

The configuration multi-edge model: Assessing the effect of fixing node strengths on weighted network magnitudes

This content has been downloaded from IOPscience. Please scroll down to see the full text.

2014 EPL 107 38002

(<http://iopscience.iop.org/0295-5075/107/3/38002>)

View [the table of contents for this issue](#), or go to the [journal homepage](#) for more

Download details:

IP Address: 161.116.100.31

This content was downloaded on 01/08/2014 at 13:47

Please note that [terms and conditions apply](#).

The configuration multi-edge model: Assessing the effect of fixing node strengths on weighted network magnitudes

O. SAGARRA¹, F. FONT-CLOS^{2,3}, C. J. PÉREZ-VICENTE¹ and A. DÍAZ-GUILERA¹

¹ *Departament de Física Fonamental, Universitat de Barcelona, Facultat de Física - C. Martí i Franquès 1, 08028 Barcelona, Spain*

² *Centre de Recerca Matemàtica, Campus de Bellaterra, Edifici C - 08193 Bellaterra (Barcelona), Spain*

³ *Departament de Matemàtiques, Universitat Autònoma de Barcelona, Edifici C 08193 Bellaterra (Barcelona), Spain*

received 22 May 2014; accepted in final form 10 July 2014

published online 25 July 2014

PACS 89.75.Da – Systems obeying scaling laws

PACS 89.75.Fb – Structures and organization in complex systems

PACS 89.75.Hc – Networks and genealogical trees

Abstract – Complex networks grow subject to structural constraints which affect their measurable properties. Assessing the effect that such constraints impose on their observables is thus a crucial aspect to be taken into account in their analysis. To this end, we examine the effect of fixing the strength sequence in multi-edge networks on several network observables such as degrees, disparity, average neighbor properties and weight distribution using an ensemble approach. We provide a general method to calculate any desired weighted network metric and we show that several features detected in real data could be explained solely by structural constraints. We thus justify the need of analytical null models to be used as basis to assess the relevance of features found in real data represented in weighted network form.

Copyright © EPLA, 2014

Introduction. – Modern complex networks theory has found many applications since its dawn. In particular, the explosion of information technologies has given rise to large-scale, high-dimensional data sets, in which hidden relations might now be uncovered. This has fostered data-driven studies in a wide spectrum of fields ranging from biology [1] to social sciences [2], including transportation studies [3], genomics [4], ecology or bibliometrics [5]. The repertoire of available networks for data modeling has thus grown accordingly: binary or weighted [6,7], directed or undirected, and also simple or multilayer [8,9] structures have been used. However, extending some of the most basic concepts and tools, such as clustering coefficient, centrality measures, and even finding suitable null models in each case has proven harder than expected [10], giving rise to multiple definitions in some cases and, consequently, to some controversy.

In fact, the need to distinguish different types of so-called *weighted* networks according to the nature of the events being represented has been pointed out recently [11]. If nodes accept multiple *distinguishable* connections, then one can speak of multi-edge networks, and it is in this scenario that we propose a flexible, general theory

for null-model generation. Our results allow to compute exact analytical expressions for network observables generated under random conditions but preserving some given properties or constraints from the original data [12,13]. This allows to quantify the relevance of the observed features, which, given the high dimensionality of the studied data sets in real complex networks, might not be trivial to detect otherwise [14]. A classical example where this circumstance is important are Origin-Destination matrices (OD), where the mobility of agents between departure and arrival nodes is represented as weighted networks: usually the average total flow incoming or outgoing from each site, which corresponds to the strength of a node, is constrained by several factors such as population or density of commercial areas. This fact needs to be assessed on the observed data as it can generate complicated spatial patterns that can produce spatial correlations such as the so-called *gravity laws of transportation* [3].

In this letter, the effect of fixing an arbitrary strengths sequence on several network observables is thoroughly studied. The obtained results serve as a complement to its binary counterpart, the classical configurational model for arbitrary degree sequences [15]. We develop the full edge

and node statistics as well as first-order correlations using an application of a general ensemble approach developed in [11], providing not only average expected values for the observables but also precise bounds for its fluctuations and we compare the obtained results with simulations, yielding excellent agreement. By particularizing our general results to the case of power-law-distributed strengths, commonly found in real data [16–20], we demonstrate how the null-model expectations of some widely used weighted network metrics, which are generally considered a sign of relevant correlations (see [21] and references therein), can instead be seen as just a consequence of the particular form of the imposed strength sequence, and hence may not represent any unexpected property of the network under study.

Theoretical framework. – We start by considering a multi-edge undirected network with a fixed number of nodes N and hence a total of $L = N(N + 1)/2$ possible edges. Each edge has an associated integer weight $t_{ij} \in \{0, 1, 2, \dots\}$, and the strength of a node i is defined as $s_i = \sum_j t_{ij}$. We further consider a fixed strength sequence $\{\hat{s}_i\}; i \in 1, N$, so that we have a total of $\hat{T} = \sum_i \hat{s}_i$ distinguishable events to be randomly allocated among L edges. Note that the difference between distinguishable and non-distinguishable events is crucial, as it happens usually in statistical mechanics, because the resulting statistics of the ensembles one can consider depend on this property (see [11,22] for an extended discussion). As for the notation, note also that in general we will use \hat{x} to denote the value to which a variable x is being constrained.

To clarify the elements of our system, we can specify them for the case of an OD study as an illustrative example: Nodes correspond to different locations, events to individual trips between locations, the strengths of nodes to the fixed amount of travelers departing/entering each location and the weights correspond to the observed flows between locations. In this work we consider the case where self-loops are allowed for simplicity but the methodology can be easily extended to the case where no self-loops are accepted or even to directed networks.

Under the circumstances described, in analogy to statistical mechanics for classical systems, one can consider the Grand-Canonical (GC) ensemble of graphs (see [23,24] for different examples of other network ensembles) fulfilling *on average* the proposed constraints, *i.e.* $\{\langle s_i \rangle = \sum_j \langle t_{ij} \rangle = \hat{s}_i\}$ (throughout this letter $\langle x \rangle$ will refer to the ensemble average of random variable x whereas \bar{x} will refer to the graph average of variable x over a single realization). This means that both the occupation numbers or weights t_{ij} and the node strengths $s_i = \sum_j t_{ij}$ are integer random variables, and that each network belonging to the ensemble corresponds to a different realization of such variables. However, relative fluctuations around the constraints $\langle s_i \rangle = \hat{s}_i$ vanish in the large event limit [11]. The GC ensemble yields independent Poisson statistics for the occupation numbers t_{ij} with mean $\langle t_{ij} \rangle = \beta x_i x_j$, where $\{x_i\}$ can be considered as node-specific *hidden*

variables [25–27]. The values of $\{x_i\}$ can be obtained by solving the N saddle point equations $\hat{s}_i = \beta x_i \sum_j x_j$ that define the constraints of the system, which correspond to fixing the (ensemble) average strength of each node in the network. This set of equations is easily solvable analytically yielding $x_i = \hat{s}_i \forall i$ and $\beta = \hat{T}^{-1}$. In this way the expression $\langle t_{ij} \rangle = \hat{s}_i \hat{s}_j / \hat{T}$ is reached: the left-hand side is the ensemble average of a random variable, while the right-hand side is a result expressed in term of the constraints. In our case the $\{\hat{s}_i\}$ are the only fixed quantities and hence we must take them as the basic variables from which to derive the rest of weighted network properties: All nodes sharing the same strength value $\hat{s}_i = \hat{s}$ are statistically equivalent, and possess self-averaging properties (likewise all edges connecting nodes with the same pair of strength values). In what follows, we show how to proceed to obtain some particular network metrics, although the procedure is fully general and permits to obtain any desired property.

General methodology. – Many network metrics widely used in the literature can be written as a quotient of functions of the edge weights $\mathcal{M} = x(t_{ij})/y(t_{ij})$. In our framework, $\{t_{ij}\}$ are random variables and computing $\langle \mathcal{M} \rangle$ is not straightforward. We thus need to rely on approximations, expanding the expressions in Taylor series around their mean values and then taking the ensemble average of the first terms of the sum:

$$\langle \mathcal{M} \rangle \simeq \frac{\langle x \rangle}{\langle y \rangle} \left(1 + \frac{\langle y^2 \rangle}{\langle y \rangle^2} - \frac{\langle xy \rangle}{\langle x \rangle \langle y \rangle} \right), \quad (1)$$

$$\sigma_{\mathcal{M}}^2 \simeq \frac{\langle x \rangle^2}{\langle y \rangle^2} \left(\frac{\langle x^2 \rangle}{\langle x \rangle^2} + \frac{\langle y^2 \rangle}{\langle y \rangle^2} - 2 \frac{\langle xy \rangle}{\langle x \rangle \langle y \rangle} \right). \quad (2)$$

These expressions can be used to compute expected values and fluctuations of any network metric expressed as a ratio of functions of the occupation numbers $x(t_{ij}), y(t_{ij})$ provided that the moments ($\langle x \rangle, \langle x^2 \rangle, \langle y \rangle, \langle y^2 \rangle$) in the right-hand side can be evaluated. This is usually the case when $x(t_{ij}), y(t_{ij})$ are algebraic expressions of $\{t_{ij}\}$ (which are uncorrelated random variables). For most metrics \mathcal{M} considered in this letter, the calculations of the moments of $x(t_{ij})$ and $y(t_{ij})$ are lengthy, but follow from a general methodology without further difficulty¹. We thus subsume them in the appendix also explaining the general strategy used in their calculation, stating here only the key results.

Distribution of weights. – We start by computing the distribution of occupation numbers or *weights*

$$P(t) = \frac{1}{E} \sum_{ij} \delta(t, t_{ij}) \quad (3)$$

($E = \sum_{ij} \Theta(t_{ij})$ refers to the total number of existing edges), which has been reported to have broad

¹And can be easily implemented using any standard mathematical symbolic software.

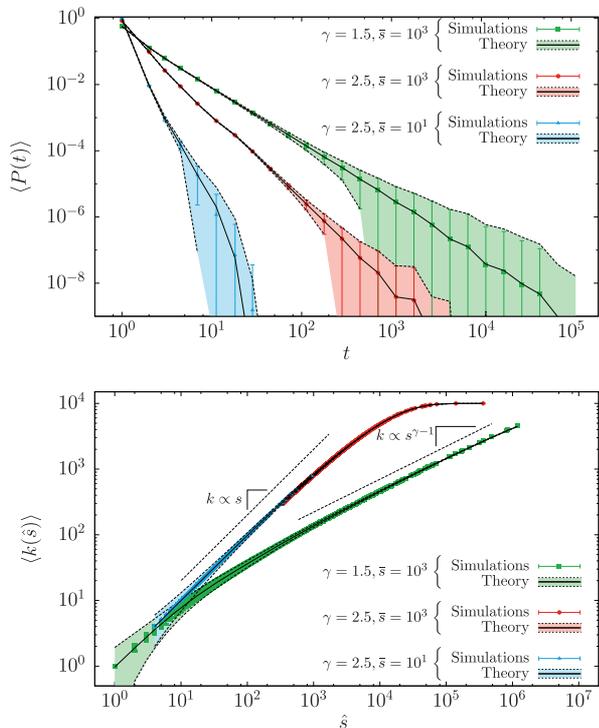


Fig. 1: (Color online) The effect of the strength sequence on network observables. Top: ensemble average of distribution of occupation numbers over existing edges (log-binned) and analytical predictions given by expression (4) and its standard deviation (see appendix, eq. (A.1)) for power-law-distributed strength sequences with $N = 10^4$ and different exponents for 1000 repetitions. The dependence on sampling $\bar{s} = \hat{T}/N$ is apparent. Bottom: degree-strength relationship for the same networks as earlier (average) and theoretical predictions from eq. (5). Standard deviation are represented as error bars and lines of constant slope are provided as a guide to the eye.

forms on empirical data for airport flow [16], cargo ship transport [17], public transport in cities [18], commuting [19] or face to face interactions [20] among others.

Applying eq. (1) to the case of $P(t)$ yields

$$\langle P(t) \rangle = \left\langle \frac{\sum_{ij} \delta(t, t_{ij})}{\sum_{ij} \Theta(t_{ij})} \right\rangle \simeq \frac{\sum_{ij} e^{-\langle t_{ij} \rangle} \langle t_{ij} \rangle^t}{t! \langle E \rangle} + \mathcal{O}(\langle E \rangle^{-2}), \quad (4)$$

being $\Theta(x)$ the Heaviside step function. Figure 1 shows the distribution of occupation numbers for existing edges and its associated standard deviation (see eq. (A.1) in the appendix) for three networks generated using power-law-distributed strength sequences ($\gamma = 1.5, 2.5$) and different graph-average strength $\bar{s} = \hat{T}/N$. We can see that the form of the resulting distribution is broad due to the imposed form of the strength sequence, and hence is not a sign *per se* of any interesting property of the multi-edge network being studied. Moreover, its shape strongly depends on the total number of observed events \hat{T} .

Degrees and strengths. – Having tackled the occupation number statistics of the network, in what follows

we consider its node-related properties. We have that the strengths $s_i = \sum_j t_{ij}$ will also be Poisson-distributed random variables, being sums of independent occupation numbers. Moreover, since the binary projection of occupation numbers $\Theta(t_{ij})$ are Bernoulli-distributed variables with parameter $P(t_{ij} > 0) = 1 - e^{-\langle t_{ij} \rangle}$ [11] one can also compute the associated degrees k_i of the nodes, which will be sums of independent Bernoulli random variables². We have that

$$\langle k(\hat{s}_i) \rangle = \left\langle \sum_j \Theta(t_{ij}) \right\rangle = \sum_j P(t_{ij} > 0) = \sum_j \left(1 - e^{-\langle t_{ij} \rangle} \right) = N - \sum_j e^{-\frac{\hat{s}_i \hat{s}_j}{\hat{T}}}, \quad (5)$$

$$\sigma_{k(\hat{s}_i)}^2 = \sum_j \sigma_{\Theta(t_{ij})}^2 = \sum_j e^{-\langle t_{ij} \rangle} \left(1 - e^{-\langle t_{ij} \rangle} \right) = N - \langle k(\hat{s}_i) \rangle - \sum_j e^{-2\frac{\hat{s}_i \hat{s}_j}{\hat{T}}}, \quad (6)$$

which constitutes an extremely accurate prediction (see fig. 1 lower panel). The asymptotic cases for the ensemble averages are easy to assess: For small strengths we have that $\hat{s} \ll \hat{T}/\hat{s}' \forall \hat{s}'$ which leads expression (5) to $\langle k(\hat{s}) \rangle \sim \hat{s}$ (converging to a Poisson Bernoulli distribution), while for large strengths one has $\hat{s} \hat{s}' \gg \hat{T} \forall \hat{s}'$ which leads to fully connected nodes $\langle k(\hat{s}) \rangle \sim N$ with vanishing variance.

Results comparing simulations and eq. (5) are shown in fig. 1 (lower panel), where an interesting transition is observed for $\gamma < 2$: The degrees are exactly equal to the strengths for small values of \hat{s} (as expected by conservation of the edges) and evolve to a scaling of the type $k(\hat{s}) \sim \hat{s}^{\gamma-1}$ that finally leads to a saturation due to the bounded nature of the observables ($k(\hat{s}) \leq N$).

The important result to take home here is that *a scaling relation of the kind $k(s) \sim s^\beta$ is not always a reliable trace of correlations*. More concretely, we have seen that in our framework, and for the case of power-law-distributed strength sequences in particular, it is solely a consequence of the imposed constraints. In other cases, it might or might not be an indicator of correlations, but one cannot assume either case *a priori*: Since this metric heavily depends on the strength sequence, it always requires comparison with a null model.

Disparity, average neighbor properties and general metrics. – In recent times, efforts have been devoted to extend well-known magnitudes on binary graphs to weighted graphs: Having appropriate null models for multi-edge graphs permits to assess the applicability of such *weighted* extensions [29]. To this end, one can use the results of the GC ensemble to compute with high accuracy any network metric expressed in terms of t_{ij} : As an

²The distribution of such variables is called *Poisson Bernoulli* and has well-studied properties [28], albeit their moments are difficult to compute. One can, however, give bounds to the error committed whenever assuming a Poisson approximation also for the degrees.

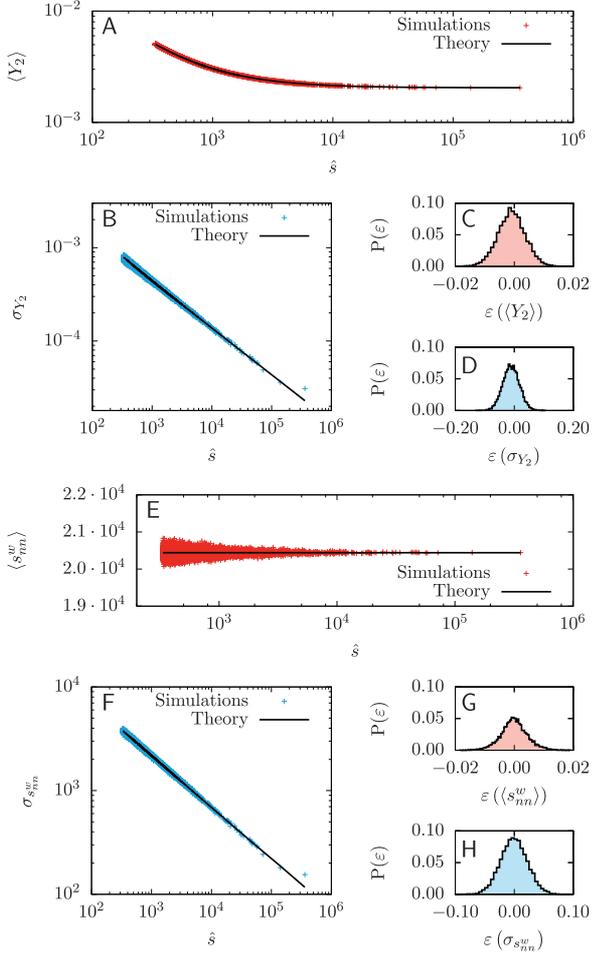


Fig. 2: (Color online) The accuracy of the GC predictions. Ensemble average (a), (e) and standard deviation (b), (f) for individual node disparity Y_2 and weighted neighbor average strength s_{nn}^w for power-law-distributed sequences with $\bar{s} = 1000$ and $\gamma = 2.5$. (c), (d), (g), (h): histogram of relative error between theory and simulations averaged over 1000 repetitions for the values in panels (a), (e) and (b), (f). A single outlier corresponding to the lowest value of σ for both the disparity and the average neighbor strength is not shown in the histogram in panels (c), (g).

example we consider widely used magnitudes such as the disparity $Y_2(s_i) = \sum_j t_{ij}^2/s_i^2$ [30] and weighted neighbor average strength $s_{nn}^w(s_i) = \sum_j t_{ij}s_j/s_i$ [31]. Using again eq. (1), and the fact that $\{t_{ij}\}$ are a set of Poisson-distributed, *independent* random variables, one reaches after some algebra the following expressions:

$$\langle Y_2(\hat{s}_i) \rangle \simeq \frac{1 + \frac{\hat{T}_2}{\hat{T}^2} \hat{s}_i}{1 + \hat{s}_i} \left(1 + \frac{(\hat{T}^2 - \hat{T}_2)(2\hat{s}_i + 3)}{(\hat{s}_i + 1)^2 (\hat{T}_2 \hat{s}_i + \hat{T}^2)} \right), \quad (7)$$

$$\langle s_{nn}^w(\hat{s}_i) \rangle \simeq \left(1 + \frac{\hat{T}_2}{\hat{T}} \right) \left(1 - \frac{\hat{T}_2 + \hat{T} \hat{s}_i - \hat{s}_i^2}{\hat{T}(\hat{T} + \hat{T}_2)} \right), \quad (8)$$

where $\hat{T}_n \equiv \sum \hat{s}_i^n$. The average values and their fluctuations are in excellent agreement with the simulations, as

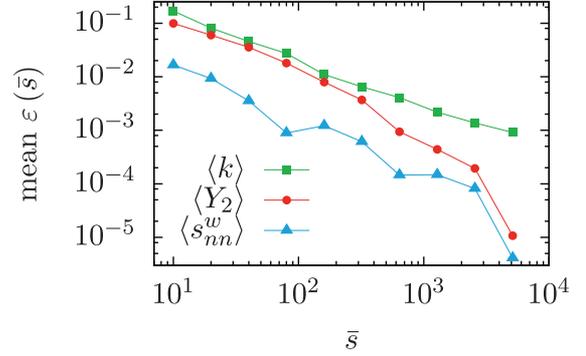


Fig. 3: (Color online) Convergence between ensembles with increased sampling. Relative error between ensemble average predictions and simulations, averaged over all nodes for degree, disparity and average neighbor strength for different values of sampling $\bar{s} = \hat{T}/N$ for 1000 repetitions each point, $\gamma = 2.5$ and $N = 2000$.

can be seen from fig. 2, panels (a), (b), (e), (f). The expressions corresponding to $\sigma_{s_{nn}^w}^2$ and $\sigma_{Y_2}^2$, and some details on how to compute them, can be found in the appendix.

The results show several interesting features: On the one hand, the expectation for the disparity is not $Y_2(\hat{s}) \sim k_i^{-1}$ as assumed under a total random allocation of edge weights [30], but rather decays as $Y_2 \sim \hat{s}_i^{-1}$ and rapidly converges to a plateau, independently of the chosen strength distribution. The weighted average neighbor strength displays an almost flat behavior which is a correct indicator of absence of correlations at the node level. On the other hand, the fluctuations of both magnitudes decay in a power-law form as the strength of the node increases: This fact can be easily understood as increased connectedness imply higher availability of sampling.

Simulations. – To quantify the precision of our predictions, we computed the histograms of the relative error generated per node when calculating a given property z , $\epsilon(z) = (\langle z \rangle_{si} - \langle z \rangle_{th}) / \langle z \rangle_{si}$, where the subindices si stand for the Micro-Canonical (MC) simulations³ and th for the GC theoretical predictions in eqs. (7) and (8). The histograms in fig. 2, panels (c), (d), (g), (h) show the accuracy of the obtained results, providing numerical evidence for the equivalence between the MC simulations and the GC predictions, which is expected in the thermodynamic limit when an infinite sampling of events $\hat{T} \rightarrow \infty$ is available. Even when this requisite is not met, the use of the theory presented here constitutes an excellent approximation as shown in fig. 3, where the relative error averaged over all nodes between ensemble expected GC magnitudes and simulations is shown for the different metrics considered for a wide range of values of sampling.

³The simulations have been performed using a general configuration model rewiring schema with allowed self-loops and multiple collections between links [32]. The code for the generation of multi-edge networks and details on the implementation can be found in [33].

Conclusions. – In the present letter, we have provided a theoretical framework for multi-edge networks with a fixed strength sequence that can be used to assess the impact of this structural constraint on the observables of real networks. We provide closed statistical forms for the edge and node statistics by means of a GC ensemble formalism, which coincide with great accuracy with simulations in the MC ensemble, for which we additionally provide the software in [33].

The general results for any given strength sequence have been studied, and precise analytical closed forms of both the average expected values and fluctuations of some widely used network metrics such as the existing occupation number distribution, node degree and strength, disparity and first-order neighbor correlations have been obtained. We have further used the case of power-law-distributed sequences as an example to show that the effect of the skewness of this distribution over some widely used network metrics can explain some of the correlations detected in real weighted complex networks.

The effects of the strength sequences are remarkable in all the metrics considered, which stresses the need for real data to be tested against appropriate null models to assess the relevance of observed properties. Interestingly, second-order neighbor correlation properties such as clustering coefficients and its different weighted versions [10] could be similarly computed with the methodology presented, albeit they lead to more complicated expressions. We leave this to future research.

The blessing of big data may also be its dearest danger: High-dimensionality data sets require sophisticated null models to detect the effects of the system constraints on the given observables and hence comparison with a null model to assess the relevance of observed features in real data is always needed. The present application of ensemble theory to networks yielding exact results for both average values and fluctuations aims to draw attention to this problem and to close this gap for the case of multi-edge networks.

We thank M. SZELL for comments and suggestions. This work has been partially supported by the EU-LASAGNE Project, Contract No. 318132 (STREP), the Spanish MINECO Grants FIS2012-38266-C02-02 and FIS2012-31324 and by the Generalitat de Catalunya 2009-SGR-00838. OS and FF have been supported by the Generalitat de Catalunya through the FI Program.

APPENDIX

We gather here the expressions for the standard deviation of all the metrics considered:

$$\sigma_P^2(t) \simeq \frac{\mathcal{Q}_1}{\langle E \rangle^2} \left(1 - \frac{\mathcal{Q}_2}{\mathcal{Q}_1} + \frac{1}{\langle E \rangle \mathcal{Q}_1} \left[1 - 2 \frac{\mathcal{R}_1}{\mathcal{Q}_1} \right] - \frac{\mathcal{R}_2}{\langle E \rangle^2 \mathcal{Q}_1} \right), \quad (\text{A.1})$$

$$\sigma_{Y_2}^2 \simeq \hat{T}_1^{-2} \left(\frac{a_3 \hat{s}_i^3 + a_2 \hat{s}_i^2 + a_1 \hat{s}_i + a_0}{(\hat{s}_i + 1)^4} \right), \quad (\text{A.2})$$

$$\sigma_{s_{nn}}^2 \simeq b_{-1} \frac{1}{\hat{s}_i} + b_0 + b_2 \hat{s}_i^2 + b_3 \hat{s}_i^3, \quad (\text{A.3})$$

where E is the number of present edges and we have defined the following notation:

$$p_{ij}(t) \equiv e^{-\langle t_{ij} \rangle} \langle t_{ij} \rangle^t / t!, \quad \langle t_{ij} \rangle = \hat{s}_i \hat{s}_j / \hat{T}, \quad (\text{A.4})$$

$$\mathcal{Q}_1 = \sum_{ij} p_{ij}(t), \quad \mathcal{R}_1 = \sum_{ij} p_{ij}(t) p_{ij}(0), \quad (\text{A.5})$$

$$\mathcal{Q}_2 = \sum_{ij} (p_{ij}(t))^2, \quad \mathcal{R}_2 = \sum_{ij} (1 - p_{ij}(0))^2, \quad (\text{A.6})$$

$$a_3 = -4 \left(\hat{T}_2^2 - \hat{T}_1 \hat{T}_3 \right), \quad b_2 = 2 \hat{T}_2 / \hat{T}_1^3, \quad (\text{A.7})$$

$$a_0 = 2 \hat{T}_1^4 - 2 \hat{T}_2 \hat{T}_1^2, \quad b_3 = -2 / \hat{T}_1^2, \quad (\text{A.8})$$

$$a_2 = 2 \left(\left[\hat{T}_1^2 - 5 \hat{T}_2 \right] \hat{T}_2 + 4 \hat{T}_1 \hat{T}_3 \right), \quad (\text{A.9})$$

$$a_1 = \hat{T}_1^4 + 2 \hat{T}_2 \hat{T}_1^2 + 4 \hat{T}_3 \hat{T}_1 - 7 \hat{T}_2^2, \quad (\text{A.10})$$

$$b_{-1} = \left(\hat{T}_1 \left[\hat{T}_2 + \hat{T}_3 \right] - \hat{T}_2^2 \right) \hat{T}_1^{-2}, \quad (\text{A.11})$$

$$b_0 = \left(-\hat{T}_2^2 + \hat{T}_1 \left[\hat{T}_2 + 3 \hat{T}_3 \right] - \hat{T}_4 \right) \hat{T}_1^{-3}, \quad (\text{A.12})$$

with $\hat{T}_n \equiv \sum_i \hat{s}_i^n$. The calculations leading to these results are admittedly tedious, but follow directly from eqs. (1) and (2) and are of no particular difficulty beyond algebraic manipulation and careful reordering of the sums. As an illustrative example, consider the disparity $Y_2(s_i)$ for node i , defined as

$$Y_2(s_i) = \sum_j \frac{t_{ij}^2}{s_i^2}. \quad (\text{A.13})$$

Identifying $x \equiv \sum_{j \neq i} t_{ij}^2$ and $y \equiv s_i^2$, eq. (1) can be readily applied. In order to approximate $\langle Y_2(s_i) \rangle$ and σ_{Y_2} , we need to compute $\langle x \rangle$, $\langle y \rangle$, $\langle x^2 \rangle$, $\langle y^2 \rangle$ and $\langle xy \rangle$. Let us show in full detail, as an illustrative example, how to compute $\langle x^2 \rangle$ in this case. First, we expand x^2 as follows:

$$x^2 = \sum_{j,k} t_{ij}^2 t_{ik}^2 = \sum_{\substack{j \neq i, \\ k \neq j, i}} t_{ij}^2 t_{ik}^2 + \sum_{j \neq i} \left[t_{ij}^4 + 2 t_{ij}^2 t_{ii}^2 \right] + t_{ii}^4, \quad (\text{A.14})$$

so that when the ensemble average is taken, all products factorize (they correspond to different pairs of values ij , which are independent),

$$\langle x^2 \rangle = \sum_{\substack{j \neq i, \\ k \neq j, i}} \langle t_{ij}^2 \rangle \langle t_{ik}^2 \rangle + \sum_{j \neq i} \left[\langle t_{ij}^4 \rangle + 2 \langle t_{ij}^2 \rangle \langle t_{ii}^2 \rangle \right] + \langle t_{ii}^4 \rangle. \quad (\text{A.15})$$

Finally, since the variables t_{ij} are Poisson-distributed, we can compute their moments ($\langle t_{ij}^2 \rangle = \langle t_{ij} \rangle (1 + \langle t_{ij} \rangle)$), and using that $\langle t_{ij} \rangle = \hat{s}_i \hat{s}_j / \hat{T}$, and after some algebra,

we get to

$$\langle x^2 \rangle = \frac{\hat{T}_2^2 \hat{s}_i^4}{\hat{T}_1^4} + \left(\frac{2\hat{T}_2}{\hat{T}_1^2} + \frac{4\hat{T}_3}{\hat{T}_1^3} \right) \hat{s}_i^3 + \left(\frac{6\hat{T}_2}{\hat{T}_1^2} + 1 \right) \hat{s}_i^2 + \hat{s}_i. \quad (\text{A.16})$$

The rest of the terms can be computed in a similar vein, leading to

$$\langle x \rangle = \frac{\hat{T}_2 \hat{s}_i^2}{\hat{T}_1^2} + \hat{s}_i, \quad \langle y \rangle = \hat{s}_i^2 + \hat{s}_i, \quad (\text{A.17})$$

$$\langle xy \rangle = \frac{\hat{T}_2 \hat{s}_i^4}{\hat{T}_1^2} + \left(\frac{5\hat{T}_2}{\hat{T}_1^2} + 1 \right) \hat{s}_i^3 + \left(\frac{4\hat{T}_2}{\hat{T}_1^2} + 3 \right) \hat{s}_i^2 + \hat{s}_i, \quad (\text{A.18})$$

$$\langle y^2 \rangle = \hat{s}_i^4 + 6\hat{s}_i^3 + 7\hat{s}_i^2 + \hat{s}_i. \quad (\text{A.19})$$

Finally, inserting eqs. (A.16)–(A.19) into eq. (1) and some simplification leads to the desired result, eqs. (7) and (A.2).

REFERENCES

- [1] SERRANO M. Á., BOGUÑÁ M. and VESPIGNANI A., *Proc. Natl. Acad. Sci. U.S.A.*, **106** (2009) 6483.
- [2] STEHLÉ J., CHARBONNIER F., PICARD T., CATTUTO C. and BARRAT A., *Soc. Netw.*, **35** (2013) 604.
- [3] BARTHÉLEMY M., *Phys. Rep.*, **499** (2011) 1.
- [4] UHLMANN S., MANNSPERGER H., ZHANG J., HORVAT E., SCHMIDT C., KÜBLBECK M., HENJES F., WARD A., TSCHULENA U., ZWEIG K., KORF U., WIEMANN S. and SAHIN O., *Mol. Syst. Biol.*, **8** (2012) 570.
- [5] CAMACHO J., GUIMERA R. and AMARAL L. A. N., *Phys. Rev. Lett.*, **88** (2002) 228102.
- [6] ZLATIĆ V., GABRIELLI A. and CALDARELLI G., *Phys. Rev. E*, **82** (2010) 066109.
- [7] NEWMAN M., *Phys. Rev. E*, **70** (2004) 056131.
- [8] GÓMEZ S., DÍAZ-GUILERA A., GÓMEZ-GARDEÑES J., PÉREZ-VICENTE C. J., MORENO Y. and ARENAS A., *Phys. Rev. Lett.*, **110** (2013) 028701.
- [9] DE DOMENICO M., SOLÉ-RIBALTA A., COZZO E., KIVELÄ M., MORENO Y., PORTER M. A., GÓMEZ S. and ARENAS A., *Phys. Rev. X*, **3** (2013) 041022.
- [10] SARAMÄKI J., KIVELÄ M., ONNELA J.-P., KASKI K. and KERTÉSZ J., *Phys. Rev. E*, **75** (2007) 027105.
- [11] SAGARRA O., PÉREZ VICENTE C. J. and DÍAZ-GUILERA A., *Phys. Rev. E*, **88** (2013) 062806.
- [12] SQUARTINI T. and GARLASCHELLI D., *New J. Phys.*, **13** (2011) 083001.
- [13] ANNIBALE A., COOLEN A., FERNANDES L., FRATERNALI F. and KLEINJUNG J., *J. Phys. A: Math. Gen.*, **42** (2009) 485001.
- [14] BIANCONI G., PIN P. and MARSILI M., *Proc. Natl. Acad. Sci. U.S.A.*, **106** (2009) 11433.
- [15] MOLLOY M. and REED B., *Random Struct. Algorithms*, **6** (1995) 161.
- [16] COLIZZA V., BARRAT A., BARTHÉLEMY M. and VESPIGNANI A., *Proc. Natl. Acad. Sci. U.S.A.*, **103** (2006) 2015.
- [17] KALUZA P., KÖLZSCH A., GASTNER M. T. and BLASIUS B., *J. R. Soc. Interface*, **7** (2010) 1093.
- [18] ROTH C., KANG S. M., BATTY M. and BARTHÉLEMY M., *PLoS ONE*, **6** (2011) e15923.
- [19] DE MONTIS A., BARTHÉLEMY M., CHessa A. and VESPIGNANI A., *Geographical*, **34** (2005) 12.
- [20] BARRAT A., CATTUTO C., COLIZZA V., GESUALDO F., ISELLA L., PANDOLFI E., PINTON J. F., RAVÀ L., RIZZO C., ROMANO M., STEHLÉ J., TOZZI A. E. and BROECK W., *Eur. Phys. J. ST*, **222** (2013) 1295.
- [21] POPOVIĆ M., ŠTEFANČIĆ H. and ZLATIĆ V., *Phys. Rev. Lett.*, **109** (2012) 208701.
- [22] GARLASCHELLI D. and LOFFREDO M. I., *Phys. Rev. Lett.*, **102** (2009) 038701.
- [23] BIANCONI G., *Phys. Rev. E*, **79** (2009) 036114.
- [24] SQUARTINI T., FAGIOLO G. and GARLASCHELLI D., *Phys. Rev. E*, **84** (2011) 046118.
- [25] GOH K.-I., KAHNG B. and KIM D., *Phys. Rev. Lett.*, **87** (2001) 278701.
- [26] CALDARELLI G., CAPOCCI A., DE LOS RIOS P. and MUÑOZ M. A., *Phys. Rev. Lett.*, **89** (2002) 258702.
- [27] BOGUÑÁ M. and PASTOR-SATORRAS R., *Phys. Rev. E*, **68** (2003) 36112.
- [28] SERFLING R., *Siam Rev.*, **20** (1978) 567.
- [29] AHNERT S. E., GARLASCHELLI D., FINK T. M. A. and CALDARELLI G., *Phys. Rev. E*, **76** (2007) 16101.
- [30] SERRANO M. A., BOGUNA M. and PASTOR-SATORRAS R., *Phys. Rev. E*, **74** (2006) 055101.
- [31] BARRAT A., BARTHÉLEMY M., PASTOR-SATORRAS R. and VESPIGNANI A., *Proc. Natl. Acad. Sci. U.S.A.*, **101** (2004) 3747.
- [32] SERRANO M. Á., *Weighted Configuration Model*, in *AIP Conf. Proc.*, Vol. **776** (AIP) 2005, pp. 101–107.
- [33] SAGARRA O., *Multi Edge Network Generator*, http://github.com/osagarra/Multi_edge_randomizer (2014).