

# Universal Features of Network Topology

K. Austin and G. J. Rodgers

Brunel University, Uxbridge, Middlesex UB8 3PH, U.K.

**Abstract.** Recent studies have revealed characteristic general features in the topology of real-world networks. We investigate the universality of mechanisms that result in the power-law behaviour of many real-world networks, paying particular attention to the Barabasi-Albert process of preferential attachment as the most successful. We introduce a variation on this theme where at each time step either a new vertex and edge is added to the network or a new edge is created between two existing vertices. This process retains a power-law degree distribution, while other variations destroy it. We also introduce alternative models which favour connections to vertices with high degree but by a different mechanism and find that one of the models displays behaviour that is compatible with a power-law degree distribution.

## 1 Introduction

A variety of real-world networks, such as communications networks, the World Wide Web and the Internet consist of many constituents that are connected by diverse interactions. Until recently, there was an absence of data on these large networks, making it difficult to extract reliable results about their topological features. However, as the accumulation of data becomes increasingly computerised, more information about large networks is becoming available. The studies that have been performed on these complex networks so far suggest that they share many common structural characteristics. Thus it has been proposed that there may be universal laws governing the structure and evolution of many diverse real-world networks.

Examples of such networks are the World Wide Web [1–3], the Internet [4], the graph of movie actors, [5], scientific collaboration networks, [6, 7], citation networks [8, 9] and the graph of long distance phone calls [10]. Many of the networks studied are found to exhibit power law degree distributions [11],  $P(k) \sim k^{-\gamma}$ , with  $\gamma$  mostly in the range  $2 \leq \gamma \leq 3$ . These findings have raised the important question of whether there is a universal mechanism by which the networks grow and self-organise that persistently results in the distributions found in real-world networks.

In this paper we introduce a mechanism that is not based on strict preferential attachment to suggest that the preferential attachment process may not be the only mechanism that results in the degree distributions observed in real-world networks.



## 2 Preferential Attachment

The process of preferential attachment was first introduced by Barabasi and Albert [5]. Preferential attachment is the process whereby new vertices and edges are introduced to the network over time in such a way that vertices with more edges increase their degree more quickly. The simplest case of linear preferential attachment is given in Ref. [13], with a vertex of degree  $k$  receiving a new edge with probability  $\Pi(k)$ , given by  $\Pi(k) \propto k$ . At each time step, a new vertex and edge are added to the network and connect to a vertex already present. The probability that a new vertex will connect to an existing vertex of degree  $k$ ,  $\Pi(k)$ , is given by  $\Pi(k) = k/M_1(t)$ . The moments  $M_\alpha(t)$  are given by  $M_\alpha(t) = \sum k^\alpha P_k(t)$ , where  $P_k(t)$  is the number of vertices of degree  $k$  at time  $t$ . The lower order moments,  $M_0(t)$  and  $M_1(t)$ , give the total number of vertices in the network and the total degree of the network, respectively. The result of this model is a power law degree distribution, where the number of vertices of degree  $k$  at time  $t$ ,  $P_k(t)$ , is given by  $P_k(t) \sim tk^{-3}$  [5, 13].

For non-linear preferential attachment, given by  $\Pi(k) \propto k^\alpha$ , where  $\alpha > 0$  is a constant, then it is found that for  $\alpha < 1$ ,  $P_k(t)$  follows a stretched exponential and for  $\alpha > 1$ , a single vertex in the network links to all other vertices [13]. In Ref. [14],  $\Pi(k)$  was measured for four large networks. This was achieved using the computerised data of network dynamics to monitor which vertices acquire new links as a function of the vertex degree. To avoid strong fluctuations, the accumulative distribution,  $\kappa(k) = \int_0^k \Pi(k)dk$ , was used instead of  $\Pi(k)$ . If  $\Pi(k)$  follows non-linear preferential attachment, then  $\kappa(k)$  should have the form  $\kappa(k) \propto k^{\alpha+1}$ . In Ref [14], the Internet and citation networks were both found to follow near linear preferential attachment, with  $\alpha = 1.05$  and  $\alpha = 0.95 \pm 0.1$ , respectively. The scientific collaboration network and the actor network were found to follow sublinear preferential attachment, with  $\alpha = 0.79 \pm 0.1$  and  $\alpha = 0.81 \pm 0.1$ , respectively. This sublinear behaviour implies that  $P(k)$  should follow a stretched exponential [13], but the measured  $P(k)$  indicate that a power law offers a better fit. The authors of Ref. [14] suggest that this is because external links to new vertices follow sub-linear preferential attachment, whereas internal links between existing vertices follow near linear preferential attachment. Since both the actor and science collaboration networks are dominated by internal links, the degree distribution exhibits the scale-free behaviour of linear preferential attachment.

Inspired by the above work, we introduce a model where at each time step a new edge is added with probability  $p$  or an edge and vertex with probability  $1 - p$ . Following the method of [13], the rate equation for  $P_k(t)$  is found to be

$$\frac{dP_k(t)}{dt} = \frac{2p + 1 - p}{M_1(t)} [(k - 1)P_{k-1} - kP_k] + (1 - p)\delta_{k,1}. \quad (1)$$

The first term on the right hand side of this equation represents the creation of a vertex of degree  $k$  from a vertex of degree  $k - 1$ . The second term represents the creation of a vertex of degree  $k + 1$  from a vertex of degree  $k$ . This happens with probability  $2p$  if a new edge is added between existing vertices, and  $1 - p$  if



a new edge and vertex are added and link to an existing vertex. The final term represents the arrival of a new vertex of degree 1, which happens with probability  $1 - p$ . Following [13], we find the total number of edges in the network,  $M_1(t)$ , obeys  $dM_1(t)/dt = 2$ . Substituting  $M_1(t) = 2t$  and  $P_k(t) = tn_k$  into Eq.(1) yields the recursion relation

$$n_k = \frac{(k-1)}{k + [2/(1+p)]} n_{k-1} \quad (2)$$

for  $k > 1$ . For large  $k$ ,  $n_k$  has the solution  $n_k \sim k^{-\gamma}$ , where  $\gamma = 1 + 2/(1+p)$ . Since  $0 \leq p \leq 1$ , then  $2 \leq \gamma \leq 3$  for large  $k$ , which is the range of the exponent found for many real-world networks.

We also consider the case where a new edge is introduced between existing vertices with linear preferential attachment, but the new vertices link to existing vertices with sublinear preferential attachment. The probability that an existing vertex will receive a link to a new vertex is then given by  $\Pi(k) \propto k^\alpha$ , where  $\alpha < 1$ . We find that when  $0 < \alpha < 1$  and  $k \rightarrow \infty$ , the degree distribution is dominated by the process whereby edges are added according to linear preferential attachment. A power-law degree distribution is obtained,  $P_k(t) \sim tk^{-\gamma}$ , with  $\gamma = 1 + 1/p$ , for large  $k$ , even when more edges are added according to the process with sub-linear preferential attachment, i.e. when  $0 \leq p < 0.5$ . However, to yield  $2 \leq \gamma \leq 3$  in agreement with experimental data [11], more edges must be added to the network from the process with linear preferential attachment, i.e. when  $0.5 \leq p \leq 1$ . This work suggests that real-world networks are dominated by the process of linear preferential attachment but that there may be other mechanisms at work.

Many other models based on linear preferential attachment have been proposed to include mechanisms such as the decay of edges, the ‘‘condensation’’ of edges and vertices with fitness [11]. The aim of these models is to maintain the scale-free nature found in real-world networks, whilst incorporating other mechanisms to model the different features that they exhibit.

### 3 Networks with Fitness

The different rates with which vertices acquire edges is often attributed to the ability of vertices to compete for new edges, called the vertex fitness [15]. In Ref. [16], models of growing random networks are studied in which the growth rates are dependent upon the vertex fitness.

#### Model A

In model A, a network is built by connecting new vertices to those already present in the network. The probability that an existing vertex will acquire a new edge is given by  $\Pi(k, \eta) \propto (k-1) + \eta$ , where  $k$  is the vertex degree and  $\eta \geq 0$  is its additive fitness chosen randomly from a probability distribution  $f(\eta)$ . For this model the degree distribution behaves as  $P_k(\eta, t) \sim tk^{-\gamma}$ , where  $\gamma = \langle \eta \rangle + 2$  and  $\langle \eta \rangle$  is the average additive fitness.



## Model B

In model B, the network is built in the same way, but an existing vertex acquires a new edge with random additive fitness,  $\eta$ , random multiplicative fitness,  $\zeta$ , and degree,  $k$ , with probability  $\Pi(k, \eta, \zeta) \propto \zeta(k-1) + \eta$ .  $\eta$  and  $\zeta$  are initially chosen from a probability distribution  $f(\eta, \zeta)$ . After some analysis, it is simple to show that the degree distribution behaves as  $P_k(\eta, \zeta, t) \sim tk^{-\gamma}$  for  $k \rightarrow \infty$ , where  $\gamma = 1 + m/\zeta$ ,  $m$  is the reduced moment, given by  $M_{11}(t) = mt$ , and  $P_k(t) = tn_k$ . For model B, specific cases are considered as the general case cannot be solved explicitly.

The case where  $f(\eta, \zeta) = 1$ ,  $0 \leq \eta \leq 1$ , and  $0 \leq \zeta \leq 1$  is first considered. For this case, the reduced moment is given by  $m = 1/(1 - e^{-2}) = 1.156$ , which yields  $n_k \sim k^{-\gamma}/\ln k$  with  $\gamma = 1 + m = 2.156$ .

For the case where  $f(\eta, \zeta) = \delta(\zeta - \eta)$  with  $0 \leq \zeta \leq 1$ , then the degree distribution follows  $n_k \sim k^{-\gamma}$  with  $\gamma = 2.255$ .

Finally, the case where  $f(\eta, \zeta)$  takes the form  $f(\eta, \zeta) = 6\zeta(1-\zeta)\delta(\zeta - \eta)$  and  $0 \leq \zeta \leq 1$  is solved to give the degree distribution  $n_k \sim k^{-\gamma}/(\ln k)^2$  for  $k \rightarrow \infty$ , where  $\gamma = 1 + m = 2.550$ .

## 4 Extremal Networks

There are many variations of the BA model that preserve the scale-free nature of the model, such as networks where the vertices have random additive fitness [15, 16]. Other variations include processes such as aging and rewiring of edges [11]. However, it is important to note that the BA method of preferential attachment may not be the only mechanism to produce power law degree distributions. Here we present a model which favours connections to vertices with high degree, but by a different mechanism.

In this model, at each time step a set of  $m(t)$  vertices are chosen from the  $N(t)$  vertices already present in the network. The vertices to be in the set are chosen at random so that any vertex can appear in the set more than once. A new vertex is then connected to the vertex in the set with the highest degree. If two or more vertices have equal highest degree, one of them is chosen at random. With these rules, the degree,  $L(t)$ , of the hub vertex, the vertex with the largest degree, obeys

$$\frac{dL(t)}{dt} = 1 - \left[1 - \frac{1}{N(t)}\right]^{m(t)}. \quad (3)$$

The second term on the right hand side of this equation represents the probability of not choosing the vertex when  $m(t)$  vertices are selected. Hence, as the hub vertex always gains another edge when selected, the rate of change of the degree of the hub vertex is one minus this probability. As  $N(t) \sim t$  for large time we have

$$\frac{dL(t)}{dt} \sim 1 - \exp\left[-\frac{m(t)}{t}\right]. \quad (4)$$



**Model A**  $m(t) = m$

When  $m(t)$  is time independent it is simple to use Eq.(4) to show that  $L(t) \sim m \ln t$  for large  $t$ . We consider four different cases of this model in turn (i)  $m = 1$ , (ii)  $m = 2$ , (iii)  $m > 2$  and (iv)  $m \rightarrow \infty$ .

For (i)  $m = 1$  the model is just the random addition of an edge and vertex to a randomly selected vertex already in the network. All vertices are selected with equal probability, independent of degree. Consequently the number of vertices of degree  $k$  at time  $t$ ,  $P_k(t)$ , obeys the equation

$$\frac{dP_k(t)}{dt} = \frac{1}{N(t)} [P_{k-1}(t) - P_k(t)] + \delta_{k,1}. \quad (5)$$

Following [13], Eq.(5) can be solved in the large time limit to yield  $P_k(t) \propto t2^{-k}$ . As one would expect from a network in which the vertices gain edges with a rate independent of their degree, this network has an exponential degree distribution.

For (ii)  $m = 2$  we choose 2 vertices at random and join the incoming edge to the vertex with the larger degree. If they both have the same degree, one of them is chosen at random. The degree distribution  $P_k(t)$  satisfies

$$\frac{dP_k(t)}{dt} = 2 \frac{P_{k-1}(t)}{N^2(t)} \sum_{r=1}^{k-2} P_r(t) - 2 \frac{P_k(t)}{N^2(t)} \sum_{r=1}^{k-1} P_r(t) + \frac{P_{k-1}^2(t)}{N^2(t)} - \frac{P_k^2(t)}{N^2(t)} + \delta_{k,1}. \quad (6)$$

From this equation it is a simple matter to check that  $dN(t)/dt = dM_0(t)/dt = 1$  and  $dM_1(t)/dt = 2$ . Introducing  $P_k(t) = tn_k$  and  $N(t) = t$  for large  $t$  into Eq.(6) gives

$$n_k = 2n_{k-1} \sum_{r=1}^{k-2} n_r - 2n_k \sum_{r=1}^{k-1} n_r + n_{k-1}^2 - n_k^2 + \delta_{k,1}. \quad (7)$$

Fig. 1. is a plot of  $n_k$  against  $k$  obtained from a simulation of this model over  $10^6$  time steps with  $m = 2$ . We see that the data suggests that  $n_k$  has an exponential dependence on  $k$ . We can test this by inserting  $n_k = (1 - e^{-\beta})e^{-\beta k}$  into Eq.(7). For large  $k$  this reveals  $\beta = \ln(3/2) = 0.405\dots$ , which is in good agreement with the simulation data, suggesting that  $\beta = 0.4306\dots$ . A similar value of  $\beta$  can be obtained by iterating Eq.(7) numerically.

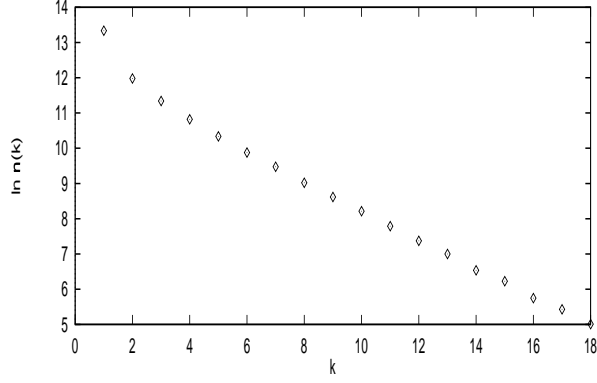
For (iii)  $m > 2$  we can show that the  $n_k$  approximately obey

$$n_k = mn_{k-1} \left[ \sum_{r=1}^{k-2} n_r \right]^m - mn_k \left[ \sum_{r=1}^{k-1} n_r \right]^m + \delta_{k,1} \quad (8)$$

which is obtained by dropping terms of order  $n_k^2$  and higher. Inserting  $n_k = (1 - e^{-\beta})e^{-\beta k}$  into this iteration for large  $k$  yields  $\beta = \ln[(m + 1)/m]$ . Thus for all finite  $m \geq 1$  we have an exponential degree distribution.

As (iv)  $m \rightarrow \infty$  this exponential degree distribution breaks down and  $\beta \rightarrow 0$ . This is because a new edge is always added to the vertex with the highest degree in the network and the network develops a star geometry. More precisely, at time





**Fig. 1.** The degree distribution when  $m = 2$  vertices are chosen at random and the incoming edge is connected to the vertex with highest degree .

$t$  it has a hub vertex with degree  $N(t) - 1$  linked to  $N(t) - 1$  vertices with degree 1. Consequently the degree distribution takes the bi-modal form

$$P_k(t) = \delta(k - N(t) + 1) + (N(t) - 1)\delta(k - 1). \quad (9)$$

**Model B**  $m(t) = \mu N(t)$

In this network model a few hubs develop with very high degree and the vast majority of the the vertices have degree 1. This can be demonstrated by some simple analysis. For instance, putting  $m(t) = \mu N(t) = \mu t$  for large  $t$  into Eq.(4) gives  $L(t) \sim (1 - e^{-\mu})t$ , which indicates that a proportion  $1 - e^{-\mu}$  of the edges in the network are connected to the vertex of the highest degree. The number of vertices with degree 1,  $P_1(t)$ , obeys

$$\frac{dP_1(t)}{dt} = 1 - \left[ \frac{P_1(t)}{t} \right]^{\mu t}. \quad (10)$$

By writing  $P_1(t) = t + p(t)$  and substituting it into Eq.(10) we can solve the resulting differential equation for  $p(t)$  in the large  $t$  limit to yield

$$P_1(t) = t - \frac{1}{\mu} \ln(t + c) \quad (11)$$

where  $c$  is a constant. Consequently the proportion of vertices with degree greater than 1 behaves as

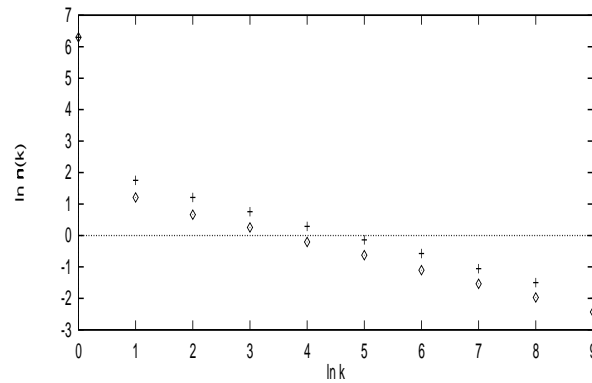
$$\frac{t - P_1(t)}{t} \sim \frac{\ln(t + c)}{t} \sim 0 \quad \text{as} \quad t \rightarrow \infty. \quad (12)$$

Thus we see that for all  $\mu > 0$  this network has  $N_{hub} \sim 1/\mu$  hub vertices with degree  $k_{hub} \sim \mu t$  and a much larger number,  $N_1 \sim t$ , of vertices with degree 1.



**Model C**  $m(t) = N^\nu(t)$

Inserting  $m(t) = N^\nu(t)$  for large  $t$  into Eq.(4) gives  $L(t) \sim t^\nu$ . Clearly for  $\nu \geq 1$  the network develops a geometry like that of Model B above. For  $0 < \nu < 1$  the precise geometry is more difficult to classify analytically. For this reason we have performed numerical simulations of this network for  $0 < \nu < 1$ . We found that the degree distribution  $P_k$  is compatible with a power-law distribution  $P_k \sim k^{-\gamma}$  with for instance  $\gamma = 1.45\dots$  when  $\nu = 3/4$  and  $\gamma = 1.46\dots$  when  $\nu = 2/3$ . This is clearer numerically for larger values of  $\nu$ , because the closer that  $\nu$  is to 1, the easier it is to obtain accurate data for a wide range of vertex degrees. A plot of the log-binned data for  $P_k$  against  $k$  for  $\nu = 3/4$  and  $\nu = 2/3$  after  $5 \times 10^6$  time steps is given in Fig. 2. Plots of  $P_k$  against  $k$  for  $10^6$  and  $2.5 \times 10^6$  time steps give the same exponents. The numerical work, and the work done on Models A and B, suggest that when we have  $m(t) \sim N(t)^\nu$ ,  $P_k \sim k^{-\gamma}$ , with  $\gamma$  a decreasing function of  $\nu$ . However, the fact that for some values of  $\nu$ ,  $\gamma < 2$ , suggests that this behaviour is destroyed for larger systems, and that  $P_k$  may follow some other functional form.



**Fig. 2.** The degree distribution when  $m(t) = N^\nu(t)$  vertices are chosen at random and the incoming edge is connected to the vertex with highest degree; (a)  $\nu = 3/4$  (*diamond*) ; (b)  $\nu = 2/3$  (*cross*).

## 5 Discussion

Many real-world networks appear to exhibit power-law degree distributions, indicating that such networks may be classified according to their scale-free nature. The process of preferential attachment introduced by the BA model, along with its variations, naturally results in these distributions. Thus it may be proposed that there are universal laws governing real-world networks and that the process of preferential attachment may prove to be one.



It seems unlikely, however, that all networks with power-law degree distributions grow according to the same mechanism. Indeed there are many variations of the BA model that preserve the scale-free nature of the model such as networks where the vertices have random additive fitness. Other variations include processes such as aging and rewiring of edges to mimic the behaviour of real world networks. This raises the question of whether preferential attachment is robust enough to incorporate different aspects of real-world networks whilst retaining their degree distributions. We have seen that the introduction of trivial, additive fitness retains power law behaviour, but multiplicative fitness destroys it.

Another important consideration is that what experimental studies indicate as power-law dependence is actually a linear fit to a comparatively narrow range on a log-log plot. It is very difficult for any experimental study to yield a functional form for the degree distribution because of strong fluctuations [11]. We introduced models that were not based on strict preferential attachment and found that one of the models has a degree distribution that is compatible with a power law, but may have a different functional form.

This work reveals important questions about the universality of complex networks: (i) Is preferential attachment robust enough to incorporate all of the diverse processes observed in real-world networks? (ii) Are there other processes that result in the degree distributions found in real-world networks?

## References

1. R. Albert, H. Jeong and A. L. Barabasi, *Nature*, **401**, 130 (1999).
2. R. Kumar, P. Raghavan, S. Rajagopalan and A. Tomkins, *Proceedings of the 25th VLDB Conference, Edinburgh, Scotland, 7-10 September*, 639-650, (1999).
3. A Broder, R. Kumar, F. Maghoui, P. Raghavan, S. Rajagopalan, R. Stata, A Tomkins and J Wiener, *Proceedings of the 9th WWW Conference, Amsterdam, 15-19 May*, 309, (2000).
4. M. Faloutsos, P. Faloutsos and C. Faloutsos, *Comput. Commun. Rev.*, **29**, 251, (1999).
5. A.L. Barabasi and R. Albert, *Science*, **286**, 509 (1999).
6. M. E. J. Newman, *Phys. Rev. E.*, **64**, 016131, (2001).
7. M. E. J. Newman, *Phys. Rev. E.*, **64**, 016132, (2001).
8. J. Laherre and D. Sornette, *Eur. Phys. J.*, **2**, 525, (1998).
9. S. Redner, *Eur. Phys. J.*, **4**, 131, (1998).
10. W. Aiello, F. Chung and L. Lu, *Proceedings of the Thirty-Second Annual ACM Symposium on Theory of Computing*, 171-180, (2000)
11. S. N. Dorogovtsev and J. F. F. Mendes, *Adv. Phys.*, **51**, 1079, (2001).
12. R. Albert, H. Jeong and A. L. Barabasi, *Nature*, **406**, 378, (2000).
13. P. L. Krapivsky, S. Redner and F. Leyvraz, *Phys. Rev. Lett.*, **85**, 4629, (2000).
14. H. Jeong, Z. Neda and A. L. Barabasi, *Europhys. Lett.*, **61**, 567, (2003).
15. G. Bianconi and A. L. Barabasi, *Europhys. Lett.*, **54**, 436, (2001).
16. G. Ergun and G. J. Rodgers, *Physica A*, **303**, 261, (2002).

