

POLITECNICO DI TORINO

Master's Degree in Physics of Complex Systems



**Politecnico
di Torino**

Master's Degree Thesis

Characterization Measurements of Metabolic Hypergraphs

Supervisors

Prof. Yamir MORENO

Dott. Guilherme F.D. ARRUDA

Candidate

Pietro TRAVERSA

October 2022

Abstract

Metabolic networks are probably among the most challenging and promising biological networks. Their study provides insight into how biological pathways work and how robust a specific organism is against an environment or therapy. Previous studies have obtained relevant results using flux balance analysis (FBA) and simulations of single gene deletion. However, the structural characterization of metabolic networks as complex networks has been proven an arduous task. Past attempts have considered graphs whose nodes are the metabolites or reactions of the metabolic network in question, and only recently has the focus shifted to higher-order structures, highlighting that simple pairwise interaction may not be sufficient for characterization. Here we show an intuitive way to map metabolic networks into hypergraphs using the bipartite representation. We introduce structural characterization measurements to analyze the metabolism of a single organism and to compare the robustness and complexity of different metabolic models. Our findings show a connection between topological and biological properties. Communicability and information-based measurements succeed in identifying relevant metabolites and reactions in the metabolic hypergraph, and different organisms exhibit an overall similar complexity but very different robustness. In particular, the *Staphylococcus aureus*, an antibiotic-resistant bacterium, displays the highest robustness.

Acknowledgements

First of all, I would like to sincerely thank my supervisors Yamir Moreno and Guilherme Ferraz de Arruda for following me through this journey and helping me to complete this thesis.

I thank Yamir's entire research team, who have been there for me as friends before being colleagues. I thank catalan Ari, for the company, laughter and advice as a biologist. I thank Henrique for being himself. Also, I thank Guilherme again since he always thanks people at least twice.

Un ringraziamento speciale va a Elena, compagna di questo tirocinio e cara amica. Il tuo supporto ha significato molto.

Ringrazio con gioia i miei amici di borgata, di una vita e acquisiti.

Una menzione speciale ai miei cani, Macchia, Sophie e Lars.

Ringrazio di cuore la mia famiglia. Mi avete sempre sostenuto senza alcuna esitazione.

Per concludere, Ari, il tuo contributo a questo percorso è stato molto di più di quel che credi. Non voglio neanche immaginare quanto vuoti sarebbero stati questi anni se non fossi stata al mio fianco. Sei una ispirazione, una speranza e una certezza. Grazie davvero, tvb.

Table of Contents

List of Figures	v
Introduction	1
1 Hypergraphs characterization	4
1.1 Hypergraphs definition	4
1.2 Hypergraphs representations	5
1.2.1 The Incidence Matrix	5
1.2.2 Hypergraph projections	5
1.2.3 Bipartite representation	7
1.3 Random Walks	10
1.3.1 Definitions on hypergraphs	10
1.3.2 Formulation and analysis	11
1.4 Laplacians	13
1.4.1 Definition on graphs	13
1.4.2 Generalization to Hypergraphs	15
1.5 Characterization measurements	16
1.5.1 Communicability	16
1.5.2 Information on graphs and hypergraphs	18
2 Metabolic Hypergraphs	21
2.1 Introduction to Metabolic Hypergraph	21
2.2 Mapping stoichiometry matrices into hypergraphs	22
2.3 The dataset and preprocessing	24
2.4 Communicability on Escherichia coli	24
2.5 Information on Escherichia coli	26
2.6 Comparison with Flux Balance Analysis	27
2.7 Comparing metabolic networks	28
2.7.1 Results	28

3 Random metabolic networks	31
3.1 Shuffle reactions vs shuffle edges	33
Conclusions	35
A Algebraic Connectivity	37
B Random metabolic networks: extra results	39
Bibliography	44
Funders	51

List of Figures

1.1	An example of two different hypergraphs A) and B) with the same counting adjacency matrix. The hypergraph in A) has hyperedges $E = \{(1,2), (2,4), (0,1,2,3)\}$ and the hypergraph in B) $E = \{(0,1,2), (2,4), (1,2,3), (0,3)\}$	7
1.2	An example of two different ways to represent hypergraphs. In the bipartite representation, vertices and hyperedges are the two partitions of the bipartite graph. An edge between a vertex and a hyperedge is present if and only if such vertex belongs to such hyperedge. In the clique representation, the hypergraph is projected onto the vertices, such that vertices belonging to the same hyperedge form a clique.	9
2.1	PFK Reaction from BiGG dataset: $atp_c + f6p_c \rightarrow adp_c + fdp_c + h_c$. In green the reagents, in blue the products, and in pink the reaction. Directionality is easily taken into account with this definition.	22
2.2	An example of a metabolic network mapped into a hypergraph with edge-dependent vertex weight. In (A), we present a small network composed of three reactions and five metabolites. The first reaction r_1 is reversible and is represented with the double arrow. In (B), we show the corresponding stoichiometry matrix. Reactants are negative and products are positive. We need to split the reversible reaction into two irreversible reactions r_1^+ and r_1^- to write it in matrix form. This stoichiometry matrix is the incidence matrix of the hypergraph with edge-dependent vertex weights shown in (C). For the sake of visualization, only the hyperedge r_1^+ is shown. The hyperedge r_1^- is just the same but with the opposite sign. Note that weights are both positive and negative, meaning that the hypergraph is directed. Indeed, we separate the head and tail of each hyperedge with a dashed line.	23

2.3	E. Coli metabolic hypergraph in the bipartite representation. Nodes are metabolites and reactions. The color of each node is proportional to the average communicability of that node, and the size is proportional to the node degree.	25
2.4	Access information vs Hide information for the metabolites (A) and reactions (B). Metabolites in blue belong to the cytosol compartment, and the ones in green belong to the extracellular space. The latter have, on average, higher access and hide information compared with the former since they belong to the outermost of the network. Each reaction is colored differently depending on the metabolic pathway to which they belong. We do not observe a separation between pathways in the Access vs Hide information plane, meaning that different pathways have, on average, the same complexity in <i>E.coli</i> . The nodes with the higher degree, h_c and <i>Biomass</i> , are highlighted with a red cross. Some nodes appear to have zero access or hide information. Those nodes cannot reach or be reached by other nodes.	26
2.5	Plot of the reactions average communicability. In purple is represented the average communicability of quasi-essential reactions, in orange the average communicability for essential reactions, and in blue for non-essential reactions	29
2.6	Average search information, in blue, and natural connectivity, in orange, computed for different BiGG models. The organisms analyzed range from simple bacteria to humans. The results show similar complexity among species, but a substantial difference in robustness. Note that the peak in robustness is obtained for the <i>Staphylococcus Aureus</i> . The models are ordered in increasing number of nodes. . .	30
3.1	Comparison between the robustness of the real metabolic hypergraph and of the randomized one. The obtained value is an average of 5 repetitions of the randomization procedure. We consider a subset of the most interesting metabolic hypergraph.	34
A.1	The algebraic connectivity of the hypergraph on the left ax, in yellow, and the robustness on the right ax, in red.	38
B.1	The average search information for some metabolic hypergraphs compared with the average search information of their randomized counterparts. The method used to randomize the hypergraph is <i>shuffle reactions</i> . We did $100 \times M$ shuffles, where M is the number of reactions.	40

B.2	The robustness for some metabolic hypergraphs compared with the robustness of their randomized counterparts. The method used to randomize the hypergraph is <i>shuffle reactions</i> . We did $100 \times M$ shuffles, where M is the number of reactions.	41
B.3	The average search information for some metabolic hypergraphs compared with the average search information of their randomized counterparts. The method used to randomize the hypergraph is <i>shuffle edges</i> . We did 5 repetitions and $100 \times M$ shuffles, where M is the number of reactions.	42

Introduction

Complex systems are ubiquitous in our everyday lives [1], which is why they are attracting more and more interest, from the study of ecological and social systems to economics and spin systems. All these systems have a large number of interactions among their components. They may be simple exchanges of messages, economic transactions, or complicated ferromagnetic interactions. Given the variety of the problem, it is helpful to abstract the concept of interaction and to represent it mathematically in a proper structure, a graph. With these motivations, complex systems and network theory have become concepts that often travel alongside [2–4]. Through the study of graphs, physics has achieved excellent results in different areas like spins models, phase transition theory, and condensed matter, as well as brought about the emergence of a thriving and ever-growing field such as neural networks [5]. Dynamic contagion systems such as the SIS model and its generalizations have been studied on social networks, a topic that has never been more important at this time in history.

However, in network sciences, only pairwise interactions are considered. This is a limitation when considering phenomena in which higher-order interactions may be present, such as in the social sciences [6–16], epidemiology [11, 14, 17–19], biology [20–26], etc. It has been shown that higher-order interactions are essential to the understanding of such phenomena [11, 25, 27–31], and thus it has become increasingly important to develop mathematical objects capable of describing these interactions [32]. One possible solution is to consider a graph whose edges are capable of connecting multiple vertices at the same time. Such a generalization of a graph is called a hypergraph [33]. It has potentially no structural limits and accounts naturally for higher-order interactions. In recent studies, different ways to represent hypergraphs have been introduced depending on the problem. One popular way of representing a hypergraph is the so-called projected representation, in which the nodes belonging to a hyperedge are represented as a clique on a graph [32, 34, 35]. However, this representation may not be sensitive to some local hypergraph structures and is especially not general. In [36], they show that for a hypergraph with edge-dependent vertex weights (EDVW), there is no clique projection, highlighting the structural limitation of this representation. Here

we propose an alternative and general way of representing hypergraphs, which projects them onto a bipartite graph. We show that this method is convenient and intuitive and potentially has no structural constraints. Even the EDVW hypergraph proposed in [36] can be represented in this way and as oriented and directed hypergraphs. The latter, in particular, are types of hypergraphs that have been little studied in the literature since the undirected counterpart is simpler but is no less important. Included in this category are metabolic networks [37, 38], which we study here, but also task or supply chains and message spreading, to give a few examples.

Of all the possible applications, here, we focus on metabolic networks. A metabolic network [39] is a set of biological processes that determines the properties of the cell. Several reactions are involved in metabolism, grouped into various metabolic pathways. A metabolic pathway is an ordered chain of reactions in which metabolites are converted into other metabolites or energy. For example, the glycolysis pathway is the set of reactions involved in transforming one glucose molecule into two pyruvate molecules, producing energy. Metabolic networks are among the most challenging and highest potential biological networks [40–42]. The way to represent a metabolic network on a graph is not unique, and several approaches have been tried. One possible way is to consider metabolites (or reactions) as nodes and connect them if and only if they share a reaction (or metabolite). The resulting graph is undirected, and this may unpleasantly change the structural properties of the network. In [43], they analyze the same dataset we examine here for *E.coli* and propose a directed graph with reactions as nodes that take into account the directionality of the reactions, highlighting the difference with the undirected counterparts. Here, however, we define the metabolic hypergraph as a hypergraph whose nodes are the metabolites and reactions are the hyperedges.

Our work contributes from a mathematical point of view to explore and compare different representations of hypergraphs and the random walks defined on them. In particular, we include directed and oriented hypergraphs in the discussion and provide a possible representation of them. We also address the problem of defining a Laplacian matrix on hypergraphs, which is very much related to the concept of random walks and diffusion, emphasizing how it becomes ill-defined on directed hypergraphs and how, although attempts have been made, at present there seems to be no clear answer in the literature.

In addition, we define and introduce characterization measures based on random walks and information theory for all the representations considered. This approach allows analysis at multiple scales: at the local scale, the access and hide information or communicability; at the global scale, the natural connectivity and average search information; and at the mesoscopic scale, by analyzing differences in pathways. From a physical point of view, we establish a connection between these types of measures and the biological properties of the studied organisms. Our findings shows

a partial connection between structurally important reactions from the network point of view and essential reactions for the metabolism. There are some aspects that we cannot capture with these measurements, like external reactions that are important for metabolism. These are leaves in the hypergraph and are considered irrelevant from the structural point of view. In addition, a global analysis of the metabolic hypergraph allowed us to compare different organism's metabolism and connect their structural organization with biology arguments. The complexity seems comparable among individuals, and this result might be motivated by evolution. We also find that an antibiotic-resistant bacterium displays the highest robustness among the organisms studied. This is a great result, but we are aware that it is still an isolated case, and more studies on more data should be carried on to be sure that such a correlation exists.

We emphasize that the representations and measurements described above are entirely general and can be applied to any type of hypergraph. In this sense, metabolic hypergraphs are a choice, not a necessity. In any case, we show a one-to-one mapping between hypergraphs and metabolic networks. This means that in using this mathematical model, no information is lost. Moreover, again as a consequence of this mapping, metabolic hypergraphs are a natural example of an EDVW hypergraph. This is relevant because it is challenging to find hypergraph data online that possess this structure and this variety of weights. In practice, weights are often arbitrary chosen by the author. Therefore, even if researchers are not directly interested in metabolic networks, they could use them as a real dataset for studying EDVW hypergraphs.

The work is organized as follows. In chapter 1, we present the methods used to characterize hypergraphs. Specifically, in subsection 1.1, we first define hypergraphs from a mathematical point of view, and in section 1.2, we compare several representations of them. In section 1.3, we explore different types of random walks for each representation and explain why we believe the bipartite representation is the best. Connecting from the random walks, in section 1.4, we generalize the Laplacian matrix for hypergraphs using the bipartite representation, comparing it with other Laplacians defined in the literature. Then in section 1.5, we introduce metrics based on random walks, such as communicability and search information. In chapter 2, we apply these measures first to the *E.coli* core metabolism and then to different metabolic networks to compare them. In chapter 3, we start exploring the possibility of defining a null model of metabolic hypergraphs. Finally, we conclude by summarizing our results and presenting possible future works.

Chapter 1

Hypergraphs characterization

1.1 Hypergraphs definition

A hypergraph $H = \{V, E\}$ is a set of vertices or nodes $v \in V$ and hyperedges $e \in E$. Each hyperedge is a subset of V such that different nodes interact with each other if and only if they belong to the same hyperedge. Thus a hyperedge represents a multiple-node interaction. If the dimension $|e|$ of the hyperedge is 2, then the hypergraph is equivalent to a normal graph. In this thesis, we will denote by $N = |V|$ the number of vertices of the hypergraph and by $M = |E|$ the number of hyperedges.

There are different types of hypergraphs. A *weighted hypergraph* $H = \{V, E, W\}$ is a set of vertices or nodes $v \in V$, hyperedges $e \in E$, and edge weights $w(e)$. If $w(e) = 1 \forall e \in E$, then the hypergraph is said to be unweighted. A *weighted hypergraph with edge-dependent vertex weights* (EDVW) $H = \{V, E, W, \Gamma\}$ is a set of vertices or nodes $v \in V$, hyperedges $e \in E$, edge weights $w(e)$ and edge-dependent vertex weight $\gamma_e(v)$. If $\gamma_e(v) = \gamma(v) \forall e \in E$, then the hypergraph is said to have *edge-independent* vertex weight. All the weights are assumed to be positive.

A hypergraph can be oriented, directed, or undirected. An oriented hypergraph is a hypergraph in which hyperedges have an orientation. We can identify the head $H(e_j)$ and tail $T(e_j)$ of the hyperedge e_j . A vertex v_i can belong only to the head or the tail of the hyperedge e_j , not both. A directed hypergraph is similar to an oriented hypergraph, but vertices can belong to both the head and the tail. Each oriented hypergraph is directed, and each directed hypergraph can be made oriented by adding additional hyperedges whenever a vertex belongs to both the head and tail.

It is also helpful to define k_v , the degree of a vertex $v \in V$, as the number of

hyperedges in which v is contained. Finally, we also define $|e|$, the cardinality, sometimes also called the degree, of the hyperedges $e \in E$, as the number of vertices belonging to e .

To be more practical, hypergraphs can be viewed as sets of sets, where sets are hyperedges and the element they contain are the nodes of the hypergraph. Indeed, using intersecting sets is a common way to visualize hypergraphs, probably the best. However, in practice, it is tough to draw hypergraphs with a vast number of hyperedges and nodes using sets, so it can be useful to present other possible representations of hypergraphs.

1.2 Hypergraphs representations

A possible hypergraph representation is a projection of a hypergraph onto a graph such that an adjacency matrix can be defined. This is practically convenient, allowing to generalize concepts from graph theory to hypergraph. Unfortunately, we show that there are different ways to project a hypergraph and that the projection operation can cause a loss of information. Also, it is important to underline that different types of representation may correspond to different types of processes.

1.2.1 The Incidence Matrix

Given an undirected hypergraph $H = \{V, E\}$ of N vertices and M hyperedges, the incidence matrix is the matrix $\mathcal{I} \in \mathbb{R}^{N \times M}$ such that

$$\mathcal{I}_{ij} = \begin{cases} 1 & \text{if } v_i \in e_j \\ 0 & \text{if } v_i \notin e_j \end{cases}. \quad (1.1)$$

Given an oriented hypergraph $H = \{V, E\}$ of N vertices and M hyperedges, the incidence matrix is the matrix $\mathcal{I} \in \mathbb{R}^{N \times M}$ such that:

$$\mathcal{I}_{ij} = \begin{cases} 1 & \text{if } v_i \in H(e_j) \\ -1 & \text{if } v_i \in T(e_j) \\ 0 & \text{if } v_i \notin e_j \end{cases}, \quad (1.2)$$

where $H(e_j)$ and $T(e_j)$ are, respectively, the head and the tail of the hyperedges e_j .

1.2.2 Hypergraph projections

Since hyperedges account for multi-body interactions, a natural and common way to represent them is as cliques in a graph. A clique of a graph is a fully connected subgraph. Therefore, the meaning of this representation is to connect

every node belonging to the same hyperedge. The adjacency matrix of a so-projected hypergraph can be defined as

$$\mathbf{A}^{count} = \mathcal{I}\mathcal{I}^T - \mathbf{D}, \quad (1.3)$$

where \mathbf{D} is the diagonal matrix whose k -element is the degree of node k . The element A_{ij} is the number of hyperedges shared by vertex i and vertex j . As a consequence, all nodes in a hyperedge are connected, i.e., they form a clique. We call this adjacency matrix the *counting adjacency matrix* [32, 34]. Another possible way to define the adjacency matrix is by normalizing each entry by the size of the hyperedges it corresponds to [35], thus

$$\mathbf{A}^{norm}_{ik} = \sum_{\substack{e_j \in E \\ v_i \neq v_k \in e_j}} \frac{1}{|e_j| - 1}, \quad (1.4)$$

which can be rewritten in matrix form as

$$\mathbf{A}^{norm} = \mathcal{I} \mathbf{E}^{-1} \mathcal{I}^T - D,$$

where $\mathbf{E} = \text{diag}(|e_i| - 1)$ and D is a diagonal matrix such that the diagonal entries of \mathbf{A}^{norm} are zeros¹. We refer to this matrix as the *normalized adjacency matrix* [35]. It is useful to highlight that in this representation, the degree can be obtained as $k_i = \sum_{k=1}^N \mathbf{A}^{norm}_{ik}$. These two representations are similar, but they describe different processes. Note that the two above-mentioned adjacency matrices are symmetric, meaning that the projected graph is undirected. It has been shown by *Citra et al.* [36] that a hypergraph with edge-dependent vertex weight cannot be represented as a clique graph. The intuition behind this statement is that if a vertex has different weights depending on which hyperedges it belongs to, this information is lost when we project on the vertices. Therefore, these representations are not general and an alternative way to represent hypergraphs is needed.

Another situation in which these projections fail is when hyperedges are contained in other bigger hyperedges. Since a k -clique is also a q -clique, with $q < k$, the smallest hyperedges are absorbed into bigger hyperedges when projected. This situation often occurs in social systems or WhatsApp groups, where private and group chats exist. One may argue that the information about the smallest hyperedges incorporated into bigger ones is in the weights of the projected edges. Consider, for example, the hypergraph A) shown in Fig. 1.1. In the projected graph, using the counting adjacency matrix, the edge connecting node 1 and node 2 has weight $w_{1,2} = 2$ because they share two hyperedges. Nevertheless, at least

¹This is to avoid lazy random walks, which correspond to self loops in the projected graph.

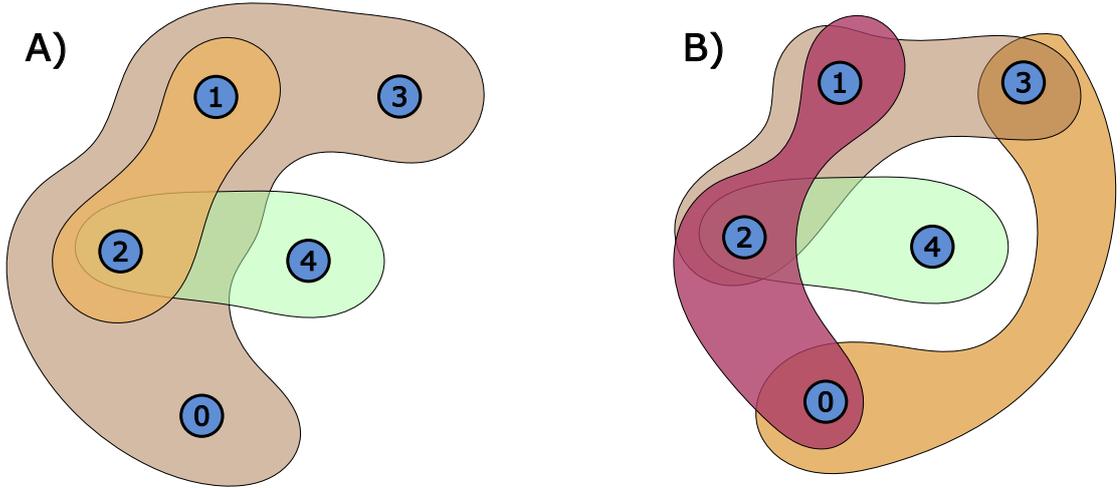


Figure 1.1: An example of two different hypergraphs A) and B) with the same counting adjacency matrix. The hypergraph in A) has hyperedges $E = \{(1,2), (2,4), (0,1,2,3)\}$ and the hypergraph in B) $E = \{(0,1,2), (2,4), (1,2,3), (0,3)\}$.

another hypergraph has the same adjacency matrix. Indeed take the hypergraph B), and one can show that the counting adjacency matrix is the same, but the structure is very different. In B), there is an extra hyperedge, and the nodes have a different degree than in A). This suggests that the normalized adjacency distinguishes between the two. Note that instead, the incidence matrix is always unique.

The fact that hypergraphs cannot be projected on graphs without losing some information indicates that they are more complex and rich than them. Still, having a way to project them helps visualize them, and it represents a good starting point to generalize the theory of networks.

1.2.3 Bipartite representation

The projection of a hypergraph $H(V, E)$ on a bipartite graph is the graph, sometimes called *the incidence graph*, $\mathcal{G} = \{V \cup E, \mathcal{E} \subset V \times E\}$, where V and E are respectively the set of all nodes and all hyperedges of the hypergraph H . An edge $(i, j) \in \mathcal{E}$ if and only if $v_i \in e_j$. The adjacency matrix of such \mathcal{G} is

$$\mathbf{A}^{bipartite} = \begin{pmatrix} \mathbf{0} & \mathcal{I} \\ \mathcal{I}^T & \mathbf{0} \end{pmatrix}, \quad (1.5)$$

where \mathcal{I} is the incidence matrix of the hypergraph H , as in Eq.(1.1). In this case, the hypergraph is not projected onto the vertices nor onto the hyperedges. This allows much more freedom in dealing with different processes while maintaining all

the information that was also contained in the previous representations. Indeed, it is easy to obtain the graph projected to the vertices or the edges from the bipartite graph. Note that since the incidence matrix is unique, the bipartite adjacency matrix is also unique.

The matrix $(\mathbf{A}^{bipartite})^2$ is a diagonal block matrix. The upper block is the adjacency matrix of the hypergraph projected on the vertices²; the bottom block is the adjacency matrix of the hypergraph projected on the hyperedges. See Fig. 1.2 for an example of the two representations.

The above representations regard undirected hypergraphs. In the case of oriented or directed hypergraphs, one can still define the adjacency matrix as \mathbf{A}^{count} as in Eq.(1.3), but the directionality is lost. Nevertheless, this can still be useful. For example, *Mulas and coworkers* [45, 46] used this adjacency to define the Hypergraph Laplacian operator. In this way, the operator is symmetrical and easier to deal with. They studied its spectral properties and applied them to chemical reaction networks.

In the case of an oriented Hypergraph, one can notice that the incidence matrix in Eq.(1.2) can be divided into

$$\mathcal{I} = \mathcal{I}_B - \mathcal{I}_F, \tag{1.6}$$

where \mathcal{I}_B and \mathcal{I}_F are matrices of positive entries [38]. The adjacency matrix of the clique graph can then be defined as

$$\mathbf{A}^{count} = \mathcal{I}_F \mathcal{I}_B^T, \tag{1.7}$$

and the adjacency matrix of the bipartite representation as

$$\mathbf{A}^{bipartite} = \begin{pmatrix} \mathbf{0} & \mathcal{I}_F \\ \mathcal{I}_B^T & \mathbf{0} \end{pmatrix}. \tag{1.8}$$

The meaning of the bipartite representation is very intuitive. The graph has a directed edge from the vertices to the hyperedges of the hypergraph when the vertices belong to the *tail* of the hyperedge, and a directed edge from the hyperedges to the vertices when the vertices belong to the *head* of the hyperedge. In this way, orientation is accounted for in a simple and natural way. For directed hypergraph, one can use the same construction as above, with the only difference that now vertices can belong to both head and tail of a hyperedge.

We conclude by making some considerations. First, notice that we pay a cost in representing the hypergraph as bipartite. The size of the projected graph, i.e.,

²The exception is the diagonal. This is equivalent to say that we allow self-loops, i.e., lazy random walks. Otherwise, a different bipartite projection is needed, see [44].

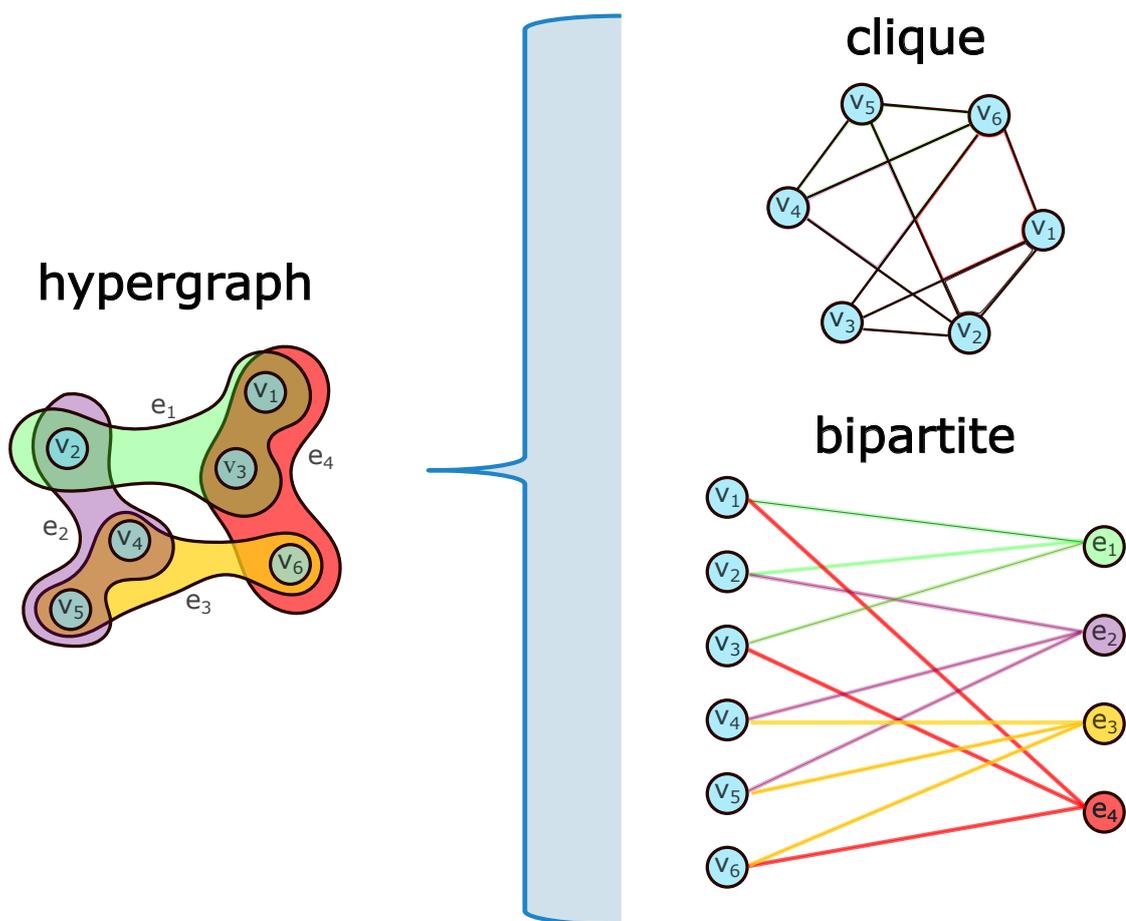


Figure 1.2: An example of two different ways to represent hypergraphs. In the bipartite representation, vertices and hyperedges are the two partitions of the bipartite graph. An edge between a vertex and a hyperedge is present if and only if such vertex belongs to such hyperedge. In the clique representation, the hypergraph is projected onto the vertices, such that vertices belonging to the same hyperedge form a clique.

its number of nodes, is $N + M$, much bigger than the size of the hypergraph, just N . Additionally, a bipartite projection corresponds to two hypergraphs. This is because of the symmetry hyperedges-nodes: a hyperedge is a set that links different nodes, but equivalently one may consider hyperedges as connected if they share at least one node. This is equivalent to saying that given a connected bipartite graph, one can always find a unique partition of nodes in two sets, but these two sets are symmetric and exchangeable. In conclusion, a hyperedge is more flexible than an edge to describe certain types of processes such as critical mass, that is, a hyperedge that becomes active when it contains a certain amount of active nodes.

However, as far as this thesis is concerned, we do not consider such processes.

1.3 Random Walks

In this section, we define different types of random walks on hypergraphs. The walk on the bipartite representation is the most intuitive one and very general.

1.3.1 Definitions on hypergraphs

A walk $w_{v_i \rightarrow v_f}$ on a hypergraph H is a sequence of tuples of vertices and hyperedges $(v_i, e_{i+1})(v_{i+1}, e_{i+2}) \dots (v_{f-1}, e_f)(e_f, v_f)$ such that $v_{i-1}, v_i \in e_i$, meaning that to move from one node to another the walker needs to travel through a hyperedge containing both nodes. So a walker that at time i is located in node v_{i-1} can go to v_i through hyperedges e_i if and only if both nodes belong to e_i . A *step* is a walk of length one.

A random walk is defined as a walk in which each step is chosen accordingly to some probability. It is thus a *stochastic process*. We are interested in *Markovian random walks*, a type of random walk such that the next step of the walk is defined based only on the current state of the process. Practically, the typical process we will consider for a random walk at time t at vertex v_t is:

- pick an edge $e \in N(v_t)$ with some probability $p_{v_t}(e)$,
- pick a vertex $v \in e$ with some probability $p_e(v)$,
- move to $v_{t+1} = v$ at time $t + 1$.

Where $N(v_t)$ is the neighborhood of v_t , i.e. the set of hyperedges containing v_t . The choice of the probability is not unique, and the way one defines the transition matrix defines the random walk. This procedure results in a two-step process and characterizes a random walk that moves on the vertices. It is simple to write the process for a random walk that moves on the hyperedges using a similar argument. Of course, one can also consider a random walk that walks alternatively both on the vertices and the hyperedges. In this case, the process will be a one-step process, exactly as on a regular graph, and the walk is no more a sequence of tuples but a sequence of nodes. This may not seem interesting from a physical point of view, but it is the intuition behind the bipartite representation.

In this thesis, we distinguish between lazy and non-lazy random walks. A lazy random walk is a process in which the probability of remaining at time $t + 1$ in the node where the walker was at time t is non-zero. In a non-lazy random walk, this probability is instead zero.

1.3.2 Formulation and analysis

We explore different types of random walks.

Given an adjacency matrix \mathcal{A} , the probability matrix is defined as

$$\mathbf{P} = \mathcal{D}^{-1}\mathcal{A}, \quad (1.9)$$

where \mathcal{D} is $\text{diag}(\sum_{k=1}^N \mathcal{A}_{ik})$, the matrix whose i -th element is the sum of the i -th row of the adjacency matrix. The matrix element \mathbf{P}_{ik} is the probability to transition from state i to state k , $P(i \rightarrow k)$ ³. Since the adjacency matrix is defined in different ways, also the random walk will be.

Take $\mathcal{A} = \mathbf{A}^{\text{count}}$ as defined in Eq.(1.3). We get what is known as a projected random walk. The i -entry of \mathcal{D} represents the number of outgoing paths from node i , while instead $\mathbf{A}_{ik}^{\text{count}}$ is the number of paths linking node i and k . Thus, $P(i \rightarrow k)$, in this case, is the number of paths from node i to k , divided by the total available paths. From this, the name $\mathbf{A}^{\text{count}}$: the matrix elements count the number of paths linking two vertices.

If one chooses $\mathcal{A} = \mathbf{A}^{\text{norm}}$ as defined in Eq.(1.4), we get a different walk. To understand its meaning, first, consider a lazy random walk. The adjacency matrix can be rewritten in matrix form as follow:

$$\mathcal{A} = \mathcal{I} \mathbf{E}^{-1} \mathcal{I}^T,$$

where $\mathbf{E} = \text{diag}(|e_i|)$. Now it is clear that the probability

$$\mathbf{P} = \mathcal{D}^{-1}\mathcal{I} \mathbf{E}^{-1} \mathcal{I}^T$$

represents a two-step process: first, the random walk moves to a hyperedge choosing uniformly between all the possible k_i linked hyperedges, and thus the associated probability is given by $\mathcal{D}^{-1}\mathcal{I}$; then from that hyperedge, the random walk chooses a node uniformly between all the possible $|e_j|$ nodes, and thus the associated probability is given by $\mathbf{E}^{-1}\mathcal{I}^T$. Since the random walk is lazy, it is equivalent to an unbiased random walk on the bipartite projection. Indeed, take $\mathcal{A} = \mathbf{A}^{\text{bipartite}}$. The random walk on this projection is a one-step process. It will move first to a node $v_i \in V$ and then to a node $e_{i+1} \in E$. This type of walk is trivially equivalent to a lazy random walk on the hypergraph since it is a sequence of alternating vertices and edges. In the case of a non-lazy random walk, the only difference is that the random walk cannot go back. This means that the two walks are not the same, but the behavior is similar. However, one can project the bipartite graph

$$\mathbf{A} = \begin{pmatrix} \mathbf{0} & \mathcal{I} \\ \mathbf{E}^{-1} \mathcal{I}^T & \mathbf{0} \end{pmatrix}, \quad (1.10)$$

³Note that, sometimes, the transition matrix is defined as the opposite: $P_{ik} = P(k \rightarrow i)$.

remove the diagonal and obtain a non-lazy random walk. There is also an extension of the bipartite representation with state nodes for non-lazy random walks [44].

Projecting the bipartite graph means taking the square of the adjacency matrix, therefore

$$\mathcal{A} = (\mathbf{A}^{bipartite})^2 = \begin{pmatrix} \mathcal{I}\mathcal{I}^T & \mathbf{0} \\ \mathbf{0} & \mathcal{I}^T\mathcal{I} \end{pmatrix}. \quad (1.11)$$

Except for the diagonal term, one can recognize the adjacency matrix as defined in Eq.(1.3). The $\mathcal{I}^T\mathcal{I}$ is the hyperedge counterpart. Indeed, there is a symmetry between hyperedges and vertices. One can equivalently define a walk on the hyperedges, obtaining a projected graph whose nodes are the hyperedges and whose weighted edges are the number of paths linking two hyperedges. In this case the diagonal will be $\mathcal{D} = \text{diag}(\sum_j |e_j|)$. The adjacency matrix so defined is block-diagonal. Indeed the two projected graphs $\mathcal{I}\mathcal{I}^T$ and $\mathcal{I}^T\mathcal{I}$ are not connected and independent. We showed that the transition probabilities are the same among representations, but a random walk on the bipartite graph can be more general. With an appropriate choice of edge weights, it has the same transition probability of a random walk on a hypergraph with edge-dependent vertex weight, a process not describable by a clique projection. In this sense, both \mathbf{A}^{count} and \mathbf{A}^{norm} are contained in $\mathbf{A}^{bipartite}$, but the latter is more general.

We show how the bipartite representation can be weighted such that a random walk on it is equivalent to a random walk on a hypergraph with edge-dependent vertex weight. In [36], the probability of transitioning from a node v to a node w of such a random walk is

$$P_{vw} = \sum_{e \in N(v)} \frac{w(e) \gamma_e(v)}{d(v) \delta(e)}, \quad (1.12)$$

or in matrix form

$$P = D_V^{-1} W D_E^{-1} \Gamma, \quad (1.13)$$

where $N(v)$ is the set of hyperedges at which e belongs, $d(v) = \sum_{e \in N(v)} w(e)$ is the weighted degree of node v and $\delta(e) = \sum_{v \in e} \gamma_e(v)$ is the weighted degree of hyperedge e . D_V and D_E are the diagonal matrices of, respectively, the weighted vertex degree and the weighted edge degree. The matrix Γ is the matrix such that each entry is

$$\Gamma_{ij} = \begin{cases} \gamma_j(i) & \text{if } i \in j \\ 0 & \text{else} \end{cases} \quad (1.14)$$

and

$$W_{ij} = \begin{cases} \omega(j) & \text{if } i \in j \\ 0 & \text{else} \end{cases}. \quad (1.15)$$

This random walk is lazy, and so Eq.(1.13) factorizes in two independent probabilities. This is what makes the bipartite projection suitable to represent this type of hypergraph. The corresponding adjacency matrix is

$$\mathbf{A} = \begin{pmatrix} \mathbf{0} & W \\ \Gamma^T & \mathbf{0} \end{pmatrix}. \quad (1.16)$$

A walk on this bipartite graph has the same transition probability of Eq.(1.12).

1.4 Laplacians

1.4.1 Definition on graphs

From the concept of random walk, it follows naturally the process of diffusion. We recall that the unbiased diffusion equations are written as

$$\partial_t n(\mathbf{x}, t) = D\Delta n(\mathbf{x}, t), \quad (1.17)$$

also known as Fick's second law of diffusion [47], where Δ is the Laplace operator, and D is the diffusion coefficient. This is related to the concept of random walk. Consider several random walkers on an undirected graph G and let us call $n_i(t)$ the number of random walkers in node i at time t . The number of random walkers in node i at time $t + dt$ will be given by

$$n_i(t + dt) = n_i(t) + dt \left(\sum_{j \neq i} w_{ij} n_j - w_{ii} n_i \right), \quad (1.18)$$

where w_{ij} is the transition rate from node j to i , and w_{ii} is the sum of all transition rates from i to j . An unbiased random walk on a graph moves only to the first nearest neighbors with equal probability. As a consequence, apart from a constant that can be absorbed in the time unit dt , the transition rate w_{ij} is given by the adjacency matrix element A_{ij} . The term w_{ii} instead is given by $\sum_{j \neq i} w_{ji}$, which in a graph will be proportional to the degree of node i . Everything can then be rewritten in a matrix form and taking the limit $dt \rightarrow 0$ as

$$\partial_t \mathbf{n} = (\mathbf{A} - \mathcal{D}) \mathbf{n} = -L\mathbf{n}, \quad (1.19)$$

where we used the convention that matrices are written with capital letters and vectors in bold. \mathcal{D} is the degree matrix, and L is the so-called Laplacian matrix. One can notice the similarity with Fick's second law of diffusion and, from that, the choice of the name. The element of the Laplacian matrix is defined as

$$L_{ij} = \begin{cases} \text{deg}(v_i) & \text{if } i = j \\ -1 & \text{if } v_i \in N(v_j) \\ 0 & \text{otherwise} \end{cases}. \quad (1.20)$$

However, in [48], the Laplacian matrix is directly introduced, explicating its relation with the incidence matrix,

$$L = BB^T = \mathcal{D} - A, \quad (1.21)$$

where B is the $|v| \times |e|$ directed incidence matrix defined element-wise as

$$B_{ij} = \begin{cases} 1 & \text{if edge } e_j \text{ enters vertex } v_i \\ -1 & \text{if edge } e_j \text{ leaves vertex } v_i . \\ 0 & \text{otherwise} \end{cases} \quad (1.22)$$

Note that undirected graphs are thought of as directed graphs in this picture. While Eq. (1.22) is a compact and convenient form when analyzing its spectrum, we are more used to the standard definition of the incidence matrix \mathcal{I} for an undirected graph, in which the entries are all positive. With this definition the adjacency matrix can be obtained from $A = \mathcal{I}\mathcal{I}^T - \mathcal{D}$. We will not follow the notation of Eq. (1.22). In this way, from the incidence matrix defined in Eqs. (1.1) and (1.2) one can recover the incidence matrix of a graph, directed or undirected, when $|e| = 2 \forall$ hyperedge e .

Note that the definition in [48] holds both for undirected and directed graphs. Nevertheless, in the case of directed graphs, the definition of the Laplacian as in Eq. (1.21) has some problems⁴. First, the adjacency matrix A is usually defined as not symmetric for a directed graph, so also the Laplacian should be. Indeed, if one thinks about how we introduced the Laplacian matrix, that is, through a diffusion process, the walk of a random walk is not symmetric in a directed graph. For this reason, we choose the definition as in Eq. (1.19), and we will refer to the Laplacian defined in Eq. (1.21) as the *symmetrized Laplacian*. Actually, for directed graphs, several definitions of the Laplacian matrix are present in the literature and their choice usually depends on the applications.

Different measures can be done using the Laplacian matrix. The most common is the algebraic connectivity, also called the Fiedler value. It is defined as the first non-zero eigenvalue of the Laplacian matrix and measures the connectivity of a graph. However, the Laplacian is not symmetric for directed graphs, so nothing ensures that it is diagonalizable and that the spectrum is real. This makes the algebraic connectivity not well defined and may be a motivation to prefer the choice of the symmetrized Laplacian. Also, there exists some generalization of the algebraic connectivity for directed graphs that avoids this confusion while maintaining most properties of the Fiedler value [49].

⁴Note that the out-degree matrix substitutes the degree matrix.

1.4.2 Generalization to Hypergraphs

Here we generalize the Laplacian matrix for hypergraphs using the representations and random walks described before. Then we compare our generalization of the Laplacian with the literature.

Since we already defined representations and used them to define different random walks, we can introduce the Laplacian matrix as the matrix that describes the transition rates of random walkers that diffuse on a hypergraph. Using the same reasoning as in the section above, we arrive at a definition of the Laplace matrix for hypergraphs equivalent to Eq. (1.19),

$$L = \mathcal{D} - A, \quad (1.23)$$

where A is the adjacency matrix of the hypergraph and $\mathcal{D} = \text{diag}(\sum_{k \neq i} A_{ik})$. We recall that different random walks can be defined and so their diffusion on the hypergraph is different. For the projected random walk $A = A^{\text{count}}$, Eq. (1.3), for the higher-order random walk $A = A^{\text{norm}}$, Eq. (1.4), and finally the diffusion on the bipartite is described by $A = A^{\text{bipartite}}$, Eq. (1.5). As anticipated, we will focus on the bipartite projection. The algebraic connectivity is then simply generalized as the first non-zero eigenvalue of the hypergraph Laplacian matrix.

Notice that this definition is also valid for directed hypergraphs, but with the same problems described above. Our Laplacian is not symmetric and so not always diagonalizable with real eigenvalues. In [45, 46], *Mulas and coworkers* define a symmetrised Laplacian matrix specifically for directed metabolic hypergraphs, generalizing Eq. (1.21). Since we are treating metabolic hypergraphs, it is interesting to make some considerations about these two Laplacians. First, the symmetrized hypergraph Laplacian describes the same process as the underlined undirected hypergraph, losing in this way all the information on the directionality of the reactions. We believe that the information on the reversibility or not of some reactions is physically too important to be neglected in the diffusion process. However, the Laplacian introduced by *Mulas* points out an important detail that our Laplacian is not able to capture. If m_1 and m_2 are two reactants belonging to a not-reversible reaction, in our representation, they are not connected and thus not interacting, while physically we know that they are taking part in the reaction and thus interacting.

For this thesis, we decided not to investigate further the several definitions of and leave it as future work. However, we still report the result of the algebraic connectivity for some bipartite metabolic hypergraphs in Appendix A.

For completeness, we also report the definition that was given in [36] for the hypergraph Laplacian matrix of a hypergraph with EDVW,

$$L = \Pi - \frac{\Pi P + P^T \Pi}{2}, \quad (1.24)$$

where P is the probability transition matrix of the hypergraph with stationary distribution π . Π is the $|V| \times |V|$ diagonal matrix with element $\Pi_{v,v} = \pi_v$.

1.5 Characterization measurements

In this section we define some measures, generalizing on hypergraphs the concepts of *communicability* [50] and *search information* [51]. They allow us analysis at multiple scales. From the communicability, we obtain the natural connectivity that gives a measure of the overall robustness of the hypergraph and, complementary, the average search information is related to the global complexity of the network. Communicability and search information are microscopic measurements of how parts of the network communicate.

1.5.1 Communicability

Estrada et al. [50, 52] define the communicability of a graph as

$$G_{ij} = \sum_{k=0}^{\infty} c_k (A^k)_{ij}, \quad (1.25)$$

where c_k is some coefficient depending on k such that the series is convergent. $(A^k)_{ij}$ is counting the number of paths of length k from i to j . The meaning of c_k is thus a “penalty” to paths that are long. A common choice for this coefficient is $c_k = \frac{1}{k!}$. With this choice, the communicability is just the exponential of the adjacency matrix

$$G = e^A. \quad (1.26)$$

They also introduce as a measure of the robustness of the graph the *natural connectivity* [50, 53]

$$\bar{\lambda} = \log(EE(G)/n), \quad (1.27)$$

where n is the number of nodes in the graph and $EE(G)$ is the *Estrada index*

$$EE(G) = Tr \{G\} = Tr \{e^A\} \quad (1.28)$$

for our choice of communicability.

We explicit the dependence of the natural connectivity on the largest eigenvalue of the adjacency matrix. Assume A diagonalizable⁵, then there exists some invertible matrix V such that $A = V\Lambda V^{-1}$, where $\Lambda = \text{diag}(\lambda_i)$ and λ_i are the eigenvalues of A . We can then write $e^A = V e^D V^{-1}$, where we recognize $e^D = \text{diag}(e^{\lambda_i})$ to be

⁵This is always true in the case of an undirected graph.

the eigenvalue of G . Assuming that $\lambda_1 > \lambda_2 > \lambda_3 > \dots$, the natural connectivity becomes

$$\begin{aligned}\bar{\lambda} &= \log \left(\sum_{i=1}^n e^{\lambda_i} \right) - \log(n) = \\ &= \log \left[e^{\lambda_1} \left(1 + \sum_{i=2}^n e^{\lambda_i - \lambda_1} \right) \right] - \log(n) = \\ &= \lambda_1 + \log \left(1 + \sum_{i=2}^n e^{\lambda_i - \lambda_1} \right) - \log(n) \\ &= \lambda_1 - \log(n) + \mathcal{O} \left(e^{-(\lambda_1 - \lambda_2)} \right).\end{aligned}$$

If the spectral gap is big enough, the natural connectivity is dominated by the largest eigenvalue. Since the correction is exponential, this approximation is expected to be quite good.

We can also define some bounds to the natural connectivity. The graph with the higher communicability possible is a complete graph whose spectrum is known to be $\{(n-1)^1, -1^{n-1}\}$, where the exponent indicates the multiplicity of the eigenvalue and n is the number of nodes. The Estrada index is then given by

$$EE^{max}(A) = Tr \{ e^A \} = e^{n-1} + (n-1)/e, \quad (1.29)$$

and so the largest possible value for the natural connectivity is well approximated by

$$\bar{\lambda}^{max} \simeq n - 1 - \log(n) \quad (1.30)$$

for large n . It is also useful to compute the natural connectivity for a complete bipartite graph B . For such graph, the spectrum is $\{\sqrt{nm}, 0^{n+m-2}, -\sqrt{nm}\}$, where n and m are the nodes of the two bipartite sets. The natural connectivity is then approximated by

$$\bar{\lambda}_{bipartite}^{max} \simeq \sqrt{nm} - \log(n+m). \quad (1.31)$$

The generalization to hypergraph is straightforward. Once one defines the hypergraph adjacency matrix, all the same definitions can be extended.

The communicability G of the bipartite projection takes a nice form. The adjacency matrix defined in Eq.(1.5) has zero main diagonal blocks. Consequently, all the even powers of the adjacency matrix are block diagonal, while the odd powers are off block-diagonal. In detail,

$$A^{2n} = \begin{pmatrix} (\mathcal{I}\mathcal{I}^T)^n & \mathbf{0} \\ \mathbf{0} & (\mathcal{I}^T\mathcal{I})^n \end{pmatrix}, \quad (1.32)$$

$$A^{2n+1} = \begin{pmatrix} \mathbf{0} & (\mathcal{I}\mathcal{I}^T)^n \mathcal{I} \\ ((\mathcal{I}^T\mathcal{I})^n \mathcal{I})^T & \mathbf{0} \end{pmatrix}, \quad (1.33)$$

where \mathcal{I} is the hypergraph incidence matrix. The communicability can be factorized in an odd and in an even part

$$G = \frac{e^A + e^{-A}}{2} + \frac{e^A - e^{-A}}{2} = \cosh A + \sinh A, \quad (1.34)$$

where $\cosh A$ contains only even powers of A and so, for the considerations above, is a diagonal block matrix. On the contrary, $\sinh A$ has zero block diagonal. We can use this result to compare it with the communicability computed using \mathbf{A}^{count} . If one considers the first block of Eq.(1.32) is the n -th power of \mathbf{A}^{count} .⁶

The $\mathbf{A}^{bipartite}$ also has some nice spectral properties. It is symmetric, so it is diagonalizable with real eigenvalues, and since the corresponding graph is bipartite, if λ_i is an eigenvalue, then $-\lambda_i$ is also an eigenvalue.

By looking at A^2 , its eigenvalues are the singular values squared of the incidence matrix, with multiplicity 2. Consequently, the eigenvalues of A are, in module, the singular values of the incidence matrix. By denoting the singular values of \mathcal{I} s_i , we get that the natural connectivity of the bipartite graph is:

$$\bar{\lambda} \simeq s_i - \log(n). \quad (1.35)$$

The natural connectivity of a *fully connected hypergraph* is the same as a fully connected bipartite graph, and so it is given by Eq.(1.31).

1.5.2 Information on graphs and hypergraphs

Rosvall et al [51] introduced the concept of search information, as a measure of complexity:

$$S(i, j) = \log_2 \left(\sum_{\{p(i, j)\}} P(p(i, j)) \right), \quad (1.36)$$

where $\{p(i, j)\}$ is the set of all shortest paths from i to j and

$$P(p(i, j)) = \frac{1}{k_i} \sum_{l \in p(i, j)} \frac{1}{k_l - 1} \quad (1.37)$$

is the probability to follow path $p(i, j)$ in the graph, and k_l is the degree of node l . $S(i, j)$ corresponds to the number of binary questions needed to reach j from i . Based on this local quantity, respectively the access, hide, and average search

⁶Note that, in this case, we allow lazy random walks.

information are introduced as

$$\begin{aligned}
 A(s) &= \frac{1}{N} \sum_t S(s, t) \\
 H(t) &= \frac{1}{N} \sum_s S(s, t) \\
 S &= \frac{1}{N^2} \sum_{s,t} S(s, t),
 \end{aligned} \tag{1.38}$$

where N is the number of nodes in the graph, s is a source node in the graph, and t is a target node in the graph.

We generalize these concepts for hypergraphs. They will be defined differently depending on which representation one is projecting the graph.

The probability of making a step for a projected random walk is given by

$$\begin{aligned}
 P(i, j) &= \frac{\# \text{ of hyperedges connecting } i \text{ and } j}{\# \text{ of hyperedges containing } i} \\
 &= \frac{\mathcal{A}_{ij}}{\sum_{k=1}^N \mathcal{A}_{ik}},
 \end{aligned}$$

where N is the number of vertices in the hypergraph.

Consider now a higher-order step. The probability of making a step will be

$$P(i, j) = \begin{cases} \frac{1}{k_i} \frac{1}{|e|-1} & \text{if } i, j \in e \\ 0 & \text{otherwise} \end{cases}, \tag{1.39}$$

where k_i is the number of hyperedges to which i belong, and $|e|$ is the size of the set e , i.e., the number of vertices it contains. Note that the projected graph has the same edges and paths in both cases, just with different weights.

It is interesting to highlight that the probability of following a path can be rewritten emphasizing that a walk on the hypergraph is a two-step process ⁷ as

$$P(p(i, j)) = \frac{1}{k_i} \frac{1}{|e_i|-1} \sum_{(v,u) \in p(i,j)} \frac{1}{k_v-1} \frac{1}{|e|-1}, \tag{1.40}$$

given that the first step is the tuple (i, e_i) . In the case of the bipartite representation, the walk is a one-step process and so the definition of the probability of a path reduces to Eq.(1.37), with no modification. This is mathematically equivalent to Eq.(1.40). So again, we see that the higher-order step is equivalent to a walk on the bipartite projection.

⁷Here, we use the higher-order step.

If the hypergraph has edge-dependent vertex weight, we showed that it is well described by the bipartite representation. We can then use the same definition but generalized for graphs with weighted edges. In this case, the probability of a random walk to make a step $i \rightarrow j$ is proportional to the weight of the edges connecting i and j

$$P(i, j) = \frac{w(e_{ij})}{\sum_{k \in N(i)} w(e_{ik})}, \quad (1.41)$$

where $N(i)$ is the neighborhood of i and $w(e_{ij})$ is the weight of the edge e_{ij} .

All the previous formulas apply for a directed graph, only with minor modifications. We replace the degree with the out-degree⁸, and we do not always subtract 1 from it. In fact, the $k_i - 1$ in the denominator of Eq.(1.37) is present because the walk cannot turn back and thus can choose between $k_i - 1$ vertices to move toward. In a directed graph, this is not always true. If the path comes from a directed edge, it cannot go back using the same edge. It follows the equation that we used for the metabolic hypergraph,

$$P(p(i, j)) = \frac{1}{k_i^{out}} \sum_{l \in p(i, j)} \frac{1}{k_l^{out} - r_l}, \quad (1.42)$$

where r_l is equal to one if the node l is a reversible reaction or if it is a metabolite that, in the path, is coming from a reversible reaction. Otherwise, it is zero.

While the formula is easily generalized to hypergraphs, it is important to point out that if the directed graph is not strongly connected, then some nodes may be unreachable or may not reach other nodes. It means that they have infinite hidden or access information. To avoid this problem, we set these cases to have zero information, but this can impact the value of the average search information.

⁸The out-degree is the number of edges leaving the vertex, denoted by k_i^{out} . If the graph is unweighted, then it is the sum of the weights of the edges exiting the vertex.

Chapter 2

Metabolic Hypergraphs

In this section, we present the applications of our theoretical work. We first analyze the metabolic hypergraph of *Escherichia coli*, and then we compare the robustness and information of different BiGG models.

2.1 Introduction to Metabolic Hypergraph

Here we define the metabolic hypergraph as the hypergraph whose nodes are the metabolites and the reactions are the hyperedges, and we use the bipartite representation. Note that one can use the representation \mathbf{A}^{count} , Eq.(1.3), and obtain a graph where metabolites are vertices and are linked if they share a reaction. The issue with this representation is that we lose directionality and we encounter the problem explained in [43]. However, using the bipartite projection, we can intuitively include the directionality: reagents of a reaction i have directed links towards the hyperedge representing reaction i , and products of reaction i have directed links incoming from reaction i . Note that this is a directed hypergraph. For example, consider the reaction *Phosphofructokinase*, BiGG-id PFK, which is a reaction belonging to the *Glycolysis/Gluconeogenesis* metabolic pathway, involved in the conversion of fructose 6-phosphate and ATP to fructose 1,6-bisphosphate and ADP. The corresponding representation as a graph is shown in Fig. 2.1.

The reaction PFK is catalyzed by the enzyme *Phosphofructokinase-1*, or *PFK-1*, from which the reaction takes its name. In chemistry, a *catalyst* is an enzyme that takes part in a reaction but is not changed by it. In practice, it is an element that appears on both sides of the reaction equation. Here we do not consider *catalyst* since they do not affect the path followed by a random walk in the metabolic network. Indeed, since reactions do not modify them, they cannot “move” on the graph. However, they can be taken into account. In [46], they define an oriented graph where *catalysts* are taken into account as self-loops in the graph.

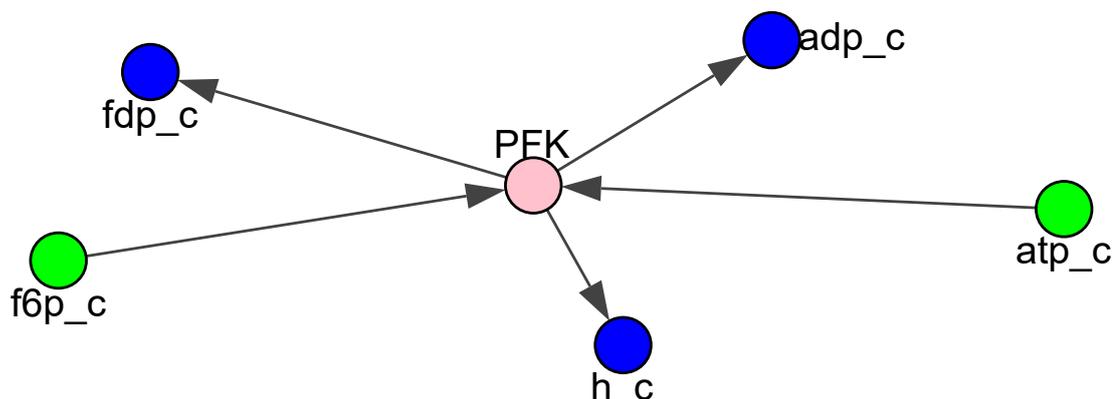


Figure 2.1: PFK Reaction from BiGG dataset: $atp_c + f6p_c \rightarrow adp_c + fdp_c + h_c$. In green the reagents, in blue the products, and in pink the reaction. Directionality is easily taken into account with this definition.

In this section, we apply the concepts of *communicability* and *search information* to the *E.coli* metabolic network [54]. Communicability can identify important reactions, and the search information measures the complexity of the network at different levels. We base our results on purely structural measures. Indeed, the chemistry is taken into account only when building the network but then is neglected. The mass of reagents/products is not conserved since stoichiometry is not taken into account as edge weights¹. This is not the approach that people usually follow to analyze metabolic networks [43, 55]. The difference is that here we are considering a higher-order structure and not a simple graph. The aim of this analysis is then to be able to understand if including this type of structure is relevant in characterizing the metabolic network.

2.2 Mapping stoichiometry matrices into hypergraphs

There are many real-world examples of systems that can be described and studied with hypergraphs. For example, one can think of social network relations as a hypergraph, where each chat or group represents a hyperedge. Contagion models can be viewed as hypergraphs too since in our everyday life our relations are not limited to interaction with just one other person. The places where the viruses spread the most are the ones where multiple people are present, and thus considering

¹The hypergraph is unweighted for simplicity.

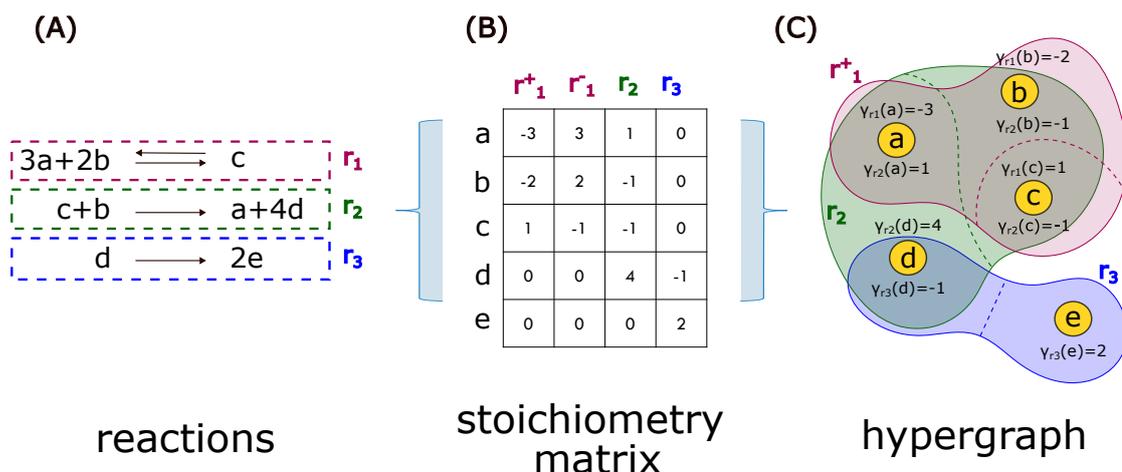


Figure 2.2: An example of a metabolic network mapped into a hypergraph with edge-dependent vertex weight. In (A), we present a small network composed of three reactions and five metabolites. The first reaction r_1 is reversible and is represented with the double arrow. In (B), we show the corresponding stoichiometry matrix. Reactants are negative and products are positive. We need to split the reversible reaction into two irreversible reactions r_1^+ and r_1^- to write it in matrix form. This stoichiometry matrix is the incidence matrix of the hypergraph with edge-dependent vertex weights shown in (C). For the sake of visualization, only the hyperedge r_1^+ is shown. The hyperedge r_1^- is just the same but with the opposite sign. Note that weights are both positive and negative, meaning that the hypergraph is directed. Indeed, we separate the head and tail of each hyperedge with a dashed line.

multi-body interaction is essential [11]. Also, biological and ecological systems may include multiple agents/species interactions [27]. In particular, we will focus on metabolic networks. Indeed, metabolic networks are naturally mapped into hypergraphs. Each network composed of N metabolites and M reactions is fully specified by the stoichiometry matrix S , a $N \times M$ matrix, such that

$$S_{i,j} = \begin{cases} 0 & \text{if metabolite } i \text{ is not in reaction } j \\ c_i^j & \text{if metabolite } i \text{ is a product in reaction } j \\ -c_i^j & \text{if metabolite } i \text{ is a reactant in reaction } j \end{cases},$$

where c_i^j is the stoichiometric coefficient of metabolite i in reaction j . This matrix is the incidence matrix of a directed hypergraph with edge-dependent vertex weight, where reactions are the hyperedges and metabolites are the vertices. Directed because the weights $\gamma_e(v)$ are positive and negative. One can indeed recognize, in the same way we defined the incidence matrix in Eq. (1.2), the head, vertices with positive weights, and the tail, vertices with negative weights, of the hyperedge.

Metabolites have different weights depending on which reaction they belong to. See Fig. 2.2 for an example. In the figure, the metabolite d is present in the reaction r_3 with weight -1 , meaning that it is a reagent, but is also the product of the reaction r_2 , with weight 4 . This particular set of weights, plus the fact that reactions are directed, makes this type of hypergraph the most general one.

2.3 The dataset and preprocessing

The metabolic hypergraphs are taken from the *BiGG Database* [56]. We analyze more than 20 different models, with an increasing number of nodes and describing different organisms. All data are publicly available on the BiGG models [57] web page in different formats. In this analysis, the *.json* format is used. The data contain information on metabolites, reactions, and genes. Metabolites are identified by a BiGG id, consisting of an abbreviation defining their type, for example, “h” for hydrogen and “ATP” for the adenosine triphosphate, and a subscript indicating the compartment to which they belong. The suffix “c” indicates the cytosol compartment and the suffix “e” the extracellular space compartment. Regarding the reactions, in addition to their ids, the metabolites belonging to them are given, with their respective stoichiometric coefficients. Usually, a negative coefficient indicates that the metabolite is a reactant, but not always. One has to look at the “upper bound” and “lower bound” parameters associated with the reaction to understand its direction and whether it is reversible or not.

2.4 Communicability on Escherichia coli

We compute the communicability of the bipartite metabolic network as Eq.(1.26). We show in Fig. 2.3 the metabolic network where each node is weighed in size and color based on the average communicability per node. One can notice that the points with higher communicability are located in the center of the graph, as expected. The metabolite with the higher communicability is h_c, so hydrogen in the cytosol compartment, as expected since it appears in many reactions. Other metabolites with higher communicability are h_e, H₂O, and ATP and NAD, which are important metabolites responsible for the energy consumption/production of the cell. In general, while this is a nice result since the highlighted metabolites are important for the organism, similar results can be derived by using the degree centrality measure: indeed, the average communicability per node and the degree distribution² correlates, with some minor differences. However, if one considers

²The degree in the directed graph is defined as the sum of out degree and in degree.

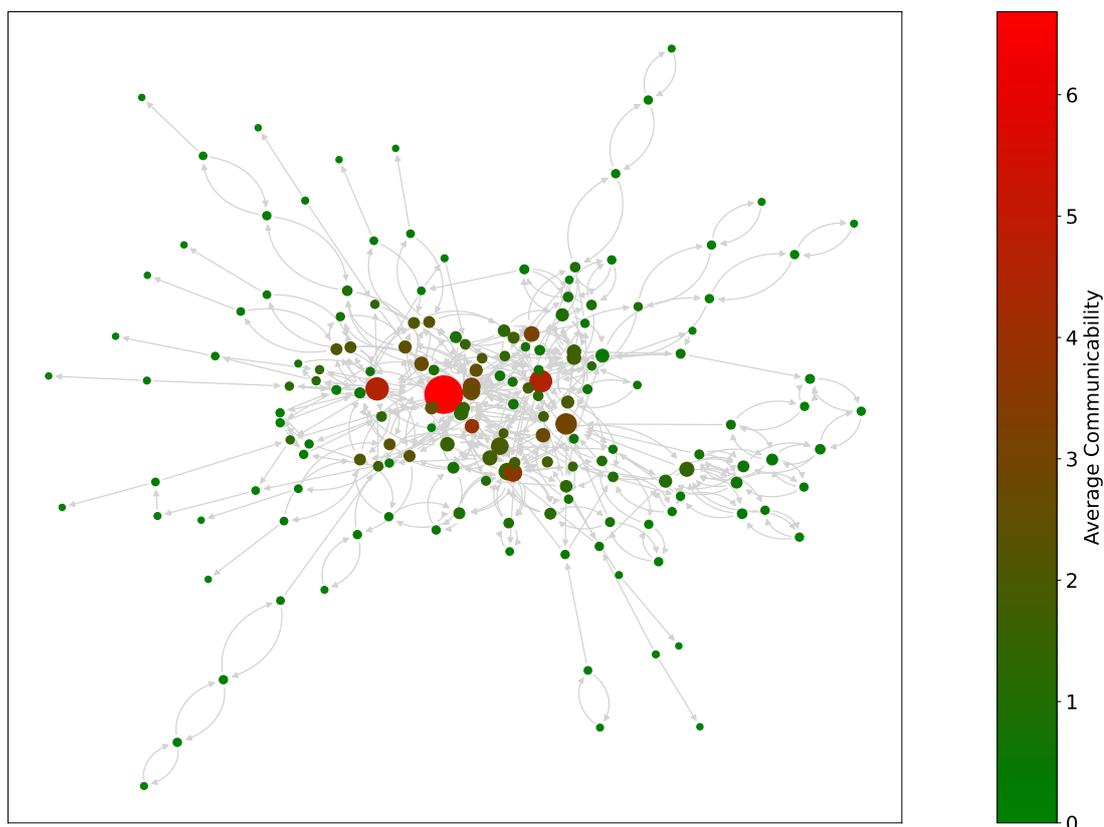


Figure 2.3: E. Coli metabolic hypergraph in the bipartite representation. Nodes are metabolites and reactions. The color of each node is proportional to the average communicability of that node, and the size is proportional to the node degree.

the reactions network, we have some differences, and the average communicability per reaction reveals something different from the degree centrality measure. The *Biomass Objective Function with GAM* is the reaction with the highest degree in the network³, but while still having high average communicability, the *ATP synthase* reaction has a higher one. This reaction is central in a metabolic process since many chemical processes require energy in the form of ATP. Another interesting fact to point out is that the metabolites or reactions that have higher communicability are the ones that are shared by many metabolic pathways. Recall that in a metabolic network, different regions can be identified based on the function they fulfill: those regions are called metabolic pathways.

³23 metabolites are involved in the reaction.

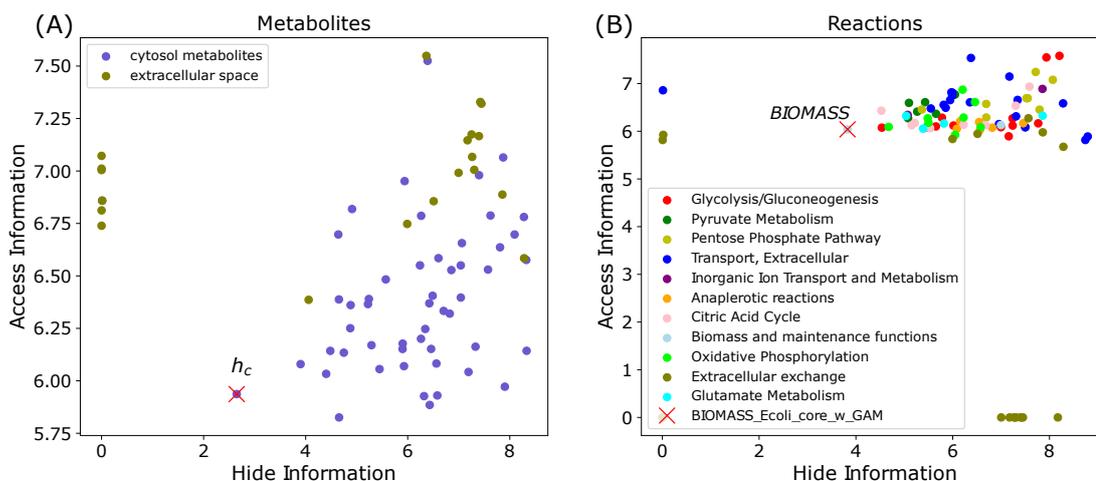


Figure 2.4: Access information vs Hide information for the metabolites (A) and reactions (B). Metabolites in blue belong to the cytosol compartment, and the ones in green belong to the extracellular space. The latter have, on average, higher access and hide information compared with the former since they belong to the outermost of the network. Each reaction is colored differently depending on the metabolic pathway to which they belong. We do not observe a separation between pathways in the Access vs Hide information plane, meaning that different pathways have, on average, the same complexity in *E. coli*. The nodes with the higher degree, h_c and *Biomass*, are highlighted with a red cross. Some nodes appear to have zero access or hide information. Those nodes cannot reach or be reached by other nodes.

2.5 Information on Escherichia coli

We apply the concepts introduced in section 1.5.2 to *E. coli* hypergraph. We compute the access and hide information for each node in the graph and we compare them. Results are shown in Fig. 2.4. The access information does change much, while the hide information varies more. This suggests that there are no reactions or metabolites that “connect” worst than others, and so that for whatever reaction or metabolite it is equally “easy” to access other metabolites or reactions in the network. With this measurement, we can identify *hubs* in the graph as nodes with low hide information and relatively high access information. Hubs are points with a high degree that are easily reachable from the other nodes in the network, and thus, they are “visible” and have low hide information. From a hub is harder to reach a specific point in the network since the degree is high, resulting in relatively high access information. On the contrary, nodes that are leaves of the graph will be better hidden from the other nodes of the network. For the metabolites, the *hub* is h_c , the hydrogen in the cytosol compartment, present in most of the core

reactions⁴; for the reaction is the *Biomass reaction*, the reaction involved in the cell growth, and thus the principal reaction for the metabolism. The points with zero hide information are just points that are not reachable and are set to zero for convenience. One can also see that both the hide and access information do not correlate with the node degree. This is because they depend more on the paths than the degree. In Fig. 2.4, metabolites are divided into metabolites belonging to the cytosol compartment, the interior of the cell, represented in blue, and metabolites belonging to the extracellular space, in green. The latter are metabolites exchanged with the environment, like cell nutrients or wastes, they are labeled with the “_e” subscript and belong to reactions of the *Transport Extracellular, Extracellular exchange and Inorganic Ion Transport and Metabolism* pathways. They are leaves of the bipartite graph. The former are the metabolites belonging to the core reactions of the cell, and they are labeled with the “_c” subscript. Metabolites belonging to the extracellular compartment have, on average, higher access and hide information with respect to the ones of the cytosol compartment.

Reactions are grouped into 12 different pathways. They do not separate in clusters, meaning that the pathways are equally complex in the network. They show a general trend: most pathways have reactions with progressively lower hide information.

2.6 Comparison with Flux Balance Analysis

In this section, we try to link the structural properties we measured through the communicability and the search information with biological properties that we can obtain with Flux Balance Analysis (FBA) [55]. It is a widely used technique that consists in solving a linear system of equations for the fluxes of metabolites. Since multiple solutions exist, we specify an objective function to maximize, that in the case of *E.coli* is chosen to be the Biomass objective function. This is a common choice for most metabolic networks since it is the reaction responsible for cell growth. To simulate FBA, we use the Cobrapy package for python [58]. With this package, we can compute the fluxes of metabolites in the network, but we can also perform some other valuable simulations like single-gene deletion and single-reaction deletion. These algorithms perform a deletion of a gene or reaction in the network and then compute the flux of metabolites toward the Biomass function. These fluxes can be the same as the optimal solution or can lead to a lower flux of metabolites, even zero. If no fluxes of metabolites arrive at the Biomass, the cell is considered dead. We can then identify with this technique essential genes or essential reactions for the metabolic network as those genes

⁴h_c has in-degree $k_{in} = 31$ and out-degree $k_{out} = 19$

or reactions that, when inhibited, cause cell death. The quasi-essential genes or reactions are those that, when inhibited, reduce the flux of metabolites by 50% or more. To help visualize the single-reaction deletion technique, there is an online Escher map [59] that allows interactively to delete a reaction and see how the fluxes distribution changes in the network and how this operation impact the cell growth. Given the information on the most important reactions in the metabolic network, we want to compare them with the most important ones identified by the communicability measure. Results are shown in Fig. 2.5. We can see that essential reactions are only partially captured by the communicability measures, but this is not unexpected. There are some reactions that are essential but have low communicability. The majority of those are leaves or are close to external reactions, so it is natural that they are ranked low. To consider these reactions, we should define a different measure that can also consider the biological properties of the network, but as already anticipated, this is outside the scope of that thesis. We want to keep the work as general as possible. There are also some reactions that are ranked with high communicability even though they are not essentials. These are fails of our measures, and the reason for that is that they are reactions containing the metabolite `h_c`, which has a large degree and biases the measure. A possible solution for this is to remove the `h_c` metabolite, since it belongs to what in the field are called pool or ubiquitous metabolites, as done in [37]. Other examples of ubiquitous metabolites are H_2O , ATP, and `CO_2`. However, this procedure does not seem to be so well-founded and introduces other problems, such as the fact that the graph can become disconnected. We leave the exploration of this approach as future work.

2.7 Comparing metabolic networks

Here, we compare the robustness and the complexity of different metabolic networks. We analyze 20 different models, with an increasing number of nodes and describing different organisms. The data is treated in the same way described in section 2.3, and each hypergraph is constructed using the bipartite projection as discussed in section 2.1. We measure the natural connectivity and average search information of each hypergraph. The results show that the complexity is almost constant across individuals of very different species, while the robustness varies significantly. In particular, an antibiotic-resistant bacterium displays the highest robustness.

2.7.1 Results

The average search information and the natural connectivity are computed for each organism and the results are shown in Fig. 2.6. As can be seen, the complexity of the organisms, as measured by the average search information, is relatively

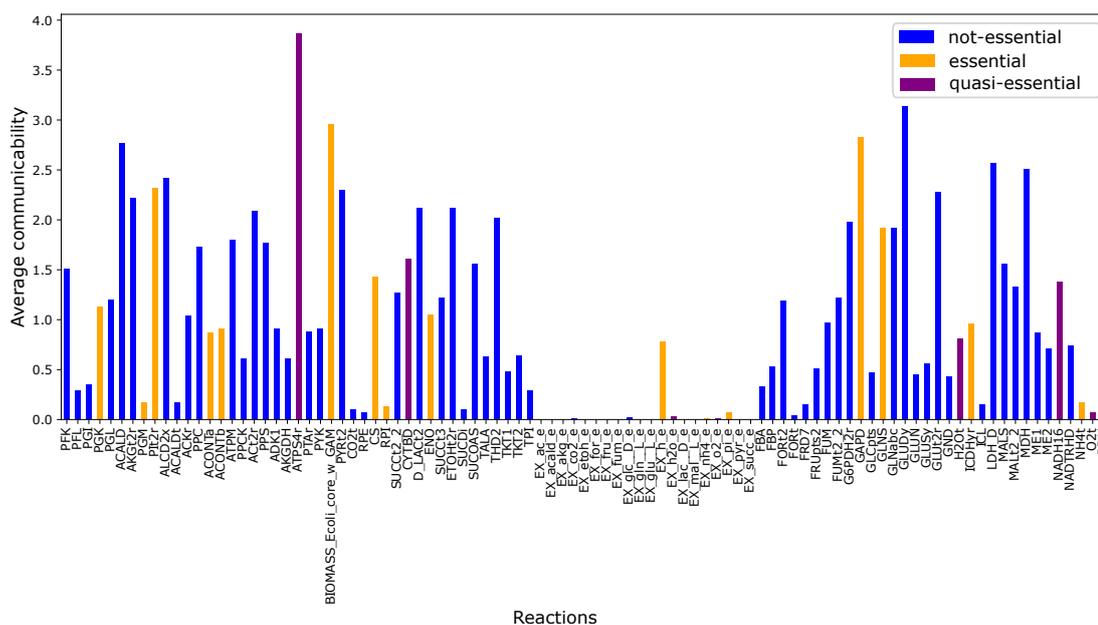


Figure 2.5: Plot of the reactions average communicability. In purple is represented the average communicability of quasi-essential reactions, in orange the average communicability for essential reactions, and in blue for non-essential reactions

constant. This is not surprising since even if the microorganisms are different, the reactions and pathways are similar [60]. On the contrary, the robustness of the network varies a lot between different organisms. We deduce that while the pathways and reactions are similar between individuals, they can be arranged differently to form more or less robust networks. A metabolic network with high robustness is expected to be efficient even if some reactions or some metabolites are not available for some reason. This means having alternative pathways, which translates into high communicability.

It is interesting to point out that antibiotics operate by targeting and inhibiting some specific reactions, without which the cell dies [61]. If the network has high natural connectivity, it is more difficult to identify reactions that, if suppressed, will cause cell death, because many alternative pathways exist. Looking at Fig. 2.6, the *Staphylococcus aureus subsp aureus N315* is the organism with the highest robustness, and it is indeed an antibiotic-resistant bacterium [62]. We stress that this result is obtained by a purely structural argument. Chemistry is taken into account just in the construction of the edges in the bipartite graph. The stoichiometric coefficients are disregarded and so is the mass conservation of reactants/products. All reactions are considered equally important, while standard *Flux Balance Analysis* (FBA) assumes a specific reaction to be optimized, usually the *Biomass reaction* [55].

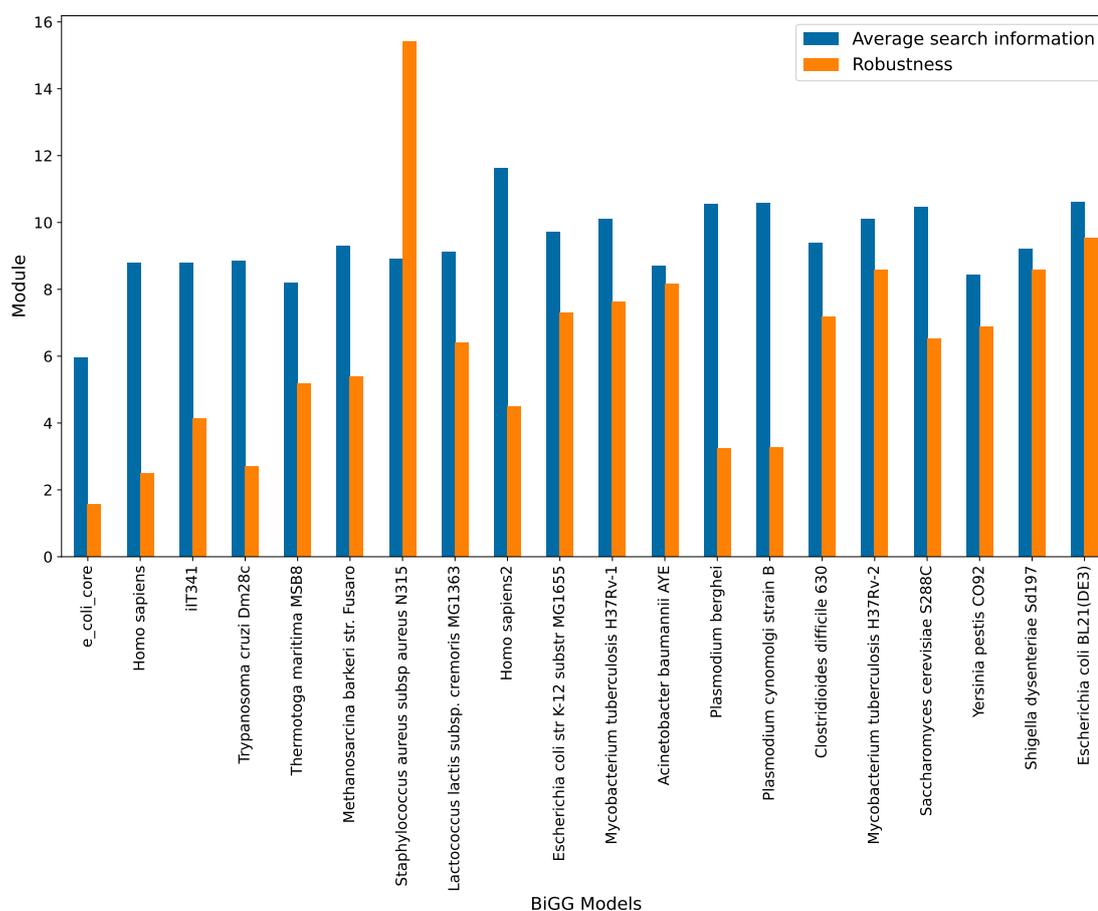


Figure 2.6: Average search information, in blue, and natural connectivity, in orange, computed for different BiGG models. The organisms analyzed range from simple bacteria to humans. The results show similar complexity among species, but a substantial difference in robustness. Note that the peak in robustness is obtained for the *Staphylococcus Aureus*. The models are ordered in increasing number of nodes.

However, by only considering structural properties, our proposed measurements are able to detect this biological feature, showing the potential of our methods.

Chapter 3

Random metabolic networks

In this chapter, we want to compare the results we obtained in the previous chapter with a null model of a metabolic hypergraph. Defining a null model is a standard procedure used to explore the statistical properties of graphs. The basic idea is to generate a random object that matches some specific properties of the object of interest, but is constructed in such a way that the model can be considered unbiased. In the case of hypergraphs, we want to build an unbiased hypergraph. So, for example, imagine having a social hypergraph, in which hyperedges form communities and we have a population divided into old and young people. They are the vertices of the hypergraph. The question that we would like to ask is: do people interact homogeneously or individual of the same category forms groups. A way to answer this question is to build a null model, a random hypergraph in which vertices are assigned randomly to hyperedges. If the assignment of the vertices is done in an unbiased way, then the null model represents an unbiased version of the original hypergraph. One can also think to preserve some structural properties of the original hypergraph, like in this case we can preserve the degree of each hyperedge. In this way, the community structure is the same for the two hypergraphs. The next point is to compare the two models, and to do so, we formulate what is called a null hypothesis. In the case of our example, the null hypothesis will be that communities have, on average, the same amount of old and young people. We can then verify or discard the hypothesis by comparing the distribution of old people in communities in the original and random hypergraph. If they are the same, then we can conclude that there is no bias in forming communities. This is an elementary example but it is similar to what we want to achieve for metabolic networks.

The idea is to see if the biological and chemical constraints of reactions induced a higher complexity or robustness compared to a hypergraph in which the chemistry is not taken into account. Our null hypothesis will then be that chemical and biological constraints do not affect the network's complexity or robustness. Although the problem seems simple and well posed, there are some subtle aspects. The first

thing to do is define how to randomize the hypergraph. To do so, we need to define the biological and chemical constraints we want to compare. Indeed, there are multiple ways in which we can define a random metabolic hypergraph.

The easiest way to randomize the network is to exchange metabolites belonging to different reactions. We are going to call this method *shuffle reactions*. The procedure is simple:

- choose a reaction r_i ,
- choose a metabolite in the chosen reaction r_i ,
- choose another reaction r_j ,
- if the chosen metabolite is a product (or a reactant), exchange it with a product (or a reactant) of r_j .

By randomizing in this way, we are changing only the chemistry of the reactions, but the degree of each reaction is left unchanged. We randomized the metabolic network $100 \times M$, where M is the number of reactions. At this point, we can build the hypergraph directly from this new random metabolic network as described in section 2.2. The resulting hypergraph will have the same amount of hyperedges and vertices and the same degree distribution.

Another possibility is to randomize directly the hypergraph. To do so, we again exploit the bipartite representation to generalize a type of shuffle widely used for directed graphs, that keeps the in-degree and out-degree fixed. We refer to this procedure as *shuffle edges* and consists in:

- select a random pair of edges $A \rightarrow B$ and $C \rightarrow D$,
- swap the links such that $A \rightarrow D$ and $C \rightarrow B$ if and only if the link is not already present.

With this procedure, the in-degree and out-degree distribution are kept fixed. We can use the same procedure on the bipartite projection of the metabolic hypergraph. We need to be careful because we want the final graph bipartite to still be interpreted as a hypergraph. So we randomize first the sub-bipartite graph with the edges going from metabolites to reactions and then the sub-bipartite graph with the edges going from reactions to metabolites. In this way, we avoid the situation in which a metabolite can be linked with another metabolite. This is equivalent to exchanging products with products and reactants with reactants, but we change the reversibility. Indeed now we are randomizing the hypergraph directly, by reshuffling the links. In this way, the double links that characterize the reversible reactions in the bipartite projection can be broken in the randomization procedure.

3.1 Shuffle reactions vs shuffle edges

We compare the results of the two null models we proposed above. We recall that the *shuffle reactions* is a randomization of the metabolic network and not directly of the generated hypergraph. This means that some chemical features of the metabolic network will be preserved, for instance, which reactions are reversible. What is swapped is the role of metabolites. In this new metabolic network, reactions have no physical meaning since they violate the mass conservation principle. Metabolites are created and destroyed without any rule as a consequence of randomization. However, consider that in the hypergraph construction, we do not take explicitly into account this factor, so we expect our null hypothesis to be true. Indeed, this is what has been observed by simulation. See the Appendix B for more details about this result.

By randomizing the hypergraph directly, we observe some differences. Notice that using the *shuffle edges* method is equivalent to randomizing the “oriented version” of the metabolic network, so randomize regardless of the reversibility. The average complexity of the network is left unchanged, see Appendix B, while the robustness displays a pattern, shown in Fig. 3.1. We see that the robustness tends to be lower in the null model, meaning that the biological properties affect the robustness of the network. Note that, for example, the robustness of the *Staphylococcus* has been almost halved by the randomization procedure. Remember that the in-degree and out-degree are preserved. The main difference with the previous method is in the reversibility, which is not maintained. This may indicate that the robustness measure is biased towards systems that have more reversible reactions. We can check what is called the reciprocity of a graph G [63],

$$r(G) = \frac{L^{\leftrightarrow}}{L}, \quad (3.1)$$

where L^{\leftrightarrow} is the total number of bidirectional links, and L is the total number of links in the graph. The generalization to the hypergraph case is simple [64]

$$r(H) = \frac{\text{Trace}\{(\mathcal{I}_F \mathcal{I}_B^T)^2\}}{\text{Trace}\{U^2\}}, \quad (3.2)$$

where A and U are, respectively, the adjacency matrix of the directed and underlying undirected hypergraph¹. The results show no correlation between reciprocity and hypergraph robustness. The value of the reciprocity is reported in the Appendix B. We can then conclude that our measurements are not biased by reversible reactions and that the biological and chemical constraints make the metabolic network more robust compared to a random network with the same degree distribution.

¹An underlying undirected hypergraph of a directed hypergraph H is defined as the hypergraph

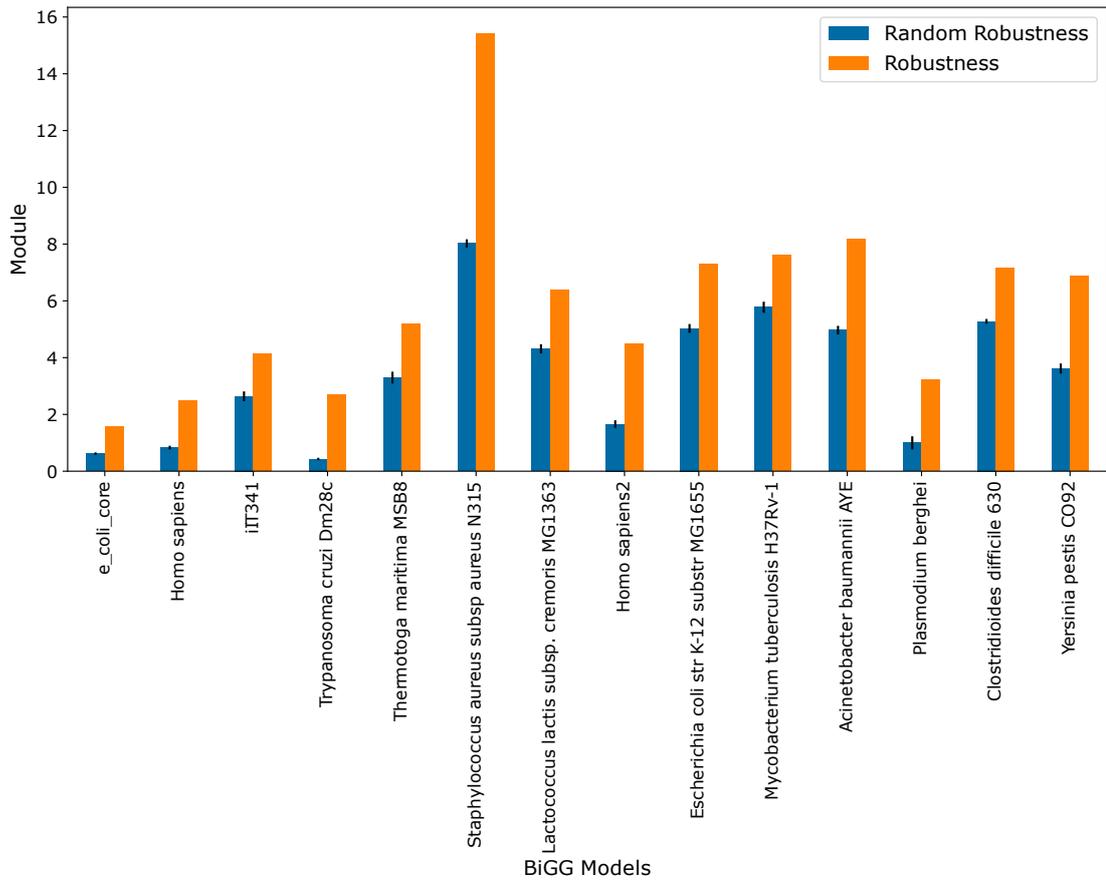


Figure 3.1: Comparison between the robustness of the real metabolic hypergraph and of the randomized one. The obtained value is an average of 5 repetitions of the randomization procedure. We consider a subset of the most interesting metabolic hypergraph.

in which each hyperedge present in H is bidirectional.

Conclusions

Metabolic networks are probably among the most challenging and promising biological networks. Their study provides insight into how biological pathways work and how resistant a given organism is to a specific environment or therapy. It is therefore critical to develop mathematical tools that can best characterize them. We propose an intuitive method to analyze metabolic hypergraphs. These are directed hypergraphs with edge-dependent vertex weights. As much as the undirected version has been explored in the literature [36, 65, 66] the theory of directed hypergraphs is lacking [37]. We propose a representation of it as a bipartite graph, define its adjacency, and analyze its structural properties by measuring the communicability and complexity of the network. A brief schematic summary of the work follows.

In chapter 1 we describe hypergraphs from a mathematical point of view. We focus on the different ways of projecting hypergraphs onto graphs and show their advantages and disadvantages. We show that the most flexible representation is the bipartite representation, and for this reason, we choose it as the basis for generalizing aspects of network theory onto hypergraphs. We compare different types of random walks and Laplacians matrices, laying the foundation for describing characterization methods. We choose to generalize communicability, to identify how different parts of the metabolism are connected, and to measure robustness. In addition, we measure the complexity of the hypergraph using probability tools: search information is defined as a kind of entropy and allows us to identify more or less easily accessible areas within the hypergraph. So far, the discussion is completely general. In the chapter 2, we apply the developed mathematical concepts to metabolic networks taken from the BiGG Database [56, 57]. We show how these can be considered EDVW hypergraphs without any loss of information. The results are discussed in detail in the paragraph below. In chapter 3 we propose two ways to define a null model for a metabolic hypergraph. This way, we can understand whether our results are derived from chemical and biological constraints or not.

We are thus able to identify important reactions and hubs in the network and characterize the complexity of different metabolic pathways. In particular, we compare several metabolic hypergraphs and obtain satisfactory results. The

complexity of the different organisms studied is almost invariant. This result is in line with the literature [60], since as varied as organisms may be, evolution has caused metabolic reactions and pathways to be similar across species. In contrast, the robustness of the network varies significantly from organism to organism, presenting a distinct peak for *Staphylococcus aureus*, an antibiotic-resistant bacterium. This result is not accidental. Antibiotics work by targeting specific reactions, inhibiting them, and causing organism death [61]. The more robust a network is, the more difficult it will be to individuate a reaction that leads to cell death because there are many alternative reactions. It would be interesting to extend this result and apply it to different antibiotic-resistant bacteria. It should be noted that antibiotics also block reactions by inhibiting the genes corresponding to the enzyme that catalyzes them. Therefore, another way for a bacterium to resist an antibiotic is to express many genes that code for the most important reactions. In this way, even if its bipartite metabolite-reaction graph is not structurally robust, the bacterium is able to endure the therapy. With these motivations, it would be interesting to use genes as hyperedges, thus obtaining a bipartite metabolite-genes graph.

As much as hypergraphs seem to be the right tool to describe metabolic networks, the mathematical theory is yet to be expanded. Network theory has several tools available that can be generalized to hypergraphs. Consider, for example, cascading failures, which seem to be strongly related to the analysis of essential reactions done via FBA. Various clustering or community detection methods could be explored and compared with the metabolic pathways. For example, Estrada in [50] proposed a community detection method based on communicability.

To conclude, we would like to emphasize that metabolic hypergraphs are a complex topic yet to be explored, and we hope to have motivated further research in this area.

Appendix A

Algebraic Connectivity

Here, in Fig. A.1, we show the results for the algebraic connectivity of the bipartite projected hypergraph Laplacian, introduced in section 1.4. We computed the first non-zero eigenvalue of the Laplacian matrix

$$L = \mathcal{D} - A^{bipartite}. \quad (\text{A.1})$$

Even if the Laplacian is not symmetric, and so the spectrum is not necessary real, the algebraic connectivity was always a real quantity.

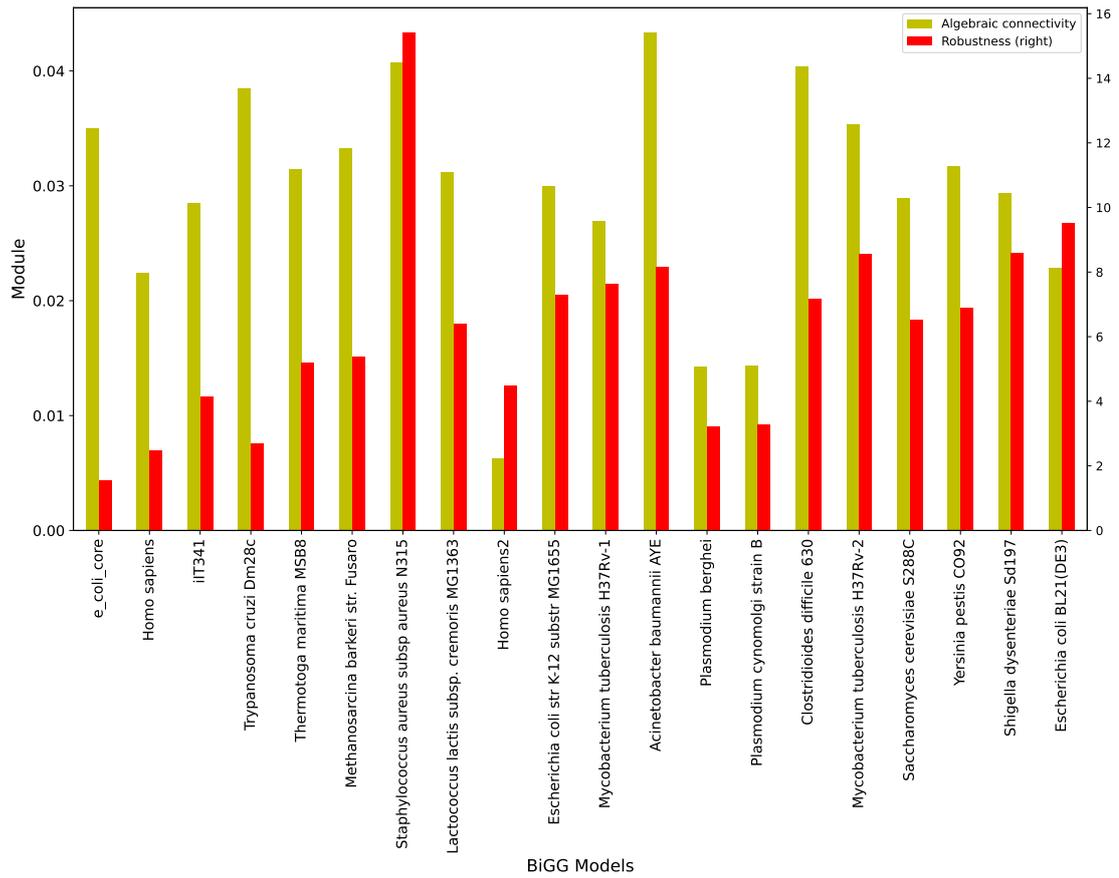


Figure A.1: The algebraic connectivity of the hypergraph on the left ax, in yellow, and the robustness on the right ax, in red.

Appendix B

Random metabolic networks: extra results

Here we report some additional results of chapter 3. In Fig. B.1 and in B.2, we show respectively the comparison between the measure of average search information and robustness for real metabolic hypergraphs and their corresponding randomized version using the method *shuffle reactions*. In Fig. B.3 we show the comparison between the measure of average search information for real metabolic hypergraphs and their corresponding randomized version using the method *shuffle edges*. In B.1 we show the value of the reciprocity. In addition, we also provide the fraction of reversible reaction for each metabolic network.

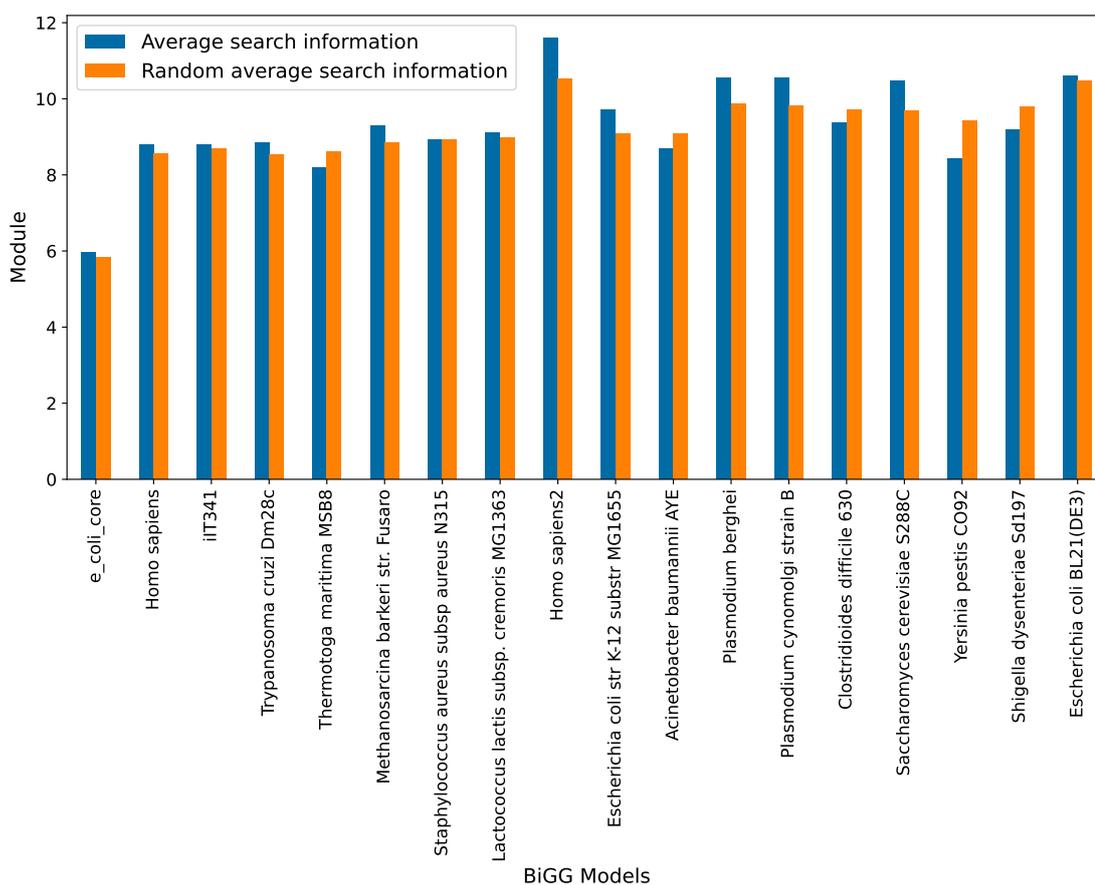


Figure B.1: The average search information for some metabolic hypergraphs compared with the average search information of their randomized counterparts. The method used to randomize the hypergraph is *shuffle reactions*. We did $100 \times M$ shuffles, where M is the number of reactions.

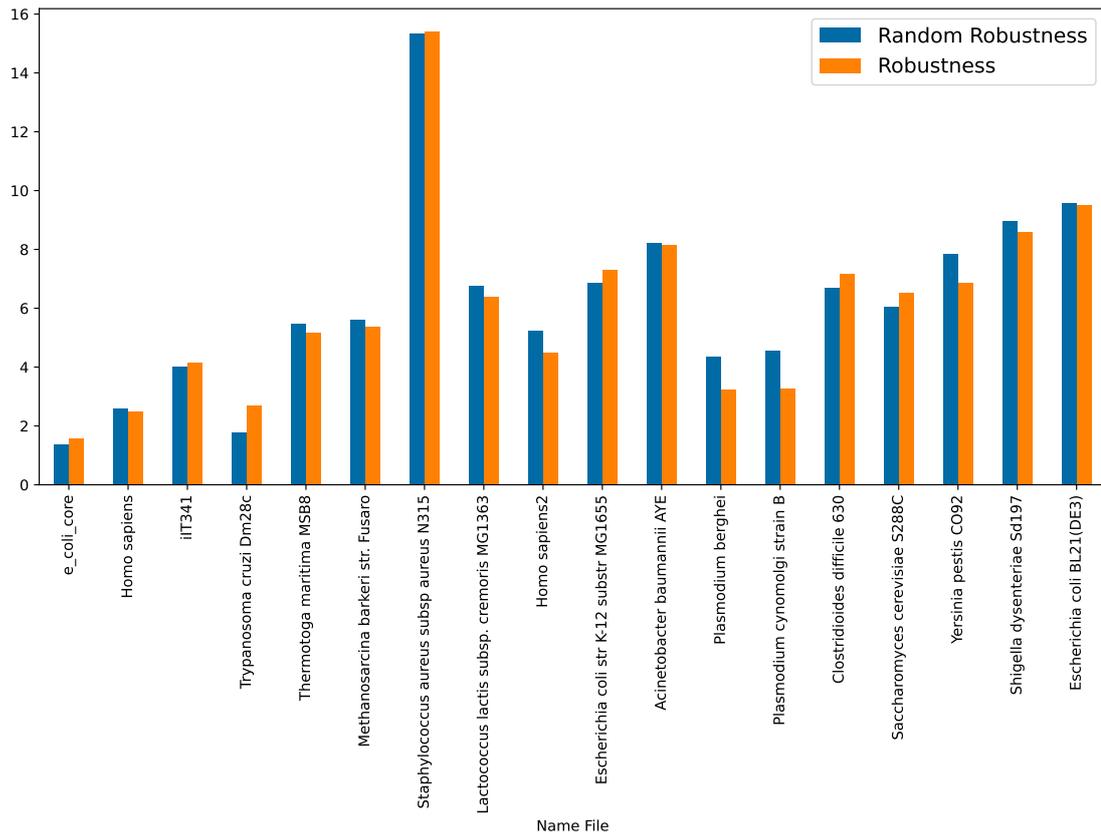


Figure B.2: The robustness for some metabolic hypergraphs compared with the robustness of their randomized counterparts. The method used to randomize the hypergraph is *shuffle reactions*. We did $100 \times M$ shuffles, where M is the number of reactions.

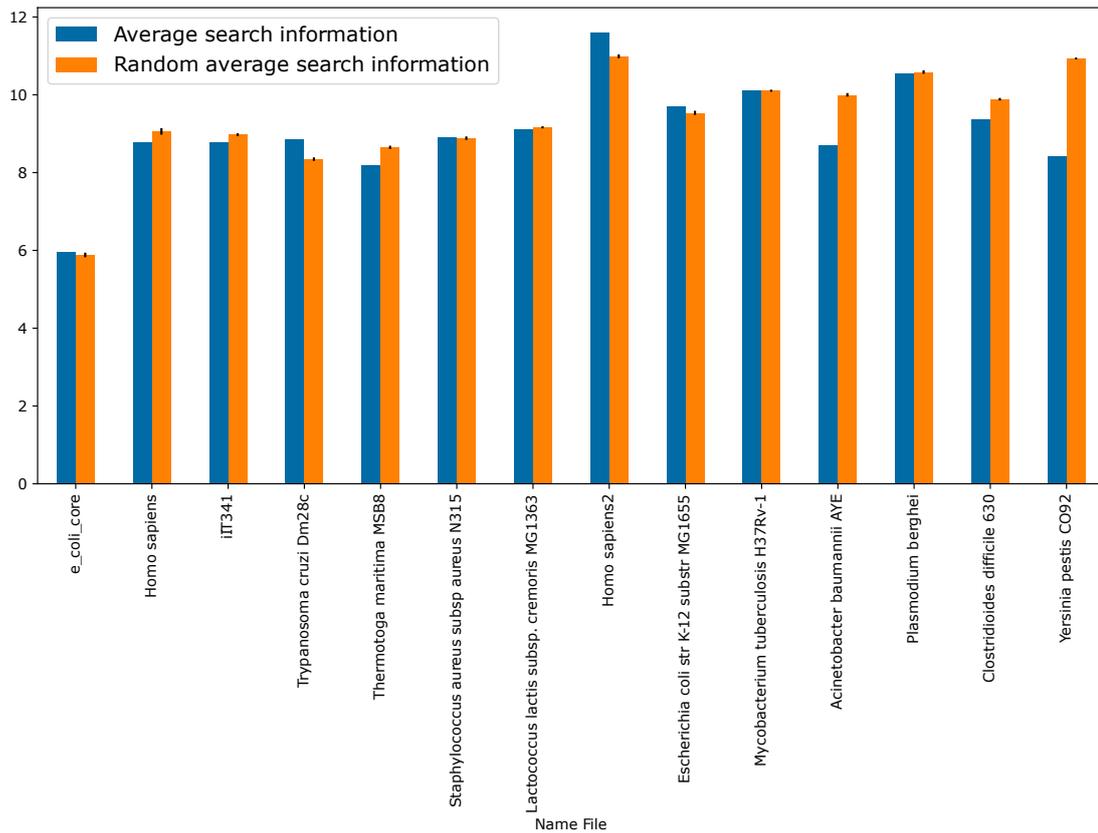


Figure B.3: The average search information for some metabolic hypergraphs compared with the average search information of their randomized counterparts. The method used to randomize the hypergraph is *shuffle edges*. We did 5 repetitions and $100 \times M$ shuffles, where M is the number of reactions.

	Name File	fraction of reversible reactions	reciprocity
0	e_coli_core	0.484211	0.425000
1	Homo sapiens	0.375267	0.289474
2	iIT341	0.330325	0.278738
3	Thermotoga maritima MSB8	0.306748	0.257384
4	Trypanosoma cruzi Dm28c	0.552987	0.438191
5	Staphylococcus aureus subsp aureus N315	0.309556	0.368752
6	Escherichia coli str K-12 substr MG1655	0.236279	0.212747
7	Homo sapiens2	0.443452	0.322267
8	Escherichia coli BL21(DE3)	0.239416	0.174967
9	Methanosarcina barkeri str. Fusaro	0.304348	0.260189
10	Acinetobacter baumannii AYE	0.417734	0.328678
11	Lactococcus lactis subsp. cremoris MG1363	0.362069	0.342130
12	Clostridioides difficile 630	0.195281	0.167685
13	Yersinia pestis CO92	0.279449	0.206161
14	Mycobacterium tuberculosis H37Rv-1	0.267317	0.215523
15	Mycobacterium tuberculosis H37Rv-2	0.226754	0.184243
16	Plasmodium berghei	0.456420	0.317741
17	Plasmodium cynomolgi strain B	0.455307	0.316309
18	Saccharomyces cerevisiae S288C	0.344392	0.295849
20	Shigella dysenteriae Sd197	0.257188	0.190624

Table B.1: Table of the fraction of reversible reactions and the reciprocity for each metabolic hypergraph studied.

Bibliography

- [1] S Boccaletti, V Latora, Y Moreno, M Chavez, and D Hwang. «Complex networks: Structure and dynamics». en. In: *Physics Reports* 424.4-5 (Feb. 2006), pp. 175–308. ISSN: 03701573. DOI: 10.1016/j.physrep.2005.10.009. URL: <https://linkinghub.elsevier.com/retrieve/pii/S037015730500462X> (visited on 10/13/2022) (cit. on p. 1).
- [2] Luciano da Fontoura Costa, Osvaldo N. Oliveira, Gonzalo Travieso, Francisco Aparecido Rodrigues, Paulino Ribeiro Villas Boas, Lucas Antiqueira, Matheus Palhares Viana, and Luis Enrique Correa Rocha. «Analyzing and modeling real-world phenomena with complex networks: a survey of applications». en. In: *Advances in Physics* 60.3 (June 2011), pp. 329–412. ISSN: 0001-8732, 1460-6976. DOI: 10.1080/00018732.2011.572452. (Visited on 10/13/2022) (cit. on p. 1).
- [3] L. da F. Costa, F. A. Rodrigues, G. Travieso, and P. R. Villas Boas. «Characterization of complex networks: A survey of measurements». en. In: *Advances in Physics* 56.1 (Jan. 2007), pp. 167–242. ISSN: 0001-8732, 1460-6976. DOI: 10.1080/00018730601170527. (Visited on 10/13/2022) (cit. on p. 1).
- [4] M. E. J. Newman. *Networks*. Second edition. Oxford, United Kingdom ; New York, NY, United States of America: Oxford University Press, 2018. ISBN: 9780198805090 (cit. on p. 1).
- [5] Judith E. Dayhoff. *Neural Network Architectures: An Introduction*. USA: Van Nostrand Reinhold Co., 1990. ISBN: 0442207441 (cit. on p. 1).
- [6] Xiang Niu, Casey Doyle, Gyorgy Korniss, and Boleslaw K. Szymanski. «The impact of variable commitment in the Naming Game on consensus formation». In: *Scientific Reports* 7.1 (Feb. 2017), p. 41750 (cit. on p. 1).
- [7] Damon Centola, Joshua Becker, Devon Brackbill, and Andrea Baronchelli. «Experimental evidence for tipping points in social convention». In: *Science* 360.6393 (2018), pp. 1116–1119. ISSN: 0036-8075. DOI: 10.1126/science.aas8827 (cit. on p. 1).

- [8] Andrea Baronchelli. «The Emergence of Consensus: A Primer». In: *Royal Society Open Science* 5 (Mar. 2018), p. 172189 (cit. on p. 1).
- [9] Austin R. Benson, Rediet Abebe, Michael T. Schaub, Ali Jadbabaie, and Jon Kleinberg. «Simplicial closure and higher-order link prediction». In: *Proceedings of the National Academy of Sciences* (2018). ISSN: 0027-8424. DOI: 10.1073/pnas.1800683115 (cit. on p. 1).
- [10] Iacopo Iacopini, Giovanni Petri, Alain Barrat, and Vito Latora. «Simplicial models of social contagion». In: *Nature Communications* 10.1 (2019), pp. 1–9 (cit. on p. 1).
- [11] Guilherme Ferraz de Arruda, Giovanni Petri, and Yamir Moreno. «Social contagion models on hypergraphs». In: *Physical Review Research* 2.2 (Apr. 2020), p. 023032. DOI: 10.1103/PhysRevResearch.2.023032. (Visited on 06/21/2022) (cit. on pp. 1, 23).
- [12] Nicholas W Landry and Juan G Restrepo. «The effect of heterogeneity on hypergraph contagion models». In: *Chaos* 30.10 (2020), p. 103117 (cit. on p. 1).
- [13] Alain Barrat, Guilherme Ferraz de Arruda, Iacopo Iacopini, and Yamir Moreno. *Social contagion on higher-order structures*. 2021. arXiv: 2103.03709 [physics.soc-ph] (cit. on p. 1).
- [14] Guilherme Ferraz de Arruda, Michele Tizzani, and Yamir Moreno. «Phase transitions and stability of dynamical processes on hypergraphs». In: *Communications Physics* 4.1 (Feb. 2021), p. 24. ISSN: 2399-3650. DOI: 10.1038/s42005-021-00525-3 (cit. on p. 1).
- [15] Unai Alvarez-Rodriguez, Federico Battiston, Guilherme Ferraz de Arruda, Yamir Moreno, Matjaž Perc, and Vito Latora. «Evolutionary dynamics of higher-order interactions in social networks». In: *Nature Human Behaviour* 5.5 (May 2021), pp. 586–595 (cit. on p. 1).
- [16] Leonie Neuhäuser, Andrew Mellor, and Renaud Lambiotte. «Multibody interactions and nonlinear consensus dynamics on networked systems». In: *Phys. Rev. E* 101 (3 Mar. 2020), p. 032310. DOI: 10.1103/PhysRevE.101.032310. URL: <https://link.aps.org/doi/10.1103/PhysRevE.101.032310> (cit. on p. 1).
- [17] Ágnes Bodó, Gyula Y. Katona, and Péter L. Simon. «SIS Epidemic Propagation on Hypergraphs». In: *Bulletin of Mathematical Biology* 78.4 (Apr. 2016), pp. 713–735 (cit. on p. 1).
- [18] Desmond J. Higham and Henry-Louis de Kergorlay. *Mean Field Analysis of Hypergraph Contagion Model*. 2021. arXiv: 2108.05451 [math.DS] (cit. on p. 1).

- [19] Desmond J. Higham and Henry-Louis de Kergorlay. «Epidemics on hypergraphs: spectral thresholds for extinction». In: *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences* 477.2252 (2021), p. 20210232 (cit. on p. 1).
- [20] Ian Stewart, Martin Golubitsky, and Marcus Pivato. «Symmetry Groupoids and Patterns of Synchrony in Coupled Cell Networks». In: *SIAM Journal on Applied Dynamical Systems* 2.4 (2003), pp. 609–646. DOI: 10.1137/S1111111103419896 (cit. on p. 1).
- [21] Martin Golubitsky, Ian Stewart, and Andrei Török. «Patterns of Synchrony in Coupled Cell Networks with Multiple Arrows». In: *SIAM Journal on Applied Dynamical Systems* 4.1 (2005), pp. 78–100. DOI: 10.1137/040612634 (cit. on p. 1).
- [22] Martin Golubitsky and Ian Stewart. «Nonlinear dynamics of networks: the groupoid formalism». In: *Bull. Amer. Math. Soc* 43 (2006), pp. 305–364 (cit. on p. 1).
- [23] Shan Yu, Hongdian Yang, Hiroyuki Nakahara, Gustavo S. Santos, Danko Nikolić, and Dietmar Plenz. «Higher-Order Interactions Characterized in Cortical Activity». In: *Journal of Neuroscience* 31.48 (2011), pp. 17514–17526. ISSN: 0270-6474. DOI: 10.1523/JNEUROSCI.3127-11.2011. URL: <https://www.jneurosci.org/content/31/48/17514> (cit. on p. 1).
- [24] Eyal Bairey, Eric D. Kelsic, and Roy Kishony. «High-order species interactions shape ecosystem diversity». In: *Nature communications* 7 (Aug. 2016), pp. 12285–12285. ISSN: 2041-1723. DOI: 10.1038/ncomms12285. URL: <https://doi.org/10.1038/ncomms12285> (cit. on p. 1).
- [25] Federico Battiston et al. «The physics of higher-order interactions in complex systems». In: *Nature Physics* 17.10 (Oct. 2021), pp. 1093–1098 (cit. on p. 1).
- [26] Alba Cervantes-Loreto, Carolyn A. Ayers, Emily K. Dobbs, Berry J. Brosi, and Daniel B. Stouffer. «The context dependency of pollinator interference: How environmental conditions and co-foraging species impact floral visitation». In: *Ecology Letters* 24.7 (2021), pp. 1443–1454. DOI: <https://doi.org/10.1111/ele.13765>. URL: <https://onlinelibrary.wiley.com/doi/abs/10.1111/ele.13765> (cit. on p. 1).
- [27] Jacopo Grilli, György Barabás, Matthew J. Michalska-Smith, and Stefano Allesina. «Higher-order interactions stabilize dynamics in competitive network models». In: *Nature* 548.7666 (Aug. 2017), pp. 210–213. ISSN: 1476-4687. DOI: 10.1038/nature23273. (Visited on 06/21/2022) (cit. on pp. 1, 23).

- [28] Renaud Lambiotte, Martin Rosvall, and Ingo Scholtes. «From networks to optimal higher-order models of complex systems». In: *Nature Physics* 15.4 (2019), pp. 313–320. ISSN: 1745-2481. DOI: 10.1038/s41567-019-0459-y (cit. on p. 1).
- [29] Ana P. Millán, Joaquín J. Torres, and Ginestra Bianconi. «Explosive Higher-Order Kuramoto Dynamics on Simplicial Complexes». In: *Phys. Rev. Lett.* 124 (21 May 2020), p. 218301. DOI: 10.1103/PhysRevLett.124.218301 (cit. on p. 1).
- [30] Michael T. Schaub, Austin R. Benson, Paul Horn, Gabor Lippner, and Ali Jadbabaie. «Random Walks on Simplicial Complexes and the Normalized Hodge 1-Laplacian». In: *SIAM Review* 62.2 (2020), pp. 353–391. DOI: 10.1137/18M1201019 (cit. on p. 1).
- [31] Guilherme Ferraz de Arruda, Giovanni Petri, Pablo Martín Rodríguez, and Yamir Moreno. *Multistability, intermittency and hybrid transitions in social contagion models on hypergraphs*. 2021. arXiv: 2112.04273 [physics.soc-ph] (cit. on p. 1).
- [32] Federico Battiston, Giulia Cencetti, Iacopo Iacopini, Vito Latora, Maxime Lucas, Alice Patania, Jean-Gabriel Young, and Giovanni Petri. «Networks beyond pairwise interactions: Structure and dynamics». In: *Physics Reports* 874 (2020). Networks beyond pairwise interactions: Structure and dynamics, pp. 1–92. ISSN: 0370-1573 (cit. on pp. 1, 6).
- [33] C. Berge. *Hypergraphs: Combinatorics of Finite Sets*. North-Holland Mathematical Library. Elsevier Science, 1984. ISBN: 9780080880235 (cit. on p. 1).
- [34] Timoteo Carletti, Federico Battiston, Giulia Cencetti, and Duccio Fanelli. «Random walks on hypergraphs». In: *Phys. Rev. E* 101 (2 Feb. 2020), p. 022308. DOI: 10.1103/PhysRevE.101.022308 (cit. on pp. 1, 6).
- [35] Anirban Banerjee. «On the spectrum of hypergraphs». In: *Linear Algebra and its Applications* 614 (2021). Special Issue ILAS 2019, pp. 82–110. ISSN: 0024-3795. DOI: <https://doi.org/10.1016/j.laa.2020.01.012> (cit. on pp. 1, 6).
- [36] Uthsav Chitra and Benjamin Raphael. «Random Walks on Hypergraphs with Edge-Dependent Vertex Weights». In: *Proceedings of the 36th International Conference on Machine Learning*. Ed. by Kamalika Chaudhuri and Ruslan Salakhutdinov. Vol. 97. Proceedings of Machine Learning Research. PMLR, Sept. 2019, pp. 1172–1181. URL: <https://proceedings.mlr.press/v97/chitra19a.html> (cit. on pp. 1, 2, 6, 12, 15, 35).

- [37] Adèle Weber Zendrera, Nataliya Sokolovska, and Hédi A. Soula. «Robust structure measures of metabolic networks that predict prokaryotic optimal growth temperature». In: *BMC Bioinformatics* 20.1 (Oct. 2019), p. 499. ISSN: 1471-2105. DOI: 10.1186/s12859-019-3112-y. (Visited on 06/22/2022) (cit. on pp. 2, 28, 35).
- [38] Nicole Pearcy, Jonathan Crofts, and Nadia Chuzhanova. «Hypergraph models of metabolism». In: *Chuzhanova, International Journal of Biological, Veterinary, Agricultural and Food Engineering* 8 (Jan. 2014) (cit. on pp. 2, 8).
- [39] Andreas Dräger and Hannes Planatscher. «Metabolic Networks». In: *Encyclopedia of Systems Biology*. Ed. by Werner Dubitzky, Olaf Wolkenhauer, Kwang-Hyun Cho, and Hiroki Yokota. New York, NY: Springer New York, 2013, pp. 1249–1251. ISBN: 9781441998620 9781441998637. DOI: 10.1007/978-1-4419-9863-7_1277. (Visited on 06/23/2022) (cit. on p. 2).
- [40] Roger Guimerà and Luís A. Nunes Amaral. «Functional cartography of complex metabolic networks». In: *Nature* 433.7028 (Feb. 2005), pp. 895–900. ISSN: 0028-0836, 1476-4687. DOI: 10.1038/nature03288. (Visited on 06/23/2022) (cit. on p. 2).
- [41] Hongwu Ma and An-Ping Zeng. «Reconstruction of metabolic networks from genome data and analysis of their global structure for various organisms». In: *Bioinformatics* 19.2 (Jan. 2003), pp. 270–277. ISSN: 1367-4803. DOI: 10.1093/bioinformatics/19.2.270 (cit. on p. 2).
- [42] Alina Renz, Reihaneh Mostolizadeh, and Andreas Dräger. «Clinical Applications of Metabolic Models in SBML Format». In: *Systems Medicine*. Ed. by Olaf Wolkenhauer. Oxford: Academic Press, 2021, pp. 362–371. ISBN: 978-0-12-816078-7. DOI: <https://doi.org/10.1016/B978-0-12-801238-3.11524-7> (cit. on p. 2).
- [43] Mariano Beguerisse-Díaz, Gabriel Bosque, Diego Oyarzún, Jesús Picó, and Mauricio Barahona. «Flux-dependent graphs for metabolic networks». In: *npj Systems Biology and Applications* 4.1 (Aug. 2018), pp. 1–14. ISSN: 2056-7189. DOI: 10.1038/s41540-018-0067-y. (Visited on 06/13/2022) (cit. on pp. 2, 21, 22).
- [44] Anton Eriksson, Daniel Edler, Alexis Rojas, Manlio de Domenico, and Martin Rosvall. «How choosing random-walk model and network representation matters for flow-based community detection in hypergraphs». In: *Communications Physics* 4.1 (June 2021), pp. 1–12. ISSN: 2399-3650. DOI: 10.1038/s42005-021-00634-z. (Visited on 06/08/2022) (cit. on pp. 8, 12).

- [45] Raffaella Mulas and Dong Zhang. «Spectral theory of Laplace operators on oriented hypergraphs». In: *Discrete Mathematics* 344.6 (June 2021), p. 112372. ISSN: 0012-365X. DOI: 10.1016/j.disc.2021.112372. (Visited on 06/08/2022) (cit. on pp. 8, 15).
- [46] Jürgen Jost and Raffaella Mulas. «Hypergraph Laplace operators for chemical reaction networks». In: *Advances in Mathematics* 351 (July 2019), pp. 870–896. ISSN: 0001-8708. DOI: 10.1016/j.aim.2019.05.025. (Visited on 06/08/2022) (cit. on pp. 8, 15, 21).
- [47] Adolf Fick. «Ueber Diffusion». In: *Annalen der Physik* 170.1 (1855), pp. 59–86. DOI: <https://doi.org/10.1002/andp.18551700105>. eprint: <https://onlinelibrary.wiley.com/doi/pdf/10.1002/andp.18551700105>. URL: <https://onlinelibrary.wiley.com/doi/abs/10.1002/andp.18551700105> (cit. on p. 13).
- [48] Piet van Mieghem. *Graph Spectra for Complex Networks*. Cambridge University Press, 2010. DOI: 10.1017/CB09780511921681 (cit. on p. 14).
- [49] Chai Wah Wu. «Algebraic connectivity of directed graphs». en. In: *Linear and Multilinear Algebra* 53.3 (June 2005), pp. 203–223. ISSN: 0308-1087, 1563-5139. DOI: 10.1080/03081080500054810 (cit. on p. 14).
- [50] Ernesto Estrada, Naomichi Hatano, and Michele Benzi. «The physics of communicability in complex networks». In: *Physics Reports. The Physics of Communicability in Complex Networks* 514.3 (May 2012), pp. 89–119. ISSN: 0370-1573. DOI: 10.1016/j.physrep.2012.01.006. (Visited on 06/05/2022) (cit. on pp. 16, 36).
- [51] M. Rosvall, A. Trusina, P. Minnhagen, and K. Sneppen. «Networks and Cities: An Information Perspective». In: *Physical Review Letters* 94.2 (Jan. 2005), p. 028701. DOI: 10.1103/PhysRevLett.94.028701. (Visited on 06/05/2022) (cit. on pp. 16, 18).
- [52] Ernesto Estrada and Naomichi Hatano. «Communicability in complex networks». In: *Phys. Rev. E* 77 (3 Mar. 2008), p. 036111. DOI: 10.1103/PhysRevE.77.036111. URL: <https://link.aps.org/doi/10.1103/PhysRevE.77.036111> (cit. on p. 16).
- [53] Wu Jun, Mauricio Barahona, Tan Yue-Jin, and Deng Hong-Zhong. «Natural Connectivity of Complex Networks». In: *Chinese Physics Letters* 27.7 (July 2010), p. 078902. ISSN: 0256-307X, 1741-3540. DOI: 10.1088/0256-307X/27/7/078902. (Visited on 06/25/2022) (cit. on p. 16).

- [54] Jeffrey D. Orth, R. M. T. Fleming, and Bernhard ø. Palsson. «Reconstruction and Use of Microbial Metabolic Networks: the Core *Escherichia coli* Metabolic Model as an Educational Guide». In: *EcoSal Plus* 4.1 (Jan. 2010). Ed. by Peter D. Karp, ecosalplus.10.2.1. ISSN: 2324-6200. DOI: 10.1128/ecosalplus.10.2.1. (Visited on 06/23/2022) (cit. on p. 22).
- [55] Jeffrey D. Orth, Ines Thiele, and Bernhard ø. Palsson. «What is flux balance analysis?» In: *Nature biotechnology* 28.3 (Mar. 2010), pp. 245–248. ISSN: 1087-0156. DOI: 10.1038/nbt.1614. (Visited on 06/15/2022) (cit. on pp. 22, 27, 29).
- [56] Zachary A. King, Justin Lu, Andreas Dräger, Philip Miller, Stephen Federowicz, Joshua A. Lerman, Ali Ebrahim, Bernhard O. Palsson, and Nathan E. Lewis. «BiGG Models: A platform for integrating, standardizing and sharing genome-scale models». In: *Nucleic Acids Research* 44.D1 (Oct. 2015), pp. D515–D522. ISSN: 0305-1048. DOI: 10.1093/nar/gkv1049 (cit. on pp. 24, 35).
- [57] *BiGG models*. <http://bigg.ucsd.edu/models>. Accessed: 2022-06-20 (cit. on pp. 24, 35).
- [58] Ali Ebrahim, Joshua A Lerman, Bernhard O Palsson, and Daniel R Hyduke. «COBRAPy: COntstraints-Based Reconstruction and Analysis for Python». en. In: *BMC Systems Biology* 7.1 (Dec. 2013), p. 74. ISSN: 1752-0509. DOI: 10.1186/1752-0509-7-74. (Visited on 10/13/2022) (cit. on p. 27).
- [59] Elliot Rowe, Bernhard O. Palsson, and Zachary A. King. «Escher-FBA: a web application for interactive flux balance analysis». en. In: *BMC Systems Biology* 12.1 (Dec. 2018), p. 84. ISSN: 1752-0509. DOI: 10.1186/s12918-018-0607-5. URL: <https://bmcsystbiol.biomedcentral.com/articles/10.1186/s12918-018-0607-5> (visited on 09/21/2022) (cit. on p. 28).
- [60] H. Jeong, B. Tombor, R. Albert, Z. N. Oltvai, and A.-L. Barabási. «The large-scale organization of metabolic networks». In: *Nature* 407.6804 (Oct. 2000), pp. 651–654. ISSN: 1476-4687. DOI: 10.1038/35036627. (Visited on 06/20/2022) (cit. on pp. 29, 36).
- [61] Neil A. Campbell, Lisa A. Urry, Michael L. Cain, Steven A. Wasserman, Peter V. Minorsky, and Jane B. Reece. *Biology: a global approach*. Eleventh edition, global edition. New York: Pearson, 2018. ISBN: 9781292170435 (cit. on pp. 29, 36).
- [62] Scott A. Becker and Bernhard ø Palsson. «Genome-scale reconstruction of the metabolic network in *Staphylococcus aureus* N315: an initial draft to the two-dimensional annotation». In: *BMC microbiology* 5 (Mar. 2005), p. 8. ISSN: 1471-2180. DOI: 10.1186/1471-2180-5-8 (cit. on p. 29).

- [63] Diego Garlaschelli and Maria I. Loffredo. «Patterns of Link Reciprocity in Directed Networks». In: *Phys. Rev. Lett.* 93 (26 Dec. 2004), p. 268701. DOI: 10.1103/PhysRevLett.93.268701 (cit. on p. 33).
- [64] Nicole Pearcy, Jonathan Crofts, and Nadia Chuzhanova. «Hypergraph models of metabolism». In: *Chuzhanova, International Journal of Biological, Veterinary, Agricultural and Food Engineering* 8 (Jan. 2014) (cit. on p. 33).
- [65] Yu Zhu, Boning Li, and Santiago Segarra. «Hypergraphs with Edge-Dependent Vertex Weights: Spectral Clustering Based on the 1-Laplacian». In: *ICASSP 2022 - 2022 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. Singapore, Singapore: IEEE, May 2022, pp. 8837–8841. ISBN: 9781665405409. DOI: 10.1109/ICASSP43922.2022.9746363 (cit. on p. 35).
- [66] Yu Zhu and Santiago Segarra. «Hypergraph Cuts with Edge-Dependent Vertex Weights». In: *arXiv:2201.06084 [cs]* (Jan. 2022). arXiv: 2201.06084 (cit. on p. 35).

Funders

The author acknowledge the financial support of Soremartec S.A. and Soremartec Italia, Ferrero Group. The funders had no role in study design, data collection, and analysis, decision to publish, or preparation of the manuscript.