WHEN IS A NETWORK COMPLEX? CONNECTANCE AS A DRIVER OF DEGREE DISTRIBUTION

T. POISOT

1. Introduction

- complexity in networks
- degree distribution vs. connectance
- “physical” argument

2. Statistical argument

Assuming and ecological network made of n species, and assuming undirected interactions with no self-edges (*e.g.* no cannibalism), there can be at most $M = n(n-1)/2$ interactions in this network, in which case it is a complete graph (the results presented below hold for both directed graphs, and graphs in which self-edges are allowed). This maximal number of links, M_n, represent the whole space of possible links. With this information in hand, it is possible to know the total number of possible networks given a number l of interactions.

If we term S_n the set of all possible M_n edges in a n-node network, then the number $G_{n,l}$ of possible networks with l links is the number of l-combinations of S_n, meaning that $G_{n,l} = C_l^{M_n}$, (where C_x^y is the binomial coefficient, *i.e.* the number of possible ways to pick x elements among y) or

$$G_{n,l} = \frac{M_n!}{l!(M_n-l)!}$$

Note that this number of possible networks include some graphs in which nodes have a degree of 0, and that in most ecological studies, such nodes will be discarded. In addition, in a null-model context [1,2], having unconnected nodes in random replicates will change the richness of the community, thus possibily biasing the value of randomized emerging properties. Finding out the number of graphs in which some nodes have a degree of 0 is similar to finding out how many networks exist with l links between $n-1$ nodes. If one node is removed from the network, there are C^{n}_{n-1} possible combinations of nodes (this simplifies to n). For each of these, there are $G_{n-1,l}$ possible networks configurations. Note that these networks will also include situations in which *more* than one species has a degree of 0, so that evaluating $G_{n-2,l}$ and so forth is not necessary. Calling $R_{n,l}$ the number of networks with n nodes and l edges in which all nodes have at least one edge attached, we can write

$$R_{n,l} = G_{n,l} - C^{n}_{n-1} \times G_{n-1,l}$$

We call the quantities R and G, respectively, the *realized* and *total* network space. They tell how many networks of n nodes and l edges exists. Based on these informations, we can make two predictions.

Prediction 1: Because $C^{y}_{x} = C^{y}_{y-x}$, it comes that the total network space is largest when $l = M_n/2$. As in this context the maximal number of edges is M_n, we define connectance as l/M_n, so $\max(G_{n,l})$ is reached at $Co = 1/2$. The algebraic expression of the maximum value of $R_{n,l}$ is hard to find, but simulations show that it also occurs around $Co = 1/2$. In other words, regardless of the number of nodes in a network, the "degrees of freedom" on network structure, as indicated by the size of the realized and total network spaces, are maximized for intermediate connectances.

Prediction 2: $R_{n,l}$ will become asimptotically closer to $G_{n,l}$ when l is close to M_n. In other words, there is only one way to fill a network of n nodes with M_n interactions, and in this situation there is no possibility to have nodes with a degree of 0. In the situation in which $l = M_n$, $G_{n,l} = C^{M_n}_{M_n} = 1$, given that $M_n > M_{n-1}$, it comes that $G_{n,l} = R_{n,l} = 1$.

We now illustrate these predictions using networks of 10 nodes, with a number of edges varying from 10 to M_{10} (*i.e.* 45 edges). As illustrated in Fig. 1, the size of the network space has a hump-shaped relationship with connectance, and the size of the realized network space becomes closer to the size of the total network space when connectance increases.

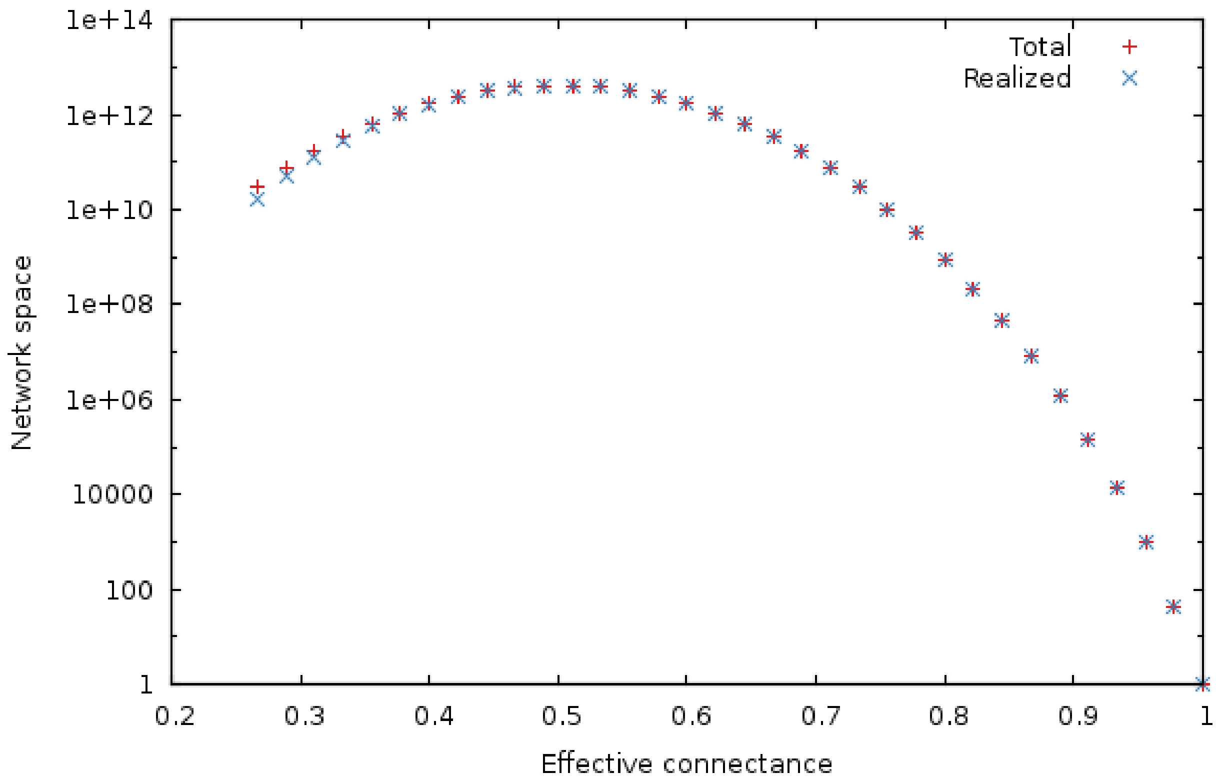

FIGURE 1. Size of the total and realized network space for $n = 10$. As predicted in the main text, (1) the size of network spaces peaks at $Co = 1/2$, and (2) the size of the realized network space becomes asymptotically closer to the size of the total network space when connectance increases.

In Fig. 2, we show that regardless of the network size, the relative size of the realized network space increases with connectance. The rate at which this increase happens is higher for networks with more nodes. However, in all cases, when connectance is low, there are only a very small proportion of total networks in which all nodes have at least one edge. This suggest that the structure of extremely sparse networks is also strongly constrained.

3. SIMULATIONS

In the previous part, we show mathematically that connectance (the number of realized *vs.* possible interaction), relative to the network size, determined the size of the *network sapce*, *i.e.* how many possible network combinations exist. Based on this, we can therefore

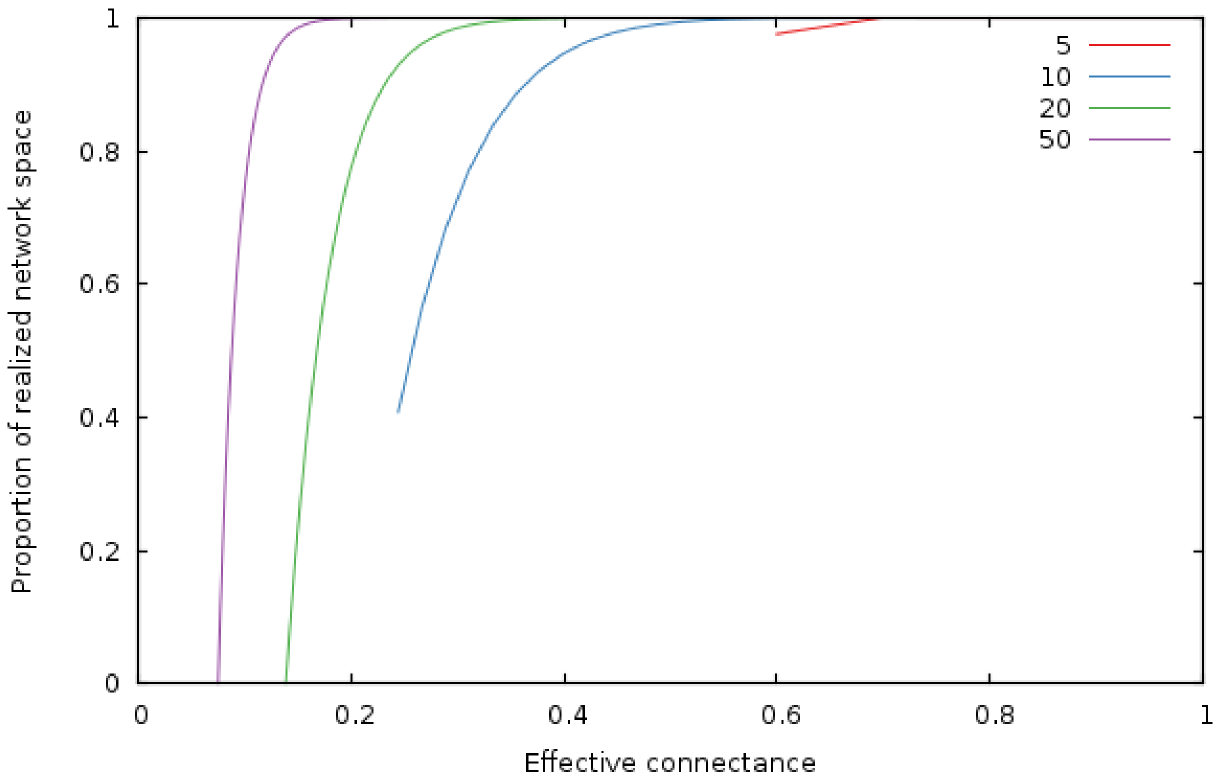

FIGURE 2. Relative size of the realized network space compared to the total network space when connectance increases, for four different network sizes.

predict that the degree distribution will be contingent upon network connectance. Specifically, we expect that the variance of the degree distribution, which is often used [3], will display a hump-shaped relationship with connectance. The mean, kurtosis, and skewness of the degree distribution should all vary in a monotonous way with connectance.

In the simulations below, we use a network of 30 nodes, filled with 35 to M_{30} interactions. The number of edges in the networks increases by steps of 10, and 500 networks are generated for each number of edges. The graphs generated are Erdős-Rényi ones, meaning that every potential interaction has the same probability of being realized [4]. We use an algorithm inspired by Reference 5, allowing to fix the number of edges in the graph. A total of 19000 networks are generated this way.

For each replicate, we measure the degree of all nodes (the degree distribution), and measure its variance, coefficient of variation, kurtosis, and skewness. In addition, for each network, we fit a power-law distribution on the sorted degree distribution using the least-squares method; we report the power-law exponent.

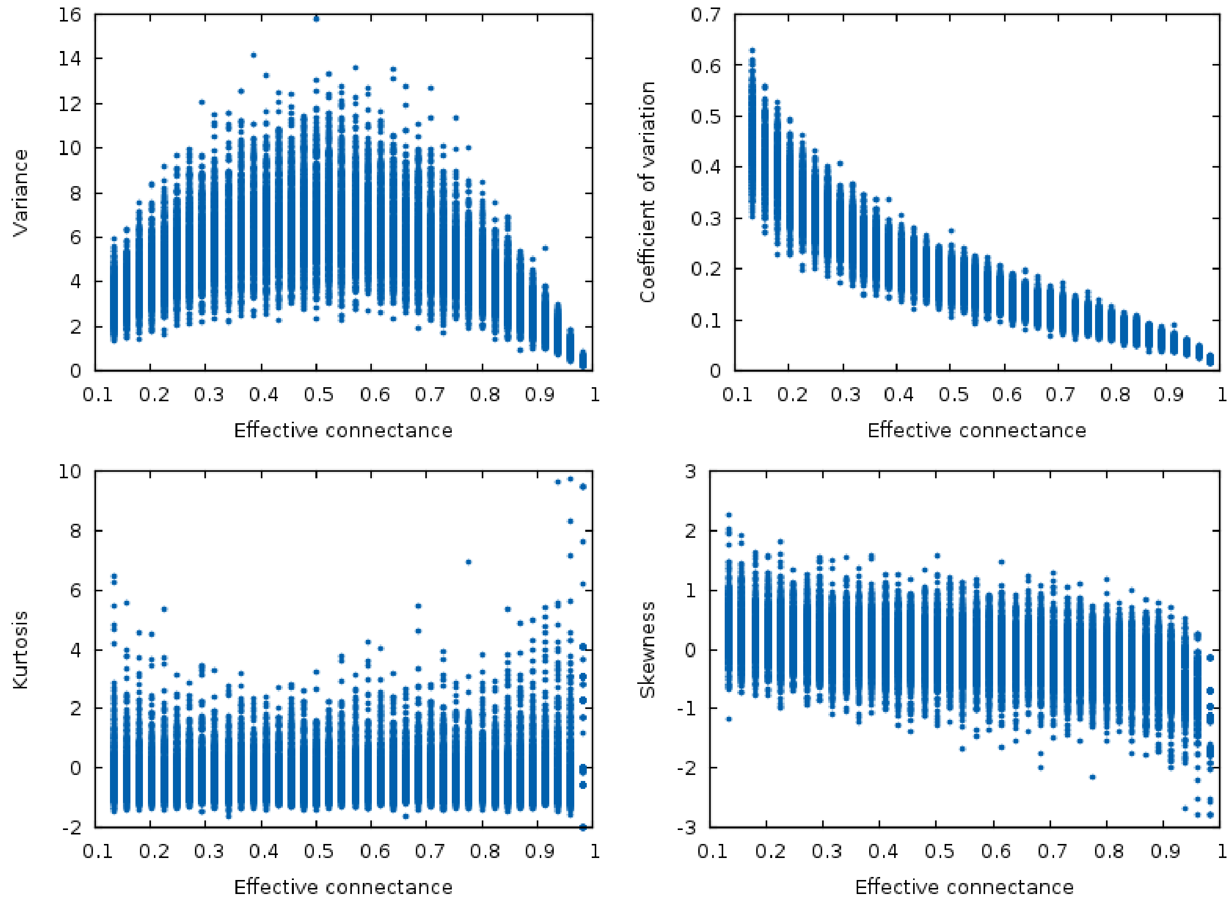

FIGURE 3. Statistical descriptors of the degree distribution of randomized networks, $n = 30$, increasing connectance. These results clearly show that central properties of the degree distribution are contingent upon connectance, at a given network size.

With the exception of the kurtosis, *all* statistical descriptors of the degree distribution were influenced by the effective connectance (Fig. 3). As predicted in the previous part, variance on the degree distribution is hump-shaped with regard to connectance ()

4. Practical consequences

- null models
- swap

1 Bascompte, J. *et al.* (2003) The nested assembly of plant–animal mutualistic networks. *Proceedings of the National Academy of Sciences of the United States of America* 100, 9383–9387

2 Fortuna, M.A. and Bascompte, J. *et al.* (2006) Habitat loss and the structure of plant-animal mutualistic networks. *Ecology Letters* 9, 281–286

3 Fortuna, M.A. *et al.* (2010) Nestedness versus modularity in ecological networks: two sides of the same coin? *Journal of Animal Ecology* 78, 811–817

4 Erdos, P. and Rényi, A. *et al.* (1959) On random graphs. *Publicationes Mathematicae Debrecen* 6, 290–297

5 Knuth, D.E. *et al.* (1997) Volume 2: Seminumerical Algorithms. *The Art of Computer Programming*