Generalized inference for the efficient reconstruction of weighted networks

entropy 2021

Federica Parisi, Tiziano Squartini, Diego Garlaschelli
IMT School for Advanced Studies Lucca (Italy)

Introduction

Network reconstruction is an active field of research. Among the methods proposed so far, some assume that the binary and weighted constraints jointly determine the reconstruction output; others consider the weights estimation step as completely unrelated to the binary one. Amidst the former ones, a special mention is deserved by the Enhanced Configuration Model; the algorithms of the second group, instead, are those iteratively adjusting the link weights on top of some previously-determined topology.

Methods

Our method generalizes the traditional reconstruction framework, by maximizing the *conditional Shannon entropy* in a constrained fashion. Our recipe satisfies the following requirements:

- it allows for *any* probability distribution (over purely binary graphs) to be acceptable as input for the preliminary topology reconstruction step;
- it allows for the generation of continuous weights, in order to let a real, unobserved network to be generated with positive likelihood;
- it satisfies the constraints that are usually imposed by the availability of limited information (i.e. the out- and in-strength sequences).

Conclusions

The knowledge of the structure of a financial network gives valuable information for estimating the systemic risk. However, since financial data are typically subject to confidentiality, network reconstruction techniques become necessary to infer both the presence of connections and their intensity. Recently, several 'horse races' have been conducted to compare the performance of these methods. Here, we establish a generalised likelihood approach to rigorously define and compare methods for reconstructing weighted networks: the best one is obtained by 'dressing' the best-performing, available binary method (i.e. the density-corrected Gravity Model) with an exponential distribution of weights.

References

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

[2] T. Squartini, G. Caldarelli, G. Cimini, A. Gabrielli, D. Garlaschelli, *Reconstruction methods for networks: the case of economic and financial systems*, Phys. Rep. 757, 1-47 (2018).

[3] G. Cimini, T. Squartini, F. Saracco, D. Garlaschelli, A. Gabrielli, G. Caldarelli, *The statistical physics of real-world networks*, Nat. Rev. Phys. 1(1), 58-71 (2019).

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

An overview of our reconstruction method

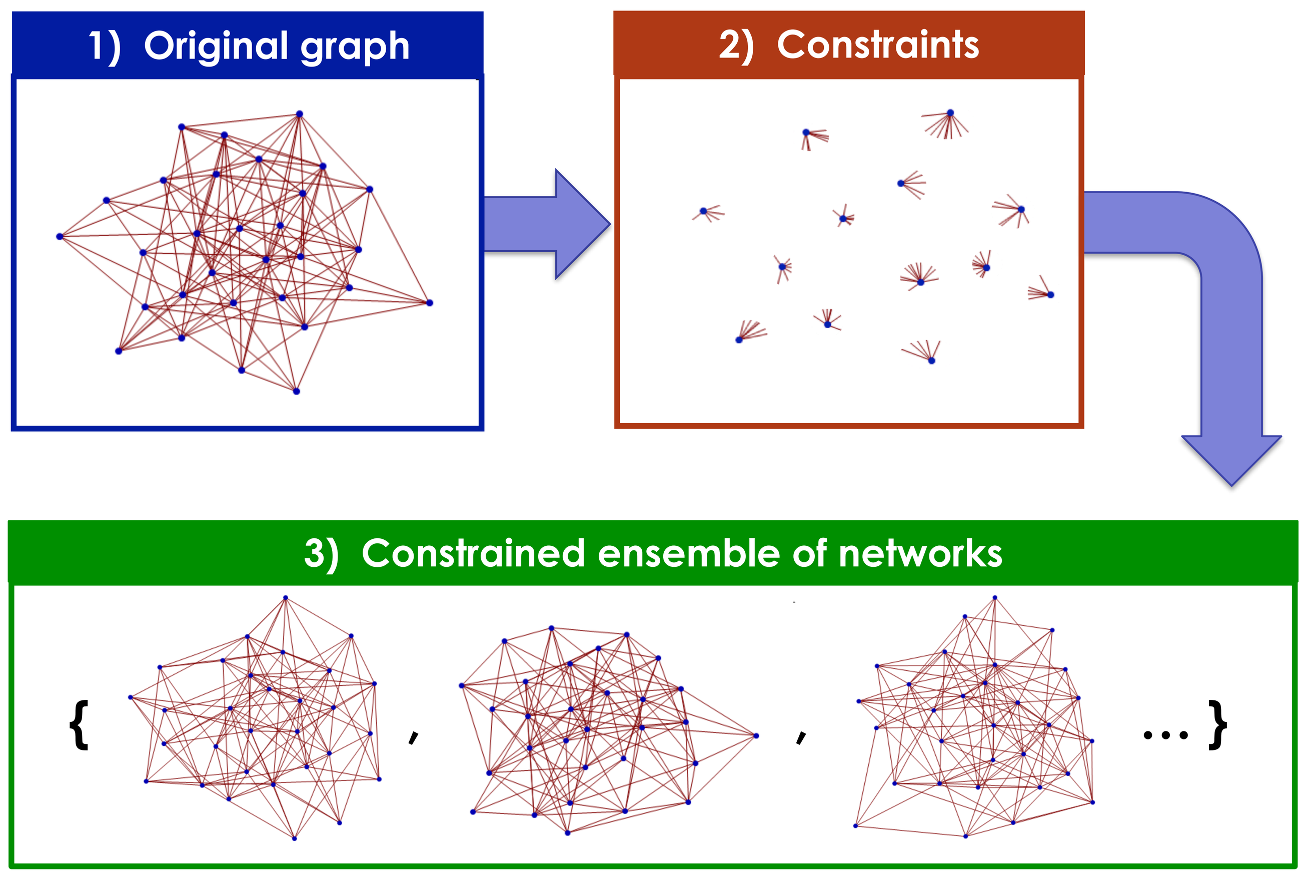

Figure 1: Graphic summary of the reconstruction procedure.

Our method generalizes the traditional reconstruction framework, by maximizing the *conditional Shannon entropy*

$$S(\mathcal{W}|\mathcal{A}) = -\sum_{\mathbf{A}\in\mathbb{A}} P(\mathbf{A}) \int_{\mathbb{W}_\mathbf{A}} Q(\mathbf{W}|\mathbf{A}) \ln Q(\mathbf{W}|\mathbf{A}) d\mathbf{W}$$

in a constrained fashion. Upon imposing the normalization condition $\int_{\mathbb{W}_\mathbf{A}} Q(\mathbf{W}|\mathbf{A})d\mathbf{W} = 1, \forall \mathbf{A} \in \mathbb{A}$ and the proper set of constraints $\sum_{\mathbf{A}\in\mathbb{A}} P(\mathbf{A}) \int_{\mathbb{W}_\mathbf{A}} Q(\mathbf{W}|\mathbf{A}) C_\alpha(\mathbf{W}) d\mathbf{W} = C^*_\alpha$, $\forall\,\alpha$, the conditional, exponential distribution

$$Q(\mathbf{W}|\mathbf{A}) = \frac{e^{-H(\mathbf{W})}}{Z_\mathbf{A}}, \quad \mathbf{W} \in \mathbb{W}_\mathbf{A}$$

is recovered. Its conditioning is on a set of binary configurations that represents the *a priori* available, topological information. The solution to the aforementioned problem, in the case of continuous weights, reads

$$q(w|a_{ij}=1) = \begin{cases} (\beta_i^{out} + \beta_j^{in}) e^{-(\beta_i^{out}+\beta_j^{in}) w_{ij}} & w > 0 \\ 0 & w \leq 0 \end{cases}$$

for each positive weight w_{ij}, showing that each pair-specific weight distribution conditional on the existence of the link is exponential with parameter $\beta_i^{out} + \beta_j^{in}$.

In order to estimate the parameters of the conditional distribution above, the likelihood-maximization step needs to be generalied as well. To this aim, we consider the *generalized likelihood functional*

$$\mathcal{G} = -H(\mathbf{W}^*) - \sum_{\mathbf{A}\in\mathbb{A}} P(\mathbf{A}) \ln Z_\mathbf{A}$$

whose optimization leads to the system of equations

$$\begin{cases} s_i^{out,*} &= \frac{f_{ij}}{\beta_i^{out}+\beta_j^{in}}, \quad \forall\, i \\ s_i^{in,*} &= \frac{f_{ji}}{\beta_j^{out}+\beta_i^{in}}, \quad \forall\, i \end{cases}$$

with $f_{ij} = \sum_{\mathbf{A}\in\mathbb{A}} P(\mathbf{A}) a_{ij} = \langle a_{ij} \rangle$ being the expectation of the binary random variable a_{ij} under a generic probability distribution.

The framework above has been tested in a couple of particular cases, i.e. $f_{ij} = \frac{z s_i^{out} s_j^{in}}{1 + z s_i^{out} s_j^{in}}$ and $f_{ij} = a_{ij}$. The results are shown in the figure below.

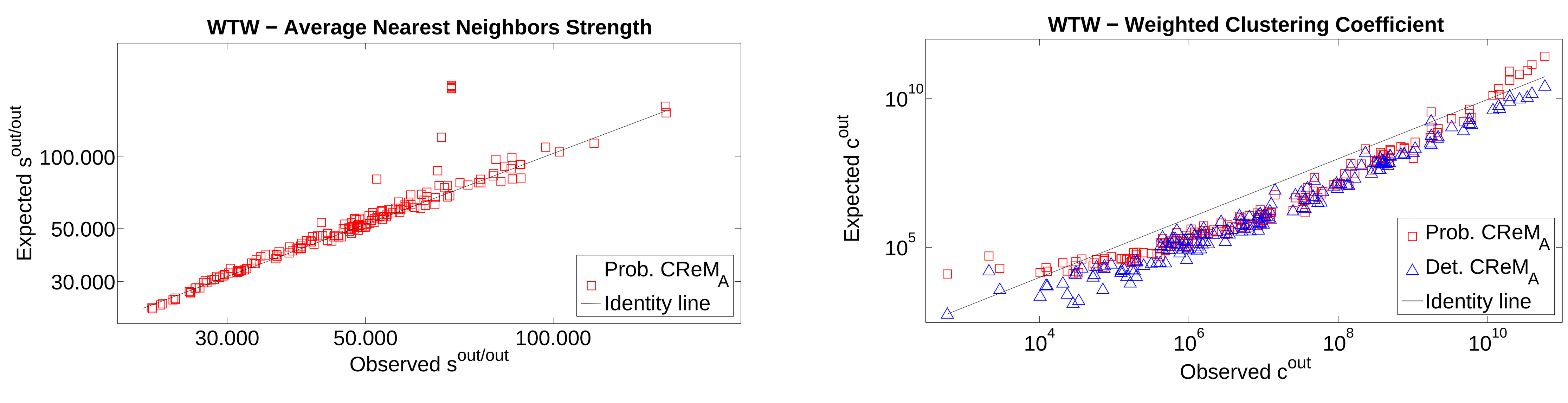

Figure 2: Test of the effectiveness of the CReMA model in reproducing the average nearest neighbors strength (left panel) and the weighted clustering coefficient (right panel) for the World Trade Web in the year 1990. The chosen probability distributions for the binary estimation step are the one defining the density-corrected Gravity Model (red squares) and the one defining the actual configuration (blue triangles). The latter choice perfectly recovers the observed values of the ANNS that lie on the identity line (drawn as a black, solid line); the WCC is reproduced with a much higher accuracy as well.