Assortative mating and mutation diffusion in spatial evolutionary systems

C. J. Paley,[1] S. N. Taraskin,[2,1] and S. R. Elliott[1]
[1]*Department of Chemistry, University of Cambridge, Lensfield Road, Cambridge CB2 1EW, United Kingdom*
[2]*St. Catharine's College, Cambridge CB2 1RL, United Kingdom*

(Received 26 May 2009; revised manuscript received 8 March 2010; published 14 April 2010)

The influence of spatial structure on the equilibrium properties of a sexual population model defined on networks is studied numerically. Using a small-world-like topology of the networks as an investigative tool, the contributions to the fitness of assortative mating and of global mutant spread properties are considered. Simple measures of nearest-neighbor correlations and speed of spread of mutants through the system have been used to confirm that both of these dynamics are important contributory factors to the fitness. It is found that assortative mating increases the fitness of populations. Quick global spread of favorable mutations is shown to be a key factor increasing the equilibrium fitness of populations.

DOI: 10.1103/PhysRevE.81.041912 PACS number(s): 87.23.−n, 05.40.−a, 68.90.+g

Spatial structure changes the fitness of populations in evolutionary biology and the dynamics of alleles spreading through them [1–7]. It has also been shown to change the dynamics of cooperative games [8–10]. Competition in the wild is mostly local and spatial [11]. Nevertheless, most population genetic models are nonspatial [12–15]. It is also true that, while it is increasingly recognized that space is an important factor in evolutionary studies and networks may be a useful way of studying this theoretically [16–19], an understanding of the role of space in the dynamics is still rudimentary.

Assortative mating, by which individuals mate with similar individuals, is also known to have an important effect on population fitness [20]. This may be considered to be a result of selection [21], it may be caused by the environment [22], or breeding pairs may even be sorted by age [23].

There are therefore two known, important effects of spatial structure: that which has an impact on the mating of individuals and that which has consequences for the spread of genes through offspring. In standard approaches, these two types of networks are not generally distinguished. In this paper, we allow for the mating and offspring-placement networks to be distinguished in order to mimic any possible structural effects and examine the effects of possibly different time scales (and thus different networks) for mating and offspring-placement processes.

Concerning the spatial structure of networks, there are three generic models: (i) fully connected network (each node is connected to every other node); (ii) random graph (nodes in (i) are connected with a certain probability); (iii) lattice model (e.g., nearest-neighbor nodes are connected). However, none of these limiting cases would seem to describe the inherent spatial structures underlying real biological populations. The actual underlying networks of biological mating and offspring dispersal are unknown. Therefore, an approach, in which one can controllably vary the network connectivity, e.g., by imposing long-range connections onto short-range structure, would be useful to investigate structure effects on population fitness.

Rewiring the links between nodes typical of the much-studied small-world-type networks [24–31] can mimic, e.g., migrations of individuals in spatial populations and allows the topology of the network to be changed continuously from that of the lattice to that of the random graph. Here, small-world-type networks (for both mating and offspring placement) are used as an investigative tool to examine an evolutionary system and to understand what factors might be important in determining the fitness of populations defined on a wider range of networks. By varying the structure of the offspring-placement and mate-choice networks independently, we are able to investigate the relative importance of spatial structure in the two and their interaction. By varying the correlation between the two networks, we have been able to vary the amount of assortative mating, without affecting offspring dispersal, across a range of spatial structures.

In this paper, we demonstrate that: (i) the impact of topology on global mutant spread through the population is important and when the time of mutant spread is reduced, the fitness is increased; (ii) assortative mating caused by local spatial structure also has a significant impact on the fitness. Below, we modify a model which has already been investigated for spatial lattices and fully connected graphs [4,32] in regard to structural effects. In this paper, the model is defined on a pair of either independent or related small-world-like networks: one which defines possible sexual partners and one which defines possible offspring placement. The use of the small-world topology allows the concentration of random connections in the system to be varied continuously and thus the topology of the system can be changed smoothly from a regular lattice to a random graph. Because we distinguish between the network from which sexual partners are selected and the network on which offspring are placed, and allow the two networks to vary in either a correlated or uncorrelated way, the level of assortative mating between sexual partners can be varied continuously along with, and separate from, the speed with which new mutants can spread through the population by birth. As these two effects happen in different ranges in the relevant parameter space, it is possible to observe directly their contributions to the fitness.

We have modified the "evolutionary-graph theory" model [17] in this paper. Each vertex i of a graph with N nodes represents an individual. The individuals have fitness r_i. On each turn (time step), an individual on a node i is selected for reproduction, with a probability, P_i, proportional to its fitness, r_i, i.e., $P_i = r_i / \Sigma_i^N r_i$. A clone of the selected individual is

placed onto one of the nodes, j, connected to it by the network. As all of the networks we consider have connectivity of four and equally weighted vertices, when an individual is selected for reproduction it will occupy one of the nodes connected to it by the network with equal probability of being placed onto any of the connecting nodes. The individual that previously occupied node j is replaced by the offspring, thus conserving the population size. This process is then repeated.

We extended the evolutionary-graph model in two ways to make our investigation of mating and offspring spread possible. The first way in which we extended it was to endow each individual with a set of genes (rather than a single number fitness) which may be affected by mutations [4,32]. This allows mating rules to be defined. The fitness of each individual, i, is determined by a set of G genes (genome) characterized by quality factors, q_{gi} (initially taken to be unity), so that the fitness is $r_i=\Sigma_g^G q_{gi}\geq 0$ (if r_i becomes negative in the process of evaluation, then it is set to zero and there is no upper bound on r_i). The subscript g cycles over the genes in an individual and the subscript i over individuals in the population.

Each time that an offspring is produced, it is subject to possible mutations. Mutations affect quality factors. At the time of reproduction, each quality factor will independently be mutated with a probability μ (therefore, μ is a per-gene mutation rate). The quality factor will increase by Δ_0 with a probability p (an advantageous mutation). The quality factor will decrease by Δ_0 with a probability $1-p$ (a deleterious mutation). This can formally be described by stochastic changes in quality factors, $q_{gi}\rightarrow q_{gi}+\Delta$, where Δ is a random variable characterized by the following probability density, $\rho(\Delta)=(1-\mu)\delta(\Delta)+\mu[p\,\delta(\Delta+\Delta_0)+(1-p)\,\delta(\Delta-\Delta_0)]$. This process is then repeated for each of the G genes of the genome.

The model thus modified has already been investigated. The population evolves to an equilibrium fitness if $p<0.5$ [32], caused by mutation-selection balance [4,33], proportional to Δ_0 provided that the initial dynamics do not lead to extinction (in which all individuals' fitness reach zero). The presence of an equilibrium is due to the fact that deleterious mutations outnumber advantageous mutations. As the population fitness increases, the relative effect of each new advantageous mutation is proportionally smaller. Therefore, the accumulation of negative mutations balances the spread of new advantageous mutations.

The model allows two ways of reproduction: asexual and sexual. In the asexual case, the offspring genome $(q_{1i},\ldots,q_{Gi})$ is a clone of its parent. For example, the following sequence, $(11,7,3,12,7)\rightarrow(11,7,3,12,7)\rightarrow(12,7,2,12,7)$, describes the reproduction of an individual with initial fitness $r=40$ and $G=5$ genes. Its offspring is subject to two mutations (of magnitude $\Delta_0=1$): an advantageous mutation in $g=1$ and a deleterious mutation in $g=3$. The asexual way of reproduction for this model defined on a fully connected graph was first studied in Ref. [34].

In the sexual case, when an individual is chosen for reproduction, it selects a partner at random from the nodes connected to it on the graph, with a probability proportional to potential mates' fitness, and the offspring is produced by a

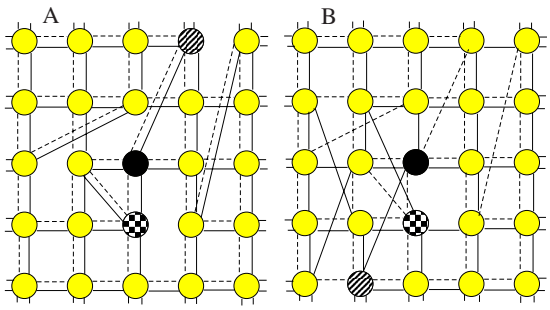

FIG. 1. (Color online) In a sexual population, two networks define the interactions between individuals in the population. The partner-selection and offspring-placement networks are shown by dashed and solid lines, respectively. A parent is selected according to its fitness (solid shading). A second parent is chosen (squared shading) from the nearest neighbors in the partner-selection network and the offspring is placed onto a node (striped shading) connected to the first parent by the offspring-placement network. In (a), the two networks are the same. In (b), the partner-selection network and offspring-placement networks are formed in the same way, with the same value of p_c, but independently. For small values of p_c, both networks in (b) are topologically correlated due to a significant proportion of nonbroken square-lattice connections.

Mendelian shuffling of the genes. For example, the following sequence, $(11,7,3,12,7)+(12,5,4,11,9)\rightarrow(11,7,4,12,9)\rightarrow(11,6,4,12,8)$, corresponds to sexual reproduction and mutation. An individual with fitness $r=40$ was selected for reproduction. From among possible mates connected to it by the partner-selection network, it selected a partner which had fitness $r=41$. An offspring was produced, each quality factor q_g was chosen at random from its two parents (so it received $g=1,2,4$ from one parent and $g=3,5$ from the other). It was then subject to mutations in $g=2,5$, both of them deleterious ($\Delta_0=1$). It has a resulting fitness $r=41$. The above criteria for sexual reproduction define hermaphrodite haploids.

The networks from which partners are selected and on which offspring are placed are defined as follows. A square lattice with periodic boundary conditions is used as a base network for the small-world-like networks and constructed in the following manner (see Fig. 1). With probability p_c, each link on the lattice is disconnected. The nodes with missing links are then connected randomly, allowing connections to reform, but avoiding double links. The value of p_c can be varied in the range $0\leq p_c\leq 1$, with $p_c=0$ corresponding to the square lattice and $p_c=1$ to the 4-regular graph (4 is the node coordination number). Unlike the Watts-Strogatz small-world network [24], the network used here maintains the constant connectivity of the nodes. The system is described by a symmetric stochastic matrix, ensuring that the probability of a mutant fixing in the system is the same in all of the networks studied [17].

In an asexual model, the network is used only for placing an offspring. In a sexual model, the network is used both for partner selection and for placing offspring. These two networks can be either identical or different if $p_c>0$ (see Fig. 1). If p_c is the same in both networks and the two networks are different, then topological correlations are gradually destroyed as p_c increases. When p_c is low, the two net-

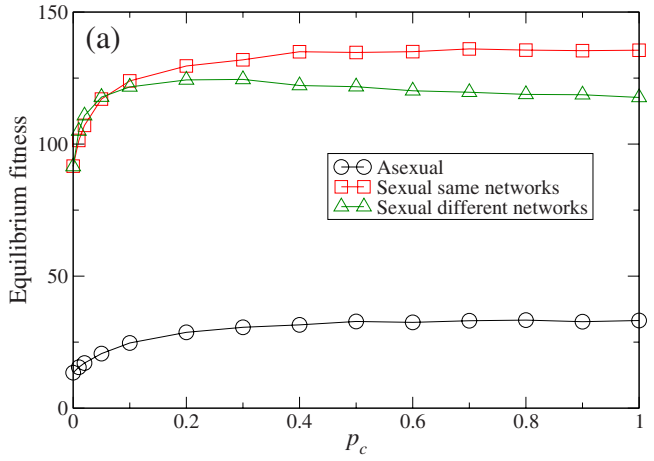

FIG. 2. (Color online) Relationship between mean equilibrium fitness, $N^{-1}\Sigma_i^N r_i$, and reconnection probability, p_c, in the small-world-like network for a sexual model when the parent and offspring networks are the same (squares) or different (triangles). The fitness of the asexual model (with only one network for offspring) is shown for comparison (circles). The parameters used in the simulation are $G=10$, $\mu=0.13$, $p=0.25$, $\Delta_0=1$, and $N=2500$. Data were averaged over 400 realizations of the system after equilibrium was reached. The errors are smaller than the symbols used. Lines are guides to the eyes.

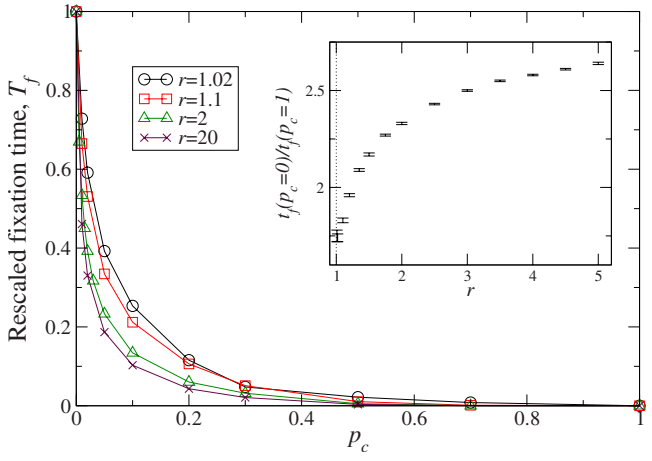

FIG. 3. (Color online) Variation of rescaled time to fixation, T_f, for a mutant on a network with link-reconnection probability p_c for an evolutionary-graph model with varying r, as marked by different symbols ($N=2500$). Where error bars are not shown, the errors are smaller than the symbols used. Lines are guides to the eyes only. Each point has been obtained by averaging over at least 850 fixation events. Inset shows the ratio between the fixation time (before rescaling) in the square lattice ($p_c=0$) and in the 4-regular graph ($p_c=1$) vs the mutant fitness.

works are highly correlated because they are both similar to square lattices. However, when p_c is high, the two networks approach independently created 4-regular graphs. Alternatively, if the partner-selection and offspring-placement networks are the same, then correlations between the two networks are high whatever value p_c takes. This is important because reducing the amount of correlation between the two networks must also reduce the amount of assortative mating. Individuals will select partners who are less likely to be closely related to them. By reducing the amount of assortative mating in this way, without affecting the spread of mutations through the population by other means, we have been able to separate the fitness effects of assortative mating.

These correlations significantly influence the equilibrium fitness of sexual populations. Figure 2 shows how the fitness of populations changes with p_c in sexual populations using both the same networks and different, overlaid networks (cf. the curves plotted with triangles and squares) and also in an asexual population for comparison (marked by circles). A small amount of global reconnection corresponding to an increasing value of p_c quickly leads to increased fitness in all three networks. The steep increase in fitness at very small $p_c(\ll 1)$ is consistent with a view that long-range correlations cause a much quicker global spread of advantageous mutants, an effect that typifies small-world networks [24]. This benefit due to the shortcuts in the network quickly saturates at $p_c \simeq 0.2$. Thereafter, in the sexual system where the two networks are varied independently, there is a slow decrease in fitness with p_c for $p_c \gtrsim 0.2$ (see triangles in Fig. 2). This decrease can be qualitatively understood as follows. As the reconnection level increases further, the local structure is broken up when the networks are independent. The genetic relatedness of sexual partners would be expected to decrease, reducing the degree of assortative mating and thus leading to a gradual decrease in fitness. This effect is not observed in the system where the partner-selection and mate-placement networks are identical (see the curve labeled by squares in Fig. 2).

We now investigate separately the rapid increase in fitness on increasing p_c from low values and the diverging fitness of the sexual networks for higher p_c. We demonstrate that the first behavior is consistent with an increase in rapidity of spread of new mutations due to a reduction in path length. The diverging fitness does indeed correspond to a decrease in assortative mating when the partner-selection and offspring-selection networks are varied independently.

The Watts-Strogatz small-world network [24], on which our network is based, is known to cross over from a large path length (minimum number of links between two nodes) to a short one at relatively low levels of rewiring. This might lead to shorter times $t_f(N,r,p_c)$ for an advantageous mutation to spread through the population. To study the speed of spread alone, we use the evolutionary-graph theory model [17] discussed above, in which a single mutant (of fitness r) is placed in an otherwise homogenous population (fitness of all other individuals is set to unity) and further mutation is suppressed. The population used is asexual such that dispersal via offspring placement alone contributes to spread. In this paradigm, fixation occurs when the single mutation present in the system has occupied all nodes.

The anticipated behavior is indeed demonstrated in Fig. 3 which shows how the fixation time for mutants of fitness r in the system depends on the reconnection probability. The rescaled fixation time, $T_f=[t_f(N,r,p_c)-t_f(N,r,1)]/[t_f(N,r,0)-t_f(N,r,1)]$, plotted in Fig. 3 vs p_c, allows all the curves for different r to be shown on the same scale. These plots show that the mutant spread time rapidly decreases on increasing

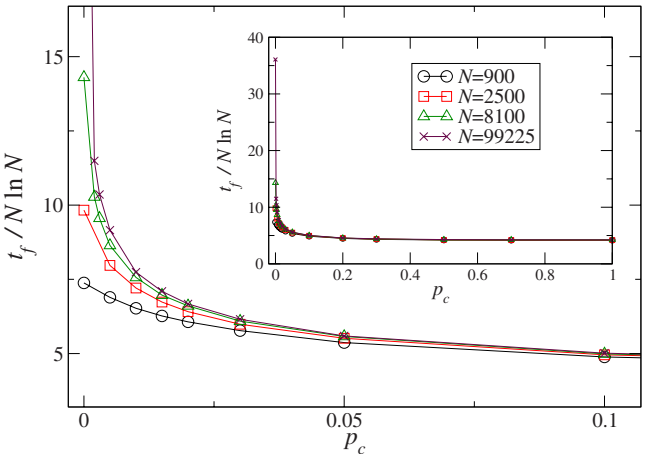

FIG. 4. (Color online) Dependence of the normalized fixation time, $t_f/N \ln N$, on the reconnection probability for networks of different size ($r=2$). The main graph is an expansion of the small p_c region of the inset. The errors are smaller than the symbols used. Lines are guides to the eyes only.

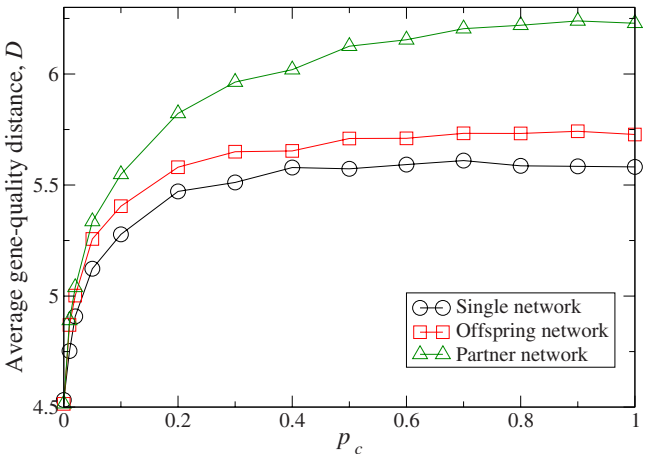

FIG. 5. (Color online) Relationship between the mean inter-neighbor gene-quality distance, D, and the reconnection probability for different networks (as marked). The averaging has been done over 400 equilibrium realizations. The errors are smaller than the symbols used. Lines are guides to the eyes only.

p_c from zero. This is consistent with our explanation for the sudden rise in fitness for small p_c in Fig. 2 being caused by the quicker spread of advantageous mutants. The inset in Fig. 3 shows the ratio $t_f(N,r,0)/t_f(N,r,1)$ as a function of the mutant fitness r, demonstrating that the behavior is qualitatively similar for a wide range of values of r.

For the square lattice, it has been shown [4] that the fixation time scales with population size as $t_f \propto N^2$ and for the fully connected system, the scaling is $t_f \propto N \ln N$. Bearing in mind the similarity in topology between fully connected (complete) and Z-regular graphs, it may be anticipated that the same scaling will hold for both types of networks. Indeed, $N \ln N$ scaling has been confirmed numerically for the Z-regular graph, as demonstrated in the inset in Fig. 4 (the rescaled fixation time for $p_c=1$ does not depend on N). This graph shows that the scaling law found in the fully connected graph holds even for small p_c ($\gtrsim 0.05$) for a wide range of system sizes. For even smaller values of $p_c \lesssim 0.05$, a change to the scaling regime found for the square lattice ($t_f \propto N^2$ for $p_c=0$) eventually occurs.

In order to investigate the degree of assortative mating, we define a "genetic distance," D_{ij}, between two individuals in terms of their gene-quality factors as $D_{ij}=[G^{-1}\Sigma_{g=1}^{g=G}(q_{ig}-q_{jg})^2]^{1/2}$, where the sum is taken over all the genes for individuals on nodes i and j (below, i and j refer to nearest neighbors only). Such a definition of the genetic distance implies that the larger this distance, the greater the level of genetic difference between the two individuals. If the individuals connected on the partner-selection network have a greater distance, then there is a smaller degree of assortative mating. The average genetic distance between nearest neighbors depends on the probability of reconnection (see Fig. 5). It can be seen from Fig. 5 that, when there is a single network, with partners being chosen across the same link as offspring are placed, the distance grows in line with the fitness (cf. the line marked by the circles in Fig. 5 with the lines marked by the squares in Fig. 2). When there are two different overlaid networks for placing offspring and selecting partners, thereby breaking up the correlations, the situation changes. For the offspring network, just a slight increase in the genetic distance is observed (cf. the curves marked by squares and circles in Fig. 5). However, the genetic distance between sexual partners increases significantly and steadily as p_c is increased (cf. the curve marked by triangles with the curves marked by the squares and circles). This demonstrates, as expected, that the degree of assortative mating decreases with increasing p_c over the scale that the fitness decrease is seen in Fig. 2. It may be argued that, in real biological systems, a decrease in assortative mating would be a positive thing, leading to less inbreeding. This effect is not considered in this simple model and, as with real systems, the situation is likely to be rather complicated. However, the method of analysis presented above highlights the importance of assortative mating for the model considered and gives a means of quantifying the effect. Ignoring spatial effects would lead to an important contributor to fitness—namely, assortative mating—being missed.

To conclude, we have presented a technique to examine the ways in which the fitness of sexual and asexual populations is affected by spatial structure in the population. It is found that the presence of space-induced assortative mating leads to fitness benefits. It has also been shown that fitness is implicitly correlated with the speed of spread of mutants through the system—a faster global spread increases the fitness. Technically, we have investigated the above effects by using paired small-world-like networks for different processes in the model. This method is rather general and could be used and developed in other areas where spatial structure is important.

C.J.P. would like to thank the EPSRC-GB for financial support. We are grateful to the Cambridge University Condor Grid for the provision of computational resources.

[1] M. Whitlock, Genetics **164**, 767 (2003).
[2] T. Maruyama, Theor Popul. Biol. **1**, 273 (1970).
[3] M. Slatkin, Evolution (Lawrence, Kans.) **35**, 477 (1981).
[4] C. J. Paley, S. N. Taraskin, and S. R. Elliott, Phys. Rev. Lett. **98**, 098103 (2007).
[5] M. G. J. L. Habets, T. Czárán, R. F. Hoekstra, and J. A. G. M. de Visser, Proc. R. Soc. London, Ser. B **274**, 2139 (2007).
[6] L. Perfeito, M. Pereira, P. R. Campos, and I. Gordo, Biol. Lett. **4**, 57 (2008).
[7] M. Broom and J. Rychtár, Proc. R. Soc. London, Ser. A **464**, 2609 (2008).
[8] M. A. Nowak and R. M. May, Int. J. Bifurcation Chaos Appl. Sci. Eng. **3**, 35 (1993).
[9] C. Hauert and M. Doebeli, Nature (London) **428**, 643 (2004).
[10] G. Szabó and G. Fáth, Phys. Rep. **446**, 97 (2007).
[11] *Spatial Ecology: The Role of Space in Population Dynamics and Interspecific Interactions*, edited by D. Tilman and P. Kareiva (Princeton University Press, Princeton, NJ, 1997).
[12] M. Kimura, J. Appl. Probab. **1**, 177 (1964).
[13] M. Kimura and T. Maruyama, Genetics **54**, 1337 (1966).
[14] A. Kondrashov, Nature (London) **336**, 435 (1988).
[15] E. Cohen, D. A. Kessler, and H. Levine, Phys. Rev. Lett. **94**, 098102 (2005).
[16] S. Proulx, Trends Ecol. Evol. **20**, 345 (2005).
[17] E. Lieberman, C. Hauert, and M. Nowak, Nature (London) **433**, 312 (2005).
[18] L. Berec, Ecol. Modell. **150**, 55 (2002).
[19] J. Krause, D. Lusseau, and R. James, Behav. Ecol. Sociobiol. **63**, 967 (2009).
[20] K. Jaffe, Acta Biotheor. **47**, 29 (1999).
[21] W. O. McMillan, L. A. Weigt, and S. R. Palumbi, Evolution (Lawrence, Kans.) **53**, 247 (1999).
[22] M. Doebeli and U. Dieckmann, Nature (London) **421**, 259 (2003).
[23] M. Ferrer and V. Peneteriani, Anim. Behav. **66**, 137 (2003).
[24] D. Watts and S. Strogatz, Nature (London) **393**, 440 (1998).
[25] S. Dorogovtsev and J. Mendes, *Evolution of Networks: From Biological Nets to the Internet and WWW* (Oxford University Press, New York, 2002).
[26] R. Albert and A. Barabási, Rev. Mod. Phys. **74**, 47 (2002).
[27] M. Barthélémy and L. A. Nunes Amaral, Phys. Rev. Lett. **82**, 3180 (1999).
[28] L. F. Lago-Fernández, R. Huerta, F. Corbacho, and J. A. Sigüenza, Phys. Rev. Lett. **84**, 2758 (2000).
[29] X. F. Wang and G. Chen, Int. J. Bifurcation Chaos Appl. Sci. Eng. **12**, 187 (2002).
[30] J. Davidsen, H. Ebel, and S. Bornholdt, Phys. Rev. Lett. **88**, 128701 (2002).
[31] J. Zhang and M. Small, Phys. Rev. Lett. **96**, 238701 (2006).
[32] C. Paley, S. Taraskin, and S. Elliott, Naturwiss. **94**, 505 (2007).
[33] R. Bürger, *The Mathematical Theory of Selection, Recombination, and Mutation* (Wiley, New York, 2000).
[34] D. Ridgway, H. Levine, and D. Kessler, J. Stat. Phys. **90**, 191 (1998).