A network model of freight data with spatial dependence

Aurélien Hazan

IMA Journal of Complex Networks (2020) Page 1 of 12
doi:10.1093/comnet/xxx000

A network model of freight data with spatial dependence

AURÉLIEN HAZAN*
Université Paris-Est Créteil, F-77567 Lieusaint, France
CAMS, EHESS, 54 bd Raspail, 75006 Paris, France
*Corresponding author: aurelien.hazan@u-pec.fr

[Received on 27 May 2020]

In this paper we analyze the origin-destination matrix arising from freight flows that occur in single-mode transport networks and compare unbiased maximum-entropy models of the corresponding networks. An original model based on earlier results allows to reconstruct a weighted network, from degree and strength sequences, taking distances into account. As an application, the properties of the European railroad freight are analyzed in detail in year 2010, with a focus on spatial effects.

Keywords: rail, freight, maximum entropy, null model, spatial network

2000 Math Subject Classification: 34K30, 35K57, 35Q80, 92D25

1. Introduction

In this article rail freight data aggregated at the regional scale are represented as a weighted directed network, following earlier works on airline networks [2], rail infrastructure and passenger traffic [15], cargo ships [14].

Publicly available origin-destination (OD) matrices by Eurostat summing the freight mass carried from some area i to another j over a year are represented as complex networks. In contrast with human mobility studies where precisely geocoded data are commonplace (social networks, phone, smart card), such data are aggregated at the regional level.

Furthermore we propose a maximum-entropy point of view, following trade studies where international trade networks have long been studied in that way.

In section 2 the network is characterized and compared to observations in the literature, and spatial effects are evidenced. In section 3 the properties of random network null models are compared to empirical ones. In section 4 we discuss the results and conclude. The computer code to reproduce all experiments is available[1].

2. European freight origin-destination networks

Transportation system have been actively studied as spatial networks in the complex networks community, both at the infrastructure, and traffic flow level.

While infrastructure network (rail, roads) are spatial planar networks where edges are not allowed to intersect, and for which specific results exist [6], airline [3] and cargo [14] were modeled as non-planar spatial networks. As shown in [15] railroad traffic flows may be uncorrelated to the topology of the physical infrastructure. Even though a complete picture of a transportation system requires the

[1] `https://gitlab.com/hazaa/mplex`

knowledge of both infrastructure and traffic[2], we will focus here on traffic flows.

The OD matrix $W = \{w_{ij}\}_{1 \leqslant i,j \leqslant N}$ with $w_{ij} \in \mathbb{N}$ is a possible starting point to analyze traffic flows. The geographical area of interest is divided into $i \in [1,N]$ zones. The elements in the OD matrix represent the count of transported units between zones i and j in a given time lapse. W defines a directed weighted network $G = (V,E)$, without loops, and a corresponding adjacency matrix $\mathbf{A} = \{a_{ij}\}_{1 \leqslant i,j \leqslant N}$ with $a_{ij} \in \{0,1\}$. OD matrix are found in epidemiology, human mobility, transport engineering. The route assignment of the transported units is unknown, as well as the traffic of elementary physical segments (for example between two rail stations).

OD matrix for European rail freight are publicly available from Eurostat under the label `tran_r_rago` and represent the annual freight mass in tonnes between and origin an a destination region with detail up to NUTS2 subdivision[3] with population ranging between 800000 and 3 millions. Cleaned and consistent data were prepared by the ETIS+ project[4] for years 2005 and 2010, and will be used below.

We then turn to the comparison between the expected and observed properties of G in a complex networks perspective. From previous studies in the field concerning non-planar spatial networks representing transport flow:

- the distribution $P(k_i)$ of node degree $k_i = \sum_{j \neq i} a_{ij}$ is expected to be peaked for infrastructure networks, to be power-law distributed (with a cutoff) in airline, and right-skewed for passenger train traffic, which reveals a heterogeneous topology with a few highly connected nodes [15].

- the distribution $P(s_i)$ of strength $s_i = \sum_{j \neq i} w_{ij}$ was found to be broad, for airlines and train traffic [15].

- a nonlinear dependence was reported between topology and strength. Empirically it was established [3] that s scales as $s \propto k^{\beta_w}$. Larger units are thus expected to exchange larger quantities, more than linearly with their size. This dependence was later modeled in [20] where the values $\beta_w = \beta_d = \frac{3}{2}$ were obtained analytically in a simple and robust setting, and compared to several empirical values.

- the effect of spatial constraint on topology can be measured by the distance strength $s_i^d = \sum_{j \neq i} a_{ij} d_{ij}$, with d_{ij} standing for the geodesic distance between regions i and j. A power-law dependence of the form $s^d \propto k^{\beta_d}$ was noticed. In the case of the North-American airline network, the value $\beta_d = 1.4$ was measured [3].

- a slightly disassortative behavior was reported in airline networks, and is prevalent in technological networks, as opposed to social networks. This is measured by the average nearest neighbor degree (ANND) $k_i^{nn} = \frac{\sum_{j \neq i} a_{ij} k_j}{k_i}$

Basic network measures for the undirected network associated to G are summarized in Tab. 1, and compared to non-planar spatial networks that represent well-studied transportation systems. As a consequence of coarse geographic resolution, G is smaller, denser and more clustered than other networks. The degree distribution $P(k)$ is represented in inset of Fig. 1(left). While peaked, it clearly differs from the planar case, that has a low cutoff value. It can be fit by a lognormal density, if k is approximated by

[2] because traffic flows do not indicate the most loaded and vulnerable segments.

[3] NUTS stands for Nomenclature of Territorial Units for Statistics. It is a Eurostat geocode standard, that often coincides with national administrative boundaries of EU countries.

[4] `https://ftp.demis.nl/outgoing/etisplus/datadeliverables/`

a continuous random variable. The empirical tail behavior is compared to an exponential, (associated to random graphs), a lognormal and a power-law distribution. The best fit is found with a lognormal distribution, as shown by the complementary cumulated density function (CCDF) $k \to P(K > k)$ in Fig. 1(left). Decreasing slower than an an exponential, this distribution is heavy-tailed, which can be explained by the presence of hub regions. Since it decreases faster than a power-law, it is not fat-tailed. These observations are confirmed by a likelihood ratio test, that doesn't reject the lognormal null hypothesis against a power-law. Thus the network is not scale-free, and a multiplicative generative mechanism may be searched for, rather than a preferential attachment one, as in scale-free networks. Similarly, the best fit for the distributions of node strengths $P(s)$ and edge weights $P(w)$ are lognormally distributed, which is confirmed by a likelihood ratio test. The scaling of strength s and distance strength s^d with respect to topology measured by the degree k is represented in Fig. 2. A superlinear scaling $s \propto k^{\beta_w}$ and $s^d \propto k^{\beta_d}$ is found with exponents β_w and β_d larger than 1. This behavior, that translates the influence of geographical constraints on G itself, is consistent with expectations as mentioned above. Larger regions do exchange larger quantities, with more distant regions. Whether or not the value of β_d constitutes a direct measurement of the magnitude of geographical constraints is unclear. In the growing network model [3] β_d varies, depending on this magnitude. On the opposite, in the fitness model [20], β_d is fixed with the value $\frac{3}{2}$. Furthermore, as remarked in [5], the scaling exponents were reported to depend on the level of spatial aggregation, which calls for further studies, that would control the effect of the level of aggregation.

Concerning directional effects, 31% of the links in the directed network are not reciprocated. This fact was noticed in [5], as characterizing distribution networks by contrast with transportation networks[5]. We didn't find imbalances between in-degree and out-degree, nor between directed average nearest-neighbor degree (see Appendix for definitions of directed quantities): in-degree is highly correlated to out-degree, in/in ANND is highly correlated to out/out ANND.

Notation	Name	Rail ETIS+	Sea GCSN	Air WAN
	Non-spatial measures			
	directed	yes	yes	no
	number of nodes	289	951	3880
	number of edges	7248	36351	18810
	density	0.17	0.04	0.0002
	clustering	0.67	0.49	
	reciprocity %	69	59	
β_w	scaling exponent $s \propto k^{\beta_w}$	1.26	1.46	1.5
$\langle k \rangle$	average degree	50.1	76.5	9.7
	Spatial measures			
β_d	scaling exponent $s^d \propto k^{\beta_d}$	1.25		

Table 1. Transport networks measures of non-planar spatial networks. GCSN: Global Cargo Ship Network [14]. WAN: world-wide airport network [2]. ETIS+ is converted to undirected network for comparison, except for reciprocity.

[5] "basically every individual performs a round trip implying symmetrical weights" in [5]

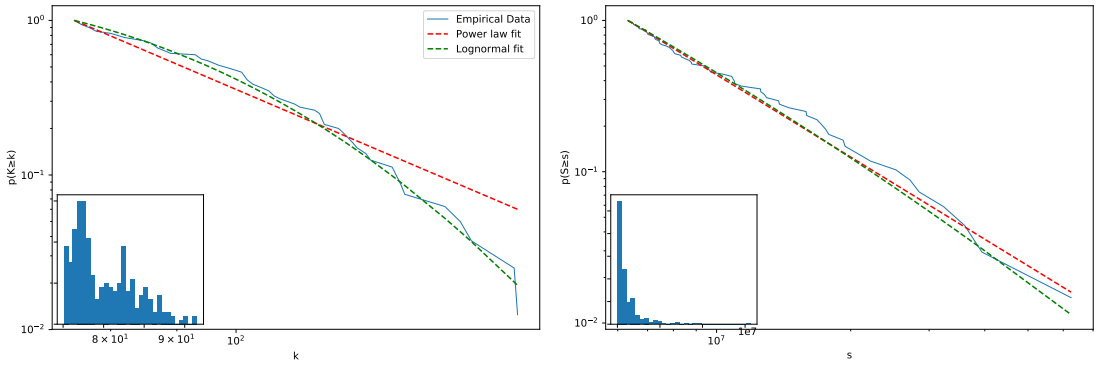

FIG. 1. Complementary CDF (left) degree k_i; (right) strength s_i. Histograms are in inset.

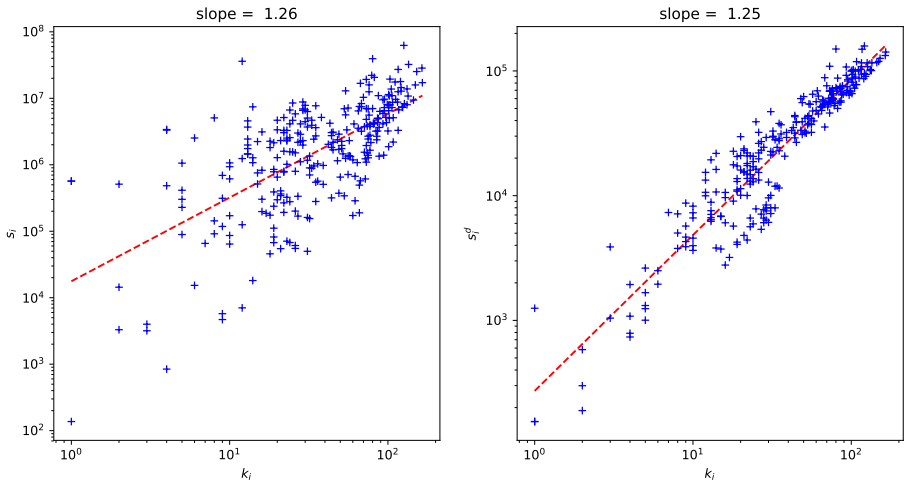

FIG. 2. Scaling of s_i and s_i^d with k_i.

The assortativity, measured by the ANND shown in Fig. 5(top, red dots) is decreasing with k, as expected. This differs from the flat behavior found in planar spatial networks. Furthermore an unexpected cluster is observed in the inferior left corner, which presence can be traced back to geographical constraints. Indeed a core/periphery structure is noticed in the choropleth in Fig.3, where color represent values of k_i and k_i^{nn} in each region. Top-ranking regions by degree are listed in Tab. 2, and are all located in central Europe or in northern Italy, except Lorraine. Peripheral regions (with respect to continental Europe) have a lower degree. Neighboring region have correlated degrees, despite local disparities.

Apart from their effect on G, the influence of geographical constraints has also been characterized in former studies by the distribution $P(d_{ij})$ where $D = \{d_{ij}\}_{1 \leqslant i,j \leqslant N}$ stands for the distance matrix. In [3], $P(d_{ij})$ was exponential which was explained by "the existence of physical and economical restrictions on airline planning in a continental setting". Fig. 4 represents the histogram in inset which can be fit

Region name	Country	Degree
Steiermark	AT	165
Oberosterreich	AT	164
Lombardia	IT	152
Niederosterreich	AT	149
Stredni Cechy	CZ	145
Moravskoslezsko	CZ	135
Lorraine	FR	134
Friuli-Venezia Giulia	IT	132
Veneto	IT	132
Emilia-Romagna	IT	127

Table 2. Top-ranking EU regions by degree. Data from ETIS+, rail freight in tonnes, year 2010.

by a lognormal distribution, and the tail behavior that decreases faster than the best fit lognormal. We found no explanation for this distribution in the literature. In [6], theoretical results on the distance distribution between the centers of cells in a Poisson-Voronoi tessellation are discussed and may be helpful if administrative boundaries of regions are considered as resulting from a tesselation, taking the regional capital as the cell center. Empirically, inter-city distance distribution have been studied (although, far less than human mobility distance distribution). Studies in economy address the relationship between city-size and spatial distribution of cities. Recent empirical works [13], in the case of the USA group cities in bins based on their population and analyze their distribution. In [9] the authors compare several intra-country between-city distributions (some of which are similar to results reported in Fig. 4) and propose a generative mechanism.

Lastly we remark that while working with networks built from OD matrices, measures based on paths (such as shortest path or betweenness centrality) have an unclear interpretation. They will not be considered in this article, as well as the small-world property.

In section 3, maximum entropy models of the spatial network G, able to reproduce local properties, are considered.

3. Maximum entropy null models

In section 2 the empirical properties of G were presented. In the present section a comparison is made with null models, built on a minimal set of assumptions. Differences between expectations based on the model and empirical measures may be informative concerning the nature of the observed phenomenon.

The model of [20] discussed in section 2 belongs to the family of *hidden variables models*, and offers a spatial and weighted generalization of *fitness models* [8]. In [20] the classical hypothesis $p_{ij} \propto x_i x_j$ is modified, and is true only if the product of fitnesses is greater that some distance-dependent cost $x_i x_j > c(d_{ij})$.

Maximum entropy models are part of the same family: in [18], the authors discuss the properties of exponential random graphs (ERG) derived from the maximum entropy methodology, that defines constraints in an average sense over the probability distribution of networks $P(\mathbf{W})$. In [11] a maximum-likelihood method to estimate hidden variables in binary networks from the degree sequence $\{k_i\}_{i\in[1,N]}$ is presented. In [7], the authors consider a spatial binary model that accounts for the distance bin that

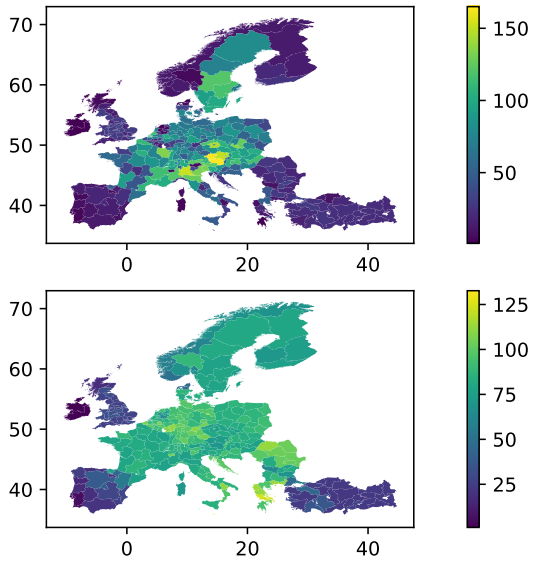

FIG. 3. Choropleth. Measures variables plotted per region (left) k_i ; (right) k_i^{nn}. Data from ETIS+, rail freight in tonnes, year 2010.

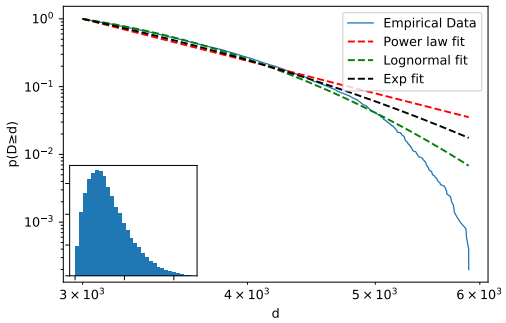

FIG. 4. CCDF of d_{ij}. Histogram in inset, linear coordinates.

a specific edge $e(i, j)$ associated to the distance d_{ij} falls into. The maximum likelihood method was later extended to weighted networks with constrained degree and strength sequences $\{k_i\}_{i\in[1,N]}$ and $\{s_i\}_{i\in[1,N]}$ [16], and to spatial binary and weighted networks where inter-node distances $\{d_{ij}\}_{i,j\in[1,N]}$ are fixed. In the case of the space-dependent International Trade Network (ITN), the authors in [1] state that "the topology of the ITN must have an immediate effect on the expected volume of trade between two countries" while "the expected topology of the ITN is independent of the expected volume of trade". This justifies our separate treatment of both aspects below. Furthermore it was shown that ERG models succeed in reproducing average higher-order properties such as the ANND, and that the function $k_i^{nn} = f(k_i)$ could be explained by the sole local constraints exerted on the degree sequence $\{k_i\}_{i\in[1,N]}$.

These models are classically compared to the gravity model (GM) that is suited for complete weighted graphs where all nodes are connected to each other, but not appropriate to recover the topology of a network, as a consequence of the hypothesis $w_{ij} \propto s_i s_j$. The GM was successfully used for example in cargo trade network by [14], however significant errors associated to non-existent links are evidenced. Other properties of the GM are further discussed in [5] and [4].

In the present article, two undirected weighted models that constrain topology and distance will be fit to the rail freight data. Firstly, the Enhanced Gravity Model (EGM) [1] is a state-of-the-art weighted network model, that circumvents the drawbacks of the GM. It derives from the Bose-Fermi distribution in [12], associated to the following Hamiltonian that constrains $\langle a_{ij} \rangle$ and $\langle w_{ij} \rangle$:

$$H(\mathbf{W}) = \sum_{i<j} \alpha_{ij} \Theta(w_{ij}) \beta_{ij} w_{ij} \tag{3.1}$$

where $\Theta(.)$ is the Heavyside function. The maximum-entropy probability $P(\mathbf{W})$ is:

$$P(\mathbf{W}) = \prod_{i<j} q_{ij}(w_{ij}) \tag{3.2}$$

The probability $q_{ij}(w)$ was shown in [12] to be:

$$q_{ij}(w) = \prod_{i<j} \frac{x_{ij}^{\Theta(w)} y_{ij}^{w} (1 - y_{ij})}{1 - y_{ij} + x_{ij} y_{ij}} \tag{3.3}$$

under the hypothesis that $w \in (O, \infty)$, where $\{x_{ij}\}_{1 \leqslant i,j \leqslant N}$ and $\{y_{ij}\}_{1 \leqslant i,j \leqslant N}$ are coefficients associated to the Lagrange multipliers α_{ij} and β_{ij}. So that x_{ij} and y_{ij} are related to the strengths s_i and distances d_{ij} we follow the method in [1], and choose the following functional forms. First, the edge probability are related to strengths s_i as in the Fitness-induced Configuration Model (FiCM) [10]:

$$p_{ij} = \frac{\delta s_i s_j}{1 + \delta s_i s_j} \tag{3.4}$$

More detail is given about this model in Appendix. Remark that the topology derived from eq.(3.4) depends on strength only, thus topological measures are expected not to depend on d_{ij}.

Then, the conditional average weight given that the edge $e(i, j)$ exists, is written:

$$\langle w_{ij} | a_{ij} = 1 \rangle = c (s_i s_j)^{\alpha} d_{ij}^{-\gamma} \tag{3.5}$$

Parameters $\delta, c, \alpha, \gamma$ are found solving a maximum-likelihood problem numerically. As regards topology, the algorithm is guaranteed to respect the empirical number of link L, but does not enforce the degree sequence as a constraint.

Secondly, we build a alternative model using two methods found in the maximum-entropy literature:

- the Degree-Corrected Gravity Model (DCGM) [22] allows one to reconstruct and sample a weighted network from the strength sequence s_i and the number of links L. Internally it also rests upon the FiCM to reconstruct the edge probability p_{ij} from the strength sequence s_i and the link density. We propose here to replace this probability by an expression that accounts for distances d_{ij}.
- the algorithm in [7], labeled DiBCM henceforth, is a distance-preserving variation of the classical binary configuration model in [18]. It assigns distance-dependent weights W_d to edges $e(i, j)$, with a functional form similar to:

$$p_{ij} = \frac{x_i x_j W_d(d_{ij})}{1 + x_i x_j W_d(d_{ij})} \tag{3.6}$$

The resulting algorithm, labeled DCGM+DiBCM, takes as inputs the degree and strength sequences, the distance matrix $D = \{d_{ij}\}_{1 \leqslant i,j \leqslant N}$, and returns sampled weight matrices $\mathbf{W}$ that are used in turn to compute averages such as $\langle w_{ij} \rangle$, $\langle k_i \rangle$, $\langle s_i \rangle$, $\langle s_i^d \rangle$ and $\langle k_i^{nn} \rangle$. Details about DCGM and DiBCM are available in the Appendix.

In Fig. 5 empirical measurements are compared to averages under the null models just presented. The $k_i^{nn} = f(k_i)$ scatter plot of experimental measures is approximated by a smooth function of k_i in the case of the EGM model, that corresponds well to the upper part of the points. By contrast, DiBCM average points are more evenly spread over the set of observed points. This is expected because p_{ij} in the DiBCM has one more degree of freedom since it depends on d_{ij}, unlike in the EGM case. For example, the low-k/low-k^{nn} zone is ignored in the EGM case, while it is better accommodated by the DiBCM. Both models find the expected disassortative trend, due to the local constraints of the degree sequence. For the sake of completeness, the inference was done with another model, the Configuration Model with Distances (DDCM) [19] but results were not found to be superior.

In the weighted case $s_i^{nn} = f(k_i)$ the DGCM+DiBCM again performs a better spread over the observed points, except in the low-s^{nn} zone and despite the correct reconstruction of k^{nn} seen above. This calls for a detailed analysis of the weight reconstruction in this zone. The tendency of the EGM to overestimate s_i^{nn} and s_i^d is even stronger than in the binary case.

The distance strength $s_i^d = f(k_i)$ is nicely reconstructed by the DGCM+DiBCM algorithm, except for the low-k /low-s^d zone in which EU regions have less rail trade neighbors, that are on average closer to each other, and that are less connected. This interesting discrepancy between the model and the measurement again justifies the quest for a good null model.

For example, a degree-dependent difference has been observed in [21] between observed and modeled s_i^d and explained by a tendency of poorly connected countries to trade more locally. However, we do not observe such degree-dependent difference here.

Finally, the good agreement obtained using DCGM+DiCM allows us to conclude that in first approximation higher order topological properties of G can be explained by $\{k_i\}_{i \in [1,N]}$ and D only, with no need for other explanatory mechanism. More work is needed however to capture faithfully the weight-dependent properties.

4. Discussion and conclusion

In this article, material flows between EU zones are studied as an origin-destination matrix represented by a directed network, and are proxied by European rail freight data at the regional NUTS2 level, prepared by Eurostat and the ETIS+ project.

The dataset is studied in a complex-network perspective, which had not been done previously to the best of our knowledge. Several empirical findings concerning the non-planar spatial weighted network are similar to national-level observations for trade data: $P(k)$ and $P(s)$ have a heavy subexponential tail, but are not fat-tailed.

Geographical effects are evidenced on the region-center distance distribution, and on the network itself: s and s^d scale in a superlinear way with k, with exponents β_w, β_d, larger than 1 and close the value $\frac{3}{2}$ predicted in [20]. Moreover, an interesting core/periphery structure appears, with respect to node degree. Even though directional effects are observed, the input and output degree sequences are highly correlated.

Several maximum entropy null models are fit to available data in order to check if the network's characteristics can be explained only by degree sequence, the strength sequence and the inter-distance matrix, or if other mechanisms must be looked for. The Enhanced Gravity Model is used as a bench-

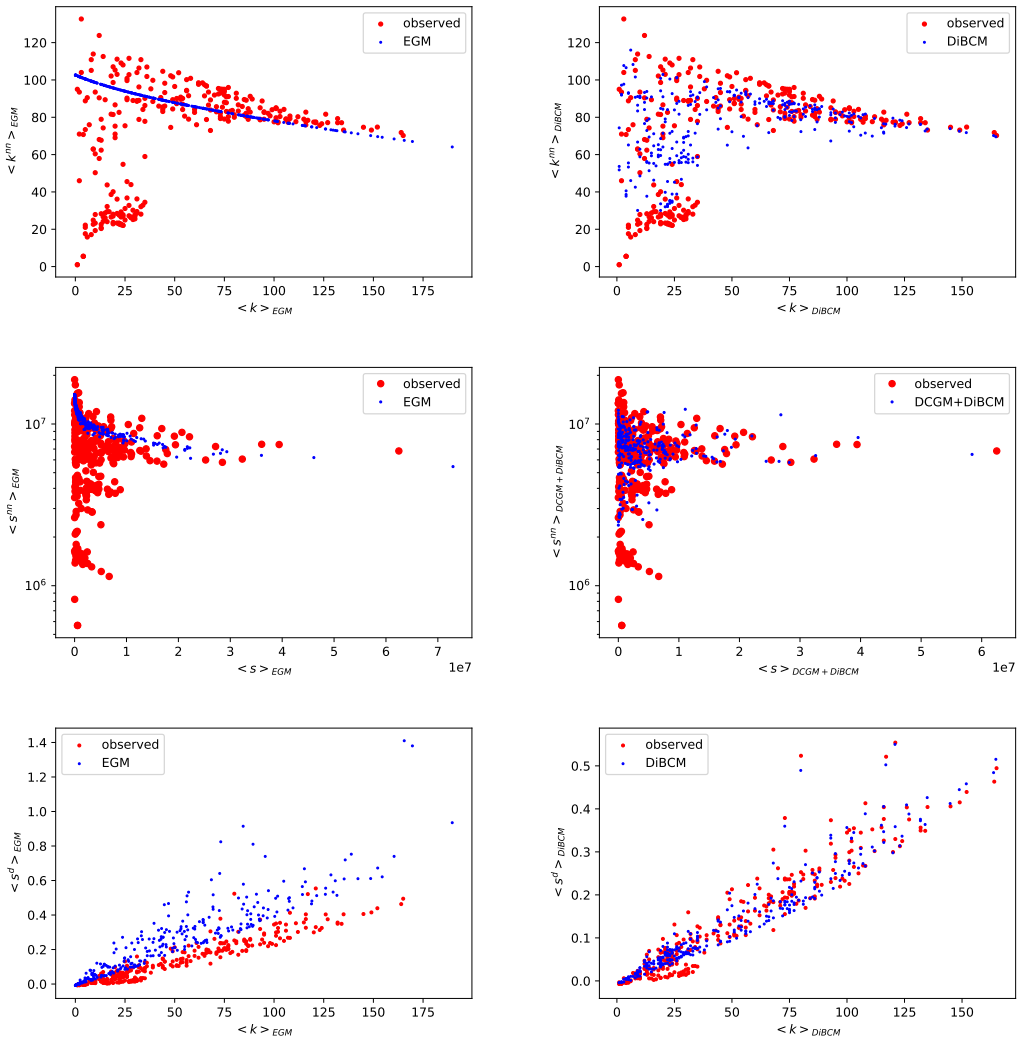

FIG. 5. Maximum entropy models. (left) EGM model; (right) DCGM+DiBCM model; (top) $k_i^{nn} = f(k_i)$; (middle) $s_i^{nn} = f(k_i)$; (bottom) $s_i^{d} = f(k_i)$

mark, and compared to a custom model labeled DCGM+DiCM that mixes the Degree-Corrected Gravity Model and a distance-weighted configuration model. This strategy allows to separate the problem-specific topology reconstruction (that depends on distance) from the weight reconstruction. DCGM+DiCM allows a better agreement with data than EGM for higher-order measures of topological structure and for distance strength, while weighted measures need to be improved.

Future works first need to address the problem of coarse geographic resolution, for example taking advantage of the literature on network renormalization, in order to quantify the spatial effect in an accurate manner, and to compare to existing fitness of growth models. Furthermore, maximum entropy benchmark models will include recent propositions in the field to better reflect the weight structure, and the distance dependency.

Lastly, results from related fields that currently tend to increase their geographical level of detail (Material Flow Accounting, Multi-Regional Input-Output, Life-cycle Inventory and Carbon accounting) could be compared to the picture given above.

A. Directed ANND

The definitions of directed average nearest-neighbor degree ANND are reproduced from [24, §3.1.2] below:

$$\begin{aligned} k_i^{nn,in/in} &= \frac{\sum_{j\neq i} a_{ji} k_j^{in}}{k_i^{in}} \\ k_i^{nn,out/out} &= \frac{\sum_{j\neq i} a_{ij} k_j^{out}}{k_i^{out}} \end{aligned}$$

$k_i^{nn,out/out}$ measures the correlation between the out-degree of node i and the out-degree of nodes that i is pointing to. More definitions are available from the same reference.

B. FiCM

In the Fitness-Induced Configuration Model (FiCM), similarly to fitness (or hidden-variables) models [8], the topology of the network derives from intrinsic properties of the nodes.

Unlike in the case of the Configuration Model (CM), no information about node degrees is available. Fitnesses are known a priori from empirical unnormalized measures, for example the GDP of countries: $x_i = \frac{GDP_i}{\sum_i GDP_i}$.

In the undirected case the connection probabilities are thus determined by the equation:

$$p_{ij} = \frac{\delta x_i x_j}{1 + \delta x_i x_j} \tag{A.1}$$

which can be compared to the expression $p_{ij} = \frac{x_i x_j}{1+x_i x_j}$ for binary network under the configuration model in [17].

There is only one free parameter: δ. As explained in [10], only the empirical network density is necessary to solve for δ. The probability of the graph associated to the incidence matrix $\mathbf{A}$ is then [23]:

$$P(\mathbf{A}) = \prod_{i<j} p_{ij}^{a_{ij}} (1 - p_{ij})^{1-a_{ij}} \tag{A.2}$$

C. DCGM

The Degree-Corrected Gravity Model in [22] is aiming at reconstructing both the topological and the weights structure of a directed network, from strength sequences $\{s_i^{in}\}_{i\in[1,N]}$ and $\{s_i^{out}\}_{i\in[1,N]}$ and the number of links only.

As explained in the main text, reconstructing the topology is done using the FiCM. Then, the weights are found using the formula:

$$w_{ij} = \begin{cases} 0 & \text{with probability } 1-p_{ij}, \\ \frac{s_i^{out}s_j^{in}}{Wp_{ij}} & \text{with probability } p_{ij} \end{cases} \tag{A.1}$$

with W the sum of weights. This recovers the GM specification $\langle w_{ij}\rangle = \frac{s_i^{out}s_j^{in}}{W}$. The average strength sequence does not correspond exactly to the input constraint $\{s_i^{in}\}$, $\{s_i^{out}\}$, and an additionnal correction term is necessary. The interested reader will find the details in [22].

D. DiBCM

The distance-based model in [7], that we labeled DiBCM for distance-preserving binary configuration model in this article, is a variation over the classical configuration model.

The degree sequence $\{k_i\}_{i\in[1,N]}$ and the distance structure $D = \{d_{ij}\}_{1\leqslant i,j\leqslant N}$ constrain the Lagrange multipliers and thus the values of hidden variables x_i below.

The distances in D are binned in intervals $I_l = (d_{l-1}, d_l), l \in [1, L]$. $\chi_l(.)$ is the indicating function for interval I_l.

The connection probabilities are written:

$$p_{ij} = \frac{x_i x_j \sum_l \chi_l(d_{ij}) W(d_l)}{1 + x_i x_j \sum_l \chi_l(d_{ij}) W(d_l)} \tag{A.1}$$

The weights $\{W(d_l)\}_{l\in[1,L]}$ are distances-related hidden variables. All hidden variables are numerically approximated.

REFERENCES

1. Almog, A., Bird, R. & Garlaschelli, D. (2019) Enhanced Gravity Model of Trade: Reconciling Macroeconomic and Network Models. *Frontiers in Physics*, **7**, 55.
2. Barrat, A., Barthelemy, M., Pastor-Satorras, R. & Vespignani, A. (2004) The architecture of complex weighted networks. *Proceedings of the National Academy of Sciences*, **101**(11), 3747–3752.
3. Barrat, A., Barthélemy, M. & Vespignani, A. (2005) The effects of spatial constraints on the evolution of weighted complex networks. *Journal of Statistical Mechanics: Theory and Experiment*, **2005**(05), P05003.
4. Barthelemy, M. (2016) *The structure and dynamics of cities: urban data analysis and theoretical modeling*. Cambridge University Press, Cambridge, UK.
5. Barthélemy, M. (2011) Spatial networks. *Physics Reports*, **499**(1-3), 1–101.
6. Barthélemy, M. (2017) *Morphogenesis of spatial networks*. Springer Berlin Heidelberg, New York, NY.
7. Bianconi, G., Pin, P. & Marsili, M. (2009) Assessing the relevance of node features for network structure. *Proceedings of the National Academy of Sciences*, **106**(28), 11433–11438.
8. Caldarelli, G., Capocci, A., De Los Rios, P. & Muñoz, M. A. (2002) Scale-Free Networks from Varying Vertex Intrinsic Fitness. *Physical Review Letters*, **89**(25).
9. Fernández-Rosales, I. Y., Angulo-Brown, F., Pérez-Campuzano, E. & Guzmán-Vargas, L. (2020) Distance distributions of human settlements. *Chaos, Solitons & Fractals*, **136**, 109808.

10. Garlaschelli, D. & Loffredo, M. I. (2004) Fitness-Dependent Topological Properties of the World Trade Web. *Physical Review Letters*, **93**(18).
11. Garlaschelli, D. & Loffredo, M. I. (2008) Maximum likelihood: Extracting unbiased information from complex networks. *Physical Review E*, **78**(1), 015101.
12. Garlaschelli, D. & Loffredo, M. I. (2009) Generalized Bose-Fermi Statistics and Structural Correlations in Weighted Networks. *Physical Review Letters*, **102**(3), 038701.
13. González-Val, R. (2019) The spatial distribution of US cities. *Cities*, **91**, 157–164.
14. Kaluza, P., Kölzsch, A., Gastner, M. T. & Blasius, B. (2010) The complex network of global cargo ship movements. *Journal of The Royal Society Interface*, **7**(48), 1093–1103.
15. Kurant, M. & Thiran, P. (2006) Extraction and analysis of traffic and topologies of transportation networks. *Physical Review E*, **74**(3), 036114.
16. Mastrandrea, R., Squartini, T., Fagiolo, G. & Garlaschelli, D. (2014) Enhanced reconstruction of weighted networks from strengths and degrees. *New Journal of Physics*, **16**(4), 043022.
17. Park, J. & Newman, M. E. J. (2003) Origin of degree correlations in the Internet and other networks. *Physical Review E*, **68**(2), 026112.
18. Park, J. & Newman, M. E. J. (2004) Statistical mechanics of networks. *Physical Review E*, **70**(6).
19. Picciolo, F., Squartini, T., Ruzzenenti, F., Basosi, R. & Garlaschelli, D. (2012) The Role of Distances in the World Trade Web. In *2012 Eighth International Conference on Signal Image Technology and Internet Based Systems*, pages 784–792, Naples. IEEE.
20. Popović, M., Štefančić, H. & Zlatić, V. (2012) Geometric Origin of Scaling in Large Traffic Networks. *Physical Review Letters*, **109**(20), 208701.
21. Ruzzenenti, F., Picciolo, F., Basosi, R. & Garlaschelli, D. (2012) Spatial effects in real networks: Measures, null models, and applications. *Physical Review E*, **86**(6), 066110.
22. Squartini, T., Cimini, G., Gabrielli, A. & Garlaschelli, D. (2017) Network reconstruction via density sampling. *Applied Network Science*, **2**(1), 3.
23. Squartini, T. & Garlaschelli, D. (2011) Analytical maximum-likelihood method to detect patterns in real networks. *New Journal of Physics*, **13**(8), 083001.
24. Squartini, T. & Garlaschelli, D. (2017) *Maximum-Entropy Networks*. SpringerBriefs in Complexity. Springer International Publishing, Cham.