



IST AUSTRIA

Institute of Science and Technology

Amplification on Undirected Population Structures: Comets Beat Stars

Andreas Pavlogiannis and Josef Tkadlec and Krishnendu Chatterjee and Martin A Nowak

Technical Report No. IST-2016-648-v1+1
Deposited at 09 Nov 2016 08:33
<https://repository.ist.ac.at/648/1/tr.pdf>

IST Austria (Institute of Science and Technology Austria)
Am Campus 1
A-3400 Klosterneuburg, Austria

Copyright © 2012, by the author(s).
All rights reserved.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, to republish, to post on servers or to redistribute to lists, requires prior specific permission.

Amplification on Undirected Population Structures: Comets Beat Stars

Andreas Pavlogiannis^{1*}, Josef Tkadlec^{1*},
Krishnendu Chatterjee¹, Martin A. Nowak²

November 8, 2016

Abstract

The fixation probability is the probability that a new mutant introduced in a homogeneous population eventually takes over the entire population. The fixation probability is a fundamental quantity of natural selection, and known to depend on the population structure. Amplifiers of natural selection are population structures which increase the fixation probability of advantageous mutants, as compared to the baseline case of well-mixed populations. In this work we focus on symmetric population structures represented as undirected graphs. In the regime of undirected graphs, the strongest amplifier known has been the Star graph, and the existence of undirected graphs with stronger amplification properties has remained open for over a decade. In this work we present the *Comet* and *Comet-swarm* families of undirected graphs. We show that for a range of fitness values of the mutants, the Comet and Comet-swarm graphs have fixation probability strictly larger than the fixation probability of the Star graph, for fixed population size and at the limit of large populations, respectively.

1 Introduction

Evolutionary dynamics study populations of reproducing individuals and the composition of the population over the course of time. A fundamental quantity is the *fixation probability* [1, 2, 3, 4, 5, 6, 7, 8, 9, 10], which characterizes the chances of an invading mutant to get fixed in a homogeneous population of residents. The most well-known mathematical model for studying evolutionary dynamics on finite populations is the birth-death Moran process [11]. Initially, a population of N individuals consists of two types: $N - 1$ residents, and 1 invading mutant. The residents are associated with a normalized fitness of 1, whereas the invading mutant has a fitness advantage $r > 1$, which is constant and independent of the composition of the population. The population size remains fixed over the course of time. At each time point, an individual is chosen for reproduction with probability proportional to its fitness, and its offspring replaces an individual chosen uniformly at random. In this setting, the population is *well-mixed*, as the reproducing individual may replace any other individual. The *fixation probability* is defined as the probability that the Moran process results in a population of N mutants (i.e., the mutants get *fixed* in the population). The fixation probability, ρ , for well-mixed populations is a function of r and N , and equals to

$$\rho(r, N) = \frac{1 - r^{-1}}{1 - r^{-N}}.$$

It is well-known that population structure affects the evolutionary dynamics [12, 13, 14, 5, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24]. Evolutionary graph theory models the population structure as a graph, where each vertex of the graph is occupied by one individual [5, 3, 25]. The edges of each vertex define the neighboring sites of that vertex in space. The *generalized* Moran process on a graph is identical to the Moran process on well-mixed populations, with the exception that each offspring can only replace a neighbor of the reproducing individual. The well-mixed population then follows as a special case of the generalized Moran process, where the individuals are spread on the vertices of a Clique (or

complete graph) K_N . A graph of N vertices G_N is said to *amplify selection* [5], if the fixation probability $\rho(r, G_N)$ of a randomly placed initial mutant on G_N is larger than the fixation probability on a well-mixed population of the same size (i.e., if $\rho(r, G_N) > \rho(r, K_N)$). The emerging question is then to what extent the population structure can amplify the fixation probability [5, 26, 27, 20, 21].

In this work, we focus on the most commonly studied case, where the population structure is modeled as an undirected graph, and the initial mutant arises with uniform probability on each vertex. Ever since the landmark work of [5], there has been immense interest in identifying selection amplifiers in this regime [15, 28, 26, 29, 16, 30]. Due to its combinatorial nature, the focus of such work has been primarily on simple structures with high degree of symmetry (e.g., Paths, Stars and Cycles). The intricacy of the problem has also given rise to computational approaches [31, 32, 33, 34, 35, 36, 37], which rely on numerical calculations and Monte Carlo simulations to search for amplifiers among small graphs.

Over a decade of active study, the strongest amplifier known for undirected graphs by either mathematical or computational approaches has been the Star graph S_N , with fixation probability

$$\rho(r, S_N) \simeq \frac{1 - r^{-2}}{1 - r^{-2N}}.$$

As $N \rightarrow \infty$, the fixation probability on the Star becomes $\rho(r, S_\infty) = 1 - r^{-2}$. In contrast, the corresponding probability for the well-mixed population is $\rho(r, K_\infty) = 1 - r^{-1}$. Hence, the Star is a *quadratic amplifier*, as it effectively amplifies the selective advantage of mutants from r to r^2 , where the well-mixed population is used as the basis of comparison. While for directed graphs stronger amplifiers are known (such as the Superstar [5]), the absence of stronger undirected amplifiers as compared to the Star graph has led to the conjecture that among undirected graphs the Star graph is the strongest amplifier [38]. The conjecture can be formalized as follows:

Conjecture 1.1. For all values of $r \geq 1$, for all undirected graphs G_N of N vertices, we have $\rho(r, G_N) \leq \rho(r, S_N)$.

In this work we refute the above conjecture. First, we present a graph G_N for a fixed size N , and show that there exist values of $r > 1$ such that $\rho(r, G_N) > \rho(r, S_N)$. Thus the refutation is wrt a fixed population size. Second, we present a family of graph $(M_N)_{N \geq 1}$ and show that there exist values of $r > 1$ such that $\lim_{N \rightarrow \infty} \rho(r, M_N) > \lim_{N \rightarrow \infty} \rho(r, S_N)$. This refutes the conjecture at the limit of large populations.

2 The Generalized Moran Process

We denote by $G_N = (V_N, E_N)$ an undirected graph of N vertices, which is connected. Given a vertex $u \in V_N$, we denote by $\text{Nh}(u)$ the set of *neighbors* of u , i.e., the vertices $v \in V_N$ such that $(u, v) \in E_N$. The *degree* of u is the number of neighbors of u , i.e., $\text{deg}(u) = |\text{Nh}(u)|$. A population of N individuals is spread on the vertices of G_N . Each individual is either a *resident* or a *mutant*. Mutants are associated with a *fitness advantage* $r \geq 1$, whereas the fitness of residents is normalized to 1. A *configuration* $S \subseteq V_N$ of G_N is the set of vertices of G_N that are occupied by mutants. The generalized Moran process on G_N is a discrete-time random process. Given a configuration S_i at time i , the next configuration at time $i + 1$ is determined by the following two events in succession.

1. One individual is chosen at random to reproduce, with probability proportional to its fitness. That is, the probability to reproduce is $r/F(S_i)$ for a mutant, and $1/F(S_i)$ for a resident, where

$$F(S_i) = r \cdot |S_i| + N - |S_i|$$

is the total population fitness. Let u be the vertex occupied by the reproducing individual.

2. A neighbor $v \in \text{Nh}(u)$ is chosen uniformly at random. The individual occupying v dies, and the offspring of the reproducing individual is placed on v .

The mutants *reach fixation* in G_N if at some time point i we reach $S_i = V$, i.e., all vertices of G_N are occupied by mutants. The mutants *reach extinction* if at some time point i we reach $S_i = \emptyset$, i.e., all vertices of G_N are occupied

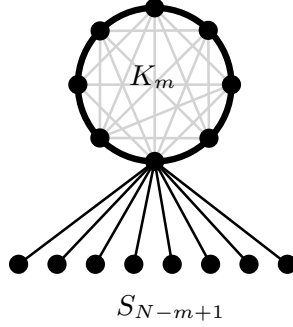


Figure 1: The Comet graph C_N^m consists of a Clique K_m and a Star S_{N-m+1} graph.

by residents. We denote by $\rho(r, G_N)$ the probability that the mutants reach fixation in the generalized Moran process starting with a single, uniformly placed mutant on G_N .

The Clique and Star graphs. The Clique graph K_N consists of N vertices and an edge between each pair of vertices. The Star graph S_N consists of a single *root* vertex and $N - 1$ leaf vertices, and an edge between the root and each of the leaves. It is known that [3]

$$\rho(r, K_N) = \frac{1 - r^{-1}}{1 - r^{-N}} \quad \text{and} \quad \rho(r, S_N) \simeq \frac{1 - r^{-2}}{1 - r^{-2N}}.$$

3 The Comet Family: Refutation for Fixed Population Size

In this section we refute Conjecture 1.1 for a fixed population size. We introduce a new graph family called the *Comet* graph, and show that for some fixed population sizes and values of r , Comets amplify selection more strongly than Stars.

The Comet graph C_N^m . Let m be any integer with $1 \leq m \leq N$. The Comet graph C_N^m consists of a Clique K_m of m vertices, where one vertex of the Clique is the root of a Star S_{N-m+1} of $N - m + 1$ vertices. Figure 1 shows an illustration. We refer to the Clique-part and the leaves of the Star-part of C_N^m as the *head* and the *tail* of the Comet, respectively. Observe how the Clique and Star graphs of N vertices are a special case of the Comet graph, without a tail (i.e., $K_N = C_N^N$) and the largest possible tail 1 (i.e., $S_N = C_N^1$), respectively.

Amplification on Comet graphs. The Comet C_N^m has the surprising property that for some values of m and r , it amplifies selection more strongly than the Star graph. Figure 2 shows the fixation probabilities on Comet graphs produced by keeping the population size N fixed, and varying the portion of the vertices that appear in the tail of the Comet. Remarkably, there is a range of graphs in between the two endpoints which amplify selection more strongly than the Star. For instance, we have $\rho(1.05, C_{200}^{120}) \geq 0.113$ and $\rho(1.05, S_{200}) < 0.093$ which leads to the following theorem.

Theorem 1. *There exists $r > 1$, N and m such that $\rho(r, C_N^m) > \rho(r, S_N)$.*

The above theorem refutes Conjecture 1.1 for fixed population size.

4 The Metastar Family: Refutation for the Limit of Large Population

In this section we refute Conjecture 1.1 for the limit of large population. We introduce the *Metastar* graph, and compute the fixation probability of new mutants arising uniformly at random. Intuitively, the Metastar is identical to

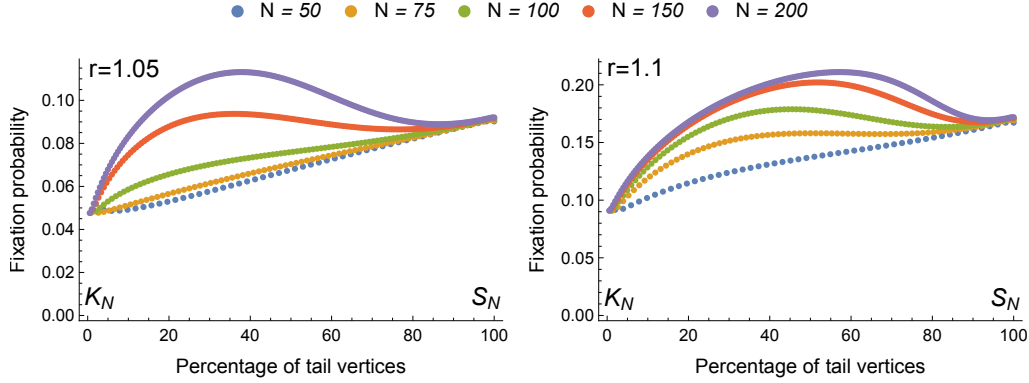


Figure 2: Fixation probabilities on the Comet graphs C_N^m . The Clique graph K_N and the Star S_N graph appear in the leftmost and rightmost points respectively. The X-axis shows the percentage of the vertices that appear in the tail of the Comet, with the remaining vertices appearing in the head of the Comet. In each case, all data-points which appear higher than the rightmost point of the plot correspond to Comets which amplify selection more strongly than the Star.

the Star, where each leaf vertex is replaced by a graph of small size. We will afterwards show how the Metastar family can be instantiated with such small graphs to refute Conjecture 1.1 for the limit of large population. We start with defining a variant of the generalized Moran process, called the lazy generalized Moran process.

4.1 The v -lazy generalized Moran Process

The v -lazy generalized Moran process. Given a distinguished vertex $v \in V_N$, the v -lazy generalized Moran process on G_N is identical to the generalized Moran process on G_N , except for the following modification. Whenever the reproducing individual occupies v , a biased coin with probability of heads $1/(\deg(v) + 1)$ is flipped, so that

1. if the coin comes up heads, the individual replaces itself (i.e., the population remains unchanged);
2. if the coin comes up tails, the individual replaces one of its neighbors, chosen uniformly at random, as in the generalized Moran process.

Intuitively, the vertex v is considered a neighbor to itself when it comes to replacing a neighboring individual.

Fixation probabilities. We consider fixation probabilities in the v -lazy generalized Moran process under two particular scenarios: (i) the initial mutant is placed on a vertex chosen uniformly at random, and (ii) the initial mutant is placed on a specific vertex. To refer to such events, we rely on the following notation.

- $\rho(r, G_N, v)$ is the probability that the mutants reach fixation in the v -lazy generalized Moran process starting with a single, uniformly placed mutant on G_N ;
- $\rho^+(r, G_N, v)$ is the probability that the mutants reach fixation in the v -lazy generalized Moran process starting with a single mutant placed on v ;
- $\rho^-(r, G_N, v)$ is the probability that the mutants reach extinction in the v -lazy generalized Moran process starting with a single resident placed on v .

4.2 The Metastar Family

Let $G_m = (V_m, E_m)$ be any fixed graph of m vertices, and distinguish some $v \in V_m$ as the *attachment vertex* of G_m . Given some $n \in \mathbb{N}^+$, we let $N(n) = n \cdot m + 1$, and construct the Metastar graph $M_{N(n)}^{G_m}$ parameterized by G_m as follows.

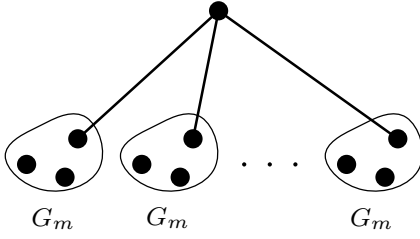


Figure 3: The Metastar graph $M_{N(n)}^{G_m}$ is identical to the Star graph, where every leaf is replaced by a small graph G_m .

1. We introduce n copies of G_m , and a new *root vertex* s .
2. We add an edge between the attachment vertex v of each copy of G_m and the root vertex s .

Figure 3 provides an illustration. From this point, we identify the i -th leaf of $M_{N(n)}^{G_m}$ with the i -th copy of G_m .

4.3 Fixation Probabilities on the Metastar

We now focus on the fixation probability on the Metastar. Since the graph is parameterized by G_m , this probability depends on G_m . However, because of the structure of $M_{N(n)}^{G_m}$, it does so in a modular way. We first introduce some terminology which will help with the exposition of these ideas.

1. A leaf of $M_{N(n)}^{G_m}$ is called *heterogeneous* if mutants and residents coexist in that leaf, and *homogeneous* otherwise. A *mutant leaf* (resp. *resident leaf*) is a homogeneous leaf that contains only mutants (resp. residents).
2. We say that a leaf i *hits the root* s when the individual placed on the attachment vertex of the i -th copy of G_m places an offspring on s . Similarly, the root s *hits leaf* i when the individual placed on s places an offspring on the attachment vertex of the i -th copy of G_m . We also say that a leaf i *hits another leaf* j at times (t_1, t_2) with $t_1 < t_2$ if leaf i hits the root at time t_1 and the root hits leaf j at time t_2 , and the root is not hit in the interval $[t_1, t_2]$.

Key idea. The key idea in analyzing the fixation probability on $M_{N(n)}^{G_m}$ is to show that as $n \rightarrow \infty$, every time the root hits a leaf i , or some leaf i hits another leaf j , the involved leaves are homogeneous with high probability. This is formally captured in the following two lemmas.

Lemma 1. *Consider that at some point the root hits a leaf i . The probability that the i -th leaf is heterogeneous the next time the root hits leaf i is $O(1/\sqrt{n})$.*

Proof. Since the graph G_m of leaf i has constant size, the expected time for leaf i to reach a homogeneous state is $O(n)$. On the other hand, the root s will need in expectation $\Omega(n^2)$ rounds to hit leaf i , as (i) s has n neighbors, and (ii) s reproduces approximately once every $N(n) = \Omega(n)$ rounds. The desired result then follows by applying concentration bounds. \square

Note that the complementary case of Lemma 1 does not hold, i.e., a heterogeneous leaf i will hit the root several times before leaf i becomes homogeneous. However, most of these events have no effect, as an offspring placed on the root by leaf i will be replaced by offsprings of other leaves, with high probability. The crucial event is the one in which a heterogeneous leaf i hits the root, and subsequently the root hits another leaf j *before* the root is hit again. Consider that leaf i becomes heterogeneous at some time t , and leaf i hits leaf j at times (t_1, t_2) , with $t_1 > t$. We call times (t, t_1, t_2) a *heterogeneous hit* if leaf i has remained heterogeneous in the interval $[t, t_1]$. The following lemma states that heterogeneous hits are rare.

Lemma 2. *Consider that at some time t the i -th leaf is heterogeneous. The probability of a heterogeneous hit (t, t_1, t_2) is $O(1/\sqrt{n})$.*

Proof. Note that in order for leaf i to hit leaf j , the following two events need to occur in succession.

- (A) Leaf i hits the root s , and afterwards
- (B) the root s reproduces before it is hit.

First, we rely on Lemma 1 to conclude that with high probability, the root s does not hit leaf i before the latter becomes homogeneous. Hence, the probability that leaf i has remained heterogeneous in the interval $[t, t_1]$ is approximately the probability that the v -lazy generalized Moran process on G_m has not reached a homogeneous state.

Since s has n neighbors, the probability of event B happening in each round is $O(1/n)$. Hence, in expectation, event A will need to happen $\Omega(n)$ times before leaf i hits leaf j . On the other hand, event A occurs with rate $O(1/n)$. Thus the expected time required for leaf i to hit leaf j is $\Omega(n^2)$. Finally, since the graph G_m occupying leaf i has constant size, the expected time to reach a homogeneous state is only $O(n)$. The desired result then follows by applying concentration bounds. \square

We are now ready to sketch the behavior of the Metastar. The initial mutant arises with high probability in one of the leaves, and is placed uniformly at random on one vertex of the corresponding graph G_m . Lemma 1 implies that we can focus on that leaf in isolation. Since v is attached to the root s , the corresponding evolutionary process on G_m alone is the v -lazy generalized Moran process, and hence the invading mutant fixates in the initial leaf with probability $\rho(r, G_m, v)$. From that point on, Lemma 1 and Lemma 2 guarantee that the Metastar behaves like the Star, with the exception that

1. when the root hits a resident leaf with a mutant offspring, the leaf turns mutant with probability approximately $\rho^+(r, G_m, v)$, and
2. when the root hits a mutant leaf with a resident offspring, the leaf turns resident with probability approximately $\rho^-(r, G_m, v)$.

In the case of the Star, both probabilities equal 1, since each leaf consists of a single vertex. Thus, if we focus on the ratio of probabilities of increasing the number of mutant leaves by one over decreasing it by one, this forward bias is amplified from r^2 (in the case of the Star) to $r^2 \cdot \rho^+(r, G_m, v) / \rho^-(r, G_m, v)$. We refer to the SI for the formal proof. The following theorem states the fixation probability on the Metastar.

Theorem 2. *Let G_m be a fixed graph and v the attachment vertex of G_m . Denote $p = \rho(r, G_m, v)$ and $\alpha = \rho^-(r, G_m, v)$ and $\beta = \rho^+(r, G_m, v)$. The fixation probability of a single mutant placed uniformly at random on $M_{N(n)}^{G_m}$ is*

$$\rho\left(r, M_{N(n)}^{G_m}\right) \geq p \cdot \frac{1 - r^{-2} \cdot (\alpha/\beta)}{1 - (r^{-2} \cdot (\alpha/\beta))^n} \cdot (1 + o(1)) \quad (1)$$

Note that for the special case where $m = 1$ and G_m consists of a single vertex $G_m = (\{v\}, \emptyset)$, we have $p = \alpha = \beta = 1$, and Eq. (1) gives the fixation probability on the Star graph. As $n \rightarrow \infty$, we have $N \rightarrow \infty$, and obtain that

$$\lim_{N \rightarrow \infty} \rho\left(r, M_N^{G_m}\right) \geq p \cdot (1 - r^{-2} \cdot (\alpha/\beta)). \quad (2)$$

4.4 Instances of the Metastar Family

In this section we present instances of the Metastar family. In particular we will instantiate the graphs G_m of the Metastar family with Comet graphs of Section 3.

Metastar: The Comet-swarm $M_{N(n)}^{C_{200}^{100}}$. We consider the Metastar $M_{N(n)}^{C_{200}^{100}}$ where each of the n leaves is a fixed-sized Comet C_{200}^{100} , and the attachment vertex v of C_{200}^{100} is some arbitrary vertex of its tail (Figure 4). We refer to this graph as the *Comet-swarm*, and obtain instances of various population sizes by increasing the number of leaves n .

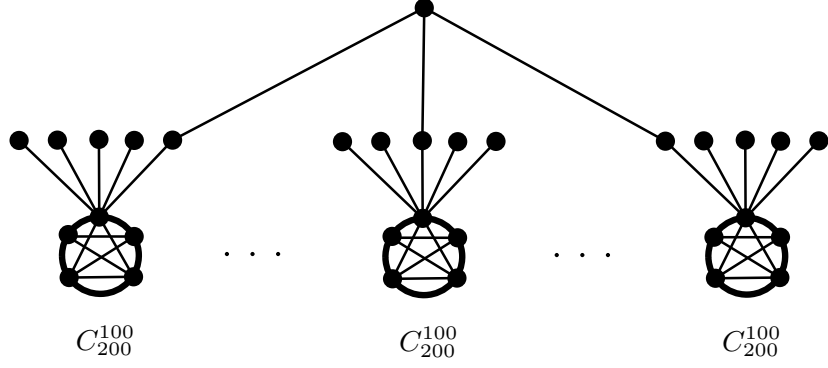


Figure 4: The Comet-swarm $M_N^{C_{200}^{100}}$.

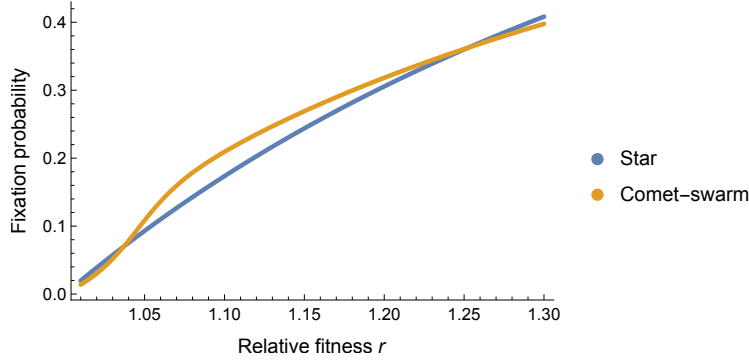


Figure 5: Fixation probabilities on $M_N^{C_{200}^{100}}$ for different values of relative fitness r and as $N \rightarrow \infty$. For a range of values of r the Metastar amplifies more strongly than the Star.

As the size of C_{200}^{100} is fixed, we can obtain the probabilities $p = \rho(r, C_{200}^{100}, v)$ and $\alpha = \rho^-(r, C_{200}^{100}, v)$ and $\beta = \rho^+(r, C_{200}^{100}, v)$ for any r , by direct calculations. Figure 5 shows the fixation probability $\lim_{N \rightarrow \infty} \rho(M_N^{C_{200}^{100}})$ obtained from Eq. 2 for various values of r . In particular, we have

$$\lim_{N \rightarrow \infty} \rho(1.1, M_N^{C_{200}^{100}}) = 0.209 \quad \text{and} \quad \lim_{N \rightarrow \infty} \rho(1.12, S_N) = 0.203 ;$$

and thus obtain the following refutation of Conjecture 1.1 for the limit of large population.

Theorem 3. *For any $r \in [1.1, 1.12]$, we have that*

$$\lim_{N \rightarrow \infty} \rho(r, M_N^{C_{200}^{100}}) > \lim_{N \rightarrow \infty} \rho(r, S_N).$$

5 Discussion

The generalized Moran process studies the evolution of populations on spatial structures. To understand the impact of the underlying topology, efforts have focused on characterizing the extremes of this process, i.e., the maximum amplification of selection that can be attained. The combinatorial nature of the problem makes it difficult for mathematical analysis, and most works focus on either simple graphs or asymmetric topologies, represented as directed graphs [5]. Directed graphs can exhibit extreme behavior, from strongly amplifying selection (fixation with probability 1) to strongly suppressing it (fixation with probability 0). There even exist directed graphs where neither fixation

nor extinction is possible. On the other hand, symmetric structures enjoy smoother behavior, as the population always resolves to a homogeneous state. In many cases symmetry is a very natural property, i.e., if an individual A can influence and individual B , then B can also influence A . Thus, amplification on undirected graphs is a very natural question to study.

It has been conjectured that the Star graph is the strongest amplifier of natural selection among undirected population structures. In this work we refute the conjecture both for fixed population sizes (with the Comet graph) and limit of large population (with the Comet-swarm Metastar family of graphs), for a range of values of r . Our results shed new light into the world of selection amplifiers, and inspire new research questions. While we establish stronger amplification for specific values of r , a concrete open question is as follows: Does there exist an undirected graph G_N such that for all values of r we have $\rho(r, G_N) > \rho(r, S_N)$? The above question asks for undirected graphs that amplify over Star graphs for all values of r . More generally, whether there exists a strongest amplifier among all undirected graphs is another open question.

References

- [1] Z Patwa and Lindi M Wahl. The fixation probability of beneficial mutations. *J. R. Soc. Interface*, 5(28):1279–1289, November 2008.
- [2] Martin A Nowak, Akira Sasaki, Christine Taylor, and Drew Fudenberg. Emergence of cooperation and evolutionary stability in finite populations. *Nature*, 428(6983):646–650, April 2004.
- [3] Martin A Nowak. *Evolutionary dynamics : exploring the equations of life*. Belknap Press of Harvard University Press, Cambridge, Massachusetts, 2006.
- [4] Drew Fudenberg, Martin A Nowak, Christine Taylor, and Lorens A Imhof. Evolutionary game dynamics in finite populations with strong selection and weak mutation. *Theor. Popul. Biol.*, 70(3):352–63, December 2006.
- [5] Erez Lieberman, Christoph Hauert, and Martin A Nowak. Evolutionary dynamics on graphs. *Nature*, pages 312–316, 2005.
- [6] Lorens A Imhof and Martin A Nowak. Evolutionary game dynamics in a Wright-Fisher process. *J. Math. Biol.*, 52(5):667–681, 2006.
- [7] Arne Traulsen, Jorge M Pacheco, and Martin A Nowak. Pairwise comparison and selection temperature in evolutionary game dynamics. *J. Theor. Biol.*, 246(3):522–529, June 2007.
- [8] Corina E Tarnita, Hisashi Ohtsuki, Tibor Antal, Feng Fu, and Martin A Nowak. Strategy selection in structured populations. *J. Theor. Biol.*, 259(3):570–81, August 2009.
- [9] Naoki Masuda and Hisashi Ohtsuki. Evolutionary dynamics and fixation probabilities in directed networks. *New J. Phys.*, 11(3):33012, 2009.
- [10] Ben Adlam and Martin A. Nowak. Universality of fixation probabilities in randomly structured populations. *Sci. Rep.*, 4, July 2014.
- [11] Patrick Alfred Pierce Moran. *The statistical processes of evolutionary theory*. Oxford University Press, Oxford, England, 1962.
- [12] Simon A Levin and Robert T Paine. Disturbance, patch formation, and community structure. *Proc. Natl. Acad. Sci.*, 71(7):2744–2747, 1974.
- [13] Simon A Levin. Population dynamic models in heterogeneous environments. *Annu. Rev. Ecol. Syst.*, 7(1):287–310, 1976.
- [14] Richard Durrett and Simon A Levin. Stochastic spatial models: a user’s guide to ecological applications. *Philos. Trans. R. Soc. London. Ser. B Biol. Sci.*, 343(1305):329–350, 1994.

- [15] Mark Broom and Jan Rychtář. An analysis of the fixation probability of a mutant on special classes of non-directed graphs. *Proc. R. Soc. A Math. Phys. Eng. Sci.*, 464(2098):2609–2627, October 2008.
- [16] Marcus Frean, Paul B Rainey, and Arne Traulsen. The effect of population structure on the rate of evolution. *Proc. R. Soc. B Biol. Sci.*, 280(1762):20130211, July 2013.
- [17] NH Barton. The probability of fixation of a favoured allele in a subdivided population. *Genet. Res.*, pages 149–157, 1993.
- [18] MC Whitlock. Fixation probability and time in subdivided populations. *Genetics*, 779(June):767–779, 2003.
- [19] Bahram Houchmandzadeh and Marcel Vallade. The fixation probability of a beneficial mutation in a geographically structured population. *New J. Phys.*, 13:073020, July 2011.
- [20] George B. Mertzios and Paul G. Spirakis. Strong bounds for evolution in networks. ICALP, pages 669–680, Berlin, Heidelberg, 2013. Springer-Verlag.
- [21] Andreas Galanis, Andreas Göbel, Leslie Ann Goldberg, John Lapinskas, and David Richerby. Amplifiers for the Moran Process. In *43rd International Colloquium on Automata, Languages, and Programming (ICALP 2016)*, volume 55 of *Leibniz International Proceedings in Informatics (LIPIcs)*, pages 62:1–62:13, Dagstuhl, Germany, 2016. Schloss Dagstuhl–Leibniz-Zentrum fuer Informatik.
- [22] Dirk Helbing, Attila Szolnoki, Matjaž Perc, and György Szabó. Punish, but not too hard: how costly punishment spreads in the spatial public goods game. *New J. Phys.*, 12(8):83005, 2010.
- [23] Dirk Helbing, Attila Szolnoki, Matjaž Perc, and György Szabó. Evolutionary Establishment of Moral and Double Moral Standards through Spatial Interactions. *PLoS Comput Biol*, 6(4):e1000758, April 2010.
- [24] György Szabó and Gábor Fáth. Evolutionary games on graphs. *Physics Reports*, 446(4–6):97 – 216, 2007.
- [25] Karan Pattni, Mark Broom, Jan Rychtář, and Lara J. Silvers. Evolutionary graph theory revisited: when is an evolutionary process equivalent to the moran process? *Proceedings of the Royal Society of London A: Mathematical, Physical and Engineering Sciences*, 471(2182), 2015.
- [26] M. Broom, J. Rychtář, and 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):369–381, 2011.
- [27] B Adlam, K Chatterjee, and MA Nowak. Amplifiers of selection. In *Proc. R. Soc. A*, volume 471, page 20150114. The Royal Society, 2015.
- [28] M. Broom, C. Hadjichrysanthou, J. Rychtář, and B. T. Stadler. Two results on evolutionary processes on general non-directed graphs. *Proceedings of the Royal Society of London A: Mathematical, Physical and Engineering Sciences*, 466(2121):2795–2798, 2010.
- [29] George B. Mertzios, Sotiris Nikolettseas, Christoforos Raptopoulos, and Paul G. Spirakis. Natural models for evolution on networks. *Theor. Comput. Sci.*, 477:76–95, 2013.
- [30] T Monk, P Green, and M Paulin. Martingales and fixation probabilities of evolutionary graphs. *Proc. R. Soc. A Math. Phys. Eng. Sci.*, 470(2165):20130730, 2014.
- [31] M. Broom, J. Rychtář, and B. Stadler. Evolutionary dynamics on small-order graphs. *Journal of Interdisciplinary Mathematics*, 12(2):129–140, 2009.
- [32] Paulo Shakarian and Patrick Roos. *Fast and deterministic computation of fixation probability in evolutionary graphs*, pages 97–104. 2011.
- [33] Christoforos Hadjichrysanthou, Mark Broom, and Istvan Z. Kiss. Approximating evolutionary dynamics on networks using a neighbourhood configuration model. *Journal of Theoretical Biology*, 312:13 – 21, 2012.
- [34] Josep Díaz, Leslie Ann Goldberg, George B Mertzios, David Richerby, Maria Serna, and Paul G Spirakis. Approximating fixation probabilities in the generalized moran process. *Algorithmica*, 69(1):78–91, 2014.

- [35] Laura Hindersin and Arne Traulsen. Counterintuitive properties of the fixation time in network-structured populations. *J. R. Soc. Interface*, 11(99), August 2014.
- [36] Laura Hindersin and Arne 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(11):1–14, 11 2015.
- [37] Rasmus Ibsen-Jensen, Krishnendu Chatterjee, and Martin A Nowak. Computational complexity of ecological and evolutionary spatial dynamics. *Proceedings of the National Academy of Sciences*, 112(51):15636–15641, 2015.
- [38] Paul Spirakis. Evolutionary dynamics on undirected networks, 2012. Summer school on Algorithmic Game Theory, Samos, Greece.

Supplementary Information for ‘Amplification on Undirected Population Structures: Comets Beat Stars’

Andreas Pavlogiannis^{1*†}, Josef Tkadlec^{1*},
Krishnendu Chatterjee¹, Martin A. Nowak²

¹IST Austria, Klosterneuburg, A-3400, Austria

²Program for Evolutionary Dynamics, Department of Organismic and Evolutionary Biology,
Department of Mathematics, Harvard University, Cambridge, MA 02138, USA

*These authors contributed equally to this work.

†To whom correspondence should be addressed; E-mail: pavlogiannis@ist.ac.at.

In this document we introduce the Metastar family of graphs and establish formally the fixation probability of mutants arising uniformly at random on a Metastar. The Metastar family is parametric wrt a small graph of fixed size, and the fixation probability depends on this fixed graph. We refer to the main article for particular instances of the Metastar family with specific fixed graphs, which result in Metastar graphs that amplify selection more strongly than Star graphs, for various fitness values of the invading mutants.

1 Preliminaries

In this section we define formally the Moran Process on graphs, and introduce several definitions and notation that will help with the exposition of the ideas in this work.

1.1 The Moran Process on Structured Populations

We denote by $G_N = (V_N, E_N)$ an undirected graph of N vertices, which is connected. Given a vertex $u \in V_N$, we denote by $\text{Nh}(u)$ the set of *neighbors* of u , i.e., the vertices $v \in V_N$ such that $(u, v) \in E_N$. The *degree* of u is the number of neighbors of u , i.e., $\text{deg}(u) = |\text{Nh}(u)|$. A population of N individuals is spread on the vertices of G_N . Each individual is either a *resident* or a *mutant*. Mutants are associated with a *fitness advantage* $r \geq 1$, whereas the fitness of residents is normalized to 1. A *configuration* $S \subseteq V_N$ of G_N is the set of vertices of G_N that are occupied by mutants. The generalized Moran process on G_N is a discrete-time random process. Given a configuration S_i at time i , the next configuration at time $i + 1$ is determined by the following two events in succession.

1. One individual is chosen at random to reproduce, with probability proportional to its fitness. That is, the probability to reproduce is $r/F(S_i)$ for a mutant, and $1/F(S_i)$ for a resident, where

$$F(S_i) = r \cdot |S_i| + N - |S_i|$$

is the total population fitness. Let u be the vertex occupied by the reproducing individual.

2. A neighbor $v \in \text{Nh}(u)$ is chosen uniformly at random. The individual occupying v dies, and the offspring of the reproducing individual is placed on v .

The v -lazy Moran process. Given an undirected graph $G = (V, E)$ and a distinguished node $v \in V$, the v -lazy Moran process on G is the regular Moran process on G with the following modification. Whenever the reproducing individual is one that occupies the node v , a biased coin with probability of heads $1/(\deg(v) + 1)$ is flipped, so that

1. if the coin comes up heads, the individual replaces itself (i.e., the population remains unchanged);
2. if the coin comes up tails, the individual replaces one of its neighbors, chosen uniformly at random (as in the regular Moran process).

Intuitively, the node v is considered a neighbor of itself when it comes to replacing a neighboring individual. We will often call v the *lazy node* in this process.

1.2 Fixation Probabilities

The mutants *reach fixation* in G_N if at some time point i we reach $S_i = V$, i.e., all vertices of G_N are occupied by mutants. The mutants *reach extinction* if at some time point i we reach $S_i = \emptyset$, i.e., all vertices of G_N are occupied by residents. We denote by $\rho(r, G_N)$ the probability that the mutants reach fixation in the generalized Moran process starting with a single, uniformly placed mutant on G_N . Given a heterogeneous population (where mutants and residents coexist) spread out on a graph $G = (V, E)$ the Moran process on G almost surely reaches a state where the mutants either fixate in the population or go extinct. Our interest is on the probability that starting from a state where a single mutant coexists with $N - 1$ residents, the mutant eventually fixates. In general, this probability depends on the node that the mutant occupies initially. We consider two particular scenarios: (i) the initial mutant is placed on a node chosen uniformly at random, and (ii) the initial mutant is placed on a specific node. To refer to such events, we rely on the following notation.

- $\rho(r, G_N)$ is the probability that the mutants reach fixation in the generalized Moran process starting from a single, uniformly placed mutant on G_N ;
- $\rho(r, G_N, v)$ is the probability that the mutants reach fixation in the v -lazy generalized Moran process starting with a single, uniformly placed mutant on G_N ;
- $\rho^+(r, G_N, v)$ is the probability that the mutants reach fixation in the v -lazy generalized Moran process starting with a single mutant placed on v ;
- $\rho^-(r, G_N, v)$ is the probability that the mutants reach extinction in the v -lazy generalized Moran process starting with a single resident placed on v .

The Clique and Star graphs. The Clique graph K_N consists of N vertices and an edge between each pair of vertices. The Star graph S_N consists of a single *root* vertex and $N - 1$ leaf vertices, and an edge between the root and each of the leaves. It is known that [1]

$$\rho(r, K_N) = \frac{1 - r^{-1}}{1 - r^{-N}} \quad \text{and} \quad \rho(r, S_N) \simeq \frac{1 - r^{-2}}{1 - r^{-2N}}.$$

2 The Metastar Family of Selection Amplifiers

In this section we introduce the Metastar family of graphs, and prove a general theorem about the fixation probability of mutants on the Metastar. The family is parameterized by a small graph of fixed size and naturally, the fixation probability depends on this parameter. However, because of the structure of the Metastar, it does so in a modular way.

The Metastar family $M_{N(n)}^{G_m}$. Let $G_m = (V_m, E_m)$ be any fixed graph of m vertices, and distinguish some $v \in V_m$ as the *attachment vertex* of G_m . Given some $n \in \mathbb{N}^+$, we let $N(n) = n \cdot m + 1$, and construct the Metastar graph $M_{N(n)}^{G_m}$ parameterized by G_m as follows.

1. We introduce n copies of G_m , and a new *root vertex* s .
2. We add an edge between the attachment vertex v of each copy of G_m and the root vertex s .

and we identify the i -th leaf of $M_{N(n)}^{G_m}$ with the i -th copy of G_m .

2.1 Random Trajectories and the Modified Moran Process on $M_{N(n)}^{G_m}$

We first introduce some notation that will help with the exposition of the ideas in this section.

1. Given a configuration S , a leaf of $M_{N(n)}^{G_m}$ is called *heterogeneous* in X if mutants and residents coexist in that leaf, and *homogeneous* otherwise. A *mutant leaf* (resp. *resident leaf*) is a homogeneous leaf that contains only mutants (resp. residents).
2. A *trajectory* ψ generated by the Moran process on $M_{N(n)}^{G_m}$ is a sequence of events $\psi = e_0, e_1, \dots, e_k$, where e_0 is the vertex of $M_{N(n)}^{G_m}$ that contains the initial mutant, and for $1 \leq i \leq k$, $e_i = (a_i, b_i)$ is a pair denoting the vertices a_i and b_i occupied by the reproducing and dying individuals respectively. Given some $i > 0$, we denote by ψ_i the prefix of ψ up to position $i - 1$. We let \mathcal{X}_i be a random variable that indicates the i -th event of ψ , and write $\mathbb{P}[\mathcal{X}_i = e_i | \psi_i]$ to denote the probability that \mathcal{X}_i is realized to e_i given the trajectory ψ_i up to position $i - 1$.
3. A *modified trajectory* π is obtained from a trajectory ψ by removing some events $e_i = (a_i, b_i)$ where $a_i = s$ is the root of $M_{N(n)}^{G_m}$.
4. We say that a leaf i *hits the root* s at time t in a trajectory ψ if $e_t = (a_t, b_t)$ and a_t is the attachment vertex of the i -th copy of G_m , and b_t is the root s . Similarly, we say that the root s *hits leaf* i at time t in ψ if $e_t = (a_t, b_t)$ and a_t is the root s and b_t is the attachment vertex of the i -th copy of G_m . We also say that a leaf i *hits another leaf* j at times (t_1, t_2) with $t_1 < t_2$ if leaf i hits the root at time t_1 and the root hits leaf j at time t_2 , and the root is not hit in the interval $[t_1, t_2]$. In such a case, we call (t_1, t_2) a *hitting pair*. A *heterogeneous hit* from leaf i to leaf j occurs at times (t_1, t_2, t_3) if
 - (a) The root hits leaf i at time t_1 , and
 - (b) leaf i hits leaf j at times (t_2, t_3) , and
 - (c) leaf i is heterogeneous in the interval $(t_1, t_2]$.

The modified Moran process. The *modified* Moran process on $M_{N(n)}^{G_m}$ consists of the regular Moran process with the following modifications:

1. if at any point the root hits a heterogeneous leaf i , then leaf i becomes instantaneously a resident leaf, and
2. if at any point a heterogeneous leaf i hits a leaf j , then leaf j becomes instantaneously a resident leaf.

Observe that every time Item 1 or Item 2 applies, the modified Moran process transitions to a configuration S' while the regular Moran process would transition to a configuration S and such that set of vertices occupied by mutants in X' is a subset of the set of vertices occupied by mutants in S . Thus the fixation probability from S is at least as large as the fixation probability from S' . We will use the modified Moran process on $M_{N(n)}^{G_m}$ to underapproximate the fixation probability $\rho(r, M_{N(n)}^{G_m})$.

2.2 The Interference of Heterogeneous Leaves

In this section we prove some useful lemmas regarding the modified Moran process on $M_{N(n)}^{G_m}$. In particular, we show that every time the root hits a leaf i , or a leaf i hits another leaf j , the involved leaves can be considered homogeneous whp (Lemma 1 and Lemma 2). Additionally, we show that if the root reproduces in any two times t_1 and t_2 , with $t_1 < t_2$, the root is hit in the interval (t_1, t_2) whp (Lemma 3). Finally, in Lemma 4 we characterize the probability that a hitting pair (t_1, t_2) is such that the individual reproducing in time t_1 comes from a mutant leaf, over the probability that this individual comes from a resident leaf.

The following lemma states that once a homogeneous leaf is hit by the root, w.h.p. that leaf will evolve independently of the root until it becomes homogeneous again.

Lemma 1. *Let π be a random modified trajectory, and consider that the root hits a leaf i at some time t . The probability that the i -th leaf is heterogeneous the next time it is hit by the root is $O(1/\sqrt{n})$.*

Proof. Let X be the random variable that indicates the number of rounds until the root hits leaf i for the first time after time t . Note that in a random modified trajectory the root reproduces with rate no larger than the rate with which it reproduces in a random regular trajectory. We lower-bound X by a random variable X' which realizes the same event while assuming that the root is always occupied by a mutant, and the rest of the population consists only of residents, and additionally π is a random regular trajectory. In particular, we have $\mathbb{P}[X \leq \alpha] \leq \mathbb{P}[X' \leq \alpha]$ for all α , where X' is geometrically distributed with rate r/n^2 , which is an upper bound on the probability of the root reproducing and hitting leaf i , that is

$$X' \sim \text{GM}(p_X) \quad \text{where} \quad p_X = \frac{r}{n^2}$$

Let Y_y be the random variable that denotes the number of rounds required for y reproduction events to take place in leaf i after time t . We upper-bound Y_y by another random variable Y'_y which realizes the same event while assuming that leaf i is a resident leaf, and the rest of the population contains only mutants. We have $\mathbb{P}[Y_y \geq \alpha] \leq \mathbb{P}[Y'_y \geq \alpha]$ for all α , with Y'_y drawn from the negative binomial distribution of receiving y failures with success rate equal to the probability of choosing an individual to reproduce that does not belong to that leaf, that is

$$Y'_y \sim \text{NB}(y, p_Y) \quad \text{where} \quad p_Y = \frac{((n-1) \cdot m + 1) \cdot r}{((n-1) \cdot m + 1) \cdot r + m} \quad \mathbb{E}[Y'_y] = \frac{p_Y \cdot y}{1 - p_Y} \quad \text{Var}[Y'_y] = \frac{p_Y \cdot y}{(1 - p_Y)^2}$$

Let $y = \sqrt{n}$. The event of leaf i being heterogeneous when hit by the root requires that either (i) it is hit by the root before y reproduction events have occurred locally, or (ii) it has remained heterogeneous after y reproduction events have occurred locally. Since $m = O(1)$, event (ii) happens with probability $O(1/\sqrt{n})$, while event (i) happens with probability $\mathbb{P}[X < Y_y]$. Let $\alpha = c \cdot n^{3/2}$ where c is a sufficiently large constant, and note that

$$\mathbb{P}[X < Y_y] = \mathbb{P}[X < Y_y \leq \alpha] + \mathbb{P}[X < \alpha < Y_y] + \mathbb{P}[\alpha \leq X < Y_y] \leq \mathbb{P}[Y'_y \geq \alpha] + \mathbb{P}[X' \leq \alpha] \quad (1)$$

where $X' \sim \text{GM}(p_X)$ and $Y'_y \sim \text{NB}(y, p_Y)$ as defined above. We have

$$\mathbb{P}[X' \leq \alpha] = 1 - (1 - p_X)^\alpha = 1 - \left(1 - \frac{r}{n^2}\right)^{c \cdot n^{3/2}} = O(1/\sqrt{n}) \quad (2)$$

and

$$\mathbb{P}[Y'_y \geq \alpha] = \mathbb{P}[Y'_y - \mathbb{E}[Y'_y] \geq \alpha - \mathbb{E}[Y'_y]] \leq \frac{\text{Var}[Y'_y]}{(\alpha - \mathbb{E}[Y'_y])^2} \quad (3)$$

Note that

$$\text{Var}[Y'_y] = \frac{\frac{((n-1) \cdot m + 1) \cdot r \cdot \sqrt{n}}{((n-1) \cdot m + 1) \cdot r + m}}{\left(\frac{m}{((n-1) \cdot m + 1) \cdot r + m}\right)^2} = O(n^{5/2})$$

and

$$\mathbb{E}[Y'_y] = \frac{\frac{((n-1) \cdot m + 1) \cdot r \cdot \sqrt{n}}{((n-1) \cdot m + 1) \cdot r + m}}{\frac{m}{((n-1) \cdot m + 1) \cdot r + m}} = O(n^{3/2})$$

thus

$$(\alpha - \mathbb{E}[Y'_y])^2 = (c \cdot n^{3/2} - O(n^{3/2}))^2 = \Omega(n^3)$$

Substituting to Eq. (3), we obtain $\mathbb{P}[Y'_y \geq \alpha] = O(1/\sqrt{n})$, and combining Eq. (2) and Eq. (3) with Eq. (1) we have $\mathbb{P}[X < Y_y] = O(1/\sqrt{n})$. By the union bound, the desired event happens with probability $O(1/\sqrt{n}) + O(1/\sqrt{n})$. The desired result follows. \square

The following lemma states that when the root hits a leaf, the individual in the root is an offspring that came from a homogeneous leaf.

Lemma 2. *Let π be a random modified trajectory, and consider that at some time t_1 the i -th leaf is hit by the root. Let $t_2 > t_3$ be the random variables which indicate that leaf i hits some leaf j at times (t_2, t_3) . The probability that (t_1, t_2, t_3) is a heterogeneous hit is $O(1/\sqrt{n})$.*

Proof. Note that in a random modified trajectory the root reproduces with rate no larger than the rate with which it reproduces in a random regular trajectory. Let Z be the random variable that counts the number of times leaf i hits the root in the interval $(t_1, t_2]$. We obtain a lower bound on Z by assuming that the root is the only mutant in the population after it is hit by leaf i , and π is a random regular trajectory. In particular, we have $\mathbb{P}[Z \leq \alpha] \leq \mathbb{P}[Z' \leq \alpha]$ for all α , where Z' is geometrically distributed with rate $(2 \cdot r)/n$, which is an upper bound on the probability that the root reproduces before it is hit again, i.e.

$$Z' \sim \text{GM}(p_Z) \quad \text{where} \quad p_Z = \frac{2 \cdot r}{n}$$

Then

$$\mathbb{P}[Z \leq \sqrt{n}] \leq \mathbb{P}[Z' \leq \sqrt{n}] = 1 - (1 - p_Z)^{\sqrt{n}} = O(1/\sqrt{n}) \quad (4)$$

Hence, the probability to observe the desired event when leaf i is chosen for reproduction at most \sqrt{n} times in the interval $(t_1, t_2]$ is $O(1/\sqrt{n})$. On the other hand, the probability to observe the desired event when leaf i is chosen for reproduction at least \sqrt{n} times requires that the leaf has remained heterogeneous after at least \sqrt{n} reproduction events have occurred locally. In turn, this event requires that leaf i is hit by the root before it becomes homogeneous, or it has remained heterogeneous after \sqrt{n} reproduction events have occurred locally, given that it has not been hit by the root. By Lemma 1, the probability that leaf i is hit by the root before it becomes homogeneous is $O(1/\sqrt{n})$. Finally, since $m = O(1)$, if we condition on the fact that leaf i is not hit by the root before it becomes homogeneous, the probability that it has remained heterogeneous after it has been chosen for reproduction \sqrt{n} times is $O(1/\sqrt{n})$. The desired result follows. \square

Lemma 3. *Let π be a random modified trajectory, any position t of π . The probability that the root reproduces after t before it is hit by a leaf is $O(1/n)$.*

Proof. Note that in a random modified trajectory the root reproduces with rate no larger than the rate with which it reproduces in a random regular trajectory. In a random regular trajectory, the root is chosen for reproduction with rate at most r/n , whereas it is hit by the leaves with rate at least ε , for some constant ε . Then the probability that the root is chosen for reproduction before it is hit by a leaf is at most

$$\frac{\frac{r}{n}}{\frac{r}{n} + \varepsilon} = O(1/n)$$

Hence the event in consideration occurs with probability $O(1/n)$. \square

Lemma 4. Let π be a random modified trajectory, and t any position of π . Let t_1, t_2 be the random variables that indicate a hitting pair (t_1, t_2) , with $t_1 > t$. Let $e_{t_1} = (a_{t_1}, b_{t_1})$ and $e_{t_2} = (a_{t_2}, b_{t_2})$. If a_{t_1} and b_{t_2} belong to homogeneous leaves of different types, then the probability that a_{t_1} belongs to a resident leaf is at most $1/(r^2 + 1) + O(1/n)$.

Proof. The proof is by showing that for any modified trajectory π^- up to time t_2 , and in which a_1 belongs to a resident leaf, if $t_2 - t_1$ is “reasonably small”, then there exists a modified trajectory π^+ up to time t_2 where a_{t_1} belongs to a mutant leaf and such that

$$\frac{\mathbb{P}[\pi^-]}{\mathbb{P}[\pi^+]} \leq \frac{1}{r^2} \cdot (1 + O(1/n))$$

Let $\pi^- = e_1^-, \dots, e_{t_2}^-$. Observe that in every round the root is hit with probability at least ε , for some constant $\varepsilon > 0$. Let $\gamma = -\log(1 - \varepsilon)$. First we show that the probability that $t_2 - t_1 - 1 > (1/\gamma) \cdot \log n$ is $O(1/n)$. Indeed, since (t_1, t_2) is a hitting pair, the root is not hit in any e_i^- for $t_1 < i < t_2$, and the probability of this event happening is at most $(1 - \varepsilon)^{t_2 - t_1 - 1}$. Then, for $t_2 - t_1 - 1 > (1/\gamma) \cdot \log n$, this event happens with probability at most

$$(1 - \varepsilon)^{\frac{1}{\gamma} \cdot \log n} = \left(\frac{1}{1 - \varepsilon} \right)^{-\frac{1}{\gamma} \cdot \log n} = 2^{-\gamma \cdot \frac{1}{\gamma} \cdot \log n} = 2^{-\log n} = O(1/n) \quad (5)$$

Now assume that $t_2 - t_1 - 1 \leq (1/\gamma) \cdot \log n$. The probability of π^- is

$$\mathbb{P}[\pi^-] = \prod_{i=1}^{t_2} \mathbb{P}[\mathcal{X}_i = e_i^- | \pi_i^-] = \left(\prod_{i=1}^{t_1} \mathbb{P}[\mathcal{X}_i = e_i^- | \pi_i^-] \right) \cdot \left(\prod_{i=t_1+1}^{t_2} \mathbb{P}[\mathcal{X}_i = e_i^- | \pi_i^-] \right)$$

We obtain π^+ by replacing e_{t_1} and e_{t_2} with e'_{t_1} and e'_{t_2} respectively, where $e'_{t_1} = (b_{t_2}, a_{t_2})$ and $e'_{t_2} = (b_{t_1}, a_{t_1})$. Let $\pi^+ = e_1^+, \dots, e_{t_2}^+$, and then (t_1, t_2) is a hitting pair in π^+ , in which the reproducing leaf $e_{t_1}^+$ is mutant.

We have

$$\mathbb{P}[\pi^+] = \prod_{i=1}^k \mathbb{P}[\mathcal{X}_i = e_i^+ | \pi_i^+] = \left(\prod_{i=1}^{t_1} \mathbb{P}[\mathcal{X}_i = e_i^+ | \pi_i^+] \right) \cdot \left(\prod_{i=t_1+1}^{t_2} \mathbb{P}[\mathcal{X}_i = e_i^+ | \pi_i^+] \right)$$

Since $\pi_{t_1}^+ = \pi_{t_1}^-$, we have

$$\prod_{i=1}^{t_1} \frac{\mathbb{P}[\mathcal{X}_i = e_i^- | \pi_i^-]}{\mathbb{P}[\mathcal{X}_i = e_i^+ | \pi_i^+]} = \frac{\mathbb{P}[\mathcal{X}_i = e_{t_1}^- | \pi_{t_1}^-]}{\mathbb{P}[\mathcal{X}_i = e_{t_1}^+ | \pi_{t_1}^+]} = \frac{\mathbb{P}[\mathcal{X}_i = e_{t_1}^- | \pi_{t_1}^-]}{\mathbb{P}[\mathcal{X}_i = e'_{t_1} | \pi_{t_1}^-]} = \frac{1}{r} \quad (6)$$

Let f_i^- and f_i^+ be the fitness of the population right before events e_i^- and e_i^+ occur, respectively. Since after position t_1 there is one more mutant in π^+ than π^- , we have $f_i^+ = f_i^- + r - 1$ for $i \geq t_1$. Since $f_i^- = \Omega(n)$, we have

$$\prod_{i=k'+1}^k \frac{f_i^+}{f_i^-} = \prod_{i=k'+1}^k \frac{(f_i^- + r - 1)}{f_i^-} \leq (1 + O(1/n))^{k-k'+1} \leq (1 + O(1/n))^{(1/\gamma) \cdot \log n + 1} \leq 1 + O(1/n)$$

thus

$$\prod_{i=k'+1}^k \frac{\mathbb{P}[\mathcal{X}_i = e_i^- | \pi_i^-]}{\mathbb{P}[\mathcal{X}_i = e_i^+ | \pi_i^+]} \leq \frac{1}{r} \cdot \prod_{i=k'+1}^k \frac{f_i^+}{f_i^-} \leq \frac{1}{r} \cdot (1 + O(1/n)) \quad (7)$$

Using Eq. (6) and Eq. (7) we obtain

$$\frac{\mathbb{P}[\pi^-]}{\mathbb{P}[\pi^+]} = \left(\prod_{i=1}^{t_1} \frac{\mathbb{P}[\mathcal{X}_i = e_i^- | \pi_i^-]}{\mathbb{P}[\mathcal{X}_i = e_i^+ | \pi_i^+]} \right) \cdot \left(\prod_{i=t_1+1}^{t_2} \frac{\mathbb{P}[\mathcal{X}_i = e_i^- | \pi_i^-]}{\mathbb{P}[\mathcal{X}_i = e_i^+ | \pi_i^+]} \right) \leq \frac{1}{r} \cdot \frac{1}{r} \cdot (1 + O(1/n)) = \frac{1}{r^2} \cdot (1 + O(1/n)) \quad (8)$$

Overall, the probability that a_{t_1} belongs to a resident leaf is bounded by the sum of the probabilities of this event happening when (i) $t_2 - t_1 - 1 > (1/\gamma) \cdot \log n + 1$ and (ii) $t_2 - t_1 - 1 \leq (1/\gamma) \cdot \log n + 1$. By Eq. (5), case (i) happens with probability $O(1/n)$, whereas by Eq. (8) case (ii) happens with probability

$$\frac{1}{1 + \frac{r^2}{1+O(1/n)}} \leq \frac{1}{r^2 + 1} + O(1/n)$$

Thus the event under consideration happens with probability at most

$$O(1/n) + \frac{1}{r^2 + 1} + O(1/n) = \frac{1}{r^2 + 1} + O(1/n)$$

The desired result follows. \square

2.3 A Coupling Argument for the Fixation Probability on $M_{N(n)}^{G_m}$

In this section we introduce a simple Markov chain \mathcal{M} and use a coupling argument to argue that the fixation probability on the Metastar is underapproximated by the probability that a random walk on \mathcal{M} gets absorbed in a particular state.

To simplify notation, we let $\alpha = \rho^-(r, G_m, v)$ and $\beta = \rho^+(r, G_m, v)$. We define a Markov chain $\mathcal{M} = (S, \delta)$ which consists of the set of states S :

1. s_i , for $0 \leq i \leq n$,
2. η_i , for $0 < i < n$,
3. θ_i , for $0 < i < n$.

Let $z = c/\sqrt{n}$ where c is a large enough constant. The transition probability function $\delta : S \times S \rightarrow [0, 1]$ is defined such that for all $0 < i < n$ we have:

1. $\delta(s_i, \eta_i) = \frac{r^2}{r^2+1} - 2 \cdot z$,
2. $\delta(s_i, \theta_i) = \frac{1}{r^2+1} + z$,
3. $\delta(s_i, s_{i-1}) = z$,
4. $\delta(\eta_i, s_{i+1}) = \beta - z$,
5. $\delta(\eta_i, s_i) = 1 - \beta + z$,
6. $\delta(\theta_i, s_{i-1}) = \alpha + z$,
7. $\delta(\theta_i, s_i) = 1 - \alpha - z$,

whereas $\delta(s_0, s_0) = \delta(s_n, s_n) = 1$ (i.e., the states s_0 and s_n are absorbing).

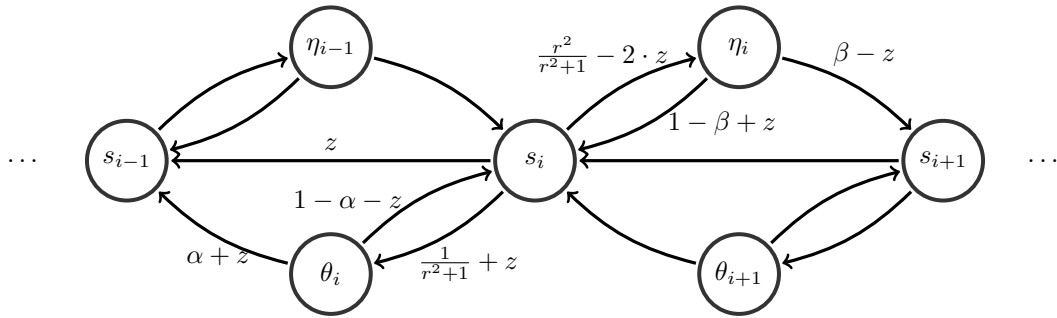


Figure 1: .

Intuition. Intuitively, a random walk on \mathcal{M} starting from state s_1 models the modified Moran process on $M_{N(n)}^{G_m}$ starting from a mutant leaf. Whenever the random walk is on some state s_i , there are at most $n - i$ resident leaves in

$M_{N(n)}^{G_m}$. A transition to state η corresponds to the event of a mutant leaf hitting a resident leaf. A transition to state θ corresponds to the event of a resident leaf hitting a mutant leaf.

Lemma 5. *Consider that at some time t^* the modified Moran process on $M_{N(n)}^{G_m}$ reaches a configuration S_{t^*} which contains a mutant leaf. Then the fixation probability on $M_{N(n)}^{G_m}$ is at least the probability that a random walk on \mathcal{M} starting from state s_0 eventually gets absorbed in state s_n .*

Proof. Let ψ be a fixed trajectory of the modified Moran process of to time t^* , which generates the desired configuration S_{t^*} . Let Π_1 be the modified Moran process on $M_{N(n)}^{G_m}$ starting from S_{t^*} . Similarly, let Π_2 be the random process on \mathcal{M} starting from s_1 . We couple Π_1 and Π_2 , so that whenever Π_2 is on state s_i of \mathcal{M} , there are at most $n - i$ (homogeneous) resident leaves of $M_{N(n)}^{G_m}$ in Π_1 . We do so by first extending ψ indefinitely, and then using ψ with some biased coins as the source of randomness for Π_2 .

We now describe the process of associating certain events in ψ with events in Π_2 . In this process, we will be erasing some events e_i of ψ where the reproducing individual of e_1 occupies the root of $M_{N(n)}^{G_m}$. Thus we will be working in general with a modified trajectory π . Initially, π is identical to ψ .

Whenever Π_2 transitions to some state η_i or θ_i , this will correspond to some leaf j of $M_{N(n)}^{G_m}$ being hit by the root in π . Then we will refer to leaf j as the *active leaf* while in η_i or θ_i . In particular, an active leaf in η_i is one that is hit with a mutant and was a resident leaf in the last round, whereas an active leaf in θ_i is one that is hit with a resident, and was a homogeneous mutant leaf in the last round.

1. Starting from s_1 , we scan π from left to right until we find the first occurrence in π where some leaf j becomes heterogeneous. Let \mathcal{E}_1 be the event that the root reproduces before it is hit, and by Lemma 3, we have $\mathbb{P}[\mathcal{E}_1] \leq z$. If \mathcal{E}_1 does not hold, we flip a coin with probability of heads $z - \mathbb{P}[e_1]$, and let \mathcal{E}_2 be the event that the coin comes up heads. We make Π_2 take the transition $s_1 \rightarrow s_0$ if either \mathcal{E}_1 or \mathcal{E}_2 hold. Note that the second event is conditioned on the failure of the first (i.e., we assign $\mathbb{P}[\mathcal{E}_2|\mathcal{E}_1] = 0$), thus \mathcal{E}_1 and \mathcal{E}_2 are disjoint, and hence their union occurs with probability

$$\mathbb{P}[\mathcal{E}_1] + \mathbb{P}[\mathcal{E}_2] = z$$

Now assume that both \mathcal{E}_1 and \mathcal{E}_2 fail. Since \mathcal{E}_1 does not occur, the leaf j is turned heterogeneous by a hitting pair (t, t') , and let $e_t = (a_t, b_t), e_{t'} = (a_{t'}, b_{t'})$. Let \mathcal{E}_3 be the event that a_t belongs to a resident leaf, and by Lemma 4,

$$\mathbb{P}[\mathcal{E}_3] \leq \frac{1}{r^2 + 1} + z$$

If \mathcal{E}_3 does not hold, we flip a coin with probability of heads $z - \mathbb{P}[\mathcal{E}_3]$, and let \mathcal{E}_4 be the event that the coin comes up heads. We make Π_2 take the transition $s_1 \rightarrow \theta_1$ if either \mathcal{E}_3 or \mathcal{E}_4 occurs. Additionally, we mark leaf j as the active leaf in θ_1 . Finally, we make Π_2 take the transition $s_1 \rightarrow \eta_1$ if none other transition has been taken, i.e. with probability

$$1 - \left(\frac{1}{r^2 + 1} + z \right) = \frac{r^2}{r^2 + 1} - z$$

and mark leaf j as the active leaf in η_1 . Additionally, we erase event $e_{t'}$ from π .

2. While in state η_i with active leaf j , let \mathcal{E}_1 be the event that leaf j is hit by the root at some time t before it becomes homogeneous, and by Lemma 1, we have $\mathbb{P}[\mathcal{E}_1] \leq z$. If \mathcal{E}_1 occurs, we erase the event e_t from π . If \mathcal{E}_1 does not occur, let \mathcal{E}_2 be the event that the j -th leaf becomes a resident leaf the next time it becomes homogeneous. Since this leaf has been hit with a mutant, we have $\mathbb{P}[\mathcal{E}_2] = 1 - \beta$. Finally, if neither \mathcal{E}_1 nor \mathcal{E}_2 hold, we flip a coin with probability of heads $z - \mathbb{P}[\mathcal{E}_1]$, and let \mathcal{E}_3 be the event that the coin comes up heads. We make Π_2 take the transition $\eta_i \rightarrow s_i$ if any of the events $\mathcal{E}_1, \mathcal{E}_2$ and \mathcal{E}_3 occurs, which happens with probability

$$\mathbb{P}[\mathcal{E}_1] + \mathbb{P}[\mathcal{E}_2] + \mathbb{P}[\mathcal{E}_3] = 1 - \beta + z$$

and make Π_2 take the transition $\eta_i \rightarrow s_{i+1}$ otherwise, i.e. with probability $\beta - z$.

3. While in state θ_i with active leaf j , let \mathcal{E}_1 be the event that leaf j is hit by the root at some time t before it becomes heterogeneous, and by Lemma 1, we have $\mathbb{P}[\mathcal{E}_1] \leq z$. If \mathcal{E}_1 does not occur, let \mathcal{E}_2 be the event that the j -th leaf becomes a resident leaf the next time it becomes homogeneous. Since this leaf has been hit with a resident and was a homogeneous mutant leaf before, we have $\mathbb{P}[\mathcal{E}_2] = \alpha$. Finally, if neither \mathcal{E}_1 nor \mathcal{E}_2 hold, we flip a coin with probability of heads $z - \mathbb{P}[\mathcal{E}_1]$, and let \mathcal{E}_3 be the event that the coin comes up heads. We make Π_2 take the transition $\theta_i \rightarrow s_{i-1}$ if any of the events $\mathcal{E}_1, \mathcal{E}_2$ and \mathcal{E}_3 occurs, which happens with probability

$$\mathbb{P}[\mathcal{E}_1] + \mathbb{P}[\mathcal{E}_2] + \mathbb{P}[\mathcal{E}_3] = \alpha + z$$

and make Π_2 take the transition $\theta_i \rightarrow s_i$ otherwise, i.e. with probability $1 - \alpha - z$.

4. While in state s_i , let j be the last active leaf (i.e., the leaf that was active the last time Π_2 was in either a state $\eta_{i'}$ or $\theta_{i'}$), and t_1 the position in π that turned leaf j heterogeneous. Let t_2 be the first time after t_1 such that π has a hitting pair (t_2, t_3) , and \mathcal{E}_1 the event that (t_1, t_2, t_3) constitutes a heterogeneous hit. By Lemma 2, we have $\mathbb{P}[\mathcal{E}_1] \leq z$. If \mathcal{E}_1 occurs, we erase from π the event e_{t_3} . If \mathcal{E}_1 does not occur, let \mathcal{E}_2 be the event that the root reproduces at some time $t_4 > t_1$ and it is not hit in the interval (t_1, t_4) . By Lemma 3, we have $\mathbb{P}[\mathcal{E}_2] \leq z$. If \mathcal{E}_2 occurs, we erase from π the event e_{t_4} . If neither \mathcal{E}_1 nor \mathcal{E}_2 occurs, we flip a coin with probability of heads $z - \mathbb{P}[\mathcal{E}_1] - \mathbb{P}[\mathcal{E}_2]$, and let \mathcal{E}_3 be the event that the coin comes up heads. We make Π_2 take the transition $s_i \rightarrow s_{i-1}$ if either \mathcal{E}_1 or \mathcal{E}_2 occurs, which happens with probability

$$\mathbb{P}[\mathcal{E}_1] + \mathbb{P}[\mathcal{E}_2] + \mathbb{P}[\mathcal{E}_3] = z$$

If none of $\mathcal{E}_1, \mathcal{E}_2$ and \mathcal{E}_3 occur, we scan π to the right from position t_1 , and find the next hitting pair (t, t') , and let $e_t = (a_t, b_t), e_{t'} = (a_{t'}, b_{t'})$. Note that a_t and $b_{t'}$ necessarily belong to homogeneous leaves. Let \mathcal{E}_4 be the event that a_t belongs to a resident leaf, and by Lemma 4 we have

$$\mathbb{P}[\mathcal{E}_4] \leq \frac{1}{r^2 + 1} + z$$

If \mathcal{E}_4 does not occur, we flip a coin with probability of heads $z - \mathbb{P}[\mathcal{E}_4]$, and let \mathcal{E}_5 be the event that the coin comes up heads. We make Π_2 take the transition $s_i \rightarrow \theta_i$ if either \mathcal{E}_4 or \mathcal{E}_5 occurs, which happens with probability

$$\mathbb{P}[\mathcal{E}_4] + \mathbb{P}[\mathcal{E}_5] = \frac{1}{r^2 + 1} + z$$

Additionally, we mark leaf j as the active leaf in θ_i . Finally, we make Π_2 take the transition $s_i \rightarrow \eta_i$ if none other transition has been taken, i.e. with probability

$$1 - \left(\frac{1}{r^2 + 1} + z + z \right) = \frac{r^2}{r^2 + 1} - 2 \cdot z$$

and mark leaf j as the active leaf in η_i . Additionally, we erase event $e_{t'}$ from π .

The desired result follows directly from the coupling process. \square

The following lemma establishes the forward bias on the Markov chain \mathcal{M} , i.e. given a current state s_i , the ratio of the probabilities of transitioning to state s_{i+1} over transitioning to state s_{i-1} .

Lemma 6. *For any $0 < i < n$, let x_{s_i} be the probability that a random walk on \mathcal{M} starting from state s_i transitions to state s_{i+1} before it transitions to state s_{i-1} . We have*

$$\frac{x_{s_i}}{1 - x_{s_i}} = \frac{\frac{r^2}{r^2+1} \cdot \beta - z \cdot \left(\frac{r^2}{r^2+1} + \beta - z \right)}{\frac{1}{r^2+1} \cdot \alpha + z \cdot \frac{r^2+2}{r^2+1}}$$

Proof. Given that a random walk is in state s_i (resp. η_i, β_i), $0 < i < n$, let x_{s_i} (resp. x_{η_i}, x_{β_i}) be the probability that the walk transitions to s_{i+1} before it transitions to s_{i-1} . Then we have the following system

$$\begin{aligned}
\left. \begin{aligned}
x_{s_i} &= \left(\frac{r^2}{r^2+1} - z \right) \cdot x_{\eta_i} + \frac{1}{r^2+1} \cdot x_{\beta_i} \\
x_{\eta_i} &= \beta - z + (1 - \beta + z) \cdot x_{s_i} \\
x_{\beta_i} &= (1 - \alpha - z) \cdot x_{s_i}
\end{aligned} \right\} \implies x_{s_i} = \frac{\left(\frac{r^2}{r^2+1} - z \right) \cdot (\beta - z)}{1 - \left(\frac{r^2}{r^2+1} - z \right) \cdot (1 - \beta + z) - \frac{1}{r^2+1} \cdot (1 - \alpha - z)} \\
\implies \frac{x_{s_i}}{1 - x_{s_i}} = \frac{\frac{r^2}{r^2+1} \cdot \beta - z \cdot \left(\frac{r^2}{r^2+1} + \beta - z \right)}{\frac{1}{r^2+1} \cdot \alpha + z \cdot \frac{r^2+2}{r^2+1}}
\end{aligned}$$

The desired result follows. \square

The following theorem captures the fixation probability on the metastar family.

Theorem 1. *Let G_m be a fixed graph and v the attachment vertex of G_m . Denote $p = \rho(r, G_m, v)$ and $\alpha = \rho^-(r, G_m, v)$ and $\beta = \rho^+(r, G_m, v)$. The fixation probability of a single mutant placed uniformly at random on $M_{N(n)}^{G_m}$ is*

$$\rho\left(r, M_{N(n)}^{G_m}\right) \geq p \cdot \frac{1 - r^{-2} \cdot (\alpha/\beta)}{1 - (r^{-2} \cdot (\alpha/\beta))^n} \cdot (1 + o(1)) \quad (9)$$

Proof. First, note that a mutant placed uniformly at random on $M_{N(n)}^{G_m}$ will be placed in a leaf with probability $1 - O(1/n)$. Then $\rho\left(r, M_{N(n)}^{G_m}\right)$ is lowerbounded by the probability of that mutant fixating in the initial leaf, times the probability that the mutants fixate in $M_{N(n)}^{G_m}$ starting from a mutant leaf. The former event occurs with probability $p = \rho(r, G_m, v)$. By Lemma 5, the probability of the latter event is lowerbounded by the probability ϕ that a random walk on \mathcal{M} starting from s_1 will result in s_n . Let x_{s_i} be the probability that a random walk on \mathcal{M} starting from state s_i transitions to state s_{i+1} before it transitions to state s_{i-1} , and $\gamma = \frac{x_{s_i}}{1-x_{s_i}}$. Using Lemma 6, the probability that a random walk of \mathcal{M} from s_1 will get absorbed in s_n is

$$\phi = \frac{1}{\sum_{i=0}^{n-1} \left(\frac{1-x_{s_i}}{x_{s_i}} \right)^i} = \frac{1}{\sum_{i=0}^{n-1} \gamma^{-i}} = \frac{1 - \gamma^{-1}}{1 - \gamma^{-n}} = \frac{1 - r^{-2} \cdot (\alpha/\beta)}{1 - (r^{-2} \cdot (\alpha/\beta))^n} \cdot (1 + o(1))$$

The desired result follows. \square

References

- [1] Martin A Nowak. *Evolutionary dynamics : exploring the equations of life*. Belknap Press of Harvard University Press, Cambridge, Massachusetts, 2006.