Reciprocity of networks with degree correlations and arbitrary degree sequences

Gorka Zamora-López,[1] Vinko Zlatić,[2,][*] Changsong Zhou,[3] Hrvoje Štefančić,[2] and Jürgen Kurths[1]

[1]*Institute of Physics, University of Potsdam, PF 601553, 14415 Potsdam, Germany*
[2]*Theoretical Physics Division, Rudjer Bošković Institute, P.O. Box 180, HR-10002 Zagreb, Croatia*
[3]*Department of Physics, Hong Kong Baptist University, Kowloon Tong, Hong Kong*

Although most of the real networks contain a mixture of directed and bidirectional (reciprocal) connections, the reciprocity r has received little attention as a subject of theoretical understanding. We study the expected reciprocity of networks with arbitrary input and output degree sequences and given 2-node degree correlations by means of statistical ensemble approach. We demonstrate that degree correlations are crucial to understand the reciprocity in real networks and a hierarchy of correlation contributions to r is revealed. Numerical experiments using network randomization methods show very good agreement to our analytical estimations.

I. INTRODUCTION

Most of the real networks combine both unidirectional and bidirectional (reciprocal) connections. This directed nature is often obviated, e.g., networks are symmetrized for algorithmic convenience and network models largely ignore the directionality for analytical simplicity. However, the directionality is known to be relevant, e.g., robustness against environmental changes of metabolic networks seems to arise from evolutionary pressure on the directions and weights of the metabolic fluxes [1]. The formation of functional communities and hierarchies in the cerebral cortex is mediated by the presence of reciprocal and unidirectional connections [2]. The dynamical stability in complex networks, e.g., ecological systems [3,4] and synchronization of coupled oscillators, is commonly assessed by the eigenvalue space of the Jacobian or Laplacian matrices [5]. When networks are directed, these matrices will have complex eigenvalues, which influences both the stability and the dynamical organization far from the equilibrium state.

The network reciprocity r is classically defined as $r = \frac{L^{\leftrightarrow}}{L}$ [6] where $L^{\leftrightarrow}$ is the number of directed links $s \to t$ that also have a reciprocal (bidirectional) counterpart $s \leftarrow t$, and L is the total number of directed links. In networks without self-loops, reciprocal links form the cycles of lowest order and are therefore, important as a natural measure of feedback in the network. Recently, r of the Wikipedia networks [29] (for different languages) was found to be very stable over a wide range of network sizes [7] which signals its relevance for the structure or functionality of the networks. In [8] it was shown that reciprocal connections carry most of the topological information of the world wide web. The formation of the giant component in directed networks is facilitated by reciprocal connections [9]. Despite the extensive modeling efforts during the recent years to reproduce realistic features of networks, models have largely ignored reciprocity. Only in [10] a general class of random networks with prescribed r has been presented.

When analyzing real networks it is important to test whether measured values are significant or not. Usually, the properties of a real network are compared to those of complete random networks of the same size N and number of links L. However, the degree distribution of most real networks largely differs from Poissonian, i.e., the degree distribution of random graphs, and more reasonable comparison is desired. Obtaining analytical expressions for expected measures under conditions of arbitrary degree sequence is difficult and further assumptions are typically introduced to derive approximate estimations, e.g., scale-free [11] or exponential degree distributions [12]. In real applications a numerical approach is, most of the time, the only solution. Ensembles of maximally random networks can be generated with the same degree sequences and the ensemble average properties can be calculated. Unfortunately, this is often computationally very demanding. In this paper we present analytical expressions for the expected reciprocity of directed networks that can be evaluated using only information measured from the real network under study, and thus, overcome the problems discussed above.

In the context of social networks, several redefinitions of r have been introduced that account for biases of the experimental conditions [6,13]. Recently, a similar redefinition based on the correlation between the adjacency matrix and its transpose has been presented [14]. It evaluates the reciprocity with respect to the density of connections $\bar{a} = L/[N(N-1)]$. This results from the fact that in a random digraph with N nodes and L links, $r = \bar{a}$. However, the specific degree sequence is expected to affect the number of reciprocal links. A node with both large input degree k_i and large output degree k_o has a higher tendency to form reciprocal links. This is expressed by the 1-node degree correlations between k_i and k_o of individual nodes. Similarly, as r involves the pairwise connectivity of nodes, the correlation between the degrees of neighboring nodes, 2-node degree correlations, should also be relevant. Imagine a link $s \to t$ connecting a source node s with input and output degrees (k_i, k_o) to a target node t with degrees (q_i, q_o). The 2-node degree correlations exist when any of the (k_i, k_o) degrees of node s are correlated with any of the (q_i, q_o) degrees of node t. When all four values are correlated, then both 1-node ($1n$) correlations and the 2-node 4-degree ($2n4d$) correlations are present. The class of 2-node 2-degree ($2n2d$) correlations is depicted in Fig. 1.

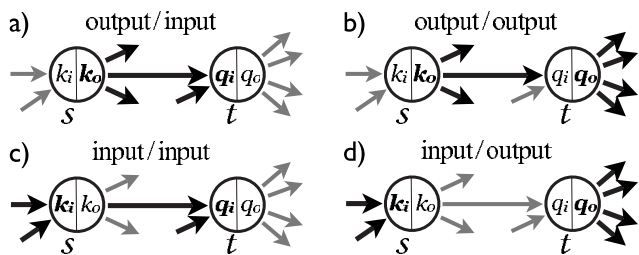*Corresponding author. vzlatic@irb.hr

FIG. 1. 2-node 2-degree correlations ($2n2d$) of neighboring nodes in directed networks. Links corresponding to correlated degrees are colored black.

In this paper we study the expected reciprocity $\langle r \rangle$ of networks with prescribed degree sequences and arbitrary 2-node degree correlations. We consider the complex networks as members of the statistical ensemble with given node degree sequence and degree correlations, and we calculate the expected reciprocity of such ensembles in the thermodynamical limit [30]. We find that degree correlations explain almost completely the observed r of some real networks. In other examples, larger discrepancies indicate the presence of additional internal structure.

II. GENERAL RESULTS

In order to analytically estimate $\langle r \rangle$ under different correlation structures, we characterize the real directed networks by: the number of nodes N, the number of links L, the number of nodes $N(k_i, k_o) = N(\mathbf{k})$ having in-degree k_i and out-degree k_o, and the number of directed links $L(\mathbf{k} \rightarrow \mathbf{q})$ pointing from nodes with degrees (k_i, k_o) to nodes with degrees (q_i, q_o). All these properties are easy to measure in a real directed network and contain all the relevant information about the degree correlations. We use frequencies of these properties as their probabilities and calculate the expected number of reciprocal links $\langle L^{\leftrightarrow} \rangle$. We remind the reader that by definition, $\langle r \rangle$ is related to $\langle L^{\leftrightarrow} \rangle$ by $\langle r \rangle = \langle L^{\leftrightarrow} \rangle / L$.

Under the class of 1-node and 2-node degree correlations here assumed, a network is considered as maximally random when any of the nodes with degrees $\mathbf{k}$ are equally likely connected to any of the nodes with degrees $\mathbf{q}$. If a network contains $L(\mathbf{k} \rightarrow \mathbf{q})$ links of the type $\mathbf{k} \rightarrow \mathbf{q}$, the probability that any of them connects a pair of randomly chosen nodes with degrees $\mathbf{k}$ and $\mathbf{q}$ is, in the thermodynamical limit,

$$p(\mathbf{k} \rightarrow \mathbf{q}) = \frac{L(\mathbf{k} \rightarrow \mathbf{q})}{N(\mathbf{k})N(\mathbf{q})}. \quad (1)$$

The denominator $N(\mathbf{k})N(\mathbf{q})$ is the number of all possible connections between nodes with degrees $\mathbf{k}$ and nodes with degrees $\mathbf{q}$.

Again, if the network has $L(\mathbf{k} \rightarrow \mathbf{q})$ links of the type $\mathbf{k} \rightarrow \mathbf{q}$, then the expected number of reciprocal $\mathbf{k} \leftrightarrow \mathbf{q}$ links is $\langle L(\mathbf{k} \leftrightarrow \mathbf{q}) \rangle = L(\mathbf{k} \rightarrow \mathbf{q}) p(\mathbf{k} \leftarrow \mathbf{q})$. The overall expected reciprocity r_{1n2n} of the network is obtained by summing $\langle L(\mathbf{k} \leftrightarrow \mathbf{q}) \rangle$ over all $\mathbf{k}, \mathbf{q}$ combinations,

$$r_{1n2n} = \frac{1}{L} \sum_{k,q} \frac{L(\mathbf{k} \rightarrow \mathbf{q})L(\mathbf{k} \leftarrow \mathbf{q})}{N(\mathbf{k})N(\mathbf{q})}. \quad (2)$$

Note that in general $L(\mathbf{k} \rightarrow \mathbf{q}) \neq L(\mathbf{k} \leftarrow \mathbf{q})$. Taking frequencies of nodes $P(\mathbf{k}) = N(\mathbf{k})/N$ and frequencies of links $\mathcal{P}(\mathbf{k} \rightarrow \mathbf{q}) = L(\mathbf{k} \rightarrow \mathbf{q})/L$ as probabilities in the thermodynamical limit, Eq. (2) reads as

$$r_{1n2n} = \frac{L}{N^2} \sum_{k,q} \frac{\mathcal{P}(\mathbf{k} \rightarrow \mathbf{q})\mathcal{P}(\mathbf{k} \leftarrow \mathbf{q})}{P(\mathbf{k})P(\mathbf{q})}. \quad (3)$$

The contribution of the correlation structure is accounted by the sum, $\mathcal{P}(\mathbf{k} \rightarrow \mathbf{q})$ accounts for both 1-node and 2-node correlations, and $P(\mathbf{k})$ only for the 1-node correlations. When all four degrees are independent $\mathcal{P}(\mathbf{k} \rightarrow \mathbf{q}) = P(k_i)k_o P(k_o)q_i P(q_i)P(q_o)/\langle k \rangle^2$ where $\langle k \rangle = L/N$ is the average degree and $P(\mathbf{k}) = P(k_i)P(k_o)$. Then reciprocity reduces to the density of links L/N^2, i.e., the expected reciprocity of uncorrelated random networks.

III. SPECIAL CORRELATION CLASSES

Equations (2) and (3) are general formulas that account for all 1-node and 2-node degree correlations. These equations can be reduced to consider only desired special classes of correlations and thus explore the contribution of individual correlation types to r. In this section we present detailed derivations for all eight possible combinations of 1-node and 2-node correlations and in Table I the main results are summarized. Note that in Table I only r_c, the contribution of the correlation structure is shown. The expected reciprocities are obtained by multiplying with the density of links $\langle r \rangle = \frac{L}{N^2} r_c$. Along this section, the usual product rule of joint probabilities in terms of conditional probabilities will be used,

$$P(X_1, X_2, \ldots, X_n) = P(X_1)P(X_2|X_1)P(X_n|X_1, X_2, \ldots, X_{n-1}).$$

With this rule in mind, the joint probabilities associated with the link statistics $\mathcal{P}(\mathbf{k} \rightarrow \mathbf{q}) \equiv p(k_i, k_o, q_i, q_o)$ can be expressed, for example, as

$$\mathcal{P}(\mathbf{k} \rightarrow \mathbf{q}) = \mathcal{P}(k_o \rightarrow q_i)P(k_i|k_o \rightarrow q_i)P(q_o|k_o \rightarrow q_i, k_i).$$

A. 2-node output/input degree correlations

Suppose that a network has significant 1-node correlations and 2-node correlations only between the out-degree k_o of source nodes and the in-degree q_i of the target nodes, see Fig. 1(a), while other possible 2-node correlations are negligible. In this case, the probabilities can be approximated by

$$\mathcal{P}(\mathbf{k} \rightarrow \mathbf{q}) \approx \mathcal{P}(k_o \rightarrow q_i)P(k_i|k_o)P(q_o|q_i),$$

$$\mathcal{P}(\mathbf{k} \leftarrow \mathbf{q}) \approx \mathcal{P}(k_o \leftarrow q_i)P(k_o|k_i)P(q_i|q_o).$$

Using conditional probabilities for the degrees of individual nodes, e.g., $P(k_i|k_o) = P(\mathbf{k})/P(k_o)$, Eq. (3) reduces to

TABLE I. Reduced formulas for the expected reciprocity of different combinations of 1-node and 2-node correlations. The expected reciprocity is $\langle r \rangle = \frac{L}{N^2} r_c$. Note that the result for the 2-node out-out and in-in correlations are independent of the 1-node correlations.

1-node	2-node	r_c, contribution of the correlations
No	No	1
Yes	No	$\dfrac{\langle k_i k_o \rangle^2}{\langle k \rangle^4}$
No	Output/input	1
Yes	Output/input	$\displaystyle\sum_{\mathbf{k,q}} \dfrac{\mathcal{P}(k_o \to q_i)\mathcal{P}(k_i \leftarrow q_o)P(k_i,k_o)P(q_i,q_o)}{P(k_i)P(k_o)P(q_i)P(q_o)}$
Yes and/or no	Output/output	$\displaystyle\sum_{k_o,q_o} \dfrac{\mathcal{P}(k_o \to q_o)\mathcal{P}(k_o \leftarrow q_o)}{P(k_o)P(q_o)}$
Yes and/or no	Input/input	$\displaystyle\sum_{k_i,q_i} \dfrac{\mathcal{P}(k_i \to q_i)\mathcal{P}(k_i \leftarrow q_i)}{P(k_i)P(q_i)}$
No	Input/output	$\dfrac{(\langle k_i q_o \rangle_{\mathcal{P}})^2}{\langle k \rangle^4}$ [a]
Yes	Input/output	$\displaystyle\sum_{\mathbf{k,q}} k_i k_o q_i q_o \dfrac{\mathcal{P}(k_i \to q_o)\mathcal{P}(k_o \leftarrow q_i)P(k_i,k_o)P(q_i,q_o)}{\bar{k}_{i,k_o}\bar{k}_{o,k_i}\bar{q}_{i,q_o}\bar{q}_{o,q_i}}$ [b]

[a] The averaging in the formulas is performed over in-degree of source nodes and the out-degree of the target nodes.

[b] $\bar{k}_{o,k_i} = \sum_{k'_o} k'_o P(k'_o, k_i)$ and similar for all other averages of this type.

$$r_{1n2n:o/i} = \frac{L}{N^2} \sum_{\mathbf{k,q}} \frac{\mathcal{P}(k_o \to q_i)\mathcal{P}(k_i \leftarrow q_o)P(\mathbf{k})P(\mathbf{q})}{P(k_i)P(k_o)P(q_i)P(q_o)}. \quad (4)$$

Additionally, if the 1-node correlations are negligible, the degrees of individual nodes become independent: $P(\mathbf{k}) = P(k_i, k_o) = P(k_i)P(k_o)$. Then Eq. (4) becomes

$$r_{2n:o/i} = \frac{L}{N^2} \sum_{\mathbf{k,q}} \mathcal{P}(k_o \to q_i)\mathcal{P}(k_i \leftarrow q_o) = \frac{L}{N^2}, \quad (5)$$

that equals the density of connections $\bar{a}$ in the thermodynamical limit. This means that the 2-node output/input degree correlations do not contribute to reciprocity in the absence of 1-node correlations. This is not a general case since other classes of 2-node correlations largely contribute to r.

B. 1-node degree correlations

Starting from Eq. (4), we can alternatively remove the remaining 2-node input/output degree correlations and obtain a general expression for the expected reciprocity r_{1n} due to

the 1-node correlations alone. In this case, the number of all possible output connections from source nodes with degree k_o is $k_o N(k_o)$ and the number of all possible input connections to target nodes with in-degree q_i is $q_i N(q_i)$. Hence, the probability that one link connects a node with out-degree k_o to a node with in-degree q_i is $\mathcal{P}(k_o \to q_i) = \frac{k_o N(k_o)}{L} \frac{q_i N(q_i)}{L} = \frac{k_o P(k_o) q_i P(q_i)}{\langle k \rangle^2}$. Equation (4) reduces to

$$r_{1n} = \frac{L}{N^2} \frac{\langle k_i k_o \rangle^2}{\langle k \rangle^4}. \quad (6)$$

Here r, a 2-node property, is determined by the single-node characteristics arising from the specific input and output degree sequences. We remind the reader that, in the literature, it is common to randomize networks by methods that conserve the degree sequences in order to obtain expected values accounting for the real degree distribution. Equation (6) is of relevance for significance testing because it is the theoretical estimation of the expected reciprocity in such a typical case.

C. 2-node output/output degree correlations

Following a similar approximation we can calculate the expected reciprocity $r_{2n:o/o}$ due to the 2-node correlations, Fig. 1(b), between the output degree k_o of the source node and the output degree q_o of the target node. However, in this case a few steps need to be carefully considered. We rewrite the link probabilities as

$$\mathcal{P}(\mathbf{k} \to \mathbf{q}) = \mathcal{P}(k_o \to q_o)P(k_i|k_o \to q_o)P(q_i|k_i, k_o \to q_o).$$

The term $P(q_i|k_i, k_o \to q_o)$ introduces dependence of in-degree of the target node q_i on the in-degree of the source node k_i. We are assuming that such correlations are negligible and therefore this term can be rewritten as $P(q_i|k_o \to q_o)$. The second term, $P(k_i|k_o \to q_o)$, can be written as $P(k_i|k_o)$ because the in-degree of the target node q_i and the in-degree of the source node k_i are, again, not correlated. To proceed, it is necessary to use the fact that we are calculating expectations on graphs, and that the final expression has a form which can be calculated using only the assumed knowledge of the network structure, i.e., the frequencies of nodes $P(k) = N(k)/N$ and the frequencies of links $\mathcal{P}(\mathbf{k} \to \mathbf{q}) = L(\mathbf{k} \to \mathbf{q})/L$. Following this line of reasoning the third, already approximated term, should be carefully rewritten. What is the probability that a node will have in-degree q_i given that (i) it has an input link and (ii) it has an out-degree q_o? Without the information (i) this probability is simply $P(q_i|q_o)$, but we have information that such a link does exist. The probability that, following a link between nodes with degrees k_o and q_o, it will run into a node with in-degree q_i is just the number of links $L(\to \mathbf{q})$ which enter all nodes with degrees $\mathbf{q}$ divided by the number of all links $L(\to q_o)$ that enter the nodes with out-degree q_o. These numbers of links can be expressed as $q_i N(q_i, q_o)$ and $\Sigma_{q_i'} q_i' N(q_i', q_o)$, respectively. Thus,

$$P(q_i|k_o \to q_o) = \frac{q_i N(q_i, q_o)}{\displaystyle\sum_{q_i'} q_i' N(q_i', q_o)} = \frac{q_i P(q_i, q_o)}{\displaystyle\sum_{q_i'} q_i' P(q_i', q_o)}.$$

Substituting all these approximations in Eq. (3),

$$r_{2n:o/o} = \frac{L}{N^2} \sum_{\mathbf{k}, \mathbf{q}} \frac{\mathcal{P}(k_o \to q_o)\mathcal{P}(k_o \leftarrow q_o)k_i P(\mathbf{k})q_i P(\mathbf{q})}{\displaystyle\sum_{k_i', q_i'} k_i' P(k_i', k_o)q_i' P(q_i', q_o)P(k_o)P(q_o)}.$$

Finally, the summation terms over in-degrees in the numerator and the denominator cancel out and we obtain a final expression for the expected reciprocity under 2-node output/output degree correlations,

$$r_{2n:o/o} = \frac{L}{N^2} \sum_{k_o, q_o} \frac{\mathcal{P}(k_o \to q_o)\mathcal{P}(k_o \leftarrow q_o)}{P(k_o)P(q_o)}. \quad (7)$$

Interestingly, this expression is independent of the 1-node correlations even if we did not explicitly assume it. The expression for 2-node input/input correlations conserved, Fig. 1(c), is the same only with k_o and q_o replaced by k_i and q_i, see Table I.

D. 2-node input/output degree correlations

We now describe the case in which significant in-out correlations, Fig. 1(d), are present in the network. Such correlations are supposed to influence reciprocity considerably because the probability that a given link of the type $\mathbf{k} \to \mathbf{q}$ has a reciprocal counterpart $\mathbf{k} \leftarrow \mathbf{q}$ is directly proportional to the in-degree k_i of the source node and the out-degree q_o of the target node. Following the previous line of reasoning, the link probability can be expressed in this case as

$$\mathcal{P}(\mathbf{k} \to \mathbf{q}) = \mathcal{P}(k_i \to q_o)P(k_o|k_i \to q_o)P(q_i|k_i \to q_o, k_o),$$

and the conditional terms as

$$P(k_o|k_i \to q_o) = \frac{k_o N(k_o, k_i)}{\displaystyle\sum_{k_o'} k_o' N(k_o', k_i)} = \frac{k_o P(k_o, k_i)}{\displaystyle\sum_{k_o'} k_o' P(k_o', k_i)} = \frac{k_o P(k_o, k_i)}{\bar{k}_{o,k_i}},$$

$$P(q_i|k_i \to q_o) = \frac{q_i N(q_o, q_i)}{\displaystyle\sum_{q_i'} q_i' N(q_o, q_i')} = \frac{q_i P(q_o, q_i)}{\displaystyle\sum_{q_i'} q_i' P(q_o, q_i')}$$

$$= \frac{q_i P(q_o, q_i)}{\bar{q}_{i,q_o}}.$$

Note that the expression $\bar{k}_{o,k_i} = \Sigma_{k'} k_o' P(k_o', k_i)$ is not the average out-degree of nodes with in-degree k_i, because $P(k_o', k_i)$ is the joint probability and not the conditional probability. The average out-degree of nodes with in-degree k_i is $\langle k_o \rangle_{k_i} = \frac{\bar{k}_{o,k_i}}{P(k_i)}$. Using the above relationships we obtain from Eq. (3) an expression for the expected reciprocity $r_{1n2n:i/o} = r_{i/o}$ due to 1-node and 2-node input/output degree correlations,

$$r_{i/o} = \frac{L}{N^2} \sum_{\mathbf{k}, \mathbf{q}} \frac{k_i k_o q_i q_o \mathcal{P}(k_i \to q_o)\mathcal{P}(k_o \leftarrow q_i)P(\mathbf{k})P(\mathbf{q})}{\bar{k}_{i,k_o} \bar{k}_{o,k_i} \bar{q}_{i,q_o} \bar{q}_{o,q_i}}. \quad (8)$$

Additionally, if the 1-node correlations are removed, the expected reciprocity $r_{2n:i/o}$ due to the 2-node input/output degree correlations alone is

$$r_{2n:i/o} = \frac{L}{N^2} \sum_{\mathbf{k}, \mathbf{q}} \frac{k_i k_o q_i q_o \mathcal{P}(k_i \to q_o)\mathcal{P}(k_o \leftarrow q_i)}{\langle k_o \rangle_{k_i} \langle k_i \rangle_{k_o} \langle q_o \rangle_{q_i} \langle q_i \rangle_{q_o}} = \frac{L^2}{N^2} \frac{\langle k_i q_o \rangle^2}{\langle k_i \rangle^4}. \quad (9)$$

Note that without 1-node correlations $\langle k_o \rangle_{k_i} = \Sigma_{k_o'}[k_o' P(k_o', k_i)/P(k_i)] = \langle k_o \rangle$.

IV. APPLICATION TO REAL NETWORKS

The class of degree correlations chosen in this paper is not only very interesting from the theoretical point of view, it is also of practical relevance. In this section, first, our results are applied to several real networks and the impact of degree correlations on reciprocity is discussed. Second, in order to prove the validity of our equations, the theoretical estimations are compared to empirical ensemble averages of proper random networks. Therefore, we present algorithms to gen-

TABLE II. Measured reciprocity r of several real networks, and theoretically expected reciprocities due to different correlation structures. (i) 1-node and all 2-node correlations r_{1n2n}. (ii) 2-node out-out $r_{2n:o/o}$ (no 1-node). (iii) 1-node correlations r_{1n} (no 2-node). (iv) The density of connections $\bar{a}$.

Network	r	r_{1n2n}	$r_{2n:o/o}$	r_{1n}	$\bar{a}$
World trade webs [15]					
Year 1948	0.823	0.812	0.768	0.707	0.382
Year 2000	0.980	0.958	0.883	0.813	0.560
Neural networks					
C. elegans [16]	0.433	0.329	0.071	0.060	0.033
Cortical networks					
Cat [17]	0.734	0.659	0.473	0.390	0.300
Macaque [18]	0.750	0.645	0.287	0.230	0.155
Food webs [19]					
Little Rock Lake	0.0339	0.0323	0.0365	0.0501	0.0743
Grassland	0.0	0.0	0.0077	0.0079	0.0179
St. Marks Sea	0.0	0.0075	0.0500	0.0703	0.0948
St. Martin Island	0.0	0.0016	0.0419	0.06765	0.1131
Silwood Park	0.0	0.0	0.0002	0.0160	0.0155
Ythan Estuary	0.0034	0.0050	0.0335	0.0531	0.0330
Wikipedia [7]					
Spanish	0.3517	0.1466	0.0322	0.0056	0.0004
Portuguese	0.3563	0.1207	0.0168	0.0084	0.0004
Chinese	0.3668	0.1556	0.0256	0.0096	0.0010

erate random networks conditional on different correlation classes.

A. Reciprocity of real networks

All the analytical expressions summarized in Table I can be directly estimated by measuring the necessary statistics out of a real network. For the distinct correlation classes, different quantities are counted: $N(k_i)$, $N(k_o)$, $N(\boldsymbol{k})$ or $L(k_o \to q_i)$, $L(k_i \to q_o)$, $L(\boldsymbol{k} \to \boldsymbol{q})$, etc. Then, the frequencies $\mathcal{P}(\boldsymbol{k} \to \boldsymbol{q}) = L(\boldsymbol{k} \to \boldsymbol{q})/L$, $P(k_i) = N(k_i)/N$, etc., are introduced in the formulas to calculate the expected reciprocity.

For all the real networks in Table II we have calculated the expected reciprocities after Eqs. (3) and (5)–(7). For comparison, we also show the density of connections $\bar{a}$, i.e., the reciprocity of uncorrelated random networks of the same size N and the same number of links L. As observed in Table II the general estimation r_{1n2n} alone can almost completely explain the real r of many networks, e.g., world trade webs and most of the food webs. It also makes a very good approximation for cortical and neural networks. The large discrepancy in the case of Wikipedia websites suggests the presence of additional internal structure apart from the 1-node and 2-node degree correlations. For example, the *C. elegans* neural network, the cortical networks, and the Wikipedias are known to have modular and hierarchical structure. In all of these cases the real r is larger than the expected r_{1n2n}. Nevertheless, our results demonstrate that degree correlations are crucial to understand the reciprocity of real networks. The expected r_{1n2n} of the world trade webs is roughly 2 times as large as $\bar{a}$, the expected reciprocity of an equivalent uncorrelated random network. In other cases r_{1n2n} is up to 3 orders of magnitude larger than $\bar{a}$.

The values of r_{1n2n} are always followed by the contribution of 2-node output/output correlations $r_{2n:o/o}$, Eq. (7). With the exception of Silwood Park and Ythan Estuary food webs, values of r_{1n}, Eq. (6), are all closer to the real r than the density of links $\bar{a}$. These results reveal a hierarchy of degree correlation classes according to their contribution to reciprocity.

B. Numerical corroboration

Finally, we prove in this section that our analytical expressions are valid expectation values of the network reciprocity. Using the correlation structure of the real networks in Table II, we have generated ensembles of maximally random networks under different conditions of 1-node and 2-node degree correlations. The ensemble average reciprocities have been compared to the analytical results showing excellent agreement between experimental and theoretical expected values, Fig. 2. The generation of random networks with desired degrees and correlations is not trivial, but we have developed three algorithms for that purpose. In the absence of analytical results for other graph measures, e.g., clustering coefficient, average path length, etc., the following algorithms are useful tools to empirically calculate their expected values under desired conditions of degree correlations.
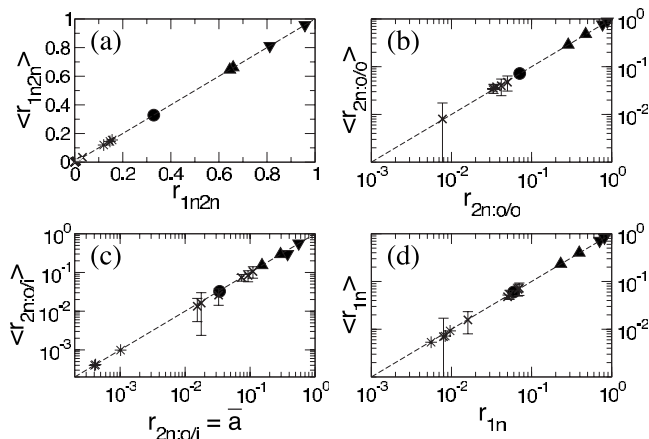
FIG. 2. Numerical corroboration of theoretical estimations. Experimentally measured reciprocities (vertical axes) and our theoretical estimations (horizontal axes). (a) All 1-node and 2-node correlations conserved. (b) Only 2-node out-out correlations conserved. (c) Only 2-node out-in correlations. (d) 1-node correlations. All data points are averages of 100 realizations. Food webs ($\times$), *C. elegans* ($\bullet$), world trade webs ($\blacktriangledown$), cortical networks ($\blacktriangle$), and Wikipedias ($\ast$). Dashed lines are references of perfect coincidence.

1. Random networks with desired degree sequences and 2-node degree correlations, r_{1n2n}

Given a real network of size N and L links, we can measure its degree distribution $N(\boldsymbol{k})$ and the 2-node correlation structure $L(\boldsymbol{k} \rightarrow \boldsymbol{q})$. With this information in hand, it is possible to generate maximally random networks that possess such properties. To an initially empty network of size N, its nodes are randomly assigned their final degrees $\boldsymbol{k}$ following the distribution $N(\boldsymbol{k})$. Then, links are introduced at random but following carefully considered steps. First, one source node s is chosen at random. We know that s has been assigned to have final degrees $\boldsymbol{k}'$. Because of the 2-node correlations, s can only connect to nodes with particular degrees $\boldsymbol{q}'$ such that $L(\boldsymbol{k}' \rightarrow \boldsymbol{q}') > 0$. A list of possible target nodes is constructed by taking only those nodes assigned to have final degrees $\boldsymbol{q}'$ where $L(\boldsymbol{k}' \rightarrow \boldsymbol{q}') > 0$. From this list one target node t is chosen at random and the link $s \rightarrow t$ is introduced into the initially empty network. Note that a node with degrees $\boldsymbol{k} = (k_i, k_o)$ can only be chosen k_o times as source and k_i time as target, otherwise the distribution $N(\boldsymbol{k})$ will not be conserved. Thus, if the quantities $L(\boldsymbol{k} \rightarrow \boldsymbol{q})$, and k_i and k_o of each node are adequately updated during the process, only L iterations are required to construct the maximally random network. Described as it is, this method allows for the introduction of self-loops and multiple links. Avoiding them is far from trivial.

To prove the validity of our general theoretical result r_{1n2n}, Eq. (3), ensembles of 100 random networks have been generated. The results in Fig. 2(a) show an excellent agreement between r_{1n2n} (horizontal axis) and the empirical ensemble average reciprocities $\langle r_{1n2n} \rangle$ (vertical axis). All of the generated random realizations have been positively tested to have the same $N(\boldsymbol{k})$ and $L(\boldsymbol{k} \rightarrow \boldsymbol{q})$ as the original real networks. Apart from the *C. elegans*, none of the real networks
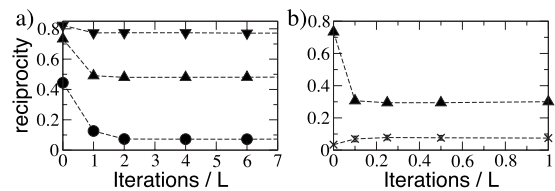

FIG. 3. Stability of rewiring algorithms. With increasing number of randomizing iterations structural properties reach stable values, here r is shown. (a) Algorithm that conserves the 2-node output/output degree correlations. (b) Algorithm that conserves the 2-node output/input correlations. All data points are averages of 50 realizations. Error bars are very small. Food webs ($\times$), *C. elegans* ($\bullet$), world trade webs ($\blacktriangledown$), cortical networks ($\blacktriangle$).

here studied contain self-loops or multiple links. In the random realizations out of the Silwood Park food web, on average, only 4.5% of the L links formed self-loops or multiple links. The realizations out of the Wikipedias contain less than 2% of such links, and the realizations out of the cortical networks and the world trade webs contain less than 1%. In all cases the reciprocity was measured counting the multiple links as a unique link and ignoring the self-loops. The small fraction of nondesired links have very little impact as observed in the excellent agreement between the theoretical expectation r_{1n2n} and the experimental ensemble averages $\langle r_{1n2n} \rangle$, Fig. 2(a).

2. Rewiring method that conserves 2-node output/output degree correlations, $r_{2n:o/o}$

In this case, we opted for a rewiring algorithm. Such methods start from a given real network and stepwise randomize its connections by specific rules that conserve the desired properties. In order to obtain maximally random networks that conserve uniquely the original 2-node output/output degree correlations, Fig. 1(b), we proceed in the following manner. One of the links in the network is chosen at random $s \rightarrow t_1$. Nodes s and t_1 have out-degrees k_o and q_o, respectively. From all the nodes in the network with out-degree q_o a new target node t_2 is randomly chosen. If all conditions to avoid the introduction of self-loops and multiple links are satisfied, the old link is destroyed $s \nrightarrow t_1$ and the new link $s \rightarrow t_2$ is created.

After several iterations, all 2-node degree correlations are randomized except for the output/output degree. A relevant question arises when applying rewiring algorithms: how long should the process run so that resulting networks are maximally random? After some finite number of iterations the network reaches a state where any successive rewiring leads to a statistically equivalent random network. Once this state is reached, the network measures converge to their expected value [20]. We have applied several levels of rewiring and the reciprocity has been measured at each level. As observed in Fig. 3(a) the ensemble average of r reaches a stable point. Any other network measure follows the same behavior. Obviously, the number of necessary iterations is proportional to the number of links L and, for this particular method, all networks reach a maximally random state after $2L$ iterations. For security, $4L$ iterations are recommended.

To prove the validity of our theoretical expression $r_{2n:o/o}$ in Eq. (7), e.g., the expected reciprocity of networks with prescribed 2-node output/output degree correlations, we have generated ensembles of 100 rewired networks out of the real networks in Table II. The comparison between $r_{2n:o/o}$ and the empirical ensemble averages $\langle r_{2n:o/o} \rangle$ shows again excellent agreement, Fig. 2(b).

3. Rewiring method that conserves 2-node output/input degree correlations, $r_{2n:o/i}$

We designed yet another rewiring algorithm that conserves uniquely the original 2-node output/input correlations, Fig. 1(a), of a real network. First, two nodes are selected at random, s_1 and s_2. These nodes are divided in two halves, one containing all in links and the other containing all the out links, i.e., $s_1 = s_1^{in} \cup s_1^{out}$ and $s_2 = s_2^{in} \cup s_2^{out}$. Then, the in-halves and the out-halves are switched to form two new nodes, $s_3 = s_1^{in} \cup s_2^{out}$ and $s_4 = s_2^{in} \cup s_1^{out}$. Cases that would introduce self-loops are carefully discarded. The resulting randomized networks conserve the in-degree $N(k_i)$ and the out-degree $N(k_o)$ distributions, and the number of links $L(k_o \rightarrow q_i)$ from the original network while the rest of correlations are randomized.

The stability of the rewiring process has been tested and the results are shown in Fig. 3(b). In this case, each iteration step rewires several links so that the algorithm is much faster. After only $0.3L$ iterations all networks reach a maximally random state. The theoretical expected reciprocity $r_{2n:o/i}$, Eq. (5), is compared to the empirical ensemble averages $\langle r_{2n:o/i} \rangle$ of 100 rewired networks and the results shown in Fig. 2(c). All networks were rewired $0.4L$ times.

4. Conserving degree sequences, r_{1n}

Finally, a well-known rewiring method was used to randomize all 2-node correlations while degree sequences are conserved [21]. The method arises from earlier approaches [22–24] and consists in randomly choosing two links, $s_1 \rightarrow t_1$ and $s_2 \rightarrow t_2$, and exchanging them, $s_1 \rightarrow t_2$ and $s_2 \rightarrow t_1$. In Fig. 2(d) the expected reciprocities $\langle r_{1n} \rangle$ from the ensemble averages are compared to our analytical estimations r_{1n}, Eq. (6).

All of the results in this section prove that our analytical expressions are valid expected reciprocities under the correlation conditions considered in this paper. The equations are valid even for small networks such as the cat cortical network ($N = 53$, density of links $\bar{a} \approx 0.3$) despite the fact that formulas are derived in the thermodynamical limit. Error bars in most cases are very small and only food webs exhibit some larger fluctuations.

V. SUMMARY AND DISCUSSION

In summary, we have studied the influence of 1-node and 2-node degree correlations on the reciprocity of networks with arbitrary degree sequence. We find that degree correlations account for a very large part of the observed reciprocity, explaining it almost completely in typical cases. The level of contribution to r is nontrivial and largely depends on the type of correlations involved, revealing a hierarchy of correlation classes. As observed, these contributions can span over orders of magnitude. Our analytical estimations are proved as valid expectation values of reciprocity by comparison to ensemble averages of random networks which preserve desired correlation structure. Both from a theoretical and a practical point of view, it would be highly desirable to extend the current work and obtain expected values of other network measures, e.g., clustering coefficient, average path length, motif profiles, etc., following a similar philosophy: the analytical expressions should be computable using only information directly measured from the real network under study. In the absence of such results the numerical methods introduced in this paper are useful tools for obtaining expectation values under conditions of prescribed degree sequences and degree correlations.

The dynamical influence of degree correlations has been studied for epidemic spreading [25] and for synchronization in undirected networks [26]. In the light of our results, which clearly relate reciprocity and degree correlations, it is important to also investigate the influence of reciprocity on those phenomena for the broader class of directed networks.

In order to perform satisfactory modeling, the key parameters governing network growth and evolution need to be identified. Therefore, theoretical understanding of the interplay between different topological properties is necessary to distinguish between the significant measured values and those expected as by-products of other properties. In this paper, we have explored the interplay between degree correlations and network reciprocity. Previous efforts in this direction include relations between degree correlations and clustering coefficient in undirected graphs [27,28]. Degree correlations are also expected to be relevant for other network measures, e.g., path length, network motifs, modularity, etc. Therefore, further work is desirable to extend the analytical and experimental approaches presented here. No doubt, quantification of similar structural interrelations will significantly elucidate the essence of the structure-function-evolution interplay in complex networks.

[1] E. Fischer and U. Sauer, Nat. Genet. **37**, 636 (2005).

[2] C. Zhou, L. Zemanova, G. Zamora, C. C. Hilgetag, and J. Kurths, Phys. Rev. Lett. **97**, 238103 (2006).

[3] M. R. Gardner and W. R. Ashby, Nature (London) **228**, 784 (1970).

[4] R. M. May, Nature (London) **238**, 413 (1972).

[5] L. M. Pecora and T. L. Carroll, Phys. Rev. Lett. **80**, 2109 (1998).

[6] S. Wasserman and K. Faust, *Social Network Analysis* (Cambridge University Press, Cambridge, 1994).

[7] V. Zlatić, M. Bozicevic, H. Stefancic, and M. Domazet, Phys. Rev. E **74**, 016115 (2006).

[8] M. A. Serrano *et al.*, ACM Trans. Web. **1**(2), 10 (2007).

[9] M. Boguñá and M. A. Serrano, Phys. Rev. E **72**, 016106 (2005).

[10] D. Garlaschelli and M. I. Loffredo, Phys. Rev. E **73**, 015101(R) (2006).

[11] N. Schwartz, R. Cohen, D. ben-Avraham, A. L. Barabasi, and S. Havlin, Phys. Rev. E **66**, 015104(R) (2002).

[12] T. A. B. Snijders, P. E. Pattison, and M. S. Handcock, Sociol. Methodol. **36**, 99 (2006).

[13] M. Mandel, Soc. Networks **22**, 285 (2000).

[14] D. Garlaschelli and M. I. Loffredo, Phys. Rev. Lett. **93**, 268701 (2004).

[15] K. S. Gleditsch, J. Conflict Resolut. **46**, 712 (2002).

[16] J. G. White *et al.*, Philos. Trans. R. Soc. London, Ser. B **314**, 1 (1986).

[17] J. W. Scannell *et al.*, Cereb. Cortex **9**, 277 (1999).

[18] O. Sporns and J. D. Zwi, Neuroinformatics **2**, 145 (2004).

[19] http://www.cosin.org/extra/data/foodwebs/

[20] G. Zamora-López *et al.*, in *Lectures in Supercomputational Neuroscience*, edited by P. Beim Graben, C. S. Zhou, M. Thiel, and J. Kurths (Springer, New York, 2008), Chap. 3.

[21] S. S. Shen-Orr *et al.*, Nat. Genet. **31**, 64 (2002).

[22] A. R. Rao and S. Bandyopadhyay, Sankhya, Ser. A **58**, 225 (1996).

[23] J. M. Roberts, Soc. Networks **22**, 273 (2000).

[24] T. A. B. Snijders, Psychometrika **56**, 397 (1991).

[25] Y. Moreno, J. B. Gómez, and A. F. Pacheco, Phys. Rev. E **68**, 035103(R) (2003).

[26] F. Sorrentino *et al.*, Physica D **224**, 123 (2006).

[27] S. N. Dorogovtsev, Phys. Rev. E **69**, 027104 (2004).

[28] P. Holme and J. Zhao, Phys. Rev. E **75**, 046111 (2007).

[29] Wikipedia networks are constructed out of the presented documents under the http://wikipedia.org website. Only internal links within the Wikipedia articles are considered.

[30] For the thermodynamical limit it is understood that we have the limit of large networks ($N \to \infty$) with sparse connectivity (density $\bar{a} \to 0$). In such a limit the probability of random introduction of multiple links can be discarded.