Effect of Graph Structures on Selection for a Model of a Population on an Undirected Graph

Wanting Chen
Advisor: Jason Schweinsberg

May 10, 2016

Abstract

This research focuses on analyzing selection amplifiers in population genetics. Since the structure of a population graph can influence selection, this paper focuses on finding undirected graphs that can amplify selection. To clearly demonstrate the idea, the paper will briefly discuss the basic Moran model at first. Then it will analyze the star graph as an example of selection amplifiers that can increase the fixation probability of mutants compared to the simple Moran model. Finally, it will discuss the specific kind of undirected graphs that amplify selection and have fixation probability close to 1.

1 Introduction

In population genetics, the Moran model is the most fundamental model to mimic how populations evolve over time considering the effect of selection. The Moran model assumes fixed population size of N and each individual in the population is classified as either a mutant or a non-mutant. Fitness measures the survival probability of each individual, and the fitness of mutants and non-mutants are always different. In different settings, the mutants may be more likely or less likely to be selected to reproduce and replace other individuals. Since originally the population has no mutants, we normalize the fitness of the non-mutant individuals in the population to be 1 and let the fitness of the mutants to be r. In this paper, since we are only interested in the case that mutants are advantageous, the fitness of mutants is always larger than 1, i.e. $r = 1 + s$ and $0 \leqslant s \leqslant 1$. When the fitness r is high, it is more likely for the mutant to replace other individuals. Each individual chooses to reproduce and replace a randomly chosen other individual in the population at rate equal to its fitness. For such an event to happen, we have to wait for an exponential distribution of time with rate r. Let states represent the number of mutants in the population, so state i is the circumstance that there are i mutants in the population. Suppose there are i mutants initially in the population, the rate for the model to go from state i to state $i+1$ and the rate to go from state i to state $i-1$ are as follows.

$$i \to i+1, \text{ at rate } \frac{ri(n-i)}{n},$$
$$i \to i-1, \text{ at rate } \frac{i(n-i)}{n}.$$

Thus, the rate at which the process leaves the state i is

$$q(i) = \frac{(1+r)i(n-i)}{n} = \frac{(2+s)i(n-i)}{n}. \tag{1}$$

In order to better present and analyze population structures, we will represent different population models using graphs $G = (V, E)$. The V stands for the set of vertices or nodes in the graph and the E stands for the set of edges connecting those nodes in the graph. When an individual reproduces, its offspring replaces the individual at a randomly chosen neighboring vertex. Since each node can be replaced by any of the other nodes in the population, the Moran model we discussed above can be represented by a complete graph with all nodes strongly connected by all possible edges. The mutants in the graph have rate r to reproduce and replace other nodes and the non-mutants have rate 1 of reproduction. The process of the Moran model is repeated until eventually the population has reached homogeneity, the state that all individuals are non-mutants or mutants. In either case we say that the population reaches fixation. In this paper, the mutant fixation refers to the probability that all individuals in the graph become mutants, if no further explanation.

We are interested in the effect of selection in such model, and in particular, we are interested in the case that there is only a single mutant in the population initially. According to Lieberman et al.[4], for a complete graph under the Moran model with size N, the probability that all nodes will eventually become mutants with the initial occurrence of a single mutant in the population is

$$\rho = \frac{1 - \frac{1}{r}}{1 - \frac{1}{r^N}}. \tag{2}$$

However, when the population graph is not complete, not all nodes are connected by all possible edges, so the fixation probability becomes difficult to study. One such example is the star graph.

2 Star Graph

Consider a star graph with 1 node in the center and n leaf nodes connected to the center and to no other vertex. In total, such a star graph has $N = n + 1$ nodes. In the following discussion, we will look at the node in the center and nodes on the spokes separately.

In 2008, M. Broom and J. Rychtar[1] have found the exact fixation probability of such star graph with n leaf nodes and fitness r to be

$$Q = \frac{\frac{n^2 r}{nr+1} + \frac{r}{r+n}}{(n+1)\left[1 + \frac{n}{n+r}\sum_{j=1}^{n-1}\left(\frac{n+r}{r(nr+1)}\right)^j\right]}. \tag{3}$$

In this paper, however, we will look at the fixation probability of the star graph by finding its upper and lower bound when n goes to infinity, and intuitively see why it has larger selection amplification effect than the Moran model.

Consider an advantageous mutant with fitness $r = 1 + s$ and $0 \leqslant s \leqslant 1$ appearing on one of the leaf nodes. We want to find the fixation probability of the star graph with such initial mutant. Let i be the number of mutants on the leaf nodes. There are only four events possible to happen: the possibility of having $i+1$ mutants and $i-1$ mutants on the leaf nodes in the next generation when the center is a mutant; and the possibility of having $i + 1$ mutants and $i - 1$ mutants on the leaf nodes in the next generation when the center is not a mutant. However, when the center node is a mutant, there is no way to make any mutants on the leaf nodes disappear in the next generation; when the center node is a non-mutant, there is no way for it to give birth

to any mutants at the leaf nodes in the next generation. Therefore, the above two events can be eliminated and the only two possible events that can happen with their rates are as follows.

$$i \to i+1, \text{ at rate } \frac{r(n-i)}{n} \text{ if the center is a mutant;}$$
$$i \to i-1, \text{ at rate } \frac{i}{n} \text{ if the center is a non-mutant.}$$

If the center is a mutant, it has rate r to infect all the other n nodes connected to it, since there are only $n-i$ non-mutants in the graph, the rate from state i to state $i+1$ is $\frac{r(n-i)}{n}$. Similarly, when the center is a non-mutant, it can only infect others to become non-mutants with rate 1. Since there are i out of n mutants that could be infected to become non-mutants, the rate from state i to state $i-1$ is $\frac{i}{n}$.

In order to understand the dynamics of the star graph, we divide the problem into a sequence of four-state Markov Chains. Let state 1 be the state that there are $i-1$ mutants on the spokes regardless of what the center node is. Let state 2 be the state that there are i mutants on the spokes and center of the star graph is mutant. Let state 3 be the state that there are i mutants on the spoke and the center of the star graph is non-mutant. Let state 4 be the state that there are $i+1$ mutants on the spoke regardless of what the center node is. So the transition rates are

State 3 $\to$ State 1, $\frac{i}{n}$;

State 2 $\to$ State 3, $n-i$;

State 3 $\to$ State 2, ri;

State 2 $\to$ State 4, $\frac{r(n-i)}{n}$.

Let $g(i)$ be the probability that the population reaches state 4 before state 1, starting from state i. We can obtain the following equations.

$$\begin{aligned} g(2) &= \frac{n-i}{\frac{r(n-i)}{n}+(n-i)}g(3) + \frac{\frac{r(n-i)}{n}}{\frac{r(n-i)}{n}+(n-i)}g(4) \\ &= \frac{n}{n+r}g(3) + \frac{r}{n+r}g(4) \ . \end{aligned}$$

Likewise,

$$g(3) = \frac{\frac{i}{n}}{\frac{i}{n}+ri}g(1) + \frac{ri}{ri+\frac{i}{n}}g(2) = \frac{1}{1+rn}g(1) + \frac{rn}{1+rn}g(2) \ .$$

Solve the above two equations with $g(1)=0$ and $g(4)=1$, we get

$$g(2) = \frac{r+r^2n}{n+r+r^2n} \text{ , the upper bound for reaching state 4 before state 1.}$$

$$g(3) = \frac{r^2n}{n+r+r^2n} \text{ , the lower bound for reaching state 4 before state 1.}$$

Let $f(i)$ be the probability that the population reaches state 1 before state 4. Then the lower bound for reaching state 1 before state 4 is $1 - g(2)$,

$$f(2) = \frac{n}{n + r + r^2 n} \ .$$

The upper bound for reaching state 1 before state 4 is $1 - g(3)$

$$f(3) = \frac{n + r}{n + r + r^2 n} \ .$$

Then, we will use these four bounds for the simple 4-state Markov Chain to calculate the upper and lower bound for the star graph.

Let $T_y = \min\{t : X_t = y\}$ be the first hitting time of state y. Let $h(i) = P_i(T_n < T_0)$ be the probability that the mutants fixate when there are initially i mutants in the graph. We will use the following standard result about asymmetric random walks from [6].

Lemma 1. *Let p be the probability that the number of mutants in the graph goes from i to $i+1$, and assume that this probability is the same for all i. Then*

$$P_i(T_n < T_0) = \frac{(\frac{1-p}{p})^i - 1}{(\frac{1-p}{p})^n - 1}. \tag{4}$$

Proof. Since p is the probability that the number of mutants in the graph goes from i to $i+1$, the probability that the number of mutants in the graph goes from i to $i-1$ is $1-p$. According to the first step conditioning method, the fixation probability with initial state i can be divided into two cases, the probability of fixation with initial state $i+1$ and probability of fixation with initial state $i-1$, with corresponding probability that goes from state i to each of them respectively.

$$h(i) = ph(i+1) + (1-p)h(i-1).$$

Rearrange the equation, we could get

$$\begin{aligned} ph(i) + (1-p)h(i) &= ph(i+1) + (1-p)h(i-1) \\ (1-p)[h(i) - h(i-1)] &= p[h(i+1) - h(i)] \\ h(i+1) - h(i) &= \frac{1-p}{p}[h(i) - h(i-1)]. \end{aligned}$$

Starting from state 0, it is impossible to get to state n, so $h(0) = 0$. Let $c = h(1)$.

$$\begin{aligned} h(j) &= \sum_{i=0}^{j-1} c\left(\frac{1-p}{p}\right)^i \\ &= c\frac{1 - (\frac{1-p}{p})^j}{1 - \frac{1-p}{p}}. \end{aligned}$$

Since $h(n) = 1$,

$$c = \frac{1 - \frac{1-p}{p}}{1 - (\frac{1-p}{p})^n},$$

then

$$h(1) = c = \frac{1 - \frac{1-p}{p}}{1 - (\frac{1-p}{p})^n}.$$

Thus,

$$\begin{aligned} P_i(T_n < T_0) &= \frac{1 - \frac{1-p}{p}}{1 - (\frac{1-p}{p})^n} \cdot \frac{1 - (\frac{1-p}{p})^i}{1 - \frac{1-p}{p}} \\ &= \frac{1 - (\frac{1-p}{p})^i}{1 - (\frac{1-p}{p})^n}. \end{aligned}$$

□

We can use this result to calculate the upper bound and lower bound of the fixation probability of a star graph. The upper bound can be calculated using the equations $g(2)$ and $f(2)$ that we previously obtained. So the probability of getting an additional mutant on the spoke is $\frac{r+r^2n}{n+r+r^2n}$ and the probability of removing an existing mutant on the spoke is $\frac{n}{n+r+r^2n}$. Then

$$\frac{1-p}{p} = \frac{\frac{n}{n+r+r^2n}}{\frac{r+r^2n}{n+r+r^2n}} = \frac{n}{r + r^2n} = \frac{1}{\frac{r}{n} + r^2}.$$

Let $u(i)$ be the upper bound of $h(i)$, then by Lemma 1, the upper bound of the fixation probability is

$$u(1) = \frac{1 - \frac{1}{\frac{r}{n}+r^2}}{1 - (\frac{1}{\frac{r}{n}+r^2})^n}.$$

When n goes to infinity, $\frac{r}{n}$ tends to 0. Then the upper bound of the fixation probability $u(1)$ tends to

$$u(1) \to 1 - \frac{1}{r^2}.$$

Similarly, we can calculate the lower bound of the fixation probability, $l(1)$, by equations $g(3)$ and $f(3)$. By Lemma 1, the lower bound of the fixation probability is

$$l(1) = \frac{1 - \frac{1+\frac{r}{n}}{r^2}}{1 - (\frac{1+\frac{r}{n}}{r^2})^n}.$$

When n goes to infinity, $\frac{r}{n}$ tends to 0, the lower bound of the fixation probability tends to

$$l(1) \to 1 - \frac{1}{r^2}.$$

Since when n goes to infinity, both of the upper bound and lower bound of the fixation probability goes to

$$1 - \frac{1}{r^2},$$

when n tends to infinity, the fixation probability of the star graph tends to

$$1 - \frac{1}{r^2}. \tag{5}$$

Comparing formulas (2) and (5), we can see that the fixation probability of the star graph is higher than the Moran model. Therefore, we can conclude that compared to the Moran model, the star graph has greater selection amplification effect.

3 Graph Description

In the literature, researchers have already found directed graphs such as superstars that have a large selection amplification effect which greatly increases fixation probability[3]. The initial mutation could happen randomly in any place in the graph to obtain a fixation probability that tends to 1. However, in this paper we are interested in finding undirected graphs that could amplify selection rate. We come up with a specific graph $G_{m,n}$ which has a node in the center with m complete graphs of size n connected to it. In order to reach mutant fixation, the original mutant has to start from the center of the graph. The mutant in the center can quickly infect all m nodes around it, and each of those m nodes can then infect all the other nodes in each complete graph. Our study shows that this undirected graph $G_{m,n}$ has fixation probability converging to 1 when the size of each complete graph n goes to infinity and m tends to infinity at an appropriate rate.

Theorem 2. *The undirected graph $G_{m,n}$ has fixation probability converging to 1 when n goes to infinity and $m = \log n$.*

4 Proof of Fixation

Firstly, we would like to compute the expected time to fixation of a complete graph under Moran model following the argument from Durrett's book [5]. Let τ be the time one complete graph fixates, which includes both the cases that all individuals are mutants and all are non-mutants. Let state number be the number of mutants in the graph at a particular time. Let S_j be the amount of time spent at state j before time τ. The expected fixation time with initial starting state i, denoted as E_i, is the sum of time spent at each state j. So

$$E_i[\tau] = \sum_{j=1}^{n-1} E_i[S_j] \ . \tag{6}$$

Lemma 3. *The expected time for a complete graph to fixate with initially one mutant in the graph can be upper bounded by $\frac{4}{s}(1 + \log n)$, i.e. $E_1[\tau] \leqslant \frac{4}{s}(1 + \log n)$.*

Proof. Let N_j be the number of visits to state j. Let $q(j) = \frac{(2+s)j(N-j)}{N}$ be the rate the chain leaves state j. Since each visit to j lasts an exponential amount of time with mean $\frac{1}{q(j)}$, we have

$$E_i[S_j] = \frac{1}{q(j)} E_i[N_j] \ . \tag{7}$$

Let $T_j = \min\{t : X_t = j\}$ be the time the first visit to state j and R_j=min $\{t : X_t = j$ and $X_s \neq j$ for some $s < t\}$ be the time of the first return to state j, then according to [5],

$$E_i[N_j] = \frac{P_i(T_j < \infty)}{P_j(R_j = \infty)} \ .$$

When $n = j$, $0 \leqslant i \leqslant j$, according to Lemma 1,

$$P_i(T_j < T_0) = \frac{1 - (\frac{1-p}{p})^i}{1 - (\frac{1-p}{p})^j} \ , \tag{8}$$

$$P_i(T_0 < T_j) = 1 - P_i(T_j < T_0) = \frac{(\frac{1-p}{p})^i - (\frac{1-p}{p})^j}{1 - (\frac{1-p}{p})^j}. \tag{9}$$

Likewise, when $j \leqslant i \leqslant n$,

$$P_i(T_n < T_j) = \frac{1 - (\frac{1-p}{p})^{i-j}}{1 - (\frac{1-p}{p})^{n-j}} \ , \tag{10}$$

$$P_i(T_j < T_n) = 1 - P_i(T_n < T_j) = \frac{(\frac{1-p}{p})^{i-j} - (\frac{1-p}{p})^{n-j}}{1 - (\frac{1-p}{p})^{n-j}}.$$

Under our Moran model,

$$\frac{1-p}{p} = \frac{(n-j)\frac{j}{n}}{\frac{rj(n-j)}{n}} = \frac{1}{r} = \frac{1}{1+s}.$$

Using equations (9) and (10),

$$\begin{aligned}
P_j(R_j = \infty) &= \frac{1+s}{2+s} \cdot P_{j+1}(T_n < T_j) + \frac{1}{2+s} \cdot P_{j-1}(T_0 < T_j) \\
&= \frac{1+s}{2+s} \cdot \frac{1 - (\frac{1-p}{p})^{j+1-j}}{1 - (\frac{1-p}{p})^{n-j}} + \frac{1}{2+s} \cdot \frac{(\frac{1-p}{p})^{j-1} - (\frac{1-p}{p})^j}{1 - (\frac{1-p}{p})^j} \\
&= \frac{1+s}{2+s} \cdot \frac{1 - (\frac{1}{1+s})}{1 - (\frac{1}{1+s})^{n-j}} + \frac{1}{2+s} \cdot \frac{(\frac{1}{1+s})^{j-1} - (\frac{1}{1+s})^j}{1 - (\frac{1}{1+s})^j} \\
&= \frac{1}{2+s} \cdot \left[\frac{1+s-1}{1 - (\frac{1}{1+s})^{n-j}} + \frac{1+s-1}{(1+s)^j - 1} \right] \\
&= \frac{s}{2+s} \cdot \left[\frac{1}{1 - (\frac{1}{1+s})^{n-j}} + \frac{1}{(1+s)^j - 1} \right] \\
&= \frac{s}{2+s} \cdot \left[\frac{(1+s)^{n-j}}{(1+s)^{n-j} - 1} + \frac{1}{(1+s)^j - 1} \right].
\end{aligned}$$

Since $\frac{(1+s)^{n-j}}{(1+s)^{n-j}-1} \geqslant 1$ and $\frac{1}{(1+s)^j-1} \geqslant 0$,

$$\frac{(1+s)^{n-j}}{(1+s)^{n-j} - 1} + \frac{1}{(1+s)^j - 1} \geqslant 1.$$

Therefore,

$$P_j(R_j = \infty) \geqslant \frac{s}{2+s}. \tag{11}$$

According to equation (8),

$$P_1(T_j < T_0) = \frac{1-(\frac{1-p}{p})}{1-(\frac{1-p}{p})^j} = \frac{1-(\frac{1}{1+s})}{1-(\frac{1}{1+s})^j} = \frac{(1+s)^j-(1+s)^{j-1}}{(1+s)^j-1} = \frac{s(1+s)^{j-1}}{(1+s)^j-1}.$$

Since $P_1(T_j < T_0) = P_1(T_j < \infty)$,

$$P_1(T_j < \infty) = \frac{s(1+s)^{j-1}}{(1+s)^j-1}. \tag{12}$$

According to equations (11) and (12),

$$\begin{aligned} E_1[N_j] &= \frac{P_1(T_j < \infty)}{P_j(R_j = \infty)} \\ &\leqslant \frac{s(1+s)^{j-1}}{(1+s)^j-1} \cdot \frac{2+s}{s} \ . \end{aligned}$$

Since $q(j) = \frac{(2+s)j(n-j)}{n}$,

$$E_1[S_j] = \frac{1}{q(j)} E_1[N_j] \leqslant \frac{n(1+s)^{j-1}}{j(n-j)[(1+s)^j-1]} \ . \tag{13}$$

Then using (7) and (13),

$$\begin{aligned} E_1[\tau] &\leqslant \sum_{j=1}^{n-1} \frac{n(1+s)^{j-1}}{j(n-j)[(1+s)^j-1]} \\ &= \sum_{j=1}^{\lfloor \frac{n}{2} \rfloor} \frac{n}{j(n-j)} \frac{(1+s)^{j-1}}{[(1+s)^j-1]} + \sum_{j=\lfloor \frac{n}{2} \rfloor+1}^{n-1} \frac{n}{j(n-j)} \frac{(1+s)^{j-1}}{[(1+s)^j-1]} \ . \end{aligned}$$

Among $1 \leqslant j \leqslant \lfloor \frac{n}{2} \rfloor$, $\frac{n}{n-j}$ can be upper bounded when $j = \lfloor \frac{n}{2} \rfloor$ and among $\lfloor \frac{n}{2} \rfloor + 1 \leqslant j \leqslant n-1$, $\frac{n}{j}$ can be upper bounded when $j = \lfloor \frac{n}{2} \rfloor + 1$. Also $\frac{(1+s)^{j-1}}{(1+s)^j-1} = \frac{1}{(1+s)-(1+s)^{-j+1}}$ can be upper bounded by $\frac{1}{s}$,

$$\begin{aligned} E_1[\tau] &\leqslant 2 \sum_{j=1}^{\lfloor \frac{n}{2} \rfloor} \frac{1}{sj} + 2 \sum_{j=\lfloor \frac{n}{2} \rfloor+1}^{n-1} \frac{1}{n-j} \frac{1}{s} \\ &= \frac{2}{s} \Big[\sum_{j=1}^{\lfloor \frac{n}{2} \rfloor} \frac{1}{j} + \sum_{j=\lfloor \frac{n}{2} \rfloor+1}^{n-1} \frac{1}{n-j} \Big] \\ &\leqslant \frac{4}{s}(1+\log n), \text{ using the fact that } \sum_{j=1}^{n} \frac{1}{j} \leqslant 1+\log n. \end{aligned}$$

□

The above upper bound we calculated is the upper bound for the expected time for a particular complete graph to fixate with initially 1 mutant. For our graph $G_{m,n}$, the center node is the only initial mutant and it takes time for it to infect all the complete graphs connected to it. On average, the expected time for the center to infect a particular complete graph is $\frac{m}{r}$, since there are m such complete graphs and the rate of infection of the center node is r.

For a complete graph, if all mutants die out, there is no way for it to reach mutant fixation. However, one benefit of our graph $G_{m,n}$ is that it provides multiple chances or attempts for each complete graph to fixate. If one complete graph has no mutants, as long as the center is still a mutant, the complete graph can be reinfected. Then we can begin our proof for Theorem 2.

Proof of Theorem 2. According to Lemma 3, for each attempt, the upper bound for the expected time each complete graph takes to fixation is $\frac{4}{s}(1+\log n)$, which considers both the cases that all mutants die out and all mutants fixate, and the expected time for the center to infect a particular complete graph is $\frac{m}{r}$, so the expected time for each attempt a particular graph takes is upper bounded by

$$\frac{m}{r}+\frac{4}{s}(1+\log n).$$

Since the probability of fixation for each complete graph is

$$\frac{1-\frac{1}{r}}{1-\frac{1}{r^n}}\ ,$$

on average, the expected number of attempts needed for each complete graph to fixate is

$$\frac{1}{\text{probability of fixation}}=\frac{1-\frac{1}{r^n}}{1-\frac{1}{r}}.$$

Let T_i be the time for a complete graph i to fixate. With m such complete graphs, the expected time for a particular complete graph with size n to fixate is

$$E[T_i]\leqslant\left[\frac{m}{r}+\frac{4}{s}(1+\log n)\right]\cdot\frac{1-\frac{1}{r^n}}{1-\frac{1}{r}}\ .$$

Then by Markov's inequality, the probability that one graph takes longer than time t to fixate is

$$\begin{aligned}P(T_i>t)&\leqslant\frac{E[T_i]}{t}\\&\leqslant\frac{1}{t}\left\{\left[\frac{m}{r}+\frac{4}{s}(1+\log n)\right]\cdot\frac{1-\frac{1}{r^n}}{1-\frac{1}{r}}\right\}.\end{aligned}$$

Then by Boole's inequality, the probability that at least one graph takes longer than time t to fixate is

$$\begin{aligned}P\left(\bigcup_{i=1}^{m}\{T_i>t\}\right)&\leqslant\sum_{i=1}^{m}\frac{1}{t}\left\{\left[\frac{m}{r}+\frac{4}{s}(1+\log n)\right]\cdot\frac{1-\frac{1}{r^n}}{1-\frac{1}{r}}\right\}\\&\leqslant\frac{m}{t}\left\{\left[\frac{m}{r}+\frac{4}{s}(1+\log n)\right]\cdot\frac{1-\frac{1}{r^n}}{1-\frac{1}{r}}\right\}.\end{aligned}$$

To simplify the inequality, there exist constants c and d such that

$$P\Big(\bigcup_{i=1}^{m}\{T_i > t\}\Big) \leqslant \frac{c}{t}\Big[m^2 + m(1+\log n)\Big]$$
$$\leqslant \frac{d}{t}\Big[m^2 + m\log n\Big].$$

The above computation assumes that the center is always mutant. However, the central mutant could also die and be replaced by a non-mutant. So we would like to investigate the probability that the center dies before time t and compares it with the probability that there is at least one complete graph has not fixated by that time t, which we obtained above.

The rate of a non-mutant infection is 1 and the probability for the center to be infected by a non-mutant in a unit time, the probability the center dies, is $\frac{m}{n}$, since there are m complete graphs connected to the center, each with probability $\frac{1}{n}$ to infect the center node. Let D be the time the center dies. The probability that the center dies before time t is

$$P(D \leqslant t) \leqslant \frac{m}{n}t.$$

In order to make both equations above to be small, pick $m = \log n$, $t = 2(\log n)^3$. Let T be the time that all complete graphs fixate. Then

$$P(D \leqslant t) \leqslant \frac{2(\log n)^4}{n}$$

and

$$P(T > t) \leqslant d\frac{2(\log n)^2}{2(\log n)^3}$$
$$\leqslant d\frac{1}{\log n}.$$

We can see that both of the equations go to 0 when n goes to infinity. Thus, when n goes to infinity, by time t, the center of the graph dies with probability converging to 0, and all graphs fixate with probability converging to 1, which means that our graph $G_{m,n}$ will fixate with probability tending to 1. Thus, the undirected graph $G_{m,n}$ amplifies selection and has fixation probability close to 1 when the size of each sub-graph, n, goes to infinity. □

When $s = 0$, $r = 1$, there is no selection in the population. We are interested in the fixation probability of our graph $G_{m,n}$ for that case to guarantee that our graph can truly amplify selection rate.

Proposition 4. *If there is no selection, i.e. $s = 0$ and $r = 1$, the fixation probability of the graph $G_{m,n}$ goes to 0 when n goes to infinity.*

Proof. Given an undirected graph $G = (V, E)$, let V denote the set of all vertices and E denote the set of all edges in the graph. Also, let C be the set of all mutant nodes and $C \subseteq V$. Let P_C be the probability of mutant fixation, and d_m be the degree for each node m, which is the

number of nodes connected to m. According to the formula given by Broom et al.[2], for fitness r=1,

$$P_C = \frac{\sum_{i\in C} d_i^{-1}}{\sum_{k\in V} d_k^{-1}}. \tag{14}$$

In our graph $G_{m,n}$, the center node has m nodes connected to it so the degree of the center node is m. Since initially there is only one mutant in the graph $G_{m,n}$ which is the center node, the summation of the inverse of the degrees of all mutants is

$$\sum_{i\in C} d_i^{-1} = \frac{1}{m}.$$

In addition, there are m complete graphs connected with the center node and the degree for each node in each complete graph is n. Meanwhile, there are n nodes in each complete graph and there are m such complete graphs. For all nodes that are not connected to the center in a complete graph, their degree is $n-1$; for the node that connected to the center, its degree is n. So the summation of the inverse of the degrees of all nodes in graph $G_{m,n}$ is

$$\sum_{k\in V} d_k^{-1} = \frac{1}{m} + m\left[m\frac{1}{n} + (n-m)\frac{1}{n-1}\right],$$

According to formula (14),

$$\begin{aligned} P_C &= \frac{\sum_{i\in C} d_i^{-1}}{\sum_{k\in V} d_k^{-1}} \\ &= \frac{\frac{1}{m}}{\frac{1}{m} + m\left[m\frac{1}{n} + (n-m)\frac{1}{n-1}\right]} \\ &= \frac{1}{1 + m^2\left[\frac{m}{n} + \frac{n-m}{n-1}\right]} \end{aligned}$$

which can be bounded above by

$$P_C = \frac{1}{1+m^2}. \tag{15}$$

When n goes to infinity, $m = \log n$ goes to infinity, and the fixation probability P_C also goes to 0, which means that $G_{m,n}$ cannot reach mutant fixation without selection. Therefore, our graph $G_{m,n}$ is indeed a selection amplifier. □

References

[1] Broom, M., Rychtar, J. (2008): An analysis of the fixation probability of a mutant on special classes of non-directed graphs. *Proceedings of the Royal Society A* **464**, 2609-2627. doi:10.1098/rspa.2008.0058

[2] Broom, M., Hadjichrysanthou, C., Rychtar, J. and Stadler, B.T. (2010): Two results on evolutionary processes on general non-directed graphs. *Proceedings of the Royal Society A*. doi:10.1098/rspa.2010.0067

[3] Jamieson-Lane, A., Hauert, C. (2015): Fixation probabilities on superstars, revisited and revised. *Journal of Theoretical Biology* Volume **382**, 4456. doi:10.1016/j.jtbi.2015.06.029

[4] Lieberman, E., Hauert, C., Nowak, M.A.(2005): Evolutionary Dynamics on Graphs. *Nature* **433**, 312316

[5] Rick Durrett(2008): Probability Models for DNA Sequence Evolution. *Springer*

[6] Rick Durrett(1999): Essentials of Stochastic Processes. *Springer*