Significance of Graph Measures and Determinism in Random Networks

G. Zamora-López[1]*, V. Zlatic[2], C.S. Zhou[3] and J. Kurths[4]

[1]Center for Dynamics of Complex Systems, University of Potsdam

[2]Rudjer Boskovic Institute, Zagreb

[3]Department of Physics, Hong Kong Baptist University

[4]Potsdam Institute for Climate Impact Research

*gorka@agnld.uni-potsdam.de

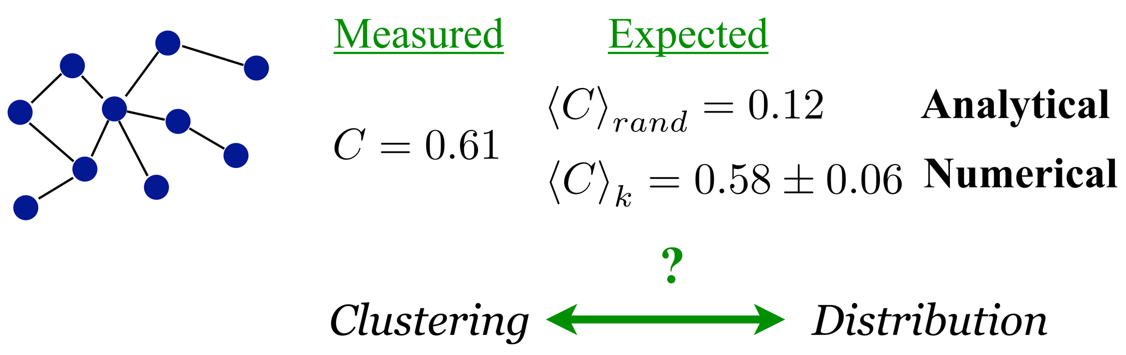
ABSTRACT

Beyond describing the structure of networks, we also want to uncover their underlying principles of organisation. To capture the true driving forces in the organisation and development, we still need to understand the statistical interdependencies between the graph measures. In our work, we have derived analytical formulae to calculate the expected reciprocity in directed networks, conditional on the degree distribution and/or the degree-degree correlations [1]. We find that, in many empirical networks, the observed and the expected reciprocities are equal, opening the question whether the observed reciprocity arises as a by-product of the degree correlations, or on the contrary, the observed degree correlations arise as a functional necessity for reciprocal links. The application of constrains (e.g. degree-correlations) to random networks can drastically reduce their ``ammount of randomness'' [2]. Within the empirical networks we analysed, the degree correlations alone determine between 30% and 95% of the connectivity.

MOTIVATION

The graph measures are statistical measures applied to the same mathematical object (the network) hence, their outcome is not independent of each other. Understanding the statistical dependence between the graph measures is still a relevant open question, only understanding of this interdependencies will we be able to infer with certainty the causal relations governing the interplay between structure, dynamics and function.

What is significant?



Figure–1: (IN-)DEPENDENT MEASURES: Under the current theoretical understanding, it is very difficult to identify which are the "true driving-forces" in the evolution of a real network, i.e. which network characteristics are optimised by function and which not.

LOCAL GRAPH MEASURES

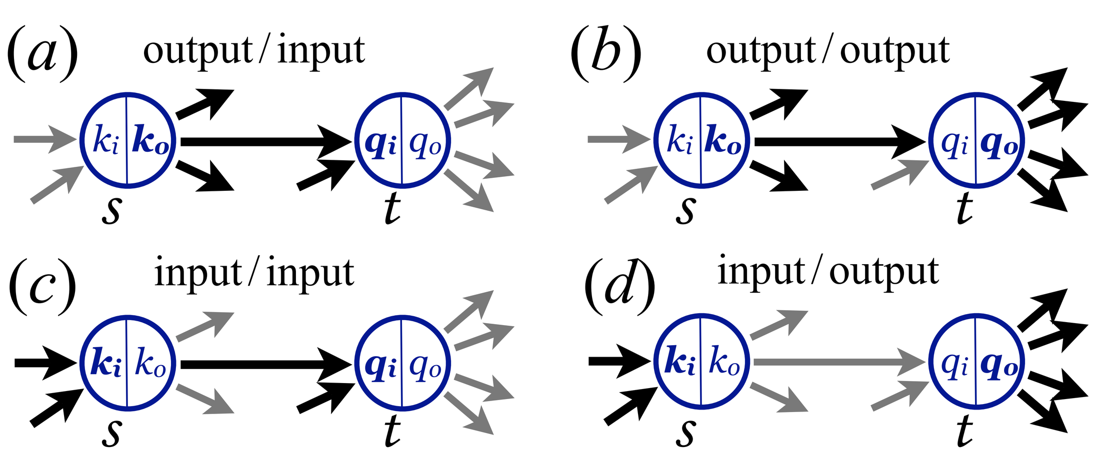
In *directed networks*, the degree of nodes is splitted into the input k_i and the output k_o degrees, hence, the **degree distribution** becomes a 2-dimensional statistic:

- $N(\boldsymbol{k})$ = number of nodes with degree $\boldsymbol{k} = (k_i, k_o)$.

The **1-node degree correlation (1n)** is the correlation between the input and the output degrees of the same node.

For a link $s \to t$, connecting two nodes with degrees $\boldsymbol{k} = (k_i, k_o)$ and $\boldsymbol{q} = (q_i, q_o)$, the **2-node degree correlation (2n)** is a 4-dimensional statistic. We charaterise it as:

- $L(\boldsymbol{k}{\to}\boldsymbol{q}) = L(k_i, k_o, q_i, q_o)$ = number of links projeting from nodes with degrees $\boldsymbol{k}$ to nodes with degree $\boldsymbol{q}$.



Figure–2: 2n2d DEGREE CORRELATIONS: In directed networks, there are 4 classes of 2-node, 2-degree (2n2d) correlations. Altogether, there are up to 10 combinations of 1-node and 2-node degree correlations.

The **reciprocity** of a directed network is defined as the probability that for a randomly chosen link s → t, the opposite s ← t link also happens.

- $r = L^{\leftrightarrow}/L$, where $L^{\leftrightarrow}$ is the number of reciprocal links and L is the number of links in the network.

EXPECTED RECIPROCITY

Under the class of degree correlations here assumed, a network is considered to be maximally random when any of the nodes with degree $\boldsymbol{k}$ are equally connected to any of the nodes with degree $\boldsymbol{q}$. If a network contains $L(\boldsymbol{k}{\to}\boldsymbol{q})$ links of the type $\boldsymbol{k}{\to}\boldsymbol{q}$, then the probability is, in the thermodynamical limit:

$$p(\boldsymbol{k} \to \boldsymbol{q}) = \frac{L(\boldsymbol{k} \to \boldsymbol{q})}{N(\boldsymbol{k})N(\boldsymbol{q})}$$

The expected number of reciprocal links of the type $\boldsymbol{k}{\leftrightarrow}\boldsymbol{q}$ is $\langle L(\boldsymbol{k}{\leftrightarrow}\boldsymbol{q})\rangle = L(\boldsymbol{k}{\to}\boldsymbol{q})\, p(\boldsymbol{k}{\leftarrow}\boldsymbol{q})$. Thus we compute the **expected reciprocity under all the 1-node and 2-node degree correlations**:

$$r_{1n2n} = \frac{1}{L}\sum_{\boldsymbol{k},\boldsymbol{q}} \frac{L(\boldsymbol{k} \to \boldsymbol{q})L(\boldsymbol{k} \leftarrow \boldsymbol{q})}{N(\boldsymbol{k})N(\boldsymbol{q})} = \frac{L}{N^2}\sum_{\boldsymbol{k},\boldsymbol{q}} \frac{\mathcal{P}(\boldsymbol{k} \to \boldsymbol{q})\mathcal{P}(\boldsymbol{k} \leftarrow \boldsymbol{q})}{P(\boldsymbol{k})P(\boldsymbol{q})}$$

Considering only random networks with prescribed degree distribution **(1-node correlations)**:

$$r_{1n} = \frac{L}{N^2}\frac{\langle k_i k_o \rangle^2}{\langle k \rangle^4}$$

And in **random networks**:

$$r_{rand} = \frac{L}{N^2} = \text{density}$$

EMPIRICAL NETWORKS

We find that in many networks, **the 1-node and the 2-node degree correlations "explain" the observed reciprocity**.

Network	r_{real}	r_{1n2n}	r_{1n}	r_{rand}
World Trade Webs				
Year 1948	0.823	0.812	0.707	0.382
Year 2000	0.980	0.958	0.813	0.560
Neural Networks				
C. Elegans	0.433	0.329	0.060	0.033
Cortical Networks				
Cat	0.734	0.659	0.390	0.300
Macaque	0.750	0.645	0.230	0.155
Food Webs				
Little Rock lake	0.0339	0.0323	0.0501	0.0743
Grassland	0.0	0.0	0.0079	0.0179
St. Marks sea.	0.0	0.0075	0.0703	0.0948
St. Martin Isl.	0.0	0.0016	0.06765	0.1131
Silwood Park	0.0	0.0	0.0160	0.0155
Ythan estuary	0.0034	0.0050	0.0531	0.0330
Wikipedia Website				
Spanish	0.3517	0.1466	0.0056	0.0004
Portuguese	0.3563	0.1207	0.0084	0.0004
Chinese	0.3668	0.1556	0.0096	0.0010

Table–I: RECIPROCITY OF REAL NETS: After measuring the reciprocity for several empirical networks of different characteristics, the reciprocity has been compared to the expected reciprocity under different constraints.

NULL MODELS

The study of significance of graph measures and their statistical interdependence lies in the formulation of proper null-hypothesis. We are interested in uncovering what are the expected values that graph measures obtain in maximally random graphs which conserve desired statistics.

The random graph (Erdös-Rényi) is the most basic null-model, it consists of the set of maximally random graphs of size N and number of links L. Further popular constraints are to conserve also the degree sequence $N(\boldsymbol{k})$ (degree distribution) or, as in our work, the degree-degree correlations $L(\boldsymbol{k}{\to}\boldsymbol{q})$.

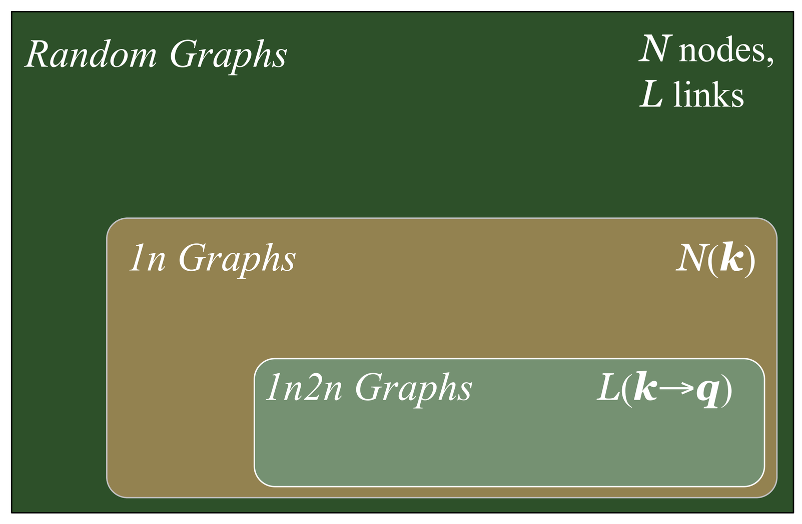


Figure-4: SPACE OF RANDOM GRAPHS: By imposing statistical constraints to the generation of random networks drastically reduces the degrees of freedom of a null-model and hence, the space of accessible random graphs.

In the case of random graphs with desired degree correlations (1n2n random digraphs) we find that when the condition

$$L(\boldsymbol{k}{\to}\boldsymbol{q}) = N(\boldsymbol{k})\, N(\boldsymbol{q}) \qquad (1)$$

holds, those links become deterministic, i.e. they cannot not be randomised.

Network	N	L	L_{det}	L_{det}/L
Cortical Networks				
Cat	53	826	654	0.792
Macaque	70	747	569	0.762
Food Webs				
St. Martin Isl.	45	224	139	0.621
St. Marks sea.	49	223	146	0.655
Grassland	88	137	9	0.0657
Ythan estuary	135	597	267	0.447
Silwood Park	154	365	33	0.315
Little Rock lake	183	2476	2149	0.868
World Trade Webs				
Year 1948	82	2539	2433	0.958
Year 2000	190	20105	19138	0.952
Wikipedia Website				
Chinese	18089	332434	96611	0.291
Portuguese	30374	373215	78152	0.209
Spanish	39562	655615	166073	0.253

Table–II: DETERMINISM IN RANDOM NETS: Number of deterministic links (L_{det}) found in real networks after condition (1) is satisfied.

We find that ***in many real networks the 1-node and the 2-node correlations deternine the structure of the network almost completely***, while in other cases (e.g. the Wikipedias) higher order structures must be present.

ORIGINAL PUBLICATIONS

[1] Zamora-López, V. Zlatic et al. *Phys. Rev. E* **77**, 016106 (2008).

[2] Zamora-López, V. Zlatic et al. *J. Phys. A* **41**, 224006 (2008).