OXFORD GENETICS

Evolutionary graph theory beyond single mutation dynamics: on how network structured populations cross fitness landscapes

Yang Ping Kuo[1,2] and Oana Carja [1,*]

[1]Computational Biology Department, School of Computer Science, Carnegie Mellon University, Pittsburgh, PA, USA
[2]Joint Carnegie Mellon University-University of Pittsburgh Ph.D. Program in Computational Biology

*Corresponding author: Computational Biology Department, School of Computer Science, Carnegie Mellon University, Pittsburgh, PA, USA. Email: oana.carja@gmail.com or ocarja@andrew.cmu.edu.

Abstract

Spatially-resolved datasets are revolutionizing knowledge in molecular biology, yet are under-utilized for questions in evolutionary biology. To gain insight from these large-scale datasets of spatial organization, we need mathematical representations and modeling techniques that can both capture their complexity, but also allow for mathematical tractability. Evolutionary graph theory utilizes the mathematical representation of networks as a proxy for heterogeneous population structure and has started to reshape our understanding of how spatial structure can direct evolutionary dynamics. However, previous results are derived for the case of a single new mutation appearing in the population and the role of network structure in shaping fitness landscape crossing is still poorly understood. Here we study how network structured populations cross fitness landscapes and show that even a simple extension to a two-mutational landscape can exhibit complex evolutionary dynamics that cannot be predicted using previous single-mutation results. We show how our results can be intuitively understood through the lens of how the two main evolutionary properties of a network, the amplification and acceleration factors, change the expected fate of the intermediate mutant in the population and further discuss how to link these models to spatially-resolved datasets of cellular organization.

Keywords: evolutionary graph theory; spatial structure; rugged fitness landscapes; stochastic tunneling; cell graphs

Introduction

In recent years, evolutionary graph theory has started to reshape our understanding of how spatial structure can direct evolutionary dynamics (Lieberman *et al.* 2005; Ohtsuki and Nowak 2006; Santos *et al.* 2006; Ohtsuki *et al.* 2006, 2007; Poncela *et al.* 2007; Allen *et al.* 2013; Maciejewski *et al.* 2014; Jiang *et al.* 2014; Leventhal *et al.* 2015; Kuo *et al.* 2021; Kuo and Carja 2021; Su *et al.* 2022). Using tools from network theory, these models allow us to tune the heterogeneity of spatial structure, beyond what is possible with deme-based or lattice-based models (Wright 1943; Kimura and Weiss 1964; Carja *et al.* 2014; Maruyama 1970a; Slatkin 1981; Whitlock and Barton 1997; Whitlock 2003). The nodes of the graph represent individuals in the population and edges are proxies for the local pattern of replacement and substitution. Nodes can also be interpreted as genetically homogeneous subpopulations where genetic drift is ignored and a new beneficial mutation arriving in the node subpopulation is assumed to fix immediately (but see also studies where this assumption is relaxed Yagoobi and Traulsen (2021); Marrec *et al.* (2021); Yagoobi *et al.* (2023)).

By incorporating heterogeneity in spatial structure, these models have been shown to greatly extend the range of possible evolutionary outcome of a population and can boost the selective benefit of new mutations and, reversely, suppress the spread of deleterious mutants (Adlam *et al.* 2015; Hindersin and Traulsen 2015; Kuo *et al.* 2021; Allen *et al.* 2021; Tkadlec *et al.* 2020). Network structures have also been shown to affect times to fixation of new mutants in the population (Hindersin and Traulsen 2014; Tkadlec *et al.* 2019; Kuo and Carja 2021). One of the main obstacles to theoretical progress in this area has been the analytical difficulty of multidimensional stochastic models and previous results have been restricted to studying times and probabilities of fixation for one single mutation appearing in the population. These previous single-mutation frameworks can therefore only be used to predict sequential fixation dynamics, in the limit of small mutation rates.

However, complex traits often arise from interactions between multiple gene products and, in order to reach a fitness peak, populations often need to cross fitness valleys or plateaus (Wright *et al.* 1932; Kvitek and Sherlock 2011; Burch and Chao 1999; Wood *et al.* 2007; Vogelstein *et al.* 2013; Huang 2013; Rogers *et al.* 2018; Acar *et al.* 2020; Salehi *et al.* 2021). Well-mixed large populations have been shown to cross wide fitness valleys remarkably quickly, suggesting valley-crossing dynamics is common, even when mutations that directly increase fitness are available (Weissman *et al.* 2009; Weinreich and Chao 2005; Jain and Krug 2007). Furthermore, under elevated mutation rates, crossing fitness valleys occurs not through sequential fixation, but mostly through stochastic tunneling, especially when the population size is large and when the intermediate mutant is deleterious (Nowak *et al.* 2002; Komarova *et al.* 2003; Iwasa *et al.* 2004; Weissman *et al.* 2009). This makes previous evolutionary graph theory results in the limit of sequential fixation no longer predictive for populations where stochastic tunneling and fitness

valley crossing is prevalent.

Previous deme-based and lattice-based spatial models have shown that these spatial structures can accelerate the crossing of fitness valleys and plateaus, while delaying the evolution of complex traits with advantageous intermediates (Komarova 2014; Bitbol and Schwab 2014; Durrett and Moseley 2015). However, in contrast to more heterogenous graph spatial topologies, unless local differences in the nodes or demes are assumed, these regular, lattice-based structures do not change probabilities of fixation for intermediate mutants, and only affect valley crossing through changes to the extinction or fixation time of the intermediate (Pollak 1966; Maruyama 1970b; Lande 1979). Here we study the role of network topologies in shaping multi-mutational dynamics and show that the complex interplay between network properties that amplify or suppress selection on the intermediate mutant, combined with the effects of network motifs that accelerate or decelerate mutational spread gives rise to rich patterns of previously unexplored evolutionary outcomes.

In agreement with previous lattice and deme-based models (Komarova 2014; Bitbol and Schwab 2014), our results show that heterogenous spatial structure can promote fitness landscape crossing by allowing intermediate mutants to persist for longer, until the final beneficial mutation can appear in the population. However, in contrast with previous models, we show that more complex population structures rarely purely promote or suppress fitness landscape crossing and instead, switch across regimes as a function of the fitness of the intermediate mutants. We study the role of network properties in shaping rates of fitness valley crossing and compare these rates across well-studied families of graphs.

We also discuss how to apply these network-based approaches to large-scale datasets of spatial organization. We use previously studied datasets of the cellular networks of the stem cell niches in the bone marrow and apply our evolutionary model to study rates of mutation accumulation and leukemia initiation. Our results show that these cellular spatial architectures reduce the probability of neoplasm initiation across biologically relevant mutation rates and fitness distributions, compared to well-mixed populations. However, we show that there exists a threshold mutation rate (under exposure to carcinogens, for example) above which the bone marrow structure shifts from a suppressor into a promoter of neoplasm initiation. Our results provide the groundwork for further exploration of the role of heterogeneous spatial topology in shaping rates of mutation accumulation in both engineering and biological settings.

Model

We consider an asexual population of N haploid individuals and study the process by which this population acquires a beneficial trait that requires mutations at multiple loci. This process is usually referred to as the 'K-hit' mutational process. Here we study the $K = 2$ case (illustrated in **Figure 1**). We assume that, at time $t = 0$, a single first mutant, with reproductive fitness $(1 + s)$, appears in a population consisting of only wild-type individuals of fitness equal to one. Depending on the sign of the selection coefficient s, this first mutation can be beneficial (i.e., $s > 0$), neutral (i.e., $s = 0$), or deleterious (i.e., $s < 0$). With the probability of mutation upon reproduction μ, individuals carrying the first mutation can acquire a second mutation assumed to be extremely beneficial, such that fixation of double-mutant individuals is guaranteed with probability one. Therefore, the probability that all the individuals acquire the beneficial trait is equal to the probability of acquiring the second mutation. The second mutant can spread through the population either through a step-wise fixation process or through stochastic tunneling, without ever visiting a population fixed on the intermediate mutant (**Figure 1C**).

To represent heterogeneous population spatial structure, we use unweighted and undirected graphs, where each node in the graph represents one individual in the population and edges are proxies for the local pattern of replacement and substitution. A node can also represent a homogeneous subgroup of individuals and the edges as migration corridors between them, with the assumption that the timescale of a mutation traveling between nodes is much larger than the time scale of fixation within a node. We study the role of population structure in shaping rates of fitness valley crossing by analyzing how graph properties determine the probability that the population reaches the fitness peak.

To study graph properties independently of one another, we use well-known graph generation algorithms implemented in NetworkX (Hagberg *et al.* 2008), as well as algorithms that allow us to systematically tune network parameters (Kuo *et al.* 2021; Kuo and Carja 2021). We analyze a wide variety of graph families including preferential attachment graphs (Barabási and Albert 1999), bipartite graphs (Asratian *et al.* 1998), detour graphs (Möller *et al.* 2019), star-like graph (Tkadlec *et al.* 2019), small world networks (Watts and Strogatz 1998) used to model properties of social networks, and random geometric networks (Waxman 1988; Penrose *et al.* 2003), mathematical representations of populations embedded in Euclidean space. We also introduce regular island graphs which are created by adding random edges to two separate k-regular graphs and shortcut graphs which are created by adding edges connecting the loop of a detour graph. We present a detailed list of the graphs used in this study in the **Materials and methods** section.

We write analytic predictions for the probability of fitness valley crossing as a function of the population size N, the mutation rate μ, the intermediate mutant selection coefficient s, and the statistical properties of the spatial structure and compare these analytic predictions with results from Monte Carlo simulations. These analytic results allow us to generalize our insights beyond the graph families discussed here. To simulate the evolutionary trajectories of the population, at time $t = 0$, we introduce one intermediate mutant on a random node of the network and use a Moran Birth-death (Bd) process to track mutant frequency changes (**Figure 1B**). The birth-death process assumes that at each time step, an individual from the population is selected to reproduce proportional to fitness and a random neighbor node is selected to die, leaving an unoccupied node for the offspring of the reproducing node. This is the essential difference that allows us to study the role of local population structure, in comparison to a well-mixed population: in a well-mixed population, a node from the entire network would be randomly selected for death. Each time an intermediate mutant is selected to reproduce, with probability μ, it can acquire the second beneficial mutation. The simulation ends when either the intermediate mutant dies out or the second mutation is acquired. The fitness landscape crossing probability is then determined by the fraction of simulations that end with acquisition of the second mutation. Landscape crossing time is determined by averaging over the time it takes for the beneficial trait to fix, given fixation. Landscape crossing probabilities and times are computed using at least 10^7 simulations, for any given network structure.

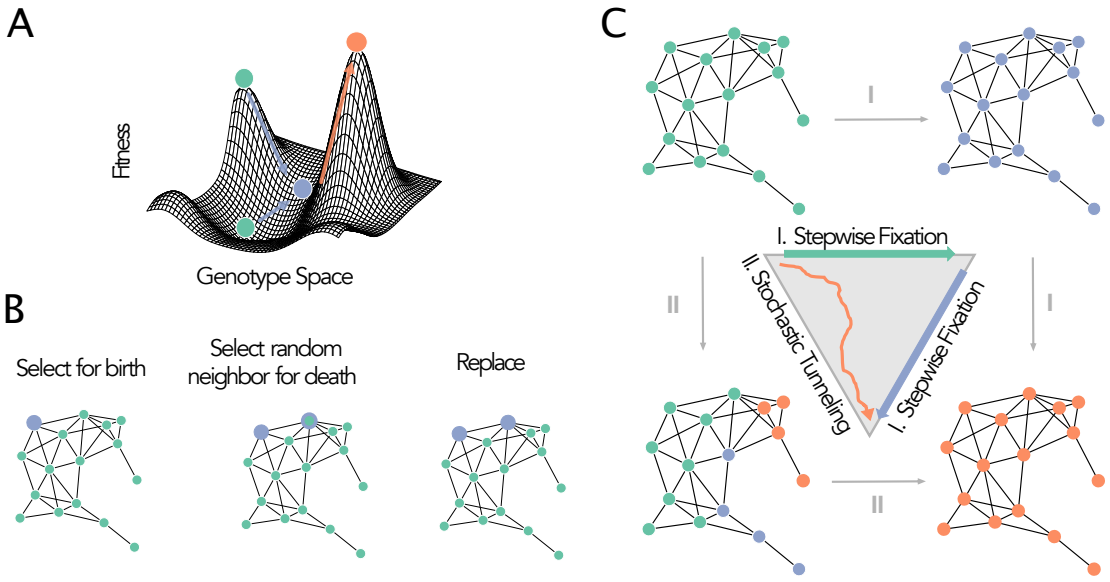

Figure 1 Illustration of the 2-hit process on network structured populations. Panel A. A two-step mutational process. The first mutation can be beneficial, neutral, or deleterious. The second mutation is assumed to be extremely beneficial, such that fixation is guaranteed. **Panel B.** The Bd (Birth-death) update rules. **Panel C.** There are two ways for the second mutation to fix in the population. The first is through step-wise fixation, where the intermediate mutant reaches fixation before the second mutation occurs (I). The second is through stochastic tunneling, where the second mutation is acquired before fixation of the intermediate (II).

Results

Landscape crossing is fundamentally shaped by both the probability and the time it takes for the first mutant to spread in the population and create opportunities for the appearance of the second mutant. Previous work has shown that, for single mutations, the evolutionary role of complex spatial structure can be quantified by analyzing two essential network properties: the network amplification factor, which shapes probabilities of fixation compared to well-mixed populations (Lieberman *et al.* 2005; Kuo *et al.* 2021) and the network acceleration factor, which shapes the time to fixation for new mutations in the population (Kuo and Carja 2021). The network amplification factor α quantifies how to rescale the selection coefficient s for the well-mixed model to obtain the same probability of fixation as an allele with selection coefficient s on a network population. In other words, a mutation with a selective benefit s in a graph population, would have a probability of fixation corresponding to an equivalent selection coefficient αs in a well-mixed population. If $\alpha > 1$, the graph amplifies selection, if $\alpha < 1$ the graph suppresses selection and if $\alpha = 1$ the graph does not change probabilities of fixation compared to the well-mixed. Similarly, we previously introduced the network acceleration factor, λ (Kuo and Carja 2021), and showed that this network property shapes times to fixation, making the evolutionary dynamics λ times faster than in a well-mixed population, for $\lambda > 1$.

While the amplification and acceleration have closed-form approximations using network descriptors, such as the mean and variance in degree (Kuo *et al.* 2021; Kuo and Carja 2021), they can also be computed using simulations and the definition from Lieberman *et al.* (2005) and solving

$$\Phi = \frac{1-(1+s)^{-\alpha}}{1-(1+s)^{-\alpha N}} \quad (1)$$

for α, where Φ is the single mutation fixation probability. The amplification factor is constant only for weak selection (McAvoy and Allen 2021), and varies with selection when selection is strong for many graphs (Voorhees 2013; Voorhees and Murray 2013; Allen *et al.* 2020). In this paper, we assume amplification factor is in the constant regime and we estimate it using $Ns = 1$. See more detailed description of how to compute network amplification and acceleration in **Materials and methods**.

The acceleration factor is defined as the ratio of conditional mean time to fixation for the equivalent well-mixed model (with the mutant having selection coefficient αs) and the network model with selective coefficient s (Kuo and Carja 2021), $\lambda = \frac{T_{wm}(\alpha s)}{T_{graph}(s)}$.

We will show that the interplay between how the network structure changes the fixation probability, the fixation time of the intermediate mutant and the mutation rate μ leads to rich landscape crossing dynamics that cannot be predicted or interpreted through the simpler lens of single mutation results. We start by presenting heuristic descriptions of the roles of these two evolutionary network parameters, α and λ, in shaping fitness landscape crossing and, using this intuition, we then present our full analytic results, with a detailed description of the analytic approach in the **Supplementary Material**.

The role of the network acceleration factor

Let us assume a network that only affects time to fixation and not the probability of fixation ($\alpha = 1$), for a single mutant. Any k-regular graph, for example, satisfies this property. There are two independent evolutionary scenarios that can lead to the acquisition of the second mutant.

The first scenario occurs when the intermediate mutant lineage is poised for extinction and the second mutation appears in the population before the lineage goes extinct. In a well-mixed

population, the lineage reaches a size T, smaller than the establishment threshold of $1/s$, in T generations and then drifts to extinction in another T generations, with probability $1/T$. The overall shape of such a trajectory is shown in **Figure 2A** and the total number of mutational opportunities to acquire the second mutation depends on the area under the trajectory, $W(T) = T^2$ (Weissman *et al.* 2009). The probability that a second mutation arises on the background of this mutant lineage is the expectation of the probability that a second mutation occurs given a trajectory that reaches maximum size of T, $\Phi(W(T))$, over all possible such trajectories,

$$P = \int \Phi(W(T))p(T)dT. \quad (2)$$

When the magnitude of s is strong enough ($|s| > \sqrt{\mu}$), the rate at which double-mutants are produced is dominated by rare lucky intermediate mutant lineages that survived for $T \sim 1/s$ (Weissman *et al.* 2009). Therefore, the probability of landscape crossing under this scenario is

$$P \sim \mu W\left(T = \frac{1}{s}\right) p\left(T = \frac{1}{s}\right) = \frac{\mu}{s}. \quad (3)$$

A network-structured population with acceleration factor λ effectively stretches the time the lineage reaches a size of T to T/λ generations, and also the time to extinction to $2T/\lambda$ generations. The area under the trajectory thus becomes T^2/λ and, using equation (3), the rate of landscape crossing is changed by a factor of λ^{-1}, to $\mu/\lambda s$.

The second scenario occurs when the intermediate mutant lineage would reach fixation on its own. In this scenario, the trajectory behaves stochastically below the establishment threshold, $1/s$. As soon as the mutant frequency crosses this threshold, the population dynamics becomes essentially deterministic (**Figure 2B**). At any point along the trajectory, the second mutation can occur and carry the mutant lineage to fixation. Since this scenario is conditioned on the first mutant fixing, the population is thus guaranteed to eventually acquire the second mutation and therefore, in this regime, the network structure does not change the probability of crossing the fitness valley. The total probability of crossing the fitness landscape is the sum of the probabilities of acquiring the second mutation under the two independent evolutionary processes (**Figure 2C**). When the first mutant is strongly deleterious, the population depends on the second mutation appearing in time to cross the fitness landscape and the acceleration factor of the network changes the rate of fitness valley crossing by a factor of λ^{-1}. Otherwise, the probability of stochastic tunneling remains unchanged by the network structure.

The role of the network amplification factor

Let us now assume a network structure that changes the fixation probability of a new mutant, but not the time to fixation ($\lambda = 1$) (**Figure 2**). An example of this type of structure are the shortcut graphs we present in **Figure 2F**.

For a network with amplification factor α, the mutant experiences an effective selection strength of αs in a well-mixed population. Therefore, if the intermediate mutant lineage is on its way to extinction, with probability $\frac{\alpha}{T}$, the mutant reaches an intermediate mutant size of $\frac{T}{\alpha} << \frac{1}{\alpha s}$ in $\frac{T}{\alpha}$ generations and goes to extinction in another $\frac{T}{\alpha}$ generations. The network amplification factor does not change the expected trajectory that reaches size $\frac{T}{\alpha}$, but instead changes the limit on the size that an intermediate mutant trajectory can reach, from a maximum of $\frac{1}{s}$ to $\frac{1}{\alpha s}$. Using equation (2), the rate of landscape crossing under this scenario becomes $\mu/\alpha s$.

In contrast, if the intermediate mutant linage is destined to reach fixation, the crossing probability is the same as the probability of fixation of the first mutant with effective fitness αs, since the population is guaranteed to acquire the second mutation. In this regime, the network structure does not change the probability of crossing the fitness valley, beyond modifying the effective selective coefficient of the intermediate mutant. Considering the two scenarios together, when the first mutant is strongly deleterious, the population depends on the first scenario to cross the fitness landscape, and the rate of landscape crossing is changed by a factor of $1/\alpha$. In the limit where the first mutant is strongly beneficial, fixation of the intermediate becomes the dominant process, and the crossing probability is approximately αs (**Figure 2F**).

The general analytic approximation

The previous sections provide heuristic intuition into the independent roles of the network amplification and acceleration factors in shaping probabilities of valley crossing. We detail the general analytic derivation in the **Supplementary Material** and show how to derive approximations for the two quantities of interest (as proxies for reaching the fitness peak): the probability that a second mutation arises in the population, and the expected time for that to occur. To do so, we use the diffusion approximation. Direct application of the diffusion approximation is difficult due to the complex dynamics that can occur on a graph. For that reason, the analysis is split into two parts.

In **Supplementary Material Section 1.1**, we begin by describing the deterministic population dynamics and identify the neutral equilibrium of the intermediate mutant in the population. This neutral equilibrium describes the distribution of the intermediate mutant over the nodes of graph, assuming zero stochastic fluctuations. To compute this neutral equilibrium, we show that we need to keep track of both the node and edge dynamics. In a well-mixed population or on the complete graph, keeping track of the mutant frequencies on the nodes is sufficient for determining the evolutionary fate of the population. However, here, the replacement between mutant and wild-type also depends on the frequency of edges that connect mutants and wild-type individuals.

In the second part of the analysis (**Supplementary Material Section 1.2**), we use the quasi-equilibrium to derive the stochastic equation for our model, which, in the limit of weak selection $s \ll 1$, weak mutation $N\mu \ll 1$, and large population size $1/N \ll 1$, can be written as

$$\frac{\partial \Phi}{\partial t} = \lambda q_a(1-q_a)\left(\frac{1}{N}\frac{\partial^2 \Phi}{\partial q_a^2} + \alpha s \frac{\partial \Phi}{\partial q_a}\right) + N\mu q_a(1-\Phi), \quad (4)$$

where $\Phi(q_a = \frac{1}{N}, t)$ is the probability of landscape crossing, when starting with one intermediate mutant in a random node on the graph (see also **Supplementary Material** equation (39)).

We next show that this equation is equivalent to the one governing the Moran process in a well-mixed population with the following changes: an effective mutant selection coefficient αs and an effective extended fixation time T_{fix}/λ (see **Supplementary Material Section 1.2.2** and, specifically, equation (48)). Thus, the arsenal of tools developed for well-mixed evolutionary theory can be applied to study landscape crossing in a network

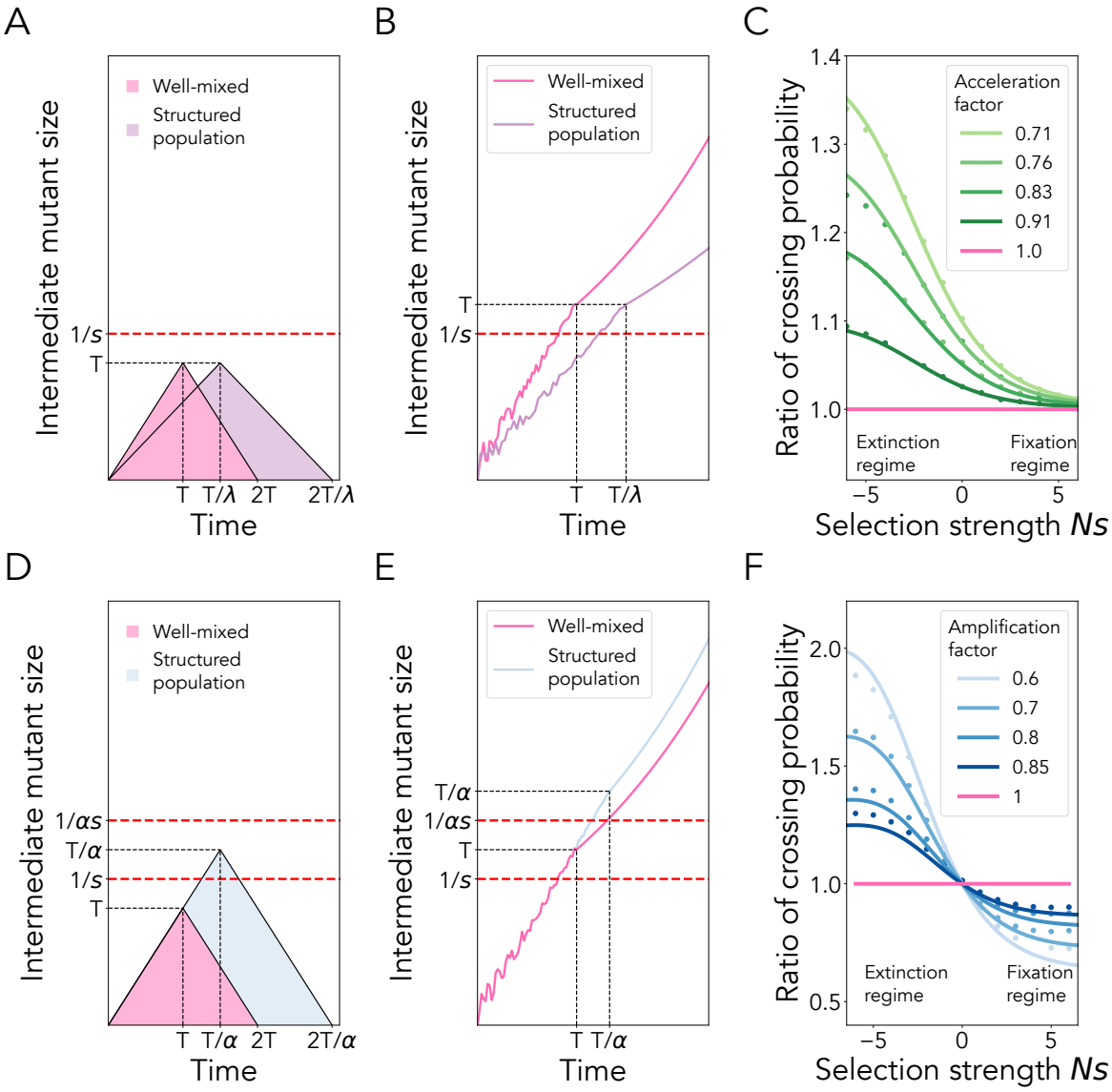

Figure 2 The role of the network amplification and acceleration factors in shaping the frequency of the intermediate mutant and fitness valley crossing probability. Top panels show dynamics for networks of constant amplification equal to 1. Bottom panels show dynamics for networks with acceleration equal to 1. **Panel A.** Typical trajectory of the intermediate mutant frequency, given that the intermediate mutant goes extinct. Population structures that increase extinction time lead to an increased probability of acquiring the second mutation. **Panel B.** Typical trajectory of the intermediate mutant frequency, given that the intermediate mutant fixes. **Panel C.** Ratio of crossing probability between k-regular graphs and the well-mixed model. Here, $\mu = 10^{-4}$ and the lines represent our numerical approximation to equation (4) in the limit of infinite time, using Supplementary equation (54), while the dots are results from 10^7 simulations per parameter and network. Network acceleration factors calculated using equation (10). **Panel D.** Typical trajectory of the intermediate mutant frequency, given the intermediate mutant goes extinct. A suppressor increases the line of transition between the stochastic and the deterministic dynamics, allowing the mutant to drift for longer and to a higher frequency. The structure would, therefore, increase the probability of acquiring the second mutation. **Panel E.** Typical trajectory of the intermediate mutant frequency, given the intermediate fixes. **Panel F.** Ratio of crossing probability between shortcut graphs with acceleration factor equal to one and the well-mixed model. Here, $\mu = 10^{-4}$ and the lines represent our numerical approximations to equation (4) (Supplementary equation (54)), while the dots are results from 10^7 simulations per parameter and network. Network amplification factors empirically determined using equation (9).

population and the crossing probability and time can be found by simply utilizing the transition matrix for the analogous well-mixed system. We further show how to use finite difference numerical approximations to solve for the crossing probabil-

ity (**Supplementary Material** equation (54)) and crossing time (**Supplementary Material** equation (64)).

Lastly, we show that the dynamics of fitness landscape crossing reduce to three simple cases when the population size becomes large. As N becomes large, we can approximate the tunneling probability using continuous time branching processes (Weissman *et al.* 2009) (see **Supplementary Material Section 1.2.3**) and write

$$\Phi = \frac{\alpha s - \lambda^{-1}\mu + \sqrt{(\alpha s + \lambda^{-1}\mu)^2 + 4\lambda^{-1}\mu}}{2(1+\alpha s)}. \tag{5}$$

Under weak mutation, series expansion in μ (see **Supplementary Material** equations (52)-(56)) reduces equation (5) to

$$\Phi = \begin{cases} \frac{\mu}{\alpha\lambda|s|} & \text{for } s \ll -2\sqrt{\mu} \\ \sqrt{\lambda^{-1}\mu} & \text{for } -2\sqrt{\mu} \ll s \ll 2\sqrt{\mu} \\ \alpha s & \text{for } s \gg 2\sqrt{\mu}. \end{cases} \tag{6}$$

In the case of a deleterious intermediate mutant, $s \ll -2\sqrt{\mu}$, equation (6) recaptures the crossing probability of $\mu/\lambda|s|$, when $\alpha = 1$, and $\mu/\alpha|s|$ when $\lambda = 1$. A network structure with amplification parameter α and acceleration parameter λ changes the crossing probability of the well-mixed population, $\mu/|s|$, by $1/\alpha\lambda$. Therefore, an amplifier, as defined in the single mutant case, can promote the fixation of a mutant linage with deleterious intermediates, as long as the the product of amplification and acceleration parameters is less than one. This is in stark contrast to the case of sequential fixation, where, in the case of a deleterious intermediate mutant, an amplifier always reduces the acquisition probability of the second mutation (**Figure 3A**). With neutral intermediate mutants, $-2\sqrt{\mu} \ll s \ll 2\sqrt{\mu}$, selection on these mutants has a negligible effect on the crossing probability: equation (6) only depends on λ and not α. In the case of a beneficial intermediate mutation, $2\sqrt{\mu} \gg s$, the population crosses the fitness landscape through intermediate mutants that are destined to fix. Since λ only affects the probability of acquiring the second mutation before fixation of the first, the crossing probability only depends on α and the crossing dynamics reduce to the sequential mutation case.

These results highlight that the intersection between the crossing probability of the structured population and that of the well-mixed case can shift from neutrality. For example, for the amplifier of selection in **Figure 3B**, there exists a region between the new intercept and neutrality, where the population structure behaves as a suppressor (see insert). In addition, for very deleterious intermediates, there exists an additional intercept, where the population structure acts as a suppressor on the left and an amplifier on the right. Single mutation fixation probabilities can therefore lead to erroneous inferences of multi-mutational landscape crossing dynamics.

The network structure can still act as an overall suppressor or amplifier, but only in a small region of the parameter space where the acceleration factor is exactly one. Beyond this knife-edge case, the general dynamics become more complicated, with the existence of piece-wise amplification or suppression, and we observe seven additional categories of graphs as the acceleration factor shifts away from one (**Figure 3D**).

These additional categories of graphs can be understood using the intuition developed in **Figure 2** and equation (6). We start from the largest values on the x-axis and go clockwise (light blue regime). In the regime where $\lambda < 1$, $\alpha > 1$ and $\alpha\lambda > 1$, the population structure is an amplifier except for weakly deleterious intermediate mutation. As λ decreases and $\alpha\lambda$ starts decreasing below one, a population structure can act as an amplifier for moderately deleterious mutations, but behaves like a suppressor for weakly and strongly deleterious intermediate mutants. As λ decreases further, $\alpha\lambda \ll 1$ and the population structure will increase crossing probability compared to a well-mixed population, regardless of the fitness of the intermediate (darker blue regime).

A network structure with $\alpha < 1$ and $\alpha\lambda \ll 1$ acts as a suppressor, increasing the crossing rate for a deleterious intermediate mutant. For a neutral mutation, the crossing probability depends on $1/\sqrt{\lambda}$ and, for $\lambda < 1$, the population structure acts as an amplifier for weakly beneficial mutations. As the strength of selection keeps increasing however, the crossing probability depends on $1/\alpha$ and the population structure starts acting as a suppressor, decreasing the rate of valley crossing (red regime). As λ increases above 1, but $\alpha\lambda < 1$, the population structure becomes an amplifier for weakly deleterious mutations (orange regime). There is a further sub-category of behavior where the ratio of crossing probabilities is higher than one for a moderately deleterious intermediate mutant. As λ increases even further, $\alpha\lambda$ is no longer smaller than one (light orange regime). The population structure starts behaving like an amplifier for deleterious mutations, crosses neutrality with a probability below the well-mixed, and behaves as a suppressor for beneficial mutations. No graphs exist in the upper right corner of the parameter space, since networks that are amplifiers have longer fixation time than well-mixed populations (Tkadlec *et al.* 2019).

Fitness valley crossing for different network families

Our results provide a unifying way of predicting multi-mutational dynamics by distilling properties of complex network organization through two single parameters: the acceleration and amplification factors. We can use this framework to compute the rate of fitness landscape crossing across known network families and we can design networks with the right ratio of amplification and acceleration to optimize for the rate of valley crossing. For example, for the case in equation (6) where the dynamics is controlled by $1/\alpha\lambda$, the graph with the highest crossing probability is the star graph (**Figure 4A**). Since the intermediate mutant lineage must wait for $\sim 1/\alpha\lambda$ for the second mutation to occur, the crossing time also scales with $\alpha\lambda$ (**Figure 4B**).

Using the crossing probability Φ and crossing time τ, we can approximate the effective rate of crossing the fitness landscape by $\left(\frac{1}{N\mu\Phi} + \tau\right)^{-1}$ (Frean *et al.* 2013). Using this approximation for the rate of evolution, we compare the main families of networks in **Figure 4C**. Lattice graphs and detour graphs have the highest rate of crossing the valley, out of all graphs we explored, overtaking strong amplifiers like the star graph. Another graph family of interest is the bipartite graph. These graphs have the lowest fixation time, for a given fixation probability, in the case of a single mutation. However, we show that they can span a wide range of crossing rates, showcasing that a network that is Pareto optimal for a single mutation in terms of fixation time and probability does not guarantee the best performance when crossing a fitness valley. The best-performing graph is inherently dependent on the underlying mutational landscape.

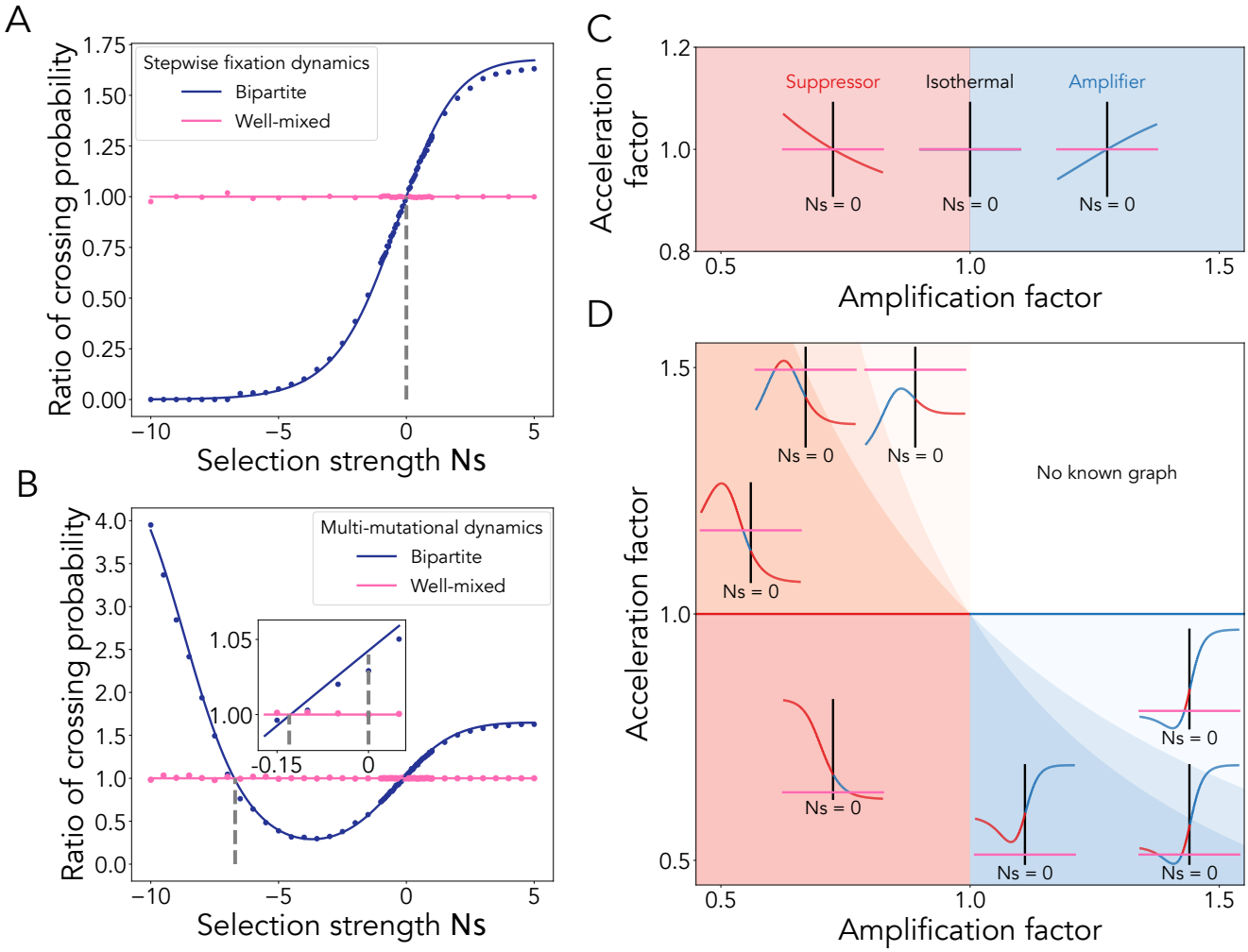

Figure 3 Network structures can switch from promoting to suppressing fitness valley crossing as a function of their amplification and acceleration properties. Panel A. We compare fixation probabilities of the second mutation in a structured population (a bipartite graph, with $\alpha = 1.7$, $\lambda = 0.1$) and a well-mixed population, under stepwise fixation dynamics ($\mu = 10^{-12}$). Here, dots represent simulation results from 10^7 replicates, and lines are calculated using Supplementary equation (54). The population size is $N = 100$. **Panel B.** Fixation probabilities of the second mutation in a bipartite graph, $\alpha = 1.7$, $\lambda = 0.1$, compared to those for a well-mixed population under a stochastic tunneling scenario ($\mu = 10^{-7}$). Here, dots represent simulation results from 10^7 replicate simulations per parameter and network, and lines represent analytic approximations from Supplementary equation (54). In this regime, the population structure crosses the well-mixed line twice, for negative selection strengths, and switches from promoting to suppressing fitness landscape crossing. Zoomed in version around $Ns = 0$ in the insert. The population size is $N = 100$. **Panel C.** The general amplifier suppressor classification for the single-mutation case. **Panel D.** The seven different scenarios for the ratio of crossing probabilities between a network over the well-mixed population, as a function of selection strength Ns and the two main evolutionary properties of the networks, the amplification and the acceleration factor. Here, the lines represent analytic approximations using equation (6) and the red and blue colors denote suppression and amplification of selection, respectively. The population size is $N = 1000$.

The roles of other network parameters

While the network amplification and acceleration factors play a unifying role in shaping evolutionary dynamics on the network, we also analyze how other network properties, in particular network connectivity and node heterogeneity, shape the fitness landscape crossing probability.

We use networks with preferential attachment, which allow for easy independent tuning of the number of edges and the shape of the degree distribution. We find that strong amplifiers of selection (preferential attachment graphs with low mean and high variance in degree) can increase the probability of fitness valley crossing compared to well-mixed populations (**Figure 5A**). This is contrary to expectation, since these graphs suppress deleterious intermediates in the case of sequential fixation. In evolutionary scenarios where the mutation rate is high or the population size is large, the rate of evolutionary dynamics in the population depends not only on the probability of crossing, but also on the expected time to cross. For the same mutational landscape as **Figure 5A**, the crossing time decreases with increasing mean degree and increases with increasing network node

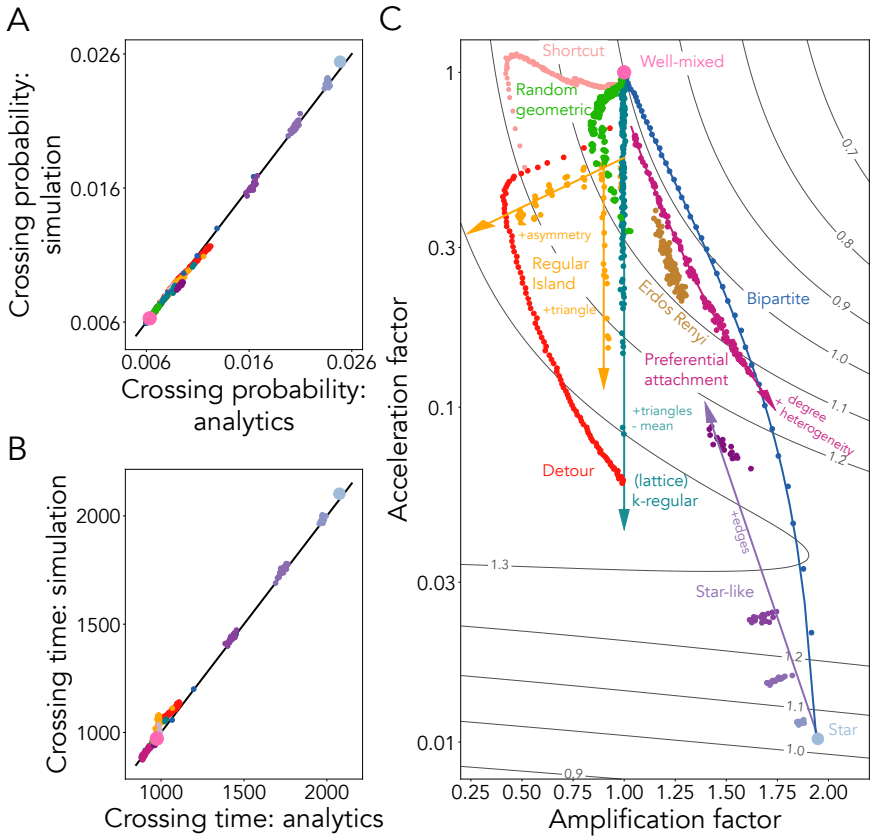

Figure 4 The rate of fitness valley crossing across network families. **Panel A.** Comparison of our analytic approximation (using Supplementary equation (54)) and simulations (using 10^7 replicate Monte Carlo simulations per parameter and network) for known graph families, colors as in Panel C. **Panel B.** Comparison of our analytic approximation for crossing time (Supplementary equation (64)) and simulations (10^7 replicate Monte Carlo simulations per parameter and network) for known graph families. **Panel C.** The rate of fitness valley crossing, as a function of the amplification and acceleration factor for graphs of size 100. The mutation rate is set to 10^{-4}. Here, $Ns = -1$. The black lines are contour lines of the ratio of crossing rate between the network and a well-mixed population and the dots represent individual networks of the families depicted. Network amplification factors empirically determined using equation (9). Network acceleration factors empirically determined using equation (10).

heterogeneity (**Figure 5B**, see **Supplementary Material** for the analytic derivation).

When the mutation rate is low, sequential fixation dominates and the relative crossing probability is equal to the relative fixation probability of the intermediate mutant. However, as the mutation rate increases, stochastic tunneling starts to become a dominant contributor to the crossing probability and the amplifier networks begin to act as suppressors of a deleterious intermediate mutation, increasing the crossing probability compared to the well-mixed model (**Figure 5C**). Our approximation holds for a range of mutational landscapes, ranging from fitness valleys to fitness plateaus and fitness hills (**Figure 5D**). In the strongly deleterious valley crossing regime, we see the largest deviation from the known theory of single mutation dynamics, since strongly deleterious mutations are unlikely to fix and successful crossing of the landscape depends solely on stochastic tunneling.

Application to bone marrow stem cell population architectures: rates of neoplasm initiation

We apply our model to study the propagation of somatic mutations in the stem cell architectures of the bone marrow to understand spatial factors shaping rates of leukemia initiation. We study a simple two-hit process. While this is a highly simplified model, it nonetheless highlights how the properties of spatial structure, in interplay with details of the mutational landscape, offer predictive power that is impossible to obtain with single mutation models. We assume there exists a set of cancer driver mutations with different effects on cell fitness and that the acquisition of any pair of these genes in the set can lead to neoplasm initiation (**Figure 6A**). The second cancer-initiating mutant is assumed to have an extremely high somatic fitness and to fix upon introduction and we assume a negative gamma distribution for the distribution of intermediate mutational effects (**Figure 6B**), shifted to the right to account for beneficial mutations (Piganeau and Eyre-Walker 2003):

$$P(s|\alpha,\beta,s_0) = \frac{\beta^\alpha}{\Gamma(\alpha)}(s_0 - s)^{\alpha-1}e^{\beta(s-s_0)}, \qquad (7)$$

where s_0 is the maximum fitness effect of the mutation, and α and β are parameters of the Gamma distribution.

We use published datasets that provide the spatial location of hematopoietic stem and progenitor cells in four samples of mouse tibia (Coutu *et al.* 2018) and the spatial locations of 8 bone

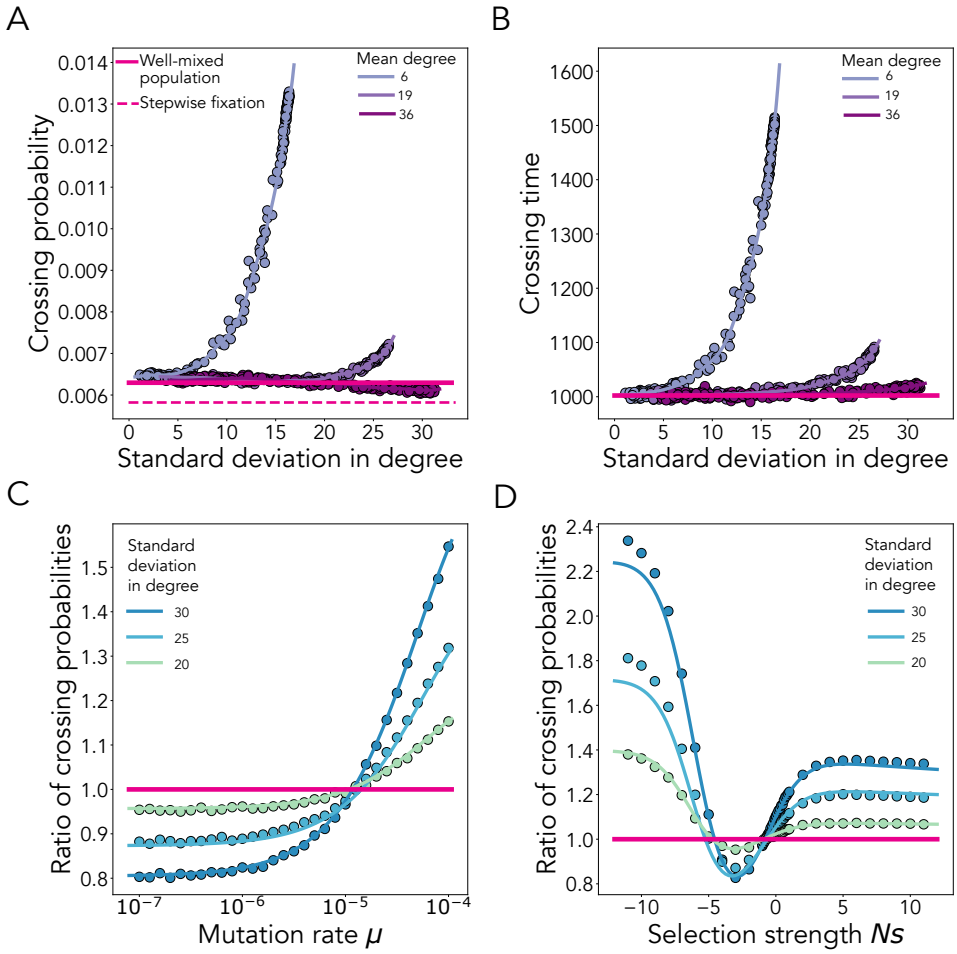

Figure 5 Amplifiers of selection can promote crossing of fitness valleys, even if the intermediate mutation is deleterious. The dots represent ensemble averages across 10^7 replicate Monte Carlo simulations per parameter set and network. Here we use preferential attachment graphs and approximations are obtained by fitting the degree standard deviation to the network amplification and acceleration factors (see **Supplementary Figure S3**). Lines show the results of substituting the fitting functions into Supplementary equation (54). **Panel A.** Here, for dots of the same color, the mean degree of the network is held constant (as in the legend) and we vary the variance of the degree distribution. Population size $N = 100$, $\mu = 10^{-5}$ and $s = -0.01$, such that $Ns = -1$. **Panel B.** The crossing time for networks with different variance of the degree distribution. Here, $N = 100$, $\mu = 10^{-5}$ and $s = -0.01$, such that $Ns = -1$. Lines show the results of substituting the fitting functions into Supplementary equation (64). **Panel C.** The ratio of the crossing probability for the graph-structured population over the crossing probability of a well-mixed population, as a function of mutation rate. Here, $N = 100$, mean degree is 19 and $s = -0.01$. **Panel D.** The ratio of the crossing probability for the graph-structured population over the crossing probability of a well-mixed population, as a function of the selective coefficient of the intermediate mutant. Here, $N = 100$, $\mu = 10^{-5}$ and mean degree is 19.

marrow samples of CXCL12-abundant reticular cells (which critically modulate hematopoiesis at various levels, including hematopoietic stem cell maintenance), each with two images of two anatomically distinct regions (the diaphysis and the metaphysis), in total 16 cellular populations (Gomariz *et al.* 2018).

We use previously built cellular networks of the stem cell niches of the bone marrow (Kuo *et al.* 2021), where every Hematopoietic Stem Cell (HSC) niche constitutes a node in the graph and an edge is added between two nodes if the distance between them is less than a cut-off radius, similar to the generation of a random geometric graph. Kuo *et al.* (2021) showed that, across a wide variety of parameters and regardless of the birth death process used, these networks are strong suppressors of selection, potentially delaying mutation accumulation in this tissue.

For a range of biologically relevant mutation rates, and using these previous estimates of amplification and acceleration factor for these networks, we use our analytic results for the rate of fitness landscape crossing to compute the total rate of neoplasm initiation,

$$N\mu \int p(s)\Phi\Big(\frac{1}{N}\Big|s,\mu,N\Big)\,ds. \tag{8}$$

We find that the change in the rate of cancer initiation, compared to a well-mixed population, depends heavily on the assumed fraction of incoming beneficial mutations into the population (**6C**). At smaller, native mutation rates, the bone marrow mostly suppresses the rate of leukemia initiation, but at higher mutation rates, under environmental carcinogens, for example,

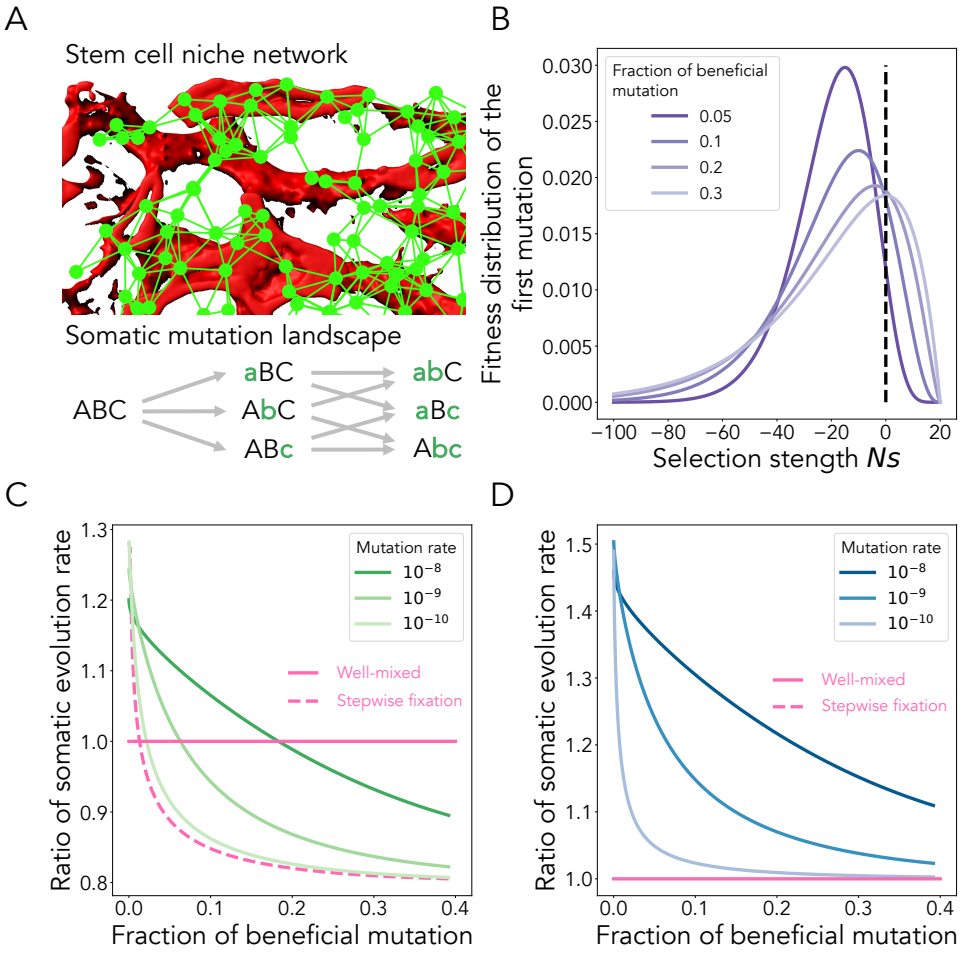

Figure 6 A 2-hit model of tumor initiation for bone marrow cellular architectures. **Panel A.** Illustration of the hematopoietic stem cells niche architecture (for a full description see (Kuo *et al.* 2021)) and a 2-hit model. In this example, there are 3 genes where any combination of two mutations leads to neoplasm initiation. The letters represent a genotype and the arrows represents mutation paths. **Panel B.** We assume the intermediate mutations to have different fitness effects, as represented by a shifted negative gamma distribution. Here, the gamma distribution has parameters $s_0 = 4 \times 10^{-4}$, $\alpha \in \{2, 2.5, 4, 8\}$, and $\beta \in \{4 \times 10^{-4}, 3.2 \times 10^{-4}, 2 \times 10^{-4}, 1 \times 10^{-4}\}$, such that the mean of all four distributions presented is equal to $(s_0 - \alpha\beta) = -4 \times 10^{-4}$. The fraction of beneficial mutations is proportional to parameter β, such that $\beta = 4 \times 10^{-4}$ corresponds to the darkest purple line and $\beta = 1 \times 10^{-4}$ corresponds to the lightest purple line. The population size is N=50000. **Panel C.** Ratio of cancer initiation rate between the bone marrow tissue cellular architecture and the well-mixed population, as a function of the shape of the fitness distribution and the mutation rate. Ratios are determined using Supplementary equation (54). Here, as above, the gamma distribution has a fixed mean selection strength of $s = -4 \times 10^{-4}$. The population size is $N = 50000$. **Panel D.** The ratio of cancer initiation rate between a 4-regular graph and the well-mixed population, as a function of the shape of the fitness distribution and the mutation rate. Here, the 4-regular graph has 0 triangles. The acceleration factor is 0.66, calculated using Supplementary equation (73). As above, the gamma distribution has a fixed mean selection strength of $s = -4 \times 10^{-4}$. The population size is $N = 50000$. Here, the well-mixed line overlaps the stepwise fixation line, since a regular graph does not affect single mutation probability of fixation.

the cellular structure can increase the rate of cancer initiation compared to the well-mixed population, for a much wider range in the fraction of incoming beneficial intermediates (**Figure 6C**). This is in contrast to lattice-based populations, where the population topology promotes mutation accumulation compared to well-mixed, regardless of the mutation rate and the fitness distribution of the mutation (**Figure 6D**). These results show that spatial heterogeneity in tissue architectures can reduce the rates of mutation accumulation and tumor initiation in ways that can be overlooked by previous lattice-based spatial models.

Discussion

Here we study how network structure shapes fitness landscape crossing, compared to well-mixed populations. Our results extend previous well-mixed, lattice or deme-based models (Komarova 2014; Bitbol and Schwab 2014) and analyze non-symmetric spatial structures of controllable complexity. These previous models show that, by changing the time intermediate mutants can persist until the beneficial final mutation occurs, these spatial structures can help populations cross fitness valleys, compared with well-mixed populations. Here, we recover previous results for lattice-based populations by studying reg-

ular graphs with amplification factor $\alpha = 1$ (Komarova 2014; Bitbol and Schwab 2014). However, more complex network structures can change not just the time to fixation, but also the probability of fixation for intermediate variants and the interplay between how the structure affects probabilities and times to fixation and the mutation rate towards the second mutant can give rise to complex evolutionary dynamics, not observed in previous models.

Using simulation and analytic approximations, we show that these complex evolutionary dynamics can be intuitively explained by analyzing how two main network properties, the amplification and acceleration factors, change the expected fate of the intermediate mutant in the population. This intuition can be used to inform on the design of population topologies that optimize rates of crossing fitness landscapes: 1) network-structured populations with low acceleration and amplification factors cross fitness valleys and plateaus more effectively, while 2) network populations with high amplification factors cross fitness hills more effectively. The first observation can be understood and interpreted using Wright's shifting balance theory, since through either decreased effective selection or increasing hindrance to gene flow, an increase in the relative force of drift leads to an increased window in which the final beneficial mutation can rescue the mutant lineage (Wright *et al.* 1932).

We find that most network structures rarely purely promote or purely suppress fitness valley crossing, instead transitioning across regimes depending on the selective coefficients of the intermediate mutants. For example, our results show that single-mutation amplifiers, like bipartite graphs and preferential attachment graphs, can increase valley crossing probability for strongly and weakly deleterious intermediate mutations, yet decrease valley crossing for nearly neutral intermediates. Overall, we find seven different network categories that classify evolutionary dynamics of fitness landscape crossing for network-based populations.

For mathematical tractability, obtaining analytical approximations on large, heterogeneous graph structures comes at the cost of simplifying the evolutionary process assumed on the graph. While in this study we compare our results with deme-based or lattice-based models of the Moran process, where all nodes are assumed to be equivalent and probabilities of fixation for the mutant intermediate are not changed, there is also a large literature studying more complex dynamics on populations of demes or on regular graph structures (Barton 1993; Yagoobi and Traulsen 2021; Marrec *et al.* 2021; Marrec 2023). For example, in Barton (1993) random deme extinction is introduced as an extra component of sampling drift, which can shape intermediate mutant fixation probabilities. Other models, such as Bitbol and Schwab (2014), allow for the uncoupling of migration, birth, and death and, thus, exploration over a broader parameter space. Further work is needed to understand how relaxing the modeling assumptions we make here interplays with heterogenous network structure to shape landscape crossing dynamics.

One key limitation of our analysis is the focus on low selection strength for the intermediate mutant. In this selection regime, the amplifier / suppressor framework, with constant strength of amplification with respect to selection is a valid first-order approximation (McAvoy and Allen 2021). Under a broader range of selection however, there is a wide range of graphs that are piece-wise or transient amplifiers / suppressors, with graphs having selection-dependent amplification factors (Voorhees 2013; Voorhees and Murray 2013; Allen *et al.* 2020). A notable example is a reducer, a network that consistently decreases the fixation probability compared to a well-mixed model, irrespective of the selection strength of the mutant (Allen *et al.* 2020; Hindersin *et al.* 2016). Finer categorizations of amplifiers also exist when considering different mutant initializations, where the mutant does not appear uniformly in the population (Allen *et al.* 2021). Extending the analytic treatment of crossing dynamics to these selection regimes is challenging and further study is required. Previous work studying single mutation probabilities of fixation has also shown that the type of birth death process assumed on the network can significantly shape the observed dynamics and further work is needed to understand landscape crossing for other processes, such as death-Birth (Hindersin and Traulsen 2015; Kuo *et al.* 2021).

Our model can be applied to understand the evolutionary trajectory of cellular systems with complex spatial architectures, and we use an existing dataset of cellular structure to study how the spatial organization of the stem cell niches in the bone marrow shapes rates of leukemia initiation. We find that the relative rate towards neoplasm initiation in this spatial cellular architecture depends critically on the mutation rate, as well as the proportion of beneficial mutations coming into the population. We show that a structure that suppresses cancer initiation can instead transition into one promoting cancer initiation, depending on the distribution of mutational effects. A previous analysis of nearly 50,000 healthy individuals showed that most somatic mutations confer a fitness advantage (Watson *et al.* 2020). Under a regime of prevalent beneficial driver mutations, we show that the bone marrow architecture suppresses accumulation of these mutations compared to well-mixed populations. However, under exposure to carcinogens (increased rate of mutation (Nowak *et al.* 2002; Pino and Chung 2010; Tomlinson *et al.* 1996)), the architecture of the bone marrow can instead increase rates of neoplasm initiation. As spatial datasets of cellular and molecular organization become increasingly available, linking these data sets with theoretical models of evolutionary spread will be crucial for a rigorous understanding of how complex spatial structure shapes evolutionary dynamics at these levels of organization.

Materials and methods

List of network families used in the study

Linking network topology to evolutionary dynamics and understanding which network properties shape rates of evolution is complicated by the fact that these properties are often correlated, hard to tune independently and differ across many network families (Kuo *et al.* 2021). In this study, we explore both well-known network families using built-in generators from NetworkX (Hagberg *et al.* 2008) and also design graphs that allow us to tune properties independently, as detailed below.

***k*-regular graphs.** A k-regular graph is a graph where each node has the same number of neighbors, k. We generate k-regular graphs using built-in generators from NetworkX (Hagberg *et al.* 2008), with $k \in \{3, 5, 10, 20\}$. For some of these graphs, to generate Figure 4, we further tune the number of triangles and change their acceleration factor (while keeping mean degree constant at k), using methods perviously outlined in Kuo and Carja (2021).

Erdős Rényi random networks. The Erdős Rényi model starts with a set of N isolated nodes, and connects two nodes with

probability p. We generate Erdős Rényi random networks using built-in generators from NetworkX (Hagberg *et al.* 2008).

Preferential attachment graphs. For graphs with preferential attachment, nodes are added sequentially starting from one initial node until the population reaches size N. Each new node is added to the network and connected to other individuals with a probability proportional to the individual's current degree to the power of a given parameter β. Using k and β, this family of graphs allows for straightforward independent tuning of mean and variance in degree. We generate these networks using NetworkX (Hagberg *et al.* 2008) with $k \in \{3, 5, 20\}$, and $\beta \in [-3, 3]$.

Small world networks. We generate small world networks using built-in generators from NetworkX (Hagberg *et al.* 2008) and the Watts–Strogatz model with parameters: number of nodes N, the mean degree k, and rewire probability p. We use $N = 100$, $k \in \{8, 12, 16\}$, and $p \in [0, 1]$.

Bipartite graphs. Bipartite graphs have two node sets n_1 and n_2. Edges in the graph only connect nodes from opposite sets. We generate bipartite graphs using built-in generators from NetworkX (Hagberg *et al.* 2008). We vary n_1 from 1 to 50, and $n_2 = 100 - n_1$.

Random geometric graphs. The random geometric graph model places N nodes at random following a probability distribution. Two nodes are joined by an edge if the distance between the nodes is below a predefined cut-off radius. For the graphs in Figure 4, we place nodes in 2-dimensional space. For the x-position, 50 nodes are drawn from a normal distribution $\mathcal{N}(0, 0.25)$, and 50 nodes are drawn from $\mathcal{N}(3, 0.25)$. The y-position is drawn from a normal distribution $\mathcal{N}(0, 0.25)$. We start with the lowest cut-off radius, resulting in a connected graph, and increase the cut-off radius until the complete graph is formed.

Detour graphs. A detour graph is formed by starting with a complete graph of size n_1 and replacing one of the edges with a path of length $n_2 + 1 \geq 2$ (Möller *et al.* 2019). To generate the detours in Figure 4, we vary n_1 from 3 to 99, and $n_2 = 100 - n_1$.

Shortcut graphs. We design the shortcut graphs that allow us to vary amplification, while keeping acceleration equal to one used in Figure 2F and Figure 4, by starting with a detour graph and adding edges (shortcuts) between two nodes. We add 1 to 255 edges to the detour graph with $n_1 = 90$ and $n_2 = 10$.

Star and star-like graphs. The star graph consists of one center node connected to $N - 1$ outer nodes. We generate star graphs using built-in generators from NetworkX (Hagberg *et al.* 2008). A star-like graph is constructed by adding random edges to the star graph (Tkadlec *et al.* 2019). We generate 20 graphs each for 50, 100, 200, and 300 random edges added to a star-graph of size 100.

Regular islands. A regular island is the result of connecting two k-regular graphs. A random edge is picked from each k-regular graph, and edge swap is performed to connect the two k-regular graphs.

Computing the amplification and acceleration parameters for a given network

Here we show that the amplification and acceleration factors of a network, computed from single mutation analyses, if used together, can be predictive of the landscape crossing behavior of network structured populations. In previous work, we showed how to analytically compute the amplification and acceleration factors (Kuo *et al.* 2021; Kuo and Carja 2021). The approximation for the amplification factor from Kuo *et al.* (2021) is very accurate for amplifiers and can slightly deviate from the true amplification factor for suppressors (see also equation (37) in the **Supplementary Material**). Similarly, the approximations described in Kuo and Carja (2021) for computing the acceleration factor are very accurate if triangles in the network are distributed equally amongst the triplet types (see also equation (38) in the **Supplementary Material**).

These factors can also be computed empirically for all network families and types using single-mutation simulations, as detailed below. Using the Birth-death process, and starting with one single mutant with fitness $(1 + s)$ invading a population of wild-type individuals with fitness 1, the amplification factor can be computed using the definition from (Lieberman *et al.* 2005) and solving

$$\Phi = \frac{1 - (1 + s)^{-\alpha}}{1 - (1 + s)^{-\alpha N}} \tag{9}$$

for α, where Φ is the mutant's fixation probability.

For graphs in this paper, we estimate α using $Ns = 1$. When graph size is $N = 100$, this equates to $s = 0.01$. This value is large enough such that the difference in fixation probability compared to the well-mixed model is big enough, while also small enough to be in the constant regime. The amplification factor for a given network is estimated once and used to study landscape crossing with varying selection on the intermediate mutant.

For the acceleration factor, we use the following definition

$$\lambda = \frac{T_{fix,wm}(\alpha s)}{T_{fix,graph}(s)}, \tag{10}$$

which is the ratio of conditional mean fixation time for an equivalent mutant on the well-mixed population (with selection coefficient αs) and the one on the network population with mutant selection coefficient s (Kuo and Carja 2021). For the graphs presented here, we evaluate λ at $s = 0$, since there is a closed-form expression for $T_{fix,wm}(s = 0)$, $(N - 1)$ generations (Ewens 2004). For weak s, λ varies no more than $\mathcal{O}(s)$.

Building the cellular networks of the stem cell niches of the bone marrow

The samples vary in dimensions, number of cells, and segmentation techniques. We normalize the data by expressing the distance in units of the average shortest distance between pairs of cells (62.72 μm for hematopoietic stem cells). We use networks generated using cut-off distance of 15, which is the closest to our estimated biological interaction range. We interpret the cut-off distance as the maximum distance a HSC could travel in its entire lifespan. Live-animal tracking of individual hematopoietic stem cells in their niche showed MFG cells, a largely quiescent population with long-term self-renewal capability, displacing an average distance of 8.69μm in a 2.5 hour period (Christodoulou *et al.* 2020). HSCs have median replication time (the time when 50% of HSCs have divided) of 1.7 weeks (Abkowitz *et al.* 2000). During homeostasis, the rate of replication should balance the rate of depletion. This leads to the estimated interaction range of 1028μm which corresponds to 16.4$\times$ the distance between the shortest pairs. More details on network generation and robustness results are presented in Kuo *et al.* (2021).

The amplification and acceleration factors of the bone marrow network structures are then calculated using Monte Carlo simulation of single mutation fixation of 10^7 replicate simulations at $Ns = 0.01$. The two factors are used to derive the analytical approximation for the crossing probability in these cellular populations, shown here together with simulations (**Supplementary Figure 6A**). Although the image data provides precise spatial description of the hematopoietic stem cell niches, these only represent a small portion of the HSC population. Previous models estimate the number of hematopoietic stem cells that are actively making white blood cells at any one time to be in the range of 50,000-200,000 (Lee-Six *et al*. 2018). Here, we assume the overall spatial distribution of the HSC population across the bone marrow is identical to our sample. We use linear regression to extrapolate the amplification factor to the lower bound population of 50,000 (**Supplementary Figure 6B**). The acceleration factors inferred from data are close to 1, so we assume the larger population also has acceleration factor of 1.

Data availability

The graphs used in this study, information about the C++ and Python code used to simulate and analyze the model presented here are available at https://github.com/yangpingkuo/Evolutionary-graph-theory-beyond-single-mutation-dynamics.

Acknowledgments

We gratefully acknowledge support from the NIH National Institute of General Medical Sciences (award no. R35GM147445), the United States-Israel Binational Science Foundation (award no. 2019266) and from the NIH T32 training grant (no. T32 EB009403). This research was done using resources provided by the Open Science Grid, which is supported by the National Science Foundation award 1148698, and the U.S. Department of Energy's Office of Science.

Conflicts of interest

The authors declare no conflicts of interest.

Literature cited

Abkowitz JL, Golinelli D, Harrison DE, Guttorp P. 2000. In vivo kinetics of murine hemopoietic stem cells. Blood, The Journal of the American Society of Hematology. 96:3399–3405.

Acar A, Nichol D, Fernandez-Mateos J, Cresswell GD, Barozzi I, Hong SP, Trahearn N, Spiteri I, Stubbs M, Burke R *et al*. 2020. Exploiting evolutionary steering to induce collateral drug sensitivity in cancer. Nature Communications. 11:1923.

Adlam B, Chatterjee K, Nowak MA. 2015. Amplifiers of selection. Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences. 471:20150114.

Allen B, Gore J, Nowak MA. 2013. Spatial dilemmas of diffusible public goods. Elife. 2:e01169.

Allen B, Sample C, Jencks R, Withers J, Steinhagen P, Brizuela L, Kolodny J, Parke D, Lippner G, Dementieva YA. 2020. Transient amplifiers of selection and reducers of fixation for death-birth updating on graphs. PLoS Computational Biology. 16:e1007529.

Allen B, Sample C, Steinhagen P, Shapiro J, King M, Hedspeth T, Goncalves M. 2021. Fixation probabilities in graph-structured populations under weak selection. PLoS Computational Biology. 17:e1008695.

Asratian AS, Denley TM, Häggkvist R. 1998. *Bipartite graphs and their applications*. volume 131. Cambridge University Press.

Barabási AL, Albert R. 1999. Emergence of scaling in random networks. Science. 286:509–512.

Barton NH. 1993. The probability of fixation of a favoured allele in a subdivided population. Genetical Research. 62:149–157.

Bitbol AF, Schwab DJ. 2014. Quantifying the role of population subdivision in evolution on rugged fitness landscapes. PLoS Computational Biology. 10:e1003778.

Burch CL, Chao L. 1999. Evolution by small steps and rugged landscapes in the RNA virus ϕ6. Genetics. 151:921–927.

Carja O, Liberman U, Feldman MW. 2014. Evolution in changing environments: Modifiers of mutation, recombination, and migration. Proceedings of the National Academy of Sciences. 111:17935–17940.

Christodoulou C, Spencer JA, Yeh SCA, Turcotte R, Kokkaliaris KD, Panero R, Ramos A, Guo G, Seyedhassantehrani N, Esipova TV *et al*. 2020. Live-animal imaging of native haematopoietic stem and progenitor cells. Nature. 578:278–283.

Coutu DL, Kokkaliaris KD, Kunz L, Schroeder T. 2018. Multicolor quantitative confocal imaging cytometry. Nature Methods. 15:39–46.

Durrett R, Moseley S. 2015. Spatial moran models i. stochastic tunneling in the neutral case. The annals of applied probability: an official journal of the Institute of Mathematical Statistics. 25:104.

Ewens WJ. 2004. *Mathematical population genetics: theoretical introduction*. volume 27. Springer.

Frean M, Rainey PB, Traulsen A. 2013. The effect of population structure on the rate of evolution. Proceedings of the Royal Society B: Biological Sciences. 280:20130211.

Gomariz A, Helbling PM, Isringhausen S, Suessbier U, Becker A, Boss A, Nagasawa T, Paul G, Goksel O, Székely G *et al*. 2018. Quantitative spatial analysis of haematopoiesis-regulating stromal cells in the bone marrow microenvironment by 3d microscopy. Nature Communications. 9:2532.

Hagberg AA, Schult DA, Swart PJ. 2008. Exploring network structure, dynamics, and function using networkx. In: Varoquaux G, Vaught T, Millman J, editors, *Proceedings of the 7th Python in Science Conference*. pp. 11 – 15. Pasadena, CA USA.

Hindersin L, Traulsen A. 2014. Counterintuitive properties of the fixation time in network-structured populations. Journal of The Royal Society Interface. 11:20140606.

Hindersin L, Traulsen A. 2015. Most undirected random graphs are amplifiers of selection for birth-death dynamics, but suppressors of selection for death-birth dynamics. PLoS Computational Biology. 11:e1004437.

Hindersin L, Werner B, Dingli D, Traulsen A. 2016. Should tissue structure suppress or amplify selection to minimize cancer risk? Biology Direct. 11:1–11.

Huang S. 2013. Genetic and non-genetic instability in tumor progression: link between the fitness landscape and the epigenetic landscape of cancer cells. Cancer and Metastasis Reviews. 32:423–448.

Iwasa Y, Michor F, Nowak MA. 2004. Stochastic tunnels in evolutionary dynamics. Genetics. 166:1571–1579.

Jain K, Krug J. 2007. Deterministic and stochastic regimes of asexual evolution on rugged fitness landscapes. Genetics. 175:1275–1288.

Jiang C, Chen Y, Liu KR. 2014. Evolutionary dynamics of information diffusion over social networks. IEEE Transactions on

Signal Processing. 62:4573–4586.
Kimura M, Weiss GH. 1964. The stepping stone model of population structure and the decrease of genetic correlation with distance. Genetics. 49:561.
Komarova NL. 2014. Spatial interactions and cooperation can change the speed of evolution of complex phenotypes. Proceedings of the National Academy of Sciences. 111:10789–10795.
Komarova NL, Sengupta A, Nowak MA. 2003. Mutation–selection networks of cancer initiation: tumor suppressor genes and chromosomal instability. Journal of Theoretical Biology. 223:433–450.
Kuo YP, Arrieta CN, Carja O. 2021. A theory of evolutionary dynamics on any complex spatial structure. bioRxiv. .
Kuo YP, Carja O. 2021. Evolutionary graph theory beyond pairwise interactions: higher-order network motifs shape times to fixation in structured populations. bioRxiv. .
Kvitek DJ, Sherlock G. 2011. Reciprocal sign epistasis between frequently experimentally evolved adaptive mutations causes a rugged fitness landscape. PLoS Genetics. 7:e1002056.
Lande R. 1979. Effective deme sizes during long-term evolution estimated from rates of chromosomal rearrangement. Evolution. pp. 234–251.
Lee-Six H, Øbro NF, Shepherd MS, Grossmann S, Dawson K, Belmonte M, Osborne RJ, Huntly BJ, Martincorena I, Anderson E *et al*. 2018. Population dynamics of normal human blood inferred from somatic mutations. Nature. 561:473–478.
Leventhal GE, Hill AL, Nowak MA, Bonhoeffer S. 2015. Evolution and emergence of infectious diseases in theoretical and real-world networks. Nature Communications. 6:6101.
Lieberman E, Hauert C, Nowak MA. 2005. Evolutionary dynamics on graphs. Nature. 433:312–316.
Maciejewski W, Fu F, Hauert C. 2014. Evolutionary game dynamics in populations with heterogenous structures. PLoS Computational Biology. 10:e1003567.
Marrec L. 2023. Quantifying the impact of genotype-dependent gene flow on mutation fixation in subdivided populations. bioRxiv. .
Marrec L, Lamberti I, Bitbol AF. 2021. Toward a universal model for spatially structured populations. Physical review letters. 127:218102.
Maruyama T. 1970a. Effective number of alleles in a subdivided population. Theoretical population biology. 1:273–306.
Maruyama T. 1970b. On the fixation probability of mutant genes in a subdivided population. Genetics Research. 15:221–225.
McAvoy A, Allen B. 2021. Fixation probabilities in evolutionary dynamics under weak selection. Journal of Mathematical Biology. 82:1–41.
Möller M, Hindersin L, Traulsen A. 2019. Exploring and mapping the universe of evolutionary graphs identifies structural properties affecting fixation probability and time. Communications Biology. 2:1–9.
Nowak MA, Komarova NL, Sengupta A, Jallepalli PV, Shih IM, Vogelstein B, Lengauer C. 2002. The role of chromosomal instability in tumor initiation. Proceedings of the National Academy of Sciences. 99:16226–16231.
Ohtsuki H, Hauert C, Lieberman E, Nowak MA. 2006. A simple rule for the evolution of cooperation on graphs and social networks. Nature. 441:502–505.
Ohtsuki H, Nowak MA. 2006. The replicator equation on graphs. Journal of Theoretical Biology. 243:86–97.
Ohtsuki H, Pacheco JM, Nowak MA. 2007. Evolutionary graph theory: Breaking the symmetry between interaction and replacement. Journal of Theoretical Biology. 246:681–694.
Penrose M *et al*. 2003. *Random geometric graphs*. volume 5. Oxford university press.
Piganeau G, Eyre-Walker A. 2003. Estimating the distribution of fitness effects from DNA sequence data: implications for the molecular clock. Proceedings of the National Academy of Sciences. 100:10335–10340.
Pino MS, Chung DC. 2010. The chromosomal instability pathway in colon cancer. Gastroenterology. 138:2059–2072.
Pollak E. 1966. On the survival of a gene in a subdivided population. Journal of Applied Probability. 3:142–155.
Poncela J, Gómez-Gardenes J, Floría LM, Moreno Y. 2007. Robustness of cooperation in the evolutionary prisoner's dilemma on complex networks. New Journal of Physics. 9:184.
Rogers ZN, McFarland CD, Winters IP, Seoane JA, Brady JJ, Yoon S, Curtis C, Petrov DA, Winslow MM. 2018. Mapping the in vivo fitness landscape of lung adenocarcinoma tumor suppression in mice. Nature Genetics. 50:483–486.
Salehi S, Kabeer F, Ceglia N, Andronescu M, Williams MJ, Campbell KR, Masud T, Wang B, Biele J, Brimhall J *et al*. 2021. Clonal fitness inferred from time-series modelling of single-cell cancer genomes. Nature. 595:585–590.
Santos FC, Pacheco JM, Lenaerts T. 2006. Evolutionary dynamics of social dilemmas in structured heterogeneous populations. Proceedings of the National Academy of Sciences. 103:3490–3494.
Slatkin M. 1981. Fixation probabilities and fixation times in a subdivided population. Evolution. pp. 477–488.
Su Q, Allen B, Plotkin JB. 2022. Evolution of cooperation with asymmetric social interactions. Proceedings of the National Academy of Sciences. 119:e2113468118.
Tkadlec J, Pavlogiannis A, Chatterjee K, Nowak MA. 2019. Population structure determines the tradeoff between fixation probability and fixation time. Communications Biology. 2:1–8.
Tkadlec J, Pavlogiannis A, Chatterjee K, Nowak MA. 2020. Limits on amplifiers of natural selection under death-birth updating. PLoS Computational Biology. 16:e1007494.
Tomlinson IP, Novelli M, Bodmer W. 1996. The mutation rate and cancer. Proceedings of the National Academy of Sciences. 93:14800–14803.
Vogelstein B, Papadopoulos N, Velculescu VE, Zhou S, Diaz Jr LA, Kinzler KW. 2013. Cancer genome landscapes. science. 339:1546–1558.
Voorhees B. 2013. Birth–death fixation probabilities for structured populations. Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences. 469:20120248.
Voorhees B, Murray A. 2013. Fixation probabilities for simple digraphs. Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences. 469:20120676.
Watson CJ, Papula A, Poon GY, Wong WH, Young AL, Druley TE, Fisher DS, Blundell JR. 2020. The evolutionary dynamics and fitness landscape of clonal hematopoiesis. Science. 367:1449–1454.
Watts DJ, Strogatz SH. 1998. Collective dynamics of small world networks. Nature. 393:440–442.
Waxman BM. 1988. Routing of multipoint connections. IEEE journal on selected areas in communications. 6:1617–1622.
Weinreich DM, Chao L. 2005. Rapid evolutionary escape by large populations from local fitness peaks is likely in nature. Evolution. 59:1175–1182.
Weissman DB, Desai MM, Fisher DS, Feldman MW. 2009. The

rate at which asexual populations cross fitness valleys. Theoretical Population Biology. 75:286–300.
Whitlock MC. 2003. Fixation probability and time in subdivided populations. Genetics. 164:767–779.
Whitlock MC, Barton N. 1997. The effective size of a subdivided population. Genetics. 146:427–441.
Wood LD, Parsons DW, Jones S, Lin J, Sjoblom T, Leary RJ, Shen D, Boca SM, Barber T, Ptak J *et al*. 2007. The genomic landscapes of human breast and colorectal cancers. Science. 318:1108–1113.
Wright S. 1943. Isolation by distance. Genetics. 28:114.
Wright S *et al*. 1932. The roles of mutation, inbreeding, crossbreeding, and selection in evolution. .
Yagoobi S, Sharma N, Traulsen A. 2023. Categorizing update mechanisms for graph-structured metapopulations. Journal of the Royal Society Interface. 20:20220769.
Yagoobi S, Traulsen A. 2021. Fixation probabilities in network structured meta-populations. Scientific Reports. 11:17979.