Constructing and sampling graphs with specified joint-degree matrix

Charo I. del Genio

Warwick Mathematics Institute
Centre for Complexity Science
Warwick Infectious Disease Epidemiology Research (WIDER) Centre

University of Warwick

Complex networks

Roads, power-grids, gene regulation, databases, airline connections, the Internet, epidemics, metabolism: all can be modelled as networks.

The structure of a network is what is known in mathematics as a graph: as a set of connections, called edges or links, between pairs from a discrete set, called vertices or nodes.

As its applications are, so network science is a naturally multidisciplinary field, encompassing diverse subjects such as mathematics, physics, computer science, and sociology.

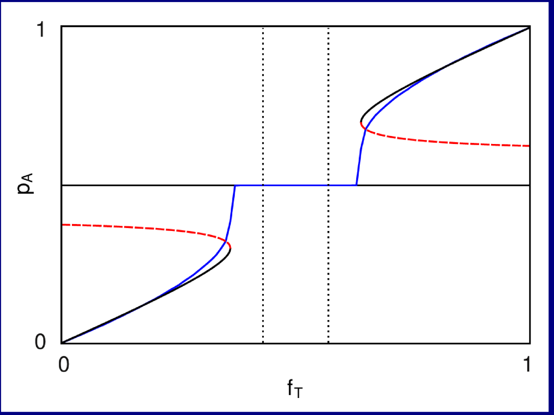
THE UNIVERSITY OF WARWICK

centre for complexity science

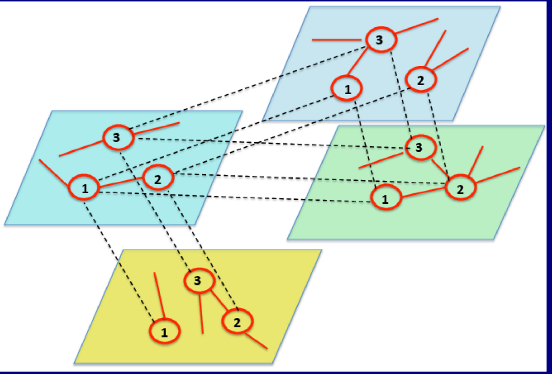
The network model

In the last 15 years, the network model has proved a successful tool in gaining deeper insight into complex systems.

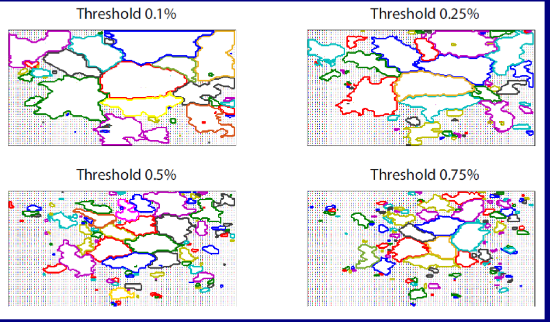
Network representation has allowed researchers to gain insight into the behaviour of systems of diverse nature, including the Internet, power grids, transport networks, and epidemic spreading.



C. I. del Genio and T. Gross, New J. Phys. **13**, 103038



S. Boccaletti, C. I. del Genio *et al.*, Phys. Rep. **544**, 1



F. Botta and C. I. del Genio, *in preparation*

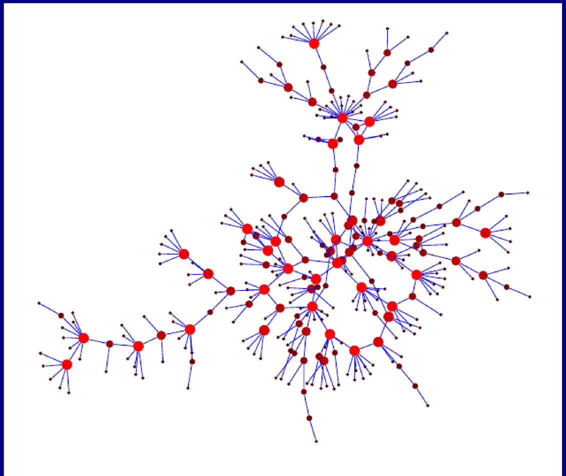
THE UNIVERSITY OF WARWICK

centre for complexity science

Constrained graph sampling

In network modeling, one often generates ensembles of graphs obeying a given constraint. The node degrees, that is the number of edges each node forms, are a typical constraint.

Creating and sampling graphs with a given degree sequence is a challenging problem that has attracted considerable interest amongst researchers.



J. Potterat *et al.*, Sex. Trans. Inf. **78**, i159

The two main classes of graph construction methods are the “edge-swap” and the “stub-matching”.

Shortcomings of existing methods

Unfortunately, the mixing times of edge-swap schemes are unknown in the general case, and they quickly increase with the number of nodes in the few known cases.

Also, random placement of edges as often performed in stub-matching methods easily generates multiple edges and self-connections, forcing a rejection of the sample to avoid biases.

This effectively eliminates the main advantage of fully randomized configuration-model-like algorithms, as the networks for which they are exponentially more likely to incur in these problems are those of highest scientific interest.

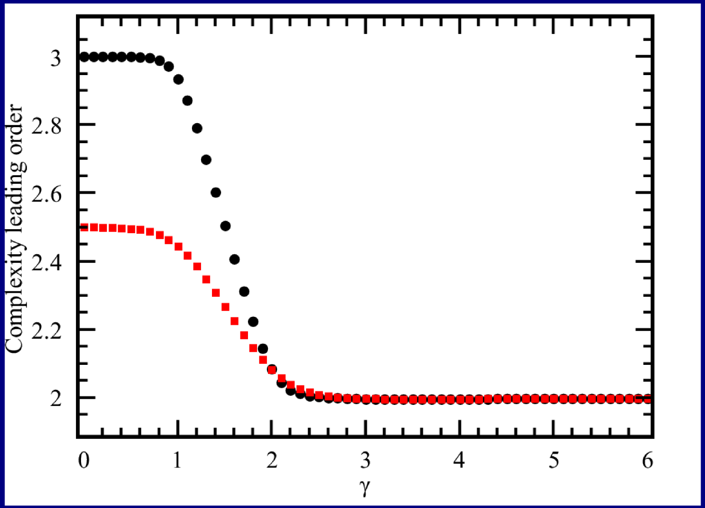
THE UNIVERSITY OF WARWICK

centre for complexity science

Degree-based graph construction

Recently, I introduced a direct graph construction algorithm that efficiently produces rejection-free, unbiased, statistically independent samples of a given sequence.

The method systematically creates links between nodes, verifying that the placement of a link will not lead to multiple edges or loops.



C. I. del Genio *et al.*, PLoS One **5** (4), e10012
F1000 top 2% in biology and medicine

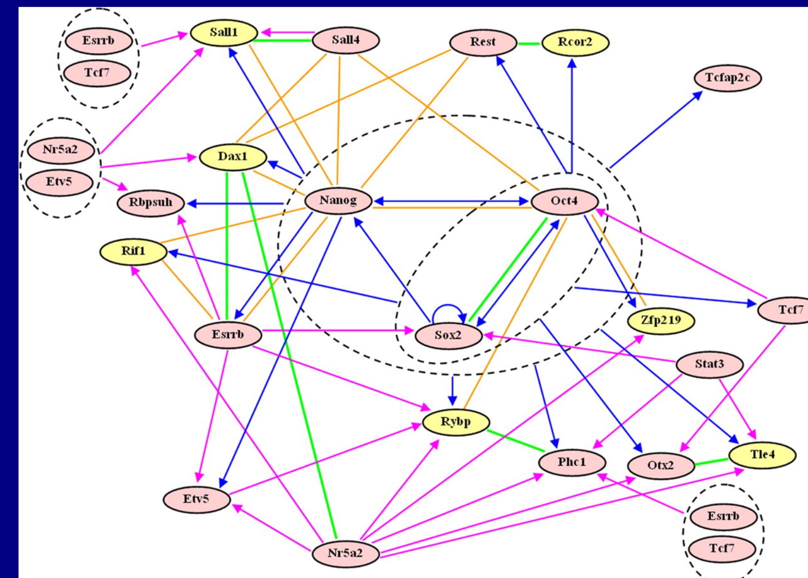
All the evaluations are efficiently implemented, guaranteeing the construction of a sample in cubic time in the worst-case.

Directed graphs

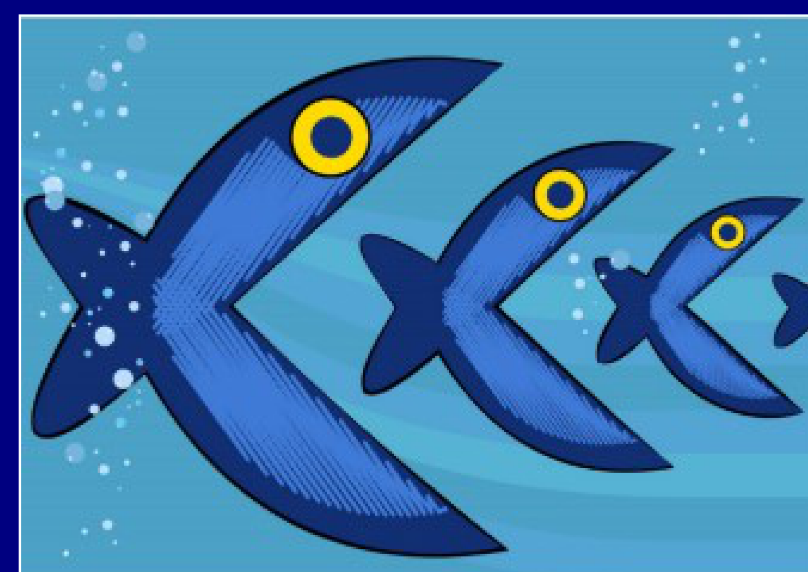
Nevertheless, in many systems the interaction between two entities has a direction from one to the other. Gene regulatory networks and food webs provide an example.

Such systems cannot be described by undirected graphs, although undirected graphs can be thought of in terms of directed ones.

Therefore, a method for generating and sampling *directed* graphs with given degrees is also needed.



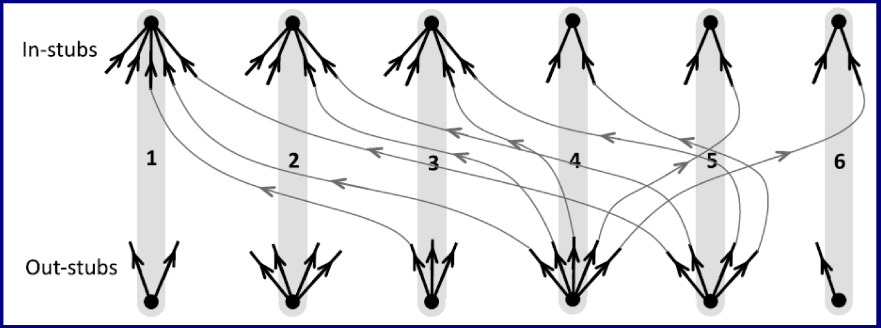
Q. Zhou *et al.*, PNAS, **104** (42), 16438 (2007)



THE UNIVERSITY OF WARWICK

centre for complexity science

Directed graphs



H. Kim, C. I. del Genio *et al.*, New J. Phys. **14**, 023012

Fortunately, proving new theorems and introducing a new formalism, an efficient and exact solution could be obtained also in the case of directed networks.

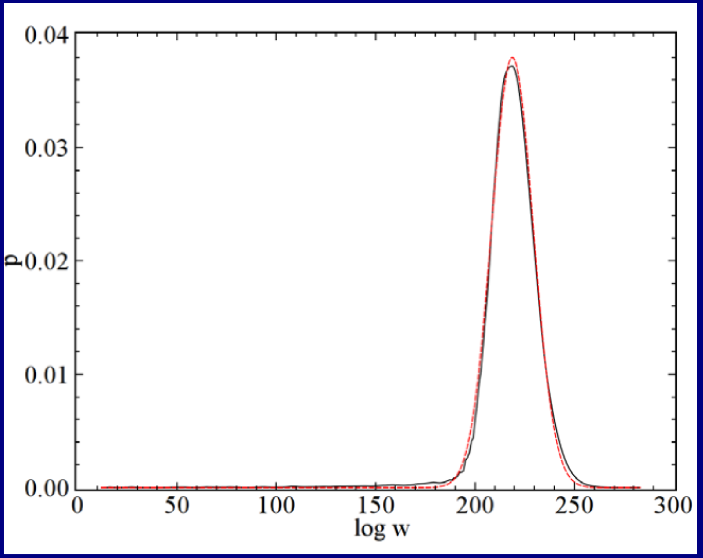
Notably, this approach also solves the problem for bipartite graphs, since they can always be represented as directed networks in which some vertices have vanishing out-degree, and the remaining have vanishing in-degree.

THE UNIVERSITY OF WARWICK

centre for complexity science

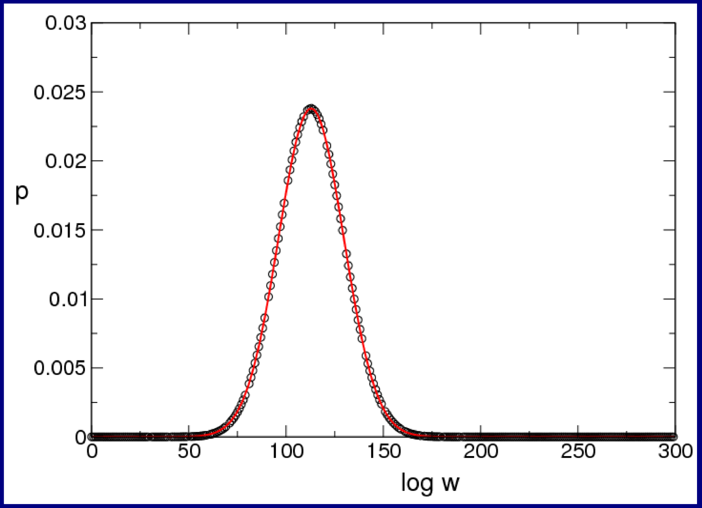
Biased sampling

Both algorithms generate an independent sample (di)graph every time they are run. However, the samples are not realized with a uniform probability.



C. I. del Genio *et al.*, PLoS One **5** (4), e10012

Nevertheless, the methods provide a weight for each sample, allowing the calculation of observables as if they were obtained uniformly, or from any other chosen distribution.



H. Kim, C. I. del Genio *et al.*, New J. Phys. **14**, 023012

Biased sampling

All that is needed for this purpose is to apply the customary importance sampling equation, for any quantity of interest Q, using the weights w provided by the method.

$$\langle Q \rangle = \frac{\sum_{i=1}^{M} w(s_i) Q(s_i)}{\sum_{i=1}^{M} w(s_i)}$$

$$w_u(s) = \prod_{i=1}^{m} \frac{1}{\bar{d}_i!} \prod_{j=1}^{\bar{d}_i!} |A_i(j)| \qquad w_d(s) = \prod_{i} \prod_{j=1}^{d_i^+} |A_i(j)|$$

THE UNIVERSITY OF WARWICK

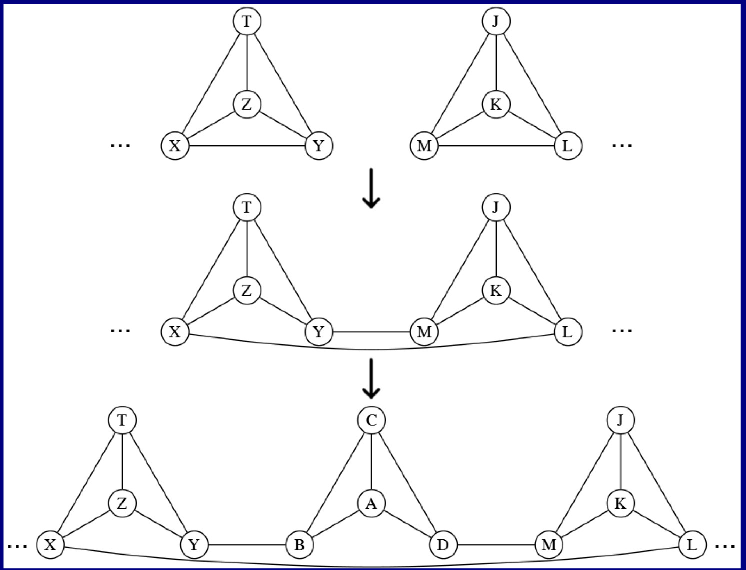
centre for complexity science

Correlations

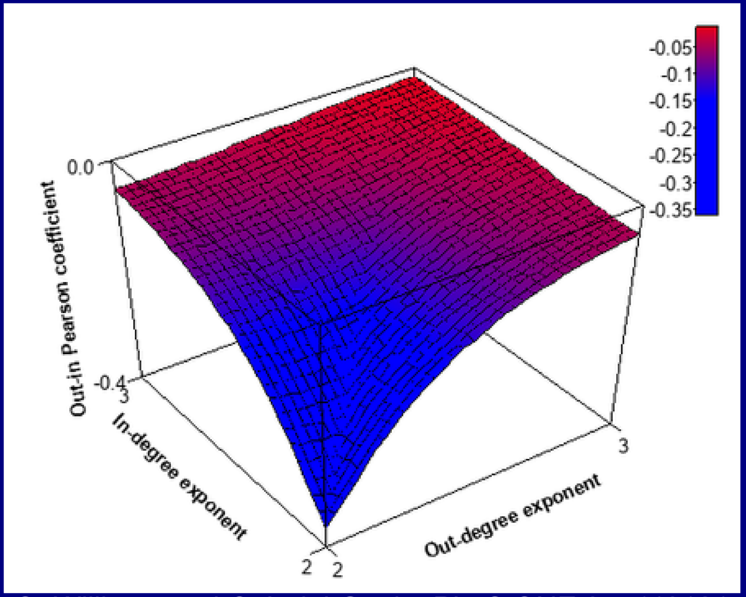
Graph sampling is even more difficult with structural constraints of higher order, such as correlations amongst the node degrees.

These are probability distributions, defined for each degree *d* in an ensemble of graphs, so that $P(d'|d)$ is the probability that a neighbour of a node of degree *d* has degree *d'*.

Degree correlations are known to affect structural and dynamical properties of networks and the processes they support.



C. I. del Genio and T. House, Phys. Rev. E **88**, 040801(R)



O. Williams and C. I. del Genio, PLoS ONE **9**, e110121

THE UNIVERSITY OF WARWICK

centre for complexity science

Joint-degree matrices

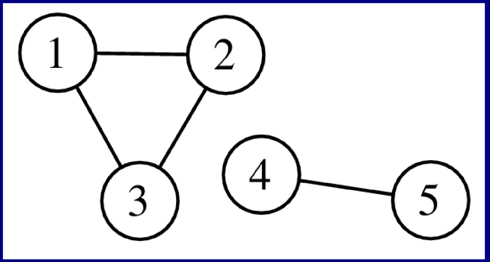
In a single graph, the correlation information can be represented as a symmetric matrix, called a joint-degree matrix (JDM), whose (α,β) element is the number of edges between nodes of degree α and nodes of degree β.

Then, degree correlations can be thought of the average of the individual JDMs of an ensemble of graphs, each normalized by its number of edges.

Our aim is to describe an exact and efficient algorithm to construct and sample ensembles of graphs with a specified joint-degree matrix.

Graphicality

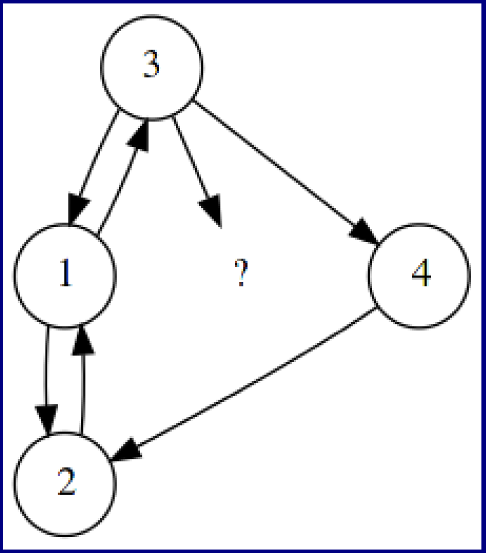
Not all sequences of integers are the degree sequence of some graph. Those that are, are said to be *graphical*. The same is true for JDMs.



{2, 2, 2, 1, 1}

Before discussing the graphicality of JDMs, note that a JDM J of largest degree Δ induces a single degree sequence, as the number of nodes with degree α is

$$|V_\alpha| = \frac{1}{\alpha}\left(J_{\alpha\alpha} + \sum_{\beta=1}^{\Delta} J_{\alpha\beta}\right).$$



{(2,2), (2,1), (1,3), (1,1)}

Graphicality

A symmetric $\Delta\times\Delta$ matrix J is a graphical JDM if and only if:

1) $|V_\alpha|$ is an integer $\quad \forall\, 1\le\alpha\le\Delta$
2) $J_{\alpha\alpha}\le\binom{|V_\alpha|}{2} \quad \forall\, 1\le\alpha\le\Delta$
3) $J_{\alpha\beta}\le|V_\alpha||V_\beta| \quad \forall\, 1\le\alpha,\beta\le\Delta$ and $\alpha\ne\beta$.

With this result, one could consider using an approach similar to the one that worked for the degree sequences, based on a sequence of graphicality-conserving swaps.

However, this kind of swaps can modify the JDM, and so they cannot be used in this case.