Are Copying and Innovation Enough?

T.S. Evans[1,2], A.D.K. Plato[2], and T.You[1]

[1] Theoretical Physics, Imperial College London, London SW7 2AZ, UK
T.Evans@imperial.ac.uk, tevong@gmail.com
[2] Institute for Mathematical Sciences, Imperial College London, London SW7 2PG, UK, Alexander.Plato@imperial.ac.uk

Summary. 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.

1 Introduction

There are many situations where an 'individual' chooses only one of many 'artifacts' but where their choice depends in part on the current choices of the community. Names for new babies and registration rates of pedigree dogs often reflect current popular choices [10, 11]. The allele for a particular gene carried ('chosen') by an individual reflects current gene frequencies [8]. In Urn models the probabilities controlling the urn chosen by a ball can reflect earlier choices [9]. In all cases copying the state of a neighbour, as defined by a network of the individuals, is a common process because it can be implemented without any global information [7]. At the other extreme, an individual might pick an artifact at random.

2 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, $n(k,t)$, satisfies the following equation:-

$$\begin{aligned} n(k,t+1) = n(k,t) &+ n(k+1,t)\Pi_R(k+1,t)\,(1-\Pi_A(k+1,t)) \\ &- n(k,t)\Pi_R(k,t)\,(1-\Pi_A(k,t)) - n(k,t)\Pi_A(k,t)\,(1-\Pi_R(k,t)) \\ &+ n(k-1,t)\Pi_A(k-1,t)\,(1-\Pi_R(k-1,t))\,, \quad (E \geq k \geq 0)\,, \end{aligned} \quad (1)$$

A.D. Fitt et al. (eds.), *Progress in Industrial Mathematics at ECMI 2008*,
Mathematics in Industry 15, DOI 10.1007/978-3-642-12110-4_132,

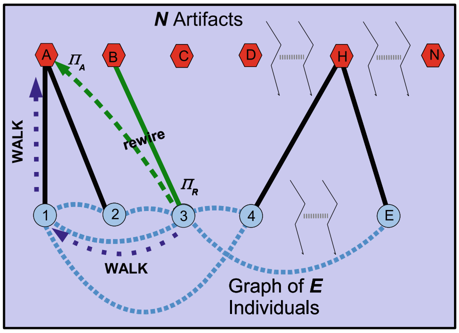

Fig. 1. The bipartite network of E individual vertices, each connected by a single edge (*solid lines*) to any one of N artifacts. The *dashed lines* below the individuals are a social network. In the event shown individual **3** updates their choice, making **B** the departure artifact. They do this 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**

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 [5]. We will use the most general Π_R and Π_A for which (1) is exact, namely

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

This is equivalent to using a complete graph with self loops for the social network at this stage but these preferential attachment forms emerge naturally when using a random walk on a general network [7]. This choice for Π_A has two other special properties: one involves the scaling properties [5] and the second is that these exact equations can be solved analytically [3–6]. 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 [5]

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

where $E' = E$. These solutions are well known in theoretical population genetics as those of the Moran model [8] and one may map the bipartite model directly onto a simple model of the genetics of a haploid population [5].

The equilibrium result for the degree distribution [3, 5] is proportional to $\frac{\Gamma(k+a)}{\Gamma(k+1)} \frac{\Gamma(E+a(N-1)-1-k)}{\Gamma(E+1-k)}$. This has three typical regions. 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

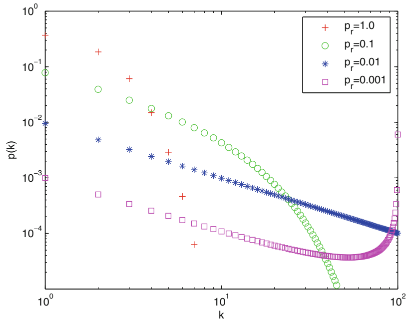

Fig. 2. The equilibrium degree probability distribution function $p(k) = n(k)/N$ for $N = E = 100$. Shown are (from top to bottom at low k) $p_r = 1$ (*crosses*), $10/E$ (*circles*), $1/E$ (*stars*) and $0.1/E$ (*squares*)

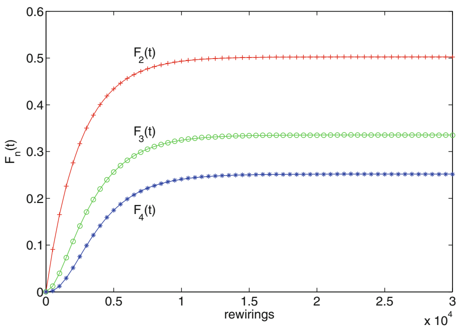

Fig. 3. Plots of various $F_n(t)$ for $E = N = 100$, $p_r = 0.01$. The *points* are averages over 10^5 runs while the *lines* are the exact theoretical results. From top to bottom we have: $F_2(t)$ (*crosses*), $F_3(t)$ (*circles*), $F_4(t)$ (*stars*)

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 seen in a diverse range of situations (e.g. see [1, 12]; Fig. 2).

One of the best ways to study the evolution of the degree distribution [5, 6] is through the *Homogeneity Measures*, F_n. This is the probability that n distinct edges chosen at random are connected to same artifact, and is given by $F_n(t) := (\Gamma(E + 1 - n)/\Gamma(E + 1))(d^n G(z, t)/dz^n)_{z=1}$. 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 $\tau_2 = -1/\ln(\lambda_2)$ (Fig. 3).

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 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 [6].

4 Complex Social Networks

An obvious generalisation is to use a complex network as the Individual's social network [6]. 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, to a good approximation, proportional to the degree distribution [7]. The alternative process of innovation, 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. However this process could also be a first approximation for other unknown processes used for artifact choice.

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 [13] 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 [6].

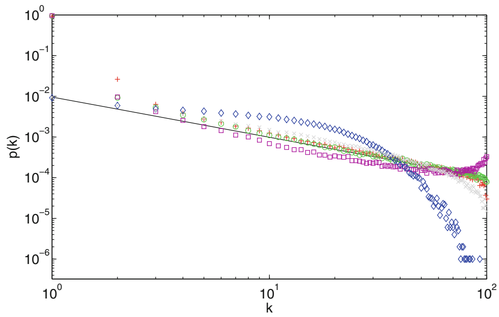

Fig. 4. The degree distributions $p(k)$ averaged over 10^4 runs for different social networks of average degree of 4: Erdős-Réyni (*pluses*), Exponential (random with $p(k) \propto \exp(-\zeta k)$, *circles*), Scale Free (random with $p(k) \propto k^{-3}$, *squares*), periodic lattices of two (*grey crosses*) and one (*diamonds*) dimension. The line is the analytic result where the social network is a complete graph with self loops. $N = E = 100$, $p_r = 1/E$

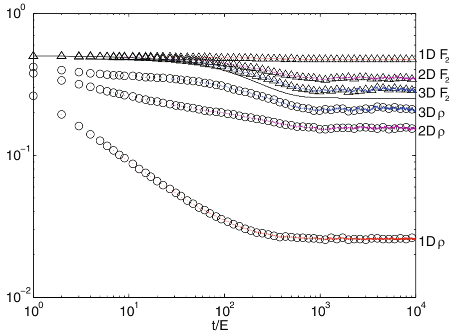

Fig. 5. Inhomogeneity measures for various lattices against t/E. The *black solid line* represents the analytic result $(1 - F_2(t))$ for $N = 2$, $p_r = 1/E$ and $E = 729$. Numerical results for $(1 - F_2(t))$ (*triangles*) and for the average probability that a neighbour has a different artifact, $\rho(t)$ (*circles*) shown for social networks which are lattices of different dimensions. Averaged over 1,000 runs

5 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.

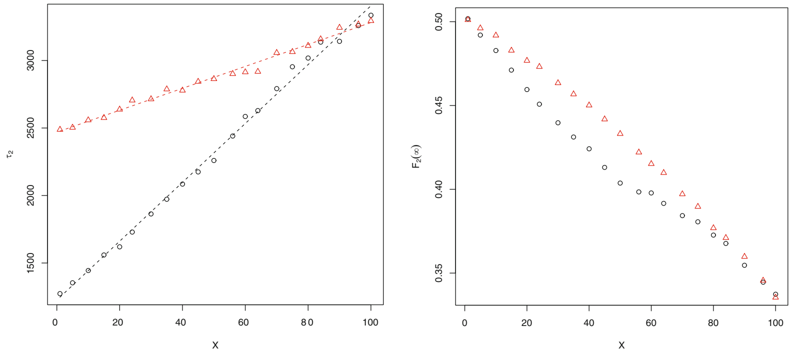

Fig. 6. $\tau_2 = -1/\ln(\lambda_2)$ (*left*) and $F_2(\infty)$ (*right*) obtained by fitting $A + B(\lambda_2)^t$ to the data for $F_2(t)$. For sequential ($m = 4$ circles, *lower lines*) and random ($m = 6$ triangles, *upper lines*) updates of X individuals at a time. $N = E = 100$, $p_r = 1/E = 0.01$ and averaged over 10^4 runs. The *dashed lines* represent the best linear fit with $\tau_2 \approx 1{,}230(20) + 21.8(3)X$ for $m = 4$ and $\tau_2 \approx 2{,}470(10) + 8.1(2)X$ for $m = 6$. Theoretical values are $\tau_2 \approx 2{,}512.1$ and $F_2(\infty) \approx 0.50251$ for $X = 1$ random update and $\tau_2 \approx 3{,}316.6$ and $F_2(\infty) \approx 0.33669$ for $X = 100$ either update

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 [2,10,11] to model various data sets on cultural transmission. It is also the classic Fisher-Wright model of population genetics [8]. 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 ($m < n$). Non trivial information again comes first from $F_2(t) = F_2(\infty) + (\lambda_2)^t\,(F_2(0) - F_2(\infty))$ where

$$F_2(\infty) = \frac{p_p^2 + (1-p_p^2)\langle k\rangle}{p_p^2 + (1-p_p^2)E}\,, \qquad \lambda_2 = \frac{p_p^2(E-1)}{E}\,. \tag{4}$$

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.

References

1. Anghel, M., et al.: Phys. Rev. Lett. **92**, 058701 (2003)
2. Bentley, R.A., et al.: Evol. Hum. Behav. **28**, 151 (2007)
3. Evans, T.S.: Eur. Phys. J. B **56**, 65 (2007)
4. Evans, T.S.: arXiv:0711.0603
5. Evans, T.S., Plato, A.D.K.: Phys. Rev. E **75**, 056101 (2007)
6. Evans, T.S., Plato, A.D.K.: Netw. Heterogen. Media **3**, 221 (2008)
7. Evans, T.S., Saramäki, J.P.: Phys. Rev. E **72**, 026138 (2005)
8. Ewens, W.J.: Mathematical population genetics. I. Theoretical introduction, 2nd edn. Springer, New York (2004)
9. Godreche, C., Luck, J.M.: J. Phys.: Condens. Matter **14**, 1601 (2002)
10. Hahn, M.W., Bentley, R.A.: Proc. R. Soc. Lond. B **270**(Suppl. 1), S120 (2003)
11. Herzog, H.A., Bentley, R.A., Hahn, M.W.: Proc. R. Soc. Lond. B **271**(Suppl.), S353 (2004)
12. Laird, S., Jensen, H.J.: Europhys. Lett. **76**, 710 (2006)
13. Liggett, T.M.: Stochastic Interacting Systems: Contact, Voter and Exclusion Processes. Springer, New York (1999)