Autocatalytic Replication of Polymers Revisited

Ben Blundell[1]

[1]Evolutionary and Adaptive Systems
University of Sussex, Brighton, BN1 9RH, UK
ben@benblundell.com

Abstract

A simple computational model for the emergence of autocatalytic sets as described in (Farmer et al., 1986) is re-implemented. Results are found to generally agree with the major theme in the original work: increasing the initial polymer variety in a toy chemical soup scenario increases the likelihood that a complex autocatalytic set will suddenly bootstrap itself into existence. Quantitatively, however, critical probabilities derived from this careful re-implementation are very much higher than those reported in the original work. A full resolution is not reached, but a theoretical argument supports the simulation results gained in this instance.

Introduction

The principle of an autocatalytic set, a set of molecules which collectively catalyses its own production, holds intuitive interest. There exists obvious relations to primitive metabolic systems, and contemporary minimal definitions of life such as autopoiesis (McMullin, 1999).

By achieving catalytic closure, a set of relatively inert molecules can organise into a self-sustaining identity, a persistent presence in a chemical soup.

Different questions can be asked of autocatalytic sets. In general, one might be interested in (1) *how a set came to be* and the preconditions necessary for it's emergence, (2) *which critical molecular species the set consists of*, or (3) *how the set chemically operates in real physical space and time*.

Original work on autocatalytic sets (Kauffman, 1986; Farmer et al., 1986) pursued the first question as the main point of interest, although all questions are inter-related to some extent. Question 2 has recently been given a deep formal treatment (Hordijk and Steel, 2004). Question 3 is of considerable depth and of most contemporary interest, involving concepts such as dynamics, spatial compartmentalisation, reaction kinetics and concentrations in particular physical autocatalytic instantiations (see, for example (Ono and Ikegami, 2000) for application in a spatial abstract cell model).

This paper describes a careful re-implementation of the original (graph theoretic) model investigating the inevitable emergence of complex autocatalytic sets (Farmer et al., 1986). Section 1 recaps the motivations and assumptions of the original model. Section 2 describes in detail the re-implementation carried out. Remaining sections present and discuss the results, which generally follow the same qualitative pattern as original results, but differ by a factor some 100 in quantitative predictions of the critical probability of autocatalysis.

1 Original Work

The original work on autocatalytic sets by Stuart Kauffman is concerned with making a tentative link to the grand problem of the origin of life itself (Kauffman, 1986, 1993) and levelling a respectable argument against entrenched explanations of template based replication.

In the original model (Farmer et al., 1986), the emergence of autocatalytic sets is investigated as a connectivity feature of directed graphs.

A reaction graph captures the core chemical relationships in a system of polymers, expressing the *reaction possibilities* in that system. Operational details such as space, time and quantity are not represented in this canonical description. The chemical system is assumed to exist in a well-stirred overflowing reactor environment.

The central idea is built upon the *phase transition* phenomena in connectivity problems. As systems become increasingly connected a critical limit is reached when, very suddenly, each component of the system is connected directly or indirectly to every other. A large component crystallises from a mass of independent sub-systems.

By the same logic, when a reaction network is expressed as a reaction graph, there must exist some critical catalytic connectivity beyond which each polymer will directly or indirectly catalyse every other - at which point the existence of a complex autocatalytic set can be inferred with almost certainty.

The original model focuses on finding this critical connectivity. A basic reaction system - where polymers consist of directional strings of characters - is successively grown from an original 'firing disk' (food set). In this scenario,

the reaction system is always autocatalytic in a strict sense, but criticality is judged when the rate of change of polymer species becomes exponential (autocatalytic networks which continually create large complex proteins were of prime interest to the authors).

Significant assumptions of the model include the prerequisite of flow reactor conditions and the assumption that the distribution of catalytic capacities in peptide space can be modelled by a fixed probability *P* that any one polymer will catalyse any other.

Farmer et al. find that that the critical value of *P* required for an autocatalytic set decreases as the initial polymer variation in the system increases, lending support to their general autocatalytic account for the origin of life from a sufficiently diverse pre-biotic chemical soup.

2 Re-implementation

Original Graph Growth Algorithm

For clarity, the original graph growth algorithm is presented below. Square braces represent cross-references to the more detailed implementation to follow.

> *Our rule for random assignment of reactions is implemented as follows: For a given starting list of molecular species, we compute the maximum number of allowed condensation and cleavage reactions by counting the number of distinguishable combinations of string concatenations and string cleavages* **[see Note 2]**. *The number of reactions that we actually assign is obtained by multiplying the number of allowed reactions of each type by a probability P. To assign condensation reactions, we chose two molecules at random* **[Box 1]**, *while for cleavage reactions we chose a molecule and a cleavage point at random* **[Box 2]**. *In both cases enzymes are chosen at random from the set of species currently present.*
>
> *Assignment of reactions can be viewed as a dynamical process. We initialise the system by choosing a starting list, called the "firing disk", typically chosen to be all possible strings shorter than a given length L* **[Iteration 0, Step 1]**. *Reactions within the firing disk are assigned as described above* **[Iteration 0, Step 2]**. *Condensation reactions may generate new species outside the firing disk, thereby expanding the list* **[Iteration 0, Steps 3 and 4]**. *The introduction of new species creates new reaction possibilities; to take these into account, on the next time step we count the number of combinatorial possibilities involving the new species* **[Iterations 1 to 1000, Steps 2 and 3]**. *Multiplying by P gives the number of new reactions* **[Iterations 1 to 1000, Steps 6,7,8,9]**. *This process is repeated on subsequent time steps. As long as new species are created on each step the graph continues to grow; otherwise growth stops.* (Farmer et al., 1986), p. 54

Graph Growth Algorithm as Implemented

Definitions

S Set of all distinct polymer species currently in the system. Initially empty.

N Set of distinct polymer species, new on the current iteration. Initially empty.

s_n Size of set S. Number of distinct polymer species in the system.

n_n Size of set N. Number of new distinct polymer species on current iteration.

B Alphabet size of polymers.

M Order of initial firing disk (or 'maximum sized polymer' in firing disk, also referred to as L_f elsewhere in this paper).

P Probability that a random polymer catalyses an arbitrary reaction.

Iteration 0

1. Seed firing disk. Make set S contain all possible polymers of alphabet size B up to length M. S will contain a total of $s_n = \sum_{L=1}^{M} B^L$ polymers.
2. Calculate the number of *distinct* condensation reactions *possible* in the firing disk, $R^*_{cond} = s_n \times s_n$. (See Note 2 below).
3. Calculate a number of random condensation reactions to *assign*, $R^+_{cond} = P \times R^*_{cond}$.
4. Assign R^+_{cond} condensation reactions to the firing disk as in Box 1 below, thereby expanding the disk. Cleavage reactions need not be assigned here. The graph currently consists of all possible polymers up to size M, and thus cleavage reactions cannot introduce any new species at this stage.

Box 1: condensation reaction assignment

1. Pick a random polymer of sequence a from S
2. Pick another random polymer of sequence b from S ($a = b$ is allowed)
3. Concatenate a and b to create polymer sequence $p = a+b$
4. If $p \notin S$ then add p to N (the set of new polymers created)
5. Otherwise, disregard reaction. The condensation does not create a new species and thus is of no significance. Another reaction is not assigned in place.

Iterations 1 to 1000

1. Record n_n, the number of new species for this iteration.
2. Calculate the number of *possible* cleavage reactions that the set of new polymers N introduces to the system. $R^*_{cleav} = \sum_{L=2}^{J}[N(L) \times (L-1)]$, where J is the maximum length polymer in set N, and $N(L)$ is the total number of polymers of size L in N. (See Note 2 below).
3. Calculate the number of *possible* condensation reactions that the set of new polymers N introduces to the system. $R^*_{cond} = n_n n_n + 2n_n s_n$ (See Note 2 below).
4. Add the new polymers N to the total species set S, so S becomes $S \cup N$.
5. Set $N = \phi$
6. Calculate a number of random condensation reactions to assign for this iteration, $R^+_{cond} = P \times R^*_{cond}$.
7. Calculate a number of random cleavage reactions to assign for this iteration, $R^+_{cleav} = P \times R^*_{cleav}$.
8. Assign R^+_{cond} condensation reactions as in Box 1 above.
9. Assign R^+_{cleav} cleavage reactions as in Box 2 below.
10. Goto step 1. Perform next iteration.

Box 2: cleavage reaction assignment

1. Pick a random polymer of sequence p from S
2. If the length of $p < 2$, disregard.
3. Pick a random break point on p, splitting it into two fragment substrate sequences $p = a + b$
4. If $a \notin S$ then add a to N, otherwise disregard
5. If $b \notin S$ then add b to N, otherwise disregard

Determining Criticality of Graph Graph is SUPRA critical if, during the 1000 graph iterations,

1. The number of polymer species in the system, $n_s > 10^5$
2. A polymer in the system exceeds a length of 1000
3. The new condensation reactions *possible* on any iteration, $R^*_{cond} > 2 \times 10^9$
4. The new condensation reactions *assigned* on any iteration $R^+_{cond} > 5 \times 10^5$

If 1000 iterations are completed, the graph is still judged SUPRA critical if $n_s > 5s_f$, i.e. if the number of species in the system after the last iteration are five times greater the number of species in the firing disk, s_f.

Otherwise, the graph is judged as SUB critical.

Note 1: Distinct Reactions In a fairly common sense way, this study regards two reactions $a + b \rightleftharpoons c$ and $x + y \rightleftharpoons z$ as *distinct* if their substrates do not match. That is to say, they are only the *same* reaction if $a = x$ and $b = y$ (and thus $c = z$).

Two reactions may of course have the *same product* (e.g $aaa + a \rightleftharpoons aaaa$ and $aa + aa \rightleftharpoons aaaa$) and still remain distinct. If two reactions have *different products*, the reactions will *certainly* be distinct.

Note 2: Counting Distinct Condensation and Cleavage Reactions Although the original description does not specify exactly, this paper calculates the number of new condensation and cleavage reaction possibilities at each iteration in a straightforward way.

The number of *new possible cleavage reactions* introduced into the system is simply the sum of the number of ways the new species can be broken apart. Each new species, by definition, is a product not encountered before. By Note 1 above, breaking this new product on any of it's bonds will in turn reveal a substrate *combination* not encountered before, and thus a distinct reaction.

The number of *new possible condensation reactions* introduced into the system by the new species can be calculated in two parts. Firstly, each new species polymer can be combined in two ways with every member of the existing species (by concatenating to the left and right hand side of the existing species). By Note 1 above, both of these condensation reactions are distinct, new reactions, because either the left or right hand substrate is a new species. New species thus make possible at total of $2n_n s_n$ new condensation reactions with the existing species.

Secondly, the new species can be combined amongst themselves. Each new species polymer can be appended to the left or right hand side of every other new species polymer, including itself. However, because *both* substrates in these reactions are new species, there will be some double counting.

For example, new species g may be combined with new species h by left concatenation $g + h$ or by right concatenation $h + g$, yielding two distinct concatenations. When h is considered and combined with g to the right and left the mirror is true, yielding two non-distinct concatenations. The distinct concatenations are thus just the set of left-hand concatenations between the new species, a total of $n_n n_n$ reactions.

Estimation of P_{crit} algorithm

As in the original paper, estimation of the critical probability of catalysis, P_{crit}, is performed by using a simple trial-and-error algorithm.

For a graph of alphabet size B and firing disk order M, 10 independent estimates of P_{crit} are made as in Box 3 below and then averaged to provide a more reliable result.

Box 3: To estimate P_{crit}

1. Set P_{min} to a low probability, known to cause the graph to go firmly SUB critical.
2. Set P_{max} to be a higher probability known to cause the graph to go firmly SUPRA critical.
3. Add a little random noise to P_{max}.
4. Perform 15 iterations of gradient descent, where on each iteration
   (a) Set P to the value in-between P_{min} and P_{max}
   (b) Grow the graph at P three times
   (c) If the graph goes SUPRA critical at least 2 out of 3 times, set $P_{max} = P$
   (d) If the graph goes SUB critical at least 2 out of 3 times, set $P_{min} = P$

Notes:

- After 15 iterations, the value of P is found to be sufficiently converged to the critical probability.
- P_{min} and P_{max} are initially set to be fairly close to the critical probability in order that the 15 iterations of gradient descent may be usefully spent.
- The noise initially introduced to P_{max} introduces some variability into the 'halving' of P.

Computational Considerations

For a quality source of random numbers, an implementation of the *Mersenne Twister* random number generator (Matsumoto and Nishimura, 1998) was used. The random number generator was re-seeded on at the beginning of each graph growth run. In this way, no two simulation sessions used the same sequence of random numbers.

3 Results and Discussion

Figure 1 directly compares main results from this study (blue lines) to the main results in the original paper (red lines). Both data sets are in qualitative agreement insofar as the *central idea* goes. The downward trend of lines indicates that increasing the size of the firing disk, or increasing the alphabet size of the polymers lowers the critical probability that a complex autocatalytic set (supra-critical graph) will spontaneously emerge (in a well-stirred environment).

The main point of departure from the original results lies in the actual values of critical probability. Values gained in this study are typically two factors of ten *higher* than those gained in the original experiment. To validate their results, Farmer et al. provide a theoretical estimation of P_{crit} for a chemistry with alphabet size $B = 2$. However, the mathematical derivation is largely unexplained and hard to follow, and thus of limited insightful use.

To support results gained here, it is clear that a critical probability for autocatalysis P_{crit} has to satisfy the greater of the two following conditions:

1. Firstly, as a bare minimum, the value of P_{crit} must be able to create at least one new species outside the firing disk. In a firing disk of s_f species, there exist ${s_f}^2$ viable condensation reactions, and so it follows that $P_{crit} > \frac{1}{{s_f}^2}$ in order to catalyse at least one of these.
2. Furthermore, the value of P_{crit} must be set such that at least one new species *continues* to be created at each iteration. The graph must exhibit continual growth.

At this stage, it must be noted that the values of P_{crit} reported in the original results do not satisfy the first condition. In the simplest case, for example, when the firing disk consists of the two monomers a and b, the critical probability is cited to be less than 10^{-1}. However, with four condensation reactions aa, ab, ba, bb being initially viable in this scenario, this probability would assign around 0.4 reactions on iteration 1, which would be computationally truncated to 0. The graph would stop growing immediately, and would have no chance of being critical.

Additionally, the value of P_{crit} for condition 2 is often higher than that for condition 1. As condition 2 concerns the time behaviour of the graph leading to it's eventual fate, it can be calculated by equating the growth of the graph to the evolution of a discrete dynamical system (and finding the bifurcation point in that system).

The formulation of this dynamical system is possible providing that the following assumptions are made about the graph growth procedure described in Section 2:

1. Cleavage reactions can be ignored. (This assumption is feasible since, at each graph iteration, the number of viable cleavage reactions is very much smaller than the number of viable condensation reactions. Furthermore (through general simulation observations), of the cleavage reactions assigned, even fewer produce new polymer species outside the current system).
2. Every condensation reaction assigned creates a new species previously not in the system.

Assuming all assigned condensation reactions produce new polymer species leads to the "luckiest case" of the reaction graph, which is in fact the situation desired, whereby the graph grows at the *absolute minimum value* of P possible.

The dynamical system describing the graph growth, then, starts at iteration 0 with the total number of species in the firing disk and the total number of condensation reactions these species will have assigned amongst them:

$$n_s = s_f = \sum_{L=1}^{M} B^L$$

$$n_n = trunc\left[P{n_s}^2\right]$$

where $trunc$ denotes a truncation to the nearest integer (polymers only exist as wholes).

At each successive iteration, the total number of polymer species n_s is equal to the total number of species at the beginning of the last iteration, plus the number of new species assigned on the last iteration

$$n_s(t+1) = n_s(t) + n_n(t)$$

and the number of new species n_n is equal to the number of new condensation reactions viable on the last iteration multiplied by P (since we are assuming every condensation reaction assigned creates a new species):

$$n_n(t+1) = trunc\left[P(n_n(t)^2 + 2n_s(t)n_n(t))\right]$$

The dynamical system is parameterised by the alphabet size B, the initial order of the firing disk M and the probability of catalysis, P. Increasing the parameter P over a critical value causes a bifurcation in the dynamics. Below this bifurcation point, the reaction graph goes sub-critical and settles to a fixed point whereby the number of new polymers n_n per iteration become 0, and the total number of species n_s rest at an arbitrary value. Above the bifurcation point, the reaction graph grows exponentially. The bifurcation point thus corresponds to the phase transition in this scenario, and the value of P at which it occurs is the critical probability P_{crit}.

Figure 2 shows the nature of the bifurcation point for the dynamical system described above with firing disk $B = 2$, $M = 6$. More importantly, Figure 3 shows that the theoretical bifurcation points for different alphabet sizes and firing disk orders correspond more or less precisely to the results derived from the simulation in Section 2 (even though the simulation *does* allow cleavage reactions and *does not* require condensation reactions to necessarily produce new species).

Despite the described differences in critical probability estimates to original work, reaction graph growth curves derived from this work (Figures 4-6) tally fairly well with those presented for the original model ((Farmer et al., 1986), p55, Fig 2) both in terms of form and numerical axis values.

Figure 4 shows that a low probability of catalysis leads the reaction graph (which has an initial firing disk of size $L_f = 6$ and alphabet $B = 2$) to decay until there is no further growth. By contrast, increasing the probability of catalysis past the critical threshold leads to supra-critical growth

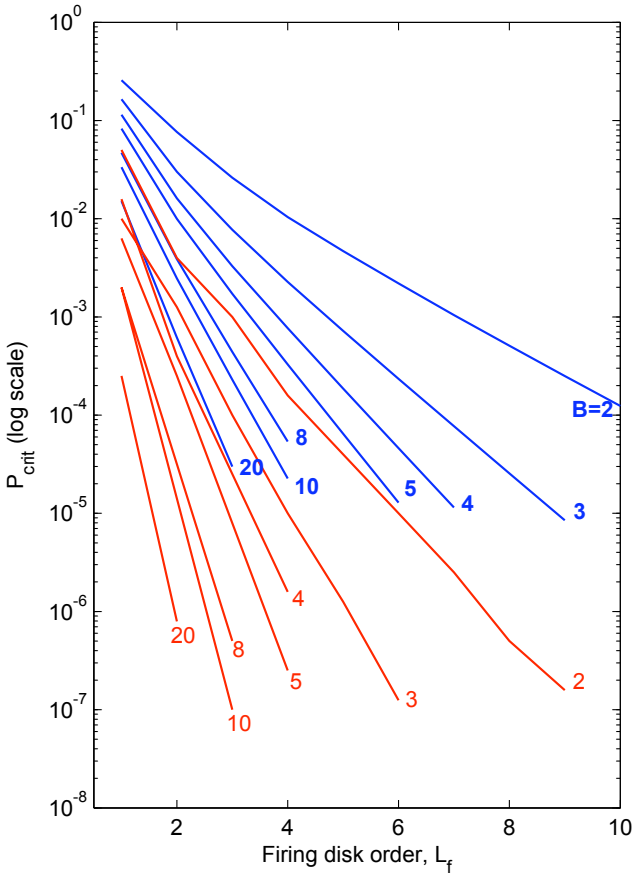

Figure 1: Main comparison of results [blue lines] with those obtained by (Farmer et al., 1986) [red lines]. Graph shows how critical probability of catalysis P_{crit} scales with order of firing disk L_f for different alphabet sizes (labels on lines). Red lines should only be viewed as an estimate of original data.

(Figure 6) where a small initial decay is followed by exponential growth without bound.

Right on the critical threshold, the reaction graph was found to be incredibly fickle, sometimes turning sub-critical, and sometimes supra-critical (Figure 7). Figure 5 was obtained by running the reaction system with firing disk $B = 2$, $L_f = 6$ over many trials at the critical threshold $P_{crit} = 0.002194$ and recording the longest instance of an eventually supra-critical graph. In this study, reaction graphs surviving for any length of time at the critical threshold were delayed supra-critical graphs where fortuitous assignment of condensation reactions meant that only a single new species would be assigned at each iteration (the other condensation reactions being assigned to produce polymers already in the system). This marks another minor departure from the original work where the number of new species per iteration is reported to be erratic at criticality.

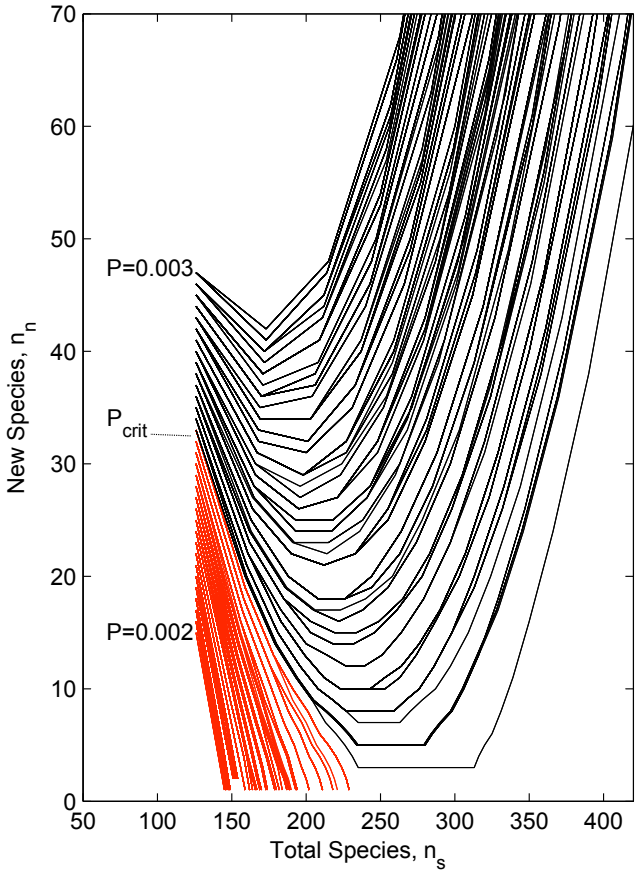

Figure 2: Overlay of phase portraits for the graph growth dynamical system with firing disk $B = 2$, $L_f = 6$. Each trajectory corresponds to the system portrait at a different value of the parameter P. When P approaches P_{crit}, the dynamical system bifurcates and instead of settling to a fixed point (corresponding to a sub-critical graph - red lines), the system spirals to infinity (corresponding to a supra-critical graph - black lines).

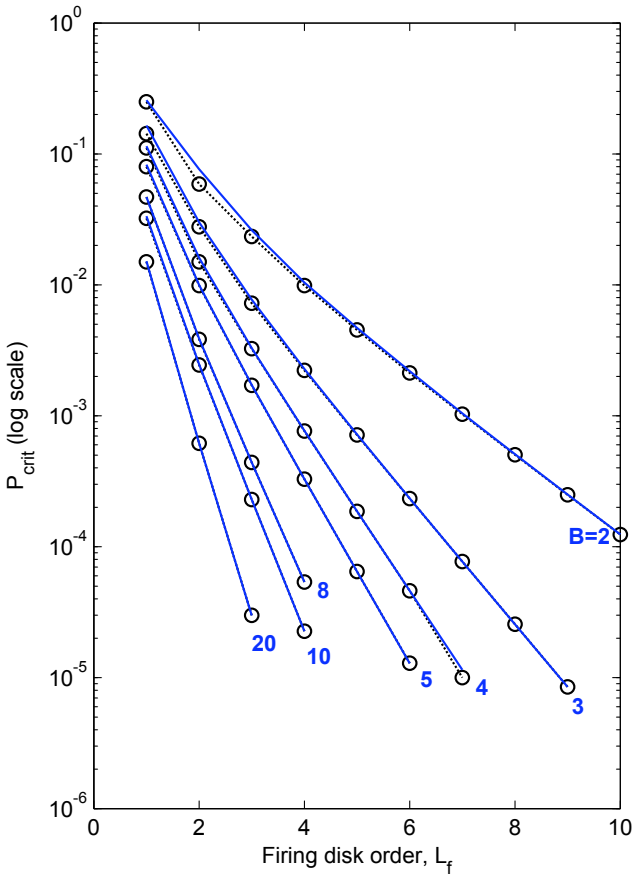

Figure 3: Similarity between theoretical and simulation results. Bifurcation values of P for the graph growth dynamical system (black dotted lines with circle markers) coincide nearly exactly with results obtained in the re-implemented simulation (blue lines) - so much so that the black dotted lines are often obscured on this plot.

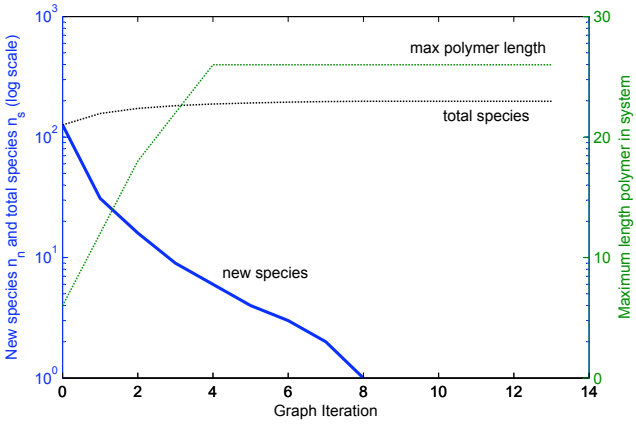

Figure 4: $B = 2$, $L_f = 6$, $P = 0.002000$. Sub-critical. Graph decays. After 8 iterations, graph stops growing.

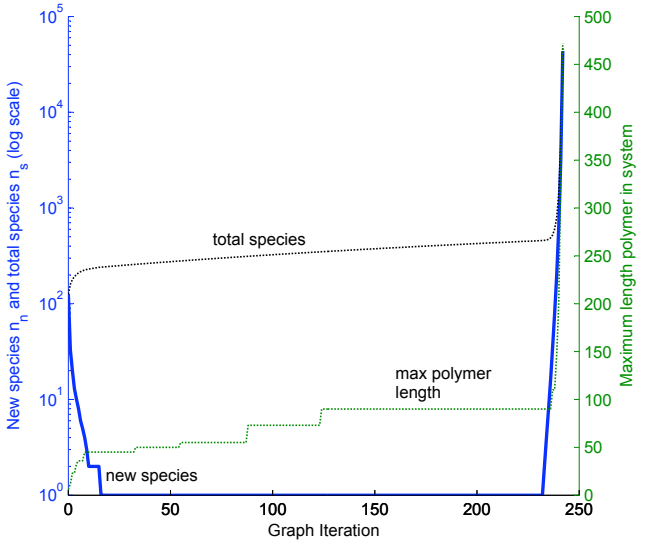

Figure 5: $B = 2$, $L_f = 6$, $P = 0.002194$. "Critical". Graph initially decays to a steady growth rate of 1 polymer per graph iteration (on the line of the x-axis) until an eventual explosion happens around iteration 230.

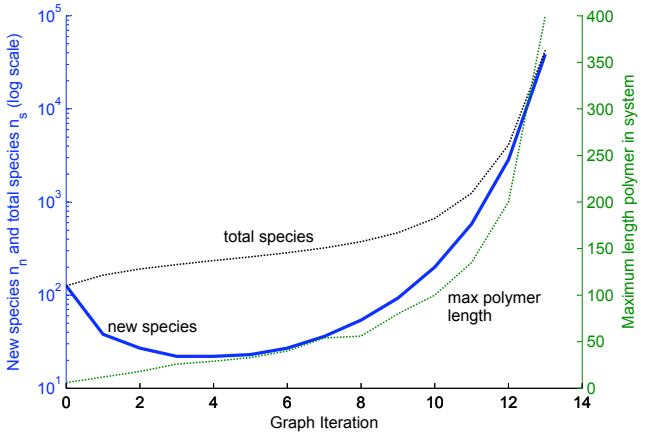

Figure 6: $B = 2, L_f = 6, P = 0.0025000$. Supra-critical. The graph initially decays, but quickly recovers and then snowballs.

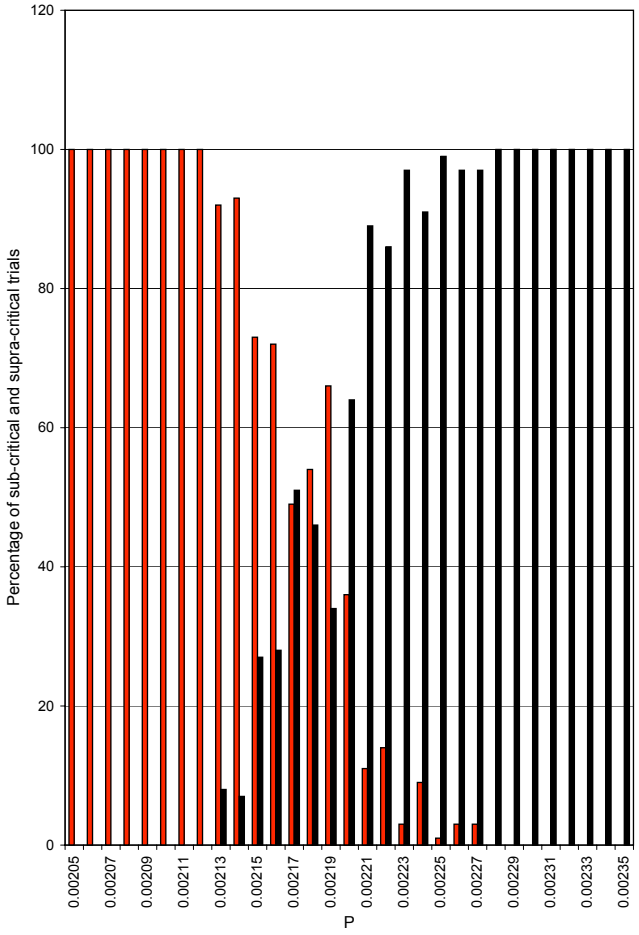

Figure 7: Characterising the phase transition for firing disk $B = 2, L_f = 6$ in terms of trial frequency. The reaction graph was grown 100 times for all values of P between 0.00205 to 0.00235 with interval 0.00001. Red bars represent how many trials went sub-critical, and black bars represent how many trials went supra-critical. The phase transition is clearly visible as a disjoint region separating two regions of stable sub and supra-critical behaviour.

4 Conclusions

This work has sought to conduct a careful re-implementation of one of the first models investigating network autocatalysis (Farmer et al., 1986). Results here follow the same qualitative pattern, and thus do not invalidate the general abstract aims of the original work, but they do present a (fairly major) quantitative discrepancy to the critical catalytic probabilities reported from the original model.

Such discrepancies probably do not hold significant connotations for subsequent work published in the last 20 years, since the spirit of the Farmer et. al. study is one of proving a very general point, but nevertheless they would be nice to resolve. Indeed, the introduction of the original work states that many of the results should be experimentally testable.

In theoretical support of critical catalytic probabilities presented here, a simple discrete dynamical systems model is proposed as an approximation to the more involved reaction graph growth algorithm. The critical catalytic probabilities of the original work can be seen as too low to produce bifurcations in this dynamical model, whereas the bifurcations correspond more or less exactly to the simulation results of this study.

The source of the discrepancy is not ultimately resolved, but it seems that the most outstanding grey area lies with the calculation of the number of new condensation and cleavage reactions at every iteration. However, even with no explicit details mentioned in the original publication, there is little room for manoeuvre, and this study implements a straightforward common-sense interpretation.

Specifics aside, it is worth finally noting that in the last two decades, models relating to the origin of life have gained (considerably) in fidelity from pure abstract autocatalytic notions. Whilst autocatalytic sets remain an important cornerstone, one branch of enquiry for instance ((Mavelli and Ruiz-Mirazo, 2007), (Ruiz-Mirazo and Mavelli, 2008)) focuses on the origins of minimal cells in terms of how active self-producing 'proto-cellular' systems could have started to couple internal chemical reactions to membrane processes. Such efforts are beginning to address the deeper issues raised in Question 3 of the Introduction.

References

Farmer, J., Kauffman, S., and Packard, N. (1986). Autocatalytic replication of polymers. *Physica D*, 22:50–67.

Hordijk, W. and Steel, M. (2004). Detecting autocatalytic, self-sustaining sets in chemical reaction systems. *J. theor. Biol.*, 227:451–461.

Kauffman, S. (1986). Autocatalytic sets of proteins. *J. theor. Biol.*, 119:1–24.

Kauffman, S. (1993). *The Origins of Order: Self-Organization and Selection in Evolution*. Oxford University Press, USA.

Matsumoto, M. and Nishimura, T. (1998). Mersenne twister: A 623-dimensionally equidistributed uniform pseudorandom number generator. *ACM Trans. on Modeling and Computer Simulation*, 8:3–30.

Mavelli, F. and Ruiz-Mirazo, K. (2007). Stochastic simulations of minimal self-reproducing cellular systems. *Phil. Trans. R. Soc. B.*, 362:1789–1802.

McMullin, B. (1999). Some remarks on autocatalysis and autopoiesis. Presented at the workshop: Closure: Emergent Organizations and their Dynamics, May 3-5, 1999, University of Ghent, Belgium.

Ono, N. and Ikegami, T. (2000). Self-maintenance and self-reproduction in an abstract cell model. *J. theor. Biol.*, 206:243–253.

Ruiz-Mirazo, K. and Mavelli, F. (2008). On the way towards 'basic autonomous agents': Stochastic simulations of minimal lipid-peptide cells. *BioSystems*, 91(2):374–387.