Statistical Complexity of Random Graphs: Or, What Makes a Small-World Graph a "Complex Network" Anyway?

Cosma Rohilla Shalizi[1]
[1] *Center for the Study of Complex Systems, 4485 Randall Laboratory, University of Michigan, Ann Arbor, MI 48109*[*]

(Dated: Notes begun 1 February 2004; last revised 8 March 2004)

I give you a complexity measure for random graphs. Unfortunately it does not seem at all easy to calculate this for anything of interest. Still, it may be useful someday.

The field of "complex networks" studies those which lie "between order and randomness". It's easy to see, in a vague way, how this applies to things like Watts-Strogatz [1] "small world" graphs, which interpolate between completely random Erdos-Renyi graphs, and completely regular lattices, since Watts-Strogatz graphs are obtained by adding a small number of long-range connections to regular lattices. It is less clear how this idea applies to, say, Barabasi-Albert scale-free networks [2], or how one might use it to say exactly which networks are complex, or *how* complex they are. This might be of some moment for questions in ecology relating food-web complexity to stability, resilience, etc., and similar questions regarding the development of social complexity. Or it may be pointless quantification-mongering; who's to say?

Let us consider an ensemble of random graphs, with a fixed number of nodes n. We shall worry about the interacction or dependencies between the *edges* of the graph. In an Erdos-Renyi graph, the edges are (by definition) all completely independent. In a regular lattice, the edges are *also* probabilistically independent, since they are always either present or absent. There *is* dependence between the edges in both Albert-Barabasi networks (due to the preferential attachment mechanism) and the original Watts-Strogatz networks (due to the random re-wiring mechanism).

Notation. Define G_{ij} as the binary random variable which indicates whether there is an edge connecting node i to node j, or not. (For simplicity, I'll assume undirected edges, but I don't think it makes a big difference. Similarly, we should be able to accomodate multiple types of edge with minor tweaking.) Define $\overline{G_{ij}}$ as the set of all G_{kl} such that $kl \neq ij$, i.e., all the edge variables except G_{ij}.

Now one measure of the dependence between edges is simply how much information $\overline{G_{ij}}$ gives us about G_{ij}. This is the "effective measure complexity" of Grassberger [3], or, more intuitively, the "predictive information" of Bialek, Tishby and Nemenmann [4, 5],

$$\begin{aligned} I_{\text{pred}}(ij) &= I(G_{ij};\overline{G_{ij}}) \\ &= H(\Pr(G_{ij}=1)) \\ &\quad -\sum_{\overline{g_{ij}}} \Pr(\overline{G_{ij}}=\overline{g_{ij}}) H(\Pr(G_{ij}=1|\overline{G_{ij}}=\overline{g_{ij}})) \end{aligned}$$

where $H(p)$ is the entropy of a binary random variable with probability of success p. Manifestly, it is zero for both Erdos-Renyi and completely ordered graphs. In statistical terms, it is the uncertainty remaining in G_{ij} when all available data about the graph is optimally utilized.

In the context of time series [6] and spatiotemporal dynamics [7], it has proved better to know how much information is required for such optimal forecasts, rather than the information such forecasts provide. This is the "forecast" or "statistical" complexity of Grassberger [3] and Crutchfield and Young [8]. It, too, can be defined in for random graphs. Define a function ϵ_{ij} for each pair of nodes:

$$\epsilon_{ij}(\overline{g_{ij}}) = \left\{\overline{g_{ij}}' \ : \ \Pr(G_{ij}|\overline{G_{ij}}=\overline{g_{ij}}) = \Pr(G_{ij}|\overline{G_{ij}}=\overline{g_{ij}}')\right\}$$

That is, ϵ associates a configuration of the graph to all the other configurations which lead to the same probability distribution for G_{ij}. If certain edges are simply irrelevant to that probability, they are not actually needed to compute ϵ. It is the minimal sufficient statistic [9] for predicting G_{ij}. (The argument for minimal sufficiency follows the form given in [6]).

The forecasting complexity associated with the pair ij is then

$$C_{ij} = I(\epsilon(\overline{G_{ij}});\overline{G_{ij}})$$

and that of the overall graph is the mean over all possible pairs:

$$C = \frac{2}{n(n-1)}\sum_{i,j} C_{ij}$$

It is easily checked that $C_{ij} \geq I_{\text{pred}}(ij)$, and so too for the graph averages.

Observation made by Cris Moore: In any model where the rest of the graph determines, with probability 1, whether there is an edge between i and j, there are just two equivalence classes over $\overline{G_{ij}}$. (This is the case, for instance, if each node has a fixed coordination number n_i. We can then explicitly calculate the minimal sufficient statistic, $\Gamma_{ij} \equiv \sum_{k\neq i,\neq j} G_{ik}$, and say that $G_{ij} = n_i - \Gamma_{ij}$.) Hence C is at most 1 bit. In fact in this case, $C_{ij} = H[G_{ij}]$.

Observation made by CRS: Another case where an explicit sufficient statistic is available is in the case of an equilibrium ensemble of random graphs [10, sec. V]. Here the probability of a given graph g is proportional to

---

[*] Electronic address: `cshalizi@umich.edu`

$e^{-U(g)}$, for some function U taking the role of a Hamiltonian. Hence

$$\Pr(G_{ij} = 1|\overline{G_{ij}} = \overline{g_{ij}}) = \frac{e^{-U(G_{ij}=1,\overline{g_{ij}})}}{e^{-U(G_{ij}=1,\overline{g_{ij}})} + e^{-U(G_{ij}=0,\overline{g_{ij}})}}$$

Define $\Delta U(\overline{g_{ij}}) \equiv U(G_{ij} = 1, \overline{g_{ij}}) - U(G_{ij} = 0, \overline{g_{ij}})$. Then

$$\Pr(G_{ij} = 1|\overline{G_{ij}} = \overline{g_{ij}}) = \frac{1}{1 + e^{\Delta U(\overline{g_{ij}})}}$$

and this tells us that $\Delta U(\overline{g_{ij}})$ is the minimal sufficient statistic (since there is a bijection between the conditional probability and ΔU).

Queries:

1. Are there values of any of the usual graph parameters [10], e.g. degree correlation, which imply positive statistical complexity? Note that, following the next item, positive clustering is compatible with zero statistical complexity.

2. Can statistical complexity be calculated analytically for any of the common network models, such as the Watts-Strogatz or Albert-Barabasi model? Note that the Newman-Watts model [11, 12], intended as a more-tractable replacement of the original Watts-Strogatz model, is not helpful here, or at any rate not encouraging. In this model, all the original edges remain, and long-range shortcuts are added completely at random. The effect is that every edge is still statistically independent of every other edge, though the network shows nonzero clustering. This shows that the question in (1) must be answered with some care. It also, frankly, casts some doubt on the value of the complexity measure proposed here!

3. Can the forceast complexity be calculated for the asmyptotic distribution of the Clauset-Moore model [13]? If so, this would let us answer whether that model really is self-organizing (following the argument in [14]).

Acknowledgments. Thanks to Cris Moore, Mark Newman and Aaron Clauset.

---

[1] D. J. Watts and S. H. Strogatz, Nature **393**, 440 (1998).

[2] A.-L. Barabási and R. Albert, Science **286**, 509 (1999), URL `http://arxiv.org/abs/cond-mat/9910332`.

[3] P. Grassberger, International Journal of Theoretical Physics **25**, 907 (1986).

[4] W. Bialek and N. Tishby, *Predictive information*, E-print, arxiv.org (1999), URL `http://arxiv.org/abs/cond-mat/9902341`.

[5] W. Bialek, I. Nemenman, and N. Tishby, Neural Computation **13**, 2409 (2001), URL `http://arxiv.org/abs/physics/0007070`.

[6] C. R. Shalizi and J. P. Crutchfield, Journal of Statistical Physics **104**, 817 (2001), URL `http://arxiv.org/abs/cond-mat/9907176`.

[7] C. R. Shalizi, Discrete Mathematics and Theoretical Computer Science **AB(DMCS)**, 11 (2003), URL `http://arxiv.org/abs/math.PR/0305160`.

[8] J. P. Crutchfield and K. Young, Physical Review Letters **63**, 105 (1989), URL `http://www.santafe.edu/~cmg/compmech/pubs/ISCTitlePage.htm`.

[9] S. Kullback, *Information Theory and Statistics* (Dover Books, New York, 1968), 2nd ed.

[10] M. E. J. Newman, SIAM Review **45**, 167 (2003), URL `http://arxiv.org/abs/cond-mat/0303516`.

[11] M. E. J. Newman and D. J. Watts, Physics Letters A **263**, 341 (1999), URL `http://www.arxiv.org/abs/cond-mat/9903357`.

[12] M. E. J. Newman and D. J. Watts, Physical Review E **60**, 7332 (1999), URL `http://www.arxiv.org/abs/cond-mat/9904419`.

[13] A. Clauset and C. Moore, Physical Review Letters **submitted** (2003), URL `http://www.arxiv.org/abs/cond-mat/0309415`.

[14] C. R. Shalizi, K. L. Klinkner, and R. Haslinger, Physical Review Letters **93**, 118701 (2004), URL `http://arxiv.org/abs/nlin.AO/0409024`.