

Exploring Exploration Catastrophes in Various Network Models

David Shorten¹ and Geoff Nitschke²

¹University of Sydney, Sydney

²University of Cape Town, Rondebosch
gnitschke@cs.uct.ac.za

Abstract

It has been argued that much of evolution takes place in the absence of fitness gradients. Such periods of evolution can be analysed by examining the mutational network formed by sequences of equal fitness, that is, the neutral network. It has been demonstrated that, in large populations under a high mutation rate, the population distribution over the neutral network and average mutational robustness are given by the principal eigenvector and eigenvalue, respectively, of the network's adjacency matrix. However, little progress has been made towards understanding the manner in which the topology of the neutral network influences the resulting population distribution and robustness. In this work, we use numerical methods and network models to enhance our understanding of how populations distribute themselves over neutral networks. We demonstrate that, in the presence of certain topological features, the population will undergo an *exploration catastrophe* and become confined to a small portion of the network. These results provide insight into the behaviour of populations on neutral networks, demonstrating that neutrality does not necessarily lead to an exploration of genotype/phenotype space or an associated increase in population diversity.

Introduction

When an entity undergoes evolutionary change, much of this change may not be due to a response to selective pressure, but rather due to the discovery of variants with equivalent fitness. It has been argued that the majority of genetic change in natural organisms is due to such neutral mutations (Kimura, 1983). In *Evolutionary Computing* (EC) (Eiben and Smith, 2015), it has been found that many fitness functions result in a substantial proportion of mutations being neutral (Galván-López et al., 2011).

A variety of authors have demonstrated the substantial impact of neutrality on evolutionary dynamics (Koelle et al., 2006; Van Nimwegen et al., 1999; Newman and Engelhardt, 1998). Much of this analysis has focused on how, in instances where no advantageous mutations exist, neutrality prevents the population from getting stuck at a certain point in sequence space. Instead, it can explore the neutral network until it finds an advantageous phenotype lying adjacent to the network (Fontana and Schuster, 1998; Gavrillets,

1997). Moreover, it has been demonstrated that larger neutral networks allow for more such “stepping off points” (Wagner, 2008), facilitating the discovery of adaptive and innovative phenotypes. It has further been shown that large neutral networks allow the population to spread out and gain standing variation. This facilitates the population's adaptive response to changes in its environment (Masel and Trotter, 2010). However, there is some ambiguity as to whether neutrality is universally beneficial to evolution (Cuevas et al., 2009; Elena and Sanjuán, 2008; Galván-López et al., 2011).

The seminal work in the modeling of evolutionary dynamics is that of Erik van Nimwegen, James P. Crutchfield and Martijn Huynen (1999). By employing a straightforward model of neutral evolution, the authors demonstrated the existence of two distinct behavioural regimes. If $M\mu \ll 1$, where M is the population size and μ is the per genome mutation rate, then the population is *monomorphic* (Bloom et al., 2007). Mutations either fix or go extinct, that is they either become present in the entire population or disappear from it completely. Conversely, if $M\mu \gg 1$, then the population is *polymorphic* and mutations do not fix. The population distributes itself over a number of nodes in the network. More specifically, the population's distribution is given by the network's principal eigenvector and its average robustness (number of neutral neighbours) is given by the network's principal eigenvalue. Random walks are a very well described phenomenon (Lovász, 1993), and so this work focuses exclusively on the, more interesting, polymorphic case.

As closed-form solutions to the eigenvalues and eigenvectors of graphs do not exist, these are somewhat opaque quantities. However, various authors have been able to draw some conclusions from this result. Firstly, the population spreads out, or diffuses, over the neutral network, gaining variation (Manrubia and Cuesta, 2010; Crutchfield and Schuster, 2003; Hu et al., 2011; Masel and Trotter, 2010). Secondly, the population will become more concentrated on the “most connected” nodes and, in so doing, increase the average robustness of the population (van Nimwegen, 2006; Van Nimwegen et al., 1999; Banzhaf and Leier, 2006).

These conclusions are well founded, as the average degree of a network is a lower bound on the principal eigenvalue (Cioabă et al., 2010) and the principal eigenvector is a measure of centrality in a network, the *eigenvector centrality* (Bonacich, 1972), and, as such, assigns a non-zero centrality score to each node.

In this paper, we numerically explore instances which demonstrate that this description of the behaviour of polymorphic populations can be refined. Although the average population robustness will always be higher than the network’s average degree, we can construct examples where the population concentrates on a region of the network which does not agree with our intuition of “most connected”. Moreover, networks can be constructed where the population concentrates on a small number of vertices and does not spread out, or diffuse, over it. Take, for instance, the two networks shown in figure 1. Both of these networks consist of an Erdős-Renyi network (Erdős and Renyi, 1959) with 400 vertices and 1200 edges connected to a hub (star network), where the connection to the hub is made via one of its peripheral vertices. In the first network, the hub is of degree 45 and in the second it is of degree 70. Despite the similarity of these two networks, the equilibrium distribution of the population over them is vastly different. In the first network, the population behaves roughly as we would expect and distributes itself fairly evenly over the network, being more concentrated on the more central nodes of the Erdős-Renyi component. It is worth noting that only a very small proportion of the population (around 0.5%) is found on the hub or its neighbours. However, in the second network, around 99.5% of the population is concentrated on the hub and its neighbours. This behaviour is observed regardless of the size of the Erdős-Renyi component, so long as the average degree of this component is kept constant.

The principal eigenvectors and eigenvalues of graphs are of great importance to a variety of problems (Restrepo et al., 2007), principally synchronization phenomena and the spread of epidemics. Since the publication of van Nimwegen et. al.’s work, there has been substantial progress in approximating these quantities in terms of network properties (Goltsev et al., 2012). In this work, we build on these results in order to incorporate the above observations and intuitions into a more complete understanding of the evolution of polymorphic populations on neutral networks.

Ancel and Fontana (2000) demonstrated that, for evolving populations of RNA sequences with plastogenic congruence, the population could undergo an *exploration catastrophe*, whereby it would be confined to a small portion of the neutral network. It has recently been demonstrated (Martin et al., 2014) that the principal eigenvector is a poor measure of centrality in networks. This is due to the fact that certain structural heterogeneities can cause the eigenvector to localise on certain portions of the network, assigning almost all of its weight to these portions and very little to the rest. We

make the argument here that this localisation phenomenon has important implications for the neutral evolution of asexual populations at high mutation rates. Specifically, in neutral networks with certain topological features, the population will undergo an *exploration catastrophe*. Moreover, this phenomenon will occur without the presence of special properties of the genotypes or phenotypes, such as plastogenic congruence, and occurs independent of mutation rate. We use computational methods to confirm that this localisation of the eigenvector occurs in biologically plausible neutral networks. We further demonstrate novel modes of eigenvector localisation not yet explored in the literature.

Localisation

In the context of graph spectra, localization refers to the phenomenon whereby the normalisation weight of an eigenvector ($\sum f_i^2(\lambda)$, where λ is the eigenvalue and $\mathbf{f}(\lambda)$ is the eigenvector) is concentrated on a small number of nodes that does not scale with the size of the network (Pastor-Satorras and Castellano, 2016). Some authors have suggested using the inverse participation ratio $Y(\lambda)$.

$$Y(\lambda) = \sum_{i=1}^N f_i^4(\lambda) \quad (1)$$

as a quantitative measure of localization where, in this case, $\mathbf{f}(\lambda)$ is the normalised eigenvector. If, in the limit $N \rightarrow \infty$, $Y(\lambda) \sim 1$ then the state is localized. On the other hand, if $Y(\lambda) \rightarrow 0$ then the state is delocalized.

Results relating aspects of network topology to localization have been derived by Chung et al. (2003), Goltsev et al. (2012) and Martin et al. (2014). The result of Goltsev et al. (2012) applies to scale-free networks, whereas the result of Martin et al. (2014) applies to hubs connected to Erdős-Renyi networks (Erdős and Renyi, 1959). On the other hand, the result derived by Chung et al. (2003) is more general and applies to any network model characterised by a degree distribution. However, this result is not tight. It can only confirm localisation in the case of extremely high-degree hubs and, similarly, can only preclude localisation in the case of networks with a very homogeneous degree-distribution. For these reasons, the study of localisation in networks that are not scale-free, or hubs connected to Erdős-Renyi networks, must be done numerically.

Pastor-Satorras and Castellano (2016) demonstrated a different form of localisation which does not result in the concentration of the eigenvector on a hub. Instead, the eigenvector localises on the maximum K-core.

Definition of Localisation

As mentioned by Pastor-Satorras and Castellano (2016), there does not exist a non-arbitrary definition for localisation of the eigenvector in single network instances. However, for the purposes of this paper it will be useful to define

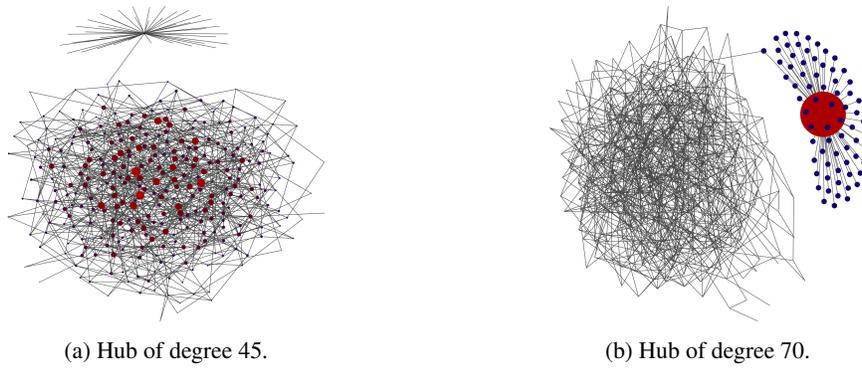


Figure 1: A localisation transition. A hub (star network) is connected to an Erdős-Renyi network by adding an edge between one of the star’s peripheral nodes and a random node of the Erdős-Renyi network. The original Erdős-Renyi networks contained 400 vertices and 1200 edges. Node sizes are proportional to the corresponding component of the principal eigenvector of the adjacency matrix which is equal to the proportion of the population found on the node.

some threshold separating localised and delocalised population distributions. We define two such thresholds and use them in different instances.

Localisation is somewhat easy to define in the case that the population has concentrated around a single hub. Here, we choose to say that if 90% or more of the population is distributed on the hub and its immediate neighbours, then the population is localised.

Trying to describe a population that is highly concentrated, but not on a hub, is slightly more challenging. Given that localisation for classes of networks is defined in terms of the inverse participation ratio, this would be natural metric with which to define localisation. However, as Pastor-Satorras and Castellano (2016) point out, in the delocalised case we expect the inverse participation ratio to be proportional to N^{-1} . On the other hand, in the localised case, we expect it to be proportional to $N^{-\beta}$, where $\beta < 1$. This dependence on the network size N , is unfortunate for our purposes, as we desire a single threshold which applies to networks of all sizes.

In order to reduce the impact of the network size N on our threshold, we define the *relative inverse participation ratio*. This is, simply, the ratio of the inverse participation ratio of the network’s principal eigenvector to what the inverse participation ratio would be if the eigenvector was distributed uniformly over the network’s nodes. If the eigenvector is distributed uniformly, then the inverse participation ratio is $1/N$. This implies that the relative inverse participation ratio can be easily calculated by multiplying the inverse participation ratio by N .

We choose to define localisation as occurring when the relative inverse participation ratio is greater than 30. In preliminary testing it was found that this value corresponded with the authors’ intuition of localisation. For comparison, Pastor-Satorras and Castellano (2016) reported on a number of real-world networks exhibiting localisation. The

lowest relative inverse participation ratio of these networks was 46.7 (The HEP network, table 1 of (Pastor-Satorras and Castellano, 2016)). The choice of 30 is fairly arbitrary. However, were a higher value used, networks with a relative inverse participation ratio higher than 30 but lower than this other value would still be exhibiting a similar phenomenon to those networks above this higher threshold. It is only the intensity of this phenomenon which would be slightly diminished. Moreover, in the authors’ preliminary work, it was found that principal eigenvectors with inverse participation ratios higher than 30 were very substantially more concentrated than the principal eigenvectors of, say, Erdős-Renyi networks (Erdős and Renyi, 1959).

Network Models

All analysis was conducted using the Python package `igraph` (Csardi and Nepusz, 2006). The calculation of the eigenvalues and eigenvectors of the adjacency matrices of graphs in `igraph` is performed using the FORTRAN 77 package ARPACK (Lehoucq et al., 1998). ARPACK implements the *implicitly restarted Arnoldi method* (Lehoucq and Sorensen, 1996) to find the eigenvalues and eigenvectors of matrices. `igraph`’s default parameters for ARPACK were used.

Barábasi-Albert Preferential Attachment

A focus of this work is the investigation of the population behaviour on around hubs. As such, it is valuable to interrogate the population distribution in network models which naturally contain hubs, and which might contain multiple hubs connected to one another. Scale-free networks (Barabási, 2016) contain multiple hub nodes. These are networks with a power-law degree distribution, that is $p(k) \sim k^{-\gamma}$ for some value of the parameter γ (usually, $2 \leq \gamma \leq 3$).

The popular Barábasi-Albert preferential attachment model (Barabási, 2016) generates connected networks

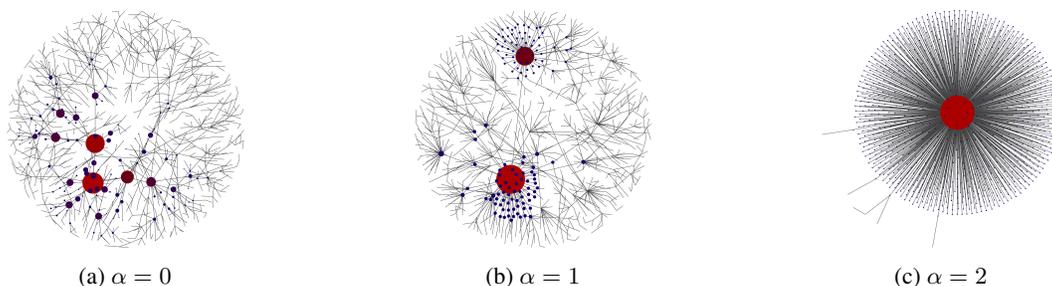


Figure 2: Barabási-Albert preferential attachment networks with $N = 200$ nodes, and three different values of the attachment parameter α . The node size is proportional to the proportion of the population that is located on it. The layout was determined by the Fruchterman-Reingold force directed layout (Fruchterman and Reingold, 1991).

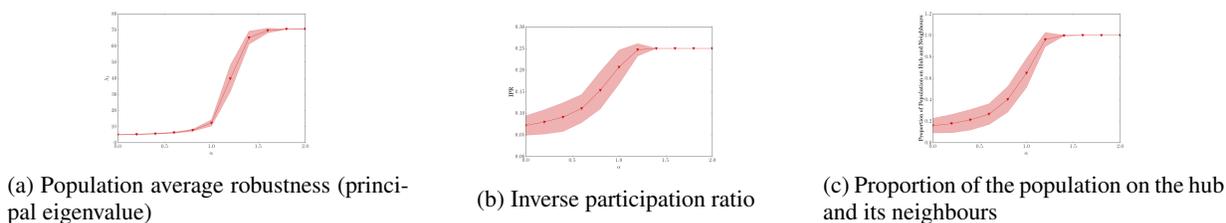


Figure 3: The average population average robustness, the inverse participation ratio, and the proportion of population on the hub node and its neighbours in Barabási-Albert preferential attachment networks. The shaded region shows the standard deviation.

which have a power-law degree distribution, although it does not uniformly sample the space of scale-free networks. Moreover, this algorithm is able to generate networks which contain hubs, but have degree distributions steeper or shallower than a power-law.

Figure 2 shows diagrams of networks generated according to this model for the three values of $\alpha = [0, 1, 2]$.

In order to examine localisation in this network model, 1000 networks were generated with $N = 5000$ nodes for the values $\alpha = [0, 0.2, 0.4, 0.6, 0.8, 1.0, 1.2, 1.4, 1.6, 1.8, 2.0]$. Figure 3 plots the principal eigenvalue, inverse participation ratio and the proportion of the population on the maximum degree node (hub) and its neighbours. We find that the population is highly localised for $\alpha > 1.2$.

Poorly Connected Random Subgraphs of Hypercubes

The realisable topology of neutral networks is constrained by the fact that the genotypes are encoded by strings of characters and that edges can only be placed between vertices whose corresponding genotypes differ by a single character. That is, are a hamming distance of one apart. Here, we analyse how this constraint influences the neutral evolution of populations. In particular, we are interested in the types of localization behaviour which can be observed.

Random subgraphs of hypercubes (or n -cubes) are well studied Reidys et al. (1997); Reidys (2009). In these models, neutral networks are created by including each node in

the neutral network with a probability θ (the symbol λ is usually used for this probability, however, we use θ to avoid confusion with the principal eigenvalue, which we denote with λ_1). Once the nodes have been assigned as being on or off the network, the connected components of the neutral network can be extracted. In the following, we study only the largest connected component of the networks, as we are interested in the exploratory behaviour of population on networks which extend over large parts of sequence space.

We hypothesize that, for sufficiently low θ , the connectivity of the network will be low enough that the population will be confined to areas of it, rather than spread evenly. In order to test this hypothesis, we generated random subgraphs of the hypercube formed by using strings of length $L = 6$ over an alphabet consisting of $A = 4$ distinct characters. The values of θ from the set $[0.1, 0.15, 0.2, 0.25, 0.3, 0.4]$ were used. Fewer larger values were chosen as preliminary experiments showed that, for large values of θ , the resulting networks were substantially larger, making analysis computationally expensive. Furthermore, figure 4 shows that the behaviour of the population is less interesting for larger values of θ . For each value of θ , 100 networks were instantiated. Various properties relating to the principal eigenvalue and eigenvector were measured. These properties are plotted in figure 4. Figure 5 shows diagrams of representative networks, with the population distribution displayed through vertex size and colour.

Figure 5 demonstrates that, at least for the selected repre-

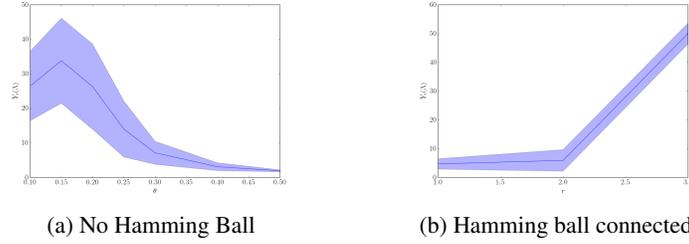


Figure 4: Relative inverse participation ratio ($Y_r(\lambda)$) of the principal eigenvector over the largest connected component of random subgraphs of an n -cube. For figure b, a hamming ball of radius ρ has been connected.

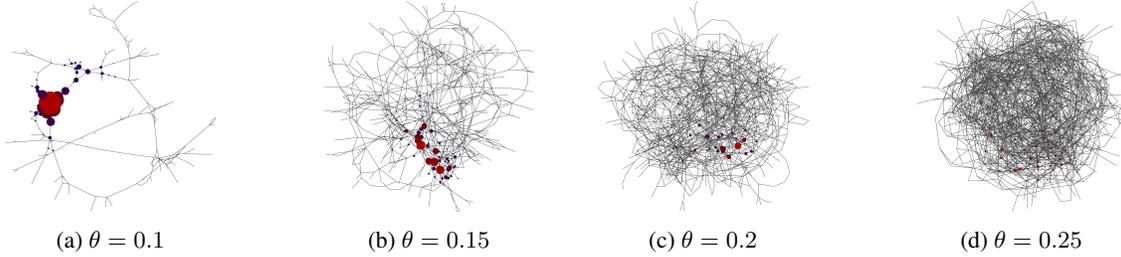


Figure 5: Network diagrams of the largest connected component of random subgraphs of an n -cube. The size of the nodes is proportional to the proportion of the principal eigenvector which is located on them.

representative networks, the population is highly concentrated on a small number of nodes for small values of θ . However, for larger values of θ it is distributed over a substantially larger number of nodes. This behaviour can be observed in figure 4a, where we see that the relative inverse participation ratio $Y_r(\lambda)$ is high enough to justify localization for low values of θ . However, it drops rapidly for increasing values of θ . It is interesting to note that $Y_r(\lambda)$ increases between $\theta = 0.1$ and $\theta = 0.15$. Further investigation revealed that the networks produced for $\theta = 0.1$ were very small (this can be observed in figure 5). We suspect that the small size of the networks prevents $Y_r(\lambda)$ from being very large, as their small size prevents the population from concentrating on a very small fraction of the network.

We also wanted to interrogate whether this mode of localisation is dissimilar from localization on a K-core (Pastor-Satorras and Castellano, 2016). In order to do this, we generated a further 20 networks each for $\theta = 0.15$ and $\theta = 0.2$ and recorded the proportion of the population residing on the maximum K-core. For $\theta = 0.15$ this value varied between 0.16 and 0.54, with a mean of 0.35 and a standard deviation of 0.12. For $\theta = 0.2$ it varied between 0.01 and 0.88, with a mean of 0.7 and a standard deviation of 0.23. The existence of networks in which the eigenvector is so weakly concentrated on the maximum K-core indicates that, at least in some cases, the mode of localization is slightly different to that described by Pastor-Satorras and Castellano (2016).

We were interested as to whether weak connectivity between parts of the network would have a similar effect on

the eigenvectors of graphs that are not embedded in hamming space. To this end, we generated pairs of Erdős-Renyi (Erdős and Renyi, 1959) networks, each with $|V| = 100$ vertices and $|E| = 200$ edges. These pairs of networks were connected by a single edge, connecting two randomly chosen vertices. Although, in some instances, the eigenvector was spread out evenly over the two original networks, in others, it was almost completely concentrated on a single network. Figure 6 shows a diagram of an instance where the eigenvector was heavily concentrated on a single network in the pair. Further analysis showed that, when analysed independently (without the single connecting edge), the one network had a slightly higher principal eigenvalue than the other, due to the randomness involved in the generation of the networks. The eigenvector of the combined network was concentrated on this network. This effect makes sense in terms of natural evolution, as the population is able to achieve a higher level of robustness on one of the pair of networks. Therefore, were a fraction of the population to be located on the sub-network with a lower principal eigenvalue, it would be out-competed by the fraction of the population located on the other sub-network. Moreover, the single connecting edge is insufficient to allow a large flow of mutants from the one sub-network to the other.

Hamming Balls on Random Subgraphs of Hypercubes

It is worthwhile querying whether the eigenvectors of random subgraphs of a hypercube can undergo a localization

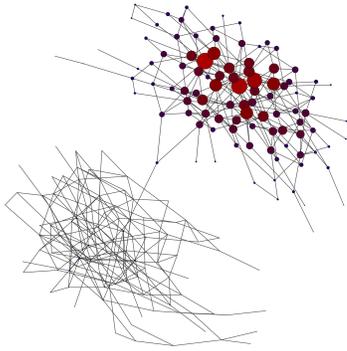


Figure 6: Network formed by connecting two Erdős-Renyi (Erdős and Renyi, 1959) networks, each with $|V| = 100$ vertices and $|E| = 200$ edges by a single edge. The size of the nodes is proportional to the proportion of the principal eigenvector which is located on them.

onto a hub, as already discussed for networks in general. Of the analytic results for localization discussed earlier, the weakest was that of Martin et al. (2014), where it was required that $\sqrt{k_{max}} > \langle q \rangle$, where $\langle q \rangle$ is the average degree of the network excluding the hub. For random subgraphs of the hypercube constructed from sequences of length L and an alphabet of size A , where a given vertex is included in the network with probability θ , if we connect a hub of maximum possible degree then this condition implies $\theta < 1/\sqrt{L(A-1)}$. For the networks which we studied in the previous section, with $L = 6$ and $A = 4$, this would imply that $\theta < 0.24$. In that section we found that, for that value of θ , the eigenvector was already somewhat localized (see figure 4a). Moreover, given that increasing L and A will decrease the bound on θ , similar effects are probable for the random subgraphs of larger hypercubes. As the eigenvectors of these networks are already under the influence of a certain mode of localisation, studying the effects of connecting hubs to them could lead to ambiguous results due to the multiple modes of localisation.

There is, however, a natural generalisation to a hub when considering subgraphs of hypercubes: the hamming ball. A hamming ball is the network composed of all nodes in the hypercube within a certain radius ρ of a specific sequence. A star of maximum possible degree in the hypercube is then a hamming ball of radius $\rho = 1$. We would expect hamming balls to produce populations with high average genetic robustness. Bornberg-Bauer and Chan (1999) studied them as an abstraction for the structure of protein neutral networks. They found that the population tended to concentrate on the inner nodes of the ball, increasing the population’s average robustness. More recently, Bollobás et al. (2016) showed that, for a given number of nodes, a hamming ball arrangement maximised the principal eigenvalue of the resulting network.

We, therefore, thought it worthwhile to investigate whether, by connecting hamming balls to otherwise delocalised graphs, localisation could occur.

We generated random subgraphs of the hypercube formed by strings of alphabet size $A = 2$ and length $L = 13$. The smaller alphabet size was chosen as, in preliminary testing, it was found that the size of the hamming balls increased too rapidly for larger values of A . This made the analysis too computationally expensive. The longer length was chosen to allow for large subgraphs, given the small size of the alphabet. The high value of $\theta = 0.4$ was chosen to discourage localisation behaviour of the eigenvector without the presence of the hamming ball. To each graph was connected a hamming ball of radius ρ . The values of ρ from the set $[0, 1, 2]$ were used. For each value of ρ , 100 networks were generated. Diagrams of the networks are not shown as it was found that the resulting networks were too large to allow for these diagrams to be informative.

Figure 4b shows the relative inverse participation ratio $Y_r(\lambda)$. Between $\rho = 2$ and $\rho = 3$ we see a sharp increase in this value, representing a localisation transition.

Discussion

In this work, we set out to incorporate and build upon recent results concerning the behaviour of the principal eigenvectors, and associated eigenvalues, of the adjacency matrices of networks in the context of the study of the dynamics of polymorphic populations evolving asexually on neutral networks.

Much of the discussion surrounding neutral evolution has functioned on the assumption that the population spreads out over the network, gaining variation and exploring sequence space (Lauring and Andino, 2010). The population will, further, “evolve toward regions denser in neutral genotypes”(Aguirre et al., 2009). However, the behaviour on networks with a heterogeneous structure can be substantially different. In the presence of certain structural heterogeneities, the principal eigenvector of the adjacency matrix of the network localises. This localisation of the principal eigenvector leads to an exploration catastrophe as described by AnceI et al. (2000), whereby the population becomes concentrated on a small region of the network. It is interesting to note that the localisation transition described here is dependent on the topology of the network and is independent of the mutation rate. On the other hand, the localisation transition described by AnceI et al. (2000), along with other localisation-delocalisation transitions studied in quasispecies theory (Tejero et al., 2011; Summers and Litwin, 2006), are dependent on the mutation rate and are studied in the context of a fixed fitness landscape. Nevertheless, such an error catastrophe has important ramifications for the study of populations evolving at high mutation rates.

Given that it is suspected that much of evolution occurs on neutral networks (Nei, 2005) along with the importance

of mutational robustness to the survival of organisms and its relationship with evolvability, understanding the impact of the topology of neutral networks on the dynamics of neutral evolution and the resulting robustness of organisms is of great importance. This work has provided insight into these issues in the case of polymorphic populations: large populations evolving at high mutation rate. The directed, neutral, evolution of bio-molecules (Currin et al., 2015; Jäckel and Hilvert, 2010) along with viruses overcoming immunity through neutral evolution (van Nimwegen, 2006) fall within this category. These results have potential applicability to these problems. For instance, the neutral evolution of large libraries of molecules (Kaltenbach and Tokuriki, 2014) will be greatly aided by delocalization, whereas a virus's attempt to escape immunity might be thwarted if its population localizes on a hub.

Probably the largest limitation of the work presented here is that all the studied networks were artificially generated from network models. Future work will focus on the applicability of these results to biological evolution. Progress has already been made on the study of the neutral networks of influenza, and this work is presented in the master's thesis of the first author (Shorten, 2017). The present focus is on obtaining a high-resolution data-set of the sequences of an influenza quasispecies evolving within a given host. Future work will also investigate the transient behaviour of populations approaching their equilibrium distribution.

Conclusion

This paper investigated the manner in which neutral network topology influences the resulting population distribution and robustness during neutral evolution at high mutation rates in large populations without recombination. In such cases, the population distribution is given by the principal eigenvector of the adjacency matrix of the neutral network and, similarly, the average mutational robustness of the individuals in the population is given by the principal eigenvalue (Van Nimwegen et al., 1999). Hence, we utilized, and built upon, recent results concerning the behaviour of these values from studies concerning the spread of epidemics on networks (Goltsev et al., 2012) as well as more general work (Martin et al., 2014).

For neutral networks with certain structural heterogeneities, it was found that the population could undergo an exploration catastrophe, whereby it becomes localised on a small number of nodes in the network. These results are particularly relevant to various arguments concerning the relationship between robustness and evolvability (Masel and Trotter, 2010; Wagner, 2008), which make the assumption that populations evolving at high mutation rate disperse over their neutral networks.

These results are relevant to the directed evolution of bio-molecules (Currin et al., 2015; Jäckel and Hilvert, 2010), where they can be used to evolve more robust molecules as

well as facilitate the evolution of greater variety. Moreover, they can also further our understanding of the factors that allow viruses to escape immunity along neutral networks (van Nimwegen, 2006).

Acknowledgements

Computations were performed using facilities provided by the University of Cape Town's ICTS High Performance Computing team: <http://hpc.uct.ac.za>

References

- Aguirre, J., Buldú, J. M., and Manrubia, S. C. (2009). Evolutionary dynamics on networks of selectively neutral genotypes: Effects of topology and sequence stability. *Physical Review E*, 80(6):066112.
- Ancel, L. and Fontana, W. (2000). Plasticity, evolvability, and modularity in rna. *Journal of Experimental Zoology*, 288(1):242–283.
- Ancel, L. W., Fontana, W., et al. (2000). Plasticity, evolvability, and modularity in rna. *Journal of Experimental Zoology*, 288(3):242–283.
- Banzhaf, W. and Leier, A. (2006). Evolution on neutral networks in genetic programming. *Genetic programming theory and practice III*, pages 207–221.
- Barabási, A.-L. (2016). *Network science*. Cambridge university press.
- Bloom, J. D., Lu, Z., Chen, D., Raval, A., Venturelli, O. S., and Arnold, F. H. (2007). Evolution favors protein mutational robustness in sufficiently large populations. *BMC biology*, 5(1):1.
- Bollobás, B., Lee, J., and Letzter, S. (2016). Eigenvalues of subgraphs of the cube. *arXiv preprint arXiv:1605.06360*.
- Bonacich, P. (1972). Factoring and weighting approaches to status scores and clique identification. *Journal of Mathematical Sociology*, 2(1):113–120.
- Bornberg-Bauer, E. and Chan, H. S. (1999). Modeling evolutionary landscapes: mutational stability, topology, and superfunnels in sequence space. *Proceedings of the National Academy of Sciences*, 96(19):10689–10694.
- Chung, F., Lu, L., and Vu, V. (2003). Spectra of random graphs with given expected degrees. *Proceedings of the National Academy of Sciences*, 100(11):6313–6318.
- Cioabă, S. M., Van Dam, E. R., Koolen, J. H., and Lee, J.-H. (2010). A lower bound for the spectral radius of graphs with fixed diameter. *European Journal of Combinatorics*, 31(6):1560–1566.
- Crutchfield, J. P. and Schuster, P. (2003). *Evolutionary dynamics: exploring the interplay of selection, accident, neutrality, and function*. Oxford University Press, USA.
- Csardi, G. and Nepusz, T. (2006). The igraph software package for complex network research. *InterJournal, Complex Systems*:1695.

- Cuevas, J., Moya, A., and Sanjuán, R. (2009). A genetic background with low mutational robustness is associated with increased adaptability to a novel host in an rna virus. *Journal of evolutionary biology*, 22(10):2041–2048.
- Currin, A., Swainston, N., Day, P. J., and Kell, D. B. (2015). Synthetic biology for the directed evolution of protein biocatalysts: navigating sequence space intelligently. *Chemical Society Reviews*, 44(5):1172–1239.
- Eiben, A. and Smith, J. (2015). *Introduction to Evolutionary Computing*, volume 53. Springer.
- Elena, S. F. and Sanjuán, R. (2008). The effect of genetic robustness on evolvability in digital organisms. *BMC evolutionary biology*, 8(1):284.
- Erdős, P. and Renyi, A. (1959). On random graphs i. *Publ. Math. Debrecen*, 6:290–297.
- Fontana, W. and Schuster, P. (1998). Continuity in evolution: on the nature of transitions. *Science*, 280:1451–1455.
- Fruchterman, T. M. and Reingold, E. M. (1991). Graph drawing by force-directed placement. *Software: Practice and experience*, 21(11):1129–1164.
- Galván-López, E., Poli, R., Kattan, A., Neill, M., and Brabazon, A. (2011). Neutrality in evolutionary algorithms... what do we know? *Evolving Systems*, 2(3):145–163.
- Gavrilets, S. (1997). Evolution and speciation on holey adaptive landscapes. *Trends in ecology & evolution*, 12(8):307–312.
- Goltsev, A. V., Dorogovtsev, S. N., Oliveira, J., and Mendes, J. F. (2012). Localization and spreading of diseases in complex networks. *Physical review letters*, 109(12):128702.
- Hu, T., Payne, J., Banzhaf, W., and Moore, J. (2011). Robustness, evolvability, and accessibility in linear genetic programming. *Genetic Programming*, pages 13–24.
- Jäckel, C. and Hilvert, D. (2010). Biocatalysts by evolution. *Current opinion in biotechnology*, 21(6):753–759.
- Kaltenbach, M. and Tokuriki, N. (2014). Generation of effective libraries by neutral drift. *Directed Evolution Library Creation: Methods and Protocols*, pages 69–81.
- Kimura, M. (1983). *The neutral theory of molecular evolution*. Cambridge University Press.
- Koelle, K., Cobey, S., Grenfell, B., and Pascual, M. (2006). Epochal evolution shapes the phylodynamics of inter-pandemic influenza a (h3n2) in humans. *Science*, 314(5807):1898–1903.
- Lauring, A. S. and Andino, R. (2010). Quasispecies theory and the behavior of rna viruses. *PLoS pathogens*, 6(7):e1001005.
- Lehoucq, R. B. and Sorensen, D. C. (1996). Deflation techniques for an implicitly restarted arnoldi iteration. *SIAM Journal on Matrix Analysis and Applications*, 17(4):789–821.
- Lehoucq, R. B., Sorensen, D. C., and Yang, C. (1998). *ARPACK users' guide: solution of large-scale eigenvalue problems with implicitly restarted Arnoldi methods*. SIAM.
- Lovász, L. (1993). Random walks on graphs. *Combinatorics, Paul erdos is eighty*, 2:1–46.
- Manrubia, S. C. and Cuesta, J. A. (2010). Neutral networks of genotypes: Evolution behind the curtain. *arXiv preprint arXiv:1002.2745*.
- Martin, T., Zhang, X., and Newman, M. (2014). Localization and centrality in networks. *Physical Review E*, 90(5):052808.
- Masel, J. and Trotter, M. V. (2010). Robustness and evolvability. *Trends in Genetics*, 26(9):406–414.
- Nei, M. (2005). Selectionism and neutralism in molecular evolution. *Molecular biology and evolution*, 22(12):2318–2342.
- Newman, M. E. and Engelhardt, R. (1998). Effects of selective neutrality on the evolution of molecular species. *Proceedings of the Royal Society of London B: Biological Sciences*, 265(1403):1333–1338.
- Pastor-Satorras, R. and Castellano, C. (2016). Distinct types of eigenvector localization in networks. *Scientific reports*, 6.
- Reidys, C., Stadler, P. F., and Schuster, P. (1997). Generic properties of combinatory maps: neutral networks of rna secondary structures. *Bulletin of mathematical biology*, 59(2):339–397.
- Reidys, C. M. (2009). Large components in random induced subgraphs of n-cubes. *Discrete Mathematics*, 309(10):3113–3124.
- Restrepo, J. G., Ott, E., and Hunt, B. R. (2007). Approximating the largest eigenvalue of network adjacency matrices. *Physical Review E*, 76(5):056119.
- Shorten, D. (2017). Spectral analysis of neutral evolution. Master's thesis, University of Cape Town.
- Summers, J. and Litwin, S. (2006). Examining the theory of error catastrophe. *Journal of virology*, 80(1):20–26.
- Tejero, H., Marín, A., and Montero, F. (2011). The relationship between the error catastrophe, survival of the flattest, and natural selection. *BMC evolutionary biology*, 11(1):2.
- van Nimwegen, E. (2006). Influenza escapes immunity along neutral networks. *Science*, 314(5807):1884–1886.
- Van Nimwegen, E., Crutchfield, J., and Huynen, M. (1999). Neutral evolution of mutational robustness. *Proceedings of the National Academy of Sciences*, 96(17):9716–9720.
- Wagner, A. (2008). Robustness and evolvability: a paradox resolved. *Proceedings of the Royal Society of London B: Biological Sciences*, 275(1630):91–100.