A Simple Model of Copying and Innovation

T.S. Evans · A.D.K. Plato · T. You

Abstract Exact analytic solutions and various numerical results for the rewiring of bipartite networks are discussed. An interpretation in terms of copying and innovation processes make this relevant in a wide variety of physical contexts.

Introduction

There are many situations where an 'individual' chooses an 'artifacts' by copying the existing choice of another individual. Names for new babies and registration rates of pedigree dogs often reflect current popular choices [1,2]. The allele for a particular gene carried ('chosen') by an individual reflects current gene frequencies [3]. In Urn models the probabilities controlling the urn chosen by a ball can reflect earlier choices [4]. The individuals in a Voter models [12] copy one of two choices made by a neighbour, as defined by a network of the individuals. Such copying is so common process because it can be implemented without any global information [5]. We will consider the simplest such model of copying and look at how this can be applied and extended.

The Basic Model

We first consider a non-growing bipartite network in which E 'individual' vertices are each attached by a single edge to one of N 'artifact' vertices. At each time step we choose to rewire the artifact end of one edge, the *departure* artifact chosen with probability Π_R. This is attached to an *arrival* artifact chosen with probability Π_A. Only after both choices are made is the graph rewired as shown in Fig. 1. The degree distribution of the artifacts when averaged over many runs of this model, $n(k,t)$,

T.S. Evans
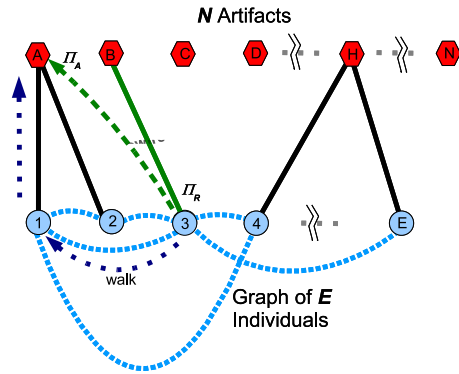Theoretical Physics, Imperial College London, SW7 2AZ, U.K.
Present address: Institute for Mathematical Sciences, Imperial College London, SW7 2PG London, UK

A.D.K.Plato
Institute for Mathematical Sciences, Imperial College London, SW7 2PG London, UK

T.You
Theoretical Physics, Imperial College London, SW7 2AZ, U.K.

Fig. 1 The bipartite network of E individual vertices (circles), each connected by a single edge (solid lines) to any one of N artifacts (hexagons). The dashed lines below the individuals are a social network. In the event shown individual **3** updates their choice, making **B** the departure artifact. They do this by copying the choice of a friend, friend of a friend, etc., found by making a random walk on the social network. Here this produces **A** as the arrival artifact so edge **3B** is rewired to become edge **3A**.

satisfies the following equation:-

$$n(k, t+1) = n(k,t) + n(k+1,t)\Pi_R(k+1,t)\left(1 - \Pi_A(k+1,t)\right)$$
$$-n(k,t)\Pi_R(k,t)\left(1 - \Pi_A(k,t)\right) - n(k,t)\Pi_A(k,t)\left(1 - \Pi_R(k,t)\right)$$
$$+n(k-1,t)\Pi_A(k-1,t)\left(1 - \Pi_R(k-1,t)\right), \quad (E \geq k \geq 0), \qquad (1)$$

where $n(k) = \Pi_R(k) = \Pi_A(k) = 0$ for $k = -1, (E+1)$. If Π_R or Π_A have terms proportional to k^β then this equation is exact only when $\beta = 0$ or 1 [6]. We will use the most general Π_R and Π_A for which (1) is exact, namely

$$\Pi_R = \frac{k}{E}, \qquad \Pi_A = p_r\frac{1}{N} + p_p\frac{k}{E}, \qquad p_p + p_r = 1 \qquad (E \geq k \geq 0).\qquad (2)$$

provided the social network is a complete graph with self-loops.

This form for Π_A (2) contains two processes. The preferential attachment p_p term means an individual copies (inherits) the choice of another. This emerges naturally when using a random walk on a general network [5]. The random choice p_r term can be thought of as innovation or mutation, or may be just be a first approximation for other unspecified. This choice for Π_A has two other special properties: one involves the scaling properties [6] and the second is that these exact equations can be solved analytically [7,6,8,9]. The generating function $G(z,t) = \sum_k z^k n(k,t)$ is decomposed into eigenmodes $G^{(m)}(z)$ through $G(z,t) = \sum_{m=0}^{E} c_m(\lambda_m)^t G^{(m)}(z)$. From (1) we find a second order linear differential equation for each of the eigenmodes with solution [6]

$$G^{(m)}(z) = (1-z)^m {}_2F_1(a+m, b+m; c; z), \quad \lambda_m = 1 - \frac{m(m-1)}{b\,d} - \frac{m(d-b)}{b\,d},$$
$$a = \frac{p_r}{p_p}\frac{E'}{N}, \quad b = -E, \quad c = 1 + a + d, \quad d = b - \frac{p_r E'}{p_p}, \qquad 0 \leq m \leq E, \quad (3)$$

where $E' = E$. These solutions are well known in theoretical population genetics as those of the Moran model [3] and one may map the bipartite model directly onto a simple model of the genetics of a haploid population [6].

The degree distribution is proportional to

$$p(k) \propto \frac{\Gamma(k+a)\,\Gamma(1-c-k)}{\Gamma(k+1)\,\Gamma(1-b-k)} \qquad (4)$$
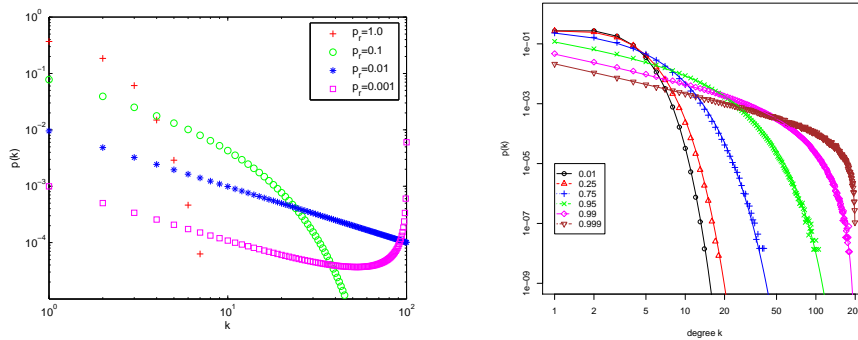
(a) The equilibrium degree probability distribution function $p(k) = n(k)/N$ for $N = E = 100$. Shown are (from top to bottom at low k) $p_r = 1$ (red crosses), $10/E$ (green circles), $1/E$ (blue stars) and $0.1/E$ (magenta squares).

(b) The points are results from the network version of the Tangled Nature model [11]. The lines are the best fits achieved by eye of the analytic equilibrium solution of the basic model (4).

Fig. 2 Equilibrium distributions.

in the equilibrium limit [7,6]. This has three typical regions illustrated in Fig. 2(a). We have a condensate, where most of the edges are attached to one artifact $p(k = E) \sim O(N^0)$, for $p_r \ll (E + 1 - \langle k \rangle)^{-1}$. On the other hand when $p_r \gg (1 + \langle k \rangle)^{-1}$ we get a peak at small k with an exponential fall off, a distribution which becomes an exact binomial at $p_r = 1$. In between we get a power law with an exponential cutoff, $p(k) \propto (k)^{-\gamma} \exp\{-\zeta k\}$ where $\gamma \approx (1 - \frac{p_r}{p_p}\langle k \rangle)$ and $\zeta \approx -\ln(1 - p_r)$.

For many parameter values the power γ will be indistinguishable from one and this is a characteristic signal of an underlying copying mechanism. For instance in the Laird and Jensen network version of the Tangled Nature model [11], the number of links between species is well fitted by the form of the analytic equilibrium solution (4). One may try to estimate these values from first principles. As the network in [11] is a simple graph one must set $-b = (N-1)$, the largest possible degree. In our basic model $a/\langle k \rangle$ is the ratio of edges attached by innovation divided by those added by copying. However the best fit for a to the Tangled Nature network results is roughly double what one would estimate from the ratio of random to inherited (copied) links in [11]. This suggests renormalisation is needed to account for additional fluctuations found in more complicated situations. Similar features can be seen in a simple econophysics model [10].

One of the best ways to study the evolution of the degree distribution [6,8] is through the Homogeneity Measures, $F_n := (\Gamma(E+1-n)/\Gamma(E+1))(d^n G(z,t)/dz^n)_{z=1}$ (rescaled factorial moments). This is the probability that n distinct edges chosen at random are connected to same artifact. Further, each F_n depends only on the modes numbered 0 to n so they provide a practical way to fix the constants c_n in the mode expansion. Since $F_0 = E$ and $F_1 = 1$, we find $c_0 = 1$ and $c_1 = 0$ while equilibration occurs on a time scale of $\tau_2 = -1/\ln(\lambda_2)$ (see Fig. 3).

Communities

Our first generalisation of the basic model is to consider two distinct communities of individuals, say E_x (E_y) of type X (Y). The individuals of type X can now copy the choices made by their own community X with probability p_{pxx}, a different rate
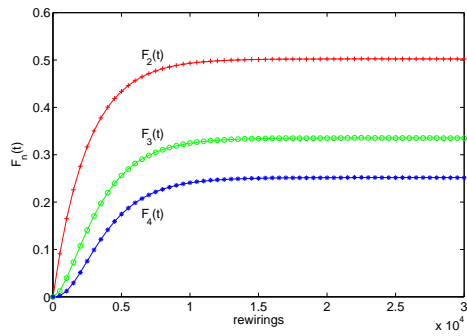
Fig. 3 Plots of various $F_n(t)$ for $E = N = 100$, $p_r = 0.01$. The points are averages over 10^5 runs while the lines are the exact theoretical results. From top to bottom we have: $F_2(t)$ (crosses), $F_3(t)$ (circles), $F_4(t)$ (stars).

which is used when an X copies the choice made by somebody in community Y, p_{pxy}. An X individual will then innovate with probability $(1 - p_{pxx} - p_{pxy})$. Another two independent copying probabilities can be set for the Y community. At each time step we choose to update the choice of a member of community X (Y) community with probability p_x $(1 - p_x)$. Complete solutions are not available but one can find exact solutions for the lowest order Homogeneity measures and eigenvalues using similar techniques to those discussed above. The unilluminating details are given in [8].

Complex Social Networks

An obvious generalisation is to use a complex network as the Individual's social network [8]. When copying, done with probability p_p, an individual does a random walk on the social network to choose another individual and finally to copy their choice of artifact, as shown in Fig. 1. The random walk is an entirely local process, no global knowledge of the social network is needed, so it is likely to be a good approximation of many processes found in the real world. It also produces an attachment probability which is approximately proportional to the degree distribution [5]. The random attachment process, followed with probability p_r, involves global knowledge through its normalisation N in (2). However when $N \gg E$ this can represent innovation of new artifacts as it is likely that the arrival artifact has never been chosen before.
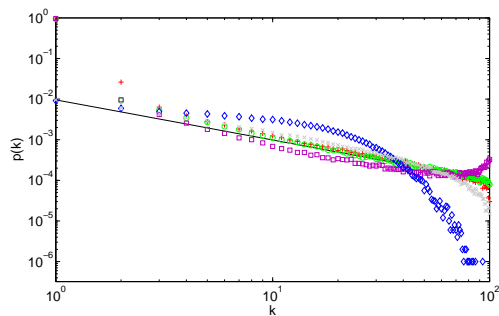
Fig. 4 The degree distributions $p(k)$ averaged over 10^4 runs for different social networks of average degree of 4: Erdős-Réyni (red pluses), Exponential (random with $p(k) \propto \exp(-\zeta k)$, green circles), Scale Free (random with $p(k) \propto k^{-3}$, purple squares), periodic lattices of two (grey crosses) and one (blue diamonds) dimension. The line is the analytic result where the social network is a complete graph with self loops. $N = E = 100$, $p_r = 1/E$.

Results shown in Fig.4 show that the existence of hubs in the Scale Free social network enhances the condensate while large distances in the social networks, as with the lattices, suppress the condensate.

An interesting example is the case of $N = 2$ which is a Voter Model [12] with noise (innovation $p_r \neq 0$) added. One can then compare the probability that a neighbour has a different artifact (the interface density) $\rho(t)$, a local measure of the inhomogeneity, with our global measure $(1 - F_2(t))$. These coincide when the social network is a complete graph. However as we move from 3D to 1D lattices, keeping N, E and p_r constant, we see from Fig. 5 that both these local and global measures move away from the result for the complete graph but in opposite directions [8].
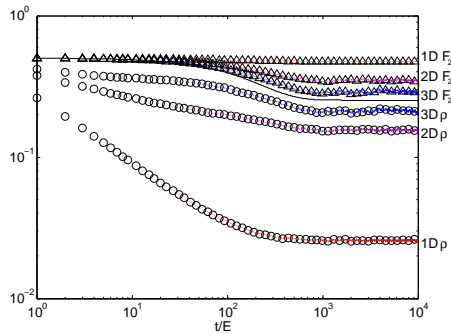
Fig. 5 Inhomogeneity measures for various lattices against t/E. The black solid line represents the analytic result $(1 - F_2(t))$ for $N = 2$, $p_r = 1/E$ and $E = 729$. Numerical results for $(1 - F_2(t))$ (triangles) and for the average probability that a neighbour has a different artifact, $\rho(t)$ (circles) shown for social networks which are lattices of different dimensions. Averaged over 1000 runs.

Different Update Methods

Another way we can change the model is to change the nature of the update. Suppose we first select the edge to be rewired and immediately remove it. Then, based on this network of $E' = (E - 1)$ edges, we choose the arrival artifact with probability $\Pi_A = (p_r/N) + (1 - p_r)k/E'$. The original master equation (1) is still valid and exact. Moreover it can still be solved exactly giving exactly the same form as before, (3), but with $E' = (E - 1)$ not E. This gives very small differences of order $O(E^{-1})$ when compared to the original simultaneous update used initially.

Instead we will consider the simultaneous rewiring of X edges in our bipartite graph at each step. We will choose the individuals, whose edges define the departure artifacts, in one of two ways: either sequentially or at random. The arrival artifacts will be chosen as before using Π_A of (2).

The opposite extreme from the single edge rewiring case we started with $(X = 1)$ is the one where all the edges are rewired at the same time, $X = E$. This is the model used in [1,2,13] to model various data sets on cultural transmission. It is also the classic Fisher-Wright model of population genetics [3]. The evolution equation is then

$$G(z, t + E) = \sum_{k'=0}^{E} [1 + (z - 1)\Pi_A(k')]^E n(k', t). \tag{5}$$

From this each homogeneity measure F_n and the n-th eigenvector λ_n may be calculated in terms of lower order results F_m $(m < n)$. For instance

$$F_2(t) = F_{2\infty} + (\lambda_2)^t (F_2(0) - F_{2\infty}), \quad F_{2\infty} = \frac{p_p^2 + (1 - p_p^2)\langle k \rangle}{p_p^2 + (1 - p_p^2)E}, \quad \lambda_2 = \frac{p_p^2(E - 1)}{E}.$$
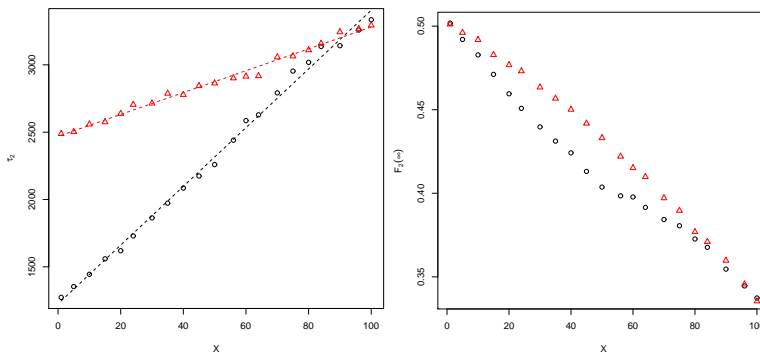
Fig. 6 $\tau_2 = -1/\ln(\lambda_2)$ (left) and $F_2(\infty)$ (right) obtained by fitting $A + B(\lambda_2)^t$ to the data for $F_2(t)$. For sequential ($m = 4$ black circles, lower lines) and random ($m = 6$ red triangles, upper lines) updates of X individuals at a time. $N = E = 100$, $p_r = 1/E = 0.01$ and averaged over 10^4 runs. The dashed lines represent the best linear fit with $\tau_2 \approx 1230(20) + 21.8(3)X$ for $m = 4$ and $\tau_2 \approx 2470(10) + 8.1(2)X$ for $m = 6$. Theoretical values are $\tau_2 \approx 2512.1$ and $F_2(\infty) \approx 0.50251$ for $X = 1$ random update and $\tau_2 \approx 3316.6$ and $F_2(\infty) \approx 0.33669$ for $X = 100$ either update.

Comparing with the results for $X = 1$ we see that there are large differences in the equilibrium solution and in the rate at which this is approached (measured in terms of number of the rewirings made). For intermediate values of X we have not obtained any analytical results so for these numerical simulations are needed, as shown in Fig.6.

Conclusions

We have considered a simple model of bipartite network rewiring. Since copying is a universal process, this model can be mapped onto problems in a wide variety of fields. We've noted that the simplest version can be solved exactly and used these solutions to gain a qualitative understanding of much more complicated models in other areas [10, 11]. We've also seen how the results change as we consider a wide variety of extensions to the basic model.

TSE would like to thank Prof. Jensen for providing the data from [11] for Fig.2(b).

References

1. M.W. Hahn and R.A. Bentley, Proc.R.Soc.Lon. B **270** (2003) s120.
2. H.A. Herzog, R.A. Bentley, and M.W. Hahn, Proc.R.Soc.Lon B (Suppl.) **271** (2004) s353.
3. W.J. Ewens, *Mathematical population genetics: I. theoretical introduction*, 2nd ed., Springer-Verlag New York Inc., 2004.
4. C. Godreche and J.M. Luck, J. of Phys.Cond.Matter **14** (2002) 1601.
5. T.S. Evans and J.P. Saramäki, Phys.Rev. E **72** (2005) 026138.
6. T.S. Evans and A.D.K. Plato, Phys.Rev. E **75** (2007) 056101.
7. T.S. Evans, Eur. Phys. J. B **56** (2007) 65.
8. T.S. Evans and A.D.K. Plato, Networks and Heterogeneous Media **3** (2008) 221.
9. T.S. Evans, *Randomness and complexity in networks*, `arXiv:0711.0603`.
10. M. Anghel et al., Phys.Rev.Lett. **92** (2003) 058701.
11. S. Laird and H.J. Jensen, Europhysics Letters **76** (2006) 710.
12. T.M. Liggett, *Stochastic interacting systems: Contact, voter and exclusion processes*, Springer-Verlag, New York, 1999.
13. R.A. Bentley et al., Evolution and Human Behavior **28** (2007) 151.