A faster horse on a safer trail: generalized inference for the efficient reconstruction of weighted networks

F. Parisi, T. Squartini, and D. Garlaschelli

1IMT School for Advanced Studies Lucca
2Lorentz Institute for Theoretical Physics, Leiden University

The knowledge of the structure of a financial network gives valuable information for the estimation of systemic risk and other important properties. However, since financial data are typically subject to confidentiality, network reconstruction techniques become necessary to infer both the presence of connections (i.e. the topology) and their intensity (i.e. the link weights) from partial information. Recently, various horse races have been conducted to compare the performance of these methods. These comparisons were based on arbitrarily chosen metrics of network similarity. Here we establish a generalised likelihood approach to the rigorous definition, estimation and comparison of methods of reconstruction of weighted networks. The crucial ingredient is the possibility of separately constraining the topology and the link weights to follow certain “tight” empirical relationships that connect them to the observed marginals. We find that a statistically rigorous estimation of the probability of each link weight should incorporate the whole distribution of reconstructed network topologies, and we provide a general procedure to do this. Our results indicate that the best method is obtained by “dressing” the best-performing method available to date with a certain exponential distribution of link weights. The method is fast, unbiased and accurate and reproduces the empirical networks with highest generalised likelihood.

Network reconstruction is an active field of research within the broader field of complex networks [20]. Addressing the network reconstruction problem means facing the double challenge represented by the estimation of topology and link weights. The task at hand consists in determining both binary and weighted ensemble distributions, and to understand the interplay between them. Among the methods proposed so far, some assume that the binary and weighted constraints jointly determine the final configuration in terms of both topology and weights while others attribute weights to the binary configuration using a completely separate methodology [17, 21]. Amidst the former ones, a special mention is deserved by the Enhanced Configuration Model [15]. This is defined by simultaneously constraining the degrees and the strengths of nodes which jointly affect the estimation of the two sets of quantities, the linkage probabilities and the weight estimates. Since these are jointly determined on the basis of the same information (i.e. constraints), this implies the impossibility to include purely topological additional information. Examples of algorithms belonging to the second group are those iteratively adjusting the link weights (e.g. via the RAS recipe [19]) on top of some previously-determined topological structure, in such a way to satisfy the constraints concerning strengths a posteriori. This approach has encountered critiques in [10]. It is important to notice that this kind of procedure assigns weights deterministically, and therefore the likelihood of observing any real matrix is exactly zero, assuming continuous weights. There exist also two-steps algorithms [6] that attempt to overcome the lack of binary information for the topological estimation. However, they have an elevated complexity (the second step requires the solution of the ECM) and the two step procedure is only heuristically motivated. Here we develop a theoretical framework that provides an analytical, unbiased procedure to estimate the weighted structure of a network, once its topology has been determined, thus extending the Exponential Random Graph framework to deal with cases that appeared tractable only via heuristic approaches. The prior information on the topological structure (either already available or obtained using a method of choice) and a number of weighted constraints are taken as input. Subsequently, we derive the unbiased link weight probability. The latter is determined by the maximization of the key quantity of our approach, i.e. the conditional entropy of the weighted distribution given the binary one.

As it will be proven, once the weights are treated as continuous random variables, their distribution, conditional on the existence of a link, is exponential. This consideration also allows to determine confidence intervals for the weights estimates. Another desirable property of the model, in its second specification, is its computational simplicity, as it does not require the numerical solution of several coupled non-linear equations.

We indicate the binary adjacency matrix of the network as \( \mathbf{A} \), and we assume that it is a realization from a random variable \( A \). Analogously, the weighted adjacency matrix associated with the graph, \( \mathbf{W} \), comes from a random variable \( W \). The probability mass function of the event \( \{ A = \mathbf{A} \} \) is denoted with \( P(\mathbf{A}) \), while \( Q(\mathbf{W}|\mathbf{A}) \) indicates the conditional multivariate probability density function of \( \mathbf{W} \) belonging to a neighbourhood of \( \mathbf{W} \), given \( \mathbf{A} = \mathbf{A} \). We denote the entries of the weighted adjacency matrix, \( \mathbf{W} \), as \( w_{ij} \) and their binary counterparts as \( a_{ij} = \Theta(w_{ij}) \), where \( \Theta(\cdot) \) represents the Heaviside step function.
**Conditional Reconstruction Method (CReM)**

**Input** The procedure takes as input \(P(A)\), the distribution over the space of the binary configurations. This can be available as prior (probabilistic) information or computed using any available method. In the appendix we discuss a few options for the derivation of \(P(A)\), but it is important to notice that the following results are valid for any choice of \(P(A)\). Moreover, the CReM requires a set of weighted constraints \(\hat{C}(W)\) to be imposed when deriving the weighted distribution. The unconditional ensemble average of such quantities will correspond to the desired target value. This can be chosen to be an observed quantity or any other value of choice.

**Output** The goal of this second step is to derive the distribution of the weighted configurations conditional on the binary one. To achieve this, we maximize over \(Q(W|A)\) the conditional entropy [2]

\[
S(W|A) = - \sum_{A} P(A) \int_{W_A} Q(W|A) \log Q(W|A) dW. \tag{1}
\]

We assume that the weights are continuous. We stress that the distribution \(Q(W|A)\) determined by the second step is relative only to the \(W\) compatible with \(A\), that is such that \(\Theta(W) = A\). We will use the compact notation \(W_A\) to indicate all the \(W\) s.t. \(\Theta(W) = A\). For all \(W' \notin W_A\), we are defined \(Q(W'|A) = 0\).

Our goal is to maximize, with respect to \(Q(W|A)\), the expression [1] under a set of constraints:

1. \(\int_{W_A} Q(W|A) dW = 1 \quad \forall A\)
2. \(\sum_{A} \int_{W_A} P(A) Q(W|A) C_{a}(W) dW = C_{a}^{*}\).

The first equation imposes the normalization of the conditional probabilities: since also \(P(A)\) is normalized, notice that condition 1. implies \(\int_{W} Q(W) dW = 1\). In the second equation \(C_{a}(W)\) represent the desired set of constraints taken as input.

The problem Lagrangian can be written as

\[
\mathcal{L} = - \sum_{A} \int_{W_A} P(A) Q(W|A) \log Q(W|A) dW + \mu(A) \left(1 - \int_{W_A} Q(W|A) dW\right) + \sum_{a} \lambda_{a} \left(C_{a}^{*} - \sum_{A} \int_{W_A} P(A) Q(W|A) C_{a}(W) dW\right) \tag{2}
\]

Differentiating this expression with respect to \(Q(W|A)\) and equating to zero, we obtain

\[
Q(W|A) = \frac{e^{-\sum_{a} \lambda_{a} C_{a}(W)}}{\int_{W_A} e^{-\sum_{a} \lambda_{a} C_{a}(W)} dW} = \frac{e^{-H(W)}}{Z_{A}}, \tag{3}
\]

where \(H(W) = \sum_{a} \lambda_{a} C_{a}(W)\) is the Hamiltonian and the parameter \(\mu\) guarantees that condition 1. is satisfied. The final form for the conditional distribution of the weighted configuration is obtained choosing which constraints we wish to impose.

**Functional form of likelihood function** In alignment with previous results on maximum entropy methods [1] [9] [14], we wish to formulate our problem in terms of maximum likelihood estimation. That is, we want to be able to re-state the problem is such a form that allows to determine the values of the Lagrange multipliers \(\bar{\lambda}\) of (2) by solving a maximum likelihood problem. See Appendix for the standard case.

Let us consider the integral term of the conditional entropy defined in [1]. Using (3) we write

\[
\int_{W_A} Q(W|A) \log Q(W|A) dW = \int_{W_A} Q(W|A)[-H(W) - \log Z_{A}] = \langle H \rangle_{A} - \log Z_{A}. \tag{4}
\]

Let us now define \(\bar{\lambda}\) as the value of the parameters for which the constraints are realized, that is

\[
\langle \hat{C}(\bar{\lambda}) \rangle = \sum_{A} P(A) \int_{W_A} \hat{C}(W) Q(W|A, \bar{\lambda}) dW = \hat{C}^{*}. \tag{5}
\]

When we evaluate the conditional entropy for such parameter choice, since for both method specifications \(H\) is linear in \(w_{ij}\), we obtain

\[
S(\bar{\lambda}) = - \sum_{A} P(A) \langle (H)_{A}(\bar{\lambda}) - \log Z_{A}(\bar{\lambda}) \rangle = - \sum_{A} P(A) \log Q(\langle W \rangle^{*}|A), \tag{6}
\]

where \(\langle W \rangle^{*}\) indicates the unconditional ensemble average of \(W\) when the desired constraints are satisfied. The last equation conveys meaningful information: we started from the expression of the conditional entropy which averages both with respect to the binary configuration and to the weighted one. Imposing certain constraints to be satisfied on average, we determine the value of the vector of parameters \(\bar{\lambda}\). When (1) is evaluated in \(\bar{\lambda}\), it does no longer contain the averaging with respect to weighted configurations, but instead a single term, \(\log Q(\langle W \rangle^{*}|A)\), then is then averaged over the space of binary configurations.

Therefore, the functional form of the likelihood function that preserves the dual relation between entropy and likelihood that we have in the standard case is given by the generalized likelihood

\[
G(\bar{\lambda}) = \sum_{A} P(A) \log Q(\langle W \rangle|A). \tag{7}
\]

This correspondence between entropy and likelihood has another important consequence: the entropy, when evaluated in \(\bar{\lambda}\), itself becomes a measure of goodness of fit
of the model. Let us observe that the estimation of parameter is a more general problem, arising for instance in the case of parameters estimation for the generative model of a known system, as treated in [14].

We are now going to derive the full specification of the model with two different constraint choices: the strength sequences (CReM_A) and the conditional weights (CReM_B).

CReM_A For each node $i$ the in and out degree and in and out strength can be expressed as $k_i^{out} = \sum_j \Theta(w_{ij}), k_i^{in} = \sum_j \Theta(w_{ji}), s_i^{out} = \sum_j w_{ij}$ and $s_i^{in} = \sum_j w_{ji}$. Imposing the preservation of the strength sequences results in an Hamiltonian function of the form:

$$H(W) = \sum_i (\beta_i^{out} s_i^{out} + \beta_i^{in} s_i^{in}) = \sum_{i \neq j} (\beta_i^{out} + \beta_j^{in}) w_{ij}.$$ (8)

The partition function can be expressed as

$$Z_A = \prod_{i \neq j} \left[ \int_0^\infty e^{-(\beta_i^{out} + \beta_j^{in}) w_{ij}} dw_{ij} \right]^{a_{ij}} = \prod_{i \neq j} \left( \frac{1}{\beta_i^{out} + \beta_j^{in}} \right)^{a_{ij}}$$ (9)

Finally, using expression [3], we can write

$$Q(W|A) = \prod_i e^{-\left( \beta_i^{out} s_i^{out} + \beta_i^{in} s_i^{in} \right)} \prod_{i \neq j} (\beta_i^{out} + \beta_j^{in})^{a_{ij}},$$ (10)

which implies

$$q_{ij}(w|a_{ij}) = (\beta_i^{out} + \beta_j^{in})^{a_{ij}} e^{-(\beta_i^{out} + \beta_j^{in}) w}$$ (11)

that shows that the weight distribution, conditional on the existence of a link, $q_{ij}(w|a_{ij} = 1)$ follows an exponential distribution of parameter $\beta_i^{out} + \beta_j^{in}$.

In order to determine the values of the vector of parameters $\beta^{out}$ and $\beta^{in}$, we maximize the expression of the conditional likelihood, that, substituting the results just derived into [7], reads as follows

$$G_A = -\sum_i \left( s_i^{out} \beta_i^{out} s_i^{in} \beta_i^{in} \right) + \sum_{i \neq j} f_{ij} \log(\beta_i^{out} + \beta_j^{in})$$ (12)

The expression $f_{ij} = \sum_A P(A) a_{ij}$ represents the value of $a_{ij}$ averaged on the ensemble of binary configurations. This general formulation implies no assumption on the structure of link interdependencies given by $P(A)$. For instance, in the case of the micro-canonical ensemble, where the degree are constrained sharply and not on average on the ensemble, $P(A)$ is constant on all graphs with the same degree sequence. When $P(A)$ factorizes $f_{ij}$ corresponds to the quantity denoted with $p_{ij}$ in the literature [20], that is the linkage probability the is independent from link to link.

Differentiating [12] with respect to $\beta_i^{out}$ and $\beta_i^{in}, \forall i$, yields the system of $2N$ coupled equations

$$\begin{cases} s_i^{out} = \sum_j \frac{f_{ij}}{\beta_i^{out} + \beta_j^{in}} \\ s_i^{in} = \sum_j \frac{f_{ij}}{\beta_i^{out} + \beta_j^{in}} \Rightarrow \langle w_{ij} \rangle = \frac{f_{ij}}{\beta_i^{out} + \beta_j^{in}}. \end{cases}$$ (13)

where $f_{ij}$ is taken as given and therefore excluded from the estimation procedure.

CReM_B In Figure 1 we can see, in blue circles, the comparison between the weight estimates delivered from the CReM_A model versus the real weights in blue. The red crosses represent the same comparison but with respect to the weight estimates taken from the degree-corrected Gravity Model [13]

$$w_{ij} = \frac{s_i^{out} s_j^{in}}{W_{tot}},$$ (14)

where $W_{tot} = \sum_i s_i^{out} = \sum_i s_i^{in}$ is the total weight of the matrix. We can see that the latter estimates shows a better agreement with the data. Therefore, we wish to find a different model specification that allows us to introduce such a functional form for the weight estimates. To do so, in the conditional entropy maximization we formally constrain the values of the weights. The Hamiltonian reads

$$H(W) = \sum_{i \neq j} \beta_{ij} W_{ij}.$$ (15)

The derivation is then analogous to the previous case. Given [3], the conditional weight distribution will then be

$$Q(W|A) = \prod_{i \neq j} e^{-\beta_{ij} w_{ij}} = \prod_{i \neq j} e^{-\beta_{ij} w_{ij} (\beta_{ij})^{a_{ij}}}.$$ (16)

As a consequence, also in this case the conditional weight distribution follows an exponential distribution of parameter $\beta_{ij}$. The generalized likelihood function, given [7], can be expressed as

$$G_B = \sum_A P(A) \left[ -\sum_{i \neq j} \left( w_{ij}^* \beta_{ij} + a_{ij} \log \beta_{ij} \right) \right]$$ (16)

$$= \sum_{i \neq j} \left( -w_{ij}^* \beta_{ij} + f_{ij} \log \beta_{ij} \right).$$ (17)

Differentiating with respect to $\beta_{ij}$ leads to the equation

$$w_{ij}^* = \frac{f_{ij}}{\beta_{ij}}.$$ (17)
Clearly, we cannot really observe the link weights (or there would be no need for reconstruction), but we formally imposed such constraints in order to be able to use the weight estimates from (14), which shows good agreement with the data. Such a choice uses as input only the strength sequences of the graph. As a consequence the sufficient statistics for CReM\(_A\) and CReM\(_B\) are the same: the strength sequences.

If we substitute the values \(w_{ij}^*\) in (17) with the expression from (14), to determine the matrix of coefficients \(\beta\) we only have to compute

\[
\beta_{ij} = \frac{W_{\text{tot}} f_{ij}}{s_{ij}^*}. \quad (18)
\]

**Directed Enhanced Configuration Model** In order to compare the results of the reconstruction models and to underline the different assumption regarding the network formation process, we will outline the characteristics of a model belonging to the class of algorithms jointly determining topological and weighted structure, i.e. the Directed Enhanced Configuration Model [2]. For consistency, we will consider the continuous version of this model, whose probability distribution reads

\[
Q(W) = \prod_{i,j} q_{ij}(w) = \prod_{i,j} \left( x_{ij}^\text{out} x_{ij}^\text{in} \right)^{\Theta[w]} e^{-(\beta_{ij}^\text{out} + \beta_{ij}^\text{in})w}, \quad (19)
\]

where \(\{\alpha_i^\text{out}, \alpha_i^\text{in}, \beta_i^\text{out}, \beta_i^\text{in}\}_{i=1}^N\) are the Lagrange multipliers associated to the constraints and \(x_i = e^{-\alpha_i}\). This implies

\[
q_{ij}(w > 0) = p_{ij}(\beta_{ij}^\text{out} + \beta_{ij}^\text{in}) e^{-(\beta_{ij}^\text{out} + \beta_{ij}^\text{in})w}, \quad (20)
\]

where \(q_{ij}(w > 0)\) is the probability that there is a link of any positive weight \(w\) between nodes \(i\) and \(j\), while \(p_{ij} = 1 - q_{ij}(0) = \left[1 + (\beta_{ij}^\text{out} + \beta_{ij}^\text{in})/(x_{ij}^\text{out} x_{ij}^\text{in})\right]^{-1}\) is the probability that a link of any weight exists.

Observe that we retained the literature notation for the linkage probability \(p_{ij}\), referring to the special case in which \(P(A)\) factorizes and as a consequence the linkage probabilities are independent from each other.

Finally, the likelihood function can be expressed as

\[
\mathcal{L} = \sum_i \left( k_i^\text{out} \ln x_i^\text{out} + k_i^\text{in} \ln x_i^\text{in} - s_i^\text{out} \beta_{i}^\text{out} - s_i^\text{in} \beta_{i}^\text{in} \right) - \sum_{i,j} \ln \left( 1 + \frac{x_{ij}^\text{out} x_{ij}^\text{in}}{\beta_{ij}^\text{out} + \beta_{ij}^\text{in}} \right). \quad (21)
\]

All the details and the derivations of the DECM model can be found in the Appendix.

Let us now come to a procedure that will enable us to compare different reconstruction models.

**Likelihood-based comparison** The first possible approach is to compare the models through their (generalized) likelihood \(G\). Given any observed strength sequences (or strength and degree for the DECM) we can solve the constrained entropy maximization. In particular, we solve the likelihood equations characterizing the two models, i.e. for model A we determine the vectors \(\beta^\text{out}^A\) and \(\beta^\text{in}^A\) by solving the system of equations (13) and for model B, we compute the quantities (18). For both models A and B, we used as prior binary distribution the linkage probabilities from the dcGM model (See (25) in the Appendix for details), since methods using such probability form have won several independent horse races [10][12].

Once we have obtained those parameters, we can compute the likelihood of observing the original network by reconstructing it through the different models. In Figure 2 we can see the comparison of the values of the (generalized) likelihood relative to models CReM\(_A\), CReM\(_B\) and DECM, on several snapshots of two different real-world datasets, the World Trade Web and the E-mid [8]. Since \(G\) is the (weighted average of) the logarithm of a probability, takes negative values. The closer its value is to zero, the better the model explains the data. Therefore, when using the likelihood to compare two models, we shall prefer the model with the highest values of \(G\). As we can see from the figure, the CReM\(_B\) performs better in the WTW dataset, while the performances are extremely similar on the E-mid one.

**Generalized AIC test** A very common test for model selection is the Akaike Information Criterion (AIC) [3]. For a given set of data, for a generic model \(m\), the standard AIC recipe can be generalized as

\[
\text{AIC}_m = 2k_m - 2L_m, \quad \Rightarrow \quad \text{GAIC}_m = 2k_m - 2G_m \quad (22)
\]

where \(k\) is the number of estimated parameters of the model and \(G\) the value of the model generalized likelihood computed on the observed data. The number of parameters estimated by each model is \(k_{\text{DECM}} = 4N\), \(k_{\text{CReM}\_A} = 2N + 1\), and \(k_{\text{CReM}\_B} = 2N + 1\).

The additional parameter comes from constraining the link density, as with the dcGM recipe (see Appendix for details). For both CReM\(_A\) and CReM\(_B\) the sufficient statistics are the vectors of and out strengths and the total density of links. Given this observation, we note that the AIC test will yield the same model ranking as the pure likelihood comparison, since the best performance was awarded to CReM\(_B\), followed by CReM\(_A\) and finally by the DECM. Since also in terms of complexity the ranking is coincident, the AIC comparison ranks again the CReM\(_B\) first, followed by CReM\(_A\) and finally by the DECM.

Given the good performance of model B and its low complexity, we recommend this choice and we will soon release a code to make its application fast and straightforward.
Discussion

The method introduced in this paper aims at filling a methodological gap. Several proposed methods for network reconstruction, combine methodologically different steps to carry out the estimation of topology and of weights, thus introducing biases in the whole procedure. A first source of bias is encountered when a probabilistic recipe for topological reconstruction is forced to produce a single outcome instead of considering the entire ensemble of admissible configurations. This choice implies a null likelihood of reproducing the actual network. A second source of bias is encountered when the weights structure is deterministically imposed via a recipe like the RAS one. Again, this recipe ensures a zero likelihood of reproducing the real underlying network: the probability of correctly "guessing" all the link weights is null. Here we reconcile the two aspects, by providing a recipe that clarifies how weights should be determined, once an algorithm for determining the topology of a given network is implemented (be it either probabilistic or deterministic). Notice that the key concept of our approach, i.e. conditional entropy, generalizes traditional approaches which, instead, aim at jointly determine a given network structure: this is immediately seen by rewriting the probability distribution coming from Shannon entropy maximization as in (3).

However, although it is clear how the topology of a network should represent a constrain for the weighted configuration, it is not necessarily true that the weighted configuration has as well an impact of the topology. The CReM takes a different approach. It assumes only the first dependency, but allows the network topology to be unaware of the weighted configuration.

On a more practical level, the solution of both model A and B require a lower complexity with respect to the DECM. The latter one, in fact, requires the solution of a system of $4N$ four-wise dependent equation, while model A only involves a system of $2N$ paired equations. Model B on the other hand, despite being formally similar in derivation, does not require to solve any system of equations, since the involved parameters are computed via the recipe [1]. In addition to this, model B shows better or comparable performance with respect the alternative methods, both in terms of likelihood and confidence interval-based comparisons.

---

* federica.parisi@imtlucca.it

[1] Squartini, T., Garlaschelli, D. Analytical maximum-likelihood method to detect patterns in real networks, New Journal of Physics, 13(8) (2011).
[2] Cimini, G., Mastrandrea, R., Caldarelli, G., Gabrielli, A. The Grand Canonical Ensemble of Weighted Networks, in preparation, (2018).
[3] Cover, T. M., Thomas, J. A. Elements of Information Theory, John Wiley & Sons, (2006).
[4] Almog, A., Squartini, T., Garlaschelli, D. A GDP-driven model for the binary and weighted structure of the International Trade Network, New Journal of Physics, 17(1) (2015).
[5] Akaike, H. A new look at the statistical model identification, IEEE transactions on automatic control, 19(9) (1974).
[6] Cimini, G., Squartini, T., Gabrielli, A., Garlaschelli, D. Estimating topological properties of weighted networks from limited information, Physical Review E, 92(4) (2015).
[7] Squartini, T., Cimini, G., Gabrielli, A., Garlaschelli, D. Network reconstruction via density sampling, Applied Network Science, 2(1) (2017).
[8] Iori, G., Jafarey, S., Padilla, F. G. Systemic risk on the interbank market, Journal of Economic Behavior, Organization, 61(4) (2006).
[9] Squartini, T., Garlaschelli, D. Reconnecting statistical physics and combinatorics beyond ensemble equivalence, arXiv:1710.11422 (2017).
[10] Anand, K., van Lelyveld, I., Banai, Á., Friedrich, S., Garrant, R., Halaj, G., Figue, J., Hansen, I. and Jaramillo, S. M., Lee, H., others. The missing links: A global study on uncovering financial network structures from partial data, Journal of Financial Stability, 35 (2018).
[11] Mazzarisi, P., Lillo, F. Methods for reconstructing interbank networks from limited information: a comparisons, N Econophysics and Sociophysics: Recent Progress and Future Directions, 201-215 (2017).
[12] Ramadiah, A., Caccioli, F., Fricke, D. Reconstructing and Stress Testing Credit Networks, (2017).
APPENDIX

Choices for P(A)

The Conditional Reconstruction Method works for any choice of procedure to determine P(A). The overall goodness of fit will depend also on the correctness of the binary model chosen. Here we propose two possible methods, from the literature, to be selected depending on the available information. One possibility is to derive P(A) by maximizing the Shannon entropy

\[ S_A = - \sum_A P(A) \log P(A), \]  

constraining the in and out degree sequences \( \{k^\text{out}\}_i = 1, \ldots, N \) and \( \{k^\text{in}\}_i = 1, \ldots, N \). This step corresponds to solving the DBCM model [20]. The resulting probability takes the form

\[ P(A) = \prod_{ij} p_{ij}^{a_{ij}} (1 - p_{ij})^{1 - a_{ij}}, \]  

\[ p_{ij} = \frac{x_{ij}^{\text{out}} x_{ij}^{\text{in}}}{1 + x_{ij}^{\text{out}} s_{ij}^a}, \]  

(24)

assuming not to know the degree distribution and using the empirically-observed correlation between degree-induced Lagrange multipliers and strengths [13], we can apply the transformations \( x_{ij}^{\text{out}} = \sqrt{a} s_{ij}^{\text{out}} \) and \( x_{ij}^{\text{out}} = \sqrt{b} s_{ij}^{\text{in}} \), that implies, for \( z = \sqrt{ab} \),

\[ p_{ij} = \frac{z s_{ij}^{\text{out}} s_{ij}^{\text{in}}}{1 + z s_{ij}^{\text{out}} s_{ij}^{\text{in}}}, \]  

(25)

where the parameter \( z \) is determined in such a way to keep the expected link density, \( \langle L \rangle \), equal to the observed one, \( L \). This gives \( \langle L \rangle = \sum_i p_{ij} = \sum_i \frac{z s_{ij}^{\text{out}} s_{ij}^{\text{in}}}{1 + z s_{ij}^{\text{out}} s_{ij}^{\text{in}}} = L \).

This corresponds to the use of the linkage probabilities defining the density-corrected Gravity Model (dcGM) [13]. We have just shown some possible choices for the functional form of \( p_{ij} \). Other possibilities are the discussed in [20].

Relation between Entropy and Likelihood

Let us consider the case of the maximization of the Shannon entropy

\[ S(P) = - \sum_G P(G) \log P(G), \]  

(26)

under a set of constraints \( \bar{C} \), where \( P \) is the distribution we wish to determine over the space of graph \( G \in \mathcal{G} \). It can be shown that \( P \) has a functional form given by the Exponential Random Graph model \( P(G) = e^{-H(G)}/Z \), where \( H(G) = \sum_{\alpha} \lambda_{\alpha} C_{\alpha} \), is a linear combination of the constraints and \( Z = \sum_G e^{-H(G)} \) is the partition function. We define as \( \bar{\alpha}^* \) the value of the parameters such that the constraints are satisfied, that is such that

\[ \langle \bar{C} \rangle^* = \langle \bar{C} \rangle^{\bar{\alpha}^*} = \sum_G \bar{C}(G) P(G|\bar{\alpha}^*) = \bar{C}(\bar{\alpha}^*) = \bar{C}^*. \]  

(27)

Now, we can rewrite (26) as

\[ S(P) = - \sum_G P(G)(-H(G) - \log Z) \]  

\[ = \sum_G P(G)H(G) + \log Z \]  

\[ = \langle H \rangle + \log Z \]  

\[ = \langle H \rangle - H(G) - \log P(G). \]  

(28)

When we evaluate this in \( \bar{\alpha} = \bar{\alpha}^* \), we obtain

\[ S(\bar{\alpha}^*) = S(P_{\bar{\alpha}^*}) = \langle H \rangle(\bar{\alpha}^*) - H^* - \log P(G|\bar{\alpha}^*). \]  

(29)

Now, using (27), we have

\[ \langle H \rangle(\bar{\alpha}^*) = \sum_{\alpha} \lambda_{\alpha} \langle C_{\alpha} \rangle(\bar{\alpha}^*) = \sum_{\alpha} \lambda_{\alpha}^* C_{\alpha}^* = H^*. \]  

(30)

Therefore, (29) becomes

\[ S(\bar{\alpha}^*) = - \log P(G|\bar{\alpha}^*). \]  

(31)
Since the expression $\mathcal{L}(\tilde{x}) = -\log P(G|\tilde{x})$ is the likelihood associated with that parameter choice that assures that the chosen constraints are realized on average on the ensemble, the last equation tells us that when evaluated in $\tilde{x}$, the entropy itself can be used to assess the goodness of fit of a model.

The DECM model for continuous weights

The Directed Enhanced Configuration Model, with continuous weights, is obtained maximizing the Shannon entropy

$$ S = -\int W Q(W) \log Q(W), \quad (32) $$

relative to the probability distribution over the space of the weighted graph configuration with given number of nodes. The constraints derive from keeping constant, on average, the in and out degree and strength sequences:

$$ (k_i^{\text{out}}) = \int W Q(W) k_i^{\text{out}}(W) = k_i^{\text{out}*} $$
$$ (k_i^{\text{in}}) = \int W Q(W) k_i^{\text{in}}(W) = k_i^{\text{in}*} $$

$$ (s_i^{\text{out}}) = \int W Q(W) s_i^{\text{out}}(W) = s_i^{\text{out}*} $$
$$ (s_i^{\text{in}}) = \int W Q(W) s_i^{\text{in}}(W) = s_i^{\text{in}*} \quad (33) $$

The probability distribution, as in the general Exponential Random Graph framework [18], takes the functional form

$$ Q(W) = \frac{e^{-H(W)}}{Z}, \quad (34) $$

where $Z = \int W e^{-H(W)}$ is the partition function and

$$ H(W) = \sum_{i=1}^{N} \left[ \alpha_i^{\text{out}} k_i^{\text{out}}(W) + \alpha_i^{\text{in}} k_i^{\text{in}}(W) + \beta_i^{\text{out}} s_i^{\text{out}}(W) + \beta_i^{\text{in}} s_i^{\text{in}}(W) \right] $$

$$ = \sum_{i,j} \left[ (\alpha_i^{\text{out}} + \alpha_j^{\text{in}}) \Theta[w_{ij}] + (\beta_i^{\text{out}} + \beta_j^{\text{in}}) w_{ij} \right] $$

From this we can compute

$$ Z = \int W e^{-H(W)} = $$

$$ = \prod_{i,j} \int_0^\infty \left[ \delta_{w_{ij},0} + \Theta(w_{ij}) \right] e^{-(\alpha_i^{\text{out}} + \alpha_j^{\text{in}}) \Theta[w_{ij}]} \cdot e^{-(\beta_i^{\text{out}} + \beta_j^{\text{in}}) w_{ij}} dw_{ij} = $$

$$ = \prod_{i,j} \left[ 1 + e^{-(\alpha_i^{\text{out}} + \alpha_j^{\text{in}})} \int_0^\infty e^{-(\beta_i^{\text{out}} + \beta_j^{\text{in}}) w_{ij}} dw_{ij} \right] = $$

$$ = \prod_{i,j} \left[ 1 + e^{-\beta_i^{\text{out}} + \beta_j^{\text{in}}} \right]. \quad (35) $$

which implies

$$ Q(W) = \prod_{i,j} q_{ij}(w) = \prod_{i,j} \left( \frac{\lambda_i^{\text{out}} \lambda_j^{\text{in}} \Theta[w]}{1 + \lambda_i^{\text{out}} \lambda_j^{\text{in}} / (\beta_i^{\text{out}} + \beta_j^{\text{in}})} e^{-(\beta_i^{\text{out}} + \beta_j^{\text{in}}) w} \right), \quad (36) $$

The likelihood function to maximize to estimate the model parameter reads

$$ \mathcal{L}(W^*) = \ln Q(W^*) = \sum_{i,j} \ln q_{ij}(w_{ij}) = $$

$$ = \sum_i \left( k_i^{\text{out}*} \ln x_i^{\text{out}} + k_i^{\text{in}*} \ln x_i^{\text{in}} - s_i^{\text{out}*} \beta_i^{\text{out}} - s_i^{\text{in}*} \beta_i^{\text{in}} \right) + $$

$$ - \sum_{i,j} \ln \left( 1 + \frac{x_i^{\text{out}} x_j^{\text{in}}}{\beta_i^{\text{out}} + \beta_j^{\text{in}}} \right). \quad (37) $$

Confidence Intervals-based comparison

In order to introduce the second comparison method, we need to make some observations about the conditional weight distribution $Q(W|A)$. For all the methods, this distribution factorizes in the product of the individual terms $q_{ij}(w|a_{ij})$. For CReMA, this term reads

$$ q_{ij}(w|a_{ij}) = (\beta_i^{\text{out}} + \beta_j^{\text{in}}) a_{ij} e^{-(\beta_i^{\text{out}} + \beta_j^{\text{in}}) w} \quad (38) $$

given the compatibility requirement we have $q_{ij}(w|a_{ij} = 0) = \delta_{w,0}$. Moreover we observe that

$$ q_{ij}(w|a_{ij} = 1) = (\beta_i^{\text{out}} + \beta_j^{\text{in}}) e^{-(\beta_i^{\text{out}} + \beta_j^{\text{in}}) w}, \quad (39) $$

which means that the weight distribution, conditional on the existence of a link, follows an exponential distribution of parameter $\beta_i^{\text{out}} + \beta_j^{\text{in}}$. The same observation holds for CReMB, where the exponential parameter is now given by $\beta_{ij}$. We observe then
FIG. 4. (Colour online) Comparison on the proportion of weights falling into the confidence interval relative to the considered method, for the WTW and E-mid datasets respectively, with value $q^+ = q^- = 0.25$. Red line indicates CReM$_A$ and blue line CReM$_B$.

that both models show the same functional form of the weights conditional distribution as the DECM (see (20)). However the numerical values of the parameters characterizing the distribution are different. Let us denote the parameter of the exponential distribution as $\lambda_{ij}$ so to keep the reasoning general.

Given the knowledge on the weight distribution we can build confidence intervals around the expected value of the weight. Figure 3 gives an illustration of the procedure. The conditional expected value of the weight estimate is indicated as $\tilde{w}$. The blue curve represents the probability density function of the exponential distribution of parameter $\lambda_{ij}$. Our goal is to determine the extreme of the confidence interval, $w^-$ and $w^+$, in such a way that the area under the pdf comprised between the expected weight and $w^-$ is equal to the desired confidence level $q^-$, and analogously for the upper part of the interval. This is achieved by solving

$$
\int_{w^-}^{(\langle w_{ij}\mid a_{ij}=1\rangle)} \lambda_{ij} e^{-\lambda_{ij}w_{ij}} dw_{ij} = q^-
$$

leading to

$$
w^- = -\ln\left[e^{-\lambda_{ij}\langle w_{ij}\mid a_{ij}=1\rangle} + q^-\right] = -\frac{\ln[e^{-1} + q^-]}{\lambda_{ij}}
$$

and

$$
\int_{w=(\langle w_{ij}\mid a_{ij}=1\rangle)}^{w^+} \lambda_{ij} e^{-\lambda_{ij}w_{ij}} dw_{ij} = q^+
$$

leading to

$$
w^+ = -\frac{\ln[e^{-\lambda_{ij}\langle w_{ij}\mid a_{ij}=1\rangle} - q^+]}{\lambda_{ij}} = -\frac{\ln[e^{-1} - q^+]}{\lambda_{ij}}
$$

In this way, upon deciding on the desired confidence levels $q^-$ and $q^+$, we are able to define confidence interval for the conditional expected weights. Notice that such a confidence interval is not symmetric, given the peculiar form of the underlying probability distribution (i.e. exponential). Although both the ECM and the DECM can provide an error estimation, its computation is much easier within the novel, continuous framework considered here. Finally, given the confidence intervals for each weight estimate, we can compute the proportion of real weights that fall into the confidence interval relative to their estimate. We can do this for all methods, using the same $q^+$ and $q^-$ and the different $\lambda_{ij}$. The results are shown in Figure 4. As for the likelihood-based comparison, model B shows a comparable or better performance than method A, thus confirming our previous considerations. Notice that both the ECM \cite{15} and the DECM provide an error estimation as well; its computation, however, is much easier within the novel, continuous framework considered here.