

Contents

1	Random and scale-free networks	1
1.1	Food webs	1
1.2	Random networks	3
1.3	Scale-free networks	5

1 Random and scale-free networks

Our study of percolation examines connectivity on a lattice.

We now proceed to an examination of aspects of connectivity without regard to a fixed underlying geometry.

We begin with an example from the natural world: food webs.

1.1 Food webs

Food webs describe which kind of organisms eat which other kinds.

The simplest food web is a *food chain*.

At the base of the food chain are *producers*, or *autotrophs*.

At the next *trophic level* one finds *consumers*, or *heterotrophs*, which live off of primary producers.

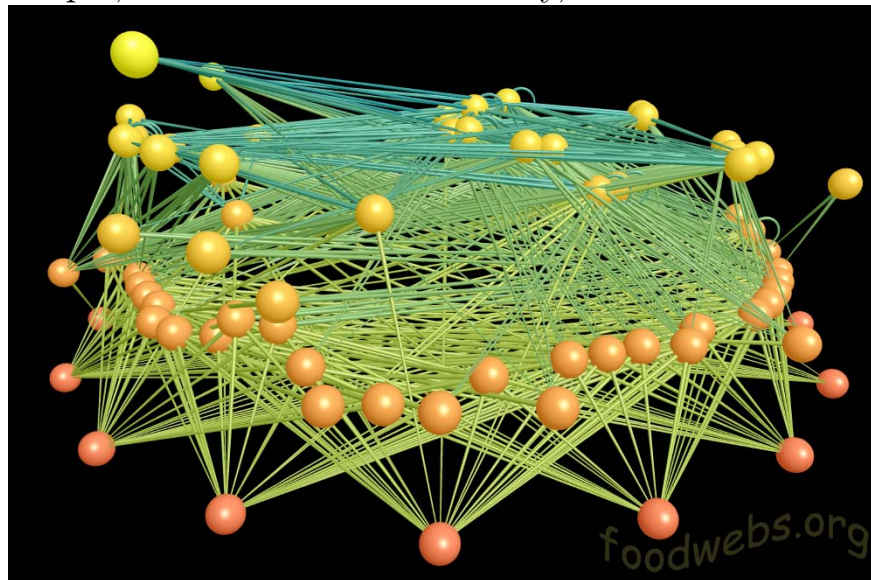
At the next higher trophic level are the consumers of consumers, etc.

We can group organisms into *trophic species*, i.e. functional groups that con-

tain organisms that appear to eat and be eaten by the exact same species within a food web.

Connections between trophic species then provide a simple representation of a food web.

Here's an example, from the East River Valley, 120 miles southwest of Denver:



www.foodwebs.org*

Image courtesy of Pacific Ecoinformatics and Computational Ecology Lab. Used with permission.

The web structure in the image is organized vertically, with node color representing trophic level. Red nodes represent basal species, such as plants and detritus, orange nodes represent intermediate species, and yellow nodes represent top species or primary predators.

Note that this image has 4 trophic levels, with some species well connected and others not so much.

In what follows, we focus on models that predict distributions of connectivity to see if we can learn more.

*Image produced with FoodWeb3D, written by R.J. Williams and provided by the Pacific Ecoinformatics and Computational Ecology Lab [1].

1.2 Random networks

Reference: Newman [2].

A perfectly plausible null hypothesis is to imagine that species represent nodes and predator-prey relationships represent links (edges) that are chosen randomly.

For simplicity, we ignore the *directed* nature of the predator-prey link, and seek information only about the existence of the link.

We then define a *random network* as follows:

- We specify n nodes (i.e., species).
- We specify the probability p that any two nodes are connected.

Note that for any given node, there are

$n - 1$ possible connections.

Since there are n possible starting points for those $n - 1$ connections, there are

$$\frac{n(n-1)}{2} \text{ possible links,}$$

where the factor of $1/2$ arises because we care only about the existence of the link, not its direction (i.e, we don't count links twice).

The expected number of links $\ell(n)$ in the random graph is

$$\begin{aligned} \ell(n) &= p \cdot (\text{number of possible links}) \\ &= p \frac{n(n-1)}{2}. \end{aligned}$$

Define

$$k_i = \text{number of links to node } i.$$

k_i is called the *degree* of node i .

The mean number of links attached to a node is

$$\langle k \rangle = 2 \frac{\ell(n)}{n}$$

The factor of 2 arises because each link is attached to 2 nodes. We thus have

$$\langle k \rangle = \frac{2pn(n-1)}{2n} = p(n-1) \simeq pn.$$

$\langle k \rangle$ is called the *mean degree* of the random network.

We seek the *degree probability distribution*: the probability of observing a node with degree k .

To obtain a node with degree k , we must have k “successful” connections, each with probability p , and $(n-1-k)$ unsuccessful connections.

The number of possible combinations of such connections is given by the binomial coefficient

$$\binom{n-1}{k} = \frac{(n-1)!}{(n-1-k)! k!}.$$

The probability P_k of observing k connections at a given node is then given by the binomial distribution

$$P_k = \binom{n-1}{k} p^k (1-p)^{n-1-k}.$$

The binomial distribution is a bell-shaped curve. If $n \rightarrow \infty$ and $p \rightarrow 0$ while $\langle k \rangle = np$ remains constant, then the binomial distribution converges to the Poisson distribution

$$P_k = \frac{\langle k \rangle^k e^{-\langle k \rangle}}{k!},$$

the mean and variance of which are both $\langle k \rangle$.

The *random-network* null-hypothesis thus makes a specific prediction: degree distributions are bell-shaped, with a mean and variance of $\langle k \rangle$.

Real food web data is, however, inconsistent with this prediction [3].

1.3 Scale-free networks

Reference: Albert and Barabási [4]

We now consider a model of a *growing* network. It too should be considered a null hypothesis rather than a mechanistic model of food webs.

There are two ingredients:

- *Growth.* We start with m_0 nodes with pre-existing connections. At each time step we add a new node with $m \leq m_0$ links to pre-existing nodes.
- *Preferential attachment.* New links tend to attach to already well-attached nodes. Specifically: the probability Π that a new node is connected to node i is proportional to k_i , the degree of the i th node:

$$\Pi(k_i) = \frac{k_i}{\sum_j k_j}.$$

We can easily see that after t time steps, there are

$$t + m_0 \text{ nodes and } mt \text{ links.}$$

To derive the degree distribution, we assume that k_i and t are continuous.

Then the rate at which k_i changes is proportional to $\Pi(k_i)$:

$$\frac{dk_i}{dt} = m\Pi(k_i).$$

The factor of m arises by identifying the units of time with the interval between time steps and recalling that m links are added at each time step.

We next substitute for $\Pi(k_i)$:

$$\frac{dk_i}{dt} = m \frac{k_i}{\sum_j k_j}.$$

Since there are mt links after t units of time, the sum

$$\sum_j k_j(t) = 2mt,$$

where the factor of 2 comes from counting each link twice (both “incoming” and “outgoing”), and we have ignored the small correction (for large t) that would arise from not counting the initial links.

Substituting the sum into the growth equation above, we obtain

$$\frac{dk_i}{dt} = \frac{k_i}{2t}.$$

Separating variables, we have

$$\frac{dk_i}{k_i} = \frac{dt}{2t}.$$

Integrating both sides, we have

$$\ln k_i(t) = \frac{1}{2} \ln t + \text{const.} \quad \Rightarrow \quad k_i = Ct^{1/2}.$$

Now define

$$t_i = \text{time of inception of the } i\text{th node.}$$

The initial condition is then $k_i(t_i) = m$, and therefore

$$k_i(t) = m \left(\frac{t}{t_i} \right)^{1/2}. \quad (1)$$

The number of links attached to the i th node therefore grows like $t^{1/2}$, but the prefactor $m/\sqrt{t_i}$ depends on its time of inception.

We seek the continuous probability density function $p(k)$ that a node has degree k .

To do so, we first define the cumulative probability distribution function

$$P[k_i(t) < k] = \text{probability that the } i\text{th node has fewer than } k \text{ links.}$$

We rewrite this expression using (1):

$$P[k_i(t) < k] = P \left[m \left(\frac{t}{t_i} \right)^{1/2} < k \right] \quad (2)$$

$$= P \left(t_i > \frac{m^2 t}{k^2} \right). \quad (3)$$

Now recall that there are $m_0 + t$ nodes, and one node is added at each time step.

Thus the time t_i at which the i th node is added is uniformly distributed between 0 and $m_0 + t$ (assuming the same rate for the first m_0 nodes).

The probability density function $p(t_i)$ is therefore constant over that interval. Since it must integrate to unity, we have

$$p(t_i) = \frac{1}{m_0 + t}.$$

Consequently the RHS of (3) is

$$\begin{aligned} P\left(t_i > \frac{m^2 t}{k^2}\right) &= \int_{m^2 t/k^2}^{m_0+t} p(t_i) dt_i \\ &= 1 - \int_0^{m^2 t/k^2} \frac{dt_i}{m_0 + t} \\ &= 1 - \frac{m^2 t}{k^2(m_0 + t)}. \end{aligned}$$

Inserting this into (3), we obtain

$$P[k_i(t) < k] = 1 - \frac{m^2 t}{k^2(m_0 + t)}.$$

We can now obtain $p(k)$ by noting that

$$\begin{aligned} p(k) &= \frac{d}{dk} P[k_i(t) < k] \\ &= -\frac{d}{dk} \left(\frac{m^2 t}{k^2(m_0 + t)} \right) \\ &= \frac{2m^2 t}{m_0 + t} k^{-3}, \end{aligned}$$

which in the limit of large t becomes

$$p(k) \sim 2m^2 k^{-3}.$$

This degree distribution is distinctive for two reasons:

- Unlike the bell-curve of random networks centered at $\langle k \rangle$, here the degree distribution is one-sided, with its maximum at 0.
- The power-law form of the distribution means that it is *scale-free*.

We say that the distribution is scale-free because its form is independent of scale (and therefore time).

To be precise, note that if we have a power-law distribution

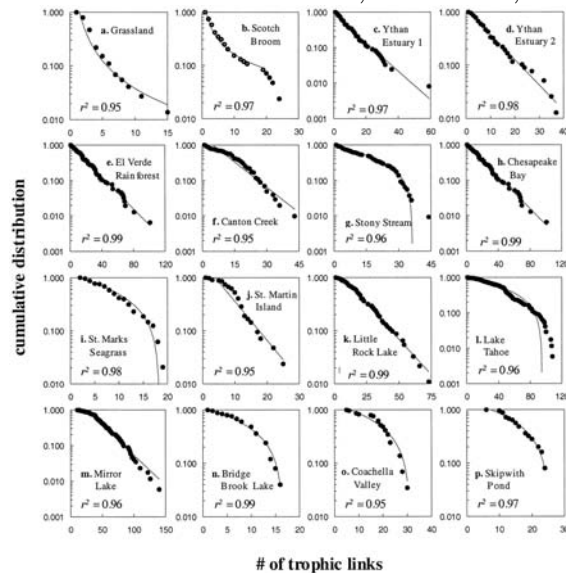
$$p(x) = x^{-\alpha},$$

then it is invariant, except for a prefactor, under the change of scale $x \rightarrow bx$:

$$p(bx) = (bx)^{-\alpha} = b^{-\alpha}p(x).$$

The power law is the only distribution of this type [5].

Are real food webs scale-free? The answer, it seems, is *sometimes*:



Dunne et al. [3]

Copyright (2002, 2006) National Academy of Sciences, U.S.A. Used with permission.

In this figure, power-law degree distributions are concave upward, exponential distributions are straight lines, and uniform distributions are concave downward.

Though hardly the rule in ecology, a great deal of other networks are apparently scale free. Some examples [4]:

- The World Wide Web and the internet.
- Scientific citations and co-authorship networks.
- Metabolic networks. (One considers the nodes to be substrates (e.g., ATP, H₂O) and the links to represent chemical reactions.)

The ubiquity of such networks leads one to ask if the power-law behavior is truly significant. For an interesting perspective on this question, see Ref. [6].

Finally, if indeed metabolism is in some way scale-invariant, might that property carry over to the scale of entire ecosystems and therefore biogeochemical cycles?

We don't know, but we now turn to discuss some interesting ways in which metabolism expresses itself at the scale of ecosystems.

References

- [1] Yoon, I. *et al.* Webs on the Web (WoW): 3D visualization of ecological networks on the WWW for collaborative research and education. In *Proceedings of the IS&T/SPIE Symposium on Electronic Imaging, Visualization and Data Analysis 5295*, 124–132 (2004).
- [2] Newman, M. E. J. *Networks: An Introduction* (Oxford University Press, Oxford, U.K., 2010).
- [3] Dunne, J. A., Williams, R. J. & Martinez, N. D. Food-web structure and network theory: The role of connectance and size. *Proceedings of the National Academy of Sciences USA* **99**, 12917–12922 (2002).
- [4] Albert, R. & Barabási, A.-L. Statistical mechanics of complex networks. *Reviews of Modern Physics* **74**, 47–97 (2002).
- [5] Newman, M. E. J. Power laws, Pareto distributions, and Zipf's law. *Contemporary Physics* **46**, 323–351 (2005).
- [6] Fox Keller, E. Revisiting “scale-free” networks. *BioEssays* **27**, 1060–1068 (2005).

MIT OpenCourseWare
<http://ocw.mit.edu>

12.086 / 12.586 Modeling Environmental Complexity
Fall 2014

For information about citing these materials or our Terms of Use, visit: <http://ocw.mit.edu/terms>.