

Abundant reactions in autocatalytic networks with locally ranked supplier cores

14 September 2026

Abstract

A reflexively autocatalytic and food-generated set (RAF) of a catalytic reaction system is a set of reactions that can be built up from a food set and in which every reaction is catalysed from within. With the empty set adjoined, the RAFs of a finite system form a union-closed family, and Steel asked whether some reaction always lies in at least half of them. We prove this for every system containing a *locally ranked complete-supplier core*: a nonempty set U of reactions in which each reaction has one internal supplier producing all of its non-food reactants and, unless food already catalyses it, a catalyst, and in which the internal substrate-production digraph is acyclic. Nothing is assumed outside U : the surrounding system may contain arbitrary substrate cycles, alternative producers and catalytic feedback into the core, and the core itself need not be elementary. The proof identifies each exterior fibre of the RAF family exactly, through food closure of the combined selected set, with a supported family under an upward constraint, and applies the dependency counting injection of Lozin and Zamaraev. The resulting occupancy inequality holds in every exterior context and with the empty set counted, and it yields exterior-weighted and bounded-distortion sampling statements with the sharp constant $1/(1 + \kappa)$, deterministic and expected bounds for the number of RAFs destroyed by deleting one core reaction, transport of the certificate across exterior edits, and distinct abundant reactions from disjoint cores. Every elementary core is a locally ranked supplier core, so the theorem strictly extends the elementary-core abundance theorem; a three-reaction system separates it from any hypothesis on the global substrate digraph. A uniform three-chain family with no elementary RAF shows that the core average is sharp while individual core reactions can be arbitrarily rare, and that the RAF family need not be the support family of any digraph. The main theorem and the source-level consequences are verified in Lean 4 against Mathlib.

Keywords. autocatalytic sets; RAF theory; union-closed sets conjecture; Horn functions; dependency property; formal verification.

MSC 2020. 05D05, 92C42, 06A15, 68V20.

1 Introduction

1.1 Autocatalytic sets and Steel’s question

Collectively autocatalytic sets were proposed by Kauffman [13, 14] as a route by which a mixture of simple molecules could organise into a self-sustaining chemistry. Hordijk and Steel [11] made the idea precise in the RAF formalism. A *catalytic reaction system* consists of molecule types, reactions with reactant and product sets, a distinguished *food* set, and a catalysis relation. A set of reactions is *food-generated* if all of its reactants can be produced from food, in some order, using only the selected reactions, and it is *reflexively autocatalytic* if every selected reaction has a catalyst that is food or is produced by the selected reactions. A RAF is

a nonempty set that is both. The largest RAF inside any given reaction set is computable in polynomial time, and the formalism has been developed extensively [12, 16, 22, 23].

The union of two RAFs is again a RAF. Adjoining the empty set, the RAFs of a finite system therefore form a *union-closed family* of subsets of the reaction set, and the map sending an available reaction set to the largest RAF it contains is an interior operator. Steel [21] studied this operator and asked, in his concluding section, whether there is always a reaction that lies in at least half the RAFs, for either an elementary or a general catalytic reaction system. A positive answer for all systems would follow from Frankl’s union-closed sets conjecture [2, 9], which asserts that every finite union-closed family other than $\{\emptyset\}$ has an element lying in at least half of its members. That conjecture is open. After a long period in which only logarithmic bounds were known [25], Gilmer [10] proved a constant lower bound by an entropy argument, quickly improved to $(3 - \sqrt{5})/2 \approx 0.382$ and beyond [1, 3, 4, 20]; the constant $1/2$ remains out of reach.

A companion paper [6] shows that the RAF version of the question has two faces. If the system contains an *elementary core*, a nonempty RAF all of whose reactants are food, then a reaction of that core lies in at least half of all RAFs of the whole system, whatever the remaining reactions do. On the other hand, every finite union-closed family containing the empty set is the RAF family of a system without an elementary core, so a half-frequency theorem for the coreless class is equivalent to the unrestricted conjecture. The companion paper therefore asks which structural hypotheses on a system *without* an elementary core imply abundance, and observes that such hypotheses must constrain reactant requirements, because the universal construction already makes the catalytic structure trivial.

This paper answers that question for a natural class. The elementary hypothesis says that core reactions need nothing but food. We replace it by two conditions on reactant supply *inside the core only*: every core reaction has a single internal supplier providing all of its non-food reactants (and a catalyst, unless food already catalyses it), and the internal substrate-production digraph is acyclic. Both conditions are checked on the core incidences alone. Outside the core nothing is assumed.

1.2 Main results

Throughout, $\mathcal{F}$ denotes the family of RAFs of a finite catalytic reaction system together with the empty set, $M = |\mathcal{F}|$, $N = M - 1$ is the number of nonempty RAFs, and $f(r)$ is the number of members of $\mathcal{F}$ containing the reaction r . A reaction is *abundant* if $2f(r) \geq M$. Because the empty set contributes to the denominator but to no numerator, abundance is slightly stronger than the half-frequency statement over nonempty RAFs alone; Steel’s question is the weaker form, and all our positive results give the stronger one.

A reaction p is a *complete supplier* of a reaction r if p produces every non-food reactant of r and, unless some food molecule catalyses r , some product of p catalyses r . A nonempty set U of reactions is a *complete-supplier core* if every $r \in U$ has a complete supplier in U , and it is *locally ranked* if there is a map $\lambda : U \rightarrow \mathbb{N}$ with $\lambda(p) < \lambda(r)$ whenever $p, r \in U$ and some product of p is a non-food reactant of r (Definition 2.2). The rank condition inspects every internal substrate-production incidence, including self-incidences, not only the chosen suppliers; catalytic incidences are not part of it, and catalytic cycles inside the core are what typically make the core autocatalytic.

Theorem 1.1 (Locally ranked supplier-core abundance; Theorem 2.5). *Let U be a locally ranked complete-supplier core of a finite catalytic reaction system. Then U is a RAF, and some reaction of U is abundant in $\mathcal{F}$. More precisely, for every set T of reactions outside U , the core parts $S = W \cap U$ of the members $W \in \mathcal{F}$ with $W \setminus U = T$ have average size at least $|U|/2$,*

and consequently

$$\sum_{W \in \mathcal{F}} |W \cap U| \geq \frac{1}{2} |U| M.$$

The fibrewise form is the real content. The proof has two steps. The first (Lemma 3.1) is an exact description of each exterior fibre: for a fixed exterior selection T , a set $S \subseteq U$ satisfies $S \cup T \in \mathcal{F}$ if and only if the exterior reactions of T are supported in the food closure of $S \cup T$ and every selected core reaction has its non-food reactants and a catalyst produced by some selected reaction. The rank hypothesis is what turns these local producer requirements into genuine food generation; the exterior condition is evaluated through the actual closure of the combined set and is monotone in S , but nothing is assumed about T on its own. The second step (Lemma 3.2) is the aggregate form of the counting theorem of Lozin and Zamaraev [15] for Horn functions with a dependency property: a family of subsets of U defined by an upward constraint together with per-element monotone support requirements, each of which is met whenever a designated companion is selected, has average member size at least $|U|/2$. The complete supplier is the companion. Summing over exterior fibres retains every exterior completion multiplicity and gives the global inequality, from which the witness follows by averaging.

The theorem strictly extends the elementary-core theorem of [6]: every elementary core is a locally ranked supplier core with constant rank (Proposition 2.3), while the three-reaction system of Section 5.1 has a locally ranked core that is not elementary, and its ambient substrate digraph has a cycle, so no hypothesis on the global substrate digraph covers it. The three-chain family of Section 5.2 has no elementary RAF at all.

The fibrewise inequality has consequences that a bare existence statement would not give (Section 4).

- *Transport and disjoint witnesses.* The hypotheses inspect only the core incidences, the food set and the catalysis relation, so the conclusion survives arbitrary edits of exterior reactions, enlargement of food and addition of catalytic incidences (Corollary 4.1). Pairwise disjoint cores give distinct abundant reactions.
- *Weighting.* The half bound survives any nonnegative reweighting of exterior contexts (Corollary 4.2). If instead each member of $\mathcal{F}$ receives a weight that varies by a factor at most κ within every exterior fibre, the expected core occupancy is at least $|U|/(1 + \kappa)$ and some core reaction has weighted frequency at least $1/(1 + \kappa)$; the constant is sharp (Theorem 4.3).
- *Deletion.* Some core reaction, when deleted, destroys at least half of the nonempty RAFs, leaving at most $\lfloor (N - 1)/2 \rfloor$; a uniformly random core reaction leaves at most $(N - 1)/2$ in expectation (Proposition 4.4).

Section 5 gives the examples and sharpness statements. A uniform family with three chains of length ℓ and three gates (Section 5.2) has exactly $\ell + 8$ members in $\mathcal{F}$, its designated core B has average occupancy exactly one half, and yet its individual core reactions have frequencies ranging from $(\ell + 4)/(\ell + 8) \rightarrow 1$ down to $4/(\ell + 8) \rightarrow 0$: the certificate identifies a set containing a witness, not the witness. The family has no elementary RAF, has transversal number three, has exactly three irreducible RAFs whose unions do not exhaust $\mathcal{F}$, and is not the support family of any directed graph on its reactions, so the digraph-support characterisation of [5] does not apply to it. It admits a conservation bookkeeping that leaves $\mathcal{F}$ unchanged (Proposition 5.5). Section 6 separates candidate validation, which is quadratic in the core and needs no enumeration, from discovery and from witness identification. Section 7 lists what is not claimed and the open questions this leaves.

1.3 Relation to prior work and what is new

The RAF framework and the question are due to Hordijk and Steel [11] and Steel [21]. The counting kernel is Theorem 4 of Lozin and Zamaraev [15], which proves Frankl’s conjecture for Boolean functions admitting a Horn DNF in which every variable occurring negatively has a companion occurring positively in each term where it is negative. The elementary-core theorem, its fibrewise form and its exterior-weighted, cycle and disjoint-core consequences are proved in the companion paper [6]; that paper also introduced the exterior-conditioned application of the counting injection. What is new here is (i) the localisation of the hypothesis to reactant supply inside a designated core, which admits non-food reactants in the core and arbitrary substrate cycles outside it; (ii) the exact closure-conditioned fibre representation through the combined selected set, which is what makes the localisation sound, together with the rank induction that establishes food generation; (iii) the bounded-distortion sampling theorem with its sharp constant; (iv) the deletion bounds; and (v) the uniform three-chain family, with its exact frequencies, its structural properties and its conservation lift. An earlier version of the theorem required a rank on the substrate digraph of the whole system; the three-reaction example of Section 5.1 shows that the local version is strictly stronger, and both versions are compiled. We make no claim on the unrestricted conjecture, on systems with three non-food-ready reactions in general, or on the kinetics of RAF operation.

The classical observation that a union-closed family containing a singleton or a two-element set satisfies the conjecture is due to Sarvate and Renaud [19] (see also Poonen [17]). Reimer’s average set size theorem [18] gives average member size at least $\frac{1}{2} \log_2 N$ for any union-closed family; our occupancy inequalities are of the same average-size type but give the bound $|U|/2$ under the structural hypothesis, for the core coordinates, and in every exterior fibre separately.

1.4 Formal verification

The main theorem and the source-level consequences marked in Appendix A have been formalised in Lean 4 [8] against Mathlib [24] directly from the literal RAF semantics of Section 2, compiled with warnings as errors, and depend only on the standard axioms `Classical.choice`, `Quot.sound` and `propext`. The formal main theorem takes an arbitrary system, a core, a rank and the supplier data as inputs and returns RAF status, the fibre inequalities, the global inequality and the witness; the desired average is not an input. The paper is written to be read without the formal development; the appendix states exactly which forms were compiled and which statements (the parametric family, the conservation lift, the general exterior-edit transport and the deletion interpretation) are ordinary mathematics only.

2 Ordinary RAF semantics and the source criterion

Definition 2.1 (Catalytic reaction system). A finite catalytic reaction system $\mathcal{Q}$ consists of a finite molecule set X , a finite reaction set R , a food set $F \subseteq X$, reactant and product sets $\rho(r), \pi(r) \subseteq X$ for each $r \in R$, and catalyst sets $C(r) \subseteq X$. Distinct reactions are distinct even if their incidence data coincide. For $W \subseteq R$ the *food closure* is

$$H_0(W) = F, \quad H_{j+1}(W) = H_j(W) \cup \bigcup_{r \in W, \rho(r) \subseteq H_j(W)} \pi(r), \quad H(W) = \bigcup_{j \geq 0} H_j(W).$$

Catalysis plays no role in this iteration. A nonempty $W \subseteq R$ is a *RAF* if for every $r \in W$,

$$\rho(r) \subseteq H(W) \quad \text{and} \quad C(r) \cap H(W) \neq \emptyset.$$

The first condition is food generation, the second is reflexive autocatalysis evaluated in the closure of the selected set. Write

$$\mathcal{F} = \{\emptyset\} \cup \{W \subseteq R : W \text{ is a RAF}\}, \quad M = |\mathcal{F}|, \quad N = M - 1, \quad f(r) = |\{W \in \mathcal{F} : r \in W\}|. \quad (1)$$

A reaction is *abundant* if $2f(r) \geq M$. A reaction is *food-ready* if $\rho(r) \subseteq F$, and an *elementary core* is a nonempty RAF consisting of food-ready reactions [6]. Since a union of RAFs is a RAF, $\mathcal{F}$ is union-closed, and an abundant reaction is exactly a witness for Frankl's conjecture in $\mathcal{F}$.

Definition 2.2 (Complete suppliers, cores and local rank). A reaction p is a *complete supplier* of a reaction r if

$$\rho(r) \setminus F \subseteq \pi(p) \quad \text{and} \quad (C(r) \cap F \neq \emptyset \text{ or } C(r) \cap \pi(p) \neq \emptyset). \quad (2)$$

A nonempty set $U \subseteq R$ is a *complete-supplier core* if for each $r \in U$ there is $d(r) \in U$ that is a complete supplier of r . It is *locally ranked* if there is a map $\lambda : U \rightarrow \mathbb{N}$ such that

$$p, r \in U, \quad \pi(p) \cap (\rho(r) \setminus F) \neq \emptyset \implies \lambda(p) < \lambda(r). \quad (3)$$

A *certificate* for U is a choice of λ and of the supplier map d .

Condition (3) says that the *internal substrate-production digraph* of U , with an edge $p \rightarrow r$ whenever some product of p is a non-food reactant of r , is acyclic; self-loops are excluded in particular. It inspects every internal producer alternative, not only the chosen suppliers. Catalytic incidences do not define its edges, and catalytic cycles inside U are allowed and typical. The supplier condition is an existence condition: other producers and other catalysts of r , inside or outside U , remain allowed. Nothing whatever is required of the reactions outside U .

Proposition 2.3 (Elementary cores are locally ranked supplier cores). *Every elementary core is a locally ranked complete-supplier core, with constant rank.*

Proof. Let U be a nonempty RAF of food-ready reactions. Since every $r \in U$ has $\rho(r) \setminus F = \emptyset$, the reactant clause of (2) holds for any p , and there are no internal substrate-production incidences, so (3) holds for $\lambda \equiv 0$. For the catalyst clause, U is a RAF, so $C(r) \cap H(U) \neq \emptyset$. Every molecule of $H(U)$ is food or a product of some reaction of U , because the closure iteration only adds products of selected reactions. Hence either food catalyses r or some $p \in U$ has $C(r) \cap \pi(p) \neq \emptyset$; in the first case take $d(r) = r$, in the second $d(r) = p$. $\square$

The converse fails: the core of Section 5.1 contains a reaction with a non-food reactant, and the family of Section 5.2 has no elementary RAF.

Remark 2.4 (What the hypothesis excludes). Two features of Definition 2.2 are forced. First, the rank condition cannot be dropped. With $F = \{f\}$ and two reactions $p : y \rightarrow x$ and $q : x \rightarrow y$, both catalysed by f , each is a complete supplier of the other, but no nonempty set is food-generated, so $\mathcal{F} = \{\emptyset\}$ and there is no RAF at all. The supplier condition alone describes support relations that are consistent but never seeded. Second, the single-supplier condition cannot be weakened to independent suppliers of different reactants. The universal producer–gate construction of [6], which realises every union-closed family, has an acyclic substrate digraph, one catalyst per reaction and one food molecule; its gates need marker molecules from several different producers, and the producer that supplies a gate's catalyst supplies none of its markers. No single reaction meets every requirement of a gate, and this is precisely where the counting injection of Section 3.2 loses the common coordinate that prevents collisions. Any sufficient condition satisfied by that construction would settle the unrestricted conjecture.

Theorem 2.5 (Locally ranked supplier-core abundance). *Let U be a locally ranked complete-supplier core of a finite catalytic reaction system. Then U is a RAF. For every $T \subseteq R \setminus U$ let*

$$\mathcal{B}_T = \{S \subseteq U : S \cup T \in \mathcal{F}\} \quad (4)$$

be the exterior fibre over T . Then

$$|U| |\mathcal{B}_T| \leq 2 \sum_{S \in \mathcal{B}_T} |S| \quad \text{for every } T \subseteq R \setminus U, \quad (5)$$

$$|U| M \leq 2 \sum_{W \in \mathcal{F}} |W \cap U|, \quad (6)$$

$$\exists r \in U : M \leq 2f(r). \quad (7)$$

No restriction is placed on the substrate or supplier structure outside U .

Inequality (5) is trivially true when $\mathcal{B}_T$ is empty, and it says that in a nonempty fibre the average core occupancy is at least $|U|/2$. The empty fibre $T = \emptyset$ contains $\emptyset$ and U itself.

3 Proof of the main theorem

3.1 The exact closure adapter

For a reaction r and a set $W \subseteq R$, let $L_r(W)$ denote the conjunction

$$\forall x \in \rho(r) \setminus F \quad \exists p \in W : x \in \pi(p), \quad \text{and} \quad C(r) \cap F \neq \emptyset \quad \text{or} \quad \exists p \in W : C(r) \cap \pi(p) \neq \emptyset. \quad (8)$$

This is the *local support predicate*: every non-food reactant of r has a selected producer, and r has a food catalyst or a selected producer of a catalyst. It is monotone in W . It looks only at products of selected reactions and not at whether those reactions are themselves enabled; that is the gap that the rank hypothesis will close. A complete supplier of r meets every clause of L_r at once:

$$d(r) \in W \implies L_r(W). \quad (9)$$

Fix $T \subseteq R \setminus U$ and define the *exterior predicate* on subsets $S \subseteq U$ by

$$K_T(S) \iff \forall t \in T : \rho(t) \subseteq H(S \cup T) \text{ and } C(t) \cap H(S \cup T) \neq \emptyset. \quad (10)$$

Food closure is monotone under addition of selected reactions, so K_T is upward in S . It is evaluated through the closure of the combined set $S \cup T$; nothing is asserted about T in isolation, and the example of Section 5.1 shows that T need not be food-generated on its own.

Lemma 3.1 (Exact fibre membership). *Let U satisfy the rank condition (3), let $T \subseteq R \setminus U$, and let $S \subseteq U$. Then*

$$S \cup T \in \mathcal{F} \iff K_T(S) \text{ and } \forall r \in S : L_r(S \cup T). \quad (11)$$

Proof. Write $W = S \cup T$. If $W = \emptyset$ both sides hold vacuously. Suppose W is a nonempty RAF. The exterior clauses of $K_T(S)$ are the RAF conditions for the reactions of T . For $r \in S$, every non-food reactant x of r lies in $H(W)$, and every molecule of $H(W)$ is food or a product of a reaction of W , so x has a selected producer; likewise a catalyst of r in $H(W)$ is food or a product of a selected reaction. Hence $L_r(W)$.

Conversely assume the right-hand side and $W \neq \emptyset$. First, every product of a reaction $t \in T$ lies in $H(W)$: the reactants of t lie in $H(W)$ by $K_T(S)$, there are finitely many of them, so they all lie in some stage $H_j(W)$, and then $\pi(t) \subseteq H_{j+1}(W)$. Next we show by strong induction

on $\lambda(r)$ that every $r \in S$ has $\rho(r) \subseteq H(W)$. Let $x \in \rho(r) \setminus F$. By $L_r(W)$ there is $p \in W$ with $x \in \pi(p)$. If $p \in T$ then $x \in H(W)$ by the first step. If $p \in S$ then $x \in \pi(p) \cap (\rho(r) \setminus F)$, so (3) gives $\lambda(p) < \lambda(r)$; by induction $\rho(p) \subseteq H(W)$, so p is enabled at some finite stage and $x \in \pi(p) \subseteq H(W)$. Food reactants are in $H_0(W)$. Thus W is food-generated, and every product of every reaction of W lies in $H(W)$. Now the catalyst clause of $L_r(W)$ for $r \in S$ gives a catalyst of r that is food or a product of a selected reaction, hence lies in $H(W)$; and the catalyst clauses for $t \in T$ are part of $K_T(S)$. So W is a RAF.

The induction can be made explicit in closure stages: let J be a stage by which the reactants of every $t \in T$ are available, so that $\pi(t) \subseteq H_{J+1}(W)$; then the reactants of a selected core reaction of rank n lie in $H_{J+n+1}(W)$. This is the form used in the formal proof. $\square$

The rank is used only inside S : exterior producers are handled by actual closure, whatever their own structure, and the induction descends through core producers only. This is why no acyclicity is needed outside U .

3.2 The common-dependency counting argument

The counting statement below is the aggregate form of Theorem 4 of Lozin and Zamaraev [15], who proved Frankl's conjecture for the families defined by Horn functions with their dependency property. In the present formulation the members of the family are the sets S satisfying an upward constraint K together with, for every selected element r , a monotone support requirement L_r ; the dependency property is that each L_r is met whenever a designated companion $d(r)$ is selected. An upward constraint corresponds to terms without a negative literal, which satisfy the dependency property vacuously, and grouping all Horn terms with the same head into one monotone predicate changes nothing in the injection. We include the proof because the paper is meant to be self-contained and because the formalisation follows it literally. The companion paper [6] uses the same lemma.

Lemma 3.2 (Conditioned occupancy). *Let U be a finite set, let K and L_r ($r \in U$) be upward predicates on subsets of U , and let $d : U \rightarrow U$ satisfy $d(r) \in S \Rightarrow L_r(S)$ for all r and S . Put*

$$\mathcal{B} = \{S \subseteq U : K(S) \text{ and } \forall r \in S, L_r(S)\}.$$

Then $|U| |\mathcal{B}| \leq 2 \sum_{S \in \mathcal{B}} |S|$.

Proof. Work with deletion sets $D = U \setminus S$ and define

$$P(D) = \neg K(U \setminus D), \quad B_r(D) = \neg L_r(U \setminus D).$$

Both are upward in D , and $B_r(D)$ implies $d(r) \in D$ (otherwise $d(r) \in U \setminus D$ would force L_r). Call D *bad* if $P(D)$ holds or some $r \notin D$ has $B_r(D)$; the bad sets are exactly the complements of the sets outside $\mathcal{B}$. A *cell* is a pair (D, a) with D bad and $a \in U$; it is a *zero cell* if $a \notin D$ and a *one cell* if $a \in D$. We construct an injection from zero cells to one cells.

Let (D, a) be a zero cell. If $P(D)$ holds, or some $q \neq a$ with $q \notin D$ has $B_q(D)$, map it *horizontally* to $(D \cup \{a\}, a)$; by upwardness P or B_q persists, and $q \notin D \cup \{a\}$, so the image row is bad, and it is a one cell. Otherwise $P(D)$ fails and a is the unique element outside D whose requirement fails, so $B_a(D)$ holds; map *vertically* to $(D, d(a))$, which is a one cell because $B_a(D) \Rightarrow d(a) \in D$.

Two horizontal images coincide only if their marks coincide and their rows, obtained by removing the mark, coincide. Two vertical images coincide only if their rows coincide, and then their marks are the unique failing elements of that row, hence coincide. Suppose finally that a horizontal image $(D \cup \{a\}, a)$ equals a vertical image $(D', d(b))$ of a zero cell (D', b) . Then $D' = D \cup \{a\}$ and $a = d(b)$, and by verticality $P(D')$ fails while b is the unique failing

element outside D' . Since (D, a) went horizontally, either $P(D)$ holds, which would persist to D' , or some $q \neq a$ with $q \notin D$ has $B_q(D)$; then $q \notin D'$ and $B_q(D')$, so $q = b$. But $B_b(D)$ forces $d(b) = a \in D$, contradicting $a \notin D$. Hence the map is injective.

Over the full Boolean cube 2^U there are as many zero cells as one cells, by complementation. The bad rows contribute at most as many zero cells as one cells, so the remaining rows, the complements of the members of $\mathcal{B}$, contribute at least as many zero cells as one cells: $\sum_{S \in \mathcal{B}} |S| \geq \sum_{S \in \mathcal{B}} (|U| - |S|)$, which is the claim. It holds trivially if $\mathcal{B}$ is empty. $\square$

3.3 Exact aggregation

Proof of Theorem 2.5. Fix $T \subseteq R \setminus U$. Apply Lemma 3.2 to the ground set U , the upward exterior predicate K_T of (10), and the internal predicates $S \mapsto L_r(S \cup T)$, which are upward in S and satisfy the dependency condition by (9). Lemma 3.1 identifies the family $\mathcal{B}$ of the counting lemma with the fibre $\mathcal{B}_T$ of (4), which proves (5).

For $T = \emptyset$ and $S = U$ the exterior predicate is vacuous and every internal predicate holds by (9), so Lemma 3.1 and $U \neq \emptyset$ show that U is a RAF.

Every $W \in \mathcal{F}$ decomposes uniquely as $W = (W \cap U) \cup (W \setminus U)$ with $W \cap U \in \mathcal{B}_{W \setminus U}$, so $\mathcal{F}$ is the disjoint union over T of the sets $\{S \cup T : S \in \mathcal{B}_T\}$, with $|\{S \cup T : S \in \mathcal{B}_T\}| = |\mathcal{B}_T|$. Summing (5) over T gives (6) with no change of measure. Finally $\sum_{W \in \mathcal{F}} |W \cap U| = \sum_{r \in U} f(r)$; if $2f(r) < M$ for every $r \in U$, summing the strict inequalities contradicts (6). This proves (7). $\square$

Remark 3.3. The proof never assumes that the exterior selection T is a RAF, that it is food-generated on its own, or that U is the largest RAF. It retains each original reaction subset and each exterior completion multiplicity: fibres with many completions weigh more in (6), exactly as they do in $f(r)$. Replacing the exterior by a separately food-generated subsystem, a feasible trace with uniform weight, or a list of irreducible RAFs would change the measure and break the identity $\sum_W |W \cap U| = \sum_{r \in U} f(r)$.

4 Consequences

4.1 Transport across exterior edits and distinct witnesses

Corollary 4.1 (Transport and disjoint cores). *Let U be a locally ranked complete-supplier core of $\mathcal{Q}$ with certificate (λ, d) .*

- (a) *If $\mathcal{Q}'$ is obtained from $\mathcal{Q}$ by adding, deleting or editing reactions outside U , by enlarging the food set, and by adding catalytic incidences, while the reactant and product sets of the reactions of U are unchanged, then (λ, d) certifies U in $\mathcal{Q}'$; hence U is a RAF of $\mathcal{Q}'$ and contains a reaction abundant in $\mathcal{F}(\mathcal{Q}')$.*
- (b) *If $U_1, \dots, U_k$ are pairwise disjoint locally ranked complete-supplier cores of $\mathcal{Q}$, then $\mathcal{Q}$ has k distinct abundant reactions $r_i \in U_i$.*

Proof. (a) Conditions (2) and (3) mention only F , the incidences of reactions in U and the catalysis relation. Enlarging F shrinks each $\rho(r) \setminus F$, which can only remove reactant requirements and internal substrate incidences; enlarging the catalysis relation can only add catalyst options; exterior edits change nothing that is inspected. So the certificate remains valid and Theorem 2.5 applies to $\mathcal{Q}'$. (b) Choose an abundant $r_i \in U_i$ by Theorem 2.5; disjointness makes the choices distinct. No independence between the cores is needed. $\square$

The witness in (a) may change with the surrounding system: the companion paper gives an elementary core $\{a, b\}$ whose abundant reaction is a in one exterior and b in another [6, Example 5.2]. No single reaction is asserted to work for all embeddings.

4.2 Exterior weighting and bounded distortion

Corollary 4.2 (Exterior weights). *Let $h : 2^{R \setminus U} \rightarrow [0, \infty)$ and suppose $Z = \sum_{W \in \mathcal{F}} h(W \setminus U) > 0$. Put $\mu(W) = h(W \setminus U)/Z$. Then $\mathbb{E}_\mu |W \cap U| \geq |U|/2$, and some $r \in U$ satisfies $\mu(r \in W) \geq 1/2$.*

Proof. Multiply (5) by $h(T)$, sum over T , reindex by $W = S \cup T$ and divide by Z . The identity $|W \cap U| = \sum_{r \in U} \mathbf{1}_{\{r \in W\}}$ gives the witness by averaging over U . $\square$

Indicator weights condition on any collection of exterior contexts, including a single fibre. Within a fibre all compatible core subsets receive equal weight. The next statement relaxes this.

Theorem 4.3 (Bounded within-fibre distortion). *Let $\kappa \geq 1$ and let $w : \mathcal{F} \rightarrow [0, \infty)$ have positive total mass Z . Suppose that for every $T \subseteq R \setminus U$ there is $b_T \geq 0$ with*

$$b_T \leq w(S \cup T) \leq \kappa b_T \quad \text{for all } S \in \mathcal{B}_T. \quad (12)$$

Then for $\mu = w/Z$,

$$\mathbb{E}_\mu |W \cap U| \geq \frac{|U|}{1 + \kappa}, \quad \exists r \in U : \mu(r \in W) \geq \frac{1}{1 + \kappa}. \quad (13)$$

The constant $1/(1 + \kappa)$ is sharp for every $\kappa \geq 1$. In particular (13) applies to weights $w(S \cup T) = h(T) a_T(S)$ whenever the positive within-fibre factors a_T have maximum-to-minimum ratio at most κ .

Proof. Fix T with $b_T > 0$; fibres with $b_T = 0$ carry no mass. Let $n = |U|$, write $X(S) = |S|/n \in [0, 1]$, and let $\mathbb{E}_T$ denote uniform expectation over $\mathcal{B}_T$, so that $\mathbb{E}_T X \geq 1/2$ by (5). Put $a(S) = w(S \cup T)/b_T \in [1, \kappa]$, $A = \mathbb{E}_T[aX]$ and $B = \mathbb{E}_T[a(1 - X)]$. Then

$$A \geq \mathbb{E}_T X, \quad B \leq \kappa \mathbb{E}_T[1 - X] = \kappa(1 - \mathbb{E}_T X) \leq \kappa \mathbb{E}_T X \leq \kappa A,$$

so $A/(A + B) \geq 1/(1 + \kappa)$. Since $A/(A + B) = \sum_{S \in \mathcal{B}_T} w(S \cup T)|S| / (n \sum_{S \in \mathcal{B}_T} w(S \cup T))$, this says $\sum_{S \in \mathcal{B}_T} w(S \cup T)|S| \geq \frac{n}{1 + \kappa} \sum_{S \in \mathcal{B}_T} w(S \cup T)$. Summing over T and dividing by Z gives the first inequality of (13); the coordinate identity gives the witness. For sharpness take $\mathcal{F} = \{\emptyset, U\}$ (for instance the all-or-none module of Example 5.4 with $k = 1$), which is a single fibre, with $w(\emptyset) = \kappa$ and $w(U) = 1$: then $\mathbb{E}_\mu |W \cap U| = |U|/(1 + \kappa)$ and every $r \in U$ has $\mu(r \in W) = 1/(1 + \kappa)$. $\square$

These are structural sampling statements about the set family $\mathcal{F}$. A kinetic interpretation would require an independent derivation of the sampling weights and of the distortion bound, and each chosen weighting has its own witness.

4.3 Reaction deletion

Deleting a reaction r from $\mathcal{Q}$ means removing it from R while retaining the food set and the incidence data of all other reactions.

Proposition 4.4 (Deletion bounds). *Let U be a locally ranked complete-supplier core and let N be the number of nonempty RAFs of $\mathcal{Q}$. For any reaction r , the nonempty RAFs of $\mathcal{Q}$ with r deleted are exactly the nonempty RAFs of $\mathcal{Q}$ not containing r , so their number is $N_{\text{after}}(r) = N - f(r)$. Consequently there is $r \in U$ with*

$$N_{\text{after}}(r) \leq \left\lfloor \frac{N - 1}{2} \right\rfloor, \quad (14)$$

and if r is chosen uniformly at random from U then

$$\mathbb{E} N_{\text{after}}(r) = N - \frac{1}{|U|} \sum_{r \in U} f(r) \leq \frac{N - 1}{2}. \quad (15)$$

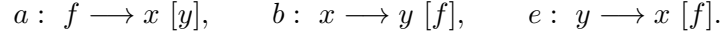
Proof. For $W \subseteq R \setminus \{r\}$, the closure iteration of Definition 2.1 uses only the reactions of W and their incidence data, which are the same before and after the deletion; so $H(W)$, and hence the RAF status of W , is unchanged. Thus the surviving RAFs are the original RAFs omitting r , and $N_{\text{after}}(r) = N - f(r)$. By (7) some $r \in U$ has $2f(r) \geq M = N + 1$, so $f(r) \geq \lceil (N + 1)/2 \rceil$ and $N - f(r) \leq \lfloor (N - 1)/2 \rfloor$. Averaging the identity over U and applying (6) gives (15). $\square$

The deterministic bound identifies a set containing a successful deletion; the expectation bound requires no individual frequencies. Both concern the *number of RAF organisations*, not viability, a specified product, flux or persistence. Repeated halving requires a fresh certificate after each deletion, since deleting a core reaction destroys the core. Section 5.4 shows that individual core deletions can perform much worse than the expectation.

5 Examples and sharpness

5.1 A strict extension beyond global substrate rank

Example 5.1. Let $F = \{f\}$ and take three reactions, with catalysts in brackets:



The set $U = \{a, b\}$ is a locally ranked complete-supplier core: $d(a) = b$ supplies the catalyst y (and a has no non-food reactant), $d(b) = a$ supplies x (and food catalyses b), and the only internal substrate incidence is $a \rightarrow b$, so $\lambda(a) = 0$, $\lambda(b) = 1$. The core is not elementary, since b consumes x . The ambient substrate digraph contains the cycle $b \rightarrow e \rightarrow b$, so no rank on all of R exists. The family is

$$\mathcal{F} = \{\emptyset, \{a, b\}, \{a, b, e\}\}, \quad M = 3, \quad f(a) = f(b) = 2, \quad f(e) = 1. \quad (16)$$

Indeed, without a no selected reaction can produce a non-food molecule, so $\{e\}$, $\{b\}$ and $\{b, e\}$ are not food-generated; selecting a requires its catalyst y , produced only by b ; and once a, b are selected, e is optional. Both core reactions are abundant.

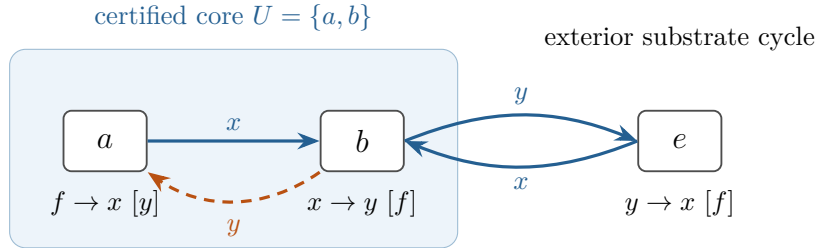


Figure 1: Example 5.1. Solid blue arrows are substrate-production incidences; the dashed orange arrow is non-food catalytic support. Food catalyses b and e . The cycle $b \rightarrow e \rightarrow b$ in the ambient substrate digraph does not affect the internal rank of U .

The example also illustrates why the exterior predicate (10) must be evaluated through the combined closure: the exterior selection $T = \{e\}$ is not food-generated on its own, yet $\{a, b\} \cup \{e\} \in \mathcal{F}$. The fibre over $T = \{e\}$ is $\{\{a, b\}\}$ and the fibre over $T = \emptyset$ is $\{\emptyset, \{a, b\}\}$; both satisfy (5). An earlier form of Theorem 2.5, which required a rank on the substrate digraph of the whole system, does not apply here, and the formal development records both the certificate of U and the nonexistence of a global rank (Appendix A).

5.2 A uniform three-chain family

Fix an integer $\ell \geq 3$ and let $F = \{f\}$. There are three chains $a_1, \dots, a_\ell$, $b_1, \dots, b_\ell$, $c_1, \dots, c_\ell$ and three gates g_A, g_B, g_C . Every chain reaction consumes only f . For $j < \ell$ the reaction a_j produces a private molecule y_j^A ; the last chain reactions and the gates are

Reaction	Transformation	Catalyst
a_ℓ	$f \rightarrow x_A + u$	$y_{\ell-1}^A$
b_ℓ	$f \rightarrow u + v$	$y_{\ell-1}^B$
c_ℓ	$f \rightarrow v + x_C$	$y_{\ell-1}^C$
g_A	$x_A \rightarrow z_A$	f
g_B	$u + v \rightarrow z_B$	f
g_C	$x_C \rightarrow z_C$	f

The first chain reaction a_1 is catalysed only by the gate product z_A , and each a_j with $j \geq 2$ is catalysed only by y_{j-1}^A ; the B and C chains are analogous. The system has $3\ell + 3$ reactions and $3\ell + 5$ molecules, one food molecule, one catalyst per reaction, at most two reactants and two products per reaction, and its food closure stabilises after two rounds for every selected set.

Let A denote the A chain together with g_A , and define B and C similarly. Each is a locally ranked complete-supplier core: $d(a_1) = g_A$, $d(a_j) = a_{j-1}$ for $j \geq 2$, and $d(g_A) = a_\ell$, which supplies x_A while food catalyses the gate; the same for C ; and for B , $d(g_B) = b_\ell$ supplies both u and v . The only internal substrate incidences are $a_\ell \rightarrow g_A$, $b_\ell \rightarrow g_B$, $c_\ell \rightarrow g_C$, so rank 0 on chain reactions and 1 on gates is valid, here even globally.

Proposition 5.2 (Exact family). *$\mathcal{F}$ consists exactly of the eight unions of subcollections of $\{A, B, C\}$ together with the ℓ sets*

$$A \cup C \cup \{g_B\} \cup \{b_1, \dots, b_j\}, \quad 0 \leq j < \ell. \quad (17)$$

Consequently

$$M_\ell = \ell + 8, \quad (18)$$

$$f(a_j) = f(c_j) = f(g_A) = f(g_B) = f(g_C) = \ell + 4, \quad (19)$$

$$f(b_j) = \ell + 4 - j \quad (1 \leq j \leq \ell). \quad (20)$$

Proof. Let W be a RAF. Within a chain, the selected reactions form a prefix, because each chain reaction after the first is catalysed only by the product of its predecessor, and a nonempty prefix requires the chain's gate, because the first chain reaction is catalysed only by the gate product. The gate g_A requires x_A , produced only by a_ℓ , so $g_A \in W$ forces all of A ; likewise for C . The gate g_B requires u and v : either $b_\ell \in W$, which forces all of B , or $a_\ell, c_\ell \in W$, which forces all of A and C , and then any proper prefix of the B chain may accompany g_B , including the empty prefix. Conversely every set of the stated form is a RAF: all selected chain reactions are food-ready, every selected gate has its substrates from a selected final chain reaction, every gate is food-catalysed, and every selected chain reaction has its catalyst produced by a selected reaction, all within the closure. The ℓ sets (17) are distinct from each other and from the eight whole-core unions, since they contain g_B but not all of B ; the union with $j = \ell$ is $A \cup B \cup C$.

Every reaction of A , of C , and every gate lies in four whole-core unions and in all ℓ sets (17). The reaction b_j lies in the four whole-core unions containing B and in the $\ell - j$ sets (17) with prefix length at least j . $\square$

Corollary 5.3 (Sharp average, rare members). *In the three-chain family the core B has average occupancy exactly one half, while its rarest reaction has occupancy tending to zero and*

its gate has occupancy tending to one:

$$\frac{\sum_{r \in B} f(r)}{|B| M_\ell} = \frac{1}{2}, \quad \frac{f(b_\ell)}{M_\ell} = \frac{4}{\ell + 8} \rightarrow 0, \quad \frac{f(g_B)}{M_\ell} = \frac{\ell + 4}{\ell + 8} \rightarrow 1.$$

Exactly the reactions b_j with $2j \leq \ell$ are abundant among the B -chain reactions.

Proof. With $|B| = \ell + 1$ and (20), $\sum_{r \in B} f(r) = (\ell + 4) + \sum_{j=1}^{\ell} (\ell + 4 - j) = \frac{(\ell+1)(\ell+8)}{2}$. The limits are immediate, and $2(\ell + 4 - j) \geq \ell + 8$ is equivalent to $2j \leq \ell$. $\square$

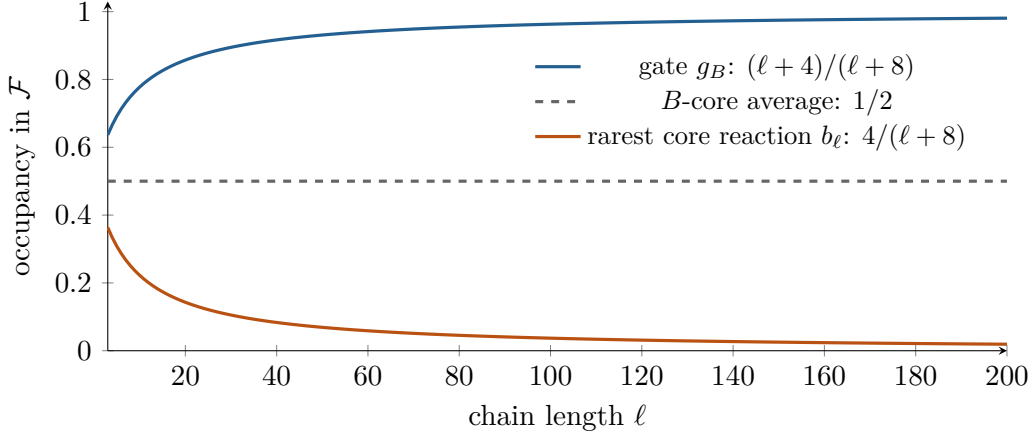


Figure 2: Occupancies in the three-chain family, from (18)–(20). The average over the certified core B equals one half for every ℓ , while individual core reactions range from almost always present to almost never present. The sharpness concerns the average over a designated core, not the maximum frequency in the whole system: the reactions of A and C have occupancy $(\ell + 4)/(\ell + 8)$.

The family has several structural properties that place it outside earlier positive results.

- *No elementary RAF.* Without gates no first chain reaction has a catalyst, so no set of food-ready reactions is a RAF. The elementary-core theorem of [6] does not apply; Theorem 2.5 applies to each of A , B and C , and Corollary 4.1(b) gives three distinct abundant reactions.
- *Transversal number three.* Every nonempty member of $\mathcal{F}$ contains a whole core, so the minimum RAF size is $\ell + 1$, although only the three gates are not food-ready. The three disjoint cores show that no two reactions meet every nonempty RAF; the three gates do. The two-element transversal argument of [6, Lemma 5.6] therefore does not apply.
- *Irreducible RAFs do not determine $\mathcal{F}$.* The only irreducible RAFs are A , B and C ; their unions omit the ℓ sets (17), which are what make the B -chain frequencies nonuniform. A count restricted to unions of irreducible RAFs would report every reaction with frequency 4 out of 8.
- *Not a digraph support family.* Call a family of subsets of a vertex set the support family of a digraph if a set belongs to it exactly when every member has an in-neighbour in the set (loops allowed). $\mathcal{F}$ is not the support family of any digraph on the $3\ell + 3$ reactions: A and C are members, $A \cup \{g_B\}$ and $C \cup \{g_B\}$ are not, but $A \cup C \cup \{g_B\}$ is; a digraph in-neighbour of g_B inside the last set lies in A , in C , or is g_B itself, and each possibility makes one of the two excluded sets supported. The intrinsic characterisation of digraph support families in [5] thus does not cover RAF families with joint reactant requirements.

5.3 Sharpness of the witness fraction

Example 5.4 (All-or-none modules). For $k \geq 1$ take k modules with private molecules x_i, y_i and reactions $p_i : f \rightarrow x_i [y_i]$, $q_i : x_i \rightarrow y_i [f]$. Each $\{p_i, q_i\}$ is a locally ranked complete-supplier core ($d(p_i) = q_i$, $d(q_i) = p_i$, $\lambda(p_i) = 0$, $\lambda(q_i) = 1$), and a set is a RAF exactly when it is a nonempty union of whole modules. Hence $M = 2^k$ and $f(r) = 2^{k-1}$ for every reaction: the empty-inclusive bound $2f(r) \geq M$ holds with equality for every r , and the nonempty-only fraction $2^{k-1}/(2^k - 1)$ tends to $1/2$ from above. With $k = 1$ this is the two-member family used for the sharpness of Theorem 4.3.

The constant $1/2$ in Theorem 2.5 therefore cannot be improved, either for the core average (Corollary 5.3) or for the best individual witness (Example 5.4), and the two kinds of sharpness are realised by different mechanisms.

5.4 The twelve-reaction instance and a conservation lift

For $\ell = 3$ the family has $M = 11$, $N = 10$. Figure 3 shows the supplier structure and Table 1 the full frequency table. The proposition proves these values uniformly in ℓ ; an independent exhaustive food-closure computation over all 2^{12} subsets confirms them for this instance (Appendix A).

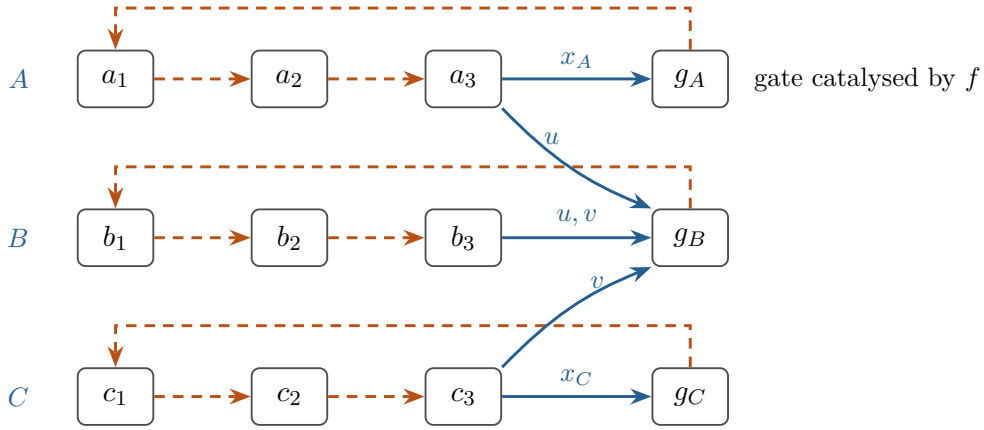


Figure 3: The three-chain family at $\ell = 3$. Dashed orange arrows are catalytic support along each chain, with the gate product catalysing the first chain reaction. Solid blue arrows supply gate substrates. The cross incidences $a_3 \rightarrow g_B$ and $c_3 \rightarrow g_B$ jointly permit the three proper-prefix members (17). Every chain reaction consumes only food.

Table 1: Frequencies in the twelve-reaction instance ($\ell = 3$, $M = 11$). The abundance threshold is $M/2 = 5.5$; b_2 and b_3 are rare although they belong to a certified core.

Core	first chain reaction	second	third	gate
A	$f(a_1) = 7$	$f(a_2) = 7$	$f(a_3) = 7$	$f(g_A) = 7$
B	$f(b_1) = 6$	$f(b_2) = 5$	$f(b_3) = 4$	$f(g_B) = 7$
C	$f(c_1) = 7$	$f(c_2) = 7$	$f(c_3) = 7$	$f(g_C) = 7$

Deleting one reaction of B gives the following exact panel, by Proposition 4.4:

Deleted reaction	RAFTs destroyed	nonempty RAFTs remaining
g_B	7	3
b_1	6	4
b_2	5	5
b_3	4	6

The expected number remaining under uniform choice in B is $4.5 = (N - 1)/2$, attaining (15); the individual choices b_2 and b_3 destroy fewer than half of the RAFTs, while g_B and b_1 satisfy (14).

The construction admits a conservation bookkeeping without any change to $\mathcal{F}$. Assign two units to f , one unit to every private chain product and to x_A, u, v, x_C, z_A, z_C , and two units to z_B . Introduce an inert one-unit molecule ω and replace every chain reaction $f \rightarrow p$ with a single product by $f \rightarrow p + \omega$. Then every reaction conserves units: the final chain reactions produce two one-unit molecules from f , and the gates $x_A \rightarrow z_A, u + v \rightarrow z_B, x_C \rightarrow z_C$ balance. Catalysts appear on both sides of a stoichiometric equation and cancel.

Proposition 5.5 (Inert-output preservation). *Let $\mathcal{Q}'$ be obtained from $\mathcal{Q}$ by adding, to the product sets of some reactions, fresh molecules that are reactants or catalysts of no reaction. Then $\mathcal{Q}'$ and $\mathcal{Q}$ have the same RAFT family on the same reaction identities.*

Proof. Let X be the original molecule set. We show by induction on the stage that $H'_j(W) \cap X = H_j(W)$ for every W and j , where H' is the closure in $\mathcal{Q}'$. The base case is F . A reaction is enabled at stage j in $\mathcal{Q}'$ if and only if its reactants, which lie in X , are in $H'_j(W) \cap X = H_j(W)$, so exactly the same reactions fire, and each adds its original products together with inert molecules outside X . Hence $H'(W) \cap X = H(W)$. Food generation tests reactants, and catalysis tests catalysts, both subsets of X ; so every W has the same status in both systems. $\square$

This gives a balanced schematic construction, not a chemical model: it establishes no empirical catalysts, thermodynamic feasibility, flux or persistence, and a biochemical reconstruction would have to supply those ingredients separately.

6 Certificates and computation

A certificate for Theorem 2.5 consists of a nonempty candidate U , a rank λ on U and a supplier $d(r) \in U$ for each $r \in U$. Validation checks (2) for each pair $(d(r), r)$ and (3) for every internal incidence, including self-incidences. Nothing outside U is read, no RAFT is enumerated, and no closure is computed: the theorem supplies RAFT status and abundance. In particular the validation cost is independent of the size of the exterior.

With $k = |U|$ and molecules encoded as bitsets of m coordinates and word size w , constructing the internal substrate digraph costs $O(k^2 \lceil m/w \rceil)$ word operations and a topological sort costs $O(k + e)$ for e internal edges, which also produces a rank if one exists. Checking the supplied suppliers costs $O(k \lceil m/w \rceil)$; searching for suppliers among all internal pairs, when none are supplied, has the same quadratic bound as the graph construction. Input parsing is additional. With set-based representations the same bounds hold in units of set intersection and subset tests.

Three capabilities must be kept apart. *Validation* checks a supplied candidate, as above. *Discovery* proposes candidates; a directed cycle in the complete-supplier digraph, with edges $p \Rightarrow r$ whenever p is a complete supplier of r , is a natural candidate, but any nonempty internally ranked, internally supplied set is admissible, and the theorem does not claim polynomial-time complete discovery of all certifiable cores. *Witness identification* chooses an individually abundant reaction inside U ; the certificate does not compute frequencies, and Table 1 shows that a certified core can contain rare reactions. A failed candidate means only

that this certificate failed, not that no RAF or no abundant reaction exists. The accompanying implementation [7] validates a supplied core locally, accepts Example 5.1 although the global-rank test rejects it, and rejects a missing supplier and an internal self-loop; individual frequencies in its worked output come from a separately labelled exhaustive check.

7 Scope and open problems

The results are statements about finite set families. They assert nothing about concentrations, fluxes, persistence under dynamics, thermodynamic feasibility or the chemical realisability of the molecules in the examples, and a reaction that lies in half of all RAFs is not thereby essential to any of them. The weighted results allow arbitrary nonnegative weights on exterior contexts, and weights on individual members with bounded within-fibre distortion; arbitrary weights on individual RAFs are outside the theorem.

The sufficient condition proved here is the existence of a locally ranked complete-supplier core. It contains the elementary cores of [6], whose universal construction shows that no condition satisfied by the producer–gate source can be sufficient short of the full conjecture. Between the two lie the following questions.

Question 7.1. The rank condition (3) is needed in some form (Remark 2.4), but it excludes cores in which a non-food reactant is produced by its own consumer. Is there a seeded variant, requiring for instance that every internal substrate cycle contain a reaction whose reactants are available from outside the cycle, under which Lemma 3.1 survives?

Question 7.2. The single-supplier condition asks one reaction to meet all requirements of r . The universal construction shows that fully independent suppliers cannot be allowed. Is there an intermediate condition, for instance two suppliers with a common coordinate in every failing configuration, under which an injection with the collision structure of Lemma 3.2 still exists?

Question 7.3. The three-chain family and its relatives have exactly three non-food-ready reactions, each a gate of a core. Does every system with a RAF and exactly three non-food-ready reactions have an abundant reaction? The transversal argument of [6] stops at two, and the family shows that the three gates can all be non-rare while the cores contain rare reactions.

Question 7.4. Which certified cores admit a witness that is stable across a specified family of exterior contexts, or an individually abundant reaction identifiable from the certificate alone? Corollary 5.3 shows that the gate of a chain core is asymptotically the right choice in the three-chain family; whether a general rule exists is open.

The strongest positive statement known for general RAF families remains the constant of the entropy method [1, 3, 4, 10, 20], a reaction in a 0.38 fraction of all RAFs of any system with a RAF. Whether the RAF structure can push that constant to 1/2 is, by the equivalence theorem of [6], exactly the union-closed sets conjecture.

A Formal verification

A.1 Environment and receipts

The formal development is written in Lean 4 (toolchain `leanprover/lean4:v4.30.0`) against Mathlib at commit `c5ea0035` (the `v4.30.0` tag), in the directory `proofs/RAF/Frank1/` of the companion repository [7]. The public entry module `LocalSupplierPaper.lean` imports the four result modules. It was compiled with `-DwarningAsError=true`; the compiler produced no output. The verification receipt records the SHA-256 of the root source, of the 26 local dependency sources and of the Lake manifest, together with an elaboration probe of each

exported declaration; every exported theorem depends only on `Classical.choice`, `Quot.sound` and `propext`, with no `sorry` and no `native_decide`. The receipt used here is:

Root module	Declarations	Root SHA-256 (prefix)	Elaboration time
RAF/Frankl/LocalSupplierPaper.lean	7	5216507abc656138	154.4 s

The earlier global-rank form of the theorem, exported as `abundance_beyond_elementary` from the module `SupplierCycleCertificate`, is retained unchanged with its own receipt; Example 5.1 separates the two.

The RAF semantics is defined literally as in Definition 2.1: a structure `CRS M R` with `food`, `inputs` and `outputs`, a relation `Catalysis M R`, the stage closure `closureAt Q S k`, the predicates `FoodGenerated`, `CatalyzedFromClosure` and `IsRAF`, and the finite families `rafFamily` (nonempty RAFs) and `fixedFamily` (the family $\mathcal{F}$). The hypotheses of Definition 2.2 are `CompleteSupplier`, `SupplierCore` and `InternalSubstrateRank`; the exterior fibre (4) is `extensionFibre`, and the local predicate (8) is `LocalSupply`.

A.2 Correspondence

Article statement	Lean declaration (namespace <code>RAF.Frankl</code>) and module
Lemma 3.1, rank induction	<code>internal_rank_foodGenerated (LocalRankedClosure)</code>
Lemma 3.1, equivalence (11)	<code>local_rank_fibre_iff</code> ; monotonicity of (10) is <code>closureExterior_mono</code>
Lemma 3.2	<code>CommonSupplier.supported_upward_average (CommonSupplierConditioning)</code> ; the injection is <code>conditionedCellMap</code>
Theorem 2.5, all four conclusions	<code>locally_ranked_supplier_abundance (LocallyRankedSupplierCore)</code> ; components <code>local_supplierCore_isRAF</code> , <code>localSupplierFibre_average</code> , <code>local_supplier_core_global_average</code> , <code>local_supplier_core_exists_abundant</code>
Corollary 4.1(a), food enlargement and catalysis enlargement with fixed core incidences	<code>local_certificate_transport (LocalSupplierEmbedding)</code>
Corollary 4.1(b)	<code>disjoint_local_cores_witnesses</code>
Corollary 4.2 (unnormalised)	<code>local_supplier_exterior_weights (LocalSupplierWeights)</code>
Theorem 4.3, inequalities (13) (unnormalised)	<code>local_supplier_bounded_weights</code> ; the fibre aggregation is <code>weighted_fibre_aggregation (SupplierWeightedCounting)</code>
Proposition 4.4, bound (14) as a count of original RAFs omitting r	<code>local_supplier_deletion (SupplierDeletion)</code> ; the partition $N = N_{\text{after}} + f(r)$ is <code>deletion_partition</code>
Example 5.1, family (16)	<code>LocalSupplierExample.exact_fixed_family (LocalSupplierExamples)</code> ; also certified, <code>no_global_rank</code> , <code>exterior_not_generated</code>

A.3 What is and is not formalised

The formal statements are literal. `locally_ranked_supplier_abundance` takes an arbitrary `CRS`, an arbitrary catalysis relation, a finite set U , a rank function with `InternalSubstrateRank`, and `SupplierCore`, and returns the conjunction of `IsRAF Q C U`, the fibre inequality (5) for every T , the global inequality (6) and the witness (7), all for the full system; no projection, feedback or average hypothesis appears. The weighted exports are the unnormalised inequalities

$|U| \sum_W w(W) \leq (1 + \kappa) \sum_W w(W) |W \cap U|$ and their witnesses, over real weights bounded fibrewise by $b_T \leq w \leq \kappa b_T$; division by the positive total mass, which gives the probability statements of Section 4, is ordinary arithmetic. The deletion export counts original RAFs omitting r ; the identification of these with the RAFs of the reduced system, and the expectation bound, are proved in Proposition 4.4. The transport export covers enlargement of food and of the catalysis relation on a fixed reaction type with unchanged core incidences; addition and deletion of exterior reactions, which change the reaction type, are covered by the ordinary argument of Corollary 4.1. Proposition 2.3, Remark 2.4, Proposition 5.2, Corollary 5.3, Example 5.4, the structural properties of the three-chain family and Proposition 5.5 have ordinary proofs above and are not claimed as separately kernel-verified parametric constructions. The finite data of Section 5.4 (the family for $\ell = 3$, the frequencies, the deletion panel, the unit balance of the lifted system and the identity of its family) are replayed by an independent exhaustive script in the paper’s source folder, which also checks Proposition 5.2 for $3 \leq \ell \leq 7$ and Example 5.4 for $k \leq 5$; that script is a diagnostic, not a proof substitute.

A.4 Reproduction

From the repository root, with the pinned toolchain, the strict verification of the entry module is

```
python scripts/verify_proof.py proofs/RAF/Frankl/LocalSupplierPaper.lean
-declaration RAF.Frankl.locally_ranked_supplier_abundance
-declaration RAF.Frankl.local_supplier_bounded_weights
-declaration RAF.Frankl.local_supplier_exterior_weights
-declaration RAF.Frankl.LocalSupplierExample.exact_fixed_family
-declaration RAF.Frankl.local_supplier_deletion
-declaration RAF.Frankl.disjoint_local_cores_witnesses
-declaration RAF.Frankl.local_certificate_transport
-timeout 360 -output LocalSupplierPaper.verify.json
```

The bridge fixes the frozen Mathlib working directory, treats warnings as errors and records the source hashes and the axioms of each declaration.

Acknowledgements

To be inserted.

References

- [1] Ryan Alweiss, Brice Huang, and Mark Sellke. Improved lower bound for Frankl’s union-closed sets conjecture, 2022. arXiv:2211.11731.
- [2] Henning Bruhn and Oliver Schaudt. The journey of the union-closed sets conjecture. *Graphs and Combinatorics*, 31(6):2043–2074, 2015. doi: 10.1007/s00373-014-1515-0.
- [3] Stijn Cambie. Better bounds for the union-closed sets conjecture using the entropy approach, 2022. arXiv:2212.12500.
- [4] Zachary Chase and Shachar Lovett. Approximate union closed conjecture, 2022. arXiv:2211.11689.
- [5] Anonymous. An intrinsic interval-deletion characterization of digraph support families. Manuscript, September 2026.

- [6] Anonymous. Elementary cores, abundant reactions, and the union-closed boundary of autocatalytic reaction systems. Manuscript, 13 September 2026.
- [7] Anonymous. Companion RAF proof and experiment repository. <https://github.com/mmislán/Erdos>, 2026. Lean sources under `proofs/RAF/Frankl/`; verification receipts, fixtures and checks under `problem_workspaces/RAF_abundance_beyond_elementary_core_coverage/`. Local snapshot of 14 September 2026.
- [8] Leonardo de Moura and Sebastian Ullrich. The Lean 4 theorem prover and programming language. In *Automated Deduction – CCADE 28*, volume 12699 of *Lecture Notes in Computer Science*, pages 625–635. Springer, 2021. doi: 10.1007/978-3-030-79876-5_37.
- [9] Peter Frankl. Extremal set systems. In Ronald L. Graham, Martin Grötschel, and László Lovász, editors, *Handbook of Combinatorics*, volume 2, pages 1293–1329. Elsevier, Amsterdam, 1995.
- [10] Justin Gilmer. A constant lower bound for the union-closed sets conjecture, 2022. arXiv:2211.09055.
- [11] Wim Hordijk and Mike Steel. Detecting autocatalytic, self-sustaining sets in chemical reaction systems. *Journal of Theoretical Biology*, 227(4):451–461, 2004. doi: 10.1016/j.jtbi.2003.11.020.
- [12] Wim Hordijk and Mike Steel. Chasing the tail: the emergence of autocatalytic networks. *BioSystems*, 152:1–10, 2017. doi: 10.1016/j.biosystems.2016.12.002.
- [13] Stuart A. Kauffman. Autocatalytic sets of proteins. *Journal of Theoretical Biology*, 119(1):1–24, 1986. doi: 10.1016/S0022-5193(86)80047-9.
- [14] Stuart A. Kauffman. *The Origins of Order: Self-Organization and Selection in Evolution*. Oxford University Press, New York, 1993.
- [15] Vadim Lozin and Viktor Zamaraev. Union-closed sets and Horn Boolean functions. *Journal of Combinatorial Theory, Series A*, 202:105818, 2024. doi: 10.1016/j.jcta.2023.105818.
- [16] Elchanan Mossel and Mike Steel. Random biochemical networks: the probability of self-sustaining autocatalysis. *Journal of Theoretical Biology*, 233(3):327–336, 2005. doi: 10.1016/j.jtbi.2004.10.011.
- [17] Bjorn Poonen. Union-closed families. *Journal of Combinatorial Theory, Series A*, 59(2): 253–268, 1992. doi: 10.1016/0097-3165(92)90068-7.
- [18] David Reimer. An average set size theorem. *Combinatorics, Probability and Computing*, 12(1):89–93, 2003. doi: 10.1017/S0963548302005539.
- [19] Dinesh G. Sarvate and Jean-Claude Renaud. On the union-closed sets conjecture. *Ars Combinatoria*, 27:149–154, 1989.
- [20] Will Sawin. An improved lower bound for the union-closed set conjecture, 2022. arXiv:2211.11504.
- [21] Mike Steel. Interior operators and their relationship to autocatalytic networks. *Acta Biotheoretica*, 71(4):21, 2023. doi: 10.1007/s10441-023-09472-8. arXiv:2305.16714.
- [22] Mike Steel, Wim Hordijk, and Joshua Smith. Minimal autocatalytic networks. *Journal of Theoretical Biology*, 332:96–107, 2013. doi: 10.1016/j.jtbi.2013.04.032.

- [23] Mike Steel, Wim Hordijk, and Joana C. Xavier. Autocatalytic networks in biology: structural theory and algorithms. *Journal of the Royal Society Interface*, 16(151):20180808, 2019. doi: 10.1098/rsif.2018.0808.
- [24] The mathlib Community. The Lean mathematical library. In *Proceedings of the 9th ACM SIGPLAN International Conference on Certified Programs and Proofs (CPP 2020)*, pages 367–381. ACM, 2020. doi: 10.1145/3372885.3373824.
- [25] Piotr Wójcik. Union-closed families of sets. *Discrete Mathematics*, 199(1–3):173–182, 1999. doi: 10.1016/S0012-365X(98)00292-0.