When is a network complex? Connectance drives degree distribution and emerging network properties

T. Poisot & D. Gravel

June 27, 2013

1 Introduction

Ecologists developped a strong interest for network theory, as it allowed to make sense of some of the complexity of ecological communities. In constrast to early approaches suchs as community modules (Holt 1997), using networks allows one to work at the whole community scale (Dunne 2006), thus accounting for feedbacks in species interactions (Berlow *et al.* 2009). Networks have often been called "complex" (Williams & Martinez 2000), on account of the fact that they represent objects (ecological communities) with complex (non-linear, sensitive to indirect interactions) dynamics. Because networks are multi-faceted objects with a rich range of structure, ecologists have been looking for emerging properties that can be easily measured and analyzed, and that relate to ecological properties and processes.

Since the beginning of ecological network litterature, connectance, *i.e.* the relative number of ecological interactions over the potential number, usually defined at the squared richness, has been recognized as a central network property (Yodzis 1980; Martinez 1992). In part, this success is to be attributed to the fact that connectance relates to early definitions of network complexity (Pimm 1982), and to the fact that connectance predicts reasonnably well some key dynamical properties of ecological networks (Dunne *et al.* 2002a; 2002b). More recently, attention shifted from connectance to degree distribution, that is the statistical properties of the distrubtion of number of interactions per species. Variation of degree distribution among networks has often been taken as evidence that assembly or interaction mechanisms differ (Vázquez 2005; Williams 2011), and increasingly refined method to estimate degree distribution have been devised (Williams 2009). Some authors proposed that degree distribution, rather than connectance, are driving the values of nestedness or modularity, which are important drivers of network dynamics (Fortuna *et al.* 2010).

However, it is worth asking if we were not too quick in discarding connectance in profit of degree distribution. A network, ecological or otherwise, can be viewed as a physical space that edges (interactions) occupy, the size of which is limited by the number of nodes (species). This means that there are physical constraints on the filling of a network, due to the fact that placing the first edge will limit the number of ways to place the remaining edges, and so on. For example, there is only one way to have a fully connected network, and there are a limited number of ways to have a network with the lower possible connectance. For this reason, and given the importance that degree distribution took in the recent years, it is important that we clearly understand how constrained degree distribution actually is, in relation to connectance. In this contribution, using an argument from combinatorial statistics and simulations of random networks under two different models, we present strong evidences that degree distribution, along with emerging network properties, are constrained (and can be predicted) by connectance. We discuss the consequences of our results for the comparison of different ecological networks, and for the generation of random networks in null-model analyses.

2 Statistical argument

Assuming an 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 qualitatively 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 (Bascompte *et al.* 2003; Fortuna & Bascompte 2006), 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 effective 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.

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. This is congruent with historical findings by Erdos & Rényi (1959), namely that the probabilty of each node being connected to the graph giant component increases with average degree (thus for high connectances, all nodes are likely to be connected to the giant component, hence no node has a degree of 0).

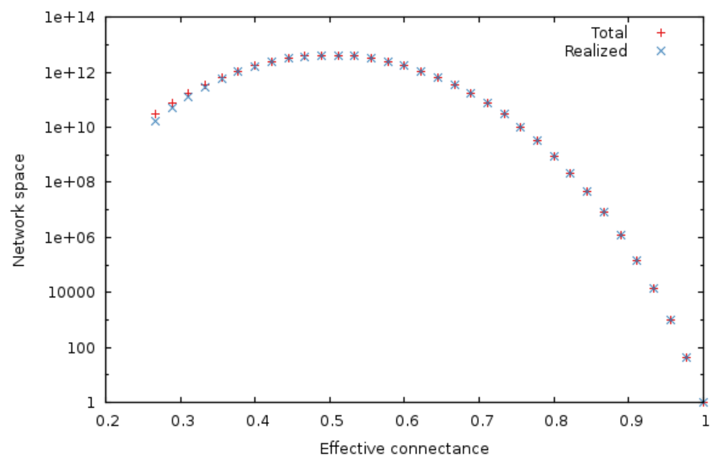

Figure 1: Size of the total and realized network space for $n = 10$. As predicted in the main text, (1) the size of network spaces peaks at $Co = 1/2$, and (2) the size of the realized network space becomes asymptotically closer to the size of the total network space when connectance increases.

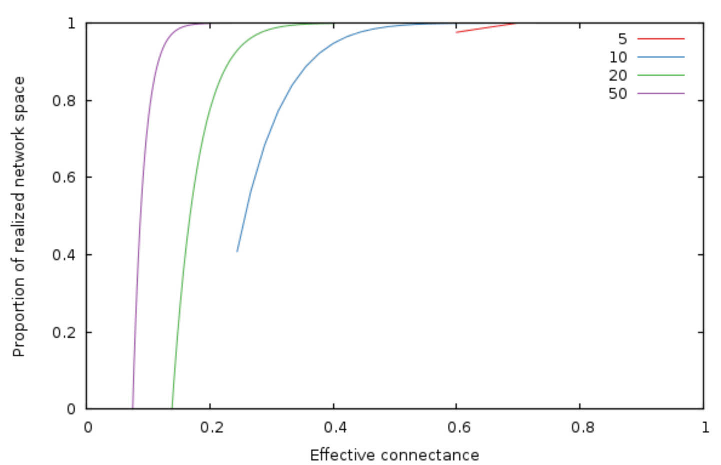

Figure 2: Relative size of the realized network space compared to the total network space when connectance increases, for four different network sizes.

3 Simulations

In the previous part, we show mathematically that connectance (the number of realized *vs.* possible interactions), relative to the network size, determined the size of the *network space*, *i.e.* how many possible network combinations exist. Based on this, we can therefore 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 (Fortuna *et al.* 2010), 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. We use two different routines to generate networks, that are contrasted in the way they distribute edges among nodes. First, we generate Erdős-Rényi graphs, meaning that every potential interaction has the same probability of being realized (Erdos & Rényi 1959). We use an algorithm inspired by Knuth (1997), allowing to fix the number of edges in the graph rather than the probability of an edge occuring, although the generated graphs have the same properties as the original model. A total of 19000 networks are generated this way. Second, we use the niche model of food webs (Williams & Martinez 2000), which generates networks under rules representing hypothetized mechanisms of prey-selection in empirical ecosystems. This particular model assumes that the existence of interactions is constrained by the position of species along a "niche" axis, for example body size. Other randomization methods for food webs exists, but as Stouffer et al. (2005) showed them to yield distribution of degree equivalent to the niche model under most conditions, we will not use them here. A total of 500 replicates for each level of number of links are generated. All networks generated with the two models satisfy the same criteria from the previous part, *i.e.* there are no self-edges and no nodes with a null degree.

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.

Qualitatively, both the random graphs and the niche networks behave exactly the same. 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, which implies that as average degree increases with connectance, the coefficient of variation of the degree distribution decreases at high connectances. Note also that the range of variances in the degree

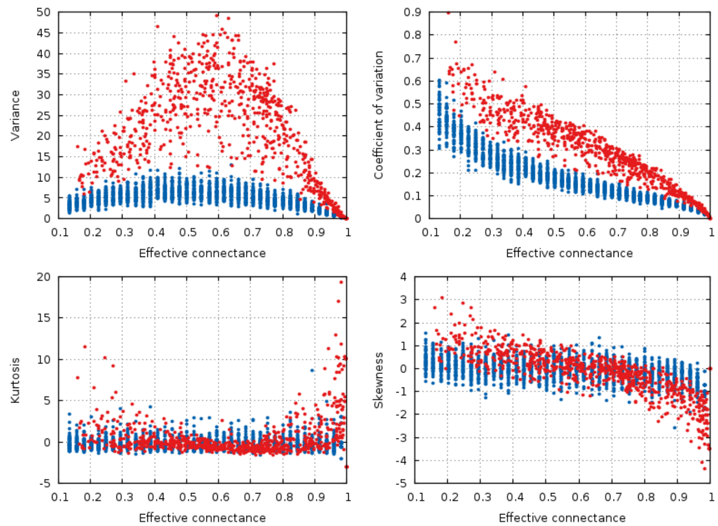

Figure 3: Statistical descriptors of the degree distribution of randomized networks, $n = 30$, increasing connectance. These results clearly show that central properties of the degree distribution are contingent upon connectance, at a given network size, and under a given network generation model.

distribution is higher at intermediate connectances, but lower at the extreme. Due to the fact that the Erdős-Rényi graphs we simulate are essentially Poisson random graphs, it is expected that the variance of their degree distribution would be lower than for the niche model, which in contrast *forces* strong difference in the degree of species according to their niche position.

Kurtosis seems to be unaffected by connectance. On the other hand, skewness decreases when connectance increases. This result is expected. Positively skewed distribution have longer or fatter right tails, indicating mostly low values (low degree): unconnected networks are made mostly of species with a weak generality (Schoener 1989). On the other hand, negative skewness indicate that most of the values in the distribution are high. Ecologically, it means that most species are wide-range generalists, which happens in densely connected networks. This bears importants ecological consequences, as it indicates that due to physical constraints acting on the filling of interactions within the graph, the specialists and generalist species are expected to be found together at intermediate connectances.

The estimate of the power-law exponent increases when connectance increases (Fig. 4). This indicates that the degree distribution flattens when connectance increases. Taken with the elements presented above, we show that all of the estimators of the degree distribution vary strongly with connectance of the network.

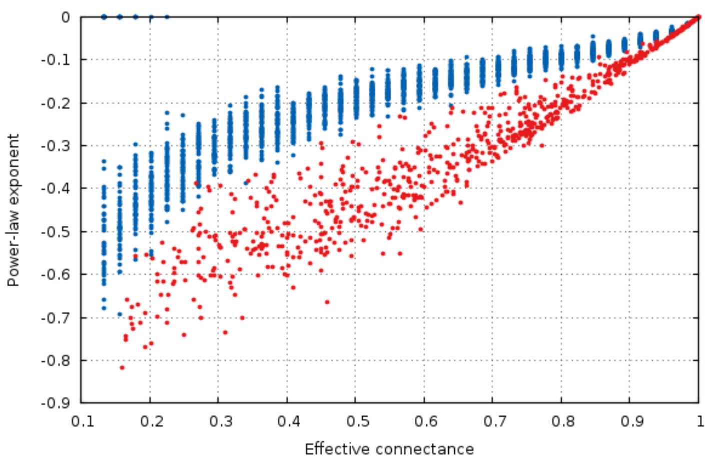

Figure 4: The estimate of the power-law exponent increases with connectance, arriving to a flat distribution for complete graphs.

4 Practical consequences

Randomized null models are often used to estimate how much a given emerging property deviates from its random expectation (Flores *et al.* 2011). Our results show two things. First, except for extremely high or low connectance, the proportion of the network space that will be explored using 10^3 or 10^4 replicates is orders of magnitude smaller than the *realized* network space. Although this is somewhat compensated by the fact that a part of these networks are isomorphic, the risk of infering deviation from the random expectation based on a drastically small sampling of the network space is real.

Second, generating null models with a low connectance is a computationally intensive task. When connectance decreases, the *realized* network space decreases faster than the *total* network space, meaning that the probability of picking a network with no un-attached nodes (which is simply $R_{n,l}/G_{n,l}$) goes toward zero. For this reason, classical rejection sampling (accept the random network if no nodes have no edges, reject else) if bound to take an unreasaonable amount of time in networks with low connectance. For this reason, using a purely random matrix shuffling as a starting point, then swapping interactions until no free nodes remain, seems to be a promising way to adress this problem.

5 Conclusions

Through statistical reasoning and simple simulations using models of random networks, we show that for a given number of species, the connectance of the network drives (i) how many different networks can beg enerated, and (ii) some key elements of the degree distribution. We observed both among and between model quantitative changes in degree distribution along a connectance gradient. The niche model is a particularly striking example of this, with the variance in the degree distribution increasing 50-fold when connectance moves from 0.1 to 0.5. This result has extremely practical implications for the comparison of networks, and network properties. As descriptors of degree distribution vary with connectance, connectance should be a covariate in all analyses. To some extent, the impact of connectance is lesser in the 0.05-0.3 range where most empirical food webs lies (although bipartite networks can have much higher connectances), but the effect is high enough that it should not be ignored: at equal number of species, networks with different connectances are expected to have different degree distributions.

- networks with a lot or a few interactions are actually simple, because extremely constrained

Finally, this analysis raises interesting ecological questions. Early analyses focusing on degree distribution argued that ecological mechanisms were responsible for the distribution shape (Vázquez 2005; Fortuna *et al.* 2010; Williams 2011). In this contribution, we show that connectance will impose a lower and higher limit for the shape of the degree distribution.Given this information, it's time to bring the debate fullcircle: is connectance the cause of observed network properties, or an emergent property of pairwise species interactions? As the later seems far more likely, it now makes sense to focus on why some networks deviate, or not, from the expected degree distribution knowing their connectance.

6 References

14.Bascompte, J., Jordano, P., Melián, C.J. & Olesen, J.M.*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.

3.Berlow, E.L., Dunne, J. a, Martinez, N.D., Stark, P.B., Williams, R.J. & Brose, U.*et al.* (2009). Simple prediction of interaction strengths in complex food webs. *Proceedings of the National Academy of Sciences of the United States of America*, 106, 187–91.

2.Dunne, J.A. *et al.* (2006). The Network Structure of Food Webs. In: *Ecological networks: Linking structure and dynamics* (eds. Dunne, J.A. & Pascual, M.). Oxford University Press, pp. 27–86.

8.Dunne, J.A., Williams, R.J. & Martinez, N.D.*et al.* (2002a). Food-web structure and network theory: the role of connectance and size. *Proceedings of the National Academy of Sciences of the United States of America*, 99, 12917.

9.Dunne, J.A., Williams, R.J. & Martinez, N.D.*et al.* (2002b). Network structure and biodiversity loss in food webs: robustness increases with connectance. *Ecology Letters*, 5, 558–567.

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

20.Flores, C.O., Meyer, J.R., Valverde, S., Farr, L. & Weitz, J.S.*et al.* (2011). Statistical structure of host-phage interactions. *Proceedings of the National Academy of Sciences of the United States of America*, 108, E288–297.

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

13.Fortuna, M.A., Stouffer, D.B., Olesen, J.M., Jordano, P., Mouillot, D. & Krasnov, B.R.*et al.* (2010). Nestedness versus modularity in ecological networks: two sides of the same coin?. *Journal of Animal Ecology*, 78, 811–817.

1.Holt, R.D. *et al.* (1997). Community modules. In: *Multitrophic Interactions in Terrestrial Ecosystems* (eds. Gange, A.C. & Brown, V.K.). Blackwell Science, pp. 333–349.

17.Knuth, D.E. *et al.* (1997). Volume 2: Seminumerical Algorithms. *The Art of Computer Programming*, 192.

6.Martinez, N.D. *et al.* (1992). Constant connectance in community food webs. *The American Naturalist*, 139, 1208–1218.

7.Pimm, S.L. *et al.* (1982). *Food Webs*. Chapman & Hall, London.

19.Schoener, T.W. *et al.* (1989). Food webs from the small to the large. *Ecology*, 70, 1559–1589.

18.Stouffer, D.B., Camacho, J., Guimera, R., Ng, C.A. & Amaral, L.A.N.*et al.* (2005). Quantitative patterns in the structure of model and empirical food webs. *Ecology*, 86, 1301–1311.

11.Vázquez, D.P. *et al.* (2005). Degree distribution in plant-animal mutualistic networks: forbidden links or random interactions?. *Oikos*, 108, 421–426.

12.Williams, R.J. *et al.* (2009). Simple MaxEnt models explain food web degree distributions. *Theoretical Ecology*, 3, 45–52.

10.Williams, R.J. *et al.* (2011). Biology, Methodology or Chance? The Degree Distributions of Bipartite Ecological Networks. *PLoS One*, 6, e17645.

4.Williams, R.J. & Martinez, N.D. *et al.* (2000). Simple rules yield complex food webs. *Nature*, 404, 180–183.

5.Yodzis, P. *et al.* (1980). The connectance of real ecosystems. *Nature*, 284, 544–545.