

REPORT

# The evolution of network topology by selective removal

Marcel Salathé<sup>1,†</sup>, Robert M. May<sup>2</sup>  
and Sebastian Bonhoeffer<sup>1</sup>

<sup>1</sup>*Ecology & Evolution, ETH Zürich, ETH-Zentrum  
CHN, CH 8092, Switzerland*

<sup>2</sup>*Mathematical Biology Group, The Tinbergen Building,  
South Parks Road,  
Oxford OX1 3PS, UK*

The topology of large social, technical and biological networks such as the World Wide Web or protein interaction networks has caught considerable attention in the past few years (reviewed in Newman 2003), and analysis of the structure of such networks revealed that many of them can be classified as broad-tailed, scale-free-like networks, since their vertex connectivities follow approximately a power-law. Preferential attachment of new vertices to highly connected vertices is commonly seen as the main mechanism that can generate scale-free connectivity in growing networks (Watts 2004). Here, we propose a new model that can generate broad-tailed networks even in the absence of network growth, by not only adding vertices, but also selectively eliminating vertices with a probability that is inversely related to the sum of their first- and second order connectivity.

**Keywords:** scale-free networks; network topology; preferential attachment

The most common explanation for the emergence of scale-free structures in networks is a process referred to as preferential attachment. In a nutshell, preferential attachment assumes that new vertices attach preferentially to already well-connected vertices in a network (Simon 1955; Barabási & Albert 1999). Various mechanisms leading to preferential attachment have been described, some of them including conscious choice (e.g. the world wide web or collaboration networks; Watts 2004), others assuming an ongoing duplication of nodes and edges (e.g. protein and gene networks; Barabási & Oltvai 2004; Berg *et al.* 2004). However, these mechanisms usually require that the networks are growing, a precondition that is certainly not given in all networks (Dorogovtsev & Mendes 2000). We propose an alternative model that results in the emergence of broad-tailed connectivity

distributions that are best described by a power-law (hereafter referred to as scale-free-like networks), both in growing and non-growing networks, without assuming any form of cognition, choice or duplication, leading to apparent preferential attachment.

In the model presented here, the sum of first- and second-order connections of a vertex determines the likelihood of selective removal. A randomly connected network with initially  $m$  vertices is subjected to selection during  $t$  generations. In each generation, a proportion ( $p$ ) of the vertices is removed from the network. To simulate a selective process, a random vertex is chosen and removed with a probability of  $1/c^a$ , where  $c$  is the sum of its first- and second-order connectivity, that is, the number of vertices with a maximal distance of 2 (i.e. the number of neighbours and secondary neighbours), and  $a$  measures the strength of selection, since higher values of  $a$  result in stronger selection against low values of  $c$ . This procedure is repeated until  $r=pN$  vertices are removed from the network (where  $N$  is the number of all nodes in the network). Additionally, random vertices can be removed independent of their fitness value with a natural vertex death rate  $d$ . If after removal the system contains a number of unconnected vertices (denoted by  $u$ ), these are removed too. Finally,  $v$  new vertices are added and connected to random vertices of the network. The resulting network grows when  $v > r + u$  or is constant in size when  $v = r + u$ . The net number of added vertices per generation is denoted by  $n = v - (r + u)$ . Thus, after  $t$  generations, the number of vertices in the network is  $m + tn$ , where  $m$  is the initial network size. Note that no preferential attachment is occurring during the entire process.

We first looked at growing networks by running simulations over 3000 time steps, starting on a network consisting of only two vertices connected to each other. At each time step, the network grew by one vertex, resulting in a final network of 3002 vertices. The size of these simulated networks is of the general order of larger real-world datasets. Simulations of the described selective removal process resulted in the self-organization of networks with vertex connectivities  $p(k)$  that are best described by a power-law in a wide range of parameter values (figure 1 and table 1). Since network growth has been shown to be an essential component of the preferential attachment mechanism, we wanted to know if the selective removal mechanism would create scale-free-like distributions when applied to non-growing networks. Simulations were started on initially random networks with 3000 vertices, characterized by a Poisson connectivity distribution, and at each time step, the number of randomly attached vertices was equal to the number of removed vertices (thus  $n=0$ ). We found that the connectivity distribution was best described by a power-law very quickly (less than 500 time steps for most of the simulations—for a typical distribution, see figure 1). The scaling of the simulated networks was conserved even over large time-scales ( $t=100\,000$ ; table 1); thus growth does not appear to be an essential component for the selective removal mechanism to generate scale-free-like networks.

<sup>†</sup>Author for correspondence (marcel.salathe@env.ethz.ch).

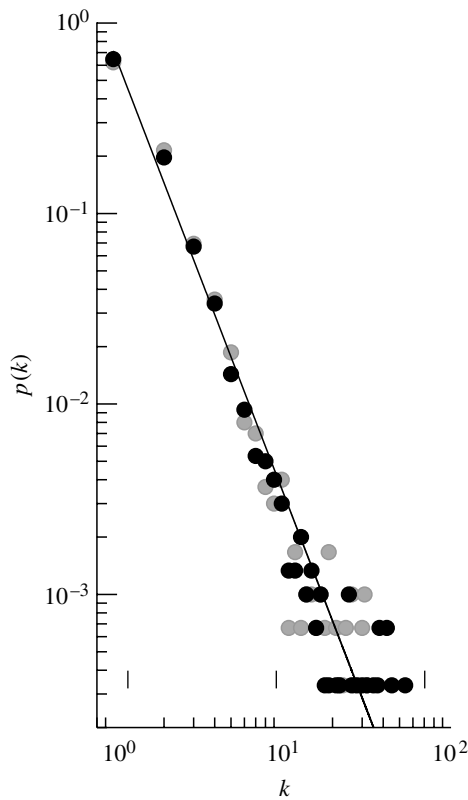


Figure 1. A typical connectivity distribution function in a non-growing network with  $a=2.5$  after 1000 (grey circles) and 100 000 (black circles) time steps.  $p(k)$  represents the frequency of vertices with exactly  $k$  connections. The straight line has a slope of  $\gamma=2.3$ . Simulation parameters:  $r=0.05$ ,  $d=0.001$ ,  $m=3000$ ,  $n=0$ . The distributions are best fit by a power-law (see text).

While many investigated network distributions have only been fitted by eye (Dorogovtsev & Mendes 2003), we follow the approach of Stumpf (submitted). This approach employs the Akaike-information criterion (AIC), which uses the maximum likelihood estimates (MLE) and the number of parameters of a model, to find the distribution that best describes the data (Akaike 1983; Burnham & Anderson 2002). Here, we consider the Poisson, exponential and power-law distribution. Since all of them can be described by one parameter, it is sufficient to compare only the MLE to identify the best model. Our results (table 1) show that scale-free-like structures emerge if selection is strong enough (threshold value for  $a \approx 1.5$ ). However, an analytical approximation (see appendix A) shows that for the limiting case where selection is extremely strong, the asymptotic distribution is again exponential.

In scale-free networks, the probability  $p(k)$  that a vertex is connected to  $k$  other vertices decays with  $p(k) \sim k^{-\gamma}$ . Empirical evidence from real-world networks has shown that the degree exponent  $\gamma$  lies in the range of about 2 to 4. The scale-free-like networks that emerge by selective removal have connectivity distributions following power-laws with degree exponents  $1.8 < \gamma < 3.9$  depending on the parameter values used. Thus, selective removal can account for all degree exponents observed in real-world networks.

Interestingly, the described selective removal mechanism does not produce scale-free-like networks if fitness is defined as first-order connectivity only (no second-order connectivity). This seems surprising, since first-order connectivity has been linked to several important properties of vertices, such as essentiality of

Table 1. Number of networks with a distribution that is best described by a power-law (see text).

(For each pair of  $a$  and  $d$  values, 10 networks were simulated. Hence, a table entry of seven means that seven out of 10 networks are best described by a power-law. For growing networks,  $m=2$ ,  $n=1$  and  $t=3000$ . For non-growing networks,  $m=3000$ ,  $n=0$  and  $t=100\ 000$ .)

$d$	$a$									
	1	1.5	2	2.5	3	3.5	4	4.5	5	10
growing networks ( $p=0.01$ )										
0	0	0	10	10	10	10	10	10	10	10
0.0001	0	0	10	10	10	10	10	10	10	10
0.001	0	0	0	8	10	10	10	10	10	10
0.01	0	0	0	0	0	0	0	0	0	0
growing networks ( $p=0.05$ )										
0	0	0	10	10	10	10	10	10	10	10
0.0001	0	0	10	10	10	10	10	10	10	10
0.001	0	0	0	8	10	10	10	10	10	10
0.01	0	0	0	0	0	0	0	0	0	0
non-growing networks ( $p=0.01$ )										
0	0	4	8	10	10	10	10	10	10	10
0.0001	0	0	10	10	10	10	10	10	10	10
0.001	0	0	0	10	10	10	10	10	10	10
0.01	0	0	0	0	0	0	0	0	0	0
non-growing networks ( $p=0.05$ )										
0	0	7	10	10	10	10	10	10	10	10
0.0001	0	9	10	10	10	10	10	10	10	10
0.001	0	0	10	10	10	10	10	10	10	10
0.01	0	0	0	0	0	10	10	10	10	10

proteins (Jeong *et al.* 2001) or influence and status of individuals in social networks. However, it is easy to see how removal based on second-order connectivity leads to *apparent* preferential attachment, resulting in scale-free-like networks. Imagine an extreme case of the model where the network size is doubled by creating new vertices and attaching all of them randomly. Then, all of the freshly attached vertices are removed except one, based on the selective removal criteria. As the second-order connectivity of the remaining new vertex is quite high (or else it had probably been removed), it appears as if it had been attached preferentially to an already well-connected node. Yet, the attachment was perfectly random.

Removal of vertices is certainly a decisive process in the dynamics of networks especially in the absence of network growth, and we showed here that if removal probability correlates inversely with second-order connectivity, *apparent* preferential attachment will lead to scale-free-like network topologies. The model can be applied to dynamic networks where the fitness of a vertex is not only defined by the number of direct connections, but also by the number of connections of the connecting vertices. In social networks, for instance, it is not only important how many people one knows, but also how many people those people know (Katz 1953). This is certainly obvious, for example, in economic or political networks. Also, second-order connectivity is somewhat related to the measures of betweenness and closeness (Wasserman & Faust 1994) whose importance is well known in social networks. Thanks to both the advances in network theory and the increasing amount of network data available, we are beginning to reveal many different processes that may lead to similar network connectivity distributions (Ferrer i Cancho & Solé 2001; Caldarelli *et al.* 2002; Eppstein & Wang 2002; Alava & Dorogovtsev 2005). Hence, a pluralistic view should certainly be adopted if we want to fully understand the emergence of network topologies.

We thank Michael P. H. Stumpf for help with the AIC calculations, and Thomas Pfeiffer, Jacob C. Koella and Roger Koyous for valuable comments.

## APPENDIX A

Here, we obtain an analytic solution for the asymptotic connectivity distribution for the limiting case where  $a$  is very large. In this limit, only vertices with  $c=2$ , i.e. vertices with one first-order and one second-order connection, are deleted. The direct effect of deleting  $r=pN$  vertices with  $c=2$  is to leave  $N_1$  unaltered (where  $N_i$  is the number of vertices with  $i$  direct links), and to lose  $r$  vertices with  $i=2$ . Also, there are no isolated nodes created by such deletions and thus  $u=0$ . We now add  $v=n+r$  new vertices (with one link) to the network at random. The direct effect of this is to add  $v$  vertices with  $i=1$ , but there are indirect effects on existing vertices by virtue of creating new links. Let  $N_i(t)$  be the number of vertices in generation  $t$  with  $i$  links. Then, of the  $v$  new links,  $v_i=vN_i/N$  will attach to nodes with  $i$  links, moving them up to the  $(i+1)$ -link category.

We thus have the simple, linear difference equation

$$N_i(t+1) = N_i(t) + v\delta_{i1} - r\delta_{i2} + \frac{v}{N(t)}\{N_{i-1} - N_i\}, \quad (\text{A } 1)$$

with  $N(t) = m + nt$ . Approximating this difference equation by a differential equation, we obtain

$$\frac{dN_i}{dt} = \frac{v}{m + nt}(N_{i-1} - N_i) + v\delta_{i1} - r\delta_{i2}. \quad (\text{A } 2)$$

We can now show that the asymptotic link distribution is exponential for large  $t$ , with the 1-links being predominant. If we substitute

$$f_i = \frac{N_i}{N},$$

we obtain

$$\dot{f}_i = \frac{v}{N}(f_{i-1} - f_i) + \frac{v}{N}\delta_{i1} - \frac{r}{N}\delta_{i2} - \frac{n}{N}f_i. \quad (\text{A } 3)$$

Substituting  $r=v-n$  we have at equilibrium

$$f_1 = \frac{v}{v+n}, \quad (\text{A } 4a)$$

and for  $(i \geq 2)$

$$f_i = \left(\frac{n}{v}\right)^2 \left(\frac{v}{v+n}\right)^i. \quad (\text{A } 4b)$$

## REFERENCES

- Akaike, H. 1983 Information measures and model selection. In *Proc. 44th Session of the Int. Statistical Institute*, pp. 277–291. Madrid: International Statistical Institute.
- Alava, M. J. & Dorogovtsev, S. N. 2005 Complex networks created by aggregation. *Phys. Rev. E* **71**, 036 107. (doi:10.1103/PhysRevE.71.036107.)
- Barabási, A. L. & Albert, R. 1999 Emergence of scaling in random networks. *Science* **286**, 509–512. (doi:10.1126/science.286.5439.509.)
- Barabási, A. L. & Oltvai, Z. N. 2004 Network biology: understanding the cell's functional organization. *Nat. Rev. Genet.* **5**, 101–115. (doi:10.1038/nrg1272.)
- Berg, J., Lässig, M. & Wagner, A. 2004 Structure and evolution of protein interaction networks: a statistical model for link dynamics and gene duplications. *BMC Evol. Biol.* **4**, 51. (doi:10.1186/1471-2148-4-51.)
- Burnham, K. P. & Anderson, D. R. 2002 *Model selection and multimodal inference: a practical information-theoretic approach*. New York: Springer.
- Caldarelli, G., Capocci, A., De Los Rios, P. & Muñoz, M. A. 2002 Scale-free networks from varying vertex intrinsic fitness. *Phys. Rev. Lett.* **89**, 258 702. (doi:10.1103/PhysRevLett.89.258702.)
- Dorogovtsev, S. N. & Mendes, J. F. F. 2000 Scaling behaviour of decaying networks. *Europhys. Lett.* **52**, 33–39. (doi:10.1209/epl/i2000-00400-0.)
- Dorogovtsev, S. N. & Mendes, J. F. F. 2003 *Evolution of networks*. Oxford: Oxford University Press.
- Eppstein, D. & Wang, J. Y. 2002 A steady state model for graph power laws. *ACM Computing Research Repository*, 2nd Int. Workshop on Web Dynamics, May, 2002.
- Ferrer i Cancho, R. & Solé, R. V. 2001 Optimization in complex networks. Available at <http://arXiv.org/abs/cond-mat/0111222>.

- Jeong, H., Mason, S. P., Barabási, A. L. & Oltvai, Z. N. 2001 Lethality and centrality in protein networks. *Nature* **411**, 41–42. (doi:10.1038/35075138.)
- Katz, L. 1953 A new status index derived from sociometric analysis. *Psychometrika* **18**, 39–43.
- Newman, M. E. J. 2003 The structure and function of complex networks. *SIAM Rev.* **45**, 167–256.
- Simon, H. A. 1955 On a class of skew distribution functions. *Biometrika* **42**, 425–440.
- Wasserman, S. & Faust, K. 1994 *Social network analysis*. Cambridge: Cambridge University Press.
- Watts, D. J. 2004 The ‘new’ science of networks. *Annu. Rev. Sociol.* **30**, 243–270. (doi:10.1146/annurev.soc.30.020404.104342.)