arXiv:cond-mat/0607196v3 [cond-mat.stat-mech] 11 Apr 2007

Imperial/TP/06/TSE/1
cond-mat/0607196
Final version 21st July 2006
Eur. Phys. J. B **56**, (2007) 65-69
DOI: 10.1140/epjb/e2007-00084-8
Original publication available at
www.europhysj.org

Exact Solutions for Network Rewiring Models

T.S. Evans*,

Theoretical Physics, Blackett Laboratory, Imperial College London,
South Kensington campus, London, SW7 2AZ, U.K.

Abstract

Evolving networks with a constant number of edges may be modelled using a rewiring process. These models are used to describe many real-world processes including the evolution of cultural artifacts such as family names, the evolution of gene variations, and the popularity of strategies in simple econophysics models such as the minority game. The model is closely related to Urn models used for glasses, quantum gravity and wealth distributions. The full mean field equation for the degree distribution is found and its exact solution and generating solution are given.

Networks with a constant number of edges that evolve only through a rewiring of those edges are of great importance, as exemplified by Watts and Strogatz [1]. Many different applications may be modelled as a network rewiring: the transmission of cultural artifacts such as pottery designs, dog breed and baby name popularity [3, 4, 6, 5], the distribution of family names in constant populations [7], the diversity of genes [8, 9] and the popularity of minority game strategies [10]. There is a close link to some models of the zero range process [11] and the closely related Urn type-models used for glasses [12, 13], simplicial gravity [14] and wealth distributions [15]. The rewiring of networks is also studied in its own right [1, 16, 17].

However previous analytic results for network rewiring models are based on incomplete mean field equations and their approximate solutions. In this letter I give the full equations for linear removal and attachment probabilities with their exact solution. This means the analytic results for rewiring models can match the status of those for random graph and growing network models (e.g. see [2]).

Consider the degree distribution of the artifact vertices[1], $n(k)$, in the bipartite graph of Fig. 1. At each time step I make two choices then alter the network.

*http://www.imperial.ac.uk/people/t.evans

[1]The artifacts may be a dog breed, baby name or pottery style with each individual choosing one type of artifact as indicated by its edge [3, 4, 6, 5]. For family names the individuals are those who inherit

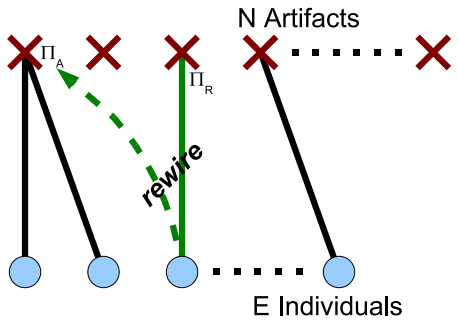


Figure 1: In the abstract form of the model, there are E vertices of one type—'individual' vertices. Each has one edge which runs to one of N vertices of a second type—the 'artifacts'. The degree of the artifact vertices is k indicating that one artifact has been chosen by k distinct individuals. The rewiring will be of the artifact ends of the edges, so each individual always has one edge.

First I choose one of the E individuals at random[2], aiming to rewire the artifact end of the chosen individual's one edge. Thus the edge to be rewired is from an artifact chosen by 'preferential removal'.

Second, this edge will be reattached to one of the N artifact vertices chosen with probability Π_A. With probability p_p preferential attachment is used to choose the artifact. This is equivalent to choosing an individual at random and *copying* the current artifact choice of that individual. Alternatively with probability p_r an artifact is chosen at random to receive the rewired edge. This corresponds to *innovation* in the context of cultural transmission [3, 5], in gene evolution it is *mutation* [8, 9].

With only these types of event, $p_p + p_r = 1$, the number of artifacts N is constant, and

$$\Pi_R = \frac{k}{E}, \qquad \Pi_A = p_r \frac{1}{N} + p_p \frac{k}{E}, \qquad (0 \leq k \leq E). \tag{1}$$

After these choices have been made the rewiring takes place. The mean field equation for the degree distribution for $(0 \leq k \leq E)$ is therefore

$$\begin{aligned} n(k,t+1) - n(k,t) = & \\ & n(k+1,t)\Pi_R(k+1)\,(1-\Pi_A(k+1)) \\ - \;& n(k,t)\Pi_R(k)\,(1-\Pi_A(k)) - n(k,t)\Pi_A(k)\,(1-\Pi_R(k)) \\ + \;& n(k-1,t)\Pi_A(k-1)\,(1-\Pi_R(k-1))\,. \end{aligned} \tag{2}$$

This equation holds at the boundaries $k = 0$ and $k = E$ provided $n(k) = \Pi_R(k) = \Pi_A(k) = 0$ for $k = -1$ and $k = (E+1)$ are chosen. Note that by including the

the name from their partner, the edges are the partners who retain their family name while the artifacts represent different family names. For a model of gene distributions [8, 9] in a haploid population, the artifacts are the alleles while the individuals are the organisms. A diploid population may be modelled in a similar manner. In Urn/Backgammon/Balls-in-Boxes models [12, 13, 14, 15] the individuals represent the balls while the artifacts are the boxes. The rewiring of an undirected network is described by the same equations provided E is replaced by $2E$.

[2]In this letter 'random' without further qualification indicates that a uniform distribution is used to draw from the set under discussion.

factors of $(1-\Pi_A)$ and $(1-\Pi_R)$ on the right hand side I am explicitly excluding events where the same vertex is chosen for removal and attachment in any one rewiring event since such events do not change the distribution. More importantly these terms ensure a rigid upper boundary $n(k>E,t)=0$. Contrast this with the equations for growing networks (for instance in [2]) where there is no rigid upper boundary in the long time limit and such terms are absent . These additional factors are only significant for $k\sim E$ but they are missing in other discussions of rewiring models. The condition $E\gg k$ is usually sufficient for these factors to be negligible and results match the literature in this regime.

These equations (2) have a stationary solution if

$$\begin{aligned} & n(k+1)\Pi_R(k+1)\left(1-\Pi_A(k+1)\right) \\ = \;& n(k)\Pi_A(k)\left(1-\Pi_R(k)\right) \end{aligned} \tag{3}$$

which gives the static solution $n(k)$ for $(E\geq k\geq 0)$ as

$$n(k) \;=\; A\frac{\Gamma(k+\widetilde{K})}{\Gamma(k+1)}\frac{\Gamma(E+\widetilde{E}-\widetilde{K}-k)}{\Gamma(E+1-k)}, \tag{4}$$

$$\widetilde{K}:=\frac{p_r}{p_p}\langle k\rangle,\;\widetilde{E}:=\frac{p_r}{p_p}E\;. \tag{5}$$

The normalisation constant A can be found from (10) below and $\langle k\rangle = E/N$ is the average artifact degree.

For $k\gg 1,\widetilde{K}$, the first ratio of Gamma functions gives

$$\frac{\Gamma(k+\widetilde{K})}{\Gamma(k+1)} \;\propto\; k^{-\gamma}\left(1+O(\frac{1}{k},\frac{\widetilde{K}}{k})\right), \tag{6}$$

where $\gamma=1-\widetilde{K}\leq 1$. This is consistent with previous results which are usually given in a small mutation, $p_r\approx 0$, and/or low average degree $\langle k\rangle\ll 1$ limit.

The novel aspects in the present formulation are the extra factors of $(1-\Pi_A)$ and $(1-\Pi_R)$ in (2). These lead directly to the second ratio of Gamma functions in (4) which for $p_rE\gtrsim 1$ decays exponentially:

$$\frac{\Gamma\left(E+\widetilde{E}-\widetilde{K}-k\right)}{\Gamma(E+1-k)}\propto\exp\{-\zeta k\}\left(1+O(\frac{k}{E})\right)\;, \tag{7}$$

where $\zeta=-\ln(p_p)+O(E^{-1})$.

However, when $p_rE\lesssim 1$ the numerator grows with k. In fact at a critical random attachment probability, p_r^*, the total distribution stops decreasing at the upper boundary, so $n(E)=n(E-1)$. This occurs at

$$p_r^* \;=\; \frac{E-1}{E^2+E(1-\langle k\rangle)-1-\langle k\rangle} \tag{8}$$

Therefore when $p_r<p_r^*\sim 1/E$ the degree distribution will increase near $k=E$.

Thus there are two types of behaviour. For large innovation or mutation, for $1>p_r\gtrsim E^{-1}$ the distribution is approximately $n(k)\propto(k)^{-\gamma}\exp\{-\zeta k\}$, a gamma distribution, with an exact binomial distribution at $p_r=1$, the random graph case of [1]. This gives a power law for small degree, $k\lesssim\ln(p_p^{-1})$, with an exponential

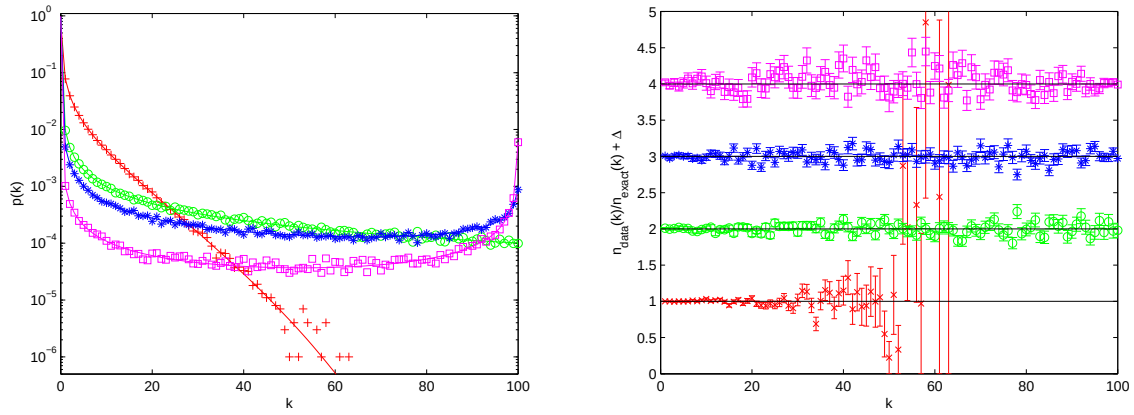


Figure 2: Plots of the degree probability distribution function $p(k) = n(k)/N$ and the fractional error shifted by Δ of the data w.r.t. the exact solution. For $N = E = 100$ and various $p_r = (1 - p_p) = 0.1$ (crosses, $\Delta = 0$), 0.01 (circles, $\Delta = 1$), 0.005 (stars, $\Delta = 2$) and 0.001 (squares, $\Delta = 3$), while lines are the exact solutions. Measured after 10^5 rewiring events, averaged over 10^4 runs. Started with $n(k = 1) = E$ but otherwise $n(k) = 0$. The error bars are mostly smaller or similar in size to the symbol in the first plot.

cutoff for higher degrees. Such behaviour is noted in the literature under various approximations [6, 5, 8, 9, 16] and those results are consistent the exact solution (4). However since $1 \ll \zeta$ implies $(\gamma - 1) \ll \langle k \rangle$, if one had only one data set of a typical size, any power law section of reasonable length ($k \lesssim \zeta^{-1}$) will have a power γ indistinguishable from the value one (c.f. growing networks where $\gamma > 2$).

The second regime is where $p_r E \lesssim 1$, i.e. there is usually no mutation or innovation over a time period when most edges have been rewired once. Here the tail of the distribution rises and one artifact will be linked to almost all of the individuals. It is the *condensation* of [13, 15] and *fixation* in [9] but again those results were given only for the equivalent of large E. Similar behaviour has been discussed for growing networks, for example in [2], but not as an explicit network rewiring problem.

In this simple model, there are no correlations between the degree of vertices. Indeed one need not impose a network structure as in [8, 9, 13, 14, 15]. Thus the mean-field equations should be an excellent approximation to the actual results. Numerical simulations confirm this as Fig. 2 and 3 show.

Given this exact solution for the degree distribution, its generating function

$$G(z) := \sum_{k=0}^{E} n(k) z^k, \tag{9}$$

may be obtained exactly in terms of the hypergeometric function $F(a, b; c; z)$:

$$G(z) \quad = \quad n(0) F(\widetilde{K}, -E; 1 + \widetilde{K} - E - \widetilde{E}; z). \tag{10}$$

The m-th moments of the degree distribution are then

$$\frac{1}{G(1)} \frac{d^m G(z)}{dz^m}\bigg|_{z=1} = \frac{\Gamma(\widetilde{K} + m)\Gamma(-E + m)\Gamma(1 - \widetilde{E} - m)}{\Gamma(\widetilde{K})\Gamma(-E)\Gamma(1 - \widetilde{E})} \tag{11}$$

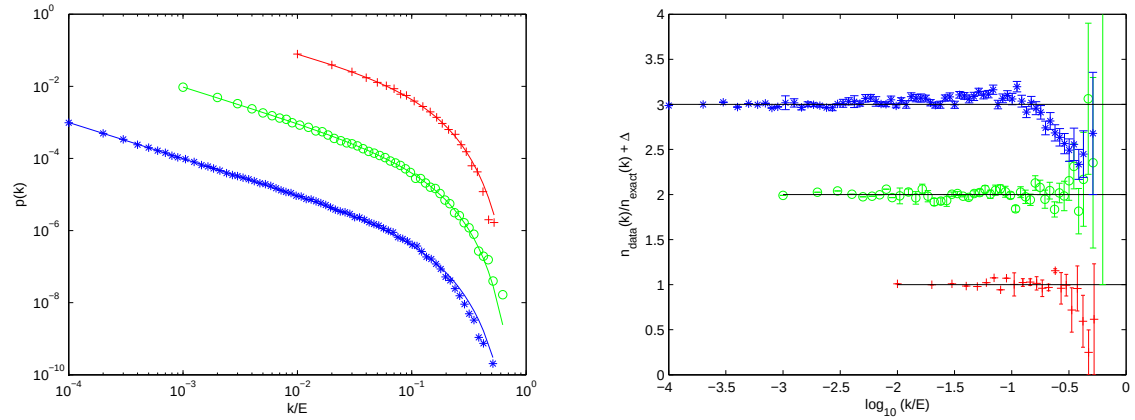


Figure 3: The degree probability distribution function $p(k) = n(k)/N$ and the fractional error shifted by Δ w.r.t. the exact solution for $N = E$, $Ep_r = 10.0$ and $p_r = 10^{-2}$ (crosses, $\Delta = 0$), 10^{-3} (circles, $\Delta = 1$) and 10^{-4} (stars, $\Delta = 2$). Measured after 10^7 rewiring events, averaged over 10^3 runs. Note that for $p_r = 10^{-4}$ there are signs the model may not have quite reached equilibrium. Started with $n(k = 1) = E$ but otherwise $n(k) = 0$.

In particular the case $m = 0$ fixes the normalisation, A, of $n(k)$ in (4), while $m = 1$ confirms the results are completely consistent in the determination of $\langle k \rangle$.

There is another important attachment process that may be included in this model. Suppose that with probability $\bar{p} = 1 - p_r - p_p$ a new artifact vertex is added to the network. The new artifact receives the edge removed from an existing artifact on the same time step. The cultural transmission models [3, 4, 5, 6] and the gene pool model studied in [8, 9] include this process. In the long time limit the number of artifacts becomes infinite, and the random attachment then becomes completely equivalent to this process of new artifact addition. This is the large N, zero $\langle k \rangle$ limit of the discussion above. Care is needed as $n(0)$ diverges and an alternative normalisation is needed. The degree distribution for $k \geq 1$ behaves in exactly the same way as before, a simple inverse degree power law cutoff by an exponential for $E(1 - p_p) \gtrsim 1$ but for $E(1 - p_p) \lesssim 1$ a single artifact is chosen by most individuals. Intriguingly for this model when $E(1 - p_p) = p_p$ the degree distribution an exact inverse power law for the *whole* range of non-zero degrees. The exact solution to the mean field equations again provides an excellent fit to the data as Fig. 4 shows.

These results have several implications. First I have noted that many apparently different models are all equivalent to this simple bipartite network model. Then in terms of mathematical detail, previous mean field equations did not include the $(1-\Pi)$ terms of (2). Thus exact solutions given here are novel. The various forms for the asymptotic behaviour found in the literature can now be seen to be various small p_r and/or large E approximations to the exact results, e.g. descriptions elsewhere of the condensation regime $p_r E \ll 1$ are for large E. Further the calculation of the generating function shows that all aspects of this model appear to be analytically tractable so this rewiring model may prove to be as useful the Erdős-Réyni random graph.

As noted in the introduction, the model also has a wide range of practical

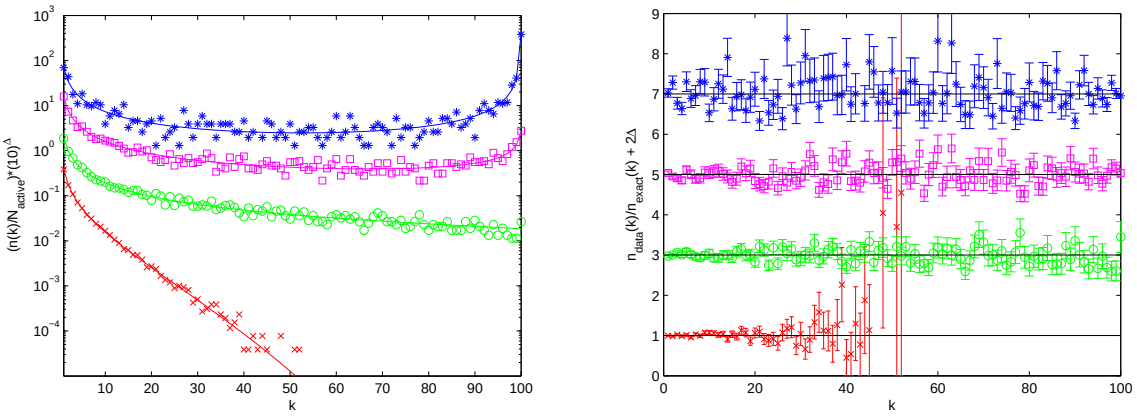


Figure 4: Plots of the degree distribution, normalised by the sum of values for degree greater than zero, and the fractional error of the data w.r.t. the exact solution. For each parameter set plots shifted by a constant controlled by Δ for clarity. For $E = 100$ but with new artifacts added with probability $\bar{p} = 1 - p_p$ ($p_r = 0$) where $\bar{p} = 0.1$ (crosses, $\Delta = 0$), 0.01 (circles, $\Delta = 1$), 0.005 (squares, $\Delta = 2$) and 0.001 (stars, $\Delta = 3$). The lines are the relevant equivalent mean field solutions. Measured after 10^5 rewiring events, and averaged over 10^4 runs. Started with $n(k = 1) = E$ but otherwise $n(k) = 0$. Errors on the degree distribution are not shown.

applications [3, 4, 5, 6, 8, 9, 12, 14, 15]. While it may be too simple in practice, it does at worst give a useful null model against which to test other hypotheses.

However copying the choice of others could also be a genuine strategy, even if it emerges as the result of a more fundamental process. Suppose that the individuals are connected to each other by a second network. When making the choice of artifact for attachment, the individual which is rewiring its edge could consult its acquaintances as represented by this network, and may well choose to follow their recommendation, i.e. copy their artifact. Such random walks on a network, even when of length one, lead naturally to the emergence of preferential attachment in most cases, [18]. This explains results in a model of the Minority Game [10]. There individuals are connected by a random graph and choose a strategy (the artifact) by copying the 'best' of their neighbours. If what is best is continually changing then for the degree distribution this will be statistically equivalent to copying the strategy of a random individual. It is no surprise then that the results for the popularity of strategies in [10] follows a simple inverse power law with a large degree cutoff.

Finally one may consider the scaling properties of the model. In examples such as pottery styles or dog breeds, the categories assigned by investigators are a coarse graining imposed on a collection where each individual is really unique at some level. However one would hope that the results are largely independent of this categorisation. So suppose the artifacts are paired off at random. The decision to copy or to innovate on a given event do not change, so p_r, p_p and $\bar{p}$ remain the same. Because preferential attachment is *linear* in degree, the probability of preferentially attaching to a given pair of artifact vertices is just proportional to the sum of their degrees which in turn is just the degree of the artifact pair vertex. Thus we retain

preferential attachment. The probability of choosing one of a pair of artifacts at random is double choosing just one at random but this reflects that the number of artifact pairs $N_2 = N/2$ is just half the original number of artifact vertices. Overall, the form of the equations for the degree distribution of these pairs, $n_2(k)$, is exactly as before and the only parameters which change are $N \to N/2$ and $\langle k \rangle \to \langle k \rangle /2$. Thus the generic form of the distribution of artifact choice is independent of how artifacts are classified though the detailed prescription changes in a simple manner.

I thank D.Brody, H.Morgan, A.D.K.Plato and W.Swanell for useful conversations.

References

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

[2] S.N.Dorogovtsev and J.F.F.Mendes, Adv.Phys. **51**, 1079 (2002); P.L.Krapivsky and S.Redner, Phys.Rev.E **63**, 066123 (2001).

[3] F.D.Neiman, American Antiquity **60**, 1 (1995).

[4] R.A.Bentley and S.J.Shennan, American Antiquity **68**, (2003) 459; M.W.Hahn and R.A.Bentley, Proc.R.Soc.Lon.B **270**, S120 (2003).

[5] R.A.Bentley, M.W.Hahn and S.J.Shennan, Proc.R.Soc.Lon.B**271**, 1443 (2004).

[6] H.A.Herzog, R.A.Bentley and M.W.Hahn, Proc.R.Soc.Lon B (Suppl.) **271**, S353 (2004).

[7] D.Zanette and S.Manrubia, Physica A **295**, 1 (2001).

[8] M.Kimura and J.F.Crow, Genetics **49**, 725 (1964).

[9] J.F.Crow and M.Kimura, *An Introduction to Population Genetics Theory* (Harper and Row, New York, 1970).

[10] M.Anghel, Z.Toroczkai, K.E.Bassler and G.Korniss, Phys.Rev.Lett. **92**, 058701 (2004).

[11] M.R.Evans and T.Hanney, J.Phys.A. **38**, R195-R240 (2005).

[12] F.Ritort, Phys.Rev.Lett. **75**, 1190 (1995).

[13] P.Bialas, Z.Burda and D.Johnston, Nucl.Phys.B **493**, 505 (1997).

[14] P.Bialas, Z.Burda and D.Johnston, Nucl.Phys.B **542**, 413 (1999).

[15] Z.Burda *et al.*, Phys.Rev.E **65**, 026102 (2002).

[16] K.Park, Y.-C.Lai and N.Ye, Phys.Rev.E **72**, 026131 (2005); Y.-B.Xie, T.Zhou and B-.H.Wang, Report `cond-mat/0512485`.

[17] J.Ohkubo, K.Tanaka and T.Horiguchi, Phys.Rev.E **72**, 036120 (2005); J.Ohkubo, M.Yasuda and K.Tanaka, Phys.Rev.E **72**, 065104R (2005).

[18] A.Vázquez, Phys.Rev.E **67**, 056104 (2003); J.Saramäki and K.Kaski, Physica A **341**, 80 (2004); T.S.Evans and J.P.Saramäki, Phys.Rev.E **72**, 026138 (2005).