Author for correspondence:
Susanna Manrubia
e-mail: smanrubia@cnb.csic.es

Evolution on neutral networks accelerates the ticking rate of the molecular clock

Susanna Manrubia[1,2] and José A. Cuesta[1,3,4]

[1]Grupo Interdisciplinar de Sistemas Complejos (GISC), Madrid, Spain
[2]Systems Biology Programm, National Centre for Biotechnology (CSIC), c/ Darwin 3, 28049 Madrid, Spain
[3]Department of Mathematics, Universidad Carlos III de Madrid, 28911 Leganés, Madrid, Spain
[4]Instituto de Biocomputación y Física de Sistemas Complejos (BIFI), Universidad de Zaragoza, 50009 Zaragoza, Spain

Large sets of genotypes give rise to the same phenotype, because phenotypic expression is highly redundant. Accordingly, a population can accept mutations without altering its phenotype, as long as the genotype mutates into another one on the same set. By linking every pair of genotypes that are mutually accessible through mutation, genotypes organize themselves into neutral networks (NNs). These networks are known to be heterogeneous and assortative, and these properties affect the evolutionary dynamics of the population. By studying the dynamics of populations on NNs with arbitrary topology, we analyse the effect of assortativity, of NN (phenotype) fitness and of network size. We find that the probability that the population leaves the network is smaller the longer the time spent on it. This progressive 'phenotypic entrapment' entails a systematic increase in the overdispersion of the process with time and an acceleration in the fixation rate of neutral mutations. We also quantify the variation of these effects with the size of the phenotype and with its fitness relative to that of neighbouring alternatives.

1. Introduction

The relationship between genotype and phenotype, the ways in which this map conditions the adaptive dynamics of populations, or the imprints that life histories leave in the genomes of organisms are essential questions to be solved before a complete evolutionary theory can be achieved. Genotypes, which encode much of the information required to construct organisms, are occasionally affected by mutations that modify the phenotype, their visible expression and the target of natural selection. Many mutations are neutral instead [1], varying the regions of the space of genotypes that can be accessed by the population [2] and conditioning its evolvability [3] but leaving the phenotype unchanged. The relation between genotype and phenotype is not one-to-one, but many-to-many. In particular, genotypes encoding a specific phenotype may form vast, connected networks that often span the whole space of possible genotypes. The existence of these networks in the case of proteins was postulated by Maynard Smith [4] as a requirement for evolution by natural selection to occur. Subsequent research has shown that these networks do exist for functional proteins [5], for other macromolecules like RNA [6], and generically appear in simple models of the genotype–phenotype map-mimicking regulatory gene networks [7], metabolic reaction networks [8] or the self-assembly of protein quaternary structure [9].

Nevertheless, systematic explorations of the topological structure of neutral networks (NNs) have been undertaken only recently, despite the fact that some of the implications of NN structure on sequence evolution were identified long ago. For instance, Kimura's neutral theory [1] postulated that the number of neutral substitutions in a given time interval was Poisson distributed. That assumption had an underlying hypothesis that was not explicitly stated at the time, namely that the number of neutral mutations available to any genotype was constant, independent of the precise genotype, of time or of the expressed phenotype. In other words, NNs were assumed to be homogeneous in degree. A consequence was that the variance of the number of mutations

accumulated should equal the mean, and the dispersion index (R, the ratio between the variance and the mean) must then be equal to 1. Very early, however, it was observed that R was significantly larger than 1 in almost all cases analysed [10–12]. The appearance of short bursts of rapid evolution was ascribed to episodes of positive Darwinian selection [12] that may reflect fluctuations in population size in quasi-neutral environments, where epistatic interactions would become relevant [13].

The fact is that NNs are highly heterogeneous. Some genotypes are brittle and easily yield a different phenotype under mutation. They have one or few neighbours within the NN. Other genotypes, instead, are robust and can stand a very large number of mutations while maintaining their biological or chemical function. The existence of variations in the degree of neutrality of genotypes (in their robustness) was soon put forth as a possible explanation for the overdispersion of the molecular clock [13]. Nowadays, the distributions of robustness of the genotypes in several different NNs have been measured, turning out to be remarkably broad [14–16]. The effects of fluctuating neutral spaces in the overdispersion of the molecular clock have been investigated in realistic models of evolution for proteins [17] and quasi-species [18], and also from a theoretical viewpoint [19].

In this contribution, we explore three features of NNs whose consequences in the rate of fixation of mutations have not been systematically investigated. They are (i) the correlations in neutrality between neighbouring genotypes, (ii) the degree of redundancy of phenotypes (the size of NN), and (iii) the fitness of the current phenotype in relation to accessible alternatives. First, the degree of neutrality is not randomly distributed in the NN. Thermodynamical arguments [20,21] and analyses of full NNs [22] indicate that genotypes tend to cluster as a function of their robustness, implying that NNs belong to the class of the so-called assortative networks. It is known that populations evolving on NNs tend to occupy maximally connected regions in order to minimize the number of mutations changing their phenotype [2]. In assortative networks, neutral drift entails a canalization towards mutation–selection equilibrium, progressively increasing the rate of fixation of neutral mutations through a dynamical process that we dub *phenotypic entrapment*. Second, there is abundant evidence coming from computational genotype–phenotype maps [9,23–24] and from empirical reconstruction of NNs [25] that the average robustness of a given phenotype grows with the size of its associated NN: here, we quantify the effect of a systematic difference in average degree on the probability of fixation of neutral mutations. Third, the difference in fitness between genotypes in the current NN and their mutational neighbours affects the probability that a mutation (be it neutral, beneficial or deleterious) gets fixed in the population, and with it, as we explicitly show, the rate of the molecular clock during intervals of strictly neutral drift.

With the former goals in mind, we develop an out-of-equilibrium formal framework to describe the dynamics of homogeneous, infinite populations on generic NNs. We demonstrate that the population keeps memory of its past history because, as time elapses, the likelihood that it visits genotypes of increasingly higher robustness augments in a precise way that we calculate. This is a consequence of assortativity and a dynamic manifestation of the 'friendship paradox' [26] described in social networks (your friends have more friends than you). As a result, the probability that the population leaves the network explicitly depends on the elapsed time. Further, the decline of this probability with time entails a systematic acceleration of the rate of accumulation of neutral mutations. The degree of entrapment is higher, the larger the NN and the broader the difference between the fitness of the current phenotype and that of accessible alternatives. These results are fairly general and have implications in the derivation of effective models of phenotypic change and in the calibration of molecular clocks.

2. Dynamical model

2.1. Description of neutral networks and populations

In the following sections, we employ terms related to dynamics on complex networks to describe the dynamical process undergone by a homogeneous population on an NN. Genotypes are the nodes of the network, and two nodes are linked if the corresponding genotypes can be mutually accessed through a single mutation. Links are all identical, so we assume that the mutational process affects all genotypes in the same way. The robustness of a node is a quantity proportional to its degree k, that is, to the number of links leading to other nodes within the NN (or the number of neutral mutations).

For the sake of analytical tractability, we use a mean-field-like description of the NN. This amounts to assuming that all nodes of the same degree (robustness) are equivalent with respect to the dynamics, and therefore the network can be characterized only through its degree distribution, $p(k)$ (the probability that a node has degree k or the distribution of robustness values), and its nearest-neighbour degree distribution $p_{nn}(k|k')$ (the probability that a degree k' node's neighbour has degree k), where $k, k' = 1, \ldots, z$. We are assuming that there is a maximum degree z in the NN. This is always true if genotypes are sequences whose loci are taken from a finite alphabet (e.g. $\{0, 1\}$, nucleotides or amino acids). If there are A different types in the alphabet, sequences have length L, and mutations refer just to point mutations (the most widely studied mutational process), then $z = (A - 1)L$. Typically, thus, $z \gg 1$.

The strength of correlation between the degree of neighbouring nodes is measured through the *degree–degree correlation coefficient* r. This is a quantity that characterizes the assortativity of a network and is defined as the Pearson correlation coefficient (cf. the covariance) between the degree of nodes and the degree of their neighbours. For a network with $i = 1, \ldots, M$ nodes, each with degree k_i and whose neighbours have on average degree k_i^{nn}, r is defined as

$$r = \frac{\sum_{i=1}^{M}(k_i - \overline{k})(k_i^{nn} - \overline{k^{nn}})}{\sqrt{\sum_{i=1}^{M}(k_i - \overline{k})^2}\sqrt{\sum_{i=1}^{M}(k_i^{nn} - \overline{k^{nn}})^2}}, \tag{2.1}$$

with $\overline{k} = M^{-1}\sum_i k_i$, $\overline{k^{nn}} = M^{-1}\sum_i k_i^{nn}$. The largest value of assortativity $r = 1$ is achieved when the average degree of neighbours of node i equals its degree, $k_i^{nn} = k_i$. Anticorrelations in degree yield negative values of r. By definition, $-1 \leq r \leq 1$.

We assume that all genotypes in an NN have the same fitness, so evolution is strictly neutral within the network. We should only take into account non-neutral mutations that get fixed, that is, that move the population from a node in

its current NN to a node of a different NN. As of neutral evolution, population dynamics distinguishes two basic regimes. If μ denotes the mutation rate and N_e the (effective [27]) population size, these correspond to low ($\mu N_e \ll 1$) and high ($\mu N_e \gg 1$) mutation rates. For high mutation rates, populations are heterogeneous, and we need an ensemble of individuals coupled through replication and mutation to account for their evolution. This is a complex situation where quasi-species dynamics comes into play, and one has to resort to computational simulations to describe its dynamics [18]. In this study, we work in the limit $\mu N_e \ll 1$, where every mutation either gets fixed in or disappears from the population before a new mutation occurs, so the population is always homogeneous and its evolution is effectively described as a single *random walk* (RW) on the NN.

2.2. Dynamics

We describe the dynamical process that begins with the arrival of a random walker to the NN and ends when it leaves to a neighbouring NN. In particular, we aim at describing the entrapment effect that occurs while the walker remains within the NN. Let us mention, at this point, a fact that will become clear later on: because of frequent jumps in and out of NNs, populations will hardly spend a long enough time within a single NN to reach the equilibrium distribution. Therefore, the relevant process to describe is the transient dynamics that starts with arrival at the NN, instead of the asymptotic behaviour within the phenotype. For this reason, we require a good estimation of the initial probability distribution of arrival at any node of the NN.

The process starts as soon as the walker jumps to a node of the NN from outside. We assume that the initial jump is equally likely to occur through any link of the NN pointing outwards, so the probability that a node of degree k is chosen as the entry node should be proportional to the number of outbound links $z - k$ and to the abundance of nodes of degree k. This yields the initial condition

$$P_0(0) = 0, \quad P_k(0) = \frac{z-k}{z - \langle k \rangle} p(k), \quad k = 1, \ldots, z. \quad (2.2)$$

The state $i = 0$ does not describe any node, but it proves convenient to introduce it to represent the exterior of the NN. Hence, $P_0(t)$ is the probability that the process ends at any time $t > 0$; the choice $P_0(0) = 0$ means that the process begins as soon as it 'leaves' the exterior.

Because the mean-field description assumes that all nodes of the same degree are equivalent, the RW on the NN can be further described as a stochastic process $X(t)$ taking values in the set $\{0, 1, \ldots, z\}$. These values represent the degree of the node the walker is in at time t, with state 0 corresponding to the exterior of the NN, as just described. When the walker jumps to state 0 the phenotype changes, and other considerations are needed to follow up the evolution. For our purposes, the process ends at this point, so 0 is an *absorbing state* (the only one in this process).

The stochastic process just introduced is a continuous-time Markov chain with transition rate

$$\Pr\{X(t + \delta t) = k | X(t) = k'\} = \mu_k W_{kk'} \delta t + o(\delta t). \quad (2.3)$$

Element $W_{kk'}$ of matrix $\mathbf{W}$ is the conditional probability that, given that a jump occurs when the process is at a node of degree k', it ends up at a node of degree k. Within our mean-field description, this probability can be simply obtained as

$$W_{kk'} = \frac{k'}{z} p_{nn}(k|k'), \quad k, k' > 0, \quad W_{0k'} = 1 - \frac{k'}{z}, \quad (2.4)$$
$$k' > 0, \quad W_{00} = 0.$$

As mentioned above, the rationale for this choice is that $z - k'$ links of the node where the RW is located point outside the NN and k' point inside, and if the process stays within the NN (probability k'/z), the probability that the node it jumps to has degree k is $p_{nn}(k|k')$. The nearest-neighbour degree distribution $p_{nn}(k|k')$ is zero if $p(k') = 0$ and is normalized otherwise; therefore,

$$\sum_{k=0}^{z} W_{kk'} = 1, \quad \text{for all } k' \text{ such that } p(k') > 0. \quad (2.5)$$

2.3. Leaving the neutral networks

Let us now discuss the meaning of the transition rates μ_k. If $k > 0$, the RW jumps to a node that belongs to the NN, and thus the mutation is neutral. Hence, $\mu_k = \mu$, because, in that case, the mutation rate is also the rate of fixation of neutral mutations in a population of arbitrary size [28]. On the other hand, μ_0 implies a jump outside the NN, hence a phenotypic change. The new phenotype will have, in general, a different fitness, and the rate at which non-neutral mutations go to fixation in a population of effective size N_e is $\mu_0 = \mu N_e (f - 1)/(f^{N_e} - 1)$, where f is the fitness of the current phenotype relative to that of the phenotype the process jumps to [27]. The ratio $\phi \equiv \mu_0/\mu$ can take any value in the interval $0 \leq \phi \leq N_e$. The lower bound occurs, if purifying selection eliminates any phenotype other than the current one; the upper bound is the rate at which the population adopts a highly fitter new phenotype. Neutral changes of phenotype correspond to $\phi = 1$.

There is a simplification implicit in the expression of ϕ, though. If the neighbouring NNs are diverse in fitness values, ϕ should be replaced by an appropriate average over this diversity—weighted with the probabilities of jumping to these other NNs. In any case some factor, $0 < \phi < N_e$ accounts for the differences in the fitness between the current NN and those of its neighbouring NNs. For the discussion to come, we do not need to be more specific about its precise form.

2.4. Master equation

The probability distribution of the process is $P_k(t) \equiv \Pr\{X(t) = k\}$, which stands for the probability that at time t the process sits on a node of degree $k = 1, \ldots, z$, or is outside the NN, if $k = 0$. This probability satisfies the master equation

$$\dot{P}_k = \sum_{j=0}^{z} [\mu_k W_{kj} P_j - \mu_j W_{jk} P_k], \quad k = 0, 1, \ldots, z. \quad (2.6)$$

A few considerations transform (2.6) into a handier expression for further calculations. To begin with, we can set $\mu = 1$, i.e. measure time in units of μ^{-1}. Let us also introduce the *support* of $p(k)$, i.e. the set $\mathcal{S} \equiv \{k : 1 \leq k \leq z, \ p(k) > 0\}$. Thus, any $k \notin \mathcal{S}$ has $p(k) = 0$, $p_{nn}(j|k) = 0$, and, given definition (2.2), $P_k(0) = 0$. Therefore, according to (2.4), for any $k \notin \mathcal{S}$, the master equation (2.6) becomes $\dot{P}_k = -[\phi + (1 - \phi)k/z]P_k$, whose solution is $P_k(t) = P_k(0)e^{-[\phi+(1-\phi)k/z]t} = 0$.

Because of this result, we can ignore in (2.6) any index $k \notin \mathcal{S}$, that is, those degrees that are not represented by any

node in the NN. Accordingly, if we introduce vector $\mathbf{P} \equiv (P_k)^\top$, $k \in \mathcal{S}$ and matrices

$$\mathbf{F} = \mathrm{diag}\left(\phi + (1-\phi)\frac{k}{z}\right), \quad \mathbf{A} = \left(p_{\mathrm{nn}}(k|j)\frac{j}{z}\right), \quad k, j \in \mathcal{S}, \tag{2.7}$$

the master equation (2.6) simply becomes

$$\dot{\mathbf{P}} = -(\mathbf{F} - \mathbf{A})\mathbf{P}. \tag{2.8}$$

3. Results

3.1. Time spent on the neutral networks

An important quantity characterizing the process is the probability that, after a time t, the population is still on the NN. This probability can be calculated as $Q(t) \equiv 1 - P_0(t) = \mathbf{1} \cdot \mathbf{P}(t)$, where $\mathbf{1} = (1, \ldots, 1)^\top$. For $k = 0$, equation (2.6) yields $\dot{Q} = \phi[\boldsymbol{\kappa} \cdot \mathbf{P}(t) - Q]$, where $\boldsymbol{\kappa} = (k/z)^\top$, $k \in \mathcal{S}$, a differential equation that can be rewritten as the integral equation

$$Q(t) = \mathrm{e}^{-\phi t} + \phi \int_0^t \mathrm{e}^{-\phi(t-\tau)} \boldsymbol{\kappa} \cdot \mathbf{P}(\tau)\, \mathrm{d}\tau. \tag{3.1}$$

In appendix A, we prove that the eigenvalues α_k of matrix $\mathbf{F} - \mathbf{A}$ are all real and different. Therefore, denoting $\mathbf{u}_k$ and $\mathbf{v}_k$ the corresponding right and left eigenvectors, respectively, we can write

$$Q(t) = \sum_{k \in \mathcal{S}} q_k \mathrm{e}^{-\alpha_k t}, \quad q_k \equiv \frac{[\mathbf{v}_k \cdot \mathbf{P}(0)](\mathbf{u}_k \cdot \mathbf{1})}{(\mathbf{v}_k \cdot \mathbf{u}_k)}, \quad k \in \mathcal{S}. \tag{3.2}$$

(Note that $\sum_{k \in \mathcal{S}} q_k = Q(0) = 1$.)

Equation (3.2) makes apparent that, in general, the lifetime distribution of the process before leaving the NN is not exponential, but displays many different time scales, as many as elements in the set $\{\alpha_k\}$. Asymptotically, $Q(t) \sim q_{\mathrm{min}} \mathrm{e}^{-\alpha_{\mathrm{min}} t}$, with α_{min} the smallest eigenvalue. Thus, the rate at which the process leaves the NN tends to a minimum as time elapses—in other words, it gets more and more *trapped* within the NN. Eventually, mutation–selection equilibrium is reached, and the process can be described through a unique time scale, α_{min}.

3.2. Mutation accumulation

Perhaps the most immediate consequence of this entrapment of the evolutionary process in an NN is that the observed rate at which neutral mutations accumulate increases with time. Together with the stochastic nature of the dynamics, the consequence is twofold: on the one hand, the molecular clock gets accelerated during those periods of strictly neutral evolution; on the other hand, overdispersion of the molecular clock varies non-monotonically with time, as we show in the following.

In order to analyse how mutations accumulate with time, we need to introduce $P_k(m, t)$, the probability that the process is at time t at a node of degree k of the NN, having undergone exactly m mutations. Introducing $\mathbf{P}_m(t) \equiv (P_k(m, t))^\top$, $k \in \mathcal{S}$, the dynamic equation for this vector can be written as

$$\dot{\mathbf{P}}_m = \mathbf{A}\mathbf{P}_{m-1} - \mathbf{F}\mathbf{P}_m, \tag{3.3}$$

which is valid for all integers m if we assume $\mathbf{P}_m(t) = 0$ for all $m < 0$. Obviously $\mathbf{P}_m(0) = \mathbf{P}(0)\delta_{m,0}$, with $\mathbf{P}(0)$ the initial probability distribution (2.2).

In order to extract information on the mean and variance of the number of mutations accumulated at time t, it is useful to introduce the moment-generating function

$$\mathbf{G}(\theta, t) \equiv \sum_m \mathbf{P}_m(t)\mathrm{e}^{m\theta}, \tag{3.4}$$

so that

$$m_r(t) \equiv \langle m^r \rangle(t) = \lim_{\theta \to 0} \frac{\partial^r}{\partial \theta^r}\left(\sum_{k \in \mathcal{S}} G_k(\theta, t)\right). \tag{3.5}$$

Multiplying equation (3.3) by $\mathrm{e}^{m\theta}$ and summing on m, we obtain the dynamic equation for $\mathbf{G}(\theta, t)$

$$\frac{\partial}{\partial t}\mathbf{G}(\theta, t) = (\mathrm{e}^\theta \mathbf{A} - \mathbf{F})\mathbf{G}(\theta, t). \tag{3.6}$$

Setting $\theta = 0$ shows that $\mathbf{G}(0, t) = \mathbf{P}(t)$, the solution of (2.8), hence $m_0(t) = Q(t)$.

Differentiating (3.6) with respect to θ yields, for $\theta = 0$,

$$\left.\begin{aligned}
&\frac{\partial}{\partial t}\mathbf{G}_\theta(0, t) = (\mathbf{A} - \mathbf{F})\mathbf{G}_\theta(0, t) + \mathbf{A}\mathbf{G}(0, t) \\
\text{and } &\frac{\partial}{\partial t}\mathbf{G}_{\theta\theta}(0, t) = (\mathbf{A} - \mathbf{F})\mathbf{G}_{\theta\theta}(0, t) + \mathbf{A}[2\mathbf{G}_\theta(0, t) + \mathbf{G}(0, t)],
\end{aligned}\right\} \tag{3.7}$$

(subscripts θ denote partial derivatives).

Solving simultaneously equations (2.8), and (3.7), and using (3.5) yields $m_0(t) = Q(t)$, $m_1(t)$ and $m_2(t)$. We are not yet done, because these are not the quantities required to estimate magnitudes related to the molecular clock: actual measurements are always performed on extant populations, so we need the mean $m(t)$ and variance $v(t)$ in the number of mutations conditioned on remaining within the NN at time t. These are obtained as

$$m(t) = \frac{m_1(t)}{m_0(t)}, \quad v(t) = \frac{m_2(t)}{m_0(t)} - m(t)^2. \tag{3.8}$$

Correspondingly, the overdispersion of the molecular clock at time t should be computed as

$$R(t) \equiv \frac{v(t)}{m(t)} = \frac{m_2(t)}{m_1(t)} - \frac{m_1(t)}{m_0(t)}. \tag{3.9}$$

4. Examples

In this section, we compare analytical and numerical results for the dynamics of a RW (representing a homogeneous population) on networks differing in size, fitness and topological features. Our aim is to better understand the quantitative effect of these variables on the dynamics of the population, paying particular attention to the time scales involved, to the change in overdispersion with time and to variations in the rate of fixation of mutations. We analyse a random network (with near zero assortativity and a well-defined average degree) as well as two highly assortative networks with either only two degrees or a constant degree distribution. Our final example comes from secondary structure RNA networks, which share several properties with other natural NN.

4.1. Random network

Despite its apparent simplicity, equation (2.8) cannot be analytically solved in general. An interesting exception is the case of a random network constructed by randomly drawing links

between pairs of nodes. In this situation, the degree follows a Poisson distribution and, consequently, the average degree is well defined. Even this special case is exactly solvable only for $\phi = 1$.

In a random network, there is no correlation between the degree of a given node and that of its nearest neighbours. If the network is infinitely large, its assortativity $r = 0$—though a residual assortativity will be typically obtained when the number of nodes is finite. This lack of correlation reflects in the probability that a node's neighbour has degree k only depends on its own degree. As a matter of fact, $p_{nn}(k|j) = (k/\langle k \rangle)p(k)$. (Here, and in what follows, $\langle k^r \rangle \equiv \sum_k k^r p(k)$). For this special case,

$$\mathbf{F} = \mathbf{I}, \quad \mathbf{A} = \left(\frac{kj}{z\langle k \rangle}p(k)\right), \quad (4.1)$$

and deriving an expression for $Q(t)$ is straightforward (see appendix B). The resulting formula is

$$Q(t) = \frac{z\rho - \langle k \rangle}{\rho(z - \langle k \rangle)}e^{-t} + \frac{(1-\rho)\langle k \rangle}{\rho(z-\langle k \rangle)}e^{-(1-\rho)t}, \quad \rho \equiv \frac{\langle k^2 \rangle}{z\langle k \rangle}. \quad (4.2)$$

Despite the ease of this case, and the fact that the NN has no fitness advantage with respect to the average neighbouring NN ($\phi = 1$), we can already observe what turns out to be a generic property of heterogeneous networks: the lifetime of the RW within the NN is not exponentially distributed. For random networks, there are only two time scales: the 'standard' one, μ^{-1}, and a slower one, with characteristic time $(1 - \rho)^{-1}\mu^{-1}$. The reason for this non-exponential behaviour—and for the equivalent effect that will arise in any heterogeneous network—is that the RW comes in the NN preferentially through nodes of low inner connectivity (high outer connectivity) and as time passes without leaving the NN it progressively moves towards the regions of higher connectivity (note that the mean degree of a node's neighbour is $\langle k^2 \rangle/\langle k \rangle > \langle k \rangle$), from which escaping is less probable. Thus, the first scale corresponds to walkers which leave the network in very few steps, whereas the second (slow) scale characterizes walkers that have spent some time in the NN.

Further insights can be gained by calculating the overdispersion (3.9) (see appendix B), whose expression for this case becomes

$$R(t) = 1 + \frac{(z\rho - \langle k \rangle)\rho t e^{-\rho t}}{(1 - \rho)\langle k \rangle + (z\rho - \langle k \rangle)\rho t e^{-\rho t}}. \quad (4.3)$$

Figure 1 shows the probability $Q(t)$ to remain on the network at time t, the overdispersion $R(t)$ of the molecular clock with time and the mutation rate (measured as the time derivative of the mean number of mutations) for an Erdös–Renyi random network. In figure 1a, as well as in all other plots, $Q(t)$ is represented as a function of the rescaled variable ϕt. The reason is that, as can be seen, there is a partial collapse of all $Q(t)$ curves (but see the Discussion for forthcoming examples). In addition, the case $\phi = 0$ is not included, because $\phi = 0$ implies that the RW cannot leave the NN and, therefore, $Q(t) = 1$, for all t. For the Erdös–Renyi random networks, the fast time scale is barely visible, due to the fact that the RW becomes rapidly trapped. The slow time scale actually dominates the dynamics: for the values chosen, $(1 - \rho) \simeq 0.47$ (case $\phi = 1$) and the coefficient of the corresponding exponential in $Q(t)$ is about 10 times larger than that of the fast time scale.
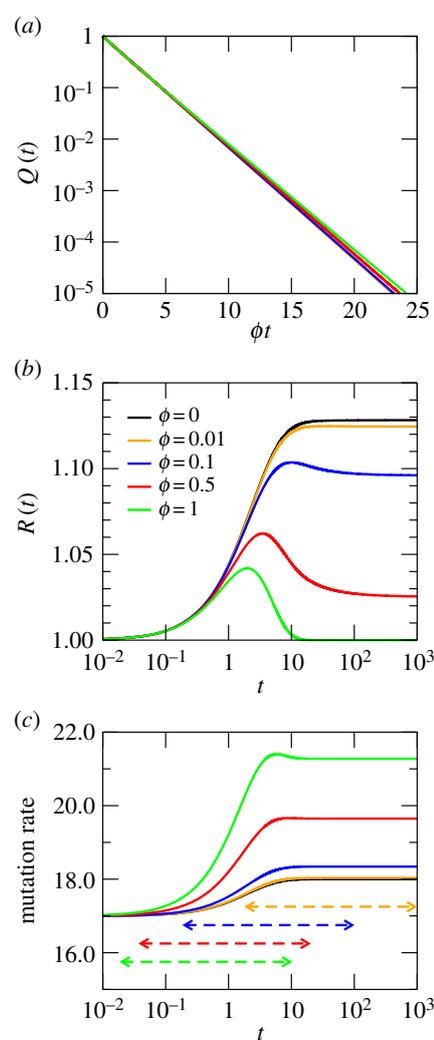
Figure 1. Dynamics of RWs on an Erdös–Renyi random network with $M = 20\,000$ nodes and average degree $\langle k \rangle = 18$. (a) Probability $Q(t)$ that the RW stays on the network after a time t. (b) Overdispersion of the network. (c) Mutation rate or acceleration of the molecular clock. The time interval through which the number of trajectories remaining on the network decreases from 99% to 1% is represented as a horizontal dashed line limited by two arrows, with the same colour as the ϕ curve they correspond to. The same code is used in subsequent corresponding panels for other networks. The measured assortativity for this finite network is $r \approx 0.018$.

Figure 1b shows the typical behaviour of the overdispersion on heterogeneous NN: an initial increase up to a maximum followed by a decay to an asymptotic value. Interestingly, even for $\phi = 1$ (corresponding to equation (4.3)), a 4% overdispersion is reached before it decays to its asymptotic value, a process that requires about 10 mutations ($t \sim 10\,\mu^{-1}$). As values of ϕ diminish, the asymptotic value of $R(t)$ monotonically increases. Three features that will show up in the remaining examples as well should be highlighted: first, ϕ has a strong, non-trivial effect on overdispersion; second, it is in the transient where overdispersion reaches its largest value; third, the rate of accumulation of mutations increases with ϕ (it therefore diminishes the larger the difference between the fitness of the current NN and its neighbouring phenotypes).

In figure 1c, we can appreciate how the ticking rate of the molecular clock undergoes an acceleration from its initial value, eventually reaching a steady rate higher than the one it started off from. The change occurs at a time $t \sim 10\,\mu^{-1}$,

which corresponds to the number of mutations accumulated. Note that, owing to the heterogeneity of the network, the initial mutation rate is already higher than μ. As we stated earlier, the stationary state is not attained in general. Now, we can quantify this statement by estimating the time $t_{1\%}(\phi)$ by which only 1% of the RWs remain on the network. This time can be obtained by numerically solving $Q(t_{1\%}) = 0.01$—it can be approximately read from figure 1a. In order to estimate the range of t values where typical trajectories on the NN should be found, we have calculated the time $t_{99\%}(\phi)$ as well, when only 1% of the trajectories have left the network. The interval $(t_{99\%}(\phi), t_{1\%}(\phi),)$ is represented as a dashed line within two arrows in figure 1c (and all subsequent (c) panels). As can be seen, the maximum variation in $R(t)$ and in the mutation rate is achieved within that time interval, whereas mutation–selection equilibrium is rarely attained. Note that $t_{99\%}(\phi)$ and $t_{1\%}(\phi)$ decrease with increasing ϕ. The acceleration of the molecular clock along neutral evolution, which is higher the smaller ϕ, is a feature that will reappear in all subsequent examples, and probably stands as the best illustration of the phenotypic entrapment produced by NN. Typically, most RWs will leave the network before reaching the asymptotic neutral ticking rate, and certainly before the RW has had time to explore a significant fraction of the NN [29].

4.2. Two-degree network

Consider now a network with M nodes such that $p(k) = 1/2$ for $k = k_1, k_2$ ($k_1 < k_2$), and 0 otherwise. The $M/2$ nodes of each degree are randomly connected between them, except for L links that connect (also randomly) nodes of the two different degrees. Accordingly, $p_{nn}(k_i|k_i) = 1 - 2L/(Mk_i)$ and $p_{nn}(k_j|k_i) = 1 - 2L/(Mk_i)$, for $j \neq i$. The nearest-neighbour degree distribution clearly reflects, in this case, the trapping power of more connected regions: note that the probability of entering the group of high connectivity (k_2) from that of low connectivity is higher than the probability of leaving it, $p_{nn}(k_1|k_2) < p_{nn}(k_2|k_1)$. The asymmetry of p_{nn} reflects the entrapment by progressively more connected regions of the NN.

In terms of the variables $\ell \equiv 2L/(Mz)$, $\kappa = (k_2 + k_1)/2z$ and $\delta = (k_2 - k_1)/(2z)$, matrix $\mathbf{A} - \mathbf{F}$ becomes

$$\mathbf{A} - \mathbf{F} = \begin{pmatrix} \phi(\kappa - \delta) - \ell - \phi & \ell \\ \ell & \phi(\kappa + \delta) - \ell - \phi \end{pmatrix}, \quad (4.4)$$

and the resulting probability (3.2) of remaining trapped within the NN at time t is

$$Q(t) = q_+ e^{-\alpha_+ t} + q_- e^{-\alpha_- t}, \quad (4.5)$$

where

$$q_{\pm} = \frac{1}{2}\left[1 \pm \frac{\ell - (1-\kappa)^{-1}\phi\delta^2}{\Delta}\right], \quad (4.6)$$
$$\alpha_{\pm} = \ell + \phi(1 - \kappa) \mp \Delta,$$

with $\Delta \equiv \sqrt{\ell^2 + \phi^2\delta^2}$. Thus, the dynamics on a two-degree network is characterized by two time scales, with coefficients $q_{\pm}$ further determining the relative weight of each one in the process between arriving at the NN and achieving maximum entrapment.

Despite the formal similarity of $Q(t)$ with the solution for an Erdös–Renyi network, there are important quantitative differences between the two cases. Figure 2a illustrates the presence of the two time scales, here clearly visible (note
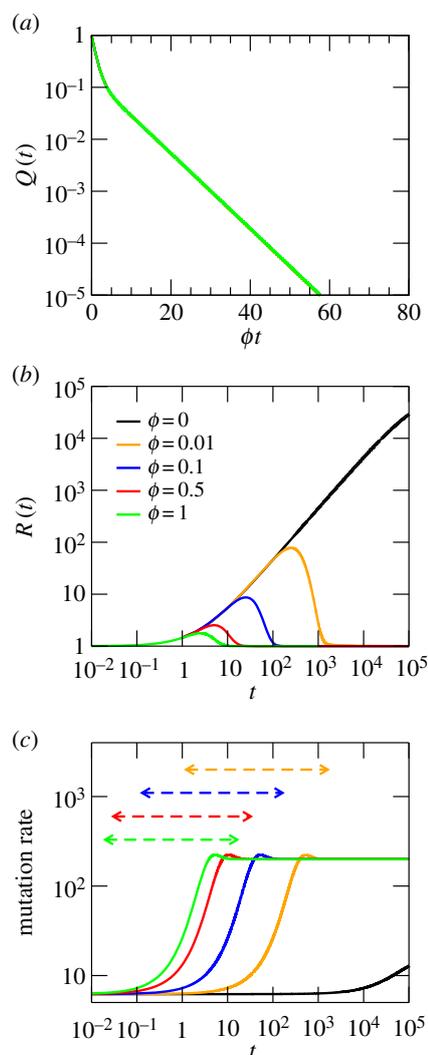
Figure 2. Dynamics of a RW on a two-degree network with $k_1 = 2$, $k_2 = 30$, $L = 2$ and $M = 20\,000$ nodes. (a) Probability $Q(t)$ that the RW stays on the network after a time t. (b) Overdispersion of the network. (c) Mutation rate. By construction, the assortativity is very high, $r \approx 0.99995$.

the change in the horizontal scale with respect to figure 1a). The probability to stay on this two-degree network is also significantly higher after a fixed time, which speaks for longer trajectories in comparison with those on a random network. The average degree does not play an essential role in this behaviour, because it has value 16, slightly below the random network previously studied; this nonetheless, the two-degree network is more efficient at trapping the RW. Let us now explain why the curves for $Q(t)$ collapse so nicely in this case. This is owing to the very low value of $\ell \simeq 5 \times 10^{-6}$ for the parameters chosen. As a result, the two time scales are well approximated as $\alpha_- \simeq \phi(1 - k_2/z)$ and $\alpha_+ \simeq \phi(1 - k_1/z)$, thus nearly scaling with ϕ as long as $\phi \gtrsim \ell$. Hence, the collapse of the $Q(t)$ curves is a particular feature of this network and does not hold in general, as the next example will show.

Differences in overdispersion between the random and the two-degree network are still more remarkable. Figure 2b (note the logarithmic scale in the y-axis) represents $R(t)$, showing an extraordinary increase in overdispersion with time, especially as ϕ decreases. Although not shown in the picture, $R(t)$ eventually converges to 1, albeit the time required is huge. This is another important difference with

the random network, where the overdispersion saturates at values $R(t \to \infty) > 1$. This is due to the persistence of heterogeneity in the degree *from the viewpoint of the RW*. Whereas in the random network, the walker visits nodes of different degree once mutation–selection equilibrium has been reached, in the two-degree network the region of low connectivity is mostly invisible to the dynamics, trapped in the large cluster of high-connectivity nodes. Nonetheless, there is an enormous dispersion in the time required for a particular trajectory to become trapped; this fact explaining the growth of overdispersion, the delayed achievement of its maximum value and the slow convergence to equilibrium.

Similar differences can be observed in the mutation rate, depicted in figure 2c, compared with that for the Erdös–Renyi network. The acceleration of the molecular clock for this network is very large (over an order of magnitude with respect to the initial ticking rate). In addition, as ϕ decreases, the acceleration occurs at earlier times, an effect that was much milder in the Erdös–Renyi network. As a matter of fact, these curves collapse nicely when they are represented as a function of ϕt, as happened with $Q(t)$. The asymptotic mutation rate is independent of ϕ, because, similar to what happens with overdispersion, the RW does not see any difference in the degrees visited at large times, even for different values of ϕ. The next example clarifies how ϕ affects the trapping strength of nodes with different degree, and the resulting variation in the asymptotic mutation rate.

4.3. Constant degree distribution network

This example shares some similarities with the previous one, but now nodes of all degrees between a minimum $k_{min} = 2$ and a maximum value appear in equal amounts. In addition, the fraction of links connecting nodes of different degrees is notably larger. Consider a network with a degree distribution $p(k) = 1/50$, $2 \leq k \leq 51$, $N = 20\,000$ nodes, and $z = 52$. In order to generate an assortative network, we proceed as follows. First, we assign an arbitrary index to the nodes (with no meaning whatsoever) and divide them up into blocks of $400(= 20\,000/50)$ consecutive nodes. We then assign degree 2 to the first block, degree 3 to the second block and so on until exhausting the nodes. Now, we randomly connect the nodes, in such a way that the probability to connect a pair of nodes at positions i and j is taken proportional to $\exp\{-d/200\}$, where $d = |i - j|$. With this procedure, nodes with similar degree are preferentially connected, and high assortativity is ensured. It is not possible to derive any exact expression for $Q(t)$ or $R(t)$, though the results can be interpreted in the light of our previous examples.

Figure 3a shows the presence of many different time scales that become dominant at different times and cause a systematic bending of $Q(t)$. The visibility of the different time scales is related to the presence of significant fractions of nodes of different degrees together with the canalization of the dynamics towards regions of increasingly higher connectivity—a consequence of assortativity. This behaviour differs from that of the two-degree network, where the curvature region was small and owing to the cross-over between the only two time scales involved. Overdispersion can reach very high values, even once mutation–selection equilibrium has been reached, and so does the mutation rate (figures 3b,c).
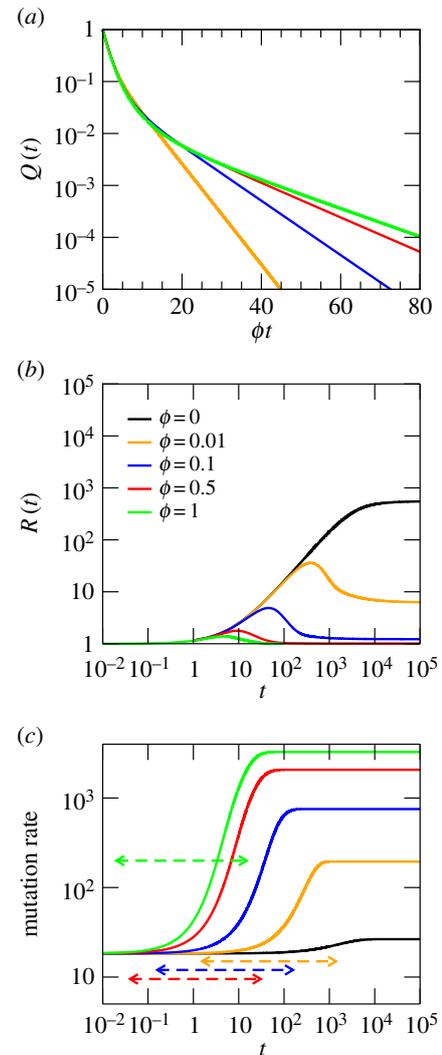
Figure 3. Dynamics of a RW on a network with a constant degree distribution, with parameters described in the main text. (a) Probability $Q(t)$ that the RW stays on the network after a time t. (b) Overdispersion of the network. (c) Mutation rate. The assortativity is $r \approx 0.9997$. For the sake of comparison, these plots share scale with figure 2.

In a network of this kind, the larger ϕ the less diverse is the set of different degrees visited by the RW in the asymptotic state. This responds to the fact that trajectories that survive for long times are preferentially concentrated in highly connected regions. As ϕ decreases, nodes of lower degree become more competent at trapping the dynamics, causing two effects. First, the asymptotic time scale α_{min} changes, decreasing in absolute value with ϕ. Second, differences between the trajectories are less pronounced the larger ϕ, such that the maximum in the overdispersion and the largest acceleration of the clock are reached at earlier times. Third, the asymptotic values of $R(t)$ decrease with ϕ, and those of the mutation rate increase. At mutation–selection equilibrium, however, nodes with the highest connectivity cannot fully trap the RW, such that nodes of lower degree are visited frequently enough to yield $R(t) > 1$, attaining larger values the lower ϕ. In the asymptotic state, thus, the RW is more delocalized (in terms of the visited degree), the lower the value of ϕ. The fact that, at high ϕ, those RW remaining on the NN preferentially visit nodes of high degree is responsible for the larger ticking rates observed—reaching values orders of magnitude bigger than initially.

4.4. RNA networks

Secondary structure RNA NNs (here RNA networks for short) are a paradigm of the genotype–phenotype map, and many important properties on their topology and on genotype–phenotype mapping are known in detail. An RNA network is formed in principle by all the sequences that yield the same minimum free energy secondary structure. Usually, links are drawn between pairs of sequences in the set that differ in a single point mutation. In general, this leads to a number of disjoint connected components. Because dynamics are well defined on connected networks, we will assume that we work with subsets of the genotype–phenotype map that fulfil this condition. Among other properties known, there is a broad variability in the number of sequences that fold into a given phenotype for a fixed sequence length (several orders of magnitude) [6,30]; the average degree of an RNA network grows logarithmically with its size (number of genotypes leading to the same phenotype) [22], and RNA networks are highly assortative, with an index r that increases on average with size and approaches 1 [22]. In this section, we discuss the results obtained for RNA networks of sequences of length 12. The genotype–phenotype map for this case is fully known, with networks of sizes spanning four and a half orders of magnitude [22].

We have chosen two connected NNs formed by genotypes that belong to different phenotypes. In dot–bracket notation, these are ...((.....)) and (((.....))).; the first phenotype can be obtained from 14 675 different sequences which are grouped into four different connected subnetworks. As an example, we have chosen a connected component of size $M = 1965$. The second phenotype is more abundant, corresponding to the minimum free energy folded state of 142 302 sequences. As a second example, we have chosen one of its largest 17 connected subnetworks, of size $M = 21\,908$. Curves for $Q(t)$, $R(t)$ and the mutation rate are plotted in figure 4. As can be seen, the behaviour of these networks is quantitatively comparable to the Erdös–Renyi random network. In RNA, there is a monotonic change in the slower time scale and in the equilibrium value of overdispersion with network size. This responds to the particular relationship between size and degree exhibited by secondary structure RNA networks, where $\langle k \rangle \propto \ln M$. The connection between the slower time scale and average degree can be made explicit in the calculations for random networks (equations (B 1) and (4.2)), and despite the high assortativity of RNA networks, it also shows up here. The degree distribution $p(k)$ of RNA networks is very peaked for M small, though its variance increases with M faster than it does for a Poisson degree distribution [22]. Whether these differences lead, for realistic values of M, to significant departures from the behaviour of random networks remains to be seen.

5. Discussion

Certain features of the dynamics of homogeneous populations on NNs are generic, resulting from two ubiquitous topological properties: heterogeneity and assortativity. In general, the dynamical process that begins when the population enters the NN and finishes when mutation–selection equilibrium is achieved is characterized by as many different time scales as different degrees of robustness are found in the NN. These time scales successively show up along the
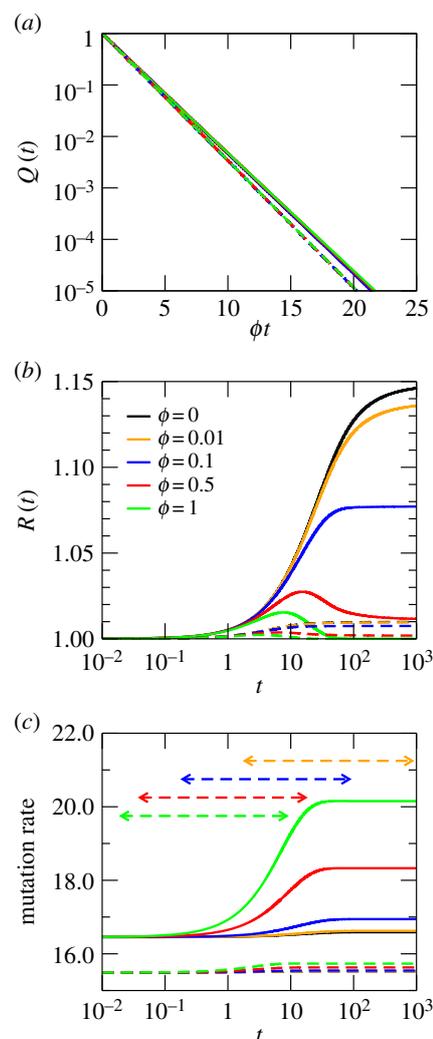
Figure 4. Dynamics of a RW on two different RNA secondary structure NNs with $z = 36$. Sizes are $M = 1965$ (dashed lines) and $M = 21\,908$ (solid lines). (a) Probability $Q(t)$ that the RW stays on the network after a time t. (b) Overdispersion of the network. (c) Mutation rate. Times $t_{1\%}(\phi)$ and $t_{99\%}(\phi)$ are only shown for the largest network. The assortativity is $r > 0.90$. For the sake of comparison, these plots share scale with figure 1.

transient and affect the process of mutation fixation. As time elapses, regions of increasingly higher degree are visited, causing entrapment of the population. The actual time spent on a NN should increase as the difference between the fitness of the current phenotype and that of neighbouring alternatives grows, because fitter phenotypes trap populations for longer times.

Some of the features here described directly affect evolutionary dynamics on realistic NNs and, therefore, quantitative properties of the actual mutation substitution process. We have used as an example RNA secondary structure NNs, because this is the only case we are aware of where exhaustive studies of the genotype–phenotype map encompassing the whole space of genotypes have been performed [22,31]. Still, the networks we may use to compute quantities such as overdispersion are limited by current computational power. The average size of an RNA network for sequences of length l grows as $0.673l^{3/2}2.1636^l$, which is an astronomically large number even for moderate values of l [6]. For instance, a natural tRNA 76 nt long should have a NN formed by over 10^{28} different sequences. There is partial evidence that, as the

length of genotypes increases, the degree distribution of the corresponding NNs broadens, whereas assortativity grows with size [22]. If this is so, several different time scales should be visible and relevant in the dynamic process of mutation fixation in any realistic scenario. An additional example is provided by the NNs of natural proteins with known folds that have been computationally studied. Although the techniques used only permit a partial exploration of those immense NNs, the degree distributions obtained are remarkably broad [14]. There is sufficient evidence as well that protein NNs are highly assortative [20].

Network heterogeneity, together with stochasticity in the appearance and fixation of mutations, translates into potentially large values of $R(t)$. In most cases, overdispersion changes non-monotonically, presenting a maximum value at intermediate times, before equilibrium is reached. The value of evolutionary time in our model yields an upper bound to the number of mutations accumulated up to time t, because we have defined time in units of μ^{-1}, where μ represents the rate of attempted mutations. Therefore, typical trajectories accumulate between a few and a few tens of mutations before leaving the network, a range where the strongest variation in $R(t)$ and mutation rates occurs. The non-monotonicity of $R(t)$ has been observed in realistic models of protein evolution [14]. In an exhaustive analysis of thousands of protein sequences in several *Drosophila* species [32], it has been demonstrated that $R(t)$ is time- and gene-dependent, and that it should be described as a function of a phylogeny's level of divergence. The former observations are in full qualitative agreement with our results.

For any time t, $R(t) > 1$ as long as genomes with different robustness (nodes with different degree) are visited with non-zero probability. There might be some extreme cases, such as the two-degree networks we have analysed, when only nodes with the same degree are visible to the population at mutation–selection equilibrium. In those special situations (and, trivially, if the network is homogeneous), $R(t) = 1$ is recovered, in agreement with Kimura's initial guess on the Poissonian nature of the substitution process. However, this is a very atypical situation that cannot be expected to occur in any realistic case. That network heterogeneity increases overdispersion had already been acknowledged before [19]. What had not been realized is the fact that this effect is stronger during transients; that transients are unavoidable, because evolution on NN begins preferentially in regions of low robustness; and that—depending on the value of ϕ— equilibrium may be virtually unreachable before a mutation hits a fitter phenotype, driving the process again far from equilibrium. Indeed, the number of mutations that have to be accumulated to attain equilibrium vary from slightly less than a hundred to a few thousands, a situation not reached by any typical trajectory unless ϕ is close to 1. It would be interesting to consider other mutational mechanisms, such as recombination, which might affect the nature of equilibrium and the length of the transient. For example, it has been shown that, under recombination, the mutation–selection equilibria of populations replicating on NNs display a higher degree of robustness (are thus more entrapped) than those corresponding to simple point mutations [29].

There might be some natural systems where the accumulation of mutations, and therefore the transient time before equilibrium is achieved, proceeds at a pace significantly higher than that in most genes or proteins, namely RNA viruses, viroids and miRNAs. These three systems are characterized by high mutation rates and, in RNA viruses and miRNAs, a correlation between selection for thermostability and increased robustness to mutations has been detected (a phenomenon called plastogenetic congruence) [33,34]. Viroids, which share several structural properties with miRNAs, also increase robustness along evolution [35] and present a limited effect of mutations on their structure [36]. The higher robustness of these systems with respect to randomized counterparts is an indirect evidence of phenotypic entrapment and, therefore, of populational states closer to mutation–selection equilibrium. Their corresponding phylogenies might be excellent testing beds for the theoretical results here derived.

Evolutionary dynamics on heterogeneous and assortative networks with phenotype-dependent size has important consequences regarding the molecular clock [37]. First, there is an increase in the rate at which mutations are fixed as time elapses, with an occasional maximum in its value during the transient. When the population first enters the network, the probability that it stays is relatively low. However, if it remains, the neutral mutation that is consequently fixed does it in a typical time that depends on the robustness of the current genotype. Because robustness systematically increases along the transient in assortative networks, the age of a population which has accumulated few neutral mutations will be systematically underestimated if a regularly ticking molecular clock is assumed. This prediction is in agreement with observations of a steady increase in $R(t)$ along phylogenies [32] where, moreover, a gene-dependent $R(t)$ was identified. Second, the ticking rate of the clock depends on ϕ, that is, on the fitness value of the current phenotype (or NN) relative to neighbouring phenotypes: the smaller ϕ, the higher the trapping power of the network, a situation that causes larger overdispersion and lower asymptotic mutational rates. Finally, even sequences of the same length may belong to NNs that differ in size in several orders of magnitude. The ticking rate of the clock is unavoidably dependent on robustness, which in its turn is a function of phenotype size in all genotype–phenotype maps studied [9,22–25]. Reliable estimations of NN sizes, at hand with appropriate computational algorithms [38], should aid in the calibration of neutral evolutionary rates. This information could be combined with knowledge of $R(t)$ in homologous genes with the same function (thus presumably characterized by the same NN). Finally, the $R(t)$ and the NN sizes of different genes in diverging lineages could be compared in order to disentangle time-dependent changes in $R(t)$ from variation owing to phenotype sizes, extending studies such as those in reference [32].

The results here obtained do not contradict observations revealing a decrease in the ticking rate of the molecular clock along phylogenetic branches [39,40], because most fixed mutations are in those cases subjected to selection. In this sense, our model addresses a partial aspect of phylogenies, namely strictly neutral evolution. A tractable model of how NNs of different fitness intermingle in the space of genomes may bring about the formulation of mean-field dynamical models, such as the one here presented, that simultaneously take into account neutral drift and selection without discarding the complex architecture of the space of genotypes. Actually, it is plausible that the integration of several different

structural features of phenotypes may clarify why the molecular clock depends so strongly on specific cases analysed [41]. Indeed, there are several mechanisms known to affect the uniformity of rates of molecular evolution [42]. In the light of the results here presented, also the particular topology of NNs, the degree of adaptation of populations and the abundance of the current phenotype should be taken into account when calibrating molecular clocks.

Appendix A. Solving for $Q(t)$ in general

According to equation (3.1), to obtain $Q(t)$, we first need to compute $\mathbf{P}$. Except for very specific choices of $p_{nn}(k|j)$ and ϕ, the master equation (2.8) needs to be solved numerically. We prove here that matrix $\mathbf{F} - \mathbf{A}$ has s (the cardinal of set $\mathcal{S}$) different real eigenvalues—which justifies formula (3.1).

Let us define $\mathbf{M}(x) \equiv \mathbf{F} - x\mathbf{A}$, where x is a complex variable. Standard perturbation theory [43] tells us that the eigenvalues of $\mathbf{M}(x)$ form a set of analytic functions of x with constant cardinality, except for a finite set of *singular* points in the complex plane—where at least two of the eigenvalues coalesce. But $\mathbf{M}(0) = \mathrm{diag}(\phi + (1 - \phi)k/z)$, $k \in \mathcal{S}$, so it has s different real eigenvalues (s being the cardinal of $\mathcal{S}$), provided $\phi \neq 1$. Therefore, $\mathbf{M}(x)$ will also have s different eigenvalues—hence will be diagonalizable—except when x is one of those singular points. Now, $x = 1$ will not be a singular point for almost any probability distributions $p(k|j)$. But if $p(k|j)$ happens to be one of those singular distributions, then the analysis to come is still valid for $x = 1 + \epsilon$, for arbitrarily small ϵ (because the singular points are isolated in the complex plane). Thus, we can eventually take the limit $\epsilon \to 0$ without changing the conclusions.

We now show that all eigenvalues of $\mathbf{M}(x)$ are real. To that aim let us introduce the (non-singular) matrix $\mathbf{D} = \mathrm{diag}((z/k)p(k))$, $k \in \mathcal{S}$. Then, the elements of $\mathbf{T}(x) \equiv \mathbf{M}(x)\mathbf{D}$ are $[(1 - \phi) + \phi z/k]p(k)\delta_{kj} - xp_{nn}(k, j)$, where $p_{nn}(k, j)$ is the fractions of links of the NN joining two nodes of degrees k and j. Hence, $\mathbf{T}(x)$ is a symmetric matrix and therefore all its eigenvalues are real. But then so are the eigenvalues of $\mathbf{M}(x)$, because $\mathbf{M}(x) = \mathbf{T}(x)\mathbf{D}^{-1}$, which has exactly the same eigenvalues as the symmetric matrix $\mathbf{D}^{-1/2}\mathbf{T}(x)\mathbf{D}^{-1/2}$.

Let $\mathbf{u}_j$ and $\mathbf{v}_j$, $j \in \mathcal{S}$, be the right and left eigenvectors of $\mathbf{M}(1)$ corresponding to the eigenvalue α_j. They verify the orthogonality condition $\mathbf{v}_k \cdot \mathbf{u}_j \propto \delta_{kj}$. Then, the solution of $\dot{\mathbf{P}} = -\mathbf{M}(1)\mathbf{P}$ can be obtained as

$$\mathbf{P}(t) = \sum_{k \in \mathcal{S}} \frac{\mathbf{v}_k \cdot \mathbf{P}(0)}{\mathbf{v}_k \cdot \mathbf{u}_k} e^{-\alpha_k t} \mathbf{u}_k. \tag{A 1}$$

Substituting this result into (3.1) yields (3.2).

Appendix B. $Q(t)$ and $R(t)$ for random networks

To obtain $Q(t)$, we left-multiply (2.8) by $\boldsymbol{\kappa}^{\top}$. This leads to

$$\frac{\mathrm{d}}{\mathrm{d}t}\boldsymbol{\kappa} \cdot \mathbf{P} = -(1 - \rho)\boldsymbol{\kappa} \cdot \mathbf{P}, \quad \rho \equiv \frac{\langle k^2 \rangle}{z\langle k \rangle}, \tag{B 1}$$

whose solution is simply

$$\boldsymbol{\kappa} \cdot \mathbf{P}(t) = \boldsymbol{\kappa} \cdot \mathbf{P}(0)e^{-(1-\rho)t}, \quad \boldsymbol{\kappa} \cdot \mathbf{P}(0) = \frac{(1 - \rho)\langle k \rangle}{z - \langle k \rangle}. \tag{B 2}$$

Substituting into (3.1) finally yields (4.2).

As for overdispersion, from equations (3.7), we can derive

$$\left.\begin{array}{l} \dot{m}_1 = -m_1 + G_1 + \boldsymbol{\kappa} \cdot \mathbf{P}, \\[4pt] \dot{G}_1 = -(1 - \rho)G_1 + \rho\boldsymbol{\kappa} \cdot \mathbf{P}, \\[4pt] \dot{m}_2 = -m_2 + G_2 + 2G_1 + \boldsymbol{\kappa} \cdot \mathbf{P} \\[4pt] \dot{G}_2 = -(1 - \rho)G_2 + \rho(2G_1 + \boldsymbol{\kappa} \cdot \mathbf{P}), \end{array}\right\} \tag{B 3}$$

and

with the initial condition $m_1(0) = m_2(0) = G_1(0) = G_2(0) = 0$. Combining the former two equations and the latter two, we obtain

$$\frac{\mathrm{d}}{\mathrm{d}t}(G_i - \rho m_i) = -(G_i - \rho m_i), \quad i = 1, 2, \tag{B 4}$$

which, given the null initial conditions, yield $G_i = \rho m_i$. Therefore, the four equations (B 3) reduce to

$$\left.\begin{array}{l} \dfrac{\mathrm{d}}{\mathrm{d}t}m_1 = -(1 - \rho)m_1 + \boldsymbol{\kappa} \cdot \mathbf{P} \\[8pt] \dfrac{\mathrm{d}}{\mathrm{d}t}(m_2 - m_1) = -(1 - \rho)(m_2 - m_1) + 2\rho m_1. \end{array}\right\} \tag{B 5}$$

and

Using (B 2), this linear system can be readily solved as

$$\left.\begin{array}{l} m_1(t) = \dfrac{(1 - \rho)\langle k \rangle}{z - \langle k \rangle} t e^{-(1-\rho)t} \\[10pt] m_2(t) = \dfrac{(1 - \rho)\langle k \rangle}{z - \langle k \rangle}(\rho t^2 + t)e^{-(1-\rho)t}, \end{array}\right\} \tag{B 6}$$

and

from which overdispersion can be obtained as (4.3).

References

1. Kimura M. 1968 Evolutionary rate at the molecular level. *Nature* **217**, 624–626. (doi:10.1038/217624a0)
2. van Nimwegen E, Crutchfield JP, Huynen M. 1999 Neutral evolution of mutational robustness. *Proc. Natl Acad. Sci. USA* **96**, 9716–9720. (doi:10.1073/pnas.96.17.9716)
3. Draghi JA, Parsons TL, Plotkin JB. 2011 Epistasis increases the rate of conditionally neutral substitution in an adapting population. *Genetics* **187**, 1139–1152. (doi:10.1534/genetics.110.125997)
4. Smith JM. 1970 Natural selection and the concept of a protein space. *Nature* **225**, 563–564. (doi:10.1038/225563a0)
5. Lipman DJ, Wilbur WJ. 1991 Modelling neutral and selective evolution of protein folding. *Proc. R. Soc. Lond. B* **245**, 7–11. (doi:10.1098/rspb.1991.0081)
6. Schuster P, Fontana W, Stadler PF, Hofacker IL. 1994 From sequences to shapes and back: a case study in RNA secondary structures. *Proc. R. Soc. Lond. B* **255**, 279–284. (doi:10.1098/rspb.1994.0040)
7. Ciliberti S, Martin OC, Wagner A. 2007 Innovation and robustness in complex regulatory gene networks. *Proc. Natl Acad. Sci. USA* **104**, 13 595–13 596. (doi:10.1073/pnas.0705396104)
8. Rodrigues JFM, Wagner A. 2009 Evolutionary plasticity and innovations in complex metabolic reaction networks. *PLoS Comput. Biol.* **5**, e1000613. (doi:10.1371/journal.pcbi.1000613)
9. Greenbury SF, Johnston IG, Louis AA, Ahnert SE. 2014 A tractable genotype–phenotype map for the

self-assembly of protein quaternary structure. *JRSI* **11**, 20140249.

10. Kimura M, Ohta T. 1971 On the rate of molecular evolution. *J. Mol. Evol.* **1**, 1–17. (doi:10.1007/BF01659390)

11. Langley CH, Fitch WM. 1974 An examination of the constancy of the rate of molecular evolution. *J. Mol. Evol.* **3**, 161–177. (doi:10.1007/BF01797451)

12. Gillespie JH. 1984 The molecular clock may be an episodic clock. *Proc. Natl Acad. Sci. USA* **81**, 8009–8013. (doi:10.1073/pnas.81.24.8009)

13. Takahata N. 1987 On the overdispersed molecular clock. *Genetics* **116**, 169–179.

14. Bastolla U, Porto M, Roman HE, Vendruscolo M. 2003 Connectivity of neutral networks, overdispersion, and structural conservation in protein evolution. *J. Mol. Biol.* **56**, 243–254.

15. Aguirre J, Buldú JM, Manrubia SC. 2009 Evolutionary dynamics on networks of selectively neutral genotypes: effects of topology and sequence stability. *Phys. Rev. E* **80**, 066112. (doi:10.1103/PhysRevE.80.066112)

16. Wagner A. 2011 *The origins of evolutionary innovations*. Oxford, UK: Oxford University Press.

17. Bastolla U, Roman HE, Vendruscolo M. 1999 Neutral evolution of model proteins: diffusion in sequence space and overdispersion. *JTB* **200**, 49–64. (doi:10.1006/jtbi.1999.0975)

18. Wilke CO. 2004 Molecular clock in neutral protein evolution. *BMC Genet.* **5**, 25. (doi:10.1186/1471-2156-5-25)

19. Raval A. 2007 Molecular clock on a neutral network. *Phys. Rev. Lett.* **99**, 138104. (doi:10.1103/PhysRevLett.99.138104)

20. Bornberg-Bauer E, Chan HS. 1999 Modeling evolutionary landscapes: Mutational stability, topology, and superfunnels in sequence space. *Proc. Natl Acad. Sci. USA* **96**, 10 689–10 694. (doi:10.1073/pnas.96.19.10689)

21. Wuchty S, Fontana W, Hofacker IL, Schuster P. 1999 Complete suboptimal folding of RNA and the stability of secondary structures. *Biopolymers* **49**, 145–165. (doi:10.1002/(SICI)1097-0282(199902)49:2<145::AID-BIP4>3.0.CO;2-G)

22. Aguirre J, Buldú JM, Stich M, Manrubia SC. 2011 Topological structure of the space of phenotypes: the case of RNA neutral networks. *PLoS ONE* **6**, e26324. (doi:10.1371/journal.pone.0026324)

23. Li H, Helling R, Tang C, Wingreen N. 1996 Emergence of preferred structures in a simple model of protein folding. *Science* **273**, 666–669. (doi:10.1126/science.273.5275.666)

24. Bloom JD, Raval A, Wilke CO. 2007 Thermodynamics of neutral protein evolution. *Genetics* **175**, 255–266. (doi:10.1534/genetics.106.061754)

25. Dall'Olio GM, Bertranpetit J, Wagner A, Laayouni H. 2014 Human genome variation and the concept of genotype networks. *PLoS ONE* **9**, e107347. (doi:10.1371/journal.pone.0107347)

26. Feld SL. 1991 Why your friends have more friends than you do. *Am. J. Sociol.* **96**, 1464–1477. (doi:10.1086/229693)

27. Ewens WJ. 2004 *Mathematical population genetics*, 2nd edn. New York, NY: Springer.

28. Gillespie JH. 1994 *The causes of molecular evolution*. New York, NY: Oxford University Press.

29. Szöllősi GJ, Derényi I. 2008 The effect of recombination on the neutral evolution of genetic robustness. *Math. Biosci.* **214**, 58–62. (doi:10.1016/j.mbs.2008.03.010)

30. Stich M, Briones C, Manrubia SC. 2008 On the structural repertoire of pools of short, random RNA sequences. *J. Theor. Biol.* **252**, 750–763. (doi:10.1016/j.jtbi.2008.02.018)

31. Grüner W, Giegerich R, Strothmann D, Reidys C, Weber J, Hofacker IL, Stadler PF, Schuster P. 1996 Analysis of RNA sequence structure maps by exhaustive enumeration. I. Neutral networks. *Monatshefte F. Chem.* **127**, 355–374. (doi:10.1007/BF00810881)

32. Bedford T, Hartl DL. 2008 Overdispersion of the molecular clock: temporal variation of gene-specific substitution rates in *drosophila*. *Mol. Biol. Evol.* **25**, 1631–1638. (doi:10.1093/molbev/msn112)

33. Domingo-Calap P, Pereira-Gómez M, Sanjuán R. 2010 Selection for thermostability can lead to the emergence of mutational robustness in an RNA virus. *J. Evol. Biol.* **23**, 2453–2460. (doi:10.1111/j.1420-9101.2010.02107.x)

34. Szöllősi GJ, Derényi I. 2009 Congruent evolution of genetic and environmental robustness in micro-RNA. *Mol. Biol. Evol.* **26**, 867–874. (doi:10.1093/molbev/msp008)

35. Sanjuán R, Forment J, Elena SF. 2006 *In silico* predicted robustness of viroids RNA secondary structures. I. The effect of single mutations. *Mol. Biol. Evol.* **23**, 1427–1436. (doi:10.1093/molbev/msl005)

36. Manrubia SC, Sanjuán R. 2013 Shape matters: effect of point mutations on RNA secondary structure. *Adv. Comp. Syst.* **16**, 1250052. (doi:10.1142/S021952591250052X)

37. Zuckerkandl E, Pauling L. 1965 Evolutionary divergence and convergence in proteins. In *Evolving genes and proteins* (eds V Bryson, H Vogel), pp. 97–166. New York, NY: Academic Press.

38. Jörg T, Martin OC, Wagner A. 2008 Neutral network sizes of biological RNA molecules can be computed and are atypically large. *BMC Bioinform.* **9**, 464. (doi:10.1186/1471-2105-9-464)

39. Pereira SL, Baker AJ. 2006 A mitogenomic timescale for birds detects variable phylogenetic rates of molecular evolution and refutes the standard molecular clock. *Mol. Biol. Evol.* **23**, 1731–1740. (doi:10.1093/molbev/msl038)

40. Ho SYW, Lanfear R, Bromham L, Phillips MJ, Soubrier J, Rodrigos AG, Cooper A. 2011 Time-dependent rates of molecular evolution. *Mol. Ecol.* **20**, 3087–3101. (doi:10.1111/j.1365-294X.2011.05178.x)

41. Kumar S. 2005 Molecular clocks: four decades of evolution. *Nat. Rev. Genet.* **6**, 654–662. (doi:10.1038/nrg1659)

42. Liò P, Goldman N. 1998 Models of molecular evolution and phylogeny. *Genome Res.* **8**, 1233–1244.

43. Kato T. 1980 *Perturbation theory for linear operators*. Heidelberg, Germany: Springer.