BioSystems 129 (2015) 25–35

Contents lists available at ScienceDirect

BioSystems

journal homepage: www.elsevier.com/locate/biosystems

Fast and asymptotic computation of the fixation probability for Moran processes on graphs

F. Alcalde Cuesta [a], P. González Sequeiros [a,b], Á. Lozano Rojo [a,c,d,*]

[a] GeoDynApp – ECSING Group, Spain
[b] Dpto. de Didáctica das Ciencias Experimentais, Facultade de Formación do Profesorado, Universidade de Santiago de Compostela, Avda. Ramón Ferreiro, 10, E-27002 Lugo, Spain
[c] Centro Universitario de la Defensa de Zaragoza, Academia General Militar, Ctra. Huesca s/n, E-50090 Zaragoza, Spain
[d] Instituto Universitario de Matemáticas y Aplicaciones, Universidad de Zaragoza, Spain

ARTICLE INFO

Article history:
Received 30 September 2013
Received in revised form 19 January 2015
Accepted 23 January 2015
Available online 24 January 2015

MSC:
05C81
60J20
92D15

Keywords:
Evolutionary dynamics
Markov chain
Monte Carlo methods
Fixation probability
Expected fixation time
Star and bipartite graphs

ABSTRACT

Evolutionary dynamics has been classically studied for homogeneous populations, but now there is a growing interest in the non-homogeneous case. One of the most important models has been proposed in Lieberman et al. (2005), adapting to a weighted directed graph the process described in Moran (1958). The Markov chain associated with the graph can be modified by erasing all non-trivial loops in its state space, obtaining the so-called Embedded Markov chain (EMC). The fixation probability remains unchanged, but the expected time to absorption (fixation or extinction) is reduced. In this paper, we shall use this idea to compute asymptotically the average fixation probability for complete bipartite graphs $K_{n,m}$. To this end, we firstly review some recent results on evolutionary dynamics on graphs trying to clarify some points. We also revisit the 'Star Theorem' proved in Lieberman et al. (2005) for the star graphs $K_{1,m}$. Theoretically, EMC techniques allow fast computation of the fixation probability, but in practice this is not always true. Thus, in the last part of the paper, we compare this algorithm with the standard Monte Carlo method for some kind of complex networks.

1. Introduction and motivation

Population genetics studies the genetic composition of biological populations, and the changes in this composition that result from the action of four different processes: *natural selection*, *random drift*, *mutation* and *migration*. The *modern evolutionary synthesis* combines Darwin's thesis on natural selection and Mendel's theory of inheritance. According to this synthesis, the central object of study in evolutionary dynamics is the frequency distribution of the alternative forms (*allele*) that a hereditary unit (*gene*) can take in a population evolving under these forces.

Many mathematical models have been proposed to understand evolutionary process. Introduced in Moran (1958), the *Moran model* describes the change of gene frequency by random drift on a population of finite fixed size. This model has many variants, but we assume for simplicity that involved organisms are haploids with only two possible alleles a and A for a given locus. Suppose there is a single individual with a copy of the allele A. At each unit of time, one individual is chosen at random for reproduction and its clonal offspring replaces another individual chosen at random to die. To model natural selection, individuals with the advantageous allele A are assumed to have relative fitness $r > 1$ as compared with those with allele a of fitness 1.

Evolutionary dynamics has been classically studied for homogeneous populations, but it is a natural question to ask how non-homogeneous structures affect this dynamics. In Lieberman et al. (2005), a generalisation of the Moran process was introduced by arranging the population on a directed graph, see also Nowak (2006), Shakarian et al. (2012) and Shakarian et al. (2013). In this model, each vertex represents an individual in the population, and the offspring of each individual only replace direct successors, i.e.

* Corresponding author at: Centro Universitario de la Defensa de Zaragoza, Academia General Militar, Ctra. Huesca s/n, E-50090 Zaragoza, Spain. Tel.: +34 976739856.
E-mail addresses: fernando.alcaldecuesta@gmail.com (F. Alcalde Cuesta), pablo.gonzalez.sequeiros@usc.es (P. González Sequeiros), alvarolozano@unizar.es (Á. Lozano Rojo).

http://dx.doi.org/10.1016/j.biosystems.2015.01.007
0303-2647/

end-points of edges with origin in this vertex. The *fitness* of an individual represents again its reproductive rate which determines how often offspring takes over its neighbour vertices, although these vertices do not have to be replaced in an equiprobable way. The evolutionary process is described by the choice of stochastic matrix $W = (w_{ij})$ where w_{ij} denotes the probability that individual i places its offspring into vertex j. In fact, further generalisations can be considered assuming that the probability above is proportional to the product of a weight w_{ij} and the fitness of the individual i. In this case, W does not need to be stochastic, but non-negative. The *fixation probability* of the single individual i is the probability that the progeny of i takes over the whole population. Several interesting and important results are shown in Lieberman et al. (2005).

- Different graph structures support different dynamical behaviours amplifying or suppressing the reproductive advantage of *mutant* individuals (with the advantageous allele A) over the *resident* individuals (with the disadvantageous allele a).
- An evolutionary process on a weighted directed graph (G, W) is *equivalent to a Moran process* (i.e. there is a fixation probability well-defined for any individual, which coincides with the fixation probability in a homogeneous population) if and only if (G, W) is *weight-balanced*, i.e. for any vertex i the sum of the weights of entering edges $w_-(i) = \sum_{j=1}^{N} w_{ji}$ and that of leaving edges $w_+(i) = \sum_{j=1}^{N} w_{ij}$ are equal. This is called the *Circulation Theorem* in Lieberman et al. (2005) and Nowak (2006).

As in the classical setting, mutant individuals will either become extinct or take over the whole population, reaching one of the two *absorption* states (*extinction* or *fixation*), when a finite population is arranged on an undirected graph or on a *strongly connected* directed graph (where two different vertices are always connected by an edge-path). Even in the first case, the fixation probability depends usually on the starting position of the mutant. The effect of this initial placement on mutant spread has been discussed in Broom et al. (2009, 2011). In the present paper, we start by summarising some fundamental ideas and results on evolutionary dynamics on graphs. In this context, most work involves computing the (average) fixation probability, but doing so in general requires solving a system of 2^N linear equations. In the example of the *star graph* described in Lieberman et al. (2005), like for other examples described in Broom and Rychtář (2008), Hadjichrysanthou (2012) and Lieberman et al. (2005), a high degree of symmetry reduces the size of the linear system to a set of $2N$ equations, which becomes asymptotically equivalent to a linear system with N equations. We revisit this example that will be useful in addressing the study of complete bipartite graphs. Another research direction has been to use Monte Carlo techniques to implement numerical simulations, but often limited to small graphs (Broom et al., 2009), small random modification of regular graphs (Rychtář and Stadler, 2008) or graphs evolving under random drift (Shakarian and Roos, 2011).

Our aim is to show how to modify the stochastic process associated with a weighted directed graph to simplify the evolutionary process both analytically and numerically. Recall that an evolutionary process on a weighted directed graph (G, W) with N vertices is a Markov chain with 2^N states representing the vertex sets inhabited by mutant individuals and transition matrix P derived from W. The non-zero entries of P can be used to see the state space as a (weighted) directed graph. We call *loop-erasing* the loop suppression in this graph $\mathcal{S}$, avoiding to remain in the same state in two consecutive steps and providing the *Embedded Markov chain* (EMC) associated to the process. This technique is used here to compute asymptotically the average fixation probability for complete bipartite graphs, generalising the Star Theorem of Lieberman et al. (2005), see also Banerjee (2012), Houchmandzadeh and Vallade (2013) and Tan and Lu (2014). Expected time to *absorption* (fixation or extinction) of this EMC has been studied for circular, complete and star graphs in Hadjichrysanthou (2012). Here we compare numerically the expected absorption time of both chains on some kinds of complex networks. This method can be combined with other approximation methods (like the FPRAS method described in Díaz et al. (2012) for undirected graphs) to obtain a fast approximation scheme.

The paper is organised as follows. In Section 2, we review the Moran model for homogeneous and non-homogeneous populations. In Section 3, we revisit the Star Theorem giving an alternative proof of it. In Section 4, we briefly explain the machinery of the loop-erasing method and we use this idea to describe the asymptotic behaviour of the fixation probability on the complete bipartite graphs family. At the end, in Section 5, we include some numerical experiments to evaluate the performance of the Monte Carlo method on both the standard and the loop-erased chains for different complex networks.

2. Review of Moran process

The *Moran process* models random drift and natural selection for finite homogeneous populations (Moran, 1958). As indicated before, we consider a haploid population of N individuals having only two possible alleles a and A for a given locus. At the beginning, all individuals have the allele a. Then one resident individual is chosen at random and replaced by a mutant having the neutral or advantageous allele A. At successive steps, one randomly chosen individual replicates with probability proportional to the fitness $r \geq 1$ and its offspring replaces one individual randomly chosen to be eliminated, see Fig. 1. Since the future state depends only on the present state, the Moran process is a Markov chain X_n with state space $\mathcal{S} = \{0, \ldots, N\}$ representing the number of mutant individuals with the allele A at the time step n. This is a stationary process because the probability $P_{i,j} = \mathbb{P}[X_{n+1} = j | X_n = i]$ to pass from i to j mutant individuals does not depend on the time n. In fact, the number of mutant individuals can change at most by one at each step and hence the *transition matrix* $P = (P_{i,j})$ is a tridiagonal matrix where $P_{i,j} = 0$ if $j \neq i-1, i, i+1$. As $P_{0,0} = P_{N,N} = 1$, the states $i = 0$ and $i = N$ are *absorbing*, whereas the other states are *transient*.

The *fixation probability* of i mutant individuals

$$\Phi_i = \Phi_i(r) = \mathbb{P}[\exists n \geq 0 : X_n = N | X_0 = i]$$

is the solution of the system of linear equations:

$$\begin{aligned} \Phi_0 &= 0 \\ \Phi_i &= P_{i,i-1}\Phi_{i-1} + P_{i,i}\Phi_i + P_{i,i+1}\Phi_{i+1} \\ \Phi_N &= 1 \end{aligned} \tag{1}$$

where $P_{i,i} = 1 - P_{i,i-1} - P_{i,i+1}$. In particular, the probability of a single mutant to reach fixation $\Phi_1 = \Phi_1(r)$ is usually referred to as the *fixation probability* in short. To solve (1), we define $y_i = \Phi_i - \Phi_{i-1}$ which verifies $\sum_{i=1}^{N} y_i = \Phi_N - \Phi_0 = 1$. Then, dividing each side of (1) by $P_{i,i+1}$, we have $y_{i+1} = \gamma_i y_i$ where $\gamma_i = P_{i,i-1}/P_{i,i+1}$ is the *death-birth rate*. It follows $y_i = \Phi_1 \prod_{j=1}^{i-1} \gamma_j$, and hence the fixation probability is

$$\Phi_1 = \frac{1}{1 + \sum_{i=1}^{N-1} \prod_{j=1}^{i} \gamma_j} \tag{2}$$

see Karlin and Taylor (1975), Taylor et al. (2004) and Nowak et al. (2004).

If neither of alleles a and A is advantageous reproductively, the *random drift* phenomenon is modelled by the Moran process with fitness $r = 1$, and (2) becomes $\Phi_1 = 1/N$. On the contrary, if mutant