

# An accurate database of the fixation probabilities for all undirected graphs of order 10 or less

Fernando Alcalde Cuesta<sup>1</sup>, Pablo González Sequeiros<sup>1</sup>, Álvaro Lozano Rojo<sup>2,3</sup>,  
and Rubén Vígara Benito<sup>2,3</sup>

<sup>1</sup> University of Santiago de Compostela, Spain.

{fernando.alcalde, pablo.gonzalez.sequeiros}@usc.es

<sup>2</sup> Centro Universitario de la Defensa Zaragoza, Spain.

{alozano, rvigara}@usc.es

<sup>3</sup> IUMA - University of Zaragoza. Spain.

**Abstract.** We present an extremely precise database of the fixation probabilities of mutant individuals in a non-homogeneous population which are spatially arranged on a small graph. We explore what features of a graph increase the chances of a beneficial allele of a gene to spread over a structured population.

## 1 Introduction

The fixation probability is a fundamental concept in evolutionary dynamics, representing the probability that a gene spreads over a whole population. An interesting model to study both neutral drift and natural selection on homogeneous population was introduced by P. A. P. Moran in [19]. The Moran process is a Markov chain whose states are the number of individuals with a mutant allele  $A$  of a gene starting from an initial population of  $N$  residents having the same allele  $a$  of the same gene at some locus. At each step of time, one individual is selected for reproduction with probability proportional to its *relative fitness* with respect to the resident ones,  $r$  or 1 depending on whether it is mutant or resident. Then another individual is randomly chosen (with uniform probability) to be replaced by an identical offspring of the first individual. This Markov chain has two absorbing states corresponding to the *extinction* and *fixation* of the mutant allele  $A$  (where all the individuals have the allele  $a$  or  $A$  respectively). Starting with a single mutant individual in an initial population of  $N$  individuals, the fixation probability is

$$\Phi = \frac{r^{N-1}}{r^{N-1} + r^{N-2} + \dots + 1}, \quad (1)$$

for any relative fitness  $r$ .

E. Lieberman, C. Hauert and M. A. Nowak generalized this process to structured populations [21]. Let  $G = (V, E)$  be an undirected connected graph with

vertex set  $V = \{0, 1, \dots, N - 1\}$  with no loops or multiple edges. Denote by  $d_i$  the degree of the vertex  $i$ . The Moran process on  $G$  with fitness  $r$  is the Markov chain obtained as follows. Like for a homogeneous population, we start with a population of  $N$  resident individuals occupying the vertex set  $V$ . Afterward, one single vertex  $i_0$  is chosen to become occupied by a mutant. At successive steps one vertex  $i$  is selected at random with probability  $r_i / (r m + N - m)$ , where  $r_i$  is the relative fitness of the individual occupying  $i$  ( $r$  or  $1$ ), and  $m$  is the number of vertices occupied by mutants in that moment. Next, a neighbor of  $i$ , randomly chosen with uniform probability, is replaced by an identical offspring of  $i$ . Now, the fixation probability depends on where the first mutant is placed inside the graph [8], and hence the *average fixation probability* on  $G$ , denoted  $\Phi_G^r$ , is the main concept in the theory of evolutionary graphs. These kind of models have had impact not only in evolutionary genetics but in other areas like invasion dynamics, epidemics of disease, tumor growth or economics and management [2, 11, 16, 20, 22]. Actually, if the spreading of favorable innovations can be enhanced by network structures amplifying selection [26], as counterpart, we can find structural properties that increase the robustness of a complex network against invasion [2].

In the case of general graphs there is no closed form for the fixation probability. Some techniques have been proposed to compute this quantity [1, 6, 9, 14, 18] and there has been some calculations on small graphs [1, 7, 12, 25]. However, there is no accurate available data for degrees less than or equal to 8, and there is no data available at all for degrees greater than 8.

Here we present an accurate database of the fixation probabilities for all undirected graphs with 10 or less vertices. This dataset could allow to find families of graph structures with interesting evolutionary properties, as already has been done in [3]. Moreover, the database has been enriched with some graph invariants which have been related to the fixation probability, see [2] and the references therein. In this way, exploring these data, it would be possible to shed light on the structural properties of graphs increasing or decreasing their fixation probabilities and the transitions among different evolutionary types. This is a particularly interesting property for biological networks like brain networks or PPI interactomes, as well as for technological ones. Finally, due to its precision, it also could be used as a testing dataset for new methods or computing libraries.

## 2 The database

### 2.1 The computation

The main steps of the computation process are the following:

*Generation of the graphs.* The generation of the edge lists of all undirected graphs (up to isomorphism) with 10 vertices or less was done with Sage [17, 23]. Since this is a relatively short computation we made no attempt to parallelize this process. Afterwards we have a binary file with the 11,989,763 connected graphs of order  $\leq 10$  (we dropped the trivial graph of one vertex).

*Description of the fixation probability.* Consider a connected undirected graph  $G = (V, E)$  of order  $N \leq 10$ , and fix a relative fitness  $r > 0$  for the invader mutants. The Markov chain described in the introduction is formalized as follows. The set of states  $\mathcal{S}_G$  of the chain is the power set of  $V$ , where each set  $S \in \mathcal{S}_G$  contains the nodes occupied by mutant individuals. The transition probabilities between  $S, S' \in \mathcal{S}_G$  are given by

$$P_{S,S'}^r = \begin{cases} \frac{r \sum_{\{i \in S | (i,j) \in E\}} \frac{1}{d_i}}{r \#S + N - \#S} & \text{if } S' \setminus S = \{j\}, \\ \frac{\sum_{\{i \in V \setminus S | (i,j) \in E\}} \frac{1}{d_i}}{r \#S + N - \#S} & \text{if } S \setminus S' = \{j\}, \\ \frac{r \sum_{\{(i,j) \in E | i,j \in S\}} \frac{1}{d_i} + \sum_{\{(i,j) \in E | i,j \notin S\}} \frac{1}{d_i}}{r \#S + N - \#S} & \text{if } S = S', \\ 0 & \text{otherwise.} \end{cases} \quad (2)$$

These values can be arranged in a  $2^N \times 2^N$  matrix  $\mathbf{P}_G^r = (P_{S,S'}^r)$ . On the other hand, the fixation probabilities  $\phi_S^r$  associated to all sets  $S \in \mathcal{S}_G$  inhabited by mutant individuals are determined by the system of linear equations  $\phi_S^r = \sum_{S' \in \mathcal{S}_G} P_{S,S'}^r \phi_{S'}^r$ , with the boundary conditions  $\phi_\emptyset^r = 0$  and  $\phi_V^r = 1$ . Equivalently

$$\mathbf{P}_G^r \cdot \begin{pmatrix} 0 \\ \phi_S^r \\ \vdots \\ \phi_{S'}^r \\ 1 \end{pmatrix} = \begin{pmatrix} 1 & | & 0 & \cdots & 0 & | & 0 \\ P_{S,\emptyset}^r & & & & & & P_{S,V}^r \\ \vdots & & \widehat{\mathbf{P}}_G^r & & & & \vdots \\ P_{S',\emptyset}^r & & & & & & P_{S',V}^r \\ 0 & | & 0 & \cdots & 0 & | & 1 \end{pmatrix} \cdot \begin{pmatrix} 0 \\ \phi_S^r \\ \vdots \\ \phi_{S'}^r \\ 1 \end{pmatrix} = \begin{pmatrix} 0 \\ \phi_S^r \\ \vdots \\ \phi_{S'}^r \\ 1 \end{pmatrix}. \quad (3)$$

With this notation, the (average) fixation probability is then

$$\Phi_G^r = \frac{1}{N} \sum_{i \in V} \phi_{\{i\}}^r. \quad (4)$$

To compute this value, we need to solve (3), which can be written as a linear system

$$(\mathbf{I} - \widehat{\mathbf{P}}_G^r) \cdot \boldsymbol{\phi} = \begin{pmatrix} P_{S,V}^r \\ \vdots \\ P_{S',V}^r \end{pmatrix}$$

where  $\mathbf{I}$  is the identity matrix of size  $2^N - 2$ . Looking at (2) it is clear that it is possible to multiply each equation (associated to a state  $S$ ) by the *reproductive weight*  $W_S^r = r \#S + N - \#S$  of  $S$  obtaining the final equation  $\mathbf{Q}_G^r \cdot \boldsymbol{\phi} = \mathbf{b}_G^r$ , where

$$\mathbf{Q}_G^r = \begin{pmatrix} W_S^r & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & W_{S'}^r \end{pmatrix} \cdot (\mathbf{I} - \widehat{\mathbf{P}}_G^r), \quad \mathbf{b}_G^r = \begin{pmatrix} W_S^r P_{S,V}^r \\ \vdots \\ W_{S'}^r P_{S',V}^r \end{pmatrix}.$$

By construction, the coordinates of  $\mathbf{Q}_G^r$  and  $\mathbf{b}_G^r$  are degree 1 polynomials on  $r$  with rational coefficients.

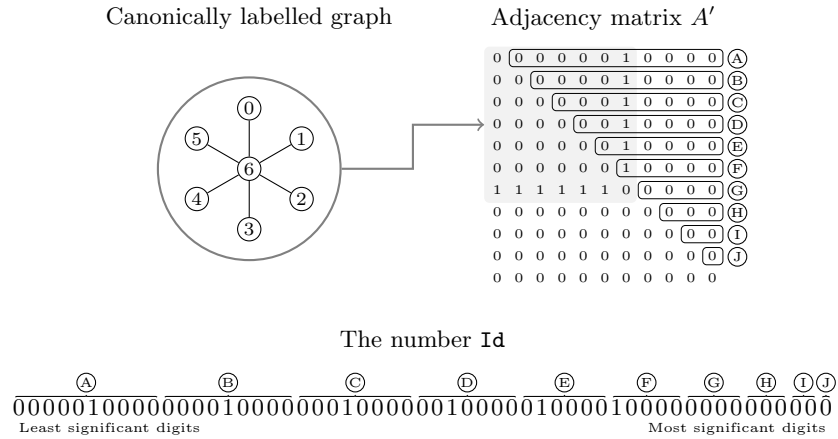
*Computation of the fixation probability.* The computation of  $\Phi_G^r$  according to (4) was written in the C programming language. It runs as follows: consider a connected graph  $G$  of order  $N \leq 10$ . Each element in  $\mathcal{S}_G$  can be represented by a number of  $N$  bits in length, that is an integer between 0 and  $2^N - 1$ . Given one of those states  $S$ , identified with the corresponding number, there are at most  $N+1$  possible transitions: a change in any of the bits of the number, or the number itself if there is no change. Hence,  $\mathbf{Q}_G^r$  has at most  $2^N \times (N+1)$  non-zero entries. To compute the row associated to the state  $S$  of this matrix, all the possible elections for reproduction are performed and the results are accumulated in the correct positions. Since coordinates of  $\mathbf{Q}_G^r$  and  $\mathbf{b}_G^r$  are of the form  $ar + b$  with  $a, b \in \mathbb{Q}$ , they can be easily represented as a couple of pairs of 64 bits integers. Therefore, we can exactly compute the fundamental matrix  $\mathbf{Q}_G^r$  and the vector  $\mathbf{b}_G^r$ .

Finally, for each value of  $r \in \{0.25, 0.5, \dots, 10\}$ , we should find the solution of  $\mathbf{Q}_G^r \cdot \phi = \mathbf{b}_G^r$  to compute  $\Phi_G^r$ , which is the mean of the entrances of  $\phi$  corresponding to states with only one mutant (that is, states with a single bit set). To do so, we construct dense matrices evaluating  $\mathbf{Q}_G^r$  and  $\mathbf{b}_G^r$  for the given fitness  $r$  using 64 bit floating point numbers. Note that we construct them dense instead of sparse because of the used linear solver.

To guarantee a high relative precision of the solution that does not depend on the condition number of the matrix  $\mathbf{Q}_G^r$ , we use a special *LDU* factorization algorithm for  $M$ -matrices due to Barreras and Peña [5]. This algorithm is slower than iterative algorithms since it asymptotically has the complexity of the Gaussian elimination. However, it is reasonably quick for a single matrix of size  $1022 \times 1022$ , although the amount of graphs and possible values of  $r$  forced the parallelization of the computation. This step was trivial since the computations for different graphs are independent. The code is available at [24].

The actual computation was done in the Supercomputer FinisTerra2 located at CESGA (Spain) using 1024 cores of Haswell 2680v3 CPUs for almost 3 days.

*Building the database.* Finally, the database of fixation probabilities already computed was enriched with some graph properties which have been related with its evolutionary behavior. For each undirected graph  $G = (V, E)$  we considered the *order*, the number of vertices  $N = |V|$ , the *size*, the number of edges  $|E|$ , and some statistics related with the degree distribution. We also computed some global scale measures like the *diameter*  $\Delta = \max\{d(i, j)\}_{i, j \in V}$  and the *average path length*  $L = \sum_{i, j \in V} \frac{d(i, j)}{N(N-1)}$ , where  $d(i, j)$  is the length of the shortest path joining the vertices  $i$  and  $j$ . We also added other measures of ‘small-worldness’ like the *clustering coefficient*  $C = \frac{1}{N} \sum_{i \in V} \frac{2|E_i|}{d_i(d_i-1)}$ , with  $G_i = (V_i, E_i)$  the subgraph of neighbors of  $i$  with the edges between them [29], and the *transitivity*  $T$ , which is the ratio of the number of complete subgraphs of order 3 over the number of connected subgraphs of order 3 [28]. Additionally, we added the *heat heterogeneity* of the graph, defined as the variance of the temperature distribution, and the *temperature entropy* which is the entropy of that distribution [26]. Recall that the *temperature* of the vertex  $i$  is defined as  $T_i = \frac{1}{d_i} \sum_{j \sim i} 1/d_j$  where



**Fig. 1.** Generating the Id of a graph.

$d_i$  denotes the degree of  $i$  [27]. We completed this property set with three qualitative variables: whether if the graph is a *tree*, *bipartite* and/or *biconnected*. These quantities were computed using the NetworkX software [10].

On the other hand, each graph is identified with a unique 64 bits unsigned integer (the Id) which contains the adjacency matrix of the graph as follows (see Figure 1): Consider a graph  $G$  with  $N \leq 11$  vertices. After canonically labeling it (we used the bliss software [15]), consider its adjacency matrix  $A$ . Extend this matrix to the right and below as  $A' = \begin{pmatrix} A & 0 \\ 0 & 0 \end{pmatrix}$  to make it  $11 \times 11$ . In this way we get the adjacency matrix of a graph of order 11 such that the connected component of the first vertex agrees with  $G$ , and the rest of connected components are singletons. Since  $A'$  is symmetric and the diagonal is always null, all the relevant information belongs to the off-diagonal upper part of  $A'$ . Ordering those bits from left to right and from top to bottom we build a number, the Id, with at most 55 bits set. Obviously, the Id constructed in this way is unique for each (isomorphism class of a) graph of order  $\leq 11$ . The step of extending the matrix is required in order to guarantee this condition. In any case, it is trivial to recover the edge list from the Id without previous knowledge of the order or size of the graph (see <https://bitbucket.org/snippets/geodynapp/AppyX> for C and Python implementations).

## 2.2 On the accuracy

Even when the accuracy of the database is granted by the method used to solve the linear problem, we checked the result with the known closed forms for average fixation probabilities. As it has been pointed out the fixation probability for the complete graph  $K_N$  is given in (1) for any  $r > 0$ . Moreover, the Isothermal Theorem [21] implies that constant degree graphs have the same fixation probability as the complete one of the same order. Monk, Green and Paulin devised

**Table 1.** Estimated maximal absolute and relative errors for the known values of the fixation probability.

<b>For <math>r = 1</math></b>		<b>For <math>K_{n,m}</math> with <math>n, m \neq 1</math></b>	
maximal absolute error	$2.914 \cdot 10^{-16}$	maximal absolute error	$1.665 \cdot 10^{-15}$
maximal relative error	$2.887 \cdot 10^{-15}$	maximal relative error	$1.998 \cdot 10^{-15}$
<b>For Isothermal graphs</b>		<b>For <math>K_{1,n}</math></b>	
maximal absolute error	$1.887 \cdot 10^{-15}$	maximal absolute error	$2.220 \cdot 10^{-15}$
maximal relative error	$2.109 \cdot 10^{-15}$	maximal relative error	$2.665 \cdot 10^{-15}$

a technique to compute closed forms of the fixation probability of highly symmetric graphs [18] (see also [27] for star graphs  $K_{1,n}$ ). And finally, in the case of neutral drift ( $r = 1$ ) the fixation probability for a graph of order  $N$  is  $1/N$  [25]. In those cases it is possible to estimate the absolute and relative errors. Table 1 contains the maximal observed errors.

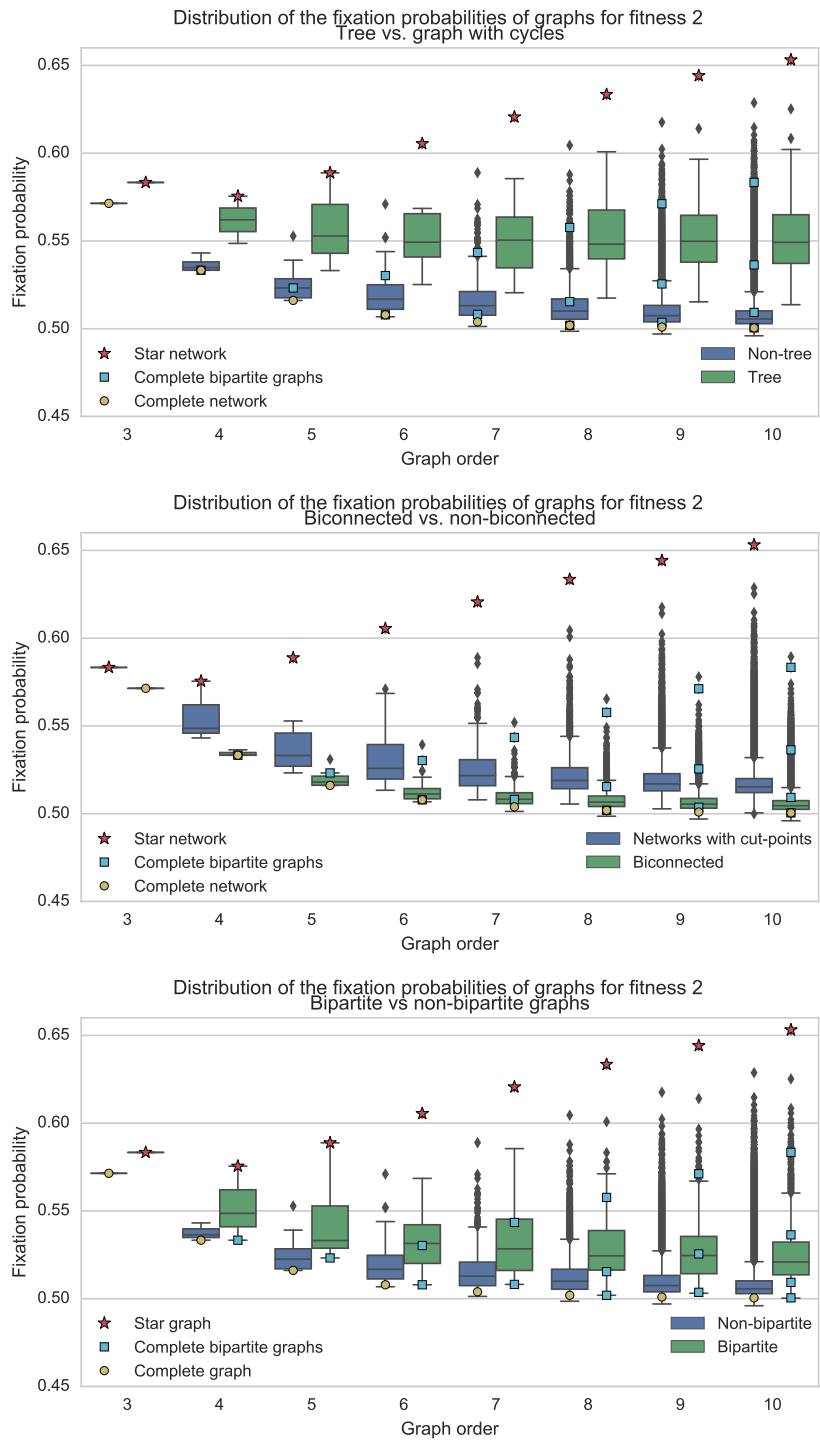
### 2.3 The database file

The database is presented as a single HDF5 file [13] available at [4]. The root group has a dataset called **FP** which contains the data computed above. The HDF5 format enables us to save column names and data types for each column (composite datatype in the HDF5 jargon). Fractional data is saved as IEEE-754 little-endian 64 bits numbers. Integer are saved with the smallest possible number of bytes, hence, all are one byte little-endian unsigned integers except the **Id** which uses 64 bits. The dataset **FP** contains the 58 columns described in Table 2. Notice that the case random drift  $r = 1$  has been dropped from the dataset since it is known [25].

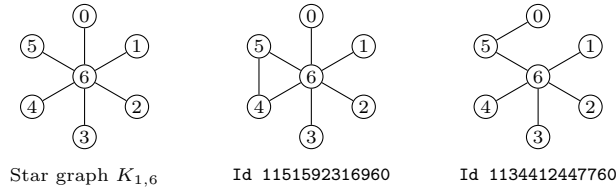
## 3 Results

A careful analysis of this dataset could shed some light to the factors that increase or decrease the fixation probabilities of advantageous mutations in structured population. For the rest of the section we fix a fitness of  $r = 2$ . By fixing a different fitness we would obtain similar results.

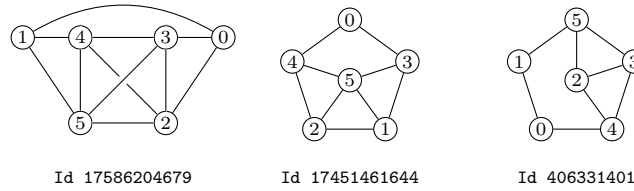
*Qualitative variables.* Figure 2 shows the effect in the distribution of the fixation probability of different qualitative properties of graphs. The data is disaggregated by order of the graph. It is remarkable that trees tend to be amplifiers of selection at any order, that is, they have higher fixation probability than the complete graph  $K_N$  with the same order  $N$ . In fact, in each order, all the trees are above the 75% of the non-tree graphs. Moreover, the star graph  $K_{1,n}$  is maximal in its order  $N = n + 1$ . In the same direction, graphs with cut-points (removing this vertex disconnects the graph) tend to have higher fixation probabilities.



**Fig. 2.** Qualitative properties of the graphs and the fixation probability.



**Fig. 3.** A family of amplifiers of selection ordered from left to right by  $\Phi_G^2$ .



**Fig. 4.** The suppressors of selection of order 6 for  $r = 2$ .

Probably, these cut-points act as *defenses* of the advantageous allele during its early life, that is, in the early steps of the invasion. All nodes of a tree are cut-points except the leafs. The extremal case is again the star graph, where the center acts as the defense of the mutant alleles located at the periphery [2]. Something similar happens with the bipartite ones.

It is also notable the distance between  $K_{1,n}$  and the rest of the graphs. This distance becomes smaller as  $n$  increases. Moreover, for each order greater than 6 two isolated graphs appear between  $K_{1,n}$  and the rest. They are more visible in the figure about biconnected graphs but can be spotted in the other two plots. They correspond to small modifications to  $K_{1,n}$  where an edge is added between two peripheral vertices which could be or not be disconnected from the central one, see Figure 3.

There are also some suppressors of selection, that is, those with fixation probability (strictly) less than the one of  $K_N$  for  $r = 2$ . For order 6 there are 3 of these graphs (see figure 4). This quantity increases rapidly reaching hundreds of thousands for order 10. In [3] a family of graphs obtained from this data has been proved to be a *global suppressor of selection*, i.e., the fixation probability of each of them is below the fixation probability of the complete graph of the same order for any fitness  $r > 1$ . The discovery of this family is a first step towards finding structural properties increasing the robustness of non homogeneous populations against mutation.

*Quantitative variables.* Among the computed quantitative variables the least interesting one is the temperature entropy [27] which is constant for graphs with the same order. Figure 5 shows the effects of the most interesting quantitative variables on the fixation probability of the graphs of order 10. Similar pictures could be obtained for other orders. It is clear that one by one only the heat



heterogeneity [26] explains effect of the population structure in the spread of the mutants, showing a high correlation.

## 4 Conclusions

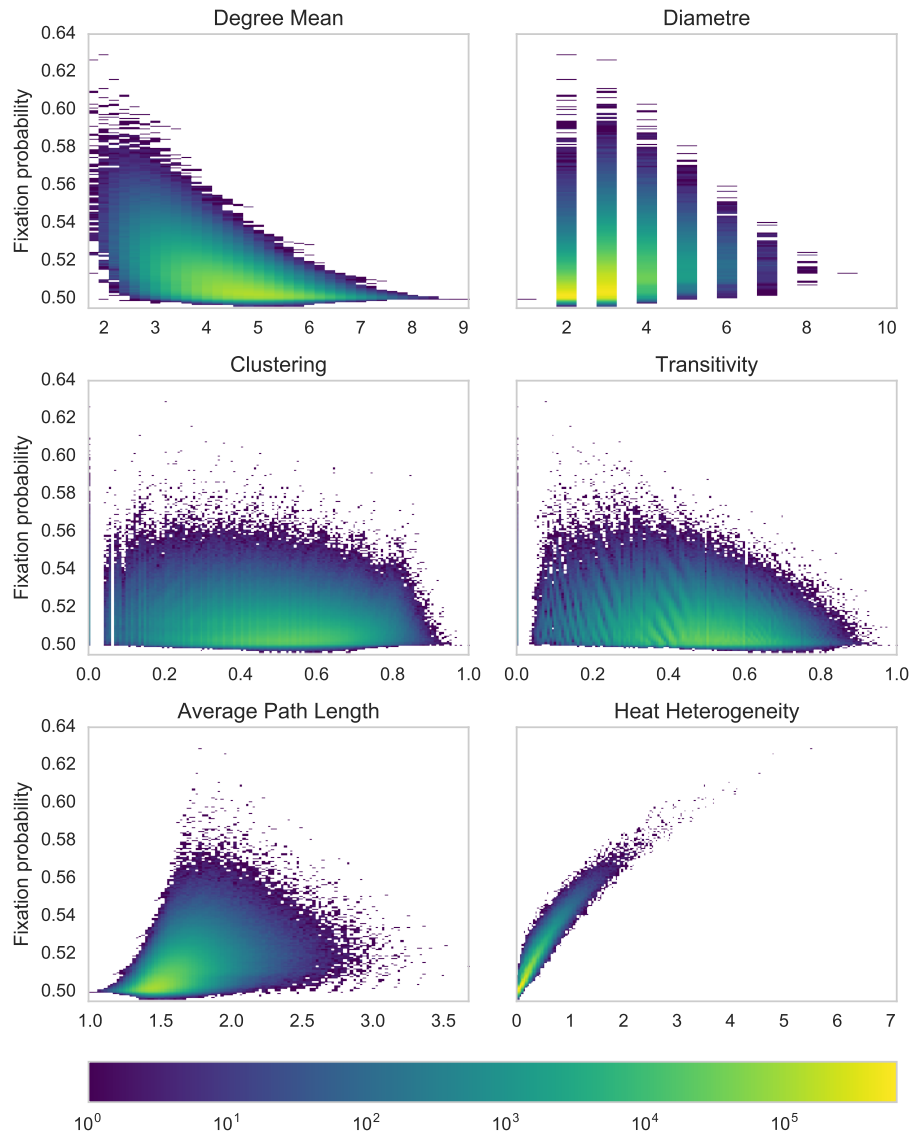
We present an extremely precise database of the fixation probabilities on small graphs of interest in research areas like invasion dynamics, epidemics of disease, tumor growth or economics and management [2, 11, 16, 20, 22]. Other authors have explored the evolutionary dynamics of small graphs [7, 12], but this is the first systematic computation for a large family of small graphs, namely all undirected graphs with 10 or less vertices. Exploring this dataset, evolutionary biology researchers could shed some light on the factors and structures affecting the spread of beneficial alleles in larger populations arranged on graphs. Deep statistical techniques are still needed to understand how to design population arrangements with a desired dynamical behavior and this dataset represents the first step in that direction.

## 5 Acknowledgments

We gratefully thank CESGA for providing access to the FinisTerra2 super-computer. FA, PG and ÁL are supported by Spanish Ministry of Economy and Competitiveness and European Social Fund (Grant MTM2013-46337-C2-2-P). ÁL and RV are supported by DGA and European Social Fund (Grant E15 Geometría) and CUDZ (Grant CUD 2015-10). RV is supported by Spanish Ministry of Economy and Competitiveness and European Social Fund (Grant MTM2013-45710-C2).

## References

1. F. Alcalde, P. G. Sequeiros, Á. Lozano. Fast and asymptotic computation of the fixation probability for Moran processes on graphs. *BioSystems* **129** (2015) 25–35.
2. F. Alcalde, P. G. Sequeiros, Á. Lozano. Exploring the topological sources of robustness against invasion in biological and technological networks. *Scientific Reports* **6** (2016), 20666.
3. F. Alcalde, P. G. Sequeiros, Á. Lozano. Suppressors of selection. arXiv 1607.04469.
4. F. Alcalde, P. G. Sequeiros, Á. Lozano, R. Vígara, Data of “An accurate database of the fixation probabilities for all undirected graphs of order 10 or less”, *Mendeley Data*, v2, 2017. <http://dx.doi.org/10.17632/587bnf6mt3.2>
5. A. Barreras, J. M. Peña, Accurate and efficient *LDU* decomposition of diagonally dominant *M*-matrices, *Electronic J. Linear Algebra* **24** (2012/13), 152–167.
6. M. Broom, J. Rychtář, An analysis of the fixation probability of a mutant on special classes of non-directed graphs, *Proc. of the Royal Soc. of London A* **464** 2098 (2008), pp. 2609–2627
7. M. Broom, J. Rychtář, B. T. Stadler. Evolutionary dynamics on small-order graphs. *Journal of Interdisciplinary Mathematics* **12** 2 (2009) pp. 129–140.



**Fig. 5.** Quantitative variables versus fixation probability for graphs of order 10 and fitness 2. Color represents the amount of graphs in the corresponding region of the plot.

8. M. Broom, J. Rychtář, B. T. Stadler. Evolutionary dynamics on graphs - the effect of graph structure and initial placement on mutant spread. *Journal of Statistical Theory and Practice*, **5** 3 (2011), 369–381.
9. J. Díaz, L. A. Goldberg, G. B. Mertzios, D. Richerby, M. Serna, P. G. Spirakis. Approximating fixation probabilities in the generalized Moran process. In: *Proceedings of the Twenty-Third Annual ACM-SIAM Symposium on Discrete Algorithms. SODA '12*. SIAM (2012), 954–960.
10. A. Hagberg, D. Schult, P. Swart, Exploring Network Structure, Dynamics, and Function using NetworkX, *Proceedings of the 7<sup>th</sup> Python in Science conference (SciPy 2008)*, G. Varoquaux, T. Vaught, J. Millman (Eds.), 11–15.
11. L. Hindersin, B. Werner, D. Dingli, A. Traulsen. Should tissue structure suppress or amplify selection to minimize cancer risk? *Biology Direct* (2016), 11:41
12. L. Hindersin, A. Traulsen. Most Undirected Random Graphs Are Amplifiers of Selection for Birth-Death Dynamics, but Suppressors of Selection for Death-Birth Dynamics, *PLoS Comput Biol* **11** (2015) 1–14.
13. The HDF Group. *Hierarchical Data Format, version 5*, 1997-2016. <http://www.hdfgroup.org/HDF5/>.
14. B. Houchmandzadeh, M. Vallade. The fixation probability of a beneficial mutation in a geographically structured population. *New Journal of Physics* **13** (2011), 073020.
15. T. Junttila, P. Kaski. Engineering an efficient canonical labeling tool for large and sparse graphs. In *Proceedings of the Ninth Workshop on Algorithm Engineering and Experiments (ALENEX07)*, SIAM, 2007, 135–149.
16. N. L. Komarova, A. Sengupta, M. A. Nowaka. Mutation-selection networks of cancer initiation: tumor suppressor genes and chromosomal instability. *Journal of Theoretical Biology* **223** (2003), 433–450.
17. B. D. McKay, Isomorph-Free Exhaustive generation. *Journal of Algorithms* **26** 2 (1998), 306–324.
18. T. Monk, P. Green, M. Paulin. Martingales and fixation probabilities of evolutionary graphs. *Proc. R. Soc. A* **470** 20130730.
19. P. A. P. Moran. Random processes in genetics. *Proc. Cambridge Philos. Soc.*, **54** (1958), 60–71.
20. M. A. Nowak. *Evolutionary Dynamics*. Harvard University Press, 2006.
21. E. Lieberman, C. Hauert, M. A. Nowak. Evolutionary dynamics on graphs. *Nature* **433** (2005), 312–316.
22. V. Salas-Fumás, C. Sáenz, Á. Lozano. Organisational Structure and Performance of Consensus Decisions through Mutual Influences: A Computer Simulation Approach. *Decision Support Systems* **86** (2016) 61–72.
23. W. A. Stein et al., *Sage Mathematics Software (Version 7.0)*, The Sage Development Team, 2016, <http://www.sagemath.org>.
24. *MoM software*. Available at <https://bitbucket.org/geodynapp/mom>
25. P. Shakarian, P. Roos, A. Johnson. A review of evolutionary graph theory with applications to game theory. *BioSystems* **107** 2 (2012), 66–80.
26. S. Tan, J. Lu. Characterizing the effect of population heterogeneity on evolutionary dynamics on complex networks. *Sci. Rep.* **4** (2014) 5034.
27. B. Voorhees, A. Murray. Fixation probabilities for simple digraphs. *Proc. R. Soc. A* **469** (2013), 20120676.
28. S. Wasserman, K. Faust. *Social Network Analysis: Methods and Applications*. Cambridge, 1994.
29. D. J. Watts, S. H. Strogatz. Collective dynamics of “small-world” networks. *Nature* **393** (1998) 440–442.

**Table 2.** The columns of the dataset FP. All datatypes are little-endian and denoted as C types.

<b>Name</b>	<b>Description</b>	<b>Type</b>
<code>Id</code>	The id of the graph, see §2.1 and Figure 1	<code>uint64_t</code>
<code>FP_r</code>	The average fixation probability of the graph for fitness $r \in \{0.25, 0.5, \dots, 10.0\}$ and $r \neq 1.0$	<code>double</code>
<code>order</code>	The number of vertices of the graph	<code>uint8_t</code>
<code>size</code>	The number of edges of the graph	<code>uint8_t</code>
<code>degree_mean</code>	The mean degree of the graph	<code>double</code>
<code>degree_median</code>	The median of the degree distribution	<code>double</code>
<code>degree_var</code>	The variance of the degree distribution	<code>double</code>
<code>degree_skewness</code>	Pearson’s skewness of the degree distribution	<code>double</code>
<code>degree_kurtosis</code>	The kurtosis coefficient of the degree distribution	<code>double</code>
<code>degree_min</code>	The minimum number of neighbors of a vertex in the graph	<code>uint8_t</code>
<code>degree_max</code>	The maximum number of neighbors of a vertex in the graph	<code>uint8_t</code>
<code>diametre</code>	The distance between two vertices in the graph	<code>uint8_t</code>
<code>average_path_length</code>	The mean of the distances in the graph	<code>double</code>
<code>clustering</code>	The clustering coefficient [29]	<code>double</code>
<code>transitivity</code>	The fraction of possible triangles present in the graph. Also called global clustering coefficient [28, §4.10.3 and §6]	<code>double</code>
<code>heat_heterogeneity</code>	The variance of the temperature distribution, where the temperature of a vertex is the sum of inverses of the degrees of its neighbors [26]	<code>double</code>
<code>temperature_entropy</code>	The entropy of the temperature distribution [27]	<code>double</code>
<code>is_tree</code>	1 if the graph is a tree, 0 otherwise	<code>uint8_t</code>
<code>is_bipartite</code>	1 if the graph is bipartite, 0 otherwise	<code>uint8_t</code>
<code>is_biconnected</code>	1 if the graph is biconnected, 0 otherwise	<code>uint8_t</code>