Suplemento de la Revista Mexicana de Física **6** 011315 (2025) 1–8

Topological exploration of chemical hypergraphs using Information theory

H. G. Laguna and A. García-Chung

Departamento de Química, Universidad Autónoma Metropolitana Iztapalapa, Ciudad de México, 09310, México.

F. Betancourt

Departamento de Estudios en Ingeniería para la Innovación, Universidad Iberoamericana, Ciudad de México, 01219, México.

G. Restrepo

Max Planck Institute for Mathematics in the Sciences, Leipzig, Germany.

Received 14 March 2025; accepted 16 July 2025

A chemical hypergraph represents a set of chemical reactions. Hypernodes consist of sets of substances that act as reactants or products, while hyperedges correspond to chemical reactions, linking reactants to products. Another key structure in the hypergraph is the intersection of hypernodes, representing substances that participate in multiple reactions. In this work, we study a random walker on a chemical hypergraph under two different transition probability regimes. We characterize the random walker using network entropy, highlighting differences between these regimes. Additionally, we examine the structure of hypernodes by defining chemically inspired random variables and analyzing their joint and marginal Shannon entropies, as well as their mutual information. For large N, we observe bounds in these quantities.

Keywords: Shannon entropy; mutual information; network entropy; chemical hypergraph.

DOI: https://doi.org/10.31349/SuplRevMexFis.6.011315

1. Introduction

A graph G is a mathematical representation of a network consisting of a set of vertices $v \in V$ connected by edges $e \in E$. A defining characteristic of a graph is that each edge connects exactly two vertices, *i.e.*, $e = \{v_1, v_2\}$ where $v_1, v_2 \in V$. In a pictorial representation of G, vertices are depicted as points, and edges as lines connecting these points [1–3]. An illustration of a graph is shown in Fig. 1.

A hypergraph is a generalization of a graph that allows relationships between more than two vertices [4]. These multivertex relationships are called hyperedges and are typically represented pictorially as curves enclosing the related vertices.

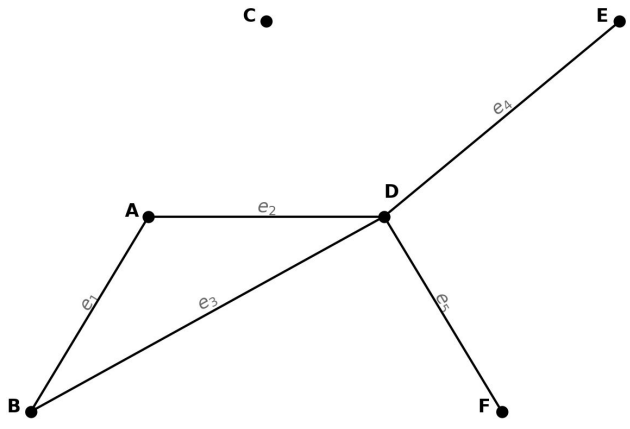

FIGURE 1. A graph with vertex set $V = \{A, B, C, D, E, F\}$ and edge set $E = \{e_1, e_2, e_3, e_4, e_5\}$.

A hypergraph model accounts for higher-order relationships (involving more than two vertices) and enables the study of features that are not apparent in traditional graphs. For example, consider a citation network where each author is represented as a vertex-multiple co-authors on a paper can be naturally represented by a hyperedge. Moreover, every graph can be represented as a hypergraph, but the converse is not always true, indicating that hypergraphs provide a more general framework for modeling real-world systems (networks). This generality is particularly relevant because real systems often extend beyond simple pairwise (binary) relationships. Therefore, if the goal is to model real-world systems accurately, it is essential to develop appropriate models for higher-order interactions.

Figure 2 illustrates a hypergraph. Mathematically, a hypergraph is defined as the pair (V, H) where V is the set of

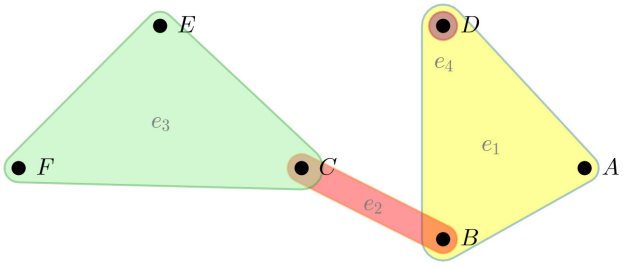

FIGURE 2. The hypergraph with vertex set $V = \{A, B, C, D, E, F\}$ and hyperedge set $H = \{e_1, e_2, e_3, e_4\}$. Note the different size of hyperedges e_i. For instance, e_1 and e_3 have size 3 (three vertices belong to each of them) while e_2 and e_4 have sizes 2 and 1, respectively.

$$\mathrm{r}_1 : D \to C + F + E$$
$$\mathrm{r}_2 : C + E + F \to A + B + D$$
$$\mathrm{r}_3 : C + B \to D$$

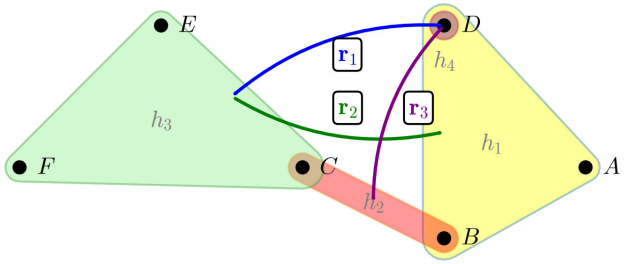

FIGURE 3. Chemical hypergraph for chemical reactions r_1 to r_3. Substances (nodes) are represented as black dots and reactions (hyperedges) as coloured lines labelled as r_i, hypernodes correspond to coloured regions h_1 to h_4.

nodes, and $H \subseteq \mathcal{P}(V)\backslash\emptyset$ is a collection of non-empty subsets of V, forming the hyperedge set.

Another structure, known as the *chemical hypergraph*, extends the concept of hypergraphs by allowing additional types of relationships beyond those present in a standard hypergraph and, consequently, in a graph [5, 6]. Figure 3 illustrates an example of a chemical hypergraph (ChemHy). A ChemHy, is the pair (V, H) such that

$$H \subseteq \{\{X, Y\} : X, Y \in \mathcal{P}(V)\backslash\emptyset \text{ and } X \cap Y = \emptyset\}. \quad (1)$$

As can be seen, a ChemHy has

- Nodes: denoted as $v \in V$ which are called substances in chemical contexts.
- Hypernodes: denoted as $X \in \mathcal{P}(V)\backslash\emptyset$, which are called collections of substances, again in connection with their application in chemistry, and notice that $\emptyset$ is not allowed.
- Hyperedges: connections between two hypernodes. These represent non-autocatalytic chemical reactions.

There is another structure that models decision-making processes in laboratories such as the selection of a substance to trigger a further chemical reaction. These decisions are modeled as intersections between hypernodes. We call these decisions chemical jumps [7]. It is worth mentioning that there are several approaches to the study of chemical reaction networks by using network science and hypergraphs [8, 9].

In the ChemHy model, features are defined to represent both the properties of substances and chemical reactions. One of the key features is the degree of a substance, which refers to the number of reactions in which a substance participates. Specifically, it corresponds to the number of hypernodes to which a substance belongs. Reactions are characterized by their size, which is determined by the number of substances involved in the transformation.

In this work, we represent idealized chemistry using a chemical hypergraph and study, on one hand, the probability of transition between hypernodes based on their connectivity and structure, and on the other hand, random variables related to the structure of the chemical reactions and their associated probability distributions. This approach reveals several probability distributions related to the topological structure of the ChemHy, making it suitable to apply information theory to study various topological features of the ChemHy.

2. The model

The ChemHy model represents an idealized chemical space[i] where all possible non-autocatalytic reactions occur. The model is constructed as follows: given a set of N substances denoted as

$$V = \{v_1, v_2, ..., v_N\}, \quad (2)$$

consider the set E of all possible non-autocatalytic reactions. It can be checked [5] that the number of all the possible chemical reactions is given by

$$N_r = \frac{1}{2}\left(3^N - 2^{N+1} + 1\right), \quad (3)$$

which is a huge number if one considers that the number of substances reported by Reaxys [10] is around 27 millions. Reaxys is one of the largest databases of chemical information. If we consider this *complete* ChemHy then we can construct its adjacency matrix, which is given as

$$M_{i,j} = \begin{cases} 1 & \text{if} \quad r(X_i, X_j) \in E \\ 0 & \text{if} \quad r(X_i, X_j) \notin E \end{cases}. \quad (4)$$

Let us provide an example to better illustrate this construction. Consider $N = 5$ and let us say that $V = \{A, B, C, D, E\}$. The number of chemical reactions in the complete ChemHy is $N_r = 90$. The reactions

$$A + B \longrightarrow C, \quad (5)$$
$$A \longrightarrow B, \quad (6)$$
$$A + C + D \longrightarrow E + B, \quad (7)$$

are allowed, but the auto-autocatalytic reactions, such as

$$A + B \longrightarrow A + C, \quad (8)$$
$$B + C + D + E \longrightarrow B + A, \quad (9)$$

are not allowed since at least one of the substances appears among the products and among the reactants. The adjacency matrix for this example is shown in Fig. 4, where reactants are represented in rows while products in columns. Examination of this adjacency matrix gives us information about the connection of the hypernodes.

It is worth mentioning that the direction of the reaction, although explicitly written for this example, is not actually a variable in our model.

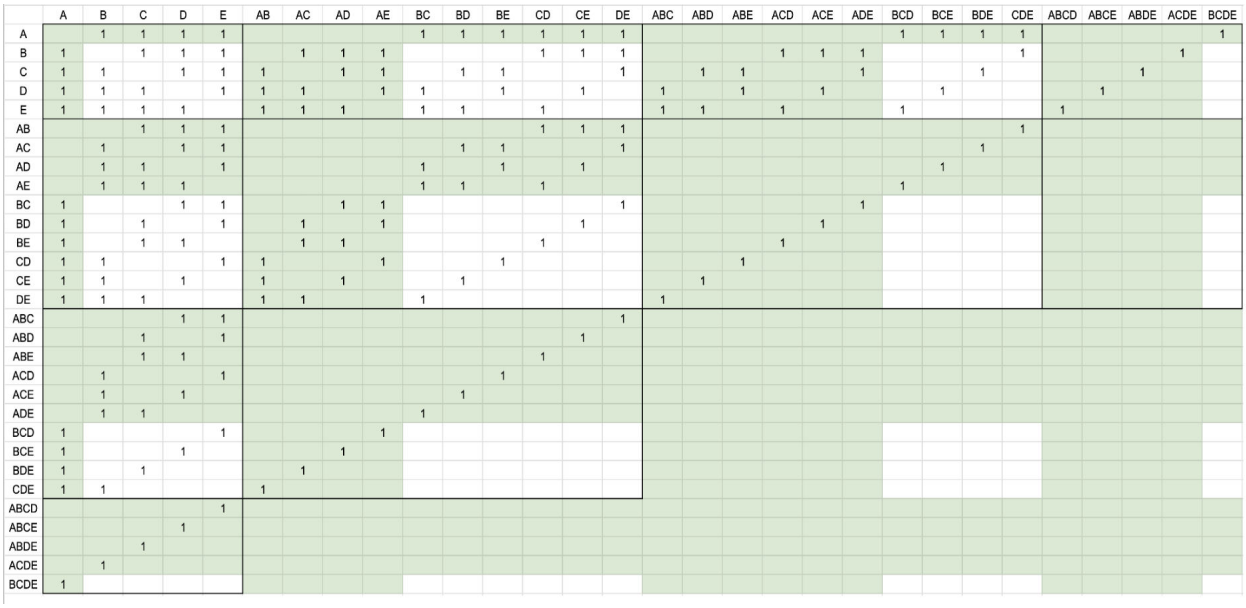

FIGURE 4. Adjacency matrix of the idealized complete chemistry with 5 substances. Empty cells denote not allowed reactions.

3. Network entropy

The topological exploration of the **hyperedge structure of ChemHy** and its interplay with the **internal hypernode structure** can be carried out using a random walker that moves between hypernodes based on chemical reactions. In this context, a random walk over the reaction hypergraph can be interpreted as a stochastic exploration of plausible reaction pathways. This framework enables the identification of highly connected substances or reaction motifs that may act as hubs or bottlenecks in synthetic routes. Furthermore, by adjusting the transition probabilities according to chemically inspired constraints –such as thermodynamic favorability, reaction type, or known kinetic preferences– the walk can be biased to reflect more realistic pathways. Therefore, random walks offer a way to simulate exploratory synthesis and analyze the structure of feasible transformation pathways in complex chemical systems. At this stage, we apply this framework to study the influence of connectivity and the number of substances on the walk. Let us now introduce the definition in the context of a network –a graph [11]– as follows.

Consider a network with N nodes and let k_i be the degree of node i,[ii] define the probability of going from node i to node j as

$$p_{i\to j} = \begin{cases} 0, & \text{for} \quad a_{ij} = 0 \\ \frac{1}{k_i}, & \text{for} \quad a_{ij} = 1 \end{cases}, \tag{10}$$

where a_{ij} is the adjacency matrix element that indicates if the nodes i and j are connected ($a_{ij} = 1$) or not ($a_{ij} = 0$). Note that for each node i, $\sum_j p_{i\to j} = 1$ if and only if the node is not isolated, that is, $k_i = 0$.

The entropy of each node can be defined as

$$\ell[p_{i\to\cdot}] = -\sum_{j=1}^{N-1} p_{i\to j} \ln p_{i\to j} = \ln k_i, \tag{11}$$

where we formally consider that

$$\lim_{p_{i\to j}\to 0} p_{i\to j} \ln p_{i\to j} = 0. \tag{12}$$

The normalized node entropy is calculated by dividing the entropy of each node by the maximum possible entropy that a node can have, that is, $\ln(N-1)$:

$$s^{(i)} = \frac{\ell[p_{i\to\cdot}]}{\ln(N-1)} = \frac{\ln k_i}{\ln(N-1)}, \tag{13}$$

and the normalized network entropy is calculated averaging the normalized node entropy

$$S = \frac{1}{N}\sum_{i=1}^{N} s^{(i)} = \frac{1}{N\ln(N-1)}\sum_{i=1}^{N} \ln k_i. \tag{14}$$

Let us provide some insights about the role of both, $s^{(i)}$ and S in the exploration of the topological aspects of the chemical hypergraph. Consider, for example, the case of a system with a discrete probability distribution $P = \{1/2, 1/2\}$. It is already known that the Shannon entropy for P, is given by $S_e[P] = \log(2)$, which, when the logarithm is written in base 2, gives just $S_e[P] = 1$. In this regard, the Shannon entropy is providing information about the number of bits, 1 in this case, within the system.

Consider now the expression for $s^{(i)}$. Using logarithm rules, it can be rewritten in the form

$$s^{(i)} = \log_{N-1} k_i, \tag{15}$$

hence indicating what we can call the number of *chemical bits* stored in the node (i) of the chemical hypergraph.

When using this approach in the expression for S, the interpretation is thus the mean number of *chemical bits* stored in the chemical hypergraph. Of course, this mean value is calculated per substance, because N is the number of substances in the chemical hypergraph.

To apply these entropic tools to the ChemHy case, we first need to introduce the degree of the hypernodes. This requires an additional structure. Let us begin by constructing the set $\mathcal{H}_{CH}$ of hypernodes in a ChemHy, defined as follows:

$$\mathcal{H}_{CH} = \{X : X \in E\}. \tag{16}$$

Then, the degree of a hypernode $X \in \mathcal{H}_{CH}$ is given as the map

$$k : \mathcal{H}_{CH} \to \mathbb{N}; X \mapsto k_X = \sum_{r \in E} \delta_{X \in r}, \tag{17}$$

where $\delta_{X \in r} = 1$ if the node X participates in the chemical reaction r (which means that it is counting the connection of one hypernode with the others). Consider now the hypernode cardinality which is given as the map

$$\lambda : \mathcal{H}_{CH} \to \mathbb{N}; X \mapsto \lambda_X = |X|. \tag{18}$$

We can now use these quantities to study the probability of transition from one hypernode (i) to another (j), thereby exploring both the hyperedge structure of the hypergraph and the internal structure of the hypernodes. Let us introduce two probabilities of transition between hypernodes. The first is the chemical probability of transition

$$P_{i \to j} = \frac{1}{k_i}, \tag{19}$$

which in this case, was considered as uniform among the hypernodes connected by a chemical reaction, hence the name.

The other one is the cardinality probability of transition

$$P_{i \to j} = \frac{1}{\lambda_j}, \tag{20}$$

which states that the transition depends on the cardinality of the neighboring hypernode j independently of the cardinality or connectivity of the hypernode i. The larger the number of substances, the lower the probability of visiting the hypernode j.

Figure 5 shows the behavior of the network entropy computed for several complete ChemHy models with different numbers of chemicals. It can be observed that in both cases, the entropies approach an asymptotic value. In the case where the probability depends on the cardinality of the destination node, a maximum is observed.

In the case of $1/k_i$, a global maximum is reached in the vicinity of $N \to 10$. In the case of $1/\lambda_j$, the supremum is lower than the maximum value of the black curve. This result indicates that indeed, the entropy of the path followed by the random walker is affected by the jump probability of the random walker. Selecting different probabilities for the jump,

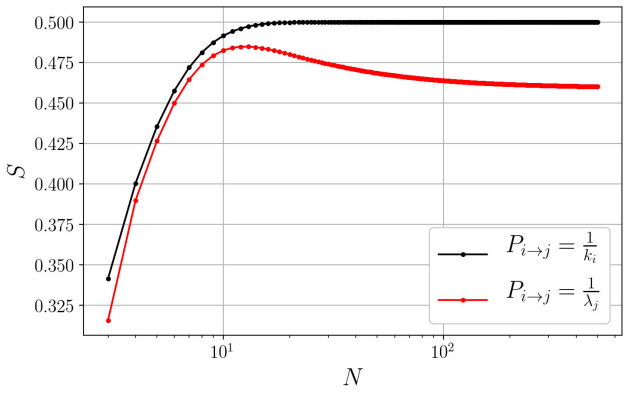

FIGURE 5. Network entropy computed for two different probability regimes for a random walker transition from hypernode i towards hypernode j. Black: probability depends on the degree of the initial hypernode (i). Red: probability depends on the size of the final hypernode (j).

will tell us different information about the topological structure of the network: selecting $1/k_i$ covers the network *faster* than selecting $1/\lambda_j$.

Chemically speaking, this finding suggests that the random walker's behavior is influenced by both the information contained within a hypervertex and its intrinsic properties, specifically, whether it is the origin or destination of the jump. Notably, knowing the size of the hypervertex to which the walker will jump leads to a different outcome than knowing the size of the hypervertex from which the jump originates. In the former case, the entropy is lower compared to the latter; in other words, there is less knowledge when the destination is unknown (*i.e.*, what is to be chemically obtained) than when it is already known. Moreover, a random walker covering the network with $1/\lambda_j$ provides information about a fraction of the *chemical qubits* of the network when compared with the *chemical qubits* accessible with $1/k_i$. This is a clear manifestation of how the entropy can be used to detect structural topological features of the ChemHy.

One might wonder how these patterns are affected when considering realistic chemical reaction sets, and whether these limiting values can be interpreted as bounds for realistic networks. Furthermore, how are they influenced by limiting the number of substances that can participate in a chemical reaction? These and other related questions can be explored within this framework.

Other options for the probability of transition worth considering are the following:

$$P_{i \to j} = \frac{1}{k_j}. \tag{21}$$

Here, the transitions depend on the number of connections of the neighboring hypernode j rather than on the hypernode i. Hypernodes that participate in a larger number of chemical reactions are less likely to be visited.

Another case is

$$P_{i \to j} = \frac{\beta}{\lambda_j}. \tag{22}$$

Here, the cardinality probability of transition is modified by a parameter β. This parameter can be used to model other effects on the probability of transition. For example, β could depend on the number of times a substance, say A, participates in chemical reactions (its degree). If *chemically relevant* substances are involved in the transition, the transition becomes more likely (with a larger β), allowing the study of the interplay between these two features: the connectivity of the hypernode and the degree of a substance. This approach is grounded in the fact that, historically, few substances appear in a large number of chemical reactions, whereas most substances participate in only a few reactions [10]. Other effects and considerations could be incorporated in a similar way using β.

Some open questions connected to these explorations are the following:

- Can the parameter β be adjusted to produce a distinct pattern formation?
- Is it possible to define a correlation measure between hypernodes?
- Exploring paths in a hypergraph that represents the chemical space is crucial for synthetic routes and design. What insights can entropy provide about the paths in the ChemHy?

Up to this point, the topological explorations have focused on general information about hypernodes, hyperedges, the number of substances present, and the number of hyperedges for each hypernode. However, we are also interested in more specific information about the hypernodes that could be chemically relevant. The purpose of the following sections is to discuss some ideas and proposals in this regard.

4. Information theory

Let X and Y be two discrete random variables with a joint probability distribution $P(X,Y)$. Mutual information [12, 13] is a special case of the Kullback-Leibler [14] distance and quantifies the statistical correlation between two variables. It is defined as:

$$I_{XY} = \sum_X \sum_Y P(X,Y) \ln \frac{P(X,Y)}{P(X)P(Y)} = s_X + s_Y - S_{XY} \geq 0, \tag{23}$$

with the joint Shannon entropy [12], S_{XY}, defined by

$$S_{XY} = -\sum_X \sum_Y P(X,Y) \ln P(X,Y), \tag{24}$$

and the marginal Shannon entropies, S_X and S_Y, by

$$S_X = -\sum_X P(X) \ln P(X), \tag{25}$$

$$S_Y = -\sum_Y P(Y) \ln P(Y), \tag{26}$$

where the marginal probability distributions are given by

$$P(X) = \sum_Y P(X,Y), \tag{27}$$

$$P(Y) = \sum_X P(X,Y). \tag{28}$$

Shannon entropies are interpreted as measures of uncertainty associated with the probability distribution. Its maximum value ($\ln n$ with n the number of events) is achieved when all events are uniformly distributed. Its minimum value (0) is obtained when only one event has probability 1 and the others have probability 0.

If X and Y are independent random variables, $I_{XY} = 0$. If X and Y are not independent, some information about X can be inferred from Y and viceversa, and $I_{XY} > 0$.

In order to clarify the above-mentioned notions, let us use an example of joint probability. Consider a dice (six outcomes: $1, 2, 3, 4, 5, 6$) and a coin (two outcomes: $0, 1$). Throw the dice and the coin and each time register both as a pair (this is one event), it is possible to obtain one of the twelve possible events:

$$\{(1,0),(2,0),(3,0),(4,0),(5,0),(6,0), (1,1),(2,1),(3,1),(4,1),(5,1),(6,1)\}.$$

In this case, both variables are independent and all have the same probability; hence, $S_{X,Y} = \ln 12$, $s_X = \ln 6$, $s_Y = \ln 2$, $I_{XY} = 0$.

Considering that any deviation from zero for mutual information is indicative of the statistical correlation between the variables, we wonder if some *chemical information* can be obtained from this correlation. For example, if the presence of a specific substance A influences the selection of another chemical to trigger a chemical reaction. In order to perform this analysis, we need to consider the adjacency matrix and focus our attention on the structure of the hypernodes. An example of this approach is presented in the next section. It should be mentioned that other possibilities can likewise be explored.

5. Chemically inspired random variables

A topological exploration of the **reaction structure** can be achieved by defining chemically inspired random variables for the chemical reactions presented in the adjacency matrix, as in the illustration of Fig. 6, where we present a specific example. In this case, the random variable X is built by classifying reactions depending on the number of reactants (N_R) and products (N_P) as follows:

$$X = \begin{cases} 0 \text{ if } N_R = N_P & \text{(exchange-type reactions)} \\ 1 \text{ if } N_R < N_P & \text{(decomposition-type reactions)} \\ 2 \text{ if } N_R > N_P & \text{(synthesis-type reactions)} \end{cases}, \tag{29}$$

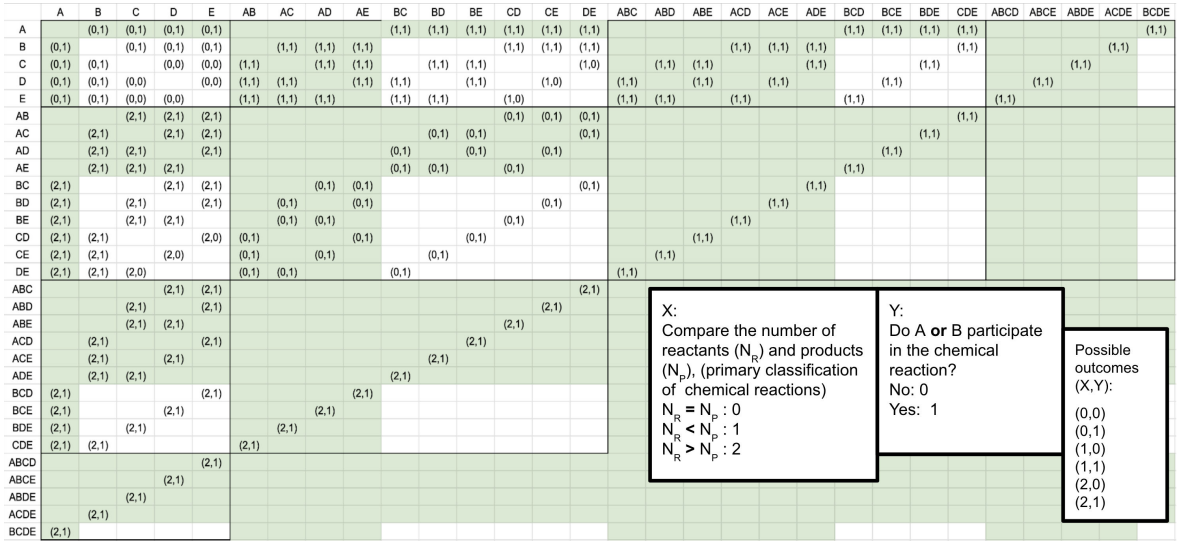

FIGURE 6. Definition of two specific random variables on the idealized chemistry with five substances.

The random variable Y contains information about the presence of specific substances:

$$Y = \begin{cases} 1 & \text{if A } \textbf{or} \text{ B are present in the reaction} \\ 0 & \text{if neither A nor B are present in the reaction} \end{cases} . \quad (30)$$

Note that random variables can be inspired by a variety of situations, for example, inquiring if one specific substance (A) participates in the chemical reaction, also if two (A **and** B), three (A **and** B **and** C), etc. The options (A **or** B), (A **or** B **or** C) can also be considered. Of course, as can be inferred, different combinations are possible.

It remains an open question which random variables can be used to explore the chemical information contained in the set of reactions under consideration, particularly in the case of realistic reaction sets. For instance, once a chemical is used by chemists in one reaction, it is reasonable to assume that they will test this chemical with other substances in pursuing similar reactions. Therefore, its presence could serve as a marker for specific chemical reactions.

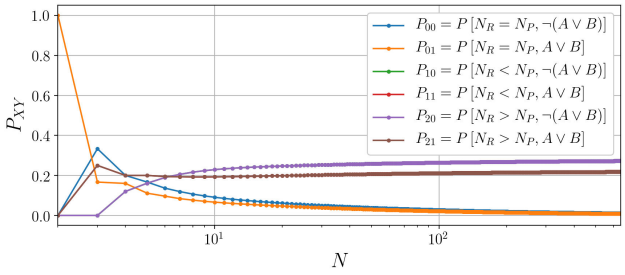

FIGURE 7. Joint probability distribution of the events $(X, Y) = (0,0), (0,1), (1,0), (1,1), (2,0), (2,1)$, for the random variables defined in the text. x-axis is the logarithm of the number of substances ($\log(N)$). Green (P_{10}) and violet (P_{20}) curves are superimposed due to the symmetry of the adjacency matrix in the complete chemistry model, the same happens between the red (P_{11}) and brown (P_{21}) curves.

In the example presented here, a joint event (X, Y) is associated with each valid reaction, and the six possible outcomes are $(0,0), (0,1), (1,0), (1,1), (2,0), (2,1)$. These outcomes are not equiprobable, and the probability of occurrence depends on the number of substances considered.

Figure 7 shows the joint probability distribution of random variables X and Y. It is important to note that there are bounds for the probabilities. At small N, there are variations in the order of the probabilities, which can impact the results, particularly when dealing with sets of reactions involving a small number of substances. For example, this could be relevant when applying the model to a specific metabolic pathway.

Figure 8 shows the Shannon entropy of the joint distribution. The maximum observed for small N reflects the variations and intersections between the probabilities depicted in Fig. 7 (the more balanced the probabilities, the higher the entropy). It is important to note that there appears to be an entropy bound at higher values of N. How does this bound change if we define another chemically inspired random variable, as discussed earlier in the text?

Figure 9 shows the marginal probabilities of the random variable X, while those of Y are omitted for the sake of brevity. Notably, the probabilities tend toward an asymptotic value after a certain crossing point.

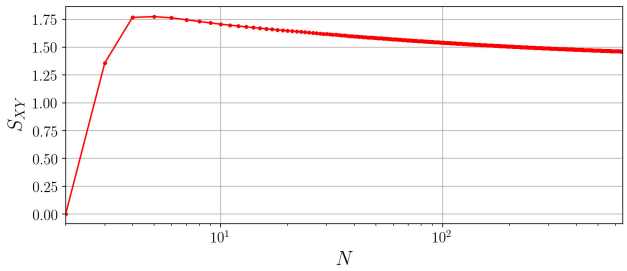

FIGURE 8. Shannon entropy of the joint probability distribution depicted in Fig. 7. x-axis is in log-scale.

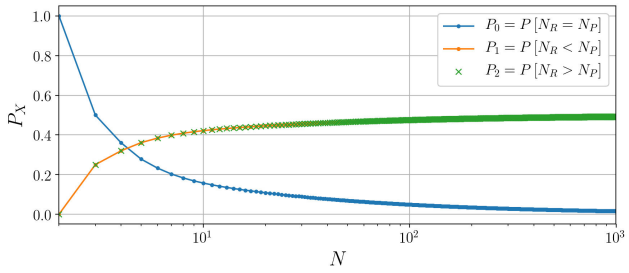

FIGURE 9. Marginal probability distribution of the random variable X defined in the text. The x-axis is in log-scale.

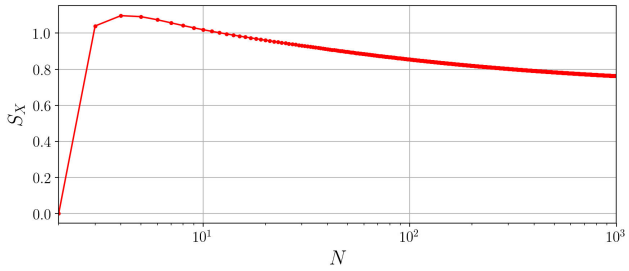

FIGURE 10. Shannon entropy of the marginal probability distribution depicted in Fig. 9. The x-axis is in log-scale.

Figure 10 shows the Shannon entropy of the probability distribution depicted in Fig. 9. The entropy reaches a maximum due to the observed crossing point in the probabilities and also tends toward an asymptotic value. How these trends are influenced by the definition of the random variables, and what chemical information can be inferred from the interpretation of such bounds, are two open questions in this analysis.

Mutual information is shown in Fig. 11, where the monotonic decreasing behavior indicates that the information one variable provides about the other diminishes as the number of substances increases. In other words, as the number of substances grows, it becomes harder to associate the presence of specific substances (*e.g.*, A, B) with the type of chemical reaction (*e.g.*, exchange type, decomposition type, or synthesis type). This raises the question of how these two random variables correlate in realistic sets of chemical reactions, as certain substances are specific to synthesis reactions and others to decomposition reactions. How is mutual information affected in these more realistic scenarios, and can it be used to characterize specific regions of the chemical hypergraph, *i.e.*, particular sets of chemical reactions?

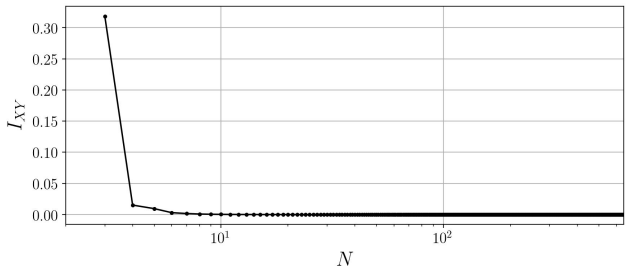

FIGURE 11. Mutual information between random variables X and Y. x-axis is in log-scale.

Variations in mutual information and other entropic measures were observed under different choices of random variables. However, these findings are not included, as the focus of the present manuscript is to outline the initial concepts underlying these topological approaches. Future research will aim to elucidate the chemical significance of these behaviors.

There are many open questions related to these topics. Two of the most interesting ones are whether these quantities can be used to detect pattern formation in chemical hypergraphs and whether they can help distinguish between different types of patterns, such as cycles (closed regions) formed within the chemical hypergraph.

6. Conclusions

Information-theoretical tools seem to be reliable in performing topological explorations on chemical hypergraphs.

Different transition probability regimes can be explored through network entropy. These probabilities can be designed to account for various features of hypernodes (sets of substances), allowing for the representation of realistic chemical reaction sets. In this work, we show that different regimes lead to distinct entropy behaviors. Further research is needed to interpret the bounds of network entropy and the occurrence of maxima in transition probability, which depend on the cardinality of one of the hypernodes.

In the same spirit, parametric definitions of probabilities can be used to test other characteristics of reaction networks.

Chemically inspired random variables were introduced to explore different internal hypernode structures. In this work, a specific example is presented to illustrate this approach. Various chemical properties can be incorporated into this framework, such as the presence of specific substances and their relationships with other features of the hypergraph. Detecting topological patterns is one of the most intriguing questions raised by this exploration.

Several open questions emerged from the analyses conducted in this work. In fact, unanswered questions outnumber the resolved ones. Those related to pattern detection could be of interest to various scientific communities and will be a focus of future research.

It remains an open question whether the approach developed in this work, grounded in the internal structures of hypernodes and hyperedges, can be used alongside existing models of chemical reaction networks to more effectively capture chemical phenomena.

One of our main interests is the nature of the random variables used in these explorations. In this work, the choice of variables preceded the analysis of the data represented in the chemical hypergraph. Our goal is to develop models that allow the data to "speak" and reveal patterns, helping us identify the most informative random variables. This approach could be highly useful for detecting biases in the data.

Acknowledgements

H.G.L. acknowledges the Dirección de Apoyo a la Investigación de la Universidad Autónoma Metropolitana (DAI-UAM) for financial support. A.G.-C. acknowledges CONAHCyT for a postdoctoral fellowship.

---

i. In this work, the notion of chemical space refers to the set of substances and the set of chemical reactions between these substances.

ii. The number of edges connected to this node.

1. N. Biggs, *Algebraic graph theory*, (Cambridge university press, 1993).
2. M. Barthélemy, *Spatial networks*, *Physics reports*, **499** (2011) 1, https://doi.org/10.1016/j.physrep.2010.11.002
3. M. E. J. Newman, Networks An Introduction (Oxford University Press, 2010)
4. A. Bretto, Hypergraph Theory An Introduction (Springer Cham, 2013), https://doi.org/10.1007/978-3-319-00080-0.
5. A. García-Chung, M. Bermúdez-Montaña, P. F. Stadler, J. Jost, and G. Restrepo, Chemically inspired Erdös-Rényi oriented hypergraphs, *J. Math. Chem.* **62** (2024) 1357, https://doi.org/10.1007/s10910-024-01595-8
6. J. Jost and R. Mulas, Hypergraph Laplace operators for chemical reaction networks, *Adv. Math.* **351** (2019) 870, https://doi.org/10.1016/j.aim.2019.05.025
7. F. Betancourt-Moreno, *et. al.*, Oriented Erdõs-Rényi Hypergraphs: A Computational Analysis. In progress.
8. J. L. Andersen, S. Banke, R. Fagerberg, C. Flamm, D. Merkle, and P. F. Stadler, Pathway Realizability in Chemical Networks, *J. Comp. Bio.* **32** (2025) 164, https://doi.org/10.1089/cmb.2024.0521
9. K. Molga, S. Szymkuc, and B. A. Grzybowski, Chemist ex machina: advanced synthesis planning by computers, *Acc. Chem. Res.*, **54** (2021) 1094, https://doi.org/10.1021/acs.accounts.0c00714
10. E. J. Llanos, W. Leal, D. H. Luu, J. Jost, P. F. Stadler, and G. Restrepo, Exploration of the chemical space and its three historical regimes, *Proc. Natl. Acad. Sci. U.S.A.* **116** (2019) 12660, https://doi.org/10.1073/pnas.1816039116
11. C. G. S. Freitas, A. L. L. Aquino, H. S. Ramos, A. C. Frery and O. A. Rosso, A detailed characterization of complex networks using Information Theory, *Sci. Rep.* **9** (2019) 16689, https://doi.org/10.1038/s41598-019-53167-5
12. C. E. Shannon, A mathematical theory of communication, *Bell Syst. Tech. J.* **27** (1948) 379, https://doi.org/10.1002/j.1538-7305.1948.tb01338.x
13. T. M. Cover, J. A. Thomas, *Elements of Information Theory*, (Wiley, New York, 1991).
14. S. Kullback, R. A. Leibler, On information and sufficiency, *Ann. Math. Stat.* **22** (1951) 79, https://doi.org/10.1214/aoms/1177729694