Competing evolutionary paths in growing populations with applications to multidrug resistance

Michael D. Nicholson[1] and Tibor Antal[2]

[1] School of Physics and Astronomy, University of Edinburgh
[2]School of Mathematics, University of Edinburgh

Abstract

Investigating the emergence of a particular cell type is a recurring theme in models of growing cellular populations. The evolution of resistance to therapy is a classic example. Common questions are: when does the cell type first occur, and via which sequence of steps is it most likely to emerge? For growing populations, these questions can be formulated in a general framework of branching processes spreading through a graph from a root to a target vertex. Cells have a particular fitness value on each vertex and can transition along edges at specific rates. Vertices represents cell states, say genotypes or physical locations, while possible transitions are acquiring a mutation or cell migration. We focus on the setting where cells at the root vertex have the highest fitness and transition rates are small. Simple formulas are derived for the time to reach the target vertex and for the probability that it is reached along a given path in the graph. We demonstrate our results on several scenarios relevant to the emergence of drug resistance, including: the orderings of resistance-conferring mutations in bacteria and the impact of imperfect drug penetration in cancer.

Contents

1 Introduction **2**

2 Results **3**
2.1 General framework . . . 3
2.2 Time until target vertex is populated . . . 5
2.3 Path distribution . . . 6
2.4 Extensions . . . 9
2.4.1 Cyclic graphs . . . 9
2.4.2 Distribution of fitness effects . . . 10
2.4.3 Alternative offspring distribution . . . 10

3 Discussion **11**
3.1 Applications . . . 11
3.1.1 Imperfect drug penetration: monotherapy . . . 11
3.1.2 Imperfect drug penetration: combination therapy . . . 13
3.1.3 Path to resistance in chronic myeloid leukemia . . . 16
3.1.4 Antibacterial multidrug resistance . . . 16
3.2 Concluding remarks . . . 17
3.2.1 Related work . . . 17
3.2.2 Summary . . . 18

4 Materials and Methods **19**
4.1 Population growth . . . 19

5 Supplementary Material **20**
5.1 Model recap . . . 20
5.2 Population growth . . . 21
5.3 Time until target vertex is populated: general case . . . 23
5.4 Path graph and acyclic graphs . . . 28
5.5 Distribution of the path to the target . . . 30
5.6 Cyclic graphs . . . 31
5.7 Simulation for alternative transition formulation . . . 34
5.8 Derivation of select formulas in Section 3.1.2 . . . 34

1 Introduction

The timing and manner in which a particular phenotype arises in a population is a central question of theoretical biology [8, 7, 75, 76, 27, 32, 80, 33, 20, 67, 39, 60, 58]. A typical scenario is to consider an initially monomorphic, wild type population, composed of cells that can acquire mutations, for example single site substitutions on the genome. The phenotype of interest comes to exist after a cell has accrued a specific set of mutations. The interpretations of this event are application dependent, but examples are the genesis of cancer instigated by mutations in a pair of tumour suppressor genes, or the emergence of multidrug resistance via alterations to the genes coding for the target proteins. Regardless of context, the questions of when, and how, the phenotype emerges are of significant interest.

It is commonly assumed that the population under consideration is a fixed size. However, with the aim of characterising disease progression, an increasing body of research has been developed to examine the evolutionary dynamics of a growing population. These studies have provided insights on a range of applications, including; cancer genetics [78, 12, 16, 31], metastasis formation [36, 35, 59, 23, 48, 63], drug resistance [14, 46, 48, 29], phylogenetics [77], and the impact of poor drug penetration [34, 38, 30, 57].

Here we continue in the same vein by considering a stochastically growing cellular population, where cells can transition in such a fashion so as to alter their, and their offsprings, reproductive capabilities. Such a transition might be due to the acquisition of a (epi)genetic alteration or migration into a new environment. As before, suppose we have a cellular state of interest, for example; a given genotype, a spatial location, or a combination of both. Will this state ever be reached? If it is, when is it reached? And by which sequence of intermediate states?

To make the discussion clear, let us consider an example application: the emergence of multidrug resistant bacteria. Suppose an infection begins with a single pathogenic bacterium which is sensitive to two antibiotic therapies, drug A and drug B. In the absence of either drug, the initial bacterium initiates a growing colony. Within the colony, when a cell replicates one of the daughter cells can acquire a mutation yielding resistance to either of the drugs. Here our questions are: how long does it take for multidrug resistance to emerge? En route, is resistance to drug A or drug B more likely to arise first? An ability to answer such questions is key to understanding pre-existing resistance, a common cause of therapy failure in various settings [13] The scenario is illustrated in Fig. 1a. There each vertex represents a cellular type, in this case its resistance profile. The edges represent cell transitions via mutation upon replication.

In this article we focus on the setting where the intermediate states have reduced reproductive ability (fitness) relative to the initial cells. This is primarily motivated by the commonly observed cost of resistance [2, 40, 28], whereby cellular populations that are resistant to a given drug grow more slowly than their sensitive counterparts. A second scenario where we expect a reduction in fitness is drug sensitive cells becoming exposed to toxins, which increases the cell's rate of dying (cytotoxic) or decreases the rate of replication (cytostatic). As in many biological applications the relevant transition rates are small (some

[1]mdnicholson5@gmail.com
[2]Tibor.Antal@ed.ac.uk

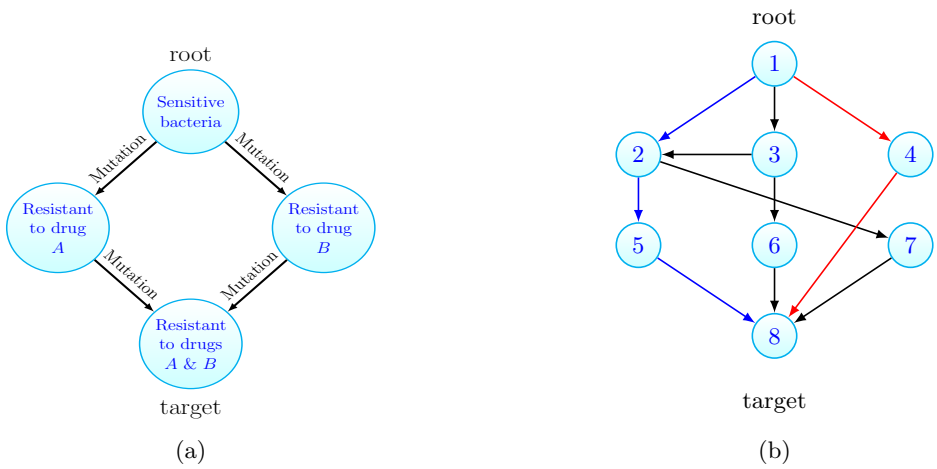


Figure 1: (a) Motivating example: consider a population of growing drug sensitive bacteria. Each vertex represents a cell type and cells can mutate to acquire resistance. Starting from one sensitive bacterium that can replicate, die or mutate, how long does it take to traverse the graph until multidrug resistance emerges? Is resistance to drug A then B more likely than the converse? (b) For the general case, the growing population is composed of different cell types that can be associated to vertices on a graph. Cells at vertex x can transition to vertex y if the edge (x, y) exists. Starting with cells at the root vertex, how long until the target is populated? Is the blue or red path more likely to initiate the target population? Here $N = 8$.

representative examples are: measurements for the point mutation rate per cell division of 10^{-9} in cancer [71] and 10^{-10} for bacteria [51], while the dissemination rate of pancreatic cancer cells from the primary tumour was estimated as 10^{-7} per cell division in [35]), we concentrate our efforts on the regime of low transition rates. Our main contribution is to provide simple, intuitive formulas that answer the questions posed in the small transition rate limit. These formulas show explicitly the contribution of the model parameters, e.g. transition rates and fitness reductions, which allow them to be probed for biological insight. This provides relationships which would be difficult to deduce from simulation alone.

We now move to detail the general framework we study, which will be seen to encompass models introduced in previous works as special cases [30, 48, 29, 18]. Our main results concerning when and how a particular cell type emerges are then presented, and we follow this by demonstrating our method on several applications.

2 Results

2.1 General framework

Our interest will always be in the emergence of a particular cell type in a growing population which is modelled as a specific form of a multitype branching process [4]. It is convenient to picture the population on a finite, simple, directed graph $G = (V, E)$, see Fig. 1b, where each vertex of the graph represents a cell type or state. The number of types in the system is denoted N and so we let $V = \{1, 2, \ldots, N\}$. Thus E is a subset of the set of ordered pairs $\{(i, j) : i, j \in V, i \neq j\}$. For any given cell type there is an associated vertex in V. Henceforth we will refer to cells of type x as cells residing at vertex x. We will be concerned with the timing, and fashion, that vertex N, which we refer to as the target, is populated, when we initiate with cells at vertex 1, which we denote the root. In the example of multidrug resistance, illustrated in Fig. 1a, the cells at the root are sensitive bacteria while the target population is resistant to both drugs.

Let (x) represent a cell at vertex x and $\varnothing$ symbolise a dead cell. Then, with all cells behaving indepen-

dently, our cell level dynamics can be graphically represented as:

$$(x) \to \begin{cases} (x),(x) & \text{at rate } \alpha(x) \\ \varnothing & \text{at rate } \beta(x) \\ (x),(y) & \text{at rate } \nu(x,y) \text{ if } (x,y) \in E. \end{cases}$$

That is: cells divide at rate $\alpha(x)$, die at rate $\beta(x)$, and transition to a cell at vertex y at rate $\nu(x,y)$ if the edge (x,y) exists. We will denote the fitness (growth rate) of cells at vertex x by $\lambda(x) = \alpha(x) - \beta(x)$, the difference of birth and death rates of vertex x cells. The parameters associated with the vertex 1 population feature prominently and so for convenience we let $\alpha = \alpha(1)$, $\beta = \beta(1)$ and $\lambda = \lambda(1)$. As mentioned above, in this article we focus on the setting where the vertex 1 population has the largest fitness, which is also positive. That is, we assume that $\lambda > 0$ and for $2 \leq x \leq N-1$, $\lambda(x) < \lambda$. We do not specify the fitness of the target population (cells at vertex N).

A common variant when modeling transitions (or mutations) is to specify that cells of type x divide at rate $\alpha'(x)$, and then with probability $\nu'(x,y)$ a transition occurs to vertex y. For a fixed value of parameters, this formulation of transitions is equivalent to that given above upon letting $\nu(x,y) = \alpha'(x)\nu'(x,y)$, $\alpha(x) = \alpha'(x)(1-\sum_{y:(x,y)\in E} \nu'(x,y))$. Also in the model as currently stated, transitioned cells have deterministically fixed new birth and death rates. However including a finite distribution of fitness effects, when a transitioned cell is assigned random birth and death rates, is actually covered by our model as given thus far. We discuss this point further in Section 2.4.

At a population level, the number of cells at vertex x at time t will be denoted $Z_x(t)$. We shall always assume that the system is initiated with only vertex 1 cells, at a quantity z, that is $Z_x(0) = z\delta_{x,1}$ with the Kronecker delta function $\delta_{x,y}$.

We have two primary questions. The first is, having initiated the system with z vertex 1 cells, how long does it take for the population at the target vertex to arise? That is we concern ourself with the distribution of the target hitting time, defined as

$$T = \inf\{t \geq 0 : \, Z_N(t) > 0\}. \tag{1}$$

Now assuming that the target vertex is populated by a founding cell, we ask from which path through the graph G did this founding cell come? This gives rise to a distribution over the set of paths (or walks) from the root vertex to the target vertex, which we aim to characterise. This second question is more precisely formulated in Section 2.3. Throughout we will assume that we wait for the first cell of vertex N to arise, however it may be that the population founded by the initial cell goes extinct. If instead one wishes to wait for the first 'successful' cell, that is the first cell of at vertex N to exist whose progeny survives, then all the results presented below hold so long as the mapping $\nu(x,N) \mapsto \nu(x,N)\lambda(N)/\alpha(N)$ for all target adjacent vertices x, is applied.

Our results are most clearly understood when G is acyclic, which amounts to excluding reverse transitions. Therefore in the following presentation this will be assumed. The setting when this assumption is relaxed will be discussed in Section 2.4.

Consider any path from the root to the target vertex. Intuitively, if the transition rates encountered along this path are larger, the target population will be seeded from this path quicker. Conversely, the smaller the growth rates along this path are, the slower the target will be reached. It will transpire that this intuition is indeed correct. The key quantities that show how the time to the target, and which path is taken, depend on these competing factors will be seen to be the path weights, which we now introduce.

Let the set of paths between the root and the target be be denoted $\mathcal{P}_{1,N}$. Any $p \in \mathcal{P}_{1,N}$, will be a sequence $p = (p_1, p_2, \ldots, p_l, p_{l+1})$, where each $1 \leq p_i \leq N$, l is the path length (the number of edges in p), all p_i are distinct (as presently G is acyclic), and $p_1 = 1$, $p_{l+1} = N$. Along the path p, let us call the difference between the growth rate at the root and the vertices encountered the fitness costs. Then we define the weight of path p as

$$w(p) = \nu(p_1, p_2) \prod_{i=2}^{l} \frac{\nu(p_i, p_{i+1})}{\lambda - \lambda(p_i)}, \tag{2}$$

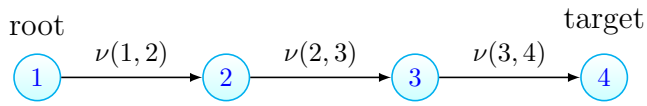


Figure 2: A path graph of length 3, with $N = 4$.

that is the product of the transition rates along the path divided by the fitness costs along the path. Throughout the empty product is set to 1. Further, we let the total weight of the target be

$$\phi_N = \sum_{p \in \mathcal{P}_{1,N}} w(p). \tag{3}$$

We will take the case of a path graph as a running example. A path graph, as illustrated in Fig. 2, is the scenario in which the only edges are $(i, i+1)$, for $1 \leq i \leq N-1$. There is only one path p between the root and the target and so in this case, $\phi_N = w(p) = \nu(1,2) \prod_{i=2}^{N-1} \frac{\nu(i,i+1)}{\lambda - \lambda(i)}$.

2.2 Time until target vertex is populated

How long do we wait until the target vertex is populated? We answer this question by characterising the distribution of T, defined in (1), when the target seeding transition rates (the transition rates associated with edges into the target vertex) are small. The key tool required is the long-time population numbers at the initial and intermediate vertices, which is discussed further in Section 4. Using this we are able to prove the following theorem regarding the target hitting time, whose proof can be found in the supplementary material (see Sections 5.3 and 5.4).

Theorem 1. *As the target seeding transition rates tend to 0, we have*

$$\mathbb{P}(T - \mu > t) \to \left(\frac{\lambda/\alpha}{1 + e^{\lambda t}} + \beta/\alpha \right)^z \tag{4}$$

where

$$\mu = \frac{1}{\lambda} \log \frac{\lambda^2}{\alpha \phi_N}.$$

Heuristically, the time until the target is populated is approximately μ plus some noise, where the distribution of the noise is given by the right hand side of (4). Note that this distribution depends only on parameters associated with vertex 1 cells, while all other parameters of the system are bundled into μ. For practical purposes the theorem yields the following approximation

$$\mathbb{P}(T > t) \approx \left(\frac{\lambda/\alpha}{1 + e^{\lambda t} \phi_N \alpha / \lambda^2} + \beta/\alpha \right)^z \tag{5}$$

This approximation will be valid if the target seeding transition rates are small. We do not provide estimates on the approximation error, however simulations (for example see Fig. 3b) demonstrate this approximation holds even for only moderately small transition rates.

Notice that it is possible that the target is never populated ($T = \infty$). In the supplementary material (Proposition 1) we show that if the vertex 1 population survives forever then the target hitting time is finite with probability one. Furthermore in many relevant cases, namely a large initial population, low death rate, or small transition rates leaving vertex 1, that the target is eventually populated is essentially equivalent to the vertex 1 population surviving. The probability that the vertex 1 population survives is obtained in the supplementary material (Eq. (25)), and this quantity thus gives an approximation for the probability that the target population will ever arise, which is

$$\mathbb{P}(T < \infty) \approx 1 - (\beta/\alpha)^z. \tag{6}$$

This also explains the β/α term in (5), which arises due to lineages stemming from the initial vertex 1 cells going extinct. The corresponding approximate distribution for the target hitting time, when we condition that the vertex 1 population survives is given in the supplementary material (Eq. (40)). Differentiating the conditional hitting time distribution yields the density which is compared with simulations in Fig. 3c.

Let us now discuss the time centring term μ, as given in Theorem 1. In the case of $\beta = 0$, $z = 1$, it can be easily seen (recall $\lambda = \alpha - \beta$) that the median of our approximate distribution (5), is exactly μ. More generally, if we let $t_{1/2}$ be the median time to hit the target, then if the vertex 1 population survives, we have

$$t_{1/2} \sim \mu - h(z). \tag{7}$$

as the target seeding transition rates tend to 0. The shift $h(z)$, whose precise form can be found in the supplementary material (Corollary (7.2)), has the behaviour

$$h(1) = 0, \quad h(z) \sim \frac{1}{\lambda} \log \frac{z\lambda}{\alpha}, \quad z \to \infty. \tag{8}$$

The shift exists as initiating the system with a larger number of cells at vertex 1, leads to the target being reached faster. In terms of notation $f \sim g$ means that $f/g \to 1$, when the limit under consideration is taken.

We now return to the running example of the path graph setting which was introduced at the end of Section 2.1. With $z = 1$, (7) yields that the median time for the target population to appear is

$$t_{1/2} \approx \mu = \frac{1}{\lambda} \log \frac{\lambda^2}{\alpha\nu(1,2)} + \frac{1}{\lambda} \sum_{i=2}^{N-1} \log \frac{\lambda - \lambda(i)}{\nu(i, i+1)}, \tag{9}$$

for small $\nu(N-1, N)$. The first summand on the right hand side comes from waiting for the first transition from the vertex 1 population, while the second is due transitions between the remaining vertices. The first summand is distinct as the vertex 1 growth is the main cause of stochasticity, as discussed in Section 4. The population growth and target hitting time is illustrated for the path graph case in Fig. 3 .

2.3 Path distribution

We now move to our second question: which path leads to the target population arising? Our naive expectation is that paths with larger weights will be more likely. This simple conjecture turns out to be true. To show this we first introduce some notation.

To any existing cell, say at vertex x, we may define the cell's vertex lineage. This tracks the vertices of cells in the ancestral lineage of the cell under consideration and is a sequence of the form $(1, \ldots, x)$. For example a cell with vertex lineage $(1, 3, 4)$ is at vertex 4, and descended from a vertex 3 cell, who in turn descended from a vertex 1 cell. Any vertex lineage is a path in G and we denote the number of cells with vertex lineage q at time t as $X(q, t)$. Note that while there may be multiple ancestors at the same vertex in a given cell's ancestral lineage, for example several generations of vertex 3 cells before the first vertex 4 cell, we do not record these in the vertex lineage (more precisely $q_i \neq q_{i+1}$).

Now for any root to target path $p \in \mathcal{P}_{1,N}$, that is a path of the form $p = (1, p_2, p_3, \ldots, N)$, the first time that p is traversed can be defined as

$$T(p) = \inf\{t \geq 0 : X(p, t) > 0\}. \tag{10}$$

Observe that path p populating the target first is equivalent to $T(p) = T$. The question of which path initiated the target only makes sense if the target is eventually populated. To ensure that this occurs we condition on a surviving vertex 1 population, and hence let

$$\mathbb{P}_1(\cdot) = \mathbb{P}(\cdot \,|\text{vertex 1 population survives}).$$

Recall that the vertex 1 population surviving is used as a proxy for T being finite (the target is eventually populated), as discussed in the paragraph following (5). We can now state the answer to our second question,

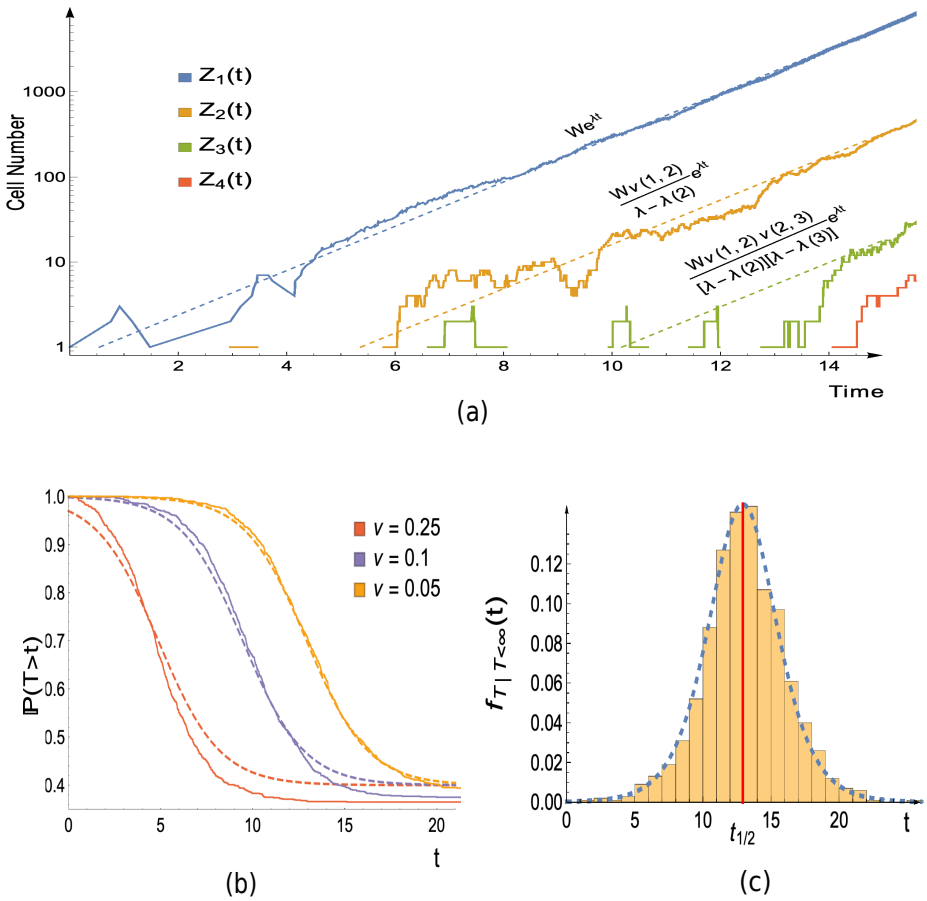
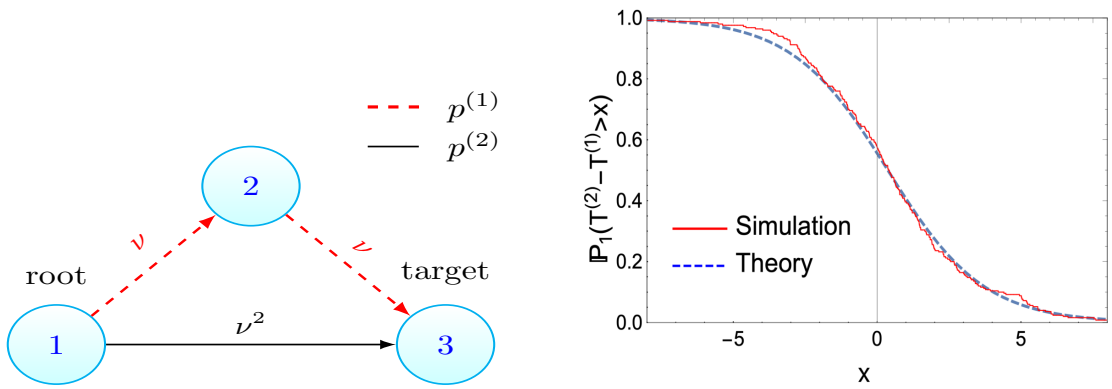


Figure 3: Simulation versus theory in the case of a path graph with 4 vertices, as in Fig. 2. (a) Joined lines show a single realisation of the population growth obtained via simulation. Dashed lines are the asymptotic growth of the populations, which follows the functions given immediately above the dashed lines. These functions are a consequence of Theorem 3, which can be found in Section 4. W is a random variable whose value will differ between simulations. Its value here is estimated by the realised value of $e^{-\lambda t_f} Z_1(t_f)$, with $t_f = 14.5$. (b) The distribution of the hitting time for vertex 4: Joined lines show the empirical distribution obtained from 1000 realisations, dashed lines display the theoretical approximation (5). The error in the approximation decreases as $\nu \to 0$, but still shows good agreement for moderate values of ν. (c) Probability density function for hitting time of vertex 4, conditioned that it is finite: bars display normalised histogram constructed from 1000 realisations, while dashed line shows approximation which follows from differentiating the conditional variant of (5) (precisely given in (41) of the supplementary material). The approximate median time for the type 4 population to arise, as given (9), is indicated by the red vertical line. Parameters: $\alpha = \alpha(2) = \alpha(3) = \alpha(4) = 1, \beta = 0.4, \beta(2) = \beta(3) = 1.3, \beta(4) = 0.6, z = 1$; for (a) and (c), $\nu(1,2) = \nu(2,3) = \nu(3,4) = 0.05$; for (b), legend displays value of $\nu = \nu(1,2) = \nu(2,3) = \nu(3,4)$ used.



(a) (b)

Figure 4: (a) Crossing valleys can be faster. Despite the red-dashed path containing a fitness valley (cells at vertex 2 have a reduced growth rate compared to the root population), the target population is more likely to arise via this path if $\lambda - \lambda(2) < 1$, as derived in (12). This is confirmed by the simulations displayed in (b), where $\lambda - \lambda(2) = 0.8 < 1$ and the Monte-Carlo estimate for $\mathbb{P}_1(T^{(2)} > T^{(1)})$ (the probability path 1 seeds the target first) is 0.58, while the analytic result yields 0.56. (b) Stochastic simulations of system in (a) vs theory (12). Parameters $\alpha = 0.9$, $\beta = 0.3$, $\lambda = 0.6$, $\alpha(2) = 0.2$, $\beta(2) = 0.4$, $\lambda(2) = -0.2$, $\nu = 0.1$, runs $= 250$.

which is a special case of Theorem 2 presented below. The probability that the target is populated via path p is simply the path weight of p, suitably normalised, that is

$$\mathbb{P}_1(T(p) = T) \approx \frac{w(p)}{\sum_{q \in \mathcal{P}_{1,N}} w(q)} = \frac{w(p)}{\phi_N} \tag{11}$$

for small target seeding transition rates. One may ask not only if a particular path populated the target, but whether the target was initiated from a given set of paths, for an example see Section 3.1.2. In this case the probability that the target is populated via a particular set of paths is given by summing the normalised path weights over each path in the set.

If the vertex 1 population avoids extinction then $T(p)$ will be finite for each root to target path p. Therefore instead of only asking which path populated the target first, one might be interested in whether a particular path is traversed significantly faster. For example, does multidrug resistance obtained via one of the paths in Fig. 1a occur t days before resistance from the other path. In order to also consider this case, for a root to target path p, we define

$$T(\neg p) = \min\{T(q) : q \in \mathcal{P}_{1,N} \backslash \{p\}\},$$

that is the first time the target is reached along any path except p.

We can now quantify the probability of reaching the target via path p more than t time units before any other path. To avoid discussing a technical assumption, which excludes pathological scenarios, the theorem is stated in approximate form. The more precise version can be found in the supplementary material (Theorem 8 in Section 5.5).

Theorem 2. *For small target seeding transition rates, and $t \in \mathbb{R}$,*

$$\mathbb{P}_1(T(\neg p) - T(p) > t) \approx \frac{w(p)}{w(p) + e^{\lambda t} \sum_{q \in \mathcal{P}_{1,N} \backslash \{p\}} w(q)}.$$

Note that letting $t = 0$ gives (11). In the running example of a path graph we cannot ask which path seeded the target vertex. However we briefly illustrate the usefulness of the above theorem by considering

the question of whether the target vertex is populated via a fitness valley (where by fitness valley we mean a path such that cells at the internal vertices have smaller growth rates than the root population) .

Let us consider the case depicted in Fig. 4a where two paths exist to the target, a direct path and an indirect path. Here $N = 3$, with paths $p^{(1)} = (1, 2, 3)$, $p^{(2)} = (1, 3)$ and $\nu = \nu(1, 2) = \nu(2, 3)$, $\nu(1, 3) = \nu^2$. Naively it may be expected that as the product of transition rates along both these paths are equal, and the indirect path ($p^{(1)}$) contains a vertex with a fitness cost, the direct path ($p^{(2)}$) is more likely. However, for small ν, Theorem 2 informs us that the target is populated via the indirect path t time units before the direct path, with probability

$$\mathbb{P}_1(T^{(2)} - T^{(1)} > t) \approx [1 + e^{\lambda t}(\lambda - \lambda(2))]^{-1} \tag{12}$$

where $T^{(i)} = T(p^{(i)})$. Thus the indirect path is more probable, that is $\mathbb{P}_1(T^{(1)} < T^{(2)}) > \mathbb{P}_1(T^{(2)} < T^{(1)})$, if $\lambda - 1 < \lambda(2)$.

Further applications of the results presented thus far will be given in Section 3.1. Before this, we discuss the case when G is cyclic, and some extensions to the initial model for which our results still hold.

2.4 Extensions

2.4.1 Cyclic graphs

In the above exposition we have assumed that G is acyclic, that is no path exists which contains cycles. We now briefly summarise how the results are altered if G is cyclic, with details given in the supplementary material. We will however maintain that the target is a sink vertex (has no outgoing edges) and the root is a source vertex (has no incoming edges). The condition on the root is perhaps unnatural, as we now allow back transitions between vertices except to the root, but is required for technical reasons which are discussed in the supplementary material (see the remarks following the proofs of Theorem 6 and Proposition 2). We aim to drop this condition in future work..

First we introduce the $N - 1 \times N - 1$ matrix A with entries

$$a_{ij} = \begin{cases} \lambda(i) & j = i \\ \nu(j, i) & \text{if } (j, i) \in E \\ 0 & \text{otherwise} \end{cases}$$

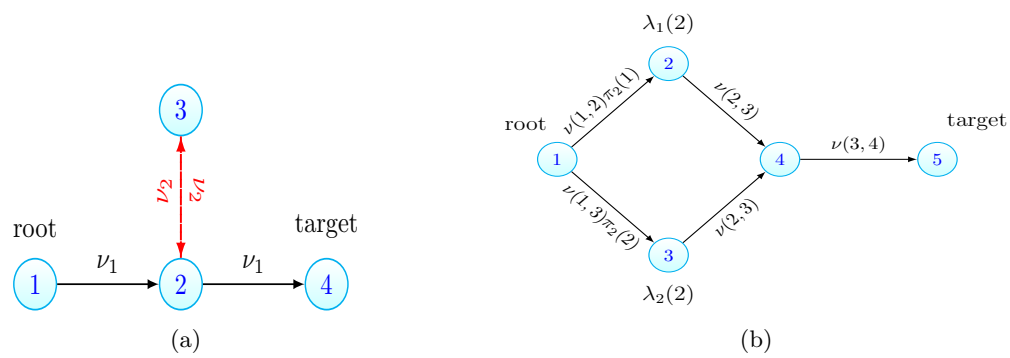


Figure 5: Example systems for which our extended results, Section 2.4, are useful. (a) Walks containing cycles can seed the target. The walk of length 4 that traverses the red edges with transition rate ν_2 can still be faster than the length 2 path. See (13). (b) Incorporating a distribution of fitnesses. Take the path graph from Fig. 2. Suppose that instead of a fixed $\lambda(2)$, after the first transition cells have fitness $\lambda_1(2)$ or $\lambda_2(2)$ with probabilities $\pi_2(1)$, $\pi_2(2)$, with $\lambda_i(2) < \lambda$. Then instead of the original path graph, we consider the enlarged graph shown here, where vertex 2 is replaced by two vertices with associated growth rates $\lambda_1(2)$, $\lambda_2(2)$. The transition rates to these new vertices are $\nu(1, 2)\pi_2(i)$, where $\nu(1, 2)$ is the transition rate between vertices 1 and 2 in the original setting. All results can then be applied to the new enlarged graph.

and we denote the largest real eigenvalue of A as λ^*.

In order that a modified version of Theorem 1 holds it is required that $\lambda^* = \lambda$ and λ^* is a simple eigenvalue. This condition is guaranteed for acyclic graphs, and will also be satisfied if the transition rates throughout the graph are small enough (see Lemma 5 in the supplementary material). Assuming this eigenvalue condition, Theorem 1 holds but with ϕ_N replaced by summing particular elements of the λ-associated eigenvector of A (see Theorem 7). Theorem 3, which concerns the population growth at the initial and intermediate vertices and is presented below in Section 4, is similarly modified under the same condition (see Theorem 6).

Turning to Theorem 2, we can now consider walks that contain cycles as opposed to only paths in the graph G. If we consider a particular set of root to target walks of a finite length, analogously to Theorem 2, we can ask whether the target is reached via one particular walk more than t units before any other walk in the specified set. A near identical result to Theorem 2 is obtained (see Proposition 2 in the supplementary material). However the path weights should now be replaced by the walk weights (defined analogously to the path weight; the product of the transition rates along the walk divided by the fitness costs along the walk). In particular for two given walks of finite length we can still approximate the probability the target population arises via one of the walks versus the other.

For example we can use this to explore when a walk, containing a cycle, will seed the target before a path. Consider the graph and edge parameters displayed in Fig. 5a. Let $p = (1, 2, 4)$ and $\omega = (1, 2, 3, 2, 4)$. Despite the extra length, by using the modified Theorem 2, for small ν_1 we see

$$\mathbb{P}_1(T(\omega) < T(p)) \approx \left(1 + \frac{(\lambda - \lambda(2))(\lambda - \lambda(3))}{\nu_2^2}\right)^{-1}. \tag{13}$$

Hence the target can be seeded via the walk, if ν_2 is large enough. On the other hand, by using the modified Theorem 2 it can be seen that when all the transition rates are small in the system, the target population will arise from a path and not a walk containing cycles (see Proposition 3) This further justifies our focus on the acyclic case.

2.4.2 Distribution of fitness effects

Let us now expand on the comments in Section 2.1 where we claimed that having a finite distribution of fitness effects is covered by our model. This may be of interest, for example, when a cell acquires drug resistance via a mutation, but the cost of resistance is stochastic. More generally, suppose we wish to consider cells transitioning from vertex x to vertex y but require that the new growth rate of cells to be $\lambda_i(y)$ with probability $\pi_y(i)$, where $\pi_y(i)$ is a finite distribution and every $\lambda_i(y) < \lambda$. Then we can simply replace the vertex y with new vertices associated to each of the $\lambda_i(y)$ and have cells transition from vertex x to the new vertices at rate $\nu(x, y)\pi_y(i)$. We may freely apply the results of Section 2 to the graph containing the new vertices and edges. See figure 5b for an example.

2.4.3 Alternative offspring distribution

We have focused on the setting where cells can divide, die or transition, however Theorem 2 would also hold for a slightly more general branching process. Keeping the transition process as it is, we can have a vertex x cell being replaced by j vertex x cells at a rate $\alpha(x)\sigma_{x,j}$ where for each x between 1 and $N - 1$, $(\sigma_{x,j})_{j\geq 0}$ is a probability distribution with finite second moment and mean $\bar{\sigma}_x$. We now have no need for $\beta(x)$ as cell death is contained in $\sigma_{x,0}$. This scenario may be relevant to the case of viral dynamics, where an infected cell can release multiple virions upon bursting [61]. In this setting if we redefine $\lambda(x) = \alpha(x)(\bar{\sigma}_x - 1)$ then the path distribution will be unchanged. This is due to the fact that the key result of [44] used in the proof of Theorem 3 holds then also. However Theorem 1 is no longer true in this setting. This is due to a lack of understanding of the large time behaviour of the root population (see the remarks following the proof of Theorem 6 in the supplementary material).

3 Discussion

Before discussing related work and summarising this study, we demonstrate how to apply the results of Section 2 on some applications. The approximate formulas given in Section 2 will be our key tools, and we now briefly remark on these.

The results of Section 2 hold when the final transition rates in each path tends to 0. Therefore in our applications all transition rates, in particular the mutation and migration rates discussed below, will be taken to be small and statements are to be interpreted as approximations that hold true in this limiting regime. We also note that in our model, when transitions represent migrations, they also occur at cell divisions, $(x) \to (x), (y)$, instead of at any time, as $(x) \to (y)$. This might appear strange from a biological point of view, but the former formulation for transitions has been chosen, as it simplifies the mathematical treatment. More importantly, we expect that for small transition rates these formulations lead to very similar target hitting times. Indeed, simulations support this claim, as presented in Fig. 9. In all applications considered we shall neglect the role of back transitions. This is for simplicity and is in keeping with the previous works that we compare with. The effect of (finitely many) back transitions could be included by using our results on cyclic graphs.

3.1 Applications

First we consider the impact of imperfect drug penetration on the emergence of resistance. Two recent publications have explored how resistance spreads in this setting, and have shown that poor penetration can accelerate resistance [30, 57]. We are able to recover and extend some of their findings in a rigorous fashion. Next, the ordering by which resistance-causing mutations accrue is investigated, firstly in the setting of cancer and then in bacterial infections. In the case of bacterial infections we examine how the risk of multidrug resistance depends on mutation rates, recovering simulation results in [29]. A particular aim of this section is to illustrate how to apply the results of Section 2, and so, for the readers' convenience, we have collected the key formulas in Table 1.

Table 1: Key formulas needed for Section 3.1. These approximations will be valid for small transition rates. The approximation for $t_{1/2}$ assumes a large initial number of cells at the root, $z \gg 1$. For small z see Corollary 7.2 in the supplementary material.

Description	Formula	Reference
Weight of path p	$w(p) = \nu(p_1, p_2) \prod_{i=2}^{l} \frac{\nu(p_i, p_{i+1})}{\lambda - \lambda(p_i)}$	Eq. (2)
Total weight of target	$\phi_N = \sum_{p \in \mathcal{P}_{1,N}} w(p)$	Eq. (3)
Distribution of target hitting time	$\mathbb{P}(T > t) \approx \left(\frac{\lambda/\alpha}{1+e^{\lambda t}\phi_N \alpha/\lambda^2} + \beta/\alpha \right)^z$	Eq. (5)
Median of target hitting time	$t_{1/2} \approx \frac{1}{\lambda} \log \frac{\lambda}{z\phi_N}$	Eq. (7)
Probability target populated via path p	$\mathbb{P}_1(T(p) = T) \approx \frac{w(p)}{\phi_N}$	Eq. (11)

3.1.1 Imperfect drug penetration: monotherapy

Our first two applications concern imperfect drug penetration and so the language introduced here will hold in the next section also. The first scenario we examine is exactly the same as that considered in [30].

We consider a pathogenic population, with cancer cells or bacteria as example populations, being treated with one or more drugs with imperfect penetration profiles. The imperfect penetration results in low drug

concentration in spatial locations, such that cells in these locations may still proliferate. We term the cumulation of these spatial regions the *sanctuary*. Sanctuaries have been observed in both bacterial infections [22, 74] and tumours [62]. We consider resistance to have arisen when cells with sufficiently many mutations such that all therapies are ineffective come to exist in areas where the drugs have penetrated. The possibility of resistance occurring prior to treatment is excluded and we consider only the case when resistance arises from the sanctuary. Further, we do not consider cells acquiring mutations via gene transfer.

Throughout this section and the next we shall be comparing paths composed of mutation and migration events. Growing cells in the sanctuary will be the vertex 1 cells from Section 2.1 and thus have birth and death rates α, β, with growth rate λ. The lack of exposure to drug motivates $\lambda > 0$. We shall assume all cells have birth rate α. Across all cell types, i.e. regardless of location or drug presence, the per cell mutation and migration rates will be denoted ν and m respectively. We fix the initial number of cells in the sanctuary to be z. We will often suppose that z is dependent on how penetrative a particular drug is (as this will affect the sanctuary size).

In the case of monotherapy, there are two paths from which resistance might arise. Firstly, upon replication a cell in the sanctuary may produce a mutant able to grow in the presence of the drug. Then a mutated cell in the sanctuary can migrate to the drug compartment. We term this the mutation-migration path. Analogously we have the migration-mutation path which involves a susceptible cell migrating from the sanctuary, from whose progeny a resistant mutant emerges. We suppose a resistance-conferring mutation carries a fitness cost which will be denoted s. Resistance costs may not always exist but are well documented and often assumed in viruses, bacteria and cancer [2, 40, 28]. Due to the fitness cost of resistance the death rate of mutated cells in the sanctuary is increased to $\beta + s$, while the death rate of migrated susceptible cells in the drug compartment is $\beta + d$, with $s, d > 0$.

The first quantity to determine is the probability of escaping resistance, and its dependence on the initial number of cells in the sanctuary. The approximation of (6) shows that the probability resistance never occurs decreases exponentially with the initial number of cells in the sanctuary as $\mathbb{P}(\text{resistance never occurs}) \approx (\beta/\alpha)^z$. For the rest of the discussion we will assume that resistance does occur. Further we suppose $z \gg 1$, so that we may use the approximate form for $h(z)$ given in (8). The case when z is small could be treated similarly using the exact expression for $h(z)$.

We now ask when, and via which path, will resistant cells come to exist in the drug compartment? From Section 2 we know the key quantities required to answer these questions are the path weights. For this system these are given by

$$w(\text{mutation-migration path}) = \frac{\nu m}{s}, \quad w(\text{migration-mutation path}) = \frac{\nu m}{d}. \tag{14}$$

If we now let $t_{1/2}$ be the median time resistance occurs, then using the path weights with (7), we immediately have

$$t_{1/2} \approx \frac{1}{\lambda} \log \frac{\lambda^2}{\alpha} - \frac{1}{\lambda} \log \left(\frac{\nu m}{s} + \frac{\nu m}{d} \right) + \frac{1}{\lambda} \log \frac{\alpha}{z\lambda} \tag{15}$$

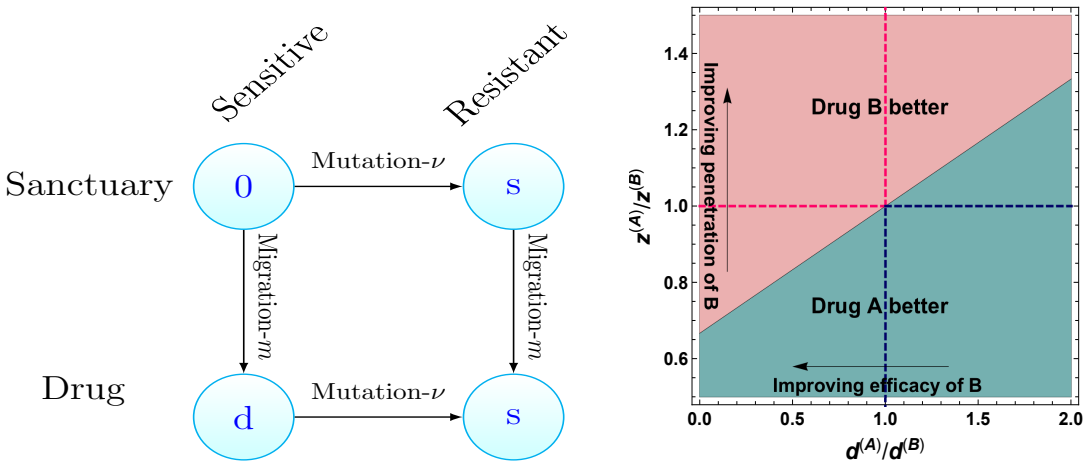
Of note is that the sanctuary size, mutation and migration rate all have equivalent impact. This simple result can now be explored under different biological scenarios.

For example, suppose we are currently using drug A with fitness cost $d^{(A)}$ and penetration profile leaving $z^{(A)}$ initial cells in the sanctuary. We consider instead using a second drug which results in sanctuary size $z^{(B)}$ and has fitness cost $d^{(B)}$. Let the median resistance times under the first and second drug be denoted $t_{1/2}^{(A)}$ and $t_{1/2}^{(B)}$ respectively. Then, using (15) and examining $t_{1/2}^{(B)} > t_{1/2}^{(A)}$, we find we should switch to drug B, if

$$\frac{z^{(A)}}{z^{(B)}} > \frac{1 + s/d^{(B)}}{1 + s/d^{(A)}}. \tag{16}$$

This inequality is illustrated in Fig. 6b.

Note as $d^{(B)} \to \infty$, implying drug B is far more effective, the condition tends to $z^{(A)}/z^{(B)} > (1 + s/d^{(A)})^{-1}$. This is due to the fact that when $d^{(B)}$ is large resistance will always arise via the mutation-migration path (which we show below), so this provides an upper bound to the gain that can be obtained



(a) (b)

Figure 6: (a) In the case of monotherapy, there are two paths by which drug resistance cells can arise in the region containing the drug. Horizontal edges represent changes in genotype (gaining resistance) whereas vertical edges represent changes in spatial location. The values of vertices are the fitness cost cells experience in that state and μ, m are the per cell mutation and migration rates. (b) Comparison of two drugs with differing efficacies and penetration profiles using (16). Arrows pertain to characteristics of drug B relative to drug A. Top left and bottom right quadrants illustrate region where drugs are trivially better, i.e. have superior penetration and efficacy. Parameters: $d^{(A)} = 10^{-2}$, $s = 5 \times 10^{-3}$.

from increasing the drug efficacy. The above is under the assumption resistance will occur but note the probability of no resistance would decrease exponentially with larger sanctuary size.

Having discussed the time resistance will occur we now ask by which of the two paths it will occur. Let $T^{\nu m}$ be the first time resistance arises via the mutation-migration path and $T^{m\nu}$ be defined analogously for the migration-mutation path. Then Theorem 2 gives us

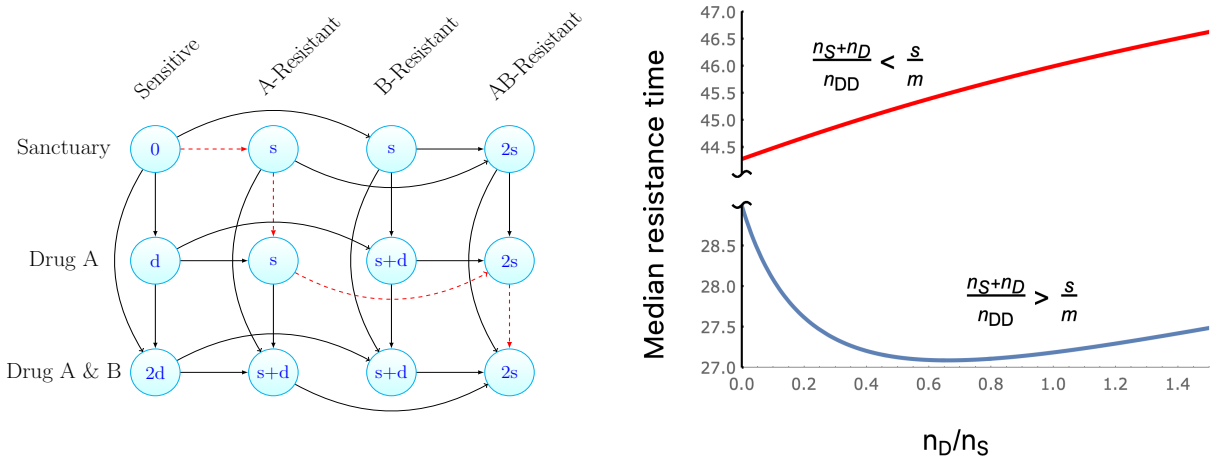
$$\mathbb{P}(T^{\nu m} < T^{m\nu}) \approx \frac{d}{s+d}.$$

Hence resistance arising via the mutation-migration path is more probable, that is the above probability is greater than 1/2, if $d > s$. As most drugs are designed to at least halt cell growth, while many resistant cell lines can still proliferate, we expect this condition to be satisfied for most drugs. Note that neither m nor ν appear in the above expression as the numerator of both path weights is equal, see (14), and hence cancels.

A similar question was asked in [30]. There they considered an approximate stochastic process and sought the parameter regime such that $T^{m\nu}$ stochastically dominated $T^{\nu m}$ under the approximate process. For the full process considered here, we can ask the same question but on the limit of the centred (in the sense of Theorem 1) resistance times. The resulting condition for stochastic dominance is again $d > s$. The difference in the condition given here ($d > s$) and that stated in equation (28) in [30] exists as there ν is defined as the probability of mutation per birth event. Accounting for this leads to the same condition.

3.1.2 Imperfect drug penetration: combination therapy

There are at least two reasons for extending our analysis to include two drugs, say drugs A and B, being used in tandem. Firstly, combination therapies are widely implemented in HIV and bacterial infections, and there is a growing appreciation for their use in cancer [17, 68, 14]. Secondly, interesting new phenomena emerges, especially when a region still exists with only one drug present. The work presented in this section is inspired by the model and questions of [57].



(a) (b)

Figure 7: (a) Illustration of the paths to multidrug resistant cells arising in the region containing both drugs A and B, as in Fig. 6a. Note a spatial location exists containing only drug A. The red, dashed path is an example of stepwise evolution, where multidrug resistant cells come to exist in the sanctuary via the region containing only a single drug. The vertex labels display the fitness cost (increase to the cellular death rate) of cells in that state. (b) Exploration of the median time to resistance as the size of the drug A region increases relative to the sanctuary size. Here we plot $t_{1/2}$, as given in Table 1, for the system displayed in (a). We observe that, for a subset of the parameter space, the existence of a single drug region can decrease the time to resistance, despite diminishing the sanctuary size. If both drugs penetrate the majority of the system and the drugs have a greater effect than the cost of resistance, this subset of parameter space is identified by the inequality (17). Within this parameter subset, resistance is fastest at a particular size of single drug compartment, given by (18). Parameters; (top) $\nu = 10^{-6}$, $m = 0.05$, $d = 0.9$, $s = 10^{-3}$, $\lambda = 0.4, \alpha = 0.3$, $n_{\text{Tot}} = 10^7$, $n_{\text{DD}} = 0.9 n_{\text{Tot}}$, $\gamma = 10^{-3}$; (bottom) all same except $s = 10^{-2}$.

We will only consider the case where the penetration profile for drug B is a subset of drug A, and thus a single drug region exists only for drug A. This allows us to clearly demonstrate the impact of unequal penetration profiles. In this setting there are 12 possible states that a cell may inhabit, defined by which of the drugs the cell is exposed to and the resistance profile it has obtained. These states are illustrated in Fig. 7a. We further assume that the mutation rate bringing resistance to either drug is ν. This is for simplicity only, and dealing with differing mutation rates is straightforward. Mutations that confer resistance will again have a fitness cost s, and sensitive cells coming into contact with drugs invokes a fitness cost d. These fitness costs increase the cellular death rate additively. The increase to the death rate for a cell in any given state is displayed as the vertex labels in Fig.7a. Drug interactions, mutation specific fitness costs, and epistasis are neglected but may easily be included.

To investigate the effect of the single drug region, we introduce further notation. Let n_{Tot} be the number of cells that may reside in our system of interest (e.g. the entire patient or tumour) in the absence of drug. Further, let the regions where both drugs act, the single drug (drug A) acts and the sanctuary have their respective cell capacities denoted by n_{DD}, n_{D}, and n_{S}. We assume that all these capacities are large, and hence we may still investigate the system with our branching process model. These capacities enter the dynamics of the process in the following manner. When a cell leaves the region it is in, at rate m, it now may migrate to one of two regions. Of these two regions, we specify that it migrates to a particular region at a frequency proportional to the regions capacity. For example a cell in the sanctuary migrates to the double drug region at rate $m n_{\text{DD}}/(n_{\text{D}} + n_{\text{DD}})$. Resistance mutations and the presence of the drug has the same additive effect as in Section 3.1.1.

Observe that there are 18 possible paths by which sensitive cells in the sanctuary can eventually produce

fully drug resistant cells in the double drug compartment. To see this it is convenient to separate the paths into those that pass through the single drug compartment, which following [57] we term stepwise evolution, and those that involve migration from the sanctuary directly to the double drug compartment, which we collectively call direct evolution. For the direct evolution paths there are three transitions needed, mutations yielding resistance to drug A and B and then the migration to the double drug region. From the number of permutations of these events, this yields 6 direct evolution paths. In the stepwise evolution case four transitions are needed, both mutations and then the two requisite migration events. Notice that the migration from the sanctuary to the single drug region must precede the migration from the single drug region to the double drug region. This leads to 12 stepwise evolution paths. An exemplar stepwise evolution path is a mutation to drug A, followed by migration to the single drug region, then a mutation to drug B, and finally a migration to the double drug region. The edges of this path are red and dashed in Fig. 7a.

Suppose, momentarily, that the penetration profiles of drugs A and B exactly match (but are still imperfect). In this case, resistance in the double drug region can be obtained only via direct evolution. If we now increase the penetration of drug A only, 12 new possible paths to resistance open up (the stepwise evolution paths of the previous paragraph), accelerating the time until resistance. However, the increased penetration also leads to a reduction in the sanctuary size, which should delay resistance. Hence a trade-off may exist with respect to the penetration profile of drug A. To study this we fix $n_{\mathrm{Tot}} = n_{\mathrm{S}} + n_{\mathrm{D}} + n_{\mathrm{DD}}$ and specify that the initial number of cells in the sanctuary is related to its size via $z = \gamma n_{\mathrm{S}}$ with $0 < \gamma \ll 1$. Further let $t_{1/2}(n_{\mathrm{D}}/n_{\mathrm{S}})$ be the median time to resistance with fixed $n_{\mathrm{D}}/n_{\mathrm{S}}$. Using the approximation provided by (7) we can explore the behaviour of $t_{1/2}(n_{\mathrm{D}}/n_{\mathrm{S}})$. As can be seen in Fig. 7b it is possible, in certain regimes of the parameter space, that the existence of the single drug compartment accelerates resistance (by reducing $t_{1/2}$). We now seek the conditions under which this is the case. Brief derivations of the presented formulas below (17, 18, 19) are given in the supplementary material (Section 5.8) ,

As mentioned previously, it is reasonable to expect the efficacy of either drug is greater than the cost to resistance. Also we may hope that both drugs penetrate the majority of the target system. This motivates taking $s \ll d$ and $(n_{\mathrm{S}} + n_{\mathrm{D}}) \ll n_{\mathrm{Tot}}$. Under this limit, by examining the sign of $\frac{d}{d(n_{\mathrm{D}}/n_{\mathrm{S}})} t_{1/2}(0)$, we find a simple condition for when increasing the single drug compartment accelerates resistance, namely

$$\frac{n_{\mathrm{S}} + n_{\mathrm{D}}}{n_{\mathrm{Tot}}} > \frac{s}{m}. \tag{17}$$

However, when (17) holds, as $n_{\mathrm{D}}/n_{\mathrm{S}}$ increases, eventually a minimal resistance time is achieved. This represents a worse cast scenario yielding fastest possible resistance. For $s \ll d$, this minimal time occurs when

$$\frac{n_{\mathrm{D}}}{n_{\mathrm{S}}} \approx 1 - \frac{2 s n_{\mathrm{Tot}}}{m(n_{\mathrm{S}} + n_{\mathrm{D}})}. \tag{18}$$

We now examine whether resistance arises via stepwise evolution or direct evolution. Let T^{SE} be the first time resistance occurs via stepwise evolution, and T^{DE} defined analogously for direct evolution. Then using Theorem 2, again with $s \ll d$, we find

$$\mathbb{P}(T^{\mathrm{SE}} < T^{\mathrm{DE}}) \approx \left[1 + \frac{s(n_{\mathrm{DD}} + n_{\mathrm{S}})}{m n_{\mathrm{D}}}\right]^{-1}. \tag{19}$$

Therefore the larger the proportional size of the single drug region, the more likely resistance will arise via stepwise evolution. A qualitatively similar result was derived in [57] (e-page 2878) by comparing the most likely stepwise evolution path with the most likely directed evolution path.

Furthermore, in [57] a minimal resistance time, in the sense of (18), was observed via simulations for the mutation$_A$-migration-mutation$_B$-migration path (the red and dashed path in Fig. 7a). Using Theorem 2 we can show that this is the most probable (in the sense of the mode of the path distribution, see (11)) stepwise evolution path whenever $d > s$. Further, by differentiating the median time along this path, we see a minimal time occurs at $n_{\mathrm{D}} \approx n_{\mathrm{S}}$, which recovers the simulation result in [57] (see their Fig. 3). That this path is the most probable stepwise evolution path also gives insight into the recurring s/m term in (17) and

(19). Relative to direct evolution this path has an extra migration step, so higher migration should increase the probability of seeding the target via this path. However fitness costs along this paths are incurred only via the cost of resistance, and a smaller s diminishes these fitness costs. While the above is motivated by and compared with the study [57], in contrast to that work, we do not consider a size-dependent branching process (the population was constrained by a large carrying capacity in [57]). Despite this we are able to derive similar qualitative, and indeed quantitative (e.g. the recovery of the simulation result discussed above), results. These features may be general properties of such models or are perhaps due to the size of the carrying capacity in [57].

3.1.3 Path to resistance in chronic myeloid leukemia

Let us discuss the example application of [48], concerning the mechanism by which resistance to the targeted therapy imatinib occurs in chronic myeloid leukemia (CML). The two main resistance mechanisms are via gene amplification of the BCR-ABL fusion gene, which drives the CML, or a point mutation resulting in a modification to the target protein of imatinib. Approximately 100 point mutations have been identified conferring resistance, and a rough estimate for the probability of each of these occurring during a cell division is 10^{-7} [50]. Thus the probability of a point mutation conferring resistance is approximately 10^{-5} per division. The analogous quantity for resistance-causing gene amplifications, in a similar system, was found to be approximately 10^{-4} [70]. With the birth rate of leukemic stem cells as $\alpha = 0.008$ day^{-1} [55] we have the rate of point mutations as $\nu_{pm} = 8 \times 10^{-8}$ day^{-1}, while the rate of resistance due to gene amplifications is $\nu_{ga} = 8 \times 10^{-7}$ day^{-1}. Therefore, using this information alone, the probability of resistance arising from a point mutation first is estimated to be 1/11. However, in the majority of cases, the primary mechanism of resistance is found to be point mutations [9].

Within our framework there are two possible explanations we can explore, both related to the reproductive capabilities of resistant cells. The first is simply that more lineages of amplified cells go extinct. The paragraph directly following (1) indicates how to treat this case. Indeed, if ρ_{pm} is the average survival probability for a population initiated by a single cell with a point mutation and ρ_{ga} the analogous quantity for gene amplification, then we find resistance arising first from a point mutation is more probable if $\rho_{pm} > 10\rho_{ga}$. This is possible as point mutations may be deleterious or advantageous [50] whereas gene amplifications appear deleterious [69].

The second explanation, that suggested in [48], is that several gene amplifications are required to attain resistance. Suppose a gene amplification increases the death rate of cells by s per day. Then if two gene amplifications are required, ignoring the survival probability from before, the probability that resistance arises via point mutation is now

$$\frac{\nu_{pm}}{\nu_{pm} + \nu_{ga}^2/s} = (1 + 8 \times 10^{-6}/s)^{-1}.$$

Resistance via points mutations is now more likely if $s > 10^{-6}$. Note that Fig. 4a serves as a schematic for this scenario, reinterpreting the fitness valley path as resistance arising via gene amplifications and the direct path relating to resistance via point mutations. Further explanations for the primacy of point mutation mediated resistance exist and the assumptions above may not hold. But when they do hold, our results offer a clear framework to investigate such questions.

3.1.4 Antibacterial multidrug resistance

A timely problem is that posed by the emergence of multidrug resistant bacteria. We now explore some specific examples of our general framework that pertain to this issue. In particular we consider the accumulation of resistance-conferring mutations as a bacterial colony grows in the absence of drugs. As before we assume resistance incurs a fitness cost and thus the wild-type (unmutated cells) has the highest growth rate. Using Theorem 2 and measured fitness costs and mutation rates we can predict the most likely path to multidrug resistance.

We consider the acquisition of resistance to both rifampicin and streptomycin in *P. aeruginosa*. Recently the fitness costs associated with resistance to rifampicin [37] and streptomycin [73] have been reported. Fitness costs were determined by the maximum growth rate of the bacteria undergoing exponential growth on nutrient-rich plates. From [73] the ratio of streptomycin resistant bacteria to the wild-type's growth rate was 0.71 (here the growth rate of the wild-type is estimated from Fig. 2.a in [73]). The same value for resistance to rifampicin, in [37], was 0.88. Let T^{RS} be the time when multidrug resistance emerges assuming rifampicin resistance is acquired first, and T^{SR} defined analogously for streptomycin resistance first. Then from Theorem 2, we have that

$$\mathbb{P}(T^{RS} < T^{SR}) \approx \frac{1-0.71}{(1-0.88)+(1-0.71)} = 0.7.$$

Hence, under this model, we would predict that rifampicin resistance emerges first en route to multidrug resistance. We note that above we have taken the average fitness costs associated with the differing antibiotics. Fluctuations have not been taken into account.

Instead of considering the the acquisition of resistance to different drugs within a bacterial strain, we can also examine the impact of differing mutation rates across strains. In [29] the heightened occurrence of multidrug resistance in lineage 2 vs lineage 4 from *M. tuberculosis* was investigated. Considering in particular resistance to rifampicin and isoniazid, via simulations they deduced that the heightened occurrence of resistance was due to elevated mutation rates in lineage 2 strains. The model simulated can be considered a discrete time counterpart to ours in the case of two paths of length 2 (see Fig. 1a). One of the main quantities of interest was the relative probability of resistance, when the only difference between the strains was the mutation rates. For a particular strain, initiated with one wild-type cell with wild-type birth and growth rates, α, λ, suppose we have mutation rates $\nu_{\mathrm{R}}, \nu_{\mathrm{I}}$ to rifampicin and isoniazid, each incurring the same fitness cost s. Then from Theorem 1, we have

$$\mathbb{P}(T \le t) \approx 1 - \left(\frac{\lambda/\alpha}{1+e^{\lambda(t-\mu)}} + \beta/\alpha \right), \quad \mu = \frac{1}{\lambda}\log\frac{\lambda^2}{\alpha} + \frac{1}{\lambda}\log\frac{s}{2\nu_{\mathrm{R}}\nu_{\mathrm{I}}}. \tag{20}$$

For lineage-i ($i = 2, 4$) let $\nu_{\mathrm{R}}^{(i)}$ and $\nu_{\mathrm{I}}^{(i)}$ be the mutation rates to rifampicin and isoniazid. As in [29], keeping all other parameters equal between the lineages we have the relative probability of resistance

$$\frac{\mathbb{P}(T^{(2)} \le t)}{\mathbb{P}(T^{(4)} \le t)} \approx \frac{\nu_{\mathrm{R}}^{(2)}\nu_{\mathrm{I}}^{(2)}}{\nu_{\mathrm{R}}^{(4)}\nu_{\mathrm{I}}^{(4)}} \frac{(1+2\nu_{\mathrm{R}}^{(4)}\nu_{\mathrm{I}}^{(4)}e^{\lambda t}\alpha/\lambda^2 s)}{(1+2\nu_{\mathrm{R}}^{(2)}\nu_{\mathrm{I}}^{(2)}e^{\lambda t}\alpha/\lambda^2 s)}.$$

For times such that $\nu_{\mathrm{R}}^{(i)}\nu_{\mathrm{I}}^{(i)}e^{\lambda t} \ll 1$, $i = 2, 4$, we see the relative probability is simply $\nu_{\mathrm{R}}^{(2)}\nu_{\mathrm{I}}^{(2)}/\nu_{\mathrm{R}}^{(4)}\nu_{\mathrm{I}}^{(4)}$. Taking the mutation rates measured in [29], this gives the relative probability of 22.06. This agrees with the simulation results in [29] who reported an approximately 22 fold increase. The same problem was treated in [18] but for multidrug resistance to two drugs in a fixed order (i.e. resistance to rifampicin followed by isoniazid). Staying in the regime $\nu_{\mathrm{R}}^{(i)}\nu_{\mathrm{I}}^{(i)}e^{\lambda t} \ll 1$, if in (20), we replace the 2 by 1 (as only one path to resistance), and alter both λ^2/α by λ, and $e^{\lambda t}$ with M, our expression is the same to leading order as that given in [18] (equation 2) for the probability of resistance in a population of size M. This difference between fixed population size and fixed time results is the same as given in [16, 26].

3.2 Concluding remarks

3.2.1 Related work

Many previous studies have considered the probability of particular cell type emerging in a growing population, typically by a fixed time after the process starts or when the total population reaches a given size. The majority deal with the target population being one or two transitions away from the initial population (the root vertex in this paper) [52, 43, 14, 18, 36, 16, 3, 21]. In particular we single out the pioneering work of Luria and Delbrück [52], which demonstrated the spontaneous nature of mutations by combining

an appropriate mathematical model with bacterial experiments on phage resistance. The original model of [52], and its various incarnations [49, 5] have been extensively studied [54, 56, 53, 3, 47, 45, 16]. Its fully stochastic formulation, which is identical to our model for two vertices, is due to Bartlett [5]. This full model admits an explicit solution [3], and the model's asymptotic behavior has been recently explored [16, 47]. One of the simplest quantities of interest is the probability of no resistant cells at a fixed time, often called p_0. It is a closely related quantity to our target hitting time described in (5), that is no resistant cell arises by a fixed time. Understanding p_0 provides a method to infer the per cell-rate at which resistance-conferring mutations are acquired, often termed the p_0-method [64].

Some notable exceptions which consider the target population being greater than two transitions away are [11, 48, 6, 26, 25, 57, 15] (ref. [57] is discussed above in Section 3.1.2). In [48], the same model as that presented here is numerically explored when all vertices have the same fitness ($\alpha(x) = \alpha$, $\beta(x) = \beta$ for all vertices x), and the implications on multidrug therapy failure is emphasised. An efficient numerical method to compute the distribution of target hitting time and path probabilities via the iteration of generating functions is given in [6] with a focus on cancer initiation. Both [11, 26] are motivated by the accumulation of driver mutations in cancer, and so each transition leads to a fitness increase. For [11] the mean time of the kth driver mutation derived and compared to genetic data. The distribution of the time until the kth driver is sought in [26], whose methods are the closest to those used in this paper. There the authors employed an appealing approximation of the model studied in this paper in the path graph case (the approximation is that the seeding rate into vertex $x+1$ from vertex x, which is exactly $\nu(x, x+1)Z_x(t)$, is approximated by $\nu(x, x+1)W_x^* e^{\lambda(x)t}$, for some sensibly chosen random variables W_x^*). Notably, again for small transition rates, the functional form of the distribution of the target hitting time is the same as that given in Theorem 1, however with a different median. The altered median derived in [26] demonstrates that when transitions lead to a fitness increase, the growth rates associated with the intermediate vertex populations have a far greater effect on the target hitting time when compared to the setting considered in this paper. The same type of approximate model was further used in [25] to investigate the target hitting time when transitions bring a fitness increase which is a random variable. Target hitting times on a path graph for a branching process with general offspring distribution, were also discussed in the very recent paper [15], but their main explicit results exclude cell death, and hold when successive transition rates along the path are assumed to be getting (infinitely) stronger.

The model and questions considered in this paper also arise frequently in the evolutionary emergence or evolutionary escape literature [41, 42, 66, 65, 1], with the notable distinction that in the evolutionary escape setting the root and intermediate vertex populations are destined to go extinct ($\lambda(x) < 0$ for $1 \leq x \leq N-1$). This scenario is of interest when, for example, a homogeneous, pathogenic population (all residing at the root vertex in the language used in the present study) is treated with a therapy and must acquire sufficiently many mutations away from the root so as to become resistant (populate the target vertex). As in the setup of the present paper, the target hitting time is strongly controlled by the growth of the root population, $Z_1(t)$, which has positive growth rate, $\lambda > 0$, the target hitting times in the evolutionary escape setting are distinct from those given in Section 2.2. However for the escape probability (which is the probability of reaching the target), if there are multiple paths from the root to the target, the contribution of each path to the escape probability (termed the path value in [41, 42]) has an expression strikingly similar to the path weights discussed here for small transition rates (compare (2) with Eqs 6a-c in [42]). We might conjecture that for a specific path, the path value, as given in [42], is the unnormalised probability of reaching the target via the specified path, as we have demonstrated is the case with the path weights in Theorem 2 (this is implied in [42] Sec 2.5). Further connections surrounding the path distribution in these differing regimes is an interesting avenue for future work.

3.2.2 Summary

In this article we have considered a continuous time birth-death process with transitions. The time until a population of a particular target state arises was investigated. This target state is accessible via one or more transitions from the initial population. Which sequence of transitions leads to the target population arising was explored. This gave rise to a probability distribution over the paths composed of transitions

between differing vertices. Motivated by applications, the setting when the initial population was most fit was focused on. The important factor in determining whether the target was reached via a specific path was seen to be the weight of the path. The weight of the path is composed of the transition rates and fitness costs associated with the intermediate vertices along the considered path.

Our primary contributions are the simple, explicit formulas for the target hitting time and the distribution of paths to the target, which are valid for small transition rates. When compared with previously given expressions for hitting times in branching processes, due to the reduction in complexity, we believe our results are easily interpretable and widely applicable. Further, to the best of our knowledge, the path distribution presented is the first analytic characterisation of evolutionary paths through a multitype branching process. The utility of our formulas was demonstrated on a variety of scenarios pertaining to the emergence of drug resistance. The biological relations revealed would have been difficult to deduce by a simulation based approach.

A shortcoming of our model is that the potential for unlimited growth and the lack of interactions between the cells is unrealistic. However in systems with large carrying capacities and abundant resources, our framework provides a convenient approximation and has been widely used. One possible route for assessing the importance of these neglected factors is to embed our questions in a similar setup to that used in the recent work [10], which included competition between cell types and a carrying capacity which the population could fluctuate around. The initial conditions in [10] were such that the model does not represent a growing population, so we do not compare their findings with the results of Section 2. Embedding our questions in a variant of the model of [10], such that the population is growing, might offer insight on the effect of these neglected aspects. Whenever these factors are deemed unimportant, or as a starting point for more complex models, this article provides quick and accessible results which may be used to guide and develop biological insight.

4 Materials and Methods

4.1 Population growth

The main ingredient in proving Theorem 1, and consequently Theorem 2, is the long-time behaviour of the population of the initial and intermediate vertices. In this section we discuss this asymptotic behaviour.

Firstly we extend the definition of the total path weight given in Section 2.1. Let the set of paths between the root and vertex x be denoted $\mathcal{P}_{1,x}$. Then using the definition of path weight (2), we let the total weight for vertex x be $\phi_x = \sum_{p\in\mathcal{P}_{1,x}} w(p)$ for $2 \leq x \leq N$. Further, let us denote the ratio of the total weight to fitness cost for vertex x as

$$\Phi_x = \frac{\phi_x}{\lambda - \lambda(x)}$$

for $2 \leq x \leq N-1$. To allow us to conveniently describe the growth at vertex 1 also, we set $\Phi_1 = 1$

The target population can only be founded by transitions from cells residing at the neighbouring vertices (vertices connected to the target by an edge). Therefore, understanding the population growth of cells at these vertices is needed to discuss the timing and manner in which cells at the target arise. At large times, this understanding is provided by the following theorem, concerning the population at the initial and intermediate vertices (it is useful to recall that $Z_x(t)$ is the number of cells at vertex x at time t and that we initiate with z cells at vertex 1). It follows from a more general result due to Janson [44], which we have tailored to our particular setting.

Theorem 3. *With probability one*

$$\lim_{t\to\infty} e^{-\lambda t}(Z_x(t))_{x=1}^{N-1} = W(\Phi_x)_{x=1}^{N-1}.$$

Here W is distributed as the sum of K independent exponential random variables with parameter λ/α, where K is binomial with z trials and success probability λ/α.

A short proof, demonstrating how Theorem 3 can be deduced from [44] is given in Section 5.2 of the supplementary material (see Theorem 6). The distribution of W may be called a binomial-Erlang mixture, being an Erlang distribution with a binomially distributed shape parameter. For the running example of a path graph, we see that asymptotically $Z_{i+1}(t)$ is related to $Z_i(t)$ by a multiplicative factor of $\frac{\nu(i,i+1)}{\lambda-\lambda(i+1)}$. This is demonstrated in Fig. 3a. We note that in the case of $N = 3$, Theorem 3 was previously given in [16] (Theorem 3.2).

Using Theorem 3 and that the immigration rate into the target vertex from any target neighbouring vertex x is $Z_x(t)\nu(x, N)$ we are able to prove the results of Section 2. Full proofs are provided in the supplementary material.

Acknowledgments

We thank David Cheek, Stefano Avanzini and the Edinburgh physics population dynamics group for helpful discussions and feedback. MDN acknowledges support from EPSRC via a studentship.

5 Supplementary Material

The outline of this supplement is as follows. First we recap the model, introducing notation that will be needed. Next we prove versions of Theorem 3 and Theorem 1, which do not require G to be acyclic. After this we specialise to the acyclic case. Staying with the acyclic case we prove Theorem 2. Finally, how the path distribution is altered when cycles are permitted is considered.

5.1 Model recap

We recap our framework. Our model is a specific form of multitype branching process [4] in which each population will evolve according to a Markovian linear birth-death process with transitions. We will assume each population is comprised of cells. As a conceptual framework we associate the multitype branching process with a simple, finite, rooted, directed, graph $G = (V, E)$ containing N vertices ($N = |V|$). Labels of the vertices take values in $\{1, \ldots, N\}$ and E is a subset of the set of ordered pairs $\{(i, j) : i, j \in V, i \neq j\}$. We shall often refer to vertex 1 as the root and vertex N as the target. Each vertex is reachable from the root and the target is reachable from any other vertex. We further assume the root is a source vertex and the target vertex is a sink. Letting the set of incoming neighbours for vertex i be $\mathcal{N}^-(i) = \{k \in V : (k, i) \in E\}$ and the set of outgoing neighbours $\mathcal{N}^+(i) = \{k \in V : (i, k) \in E\}$, the previous assumption is $\mathcal{N}^-(1) = \mathcal{N}^+(N) = \emptyset$. Each type in the branching process is uniquely mapped to a vertex, and so the number of types is N. Hence any cell may be described by its type or the vertex associated to that type.

Take any cell at vertex x. We assume this cell divides (replaced by two identical copies) at rate $\alpha(x)$, dies (removed from system) at rate $\beta(x)$ and transitions to a cell at vertex y (replaced by an identical copy and a cell at vertex y) at rate $\nu(x, y)$ if edge (x, y) is contained in E. The growth rate of cells at vertex x will be denoted $\lambda(x) = \alpha(x) - \beta(x)$. The parameters associated with the vertex 1 population feature prominently and so for convenience we let $\alpha = \alpha(1)$, $\beta = \beta(1)$ and $\lambda = \lambda(1)$. All cells are independent. We will focus on the setting where the vertex 1 population is the most fit and has positive growth rate. Therefore, we henceforth assume that $\lambda > 0$ and for $2 \leq x \leq N - 1$, $\lambda(x) < \lambda$. We do not specify the relative fitness of the vertex N population. The cell level dynamics may summarised as

$$(x) \to \begin{cases} (x), (x) & \text{rate } \alpha(x) \\ \varnothing & \text{rate } \beta(x) \\ (x), (y) & \text{rate } \nu(x, y) \text{ if } (x, y) \in E \end{cases}$$

where (x) represents a cell at vertex x and $\varnothing$ symbolises a dead cell. At a population level, the number of cells at vertex x at time t will be denoted $Z_x(t)$. We shall always assume $Z_x(0) = z\delta_{x,1}$, where $\delta_{x,y}$ is the

Kronecker delta function. The population growth of the initial and intermediate vertices is crucial and so the notation $\mathcal{Z}(t) = (Z_x(t))_{1\leq x\leq N-1}$ will be used.

We also recap some relevant definitions from the main text that will be used repeatedly. First, the target hitting time is defined to be

$$T = \inf\{t \geq 0 : \, Z_N(t) > 0\}. \tag{21}$$

Now let the set of paths between the root and any vertex $x \in V$ be be denoted $\mathcal{P}_{1,x}$. Then we define the weight of the path $p \in \mathcal{P}_{1,x}$ to be

$$w(p) = \nu(p_1, p_2) \prod_{i=2}^{l} \frac{\nu(p_i, p_{i+1})}{\lambda - \lambda(p_i)}, \tag{22}$$

Throughout the empty product is set to 1. Further, we let the total weight of the the vertex x be

$$\phi_x = \sum_{p\in\mathcal{P}_{1,x}} w(p). \tag{23}$$

5.2 Population growth

In this section we prove Theorem 3, but G is not assumed to be acyclic. Specialising to the acyclic case is in Section 5.4.

As the root vertex is a source, and hence no type transitions into vertex 1 cells, $Z_1(t)$ follows a linear birth-death process. Its asymptotic behaviour is described by the following classic result [4, 24].

Lemma 4.

$$\lim_{t\to\infty} Z_1(t)e^{-\lambda t} = W \quad a.s.$$

Here W is distributed as the sum of K independent exponential random variables with parameter λ/α, where K is binomial with z trials and success probability λ/α.

W, as given in the above theorem, is positive if and only if the vertex 1 population survives (almost surely). It will be convenient to define this event and so we let

$$S_1 = \{Z_1(t) > 0 \text{ for all } t \geq 0\}. \tag{24}$$

From the distribution of W, in particular using that $\mathbb{P}(S_1) = \mathbb{P}(W > 0) = \mathbb{P}(K > 0)$ with K as in Lemma 4, we immediately have

$$\mathbb{P}(S_1) = 1 - (\beta/\alpha)^z. \tag{25}$$

To obtain a similar result to Lemma 4 for $\mathcal{Z}(t)$, we recall the $N-1 \times N-1$ matrix A from Section 2.4, with entries

$$a_{ij} = \begin{cases} \lambda(i) & j = i \\ \nu(j,i) & \text{if } (j,i) \in E \\ 0 & \text{otherwise.} \end{cases}$$

Note that the off diagonal elements of the transpose of A are positive whenever the corresponding element of the adjacency matrix of G is one.

We will be using a result from [44] which holds when the largest real eigenvalue of A, λ^*, is simple (i.e. has algebraic multiplicity 1) and equal to λ. Below we give two sufficient conditions for this to be true. The second is motivated by applications when typically all transition rates are small.

Lemma 5. *Two sufficient conditions such that $\lambda^* = \lambda$ and is simple are: 1) G is acyclic. 2) For all $2 \leq i \leq N-1$, $\lambda(i) + \sum_{k\in\mathcal{N}^-(i)} \nu(k,i) < \lambda$.*

Proof. 1) As G is directed and acyclic, a relabelling of the vertices exists (often called the topological sorting, see [19] chapter 22.4) such that the underlying adjacency matrix of the relabelled graph is upper triangular. We use this relabelling with vertex 1 as first in the labelling, π. Thus π is a permutation of $\{1, \ldots, N\}$ with $\pi(1) = 1$. The permutation matrix associated with π and its transpose are

$$P_1 = \begin{bmatrix} \mathbf{e}_{\pi(1)}^{\mathrm{T}} \\ \vdots \\ \mathbf{e}_{\pi(n)}^{\mathrm{T}} \end{bmatrix}, \quad P_1^{\mathrm{T}} = \begin{bmatrix} \mathbf{e}_{\pi(1)}, \cdots, \mathbf{e}_{\pi(N)} \end{bmatrix}.$$

Under π we have the new matrix

$$A' = P_1 A P_1^{-1}.$$

Due to the relation between the adjaceny matrix of G and A, we see A' is lower triangular and has the $\lambda(i)$ as diagonal elements. Thus λ is the largest eigenvalue of A'. As A and A' are similar matrices, and hence share eigenvalues, we can conclude the statement.

2) By considering the left eigenvector $\mathbf{e}_1^{\mathrm{T}}$, λ is indeed an eigenvalue. We now demonstrate all other eigenvalues have real part smaller than λ, which implies λ is simple. Observe that the assumed condition implies the bound on the row sums

$$\sum_{j=1}^{N-1} a_{i,j} < \lambda. \tag{26}$$

As A is reducible (due to the root being a source vertex) there exists a permutation matrix P_2 such that

$$P_2 A P_2^{-1} = \begin{pmatrix} \lambda & & & \\ R_{2,1} & R_{2,2} & & 0 \\ \vdots & & \ddots & \\ R_{M,1} & \ldots & & R_{M,M} \end{pmatrix}$$

with each submatrix $R_{i,i}$ square and irreducible and $2 \leq M \leq N-1$ (see [72] section 2.3). The eigenvalues of A comprise λ and the eigenvalues of the $R_{i,i}$. For any of the $R_{i,i}$ let $r_j^{(i)}$ be sum of the jth row of $R_{i,i}$. By Gershgorin's Disc Theorem ([72] Theorem 1.5) the real part of the eigenvalues of $R_{i,i}$ are bounded by $\max_j\{r_j^{(i)}\}$. Further, the bound (26) is preserved under permutations and so each $r_j^{(i)} < \lambda$. As this holds for each $R_{i,i}$ the claim is shown. □

Henceforth we assume either of the conditions given in Lemma 5 hold and thus $\lambda^* = \lambda$ and is simple. Due to this assumption, results of [44] show that the vertex 1 cells drives the entire population growth. To state this result, we introduce the vector ψ with

$$\psi_i = \alpha(i) + \beta(i) + \sum_{k \in \mathcal{N}^+(i)} \nu(i,k).$$

Now let $\tilde{u}, \tilde{v}$ be the left and right eigenvectors of A corresponding to λ, scaled such that

$$\psi \cdot \tilde{v} = 1, \quad \tilde{u} \cdot \tilde{v} = 1. \tag{27}$$

Note, from the structure of A, $\tilde{u}_i > 0$ if $i = 1$ and 0 otherwise. Further let

$$\Phi = \tilde{u}_1 \tilde{v}$$

Then the following result, which is the more general form of Theorem 3, holds.

Theorem 6. *With W as in Lemma 4,*

$$\lim_{t\to\infty} e^{-\lambda t}\mathcal{Z}(t) = W(\Phi_x)_{x=1}^{N-1} \quad a.s.$$

Proof. We demonstrate how to apply the result in [44] to our setting. From Theorem 3.1 in [44], with probability one

$$\lim_{t\to\infty} e^{-\lambda t}\mathcal{Z}(t) = \hat{W}\tilde{v} \tag{28}$$

where $\hat{W}$ is a non-negative random variable, with as yet unknown distribution. However from Lemma 4 almost surely $e^{-\lambda t}Z_1(t) \to W$, with the distribution of W given in the lemma. This implies $\hat{W} = W/\tilde{v}_1 = \tilde{u}_1 W$ with the last equality following from (27) and that only the first entry of $\tilde{u}$ is positive. □

The proof of Theorem 6 indicates our primary reason for taking the root vertex to be a source (no incoming edges). If transitions were permitted into the vertex 1 population, but still $\lambda^* = \lambda$ and is simple, then (28) still applies but we do not know the distribution of $\hat{W}$. Additionally, regardless of whether the root is a source, if we consider the the more general offspring distribution discussed in Section 2.4.3, we similarly would have no relevant knowledge of the distribution of $\hat{W}$ (the Laplace transform of $\hat{W}$ is known to satisfy an implicit integral equation [4], but we cannot see how to use this information here).

5.3 Time until target vertex is populated: general case

We now turn our attention to T, the time at which the target population arises, as defined in (21). So that we can control all target seeding transition rates we let

$$\nu^* = \max_{i\in\mathcal{N}^-(N)} \nu(i,N)$$

i.e. the largest of the transition rates into the target. We further give the definition of μ in this more general setting as

$$\mu = \frac{1}{\lambda}\log\frac{\lambda^2}{\alpha\sum_{i\in\mathcal{N}^-(N)}\nu(i,N)\Phi_i}. \tag{29}$$

Then the following holds.

Theorem 7. *Let W be as in Theorem 6, then with probability one*

$$\lim_{\nu^*\to 0}\mathbb{P}(\lambda(T-\mu) > t|(\mathcal{Z}(t))_{t\geq 0})) = \exp\left(-e^t W\lambda/\alpha\right).$$

It follows that

$$\lim_{\nu^*\to 0}\mathbb{P}(\lambda(T-\mu) > t) = \left(\frac{\lambda/\alpha}{1+e^t} + \beta/\alpha\right)^z$$

Proof. Let

$$\sigma = \sum_{i\in\mathcal{N}^-(N)}\nu(i,N)\Phi_i.$$

Then, as for given $(\mathcal{Z}(t))_{t\geq 0}$ immigration into vertex N occurs as a non-homogeneous Poisson process,

$$\mathbb{P}(T-\lambda^{-1}\log(\sigma^{-1}) > t|(\mathcal{Z}(t))_{t\geq 0}) = \exp\left(-\int_0^{t+\lambda^{-1}\log(\sigma^{-1})}\sum_{i\in\mathcal{N}^-(N)}\nu(i,N)Z_i(s)\,ds\right). \tag{30}$$

The variable change $u = s - \lambda^{-1}\log(\sigma^{-1})$ yields

$$\exp\left(-\int_{-\lambda^{-1}\log(\sigma^{-1})}^{t}\sum_{i\in\mathcal{N}^-(N)}\nu(i,N)Z_i(u+\lambda^{-1}\log(\sigma^{-1}))\,du\right). \tag{31}$$

Now observe that for any $i \in \mathcal{N}^-(N)$, by multiplying by $1 = \frac{e^{\lambda(u+\lambda^{-1}\log(\sigma^{-1}))}}{e^{\lambda(u+\lambda^{-1}\log(\sigma^{-1}))}}$,

$$\nu(i,N)Z_i(u+\lambda^{-1}\log(\sigma^{-1})) = \frac{Z_i(u+\lambda^{-1}\log(\sigma^{-1}))}{e^{\lambda(u+\lambda^{-1}\log(\sigma^{-1}))}}\frac{\nu(i,N)e^{\lambda u}}{\sigma}. \tag{32}$$

Consider the limit of (32) as $\nu^* \to 0$, which results in $\log(\sigma^{-1}) \to \infty$. As the $(Z_i(t))_{i\in\mathcal{N}^-(N)}$ are independent of $(\nu(i,N))_{i\in\mathcal{N}^-(N)}$, Theorem 6 informs us that, almost surely,

$$\lim_{\nu^*\to 0}\frac{Z_i(u+\lambda^{-1}\log(\sigma^{-1}))}{e^{\lambda(u+\lambda^{-1}\log(\sigma^{-1}))}} = We^{\lambda u}\Phi_i \tag{33}$$

Upon considering $(\sigma/\nu(i,N))^{-1}$, we see $\nu(i,N)/\sigma < \infty$ for each i. Hence taking limits across (32) yields, with probability one,

$$\lim_{\nu^*\to 0}\sum_{i\in\mathcal{N}^-(N)}\frac{Z_i(u+\lambda^{-1}\log(\sigma^{-1}))}{e^{\lambda(u+\lambda^{-1}\log(\sigma^{-1}))}}\frac{\nu(i,N)e^{\lambda u}}{\sigma} = We^{\lambda u}\lim_{\nu^*\to 0}\sum_{i\in\mathcal{N}^-(N)}\frac{\nu(i,N)\Phi_i}{\sigma}.$$

By the definition of σ this last sum is one and hence

$$\lim_{\nu^*\to 0}\sum_{i\in\mathcal{N}^-(N)}\nu(i,N)Z_i(u+\lambda^{-1}\log(\sigma^{-1})) = We^{\lambda u} \quad a.s. \tag{34}$$

We seek to apply the dominated convergence theorem to the integral in (31). The limit demonstrated in (34) implies that for any realisation, we may find x such that for $\nu^* \le x$

$$\sum_{i\in\mathcal{N}^-(N)}\nu(i,N)Z_i(u+\lambda^{-1}\log(\sigma^{-1})) \le 2We^{\lambda u}$$

which is integrable over $(-\infty, t]$. If $\nu^* > x$ then $\sigma > 0$. Thus the interval $[-\lambda^{-1}\log(\sigma^{-1}), t]$ is finite. As each $Z_i(t)$ is cadlag, it is bounded over finite intervals. Hence for all (ν^*, u) the integrand is dominated by an integrable function. Therefore, taking the limit of (31) and using (34) leads to, with probability one,

$$\lim_{\nu^*\to 0}\mathbb{P}(T-\lambda^{-1}\log(\sigma^{-1}) > t|(\mathcal{Z}(t))_{t\ge 0}) = \exp\left(-We^{\lambda t}/\lambda\right).$$

Using that

$$\mu = \lambda^{-1}\log(\lambda^2/\alpha) + \lambda^{-1}\log(\sigma^{-1})$$

so that we centre appropriately, leads to the first stated result.

For the second statement, we must derive an expression for

$$\mathbb{E}\left[\exp\left(-e^t W\lambda/\alpha\right)\right]$$

From Lemma 4

$$W\lambda/\alpha \stackrel{d}{=} \sum_{i=1}^{z}\chi_i\xi_i'$$

where $\chi_i \sim$ Bernoulli(λ/α), $\xi_i' \sim$ Exponential(1) and are independent. Thus

$$\mathbb{E}\left[\exp\left(-e^t W\lambda/\alpha\right)\right] = \left(\mathbb{E}\left[\exp\left(-e^t\chi_1\xi_1'\right)\right]\right)^z.$$

Using a one-step conditioning argument on χ_1 and considering the Laplace transform for a unit rate exponential variable leads to the stated result. □

Before proceeding, we comment on a limitation of our proof approach. A key point in the proof of Theorem 7 is that the stochastic process $(Z_i(t))_{1\leq i\leq N-1,t\geq 0}$ is independent of the target seeding transition rates $(\nu(i,N))_{i\in\mathcal{N}^-(N)}$. This permits the limit displayed in (33) to hold. If we considered a limit where transition rates associated with non-target adjacent edges also tended to 0, the limit (33) would not hold. This is as Theorem 6 is true for fixed $(\alpha(x))_{x=1}^{N-1}$, $(\beta(x))_{x=1}^{N-1}$ and transition rates associated with non-target adjacent edges (the parameters controlling the growth of the populations at vertices $1,\ldots,N-1$). Our inability to alter these parameters is also the reason, that in a precise sense, our results do not cover the formulation of transitions where type x divide at rate $\alpha'(x)$, and then with probability $\nu'(x,y)$ a transition occurs to vertex y. In that formulation, taking the limit when $\nu'(x,N)$ tends to 0, for $x\in\mathcal{N}^-(N)$, alters the division rate of cells at vertices in $\mathcal{N}^-(N)$, again forbidding the use of Theorem 6. We hope to improve upon these limitations in future work.

The first statement in Theorem (7) will be key to understanding the path distribution, discussed in Section 2.3, which justifies its prominence. As it implies that, with probability one,

$$\lim_{\nu^*\to 0}\mathbb{P}(\exp\left[\lambda(T-\mu)\right]>t|(\mathcal{Z}(t))_{t\geq 0}))=\exp\left(-tW\lambda/\alpha\right),$$

the first statement in Theorem (7) can be interpreted as indicating that if we consider the conditional distribution of T around μ, then rescale time, an exponential distribution appears.

As discussed in Section 2.2, it is natural to consider the distribution of T conditioned that it is finite (we cannot discuss which path populated the target vertex if $T=\infty$). However it is technically more convenient to condition on S_1, defined in (24). The following proposition states that in many relevant cases, namely large initial population, low death rate or small transition rates leaving vertex 1. these events are similar. Below, the superscript c denotes the complement of the set.

Proposition 1.

$$\mathbb{P}(\{T<\infty\}^c\cap S_1)=0.$$

$$\mathbb{P}(\{T<\infty\}\cap S_1^c)\leq\begin{cases}O((\beta/\alpha)^z), & \beta\to 0\\ O((\beta/\alpha)^z), & z\to\infty\\ O(\nu_1), & \nu_1\to 0\end{cases}$$

where $\nu_1=\sum_{x\in\mathcal{N}^+(1)}\nu(1,x)$.

Proof. As $\mathbb{P}(\{T<\infty\}^c\cap S_1)=\mathbb{P}(T=\infty|S_1)\mathbb{P}(S_1)$ and $\mathbb{P}(S_1)>0$ we initially demonstrate

$$\mathbb{P}(T<\infty|S_1)=1. \tag{35}$$

Firstly recall that with W as in Lemma 4, up to null events, $\{W>0\}=S_1$. Now T is the first arrival in a Cox process with intensity $\sum_{x\in\mathcal{N}^-(N)}\nu(x,N)Z_x(s)$, and hence $T<\infty$ with probability 1 on S_1 if

$$\int_0^\infty\sum_{x\in\mathcal{N}^-(N)}\nu_{x,N}Z_x(s)ds=\infty \tag{36}$$

with probability 1 on S_1. Observe that (36) is true as, by Theorem 6, for any realisation we may find $t_1\geq 0$ such that

$$\int_{t_1}^\infty\sum_{x\in\mathcal{N}^-(N)}\nu(x,N)Z_x(s)ds\geq 2^{-1}\int_{t_1}^\infty\sum_{x\in\mathcal{N}^-(N)}\nu(x,N)W\Phi_x e^{\lambda s}ds.$$

As each $\Phi_x>0$, due to each entry of $\tilde{v}$ also being positive [44], the right hand side of the above inequality is infinite a.s. on S_1 which gives (35). It remains to bound $\mathbb{P}(\{T<\infty\}\cap S_1^c)$.

Let us introduce $Z^*(t)=\sum_{i\in\mathcal{N}^+(1)}Z_i(t)$ and

$$\tau=\inf\{t\geq 0:Z^*(t)>0\}.$$

Note that we cannot hope to populate the target vertex without $Z^*(t)$ becoming positive. Therefore, $\{T < \infty\} \subseteq \{\tau < \infty\}$ which implies

$$\mathbb{P}(\{T < \infty\} \cap S_1^c) \leq \mathbb{P}(\{\tau < \infty\} \cap S_1^c).$$

Hence our interest turns to the right hand side of the preceding inequality. The distribution of τ is invariant to the growth at vertices $i \in \mathcal{N}^+(1)$, and so we let $\alpha(i) = \beta(i) = 0$ for such i. The process $(Z_1(t), Z^*(t))$ is a two-type branching process where cells of the second type (that contribute to $Z^*(t)$) simply accumulate. As

$$\begin{aligned} \mathbb{P}(\{\tau < \infty\} \cap S_1^c) &= (1 - \mathbb{P}(\{\tau = \infty\}|S_1^c))\mathbb{P}(S_1^c) \\ &= \mathbb{P}(S_1^c) - \lim_{t\to\infty} \mathbb{P}(Z_1(t) = 0, Z^*(t) = 0), \end{aligned} \tag{37}$$

and from (25) we know $\mathbb{P}(S_1^c) = \beta/\alpha$, we now solve for the joint distribution of $(Z_1(t), Z^*(t))$.

For now assume $z = 1$. We introduce the generating functions

$$\begin{aligned} G_1(x, y, t) &= \mathbb{E}[x^{Z_1(t)} y^{Z^*(t)} | (Z_1(0), Z^*(0)) = (1, 0)] \\ G^*(x, y, t) &= \mathbb{E}[x^{Z_1(t)} y^{Z^*(t)} | (Z_1(0), Z^*(0)) = (0, 1)]. \end{aligned}$$

Immediately $G^*(x, y, t) = y$ for all t. The backward Kolmogorov equation for G_1 is

$$\frac{\partial}{\partial t} G_1 = \alpha G_1^2 + \beta + \nu_1 G_1 G^* - (\alpha + \beta + \nu_1) G_1.$$

By solving this equation (separation of variables with a partial fraction decomposition), and using the initial condition we find

$$\mathbb{P}(Z_1(t) = 0, Z^*(t) = 0) = G_1(0, 0, t) = \frac{r_2(1 - e^{\alpha(r_2 - r_1)t})}{1 - r_2 e^{\alpha(r_2 - r_1)t}/r_1}.$$

With

$$\begin{aligned} r_1 &= \frac{1}{2}\left(1 + q_1 + \nu_1' + \sqrt{(1 + q_1 + \nu_1')^2 - 4q_1}\right) \\ r_2 &= \frac{1}{2}\left(1 + q_1 + \nu_1' - \sqrt{(1 + q_1 + \nu_1')^2 - 4q_1}\right) \end{aligned}$$

where $q_1 = \beta/\alpha$ and $\nu_1' = \nu_1/\alpha$. Observe that $(1 + q_1 + \nu_1')^2 - 4q_1 > (1 - q_1)^2$ which implies $r_2 - r_1 < 0$. Hence

$$\lim_{t\to\infty} \mathbb{P}(Z_1(t) = 0, Z^*(t) = 0) = r_2.$$

We now have expressions for the final terms in (37) for $z = 1$. For $z \geq 1$, we can use independence, and this leads to

$$\mathbb{P}(\{\tau < \infty\} \cap S_1^c) = q_1^z - r_2^z.$$

By examining the leading order with respect to the relevant variables we can conclude.

□

We remark that if we instead wish to look at the case when the vertex N population arises and does not go extinct then $\mathbb{P}(\text{All populations go extinct}) = \mathbb{P}(\tilde{T} = \infty)$ where $\tilde{T}$ is the first time the target is founded by a cell whose progeny does not go extinct. Recall that our results for T hold identically for $\tilde{T}$ so long as the mapping $\nu(x, N) \mapsto \nu(x, N)\lambda(x)/\alpha(x)$ for all $x \in \mathcal{N}^-(N)$ is applied . The probability of population extinction is known to be given by an implicit equation involving the offspring distribution's generating function [4]. This was the approach taken by [42]. The approximate solution for small transition rates was derived and agrees with the result presented above (see Eq. A.4a in Appendix A of [42]).

The above proposition yields $\mathbb{P}(T < \infty) \approx \mathbb{P}(S_1)$. Using this with (25) gives (6) in the main text. We now give consider the distribution of T given S_1 (which we use as proxy for conditioning on a finite hitting time).

Corollary 7.1. *With the notation of Theorem 7,*

$$\lim_{\nu^*\to 0} \mathbb{P}(\lambda(T-\mu) > t|S_1) = \frac{\left[\lambda/\alpha(1+e^t)^{-1} + \beta/\alpha\right]^z - (\beta/\alpha)^z}{1-(\beta/\alpha)^z}$$

Proof. Starting from the first statement in Theorem 7, we only need apply $\mathbb{E}[\cdot|S_1]$. The distribution of $W\lambda/\alpha$ can be rewritten as

$$W\lambda/\alpha \stackrel{d}{=} \sum_{i=1}^{K} \xi_i'$$

where ξ_i' are rate one exponential variables, and $K \sim \text{Binomial}(z, \lambda/\alpha)$, independent of all ξ_i'. Conditioning on S_1 ensures that $K > 0$.

Thus we must calculate

$$\mathbb{E}\left[\exp\left(-e^t W\lambda/\alpha\right)|S_1\right] = \mathbb{E}\left[(1+e^t)^{-K}|K>0\right].$$

The binomial theorem leads to the stated expression. □

We now prove a corollary regarding μ (defined in (29)), which provides justification to (7).

Corollary 7.2. *Let $t_{1/2}$ be the median time for the vertex N population to arise, conditioned on S_1. That is $t_{1/2}$ satisfies*

$$\mathbb{P}(T > t_{1/2}|S_1) = \frac{1}{2}. \tag{38}$$

Then

$$\lim_{\nu^*\to 0} |t_{1/2} - (\mu - h(z))| = 0$$

where

$$h(z) = -\lambda^{-1}\log\left(\frac{\lambda/\alpha}{2^{-1/z}(1+(\beta/\alpha)^z)^{1/z} - \beta/\alpha} - 1\right).$$

This also implies $t_{1/2} \sim \mu - h(z)$. Regarding $h(z)$ we have

$$h(1) = 0, \quad h(z) = \lambda^{-1}\log\left(\frac{z\lambda}{\alpha}\right) + O(z)^{-1}, \quad z\to\infty.$$

Proof. We firstly note, as is apparent from the cumulative distribution function of T (see (30) for instance), that $t_{1/2}$ is monotone increasing as ν^* decreases. By (38)

$$\frac{1}{2} = \mathbb{P}(\lambda(T-\mu) > \lambda(t_{1/2}-\mu)|S_1).$$

Taking the $\nu^*\to 0$ limit we find $\lambda(t_{1/2}-\mu)\to x$, where the convergence is guaranteed as both $t_{1/2}$, μ are monotone increasing. Further, from Corollary 7.1, the limit x satisfies

$$\frac{[\lambda/\alpha(1+e^x)^{-1} + \beta/\alpha]^z - (\beta/\alpha)^z}{1-(\beta/\alpha)^z} = \frac{1}{2}.$$

Solving for x leads to $x = -\lambda h(z)$ as stated. For the asymptotic result we firstly consider

$$d(z) = \frac{\lambda/\alpha}{2^{-1/z}(1+(\beta/\alpha)^z)^{1/z} - \beta/\alpha} - 2^{\frac{\alpha}{\lambda z}}$$

Expansions on the numerator and denominator show that $d(z)$ is $O(z^{-2})$. Then

$$\begin{aligned} -\lambda h(z) &= \log(2^{\frac{\alpha}{\lambda z}} - 1 + d(z)) \\ &= \log(2^{\frac{\alpha}{\lambda z}} - 1) + \frac{d(z)}{2^{\frac{\alpha}{\lambda z}} - 1} + \dots \end{aligned}$$

As $(2^{\frac{\alpha}{\lambda z}} - 1)^{-1} = O(z)$ we have shown the claimed asymptotic behaviour.

□

5.4 Path graph and acyclic graphs

In this section we specialise the results of Sections 5.2 and 5.3 firstly to the case of a path graph and then to acyclic graphs.

We first consider a path graph with N vertices and let $\tilde{p}_{1:x} = (1, 2, \dots x)$. In this setting the matrix A from Section 5.2 has entries

$$a_{ij} = \begin{cases} \lambda(i) & j = i \\ \nu(j,i) & j = i-1 \\ 0 & \text{otherwise,} \end{cases}$$

and the vector ψ has entries $\psi_i = \alpha(i) + \beta(i) + \nu(i, i+1)$ for $i = 1, \dots, N-1$. We want a useful representation of the sequence $(\Phi_x)_{x=1}^{N-1}$, which we recall is given in terms of the entries of the suitably normalised left and right eigenvectors of A (see (27)). Because they are easier to work with (i.e. check they indeed are eigenvectors) we first introduce the unnormalised left and right vectors u', v' with entries

$$u'_i = \begin{cases} 1, & i = 1 \\ 0 & 2 \le i \le N-1 \end{cases} \qquad v'_i = \prod_{j=2}^{i} \frac{\nu(j-1,j)}{\lambda - \lambda(j)}, \quad 1 \le i \le N-1$$

where the empty product is again set equal to 1. Note for $2 \le x \le N-1$ we have

$$v'_x = \frac{w(\tilde{p}_{1:x})}{\lambda - \lambda(x)},$$

here $w(\cdot)$ is again the path weight defined in (22) . It may be directly verified that these are left and right eigenvectors for A with eigenvalue λ (for v' it is helpful to note that $Av' = \lambda v'$ implies $\nu(i-1,i)v'_{i-1} + \lambda(i)v'_i = \lambda v'_i$) . Now we normalise these, so that they agree with (27) via

$$\tilde{u} = \frac{\psi \cdot v'}{v'_1} u', \quad \tilde{v} = \frac{1}{\psi \cdot v'} v'.$$

Our aim is to obtain $\Phi = \tilde{u}_1 \tilde{v}$, and as both $u'_1 = v'_1 = 1$, we see that $\Phi = v'$. Using this, the following corollary demonstrates how the results of the previous section on population growth and time to reach the target simplifies in the path graph setting.

Corollary 7.3. *The results of Sections 5.2 and 5.3 hold with the explicit representations*

$$\Phi_x = \begin{cases} 1 & x = 1 \\ \frac{w(\tilde{p}_{1:x})}{\lambda - \lambda(x)} & 2 \le x \le N-1 \end{cases}$$

$$\nu^* = \nu(N-1, N),$$

and

$$\mu = \lambda^{-1} \log \left[\lambda^2 \left(\alpha \phi_N \right)^{-1} \right].$$

The appearance of ϕ_N in the expression for μ in Corollary (7.3) is due to $\phi_N = w(\tilde{p}_{1:N}) = \nu(N-1, N)\Phi_{N-1}$ (to be compared with the general definition of μ (29)). For referencing, note the definition for the weight of a path p, $w(p)$, and total weight of N, ϕ_N, are in (22) and (23). We now move to the case where G is acyclic.

Recall the concept of vertex lineage from Section 2.3 in the main text (which for a chosen cell, is a sequence recording the vertices of cells in the ancestral lineage of our chosen cell). We denote the number of cells at time t with vertex lineage p as $X(p,t)$. The number of cells at a vertex is related to the cells with vertex lineages ending at that vertex via

$$Z_x(t) = \sum_{p \in \mathcal{P}_{1,x}} X(p,t). \tag{39}$$

For any path p, as above we let the first j terms in p be denoted $p_{1:j}$. The process $(X(p_{1:j},t))_{j=1}^{l+1}$ is a multitype branching process, and $X(p_{1:j},t)$ represents the number of cells that have progressed to step j along path p at time t. Births occur to individuals of type i at at rate $\alpha(p_i)$, deaths at rate $\beta(p_i)$ and individuals of type i transition to type $i+1$ at rate $\nu(p_i,p_{i+1})$. Thus we may use our results for path graphs with vertices labelled $1,\dots,l+1$. Hence via Theorem 6 and Corollary 7.3 we have the following.

Corollary 7.4.

$$\lim_{t\to\infty} e^{-\lambda t}X(p,t) = Ww(p) \quad a.s.$$

where $w(p)$ is the path weight given in (22) *and W is as in Lemma 4.*

Analogously to Corollary 7.3 we have the following statement regarding the growth of the population at the initial and intermediate vertices.

Corollary 7.5. *The results of Sections 5.2 and 5.3 hold with the explicit representations*

$$\Phi_x = \begin{cases} 1 & x=1 \\ \frac{1}{\lambda-\lambda(x)}\sum_{p\in\mathcal{P}_{1,x}} w(p) & 2\le x\le N-1 \end{cases}$$

$$\mu = \lambda^{-1}\log\left(\lambda^2\left[\alpha\sum_{p\in\mathcal{P}_{1,N}} w(p)\right]^{-1}\right).$$

where $w(p)$ is the path weight given in (22).

Note that as the empty product is defined to be 1, $\phi_1 = 1$.

Proof of Corollary 7.5. Due to Theorem 6 we must demonstrate

$$\lim_{t\to\infty} e^{-\lambda t}Z_x(t) = W\sum_{p\in\mathcal{P}_{1,x}} w(p) \quad a.s.$$

By (39) it is enough to show

$$\mathbb{P}\left(\bigcap_{p\in\mathcal{P}_{1,x}} \lim_{t\to\infty} e^{-\lambda t}X(p,t) = W\right) = 1.$$

Observe that for any events $(A_i)_{i=1}^n$ such that for all i, $\mathbb{P}(A_i)=1$ then $\mathbb{P}(\cap_i^n A_i)=1$. Use this with events $A_p = \{\lim_{t\to\infty} e^{-\lambda t}X(p,t) = W\}$ to conclude. □

One particular consequence, by coupling Corollaries 7.1 and 7.5 is that for ν^* small

$$\mathbb{P}(T>t|S_1) \approx \frac{\left[\lambda/\alpha(1+e^{\lambda t}\phi_N\alpha/\lambda^2)^{-1}+\beta/\alpha\right]^z - (\beta/\alpha)^z}{1-(\beta/\alpha)^z} \tag{40}$$

which is the conditional version of (5). Differentiating yields an approximation for the conditional density

$$f_{T|T<\infty}(t) \approx \frac{ze^{\lambda t}\phi_N\left(\frac{\lambda/\alpha}{1+e^{\lambda t}\phi_N\alpha/\lambda^2}+\beta/\alpha\right)^{z-1}}{(1-(\beta/\alpha)^z)\left(1+\frac{\alpha e^{\lambda t}\phi_N}{\lambda^2}\right)^2}. \tag{41}$$

5.5 Distribution of the path to the target

We continue to take G acyclic. Note that $|\mathcal{P}_{1,N}|$ is finite, and so we let $n = |\mathcal{P}_{1,N}|$. The elements of $\mathcal{P}_{1,N}$ can be enumerated and we use the following notation to do so $\mathcal{P}_{1,N} = \{p^{(1)}, \ldots, p^{(n)}\}$. For each $p^{(i)} \in \mathcal{P}_{1,N}$ we let: the path lengths be $l^{(i)} = |p^{(i)}| - 1$, $u^{(i)} = \nu(p^{(i)}_{l^{(i)}}, N)$ be the final transition rates, and $\eta^{(i)} = w(p^{(i)})/u^{(i)}$. Consider $p^{(i)} \in \mathcal{P}_{1,N}$ and recall that $X(p^{(i)}_{1:j}, t)$ represents the number of cells that have progressed to step j along path i at time t. If we apply the first statement in Theorem 7 to the branching process $(X(p^{(i)}_{1:j}, t)))_{j=1}^{l^{(i)}+1}$ we can deduce that under a time rescaling, the time to traverse path i is approximately exponentially distributed, conditional on the population growth along that path. Therefore the question of which path the target population arises from becomes equivalent to the minimum of a set of exponential random variables. To simplify notation we let,

$$T^{(i)} = T(p^{(i)})$$

where $T(p^{(i)})$ is defined in (10). Also recall the notation

$$T(\neg p) = \min\{T(q) : q \in \mathcal{P}_{1,N} \backslash \{p\}\}.$$

We can now prove the precise version of Theorem 2

Theorem 8. *Assume the limits* $\lim_{\nu^* \to 0} \frac{u^{(i)}}{u^{(j)}}$ *exist, but may be infinite, and further that there exists* $i^* \in \{1, \ldots, n\}$ *such that* $\lim_{\nu^* \to 0} \frac{u^{(i)}}{u^{(i^*)}} < \infty$ *for all* $i = 1, \ldots n$. *Then, for* $p \in \mathcal{P}_{1,N}$ *and* $t \in \mathbb{R}$,

$$\lim_{\nu^* \to 0} \mathbb{P}(T(\neg p) - T(p) > t | S_1) = \lim_{\nu^* \to 0} \frac{w(p)}{w(p) + e^{\lambda t} \sum_{q \in \mathcal{P}_{1,N} \backslash \{p\}} w(q)}.$$

Proof. Consider initially the case when $n = 2$, that is there are only two paths from vertex 1 to vertex N. These two paths must diverge before N. If the population growth along these paths until the target vertex (but excluding the population growth at the target vertex itself) is given, then $T^{(1)}$ and $T^{(2)}$ are the first arrival times in two independent non-homogeneous Poisson processes. Returning to arbitrary n, to use this fact we define the population numbers along all the paths up until the target vertex as

$$\mathcal{X}(s) = (X(p^{(i)}_{1:j}, s))_{i=1,j=1}^{n,l^{(i)}}.$$

So that we can use this conditional independence, and prior results, we note

$$\mathbb{P}(T(\neg p) - T(p) > t | S_1) = \frac{\mathbb{E}[\mathbb{I}_{S_1} \mathbb{P}(T(\neg p) - T(p) > t | (\mathcal{X}(s))_{s \geq 0})]}{\mathbb{P}(S_1)}. \tag{42}$$

We now focus on the argument of the above expectation. Let the path under consideration be path 1 ($p = p^{(1)}$) and

$$\kappa^{(i)} = \lim_{\nu^* \to 0} \frac{u^{(i)}}{u^{(i^*)}}. \tag{43}$$

We know, by Theorem 7, that each $T^{(i)}$ appropriately centred converges. If we centre each $T^{(i)}$ by the same quantity, then the ordering is preserved. That is, clearly $T^{(1)} < T^{(2)} \iff T^{(1)} - \mu^{(i)} < T^{(2)} - \mu^{(i)}$ for any $\mu^{(i)}$. As $T^{(i)}$ is the target time along the path graph $p^{(i)}$, Corollary 7.3 gives the appropriate centring. In particular with

$$\mu^{(i)} = \lambda^{-1} \log(\lambda^2 [\alpha \eta^{(i)} u^{(i)}]^{-1})$$

we have, by the first statement of Theorem 7 and Corollary 7.3,

$$\lim_{\nu^* \to 0} \mathbb{P}(\lambda(T^{(i)} - \mu^{(i)}) > x | (\mathcal{X}(s))_{s \geq 0}) = \exp(-e^x W \lambda / \alpha) \, a.s.$$

For order preservation (of the $T^{(i)}$s) we take $\mu^{(i^*)}$ as a common centring of all $T^{(i)}$, and use $T^{(i)} - \mu^{(i^*)} = T^{(i)} - \mu^{(i)} + \mu^{(i)} - \mu^{(i^*)}$ and

$$\lim_{\nu^* \to 0} \mu^{(i)} - \mu^{(i^*)} = \lambda^{-1} \log \left(\frac{\eta^{(i^*)}}{\eta^{(i)} \kappa^{(i)}} \right).$$

Note the above expression may be $-\infty$. Using this, and exponentiating (which again preserves order), leads to

$$\lim_{\nu^* \to 0} \mathbb{P}(\exp[\lambda(T^{(i)} - \mu^{(i^*)})] > t | (\mathcal{X}(s))_{s \geq 0}) = \exp \left(- \frac{tW\lambda \eta^{(i)} \kappa^{(i)}}{\alpha \eta^{(i^*)}} \right) \ a.s.$$

At this point we note that if $\kappa^{(i)} = 0$, then for fixed $(\mathcal{X}(s))_{s \geq 0}$, the random variables $\exp[\lambda(T^{(i)} - \mu^{(i^*)})]$ converges to $+\infty$ almost surely. Recall the $T^{(i)}$ are conditionally independent, and hence convergence of the marginal distributions implies joint convergence. Therefore for $(x_j)_{j=1}^n$ with each $x_j \geq 0$, almost surely we have

$$\lim_{\nu^* \to 0} \mathbb{P} \left[\left(\exp[\lambda(T^{(j)} - \mu^{(i^*)})] > x_j \right)_{j=1}^n | (\mathcal{X}(s))_{s \geq 0} \right] = \prod_{j=1}^n \exp \left(- \frac{x_j W \kappa^{(j)} \eta^{(j)} \lambda}{\eta^{(i^*)} \alpha} \right).$$

Including the shift t, we have for fixed $(\mathcal{X}(s))_{s \geq 0}$ the random vector $(\exp[\lambda(T^{(1)} + t - \mu^{(i^*)})], \exp[\lambda(T^{(j)} - \mu^{(i^*)})])_{j=2}^n$ jointly tends in distribution to the random vector $(U_i)_{i=1}^n$. Note that only $T^{(1)}$ is shifted, due to the path under consideration being $p^{(1)}$. If $W = 0$, all the $U_i = +\infty$ a.s., instead if $W > 0$ each U_i is independent and for $2 \leq i \leq n$ has distribution

$$U_i \stackrel{d}{=} \begin{cases} +\infty \text{ if } \kappa^{(i)} = 0 \\ E_i \text{ if } 0 < \kappa^{(i)} < \infty \end{cases}$$

where E_i is an exponential random variable with parameter $\frac{W \kappa^{(i)} \eta^{(i)} \lambda}{\eta^{(i^*)} \alpha}$. For $i = 1$, instead we have $U_1 \stackrel{d}{=} +\infty$ if $\kappa^{(1)} = 0$, else U_1 is distributed as an exponential with parameter $\frac{e^{-\lambda t} W \kappa^{(1)} \eta^{(1)} \lambda}{\eta^{(i^*)} \alpha}$. As, up to null events, $\{W > 0\} = S_1$, and using standard results for the ordering of exponential variables (looking at the minimum of the U_i), we have

$$\lim_{\nu^* \to 0} \mathbb{I}_{S_1} \mathbb{P}(T(p) + t < T(\neg p) | (\mathcal{X}(s))_{s \geq 0}) = \mathbb{I}_{S_1} \frac{\kappa^{(1)} \eta^{(1)}}{\kappa^{(1)} \eta^{(1)} + e^{\lambda t} \sum_{j=2}^n \kappa^{(j)} \eta^{(j)}}. \tag{44}$$

The right hand side of (44) is an element of $[0, 1]$. 0 is obtained when $\kappa^{(1)} = 0$ (the final transition rate along the path is small, in the sense of the limit (43), relative to other paths) and 1 is obtained if $\kappa^{(i)} = \delta_{1,i}$ (the final transition rate along the path is the largest relative to other paths). Coupling the above with (42) gives the result. □

5.6 Cyclic graphs

Suppose G is as before, i,e. Section 5.5, but without the acyclic assumption. For cyclic graphs in place of $\mathcal{P}_{1,N}$ we must consider $\mathcal{W}_{1,N}$, which we define to be the set of walks between the root and vertex N. The definition of vertex lineages holds identically for walks. Therefore for $\omega \in \mathcal{W}_{1,N}$ we let

$$T(\omega) = \inf\{t \geq 0 : X(\omega, t) > 0\} \tag{45}$$

where $X(\omega, t)$ is as given in Section 2.3. Here we show that so long as the number of transitions between vertices is finite, which corresponds to finitely many back transitions, we may map a graph containing cycles to an acyclic one and then use the results given above.

The set of walks between the root and the target vertex of length at most $i - 1$ will be denoted

$$\mathcal{W}_{1,N}^{(i)} = \{\omega \in \mathcal{W}_{1,N} : |\omega| \leq i\}.$$

Further for a graph $G = (V, E)$, let the vertex parameters be the birth and death rates associated to the vertices of G and the edge parameters the transition rates associated with the edges.

Proposition 2. *For a given graph $G = (V, E)$ with associated vertex and edge parameters, and $i \geq 1$ which is the upper-length of walks that we consider on G, we can construct an acyclic graph $G' = (V', E')$ (dependent on i), which coupled with appropriate vertex and edge parameters, permits a bijection $g : \mathcal{W}_{1,N}^{(i)} \mapsto \mathcal{P}'_{1,|V'|}$. This bijection possesses the property that for $\omega \in \mathcal{W}_{1,N}^{(i)}$*

$$T(\omega) \stackrel{d}{=} T'(g(\omega))$$

where $T'(g(\omega))$ is defined as in (45) but with the process on G' with its associated parameters.

Before giving the proof we briefly demonstrate this on the graphs shown in Fig. 8. This elucidates the general idea.

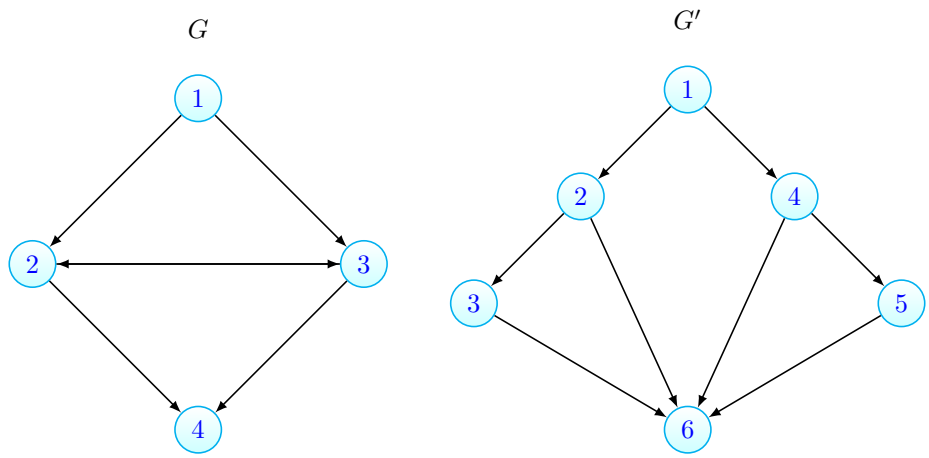


Figure 8

Here

$$\mathcal{W}_{1,N}^{(4)} = \{(1,2,4),(1,2,3,4),(1,3,4),(1,3,2,4)\}.$$

It is clear these may be mapped to the paths in G'

$$\mathcal{P}'_{1,|V'|} = \{(1,2,6),(1,2,3,6),(1,4,6),(1,4,5,6)\}.$$

Further $T((1,2,4)) \stackrel{d}{=} T'((1,2,6))$ so long as the transition rates associated with edges $(1,2)$ and $(2,4)$ on G are equal to the rates with $(1,2)$ and $(2,6)$ on G' and that the birth and death rates at vertices 1 and 2 are equal in both cases.

Proof of Proposition 2. To avoid cumbersome notation and as the idea is illustrated in Fig. 8 we sketch an algorithm. We enumerate the set $\mathcal{W}_{1,N}^{(i)} = \{\omega^{(1)}, \omega^{(2)}, \ldots \omega^{(\mathcal{M})}\}$, where $\mathcal{M} = |\mathcal{W}_{1,N}^{(i)}|$. Again we use the notation $l^{(j)}$ for the length of the jth walk, that is $l^{(j)} = |\omega^{(j)}| - 1$. The graph G' is constructed iteratively. Let $V'_1 = \{1, 2, \ldots, l^{(1)}, N'\}$, where N' is a placeholder and will ultimately be $|V'|$. Similarly $E'_1 = \{(1,2), \ldots, (l^{(1)}, N')\}$ and $p^{(1)} = (1, 2, \ldots, l^{(1)}, N')$.

Then for each further $\omega^{(j)}$, $2 \leq j \leq \mathcal{M}$, from the previously considered walks select the walk that is identical to $\omega^{(j)}$ for the greatest number of vertices, say $\omega^{(k)}$, $1 \leq k < j$. If $\omega^{(j)}$ and $\omega^{(k)}$ agree up until their mth element, and $l^{(j)} - m > 0$ then add $l^{(j)} - m$ elements to V'_{j-1} (new vertices starting with $|V'_{j-1}|$). This creates V'_j. Denote the new vertices as $(v_x^{(j)})_{x=1}^{l^{(j)}-m}$. As $\omega^{(k)}$ has previously been considered it will have an associated path $p^{(k)}$, whose final element we will require to construct E_j: to E_{j-1} add the edges $(p_m^{(k)}, v_1^{(j)})$, $\{(v_x^{(j)}, v_{x+1}^{(j)})\}_{x=1}^{l^{(j)}-m-1}$ and $(v_{l^{(j)}-m}^{(j)}, N')$. This creates E_j. For $p^{(j)}$ we let $p_{1:m}^{(j)} = p_{1:m}^{(k)}$ and then add the new vertices of V'_j, and terminate the path at N'. If $l^{(j)} = m$ (and so the walk $\omega^{(j)}$ agrees with $\omega^{(k)}$

until its penultimate element) let $V'_j = V'_{j-1}$ and add the edge $(p^{(k)}_m, N')$ to E'_{j-1}. Then $p^{(j)}$ is equal to $p^{(k)}_{1:m}$ concatenated with N'. Finally our constructed graph is $G' = (V', E') = (V'_{\mathcal{M}}, E'_{\mathcal{M}})$ with $N' = |V'|$.

Identifying $g(\omega^{(j)}) = p^{(j)}$, we see that if $\omega^{(j)}$, $\omega^{(k)}$ agree until their mth element then so also will $g(\omega^{(j)})$, $g(\omega^{(k)})$, but after this the paths diverge. We further require that the birth and death rates for the population type represented by the kth element of ω should be the same as the kth element of $g(\omega)$. Similarly for the transition rates between the kth and $k+1$th element of ω and $g(\omega)$.

□

Coupling the above proposition with the results of Section 5.4 allows one to characterise

$$\min_{\omega \in \mathcal{W}^{(k)}_{1,N}} T(\omega)$$

in terms of the weights of the walks. Further for a set of walks $\mathcal{W}^{(i)}_{1,N}$, Theorem 8 holds but now with the walk weights, defined analogously to the definition given in (22). Note that the Proposition 2 holds regardless of whether the root is a source vertex, or whether the vertex 1 population is the most fit. However, if transitions are permitted into the root vertex, then for any given walk, there is no guarantee that the cells associated with the first element of that walk are the most fit. This would leave us unable to apply the results derived in earlier sections.

While the above proposition allows us to consider back transitions, the following result demonstrates that when the transition rates between vertices are small, the first initiation of the target is dominated by the paths, as opposed to walks containing cycles. This offers a secondary justification for mainly focusing on acyclic graphs. To state the result let us define $q : \mathcal{W}_{1,N} \mapsto \mathcal{P}_{1,N}$ as the operation reducing walks to their respective paths. This may be accomplished recursively by searching for the first cycle in ω and replacing this with the first element in the cycle. Then repeating the same procedure on the reduced walk until it is a path. This operation is the same as that which maps trajectories to their corresponding trajectory class in [67]. Further, for $p \in \mathcal{P}_{1,N}$ let

$$\mathcal{W}^{(i)}[p] = \{\omega \in \mathcal{W}^{(i)}_{1,N} : q[\omega] = p, \omega \neq p\}$$

that is the walks of length at most $i-1$ which when reduced to paths are p, but excluding p itself. Then with

$$\nu_{\max} = \max\{\nu(i,j) : (i,j) \in E\}$$

we have the following.

Proposition 3. *For $i \geq 1$, $p \in \mathcal{P}_{1,N}$ such that $\mathcal{W}^{(i)}[p]$ is non-empty, and all $\omega \in \mathcal{W}^{(i)}[p]$*

$$\lim_{\nu_{\max} \to 0} \lim_{\nu^* \to 0} \mathbb{P}(T(p) < T(\omega)|S_1) = 1.$$

Proof. We construct an acyclic graph with appropriate vertex and edge parameters as in Proposition 2 and use the mapping g from there with domain $\mathcal{W}^{(i)}_{1,N}$. As before let $\mathcal{P}'_{1,|V'|}$ be the codomain of g. Then for path $p \in \mathcal{W}^{(i)}_{1,N}$ and walk $\omega \in \mathcal{W}^{(i)}[p]$, $g(p)$, $g(\omega) \in \mathcal{P}'_{1,|V'|}$. Note due to the manner in which the edge parameters are chosen $\{\nu(i,N) : i \in \mathcal{N}^-(N)\} = \{\nu(i,|V'|) : i \in \mathcal{N}^-(|V'|)\}$ and hence ν^* remains unchanged. Similarly $\nu_{\max}$ is unchanged. Therefore,

$$\lim_{\nu^* \to 0} \mathbb{P}(T(p) < T(\omega)|S_1) = \lim_{\nu^* \to 0} \mathbb{P}(T'(g(p)) < T'(g(\omega))|S_1).$$

Furthermore each vertex and edge in p is also present in ω. Hence the weight of $g(\omega)$ satisfies $w(g(\omega)) = w(g(p))\tilde{w}(g(\omega))$ where $\tilde{w}(g(\omega))$ is the product of transition rates along the edges existing in cycles in ω divided by the costs of vertices in the cycles. Then by Theorem 2

$$\lim_{\nu^* \to 0} \mathbb{P}(T'(g(p)) < T'(g(\omega))|S_1) = \frac{1}{1 + \tilde{w}(g(\omega))}.$$

As $\tilde{w}(g(\omega))$ contains transition rates, this leads to the stated result upon taking $\nu_{\max} \to 0$. □

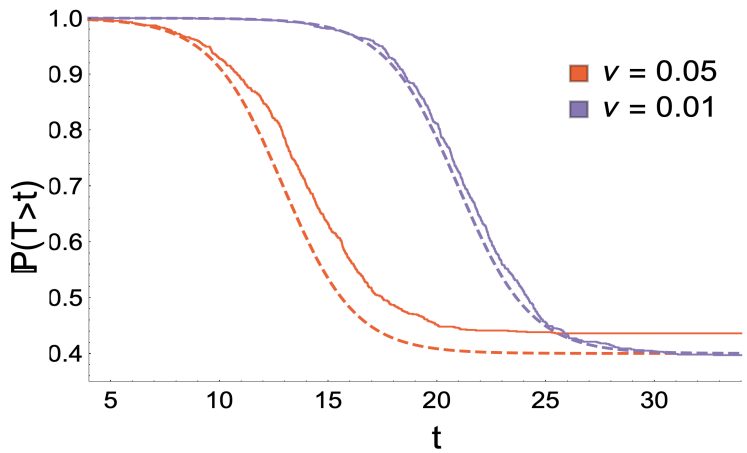


Figure 9: Alternative transition formulation: Here we compare analogous simulations to Fig 3(b), except that transitions follow $(x) \to (y)$, with the theoretical approximation of (5). Dashed lines give the theoretical result, joined lines represents the empirical distribution obtained from 1000 realisations. Despite (5) begin derived under the formulation when transitions follow $(x) \to (x), (y)$, there is excellent agreement between the simulations and theory. Parameters: $\alpha = \alpha(2) = \alpha(3) = \alpha(4) = 1, \beta = 0.4, \beta(2) = \beta(3) = 1.3, \beta(4) = 0.6, z = 1$; legend displays value of $\nu = \nu(1,2) = \nu(2,3) = \nu(3,4)$ used.

5.7 Simulation for alternative transition formulation

See Fig. 9 for a comparison of simulations with transitions following $(x) \to (y)$ with the hitting time results derived in Section 2 of the main text.

5.8 Derivation of select formulas in Section 3.1.2

Here we outline the derivation of the expressions contained in Section 3.1.2, in particular formulas (17)(18)(19). These expressions were derived with the assistance of the Mathemetica software [79].

We start with (17). Let the set of stepwise paths be $\mathcal{P}_{\text{SE}}$ and the set of directed paths be $\mathcal{P}_{\text{DE}}$. For a path $p = (p_1, \dots p_{l+1})$ with number of edges (path length) l, write $c(p) = \prod_{i=2}^{l}(\lambda - \lambda(p_i))$. Note for any path considered here, if we follow the path on Fig. 7a, then $c(p)$ is the product of the vertex labels, taken over all intermediate vertices (excluding the root and the target). Further, let us introduce $r = n_{\text{D}}/n_{\text{S}}$ and $f = (n_{\text{D}} + n_{\text{S}})/n_{\text{Tot}}$. This implies $n_{\text{DD}} = (1-f)n_{\text{Tot}}$, $n_{\text{S}} = f n_{\text{Tot}}/(1+r)$, $n_{\text{D}} = r f n_{\text{Tot}}/(1+r)$. Then

$$\begin{aligned}\phi_N &= \frac{n_{\text{D}}}{n_{\text{D}} + n_{\text{DD}}} \frac{n_{\text{DD}}}{n_{\text{S}} + n_{\text{DD}}} m^2 \nu^2 \sum_{p \in \mathcal{P}_{\text{SE}}} (c(p))^{-1} + \frac{n_{\text{DD}}}{n_{\text{DD}} + n_{\text{D}}} \nu^2 m \sum_{p \in \mathcal{P}_{\text{DE}}} (c(p))^{-1} \\ &= \left[1 + \frac{(1-f)(1+r)}{rf}\right]^{-1} \left[1 + \frac{f}{(1+r)(1-f)}\right]^{-1} m^2 \nu^2 \sum_{p \in \mathcal{P}_{\text{SE}}} (c(p))^{-1} + \\ &\quad \left[1 + \frac{r}{(1+r)(1-f)}\right]^{-1} \nu^2 m \sum_{p \in \mathcal{P}_{\text{DE}}} (c(p))^{-1}.\end{aligned} \tag{46}$$

Using the approximate formula (7), we have $t_{1/2}(r) = -\lambda^{-1} \log(\lambda^{-1} \phi_N z))$. Also recall $z = \gamma n_s = \gamma f n_{\text{Tot}}/(1+r)$. Differentiating $t_{1/2}(r)$ and evaluating at $r = 0$ leads to

$$\frac{d}{dr} t_{1/2}(0) = \frac{\frac{1}{1-f} - \frac{fm(d+s)^3}{ds(d^2+3ds+s^2)}}{\lambda}.$$

Increasing the relative size of the single drug compartment speeds up resistance when $\frac{d}{dr}t_{1/2}(0) < 0$ or equivalently when

$$\frac{ds(d^2+3ds+s^2)}{1-f} > fm(d+s)^2.$$

Expansions in f and s/d (recall $f \ll 1, s/d \ll 1$), keeping only leading order terms and using the definition of f leads to (17).

Turning to (18), we wish to find r such that $\frac{d}{dr}t_{1/2}(r) = 0$. The numerator of $\frac{d}{dr}t_{1/2}(r)$ is quadratic in r. After solving for both roots and expanding in powers of s/d, we see the leading term for one of the roots is -1. Thus the root of interest is the other root, which to leading order is the expression in (18).

For (19), we use our main result (11). With ϕ_N as given above in (46), this yields

$$\begin{aligned}\mathbb{P}(T^{\mathrm{SE}} < T^{\mathrm{DE}}) &\approx \frac{\frac{n_{\mathrm{D}}}{n_{\mathrm{D}}+n_{\mathrm{DD}}}\frac{n_{\mathrm{DD}}}{n_{\mathrm{S}}+n_{\mathrm{DD}}}m^2\nu^2\sum_{p\in\mathcal{P}_{\mathrm{SE}}}(c(p))^{-1}}{\phi_N} \\ &= \left(1+\frac{(n_{\mathrm{DD}}+n_{\mathrm{S}})}{n_{\mathrm{D}}m}\frac{d^3s+3s^2d^2+ds^3}{(d+s)^3}\right)^{-1}.\end{aligned}$$

Upon expanding in s/d, to leading order we have (19).

References

[1] H. K. Alexander, *Conditional distributions and waiting times in multitype branching processes*, Advances in Applied Probability, 45 (2013), p. 692–718.

[2] D. I. Andersson and D. Hughes, *Antibiotic resistance and its cost: is it possible to reverse resistance?*, Nature Reviews Microbiology, (2010).

[3] T. Antal and P. L. Krapivsky, *Exact solution of a two-type branching process: models of tumor progression*, Journal of Statistical Mechanics: Theory and Experiment, 2011 (2011), p. P08018.

[4] K. B. Athreya and P. E. Ney, *Branching Processes*, Dover Publications, 2004.

[5] M. Bartlett, *An Introduction to Stochastic Processes*, Cambridge University Press, 1955.

[6] B. Bauer and C. S. Gokhale, *Repeatability of evolution on epistatic landscapes*, Scientific Reports, 5 (2015), pp. 9607 EP –.

[7] N. Beerenwinkel, T. Antal, D. Dingli, A. Traulsen, K. W. Kinzler, V. E. Velculescu, B. Vogelstein, and M. A. Nowak, *Genetic progression and the waiting time to cancer*, PLOS Computational Biology, 3 (2007), pp. 1–8.

[8] J. Berestycki, E. Brunet, and Z. Shi, *The number of accessible paths in the hypercube*, Bernoulli, 22 (2016), pp. 653–680.

[9] D. Bixby and M. Talpaz, *Seeking the causes and solutions to imatinib-resistance in chronic myeloid leukemia*, Leukemia, 25 (2010), p. 7.

[10] A. Bovier, L. Coquille, and C. Smadi, *Crossing a fitness valley as a metastable transition in a stochastic population model*, arXiv e-prints, (2018), p. arXiv:1801.06473.

[11] I. Bozic, T. Antal, H. Ohtsuki, H. Carter, D. Kim, S. Chen, R. Karchin, K. W. Kinzler, B. Vogelstein, and M. A. Nowak, *Accumulation of driver and passenger mutations during tumor progression*, Proceedings of the National Academy of Sciences, 107 (2010), pp. 18545–18550.

[12] I. Bozic, J. M. Gerold, and M. A. Nowak, *Quantifying clonal and subclonal passenger mutations in cancer evolution*, PLOS Computational Biology, 12 (2016), pp. 1–19.

[13] I. Bozic and M. A. Nowak, *Resisting resistance*, Annual Review of Cancer Biology, 1 (2017), pp. 203–221.

[14] I. Bozic, J. G. Reiter, B. Allen, T. Antal, K. Chatterjee, P. Shah, Y. S. Moon, A. Yaqubie, N. Kelly, D. T. Le, E. J. Lipson, P. B. Chapman, J. Diaz, Luis A, B. Vogelstein, and M. A. Nowak, *Evolutionary dynamics of cancer in response to targeted combination therapy*, eLife, 2 (2013), p. e00747.

[15] L. Chaumont and T. N. A. Nguyen, *On mutations in the branching model for multitype populations*, Advances in Applied Probability, 50 (2018), p. 543–564.

[16] D. Cheek and T. Antal, *Mutation frequencies in a birth–death branching process*, Ann. Appl. Probab., 28 (2018), pp. 3922–3947.

[17] T. Cihlar and M. Fordyce, *Current status and prospects of hiv treatment*, Current Opinion in Virology, 18 (2016), pp. 50 – 56. Antiviral strategies • Virus structure and expression.

[18] C. Colijn, T. Cohen, A. Ganesh, and M. Murray, *Spontaneous emergence of multiple drug resistance in tuberculosis before and during therapy*, PLOS ONE, 6 (2011), pp. 1–7.

[19] T. H. Cormen, C. E. Leiserson, R. L. Rivest, and C. Stein, *Introduction to Algorithms, Third Edition*, vol. 7, The MIT Press, Cambridge, Massachusetts., third ed., 2009.

[20] J. A. G. M. de Visser and J. Krug, *Empirical fitness landscapes and the predictability of evolution*, Nature Reviews Genetics, 15 (2014), p. 480.

[21] J. Denes and D. Krewski, *An exact representation for the generating function for the moolgavkar-venzon-knudson two-stage model of carcinogenesis with stochastic stem cell growth*, Mathematical Biosciences, 131 (1996), pp. 185 – 204.

[22] S. Deresinski, *Vancomycin in Combination with Other Antibiotics for the Treatment of Serious Methicillin-Resistant Staphylococcus aureus Infections*, Clinical Infectious Diseases, 49 (2009), pp. 1072–1079.

[23] A. Dewanji, J. Jeon, R. Mexa, and E. G. Luebeck, *Number and size distribution of colorectal adenomas under the multistage clonal expansion model of cancer*, PLoS Computational Biology, 7 (2011).

[24] R. Durrett, *Branching Process Models of Cancer*, Stochastics in Biological Systems, Springer, 2015.

[25] R. Durrett, J. Foo, K. Leder, J. Mayberry, and F. Michor, *Intratumor heterogeneity in evolutionary models of tumor progression*, Genetics, 188 (2011), pp. 461–477.

[26] R. Durrett and S. Moseley, *Evolution of resistance and progression to disease during clonal expansion of cancer*, Theoretical Population Biology, 77 (2010), pp. 42–48.

[27] R. Durrett, D. Schmidt, and J. Scweinsberg, *A waiting time problem arising from the study of mutli-stage carcinogenesis*, The Annals of Applied Probability, 19 (2009), pp. 676–718.

[28] P. M. Enriquez-Navas, Y. Kam, T. Das, S. Hassan, A. Silva, P. Foroutan, E. Ruiz, G. Martinez, S. Minton, R. J. Gillies, and R. A. Gatenby, *Exploiting evolutionary principles to prolong tumor control in preclinical models of breast cancer*, Science Translational Medicine, 8 (2016), pp. 327ra24–327ra24.

[29] C. B. Ford, R. R. Shah, M. K. Maeda, S. Gagneux, M. B. Murray, T. Cohen, J. C. Johnston, J. Gardy, M. Lipsitch, and S. M. Fortune, *Mycobacterium tuberculosis mutation rate estimates from different lineages predict substantial differences in the emergence of drug-resistant tuberculosis*, Nature Genetics, 45 (2013), pp. 784–790.

[30] F. Fu, M. A. Nowak, and S. Bonhoeffer, *Spatial Heterogeneity in Drug Concentrations Can Facilitate the Emergence of Resistance to Cancer Therapy*, PLoS Computational Biology, 11 (2015).

[31] R. Gao, A. Davis, T. O. McDonald, E. Sei, X. Shi, Y. Wang, P.-C. Tsai, A. Casasent, J. Waters, H. Zhang, F. Meric-Bernstam, F. Michor, and N. E. Navin, *Punctuated copy number evolution and clonal stasis in triple-negative breast cancer*, Nature Genetics, 48 (2016), p. 1119.

[32] J. H. Gillespie, *A simple stochastic gene substitution model*, Theoretical Population Biology, 23 (1983), pp. 202 – 215.

[33] C. S. Gokhale, Y. Iwasa, M. A. Nowak, and A. Traulsen, *The pace of evolution across fitness valleys*, Journal of Theoretical Biology, 259 (2009), pp. 613–620.

[34] P. Greulich, B. Waclaw, and R. J. Allen, *Mutational pathway determines whether drug gradients accelerate evolution of drug-resistant cells*, Phys. Rev. Lett., 109 (2012), p. 088101.

[35] H. Haeno, M. Gonen, M. Davis, J. Herman, C. Iacobuzio-Donahue, and F. Michor, *Computational modeling of pancreatic cancer reveals kinetics of metastasis suggesting optimum treatment strategies*, Cell, 148 (2012), pp. 362 – 375.

[36] H. Haeno, Y. Iwasa, and F. Michor, *The evolution of two mutations during clonal expansion*, Genetics, 177 (2007), pp. 2209–2221.

[37] A. R. Hall, J. C. Iles, and R. C. MacLean, *The fitness cost of rifampicin resistance in Pseudomonas aeruginosa depends on demand for RNA polymerase*, Genetics, 187 (2011), pp. 817–822.

[38] R. Hermsen, J. B. Deris, and T. Hwa, *On the rapidity of antibiotic resistance evolution facilitated by a concentration gradient*, Proceedings of the National Academy of Sciences, 109 (2012), pp. 10775–10780.

[39] R. Hermsen and T. Hwa, *Sources and sinks: A stochastic model of evolution in heterogeneous environments*, Phys. Rev. Lett., 105 (2010), p. 248104.

[40] D. Hughes and D. I. Andersson, *Evolutionary consequences of drug resistance: Shared principles across diverse targets and organisms*, Nature Reviews Genetics, 16 (2015), pp. 459–471.

[41] Y. Iwasa, F. Michor, and M. A. Nowak, *Evolutionary dynamics of escape from biomedical intervention*, Proceedings of the Royal Society of London. Series B: Biological Sciences, 270 (2003), pp. 2573–2578.

[42] Y. Iwasa, F. Michor, and M. A. Nowak, *Evolutionary dynamics of invasion and escape*, Journal of Theoretical Biology, 226 (2004), pp. 205–214.

[43] Y. Iwasa, M. A. Nowak, and F. Michor, *Evolution of resistance during clonal expansion*, Genetics, 172 (2006), pp. 2557–2566.

[44] S. Janson, *Functional limit theorems for multitype branching processes and generalized Pólya urns*, Stochastic Processes and their Applications, 110 (2004), pp. 177–245.

[45] P. Keller and T. Antal, *Mutant number distribution in an exponentially growing population*, J. Stat. Mech. P01011, (2015).

[46] D. A. Kessler, R. H. Austin, and H. Levine, *Resistance to chemotherapy: patient variability and cellular heterogeneity.*, Cancer research, 74 (2014), pp. 4663–70.

[47] D. A. Kessler and H. Levine, *Large population solution of the stochastic {L}uria-{D}elbrück evolution model*, Proceedings of the national Academy of Science USA, 110 (2013), pp. 11682–11687.

[48] N. L. Komarova and D. Wodarz, *Drug resistance in cancer: Principles of emergence and prevention*, Proceedings of the National Academy of Sciences, 102 (2005), pp. 9714–9719.

[49] D. E. Lea and C. A. Coulson, *The distribution of the numbers of mutants in bacterial populations*, Journal of Genetics, 49 (1949), pp. 264–285.

[50] K. Leder, J. Foo, B. Skaggs, M. Gorre, C. L. Sawyers, and F. Michor, *Fitness conferred by bcr-abl kinase domain mutations determines the risk of pre-existing resistance in chronic myeloid leukemia*, PLOS ONE, 6 (2011), pp. 1–11.

[51] H. Lee, E. Popodi, H. Tang, and P. L. Foster, *Rate and molecular spectrum of spontaneous mutations in the bacterium escherichia coli as determined by whole-genome sequencing*, Proceedings of the National Academy of Sciences, 109 (2012), pp. E2774–E2783.

[52] S. E. Luria and M. Delbrück, *Mutations of bacteria from virus sensitivity to virus resistance*, Genetics, 48 (1943), pp. 491–511.

[53] W. T. Ma, G. V. Sandri, and S. Sarkar, *Analysis of the {L}uria-{D}elbrück Distribution using discrete convolution powers*, Journal of Applied Probability, 29 (1992), pp. 255–267.

[54] B. Mandelbrot, *A Population Birth-and-Mutation Process, I: Explicit Distributions for the Number of Mutants in an Old Culture of Bacteria*, Journal of Applied Probability, 11 (1974), pp. pp. 437–444.

[55] F. Michor, T. P. Hughes, Y. Iwasa, S. Branford, N. P. Shah, C. L. Sawyers, and M. A. Nowak, *Dynamics of chronic myeloid leukaemia.*, Nature, 435 (2005), pp. 1267–1270.

[56] M. Möhle, *Convergence Results for Compound Poisson Distributions and Applications to the Standard Luria-Delbrück Distribution*, Journal of Applied Probability, 42 (2005), pp. pp. 620–631.

[57] S. Moreno-Gamez, A. L. Hill, D. I. S. Rosenbloom, D. A. Petrov, M. A. Nowak, and P. S. Pennings, *Imperfect drug penetration leads to spatial monotherapy and rapid evolution of multidrug resistance*, Proceedings of the National Academy of Sciences, 112 (2015), pp. E2874–E2883.

[58] D. Nichol, P. Jeavons, A. G. Fletcher, R. A. Bonomo, P. K. Maini, J. L. Paul, R. A. Gatenby, A. R. Anderson, and J. G. Scott, *Steering evolution with sequential therapy to prevent the emergence of bacterial antibiotic resistance*, PLOS Computational Biology, 11 (2015), pp. 1–19.

[59] M. D. Nicholson and T. Antal, *Universal asymptotic clone size distribution for general population growth*, Bull Math Biol, 78 (2016), p. 2243–2276.

[60] M. A. Nowak, N. L. Komarova, A. Sengupta, P. V. Jallepalli, I.-M. Shih, B. Vogelstein, and C. Lengauer, *The role of chromosomal instability in tumor initiation*, Proceedings of the National Academy of Sciences, 99 (2002), pp. 16226–16231.

[61] J. E. Pearson, P. Krapivsky, and A. S. Perelson, *Stochastic theory of early viral infection: Continuous versus burst production of virions*, PLOS Computational Biology, 7 (2011), pp. 1–17.

[62] A. J. Primeau, A. Rendon, D. Hedley, L. Lilge, and I. F. Tannock, *The distribution of the anticancer drug doxorubicin in relation to blood vessels in solid tumors*, Clinical Cancer Research, 11 (2005), pp. 8782–8788.

[63] J. G. Reiter, A. P. Makohon-Moore, J. M. Gerold, A. Heyde, M. A. Attiyeh, Z. A. Kohutek, C. J. Tokheim, A. Brown, R. M. DeBlasio, J. Niyazov, A. Zucker, R. Karchin, K. W. Kinzler, C. A. Iacobuzio-Donahue, B. Vogelstein, and M. A. Nowak, *Minimal functional driver gene heterogeneity among untreated metastases*, Science, 361 (2018), pp. 1033–1037.

[64] W. A. Rosche and P. L. Foster, *Determining mutation rates in bacterial populations*, Methods, 20 (2000), pp. 4–17.

[65] S. Sagitov and M. Conceição Serra, *Multitype bienaymé–galton–watson processes escaping extinction*, Advances in Applied Probability, 41 (2009).

[66] M. C. Serra and P. Haccou, *Dynamics of escape mutants*, Theoretical Population Biology, 72 (2007), pp. 167 – 178.

[67] A. C. Tadrowski, M. R. Evans, and B. Waclaw, *Phenotypic Switching Can Speed up Microbial Evolution*, Scientific Reports, 8 (2018), p. 8941.

[68] P. D. Tamma, S. E. Cosgrove, and L. L. Maragakis, *Combination therapy for treatment of infections with gram-negative bacteria*, Clinical Microbiology Reviews, 25 (2012), pp. 450–470.

[69] A. J. Tipping, F. X. Manon, V. Lagarde, J. M. Goldman, and J. V. Melo, *Restoration of sensitivity to STI571 in STI571-resistant chronic myeloid leukemia cells*, Blood, 98 (2001), pp. 3864–3867.

[70] T. D. Tlsty, B. H. Margolin, and K. Lum, *Differences in the rates of gene amplification in nontumorigenic and tumorigenic cell lines as measured by Luria-Delbruck fluctuation analysis*, Proc Natl Acad Sci U S A, 86 (1989), pp. 9441–9445.

[71] C. Tomasetti, B. Vogelstein, and G. Parmigiani, *Half or more of the somatic mutations in cancers of self-renewing tissues originate prior to tumor initiation*, Proceedings of the National Academy of Sciences, (2013).

[72] R. S. Varga, *Matrix Iterative Analysis*, Prentice-Hall International, 1962.

[73] H. Ward, G. G. Perron, and R. C. MacLean, *The cost of multiple drug resistance in Pseudomonas aeruginosa*, Journal of Evolutionary Biology, 22 (2009), pp. 997–1003.

[74] D. F. Warner and V. Mizrahi, *Tuberculosis chemotherapy: The influence of bacillary stress and damage response pathways on drug efficacy*, 2006.

[75] D. M. Weinreich, N. F. Delaney, M. A. DePristo, and D. L. Hartl, *Darwinian evolution can follow only very few mutational paths to fitter proteins*, Science, 312 (2006), pp. 111–114.

[76] D. B. Weissman, M. M. Desai, D. S. Fisher, and M. W. Feldman, *The rate at which asexual populations cross fitness valleys*, Theoretical Population Biology, 75 (2009), pp. 286 – 300. Sam Karlin: Special Issue.

[77] R. D. Wilkinson and S. Tavaré, *Estimating primate divergence times by using conditioned birth-and-death processes*, Theoretical Population Biology, 75 (2009), pp. 278 – 285. Sam Karlin: Special Issue.

[78] M. J. Williams, B. Werner, C. P. Barnes, T. A. Graham, and A. Sottoriva, *Identification of neutral tumor evolution across cancer types*, Nature Genetics, 48 (2016), pp. 238–244.

[79] Wolfram Research, Inc., *Mathematica, Version 11.3*. Champaign, IL, 2018.

[80] M. Zagorski, Z. Burda, and B. Waclaw, *Beyond the hypercube: Evolutionary accessibility of fitness landscapes with realistic mutational networks*, PLOS Computational Biology, 12 (2016), pp. 1–18.