports 5, 10265 (2015).
38. Lambiotte, R. et al. Geographical dispersal of mobile communication networks. Physica A: Statistical Mechanics and its Applications 387, 5317–5325 (2008).
39. Onnela, J.-P., Arbesman, S., González, M. C., Barabási, A.-L. & Christakis, N. A. Geographic constraints on social network groups. *PLoS one***6**, e16939 (2011).
40. Walsh, F. & Pozdnoukhov, A. Spatial structure and dynamics of urban communities (2011).
41. Büchel, K. & Ehrlich, M. V. Cities and the structure of social interactions: Evidence from mobile phone data. *Journal of Urban Economics***119**, 103276 (2020).
42. Goldenberg, J. & Levy, M. Distance is not dead: Social interaction and geographical distance in the internet era. arXiv:0906.3202 (2009).
43. Krackhardt, D., Nohria, N. & Eccles, B. The strength of strong ties. Netw. Knowl. Econ. 82:1 (2003).
44. Bernaschi, M., Celestini, A., Guarino, S., Lombardi, F. & Mastrostefano, E. Spiders like onions: On the network of tor hidden services. In *The World Wide Web Conference*, 105–115 (2019).
45. Newman, M. *Networks: An Introduction* (OUP, Oxford, 2010).
46. Onnela, J.-P. et al. Analysis of a large-scale weighted network of one-to-one human communication. New journal of physics 9, 179, 2007, DOI: 10.1088/1367-2630/9/6/179.

47. Cointet, J.-P. & Roth, C. How realistic should knowledge diffusion models be?. *J. Artif. Soc. Soc. Simul.* **10**, 1–11 (2007).
48. Iskhakov, L., Kamiński, B., Mironov, M., Prałat, P. & Prokhorenkova, L. Local clustering coefficient of spatial preferential attachment model. J. Complex Netw. **8**, 019 (2020).
49. Robins, G., Snijders, T., Wang, P., Handcock, M. & Pattison, P. Recent developments in exponential random graph (p*) models for social networks. Soc. Netw. 29, 192–215. 2007, doi: 10.1016/j.socnet.2006.08.003.
50. Daraganova, G. et al. (2012) Networks and geography: Modelling community network structures as the outcome of both spatial and network processes. Soc. Netw. 34, 6–17. doi: 10.1016/j.socnet.2010.12.001.
51. McCallum, A., Wang, X. & Corrada-Emmanuel, A. Topic and role discovery in social networks with experiments on Enron and academic email. *J. Artif. Intell. Res.* **30**, 249–272 (2007).
52. Zhou, D., Manavoglu, E., Li, J., Giles, C. L. & Zha, H. Probabilistic models for discovering e-communities. In *Proceedings of the 15th International Conference on World Wide Web*, 173–182 (2006).
53. Barthélemy, M. Spatial networks. Physics Reports 499, 1–101, 2011, doi: 10.1016/j.physrep.2010.11.002.
54. Alizadeh, M., Cioffi-Revilla, C. & Crooks, A. Generating and analyzing spatial social networks. Computational and Mathematical Organization Theory 23, 362–390, 2017, DOI: 10.1007/s10588-016-9232-2.
55. Krioukov, D., Papadopoulos, F., Kitsak, M., Vahdat, A. & Boguná, M. Hyperbolic geometry of complex networks. *Physical Review E* **82**, 036106 (2010).
56. Boguná, M., Papadopoulos, F. & Krioukov, D. Sustaining the internet with hyperbolic mapping. Nature communications 1, 1–8 (2010).
57. Serrano, M. A., Krioukov, D. & Boguná, M. Self-similarity of complex networks and hidden metric spaces. *Physical review letters* **100**, 078701 (2008).
58. Papadopoulos, F., Kitsak, M., Serrano, M., Boguná, M. & Krioukov, D. Popularity versus similarity in growing networks. Nature 489, 537–540 (2012).
59. Park, J. & Newman, M. E. J. Statistical mechanics of networks. Physical Review E 70, 66117, 2004, DOI: 10.1103/PhysRevE.70.066117.
60. Jaynes, E. Information theory and statistical mechanics. *The Physical Review* **106**, 181–218. https://doi.org/10.1103/PhysRev.106.620 (1957).
61. Garlaschelli, D. & Loffredo, M. I. Fitness-dependent topological properties of the world trade web. *Physical Review Letters* **93**, 188701. https://doi.org/10.1103/PhysRevLett.93.188701 (2004).
62. Squartini, T. & Garlaschelli, D. Analytical maximum-likelihood method to detect patterns in real networks. New J. Phys. 2011, doi: 10.1088/1367-2630/13/8/083001.
63. Straka, M. M. J., Caldarelli, G., Squartini, T. & Saracco, F. From ecology to finance (and back?): A review on entropy-based null models for the analysis of bipartite networks. Journal of Statistical Physics 173, 1252–1285, 2018, DOI: 10.1007/s10955-018-2039-4.
64. Squartini, T., Picciolo, F., Ruzzenenti, F. & Garlaschelli, D. Reciprocity of weighted networks. Sci. Rep. 2013, **3**, 1–9, https://doi.org/10.1038/srep02729 (2013).
65. Mastrandrea, R., Squartini, T., Fagiolo, G. & Garlaschelli, D. Enhanced reconstruction of weighted networks from strengths and degrees. New J. Phys. 2014, doi: 10.1088/1367-2630/16/4/043022.
66. Squartini, T., Mastrandrea, R. & Garlaschelli, D. Unbiased sampling of network ensembles. New J. Phys. 2015, doi: 10.1088/1367-2630/17/2/023052.
67. Saracco, F., Clemente, R. D., Gabrielli, A. & Squartini, T. Randomizing bipartite networks: the case of the world trade web. *Sci. Rep.* **5**, 10595. https://doi.org/10.1038/srep10595 (2015).
68. Di Gangi, D., Lillo, F. & Pirino, D. Assessing systemic risk due to fire sales spillover through maximum entropy network reconstruction. Journal of Economic Dynamics and Control 94, 117–141, 2018, doi: 10.1016/j.jedc.2018.07.001.
69. Squartini T, Caldarelli G, Cimini G, Gabrielli A, Garlaschelli D (2018) Reconstruction methods for networks: The case of economic and financial systems. Phys. Rep. 757, 1–47. DOI: 10.1016/j.physrep.2018.06.008.
70. Bardoscia, M. et al. The physics of financial networks. Nat. Rev. Phys. 2021, 3, 490–507. 2021, doi: 10.1038/s42254-021-00322-5.
71. Gandy, A. & Veraart, L. A. M. Adjustable network reconstruction with applications to cds exposures. *SSRN Electron. J.* https://doi.org/10.2139/SSRN.2895754 *(2017)*.
72. Ramadiah, A., Caccioli, F. & Fricke, D. Reconstructing and stress testing credit networks. SSRN Electronic Journal 2017, 10.2139/SSRN.3084543.
73. Anand, K. et al. The missing links: A global study on uncovering financial network structures from partial data. Journal of Financial Stability 35, 107–119, 2018, DOI: 10.1016/J.JFS.2017.05.012.

Acknowledgements

Author contributions

All authors designed the study. S.G. and F.S. developed the methods. A.C. and E.M. acquired the data. M.B., A.C, S.G and E.M. created the software. S.G. interpreted the experimental results and wrote most of the paper. All authors contributed to the writing and the revision of the article, and they all read and approved the final manuscript.

Competing interests

The authors declare no competing interests.

Additional information

Correspondence and requests for materials should be addressed to S.G.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.