Fast enumeration of non-isomorphic chemical reaction networks

Carlo Spaccasassi, Boyan Yordanov, Andrew Phillips, and Neil Dalchau

Biological Computation group, Microsoft Research, Cambridge, CB1 2FB, UK
{t-caspac,yordanov,aphillip,ndalchau}@microsoft.com

Abstract. Chemical reaction networks (CRNs) have been applied successfully to model a wide range of phenomena and are commonly used for designing molecular computation circuits. Often, CRNs with specific properties (oscillations, Turing patterns, multistability) are sought, which entails searching an exponentially large space of CRNs for those that satisfy a property. As the size of the CRNs being considered grows, so does the frequency of isomorphisms, by up to a factor $N!$, where N is the number of species. Accordingly, being able to *generate* sets of non-isomorphic CRNs within a class can lead to large computational savings when carrying out global searches. Here, we present a bijective encoding of bimolecular CRNs into novel vertex-coloured digraphs called Complex-Species graphs. The problem of enumerating non-isomorphic CRNs can then be tackled by leveraging well-established computational methods from graph theory [20]. In particular, we extend Nauty, the graph isomorphism tool suite by McKay [22]. Our method is highly parallelisable and more efficient than competing approaches, and a software package (`genCRN`) is freely available for reuse. Non-isomorphs are generated directly by `genCRN`, alleviating the need to store intermediate results. We provide the first complete count of all 2-species bimolecular CRNs and extend previous known counts for classes of CRNs of special interest, such as mass-conserving and reversible CRNs.

1 Introduction

Chemical reaction networks (CRNs) are widely recognised as a convenient formalism for modelling and analysing a broad range of biochemical systems [17,1]. In recent years, they have also been used for designing synthetic systems with specified behaviours, such as distributed consensus networks [9], oscillators [34] and feedback control circuits [27]. CRNs provide a convenient abstraction for modelling synthetic biological systems, while also supporting a mapping to biological implementations in both molecular [33] and genetic [28] circuits.

CRNs also support a broad range of analysis methods, which can be used to check the desired properties of a system prior to its implementation. In particular, a promising approach is to encode a CRN as a graph and analyse its properties using graph-theoretic methods. A CRN is essentially a map from a multiset of *reactant* species to a multiset of *product* species, which can be can be written as

$$\alpha_{ji}A_i \xrightarrow{k_j} \beta_{ji}A_i, \quad j \in \{1,\ldots,M\}, i \in \{1,\ldots,N\} \tag{1}$$

where N denotes the number of species, M denotes the number of reactions, α_{ji} and β_{ji} denote the multiplicity of species A_i in the reactants and products of reaction j, respectively, and $k_j \in \mathbb{R}_+$ denotes the rate constant of reaction j. In a pair of landmark papers [15,16], Feinberg encoded CRNs as *complex graphs* – where each vertex represents a complex and each directed edge represents a reaction – and related the *deficiency* of a CRN to the existence of positive steady states. Alternative graph encodings have also been developed, including *species-reaction* (SR) graphs, which are directed bipartite graphs whose vertices are either species or reactions. These SR graphs are used to check for the existence of multiple equilibria that can be determined from network structure alone [11,12]. Graph-theoretic properties have also been developed for detecting oscillations [24] and Turing instabilities [23], and for assessing concentration robustness [32].

As the number of species N increases, the number of possible reactions grows such that the number of bimolecular CRNs grows as $\mathcal{O}(2^{N^4})$ (see Lemma 4 in Appendix B). As a result, the design space of CRNs with more than a few species is intractable to explore systematically [26], and the design of CRNs with specified behaviours remains largely an artisanal process. One approach to exploring the design space of CRNs more efficiently is to filter out CRNs that are *isomorphic* and therefore exhibit identical behaviour. In principle, the enumeration of these non-isomorphic CRNs can make it possible to exhaustively explore an otherwise intractable space, since as the number of species N increases, the number of isomorphic CRNs also increases substantially. More generally, enumerating non-isomorphic CRNs can be used to determine which CRNs satisfy a property in a *complete* sense, such as determining the complete set of 2-species CRN oscillators [4], or the smallest CRN admitting bistability [35].

The problem of enumerating non-isomorphic CRNs is related to the problem of enumerating non-isomorphic graphs, which is NP-hard and, worse still, considered to be a pathology of computer science research [30]. Several methods for working with graph isomorphisms already exist, the most notable of which is NAUTY [22], which can efficiently compute a canonical form of a graph, find its automorphism group and its generators. NAUTY also provides enumeration tools for graphs, digraphs and vertex colouring, among others. Also related is Polya's enumeration theorem, which *counts* non-isomorphic graphs *without* constructing them [29]. The most promising method for working with isomorphic CRNs was introduced in [3], which uses NAUTY to encode a CRN as a *species-reaction Petri net* [2]. This is similar to an SR-graph, except that edges are directed and weighted. It was inspired by attempts to enumerate CRNs in [13], which also leverages NAUTY. However, species-reaction Petri nets need to be encoded as multidigraphs, which are not supported natively in NAUTY. They can be encoded in terms of digraphs, but such an encoding is not enumerable in NAUTY without also generating invalid multidigraphs. This requires storing and then filtering out non-isomorphs after enumeration, which penalises run-time. Time measurements or a software tool are not available in [3], so it is hard to quantify the number of non-isomorphs and their negative impact on performance. The maximum counts reported in these works are for the bimolecular CRN classes of size (N, M) from (2,7), (3,6), (4,5), (5,4) to (9,4) and (10,3), with a maximum running time of 20 days [13]; (5,5) is reported in [3].

The other major challenge of checking large sets of CRNs is that storing the set of CRNs in the memory of a computer becomes impossible beyond some problem size (N, M), even when using a memory-efficient representation of the CRN. For example, the (5,5) class stored in the encoding of [3] takes 64.4 gigabytes of disk space; the (5,6) class of reversible CRNs takes 198 gigabytes. As such, the only way to proceed practically is to directly generate non-isomorphic CRNs using the *canonical construction path* method [21], check whether the CRN satisfies the predicate, and write those to file (or store in memory if the satisfying subset happens to be small enough).

In this paper, we present an efficient method for *generating* non-isomorphic bimolecular CRNs. Our method can determine the complete subset of CRNs of a given size that satisfy a specified property, without the need to enumerate and store in memory all non-isomorphic CRNs of that size. By creating such a generator, our method can be used to ask complete questions for larger CRN sizes than was previously possible, since memory is no longer limiting. Instead, it is limited only by the computation time of testing each non-isomorphic CRN. Our approach is based on a new graph encoding of CRNs that we name the *Complex-Species graph* (*CS-graph*), and we prove that isomorphisms of bimolecular CRNs are equivalent to isomorphisms of CS-graphs. Our method also facilitates a tighter relationship with NAUTY than previous methods, leading to efficiency benefits and high parallelisation. We are the first to report that there are precisely 535,852,102 bimolecular 2-species CRNs, and extend the counts of non-isomorphic CRNs with more than 2 species beyond what has been reported previously, including counts for (10,5), (5,6) and (4,7) in less than a day. We provide execution times of all enumerations, and a new computational tool (`genCRN`) for enumerating non-isomorphic CRNs with several filters. Using `genCRN`, it is now possible to explore the design space of larger CRNs satisfying a given set of properties.

2 Methods

We first present the *Complex-Species graph* (CS-graph), an encoding of bimolecular CRNs into directed coloured graphs, and prove that CS-graphs faithfully encode bimolecular CRNs up to isomorphism, in the sense that two bimolecular CRNs are isomorphic if and only if their CS-graphs are isomorphic. We then explain how CS-graphs facilitate the fast enumeration of the set of all CRNs that are non-isomorphic to one another.

2.1 Complex-Species graph encoding

We begin with a formal definition of CRNs and CRN isomorphism. We define $\mathcal{S}$ to be a set of species and $\mathbb{C}$ to be the space of complexes, which is any combination of species that may appear as the reactant or product set in a reaction. A set $\mathcal{C} \subset \mathbb{C}$ is a set of multisets, where $c \in \mathcal{C}$ is a pair (A, m) with $A \in \mathcal{S}$ and $m \in \mathbb{N}$.

Definition 1 (CRN). *A chemical reaction network* $\mathcal{N} = (\mathcal{S}, \mathcal{C}, \mathcal{R})$ *consists of a set of species* $\mathcal{S}$*, a set of complexes* $\mathcal{C} \subset \mathbb{C}$*, and a set of reactions* $\mathcal{R} \subset \mathcal{C} \times \mathcal{C}$ *with* $(y, y) \notin \mathcal{R}$ *for any* $y \in \mathcal{C}$*.*

Two CRNs are isomorphic when they are identical under species renaming:

Definition 2 (CRN isomorphism). *Let* $\mathcal{N}_1$ *and* $\mathcal{N}_2$ *be chemical reaction networks.* $\mathcal{N}_1$ *is* isomorphic *with* $\mathcal{N}_2$*, or* $\mathcal{N}_1 \cong \mathcal{N}_2$*, if there exists a permutation* π *over* $\mathcal{S}$ *such that* $\mathcal{N}_1\pi = \mathcal{N}_2$*.*

Here, we have written the function application $\mathcal{N}_1\pi$ in postfix notation. Note that the reaction rates are not relevant in CRN isomorphism, so are not included in this definition and are omitted from the remainder of this paper.

Before introducing CS-graphs, we introduce a technical device to more conveniently index the elements of a set:

Definition 3 (Indexed set). *Let* I *and* S *be sets, and* f *be a bijection* $I \to S$*. The set* $\{S_i\}_{i \in I} \triangleq \{S_i \mid S_i = f(i), i \in I\}$ *is an* indexed set*, and* I *is the* indexing set*. We write* S_i *for* $S_i = f(i)$ *with* $i \in I$ *when* f *is clear from the context.*

If $\mathcal{N} = (\mathcal{S}, \mathcal{C}, \mathcal{R})$ is a CRN, we indicate with $\{c_i\}_{i \in I}$ the indexed set of complexes occurring in $\mathcal{N}$, and with $\{A_j\}_{j \in J}$ the indexed set of species occurring in $\mathcal{N}$. For the remainder of this section, the indexing sets I and J always index respectively the complexes and the species of a CRN; moreover, we assume that $I \cap J = \emptyset$ and $I, J \subset \mathbb{N}$. We are now ready to define the CS-graph of a bimolecular CRN:

Definition 4 (Complex-Species graph). *Let* $\mathcal{N}$ *be a bimolecular CRN with indexed sets* $\{c_i\}_{i \in I}$ *and* $\{A_j\}_{j \in J}$*. The* Complex-Species graph $[\![\mathcal{N}]\!]^I_J$ *is the quadruple* $\langle V, E, \sigma, \rho \rangle$*, where:*

$$\begin{aligned}
V &= I \cup J && \textit{(Vertices)}\\
E &= \{(j, i) \mid A_j \textit{ occurs in } c_i\} && \textit{(Edges)}\\
&\quad \cup \{(i_1, i_2) \mid c_{i_1} \to c_{i_2} \textit{ occurs in } \mathcal{N}\} &&\\
\sigma(i) &= \begin{cases} \emptyset & \textit{if } c_i = \emptyset \\ \square & \textit{if } c_i = A_j \\ 2\square & \textit{if } c_i = 2A_j \\ \square\square & \textit{if } c_i = A_{j_1} + A_{j_2} \end{cases} && \textit{(Stoichiometry function)}\\
\rho(j) &= A_j && \textit{(Labelling function)}
\end{aligned}$$

for $i, i_1, i_2 \in I$ *and* $j, j_1, j_2 \in J$*.*

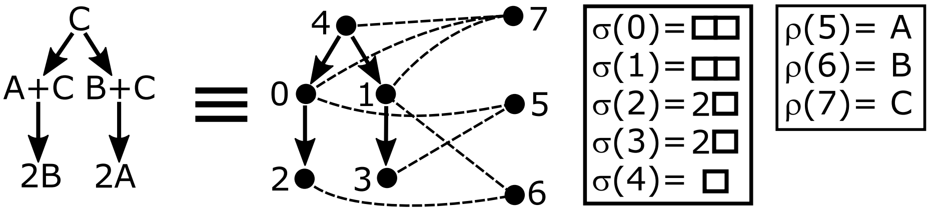


Fig. 1. Complex-Species graph of a bimolecular CRN. The set $I = \{0, 1, 2, 3, 4\}$ is the indexing set for the complexes of the CRN, while $J = \{5, 6, 7\}$ is the indexing set for its species. The concrete names of the indexes are unimportant; any disjoint set I and J can be used. Set I indexes the complexes of a CRN, the stoichiometry function σ assigns a multiplicity to each index (e.g. σ assigns heterodimer to node 0, homodimer to node 2 and monomer to node 4). An edge between two nodes in I represents a CRN reaction. Set J indexes the CRN species, with labelling function ρ assigning them concrete species names. A dashed edge from node $j \in J$ to node $i \in I$ means that species $\rho(j)$ occurs in complex i.

Fig. 1 provides a visual representation of a Complex-Species graph. Notice that it is not possible to distinguish monomers from homodimers using the encoding's vertices and edges alone; this is accomplished by σ. The indexing set I in the figure is the same indexing set returned by NAUTY for that digraph. Appendix C shows an extension to CS-graphs to encode CRNs with higher molecularity.

Two CS-graphs are isomorphic when their underlying graphs are isomorphic and have the same stoichiometry:

Definition 5 (CS-graph isomorphism). *Let* $[\![\mathcal{N}_1]\!]^{I_1}_{J_1} = \langle V_1, E_1, \sigma_1, \rho_1 \rangle$ *and* $[\![\mathcal{N}_2]\!]^{I_2}_{J_2} = \langle V_2, E_2, \sigma_2, \rho_2 \rangle$. *Complex-Species graph* $[\![\mathcal{N}_1]\!]^{I_1}_{J_1}$ *and* $[\![\mathcal{N}_2]\!]^{I_2}_{J_2}$ *are* isomorphic, *or* $[\![\mathcal{N}_2]\!]^{I_2}_{J_2} \cong [\![\mathcal{N}_2]\!]^{I_2}_{J_2}$, *if there exist bijections* $\alpha : I_1 \to I_2$ *and* $\beta : J_1 \to J_2$ *such that:*

1. $V_1 \alpha\beta = V_2$
2. $E_1 \alpha\beta = E_2$
3. $\sigma_1 \alpha = \sigma_2$

where $\alpha\beta$ *stands for the function composition of* α *and* β.

As already pointed out, the actual indexing sets used in a CS-graph are unimportant. As a matter of fact, we can show that CS-graphs of the same CRN are all isomorphic with each other:

Lemma 1. *Let* $\mathcal{N}$ *be a bimolecular CRN. Then* $[\![\mathcal{N}]\!]^{I_1}_{J_1} \cong [\![\mathcal{N}]\!]^{I_2}_{J_2}$ *holds for any indexing sets* I_1, I_2, J_1, J_2.

Proof. The lemma is proved by explicitly constructing bijections $\alpha = \{(i_1, i_2) \mid c_{i_1} = c_{i_2} \text{ for } i_1 \in I_1, i_2 \in I_2\}$ and $\beta = \{(j_1, j_2) \mid A_{j_1} = A_{j_2} \text{ for } j_1 \in J_1, j_2 \in J_2\}$ that satisfy Def. 5. See Appendix B for more details. □

Having proved this result, and since CRN isomorphism provides a permutation of species π such that two CRNs become equal, it is easy to show that CRN isomorphism implies CS-graph isomorphism:

Lemma 2. *Let* $\mathcal{N}_1$ *and* $\mathcal{N}_2$ *be bimolecular CRNs. If* $\mathcal{N}_1 \cong \mathcal{N}_2$*, then* $[\![\mathcal{N}_1]\!]^{I_1}_{J_1} \cong [\![\mathcal{N}_2]\!]^{I_2}_{J_2}$ *for any indexing sets* I_1, I_2, J_1 *and* J_2*.*

Proof. By Def. 2, there exists a permutation π over the species of $\mathcal{N}_1$ such that $\mathcal{N}_1\pi = \mathcal{N}_2$. Notice that by Lem. 1 we can deduce $[\![\mathcal{N}_1\pi]\!]^{I_1}_{J_1} \cong [\![\mathcal{N}_2]\!]^{I_2}_{J_2}$ for any indexing sets I_1, I_2, J_1 and J_2. The lemma is proved by taking $\alpha = \{(i_1, i_2) \mid c_{i_1}\pi = c_{i_2} \text{ for } i_1 \in I_1, i_2 \in I_2\}$ and $\beta = \{(j_1, j_2) \mid A_{j_1}\pi = A_{j_2} \text{ for } j_1 \in J_1, j_2 \in J_2\}$. □

When two CS-graphs are isomorphic, the indexed sets of complexes and species provide enough information to reconstruct an isomorphism π for their original CRNs:

Lemma 3. *Let* $\mathcal{N}_1$ *and* $\mathcal{N}_2$ *be bimolecular CRNs with indexing sets respectively* I_1, J_1 *and* I_2, J_2*. If* $[\![\mathcal{N}_1]\!]^{I_1}_{J_1} \cong [\![\mathcal{N}_2]\!]^{I_2}_{J_2}$*, then* $\mathcal{N}_1 \cong \mathcal{N}_2$*.*

Proof. Let $[\![\mathcal{N}_1]\!]^{I_1}_{J_1} = \langle V_1, E_1, \sigma_1, \rho_1\rangle$ and $[\![\mathcal{N}_2]\!]^{I_2}_{J_2} = \langle V_2, E_2, \sigma_2, \rho_2\rangle$, such that $[\![\mathcal{N}_1]\!]^{I_1}_{J_1} \cong [\![\mathcal{N}_2]\!]^{I_2}_{J_2}$. By hypothesis, $[\![\mathcal{N}_1]\!]^{I_1}_{J_1} \cong [\![\mathcal{N}_2]\!]^{I_2}_{J_2}$ implies the existence of bijections α and β that satisfy conditions 1 to 3 in Def. 5. The lemma is proved by taking $\pi = \{(A_{j_1}, A_{j_2}) \mid j_1\beta = j_2\} \circ \pi_I$ where π_I is the identity function over $\mathcal{S}$. See Appendix B for more details. □

We can now show that CS-graphs are a faithful encoding of bimolecular CRNs up to isomorphism:

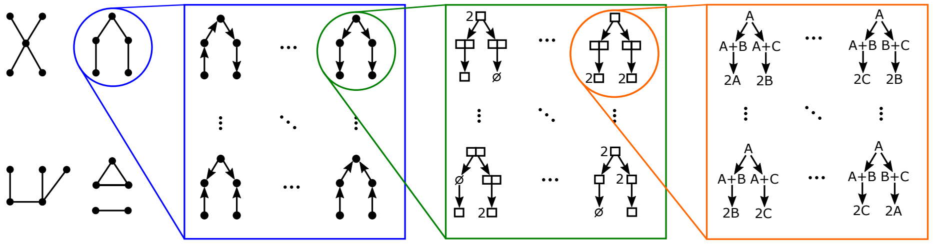
Theorem 1 (Faithful encoding). *Let* $\mathcal{N}_1$ *and* $\mathcal{N}_2$ *be bimolecular CRNs with indexing sets respectively* I_1, J_1 *and* I_2, J_2*. Then* $\mathcal{N}_1 \cong \mathcal{N}_2$ *if and only if* $[\![\mathcal{N}_1]\!]^{I_1}_{J_1} \cong [\![\mathcal{N}_2]\!]^{I_2}_{J_2}$*.*

Proof. By Lem. 2 and Lem. 3. □

2.2 Isomorph-free Complex-Species graphs enumeration

Our non-isomorphic CS-graph enumeration method entails the generation of all non-isomorphic bimolecular CRNs, by virtue of Theorem 1. Our method inputs are the numbers of complexes L, reactions M and species N of the output CRNs. CS-graphs are generated through four successive enumeration stages, where each stage turns a structure generated in the previous stage into a list of more refined non-isomorphic structures (Fig. 2).

The first stage enumerates all undirected graphs with L nodes and M edges. Each undirected graph represents the topology of a CRN. The second stage orients the edges of an undirected graphs in all possible directions, including both directions at the same time (by replacing an undirected edge with two opposite directed edges). A directed edge between two nodes represents a reaction between two complexes; two opposite edges mean a reversible reaction. For each directed graph, the third stage assigns all possible stoichiometries σ to the nodes, generating a list of *unassigned CS-graphs*, that is CS-graphs without species nodes. The fourth stage finds all possible assignments for N species nodes to an unassigned CS-graph, therefore listing non-isomorphic CRNs by Thm. 1.



A 4 graphs **B** 45 digraphs **C** 312 unassigned Complex-Species graphs **D** 10 CRNs, as Complex-Species graphs

Fig. 2. Isomorph-free generation of 3-species bimolecular CRNs with 5 complexes and 4 reactions. CRN enumeration comprises four stages. It starts with generating the list of all non-isomorphic undirected graphs with 5 vertices and 4 edges (**A**); for each graph the list of all non-isomorphic directed graphs (**B**); for each directed graph the list of all non-isomorphic unassigned CS-graphs (**C**), and finally for each unassigned CS-graph the list of all non-isomorphic CRNs (**D**). The total count of elements enumerated from a single element in each stage is reported. There are 428,502 non-isomorphic CRNs against 635,040 isomorphs in total.

The state-structured enumeration we present (Fig. 2) is based on McKay's *canonical construction path* method [21], whereby a generation of larger structures is first constructed from a previous generation of smaller non-isomorphic structures, and then filtered out by some *canonical form* function f. This function maps all structures in an isomorphism class to the same structure in that class, which is called *canonical*. Only the canonical form is retained from the generated structures.

For example, let G be a graph of size n. A new graph G' of size $n+1$ can be obtained by adding a new node to G and a new set of edges between the new node and any subset of nodes in G. The new graph G' is discarded unless G' is in canonical form, i.e. $G' = f(G')$ for a canonical function f. A simple but inefficient example of $f(G')$ is to apply all possible node permutations to G', sort the resulting graphs by lexicographic order on their edges, and return the least graph in the sorting. Starting from the empty graph, it is then possible to enumerate all graphs by iteratively constructing and filtering larger non-isomorphic structures.

2.3 Enumeration invariants and implementation details

Although the generation of classes of larger structures might grow combinatorially, in practice the judicious use of graph invariants reduces this number greatly [21,6]. A notable example is the graph isomorphism tool suite by McKay, based on NAUTY [22]. NAUTY is a fast coinductive algorithm to find a graph's canonical labelling and its automorphism group [31]. A graph automorphism is a permutation of vertices that maps the graph onto itself. The actions of the automorphism group can generate a graph's isomorphism class very efficiently. Moreover, the generators can be used early to avoid generating non-canonical candidate structures immediately.

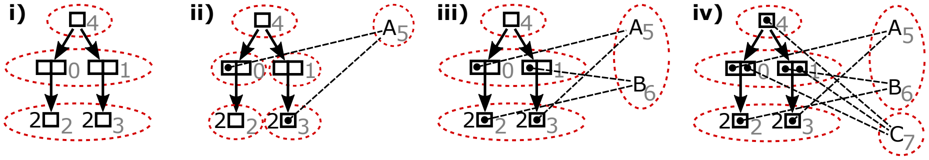


Fig. 3. Assignment of three species to an unassigned CS-graph in four steps. To simplify the visualisation, complex multiplicities have been overlaid over the complex nodes, where complex and species indices are subscript next to each node in grey. The red dashed circles indicate the existence of an *automorphism*, a permutation of nodes that maps the CS-graph to itself. For example, in step i) the permutation $[0 \mapsto 1, 1 \mapsto 0, 2 \mapsto 3, 3 \mapsto 2, 4 \mapsto 4]$, or (01)(23)(4) in permutation cycle notation [31], returns the same CS-graph. At step ii) the only automorphism is the identity permutation. The resulting CS-graph represents the CRN from Fig. 1; the automorphism reveals that species A and B are symmetric in the CRN.

Graphs and directed graphs in the first and second stage of Fig. 2 are generated respectively by `geng` and `directg`, two enumeration programs available in the NAUTY tool suite. The enumeration of unassigned CS-graphs can be encoded as graph-vertex colouring problems. We use four colours, one per stoichiometry type (naught, monomer, homodimer, heterodimer). The enumeration of a coloured graph is performed by another NAUTY tool, `vcolg`. However, not all graph colourings result in a valid CRN stoichiometry: for example, a 3-species CRN cannot have 4 monomers, since there are only 3 species available to make a monomer from. We have thus modified `vcolg` to enumerate valid stoichiometries only, by providing an upper bound for the number of nodes with each specific colour. A valid CRN can have at most one naught complex, n monomers, n homodimers and $\binom{n}{2}$ heterodimers. The total number of complexes for a given maximum number of reactants p is given by the sum of the multiset coefficients:

$$L_p(n) := \sum_{i=0}^{p} \left(\!\!\binom{n}{i}\!\!\right) = \sum_{i=0}^{p} \binom{n+i-1}{i} = \binom{n+p}{p} \tag{2}$$

which for bimolecular CRNs with N species is $\binom{N+2}{2}$.

The last stage is the enumeration of CRNs from an unassigned CS-graph, for which we have developed a custom algorithm following the canonical construction path method [21]. The structures we augment are partially assigned CS-graphs, starting from an unassigned one. Larger structures are obtained by adding a new species node, together with a set of edges that assign the new species to a subset of the complex nodes in the graph, until all N have been added and all complexes are valid. We call *species assignment* the set of complex nodes targeted by the new species; for example, in step ii) of Fig. 3, $\{0, 3\}$ is the species assignment for A.

As pointed out in [20], it is crucial to exploit graph invariants in order to curb the number of larger structures to test for canonicity. We adopt some of `geng`'s invariants in our enumeration method when adding a new species node. In order to avoid constructing the same graph more than once by adding the same species assignments in a different order, we impose a lexicographic order on the species assignments. For example, the choice of species assignment $\{1, 2\}$ at step iii) of Fig. 3 is allowed, because it is greater than the previous assignment $\{0, 3\}$ at step ii); if $\{1, 2\}$ is chosen first, $\{0, 3\}$ breaks the lexicographic order and is illegal. Similarly, the cardinality of the species assignment must be equal or greater than the previous one. As for vertex colouring, not all species assignments are valid; for example, the same species cannot be assigned to two different monomer complexes, or a species to naught. Such assignments are discarded immediately.

After augmenting a CS-graph with a new species, it is tested for canonicity. The test applies all possible automorphisms α and β to the current CS-graph G: if G is the least graph of all $G\alpha\beta$ graphs by lexicographic order, then G is in canonical form, and used to assign more species to it. When the automorphism group is trivial (the only automorphism is the identity) any species assignment added in lexicographic order is already canonical.

The fourth stage does not add new complex nodes or edges, therefore the automorphisms β over complexes are either the same or they decrease after adding new species assignment, which might introduce asymmetries in the graph. For example, the species assignments of A in step ii) of Fig. 3 introduces an asymmetry that renders the automorphism graph trivial (the only automorphism is the identity); however adding a species assignment B of the same cardinality in step iii) restores the group. For this reason we only recompute a CS-graph's automorphism group after increasing the cardinality of its species assignments (which, as previously pointed out, are only added in increasing order).

An implementation of our method is available online at https://github.com/CSpaccasassi/genCRN for Windows and Unix systems. Our tool, called `genCRN`, implements the third and fourth stage of Fig. 2, and relies on inputs from `geng` and `directg`. `genCRN` is based on version 2.6 of NAUTY, where `geng` only generates graphs with a maximum size of 32 nodes. Our tool has the same limitation, it can only produce CS-graphs of size $|I \cup J| \leq 32$. Later versions of NAUTY raise this limit to 64 nodes; we leave the extension of our implementation to 64 nodes for future work.

3 Results

3.1 Complete enumeration of non-isomorphic 2-species CRNs

We applied our CRN enumeration technique to count how many non-isomorphic CRNs there exist with specified numbers of species and reactions. When considering only 2 species, we are able to provide a complete construction, covering all possible numbers of reactions (Fig. 4). Overall, we find that there are 536,884,871 non-isomorphic CRNs with 2 species.

The practical utility of non-isomorphic CRN enumeration is that it enables testing of properties of CRNs against a smaller set. Compared to the naive approach of writing all possible reactions among the N and picking all combinations of size M, checking only the non-isomorphic subset amounts to a computational

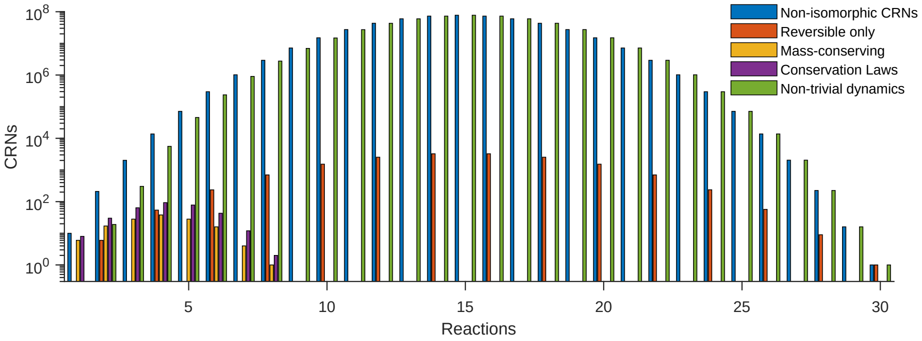


Fig. 4. Enumeration of non-isomorphic CRNs with 2 species. All non-isomorphic 2-species CRNs were enumerated (blue bars), and then filtered according to four criteria (separately): CRNs with only reversible reactions (red bars), mass-conserving CRNs (orange bars), CRNs with conservation laws (purple bars) and CRNs with non-trivial dynamics (green bars).

saving of at least $N!$. We also considered four simple filters of properties over the 2-species CRNs. Due to the tree structure of the approach (Fig. 2), applying filters as early as possible is preferable.

The first filter we checked was to restrict the enumeration to CRNs that have only reversible reactions. To achieve this, we modified our approach (Fig. 2) to skip part B (enumeration of directed graphs from undirected graphs), and immediately constructed undirected unassigned CS-graphs. Accordingly, each edge can be viewed as a reversible reaction. This enabled us to rapidly compute all reversible-only 2-species CRNs, of which there were exponentially fewer examples, yet following a similar Gaussian-like distribution over M (Fig. 4).

The second filter we applied was to identify *mass-conserving* CRNs, using the defining feature that there exists a vector $v \in \mathbb{R}^N_{>0}$ (all entries strictly positive) such that $v.\Gamma_k = 0$ for all k, where Γ is the stoichiometry matrix. For example, the CRN $A \rightarrow B$ is mass-conserving. It's stoichiometry matrix is $\Gamma = [-1, 1]^\top$, and so $v = [1, 1]$ can satisfy the property. To test for the existence of such a v in general, we used a Fourier-Motzkin algorithm to identify invariants $v \in \mathbb{R}^N$ [10]. As such invariants may include zero entries, we do an additional check to see whether all species participate in an invariant. E.g. strictly positive v can be constructed from the set of invariants. Rather than applying the filter directly to the complete set of non-isomorphic CRNs, we can obtain a computational saving by first removing the naught complex $\emptyset$ from $\mathcal{C}$, since any reaction involving $\emptyset$ would not be mass-conserving. In total, there were only 138 mass-conserving CRNs with 2 species, the largest of which had 8 reactions.

$$\begin{array}{llll} A \rightarrow B & A + B \rightarrow 2A & 2A \rightarrow A + B & 2A \rightarrow 2B \\ B \rightarrow A & A + B \rightarrow 2B & 2B \rightarrow A + B & 2B \rightarrow 2A \end{array}$$

This CRN simply includes all reactions that preserve the total molecule count. However, they are not the only reactions that are mass-conserving on their own. For example, the CRN $2A \rightarrow B$ is also mass-conserving, though now B has equivalent mass to 2 copies of A (e.g. this is simply homo-dimerisation). Accordingly, the counting of mass-conserving CRNs is not trivial.

We next identified CRNs for which there exists any conservation law, e.g. there exists a vector v such that $v.\Gamma_k = 0$ for all k, but in contrast to strictly mass-conserving CRNs, now v can include zero entries, as not all species need to participate in a conservation law for one to exist. Single-reaction examples include $A \to B$ ($A + B$ is conserved), $A \to A + B$ (A is conserved) and $A \to 2B$ ($2A + B$ is conserved). As before, we used the Fourier-Motzkin algorithm, but this time simply as a filter applied to the same enumeration approach for the full non-isomorphic set. We found 330 such CRNs (Fig. 4). As for the mass-conserving CRNs, there were no CRNs with more than 8 reactions, though this time an additional CRN was found:

$$\begin{array}{cccc} \emptyset \to B & \emptyset \to 2B & B \to 2B & A \to A + B \\ B \to \emptyset & 2B \to \emptyset & 2B \to B & A + B \to A \end{array}$$

Notably, this CRN includes $\emptyset$ complexes, but these only appear in reactions not interacting with the species A. Instead, the species A only participates by catalysing the production and degradation of B, and is not produced or consumed in these reactions. As such, A is conserved in this CRN.

Finally, we considered "dynamically non-trivial" CRNs [3], which can give rise to positive equilibria, periodic orbits, and other "interesting" properties. Dynamically trivial CRNs, in contrast, have no limit sets. e.g. trajectories grow unbounded in phase space. To enumerate dynamically non-trivial CRNs, we use the definition in [3], that a CRN $\mathcal{N}$ is dynamically trivial if there exists a vector $q > 0$ in $\operatorname{Im} \Gamma^\top$. [14] Accordingly, we find the reduced row echelon form of Γ and ask whether any row contains only non-negative entries (though not all zero). As done in [3], we take the set of non-isomorphic CRNs, and then check each CRN. Applying this filter to the 2-species CRNs reveals a considerably smaller number of CRNs with non-trivial dynamics than the full set, when there are few reactions (Fig. 4). As the number of reactions increases, the fraction of CRNs that are dynamically non-trivial tends to 1.

3.2 Enumeration of non-isomorphic CRNs with more than 2 species

Owing to the combinatorial nature of CRNs, simply extending to 3 species leads to an exponential increase in the number of possible CRNs. Using our enumeration method, we found that there are 1,244,363,180 bimolecular 3-species CRNs with $M = 7$ reactions, more than twice the number of all 2-species CRNs (Fig. 5A). Adding another reaction ($M = 8$) increases by a factor of 10 (12,916,870,803) and for $M = 9$ another factor of 10 (117,703,409,335). We enumerated and counted non-isomorphic CRNs with up to 10 species and with a number of reactions that could be evaluated within approximately 2 days of computation. In doing so, we have extended the known number of non-isomorphic CRNs beyond what was previously evaluated in [3], and have tabulated these values in Appendix D (Table S1). We have also evaluated non-isomorphic reversible

CRNs with up to 8 species (Fig. S2, Table S2) and non-isomorphic CRNs with non-trivial dynamics (Fig. S3, Table S3).

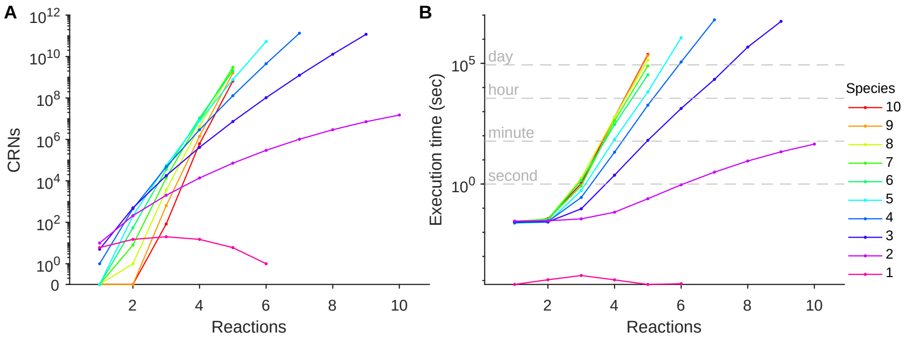


Fig. 5. Counting non-isomorphic CRNs. A. The total number of non-isomorphic CRNs is quantified for up to 10 species and up to 10 reactions. **B.** Execution time is quantified as if running on a single-core computer, by summing times over a parallel execution on an Intel Xeon Platinum 8168 2.70GHz machine with 72 cores.

While there is no reason our enumeration method cannot handle more reactions, additional combinatorial complexity leads to longer run times (Fig. 5B, S2B, S3B). We have quantified the execution times for each combination of species and reactions (N, M) by reporting the values as if the calculation was run sequentially on a single-core machine. In practice, we perform calculations in parallel, enumerating each digraph independently and collecting results.

McKay's labelling algorithm is known to have exponential complexity in the worst-case [25] but is well-behaved in practice. Similarly, there is no precise complexity for the canonical construction path method, although it depends on the size of the graph's automorphism groups [21]. In line with this, we found considerable differences in the execution times of each digraph (Fig. 6). The most adversarial digraphs are those which contain the highest number of disconnected sub-digraphs. Examples of such digraphs are those with $2n$ nodes and n edges, resulting in n sub-digraphs. The topology of such digraphs poses little constraints on the topology of the CRNs, and therefore the number of possible CRNs arising from such digraphs is combinatorially larger.

3.3 The non-isomorphic CRNs fraction

To gain a more quantitative understanding of the frequency of isomorphisms among sets of CRNs, we computed an *isomorph ratio*, defined simply as the number of non-isomorphic CRNs found, divided by the total number T of CRNs, for a given number of species and reactions. Using

$$T = \binom{L_p(N)(L_p(N)-1)}{M} - \sum_{k=1}^{s-1} \binom{L_p(k)(L_p(k)-1)}{M}, \tag{3}$$

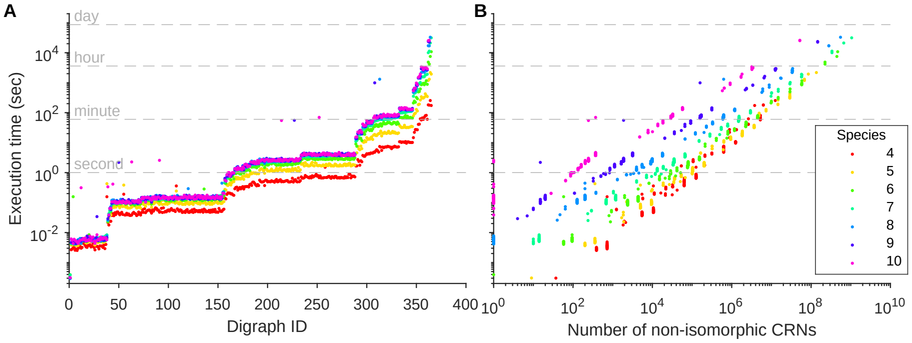


Fig. 6. Execution times vary considerably with input digraph. The execution times of enumerating non-isomorphic CRNs are shown for each of the 365 digraphs of CRNs with 5 reactions. **A.** The digraphs are sorted by the execution times corresponding to 7 species, which illustrates that the variation in execution time is strongly influenced by digraph structure. **B.** Execution times are compared against the total number of non-isomorphic CRNs found.

where $L_p(N) = \binom{N+p}{p}$ is the number of complex nodes in the CS-graph when there are N species, we computed the isomorph ratios for bimolecular CRNs ($p = 2$) with up to 5 species and 6 reactions (Fig. 7). By considering species relabellings alone, one would naively expect a factor $N!$ saving when considering isomorphisms. As there are $\binom{N+2}{2}$ possible complexes for N species (Eq. 2), there are 6 complexes for 2 species: $\{\emptyset, A, B, 2A, 2B, A+B\}$. As there are $L(L-1)$ possible (directed) edges connecting L nodes, there are 30 possible reactions for 2 species. Without considering CRN isomorphisms, this would result in $\sum_{r=1}^{30}\binom{30}{r} = 1,073,741,823$ possible CRNs. Whereas, we found that there are 536,884,871 non-isomorphic CRNs with 2 species, which is just more than half of the concretely labelled set. There are $N!$ permutations of N species, and so for most non-isomorphic CRNs, the $N!$ species permutations leads to an $N!$ different CRNs. However, some CRNs are *species-symmetric*, for example $A \rightleftharpoons B$, which means that a species relabelling can sometimes return the exact same CRN. Because such symmetries are automatically resolved in our calculation of the number of non-isomorphic CRNs, but not incorporated into Eq. 3, the isomorph ratio can be less than $1/N!$ (Fig. 7).

3.4 Checking properties of CRNs with external tools

As mentioned above, a practical benefit of using non-isomorphic CRN enumeration is that filters can be applied to a stream of CRNs, producing subsets of CRNs satisfying a property of interest. Such a property need not be implemented in the same code base as the CRN enumerator, since results can be piped into external tools. To demonstrate this, we considered the existence of *forward bisimulations* of CRNs [7], using the ERODE tool [8] (Fig. 8). The existence of a forward bisimulation means that a subset of the species can be lumped into a single species, the result being a different CRN with fewer species but with

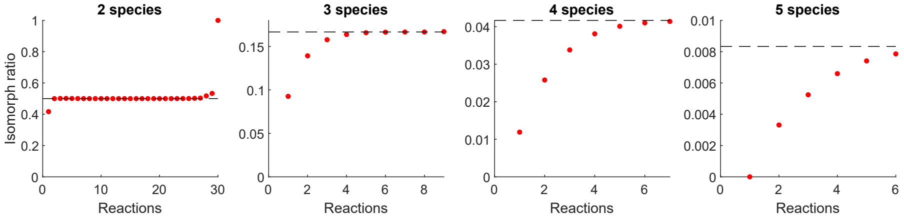


Fig. 7. The isomorph ratio. The ratio of non-isomorphic CRNs to the expected total number of CRNs was computed for different numbers of species and reactions. The expected total number of CRNs was calculated using Eq. 3 . The dashed black line indicates the value of $1/N!$, the reciprocal of the number of species relabellings in a CRN with N species.

the same behaviour. As such, the analysis of a CRN for which there exists a forward bisimulation can be considered to have been covered by equivalent analysis of CRNs with fewer species. The set of CRNs which are connected, have non-trivial dynamics and are irreducible via forward bisimulation was determined in less than 5 minutes, despite there being as many as 10^{11} CRNs initially covered by our encoding.

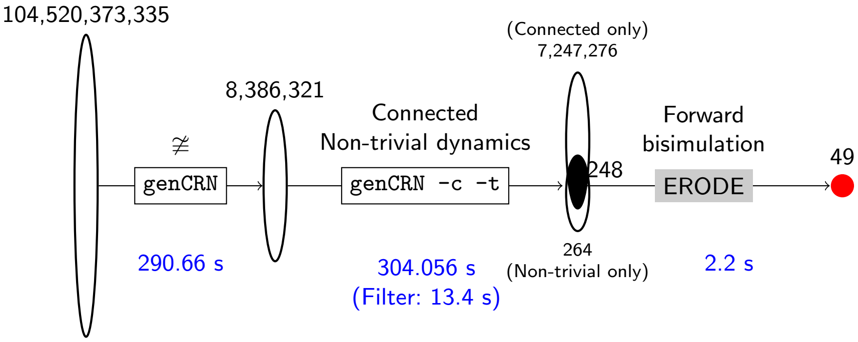


Fig. 8. Identifying non-isomorphic non-trivial connected CRNs with no forward bisimulation. CRNs with 7 species and 4 reactions were enumerated using `genCRN`, both in the presence and absence of filters for connectedness (`-c`) and non-trivial dynamics (`-t`). The resultant 251 CRNs were processed by ERODE, producing 49 CRNs for which no forward bisimulation exists.

4 Discussion

In this paper we have presented a method for the fast enumeration of non-isomorphic Chemical Reaction Networks, which enables *complete* statements on properties of classes of CRNs. The method is based on a novel encoding of CRNs into *Complex-Species graphs*, which are enumerated using established techniques such as the canonical construction path method [21,6], and implemented on top of the NAUTY tool suite [22,5]. We have shown that classes of non-isomorphic CRNs can be further specialised into classes satisfying certain properties of interest, such as mass convervation, non-trivial dynamics, reversible networks and non-lumpability under Forward Bisimulation.

We are the first, to the best of our knowledge, to report that there are precisely 535,852,102 bimolecular 2-species CRNs in total; a surprising number for just two species. The method is highly efficient, and can calculate this count in ≈ 27 minutes. It is also highly parallelisable; the same count on a 72 core machine takes ≈ 38 seconds. Currently our implementation enumerates unassigned CS-graphs and CRNs in a single step. Deeper parallelisation could be achieved by splitting this step into two, for handling more heavily combinatorial digraphs. We have extended previously known counts for CRNs in excess of 6 species or 6 reactions (see Appendix). Our counts and measurements are reproducible via `genCRN`, which is available online; comparison with other approaches [3,13] is difficult, because we have not found accompanying tools or time measurements.

Compared to other encodings, Complex-Species graphs are encoded in terms of digraphs and vertex colouring, and as such it has been easier to enumerate them with existing techniques and tools. For comparison, Species-Reaction graphs are expressed using multidigraphs or digraphs with edge labels, in order to express the multiplicity of a species in a reaction. For example, reaction $2A \rightarrow B$ is encoded either by two edges from a species node A to some reaction node R or by an edge with the label 2, to capture the fact that $2A$ is a heterodimer. NAUTY does not support multidigraphs or labelled digraphs natively [5], so CRN encodings relying on these lead to the *production* of isomorphs, which must be stored and filtered out in a subsequent step [3]. Interesting graph-theoretic results exist for complex graphs [15,16] and various directed or undirected bipartite graphs [12,23,24]; once the CRNs have been enumerated (and filtered) as CS-graphs, they can be translated into different representations for further analysis.

In future, it would be interesting to explore more advanced properties of CRNs, such as lumpability, multistability and limit cycles. Our tool allows testing sets of CRNs for overlapping properties, such as CRNs with no conservation laws and lumpable under Forward Bisimulation, or mass-conserving CRNs which are not lumpable under Forward Bisimulation, such as $2A \rightleftharpoons B$. Moreover, it would be interesting to verify properties on the unassigned CS-graphs; we conjecture that trivial dynamics occur when any species, independently from the others, is only assigned to *multiplicity monotonic paths* in an unassigned CS-graph, that is non-cyclic paths where the multiplicity of that species is ever increasing.

Our approach could be beneficial to the study of non-mass-action reaction systems such as Gene Regulatory Networks [18], or reaction-diffusion systems [19]. For example, special species roles as *fast diffuser*, *slow diffuser* and *other* could be encoded as ulterior nodes in a CS-graph, connected to species nodes and used as a further enumeration step. Broader applications in computer science might also be possible, to enumerate programs against a set of primitives, unique up to α-conversion. Species nodes might represent variables in the lambda calculus, or channels names in the π-calculus or CCS.

Acknowledgements

We would like to thank Brendan McKay, who extended NAUTY's vertex-colouring algorithm to directed graphs on our request, without which this work would have not been possible. We would also like to thank Andrea Vandin and Mirco Tribastone for providing a command-line version of ERODE and for useful discussions. Finally, we thank an anonymous reviewer for helping us to identify an error in the algorithm used for filtering dynamically trivial networks.

References

1. Angeli, D.: A tutorial on chemical reaction network dynamics. European journal of control **15**(3-4), 398–406 (2009)
2. Angeli, D., De Leenheer, P., Sontag, E.: A Petri Net Approach to Persistence Analysis in Chemical Reaction Networks, pp. 181–216. Springer Berlin Heidelberg, Berlin, Heidelberg (2007)
3. Banaji, M.: Counting chemical reaction networks with NAUTY. arXiv e-prints arXiv:1705.10820 (May 2017)
4. Bayramov, S.K.: New theoretical schemes of the simplest chemical oscillators. Biochemistry (moscow) **70**(12), 1377–1384 (2005)
5. Brendan D. McKay, A.P.: nauty and Traces User's Guide
6. Brinkmann, G.: Isomorphism rejection in structure generation programs. DIMACS Series in Discrete Mathematics and Theoretical Computer Science **51**(3), 25–38 (2000)
7. Cardelli, L., Tribastone, M., Vandin, A., Tschaikowski, M.: Forward and backward bisimulations for chemical reaction networks. In: CONCUR 2015 (2015)
8. Cardelli, L., Tribastone, M., Tschaikowski, M., Vandin, A.: Erode: A tool for the evaluation and reduction of ordinary differential equations. In: International Conference on Tools and Algorithms for the Construction and Analysis of Systems. pp. 310–328. Springer (2017)
9. Chen, Y.J., Dalchau, N., Srinivas, N., Phillips, A., Cardelli, L., Soloveichik, D., Seelig, G.: Programmable chemical controllers made from dna. Nature Nanotechnology **8**(10), 755 (2013)
10. Colom, J.M., Silva, M.: Convex geometry and semiflows in p/t nets. a comparative study of algorithms for computation of minimal p-semiflows. In: Rozenberg, G. (ed.) Advances in Petri Nets 1990. pp. 79–112. Springer Berlin Heidelberg, Berlin, Heidelberg (1991)
11. Craciun, G., Feinberg, M.: Multiple equilibria in complex chemical reaction networks: I. the injectivity property. SIAM Journal on Applied Mathematics **65**(5), 1526–1546 (2005)
12. Craciun, G., Feinberg, M.: Multiple equilibria in complex chemical reaction networks: Ii. the species-reaction graph. SIAM Journal on Applied Mathematics **66**(4), 1321–1338 (2006)
13. Deckard, A.C., Bergmann, F.T., Sauro, H.M.: Enumeration and Online Library of Mass-Action Reaction Networks. arXiv e-prints arXiv:0901.3067 (Jan 2009)
14. Farkas, J.: Uber die theorie der einfachen ungeichungen. J. Reine Angew. Math. **124**, 1–24 (1902)
15. Feinberg, M.: Chemical reaction network structure and the stability of complex isothermal reactors—i. the deficiency zero and deficiency one theorems. Chemical Engineering Science **42**(10), 2229–2268 (1987)

16. Feinberg, M.: Chemical reaction network structure and the stability of complex isothermal reactors—ii. multiple steady states for networks of deficiency one. Chemical Engineering Science **43**(1), 1–25 (1988)
17. Hucka, M., Finney, A., Sauro, H.M., Bolouri, H., Doyle, J.C., Kitano, H., Arkin, A.P., Bornstein, B.J., Bray, D., Cornish-Bowden, A., et al.: The systems biology markup language (sbml): a medium for representation and exchange of biochemical network models. Bioinformatics **19**(4), 524–531 (2003)
18. Karlebach, G., Shamir, R.: Modelling and analysis of gene regulatory networks. Nature Reviews Molecular Cell Biology **9**(10), 770 (2008)
19. Kondo, S., Miura, T.: Reaction-diffusion model as a framework for understanding biological pattern formation. science **329**(5999), 1616–1620 (2010)
20. McKay, B.D.: Isomorph-free exhaustive generation. Journal of Algorithms **26**(2), 306 – 324 (1998)
21. McKay, B.D.: Isomorph-free exhaustive generation. J. Algorithms **26**(2), 306–324 (1998)
22. McKay, B.D., Piperno, A.: Practical graph isomorphism, II. CoRR **abs/1301.1493** (2013)
23. Mincheva, M., Roussel, M.R.: A graph-theoretic method for detecting potential turing bifurcations. The Journal of Chemical Physics **125**(20), 204102 (2006)
24. Mincheva, M., Roussel, M.R.: Graph-theoretic methods for the analysis of chemical and biochemical networks. i. multistability and oscillations in ordinary differential equation models. Journal of Mathematical Biology **55**(1), 61–86 (2007)
25. Miyazaki, T.: The complixity of mckay's canonical labeling algorithm. In: Groups and Computation, Proceedings of a DIMACS Workshop, New Brunswick, New Jersey, USA, June 7-10, 1995. pp. 239–256 (1995)
26. Murphy, N., Petersen, R., Phillips, A., Yordanov, B., Dalchau, N.: Synthesizing and tuning stochastic chemical reaction networks with specified behaviours. Journal of The Royal Society Interface **15**(145), 20180283 (2018)
27. Oishi, K., Klavins, E.: Biomolecular implementation of linear i/o systems. IET systems biology **5**(4), 252–260 (2011)
28. Pedersen, M., Phillips, A.: Towards programming languages for genetic engineering of living cells. Journal of the Royal Society Interface **6**(suppl_4), S437–S450 (2009)
29. Pólya, G., Read, R.: Combinatorial enumeration of groups, graphs, and chemical compounds. Springer-Verlag, New York (1987)
30. Read, R.C., Corneil, D.G.: The graph isomorphism disease. Journal of Graph Theory **1**(4), 339–363 (1977)
31. Rosen, K.H.: Handbook of Discrete and Combinatorial Mathematics, Second Edition. Chapman & Hall/CRC, 2nd edn. (2010)
32. Shinar, G., Feinberg, M.: Structural sources of robustness in biochemical reaction networks. Science **327**(5971), 1389–1391 (2010)
33. Soloveichik, D., Seelig, G., Winfree, E.: Dna as a universal substrate for chemical kinetics. Proceedings of the National Academy of Sciences **107**(12), 5393–5398 (2010)
34. Srinivas, N., Parkin, J., Seelig, G., Winfree, E., Soloveichik, D.: Enzyme-free nucleic acid dynamical systems. Science **358**(6369), eaal2052 (2017)
35. Wilhelm, T.: The smallest chemical reaction system with bistability. BMC systems biology **3**(1), 90 (2009)

A Definitions

This section introduces definitions for automorphisms, orbits and the automorphism group for graphs, following [31].

Definition 6 (Permutation). *A permutation of a set S is a total function from S to itself.*

Definition 7 (Cyclic permutation). *A permutation π of the form:*

$$\begin{pmatrix} x & \pi(x) & \pi^2(x) & \cdots & \pi^{p-2}(x) & \pi^{p-1}(x) \\ \pi(x) & \pi^2(x) & \pi^3(x) & \cdots & \pi^{p-1}(x) & x \end{pmatrix}$$

is said to be cyclic permutation of period p.

Definition 8 (Disjoint cycle representation). *A* disjoint cycle representation *of a permutation π on a set S is a composition of cyclic permutations on subsets of S that constitute a partition of S, one cyclic permutation for each subset in the partition.*

Definition 9 (Group). *An algebraic system $< U, \star >$ is a called a* group *if it has the following properties:*

1. *the operation $\star$ is associative,*
2. *there is an identity element,*
3. *every element of U has an inverse.*

Definition 10 (Permutation group). *A closed non-empty collection P of permutations on a set Y of objects that forms a group under the operation of composition is called a* permutation group. *The combined structure may be denoted $V = [P : V]$. It is often denoted P when the set of Y objects is understood from context.*

Definition 11 (Orbit). *Let $\mathcal{P} = [P : Y]$ be a permutation group, and let $y \in Y$. The* orbit *of the object y under the action of P is the set $\{\pi(y) \mid \pi \in P\}$.*

Corollary 1. *Let $\mathcal{P} = [P : Y]$ be a permutation group. Then being coorbital is an equivalence relation*

Proof. Identity: by the identity permutation. Commutativity: because each π is invertible. Transitivity: by function composition $\circ$.

B Proofs

Lemma 1. *Let $\mathcal{N}$ be a bimolecular CRN. Then $[\![\mathcal{N}]\!]_{J_1}^{I_1} \cong [\![\mathcal{N}]\!]_{J_2}^{I_2}$ holds for any indexing sets I_1, I_2, J_1, J_2.*

Proof. The lemma can be proved by explicitly constructing bijections α and β required by Def. 5. Recall that we indicate with $\{c_i\}_{i\in I}$ and $\{A_j\}_{j\in J}$ respectively the indexed set of the complexes and of the species in $\mathcal{N}$.

Let $\alpha = \{(i_1, i_2) \mid c_{i_1} = c_{i_2} \text{ for } i_1 \in I_1, i_2 \in I_2\}$ and $\beta = \{(j_1, j_2) \mid A_{j_1} = A_{j_2} \text{ for } j_1 \in J_1, j_2 \in J_2\}$. These functions are well-defined because the indexing sets all target the same CRN $\mathcal{N}$. It is also easy to show that they are bijections.

The lemma is proved by verifying that α and β satisfy Condition 1 to 3 of Def. 5:

1. $V_1\alpha\beta = V_2$ because α and β are bijections over the indexed sets;
2. $$\begin{aligned} E_1\alpha\beta &= \{(j_1, i_1) \mid A_{j_1} \in c_{i_1}\}\alpha\beta \cup \{(i_1, i_1') \mid c_{i_1} \to c_{i_1'} \in \mathcal{R}\}\alpha\beta \quad \text{by Def. 4} \\ &= \{(j_2, i_2) \mid A_{j_2} \in c_{i_2}\} \cup \{(i_2, i_2') \mid c_{i_2} \to c_{i_2'} \in \mathcal{R}\} \quad \text{by def. of } \alpha, \beta. \\ &= E_2 \end{aligned}$$
   which proves the case.
3. Let i_1 be such that $\sigma_1(i_1) = \emptyset$. By Def. 4, $c_{i_1} = \emptyset$, and since $\alpha(i_1) = i_2$ implies that $c_{i_1} = c_{i_2}$, then $c_2 = \emptyset$ as well. Therefore $\sigma(i_2) = \emptyset$ holds by Def. 4, which implies $\sigma(i_1)\alpha = \sigma_2(i_2)$. The proof for the remaining cases (monomers, homodimers and heterodimers) is similar. ☐

Lemma 2. *Let $\mathcal{N}_1$ and $\mathcal{N}_2$ be bimolecular CRNs. If $\mathcal{N}_1 \cong \mathcal{N}_2$, then $[\![\mathcal{N}_1]\!]_{J_1}^{I_1} \cong [\![\mathcal{N}_2]\!]_{J_2}^{I_2}$ for any indexing sets I_1, I_2, J_1 and J_2.*

Proof. By definition of CRN isomorphism (Def. 2), there exists a permutation π over the species of $\mathcal{N}_1$ such that $\mathcal{N}_1\pi = \mathcal{N}_2$. Notice that by Lem. 1 we can deduce $[\![\mathcal{N}_1\pi]\!]_{J_1}^{I_1} \cong [\![\mathcal{N}_2]\!]_{J_2}^{I_2}$ for any indexing sets I_1, I_2, J_1 and J_2. The proof of this lemma is similar to Lem. 1, by defining $\alpha = \{(i_1, i_2) \mid c_{i_1}\pi = c_{i_2} \text{ for } i_1 \in I_1, i_2 \in I_2\}$ and $\beta = \{(j_1, j_2) \mid A_{j_1}\pi = A_{j_2} \text{ for } j_1 \in J_1, j_2 \in J_2\}$. ☐

Lemma 3. *Let $\mathcal{N}_1$ and $\mathcal{N}_2$ be bimolecular CRNs with indexing sets respectively I_1, J_1 and I_2, J_2. If $[\![\mathcal{N}_1]\!]_{J_1}^{I_1} \cong [\![\mathcal{N}_2]\!]_{J_2}^{I_2}$, then $\mathcal{N}_1 \cong \mathcal{N}_2$.*

Proof. Let $[\![\mathcal{N}_1]\!]_{J_1}^{I_1} = \langle V_1, E_1, \sigma_1, \rho_1\rangle$ and $[\![\mathcal{N}_2]\!]_{J_2}^{I_2} = \langle V_2, E_2, \sigma_2, \rho_2\rangle$, such that $[\![\mathcal{N}_1]\!]_{J_1}^{I_1} \cong [\![\mathcal{N}_2]\!]_{J_2}^{I_2}$. By hypothesis, $[\![\mathcal{N}_1]\!]_{J_1}^{I_1} \cong [\![\mathcal{N}_2]\!]_{J_2}^{I_2}$ implies the existence of bijections α and β that satisfy conditions 1 to 3 in Def. 5. Let us define the following permutation of $\mathcal{S}$:

$$\pi = \{(A_{j_1}, A_{j_2}) \mid j_1\beta = j_2\} \circ \pi_I$$

where π_I is the identity function over $\mathcal{S}$. Since β and π_I are bijections, then π is also a bijection; since its domain and range are $\mathcal{S}$, π is a well-defined permutation.

Let $c_{i_1} \to c_{i_1'}$ be a reaction in $\mathcal{N}_1$. By Def. 4 E_1 contains the edge (i_1, i_1'). Since $[\![\mathcal{N}_1]\!]_{J_1}^{I_1}$ and $[\![\mathcal{N}_1]\!]_{J_2}^{I_2}$ are isomorphic by hypothesis, it follows by definition that $E_2 = E_1\alpha\beta$, therefore the edge $(i_1, i'_1)\alpha = (i_2, i'_2)$ also exists in E_2 for $i_2, i'_2 \in I_2$. Because of this, the reaction $c_{i_2} \rightarrow c_{i'_2}$ exists in $\mathcal{N}_2$; moreover, by Condition 3 of Def. 5, the complexes have the same stoichiometry.

Similarly, let (j_1, i_1) be an edge in E_1 such that A_{j_1} occurs in c_{i_1}. By Condition 2 of Def. 5, E_2 contains the edge $(j_1, i_1)\alpha\beta = (j_1\alpha, i_1\beta) = (i_2, j_2)$, which means that A_{j_2} occurs in c_{i_2}. By definition of π, $A_{j_1}\pi = A_{j_1\beta} = A_{j_2}$; since c_{i_1} and c_{i_2} also have the same multiplicity by Condition 3 of Def. 5, this implies that $c_{i_1}\pi = c_{i_2}$. A similar line of reasoning shows that $c_{i'_1}\pi = c_{i'_2}$. Therefore $(c_{i_1} \rightarrow c_{i'_1})\pi = c_{i_2} \rightarrow c_{i'_2}$. Generalising this result to all reactions in $\mathcal{N}_1$, we obtain $\mathcal{N}_1\pi = \mathcal{N}_2$, which concludes the proof. □

Lemma 4. *The number of p-CRNs (reactions have up to p reactants/products) with up to N species grows as* $\mathcal{O}(2^{N^{2p}})$.

Proof. Following equation 3, the total number of p-CRNs with up to N and specifically M reactions is given by $\binom{L_p(N)(L_p(N)-1)}{M}$. Given that

$$L(N)(L(N)-1) = \frac{(N+p)\dots(N+1)}{p!} . \frac{(N+p)\dots(N+1) - p!}{p!} = \mathcal{O}(N^{2p}),$$

we can use the fact that $\sum_{i=1}^{k} \binom{n}{k} = 2^n$ to characterise the total number of bimolecular CRNs as $\mathcal{O}(2^{N^{2p}})$. □

C Complex-Multiplicity-Species Graph

Section 2.1 has shown how to encode bimolecular CRNs as vertex-coloured digraphs. It is natural to wonder whether this encoding extends to more than bimolecularity. Unfortunately CS-graph cannot encode higher molecularities than 2, however we propose in this section a more general encoding of CRNs called the *Complex-Multiplicity-Species graph* (or CMS graph).

We begin by showing that CS-graphs cannot encode trimolecular reactions. Consider in fact the reaction $2A+B \rightarrow A+2B$. If we added a new color "2□+□" and connect two node species A and B to it, there would be no way to tell which of the two species is actually the homodimer and which one is the monomer.

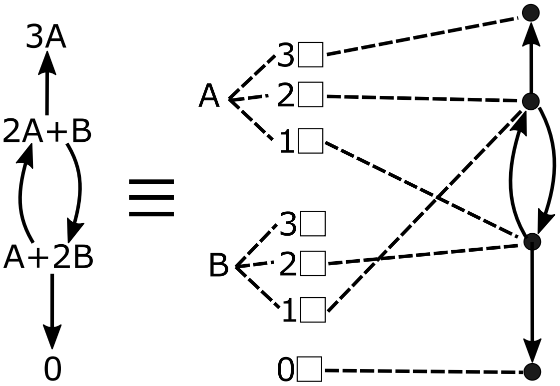


Fig. S1. Complex-Multiplicity-Species graph encoding of a CRN.

To overcome this issue, we propose Complex-Multiplicity-Species graphs, which extend CS-graphs with *multiplicity nodes*, that is distinct coloured nodes that point out the multiplicity of a species in a reaction. If m is the molecularity of interest, then there are $m+1$ kinds of multiplicity nodes: naught, □, 2□, 3□ and so on. Each species node is connected to m multiplicity nodes, signifying for example $A, 2A, 3A$ etc. Naught is a separate multiplicity node that cannot be connected to any species. In turn, multiplicity nodes are connected to complex nodes to represent the original CRN's reaction. Fig. S1 show an example of a CMS graph; notice that no confusion is possible between complexes $2A+B$ and $A+2B$.

We believe that CMS graphs are a general encoding of CRNs with any molecularity, but we leave a formal definition and proofs for future work.

D Counts of non-isomorphic CRNs

In this appendix, we tabulate the numbers of non-isomorphic CRNs found using `genCRN`. The tables can be compared against values reported at https://reaction-networks.net/networks, at the time of writing, which were evaluated using the method in [3]. In each case, we report values for *genuine* CRNs, those which use all N species.

D.1 No filters

Here, we consider the total number of non-isomorphic CRNs for N species and M reactions. The results are graphically depicted in Fig. 5, but tabulated below (Table S1).

Table S1. Genuine non-isomorphic CRNs. The number of non-isomorphic CRNs is shown for different numbers of species and reactions. Coloured in blue are those counts not available at https://reaction-networks.net/networks at time of writing.

	Reactions						
Species	1	2	3	4	5	6	7
1	6	15	20	15	6	1	0
2	10	210	2,024	13,740	71,338	297,114	1,018,264
3	5	495	17,890	414,015	7,262,666	103,511,272	1,244,363,180
4	1	451	47,323	2,900,934	128,328,834	4,518,901,463	133,379,120,523
5	0	204	55,682	7,894,798	763,695,711	56,929,248,832	
6	0	54	35,678	10,704,289	2,069,783,947		
7	0	8	13,964	8,386,321	3,041,467,242		
8	0	1	3,594	4,182,295	2,715,774,734		
9	0	0	639	1,417,784	1,595,551,325		
10	0	0	83	618,885	653,346,685		

D.2 Reversible CRNs

To generate reversible CRNs, we generate undirected graphs of a suitable size and feed these into `genCRN`in the same way as for general CRNs with irreversible reactions. Reported below are counts for M reversible reactions. As such, the CRNs found have $2M$ unidirectional reactions.

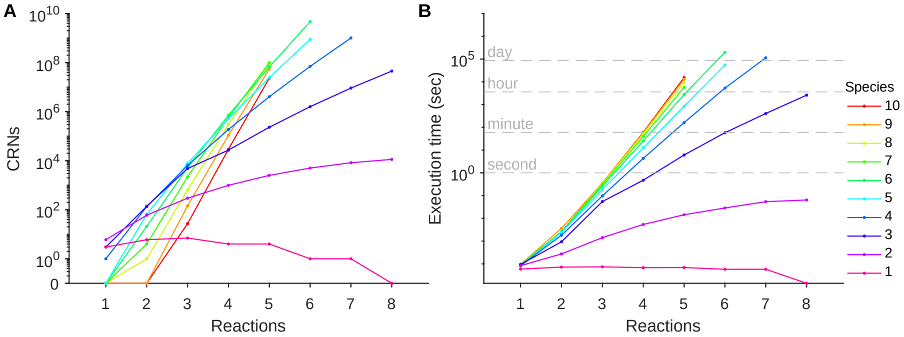


Fig. S2. Counts and execution times for enumeration of genuine non-isomorphic reversible CRNs.

Table S2. Genuine non-isomorphic reversible CRNs. The number of non-isomorphic CRNs with only reversible reactions is shown for different numbers of species and reactions. Coloured entries correspond to comparisons with the counts available at https://reaction-networks.net/networks at time of writing. Blue indicates values not available, and red indicates values that differ.

	Reactions							
Species	1	2	3	4	5	6	7	8
1	3	6	7	4	4	1	1	0
2	6	60	296	989	2,516	4,997	8,241	11,271
3	3	138	4,788	26,988	230,595	1,589,808	9,161,056	45,107,712
4	1	134	6,354	187,005	4,048,219	69,982,180	1,011,965,511	
5	0	65	7,677	513,036	24,186,053	888,323,405		
6	0	21	5,178	709,212	66,152,034	4,674,311,477		
7	0	4	2,188	572,058	98,576,689			
8	0	1	648	298,030	89,754,652			

D.3 Non-trivial dynamics

As described in the main text, one can test whether a CRN has non-trivial dynamics. To apply this filter to the enumerated non-isomorphic CRNs, one can use the `-t` flag for `genCRN`.

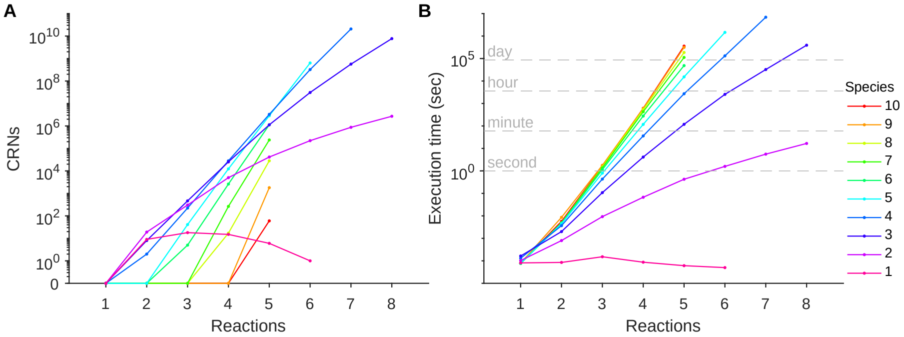


Fig. S3. Counts and execution times for enumeration of genuine non-isomorphic CRNs with non-trivial dynamics.

Table S3. Genuine non-isomorphic CRNs with non-trivial dynamics. The number of non-isomorphic CRNs with non-trivial dynamics is shown for different numbers of species and reactions. Coloured entries correspond to comparisons with the counts available at https://reaction-networks.net/networks at time of writing. Blue indicates values not available.

	Reactions							
Species	1	2	3	4	5	6	7	8
1	0	9	18	15	6	1	0	0
2	0	19	304	5,016	41,500	221,728	871,330	2,700,277
3	0	8	464	25,272	1,125,465	30,806,874	563,453,020	7,675,100,687
4	0	2	223	28,052	3,279,132	321,921,288	20,669,624,467	
5	0	0	41	12,340	2,845,389	633,623,890		
6	0	0	5	2,606	1,127,294			
7	0	0	0	264	238,105			
8	0	0	0	17	28,191			
9	0	0	0	0	1,795			
10	0	0	0	0	60			