Application of Small-World Mutation Topologies to an Artificial Life System

Alastair Droop[1,2] and Simon Hickinbotham[1,3]

1: YCCSA, Ron Cooke Hub
2: YCR Cancer Research Unit, Department of Biology
3: Department of Computer Science
The University of York, Heslington, York YO10 5DD, UK
email: alastair.droop@york.ac.uk & sjh@cs.york.ac.uk

Abstract

The mutation networks observed in biological systems have the properties of small-world networks. These properties of short average path length and high transitivity confer a favourable exploration of mutation space. Any evolvable string-based ALife system (for example stringmol, typogenetics, Tierra, or Avida) uses a substitution network either implicitly or explicitly. Current ALife simulations use either regular or random mutations schemes. We have previously discussed the requirement for small-world substitution networks for ALife simulations. In this paper, we explore the effects of rewiring the stringmol mutation lattice on the evolution of a self-replicating molecule.

Introduction

Mutation is an essential component of any evolvable system, allowing it to explore its fitness landscape and therefore to evolve. The evolutionary dynamics of a system are thus critically dependent upon its mutation strategy.

Amino acid substitution matrices (Dayhoff et al., 1978; Henikoff and Henikoff, 1992) give an indication of the likelihood of observing an amino acid substitution in homologous proteins. Ideally, a substitution matrix should allow any token to mutate to any other token relatively easily (thus allowing a rapid exploration of the fitness landscape); whilst simultaneously favouring mutations to tokens of similar function (thus minimising the chance of deleterious mutations). Networks that exhibit these properties of short average path length and high clustering coefficient were described by Watts and Strogatz (Watts and Strogatz, 1998). We have previously demonstrated that biological mutation networks exhibit these small-world properties (Droop and Hickinbotham, 2011).

Although biological mutation networks exhibit small-world properties, ALife simulation mutation schemes do not. The typogenetics (Gwak and Wee, 2007) and Avida (Johnson and Wilke, 2004) systems use essentially random mutation schemes. By contrast, the stringmol (Hickinbotham et al., 2010a,b) and Tierra (Ray, 1991) systems use regular mutation networks. Figure 1 shows the mutation lattices used by the stringmol and Tierra systems.

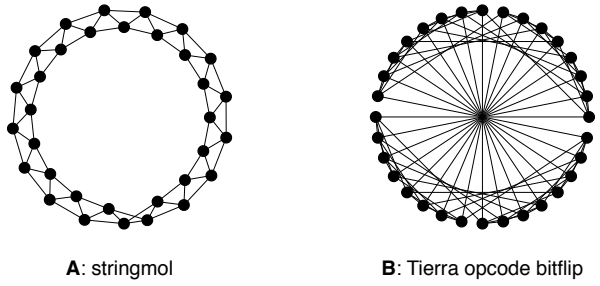

Figure 1: The mutation networks used by the stringmol and Tierra ALife simulations. The stringmol lattice shown here has $k = 4$.

The stringmol mutation network is constructed as a complete lattice with each node connected to its k nearest neighbours (in this case $k = 4$). The Tierra network is based upon binary bit flip operations: each opcode can flip a single binary digit. Any mutation scheme can be represented by a graph where possible substitutions between tokens are represented as edges. The Tierra opcode lattice in figure 1 shows that although the neighbours for each opcode were carefully chosen to allow 'sensible' mutations, the mutation network topology is nonetheless regular.

To test this idea, we implemented a small-world mutation network topology for the stringmol system using the rewiring scheme devised by Watts and Strogatz (Watts and Strogatz, 1998). Here, we present the findings from this study, and discuss the effects of different mutation models upon the stringmol system.

Methods

Multiple networks were created for the stringmol system using the Watts and Strogatz (1998) rewiring model as described previously (Droop and Hickinbotham, 2011). The stringmol regular mutation lattice with $k = 4$ was used to perform all analyses. The mutation rate is fixed for all trials. Four different rewiring probabilities ($p_{reg} = 0$, $p_{small} = 0.1$, $p_{mid} = 0.3$ and $p_{rand} = 1$) were used when cre-

ating networks. The p_{reg} network is the same regular lattice as used previously in stringmol. The p_{small} and p_{mid} networks fall within in the small-world region for networks of this size (Droop and Hickinbotham, 2011). The p_{rand} network is completely random. 100 rewiring replicates were created for p_{small}, p_{mid} and p_{rand}; whilst only a single trial of p_{reg} was performed (as there is only one possible network). For each replicate experiment, the stringmol simulation run for a maximum of 1.2×10^9 time steps with 300 trials.

Summary statistics were collected for each trial. Four statistics were used to describe each individual trial. The *life time* is the total number of time steps over which the trial survives. The *epoch count* is the number of epochs (defined as a continuous period of time in which a particular string species is the most common). The *epoch length* is the length of each epoch. The *edit distance* is a measure of the mutational distance that the trial has been able to cover. The edit distance is calculated as the Smith-Waterman alignment score between the initial string and the dominant string (only counting strings greater than 10 letters, thus ignoring short 'pathogenic' strings) for the last epoch of length ≥ 50000, normalised to the length of the string.

Results

Box plots (McGill et al., 1978) of the results for the summary statistics outlined above are given in figure 2. The trend across figures 2**A**, **B** and **C** is consistent: the more random the mutation network, the shorter the simulation life time and fewer (shorter) epochs are present. The reduction in life time indicates that large amounts of energy are wasted in the simulation (presumably on harmful mutations). Similarly, the reduction in the epoch count indicates that fewer successful species are created during the trial. Figure 2**D**, however, shows that the mutational distance covered by the runs *increases* with increasing randomness in the mutation networks: this demonstrates that although shorter, the runs with more random mutation schemes can produce more varied successful species.

Summary & Conclusion

Taken together, the results shown here suggest that the optimal mutation strategy for the stringmol system is neither at p_{reg} or p_{rand}; rather somewhere in between: in the small-world rewiring region. This work provides an experimental validation of the argument presented previously (Droop and Hickinbotham, 2011).

Acknowledgements

Alastair Droop is funded by Yorkshire Cancer Research. Simon Hickinbotham is funded by the Plazzmid project, EPSRC grant EP/F031033/1.

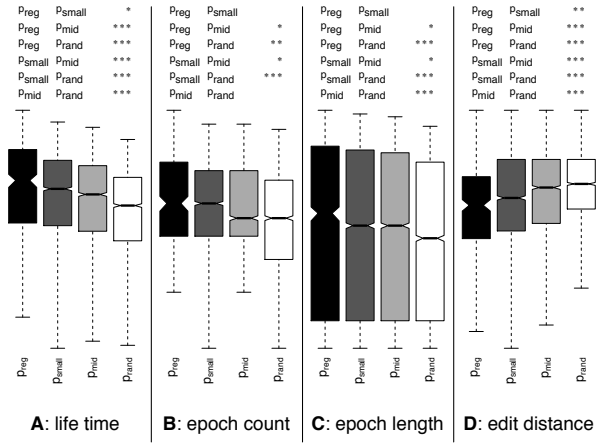

Figure 2: Box plots of four summary statistics for each experiment. Each plot is drawn using a logarithmic y-axis (values omitted for clarity). 1, 2 and 3 stars represent t-test p-values of ≤ 0.05, 0.005 and 0.0005 respectively.

References

Dayhoff, M. O., Schwartz, R., and Orcutt, B. C. (1978). A model of evolutionary change in proteins. *Atlas of protein sequence and structure*, 5(suppliment 3):345–358.

Droop, A. and Hickinbotham, S. (2011). Properties of biological mutation networks and their implications for artificial life. *ALife Journal*, (in press).

Gwak, C. and Wee, K. (2007). Construction of hypercycles in typogenetics with evolutionary algorithms. In *Proceedings of the 9th European conference on Advances in artificial life*, ECAL'07, pages 1060–1068, Berlin, Heidelberg. Springer-Verlag.

Henikoff, S. and Henikoff, J. G. (1992). Amino acid substitution matrices from protein blocks. *Proceedings of the National Academy of Sciences of the United States of America*, 89(22):10915–10919.

Hickinbotham, S., Clark, E., Stepney, S., Clarke, T., Nellis, A., Pay, M., and Young, P. (2010a). Diversity from a monoculture: Effects of mutation-on-copy in a string-based artificial chemistry. In *ALife XII*, pages 24–31. MIT Press.

Hickinbotham, S., Clark, E., Stepney, S., Clarke, T., Nellis, A., Pay, M., and Young, P. (2010b). Specification of the stringmol chemical programming language version 0.2. Technical Report YCS-2010-458, The University of York.

Johnson, T. J. and Wilke, C. O. (2004). Evolution of resource competition between mutually dependent digital organisms. *Artificial Life*, 10(2):145–156.

McGill, R., Tukey, J. W., and Larsen, W. A. (1978). Variations of box plots. *The American Statistician*, 32(1):12–16.

Ray, T. S. (1991). An approach to the synthesis of life. In *Artificial Life II*, pages 371–408. Addison-Wesley.

Watts, D. J. and Strogatz, S. H. (1998). Collective dynamics of 'small-world' networks. *Nature*, 393(6684):440–442.