2016

Some Extremal And Structural Problems In Graph Theory

Taylor Mitchell Short
University of South Carolina

Some extremal and structural problems in graph theory

by

Taylor Mitchell Short

Bachelor of Science
College of William and Mary 2008

Master of Science
Virginia Commonwealth University 2011

Submitted in Partial Fulfillment of the Requirements

for the Degree of Doctor of Philosophy in

Mathematics

College of Arts and Sciences

University of South Carolina

2016

Accepted by:

László Székely, Major Professor

Éva Czabarka, Committee Member

Linyuan Lu, Committee Member

Csilla Farkas, Committee Member

Lacy Ford, Senior Vice Provost and Dean of Graduate Studies

Acknowledgments

First and foremost, thanks to my advisor László Székely. I have learned so much from not only the lessons you taught me, but also from simply being around you. Thank you for your patience, support, and inspiration.

Thank you to Éva Czabarka. Your advice has been invaluable. Working closely with you has been an absolute pleasure.

To Tomislav Došlić, thank you for hosting me during the Summer 2015. Your kindness and hospitality made my visit to Croatia an experience I will never forget.

This work was partially supported by a SPARC Graduate Research Grant from the Office of the Vice President for Research at the University of South Carolina and by the NSF DMS under contract 1300547.

Abstract

This work considers three main topics. In Chapter 2, we deal with König-Egerváry graphs. We will give two new characterizations of König-Egerváry graphs as well as prove a related lower bound for the independence number of a graph. In Chapter 3, we study joint degree vectors (JDV). A problem arising from statistics is to determine the maximum number of non-zero elements of a JDV. We provide reasonable lower and upper bounds for this maximum number. Lastly, in Chapter 4 we study a problem in chemical graph theory. In particular, we characterize extremal cases for the number of maximal matchings in two linear polymers of chemical interest: the polyspiro chains and benzenoid chains. We also enumerate maximal matchings in several classes of these linear polymers and use the obtained results to determine the asymptotic behavior of these matchings.

Table of Contents

List of Figures

Chapter 1

Introduction

This work concerns the results from [45, 13, 17], each addressing problems from different areas within graph theory. In this chapter, we give a brief introduction to the topics considered.

1.1 König-Egerváry Graphs

Every n-vertex, simple graph G satisfies $\alpha(G) + \mu(G) \leq n$, where $\alpha(G)$ is the independence number of G and $\mu(G)$ is the matching number of G. We say G is a König-Egerváry graph if $\alpha(G) + \mu(G) = n$. The classical König-Egerváry theorem implies that every bipartite graph is a König-Egerváry graph, however, there are non-bipartite graphs which are König-Egerváry as well. Therefore, it is natural to question whether a characterization of König-Egerváry graphs exists.

Deming [14] was one of the first to work toward a characterization of König-Egerváry graphs, however, his result only applied to graphs with a perfect matching. Around the same time, Sterboul [47] produced an equivalent result.

Larson [34] gave a characterization of König-Egerváry graphs involving critical indepedent sets. Jarden, Levit, and Mandrescu [27, 28] questioned whether a more global characterization existed, involving unions and intersections of critical independent sets. In Chapter 2, we address their questions as well as answer a related conjecture on the independence number of a graph.

1.2 Network Models

Degree sequences and degree distributions have been subjects of study in graph theory and many other fields in the past decades. In particular, in social network analysis, they have been shown to possess a great expressive power in representing and statistically modeling networks.

Joint degree distributions are a generalization of degree distributions that deal with higher order induced subgraphs than just the nodes of the graph. Different networks may have the same degree distribution (e.g. scale-free distribution for social and biological networks) but different assortativity (social networks are assortative while biological networks are disassortative). Joint degree distributions capture the assortativity of networks, therefore they are of interest to network scientists.

A special case of the joint degree distribution is the bidegree distribution, which describes the probability that a randomly selected edge of the graph connects vertices of degree i and j. However, for this model the general conditions for the existence of the maximum likelihood estimation (MLE) are not known. As a sufficient condition, it is known that when there is only one observation of the network available, so the parameters corresponding to zeros on the bidegree vector are not estimable [43]. This motivates us to find the maximum possible number of non-zero elements on the bidegree vector of a graph, and consequently the maximum number of estimable parameters with an observed network. However, this problem seems quite challenging. In Chapter 3, we provide reasonable lower and upper bounds for this maximum number.

1.3 Chemical Graph Theory

Chemical graphs are the representation of the structural formula of a chemical compound using graph theory. The atoms of a compound are represented by vertices in a graph and the bonds between molecules represented by edges connecting the vertices (hydrogen-depleted chemical graphs have the hydrogen vertices deleted). For instance, benzene (molecular formula C_6H_6) is an important building block of carbon nanostructures and is represented by cycle graph on six vertices.

Of particular interest in chemical graphs are enumerative and structural results of matchings, sets of edges of a graph that do not share a vertex. Matchings most commonly serve as a model for bonding in chemical molecules. This idea is largely based on the work of the famous chemist August Kekulé. In 1865, Kekulé [32] claimed the structure of benzene consisted of a six-membered ring of carbon atoms with alternating single and double bonds. In the chemical graph of benzene, the edges in a perfect matching model possible locations for the double bonds. In the chemical literature, perfect matchings in graphs are also refered to as *Kekulé structures*.

There is a large and growing literature concerning perfect matchings and maximum matchings in chemical graphs, however, maximal matchings are less studied than their maximum counterparts. Maximal matchings serve as models of adsorption of dimers to a substrate or a molecule; when that process is random, it is clear that the substrate can get ''clogged'' by a number of dimers way below the theoretical maximum.

In Chapter 4, we consider the number of maximal matchings in two types of connected, plane graphs with underlying hexagonal substructure: *hexagonal chain cacti* (also known as *polyspiro chains* in the chemical literature) and *benzenoid chains*. In both types of graphs, every face is a hexagon (except the unbounded one). We characterize extremal structures for the number of maximal matchings in these graphs. We also enumerate maximal matchings in several classes of these linear polymers and use the obtained results to determine the asymptotic behavior of these matchings.

Chapter 2

König-Egerváry Theory

2.1 Introduction

The König-Egerváry theorem is a classical result in graph theory that states in a bipartite graph, the size of a maximum matching equals the cardinality of a minimum vertex cover. For a graph G, let $\alpha(G)$ be the independence number and $\mu(G)$ be the matching number. It is well-known that any n-vertex graph G satisfies the inequality

$$\alpha(G) + \mu(G) \leq n. \tag{2.1}$$

The König-Egerváry theorem is equivalent to the statement that equality holds in (2.1) for all bipartite graphs, however the converse of this statement is not true, see G_1 in Figure 2.2 for an example. In 1979, Deming [14] generalized the König-Egerváry theorem by defining a *König-Egerváry graph* to be a graph G such that equality holds in (2.1). In the same year, Sterboul [47] studied such graphs as well.

In this chapter G is a simple graph with vertex set $V(G)$, $|V(G)| = n$, and edge set $E(G)$. The set of neighbors of a vertex v is $N_G(v)$ or simply $N(v)$ if there is no possibility of ambiguity. If $X \subseteq V(G)$, then the set of neighbors of X is $N(X) = \cup_{u \in X} N(u)$, $G[X]$ is the subgraph induced by X, and X^c is the complement of the subset X. For sets $A, B \subseteq V(G)$, we use $A \setminus B$ to denote the vertices belonging to A but not B. For such disjoint A and B we let (A, B) denote the set of edges such that each edge is incident to both a vertex in A and a vertex in B.

A *matching* M is a set of pairwise non-incident edges of G. A matching of maximum cardinality is a *maximum matching* and $\mu(G)$ is the cardinality of such a maximum matching. For a set $A \subseteq V(G)$ and matching M, we say A is *saturated* by M if every vertex of A is incident to an edge in M. For two disjoint sets $A, B \subseteq V(G)$, we say there is a matching M of A into B if M is a matching of G such that every edge of M belongs to (A, B) and each vertex of A is saturated. An *M-alternating path* is a path that alternates between edges in M and those not in M. An *M-augmenting path* is an M-alternating path which begins and ends with an edge not in M.

A set $S \subseteq V(G)$ is *independent* if no two vertices from S are adjacent. An independent set of maximum cardinality is a *maximum independent set* and $\alpha(G)$ is the cardinality of such a maximum independent set. For a graph G, let $\Omega(G)$ denote the family of all its maximum independent sets, let

$$\mathrm{core}(G) = \bigcap\{S : S \in \Omega(G)\}, \qquad \text{and} \qquad \mathrm{corona}(G) = \bigcup\{S : S \in \Omega(G)\}.$$

See [36, 7, 44] for background and properties of core(G) and corona(G).

For a graph G and a set $X \subseteq V(G)$, the *difference* of X is $d(X) = |X| - |N(X)|$ and the *critical difference* $d(G)$ is $\max\{d(X) : X \subseteq V(G)\}$. Zhang [52] showed that $\max\{d(X) : X \subseteq V(G)\} = \max\{d(S) : S \subseteq V(G) \text{ is an independent set}\}$. The set X is a *critical set* if $d(X) = d(G)$. The set $S \subseteq V(G)$ a *critical independent set* if S is both a critical set and independent. A critical independent set of maximum cardinality is called a *maximum critical independent set.* Note that for some graphs the empty set is the only critical independent set, for example odd cycles or complete graphs. See [52, 9, 35, 34] for more background and properties of critical independent sets.

Finding a maximum independent set is a well-known **NP**-hard problem. Zhang [52] first showed that a critical independent set can be found in polynomial time. Butenko and Trukhanov [9] showed that every critical independent set is contained in a maximum independent set, thereby directly connecting the problem of finding a critical independent set to that of finding a maximum independent set.

Returning to consider König-Egerváry graphs, we adopt the convention that the empty graph K_0, without vertices, is a König-Egerváry graph. In [34] it was shown that König-Egerváry graphs are closely related to critical independent sets.

Theorem 2.1. *[34] A graph G is König-Egerváry if, and only if, every maximum independent set in G is critical.*

Theorem 2.2. *[34] For any graph G, there is a unique set $X \subseteq V(G)$ such that all of the following hold:*

(i) $\alpha(G) = \alpha(G[X]) + \alpha(G[X^c])$,

(ii) $G[X]$ *is a König-Egerváry graph,*

(iii) for every non-empty independent set S in $G[X^c]$, $|N(S)| \geq |S|$, and

(iv) for every maximum critical indendent set I of G, $X = I \cup N(I)$.

Larson in [35] showed that a maximum critical independent set can be found in polynomial time. So the decomposition in Theorem 2.2 of a graph G into X and X^c is also computable in polynomial time. Figure 2.1 gives an example of this decomposition, where both the sets X and X^c are non-empty. Recall, for some graphs the empty set is the only critical independent set, so for such graphs the set X would be empty. If a graph G is a König-Egerváry graph, then the set X^c would be empty. We adopt the convention that if K_0 is empty graph, then $\alpha(K_0) = 0$.

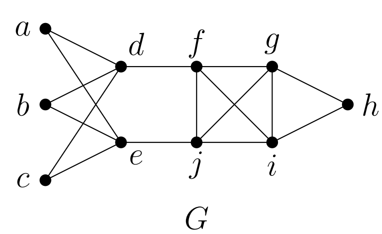

Figure 2.1 G has maximum critical independent set $I = \{a, b, c\}$. Theorem 2.2 gives that $X = \{a, b, c, d, e\}$ and $X^c = \{f, g, h, i, j\}$.

In [37, 28] the following concepts were introduced: for a graph G,

$$\ker(G) = \bigcap\{S : S \text{ is a critical independent set in } G\},$$
$$\text{diadem}(G) = \bigcup\{S : S \text{ is a critical independent set in } G\}, \text{ and}$$
$$\text{nucleus}(G) = \bigcap\{S : S \text{ is a maximum critical independent set in } G\}.$$

However, the following result due to Larson allows us to use a more suitable definition for $\text{diadem}(G)$.

Theorem 2.3. *[35] Each critical independent set is contained in some maximum critical independent set.*

For the remainder of this paper we define

$$\text{diadem}(G) = \bigcup\{S : S \text{ is a maximum critical independent set in } G\}.$$

Note that if G is a graph where the empty set is the only critical indepedent set (including the case $G = K_0$, the empty graph), then $\ker(G), \text{diadem}(G)$, and $\text{nucleus}(G)$ are all empty. See Figure 2.2 for examples of the sets $\ker(G)$, $\text{diadem}(G)$, and $\text{nucleus}(G)$.

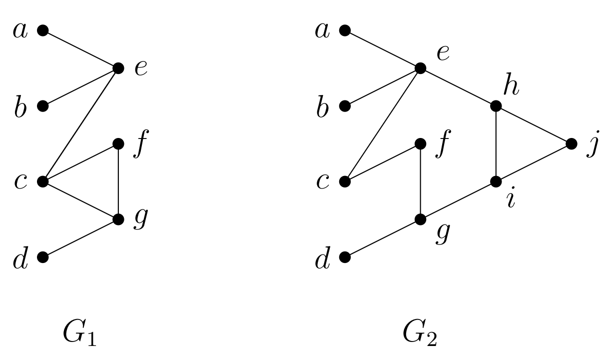

Figure 2.2 G_1 is a König-Egerváry graph with
$\ker(G_1) = \{a, b\} \subsetneq \text{core}(G_1) = \text{nucleus}(G_1) = \{a, b, d\}$ and
$\text{diadem}(G_1) = \text{corona}(G_1) = \{a, b, c, d, f\}$. G_2 is not a König-Egerváry graph and has $\ker(G_2) = \text{core}(G_2) = \{a, b\} \subsetneq \text{nucleus}(G_2) = \{a, b, d\}$ and
$\text{diadem}(G_2) = \{a, b, c, d, f\} \subsetneq \text{corona}(G) = \{a, b, c, d, f, g, h, i, j\}$.

In [27, 28], the following necessary conditions for König-Egerváry graphs were given:

Theorem 2.4. *[27] If G is a König-Egerváry graph, then*

(i) $\mathrm{diadem}(G) = \mathrm{corona}(G)$*, and*

(ii) $|\ker(G)| + |\mathrm{diadem}(G)| \leq 2\alpha(G)$.

Theorem 2.5. *[28] If G is a König-Egerváry graph, then* $|\mathrm{nucleus}(G)|+|\mathrm{diadem}(G)| = 2\alpha(G)$.

In [27] it was conjectured that condition (i) of Theorem 2.4 is sufficient for König-Egerváry graphs and in [28] it was conjectured the necessary condition in Theorem 2.5 is also sufficient. The purpose of this paper is to affirm these conjectures by proving the following new characterizations of König-Egerváry graphs.

Theorem 2.6. *For a graph G, the following are equivalent:*

(i) *G is a König-Egerváry graph,*

(ii) $\mathrm{diadem}(G) = \mathrm{corona}(G)$*, and*

(iii) $|\mathrm{diadem}(G)| + |\mathrm{nucleus}(G)| = 2\alpha(G)$.

The paper [27] gives an upper bound for $\alpha(G)$ in terms of unions and intersections of maximum independent sets, proving

$$2\alpha(G) \leq |\mathrm{core}(G)| + |\mathrm{corona}(G)|$$

for any graph G. It is natural to ask whether a similar lower bound for $\alpha(G)$ can be formulated in terms of unions and intersections of critical independent sets. Jarden, Levit, and Mandrescu in [27] conjectured that for any graph G, the inequality $|\ker(G)| + |\mathrm{diadem}(G)| \leq 2\alpha(G)$ always holds. We will prove a slightly stronger statement. By Theorem 2.3 we see that $\ker(G) \subseteq \mathrm{nucleus}(G)$ holds implying that $|\ker(G)| + |\mathrm{diadem}(G)| \leq |\mathrm{nucleus}(G)| + |\mathrm{diadem}(G)|$. In section 2.4 we will prove the following statement, resolving the cited conjecture:

Theorem 2.7. *For any graph G,*

$$|\operatorname{nucleus}(G)| + |\operatorname{diadem}(G)| \leq 2\alpha(G).$$

It would be interesting to know whether the sets $\operatorname{nucleus}(G)$ and $\operatorname{diadem}(G)$, or their sizes, can be computed in polynomial time.

2.2 Some structural lemmas

Here we prove several lemmas which will be needed in our proofs. Our results hinge upon the structure of the set X as described in Theorem 2.2.

Lemma 2.8. *Let I be a maximum critical independent set in G and set $X = I \cup N(I)$. Then $\operatorname{diadem}(G) \cup N(\operatorname{diadem}(G)) = X$.*

Proof. By Theorem 2.2 the set X is unique in G, that is, for any maximum critical independent set S, $X = S \cup N(S)$. Then $\operatorname{diadem}(G) = X$ follows by definition. □

Lemma 2.9. *Let I be a maximum critical independent set in G and set $X = I \cup N(I)$. Then $\operatorname{diadem}(G) \subseteq \operatorname{diadem}(G[X])$ and $\operatorname{nucleus}(G[X]) \subseteq \operatorname{nucleus}(G)$.*

Proof. Let S be a maximum critical independent set in G. Using Theorem 2.2 we see that S is a maximum independent set in $G[X]$ and also $G[X]$ is a König-Egerváry graph. Then Theorem 2.1 gives that S must also be critical in $G[X]$, which implies that $\operatorname{diadem}(G) \subseteq \operatorname{diadem}(G[X])$.

Now let $v \in \operatorname{nucleus}(G[X])$. Then v belongs to every maximum critical indepedent set in $G[X]$. As remarked above, since every maximum critical independent set in G is also a maximum critical independent set in $G[X]$, then v belongs to every maximum critical independent set in G. This shows that $v \in \operatorname{nucleus}(G)$ and $\operatorname{nucleus}(G[X]) \subseteq \operatorname{nucleus}(G)$ follows. □

Lemma 2.10. *Suppose I is a non-empty maximum critical independent set in G, set $X = I \cup N(I)$, let $A = \operatorname{nucleus}(G) \setminus \operatorname{nucleus}(G[X])$, and let S be a maximum* *independent set in $G[X]$. For $S' \subseteq S \cap N(A)$, if there exists $A' \subseteq A$ such that $N(A') \cap S \subseteq S'$, then $|S'| \geq |A'|$.*

Proof. For $S' \subseteq S \cap N(A)$ suppose such an A' exists. For sake of contradiction, suppose that $|S'| < |A'|$. Since $A' \subseteq \text{nucleus}(G)$, then A' is an independent set. Also since $A' \subseteq \text{nucleus}(G) \subseteq \text{diadem}(G)$, by Lemma 2.8 we have $A' \subseteq X$. Furthermore, since $N(A') \cap S \subseteq S'$ then $A' \cup (S \setminus S')$ is an independent set in $G[X]$. Now by assumption $|S'| < |A'|$, so $A' \cup (S \setminus S')$ is an independent set in $G[X]$ larger than S, which cannot happen. Therefore we must have $|S'| \geq |A'|$ as desired. $\square$

Lemma 2.11. *Let I be a maximum critical independent set in G and set $X = I \cup N(I)$. Then*

$$|\,\text{nucleus}(G)| + |\,\text{diadem}(G)| \leq |\,\text{nucleus}(G[X])| + |\,\text{diadem}(G[X])|.$$

Proof. First note that if the set X is empty, then by Lemma 2.8 both sides of the inequality are zero. So let us assume that X is non-empty. Now consider the set $A = \text{nucleus}(G) \setminus \text{nucleus}(G[X])$. If this independent set is empty, then $\text{nucleus}(G) = \text{nucleus}(G[X])$ and there is nothing to prove since $\text{diadem}(G) \subseteq \text{diadem}(G[X])$ holds by Lemma 2.9. If A is non-empty, for each $v \in A$ there is some maximum independent set S of $G[X]$ which doesn't contain v. Since S is a maximum independent set there exists $u \in N(v) \cap S$. Since $v \in \text{nucleus}(G)$, then u does not belong to any maximum critical independent set in G. Recall by Theorem 2.2 (ii) $G[X]$ is a König-Egerváry graph, so Theorem 2.1 gives that S is a maximum critical independent set in $G[X]$. It follows that $u \in \text{diadem}(G[X]) \setminus \text{diadem}(G)$, which shows each vertex in A is adjacent to at least one vertex in $\text{diadem}(G[X]) \setminus \text{diadem}(G)$.

Now we will show there is a maximum matching from A into $\text{diadem}(G[X]) \setminus \text{diadem}(G)$ with size $|A|$. For sake of contradiction, suppose such a matching M has less than $|A|$ edges. Then there exists some vertex $v \in A$ not saturated by M. By the above, v is adjacent to some vertex $u \in \text{diadem}(G[X]) \setminus \text{diadem}(G)$. Since M is maximum, u is matched to some vertex $w \in A$ under M. Now let S be a maximum independent set of $G[X]$ containing u. We now restrict ourselves to the subgraph induced by the edges $(A \cap N(S), S \cap N(A))$, noting this subgraph is bipartite since both $A \cap N(S)$ and $S \cap N(A)$ are independent. In this subgraph, consider the set $\mathcal{P}$ of all M-alternating paths starting with the edge vu. Note that all such paths must start with the vertices v, u, then w. Also, such paths must end at either a matched vertex in $A \cap N(S)$ or an unmatched vertex in $S \cap N(A)$.

We wish to show that there is some alternating path ending at an unmatched vertex in $S \cap N(A)$. For sake of contradiction, suppose all alternating paths end at a matched vertex in $A \cap N(S)$ and let $V(\mathcal{P})$ denote the union of all vertices belonging to such an alternating path. We aim to show this scenario contradicts Lemma 2.10. Now clearly we must have $N(V(\mathcal{P}) \cap A) \cap S \subseteq V(\mathcal{P}) \cap S$, else we could extend an alternating path to any vertex in $(N(V(\mathcal{P}) \cap A) \cap S) \setminus (V(\mathcal{P}) \cap S)$. Also, since all paths in $\mathcal{P}$ end at a matched vertex in $A \cap N(S)$, then every vertex of $V(\mathcal{P}) \cap S$ is matched under M, and such a situation should look as in Figure 2.3.

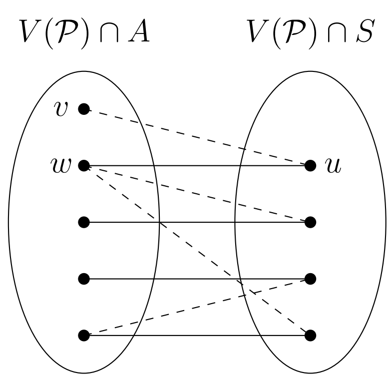

Figure 2.3 What the M-alternating paths could look like between $V(\mathcal{P}) \cap A$ and $V(\mathcal{P}) \cap S$, where solid lines represent matched edges in M and dotted lines represent the unmatched edges.

From this it follows that $|V(\mathcal{P}) \cap S| < |V(\mathcal{P}) \cap A|$. The previous statements exactly contradict Lemma 2.10, so there is some alternating path P ending at an unmatched vertex $x \in S \cap N(A)$. This means that P is an M-augmenting path. A well-known theorem in graph theory states that a matching is maximum in G if, and only if, there is no augmenting path [50]. So P being an M-augmenting path contradicts our assumption that M is a maximum matching.

Therefore there is a matching M from A into $\operatorname{diadem}(G[X]) \setminus \operatorname{diadem}(G)$. This matching implies that $|\operatorname{nucleus}(G) \setminus \operatorname{nucleus}(G[X])| \leq |\operatorname{diadem}(G[X]) \setminus \operatorname{diadem}(G)|$. Since both $\operatorname{nucleus}(G[X]) \subseteq \operatorname{nucleus}(G)$ and $\operatorname{diadem}(G) \subseteq \operatorname{diadem}(G[X])$ by Lemma 2.9, the lemma follows. $\square$

2.3 New characterizations of König-Egerváry graphs

Proof (of Theorem 2.6). First we prove $(ii) \Rightarrow (i)$. Suppose that $\operatorname{diadem}(G) = \operatorname{corona}(G)$ holds and let I be a maximum critical independent set with $X = I \cup N(I)$. We will use the decomposition in Theorem 2.2 to show that X^c must be empty and hence, $G = G[X]$ is a König-Egerváry graph. By Lemma 2.8 we have $\operatorname{corona}(G) = \operatorname{diadem}(G) \subseteq X$, in other words every maximum independent set in G is contained in X. This implies that $|I| = \alpha(G[X]) = \alpha(G)$. Now by Theorem 2.2 (i), $\alpha(G) = \alpha(G[X]) + \alpha(G[X^c])$ showing that we must have $\alpha(G[X^c]) = 0$. Now clearly the result follows, since $\alpha(G[X^c]) = 0$ implies that X^c must be empty.

To prove $(iii) \Rightarrow (i)$, again we will use the decomposition in Theorem 2.2 to show that X^c must be empty and hence, G is a König-Egerváry graph. So suppose that $|\operatorname{diadem}(G)| + |\operatorname{nucleus}(G)| = 2\alpha(G)$ and let I be a maximum critical independent set in G with $X = I \cup N(I)$. Lemma 2.11 implies that

$$2\alpha(G) = |\operatorname{diadem}(G)| + |\operatorname{nucleus}(G)| \leq |\operatorname{diadem}(G[X])| + |\operatorname{nucleus}(G[X])|.$$

Theorem 2.2 (ii) gives that $G[X]$ is König-Egerváry , so by Corollary 2.5 we have $|\operatorname{diadem}(G[X])| + |\operatorname{nucleus}(G[X])| = 2\alpha(G[X])$ implying that $\alpha(G) \leq \alpha(G[X])$. It follows by Theorem 2.2 (i) we must have $\alpha(G) = \alpha(G[X])$, so again we know that $\alpha(G[X^c]) = 0$ which finishes this part of the proof.

The implications $(i) \Rightarrow (ii)$ and $(i) \Rightarrow (iii)$ are given in Theorem 2.4 and in Theorem 2.5. $\square$

2.4 A bound on $\alpha(G)$

Proof (of Theorem 2.7). Let I be a maximum critical independent set in G and $X = I \cup N(I)$. By Theorem 2.2 (ii), $G[X]$ is a König-Egerváry graph so by Theorem 2.5 we have

$$|\operatorname{nucleus}(G[X])| + |\operatorname{diadem}(G[X])| = 2\alpha(G[X]) \leq 2\alpha(G).$$

Now by Lemma 2.11 we must have

$$|\operatorname{nucleus}(G)| + |\operatorname{diadem}(G)| \leq |\operatorname{nucleus}(G[X])| + |\operatorname{diadem}(G[X])|$$

and the theorem follows. $\square$

Combining Theorem 2.7 and the inequality $2\alpha(G) \leq |\operatorname{core}(G)| + |\operatorname{corona}(G)|$ proven in [27], the following corollary is immediate.

Corollary 2.12. *For any graph G,*

$$|\operatorname{nucleus}(G)| + |\operatorname{diadem}(G)| \leq 2\alpha(G) \leq |\operatorname{core}(G)| + |\operatorname{corona}(G)|.$$

These upper and lower bounds are quite interesting. The fact that every critical independent set is contained in a maximum independent set implies that $\operatorname{diadem}(G) \subseteq \operatorname{corona}(G)$ for all graphs G. However, the graph G_2 in Figure 2.2 has $\operatorname{core}(G_2) \subsetneq \operatorname{nucleus}(G_2)$ while the graph G in Figure 2.1 has $\operatorname{nucleus}(G) = \{a, b, c\} \subsetneq \operatorname{core}(G) = \{a, b, c, h\}$.

Chapter 3

Joint Degree Vectors

3.1 Introduction

Degree sequences and *degree distributions* have been subjects of study in graph theory and many other fields in the past decades. In particular, in social network analysis, they have been shown to possess a great expressive power in representing and statistically modeling networks; see, e.g., [40] and [25].

Generally in this context, models are in exponential family form [4], and hence known as *exponential random graph models* (ERGMs) [23, 49]. The degree sequences and distributions, act as the sufficient statistics of ERGMs, i.e. the only information that the ERGM gathers from an observed network. When the sufficient statistic is the degree sequence of a network, the corresponding ERGM is known as the *beta model*, properties of which have been extensively studied in the recent literature; see [6], [10], and [42]. Degree distributions have also been used as sufficient statistics; see [43].

Joint degree distributions are a generalization of degree distributions that deal with higher order induced subgraphs than nodes of the graph. They are usually represented in vector form and have been used as a class of network statistics. The graphs generated from such distributions are called *dK-graphs* in the computer science literature, where d indicates the number of nodes of the concerned subgraphs. The class of dK-graphs was originally proposed by [39], formulated as a means to capture increasingly refined properties of networks in a hierarchical manner based on higher order interactions among node degrees (see, e.g., [15]).

For the case of $d = 2$, the sufficient statistic of the ERGM is the special case of the joint degree distribution, known as the *bidegree distribution*. This model has been formalized in [43]. In essence, the bidegree distribution describes the probability that a randomly selected edge of the graph connects vertices of degree k and l.

However, model selection for this model is quite challenging, and, as will be discussed in the next section, the general conditions for the existence of the *maximum likelihood estimation* (MLE) are not known, and seem difficult to obtain. As a sufficient condition, it is known that when there is only one observation of the network available, the parameters corresponding to zeros on the bidegree vector are not estimable [43]. This motivates us to find the maximum possible number of non-zero elements on the bidegree vector of a graph, and consequently the maximum number of estimable parameters with an observed network.

On the other hand, [41], [2] and [46] introduced the *joint degree matrix* (JDM), which is a non-normalized version of the bidegree vector in matrix form, i.e. the elements of JDM represents the exact number of edges between a pair of vertices. Conditions for a given matrix to be the JDM of a graph were provided in [41], [46] and [12].

Finding the maximum possible number of non-zero elements of a JDM for a fixed number of nodes seems quite challenging. In this paper we shall use the conditions in [12] as well as other methods and constructions, in order to come up with reasonable lower and upper bounds for this value.

The structure of the paper is as follows: In the next section, we provide basic graph theoretical as well as statistical definitions and preliminary results needed in this paper. In Section 3.3, we provide a lower bound for the maximum possible number of non-zero elements of a JDM by constructing a family of graphs that reaches this bound. In Section 3.4, we use two different approaches to present upper bounds for this desired value.

3.2 Definitions and preliminary results

In this chapter we consider simple graphs without isolated vertices. Let $G = (V, E)$ be such an n-vertex graph and for $1 \leq i \leq n-1$ let V_i be the set of vertices of degree i. The *joint degree vector* (JDV) $s(G) = (j_{11}(G), j_{12}(G), \ldots, j_{n-1,n-1}(G))$ of the graph G is a $\binom{n}{2}$ length vector where for all $1 \leq i \leq k \leq n-1$ we have $j_{ik} = |\{xy \in E(G) : x \in V_i, y \in V_k\}|$. If, for some vector $\mathbf{m}$ there exists a graph G such that $\mathbf{s}(G) = \mathbf{m}$, then m is called a graphical JDV. Note that the degree sequence of a graph is determined by its JDV in that

$$|V_i| = \frac{1}{i}\left(\sum_{k=1}^{i} j_{ki} + \sum_{k=i}^{n-1} j_{ik}\right).$$

The following characterization for a vector $\mathbf{m}$ with integer entries to be a graphical JDV is proved by [41], [46], and [12]. As it provides simple necesssary and sufficient conditions for a vector to be realized as a graphical JDV, we call the result an Erdös-Gallai type theorem.

Theorem 3.1. *(Erdös-Gallai type theorem for a JDV) The* $\binom{n}{2}$ *size vector* $\mathbf{m} = (m_{11}, m_{12}, \ldots, m_{n-1,n-1})$ *is a JDV of some graph* G *if and only if the following holds:*

(i) for all i: $n_i := \dfrac{1}{i}\left(\sum_{k=1}^{i} m_{ik} + \sum_{k=i}^{n-1} m_{ik}\right)$ *is an integer,*

(ii) for all i: $m_{ii} \leq \dbinom{n_i}{2}$,

(iii) for all $i < k$: $m_{ik} \leq n_i n_k$.

Moreover, n_i *gives the number of vertices of degree* i *in the graph* G.

In an *exponential random graph model* (ERGM), the node set I is finite and the probability of observing a network G can be written as

$$P(G) = \exp\{\sum_{i \in I} s_i(G)\theta_i - \psi(\theta)\}, \tag{3.1}$$

where $s_i(G)$ are *canonical sufficient statistics*, which capture some important feature of G, and $\psi(\theta)$ is the *normalizing constant*, which ensures that probabilities add to 1 when summing over all possible networks.

The model is in exponential family form. Hence, the likelihood function $l(\theta) = P(g_1, \ldots, g_m)$, for generic observed networks $g_1, \ldots, g_m$, is concave, and, therefore, has a unique maximum if it exists. For distributions in exponential families, the following result [4, 8] provides an equivalent condition for the existence of the MLE.

Suppose that there are networks $G_1, \ldots, G_m$ observed. The *average observed sufficient statistic* $\bar{s}$ is a vector whose elements are the average of the corresponding elements of sufficient statistics (of dimension d), i.e. $\bar{s}_i = \frac{1}{m}\sum_{j=1}^{m} s_i(G_j)$, $1 \leq i \leq d$. We also define *the model polytope* to be the convex hull of all the points in a d-dimensional space that correspond to the sufficient statistics of all graphs with n nodes. We then have the following:

Proposition 3.2. *For an ERGM, the MLE exists if and only if the average observed sufficient statistic lies on the interior of the model polytope.*

In addition, for exactly those elements that lie on a surface that contains an extreme point corresponding to an element i, the corresponding parameter θ_i is not estimable. In network analysis, there is usually only one network G observed, and therefore, the average observed sufficient statistic is simply $s(G)$.

In the so-called 2K-model, an element of sufficient statistic in (3.1) is the bidegree vector $s(G) = (j_{11}(G), j_{12}(G), \ldots, j_{n-1,n}(G), j_{nn}(G))$, where the length of s is $\binom{n}{2}$. It is easy to show that [43] if $s_i(G) = 0$ then θ_i is not estimable. It is also easy to observe that for every graph, there are always some elements of the bidegree vector that are zero. In the next sections, we investigate how many elements of the bidegree vector are always zero.

3.3 Lower bound construction

Let H_n denote an n-vertex graph with vertex set $V(H_n) = \{v_1, v_2, \ldots, v_n\}$ and edge set $E(H_n) = \{v_i v_j : i + j > n \text{ and } i \neq j\}$. This graph, which is known as the *half graph*, has degree sequence $n-1, n-2, \ldots, \left\lfloor \frac{n}{2} \right\rfloor, \left\lfloor \frac{n}{2} \right\rfloor, \ldots, 2, 1$. See Figure 3.1 for an example half graph, H_7. Since it is obvious that nodes with degrees 0 and $n-1$ cannot co-exist in the same graph, the half graph attains the maximum number of distinct degrees.

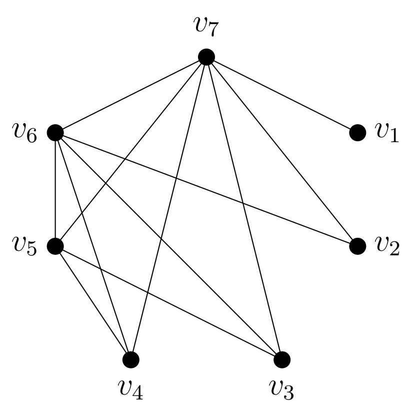

Figure 3.1 The half graph on seven vertices, H_7.

For a graph G, denote by $A(G)$ the number of non-zero elements in the JDV of G. By routine counting, we see that $A(H_n) = n(n-2)/4 + 1$ if n is even and $A(H_n) = (n-1)^2/4$ if n is odd. Hence we have that

$$\lim_{n\to\infty} \frac{A(H_n)}{\binom{n}{2}} = \frac{1}{2},$$

so about half the elements of the JDV of the half graph are non-zero. However, there are constructions which achieve a higher number of non-zero elements in the JDV than the half graph. Consider the graph H_n with $n \geq 7$ odd. If one connects the degree 1 vertex to one of the vertices with degree $(n-1)/2$, the JDV element $j_{1,n-1}$ is lost, but one gains new elements $j_{2,(n+1)/2}$ and $j_{(n+1)/2,(n+1)/2}$. We found even better such constructions but all of these only improve $A(H_n)$ by a linear additive term.

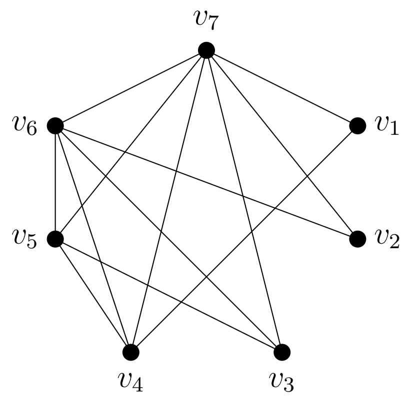

Figure 3.2 A graph on seven vertices that achieves a higher number of non-zero elements in the JDV than the half graph, H_7.

3.4 Two upper bounds

In this section, we provide two upper bounds that provide numerically very close upper bounds, but use entirely different methods. Although we tried, we were unable to combine these two proof techniques. We think that it is instructive to show both of them. We note here that it was Aaron Dutle who first gave a non-trivial upper bound $(1-1/e+o(1))n^2 = 0.63212...n^2$ for the number of non-zero entries in an $n \times n$ JDM. His proof was somewhat similar to Subsection 3.4 but missed the symmetry that is key to that subsection.

Continuous optimization

Let $P_n = \{(i,j) \in \mathbb{N}^2 : 1 \le i \le j \le n-1\}$. For any graph G, let $A(G) := \{(k_1,k_2) \in P_n : j_{k_1k_2}(G) > 0\}$. The following identity is a simple consequence of the characterization of JDM matrices was written in this form in [43]:

Proposition 3.3. *For any graph G,*

$$\sum_{(k_1,k_2)\in A(G)} \frac{k_1+k_2}{k_1k_2} n_{k_1k_2}(G) = n - n_0(G), \tag{3.2}$$

where $n_0(G)$ is the number of isolated nodes in G.

To see this, by Theorem 3.1 part (i) we have that

$$\begin{aligned} n - n_0(G) &= \sum_{i=1}^{n-1} n_i(G) = \sum_{i=1}^{n-1} \frac{1}{i} \left(\sum_{k=1}^{i} j_{ki}(G) + \sum_{k=i}^{n-1} j_{ik}(G) \right) \\ &= \sum_{(k_1,k_2)\in A} \left(\frac{1}{k_2} + \frac{1}{k_2} \right) j_{k_1k_2}(G) = \sum_{(k_1,k_2)\in A} \frac{k_1+k_2}{k_1k_2} j_{k_1k_2}(G) \end{aligned}$$

In the end, we are only interested in graphs without isolated nodes, since isolated nodes do not contribute any edges. If a graph has more than one isolated node, then we can connect the isolated nodes among each other. This can only increase the support of the vector of bidgrees. Thus, there are optimal graphs with at most one isolated node. On the other hand, connecting a single isolated node to a node of degree $d < n$ can reduce the support of the vector of bidegrees by at most d. As we have seen in Section 3.3, there are graphs where the support is of size $O(n^2)$. Hence, asymptotically, we can ignore single isolated nodes.

Corollary 3.4. *For any graph G,*

$$\sum_{(k_1,k_2)\in A(G)} \frac{k_1+k_2}{k_1k_2} \leq \sum_{(k_1,k_2)\in A(G)} \frac{k_1+k_2}{k_1k_2} n_{k_1k_2}(G) \leq n.$$

The original optimization problem can be formulated as follows:

- *Maximize $|A(G)|$ among all graphs G with n vertices.*

Using the corollary, we relax this optimization problem and study the following problem, which we call the *discrete relaxation*:

- *Maximize the cardinality among all subsets $A \subseteq P_n$ under the constraint $\sum_{(k_1,k_2)\in A} \frac{k_1+k_2}{k_1k_2} \leq n$.*

For any n, the cardinality of a subset that solves the is an upper bound for the original optimization problem.

The discrete relaxation can be solved on a computer as follows: First, compute all values $(k_1 + k_2)/(k_1k_2)$ on P_n. Second, order the values. Third, start adding them up as long as the sum does not exceed n. Finally, count the number of elements that have been added. To compare the values for different n, let α_n be the cardinality of a solution A of the discrete relaxation divided by $\binom{n}{2}$, the cardinality of P_n. The values of α_n are plotted in Figure 3.3. As a function of n, the optimum α_n decreases roughly (though not strictly) and reaches values below 0.56 for large n.

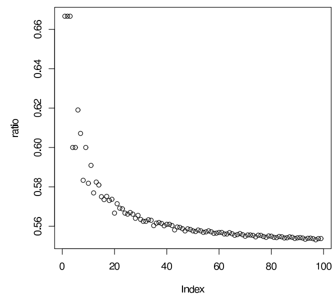

Figure 3.3 An upper bound for the ratio of maximum non-zero elements of the bi-degree vector to its length.

The limit for $n \to \infty$ can be computed by approximating the discrete relaxation by the following optimization problem, which we call the *continuous relaxation*:

- *Maximize* $\frac{|A'|}{(n-1)^2}$ *among all subsets* $A' \subseteq [1, n] \times [1, n]$ *that satisfy* $\iint_{A'} \frac{1}{x} \mathrm{d}x\mathrm{d}y \leq n$.

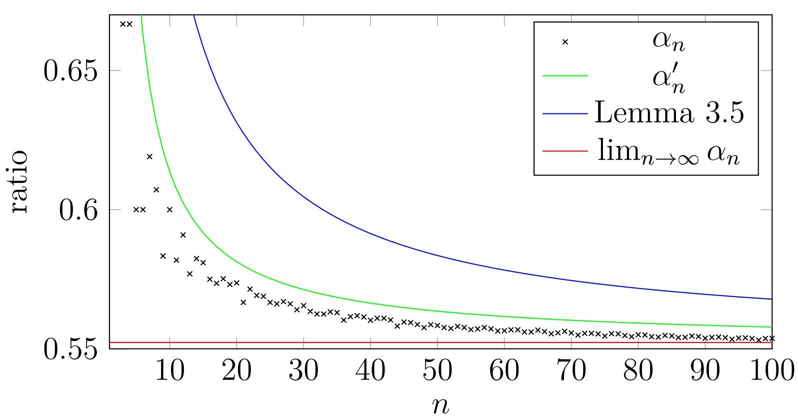

Figure 3.4 The solution of the discrete and continuous relaxation. The blue line plots the upper bound on α_n from Lemma 3.5. The red line is the limit for large n of α_n and α'_n.

Let α'_n be the maximum of the continuous relaxation.

Lemma 3.5. $\alpha_n \leq \frac{n-1}{n}\alpha'_n + \frac{1}{n}$.

Proof. To each $(i,j) \in P_n$ associate the two squares $A_{i,j} := [i, i+1) \times [j, j+1)$ and $A_{j,i} := [j, j+1) \times [i, i+1)$. For $A \subseteq P_n$ let $A'' = \bigcup_{(i,j)\in A} A_{i,j}$ and $A' = A'' \cup \bigcup_{(i,j)\in A} A_{j,i}$. Then

$$\sum_{(i,j)\in A} \frac{i+j}{i\cdot j} \geq \sum_{(i,j)\in A} \iint_{A_{i,j}} \frac{x+y}{x\cdot y}\mathrm{d}x\mathrm{d}y = \iint_{A''} \frac{x+y}{x\cdot y}\mathrm{d}x\mathrm{d}y$$
$$\geq \frac{1}{2}\iint_{A'} \frac{x+y}{x\cdot y}\mathrm{d}x\mathrm{d}y = \iint_{A'} \frac{1}{x}\mathrm{d}x\mathrm{d}y.$$

Here, the first inequality follows from the fact that the maximum of $\frac{x+y}{xy} = \frac{1}{x} + \frac{1}{y}$ over $A_{i,j}$ is at $(x,y) = (i,j)$. The second inequality follows by not double-counting the set $A_d := \bigcup_{(i,i)\in A} A_{i,i}$ corresponding to the diagonal elements of A. The last equality follows since $\frac{x+y}{x\cdot y} = \frac{1}{x} + \frac{1}{y}$ and since A' is symmetric along the diagonal. Therefore, if A is feasible for the discrete relaxation, then A' is feasible solution for the continuous relaxation. Now,

$$|A| = |A''| = \frac{1}{2}(|A'| + |A_d|) \leq \frac{|A'|}{2} + \frac{n-1}{2},$$

and so

$$\alpha_n \leq \frac{(n-1)^2}{2\binom{n}{2}}\alpha'_n + \frac{n-1}{2\binom{n}{2}}. \qquad \square$$

Corollary 3.6. $\limsup_{n\to\infty} \alpha_n \leq \limsup_{n\to\infty} \alpha'_n$.

It is not difficult to see that, actually, $\lim_{n\to\infty} \alpha_n = \lim_{n\to\infty} \alpha'_n$. Figure 3.3 shows that the upper bound from Lemma 3.5 is not very close and suggests that $\alpha_n \leq \alpha'_n$; at least for $n \leq 100$.

Next, we want to solve the continuous relaxation. The idea is the following: As the set A' it is advantageous to choose a sublevel set of the function $\frac{x+y}{x\cdot y}$. For $c > 0$ let

$$A_c := \left\{(x,y) \in [1,n]^2 : \frac{x+y}{x\cdot y} \leq c\right\}.$$

Let

$$y_c(x) = \frac{1}{c - \frac{1}{x}} = \frac{x}{xc-1}, \qquad x_1(c) = \frac{1}{c - \frac{1}{n}} = \frac{n}{nc-1}.$$

Lemma 3.7. $A_c = \Big\{(x,y) \in [1,n]^2 : x_1(c) \leq x \leq n,\ y_c(x) \leq y \leq n\Big\}$. *In particular, $A_c \neq \emptyset$ if and only if $nc \geq 2$.*

Proof. If $x < x_1(c)$ and $1 \leq y \leq n$, then $\frac{1}{x} + \frac{1}{y} > c - \frac{1}{n} + \frac{1}{n} = c$. If $x_1(c) \leq x \leq n$ and $1 \leq y < y_c(x)$, then $\frac{1}{x} + \frac{1}{y} > c - \frac{1}{x} + \frac{1}{x} = c$. For the second statement observe that $x_1(c) \leq n$ if and only if $nc \geq 2$. Similarly, $y_c(x) \leq n$ if and only if $x \geq x_1(c)$. $\square$

Lemma 3.8. *Assume that c is such that $x_1(c) \geq 1$. Then $y_c(x) \geq 1$ for all $x \in [1,n]$.*

Proof. $y_c(x)$ decreases monotonically with x. Therefore, $y_c(x) \geq y_c(n) = x_1(c)$ for all $x \in [1,n]$. $\square$

Lemma 3.9. *Let $n \geq 3$. The set A_c is feasible for the continuous relaxation if and only if*

$$(nc-2)\log(nc-1) \leq nc \tag{3.3}$$

Proof. Assume that c is such that $x_1(c) \geq 1$. Then

$$\begin{aligned}\iint_{A_c} \frac{1}{x}\mathrm{d}x\mathrm{d}y &= \int_{x_1(c)}^{n} \mathrm{d}x \int_{y_c(x)}^{n} \mathrm{d}y \frac{1}{x} = \int_{x_1(c)}^{n} \mathrm{d}x \frac{n - y_c(x)}{x} \\ &= \int_{x_1(c)}^{n} \mathrm{d}x \left(\frac{n}{x} - \frac{1}{xc-1}\right) = n \log \frac{n}{x_1(c)} - \frac{1}{c} \log \frac{nc-1}{cx_1(c)-1}.\end{aligned}$$

Now,

$$cx_1(c) - 1 = \frac{cn - nc + 1}{nc-1} = \frac{1}{nc-1},$$

and so

$$\iint_{A_c} \frac{1}{x}\mathrm{d}x\mathrm{d}y = n\log(nc-1) - \frac{1}{c}\log(nc-1)^2 = (n - \frac{2}{c})\log(nc-1).$$

Hence, A_c is feasible if and only if

$$(nc-2)\log(nc-1) \leq nc.$$

Now suppose that $n > e$. If c satisfies (3.3), then

$$x_1(c) \geq \frac{n}{\exp(nc/(nc-2))} > \frac{n}{e} > 1.$$

Thus, the above calculation is valid and shows that A_c is feasible. On the other hand, if $n > e$ and if c violates (3.3), then A_c is not feasible. □

To find the solution of the continuous relaxation, we need to find the value of c that solves (3.3) with equality. Consider the equation

$$\log(\beta - 1) = \frac{\beta}{\beta - 2}.$$

Both the left and the right hand side change sign at $\beta = 2$. For $\beta > 2$, both sides are positive, and for $\beta < 2$ they are negative. By Lemma 3.7, we are looking for a solution larger than 2. For $\beta > 2$, the right hand side is decreasing, while the left hand side is increasing. It follows that there is a unique solution $\beta_0 > 2$. Numerically, $\beta_0 \approx 5.68050$. Thus, A_c is feasible if and only if $c \leq \beta_0/n$, and in order to maximize $|A_c|$, we have to choose $c = \beta_0/n$.

Lemma 3.10. *$x_1(\beta_0/n) > 1$ for n large enough.*

Proof. $x_1(\beta_0/n) - 1 = \frac{n-\beta_0+1}{\beta_0-1} > 0$ for n large enough. □

It remains to compute the maximum value of the continuous relaxation. If $x_1(c) \geq 1$, then

$$|A_c| = \iint_A \mathrm{d}x\mathrm{d}y = \int_{x_1(c)}^n \mathrm{d}x \int_{y_c(x)}^n \mathrm{d}y = \int_{x_1(c)}^n \mathrm{d}x(n - y_c(x)).$$

Now,

$$y_c(x) = \frac{1}{c}\left(\frac{x}{x - \frac{1}{c}}\right) = \frac{1}{c}\left(1 + \frac{1/c}{x - \frac{1}{c}}\right) = \frac{1}{c}\left(1 + \frac{1}{cx - 1}\right),$$

and so

$$\begin{aligned} |A_c| = \int_{x_1(c)}^n \mathrm{d}x(n - \frac{1}{c} - \frac{1/c}{cx-1}) &= (n - \frac{1}{c})(n - x_1(c)) - \frac{1}{c^2}\log\frac{cn-1}{cx_1(c)-1} \\ &= n^2\frac{nc-1}{nc}\frac{nc-2}{nc-1} - \frac{2}{c^2}\log(nc-1). \end{aligned}$$

Therefore,

$$\alpha'_n = \frac{|A_{\beta_0/n}|}{(n-1)^2} = \frac{n^2}{(n-1)^2}\left[\frac{\beta_0-2}{\beta_0} - \frac{2}{\beta_0^2}\frac{\beta_0}{\beta_0-2}\right] = \frac{n^2}{(n-1)^2}\frac{(\beta_0-2)^2-2}{\beta_0(\beta_0-2)}.$$

Numerically, $\alpha'_n \approx n^2/(n-1)^2 0.55225694$.

Second Bound

Let $G = (V, E)$ be an n-vertex graph and let $A(G)$ be defined as in Section 3.4. Let n_i denote the number of vertices with degree i, with some m total number of distinct vertex degrees. We call i a single if $n_i = 1$ and multiple if $n_i \geq 2$, noting that some i are neither single nor multiple, they just don't occur as degrees. As before, for $1 \leq i \leq k \leq n-1$, let j_{ik} be the number of edges between the ith and kth degree classes and $\chi_{ik} = 1$ if $j_{ik} > 0$, and 0 otherwise. It is easy to see that $|A(G)| = \sum_{i=1}^{n-1}\sum_{k=i}^{n-1}\chi_{ik}$. Now we set set $D_i = \sum_{k=1}^{i}\chi_{ki} + \sum_{k=i+1}^{n-1}\chi_{ik}$ and $B(G) = \sum_{i=1}^{n-1} D_i$. Note that for $k \neq i$, D_i counts $\chi_{ki} = \chi_{ik}$ twice but χ_{ii} is counted only once, so we get $|A(G)| \leq \frac{B(G)+n-1}{2}$ and therefore

$$\frac{|A(G)|}{\binom{n}{2}} \leq \frac{B(G)+n-1}{2} \cdot \frac{2}{n(n-1)} = (1+o(1))\frac{B(G)}{n^2}.$$

Our goal for this section is to prove the following theorem but first we proceed with the proofs of several necessary lemmas.

Theorem 3.11. *For any graph G,*

$$\frac{|A(G)|}{\binom{n}{2}} \leq (1+o(1))\frac{13}{24}.$$

Lemma 3.12.

$$\sum_{i=1}^{n-1} D_i \leq \sum_{i:\, i\ \textit{single}} \min(m,i) + \sqrt{m}\sqrt{\sum_{i:\, i\ \textit{multiple}} \min(m,i)}\sqrt{\sum_{i:\, i\ \textit{multiple}} n_i}.$$

Proof. Observe that $D_i \leq m$ and $D_i \leq in_i$, and hence $D_i \leq \min(m, in_i)$. By case analysis it follows that $\min(m, in_i) = \min(m, \min(m,i) \cdot n_i)$. Then since the minimum of two elements is less than their average we get

$$D_i \leq \min(m, \min(m,i) \cdot n_i) \leq \sqrt{m \cdot \min(m,i) \cdot n_i}. \tag{3.4}$$

Note that if i is single we have

$$D_i \leq \min(m,i). \tag{3.5}$$

Employing (3.4) and (3.5) we get that

$$\begin{aligned}
\sum_{i=1}^{n-1} D_i &\leq \sum_{i:\, i \text{ single}} D_i + \sum_{i:\, i \text{ multiple}} D_i \\
&\leq \sum_{i:\, i \text{ single}} \min(m,i) + \sqrt{m} \sum_{i:\, i \text{ multiple}} \sqrt{\min(m,i) \cdot n_i} \\
&\leq \sum_{i:\, i \text{ single}} \min(m,i) + \sqrt{m} \sqrt{\sum_{i:\, i \text{ multiple}} \min(m,i)} \sqrt{\sum_{i:\, i \text{ multiple}} n_i},
\end{aligned} \tag{3.6}$$

where the last inequality follows from Cauchy-Schwarz. □

We wish to upper bound the term from (3.6) over all graphs G. From our lower bound construction we know that $|A(G)| \geq (1 - o(1))\frac{1}{2}n^2$. So we may assume that $m > n/\sqrt{2}$, else we'd have $|A(G)| \leq m^2 \leq n^2/2$ and our estimation of $|A(G)|$ would be complete.

Lemma 3.13. $\displaystyle\sum_{i: n_i > 0} \min(m,i) \leq m(n - m - 1) + \frac{n(2m - n + 1)}{2}.$

Proof. We wish to upper bound $\sum_{i:n_i>0} \min(m,i)$ over all graphs. So assume the m highest possible degrees occur in our graph:

$$n - 1, n - 2, \ldots, n - m + 1, n - m.$$

Now our assumption $m > n/\sqrt{2}$ gives that $m > n - m + 1$, so the value of m has to appear in the list of degrees above. There are $n - 1 - m$ terms strictly larger than m in this list and each contributes $\min(m,i) = m$. The remaining terms sum up exactly

$\sum_{i=n-m}^{m} i$, and hence

$$\sum_{i:n_i>0} \min(m,i) \leq m(n-m-1) + \sum_{i=n-m}^{m} i$$
$$= m(n-m-1) + \frac{n(2m-n+1)}{2}. \tag{3.7}$$

Now if the m highest degrees do not occur in our graph, then some degree less than $n-m+1$ must occur which clearly gives something smaller than the term in (3.7). □

Now let s be the number of degrees i that are singles. Observe we must have $s \leq m$ and $s + 2(m-s) \leq n$, implying that $s \leq m \leq \frac{n+s}{2}$. Now using $s + \sum_{i:i \text{ multiple}} n_i = n$ and substituting

$$y = \sum_{i:i \text{ single}} \min(m,i)$$
$$z = \sqrt{\sum_{i:i \text{ multiple}} \min(m,i)},$$

we can write the term in (3.6) as

$$g(y,z,s,m) = y + \sqrt{m} \cdot z\sqrt{n-s}.$$

We wish to maximize G subject to the constraints

1. All variables are non-negative and $s \leq n$,

2. $s \leq m \leq \frac{n+s}{2}$, and

3. $y + z^2 \leq m(n-m-1) + \frac{n(2m-n+1)}{2}$,

where constraint 3 follows from Lemma 3.13.

Note that $g(y,z,s,m) = O(n^2)$, so we wish to determine how large the coefficient of n^2 in g can be as $n \to \infty$. To do this we set $S = s/n$, $M = m/n$,

$Y = \sum_{i:i \text{ single}} \min(m,i)/n$, and $Z = \sqrt{\sum_{i:i \text{ multiple}} \min(m,i)/n}$ and turn to the following numeric optimization problem: maximize

$$f(Y,Z,S,M) = Y + \sqrt{M} \cdot Z\sqrt{1-S}$$

subject to the constraints

(a) All variables are non-negative and $S \leq 1$,

(b) $S \leq M \leq \frac{1+S}{2}$, and

(c) $Y + Z^2 \leq M(1-M) + \frac{2M-1}{2}$.

By routine arguments, it follows that $g(y,z,s,m) \leq (1+o(1))\, f(Y,Z,S,M)n^2$.

Lemma 3.14. *If constraints (a), (b), and (c) hold, then*

$$f(Y,Z,S,M) \leq \frac{13}{24}.$$

Proof. For fixed values of S, M and Z, the function f is monotone in Y. Therefore, we have to choose Y as large as possible, which, according to the last constraint, implies that $Y = -M^2 + 2M - \frac{1}{2} - Z^2$. Also the right hand side of constraint (c) is non-negative if and only if $1 - \sqrt{2}/2 \leq M \leq 1 + \sqrt{2}/2$, so $1 - \sqrt{2}/2 \leq M \leq 1$.

Now for fixed values of Z and M the target function f decreases with S. Hence we need to choose S as small as possible. The constraints imply $S \geq \max\{0, 2M-1\}$ so we consider the following two cases.

If $M \leq 1/2$, then $S = 0$. In this case, we need to optimize

$$f(Z,M) = -M^2 + 2M - \frac{1}{2} - Z^2 + \sqrt{M} \cdot Z$$

subject to

1. $1 - \sqrt{2}/2 \leq M \leq 1/2$,

2. $Z^2 \leq -M^2 + 2M - \frac{1}{2}$.

The target function f is quadratic in Z with maximum at $Z_0(M) = \sqrt{M}/2$. Observe that

$$f(Z_0(M), M) = -M^2 + \frac{9}{4}M - \frac{1}{2}.$$

This function is quadratic in M, with maximum at $M = \frac{9}{8} > \frac{1}{2}$. Therefore, it is maximized by the largest feasible $M = 1/2$. In total,

$$f(Z, M) \leq f(Z_0(M), M) \leq f(Z_0(1/2), 1/2) = \frac{3}{8}$$

for all feasible values of (Z, M). Finally, observe that $(Z_0(1/2), 1/2)$ is feasible, since

$$Z_0(1/2)^2 = \frac{1}{8} < \frac{1}{4} = -\frac{1}{4} + 1 - \frac{1}{2}.$$

For the second case, suppose now that $M \geq 1/2$. Then $S = 2M - 1$, and we need to optimize

$$f(Z, M) = -M^2 + 2M - \frac{1}{2} - Z^2 + \sqrt{M} \cdot Z \cdot \sqrt{2(1-M)}$$

subject to

1. $1/2 \leq M \leq 1$,

2. $Z^2 \leq -M^2 + 2M - \frac{1}{2}$.

Now f is quadratic in Z, with maximum at $Z_0(M) = \sqrt{M(1-M)/2}$. To see that $(Z_0(M), M)$ is feasible, we have to check that

$$0 \leq -M^2 + 2M - \frac{1}{2} - Z_0(M)^2 = -\frac{1}{2}M^2 + \frac{3}{2}M - \frac{1}{2}.$$

The right hand side is a quadratic polynomial with zeros at $(3-\sqrt{5})/2 < 1/2$ and $(3+\sqrt{5})/2 > 1$, which proves that $(M, Z_0(M))$ satisfies all constraints.

Therefore, we need to maximize the quadratic function

$$f(M) = -M^2 + 2M - \frac{1}{2} + \frac{1}{2}M(1-M) = -\frac{3}{2}M^2 + \frac{5}{2}M - \frac{1}{2}$$

with $1/2 \leq M \leq 1$. The maximum is at $M = 5/6$, where the value is

$$f(5/6) = -\frac{75}{72} + \frac{25}{12} - \frac{1}{2} = \frac{13}{24} = 0.541\overline{6}.$$

$\square$

Proof (of Theorem 3.11). By Lemma 3.12 and Lemma 3.14,

$$\begin{aligned}
B(G) &= \sum_{i=1}^{n-1} D_i \\
&\leq \sum_{i:i \text{ single}} \min(m,i) + \sqrt{m}\sqrt{\sum_{i:i \text{ multiple}} \min(m,i)}\sqrt{\sum_{i:i \text{ multiple}} n_i} \\
&= g(y,z,s,m) \\
&\leq (1+o(1))\, f(Y,Z,S,M)n^2 \\
&\leq (1+o(1))\, \frac{13}{24}n^2,
\end{aligned}$$

implying that

$$\frac{|A(G)|}{\binom{n}{2}} = (1+o(1))\frac{B(G)}{n^2} \leq (1+o(1))\, \frac{13}{24}.$$

$\square$

Chapter 4

Maximal matchings in polyspiro and benzenoid chains

4.1 Introduction

Recall a matching in a graph is a collection of its edges such that no two edges in this collection have a vertex in common. Matchings in graphs serve as successful models of many phenomena in engineering, natural and social sciences. A strong initial impetus to their study came from the chemistry of benzenoid compounds after it was observed that the stability of benzenoid compounds is related to the existence and the number of perfect matchings in the corresponding graphs. That observation gave rise to a number of enumerative results that were accumulated over the course of several decades; we refer the reader to monograph [11] for a survey. Further motivation came from the statistical mechanics *via* the Kasteleyn's solution of the dimer problem [30, 31] and its applications to evaluations of partition functions for a given value of temperature. In both cases, the matchings under consideration are perfect, i.e., their edges are collectively incident to all vertices of G. It is clear that perfect matchings are as large as possible and that no other matching in G can be "larger" than a perfect one. It turns out that in all other applications we are also interested mostly in large matchings.

Basically, there are two ways to quantify the largeness of a matching. One way, by using the number of edges, gives rise to the idea of *maximum matchings*. Maximum matchings are well researched and well understood; there is a well developed structural theory and enumerative results are abundant. The classical monograph by Lovász and Plummer [38] is an excellent reference for all aspects of the theory.

An alternative way is to say that a matching is large if no other matching contains it as a proper subset; this gives rise to the concept of *maximal matchings*. Every maximum matching is also maximal, but the opposite is usually not true. Unlike their maximum counterparts, maximal matchings can have different cardinalities (unless the graph is *equimatchable*; see [24]) and the recurrences used for their enumeration are essentially non-local. As a consequence, maximal matchings are much less understood then the maximum ones. There is nothing analogous to the structural theory of maximum matchings and the enumerative results are scarce and scattered through the literature [20, 33, 48].

In spite of their obscurity, maximal matchings are natural models for several problems connected with adsorption of dimers on a structured substrate and block-allocation of a sequential resource. One can find them also in the context of polymerization of organic molecules, as witnessed by an early paper of Flory [22]. A probabilistic approach to the same problem can be found in [26]. We refer the reader to papers [3, 16, 19, 20] for some structural and enumerative results on those models. In this paper our goal is to further the line of research of reference [20] by considering graphs with more complicated connectivity patterns and richer structure of basic units. We provide enumerative and extremal results on maximal matchings in two classes of linear polymers of chemical interest: the polyspiro chains and benzenoid chains. We extablish the recurrences and generating functions for the enumerating sequences of maximal matchings in three classes of uniform polyspiro chains. We then use the obtained results to determine the asymptotic behavior and to find the extremal chains. Further, we also enumerate maximal matchings in three classes of benzenoid chains and show that one of them is extremal with respect to the number of maximal matchings. Our results show that maximal matchings behave in a radically different way that the perfect matchings; chains rich in maximal matchings are poor in perfect matchings and vice versa. We end by comparing our results with enumerative results for other type of structures in similar polymers and by discussing some possible directions of future research.

4.2 Preliminaries

The terminology and notations in this chapter are mostly standard and taken from [38, 50]. All graphs G considered in this paper will be finite and simple, with vertex set $V(G)$ and edge set $E(G)$. For a subset of vertices S of $V(G)$, we make use of the notation $G - S$ (or $G - v$ if $S = \{v\}$) to denote the subgraph of G obtained by deleting the vertices of S and all edges incident to them. For a graph G and subset of edges X of G, we use the notation $G \setminus X$ (or $G \setminus e$ if $X = \{e\}$) to denote the subgraph of G obtained by deleting the endpoints of the edges in X as well as all incident edges to these endpoints.

A *matching* M in G is a set of edges of G such that no two edges from M have a vertex in common. The number of edges in M is called its *size*. A matching in G with the largest possible size is called a *maximum matching*. If a matching in G is not a subset of a larger matching of G, it is called a *maximal matching*. Let $\Psi(G)$ denote the number of maximal matchings of G.

In this chapter we are mainly concerned with counting maximal matchings in two classes of linear polymers (or *facsiagraphs*, [29]) with simple connectivity patterns. The first class are 6-uniform cactus chains. Chain cacti are in chemical literature known as *polyspiro chains*.

A *cactus graph* is a connected graph in which no edge lies in more than one cycle. Consequently, each block of a cactus graph is either an edge or a cycle. If all blocks of a cactus G are cycles of the same length m, the cactus is *m-uniform*.

A *hexagonal cactus* is a 6-uniform cactus, i.e., a cactus in which every block is a hexagon. A vertex shared by two or more hexagons is called a *cut-vertex*. If each hexagon of a hexagonal cactus G has at most two cut-vertices, and each cut-vertex is shared by exactly two hexagons, we say that G is a *chain hexagonal cactus*. The number of hexagons is called the *length* of the chain. An example of a chain hexagonal cactus is shown in Figure 4.1.

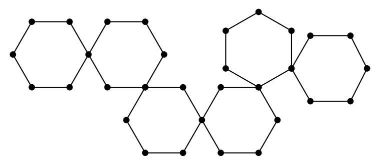

Figure 4.1 A chain hexagonal cactus of length 6.

Furthermore, any chain hexagonal cactus of length greater than one has exactly two hexagons with only one cut-vertex; such hexagons are called *terminal* and all other hexagons with two cut-vertices are called *internal*.

Internal hexagons can be one of three types depending upon the distance between its cut-vertices: in an *ortho-hexagon* cut vertices are adjacent, in a *meta-hexagon* they are at distance two, and in a *para-hexagon* cut-vertices are at distance three. The terminology is borrowed from the theory of benzenoid hydrocarbons; see [18, 19, 20] for more details. These give rise to the following three types of hexagonal cactus chains of length n: the unique chain whose internal hexagons are all para-hexagons is P_n, the unique chain whose internal hexagons are all meta-hexagons is M_n, and the unique chain whose internal hexagons are all ortho-hexagons is O_n.

The other class of unbranched polymers we consider are benzenoid chains. A *benzenoid system* is a is a connected, plane graph without cut-vertices in which all faces, except the unbounded one, are hexagons. Two hexagonal faces are either disjoint or they share exactly one common edge (adjacent hexagons). A vertex of a benzenoid system belongs to at most three hexagonal faces and the benzenoid system is called *catacondensed* if it does not posses such a vertex. If no hexagon

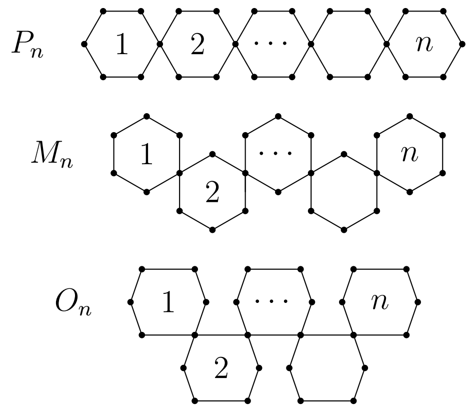

Figure 4.2 The hexagonal cactus chains P_n, M_n, and O_n.

in a catacondensed benzenoid is adjacent to three other hexagons, we say that the benzenoid is a *chain* see Figure 4.3.

The number of hexagons in a benzenoid chain is called its *length*. In each benzenoid chain there are exactly two hexagons adjacent to one other hexagon; those two hexagons are called *terminal*, while any other hexagons are called *interior*. An interior hexagon has two vertices of degree 2. If these two vertices are not adjacent, then hexagon is called *straight*. If the two vertices are adjacent, then the hexagon is called *kinky*.

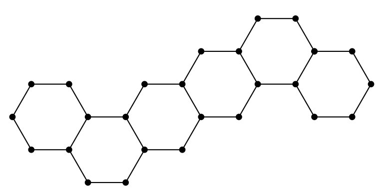

Figure 4.3 A benzenoid chain of length 6.

If all $n-2$ interior hexagons of a benzenoid chain with n hexagons are straight, we call the chain a *polyacene* and denote it by L_n. If all interior hexagons are kinky, the chain is called a *polyphenacene*. Since the number of perfect matchings in a polyphenacene of length n is equal to the $(n+2)$-nd Fibonacci number F_{n+2}, these chains are also known as *fibonacenes* [11]. We consider two specific families of polyphenacenes depicted in Figure 4.4: the *zig-zag polyphenacene*, Z_n, and *helicene*, H_n.

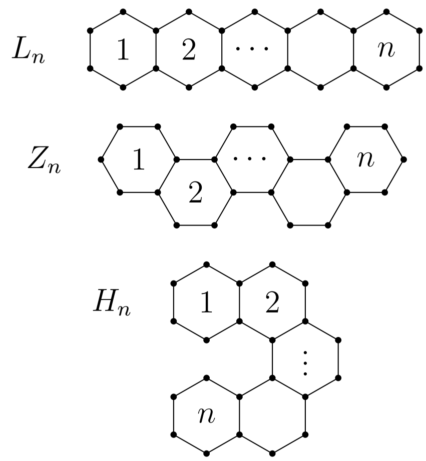

Figure 4.4 The polyacene, zig-zag polyphenacene, and helicene chains.

4.3 Chain hexagonal cacti

Generating functions

In this section, we obtain ordinary generating functions for the number of maximal matchings in the hexagonal chain cacti P_n, M_n, and O_n. To do this, we first find recursions for the number of maximal matchings using auxiliary graphs where the initial conditions are obtained by direct counting and extending the recursions backwards. By introducing generating functions for the number of maximal matchings in each auxiliary graph, the recursions can be transformed into a solvable system of equations in terms of unknown generating functions. Finally, we solve this system of equations for the desired generating function.

Lemma 4.1. *Let p_n be the number of maximal matchings in P_n and p_n^i be the number of maximal matchings in the auxiliary graph P_n^i in Figure 4.5. Then*

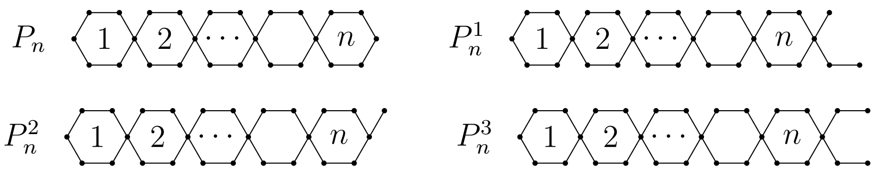

Figure 4.5 Auxiliary graphs for P_n.

(*i*) $p_n = 2p^1_{n-1} + p_{n-1}$,

(*ii*) $p^1_n = p^2_n + p^3_{n-1}$,

(*iii*) $p^2_n = p^3_{n-1} + 2p^1_{n-1}$,

(*iv*) $p^3_n = p_n + 2p^3_{n-1}$,

with the initial conditions $p_0 = 1$*,* $p^1_0 = 2$*,* $p^2_0 = 1$*, and* $p^3_0 = 3$*.*

Proof.

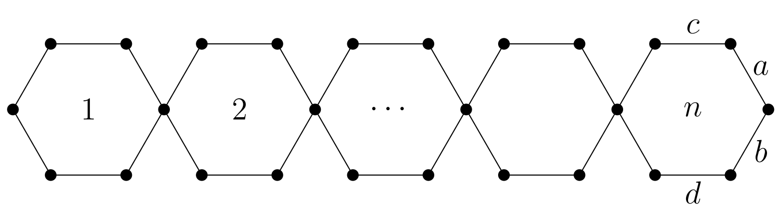

Figure 4.6 P_n with labeled edges a, b, c, and d.

Part (*i*) Refering to Figure 4.6, any maximal matching of P_n must contain exactly one of the following sets of edges: a, b, or $\{c, d\}$. Observe that for $1 \leq i \leq 3$, P^i_{n-1} is a subgraph of P_n and we will make use of similar facts from this point forward. Now for any maximal matching containing the edge a, the remaining edges must be a maximal matching of the subgraph P^1_{n-1}. The same holds for any maximal matching containing b. For any maximal matching containing the edges $\{c, d\}$, the remaining edges must be a maximal matching of the subgraph P_{n-1}. Hence the number of maximal matchings of P_n is given by $2p^1_{n-1} + p_{n-1}$, proving the recursion (i).

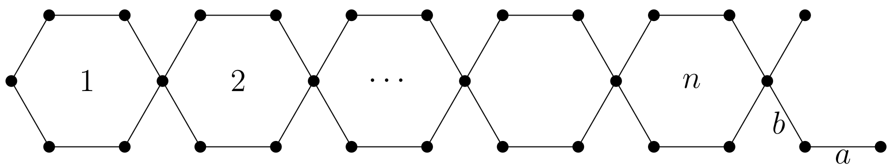

Figure 4.7 P^1_n with labeled edges a and b.

Part (*ii*) As in Figure 4.7, any maximal matching of P^1_n must contain either the edge a or the edge b. If such a maximal matching contains a, then the remaining edges in the matching must give a maximal matching of the subgraph P^2_n. If a maximal matching contains the edge b, then the remaining edges must be a maximal matching of P^3_{n-1}, proving the claimed recursion.

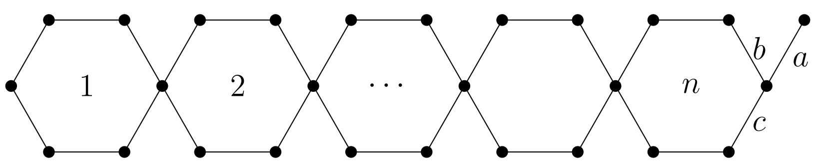

Figure 4.8 P^2_n with labeled edges a, b, and c.

Part (*iii*) Refering to Figure 4.8, any maximal matching of P^2_n must contain exactly one of the following edges: a, b, or c. If such a maximal matching contains a, then the remaining edges in the matching must give a maximal matching of the subgraph P^3_{n-1}. If a maximal matching contains the edge b or c, then the remaining edges must be a maximal matching of P^1_{n-1}. Hence the recursion (iii) holds.

Part (*iv*) Refering to Figure 4.9, any maximal matching of P^3_n must contain exactly one of the following sets of edges: $\{a, b\}$, $\{a, d\}$, or $\{b, c\}$. If such a maximal matching contains $\{a, b\}$, then the remaining edges in the matching must give a maximal matching of the subgraph P_n. If a maximal matching contains the edges

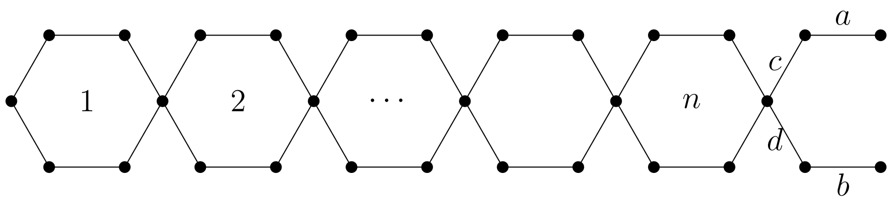

Figure 4.9 P_n^3 with labeled edges a-d.

$\{a, d\}$ or $\{b, c\}$, then the remaining edges must be a maximal matching of P_{n-1}^3, finishing the proof. □

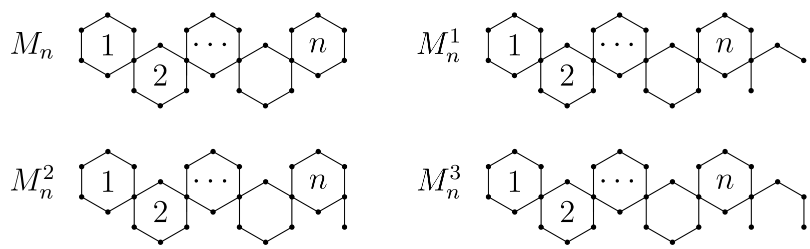

Figure 4.10 Auxiliary graphs for M_n.

Lemma 4.2. *Let m_n be the number of maximal matchings in M_n and m_n^i be the number of maximal matchings in the auxiliary graph M_n^i in Figure 4.10. Then*

(i) $m_n = 2m_{n-1}^1 + m_{n-1}$,

(ii) $m_n^1 = m_n^2 + m_{n-1}^3$,

(iii) $m_n^2 = m_{n-1}^3 + m_{n-1}^1 + m_{n-1}^2 + m_{n-1}$,

(iv) $m_n^3 = 2m_{n-1}^3 + m_{n-1}^1 + m_{n-1}^2 + m_{n-1} + m_n^2$,

with the initial conditions $m_0 = 1$, $m_0^1 = 2$, $m_0^2 = 1$, *and* $m_0^3 = 3$.

Proof.

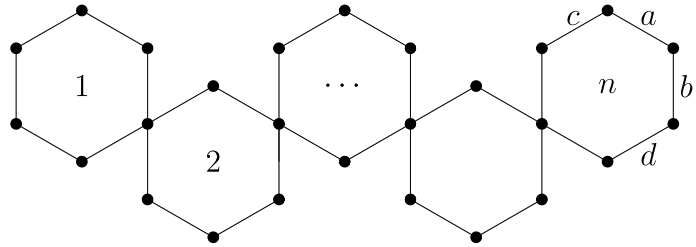

Figure 4.11 M_n with labeled edges a-d.

Part (*i*) Refering to Figure 4.11, any maximal matching of M_n must contain exactly one of the following sets of edges: a, b, or $\{c, d\}$. Any maximal matching containing the edge a, the remaining edges must be a maximal matching of the subgraph M^1_{n-1}. The same holds for any maximal matching containing b. For any maximal matching containing the edges $\{c, d\}$, the remaining edges must be a maximal matching of the subgraph M_{n-1}. Hence the recursion (i) holds.

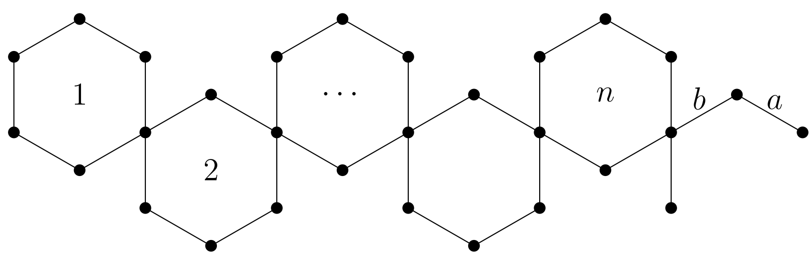

Figure 4.12 M^1_n with labeled edges a and b.

Part (*ii*) As in Figure 4.12, any maximal matching of M^1_n must contain either the edge a or the edge b. If such a maximal matching contains a, then the remaining edges in the matching must give a maximal matching of the subgraph M^2_n. If a maximal matching contains the edge b, then the remaining edges must be a maximal matching of M^3_{n-1}, proving the claimed recursion.

Part (*iii*) Refering to Figure 4.13, any maximal matching of M^2_n must contain exactly one of the following sets of edges: a, b, $\{c, d\}$, or $\{c, e\}$. If such a maximal

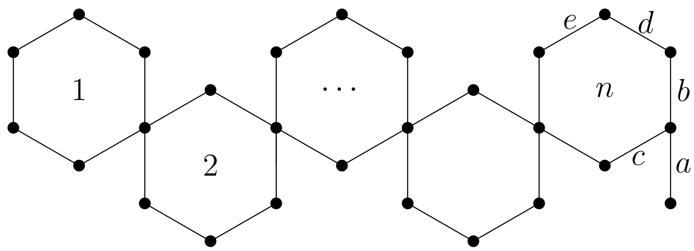

Figure 4.13 M_n^2 with labeled edges *a*-*e*.

matching contains a, then the remaining edges in the matching must give a maximal matching of the subgraph M_{n-1}^3. If a maximal matching contains the edge b, then the remaining edges give a maximal matching of M_{n-1}^1. If the maximal matching contains the sets $\{c, d\}$ or $\{c, e\}$, then the remaining edges must be a maximal matching of M_{n-1}^2 or M_{n-1}, respectively. Hence the recursion (iii) holds.

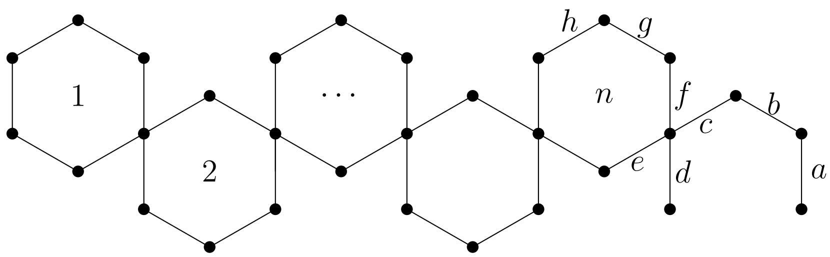

Figure 4.14 M_n^3 with labeled edges *a*-*h*.

Part *(iv)* Refering to Figure 4.14, any maximal matching of M_n^3 must contain exactly one of the following sets of edges: b, $\{a, c\}$, $\{a, d\}$, $\{a, f\}$, $\{a, e, g\}$, or $\{a, e, h\}$. If such a maximal matching contains b, then the remaining edges in the matching must give a maximal matching of the subgraph M_n^2. If a maximal matching contains the edges $\{a, c\}$ or $\{a, d\}$, then the remaining edges must be a maximal matching of M_{n-1}^3. If a maximal matching contains the edges $\{a, f\}$, then the remaining edges give a maximal matching of M_{n-1}^1. Lastly if a maximal matching contains the edges $\{a, e, g\}$, or $\{a, e, h\}$, then the remaining edges give a maximal matching of the subgraphs M_{n-1}^2 or M_{n-1}, respectively. Hence we get that $m_n^3 = 2m_{n-1}^3 + m_{n-1}^1 + m_{n-1}^2 + m_{n-1} + m_n^2$. $\square$

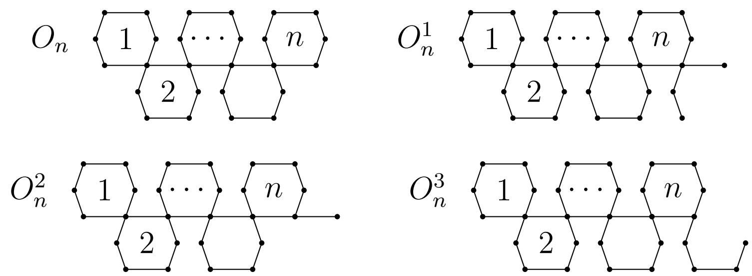

Figure 4.15 Auxiliary graphs for O_n.

Lemma 4.3. *Let o_n be the number of maximal matchings in O_n and o_n^i be the number of maximal matchings in the auxiliary graph O_n^i in Figure 4.15. Then*

(i) $o_n = 2o_{n-1}^1 + o_{n-1}$,

(ii) $o_n^1 = o_n^2 + o_{n-1}^3$,

(iii) $o_n^2 = o_{n-1}^3 + o_{n-1}^2 + o_{n-1} + 2o_{n-2}^3$,

(iv) $o_n^3 = o_n + o_{n-1}^3 + o_n^2$,

with the initial conditions $o_0 = 1$, $o_0^1 = 2$, $o_0^2 = 1$, $o_1^2 = 7$, *and* $o_0^3 = 3$.

Proof.

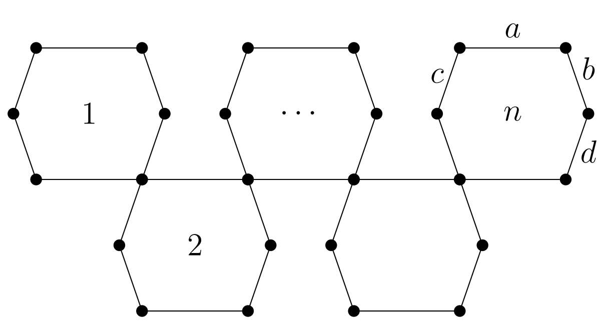

Figure 4.16 O_n with labeled edges a-d.

Part (*i*) Refering to Figure 4.16, any maximal matching of O_n must contain exactly one of the following sets of edges: a, b, or $\{c, d\}$. Any maximal matching containing the edge a, the remaining edges must be a maximal matching of the subgraph O^1_{n-1}. The same holds for any maximal matching containing b. For any maximal matching containing the edges $\{c, d\}$, the remaining edges must be a maximal matching of the subgraph O_{n-1}, proving the recursion (i) holds.

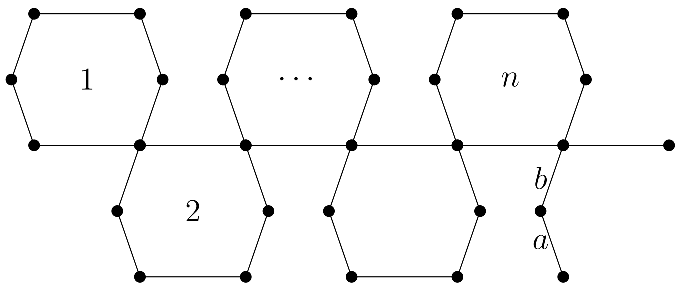

Figure 4.17 O^1_n with labeled edges a and b.

Part (*ii*) As in Figure 4.17, any maximal matching of O^1_n must contain either the edge a or the edge b. If such a maximal matching contains a, then the remaining edges in the matching must give a maximal matching of the subgraph O^2_n. If a maximal matching contains the edge b, then the remaining edges must be a maximal matching of O^3_{n-1}, which proves (ii).

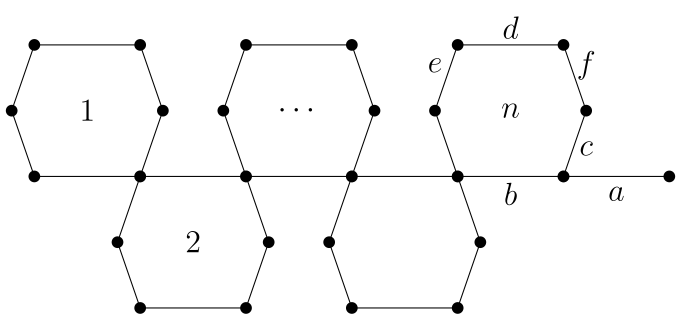

Figure 4.18 O^2_n with labeled edges a-e.

Part (*iii*) Refering to Figure 4.18, any maximal matching of O^2_n must contain exactly one of the following sets of edges: a, $\{b, d\}$, $\{b, e, f\}$, $\{c, d\}$, or $\{c, e\}$. If such a maximal matching contains a, then the remaining edges in the matching must give a maximal matching of the subgraph O^3_{n-1}. If a maximal matching contains the edge $\{b, d\}$ or $\{b, e, f\}$, then the remaining edges give a maximal matching of O^3_{n-2}. If the maximal matching contains the sets $\{c, d\}$ or $\{c, e\}$, then the remaining edges must be a maximal matching of O^2_{n-1} or O_{n-1}, respectively. Hence the recursion (iii) holds.

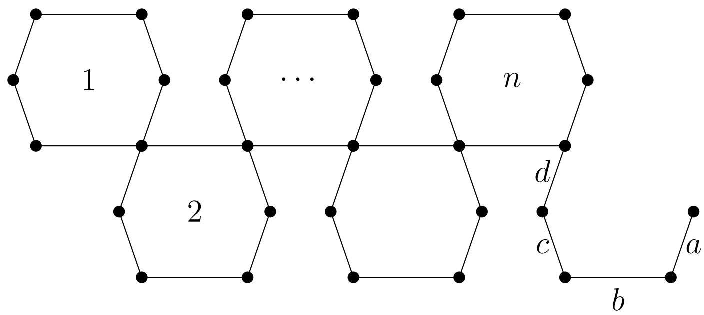

Figure 4.19 O^3_n with labeled edges a-d.

Part *(iv)* Refering to Figure 4.19, any maximal matching of O^3_n must contain exactly one of the following sets of edges: $\{a, c\}$, $\{a, d\}$, or b. If such a maximal matching contains $\{a, c\}$ or $\{a, d\}$, then the remaining edges in the matching must give a maximal matching of the subgraph O_n or O^3_{n-1}, respectively. If a maximal matching contains the edge b, then the remaining edges must be a maximal matching of O^2_n, finishing the proof $\square$

Theorem 4.4. *Let $P(x)$, $M(x)$, and $O(x)$ be the ordinary generating functions for the sequences p_n, m_n, and o_n, respectively. Then*

(i)

$$P(x) = \frac{1 + 4x^2}{1 - 5x + 4x^2 - 4x^3},$$

(ii)

$$M(x) = \frac{1 - x - 2x^2}{1 - 6x + 3x^2 - 2x^3},$$

(iii)

$$O(x) = \frac{1 + x + x^2}{1 - 4x - 4x^2 - x^3}.$$

Proof.

Part *(i)* Let $P^i(x)$ be the ordinary generating function for the sequences p_n^i for $i = 1, 2, 3$. Then Lemma 4.1 implies the following system of equations hold:

$$\begin{aligned}
\frac{P(x) - 1}{x} &= 2P^1(x) + P(x) \\
\frac{P^1(x) - 2}{x} &= \frac{P^2(x) - 1}{x} + P^3(x) \\
\frac{P^2(x) - 1}{x} &= P^3(x) + 2P^1(x) \\
\frac{P^3(x) - 3}{x} &= \frac{P(x) - 1}{x} + 2P^3(x).
\end{aligned}$$

Solving this system yields (i).

Part *(ii)* Let $M^i(x)$ be the ordinary generating function for the sequences m_n^i for $i = 1, 2, 3$. Then Lemma 4.2 implies the following system of equations hold:

$$\begin{aligned}
\frac{M(x) - 1}{x} &= 2M^1(x) + M(x) \\
\frac{M^1(x) - 2}{x} &= \frac{M^2(x) - 1}{x} + M^3(x) \\
\frac{M^2(x) - 1}{x} &= M^3(x) + M^1(x) + M^2(x) + M(x) \\
\frac{M^3(x) - 3}{x} &= 2M^3(x) + M^1(x) + M^2(x) + M(x) + \frac{M^2(x) - 1}{x}.
\end{aligned}$$

Solving this system yields (ii).

Part *(iii)* Let $O^i(x)$ be the ordinary generating function for the sequences o_n^i for $i = 1, 2, 3$. Then Lemma 4.3 implies the following system of equations hold:

$$\begin{aligned}
\frac{O(x) - 1}{x} &= 2O^1(x) + O(x) \\
\frac{O^1(x) - 2}{x} &= \frac{O^2(x) - 1}{x} + O^3(x) \\
\frac{O^2(x) - 1 - 7x}{x^2} &= \frac{O^3(x) - 3}{x} + \frac{O^2 - 1}{x} + \frac{O(x) - 1}{x} + 2O^3(x) \\
\frac{O^3(x) - 3}{x} &= \frac{O(x) - 1}{x} + O^3(x) + \frac{O^2(x) - 1}{x}.
\end{aligned}$$

Solving this system yields (iii). $\square$

Since $P(x)$, $M(x)$, and $O(x)$ are rational functions, we can conclude that the numbers p_n, m_n, and o_n each satisfy a third order linear recurrence with constant coefficients. The initial conditions can be verified by direct computations.

Corollary 4.5.

(i) $p_n = 5p_{n-1} - 4p_{n-2} + 4p_{n-3}$

with initial conditions $p_0 = 1,\ p_1 = 5,\ p_2 = 25,$

(ii) $m_n = 6m_{n-1} - 3m_{n-2} + 2m_{n-3}$

with initial conditions $m_0 = 1,\ m_1 = 5,\ m_2 = 25,$

(iii) $o_n = 4o_{n-1} + 4o_{n-2} + o_{n-3}$

with initial conditions $o_0 = 1,\ o_1 = 5,\ o_2 = 25.$

None of the obtained sequences appear in *The On-Line Encyclopedia of Integer Sequences* [1].

Now we can apply a version of Darboux's theorem to deduce the asymptotic behavior of the sequences p_n, m_n, and o_n. We refer the reader to any of standard books on generating functions, such as [5, 51] for more information on these techniques.

Theorem 4.6 (Darboux). *Let* $f(x) = \sum_{n=0}^{\infty} a_n x^n$ *denote the ordinary generating function of a sequence* a_n. *If* $f(x)$ *can be written as*

$$f(x) = \left(1 - \frac{x}{w}\right)^{\alpha} g(x),$$

where w is the smallest modulus singularity of f and g is analytic at w, then

$$a_n \sim \frac{g(w)}{\Gamma(-\alpha)} w^{-n} n^{-\alpha-1}.$$

Here $\Gamma(x)$ denotes the gamma function.

Corollary 4.7.

(i) $p_n \sim 1.37804 \cdot 4.28428^n$,

(ii) $m_n \sim 0.81408 \cdot 5.52233^n$,

(iii) $o_n \sim 1.05177 \cdot 4.86454^n$.

The characteristic equations of the three recurrences can be solved exactly, but the resulting formulas tend to be too cumbersome to be of any use. The equation for meta-chains, however, allows a compact formula for the smallest (and the only) positive root: it is equal to $\frac{1}{2}(1 + \sqrt[3]{3} - \sqrt[3]{9})$.

The obtained asymptotics suggest that meta-chains could be the richest and para-chains the poorest in maximal matchings among all polyspiro chains of the same length. In the next subsection we prove that this is, indeed, the case.

Extremal structures

Theorem 4.8. *Let G_n be a hexagonal cactus of length n. Then*

$$\Psi(P_n) \leq \Psi(G_n) \leq \Psi(M_n).$$

Let G_m be an arbitrary hexagonal cactus of length m. Observe that we can always draw G_m as in Figure 4.20, where h_m is a terminal hexagon and the hexagon adjacent to the left of h_{m-1} may attach at any of the vertices b, a, k, j, or i. Let us assume the hexagons of G_m are labeled $h_1, \ldots, h_m$ according to their ordering in Figure 4.20 where (h_1 is the other terminal hexagon).

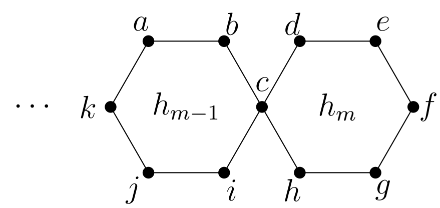

Figure 4.20 A terminal hexagon, h_m, and its adjacent hexagon, h_{m-1}, in the hexagonal chain cactus G_m.

In what follows, for $1 \le \ell, p \le m$ let H_ℓ be the subgraph of G_m induced by the vertices of the hexagons $h_1, \ldots, h_\ell$ and let $H_{\ell,p}$ denote the subgraph of G_m induced by the vertices of the two hexagons h_ℓ and h_p. We will need the following lemmas. The proof of the first lemma is immediate.

Lemma 4.9. *If H is a subgraph of the graph G, then $\Psi(H) \le \Psi(G)$.*

Lemma 4.10. *Any maximal matching in G_m must contain exactly one of the edges $cb, cd, ch,$ or ci, or the maximal matching must contain all the edges $ab, de, ji,$ and hg.*

Proof. Take a maximal matching M in G_m. For sake of contradiction, suppose that M does not contain any of the edges cb, cd, ch, or ci and that M does not contain all of the edges ab, de, ji, and hg. Then at least one of the edges ab, de, ji, and hg is missing, say ab. Since ab is not in M, then we can add the edge bc to M, which is a contradiction to the fact that M is a maximal matching. The lemma follows. $\square$

Lemma 4.11. *For the subgraph H_{m-1} of G_m, at least one of the following holds:*

(*i*) $2 \cdot \Psi(H_{m-1} - \{b, c\}) \ge \Psi(H_{m-1} - c)$

(*ii*) $2 \cdot \Psi(H_{m-1} - \{c, i\}) \ge \Psi(H_{m-1} - c)$

Proof. The proof depends on where the hexagon h_{m-2} attaches to h_{m-1}. By symmetry, suppose that h_{m-2} attaches at either i, j, or k (the case a, b, k is similar). Consider a maximal matching of $H_{m-1} - c$. If such a matching contains the edge ab, then the remaining edges give a maximal matching of $H_{m-1} - \{a, b, c\}$. If a maximal matching does not contain the edge ab, then the matching must also be maximal in the graph $H_{m-1} - \{b, c\}$. Thus by Lemma 4.9 we have

$$\begin{aligned}\Psi(H_{m-1} - \{c\}) &= \Psi(H_{m-1} - \{a, b, c\}) + \Psi(H_{m-1} - \{b, c\})\\ &\leq 2 \cdot \Psi(H_{m-1} - \{b, c\}).\end{aligned}$$

□

Proof (of Theorem 4.8). Take a hexagonal cactus C of length $n - 1$. Let us set $m = n - 1$ and suppose that C is drawn as in Figure 4.20 with vertices labeled as such, so that we may refer to this picture to aid this proof. We consider three cases of extending C by an nth hexagon h_n.

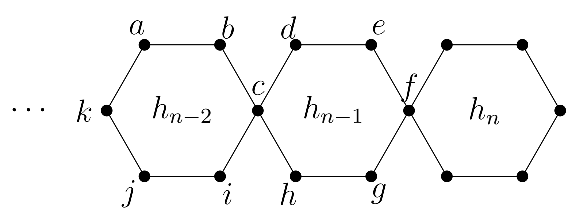

Figure 4.21 The hexagonal cactus CP.

Case 1. The hexagon h_n attaches in the para position to the vertex f and let us denote the resulting graph by CP, see Figure 4.21. To compute $\Psi(CP)$ we make use of Lemma 4.10. Consider maximal matchings in CP containing the edge bc. The remaining edges of the matching must be a maximal matching of $H_{n-2} - \{b, c\}$ and a maximal matching of $H_{n-1,n} - c$. By direct counting, we find that $\Psi(H_{n-1,n} - c) = 11$ and hence, the number of maximal matchings containing the edge bc is $11 \cdot \Psi(H_{n-2} -$

$\{b, c\}$). We count the maximal matchings containing the edges ci, cd, or ch as well as the maximal matchings containing all the edges ab, de, ji, and hg similarly, to obtain

$$\begin{aligned}\Psi(CP) =& 11(\Psi(H_{n-2} - \{b, c\}) + \Psi(H_{n-2} - \{c, i\})) + 20 \cdot \Psi(H_{n-2} - c) \\ &+ 5 \cdot \Psi(H_{n-2} - \{a, b, c, i, j\}).\end{aligned}$$

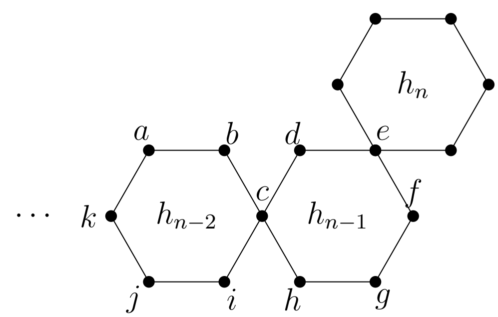

Figure 4.22 The hexagonal cactus CM.

Case 2. The hexagon h_n attaches in the meta position to the vertex e and let us denote the resulting graph by CM, see Figure 4.22. Counting similarly to Case 1 above we obtain

$$\begin{aligned}\Psi(CM) =& 17(\Psi(H_{n-2} - \{b, c\}) + \Psi(H_{n-2} - \{c, i\})) + 22 \cdot \Psi(H_{n-2} - c) \\ &+ 3 \cdot \Psi(H_{n-2} - \{a, b, c, i, j\}).\end{aligned}$$

Case 3. The hexagon h_n attaches in the ortho position to the vertex d and let us denote the resulting graph by CO, see Figure 4.23. Counting as in Cases 1 and 2,

$$\begin{aligned}\Psi(CO) =& 15(\Psi(H_{n-2} - \{b, c\}) + \Psi(H_{n-2} - \{c, i\})) + 18 \cdot \Psi(H_{n-2} - c) \\ &+ 3 \cdot \Psi(H_{n-2} - \{a, b, c, i, j\}).\end{aligned}$$

Now $\Psi(CM) \geq \Psi(CO)$ follows immediately by comparing terms. By Lemma 4.9, we have $\Psi(H_{n-2} - c) \geq \Psi(H_{n-2} - \{a, b, c, i, j\})$ and by comparing the remaining terms

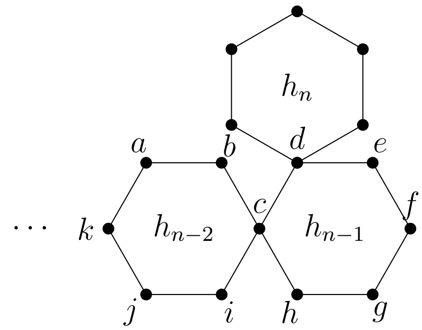

Figure 4.23 The hexagonal cactus CO.

we see that $\Psi(CM) \geq \Psi(CP)$. The preceding shows that attaching a hexagon in the meta position yields the most maximal matchings, implying

$$\Psi(G_n) \leq \Psi(M_n)$$

as desired.

To get the remaining inequality of our theorem, we need only show that $\Psi(CO) \geq \Psi(CP)$. Now we must have either (i) or (ii) of Lemma 4.11, say (i) holds. Then $4 \cdot \Psi(H_{n-2} - \{b, c\}) \geq 2 \cdot \Psi(H_{n-2} - c)$ and by Lemma 4.9 we have $\Psi(H_{n-2} - \{c, i\}) \geq \Psi(H_{n-2} - \{a, b, c, i, j\})$, showing that

$$\begin{aligned}\Psi(CO) \geq\ & 11\Psi(H_{n-2} - \{b, c\}) + 13\Psi(H_{n-2} - \{c, i\}) + 20 \cdot \Psi(H_{n-2} - c) \\ & + 5 \cdot \Psi(H_{n-2} - \{a, b, c, i, j\}).\end{aligned} \tag{4.1}$$

Now by comparing the terms of $\Psi(CP)$ with the inequality (4.1), it follows that $\Psi(CO) \geq \Psi(CP)$, which completes the proof. $\square$

It is instructive to compare the above results with the corresponding results for all matchings and for independent sets from reference [18] (Theorems 3.23 and 4.14, respectively). It can be seen that with respect to the richest chains, the number of maximal matchings behaves more like the number of independent sets than the number of all matchings. A possible explanation might be the fact that maximal matchings in any graph G are in a bijective correspondence with *nice* independent sets in G. (A set of vertices S is *nice* if $G - S$ has a perfect matching.)

4.4 Benzenoid chains

Generating functions

Now we turn our attention to benzenoid chains. Here the connectivity increases to two, and one can expect that this will result in longer recurrences, as indicated in [20]. This is, indeed, the case.

Using the same techniques outlined in subsection 4.3, we obtain ordinary generating functions for the number of maximal matchings in the benzenoid chains L_n, Z_n, and H_n.

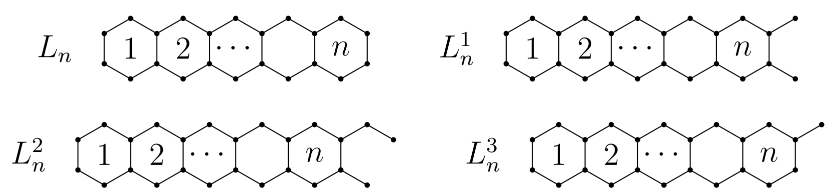

Figure 4.24 Auxiliary graphs for L_n.

Lemma 4.12. *Let ℓ_n be the number of maximal matchings in L_n and ℓ_n^i be the number of maximal matchings in the auxiliary graph L_n^i in Figure 4.24. Then*

(i) $\ell_n = \ell_{n-1}^1 + \ell_{n-1} + 2\ell_{n-2}^2$,

(ii) $\ell_n^1 = 2\ell_{n-1}^1 + \ell_{n-1} + 2\ell_{n-1}^3$,

(iii) $\ell_n^2 = \ell_n^3 + \ell_{n-1}^1 + \ell_{n-1}^3$,

(*iv*) $\ell_n^3 = 2\ell_{n-1}^1 + \ell_{n-1}^2 + \ell_{n-2}^2$,

with the initial conditions $\ell_0 = 1$, $\ell_1 = 5$, $\ell_0^1 = 2$, $\ell_0^2 = 3$, $\ell_0^3 = 2$, *and* $\ell_1^3 = 7$.

Proof.

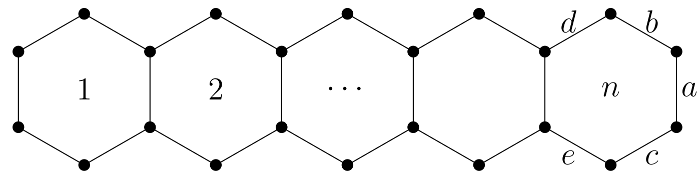

Figure 4.25 L_n with labeled edges *a*-*c*.

Part (*i*) Refering to Figure 4.25, any maximal matching of L_n must contain exactly one of the following sets of edges: a, b, c, or $\{b, c\}$. For any maximal matching containing the edge a, the remaining edges must be a maximal matching of the subgraph L_{n-1}^1. If a maximal matching contains the edges $\{b, c\}$, the remaining edges must be a maximal matching of L_{n-1}. For any maximal matching containing the edge b, the remaining edges must be a maximal matching of L_{n-2}^2 and the same holds for any maximal matching containing c. Hence the recursion (i) holds.

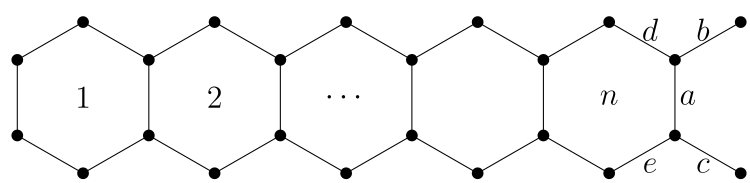

Figure 4.26 L_n^1 with labeled edges *a*-*e*.

Part (*ii*) As in Figure 4.26, any maximal matching of L_n^1 must contain exactly one of the following sets of edges: a, $\{b, c\}$, $\{d, e\}$, $\{b, e\}$, or $\{c, d\}$. If such a maximal matching contains a, then the remaining edges in the matching must give a maximal matching of the subgraph L_{n-1}^1; the same holds for any maximal matching containing $\{b, c\}$. If a maximal matching contains the edges $\{d, e\}$, then the remaining edges must be a maximal matching of M_{n-1}. If a maximal matching contains the edges $\{b, e\}$, then the remaining edges in the matching give a maximal matching of L^3_{n-1}; the same holds for any maximal matching containing the edges $\{c, d\}$. Thus the claimed recursion holds.

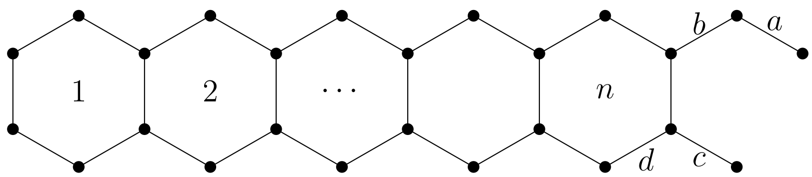

Figure 4.27 L^2_n with labeled edges *a*-*d*.

Part *(iii)* Refering to Figure 4.27, any maximal matching of L^2_n must contain exactly one of the following sets of edges: a, $\{b, c\}$, or $\{b, d\}$. If such a maximal matching contains a, then the remaining edges in the matching must give a maximal matching of the subgraph L^3_n. If a maximal matching contains the edges $\{b, c\}$, then the remaining edges give a maximal matching of L^1_{n-1}. If the maximal matching contains the edges $\{b, d\}$, then the remaining edges must be a maximal matching of L^3_{n-1}. Hence the recursion (iii) holds.

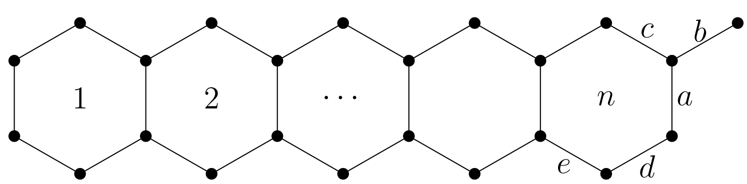

Figure 4.28 L^3_n with labeled edges *a*-*e*.

Part *(iv)* Refering to Figure 4.28, any maximal matching of L^3_n must contain exactly one of the following sets of edges: a, b, $\{c, d\}$, or $\{c, e\}$. If such a maximal matching contains a or $\{c, d\}$, then the remaining edges in the matching must give a maximal matching of the subgraph L^1_{n-1}. If a maximal matching contains the edge b, then the remaining edges must be a maximal matching of L^2_{n-1}. If a maximal matching contains the edges $\{c, e\}$, then the remaining edges give a maximal matching of L^2_{n-2}. Hence we get the recursion (iv). $\square$

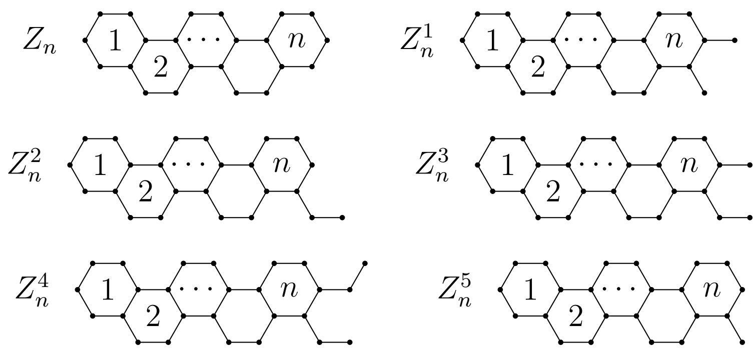

Figure 4.29 Auxiliary graphs for Z_n.

Lemma 4.13. *Let z_n be the number of maximal matchings in Z_n and z_n^i be the number of maximal matchings in the auxiliary graph Z_n^i in Figure 4.29. Then*

(i) $z_n = z^1_{n-1} + z^2_{n-1} + z^3_{n-2}$,

(ii) $z^1_n = 2z^2_{n-1} + z^4_{n-2} + z^5_{n-1} + z^3_{n-2} + z^2_{n-2}$,

(iii) $z^2_n = z_n + z^5_{n-1} + z_{n-1}$,

(iv) $z^3_n = 2z^2_{n-1} + z^3_{n-1} + z^1_{n-1} + z^5_{n-1}$,

(v) $z^4_n = z_n + z^5_{n-1} + z_{n-1} + z^2_{n-1} + z^3_{n-1}$,

(vi) $z^5_n = z^5_{n-1} + z^4_{n-2} + z^2_{n-1} + z^3_{n-2} + z_{n-1}$,

with the initial conditions $z_0 = 1$, $z_1 = 5$, $z^1_0 = 2$, $z^1_1 = 9$, $z^2_0 = 2$, $z^3_0 = 3$, $z^4_0 = 4$, $z^5_0 = 2$, and $z^5_1 = 7$.

Proof.

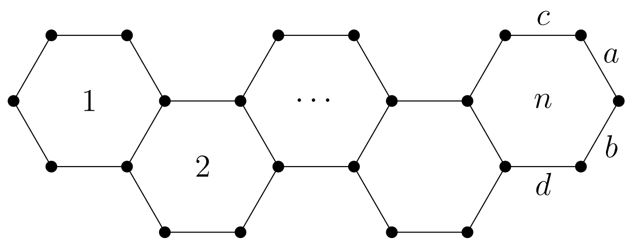

Figure 4.30 Z_n with labeled edges *a*-*d*.

Part (*i*) Refering to Figure 4.30, any maximal matching of Z_n must contain exactly one of the following sets of edges: a, b, or $\{c, d\}$. For any maximal matching containing the edge a, the remaining edges must be a maximal matching of the subgraph Z^1_{n-1}. If a maximal matching contains the edge b, then the remaining edges must give a maximal matching of Z^2_{n-1}. Lastly, if a maximal matching contains the edges $\{b, c\}$, the remaining edges must be a maximal matching of Z^3_{n-2}. Hence the recursion (i) holds.

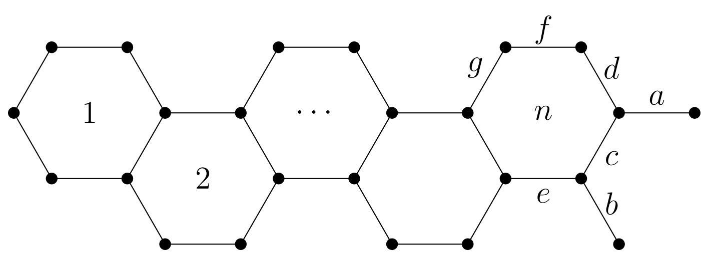

Figure 4.31 Z^1_n with labeled edges *a*-*g*.

Part (*ii*) As in Figure 4.31, any maximal matching of Z^1_n must contain exactly one of the following sets of edges: $\{a, b\}$, c, $\{d, e\}$, $\{a, e, f\}$, $\{a, e, g\}$, or $\{b, d\}$. If such a maximal matching contains $\{a, b\}$ or c, then the remaining edges in the matching must give a maximal matching of the subgraph Z^2_{n-1}. If a maximal matching contains the edges $\{d, e\}$, then the remaining edges must be a maximal matching of Z^4_{n-2}. If a maximal matching contains the edges $\{a, e, f\}$ or $\{a, e, g\}$, then the remaining edges in the matching give a maximal matching of Z^3_{n-2} or Z^2_{n-2}, respectively. If a maximal matching contains the edges $\{b, d\}$, then the remaining edges must be a maximal matching of Z^5_{n-1}. Thus the claimed recursion holds.

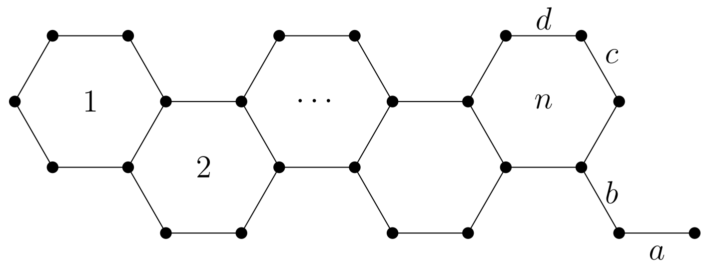

Figure 4.32 Z^2_n with labeled edges a-d.

Part *(iii)* Refering to Figure 4.32, any maximal matching of Z^2_n must contain exactly one of the following sets of edges: a, $\{b, c\}$, or $\{b, d\}$. If such a maximal matching contains a, then the remaining edges in the matching must give a maximal matching of the subgraph Z_n. If a maximal matching contains the edges $\{b, c\}$, then the remaining edges give a maximal matching of Z^5_{n-1}. If the maximal matching contains the edges $\{b, d\}$, then the remaining edges must be a maximal matching of Z_{n-1}. Hence the recursion (iii) holds.

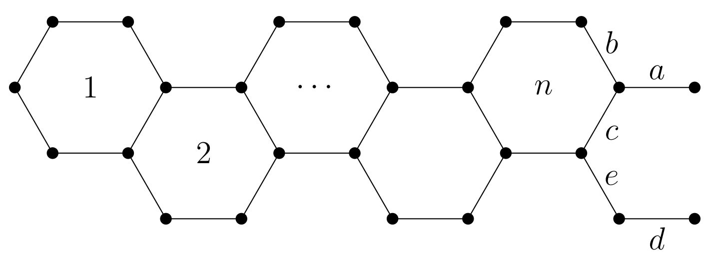

Figure 4.33 Z^3_n with labeled edges a-e.

Part *(iv)* Refering to Figure 4.33, any maximal matching of Z^3_n must contain exactly one of the following sets of edges: $\{a, d\}$, $\{a, e\}$, $\{b, d\}$, $\{b, e\}$, or $\{c, d\}$. If such a maximal matching contains $\{a, d\}$, then the remaining edges in the matching must give a maximal matching of the subgraph Z^3_{n-1}. If a maximal matching contains the edges $\{a, e\}$, then the remaining edges must be a maximal matching of Z^2_{n-1}. If a maximal matching contains the edge $\{b, d\}$, then the remaining edges give a maximal matching of Z^1_{n-1}. If a maximal matching contains the edges $\{b, e\}$, then the remaining edges give a maximal matching of Z^5_{n-1}. If a maximal matching contains the edges $\{c, d\}$, then the remaining edges give a maximal matching of Z^2_{n-1}. Hence we get the recursion (iv).

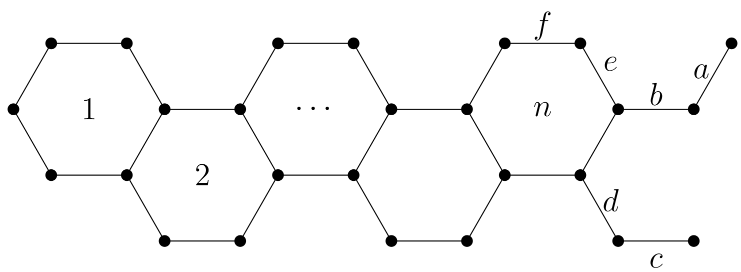

Figure 4.34 Z^4_n with labeled edges a-f.

Part (*v*) Refering to Figure 4.34, any maximal matching of Z^4_n must contain exactly one of the following sets of edges: $\{a, c\}$, $\{a, d, e\}$, $\{a, d, f\}$, $\{b, c\}$, or $\{b, d\}$. If such a maximal matching contains $\{a, c\}$, then the remaining edges in the matching must give a maximal matching of the subgraph Z_n. If a maximal matching contains the edges $\{a, d, e\}$, then the remaining edges give a maximal matching of Z^5_{n-1}. If the maximal matching contains the edges $\{a, d, f\}$, then the remaining edges must be a maximal matching of Z_{n-1}. If the maximal matching contains the edges $\{b, c\}$, then the remaining edges must be a maximal matching of Z^3_{n-1}. If the maximal matching contains the edges $\{b, d\}$, then the remaining edges must be a maximal matching of Z^2_{n-1}. Hence the recursion (iii) holds.

Part (*vi*) Refering to Figure 4.35, any maximal matching of Z^5_n must contain exactly one of the following sets of edges: $\{a, c\}$, $\{a, e\}$, b, $\{c, d\}$, or $\{d, e\}$. If such a maximal matching contains $\{a, c\}$, then the remaining edges in the matching must give a maximal matching of the subgraph Z^5_{n-1}. If a maximal matching contains

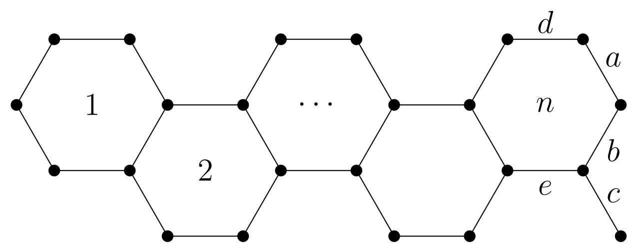

Figure 4.35 Z_n^5 with labeled edges *a*-*e*.

$\{a, e\}$, then the remaining edges in the matching must give a maximal matching of Z_{n-2}^4. If a maximal matching contains b, then the remaining edges in the matching must give a maximal matching of Z_{n-1}^2. If a maximal matching contains $\{c, d\}$, then the remaining edges in the matching must give a maximal matching of Z_{n-1}. If a maximal matching contains $\{d, e\}$, then the remaining edges in the matching must give a maximal matching of Z_{n-2}^3. The recursion (iv) now follows. □

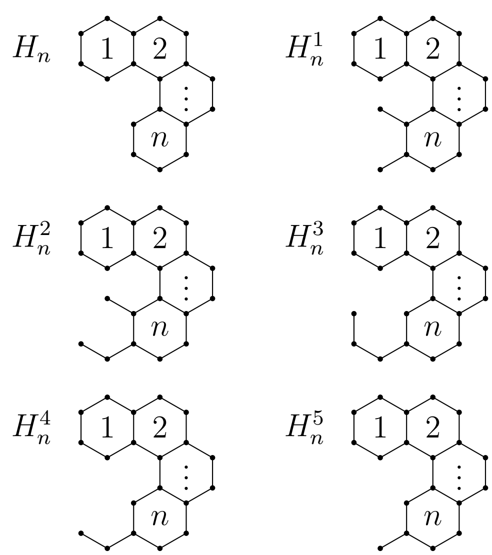

Figure 4.36 Auxiliary graphs for H_n.

Lemma 4.14. *Let h_n be the number of maximal matchings in H_n and h_n^i be the number of maximal matchings in the auxiliary graph H_n^i in Figure 4.36. Then* (*i*) $h_n = h_{n-1} + h^1_{n-1} + h^2_{n-2} + h^3_{n-2}$,

(*ii*) $h^1_n = 2h^4_{n-1} + h^5_{n-1} + h^3_{n-2} + 2h^4_{n-2} + h^5_{n-2}$,

(*iii*) $h^2_n = h^3_{n-1} + 2h^4_{n-1} + 2h^4_{n-2} + 2h^3_{n-2} + h^5_{n-2}$,

(*iv*) $h^3_n = h^5_n + h_n$,

(*v*) $h^4_n = h_n + h^2_{n-1}$,

(*vi*) $h^5_n = h^2_{n-1} + h^4_{n-1} + h^1_{n-1}$,

with the initial conditions $h_0 = 1$, $h_1 = 5$, $h^1_0 = 2$, $h^1_1 = 9$, $h^2_0 = 3$, $h^2_1 = 11$, $h^3_0 = 3$, $h^4_0 = 2$, *and* $h^5_0 = 2$.

Proof.

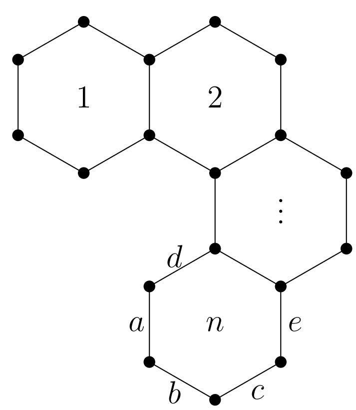

Figure 4.37 H_n with labeled edges *a*-*e*.

Part (*i*) Refering to Figure 4.37, any maximal matching of H_n must contain exactly one of the following sets of edges: $\{a, c\}$, b, $\{a, e\}$, or $\{c, d\}$. For any maximal matching containing the edges $\{a, c\}$, the remaining edges must be a maximal matching of the subgraph H_{n-1}. If a maximal matching contains the edge b, then the remaining edges must give a maximal matching of H_{n-1}^1. Lastly, if a maximal matching contains the edges $\{a, e\}$ or $\{c, d\}$, the remaining edges must be a maximal matching of H_{n-2}^2 or H_{n-2}^3. Hence the recursion (i) holds.

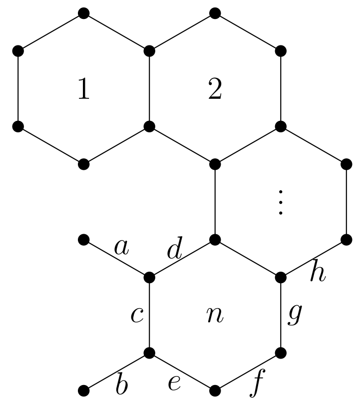

Figure 4.38 H_n^1 with labeled edges a-h.

Part *(ii)* As in Figure 4.38, any maximal matching of H_n^1 must contain exactly one of the following sets of edges: $\{a, b\}$, c, $\{d, e, g\}$, $\{d, e, h\}$, $\{a, e\}$, $\{b, d, f\}$, or $\{b, d, g\}$. If such a maximal matching contains $\{a, b\}$ or c, then the remaining edges in the matching must give a maximal matching of the subgraph H_{n-1}^4. If a maximal matching contains the edges $\{d, e, g\}$, then the remaining edges must be a maximal matching of H_{n-2}^4. If a maximal matching contains the edges $\{d, e, h\}$, then the remaining edges in the matching give a maximal matching of H_{n-2}^5. If a maximal matching contains the edges $\{a, e\}$, then the remaining edges must be a maximal matching of H_{n-1}^5. If a maximal matching contains the edges $\{b, d, f\}$ or $\{b, d, g\}$, then the remaining edges must be a maximal matching of H_{n-2}^3 or H_{n-2}^4, respectively. Thus the claimed recursion holds.

Part *(iii)* Refering to Figure 4.39, any maximal matching of H_n^2 must contain exactly one of the following sets of edges: $\{a, d\}$, $\{a, c\}$, $\{a, e, f, h\}$, $\{a, e, f, i\}$, $\{a, e, g\}$, $\{b, d\}$, $\{b, e, g\}$, or $\{b, e, h\}$. If such a maximal matching contains $\{a, d\}$, then the re-

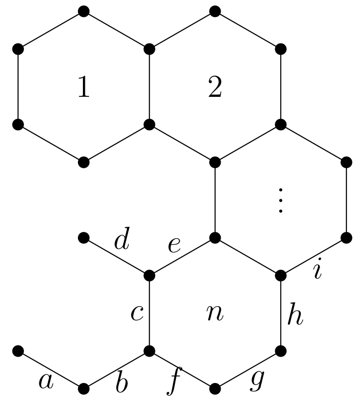

Figure 4.39 H_n^2 with labeled edges a-i.

maining edges in the matching must give a maximal matching of the subgraph H_{n-1}^3. If a maximal matching contains the edges $\{a, c\}$, then the remaining edges give a maximal matching of H_{n-1}^4. If a maximal matching contains the edges $\{a, e, f, h\}$, then the remaining edges must be a maximal matching of H_{n-2}^4. If a maximal matching contains the edges $\{a, e, f, i\}$, then the remaining edges must be a maximal matching of H_{n-2}^5. If a maximal matching contains the edges $\{a, e, g\}$, then the remaining edges must be a maximal matching of H_{n-2}^3. If a maximal matching contains the edges $\{b, d\}$, then the remaining edges must be a maximal matching of H_{n-1}^4. If a maximal matching contains the edges $\{b, e, g\}$, then the remaining edges must be a maximal matching of H_{n-2}^3. If a maximal matching contains the edges $\{b, e, h\}$, then the remaining edges must be a maximal matching of H_{n-2}^4. Hence the recursion (iii) holds.

Part (*iv*) Refering to Figure 4.40, any maximal matching of H_n^3 must contain either the edge a or the edge b. If such a maximal matching contains a, then the remaining edges in the matching must give a maximal matching of the subgraph H_n^5. If a maximal matching contains the edge b, then the remaining edges must be a maximal matching of H_n, proving (iv).

Part (*v*) Refering to Figure 4.41, any maximal matching of H_n^4 must contain

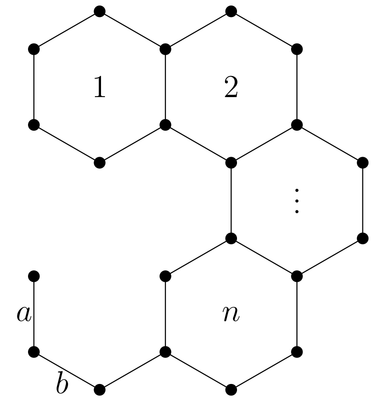

Figure 4.40 H_n^3 with labeled edges a and b.

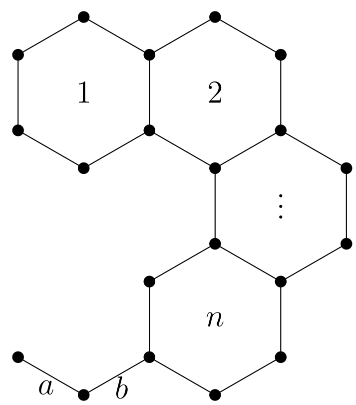

Figure 4.41 H_n^4 with labeled edges a and b.

either the edge a or the edge b. If such a maximal matching contains a, then the remaining edges in the matching must give a maximal matching of the subgraph H_n. If a maximal matching contains the edge b, then the remaining edges must be a maximal matching of H_{n-1}^2. Hence the recursion (iii) holds.

Part (vi) Refering to Figure 4.42, any maximal matching of H_n^5 must contain exactly one of the following edges: a, b, or c. If such a maximal matching contains a, then the remaining edges in the matching must give a maximal matching of the subgraph H_{n-1}^2. If a maximal matching contains b, then the remaining edges in the matching must give a maximal matching of H_{n-1}^4. If a maximal matching contains c, then the remaining edges in the matching must give a maximal matching of H_{n-1}^1.

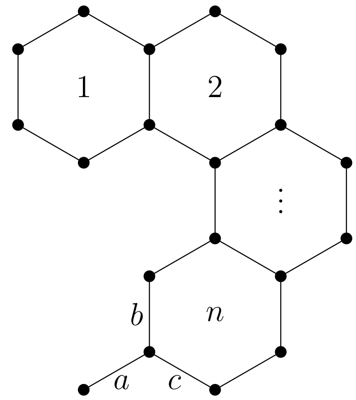

Figure 4.42 H_n^5 with labeled edges a-c.

The recursion (iv) now follows. $\square$

Theorem 4.15. *Let $L(x)$, $Z(x)$, and $H(x)$ be the ordinary generating functions for the sequences ℓ_n, z_n, and h_n, respectively. Then*

(i)

$$L(x) = \frac{1+x-x^3}{1-4x-x^4-x^5},$$

(ii)

$$Z(x) = \frac{1+2x+4x^2+4x^3+6x^4+4x^5+x^6}{1-3x-x^2-6x^3-7x^4-7x^5-5x^6-x^7},$$

(iii)

$$H(x) = \frac{1+4x+8x^2+8x^3+7x^4+4x^5+2x^6}{1-x-7x^2-12x^3-6x^4-7x^5-4x^6-2x^7}.$$

Proof.

Part *(i)* Let $L^i(x)$ be the ordinary generating function for the sequences ℓ_n^i for

$i = 1, 2, 3$. Then Lemma 4.12 implies the following system of equations hold:

$$\begin{aligned}
\frac{L(x) - 1 - 5x}{x^2} &= \frac{L^1(x) - 2}{x} + \frac{L(x) - 1}{x} + 2L^2(x) \\
\frac{L^1(x) - 2}{x} &= 2L^1(x) + L(x) + 2L^3(x) \\
\frac{L^2(x) - 3}{x} &= \frac{L^3(x) - 2}{x} + L^1(x) + L^3(x) \\
\frac{L^3(x) - 2 - 7x}{x^2} &= 2\left(\frac{L^1(x) - 2}{x}\right) + \frac{L^2(x) - 3}{x} + L^2(x).
\end{aligned}$$

Solving this system yields (i).

Part *(ii)* Let $Z^i(x)$ be the ordinary generating function for the sequences z_n^i for $i = 1, 2, \ldots, 6$. Then Lemma 4.13 implies the following system of equations hold:

$$\begin{aligned}
\frac{Z(x) - 1 - 5x}{x^2} &= \frac{Z^1(x) - 2}{x} + \frac{Z^2(x) - 2}{x} + Z^3(x) \\
\frac{Z^1(x) - 2 - 9x}{x^2} &= 2\left(\frac{Z^2(x) - 2}{x}\right) + Z^4(x) + \frac{Z^5(x) - 2}{x} + Z^3(x) + Z^2(x) \\
\frac{Z^2(x) - 2}{x} &= \frac{Z(x) - 1}{x} + Z^5(x) + Z(x) \\
\frac{Z^3(x) - 3}{x} &= 2Z^2(x) + Z^3(x) + Z^1(x) + Z^5(x) \\
\frac{Z^4(x) - 4}{x} &= \frac{Z(x) - 1}{x} + Z^5(x) + Z(x) + Z^2(x) + Z^4(x) \\
\frac{Z^5(x) - 2 - 7x}{x^2} &= \frac{Z^5(x) - 2}{x} + Z^4(x) + \frac{Z^2(x) - 2}{x} + Z^3(x) + \frac{Z(x) - 1}{x}.
\end{aligned}$$

Solving this system yields (ii).

Part *(iii)* Let $H^i(x)$ be the ordinary generating function for the sequences h_n^i

for $i = 1, 2, \ldots, 6$. Then Lemma 4.14 implies the following system of equations hold:

$$\begin{aligned}
\frac{H(x) - 1 - 5x}{x^2} &= \frac{H(x) - 1}{x} + \frac{H^1(x) - 2}{x} + H^2(x) + H^3(x) \\
\frac{H^1(x) - 2 - 9x}{x^2} &= 2\left(\frac{H^4(x) - 2}{x}\right) + \frac{H^5(x) - 2)}{x} + H^3(x) + 2H^4(x) + H^5(x) \\
\frac{H^2(x) - 3 - 11x}{x^2} &= \frac{H^3(x) - 3}{x} + 2\left(\frac{H^4(x) - 2}{x}\right) + 2H^4(x) + 2H^3(x) + H^5(x) \\
H^3(x) &= H(x) + H^5(x) \\
\frac{H^4(x) - 2}{x} &= \frac{H(x) - 1}{x} + H^2(x) \\
\frac{H^5(x) - 2}{x^2} &= H^2(x) + H^4(x) + H^1(x).
\end{aligned}$$

Solving this system yields (i).

□

Since $L(x)$, $Z(x)$, and $H(x)$ are rational functions, we can examine their denominators to obtain linear recurrences for the sequences ℓ_n, z_n, and h_n. The initial conditions can be verified by direct computations.

Corollary 4.16.

(i) $\ell_n = 4\ell_{n-1} + \ell_{n-4} + \ell_{n-5}$

with initial conditions $\ell_0 = 1$, $\ell_1 = 5$, $\ell_2 = 20$, $\ell_3 = 79$, *and* $\ell_4 = 317$,

(ii) $z_n = 3z_{n-1} + z_{n-2} + 6z_{n-3} + 7z_{n-4} + 7z_{n-5} + 5z_{n-6} + z_{n-7}$

with initial conditions $z_0 = 1$, $z_1 = 5$, $z_2 = 20$, $z_3 = 75$, $z_4 = 288$, $z_5 = 1105$, *and* $z_6 = 4234$,

(iii) $h_n = h_{n-1} + 7h_{n-2} + 12h_{n-3} + 6h_{n-4} + 7h_{n-5} + 4h_{n-6} + 2h_{n-7}$

with initial conditions $h_0 = 1$, $h_1 = 5$, $h_2 = 20$, $h_3 = 75$, $h_4 = 288$, $h_5 = 1094$, *and* $h_6 = 4171$.

Again we can use Darboux's Theorem to deduce the asymptotics of the sequences ℓ_n, z_n, and h_n. The smallest modulus singularity of $L(x)$ is approximately $x = 0.248804$. Hence, the asymptotic behavior of ℓ_n is given by $\ell_n \sim 4.01923^{n+1}$ for large n. Similarly, we deduce that $z_n \sim 3.83256^{n+1}$ and $h_n \sim 3.81063^{n+1}$ for large n.

Extremal structure

In this subsection, we prove the linear polyacene has most maximal matchings among all benzenoid chains of the same length.

Theorem 4.17. *Let* G_n *be a benzenoid chain of length* n. *Then*

$$\Psi(G_n) \leq \Psi(L_n).$$

Let G_m be an arbitrary benzenoid chain of length m. Observe that we can always draw G_m as in Figure 4.43, where h_m is a terminal hexagon and the hexagon adjacent to the left of h_{m-1} may attach at any of the edges f, g, or h. Let us assume the hexagons of G_m are labeled $h_1, \ldots, h_m$ according to their ordering in Figure 4.43 where (h_1 is the other terminal hexagon).

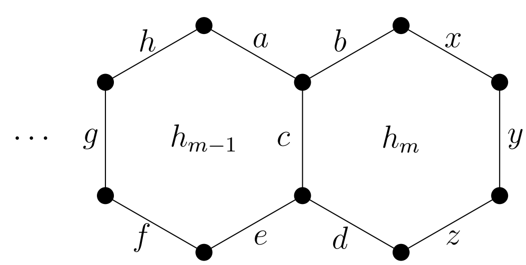

Figure 4.43 A terminal hexagon, h_m, and its adjacent hexagon, h_{m-1}, in the benzenoid chain G_m.

In what follows, let us adopt all of the same notation introduced in section 4.3. We also make use of Lemma 4.9 introduced previously, since this holds for arbitrary graphs.

Lemma 4.18. *Any maximal matching of G_m must contain at least one of the edges a, b, c, d or e. Moreover, any maximal matching of G_m contains exactly one of these edges, or contains exactly one of the following pairs of edges: a and e, a and d, b and e, or b and d.*

Proof. Take a maximal matching M. For sake of contradiction, suppose M contains none of the edges a, b, c, d or e. Then we could add the edge c to M, which is a contradiction to M being a maximal matching. Hence at least one of the edges a, b, c, d or e. The remaining part of the lemma follows by considering which pairs of edges can belong to the same matching. □

Proof. (of theorem 4.17). Take a benzenoid chain B of length $n-1$. Let us set $m = n-1$ and suppose that B is drawn as in Figure 4.43 with edges labeled as such, so that we may refer to this picture to aid this proof. We consider two cases of extending B by an nth hexagon h_n.

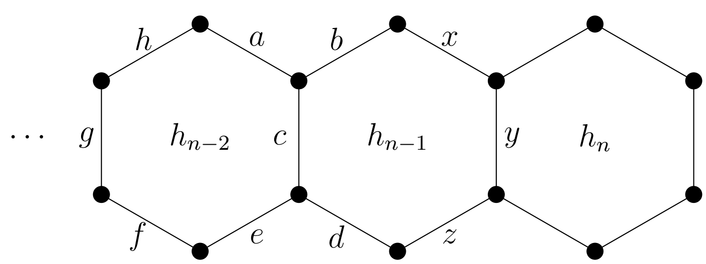

Figure 4.44 The benzenoid chain BL.

Case 1. The hexagon h_n attaches in the linear position to the edge y and let us denote the resulting graph by BL, see Figure 4.44. To compute $\Psi(BL)$ we make use of Lemma 4.18 and count matchings based on which of the edges a, b, c, d, e are saturated. Of the possibilities in Lemma 4.18, consider the maximal matchings of BL

containing only the edge a. Such a matching must also contain the edges f and z, else this matching would contain one of the other edges d or e. The remaining edges of the matching must be a maximal matching of $H_{n-2} \setminus \{a, f\}$ and a maximal matching of $H_{n-1,n} \setminus z$. By directly counting, we find that $\Psi(H_{n-1,n} \setminus z) = 4$ and hence, the number of maximal matchings containing only the edge a is $4 \cdot \Psi(H_{n-2} \setminus \{a, f\})$. We count the remaining cases from Lemma 4.18 similarly. We note that a $H_{n-1} \setminus c$ is used to count maximal matchings containing the edges b or d, since these edges do not belong to the subgraph H_{n-2}. For example, the number of maximal matchings containing only the edge b is $3 \cdot \Psi(H_{n-2} \setminus \{c, f\})$. Thus

$$\begin{aligned}\Psi(BL) = {} & 4 \cdot \Psi(H_{n-2} \setminus \{a, f\}) + 3 \cdot \Psi(H_{n-2} \setminus \{c, f\}) + 14 \cdot \Psi(H_{n-2} \setminus c) \\ & + 4 \cdot \Psi(H_{n-2} \setminus \{e, h\}) + 3 \cdot \Psi(H_{n-2} \setminus \{c, h\}) + 9 \cdot \Psi(H_{n-2} \setminus \{a, e\}) \\ & + 7 \cdot \Psi(H_{n-2} \setminus \{a, c\}) + 7 \cdot \Psi(H_{n-2} \setminus \{c, e\}).\end{aligned}$$

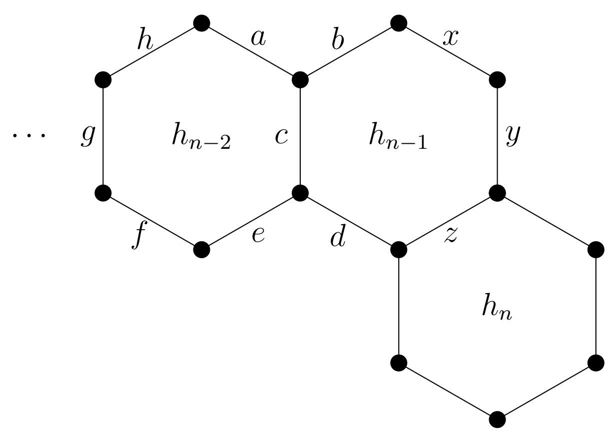

Figure 4.45 The benzenoid chain BK.

Case 2. The hexagon h_n attaches in the kinky position to the edge z and let us denote the resulting graph by BK, see Figure 4.45. Counting as in Case 1 above we obtain

$$\begin{aligned}\Psi(BK) = 6 \cdot \Psi(H_{n-2} \setminus \{a, f\}) &+ 5 \cdot \Psi(H_{n-2} \setminus \{c, f\}) + 12 \cdot \Psi(H_{n-2} \setminus c)\\ &+ 5 \cdot \Psi(H_{n-2} \setminus \{e, h\}) + 3 \cdot \Psi(H_{n-2} \setminus \{c, h\}) + 8 \cdot \Psi(H_{n-2} \setminus \{a, e\})\\ &+ 5 \cdot \Psi(H_{n-2} \setminus \{a, c\}) + 7 \cdot \Psi(H_{n-2} \setminus \{c, e\}).\end{aligned}$$

Now considering the terms in $\Psi(BL)$, by Lemma 4.9 we have

$$\begin{aligned}\Psi(H_{n-2} \setminus \{a, c\}) &\geq \Psi(H_{n-2} \setminus \{a, f\}),\\ \Psi(H_{n-2} \setminus \{c\}) &\geq \Psi(H_{n-2} \setminus \{c, f\}), and\\ \Psi(H_{n-2} \setminus \{a, e\}) &\geq \Psi(H_{n-2} \setminus \{e, h\}),\end{aligned}$$

implying that

$$\begin{aligned}\Psi(BL) \geq &6 \cdot \Psi(H_{n-2} \setminus \{a, f\}) + 5 \cdot \Psi(H_{n-2} \setminus \{c, f\}) + 12 \cdot \Psi(H_{n-2} \setminus c)\\ &+ 5 \cdot \Psi(H_{n-2} \setminus \{e, h\}) + 3 \cdot \Psi(H_{n-2} \setminus \{c, h\}) + 8 \cdot \Psi(H_{n-2} \setminus \{a, e\})\\ &+ 5 \cdot \Psi(H_{n-2} \setminus \{a, c\}) + 7 \cdot \Psi(H_{n-2} \setminus \{c, e\})\\ &\geq \Psi(BK).\end{aligned}$$

The above proves that attaching a hexagon linearly gives more maximal matchings than attaching a hexagon in the kinky position. The inequality stated in the theorem follows. □

Again, we can see that the number of maximal matchings follows the same pattern as the number of independent sets, contrary to the number of all and of perfect matchings. While the last two increase with the number of kinky hexagons, the number of maximal matchings decreases. Further, unlike the number of perfect matchings which does not discriminate between left and right kinks, the number of maximal matchings seems to be sensitive to the direction of successive turns. It seems that the helicenes have the smallest number of maximal matchings among all benzenoid chains of the same length.

4.5 Further developments

In this last section we list some unresolved problems and indicate some possible directions of future research. We start by stating a conjecture about the extremal benzenoid chains.

Conjecture 4.19. *Let B_n be a benzenoid chain of length n. Then $\Psi(H_n) \leq \Psi(B_n)$.*

Now we turn to some structural properties. The cardinality of any smallest maximal matching in G is called the *saturation number* of G. The saturation number is of interest in the context of random sequential adsorption, since it gives the information on the worst possible case of clogging the substrate; see [20] for a discussion and [3, 19, 16] for some specific cases. However, it is not enough to know the size of the worst possible case; it is also imprtant to know how (un)likely is it to happen. This brings us back to enumerative problems, since the answer to this question depends on the ability to count maximal matchings of a given size. A neat way to handle information about maximal matchings of different sizes is to use the *maximal matching polynomial.* It was introduced in [20] and some of its basic properties were established there. There are, however, many open questions about this polynomial. For example, for ordinary (generating) matching polynomials [21, 38] we know that their coefficients are log-concave. Is this valid also for maximal matching polynomials? We have computed maximal matching polynomials explicitly for several families of graphs, and we have enumerated maximal matchings in several other families. So far, no counterexample has been found, but the proof still eludes us.

Another interesting thing to do would be to look at the dynamic aspect of the problem, emulating the approach of Flory [22].

Finally, it would be interesting to extend our results on other classes of graphs, such as rotagraphs, branching polymers, composite graphs and finite portions of various lattices.

Bibliography

[1] *The on-line encyclopedia of integer sequences*, published electronically at http://oeis.org.

[2] Y. Amanatidis, B. Green, and M. Mihail, *Graphic realizations of joint-degree matrices*, In preprint (2011).

[3] V. Andova, F. Kardoš, and R. Škrekovski, *Sandwiching saturation number of fullerene graphs*, MATCH Commun. Math. Comput. Chem. **73** (2015), 501–518.

[4] O. E. Barndorff-Nielsen, *Information and exponential families in statistical theory*, John Wiley and Sons, Chichester, UK, 1978.

[5] E. A. Bender and S. Gill Williamson, *Foundations of combinatorics with applications*, Dover, 2006.

[6] Joseph Blitzstein and Persi Diaconis, *A sequential importance sampling algorithm for generating random graphs with prescribed degrees*, Internet Math. **6** (2010), no. 4, 489–522.

[7] E. Boros, M. C. Golumbic, and V. E. Levit, *On the number of vertices belonging to all maximum stable sets of a graph*, Discrete Applied Mathematics **124** (2002), 17–25.

[8] L. Brown, *Fundamentals of statistical exponential families*, vol. 9, IMS Lecture Notes—Monograph Series, Hayward, CA, USA, 1986.

[9] S. Butenko and S. Trukhanov, *Using critical sets to solve the maximum independent set problem*, Operations Research Letters **35** (2007), 519–524.

[10] Sourav Chatterjee, Persi Diaconis, and Allan Sly, *Random graphs with a given degree sequence*, Annals of Applied Probability **21** (2011), no. 4, 1400–1435.

[11] S. J. Cyvin and I. Gutman, *Kekulé structures in benzenoid hydrocarbons*, Lec. Notes in Chemistry 46, Springer, Heidelberg **46** (1988).

[12] E. Czabarka, A. Dutle, P. L. Erdős, and I. Miklós, *On realizations of a joint degree matrix*, Discrete Applied Mathematics **181** (2015), no. 0, 283 – 288.

[13] É. Czabarka, J. Rauh, K. Sadeghi, T. Short, and L. A. Székely, *On the number of non-zero elements of joint degree vectors*, arXiv:1511.01035 [math.CO], Preparing to submit to Discrete Applied Math.

[14] R. W. Deming, *Independent number of graphs - an extension of the königegerváry theorem*, Discrete Mathematics **27** (1979), 23–33.

[15] Xenofontas Dimitropoulos, Dmitri Krioukov, Amin Vahdat, and George Riley, *Graph annotations in modeling complex network topologies*, ACM Transactions on Modeling and Computer Simulation **19** (2009), no. 4, 17:1–17:29.

[16] T. Došlić, *Saturation number of fullerene graphs*, J. Math. Chem. **43** (2008), 647–657.

[17] T. Došlić and T. Short, *Maximal matchings in polyspiro and benzenoid chains*, arXiv:1511.00590 [math.CO], Preparing to submit to MATCH Commun. Math. Comput. Chem.

[18] T. Došlić and F. Maløy, *Chain hexagonal cacti: Matchings and independent sets*, Discrete Math. (2010), 1676–1690.

[19] T. Došlić and I. Zubac, *Saturation number of benzenoid graphs*, MATCH Commun. Math. Comput. Chem. **73** (2015), 491–500.

[20] ———, *Counting maximal matchings in linear polymers*, Ars Math. Contemp. (to appear).

[21] E. J. Farrell, *Introduction to matching polynomials*, J. Comb. Theory B **27** (1979), 75–86.

[22] P. J. Flory, *Intramolecular reaction between neighboring substituents of vinyl polymers*, J. Amer. Chem. Soc. (1939), 1518–1521.

[23] O. Frank, *Statistical analysis of change in networks*, Statistica Neerlandica **45** (1991), no. 3, 283–293.

[24] A. Frendrup, B. Hartnell, and P. D. Vestergaard, *A note on equimatchable graphs*, Australas. J. Combin. **46** (2010), 185–190.

[25] M. S. Handcock and M. Morris, *A simple model for complex networks with arbitrary degree distribution and clustering*, Statistical Network Analysis: Models, Issues and New Directions **4503** (2007), 103–114.

[26] J. L. Jackson and E. W. Montroll, *Free radical statistics*, J. Chem. Phys. **28** (1958), 1101–1109.

[27] A. Jarden, V. E. Levit, and E. Mandrescu, *Critical and Maximum Independent Sets of a Graph*, arXiv:1506.00255 [cs.DM] (2015), 12 pp.

[28] ———, *Monotonic Properties of Collections of Maximum Independent Sets of a Graph*, arXiv:1506.00249 [cs.DM] (2015), 15 pp.

[29] M. Juvan, B. Mohar, A. Graovac, S. Klavžar, and J. Žerovnik, *Fast computation of the wiener index of fasciagraphs and rotagraphs*, J. Chem. Inf. Comput. Sci. **35** (1995), 834–840.

[30] P. W. Kasteleyn, *The statistics of dimers on a lattice. i. the number of dimer arrangements on a quadratic lattice*, Physica **27** (1961), 1209–1225.

[31] ———, *Dimer statistics and phase transitions*, J. Math. Phys. **4** (1963), 287–293.

[32] A. Kekulé, *Sur la constitution des substances aromatiques*, Bulletin de la Societe Chimique de Paris **3** (1865), no. 2, 98–110.

[33] M. Klazar, *Twelve countings with rooted plane trees*, Eur. J. Comb. **18** (1997), 195–210.

[34] C. E. Larson, *The critical independence number and an independence decomposition*, European Journal of Combinatorics **32** (2011), 294–300.

[35] C.E. Larson, *A Note on Critical Independence Reductions*, Bulletin of the Institute of Combinatorics and its Applications **51** (2007), 34–46.

[36] V. E. Levit and E. Mandrescu, *Combinatorial properties of the family of maximum stable sets of a graph*, Discrete Applied Mathematics **117** (2002), 149–161.

[37] ———, *Vertices belonging to all critical independent sets of a graph*, SIAM Journal on Discrete Mathematics **26** (2012), 399–403.

[38] L. Lovász and M. D. Plummer, *Matching theory*, North-Holland, Amsterdam, 1986.

[39] P. Mahadevan, D. Krioukov, K. Fall, and A. Vahdat, *Systematic Topology Analysis and Generation Using Degree Correlations*, SIGCOMM 2006 **36** (2006), no. 4, 135–146.

[40] M. E. J. Newman, *The structure and function of complex networks*, SIAM Review **45** (2003), 167–256.

[41] A.N. Patrinos and S.L. Hakimi, *Relations between graphs and integer-pair sequences*, Discrete Math. (1976).

[42] Alessandro Rinaldo, Sonja Petrović, and Stephen E. Fienberg, *Maximum lilkelihood estimation in the beta model*, Annals of Statistics **41** (2013), no. 3, 1085–1110.

[43] Kayvan Sadeghi and Alessandro Rinaldo, *Statistical models for degree distributions of networks*, NIPS workshop (Montreal, Canada), 2014.

[44] T. Short, *KE theory & the number of vertices belonging to all maximum independent sets in a graph*, Master's thesis, Virginia Commonwealth University, http://scholarscompass.vcu.edu/etd/2353/, 2011.

[45] ______, *On some conjectures concerning critical independent sets of a graph*, arXiv:1509.05057 [math.CO], Submitted to the Electronic Journal of Combinatorics (2015).

[46] Isabelle Stanton and Ali Pinar, *Constructing and sampling graphs with a prescribed joint degree distribution*, J. Exp. Algorithmics **17** (2012), 3.5:3.1–3.5:3.25.

[47] F. Sterboul, *A characterization of the graphs in which transversal number equals the matching number*, Journal of Combinatorial Theory Series B **27** (1979), no. 228-229.

[48] S. G. Wagner, *On the number of matchings of a tree*, Eur. J. Comb. **28** (2007), 1322–1330.

[49] Stanley Wasserman and Philippa Pattison, *Logit models and logistic regressions for social networks: I. an introduction to Markov graphs and p^**, Psychometrika **61** (1996), no. 3, 401–425 (English).

[50] D. West, *Introduction to Graph Theory*, 2 ed., Prentice Hall, Inc., Upper Saddle River, NJ, 2001.

[51] H. S. Wilf, *generatingfunctionology*, Academic Press, 1990.

[52] C. Q. Zhang, *Finding critical independent sets and crtitical vertex subsets are polynomial problems*, SIAM Journal on Discrete Mathematics **3** (1990), 431–438.