UNIVERSITÀ DEGLI STUDI DI TRIESTE

XXXVII CICLO DEL DOTTORATO DI RICERCA IN

Scienze della terra, Fluidodinamica e Matematica. Interazioni e Metodiche

Combinatorial Aspects of Discrete Structures – Graphs and Latin Squares

Settore scientifico-disciplinare: INFO-01/A

DOTTORANDO / A
Giuliamaria Menara

COORDINATORE
PROF. Stefano Maset

SUPERVISORE DI TESI
PROF. Luca Manzoni

CO-SUPERVISORE DI TESI
PROF. Chad Giusti

ANNO ACCADEMICO 2023/2024

GIULIAMARIA MENARA

COMBINATORIAL ASPECTS OF DISCRETE STRUCTURES – GRAPHS AND LATIN SQUARES

Combinatorics is the art and science
of distilling a complex mathematical structure
into simple attributes,
and developing from this
a deeper understanding of the original structure.

– Josephine Yu, (2016)

ABSTRACT

Discrete structures are fundamental to both mathematics and computer science, serving as the basis for numerous computational and algorithmic frameworks. This dissertation explores two key areas where combinatorial methods provide deep insights into the properties and applications of discrete structures: combinatorial topology in graph theory and cellular automata (CA)-based Latin squares in cryptography.

In the first part, we investigate eulerian magnitude homology, a novel extension of standard magnitude homology, and its application to Erdős-Rényi random graphs and random geometric graphs. We develop new tools, including the eulerian Asao-Izumihara complex, to study torsion in these homology groups and establish conditions under which the eulerian magnitude homology of Erdős-Rényi random graphs is torsion-free. Additionally, we introduce an efficient algorithm to compute the ranks of first-diagonal eulerian magnitude homology groups, proving its computational complexity.

The second part of the dissertation shifts focus to cellular automata (CA), where we explore their use in generating Latin squares through bipermutive rules. These CA-based Latin squares find applications in cryptography, particularly in designing secure cryptographic functions and correlation-immune functions for coding theory.

This thesis demonstrates how discrete mathematical structures, through combinatorial techniques, contribute to both theoretical advancements and practical applications in modern computation and cryptography.

PUBLICATIONS AND PREPRINTS

Some ideas and figures have appeared in the following preprints:

1. Chad Giusti and Giuliamaria Menara.
   Eulerian magnitude homology: subgraph structure and random graphs.
   arXiv:2403.09248 [math.CO].
   Submitted to *Discrete & Computational Geometry*.

2. Giuliamaria Menara.
   On torsion in eulerian magnitude homology of Erdos-Renyi random graphs.
   arXiv:2409.03472 [math.CO].
   Submitted to *Homology, Homotopy and Applications*.

3. Giuliamaria Menara and Luca Manzoni.
   Computing eulerian magnitude homology.
   arXiv:2410.10376 [cs.CC].
   Submitted to *Acta Infomatica*.

Other publications realized during the doctorate are the followings:

1. Alberto Dennunzio, Enrico Formenti, Luca Manzoni, Luciano Margara, Giuliamaria Menara.
   A topology for P systems with active membranes.
   In *Journal of Membrane Computing*, pages 1–12, 2023.

2. Gloria Pietropolli, Giuliamaria Menara, Mauro Castelli.
   A genetic programming based heuristic to simplify rugged landscapes exploration.
   In *Emerging Science Journal*, volume 7, issue 4, pages 1037–1051, 2023.

3. Mauro Castelli, Luca Manzoni, Luca Mariot, Giuliamaria Menara, and Gloria Pietropolli.
   The Effect of Multi-Generational Selection in Geometric Semantic Genetic Programming.
   In *Applied Sciences*, volume 12, issue 10, page 4836, 2022.

ACKNOWLEDGMENTS

There are many people that I want and need to thank for their support in the years of my doctorate.

Firstly, I would like to thank my advisors. Profs. Luca Manzoni and Chad Giusti have been extraordinarily generous with their attention, mentoring, advice, and thoughtful conversations. Thank you for taking a chance on me and giving the support I needed during the past three years. I would also like to thank Prof. Mauro Castelli, who was my supervisor during my research position at the NOVA University Lisbon - it has been an honor collaborating and learning from you.

Next, I would like to thank my professors at both the University of Trieste and the University of Delaware who I had the pleasure of learning from. Their expertise allowed me to grow exponentially as a mathematician and a critical thinker.

Thank you to Rocco for being such a valuable figure, pushing me to perform at the highest level I could, and allowing me to sometimes annoy you with math and life doubts (maybe more than sometimes). Thanks also to Alexandra, Annamaria, Eleonora, Jerome, Kenza, Lindsey, Luigi, Martina, Mary, Melinda, Michele, Paul, Sharon, Sonia, Teresa, Umberto, Valeria, and so many others who have been fantastic friends and colleagues throughout the years.

While having a support base inside the department is crucial, I have been honored to have such a prominent support base outside of it. To my parents, thank you for your endless love and encouragement, which have guided me throughout my academic journey. Your unwavering belief in me, even during the most challenging moments, has been a constant source of motivation. Thank you for your sacrifices, your wisdom, and for always being my biggest supporters. This accomplishment is as much yours as it is mine.

I would like to express my heartfelt gratitude to my boyfriend, Leo, for his unwavering support, patience, and understanding throughout this journey. Your humor, intelligence and constant encouragement and belief in me have been a consistent source of strength, especially during the challenging moments. Thank you for always being there to listen, offer advice, and remind me to take breaks when I needed them most. This accomplishment would not have been possible without your love and companionship.

Last, but certainly not least, I would like to thank Annetta, Denis, Eliana, Maria and Merisa. Thank you for believing in me, allowing me to be myself, and being there for me whenever I fall. You have made me a better person, and may we always strive for the highest.

To my beloved aunt, Angela,
whose love, wisdom, and guidance
have always been a beacon of light in my life.
Though you are no longer with us,
your memory continues to inspire me.

CONTENTS

LIST OF FIGURES

INTRODUCTION

Discrete structures form the bedrock of numerous mathematical and computational theories. These structures, which include sets, graphs, permutations, and sequences, are characterized by their non-continuous nature, making them inherently distinct from the continuous objects studied in calculus and real analysis. The study of discrete structures is crucial not only in pure mathematics but also in computer science, where discrete mathematics underpins algorithms, data structures, and computational complexity.

Among the various branches of mathematics, combinatorics stands out as the discipline that focuses on counting, arranging, and optimizing these discrete structures, and its significance is evident in its applications across diverse fields, including statistical mechanics, cryptography, and network theory.

Combinatorial methods offer powerful and diverse techniques for exploring the properties of discrete structures. For example, in graph theory, combinatorial approaches are used to determine the existence of certain subgraphs, to optimize network flows, or to find paths and cycles. These problems, while seemingly simple, are often computationally complex, making combinatorics an essential tool in both the theoretical and applied setting. The importance of combinatorics is further underscored by its connections to other mathematical areas. For instance, combinatorial designs are used in statistics, while combinatorial group theory has applications in topology and geometry. Combinatorics also plays a crucial role in computer science, especially in areas like algorithm design, cryptography, and network theory. Latin squares, a specific combinatorial structure, are widely used in applications such as error detection and correction, cryptographic algorithms, and parallel processing. In cryptography, for example, Latin squares help create secure ciphers by ensuring structured, non-repetitive combinations. Additionally, in experimental design and database theory, they assist in minimizing error and bias by providing systematic arrangements of variables.

Considering the points discussed, it is evident that combinatorics has extensive applicability across a wide array of fields, and in this dissertation our objective is to explore and offer new perspectives on the various areas where combinatorics can be applied, as well as innovative approaches to how these applications can be implemented.

PART 1: EULERIAN MAGNITUDE HOMOLOGY

In the first part part of this dissertation our primary focus will be on the connection between graph theory and homology.

Graph theory has emerged as a central area of study within combinatorics. Originating with Euler's solution to the Königsberg bridge problem in 1736, graph theory has grown into a rich field that explores the properties of networks.

Homology, a tool originally developed in algebraic topology to study the shapes of spaces, provides a way to associate algebraic objects, such as groups, with topological spaces, revealing information about their structure. The application of homology to graphs, particularly in the form of *magnitude homology*, is a relatively recent development that offers new insights into the combinatorial structure of graphs.

Magnitude is a concept introduced by Leinster in [81] as a measurement of the "size" of a metric space analogous to the Euler characteristic of a category [80]. Magnitude provides a measure of size in the sense that it satisfies

$$\text{Size}(A \cup B) = \text{Size}(A) + \text{Size}(B) - \text{Size}(A \cap B)$$

$$\text{Size}(A \times B) = \text{Size}(A) \times \text{Size}(B),$$

in analogy with other common notions of size like cardinality of sets, volumes of subsets of $\mathbb{R}^n$, dimensions of vector spaces, and Euler characteristic of topological spaces. Because it is intimately related to Euler characteristic, it is natural to consider the categorification of magnitude to a homology theory, as described in [62, 85]. Homology theories for metric spaces are of general interest to the applied topology community, as evidenced by the success of persistent homology. The case of graphs equipped with the path metric is of particular utility, as new methods for interpreting and understanding the structure of networks are in general demand.

Explicitly, the magnitude chain complex of a simple graph is a bigraded complex described in terms of k-*trails*, lists of $(k+1)$ *landmark* vertices in the graph. A k-trail is of length ℓ if there is a walk in the graph of of length ℓ which passes through the listed landmarks in order, and this walk is minimal in length among all such. The differential in this chain complex deletes individual landmarks, with non-zero coefficient precisely when the deletion does not change the length of the trail. Such nonzero terms thus indicate the corresponding landmark is non-essential; a trail following the other landmarks must pass through at least as many vertices, even without this instruction. One can think of this as a form of "non-convexity", as there are no shorter paths between some pair of landmarks [83]. However, beyond this term-wise intuition, the structure of magnitude homol-

ogy groups remains hard to interpret. Recent developments in this direction are achieved in [5].

In this thesis, we will make some progress toward elucidating the connection between magnitude homology of simple graphs equipped with the path metric and their combinatorial structure. In 2018 Gu confirmed the existence of graphs with the same magnitude but different magnitude homology [54], while Kaneta and Yoshinaga have analyzed the structure and implications of torsion in magnitude homology [69]. Torsion in the magnitude homology of graphs was also studied by Sazdanovic and Summers in [132] and by Caputi and Collari in [20]. In [7] the authors explore the interaction between the magnitude homology of a graph and its diagonality and girth. Tajima and Yoshinaga investigate the connection between the homotopy type of the Asao-Izumihara CW complex (c.f. [6]) and the diagonality of magnitude homology groups, [143]. As demonstrated by Cho in [27], magnitude and persistent homology can be thought of as endpoints on a spectrum of such theories. Also, it is evident from [62] that calculating the magnitude homology of a graph is intricate and challenging, leading to the emergence of several technical approaches for its computation [6, 54, 132].

The approach we take in this work is to observe that a large portion of the magnitude chain complex is redundant, in the sense that the chains reflect combinatorial structure already recorded by chains of lower bigrading. To leverage this observation, in Chapter 2, we define the subcomplex of *eulerian magnitude chains*, supported on trails with no repeated landmarks. Focusing on the $k = \ell$ line, where the list of involved landmarks completely determines a trail, we develop the notion of a *structure graph* for families of chains, which captures which chains share terms in their differentials. In Theorem 2.2.2, we demonstrate that cycles in this complex can be decomposed into cycles with simple structure graphs. We leverage in this decomposition Corollary 2.2.3 to give a simple characterization of a generating set for (k, k)-graded eulerian magnitude homology of a graph. Combining this generating set with the language of *class graphs* from [152], we thus characterize which subgraphs of a graph support non-trivial cycles in $EMH_{k,k}(G)$ in terms of the corresponding structure graphs, providing a framework for computing these groups for graphs of interest.

Equipped with these observations, we consider the complementary set of trails, which must revisit a landmark, which we call the *discriminant magnitude chains*, by analogy to the study of singular maps in the study of embeddings. Leveraging the corresponding long exact sequence in homology, in Theorems 2.3.1 and 2.3.3, we observe that the eulerian magnitude homology groups of lower bifiltration control the structure of the discriminant magnitude homology groups, and when lower eulerian magnitude homology groups vanishes we obtain

a complete characterization of the generators of the (k,k)-magnitude homology groups.

In the interest of exploring what features of a graph the (eulerian) magnitude homology groups capture, we then turn our attention to two classes of random graphs: in Chapter 3, we study Erdös-Rényi random graphs and random geometric graphs on the standard torus. By understanding these "unstructured" examples, we aim to provide a backdrop for interpreting magnitude homology of "structured" graphs including those observed in applications. In each context, we derive a vanishing threshold for the limiting expected rank of the (k,k)-eulerian magnitude homology in terms of the density parameter (Theorems 3.1.4 and 3.2.6). In combination with our observation above about the relationship to discriminant magnitude homology, this provides a characterization of the structure of expected (k,k)-magnitude homology groups in this range. Further, adapting tools from [67], we develop a characterization of the limiting expected Betti numbers of the (k,k)-eulerian magnitude homology groups in terms of density (Theorems 3.1.6 and 3.2.8) and corresponding central limit theorems.

Moving into Chapter 4, we turn our attention to the presence of torsion in eulerian magnitude homology groups, and as a first step towards exploring whether graphs have torsion in their eulerian magnitude homology groups, we consider the Erdös-Rényi model for random graphs. Adapting the construction introduced by Asao and Izumihara in [6] to the context of eulerian magnitude homology, we are able to produce for every pair of vertices $(a,b) \in G$ two simplicial complexes $ET_{\leqslant \ell}(a,b)$ and $ET_{\leqslant \ell-1}(a,b)$ such that the homology of the quotient CW complex $ET_{\leqslant \ell}(a,b)/ET_{\leqslant \ell-1}(a,b)$ is isomorphic to a direct summand of the eulerian magnitude homology $EMC_{*,\ell}(G)$ up to degree shift. Therefore, producing a shellability result of the complexes $ET_{\leqslant \ell}(a,b)$ and $ET_{\leqslant \ell-1}(a,b)$ will in turn determine a torsion-free result for $EMC_{*,\ell}(G)$. In Theorem 4.3.1 we achieve such shellability result for $ET_{\leqslant \ell}(a,b)$ in terms of the density parameter of the considered Erdös-Rényi random graph G. Further, in Corollary 4.3.5 we link the torsion-free result for eulerian magnitude homology groups stated in Theorem 4.3.3 with the vanishing threshold for eulerian magnitude homology groups produced in 3.1.4, determining sufficient conditions under which if eulerian magnitude homology is non-vanishing, then it is also torsion-free.

Finally, in Chapter 5 we turn our attention to the computational aspects of eulerian magnitude homology and we propose an efficient algorithm to compute first-diagonal groups. In order to do this, we will rely on the fact (proved in Chapter 2) that searching for all $(k+1)$-tuples inducing a path of length k and generating homology is the same as looking for specific substructures in the graph G. In other words, computing the rank of eulerian magnitude homology groups

$EMH_{k,k}(G)$ is the same as enumerating subgraphs of G belonging to a certain family $\mathcal{H}$. This problem is called the *subgraph isomorphism problem* and there is an extensive literature studying its computational complexity. For example, it is shown in [59] that for any fixed simple graph G, the problem of whether there exists an isomorphism from another graph H to G is solvable in polynomial time if G is bipartite, and NP-complete if G is not bipartite. Also, the work by Dyer and Greenhill [37] proves that polynomial-time solvable cases arise only when G is an isolated vertex, a complete graph with all loops present, a complete bipartite graph without loops, or a disjoint union of these graphs. Moreover, in [4] Amini, Fomin and Saurabh relate counting subgraphs to counting graph isomorphisms. They provide exact algorithms for several problems (compute the number of optimal bandwidth permutations of a graph on n vertices excluding a fixed graph as a minor, counting all subgraphs excluding a fixed graph M as a minor, counting all subtrees with a given maximum degree) and all of them can be solved in time at least $2^{\mathcal{O}(n)}$.

In Chapter 5 of this dissertation, we will prove the intrinsic difficulty of our problem by showing that it is complete for the $\#W[1]$ complexity class. Then, we tackle this computational problem and propose a breadth-first-search-based approach to compute the eulerian magnitude homology groups $EMH_{k,k}(G)$, which results in an algorithm that is more computationally efficient than relying directly on the definition. Indeed, even if in the worst-case scenario we still have exponential computational complexity, for many graphs emerging from real-world scenarios the complexity is sub-exponential or even polynomial, as we will show in Section 5.2.2.

Outline

The first part of this dissertation is organized as follows. Chapter 1 provides a detailed background on graph theory, magnitude homology, and related concepts. This chapter serves as a foundation for the more advanced material that follows.

In Chapter 2, we introduce eulerian magnitude homology in full detail, including its definition, basic properties, and the motivation behind its development. We also discuss the relationship between eulerian and standard magnitude homology, highlighting the advantages of the former.

Chapter 3 explores the application of eulerian magnitude homology to random graphs. We derive new limit theorems and examine the asymptotic behavior of these homology groups in different random graph models.

Chapter 4 delves into the study of torsion in the eulerian magnitude homology groups of Erdös-Rényi random graph. To this end, we introduce the eulerian Asao-Izumihara complex and we produce

a vanishing threshold for a shelling of such complex. This will lead to a result establishing the regimes where eulerian magnitude homology of Erdös-Rényi random graphs is torsion-free.

Finally, Chapter 5 tackles the problem of computing the ranks of first-diagonal eulerian magnitude homology groups of a graph G. First, we analyze the computational cost of our problem and prove that it is #$W[1]$-complete. Then we develop the *first diagonal algorithm*, a breadth-first-search-based algorithm parameterized by the diameter of the graph to calculate the ranks of the homology groups of interest.

PART 2: CELLULAR AUTOMATA

In the second part of this thesis we examine *Cellular Automata (CA)*, a nature-inspired computational model. In particular, we are interested in how CA defined by bipermutive rules give rise to Latin squares, and in the possible applications of CA-based Latin squares to cryptography.

CA are one of the oldest models in Computer Science, that number among its creators some of the founders of the discipline, like von Neumann [145, 146]. They provide a discrete mathematical model where simple local rules drive the evolution of cells within a grid over time and, as one of the oldest method, they have long been studied for their computational abilities [119], for their link with language theory [136], and as a discrete time dynamical system [79].

More specifically, a CA consists of a regular grid of cells, each of which can exist in a finite number of states. The evolution of the CA is governed by rules that determine the next state of each cell based on the current state of its neighboring cells. This local rule is typically uniform across the grid and is applied to each cell simultaneously at discrete time steps. Formally, a one-dimensional CA can be defined as a tuple $\mathcal{A} = (S, N, f)$ where:

- S is a finite set of states, often represented as $\{0, 1\}$ (binary states),
- N is the neighborhood of the cell, often represented as the set of cells adjacent to the current cell,
- $f : S^N \to S$ is the local transition function, which specifies how a cell's state is updated based on the states of its neighbors.

Wolfram in [147] classified CA into four classes based on their long-term behavior:

1. Evolves to a homogeneous state (e.g., all cells reach the same state).
2. Evolves into a set of simple stable or periodic patterns.

3. Produces chaotic, seemingly random behavior.
4. Exhibits complex structures, which may evolve indefinitely or perform computation, akin to Turing machines.

Class 4, in particular, has captured much attention due to its connection with universal computation. Cellular automata such as Rule 110 and Conway's Game of Life are examples of systems capable of universal computation.

Bipermutive cellular automata are a special class of CA where the local rule, also known as the update function, exhibits specific symmetries. More formally, a CA is said to be bipermutive if its transition function f is both left and right permutive. This means that the function depends on the extremal (leftmost and rightmost) states of the neighboring cells in a bijective (invertible) manner.

Formally, a one-dimensional cellular automaton rule f is called left-permutive if, for all configurations of the middle and right neighboring cells, the map that assigns the left neighboring cell to the next state of the current cell is bijective. Similarly, f is right-permutive if the same condition holds for the right neighboring cell. Thus, bipermutive cellular automata satisfy the following condition for their local rule $f : S^N \to S$: for every fixed configuration of the middle cells, the dependence on both the leftmost and rightmost neighboring cells is bijective.

Bipermutive cellular automata have several notable properties, including their use in generating pseudorandom sequences for cryptographic purposes due to their inherent unpredictability [71]. They also have been shown to have particularly useful properties when it comes to generating uniform distributions and achieving maximum entropy configurations. Due to their symmetric and bijective nature, these automata can exhibit high levels of complexity, which has made them valuable in fields such as randomness testing and stream cipher design [149].

Latin squares are combinatorial objects that play a significant role in various areas of mathematics and computer sciences, such as algebra, combinatorics and error correcting codes. A Latin square of order n is an $(n \times n)$ array filled with n symbols such that each symbol appears exactly once in each row and column. This structure has found applications in the study of cellular automata due to their shared symmetry and bijection properties.

The bijective nature of bipermutive cellular automata bears a close resemblance to the construction of Latin squares. Both systems involve the permutation of elements, ensuring that certain patterns or configurations occur without repetition. In the case of bipermutive cellular automata, the left and right permutivity conditions create a natural correspondence to the row and column constraints in a Latin

square [101, 105]. Furthermore, the construction of Latin squares can be seen as a tool for analyzing the behavior of bipermutive CA, particularly in understanding their state transition diagrams and evolution patterns. For instance, the state space of a bipermutive CA can be mapped onto a Latin square, allowing for a clearer visualization of the system's dynamics. There are also deep algebraic connections between bipermutive CA and Latin squares. Specifically, the set of local rules governing a bipermutive CA can be associated with the Cayley tables of certain quasigroups, which are themselves closely related to Latin squares. This algebraic framework provides a method to systematically study the properties of bipermutive cellular automata and to classify their behavior in terms of well-understood combinatorial structures [9].

Another significant area of research in bipermutive cellular automata is their connection to correlation-immune Boolean functions, a concept originating in cryptography. Correlation-immune functions are those that exhibit a level of resistance to statistical attacks by remaining unaffected by the fixing of any subset of their input variables. This concept is crucial in designing secure cryptographic systems.

A Boolean function $f : \{0,1\}^n \to \{0,1\}$ is said to be correlation-immune of order k if the output of the function remains statistically independent of the values of any k input variables. Such functions are essential in the design of secure cryptographic systems, particularly in the context of stream ciphers and block ciphers, where resistance to differential cryptanalysis and linear cryptanalysis is paramount.

The relationship between bipermutive cellular automata and correlation-immune functions arises from the fact that bipermutive rules often exhibit resistance to correlation-based attacks due to their symmetric and bijective nature. Specifically, the local rule of a bipermutive CA can be constructed in such a way that it corresponds to a correlation-immune Boolean function. This connection has been explored extensively in the field of cryptography, where bipermutive cellular automata have been proposed as components of cryptographic algorithms due to their ability to generate sequences that are resistant to statistical analysis [116, 144]. Bipermutive cellular automata, when designed with correlation-immune functions, offer a promising approach to secure cryptographic systems. For example, stream ciphers based on bipermutive CA have been shown to exhibit strong cryptographic properties, including high nonlinearity and resistance to correlation attacks [87]. Furthermore, the connection to Latin squares further enhances the combinatorial complexity of these systems, making them difficult to analyze and break using conventional cryptanalytic techniques. We delve into this connection in Chapter 9, where we will exploit the relation between bipermutive CA and Latin squares to design correlation immune functions. Specifically, starting from the fact proved in [98] that the binary ex-

pansion of any set of mutually orthogonal CA is an orthogonal array of strength at least 2, we are able to extend this construction to the context of Latin hypercubes of order d, providing in Theorem 9.2.5 a framework to construct correlation immune functions of any order d starting from a family of mutually orthogonal CA.

In conclusion, cellular automata, particularly bipermutive cellular automata, represent a fascinating area of study that bridges combinatorics, algebra, and cryptography. Their inherent symmetry and bijection properties make them a powerful tool for generating complex behaviors from simple rules, and the connection between bipermutive CA and Latin squares offers a unique combinatorial framework for understanding the dynamics of these systems. Additionally, their relationship with correlation-immune Boolean functions highlights their potential in cryptographic applications, where resistance to statistical attacks is critical.

Outline

The second part of this dissertation is organized as follows. Chapter 6 provides a detailed background on cellular automata and combinatorial designs.

Then, Chapters 7 and 8 give an overview of the main theoretical results in the construction of CA-based Latin squares, and survey the corresponding applications in cryptography and coding theory.

Finally, Chapter 9 explores how cellular automata can be used to design correlation immune functions of a given weight.

APPENDIX

At the end of this dissertation, an appendix is included where we provide a summary of additional papers that, while not central to the main line of research presented here, contribute valuable insights and context. This appendix serves to offer a broader perspective by briefly discussing these supplementary works, highlighting their relevance and how they complement the primary focus of this thesis both in the topological and nature-inspired models framework.

Appendix A summarizes the paper "A Topology for P-Systems with Active Membranes" [33]. This paper explores deterministic P systems with active membranes within the framework of discrete time dynamical systems. Initially, we demonstrate that, given a fixed set of objects and labels, the collection of all possible P system configurations is countable and that chaotic dynamical behaviors are unattainable. Next, we introduce a notion of distance between membrane configurations that captures the intuitive idea of "dissimilarity" between them. We establish that functions defined by evolution, communication, and division rules are continuous with respect to this dis-

tance and that the resulting topological space is discrete, though not complete. Additionally, we naturally extend classical concepts such as sensitivity to initial conditions and topological transitivity to P systems, and we provide evidence that P systems with these newly defined properties exist. Lastly, we show that computing this distance is efficient, requiring only polynomial time relative to the size of the input configurations.

Appendix B provides an overview of the paper "A Genetic Programming Based Heuristic to Simplify Rugged Landscapes Exploration" [130]. This work focuses on optimization problems that are difficult to solve due to a considerable number of local optima, which may result in premature convergence of the optimization process. To address this issue, we propose a novel heuristic method for constructing a smooth surrogate model of the original function. The surrogate function is easier to optimize but maintains a fundamental property of the original rugged fitness landscape: the location of the global optimum. To create such a surrogate model, we consider a linear genetic programming approach coupled with a self-tuning fitness function. More specifically, to evaluate the fitness of the produced surrogate functions, we employ Fuzzy Self-Tuning Particle Swarm Optimization, a setting-free version of particle swarm optimization. To assess the performance of the proposed method, we considered a set of benchmark functions characterized by high noise and ruggedness. Moreover, the method is evaluated over different problems' dimensionalities. The proposed approach reveals its suitability for performing the proposed task. In particular, experimental results confirm its capability to find the global optimum for all the considered benchmark problems and all the domain dimensions taken into account, thus providing an innovative and promising strategy for dealing with challenging optimization problems.

Part I

EULERIAN MAGNITUDE HOMOLOGY

1

BACKGROUND

In this part of the thesis we recall some general background about graphs, magnitude homology and random models.

Magnitude is a multiscale measure of "size" for a metric space developed by Leinster in [80], and studied in the particular case of graphs in [82]. In this latter paper, Leinster develops cardinality-like properties of the magnitude of a graph, including being multiplicative with respect to the Cartesian product and having (in some more restrictive cases) an inclusion-exclusion formula for the magnitude of a union. It is also shown that the magnitude of a graph is formally both a rational function over $\mathbb{Q}$ and a power series over $\mathbb{Z}$, and that it shares similarities with the Tutte polynomial.

The *magnitude homology* of a graph $MH_{k,\ell}(G)$, first introduced by Hepworth and Willerton in [62], is a categorification of magnitude. We will now recall their definition, some simple results, and give an example of the computation of a magnitude homology group that will provide motivation for our definition of eulerian magnitude homology.

Throughout this work we adopt the notation $[m] = \{1, \dots, m\}$ and $[m]_0 = \{0, \dots, m\}$ for common indexing sets.

1.1 GRAPH TERMINOLOGY AND NOTATION

We will write $G = (V, E)$ to denote an undirected, simple graph[1], with vertices V and edges $E \subseteq \binom{V}{2}$, unordered pairs of distinct vertices. Recall that a *walk* in a graph G is an ordered sequence of vertices $x_0, x_1, \dots, x_k \in V$ such that there is an edge $\{x_i, x_{i+1}\} \in E$ for all $i \in [k]_0$, and a *path* is a walk with no repeated vertices. A *simple circuit* is a walk consisting of at least three vertices in G for which $x_0 = x_k$ and there is no other repetition of vertices. We may view the set of vertices of a graph as an extended metric space by taking the *path distance* $d(u, v)$ to be equal to the length of a shortest path in G from u to v, if such a path exists, and taking $d(u, v) = \infty$ if u and v lie in different components of G.

1 As these are the only flavor of graph we will encounter, we will simply call these "graphs".

If $W \subseteq V$, the *full* subgraph of G on W, written $G|_W$, is the subgraph containing all edges of G supported on W. Given a graph $H = (V', E')$, we write $c(G, H)$ for the number of full subgraphs of G isomorphic to H,

$$c(G, H) = |\{W \subseteq V : \ G|_W \cong H\}|.$$

Finally, write Δ_k for the complete graph on k vertices, or Δ_V for the complete graph on a given set V of vertices. A *clique* in G is a complete full subgraph of G, and a k-clique is a clique supported on k vertices.

Definition 1.1.1. Let $G = (V, E)$ be a graph, and k a non-negative integer. A k-*trail* $\bar{x}$ in G is a $(k+1)$-tuple $(x_0, \dots, x_k) \in V^{k+1}$ of vertices for which $x_i \neq x_{i+1}$ and $d(x_i, x_{i+1}) < \infty$ for every $i \in [k-1]_0$. The *length* of a k-trail $(x_0, \dots, x_k)$ in G is defined as the minimum length of a walk that visits $x_0, x_1, \dots, x_k$ in this order:

$$\operatorname{len}(x_0, \dots, x_k) = d(x_0, x_1) + \cdots + d(x_{k-1}, x_k).$$

We call the vertices $x_0, \dots x_k$ the *landmarks*, x_0 the *starting point*, and x_k the *ending point* of the k-trail. Given a k-trail $\bar{x}$, write $L(\bar{x}) = \{x_0, \dots, x_k\}$ for the corresponding set of landmarks. Write $T_{k,\ell}(G)$ for the collection of all k-trails on G of length ℓ.

1.2 MAGNITUDE HOMOLOGY

The two parameters k and ℓ for trails stratify walks in G that pass through given sequences of vertices, and we can use this stratification to define a bigraded chain complex.

Definition 1.2.1 ([62, Def 2.2])**.** Given a graph G, the (k, ℓ)-*magnitude chain group*, $MC_{k,\ell}(G) = \mathbb{Z}\langle T_{k,\ell}(G)\rangle$, is the free abelian group generated by k-trails in G of length ℓ.

Denote by $(x_0, \dots, \hat{x}_i, \dots, x_k)$ the k-tuple obtained by removing the i-th vertex from the $(k+1)$-tuple $(x_0, \dots, x_k)$. We define the *differential*

$$\partial_{k,\ell} : MC_{k,\ell}(G) \to MC_{k-1,\ell}(G)$$

to be the signed sum $\partial_{k,\ell} = \sum_{i \in [k-1]} (-1)^i \partial^i_{k,\ell}$ of chains corresponding to omitting landmarks without shortening the walk or changing its starting or ending points,

$$\partial^i_{k,\ell}(x_0, \dots, x_k) = \begin{cases} (x_0, \dots, \hat{x}_i, \dots, x_k), & \text{if } \operatorname{len}(x_0, \dots, \hat{x}_i, \dots, x_k) = \ell, \\ 0, & \text{otherwise.} \end{cases}$$

For a non-negative integer ℓ, we obtain the *magnitude chain complex*, $MC_{*,\ell}(G)$, given by the following sequence of free abelian groups and differentials

$$\cdots \to MC_{k+1,\ell}(G) \xrightarrow{\partial_{k+1,\ell}} MC_{k,\ell}(G) \xrightarrow{\partial_{k,\ell}} MC_{k-1,\ell}(G) \to \cdots$$

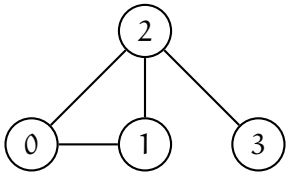


Figure 1.: This graph will be used in Examples 1.2.1 and 2.1.1 to compare computations of magnitude homology and eulerian magnitude homology.

It is shown in [62, Lemma 11] that the composition $\partial_{k,\ell} \circ \partial_{k+1,\ell}$ vanishes, justifying the name "differential", and allowing them to define the corresponding bigraded homology groups of a graph.

Definition 1.2.2 ([62, Def 2.4])**.** The (k,ℓ)-*magnitude homology group* of the graph G is

$$MH_{k,\ell}(G) = H_k(MC_{*,\ell}(G)) = \frac{\ker(\partial_{k,\ell})}{\operatorname{im}(\partial_{k+1,\ell})}.$$

The magnitude homology of a graph is a rich graph invariant. However, understanding what the groups tell us about the structure of the graph is not straightforward. Hepworth and Willerton [62, Proposition 9] show that the first two non-trivial groups simply count elements of G: $MH_{0,0}(G)$ is the free abelian group on V, and $MH_{1,1}(G)$ is the free abelian group on the set of *oriented* edges of G.

It is also straightforward to demonstrate that magnitude homology vanishes when the length of the path is too short to support the necessary landmarks.

Lemma 1.2.1 (c.f. [62, Proposition 10])**.** *Let G be a graph, and $k > \ell$ non-negative integers. Then $MC_{k,\ell}(G) \cong 0$.*

Proof. Suppose $MC_{k,\ell}(G) \neq 0$. Then, there must exist a k-trail $(x_0,\dots,x_k)$ in G so that $\operatorname{len}(x_0,\dots,x_k) = d(x_0,x_1) + \cdots + d(x_{k-1},x_k) = \ell$. However, as consecutive vertices in a k-trail must be distinct, $d(x_i,x_{i+1}) \geqslant 1$ for $i \in [k-1]_0$, so k can be at most ℓ. □

We will make extensive use of the following immediate consequence of this result.

Corollary 1.2.2. *Let G be a graph and k a non-negative integer. Then $MH_{k,k}(G) \cong \ker(\partial_{k,k})$.*

However, even for uncomplicated graphs, those magnitude homology groups which do not vanish can be quite intricate.

Example 1.2.1. Consider the graph G in Figure 1. We will compute $MH_{2,2}(G)$.

$MC_{2,2}(G)$ is generated by the 2-trails in G of length 2. There are eighteen such, consisting of all possible walks of length two in

the graph: (0,1,0), (0,1,2), (0,2,0), (0,2,1), (0,2,3), (1,0,1), (1,0,2), (1,2,0), (1,2,1), (1,2,3), (2,0,1), (2,0,2), (2,1,0), (2,1,2), (2,3,2), (3,2,0), (3,2,1), (3,2,3). Similarly, $MC_{1,2}(G)$ is generated by 1-trails in G of length 2. These are pairs of vertices for which the minimum length of a connecting path is 2, of which there are four: (0,3), (1,3), (3,0), (3,1).

Because the differential $\partial_{2,2}$ consists of omitting only the center vertex, it is straightforward to check that it is surjective, and that the kernel is generated by the 14 elements whose length diminishes when the middle vertex is removed; that is, all elements in $MC_{2,2}(G)$ except (0,2,3), (1,2,3), (3,2,0), (3,2,1).

On the other hand, by Lemma 1.2.1, $MC_{3,2}(G)$ is the trivial group, and thus the image of $\partial_{3,2}$ is $\langle 0 \rangle$. Therefore, $\operatorname{rank}(MH_{2,2}(G)) = 14$, generated by those walks that do not have vertex 3 at exactly one endpoint. Of these, eight consist of walks back-and-forth across a single edge, while the remaining six are length 2 walks between vertices of the triangle with vertices 0, 1, and 2, and record the fact that there is a shorter path between the starting and ending points of the walks.

Cycles that record “back-and-forth" trips across edges or similarly uninformative walks explode in number as the k and ℓ grow. Indeed, as we will see in Theorem 2.3.1, in some regimes along the $k = \ell$ line, such cycles make up the entirety of the magnitude homology. On the other hand, the walks around the triangle detect structure in the graph akin to “convexity" near the walk.

In comparison, in [84] the authors consider the space $\ell_1^n = \mathbb{R} \times_1 \cdots \times_1 \mathbb{R}$ (i.e. $\mathbb{R}^n$ with the taxicab metric) and prove a connection between the magnitude of a convex body $A \subseteq \ell_1^n$ and some intrinsic geometric measures of A such as volume. This suggests that if we simplify the chain complex, it may be easier to interpret the resulting homology theory in terms of informative structure in the graph and relate that structure back to properties of metric spaces.

2

INTRODUCTION TO EULERIAN MAGNITUDE HOMOLOGY

In this chapter we introduce and analyze in depth eulerian magnitude homology. In Section 2.1 we define *eulerian* magnitude homology and *discriminant* magnitude homology and highlight the advantages of the first in terms of interpretation. We then investigate the relationship between magnitude homology and the newly defined eulerian and discriminant magnitude homology.

2.1 DEFINITION AND INTUITION

In the magnitude chain complex, the differential of a single tuple $\partial_{k,\ell}(x_0,\dots,x_k)$ vanishes precisely when $\operatorname{len}(x_{i-1},\hat{x}_i,x_{i+1}) < \operatorname{len}(x_{i-1},x_i,x_{i+1})$ for each i. In other words, every landmark in the tuple enforces a longer walk than would otherwise be necessary based on the structure of the graph. So, the graph contains some smaller substructure *witnessed* by this walk, which suggests there may be a meaningful relationship between the rank of magnitude homology groups of a graph and subgraph counting problems.

However, as we observed in Example 1.2.1, the relationship between the size of the magnitude homology groups and these structures is obscured by the fact that the constituent walks may revisit regions of the graph. For example, if x_0 and x_1 are adjacent in G, $(x_0,x_1,x_0,x_1,x_0) \in MC_{5,4}(G)$. As we will see in Theorem 2.3.1, cycles made of such chains can dominate $MH_{k,k}(G)$ under certain circumstances.

With this motivation, we introduce a natural sub-complex of $MC_{k,l}(G)$.

Definition 2.1.1. Let G be a graph. We say a k-trail $(x_0,\dots,x_k) \in T_{k,\ell}(G)$ is *eulerian* if $x_i \neq x_j$ for all $i \neq j$. Denote the set of eulerian k-trails of G by $ET_{k,\ell}(G) \subseteq T_{k,\ell}(G)$.

We define the (k,ℓ)-*eulerian magnitude chain group*, $EMC_{k,\ell}(G) = \mathbb{Z}\langle ET_{k,\ell}(G)\rangle$, to be the free abelian group generated by eulerian k-trails $(x_0,\dots,x_k)$ of G of length ℓ. Throughout, we will mildly abuse notation by thinking of elements of $ET_{k,\ell}(G)$ as chains in $EMC_{k,\ell}(G)$.

Regarding $EMC_{k,\ell}(G)$ as a subgroup of $MC_{k,\ell}(G)$, it follows directly from the definition that $\partial_{k,\ell}(EMC_{k,\ell}(G)) \subseteq EMC_{k-1,\ell}(G)$, so

$EMC_{*,\ell}(G)$ is the *eulerian magnitude chain complex*, a sub-chain complex of the standard magnitude chain complex for each k. By abuse of notation, we also will write $\partial_{k,\ell}$ for the restriction $\partial_{k,\ell}|_{EMC_{k,\ell}(G)}$ unless it would create confusion.

It is worth pausing to explicitly observe that the term "eulerian" here is used to indicate that no *landmark* is repeated. This does not, in general, imply that a minimal length walk in G that visits those landmarks is eulerian; at best, it guarantees that the minimal walk between any two successive landmarks is distinct from all others.

Remark 2.1.1. However, in the special case $k = \ell$, the number of edges in a minimal walk is one less than the number of landmarks. Thus, as G is simple, there is a unique minimal walk that is entirely specified by its landmarks and so must, indeed, be eulerian (and hamiltonian).

The following simple observation about the differential on the eulerian magnitude chains will drive a great deal of what we do in this thesis.

Lemma 2.1.1. *Let $G = (V, E)$ be a graph, $k \geqslant 2$. Fix some $i \in [k-1]$ and $\bar{x} = (x_0, x_1, \ldots, x_k) \in ET_{k,k}(G) \subseteq EMC_{k,k}(G)$. Then $\partial^i_{k,\ell}(\bar{x}) = 0$ if and only if $\{x_{i-1}, x_{i+1}\} \in E$.*

Remark 2.1.2. Note, this result does not hold for standard magnitude homology, because of cycles like $(0, 1, 0)$ in Example 1, for which Lemma 2.1.1 does not hold because there is no "edge" from vertex o to itself. In Section 2.3, we will investigate this discrepancy in more detail. For now, we simply note that throughout this dissertation, results which build on Lemma 2.1.1 are in general true only in the eulerian case.

Proof. Let $G = (V, E)$ be a graph, $k \geqslant 2$, $i \in [k-1]$, and $\bar{x} = (x_0, x_1, \ldots, x_k) \in EMC_{k,k}(G)$. As $\operatorname{len}((x_0, \ldots, x_k)) = k$ and each vertex is distinct, we have $d(x_j, x_{j+1}) = 1$, so $\{x_j, x_{j+1}\} \in E$ for each $j \in [k-1]_0$. As $\partial^i_{k,k}(\bar{x}) \in ET_{k-1,k}(G)$, if $\partial^i_{k,k}(\bar{x}) = 0$, then $\operatorname{len}(\partial^i_{k,k}(\bar{x})) < \operatorname{len}(\bar{x}) = k$, which can only occur if removing the landmark x_i shortens the walk, which occurs precisely when $\{x_{i-1}, x_{i+1}\} \in E$. □

We now move on to our principal object of study.

Definition 2.1.2. The (k, ℓ)-*eulerian magnitude homology group* of a graph G is

$$EMH_{k,\ell}(G) = H_k(EMC_{*,\ell}(G)) = \frac{\ker(\partial_{k,\ell})}{\operatorname{im}(\partial_{k+1,\ell})}.$$

In this chapter, we will focus our attention on the case $k = \ell$ to facilitate explicit descriptions of the relationship between the structure of the eulerian magnitude homology and that of the graph. However, many of the tools we develop should be applicable to the study of these groups in the case $k < \ell$.

Example 2.1.1. Consider again the graph G of Figure 1. We will compute $EMH_{2,2}(G)$ to compare with the computation in Example 1.2.1.

The eulerian chain group $EMC_{2,2}(G)$ is generated by ten 2-trails: (0,1,2), (0,2,1), (0,2,3), (1,0,2), (1,2,0), (1,2,3), (2,0,1), (2,1,0), (3,2,0), (3,2,1). On the other hand, $EMC_{1,2}(G) = MC_{1,2}(G)$ is generated by the same four elements: (0,3), (1,3), (3,0), (3,1), and $EMC_{3,2}(G) \subseteq MC_{3,2}(G) = \langle 0 \rangle$. So, we have that $EMH_{2,2}(G) = \ker(\partial_{2,2})$, and the group is generated by those six elements in $MC_{2,2}(G)$ which do not visit vertex 3, and thus are precisely those six possible walks along the vertices of the triangle with vertices 0, 1, and 2.

Remark 2.1.3. Many of the definitions and properties regarding magnitude homology in [62] and [85] carry over directly to eulerian magnitude homology, as it is defined on a subcomplex of the original chain complex. In particular, one can check that $EMH_{0,0}(G) \cong MH_{0,0}(G)$ and $EMH_{1,1}(G) \cong MH_{1,1}(G)$, since the generators of the groups necessarily satisfy the condition of not revisiting vertices. Thus, these eulerian magnitude homology groups also count the number of vertices and edges in G, respectively.

Naively, Lemma 2.1.1 and Example 2.1.1, along with our experience with $EMH_{0,0}$ and $EMH_{1,1}$, may suggest that the rank of $EMH_{2,2}(G)$ should provide a count of triangles – three vertex cliques – in G. And, indeed, generators of $EMC_{2,2}(G)$ are 2-trails (x_0, x_1, x_2) in G of length 2. By Lemma 2.1.1, we see that if $\partial_{2,2}(x_0, x_1, x_2) = \partial^1_{2,2}(x_0, x_1, x_2) = 0$, then $\{x_0, x_2\} \in E$ and $G|_{\{x_0,x_1,x_2\}} \cong C_3$. Further, any 2-trail given by a permutation of these three vertices gives rise to a generator of $EMC_{2,2}(G)$ with zero differential. Conversely, if for some subset $\{x_0, x_1, x_2\} \subseteq V$ we have $G|_{\{x_0,x_1,x_2\}} \cong C_3$, we will have all six of these generators in $\ker(\partial_{2,2}) \subseteq EMC_{2,2}(G)$.

However, this is not the complete story. For, if $\partial_{2,2}(x_0, x_1, x_2) = (x_0, x_2)$, then for any trail $(x_0, x_3, x_2) \in ET_{2,2}(G)$ with $x_1 \neq x_3$, we also have $\partial_{2,2}(x_0, x_3, x_2) = (x_0, x_2)$. Thus, $(x_0.x_1, x_2) - (x_0, x_3, x_2) \in \ker(\partial_{2,2})$. However, the same will hold for any choice of another trail, $(x_0, x_3', x_2) \in ET_{2,2}(G)$.

Definition 2.1.3. Let $G = (V, E)$ be a graph and $x_i, x_j \in V$ so that $\{x_i, x_j\} \notin E$. Let $V_{x_i,x_j} = \{v \in V \mid (x_i, v, x_j) \in ET_{2,2}(G)\} \subset V$. We call the full subgraph of G induced by the vertex set $\{x_i, x_j\} \cup V_{x_i,x_j} \subset V$ the *fan subgraph of* G *at the pair* $\{x_i, x_j\}$, written $\mathrm{Fan}_{\{x_i,x_j\}}$. See Figure 2 for an example.

Now, for every pair of vertices $\{x_i, x_j\} \subseteq V$ so that $\{x_i, x_j\} \notin E$, there are $(|V_{x_i,x_j}| - 1)$ linearly independent homology generators in $EMC_{2,2}(G)$ obtained by fixing $v \in V_{x_i,x_j}$ and allowing $v' \in V_{x_i,x_j} \setminus v$ to vary, producing cycles of the form $(x_i.v, x_j) - (x_i, v', x_j) \in \ker(\partial_{2,2})$, $v, v' \in V_{x_i,x_j}$; other cycles of this form are linear combinations of those in this family. By symmetry, there are another $(|V_{x_i,x_j}| - 1)$ generators

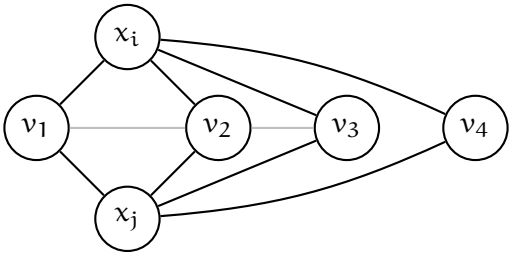


Figure 2.: Example of a fan subgraph induced by the vertex set $\{x_i, x_j, v_1, v_2, v_3, v_4\}$. Notice that any edges between the vertices v_k may or may not be present; here (v_1, v_2) and (v_2, v_3) are drawn in grey to indicate they are incidental.

for trails starting at x_j and ending at x_i. Letting the set $\{x_i, x_j\}$ vary, we obtain all other cycles in $EMH_{2,2}(G)$:

$$\operatorname{rank}(EMH_{2,2}(G)) = 6 \cdot c(G, C_3) + \sum_{\{x_i,x_j\} \in \binom{V}{2} \setminus E} 2 \cdot (|V_{x_i,x_j}| - 1).$$

For each fan subgraph Fan_{x_i,x_j} of G, the full subgraph on vertices $\{x_i, x_j, v, v'\}$ is isomorphic either to C_4 or to

$$F_4 = (\{v_0, v_1, v_2, v_3\}, \{\{v_0, v_1\}, \{v_0, v_3\}, \{v_0, v_2\}, \{v_1, v_2\}, \{v_2, v_3\}\}),$$

the cycle graph on four vertices with one diagonal added; here, the added edge is $\{v, v'\}$. In total, there are $\binom{|V_{x_i,x_j}|}{2}$ such subgraphs. The difficulty in obtaining a complete count of these subgraphs comes from the fact that if the full subgraph supported on $\{x_i, x_j, v, v'\}$ is isomorphic to C_4, then Fan_{x_i,x_j} and $\mathrm{Fan}_{v,v'}$ intersect, and there is no general way to determine the intersection pattern of such subgraphs from a naive count of homology generators.

However, we can apply the above discussion, accounting for symmetry and the fact that $(n-1) \leqslant \binom{n}{2}$ for $n \geqslant 2$, to provide a coarse bound on a subgraph count using $EMH_{2,2}(G)$.

Lemma 2.1.2. *Let $G = (V, E)$ be a graph. Write C_3 and C_4 for the cycle graphs on three and four vertices respectively, Then*

$$\operatorname{rank}(EMH_{2,2}(G)) \leqslant 6 \cdot c(G, C_3) + 4 \cdot c(G, C_4) + 2 \cdot c(G, F_4).$$

Remark 2.1.4. It is worth noting that Lemma 2.1.2 is analogous to some results proved in [52], where the authors show a connection between the path homology groups of (di)graphs and substructure like cliques, binary hypercubes and other subgraphs reminiscent of polyhedra. However, the authors have not investigated any potential connections between these results.

Generalizing Lemma 2.1.2 to higher k becomes difficult because of the increasingly complicated collection of isomorphism types of

graphs which can support eulerian trails. A first result in this direction can be directly extracted from the proof of Lemma 2.1.2. We observed that generators of $\mathrm{EMC}_{2,2}(G)$ with zero differential were precisely those eulerian trails that walk around boundaries of 3-cliques. While this is not a complete characterization of such generators in $\mathrm{EMC}_{k,k}(G)$ in general, it is the case that walks around cliques always have zero differential.

Lemma 2.1.3. *Let* G *be a graph, and let* $Z \subseteq \mathrm{ET}_{k,k}(G) \subseteq \mathrm{EMC}_{k,k}(G)$ *be the collection of generators* $\bar{x}$ *for which* $\partial_{k,k}(\bar{x}) = 0$*. Write* Δ_{k+1} *for the complete graph on* $k+1$ *vertices, then*

$$c(G, \Delta_{k+1}) \leqslant \left\lfloor \frac{|Z|}{k+1!} \right\rfloor.$$

Proof. Let G and Z be as in the statement of the lemma. Observe that if vertices $x_0, \ldots, x_k$ form a $(k+1)$-clique of G, any k-trail passing through all nodes $x_0, \ldots, x_k$ in any order has length k. Thus, for every permutation $\sigma \in \Sigma_k$, $(x_{\sigma(0)}, \ldots, x_{\sigma(k)}) \in \mathrm{ET}_{k,k}(G)$. Further, since all edges among the constituent vertices are present, removing any landmark from such a trail reduces its length. so $\partial_{k,k}(x_{\sigma(0)}, \ldots, x_{\sigma(k)}) = 0$. Thus, each such trail is an element of Z, and the number of $(k+1)$-cliques in G is bounded above by $\left\lfloor \frac{|Z|}{k+1!} \right\rfloor$. □

Remark 2.1.5. While these results provide us with some intuition about the meaning of the simplest classes in the eulerian magnitude homology groups, they also serve as a cautionary tale from a computational perspective. Any naive attempt to compute magnitude homology groups will run into clique enumeration problems, which grow exponentially in complexity with k for even moderately sized n.

One lesson that we should take away from this investigation is that the relationship between the combinatorics of the eulerian magnitude homology computation and that of subgraph counting are complicated by the fact that the presence or absence of some edges is irrelevant to computations in the differential. In the next section, we will develop language aimed at mitigating this difficulty.

2.2 FAMILIES OF GRAPHS THAT SUPPORT EULERIAN MAGNITUDE CYCLES

We will now develop the language which we will need to fully describe the structure underlying nontrivial cycles in $\mathrm{EMH}_{k,k}(G)$. To do so, we will study families of graphs that can support cycles of various forms. To this end, we introduce a useful way to describe families of graphs adopted from [152]. We have chosen to slightly alter the terminology used in that paper to more clearly distinguish between individual graphs and collections thereof.

Definition 2.2.1 (c.f. [152])**.** A *class graph* $\mathcal{G} = (V, E_S, E_B)$ consists of a vertex set V and two disjoint edge sets $E_S, E_B \subset \binom{V}{2}$. A *complete* class graph is one for which $E_S \cup E_B = \binom{V}{2}$.

The sets E_S and E_B provide us with rules for constructing a family of graphs: edges in E_S are mandatory, while edges in E_B are optional, and edges that appear in neither E_S nor E_B cannot be included.

Definition 2.2.2. Given a class graph $\mathcal{G} = (V, E_S, E_B)$, the set of *graphs of class* $\mathcal{G}$, $\Gamma(\mathcal{G})$, consists of all graphs $G = (V, E)$ such that $E_S \subseteq E \subseteq E_S \cup E_B$. Let $\mathcal{G} = (V, E_S, E_B)$ be a class graph. We denote the minimal and maximal graphs (under inclusion) in $\Gamma(\mathcal{G})$ as $\alpha(\mathcal{G}) = (V, E_S)$ and $\omega(\mathcal{G}) = (V, E_S \cup E_B)$.

We visualize a class graph $\mathcal{G}$ by drawing edges in E_S as solid lines and those in E_B as dashed lines; graphs in $\Gamma(\mathcal{G})$ must have all solid edges and may have any dashed edges, but can have no other edges. See Figure 3 for relevant examples.

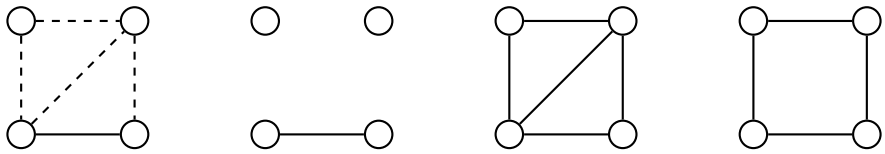


Figure 3.: (far left) A class graph $\mathcal{G}$. (middle left) the minimal graph $\alpha(\mathcal{G}) \in \Gamma(\mathcal{G})$. (middle right) the maximal graph $\omega(\mathcal{G}) \in \Gamma(\mathcal{G})$. (far right) a graph $G \in \Gamma(\mathcal{G})$.

With this language, we can characterize which graphs support cycles in $EMC_{k,k}(G)$. We begin with the simple case of a single generator with zero differential. Recall from Definition 1.1.1 that for a k-trail $\bar{x}$, $L(\bar{x})$ is the corresponding set of landmarks. In the current setting, where $k = \ell$, we have that $L(\bar{x})$ is precisely the set of vertices in the unique corresponding minimal walk in G, so in this case we will often call $L(\bar{x})$ the *support* of $\bar{x}$.

Lemma 2.2.1. *Let* $G = (V, E)$ *be a graph. Suppose* $\bar{x} = (x_0, \dots x_k) \in ET_{k,k}(G) \subseteq EMC_{k,k}(G)$ *has* $\partial_{k,k}\bar{x} = 0$. *Consider the complete class graph*

$$\mathcal{H}(\{\bar{x}\}) = \Bigg(L(\bar{x}),$$
$$E_S = \{\{x_i, x_{i+1}\}\}_{i\in[k-1]_0} \cup \{\{x_{i-1}, x_{i+1}\}\}_{i\in[k-1]},$$
$$E_B = \binom{L(\bar{x})}{2} \setminus E_S \Bigg).$$

Then $G|_{L(\bar{x})} \in \Gamma(\mathcal{H}(\{\bar{x}\}))$.

Proof. Let $G = (V, E)$ be a graph, and $\bar{x} = (x_0, \dots x_k) \in EMC_{k,k}(G)$ a k-trail. So, every edge $\{x_i, x_i + 1\}, i \in [k-1]_0$ is in E. By Lemma 2.1.1, for every $i \in [k-1]$ such that $\partial^i_{k,k}(\bar{x}) = 0$, the edge $\{x_{i-1}, x_{i+1}\}$ is in E. Thus, if $\partial_{k,k}(\bar{x}) = 0$, we have $\alpha(\mathcal{H}(\{\bar{x}\})) \subseteq G|_{L(\bar{x})}$, so $G|_{L(\bar{x})} \in \Gamma(\mathcal{H}(\{\bar{x}\}))$. □

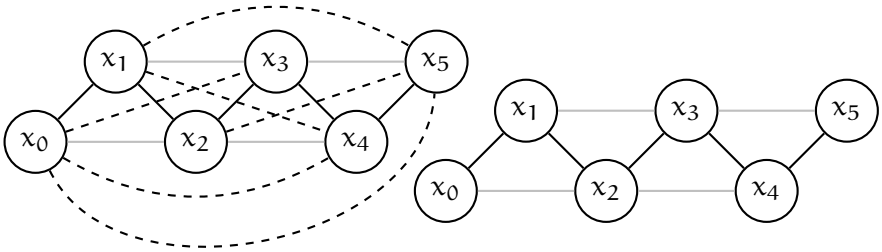


Figure 4.: (left) The class graph $\mathcal{H} = \mathcal{H}(\{(x_0, \dots, x_5)\})$. (right) The minimal element $\alpha(\mathcal{H}) \in \Gamma(\mathcal{H})$. We have $\partial_{5,5}(x_0, \dots, x_5) = 0$ in $\mathrm{EMC}_{k,k}(G)$, precisely when $G|_{\{x_0,\dots,x_5\}} \in \Gamma(\mathcal{H})$. Black edges lie in the support of the trail, while gray edges must be present to force terms in the differential to be zero. Dashed edges do not play a role in the computation of the differential.

We now consider the somewhat more intricate combinatorics that arise for linear combinations of generators whose differentials cancel. The first non-trivial case, described in the following example, is instructive.

Example 2.2.1. Let $G = (V, E)$ be a graph, and let $\bar{x}^i = (x_0^i, \dots, x_k^i), i = 1, 2$, be trails in Δ_V. We now extend the construction of $\mathcal{H}(\{\bar{x}\})$ from Lemma 2.2.1 to construct a class graph $\mathcal{H}(\{\bar{x}^1, \bar{x}^2\}) = (W^{\{1,2\}}, E_S^{\{1,2\}}, E_B^{\{1,2\}})$ so that if $G|_{W^{\{1,2\}}} \in \mathcal{H}(\{\bar{x}^1, \bar{x}^2\})$, then $\bar{x}^1$ and $\bar{x}^2$ are generators of $\mathrm{EMC}_{k,k}(G)$ for which $\partial_{k,k}(\bar{x}^i) \neq 0, i = 1, 2$ but $\partial_{k,k}(\bar{x}^1 - \bar{x}^2) = 0$.

From the definition of the differential, if $\bar{x}^1 \neq \bar{x}^2$ and $\partial_{k,k}(\bar{x}^1 - \bar{x}^2) = 0$ then both trails agree in all landmarks except one, say $x_r^1 \neq x_r^2$ for some $r \in [k-1]$, and the vertex x_r^2 cannot appear as a landmark in $\bar{x}^1$ nor vice versa. Indeed, suppose they differ in two landmarks, say $x_r^1 \neq x_r^2$ and $x_s^1 \neq x_s^2$ for some $r, s \in [k-1]$, and say $r < s$. Then if we compute the boundary $\partial_{k,k}(\bar{x}^1 - \bar{x}^2)$ we get

$$\begin{aligned}
\partial_{k,k}(\bar{x}^1 - \bar{x}^2) &= \sum_{i=1}^{k-1} (-1)^i \partial_{k,k}^i(\bar{x}^1 - \bar{x}^2) \\
&= (-1)^r \partial_{k,k}^r(\bar{x}^1 - \bar{x}^2) + (-1)^s \partial_{k,k}^s(\bar{x}^1 - \bar{x}^2) \\
&= (-1)^r((x_0^1, \dots, \hat{x}_r^1, \dots, x_s^1, \dots, x_k^1) - (x_0^2, \dots, \hat{x}_r^2, \dots, x_s^2, \dots, x_k^2)) \\
&+ (-1)^r((x_0^1, \dots, x_r^1, \dots, \hat{x}_s^1, \dots, x_k^1) - (x_0^2, \dots, x_r^2, \dots, \hat{x}_s^2, \dots, x_k^2)) \\
&\neq 0,
\end{aligned}$$

because we assumed $x_r^1 \neq x_r^2$ and $x_s^1 \neq x_s^2$, and thus the subtuples do not simplify. Note that, in particular, if this differential vanishes then $x_0^1 = x_0^2$ and $x_k^1 = x_k^2$, so the trails have the same starting and ending points.

Let $\mathcal{H}(\{\bar{x}^1\}) = (W^{\{1\}}, E_S^{\{1\}}, E_B^{\{1\}})$ be the class graph of Lemma 2.2.1. The set E_S necessarily contains all of the edges in the support of

both trails except $\{x^1_{r-1}, x^2_r\}$ and $\{x^2_r, x^1_{r+1}.\}$ Further, per the proof of Lemma 2.2.1, the edges already present in $\mathcal{H}(\{\bar{x}^1\})$ enforce $\partial^i_{k,k}(\bar{x}^1) = 0, i \neq r$, and $\partial^i_{k,k}(\bar{x}^2) = 0$ for $i \neq r-1, r, r+1$, due to agreement of the trails away from these vertices. However, we must introduce new edges to ensure $\partial^{r-1}_{k,k}(\bar{x}^2) = \partial^{r+1}_{k,k}(\bar{x}^2) = 0$. So, we define $\mathcal{H}(\{\bar{x}^1, \bar{x}^2\}) = (W^{\{1,2\}}, E_S^{\{1,2\}}, E_B^{\{1,2\}})$, where

$$
\begin{aligned}
W^{\{1,2\}} &= W^{\{1\}} \cup \{x^2_r\} \\
E_S^{\{1,2\}} &= \left(E_S^{\{1\}} \cup \{\{x^1_a, x^2_r\} : \ a \in \{r-2, r-1, r+1, r+2\} \cap [k]_0\}\right) \setminus \{\{x^1_{r-1}, x^1_{r+1}\}\} \\
E_B^{\{1,2\}} &= \binom{W^{\{1,2\}}}{2} \setminus \left(E_S^{\{1,2\}} \cup \{\{x^1_{r-1}, x^1_{r+1}\}\}\right)
\end{aligned}
$$

See Figure 5 for an illustration of the portion of this new class graph that differs from $\mathcal{H}(\{\bar{x}^1\})$. The new vertex x'_r and the two new edges $\{x^1_{r-1}, x^2_r\}$ and $\{x^1_{r+1}, x^2_r\}$ support the trail $\bar{x}^2$ and enforce the agreement of $\partial^r_{k,k}$ on the two generators. The other one or two new edges are diagonals in newly introduced subgraphs isomorphic to C_4, and so are added to enforce that $\partial^{r-1}_{k,k}(\bar{x}^i) = 0$ and $\partial^{r+1}_{k,k}(\bar{x}^i) = 0$, as needed. Finally, by Lemma 2.1.1, the edge $\{x_{r-1}, x_{r+1}\}$ must be absent from G to ensure that $\partial^r_{k,k}(\bar{x}^1) = \partial^r_{k,k}(\bar{x}^2) \neq 0$. Per Lemma 2.2.1, all other terms in the differential of both chains are zero due to edges in $E_S^{\{1\}}$.

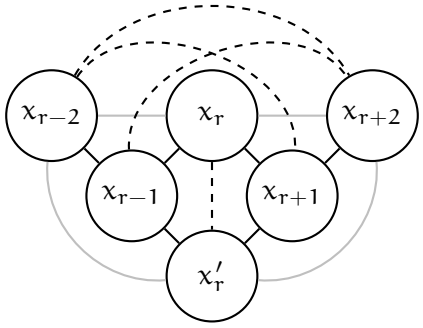


Figure 5.: Neighborhood $U \subseteq W^{\{1,2\}}$ of the vertices x_r and x'_r in the class graph $\mathcal{H}(\{\bar{x}^1, \bar{x}^2\})$. If $G|_W \in \Gamma(\mathcal{H}(\{\bar{x}^1, \bar{x}^2\}))$, then $\partial_{k,k}\left(\bar{x}^1 - \bar{x}^2\right) = 0$. Outside of this neighborhood, the class graph is identical to $\mathcal{H}(\bar{x})$ from Figure 4. Black edges lie in the support of the trail, while gray edges must be present to force terms in the differential to be zero.

Our goal now is to generalize the construction in Example 2.2.1 to characterize classes of subgraphs of a graph G which can support cycles in $EMC_{k,k}(G)$. To do so, we will first develop a decomposition of these cycles via a manageable spanning set.

Definition 2.2.3. Let G be a graph and fix integers $k, m, n \geqslant 1$. Consider a collection of eulerian k-trails $X = \{\bar{x}^i\}_{i\in[m]} \subseteq ET_{k,k}(G)$. We say such a collection X is *local* if all of the $\bar{x}^i$ have the same starting and ending points, and for each $i \in [m]$ there are $j \in [m] \setminus \{i\}$ and $r \in [k-1]$ so that $x^i_r \neq x^j_r$ and $x^i_a = x^j_a$ for all $a \in [k-1] \setminus \{r\}$.

Trails in local collections are those which can potentially share terms in their differentials, so we can hope to construct linear combinations of these trails which are cycles. However, such linear combinations may decompose into cycles supported on smaller local collections of trails. We will need to decompose cycles as much as possible to obtain our desired spanning set.

Definition 2.2.4. Let G be a graph and suppose $X = \{\bar{x}^i\}_{i\in[m]} \subseteq ET_{k,k}(G)$. If there are coefficients $a_i \in \mathbb{Z} \setminus \{0\}, i \in [m]$ so that $\partial_{k,k}(\sum_{i\in I} a_i \bar{x}^i) = 0$, we say that $\gamma = \sum_{i\in I} a_i \bar{x}^i$ is a cycle *supported on* X. If, in addition, there is no $J \subsetneq [m]$ so that $\partial_{k,k}(\sum_{j\in J} a_j \bar{x}^j) = 0$, we say that γ is an X-*minimal cycle for* G, and that γ is *minimally supported on* X.

Now, we will take the (perhaps discomfiting) step of introducing an auxiliary graph that encodes structure in cycles in $EMC_{k,k}(G)$ supported on local collections of trails, and studying its structure.

Definition 2.2.5. Let G be a graph, fix integers $k, m, n \geqslant 1$, and let $X = \{\bar{x}^i\}_{i\in[m]}$ be a local collection of eulerian k-trails in G. Define the *structure graph* for X, $s(X)$, to be the graph with vertices X and an edge $\{\bar{x}^i, \bar{x}^j\}$ if there exists $r \in [k-1]$ so that $x^i_r \neq x^j_r$ and $x^i_a = x^j_a$ for all $a \in [k-1] \setminus \{r\}$.

Suppose $X = \{\bar{x}^i\}_{i\in[m]}$ is a local collection of trails in $ET_{k,k}(G)$ for some graph G. The cliques and circuits in the structure graph $s(X)$ encode relations on coefficients of linear combinations of the trails in X for which the differential $\partial_{k,k}$ can vanish. Observe first that since every edge in the structure graph indicates that the corresponding trails differ by precisely one landmark, a maximal clique in $s(X)$ corresponds to a maximal family $\{\bar{x}^j\}_{j\in J} \subseteq X, J \subseteq [m]$ which are pairwise identical except in their rth landmark for some fixed r. Thus, since we assumed vanishing differential, there is a single nonzero term in $\partial^r_{k,k}(\sum_{i\in[m]} a_i \bar{x}^i)$, corresponding to this family, and that term vanishes precisely when $\sum_{j\in J} a_j = 0$. In the particular case when the maximal clique has two vertices – that is, it is an edge in $s(X)$ which is not included in a larger clique – this equation becomes $a_1 + a_2 = 0$, so the coefficients assigned to the two trails must differ by a sign.

Knowing that the cliques in the structure graph determine the relations between coefficients in cycles supported on a local collection X allows us to describe the class of graphs that support X-minimal cycles. To do so, for each maximal clique in $s(X)$, we assemble required and excluded edges as we did in Example 5.1.1 including a set E_{supp} of the unions of the supports of the trails, a set E_{diff} of the unions of the sets of edges which make each trail's differential cancel; and, a set E_{rem} which indicates which edges must be missing from the graph so that trails have non-zero differentials indicated by the structure of $s(X)$. If a collection X would cause a conflict between required and excluded edges, we cannot construct a graph which would minimally support the corresponding cycle, so we must exclude such collections. We collect all of these observations in Definition 2.2.6; Example 2.2.2 below describes the constituent edge sets for a particular graph and local collection X of trails.

Definition 2.2.6. Fix integers $k, m \geqslant 1$. Consider a local collection $X = \{\bar{x}^i\}_{i\in[m]} \subseteq ET_{k,k}(\Delta_n)$. Write $\bar{x}^i = (x^i_0, \ldots, x^i_k)$. Let $\{Q_t = \{\bar{x}^j\}_{j\in J_t} \subseteq X\}_{t\in[r]}$ be the collection of supports of maximal cliques in $s(X)$, so $s(X)|_{Q_t}$ is a maximal clique of $s(X)$ for all $t \in [r]$. For each t, write $b(t)$ for the landmark at which the constituent trails of Q_t differ; that is, for all $i \neq j \in J_t, x^i_{b(t)} \neq x^j_{b(t)}$ and $x^i_s = x^j_s$ for all $s \in [k]_0 \setminus \{b(t)\}$.

Now, write

$$\begin{aligned}
E_{supp} &= \{\{x^i_a, x^i_{a+1}\} : \ a \in [k-1]_0, i \in [m]\} \\
E_{diff} &= \{\{x^i_a, x^i_{a+2}\} : \ a \in [k-2]_0, i \in [m]\} \\
E_{rem} &= \{\{x^{j(t)}_{b(t)-1}, x^{j(t)}_{b(t)+1}\} : \text{ for any choice of } j(t) \in J_t, \text{ for all } t \in [r]\}
\end{aligned}$$

If $E_{supp} \cap E_{rem} = \varnothing$, we say X is *compatible*, and define the *minimal class graph supporting* X, written $\mathcal{H}(X) = (W^X, E^X_S, E^X_B)$, with the following elements:

$$\begin{aligned}
W^X &= \bigcup_{i\in[m]} \{x^i_0, \ldots, x^i_k\} \\
E^X_S &= E_{supp} \cup E_{diff} \setminus (E_{diff} \cap E_{rem}) \\
E^X_B &= \binom{W}{2} \setminus E_{rem}
\end{aligned}$$

Example 2.2.2. For the graph in Figure 6, consider the local collection $X = \{\bar{x}^i\}^4_{i=1}$, with $\bar{x}^1 = (0,1,3,5,6)$, $\bar{x}^2 = (0,1,4,5,6)$, $\bar{x}^3 = (0,2,4,5,6)$ and $\bar{x}^4 = (0,2,3,5,6)$. Applying Definition 2.2.6 to this collection, we have $k = m = 4$, while the structure graph $s(X)$ is isomorphic

to C_4, with $r = 4$ maximal cliques. From this structure, we collect:

$$\begin{aligned}
E_{\text{supp}} &= \bigcup_{i\in[4]} \{\{x^i_a, x^i_{a+1}\} : a \in [3]_0\} \\
&= \{\{0,1\},\{1,3\},\{3,5\},\{5,6\}\} \cup \{\{0,1\},\{1,4\},\{4,5\},\{5,6\}\} \\
&\quad \cup \{\{0,2\},\{2,4\},\{4,5\},\{5,6\}\} \cup \{\{0,2\},\{2,3\},\{3,5\},\{5,6\}\} \\
&= \{\{0,1\},\{0,2\},\{1,3\},\{1,4\},\{2,3\},\{2,4\},\{3,5\},\{4,5\},\{5,6\}\} \\
E_{\text{diff}} &= \bigcup_{i\in[4]} \{\{x^i_a, x^i_{a+2}\} : \ a \in [2]_0\} \\
&= \{\{0,3\},\{1,5\},\{3,6\}\} \cup \{\{0,4\},\{1,5\},\{4,6\}\} \\
&\quad \cup \{\{0,4\},\{2,5\},\{4,6\}\} \cup \{\{0,3\},\{2,5\},\{3,6\}\} \\
&= \{\{0,3\},\{0,4\},\{1,5\},\{2,5\},\{3,6\},\{4,6\}\} \\
E_{\text{rem}} &= \bigcup_{i\in[4]} \{\{x^{j(t)}_{b(t)-1}, x^{j(t)}_{b(t)+1}\} : \text{ for any choice of } j(t) \in J_t, \text{ for all } t \in [r]\} \\
&= \{\{0,3\},\{1,5\}\} \cup \{\{0,4\},\{1,5\}\} \cup \{\{0,4\},\{2,5\}\} \cup \{\{0,3\},\{2,5\}\} \\
&= \{\{0,3\},\{0,4\},\{1,5\},\{2,5\}\}
\end{aligned}$$

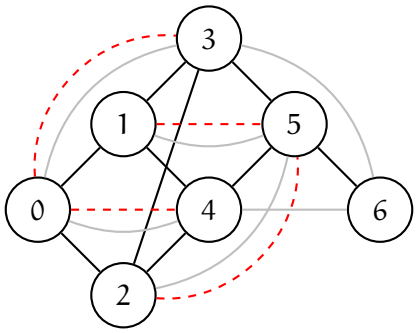


Figure 6.: Let $\bar{x}^1 = (0,1,3,5,6)$, $\bar{x}^2 = (0,1,4,5,6)$, $\bar{x}^3 = (0,2,4,5,6)$ and $\bar{x}^4 = (0,2,3,5,6)\}$. E_{supp} is black, E_{diff} is gray and E_{rem} is dashed-red.

We see from the computations above that $E_{\text{supp}} \cap E_{\text{rem}} = \emptyset$. Thus X is a compatible collection and we can define the minimal class graph supporting X, $\mathcal{H}(X)$:

$$\begin{aligned}
W^X &= \bigcup_{i\in[4]} \{x^i_0, x^i_1, x^i_2, x^i_3, x^i_3\} \\
&= \{0,1,2,3,4,5,6\} \\
E^X_S &= E_{\text{supp}} \cup E_{\text{diff}} \setminus (E_{\text{diff}} \cap E_{\text{rem}}) \\
&= \{\{0,1\},\{0,2\},\{1,3\},\{1,4\},\{2,3\},\{2,4\},\{3,5\},\{3,6\},\{4,5\},\{4,6\},\{5,6\}\}
\end{aligned}$$

$$\begin{aligned} E_B^X &= \binom{W}{2} \setminus E_{rem} \\ &= \binom{W}{2} \setminus \{\{0,3\},\{0,4\},\{1,5\},\{2,5\}\}. \end{aligned}$$

Observe that the class graph $\mathcal{H}(X)$ is "minimal" in the sense that for a cycle minimally supported on X, every edge in $\alpha(\mathcal{H}(X))$ is either in E_{supp}, and so needed to support one of the trails, or in E_{diff} but not E_{rem}, and so ensures that a boundary term that is not canceled by other trails is zero. In order to demonstrate that the family X describes at least one element in $EMC_{k,k}(\alpha(\mathcal{H}(X)))$, we will first observe that the structure graph provides us with tools for decomposing such cycles into simpler elements.

Definition 2.2.7. Let G be a graph. A simple circuit *of minimal-length* is a full subgraph of G isomorphic to some cycle graph $C_k, k \geqslant 3$. A connected structure graph G is a *clique-tree* if every minimal-length simple circuit in G has three vertices[1]. A *clique-forest* is a disjoint union of clique-trees.

Observe that two maximal cliques in a structure graph cannot intersect in more than one vertex. Indeed, we noticed that vertices $\{\bar{x}^j\}_{j\in J}$ of a maximal clique are all pairwise identical except in their rth landmark, for some fixed r. Now, suppose by contradiction that two distinct maximal cliques, Q_1 and Q_2, share two vertices $\bar{x}^a$ and $\bar{x}^b$. Notice that, because the two maximal cliques are distinct, the landmark in which vertices of Q_1 differ is different then the one in which vertices of Q_2 differ. So, say $\{\bar{x}^j\}_{j\in J} \in Q_1$ differ in their rth landmark and $\{\bar{x}^j\}_{j\in J} \in Q_2$ differ in their sth landmark, $r \neq s$. Now, because $\bar{x}^a, \bar{x}^b \in Q_1$ they both differ only in their rth landmark, and similarly because $\bar{x}^a, \bar{x}^b \in Q_2$ they both differ only in their sth landmark. But then they differ in two landmarks, so by definition of structure graph they are not connected by edge, and thus they cannot be contained in the same maximal clique, which gives us a contradiction.

Therefore, clique-trees are connected graphs for which every simple circuit is contained in a maximal clique, and so successively collapsing the maximal cliques of three or more vertices into vertices will result in a tree.

It is straightforward to construct eulerian magnitude cycles minimally supported on clique-trees. Suppose that for some local, compatible collection $X = \{\bar{x}^i\}_{i\in[m]} \subseteq ET_{k,k}(\Delta_n)$, $G \in \mathcal{H}(X)$ and $s(X)$ is a clique-tree. Select any vertex $\bar{x}^i \in X$ and assign to it a non-zero coefficient a_i. For each maximal clique $\{\bar{x}^j\}_{j\in J}$, $J \subseteq [m]$ in $s(X)$ such that $i \in J$, assign non-zero coefficients $a_j, j \in J \setminus \{i\}$ so that $\sum_{j\in J} a_j = 0$.

1 Clique trees are chordal graphs. However, as structure graphs they have a restricted, tree-like form we describe below, which motivates our choice of terminology.

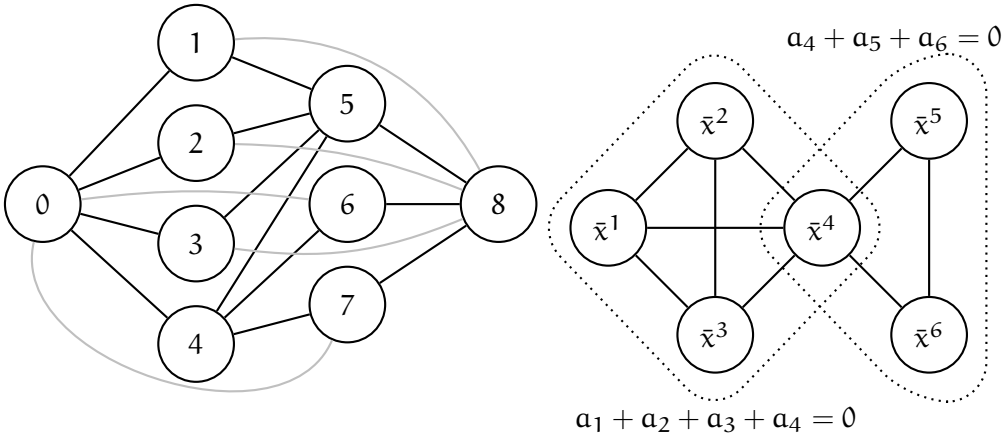


Figure 7.: Write $\bar{x}^1 = (0,1,5,8)$, $\bar{x}^2 = (0,2,5,8)$, $\bar{x}^3 = (0,3,5,8)$, $\bar{x}^4 = (0,4,5,8)$, $\bar{x}^5 = (0,4,6,8)$, $\bar{x}^6 = (0,4,7,8)$. and $X = \{\bar{x}^i\}_{i\in[6]}$. (left) In this graph $G \in \mathcal{H}(X)$, the cycle $\gamma = \bar{x}^1 - \bar{x}^2 + \bar{x}^3 - \bar{x}^4 - \bar{x}^5 + 2\bar{x}^6$ is minimally supported on X. The set E_{supp} is represented in black, while the edges in $E_{diff} \setminus (E_{diff} \cap E_{rem})$ are gray. (right) The structure graph $s(X)$ is a clique-tree, and its maximal cliques describe relations on the coefficients of supported cycles.

Repeat inductively for every maximal clique containing a vertex assigned a coefficient by this process. Since $s(X)$ is connected this process will assign a coefficient to every element of X. By construction, the resulting linear combination will be a cycle in $EMC_{k,k}(G)$. In addition, the cycle is X-minimal because we have constructed G to force other terms in the boundary to vanish, so terms in the differential corresponding to cliques in $s(X)$ cancel only due to the relations on the coefficients. See Figure 7 for an example.

We now come to one of the core results of this chapter: given a general local, compatible collection X of k-trails in G, we can decompose cycles minimally supported on X into cycles minimally supported on clique-trees and cycles minimally supported on minimal circuits of $s(X)$. This decomposition bears more than a passing resemblance to finding a minimal spanning tree for $s(X)$, though here we will remove vertices in cycles rather than edges.

Theorem 2.2.2. *Let G be a graph and $X = \{\bar{x}^i\}_{i\in[m]} \subseteq ET_{k,k}(G)$ a local, compatible collection of trails in G. Suppose $\gamma \in EMC_{k,k}(G)$ is a cycle minimally supported on X. Then there is a (possibly empty) family of local, compatible collections $Y_1, \dots Y_r \subseteq X$ and disjoint subsets $Z_a \subseteq Y_a, a \in [r]$ so that*

1. *$s(X)|_{Y_a}$ is a minimal-length simple circuit for each $a \in [r]$ with $|Y_a| \geqslant 4$ and even,*
2. *writing $Y_T = X \setminus \left(\bigsqcup_{a\in[r]} Z_a\right)$, $s(X)|_{Y_T}$ is a clique-forest composed of disjoint local, compatible clique-trees $s(X)|_{Y_T^1}, \dots s(X)|_{Y_T^p}$, and*

3. *there are cycles* $\gamma_a \in \mathsf{EMC}_{k,k}(G), a \in [r]$ *minimally supported on* Y_a *and* $\gamma^b_T \in \mathsf{EMC}_{k,k}(G), b \in [p]$ *minimally supported on* Y^b_T*, so that* $\gamma = \sum_{a\in[r]} \gamma_a + \sum_{b\in[p]} \gamma^b_T$.

Proof. Fix a graph G and $X = \{\bar{x}^i\}_{i\in[m]} \subseteq \mathsf{ET}_{k,k}(G)$ a local, compatible collection of k-trails in G, and let $\gamma = \sum_{i\in[m]} a_i\bar{x}$, $a_i \neq 0$, be a cycle in $\mathsf{EMC}_{k,k}(G)$ which is minimally supported on X.

Suppose $s(X)$ contains a minimal-length simple circuit involving four or more vertices, so we have $Y = \{\bar{x}^j\}_{j\in J}$ for $J \subseteq [m], |J| \geqslant 4$ such that $s(X)|_Y \cong C_{|J|}$. Thus, each vertex $\bar{v}$ of $s(X)|_Y$ is implicated in two edges $(\bar{u},\bar{v})$ and $(\bar{v},\bar{w})$, and these edges cannot lie in the same maximal clique in $s(X)$ or the edge $(\bar{u},\bar{w})$ would also be present in the subgraph $s(X)|_Y$.

As before, because $s(X)|_Y$ is a cycle graph, each of the chains $\bar{x}^{j_1}$ and $\bar{x}^{j_2}$ has exactly two non-vanishing terms in its differential. And, because they share all landmarks except $x^{j_1}_t$ or $x^{j_2}_{t'}$ with $\bar{x}^{j_0}$, the missing edges in G force those to be the tth and t'th terms of the differential, per Lemma 2.1.1. There are two possibilities: Fixing $j_0 \in J$, the fact that the edge $(\bar{u},\bar{w})$ is not present in the subgraph $s(X)|_Y$ says that there are precisely two terms in the differential $\partial_{k,k}(\bar{x}^{j_0})$ which do not vanish, say $\partial^t_{k,k}(\bar{x}^{j_0})$ and $\partial^{t'}_{k,k}(\bar{x}^{j_0})$, $t < t'$. Thus, because γ is a cycle and therefore $\partial(\gamma) = 0$, from Example 2.2.1 we see that there are $\bar{x}^{j_1}$ and $\bar{x}^{j_2}$ in Y so that $\bar{x}^{j_0}$ and $\bar{x}^{j_1}$ differ precisely in their tth landmark, and similarly $\bar{x}^{j_0}$ and $\bar{x}^{j_2}$ differ in their t'th landmark. Thus, G contains the walks $\bar{x}^{j_0} = (x^{j_0}_0,\dots x^{j_0}_t,\dots,x^{j_0}_{t'},\dots,x^{j_0}_k)$, $\bar{x}^{j_1} = (x^{j_0}_0,\dots x^{j_1}_t,\dots,x^{j_0}_{t'},\dots,x^{j_0}_k)$, and $\bar{x}^{j_2} = (x^{j_0}_0,\dots x^{j_0}_t,\dots,x^{j_2}_{t'},\dots,x^{j_0}_k)$, and necessarily does not contain the edges $\{x^{j_0}_{t-1}, x^{j_0}_{t+1}\}$ and $\{x^{j_0}_{t'-1}, x^{j_0}_{t'+1}\}$.

1. The trail $\bar{x}^{j_3} = (x^{j_0}_0,\dots x^{j_1}_t,\dots,x^{j_2}_{t'},\dots,x^{j_0}_k)$ is in Y. In this case, by minimality of the circuit, $s(X)|_Y \cong C_4$. An example is illustrated in Figure 8.

2. The set Y contains two distinct trails $\bar{x}^{j'_1} = (x^{j_0}_0,\dots x^{j'_1}_t,\dots,x^{j_2}_{t'},\dots,x^{j_0}_k)$ and $\bar{x}^{j'_2} = (x^{j_0}_0,\dots x^{j_1}_t,\dots,x^{j'_2}_{t'},\dots,x^{j_0}_k)$ in order to cancel these terms. In this case, we have an equivalent pair of options again for canceling the corresponding differential terms, and this process proceeds inductively. As $|J|$ is finite, eventually we must select the option where a single trail shares both of the unresolved differential terms. As each step added an even number of trails to Y, we conclude that $|J|$ is even. Such a circuit is illustrated in Figure 9.

Select any $\bar{x}^{j_0} \in Y$, and take $\gamma_Y = \sum_{j\in J}(-1)^{d(\bar{x}^j,\bar{x}^{j_0})}a_{j_0}\bar{x}^j$, where d is the path distance in $s(X)|_Y$. This is a cycle in $\mathsf{EMC}_{k,k}(G)$ minimally supported on Y, per the discussion following Definition 2.2.5, as illustrated in Figures 8 and Figures 9.

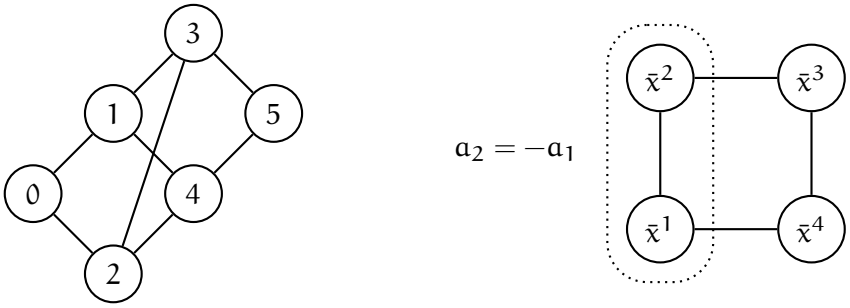


Figure 8.: Let $\bar{x}^1 = (0,1,3,5)$, $\bar{x}^2 = (0,1,4,5)$, $\bar{x}^3 = (0,2,4,5)$ and $\bar{x}^4 = (0,2,3,5)\}$, and let $Y = \{\bar{x}^1, \bar{x}^2, \bar{x}^3, \bar{x}^4\}$. (left) A graph G for which the cycle $\gamma = \bar{x}^1 - \bar{x}^2 - \bar{x}^3 + \bar{x}^4$ is minimally supported on Y. (right) The structure graph $s(Y)$ is isomorphic to C_4. Its maximal cliques, the edges, induce alternating signs on the coefficients of the cycle γ.

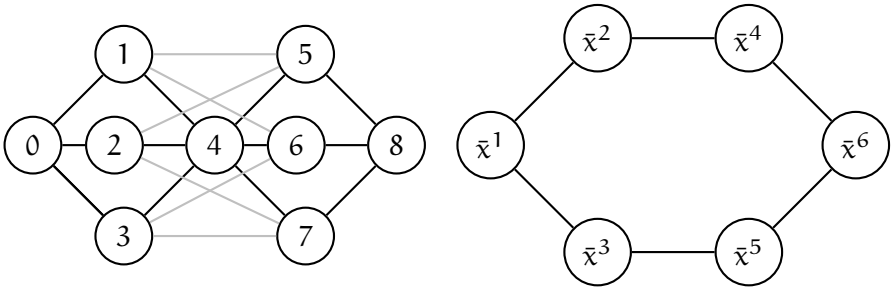


Figure 9.: Take $\bar{x}^1 = (0,1,4,5,8), \bar{x}^2 = (0,2,4,5,8), \bar{x}^3 = (0,1,4,6,8), \bar{x}^4 = (0,2,4,7,8), \bar{x}^5 = (0,3,4,6,8), \bar{x}^6 = (0,3,4,7,8)$ and $Y = \{\bar{x}^1, \bar{x}^2, \bar{x}^3, \bar{x}^4, \bar{x}^5, \bar{x}^6\}$. (left) A graph G for which the cycle $\gamma = \bar{x}^1 - \bar{x}^2 - \bar{x}^3 + \bar{x}^4 + \bar{x}^5 - \bar{x}^6$ is minimally supported on Y. The set E_{supp} is represented in black, while the edges in $E_{diff} \setminus (E_{diff} \cap E_{rem})$ are gray. (right) The structure graph $s(Y)$.

There is a nonempty collection of trails $Z \subseteq Y$ for which the coefficients in γ_Y matches their coefficient in γ. Thus, the cycle γ can be written as $\gamma = \gamma' + \gamma_Y$ for some γ' minimally supported on $X \setminus Z$, which does not contain the circuit $s(X)|_Y$.

The structure graph $s(X \setminus Z)$ has strictly fewer minimal-length simple circuits than $s(X)$. By repeating this process inductively, we can decompose any cycle supported on a local, compatible family X into summands minimally supported on local subsets $Y_1, \dots Y_r$ for which $s(X)_{Y_a}$ is a minimal-length simple circuit or such a circuit along with an adjoined 3-clique, along with a summand supported on a clique-forest subgraph of $s(X)$. This final summand decomposes into summands minimally supported on its constituent clique-trees, which each local and compatible by construction. □

Remark 2.2.1. Theorem 2.2.2 does not quite provide a cycle basis for $EMH_{k,k}$, but it does allow us to enumerate possible cycles. In the special case $k = 2$, the decomposition produces a structure graph

with a single clique, as there is only one vertex which can differ from trail to trail in any given local compatible family of 2-trails. Each edge in this clique witnesses a pair of trails which are supported on a full subgraph of G isomorphic to either C_4 or F_4. Thus, applying this decomposition to the collection of all such families in G recovers Lemma 2.1.2.

Careful accounting of the way such subgraphs interact could in principle be used to generalize Lemma 2.1.2 to higher k, but the necessary combinatorics quickly become complicated. For example, in the case $k = 3$, the only possible minimal circuits in the structure graph for a local compatible family are isomorphic to C_4, as illustrated in Figure 8. These witness subgraphs which are either isomorphic to the graph illustrated in that figure, or to variants of that graph which include edges $\{1,2\}$ and/or $\{3,4\}$. The cycles resulting from such families may interact in non-trivial ways. Rather than pursuing these combinatorics directly, our goal will be to apply these decompositions as obstructions, allowing us to study when $EMH_{k,k}(G)$ vanishes.

Remark 2.2.2. In the proof of Theorem 2.2.2, when considering minimal length circuits in $s(X)$, we saw that each such circuit would have had minimal length 4 if the trail $\bar{x}^{j_3} = (x_0^{j_0}, \dots x_t^{j_1}, \dots, x_{t'}^{j_2}, \dots, x_k^{j_0})$ were in the collection Y. It is interesting to observe that this trail is always in $T_{k,k}(G)$ for G which supports the three trails $\bar{x}^{j_0}, \bar{x}^{j_1}$, and $\bar{x}^{j_2}$ in the proof, so there are always eulerian magnitude cycles with structure graphs isomorphic to C_4 whenever larger cyclic structure graphs can be found.

If we combine Theorem 2.2.2 with the observation that every cycle in $EMC_{k,k}(G)$ can be decomposed into cycles minimally supported on disjoint local collections of trails[2], we obtain our desired spanning set for $\ker(\partial_{k,k}) \cong EMH_{k,k}(G)$.

Corollary 2.2.3. *Let G be a graph. Then $\ker(\partial_{k,k}) \subseteq EMC_{k,k}(G)$ is spanned by cycles γ minimally supported on local, compatible collections $X_\gamma \subseteq ET_{k,k}(G)$ so that $s(X_\gamma)$ is a clique-tree, or $s(X_\gamma) \cong C_d$ for d even.*

For general graphs G, it is difficult to count the number of subgraphs for which the corresponding structure graphs lie in these classes, or to count the number of decompositions a particular cycle might have along these lines. However, as we will see in Sections 3.1 and 3.2, for some important families of random graphs it is possible to obtain useful bounds on when such subgraphs can arise. To apply those results to the study of magnitude homology, however, we will need to think about how it is related to this new object.

2 Due to the the block structure of the differential.

2.3 RELATIONSHIP TO MAGNITUDE HOMOLOGY

As the eulerian magnitude chain complex is a subcomplex of the usual magnitude chain complex, we can construct a long exact sequence relating the two in the usual way. As the generators for the eulerian magnitude chain groups are a subset of those for magnitude chains, it is relatively easy to characterize the quotient: it is generated by those k-trails which repeat at least one vertex. Drawing inspiration from the study of spaces of singular curves, we make the following definition.

Definition 2.3.1. Let G be a graph. The *discriminant* (k,ℓ)*-magnitude chain group* $\mathrm{DMC}_{k,\ell}(G)$ is the quotient

$$\mathrm{DMC}_{k,\ell}(G) = \frac{\mathrm{MC}_{k,\ell}(G)}{\mathrm{EMC}_{k,\ell}(G)}$$

equipped with the usual quotient differential.

By definition of $\mathrm{EMC}_{k,\ell}(G)$ and $\mathrm{DMC}_{k,\ell}(G)$ we have the following short exact sequence of chain complexes

$$0 \to \mathrm{EMC}_{*,\ell}(G) \xrightarrow{\iota} \mathrm{MC}_{*,\ell}(G) \xrightarrow{\pi} \mathrm{DMC}_{*,\ell}(G) \to 0.$$

where ι and π are the induced inclusion and quotient maps, respectively. Therefore, we obtain a long exact sequence in homology

$$\cdots \to \mathrm{DMH}_{k+1,\ell}(G) \xrightarrow{\delta_{k+1}} \mathrm{EMH}_{k,\ell}(G) \xrightarrow{\iota_*} \mathrm{MH}_{k,\ell}(G) \xrightarrow{\pi_*} \mathrm{DMH}_{k,\ell}(G) \xrightarrow{\delta_k} \cdots, \tag{1}$$

where the map δ_{k+1} as given by the Snake Lemma is defined as

$$\begin{array}{rccc} \delta_{k+1}: & \mathrm{DMH}_{k+1,\ell}(G) & \to & \mathrm{EMH}_{k,\ell}(G) \\ & [(x_0,\ldots,x_{k+1})] & \mapsto & [\bar{\partial}_{k+1,\ell}(x_0,\ldots,x_{k+1})], \end{array}$$

where $\bar{\partial}_{k+1,\ell} = \sum_{i\in[k-1]}(-1)^i \bar{\partial}^i_{k+1,\ell}$ with

$$\bar{\partial}^i_{k+1,\ell}(x_0,\ldots,x_{k+1}) = \begin{cases} (x_0,\ldots,\hat{x}_i,\ldots,x_k) & \text{if } (x_0,\ldots,\hat{x}_i,\ldots,x_k) \in \mathrm{ET}_{k,\ell}(G) \\ 0 & \text{otherwise.} \end{cases}$$

Note that, in general, it is not true that $\mathrm{EMH}_{k,\ell}(G)$ is a subgroup of $\mathrm{MH}_{k,\ell}(G)$, so the long exact sequence in equation (1) does not split. Consider, for example the graph G in Figure 10 and consider the element $(0,1,2,3,1,4) \in \mathrm{MC}_{5,5}(G)$. Applying the differential map, we have

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

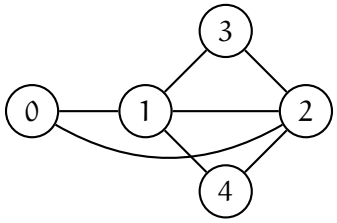


Figure 10.: In this graph, the element $[(0,1,2,3,4)]$ is trivial in $MH_{4,5}(G)$ but not in $EMH_{4,5}(G)$.

So the generator $(0,1,2,3,4) \in MC_{4,5}(G)$ is a boundary and will be trivial in $MH_{4,5}(G)$. On the other hand, the same tuple $(0,1,2,3,4)$ cannot be a boundary in $EMC_{4,5}(G)$ since the 5-trail already contains all vertices of G. Hence, in general the map $EMH_{k,\ell} \to MH_{k,\ell}$ is not injective.

Despite the fact that the long exact sequence in equation (1) fails to split in general, along the $k = \ell$ line we can still leverage the relationships between generators to make some interesting observations. As $EMH_{k,k}(G)$ is the first potentially non-trivial group in the long exact sequence of equation (1), for every k it holds that the map $i_* : EMH_{k,k}(G) \to MH_{k,k}(G)$ is injective. For the same reason, if $EMH_{k,k}(G)$ is trivial, then $MH_{k,k}(G)$ is a subgroup of $DMH_{k,k}(G)$. In fact, in the regime where $EMH_{k,k}(G)$ is trivial, we are able to explicitly describe the generators of $MH_{k,k}(G)$ and, if $k \geqslant 5$ establish an isomorphism $MH_{k,k}(G) \cong DMH_{k,k}(G)$.

Theorem 2.3.1. *Fix an integer $k \geqslant 2$. Let $G = (V,E)$ be a graph and suppose $EMH_{n,n}(G) \cong \langle 0 \rangle$ for $2 \leqslant n \leqslant k$. Then every generator of $MH_{k,k}(G) \cong \ker(\partial_{k,k})$ is a k-trail of the form $(x_0, x_1, x_0, x_1, \dots)$ which visits only one pair of vertices $x_0 \neq x_1 \in V$,*

In Sections 3.1 and 3.2, we will describe regimes in which this theorem holds for important classes of random graphs.

Proof. Fix $k \geqslant 2$. Consider a graph $G = (V,E)$ and assume $EMH_{n,n}(G) \cong \langle 0 \rangle$ for $2 \leqslant n \leqslant k$.

First, we consider the case of cycles supported on individual trails. Let $\bar{x} = (x_0, \dots, x_k) \in MC_{k,k}(G)$ and suppose $\partial_{k,k}(\bar{x}) = 0$. Suppose we have $i < j \in [k]_0$ such that there is an eulerian sub-trail $(x_i, \dots, x_j) \in ET_{j-i,j-i}(G)$ of $\bar{x}$. If $j - i \geqslant 2$, then as $EMH_{j-i,j-i}(G) \cong \ker(\partial_{j-i,j-i}|_{EMC_{j-i,j-i}(G)})$ vanishes, for some $a \in [j-i-1]$ we have $\partial^a_{j-i,j-i}(x_i, x_{i+1}, \dots x_j) \neq 0$, and thus $\partial^{i+a}_{k,k}\bar{x} \neq 0$. Thus $j = i+1$. However, for any k-trail, pairs of sequential vertices are distinct and thus must form a maximal eulerian subtrail, so $x_i = x_{i+2}$ for every $i \in [k-1]_0$. Thus, $\bar{x} = (x_0, x_1, x_0, \dots)$.

Consider now an element $\sum_{i\in[m]} a_i \bar{x}^i \in MC_{k,k}(G)$ so that $\partial_{k,k}(\sum_{i\in[m]} a_i \bar{x}^i) = 0$, and for which $\partial_{k,k}(\bar{x}^i) \neq 0$ for each $i \in [m]$. By the contrapositive of the argument above, we see that each

involved k-trail $\bar{x}^i$ visits at least three vertices, and indeed non-vanishing terms in the differential can only occur away from subtrails which repeatedly cross the same edge. Suppose $\partial^r_{k,k}(\bar{x}^i) \neq 0$ then $r \in [k-1]$, x^i_{r-1}, x^i_r, and x^i_{r+1} are all distinct. Consider $(x^i_{r-1}, x^i_r, x^i_{r+1}) \in ET_{2,2}(G) \subseteq EMC_{2,2}(G)$, and observe that $\partial^r_{k,k}(\bar{x}^i) \neq 0$ implies $\partial^1_{2,2}((x^i_{r-1}, x^i_r, x^i_{r+1})) \neq 0$. However, if this boundary is to cancel, there must be $j \neq i \in [m]$ so that $\partial^r_{k,k}(\bar{x}^i) = \partial^r_{k,k}(\bar{x}^j) \neq 0$. These two trails agree except at the term $x^i_r \neq x^j_r$, so we have $(x^i_{r-1}, x^j_r, x^i_{r+1}) \in ET_{2,2}(G)$ as well, and their difference forms a non-zero homology class, contradicting the fact that $EMH_{2,2}(G) \cong \langle 0 \rangle$. Thus, no such elements can exist, so all cycles in $MH_{k,k}(G)$ are of the required form. □

The following is an immediate consequence of Theorem 2.3.1.

Corollary 2.3.2. *Fix an integer* $k \geqslant 2$. *If* $G = (V, E)$ *is such that* $EMH_{n,n}(G) = \langle 0 \rangle$ *for* $2 \leqslant n \leqslant k$, *then* $|MH_{k,k}(G)| = 2|E| = |MH_{1,1}(G)|$.

In a similar spirit to Theorem 2.3.1, under weaker vanishing conditions, we can prove the following isomorphism.

Theorem 2.3.3. *Fix* $k \geqslant 5$. *Suppose* $EMH_{2,2}(G) \cong EMH_{k,k}(G) \cong \langle 0 \rangle$, *then* $MH_{k,k}(G) \cong DMH_{k,k}(G)$.

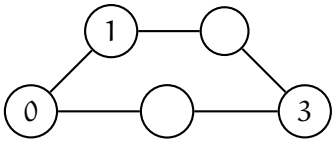
Proof. Let $k \geqslant 5$ and suppose $EMH_{2,2}(G) \cong EMH_{k,k}(G) \cong \langle 0 \rangle$. By the long exact sequence in equation (1), to prove that $MH_{k,k}(G) \cong DMH_{k,k}(G)$ it suffices to show that $EMH_{k-1,k}(G) \cong \langle 0 \rangle$.

Suppose that $\bar{x} = (x_0, \ldots, x_{k-1}) \in ET_{k-1,k} \subseteq EMC_{k-1,k}(G)$. Observe that in such a $(k-1)$-trail, we must have $d(x_i, x_{i+1}) = 1$ for all but one pair of consecutive vertices, for which $d(x_j, x_{j+1}) = 2$. Since $k \geqslant 5$, regardless of the value of $j \in [k-1]_0$, the trail $\bar{x}$ will have at least three consecutive vertices $x_{r-1}, x_r, x_{r+1}, r \in [k-2]$ for which $d(x_{i-r}, x_r) = d(x_r, x_{r+1}) = 1$. That is, $(x_{r-1}, x_r, x_{r+1}) \in ET_{2,2}(G)$.

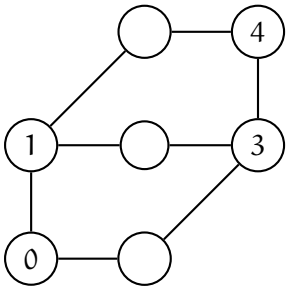
Now, suppose by contradiction $\partial_{k-1,k}(\bar{x}) = 0$. Then $\partial^r_{k-1,k}(\bar{x}) = 0$, whence $\partial^1_{2,2}(x_{r-1}, x_r, x_{r+1}) = 0$ and so $\{x_{r-1}, x_{r+1}\} \in E$ by Lemma 2.1.1. However, then $G|_{\{x_{r-1}, x_r, x_{r+1}\}} \cong C_3$, which by Lemma 2.1.2 contradicts the assumption that $EMH_{2,2}(G) \cong \langle 0 \rangle$. So, we have $\{x_{r-1}, x_{r+1}\} \notin E$ and thus $\partial^r_{k-1,k}(\bar{x}) \neq 0$.

Suppose now that there exists some linear combination $\sum_{i \in [m]} a_i \bar{x}^i \in EMC_{k-1,k}(G)$ for which $\partial_{k-1,k}(\sum_{i \in [m]} a_i \bar{x}^i) = 0$. Consider $\bar{x}^1 = (x^1_0, \ldots, x^1_{k-1})$ and apply the above argument to again find $r \in [k-1]$ so that $d(x^1_{i-r}, x^1_r) = d(x^1_r, x^1_{r+1}) = 1$. As $\partial^r_{k-1,k}(\bar{x}^1) \neq 0$, in order to cancel this term there must be some $\bar{x}^j$, $j \neq 1$ for which $\partial^r(\bar{x}^j) = \partial^r(\bar{x}^1)$, $x^1_r \neq x^j_r$ and $x^1_i = x^j_1$ for $i \neq r$. However, then $G|_{\{x^1_{r-1}, x^1_r, x^1_{r+1}, x^j_r\}}$ is isomorphic to either C_4 or F_4, as in Lemma 2.1.2, in either case again contradicting the assumption that $EMH_{2,2}(G) \cong \langle 0 \rangle$. Thus, we must have $EMH_{k-1,k}(G) \cong \langle 0 \rangle$ as required. □

Note that in Theorem 2.3.3 the hypothesis $k \geqslant 5$ was essential. In $EMC_{2,3}(G)$ and $EMC_{3,4}(G)$, graphs such as those illustrated in Figure 11 support trails which generate nontrivial homology classes.



(a) In the graph C_5, the generator $(0,1,3) \in EMC_{2,3}(G)$ descends to a nontrivial homology cycle in $EMH_{2,3}(G)$.



(b) In the pictured graph, the generator $(0,1,3,4) \in EMC_{3,4}(G)$ represents a nontrivial a homology cycle in $EMH_{3,4}(G)$.

Figure 11.: Graphs for which $EMH_{k,k}(G) \cong \langle 0 \rangle$ but $EMH_{k-1,k}(G)$ is nontrivial for some $k < 5$.

One last question about the relationship between eulerian and standard magnitude homology arises from the example in Figure 10, where we have a non-trivial eulerian magnitude homology cycle that becomes trivial in standard magnitude homology. Using the tools we have developed, we can provide a partial answer in the first nontrivial case, when $\ell = k+1$, determining which non-trivial eulerian magnitude homology cycles generated by a single tuple $\bar{x}$ become trivial in standard magnitude homology.

Theorem 2.3.4. *Let G be a graph and fix a nonnegative integer k. Let $\bar{x} = (x_0, \ldots, x_{k-1}) \in ET_{k-1,k}(G)$ such that $[\bar{x}] \in EMH_{k-1,k}(G)$ is nontrivial. Then $[\bar{x}]$ is trivial in $MH_{k-1,k}(G)$ if and only if there are some $i, r \in [k-2]$ with $r - i \geqslant 2$ so that the k-trail*

$$(x_0, \ldots, x_i, \ldots, x_r, x_i, x_{r+1}, \ldots, x_{k-1})$$

appears in G, in which case $G|_{\{x_i, x_{i+1}, \ldots, x_r\}} \in \Gamma(\mathcal{C}_{r-i})$, where $\mathcal{C}_{r-i}$ is the complete class graph for which $\alpha(\mathcal{C}_{r-i}) \cong C_{r-i}$ is cyclic on $(r-i)$ vertices[3].

Proof. Suppose $G = (V, E)$ is a graph and $\bar{x} = (x_0, \ldots, x_{k-1}) \in ET_{k-1,k}(G)$ such that $[\bar{x}] \in EMH_{k-1,k}(G)$ is non-trivial. So, there exists no $\bar{x}' \in ET_{k,k}(G)$ with $\partial_{k,k}(\bar{X}') = \bar{x}$. Thus, for $[\bar{x}]$ to be trivial in $MH_{k-1,k}(G)$, there must exist $\bar{x}' \in T_{k,k}(G) \setminus ET_{k,k}(G)$ so that $\partial^r_{k,k}(\bar{x'}) = \bar{x}$ for some $r \in [k-1]$. Thus, $\bar{x}' = (x_0, \ldots, x_{r-1}, y, x_r, \ldots, x_{k-1})$, as pictured in Figure 12. However, since $\bar{x}' \notin ET_{k,k}(G)$, we must have $y = x_i$ for some $i \in [k-2]$.

3 Here, we take C_2 to be the graph with two vertices and one edge to simplify the statement.

In this case, $\bar{x}'' = (x_0, \ldots, x_{i-1}, x_{i+1}, \ldots, x_{r-1}, x_i, x_r, \ldots, x_{k-1})$[4] is also an eulerian $(k-1)$-trail that could appear as a term in $\partial_{k,k}(\bar{x}')$. However, for the same reason as in Lemma 2.1.1, the fact that $\partial^i_{k-1,k}\bar{x} = 0$ implies that $\{x_{i-1}, x_{i+1}\} \in E$, so $\mathrm{len}(\bar{x}'') = k-1$. Thus, this term is zero in the differential. All other terms in $\partial_{k,k}(\bar{x}')$ are zero because they agree with those of $\bar{x}$, indeed when computing all other terms in $\partial_{k,k}(\bar{x}')$ we are removing the same vertices we remove when computing $\partial_{k,k}(\bar{x})$. Thus, $[\bar{x}] = 0$ in $MH_{k-1,k}(G)$. □

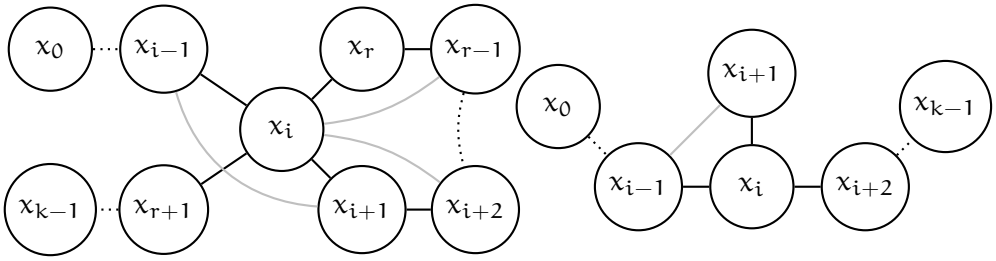


Figure 12.: Relevant subgraphs of minimal graphs in classes which support non-trivial eulerian magnitude homology cycles which become trivial in standard magnitude homology, per the proof of Theorem 2.3.4. Note the asymmetry of those gray edges around x_i which must be present to enforce the vanishing differential for $(x_0, \ldots, x_{k-1}) \in EMC_{k-1,k}(G)$.

4 Supposing $r < i$. If $i > r$, reverse the appearance order in this sequence.

3

LIMIT THEOREMS FOR RANDOM GRAPHS

In the chapter we turn our attention to the computation of eulerian magnitude homology of Erdős-Rényi random graphs (section 3.1) and random geometric graphs (section 3.2). In both cases, we will develop an in-depth analysis of the eulerian magnitude homology groups of the $k = \ell$ diagonal by identifying the subgraphs induced by homology cycles, producing a vanishing threshold and providing an explicit formula for the asymptotic size of the eulerian magnitude homology groups.

3.1 EULERIAN MAGNITUDE HOMOLOGY OF ERDŐS-RÉNYI GRAPHS

The *Erdős-Rényi (ER) model* for random graphs, $G(n, p)$, introduced in [40], is one of the most widely studied and applied models for random graphs. As the maximum entropy distribution on graphs with an expected proportion of edges, the ER model serves as a useful null model in a broad range of scientific and engineering applications. For this reason, clique complexes of ER graphs have long been an object of interest in the stochastic topology community [64, 65, 67]. We expect it to serve a similar function as a baseline model for understanding magnitude homology.

Definition 3.1.1. The *Erdős-Rényi (ER) model* $G(n, p) = (\Omega, P)$ is the probability space where Ω is the discrete space of all graphs on n vertices, and P is the probability measure that assigns to each graph $G \in \Omega$ with m edges probability

$$P(G) = p^m(1-p)^{\binom{n}{2}-m}.$$

We can sample an ER graph $G \sim G(n, p)$ on n vertices with parameter $p \in [0, 1]$ by determining whether each of the $\binom{n}{2}$ potential edges is present via independent draws from a Bernoulli distribution with probability p. In order to study the limiting behavior of these models as $n \to \infty$, it is often useful to change variables so that p is a function of n. In Section 3.1.1, we will take $p = n^{-q}$, $q \in [0, \infty)$, following [64].

In what follows, we will study the first diagonal ($k = \ell$) for the eulerian magnitude homology groups of ER graphs as $n \to \infty$. We would like to think of $EMH_{k,k}(G(n,p))$ as a random variable valued in finitely generated abelian groups. As we know from Corollary 1.2.2, in this case the groups are free abelian, and counting generators is sufficient to completely understand these groups. Write $\beta_{k,k}(G) = \operatorname{rank}(EMH_{k,k}(G)))$ for the (k,k)-EMH Betti number of G, and $\beta_{k,k}(n,p) = \operatorname{rank}(EMH_{k,k}(G(n,p)))$ for the corresponding random variable. In Theorem 3.1.4, we use the spanning set from Theorem 2.2.2 to establish a threshold in terms of q beyond which $\mathbb{E}[\beta_{k,k}(n,p)]$ vanishes, as illustrated in Figure 13. Then, in Theorem 3.1.6, we present an explicit formula for the asymptotic size of $\mathbb{E}[\beta_{k,k}(n,p)]$, and in Theorem 3.1.7 we prove a Central Limit Theorem for $\beta_{k,k}(n,p)$.

3.1.1 *A vanishing threshold for* $EMH_{k,k}(G)$

Here, we leverage the connection between eulerian magnitude chains along the $k = \ell$ line and the structure of classes $\Gamma(\mathcal{H})$ of graphs that support those chains established in Section 2.1 to study how the expected homology groups for $G(n, n^{-q})$ behave as $n \to \infty$. By counting subgraphs that can support chains in the kernel of the differential, we are able to establish a q threshold beyond which the limiting $\ker(\partial_{k,k})$ vanishes in expectation, meaning the groups $EMH_{k,k}(G(n,n^{-q}))$ vanish. We will break the argument into two stages. First, we will consider individual basis elements $\bar{x} \in ET_{k,k}(G) \subseteq EMC_{k,k}(G)$ such that $\partial_{k,k}\bar{x} = 0$, such as those we studied in Lemma 2.1.3. Then, we will demonstrate that in all other cases, the corresponding chains have a higher vanishing threshold.

First, we require a simple observation, which follows from the fact that edges are drawn independently via Bernoulli random trials.

Lemma 3.1.1. *Fix positive n and $p \in [0,1]$. Let $G = (V,E) \sim G(n,p)$ be an ER random graph, and $W \subseteq V$, let $\mathcal{H} = (W, E_S, E_B)$ a class graph. Then the probability that $G|_W \in \Gamma(\mathcal{H})$ is $p^{|E_S|}(1-p)^{\binom{|W|}{2}-|E_S|-|E_B|}$.*

Now, we can return to the issue at hand.

Lemma 3.1.2. *Fix a non-negative integer k, and let $q > \frac{k+1}{2k-1}$. Then*

$$\mathbb{E}\left[\left|\left\{\bar{x} \in ET_{k,k}(G(n,n^{-q})) \,:\, \partial_{k,k}\bar{x} = 0\right\}\right|\right] \to 0$$

as $n \to \infty$.

Proof. Sample a graph $G \sim G(n,n^{-q})$ and let $\bar{y} = (y_0, \ldots, y_k) \in ET_{k,k}(G)$ be a k-trail of length k in G. We wish to estimate the probability of the corresponding generator being a cycle in $EMC_{k,k}(G)$. Taking $\mathcal{H}(\{\bar{y}\}) = (L(\bar{y}), E_S, E_B)$ to be the complete class graph from

Lemma 2.2.1, we see that the set E_S has $2k-1$ elements, so by Lemma 3.1.1, the probability that $G|_{\{y_0,\dots,y_k\}} \in \Gamma(\mathcal{H}(\{\bar{y}\}))$ is p^{2k-1}.

However, more than one such generator may be supported on $G|_{\{y_0,\dots,y_k\}}$; each such generator corresponds to an isomorphic copy of $H_k = \alpha(\mathcal{H}(\{\bar{y}\}))$. We can bound above the number of isomorphic copies of H_k that appear on any collection of $(k+1)$ vertices in G by $c(\Delta_{k+1}, H_k)$, which we recall denotes the number of full subgraphs of Δ_{k+1} isomorphic to H_k. We apply these estimates to provide an upper bound on the expectation of the number of copies of H_k appearing in such a sampled G as $n \to \infty$.

$$\begin{aligned}
\mathbb{E}[c(G(n, n^{-q}), H_k)] &\leqslant \binom{n}{k+1} c(\Delta_{k+1}, H_k) p^{2k-1} \\
&\sim \frac{n^{k+1}}{(k+1)!} c(\Delta_{k+1}, H_k) n^{q(1-2k)} \\
&= \frac{c(\Delta_{k+1}, H_k)}{(k+1)!} n^{q(1-2k)+(k+1)} \\
&\xrightarrow{n\to\infty} \begin{cases} 0, \text{ if } q > \frac{k+1}{2k-1} \\ \infty, \text{ if } 0 < q < \frac{k+1}{2k-1}. \end{cases}
\end{aligned}$$

Thus, we conclude that no such generators are expected to exist when $q > \frac{k+1}{2k-1}$ as $n \to \infty$. □

Note that the situation described in Lemma 3.1.2, where a single tuple $\bar{x}$ generates an EMH-cycle, corresponds to a structure graph $s(\{\bar{x}\})$ consisting of a single vertex. As we will see in the following lemma, the limiting probability of observing cycles minimally supported on families of trails with more complex structure graphs goes to zero in a larger q range than this singleton case.

Lemma 3.1.3. *Fix non-negative integers k, n, and fix $m \geqslant 2$. Let $q > \frac{k+1}{2k-1}$. Let $X = \{\bar{x}^i\}_{i\in[m]} \subseteq ET_{k,k}(\Delta_n)$ be a local, compatible collection of trails so that $s(X) \cong C_d$ for some even d or $s(X)$ is a clique-tree. Then, as $n \to \infty$,*

$$\mathbb{E}[c(G(n, n^{-q}), \alpha(\mathcal{H}(X)))] \to 0.$$

Proof. Consider first the case where $s(X) \cong C_d$, which will be similar to those depicted in Figures 8 or 9, with $|X| \geqslant 4$. Taking $t < t'$ and $\bar{x}^{j_0} = (x_0^{j_0}, \dots x_t^{j_0}, \dots, x_{t'}^{j_0}, \dots, x_k^{j_0})$, as in the proof of Theorem 2.2.2, and following the construction of the cycle therein, we see that all but two of the elements of X introduce exactly one new vertex to $\mathcal{H}(X)$. The other is the initial trail, consisting of $k+1$ new vertices, and the last does not introduce anything new. So, by construction $\mathcal{H}(X)$ must have precisely $k + |X| - 1$ vertices. Further, following Definition 2.2.6, E_{supp} will consist of exactly the requisite $(k+1) + 2(|X|-2) + (|X|-3) = k + 3|X| - 6$ edges implicated in these trails, where the last $(|X|-3)$ edges connect the $(|X|-1)$ newly added vertices. E_{diff} can be

decomposed as follows. Start with the $(k-1)$ edges for the first trail $\bar{x}^1$. Iteratively, select a new maximal clique in $s(X)$ containing $\bar{x}^1$, and observe that each trail in that clique differs from $\bar{x}^1$ in precisely one vertex, and thus there are at most two edges including that vertex that must be added to E_{diff} to ensure the corresponding differential terms for those trails vanish. Iterating through cliques in this way, we see that each other trail in X adds between one and two edges, and so $(k-1)+|X|-1 \leqslant E_{\text{diff}} \leqslant (k-1)+2(|X|-1)$, depending on the values of t and t'. Finally, E_{rem} will contain precisely the edges $\{x^{j_0}_{t-1}, x^{j_0}_{t+1}\}$ and $\{x^{j_0}_{t'-1}, x^{j_0}_{t'+1}\}$. Thus, as $|E_{\text{diff}} \cap E_{\text{rem}}| \leqslant 2$, $|E^X_S| \geqslant (k+3|X|-6)+(k-1)+|X|-2 \geqslant 2k+2|X|-1$ for every $|X| \geqslant 4$. Taking W to be any collection of $k+|X|-1$ vertices in $G \sim G(n,p)$, Lemma 3.1.1 then says that the probability that $G|_W \in \Gamma(\mathcal{H}(X))$ is bounded above by $p^{2k+2|X|-1}$.

Now, we can apply the same reasoning as in the proof of Lemma 3.1.2 to count subgraphs of G isomorphic to $H_k \cong \alpha(\mathcal{H}(X))$.

$$\begin{aligned}
\mathbb{E}[c(G(n,n^{-q}),H_k)] &\leqslant \binom{n}{k+|X|} c(\Delta_{k+|X|},H_k)p^{2k+2|X|-1} \\
&\sim \frac{n^{k+|X|}}{(k+|X|)!}c(\Delta_{k+|X|},H_k)n^{q(1-2|X|-2k)} \\
&= \frac{c(\Delta_{k+2|X|},H_k)}{(k+|X|)!}n^{q(1-2|X|-2k)+(k+|X|)} \\
&\xrightarrow{n\to\infty} \begin{cases} 0, \text{ if } q > \frac{k+|X|}{2k+2|X|-1} \\ \infty, \text{ if } 0 < q < \frac{k+|X|}{2k+2|X|-1}. \end{cases}
\end{aligned}$$

In particular, as $\frac{k+1}{2k-1} > \frac{k+|X|}{2k+2|X|-1}$, we conclude that no local compatible collection of trails with cyclic structure graph is expected to be supported in G as $n \to \infty$ when $q > \frac{k+1}{2k-1}$.

The case where $s(X)$ a clique-tree, as in Figure 7, is similar. Each vertex in $s(X)$ beyond the first adds at most one vertex to $\mathcal{H}(X)$. Suppose y such vertices are added, so $\mathcal{H}(X)$ has $k+y+1$ vertices. For each such new vertex there are two edges added to E_{supp}. The set E_{rem} contains no more edges than the number of maximal cliques, r. And, E_{diff} contains at least $((k-1)+y)$ edges. Then, since $y \geqslant r$, $|E^X_S| \geqslant (k+2y)+(k-1)+y-r \geqslant 2k+2y-1$. So, again taking W to be a subset of the vertices of $G \sim G(n,p)$ of the appropriate size and invoking Lemma 3.1.1, we have the probability that $G|_W \in \Gamma(\mathcal{H}(X))$ is bounded above by $p^{2k+2y-1}$. We now apply an identical counting argument to conclude the required vanishing bound for collections of trails with clique-tree structure graphs.

□

Now, by Theorem 2.2.2, any cycle in $EMH_{k,k}(G)$ is either minimally supported on a singleton $\{\bar{x}^1\}$ or can be decomposed into cycles mini-

mally supported on collections with structure graphs given by clique-trees and cycles. Thus, Lemma 3.1.2 and Lemma 3.1.3 together provide a vanishing threshold for eulerian magnitude homology on the $k = \ell$ line.

Theorem 3.1.4. *Fix* k *and* $q > \frac{k+1}{2k-1}$. *As* $n \to \infty$,

$$\mathbb{E}\left[\beta_{k,k}(n, n^{-q})\right] \to 0.$$

Figure 13 illustrates the q range where the first diagonal of the eulerian magnitude homology of an Erdős-Rényi random graph $G(n, n^{-q})$ is expected to vanish as $n \to \infty$.

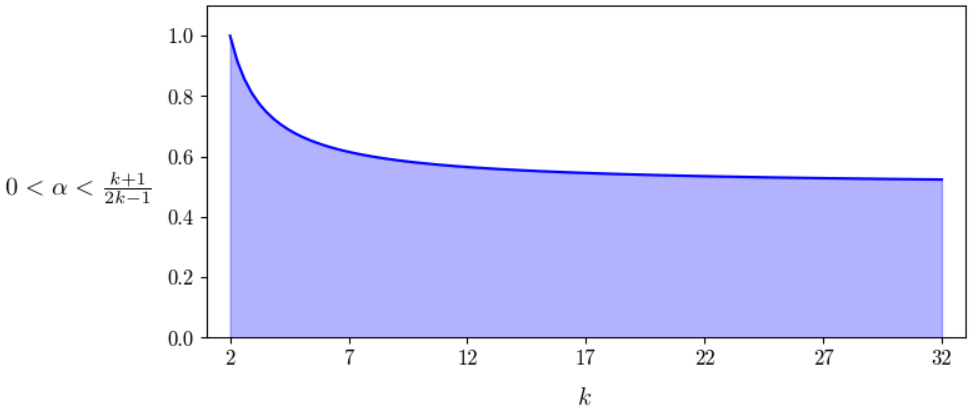


Figure 13.: The shaded region below the curve is the q vs k region for which we can have non-vanishing $EMH_{k,k}(G)$ in expectation as $n \to \infty$ for graphs $G \sim G(n, n^{-q})$. By Theorems 2.3.1 and 2.3.3, in the non-shaded region asymptotically almost surely $MH_{k,k}(G) \leqslant DMH_{k,k}(G)$, and if also $k \geqslant 5$, then $MH_{k,k}(G) \cong DMH_{k,k}(G)$.

Combining the vanishing threshold in Theorem 3.1.4 with the behavior of $MH_{k,k}(G)$ when eulerian magnitude homology vanishes as in Theorem 2.3.1, we obtain the following characterization of the expected behavior of $MH_{k,k}(G(n,p))$ as $n \to \infty$.

Corollary 3.1.5. *Let* $G = G(n, n^{-q})$ *be an Erdős-Rényi random graph. If* $q > \frac{k+1}{2k-1}$ *then the magnitude homology group* $MH_{k,k}(G)$ *is the subgroup of the discriminant magnitude homology group* $DMH_{k,k}(G)$ *generated by tuples of the form* $(x_0, x_1, x_0, \dots)$ *for which the induced path only revisits the same edge* (x_0, x_1).

3.1.2 *Asymptotic behavior of* $\beta_{k,k}(G)$ *for ER random graphs*

Given the prominence of cliques in our computations in Section 3.1.1, it may come as little surprise that the pioneering work of Kahle and Meckes on the limiting behavior of Betti numbers in random clique

complexes would have application in this context. Following [67, Theorem 2.3] and [66, Theorem 1.1], in this section we compute expectations of these values for Erdős-Rényi random graphs as $n \to \infty$, see Figure 14. Because we rely on many of the same combinatorial structures, in the name of brevity we will refer the reader to computations in these papers when the details are identical.

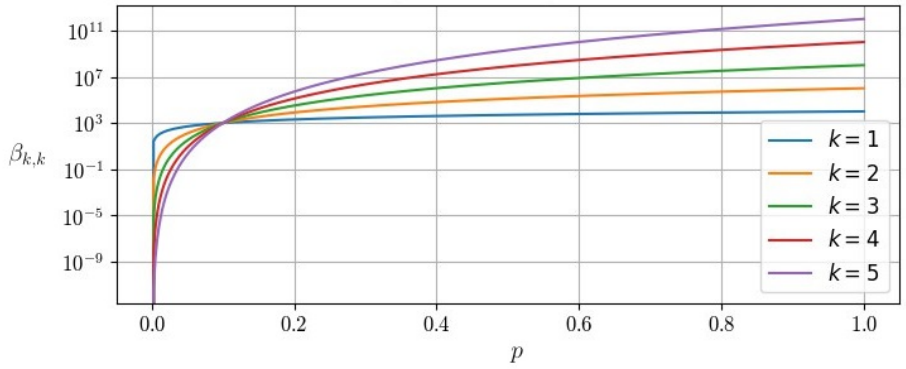


Figure 14.: The expected value for the Betti numbers $\beta_{k,k}$, $k \in [1,5]$, of an Erdős-Rényi random graph is plotted vertically against the probability p. In this example $n = 100$ and the y-axis is on a base 10 logarithmic scale to better visualize the large differences in values. Notice that the the curves all intersect at roughly $p = 0.1 = n^{-\alpha}$ with $n = 100$ and $\alpha = \frac{1}{2} \sim \frac{k+1}{2k-1}$.

Theorem 3.1.6. *Fix* k *nonnegative and write* $\beta_{k,k} = \operatorname{rank}(EMH_{k,k}(G(n,p)))$.

If $p = 1 - O\left(\frac{\ln(n)}{n}\right)$ *then*

$$\lim_{n\to\infty} \frac{\mathbb{E}[\beta_{k,k}]}{n^{k+1}p^{2k-1}} = 1.$$

Proof. In the regime where eulerian magnitude homology does not vanish we know that $0 < q < \frac{k+1}{2k-1}$. From Lemma 3.1.2 the Betti number $\beta_{k,k}$ is at least the number of single vertex structure graphs,

$$\binom{n}{k+1}(k+1)!p^{2k-1} \leqslant \beta_{k,k}, \tag{2}$$

and from Lemma and 3.1.3 it is at most the number of combinations obtained by replacing m edges (x_{i-1}, x_{i+1}) with $2m$ edges (x_{i-1}, x'_i), (x'_i, x_{i+1}):

$$\begin{aligned}\beta_{k,k} &\leqslant \sum_{m=0}^{k-1} \binom{n}{k+1+m}(k+1+m)!p^{2k-1-m+2m}(1-p)^m \\ &= \sum_{m=0}^{k-1} \binom{n}{k+1+m}(k+1+m)!p^{2k-1+m}(1-p)^m,\end{aligned} \tag{3}$$

which for $p = n^{-q}$ is equal to

$$\begin{aligned}&\sum_{m=0}^{k-1} \binom{n}{k+1+m}(k+1+m)!n^{-q(2k+m-1)}(1-n^{-q})^m \\ &\sim \sum_{m=0}^{k-1} n^{k+1+m}n^{-q(2k+m-1)}(1-n^{-q})^m \\ &= n^{(k+1)-q(2k-1)} + \sum_{m=1}^{k-1} n^{k+1+m}n^{-q(2k+m-1)}(1-n^{-q})^m.\end{aligned} \tag{4}$$

Now, the assumption that $p = 1 - O\left(\frac{\ln(n)}{n}\right)$ is telling us that p behaves asymptotically as $n^{-\frac{1}{n}}$. So $(1-n^{-q}) \sim (1-n^{-\frac{1}{n}}) \xrightarrow{n\to\infty} 0$, and we see that the only subgraphs asymptotically contributing to eulerian magnitude homology are the ones for which $m = 0$, i.e. the ones induced by a single tuple $(x_0, \ldots, x_k)$.

Therefore, putting together equations 2, 3 and 4 we obtain

$$\lim_{n\to\infty} \frac{\mathbb{E}[\beta_{k,k}]}{n^{k+1}p^{2k-1}} = 1.$$

□

We can also prove a central limit type result for $\beta_{k,k}(G(n,p))$. In what follows we will denote by $\Phi(\cdot)$ the distribution function of the standard normal distribution. Recall that a sequence $\{X_n\}_n$ of random variables converges weakly to a random variable X, and we write $X_n \Rightarrow X$, if $\lim_{n\to\infty} \mathbb{E}[f(X_n)] = \mathbb{E}[f(X)]$ for all bounded continuous functions f.

Theorem 3.1.7. *Fix k nonnegative and write $\beta_{k,k} = \operatorname{rank}(EMH_{k,k}(G(n,p)))$. If $p = 1 - O\left(\frac{\ln(n)}{n}\right)$, then as $n \to \infty$ we have that the sequence $\{\mathbb{E}[\beta_{k,k}]\}_k$ converges weakly to the Betti number $\beta_{k,k}$,*

$$\frac{\beta_{k,k} - \mathbb{E}[\beta_{k,k}]}{\sqrt{Var(\beta_{k,k})}} \Rightarrow N(0,1).$$

Proof. Let G be a graph. Write $t_{k,\ell} = |ET_{k,\ell}(G(n,p))|$ for the random variable providing the number of eulerian (k)-trails of length ℓ in a graph sampled from $G(n,p)$. By Lemma 1.2.1, $EMC_{k+1,k}(G) = \langle 0 \rangle$,

so $EMH_{k,k}(G)$ is free abelian. From the rank-nullity theorem, we conclude that

$$-t_{k-1,k} + t_{k,k} \leqslant \beta_{k,k} \leqslant t_{k,k}. \tag{5}$$

Observe that $t_{k,k}$ is the number of $(k+1)$-tuples of distinct vertices that induce in the graph G a path of length k. There are $\binom{n}{k+1}(k+1)!$ such paths possible for a graph on n vertices, and for $G \sim G(n,p)$ they each appear with probability p^k. For $t_{k-1,k}$, notice that this term counts (k)-tuples of vertices that induce in the graph G a path of length k. So we still need to account for $k+1$ visited vertices (even if one of those is not explicitly mentioned in the tuple). Further, we need to account for two facts: first, each edge of the induced path appears with probability p^k; and second, at least one of the edges (x_{i-1}, x_{i+1}) will not appear in the induced graph. This leaves us with $\binom{n}{k+1}(k+1)!p^k(1-p^{k-1})$ such tuples.

To prove the result we will need the following three intermediate claims:

1. $\lim_{n\to\infty} \frac{Var(t_{k,k})}{Var(t_{k,k}-t_{k-1,k})} = 1$
2. $\frac{t_{k,k}-\mathbb{E}[t_{k,k}]}{\sqrt{Var(t_{k,k})}} \Rightarrow N(0,1)$ as $n \to \infty$
3. $\frac{(t_{k,k}-t_{k-1,k})-\mathbb{E}[t_{k,k}-t_{k-1,k}]}{\sqrt{Var(t_{k,k}-t_{k-1,k})}} \Rightarrow N(0,1)$ as $n \to \infty$.

Deferring their proofs, we now apply these three claims to prove the theorem. Take $\alpha \in \mathbb{R}$. Then from the inequalities in equation (5), we get

$$\mathbb{P}\left[\frac{t_{k,k}-\mathbb{E}[t_{k,k}]}{\sqrt{Var(t_{k,k})}} \leqslant \alpha\right] \leqslant \mathbb{P}\left[\frac{\beta_{k,k}-\mathbb{E}[t_{k,k}]}{\sqrt{Var(t_{k,k})}} \leqslant \alpha\right] \leqslant \mathbb{P}\left[\frac{(t_{k,k}-t_{k-1,k})-\mathbb{E}[t_{k,k}]}{\sqrt{Var(t_{k,k})}} \leqslant \alpha\right].$$

Because of Claim 2, the left term $\mathbb{P}\left[\frac{t_{k,k}-\mathbb{E}[t_{k,k}]}{\sqrt{Var(t_{k,k})}} \leqslant \alpha\right]$ tends, as $n \to \infty$, to $\Phi(\alpha)$, the distribution function of the standard normal distribution. For the right-hand-side we can proceed analogously as in [67, Theorem 2.3]. Thus, note that

$$\begin{aligned}
&\mathbb{P}\left[\frac{(t_{k,k}-t_{k-1,k})-\mathbb{E}[t_{k,k}]}{\sqrt{Var(t_{k,k})}} \leqslant \alpha\right] \leqslant \mathbb{P}\left[\frac{(t_{k,k}-t_{k-1,k})-\mathbb{E}[t_{k,k}-t_{k-1,k}]}{\sqrt{Var(t_{k,k}-t_{k-1,k})}} \leqslant \alpha-\varepsilon\right] + \\
&+\mathbb{P}\left[\left|\frac{(t_{k,k}-t_{k-1,k})-\mathbb{E}[t_{k,k}-t_{k-1,k}]}{\sqrt{Var(t_{k,k}-t_{k-1,k})}} - \frac{(t_{k,k}-t_{k-1,k})-\mathbb{E}[t_{k,k}]}{\sqrt{Var(t_{k,k})}}\right| > \varepsilon\right] + \\
&+\mathbb{P}\left[\frac{(t_{k,k}-t_{k-1,k})-\mathbb{E}[t_{k,k}]}{\sqrt{Var(t_{k,k})}} \leqslant \alpha, \left|\frac{(t_{k,k}-t_{k-1,k})-\mathbb{E}[t_{k,k}-t_{k-1,k}]}{\sqrt{Var(t_{k,k}-t_{k-1,k})}}\right| \leqslant \varepsilon\right].
\end{aligned} \tag{6}$$

Using Claim 3 we see that the first summand of the right-hand-side of equation (6) tends to $\Phi(\alpha-\varepsilon)$, and the last summand is bounded above by $\Phi(\alpha+\varepsilon)-\Phi(\alpha-\varepsilon)$. For the middle term, notice that from the assumption $p = 1 - O\left(\frac{\ln(n)}{n}\right)$ we have $p \sim n^{-\frac{1}{n}}$, and thus

$$\lim_{n\to\infty} \frac{\mathbb{E}[t_{k-1,k}]}{\mathbb{E}[t_{k,k}]} = \lim_{n\to\infty} \frac{\binom{n}{k+1}(k+1)!p^k(1-p^{k-1})}{\binom{n}{k+1}(k+1)!p^k} = 0.$$

It follows that

$$\lim_{n\to\infty} \frac{\mathbb{E}[t_{k,k}]}{\mathbb{E}[-t_{k-1,k}+t_{k,k}]} = 1, \tag{7}$$

and so using the limit in equation 7 and in claim 1 we get that

$$\begin{aligned}
&\lim_{n\to\infty} \frac{\mathbb{E}[t_{k,k}]}{\sqrt{Var(t_{k,k})}} - \frac{\mathbb{E}[t_{k,k}-t_{k-1,k}]}{\sqrt{Var(t_{k,k}-t_{k-1,k})}} = \\
&= \frac{\mathbb{E}[t_{k,k}-t_{k-1,k}]}{\sqrt{Var(t_{k,k}-t_{k-1,k})}} \cdot \lim_{n\to\infty} \frac{\mathbb{E}[t_{k,k}]}{\sqrt{Var(t_{k,k})}} \frac{\sqrt{Var(t_{k,k}-t_{k-1,k})}}{\mathbb{E}[t_{k,k}-t_{k-1,k}]} - 1 \\
&= \frac{\mathbb{E}[t_{k,k}-t_{k-1,k}]}{\sqrt{Var(t_{k,k}-t_{k-1,k})}} \cdot 0 \\
&= 0.
\end{aligned}$$

It is thus possible to take n large enough so that

$$\left| \frac{\mathbb{E}[t_{k,k}]}{\sqrt{Var(t_{k,k})}} - \frac{\mathbb{E}[t_{k,k}-t_{k-1,k}]}{\sqrt{Var(t_{k,k}-t_{k-1,k})}} \right| < \frac{\varepsilon}{2}. \tag{8}$$

Now recall that, by Chebyshev's inequality, for a random variable X with finite non-zero variance σ^2 (and finite expected value μ) it holds that for any positive real number r

$$\mathbb{P}\left(|X-\mu| \geqslant r\sigma\right) \leqslant \frac{1}{r^2}.$$

This, together with the condition in equation (8), implies that

$$\begin{aligned}
&\mathbb{P}\left[\left| \frac{(t_{k,k}-t_{k-1,k})-\mathbb{E}[t_{k,k}-t_{k-1,k}]}{\sqrt{Var(t_{k,k}-t_{k-1,k})}} - \frac{(t_{k,k}-t_{k-1,k})-\mathbb{E}[t_{k,k}]}{\sqrt{Var(t_{k,k})}} \right| > \varepsilon \right] \\
&= \mathbb{P}\left[\left| \left(\frac{t_{k,k}-t_{k-1,k}}{\sqrt{Var(t_{k,k}-t_{k-1,k})}} - \frac{t_{k,k}-t_{k-1,k}}{\sqrt{Var(t_{k,k})}} \right) + \left(\frac{\mathbb{E}[t_{k,k}]}{\sqrt{Var(t_{k,k})}} - \frac{\mathbb{E}[t_{k,k}-t_{k-1,k}]}{\sqrt{Var(t_{k,k}-t_{k-1,k})}} \right) \right| > \varepsilon \right] \\
&\leqslant \mathbb{P}\left[(t_{k,k}-t_{k-1,k}) \left| \frac{1}{\sqrt{Var(t_{k,k})}} - \frac{1}{\sqrt{Var(t_{k,k}-t_{k-1,k})}} \right| > \frac{\varepsilon}{2} \right] \\
&\leqslant 4\varepsilon^{-2} \left(\frac{\sqrt{Var(t_{k,k}-t_{k-1,k})}}{\sqrt{Var(t_{k,k})}} - 1 \right)^2,
\end{aligned}$$

which by Claim 1 goes to zero for any fixed $\varepsilon > 0$. In conclusion, the

right side of equation (6) is bounded above by $\Phi(\alpha+\varepsilon)$ in the limit of $n \to \infty$, and therefore the central limit result for $\beta_{k,k}$ follows.

We proceed now to prove the three claims the above argument relies on.

Claim 1: $\lim_{n\to\infty} \frac{Var(t_{k,k})}{Var(t_{k,k}-t_{k-1,k})} = 1$

By definition, $Var(t_{k,k}) = \mathbb{E}[t_{k,k}^2] - \mathbb{E}[t_{k,k}]^2$.

Let χ_I be the indicator function taking value 1 if a tuple I spans a face in $EMC_{k,k}(G)$, i.e. if I induces in G a path p_I of length k. So,

$$t_{k,k} = \sum_{\substack{I\subseteq\{0,\dots,n\}\\ |I|=k+1\\ len(p_I)=k}} \chi_I$$

and

$$\mathbb{E}[t_{k,k}^2] = \sum_{I,J} \mathbb{E}[\chi_I\chi_J].$$

Therefore, $Var(t_{k,k}) = \sum_{I,J} \mathbb{E}[\chi_I\chi_J] - \mathbb{E}[t_{k,k}]^2$.

The second summand is, as we noticed in Theorem 3.1.6, $\left(\binom{n}{k+1}(k+1)!p^k\right)^2$. The first summand requires a little more work. Indicating by j the number of vertices that the tuples I and J might share we get

$$\sum_{I,J} \mathbb{E}[\chi_I\chi_J] = \binom{n}{k+1}(k+1)!\sum_{j=0}^{k+1}\binom{k+1}{j}\binom{n-(k+1)}{k+1-j}p^{2k-f(j)},$$

where $f(j)$ is the number of edges shared by the induced paths p_I and p_J. Since the number $f(j)$ of shared edges will always be at most k (the length of the path) and being $p < 1$ we can write

$$\begin{aligned}
&\binom{n}{k+1}(k+1)!\sum_{j=0}^{k+1}\binom{k+1}{j}\binom{n-(k+1)}{k+1-j}p^{2k-f(j)}\\
&\leqslant \binom{n}{k+1}(k+1)!\sum_{j=0}^{k+1}\binom{k+1}{j}\binom{n-(k+1)}{k+1-j}p^{2k-k}\\
&= \binom{n}{k+1}(k+1)!p^k\sum_{j=0}^{k+1}\binom{k+1}{j}\binom{n-(k+1)}{k+1-j}\\
&= \binom{n}{k+1}^2(k+1)!p^k,
\end{aligned}$$

where the last equality holds because of Vandermonde's identity.

So now,

$$\begin{aligned}
\mathrm{Var}(t_{k,k}) &= \sum_{I,J} \mathbb{E}[\chi_I \chi_J] - \mathbb{E}[t_{k,k}]^2 \\
&\leqslant \binom{n}{k+1}^2 (k+1)! p^k - \binom{n}{k+1}^2 ((k+1)!)^2 p^{2k} \\
&= \binom{n}{k+1}^2 ((k+1)!)^2 p^{2k} (((k+1)!)^{-1} p^{-k} - 1) \\
&\sim n^{2(k+1)} p^{2k} \left(\frac{p^{-k}}{(k+1)!} - 1 \right).
\end{aligned}$$

Thus

$$\lim_{n\to\infty} \frac{\mathrm{Var}(t_{k,k})}{n^{2(k+1)} p^{2k} (p^{-k}/(k+1)! - 1)} = 1.$$

Performing similar computations to estimate $\mathrm{Var}(t_{k-1,k})$ we obtain

$$\mathrm{Var}(t_{k-1,k}) \leqslant \binom{n}{k+1}^2 (k+1)!(1-p^{k-1})p^k - \binom{n}{k+1}^2 ((k+1)!)^2 (1-p^{k-1})^2 p^{2k},$$

and from this it follows that $\frac{\mathrm{Var}(t_{k-1,k})}{\mathrm{Var}(t_{k,k})} = o(1)$.

Now, call I and J tuples in $\mathrm{EMC}_{k-1,k}(G)$ and $\mathrm{EMC}_{k,k}(G)$ respectively. Expanding the same way as above we find

$$\begin{aligned}
\mathrm{Cov}(t_{k-1,k}, t_{k,k}) &= \sum_{I,J} \mathbb{E}[\chi_I \chi_J] - \mathbb{E}[t_{k-1,k}]\mathbb{E}[t_{k,k}] \\
&= \binom{n}{k+1} (k+1)!(1-p^{k-1}) \sum_{j=0}^{k+1} \binom{k}{j} \binom{n-(k+1)}{k+1-j} p^{2k-f(j)} - \\
&- \binom{n}{k+1} (k+1)! p^k (1-p^{k-1}) \binom{n}{k+1} (k+1)! p^k \\
&\leqslant \binom{n}{k+1}^2 (k+1)! p^k (1-p^{k-1}) - \binom{n}{k+1}^2 ((k+1)!)^2 p^{2k} (1-p^{k-1}) \\
&= \binom{n}{k+1}^2 ((k+1)!)^2 p^{2k} (1-p^{k-1}) \left(\frac{p^{-k}}{(k+1)!} - 1 \right).
\end{aligned}$$

Therefore

$$\lim_{n\to\infty} \frac{\mathrm{Cov}(t_{k-1,k}, t_{k,k})}{n^{2(k+1)} p^{2k} (1-p^{k-1})(p^{-k}/(k+1)! - 1)} = 1,$$

which implies that $\frac{\mathrm{Cov}(t_{k-1,k}, t_{k,k})}{\mathrm{Var}(t_{k,k})} = o(1)$, completing the proof of the claim.

Claims 2 and 3: $\frac{t_{k,k} - \mathbb{E}[t_{k,k}]}{\sqrt{\mathrm{Var}(t_{k,k})}} \Rightarrow N(0,1)$ as $n \to \infty$,

and $\frac{(t_{k,k} - t_{k-1,k}) - \mathbb{E}[t_{k,k} - t_{k-1,k}]}{\sqrt{\mathrm{Var}(t_{k,k} - t_{k-1,k})}} \Rightarrow N(0,1)$ as $n \to \infty$.

The proofs of claims 2 and 3 are a consequence of an abstract normal approximation theorem for dissociated random variables showed in [10].

We recall that, given a set A of n-tuples, a set of random variables $\{X_{\mathbf{a}} : \mathbf{a} = (a_1, \ldots, a_n) \in A\}$ is *dissociated* if there exist two subsets of random variables $\{X_{\mathbf{a}} : \mathbf{a} \in A'\}$ and $\{X_{\mathbf{a}} : \mathbf{a} \in A''\}$ that are independent whenever $(\cup_{\mathbf{a}\in A'}(a_1, \ldots, a_n)) \cap (\cup_{\mathbf{a}\in A''}(a_1, \ldots, a_n)) = \emptyset$.

Now let $W = \sum_{\mathbf{a}\in A} X_{\mathbf{a}}$, and for each $\mathbf{a} \in A$ define $D_{\mathbf{a}} = \{\boldsymbol{\alpha} \in A : (\alpha_1, \ldots, \alpha_n) \cap (a_1, \ldots, a_n) \neq \emptyset\}$, i.e. $D_{\mathbf{a}}$ is a dependency neighborhood for $\mathbf{a}$.

It is proved in [10, Theorem 1] that if $\mathbb{E}[X_{\mathbf{a}}] = 0$ and $\mathbb{E}[W^2] = 1$, then

$$d_1(W, Z) \leqslant K \sum_{\mathbf{a}\in A} \sum_{\boldsymbol{\alpha},\boldsymbol{\beta}\in D_{\mathbf{a}}} \left(\mathbb{E}|X_{\mathbf{a}}X_{\boldsymbol{\alpha}}X_{\boldsymbol{\beta}}| + \mathbb{E}|X_{\mathbf{a}}X_{\boldsymbol{\alpha}}|\mathbb{E}|X_{\boldsymbol{\beta}}|\right), \tag{9}$$

where $d_1(\cdot,\cdot)$ is the L_1-*Wasserstein distance*, Z is a standard normal random variable, K is a universal constant.

To show that $(t_{k,k} - t_{k-1,k})$ satisfies a central limit theorem, take the index set A to be the potential edge set of k-paths induced by simplices in $EMC_{k-1,k}(G)$ and $EMC_{k,k}(G)$. That is, the potential edge set of k-paths spanning a set of $(k+e)$ $(e \in \{0,1\})$ vertices in $G(n,p)$. We then associate to each $\mathbf{a} \in A$ the set $V_{\mathbf{a}}$ of corresponding vertices and we define

$$X_{\mathbf{a}} = \frac{1}{\sqrt{Var_{t_{k,k}}}}(\chi_{V_{\mathbf{a}}} - \mathbb{E}[\chi_{V_{\mathbf{a}}}]).$$

With this definition, the $\{X_{\mathbf{a}}\}$ are dissociated, $W = \sum_{\mathbf{a}\in A} X_{\mathbf{a}}$, $\mathbb{E}[W] = 0$ and $Var(W) = \mathbb{E}[W^2] = 1$. The proof of the claim is analogous to [67, Theorem 2.3], and for completeness we summarize here the main steps. Specifically, what is left to do is to bound the first and second half of the the sum in equation 9 and show that they both tend to zero as $n \to \infty$.

In both cases the technique used by Kahle and Meckes is to partition the dependency neighborhood $D_{\mathbf{a}}$ for each $\mathbf{a}$ into subsets of indices $D^e_{\mathbf{a}}$ whose corresponding spanning set of vertices $V^e_{\mathbf{a}}$ has size $k+e$. Then, decomposing the two sums (as in the variance estimate) by the size of the intersection of the vertex sets, it is possible to bound the first and second half of the sum with $o\left(\frac{1}{n}\right)$, and conclude that the whole quantity tends to zero as n tends to infinity. □

Remark 3.1.1. Note that we are referring to the original central limit theorem proof [67, Theorem 2.3] and not the erratum presented in [66, Theorem 1.1]. This is because Kahle and Meckes needed to slightly restrict the p-interval where the estimate for the expected value holds by setting $p = \omega\left(n^{-1/k+\delta}\right)$ and $p = o\left(n^{-1/(k+1)-\delta}\right)$ with $\delta > 0$ to avoid that the difference in means of the upper and lower bounds for the Betti numbers under consideration was too large relative to the

normalization. In our case we were able to leave the assumption that $p = 1 - O\left(\frac{\ln(n)}{n}\right)$.

3.2 EULERIAN MAGNITUDE HOMOLOGY FOR RANDOM GEOMETRIC GRAPHS ON T^2

Like Erdős-Rényi graphs, *random geometric graphs* are a fundamental model for random graphs across a variety of disciplines. First introduced by Edward N. Gilbert in [49] to model communications between radio stations, the original model involved sampling points from a Poisson point process and connecting points that fall within a fixed distance. Another common model involves selecting a fixed number of points independently and identically distributed on the space. The two models are closely connected, as observed by Penrose in [128, Section 1.7]. Because it is somewhat more commonly studied in the context of stochastic topology, here we will use the latter model.

Definition 3.2.1. Let (X, d, m) be a metric space equipped with be a Borel probability measure m. Given a positive real number $r > 0$ and positive integer n, the *Random Geometric Graph (RGG) model* $\mathrm{RGG}(n, r, (X, d, m))$ is the probability distribution on graphs with n vertices $\{x_1, \ldots, x_n\}$ given by

- selecting a collection of points $\{p_i\}_{i=1}^n$ in X according to the to the probability measure $m^{\otimes n}$ on X^n, and
- for any $i \neq j \in \{1, \ldots, n\}$, taking the edge $\{x_i, x_j\}$ to be in G if and only if $d(p_i, p_j) \leqslant r$.

We are particularly interested in bounded regions in Euclidean spaces with the uniform probability measure, as these are of central interest in many applications. However, to avoid boundary effects we will follow [152], and instead consider the flat torus of area $T^2_A = [0, \sqrt{A})^2$, with the uniform probability measure, $\mathrm{RGG}(n, r, (T^2_A, d_T, u))$, which we will abbreviate to $G(n, r, A)$. By moving to the torus, we homogenize the probability of a single edge being part of our random geometric graph, as every point x lies at the center of a euclidean ball of radius r, each of which is equally likely to contain other points. As we are again interested in studying limiting phenomena as $n \to \infty$, we will take $r = n^{-q}$ for a fixed parameter q. In particular, this will ensure that $r << \sqrt{A}$, so our euclidean balls around points do not self-intersect.

Our results in this context have similar flavor to those about ER graphs from Section 3.1, however they rely on somewhat different subgraph counting methods. However, in Theorem 3.2.6, we again find a threshold for q beyond which $\mathrm{EMH}_{k,k}(G(n, n^{-q}, A))$ vanishes

in expectation as $n \to \infty$. And, in Theorem 3.2.8 we again present an explicit formula for the asymptotic size of the Betti numbers $\mathbb{E}[\beta_{k,k}]$ as $n \to \infty$, as well as a corresponding Central Limit Theorem in Theorem 3.2.9.

We start by recalling results presented by Yu in [152] which will provide us with fundamental tools in our analysis of RGGs.

Theorem 3.2.1 (Theorem 1 [152]). *Let* $G \sim G(n, r, A)$ *and let* $x_i \neq x_j$ *be vertices of* G. *The probability of the edge* $\{x_i, x_j\}$ *appearing in* G *is* $\frac{\pi r^2}{|A|}$.

It is shown in [151, Theorem 2] that the occurrences of arbitrary pair-wise edges in RGGs are independent even if they share one end vertex. Combining this fact with the statement of Theorem 3.2.1 we obtain the following.

Corollary 3.2.2 (Corollary 3 [152]). *Let* $G \sim G(n, r, A)$ *with vertices* $V = \{x_1, \dots, x_n\}$, *and let* $\{x_{i_1}, x_{i_2}\} \neq \{x_{i_3}, x_{i_4}\} \in \binom{V}{2}$. *The probability of both* $\{x_{i_1}, x_{i_2}\}$ *and* $\{x_{i_3}, x_{i_4}\}$ *appearing as edges of* G *is* $\left(\frac{\pi r^2}{|A|}\right)^2$.

By inductively applying this corollary, we can extend this argument to any k-trail.

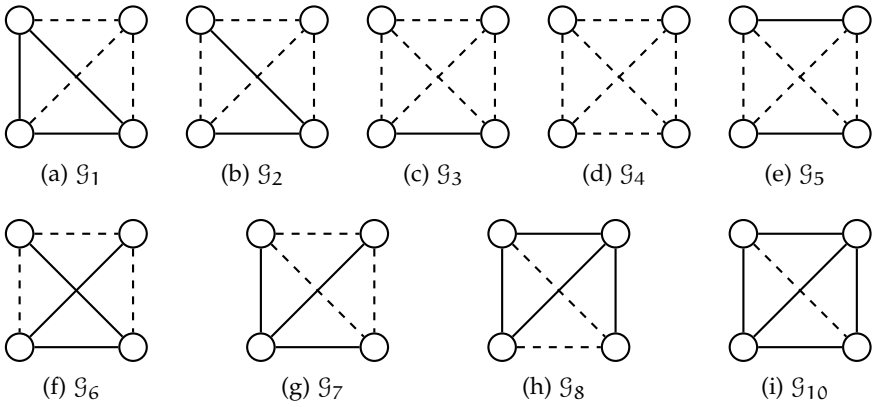
Lemma 3.2.3. *Let* $G \sim G(n, r, A)$ *with vertices* $V = \{x_1, \dots, x_n\}$, *and let* $(x_0, \dots, x_k) \in ET_k(G)$ *be a* k*-trail. The probability of the edges* $\{x_i, x_{i+1}\}_{i=0}^{k-1}$ *all appearing as edges of* G *is* $\left(\frac{\pi r^2}{|A|}\right)^k$.

Proof. Starting from the base step stated in Corollary 3.2.2, consider the eulerian $(k-1)$-trail $(x_0, \dots, x_{k-1}) \in ET_{k-1}(G)$ and assume the probability of the edges $\{x_i, x_{i+1}\}_{i=0}^{k-2}$ all appearing as edges of G is $\left(\frac{\pi r^2}{|A|}\right)^{k-1}$.

Consider now the the k-trail $(x_0, \dots, x_{k-1}, x_k) \in ET_k(G)$. Since the edge (x_{k-1}, x_k) shares only the end vertex x_{k-1} with the the $(k-1)$-trail $(x_0, \dots, x_{k-1})$, we can conclude that the probability of the edges $\{x_i, x_{i+1}\}_{i=0}^{k-1}$ all appearing as edges of G is $\left(\frac{\pi r^2}{|A|}\right)^{k-1} \cdot \frac{\pi r^2}{|A|} = \left(\frac{\pi r^2}{|A|}\right)^k$. □

To prove our vanishing thresholds, we will leverage this computation, along with families of graphs called "Y-graphs" in [152]. We require only a particular subset of this collection of families of graphs which we will call *restricted Y-classes*. In particular, our families will satisfy the conditions of [152, Corollary 18], which we will restate in our context in Corollary 3.2.4. First, we require some terminology.

Definition 3.2.2. Let $\mathcal{G} = (V, E_S, E_B)$ be a class graph, and let $V' \subseteq V$. The *full class subgraph of* $\mathcal{G}$ *on* V' is $\mathcal{G}|_{V'} = \left(V', E_S \cap \binom{V'}{2}, E_B \cap \binom{V'}{2}\right)$. The *contraction of* $\mathcal{G}$ *along* V' is the class



(a) $\mathcal{G}_1$ (b) $\mathcal{G}_2$ (c) $\mathcal{G}_3$ (d) $\mathcal{G}_4$ (e) $\mathcal{G}_5$

(f) $\mathcal{G}_6$ (g) $\mathcal{G}_7$ (h) $\mathcal{G}_8$ (i) $\mathcal{G}_{10}$

Figure 15.: Class graphs on four vertices for which $\Gamma(\mathcal{G})$ is defined to be a restricted Y-class. Numbering is preserved from [152]. Recall from Definition 2.2.2 that solid lines identify the edges in the set E_S and dashed lines denote the edges in E_B, and for every class graph $\mathcal{G}_i$ the minimal and maximal graphs (under inclusion) in $\Gamma(\mathcal{G}_i)$ are given by $\alpha(\mathcal{G}_i) = (V, E_S)$ and $\omega(\mathcal{G}_i) = (V, E_S \cup E_B)$.

graph $\mathcal{G}/V' = (V/V', E_S^{V'}, E_B^{V'})$ where, writing v' for the element of the quotient set V/V' corresponding to V', we have

$$E_S^{\mathcal{G}/V'} = \left(E_S \cap \binom{V \setminus V'}{2}\right) \cup \{\{v, v'\} : \{v, w\} \in E_S \text{ for some } w \in V'\}$$

$$E_B^{\mathcal{G}/V'} = \left(E_B \cap \binom{V \setminus V'}{2}\right) \cup \{\{v, v'\} : \{v, w\} \in E_B \text{ for some } w \in V', \text{ and, for all } w \in V', \{v, w\} \notin E_S\}.$$

Finally, we can define

Definition 3.2.3 (restricted Y-class, c.f. [152])**.** Let $\mathcal{G} = (V, E_S, E_B)$ be a class graph and $\Gamma(\mathcal{G})$ the graphs of class $\mathcal{G}$. Then $\Gamma(\mathcal{G})$ is *restricted Y-class* if $\mathcal{G}$ can be constructed according to the following two rules:

- R1 $\mathcal{G}$ is a class graph on three vertices, or one of the complete class graphs on four vertices pictured in Figure 15, or
- R2 Take $V^1, \ldots, V^k \subseteq V$ so that $V^i \cap V^j = \varnothing$ if $i \neq j$, and so that for each $i = 1, \ldots, k$, $\Gamma(\mathcal{G}|_{V^i})$ is a restricted Y-class. Write $\phi(\mathcal{G})$ for the class graph obtained by sequentially contracting $\mathcal{G}$ along each V^i. If $\alpha(\phi(\mathcal{G}))$ is a tree and $\omega(\phi(\mathcal{G}))$ is a complete graph, then $\Gamma(\mathcal{G})$ is a restricted Y-class, where $\alpha(\phi(\mathcal{G}))$ and $\omega(\phi(\mathcal{G}))$ are as in Definition 2.2.2.

Our definition of restricted Y-classes involves a subset of the rules for producing *Y-graphs* in [152], so every restricted Y-class is a Y-graph. Thus, we obtain the following corollary to [152, Corollary 18].

Corollary 3.2.4. *Suppose $\mathcal{G} = (V, E_S, E_B)$ is a class graph so that $\alpha(\mathcal{G})$ is a tree and $\Gamma(\mathcal{G})$ is a restricted Y-class. Let $G \sim G(n, r, A)$ be a geometric random graph in $\mathcal{G}$. The probability that there is some full subgraph of $H \subseteq G$ with $|V|$ vertices so that $H \cong K$ for some $K \in \Gamma(\mathcal{G})$ is $(\pi r^2/A)^{|V|-1}$.*

In order to count eulerian magnitude homology classes, we are interested in certain class graphs $\mathcal{G}$ for which graphs of the type pictured in Figure 4 appear in $\Gamma(\mathcal{G})$.

Lemma 3.2.5. *Let G be a graph and $k \geqslant 2$. Suppose $\bar{x} \in ET_{k,k}(G)$ is a k-trail in G for which $\partial_{k,k}\bar{x} = 0$. Let $H(\bar{x})$ be the graph from the proof of Lemma 3.1.2. Then there is a class graph $\mathcal{G}$ such that $G \in \mathcal{G}$, $\alpha(\mathcal{G})$ is a tree, $\Gamma(\mathcal{G})$ is a restricted Y-class, and $H \cong K$ for some $K \in \Gamma(\mathcal{G})$.*

Proof. Let $\bar{x} = (x_0, x_1, \dots x_k)$ and recall that $H(\bar{x})$ is the graph on vertices $V = \{x_0, \dots, x_k\}$ with edges $E_H = \{\{x_{i-1}, x_{i+1}\}\}_{i=0}^{k-1} \cup \{\{x_{i-1}, x_{i+1}\}_{i=1}^{k-1}$, as illustrated in Figure 4 for $k = 5$. We will proceed to build the required restricted Y-class containing $H(\bar{x})$ by (very similar) cases in k.

If $k = 2$, H is the complete graph on three vertices. This is a member of the restricted Y-class $\Gamma(\{\{x_0, x_1, x_2\}, \{\{x_0, x_1\}, \{x_0, x_2\}\}, \{\{x_1, x_2\}\}))$, whose minimal element is a tree.

If $k = 3$, H is a graph on four vertices with five edges, and so is isomorphic to a graph in each of the $\Gamma(\mathcal{G}_i)$ for $\mathcal{G}_i$ found in Figure 15. In particular, if we take $\Gamma(\mathcal{G}_7)$, the minimal element is a tree, and we have constructed the required class.

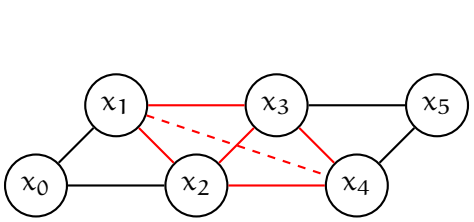
Assume now that $k \geqslant 4$. In each case, we will partition the vertices of $H(\bar{x})$ to construct the required restricted Y class using R2 of Definition 3.2.3.

If $k = 4\ell + 1$ for some $\ell > 1$, let $V^i = \{v_{1+4i}, v_{2+4i}, v_{3+4i}, v_{4+4i}\}$, $i = 0, \dots, \ell - 1$, so the V^i are disjoint and all vertices but v_0 and v_k are contained in some V^i. Each of the subgraphs $H(\bar{x})|_{V^i}$ is again a graph on four vertices with five edges. Define class graphs

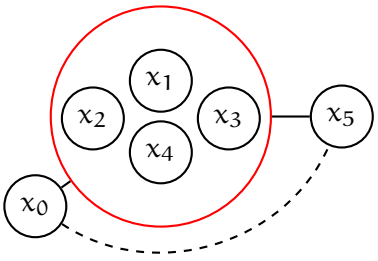
$$\mathcal{H}_i = \left(V^i, E_S^i = \binom{V^i}{2} \setminus \{\{v_{2+4i}, v_{4+4i}\}\}, E_B^i = \{\{v_{2+4i}, v_{4+4i}\}\}\right)$$

and observe that $H(\bar{x})|_{V^i} = \alpha(\mathcal{H}_i)$, as in Figure 16(a). Define $\mathcal{H}$ to be the class graph $\mathcal{H} = (V, E_S^{\mathcal{H}} = E_H, E_B^{\mathcal{H}} = \binom{V}{2} \setminus E_H)$, so again $H(\bar{x}) = \alpha(\mathcal{H})$, and $\omega(\mathcal{H})$ is the complete graph on V. Contracting $\mathcal{H}$ along each of the $\mathcal{H}_i$ sequentially, and calling the vertices corresponding to $\mathcal{H}_i$ in the contraction by h_i, we obtain the new class graph $\phi(\mathcal{H}) = (V^\phi, E_S^\phi, E_B^\phi)$ with

$$\begin{aligned} V^\phi &= \{v_0, h_1, \dots, h_\ell, v_k\}, \\ E_S^\phi &= \{\{v_0, h_1\}, \{h_\ell, v_k\}\} \cup \{\{h_i, h_{i+1}\}_{i=1}^{\ell-1}, \\ E_B^\phi &= \binom{V^\phi}{2} \setminus E_S^\phi. \end{aligned}$$



(a) In red, class graph $\mathcal{H}_1$ visualized inside $H(\bar{x})$. To expand this to the class graph $\mathcal{H}$, we add all non-pictured edges to the set $E_B^{\mathcal{H}}$.



(b) The contraction $\varphi(\mathcal{H})$. Here the large red circle is the vertex h_1, representing the quotient by the set of original vertices $\{x_1, \dots, x_4\}$. supporting $\mathcal{H}_1$.

Figure 16.: Demonstration of construction of restricted Y-class containing $H(\bar{x})$ in the proof of Lemma 3.2.5.

Here, $\alpha(\phi(\mathcal{H}))$ is a line graph, thus a tree, and $\omega(\phi(\mathcal{H}))$ is a complete graph, as required. Thus, $\mathcal{H}$ is a restricted Y-class.

The other cases follow from this same framework with slight modifications. When $k = 4\ell$, we obtain a partition of all vertices besides v_0 into sets of four vertices and obtain $\phi(\mathcal{H})$ as before, but missing vertex v_k. When $k = 4\ell + 3$, we partition the vertices $\{v_1, \dots, v_{4\ell-1}\}$ as before into sets $V^1, \dots V^{\ell-2}$ with four vertices. We then add the set $W^{\ell-1} = \{v_{4\ell+1}, v_{4\ell+2}, v_{4\ell+3}\}$, and observe that $H(\bar{x})|_{W^{\ell-1}}$ is a complete graph on three vertices, so as in the $k = 2$ case we can take the corresponding singleton restricted Y-class. The construction of $\mathcal{H}$ now proceeds as before, again omitting the vertex v_k. When $k = 4\ell + 2$, we shift the partition so we have $V^0 = \{v_0, v_1, v_2\}$, $V^i = \{v_{-1+4i}, v_{4i}, v_{1+4i}, v_{2+4i}\}, i = 1, \dots, \ell$, and proceed as before. □

We are now able to show the following.

Theorem 3.2.6. *Fix* k *and* $q > \frac{k+1}{2k}$. *Then as* $n \to \infty$,

$$\mathbb{E}[|EMH_{k,k}(G(n, n^{-q}, A)|] \to 0.$$

Proof. Let $G \sim G(n, n^{-q}, A)$, and suppose $\bar{x} = (x_0, \dots, x_k) \in ET_{k,k}(G)$ is a k-trail in G, for which $\partial_{k,k}\bar{x} = 0$. Let $\mathcal{H}$ be the resulting restricted Y-class obtained in Lemma 3.2.5. By Corollary 3.2.4, the probability of finding any subgraph $H \subseteq G|_{\{v_0,\dots,v_k\}}$ isomorphic to an element of $\Gamma(\mathcal{H})$ is $\left(\frac{\pi r^2}{|A|}\right)^{(k+1)-1} = \left(\frac{\pi r^2}{|A|}\right)^k$. Thus, this is an upper bound for the probability of observing a graph isomorphic to H_k on any subgraph on $(k+1)$ vertices. Proceeding similarly to the proof of Lemma 3.1.2 for ER graphs, call $N_{H_k}(G(n, n^{-q}, A))$ the number of subgraphs isomorphic to H_k expected to appear in a graph $G \sim G(n, n^{-q}, A)$, and a_{H_k} for the largest number of graphs isomorphic to H_k supported on a collection of $(k+1)$ vertices in G. We have

$$\begin{aligned}
N_{H_k}(G(n, n^{-q}, A)) &\leqslant \binom{n}{k+1} a_H \left(\frac{\pi r^2}{|A|}\right)^k \\
&\sim \frac{n^{k+1}}{(k+1)!} a_H \left(\frac{\pi}{|A|}\right)^k n^{-2qk} \\
&= \frac{a_H}{(k+1)!} \left(\frac{\pi}{|A|}\right)^k n^{(k+1)-2qk} \\
&\xrightarrow{n\to\infty} \begin{cases} 0, \text{ if } q > \frac{k+1}{2k} \\ \infty, \text{ if } 0 < q < \frac{k+1}{2k}. \end{cases}
\end{aligned}$$

As in the ER case, the presence of H_k is necessary to support an element in the kernel of $\partial_{k,k}$. As these vanish, so must $EMH_{k,k}(G)$.

As in the Erdős-Rényi case, the subgraph induced by a combination $\sum_{i=1}^{m}[(x_0,\dots,x_k)_i] \in EMH_{k,k}(G)$ is less likely to appear then the one induced by a single tuple. Therefore, showing that there is a restricted Y-class associated with linear combinations $\sum_{i=1}^{m} a_i \bar{x}^i$ suffices to prove that the vanishing threshold obtained from the count of single tuples $(x_0,\dots,x_k)$ generating a homology cycle holds for the whole eulerian magnitude homology group $EMH_{k,k}(G)$. To see that this is true, we will proceed similarly as in Lemma 3.2.5 by building the required restricted Y-class containing $H(\bar{x}) = \Gamma(\mathcal{H}(\bar{x}))$.

Suppose $k = 2$. Then $H(\bar{x})$ is the graph on four vertices in the restricted Y-class $\Gamma\left(\left(\{x_0,x_1,x_2,x_1'\},\{\{x_0,x_1\},\{x_1,x_2\},\{x_0,x_1'\},\{x_1',x_2\}\},\{\{x_1,x_1'\}\}\right)\right)$, which is a restricted Y-class as described in rule R1 (contained in the class $\mathcal{G}_4$), and whose minimal element is a tree.

If $k = 3$, there are two options for $H(\bar{x})$. Either it is a graph on five vertices with six edges, or it is a graph on six vertices and seven edges, as illustrated in Figure 17. In this case, we can partition the

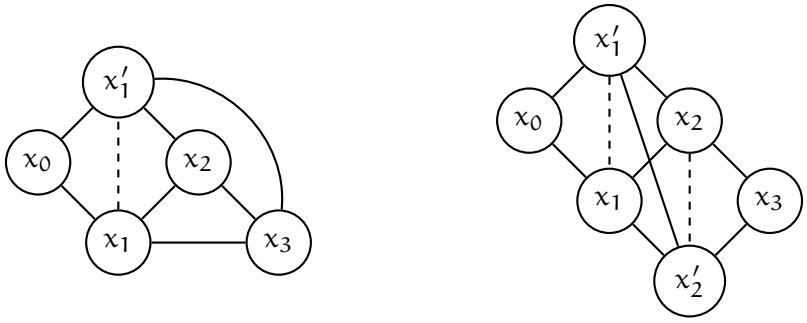


Figure 17.: (left) Subgraph $H(\bar{x})$ mandated if $H(\bar{x})$ has five vertices with six edges. (right) Subgraph $H(\bar{x})$ mandated if $H(\bar{x})$ has six vertices and seven edges.

vertices of $H(\bar{x})$ to construct the required restricted Y class using R2 of Definition 3.2.3.

In general, for $k \geqslant 4$ the subgraph $H(\bar{x})$ will look like the ones constructed in the proof of Lemma 3.1.3. Starting from the first induced 3-subgraph or 4-subgraph containing x_0, it is possible to sequentially contract (disjoint) 3-vertex Y-subgraphs of $H(\bar{x})$ or 4-vertex Y-subgraphs belonging to the $\mathcal{G}_4$ class. After this process, the class graph $\phi(\mathcal{H}(\bar{x}))$ is obtained by sequentially contracting $\mathcal{H}(\bar{x})$ is such that $\alpha(\phi(\mathcal{H}(\bar{x})))$ is a tree and $\omega(\phi(\mathcal{H}(\bar{x})))$ is a complete graph. Thus $\mathcal{H}(\bar{x})$ is a restricted Y-class.

□

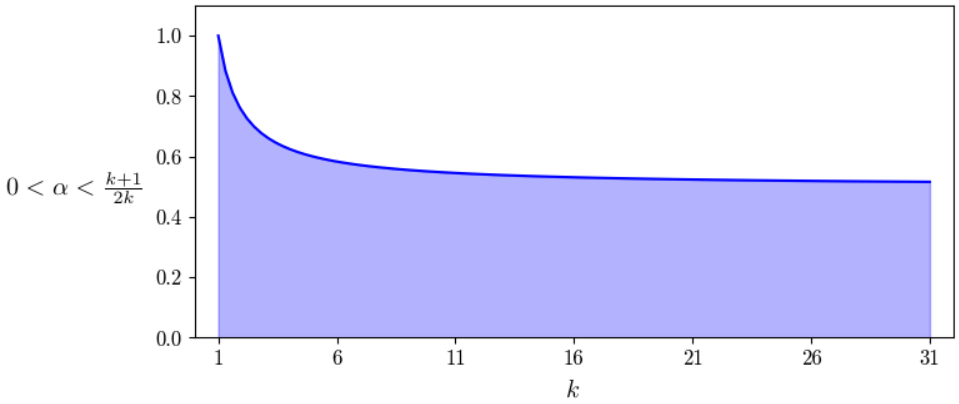


Figure 18.: The shaded region below the curve is the q vs k region for which we can have non-vanishing $EMH_{k,k}(G)$ in expectation as $n \to \infty$ for graphs $G \sim G(n, n^{-q}, A)$. By Theorems 2.3.1 and 2.3.3, in the non-shaded region asymptotically almost surely $MH_{k,k}(G) \leqslant DMH_{k,k}(G)$, and if also $k \geqslant 5$, then $MH_{k,k}(G) \cong DMH_{k,k}(G)$.

Combining this vanishing threshold with Theorem 2.3.1, we obtain the following characterization of the expected behavior of $MH_{k,k}(G(n, r, A))$ as $n \to \infty$.

Corollary 3.2.7. *Let $G = G(n, r, A)$ be a random geometric graph and take $r = n^{-q}$. If $q > \frac{k+1}{2k}$ then the magnitude homology group $MH_{k,k}(G)$ is the subgroup of the discriminant magnitude homology group $DMH_{k,k}(G)$ generated by tuples $(x_0, \ldots, x_k)$ such that the induced path always revisits the same edge (x_i, x_j).*

Using the same combinatorics, we can again provide estimates of the expected Betti numbers $\beta_{k,k}$ for geometric random graphs as $n \to \infty$, see Figure 19.

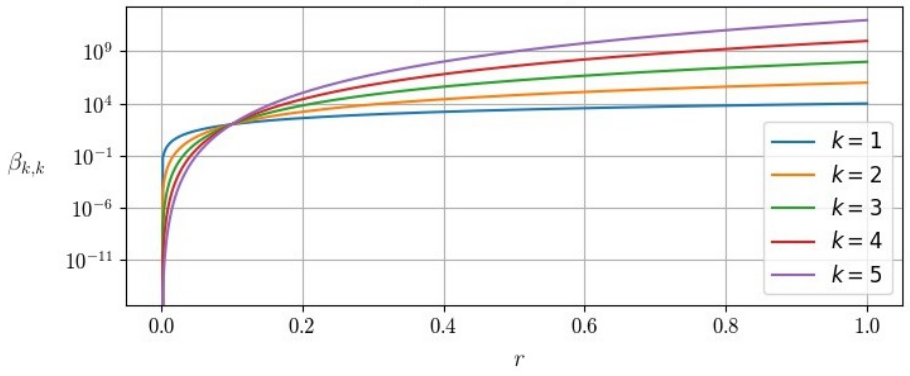


Figure 19.: The expected value for the Betti numbers $\beta_{k,k}$, $k \in [1,5]$, of a random geometric graph on a flat torus of area $T^2_A = \pi$ is plotted vertically against the radius r. In this example $n = 100$ and the y-axis is on a base 10 logarithmic scale to better visualize the large differences in values. Notice that the the curves all intersect at roughly $p = 0.1 = n^{-\alpha}$ with $n = 100$ and $\alpha = \frac{1}{2} \sim \frac{k+1}{2k}$.

Theorem 3.2.8. *Fix* $r > 0$, k *nonnegative. Write* $\beta_{k,k} = \operatorname{rank}(EMH_{k,k}(G(n,r,A)))$. *Then*

$$\lim_{n\to\infty} \frac{\mathbb{E}[\beta_{k,k}]}{n^{k+1}r^{2k}} = \left(\frac{\pi}{|A|}\right)^k.$$

Proof. This proof is essentially identical to that of Theorem 3.1.6, with a different estimate for the probability of k-trails appearing in $G \sim G(n, n^{-q}, A)$. Specifically, we have that each k-trail appears with probability $\left(\frac{\pi r^2}{|A|}\right)^k$ by Lemma 3.2.3. Since $EMH_{k,k}$ is free abelian by Lemma 1.2.1, we have

$$\lim_{n\to\infty} \frac{\mathbb{E}[t_{k,k}]}{n^{k+1}r^{2k}} = \left(\frac{\pi}{|A|}\right)^k.$$

As in the proof of Theorem 3.1.7, per Lemma 3.2.3 it holds that

$$\begin{aligned}\frac{\mathbb{E}[t_{k-1,k}]}{\mathbb{E}[t_{k,k}]} &= \frac{\binom{n}{k}k!\left(1-\left(\frac{\pi r^2}{|A|}\right)^{k-1}\right)\left(\frac{\pi r^2}{|A|}\right)^k}{\binom{n}{k+1}(k+1)!\left(\frac{\pi r^2}{|A|}\right)^k} \\ &\sim \frac{\frac{n^k}{k!}k!}{\frac{n^{k+1}}{(k+1)!}(k+1)!} = \frac{1}{n} \xrightarrow{n\to\infty} 0.\end{aligned}$$

And so it follows that

$$\lim_{n\to\infty} \frac{\mathbb{E}[t_{k,k}]}{\mathbb{E}[-t_{k-1,k} + t_{k,k}]} = 1,$$

which, together with equation (5), proves the statement. □

Finally, it is possible to prove a central limit type result for the Betti numbers of the magnitude homology groups of the first diagonal $EMH_{k,k}(G)$. The proof is identical to the result shown for the Erdős-Rényi random model in Theorem 3.1.7 by choosing $p = \left(\frac{\pi r^2}{|A|}\right)$.

Theorem 3.2.9. *Call* $\beta_{k,k} = \text{rank}(EMH_{k,k}(G(n, n^{-q}, A)))$.
Then as $n \to \infty$,

$$\frac{\beta_{k,k} - \mathbb{E}[\beta_{k,k}]}{\sqrt{Var(\beta_{k,k})}} \Rightarrow N(0,1).$$

4

HOMOTOPY TYPE OF THE EULERIAN MAGNITUDE CHAIN COMPLEX

In this chapter we investigate the regimes where an Erdös-Rényi random graph has torsion-free eulerian magnitude homology groups. To this end, we start be introducing the eulerian Asao-Izumihara complex - a CW-complex whose homology groups are isomorphic to direct summands of the graph eulerian magnitude homology group. We then proceed by producing a vanishing threshold for a shelling of eulerian Asao-Izumihara complex. This will lead to a result establishing the regimes where eulerian magnitude homology of Erdös-Rényi random graphs is torsion-free.

We start by recalling in Section 4.1 some general background about shellability. In Section 4.2 we introduce the *eulerian* Asao-Izumihara complex. We then investigate in Section 4.3 the probability regimes in which the eulerian Asao-Izumihara complex is shellable, and we conclude by producing a vanishing threshold for torsion in eulerian magnitude homology groups. Finally, in Section 4.4 we propose extensions of the current work and identify open questions that could deepen the understanding of the topic.

4.1 SHELLABLE SIMPLICIAL COMPLEXES

We recall the definition of shellable simplicial complex.

Definition 4.1.1 ([14, Definition 2.1])**.** If X is a finite simplicial complex, then a *shelling* of X is an ordering $F_1, \dots, F_t$ of the facets (maximal faces) of X such that $F_k \cap \bigcup_{i=1}^{k-1} F_i$ is a non-empty union of facets of F_k for $k \geqslant 2$. If X has a shelling, we say it is *shellable*.

In other words, we ask that the last simplex F_k meets the previous simplices along some union B_k of top-dimensional simplices of the boundary of F_k, so that X can be built stepwise by introducing the facets one at a time and attaching each new facet F_k to the complex previously built in the nicest possible fashion.

Suppose X is a non-pure simplicial complex. In this case the first facet of a shelling is always of maximal dimension. In fact, if X is shellable there is always a shelling in which the facets appear in order of decreasing dimension.

Lemma 4.1.1 ([14, Rearrangement lemma 2.6]). *Let* $F_1, F_2, \ldots, F_t$ *be a shelling of* X. *Let* $F_{i_1}, F_{i_2}, \ldots, F_{i_t}$ *be the rearrangement obtained by taking first all facets of dimension* $d = \dim X$ *in the induced order, then all facets of dimension* $d-1$ *in the induced order, and continuing this way in order of decreasing dimension. Then this rearrangement is also a shelling.*

Theorem 4.1.2 ([14, Theorem 2.9]). *Let* X *be a simplicial complex, and let* $0 \leqslant r \leqslant s \leqslant \dim X$. *Define* $X^{(r,s)} = \{\sigma \in X$ *such that* $\dim \sigma \leqslant s$ *and* $\sigma \in F$ *for some facet* F *with* $\dim F \geqslant r\}$. *If* X *is shellable, then so is* $X^{(r,s)}$ *for all* $r \leqslant s$.

Lemma 4.1.1 and Theorem 4.1.2 can be interpreted as providing a kind of "structure theorem", describing how a general shellable complex X is put together from pure shellable complexes. First there is the pure shellable complex $X^1 = X^{(d,d)}$ generated by all facets of maximal size. Then $(d-1)$-skeleton of X^1, which is also shellable, is extended by shelling steps in dimension $d-1$ to obtain $X^2 = X^{(d-1,d)}$. Then $(d-2)$-skeleton of X^2 is extended by shelling steps in dimension $(d-2)$ to obtain $X^{(d-2,d)}$, and so on until all of $X = X^{(0,d)}$ has been constructed.

A shellable simplicial complex enjoys several strong properties of a combinatorial, topological and algebraic nature. Let it suffice here to mention that it is homotopy equivalent to a wedge sum of spheres, one for each spanning simplex of corresponding dimension [44].

4.2 EULERIAN ASAO-IZUMIHARA COMPLEX

The Asao-Izumihara complex is a CW complex which is obtained as the quotient of a simplicial complex $K_\ell(a,b)$ divided by a subcomplex $K'_\ell(a,b)$, and was proposed in [6] as a geometric approach to compute magnitude homology of general graphs. Here we adapt this construction to the context of eulerian magnitude homology, providing a way of replacing the computation of the eulerian magnitude homology $EMH_{k,\ell}(G)$ by that of simplicial homology.

Let us start by recalling the Asao-Izumihara complex. Let $G = (V, E)$ be a connected graph and fix $k \geqslant 1$. For any $a, b \in V$ the set of walks with length ℓ which start with a and end with b is denoted by

$$W_\ell(a,b) := \{\bar{x} = (x_0, \ldots, x_k) \text{ walk in } G \mid x_0 = a, x_k = b, \mathrm{len}(\bar{x}) = \ell\}.$$

Definition 4.2.1 (c.f.[6, Def. 4.1]). Let G be a graph, and $a, b \in V$, $\ell \geqslant 3$.

$$\begin{aligned} K_\ell(a,b) := \{&\emptyset \neq ((x_{i_1}, i_1), \ldots, (x_{i_k}, i_k)) \subset V \times \{1, \ldots, \ell-1\} \\ &\mid (a, x_{i_1}, \ldots, x_{i_k}, b) \prec \exists (a, x_1, \ldots, x_{\ell-1}, b) \in W_\ell(a,b)\} \\ K'_\ell(a,b) := \{&((x_{i_1}, i_1), \ldots, (x_{i_k}, i_k)) \in K_\ell(a,b) \mid \mathrm{len}(a, x_{i_1}, \ldots, x_{i_k}, b) \leqslant \ell - 1\}. \end{aligned}$$

Remark 4.2.1. Following [6], we will denote $((x_{i_1}, i_1), \dots, (x_{i_k}, i_k))$ by $(x_{i_1}, \dots, x_{i_k})$ when there is no confusion.

It can also be easily seen that $K_\ell(a,b)$ is a simplicial complex and $K'_\ell(a,b)$ is a subcomplex.

Theorem 4.2.1 (c.f.[6, Thm. 4.3])**.** *Let $\ell \geqslant 3$ and $* \geqslant 0$. Then, the isomorphism*

$$(C_*(K_\ell(a,b), K'_\ell(a,b)), -\partial) \cong (MC_{*+2,\ell}(a,b), \partial)$$

of chain complexes holds.

Corollary 4.2.2 (c.f.[6, Cor. 4.4])**.** *Let $\ell \geqslant 3$.*

- *If $k \geqslant 3$, $MH_{k,\ell}(a,b) \cong H_{k-2}(K_\ell(a,b), K'_\ell(a,b))$.*
- *If $k = 2$, we also have*

$$MH_{2,\ell}(a,b) \cong \begin{cases} H_0(K_\ell(a,b), K'_\ell(a,b)) & \text{if } d(a,b) < \ell, \\ \tilde{H}_0(K_\ell(a,b)) & \text{if } d(a,b) = \ell, \end{cases}$$

  where $\tilde{H}_*$ denotes the reduced homology group.

Remark 4.2.2. Notice while both $K_{\ell-1}(a,b)$ and $K'_\ell(a,b)$ are subcomplexes of $K_\ell(a,b)$, in general $K_{\ell-1}(a,b) \subsetneq K'_\ell(a,b)$. Indeed, say v and u are two adjacent vertices, then the tuple (v,u,u) is an element of both $K_3(v,u)$ and $K'_3(v,u)$ because it is a subtuple of (v,u,v,u), but it cannot be in $K_2(v,u)$. This type of example with consecutively repeated vertices is the only one that can be constructed to show that $K_{\ell-1}(a,b)$ is a proper subset of $K'_\ell(a,b)$, and in the context of eulerian magnitude homology it cannot arise because the tuples have all different vertices. Therefore when introducing the eulerian Asao-Izumihara complex it will possible to only rely on the (eulerian versions of the) complexes $K_\ell(a,b)$ and $K_{\ell-1}(a,b)$.

Definition 4.2.2. Let $ET_{\leqslant \ell}(a,b)$ be the set of eulerian trails from a to b with length smaller than ℓ. That is, the set of all trails $(x_1, \dots, x_t) \in V^{t+1}$ such that $x_i \neq x_j$ for every $i, j \in \{1, \dots, t\}$ and

$$\operatorname{len}(a, x_1, \dots, x_t, b) \leqslant \ell.$$

The set $ET_{\leqslant \ell}(a,b)$ is clearly a simplicial complex, and the complex $ET_{\leqslant \ell-1}(a,b)$ is a subcomplex of $ET_{\leqslant \ell}(a,b)$, see Figure 20 for an illustration.

Example 4.2.1. Consider the same graph G as in example 2.1.1. Suppose we choose $(a,b) = (0,4)$ and $\ell = 4$. Then we have $ET_4(0,4) = \{(1,2,3), (1,2), (1,3), (2,3), (1), (2), (3)\}$ and $ET_3(0,4) = \{(1,2), (2,3), (1), (2), (3)\}$.

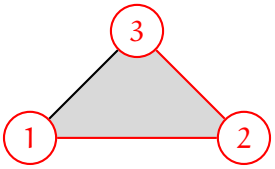


Figure 20.: The geometric realization of $ET_{\leqslant 4}(0,4)$ and $ET_{\leqslant 3}(0,4)$: $ET_{\leqslant 4}(0,4)$ is the full triangle, while $ET_{\leqslant 3}(0,4)$ is the subcomplex represented in red.

The following two results can be shown proceeding similarly to the proofs of [6, Thm. 4.3 and Cor. 4.4].

Theorem 4.2.3. *Let a, b be vertices of a graph G, and fix an integer $\ell \geqslant 3$. Then we can construct a pair of simplicial complexes $(ET_{\leqslant \ell}(a,b), ET_{\leqslant \ell-1}(a,b))$ which satisfies*

$$C_{*-2}(ET_{\leqslant \ell}(a,b), ET_{\leqslant \ell-1}(a,b)) \cong EMC_{*,\ell}(a,b).$$

Corollary 4.2.4. *Let $\ell \geqslant 3$. Then*

$$EMH_{k,\ell}(a,b) \cong H_{k-2}(ET_{\leqslant \ell}(a,b), ET_{\leqslant \ell-1}(a,b))$$

Moreover, for $k = 2$, we also have

$$EMH_{2,\ell}(a,b) \cong \begin{cases} H_0(ET_{\leqslant \ell}(a,b), ET_{\leqslant \ell-1}(a,b)) & \text{if } d(a,b) < \ell, \\ \tilde{H}_0(ET_{\leqslant \ell}(a,b)) & \text{if } d(a,b) = \ell, \end{cases}$$

where $\tilde{H}_*$ denotes the reduced homology group.

4.3 HOMOTOPY TYPE OF THE EMC COMPLEX OF ERDŐS-RÉNYI RANDOM GRAPHS

In this section we investigate the homotopy type of the eulerian magnitude chain complex of Erdős-Rényi random graphs.

Recall that the *Erdős-Rényi (ER) model* for random graphs, denoted as $G(n,p)$ and first introduced in [40], is one of the most extensively studied and utilized models for random graphs. This model represents the maximum entropy distribution for graphs with a given expected edge proportion, making it a valuable null model across a wide array of scientific and engineering fields. Consequently, the clique complexes of ER graphs have garnered significant interest within the stochastic topology community [64, 65, 67].

Definition 4.3.1. Recall that the *Erdős-Rényi (ER) model* $G(n,p) = (\Omega, P)$ is the probability space where Ω is the discrete space of all graphs on n vertices, and P is the probability measure that assigns to each graph $G \in \Omega$ with m edges probability

$$P(G) = p^m(1-p)^{\binom{n}{2}-m}.$$

We can sample an ER graph $G \sim G(n,p)$ on n vertices with parameter $p \in [0,1]$ by determining whether each of the $\binom{n}{2}$ potential edges is present via independent draws from a Bernoulli distribution with probability p. In order to study the limiting behavior of these models as $n \to \infty$, it is often useful to change variables so that p is a function of n. Here we will take $p = n^{-\alpha}$, $\alpha \in [0,\infty)$, as in [51].

In what follows, we study the homotopy type of the eulerian Asao-Izumihara complex in the context of ER random graphs. Specifically, we prove in Section 4.3.1 that, under certain assumptions, the complex $ET_{\leqslant \ell}(a,b)$ is shellable for every choice for $\ell \geqslant 3$. This will imply that $H_*(ET_{\leqslant \ell}(a,b), ET_{\leqslant \ell-1}(a,b))$ is torsion-free, and by Corollary 4.2.4 that $EMH_{*+2,\ell}(G)$ is torsion-free.

4.3.1 *Homotopy type of the eulerian Asao-Izumihara complex*

Recall from Section 4.2 that the eulerian Asao-Izumihara chain complex is the relative complex $C_*(ET_{\leqslant \ell}(a,b), ET_{\leqslant \ell-1}(a,b))$, where $ET_{\leqslant \ell}(a,b)$ is the set of eulerian tuples $(x_0 \ldots, x_k)$ such that $\operatorname{len}(a, x_0, \ldots, x_k, b) \leqslant \ell$, and $ET_{\leqslant \ell-1}(a,b)$ is defined similarly. Fix an integer $\ell \geqslant 3$.

Theorem 4.3.1. *Let $G(n, n^{-\alpha})$ be an ER graph. Suppose the facets $f_1, \ldots, f_{t-1}, f_t$ of $ET_{\leqslant \ell}(a,b)$ are ordered in decreasing dimension. Then as $n \to \infty$ $ET_{\leqslant \ell}(a,b)$ is shellable asymptotically almost surely when*

- $0 < \alpha < \prod_{i=1}^{t-1} \frac{\dim f_i + \dim f_{i+1}}{\ell + 2\dim f_{i+1} - 2}$, *if* $\dim f_1 < \frac{\ell-2}{2}$,
- $0 < \alpha < \prod_{i=1}^{k-1} \frac{\dim f_i + 3}{\ell+4} \prod_{i=k}^{t-1} \frac{\dim f_i + \dim f_{i+1}}{\ell + 2\dim f_{i+1} - 2}$, *if* $\dim f_i \geqslant \frac{\ell-2}{2}$ *for* $1 \leqslant i \leqslant k-1$ *and* $\dim f_i < \frac{\ell-2}{2}$ *for* $i \geqslant k$.

Proof. Consider the facets $f_1, \ldots, f_t$ of $ET_{\leqslant \ell}(a,b)$. Suppose they are ordered in decreasing dimension and say $\dim f_1 = d$. There are some cases we need to consider.

1. If there is a single facet f_1, then $ET_{\leqslant \ell}(a,b)$ is homotopic to a sphere S^{d-1} with $d = \dim f_1$ and we are done.
2. Say there are two different maximal facets, f_1 and f_2 and suppose they have the same dimension d.
   If f_1 and f_2 differ in one vertex, then they intersect in a $(d-1)$-face, and thus $\{f_1, f_2\}$ is a shelling.
   If f_1 and f_2 differ in two vertices u, v, then we need to distiguish the situations when u and v are adjacent and when they are not.
   a) If u and v are not adjacent, then we will have $f_1 = (a, \ldots, u, \ldots, v, \ldots, b)$ and $f_2 = (a, \ldots, u', \ldots, v', \ldots, b)$, and by construction there exists a third facet $f_3 = (a, \ldots, u', \ldots, v, \ldots, b)$ such that $\{f_1, f_3, f_2\}$ is a shelling, see Figure 21.

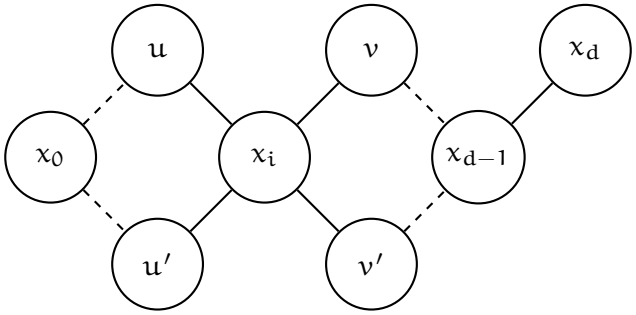


Figure 21.: In this example $f_1 = (x_0 \dots, u, x_i, v, \dots, x_d)$ and $f_2 = (x_0 \dots, u', x_i, v', \dots, x_d)$. We can define $f_3 = (x_0 \dots, u', x_i, v, \dots, x_d)$, so that $\{f_1, f_3, f_2\}$ is a shelling.

b) If u and v are adjacent, then in order to construct a facet f_3 intersecting f_1 in a $(d-1)$-face we need either the edge (u, v') or the edge (u', v) to be present (see Figure 22), and this happens with probability $p = n^{-\alpha}$.

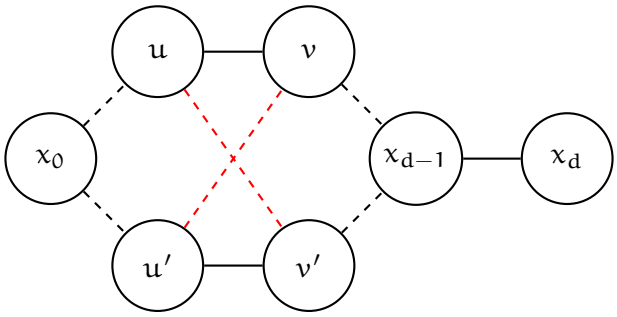


Figure 22.: In this example $f_1 = (x_0 \dots, u, v, \dots, x_d)$ and $f_2 = (x_0 \dots, u', v', \dots, x_d)$. In case one of the two dotted red edges (u, v') and (u', v) is present we can define $f_3 = (x_0 \dots, u', v, \dots, x_d)$, or $f_4 = (x_0 \dots, u, v', \dots, x_d)$, so that $\{f_1, f_3, f_2\}$ or $\{f_1, f_4, f_2\}$ is a shelling.

Now say f_1 and f_2 differ in m vertices and, indicating the facets f_1 and f_2 only by the vertices they differ in, write $f_1 = (u_1, u_2, \dots, u_m)$ and $f_2 = (u'_1, u'_2, \dots, u'_m)$. Define a partition A_i with $\bigcup_i A_i = \{u_1, \dots, u_m\}$ such that two vertices u^i_α, u^i_β belong to the same set A_i if and only if they are adjacent in G, see Figure 23. Call A'_i the corresponding partition for the vertices $(u'_1, u'_2, \dots, u'_m)$. Notice that $|A_i| = |A'_i|$ for every i. Indeed, suppose by contradiction this is not true. Then, because f_1 and f_2 have the same dimension, there exists i_1, i_2 such that $|A_{i_1}| > |A'_{i_1}|$ and $|A_{i_2}| < |A'_{i_2}|$. But then it is possible to construct a f_3 visiting vertices from A_{i_1} and A'_{i_2} thus having $\dim f_3 > \dim f_1, \dim f_2$, contradicting the fact that f_1 and f_2 are maximal facets.

Then in this case we need for every set of adjacent vertices A_i and A'_i a number $|A_i| - 1$ of edges (u^i_α, u'^i_β), $\alpha \neq \beta$, in order

to create a shelling. Indeed, we need to be able to construct a sequence of facets $f'_1, \dots, f'_m$ by changing one vertex each time so that the intersection between the j-th facet and the preceding $(j-1)$ facets is a $(d-1)$- dimensional simplex, see Figure 23. Given the fact that we also require for every set A_i a number $|A_i|+1$ of edges to connect the vertices in A_i, we obtain that the probability of all the required edges existing is

$$p^{\ell+\sum_i(|A_i|+1)+\sum_i(|A_i|-1)} = p^{\ell+2m}.$$

With $p = n^{-\alpha}$, $\alpha \in [1/2, \infty)$, we get

$$\sum_{m=2}^{d-1} \binom{n}{d+1+m} n^{-\alpha(\ell+2m)} \leqslant$$
$$(d-2)\binom{n}{d+3} n^{-\alpha(\ell+4)} \sim$$
$$(d-2)\frac{n^{d+3}}{(d+3)!} n^{-\alpha(\ell+4)} \xrightarrow{n\to\infty} \begin{cases} 0, \text{ if } \alpha > \frac{d+3}{\ell+4} \\ \infty, \text{ if } \alpha < \frac{d+3}{\ell+4}. \end{cases}$$

Notice that we assumed $\alpha \in [1/2, \infty)$ and $\frac{d+3}{\ell+4} \geqslant \frac{1}{2}$ only when $d \geqslant \frac{\ell-2}{2}$.

With $p = n^{-\alpha}$, $\alpha \in [0, 1/2)$, we get

$$\sum_{m=2}^{d-1} \binom{n}{d+1+m} n^{-\alpha(\ell+2m)} \leqslant$$
$$(d-2)\binom{n}{2d} n^{-\alpha(\ell+2d-2)} \sim$$
$$(d-2)\frac{n^{2d}}{(2d)!} n^{-\alpha(\ell+2d-2)} \xrightarrow{n\to\infty} \begin{cases} 0, \text{ if } \alpha > \frac{2d}{\ell+2d-2} \\ \infty, \text{ if } 0 < \alpha < \frac{2d}{\ell+2d-2}. \end{cases}$$

Since it holds also in this case that $\frac{2d}{\ell+2d-2} \geqslant \frac{1}{2}$ if and only if $d \geqslant \frac{\ell-2}{2}$, we can conclude that we can construct a shelling when

$$\begin{cases} 0 < \alpha < \frac{d+3}{\ell+4}, \text{ if } d \geqslant \frac{\ell-2}{2} \\ 0 < \alpha < \frac{2d}{\ell+2d-2}, \text{ if } d < \frac{\ell-2}{2}. \end{cases}$$

3. Suppose now there are two different facets, f_1 and f_2, and suppose $\dim f_2 < \dim f_1$.

   Let $\dim f_2 = d' \leqslant d-1$. Following the structure theorem for non-pure shellable complexes provided by Lemma 4.1.1 and Theorem 4.1.2, in order to produce a shelling we need to extend the (d')-skeleton of f_1 to f_2 by constructing a sequence of (d')-dimensional facets $f'_1, \dots, f'_m$ by changing one vertex each time

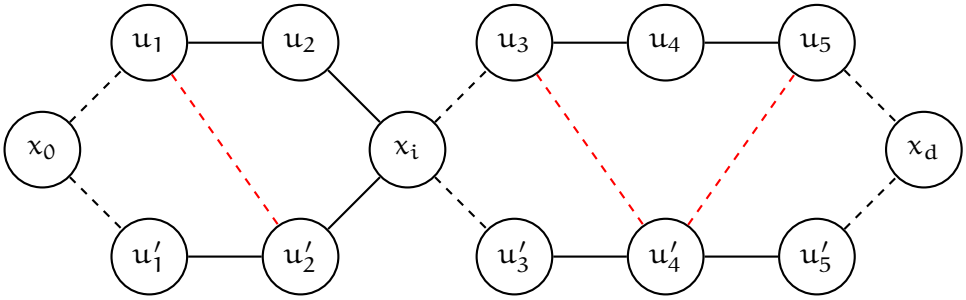


Figure 23.: In this example $A_1 = \{u_1, u_2\}$ and $A_2 = \{u_3, u_4, u_5\}$. Indicating the facets f_1 and f_2 only by the vertices they differ in we have $f_1 = (u_1, u_2, u_3, u_4, u_5)$ and $f_2 = (u'_1, u'_2, u'_3, u'_4, u'_5)$. In case all the dotted red edges are present, then we can define $f_3 = (u_1, u'_2, u_3, u_4, u_5)$, $f_4 = (u_1, u'_2, u_3, u'_4, u_5)$, $f_5 = (u_1, u'_2, u_3, u'_4, u'_5)$ and $f_6 = (u_1, u'_2, u'_3, u'_4, u'_5)$ such that $\{f_1, f_3, f_4, f_5, f_6, f_2\}$ is a shelling.

so that the intersection between the j-th facet and the preceding $(j-1)$ facets is a $(d'-1)$-dimensional simplex.

If the simplices in the (d')-skeleton of f_1 and f_2 differ in $m \leqslant d'-1$ vertices, constructing such sequence is possible if we can find $\ell + 2m$ edges joining the vertices in which f_1 and f_2 differ. This happens with probability $p^{\ell+2m}$ and therefore following the computations done in the previous point we get, for $p = n^{-\alpha}$ and $\alpha \in [1/2, \infty)$,

$$\sum_{m=2}^{d'-1} \binom{n}{d+1+m} n^{-\alpha(\ell+2m)} \leqslant$$
$$(d-3)\binom{n}{d+3} n^{-\alpha(\ell+4)} \sim$$
$$(d-3)\frac{n^{d+3}}{(d+3)!} n^{-\alpha(\ell+4)} \xrightarrow{n\to\infty} \begin{cases} 0, \text{ if } \alpha > \frac{d+3}{\ell+4} \\ \infty, \text{ if } \frac{1}{2} < \alpha < \frac{d+3}{\ell+4}. \end{cases}$$

With $p = n^{-\alpha}$, $\alpha \in [0, 1/2)$, we get

$$\sum_{m=2}^{d'-1} \binom{n}{d+1+m} n^{-\alpha(\ell+2m)} \leqslant$$
$$(d-3)\binom{n}{d+d'} n^{-\alpha(\ell+2(d'-1))} \sim$$
$$(d-3)\frac{n^{d+d'}}{(d+d')!} n^{-\alpha(\ell+2d'-2)} \xrightarrow{n\to\infty} \begin{cases} 0, \text{ if } \alpha > \frac{d+d'}{\ell+2d'-2} \\ \infty, \text{ if } 0 < \alpha < \frac{d+d'}{\ell+2d'-2}. \end{cases}$$

Again, from the fact that both inequalities $\frac{d+3}{\ell+4} \geqslant \frac{1}{2}$ and $\frac{d+d'}{\ell+2d'-2} \geqslant \frac{1}{2}$ are true if and only if $d \geqslant \frac{\ell-2}{2}$, we conclude that we can construct a shelling when

$$\begin{cases} 0 < \alpha < \frac{d+4}{\ell+4}, \text{ if } d \geqslant \frac{\ell-2}{2} \\ 0 < \alpha < \frac{d+d'}{\ell+2d'-2}, \text{ if } d < \frac{\ell-2}{2}. \end{cases}$$

4. Suppose there are t facets $f_1, \ldots, f_{t-1}, f_t$ ordered in decreasing order with $\dim f_1 = d$, then we only need to iterate the observations made in point (3).

   That is, at each step $j \in [1, \ldots t-1]$ we have a shelling when

$$\begin{cases} 0 < \alpha < \frac{\dim f_j+3}{\ell+4}, \text{ if } \dim f_j \geqslant \frac{\ell-2}{2} \\ 0 < \alpha < \frac{\dim f_j+\dim f_{j+1}}{\ell+2\dim f_{j+1}-2}, \text{ if } \dim f_j < \frac{\ell-2}{2}. \end{cases}$$

   Therefore, suppose $d = \dim f_1 < \frac{\ell-2}{2}$. Then every smaller facet f_k will be such that $\dim f_k < \frac{\ell-2}{2}$ and we will have a shelling when

$$\alpha < \prod_{i=1}^{t-1} \frac{\dim f_i + \dim f_{i+1}}{\ell + 2\dim f_{i+1} - 2}.$$

   On the other hand, if $d = \dim f_1 \geqslant \frac{\ell-2}{2}$ let f_k be the first facet in the sequence $f_1, \ldots, f_t$ such that $\dim f_k < \frac{\ell-2}{2}$. Then we will have a shelling when

$$\alpha < \prod_{i=1}^{k-1} \frac{\dim f_i + 3}{\ell + 4} \prod_{i=k}^{t-1} \frac{\dim f_i + \dim f_{i+1}}{\ell + 2\dim f_{i+1} - 2}.$$

□

Corollary 4.3.2. *Let $G(n, n^{-\alpha})$ be an ER graph. Suppose the facets $g_1, \ldots, g_{\tau-1}, g_\tau$ of $ET_{\leqslant \ell-1}(a, b)$ are ordered in decreasing dimension. Then as $n \to \infty$ $ET_{\leqslant \ell-1}(a, b)$ is shellable asymptotically almost surely when*

- $0 < \alpha < \prod_{i=1}^{\tau-1} \frac{\dim g_i+\dim g_{i+1}}{(\ell-1)+2\dim g_{i+1}-2}$, *if* $\dim g_1 < \frac{(\ell-1)-2}{2}$,
- $0 < \alpha < \prod_{i=1}^{k-1} \frac{\dim g_i+3}{(\ell-1)+4} \prod_{i=k}^{\tau-1} \frac{\dim g_i+\dim g_{i+1}}{(\ell-1)+2\dim g_{i+1}-2}$, *if* $\dim g_i \geqslant \frac{(\ell-1)-2}{2}$ *for* $1 \leqslant i \leqslant k-1$ *and* $\dim g_i < \frac{(\ell-1)-2}{2}$ *for* $i \geqslant k$.

It was shown in both [44] and [13] that a shellable simplicial complex has the homotopy type of a wedge of spheres.

Therefore using Theorem 4.3.1 and Corollary 4.3.2 we can show the following.

Theorem 4.3.3. *Let* $G(n, n^{-\alpha})$ *be an ER graph. For any pair of vertices* $(a, b) \in V^2$ *consider the eulerian Asao-Izumihara chain complex* $C_{*-2}(ET_{\leqslant \ell}(a,b), ET_{\leqslant \ell-1}(a,b)) \cong EMC_{*,\ell}(a,b)$. *Suppose the facets* $f_1, \dots, f_t$ *of* $ET_{\leqslant \ell}(a,b)$ *and* $g_1, \dots, g_\tau$ *of* $ET_{\leqslant \ell-1}(a,b)$ *are ordered in decreasing dimension. As* $n \to \infty$, *in the regimes where both* $ET_{\leqslant \ell}(a,b)$ *and* $ET_{\leqslant \ell-1}(a,b)$ *are shellable,* $EMH_{k,\ell}(a,b)$ *is torsion-free for every* k.

Proof. In the regimes where both $ET_{\leqslant \ell}(a,b)$ and $ET_{\leqslant \ell-1}(a,b)$ are shellable we can assume

$$ET_{\leqslant \ell}(a,b) \simeq \bigvee_{i=1}^{t} S_i^{n_i} \qquad \text{and} \qquad ET_{\leqslant \ell-1}(a,b) \simeq \bigvee_{j=1}^{\tau} S_j^{n_j}.$$

So, $H_k\left(ET_{\leqslant \ell}(a,b), ET_{\leqslant \ell-1}(a,b)\right) \cong H_k\left(\vee S^{n_i}, \vee S^{n_j}\right)$, and considering the long exact sequence

$$\cdots \to H_k(\vee S^{n_j}) \to H_k(\vee S^{n_i}) \to H_k\left(\vee S^{n_i}, \vee S^{n_j}\right) \to H_{k-1}(\vee S^{n_j}) \to \cdots$$

we see that

$$H_k\left(ET_{\leqslant \ell}(a,b), ET_{\leqslant \ell-1}(a,b)\right) \cong H_k\left(\vee S^{n_i}, \vee S^{n_j}\right) \cong \begin{cases} \mathbb{Z}^{m_i}, \text{ if } k = n_i, \\ \mathbb{Z}^{m_j}, \text{ if } k = n_j, \\ 0, \text{ otherwise.} \end{cases}$$

Finally, from the isomorphism theorem 4.2.3 proved in [6], we can conclude that $EMH_{k,\ell}(a,b)$ is torsion-free for every k. □

Recall that [51, Theorem 4.4] provides a vanishing threshold for the limiting expected rank of the (ℓ,ℓ)-eulerian magnitude homology in terms of the density parameter in the contexts of Erdös-Rényi random graphs.

Theorem 4.3.4 ([51, Theorems 4.4])**.** *Let* $G = G(n, n^{-\alpha})$ *be an Erdös-Rényi random graph. Fix* ℓ *and let* $\alpha > \frac{\ell+1}{2\ell-1}$. *As* $n \to \infty$, $\mathbb{E}\left[\beta_{\ell,\ell}(n, n^{-\alpha})\right] \to 0$ *asymptotically almost surely.*

Remark 4.3.1. Notice that when the smallest facet of $ET_{\leqslant \ell}(a,b)$, f_t, is such that $\dim f_t \sim \ell > \frac{\ell-2}{2}$, then $ET_{\leqslant \ell}(a,b)$ is shellable when

$$\alpha < \prod_{i=1}^{t-1} \left(\frac{\dim f_i + 3}{\ell+4}\right) \sim \prod_{i=1}^{t-1} \left(\frac{\ell+3}{\ell+4}\right) \sim 1.$$

Therefore, putting together Remark 4.3.1 with Theorems 4.3.3 and 4.3.4 we have the following.

Corollary 4.3.5. *Let* $G(n, n^{-\alpha})$ *be an Erdős-Rényi random graph. When the smallest facet* f_t *of* $ET_{\leqslant \ell}(a,b)$ *and the smallest facet* g_τ *of* $ET_{\leqslant \ell-1}(a,b)$ *are such that* $\dim f_t, \dim g_\tau \sim \ell$, *if* $EMH_{k,\ell}(G(n, n^{-\alpha}))$ *is non-vanishing it is also torsion-free.*

4.4 CONJECTURE: THE CHOICE OF ℓ

In this chapter we investigated the regimes where an Erdös-Rényi random graph G has torsion-free eulerian magnitude homology groups.

While the results presented have provided significant insights into the problem, several aspects remain unexplored, offering fertile ground for continued research.

In particular, the result stated in Corollary 4.3.5 relies on the dimension of the minimal facet f_t of $ET_{\leqslant \ell}(a,b)$ and the minimal facet g_τ of $ET_{\leqslant \ell-1}(a,b)$ being close enough to the parameter ℓ so that $\frac{\dim f_i+3}{\ell+4} \sim 1$ and $\frac{\dim g_j+3}{\ell+4} \sim 1$ for every other facet f_i, g_j.

It is thus natural to ask, how do we choose ℓ so that $\dim f_t \sim \ell$?

First, notice that the parameter ℓ cannot be too big with respect to the number of vertices n. Specifically, ℓ cannot be of the order n^2. Indeed, suppose we pick $\ell = \frac{n(n+1)}{2}$. The only way we can produce a facet f inducing a path of such length is if we have a path graph on n vertices $V = \{1,\dots,n\}$, $(a,b) = (1,\lceil n/2 \rceil)$, and we visit vertex $n-i+1$ after vertex i, $i \in \{1,\dots,\lfloor n/2 \rfloor\}$, i.e. $f = (1,n,2,n-1,\dots,\lceil n/2 \rceil)$. Then $\dim f = n < \frac{n(n+1)}{2}$. See Figure 24 for an illustration.

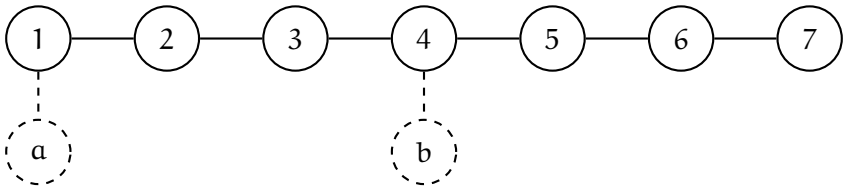


Figure 24.: In this example $(a,b) = (1,4)$ and the only facet f obtained by setting $\ell = 28$ is $(1,7,2,6,3,5,4)$, and $\dim f = 7$.

We conclude that a quadratic growth rate for ℓ with respect to n is not appropriate.

On the other hand, setting $\ell = n$ we do not encounter the same problem as before. For example, consider the path graph in Figure 24. Choosing $(a,b) = (1,4)$ and $\ell = n = 7$ we find two facets $f_1 = (1,2,3,6,5,4)$ and $f_2 = (1,2,3,5,6,4)$. Both have dimension 6 and thus $\frac{\dim f_i+3}{\ell+4} = \frac{6+3}{7+4} = \frac{9}{11} > \frac{1}{2}$.

Based on this computation, along with many other examples not displayed here, we make the following conjecture.

Conjecture 4.4.1. *Indicate the diameter of the graph G by $diam(G)$. There exists a linear function φ such that if $\ell \leqslant \varphi(diam(G))$, then $\dim f_t \sim \ell$.*

5

COMPUTING EULERIAN MAGNITUDE HOMOLOGY

In this chapter we analyze the computational cost of computing eulerian magnitude homology groups on the first diagonal and prove that it is #$W[1]$-complete. Then we develop the *first diagonal algorithm*, a breadth-first-search-based algorithm parameterized by the diameter of the graph to calculate the ranks of homology groups of interest. To do this, we leverage the close relationship between the combinatorics of the eulerian magnitude homology boundary map and the substructures appearing in the graph. We then discuss the feasibility of the presented algorithm and consider future perspectives.

5.1 EULERIAN MAGNITUDE HOMOLOGY COMPUTATIONAL COST

Relying on Definition 2.1.1 to compute eulerian magnitude homology results in a computationally unfeasible task. Indeed, consider for example the (k,k)-EMH group. The sole construction of the eulerian magnitude chain $EMC_{k,k}(G)$ requires $k!$ checks for every possible sequence of $(k+1)$ vertices, for a total of $\binom{n}{k+1}k! \sim n^{k+1}$ checks. Therefore, in the worst-case scenario of a complete graph $G = K_{n+1}$, the last computable chain $EMC_{n,n}(G)$ requires n^{n+1} computations.

One possible approach to this issue is turn it into a subgraph counting problem. Indeed, it was proven in [51] (with the language of structure graphs) that homology cycles in the eulerian magnitude chain complex can be decomposed into cycles supported on specific subgraphs H_i belonging to a family $\mathcal{H} = \{H_i\}_i$. In what follows, we recall a useful result from [51] and then we proceed by explicitly exhibiting the graphs family $\mathcal{H}$.

Lemma 5.1.1 (Lemma 2.1.1)**.** *Let* $G = (V,E)$ *be a graph and take* $k \geqslant 2$*. Fix some* $i \in [k-1]$ *and* $\bar{x} = (x_0, x_1, \ldots, x_k) \in EMC_{k,k}(G)$*. Then* $\partial^i_{k,\ell}(\bar{x}) = 0$ *if and only if* $\{x_{i-1}, x_{i+1}\} \in E$*.*

Generalizing Lemma 2.1.1 to the situation where a linear combination of tuples $\sum_i a_i \bar{x}_i$ is an eulerian magnitude homology cycle becomes difficult because of the increasingly complicated collection of isomorphism types of graphs which can support eulerian trails.

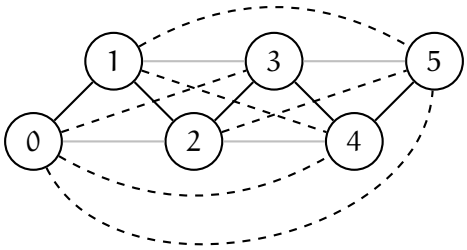


Figure 25.: Subgraph $H(\bar{x})$ induced by $[\bar{x}] = [0,1,2,3,4,5] \in EMH_{5,5}(G)$. The edges in the path $(0,1,2,3,4,5)$ are represented in black. Since the removal of each vertex causes the length of the induced path to decrease, it means all grey edges $\{x_{i-1}, x_{i+1}\}$ are contained in the induced graph. The dashed edges do not play a role in the homology computation.

A first result in this direction is the characterization of the subgraph H induced by two tuples $\bar{x}^1$ and $\bar{x}^2$ such that $\partial_{k,k}(\bar{x}^1 - \bar{x}^2) = 0$. We describe this case in the following example.

Example 5.1.1. Let $G = (V, E)$ be a graph, and let $\bar{x}^i = (x_0^i, \ldots, x_k^i), i = 1, 2$, be trails in $EMC_{k,k}(G)$. We want to construct a graph H induced by the vertex set $V(\bar{x}^1) \cup V(\bar{x}^2)$ such that the difference $(\bar{x}^1 - \bar{x}^2)$ is a non-trivial generator of $EMH_{k,k}(G)$, i.e. $\bar{x}^1$ and $\bar{x}^2$ are generators of $EMC_{k,k}(G)$ for which $\partial_{k,k}(\bar{x}^i) \neq 0, i = 1, 2$ but $\partial_{k,k}(\bar{x}^1 - \bar{x}^2) = 0$. So, say there is one landmark $\bar{x}_r^i$ such that $\partial_{k,k}(\bar{x}^i) = (-1)^r \partial_{k,k}^r(\bar{x}^i) \neq 0$.

We saw in chapter 2 that, from the definition of the differential, if $\bar{x}^1 \neq \bar{x}^2$ and $\partial_{k,k}(\bar{x}^1 - \bar{x}^2) = 0$ then both trails agree in all landmarks except one, say $x_r^1 \neq x_r^2$ for some $r \in [k-1]$, and the vertex x_r^2 cannot appear as a landmark in $\bar{x}^1$ nor vice versa.

Let $H(\bar{x}^1) = (V^{\{1\}}, E^{\{1\}})$ be the graph of Lemma 2.1.1 represented in Figure 25. The set $E^{\{1\}}$ necessarily contains all of the edges in the support of both trails except $\{x_{r-1}^1, x_r^2\}$ and $\{x_r^2, x_{r+1}^1.\}$ Further, from the proof of Lemma 2.1.1 we know that the edges already present in $H(\bar{x}^1)$ imply $\partial_{k,k}^i(\bar{x}^1) = 0, i \neq r$, and $\partial_{k,k}^i(\bar{x}^2) = 0$ for $i \neq r-1, r, r+1$, because the two trails are equal away from these vertices. However, we must introduce new edges to ensure $\partial_{k,k}^{r-1}(\bar{x}^2) = \partial_{k,k}^{r+1}(\bar{x}^2) = 0$. So, we define $H = (V_H, E_H)$, where

$$\begin{aligned} V_H =& V^{\{1\}} \cup \{x_r^2\} \\ E_H =& \Big(E^{\{1\}} \cup \{\{x_a^1, x_r^2\} : \ a \in \{r-2, r-1, r+1, r+2\} \cap [k]_0\} \Big) \\ & \setminus \{\{x_{r-1}^1, x_{r+1}^1\}\}. \end{aligned}$$

See Figure 26 for an illustration. The new vertex x_r^2 and the two new edges $\{x_{r-1}^1, x_r^2\}$ and $\{x_{r+1}^1, x_r^2\}$ support the trail $\bar{x}^2$ and imply the agreement of $\partial_{k,k}^r$ on the two generators. The other one or two new edges are diagonals in newly introduced subgraphs isomorphic to

C_4, and so are added to enforce that $\partial^{r-1}_{k,k}(\bar{x}^i) = 0$ and $\partial^{r+1}_{k,k}(\bar{x}^i) = 0$, as needed. Finally, by Lemma 2.1.1, the edge $\{x_{r-1}, x_{r+1}\}$ must be absent from G to ensure that $\partial^r_{k,k}(\bar{x}^1) = \partial^r_{k,k}(\bar{x}^2) \neq 0$. Also, from Lemma 2.1.1 all other terms in the differential of both chains are zero due to edges in $E^{\{1\}}$.

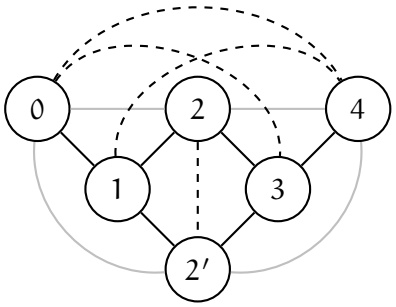


Figure 26.: Subgraph H induced by the $(4,4)$-EMH cycle $[0,1,2,3,4] - [0,1,2',3,4]$. In this case, the edge $(1,3)$ cannot be present in order to have $\partial_{4,4}(0,1,2,3,4) \neq 0$ and $\partial_{4,4}(0,1,2',3,4) \neq 0$, but $\partial_{4,4}((0,1,2,3,4) - (0,1,2',3,4)) = 0$. The black edges are in the support of the two paths, the grey edges needed to be added for the differential to vanish, and the dashed edges do not play a role in the homology computation.

It is possible to generalize the construction presented in Example 5.1.1 to the point where we remove all edges $\{x_{i-1}, x_{i+1}\}$, and in this case the graph $H = (V_H, E_H)$ would be defined as

$$\begin{aligned} V_H =& V^{\{1\}} \cup \{x'_r : r \in [k-1]\} \\ E_H =& \left(E^{\{1\}} \cup \{\{x^1_{r-1}, x'_r\}, \{x'_r, x^1_{r+1}\} : r \in [k-1]\} \cup \{x'_r, x'_{r+1} : r \in [k-2]\}\right) \\ & \setminus \{\{x^1_{r-1}, x^1_{r+1}\} : r \in [k-1]\}, \end{aligned}$$

and we see that H is homomorphic to a grid graph of dimension $(k-2) \times (k-2)$. See Figure 27 for an illustration.

Definition 5.1.1. Let $G = (V, E)$ be a graph. We call $\mathcal{H} = \{H_i\}$ the family of subgraphs of G such that H_i is induced by an (k,k)-eulerian magnitude homology cycle for every i. Notice that the minimal element of $\mathcal{H}$ is the graph in Figure 25, while the maximal element is homomorphic to the grid graph of dimension $(k-2) \times (k-2)$ as in Figure 27.

The construction above implies the following result, which we will rely on to prove the complexity of computing eulerian magnitude homology.

Theorem 5.1.2. *Let $\{\bar{x}^i\}$ be a collection of tuples in $EMC_{k,k}(G)$. Then a linear combination $\sum_i a_i \bar{x}^i$ is an eulerian magnitude homology cycle, i.e.*

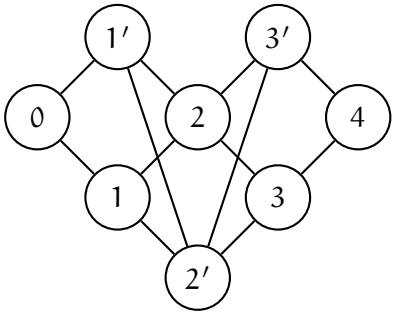


Figure 27.: Subgraph H induced by the $(4,4)$-EMH cycle $\bar{x}_1 - \bar{x}_2 + \bar{x}_3 - \bar{x}_4 + \bar{x}_5 - \bar{x}_6 + \bar{x}_7 - \bar{x}8$, where $\bar{x}_1 = (0,1,2,3,4)$, $\bar{x}_2 = (0,1',2,3,4)$, $\bar{x}_3 = (0,1',2',3,4)$, $\bar{x}_4 = (0,1',2',3',4)$, $\bar{x}_5 = (0,1',2,3',4)$, $\bar{x}_6 = (0,1,2,3',4)$, $\bar{x}_7 = (0,1,2',3',4)$, $\bar{x}_8 = (0,1,2',3,4)$. In this case, the edges $(0,2)$, $(1,3)$ and $(2,4)$ cannot be present in order to have $\partial_{4,4}\bar{x}_i \neq 0$ but $\partial_{4,4}(\sum_i (-1)^i \bar{x}_i) = 0$.

$\left[\sum_i a_i \bar{x}^i\right] \in EMH_{k,k}(G)$, *if and only if the subgraph* H *induced by the vertex set of* $\sum_i a_i \bar{x}^i$ *belongs to the family* $\mathcal{H}$ *defined in Definition 5.1.1.*

Theorem 5.1.2 (which we notice is a reformulation of [51, Theorem 3.5]) highlights the close relationship between computing the first diagonal of eulerian magnitude homology $EMH_{k,k}(G)$ and enumerating specific subgraphs H_i of G. More precisely, computing the rank of the (k,k)-eulerian magnitude homology group is equivalent to enumerating the subgraphs of G isomorphic to graphs in the family $\mathcal{H}$ defined in Definition 5.1.1. In other words, our problem is equivalent to enumerating all isomorphisms from $\mathcal{H}$ to G, and in what follows we will use the connection to prove that computing $EMH_{k,k}(G)$ is $\#W[1]$-complete.

5.1.1 $\#W[1]$-*completeness*

In this section we determined that calculating EMH is a computationally challenging problem, and in the next section we propose an algorithm to deal with this task.

Parameterized complexity theory provides a framework for a fine-grain complexity analysis of algorithmic problems that are intractable in general. The core idea of the theory is *fixed-parameter tractability*, which modifies the classical concept of polynomial time computability by allowing algorithms that run in exponential time, but only relative to a specific parameter of the problem that is typically small in practical applications. A good example of this is database query evaluation: often, the query size k is very small compared to the database size n. Therefore, an algorithm that evaluates the query in $O(2^k \cdot n)$ time may be acceptable and even efficient. In contrast, an algorithm with a $\Omega(n^k)$ evaluation time is generally not practical. Fixed-parameter tractability hinges on this principle: a parameterized

problem is considered fixed-parameter tractable if there exists a computable function f and a constant c such that the problem can be solved in $f(k) \cdot n^c$ time, where n is the input size and k is the parameter value.

To show that parameterized problems like the clique problem are not fixed-parameter tractable, a theory of *parameterized intractability* has been developed [35]. This theory has led to the creation of a complex array of parameterized complexity classes. Among these, the most significant are the $W[t]$ classes for $t \geqslant 1$, which together form the W-hierarchy. It is believed that $W[1]$ contains the class FPT (fixed-parameter tractable problems) and that the W-hierarchy is strictly ordered. Many natural parameterized problems belong to one of the classes within the W-hierarchy. For instance, the parameterized clique problem is known to be complete for the class $W[1]$.

We recall now some concepts that are needed to define the W-hierarchy.

Definition 5.1.2. A *Boolean circuit* is a directed acyclic graph with the nodes labeled as follows:

- every node of in-degree 0 is an input node
- every node with in-degree 1 is a negation node ($\neg$)
- every node with in-degree $\geqslant 2$ is either an AND-node ($\wedge$) or an OR-node ($\vee$).

Moreover, exactly one node with out-degree 0 is also labeled the output node. The *depth* of the circuit is the maximum length of a directed path from an input node to the output node. The *weft* of the circuit is the maximum number of nodes with in-degree $\geqslant 3$ on a directed path from an input node to the output node.

Given an assignment of Boolean values to the input gates, the circuit determines Boolean values at each node in the obvious way. If the value of the output node is 1 for an input assignment, we say that this assignment satisfies the circuit. The weight of an assignment is the number of inputs it sets to 1.

WEIGHTED CIRCUIT SATISFIABILITY (WCS)
Instance: A Boolean circuit C, an integer k Parameter: k Problem: Is there an assignment with weight k that satisfies C?

It is known [35] that WCS has exponential complexity.

Definition 5.1.3. The class of circuits $C_{t,d}$ contains the circuits with weft $\leqslant t$ and depth $\leqslant d$.

For any class of circuits $\mathcal{C}$, we can define the following problem.

WCS($\mathcal{C}$)
Instance: A Boolean circuit $C \in \mathcal{C}$, an integer k Parameter: k Problem: Is there an assignment with weight k that satisfies C?

Definition 5.1.4 (W and #W-hierarchy)**.** Let $t \in \{1, 2, \dots\}$. A parameterized problem Π is in the parameterized complexity class $W[t]$ if there exists a parameterized reduction from Π to $WCS[C_{t,d}]$ for some constant $d \geqslant 1$.

Similarly, we define the class #$W[t]$ for $t \geqslant 1$ as the class of all parameterized counting problems that are fixed-parameter reducible to #$WCS[C_{t,d}]$ for some constant $d \geqslant 1$.

Theorem 5.1.3 ([26, Corollary 4])**.** *Let $\mathcal{C}$ be a family of graphs. Call StrEmb($\mathcal{C}$) (strong embedding) the problem of asking whether a graph $C \in \mathcal{C}$ is isomorphic to an induced subgraph of a graph B, and call #−StrEmb($\mathcal{C}$) the counting analog. Then, if the graphs in $\mathcal{C}$ have bounded size, then StrEmb($\mathcal{C}$) $\in$ P. Otherwise, p−StrEmb($\mathcal{C}$) is complete for $W[1]$ under FPT many-one reductions. The analogue holds for the counting problem #−StrEmb($\mathcal{C}$).*

We thus have the following result.

Theorem 5.1.4. *Computing the rank of the (k, k)-eulerian magnitude homology group $EMH_{k,k}(G)$ is #$W[1]$-complete under FPT many-one reductions.*

Proof. From Theorem 5.1.2 we know that computing the rank of the group $EMH_{k,k}(G)$ is equivalent to enumerating the isomorphisms from graphs in the family $\mathcal{H}$ defined in Definition 5.1.1 to induced subgraphs of G. We noticed that the graphs in the family $\mathcal{H}$ do not have bounded size, being the maximal element homomorphic to the grid graph of dimension $(k-2) \times (k-2)$. Therefore, using Theorem 5.1.3 we can conclude that enumerating isomorphisms from graphs in $\mathcal{H}$ to induced subgraphs in G is #$W[1]$-complete under FPT many-one reductions, and thus so is the problem of computing the rank of the (k, k)-eulerian magnitude homology group. □

5.2 FIRST DIAGONAL ALGORITHM

In this section we propose a method to compute eulerian magnitude chains $EMC_{k,k}(G)$ and $EMC_{k-1,k}(G)$, and the first diagonal eulerian magnitude homology groups of a graph G, $EMH_{k,k}(G)$.

Algorithm 1 Algorithm to compute $EMC_{k,k}(G)$ and $EMC_{k-1,k}(G)$ for $2 \leqslant k \leqslant L$, where L is the diameter of G.

```
1: function FDA(graph G = (V, E), length of longest simple path L)
2:     for v_0 ∈ V do
3:         V_0 ← V \ v_0
4:         for v_1 adjacent to v_0 do
5:             V_0 ← V_0 \ v_1
6:             EMC_{1,1}(G).append(v_0, v_1)
7:             for v_2 ∈ V_0 and v_2 adjacent to v_1 do
8:                 V_0 ← V_0 \ v_2
9:                 EMC_{2,2}(G).append(v_0, v_1, v_2)
10:                if len(v_0, v_2) = 2 then EMC_{1,2}(G).append(v_0, v_2)
11:                for k in range(2, L − 1) do
12:                    for each (k + 1)-tuple in EMC_{k,k}(G) do
13:                        u ← (k + 1)-tuple[−1]
14:                        V_0 ← V_0 \ u
15:                        for w ∈ V_0 and w adjacent to u do
16:                            EMC_{k+1,k+1}(G).append((k + 1)-tuple+w)
17:                            if EMC_{k−1,k}(G) is not empty then
18:                                for k-tuple in EMC_{k−1,k}(G) do
19:                                    EMC_{k,k+1}(G).append(k-tuple+w)
20:                            if distance between entry ((k + 1)-tuple+w)[-3]
                                  and entry ((k + 1)-tuple+w)[-1]) is 2
                               then
21:                                EMC_{k,k+1}(G).append((k + 1)-tuple−u + w)
       return EMC_{k−1,k}(G), EMC_{k,k}(G) for all k ∈ [2, L]
```

The complexity of building the eulerian magnitude chains $EMC_{k,k}(G)$ highly depends on the density of the graph. Indeed, to add the i-th vertex v_i in the construction of a trail $(v_0, \ldots, v_k) \in EMC_{k,k}(G)$ we are required by definition to verify that v_i is different from the previous i vertices, and this results in $k!$ checks for each trail in $EMC_{k,k}(G)$.

In order to overcome this problem we use a breadth-first-search-based approach. Specifically, we choose a starting vertex v_0 and we build all possible k-paths starting at v_0 by performing BFS. Then we repeat the procedure n times (that is, we allow every vertex in G to be the starting point v_0). We call this algorithm First Diagonal Algorithm (FDA), and its pseudocode is shown in Algorithm 1.

For what concerns eulerian magnitude homology groups $EMH_{k,k}(G)$, our approach is to build a sparse matrix with rows and column indexed with the elements of $EMC_{k,k}(G)$ and $EMC_{k-1,k}(G)$ respectively, and compute the kernel of such matrix. The procedure is presented in Algorithm 2.

Algorithm 2 Algorithm to compute Betti numbers $\beta_{k,k}$ for a chosen k.

```
1: function EMH(EMC_{k,k}, EMC_{k-1,k})
2:     RowVec ← [ ]
3:     ColVec ← [ ]
4:     Data ← [ ]
5:     index the columns with elements of EMC_{k,k}
6:     for ChainIndex ∈ [len(EMC_{k,k}) − 1]_0 do
7:         chain ← EMC_{k,k}[ChainIndex]
8:         for VtxIdx ∈ [len(chain) − 1] do
9:             if removing any vertex does not change the length of a path
               then
10:                if the k-tuple with the vertex removed is part of EMC_{k-1,k}
                   then
11:                    RowVec.append(EMC_{k-1,k}.index(chain), VtxIdx))))
12:                    ColVec.append(ChainIndex)
13:                    Data.append(−1^{VtxIdx})
14:    ShapeMatrix ← (len(EMC_{k-1,k}),len(EMC_{k,k})))
15:    matrix ← SparseMatrix((Data, (RowVec, ColVec)), ShapeMatrix)
16:    betti ← dimension of kernel(matrix)
       return betti
```

5.2.1 *Complexity of eulerian magnitude chain computation*

The complexity of the FDA presented Algorithm 1 highly depends on the connectivity of the graph G. Indeed, apart for the first *for loop* iterating on all the n vertices, the other internal loops iterate on the neighbors of the considered vertex.

In our analysis we assume that the maximum degree of G is N_v (i.e. each vertex v has at most N_v neighbors). Proceeding with this assumption it holds that

- The second *for loop*, lines 4-21, (saving all edges (v_0, v_1) in $EMC_{1,1}(G)$) performs at most N_v iterations.
- The third *for loop*, lines 7-21, performs (at most) $(N_v)^2$ "append", $(N_v)^2$ "if" checks and $(N_v)^2$ more "append", for a total of $3(N_v)^2$ operations.
- The fourth *for loop*, lines 11-21, performs (at most) $(N_v)^3$ "$EMC_{3,3}$.append", $(N_v)^3$ "$EMC_{2,3}$.append", $(N_v)^3$ "if" checks and $(N_v)^3$ "$EMC_{2,3}$.append", for a total of (at most) $4(N_v)^3$ operations.
- In general, at the m-th step we perform at most $m \cdot (N_v)^{m-1}$ operations.

Therefore, an upper bound for the number of steps is $N_v + \sum_{i=3}^{L} i(N_v)^{i-1}$ where, we recall, L is the diameter of the graph. Thus,

$$N_v + \sum_{i=3}^{L} i(N_v)^{i-1} = N_v + \sum_{i=0}^{L} i(N_v)^{i-1} - 0 - 1 - 2N_v = (L-1)(N_v)^L - N_v,$$

and we perform at most $n \cdot ((L-1)(N_v)^L - N_v) = \mathcal{O}(n(N_v)^L)$ operations, with $2 \leqslant L \leqslant n$, where we achieve the lower bound 2 if G is a star graph and the upper bound n if G is a path graph.

5.2.2 *Discussion*

The FDA we just introduced does have exponential complexity in the worst-case scenario. Indeed, the complexity depends on the diameter L of the graph, which grows linearly with n in the case, for example, of path graphs and cycle graphs.

Nevertheless, we present in this section some examples of graphs coming from real-world situations with a much smaller diameter, making FDA an effective tool for their analysis.

The concept of the "small-world phenomenon" describes a notable trend evident in various real-world graphs: the majority of vertex pairs are linked by paths significantly shorter than the overall size of the graph. The *diameter* of an undirected graph represents the longest shortest-path distance between any pair of vertices. It serves as a familiar measure indicating the "small-world" nature of the graph. In simpler terms, it gauges how swiftly one can traverse from one side of the graph to the opposite side. The diameter is related to various processes, e.g. it is within a constant factor of the memory complexity of the depth-first search algorithm. Also, it is a natural lower bound for the mixing time of any random walk [91] and the broadcast time of the graph [57]. Mehrabian showed in 2017 [108] logarithmic upper bounds for the diameters of a variety of models, including the following well known ones: Erdös-Rényi random model [40], forest fire model [90], copying model [78], PageRank-based selection model [123], Aiello-Chung-Lu models [1], generalized linear preference model [18], directed scale-free graphs [16], Cooper-Frieze model [29], and random unordered increasing k-trees [48]. This means that in each of these models, for every pair (u, v) of vertices there exists a very short (u, v)-path connecting u and v whose length is logarithmic in the number of vertices.

A similar result was proved in the 1990s concerning the structure of the World Wide Web. Indeed, Albert, Jeong and Barabási showed in [3] that the average shortest path d between two documents (defined as the smallest number of URL links that must be followed to navigate from one document to the other) is $\langle d \rangle = 0.35 + 2.06 \log(N)$, where N is the dimension of the network, indicating that the web forms a small-world network which characterizes social or biological

systems. Also, since for a given N, d follows a gaussian distribution then $\langle d \rangle$ can be interpreted as the diameter of the web, a measure of the shortest distance between any two points in the system. Despite its huge size, estimated to be $N = 8 \times 10^8$, this result indicates that the web is a highly connected graph with an average diameter of only 19 links.

This implies, in particular, that the average diameters of many real-world models are logarithmic, which in turn signifies that in principle our algorithm FDA runs in $\mathcal{O}(n(N_v)^{\log(n)})$. Although this means that FDA's complexity is in general super-polynomial, we point out two crucial facts:

1. As noticed in Section 5.1, computing EMH using the definition potentially results in exponential computational complexity.

2. Many families of graphs have diameter L much smaller than $\log(n)$. For example: complete graphs have $L = 1$, star graphs have $L = 2$, complete bipartite graphs have $L = 2$, friendship graphs have $L = 2$, (k, λ, μ)-strongly regular graphs with $\mu > 0$ have $L = 2$, and "big enough" Cayley graph [41] have diameter $L \leqslant 3$.

Therefore, even though FDA's worst case scenario is indeed super-polynomial, it does represent both a strong improvement of the definition and a feasible computational tool.

Consider the case of a real-world graph representing a Visual Social Network, [17]. Over the past few years, there has been a tremendous surge in interest surrounding the analysis of Virtual Social Networks. This burgeoning field has attracted attention from a diverse array of disciplines including psychology, anthropology, sociology, economics, and statistics, transforming it into a truly interdisciplinary research domain.

By analyzing such networks, scientists aim to delve into the intricate web of relationships among individuals, groups, organizations, and other entities engaged in knowledge exchange across the digital landscape. To achieve this, the following two methods are generally used [38]:

1. *Socio-centric*: to examine sets of relationships between people that are regarded for analytical purposes as bounded social collectives, Figure 28.

2. *Ego-centric*: to select focal individuals (egos), and identify the nodes they are connected to, Figure 29.

In the first case, one of the most important measures characterizing the network is the density (or connectedness), which is the number of links in a network as a ratio of the total possible links. As the density grows, the diameter of the graph describing the relationships in the

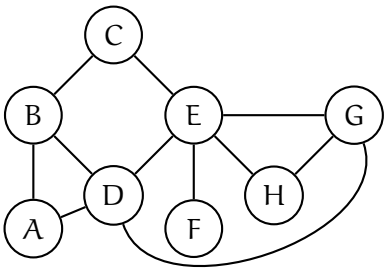


Figure 28.: Socio-centric social network with individuals A, B, C, D, E, F, G, H. The diameter of this graph is $L = 3$.

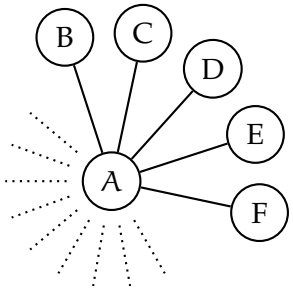


Figure 29.: Ego-centric social network with individuals A, B, C, D, E, F, G. The diameter of this graph is $L = 2$.

visual network will decrease, making the computation of eulerian magnitude homology more feasible.

If the second method is used, then we are focusing on the relations starting from one specific individual, which will produce a star graph. Therefore, also in this case eulerian magnitude homology represents a viable tool.

Part II

APPLICATIONS OF BIPERMUTIVE CELLULAR AUTOMATA TO CRYPTOGRAPHY

6

BACKGROUND

Cellular Automata (CA) are often studied as a particular type of dynamical systems, defined by shift-invariant local rules. In this thesis, however, we approach them from an algebraic perspective, emphasizing the combinatorial patterns that emerge from their short-term dynamics. In particular, we show how CA governed by bipermutive rules lead to the creation of Latin squares, a type of combinatorial design that has applications in various fields of mathematics and computer science, and we summarize the main theoretical results in the construction of CA-based Latin squares.

We start by recalling in this Chapter the background definitions and results related to cellular automata and combinatorial designs used throughout this work. We start in Section 6.1 by introducing a CA model that can be interpreted as an algebraic system. In Section 6.2 we define the main combinatorial designs of our interest, namely mutually orthogonal Latin squares and orthogonal arrays.

6.1 CELLULAR AUTOMATA

Cellular Automata (CA) are a type of parallel computing model. At its most basic level, a cellular automaton is made up of a grid of individual *cells*, each of which follows a *local rule* to decide its next state based on the states of its *neighboring* cells. The complete arrangement of all the cells at any given time is referred to as the *global state*. The CA evolves as all the cells update simultaneously in discrete time intervals, and this overall process can be described by a *global rule*, derived from the local rule governing each cell.

There are several approaches to analyzing the properties of CA. A common method is to examine a CA on an *infinite* one-dimensional grid, where each cell can take on values from a finite alphabet Σ. The set of all possible global states, represented by $\Sigma^{\mathbb{Z}}$, can be given the structure of a compact metric space using the *Cantor distance*. This leads to what is known as the *full-shift space*, which is a central topic in symbolic dynamics [93]. According to the *Curtis-Hedlund-Lyndon* theorem [58], CA are those endomorphisms $F : \Sigma^{\mathbb{Z}} \to \Sigma^{\mathbb{Z}}$ of the full-shift space that are both uniformly continuous with respect to the Cantor distance and *shift-invariant*. This means that a CA commutes

with the *shift operator* $\sigma : \Sigma^{\mathbb{Z}} \to \Sigma^{\mathbb{Z}}$, which shifts each element of a bi-infinite sequence one position to the left.

The topological approach provides a powerful framework for understanding the long-term behavior of infinite CA. However, in practical applications - such as cryptography - one is typically limited to finite arrays, since CA must be implemented on physical hardware or programmed within the constraints of software memory. This poses a challenge when updating boundary cells in the CA, as they do not have enough neighboring cells to apply the local rule. Several strategies have been proposed in the literature to address this issue. A common solution is to use a finite array of n cells with *periodic boundary conditions*, where the last (rightmost) cell is treated as adjacent to the first (leftmost) cell, forming a circular array. This ensures that each cell has a complete set of neighbors and can apply the local rule to update its state. As a result, the CA can be iterated indefinitely, although the system's long-term behavior will eventually become periodic: with n cells, the system's evolution will repeat after at most $|\Sigma|^n$ iterations of the global rule. For this reason, periodic CA are typically run for a number of steps that is polynomial in the size of the array. In some respects, the topological perspective can still be applied in this context, as periodic CA correspond to a specific subset of the full-shift space known as *spatially periodic configurations* [72].

Periodic CA have been extensively studied as finite dynamical systems for cryptographic applications such as pseudorandom number generation [148, 47, 86], block cipher design [55, 95, 142], and the construction of S-boxes [31, 11, 104]. Since periodic CA are not the focus of this thesis, we will not explore this research direction further. Readers interested in this topic can find a detailed review of related work in [106].

We point out that the topological dynamics approach can be misleading for cryptographic purposes involving CA. Many security requirements are defined through algebraic properties rather than system-theoretic ones. Moreover, in several cryptographic applications, it is unnecessary to iterate a function over multiple steps; a single application of the function is often sufficient, as it is usually combined with other types of operations. This reasoning motivates the CA model that we will explore in the remainder of this chapter.

Definition 6.1.1. Let Σ be a finite alphabet and $n, d \in \mathbb{N}$ with $n \geqslant d$. Additionally, let $f \in \Sigma^d \to \Sigma$ be a local rule of d variables. The *No Boundary Cellular Automaton (NBCA)* $F : \Sigma^n \to \Sigma^{n-d+1}$ is the function defined for all $x \in \Sigma^n$ as:

$$F(x_1, \ldots, x_n) = (f(x_1, \ldots, x_d), f(x_2, \ldots, x_{d+1}), \ldots, f(x_{n-d+1}, \ldots, x_n)).$$

In other words, a NBCA (referred to as a CA from here on) is defined as a vector function where each component function $F_i : \Sigma^n \to \Sigma$ applies the local rule f to the neighborhood formed by the i-th input

cell and the $i + d - 1$ cells to its right. Since there are $n - d + 1$ output cells in total, this ensures that the neighborhood is always complete. The parameter d is also referred to as the *diameter* of the CA.

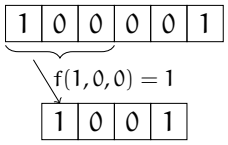
Various options are available for the choice of alphabet, with the simplest being the *binary* alphabet $\Sigma = \{0, 1\}$. More generally, we will also consider the case where $\Sigma = \mathbb{F}_q$, the *finite field* of order q, where q is a power of a prime number [92]. When $q = 2$, we have $\mathbb{F}_2 = \{0, 1\}$, which corresponds to the binary case.

Once the alphabet is chosen, there are several ways to represent the local rule of a CA. The most straightforward is the *rule table*, which specifies the next state of a cell for each possible neighborhood configuration. In the binary case, the local rule $f : \mathbb{F}_2^d \to \mathbb{F}_2$ is essentially a *Boolean function* of d variables, so the rule table is simply the *truth table* representation of f. If we impose a total order on the input vectors of $\mathbb{F}_2^d$ (for example, lexicographically), the truth table of f can be represented by its output column Ω_f, a 2^d-bit string. In CA terminology, the decimal encoding of Ω_f is also referred to as the *Wolfram code* of the local rule [147].

Another commonly used representation for the local rule of a CA is through the *de Bruijn graph* $G = (V, E)$. Here, the set of vertices $V = \Sigma^{d-1}$ consists of all possible blocks of $d - 1$ cells. Two vertices $u, v \in \Sigma^{d-1}$ are connected by an edge if and only if they *overlap* on the rightmost and leftmost $d - 1$ cells, meaning $u = u_1 t$ and $v = t v_1$, where $u_1, v_1 \in \Sigma$ and $t \in \Sigma^{d-2}$. In this situation, the string $x \in \Sigma^d$ of length d can be seen as the *fusion* of u and v, denoted as $x = u \odot v = u_1 t v_1$ [141]. A local rule $f : \Sigma^d \to \Sigma$ can then be represented as a *labeling function* $l : E \to \Sigma$ that assigns labels to the edges of the de Bruijn graph, where $l(u, v) = f(u \odot v)$ for each $(u, v) \in E$. Figure 30 illustrates an example of a binary CA $F : \mathbb{F}_2^6 \to \mathbb{F}_2^4$, which is generated by the local rule $f(x_i, x_{i+1}, x_{i+2}) = x_i \oplus x_{i+1} \oplus x_{i+2}$. The figure also shows the truth table and de Bruijn graph corresponding to the rule. The Wolfram code for this rule is 150, which corresponds to the decimal encoding of the output column 10010110, read from bottom to top.

An interesting result of the de Bruijn graph representation is that the input vector of the CA corresponds to a *path along the vertices*, combined using the fusion operator. The output vector of the CA, on the other hand, corresponds to the associated *path along the edges* of the graph. For instance, the input configuration shown in Figure 30 can be represented as the vertex sequence $10 \odot 00 \odot 00 \odot 00 \odot 01 = 100001$. The labels on the corresponding edges then produce the output configuration 1001.

One of the properties of local rules we focus on in this chapter is permutivity, as it relates to combinatorial designs. Specifically, a local rule $f : \Sigma^d \to \Sigma$ is defined as *permutive* in the i-th variable (for $i \in 1, \cdots, d$) if, when all other input coordinates are fixed to any value



x_i	x_{i+1}	x_{i+2}	$f(x)$
0	0	0	0
0	0	1	1
0	1	0	1
0	1	1	0
1	0	0	1
1	0	1	0
1	1	0	0
1	1	1	1

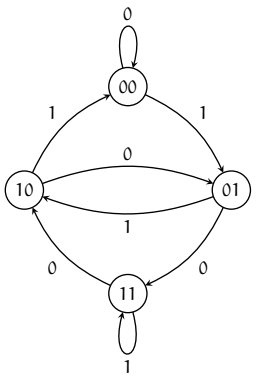


Figure 30.: Example of NBCA defined by rule 150, together with its truth table and its de Bruijn graph representations.

except the i-th one, the restricted function in the i-th variable forms a permutation over Σ. A local rule that is permutive in the first (or d-th) variable is referred to as *leftmost* (or *rightmost*) permutive. If a rule is permutive in both the first and the last variables, it is known as a *bipermutive local rule*. Rule 150 is an example of a bipermutive local rule. In fact, any bipermutive local rule over the binary alphabet is a Boolean function $f : \mathbb{F}_2^d \to \mathbb{F}_2$ of the form:

$$f(x_1, \cdots, x_d) = x_1 \oplus g(x_2, \cdots, x_{d-1}) \oplus x_d$$

for all $x = (x_1, \cdots, x_d) \in \mathbb{F}_2^d$, where $g : \mathbb{F}_2^{d-2} \to \mathbb{F}_2$ is any Boolean function of $d-2$ variables. Hence, the output of f is determined by computing the XOR of the leftmost and rightmost variables, together with a function of the $d-2$ central variables.

CA defined by permutive rules have been extensively studied in the context of topological dynamics. In [50], Gilman demonstrated that infinite one-dimensional cellular automata with (bi)permutive behavior are topologically conjugate to one-sided shifts, provided that they operate over a suitable finite state space. As a result, it can be concluded that infinite CA with (bi)permutive properties exhibit chaotic behavior as dynamical systems. Interestingly, this finding has been independently rediscovered multiple times, as seen in works such as [25, 43, 75, 134]. Furthermore, the relationship between the chaotic dynamics of permutive CA and the cryptographic aspects related to pseudorandom number generators has been explored in [47, 86, 87, 107].

When the alphabet is a finite field $\mathbb{F}_q$, it is possible to introduce the notion of linearity in CA. Specifically, a local rule $f : \mathbb{F}_q^d \to \mathbb{F}_q$ is called *linear* if it is defined as a linear combination of the neighborhood cells, or formally if there exist $a_1, \cdots, a_d \in \mathbb{F}_q$ such that:

$$f(x_1, \cdots, x_d) = a_1 x_1 + \cdots + a_d x_d$$

for all $x = (x_1, \cdots, x_d) \in \mathbb{F}_2^d$, where sum and multiplication correspond to the field operations of $\mathbb{F}_q$. A CA $F : \mathbb{F}_q^n \to \mathbb{F}_q^{n-d+1}$ equipped with a linear local rule $f : \mathbb{F}_q^d \to \mathbb{F}_q$ is a linear map from the vector space $\mathbb{F}_q^n$ to $\mathbb{F}_q^{n-d+1}$, and the global rule is defined as the matrix-vector multiplication $F(x) = M_F \cdot x^\top$ for all $x \in \mathbb{F}_q^n$, where M_F is a *transition matrix* with the following structure:

$$M_F = \begin{pmatrix} a_1 & \cdots & a_{d-1} & a_d & 0 & \cdots & \cdots & \cdots & \cdots & 0 \\ 0 & a_1 & \cdots & a_{d-1} & a_d & 0 & \cdots & \cdots & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & \cdots & \cdots & \cdots & \cdots & 0 & a_1 & \cdots & a_{d-1} & a_d \end{pmatrix} \tag{10}$$

Moreover, one can naturally define the *associated polynomial* $p_f(X) \in \mathbb{F}_q[X]$ associated to a linear rule $f : \mathbb{F}_q^d \to \mathbb{F}_q$ as :

$$p_f(X) = a_1 + a_2X + a_3X^2 + \cdots + a_dX^{d-1} \ .$$

Stated otherwise, the associated polynomial is defined by using the coefficients $a_1, \cdots, a_d$ of the local rule as the coefficients of the increasing powers of the unknown X. Notice that a linear CA is bipermutive if and only if both a_1 and a_d are not null. In this case, the associated polynomials have degree $d-1$ and a nonzero constant term.

6.2 COMBINATORIAL DESIGNS

In general terms, the central focus of the theory of *combinatorial designs* is on families of subsets drawn from a finite set, which adhere to specific balancedness conditions. Although the majority of research in this area emerged in the 20th century, some earlier results date back to Euler [42]. Today, combinatorial designs have found applications in various scientific fields, including experimental design, cryptography, and error-correcting codes.

The breadth of combinatorial design theory is vast and cannot be fully addressed in this chapter. Therefore, we concentrate on two key types of combinatorial designs that have been extensively explored in the context of CA: Latin squares and orthogonal arrays. For readers seeking a more detailed overview of combinatorial designs, standard references include [140, 28, 73, 56].

In the following, we denote by $[N] = \{1, \cdots, N\}$ the set of the first N positive integer numbers, for all $N \in \mathbb{N}$. A Latin square is formally defined as follows:

Definition 6.2.1. A *Latin square* of order $N \in \mathbb{N}$ is a $N \times N$ square matrix L with entries from $[N]$, such that $L(i,j) \neq L(i,k)$ and $L(j,i) \neq L(k,i)$ for all $i, j, k \in [N]$.

1	3	4	2
4	2	1	3
2	4	3	1
3	1	2	4

1	4	2	3
3	2	4	1
4	1	3	2
2	3	4	1

1,1	3,4	4,2	2,3
4,3	2,2	1,4	3,1
2,4	4,1	3,3	1,2
3,2	1,3	2,1	4,4

Figure 31.: Orthogonal Latin squares of order $N = 4$, and their superposition.

Intuitively, an $N \times N$ matrix is a Latin square of order N if and only if every number from 1 to N appears exactly once in each row and each column. In other words, each row and column in the matrix is a permutation of the set $[N]$.

To construct a Latin square of any order $N \in \mathbb{N}$, one can start with the elements of $[N]$ arranged in increasing order in the first row, and then generate the subsequent rows by repeatedly applying cyclic shifts. However, determining the total number of possible Latin squares remains an unsolved problem for orders $N > 11$.

The algebraic structure related to the notion of Latin square is the *quasigroup*, for which we give the following definition:

Definition 6.2.2. A *quasigroup* of order $N \in \mathbb{N}$ is a pair $(X, \circ)$ where X is a finite set of cardinality N, and $\circ$ is a binary operation over X such that for all $x, y \in X$ the two equations $x \circ z = y$ and $z \circ x = y$ admit a unique solution for all $z \in X$.

In particular, a finite algebraic structure $(X, \circ)$ is a quasigroup if and only if its *Cayley table* is a Latin square [140].

We now introduce the property of orthogonality, which takes into account two Latin squares of the same order:

Definition 6.2.3. Two Latin squares L_1 and L_2 of order N are called *orthogonal* if

$$(L_1(i_1, j_1), L_2(i_1, j_1)) \neq (L_1(i_2, j_2), L_2(i_2, j_2)) \tag{11}$$

for all distinct pairs of coordinates $(i_1, j_1), (i_2, j_2) \in [N] \times [N]$.

In other words, L_1 and L_2 are considered orthogonal if, when you overlay them on top of each other, every possible ordered pair from the Cartesian product $[N] \times [N]$ is visible. A collection of k *Mutually Orthogonal Latin Squares* (k-MOLS) consists of k Latin squares of the same order where each pair of Latin squares in the set is orthogonal to one another. For instance, Figure 31 shows a pair of orthogonal Latin squares of order 4.

Contrarily to the case of single Latin squares, it is not possible to construct a family of MOLS for any possible order. A curious result

is that the only two orders for which there are no orthogonal Latin squares are $N = 2$ and $N = 6$ [140].

A second type of combinatorial designs that we consider in this chapter are *orthogonal arrays*, closely related to orthogonal Latin squares:

Definition 6.2.4. A (N, k, s, t) *orthogonal array* (abbreviated as (N, k, s, t)–OA) is a $N \times k$ matrix with entries from a finite set X of s symbols such that, for any subset of t columns, every t–uple of symbols occurs exactly $\lambda = N/s^t$ times.

The parameter t is referred to as the *strength* of the orthogonal array (OA). When $t = 2$ and $\lambda = 1$, the resulting orthogonal array is a $N^2 \times k$ matrix where each pair of columns contains every possible ordered pair of symbols from X. In this case, the orthogonal array is simply denoted as $OA(k, \nu)$, and it corresponds to a set of $(k-2)$-MOLS. For a detailed proof of this equivalence, see [140, 56].

Families of MOLS and orthogonal arrays are utilized in cryptography and coding theory. One notable application is in constructing threshold secret sharing schemes. A *secret sharing scheme* (SSS) allows a dealer to distribute a secret value S among players $P_1, \cdots, P_n$ by giving each player a share of S. The scheme is designed so that only specific *authorized subsets* can reconstruct the secret S by combining their shares. In a (t, n) *threshold* SSS, any subset with at least t participants is authorized [133]. It can be demonstrated that a (t, n) threshold secret sharing scheme corresponds to an OA with strength t, $\lambda = 1$, and $k = n + 1$ columns [140]. Consequently, $(2, n)$ threshold schemes are equivalent to families of n-MOLS.

The connection between MOLS and OAs extends further into coding theory. As discussed in [56], *linear OAs* (OAs where the rows form a vector subspace) are related to the *duals* of linear codes. Specifically, the strength t of an OA is always strictly less than the dual distance $d^\perp$ of the associated linear code.

7

CELLULAR AUTOMATA AS ALGEBRAIC SYSTEMS

In this Chapter we describe the basic method to interpret any one-dimensional CA as an algebraic system, namely the *block transformation*, and we give a concise survey of the works that used this technique to study the dynamical properties of CA. Next, we show a related aspect of the block transformation, i.e. the preimage computation algorithm for permutive CA, and how it has been employed for cryptographic applications. Finally, we show how the algebraic structure induced by the block transformation of a bipermutive CA generates a Latin square.

7.1 THE BLOCK TRANSFORMATION

Following Definition 6.1.1, a CA is defined as a vectorial function $F : \Sigma^n \to \Sigma^{n+d-1}$ where each output coordinate is determined by the application of the local rule f on the corresponding neighborhood of diameter d. The *block transformation* allows one to re-define any CA of diameter $d > 1$ as a CA of diameter 2, where to compute the next state each cell looks only at itself and the immediate right neighbor.

The idea is to group the cells in blocks of length $d-1$, redefining the CA's alphabet as $\hat{\Sigma} = \Sigma^{d-1}$. The new local rule $f' : \hat{\Sigma}^2 \to \hat{\Sigma}$ thus maps two $(d-1)$-cell blocks to a single $(d-1)$-cell block. This can be simulated with the original local rule f by just defining f' as the NBCA $G : \Sigma^{2(d-1)} \to \Sigma^{d-1}$, where f is applied to an input vector of $2(d-1)$ cells, so that the final output vector is composed of $d-1$ cells. Figure 32 depicts an example of the block transformation on a CA of diameter $d = 5$. Assuming that the original alphabet is binary, the local rule is a Boolean function of the form $f : \mathbb{F}_2^5 \to F_2$. If we arrange the cells four by four, the new alphabet is $\hat{\Sigma} = \mathbb{F}_2^4$, i.e. 4-tuples of bits. The new local rule f' will thus have the form $f' : (\mathbb{F}_2^4)^2 \to \mathbb{F}_2^4$.

An advantage of the block transformation is that the resulting local rule f' of diameter $d = 2$ can be interpreted as a *binary operation* $* : \hat{\Sigma} \times \hat{\Sigma} \to \hat{\Sigma}$. Hence, one can consider a CA essentially as an algebraic structure $\langle \hat{\Sigma}, * \rangle$ over the state alphabet. The earlier literature that studied CA as algebraic systems focused on the relationship between the dynamic behavior of an infinite CA and the properties satisfied

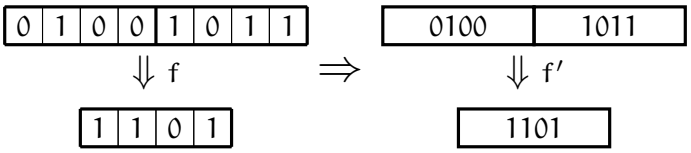


Figure 32.: Example of block transformation for a CA of diameter $d = 5$. The original local rule is defined as $f(x_1, x_2, x_3, x_4, x_5) = x_1 \oplus x_3 \oplus x_5$.

by the underlying the algebraic structure $\langle \hat{\Sigma}, * \rangle$, such as being a group or a monoid.

The block transformation was introduced by Pedersen in [127][1], where he used it to characterize infinite CA with ultimately periodic behavior to local rules that are subvarieties of groupoids. In this context, a "groupoid" is simply an algebraic structure $\langle \Sigma, * \rangle$ whose only property is closure, i.e., the result of the operation $*$ is still an element of the set to which the two operands belong to[2]. The author then investigated the relationship between different varieties of these groupoids.

Eloranta [39] considered the block transformation defined by partially permutive local rules (i.e., rules that are permutive only on subsets of the state alphabet) to explain kink-like structures exhibited by the dynamic evolution of certain CA. To the best of our knowledge, this is the first work where the block transformation of a permutive CA is related to a *quasigroup*, the algebraic structure underlying Latin squares that we discuss more in detail in Section 7.3.

The algebraic perspective on CA has been further brought forward in several works by Moore and coworkers. Moore and Drisko [115] investigated which algebraic structures give rise to efficiently predictable CA. In particular, they showed that if the binary operation defined by the block transformation satisfies one of four properties (associativity with identity, inverse property loop, anticommutativity with identity and commutativity), then the original CA local rule depends only on its leftmost and rightmost cells. Moore [112] showed that for a variety of algebraic structures such as quasigroups and Steiner systems, there exists an efficient algorithm to predict the dynamical evolution of the corresponding CA. For this reason, he termed such CA as *quasilinear*, since they obey to a law analogous to the superposition principle of linear CA. Moore and Boykett [114] considered the problem of *commuting* cellular automata, that is, under which conditions one can apply the global rules of two CA in any or-

1 Strictly speaking, the term "block transformation" actually comes from a paper by Moore and Drisko [115], although Pedersen already used it without giving it a name. The same technique was sketched even earlier by Albert and Culík [2] and Smith [137], as well as by Hedlund in his seminal work on CA [58].

2 The terminology here is not standard. In abstract algebra, groupoids are usually called *magmas*.

der and obtain the same output configuration. The main finding of that work is that linear permutive CA cannot commute with nonlinear CA. Finally, Moore [113] proved that nonlinear CA whose block transformation yields a solvable group can be decomposed into the quasidirect product of linear CA, making their prediction by parallel circuits more efficient.

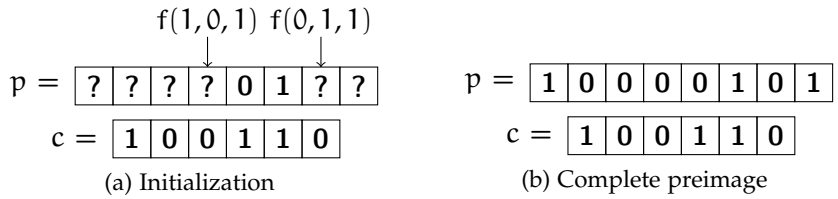
7.2 PREIMAGE COMPUTATION FOR PERMUTIVE CA

In this section we discuss a research thread that developed alongside the perspective of CA as algebraic systems, which has been specifically adopted for the design of cryptographic primitives.

Since Hedlund's work [58], it is well-known that permutive CA are surjective. The reasoning goes as follows: suppose that we have a finite configuration $y \in \Sigma^{n-d+1}$ and we want to determine one of its preimages under the action of a permutive CA $F : \Sigma^n \to \Sigma^{n-d+1}$. Without loss of generality, let us assume that the local rule $f : \Sigma^d \to \Sigma$ is rightmost permutive, and fix the leftmost $(d-1)$-cell block of a preimage $x \in \Sigma^n$ to an arbitrary value $x_{1\dots d-1} = (x_1, \cdots, x_{d-1}) \in \Sigma^{d-1}$. Then, by rightmost permutivity we know that $x_{1\dots d-1}$ determines a permutation $\pi : \Sigma \to \Sigma$ between x_d and the value of the output cell y_1. Therefore, we can compute x_d by applying the inverse permutation π^{-1} to y_1. After that, the $(d-1)$-cell block $x_{2\dots d} = (x_2, \cdots x_d)$ is fully determined, so we can obtain x_{d+1} as $\pi^{-1}(y_2)$; and so on, until the preimage is complete. A similar argument applies if the rule is leftmost permutive, by initializing the rightmost $(d-1)$-cell block of the preimage and then completing it to the left. If the rule is bipermutive the initial block of $d-1$ cells can be put in any position of the preimage, which is then completed by "expanding" the block in both directions.

The procedure above is also called the *preimage computation* or *preimage reconstruction* algorithm by some authors [94, 96]. This procedure also has an elegant description in terms of de Bruijn graphs: the basic idea is to find the path on the edges labeled by the output configuration, and then return the corresponding path on the vertices (merged together with the fusion operator) as the preimage [141, 102]. As an example, Figure 33 depicts the preimage reconstruction process for a binary NBCA $F : \mathbb{F}_2^8 \to \mathbb{F}_2^6$ equipped with the bipermutive rule 150. In the binary case, the inverse permutation π^{-1} used to determine the value of the leftmost (respectively, rightmost) input cell is simply the XOR between the output value, the rightmost (respectively, leftmost) input cell and the generating function g computed on the central $d-2$ input cells. As a matter of fact, one has:

$$y = x_1 \oplus g(x_2, \cdots, x_{d-1}) \oplus x_d \;\Leftrightarrow\; x_1 = y \oplus g(x_2, \cdots, x_{d-1}) \oplus x_d \;, \tag{12}$$



(a) Initialization (b) Complete preimage

Figure 33.: Preimage computation for $c = (1,0,0,1,1,0) \in \mathbb{F}_2^6$ using rule 150.

or equivalently,

$$y = x_1 \oplus g(x_2, \cdots, x_{d-1}) \oplus x_d \;\Leftrightarrow\; x_d = x_1 \oplus g(x_2, \cdots, x_{d-1}) \oplus y \;. \tag{13}$$

Remark that, in the no-boundary setting, the preimage computation algorithm can be iterated indefinitely, obtaining at each step a slightly larger preimage that has $d-1$ additional cells. This observation has been used by various researchers to devise cryptographic applications based on permutive CA, the first of whom being Gutowitz [55]. There, the author proposed to iterate the preimage computation algorithm with a permutive CA to design the diffusion phase of a block cipher. Oliveira et al. [118] refined Gutowitz's idea by considering bipermutive rules (there called "bi-directional toggle rules"). The underlying argument for using bipermutive rules instead of just leftmost or rightmost ones is that differences propagate in both directions, making differential cryptanalysis more difficult. The main problem of the proposal, however, remained ciphertext expansion as in Gutowitz's original design, which is a direct consequence of the iterated application of the preimage reconstruction algorithm. Later, Macêdo et al. [94] attempted to address this issue by forcing reversibility with a periodic-boundary CA.

More recently, Mariot and Leporati [96] introduced a CA-based secret sharing scheme based on the iterated preimage computation algorithm. The secret, denoted as S, is represented as a finite configuration of length m in a NBCA equipped with a bipermutive local rule. The dealer executes the preimage construction algorithm until a preimage of length $k \cdot m$ is achieved, where k is the number of players. This preimage is then divided into k blocks, each of size m, and sequentially distributed to the players. To reconstruct the secret, players must correctly order and combine their blocks, subsequently evolving the CA forward until the original secret is obtained. In this way, the resulting scheme is a (k,k)-threshold SSS, where all k players are required to pool their shares to recover the secret. A straightforward modification extends this scheme to accommodate $k+1$ players. This involves appending a *copy* of the secret to the right, resulting in a final preimage with $k+1$ blocks of size m. Both sets of players, namely $P_1, \ldots, P_k$ and $P_2, \ldots, P_{k+1}$, can then recover the secret

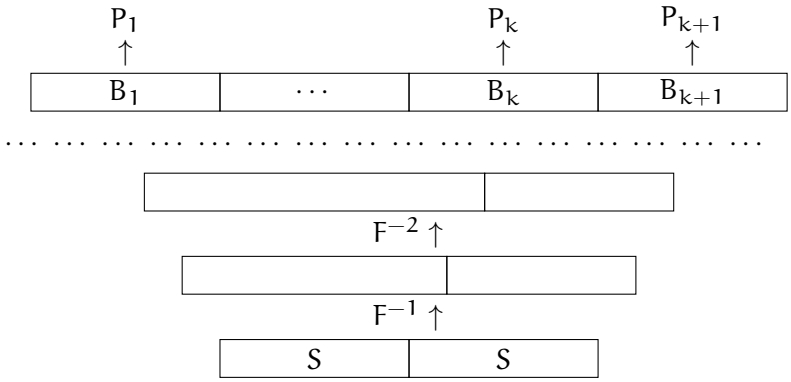


Figure 34.: Setup phase of the SSS scheme from [96] with two copies of the secret S.

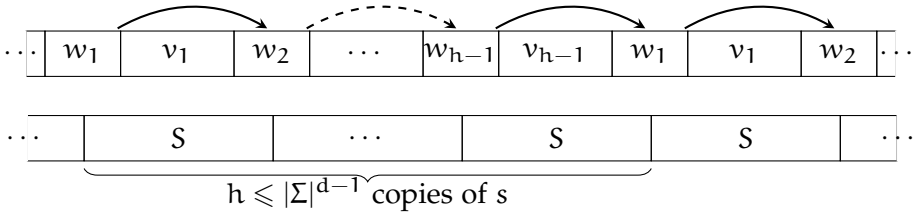


Figure 35.: Repetition of shares in the SSS proposed in [96].

using the same procedure: by combining their respective shares, the dynamic evolution of the resulting preimage pieces converges on one of the two copies of the secret. Figure 34 displays the setup phase of the SSS in this specific case.

This procedure can be generalized to establish a (k, n)-threshold scheme by concatenating k copies of the secret. Consequently, the scheme exhibits a *sequential* threshold access structure, where all minimal authorized subsets take the form $\{P_i, \ldots, P_{i+k-1}\}$. This sequentiality feature is shared by other CA-based SSS, such as the one proposed by del Rey et al. in [32]. However, a notable departure from the approach in [32] lies in the fact that in the latter shares must adhere to a *temporal adjacency* constraint, being successive configurations of a higher-order CA. On the other hand, in the scheme proposed in [96] the shares are *spatially adjacent*, since they constitute blocks of an NBCA preimage.

Furthermore, the authors of [96] remark that, as the preimage of a configuration is uniquely determined by a block of $d-1$ cells, the shares must eventually repeat after at most $|\Sigma|^{d-1}$ juxtaposed copies of the secret. Hence, the resulting access structure of the scheme is both sequential and *cyclic*. Figure 35 represents this effect for a single iteration of the preimage computation algorithm. The $d-1$-cell blocks that determine the subsequent block of m cells are denoted as $w_1, \ldots, w_{h-1}$.

Consequently, determining the maximum number of players allowed in the CA-based SSS of [96] is equivalent to studying the periods of preimages of *spatially periodic configurations* (SPC) in bipermutive CA. Indeed, a well-known fact for infinite CA is that any preimage $x \in F^{-1}(y)$ of a SPC $y \in \Sigma^{\mathbb{Z}}$ is also a SPC, whose period is a multiple of the period of y [58]. Following the SSS-inspired motivation, Mariot et al. [102] investigated more in detail the periods of preimages in SPC, first by providing some upper bounds for the case of general bipermutive CA. Then, restricting the attention to the subclass of linear bipermutive CA over a finite field $\mathbb{F}_q$, the authors provided an exact characterization of the periods of preimages, leveraging on the theory of *Linear Recurring Sequences* (LRS). Indeed, the preimage reconstruction algorithm for a linear bipermutive CA can be synthesized by a concatenation of *Linear Feedback Shift Registers* (LFSR), and thus the period of the preimage can be deduced by the minimal polynomials of the two LFSRs.

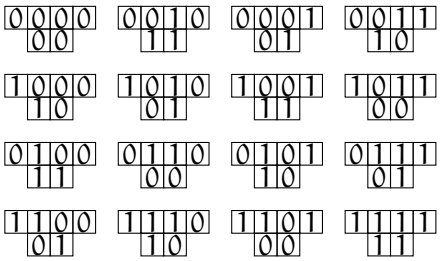
7.3 LATIN SQUARES FROM BIPERMUTIVE CA

We conclude this chapter by showing how the two research tracks presented so far—CA as algebraic systems and preimage computation for permutive CA— are actually related to one another, and provide the basis for the combinatorial designs perspective that we develop in the remainder of the chapter.

From now on, we focus on the case of an NBCA $F : \Sigma^{2(d-1)} \to \Sigma^{d-1}$ defined by a bipermutive local rule $f : \Sigma^d \to \Sigma$ of diameter d. In this way, we can directly consider the CA global rule as an algebraic structure $\langle \hat{\Sigma}, * \rangle$, where $\hat{\Sigma} = \Sigma^{d-1}$ and $* \equiv F$. In general, the *Cayley table* of a finite algebraic structure $\langle S, * \rangle$ with $N = |S|$ is the $N \times N$ matrix C_* where the rows and columns are indexed by the elements of S, and $C_*(x, y) = x * y$ for all pairs of row and column coordinates $x, y \in S$. For the CA case $\langle \hat{\Sigma}, F \rangle$, denoting $N = |\hat{\Sigma}| = |\Sigma|^{d-1}$, the corresponding Cayley table C_F is defined by using the leftmost $(d-1)$-cell block as the row coordinate, the rightmost $(d-1)$-cell block as the column coordinate, and the output configuration of the CA (itself a $(d-1)$-cell block) as the entry to be inserted at those coordinates.

More formally, let us define a total order $\leqslant$ over Σ^{d-1}, and let $\phi : \Sigma^{d-1} \to [N]$ be a monotone bijective mapping between Σ^{d-1} and $[N] = \{1, \cdots, q^{d-1}\}$, where the total order on $[N]$ is inherited from the usual order of natural numbers. We denote by $\psi = \phi^{-1}$, i.e. the inverse of ϕ. We can now give the formal definition of Cayley table associated to a CA:

Definition 7.3.1. Let Σ be an alphabet of q symbols and $d \in \mathbb{N}$, with $N = q^{d-1}$. Further, suppose that $F : \Sigma^{2(d-1)} \to \Sigma^{d-1}$ is a CA defined



1	4	3	2
2	3	4	1
4	1	2	3
3	2	1	4

Figure 36.: Latin square of order 4 generated by rule 150.

by the local rule $f : \Sigma^d \to \Sigma$. The *Cayley table* associated to F is the $N \times N$ matrix C_F with entries from $[N]$ such that

$$C_F(i,j) = \phi(F(\psi(i) \cdot \psi(j))) \ , \tag{14}$$

for all $1 \leqslant i,j \leqslant N$, where $\psi(i) \cdot \psi(j) \in \Sigma^{2(d-1)}$ denotes the *concatenation* of $\psi(i), \psi(j) \in \Sigma^{d-1}$.

The next result, originally proved by Eloranta in [39] and independently re-discovered by Mariot et al. [100], shows the connection between bipermutive CA, Latin squares and quasigroups:

Lemma 7.3.1. *The Cayley table C_F associated to a CA $F : \Sigma^{2(d-1)} \to \Sigma^{d-1}$ defined by a bipermutive local rule $f : \Sigma^d \to \Sigma$ is a Latin square of order $N = q^{d-1}$, where $q = |\Sigma|$. Equivalently, the algebraic structure $\langle \hat{\Sigma}, F \rangle$ is a quasigroup of order N.*

The proof of this lemma is a straightforward application of the preimage computation algorithm in both directions: by fixing the leftmost block of the CA, there exists a permutation between the rightmost block and the output block. This means that by fixing any row of the Cayley table, we see all numbers from 1 to N exactly once. A symmetrical argument holds by fixing the rightmost block of the CA (which means fixing a column of the table).

Figure 36 illustrates the construction of the Cayley table associated to the CA $F : \mathbb{F}_2^4 \to \mathbb{F}_2^2$ defined by rule 150. The mapping ϕ here is $\phi(00) \mapsto 1$, $\phi(10) \mapsto 2$, $\phi(01) \mapsto 3$ and $\phi(11) \mapsto 4$. Indeed, one can see that each number from 1 to 4 occurs exactly once in each row and column of the table, making it a Latin square of order 4.

8

MUTUALLY ORTHOGONAL CELLULAR AUTOMATA

This Chapter focuses on families of *Mutually Orthogonal Cellular Automata* (MOCA), which are families of bipermutive CA that form *Mutually Orthogonal Latin Squares* (MOLS). The chapter covers in particular both the well-developed theory of linear MOCA, and the few results discovered so far for the nonlinear case.

8.1 THE LINEAR CASE AND THE CONNECTION WITH COPRIME POLYNOMIALS

There is an extensive literature exploring how linear cellular automata can be used to generate complex combinatorial structures such as mutually orthogonal Latin squares (MOLS), and the key idea behind these works is that the simple, local, and deterministic nature of CA can be leveraged to generate mathematically rich objects with desirable properties. We review here some of the major works in this field.

In [100, 105] the authors undertake an investigation of combinatorial designs engendered by cellular automata, focusing in particular on orthogonal Latin squares and orthogonal arrays. First they show that any bipermutive cellular automaton of diameter d and length $2(d-1)$ induces a Latin square of order $N = q^{d-1}$, where q is the size of the alphabet. Then, using a characterization based on Sylvester matrices, they prove that two linear CA induce a pair of orthogonal Latin squares if and only if the polynomials associated to their local rules are relatively prime.

We start by introducing some notation for linear CAs over the finite field $\mathbb{F}_q$.

A CA $F : \mathbb{F}_q^n \to \mathbb{F}_q^{n-d+1}$ of diameter d is called *linear* if its local rule $f : \mathbb{F}_q^d \to \mathbb{F}_q$ is a linear combination of the cells in the neighborhood, i.e. there exist $a_0, \ldots, a_{d-1} \in \mathbb{F}_q$ such that

$$f(x_0, \ldots, x_{d-1}) = a_0 x_0 + a_1 x_1 + \cdots + a_{d-1} x_{d-1},$$

for all $x \in \mathbb{F}_q^d$, where sum and product are the field operations of $\mathbb{F}_q$. Notice that for $q = 2$, these respectively correspond to the logical

operations XOR ($\oplus$) and AND ($\wedge$). A linear CA can be seen as a linear transformation over $\mathbb{F}_q$-vector spaces described by the following $n \times (n - d + 1)$ transition matrix:

$$M_F = \begin{pmatrix} a_0 & \cdots & a_{d-1} & 0 & \cdots & \cdots & \cdots & \cdots & 0 \\ 0 & a_0 & \cdots & a_{d-1} & 0 & \cdots & \cdots & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & \cdots & \cdots & \cdots & \cdots & 0 & a_0 & \cdots & a_{d-1} \end{pmatrix} \quad (15)$$

In particular, the CA global rule is defined as the matrix-vector multiplication $F(x) = M_F \cdot x^\top$ for all $x \in \mathbb{F}_q^n$. As remarked in [97], the matrix M_F in Equation 15 is the generator matrix of a cyclic code. Hence, one can naturally define the polynomial $p_f(X) \in \mathbb{F}_q[X]$ associated to a linear CA F as the generator polynomial of degree $n \leqslant d - 1$ of the corresponding cyclic code:

$$p_f(X) = a_0 + a_1 X + \cdots + a_{d-1} X^{d-1} \in \mathbb{F}_q[X]. \quad (16)$$

It is easy to see that a linear CA is bipermutive if and only if both a_0 and a_{d-1} are not null [100].

Now, given an alphabet A of q symbols, suppose that a total order $\leqslant$ is defined over the set of $(d-1)$-uples of A^{d-1}, and that $\phi : A^{d-1} \to [N]$ is a monotone one-to-one mapping between A^{d-1} and $[N] = \{1, \cdots, q^{d-1}\}$, where the order relation on $[N]$ is the usual order on natural numbers. We denote by $\psi = \phi^{-1}$ the inverse mapping of ϕ.

The following definition introduces the notion of square associated to a CA:

Definition 8.1.1. Let A be an alphabet of q symbols. The square associated to the CA $F : A^{2(d-1)} \to A^{d-1}$ defined by the rule $f : A^d \to A$ is the square matrix S_F of size $q^{d-1} \times q^{d-1}$ with entries from $[q^{d-1}]$ defined for all $1 \leqslant i, j \leqslant q^{d-1}$ as

$$S_F(i, j) = \phi(F(\psi(i) \cdot \psi(j))), \quad (17)$$

where $\psi(i) \cdot \psi(j) \in A^{2(d-1)}$ denotes the concatenation of $\psi(i), \psi(j) \in A^{d-1}$.

Hence, the square S_F is defined by encoding the first half of the CA configuration as the row coordinate i, the second half as the column coordinate j and the output of the CA $F(\psi(i) \cdot \psi(j))$ as the entry in cell (i, j).

The next lemma shows that fixing $d - 1$ adjacent input variables in the global rules of a bipermutive CA yields a permutation between the remaining variables and the output:

Lemma 8.1.1 ([96]). *Let* $F : A^n \to A^{n-d+1}$ *be a bipermutive CA defined by local rule* $f : A^d \to A$. *Given* $\tilde{x} \in A^{d-1}$ *and* i *with* $0 \leqslant i \leqslant n-d+1$, *let* $F|_{\tilde{x},i} : A^{n-d+1} \to A^{n-d+1}$ *be the restriction of* F *obtained by fixing to* $\tilde{x}$ *the block of* $d-1$ *consecutive coordinates starting at* i *of the bipermutive CA input vector, i.e.* $x_i = \tilde{x}_0, x_{i+1} = \tilde{x}_1, \ldots, x_{i+d-2} = \tilde{x}_{d-2}$. *Then,* $F|_{\tilde{x},i}$ *is a permutation over* A^{n-d+1}.

Given Lemma 8.1.1, the authors show that the squares associated to bipermutive CA are in fact Latin squares. The proof follows the argument laid out in Lemma 2 of [96].

Lemma 8.1.2 ([100]). *Let* A *be an alphabet of* q *symbols, and* $d \geqslant 2$. *Then, the square* L_F *of the BCA* $F : A^{2(d-1)} \to A^{d-1}$ *defined by local rule* $f : A^d \to A$ *is a Latin square of order* $N = q^{d-1}$.

The next critical result presented in [100, 105] is the characterization of pairs of CA which generate orthogonal Latin squares. Specifically, the authors show that given two linear CAs of diameter d $F, G : \mathbb{F}_q^{2(d-1)} \to \mathbb{F}_q^{d-1}$ of length $2(d-1)$, respectively defined by the local rules $f, g : \mathbb{F}_q^d \to \mathbb{F}_q$ defined as:

$$f(x_0, \ldots, x_{d-1}) = a_0 x_0 + \cdots + a_{d-1} x_{d-1}, \tag{18}$$

$$g(x_0, \ldots, x_{d-1}) = b_0 x_0 + \cdots + b_{d-1} x_{d-1}, \tag{19}$$

then the Latin squares L_F and L_G of order q^{d-1} generated by F and G are orthogonal if and only if the polynomials $p_f(X), p_g(X) \in \mathbb{F}_q[X]$ associated to f and g are relatively prime.

The key in the proving of this result is exploiting the relation between the polynomials $p_f(X), p_g(X)$ associated to f and g and the transition matrices M_F and M_G of the two CAs. Indeed, the authors build a new matrix M by placing the transition matrix M_F above M_G, in such a way that the function $M(x, y)^\top$ is bijective if and only if the determinant of M is not null. But since M is a Sylvester matrix, then its determinant is the resultant of the two polynomials $p_f(X)$ and $p_g(X)$, and it is a classical result (e.g. [92]) that the resultant of two polynomials is nonzero if and only if they are relatively prime.

Further, the work presented in [46] addresses a key problem in the study of linear CA: the exhaustive generation of orthogonal cellular automata. While previous works focused on the construction of specific instances of orthogonal cellular automata for tasks such as generating MOLS or bent functions, this paper explores how to systematically generate all possible linear orthogonal CA of a given size.

Orthogonal cellular automata are defined as pairs (or sets) of cellular automata whose outputs are uncorrelated over time. In mathematical terms, two cellular automata are orthogonal if the dot product of their output sequences is zero over a given time period. This

property is crucial in applications where independence and minimal interference are required, such as in cryptographic key generation or experimental design.

Building upon the work presented in [105], the authors develop an algorithm for the exhaustive counting and enumeration of all pairs of linear Orthogonal Cellular Automata (OCA) of diameter d by counting and enumerating all pairs of coprime polynomials of degree $n = d - 1$ with a nonzero constant term.

The paper details the conditions under which pairs (or larger sets) of CA will be orthogonal, including constraints on the rule matrices that guarantee the orthogonality of the state sequences.

One of the major contributions of the paper is the demonstration that, for a given size n, there is a finite number of orthogonal CA that can be generated. The authors provide a classification of all such CA for small values of n, along with an algorithm that can be used to extend the classification to larger values. This exhaustive generation method has applications in constructing large sets of mutually orthogonal structures, such as Latin squares and cryptographic keys. The systematic approach to generating orthogonal CA outlined in the paper is particularly valuable for cryptographic applications, where the ability to generate multiple independent sequences is crucial for security. The authors of [46] also explore the combinatorial implications of these results, suggesting that orthogonal CA could be used to generate new classes of error-correcting codes with desirable properties.

8.2 THE NONLINEAR CASE

Cellular automata with nonlinear rules represent a significant departure from their linear counterparts, introducing greater complexity and unpredictability into their dynamics. Unlike linear CA, which rely on straightforward algebraic relationships for state transitions, nonlinear CA leverage intricate Boolean functions or other nonlinear mappings to dictate the evolution of states. This added complexity makes them particularly appealing for applications requiring high levels of security and robustness, such as cryptography and pseudo-random number generation.

One of the main contributions in this field is represented by the work done in [101], where the authors consider the problem of enumerating orthogonal Latin squares induced by nonlinear bipermutive CA, which could have interesting cryptographic applications. The core result presented in this paper is a necessary condition for the generation of orthogonal Latin squares. Specifically, the authors show that in order to give rise to orthogonal Latin squares two bipermutive rules must be *pairwise balanced* meaning the four pairs $(0,0)$, $(1,0)$, $(0,1)$ and $(1,1)$ must occur an equal number of times in the super-

position of their truth tables. Additionally, they derive a formula for the number of pairwise balanced bipermutive rules, and apply a combinatorial algorithm to enumerate all those pairs which generate orthogonal Latin squares up to $n = 6$ variables.

An extension of the work presented in [101] is provided by [103]. Here the authors expand (via two metaheuristics) the enumeration of all pairs of bipermutive which generate orthogonal Latin squares to $n = 7$ and $n = 8$ variables. In particular, they investigate the use of Genetic Algorithms (GA) [111] and Genetic Programming (GP) [76] to evolve orthogonal Latin squares engendered by cellular automata.

To address the challenge of optimizing the evolution of CA-based Orthogonal Latin Squares (OLS), the authors exploit the property that bipermutive rules involving n variables can be described using generating functions of $n-2$ variables. Consequently, the genotype of each individual in the population encodes a pair of generating functions, ensuring that the resulting phenotype corresponds to a pair of bipermutive CA capable of generating two Latin squares. As a result, the optimization process in both GA and GP can concentrate on enhancing the orthogonality and nonlinearity of the solutions, bypassing the need to verify the row-column permutation property. For GA, the authors propose three distinct encodings for the chromosomes representing candidate solutions. The first encoding combines the truth tables of the two generating functions into a single bitstring, allowing standard crossover and mutation operators to be directly applied. In the second encoding, the two generating functions are treated independently, with genetic operators applied separately to each function. Lastly, the approach takes advantage of two empirical observations derived from the search experiments. Firstly, bipermutive rules in CA that generate OLS must exhibit pairwise balance, ensuring that each of the four possible bit pairs appears equally often in the combined truth tables. Secondly, if the generating functions themselves are pairwise balanced, the resulting bipermutive rules will also maintain this balance. Leveraging these insights, the authors design the third encoding, where the genotype is structured as a balanced quaternary string. To ensure the balancedness property is maintained in quaternary strings, the authors develop specialized crossover and mutation operators tailored for GA. In contrast, for GP, they employ an encoding similar to the GA's double bitstring representation, where each generating function is modeled as an independent Boolean tree. Both GA and GP are evaluated on the smallest problem instances beyond the scope of exhaustive search, specifically the sets of generating function pairs with 5 and 6 variables. These instances correspond, respectively, to the sets of bipermutive CA pairs with local rules of $n = 7$ and $n = 8$ variables, or equivalently to the sets of CA-based Latin squares pairs of size 64×64 and 128×128.

9

DESIGN OF CORRELATION IMMUNE FUNCTIONS

This chapter aims to explore how cellular automata can be used to design correlation immune functions of a given weight. We use a construction of mutually orthogonal CA (MOCA)-a family of CA that generates a set of mutually orthogonal Latin squares (MOLS)-as introduced in [105]. We utilize the fact that any set of MOLS corresponds to an orthogonal array (OA) of strength 2. By analyzing any MOCA set, we prove that the binary expansion of MOCA is an OA of strength at least 2, leveraging the MOCA's characterization as orthogonal labelings on de Bruijn graphs. This result enables us to use the expanded binary OA of a MOCA family to create a Boolean function with at least 2-order correlation immunity. We then extend this construction to the context of Latin hypercubes of order d, providing a framework to construct correlation immune functions of any order d starting from a family of MOCA.

9.1 BASIC NOTIONS

In this section, we recall all relevant definitions to describe our results in the remainder of the chapter. We start by recalling basic concepts and results of Boolean functions, and how correlation immune functions can be characterized by orthogonal arrays. Then, we give a formal definition of *Latin hypercube*.

9.1.1 *Boolean Functions*

As a general reference, we follow Carlet's recent book on Boolean functions [21].

Let $\mathbb{F}_2 = \{0, 1\}$ represent the finite field with two elements, where addition and multiplication are defined as XOR (denoted by $\oplus$) and logical AND (denoted by juxtaposition), respectively. For any natural number n, the set $\mathbb{F}_2^n$ consists of all n-bit binary strings and forms a vector space. In this space, vector addition is performed component-wise using the XOR operation, and scalar multiplication for any $a \in \mathbb{F}_2$ involves multiplying a with each coordinate of a vector $x \in \mathbb{F}_2^n$ using the field multiplication. Given two vectors $x, y \in \mathbb{F}_2^n$, their Hamming distance, $d_H(x, y)$, is the number of posi-

tions where x and y differ. The scalar product of x and y is defined as $x \cdot y = \bigoplus_{i=1}^{n} x_i y_i$, which is the XOR of the pairwise products of their corresponding entries. The support of a vector $x \in \mathbb{F}_2^n$ is the set $supp(x) = i : x_i \neq 0$, indicating the indices where the coordinates of x are non-zero. The Hamming weight $w_H(x)$ is the size of this support, i.e., the number of non-zero entries in x. Alternatively, $w_H(x)$ can be viewed as the Hamming distance between x and the zero vector $\underline{0} \in \mathbb{F}_2^n$, which measures how many coordinates of x are non-zero.

For all $n \in \mathbb{N}$, a *Boolean function* of n variables is a mapping $f : \mathbb{F}_2^n \to \mathbb{F}_2$. The most straightforward way to represent f is by means of its truth table, i.e. the 2^n-bit vector that specifies for each input vector $x \in \mathbb{F}_2^n$ in the total order fixed on the vectors of $\mathbb{F}_2^n$ (e.g., the lexicographic order) the corresponding output value $f(x)$. Suppose that the lexicographic order is fixed on the vectors of $\mathbb{F}_2^n$, then the truth table of f is the vector $\Omega_f \in \mathbb{F}_2^{2^n}$ defined as:

$$\Omega_f = (f(0, \cdots, 0), f(0, \cdots, 1), \cdots, f(1, \cdots, 1)) \ . \tag{20}$$

In other words, the truth table is a 2^n-bit vector that lists the output value $f(x)$ for each input vector $x \in \mathbb{F}_2^n$, following lexicographic order. Just as we defined for binary vectors, the support of f is the set $supp(f) = x \in \mathbb{F}_2^n : f(x) \neq 0$, which consists of the inputs where f produces a non-zero output. The Hamming weight of f, denoted by $w_H(f)$, is the size of this support, or the number of non-zero outputs. Alternatively, the support and weight of f can be viewed as the set of non-zero entries in the truth table Ω_f and the number of such entries, respectively.

Another method to represent a Boolean function usually adopted in cryptography is the *Walsh transform*. Given $f : \mathbb{F}_2^n \to \mathbb{F}_2$, the Walsh transform of f is the function $W_f : \mathbb{F}_2^n \to \mathbb{Z}$ defined for all $a \in \mathbb{F}_2^n$ as:

$$W_f(a) = \sum_{x \in \mathbb{F}_2^n} (-1)^{f(x) \oplus a \cdot x} \ . \tag{21}$$

Intuitively, $W_f(a)$ quantifies the correlation between the function f and the linear function given by the scalar product $a \cdot x$. For this reason, the Walsh transform is a valuable tool for evaluating (among various cryptographic characteristics of f) nonlinearity, which refers to the Hamming distance between f and the set of all affine functions. The Walsh transform allows us to calculate the nonlinearity of f as:

$$nl(f) = 2^{n-1} - \frac{1}{2} \max_{a \in \mathbb{F}_2^n} \{|W_f(a)|\} \ . \tag{22}$$

9.1.2 *Correlation immune and resilient Boolean functions*

Other two cryptographic properties which will be of special interest for this thesis are correlation immunity and resilience.

Definition 9.1.1. A Boolean function $f : \mathbb{F}_2^n \to \mathbb{F}_2$ is *correlation immune of order* $1 \leqslant t \leqslant n$ if any subset of at $1 \leqslant k \leqslant t$ input variables is statistically independent from the output of f.

This property can also be described in terms of the Walsh transform, as first established by Xiao and Massey [150]: a function f is t-th order correlation immune if and only if $W_f(a) = 0$ for all vectors $a \in \mathbb{F}_2^n$ with Hamming weight $1 \leqslant k \leqslant t$.

Correlation immunity is crucial in defending against correlation attacks on stream ciphers that use the combiner model [135]. More recently, this criterion has become important in designing masking countermeasures to protect against side-channel analysis (SCA). In this context, the objective is to identify a t-th order correlation immune function that can withstand SCA attacks of order t. Additionally, it is desirable for this function to have the smallest possible Hamming weight to ensure that the masking countermeasure can be implemented efficiently.

In addition to the Walsh transform, correlation immune functions have a useful combinatorial interpretation through orthogonal arrays (see Definition 6.2.4). Formally, an orthogonal array with N runs, k factors, s levels, and strength t (denoted as $(N, k, s, t) - OA$) is an $N \times k$ matrix with entries from a set S of s elements, such that for any $N \times t$ subarray, every possible t-tuple from S^t appears exactly $\lambda = N/s^t$ times [56]. The value λ, known as the index of the OA, is determined by the other parameters. The connection between binary orthogonal arrays (with $s = 2$ levels) and correlation immune functions is established by the following result from [19]:

Lemma 9.1.1. *A Boolean function* $f : \mathbb{F}_2^n \to \mathbb{F}_2$ *is correlation immune of order* t *if and only if its support* $\mathrm{supp}(f) = \{x \in \mathbb{F}_2^n : f(x) \neq 0\}$ *is an* $(N, n, 2, t) - OA$.

In other words, designing n-variable, t-th order correlation immune Boolean functions for SCA masking countermeasures can be reduced to finding binary orthogonal arrays (OAs) with n factors and strength t. Minimizing the Hamming weight of the function corresponds to minimizing the number of runs N in the OA. Once such an OA is identified, the corresponding correlation immune function f can be defined by using the runs of the OA as the vectors that form the support of f.

In the context of cryptography, a special class of correlation immune functions, known as resilient functions, plays a crucial role in enhancing security. A Boolean function is called *resilient* if it is not only t-th order correlation immune but also balanced, meaning that it produces an equal number of zeros and ones in its output. We recall the precise definitions.

Definition 9.1.2. A Boolean function $f : \mathbb{F}_2^n \to \mathbb{F}_2$ is *balanced* if its output is 0 for exactly half of the inputs and 1 for the other half.

Definition 9.1.3. A Boolean function $f : \mathbb{F}_2^n \to \mathbb{F}_2$ is *resilient of order* $1 \leqslant t \leqslant n$ if it is correlation immune of order t and balanced.

Notice that the balance property ensures that the function does not favor any particular output, making it more resistant to attacks that exploit statistical biases. Resilient functions are particularly important in stream cipher design and side-channel resistance, where both correlation immunity and output uniformity are essential for maintaining security against a range of cryptanalytic and side-channel attacks.

9.1.3 *Latin (hyper)cubes*

In this work we will also rely on a generalization of the concept of Latin squares, i.e. Latin hypercubes.

Let $[n]$ denote the set $\{1, 2, \ldots, n\}$ and let $[n]^d$ denote the cartesian product $[n] \times [n] \times \cdots \times [n]$ of d copies of $[n]$. By a *hypercube* of order n and dimension d we mean a d-dimensional array of n^d cells where the cells are indexed by $[n]^d$ and each cell contains an element of $[n]$ which we will call a symbol. Suppose that H is such a hypercube and c is any cell of H. A *line* through c is a set of n cells of H whose coordinates match those of c except possibly in the k-th coordinate (there is one line for each choice of k). A *hyperplane* through c is a set of n^{d-1} cells of H whose k-th coordinate matches that of c (there is one hyperplane for each choice of k). Any hyperplane in a d-dimensional hypercube can be considered to be a $(d-1)$-dimensional hypercube, simply by dropping the common coordinate. We use vector notation such as $\vec{x}$ for an element of $[n]^d$. In a hypercube H we denote the symbol in the cell with coordinates $\vec{x} = (x_1, x_2, \ldots, x_d)$ by $H(\vec{x})$ or $H(x_1, x_2, \ldots, x_d)$.

Definition 9.1.4. We say that a hypercube H is *Latin* if :

- every permutation $(x_1, x_2, \ldots, x_d)$ of the symbols in $[n]$ appears exactly once in the H, and
- every hyperplane is a $(d-1)$-Latin hypercube.

Example 9.1.1. Taking $d = 3$ in Definition 9.1.4 we obtain a Latin cube. For example, Figure 37 represents a $4 \times 4 \times 4$ Latin cube.

9.2 CONSTRUCTION OF CORRELATION IMMUNE FUNCTIONS

This section is devoted to show how a set of k-MOCA can be used to define a binary OA. We will start in Section 9.2.1 by constructing a binary OA of strength at least 2. Then in Section 9.2.2 we show that by assuming a stronger orthogonality property we are able to define an orthogonal array A of strength 3. Finally, in Section 9.2.3 we generalize this result to construct correlation immune functions of any given order d.

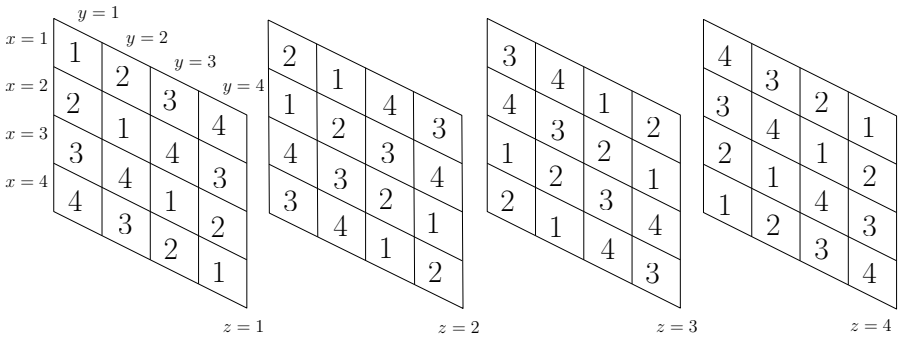


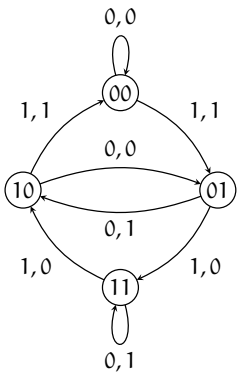
Figure 37.: Visualization of a $4 \times 4 \times 4$ cube, with each slice representing a Latin square. The cube consists of four layers, where each layer (corresponding to a different value of z) contains a distinct 4×4 Latin square, showcasing the property that each number appears exactly once in each row and column.

9.2.1 *Correlation immune functions of order at least 2*

This section is devoted to show how a set of k-MOCA can be used to define a binary OA of strength at least 2. This will allow us, in turn, to construct correlation immune functions of order at least 2. To this end, let us first review the concept of coupled de Bruijn graph introduced in Section 6.1, which will be useful in our proof.

A de Bruijn graph of order b over a set S of m symbols is defined as $G_{m,b} = (V, E)$, where the vertex set $V = S^b$, and two vertices $u, v \in V$ are connected by a directed edge if and only if they overlap on their rightmost and leftmost $b-1$ coordinates, respectively. Now, assume $S = \mathbb{F}_2$, and set $b = d-1$. A local rule $f : \mathbb{F}_2^d \to \mathbb{F}_2$ of diameter d can be described as a labeling function $l_f : E \to \mathbb{F}_2$ on the edges of $G_{2,b}$. Specifically, for each edge connecting u and v, we define $l(u, v) = f(u \odot v)$, where $u \odot v \in \mathbb{F}_2^d$ is the vector formed by concatenating the last coordinate of v to u. The output of a cellular automaton (CA) governed by rule f corresponds to a path along the edges of the de Bruijn graph, following the sequence of overlapping vertices based on the input. If f is bipermutive, then the labels of the outgoing (and incoming) edges from any vertex $v \in V$ form a permutation of $\mathbb{F}_2$. This property ensures that the CA is surjective, meaning its output can traverse all edges bidirectionally.

Now, suppose we have two labelings $l_f, l_g : E \to \mathbb{F}_2$ defined by bipermutive rules $f, g : \mathbb{F}_2^d \to \mathbb{F}_2$. The corresponding graph is called the coupled de Bruijn graph associated with f and g, as each edge $(u, v) \in E$ is now labeled with a pair of bits $(f(u \odot v), g(u \odot v))$. We define two bipermutive labelings l_f and l_g as orthogonal if, for each pair of vectors $(x, y) \in \mathbb{F}_2^b$, there exists exactly one path through the edges of the coupled de Bruijn graph that is labeled by (x, y). It can be easily shown that this provides an equivalent characterization of



$(v_1, v_2) \to (u_1, u_2)$	l_f	l_g
$00 \to 00$	0	0
$10 \to 00$	1	1
$01 \to 10$	0	1
$11 \to 10$	1	0
$00 \to 01$	1	1
$10 \to 01$	0	0
$01 \to 11$	1	0
$11 \to 11$	0	1

Figure 38.: Example of orthogonal labelings for the de Bruijn graph $G_{2,2}$ induced by the CA local rules 90 and 150 of diameter $d = 3$.

orthogonal cellular automata (CA). We present this formal statement below, as it will be useful later in our proof:

Lemma 9.2.1. *Let* $d \in \mathbb{N}$ *with* $b = d - 1$, *and* $f, g : \mathbb{F}_2^d \to \mathbb{F}_2$ *be two bipermutive rules of diameter* d. *Then, the CA* $F, G : \mathbb{F}_2^{2b} \to \mathbb{F}_2^b$ *respectively equipped with rule* f *and* g *are orthogonal if and only if the labelings on the coupled de Bruijn graph of* f *and* g *are orthogonal.*

To illustrate this, let us take $d = 3$ and consider two bipermutive local rules: $f(x_1, x_2, x_3) = x_1 \oplus x_3$ and $g(x_1, x_2, x_3) = x_1 \oplus x_2 \oplus x_3$, which correspond to rules 90 and 150 in Wolfram's notation. These two rules generate orthogonal cellular automata (CA) with Latin squares of order $2^2 = 4$, and the resulting coupled de Bruijn graph is shown in Figure 38. Now, let $F_1, F_2, \cdots, F_k : \mathbb{F}_2^{2b} \to \mathbb{F}_2^b$ be a set of k-MOCA, defined by local rules $f_1, \cdots, f_k : \mathbb{F}_2^d \to \mathbb{F}_2$ of diameter d, where $b = d - 1$. We construct an $N \times n$ array A, with $N = 2^{2b}$ and $n = kb$, as follows: for each $(x, y) \in \mathbb{F}_2^{2b}$, the row of A corresponding to (x, y) is given by:

$$A(x, y) = (F_1(x, y), F_2(x, y), \cdots, F_k(x, y)) \ . \tag{23}$$

In other words, the array A is created by concatenating the outputs of the k MOCA for every possible input pair $(x, y) \in \mathbb{F}_2^{2b}$. We will now demonstrate that this array forms an orthogonal array (OA) of strength 2.

Lemma 9.2.2. *The array* A *defined in Eq.* (23) *is an* $\mathrm{OA}(N, n, 2, 2)$, *where the number of runs is* $N = 2^{2b}$ *and the number of factors is* $n = kb$.

Proof. We need to show that in any subset of $t = 2$ columns i, j of A each pair of bits $(x_i, x_j) \in \mathbb{F}_2^2$ occurs exactly $\lambda = N/2^t = 2^{2b-2}$ times. Without loss of generality, we can assume that $k = 2$, since $F_1, \cdots F_k$ is a family of k-MOCA. Hence, we have two main cases to check for two columns $i \neq j$:

1. i, j belong to the output of the same CA F_l.
2. i, j belong to the output of two different CA F_l, F_m (which are orthogonal).

Let us start from the first case, i.e. i and j are chosen among the columns of the same CA F_l. Let $(\tilde{x}_i, \tilde{x}_j) \in \mathbb{F}_2^2$ be the value of the two bits of which we want to compute the multiplicity of occurrence in columns i and j. Since the two CA F_l and F_m are orthogonal, it means that by fixing the output of F_m to a specific vector $(y_1, \cdots, y_b) \in \mathbb{F}_2^b$, each possible vector $(x_1, \cdots, x_b) \in \mathbb{F}_2^b$ occurs exactly once as an output of F_l. Suppose now that we fix the i-th and j-th coordinates respectively to $\tilde{x}_i$ and $\tilde{x}_j$ in the output of F_l. Since we have 2^{b-2} free coordinates, it follows that the pair $(\tilde{x}_i, \tilde{x}_j)$ occurs 2^{b-2} times if we keep the output of F_m fixed to $(y_1, \cdots, y_b)$. If we consider the occurrences of $(\tilde{x}_i, \tilde{x}_j)$ in F_l across all possible outputs of F_m we need to multiply 2^{b-2} by 2^b, i.e. the number of possible ways to fix the output of F_m. Therefore the total number of occurrences is $2^{b-2} \cdot 2^b = 2^{2b-2}$.

Suppose now that i and j are columns respectively of F_l and F_m. If all output coordinates of F_l and F_m are fixed respectively to x and $y \in \mathbb{F}_2^b$, then there exists a single row of A labeled by x and y, since F_l and F_m are orthogonal, and by Lemma 9.2.1 there is a unique path on the coupled de Bruijn graph labelled by (x, y). We proceed by induction on the number of free coordinates in (x, y) to show that if we only have two of them fixed, i.e. $\tilde{x}_i$ and $\tilde{y}_j$, then there are exactly 2^{2b-2} paths on the de Bruijn graph that feature $\tilde{x}_i$ and $\tilde{y}_j$ in those coordinates. As a base case, suppose that we have only one free coordinate in the pair of paths, i.e. all other $2b-1$ are fixed. Then, since each of the two labelings is bipermutive, it follows that there are exactly 2 paths labelled by the $2b-1$ fixed coordinates. For the induction step, suppose that there are $1 \leqslant p < 2b$ free coordinates, and thus 2^p paths labelled by the remaining $2b-p$ fixed coordinates by induction hypothesis. If we free an additional coordinate, we need to multiply the number of paths with p free coordinates by 2, since each of them can be completed in 2 different ways in the additional free coordinate, due to the bipermutivity of the two rules. Hence, the number of partially labelled paths with $p+1$ free coordinates is 2^{p+1}. If we take the particular case where only 2 coordinates i and j are fixed respectively to $\tilde{x}_i$ and $\tilde{y}_j$ (or equivalently, $2b-2$ are free), it follows that there are 2^{2b-2} paths partially labeled by $\tilde{x}_i$ and $\tilde{y}_j$. □

Combining Lemma 9.1.1 and 9.2.2, we arrive at the following result, which provides a method for constructing a second-order correlation immune function using a set of k-MOCA:

Theorem 9.2.3. *Let* $F_1, \cdots, F_k : \mathbb{F}_2^{2b} \to \mathbb{F}_2^b$ *be a set of* k*-MOCA of diameter* $d = b + 1$, *and let* $n = kb$. *Then, the* n*-variable function* $f : \mathbb{F}_2^n \to \mathbb{F}_2$

whose support is defined by the array A in Eq. (23) is correlation immune of order at least 2. In particular, the Hamming weight of f is $N = 2^{2b}$.

Notice that when $k = 2$, the resulting function is trivially the constant function $f(x) = 1$ for all $x \in \mathbb{F}_2^n$. This occurs because the number of input variables is $n = kb = 2b$, meaning that the truth table of f consists of 2^{2b} entries. Simultaneously, the number of runs in the orthogonal array (OA) corresponding to the $k = 2$ MOCA is also $N = 2^{2b}$. As a result, the support of f covers the entire truth table. For this reason, we will focus primarily on the case where $k = 3$ in the subsequent discussion.

9.2.2 *Correlation immune functions of order 3*

In this section we prove a modified version of Lemma 9.2.2. In particular, by assuming a stronger orthogonality property we are able to show the array A defined in Eq. (23) is an orthogonal array of strength 3.

Theorem 9.2.4. *Let* $F_1, F_2, \cdots, F_k : \mathbb{F}_2^{2b} \to \mathbb{F}_2^b$ *be a set of k-MOCA such that any triple* F_i, F_j, F_k *induces a Latin cube and any slice of this cube is a Latin square. Then the array A defined in Eq. (23) is an* $OA(N, n, 2, 3)$, *where the number of runs is* $N = 2^{2b}$ *and the number of factors is* $n = kb$.

Proof. By contradiction, A is not an orthogonal array of strength 3. Then it is possible to find three unbalanced columns a, b, c. Notice that it can never happen that a, b, c belong to same cellular automaton F_i, so either $a, b \in F_i$ and $c \in F_j$, or $a \in F_i$, $b \in F_j$ and $c \in F_k$. In both cases, since the columns are not balanced there is at least an output triplet (x_a, x_b, x_c) which is missing. But then the cube is not containing all possible triplets. □

Remark 9.2.1. Although the assumption made in Lemma 9.2.4 that any triple F_i, F_j, F_k of k-MOCA induces a Latin cube is quite strong, it is not possible to relax it. Indeed, let $F_1, F_2, \cdots, F_k : \mathbb{F}_2^{2b} \to \mathbb{F}_2^b$ be a set of k-MOCA and suppose that any triple F_i, F_j, F_k induces a cube (not necessarily a Latin cube) such that any slice is a a Latin square. Then it is possible to construct and example where there are three columns $a, b, c \in F_i \cup F_j \cup F_k$ such that $c = a \oplus b$, and thus all three pairs $(a, b), (a, c), (b, c)$ are balanced, but since c is dependent on a and b the triple (a, b, c) is unbalanced.

9.2.3 *Correlation immune functions of order d*

In this section we extend the result proved in Lemma 9.2.4 and we show that it is possible, starting from a family $F_1, F_2, \cdots, F_k : \mathbb{F}_2^{2b} \to \mathbb{F}_2^b$ of k-MOCA to define an orthogonal array of strength d for any chosen d.

Theorem 9.2.5. *Let* $F_1, F_2, \cdots, F_k : \mathbb{F}_2^{2b} \to \mathbb{F}_2^b$ *be a set of* k*-MOCA such that any* d*-tuple* $F_{i_1}, F_{i_2}, \cdots, F_{i_d}$ *induces a Latin* d*-hypercube and any slice of this cube is a Latin square. Then the array* A *defined in Eq.* (23) *is an* $OA(N, n, 2, d)$*, where the number of runs is* $N = 2^{2b}$ *and the number of factors is* $n = kb$*.*

Proof. Suppose by contradiction that, under the assumptions of the lemma, A is not an orthogonal array of strength d. Then it is possible to find d unbalanced columns $c_1, \ldots, c_d$. Notice that it can never happen that all columns c_j, $j = 1, \ldots, d$, belong to same cellular automaton F_{i_j}, so either there exists $F_{i_{\tilde{j}}}$, $i_{\tilde{j}} \in \{i_1, \ldots, i_d\}$ such that two columns $c_1, c_2 \in F_{i_{\tilde{j}}}$, or $c_j \in F_{i_j}$, $j \in \{1, \ldots, d\}$. In both cases, since the columns are not balanced there is at least an output vector $(x_1, \ldots, x_d)$ which is missing. But then the hypercube is not containing all possible d-tuples. □

Remark 9.2.2. Similarly to what we noted in Remark 9.2.1, also in this case it is not possible to renounce the assumption that any d-tuple $F_{i_1}, F_{i_2}, \cdots, F_{i_d}$ induces a Latin d-hypercube. Indeed, if this was the case then we could choose F_{i_d} to be a function of $F_{i_1}, \ldots, F_{i_{d-1}}$ so that given d columns $c_1, \ldots, c_d \in F_{i_1} \cup \cdots \cup F_{i_d}$ every $(d-1)$-subset $(c_{i_1}, \ldots, c_{i_{d-1}})$ is balanced, but the set $(c_1, \ldots, c_d)$ is unbalanced.

Part III

FINAL REMARKS

10

CONCLUSIONS AND FUTURE WORKS

As final remarks, we summarize the contribution of this thesis and we state some future directions of research and open problems.

10.1 CONTRIBUTIONS

The contributions of this thesis must be placed into different categories.

10.1.1 *Combinatorial and Stochastic Topology*

In Chapters 2 we investigated how the structure of a graph G is represented in the magnitude homology groups of G. To strengthen this connection, we developed the theory of eulerian magnitude homology, built on a subcomplex of the magnitude chain complex that omits "singular" trails that must revisit vertices. By restricting our attention to the eulerian trails, we are able to characterize classes of graphs which can support eulerian magnitude cycles along the $k = \ell$ line. In this case, the image of the differential is zero for dimension reasons and so understanding these classes is sufficient to characterize eulerian magnitude homology completely. In Theorem 2.2.2 we produce a generating set for these homology groups, thus characterizing these groups in terms of existence of particular subgraphs. Due to its combinatorial complexity, we leave the matter of understanding how elements of these generating sets interact open for the present.

Singular trails, which repeat landmarks, generate the complementary discriminant magnitude homology. The two theories are intertwined through the corresponding long exact sequence (1), as the differential of a trail that must revisit a vertex may include trails that do not. Characterizing this interaction provides us with tools for dissecting the generators of magnitude homology, again focused along the $k = \ell$ diagonal where the combinatorics are more accessible. Here, as we saw in Theorems 2.3.3 and 2.3.4, the existence of nonsingular subtrails is regulated by eulerian magnitude homology groups of lower degree.

Further, in Chapter 3 we turn our attention to the context of Erdös-Rényi random graphs and random geometric graphs on the standard

torus, where boundary effects are removed, we leverage our understanding of classes of graphs supporting generators along the $k = \ell$ line to develop vanishing thresholds for eulerian magnitude homology in limiting expectation in Theorems 3.1.4 and 3.2.6. Combined with the long exact sequence relating eulerian and discriminant magnitude homology, these results allow us to completely characterize the magnitude homology groups in the vanishing range. Outside of the vanishing range, we provide limiting characterizations of the (k, k)-Betti numbers in Theorems 3.1.6 and 3.2.8 for both classes of graphs, and corresponding central limit theorems in Theorems 3.1.7 and 3.2.9.

The results presented in these chapters appear in the preprint "Eulerian magnitude homology: subgraph structure and random graphs" [51].

Moreover, in Chapter 4 we investigated the regimes where an Erdös-Rényi random graph has torsion free eulerian magnitude homology groups. To this end, we introduced in Section 4.2 the eulerian Asao-Izumihara complex - a quotient CW-complex whose homology groups are isomorphic to direct summands of the graph eulerian magnitude homology group. We then proceeded by producing a vanishing threshold for a shelling of the eulerian Asao-Izumihara complex 4.3.1, which led to the result stated in Theorem **??**, establishing the regimes where eulerian magnitude homology of Erdös-Rényi random graphs is torsion free.

The results presented in this chapter appear in the preprint "On torsion in eulerian magnitude homology of Erdős-Rényi random graphs" [109].

10.1.2 *Algorithms and Complexity Theory*

In Chapter 5 we investigated the computational cost of calculating the ranks of first diagonal eulerian magnitude homology groups of a graph G, $\mathrm{EMH}_{k,k}(G)$. We first prove in Theorem 5.1.4 that this problem is $\#W[1]$-complete (under FPT many-one reductions) by rewriting it as a subgraph isomorphism problem. Then, we produce the first-diagonal algorithm (Algorithm 1), a BFS-based algorithm which computes the eulerian magnitude chains $\mathrm{EMC}_{k,k}(G)$ and $\mathrm{EMC}_{k-1,k}(G)$ in super-polynomial time.

The results presented in this chapter appear in the preprint "Computing eulerian magnitude homology" [110].

10.1.3 *Cellular Automata*

In Chapter 9 we explored how cellular automata can be used to design correlation immune functions of a given weight. We first proved in Lemma 9.2.2 that the binary expansion of any set of mutually or-

thogonal CA (MOCA) is an orthogonal array (OA) of strength at least 2, leveraging the MOCA's characterization as orthogonal labelings on de Bruijn graphs. This result enables us to use the expanded binary OA of a MOCA family to create a Boolean function with at least 2-order correlation immunity, as stated in Theorem 9.2.3. We then extend this construction to the context of Latin hypercubes of order d, providing in Theorem 9.2.5 a framework to construct correlation immune functions of any order d starting from a family of MOCA.

The results presented in this chapter are the base for the manuscript "CA-based correlation immune functions" currently in preparation, [99].

10.2 OPEN PROBLEMS

While there are many directions for future research and open problems, here we want to highlight a few that we find particularly interesting.

10.2.1 *Combinatorial and Stochastic Topology*

The tools described in Chapters 2, 3 and 4 suggest a number of avenues for further work. Here we highlight a few that we find particularly interesting.

1. We heavily leverage equality between the length ℓ and number of landmarks k in a trail to obtain our combinatorial characterization of generators in Section 2.1. We believe that these results can in turn be leveraged to iteratively study the combinatorics witnessed by the $\ell = k + i$ lines for increasing i, providing more insight into the graph-theoretic meaning of the higher magnitude homology groups.

2. The expected values of Betti numbers of both Erdős-Rényi and random geometric graphs are computed here using an approach which heavily depends on the possibility of computing explicitly the probability that each edge appears in the graph. There are examples in the literature [70, 8, 15] of limit results for Betti numbers of random structures exploiting discrete Morse theory: would it be possible to use similar techniques to extend the results present in this work to more general classes of graphs?

3. The vanishing threshold and expected Betti numbers for random geometric graphs, computed in Theorem 3.2.6 and 3.2.8 respectively, are rather coarse upper bounds. Indeed, we obtained

them by bounding the number of $(k+1)$-tuples of vertices inducing a path of length k with the quantity

$$\binom{n}{k+1}\left(\frac{\pi r^2}{|A|}\right)^k,$$

which is impossible to attain in some regimes, e.g. relatively sparse and sparse regimes [36]. Is it possible to establish more refined distribution-dependent thresholds and Betti numbers estimates by taking into account the specific regime?

4. The random geometric graphs studied here are embedded in the torus $\mathbb{T}^2$ mostly as a matter of convenience. Would a similar proof for the vanishing threshold of $EMH_{k,k}(G)$ work (possibly with some restrictions on the acceptable regimes) in more general settings?

5. A natural development of the work present in Chapter 4 concerns a deterministic result about the presence of torsion in eulerian magnitude homology groups of graphs, and we believe that this kind of result can be achieved by exploiting the strong connection between the eulerian magnitude chain complex and the complex of injective words [44, 13]. Indeed, Hepworth and Roff [61] thoroughly analyzed in the context of directed graphs the *magnitude-path spectral sequence (MPSS)*, a spectral sequence whose E^1 page is exactly standard magnitude homology, path homology [53] can be identified with a single axis of page E^2, and whose target object is reachability homology [60]. Reproducing the computations proposed in [61, Section 2] using a filtration of the complex of injective words, leads to a version of the MPSS where the E^1 page is exactly eulerian magnitude homology. Since the homology of the complex of injective words, as the target object, controls the behavior of the spectral sequence, we believe this connection holds great potential and is well worth exploring.

10.2.2 *Algorithms and Complexity Theory*

The one presented in Chapter 5 is the first algorithm computing eulerian magnitude homology in non-exponential time, thus representing a high improvement of definition's computational complexity. Nevertheless, the author believes the present algorithm's efficiency can be further improved with techniques similar to the ones used in the case of Persistent Homology [121]. This would initiate a new current in the field of Topological Data Analysis based on magnitude homology and devoted to network analysis.

10.2.3 *Cellular Automata*

We believe an interesting avenue for future research concerns the development of a heuristic approach for simplifying the complex theoretical method presented in Chapter 9. Indeed, from Remark 9.2.2 we deduce that counter examples are given by $(d-1)$-resilient functions that are not d-resilient.

Asymptotic estimates for the number of d-resilient functions where provided, for example, by Thomas W. Cusick and Pantelimon Stanica in [30] and by Claude Carlet in [22]. Denoting by $R_d(n)$ the number of d-resilient functions, the ratio $\frac{R_d(n)}{R_{d-1}(n)}$ has been found to have order $O(2^{-(n-d-1)})$, suggesting that as d increases the number of d-resilient functions is exponentially smaller compared to the number of $(d-1)$-resilient functions.

Given this result, we believe it would be reasonable to define an heuristic method to construct correlation immune functions of order d that, given a set of k-MOCA $F_1, F_2, \cdots, F_k : \mathbb{F}_2^{2b} \to \mathbb{F}_2^b$, only checks that for every d-tuple $F_{i_1}, F_{i_2}, \cdots, F_{i_d}$, any pair (F_{i_a}, F_{i_b}) induces a Latin square.

In conclusion, the topics introduced in this thesis open up a wide array of possibilities for further research. Each area explored presents its own set of challenges and opportunities, offering fertile ground for deeper investigation: by building on the foundations laid here, future studies can delve into more specialized aspects, refine methodologies, and uncover new insights.

Part IV

APPENDIX

A

A TOPOLOGY FOR P-SYSTEMS WITH ACTIVE MEMBRANES

We provide here an extended summary of the paper "A Topology for P Systems with Active Membranes" [33]. This paper makes a significant contribution to the study of P systems by bridging the gap between membrane computing and dynamical systems theory. The paper introduces a new topology for P systems, based on a carefully defined distance measure between configurations, and investigates the resulting dynamical properties, including sensitivity to initial conditions and topological transitivity.

The work is both theoretical, providing rigorous proofs of key properties, and practical, ensuring that the proposed concepts are computationally feasible. The findings have important implications for the understanding of P systems, particularly in terms of their potential to model complex, non-chaotic behaviors in a controlled and analyzable way. The paper opens new avenues for research at the intersection of membrane computing and dynamical systems, with the potential to enhance the theoretical foundations and practical applications of P systems in computational biology and beyond.

In what follows we will maintain the section layout of the original work.

A.1 INTRODUCTION

The paper "A Topology for P Systems with Active Membranes" delves into a specialized area of membrane computing, focusing on the development of a topology for deterministic P systems with active membranes and analyzing their behavior within the framework of discrete-time dynamical systems. P systems, or membrane systems, are a class of distributed parallel computing models inspired by the functioning of biological cells [124]. These systems leverage the structure and processes of membranes to perform computations, making them powerful tools for addressing complex computational problems [120, 89, 88, 122, 129].

The motivation behind the paper lies in the intersection of two fields: membrane computing and dynamical systems theory [77]. While P systems have been extensively studied in terms of their

computational power and efficiency [125, 139], their behavior from the perspective of dynamical systems theory remains underexplored. This paper aims to fill that gap by defining a new topology for P systems and investigating the dynamical properties that arise from this topology.

A.2 BACKGROUND AND BASIC CONCEPTS

A.2.1 *P systems*

P systems with active membranes extend the basic P system model by allowing membranes to change their state actively, thus increasing the computational power and complexity of the system [138, 126]. These systems are characterized by the presence of multiple membranes, each with a distinct label, and objects that evolve according to specified rules. The rules dictate the behavior of the membranes and objects, including division, dissolution, and communication between membranes.

A.2.2 *Discrete-Time Dynamical Systems*

In mathematics, a discrete-time dynamical system is one in which a function describes the time evolution of a system at discrete time intervals. The behavior of such systems can be analyzed using concepts like orbits, periodicity, transitivity, and chaos [34, 24, 23]. The paper adapts these concepts to the context of P systems, where the system's state evolves according to the application of rules at each discrete time step.

A.3 COUNTABILITY OF THE CONFIGURATION SPACE

One of the foundational results in the paper is the proof that the configuration space of a P system is countable. The configuration space consists of all possible states that the system can be in, given a fixed set of objects and membrane labels. By demonstrating the countability of this space, we show that it is possible to enumerate all configurations, which has significant implications for the system's behavior.

The countability result implies that classical chaotic behavior as defined by Devaney's criteria (which require the existence of an uncountable set of configurations with sensitive dependence on initial conditions) cannot occur in these systems. Since the configuration space of P systems is countable, this type of chaotic behavior is ruled out, leading to the conclusion that while P systems can exhibit complex behavior, it is not chaotic in the traditional sense.

A.4 DISTANCE FOR CONFIGURATIONS

To further analyze the dynamical properties of P systems, we introduce the definition of a distance between configurations. This distance is designed to capture the notion of dissimilarity between different states of the system, and the measure is defined in a way that reflects how "far apart" two configurations are in terms of the number of operations (such as membrane division or object evolution) needed to transform one configuration into the other.

The distance plays a crucial role in defining the topology of the configuration space. We prove that this distance is continuous under the functions induced by the P system's rules, meaning that small changes in the configuration lead to small changes in the distance. However, the resulting topological space is discrete, meaning that each configuration is isolated from the others, and the space is not complete, which implies that there are "gaps" in the configuration space that cannot be filled by configurations.

A.5 DYNAMICAL PROPERTIES OF P SYSTEMS

The paper adapts several classical concepts from dynamical systems theory to the context of P systems, focusing on sensitivity to initial conditions and topological transitivity.

A.5.1 *Sensitivity to Initial Conditions*

In classical dynamical systems, sensitivity to initial conditions is a hallmark of chaotic systems, where small differences in the initial state can lead to vastly different outcomes. We adapt this concept to P systems, defining a system as sensitive if there exists a configuration and a small perturbation (in terms of the distance measure) such that the evolution of these two configurations diverges significantly over time.

We prove that certain P systems can exhibit sensitivity to initial conditions, though this sensitivity does not imply chaos in the classical sense due to the countability of the configuration space.

A.5.2 *Topological Transitivity*

Topological transitivity is another key concept in dynamical systems theory, indicating that the system can move from any given configuration to any other configuration under the application of the system's rules. A system that is topologically transitive is considered to be "mixing" in the sense that its configurations are thoroughly mixed over time.

The paper demonstrates that certain P systems can exhibit topological transitivity, meaning that the system's dynamics allow it to explore a wide range of configurations. This property suggests that P systems have the potential to model complex behaviors, even within the constraints of a discrete and countable configuration space.

A.6 EFFICIENT COMPUTABILITY OF THE DISTANCE MEASURE

In addition to theoretical insights, the paper addresses the practical aspect of the proposed distance measure by proving that it can be computed efficiently. We show that the distance between two configurations can be calculated in polynomial time relative to the size of the input configurations. This result is significant because it ensures that the topological properties of P systems can be analyzed without incurring prohibitive computational costs, making the study of these systems feasible for real-world applications.

A.7 FUTURE WORK

Looking forward, we suggest several avenues for future research. One potential direction is to explore the relationship between the topological properties of P systems and their computational power, particularly in the context of solving complex problems. Another area of interest is the extension of the proposed topology to other variants of P systems, such as probabilistic or stochastic P systems, where the introduction of randomness could lead to different dynamical behaviors.

B

A GENETIC PROGRAMMING BASED HEURISTIC TO SIMPLIFY RUGGED LANDSCAPES EXPLORATION

We provide here an extended summary of the paper "A Genetic Programming Based Heuristic to Simplify Rugged Landscapes Exploration" [130], delving into the paper's methodology, experimental design, results, and conclusions, providing a detailed and precise overview of we' contributions.

In what follows we will maintain the section layout of the original work.

B.1 INTRODUCTION

Optimization problems in complex and rugged fitness landscapes pose significant challenges due to the presence of numerous local optima, which often lead to premature convergence of traditional optimization algorithms. These landscapes, characterized by their "ruggedness", make it difficult for algorithms to efficiently search for global optima, as they can become trapped in suboptimal regions. This paper addresses this issue by proposing a novel heuristic that combines Genetic Programming (GP) [76] and Fuzzy Self-Tuning Particle Swarm Optimization (FST-PSO) [117] to construct a surrogate model of the original function. This model is designed to smooth the rugged landscape while preserving the global optimum's location, facilitating a more efficient and effective search process.

We introduce the concept of a fitness landscape as a representation of the search space in optimization problems, where each point in the space corresponds to a potential solution, and its height corresponds to the fitness value. In rugged landscapes, the multitude of peaks and valleys complicates the search, often leading traditional methods like gradient-based algorithms or simple heuristics to get stuck in local optima. The paper posits that by using GP to evolve a smoother surrogate model, the search process can be significantly improved, allowing for the discovery of the global optimum in a more reliable and computationally efficient manner.

B.2 LITERATURE REVIEW

In this section a comprehensive background is provided on the integration of Genetic Programming with Particle Swarm Optimization and the use of surrogate models in optimization. Specifically, the review is divided into two key areas: first, the role of Genetic Programming and Particle Swarm Optimization in complex optimization tasks, and second, the development and application of surrogate models to simplify such tasks.

B.2.1 *Genetic Programming and Particle Swarm Optimization*

Genetic Programming (GP) [76] is an evolutionary algorithm that evolves programs or models, typically represented as tree structures, which can be used to solve complex problems. GP is flexible, allowing for the evolution of models of varying sizes and complexities, and has been applied to a wide range of problems, from symbolic regression to automated design.

Particle Swarm Optimization (PSO) [74], on the other hand, is a population-based optimization algorithm inspired by the social behaviors of swarming animals. PSO optimizes a problem by iteratively improving a candidate solution concerning a given measure of quality. The Fuzzy Self-Tuning Particle Swarm Optimization (FST-PSO) is a variant of PSO introduced in [117] that incorporates fuzzy logic to dynamically adjust the algorithm's parameters during the optimization process, enhancing its ability to balance exploration and exploitation without manual parameter tuning.

The integration of GP with PSO has been explored in various contexts, with studies showing that combining these techniques can enhance the search for global optima in complex landscapes [45, 131, 68]. GP's ability to evolve complex models complements PSO's strength in efficiently exploring the search space.

However, previous studies also highlight challenges, such as the computational cost of evolving complex models and the difficulty in maintaining a balance between exploration and exploitation, particularly in high-dimensional spaces.

B.2.2 *Surrogate Modeling in Optimization*

Surrogate models [12], also known as meta-models, are approximations of the objective function used to reduce the computational cost of optimization. These models are particularly useful when the objective function is expensive to evaluate, as they allow for a more efficient exploration of the search space by approximating the true fitness landscape [63].

B.3 METHODOLOGY

This section outlines the proposed approach, which integrates Genetic Programming with Fuzzy Self-Tuning Particle Swarm Optimization to develop a surrogate model that simplifies the exploration of rugged fitness landscapes.

B.3.1 *Genetic Programming (GP)*

Genetic Programming is a form of evolutionary algorithm that evolves tree-based representations of programs or models. These tree structures can vary in size and complexity and are evolved through genetic operations such as *crossover* (recombination of the genes), *mutation* (random alteration), and *selection* (choosing the fittest individuals to reproduce). In this study, GP is used to evolve surrogate models that approximate the original rugged fitness landscape. The objective is to create a surrogate that smooths out the local optima, making the global search more straightforward while retaining the global optimum's position.

The GP process begins with an initial population of random tree structures, each representing a potential surrogate model. These models are evaluated based on their fitness, which in this context refers to how well they approximate the original landscape while providing a smoother surface for optimization. Over successive generations, the GP algorithm refines these models, gradually improving their accuracy and smoothing characteristics.

B.3.2 *Fuzzy Self-Tuning Particle Swarm Optimization (FST-PSO)*

Recall that FST-PSO is a variant of the traditional Particle Swarm Optimization algorithm that incorporates fuzzy logic to dynamically adjust its parameters. This self-tuning mechanism allows the algorithm to adapt to different stages of the search process, enhancing its ability to balance exploration and exploitation. In the context of this study, FST-PSO is used to optimize the surrogate models evolved by GP.

The PSO component of the methodology involves a population of particles, each representing a potential solution in the search space. These particles move through the space based on their own experience and the experience of neighboring particles, guided by a fitness function. The FST-PSO variant enhances this process by using fuzzy logic to adjust the particles' velocity and position parameters dynamically, allowing the algorithm to respond to the landscape's characteristics and avoid premature convergence to local optima.

B.3.3 *Integration and Surrogate Model Construction*

The integration of GP and FST-PSO in this methodology creates a powerful tool for simplifying rugged landscapes. The GP component evolves a population of surrogate models, which are then optimized using FST-PSO. The surrogate model aims to approximate the original fitness landscape in a way that smooths out the ruggedness, making the search for the global optimum more tractable.

The iterative nature of this process is key to its success. As the surrogate model evolves, it becomes increasingly accurate, allowing the FST-PSO algorithm to navigate the search space more effectively. The ultimate goal is to produce a surrogate that preserves the global optimum's location while eliminating the smaller, distracting local optima that could trap traditional optimization algorithms.

B.4 EXPERIMENTAL DESIGN AND SETUP

In this section we provide detailed information about how the proposed GP-FST-PSO model was tested and validated. We selected a set of benchmark functions known for their complexity and ruggedness to evaluate the model's performance. These benchmark functions are commonly used in the optimization literature to assess the effectiveness of optimization algorithms in handling rugged, multimodal landscapes.

B.4.1 *Benchmark Functions*

The benchmark functions used in the experiments include:

- *Rastrigin Function*: this function is known for its large number of local minima, making it a challenging test for global optimization algorithms. It is often used to evaluate an algorithm's ability to navigate through a highly rugged landscape.
- *Ackley Function*: the Ackley function is characterized by a nearly flat outer region and a deep hole at the center, requiring algorithms to effectively balance exploration and exploitation to find the global optimum.
- *Schwefel Function*: this function is complex and features many local minima, making it difficult for optimization algorithms to avoid becoming trapped in suboptimal regions.

These functions were chosen because they represent different types of rugged landscapes, allowing us to test the robustness and generalizability of the GP-FST-PSO model across various scenarios.

B.4.2 *Experimental Setup*

The experiments were designed to compare the performance of the GP-FST-PSO model against traditional optimization methods, including standard PSO, Genetic Algorithms, and other surrogate modeling techniques. The key parameters of the experiments, such as population size, mutation rates, and the number of generations, were carefully selected based on prior studies and empirical tuning.

The dimensionality of the benchmark functions was varied to assess how well the GP-FST-PSO model scales with increasing problem complexity. High-dimensional spaces are particularly challenging for optimization algorithms due to the exponential increase in the search space size, commonly referred to as the "curse of dimensionality", and by testing the model across different dimensions, we could evaluate its scalability and robustness.

Multiple runs were conducted for each benchmark function to ensure the results' statistical significance. The outcomes were measured in terms of convergence speed, accuracy in finding the global optimum, and resilience to becoming trapped in local optima.

B.5 RESULTS AND ANALYSIS

The results of the experiments are presented in detail, demonstrating the effectiveness of the GP-FST-PSO model in navigating rugged fitness landscapes. The key findings are as follows.

B.5.1 *Convergence Speed and Accuracy*

The GP-FST-PSO model consistently outperformed traditional optimization methods in terms of convergence speed, meaning it reached the global optimum faster and with fewer function evaluations. This efficiency is attributed to the surrogate model's ability to smooth the fitness landscape, reducing the number of local optima that could trap the optimization process.

B.5.2 *Robustness*

The model showed strong performance even as the dimensionality of the benchmark functions increased. This robustness is particularly noteworthy, as many optimization algorithms struggle in high-dimensional spaces due to the increased complexity and number of local optima. The GP-FST-PSO model's ability to maintain performance across different dimensions highlights its scalability and adaptability.

B.5.3 *Resilience to Local Optima*

One of the most significant advantages of the GP-FST-PSO model is its resilience to local optima. The surrogate model effectively smooths the landscape, allowing the FST-PSO algorithm to bypass local optima and continue searching for the global solution. This was particularly evident in the Rastrigin and Schwefel function results, where traditional methods often became trapped in local minima.

B.5.4 *Statistical Analysis*

A thorough statistical analysis of the experimental results is also provided, comparing the mean and standard deviation of the solutions obtained by the GP-FST-PSO model against those obtained by other methods. These analyses confirm the model's superior performance, especially in challenging optimization scenarios characterized by rugged landscapes and high-dimensional spaces.

B.5.5 *Discussion of Limitations*

While the results are promising, we also discuss potential limitations of their approach. The primary concern is the computational cost associated with evolving the surrogate model using GP, which can be significant, particularly for very high-dimensional problems. Additionally, while the GP-FST-PSO model performed well across the tested benchmark functions, its performance on other types of landscapes or real-world problems may require further tuning or modifications.

B.6 CONCLUSION AND FUTURE DIRECTIONS

In conclusion, the paper presents a novel approach to optimizing rugged fitness landscapes by combining Genetic Programming with Fuzzy Self-Tuning Particle Swarm Optimization. The proposed GP-FST-PSO model constructs a surrogate landscape that smooths the original rugged fitness landscape, making it easier to find the global optimum. The experimental results demonstrate that this approach is effective in overcoming the challenges posed by local optima, high dimensionality, and the complexity of rugged landscapes.

The success of the GP-FST-PSO model across various benchmark functions suggests its potential for broader application in real-world optimization problems, particularly those that are characterized by high noise, complex landscapes, and significant computational costs. We suggest that further research could explore the integration of additional evolutionary algorithms with their approach or refine the

GP-FST-PSO model to better adapt to specific problem domains. In particular, future research could focus on several key areas:

- Application to Real-World Problems: The GP-FST-PSO model could be applied to complex real-world problems in engineering design, financial modeling, machine learning, and other fields where optimization is challenging due to the ruggedness of the fitness landscape.
- Hybridization with Other Algorithms: The model could be enhanced by integrating it with other evolutionary algorithms, such as Differential Evolution or Genetic Algorithms, to further improve its performance in specific problem domains.
- Advanced Surrogate Modeling Techniques: We propose investigating other surrogate modeling techniques, such as deep neural networks or ensemble models, to create even more accurate and robust surrogates for complex landscapes.
- Exploration-Exploitation Balance: Further research could explore more sophisticated methods for balancing exploration and exploitation in high-dimensional search spaces, potentially improving the model's ability to navigate extremely complex landscapes.

Overall, this paper provides a comprehensive framework for addressing the challenges of rugged optimization problems, offering a new perspective on the use of surrogate models and evolutionary algorithms in complex problem-solving. The GP-FST-PSO model represents a significant advancement in the field, with the potential to be applied to a wide range of challenging optimization problems.

BIBLIOGRAPHY

[1] W. Aiello, F. Chung, and L. Lu. Random evolution in massive graphs. *Handbook of massive data sets*, pages 97–122, 2002. (Cited on page 77.)

[2] J. Albert and K. C. II. A simple universal cellular automaton and its one-way and totalistic version. *Complex Syst.*, 1(1), 1987. (Cited on page 89.)

[3] R. Albert, H. Jeong, and A.-L. Barabási. Diameter of the world-wide web. *Nature*, 401(6749):130–131, 1999. (Cited on page 77.)

[4] O. Amini, F. V. Fomin, and S. Saurabh. Counting subgraphs via homomorphisms. *SIAM Journal on Discrete Mathematics*, 26(2):695–717, 2012. (Cited on page 5.)

[5] Y. Asao and S. O. Ivanov. Magnitude homology is a derived functor. *arXiv preprint arXiv:2402.14466*, 2024. (Cited on page 3.)

[6] Y. Asao and K. Izumihara. Geometric approach to graph magnitude homology. *Homology, Homotopy and Applications*, 23(1):297–310, 2021. (Cited on pages 3, 4, 59, 60, 61, and 67.)

[7] Y. Asao, Y. Hiraoka, and S. Kanazawa. Girth, magnitude homology and phase transition of diagonality. *Proceedings of the Royal Society of Edinburgh Section A: Mathematics*, pages 1–27, 2021. (Cited on page 3.)

[8] R. Ayala, L. Fernández, D. Fernández-Ternero, and J. Vilches. Discrete morse theory on graphs. *Topology and its Applications*, 156(18):3091–3100, 2009. (Cited on page 112.)

[9] A. Barbé and F. von Haeseler. Cellular automata, quasigroups and symmetries. *Aequationes mathematicae*, 62(3):211–248, 2001. (Cited on page 8.)

[10] A. D. Barbour, M. Karoński, and A. Ruciński. A central limit theorem for decomposable random variables with applications to random graphs. *Journal of Combinatorial Theory, Series B*, 47(2):125–145, 1989. (Cited on page 48.)

[11] G. Bertoni, J. Daemen, M. Peeters, and G. V. Assche. The Keccak reference, January 2011. URL http://keccak.noekeon.org/. (Cited on page 82.)

[12] A. Bhosekar and M. Ierapetritou. Advances in surrogate based modeling, feasibility analysis, and optimization: A review. *Computers & Chemical Engineering*, 108:250–267, 2018. (Cited on page 121.)

[13] A. Björner and M. Wachs. On lexicographically shellable posets. *Transactions of the American Mathematical Society*, 277(1):323–341, 1983. (Cited on pages 66 and 113.)

[14] A. Björner and M. Wachs. Shellable nonpure complexes and posets. i. *Transactions of the American mathematical society*, 348(4):1299–1327, 1996. (Cited on pages 58 and 59.)

[15] O. Bobrowski and S. Mukherjee. The topology of probability distributions on manifolds. *Probability theory and related fields*, 161(3-4):651–686, 2015. (Cited on page 112.)

[16] B. Bollobás, C. Borgs, J. T. Chayes, and O. Riordan. Directed scale-free graphs. In *SODA*, volume 3, pages 132–139. Baltimore, MD, United States, 2003. (Cited on page 77.)

[17] D. M. Boyd and N. B. Ellison. Social network sites: Definition, history, and scholarship. *Journal of computer-mediated Communication*, 13(1):210–230, 2007. (Cited on page 78.)

[18] T. Bu and D. Towsley. On distinguishing between internet power law topology generators. In *Proceedings. twenty-first annual joint conference of the ieee computer and communications societies*, volume 2, pages 638–647. IEEE, 2002. (Cited on page 77.)

[19] P. Camion, C. Carlet, P. Charpin, and N. Sendrier. On correlation-immune functions. In J. Feigenbaum, editor, *Advances in Cryptology - CRYPTO '91, 11th Annual International Cryptology Conference, Santa Barbara, California, USA, August 11-15, 1991, Proceedings*, volume 576 of *Lecture Notes in Computer Science*, pages 86–100. Springer, 1991. (Cited on page 102.)

[20] L. Caputi and C. Collari. On finite generation in magnitude (co) homology, and its torsion. *arXiv preprint arXiv:2302.06525*, 2023. (Cited on page 3.)

[21] C. Carlet. *Boolean Functions for Cryptography and Coding Theory*. Cambridge University Press, 2021. (Cited on page 100.)

[22] C. Carlet, Y. Crama, and P. L. Hammer. Boolean functions for cryptography and error-correcting codes., 2010. (Cited on page 114.)

[23] G. Cattaneo, E. Formenti, L. Margara, and J. Mazoyer. A shift-invariant metric on s^{ZZ} inducing a non-trivial tolology. In I. Prívara and P. Ruzicka, editors, *Mathematical Foundations of Computer Science 1997, 22nd International Symposium, MFCS'97, Bratislava, Slovakia, August 25-29, 1997, Proceedings*, volume 1295 of *Lecture Notes in Computer Science*, pages 179–188, , 1997. Springer. (Cited on page 117.)

[24] G. Cattaneo, E. Formenti, L. Margara, and G. Mauri. On the dynamical behavior of chaotic cellular automata. *Theor. Comput. Sci.*, 217(1):31–51, 1999. (Cited on page 117.)

[25] G. Cattaneo, M. Finelli, and L. Margara. Investigating topological chaos by elementary cellular automata dynamics. *Theoretical computer science*, 244(1-2):219–241, 2000. (Cited on page 84.)

[26] Y. Chen, M. Thurley, and M. Weyer. Understanding the complexity of induced subgraph isomorphisms. In *Automata, Languages and Programming: 35th International Colloquium, ICALP 2008, Reykjavik, Iceland, July*

7-11, 2008, Proceedings, Part I 35, pages 587–596. Springer, 2008. (Cited on page 74.)

[27] S. Cho. Quantales, persistence, and magnitude homology. *arXiv preprint arXiv:1910.02905*, 2019. (Cited on page 3.)

[28] C. J. Colbourn and J. H. Dinitz. Combinatorial designs. In *Handbook of Discrete and Combinatorial Mathematics*. 1999. (Cited on page 85.)

[29] C. Cooper and A. Frieze. A general model of web graphs. *Random Structures & Algorithms*, 22(3):311–335, 2003. (Cited on page 77.)

[30] T. W. Cusick and P. Stanica. *Cryptographic Boolean functions and applications*. Academic Press, 2017. (Cited on page 114.)

[31] J. Daemen, R. Govaerts, and J. Vandewalle. Invertible shift-invariant transformations on binary arrays. *Applied Mathematics and Computation*, 62(2):259 – 277, 1994. (Cited on page 82.)

[32] Á. M. del Rey, J. P. Mateus, and G. R. Sánchez. A secret sharing scheme based on cellular automata. *Appl. Math. Comput.*, 170(2):1356–1364, 2005. (Cited on page 92.)

[33] A. Dennunzio, E. Formenti, L. Manzoni, L. Margara, and G. Menara. A topology for p-systems with active membranes. *Journal of Membrane Computing*, 5(4):193–204, 2023. (Cited on pages 9 and 116.)

[34] R. L. Devaney. *An Introduction to Chaotic Dynamical Systems*. Addison-Wesley advanced book program. Addison-Wesley, , 1989. (Cited on page 117.)

[35] R. G. Downey and M. R. Fellows. *Parameterized complexity*. Springer Science & Business Media, 2012. (Cited on page 73.)

[36] Q. Duchemin and Y. De Castro. Random geometric graph: Some recent developments and perspectives. *High Dimensional Probability IX: The Ethereal Volume*, pages 347–392, 2023. (Cited on page 113.)

[37] M. Dyer and C. Greenhill. The complexity of counting graph homomorphisms. *Random Structures & Algorithms*, 17(3-4):260–289, 2000. (Cited on page 5.)

[38] A. D'Andrea, F. Ferri, and P. Grifoni. *An overview of methods for virtual social networks analysis*. Springer, 2010. (Cited on page 78.)

[39] K. Eloranta. Partially permutive cellular automata. *Nonlinearity*, 6(6): 1009–1023, 1993. (Cited on pages 89 and 94.)

[40] P. Erdős, A. Rényi, et al. On the evolution of random graphs. *Publ. Math. Inst. Hung. Acad. Sci*, 5(1):17–60, 1960. (Cited on pages 37, 61, and 77.)

[41] G. Erskine and J. Tuite. Large cayley graphs of small diameter. *Discrete Applied Mathematics*, 250:202–214, 2018. (Cited on page 78.)

[42] L. Euler. *Recherches sur une nouvelle espece de quarres magiques*. Zeeuwsch Genootschao, 1782. (Cited on page 85.)

[43] F. Fagnani and L. Margara. Expansivity, permutivity, and chaos for cellular automata. *Theory of Computing Systems*, 31(6):663–677, 1998. (Cited on page 84.)

[44] F. D. Farmer. Cellular homology for posets. *Math. Japon*, 23(6):79, 1978. (Cited on pages 59, 66, and 113.)

[45] X.-T. Feng, B.-R. Chen, C. Yang, H. Zhou, and X. Ding. Identification of visco-elastic models for rocks using genetic programming coupled with the modified particle swarm optimization algorithm. *International Journal of Rock Mechanics and Mining Sciences*, 43(5):789–801, 2006. (Cited on page 121.)

[46] E. Formenti and L. Mariot. Exhaustive generation of linear orthogonal cellular automata. *arXiv preprint arXiv:2307.07505*, 2023. (Cited on pages 97 and 98.)

[47] E. Formenti, K. Imai, B. Martin, and J. Yunès. Advances on random sequence generation by uniform cellular automata. In C. S. Calude, R. Freivalds, and K. Iwama, editors, *Computing with New Resources - Essays Dedicated to Jozef Gruska on the Occasion of His 80th Birthday*, volume 8808 of *Lecture Notes in Computer Science*, pages 56–70. Springer, 2014. (Cited on pages 82 and 84.)

[48] Y. Gao. The degree distribution of random k-trees. *Theoretical Computer Science*, 410(8-10):688–695, 2009. (Cited on page 77.)

[49] E. N. Gilbert. Random plane networks. *Journal of the society for industrial and applied mathematics*, 9(4):533–543, 1961. (Cited on page 49.)

[50] R. H. Gilman. Periodic behavior of linear automata. In *Dynamical Systems: Proceedings of the Special Year held at the University of Maryland, College Park, 1986–87*, pages 216–219. Springer, 2006. (Cited on page 84.)

[51] C. Giusti and G. Menara. Eulerian magnitude homology: subgraph structure and random graphs. *arXiv preprint arXiv:2403.09248*, 2024. (Cited on pages 62, 67, 69, 72, and 111.)

[52] A. Grigor'yan, Y. Lin, Y. Muranov, and S.-T. Yau. Homologies of path complexes and digraphs. *arXiv preprint arXiv:1207.2834*, 2012. (Cited on page 19.)

[53] A. Grigor'yan, R. Jimenez, Y. Muranov, and S.-T. Yau. Homology of path complexes and hypergraphs. *Topology and its Applications*, 267: 106877, 2019. (Cited on page 113.)

[54] Y. Gu. Graph magnitude homology via algebraic morse theory. *arXiv preprint arXiv:1809.07240*, 2018. (Cited on page 3.)

[55] H. Gutowitz. Cryptography with dynamical systems. In *Cellular Automata and Cooperative Systems*, pages 237–274. Springer, 1993. (Cited on pages 82 and 91.)

[56] A. S. Hedayat, N. J. A. Sloane, and J. Stufken. *Orthogonal arrays: theory and applications*. Springer Science & Business Media, 2012. (Cited on pages 85, 87, and 102.)

[57] S. M. Hedetniemi, S. T. Hedetniemi, and A. L. Liestman. A survey of gossiping and broadcasting in communication networks. *Networks*, 18 (4):319–349, 1988. (Cited on page 77.)

[58] G. A. Hedlund. Endomorphisms and automorphisms of the shift dynamical systems. *Mathematical Systems Theory*, 3(4):320–375, 1969. (Cited on pages 81, 89, 90, and 93.)

[59] P. Hell and J. Nešetřil. On the complexity of h-coloring. *Journal of Combinatorial Theory, Series B*, 48(1):92–110, 1990. (Cited on page 5.)

[60] R. Hepworth and E. Roff. The reachability homology of a directed graph. *arXiv preprint arXiv:2312.01378*, 2023. (Cited on page 113.)

[61] R. Hepworth and E. Roff. Bigraded path homology and the magnitude-path spectral sequence. *arXiv preprint arXiv:2404.06689*, 2024. (Cited on page 113.)

[62] R. Hepworth and S. Willerton. Categorifying the magnitude of a graph. *arXiv preprint arXiv:1505.04125*, 2015. (Cited on pages 2, 3, 12, 13, 14, and 18.)

[63] Y. Jin. A comprehensive survey of fitness approximation in evolutionary computation. *Soft computing*, 9(1):3–12, 2005. (Cited on page 121.)

[64] M. Kahle. Topology of random clique complexes. *Discrete mathematics*, 309(6):1658–1671, 2009. (Cited on pages 37 and 61.)

[65] M. Kahle. Random geometric complexes. *Discrete & Computational Geometry*, 45:553–573, 2011. (Cited on pages 37 and 61.)

[66] M. Kahle and E. Meckes. Erratum: Limit theorems for betti numbers of random simplicial complexes. *arXiv preprint arXiv:1501.03759*, 2015. (Cited on pages 42 and 48.)

[67] M. Kahle and E. Meckes. Limit theorems for betti numbers of random simplicial complexes. *Homology, Homotopy and Applications*, 15(1):343–374, 2013. (Cited on pages 4, 37, 42, 44, 48, and 61.)

[68] M. Kanemasa and E. Aiyoshi. Algorithm tuners for pso methods and genetic programming techniques for learning tuning rules. *IEEJ Transactions on Electrical and Electronic Engineering*, 9(4):407–414, 2014. (Cited on page 121.)

[69] R. Kaneta and M. Yoshinaga. Magnitude homology of metric spaces and order complexes. *Bulletin of the London Mathematical Society*, 53(3): 893–905, 2021. (Cited on page 3.)

[70] H. Kannan, E. Saucan, I. Roy, and A. Samal. Persistent homology of unweighted complex networks via discrete morse theory. *Scientific reports*, 9(1):13817, 2019. (Cited on page 112.)

[71] J. Kari. Reversibility and surjectivity problems of cellular automata. *Journal of Computer and System Sciences*, 48(1):149–182, 1994. (Cited on page 7.)

[72] J. Kari. Basic concepts of cellular automata. In G. Rozenberg, T. Bäck, and J. N. Kok, editors, *Handbook of Natural Computing*, pages 3–24. Springer, 2012. (Cited on page 82.)

[73] A. D. Keedwell and J. Dénes. *Latin squares and their applications*. Elsevier, 2015. (Cited on page 85.)

[74] J. Kennedy and R. Eberhart. Particle swarm optimization. In *Proceedings of ICNN'95-international conference on neural networks*, volume 4, pages 1942–1948. IEEE, 1995. (Cited on page 121.)

[75] R. Kleveland. Mixing properties of one-dimensional cellular automata. *Proceedings of the American Mathematical Society*, 125(6):1755–1766, 1997. (Cited on page 84.)

[76] J. R. Koza. Genetic programming as a means for programming computers by natural selection. *Statistics and computing*, 4(2):87–112, 1994. (Cited on pages 99, 120, and 121.)

[77] P. Kůrka. *Topological and symbolic dynamics*. Société Mathématique de France, Paris, 2003. (Cited on page 116.)

[78] R. Kumar, P. Raghavan, S. Rajagopalan, D. Sivakumar, A. Tomkins, and E. Upfal. Stochastic models for the web graph. In *Proceedings 41st Annual Symposium on Foundations of Computer Science*, pages 57–65. IEEE, 2000. (Cited on page 77.)

[79] P. Kurka. Topological dynamics of one-dimensional cellular automata. In *Mathematical basis of cellular automata*, pages 2232–2242. Springer-Verlag, 2009. (Cited on page 6.)

[80] T. Leinster. The Euler characteristic of a category. *arXiv preprint math/0610260*, 2006. (Cited on pages 2 and 12.)

[81] T. Leinster. The magnitude of metric spaces. *Documenta Mathematica*, 18:857–905, 2013. (Cited on page 2.)

[82] T. Leinster. The magnitude of a graph. In *Mathematical Proceedings of the Cambridge Philosophical Society*, volume 166, pages 247–264. Cambridge University Press, 2019. (Cited on page 12.)

[83] T. Leinster. *Entropy and diversity: the axiomatic approach*. Cambridge university press, 2021. (Cited on page 2.)

[84] T. Leinster and M. W. Meckes. The magnitude of a metric space: from category theory to geometric measure theory. *arXiv preprint arXiv:1606.00095*, 2016. (Cited on page 15.)

[85] T. Leinster and M. Shulman. Magnitude homology of enriched categories and metric spaces. *Algebraic & Geometric Topology*, 21(5):2175–2221, 2021. (Cited on pages 2 and 18.)

[86] A. Leporati and L. Mariot. 1-resiliency of bipermutive cellular automata rules. In J. Kari, M. Kutrib, and A. Malcher, editors, *Cellular Automata and Discrete Complex Systems - 19th International Workshop, AUTOMATA 2013, Gießen, Germany, September 17-19, 2013. Proceedings*, volume 8155 of *Lecture Notes in Computer Science*, pages 110–123. Springer, 2013. (Cited on pages 82 and 84.)

[87] A. Leporati, L. Mariot, et al. Cryptographic properties of bipermutive cellular automata rules. *J. Cell. Autom.*, 9(5-6):437–475, 2014. (Cited on pages 8 and 84.)

[88] A. Leporati, L. Manzoni, G. Mauri, A. E. Porreca, and C. Zandron. A survey on space complexity of P systems with active membranes. *International Journal of Advances in Engineering Sciences and Applied Mathematics*, 10(3):221–9, 2018. URL https://doi.org/10.1007/s12572-018-0227-8. (Cited on page 116.)

[89] A. Leporati, L. Manzoni, G. Mauri, A. E. Porreca, and C. Zandron. A gentle introduction to membrane systems and their computational properties. In T. Song, P. Zheng, M. L. D. Wong, and X. Wang, editors, *Bio-Inspired Computing Models and Algorithms*, pages 1–32. World Scientific, , 2019. (Cited on page 116.)

[90] J. Leskovec, J. Kleinberg, and C. Faloutsos. Graph evolution: Densification and shrinking diameters. *ACM transactions on Knowledge Discovery from Data (TKDD)*, 1(1):2–es, 2007. (Cited on page 77.)

[91] D. A. Levin and Y. Peres. *Markov chains and mixing times*, volume 107. American Mathematical Soc., 2017. (Cited on page 77.)

[92] R. Lidl and H. Niederreiter. *Introduction to finite fields and their applications*. Cambridge university press, 1994. (Cited on pages 83 and 97.)

[93] D. A. Lind, D. Lind, and B. Marcus. *An introduction to symbolic dynamics and coding*. Cambridge university press, 2021. (Cited on page 81.)

[94] H. B. Macêdo, M. J. L. Lima, and G. M. B. Oliveira. Searching for a cryptographic model based on the pre-image calculus of cellular automata. In M. M. B. R. Vellasco, M. C. P. de Souto, and J. J. F. Cerqueira, editors, *10th Brazilian Symposium on Neural Networks (SBRN 2008), Salvador, Bahia, Brazil, October 26-30, 2008*, pages 153–158. IEEE Computer Society, 2008. (Cited on pages 90 and 91.)

[95] S. Marconi and B. Chopard. Discrete physics, cellular automata and cryptography. In S. E. Yacoubi, B. Chopard, and S. Bandini, editors, *Cellular Automata, 7th International Conference on Cellular Automata, for Research and Industry, ACRI 2006, Perpignan, France, September 20-23, 2006, Proceedings*, volume 4173 of *Lecture Notes in Computer Science*, pages 617–626. Springer, 2006. (Cited on page 82.)

[96] L. Mariot and A. Leporati. Sharing secrets by computing preimages of bipermutive cellular automata. In J. Was, G. C. Sirakoulis, and S. Bandini, editors, *Cellular Automata - 11th International Conference on Cellular Automata for Research and Industry, ACRI 2014, Krakow, Poland, September 22-25, 2014. Proceedings*, volume 8751 of *Lecture Notes in Computer Science*, pages 417–426. Springer, 2014. (Cited on pages xv, 90, 91, 92, 93, and 97.)

[97] L. Mariot and A. Leporati. A cryptographic and coding-theoretic perspective on the global rules of cellular automata. *Natural Computing*, 17:487–498, 2018. (Cited on page 96.)

[98] L. Mariot and L. Manzoni. Building correlation immune functions from sets of mutually orthogonal cellular automata. In *International Workshop on Cellular Automata and Discrete Complex Systems*, pages 153–164. Springer, 2023. (Cited on page 8.)

[99] L. Mariot, L. Manzoni, and G. Menara. CA-based correlation immune functions. *in preparation*. (Cited on page 112.)

[100] L. Mariot, E. Formenti, and A. Leporati. Constructing orthogonal latin squares from linear cellular automata. *CoRR*, abs/1610.00139, 2016. URL http://arxiv.org/abs/1610.00139. (Cited on pages 94, 95, 96, and 97.)

[101] L. Mariot, E. Formenti, and A. Leporati. Enumerating orthogonal latin squares generated by bipermutive cellular automata. In *International Workshop on Cellular Automata and Discrete Complex Systems*, pages 151–164. Springer, 2017. (Cited on pages 8, 98, and 99.)

[102] L. Mariot, A. Leporati, A. Dennunzio, and E. Formenti. Computing the periods of preimages in surjective cellular automata. *Nat. Comput.*, 16(3):367–381, 2017. (Cited on pages 90 and 93.)

[103] L. Mariot, S. Picek, D. Jakobovic, and A. Leporati. Evolutionary algorithms for the design of orthogonal latin squares based on cellular automata. In *Proceedings of the Genetic and Evolutionary Computation Conference*, pages 306–313, 2017. (Cited on page 99.)

[104] L. Mariot, S. Picek, A. Leporati, and D. Jakobovic. Cellular automata based s-boxes. *Cryptogr. Commun.*, 11(1):41–62, 2019. (Cited on page 82.)

[105] L. Mariot, M. Gadouleau, E. Formenti, and A. Leporati. Mutually orthogonal latin squares based on cellular automata. *Designs, Codes and Cryptography*, 88(2):391–411, 2020. (Cited on pages 8, 95, 97, 98, and 100.)

[106] L. Mariot, D. Jakobovic, T. Bäck, and J. C. Hernandez-Castro. Artificial intelligence for the design of symmetric cryptographic primitives. In *Security and Artificial Intelligence*, pages 3–24. 2022. (Cited on page 82.)

[107] B. Martin. A walsh exploration of elementary CA rules. *J. Cell. Autom.*, 3(2):145–156, 2008. (Cited on page 84.)

[108] A. Mehrabian. Justifying the small-world phenomenon via random recursive trees. *Random Structures & Algorithms*, 50(2):201–224, 2017. (Cited on page 77.)

[109] G. Menara. On torsion in eulerian magnitude homology of Erdos-Renyi random graphs. *arXiv preprint arXiv:2409.03472*, 2024. (Cited on page 111.)

[110] G. Menara and L. Manzoni. Computing eulerian magnitude homology. *arXiv preprint arXiv:2410.10376*, 2024. (Cited on page 111.)

[111] M. Mitchell. *An introduction to genetic algorithms*. MIT press, 1998. (Cited on page 99.)

[112] C. Moore. Quasilinear cellular automata. *Physica D: Nonlinear Phenomena*, 103(1-4):100–132, 1997. (Cited on page 89.)

[113] C. Moore. Predicting nonlinear cellular automata quickly by decomposing them into linear ones. *Physica D: Nonlinear Phenomena*, 111(1-4): 27–41, 1998. (Cited on page 90.)

[114] C. Moore and T. Boykett. Commuting cellular automata. *Complex Syst.*, 11(1), 1997. (Cited on page 89.)

[115] C. Moore and A. A. Drisko. Algebraic properties of the block transformation on cellular automata. *Complex Systems*, 10(3):185–194, 1996. (Cited on page 89.)

[116] S. Nandi, B. K. Kar, and P. P. Chaudhuri. Theory and applications of cellular automata in cryptography. *IEEE Transactions on computers*, 43 (12):1346–1357, 1994. (Cited on page 8.)

[117] M. S. Nobile, P. Cazzaniga, D. Besozzi, R. Colombo, G. Mauri, and G. Pasi. Fuzzy self-tuning pso: A settings-free algorithm for global optimization. *Swarm and evolutionary computation*, 39:70–85, 2018. (Cited on pages 120 and 121.)

[118] G. Oliveira, A. Coelho, and L. Monteiro. Cellular automata cryptographic model based on bi-directional toggle rules. *International Journal of Modern Physics C*, 15(08):1061–1068, 2004. (Cited on page 91.)

[119] N. Ollinger. Universalities in cellular automata. *Handbook of Natural Computing*, pages 189–229, 2012. (Cited on page 6.)

[120] D. Orellana-Martín, L. Valencia-Cabrera, A. Riscos-Núñez, and M. J. Pérez-Jiménez. A path to computational efficiency through membrane computing. *Theoretical Computer Science*, 777:443–453, 2019. ISSN 0304-3975. URL https://doi.org/10.1016/j.tcs.2018.12.024. (Cited on page 116.)

[121] N. Otter, M. A. Porter, U. Tillmann, P. Grindrod, and H. A. Harrington. A roadmap for the computation of persistent homology. *EPJ Data Science*, 6:1–38, 2017. (Cited on page 113.)

[122] L. Pan and M. J. Pérez-Jiménez. Computational complexity of tissue-like P systems. *Journal of Complexity*, 26(3):296–315, 2010. URL https://doi.org/10.1016/j.jco.2010.03.001. (Cited on page 116.)

[123] G. Pandurangan, P. Raghavan, and E. Upfal. Using pagerank to characterize web structure. In *International computing and combinatorics conference*, pages 330–339. Springer, 2002. (Cited on page 77.)

[124] Gh. Păun. Computing with membranes. *Journal of Computer and System Sciences*, 61(1):108–143, 2000. URL https://doi.org/10.1006/jcss.1999.1693. (Cited on page 116.)

[125] Gh. Păun. P systems with active membranes: Attacking NP-complete problems. *Journal of Automata, Languages and Combinatorics*, 6(1):75–90, 2001. (Cited on page 117.)

[126] Gh. Păun, G. Rozenberg, and A. Salomaa, editors. *The Oxford Handbook of Membrane Computing*. Oxford University Press, Oxford, 2010. (Cited on page 117.)

[127] J. Pedersen. Cellular automata as algebraic systems. *Complex Systems*, 6(3):237–250, 1992. (Cited on page 89.)

[128] M. Penrose. *Random geometric graphs*, volume 5. OUP Oxford, 2003. (Cited on page 49.)

[129] M. J. Pérez-Jiménez. A computational complexity theory in membrane computing. In Gh. Păun, M. J. Pérez-Jiménez, A. Riscos-Núñez, G. Rozenberg, and A. Salomaa, editors, *Membrane Computing, 10th International Workshop, WMC 2009*, volume 5957 of *Lecture Notes in Computer Science*, pages 125–148. Springer, , 2010. (Cited on page 116.)

[130] G. Pietropolli, G. Menara, C. Mauro, et al. A genetic programming based heuristic to simplify rugged landscapes exploration. *Emerging Science Journal*, 7(4):1037–1051, 2023. (Cited on pages 10 and 120.)

[131] R. Poli, W. B. Langdon, and O. Holland. Extending particle swarm optimisation via genetic programming. In *European Conference on Genetic Programming*, pages 291–300. Springer, 2005. (Cited on page 121.)

[132] R. Sazdanovic and V. Summers. Torsion in the magnitude homology of graphs. *Journal of Homotopy and Related Structures*, 16(2):275–296, 2021. (Cited on page 3.)

[133] A. Shamir. How to share a secret. *Commun. ACM*, 22(11):612–613, 1979. (Cited on page 87.)

[134] M. A. Shereshevsky. Bipermutative cellular automata are topologically conjugate to the one-side bernoulli shift. *Random & Computational Dynamics*, 1:91–98, 1992. (Cited on page 84.)

[135] T. Siegenthaler. Decrypting a class of stream ciphers using ciphertext only. *IEEE Trans. Computers*, 34(1):81–85, 1985. (Cited on page 102.)

[136] A. R. Smith. Cellular automata and formal languages. In *11th Annual Symposium on Switching and Automata Theory (swat 1970)*, pages 216–224. IEEE, 1970. (Cited on page 6.)

[137] A. R. Smith. Simple computation-universal cellular spaces. *J. ACM*, 18(2):339–353, 1971. (Cited on page 89.)

[138] B. Song, K. Li, D. Orellana-Martín, M. J. Pérez-Jiménez, and I. Pérez-Hurtado. A survey of nature-inspired computing: Membrane computing. *ACM Comput. Surv.*, 54(1):22:1–22:31, 2022. (Cited on page 117.)

[139] P. Sosík. P systems attacking hard problems beyond NP: a survey. *Journal of Membrane Computing*, 1:198–208, 2019. URL https://doi.org/10.1007/s41965-019-00017-y. (Cited on page 117.)

[140] D. R. Stinson. *Combinatorial designs - constructions and analysis*. Springer, 2004. (Cited on pages 85, 86, and 87.)

[141] K. Sutner. De bruijn graphs and linear cellular automata. *Complex Syst.*, 5(1), 1991. (Cited on pages 83 and 90.)

[142] M. Szaban and F. Seredynski. Cryptographically strong s-boxes based on cellular automata. In *Cellular Automata, 8th International Conference on Cellular Automata for Research and Industry, ACRI 2008, Yokohama, Japan, September 23-26, 2008. Proceedings*, pages 478–485, 2008. (Cited on page 82.)

[143] Y. Tajima and M. Yoshinaga. Magnitude homology of graphs and discrete morse theory on asao-izumihara complexes. *arXiv preprint arXiv:2110.02458*, 2021. (Cited on page 3.)

[144] M. Tomassini and M. Perrenoud. Cryptography with cellular automata. *Applied Soft Computing*, 1(2):151–160, 2001. (Cited on page 8.)

[145] L. Von Bertalanffy. Problems of general system theory. *Human biology*, 23(4):302, 1951. (Cited on page 6.)

[146] J. Von Neumann, A. W. Burks, et al. Theory of self-reproducing automata. 1966. (Cited on page 6.)

[147] S. Wolfram. Statistical mechanics of cellular automata. *Reviews of modern physics*, 55(3):601, 1983. (Cited on pages 6 and 83.)

[148] S. Wolfram. Cryptography with cellular automata. In *Advances in Cryptology - CRYPTO '85, Santa Barbara, California, USA, August 18-22, 1985, Proceedings*, pages 429–432, 1985. (Cited on page 82.)

[149] S. Wolfram and M. Gad-el Hak. A new kind of science. *Appl. Mech. Rev.*, 56(2):B18–B19, 2003. (Cited on page 7.)

[150] G. Xiao and J. L. Massey. A spectral characterization of correlation-immune combining functions. *IEEE Trans. Inf. Theory*, 34(3):569–571, 1988. (Cited on page 102.)

[151] L.-H. Yen and C. W. Yu. Link probability, network coverage, and related properties of wireless ad hoc networks. In *2004 IEEE International Conference on Mobile Ad-hoc and Sensor Systems (IEEE Cat. No. 04EX975)*, pages 525–527. IEEE, 2004. (Cited on page 50.)

[152] C. W. Yu. Computing subgraph probability of random geometric graphs with applications in quantitative analysis of ad hoc networks. *IEEE Journal on Selected Areas in Communications*, 27(7):1056–1065, 2009. (Cited on pages xiii, 3, 20, 21, 49, 50, and 51.)

Trieste, Italia (Trieste, Italy)

October MMXXIV