Gravity models of networks: Integrating maximum-entropy and econometric approaches

Marzio Di Vece [1,*] Diego Garlaschelli [1,2] and Tiziano Squartini [1,3]

[1]*IMT School for Advanced Studies Lucca, P.zza San Francesco 19, 55100 Lucca, Italy*
[2]*Lorentz Institute for Theoretical Physics, Leiden University, Niels Bohrweg 2, 2333 CA Leiden, The Netherlands*
[3]*Institute for Advanced Study (IAS), University of Amsterdam, Oude Turfmarkt 145, 1012 GC Amsterdam, The Netherlands*

The World Trade Web (WTW) is the network of international trade relationships among world countries. Characterizing both the local link weights (observed trade volumes) and the global network structure (large-scale topology) of the WTW via a single model is still an open issue. While the traditional Gravity Model (GM) successfully replicates the observed trade volumes by employing macroeconomic properties such as GDP and geographic distance, it unfortunately predicts a fully connected network, returning a completely unrealistic topology of the WTW. To overcome this problem, two different classes of models have been introduced in econometrics and statistical physics. Econometric approaches interpret the traditional GM as the expected value of a probability distribution that can be chosen largely arbitrarily and tested against alternative distributions. Statistical physics approaches construct maximum-entropy probability distributions of (weighted) graphs from a chosen set of measurable, structural constraints and test distributions resulting from different constraints. Here we compare and integrate the two approaches by considering a class of maximum-entropy models that can incorporate macroeconomic properties used in standard econometric models. We find that the integrated approach achieves an overall better performance than the purely econometric one. These results suggest that the maximum-entropy construction can serve as a viable econometric framework wherein extensive and intensive margins can be separately controlled for, by combining topological constraints and dyadic macroeconomic variables.

I. INTRODUCTION

The World Trade Web (WTW), formed by the import/export relationships between world countries, has received a lot of attention both in traditional economic studies and in more recent investigations. Indeed, as recent crises have clearly pointed out, "network effects matter" [1]: Trade linkages are, in fact, the most important channels of interaction between world countries, directly transmitting economic shocks [2–4] or indirectly transmitting financial ones [5–9]). It therefore comes with no surprise that many contributions have focused on the analysis of the *network properties* of the WTW. Indeed, the (ever-increasing) data availability over the last years has motivated researchers (coming from disciplines as different as economics, network science, and social sciences) to explore the architecture of the WTW from an empirical point of view, complementing the traditional theoretical knowledge in classical trade economics.

These efforts have produced a large wealth of literature characterizing various stylized facts of the WTW architecture [1,5,10–19]. A first stream of contributions has focused on the purely *binary* properties of the WTW, i.e., the structural features that require the knowledge of only the presence of connections (trade relationships) between countries, irrespective of their intensity. These studies have highlighted a large (compared with most of the other real-world networks) density, a right-skewed and heavy-tailed degree distribution, disassortative mixing by degree (indicating that countries having many trade partners are, on average, connected with countries having few partners), a hierarchical organization (indicating that partners of well-connected countries are less interconnected than those of poorly-connected ones), the presence of a core of countries trading with almost everyone else, and finally the presence of (a sort of) bow-tie structure [14,15,17,19,20]. A second stream of contributions has focused on the weighted properties of the WTW, where the *weight* attached to each link represents the volume of the corresponding trade relationship and the *strength* (sum of all link weights) of each node represents the overall trade volume of the corresponding country. These studies have highlighted right-skewed and heavy-tailed (generally log-normal) weight and strength distributions (indicating that few intense trade connections coexist with a majority of low-intensity ones), disassortative mixing by strength (indicating that countries whose trade volume is large are, on average, connected with countries whose trade volume is small), and a large weighted clustering coefficient (indicating that the trade volume of the partners of well-connected countries is larger

*marzio.divece@imtlucca.it

than the trade volume of the partners of poorly connected ones) [18].

From the modeling side, two broad classes of models of the international trade system can be roughly identified: *econometric* models and *network* models (the latter mainly rooted into statistical physics). Both can be traced back to the earliest model of international trade, proposed in 1962 by the physics-educated Dutch economist Jan Tinbergen (considered the founding father of econometrics together with Ragnar Frisch). Tinbergen modelled the import/export flows between countries via a relationship that is formally analogous to the law of gravity, where the "masses" of countries are replaced by GDPs and the intercountry distances are replaced by suitably defined geographic distances [21]. This is the celebrated gravity model (GM) and can be shown to reproduce the positive trade volumes quite accurately [22–25]. Although our focus here is on international trade, it is worth mentioning that the GM has been used successfully to describe the positive link weights associated to processes relevant to many other networks as well, including migration flows [26,27] (which actually represent the earliest [26] application of the GM), mobility and traffic patterns [28–30], communication streams [31], and spreading phenomena [32,33]. Despite its success, the most evident limitation of the GM consists in predicting that each country establishes a trading relationship with every other country, a result that is in stark contrast with empirical data, where "missing" trade relationships are actually found to be a significant proportion (up to one half) of the number of all possible pairs of countries. Since systematically overestimating the number of connections is known to lead to a significant miscalculation of network effects [1], one should avoid the use of the pure GM as a reliable network model of the WTW [23].

From the sixties on, a lot of work has been done in econometrics to overcome the aforementioned limitation. Initially, Eaton and Tamura [34] and Martin and Pham [35] suggested to employ a tobit-like estimation procedure, by rounding to zero the trade flows below a certain threshold. Later, the serious conceptual problem posed by the arbitrariness of the chosen threshold motivated Helpman, Melitz, and Rubinstein [36] to propose a two-step estimation procedure: First, a probit model is employed to estimate the probability of observing a trade relationship between any two nodes; subsequently, the corresponding trade flow is estimated via an ordinary-least-squares (OLS) regression whose parameters are tuned on the entire set of positive weights. An alternative algorithm is the one proposed by Silva and Tenreyro [37], who estimated the gravity equation in a multiplicative—rather than additive—fashion by employing a Poisson pseudo-maximum likelihood (PPML) method and obtained good estimates of trade flows, robust to heteroskedasticity. In the following years, PPML has been proven to be very sensitive to the "excess zeros" of the dependent variables. Hence, a different class of two-step estimation procedures has been devised, i.e., the so-called "zero-inflated" (ZI) methods [38,39]: These models are defined by a logit estimation, aimed at establishing the presence of a link, followed by either a Poisson (ZIP) or a negative binomial (ZINB) regression, aimed at estimating the corresponding trade volume. Later in 2011, Dueñas and Fagiolo [40] proved that the weighted structure of the WTW is well

predicted by GMs if and only if its topological structure is specified as well. This result marks an important shift from a perspective whose only focus use to be the prediction of trade flows to a perspective where the topological structure of the network becomes one of the reconstruction targets.

Coming to the physics-inspired approaches to international trade, the majority of proposed models are maximum-entropy models of networks [23,41–46]. The maximum-entropy (ME) framework [41,43,47–51] allows for purely structural network properties to be constrained in order to derive the maximally unbiased probability distribution compatible with those properties. A series of publications over the years [14,23,44,45, 52–54] has showed that fixing the degree sequence of world countries is enough to successfully reproduce higher-order binary properties such as the average nearest-neighbour degree and the clustering coefficient. By contrast, this result is no longer retrieved when the analysis is reformulated within a purely weighted framework: for instance, it turns out that the strength sequence is not an effective constraint in reproducing the average nearest-neighbour strength and the weighted clustering coefficient [41,55]. In fact, accuracy in the reconstruction of weighted properties is recovered only if the degrees and the strengths are constrained jointly [56–58].

Econometric and network approaches have mostly proceeded along parallel tracks, with little interaction so far. The only line of research that has systematically looked for the possibility of an integration between the two approaches is the one reformulating maximum-entropy network ensembles as *hidden variable* (or *fitness* [59]) models [14,15,23,60,61]. In this approach, the maximum-entropy construction of the network ensemble is formally preserved, but the Lagrange multipliers usually viewed as free parameters tuned in order to enforce the chosen constraints are instead identified with empirical macroeconomic factors, most notably the GDP of countries. Building on those results, as well as on the success of physics-inspired models in reconstructing various kinds of socioeconomic and financial systems [52,62], here we explore the potential of the maximum-entropy formalism to provide a viable framework for econometrics too.

The rest of the paper is organized as follows. Section II reviews the traditional GM and its performance in reproducing the positive WTW weights. Section III is devoted to the description of purely econometric models and their application to the analysis of trade (networks). Section IV is devoted to the description of maximum-entropy network models and their application to the WTW. Section V illustrates the main results of the paper, i.e., the integration of purely econometric and physics-inspired models. Section VI concludes and presents an outlook on possible future extensions.

II. MODELLING POSITIVE WEIGHTS

From a merely econometric point of view, the simplest exercise is that of reproducing the realized (positive) trade volumes (link weights) of the WTW. To this aim, we have considered two different datasets. The first one is curated by Gleditsch [63] and includes yearly trade volumes, yearly GDP values (both reported in millions of US dollars) and the (time-independent) matrix of geographic distances between capital cities of all countries in the data. The second one
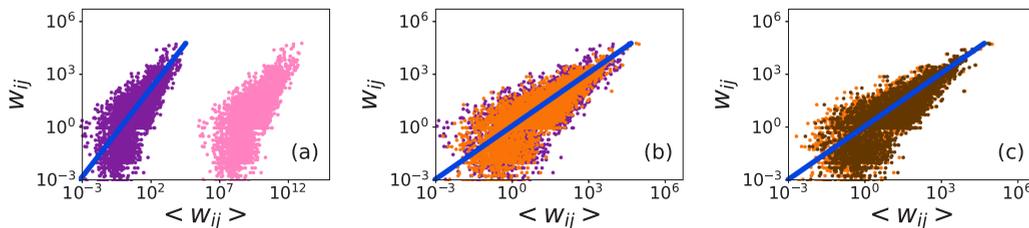
FIG. 1. Scatter plot of the entire set of positive WTW weights w_{ij} vs the values predicted by different specifications of the GM: (1) GM with $\rho = \beta = 1$ and $\gamma = 0$, depicted in •; (2) GM with $\beta = 1$, $\gamma = 0$ and ρ tuned by requiring that the total weight is reproduced, i.e., $W = \sum_{i<j} w_{ij} = \sum_{i<j} \langle w_{ij} \rangle_{\text{GM}} = \langle W \rangle_{\text{GM}}$, depicted in •; (3) GM with $\beta = 1$, $\gamma = -1$ and ρ tuned as above, depicted in •; (4) full GM where all parameters are tuned as described in Appendix A, depicted in •. While moving from a parameter-free model to a single-parameter model largely improves the goodness-of-fit, see (a), the net effect of rising the number of explanatory variables is that of decreasing the dispersion around the identity, see (b). Further rising the number parameters, instead, does not add much to the picture provided by the third specification of the GM, see (c). Results refer to the year 2000 of the dataset curated by Gleditsch [63].

is the BACI dataset, a detailed description of which can be found in [64,65]. For both datasets we have selected and analyzed eleven years: 1990–2000 for the Gleditsch dataset and 2007–2017 for the BACI one. The year 2000 of the Gleditsch dataset—i.e., the one with the largest number of countries (176)—is the snapshot we have selected to graphically illustrate all the results of our analyses. We have always considered the undirected (symmetrized) version of the weighted trade matrix, whose generic entry reads $w_{ij} = \frac{\exp_{ij} + \exp_{ji}}{2}$, i.e., w_{ij} is the bilateral trade volume defined as the arithmetic mean of the export volume from country i to country j and of the export volume from country j to country i.

Our econometric exercise is carried out by comparing the empirical, positive weights of the WTW, in the year 2000, with four different specifications of the following econometric function:

$$\langle w_{ij} \rangle_{\text{GM}} = f(\omega_i, \omega_j, d_{ij} | \underline{\theta}) = \rho(\omega_i \omega_j)^\beta d_{ij}^\gamma \tag{1}$$

where $\omega_i = \frac{\text{GDP}_i}{\overline{\text{GDP}}}$ is the GDP of country i divided by the arithmetic mean of GDPs, d_{ij} is the geographic distance between the capitals of countries i and j and $\underline{\theta} = (\rho, \beta, \gamma)$. The first specification of the model above is characterized by the assignment $\rho = \beta = 1$ and $\gamma = 0$ and its performance in reproducing the positive weights of the WTW is shown in Fig. 1(a). The overall poor performance of this version of the GM signals the presence of both a *scaling problem*—estimates are positively correlated with observations, only shifted to the right—and of a *dimensional problem*—in fact, pure numbers appear on the right-hand side while the WTW weights are measured in (multiples of) dollars.

A second specification of the GM solves both problems at once. This specification is characterized by the assignment $\beta = 1$ and $\gamma = 0$ while ρ is treated as a free parameter, tuned by requiring that the total weight is reproduced, i.e., that $W = \sum_{i<j} w_{ij} = \sum_{i<j} \langle w_{ij} \rangle_{\text{GM}} = \langle W \rangle_{\text{GM}}$: the fit is, now, much more accurate, as shown in Fig. 1(b). The model can be further enriched by adding dyadic factors such as the geographic distances between capitals: some more accuracy in the description of the empirical data is indeed gained, as the reduced dispersion of the cloud of points around the identity witness. Quite remarkably, the picture does not change much if, now, we let the entire set of parameters to be tuned as described in Appendix A [see Fig. 1(c)].

Although the GM leads to a good prediction of the positive weights, its intrinsic limitation is that it does not allow the topological structure of the WTW to be correctly recovered. In fact, by outputting only positive weights, it induces a trivial, fully-connected structure: upon defining $a_{ij} = \Theta[\langle w_{ij} \rangle_{\text{GM}}]$, $\forall i < j$, it is evident that $a_{ij} = 1$, $\forall i < j$. In order to overcome such a limitation, one can "refine" the plain gravity model by "dressing" it with a probability distribution capable of accounting for the null entries as well.

In very general terms, we need to define a *statistical network model*, i.e., a set of mathematical relationships between the random variables that are of interest for our network description. Two broad classes of such models can be identified, i.e., the *econometric* ones and the ones rooted into *statistical physics*. In what follows, we will deal with (either econometric or physics-rooted) *discrete* statistical models.

III. STATISTICAL NETWORK MODELS - I: ECONOMETRIC MODELS

Let us start with the description of some of the most representative members of the econometric class of models.

A. Poisson model

The simplest model in this class prescribes to consider $\langle w_{ij} \rangle_{\text{GM}} = \rho(\omega_i \omega_j)^\beta d_{ij}^\gamma$ as the expected value of a Poisson probability mass function:

$$q_{ij}^{\text{Pois}}(w_{ij}) = \frac{z_{ij}^{w_{ij}} e^{-z_{ij}}}{w_{ij}!}; \tag{2}$$

since $\langle w_{ij} \rangle_{\text{Pois}} = z_{ij}$, the explanatory power of the GM is retained upon requiring that $\langle w_{ij} \rangle_{\text{Pois}} = \langle w_{ij} \rangle_{\text{GM}}$, i.e., by posing

$$z_{ij} = \rho(\omega_i \omega_j)^\beta d_{ij}^\gamma. \tag{3}$$

The expected topology of the network is determined by the expected adjacency matrix entries implied by the model, which is captured by the expression $\langle a_{ij} \rangle_{\text{Pois}} = p_{ij}^{\text{Pois}} = 1 - q_{ij}^{\text{Pois}}(0) = 1 - e^{-z_{ij}}$ (see Appendix B for a detailed description of the procedure to estimate the parameters of the Poisson model).

B. Negative binomial model

The main drawback of the Poisson model is that of predicting a variance of the weights that is necessarily equal to their average value. In general, this may be different from what empirical analyses suggest. In order to overcome this problem, econometricians have considered a different probability mass function, namely the negative binomial one with the introduction of an overdispersion [66] parameter $\alpha = m^{-1}$ [67]:

$$q_{ij}^{\mathrm{NB}}(w_{ij}) = \binom{m + w_{ij} - 1}{w_{ij}} \left(\frac{1}{1 + \alpha z_{ij}}\right)^{m} \left(\frac{\alpha z_{ij}}{1 + \alpha z_{ij}}\right)^{w_{ij}}. \tag{4}$$

One finds that $\langle w_{ij} \rangle_{\mathrm{NB}} = m\alpha z_{ij} = z_{ij}$. The requirement that the expected value of the negative binomial distribution coincides with the prediction coming from the GM, i.e., $\langle w_{ij} \rangle_{\mathrm{NB}} = \langle w_{ij} \rangle_{\mathrm{GM}}$, can be, again, realized by posing $z_{ij} = \rho(\omega_i \omega_j)^{\beta} d_{ij}^{\gamma}$. Predictions about the topology are, now, carried out via the expression $\langle a_{ij} \rangle_{\mathrm{NB}} = p_{ij}^{\mathrm{NB}} = 1 - q_{ij}^{\mathrm{NB}}(0) = 1 - (\frac{1}{1 + \alpha z_{ij}})^m$ (see Appendix B for a detailed description of the procedure to estimate the parameters of the negative binomial model).

C. Zero-inflated Poisson model

The main drawback of the econometric models above is that of failing in reproducing the link density of the WTW. For instance, the latter equals $c = \frac{2L}{N(N-1)} \simeq 0.63$ in year 2000. It turns out that, while the Poisson model overestimates this quantity, the negative binomial one underestimates it, i.e.,

$$\langle c \rangle_{\mathrm{NB}} < c < \langle c \rangle_{\mathrm{Pois}} \tag{5}$$

where $\langle c \rangle_{\mathrm{Pois}} = \sum_{i<j} p_{ij}^{\mathrm{Pois}} \simeq 0.68$ and $\langle c \rangle_{\mathrm{NB}} \sum_{i<j} p_{ij}^{\mathrm{NB}} \simeq 0.60$. For this reason, econometricians have defined the so-called *zero-inflated* (ZI) models, i.e., two-step recipes whose general form reads

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

$$= \prod_{i<j} p_{ij}^{a_{ij}} (1 - p_{ij})^{1 - a_{ij}} \cdot q_{ij}(w_{ij}|a_{ij})$$

$$= P(\mathbf{A}) Q(\mathbf{W}|\mathbf{A}), \tag{6}$$

a relationship indicating that the probability of the (network represented by the) weighted adjacency matrix $\mathbf{W}$ can be obtained as the product of the probability $P(\mathbf{A})$ of observing the purely binary adjacency matrix $\mathbf{A}$ and the conditional probability $Q(\mathbf{W}|\mathbf{A})$, where, for consistency, $\mathbf{A} = \Theta[\mathbf{W}]$, i.e., $a_{ij} = \Theta[w_{ij}], \forall i < j$, the position $a_{ij} = \Theta[w_{ij}]$ meaning that $a_{ij} = 1$ whenever $w_{ij} > 0$ and $a_{ij} = 0$ if and only if $w_{ij} = 0$.

The simplest ZI model is the Poisson one, defined by the positions

$$p_{ij}^{\mathrm{ZIP}} = \frac{G_{ij}}{1 + G_{ij}} (1 - e^{-z_{ij}}), \tag{7}$$

$$q_{ij}^{\mathrm{ZIP}}(w_{ij}|a_{ij} = 1) = \begin{cases} \frac{z_{ij}^{w_{ij}} e^{-z_{ij}}}{(1 - e^{-z_{ij}}) w_{ij}!}, & w_{ij} > 0 \\ 0, & w_{ij} \leqslant 0 \end{cases}. \tag{8}$$

Notice that $1 - p_{ij}^{\mathrm{ZIP}} = \frac{1}{1 + G_{ij}} + \frac{G_{ij}}{1 + G_{ij}} e^{-z_{ij}}$, i.e., the connection between nodes i and j can be missing either because a link is not there (with probability $\frac{1}{1 + G_{ij}}$) or because a link is there but has zero weight (with probability $\frac{G_{ij}}{1 + G_{ij}} e^{-z_{ij}}$). Consistently, i and j are connected because the weight is not zero (with probability $1 - e^{-z_{ij}}$). In order to "dress" the GM, we need to identify some of the parameters of the Poisson model with the usual econometric function. Since

$$\langle w_{ij} \rangle_{\mathrm{ZIP}} = p_{ij}^{\mathrm{ZIP}} \langle w_{ij} | a_{ij} \rangle_{\mathrm{ZIP}}$$

$$= p_{ij}^{\mathrm{ZIP}} \frac{z_{ij}}{1 - e^{-z_{ij}}} = \frac{G_{ij}}{1 + G_{ij}} z_{ij}, \tag{9}$$

we can make the identification $z_{ij} = \rho(\omega_i \omega_j)^{\beta} d_{ij}^{\gamma}$. Upon doing so, we are treating z_{ij} as an "effective" conditional weight: In fact, Eq. (9) can be understood as describing an aleatory experiment that combines a logit with a full Poisson step. According to this interpretation, z_{ij} would represent a Poisson-like expected weight, conditional to the success of the logit step, i.e., $z_{ij} = \frac{\langle w_{ij} \rangle_{\mathrm{ZIP}}}{p_{ij}^{\mathrm{logit}}}$, with $p_{ij}^{\mathrm{logit}} = \frac{G_{ij}}{1 + G_{ij}}$.

A second econometric identification is, however, needed: We will proceed by imposing

$$G_{ij} = \delta \omega_i \omega_j \tag{10}$$

(see Appendix B for a detailed description of the procedure to estimate the parameters of the ZIP model).

D. Zero-inflated negative binomial model

The ZI version of the negative binomial model, instead, is defined by

$$p_{ij}^{\mathrm{ZINB}} = \frac{G_{ij}}{1 + G_{ij}} (1 - \tau_{ij}), \tag{11}$$

$$q_{ij}^{\mathrm{ZINB}}(w_{ij}|a_{ij} = 1)$$

$$= \begin{cases} \binom{m + w_{ij} - 1}{w_{ij}} \left(\frac{1}{1 - \tau_{ij}}\right) \left(\frac{1}{1 + \alpha z_{ij}}\right)^m \left(\frac{\alpha z_{ij}}{1 + \alpha z_{ij}}\right)^{w_{ij}}, & w_{ij} > 0 \\ 0, & w_{ij} \leqslant 0 \end{cases} \tag{12}$$

where, as the for the plain negative binomial model, $\alpha = m^{-1}$ and $\tau_{ij} = (\frac{1}{1 + \alpha z_{ij}})^m$. Moreover, as for the zero-inflated (ZIP) model, $1 - p_{ij}^{\mathrm{ZINB}} = \frac{1}{1 + G_{ij}} + \frac{G_{ij}}{1 + G_{ij}} \tau_{ij}$, i.e., the connection between nodes i and j can be missing either because a link is not there (with probability $\frac{1}{1 + G_{ij}}$) or because a link is there but has zero weight (with probability $\frac{G_{ij}}{1 + G_{ij}} \tau_{ij}$); consistently, i and j are connected because the weight is not zero (with probability $1 - \tau_{ij}$). In order to "dress" the GM, we need to identify some of the parameters of the negative binomial model with the usual econometric function. Upon considering that

$$\langle w_{ij} \rangle_{\mathrm{ZINB}} = p_{ij}^{\mathrm{ZINB}} \langle w_{ij} | a_{ij} \rangle_{\mathrm{ZINB}}$$

$$= p_{ij}^{\mathrm{ZINB}} \frac{z_{ij}}{1 - \tau_{ij}} = \frac{G_{ij}}{1 + G_{ij}} z_{ij}, \tag{13}$$

we can make the identification $z_{ij} = \rho(\omega_i \omega_j)^{\beta} d_{ij}^{\gamma}$ and $G_{ij} = \delta \omega_i \omega_j$.

As for the ZIP case, we are treating z_{ij} as a negative binomial-like expected weight, conditional to the success of a logit step, i.e., $z_{ij} = \frac{\langle w_{ij} \rangle_{\text{ZINB}}}{p_{ij}^{\text{logit}}}$, with $p_{ij}^{\text{logit}} = \frac{G_{ij}}{1+G_{ij}}$ (see Appendix B for a detailed description of the procedure to estimate the parameters of the ZINB model).

Let us notice that, while the ZIP model provides a better estimation of the link density than the Poisson model, the ZINB and the negative binomial ones basically perform in the same way. In fact,

$$\langle c \rangle_{\text{ZINB}} < c \simeq \langle c \rangle_{\text{ZIP}} \qquad (14)$$

since $c = \frac{2L}{N(N-1)} \simeq 0.63$, $\langle c \rangle_{\text{ZIP}} = \sum_{i<j} p_{ij}^{\text{ZIP}} \simeq 0.63$, and $\langle c \rangle_{\text{ZINB}} = \sum_{i<j} p_{ij}^{\text{ZINB}} \simeq 0.60$, a result suggesting that both variants of the negative binomial model will perform poorly in reproducing the binary properties of the WTW.

IV. STATISTICAL NETWORK MODELS - II: MAXIMUM-ENTROPY MODELS

The members of the second class of network models are the ones defined within the framework of traditional statistical mechanics. All of them can be derived by performing a constrained maximization of Shannon entropy [54] where the constraints represent the available information about the system at hand.

The simplest, yet nontrivial, ME model that can be considered comes from the maximization of the binary Shannon functional

$$S = -\sum_{\mathbf{A}} P(\mathbf{A}) \ln P(\mathbf{A}) \qquad (15)$$

constrained to reproduce the entire degree sequence, $\{k_i(\mathbf{A})\}_{i=1}^{N}$, of the network. This model is known under the name of undirected binary configuration model (UBCM) and has been shown to accurately reproduce many (binary) properties of a wide spectrum of real-world systems [44].

The UBCM is described by the probability mass function

$$P(\mathbf{A}) = \prod_{i<j} p_{ij}^{a_{ij}} (1 - p_{ij})^{1-a_{ij}}, \qquad (16)$$

which is factorized into the product of Bernoulli probability mass functions (one for each pair of nodes) with

$$p_{ij}^{\text{UBCM}} = \frac{x_i x_j}{1 + x_i x_j} \qquad (17)$$

(where x_i is the Lagrange multiplier controlling for the degree of node i). Importantly, the logit model admitting the presence of a single global constant can be derived from entropy maximization upon reparametrizing the Lagrange multipliers of the UBCM and imposing the total number of links as the only constraint [62]). The identification $x_i \equiv \sqrt{\delta}\omega_i$, in fact, leads to

$$p_{ij}^{\text{logit}} = \frac{G_{ij}}{1 + G_{ij}} = \frac{\delta\omega_i\omega_j}{1 + \delta\omega_i\omega_j}. \qquad (18)$$

Although the functional form above is not the most general one (for instance, dyadic factors such as geographic distances could be added as well), it is the form we will adopt in what follows.

In the network literature, the logit model (in its formulation above) has been popularized [14] as one particular case of the so-called fitness model [59] and as the so-called density-corrected gravity model (dcGM) [68] and has been proven to perform remarkably well for the task of reconstructing the topology of networks from partial information [62].

Since we are interested in reproducing the structural properties of weighted networks, we need to complement the purely binary step above with a recipe for reconstructing weights. The entropy-based framework handles such a requirement via the maximization of conditional Shannon functionals allowing the specification of $P(\mathbf{A})$ to be disentangled from that of $Q(\mathbf{W}|\mathbf{A})$ [51].

When discrete weighted models are considered, a useful quantity is the conditional Shannon entropy

$$S(\mathscr{W}|\mathscr{A}) = -\sum_{\mathbf{A} \in \mathbb{A}} P(\mathbf{A}) \sum_{\mathbb{W}_{\mathbf{A}}} Q(\mathbf{W}|\mathbf{A}) \log Q(\mathbf{W}|\mathbf{A}) \qquad (19)$$

where the first sum runs over all binary configurations within the ensemble $\mathbb{A}$ and the second sum runs over all weighted configurations that are compatible with each specific binary structure represented by the adjacency matrix $\mathbf{A}$, i.e., such that $\mathbb{W}_{\mathbf{A}} = \{\mathbf{W} : \Theta[\mathbf{W}] = \mathbf{A}\}$.

Conditional maximization proceeds by specifying a set of weighted constraints that, in the discrete case, reads

$$1 = \sum_{\mathbb{W}_{\mathbf{A}}} P(\mathbf{W}|\mathbf{A}), \quad \forall \, \mathbf{A} \in \mathbb{A} \qquad (20)$$

$$\langle C_\alpha \rangle = \sum_{\mathbf{A} \in \mathbb{A}} P(\mathbf{A}) \sum_{\mathbb{W}_{\mathbf{A}}} Q(\mathbf{W}|\mathbf{A}) C_\alpha(\mathbf{W}), \quad \forall \, \alpha \qquad (21)$$

the first condition ensuring the normalization of the conditional probability mass function and the vector $\{C_\alpha(\mathbf{W})\}$ representing the "proper" set of weighted constraints. Differentiating the corresponding Lagrangean functional with respect to $Q(\mathbf{W}|\mathbf{A})$ and equating the result to zero leads to

$$Q(\mathbf{W}|\mathbf{A}) = \begin{cases} \frac{e^{-H(\mathbf{W})}}{Z_{\mathbf{A}}}, & \mathbf{W} \in \mathbb{W}_{\mathbf{A}} \\ 0, & \mathbf{W} \notin \mathbb{W}_{\mathbf{A}} \end{cases} \qquad (22)$$

where $H(\mathbf{W}) = \sum_\alpha \psi_\alpha C_\alpha$ is the so-called Hamiltonian, listing the constrained, weighted quantities, and $Z_{\mathbf{A}} = \sum_{\mathbb{W}_{\mathbf{A}}} e^{-H(\mathbf{W})}$ is the partition function for fixed $\mathbf{A}$. The explicit functional form of $Q(\mathbf{W}|\mathbf{A})$ can be obtained only once the functional form of the constraints has been specified as well.

To this aim, let us consider the Hamiltonian

$$H(\mathbf{W}) = \sum_{i<j} \phi_{ij} w_{ij}, \qquad (23)$$

where weights are modelled as nonnegative integer variables, i.e., $w_{ij} \in \mathbb{N}, \forall \, i < j$. This choice induces a conditional probability mass function reading

$$Q(\mathbf{W}|\mathbf{A}) = \prod_{i<j} q_{ij}(w_{ij}|a_{ij}) = \prod_{i<j} y_{ij}^{w_{ij}-a_{ij}} (1 - y_{ij})^{a_{ij}} \qquad (24)$$

(with $e^{-\psi_{ij}} = e^{-\phi_{ij}} = y_{ij}$). Let us now turn the model above into a proper econometric one. To this aim, let us proceed by analogy. All zero-inflated econometric recipes identify z_{ij} with a conditional expected weight, a prescription that in our case, would translate into its identification with

$\langle w_{ij}|a_{ij}\rangle = \frac{1}{1-y_{ij}}$. This choice, however, would lead to an inconsistency, since $z_{ij} > 0$ is a positive real number while $\langle w_{ij}|a_{ij}\rangle$ must necessarily exceed 1, as it represents the expected weight conditional to the existence of a connection. An alternative, consistent econometric identification is

$$\frac{1}{1-y_{ij}} = 1 + z_{ij}, \qquad (25)$$

which, in turn, induces a conditional probability mass function

$$Q(\mathbf{W}|\mathbf{A}) = \prod_{i<j} q_{ij}(w_{ij}|a_{ij})$$
$$= \prod_{i<j}\left(\frac{z_{ij}}{1+z_{ij}}\right)^{w_{ij}-a_{ij}}\left(\frac{1}{1+z_{ij}}\right)^{a_{ij}}. \quad (26)$$

Models of the kind are known as hurdle models: Quite remarkably, entropy maximization allows us to recover them in a fully principled way, i.e., by eliminating the (otherwise unavoidable) ambiguity that accompanies the choice of the distribution (supposedly) describing the positive values of an economic system.

The hurdle-geometric model derived above, however, suffers from a number of limitations, the most relevant of which is that of failing in reproducing basic network quantities such as the WTW total weight. As an illustrative example, while $W = \sum_{i<j} w_{ij} \simeq 10^9$, in the year 2000, we find that $\langle W \rangle_{\text{h-g}} \simeq 10^5$. In order to overcome such a limitation, we have considered the conditional probability mass function induced by the Hamiltonian

$$H(\mathbf{W}) = \sum_{i<j}(\phi_0 + \phi_{ij})w_{ij} = \phi_0 W + \sum_{i<j}\phi_{ij}w_{ij}. \quad (27)$$

Identifying $e^{-\phi_0} = y_0$ and $e^{-\phi_{ij}} = y_{ij} = \frac{z_{ij}}{1+z_{ij}}$ [see Eq. (25)], we arrive at the modified econometric model

$$q_{ij}(w_{ij}|a_{ij}) = \left(\frac{y_0 z_{ij}}{1+z_{ij}}\right)^{w_{ij}-a_{ij}}\left(\frac{1+z_{ij}-y_0 z_{ij}}{1+z_{ij}}\right)^{a_{ij}}. \quad (28)$$

We are now ready to fully specify the suite of discrete entropy-models that we will compare with the aforementioned, purely econometric ones. To this aim, we need to fully specify the functional form

$$Q(\mathbf{W}) = \prod_{i<j} q_{ij}(w_{ij})$$
$$= \prod_{i<j} p_{ij}^{a_{ij}}(1-p_{ij})^{1-a_{ij}} \cdot q_{ij}(w_{ij}|a_{ij})$$
$$= P(\mathbf{A})Q(\mathbf{W}|\mathbf{A}); \qquad (29)$$

the two most obvious choices are represented by the models

$$Q_{\text{TSF}}(\mathbf{W}) = P_{\text{logit}}(\mathbf{A})Q(\mathbf{W}|\mathbf{A})$$
$$= \prod_{i<j}\left(p_{ij}^{\text{logit}}\right)^{a_{ij}}\left(1-p_{ij}^{\text{logit}}\right)^{1-a_{ij}} \cdot q_{ij}(w_{ij}|a_{ij}) \quad (30)$$

and

$$Q_{\text{TS}}(\mathbf{W}) = P_{\text{UBCM}}(\mathbf{A})Q(\mathbf{W}|\mathbf{A})$$
$$= \prod_{i<j}\left(p_{ij}^{\text{UBCM}}\right)^{a_{ij}}\left(1-p_{ij}^{\text{UBCM}}\right)^{1-a_{ij}} \cdot q_{ij}(w_{ij}|a_{ij}) \quad (31)$$

that combine the weighted, conditional step induced by the Hamiltonian defined in Eq. (27) with the purely binary logit model and with the undirected binary configuration model, respectively. The acronyms stand for "two-step fitness" model and "two-step" model and recall the names originally used to define them [15,61].

Less trivial choices are represented by models whose both binary and weighted portions are jointly determined by the constraints. They can all be recovered as specifications of the generic Hamiltonian

$$H(\mathbf{W}) = \sum_{i<j}[\theta_{ij}a_{ij} + (\phi_0 + \phi_{ij})w_{ij}]$$
$$= \sum_{i<j}\theta_{ij}a_{ij} + \phi_0 W + \sum_{i<j}\phi_{ij}w_{ij}; \quad (32)$$

in what follows, we will consider two different instances of such a function, defined by the choices $\theta_{ij} = \theta_0$ and $\theta_{ij} = \theta_i + \theta_j$. In other words, while we let the weighted parts of these models coincide and read as in Eq. (28), we allow for the binary part to vary, either constraining the total number of links L, or the entire degree sequence $\{k_i(\mathbf{A})\}_{i=1}^N$. In the first case, our Hamiltonian reads

$$H_{(1)}(\mathbf{W}) = \theta_0 L + \phi_0 W + \sum_{i<j}\phi_{ij}w_{ij} \quad (33)$$

and instances the model in Eq. (29) with

$$p_{ij}^{(1)} = \frac{xy_0 z_{ij}}{1+z_{ij}-y_0 z_{ij}+xy_0 z_{ij}} \quad (34)$$

$$q_{ij}^{(1)}(w_{ij}|a_{ij}) = \left(\frac{y_0 z_{ij}}{1+z_{ij}}\right)^{w_{ij}-a_{ij}}\left(\frac{1+z_{ij}-y_0 z_{ij}}{1+z_{ij}}\right)^{a_{ij}} \quad (35)$$

(having posed $e^{-\theta_0} = x$, $e^{-\phi_0} = y_0$ and $e^{-\phi_{ij}} = y_{ij} = \frac{z_{ij}}{1+z_{ij}}$); in the second case, it reads

$$H_{(2)}(\mathbf{W}) = \sum_i \theta_i k_i + \psi_0 W + \sum_{i<j}\psi_{ij}w_{ij} \quad (36)$$

and instances the model in Eq. (29) with

$$p_{ij}^{(2)} = \frac{x_i x_j y_0 z_{ij}}{1+z_{ij}-y_0 z_{ij}+x_i x_j y_0 z_{ij}} \quad (37)$$

$$q_{ij}^{(2)}(w_{ij}|a_{ij}) = \left(\frac{y_0 z_{ij}}{1+z_{ij}}\right)^{w_{ij}-a_{ij}}\left(\frac{1+z_{ij}-y_0 z_{ij}}{1+z_{ij}}\right)^{a_{ij}} \quad (38)$$

(having posed $e^{-\theta_i} = x_i$, $e^{-\phi_0} = y_0$, and $e^{-\phi_{ij}} = y_{ij} = \frac{z_{ij}}{1+z_{ij}}$). Appendix C provides a detailed description of the procedure we have adopted to estimate the parameters entering into the definition of our basket of discrete ME models.

So far, we have turned entropy-based models into econometric ones via a suitable econometric transformation of the Lagrange multipliers defining the proper "physical" models. The entropy-based formalism, however, also offers the opportunity to define statistical models in a fully data-driven fashion. To this aim, let us consider the Hamiltonian

$$H_{\text{UECM}}(\mathbf{W}) = \sum_i [\theta_i k_i + \phi_i s_i]$$
$$= \sum_{i<j}[(\theta_i+\theta_j)a_{ij} + (\phi_i+\phi_j)w_{ij}] \quad (39)$$

that constrains both degrees and strengths. The model induced by the latter ones is called undirected enhanced configuration model (UECM) and represents the best-performing one for the task of network reconstruction in presence of full information about the constraints [46,57].

Remarkably, all models considered in the previous section can be compactly derived by combining a logit-like probability mass function describing the binary network structure with the conditional expression defined in Eq. (28). To prove this, it is enough to notice that all the Bernoulli-like probability mass functions characterizing our model can be compactly rewritten as

$$p_{ij} = \frac{x_i' x_j'}{1 + x_i' x_j'} \tag{40}$$

where

$$(x_i' x_j')^{\text{logit}} = \delta \omega_i \omega_j, \tag{41}$$

$$(x_i' x_j')^{\text{UBCM}} = x_i x_j, \tag{42}$$

$$(x_i' x_j')^{(1)} = \frac{x y_0 z_{ij}}{1 + z_{ij} - y_0 z_{ij}}, \tag{43}$$

$$(x_i' x_j')^{(2)} = \frac{x_i x_j y_0 z_{ij}}{1 + z_{ij} - y_0 z_{ij}}, \tag{44}$$

$$(x_i' x_j')^{\text{UECM}} = \frac{x_i x_j y_i y_j}{1 - y_i y_j}. \tag{45}$$

V. RESULTS

Let us now test and compare the performance of our two broad classes of models in reproducing the topological properties of the World Trade Web. To this aim, let us consider both the local properties, such as the degrees and the strengths, and the higher-order ones such as the average nearest neighbors degree (ANND) and the clustering coefficient (BCC), i.e.,

$$k_i^{nn} = \frac{\sum_{j(\neq i)} a_{ij} k_j}{k_i}, \tag{46}$$

$$c_i = \frac{\sum_{j(\neq i,k)} \sum_{k(\neq i)} a_{ij} a_{jk} a_{ki}}{k_i(k_i - 1)}; \tag{47}$$

we will also consider their weighted counterparts, i.e., the average nearest-neighbors strength (ANNS) and the weighted clustering coefficient (WCC), defined as

$$s_i^{nn} = \frac{\sum_{j(\neq i)} a_{ij} s_j}{k_i}, \tag{48}$$

$$c_i^w = \frac{\sum_{j(\neq i,k)} \sum_{k(\neq i)} w_{ij} w_{jk} w_{ki}}{k_i(k_i - 1)}. \tag{49}$$

We will also test the accuracy of the reconstruction provided by the methods in our basket by considering properties like the true positive rate (TPR)

$$\langle TPR \rangle = \frac{\langle TP \rangle}{L} = \frac{\sum_{i<j} a_{ij} p_{ij}}{L}, \tag{50}$$

i.e., the percentage of links correctly recovered by a given reconstruction method, the specificity (SPC)

$$\langle SPC \rangle = \frac{\langle TN \rangle}{\binom{N}{2} - L} = \frac{\sum_{i<j}(1 - a_{ij})(1 - p_{ij})}{\binom{N}{2} - L}, \tag{51}$$

i.e., the percentage of zeros correctly recovered by a given reconstruction method, the positive predictive value (PPV)

$$\langle PPV \rangle = \frac{\langle TP \rangle}{\langle L \rangle} = \frac{\sum_{i<j} a_{ij} p_{ij}}{\langle L \rangle}, \tag{52}$$

i.e., the percentage of links correctly recovered by a given reconstruction method with respect to the total number of predicted links and the accuracy (ACC)

$$\langle ACC \rangle = \frac{\langle TP \rangle + \langle TN \rangle}{\binom{N}{2}} \tag{53}$$

measuring the overall performance of a given reconstruction method in correctly placing both links and zeros.

Figure 2 sums up the comparisons carried out between the econometric models and the ME ones. The comparison between the empirical cumulative density function (CDF) of the degrees and the ones output by the econometric models reveals the latter ones to be able to predict an overall similar functional form [see Figs. 2(a) and 2(b)]; still, the prediction obtained by any of the ME models is much closer to the empirical trend. More quantitatively, one can implement the Kolmogorov-Smirnov (KS) test to check the goodness of any of the models considered in the present work to replicate the empirical degrees: while any of the ME models provides estimates of the degrees that are compatible with the empirical CDF (at the significance level of 5%), only the ZIP model predicts degrees that are compatible with the empirical ones: in fact, the p-values of the ME models read $p_{(1)} \simeq 0.06$, $p_{(2)} \simeq 0.99$, $p_{\text{TS}} \simeq 0.99$, $p_{\text{TSF}} \simeq 0.32$ while the p-values of the econometric models read $p_{\text{Pois}} \simeq 0.001$, $p_{\text{NB}} \simeq 0.0008$, $p_{\text{ZIP}} \simeq 0.63$, $p_{\text{ZINB}} \simeq 0.001$.

Coming to the higher-order properties, it is evident that the majority of the econometric models fails to overlap with the empirical cloud of points [see Fig. 2(b)]: the one providing the best prediction is the ZIP model, whose performance represents quite an improvement with respect to the one provided by the "plain" Poisson model. While this is quite evident for what concerns the prediction of the ANND values, the performances of the ZIP and of the "plain" Poisson model become less different when tested on the BCC values. On the contrary, the performance of the ZINB model closely resembles that of the negative binomial one when tested both on the ANND and on the BCC. As for the local properties, the KS test reveals that the only model outputting predictions compatible with the empirical values (at the significance level of 5%) is the ZIP one: in fact, $p_{\text{ZIP}}^{\text{ANND}} \simeq 0.38$, $p_{\text{ZIP}}^{\text{BCC}} \simeq 0.08$.

For what concerns ME models, the ones performing best are those constraining the degrees, i.e., the model induced by $H_{(2)}$ and its two-step counterpart, whose topological estimation step is carried out by employing p_{ij}^{UBCM}. The evidence that their performances in reproducing the purely binary structure of a network are very similar lets us suspect that $p_{ij}^{(2)} \simeq p_{ij}^{\text{UBCM}}$ and conclude that the purely econometric information encoded into $p_{ij}^{(2)}$ does not add much to what is already conveyed
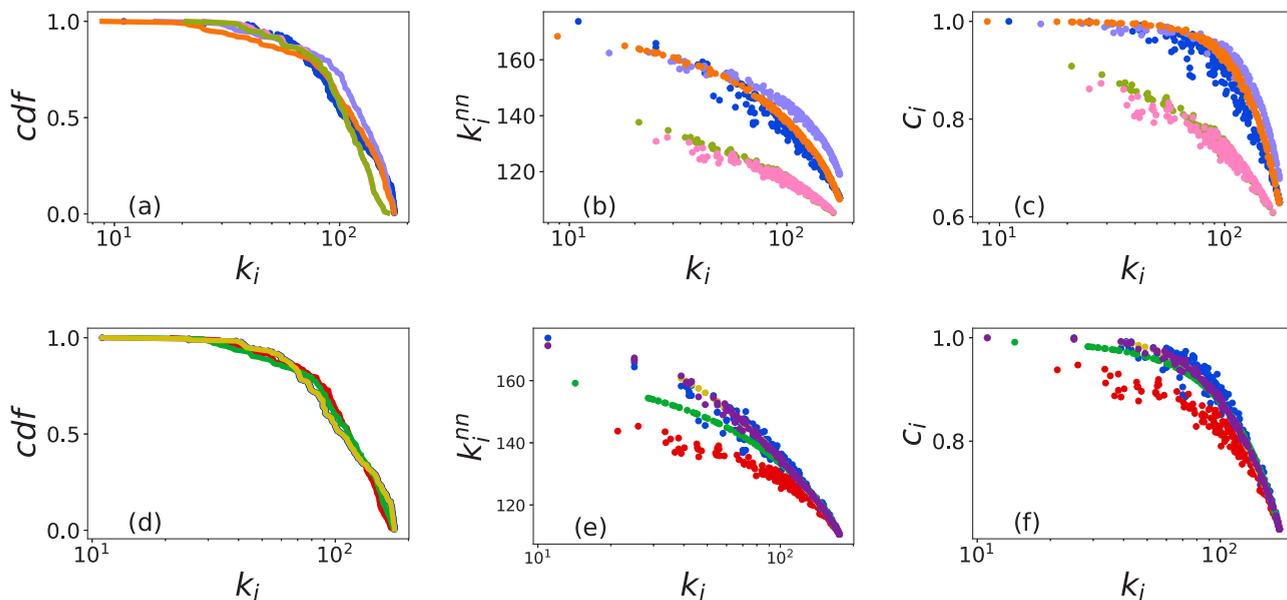
FIG. 2. Performance of econometric models (top row) vs the performance of ME models (bottom row) in reproducing: [(a),(d)] the degrees; [(b), (e)] the ANND; and [(c), (f)] the BCC. Empirical points are indicated as ●; econometric and ME models are indicated as follows: ● - Poisson; ● - Negative Binomial; ● - ZIP; ● - ZINB; ● - $H_{(1)}$; ● - $H_{(2)}$; ● - TS; ● - TSF. Results refer to the year 2000 of the dataset curated by Gleditsch [63].

by the purely topological one. On the other hand, ME models not constraining the degrees provide predictions differing from the empirical trends to quite a large extent. As the KS test reveals, the only ME model outputting predictions that are not compatible with the empirical values (at the significance level of 5%) is the one induced by $H_{(1)}$.

The overall accuracy of our models in reproducing a network topology can be proxied by the index $\Delta_L = |\langle L \rangle - L|/L$ amounting at $\Delta_L^{\text{Pois}} \gtrsim \Delta_L^{\text{NB}} = \Delta_L^{\text{ZINB}} \simeq 6\%$ while $\Delta_L^{\text{ZIP}} \simeq 0.5\%$ and $\Delta_L^{\text{ME}} = 0$ for each ME model. This is confirmed by our analysis of single link statistics: In fact, $\langle ACC \rangle_{(2)} \simeq 0.83$ attains the largest value, followed by $\langle ACC \rangle_{\text{TS}} \simeq 0.81$ and $\langle ACC \rangle_{\text{ZIP}} \simeq 0.77$. Remarkably, $\langle PPV \rangle_{(2)} \simeq 0.86$ attains the largest value, indicating that the ME model induced by $H_{(2)}$ is the one placing links best among all the models in our basket.

Let us now consider the weighted properties (see Fig. 3). Overall, the distribution of the strengths is reproduced quite well by all models considered here, although no one explicitly constrains them. This seems to indicate that the purely econometric information "feeded" into our models indeed plays a role, which, however, is limited to ensure that the intensive margins (and the related properties, as we will see) are accurately predicted. The larger explanatory power of econometric models becomes now evident: All of them output predictions that are compatible with the empirical values. Although the same result holds true for ME models, the latter ones are outperformed by purely econometric models—the best performing ones in predicting the strengths being the Poisson-like ones.

Coming to the higher-order properties, let us notice that the best performing econometric models in reproducing the ANNS values are the ZIP and the "plain" Poisson ones whose performances differ less than in the ANND case, although the

KS test lets the ZIP model win. On the other hand, the ZINB and the negative binomial models (whose performances are, again, very similar) completely fail in capturing the empirical values. All predictions from ME models overlap with the empirical ANNS values: As the KS test reveals, the only ME model outputting predictions that are not compatible with the empirical values (at the significance level of 1%) is the one induced by $H_{(1)}$. For what concerns the values of the WCC, both the econometric and the ME models perform quite satisfactorily in capturing its rising trend. However, the KS test reveals that only the econometric models and the TS model output predictions compatible with the WCC empirical values (at the significance level of 1%).

To proxy the accuracy of our models in reproducing the weighted network structure, we have considered the index $\Delta_W = |\langle W \rangle - W|/W$ amounting at $\Delta_W^{\text{ZINB}} \simeq 95\%$, $\Delta_W^{\text{NB}} = 60\%$, $\Delta_W^{\text{ZIP}} \simeq 0.3\%$ and $\Delta_W^{\text{Pois}} \simeq 0$ while $\Delta_W^{\text{ME}} = 0$ for the ME models constraining the total weight and $\Delta_W^{\text{TS}} \gtrsim \Delta_W^{\text{TSF}} \simeq 0.2\%$ for the ME two-step ones.

In order to understand if the conclusions above can be generalized, let us calculate the accuracy of all models in our basket, for all years constituting our two datasets. The results, summed up in Table I, confirm that model $H_{(2)}$ systematically outperforms all competing models. As an additional test, we have calculated the percentage of times the empirical values of our network statistics are compatible with their ensemble distributions, via KS tests at the significance level of 5%: the results, shown in Fig. 4, point out that ME models are the ones for which compatibility is largest.

Let us now ask ourselves if a criterion exist to carry out a principled comparison of the performance of the models considered in the present paper. The answer is positive and lays in the adoption of the popular *Akaike information criterion* and *Bayesian information criterion*, respectively defined
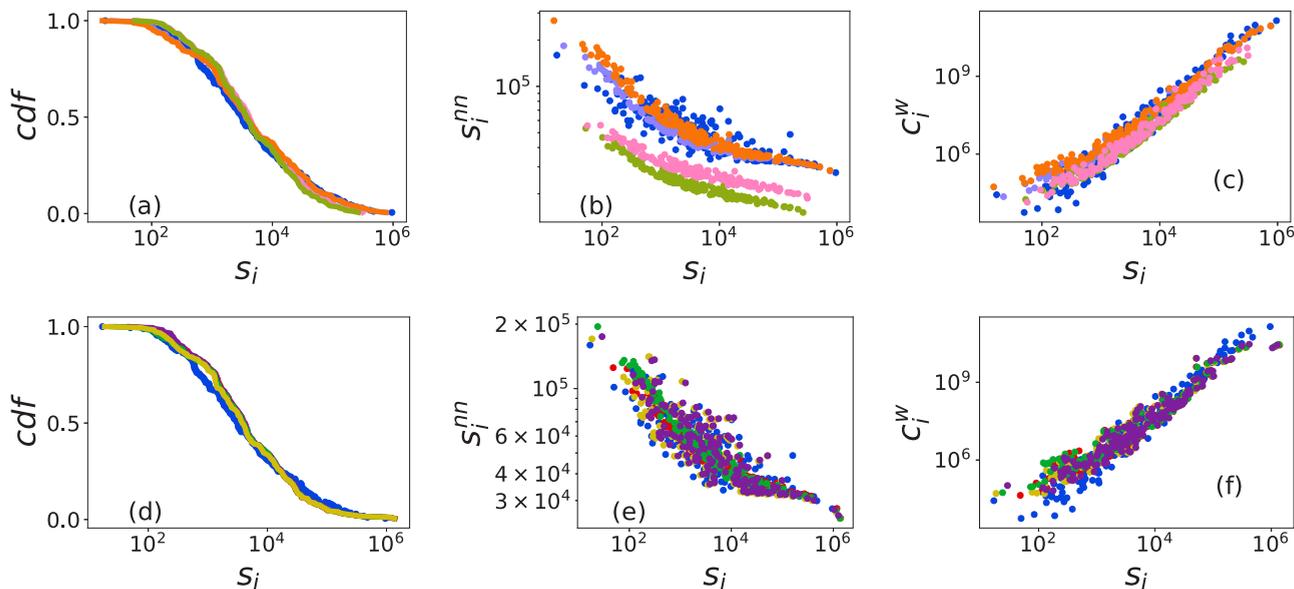
FIG. 3. Performance of econometric models (top row) vs the performance of ME models (bottom row) in reproducing: [(a), (d)] the strengths; [(b), (e)] the ANNS; and [(c), (f)] the WCC. Empirical points are indicated as •; econometric and ME models are indicated as follows: • - Poisson; • - Negative Binomial; • - ZIP; • - ZINB; • - $H_{(1)}$; • - $H_{(2)}$; • - TS; • - TSF. Results refer to the year 2000 of the dataset curated by Gleditsch [63].

as

$$\text{AIC} = 2K - 2\mathcal{L} \qquad (54)$$

and

$$\text{BIC} = K \ln n - 2\mathcal{L} \qquad (55)$$

where $\mathcal{L}$ is the log-likelihood of the tested model evaluated in its maximum, K is the number of parameters characterizing the model itself and n is the cardinality of the set of observations, estimated as $\frac{N(N-1)}{2}$ for undirected network data. Model selection based on these criteria prescribes to rank models according to (either) their AIC or BIC value and choose the

TABLE I. Accuracy of ME and econometric models in reconstructing both the Gleditsch and the BACI dataset. The best performing models are $H_{(2)}$ and TS.

Dataset	$H_{(1)}$	$H_{(2)}$	TSF	TS	Poisson	ZIP	Negative binomial	ZINB
Gleditsch$_{90}$	0.72	**0.81**	0.73	0.79	0.75	0.75	0.68	0.70
Gleditsch$_{91}$	0.69	**0.78**	0.70	0.77	0.73	0.73	0.65	0.68
Gleditsch$_{92}$	0.70	**0.79**	0.71	0.77	0.73	0.73	0.65	0.68
Gleditsch$_{93}$	0.70	**0.79**	0.71	0.78	0.74	0.74	0.66	0.69
Gleditsch$_{94}$	0.70	**0.79**	0.72	0.78	0.75	0.75	0.66	0.69
Gleditsch$_{95}$	0.70	**0.80**	0.72	**0.80**	0.75	0.76	0.68	0.69
Gleditsch$_{96}$	0.72	**0.81**	0.73	**0.81**	0.76	0.77	0.68	0.69
Gleditsch$_{97}$	0.73	**0.82**	0.74	0.81	0.77	0.77	0.69	0.71
Gleditsch$_{98}$	0.73	**0.82**	0.74	0.81	0.77	0.77	0.69	0.71
Gleditsch$_{99}$	0.73	**0.82**	0.74	0.81	0.77	0.77	0.69	0.72
Gleditsch$_{00}$	0.74	**0.82**	0.74	0.81	0.78	0.77	0.70	0.72
BACI$_{07}$	0.83	**0.88**	0.83	0.87	0.84	0.84	0.76	0.77
BACI$_{08}$	0.83	**0.88**	0.83	0.87	0.85	0.84	0.76	0.77
BACI$_{09}$	0.84	**0.88**	0.84	0.87	0.84	0.84	0.76	0.77
BACI$_{10}$	0.85	**0.89**	0.85	0.88	0.85	0.85	0.77	0.78
BACI$_{11}$	0.85	**0.89**	0.85	0.88	0.86	0.85	0.77	0.78
BACI$_{12}$	0.85	**0.89**	0.85	0.88	0.86	0.85	0.77	0.78
BACI$_{13}$	0.85	**0.90**	0.85	0.89	0.86	0.86	0.77	0.78
BACI$_{14}$	0.85	**0.90**	0.85	0.89	0.86	0.85	0.77	0.78
BACI$_{15}$	0.85	**0.90**	0.85	0.89	0.85	0.84	0.77	0.78
BACI$_{16}$	0.84	**0.90**	0.84	0.89	0.85	0.84	0.76	0.78
BACI$_{17}$	0.85	**0.90**	0.85	0.89	0.85	0.85	0.77	0.78
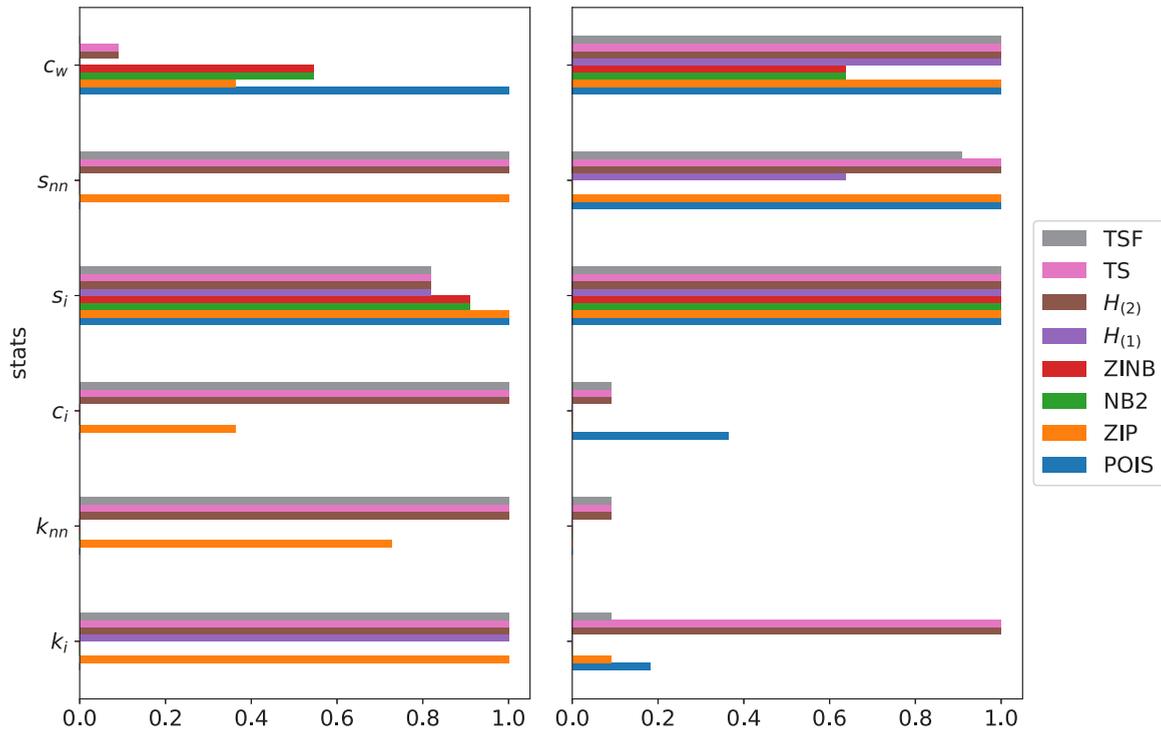
FIG. 4. Percentage of times the empirical value of a given network statistics is compatible with its ensemble distribution, according to a KS test at the significance level of 5% (left, for the Gleditsch dataset; right, for the BACI dataset): A value of 1 means that the given network statistic is compatible with the model-induced ensemble distribution across all years of the considered dataset. Generally speaking, ME models are the ones for which compatibility is largest—although econometric models are characterized by a compatibility, which is large as well whenever weighted measures are considered.

one minimizing it. Table II shows both the AIC and the BIC values for all the models considered here.

Quite surprisingly, the negative binomial model is the favoured one among the econometric models, followed by its zero-inflated version; however, its bad performance in reproducing the empirical trends makes the choice of including it among the most suitable models for modeling trade data highly questionable. On the other hand, the ZIP model performs much better in reconstructing the trends of both local

and higher-order properties although being much less parsimonious than both versions of the negative binomial model. Apparently, then, the question about which model to prefer—the favoured one by information criteria or the best performing one in reproducing trends?—cannot be properly answered by just considering purely econometric models. On the other hand, such a question can be unambiguously answered as soon as one switches to the class of maximum-entropy models: now, both the AIC- and the BIC-based rankings favour the

TABLE II. Selection of ME and econometric models, according to the AIC and BIC values. While the negative binomial model is the favoured one among the econometric models, it performs badly in reproducing both local and higher-order topological properties; on the other hand, the ZIP model reproduces the higher-order statistics quite accurately but is disfavoured by both AIC and BIC. The solution to this dilemma comes from considering a different class of reconstruction models, i.e., the maximum-entropy ones: Beside being the favoured one by information criteria, the $H_{(2)}$ model achieves a very good reconstruction of both binary and weighted network properties (since $\Delta L = |\langle L \rangle - L|/L$ and $\Delta W = |\langle W \rangle - W|/W$, the symbol 0^* indicates that the model exactly reproduces the corresponding constraint). Results refer to the year 2000 of the dataset curated by Gleditsch [63].

Model	Δ_L	Δ_W	AIC	BIC
Poisson	$\simeq 0.07$	0^*	$\simeq 5088080.1$	$\simeq 5088105.1$
ZI Poisson	$\simeq 4.5 \times 10^{-3}$	$\simeq 3.3 \times 10^{-3}$	$\simeq 5052767.3$	$\simeq 5052800.6$
Negative binomial	$\simeq 0.06$	$\simeq 0.59$	$\simeq \mathbf{175693.2}$	$\simeq \mathbf{175726.6}$
ZI Negative binomial	$\simeq 0.06$	$\simeq 0.95$	$\simeq 176071.4$	$\simeq 176113.1$
ME model $H_{(1)}$	0^*	0^*	$\simeq 182909.5$	$\simeq 182951.2$
ME model $H_{(2)}$	0^*	0^*	$\simeq \mathbf{174050.0}$	$\simeq \mathbf{175550.4}$
TS	0^*	$\simeq 6.8 \times 10^{-3}$	$\simeq 176129.1$	$\simeq 177629.5$
TSF	0^*	$\simeq 2.2 \times 10^{-3}$	$\simeq 185940.2$	$\simeq 185981.8$
UECM	0^*	0^*	$\simeq 186602.8$	$\simeq 189536.8$

model described by the Hamiltonian $H_{(2)}$—the one encoding the information about the degree sequence and the total weight, plus admitting a tunable function of the weights—i.e., precisely the most accurate in replicating many (if not all) empirical trends.

For the sake of comparison, we have included into the basket of maximum-entropy models the undirected enhanced configuration model (UECM), i.e., the model performing best in presence of complete information about the constraints—degrees and strengths, in the specific case—of a given networked system: as evident from the table, it is disfavoured with respect to the model described by the Hamiltonian $H_{(2)}$, an evidence signaling that while the information encoded into the degrees is essential (i.e., the latter ones must be explicitly constrained), the one carried by the strengths appears to be "less fundamental" since providing a good approximation of them is enough to obtain an overall good reconstruction.

In order to understand if the conclusions above can be generalized, let us calculate the Akaike weights for the models in our basket $\mathcal{M}$. The Akaike weight for the ith model is defined as

$$w_i = \frac{e^{-\Delta_i/2}}{\sum_m e^{-\Delta_m/2}}, \qquad (56)$$

with $\Delta_i = \mathrm{AIC}_i - \min\{\mathrm{AIC}_m\}_{m \in \mathcal{M}}$. Results on the dataset curated by Gleditsch show that the negative binomial and the $H_{(2)}$ models "compete", in the sense that $H_{(2)}$ performs best (i.e., $w_{H_{(2)}} \simeq 1$) in the (bunches of) years 1990–1993 and 1997–2000 while the negative binomial model performs best (i.e., $w_{NB} \simeq 1$) in the (bunch of) years 1994–1996. For what concerns the BACI dataset, instead, the competing models are three: In fact, while $H_{(2)}$ performs best in the (bunches of) years 2007, 2009, and 2015–2017, the negative binomial outperforms the others in the (bunches of) years 2008 and 2010–2014; however, the ZINB model has a positive, non-negligible Akaike weight in the years 2008, 2012, and 2014, hence performing as well as the negative binomial one.

Let us now consider a couple of additional exercises, carried out on both datasets considered in the present paper. The first one concerns link prediction and was carried out by following the reference [69]. Specifically, we have approached link prediction from a temporal perspective, inspecting the accuracy achieved by our reconstruction models at time $t + 1$ given the knowledge about the network topology (for the maximum-entropy models) and of the other exogenous variables (for the purely econometric models) at time t. In other words, we opted for a one-lagged link prediction, calculating the log-likelihood

$$\mathcal{L}_{1l} = \ln \left[\prod_{i<j} \left(p_{ij}^{(t)}\right)^{a_{ij}^{(t+1)}} \left(1 - p_{ij}^{(t)}\right)^{1 - a_{ij}^{(t+1)}} \right] \qquad (57)$$

for each statistical model in our basket, the coefficients $\{p_{ij}^{(t)}\}_{i,j=1}^N$ being the probabilities output by any model at time t and the coefficients $\{a_{ij}^{(t+1)}\}_{i,j=1}^N$ being the entries of the adjacency matrix at time $t + 1$. When carried out on the pairs of years 1993–1994, 1994–1995, 1995–1996, 1996–1997, 1997–1998, 1998–1999, and 1999–2000 of the dataset curated by Gleditsch and on the pairs of years 2008–2009, 2009–2010, 2010–2011, 2013–2014, 2014–2015, 2015–2016, and 2016–

2017 of the BACI dataset, the exercise above shows $H_{(2)}$ to outperform not only the entire class of econometric models but also the purely binary, maximum-entropy ones—as confirmed by the Akaike weights induced by the log-likelihood above.

To provide a more refined picture of the performance of the models in our basket in providing one-lagged predictions, we have also calculated their one-lagged accuracy, defined as

$$\langle ACC \rangle_{1l} = \frac{\langle TP \rangle_{1l} + \langle TN \rangle_{1l}}{\binom{N}{2}} \qquad (58)$$

with $\langle TP \rangle_{1l} = \sum_{i<j} a_{ij}^{(t+1)} p_{ij}^{(t)}$ and $\langle TN \rangle_{1l} = \sum_{i<j} (1 - a_{ij}^{(t+1)})(1 - p_{ij}^{(t)})$. The results are reported in Table III and confirm what has been previously said: $H_{(2)}$ is the one performing best.

As a second exercise, we have tested the accuracy of our models in estimating link-specific weights—hence, carrying out what we have called a "weight prediction" exercise. To this aim, we have considered each expected weight and calculated the confidence interval enclosing the 95% of total probability around it. On the practical side, we have sampled 1000 configurations from the ensemble induced by each model in our basket and calculated the (ensemble-induced) 2.5 and 97.5 percentiles for each specific weight; then, we have calculated the percentage of empirical weights "falling" within the corresponding CIs—now treated as error bars "accompanying" the point-estimate of each weight. The results of this exercise are reported in in Table IV: as it can be appreciated, maximum-entropy models compete with both the negative binomial and the ZINB ones—although the latter (slightly) outperform the former.

VI. DISCUSSION

In the present paper, we have compared the performance of two broad classes of statistical models, i.e., the ones rooted into economic theory and the ones rooted into statistical physics (in particular, the ones derived from the maximum-entropy principle) in reconstructing both the binary and the weighted network properties of an economic system such as the WTW.

Although the case study is the same, the two classes of models "reflect" the languages of two disciplines that are still deeply different: While econometricians have traditionally focused on bilateral trade volumes between countries—emphasizing the role played by common borders, language, religion, the presence of regional trade agreements, etc. on trade relationships—network scientists have, instead, paid more attention to the structural and dynamical aspects of network formation, emphasizing the role played by purely structural information in determining the topology itself. The 2008 global financial crisis has dramatically clarified that bilateral trade relationships can explain only a small fraction of the impact that an economic shock, originating in a given country, can have on another country, which is not a direct trade partner, urging researchers in economics to adopt a network-aware perspective. This, in turn, has motivated us to carry out a methodological comparison on real-world cases, with the aim of clarifying pros and cons of both approaches.

TABLE III. One-lagged accuracy, quantifying the performance of ME and econometric models in providing one-lagged predictions on both the Gleditsch and the BACI datasets. $H_{(2)}$ and TS are systematically the best performing models.

Dataset	$H_{(1)}$	$H_{(2)}$	TSF	TS	Poisson	ZIP	Negative binomial	ZINB
Gleditsch$_{94}$	0.70	**0.79**	0.72	0.77	0.75	0.74	0.66	0.69
Gleditsch$_{95}$	0.71	**0.80**	0.72	0.78	0.75	0.75	0.66	0.69
Gleditsch$_{96}$	0.72	**0.80**	0.73	**0.80**	0.76	0.76	0.67	0.70
Gleditsch$_{97}$	0.73	**0.82**	0.74	0.80	0.77	0.77	0.67	0.70
Gleditsch$_{98}$	0.73	**0.82**	0.74	0.81	0.77	0.77	0.69	0.71
Gleditsch$_{99}$	0.73	**0.82**	0.74	0.81	0.77	0.77	0.69	0.71
Gleditsch$_{00}$	0.73	**0.82**	0.74	0.81	0.77	0.77	0.69	0.71
BACI$_{09}$	0.84	**0.88**	0.84	0.87	0.85	0.84	0.76	0.77
BACI$_{10}$	0.84	**0.88**	0.84	0.87	0.85	0.84	0.76	0.77
BACI$_{11}$	0.85	**0.89**	0.85	0.88	0.86	0.85	0.77	0.78
BACI$_{14}$	0.85	**0.89**	0.85	0.88	0.86	0.85	0.77	0.78
BACI$_{15}$	0.85	**0.89**	0.85	0.88	0.86	0.85	0.77	0.78
BACI$_{16}$	0.84	**0.88**	0.84	0.88	0.85	0.84	0.77	0.78
BACI$_{17}$	0.85	**0.89**	0.85	0.88	0.85	0.84	0.77	0.78

Researchers in economics have dealt with the issue of reconstructing network topology by approaching the simpler problem of reproducing the number of missing connections—or, equivalently, the link density. For instance, as we see from the year 2000 snapshot of the dataset curated by Gleditsch [63], although the error of the Poisson model in reproducing L is overall small (amounting at $\Delta_L^{\text{Pois}} \simeq 7\%$), it can be further reduced by adopting the zero-inflated version of it. On the contrary, inflating zeros does not improve the performance of the negative binomial model in reproducing the link density since it already underestimates L. The ability of a model in reproducing a global quantity such as the link density proxies its ability in providing a good estimation of local as well as higher-order topological properties (i.e., the degrees, the ANND and the BCC): From this point of view, the zero-inflated Poisson model is the one performing best among the econometric models. However, it is largely disfavoured by information criteria such as AIC and BIC, a result suggesting that it may be not parsimonious enough.

TABLE IV. Performance of ME and econometric models in providing accurate predictions of weights, on both the Gleditsch and the BACI datasets, quantified by calculating the percentage of empirical weights falling within the confidence interval enclosing the 95% probability around the corresponding expected value. Our maximum-entropy models compete with both the negative binomial and the ZINB ones—although the latter (slightly) outperform the former.

Dataset	$H_{(1)}$	$H_{(2)}$	TSF	TS	Poisson	ZIP	Negative binomial	ZINB
Gleditsch$_{90}$	0.96	0.96	0.94	0.96	0.62	0.67	**0.98**	0.97
Gleditsch$_{91}$	0.96	0.96	0.94	0.96	0.63	0.71	**0.98**	0.97
Gleditsch$_{92}$	0.96	0.96	0.94	0.96	0.63	0.70	**0.98**	0.97
Gleditsch$_{93}$	0.96	0.96	0.94	0.96	0.64	0.69	**0.98**	0.97
Gleditsch$_{94}$	0.96	0.96	0.94	0.96	0.64	0.67	**0.98**	0.97
Gleditsch$_{95}$	0.96	0.96	0.94	0.96	0.61	0.68	**0.98**	0.97
Gleditsch$_{96}$	0.96	0.96	0.94	0.96	0.60	0.65	**0.98**	0.97
Gleditsch$_{97}$	0.96	0.96	0.94	0.96	0.60	0.63	**0.98**	0.97
Gleditsch$_{98}$	0.96	0.96	0.94	0.96	0.61	0.63	**0.98**	0.97
Gleditsch$_{99}$	0.96	0.96	0.94	0.96	0.61	0.62	**0.98**	0.97
Gleditsch$_{00}$	0.96	0.96	0.95	0.96	0.60	0.63	**0.97**	**0.97**
BACI$_{07}$	0.93	0.95	0.92	0.95	0.43	0.43	**0.97**	**0.97**
BACI$_{08}$	0.92	0.94	0.91	0.94	0.40	0.40	**0.97**	**0.97**
BACI$_{09}$	0.93	0.95	0.92	0.95	0.43	0.43	**0.97**	**0.97**
BACI$_{10}$	0.92	0.94	0.91	0.95	0.41	0.41	**0.97**	**0.97**
BACI$_{11}$	0.92	0.94	0.90	0.94	0.38	0.39	**0.97**	**0.97**
BACI$_{12}$	0.91	0.94	0.90	0.94	0.38	0.38	**0.97**	**0.97**
BACI$_{13}$	0.91	0.93	0.90	0.93	0.37	0.38	**0.97**	**0.97**
BACI$_{14}$	0.91	0.93	0.90	0.94	0.38	0.39	**0.97**	**0.97**
BACI$_{15}$	0.91	0.94	0.91	0.94	0.41	0.41	**0.97**	**0.97**
BACI$_{16}$	0.92	0.94	0.91	0.94	0.40	0.41	**0.97**	**0.97**
BACI$_{17}$	0.91	0.94	0.91	0.94	0.39	0.39	**0.97**	**0.97**
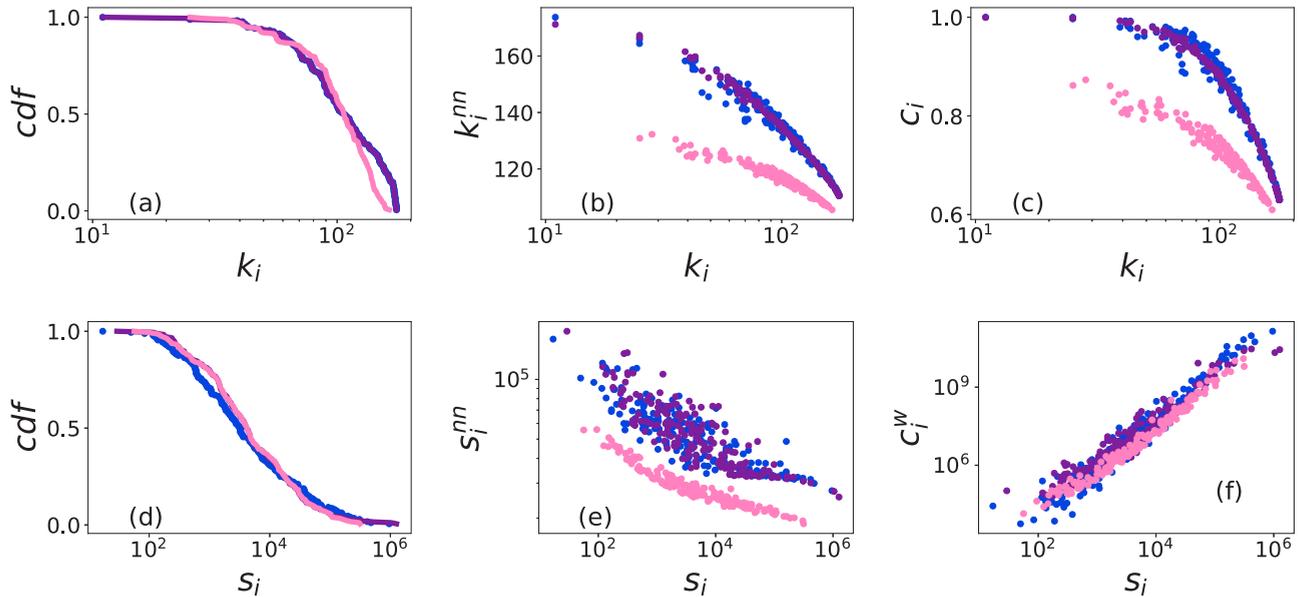
FIG. 5. Performance of the negative binomial model vs the performance of the ME model described by the Hamiltonian $H_{(2)}$, in reproducing: (a) the degree distribution; (b) the ANND; (c) the BCC; (d) the strength distribution; (e) the ANNS; and (f) the WCC. While a look at Table II suggests the negative binomial to be the econometric model performing best, this is definitely not the case as explicitly plotting its predictions against the empirical trends reveals. Notice that in (d) the distributions of strength induced by $H_{(2)}$ and by the negative binomial model overlap for a large range of values whereas, for other statistics, $H_{(2)}$ produces the best fit—confirming the importance of a good topological estimation. Empirical points are indicated •; econometric and ME models are indicated as follows: • - Negative Binomial; • - $H_{(2)}$. Results refer to the year 2000 of the dataset curated by Gleditsch [63].

Some of the problems of purely econometric models are solved by looking at a different class of statistical models, i.e., the physics-inspired ones. In particular, the model described by the Hamiltonian $H_{(2)} = \sum_i \theta_i k_i + \psi_0 W + \sum_{i<j} \psi_{ij} w_{ij}$ provides a very accurate reconstruction while being favoured by information criteria. Remarkably, although it is defined by $N + 1$ purely topological constraints, both AIC and BIC reveal that the latter are "irreducible", i.e., necessary to provide a satisfactory explanation of the network generating process. For the sake of comparison, Fig. 5 explicitly shows the performance of the models favoured by the adopted information criteria (i.e., the negative binomial model and its zero-inflated version) with that of the ME model described by $H_{(2)}$: It is evident that the ME model outperforms the purely econometric ones, still achieving a good ranking.

Looking at the class of ME models in more detail, our analysis indicates that the information carried by the strengths is not as "fundamental" as the one carried by the degrees: This is evident upon considering that (1) the UECM is always disfavoured with respect to the models just constraining the degrees, (2) the second best performing ME model is (always) the two step one, defined by a first purely topological step, accounting for the degrees, followed by an econometric-wise estimation of the weights. On top of that, we explicitly notice that structural topological information (e.g., the one provided by the link density or the degree sequence) usually correlates with node-specific economic covariates; hence, excluding such information from the model may lead to the so-called "omitted variable" bias. As proved by AIC and BIC, ME models represent the best compromise between goodness-of-fit and parsimony: In fact, they allow for structural information to be included, keeping the aforementioned type of bias low while leading to a better description of economic systems than that provided by traditional econometric models.

Our findings may indicate a route towards reconciling econometric and maximum-entropy network approaches, suggesting how to build a model that combines the pros of both: the importance of purely structural information (highlighted by physics-inspired models) can be accounted for by a model with a first step that is purely topological in nature (notice, in fact, that the TSF model is disfavoured with respect to the TS one) and a second step that takes care of estimating the weighted structure. Such an estimation can rest upon econometric considerations, driving the re-parametrization of otherwise purely structural models.

APPENDIX A: ESTIMATING THE GM PARAMETERS

Here, we consider two different specifications of the GM, i.e.,

$$\langle w_{ij} \rangle_{\text{GM}}^{(1)} = \rho(\omega_i \omega_j) d_{ij}^{-1} \qquad \text{(A1)}$$

and

$$\langle w_{ij} \rangle_{\text{GM}}^{(2)} = \rho(\omega_i \omega_j)^\beta d_{ij}^\gamma. \tag{A2}$$

The parameters appearing in both specifications of the GM can be estimated by implementing a nonlinear least squares (NLS) regression, i.e., by solving the optimization problem

$$\arg \min_\theta \left\{ \sum_{i<j} [w_{ij} - \langle w_{ij} \rangle_{\text{GM}}]^2 \right\}, \tag{A3}$$

which, in turn, translates into solving the equation

$$\sum_{i<j} [w_{ij} - \langle w_{ij} \rangle_{\text{GM}}] \frac{\partial \langle w_{ij} \rangle_{\text{GM}}}{\partial \theta_i} = 0, \quad \forall i. \tag{A4}$$

However, such a procedure is known to produce biased estimations. In fact, it leads to the set of conditions

$$\sum_{i<j} \left[w_{ij} - \langle w_{ij} \rangle_{\text{GM}}^{(1)} \right] \langle w_{ij} \rangle_{\text{GM}} = 0 \tag{A5}$$

for the first GM specification and

$$\sum_{i<j} \left[w_{ij} - \langle w_{ij} \rangle_{\text{GM}}^{(2)} \right] \langle w_{ij} \rangle_{\text{GM}} = 0$$

$$\sum_{i<j} \left[w_{ij} - \langle w_{ij} \rangle_{\text{GM}}^{(2)} \right] \langle w_{ij} \rangle_{\text{GM}} \ln(\omega_i \omega_j) = 0 \tag{A6}$$

$$\sum_{i<j} \left[w_{ij} - \langle w_{ij} \rangle_{\text{GM}}^{(2)} \right] \langle w_{ij} \rangle_{\text{GM}} \ln d_{ij} = 0$$

for the second GM specification. Since the conditions above are known to "weigh" more larger weights, Silva and Tenreyro [37] propose to employ a Poisson pseudo-maximum likelihood (PPML) estimator, leading to the set of conditions

$$\sum_{i<j} \left[w_{ij} - \langle w_{ij} \rangle_{\text{GM}}^{(1)} \right] = 0 \tag{A7}$$

for the first GM specification and

$$\sum_{i<j} \left[w_{ij} - \langle w_{ij} \rangle_{\text{GM}}^{(2)} \right] = 0$$

$$\sum_{i<j} \left[w_{ij} - \langle w_{ij} \rangle_{\text{GM}}^{(2)} \right] \ln(\omega_i \omega_j) = 0 \tag{A8}$$

$$\sum_{i<j} \left[w_{ij} - \langle w_{ij} \rangle_{\text{GM}}^{(2)} \right] \ln d_{ij} = 0.$$

for the second GM specification. Remarkably, the PPML estimator lets the purely topological condition $W = \sum_{i<j} w_{ij} = \sum_{i<j} \langle w_{ij} \rangle_{\text{GM}} = \langle W \rangle_{\text{GM}}$ (i.e., the preservation of the total weight) to be recovered for both GM specifications.

APPENDIX B: ECONOMETRIC MODELS

This Appendix is devoted to the detailed description of the econometric models considered in the present paper. When coming to estimate the parameters entering into the definition of any of the econometric models considered above, we invoke the maximum-of-the-likelihood principle, prescribing to maximize the function

$$\mathcal{L} = \ln Q(\mathbf{W}) \tag{B1}$$

where the optimization is carried out with respect to the set of parameters characterizing each specific model.

In the present contribution we have considered undirected networks, hence the parameters associated to node-specific regressors of the same economic variable (e.g., the GDPs of the country of origin and destination) are equal. In such a framework we have employed the specifications $z_{ij} = \rho(\omega_i \omega_j)^\beta d_{ij}^\gamma$ and $G_{ij} = \delta(\omega_i \omega_j)$ where $\omega_i = \frac{\text{GDP}_i}{\overline{\text{GDP}}}$, i.e., the GDP of each country is divided by the mean value of all GDPs. We would also like to stress that the parameters entering into the definition of z_{ij} are equal to those employed for the standard GM specification, i.e., $z_{ij} = e^{X \cdot \theta}$, as evident upon taking as regressors the natural logarithm of the GDPs and that of the geographic distance, i.e.,

$$z_{ij} = e^{\underline{X} \cdot \underline{\theta}} = e^{\beta \ln \omega_i + \beta \ln \omega_j + \gamma \ln d_{ij} + c} = \rho(\omega_i \omega_j)^\beta d_{ij}^\gamma \tag{B2}$$

having posed $c = \ln \rho$.

1. Poisson model

The probability mass function of the Poisson model reads

$$q_{ij}^{\text{Pois}}(w_{ij}) = \frac{z_{ij}^{w_{ij}} e^{-z_{ij}}}{w_{ij}!}, \tag{B3}$$

where $z_{ij} = \rho(\omega_i \omega_j)^\beta d_{ij}^\gamma$. In what follows, we will substitute $w_{ij}!$ with $\Gamma[w_{ij} + 1]$, as routinely done in the packages for solving econometric models. Its log-likelihood reads

$$\mathcal{L}_{\text{Pois}} = \sum_{i<j} [w_{ij} \ln z_{ij} - z_{ij} - \ln \Gamma[w_{ij} + 1]] \tag{B4}$$

whose optimization leads to the set of equations

$$\sum_{i<j} \left[\frac{w_{ij}}{z_{ij}} - 1 \right] \frac{\partial z_{ij}}{\partial \theta_i} = 0, \quad \forall i \tag{B5}$$

reading, more explicitly,

$$\sum_{i<j} [w_{ij} - \langle w_{ij} \rangle_{\text{Pois}}] = 0$$

$$\sum_{i<j} [w_{ij} - \langle w_{ij} \rangle_{\text{Pois}}] \ln(\omega_i \omega_j) = 0$$

$$\sum_{i<j} [w_{ij} - \langle w_{ij} \rangle_{\text{Pois}}] \ln d_{ij} = 0. \tag{B6}$$

Notice that the set of equations above coincides with Eqs. (A8). The Poisson model remains the most used one because it ensures that a network total weight is reproduced—a desirable feature to correctly estimate the weights of a network, as also observed for the "plain" GM.

2. Negative binomial model

The probability mass function of the negative binomial model reads

$$q_{ij}^{\text{NB}}(w_{ij}) = \binom{m + w_{ij} - 1}{w_{ij}} \left(\frac{1}{1 + \alpha z_{ij}} \right)^m \left(\frac{\alpha z_{ij}}{1 + \alpha z_{ij}} \right)^{w_{ij}} \tag{B7}$$

where $\alpha = m^{-1}$ to handle overdispersion and $z_{ij} = \rho(\omega_i \omega_j)^\beta d_{ij}^\gamma$. By replacing each binomial coefficient with the

corresponding Gamma function, one recovers the expression

$$\mathcal{L}_{\text{NB}} = \sum_{i<j}[w_{ij}\ln(\alpha z_{ij}) - (m + w_{ij})\ln(1 + \alpha z_{ij})$$
$$+ \ln\Gamma[m + w_{ij}] - \ln\Gamma[w_{ij} + 1] - \ln\Gamma[m]]; \quad \text{(B8)}$$

its optimization leads to the set of equations

$$\sum_{i<j}\left[\frac{w_{ij} - z_{ij}}{z_{ij}(1 + \alpha z_{ij})}\right]\frac{\partial z_{ij}}{\partial\theta_i} = 0, \quad \forall i \quad \text{(B9)}$$

and

$$\sum_{i<j}\left[\frac{w_{ij} - z_{ij}}{\alpha(1 + \alpha z_{ij})} - m^2\left(\frac{\Gamma'[m + w_{ij}]}{\Gamma[m + w_{ij}]} - \frac{\Gamma'[m]}{\Gamma[m]}\right)\right] = 0, \quad \text{(B10)}$$

reading, more explicitly,

$$\sum_{i<j}\left[\frac{w_{ij} - \langle w_{ij}\rangle_{\text{NB}}}{\text{Var}_{\text{NB}}[w_{ij}]}\right] = 0$$

$$\sum_{i<j}\left[\frac{w_{ij} - \langle w_{ij}\rangle_{\text{NB}}}{\text{Var}_{\text{NB}}[w_{ij}]}\right]\ln(\omega_i\omega_j) = 0$$

$$\sum_{i<j}\left[\frac{w_{ij} - \langle w_{ij}\rangle_{\text{NB}}}{\text{Var}_{\text{NB}}[w_{ij}]}\right]\ln d_{ij} = 0$$

$$\sum_{i<j}\left[\frac{w_{ij} - \langle w_{ij}\rangle_{\text{NB}}}{\alpha(1 + \alpha z_{ij})} - m^2\left(\frac{\Gamma'[m + w_{ij}]}{\Gamma[m + w_{ij}]} - \frac{\Gamma'[m]}{\Gamma[m]}\right)\right] = 0. \quad \text{(B11)}$$

Notice that the negative binomial model does not constrain the total weight of a network, whence its bad performance in reproducing the other weighted structural properties.

3. Zero-inflated Poisson model

The zero-inflated version of the Poisson model is defined by a functional form reading

$$Q(\mathbf{W}) = \prod_{i<j}q_{ij}(w_{ij})$$
$$= \prod_{i<j}p_{ij}^{a_{ij}}(1 - p_{ij})^{1-a_{ij}} \cdot q_{ij}(w_{ij}|a_{ij})$$
$$= P(\mathbf{A})Q(\mathbf{W}|\mathbf{A}) \quad \text{(B12)}$$

and inducing the following log-likelihood:

$$\mathcal{L} = \ln Q(\mathbf{W}) = \ln P(\mathbf{A}) + \ln Q(\mathbf{W}|\mathbf{A})$$
$$= \sum_{i<j}[a_{ij}\ln p_{ij} + (1 - a_{ij})\ln(1 - p_{ij})$$
$$+ \ln q_{ij}(w_{ij}|a_{ij})] \quad \text{(B13)}$$

where

$$p_{ij}^{\text{ZIP}} = \frac{G_{ij}}{1 + G_{ij}}(1 - e^{-z_{ij}}), \quad \text{(B14)}$$

$$q_{ij}^{\text{ZIP}}(w_{ij}|a_{ij}) = \left[\frac{z_{ij}^{w_{ij}}e^{-z_{ij}}}{(1 - e^{-z_{ij}})w_{ij}!}\right]^{a_{ij}} \quad \text{(B15)}$$

with $z_{ij} = \rho(\omega_i\omega_j)^\beta d_{ij}^\gamma$ and $G_{ij} = \delta\omega_i\omega_j$; hence,

$$\mathcal{L}_{\text{ZIP}} = \sum_{i<j}[a_{ij}\ln G_{ij} - a_{ij}\ln(1 + G_{ij}e^{-z_{ij}})$$
$$+ \ln(1 + G_{ij}e^{-z_{ij}}) - \ln(1 + G_{ij})$$
$$+ w_{ij}\ln z_{ij} - a_{ij}z_{ij} - a_{ij}\ln\Gamma[w_{ij} + 1]]. \quad \text{(B16)}$$

Its optimization leads to the set of equations

$$\sum_{i<j}\left[\frac{a_{ij} - p_{ij}^{\text{ZIP}}}{1 + G_{ij}e^{-z_{ij}}}\right] = 0$$

$$\sum_{i<j}\left[w_{ij} - \left(\frac{a_{ij} + G_{ij}e^{-z_{ij}}}{1 + G_{ij}e^{-z_{ij}}}\right)z_{ij}\right] = 0$$

$$\sum_{i<j}\left[w_{ij} - \left(\frac{a_{ij} + G_{ij}e^{-z_{ij}}}{1 + G_{ij}e^{-z_{ij}}}\right)z_{ij}\right]\ln(\omega_i\omega_j) = 0$$

$$\sum_{i<j}\left[w_{ij} - \left(\frac{a_{ij} + G_{ij}e^{-z_{ij}}}{1 + G_{ij}e^{-z_{ij}}}\right)z_{ij}\right]\ln d_{ij} = 0 \quad \text{(B17)}$$

with a clear meaning of the symbols.

4. Zero-inflated negative binomial model

As for the ZIP model, the zero-inflated version of the negative binomial model induces a log-likelihood reading

$$\mathcal{L} = \ln Q(\mathbf{W}) = \ln P(\mathbf{A}) + \ln Q(\mathbf{W}|\mathbf{A})$$
$$= \sum_{i<j}[a_{ij}\ln p_{ij} + (1 - a_{ij})\ln(1 - p_{ij})$$
$$+ \ln q_{ij}(w_{ij}|a_{ij})] \quad \text{(B18)}$$

where

$$p_{ij}^{\text{ZINB}} = \frac{G_{ij}}{1 + G_{ij}}(1 - \tau_{ij}), \quad \text{(B19)}$$

$$q_{ij}^{\text{ZINB}}(w_{ij}|a_{ij}) = \binom{m + w_{ij} - 1}{w_{ij}}^{a_{ij}}\left(\frac{\tau_{ij}}{1 - \tau_{ij}}\right)^{a_{ij}}\left(\frac{\alpha z_{ij}}{1 + \alpha z_{ij}}\right)^{w_{ij}}$$
$$= \left[\binom{m + w_{ij} - 1}{w_{ij}}\left(\frac{1}{1 - \tau_{ij}}\right)\left(\frac{1}{1 + \alpha z_{ij}}\right)^m\left(\frac{\alpha z_{ij}}{1 + \alpha z_{ij}}\right)^{w_{ij}}\right]^{a_{ij}} \quad \text{(B20)}$$

with $\alpha = m^{-1}$, $\tau_{ij} = (\frac{1}{1+\alpha z_{ij}})^m$, $z_{ij} = \rho(\omega_i\omega_j)^\beta d_{ij}^\gamma$ and $G_{ij} = \delta\omega_i\omega_j$. Hence,

$$\mathcal{L}_{\text{ZINB}} = \sum_{i<j}[a_{ij}\ln G_{ij} - a_{ij}\ln(1+G_{ij}\tau_{ij}) + \ln(1+G_{ij}\tau_{ij}) - \ln(1+G_{ij}) + w_{ij}\ln(\alpha z_{ij}) - w_{ij}\ln(1+\alpha z_{ij})$$
$$-ma_{ij}\ln(1+\alpha z_{ij}) + a_{ij}\ln\Gamma[m+w_{ij}] - a_{ij}\ln\Gamma[w_{ij}+1] - a_{ij}\ln\Gamma[m]] \tag{B21}$$

whose optimization leads to the set of equations

$$\sum_{i<j}\left[\frac{a_{ij} - p_{ij}^{\text{ZINB}}}{1+G_{ij}\tau_{ij}}\right] = 0$$

$$\sum_{i<j}\left[w_{ij} - \left(\frac{a_{ij}+G_{ij}\tau_{ij}}{1+G_{ij}\tau_{ij}}\right)\right]\left(\frac{z_{ij}}{1+\alpha z_{ij}}\right) = 0$$

$$\sum_{i<j}\left[w_{ij} - \left(\frac{a_{ij}+G_{ij}\tau_{ij}}{1+G_{ij}\tau_{ij}}\right)\right]\left(\frac{z_{ij}}{1+\alpha z_{ij}}\right)\ln(\omega_i\omega_j) = 0 \tag{B22}$$

$$\sum_{i<j}\left[w_{ij} - \left(\frac{a_{ij}+G_{ij}\tau_{ij}}{1+G_{ij}\tau_{ij}}\right)\right]\left(\frac{z_{ij}}{1+\alpha z_{ij}}\right)\ln d_{ij} = 0$$

$$\sum_{i<j}\left[\left(\frac{a_{ij}+G_{ij}\tau_{ij}}{1+G_{ij}\tau_{ij}}\right)\left(m^2\ln(1+\alpha z_{ij}) - \frac{mz_{ij}}{1+\alpha z_{ij}}\right) + \frac{w_{ij}}{\alpha(1+\alpha z_{ij})} - m^2 a_{ij}\left(\frac{\Gamma'[m+w_{ij}]}{\Gamma[m+w_{ij}]} - \frac{\Gamma'[m]}{\Gamma[m]}\right)\right] = 0.$$

APPENDIX C: MAXIMUM-ENTROPY MODELS

This Appendix is devoted to the detailed description of the maximum-entropy models considered in the present paper. As for the econometric ones, the estimation of the set of parameters defining each model is carried out by maximizing the corresponding log-likelihood function, $\mathcal{L} = \ln Q(\mathbf{W})$.

ME models are derived by maximizing Shannon entropy under a suitable set of constraints. The generic probability mass function reads

$$Q(\mathbf{W}) = \prod_{i<j} q_{ij}(w_{ij})$$
$$= \prod_{i<j} p_{ij}^{a_{ij}}(1-p_{ij})^{1-a_{ij}} \cdot q_{ij}(w_{ij}|a_{ij})$$
$$= \prod_{i<j} p_{ij}^{a_{ij}}(1-p_{ij})^{1-a_{ij}} \cdot y_{ij}^{w_{ij}-a_{ij}}(1-y_{ij})^{a_{ij}}$$
$$= P(\mathbf{A})Q(\mathbf{W}|\mathbf{A}) \tag{C1}$$

and induces the log-likelihood

$$\mathcal{L} = \ln Q(\mathbf{W}) = \ln P(\mathbf{A}) + \ln Q(\mathbf{W}|\mathbf{A})$$
$$= \sum_{i<j}[a_{ij}\ln p_{ij} + (1-a_{ij})\ln(1-p_{ij})$$
$$+ \ln q_{ij}(w_{ij}|a_{ij})]. \tag{C2}$$

In order to turn maximum-entropy models into proper econometric ones, we have posed $\frac{y_{ij}}{1-y_{ij}} = z_{ij} = \rho(\omega_i\omega_j)^\beta d_{ij}^\gamma$, with $\omega_i = \frac{\text{GDP}_i}{\text{GDP}}$.

Let us start instantiating the ME formalism by considering the Hamiltonian

$$H_{(1)}(\mathbf{W}) = \theta_0 L + \psi_0 W + \sum_{i<j}\psi_{ij}w_{ij} \tag{C3}$$

that leads to

$$\mathcal{L}_{(1)} = L\ln x + W\ln y_0$$
$$+ \sum_{i<j}[w_{ij}\ln z_{ij} - w_{ij}\ln(1+z_{ij}) + \ln(1+z_{ij}-y_0 z_{ij})$$
$$- \ln(1+z_{ij}-y_0 z_{ij}+xy_0 z_{ij})] \tag{C4}$$

whose optimization leads to solve the following set of equations

$$\sum_{i<j}\left[a_{ij} - p_{ij}^{(1)}\right] = 0$$

$$\sum_{i<j}[w_{ij} - \langle w_{ij}\rangle_{(1)}] = 0$$

$$\sum_{i<j}[w_{ij} - \langle w_{ij}\rangle_{(1)}]\left(\frac{1}{1+z_{ij}}\right) = 0 \tag{C5}$$

$$\sum_{i<j}[w_{ij} - \langle w_{ij}\rangle_{(1)}]\left(\frac{\ln(\omega_i\omega_j)}{1+z_{ij}}\right) = 0$$

$$\sum_{i<j}[w_{ij} - \langle w_{ij}\rangle_{(1)}]\left(\frac{\ln d_{ij}}{1+z_{ij}}\right) = 0.$$

The first equation guaranteeing that the total number of links is preserved, i.e.,

$$L = \langle L \rangle = \sum_{i<j}\frac{xy_0 z_{ij}}{1+z_{ij}-y_0 z_{ij}+xy_0 z_{ij}} = \sum_{i<j}p_{ij}^{(1)} \tag{C6}$$

and the second equation guaranteeing that the total weight is preserved, i.e.,

$$W = \langle W \rangle = \sum_{i<j}\frac{p_{ij}^{(1)}(1+z_{ij})}{1+z_{ij}-y_0 z_{ij}} = \sum_{i<j}\langle w_{ij}\rangle_{(1)}. \tag{C7}$$

On the other hand, the Hamiltonian

$$H_{(2)}(\mathbf{W}) = \sum_i \theta_i k_i + \psi_0 W + \sum_{i<j} \psi_{ij} w_{ij} \qquad (C8)$$

leads to the expression

$$
\begin{aligned}
\mathcal{L}_{(2)} = & \sum_i k_i \ln x_i + W \ln y_0 \\
& + \sum_{i<j} [w_{ij} \ln z_{ij} - w_{ij} \ln(1 + z_{ij}) + \ln(1 + z_{ij} - y_0 z_{ij}) \\
& - \ln(1 + z_{ij} - y_0 z_{ij} + x_i x_j y_0 z_{ij})]
\end{aligned}
\qquad (C9)
$$

whose optimization requires to solve the set of equations below

$$\sum_{j(\neq i)} \left[a_{ij} - p_{ij}^{(2)} \right] = 0, \quad \forall\, i$$

$$\sum_{i<j} [w_{ij} - \langle w_{ij} \rangle_{(2)}] = 0$$

$$\sum_{i<j} [w_{ij} - \langle w_{ij} \rangle_{(2)}] \left(\frac{1}{1 + z_{ij}} \right) = 0 \qquad (C10)$$

$$\sum_{i<j} [w_{ij} - \langle w_{ij} \rangle_{(2)}] \left(\frac{\ln(\omega_i \omega_j)}{1 + z_{ij}} \right) = 0$$

$$\sum_{i<j} [w_{ij} - \langle w_{ij} \rangle_{(2)}] \left(\frac{\ln d_{ij}}{1 + z_{ij}} \right) = 0.$$

The first set of N equations guaranteeing that each degree is preserved, i.e.,

$$k_i = \langle k_i \rangle = \sum_{j(\neq i)} \frac{x_i x_j y_0 z_{ij}}{1 + z_{ij} - y_0 z_{ij} + x_i x_j y_0 z_{ij}} = \sum_{j(\neq i)} p_{ij}^{(2)}$$

$$\qquad (C11)$$

and the $(N+1)$th equation guaranteeing that the total weight is preserved, i.e.,

$$W = \langle W \rangle = \sum_{i<j} \frac{p_{ij}^{(2)}(1 + z_{ij})}{1 + z_{ij} - y_0 z_{ij}} = \sum_{i<j} \langle w_{ij} \rangle_{(2)}. \qquad (C12)$$

Writing a probability mass function in the factorized form $Q(\mathbf{W}) = P(\mathbf{A})Q(\mathbf{W}|\mathbf{A})$ allows us to design two-step models, i.e., network models whose binary estimation step can be defined independently from the weighted one. To this aim, let us combine the probability distribution of the undirected binary configuration model, inducing a single-link probability coefficient reading $p_{ij}^{\mathrm{UBCM}} = \frac{x_i x_j}{1 + x_i x_j}$, with the usual weighted, conditional one, i.e., $q_{ij}(w_{ij}|a_{ij}) = (\frac{y_0 z_{ij}}{1 + z_{ij}})^{w_{ij} - a_{ij}} \cdot (\frac{1 + z_{ij} - y_0 z_{ij}}{1 + z_{ij}})^{a_{ij}}$. As a result, one obtains

$$
\begin{aligned}
\mathcal{L}_{\mathrm{TS}} = & \sum_i k_i \ln x_i + (W - L) \ln y_0 \\
& + \sum_{i<j} [- \ln(1 + x_i x_j) + (w_{ij} - a_{ij}) \ln z_{ij} \\
& - w_{ij} \ln(1 + z_{ij}) + a_{ij} \ln(1 + z_{ij} - y_0 z_{ij})] \qquad (C13)
\end{aligned}
$$

whose optimization requires to solve the set of equations below

$$\sum_{j(\neq i)} \left[a_{ij} - p_{ij}^{\mathrm{TS}} \right] = 0, \quad \forall\, i$$

$$\sum_{i<j} [w_{ij} - \langle w_{ij} \rangle_{\mathrm{TS}}] = 0$$

$$\sum_{i<j} [w_{ij} - \langle w_{ij} \rangle_{\mathrm{TS}}] \left(\frac{1}{1 + z_{ij}} \right) = 0 \qquad (C14)$$

$$\sum_{i<j} [w_{ij} - \langle w_{ij} \rangle_{\mathrm{TS}}] \left(\frac{\ln(\omega_i \omega_j)}{1 + z_{ij}} \right) = 0$$

$$\sum_{i<j} [w_{ij} - \langle w_{ij} \rangle_{\mathrm{TS}}] \left(\frac{\ln d_{ij}}{1 + z_{ij}} \right) = 0.$$

The first set of N equations guaranteeing that each degree is preserved, i.e.,

$$k_i = \langle k_i \rangle = \sum_{j(\neq i)} \frac{x_i x_j}{1 + x_i x_j} = \sum_{j(\neq i)} p_{ij}^{\mathrm{TS}} = \sum_{j(\neq i)} p_{ij}^{\mathrm{UBCM}} \quad (C15)$$

and the $(N+1)$th equation guaranteeing that the total conditional weight is preserved, i.e.,

$$W = \langle W \rangle = \sum_{i<j} \frac{a_{ij}(1 + z_{ij})}{1 + z_{ij} - y_0 z_{ij}} = \sum_{i<j} \langle w_{ij} \rangle_{\mathrm{TS}}. \qquad (C16)$$

The amount of structural information to constrain can be further reduced by imagining a two-step model whose binary estimation step encodes a larger amount of "economic" information. To this aim, let us combine the probability distribution of the density-corrected Gravity Model, inducing a single-link probability coefficient reading $p_{ij}^{\mathrm{TSF}} = \frac{G_{ij}}{1 + G_{ij}}$, with the usual weighted, conditional one, i.e., $q_{ij}(w_{ij}|a_{ij} = 1) = (\frac{y_0 z_{ij}}{1 + z_{ij}})^{w_{ij} - a_{ij}} \cdot (\frac{1 + z_{ij} - y_0 z_{ij}}{1 + z_{ij}})^{a_{ij}}$. As a result, one obtains

$$
\begin{aligned}
\mathcal{L}_{\mathrm{TSF}} = & \sum_{i<j} [a_{ij} \ln G_{ij} - \ln(1 + G_{ij}) + (W - L) \ln y_0 \\
& + (w_{ij} - a_{ij}) \ln z_{ij} - w_{ij} \ln(1 + z_{ij}) \\
& + a_{ij} \ln(1 + z_{ij} - y_0 z_{ij})] \qquad (C17)
\end{aligned}
$$

where $G_{ij} = \delta \omega_i \omega_j$. Its optimization leads to the resolution of the following set of equations

$$\sum_{i<j} \left[a_{ij} - p_{ij}^{\mathrm{TSF}} \right] = 0$$

$$\sum_{i<j} [w_{ij} - \langle w_{ij} \rangle_{\mathrm{TSF}}] = 0$$

$$\sum_{i<j} [w_{ij} - \langle w_{ij} \rangle_{\mathrm{TSF}}] \left(\frac{1}{1 + z_{ij}} \right) = 0 \qquad (C18)$$

$$\sum_{i<j} [w_{ij} - \langle w_{ij} \rangle_{\mathrm{TSF}}] \left(\frac{\ln(\omega_i \omega_j)}{1 + z_{ij}} \right) = 0$$

$$\sum_{i<j} [w_{ij} - \langle w_{ij} \rangle_{\mathrm{TSF}}] \left(\frac{\ln d_{ij}}{1 + z_{ij}} \right) = 0.$$

The first equation guaranteeing that the total number of links is preserved, i.e.,

$$L = \langle L \rangle = \sum_{j(\neq i)} \frac{G_{ij}}{1 + G_{ij}} = \sum_{j(\neq i)} p_{ij}^{\text{TSF}} = \sum_{j(\neq i)} p_{ij}^{\text{dcGM}} \quad \text{(C19)}$$

and the second equation guaranteeing that the total conditional weight is preserved, i.e.,

$$W = \langle W \rangle = \sum_{i<j} \frac{a_{ij}(1 + z_{ij})}{1 + z_{ij} - y_0 z_{ij}} = \sum_{i<j} \langle w_{ij} \rangle_{\text{TSF}}. \quad \text{(C20)}$$

[1] F. Schweitzer, G. Fagiolo, D. Sornette, and F. Vega-Redondo, Economic networks: The new challenges, Science **325**, 422 (2009).

[2] G. Fagiolo, The international-trade network: Gravity equations and topological properties, J. Econ. Interact. Coord. **5**, 1 (2010).

[3] F. Saracco, R. Di Clemente, A. Gabrielli, and T. Squartini, Detecting early signs of the 20072008 crisis in the world trade, Sci. Rep. **6**, 30286 (2016).

[4] T. Squartini and D. Garlaschelli, Stationarity, non-stationarity and early warning signals in economic networks, J. Complex Netw. **3**, 1 (2015).

[5] S. Schiavo, J. Reyes, and G. Fagiolo, International trade and financial integration: A weighted network analysis, Quant. Financ. **10**, 389 (2010).

[6] T. Squartini, I. van Lelyveld, and D. Garlaschell, Early-warning signals of topological collapse in interbank networks, Sci. Rep. **3**, (3357) (2013).

[7] R. Kali and J. Reyes, Financial contagion on the international trade network, Econ. Inq. **48**, 1072 (2010).

[8] R. Kali and J. Reyes, The architecture of globalization: A network approach to international economic integration, J. Int. Bus. Stud. **38**, 595 (2007).

[9] M. Starnini, M. Boguña, and M. Á. Serrano, The interconnected wealth of nations: Shock propagation on global trade-investment multiplex networks, Sci. Rep. **9**, 13079 (2019).

[10] M. Barigozzi, G. Fagiolo, and D. Garlaschelli, Multinetwork of international trade: A commodity-specific analysis, Phys. Rev. E **81**, 046104 (2010).

[11] A. Fronczak, Structural Hamiltonian of the international trade network, Acta Phys. Pol. B Proc. Suppl. **5**, 31 (2012).

[12] A. Fronczak and P. Fronczak, Statistical mechanics of the international trade network, Phys. Rev. E **85**, 056113 (2012).

[13] M. Á. Serrano and M. Boguña, Topology of the world trade web, Phys. Rev. E **68**, 015101(R) (2003).

[14] D. Garlaschelli and M. I. Loffredo, Fitness-Dependent Topological Properties of the World Trade Web, Phys. Rev. Lett. **93**, 188701 (2004).

[15] D. Garlaschelli and M. I. Loffredo, Structure and evolution of the world trade network, Physica A **355**, 138 (2005).

[16] G. Fagiolo, J. Reyes, and S. Schiavo, The evolution of the world trade web: A weighted-network analysis, J. Evol. Econ. **20**, 479 (2010).

[17] G. Fagiolo, J. Reyes, and S. Schiavo, On the topological properties of the world trade web: A weighted network analysis, Physica A **387**, 3868 (2008).

[18] G. Fagiolo, J. Reyes, and S. Schiavo, World-trade web: Topological properties, dynamics, and evolution, Phys. Rev. E **79**, 036115 (2009).

[19] S. Vitali, J. B. Glattfelder, and S. Battiston, The network of global corporate control, PLoS ONE **6**, e25995 (2011).

[20] J. van Lidth de Jeude, R. Di Clemente, G. Caldarelli, F. Saracco, and T. Squartini, Reconstructing mesoscale network structures, Complexity **2019**, 5120581 (2019).

[21] J. Tinbergen, *Shaping the World Economy: Suggestions for an International Economic Policy* (The Twentieth Century Fund, New York, 1962).

[22] J. Anderson, The gravity model, Annu. Rev. Econ. **3**, 133 (2011).

[23] A. Almog, R. Bird, and D. Garlaschelli, Enhanced Gravity Model of trade: reconciling macroeconomic and network models, Front. Phys. **7**, 00055 (2019).

[24] P. van Bergeijk and S. Brakman, *The Gravity Model in International Trade*, (Cambridge University Press, Cambridge, 2010).

[25] L. De Benedictis and D. Taglioni, *The Gravity Model in International Trade* (Springer, New York, 2011) p. 55-89.

[26] E. G. Ravenstein, The laws of migration, J. R. Stat. Soc. **52**, 241 (1889).

[27] G. Fagiolo and M. Mastrorillo, International migration network: Topology and modeling, Phys. Rev. E **88**, 012812 (2013).

[28] G. K. Zipf, The $P_1 P_2/D$ hypothesis: On the intercity movement of persons, Am. Sociol. Rev. **11**, 677 (1946).

[29] D. Balcan, V. Colizza, B. Gonçalves, H. Hu, J. J. Ramasco, A. Vespignani, Multiscale mobility networks and the spatial spreading of infectious diseases, Proc. Natl. Acad. Sci. USA **106**, 21484 (2009).

[30] W.-S. Jung, F. Wang, H. E. Stanley, Gravity model in the Korean highway, Europhys. Lett. **81**, 48005 (2008).

[31] G. Krings, F. Calabrese, C. Ratti, and V. D. Blondel, Urban gravity: A model for inter-city telecommunication flows, J. Stat. Mech. (2009) L07003.

[32] L.-L. Ma, C. Ma, H.-F. Zhang, and B.-H. Wang, Identifying influential spreaders in complex networks based on gravity formula, Physica A **451**, 205 (2016).

[33] Z. Li, T. Ren, X. Ma, S. Liu, Y. Zhang, and T. Zhou, Identifying influential spreaders by gravity model, Sci. Rep. **9**, 8387 (2019).

[34] J. Eaton and A. Tamura, Bilateralism and regionalism in Japanese and U. S. trade and direct foreign investment patterns, J. Jpn. Int. Econ. **8**, 478 (1994).

[35] W. J. Martin and C. S. Pham, Estimating the gravity model when zero trade flows are frequent and economically determined, Appl. Econ. **52**, 2766 (2020).

[36] E. Helpman, M. Melitz, and Y. Rubinstein, Estimating trade flows: Trading partners and trading volumes, Q. J. Econ. **123**, 441 (2008).

[37] J. M. C. Santos Silva and S. Tenreyro, The log of gravity, Rev. Econ. Stat **88**, 641 (2006).

[38] M. Burger, F. van Oort, and G. Linders, On the specification of the gravity model of trade: Zeros, excess zeros and zero-inflated estimation, Spat. Econ. Anal. **4**, 167 (2009).

[39] R. Winkelmann, *Econometric Analysis of Count Data* (Springer, Berlin, 2008).

[40] M. Dueñas and G. Fagiolo, Modeling the international-trade network: A gravity approach, J. Econ. Interact. Coord. **8**, 155 (2013).

[41] T. Squartini, G. Fagiolo, and D. Garlaschelli, Randomizing world trade. II. A weighted network analysis, Phys. Rev. E **84**, 046118 (2011).

[42] L. Bargigli, Statistical ensembles for economic networks, J. Stat. Phys. **155**, 810 (2014).

[43] T. Squartini and D. Garlaschelli, *Maximum-Entropy Networks: Pattern Detection, Network Reconstruction and Graph Combinatorics* (Springer International Publishing, New York, 2017).

[44] T. Squartini, G. Fagiolo, and D. Garlaschelli, Randomizing world trade. I. A binary network analysis, Phys. Rev. E **84**, 046117 (2011).

[45] T. Squartini and D. Garlaschelli, Analytical maximum-likelihood method to detect patterns in real networks, New J. Phys. **13**, 083001 (2011).

[46] R. Mastrandrea, T. Squartini, G. Fagiolo, and D. Garlaschelli, Enhanced reconstruction of weighted networks from strengths and degrees, New J. Phys. **16**, 043022 (2014).

[47] E. T. Jaynes, Information theory and statistical mechanics, Phys. Rev. **106**, 620 (1957).

[48] E. T. Jaynes, Information theory and statistical mechanics II, Phys. Rev. **108**, 171 (1957).

[49] E. T. Jaynes, On the rationale of maximum-entropy methods, Proc. IEEE **70**, 939 (1982).

[50] G. Cimini, T. Squartini, F. Saracco, D. Garlaschelli, A. Gabrielli, and G. Caldarelli, The statistical physics of real-world networks, Nat. Rev. Phys. **1**, 58 (2019).

[51] F. Parisi, T. Squartini, and D. Garlaschelli, A faster horse on a safer trail: Generalized inference for the efficient reconstruction of weighted networks, New J. Phys. **22**, 053053 (2020).

[52] G. Cimini, T. Squartini, and R. Mastrandrea, *Reconstructing networks* in *The structure and dynamics of complex networks*, Cambridge Elements series, edited by G. Caldarelli (Cambridge University Press, Cambridge, 2021).

[53] T. Squartini, A. Almog, G. Caldarelli, I. van Lelyveld, D. Garlaschelli, and G. Cimini, Enhanced capital-asset pricing model for the reconstruction of bipartite financial networks, Phys. Rev. E **96**, 032315 (2017).

[54] T. Squartini, R. Mastrandrea, and D. Garlaschelli, Unbiased sampling of network ensembles, New J. Phys. **17**, 023052 (2015).

[55] G. Fagiolo, T. Squartini, and D. Garlaschelli, Null models of economic networks: The case of the world trade web, J. Econ. Interact. Coord. **8**, 75 (2013).

[56] R. Mastrandrea, T. Squartini, G. Fagiolo, and D. Garlaschelli, Reconstructing the world trade multiplex: The role of intensive and extensive biases, Phys. Rev. E **90**, 062804 (2014).

[57] D. Garlaschelli, M. I. Loffredo, Generalized Bose-Fermi Statistics and Structural Correlations in Weighted Networks, Phys. Rev. Lett. **102**, 038701 (2009).

[58] A. Gabrielli, R. Mastrandrea, G. Caldarelli, and G. Cimini, Grand canonical ensemble of weighted networks, Phys. Rev. E **99**, 030301(R) (2019).

[59] G. Caldarelli, A. Capocci, P. De Los Rios, and M. A. Muñoz, Scale-Free Networks from Varying Vertex Intrinsic Fitness, Phys. Rev. Lett. **89**, 258702 (2002).

[60] A. Almog, T. Squartini, and D. Garlaschelli, A GDP-driven model for the binary and weighted structure of the International Trade Network, New J. Phys. **17**, 013009 (2015).

[61] A. Almog, T. Squartini, and D. Garlaschelli, The double role of GDP in shaping the structure of the International Trade Network, Int. J. Comput. Econ. Econom. **7**, 381 (2017).

[62] T. Squartini, G. Caldarelli, G. Cimini, A. Gabrielli, and D. Garlaschelli, Reconstruction methods for networks: The case of economic and financial systems, Phys. Rep. **757**, 1 (2018).

[63] K. Gleditsch, Expanded trade and GDP data, J. Conflict Resol. **46**, 712 (2002).

[64] K. Head, T. Mayer, and J. Ries, The erosion of colonial trade linkages after independence, J. Intl. Econ. **81**, 1 (2010).

[65] K. Head and T. Mayer, *Gravity Equations: Toolkit, Cookbook, Workhorse.*, in *Handbook of International Economics*, Vol. 4 (Elsevier, Amsterdam, 2014).

[66] In the Poisson case, one has that $\sigma^2_{\mathrm{Pois}}[w_{ij}] = z_{ij} = \langle w_{ij} \rangle_{\mathrm{Pois}}$; hence, variance cannot be adjusted independently from the mean. In the negative binomial case, instead, $\sigma^2_{\mathrm{NB}}[w_{ij}] = z_{ij}(1 + \alpha z_{ij}) = z_{ij}(1 + \frac{z_{ij}}{m}) = \langle w_{ij} \rangle_{\mathrm{NB}}(1 + \alpha \langle w_{ij} \rangle_{\mathrm{NB}}) = \langle w_{ij} \rangle_{\mathrm{NB}}(1 + \frac{\langle w_{ij} \rangle_{\mathrm{NB}}}{m})$ and the variance can be increased to overcome the problem of overestimating the link density.

[67] J. M. Hilbe, *Negative Binomial Regression* (Cambridge University Press, Cambridge, 2011).

[68] G. Cimini, T. Squartini, D. Garlaschelli, and A. Gabrielli, Systemic risk analysis on reconstructed economic and financial networks, Sci. Rep. **5**, 15758 (2015).

[69] A. Wahid-Ul-Ashraf, M. Budka, and K. Musial, How to predict social relationships - Physics-inspired approach to link prediction, Physica A **523**, 1110 (2019).