Emergence of Autocatalysis in Prebiotic Reaction Networks

Varun Varanasi*
Department of Physics, Yale University

Jun Korenaga
Department of Earth and Planetary Sciences, Yale University
(Dated: October 29, 2024)

This paper presents a novel analytical model to explain the emergence of reflexively autocatalytic and food-generated (RAF) sets in Kauffman networks. The sharp transition in the probability of observing RAF sets as a function of catalytic probability is a well-documented, but unexplained phenomenon. Our model reveals that this behavior arises from the disparity between the number of ways to select k unique molecule-reaction pairs and the number of ways to allocate minimally RAF sets among them. This perspective on RAF theory offers new insights into the underlying structure of Kauffman networks, particularly emphasizing the importance of catalytic cores. This work further predicts the number of catalytic cores proportional to $n2^{n+1}$ in realistic Kauffman networks. While this model advances the understanding of RAF set formation, it relies on the simplifying assumption of non-overlapping catalytic cores. Future work should incorporate overlap probabilities and refine the restricted partition function to better capture the behavior of RAF emergence.

I. INTRODUCTION

The first major hypotheses regarding the emergence of life from evolving chemical systems were proposed nearly a century ago [1, 2]. Since then scientific theory has developed into two major schools of thought: the RNA-first theory and the protein-first theory. Whereas both hypotheses rely on self-sustaining chemical reactions as a prerequisite, they differ on the assumed chemical properties of their constituents. The RNA-first theory suggests that life began with self-reproducing RNA molecules. Over time, these molecules evolved to encode genetic information as we understand it today. While the RNA-first theory benefits from experimental evidence of RNA's catalytic abilities [3], it lacks the necessary self-replication mechanism. In the protein-first theory, the first biological molecules were proteins and life emerged from a set of organized and self-sustaining chemical reactions. Since proteins are natural catalysts, this model is a more chemically feasible alternative than the RNA-first theory. However, it still struggles to answer how proteins can replicate without an initial template structure or genetic code [4].

Since the 1970's scientists have attempted to reconcile the two frameworks via a handful of theoretical models including hypercycles, chemeotons, autopoietic systems, (M,R) systems, and collectively autocatalytic sets [5]. Despite progress, each theory encounters issues in explaining self-replication without external templates. The autocatalytic set theory offers a promising alternative due to its basis in self-sustaining reaction networks, supported by experimental observations of early biological systems [6, 7]. The autocatalytic set model argues that given a sufficiently large set of molecules and a library of catalyzed reactions, the molecules will react within their environment and spontaneously produce a self-sustaining set. Instead of considering a particular chemical species, the autocatalytic reaction network framework considers sets of molecules that can interact to create the desired properties of self-sustenance and self-replication. These sets, referred to as autocatalytic sets, are a subset of the total set of molecules and reactions such that each molecule in the set is produced by a reaction also in the set. Furthermore, to be included in the set, each reaction must be catalyzed by a molecule in the set. These conditions, also known as catalytic closure, allow us to formalize the idea of "life-like" chemical systems [8]. The remainder of this paper will focus on this formulation of autocatalytic sets.

Research in this field has been primarily focused on the probability of observing a self-sustaining set in simple prebiotic models. Previous theoretical efforts have concentrated on bounding the required rate of catalyzation to probabilistically ensure this phenomenon. Computational work, on the other hand, has centered on the behavior of autocatalytic sets across varying chemical models and probabilistic regimes. To our knowledge, there is no analytical explanation for the observed emergent phenomena. The main motivation behind this work is to provide such an explanation and bridge the theoretical and computational gap. In particular, in this paper, we propose an analytical model to elucidate the mechanics of the emergence of self-sustaining sets as a function of catalytic probability in Kauffman networks [9].

In this work, we introduce a closed-form combinatoric model to explain the emergence of RAF sets in Kauffman networks. In Section 2, we cover the background formalization of Kauffman networks, autocatalytic sets, and their behavior as a function of catalytic probability. Section 3 outlines the proposed combinatoric model, and Section 4 describes the outcomes of the model and discusses the origins of the probability transition. Section 4

---

[*] Also at IMC Trading

also discusses the behavior of the core set and its importance in determining the RAF behavior of the system.

II. REACTION NETWORKS AND THE EMERGENCE OF AUTOCATALYTIC SETS

A. Reaction Networks

The chemical system we consider in this work is a generalization of the Kauffman model also known as the binary polymer model [4]. As outlined in [9], this model abstracts prebiotic chemical reactions to a series of addition and breaking relations between strings. Each molecule is labeled by a string defined over an alphabet of ℓ numbers. The set of reactions are all possible concatenation of pairs of strings such that the combined string is still less than a length n. The set of reactions also includes the inverse relation corresponding to a string breaking into substrings. This system is specified by 4 sets defined by 4 hyperparameters.

- **X (Molecule set)**: The set of all possible molecules is determined by the parameters n and ℓ. This set is defined as $X = \{x \in \{0, 1, ...\ell - 1\}^{\leq n}\}$ where $S^{\leq n} = \bigcup_{i=1}^{n} S^i$. The size of the alphabet is represented by ℓ while n represents the maximum size of a molecule in our set. Therefore, X is the set of all possible configurations up to length n where each index takes ℓ possible values. Typically $\ell = 2$ and we consider our molecule set as a set of binary strings up to length n for a total size $|X| = 2^{n+1}-2$. In a simple example defined by $n = 2$ and $\ell = 2$, set X is $\{0, 1, 00, 01, 10, 11\}$ with a total size of 6.

- **F (Food set)**: The set of food molecules is a subset of X determined by the parameter t. $F = \{f \in \{0, 1, ...\ell - 1\}^{\leq t}\}$ refers to the set of molecules with lengths less than or equal to t. The food set can be understood as a theoretical molecular reservoir. In the case where $t = 1$ and $\ell = 2$, F is simply $\{0, 1\}$

- **R (Reaction set)**: The reaction set is the set of all allowed reactions determined by the molecule set X. The reaction set consists of tuples of $r = (\{a, b\}, \{c\})$ or $r = (\{c\}, \{a, b\})$. The first term in the tuple refers to the reactants and the second the products. Each reaction in the reaction set corresponds to the concatenation of two molecules in X or the breaking of one molecule in X into two submolecules in X. Formally, we define the reaction set $R = \{(\{a, b\}, \{c\})$ and $(\{c\}, \{a, b\}) : a, b, c \in X, c = ab$ or $c = ba\}$. Note that if $a = b$, r corresponds to the concatenation of a with itself or the lysis of c into two a molecules. Following these rules, if we consider the model defined by $n = 2$ and $\ell = 2$, the reaction set is

$$\begin{aligned} R = \Big\{ & r_1 = (\{0,0\},\{00\}), r_2 = (\{00\},\{0,0\}), \\ & r_3 = (\{0,1\},\{01\}), r_4 = (\{01\},\{0,1\}), \\ & r_5 = (\{1,0\},\{10\}), r_6 = (\{10\},\{1,0\}), \\ & r_7 = (\{1,1\},\{11\}), r_8 = (\{11\},\{1,1\}) \Big\} \end{aligned}$$

The size of the set of reactions $|R|$ grows at the rate $n2^{n+1}$ [10].

- **C (Catalyst set)**: The set of catalyzing reactions C records which molecules can catalyze which reactions in the system. Each relation is denoted by a tuple $c = (x, r) : x \in X, r \in R$. Therefore, $C \subseteq X \times R$. A given molecule-reaction pair c is included in C with probability p. Thus, for a given system, $|C| \sim Binom(|X| \cdot |R|, p)$. Following the earlier examples defined by $n = 2$ and $\ell = 2$, for a non-zero or non-one probability parameter p, one possible realization of C is

$$\begin{aligned} C = \Big\{ & (00, r_3), (00, r_4), (00, r_7), (00, r_8), \\ & (01, r_1), (01, r_2), (11, r_1), (11, r_2) \Big\} \end{aligned}$$

For the remainder of this paper, we set $\ell = 2$ and $t = 2$ such that we only have two remaining hyperparameters: n and p. The parameters n and p correspond to, respectively, the size of the network and the probability of including each molecule-reaction catalyzation in the set of C. We also assume that if a molecule catalyzes a given reaction, it will also catalyze the corresponding reverse reaction. If $r = (\{a, b\}, \{c\})$ and $r' = (\{c\}, \{a, b\}$, then, if $(x, r) \in C \Rightarrow (x, r') \in C$. For more details see [9].

B. Autocatalytic Sets

The most common and applicable formalization of an autocatalytic set in a Kauffman network is called an RAF (Reflexivley Autocatalytic and Food-generated) set. A reaction network defined by $R' \subseteq R$ and $X' \subseteq X$ is said to be RAF if it satisfies the following conditions:

- **Reflexively Autocatalytic**: $\forall r \in R', \exists x \in supp(R') : (x, r) \in C$ where $supp(R') = \bigcup_{r \in R'} \{r_a, r_b, r_c\}$. Intuitively, this refers to the set of reactions where each reaction r is catalyzed by a molecule involved in another reaction in R' or the food set F.

- **Food-Generated**: A set is food generated if every reactant in R' can be created by repeated applications of reactions in R' and molecules in F.

Essentially, an RAF set is a set of molecules and reactions that can generate every molecule in the set using reactions that are catalyzed by other members of the set or in the food set. [9] outlines a more rigorous formalization of RAF sets and some of their key properties.

C. Emergence of Autocatalytic Sets

The earliest hypotheses on the emergence of autocatalytic sets in Kauffman networks were made by Stuart Kauffman himself in 1986 [4]. In this paper, he argued that observing an RAF set in a binary polymer model should be highly probable since the process is akin to the emergence of a giant connected component in an Erdos-Reyni graph. Later, Lifson argued against the assumptions in Kauffman's original work stating that rather than keeping the probability of catalyzation fixed across various network sizes, a more reasonable expectation would be to fix the average number of catalyzations per molecule, $f = p \cdot |R|$. Under this more realistic system, it was unclear whether an RAF set should emerge. This remained an open question until 2004 when it was computationally shown that only a linear growth rate in the expected number of catalyzations was required to spontaneously generate RAF sets [9]. This result was then theoretically proven by Steel and Mossel [11]. In fact, they proved this under stronger assumptions than required for the binary polymer model showing that a linear growth rate is more than sufficient to ensure the spontaneous emergence of RAF sets. Computational results have also found that the level of catalysis only needs to be between $1 < f < 2$ for network sizes up to 20. In fact, computationally, f grows linearly with n according to $f = 1.0970 + 0.0189n$ [12]. Recent work has moved away from studying the requirements of spontaneous autocatalytic sets into studying the structure of the RAF sets themselves. One of the concepts under investigation is the notion of irrRAFs, minimally RAF subsets of RAF networks. Hordjik et al. found that within a single RAF set, there can be exponentially many irrRAFs. They also found that these irrRAF overlap, on average, by about 50% but can range anywhere from 25% to 80% [13]. irrRAFs provide a fresh perspective on the evolvability and emergence of RAF sets. In this work, we build on the work of Hordjik et al. to propose an analytical model for the emergence of maximal RAFs as a function of the number of irrRAFs in the binary polymer model. To the best of our knowledge, this is the first proposed explanation for the emergence of RAF sets.

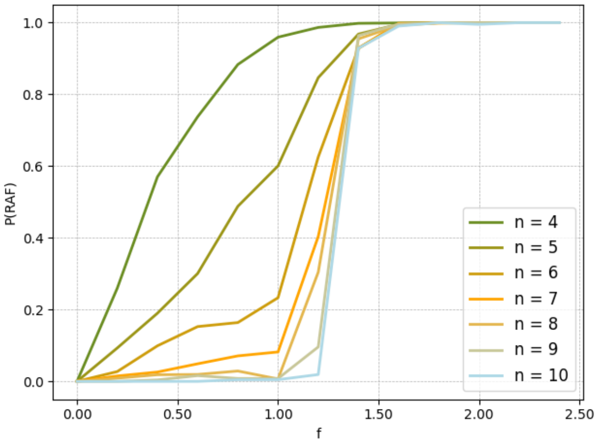

FIG. 1. Simulations of the probability of finding an RAF set across various average catalysis parameters by network size. Note the transition probability between $1 < f < 2$.

III. MODEL

For a given system with network size n and probability parameter p, we can sample various C to estimate $P(RAF)$, the probability of observing an RAF set [9]. The goal of this paper is to develop a closed-form non-asymptotic estimate for this quantity. First, we begin by framing the probability of observing an RAF set in the language of boolean analysis. For example, in a given network size n and probability parameter p, we can define C as an $|X| \cdot |R|/2$ dimensional boolean vector where each 1 indicates that the corresponding molecule-reaction catalyzation is included in our set C. For the remainder of this paper, we refer to the size of sample space as $S = |X| \cdot |R|/2$ as determined by network size n. Note that we divide the set of available vectors by 2 to account for the assumed reflexive property of catalysis. Then, we define $RAF(C)$ as an indicator function if the network of size n and catalyst set C contains an RAF set. For a given p and corresponding catalyst set C_p, we define $P(RAF)$ as

$$P(RAF) = \mathbb{E}[RAF(C_p)] \tag{1}$$

$P(RAF)$ for a given p is thus the probability of observing an RAF set with the system parameters n and p. We can characterize $P(RAF)$ by decomposing it into a summation of probabilities conditional on the number of catalysts in C.

$$P(RAF) = \sum_{i=1}^{S} P(k = i) P(RAF|k) \tag{2}$$

In the above expression, $k = \sum x_i$ is the number of bits turned on. The term $P(k = i)$ represents the probability of turning a total of i bits. Since each bit is turned on with an independent probability p, this term follows the binomial distribution.

$$\begin{aligned} P(k = i) &= \binom{S}{i} p^i (1-p)^{mR-i} \\ &\sim Binom(S, p) \end{aligned} \tag{3}$$

The second conditional probability term, $P(RAF|k)$, is where the bulk of the complexity resides.

A. Intuition

The intuition behind our method involves the notion of "catalytic cores" which can be thought of as minimally RAF catalytic sets. In other words, if a set is

RAF, the core is the set of catalyzations that contribute to the RAF set. They can be defined as the set of catalysts such that removing any catalyst will convert the set from RAF to non-RAF. Note that the notion of catalytic cores is identical to the minimal CRA introduced by [11] or irreducible RAF used in [13]. If we know the cores of the system, we can exactly write $P(RAF|k)$ as a simple combinatoric equation. However, given the difficulty of enumerating the cores of a system, we consider the simplified case where we only know the total number of cores. The new challenge in this method arises from differentiating sets of solutions that share common bits. In fig. 2, we have 2 cores. The core of length 3 is colored green while the core of length 2 is in orange. In b) we see that they share a catalyst denoted in striped orange and green. For a given k catalyzations, the probability of hitting at least one of the cores is greater in the case that the cores share a catalyst. In the Kauffman network, a priori, we do not know which solutions overlap with others. Instead, we can assume no overlap between our solutions to place a theoretical lower bound on the probability of observing an RAF set.

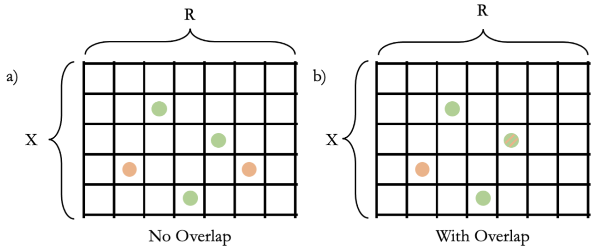

FIG. 2. Toy example with $|X| = 5$ and $|R| = 7$. The grid visualizes which molecules in X catalyze corresponding reactions in R. The system is defined by two RAF cores in green and orange. a) depicts the case with two non-overlapping cores, while b) is an example of a system with two cores sharing an overlapping catalyst (striped green and orange).

Armed with a distribution of catalytic cores, we begin with conditioning $P(RAF)$ by the number of bits turned on (see eq. 2). Within the k catalysts in our set C, we want to consider the cases where we occupy $1 \leq m \leq k$ bits with catalytic cores. We require at least one core so that the set is RAF. The number of ways we can allocate cores across m catalysts is approximated by the restricted integer partition function over the finite set of cores. The remaining $k - m$ bits must allocated to non-cores. We estimate this value by the number of ways we can select $k - m$ catalysts from the remaining set. This value is further corrected by the ways we may overcount cores in the combinatoric calculation. The remainder of Section 3 details the mechanics of this process. Note that our no overlap assumption becomes increasingly inaccurate as we increase m; however, as discussed later, we find that the relevant values of m are sufficiently small that the assumption remains useful for our purpose of understanding the phase transition.

B. Definitions

For a given network size n and probability parameter p, we sample $C \in \{0,1\}^S$ where each $x_i = 1$ with probability p. This defines a discrete hypercube in S dimensions with probability measure

$$\mathbb{P}_p(x_1, ...x_S) = p^k(1-p)^{S-k} \tag{4}$$

where $k = \sum_i x_i$. We can define $RAF(x)$ as a function of these boolean variables that returns 1 if the set of catalyzations yields an RAF set or 0 otherwise. We then define the set of solutions or RAF catalyzations as

$$\Omega = \{C \in \{0,1\}^S : RAF(C) = 1\} \tag{5}$$

Within Ω we define A, the set of cores, as the set of vectors such that flipping any of the inputs from True to False will remove the vector from Ω. Intuitively, the set of cores is the minimal set of catalyzations that remains RAF. In explicit notation, we first define the set Θ_x for a given boolean vector x as the set of boolean vectors with no additional True indices:

$$\Theta_x = \{y \in \{0,1\}^S | \forall i, y_i = 0 \text{ whenever } x_i = 0\} \tag{6}$$

Based on this definition, we define our set of cores as the set of RAF solutions, such that none of the associated Θ_x (save the x itself) is a solution.

$$A = \{x \in \Omega : \forall y \in \{\Theta_x \backslash x\}, RAF(y) = 0\} \tag{7}$$

The set A therefore defines the set of RAF solutions such that removing any of the catalyzations (flipping a True bit) destroys the solution. We refer to the size of our cores as $L = |A|$. Our final set definition, A_m, is the subset of vectors in A with m True bits

$$A_m = \{x \in A : \sum x_i = m\} \tag{8}$$

A_m is thus the set of core solutions with length m.

C. Model

Our general method for estimating $P(RAF|k)$ is to enumerate the possible vectors x with length k such that $RAF(x) = 1$.

$$\begin{gathered} P(RAF|k) = \sum_{x \in B_k} P(x) \\ B_k = \{x \in \Omega : \sum x_i = k\} \end{gathered} \tag{9}$$

Since each $x \in B_k$ has the same distribution of 0 and 1 bits, they each occur with the same probability. In other words, $P(x)$ is uniformly distributed over the $\binom{S}{k}$ possible vectors with k bits turned on.

$$P(RAF|k) = \frac{|B_k|}{\binom{S}{k}} \tag{10}$$

Approximating B_k

The problem reduces to estimating the size of set B_k. For a given boolean vector in $|B_k|$, m, the size of all of our catalytic cores can occupy values $1 \leq m \leq k$. For each given m, we consider the number of ways that we can allocate our set of catalysts A into m bits. Since we assume no overlaps between cores, we can calculate this value as a restricted integer partition over a finite set, $p_A(m)$. With the remaining $k-m$ bits we want to calculate how we can allocate them across non-solution catalysts. There are a total of $S-m$ remaining unselected catalysts. To avoid selecting catalyst cores, we want to remove all cores $|A_1|$ from the available set. Finally, to avoid selecting catalysts that constitute a core, we correct our estimate by the number of ways that we can include cores from lengths $2 \leq j \leq k-m$ via the restricted partition function. Altogether, our estimate for B_k takes the form:

$$|B_k| \approx \sum_{m=1}^{k} p_A(m) \cdot \max\left(\binom{S-|A_1|-m}{k-m} - \sum_{j=2}^{k-m} p_A(j), 0\right) \tag{11}$$

The max function ensures that when the correction term surpasses the number of ways to place the remaining $k-m$ bits, the contribution to $|B_k|$ is 0.

Restricted Partition Function

In number theory, partition functions return the number of distinct ways of representing n as a sum of positive integers. In our problem, we are interested in a variation of the partition function known as the restricted integer partition over a finite set. Unlike the traditional partition function, the restricted partition function, $p_A(n)$, provides the number of distinct ways to sum values in A to equal n.

$$p_A(n) = \left|\{a \subseteq A : \sum_{x \in a} x = n\}\right| \tag{12}$$

There are no known closed-form representations of the restricted partition function or general partition function, but there exist asymptotic expansions for variants of the problem. For a finite set A with L relatively prime elements, the asymptotic expansion of the restricted partition function over this set is proportional to n^{L-1}. The full asymptotic expansion is [14]

$$p_A(n) = \left(\frac{1}{\Pi_{a \in A} a}\right) \frac{n^{L-1}}{(L-1)!} + O(n^{L-2}) \tag{13}$$

However, the regime we are interested in is the small n regime for which the asymptotic correction factor is not appropriate. For that reason, we estimate the restricted partition function via the proportionality relation

$$p_A(n) \propto n^{L-1} \tag{14}$$

With these estimates in place, we can fully predict and model the behavior of the $P(RAF)$ function for various network sizes and catalyzation probabilities as a function of the number of cores. In the proceeding sections, we demonstrate the validity of the proposed model, explain the apparent phase transition, and comment on further model predictions and limitations.

Approximations

The heart of the proposed model is a counting argument involving combinations and the restricted partition function. These terms grow at rates that are not computationally tractable to solve exactly. Instead, when implementing this model we rely on a handful of approximations. First, for the partition function, we use the proportionality relation given in eq. 14 with a scaling constant D. The value of D is selected to produce a probability transition in the range near $f = 1.25$. We also estimate the combination functions via the upper-bound approximation

$$\binom{n}{k} \approx \left(\frac{n}{k}\right)^k \tag{15}$$

Finally, to limit the computational strain of summing over all possible k, the count of catalysts, we only consider the high probability terms within 0.25 standard deviations of the expected number of catalysts. These regions are highlighted in fig. 5.

IV. RESULTS AND DISCUSSION

Origin of Phase Transition

Using the model developed in Section 3, we first demonstrate its ability to replicate the behavior exhibited in fig. 1. Consider a Kauffman network defined by $n = 6$. The behavior shown in fig. 3 can be understood in three parts. As discussed in Section 3, the probability of observing an RAF set can be broken down into a conditional probability statement. Each $P(RAF|k)$ term can then be estimated by eq. 10. If we plot the behavior of each term in the expression, we can see where the sharp jump in probability arises. As shown in fig. 4, $|B_k|$ grows at a polynomial rate while the combinatorial term in the denominator grows exponentially in k. Intuitively, we can think of the conditional probability, $P(RAF|k)$, as proportional to the difference between the two lines. We see that at $k < 160$ the numerator is substantially smaller than the number of ways we can select k catalysts. However, for $k > 160$, we see that the numerator dominates the denominator. This implies that at low k, $P(RAF) \rightarrow 0$ and at high k $P(RAF) \rightarrow 1$. The sharp jump in probability occurs in the region surrounding $k = 160$ with the sharpness exaggerated by the

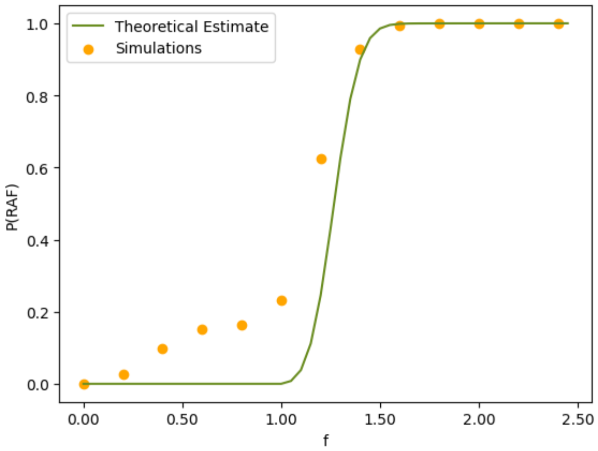

FIG. 3. Probability of observing an RAF set in a $n = 6$ Kauffman network plotted against the expected number of catalyzations per molecule. Simulation data (in orange) is plotted next to the theoretical model (in green) from Section 3.

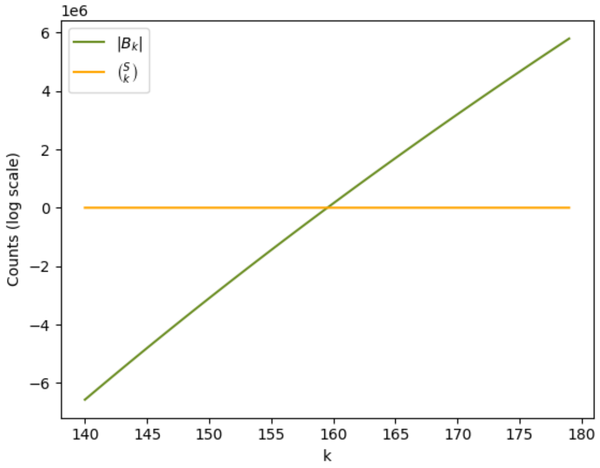

FIG. 4. Behavior of the terms in the eq. 10 in log space as a function of k catalyzations in the $n = 6$ Kauffman network with core distribution shown in fig. 3. The estimate $|B_k|$ surpasses the denominator term at approximately $k = 136$.

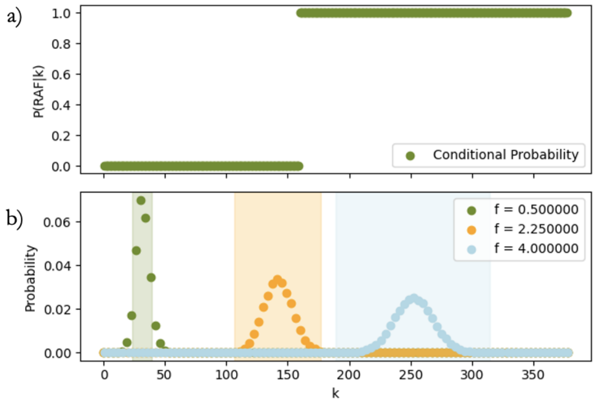

FIG. 5. a) The conditional probability of observing an RAF set in a $n = 6$ Kauffman network with the core distribution given by fig. 3. b) Distributions of k as a function of the expected number of catalysts per molecule

exponential difference in the terms. Also, note that the limitations of this model do not appropriately restrict the conditional probability terms to be ≤ 1; however, for our purposes, when the numerator exceeds the denominator, we consider the conditional probability estimate to be 1. The corresponding conditional probability behaves as shown in fig. 5. The final $P(RAF)$ estimate is calculated as the product of the conditional probability with the distribution of k. Therefore, the product almost entirely vanishes for small $f < 2$. Similarly, for large $f > 3$, each conditional probability term is effectively 1. The values of f that center near $k = 160$, or the conditional probability jump, are where we observe the phase transition behavior. Thus the phase transition occurs at the catalyzation probability where the binomial coefficient in the denominator of eq. 10 equals the estimate for $|B_k|$. The main driving force behind this action is the conditional probability step function created by the polynomial growth of $|B_k|$. Remember that this behavior is a consequence of the restricted partition function. Furthermore, since the restricted partition is a function of the core distribution, we can vary L, the size of the cores, and D, the scaling factor, to recover any desired phase transition point.

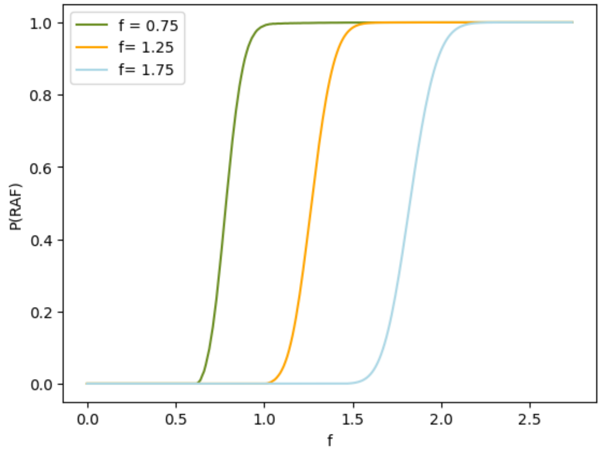

FIG. 6. Plots of the probability of observing an RAF set in an $n = 6$ Kauffman network for various core distributions. Note that the core distributions in these particular plots were constructed using various values of L, the number of catalytic cores, and an appropriate scaling factor D. (see the section "Approximations")

Behavior of Core Set

The catalytic cores clearly define the behavior of the system. In this model, we are primarily concerned with two parameters: L, the size of the core set, and M, the product of the sizes of the core set. As discussed in Section 3, the partition function is polynomial of order $L-1$ with asymptotic corrections proportional to $(L-1)!$ and M. Simulations of small Kauffman networks show that the core set tends to be negatively skewed by core size and contains at least 1 core for every length between 1 and the size of the largest core. They also confirm that $|A_1| = 96$ for networks with sizes between $n = 3$ and $n = 8$. Furthermore, if A_n represents the set of cores for a Kauffman network of size n, then $A_n \subseteq A_{n+1}$. This is a direct consequence of the definitions of Kauffman networks since every network is a subset of a larger n network. It follows that any RAF set in a smaller network will maintain this property in a larger network.

As demonstrated in fig. 6, the core set controls the phase transition point. Furthermore, as shown in fig. 1, for Kauffman networks $n \geq 6$, the probability transition roughly occurs $f = 1.25$. Previous work has found that f grows linearly with n according to $f = 1.0970 + 0.0189n$ [12]. This implies that we can understand the growth rate of the core set by calculating the required size for a phase transition at $f = 1.0970 + 0.0189n$. Consider a general network defined by size n and core distribution A. As discussed in Section 3, A consists of L relatively prime integers corresponding to core lengths and a product of core lengths, M. In the framework presented by our model, the phase transition occurs when

$$\binom{S}{k} = p_A(k) \tag{16}$$

In a realistic chemical model, we can consider networks with $20 \leq n \leq 30$. This implies an average catalyzation value of $f \approx 1.5$ and implied catalytic probability $p = 1.5/|R|$. Furthermore, this gives us a transition point defined by $k = S \cdot p$. Recall that $|X| \sim 2^{n+1}$ and $|R| \sim n2^{n+1}$. Using the asymptotic approximation for the binomial coefficient,

$$\binom{S}{k} \sim \left(\frac{Se}{k}\right)^k \cdot (2\pi k)^{-1/2} \cdot \exp\left(-\frac{k^2}{2S}\right) \tag{17}$$

Plugging in asymptotic approximations and taking the binary logarithm of both sides, we find the LHS of eq. 16 equal to

$$2^{n+1}\left(\log(n) + n + \log(e)\right) - \frac{1}{2}\log(\pi) - \frac{n}{2} - 1 - \frac{1}{2n} \tag{18}$$

where n dominates the parenthetical expression, simplifying the equation to $n2^{n+1}$. Similarly, on the RHS of eq. 16, we can approximate the asymptotic expression using eq. 13.

$$(L-1)\left(n + 1 - \log(L-1) + (L-1)\right) - \log(M) \tag{19}$$

We can simplify this expression by noticing that $L \gg n$ so that the $L-1$ term dominates. Putting it all together, we find that

$$n2^{n+1} = (L-1)^2 - \log(M) \tag{20}$$

We can further analyze M by considering the constraints of the core distribution. We can frame M as an optimization problem of allocation variables $x_1, ...x_m$ where each x_i refers to the number of cores of size i.

$$\textit{obj.} \sum_{i=1}^{m} x_i \log(i) \tag{21}$$

$$\text{s.t. } x_i \geq 1 \in \mathbb{Z} \tag{22}$$

$$\sum_{i=1}^{m} x_i = L \tag{23}$$

$$1 \leq m \leq L \tag{24}$$

We are interested in maximizing and minimizing the objective function subject to the constraints that we have at least 1 core from size 1 to m such that the total number of cores is L. Note that the constraints of this setup are looser than the actual distribution of cores, but still provide useful bounds. A simple analysis of this problem shows that the minimum allocation is to assume $x_1 = L$. This refers to the case where every core has a length of 1. In this case, $M = 1$. The maximal allocation can be found by conditioning on the number of decision variables m. For a given m core sizes, it is clear that the optimal allocation is 1 core for each size up to $m-1$ with the remaining cores allocated to x_m. The optimal M for a given m is thus

$$\sum_{i=1}^{m-1} \log(i) + (L - m + 1)\log(m) \tag{25}$$

This is a monotonic function of m, which implies that the optimal allocation is $x_i = 1$ for $1 \leq i \leq L$. It follows that $M = L!$. Therefore,

$$1 \leq M \leq L! \tag{26}$$

Plugging this bound into the eq˙ 19, we see that

$$L^2 \sim n2^{n+1} \tag{27}$$

This implies that the size of cores grows at a rate of roughly $\sqrt{|R|}$. Thus we provide an estimate of the number of irrRAFs as a function of n. We corroborate the exponential growth of irrRAFs as a function of n and provide a stronger relation than presented in [15].

Limitations

The proposed model has two primary limitations. First, the model itself relies on a lower-bound approximation based on the assumption that cores do not overlap.

This assumption is false, but we argue that the relevant k near the transition point is sufficiently small enough that the no overlap assumption does not lead to grossly inaccurate estimates. The error associated with this assumption is also likely canceled partly by the approximations of the partition function itself. Another limitation of the model is the inputted core size and associated scaling factor. This value drives the transition point of the model and can be selected to produce almost any result desired. The values we used in this paper to produce estimates of the system were selected to resemble simulation data. These values can be verified experimentally, but require considerable computational power. At this point, such verification is left as a future task. Another remedy for this limitation would be to use a more accurate local partition function. In doing so, we could remove the need for the scaling factor D and define the model by a singular parameter.

V. CONCLUSION

In this paper, we introduce a novel analytical model for understanding the emergence of RAF sets in a simple binary polymer model. Our findings demonstrate that the observed probability transition stems from the disparity between the exponential growth of possible catalytic subsets and the available ways to allocate minimally RAF subsets within those choices. This results in a sharp probability jump as the polynomial growth of the allocation quickly dominates the combinatorial denominator. Understanding RAF sets from this perspective allows us to infer the underlying structure of Kauffman networks. In particular, we see that RAF sets are built from catalytic cores whose number grows at a rate $n2^{n+1}$. This model provides a strong step forward in understanding the nature of RAF theory but relies on the assumption that none of the catalytic cores overlap. Future iterations of this model should correct this assumption to be in line with experimental data. We also recommend further work in estimating the local restricted partition function over the set of catalytic cores. A stronger approximation of this quantity would allow us to relate the necessary count of catalytic cores to the catalytic probability of observing an RAF set. Consequently, we can make meaningful claims on the chemical realism of the Kauffman model and its variants. Additionally, this framework offers potential applications beyond RAF theory, particularly in estimating probabilities for monotonic boolean functions with non-overlapping minimal solutions.

---

[1] A. I. Oparin, Russian*Proiskhozhdenie Zhizni* (Moskovskii Rabochii, Moscow, 1924) title translates to "The Origin of Life".

[2] J. B. S. Haldane, The origin of life, Rationalist Annual , 3 (1929).

[3] W. Gilbert, Origin of life: The RNA world, Nature **319**, 618 (1986).

[4] S. A. Kauffman, Autocatalytic sets of proteins, Journal of Theoretical Biology **119**, 1 (1986).

[5] W. Hordijk and M. Steel, Autocatalytic networks at the basis of life's origin and organization, Life **8**, 62 (2018).

[6] D. Sievers and G. von Kiedrowski, Self-replication of complementary nucleotide-based oligomers, Nature **369**, 221 (1994).

[7] G. F. Joyce, Cross-catalytic replication of an RNA ligase ribozyme, Chemistry & Biology **11**, 1505 (2004).

[8] F. J. Dyson, *Origins of Life*, 2nd ed. (Cambridge University Press, 1999) originally published in 1985.

[9] W. Hordijk and M. Steel, Detecting autocatalytic, self-sustaining sets in chemical reaction systems, Journal of Theoretical Biology **227**, 451 (2004).

[10] M. Steel, The emergence of a self-catalysing structure in abstract origin-of-life models, Applied Mathematics Letters **3**, 91 (2000).

[11] E. Mossel and M. Steel, Random biochemical networks: The probability of self-sustaining autocatalysis, Journal of Theoretical Biology **233**, 327 (2005).

[12] W. Hordijk, S. A. Kauffman, and M. Steel, Required levels of catalysis for emergence of autocatalytic sets in models of chemical reaction systems, International Journal of Molecular Sciences **12**, 3085 (2011).

[13] W. Hordijk, J. I. Smith, and M. A. Steel, Algorithms for detecting and analysing autocatalytic sets, Algorithms for Molecular Biology **10**, 10.1186/s13015-015-0050-6 (2015).

[14] M. B. Nathanson, Partitions with parts in a finite set, Proceedings of the American Mathematical Society **128**, 1485 (2000).

[15] W. Hordijk, S. A. Kauffman, and M. Steel, The structure of autocatalytic sets: Evolvability, enablement, and emergence, Acta Biotheoretica **60**, 379 (2012).