Ongoing Evolution on Networks

Marcus Frean
Victoria University of Wellington, New Zealand
marcus.frean@vuw.ac.nz
Gareth Baxter
Departamento de Física, Universidade de Aveiro, Portugal
Paul Rainey
NZ Institute for Advanced Study, New Zealand

In an evolving population, network structure can have striking effects on the survival probability of a mutant allele and on the rate at which it spreads. In networks with 'hubs' (representing geographic or other constraints), the heightened probability of an initially rare mutant has led to the prediction that such networks act to amplify the effects of selection over drift. But selection and mutation interplay in a subtle way in such populations: hubs also slow the mutant's rate of invasion, so that if multiple mutants are allowed to spread at the same time, more of them may be present. In other words it might be misleading to consider only the fixation probability, because new mutants spread at different rates in these networks.

Instead of following a single mutation to fixation, we give a very simple model that allows for a stream of mutations, leading to a dynamic equilibrium. In this way we take account of ongoing evolution rather than simply following a single mutant to fixation. In this model the amplifier effect is largely reversed: for much of the parameter space, networks with hubs suppress rather than amplify selection. This is explained by considering the relative time scales of mutation and selection.

1 Introduction

Evolution is the interplay between the sharpening force of selection and the diffusing force of variation. Both are required in order for evolution to proceed. When selection pressure is large but there is little or no variation introduced,

we might expect a relatively homogeneous population of highly fit individuals. On the other hand, when a lot of variation occurs (e.g. through mutation) but selection pressure is weak, we might expect a very diverse population. In this paper we first propose a simplified model of these two processes. Then we use it to examine the interplay between selection and drift in structured populations which can be thought of as forming a graph or network.

We consider extensions of Moran's simple single-locus panmictic competition model [8]. Each of a set of N sites is occupied by an individual, represented by a single allele. Each allele has a relative intrinsic fitness, taking positive integer values. At each step the occupant of one site is replaced by a duplicate of that at another site. In Moran's original model the invading site (parent) is chosen with a probability proportional to a value representing relative fitness, and the invaded site is chosen uniformly at random from the remainder of the population.

Without the introduction of variation, this process ends with a homogeneous population (all having the same allele). We therefore introduce random mutations to the population at a steady rate. There are then two rates of interest: the rate of mutation and the rate of invasion, but since no other events are being modelled here it suffices to specify the ratio between these two: we specify how likely the next event is to be a mutation (probability β) versus an invasion (probability $1-\beta$). Mutation is represented by replacing the allele at a randomly (uniformly) chosen site with an allele of relative fitness either higher or lower by 1, with probabilities α and $1-\alpha$ respectively. Since deleterious mutations are more likely than advantageous ones, $\alpha < \frac{1}{2}$. We use the example of $\alpha = \frac{1}{3}$ in what follows, meaning a mutant is twice as likely to be deleterious as to be advantageous. Mutations that would reduce fitness below 1 are ignored. Varying the parameter β allows us to control the balance between genetic drift ($\beta \to 1$) and selection ($\beta \to 0$). Note that β is not an absolute mutation rate, rather $\beta/(1-\beta)$ is the relative rate of mutation compared with invasion. This model is illustrated in Figure 1, and typical results shown in Figure 2 for a population size of $N = 40$.

We introduce structure/subdivision in the population as follows: the sites are arranged on an (arbitrary) graph, with invasions limited to neighbours in the graph; in other words, the invader can only invade sites to which it is connected. For our current purposes it is sufficient to consider only unweighted graphs, that is, the invaded/invading site is chosen uniformly from the neighbourhood.

When starting from low fitness values, the mean fitness in the population initially rises, as progressively fitter mutants appear and become dominant in the population. This increase does not continue indefinitely, as the relative benefit from an advantageous mutation decreases until an equilibrium between selection and mutation is reached. At equilibrium, the fitnesses present in the population are typically narrowly distributed around a population mean, which in turn explores a broader *ensemble* distribution (Figure 2). This equilibrium distribution is defined by the population size, the relative mutation rate parameter β and the network structure of the population. The highest mean is attained as β tends to zero, when the diversity is also lowest, as the population is entirely homogeneous

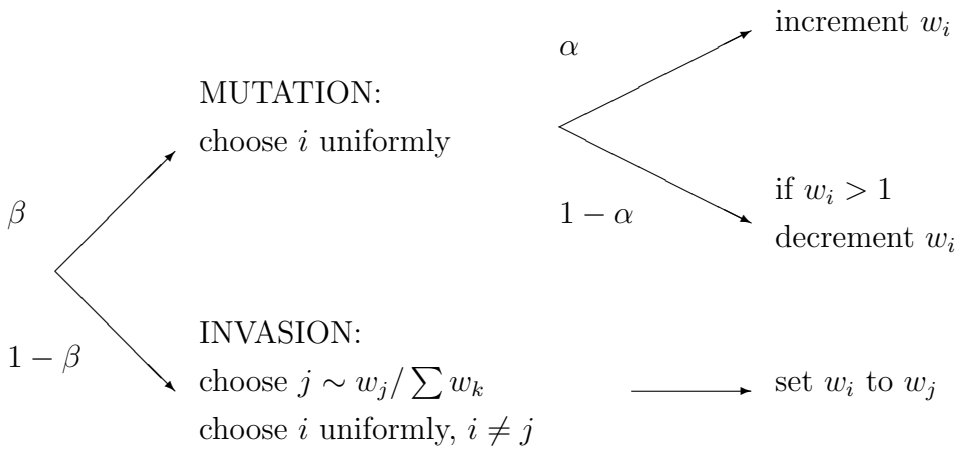

Figure 1: A simple model for evolutionary events relating to fitness: invasions (*i.e.* selection with competition) tend to drive fitnesses up, while mutations tend to drive it down, but both processes are inherently stochastic. β and α are constants: β determines the mutation rate (relative to the invasion rate), while $\alpha < \frac{1}{2}$ represents the damage (on average) to be expected from purely random mutation events. In all our simulations we used $\alpha = 1/3$. w_i is the current fitness value at site i.

for the vast majority of the time. The lower bound is when there is no invasion, i.e. pure drift. For panmictic populations, the mean equilibrium fitness grows approximately linearly with N (this would eventually plateau, but at an N that is much larger than the values of interest here).

In highly heterogeneous structures – we use the STAR graph (Figure 4b) as an exemplar – the effect of fitness is enhanced for very small mutation rates and low populations. But for for larger mutation rates or higher populations, there is suppression – the average equilibrium fitness is *lower* than in the panmictic case. For a given β we can find a population size N above which there ceases to be any enhancement, and vice versa. The bound on β becomes extremely small for large populations, hence heterogeneous structures are likely to be fitness-suppressing in large populations.

As has already been demonstrated in the one-off mutation case [2, 1] reversing the order of the invasion steps has little effect in panmictic populations, but profoundly affects the selection process in structured populations. We find that in the D-B version of our model (see below), the population behaves almost completely neutrally, for all population sizes and mutation rates. More detailed descriptions and derivations of these results are given in Section 2.

In the absence of mutation, the population (if connected) eventually fixes to a homogeneous state, and the dynamics of this process have been studied extensively, both in the panmictic case [8, 9, 11] and more recently on graphs [16, 6, 12, 2]. The more complicated case of neighbour-dependent fitness ("games"

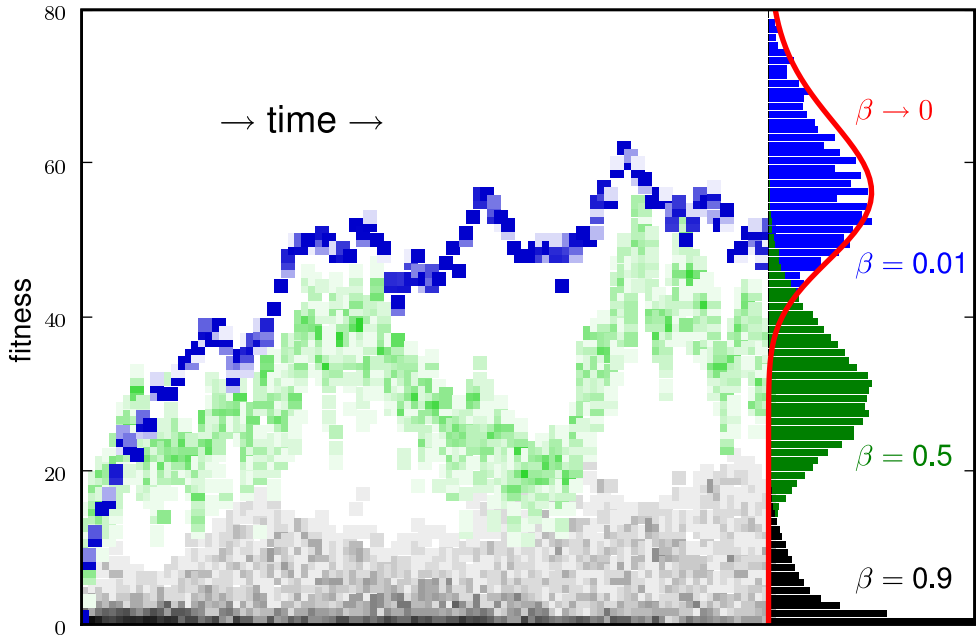

Figure 2: Fitnesses in a freely mixing population under the ongoing evolution model. The plot shows the development of fitness values in a population of 40, starting with all individuals having fitness 1. The temporal scale covers $\sim 10^6$ invasions for the $\beta = 0.01$ case, 10^5 for $\beta = 0.5$ and 10^3 for $\beta = 0.01$. The resulting equilibrium distributions are shown on the right, from data accumulated over much (over a thousandfold) longer runs. The red curve shows the theoretical limit as described in the text.

modelling cooperation for example) has also been studied [13, 14, 18, 3, 4, 17].

In a panmictic population with continuing mutation an equilibrium is reached in which a diversity of alleles is maintained. This equilibrium between mutation and selection has of course been well recognised, and extensively studied (in diploid populations), both in neutral evolution and considering the effects of selection. See for example [5, 19, 7]. In the current study we address the question of what happens when these processes take place in a structured population.

2 Analysis

2.1 Dynamics on a fully connected graph

We consider the model variant in which the birth site is chosen first and the population is panmictic (i.e. the graph is fully connected - henceforth abbreviated to just PAN). At arbitrarily low mutation rates ($\beta \to 0$) the population will always reach fixation before a new mutant is introduced. The fixation probability can be shown [10] to be

$$P_{\text{fix}}(f') = \frac{1 - f/f'}{1 - (f/f')^N} \tag{1}$$

where f' is the fitness of a lone new mutant in a population of others that have fitness f. For example, if a single mutant with fitness 2 arises in a population with fitness 1, it has about half a chance of eventually becoming the dominant type. Notably this fixation probability is only weakly dependent (through the denominator) on population size N: the first few invasions are largely independent of N, and once a fitter allele has overtaken a significant number of sites, it goes on to take over the whole population with a very high probability, regardless of the size of N.

If we repeat this process – introducing a single mutant and then waiting for fixation before introducing another mutant and so on – the mean fitness of the population does not rise indefinitely, but fluctuates according to a finite equilibrium distribution. From equation (1) the probability $P_{f\to f+1}$ that an initial population with uniform fitness f transitions to one with uniform fitness $f+1$ is proportional to $\alpha P_{\text{fix}}(f+1)$, and similarly $P_{f\to f-1}$ is proportional to $(1-\alpha)P_{\text{fix}}(f-1)$. All other transition probabilities are zero apart from the diagonal $P_{f\to f}$ required for normalisation. Within the obvious bounds the proportionality affects only the mixing rate of the associated Markov chain, not its equilibrium distribution, which is the principal eigenvector of the matrix $P_{f\to f'}$. (This gives the curve shown in red in Figure 2.) The mean can be found by solving $P_{f\to f+1} = P_{f\to f-1}$ for f, which gives $\bar{f}^{\text{eqlb}} = 1.44 \times N$.

At very high rates of mutation ($\beta \to 1$) we can ignore invasions and any population structure. In that case the fitnesses have a geometric distribution and the mean fitness is $(1-\alpha)/(1-2\alpha)$, utilising the fact that $1+\gamma+\gamma^2+... = \frac{1}{1-\gamma}$ and that $1+2\gamma+3\gamma^2+... = [1+\gamma+\gamma^2+...]^2$. In our example of $\alpha = 1/3$, the densities of fitnesses $(1, 2, 3, \ldots)$ are $\frac{1}{2}$, $\frac{1}{4}$, $\frac{1}{8}$ and so on, giving a mean fitness $\bar{f}^{\text{eqlb}} = 2$, irrespective of population size.

In general, the mean fitness lies between the extreme values of 2 and $1.44N$. We carried out numerous long-run simulations to find the mean fitness at equilibrium for different mutation rates β and population sizes N (see Figure 3). For small β, the mean fitness rises linearly with N, while for large β it is effectively independent of N. For a given intermediate value of β, the behaviour is a mixture of these: mean fitness rises at first linearly with N, but then plateauxs (consider a horizontal trajectory in Figure 3). This can be understood by considering the time to reach fixation in the limit $\beta \to 0$ which grows linearly with

N – in other words the dynamics of selection (loss of diversity) become slower as N increases. For fixed β, the mean interval between mutations is constant, and supposing it is at first much longer than the typical fixation time, it eventually becomes much *shorter* than the mean fixation time. Qualitatively, there is a transition from the dynamics being dominated by selection to one where the dynamics are dominated by mutation rate, after which the mean fitness cannot grow any further. This picture will be important when we consider the effect of network structure in the next section.

As the interval between the introduction of mutations increases to infinity, naturally the variance of the fitnesses in the population tends to zero, as only one allele is present for the vast majority of the time. As can be seen in Figure 2, for $\beta = 0.01$ the population is tightly focussed. Near the mean fitness value, there is little upward or downward pressure, so the population as a whole drifts considerably. For low values of β the ensemble distribution is approximately Gaussian as illustrated in Figure 2 (although not shown here, the variance $\approx 2N$ and is relatively insensitive to β, as can be seen in the figure). The actual distribution of fitnesses in the population at any given time however is much narrower than this, tending to be smaller for smaller β values. Thus for very low mutation rates, the model leads to a population whose mean fitness scales with the population size, but which fluctuates widely over time.

At the other extreme of high mutation rates the population variance coincides with the ensemble variance. Overall, it varies continuously between the upper bound of the ensemble variance $\frac{2\alpha(2-\alpha)}{(1-2\alpha)^2}$, which is reached only as $\beta \to 1$, and the lower bound of zero, which is reached in the $\beta \to 0$ limit.

2.2 effect of network structure

As has been noted in numerous earlier studies [16, 6, 1], network structure can have a profound effect on the dynamics of an evolutionary model. To illustrate the effect of network structure, we consider the network in which one site is connected to all the others but none of those are connected to one another (see Figure 4). We chose this because it reflects a hub-with-spokes structure that is a common motif in biological and other networks, is highly degree-heterogeneous, and has been much studied by previous authors. Leiberman et al. [16, 6] found the fixation probability of an initial rare mutant on such a STAR graph to be

$$P_{\text{fix}}(f') = \frac{1-(f/f')^2}{1-(f/f')^{2N}} \ . \tag{2}$$

Hence star graphs are "fitness amplifiers" in the sense that their fixation probability matches the panmictic case with each fitness squared. This result corresponds to $\beta \to 0$, since a single initial mutation is assumed after which the system is left to run to fixation. However fixation occurs much more slowly in the star graph than in a fully connected population, and one might wonder what the effect of subsequent mutations could be. Our model addresses this case. As

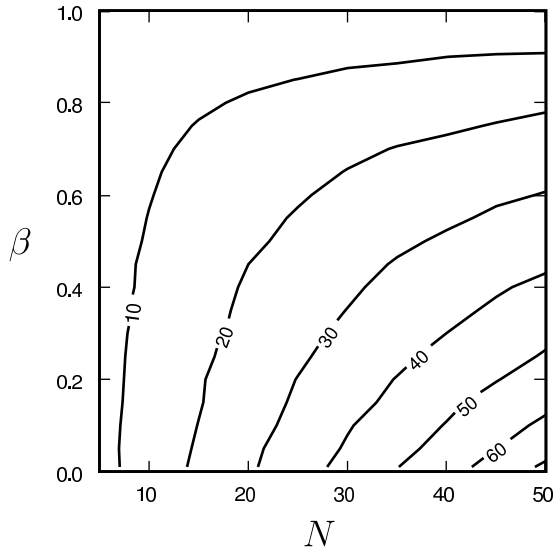

Figure 3: A contour plot of the mean equilibrium fitness $\bar{f}_{PAN}$ found in simulations of a panmictic population, as a function of the relative rate of mutation β and the population size N.

we did for the panmictic case, we carried out numerous long-run simulations for different mutation rates β and population sizes N.

Figure 5 shows the average fitness achieved by ongoing evolution on the STAR network, compared to that in a panmictic population. For small β and N, we see that the fitness is indeed enhanced.

However the time taken to reach fixation increases with population size N [15], and much more rapidly on STAR graphs than PAN, as shown in Figure 6, meaning that a new mutation is more likely to occur before fixation of the previous one. The relative tendency towards drift is thus increased, as the effect of selection is diluted. For sufficiently large N, an equilibrium is reached, so that mean fitness no longer increases with N. Similarly, with increasing β for a fixed population size, again the relative effects of drift and selection are altered, in favour of drift, and again the effect is much more pronounced on the STAR graph. The overall effect is therefore that, while selection is enhanced on a STAR graph for small N and β, for the majority of the parameter space, this graph suppresses the effect of selection.

This is further illustrated in the left graph of Figure 7. We see that $\bar{f}$ grows linearly with N for small β in the PAN network. For the STAR network, equilibrium fitness is enhanced for small N, but for large N it starts to tail off, until finally the fully connected network has a larger value of mean fitness. Similarly for a constant N, shown on the right in Figure 7, the STAR graph is

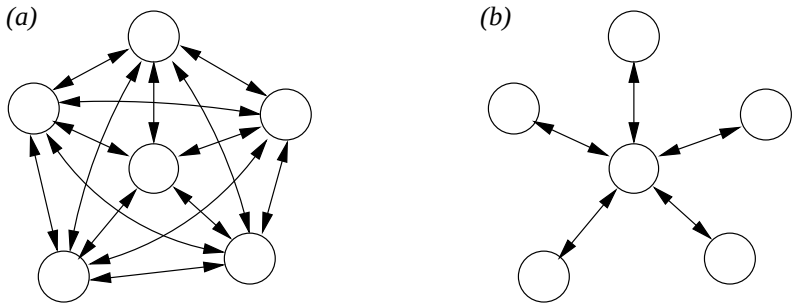

Figure 4: *(a)* A fully connected network (PAN). *(b)* The hub-and-spokes network (STAR). In both these networks all the nodes are susceptible to invasion, since they have reciprocal links. Trivially, if we were to allow one node to have no incoming links it could not be invaded and so ρ would collapse to $\leq 1/N$. We don't consider such networks here.

fitness-enhancing for small β values, but larger mutation rates soon overwhelm the slow selection dynamics, so that the STAR graph is fitness-suppressing for most β values.

Another way to illustrate this is to notice that the line where the mean fitness on the two network structures is the same – zero in Figure 5 – is decreasing with N. For larger N, the amplification effect is stronger for very small β, but the threshold below which this amplification occcurs at all gets smaller and smaller. In other words for a given β, there is always a sufficiently large population for which a star graph is *never* a fitness amplifier.

3 Discussion

In addressing the relative roles of selection and drift in evolution it is not enough to consider the fixation probability alone. This is because fixation can take much longer in some situations than in others: considering the fixation probability in isolation effectively limits us to the case where mutations are so rare as to be irrelevant at the timescale at which selection operates.

In order to be able to incorporate both fixation probabilities and the mutation rates essential to them in an appropriate way, we described a simple model of ongoing evolution in populations with a constant number of individuals along with (potentially) structured interactions. In this model selection tends to drive fitnesses up while mutations tend to drive it down, but both processes are inherently stochastic. Evolution tends towards populations in which these two tendencies balance one another on average. In panmictic populations this leads to fitnesses that are higher in large populations than small ones (though for sufficiently large populations this effect eventually saturates). Our results also show that while selection continues to restrict diversity in a population, the fitness of the population as a whole can drift considerably within bounds defined by a

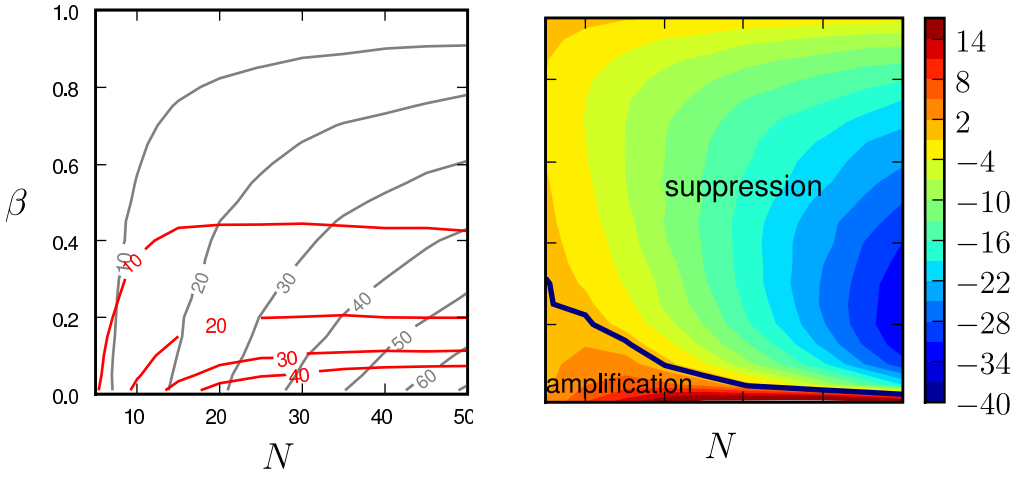

Figure 5: The STAR network. *(a)* Mean equilibrium fitnesses $\bar{f}_{STAR}$. The grey lines show $\bar{f}_{PAN}$ (ie. Figure 3) for comparison. *(b)* "Excess fitness": $\bar{f}_{STAR} - \bar{f}_{PAN}$ as a heatmap, plotted with the same axes.

broad ensemble distribution.

It has been suggested that networks with a strong hub may amplify fitness and selection (relative to drift), because the fixation probability is higher than for a panmictic population of the same size. Accounting for the rate of mutation changes this picture, as high mutation rates overwhelm any fitness enhancing effects. For large populations, the mutation rate at which this occurs may be extremely small. Thus in many circumstances, heterogeneous population structures tend to lower the fitness of a population.

Our aim in this paper has been to demonstrate the effect of network structure on mutation-selection balance. We have intentionally kept the model very simple, allowing the behaviour to be thoroughly understood and not overly sensitive to many detailed model choices, in the interests of robustness. Nevertheless, this work suggests several avenues that deserve exploration: for example, to extend the model to multiple loci (and hence Muller's ratchet), diploid organisms and more general mutation schemes.

Bibliography

[1] T. Antal, S. Redner, and V. Sood. Evolutionary dynamics on degree-heterogeneous graphs. *Physical Review Letters*, 96(18), 2006.

[2] M. R. Frean, G. J. Baxter and S. Hartley. In preparation.

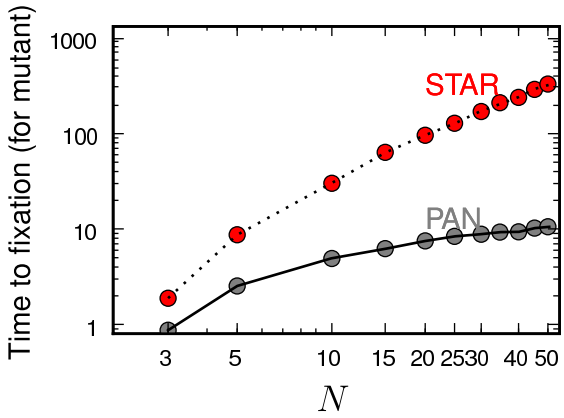

Figure 6: Mean time to fixation with $\beta = 0$ and $r = 2$. The data show how long it takes a single initial mutant to spread throughout the population, in those cases that it does so.

[3] D. Fudenberg, M.A. Nowak, C. Taylor, and L.A. Imhof. Evolutionary game dynamics in finite populations with strong selection and weak mutation. *Theor. Popn. Biol.*, 70(3):352–363, 2006.

[4] J.M. Pacheco H. Ohtsuki, M.A. Nowak. Breaking the symmetry between interaction and replacement in evolutionary dynamics on graphs. *Phys. Rev. Lett.*, 98(108106), 2007.

[5] M. Kimura and J. F. Crow. The number of alleles that can be maintained in a finite population. *Genetics*, 49:725–738, 1964.

[6] Erez Lieberman, Christoph Hauert, and Martin A. Nowak. Evolutionary dynamics on graphs. *Nature*, 433(7023):312–316, January 2005.

[7] T. Maruyama and M. Nei. Genetic variability maintained by mutation and overdominant selection in finite populations. *Genetics*, 98:441–459, 1981.

[8] P.A.P. Moran. Random processes in genetics. *Proceedings of the Cambridge Philosophical Society*, 54(60), 1958.

[9] P.A.P. Moran. *Mathematical Population Genetics*. Spinger, New York, 2004.

[10] T. Nagylaki and B. Lucier. Numerical analysis of random drift in a cline. *Genetics*, 94:497–517, 1980.

[11] Artem S. Novozhilov, Georgy P. Karev, and Eugene V. Koonin. Biological applications of the theory of birth-and-death processes. *Brief Bioinform*, 7(1):70–85, March 2006.

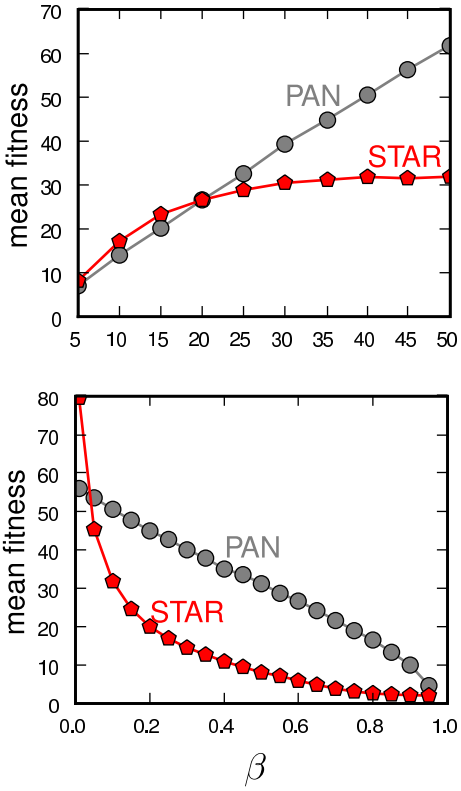

Figure 7: *(a)* Equilibrium fitness $\bar{f}$ vs N for constant mutation rate ($\beta = 0.1$) for STAR vs fully connected networks; *(b)* $\bar{f}$ vs mutation rate β for constant population size ($N = 40$).

[12] M.A. Nowak. *Evolutionary Dynamics: Exploring the Equations of Life*. Belknap Press, 2006.

[13] M.A. Nowak, A. Sasaki, C. Taylor, and D. Fudenberg. Emergence of cooperation and evolutionary stability in finite populations. *Nature*, 428:646–650, 2004.

[14] Hisashi Ohtsuki and Martin A. Nowak. The replicator equation on graphs. *Journal of Theoretical Biology*, In Press, Corrected Proof, 2006.

[15] G. C. Dick P. A. Whigham and H. G. Spencer. Genetic drift on networks: Ploidy and the time to fixation. *Theoretical Population Biology*, 74(4):283–290, December 2008.

[16] C. Taylor, D. Fudenberg, A. Sasaki, and M.A. Nowak. Evolutionary game dynamics in finite populations. *Bull. Math. Biol.*, 66:1621–1644, 2004.

[17] Peter D. Taylor, Troy Day, and Geoff Wild. Evolution of cooperation in a finite homogeneous graph. *Nature*, 447(7143):469–472, 2007.

[18] A. Traulsen, J.C. Claussen, and C. Hauert. Coevolutionary dynamics in large, but finite populations. *Phys. Rev. E*, 74(011901), 2006.

[19] G. A. Watterson. Heterosis or neutrality. *Genetics*, 85:789–814, 1977.