Mutation-selection stationary distribution in structured populations

Sedigheh Yagoobi,* Hosna Yousefi, and Keivan Aghababaei Samani†
Department of Physics, Isfahan University of Technology, Isfahan 84156-83111, Iran

(Received 6 May 2018; published 2 October 2018)

In this paper we address the evolution of finite and structured populations including two types of species. The evolutionary dynamics of the system is governed by the Moran process with constant fitnesses in the presence of mutation. We obtain the stationary distribution in a number of topologies. We also approximate the *mixing time*, i.e., the time that the system needs to reach its stationary distribution. It is observed that the mean frequency of a species in the stationary distribution is approximately independent of the population structure, and the mixing time has a power law behavior with respect to the population size whose exponent is closely related to the population structure. The obtained results indicate that more heterogeneity leads to longer mixing times.

DOI: 10.1103/PhysRevE.98.042301

I. INTRODUCTION

Traditionally, evolutionary game dynamics has been investigated in well-mixed infinite populations. Deterministic evolutionary dynamics is described by the replicator equation(s) which can be interpreted as a genetic or cultural evolution. However, in the real world we deal with structured populations of interacting individuals either in biological or social systems. Recently, the effect of population structure on evolution has attracted a lot of attention [1–4]. Moreover, the evolutionary graph theory is a promising tool for studying the impact of structure on evolutionary dynamics [5–8].

For small enough mutation rates, one of the species takes over the whole population before a new mutation arises [9]. In this case, the so-called fixation probability and the fixation time would be important [10–12]. However, in the systems where mutation occurs more often, such as in social systems, the population is a mixture of different species. Structural symmetries of underlying graphs can be very useful to obtain analytical results, such as fixation probability and fixation time for special graph topologies [13].

If we look at the evolutionary process as a Markov chain, the so-called *mixing time* as defined in the probability theory is the time needed for the chain to get close to its stationary distribution within a small error interval, say ϵ [14]. Black *et al.* [15] studied the stationary distribution and the mixing time for an evolutionary game in a well-mixed population. Their results are based on simulations and a WKB approximation of the master equation. They explored the effect of intensity of selection on the mixing time as well as the fixation time of evolutionary games with metastable states and evolutionary games with bistable states.

The effect of mutation in well-mixed populations is addressed in some studies [16–21], but researchers have not treated structured populations in much detail. Here we study the impact of structures on the evolution of their corresponding populations where each of them is composed of two types of species with constant fitnesses. We see how the mixing time of each topology is related to the mutation rate and population size. The evolution of populations is modeled by the Moran process with mutation which behaves as an ergodic Markov chain, and thus its average behavior over time tends to show a mutation-selection stationary distribution [22].

The organization of the paper is as follows. In Sec. II, a description of our model is presented. The process of obtaining the stationary distribution for complete and star graphs by using transition matrices is elucidated in Sec. III. In Sec. IV we compare stationary distribution for different topologies. Finally in Sec. V, we compute the mixing time for different topologies. Section VI is devoted to a summary and concluding remarks.

II. MODEL

Consider a structured population of size N with two types of species, say A and B, assuming their respective fitnesses equal 1 and r where $r > 1$. The evolution of this system is determined by a birth-death process. In a birth-death process, one individual is selected at each time step for birth with a probability proportional to its fitness, and then one of its randomly chosen neighbors is replaced by its offspring. In a two-type population without mutations, each individual reproduces an offspring of its own type, however in a population including two types of species with mutations, the selected individual reproduces an offspring of its own type with the probability $1-\mu$ and of the other type with the probability μ. Regardless of the topology, we have a set of states (Markov space states) determined by the number i of type B species $i \in \{0, \ldots, N\}$. According to the presented dynamics, the transitions are of three types: $i \longrightarrow i+1$, $i \longrightarrow i-1$, and $i \longrightarrow i$. Let P be the transition matrix of this process. P_{ij} is the transition probability from state i to state j. This process constructs an ergodic Markov chain in which each state is accessible from every other state (in one or a finite number of steps), and it has a unique stationary distribution π which is the left

*s.yaghoobi@ph.iut.ac.ir
†samani@cc.iut.ac.ir

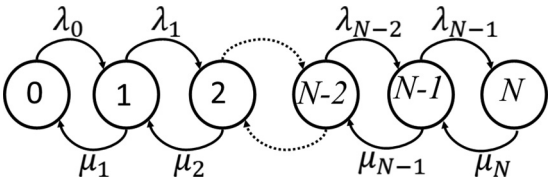

FIG. 1. The Markov states diagram of the evolutionary process on a network. State i represents all configurations with i as the number of individuals of type B, $i \in \{0, \ldots, N\}$. $\lambda_i = P_{i,i+1}$, $\mu_i = P_{i,i-1}$.

eigenvector of the transition matrix P with eigenvalue 1,

$$\pi P = \pi. \tag{1}$$

The stationary distribution is the probability distribution over the state space of the system in a long period of time. In a Markov chain, the distribution at time t equals $\nu_t = \nu_0 P^t$, where ν_0 is the initial distribution. Given enough time, the probability distribution converges to π regardless of the initial conditions. To obtain this quantity, one can easily compute P^t when $t \to \infty$ so that every row of P^t equals π.

For a structured population one can distinguish a number of configurations with the same number of species. Practically, tracing all of these configurations and transitions between them is not possible. Therefore, we represent all configurations with the same number of species of type B by a single *state* in the Markov chain (Fig. 1). This is in some sense a kind of coarse graining the original state space.

III. THE STATIONARY DISTRIBUTION OF COMPLETE AND STAR GRAPHS

Complete and star graphs are prototypes of highly homogeneous and heterogeneous networks, respectively. Therefore, they are worth studying to understand the relationship between evolutionary dynamics and structural properties of networks. This has been the subject of a number of investigations. For example, in Refs. [23,24] the fixation probability of the Moran process on star graphs is obtained analytically. The effect of heterogeneity is also studied in Ref. [25].

In this section we numerically calculate the stationary distribution for star and complete graphs. Then we compare the mean fraction of B's for complete and star graphs by approximating the master equation.

A. Probability distribution

The Markov space of a complete graph is of the general form shown in Fig. 1. In the case of a star graph, we prefer to study the state space in more detail. We consider two types of states distinguished by the type of the central node. Therefore, a Markov chain composed of $2N$ states (Fig. 2) could be obtained. The stationary distributions for complete and star graphs are depicted in Figs. 3 and 4, respectively. For these two graphs, one can calculate the exact form of the transition matrix analytically (see the next section). Therefore, we can use Eq. (1) to find the stationary distribution. Figures 3 and 4 are generated by this method. Both graphs show peaks in $i \simeq 80$ for the mutation probability $\mu = 0.1$. As mentioned

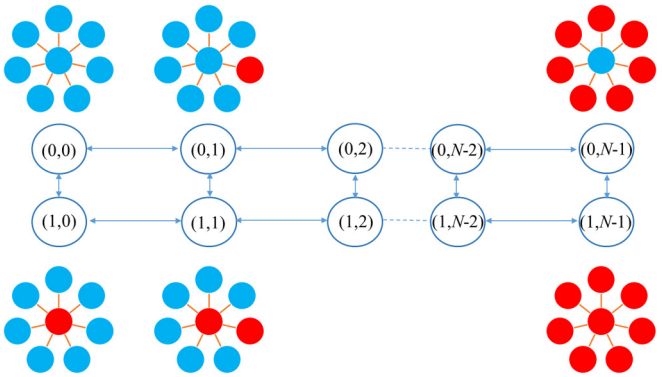

FIG. 2. Markov chain states of a star graph with N nodes. The states are divided into two groups in which the center is occupied by either type A or type B. We determine these two groups by using the notation (k, i) in which k is 0 or 1 if the center is occupied by type A or B, respectively, and i denotes the number of nodes in the periphery that are occupied by B.

before, in the star graph we can distinguish two types of states, namely, the B-centered and A-centered states. In Fig. 4 stationary (sub)distributions corresponding to these two types are shown separately as well as the overall distribution which is, in fact, the sum of A-centered and B-centered distributions. The ratio of A-centered to B-centered peaks is $\frac{p_A}{p_B} = \frac{N-m}{m}$, where p_A and p_B are the probabilities of the center occupation by A or B at the stationary distribution and m denotes the average number of B-type species in the stationary distribution.

B. Comparing the mean fraction of B's for complete and star graphs

The probability $p^\tau(i)$ of finding the system in state i at time τ obeys the master equation,

$$p^{\tau+1}(i) - p^\tau(i) = p^\tau(i-1)T^+(i-1) + p^\tau(i+1)T^-(i+1) - [T^+(i) + T^-(i)]p^\tau(i), \tag{2}$$

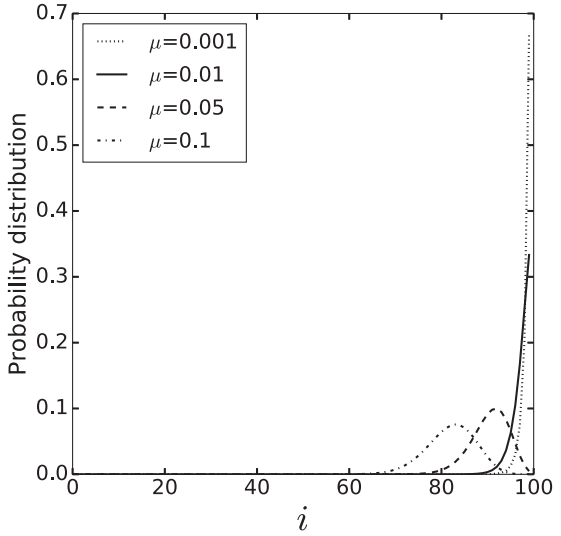

FIG. 3. The numerical stationary distribution for the complete graph with $r = 2$ and $N = 100$.

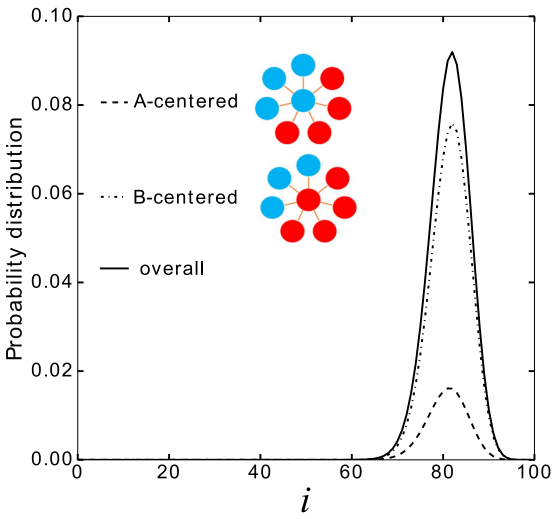

FIG. 4. The numerical stationary distribution for the star graph with parameters $r = 2$, $\mu = 0.1$, and $N = 100$. The three curves correspond to the A-centered, B-centered configurations, and the overall probability distribution. The mean value of the number of B's is $\bar{n} = 82$.

where $T^+(i)$ and $T^-(i)$ are the transition probabilities from i to $i+1$ and $i-1$, respectively. Considering both selection and mutation, there are four types of transitions that change the number of B's. Two of them increase the number of B's with the probability $T^+(i)$, and the other two decrease the number of B's with the probability $T^-(i)$. Here we explain in detail the computation of $T^+(i)$ for a complete graph. Assume the system is in state i. The transitions that increase the number of B's are of the following two types:

(a) A node of type A is chosen for reproduction with probability $\frac{N-i}{N-i+ri}$ and reproduces an offspring of type B with probability μ. Then one of its neighbors of type A is replaced by the newborn B with the probability $\frac{N-i-1}{N-1}$. As a result, the probability of increasing the number of B's in this transition is $\mu\frac{N-i}{N-i+ri}\frac{N-i-1}{N-1}$.

(b) A node of type B is chosen for reproduction with probability $\frac{ri}{N-i+ri}$ and reproduces an offspring of its own type with probability $1-\mu$. Then one of its neighbors of type A is replaced by the newborn B with the probability $\frac{N-i}{N-1}$. As a result, the probability of increasing the number of B's in this transition is $(1-\mu)\frac{ri}{N-i+ri}\frac{N-i}{N-1}$.

Therefore, $T^+(i)$ is the sum of the above two probabilities,

$$T^+(i) = \frac{\mu(N-i)(N-i-1)+(1-\mu)ri(N-i)}{(N-i+ri)(N-1)}. \quad (3)$$

Likewise, $T^-(i)$ is easily obtained as below,

$$T^-(i) = \frac{(1-\mu)(N-i)i+\mu ri(i-1)}{(N-i+ri)(N-1)}. \quad (4)$$

For the star graph there are eight possible transitions that change the number of B's such that four of them increase the number of B's and the other four decrease it. Here we explain just one of B increasing transitions. The other ones can be deduced analogously.

Consider a star graph with one B at its center and $i-1$ B's at its leaves. Let the central node be selected for reproduction with probability $\frac{r}{N-i+ri}$ and reproduce an offspring of its own type with probability $1-\mu$. Then one of its neighbors of type A is replaced by the newborn B with probability $\frac{N-i}{N-1}$. Therefore the probability of increasing the number of B's in this transition is $(1-\mu)\frac{r}{N-i+ri}\frac{N-i}{N-1}$. By considering the other types of transitions, one can obtain the total transition probabilities as follows:

$$T^+(i) = p_B(1-\mu)\frac{r}{N-i+ri}\frac{N-i}{N-1} + p_A(1-\mu)\frac{ri}{N-i+ri} + p_A\mu\frac{1}{N-i+ri}\frac{N-i-1}{N-1} + p_A\mu\frac{N-i-1}{N-i+ri}, \quad (5a)$$

$$T^-(i) = p_A(1-\mu)\frac{1}{N-i+ri}\frac{i}{N-1} + p_B(1-\mu)\frac{N-i}{N-i+ri} + \mu p_B\frac{r}{N-i+ri}\frac{i-1}{N-1} + p_B\mu\frac{r(i-1)}{N-i+ri}, \quad (5b)$$

where p_A and p_B are the fractions of A-centered and B-centered configurations, respectively, with i B's that are closely related to the initial condition and the time step. It is worth mentioning that in the stationary distribution $p_A = \frac{N-i}{N}$ and $p_B = \frac{i}{N}$.

The stationary distribution $\pi(i)$ is obtained by the condition $p^{\tau+1}(i) = p^\tau(i) = \pi(i)$ which leads to the following detailed balance condition:

$$T^-(i+1)\pi(i+1) = T^+(i)\pi(i). \quad (6)$$

The stationary distribution is obtained as follows:

$$\pi(i) = \frac{\prod_{j=0}^{i-1} R_j}{\sum_{i=0}^{N}\prod_{j=0}^{i-1} R_j}, \quad R_j = \frac{T^+(j)}{T^-(j+1)}. \quad (7)$$

This is exactly the stationary distribution π of Eq. (1). Using this stationary distribution one can obtain the mean value for the number of B's, $\bar{n} = \sum_{i=0}^{N} i\pi(i)$.

As an example for the application of the above relations we consider the evolutionary dynamics of a complete graph of size $N = 100$ with parameters $r = 2$ and $\mu = 0.1$. The mean value of the number of B's in this case is $\bar{n} = 82$ which is in agreement with the peak of the stationary distribution illustrated in Fig. 4.

Now, we use this formalism to compare the effect of mutation probability in star and complete graphs. Figure 5 shows the mean value of the fraction of B's versus mutation probability for star and complete graphs with $N = 100$ nodes. As is seen, they completely coincide. This can be explained by obtaining the large N approximation of the master equation Eq. (2), i.e., $N \gg 1$.

By defining $x := i/N$ and $t := \tau/N$ and the probability density $\rho(x,t) := Np^\tau(i)$, we can expand the master equation in a Taylor series in x and t which ultimately leads to [22]

$$\frac{d\rho(x,t)}{dt} = -\frac{d}{dx}[a(x)\rho(x,t)] + \frac{1}{2}\frac{d^2}{dx^2}[b^2(x)\rho(x,t)], \quad (8)$$

where $a(x) = T^+(x) - T^-(x)$ and $b(x) = \sqrt{\frac{1}{N}[T^+(x)+T^-(x)]}$. For large but finite values of N, this equation has the form of Fokker-Planck equation which

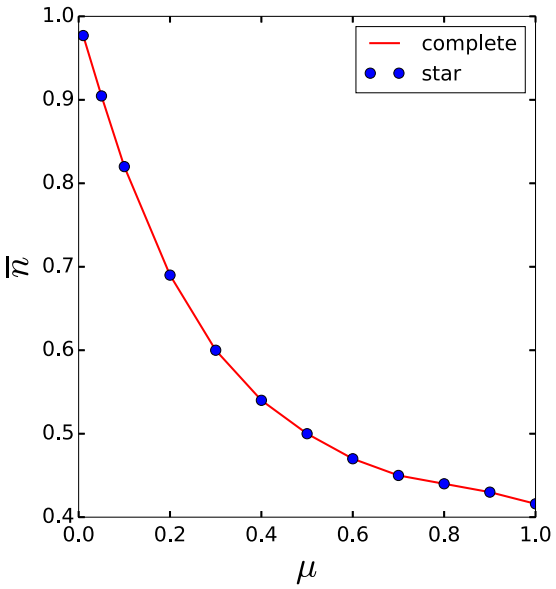

FIG. 5. The comparison of the mean fraction of B's $\bar{n}$ for the star (blue circles) and complete graph (red line) versus mutation probability μ with parameters $r = 2$ and $N = 100$. They are exactly matched.

corresponds to the following Langevin equation:

$$\dot{x} = a(x) + b(x)\xi, \tag{9}$$

where ξ is a white noise with properties $\langle \xi(t) \rangle = 0$ and $\langle \xi(t)\xi(t') \rangle = \delta(t - t')$. In large N approximation, only the drift term survives and a deterministic equation is obtained as

$$\dot{x} = a(x), \tag{10}$$

where $a(x)$ is the same for the complete and star graphs in the stationary regime,

$$a(x) = \frac{(1-r)x^2 + [r(1-\mu) - \mu - 1]x + \mu}{1 - x + rx}. \tag{11}$$

Therefore, for a large population, the mean fraction of B's versus mutation probability for both the star and the complete graphs is identical.

IV. THE EFFECT OF STRUCTURE ON THE STATIONARY DISTRIBUTION

In this section, we investigate the role of the population structure on the stationary distribution of the evolutionary process. Examining the topologies that are more complex than the star and complete graphs is more challenging because of their greater complexity and less symmetry. Here we cannot calculate the transition matrix analytically. Therefore, for these types of topologies we found the stationary distributions by simulation. The results are depicted in Figs. 6–8. According to Fig. 6, the probability distribution for the cycle graph has a maximum in a lower number of B-type individuals compared to other topologies. The star graph has a sharper maximum than other topologies. This may be traced in the special structure of the star graph because all the sites are connected to each other via the central node. Therefore, the changes happen slowly, and the system stays in the same state for a longer time compared to the other topologies.

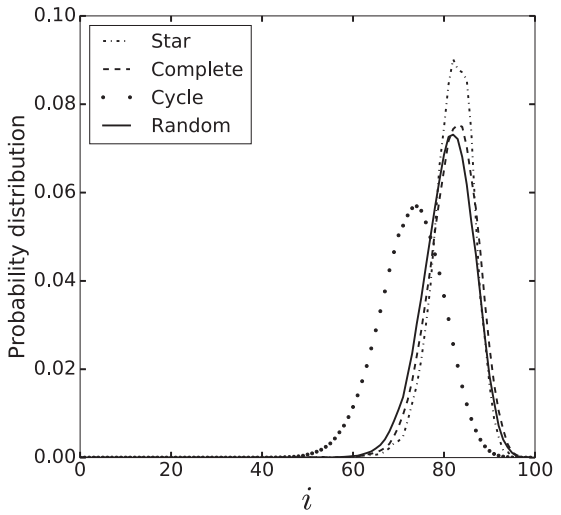

FIG. 6. The stationary distribution for the complete, star, random, and cycle graphs with $r = 2$, $\mu = 0.1$, and $N = 100$. All stationary distributions are approximately the same except for the cycle graph which has the minimum average degree.

The stationary distribution of the scale-free (SF) network based on the Barabasi-Albert (BA) model is illustrated in Fig. 7 for different values of m which denote the number of links in the BA model through which each new node connects to the present nodes. As m increases, the probability distribution converges to that of the complete and star graphs.

The stationary distribution of the Watts-Strogatz (WS) model for the small-world (SW) network is shown in Fig. 8. Two parameters are varied in this kind of graph: the node degree of initial regular network K and the rewiring

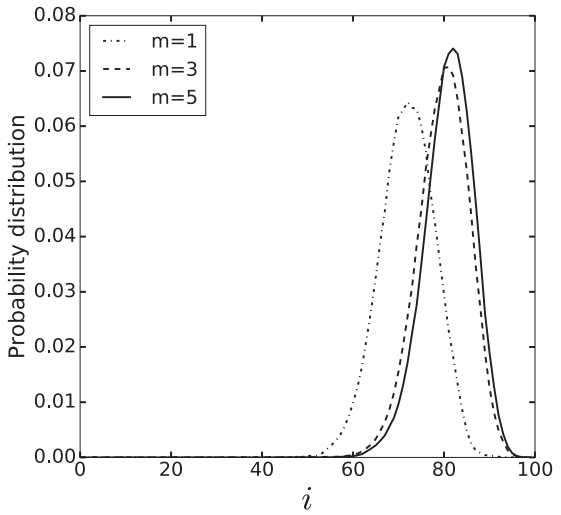

FIG. 7. The stationary distribution for the BA scale-free network for different m's, i.e., the number of links in the BA model through which each new node connects to the present nodes with $r = 2$, $\mu = 0.1$, and $N = 100$. As m increases, the probability distribution converges to the probability distribution of the complete and star graphs. The number of realization, 100.

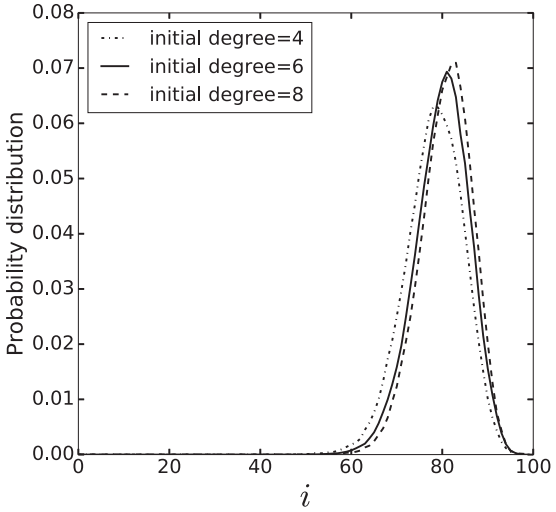

FIG. 8. The stationary distribution for a WS small-world network for different mean degrees, and the rewiring probability $p = 0.1$ with $r = 2$, $\mu = 0.1$, and $N = 100$.

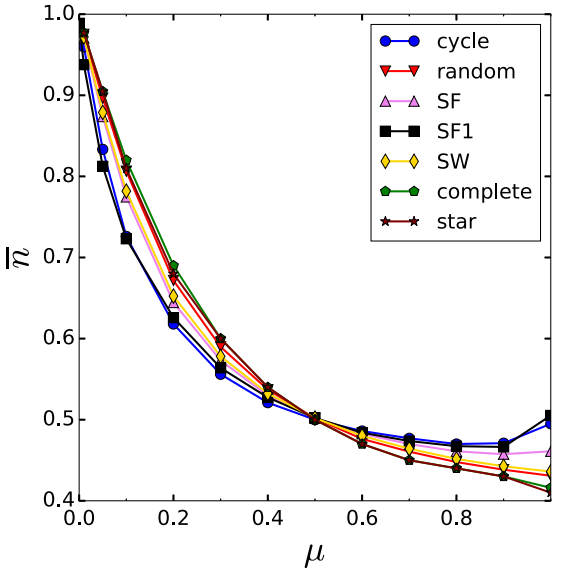

FIG. 9. The mean fraction of B's $\bar{n}$ versus the mutation probability μ for different topologies with $r = 2$, $N = 100$. The number of realizations for the SF and SF1 is taken to be 100.

probability p. To build a small-world graph based on the Watts-Strogatz algorithm, we first construct a regular graph with a circular shape in which each node is connected to its K nearest neighbors $K/2$ at each side. Then, we rewire every edge with a determined rewiring probability. For the p values between 0.001 and 0.1, the obtained graph will be a small-world network with a high clustering coefficient and a small average shortest path length. Except for $K = 4$, it is seen that the results are very similar for different average degrees. All in all, there is not much discrepancy between the probability distributions of different topologies of the small-world network with the same number of nodes.

According to these outcomes, we conclude that, by considering only the number of individuals regardless of their configuration, the topology does not play an important role in the average stationary distribution. The only examined topologies which have more discrepancies compared to the others are the cycle and scale-free graphs with the least number of edges. A plausible justification is that these two graphs have the least average degree among all the studied topologies.

We also demonstrate the mean frequency of type B individuals versus mutation probability in Fig. 9 for all topologies. When the mutation probability is increased, the frequency of individuals with a higher fitness decreases until the mutation probability reaches $\mu = 0.5$ where all the graphs meet at the mean frequency $\bar{n} = 0.5$.

V. THE MIXING TIME

The time needed for a Markov chain to get *close enough* to its stationary state is called the *mixing time* [14]. In order to obtain the mixing time, we need to define close enough more precisely. To this end, we introduce a measure named the total variation distance between two probability distributions on the state space of a Markov chain. Let α and β be two probability distributions on the state space S. The total variation distance $d(\alpha, \beta)$ between these two probability distributions is defined by

$$d(\alpha, \beta) = \frac{1}{2} \sum_{s \in S} |\alpha(s) - \beta(s)|. \tag{12}$$

According to this definition, one can measure the total variation between the probability distribution $p(s)$ of the state space at time t and the stationary distribution $\pi(s)$ as $d(t) = \frac{1}{2} \sum_{s \in S} |p(s) - \pi(s)|$. Since $d(t)$ tends to 0 when $t \longrightarrow \infty$, it makes sense to define the mixing time as below [14],

$$t_{\text{mix}}(\epsilon) = \min\{t | d(t) \leqslant \epsilon\}, \tag{13}$$

where $\epsilon > 0$ is a small parameter which determines the closeness of the probability distribution to the stationary distribution. A standard choice for ϵ which is widely used

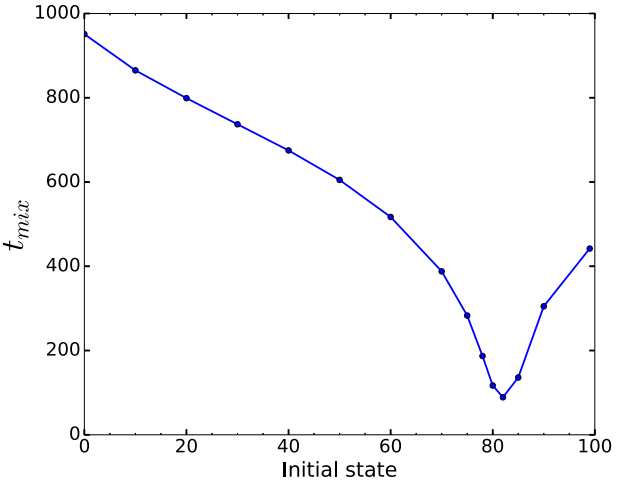

FIG. 10. The mixing time t_{mix} versus the initial state for the complete graph with $r = 2$, $\mu = 0.1$, and $N = 100$. The closer the initial state approaches the average number of B's, the sooner the dynamics converges to the stationary distribution.

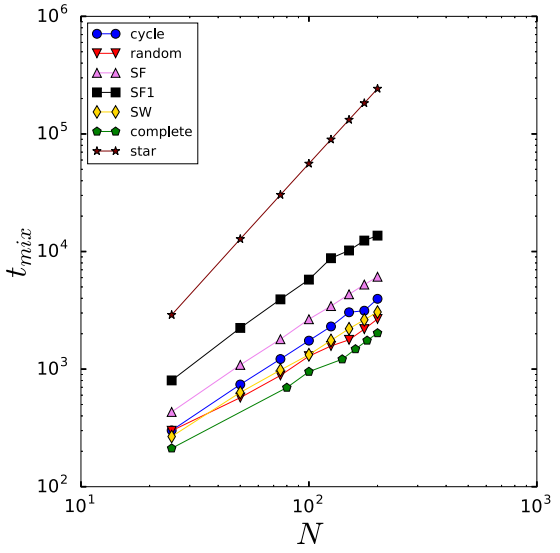

FIG. 11. Log-log plot of the mixing time t_{mix} versus the population size N for different topologies with $r = 2$ and $\mu = 0.1$. SF and SF1 denote the scale-free networks with $m > 1$ and $m = 1$, respectively, and SW denotes the small-world graph. The number of realizations for the SF and SF1 is taken to be 100.

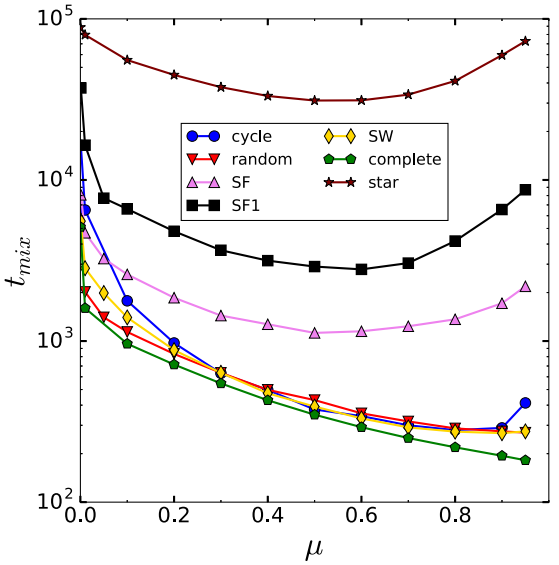

FIG. 12. Semilogarithmic plot of the mixing time t_{mix} versus the mutation probability μ for different topologies with $N = 100$ and $r = 2$. SF and SF1 denote the scale-free graphs with $m > 1$ and $m = 1$, respectively and SW denotes the small-world graph. The number of realizations for the SF and SF1 is taken to be 100.

by many authors is 1/4 [14,15]. It can be shown that smaller values of ϵ do not considerably affect $d(t)$. We approximate the mixing time for the complete, star, cycle, scale-free, small-world, and random graphs. For the complete graph, the mixing time is obtained for different initial states (Fig. 10). This figure shows that the closer the initial state approaches the average number of B's, the sooner the system finds its stationary distribution.

It is also interesting to see how the mixing time is related to the population size. Our observations show a power law relation between the mixing time and the population size $t_{\text{mix}} \sim N^{\alpha}$ (Fig. 11). The exponent α is closely related to the heterogeneity of the network (Table I).

This power law relation can be compared to the behavior of the average fixation time of the cycle and star graphs for fitness values greater than 1 [23], although the exponents are not the same.

Now we look at the effect of mutation probability on the mixing time. Figure 12 shows the mixing time as a function of mutation probability for different topologies. According to this figure, when heterogeneity increases, the mixing time increases as well. For the same values of parameters, the maximum mixing time corresponds to the star graph, whereas the minimum one belongs to the complete graph.

Clearly in homogeneous graphs, such as the complete, random, and small-world graphs, increasing the mutation probability leads to shorter mixing times. The reason lies in the behavior of $T^+ - T^-$ which determines the speed of approaching the system its stationary distribution. It is straightforward to measure the average value of $T^+ - T^-$ from the initial distribution to the stationary distribution of every mutation probability for the complete graph. It is seen that this quantity grows as the mutation probability increases. Hence, the mixing time decreases for greater mutation probabilities. It is seen that for the other topologies, such as the star and scale-free graphs that are more heterogeneous, the mixing time has a minimum in mutation probabilities less than 1. As we saw in Sec. III, approximating T^+ and T^- for the star graph is not possible because p_A and p_B are not determined exactly except in the steady state regime. Therefore, one cannot easily obtain $T^+ - T^-$ for the star graph. By comparing the mixing time versus mutation probability for homogeneous and heterogeneous topologies, one can infer that the behavior of the mixing time with respect to the mutation probability is related to heterogeneity or in other words locally starlike topologies. We do not have a rigorous explanation for this observation. It may also be related to other structural properties, such as the mean shortest path length and clustering coefficient. This can be the subject of further investigation.

In Fig. 13 we show the error bars for the mixing time over 100 realizations of BA graphs with $m = 1$ and $m = 2$. As is seen, the mixing time for different realizations of a BA network does not differ significantly. We also check how the initial configuration of a BA graph could affect the mixing time. We compare the mixing time of a BA graph when in the initial state the single B is located on the most connected node with the mixing time of that graph when in the initial state the single B is located on the less connected node. Intuitively we expect that when we start with the most connected node we reach the stationary distribution faster. Interestingly our results do

TABLE I. Exponent of the power law relation of the mixing time with respect to the network size N^{α}.

Topology	Complete	Random	Small-world	Cycle	SF	SF1	Star
α	1.07	1.04	1.16	1.23	1.26	1.4	2.13

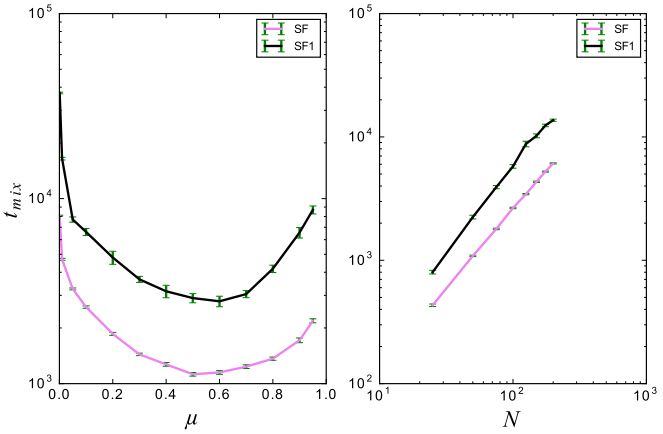

FIG. 13. The error bars for the mixing time t_{mix} over different realizations of BA graphs $N = 100$ and $r = 2$. SF and SF1 denote the scale-free graphs with $m > 1$ and $m = 1$, respectively. The number of realizations for the SF and SF1 is taken to be 100.

not fulfill this expectation, and the values of the mixing time obtained for both initial conditions are almost the same.

VI. SUMMARY AND CONCLUDING REMARKS

In this paper we studied the evolutionary dynamics in the finite structured populations with mutation. For the star and complete graphs, we could compute the exact form of the transition matrix. Using the transition matrix is an easy way to interpret the average behavior of the system in the stationary regime. For other topologies we were not able to obtain the transition matrix because of the variety of configurations that the distribution of species on the topologies have. These topologies include the scale-free graph constructed by the Barabasi-Albert model, the small-world graph constructed by the Watts-Strogatz model, the cycle graph, and the random graph. We looked at the time evolution of the number of each type of species. This way, we approximated the probability distribution at the stationary regime for each topology. We observed that the stationary distribution is almost independent of the network topology, but this is not the case for the mixing time.

The mixing time is the time that the system needs to reach its stationary distribution. The behavior of the mixing time with respect to the population size is closely related to the topology. As the heterogeneity increases from the random graph to the star graph, changing the population size has more effect on the mixing time. It is also seen that the more the heterogeneity is, the more time the system takes to reach its stationary distribution. For homogeneous topologies, such as the complete graph, as the mutation probability increases, it is expected that the system converges more rapidly to the stationary distribution. As a result, the absolute minimum of the mixing time is reached when the mutation probability is maximized. Interestingly, what we have obtained in heterogeneous graphs, such as the star graph, is different in a way that the minimum value of the mixing time in heterogeneous topologies occurs in a mutation probability less than one.

[1] S. Wright, Genetics **28**, 114 (1943).
[2] M. Slatkin, Evolution **35**, 477 (1981).
[3] M. A. Nowak, Nature (London) **359**, 826 (1992).
[4] M. A. Nowak, E. C. Tarnita, and T. Antal, Philos. Trans. R. Soc. B **365**, 19 (2010).
[5] E. Lieberman, C. Hauert, and M. A. Nowak, Nature (London) **433**, 312 (2005).
[6] F. C. Santos and J. M. Pacheco, Phys. Rev. Lett. **95**, 098104 (2005).
[7] P. D. Taylor, T. Day, and G. Wild, Nature (London) **447**, 469 (2007).
[8] L. Hindersin and A. Traulsen, J. R. Soc. Interface **11**, 20140606 (2014).
[9] B. Wu, C. S. Gokhale, L. Wang, and A. Traulsen, J. Math. Biol. **64**, 803 (2012).
[10] S. P. Otto and M. C. Whitlock, Genetics **146**, 723 (1997).
[11] Z. Patwa and L. M. Wahl, J. R. Soc. Interface **5**, 1279 (2008).
[12] N. H. Barton, Genet. Res. **62**, 149 (1993).
[13] A. McAvoy and C. Hauert, J. R. Soc. Interface **12**, 20150420 (2015).
[14] D. A. Levin and Y. Peters, *Markov Chains and Mixing Times* (American Mathematical Society, Providence, 2017).
[15] A. J. Black, A. Traulsen, and T. Galla, Phys. Rev. Lett. **109**, 028101 (2012).
[16] P. Catalán, J. M. Seoane, and M. A. Sanjuán, Commun. Nonlinear Sci. Numer. Simul. **25**, 66 (2015).
[17] C. E. Tarnita, T. Antal, and M. A. Nowak, J. Theor. Biol. **261**, 50 (2009).
[18] A. Traulsen, J. C. Claussen, and C. Hauert, Phys. Rev. E **85**, 041901 (2012).
[19] A. Traulsen, C. Hauert, H. De Silva, M. A. Nowak, and K. Sigmund, Proc. Natl. Acad. Sci. U.S.A. **106**, 709 (2009).
[20] T. Antal, M. A. Nowak, and A. Traulsen, J. Theor. Biol. **257**, 340 (2009).
[21] B. Allen and C. Tarnita, J. Math. Biol. **68**, 109 (2012).
[22] N. G. Van Kampen, *Stochastic Processes in Physics and Chemistry* (Elsevier, Amsterdam, 1992).
[23] M. Askari and K. A. Samani, Phys. Rev. E **92**, 042707 (2015).
[24] M. Broom and J. Rychtář, Proc. R. Soc. A **464**, 2609 (2008).
[25] W. Maciejewski, F. Fu, and C. Hauert, PLoS Comput. Biol. **10**, e1003567 (2014).