Supplementary Information for Universality of fixation probabilities in randomly structured populations

Ben Adlam\textsuperscript{1,2} and Martin A. Nowak\textsuperscript{1,3,4,*}

\textsuperscript{1}Program for Evolutionary Dynamics, \textsuperscript{2}School of Engineering and Applied Science, \textsuperscript{3}Department of Mathematics, \textsuperscript{4}and Department of Organismic and Evolutionary Biology, \textsuperscript{4}Harvard University, Cambridge MA 02138, USA

1 The robust isothermal theorem

The Moran process on graphs is the standard model for population structure in evolutionary dynamics.\textsuperscript{1,2}

The process is defined for a directed, weighted graph, \( G \equiv G_n = (V_n, W_n) \), where \( V_n := \left[ n \right] \) and \( W_n := [w_{ij}] \) is a stochastic matrix of edge weights. The process, denoted by \( X_t \), is Markovian with state space \( 2^V \), where each \( X_t \) is the subset of \( V \) consisting of all the vertices occupied by mutants at time \( t \). At time 0 a mutant is placed at one of the vertices uniformly at random or formally,

\[
P[X_0 = S] = \begin{cases} n^{-1} & \text{if } |S| = 1 \\ 0 & \text{otherwise} \end{cases}.
\] (1.1)

Then at each subsequent time step exactly one vertex is chosen randomly, proportional to its fitness, for reproduction: so the probability of choosing a particular wild type vertex is \( 1/(n - |X_t| + r|X_t|) \) and the probability of choosing a particular mutant vertex is \( r/(n - |X_t| + r|X_t|) \). An edge originating from the chosen vertex is then selected randomly with probability equal to its edge weight, which is well defined since \( W \) is stochastic, and the vertex at the destination of the edge takes on the type of the vertex at the origin of the edge.

Typically, there are exactly two absorbing states, \( X_t = \emptyset \) and \( X_t = V \), corresponding to the wild-type fixing in the population and the mutant fixing in the population respectively. Thus, almost surely, one of these two absorbing states is reached in finite time. The probability that the process reaches \( V \) and not \( \emptyset \) is called the fixation probability and for a graph \( G \) we denote its fixation probability for a mutant of fitness \( r > 0 \) by

\[
\rho_G(r) := P[X_t = V \text{ for some } t \geq 0].
\] (1.2)

A fundamental point of comparison is the fixation probability \( \rho_{M_n} \) for a well-mixed population structure, where the graph structure is given by

\[
w_{ij} := \begin{cases} (n - 1)^{-1} & \text{if } i \neq j \\ 0 & \text{if } i = j \end{cases}
\] (1.3)

and M stands for “Moran” or “mixed.” An easy calculation using recurrence equations shows that

\[
\rho_{M_n}(r) = \frac{1 - r^{-1}}{1 - r^{-n}}.
\] (1.4)
Graphs with exactly the same fixation probability as \( \rho_{M_n} \) are classified by the isothermal theorem, which gives sufficient conditions for a general graph \( G \) to have the same fixation probability as \( \rho_{M_n} \).

In this section we derive a generalization of the isothermal theorem and throughout we shall require that the matrix \( W \) is stochastic—that is, the row sums are all equal to 1:

\[
\sum_{j=1}^{n} w_{ij} = 1,
\]

for all \( i \in V \). Any graph with nonnegative edge weights can be normalized to produce a graph with a stochastic \( W \), so long as each row has a nonzero entry, without changing the behavior of the process as defined above. A graph \( G \) is called isothermal if all the column sums of \( W \) are identical—that is, \( W_1 = \cdots = W_n = 1 \), where

\[
W_j := \sum_{i=1}^{n} w_{ij},
\]

or equivalently, \( W \) is doubly stochastic.

For all \( S \subseteq V \), define

\[
w_{O}(S) := \sum_{i \in S} \sum_{j \notin S} w_{ij} \quad \text{and} \quad w_{I}(S) := \sum_{i \notin S} \sum_{j \in S} w_{ij}
\]

as the sum of the edge weights leaving \( S \) and entering \( S \) respectively. Then an easy calculation shows that a graph is isothermal if and only if

\[
w_{O}(S) = w_{I}(S)
\]

for all \( \emptyset \neq S \subseteq V \). The later condition and its equivalence to isothermality is at the core of the proof of the isothermal theorem. The term “isothermal” originates from an interpretation of the sum of the ingoing edge weights as temperature, with “hotter” vertices changing more frequently in the Moran process. Thus a graph satisfying (1.8) is isothermal because the ingoing and outgoing temperatures are equal and all subsets \( S \) are in “thermal equilibrium.” We now restate the forward direction of the original isothermal theorem.

**Theorem 1.1 (Isothermal Theorem).** Suppose that a graph \( G \) is isothermal, then the fixation probability of a randomly placed mutant of fitness \( r \) is equal to \( \rho_{M_n}(r) \).

We ask, can we relax the assumptions of Theorem 1.1? That is, perhaps an approximate result can be obtained for \( W \) that are only approximately doubly stochastic in the following sense:

\[
|W_j - 1| \leq \varepsilon
\]

for all \( j \in V \) and some small quantity \( \varepsilon \). However, the graph \( G_{\varepsilon} = (V, W) \), where \( 0 < \varepsilon < 1 \) and

\[
W = \begin{bmatrix}
0 & 1 - \varepsilon & 0 & \varepsilon \\
1 - \varepsilon & 0 & \varepsilon & 0 \\
0 & \varepsilon^2 & 0 & 1 - \varepsilon^2 \\
\varepsilon^2 & 0 & 1 - \varepsilon^2 & 0
\end{bmatrix},
\]

(1.10)
shows we cannot, since as $\varepsilon \to 0$

$$\frac{w_O(\{1, 2\})}{w_I(\{1, 2\})} = \frac{2\varepsilon}{2\varepsilon^2} = \varepsilon^{-1} \to \infty. \quad (1.11)$$

That is, $W$ is approximately doubly stochastic, but the ratio of the outgoing and ingoing edge weights is unbounded for some subset $S$. Moreover, it is easy to show that the fixation probability is such that

$$\lim_{\varepsilon \to 0} \rho_{G_\varepsilon}(r) = \frac{1}{2} \frac{1 - r^{-1}}{1 - r^{-2}}, \quad (1.12)$$

which is far from $\rho_{M_n}(r) = (1 - r^{-1})/(1 - r^{-4})$. Conditioning on whether the process starts at vertex 1 or 3, the difference is even more pronounced and given by

$$\lim_{\varepsilon \to 0} \rho^1_{G_\varepsilon}(r) = \frac{1 - r^{-1}}{1 - r^{-2}} \text{ and } \lim_{\varepsilon \to 0} \rho^2_{G_\varepsilon}(r) = 0 \quad (1.13)$$

respectively. Thus, we need stronger assumptions for our theorem which we state now.

**Theorem 1.2 (Robust Isothermal Theorem).** Fix $0 \leq \varepsilon < 1$. Let $G_n = (V_n, W_n)$ be a connected graph. If for all nonempty $S \subset V_n$ we have

$$\left| \frac{w_O(S)}{w_I(S)} - 1 \right| \leq \varepsilon, \quad (1.14)$$

then

$$\sup_{r > 0} |\rho_{M_n}(r) - \rho_{G_n}(r)| \leq \varepsilon. \quad (1.15)$$

**Proof.** To briefly outline the proof, we begin by projecting the process from $X_t$ to $|X_t|$. Next we consider the ratio of the probability of increasing the number of mutants to the probability of decreasing the number of mutants. By bounding this ratio we can use a coupling argument to establish that the fixation probability of the process is close to $\rho_{M_n}$. Finally, we use the mean value theorem and smoothness properties of $\rho_{M_n}$ to simplify our bound and obtain the result.

Just as in the proof of the original isothermal theorem, we make the projection of the state space of all subsets of $V$, which records exactly which vertices are mutants, to the simpler state space $\{0, 1, \ldots, n\}$, which records only the number of mutants. The problem with making this projection in general is that the transition probabilities from one subset to another can depend on the structure of a subset not merely the number of mutants. However, it is clear that the only quantities which affect the fixation probability are the ratios of the probability of increasing the number of mutants to the probability of decreasing the number of mutants in a particular state $S$.

Define $p_+(S)$ and $p_-(S)$ as the probability that in the next step the number of mutants in the population increases and decreases by one respectively. Note these quantities do not sum to 1, as the number of mutants may remain constant. Thus

$$p_+(S) = \frac{w_O(S)r}{w_O(S)r + w_I(S)} \text{ and } p_-(S) = \frac{w_I(S)}{w_O(S)r + w_I(S)}, \quad (1.16)$$
which gives, when the two equations are divided,

\[ \frac{p_+(S)}{p_-(S)} = \frac{w_0(S)}{w_I(S)}. \] (1.17)

By assumption (1.14),

\[ r(1 - \varepsilon) \leq \frac{p_+(S)}{p_-(S)} \leq r(1 + \varepsilon). \] (1.18)

This states that the ratio of the probabilities of increasing to decreasing the number of mutants in any state \( S \) is approximately proportional to \( r \).

If for some graph \( G' = (V', W') \) we have \( \frac{p_+(S)}{p_-(S)} = r(1 \pm \varepsilon) \) for all \( S \subseteq V' \), then by the standard result for fixation probabilities in birth-death processes, its fixation probability is given by

\[ \rho_{G'}(r) = \rho_{M_n}(r \pm r\varepsilon) = \frac{1 - (r \pm \varepsilon r)^{-1}}{1 - (r \pm \varepsilon r)^{-n}}. \] (1.19)

From (1.18) and (1.19) we would like to conclude that

\[ \rho_{M_n}(r - r\varepsilon) = \frac{1 - (r - \varepsilon r)^{-1}}{1 - (r - \varepsilon r)^{-n}} \leq \rho_{G'}(r) \leq \frac{1 - (r + \varepsilon r)^{-1}}{1 - (r + \varepsilon r)^{-n}} = \rho_{M_n}(r + r\varepsilon). \] (1.20)

The upper bound is given by taking the maximum allowed value for the probability of increasing the number of mutants relative to the probability of decreasing the number of mutants. For the lower bound we use the opposite.

This intuitive result can be proved with a coupling argument. We can couple the Moran process \( X_t \) of a mutant of fitness \( r \) on \( G \) with another process \( Y \) defined as follows: \( Y \) has state space \( \{0, \ldots, n\} \) (with 0 and \( n \) absorbing) and \( Y \) starts at 1. We couple \( Y \) to \( X_t \) as follows:

1. if \( |X_t| \) decreases by 1, then \( Y \) must also decrease by 1;
2. if \( |X_t| \) increases by 1, then independently \( Y \) increases by 1 with probability

\[ \frac{p_+(S) + p_-(S)}{p_+(S)} \frac{r(1 - \varepsilon)}{1 + r(1 - \varepsilon)} \] (which is less than or equal to 1 by assumption (1.18)), else \( Y \) decreases by 1;
3. otherwise \( Y \) remains constant.

Note that marginally \( Y \) is a simple random walk on \( \{0, \ldots, n\} \) with forward bias \( r(1 - \varepsilon) \), since by the law of total probability, defining \( Y^\pm \) as the event that the \( Y \) changes by \( \pm 1 \) the next time it changes, we see

\[ \mathbb{P}[Y^+] = \mathbb{P}[Y^+ | X_{t+1}] = |X_t| + 1 \mathbb{P}[|X_{t+1}| = |X_t| + 1] \] (1.22)

\[ = \frac{p_+(X_t) + p_-(X_t)}{p_+(X_t)} \frac{r(1 - \varepsilon)}{1 + r(1 - \varepsilon)} p_+(X_t) \] (1.23)

\[ = (p_+(X_t) + p_-(X_t)) \frac{r(1 - \varepsilon)}{1 + r(1 - \varepsilon)}. \] (1.24)
\[ P[Y^-] = P[Y^- \mid |X_{t+1}| = |X_t| + 1] \cdot P[|X_{t+1}| = |X_t| + 1] \]  
\[ + P[Y^- \mid |X_{t+1}| = |X_t| - 1] \cdot P[|X_{t+1}| = |X_t| - 1] \]  
\[ = \left( 1 - \frac{p_+(X_t) + p_-(X_t)}{p_+(X_t)} \frac{r(1 - \varepsilon)}{1 + r(1 - \varepsilon)} \right) p_+(X_t) + p_-(X_t) \]  
\[ = \frac{p_+(X_t) + p_-(X_t)}{1 + r(1 - \varepsilon)}. \]  
\[ (1.25) \]

Thus the probability that \( Y \) reaches \( n \) before it reaches \( 0 \) is given by
\[ \frac{1 - (r - \varepsilon r)^{-1}}{1 - (r - \varepsilon r)^{-n}}. \]  
\[ (1.29) \]

However, because the processes are coupled we have \( Y \leq |X_t| \) and thus if \( Y = n \), then the mutant has fixed in the process \( X_t \). Equation (1.29) immediately implies the lower bound in (1.20). A similar coupling yields the upper bound. Thus
\[ \rho_{M_n}(r - r\varepsilon) - \rho_{M_n}(r) \leq \rho_G(r) - \rho_{M_n}(r) \leq \rho_{M_n}(r + r\varepsilon) - \rho_{M_n}(r). \]  
\[ (1.30) \]

By the mean value theorem,
\[ \frac{\rho_{M_n}(r + r\varepsilon) - \rho_{M_n}(r)}{\varepsilon} \leq \frac{r\varepsilon}{\varepsilon} \sup_{r \leq x \leq r + r\varepsilon} \left| \rho'_{M_n}(x) \right| = r \sup_{r \leq x \leq r + r\varepsilon} \left| \rho'_{M_n}(x) \right| \]  
\[ (1.31) \]
and
\[ \frac{\rho_{M_n}(r) - \rho_{M_n}(r - r\varepsilon)}{\varepsilon} \geq -r \sup_{r - r\varepsilon \leq x \leq r} \left| \rho'_{M_n}(x) \right|. \]  
\[ (1.32) \]
Thus, it is sufficient to show for all \( r > 0 \)
\[ \sup_{n \geq 2} \left| \rho'_{M_n}(r) \right| \leq r^{-1}. \]  
\[ (1.33) \]

We note that this is not an optimal bound, however, it suffices for our applications. Calculating, one finds
\[ \rho'_{M_n}(r) = \frac{r^{n-2}(r^n - nr + n - 1)}{(r^n - 1)^2}. \]  
\[ (1.34) \]

First, when \( r \geq 1 \) we prove the stronger claim
\[ \frac{r^{n-2}(r^n - nr + n - 1)}{(r^n - 1)^2} \leq r^{-2}, \]  
\[ (1.35) \]
by noting the above is equivalent to
\[ (r - 1) \left( nr^n - \sum_{k=0}^{n-1} r^k \right) \geq 0, \]  
\[ (1.36) \]
which is true since \( r \geq 1 \). Similarly, one can prove
\[
\frac{r^{n-2} (r^n - nr + n - 1)}{(r^n - 1)^2} < 1
\] (1.37)
when \( r < 1 \). Equations (1.35) and (1.37) imply \( \sup_{n \geq 2} |\rho_{M_n}(r)| \leq r^{-1} \).

Therefore, we may conclude
\[
\sup_{r > 0} |\rho_G(r) - \rho_{M_n}(r)| \leq \varepsilon, \quad (1.38)
\]
which completes the proof. ■

Note that one can sometimes do slightly better (depending on the relative sizes of \( \varepsilon \) and \( n^{-1} \)), by showing
\[
\sup_{r \leq 1} |\rho_G_n(r) - \rho_{M_n}(r)| < \frac{2}{n} \quad (1.39)
\]
for \( \varepsilon \) small enough, using the triangle inequality and the facts that \( \rho_{M_n} \) is increasing and continuous, but again this is not important for our applications.

Theorem 1.2 is actually slightly stronger than stated, and thus we can draw a slightly stronger conclusion in Theorem 2.1. We may conclude that the fixation probability of a mutant of fitness \( r \) originating at a particular vertex\(^3\) satisfies the bound in (1.15), for exactly the same reason as in the proof of the original isothermal theorem—the bound (1.14) is for all subsets \( S \). Therefore, \textit{a fortiori}, a mutant can be started with any probability vector on the vertices (not merely uniform) and its fixation probability will still satisfy (1.15). This observation is borne through simulations too (Figure 1).

2 Evolution on random graphs

In this section we prove Theorem 2.1.

**Theorem 2.1.** Let \( (G_n)_{n \geq 1} \) be a family of random graphs as in Definition 2.4. Then there is a constant \( C > 0 \), not dependent on \( n \), such that the fixation probability of a randomly placed mutant of fitness \( r > 0 \) satisfies
\[
|\rho_{G_n}(r) - \rho_{M_n}(r)| \leq \frac{C (\log n)^{C + C \xi}}{\sqrt{n}} \quad (2.1)
\]
uniformly in \( r \) with probability greater than
\[
1 - \exp \left( -\nu (\log n)^{1+\xi} \right), \quad (2.2)
\]
for positive constants \( \xi \) and \( \nu \).

To do this we need to apply Theorem 1.2 to our random graphs by showing that its assumptions hold with high probability. We do so in several steps. First, we define precisely generalized Erdős-Rényi random graphs in Definition 2.4 and outline the necessary assumptions on the distribution of the edge weights. After reviewing some notation, we introduce an event \( \Omega \), on which
Figure 1: Fixation probability does not depend on starting location. We conducted trials where the fixation probability of a mutant of fitness $r$ starting at vertex $i$ was estimated with the Monte Carlo method of $10^4$ samples for several values of $0 \leq r \leq 10$ on a Erdős-Rényi random graph. The fixation probability was similar regardless of starting vertex, and in particular, showed no correlation with vertex temperature. We plot the Moran fixation probability (1.4) and use the error bars to illustrate the minimum and maximum empirical fixation probabilities obtained from starting at any particular vertex.

the graphs are well behaved, and show that $\Omega$ has high probability in Lemma 2.6. Then the general idea is to use large deviation estimates and concentration inequalities to show that with high probability the quantity (1.14) can be controlled. We bound both the numerator (Lemma 2.7) and denominator (Lemma 2.8) of

$$\frac{|w_\Omega(S) - w_1(S)|}{|w_1(S)|} = \frac{|w_\Omega(S)|}{w_1(S)} - 1$$

(2.3)

for all $S$, then we put everything together to prove Theorem 2.1.

Remark 2.2 (Notation). We use the large constant $C > 0$ and the small constant $c > 0$, which do not depend on the size of the graph $n$ but can depend on the distribution as outlined in Definition 2.4. We allow the constant $C$ to increase or the constants $c$, $\nu$, and $\xi$ to decrease from line to line without noting it or introducing a new notation, sometimes we even absorb other constants such as $p$, $p'$, and $\mu_1$ without noting it; as is clear from the proof, this only happens a finite number of times, and thus we end with constants $C > 0$, $c > 0$, $\nu > 0$, and $\xi > 0$.

We also make use of standard order notation for functions, $o(\cdot)$, $\mathcal{O}(\cdot)$, and $\cdot \gg \cdot$, all of which are used with respect to $n$. Moreover, in some sums it is useful to exclude particular summands, e.g.

$$\sum_{1 \leq j \leq n} \cdot \equiv \sum_{\substack{1 \leq j \leq n \\ j \notin S}} \cdot$$

(2.4)

for $S \subset V$. We abbreviate $(\{i\})$ and $(\{i, k\})$ as $(i)$ and $(i, k)$.
REM 2.3 (HIGH PROBABILITY EVENTS). We say that an $n$-dependent event $E$ holds with high probability if, for constants $\xi > 0$ and $\nu > 0$ which do not depend on $n$,

$$\mathbb{P}[E^c] \leq e^{-\nu (\log n)^{1+\xi}}$$

(2.5)

for $n \geq n_0(\nu, \xi)$. Moreover, we say an event $E$ has high probability on another event $E_0$ if

$$\mathbb{P}[E_0 \cap E^c] \leq e^{-\nu (\log n)^{1+\xi}}$$

(2.6)

In particular, this has the property that the intersection of polynomially many (in $n$, say $K n^K$ for some constant $K > 0$) events of high probability is also an event of high probability: by the union bound,

$$\mathbb{P} \left[ \bigcup_{i=1}^{Kn^K} E_i^{c} \right] = \mathbb{P} \left[ \bigcap_{i=1}^{Kn^K} E_i \right] \leq Kn^K \cdot e^{-\nu (\log n)^{1+\xi}} = Ke^{K \log n - \nu (\log n)^{1+\xi}} \leq e^{-\nu (\log n)^{1+\xi}},$$

(2.7)

with a possible increase in $n_0(\nu, \xi)$ and a decrease in the constants $\nu$ and $\xi$.

**Proof of Theorem 2.1**

Following the Erdős-Rényi model, we produce a weighted, directed graph as follows: Consider an $n \times n$ matrix $X = [x_{ij}]$ with zero for its diagonal entries and independent, identically distributed, nonnegative random variables for its off-diagonal entries. We now want to define a random, stochastic matrix $W$ of weights. The natural definition for $W = [w_{ij}]$ is

$$w_{ij} := \frac{x_{ij}}{\sum_{k=1}^{n} x_{ik}},$$

(2.8)

which is defined when at least one of the $x_{i1}, \ldots, x_{in}$ is nonzero; this happens almost surely in the limit as $n \to \infty$, when $\mathbb{P}[x_{ij} > 0] = p > 0$ is a constant:

$$\mathbb{P}[x_{i1} = x_{i2} = \cdots = x_{in} = 0] = (1 - p)^{n-1}$$

(2.9)

and by the union bound

$$\mathbb{P} \left[ \bigvee_{i=1}^{n} x_{i1} = x_{i2} = \cdots = x_{in} = 0 \right] \leq \sum_{i=1}^{n} \mathbb{P}[x_{i1} = x_{i2} = \cdots = x_{in} = 0] = n(1 - p)^{n-1} \to 0.$$  

(2.10)

However, the question is how to technically deal with the unusual event that all the entries of a row of $X$ are zero, as there are several options. We make the following choice: for $1 \leq i \leq n$

$$w_{ii} := \begin{cases} 1 & \text{if } x_{i1} = x_{i2} = \cdots = x_{in} = 0, \\ 0 & \text{otherwise} \end{cases}$$

(2.11)

and for all $1 \leq i, j \leq n$ and $i \neq j$

$$w_{ij} := \begin{cases} \frac{x_{ij}}{\sum_{k=1}^{n} x_{ik}} & \text{if } x_{ij} > 0, \\ 0 & \text{if } x_{ij} = 0. \end{cases}$$

(2.12)
This definition aligns with the definition in (2.8) with probability greater than $1 - n(1 - p)^n$. Moreover, this definition has the advantage that the events that any non-loop edge weight is 0 are independent.

**Definition 2.4 (Generalized Erdős-Rényi random graphs).** Let $\mu$ be a nonnegative distribution (not depending on $n$) with subexponential decay such that if $X \sim \mu$

$$
P[X > 0] = p > 0 \text{ and } P[X \geq x] \leq Ce^{-x^{1/C}} \tag{2.13}$$

for some positive constants $p$ and $C$ and all $x > 0$. We denote the mean and standard deviation of $X$ by $\mu_1$ and $\sigma$ respectively. We generate a family of random graphs $G_n = (V, W)$ from $\mu$ by defining the weight matrices $W_n$ according to (2.11) and (2.12), where $x_{ij}$ are independent and distributed according to $\mu$ for $i \neq j$.

The subexponential decay is necessary to control the fluctuation of the graph’s edge weights and imposes a bounded increase on the moments of $\mu$. Let $X \sim \mu$, then simple calculations show,

$$
\mu_k := \mathbb{E}X^k \leq Ck \int_0^\infty x^{k-1}P[X > x]dx \leq Ck \int_0^\infty x^{k-1}e^{-x^{1/C}}dx = C^2 k \Gamma(C(1 + k)) \leq (Ck)^C \tag{2.14}
$$

where the constant $C > 0$ depends on the constants in (2.13). Many distributions satisfy the subexponential assumption (2.13), for example any compactly supported distribution, the Gamma distribution, and the absolute value of a Gaussian distribution.

We now use the subexponential decay assumption to understand the typical behavior of the random variables $x_{ij}$.

**Definition 2.5 (Good events $\Omega$).** Let $\Omega$ be an $n$-dependent event such that the following hold:

$$
\Omega := \bigcap_{i=1}^n \left\{ x_{ii} = 0 \right\} \cap \left\{ \left| \sum_{j} (x_{ij} - \mathbb{E}x_{ij}) \right| \leq \sigma (\log n)^{C + C\xi} \sqrt{n} \right\} \cap \left\{ x_{ij} \leq C(\log n)^C \right\} \cap \left\{ G_n \text{ is connected} \right\} \tag{2.15}
$$

The conditions on $\Omega$ have natural interpretations. The first condition specifies that the normalization procedure outlined above has worked as intended and that we are not in the atypical case where the graph has a self-loop. The second condition specifies that the sums of $n$ of the $x_{ij}$s are close to their expectation $(n - 1)\mu_1$ and that they fluctuate about this value on the order of $\sqrt{n}$ as predicted by the central limit theorem. The third condition says that none of the $x_{ij}$ are too large and that typically they will all be less than $C(\log n)^C$. The last condition is self-explanatory, as the Moran process is not guaranteed to terminate on disconnected graphs.

**Lemma 2.6.** The event $\Omega$ holds with high probability.
PROOF. By Remark 2.3, it suffices to show that each conjunct holds with high probability as there are only polynomially many choices for \(i\) and \(j\). First fix \(i\). By assumption (2.13) and the fact that \(x_{ii} \neq 0\) only if \(x_{ij} = 0\) for all \(j \neq i\),

\[
P \{ x_{ii} \neq 0 \} = P \{ x_{ij} = 0 \text{ for all } j \neq i \} \leq (1 - p)^{n-1} = e^{\log(1-p)(n-1)} \leq e^{-\nu(\log n)^{1+\xi}},
\]

(2.17) since \(0 < p < 1\) and \(n - 1 \gg (\log n)^{1+\xi}\).

Now using the large deviation result, Lemma A.4, with \(a_j = x_{ij} - \mathbb{E}x_{ij}\) and \(A_j = 1\), we may verify the moment assumption (A.7): clearly

\[
\mathbb{E} (x_{ij} - \mathbb{E}x_{ij}) = 0
\]

and

\[
\mathbb{E} (x_{ij} - \mathbb{E}x_{ij})^2 = \sigma^2,
\]

then

\[
E |x_{ij} - \mathbb{E}x_{ij}|^k \leq (Ck)^Ck
\]

by Equation (2.14). Thus we get

\[
P \left[ \sum_{i=1}^n a_i A_i \geq \sigma (\log n)^{C+C_{1+\xi}} \sqrt{n} \right] \leq e^{-\nu(\log n)^{1+\xi}}.
\]

(2.19)

Now fix \(j\) too. Next, we use the subexponential decay assumption (2.13) with \(x = C (\log n)^{C+1}\) to get

\[
P \left[ x_{ij} > C (\log n)^{C-1+1} \right] \leq C \exp \left( - \left( C (\log n)^{C+1} \right)^{1/C} \right) \leq C e^{-C(C-1)(\log n)^{C-1}} \leq e^{-\nu(\log n)^{1+\xi}}.
\]

(2.20)

Thus, \(x_{ij} > C (\log n)^{C+1}\) holds with high probability since \(C^{-1} > 0\) and \(C^{C-1} > 0\).

Finally, we show that the graph \(G\) is connected with high probability, i.e. that with high probability, the graph cannot be partitioned into two disjoint sets where there are no edges going from one subset to another. This follows from an argument similar to that contained in the proof of Lemma 2.8 but without the assumption that we are on the event \(\Omega\) as we do not need a lower bound on the weights only that edges exist which they do with probability at least \(p\).

Note that by definition \(W\) is stochastic. Define the sum of the \(j\)th column as

\[
W_j := \sum_{i=1}^n w_{ij}.
\]

(2.21)

Note that while the family \(W_j\) is not independent, by symmetry, they are identically distributed. Hence

\[
\mathbb{E}W_1 = \frac{1}{n} \sum_{j=1}^n \mathbb{E}W_j = \frac{1}{n} \mathbb{E} \sum_{j=1}^n \sum_{i=1}^n w_{ij} = 1.
\]

(2.22)

This tells us that in expectation \(W\) is doubly stochastic. The next lemma shows that with very high probability it is almost \(n^{-1/2}\) close to being doubly stochastic, which is exactly the order of fluctuation we expect by the central limit theorem. The assumptions on the distribution \(\mu\) and the event \(\Omega\) guarantee that we can prove that the sum’s fluctuations are of this order.
The idea of the proof is that for fixed \( j \), the \( w_{ij} \) are independent random variables and thus we can apply a LDE to bound the fluctuations of their sum. There are complications due to the normalization required by Definition 2.4 but on \( \bar{\Omega} \) these can be overcome by relating the sum \( W_j \) to a simpler sum that may be controlled with Lemma A.4.

**Lemma 2.7.** On \( \bar{\Omega} \), there is a positive constant \( C \equiv C_\mu \), not dependent on \( n \), such that the following inequalities hold

\[
|W_j - 1| \leq \frac{C \left( \log n \right)^{C+C\xi}}{\sqrt{n}}
\]

for all \( j \in V \), with probability at least

\[
1 - e^{-\nu \left( \log n \right)^{1+\xi}}.
\]

**Proof.** Fix \( j \). First we use the fact that \( w_{ii} = x_{ii} = 0 \) for \( 1 \leq i \leq n \) on \( \bar{\Omega} \) to see

\[
W_j - 1 = \sum_i (w_{ij} - \mathbb{E}w_{ij}) = \sum_i \left( w_{ij} - \frac{1}{n-1} \right) + \mathcal{O} \left( n^2(1-p)^{-n+1} \right).
\]

By the definition of \( w_{ij} \), the above is equal to

\[
\sum_i \left( \frac{x_{ij}}{\sum_k x_{ik}} - \frac{1}{n-1} \right) + \mathcal{O} \left( c_0^{-n} \right) = \sum_i \left( \frac{x_{ij}}{\sum_k x_{ik}} - \frac{1}{n-1} \sum_k x_{ik} \right) + \mathcal{O} \left( c_0^{-n} \right),
\]

where \( c_0 < 1 \) is not dependent on \( n \). Next, using the fact that on \( \bar{\Omega} \)

\[
\frac{1}{n-1} \left| \sum_k x_{ik} - \mathbb{E}x_{ik} \right| \leq C\sigma \left( \log n \right)^{C+C\xi} \frac{1}{\sqrt{n}},
\]

for all \( 1 \leq i \leq n \), we replace the average in the numerator of (2.26) with its expectation to find it equal to

\[
\sum_i \left( \frac{x_{ij} - \mathbb{E}x_{ij}}{\sum_k x_{ik}} + C\sigma \left( \log n \right)^{C+C\xi} \frac{1}{\sqrt{n} \sum_k x_{ik}} \right) + \mathcal{O} \left( c_0^{-n} \right).
\]

Using (2.27) again, it is easy to see

\[
\sum_k x_{ik} \geq (n-1)\mathbb{E}x_{ij} - C\sigma \left( \log n \right)^{C+C\xi} \sqrt{n},
\]

which gives an upper bound on the error term in (2.28) and we find the equation equal to

\[
\sum_i \left( \frac{x_{ij} - \mathbb{E}x_{ij}}{\sum_k x_{ik}} + C\sigma \left( \log n \right)^{2C+2C\xi} \right) + \mathcal{O} \left( c_0^{-n} \right).
\]
Next we compare these two expressions to find that the absolute value their difference can be expressed as
\[
\left| x_{ij} - \mathbb{E}x_{ij} - x_{ij} - \mathbb{E}x_{ij} \right| = \frac{|x_{ij}|^2}{\sum_k^{(i)} x_{ik} \cdot \left( \sum_k^{(i)} x_{ik} - x_{ij} \right)}. \tag{2.31}
\]

However, using that on $\Omega$, for all $1 \leq i, j \leq n$, we have $x_{ij} \leq C (\log n)^C$ and using (2.27) as before, we may show the difference is bounded by
\[
\mathcal{O} \left( \frac{(\log n)^{4C+2C\xi}}{n^2} \right). \tag{2.32}
\]

We can then sum over these errors—one for each summand—to get a total error of $\mathcal{O} \left( \frac{(\log n)^{4C+2C\xi}}{n} \right)$. Thus, (2.30) may be rewritten as
\[
\sum_i^{(j)} x_{ij} - \mathbb{E}x_{ij} + \mathcal{O} \left( \frac{(\log n)^{2C+2C\xi}}{\sqrt{n}} \right), \tag{2.33}
\]

since the other error terms are dominated by the remaining one.

Note that $x_{ij}$ does not appear in the summand’s denominator and thus the denominator and numerator are independent. So we can use the large deviation estimate, Lemma A.4, with $a_i = x_{ij} - \mathbb{E}x_{ij}$ and $A_i^{-1} = \sum_k^{(i)} x_{ik}$. While the $A_i$ are random, we may condition on their values and treat them as deterministic constants, then after we have used the LDE, we can bound them using the fact that we are on $\Omega$. That is, on $\Omega$
\[
\left( \sum_k^{(i)} x_{ik} \right)^2 = (n-2)^2 (\mathbb{E}x_{ij})^2 + \mathcal{O} \left( (\log n)^{2C+2C\xi} n \sqrt{n} \right) \tag{2.34}
\]

and so
\[
\sum_k^{(j)} A_i^2 = \frac{1}{(n-2) (\mathbb{E}x_{ij})^2} + \mathcal{O} \left( (\log n)^{2C+2C\xi} \frac{1}{n \sqrt{n}} \right). \tag{2.35}
\]

Thus the LDE gives us
\[
\mathbb{P} \left[ \sum_i^{(j)} a_i A_i \right] \geq C \frac{(\log n)^{C+C\xi}}{\sqrt{n}} \leq e^{-\nu(\log n)^{1+\xi}}, \tag{2.36}
\]

which combined with (2.33)
\[
\mathbb{P} \left[ |W_j - 1| \geq C \frac{(\log n)^{2C+2C\xi}}{\sqrt{n}} \right] \leq e^{-\nu(\log n)^{1+\xi}}. \tag{2.37}
\]

The properties of high probability and the fact that we have $n$ choices for $j$ completes the proof. ■
Next we prove a lower bound on sums of edge weights, \(w_I(S)\) and \(w_O(S)\) for all \(\emptyset \neq S \subseteq V_n\). The proof relies on concentration inequalities for independent random variables and the simple fact that on \(\Omega\) there is a constant \(c > 0\) such that \(w_{ij} \geq cn^{-1} \mathbb{1}(x_{ij} \geq c)\) for all \(i,j \in V\).

**Lemma 2.8.** On \(\Omega\), for all \(\emptyset \neq S \subseteq V_n\) and some small constant \(c \equiv c_\mu > 0\), not dependent on \(n\), we have the following bound

\[
|w_I(S)| = |w_O(S^c)| \geq c_\mu \min \{|S|, n - |S|\}
\]

with probability greater than

\[
1 - e^{-\nu(\log n)^{1+\xi}}.
\]

**Proof.** First note that as in the proof of Lemma 2.7, we can argue that on \(\Omega\) the sum \(\sum_{k}^{i} x_{ik} \leq Cn\), see (2.29) for all \(1 \leq i \leq n\). Moreover, by assumption on the distribution \(\mu\), we have \(\mathbb{P}[x_{ij} > 0] = p > 0\) and thus there is a constant \(c > 0\) such that \(\mathbb{P}[x_{ij} \geq c] = p' > 0\). Therefore, on \(\Omega\)

\[
w_{ij} \geq cn^{-1} \mathbb{1}(x_{ij} \geq c).
\]

However, for each \(1 \leq i,j \leq n\) with \(i \neq j\), define \(\beta_{ij} := \mathbb{1}(x_{ij} \geq c)\) which are independent Bernoulli random variables such that

\[
\mathbb{P}[\mathbb{1}(x_{ij} \geq c) = 1] = p' > 0,
\]

since the \(x_{ij}\) are independent.

Let \(|S| = k\). By definition

\[
w_I(S) = \sum_{i \in S} \sum_{j \in S} w_{ij} = w_O(S^c).
\]

Note that no diagonal terms are in these sums. Using (2.41),

\[
\sum_{i \in S} \sum_{j \in S} w_{ij} \geq \frac{c}{n} \sum_{i \in S} \sum_{j \in S} \beta_{ij}.
\]

Note that now this is a sum of \(k(n - k)\) independent random variables. So for fixed \(\emptyset \neq S \subseteq V_n\), by the Chernoff bound, Lemma A.3, the event

\[
A_S := \left\{ \sum_{i \in S} \sum_{j \in S} \beta_{ij} \leq (1 - 1/2)p'k(n - k) \right\}
\]

has probability less than

\[
\exp \left( -\frac{1}{8} p'k(n - k) \right).
\]
Thus, by the union bound
\[
\P \left[ \bigcap_{\emptyset \neq S \subseteq V_n} A^c_S \right] = 1 - \P \left[ \bigcup_{\emptyset \neq S \subseteq V_n} A_S \right] \\
\geq 1 - \sum_{\emptyset \neq S \subseteq V_n} \P [A_S] \\
= 1 - \sum_{k=1}^{n-1} \binom{n}{k} \P [A_S] \\
= 1 - \sum_{k=1}^{\lfloor \log n \rfloor} \binom{n}{k} \P [A_S] - \sum_{k=\lfloor \log n \rfloor + 1}^{n-\lfloor \log n \rfloor - 1} \binom{n}{k} \P [A_S] - \sum_{k=\lfloor \log n \rfloor}^{n-1} \binom{n}{k} \P [A_S] \\
\geq 1 - 2n^{\log n} \exp \left( -p' n / 8 \right) - 2^n \exp \left( -p' n (\log n) / 8 \right) \\
\geq 1 - 2 \exp \left( -(p' n / 8 - (\log n)^2) \right) - \exp \left( -(p' \log n / 8 - \log 2) n \right) \\
\geq 1 - \exp \left( -cp' n \right),
\]
for some \( c > 0 \). Finally, note
\[
\frac{cp'}{2n} k(n-k) \geq \frac{cp'}{2} \min \{ |S|, n - |S| \}
\]
and
\[
\exp(-cp'n) \leq e^{-\nu(\log n)\xi}
\]
for an appropriate choice of \( \xi \) and \( \nu \).

We now complete the proof of Theorem 2.1 by putting together the results of Section 1 and
the lemmata from this section.

**Proof of Theorem 2.1.** Again let \( |S| = k \). We check that the assumptions of Theorem 1.2
hold with high probability. Observe
\[
\left| \frac{w_O(S)}{w_1(S)} - 1 \right| = \left| \frac{w_O(S) - w_1(S)}{w_1(S)} \right| = \frac{|w_O(S) - w_1(S)|}{|w_1(S)|}.
\]
Expanding the numerator, we get
\[
w_O(S) - w_1(S) = \sum_{i \in S} \sum_{j \notin S} w_{ij} - \sum_{i \notin S} \sum_{j \in S} w_{ij} = \sum_{i \in S} \sum_{j \in V} w_{ij} - \sum_{i \notin V} \sum_{j \in S} w_{ij} = \sum_{j \in S} (1 - W_j),
\]
and similarly,
\[
w_O(S) - w_1(S) = \sum_{i \in V} \sum_{j \notin S} w_{ij} - \sum_{i \notin S} \sum_{j \in V} w_{ij} = \sum_{j \notin S} (W_j - 1).
\]
Thus Lemma 2.7 implies
\[ |w_0(S) - w_1(S)| \leq \min \{ k, n - k \} \cdot \frac{C (\log n)^{C+\xi}}{\sqrt{n}}, \] (2.52)
for all \( S \) with high probability on \( \Omega \).

Lemma 2.8 implies
\[ |w_1(S)| \geq c \min \{ |S|, n - |S| \} \] (2.53)
for all \( S \) with high probability on \( \Omega \). Putting this together we see
\[ \left| \frac{w_0(S)}{w_1(S)} - 1 \right| \leq \frac{C (\log n)^{C+\xi}}{\sqrt{n}} \] (2.54)
for all \( S \) with high probability on \( \Omega \). However, by Lemma 2.6, the event \( \Omega \) holds with high probability itself and thus unconditionally (2.54) holds for all \( S \) with high probability.

Finally, applying Theorem 1.2, we get
\[ \sup_{r>0} |\rho_{G_n}(r) - \rho_{M_n}(r)| \leq \frac{C (\log n)^{C+\xi}}{\sqrt{n}} \] (2.55)
with high probability.

Remark 2.9 (On Theorem 2.1). The parameter \( p \), the probability that an edge of some weight exists between two directed vertices, can be interpreted as a measure of the sparseness of the population structure. We can ask how few interactions on average can individuals in a population have with others and still yield populations with Moran-type behavior? While \( p \) can be arbitrarily small, we have kept it constant and, in particular, not dependent on \( n \). However, could \( p \) depend on \( n \) such that \( p \to 0 \) as \( n \to \infty \) and still produce graphs which show Moran-type behavior? An obvious lower bound on the rate of \( p \)'s convergence to 0 is provided by the Erdős-Rényi model, which tells us that a graph is almost surely disconnected in the limit for \( \sqrt{p} < (1 - \varepsilon)(1/n) \log n \) for any \( \varepsilon > 0 \). This bound follows by noting that \( (1 - p)^2 \) is the probability that there is no edge, in either direction, between two vertices and then applying the usual Erdős-Rényi threshold.4,5 There is much room between this lower bound and \( p \) constant—even whether such a sharp threshold for \( p \) exists is currently unclear. The issue is difficult to approach with naive simulations as the Moran process is not guaranteed to terminate on disconnected graphs.

3 Graphs without outgoing weights summing to 1

The Moran process on graphs may be generalized to no longer require the sum of the outgoing weights of each vertex to be 1. The process is still defined for a directed, weighted graph, \( G \equiv G_n = (V_n, W_n) \), where \( V \equiv V_n := [n] \) but \( W \equiv W_n := [w_{ij}] \) need not be a stochastic matrix (it must still have nonnegative entries). Instead of sampling vertices proportional to fitness and then choosing an outgoing edge with probability equal to its weight, we rather sample edges proportional to their weights and the fitness of the individual at the beginning of the edge. Once we select an edge the type of the individual at the end of the edge becomes the same as the individual
at the beginning of the edge. It is easy to see that we get the original process as a special case of the new model. Again, the process $X_t$ is Markovian with state space $2^V$, where each $X_t$ is the subset of $V$ consisting of all the vertices occupied by mutants at time $t$. At time 0 a mutant is placed at one of the vertices uniformly at random. We then update as described above by choosing edges proportional to their weights and the individual’s type at the edges origin. For example, the probability of choosing a particular edge $(i, j)$ is

$$\frac{(r\mathbb{1}_{i \in X_t} + \mathbb{1}_{j \notin X_t}) w_{ij}}{r \sum_{i \in X_t} \sum_{j \in V} w_{ij} + \sum_{i \notin X_t} \sum_{j \in V} w_{ij}}.$$ (3.1)

In this setting the robust isothermal theorem still holds.

**Theorem 3.1 (Robust Isothermal Theorem).** Fix $0 \leq \varepsilon < 1$. Let $G^n = (V_n, W_n)$ be a connected graph. If for all nonempty $S \subseteq V_n$ we have

$$\left| \frac{w_O(S)}{w_I(S)} - 1 \right| \leq \varepsilon,$$ (3.2)

then

$$\sup_{r > 0} |\rho_{G^n}(r) - \rho_{G^n}(r)| \leq \varepsilon.$$ (3.3)

**Proof.** Just as before, we make the projection of the state space of all subsets of $V$, which records exactly which vertices are mutants, to the simpler state space \{0, 1, \ldots, n\}, which records only the number of mutants. In the new model, we still have

$$\frac{p_+(S)}{p_-(S)} = r \frac{w_O(S)}{w_I(S)}.$$ (3.4)

So the argument is exactly the same as the previous case. \hfill \blacksquare

4 More random graphs

We can introduce a random graph model where the sum of the outgoing weights are not equal, that is, where individuals can contribute differentially to the next time point in a way not dependent on the genotype they carry. We change the model by not normalizing the outgoing edge weights. Following the Erdős-Rényi model, we produce a weighted, directed graph as follows: Consider an $n \times n$ matrix $X = [x_{ij}]$ with independent, identically distributed, nonnegative random variables for its entries. Since we do not need to normalize we take, $W = X$.

**Definition 4.1 (Generalized Erdős-Rényi Random Graphs).** Let $\mu$ be a nonnegative distribution (not depending on $n$) with subexponential decay such that if $X \sim \mu$

$$\mathbb{P}[X > 0] = p > 0 \text{ and } \mathbb{P}[X \geq x] \leq Ce^{-x^{1/C}}$$ (4.1)

for some positive constants $p$ and $C$ and all $x > 0$. We denote the mean and standard deviation of $X$ by $\mu_1$ and $\sigma$ respectively. We generate a family of random graphs $G_n = (V_n, X_n \equiv W_n)$ from $\mu$, where $x_{ij}$ are independent and distributed according to $\mu$.
Now we have a similar theorem to before but the proof is easier.

**Theorem 4.2.** Let \((G_n)_{n \geq 1}\) be a family of random graphs as in Definition 4.1. Then there are constants \(C > 0\), not dependent on \(n\), such that the fixation probability of a randomly placed mutant of fitness \(r > 0\) satisfies

\[
|\rho_{G_n}(r) - \rho_{M_n}(r)| \leq \frac{C (\log n)^{C + C \xi}}{\sqrt{n}} \quad (4.2)
\]

uniformly in \(r\) with probability greater than

\[
1 - \exp\left(-\nu (\log n)^{1+\xi}\right), \quad (4.3)
\]

for positive constants \(\xi\) and \(\nu\).

**Proof of Theorem 4.2** We now use the subexponential decay assumption to understand the typical behavior of the random variables \(x_{ij}\).

**Definition 4.3 (Good Events \(\Omega\)).** Let \(\Omega\) be an \(n\)-dependent event such that the following hold:

\[
\Omega := \cap_{i,j=1}^{n} \left\{ x_{ij} \leq C (\log n)^C \right\} \cap \{G_n \text{ is connected}\}. \quad (4.4)
\]

**Lemma 4.4.** The event \(\Omega\) holds with high probability.

**Proof.** Identical to the previous proof. ■

Define the sum of the \(j\)th column as

\[
W_j := \sum_{i=1}^{n} w_{ij} \quad (4.5)
\]

and the sum of the \(i\)th row as

\[
\tilde{W}_i := \sum_{j=1}^{n} w_{ij}. \quad (4.6)
\]

**Lemma 4.5.** On \(\Omega\), there is a positive constant \(C \equiv C_{\mu},\) not dependent on \(n\), such that the following inequalities hold

\[
|W_j - \mu n| \leq C (\log n)^{C + C \xi} \sqrt{n} \quad (4.7)
\]

and

\[
|\tilde{W}_i - \mu n| \leq C (\log n)^{C + C \xi} \sqrt{n} \quad (4.8)
\]

for all \(i, j \in V\), with probability at least

\[
1 - e^{-\nu (\log n)^{1+\xi}}. \quad (4.9)
\]
PROOF. Consider $a_i = w_{ij} - \mu$ or $a_i = w_{ji} - \mu$ and apply Lemma A.4. The claim follows immediately as there are $2n$ high probability events.

LEMMA 4.6. On $\Omega$, for all $\emptyset \neq S \subseteq V_n$ and some small constant $c \equiv c_\mu > 0$, not dependent on $n$, we have the following bound

$$|w_I(S)| = |w_O(S^c)| \geq c_\mu |S| (n - |S|)$$

with probability greater than

$$1 - e^{-\nu(\log n)^{1+\xi}}.$$  

PROOF. By assumption on the distribution $\mu$, we have $P[x_{ij} > 0] = p > 0$ and thus there is a constant $c > 0$ such that $P[x_{ij} \geq c] = p' > 0$. Define $\beta_{ij} := 1 (x_{ij} \geq c)$ which are independent Bernoulli random variables such that

$$P[\beta_{ij} = 1] = p' > 0,$$

since the $x_{ij}$ are independent. Let $|S| = k$. By definition

$$w_I(S) = \sum_{i \in S} \sum_{j \in S} x_{ij} = w_O(S^c).$$

Using (2.41),

$$\sum_{i \in S} \sum_{j \notin S} x_{ij} \geq c \sum_{i \in S} \sum_{j \notin S} \beta_{ij}.$$  

Note that now this is a sum of $k(n - k)$ independent random variables. So for fixed $\emptyset \neq S \subseteq V_n$, by the Chernoff bound, Lemma A.3, the event

$$A_S := \left\{ \sum_{i \in S} \sum_{j \notin S} \beta_{ij} \leq (1 - 1/2)p'k(n - k) \right\}$$

has probability less than

$$\exp \left( -\frac{1}{8} p'k(n - k) \right).$$
Thus, by the union bound

\[
\Pr \left[ \bigcap_{\emptyset \neq S \subseteq V_n} A_S^c \right] = 1 - \Pr \left[ \bigcup_{\emptyset \neq S \subseteq V_n} A_S \right] \\
\geq 1 - \sum_{\emptyset \neq S \subseteq V_n} \Pr [A_S] \\
= 1 - \sum_{k=1}^{n-1} \binom{n}{k} \Pr [A_S] \\
= 1 - \sum_{k=1}^{\lceil \log n \rceil} \binom{n}{k} \Pr [A_S] - \sum_{k=\lceil \log n \rceil + 1}^{n-\lfloor \log n \rfloor - 1} \binom{n}{k} \Pr [A_S] \\
\geq 1 - 2n^{\log n} \exp (-p' n/8) - 2^n \exp (-p' n (\log n) / 8) \\
\geq 1 - 2 \exp \left( - (p' n/8 - (\log n)^2) \right) - 2 \exp \left( - (p' \log n/8 - \log 2) n \right) \\
\geq 1 - \exp (-cp' n) ,
\]

for some \( c > 0 \). Finally, note

\[
\exp(-cp' n) \leq e^{-\nu(\log n)^1 + \xi}
\]

for an appropriate choice of \( \xi \) and \( \nu \). 

We now complete the proof of Theorem 4.2.

**Proof of Theorem 4.2.** Again let \(|S| = k\). We check that the assumptions of Theorem 3.1 hold with high probability. Observe

\[
\left| \frac{w_O(S) - w_I(S)}{w_I(S)} - 1 \right| = \frac{|w_O(S) - w_I(S)|}{w_I(S)} .
\]

Expanding the numerator, we get

\[
w_O(S) - w_I(S) = \sum_{i \in S} \sum_{j \notin S} w_{ij} - \sum_{i \notin S} \sum_{j \notin S} w_{ij} = \sum_{i \in S} \sum_{j \in V} w_{ij} - \sum_{i \notin S} \sum_{j \in V} w_{ij} = \sum_{j \in S} (\tilde{W}_j - W_j),
\]

and similarly,

\[
w_O(S) - w_I(S) = \sum_{i \in V} \sum_{j \notin S} w_{ij} - \sum_{i \notin S} \sum_{j \in V} w_{ij} = \sum_{j \notin S} (W_j - \tilde{W}_j).
\]

Thus Lemma 4.5 implies

\[
|w_O(S) - w_I(S)| \leq \min \{ k, n - k \} \cdot C (\log n)^{C + C^\varepsilon} \sqrt{n},
\]

for all \( S \) with high probability on \( \Omega \).
Lemma 4.6 implies
\[ |w_1(S)| \geq ck (n - k) \] (4.23)
for all \( S \) with high probability on \( \Omega \). Putting this together we see
\[ \left| \frac{w_\Omega(S)}{w_1(S)} - 1 \right| \leq \frac{C (\log n)^{C+C\xi}}{\sqrt{n}} \] (4.24)
for all \( S \) with high probability on \( \Omega \). However, by Lemma 4.4, the event \( \Omega \) holds with high probability itself and thus unconditionally (2.54) holds for all \( S \) with high probability.

Finally, applying Theorem 3.1, we get
\[ \sup_{r > 0} |\rho_{G_n}(r) - \rho_{M_n}(r)| \leq \frac{C (\log n)^{C+C\xi}}{\sqrt{n}} \] (4.25)
with high probability.

\section*{A Large deviation estimates and concentration inequalities}

In this section we provide a brief review of large deviation estimates and concentration inequalities with a focus on those used above. A large deviation estimate \((\text{LDE})\) controls atypical behavior of sums of independent (or sometimes weakly dependent) random variables, whereas a concentration inequality controls the convergence of an average of independent (or sometimes weakly dependent) random variables to their mean. For a more in-depth review of \(\text{LDEs}\), see for example.\(^5,6\) Many \(\text{LDEs}\) follow directly by applying Markov’s inequality, so we state this now.

\textbf{Theorem A.1 (Markov’s Inequality).} Let \( X \) be a nonnegative random variable and \( t > 0 \). Then
\[ \mathbb{P}[X \geq t] \leq \frac{\mathbb{E}X}{t}. \] (A.1)

\textbf{Proof.} Define the indicator random variable \( 1_{X \geq t} \). Then \( t 1_{X \geq t} \leq X \), thus \( \mathbb{E}[t 1_{X \geq t}] \leq \mathbb{E}X \). Therefore,
\[ \mathbb{P}[X \geq t] = \mathbb{E}1_{X \geq t} \leq \frac{\mathbb{E}X}{t}. \]

This very simple result has lots of scope. The general idea is to define a nonnegative, increasing function \( f \) of some random variable \( X \) and note that Markov’s inequality implies
\[ \mathbb{P}[X \geq t] \leq \mathbb{P}[f(X) \geq f(t)] \leq \frac{\mathbb{E}f(X)}{f(t)}. \] (A.2)

Normally, \( f \) is chosen as \( x^k \) or \( e^{\lambda X} \) where \( k \) or \( \lambda \) is optimized to strengthen the inequality. If the random variable \( X \) is a sum of centered, independent random variables, \( \sum_{i=1}^n (X_i - \mathbb{E}X_i) \), the function takes the form
\[ \prod_{i=1}^n \exp \left( \lambda (X_i - \mathbb{E}X_i) \right). \] (A.3)

In this way we get several inequalities.
LEMMA A.2 (Hoeffding’s Inequality). Suppose that $X_1, \ldots, X_n$ are i.i.d. Bernoulli random variables with parameter $p \in [0, 1]$. Define $X := \sum_{i=1}^n X_i$. Then
\[
P \left[ |X - \mathbb{E}X| \geq \delta \sqrt{n} \right] \leq 2 \exp \left( -2\delta^2 \right) \tag{A.4}
\]
for all $\delta > 0$.

Lemma A.2 states that $X$ fluctuates about its expectation on the order of $\sqrt{n}$, and the probability of a fluctuation greater than $\delta \sqrt{n}$ decays exponentially with $\delta > 0$. The next lemma bounds fluctuation of larger orders, and thus they occur even more infrequently. We shall only need a lower bound in this case:

LEMMA A.3 (Multiplicative Chernoff Bound). Suppose that $X_1, \ldots, X_n$ are i.i.d. Bernoulli random variables with parameter $p \in [0, 1]$. Define $X := \sum_{i=1}^n X_i$. Then
\[
P \left[ X \leq (1 - \varepsilon)\mathbb{E}X \right] \leq \exp \left( -\frac{\varepsilon^2 p n}{2} \right) \tag{A.5}
\]
and
\[
P \left[ X \geq (1 + \varepsilon)\mathbb{E}X \right] \leq \exp \left( -\frac{\varepsilon^2 p n}{2} \right). \tag{A.6}
\]

We remark that far more general statements of Lemmas A.2 and A.3 are possible, but we state only the versions we use in Section 2.

Finally, we state a LDE for weighted sums of independent random variables with the following conditions on their moments:
\[
\mathbb{E}X = 0, \quad \mathbb{E}|X|^2 = \sigma^2, \quad \text{and} \quad \mathbb{E}|X|^k \leq (Ck)^{Ck}, \tag{A.7}
\]
for some positive constant $C > 0$ (not dependent on $n$ or $k$) and for $k \geq 1$.

LEMMA A.4. Suppose the independent random variables $\left(a^{(n)}_i\right)_{i=1}^n$ for $n \in \mathbb{N}$ satisfy (A.7) and that $\left(A^{(n)}_i\right)_{i=1}^n$ for $n \in \mathbb{N}$ are constants in $\mathbb{R}$. Then
\[
P \left[ \left| \sum_{i=1}^n a_i A_i \right| \geq \sigma (\log n)^{C+C\xi} \left( \sum_{i=1}^n |A_i|^2 \right)^{1/2} \right] \leq e^{-\nu (\log n)^{1+\xi}}. \tag{A.8}
\]
In words, we can bound the sum $\sum_{i=1}^n a_i A_i$ on the same order as the norm of the coefficients with high probability.

To prove this lemma we use a high-moment Markov inequality, so first we need a result bounding the higher moments of this sum.
LEMMA A.5. Suppose the independent random variables \( \left( a_i^{(n)} \right)_{i=1}^n \) for \( n \in \mathbb{N} \) satisfy (A.7) and that \( \left( A_i^{(n)} \right)_{i=1}^n \) for \( n \in \mathbb{N} \) are constants in \( \mathbb{R} \). Then
\[
\mathbb{E} \left| \sum_{i=1}^n a_i A_i \right|^k \leq (Ck)^{k/2} \left( \sum_{i=1}^n |A_i|^2 \right)^{k/2} \tag{A.9}
\]

PROOF. Without loss of generality let \( \sigma = 1 \). Let \( A^2 := \sum_i |A_i|^2 \), then by the classical Marcinkiewicz-Zygmund inequality\(^7\) in the first line, we get
\[
\mathbb{E} \left| \sum_i a_i A_i \right|^k \leq (Ck)^{k/2} \mathbb{E} \left[ \left( \sum_i |A_i|^2 a_i^2 \right)^{1/2} \right] \tag{A.10}
\]
\[
= (Ck)^{k/2} A^k \mathbb{E} \left[ \left( \sum_i \frac{|A_i|^2}{A^2} |a_i|^2 \right)^{k/2} \right] \tag{A.11}
\]
\[
\leq (Ck)^{k/2} A^k \mathbb{E} \left[ \sum_i \frac{|A_i|^2}{A^2} |a_i|^k \right] \tag{A.12}
\]
\[
= (Ck)^{k/2} A^k \sum_i \frac{|A_i|^2}{A^2} \mathbb{E} |a_i|^k \tag{A.13}
\]
\[
\leq (Ck)^{Ck+k/2} A^k \tag{A.14}
\]
\[
\leq (Ck)^{C_k} A^k \tag{A.15}
\]
where we have used Jensen’s inequality in the third line and assumption (A.7) in line 5.

PROOF OF LEMMA A.4. Without loss of generality let \( \sigma = 1 \). The proof is a simple application of Markov’s inequality, Theorem A.1. Let \( k = \nu (\log n)^{1+\xi} \), then by Lemma A.5, we get
\[
\mathbb{P} \left[ \sum_i a_i A_i \geq (\log n)^{C + C_k} \left( \sum_i |A_i|^2 \right)^{1/2} \right] = \mathbb{P} \left[ \sum_i a_i A_i \geq (\log n)^{C k + C_k \xi} \left( \sum_i |A_i|^2 \right)^{k/2} \right] \tag{A.16}
\]
\[
\leq \frac{\mathbb{E} \left| \sum_i a_i A_i \right|^k}{(\log n)^{C k + C_k \xi} \left( \sum_i |A_i|^2 \right)^{k/2}} \tag{A.17}
\]
\[
\leq \left( \frac{Ck}{(\log n)^{1+\xi}} \right)^{\frac{C k}{\nu}} \tag{A.18}
\]
\[
= (C\nu)^{\nu (\log n)^{1+\xi}} \tag{A.19}
\]
\[
\leq e^{-\nu (\log n)^{1+\xi}}, \tag{A.20}
\]
for \( \nu \leq e^{-1} \) small enough.


