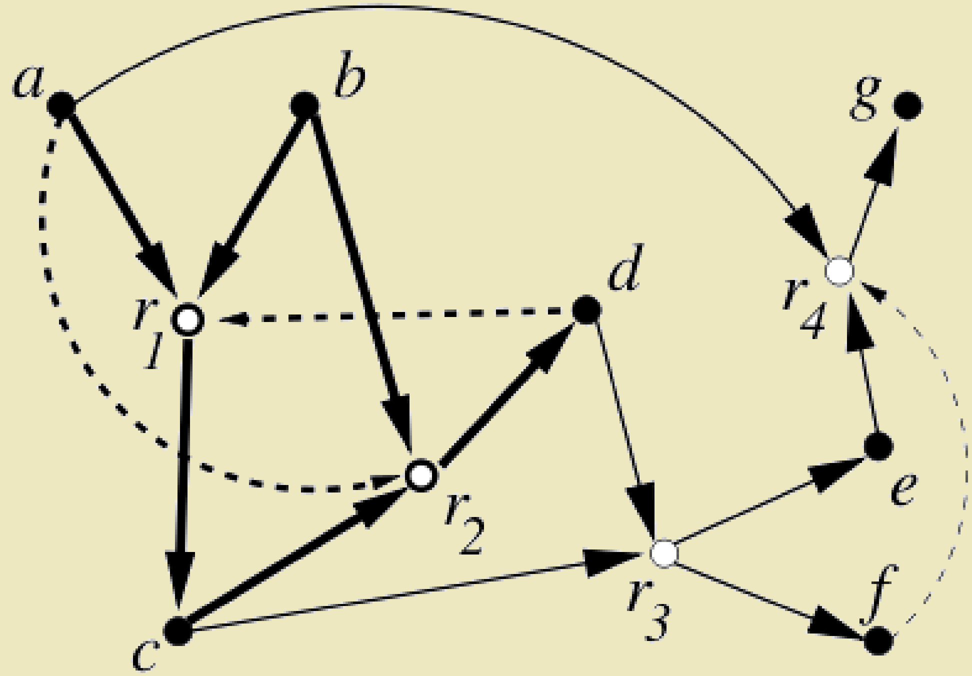
Autocatalytic Sets and the Origin of Life



Wim Hordijk, Mike Steel and Stuart Kauffman

Overview

- Origin of life problem.
- Catalytic reaction systems and autocatalytic sets.
- Algorithm for finding autocatalytic sets.
- Random catalytic reaction systems.
- Simulation results.
- Conclusions.

The Origin of Life

- Life as we know it depends on interplay between DNA & proteins.
- “Chicken & Egg” problem:
    - DNA & protein world too complex to have arisen all at once.
    - Which came first: the chicken (proteins, phenotype), or the egg (DNA, genotype)?
- Answer: *Probably neither one!*

The RNA World

- RNA is both genotype and phenotype:
  - It encodes genetic information.
  - It can act as a catalyst (ribozymes).
- Life started as collection of (self-)replicating catalytic RNA molecules.
- **Problems:**
  - How did the (building blocks of the) first RNA molecules appear?
  - How to go from RNA to DNA & proteins?

Prebiotic Metabolism

- “Universal” metabolism: citric acid cycle.
- Reverse citric acid (rTCA) cycle generates basic building blocks of organic molecules.
- rTCA cycle is an autocatalytic cycle.
- Emergence of autocatalytic cycle(s) as intermediate step between (in)organic chemistry and RNA World.

From RNA to DNA & Proteins

- RNA starts encoding and synthesizing proteins.
- Proteins take over catalytic functions, enabling longer RNA, enabling more complex proteins...
- Proteins start constructing DNA molecules.
- DNA (more stable) takes over storing genetic information, enabling even more complex proteins...
- Hypercycles, Darwin-Eigen cycle, autocatalytic sets...

Origin of Life Scenario

(In)organic chemistry

↓

Prebiotic metabolism

↓

RNA World

↓

Proteins

↓

DNA

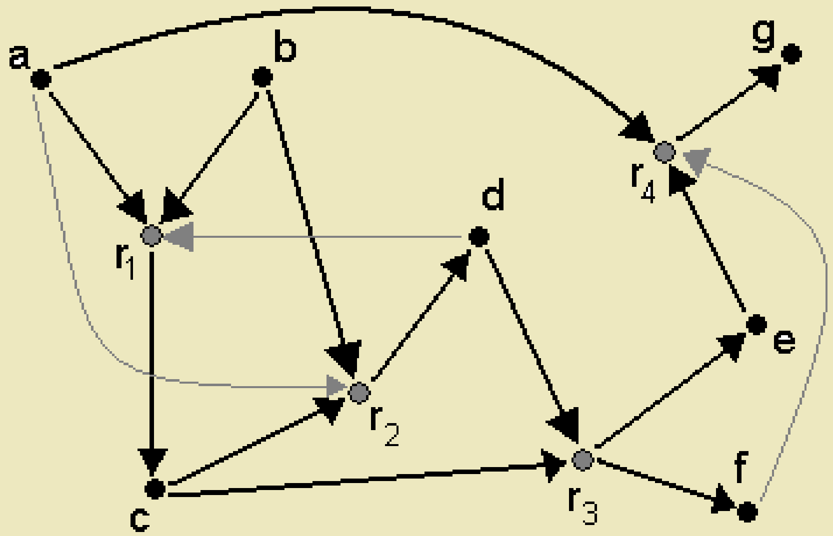
Motivation

- Autocatalytic sets seem to play an important role in several steps in the given OoL scenario.
- Unclear how likely it is that autocatalytic sets emerge “spontaneously”.

Overview

- Origin of life problem.
- Catalytic reaction systems and autocatalytic sets.
- Algorithm for finding autocatalytic sets.
- Random catalytic reaction systems.
- Simulation results.
- Conclusions.

Catalytic Reaction System



$$X = \{a, b, c, d, e, f, g\}$$

$$R = \{r_1, r_2, r_3, r_4\}$$

$$C = \{(d, r_1), (a, r_2), (f, r_4)\}$$

$$F = \{a, b\}$$

$$r_1 : a + b \xrightarrow{d} c$$

$$r_2 : b + c \xrightarrow{a} d$$

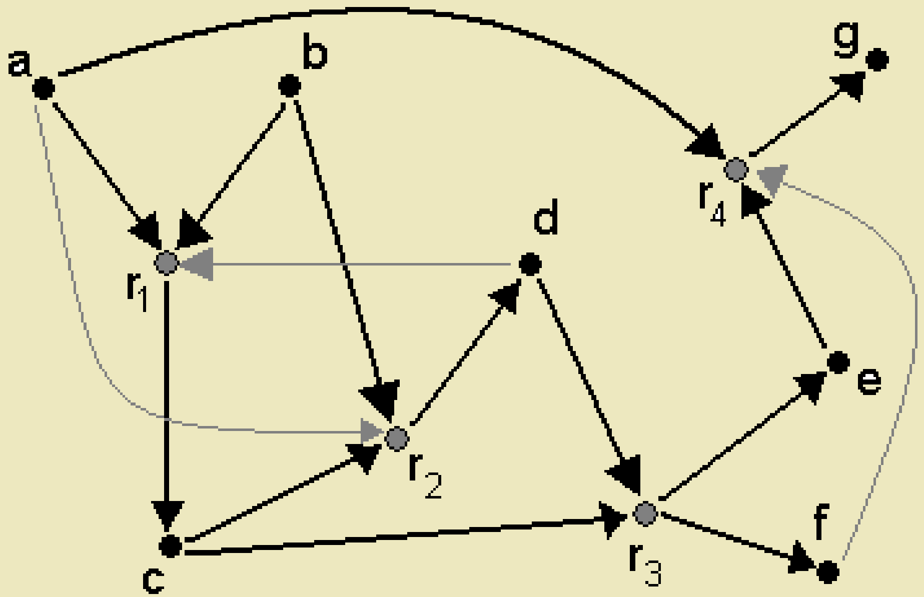
$$r_3 : c + d \longrightarrow e + f$$

$$r_4 : a + e \xrightarrow{f} g$$

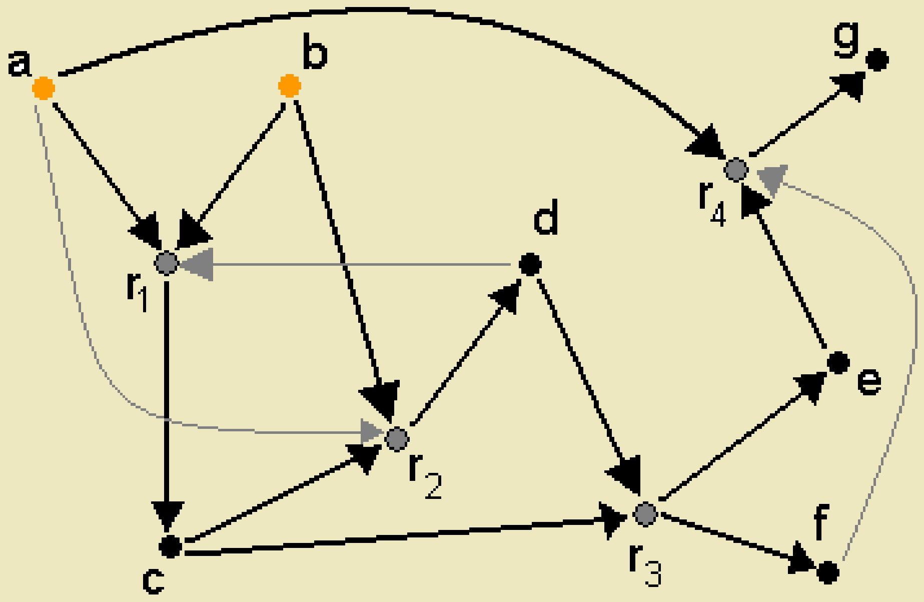
Catalytic Reaction System

- Let *R* be a set of reactions and *X* be a (sub)set of molecules:
    - The closure of *X* relative to *R* is the set of all molecules that can be constructed from *X* by repeated application of reactions in *R*.

Closure of Food Set

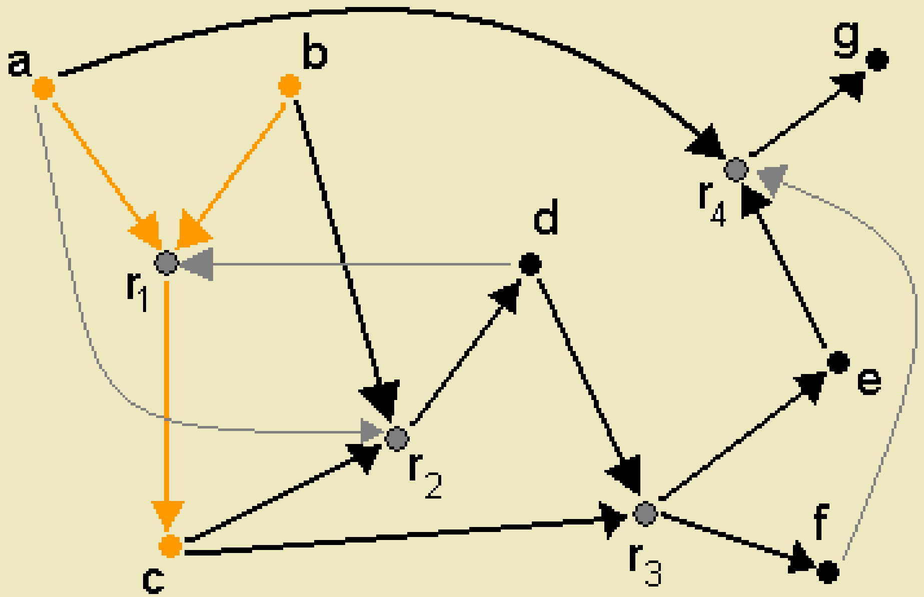


Closure of Food Set



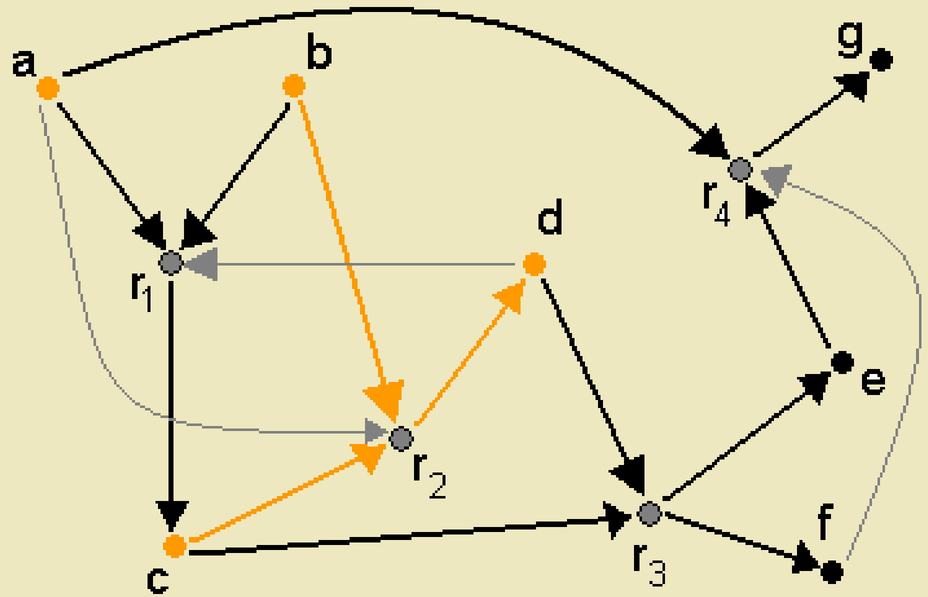
$$F = \{a, b\}$$

Closure of Food Set



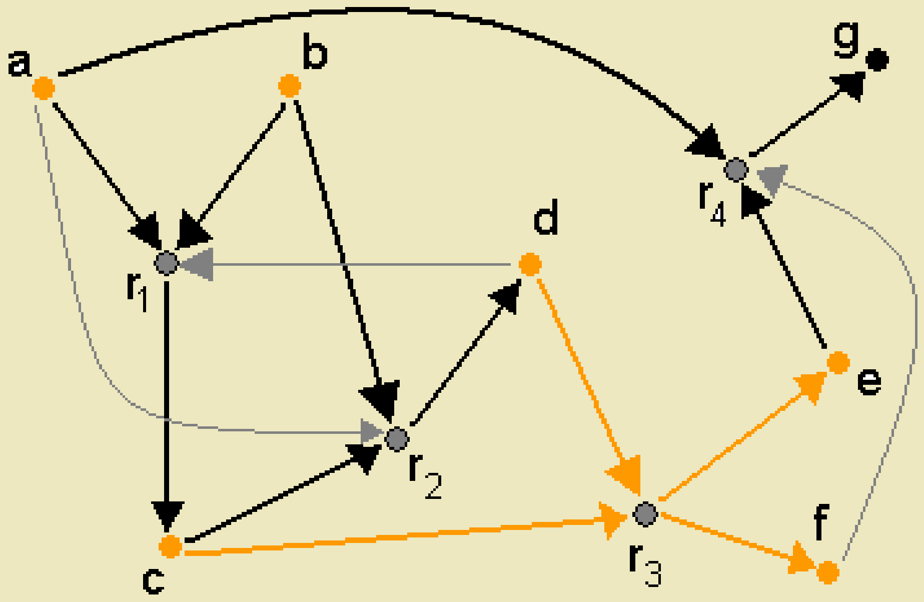
$$cl_R(F) = \{a, b, c, \ldots\}$$

Closure of Food Set



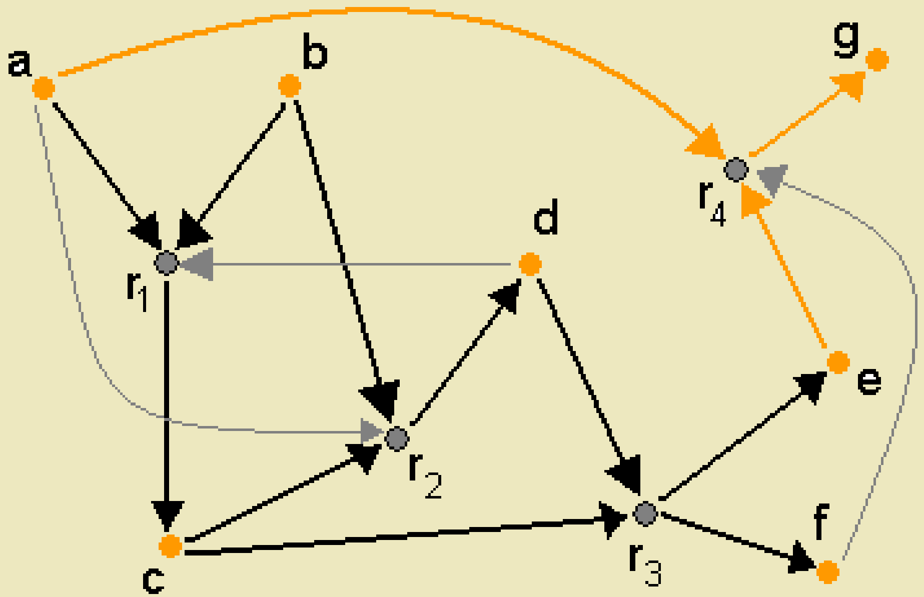
$$cl_R(F) = \{a, b, c, d, \ldots\}$$

Closure of Food Set



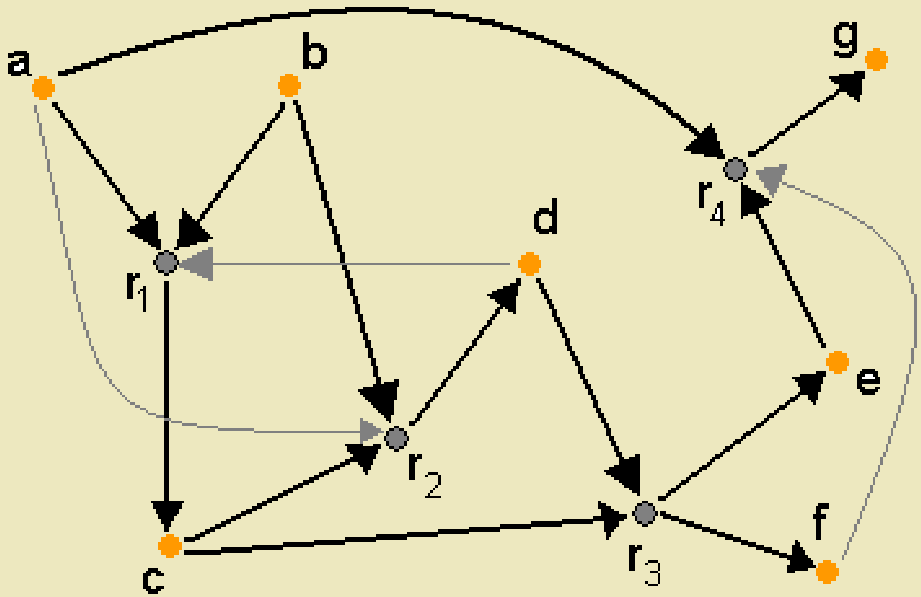
$$cl_R(F)=\{a, b, c, d, e, f, \ldots\}$$

Closure of Food Set



$$cl_R(F) = \{a, b, c, d, e, f, g, \ldots\}$$

Closure of Food Set

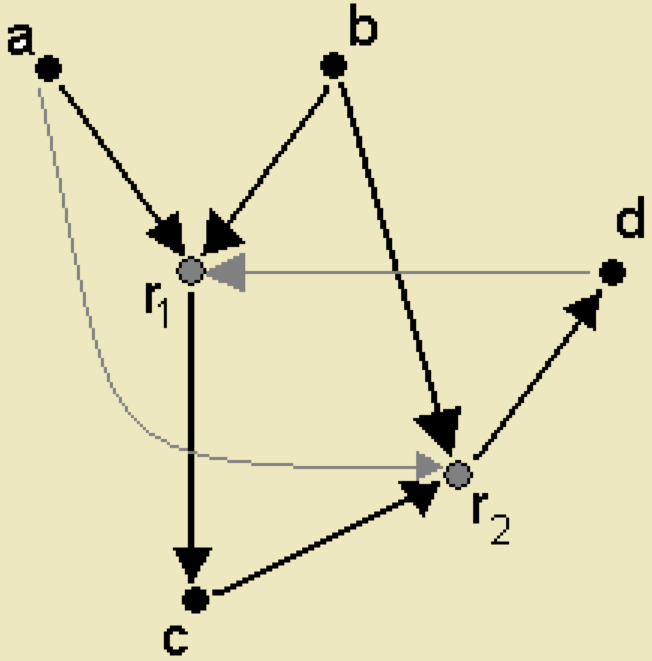


$$cl_R(F) = \{a, b, c, d, e, f, g\}$$

Autocatalytic Sets

- Given a catalytic reaction system with food set *F*, a (sub)set *R* of reactions is:
  - Reflexively autocatalytic (RA) if every reaction in *R* is catalyzed by at least one molecule in the closure of *F* relative to *R*.
  - F-generated (F) if every reactant in *R* is in the closure of *F* relative to *R*.
  - RAF if both RA and F.
- Autocatalytic cycles & sets are specific instances of an RAF set.

Example of RAF Set



$$F = \{a, b\}, R = \{r_1, r_2\}$$

$$cl_R(F) = \{a, b, c, d\}$$

Overview

- Origin of life problem.
- Catalytic reaction systems and autocatalytic sets.
- Algorithm for finding autocatalytic sets.
- Random catalytic reaction systems.
- Simulation results.
- Conclusions.

Finding RAF Sets

- Reduction rule: For each reaction *r* in *R* for which

  1) all catalysts, or

  2) one or more reactants

  are not in the closure of *F* relative to the current *R*, remove *r* from *R*.

- Reduction function:
  - $\gamma(R)$: apply above reduction rule to current *R*.

Finding RAF Sets

- Construct a sequence of reaction sets:

$$R_{i+1} = \gamma(R_i) \qquad R_{i+1} \subseteq R_i$$

- Consider R_∞ (in practice $\infty \leqslant |R|$).

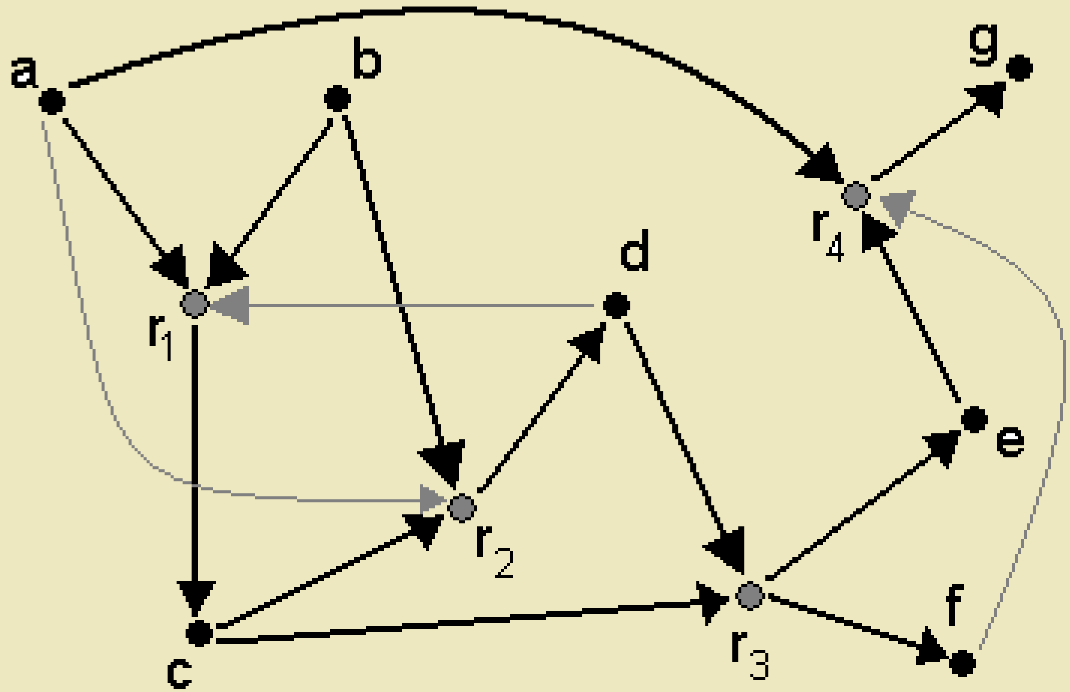
- Theorem:
  A catalytic reaction system has an RAF set if and only if R_∞ is non-empty (in which case it is the maximal RAF set).

RAF Algorithm

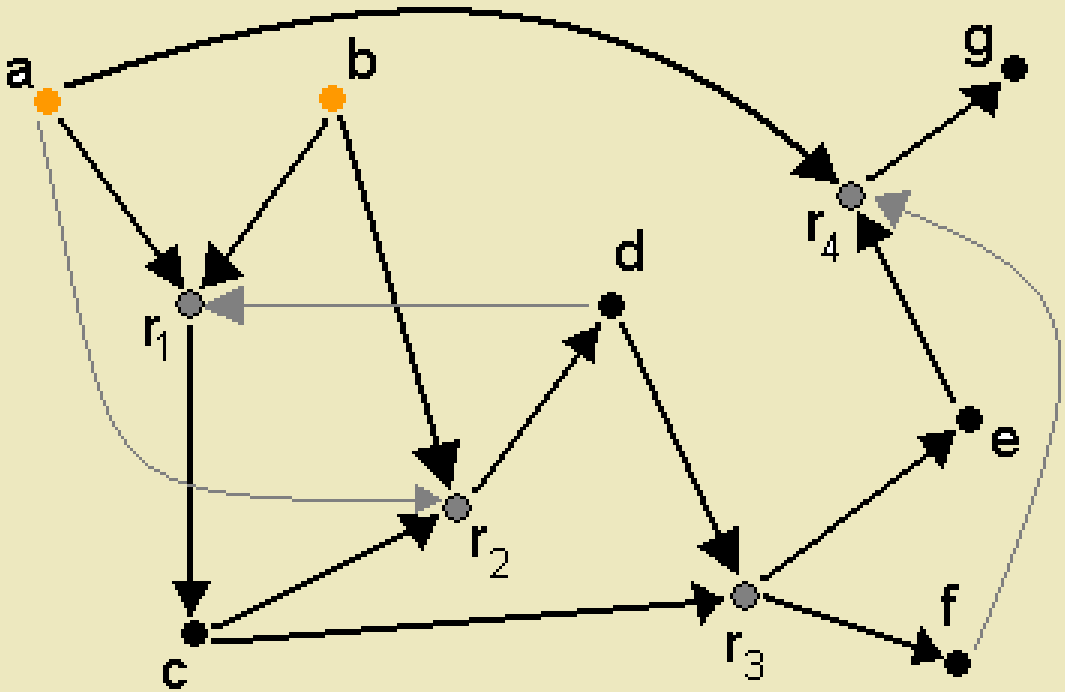
1. Start with the complete reaction set *R* and food set *F*.
2. Compute the closure of the food set *F* relative to the current reaction set *R*.
3. Apply the reduction rule to *R*.
4. Repeat steps 2 and 3 until no more reactions can be removed.

Running time: $O(|X||R|^3)$ $O(|R|^2 \log |R|)$

Example

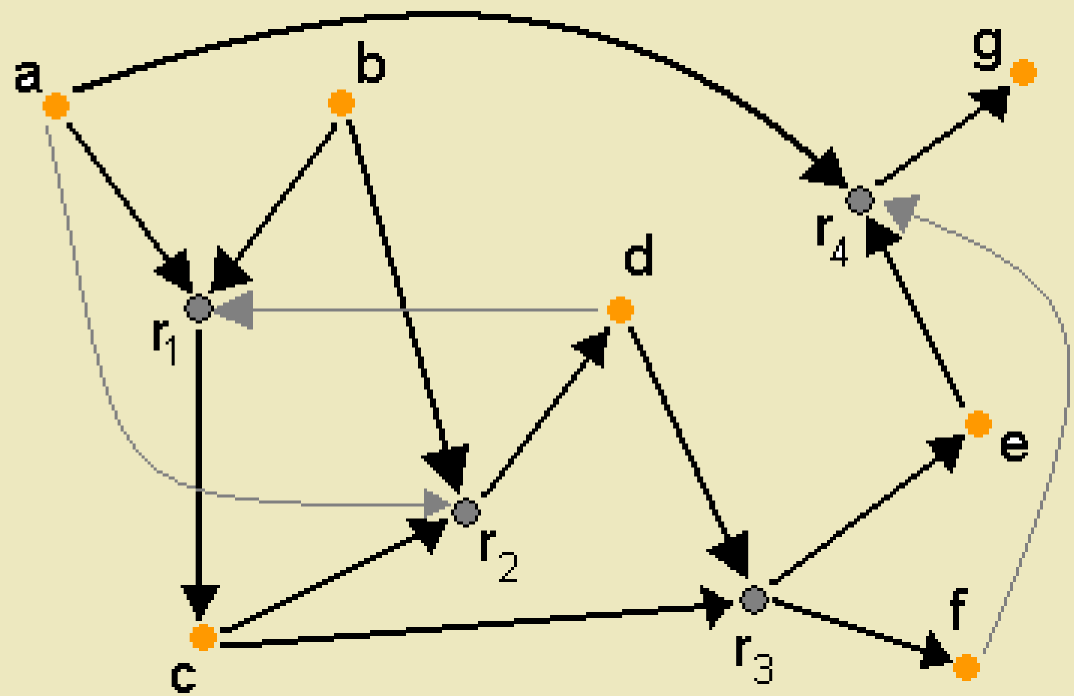


Example



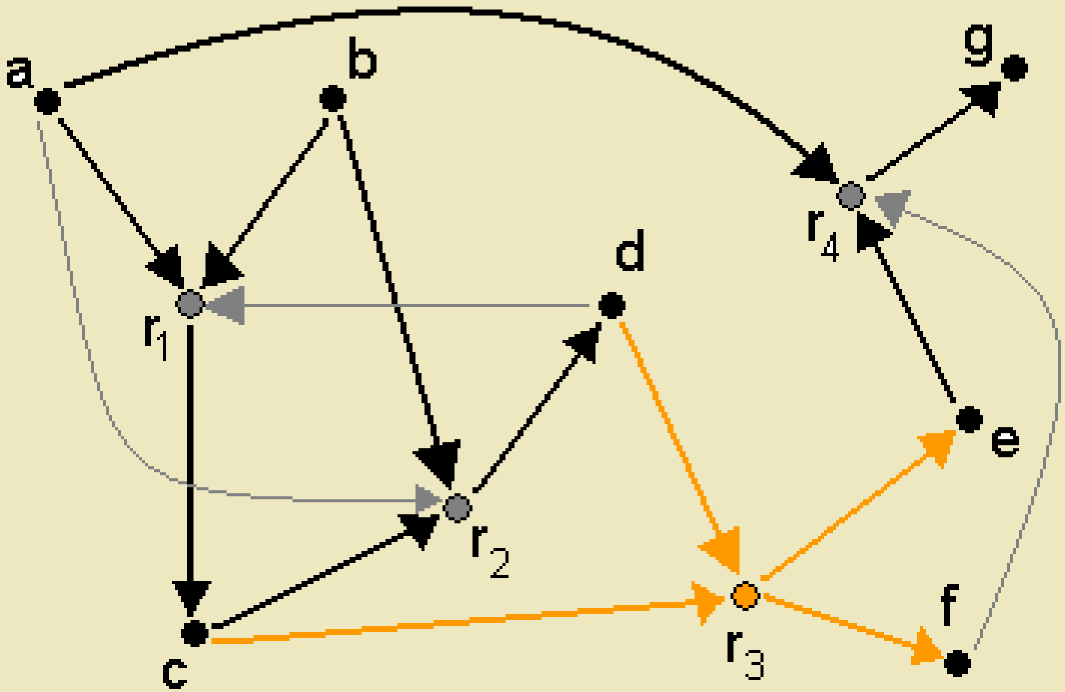
$$F = \{a, b\}$$

Example



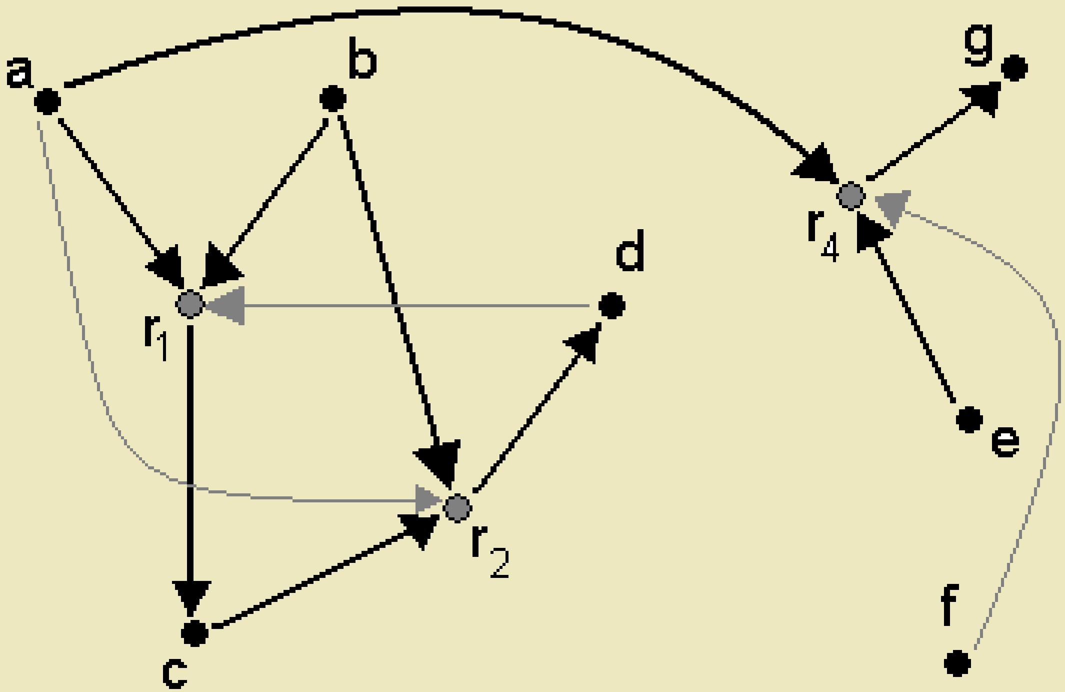
$$cl_R(F)=\{a,b,c,d,e,f,g\}$$

Example

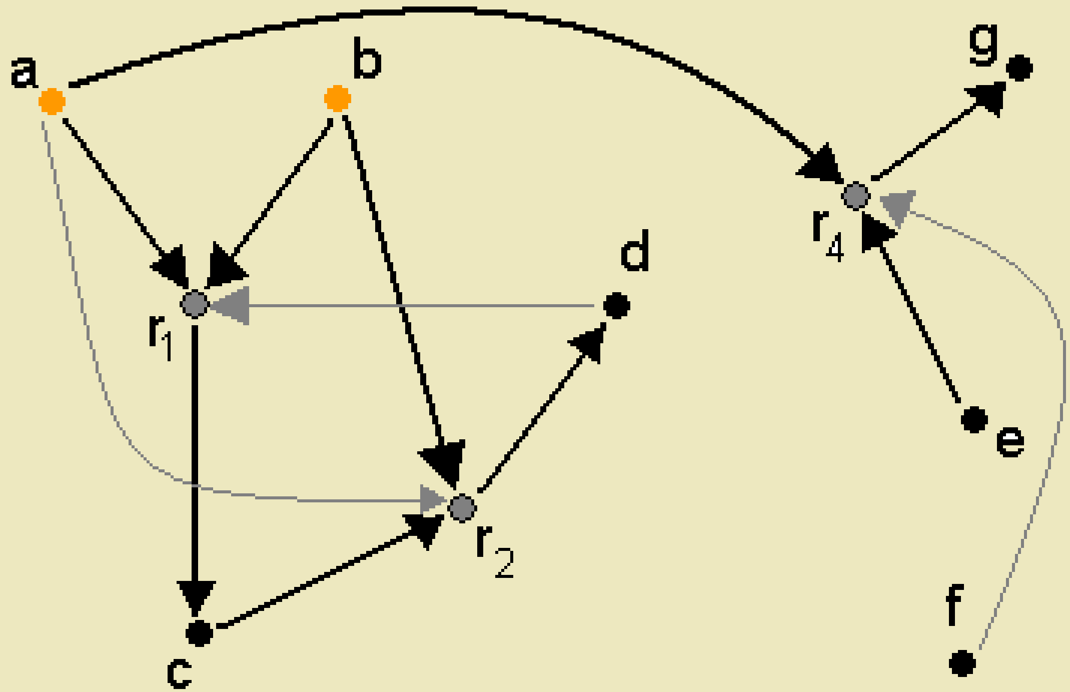


no $x \in cl_R(F) : (x, r_3) \in C$

Example

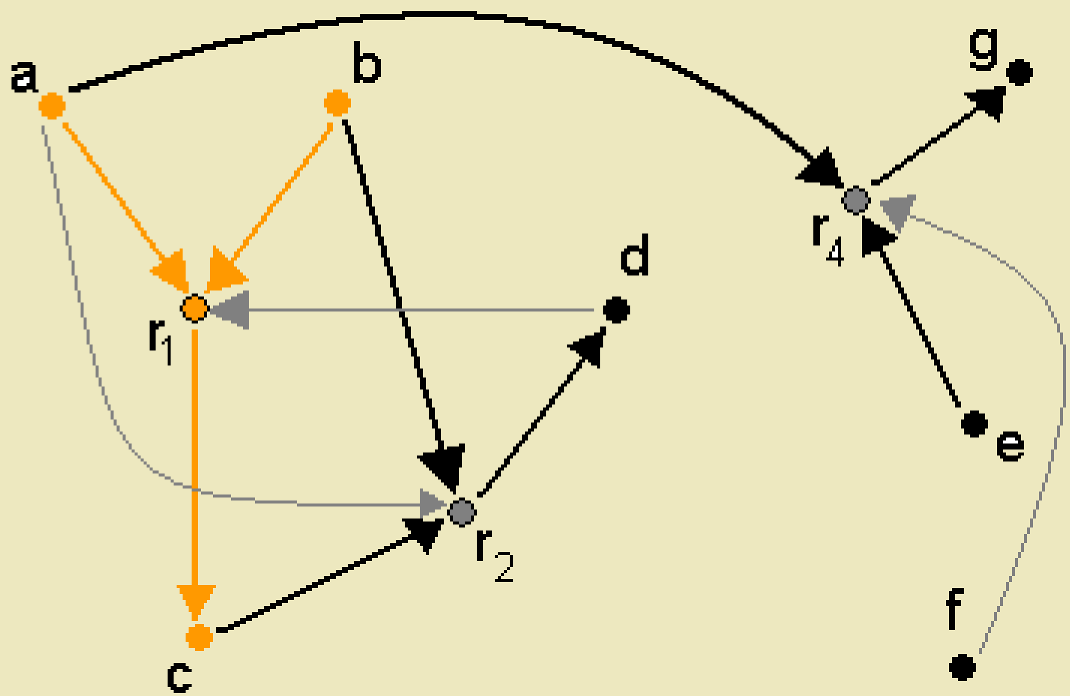


Example



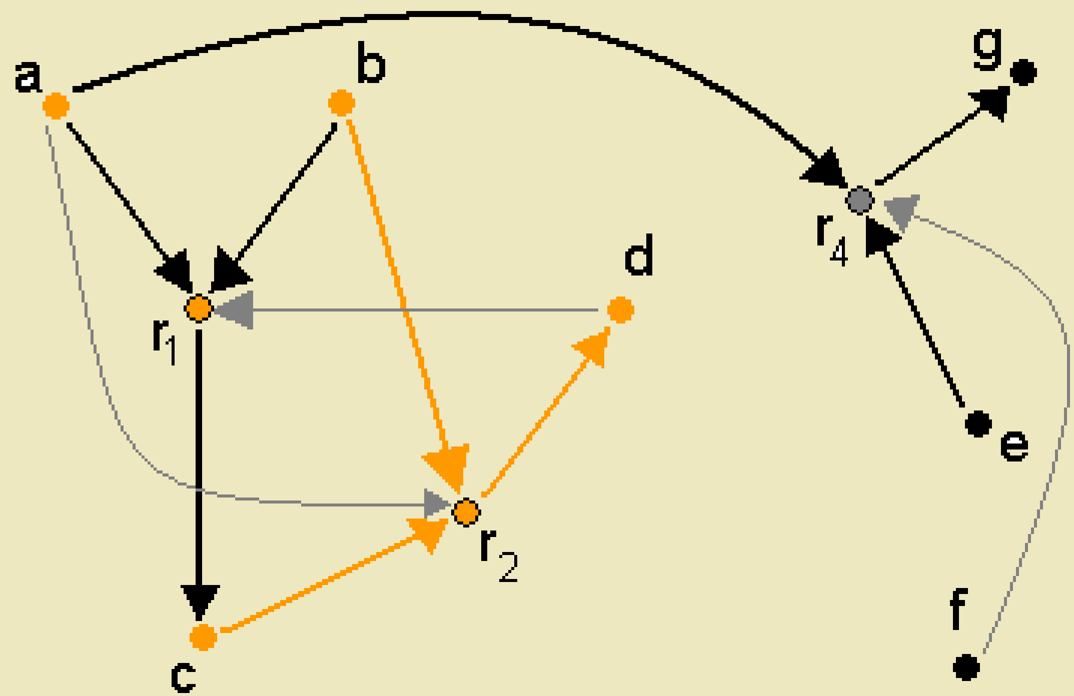
$$F = \{a, b\}$$

Example



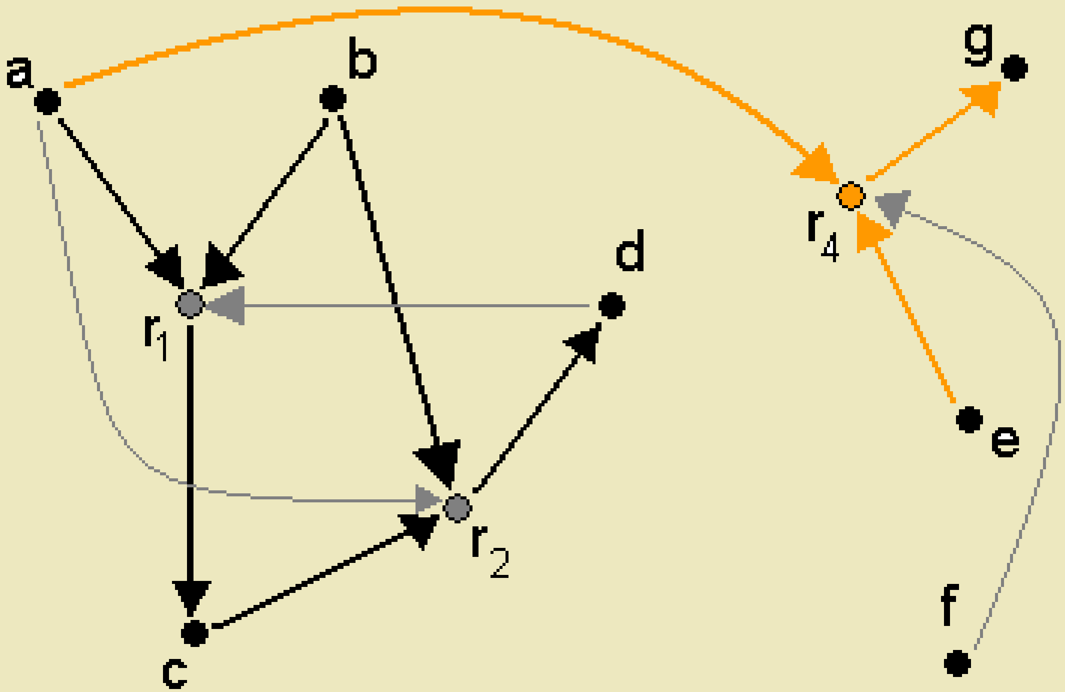
$$cl_R(F)=\{a,b,c,\ldots\}$$

Example



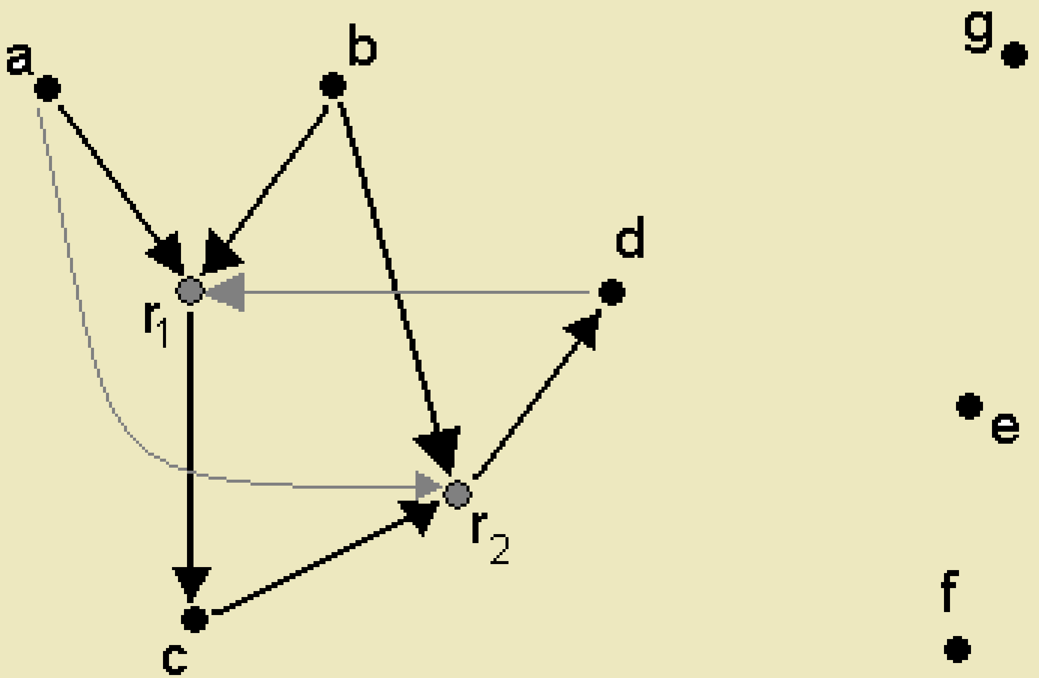
$$cl_R(F)=\{a,b,c,d\}$$

Example

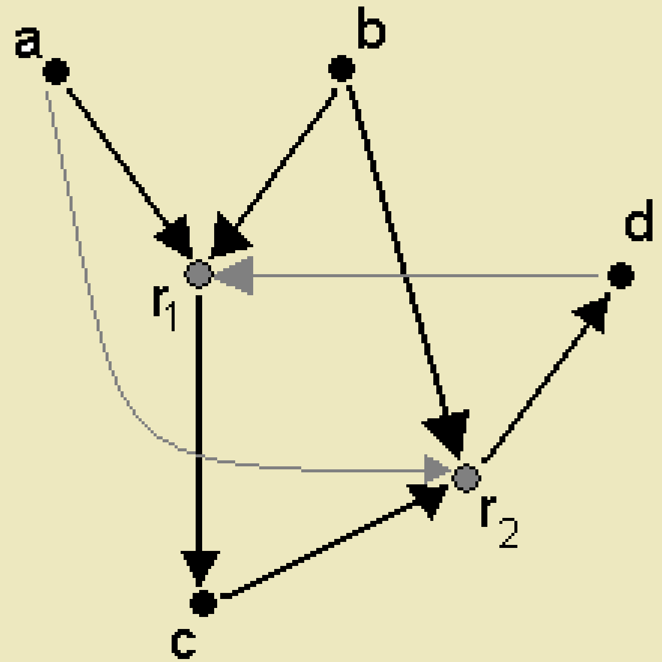


$$e \notin cl_R(F)$$

Example



Example



Maximal RAF set R_∞

Overview

- Origin of life problem.
- Catalytic reaction systems and autocatalytic sets.
- Algorithm for finding autocatalytic sets.
- Random catalytic reaction systems.
- Simulation results.
- Conclusions.

Random Catalytic Reaction Systems

- Molecules *X*:
  - Bit strings of length at most *n*.
- Reactions *R*:
  - Ligation: 000 + 11111 --> 00011111
  - Cleavage: 1010101 --> 10101 + 01
- Food set *F*:
  - All molecules of length at most *t* (<< *n*).

Random Catalytic Reaction Systems

- Catalyzations *C*:
  - Assign at random according to some probability: $Pr[(x, r) \in C] = p(n)$
- Expected (average) number of reactions catalyzed per molecule:
  - $f(n) = p(n) \times |R|$

Probability of RAF Sets

- Let P_n be the probability that a random catalytic reaction system has an RAF set and consider $P_\infty = \lim_{n \to \infty} P_n$

- Kauffman:
  - $p(n) = c \quad (f(n) \propto 2^n) \quad \Rightarrow P_\infty = 1$
- Lifson:
  - $f(n) = c \quad (p(n) \propto 1/|R|) \quad \Rightarrow P_\infty \neq 1$

Probability of RAF Sets

- Steel (2000):
  - $f(n) \geq cn^2 \Rightarrow P_\infty = 1$
  - $f(n) < \frac{1}{3} e^{-1} \Rightarrow P_\infty = 0$
- Main questions:
  - What happens between constant and quadratic level of catalyzation?
  - For fixed *n*, how does the transition from prob 0 to 1 happen with increasing *p(n)*?

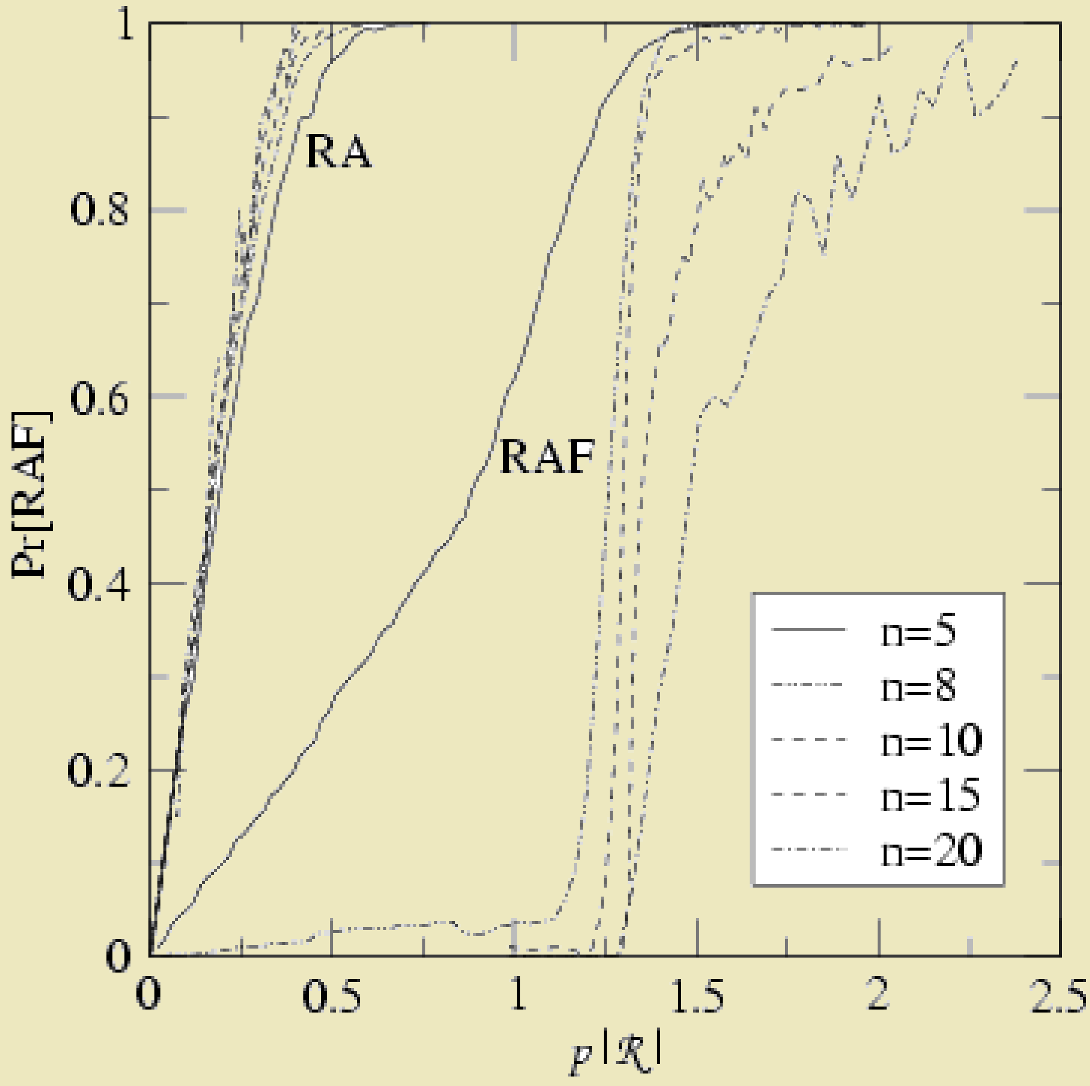
Computer Simulations

- With n up to 20 and varying values of $p(n)$.
- $t=2 \Rightarrow F=\{0,1,00,01,10,11\}$
- 100 to 10,000 instances (depending on n).
- Apply RAF algorithm to instances of random catalytic reaction systems.
- Count how many times an RAF is found.

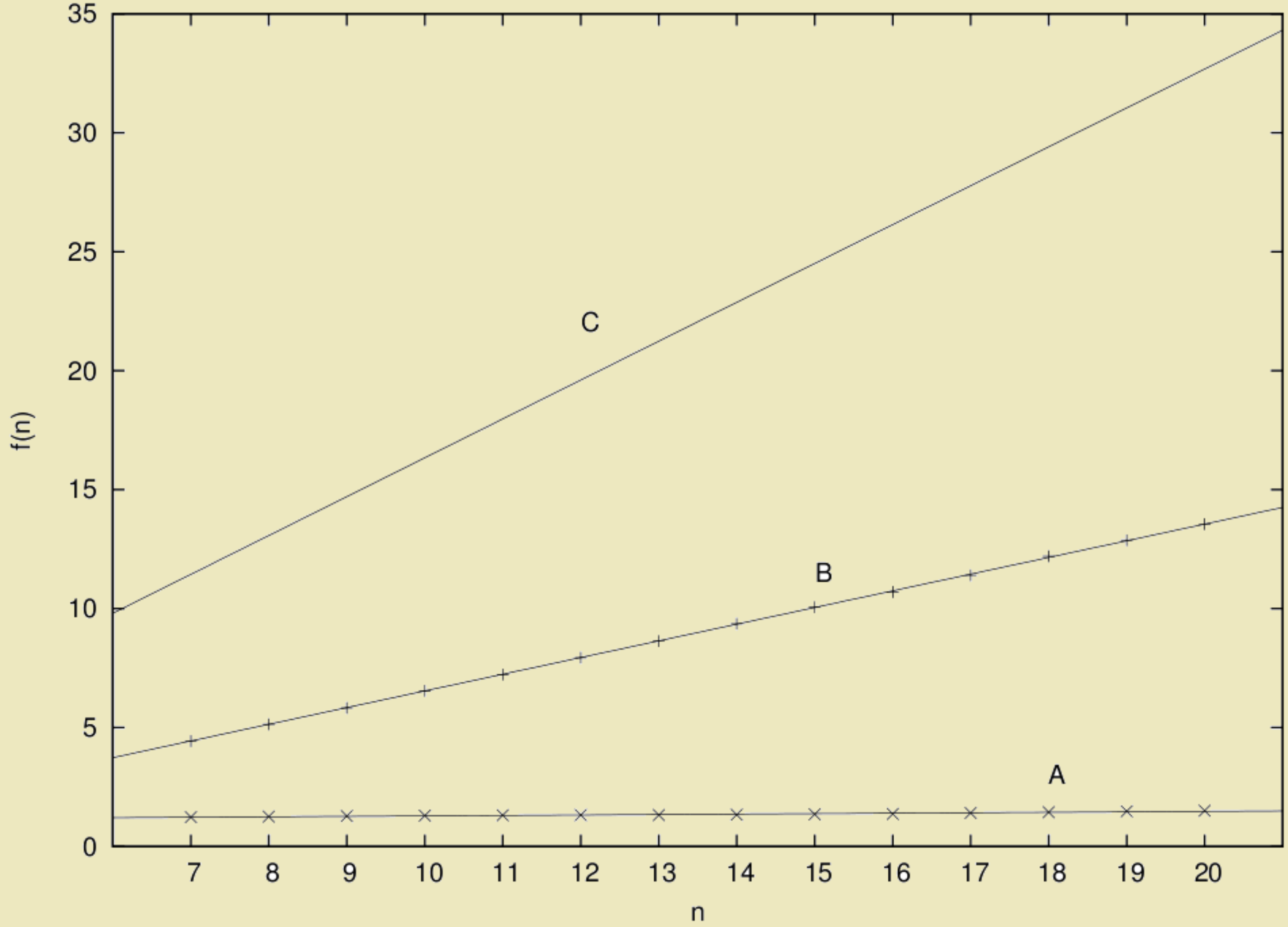
Overview

- Origin of life problem.
- Catalytic reaction systems and autocatalytic sets.
- Algorithm for finding autocatalytic sets.
- Random catalytic reaction systems.
- Simulation results.
- Conclusions.

Probability of RAF Sets

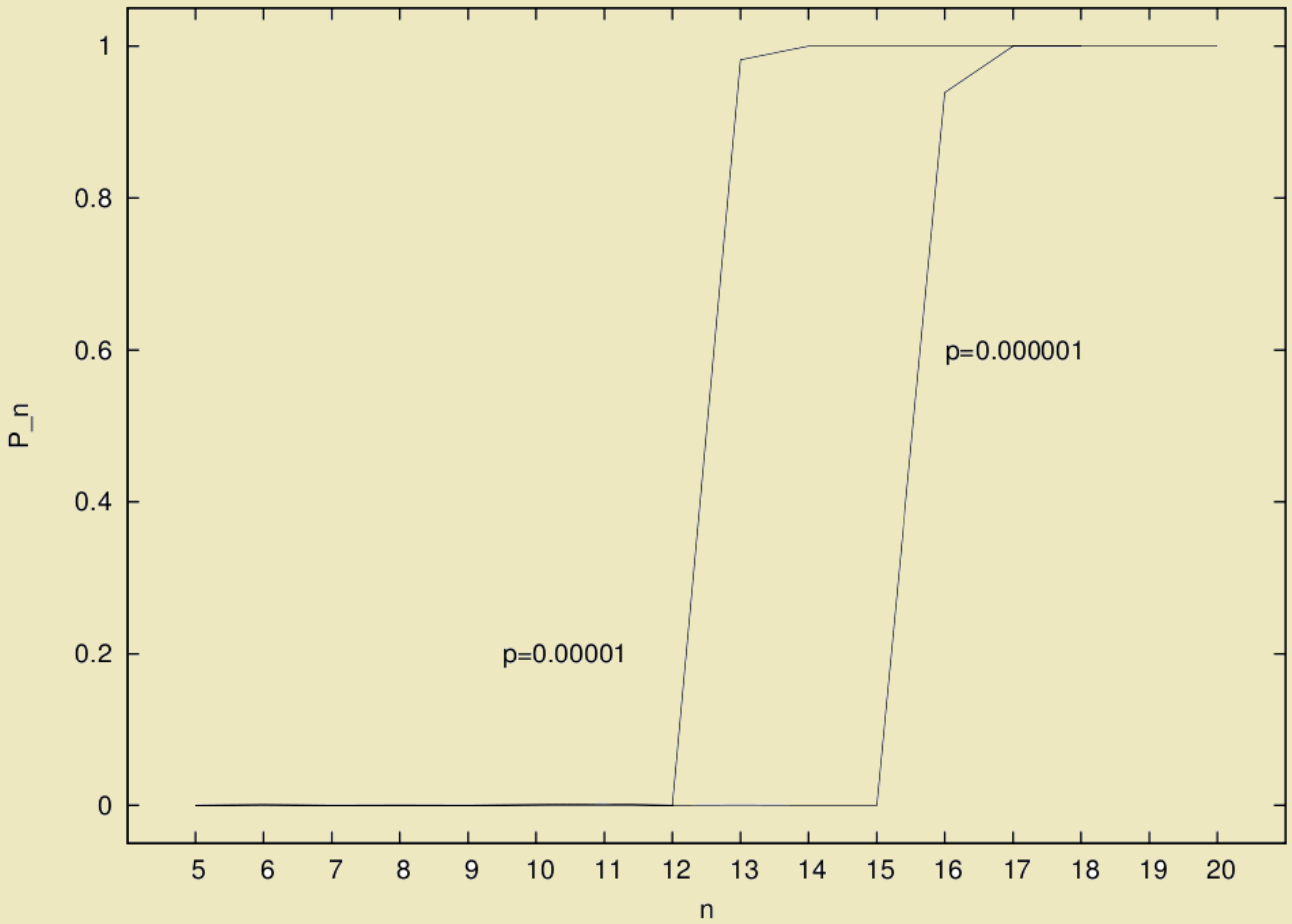


Linear Growth Rate



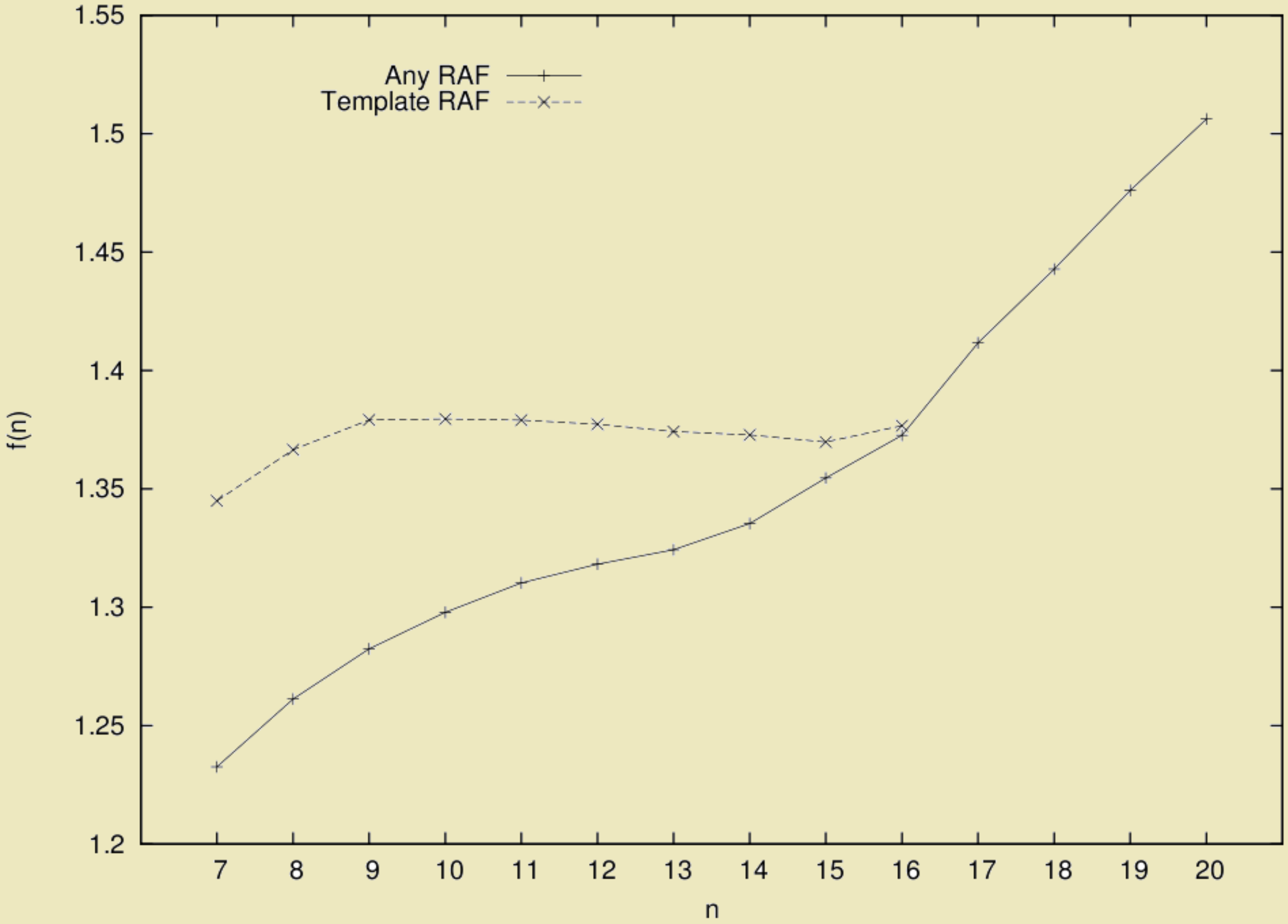
A: 1.0970+0.0189n **B:** -0.4736+0.7012n **C:** 1.6339n

Minimum Molecule Set Size



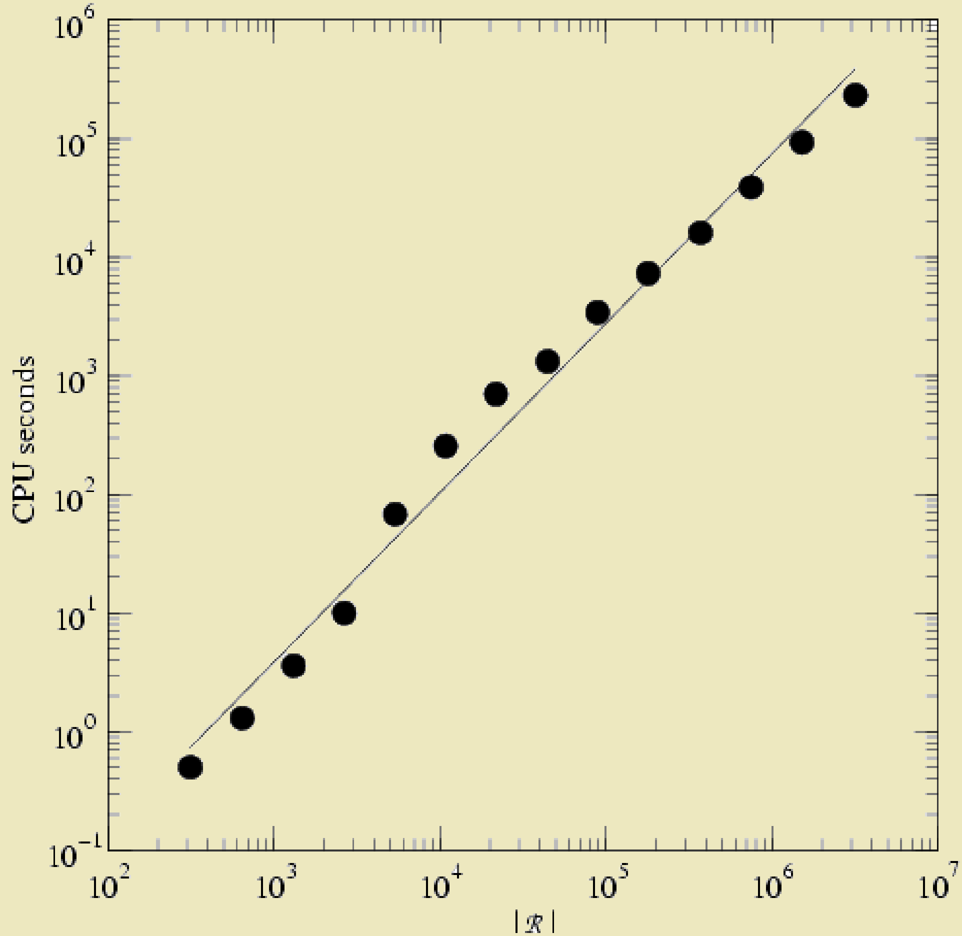
p=0.000002: n=15, |*X*|=65,000

Template-based Catalysis



Catalysts must match complement of 4-site reaction template.

Running Time



$$\tau = 0.0002\,|R|^{1.43}$$

Overview

- Origin of life problem.
- Catalytic reaction systems and autocatalytic sets.
- Algorithm for finding autocatalytic sets.
- Random catalytic reaction systems.
- Simulation results.
- Conclusions.

Conclusions

- Formalized the notion of catalytic reaction systems and autocatalytic (RAF) sets.
- Introduced a polynomial-time algorithm for finding RAF sets in catalytic reaction systems.
- Shown that a linear growth rate in level of catalysis (with size *n* of largest molecules) suffices for emergence of RAF sets.
- Derived required minimum size of molecule set given a fixed probability of catalysis.
- Provided an example of model extensions, such as template-based catalysis.

Suggestions for Future Work

- Study molecular dynamics on reaction graphs and RAF sets.
- Study real biochemical reaction networks.
- Evolve catalytic reaction networks and/or RAF sets (e.g. with genetic algorithms).

References

- M. Steel. The emergence of a self-catalysing structure in abstract origin-of-life models. *Applied Mathematics Letters* **3**, 91-95, 2000.
- W. Hordijk and M. Steel. Detecting autocatalytic, self-sustaining sets in chemical reaction systems. *Journal of Theoretical Biology* **227**, 451-461, 2004.
- E. Mossel and M. Steel. Random biochemical networks: The probability of self-sustaining autocatalysis. *Journal of Theoretical Biology* **233**, 327-336, 2005.
- W. Hordijk, J. Hein and M. Steel. Autocatalytic sets and the origin of life. *Entropy* **12**, 1733-1742, 2010.
- 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(5), 3085-3101, 2011.

Which came first...?