UNIVERSITAT POLITÈCNICA DE CATALUNYA
BARCELONATECH
Departament de Matemàtica Aplicada IV

Programa de Doctorat de Matemàtica Aplicada

Random Combinatorial Structures with low dependencies: existence and enumeration

by

Guillem Perarnau Llobet

PhD dissertation

Advisor: Oriol Serra Albó

Barcelona, July 2013

Facultat de Matemàtiques i Estadística
UNIVERSITAT POLITÈCNICA DE CATALUNYA

A l'Elena,

Acknowledgments:

This thesis is the result of a trip that, during these last years, allowed me to discover a new fascinating area. There are many people who have accompanied me and to whom I own an enormous part of this work. Their wise advice, contagious joy and unconditional support made this thesis possible.

First of all, I want to thank my advisor Oriol Serra. He has transmitted me his pure enthusiasm for knowledge, music and specially maths. Whenever I needed his advice for a new problem, he has put everything aside to discuss it, just with the help of some coffee. I will never forget the excitement on his face each time I claimed I had a new proof for that problem that didn't want to work. This thesis has a huge debt to his broad point of view and lucid ideas. With all his virtues and (not many) defects, I could not imagine anyone better to lead me all these years.

I also want to thank all my coauthors; in particular, Florent Foucaud, for all the shared moments playing board games and enjoying *vins et fromages*, Dieter Mitsche, with whom I learned more than he could imagine and Giorgis Petridis, an excellent person from who I got many useful advises. I would also like to specially thank Josep Díaz for sharing many blackboard discussions and never losing the hope in defeating a problem.

My gratitude also to Tibor Szabó and all the people in the Freie Universität, who make me feel that Berlin is a second home for me, and to Gábor Tardos for thoroughly checking this document, for his kind hospitality and for showing me Hungarian history in Budapest.

I want to give my gratitude to the non–profit help and many comments provided by colleagues and anonymous referees who enormously contributed in increasing the quality of this thesis. Specially, many thanks to Marc Noy for introducing me to permutation patterns and for our nice discussions about them.

I also want to thank all the people I have met all along these years in conferences, schools, workshops.... I would never be able to enumerate them all, but specially thanks to Andrzej, Aline, Anita, Asaf, Fiona, Henning, Jan, Kaska, Katherine, Mima, Marko, Nico, Reza, Roman, Ross, Tomas, Xavier, Will.... Thanks for the amazing moments we shared either discussing problems or learning how to toast in many languages.

Vull agrair al Department MA4 i al grup COMBGRAPH l'oportunitat de realitzar la tesi, així com el finançament per tal poder viatjar i formar-me com a matemàtic i com a persona. Agrair en especial l'energia transmesa per l'Anna Lladó i a en Josep M. Aroca, per fer que la meva experiència com a docent fos immillorable.

També vull donar les gràcies a tota la gent que algun dia m'ha preguntat, *però... i de què va el teu doctorat?*, i aquells que han despertat la meva curiositat en la ciència i en les matemàtiques, especialment al Toni Hernàndez i al Sebastià Xambó.

Tant en l'àmbit personal com professional, vull agrair a l'Arnau el seu esforç pel que faci falta, ja intentant trobar un problema de comú interés on treballar junts o convencent-me per fer un "últim" mojito. De gran suport també han estat l'Adrià, la Cris, l'Inma, la Laura, el Víctor, les múltiples vetllades renegant dels nostres doctorats i els bons moments compartits.

Gran part d'aquesta tesi li dec a la gent amb qui he compartit el dia a dia durant aquests anys i que han passat a ser molt més que excompanys de fatigues. A l'Àngela per ser una persona excel·lent, a les bromes de l'Aida, a la "mejor amiga" Cris, als comentaris del Marconi, al Morgan per fer-me riure com ningú, a les converses amb el Vena, al PD, a l'Èric, al Teixi, al Romain, a l'Hèctor i a tants d'altres que han passat pel despatx.

Gràcies a la terrasseta del Poblesec del Joan i la Núria, a la vida al piset amb el Gerard i el Sergi, a les tardes de Flohmarkt amb el Iol i la Júlia i a tanta altra gent que m'ha acompanyat durant aquests anys: Georgina, Gina, Guille, Javi, Jordi, Jud, Lara, Maria, Núria, Sergi, Tere...

Vull agrair a la meva família, avis, oncles i cosins, a aquells que hi són i a aquells que ja no, el suport sempre incondicional. Als meus pares, Tiu i Pere, per tot l'afecte que m'han donat, per haver fet de mi el que sóc, per haver-me ensenyat que la perseverança dóna els seus fruits i per intentar que sigui una mica menys desastre; sincerament, moltes gràcies. Al Martí, per tots els moments compartits d'ençà que érem petits, per ser tant bon amic com germà i pels seus rotllos de físic.

Ja per acabar, gràcies a tu, Elena, per fer-me sentir cada dia més afortunat, per les tardes de sofà, pels records passats, pels plans futurs i perquè n'estic convençut que *junts podem arribar més lluny.*

Random Combinatorial Structures with low dependencies: existence and enumeration

Abstract:

In this thesis we study different problems in combinatorics and in graph theory by means of the probabilistic method. This method was introduced by Erdős and its first applications are found in Ramsey Theory and graph colorings. It has become an extremely powerful tool to provide existential proofs for certain problems in different mathematical branches where other methods had failed utterly.

One of its main concerns is to study the behavior of random variables. In particular, one common situation arises when these random variables count the number of bad events that occur in a combinatorial structure. The idea of the Poisson Paradigm is to estimate the probability of these bad events not happening at the same time when the dependencies among them are weak or rare. If this is the case, this probability should behave similarly as in the case where all the events are mutually independent. This idea gets reflected in several well–known tools, such as the Lovász Local Lemma [52] or Suen inequality [82].

The goal of this thesis is to study these techniques by setting new versions or refining the existing ones for particular cases, as well as providing new applications of them for different problems in combinatorics and graph theory. Next, we enumerate the main contributions of this thesis.

The first part of this thesis extends a result of Erdős and Spencer on latin transversals [53]. There, the authors showed that an integer matrix such that no number appears many times, admits a latin transversal. This is equivalent to study rainbow matchings of edge–colored complete bipartite graphs. Under the same hypothesis of [53], we provide enumerating results on such rainbow matchings. Our techniques are based on the framework devised by Lu and Szekely [98].

The second part of the thesis deals with identifying codes. An identifying code is a set of vertices such that all vertices in the graph have distinct neighborhood within the set. We provide bounds on the size of a minimal identifying code in terms of the degree parameters and partially answer a question of Foucaud et al. [61]. By studying graphs with girth at least 5 and large minimum degree, we are able to estimate the size of a minimum code for random regular graphs. On a different chapter of the thesis, we show that any dense enough graph has a very large spanning subgraph that admits a small identifying code.

In some cases, proving the existence of a certain object is trivial. However, the same techniques allow us to obtain enumerative results. The study of permutation patterns is a good example of that. In the third part of the thesis we devise a new approach in order to estimate how many permutations of given length avoid a consecutive copy of a given pattern. In particular, we provide upper and lower bounds for them. One of the consequences derived from our approach is a proof of the CMP conjecture, stated by Elizalde and Noy [50] as well as some new results on the behavior of most of the patterns.

In the last part of this thesis, we focus on the Lonely Runner Conjecture, posed independently by Wills [127] and Cusick [41] and that has multiple applications in different mathematical fields. This well–known conjecture states that for any set of runners running along the unit circle with constant different speeds and starting at the same point, there is a moment where all of them are far enough from the origin. We improve the result of Chen [35] on the gap of loneliness by studying the time when two runners are close to the origin. We also extend the invisible runner result of Czerwiński and Grytczuk [44].

Estructures Combinatòries Aleatòries amb dependències febles: existència i enumeració.

Resum:

En aquesta tesi s'estudien diferents problemes en el camp de la combinatòria i la teoria de grafs, utilitzant el mètode probabilístic. Aquesta tècnica, introduïda per Erdős, ha esdevingut una eina molt potent per tal de donar proves existencials per certs problemes en diferents camps de les matemàtiques on altres mètodes no ho han aconseguit.

Un dels seus principals objectius és l'estudi del comportament de les variables aleatòries. El cas en que aquestes variables compten el nombre d'esdeveniments dolents que tenen lloc en una estructura combinatòria és de particular interès,. La idea del Paradigma de Poisson és estimar la probabilitat que tots aquests esdeveniments dolents no succeeixin a la vegada, quan les dependències entre ells són febles o escasses. En tal cas, aquesta probabilitat s'hauria de comportar de forma similar al cas on tots els esdeveniments són independents. El Lema Local de Lovász [52] o la Desigualtat de Suen [82] són exemples d'aquesta idea.

L'objectiu de la tesi és estudiar aquestes tècniques ja sigui proveint-ne noves versions, refinant-ne les existents per casos particulars o donant-ne noves aplicacions. A continuació s'enumeren les principals contribucions de la tesi.

La primera part d'aquesta tesi estén un resultat d'Erdős i Spencer sobre transversals llatins [53]. Els autors proven que qualsevol matriu d'enters on cap nombre apareix massa vegades, admet un transversal on tots els nombres són diferents. Això equival a estudiar els aparellaments multicolors en aresta–coloracions de grafs complets bipartits. Sota les mateixes hipòtesis que [53], es donen resultats sobre el nombre d'aquests aparellaments. Les tècniques que s'utilitzen estan basades en l'estratègia desenvolupada per Lu i Székely [98].

En la segona part d'aquesta tesi s'estudien els codis identificadors. Un codi identificador és un conjunt de vèrtexs tal que tots els vèrtexs del graf tenen un veïnatge diferent en el codi. Aquí s'estableixen cotes en la mida d'un codi identificador mínim en funció dels graus i es resol parcialment una conjectura de Foucaud et al. [61]. En un altre capítol, es mostra que qualsevol graf suficientment dens conté un subgraf que admet un codi identificador òptim.

En alguns casos, provar l'existència d'un cert objecte és trivial. Tot i així, es poden utilitzar les mateixes tècniques per obtenir resultats d'enumeració. L'estudi de patrons en permutacions n'és un bon exemple. A la tercera part de la tesi es desenvolupa una nova tècnica per tal d'estimar el nombre de permutacions d'una certa llargada que eviten còpies consecutives d'un patró donat. En particular, es donen cotes inferiors i superiors per a aquest nombre. Una de les conseqüències és la prova de la conjectura CMP enunciada per Elizalde i Noy [50] així com nous resultats en el comportament de la majoria dels patrons.

En l'última part de la tesi s'estudia la Conjectura Lonely Runner, enunciada independentment per Wills [127] i Cusick [41] i que té múltiples aplicacions en diferents camps de les matemàtiques. Aquesta coneguda conjectura diu que per qualsevol conjunt de corredors que corren al llarg d'un cercle unitari, hi ha un moment on tots els corredors estan suficientment lluny de l'origen. Aquí, es millora un resultat de Chen [35] ampliant la distància de tots els corredors a l'origen. També s'estén el teorema del corredor invisible de Czerwiński i Grytczuk [44].

Contents

CHAPTER 1

INTRODUCTION

The probabilistic method was initiated by Paul Erdős and it has been one of the most powerful and widely used techniques to deal with combinatorial problems. During the last 60 years it has been developed and has provided existential proofs for certain problems in different mathematical branches where other methods had failed utterly.

For instance, one may find multiple applications of it in mathematical fields like number theory, linear algebra, geometry or analysis. In particular, it is also one of the most used tools in algorithmics, connecting combinatorics with computer science. Topics like the study of randomized algorithms or property testing are good examples of that.

The probabilistic method relies on proving that a statement is true by setting a probability space and showing that the probability of this statement is strictly positive. Whereas many combinatorial proofs provide constructions of such objects in an explicit way, the probabilistic method gives only existential proofs. This is one of the reasons why this method proves to be so powerful to attack some problems where classical combinatorial arguments do not provide any interesting information. Recently, some constructive methods have been devised to give a randomized algorithm that finds the desired object and that run in polynomial time with high probability. When such algorithm exists, it is also interesting to study how to derandomize it to provide a deterministic algorithm. On this section we will comment some of these techniques.

When considering random combinatorial objects, most of their combinatorial properties can be expressed as a function of different random variables, $f(X_1, \dots, X_N)$. One of the main concerns of the probabilistic method is to study the behavior of these random variables $f(X_1, \dots, X_N)$ by understanding each random variable X_i as well as the interplay among them. For instance, one can look at the expected value of $f(X_1, \dots, X_N)$. In many situations, we want to know how likely is that $f(X_1, \dots, X_N)$ lies close to its expected value. For this reason, there is a bunch of probabilistic tools to deal with concentration of random variables around their expected value, also known as large deviation inequalities.

Nonetheless, in this thesis we will focus in the problem of estimating the probability that $f(X_1, \dots, X_N) = 0$. One common situation in the probabilistic method arises when the existence of the desired structure can be expressed in terms of the avoidance of certain set of "bad" events $\mathcal{A} = \{A_1, \dots, A_N\}$ from the probability space. For instance, a proper coloring can be understood as a random coloring that avoids the events defined by monochromatic edges. This motivates the definition of the following function $f(X_1, \dots, X_N) = X = X_1 + \cdots + X_N$, where X_i is the indicator random variable of the event $A_i \in \mathcal{A}$. Then, X counts the number of bad events in a particular instance of the probability space and the probability that $X = 0$ is exactly the probability that no bad event from $\mathcal{A}$ occurs. This property turns to be very useful. For example, the existence of at least one object satisfying our restrictions can be ensured by bounding this probability away from zero.

The study of the probability that $X = 0$ is not only related to existential results, but also to enumeration. Suppose that each element of our space appears with equal probability, that is, we have a uniform distribution. Then, the size of the probability space times the probability that a random object satisfies $X = 0$, provides the exact number of elements of this space that fulfill the desired property defined by the avoidance of $\mathcal{A}$.

In many cases it is hard to exactly compute the probability that $X = 0$, even for small probability spaces. However, meaningful asymptotic estimations can be given when the size of the probability space is large. Some applications of this idea will be seen in different parts of the thesis.

Let us focus in the tools we are going to use to bound $\Pr(X = 0)$. If the random variables X_i are non–negative and mutually independent, it is easy to compute such probability,

$$\Pr(X = 0) = \Pr\left(\bigcap_{i=1}^{N} \{X_i = 0\}\right) = \prod_{i=1}^{N} (1 - \Pr(X_i \neq 0)) \sim e^{-\mathbb{E}(X)} . \tag{1.1}$$

In particular, the distribution of the number of bad events that are satisfied, X, can be approximated by a Poisson random variable with parameter $\mathbb{E}(X)$.

Obviously, if the random variables are highly dependent this estimation can be far from being correct. Intuitively speaking, a high correlation among the variables will force $\Pr(X = 0)$ to deviate from the estimation of (1.1). Consider the following experiment. Flip just one coin and for all $i \in [N]$, set $X_i = 1$ if tail appears and $X_i = 0$ otherwise. It is clear that $\Pr(X = 0) = 1/2$, while the estimation in (1.1) gives $e^{-N/2}$, which can be arbitrarily small.

This suggests that we do not have a general intuition to rely on when computing the probability that no bad event is satisfied. Nevertheless, the estimation in (1.1) of such probability is useful when the dependencies among the random variables X_i are weak, meaning that each X_i is mutually independent to large sets of random variables; or when the join probabilities are close to the product of the individual ones.

The Poisson Paradigm states that, when there are few dependencies among the events in $\mathcal{A}$ or these dependencies are weak, the probability of these bad events not happening at the same time should behave similar to the estimation in (1.1). As a paradigm, this is not always working but provides an intuition under some circumstances on what may be the truth. The exact conditions under which we can ensure that the Poisson Paradigm is satisfied, may change depending on the techniques used and can be either local or global on the set of events $\mathcal{A}$. In this chapter, we will overview some of them. For a detailed explanation of the techniques and the conditions they require to be applied, we refer the reader to Section 2.2.2.

Let us start by introducing a well known tool to provide existential proofs of combinatorial objects, the Lovász Local Lemma (LLL). The local lemma was settled by Lovász and Erdős in

1975 (see [52]) to show the existence of 3-colorings in hypergraphs under some local restrictions on the hyperedges. Since then, it has been a useful tool for solving a great diversity of problems. As seen before, if $\mathcal{A}$ is a finite set of mutually independent events, each with probability strictly less than one, then the probability that no event occurs is always strictly positive. The LLL allows us to slightly relax the independence condition. In order to apply it, the set $\mathcal{A}$ must satisfy the following conditions: for every event $A_i \in \mathcal{A}$, if the set of all the events but $\mathcal{D}_i$ are mutually independent from A_i, then the sum of the probabilities of the events in $\mathcal{D}_i$ is not too large. If this condition is satisfied, then the probability that none of the events holds is strictly positive under some specific quantification provided by the LLL.

Because of this conditions, one can imagine the local lemma as a local union bound. The union bound implies that for every set of events $\mathcal{A}$ such that $\sum_{i=1}^{N} \Pr(A_i) < 1$, we have $\Pr(\cap_{i=1}^{N} \overline{A_i}) > 0$. The conditions needed to apply the local lemma, can be understood as a local version of the union bound.

The local lemma, not only gives the existence of these structures without bad events, but also provides and explicit exponential lower bound for it. In many cases, this bound is asymptotically tight. The power of the local lemma lies in the fact that it provides the existence of elements which have exponentially small density in a large space of combinatorial objects. Observe that some other probabilistic techniques, such as the first or the second moment, typically show the existence of elements that have constant density in the space.

As we will see in the forthcoming chapters, the local lemma takes advantage of the locality of certain defined properties in combinatorial structures. For instance, this tool is of particular interest when analyzing graphs with bounded maximum degree.

In most of the applications it is useful to study the dependency graph of the set $\mathcal{A}$ (see Definition 2.6) that captures the dependencies among the events. However, in order to illustrate the statement of the local lemma, we introduce a simplified version of it, known as the symmetric local lemma, for which the dependency graph is implicitly defined in the statement.

Let $\mathcal{A}$ be a set of events such that $\Pr(A_i) = p$ for every $A_i \in \mathcal{A}$, Suppose that each event is mutually independent from all but at most d other events. If,

$$ep(d+1) \leq 1 \ , \tag{1.2}$$

then, $\Pr(\cap_{i=1}^{N} \overline{A_i}) > 0$.

The constant e in (1.2) cannot be improved as showed by Shearer [119]. Gebauer, Szabó and Tardos [69] proved that the local lemma is also tight in the context of k-CNF formulas [69]. Their proof is based on the lopsided version of the Lovász Local Lemma, which was introduced by Erdős and Spencer [53] in the context of latin transversals of square integer matrices. In this lopsided version, the condition of mutual independence to build the dependency graph, is relaxed for events that are positively correlated, obtaining a lopsidependency graph (see Definition 2.8). We will make use of this version in the Chapter 3 of this thesis.

However, one of the main drawbacks of the local lemma is that it just takes into account the number of dependencies among the events, but not the strength of these dependencies.

For instance, if $\mathcal{A}$ is a set of “almost” independent events one would expect to be able to derive a lower bound on $\Pr(X = 0)$ similar to (1.1). Unfortunately, the local lemma is not able to provide such a result. As we will see in Section 2.2.2.3, we have other tools that allow us to give a meaningful upper bound on the probability of $X = 0$, even in the case where there are no mutually independent events.

While the local lemma is useful to provide existential results, it does not give the means to construct an object that satisfies the desired property. In this direction, Moser and Tardos [108, 109] propose an algorithmic version of the Lovász Local Lemma, following earlier work by Beck [14] and by Molloy and Reed [104]. This version provides randomized algorithms to find objects for which the standard version of the Lovász Local Lemma can prove their existence. Moreover, the algorithm is efficient; it runs in almost linear time in average. This has been a real breakthrough in the area. The method, also known as entropy compression has been useful to improve certain results where the non constructive version of the local lemma had been already applied earlier [75, 56].

Although upper bounds on the probability of avoiding a set of events $\mathcal{A}$ at the same time do not provide existential results of any kind, for some applications, like enumeration, they are particularly useful. Some applications of it will be seen in Chapters 3 and 6.

Janson [81] introduced a useful inequality, which can be also thought as a concentration inequality (see [8, Theorem 8.7.2]), that we can use to bound from above the probability that $X = 0$. Here, in contrast with the local lemma, pairwise join probabilities have an important role. However, this inequality can be applied just to a certain type of events which, in particular, should be positively correlated (see Theorem 2.15 for the complete statement).

Suen [123] propose a similar version of this inequality where the particular setting of the Janson inequality is not required anymore. This allows us to study any setting, even if the nature of the dependencies is not clear. As Janson inequality, Suen inequality is also sensible to the different pairwise relations among events by taking into account the probabilities $\Pr(A_i \cap A_j)$, when A_i and A_j are not mutually independent.

Since it is stated in a wider context, obviously, Suen inequality is not as strong as Janson inequality and can be used only in the cases where the dependencies among events are neither very numerous nor very strong. This inequality is particularly useful for some cases where dependencies are bounded. Some nice examples of the use of this inequality can be found in [66, 24]. Another good reference of it is the paper of Janson [82].

In fact, if the dependencies are strong, both Janson and Suen inequalities can be worse than the simple application of the second moment method.

A promising contribution was made by Lu and Székely [97, 98]. There, the authors consider a different version of the dependency graph, the so–called ε–near dependency graphs (see Definition 2.9), and adapt the local lemma to give an upper bound for $\Pr(X = 0)$ instead of a lower bound. Although it is not easy to show that a graph is an ε–near dependency graph, the authors provide a good example by considering the problem of finding a perfect matching in the complete graph that avoids a family of "bad" partial matchings. This example is closely related to the lopsidependency graph defined in [53]. By means of their approach, Lu and Székely manage to count the number of regular graphs, latin rectangles or permutations without k–cycles.

As in the case of the local lemma, this last approach does not consider pairwise join probabilities.

However, the definition of ε–near dependency graphs, reduces the number of dependencies to be considered, similarly as in the lopsided case, and can be very useful in some cases.

In conclusion, these tools provide upper and lower bounds for the probability that a certain set of bad events do not occur. Therefore, they can approximate the number of configurations that avoid all these bad events. Hence, they do not just provide existence results, but also enumerative ones.

Some of the material in this thesis has been already published or is to appear in journals [62, 115, 112]. Most of the remaining material has also been published as a preprint in the ArXiv server [63, 114], and is currently submitted for publication. The contributions have also been presented in several conferences and workshops.

This thesis addresses different problems in which the above framework tools are thoroughly exploited: Rainbow matchings of edge colored complete bipartite graphs (Chapter 3), Identifying codes in graphs (Chapters 4 and 5), Consecutive pattern avoiding in permutations (Chapter 6) and the Lonely Runner Conjecture (Chapter 7). In what remains of this chapter, we briefly review each problem of the thesis with an small introduction of the topic, the statement of the main results and some comments on the techniques used to prove them.

1.1 Rainbow Perfect Matchings in Complete Bipartite Graphs

A subgraph H of an edge–colored graph G is rainbow if no color appears twice in the edges of H. In particular, we will focus in rainbow perfect matchings of an edge–colored complete bipartite graph $K_{n,n}$. This case is of particular interest due to the connexion with latin transversals in integer matrices. Recall that a latin transversal is a set of n positions of an $n \times n$ matrix, no two in the same row nor the same column, that contain all different elements.

Our work is motivated by the following longstanding conjecture of Ryser on the existence of latin transversals in latin squares:

Conjecture 1.1 (Ryser Conjecture [118])**.** *Every latin square of odd size admits a latin transversal.*

This conjecture was extended by Stein [121] to $n \times n$ integer matrices containing n copies of each element in $\{1, \dots, n\}$.

An interesting approach on Stein's conjecture was given by Erdős and Spencer [53].

Theorem 1.2 ([53])**.** *Let A be an integer matrix. If every entry in A appears at most $\frac{n-1}{4e}$ times, then A has a latin transversal.*

In Chapter 3 we study the number of rainbow matchings in a given edge–coloring of $K_{n,n}$, where every color appears at most some number of times, thus extending Theorem 1.2.

The techniques used to derive these bounds are inspired by the framework devised by Lu and

Székely [98] to obtain asymptotic enumeration results using the Lovász Local Lemma. The main result in Chapter 3 is the following theorem.

Theorem 1.3. *Given an edge–coloring of $K_{n,n}$ such that no color appears more than n/k times, $k \geq 13.66$, let t be the number of pairs of non–incident edges that have the same color. Then, the number of rainbow perfect matchings is at most*

$$\exp\left(-\left(1-\frac{3}{k}-\frac{60}{k^2}\right)\frac{t}{n(n-1)}\right)n!\ ,$$

and at least

$$\exp\left(-\left(1+\frac{16}{k}\right)\frac{t}{n(n-1)}\right)n!\ .$$

Observe that the dependency on t is natural, since a coloring in which all pairs of monochromatic edges are mutually incident ($t = 0$) has $n!$ rainbow perfect matchings.

Any proper edge–coloring where each color appears exactly n/k times, satisfies $t \sim n^3/2k$.

Corollary 1.4. *Given a proper edge–coloring of $K_{n,n}$ in which each color appears exactly n/k times, $k \geq 13.66$, the number of rainbow perfect matchings is at most $\gamma_2(k)^n n!$ and at least $\gamma_1(k)^n n!$ for some constants $0 < \gamma_1(k) < \gamma_2(k) < 1$ which depend only on k.*

The second part of Chapter 3 is devoted to the study of the existence of rainbow perfect matchings in random edge–colorings. We restrict ourselves to colorings with a fixed number $s = kn$ of colors and we define two natural random models that fit with this condition: the uniform random model, $\mathcal{C}_u(n,s)$, and the regular random model, $\mathcal{C}_r(n,s)$. Analogous results to the one in Theorem 1.3 can be proved for these random models.

Proposition 1.5. *The expected number of rainbow perfect matchings in an edge–coloring of $K_{n,n}$ chosen at random from the $\mathcal{C}_u(n,s)$ (or the $\mathcal{C}_r(n,s)$) with $s = kn$ colors, $k > 1$, is*

$$\exp\left(-\left(k\left((k-1)\ln\left(\frac{k-1}{k}\right)+1\right)+o(1)\right)\frac{n^2}{s}\right)n!\ .$$

For $k = 1$, the expected number is

$$\exp\left(-(1+o(1))n\right)n!\ .$$

Since the edge–coloring is chosen at random, the probability that a perfect matching selected at random is rainbow, is more concentrated than in the case of arbitrary edge–colorings.

Finally, we show that in the $\mathcal{C}_u(n,s)$ model, an edge–coloring has a rainbow perfect matching with high probability.

Theorem 1.6. *An edge–coloring of $K_{n,n}$ chosen at random from the $\mathcal{C}_u(n,s)$ with $s \geq n$ colors, contains a rainbow perfect matching with high probability.*

This result can be easily extended to the $\mathcal{C}_r(n,s)$ using the same ideas. In particular, this implies that the conjecture posed by Stein is true for almost all edge–colorings.

The results of Chapter 3 are joint work with Oriol Serra and can be found in [115].

1.2 Bounds for identifying codes in terms of degree parameters

Given a graph G, an identifying code $\mathcal{C}$ is a dominating set such that for any two vertices, their neighborhoods within $\mathcal{C}$ are nonempty and distinct. This property can be used to distinguish all vertices of the graph from each other. Unlike dominating sets, not every graph can have an identifying code. In fact, it is easy to check that a graph has a identifying code if and only if there are no two adjacent vertices connected to the same set of vertices. For the sake of simplicity, throughout the section we will assume that G admits an identifying code.

Motivated by the applications, given a graph G we want to study the smallest size of an identifying code, also called the identifying number of G and denoted by $\gamma^{\mathrm{ID}}(G)$. The following bounds are known for this number,

$$\log_2(n+1) \leq \gamma^{\mathrm{ID}}(G) \leq n \ .$$

In Chapter 4, we provide bounds on the identifying number in terms of degree-.related graph parameters such as the minimum and maximum degree, denoted by d and Δ respectively. We also focus on the case of d–regular graphs.

In the first part of Chapter 4 we answer partially a question raised in [58] on the size of a minimum identifying code in a graph with bounded maximum degree. It was showed in [86] that if G has maximum degree Δ, then

$$\gamma^{\mathrm{ID}}(G) \geq \tfrac{2n}{d+2} \ .$$

While the proof of this result is straightforward, it does not seem so easy to provide a sharp upper bound. It was conjectured in [61] that the following upper bound holds.

Conjecture 1.7 ([61])**.** *For any connected graph G with maximum degree Δ,*

$$\gamma^{ID}(G) \leq n - \frac{n}{\Delta} + O(1) \ .$$

Graphs with maximum degree Δ that admit an identifying code of size $n - \frac{n}{\Delta}$ are known (see Section 4.5). Thus, if Conjecture 1.7 holds, it is best possible.

It was showed in [59] that $\gamma^{\mathrm{ID}}(G) \leq n - \frac{n}{O(\Delta^5)}$, and $\gamma^{\mathrm{ID}}(G) \leq n - \frac{n}{O(d^3)}$ when G is d-regular. In this thesis we prove an upper bound for $\gamma^{\mathrm{ID}}(G)$ when G has bounded maximum degree. The exact statement of this result can be find in Theorem 4.6. As corollaries of it, we have the following results.

Corollary 1.8. *For any connected graph G with maximum degree Δ,*

$$\gamma^{ID}(G) \leq n - \frac{n}{O(\Delta^3)} \ .$$

In particular, Theorem 4.6 also allows us to prove an asymptotic version of the conjecture for different classes of graphs.

Corollary 1.9 (Regular graphs)**.** *For any connected d–regular graph G,*

$$\gamma^{ID}(G) \leq n - \frac{n}{O(d)} \ .$$

Corollary 1.10 (Graphs excluding complete graphs as minors)**.** *Let $\mathcal{G}_\Delta$ be a the class of connected graphs that have maximum degree Δ and exclude K_k as a minor. Then for every graph $G \in \mathcal{G}_\Delta$,*

$$\gamma^{ID}(G) \leq n - \frac{n}{c(k)d} ,$$

for some constant $c(k)$ that depends on k.

The proof of the main theorem, uses the weighted version of Lovász Local Lemma to show the existence of an identifying code, together with standard concentration bounds, to show that the code, selected at random, is small enough.

In order to understand the behavior of $\gamma^{\text{ID}}(G)$ for d–regular graphs it is worthy to study typical d–regular graphs. The second part of this Chapter is devoted to compute the value of $\gamma^{\text{ID}}(G)$ with high probability, for a d–regular graph chosen uniformly at random. Identifying codes have been previously studied in two models of random graphs, the classic random graph model [66] and the model of random geometric graphs [110]. We will deal with random regular graphs through the so-called Configuration model (see Subsection 2.3.1).

First, it is convenient to give an upper bound for the size of an identifying code in graphs with minimum degree d, when the girth of the graph is at least 5.

Theorem 1.11. *For any graph G with minimum degree d and girth at least 5, we have*

$$\gamma^{ID}(G) \leq (1 + o_d(1))\frac{3\log d}{2d}n .$$

Since d–regular graphs do not have many triangles and 4–cycles, one can adapt the proof to show the following.

Theorem 1.12. *Let G be a d–regular graph chosen uniformly at random, $d \geq 3$. With high probability, we have*

$$\gamma^{ID}(G) = \frac{\log d + O(\log\log d)}{d}n .$$

The results of Chapter 4 are joint work with Florent Foucaud and can be found in [62].

1.3 Large spanning subgraphs admitting small identifying codes

Consider any graph parameter that is not monotone with respect to graph inclusion. Given a graph G, a natural problem in this context is to study the minimum value of this parameter over all spanning subgraphs of G. In particular, how many edge deletions are sufficient in order to obtain from G a graph with optimal value of the parameter? Herein, we study this question with respect to the identifying code number of a graph, a parameter introduced in Section 1.2.

There are very dense graphs that have a huge identifying code number; sparse graphs, such as trees and planar graphs, also have a linear identifying code number [117]. On the other hand, one can also find sparse and dense graphs with logarithmic identifying code number [106, 66], which is the smallest it can be.

This motivates the following question:

Given any sufficiently dense graph, can we delete a small number of edges to get a spanning subgraph with a small identifying code?

This question is related to the notion of resilience of a graph with respect to a graph property $\mathcal{P}$ [122].

Despite being dense, the random graph $G(n,p)$ (for $0 < p < 1$) has a logarithmic size identifying code, as with high probability,

$$\gamma^{\mathrm{ID}}(G(n,p)) = (1+o(1))\frac{2\log n}{\log(1/q)} ,$$

where $q = p^2 + (1-p)^2$ [66]. Comparing this result with the examples of dense graphs with high identifying code number, one can guess that, in a dense graph, the lack of structure implies the existence of a small identifying code number. The following theorem, formalizes this idea.

Theorem 1.13. *For every graph G on n vertices with maximum degree $\Delta = \omega(1)$ and minimum degree $d \geq 66\log\Delta$, there exists a subset of edges $F \subset E(G)$ of size*

$$|F| = O(n\log\Delta) ,$$

such that

$$\gamma^{ID}(G \setminus F) = O\left(\frac{n\log\Delta}{d}\right) .$$

The next theorem shows that Theorem 1.13 cannot be improved much.

Theorem 1.14. *For every $d \geq 2$, there exists a d–regular graph G_n^d on n vertices with the following properties.*

1. For every $M \geq 0$, there exists a constant $c > 0$ such that for every set of edges $F \subset E(G_n^d)$ satisfying $\gamma^{ID}(G_n \setminus F) \leq M\frac{n\log d}{d}$, the size of F is $|F| \geq cn\log d$.

2. For every spanning subgraph H of G_n^d, $\gamma^{ID}(H) = \Omega\left(\frac{n\log d}{d}\right)$.

When $d = \mathrm{Poly}(\Delta)$, Theorem 1.14 shows that Theorem 1.13 is asymptotically tight. Moreover, we also show that the hypothesis of Theorem 1.13 are necessary. There are graphs with bounded Δ and graphs for which $d \leq \log\Delta/2$, such that all their spanning subgraphs have a linear identifying code number.

Since the identifying code number is a non–monotone property, we also consider the case where edges can be added instead of deleted in G. For such a case, analogous results are derived.

In [12], the notion of a watching system has been introduced as a relaxation of identifying codes: in a watching system, code vertices ("watchers") are allowed to identify any subset of their closed neighborhood, and several watchers can be placed in one vertex. Under the hypothesis of Theorem 1.13 and as a corollary of it, we can provide a watching system of size $O\,(n\log\Delta/d)$.

The results of Chapter 5 are joint work with Florent Foucaud and Oriol Serra, and can be found in [63].

1.4 Consecutive pattern avoiding in permutations

A permutation $\pi \in \mathcal{S}_n$ of length n contains $\sigma \in \mathcal{S}_m$ of length m as a consecutive pattern if there exists a set of m consecutive elements in π that have the same relative order as the elements in σ. One interesting problem in the area of pattern avoidance in permutations is to determine the number of permutations in $\mathcal{S}_n$ that do not contain σ as a consecutive pattern.

This problem was introduced by Elizalde and Noy in [50], where they completely determined the asymptotic enumeration of such permutations when $m = 3$.

For every $\sigma \in \mathcal{S}_m$, let $\alpha_n(\sigma)$ be the number of permutations in S_n that avoid σ as a consecutive patterns. Elizalde [48] showed that the limit

$$\rho_\sigma = \lim_{n\to\infty} \left(\frac{\alpha_n(\sigma)}{n!}\right)^{1/n} ,$$

exists, for any $\sigma \in \mathcal{S}_m$.

In [50], the authors stress the importance of the monotone patterns, $(1, 2, \ldots, m)$ and $(m, \ldots, 2, 1)$, in this problem and pose the following conjecture.

Conjecture 1.15 (CMP conjecture [50]). *For every $\sigma \in \mathcal{S}_m$,*

$$\rho_\sigma \le \rho_{(1,2,\ldots,m)} .$$

Recently, this conjecture has been proved by Elizalde [49] using generating functions and the cluster method of Goulden and Jackson [71].

In Chapter 6 we study consecutive patterns in permutations using a completely different approach to the problem, based on the probabilistic tools we provide in Chapter 2. While our approach is not as precise as the generating function technique, it provides simpler alternative proofs of some known results, as the CMP conjecture, and allows us to obtain more general results.

Our first result gives an explicit upper bound for the number of permutations in $\mathcal{S}_n$ avoiding a given σ as a consecutive patterns, when σ is not monotone.

Theorem 1.16. *For every non monotone pattern $\sigma \in \mathcal{S}_m$,*

$$\rho_\sigma \le 1 - \frac{1}{m!} + O\left(\frac{1}{m^2 \cdot m!}\right) .$$

To prove this theorem we make use of Suen Inequality [123] since the number of dependencies in the set of bad events is small.

By comparing the upper bound given by Theorem 1.16 with the result obtained by Elizalde and Noy [51] for monotone patterns we can give an alternative proof of the CMP conjecture as a corollary. Our proof works for any value of $m \ge 5$, but does not provide meaningful results for the case $m = 4$. Moreover, the probabilistic approach also allows us to estimate the difference between the number of permutations avoiding the most and the second most avoided pattern. This last result has not been obtained by means of generating functions.

We also provide general upper and lower bounds for any pattern of length m.

Theorem 1.17. *For every $\sigma \in \mathcal{S}_m$,*

$$1 - \frac{1}{m!} - O\left(\frac{m}{(m!)^2}\right) \leq \rho_\sigma \leq 1 - \frac{1}{m!} + O\left(\frac{1}{m \cdot m!}\right) .$$

To prove the lower bound on the number of permutations avoiding a given pattern we use a one–sided version of the Lovász Local Lemma introduced by Peres and Schlag [116] (see Lemma 2.14). Both bounds are asymptotically tight. An extremal example for the upper bound is provided by monotone patterns and for the lower bound by the pattern $(1, 2, \dots, m-2, m, m-1)$.

As Theorem 1.17 gives bounds for a given σ in terms of m, a natural question is to determine how most of the patterns behave. In this direction a much stronger upper bound, close to the general lower bound, is showed to hold for most of the patterns.

Theorem 1.18. *Let σ be chosen uniformly at random from $\mathcal{S}_m$. Then, with high probability,*

$$\rho_\sigma \leq 1 - \frac{1}{m!} + O\left(\frac{c^m}{(m!)^2}\right) ,$$

where $c > 1$ is some constant.

Theorem 1.18 is stated in the thesis in a more general way (see Theorem 6.6). This theorem shows that, when m is large enough, for most of the patterns ρ_σ is concentrated close to the lower bound provided by Theorem 1.17.

The results of Chapter 6 can be found in [112].

1.5 On the Lonely Runner Conjecture

The Lonely Runner Conjecture was posed independently by Wills [127] in 1967 and Cusick [41] in 1982. Suppose that n runners are running on the unit circle with different speeds and starting at the origin. Then, the conjecture states that for each runner, there is a time where he is at distance at least $1/n$ from all the other runners. Let us denote by $\|x\|$ the distance from x to the closest integer. Then, the conjecture can be restated in the following terms,

Conjecture 1.19 (Lonely Runner Conjecture)**.** *For every $n \geq 1$ and every set of nonzero speeds $v_1, \dots, v_n$, there exists a time t such that*

$$\|tv_i\| \geq \frac{1}{n+1} ,$$

for every $i \in [n]$.

This conjecture was motivated by a problem in diophantine approximation [18, 41], but appears in many different areas such as view–obstruction problems [42], nowhere zero flows [19], chromatic numbers of distance graphs [13] or lacunary sequences [116, 46].

In [21] it is showed that the conjecture can be reduced to the case where all the speeds are integers. In such a case, observe that if the Lonely Runner Conjecture is true, we may assume that the time $t \in (0,1)$, where $(0,1)$ is the unit sphere, since at integer times all the runners are placed again at the origin.

Observe that each runner is a proportion of time 2δ at distance at most δ from the origin, independently from its speed. Here, we exactly compute the time that two runners spend at distance at most δ from the origin at the same time. This time strongly depends on the speeds. In particular, we show that there are many pairs of runners which lie a large amount of time close to the origin.

The previous result allows us to reduce the gap of loneliness. It is straightforward to see that there is a time when all the runners are at distance at least $\delta = \frac{1}{2n}$ from the origin. This result was improved by Chen [35], who showed that, for any set of n positive speeds $v_1, \dots, v_n$, there exists a time t such that

$$\|tv_i\| \geq \frac{1}{2n - 1 + \frac{1}{2n-3}} \, .$$

for every $i \in [n]$.

Our first result improves the result of Chen.

Theorem 1.20. *For every $\varepsilon > 0$, every sufficiently large n and every set of positive speeds $v_1, \dots, v_n$, there exists a time $t \in (0,1)$ such that*

$$\|tv_i\| \geq \frac{1}{2n - 2 + \varepsilon} \, ,$$

for every $i \in [n]$.

For the proof of this theorem we use the computed correlations among pairs of runners and a Bonferroni–type inequality (see Lemma 2.6).

Another consequence of our results is related to the notion of *invisible lonely runner* given by Czerwiński and Grytczuk [44].

Theorem 1.21. *For every sufficiently large n and every set of positive speeds $v_1, \dots, v_n$, there exist $t_1, t_2 \in (0,1)$ and different $j_1, j_2 \in [n]$ such that for any $\ell \in \{1,2\}$,*

$$\|t_\ell v_i\| \geq \frac{1}{n+1} \, ,$$

for any $i \neq j_\ell$.

This theorem provides de existence of two runners that leave the origin almost alone at some time, thus extending the result of [44], if n is large enough.

Finally, we use a representation of the problem through a dynamic circular interval graph, that also allows us to show the existence of two invisible lonely runners at the same time.

Theorem 1.22. *For every sufficiently large n and every set of different speeds $v_1, \dots, v_n$, there exist a time $t \in (0,1)$, $k_1, k_2 \in [n]$ and $j_1, j_2 \in [n]$ such that $k_1 \neq k_2$ and for any $\ell \in \{1,2\}$,*

$$\|t(v_i - v_{k_\ell})\| \geq \frac{1}{n} \, ,$$

for any $i \neq k_\ell, j_\ell$.

The results of Chapter 7 are joint work with Oriol Serra and can be found in [114].

CHAPTER 2

Background and Notation

The aim of this chapter is to set a notation for some basic concepts that will appear in the following chapters. Unless otherwise stated, we will also use standard terminology and notation from probability theory (see e.g. [8]) and graph theory (see e.g. [45]). A background on the probabilistic method is also provided with special emphasis on the most important techniques and tools that will be used all along the thesis.

2.1 Basic notation

We denote by $\mathbb{Z}$ the set of integer numbers, by $\mathbb{Q}$ the set of rational numbers and by $\mathbb{R}$ the set of real number. We use $\mathbb{Z}^+$, $\mathbb{Q}^+$ and $\mathbb{R}^+$ to denote the nonnegative elements of $\mathbb{Z}$, $\mathbb{Q}$ and $\mathbb{R}$ respectively. For any set S, we denote S^d, the cartesian product of d copies of S.

We denote by $[n] = \{1\dots,n\}$ the set of the first n positive integers. For every finite set S we use $|S|$ to denote its cardinality. Then, for every $0 \leq k \leq |S|$, we denote by $\binom{S}{k}$, the family of subsets of S of size k and by 2^S, the family of all the subsets of S. Observe that $\binom{S}{0} = \{\emptyset\}$, where $\emptyset$ denotes the empty set.

The notation $\log(x)$ stands for the natural logarithm of $x > 0$, while, $\log_a(x)$ will denote the logarithm in base $a > 0$ of x. We will sometimes make use of $\exp(x)$ to denote the value of e^x.

For every $f, g : \mathbb{Z}^+ \to \mathbb{R}^+$ we will use the standard asymptotic notation displayed in Table 2.1.

In some occasions we will use either $f(n) = (1 + o(1))g(n)$ or $f(n) = (1 - o(1))g(n)$ instead of $f(n) \sim g(n)$, in order to stress which function majorizes the other one in the limit.

If f and g depend on more than one variable, we use the notations $o_x, O_x, \Theta_x, \Omega_x$ and ω_x to stress the fact that the asymptotic is taken on the variable x.

Finally, we say that a sequence of events A_n in a sequence of finite probability spaces Ω_n holds

$f(n) = O(g(n))$	if $\limsup_{n\to+\infty} \frac{f(n)}{g(n)} < +\infty$.
$f(n) = o(g(n))$	if $\lim_{n\to+\infty} \frac{f(n)}{g(n)} = 0$.
$f(n) = \Omega(g(n))$	if $\limsup_{n\to+\infty} \frac{f(n)}{g(n)} > 0$.
$f(n) = \omega(g(n))$	if $\lim_{n\to+\infty} \frac{g(n)}{f(n)} = 0$.
$f(n) = \Theta(g(n))$	if $f(n) = O(g(n))$ and $f(n) = \Omega(g(n))$.
$f(n) \sim g(n)$	if $\lim_{n\to+\infty} \frac{g(n)}{f(n)} = 1$.

Table 2.1: Asymptotic notation

with high probability if

$$\lim_{n\to+\infty} \Pr(A_n) = 1 \; .$$

2.2 The probabilistic method

In this section we introduce some basic tools of what is known as the probabilistic method, with an special emphasis to the problem of avoiding a set of events. For further details, we recommend the book of Alon and Spencer [8] as a complete monograph on the topic.

2.2.1 Concentration of random variables

One of the main goals of the probabilistic method is to provide upper bounds on the probability that a random variable X is far from its expected value, in order to show that, with high probability, it will be close to $\mathbb{E}(X)$, the expected value of X. For instance, there are many useful inequalities when we can express X as a function of independent random variables X_i, $X = f(X_1, \dots, X_N)$. An interested reader may find an extensive reference on the topic in the recent book of Boucheron, Lugosi and Massart [30]. The study of concentration inequalities is not one of the main goals of this work, although we often use them in our proofs. Thus, here we present a brief overview on the ones that will be used in the forthcoming chapters.

The most fundamental concentration inequality is Markov inequality, which states than for any nonnegative random variable X and $t \geq 1$,

$$\Pr(X > t\mathbb{E}(X)) \leq \frac{1}{t} \ .$$

From it we can derive the following bound, known as Chebyshev inequality or simply the second moment method (e.g. see [8]).

Lemma 2.1 (Chebyshev inequality). *For any random variable X and $t > 0$,*

$$\Pr(|X - \mathbb{E}(X)| \geq t\sigma(X)) \leq \frac{1}{t^2} \ , \tag{2.1}$$

where σ^2 is the variance of X.

Observe that the upper bound on the probability is polynomial on t. Next, we will show that for the same deviation, an exponential bound can be achieved, by assuming some conditions on X.

As we have already mentioned, the concentration of random variables is well studied in the case when $X = f(X_1, \dots, X_N)$ and the variables X_i are independent. From now on, we will focus on the case where $X = X_1 + \dots + X_N$.

If the each random variable X_i follows a Bernoulli distribution with parameter p_i, $X_i \sim \text{Be}(p_i)$, then we can use the Chernoff inequality to bound the tails.

Lemma 2.2 (Chernoff inequality, Corollary A.1.14 in [8]). *Let $X_1, \dots, X_N$ be independent Bernoulli random variables and define $X = \sum_{i=1}^{N} X_i$. Then, for all $\varepsilon > 0$,*

$$\Pr(|X - \mathbb{E}(X)| \geq \varepsilon\mathbb{E}(X)) < 2e^{-c_\varepsilon \mathbb{E}(X)} \ ,$$

where

$$c_\varepsilon = (1+\varepsilon)\log(1+\varepsilon) - \varepsilon \ .$$

The above inequality can also be deduced from Markov inequality.

Similar results hold when the random variables X_i are bounded with probability one (see Hoeffding inequality [79]).

In particular, if the Bernoulli random variables are identically distributed, $X_i \sim \text{Be}(p)$, we can use a slightly better inequality.

Lemma 2.3 (Chernoff inequality for binomial distributions [9]). *Let $X \sim Bin(N,p)$ be a Binomial random variable, then for all $0 < \varepsilon < 1$,*

1. $\Pr(X \leq (1-\varepsilon)Np) < \exp\left(-\frac{\varepsilon^2}{2}Np\right)$.
2. $\Pr(X \geq (1+\varepsilon)Np) < \exp\left(-\frac{\varepsilon^2}{3}Np\right)$.

2.2.2 Avoiding a set of events and the Poisson Paradigm

Let $\mathcal{A} = \{A_1, \dots, A_N\}$ be a set of events in a finite probability space. As it has been seen in the previous chapter, in many cases the existence of a desired object can be expressed as the avoidance of all the events in $\mathcal{A}$ in a probability space that contains a large set of objects. The goal of this section is to show the basic probabilistic techniques to study the following probability,

$$\Pr\left(\bigcap_{i=1}^{N}\overline{A_i}\right) , \tag{2.2}$$

in different contexts.

Recall that two events A_1 and A_2 are independent if $\Pr(A_1 \cap A_2) = \Pr(A_1)\Pr(A_2)$. Then, we say that a set of of events $\mathcal{A}$ is pairwise independent if for any $i \neq j$, $\Pr(A_i \cap A_j) = \Pr(A_i)\Pr(A_j)$. An event A is mutually independent from $\{A_i\}_{j\in S}$ if and only if $\Pr(A \mid \cap_{i\in S'} A_i) = \Pr(A)$ for any $S' \subseteq S$. A set of events $\mathcal{A} = \{A_1, \dots, A_N\}$ is mutually independent if and only if $\Pr(\cap_{i\in S} A_i) = \prod_{i\in S}\Pr(A_i)$ for any $S \subseteq [N]$.

If $\mathcal{A}$ is mutually independent, we have

$$\Pr\left(\bigcap_{i=1}^{N}\overline{A_i}\right) = \prod_{i=1}^{N}(1-\Pr(A_i)) . \tag{2.3}$$

In general, the set $\mathcal{A}$ will not be mutually independent and it will not be straightforward to compute (2.2). The well known inclusion–exclusion formula provides an exact way to get the desired probability.

$$\Pr\left(\bigcap_{i=1}^{N}\overline{A_i}\right) = 1 - \Pr\left(\bigcup_{i=1}^{N} A_i\right) = \sum_{k=0}^{N}(-1)^k \sum_{S\in\binom{[N]}{k}} \Pr\big(\bigcap_{j\in S} A_j\big) . \tag{2.4}$$

The main drawback of this expression is that, in general, it is hard to provide an explicit value for the joint probabilities $\Pr\big(\bigcap_{j\in S} A_j\big)$. This implies that for most of the problems, we will not be able to give an exact expression for (2.2). However, sometimes it suffices to have a good enough estimation of such probabilities.

From (2.4) we can derive the following bounds on (2.2), known as Bonferroni inequalities. For any $0 \leq m \leq N/2$, we have

$$\Pr\left(\bigcap_{i=1}^{N}\overline{A_i}\right) \leq \sum_{k=0}^{2m}(-1)^k \sum_{S\in\binom{[N]}{k}} \Pr\big(\bigcap_{j\in S} A_j\big)$$
$$\Pr\left(\bigcap_{i=1}^{N}\overline{A_i}\right) \geq \sum_{k=0}^{2m+1}(-1)^k \sum_{S\in\binom{[N]}{k}} \Pr\big(\bigcap_{j\in S} A_j\big) .$$

These inequalities are specially useful in the case where the events are almost incompatible since then, the joint probabilities are small and they converge quickly to (2.2).

A particular case of the Bonferroni inequalities is the union bound, which allows us to give a simple lower bound on (2.2),

$$\Pr\left(\bigcup_{i=1}^{N} A_i\right) \leq \sum_{i=1}^{N}\Pr(A_i) ,$$

also written as

$$\Pr\left(\bigcap_{i=1}^{N} \overline{A_i}\right) \geq 1 - \sum_{i=1}^{N} \Pr(A_i) \ . \tag{2.5}$$

The crucial fact that makes the union bound one of the most used inequalities in probabilistic combinatorics is that one does not need to care about the dependencies among the events in $\mathcal{A}$ to deduce a lower bound on (2.2). We will repeatedly use this bound all along the thesis. Even though this bound is sharp (consider a set of disjoint events), it is not a meaningful bound if most of the events in $\mathcal{A}$ have large intersection.

The interested reader in Bonferroni–type inequalities is referred to the book of Galambos and Simonelli [68]. The following inequality can be found there and slightly improves the union bound in the case where the events are not disjoint.

Lemma 2.4 (Inequality I.11 from [68]). *For any tree T with vertex set $V(T) = [N]$, we have*

$$\Pr\left(\bigcap_{i=1}^{N} \overline{A_i}\right) \geq 1 - \sum_{i=1}^{N} \Pr(A_i) + \sum_{ij\in E(T)} \Pr(A_i \cap A_j) \ . \tag{2.6}$$

For any event $A_i \in \mathcal{A}$, it is interesting to consider its associated indicator random variable, X_i and, in particular, the following the random variable,

$$X = \sum_{i=1}^{N} X_i \ . \tag{2.7}$$

As we have already observed in the previous chapter, recall that the expression in (2.2) is equivalent to $\Pr(X = 0)$. Notice that one can upper bound the previous probability using a concentration inequality on the variable X. The probability that $X = 0$ is at most the probability that X deviates $\mathbb{E}(X)$ from $\mathbb{E}(X)$.

If no assumption on the events in $\mathcal{A}$ is done, we can use the second moment method. By setting $t = \mathbb{E}(X)/\sigma(X)$ in Lemma 2.1, we get the following corollary.

Lemma 2.5. *For any random variable X with $\mathbb{E}(X) \neq 0$,*

$$\Pr(X = 0) \leq \frac{\sigma^2(X)}{\mathbb{E}(X)^2} \ .$$

By showing that $\sigma(X)/\mathbb{E}(X) \to 0$, we know that with high probability, $X \neq 0$.

In a general context, Equation (2.5) as well as Lemma 2.5 are tight. Nevertheless, as we explained in the previous chapter, better bounds can be provided if we assume that the set $\mathcal{A}$ has a small number of dependences or these dependences are weak. If one of these conditions holds, intuitively speaking, the probability that any of the events in $\mathcal{A}$ holds, is similar to the case when the events are mutually independent (see (2.3)).

2.2.2.1 Dependency graph

To show that good approximations of (2.2) can be given under certain conditions, the definition of an underlying structure that captures the mutual independence among the elements of $\mathcal{A}$, the so–called dependency graph, is of special interest.

Definition 2.6. *A graph $H = H(\mathcal{A})$ is a dependency graph for the set of events $\mathcal{A} = \{A_1, \dots, A_N\}$ if $V(H) = [N]$, and if $i \in [N]$ is not connected to $S \subseteq [N] \setminus \{i\}$, then*

$$\Pr(A_i \mid \cap_{j\in S}\overline{A_j}) = \Pr(A_i) \ . \tag{2.8}$$

As a consequence, if H is a dependency graph for $\mathcal{A}$, each stable set $S \subseteq V(H)$ indexes a set of mutually independent events, that is, $\Pr(\cap_{i\in S} A_i) = \prod_{i\in S} \Pr(A_i)$.

Observe that the definition of H is not unique but the property of being a dependence graph of $\mathcal{A}$ is monotone by subgraph inclusion. In particular, the complete graph K_N is always a dependency graph for $\mathcal{A}$ but we will be interested in studying edge–minimal dependency graphs. Thus, each time we use a dependency graph for a set $\mathcal{A}$, we must specify its set of edges.

For the sake of convenience, we will denote by μ the expected number of events from $\mathcal{A}$ that are satisfied

$$\mu = \mathbb{E}(X) = \sum_{i=1}^{N} \Pr(A_i) \ .$$

In order to control the dependencies among the events in $\mathcal{A}$, the following two parameters are usually associated to the dependency graph H. To measure the global effect of the dependencies, we consider

$$\Delta^* = \sum_{ij\in E(H)} \Pr(A_i \cap A_j) \ ,$$

and for the local one,

$$\delta^* = \max_{1\le i\le N} \sum_{j:\, ij\in E(H)} \Pr(A_j) \ ,$$

where $E(H)$ denotes the edge set of H.

Let us now state a useful result that provides a good dependency graph if the set of events $\mathcal{A}$ satisfies a certain property.

Observation 2.7 (The Mutual Independence Principle [105])**.** *Let $\mathcal{Y} = \{Y_1, \dots Y_M\}$ be a set of independent random experiments. Suppose that $\mathcal{A} = \{A_1, \dots, A_N\}$ is a set of events where each A_i is determined by a subset of experiments indexed by $\mathcal{F}_i \subseteq [M]$. For any $i \in [N]$ and $S \subset [N]$, A_i is mutually independent from $\{A_j\}_{j\in S}$ if $\mathcal{F}_i \cap (\cup_{j\in S}\mathcal{F}_j) = \emptyset$.*

Equivalently, the graph with vertex set $[N]$ where ij is an edge if and only if $\mathcal{F}_1 \cap \mathcal{F}_2 \neq \emptyset$, is a dependency graph for the set of events $\mathcal{A}$.

Some other special types of dependency graphs will be used in this work. Erdős and Spencer [53] introduced the notion of lopsidependency.

Definition 2.8. *A graph H is a lopsidependency graph (also known in the literature as negative dependency graph [98]) for the set of events $\mathcal{A}$ if $V(H) = [N]$ and, if $i \in [N]$ and $S \subseteq [N] \setminus \{i\}$ share no edges, then*

$$\Pr(A_i \mid \cap_{j\in S}\overline{A_j}) \le \Pr(A_i) \ . \tag{2.9}$$

Observe that any dependency graph for $\mathcal{A}$ is also a lopsidependency graph. The idea behind this definition is that positive correlation among events increases the variance and thus, it increases the probability that $X = 0$. For that reason, if we want to give a lower bound on (2.2), positively correlated events do not need to be counted as "depending events".

To provide upper bounds on (2.2), it is worthy to set the notion of positive dependency. This was recently introduced by Lu and Székely [97, 98] in the context of events defined by matchings in complete graphs or complete bipartite graphs.

Definition 2.9. *For any $\varepsilon > 0$, a graph H is an ε-near positive dependency graph for the set of events $\mathcal{A} = \{A_1, \dots, A_N\}$ if $V(H) = [N]$, and the following conditions are satisfied:*

i) $\Pr(A_i \cap A_j) = 0$ *for each $ij \in E(H)$, and*

ii) if $i \in [N]$ and $S \subseteq [N] \setminus \{i\}$ share no edges, then

$$\Pr(A_i \mid \cap_{j \in S} \overline{A_j}) \geq (1 - \varepsilon) \Pr(A_i) \; . \tag{2.10}$$

Condition *i*) implies that only incompatible events can be connected. Condition *ii*) says that the non–occurrence of any set of non–neighbors can not shrink the probability of A_i too much. Thus the event A_i is almost negatively correlated from the events in the set S. The intuitive idea behind this dependency graph is that the only bad dependencies when upper bounding (2.2) are given by positively correlated events

events that are negative correlated and compatible with a given one, should not be considered.

like in the definition of lopsidependency, is that, when , the

2.2.2.2 Lower bounds and the Lovász Local Lemma

In this subsection we will show how to bound from below the probability in (2.2), conditioned to the structure of a dependency graph for $\mathcal{A}$. The classical tool for such a purpose is the Lovász Local Lemma which was introduced by Erdős and Lovász [52] in 1973.

All along this subsection we will consider that H is a dependency graph for $\mathcal{A}$, unless otherwise stated. Let us begin by giving its standard version.

Lemma 2.10 (Lovász Local Lemma (LLL))**.** *Let $\mathcal{A} = \{A_1, \dots, A_N\}$ be a set of events and let H be a dependency graph for $\mathcal{A}$.*

If there exist some constants $x_1, x_2, \dots, x_N \in (0, 1)$ such that

$$\Pr(A_i) \leq x_i \prod_{j:\, ij \in E(H)} (1 - x_j) \; , \tag{2.11}$$

for each $i \in [N]$, then, for each $T \subset [N] \setminus \{i\}$ we have

$$\Pr\big(A_i \big| \bigcap_{j \in T} \overline{A_j}\big) \leq x_i \; .$$

In particular, for each pair $S, T \subset [N]$ of disjoint sets we have

$$\Pr\big(\bigcap_{i\in S}\overline{A_i} | \bigcap_{j\in T}\overline{A_j}\big) \geq \prod_{i\in S}(1-x_i)\ , \tag{2.12}$$

and

$$\Pr\left(\bigcap_{i=1}^{N}\overline{A_i}\right) > \prod_{i=1}^{N}(1-x_i)\ . \tag{2.13}$$

The same lemma can be also stated when H is a lopsidependency graph. This was noticed by Erdős and Spencer [53].

Lemma 2.11 (Lopsided Lovász Local Lemma (LLLL))**.** *Let $\mathcal{A} = \{A_1, \dots, A_N\}$ be a set of events and let H be a lopsidependency graph for $\mathcal{A}$.*

If there exist some constants $x_1, x_2, \dots, x_N \in (0,1)$ such that

$$\Pr(A_i) \leq x_i \prod_{j:\, ij\in E(H)}(1-x_j)\ , \tag{2.14}$$

for each $i \in [N]$, then, for each $T \subset [N] \setminus \{i\}$ we have

$$\Pr\big(A_i | \bigcap_{j\in T}\overline{A_j}\big) \leq x_i\ .$$

In particular, for each pair $S, T \subset [N]$ of disjoint sets we have

$$\Pr\big(\bigcap_{i\in S}\overline{A_i} | \bigcap_{j\in T}\overline{A_j}\big) \geq \prod_{i\in S}(1-x_i)\ , \tag{2.15}$$

and

$$\Pr\left(\bigcap_{i=1}^{N}\overline{A_i}\right) > \prod_{i=1}^{N}(1-x_i)\ . \tag{2.16}$$

Next, we present a symmetric version that can be easily derived from Lemma 2.10 and which can be used when all the events in $\mathcal{A}$ play the same role.

Lemma 2.12 (Symmetric Lovász Local Lemma)**.** *Let $\mathcal{A} = \{A_1, \dots, A_N\}$ be a set of events and let H be a dependency graph for $\mathcal{A}$.*

If $\Pr(A_i) = p$ for each $i \in [N]$, H has maximum degree $\Delta = \Delta(H)$, and

$$ep(\Delta+1) \leq 1\ , \tag{2.17}$$

where $e \approx 2.718$, then

$$\Pr\left(\bigcap_{i=1}^{N}\overline{A_i}\right) \geq (1-ep)^N\ .$$

Observe that condition (2.17) is very similar to the condition $e\delta^* \leq 1$.

The following version of the local lemma is also a corollary from Lemma 2.10. It can be used in the case we can assign a weight to each event in $\mathcal{A}$ and set the value of x_i as a function of the corresponding weight.

Lemma 2.13 (Weighted Lovász Local Lemma [105]). *Let $\mathcal{A} = \{A_1, \dots, A_N\}$ be a set of events and let H be a dependency graph for $\mathcal{A}$.*

If there exist some integer weights $t_1, \dots, t_N \geq 1$ and a real $p \leq \frac{1}{4}$ such that for each $i \in [N]$:

- $\Pr(A_i) \leq p^{t_i}$, *and*
- $\sum_{j:\, ij \in E(H)} (2p)^{t_j} \leq \frac{t_i}{2}$,

then

$$\Pr\left(\bigcap_{i=1}^{N} \overline{A_i}\right) \geq \prod_{i=1}^{N} (1 - (2p)^{t_i}) . \tag{2.18}$$

By giving a total order on the set $\mathcal{A}$ we can derive a directed version of the local lemma. One particular case was proposed by Peres and Yuval [116] in the context of lacunary integer sequences.

Lemma 2.14 (One–sided Local Lemma [116]). *Let $\mathcal{A} = \{A_1, \dots, A_N\}$ be a set of events and let H be a dependency graph for $\mathcal{A}$.*

If there exist some real numbers $x_1, x_2, \dots, x_N \in (0, 1)$ such that for each $i \in [N]$ there is an integer $0 < m(i) \leq i$ such that

$$\Pr\big(A_i \mid \bigcap_{k<m(i)} \overline{A_k}\big) \leq x_i \prod_{j=m(i)}^{i-1} (1 - x_j) . \tag{2.19}$$

Then,

$$\Pr\left(\bigcap_{i=1}^{N} \overline{A_i}\right) \geq \prod_{i=1}^{N} (1 - x_i) . \tag{2.20}$$

2.2.2.3 Upper bounds

We can bound (2.2) from above by using standard concentration inequalities, around the expected value of X, as showed in Lemma 2.5. However, these bounds are far from being tight when X is the sum of almost independent variables. Here we present some of them that will be used in this thesis.

Janson inequality [81] is a good tool to deal with that kind of problems when the elements of $\mathcal{A}$ are some special type of positively correlated events. Although we will not use it explicitly in this work we think it is worthy to state it in order to compare it with Suen inequality.

Theorem 2.15 (Janson Inequality [81]). *Choose $B_1, \dots, B_N$ subsets of a finite set Ω. Let $S \subseteq \Omega$ where each element $\omega \in \Omega$ belongs to S independently with some probability. Define the events $A_i \in \mathcal{A}$ for any $i \in [N]$ as $B_i \subseteq S$ and let H be the dependency graph for $\mathcal{A}$ where ij is an edge if and only if $B_i \cap B_j \neq \emptyset$. Then,*

$$\Pr\left(\bigcap_{i=1}^{N} \overline{A_i}\right) \leq e^{-\mu+\Delta^*} .$$

Moreover, if $\Delta^* \geq \mu$,

$$\Pr\left(\bigcap_{i=1}^{N} \overline{A_i}\right) \leq e^{-\mu^2/2\Delta^*} . \tag{2.21}$$

Observe that in this case, an edge–minimal dependency graph H is given by $ij \in E(H)$ if and only if $B_i \cap B_j \neq \emptyset$. Janson [81] also gave an interesting one–way large deviation inequality extending the previous one.

A bound with the same spirit was proposed by Suen in [123], although we refer the interested reader to the nice paper of Janson [82].

Theorem 2.16 (Suen Inequality, Theorem 2 in [82]). *Let $\mathcal{A} = \{A_1, \dots, A_N\}$ be a set of events and let H be a dependency graph for $\mathcal{A}$. Then,*

$$\Pr\left(\bigcap_{i=1}^{N} \overline{A_i}\right) \leq e^{-\mu+\Delta^* e^{2\delta^*}} . \tag{2.22}$$

Since $\delta^* > 0$, Suen inequality is always worse than Janson inequality. Moreover, it is useless if $\Delta^* e^{2\delta^*} \geq \mu$ which happens to be the case for many problems where the dependency graph contain many edges or has a large maximum degree.

Nevertheless, it is much more general since no assumptions on the set of events $\mathcal{A}$ is done. Thus, it will be extremely useful for (locally) sparse dependency graphs.

One important remark that will carry some consequences in Chapter 6 is the fact that Suen inequality takes into account the pairwise join probabilities among sets of mutually depending events. That makes the inequality sensitive to the relation among pairs of events. On the other hand, the local lemma is our main tool to provide a lower bound on 2.2 when there are not many dependences in $\mathcal{A}$. Unfortunately, it can not distinguish the different nature of the dependencies among events, but just the number of local dependencies.

Finally, as we suggested in Subsection 2.2.2.1, we will show how to use ε-near positive dependency graphs to provide upper bounds on (2.2).

Theorem 2.17 (Lu and Székely [98]). *Let $\mathcal{A} = \{A_1, \dots, A_N\}$ be a set of events with an ε-near-positive dependency graph H. Then,*

$$\Pr\left(\bigcap_{i=1}^{N} \overline{A_i}\right) \leq \prod_{i=1}^{N} (1 - (1-\varepsilon)\Pr(A_i)) \leq e^{-(1-\varepsilon)\mu}.$$

The main drawback of this approach is that, typically, it is hard to prove the conditions needed for a ε-near-positive dependency graph. In the case that such graph can be settled, Theorem 2.17 provides a meaningful upper bound for (2.2) even when there are many dependencies among the events. A good example of that will be showed in Chapter 3.

2.3 Models of Random Combinatorial Structures

The idea behind the proofs that use the probabilistic combinatorics is to set a probability space on a finite set of combinatorial objects and to show that an object with the desired condition exists with some probability. For this reason, it is crucial to study different models of combinatorial structures where the previous techniques can be applied.

In Chapter 6 we will study a problem in permutations. For this reason it is worthy to consider the following model of random permutations.

A permutation $\pi = (\pi_1, \dots, \pi_n)$ is an ordering of the set $[n]$. The symmetric group on n elements, denoted by $\mathcal{S}_n$, is the set of $n!$ permutations of length n. For any sequence of distinct positive real numbers $X = (x_1, \dots, x_m)$, we define the standardization of X, $\mathrm{st}(x_1, \dots, x_m) = \sigma \in \mathcal{S}_m$ such that $x_i < x_j$ if and only if $\sigma_i < \sigma_j$, that is the permutation on m elements that maintains the same relative order. For instance, the sequence $(3, 8, 4, 1)$ standardizes to $\mathrm{st}(3, 8, 4, 1) = (2, 4, 3, 1)$.

A permutation π selected uniformly at random from $\mathcal{S}_n$ satisfies the following useful property: for each integer $m > 0$, every set of positions $i_1, \dots, i_m$ and every $\sigma \in \mathcal{S}_m$,

$$\Pr(\mathrm{st}(\pi_{i_1}, \dots, \pi_{i_m}) = \sigma) = \frac{1}{m!} .$$

For the sake of convenience, we introduce the following model that generates each permutation of $\mathcal{S}_n$ with equal probability. Let $Z_1, \dots, Z_n$ be independent uniform random variables in $(0, 1)$ and let $\pi = \mathrm{st}(Z_1, \dots, Z_n)$. Observe that with probability exactly one all the random variables are different and thus, π is well defined almost surely. This model is of special interest since we can exploit the independence property to study random permutations.

2.3.1 Models of Random Graphs

We devote this subsection to briefly review some models of random graphs that will appear throughout the thesis.

Let us first introduce the standard model of random graphs. The Erdős–Rényi random graph model, denoted as $G(n, p)$, is the probability space over the set of all labeled graphs on n vertices, where a given graph $G \in G(n, p)$ appears with probability

$$\Pr(G) = p^{|E(G)|}(1-p)^{\binom{n}{2}-|E(G)|} .$$

All the graphs with the same number of edges have the same probability to appear. In particular, in $G(n, 1/2)$ each labeled graph on n vertices appears with the same probability, $2^{-\binom{n}{2}}$. Besides, a random instance of $G(n, p)$ can be obtained by placing each edge independently with probability p. These independence is very useful when analyzing the behavior of random graphs.

Although we will not study the Erdős–Rényi model of random graphs, we will often refer to it as example. An comprehensive study of random graphs and their main parameters can be found in the monographs of Bollobás [23] and Janson, Łuczak and Ruciński [83]. We also recommend the Chapter 10 from the book of Alon and Spencer [8] for a nice introduction to the topic.

One of the goals of Chapter 4 is the study of identifying codes on random regular graphs. Thus, we want to set an easy to handle model, $G(n, d)$, that provides d–regular graphs on n vertices with a uniform distribution. Unfortunately, it is not known how to explicitly generate a uniformly random sample in $G(n, d)$.

One of the most common techniques to study random d–regular graphs is to set a probability space over a larger class of multigraphs denoted by $G^*(n, d)$. This is known as the Configuration Model [15, 22].

In this model, a d–regular multigraph on n vertices is obtained by selecting some perfect matching of K_{nd} at random (see [129] for further reference). We will only consider cases where nd is even, as otherwise there does not exist any d–regular graph on n vertices. In the Configuration Model, the set of vertices in K_{nd} is partitioned into n cells of size d and each cell W_v is associated to a vertex v of the random regular graph. An edge e of a perfect matching of K_{nd} induces either a loop in v (if it connects two elements of W_v) or an edge between v and u (if it connects a vertex from W_v to a vertex in W_u). This model produces different multigraphs with different probability, depending on the number of loops and multiedges. Since every d–regular graph on n vertices has neither loops nor multiedges, each of them is produced with the same probability.

It is showed in [129] that the probability the Configuration Model generates a simple graph satisfies

$$\Pr\big(G \text{ is simple}\big) = (1 + o(1))e^{\frac{1-d^2}{4}} \qquad \text{if } d = o(\sqrt{n}). \tag{2.23}$$

Thus, for constant d any property that holds with probability tending to 1 for $G^*(n, d)$ as $n \to \infty$, will also hold with probability tending to 1 for $G(n, d)$.

If $d \to +\infty$ when $n \to +\infty$, one must show that the statements in $G^*(n, d)$ hold with high probability; in particular, with probability large enough to beat (2.23). In this case, another useful tool for such values of d is the Switching method (see [102, 93, 39, 113]).

Finally we introduce another model of random graphs that generalizes the Erdős–Rényi model. Let G be a labeled graph and $p < 1$, then the graph G_p is an element from a probability space over the spanning subgraphs of G, where each labeled subgraph $H \subseteq G$ appears with probability

$$\Pr(G) = p^{|E(H)|}(1-p)^{|E(G)|-|E(H)|} .$$

As before, an element H of this probability space can be modeled by deleting each edge from G independently with probability $1 - p$. A well–known instance of this model is the classical Erdős–Rényi random graph $G(n, p)$, where $G = K_n$. Many results that are showed in the Erdős–Rényi model $G(d, p)$ can be translated to the G_p model, provided that G has minimum degree at least d. The model G_p has been widely studied [37, 65, 91, 92]. Alon [3] generalized this notion by deleting each edge $uv \in E(G)$ with different probabilities p_{uv}. We will use this model in Chapter 5 to show the existence of some special subgraphs.

CHAPTER 3

Rainbow Perfect Matchings in Complete Bipartite Graphs

3.1 Introduction

A subgraph H of an edge–colored graph G is *rainbow* if no color appears twice in is edges. The study of rainbow subgraphs has a large literature; see e.g. [6, 64, 84, 85, 95]. In this chapter of the thesis we deal with *rainbow perfect matchings* of an edge–colored complete bipartite graph $K_{n,n}$. Edge–colored complete bipartite graphs $K_{n,n}$ are equivalent to integer matrices of size $n \times n$ (also called *n–squares*), and the problem of finding a rainbow perfect matching is equivalent to finding a latin transversal of length n in the corresponding n–square (that is, a set of n pairwise distinct entries no two in the same row nor the same column). If an n–square contains exactly n copies of each entry, it is called *equi–n–square*. In particular, proper edge–colorings of $K_{n,n}$ with n colors are equivalent to latin squares, an interesting subclass of equi–n–squares. The following is a longstanding conjecture of Ryser [118] on the existence of latin transversals in latin squares:

Conjecture 3.1 (Ryser)**.** *Every latin square of odd order admits a latin transversal.*

The above conjecture is not true for even order latin squares. For instance, the latin square $A = (a_{ij})$ where $a_{ij} = i + j \pmod n$ contains no latin transversals for even n. Nevertheless, it was also conjectured by Brualdi that every latin square has a partial latin transversal, a set of pairwise distinct entries no two in the same row nor column, of length $n-1$. This conjecture was extended by Stein [121] to equi–n–squares. Recently, all these conjectures have been generalized by Aharoni et al. [2].

There are different approaches to address these conjectures. For instance, Hatami and Shor [77] proved that every latin square has a partial transversal of size $n - O(\log^2 n)$. Snevily [120] conjectured that every subsquare of the addition table of an abelian group of odd order has a latin transversal. This conjecture was eventually proved by Arsovski [10]. Another approach was given by Erdős and Spencer [53]. They proved the following result:

Theorem 3.2 ([53]). *Let A be an n–square. If every entry in A appears at most $\frac{n-1}{4e}$ times, then A has a latin transversal.*

In order to get the above result the authors developed the *lopsided* version of the Lovász Local Lemma. The main idea of this version is to generalize the dependency graph through the so called *lopsidependency graph*. In this graph, non edges may no longer represent mutual independence, and the hypothesis of the Local Lemma is replaced by a weaker assumption (see Definition 2.8).

In this chapter we address two problems: first, the asymptotic enumeration of rainbow perfect matchings in a given edge–coloring of $K_{n,n}$, and second, the existence of rainbow perfect matchings in random edge–colorings of $K_{n,n}$. We consider not necessarily proper edge–colorings, but the asymptotic enumeration applies to proper ones as well.

Theorem 3.2 gives sufficient conditions on the existence of at least one latin transversal. One of the goals of this work is to show that, under only slightly stronger assumptions, we can estimate the number of latin transversals. Although there is no specific conjecture on the number of latin transversals of a latin square, Vardi [125] proposed the following conjecture for the particular class of addition tables of cyclic groups.

Conjecture 3.3 ([125]). *Let $z(n)$ be the number of latin transversals in the table of the cyclic group of order n. Then, there exist two constants $0 < c_1 < c_2 < 1$ such that*

$$c_1^n n! \leq z(n) \leq c_2^n n!\ ,$$

for all odd n.

Recall that $z(n) = 0$ if n is even. In a more general setting, McKay, McLeod and Wanless [100] showed that $c_2 < 0.614$. Giving a lower bound on $z(n)$ is still an open problem. It is conjectured in [40] that

$$z(n) \sim c^n n!\ , \tag{3.1}$$

with $c \approx 0.39$.

Under the hypothesis of Theorem 3.2, we provide upper and lower bounds for the number of rainbow perfect matchings in an edge–colored $K_{n,n}$. The techniques used to derive these bounds are inspired by the framework devised by Lu and Székely [98] to obtain asymptotic enumeration results using the Lovász Local Lemma (see Section 2.2.2).

Our first result gives an asymptotic estimation of the probability that a random perfect matching is rainbow.

Theorem 3.4. *Consider an edge–coloring of $K_{n,n}$ such that no color appears more than n/k times. Let $\mathcal{M}$ denote the family of pairs of non–incident edges that have the same color and let M be a perfect matching of $K_{n,n}$ chosen uniformly at random. Denote by X_M the indicator random variable of the event that M is rainbow and let $\mu = |\mathcal{M}|/n(n-1)$.*

If $k \geq 13.66$, then there exist constants $0 < c_1(k) < 1 < c_2(k)$ depending only on k such that

$$e^{-c_2(k)\mu} \leq \Pr(X_M = 1) \leq e^{-(c_1(k)+o(1))\mu}\ .$$

In the proof of Theorem 3.4 we obtain the following explicit constants: $c_1(k) = 1 - 3/k - 60/k^2$ and $c_2(k) = 1 + 16/k$. In particular, the lower bound on Theorem 3.4 holds for every $k \geq 10.93$.

We note that the existence of a rainbow perfect matching in Theorem 3.2 is ensured with the smaller value $k \geq 4e \approx 10.87$. We also observe that the bounds on the probability of a rainbow perfect matching in Theorem 3.4 depend only on the cardinality of $\mathcal{M}$, but not on the particular structure of the pairs of monochromatic non–incident edges composing $\mathcal{M}$. The dependency on $|\mathcal{M}|$ is natural, since an edge–coloring in which all pairs of monochromatic edges are mutually incident ($|\mathcal{M}| = 0$) has $n!$ rainbow perfect matchings.

In particular, observe that any proper edge–coloring where each color appears exactly n/k times, satisfies $|\mathcal{M}| = (1 + o(1))n^3/2k$, provided that $k = o(n)$. This implies the following corollary of Theorem 3.4, which has the same form as Conjecture 3.3.

Corollary 3.5. *Let $r(C)$ be the number of rainbow perfect matchings in a proper edge–coloring C of $K_{n,n}$ in which each color appears exactly n/k times, $k \geq 13.66$. Then*

$$\gamma_1(k)^n n! \leq r(C) \leq \gamma_2(k)^n n! \ .$$

for some constants $0 < \gamma_1(k) < \gamma_2(k) < 1$ which depend only on k.

The results in Theorem 3.4 require the condition $k \geq 13.66$. By using this probabilistic approach, it seems difficult to drop this condition to $k \geq 1$, in order to cover equi–n–squares and latin squares. This prompts us to ask what can we say about most edge–colorings of $K_{n,n}$ in the case $k \geq 1$. Observe that we cannot use less than n colors. Thus we study the existence of rainbow perfect matchings in *random* edge–colorings. We restrict ourselves to colorings with a fixed number $s = kn$ of colors. We define two natural random models that fit with this condition.

In the Uniform random model, $\mathcal{C}_u(n, s)$, each edge gets one of the s colors independently and uniformly at random. In this model, every possible edge–coloring with at most s colors appears with the same probability. In the Regular random model, $\mathcal{C}_r(n, s)$, we choose an edge–coloring uniformly at random among all the equitable edge–colorings using s colors. Recall that a coloring is called *equitable* if the size of the color classes differ in at most one. For the sake of simplicity, we will assume that s divides n^2. Although the models have the same expected behavior, we consider that both are interesting to analyze. Analogous results to the one in Theorem 3.4 can be proved for these random models.

Proposition 3.6. *Let C be a random edge–coloring of $K_{n,n}$ in the model $\mathcal{C}_u(n, s)$ (or $\mathcal{C}_r(n, s)$) with $s = kn$ colors ($k > 1$) and let M any matching of $K_{n,n}$. Then,*

$$\Pr(X_M = 1) = e^{-(c(k)+o(1))\frac{n^2}{s}} \ ,$$

where $c(k) = 2k\left((k-1)\log\left(\frac{k-1}{k}\right) + 1\right)$.

For $k = 1$, we have $\Pr(X_M = 1) = e^{-(2+o(1))\mu}$.

Obviously, we have that $c_2(k) < c(k) < c_1(k)$, for any $k \geq 13.66$, where $c_1(k)$ and $c_2(k)$ are the constants appearing in Theorem 3.4. It is worthy of notice that, for any $k \geq 1$, we have $c(k) > 1$. Observe also that when $s = n$, the number of rainbow perfect matchings is *w.h.p.* around $e^{-n}n!$. Since $e^{-1} \approx 0.368$, if (3.1) holds, then an edge–coloring induced by a cyclic group of odd order would contain more rainbow perfect matchings than a typical edge–coloring of the complete bipartite graph.

Since now, the edges–colorings are random, we have a stronger concentration of the rainbow perfect matching probability than in the case of arbitrary edge–colorings. By using the random model $\mathcal{C}_u(n,s)$ we can show that for any $s \geq n$, almost all edge–colorings have a rainbow perfect matching.

Theorem 3.7. *An edge–coloring of $K_{n,n}$ chosen at random from the model $\mathcal{C}_u(n,s)$ with $s \geq n$ colors, contains a rainbow perfect matching with high probability.*

To prove Theorem 3.7 we use the second moment method on the random variable that counts the number of rainbow perfect matchings when the edge–coloring is chosen according to the model $\mathcal{C}_u(n,s)$. This result can be proved for the model $\mathcal{C}_r(n,s)$ using the same ideas. In particular, this implies that the Ryser conjecture is true with high probability for equi–n–squares.

This chapter is organized as follows. In Section 3.2 we provide a proof for Theorem 3.4. The random coloring models are defined in Section 3.3, where we also prove Proposition 3.6. Theorem 3.7 is proved in Section 3.4. Finally, in Section 3.5, we give some remarks and discuss some open problems on rainbow perfect matchings that arise from our work.

3.2 Asymptotic enumeration of rainbow matchings

In this section we prove Theorem 3.4. The theorem provides exponential upper and lower bounds for the probability that a random perfect matching in an edge–colored complete bipartite graph is rainbow.

3.2.1 Lower bound

For a given perfect matching, the property of being rainbow can be expressed in terms of the non occurrence of certain partial matchings. One of the standard tools to give a lower bound for the probability of the existence of a structure that avoids some given bad events is the Lovász Local Lemma. As it is showed in [53], it is convenient in our current setting to use the Lopsided version of Lemma 2.10.

Recall that $\mathcal{M}$ denotes the family of pairs of non–incident edges that have the same color in a given edge–coloring of $K_{n,n}$. For each such pair $\{e,f\} \in \mathcal{M}$, let $A_{e,f}$ (or $A_{f,e}$) denote the event that the pair $\{e,f\}$ belongs to the random perfect matching M. We define $\mathcal{A}_{\mathcal{M}}$ to be the set of events $A_{e,f}$ for any $\{e,f\} \in \mathcal{M}$. Consider the following dependency graph:

Definition 3.8. *The rainbow dependency graph H has the family $\mathcal{M}$ as its vertex set. Two elements in $\mathcal{M}$ are adjacent in H if they contain at least two incident edges in $K_{n,n}$, that is they are incompatible.*

Consider the graph H' obtained from the graph H by adding an edge between two matchings if they have one common edge and they are compatible. Erdős and Spencer [53] showed that H' is a lopsidependency graph for $\mathcal{M}$. By Theorem 3 in [98] we have that H is also a lopsidependency graph for the set of events A_M with $M \in \mathcal{M}$. The following lower bound can be obtained in a similar way to Lu and Székely [98, Lemma 2].

Lemma 3.9. *Given an edge–coloring of $K_{n,n}$ where each color appears at most n/k times and an arbitrary color c, let $\mathcal{M}_0 = \mathcal{M}_1 \cup \mathcal{M}_2$ be a set of matchings, where the elements of $\mathcal{M}_1$ are single edges with color c and the elements of $\mathcal{M}_2$ are monochromatic pairs of nonincident edges. We denote by m_0 the size of $\mathcal{M}_0$.*

If $k \geq 13.66$ and $n > 40000$, then for every disjoint sets $S, T \subseteq [m_0]$

$$\Pr\left(\bigcap_{i\in S} \overline{A_i} | \bigcap_{j\in T} \overline{A_j}\right) \geq e^{-(1+20/k)\sum_{i\in S}\Pr(A_i)} . \tag{3.2}$$

In particular, if $S = [m_0]$ and $T = \emptyset$, we have

$$\Pr\left(\bigcap_{i=1}^{m_0} \overline{A_i}\right) \geq e^{-(1+20/k)\mu} , \tag{3.3}$$

where $\mu = \sum_{i=1}^{m_0} \Pr(A_i)$.

Moreover, if the color c does not appear in $E(K_{n,n})$, for every constant $k \geq 10.93$ and $n > 200$,

$$\Pr\left(\bigcap_{i=1}^{m_0} \overline{A_i}\right) \geq e^{-(1+16/k)\mu} . \tag{3.4}$$

Proof. Set $\mathcal{A}_{\mathcal{M}_0} = \{A_1, \dots, A_{m_0}\}$, where $A_i = A_M$ for some $M \in \mathcal{M}_0$. Since each color does not appear many times, we have $|\mathcal{M}_1| \leq n/k$. Each of the n^2 edges belongs to at most $n/k - 1$ matchings in $\mathcal{M}_2$. Thus,

$$|\mathcal{M}_2| \leq \frac{n^2(n/k-1)}{2} = \frac{n^2(n-k)}{2k} .$$

If $M \in \mathcal{M}_1$, we have $p_1 = \Pr(A_M) = 1/n$ whereas if $M \in \mathcal{M}_2$, $p_2 = \Pr(A_M) = \frac{1}{n(n-1)}$. Therefore,

$$\mu = \frac{|\mathcal{M}_1|}{n} + \frac{|\mathcal{M}_2|}{n(n-1)} \leq \frac{n}{2k} , \tag{3.5}$$

if n is large enough with respect to k.

Let us set $t = |\mathcal{M}_1|/n + 4/k$. If no edge is colored with color c, that is $|\mathcal{M}_1| = 0$, we have $t = 4/k$. Then it can be checked that for every $k \geq 10.93$ $n \geq 200$ and $p \leq p_2$

$$pe^{(1+4t)t} < 1 - e^{-(1+4t)p} . \tag{3.6}$$

If $|\mathcal{M}_1| > 0$, then $t \leq 5/k$ and for every $k \geq 13.66$, $n > 40000$ and $p \leq p_1$, (3.6) is also satisfied.

In both cases, we can choose $x_i \in (\Pr(A_i)e^{(1+4t)t}, 1 - e^{-(1+4t)\Pr(A_i)})$ for each $i \in [m_0]$. Observe that the maximum number of matchings in $\mathcal{M}_2$ that are incompatible with a given matching M

is at most $2|M|n(n-1)/k$: given a matching $M \in \mathcal{M}_2$, there are at most $2|M|n$ possibilities to select an edge e incident to either some edge in M, and at most $n/k-1$ choices for a second edge f with the same color than e. Hence, for any $i \in [m_0]$, we have that

$$\sum_{ij \in E(H)} \Pr(A_j) \leq |\mathcal{M}_1| \cdot \frac{1}{n} + \frac{4n(n-1)}{k} \cdot \frac{1}{n(n-1)} = t\ .$$

Using the previous inequalities, for any $i \in [m_0]$, we have

$$\Pr(A_i) < x_i e^{-(1+4t)t} < x_i \prod_{ij \in E(H)} e^{-(1+4t)\Pr(A_j)} < x_i \prod_{ij \in E(H)} (1-x_j)\ . \tag{3.7}$$

Let H_0 be the graph on the set of events indexed by $\mathcal{M}_0$, where two vertices are adjacent if the corresponding matchings are incompatible. By Theorem 3 in [98], H_0 is a lopsidependency graph for $\mathcal{M}_0$. Hence, we can use the Lopsided version of the Local Lemma (see Lemma 2.11), in particular (2.15) and (2.16), to finish the proof of the lemma. □

3.2.2 Upper bound

To provide an upper bound on the number of rainbow matchings we use the new enumeration tool provided by Lu and Székely in [98] (see Theorem 2.17 in Chapter 2). For such a purpose, we must set a ε–near–positive dependency graph H (see Definition 2.9 in Chapter 2).

Lu and Székely [98] showed that an ε–near–positive dependency can be constructed using a family of matchings $\mathcal{M}$. Unfortunately, the conditions of [98, Theorem 4] which would provide an upper bound for our case, do not apply to our family $\mathcal{M}$ of matchings. We give instead a direct proof for the upper bound which is inspired by their approach.

Lemma 3.10. *With the hypothesis of Lemma 3.9, the rainbow dependency graph H is an ε–near–positive dependency graph with $\varepsilon = 1 - e^{-(3/k+60/k^2+o(1))}$.*

Proof. Set $\mathcal{A}_{\mathcal{M}} = \{A_1, \dots, A_m\}$, where $A_i = A_{e,f}$ for some $\{e,f\} \in \mathcal{M}$. The rainbow dependency graph H for $\mathcal{A}_{\mathcal{M}}$, clearly satisfies condition $i)$ in the definition of ε–near dependency graph, since two adjacent events contain incident edges and a matching is composed by a set of non–incident edges. For condition $ii)$ we want to show that, for each i and each $\mathcal{I} \subseteq \{j \mid ij \notin E(H),\, j \neq i\}$, we have the inequality

$$\Pr(A_i|B) \geq (1-\varepsilon)\Pr(A_i)\ ,$$

where $B = \cap_{j\in\mathcal{I}} \overline{A_j}$. This is equivalent to show that

$$\Pr(B|A_i) \geq (1-\varepsilon)\Pr(B)\ .$$

Let $\{a_1, \dots, a_n\}$ and $\{b_1, \dots, b_n\}$ be the vertices of the two sides of the bipartite graph $K_{n,n}$. By symmetry, we may assume that A_i consists of the event related to the 2–matching $\{a_{n-1}b_{n-1}, a_nb_n\}$. Then $\{A_j : j \in \mathcal{I}\}$ consists of some events related to the 2–matchings that coincide with $\{a_{n-1}b_{n-1}, a_nb_n\}$ in exactly one edge , indexed by $\mathcal{I}_1$; and some events related to the 2–matchings in $K_{n',n'} = K_{n,n} - \{a_{n-1}, a_n, b_{n-1}, b_n\}$, where $n' = n-2$, indexed by $\mathcal{I}_2$. The edge-coloring in $K_{n,n}$ induces an edge coloring in $K_{n',n'}$ where each color appears at most n'/k' times, where

$k' = k(1-2/n)$. Consider now the probability space of random matchings of $K_{n',n'}$ and let us define the following event $B' = \cap_{j\in\mathcal{I}}\overline{A'_j}$, where A'_j is the event that the edges in $K_{n',n'}$ corresponding to A_j are included in the random matching. Observe that if $j \in \mathcal{I}_1$ then A'_j corresponds to only one edge while if $j \in \mathcal{I}_2$ then it corresponds to two edges.

Thus,

$$\Pr(B|A_i) = \Pr(B') \ . \tag{3.8}$$

For convenience, we may split the event B corresponding to $\cap_{j\in\mathcal{I}}\overline{A_j}$ in several events depending on the perfect matching containing $\{a_{n-1}b_r, a_nb_s\}$. For any $r, s \in [n]$, $r \neq s$, let $C_{r,s}$ denote the event related to the 2–matching $\{a_{n-1}b_r, a_nb_s\}$. Define $\mathcal{I}_{r,s} \subset \mathcal{I}$ as the maximal subset of indices in $\mathcal{I}$, whose corresponding edges meet none of the two vertices b_r, b_s. Set $B_{r,s} = \cap_{j\in\mathcal{I}_{r,s}}\overline{A_j}$. Let us show that

$$\Pr(B) \leq \frac{1}{n(n-1)}\sum_{r\neq s}\Pr(B'_{r,s}) \ , \tag{3.9}$$

where, as before, $B'_{r,s} = \cap_{j\in\mathcal{I}_{r,s}}\overline{A'_j}$.

We note that, by the definition of $B_{r,s}$, we have $B \cap C_{r,s} = B_{r,s} \cap C_{r,s}$. Thus,

$$\Pr(B) = \sum_{r\neq s}\Pr(B\cap C_{r,s}) = \sum_{r\neq s}\Pr(B_{r,s}\cap C_{r,s}) \ .$$

We claim that for any $r, s \in [n]$, $r \neq s$,

$$\Pr(B_{r,s}|C_{r,s}) \leq \Pr(B_{r,s}|C_{n-1,n}) \ .$$

If $\mathcal{I}_{r,s} \subseteq \mathcal{I}_2$, by the definition of $\mathcal{I}_2$ none of the perfect matchings involved in $\cap_{j\in\mathcal{I}_2}\overline{A_j}$ meets vertices in $\{a_{n-1}a_n, b_{n-1}b_n\}$. In this case, for all r, s, $r \neq s$,

$$\Pr(B_{r,s}|C_{r,s}) = \Pr(B_{r,s}|C_{n-1,n}) \ .$$

Suppose then that $\mathcal{I}_{r,s}\cap\mathcal{I}_1 \neq \emptyset$. If some event indexed in $\mathcal{I}_1$ corresponds to a matching containing the edge a_nb_n (or $a_{n-1}b_{n-1}$) and $s \neq n$ (or $r \neq n-1$), we have

$$\Pr(B_{r,s}|C_{r,s}) = 0 \ ,$$

Otherwise, we have $\Pr(B_{r,s}|C_{r,s}) = \Pr(B_{r,s}|C_{n-1,n})$ as before.

Moreover, we observe that $\Pr(B_{r,s}|C_{n-1,n}) = \Pr(B'_{r,s})$. Therefore

$$\Pr(B) = \sum_{r\neq s}\Pr(B_{r,s}|C_{r,s})\Pr(C_{r,s}) \leq \frac{1}{n(n-1)}\sum_{r\neq s}\Pr(B_{r,s}|C_{n-1,n}) = \frac{1}{n(n-1)}\sum_{r\neq s}\Pr(B'_{r,s}) \ ,$$

giving inequality (3.9).

Let $\mathcal{M}_{\mathcal{I}}$ be the set of matchings corresponding to the events A'_j for $j \in \mathcal{I}$. By applying Lemma 3.9 with k', $\mathcal{M}_0 = \mathcal{M}_{\mathcal{I}}$, $S = \mathcal{I}\setminus\mathcal{I}_{r,s}$ and $T = \mathcal{I}_{r,s}$, we obtain

$$\Pr(B') = \Pr(B'_{r,s})\Pr(\cap_{j\in S}\overline{A'_j} \mid B'_{r,s}) \geq \Pr(B'_{r,s})e^{-(1+20/k+o(1))\sum_{i\in S}\Pr(A_i)} \ , \tag{3.10}$$

for any r, s, $r \neq s$, where x_j are given in the proof of the lemma. By combining (3.8) with (3.10) we get

$$n(n-1)\Pr(B|A_i) \geq \sum_{r\neq s} \Pr(B_{r,s}) e^{-(1+20/k+o(1))\sum_{i\in S}\Pr(A_i)} . \tag{3.11}$$

Now we give a uniform bound on $\sum_{i\in S}\Pr(A_i)$. Recall that $S = \mathcal{I} \setminus \mathcal{I}_{r,s}$ is the subset of indices in $\mathcal{I}$, whose corresponding edges are incident to either b_r or b_s. We consider the following two sets: $S_1 = \mathcal{I}_1 \setminus \mathcal{I}_{r,s}$ and $S_2 = \mathcal{I}_2 \setminus \mathcal{I}_{r,s}$. The size of both sets can be bounded independently of r and s by

$$|S_1| \leq \frac{n}{k}$$
$$|S_2| \leq 2n'\left(\frac{n'}{k'} - 1\right) \leq 2\frac{n(n-1)}{k} .$$

Then we have

$$\sum_{i\in S}\Pr(A_i) \leq \frac{|S_1|}{n} + \frac{|S_2|}{n(n-1)} \leq 3/k . \tag{3.12}$$

By using (3.11) with (3.12) and (3.9) we get

$$\Pr(B|A_i) \geq e^{-(3/k+60/k^2+o(1))}\frac{1}{n(n-1)}\sum_{r\neq s}\Pr(B_{r,s}) \geq e^{-(3/k+60/k^2+o(1))}\Pr(B) . \tag{3.13}$$

Therefore,

$$\varepsilon = 1 - e^{-(3/k+60/k^2+o(1))} ,$$

satisfies the conclusion of the lemma. □

Now we are able to prove Theorem 3.4.

Proof of Theorem 3.4. Set $\mathcal{A}_{\mathcal{M}} = \{A_1, \ldots, A_m\}$, where $A_i = A_{e,f}$ for some $\{e, f\} \in \mathcal{M}$.

By Lemma 3.10, the graph H from Definition 3.8 is an ε–near–positive dependency graph with $\varepsilon = 1 - e^{-(3/k+60/k^2+o(1))}$. It follows from Theorem 2.17 that the probability of having a rainbow perfect matching is upper bounded by

$$\Pr\left(\bigcap_{i=1}^{m}\overline{A_i}\right) \leq \prod_{i=1}^{m}(1-(1-\varepsilon)\Pr(A_i)) \leq e^{-(1-\epsilon)\mu} .$$

By plugging in our value of ε and by using $e^{-(3/k+60/k^2+o(1))} \geq 1 - \frac{3}{k} - \frac{60}{k^2} + o(1)$ we obtain

$$\Pr\left(\bigcap_{i=1}^{m}\overline{A_i}\right) \leq e^{-(1-3/k-60/k^2+o(1))\mu} .$$

Combining this upper bound with the lower bound obtained directly from (3.4) in Lemma 3.9 we obtain

$$\exp\left(-\left(1+\frac{16}{k}\right)\mu\right) \leq \Pr\left(\bigcap_{i=1}^{m}\overline{A_i}\right) \leq \exp\left(-\left(1-\frac{3}{k}-\frac{60}{k^2}+o(1)\right)\mu\right) .$$

This proves the theorem. □

Observe that when k is sufficiently large, the asymptotic estimation coincides with the one obtained by assuming that the bad events A_i are mutually independent.

3.3 Random colorings

In this section we will analyze the existence of rainbow perfect matchings in random edge–colorings of $K_{n,n}$.

Recall that, in the uniform random coloring model $\mathcal{C}_u(n,s)$, each edge of $K_{n,n}$ is given a color uniformly and independently chosen from a set of s colors, i.e. every possible coloring with at most s colors appears with the same probability.

In the regular random coloring model $\mathcal{C}_r(n,s)$ a coloring is chosen uniformly at random among all colorings of $E(K_{n,n})$ with equitable color classes of size n^2/s. Let us give a set up for this model. Consider two sets A and B each with n^2 points. Partition A in s cells $C_1,\dots,C_s$, each with n^2/s elements, representing the different colors. Let B represent the edges of $K_{n,n}$. A perfect matching between the points of A and B induces an equitable edge–coloring of the entire graph. The probability space $\mathcal{C}_r(n,s)$ of colorings is settled by choosing such a perfect matching uniformly at random. Let us show that $\mathcal{C}_r(n,s)$ is a uniform model for the set of equitable edge–colorings of $K_{n,n}$

Lemma 3.11. *Every equitable edge–coloring with s colors has the same probability in the $\mathcal{C}_r(n,s)$ model.*

Proof. We show that every equitable edge–coloring arises from the same number of perfect matchings from A to B. Let C be an equitable edge–coloring of $K_{n,n}$. Let $E_i \subset B$ be the set of edges that have color i under C. We have $|E_i| = n^2/s$ and there are $(n^2/s)!$ perfect matchings from C_i to E_i assigning color i to the edges in E_i. Therefore, there are exactly $((n^2/s)!)^s$ perfect matchings from A to B giving rise to the edge–coloring C. This number does not depend on C. □

We consider these two models since they simulate the worst situation among the colorings admitted in Theorem 3.4: the bounds on the probability that a perfect matching is rainbow only depends on the size of $\mathcal{M}$, and this set has its largest cardinality when there are few colors and the number of occurrences of each of them is maximized. This means that there are exactly $s = nk$ colors with n/k occurrences each. Observe that in both random models, the expected size of each color class is also n/k. In this sense, they are congruous to the hypothesis of Theorem 3.4.

Proof of Proposition 3.6. Consider an edge–coloring obtained using the $\mathcal{C}_u(n,s)$ model and let M denote a fixed perfect matching of $K_{n,n}$. If X_M is the random variable indicating that M is rainbow, then

$$\begin{aligned}\Pr(X_M = 1) &= \frac{s}{s}\cdot\frac{s-1}{s}\cdot\frac{s-2}{s}\cdot\ldots\cdot\frac{s-(n-1)}{s}\\ &= \prod_{i=0}^{n-1}\left(1-\frac{i}{s}\right)\,. \end{aligned} \tag{3.14}$$

For $s = n$ we can get directly from (3.14)

$$\Pr(X_M = 1) = \frac{n!}{n^n} = e^{-(2+o(1))\mu} \ .$$

Assume $s > n$. By writing $(1-x) = \exp(\log(1-x))$ for $0 < x < 1$, we have

$$\begin{aligned}\Pr(X_M = 1) &= \prod_{i=0}^{n-1} \exp\left(\log\left(1-\frac{i}{s}\right)\right) \\ &= \exp\left(\sum_{i=0}^{n-1} \log\left(1-\frac{i}{s}\right)\right) \\ &= \exp\left(\int_0^n \log\left(1-\frac{x}{s}\right)dx \ + e_1(n,s)\right) \ ,\end{aligned}$$

where $e_1(n,s)$ is the error term obtained from replacing the sum for the integral.

Since $\log\left(1-\frac{i}{s}\right)$ is decreasing, we have

$$\sum_{i=1}^{n} \log\left(1-\frac{i}{s}\right) \leq \int_0^n \log\left(1-\frac{x}{s}\right)dx \leq \sum_{i=0}^{n-1} \log\left(1-\frac{i}{s}\right) .$$

Thus, the error term can be bounded by

$$\begin{aligned}e_1(n,s) &\leq \left|\sum_{i=0}^{n-1} \log\left(1-\frac{i}{s}\right) - \int_0^n \log\left(1-\frac{x}{s}\right)dx\right| \\ &\leq \left|\sum_{i=0}^{n-1} \log\left(1-\frac{i}{s}\right) - \sum_{i=1}^{n} \log\left(1-\frac{i}{s}\right)\right| \\ &= \left|\log\left(1-\frac{n}{s}\right)\right| \\ &= \log\left(\frac{k}{k-1}\right) = O(1) \ ,\end{aligned} \tag{3.15}$$

where $k = s/n$.

Also,

$$\int_0^n \log\left(1-\frac{x}{s}\right) dx = -(s-n)\log\left(\frac{s-n}{s}\right) - n \ . \tag{3.16}$$

Using $\mu \sim \frac{n}{2k}$, we get

$$\begin{aligned}\Pr(X_M = 1) &= \exp\left(-\left((k-1)\log\left(\frac{k-1}{k}\right)+1\right)n + e_1(n,s)\right) \\ &= \exp\left(-2k\left((k-1)\log\left(\frac{k-1}{k}\right)+1+o(1)\right)\mu\right) \ ,\end{aligned}$$

proving the first part of the proposition for the $\mathcal{C}_u(n,s)$ model.

Now we study the probability that a fixed perfect matching M is rainbow in the $\mathcal{C}_r(n,s)$ model. According to the construction of the $\mathcal{C}_r(n,s)$ model, the probability of M being rainbow is

$$\begin{aligned}
\Pr(X_M = 1) &= \frac{n^2}{n^2} \cdot \frac{n^2 - \frac{n^2}{s}}{n^2 - 1} \cdot \frac{n^2 - 2\frac{n^2}{s}}{n^2 - 2} \cdot \ldots \cdot \frac{n^2 - (n-1)\frac{n^2}{s}}{n^2 - (n-1)} \\
&= \prod_{i=0}^{n-1} \left(1 - \frac{i(n^2 - s)}{s(n^2 - i)}\right) \\
&= \exp\left(\sum_{i=0}^{n-1} \log\left(1 - \frac{i(n^2 - s)}{s(n^2 - i)}\right)\right) \\
&= \exp\left(\int_0^n \log\left(1 - \frac{x(n^2 - s)}{s(n^2 - x)}\right) dx + e_2(n, s)\right) ,
\end{aligned}$$

where $e_2(n, s)$ is the error term obtained from replacing the sum for the integral.

If $s = n$ we have

$$\int_0^n \log\left(1 - \frac{x(n-1)}{(n^2 - x)}\right) dx = -n(n-1)\log\left(\frac{n}{n-1}\right) ,$$

which, by using the Taylor expansion of the logarithm, gives

$$\Pr(X_M = 1) = e^{-(2+o(1))\mu} .$$

By analogous arguments to those in (3.15), we can bound the error $e_2(n, s) = O(1)$. In the case where $s > n$, and using $k = s/n$, we have

$$\begin{aligned}
\int_0^n \log\left(1 - \frac{x(n^2 - s)}{s(n^2 - x)}\right) dx &= -\left((k-1)\log\left(\frac{k-1}{k}\right) - (n-1)\log\left(\frac{n-1}{n}\right)\right) n \\
&= -\left((k-1)\log\left(\frac{k-1}{k}\right) + 1 + o(1)\right) n .
\end{aligned}$$

Hence

$$\Pr(X_M = 1) = \exp\left(-2k\left((k-1)\log\left(\frac{k-1}{k}\right) + 1 + o(1)\right)\mu\right) .$$

□

Note that, for both models of random edge–colorings, the probability that a fixed perfect matching is rainbow is asymptotically the same. Observe that for the two random models we obtain the exact asymptotic value of the probability, while bounds provided by Theorem 3.4 (when the size $|\mathcal{M}|$ of the set of bad events is maximum) are probably not sharp, although consistent with the values for the random models.

We finally observe that for both models, when $k = 1$ we have $\Pr(X_M = 1) = e^{-(2+o(1))\mu}$, while if $k \to +\infty$, then $\Pr(X_M = 1) \to e^{-\mu}$ since

$$2k\left(1 - (k-1)\log\left(\frac{k}{k-1}\right)\right) = 1 + O\left(\frac{1}{k}\right) .$$

This reflects the idea that, when k is large, the number of bad events decreases and the model behaves as though they were independent.

3.4 Existence of rainbow perfect matchings

The aim of this section is to prove that with high probability there exists a rainbow perfect matching for any edge coloring of $E(K_{n,n})$ with $s \geq n$ colors. We only consider the $\mathcal{C}_u(n,s)$ model, but the results can be adapted to the $\mathcal{C}_r(n,s)$ model as well. The number of rainbow perfect matchings is counted by $X = \sum_M X_M$, which, according to Proposition 3.6, has expected value

$$\mathbb{E}(X) = \mathbb{E}(X) = \Pr(X_M = 1)n! = \exp\left(-2k\left((k-1)\log\left(\frac{k-1}{k}\right) + 1 + o(1)\right)\mu\right) n!\ .$$

Recall that μ is the expected size of $\mathcal{M}$, the set of pairs of non incident monochromatic edges. In order to have a rainbow perfect matching we just need to show that $X \neq 0$.

Proof of Theorem 3.7. To show that there exists some rainbow perfect matching *w.h.p.* we will use the second moment method, in particular Lemma 2.5. Observe that in our case $X = 0$ is equivalent to the non–existence of rainbow perfect matchings. Therefore, we need to compute $\sigma^2(X)$ and show that it is asymptotically smaller than $\mathbb{E}(X)^2$. Note that

$$\mathbb{E}(X^2) = \sum_{(M,N)} \mathbb{E}(X_M X_N)\ .$$

Let M and N denote two perfect matchings of $K_{n,n}$ with $|M \cap N| = z$. Then

$$\mathbb{E}(X_M X_N) = \Pr(X_M = 1)\Pr(X_N = 1 \mid X_M = 1)\ .$$

If $X_M = 1$, we know that the edges of $M \cap N$ are rainbow. In the remaining $n - z$ edges to color, we must avoid the z colors that appear in $M \cap N$. Thus,

$$\Pr(X_N = 1 \mid X_M = 1) = \prod_{i=z}^{n-1}\left(1 - \frac{i}{s}\right) \sim \exp\left(\frac{\alpha(z)z^2}{2s}\right)\Pr(X_M = 1)\ . \tag{3.17}$$

where $1 \leq \alpha(z) \leq 2$, as can be derived from (3.16). Observe that the events $X_M = 1$ and $X_N = 1$ are positively correlated.

For any perfect matching M and any integer z, such that $0 \leq z \leq n$, we claim that there exist at most $\binom{n}{z}\left(e^{-1}(n-z)! + 1\right)$ perfect matchings N such that $|M \cap N| = z$. We can assume that M is given by the identity and N by a permutation $\pi \in \mathcal{S}_n$. There are $\binom{n}{z}$ ways of choosing which edges of M will be shared by N, i.e. the set $\mathcal{I} = \{i : \pi(i) = i\}$. In order that π corresponds to a matching N with exactly z common edges with M, its restriction to $[n] \setminus \mathcal{I}$ must be a derangement. It is well known that the proportion of derangements among all the permutations of length $n - z$ is

$$\sum_{i=0}^{n-z} \frac{(-1)^i}{i!} \leq e^{-1} + \frac{1}{(n-z)!}\ .$$

Therefore there are at most $e^{-1}(n-z)! + 1$ ways to complete the perfect matching concluding our claim. Hence,

$$\mathbb{E}(X^2) = n!\sum_{z=0}^{n}\binom{n}{z}\left(e^{-1}(n-z)! + 1\right)\Pr(X_M = 1)\Pr(X_N = 1 \mid X_M = 1).$$

Since $\sigma^2(X) = \mathbb{E}(X^2) - \mathbb{E}(X)^2$,

$$\begin{aligned}\frac{\sigma^2(X)}{\mathbb{E}(X)^2} &= \frac{n!\sum_{z=0}^{n}\binom{n}{z}\left(e^{-1}(n-z)!+1\right)\Pr(X_M=1)\Pr(X_N=1\mid X_M=1)}{(n!\Pr(X_M=1))^2} - 1\\ &= e^{-1}\sum_{z=0}^{n}\frac{1}{z!}\left(1+\frac{e}{(n-z)!}\right)\frac{\Pr(X_N=1\mid X_M=1)}{\Pr(X_M=1)} - 1\ .\end{aligned}$$

By the sake of simplicity, let us define

$$f(s) = \sum_{z=0}^{n}\frac{1}{z!}\left(1+\frac{e}{(n-z)!}\right)\frac{\Pr(X_N=1\mid X_M=1)}{\Pr(X_M=1)}\ .$$

Then, using (3.17)

$$\begin{aligned}f(s) &\leq \sum_{z=0}^{n}\frac{1}{z!}\left(1+\frac{e}{(n-z)!}\right)\exp\left(\frac{\alpha(z)z^2}{2s}\right)\\ &\leq \sum_{z=0}^{\infty}\frac{1}{z!}\exp\left(\frac{\alpha(z)z^2}{2s}\right) + \frac{e}{n!}\sum_{z=0}^{n}\binom{n}{z}\exp\left(\frac{\alpha(z)z^2}{2s}\right)\\ &= \sum_{z=0}^{\infty}\frac{1}{z!}\sum_{t=0}^{\infty}\frac{1}{t!}\left(\frac{\alpha(z)z^2}{2s}\right)^t + \frac{e}{n!}\sum_{z=0}^{n}\binom{n}{z}\sum_{t=0}^{\infty}\frac{1}{t!}\left(\frac{\alpha(z)z^2}{2s}\right)^t\\ &= \sum_{t=0}^{\infty}a_t s^{-t}\ ,\end{aligned}$$

where $a_t = \frac{1}{2^t t!}\left(\sum_{z=0}^{\infty}\frac{(\alpha(z)z^2)^t}{z!} + \frac{e}{n!}\sum_{z=0}^{n}\binom{n}{z}(\alpha(z)z^2)^t\right)$. Observe that $a_0 = e\left(1+\frac{2^n}{n!}\right)$. Since $s \leq n^2$,

$$f(s) \leq e + O\left(s^{-1}\right)\ .$$

Observe that $s \geq n$, otherwise, $\Pr(X_M = 1) = 0$ in the Equation (3.14). Hence,

$$\frac{\sigma^2(X)}{\mathbb{E}(X)^2} = e^{-1}f(s) - 1 = O\left(s^{-1}\right) \to 0\ .$$

This concludes the proof. □

Corollary 3.12. *For any $\varepsilon > 0$, an equitable edge–coloring of $K_{n,n}$ with s colors, $s \geq n$, contains more than $(1-\varepsilon)c(k)^n n!$ rainbow perfect matchings with probability at least $1 - O(\varepsilon^{-2}s^{-1})$.*

Proof. It follows from the proof of Theorem 3.7 that $\Pr(X > (1-\varepsilon)\mathbb{E}(X)) \leq \frac{\sigma^2(X)}{\varepsilon^2\mathbb{E}(X)^2} = O\left(\varepsilon^{-2}s^{-1}\right) \to 0$. □

3.5 Concluding remarks and open questions

1 Theorem 3.4 provides upper and lower bounds for the number of rainbow perfect matchings of a given edge–coloring of $K_{n,n}$ such that the number of occurrences of each color is at most n/k and

$k \geq 13.66$. It is probably not true that each such edge–coloring contains $e^{-(1+o(1))c(k)\mu}n!$ rainbow perfect matchings, where $c(k)$ is defined in Proposition 3.6. However, it would be interesting to see how tight are these upper and lower bounds $c_1(k)$ and $c_2(k)$ provided in Theorem 3.4 by showing extremal examples.

2 One interesting question is how far can k be pushed down in order to still have an exponentially fraction of rainbow perfect matchings.

Question 3.13. *Which is the minimum value $k_0 > 1$ such that any edge–coloring of $K_{n,n}$ where any color appears at most n/k_0 times, has at least $c^n n!$, for some $c > 0$.*

Determining the smallest value of k with this property may shed some additional light on the open conjectures on latin transversals. Corollary 3.12 shows that almost all equitable edge–colorings with n colors contain an exponential fraction of perfect matchings that are rainbow.

In the same spirit, Wanless [126, Section 3] defines the function $f(n)$ to be the minimum number of latin transversals among all the latin squares of order n (case $k = 1$). Notice that $f(2n) = 0$ and Ryser's conjecture states that $f(2n+1) > 0$ for any $n \geq 0$. As far as we know, this function has not been studied yet.

Recently, the constant $4e \approx 10.87$ has been improved to $256/27 \approx 9.48$ by Bissacot et al. [20]. Their proof uses a Cluster version of the local lemma. Thus, it is possible that the same techniques we used in this chapter could be applied to extend our result up to a better constant.

3 When $s = n$, the proof of Theorem 3.7 shows that the probability that a random coloring in $\mathcal{C}_u(n,s)$ has no rainbow perfect matchings is

$$\Pr(\mathcal{C}_u(n,s) \text{ has no rainbow perfect matching}) = O\left(n^{-1}\right) . \tag{3.18}$$

The proportion of Latin squares among the set of square matrices with n symbols is of the order of $e^{-(1+o(1))2n^2}$, so that this estimation falls short to prove an asymptotic version of the original conjecture of Ryser. We have provided a probabilistic approach to the problem by showing that every equi–n–square admits a latin transversal with high probability. Even if there are some almost sure results on Latin squares (see e.g.[101, 32]), and some results on generating random latin squares [80, 103], to our knowledge there are no random models for latin squares, which could set the way to such an asymptotic version of the conjectures of Ryser or Brualdi on the existence of latin transversals in latin squares.

4 The following example shows how to construct exponentially many latin squares (in general edge–colorings of $K_{2k,2k}$ with $2k$ colors) which have no rainbow perfect matchings. Let k be odd. Choose two arbitrary colorings α_1, α_2 of $K_{k,k}$ with colors $\{a_1, ...a_k\}$ and two arbitrary colorings β_1, β_2 of $K_{k,k}$ with colors $\{b_1, ..., b_k\}$.

Let $\{A_1 \cup A_2, B_1 \cup B_2\}$ be the stable sets of $K = K_{2k,2k}$ with $|A_i| = |B_i| = k$ and use α_i for the edges connecting A_i with B_i, $i = 1, 2$, and β_i for the edges connecting A_i with B_j, $i \neq j$. Suppose that the resulting edge–colored graph has a rainbow perfect matching M. Since M must use the k colors $a_1, \ldots, a_k$, we may assume that it uses at least $(k+1)/2$ of these colors from the subgraph $K[A_1, B_1]$ induced by $A_1 \cup B_1$. But then each of the subgraphs $K[A_1, B_2]$ and $K[A_2, B_1]$ can only use $(k-1)/2$ colors $b_1, \ldots, b_k$ and some color b_i can not be used in the perfect matching, contradicting that M is rainbow.

It is easy to see that there are about n^{n^2} many equitable edges colorings of $K_{n,n}$. By the construction displayed above, if $n \equiv 2 \mod 4$, we can get $(k^{k^2})^4 = 2^{-n^2} n^{n^2}$ equitable edge–colorings which do not contain rainbow perfect matchings. Thus, for any coloring in $\mathcal{C}_r(n,s)$,

$$\Pr(\mathcal{C}_r(n,s) \text{ has no rainbow matching}) \geq 2^{-n^2} \gg e^{-(1+o(1))2n^2} \approx \Pr(\mathcal{C}_r(n,s) \text{ is a proper coloring}) ,$$

and there is no chance to prove that a proper edge-coloring of $K_{n,n}$ chosen uniformly at random, admits a rainbow perfect matching with high probability.

Ryser conjecture *w.h.p.* by improving the upper bound in (3.18).

CHAPTER 4

Bounds for identifying codes in terms of degree parameters

4.1 Introduction

Given a graph G, an identifying code $\mathcal{C}$ is a dominating set such that for any two vertices, their neighbourhoods within $\mathcal{C}$ are nonempty and distinct. This property can be used to distinguish all vertices of the graph one from each other. Identifying codes have found applications to various fields since the introduction of this concept in [86]. We refer to [96] for an on-line bibliography. One of the interests of this notion lies in their applications to the location of threats in facilities [1] and error-detection in computer networks [86]. One can also mention applications to routing [94], to bio-informatics [78] and to measuring the first-order logical complexity of graphs [87]. Let us also mention that identifying codes are special cases of the more general notion of *test covers* of hypergraphs, see e.g. [29, 107]. Test covers are also the implicit object of Bondy's celebrated theorem on *induced subsets* [28].

In this chapter, we address the question of providing lower and upper bounds on the size of an identifying code, thus extending earlier works (see e.g. [106, 34, 73, 59, 61]). We focus on degree–related graph parameters such as the minimum and maximum degree, and we also study the case of regular graphs. An important part of the chapter is devoted to giving the best possible upper bound for the size of an identifying code depending on the order and the maximum degree of the graph, a question raised in [58]. We also give improved bounds for graphs of girth at least 5 in terms of their minimum degree and study identifying codes in random regular graphs.

Let us first set the basic terminology on graphs we will use for Chapters 4 and 5. Unless othewise stated, we will consider G to be a simple, undirected and finite graph. The *open neighborhood* of a vertex v in G is the set of vertices in $V(G)$ that are adjacent to it, and will be denoted by $N_G(v)$. The *closed neighborhood* of a vertex v in G is defined as $N_G[v] = N_G(v) \cup \{v\}$. If the graph G is clear from the context we will write $N(v)$ and $N[v]$ instead of $N_G(v)$ and $N_G[v]$. The degree of a vertex $u \in V(G)$, is defined as $d(v) = |N_G(v)|$. Similarly, for any set $S \subseteq V(G)$, we define, $N(S) = \bigcup_{v\in S} N(v)$ and $N[S] = \bigcup_{v\in S} N[v]$. If two distinct vertices u, v are such that

$N[u] = N[v]$, they are called *twins*. If $N(u) = N(v)$ but $u \not\sim v$, u and v are called *false twins*. The symmetric difference between two sets S and T is denoted by $S \oplus T$.

Given a graph G and a subset $\mathcal{C}$ of vertices of G, $\mathcal{C}$ is called a *dominating set* if each vertex of $V(G) \setminus \mathcal{C}$ has at least one neighbor in $\mathcal{C}$. The set $\mathcal{C}$ is called a *separating set* of G if for each pair u, v of vertices of G, $N[u] \cap \mathcal{C} \neq N[v] \cap \mathcal{C}$ (equivalently, $(N[u] \oplus N[v]) \cap \mathcal{C} \neq \emptyset$). If $x \in N[u]$, we say that x *dominates* u. If $x \in N[u] \oplus N[v]$, we say that x *separates* u, v.

Definition 4.1. *A subset $\mathcal{C}$ of vertices of a graph G which is both a dominating set and a separating set is called an identifying code of G.*

It must be stressed that not every graph admits an identifying code. For instance, observe that a graph containing twin vertices does not admit a separating set, and in particular, an identifying code. In fact, a graph admits an identifying code if and only if it is *twin–free*, i.e. it has no pair of twins (one can see that if G is twin–free, $V(G)$ is an identifying code of G). Note that if for three distinct vertices u, v, w of a twin–free graph G, $N[u] \oplus N[v] = \{w\}$, then w belongs to any identifying code of G. In this case we say that w is *uv–forced*, or simply *forced*. Observe that any isolated vertex must belong to any identifying code for the reason that it must be dominated. For example, an edgeless graph needs all the vertices in any identifying code. Hence, the bounds here showed hold only for graphs with few isolated vertices. In order to shorten the statements of our results, we assume that all considered graphs have no isolated vertices.

The minimum size of a dominating set of graph G, its *domination number*, is denoted by $\gamma(G)$. Similarly, the minimum size of an identifying code of G, $\gamma^{\text{ID}}(G)$, is the *identifying code number* of G. It is known that for any twin-free graph G on n vertices having at least one edge, we have:

$$\lceil \log_2(n+1) \rceil \leq \gamma^{\text{ID}}(G) \leq n - 1.$$

The lower bound was proved in [86] and the upper bound, in [16, 73]. Both bounds are tight and all graphs reaching these two bounds have been classified (see [106] for the lower bound and [59] for the upper bound). Other papers studying bounds and extremal graphs for identifying codes are e.g. [34, 61, 62].

When considering graphs of given maximum degree Δ, it was showed in [86] that the lower bound can be improved to $\gamma^{\text{ID}}(G) \geq \frac{2n}{\Delta+2}$. This bound is tight and a classification of all graphs reaching it has been proposed in [58]. For any Δ, these graphs include some regular graphs and graphs of arbitrarily large girth.

It was conjectured in [61] that the following upper bound holds.

Conjecture 4.2 ([61])**.** *There exists a constant c such that for any nontrivial connected twin–free graph G of maximum degree Δ,*

$$\gamma^{ID}(G) \leq n - \tfrac{n}{\Delta} + c\ .$$

Graphs of maximum degree Δ such that $\gamma^{\text{ID}}(G) = n - \frac{n}{\Delta}$ are known (e.g. the complete bipartite graph $K_{d,d}$, where $\Delta = d$, and richer classes of graphs described in Section 4.5). Therefore, if Conjecture 4.2 holds, there would exist a constant c such that, for any twin–free graph G on n vertices and of maximum degree Δ, we would have $\frac{2}{\Delta+2}n \leq \gamma^{\text{ID}}(G) \leq n - \frac{n}{\Delta} + c$, with both bounds being tight.

Note that Conjecture 4.2 holds for graphs of maximum degree 2 (see [74]). It was showed in [59] that $\gamma^{\mathrm{ID}}(G) \leq n - \frac{n}{O(\Delta^5)}$, and $\gamma^{\mathrm{ID}}(G) \leq n - \frac{n}{O(\Delta^3)}$ when G is regular. It is also known that the conjecture holds asymptotically if G is triangle–free: then, $\gamma^{\mathrm{ID}}(G) \leq n - \frac{n}{\Delta(1+o_\Delta(1))}$ [61].

Identifying codes have been previously studied in the following two models of random graphs: Erdős–Rényi random graphs [66] and random geometric graphs [110]. To our knowledge random regular graphs have not been studied in the context of identifying codes.

In this chapter, we further study Conjecture 4.2 and prove that it is tight (up to constants) for large enough values of Δ and for a large class of graphs, including regular graphs and graphs of bounded clique number (Corollaries 4.7 and 4.11). In the general case, we prove that $\gamma^{\mathrm{ID}}(G) \leq n - \frac{n}{O(\Delta^3)}$ (Corollary 4.9). These results improve the known bounds given in [59] and support Conjecture 4.2. Moreover, we show that the much improved upper bound $\gamma^{\mathrm{ID}}(G) \leq (1 + o_d(1))\frac{3\log d}{2d}n$ holds for graphs having girth at least 5 and minimum degree d (Theorem 4.19). This bound is used to give an asymptotically tight bound of about $\frac{\log d}{d}n$ for the identifying code number of almost all random regular graphs (Corollary 4.23).

We summarize our results for the special case of regular graphs in Table 4.1 and compare them to the bound for the dominating set problem (the table contains references for both the bound and its tightness). All bounds are asymptotically tight. We note that identifying codes behave far from dominating sets in general, as showed by the first lines of the table: there are regular graphs having much larger identifying code number than domination number. However, for larger girth and for almost all regular graphs, the bounds for the two problems coincide asymptotically, as showed by the last lines of the table.

	Identifying codes	Dominating sets
in general	$n - \frac{n}{103d}$ Thm. 4.6, Constr. 4.25	$(1 + o_d(1))\frac{\log d}{d}n$ [8], [124]
girth 4	$n - \frac{n}{d(1+o_d(1))}$ [61], Constr. 4.26	$(1 + o_d(1))\frac{\log d}{d}n$ [8], [124]
girth 5	$(1 + o_d(1))\frac{3\log d}{2d}n$ Thm. 4.19, Thm. 4.21	$(1 + o_d(1))\frac{\log d}{d}n$ [8], [124]
almost all graphs	$(1 + o_d(1))\frac{\log d}{d}n$ Thm. 4.20, Thm. 4.21	$(1 + o_d(1))\frac{\log d}{d}n$ [8], [124]

Table 4.1: Summary of the upper bounds for d–regular graphs

In order to prove our results, we use probabilistic techniques. For some results, we use the weighted version of Lovász Local Lemma (see Section 2.2.2.2) to show the existence of an identifying code, together with the Chernoff bound (see Section 2.2.1) to show that this code is small enough. To bound the number of forced vertices in a graph we study an auxiliary directed graph that captures the underlying structure of these vertices. This new technique we introduce can be useful to study the number of forced vertices in a more general context, which is an important problem in the community of identifying codes. We also make use of other probabilistic techniques such as the Alteration Method [8] in order to give better bounds in more restricted cases. Finally, we work with the Configuration Model (see Section 2.3.1) in order to compute the identifying code number of almost all random regular graphs.

The organization of this chapter is as follows. In Section 4.2, we improve the known upper bounds on the identifying code number of graphs of maximum degree Δ. This gives new large families of graphs for which Conjecture 4.2 holds (up to constants). In Section 4.3, we give an upper bound for graphs having minimum degree d and girth at least 5. In Section 4.4, we give sharp bounds for the identifying code number of almost all d–regular graphs. A further section is dedicated to various constructions of families of graphs which show the tightness of some of our results (Section 4.5). Concluding remarks and open questions are collected in Section 4.6.

4.2 Upper bounds on the identifying code number

In this section, we improve the known upper bounds of [59] on the identifying code number by using the Weighted Local Lemma, stated in Lemma 2.13.

4.2.1 Preliminary results

First of all, we give an equivalent condition for a set to be an identifying code. This follows from the fact that for two vertices u, v at distance at least 3 from each other, $N[u] \oplus N[v] = N[u] \cup N[v]$.

Observation 4.3. *For a graph G and a set $\mathcal{C} \subseteq V(G)$, if $\mathcal{C}$ is dominating and $N[u] \cap \mathcal{C} \neq N[v] \cap \mathcal{C}$ for each pair of vertices u, v at distance at most two from each other, then $N[u] \cap \mathcal{C} \neq N[v] \cap \mathcal{C}$ for each pair of vertices of the graph.*

The next observation is immediate, but it is worth mentioning here.

Observation 4.4. *Let G be a twin–free graph and $\mathcal{C}$, an identifying code of G. Any set $\mathcal{C}'$ such that $\mathcal{C} \subseteq \mathcal{C}'$ is also an identifying code of G.*

The next proposition shows an upper bound on the number of false twins in a graph.

Proposition 4.5. *Let G be a graph on n vertices having maximum degree Δ and no isolated vertices, then G has at most $\frac{n(\Delta-1)}{2}$ pairs of false twins.*

Proof. Let us build a graph H on $V(G)$, where two vertices u, v are adjacent in H if they are false twins in G. Note that since a vertex can have at most $\Delta - 1$ false twins, H has maximum degree $\Delta - 1$. Therefore it has at most $\frac{n(\Delta-1)}{2}$ edges and the claim follows. □

Note that the bound of Proposition 4.5 is tight since in a complete bipartite graph $K_{d,d}$, $n = 2d = 2\Delta$ and there are exactly $2\binom{d}{2} = 2\binom{\Delta}{2} = \frac{n(\Delta-1)}{2}$ pairs of false twins.

4.2.2 Main theorem

In the following, given a graph G on n vertices, we will denote by $f(G)$ the proportion of non–forced vertices of G, i.e. the ratio $\frac{x}{n}$, where x is the number of non–forced vertices of G.

Theorem 4.6. *Let G be a twin–free graph on n vertices having maximum degree $\Delta \geq 3$. Then,*

$$\gamma^{ID}(G) \leq n - \frac{nf(G)^2}{103\Delta}.$$

Proof. Let F be the set of forced vertices of G, and $V' = V(G) \setminus F$. Note that $|V'| = nf(G)$. By the definition of a forced vertex, any identifying code must contain all vertices of F.

In this proof, we first build a set S in a random manner by choosing vertices from V'. Then we exhibit some "bad" configurations — if none of those occurs, the set $\mathcal{C} = F \cup (V' \setminus S)$ is an identifying code of G. Using the Weighted Local Lemma, we compute a lower bound on the (nonzero) probability that none of these bad events occurs. Finally, we use the Chernoff bound to show that with nonzero probability, the size of S is also large enough for our purposes. This shows that such a "good" large set S exists, and it can be used to build an identifying code that has a sufficiently small size.

Let $p = p(\Delta)$ be a probability which will be determined later. We build the set $S \subseteq V'$ such that each vertex of V' independently belongs to S with probability p. Therefore the random variable $|S|$ follows a binomial distribution $\mathrm{Bin}(nf(G),p)$ and has expected value $\mathbb{E}(|S|) = pnf(G)$.

Let us now define the set $\mathcal{A}$ of "bad" events of size N. These are of four types. An illustration of these events is given in Figure 4.2.

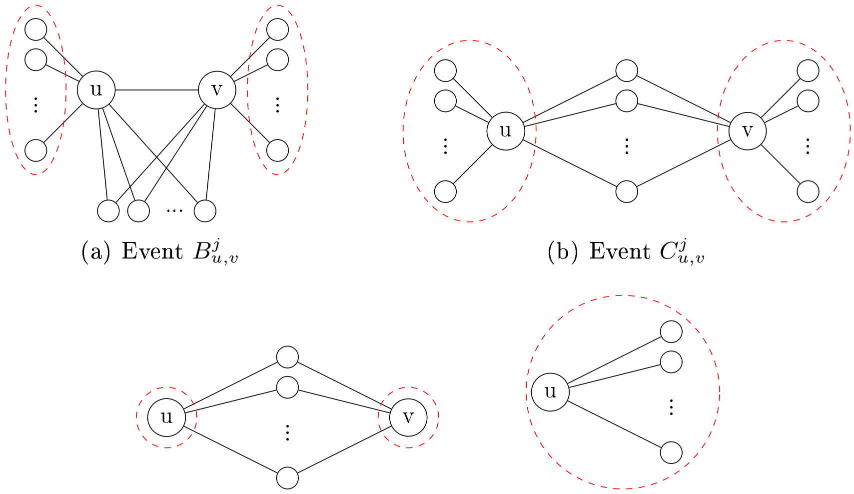
- **Type B^j ($2 \leq j \leq 2\Delta - 2$):** for each pair $\{u, v\}$ of adjacent vertices, let $B^j_{u,v}$ be the event that $|(N[u] \oplus N[v])| = j$ and $(N[u] \oplus N[v]) \subseteq S$.
- **Type C^j ($3 \leq j \leq 2\Delta$):** for each pair $\{u, v\}$ of vertices in V' at distance two from each other, let $C^j_{u,v}$ be the event that $|(N[u] \oplus N[v])| = j$ and $(N[u] \oplus N[v]) \subseteq S$.
- **Type D:** for each pair $\{u, v\}$ of false twins in V', let $D_{u,v}$ be the event that $(N[u] \oplus N[v]) = \{u, v\} \subseteq S$.
- **Type E^j ($2 \leq j \leq \Delta + 1$):** for each vertex $u \in V'$, let E^j_u be the event that $|N[u]| = j$ and $N[u] \subseteq S$.

For the sake of simplicity, we refer to the events of type B^j, C^j and E^j as events of type B, C and E respectively whenever the size of the symmetric difference is not relevant.

Events of type $B^1_{u,v}$ are not defined since then $|N[u] \oplus N[v]| = 1$ and F belongs to the code, so they never happen. Observe that the events $C^j_{u,v}$ and $D_{u,v}$ are just defined over the pairs of vertices in V' because if either u or v belongs to F, the event does not happen.

If no event of type B occurs, all pairs of adjacent vertices are separated by $V(G) \setminus S$. If no event of type C or D occurs, all pairs of vertices at distance 2 from each other are separated. If no event of type E occurs, $V(G) \setminus S$ is a dominating set of G. Thus by Observation 4.3, if no event of type B, C, D or E occurs, then $V(G) \setminus S$ is an identifying code of G.

Let $V(A_i)$ denote the set of vertices that must belong to set S so that A_i holds (see Figure 4.2, where the sets $V(A_i)$ are the ones inside the dashed circles). We will say that a vertex $v \in V(G)$ participates to A_i, if $v \in V(A_i)$. We define the weight t_i of each event $A_i \in \mathcal{A}$ as $|V(A_i)|$. For



(a) Event $B^j_{u,v}$ (b) Event $C^j_{u,v}$

(c) Event $D_{u,v}$ (d) Event E^j_u

Figure 4.2: The "bad" events. The vertices in dashed circles belong to set S.

$j \geq 2$ and for $T \in \{B^j, C^j, D, E^j, \}$, let t_T be the weight of an event of type T (for an event $A_i \in \mathcal{A}$ of type T, $t_i = t_T$). We have the following,

$$t_{B^j} = j \;, \qquad t_{C^j} = j \;, \qquad t_D = 2 \;, \qquad \text{and} \qquad t_{E^j} = j \;.$$

Some vertex x can participate to at most $\Delta(\Delta-1)$ events of type B: supposing $x \in V(B^j_{u,v})$ and u is adjacent to x, there are at most Δ ways to choose u, and at most $\Delta - 1$ ways to choose v among $N(u) \setminus \{x\}$. Observe that if $x = u$ or $x = v$, then $x \notin V(B_{u,v})$ (see Figure 4.2(a)). Similarly x can participate to at most $\Delta^2(\Delta-1)$ events of type C: for some event $C^j_{u,v}$, there are at most $\Delta(\Delta-1)$ possibilities if $x = u$ or $x = v$ and at most $\Delta(\Delta-1)^2$ if u or v is a neighbour of x. Vertex x can participate to at most $\Delta - 1$ events $D_{u,v}$ since x can have at most $\Delta - 1$ false twins. Finally, a vertex x can participate to at most $\Delta + 1$ events of type E since if it participates to some event E^j_u, then $u \in N[x]$.

For each type T of events ($T \in \{B^j, C^j, D, E^j\}$) and any vertex $v \in V(G)$, let us define $\mathrm{g}(v, T)$ to be the number of events A_i of type T such that $v \in V(A_i)$. Hence,

$$\sum_{j=2}^{2\Delta-2} \mathrm{g}(v, B^j) \leq \Delta(\Delta-1) \;, \qquad \sum_{j=3}^{2\Delta} \mathrm{g}(v, C^j) \leq \Delta^2(\Delta-1) \;,$$
$$\mathrm{g}(v, D) \leq \Delta - 1 \;, \qquad \text{and} \qquad \sum_{j=2}^{\Delta+1} \mathrm{g}(v, E^j) \leq \Delta + 1 \;. \tag{4.1}$$

Let us call A_{IC} the event that no event of $\mathcal{A}$ occurs. Using the Weighted Local Lemma, we want to show that $\Pr(A_{\mathrm{IC}}) > 0$. First of all we need to set a dependency graph H for the set of events $\mathcal{A}$. Recall that $V(H) = [N]$. For any $i, j \in [N]$, we will have $ij \in E(H)$ if and only $V(A_i) \cap V(A_j) = \emptyset$. Observe that for any event A_i and any set $T \subseteq \{j : i \not\sim j\}$, we have $\Pr(A_i \mid \cap_{j \in T} \overline{A_j}) = \Pr(A_i)$, since the vertices are included in S with independent probabilities. This means that A_i is mutually independent from the set of all events A_j for which $V(A_i) \cap V(A_j) = \emptyset$. Thus, our graph H matches Definition 2.6 and is a dependency graph for $\mathcal{A}$.

In order to apply the Weighted Local Lemma (Lemma 2.13), the following conditions must hold for each event $A_i \in \mathcal{A}$,

$$\sum_{ij \in E(H)} (2p)^{t_j} \leq \frac{t_i}{2} \,.$$

The latter conditions are implied by the following ones (for each event $A_i \in \mathcal{A}$),

$$\sum_{j=2}^{2\Delta-2} \sum_{v \in V(A_i)} \mathrm{g}(v, B^j)(2p)^{t_{B^j}} + \sum_{j=3}^{2\Delta} \sum_{v \in V(A_i)} \mathrm{g}(v, C^j)(2p)^{t_{C^j}}$$
$$+ \sum_{v \in V(A_i)} \mathrm{g}(v, D)(2p)^{t_D} + \sum_{j=2}^{\Delta+1} \sum_{v \in V(A_i)} \mathrm{g}(v, E^j)(2p)^{t_{E^j}} \leq \frac{t_i}{2} \,.$$

Which are implied by

$$t_i \cdot \max_{v \in V(A_i)} \left\{ \sum_{j=2}^{2\Delta-2} \mathrm{g}(v, B^j)(2p)^{t_{B^j}} \right\} + t_i \cdot \max_{v \in V(A_i)} \left\{ \sum_{j=3}^{2\Delta} \mathrm{g}(v, C^j)(2p)^{t_{C^j}} \right\}$$
$$+ t_i \cdot \max_{v \in V(A_i)} \left\{ \mathrm{g}(v, D)(2p)^{t_D} \right\} + t_i \cdot \max_{v \in V(A_i)} \left\{ \sum_{j=2}^{\Delta+1} \mathrm{g}(v, E^j)(2p)^{t_{E^j}} \right\} \leq \frac{t_i}{2} \,.$$

Using the bounds of Inequalities (4.1) and noting that for $p \leq 1/4$ and any j, $(2p)^{t_{B^j}} \leq (2p)^2$, $(2p)^{t_{C^j}} \leq (2p)^3$ and $(2p)^{t_{E^j}} \leq (2p)^2$, for any event A_i this equation is implied by

$$(\Delta+1)(2p)^2 + \Delta(\Delta-1)(2p)^2 + \Delta^2(\Delta-1)(2p)^3 + (\Delta-1)(2p)^2 = 4\Delta^2 p^2 + 8\Delta^3 p^3 + 4\Delta p^2 - 8\Delta^2 p^3 \leq \frac{1}{2} \,. \tag{4.2}$$

Hence, we fix $p = \frac{1}{k\Delta}$ where k is a constant to be determined later. Equation (4.2) holds for $k \geq 3.68$ for all $\Delta \geq 3$. In fact, in the following steps of the proof, we will assume that $k \geq 30$, and so Equation (4.2) will be satisfied for any $\Delta \geq 3$. Since $p \leq \frac{1}{4}$ and $\Pr(A_i) \leq p^{t_i}$ by the definition of t_i and the choice of S, the Weighted Local Lemma can be applied.

Let N_T be the number of events of type T, where $T \in \left\{B^j, C^j, D, E^j, \right\}$. By Lemma 2.13 we have

$$\Pr(A_{\mathrm{IC}}) \geq \prod_{j=2}^{2\Delta-2} \prod_{i=1}^{N_{B^j}} (1 - (2p)^{t_{B^j}}) \prod_{j=3}^{2\Delta} \prod_{i=1}^{N_{C^j}} (1 - (2p)^{t_{C^j}}) \prod_{i=1}^{N_D} (1 - (2p)^{t_D}) \prod_{j=2}^{\Delta+1} \prod_{i=1}^{N_{E^j}} (1 - (2p)^{t_{E^j}}) \,.$$

Note that $\sum_{j=2}^{2\Delta-2} N_{B^j} \leq \frac{n\Delta}{2}$ since there is exactly one event type $B^j_{u,v}$ for each edge $uv \in E(G)$ and at most $\frac{n\Delta}{2}$ edges in G. We also have that $\sum_{j=3}^{2\Delta} N_{C^j}$ is at most the number of pairs of vertices in V' at distance 2 from each other. This is also at most the number of paths of length 2 with both endpoints in V', which is upper-bounded by $\frac{nf(G)\Delta(\Delta-1)}{2}$. Moreover, N_D is the number of pairs of false twins in V', which is at most $nf(G)\frac{\Delta-1}{2}$ by Proposition 4.5. Finally, $\sum_{j=2}^{\Delta+1} N_{E^j} = nf(G)$ since by definition there exists exactly one event E^j_u for each vertex of $u \in V'$.

Hence, we have

$$\Pr(A_{\mathrm{IC}}) \geq (1 - (2p)^2)^{\frac{n\Delta}{2}} (1 - (2p)^3)^{\frac{nf(G)\Delta(\Delta-1)}{2}} (1 - (2p)^2)^{\frac{nf(G)(\Delta-1)}{2}} (1 - (2p)^2)^{nf(G)} \,.$$

Note that in Lemma 2.13, since $p \leq \frac{1}{4}$ and $(1-x) \geq e^{-(2\log 2)x}$ in $x \in [0, 1/2]$, we have

$$\Pr\left(\bigcap_{i=1}^{N} \overline{A_i}\right) \geq \exp\left(-(2\log 2)\sum_{i=1}^{N}(2p)^{t_i}\right) . \tag{4.3}$$

Since $p = \frac{1}{k\Delta}$, we obtain

$$\begin{aligned}\Pr(A_{\mathrm{IC}}) &\geq \exp\left(-(2\log 2)(2p)^2\left(\frac{\Delta}{2} + \frac{f(G)\Delta(\Delta-1)2p}{2} + \frac{f(G)(\Delta-1)}{2} + f(G)\right)n\right)\\ &\geq \exp\left(-\frac{4\log 2}{k^2\Delta}\left(1 + \frac{2f(G)}{k} + f(G) + \frac{2f(G)}{\Delta}\right)n\right) .\end{aligned}$$

Since $f(G) \leq 1$ and it is assumed that $k \geq 30$, one can check that for any $\Delta \geq 3$,

$$\Pr(A_{\mathrm{IC}}) \geq \exp\left(-\frac{164\log 2}{15k^2\Delta}n\right) .$$

Note that this bound could be strengthened by assuming Δ to be large enough. Indeed, here the term $\frac{2f(G)}{\Delta}$ can be as high as $\frac{2}{3}$ when $\Delta = 3$ and $f(G) = 1$, but can be chosen to be as low as desired by assuming Δ to be larger. However we aim to give a bound for any $\Delta \geq 3$, hence we use the weaker bound presented here.

The Weighted Local Lemma shows that the set S has the desired properties with probability $\Pr(A_{\mathrm{IC}}) > 0$, implying that such a set exists. Note that we have no guarantee on the size of S. In fact, if $S = \emptyset$ then $V(G) \setminus S = V(G)$ is always an identifying code. Therefore we need to estimate the probability that $|S|$ is far below its expected size. In order to do this, we use the Chernoff bound of Theorem 2.3 by putting $a = \frac{nf(G)}{c\Delta}$ where c is a constant to be determined. Let A_{BIG} be the event that $|S| - np > -\frac{nf(G)}{c\Delta}$. We obtain

$$\begin{aligned}\Pr(\overline{A_{\mathrm{BIG}}}) &\leq \exp\left(-\frac{\left(\frac{nf(G)}{c\Delta}\right)^2}{2pnf(G)}\right)\\ &= \exp\left(-\frac{kf(G)}{2c^2\Delta}n\right) .\end{aligned}$$

Now we have

$$\begin{aligned}\Pr(A_{\mathrm{IC}} \text{ and } A_{\mathrm{BIG}}) &= 1 - \Pr(\overline{A_{\mathrm{IC}}} \text{ or } \overline{A_{\mathrm{BIG}}})\\ &\geq 1 - \Pr(\overline{A_{\mathrm{IC}}}) - \Pr(\overline{A_{\mathrm{BIG}}})\\ &= 1 - (1 - \Pr(A_{\mathrm{IC}})) - \Pr(\overline{A_{\mathrm{BIG}}})\\ &= \Pr(A_{\mathrm{IC}}) - \Pr(\overline{A_{\mathrm{BIG}}})\\ &\geq \exp\left(-\frac{164\log 2}{15k^2\Delta}n\right) - \exp\left(-\frac{kf(G)}{2c^2\Delta}n\right) .\end{aligned}$$

Thus, $\Pr(A_{\mathrm{IC}} \text{ and } A_{\mathrm{BIG}}) > 0$ if $c < \sqrt{\frac{15}{328\log 2}} \cdot k^{3/2}f(G)^{1/2}$. We (arbitrarily) set $c = \frac{k^{3/2}f(G)^{1/2}}{\sqrt{22\log 2}}$ in order to fulfill this condition.

Now we have to check that A_{BIG} implies that S is still large enough.

$$\begin{aligned}|S| &\geq \mathbb{E}(|S|) - \frac{nf(G)}{c\Delta} \\ &= \frac{nf(G)}{k\Delta} - \frac{nf(G)}{c\Delta} \\ &= \left(\frac{1}{k} - \frac{\sqrt{22\log 2}}{k^{3/2}f(G)^{1/2}}\right)\frac{nf(G)}{\Delta}\ .\end{aligned} \tag{4.4}$$

Since $|S|$ must be positive, from Equation (4.4) we need $k^{3/2}f(G)^{1/2} > \sqrt{22\log 2}\,k$, which leads to $k = \frac{a_0}{f(G)}$ for $a_0 > 22\log 2$. Using all our previous assumptions, by derivating the expression of $|S|$, one can check that $|S|$ is maximized when $a_0 = \frac{99\log 2}{2}$. Hence we set $k = \frac{99\log 2}{2f(G)}$.

Remark that under this condition and since $f(G) \leq 1$, we have $k \geq 34$ and our assumption following Equation (4.2) that $k \geq 30$, is fulfilled.

Now, with $a_0 = \frac{99\log 2}{2}$, we can see that

$$|S| \geq \left(\frac{1}{k} - \frac{1}{c}\right)\frac{nf(G)}{\Delta} = \frac{a_0^{1/2} - \sqrt{22\log 2}}{a_0^{3/2}}\frac{f(G)^2}{\Delta}n = \frac{2}{297\log 2}\frac{f(G)^2}{\Delta}n \geq \frac{f(G)^2}{103\Delta}n\ .$$

Hence finally the identifying code $\mathcal{C} = V \setminus S$ has size

$$|\mathcal{C}| \leq n - \frac{nf(G)^2}{103\Delta}\ .$$

□

Note that for regular graphs, $f(G) = 1$, since the existence of a forced vertex implies the existence of two vertices with distinct degrees. We obtain the following result.

Corollary 4.7 (Graphs with constant proportion of non–forced vertices)**.** *Let G be a twin–free graph on n vertices with maximum degree $\Delta \geq 3$ and $f(G) = \frac{1}{\alpha}$ for some constant $\alpha \geq 1$. Then,*

$$\gamma^{ID}(G) \leq n - \frac{n}{103\alpha^2\Delta}\ .$$

In particular if G is d–regular,

$$\gamma^{ID}(G) \leq n - \frac{n}{103d}\ .$$

The next proposition will be proved in the next subsection.

Proposition 4.8. *Let G be a graph on n vertices and of maximum degree Δ. Then,*

$$f(G) \geq \tfrac{1}{\Delta+1}\ .$$

Using it, we obtain the following general result.

Corollary 4.9 (General case)**.** *Let G be a twin–free graph on n vertices having maximum degree $\Delta \geq 3$. Then,*

$$\gamma^{ID}(G) \leq n - \frac{n}{103\Delta(\Delta+1)^2} = n - \frac{n}{O(\Delta^3)}\ .$$

The next proposition will be proved in the next subsection as well.

Proposition 4.10. *Let G be a graph having no k-clique. Then, there exists a constant $\gamma(k)$ depending only on k, such that*

$$f(G) \geq \tfrac{1}{\gamma(k)} .$$

This leads to the following extension of Corollary 4.7, where $c(k) \leq 103\gamma(k)^2$.

Corollary 4.11 (Graphs with bounded clique number)**.** *There exists an integer Δ_0 such that for each twin–free graph G on n vertices having maximum degree $\Delta \geq \Delta_0$ and clique number smaller than k, we have*

$$\gamma^{ID}(G) \leq n - \frac{n}{c(k)\Delta} ,$$

for some constant $c(k)$ depending only on k. In particular this applies to triangle–free graphs, planar graphs, or more generally, graphs of bounded genus.

We remark here that the previous corollaries support Conjecture 4.2. They also lead us to think that the difficulty of the problem lies in studying the set of forced vertices.

4.2.3 Bounding the number of non–forced vertices

Here, we prove the lower bounds for function $f(G)$ of the statement of Theorem 4.6.

The following lemma was first proved in [16], and a proof can be found in [59] (as [16] is not accessible).

Lemma 4.12 ([16])**.** *If G is a finite twin–free graph without isolated vertices, then for every vertex u of G, there is a vertex $v \in N[u]$ such that $V(G) \setminus \{v\}$ is an identifying code of G.*

Let us now prove Proposition 4.8.

Proof of Proposition 4.8. Observe that a vertex v of G is not forced only if $V(G) \setminus \{v\}$ is an identifying code of G. Hence, by Lemma 4.12, the set S of non–forced vertices is a dominating set of G, and thus $|S| \geq \frac{n}{\Delta+1}$. □

Note that Proposition 4.8 is tight. Indeed, consider the graph A_k on $2k$ vertices defined in [59] as follows: $V(A_k) = \{x_1, \ldots, x_{2k}\}$ and $E(A_k) = \{x_i x_j, |i-j| \leq k-1\}$. A_k can be seen as the $(k-1)$–th power of the path P_{2k}. Construct now the graph B_k by adding a universal vertex x (i.e. x is adjacent to all vertices of A_k) in the graph A_k. One can check that all vertices from B_k but x are forced. This graph has $n = 2k+1$ vertices, maximum degree $2k$ and exactly $1 = \frac{n}{\Delta+1}$ non–forced vertex. Taking all forced vertices gives a minimum identifying code of this graph.

However, note that since for a fixed even value of Δ, we know only one such graph, it is not enough to give a counterexample to Conjecture 4.2. Indeed in this case the size of the code is $n-1 = n - \frac{n}{\Delta+1} = n - \frac{n}{\Delta} + \frac{1}{n-1} = n - \frac{n}{\Delta} + 1$.

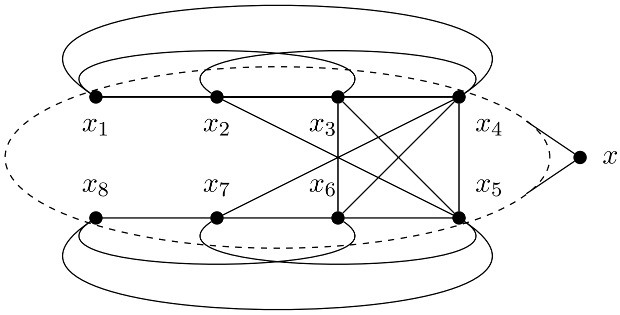


Figure 4.3: The graph B_4.

Observe that the graph B_k contains two cliques of k vertices. In fact, we can improve the bound of Proposition 4.8 for graphs having no large cliques. Let us first introduce an auxiliary structure that will be needed in order to prove this result.

Let G be a twin–free graph. We define a partial order $\preceq$ over the set of vertices of G such that $u \preceq v$ if $N[u] \subseteq N[v]$. We construct an oriented graph $\mathcal{H}(G)$ on $V(G)$ as a subgraph of the Hasse diagram of poset $(V(G), \preceq)$. The arc set of $\mathcal{H}(G)$ is the set of all arcs $\overrightarrow{uv}$ where there exists some vertex x such that $N[v] = N[u] \cup \{x\}$. Then x is uv–forced, and we note $x = f(\overrightarrow{uv})$. For a vertex v of $V(G)$, we define the set $F(v)$ as the union of v itself and the set of all predecessors and successors of v in $\mathcal{H}(G)$. Observe that $\mathcal{H}(G)$ has no directed cycle since it represents a partial order, and thus predecessors and successors are well–defined.

Lemma 4.13. *Let G be a graph having no k-clique. Then for each vertex u, $|F(u)| \leq \beta(k)$, where $\beta(k)$ is a function depending only on k.*

Proof. First of all, we prove that the maximum in-degree of $\mathcal{H}(G)$ is at most $2k-3$, and its out-degree is at most $k-2$.

Let u be a vertex of G. Suppose u has $2k-2$ in-neighbours in $\mathcal{H}(G)$. Since for each in-neighbour v of u, $|N[u] \oplus N[v]| = 1$ in G, each of them is nonadjacent in G to at most one of the other in-neighbours (in the worst case the in-neighbours of u induce in G a clique of $2k-2$ vertices minus the edges of a perfect matching). Hence they induce a clique of size at least $k-1$ in G. Together with vertex u, they form a k-clique in G, a contradiction.

Now suppose u has $k-1$ out-neighbours in $\mathcal{H}(G)$. Since for each out-neighbour v of u in $\mathcal{H}(G)$, $N[u] \subseteq N[v]$ in G, u and its out-neighbours form a k-clique in G, a contradiction.

Now, consider the subgraph of $\mathcal{H}(G)$ induced by $F(u)$. We claim that the longest directed chain in this subgraph has at most $k-1$ vertices. Indeed, all the vertices of such a chain are pairwise adjacent in G. Since G is assumed not to have any k-cliques, there are at most $k-1$ vertices in a directed chain.

Finally, we obtain that $F(u)$ has size at most $\beta(k) = \sum_{i=0}^{k-2}(2k-3)^i$ and the claim of the lemma follows. □

We now need to prove a few additional claims regarding the structure of $\mathcal{H}(G)$. In the following claims, we suppose that G is a twin–free graph.

Claim 4.14. *Let s be a forced vertex in G with $s = f(\overrightarrow{uv})$ for some vertices u and v. If t is an in-neighbour of s in $\mathcal{H}(G)$, then $v = f(\overrightarrow{ts})$. Moreover, if v is forced with $v = f(\overrightarrow{xy})$, then necessarily $y = s$.*

Proof. For the first implication, suppose s has an in-neighbour t in $\mathcal{H}(G)$. An illustration is provided in Figure 4.4. Since $u \not\sim s$, then $u \not\sim t$. Moreover, $v \not\sim t$ since $s = f(\overrightarrow{uv})$. Since $s \sim v$ the claim follows. For the other implication, suppose there exist two vertices x, y such that $v = f(\overrightarrow{xy})$. Hence $y \sim v$ but $x \not\sim v$. Therefore $u \not\sim x$ (otherwise v would be adjacent to x too) and hence $u \not\sim y$. Now the only vertex adjacent to v but not to u is s, so $y = s$. □

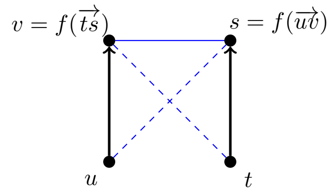


Figure 4.4: The situation of Claim 4.14. Arcs belong to $\mathcal{H}(G)$. Full thin edges belong to G only, dashed edges are nonedges in G.

Claim 4.15. *Let s be a forced vertex in G with $s = f(\overrightarrow{uv})$ for some vertices u and v. Then s has at most one in-neighbour in $\mathcal{H}(G)$.*

Proof. Suppose s has two distinct in-neighbours t and t' in $\mathcal{H}(G)$ (see Figure 4.5 for an illustration). By Claim 4.14, v is both ts–forced and $t's$–forced. But then $N[t] = N[s] \setminus \{v\} = N[t']$. Then t and t' are twins, a contradiction since G is twin–free. □

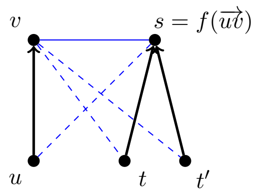


Figure 4.5: The situation of Claim 4.15. Arcs belong to $\mathcal{H}(G)$. Full thin edges belong to G only, dashed edges are nonedges in G.

Claim 4.16. *Let s be a forced vertex in G with $s = f(\overrightarrow{uv})$, and let t be a forced in-neighbour of s in $\mathcal{H}(G)$ with $t = f(\overrightarrow{xy})$ for some vertices u, v, x, y. Then $x = v$.*

Proof. Since $t \sim y$, then $s \sim y$ too. But since $t = f(\overrightarrow{xy})$, $x \sim s$ and $x \not\sim t$. Now by Claim 4.14, $v = f(\overrightarrow{ts})$, that is, v is the unique vertex such that v is adjacent to s, but not to t. Therefore $x = v$. □

We now obtain the following lemma using the previous claims.

Lemma 4.17. *Let s be a nonisolated sink in $\mathcal{H}(G)$ which is forced in G with $s = f(\overrightarrow{uv})$ for some vertices u and v. Then either s has a non–forced predecessor t in $\mathcal{H}(G)$ such that $F(s) \subseteq F(t)$, or there exists a non–forced vertex $w(s)$ such that $F(s) \subseteq N_G[w(s)]$. Moreover, if there are ℓ additional sinks $\{s_1, \ldots, s_\ell\}$ which are all nonisolated in $\mathcal{H}(G)$ and such that $w(s) = w(s_1) = \ldots = w(s_\ell)$, then there exists a set of $\ell + 1$ distinct vertices inducing a clique together with $w(s)$.*

Proof. First of all, recall that $\mathcal{H}(G)$ has no directed circuits. Suppose s has a non–forced predecessor in $\mathcal{H}(G)$ and let t be one such predecessor having the shortest distance to s in $\mathcal{H}(G)$. By Claim 4.15, predecessors of s are either successors or predecessors of t, and there is a directed path from t to s in $\mathcal{H}(G)$. Hence $F(s) \subseteq F(t)$, which proves the first part of the statement.

Now suppose all predecessors of $s = f(\overrightarrow{uv})$ are forced. By Claim 4.15, s and its predecessors form a directed path $\{t_0, \dots, t_m, s\}$ in $\mathcal{H}(G)$ (for an illustration, see Figure 4.6(a)). Note that by Claim 4.14, we have $v = f(\overrightarrow{t_m s})$. By our assumption we know that t_m is forced, say $t_m = f(\overrightarrow{x v_m})$ for some vertices x and v_m. But now by Claim 4.16, $x = v$ and $t_m = f(\overrightarrow{v v_m})$. Now, repeating these arguments for each other predecessor of s shows that there is a directed path $\{u, v, v_m, \dots, v_0\}$ with $t_m = f(\overrightarrow{v v_m})$ and for all i, $0 \leq i \leq m-1$, $t_i = f(\overrightarrow{v_{i+1} v_i})$. In particular, $t_0 = f(\overrightarrow{v_1 v_0})$. Observe also that for all $i \geq 1$, $v_i = f(\overrightarrow{t_{i-1} t_i})$. By applying Claim 4.16 on vertices v_1, v_0 and t_0, if v_0 is forced then t_0 has an in-neighbour in $\mathcal{H}(G)$, a contradiction — hence v_0 is non–forced. Also note that, since $v_0 \sim t_0$, then v_0 is adjacent to all successors of t_0 in $\mathcal{H}(G)$, that is, to all elements of $F(s)$. Therefore, putting $w(s) = v_0$, we obtain the second part of the statement.

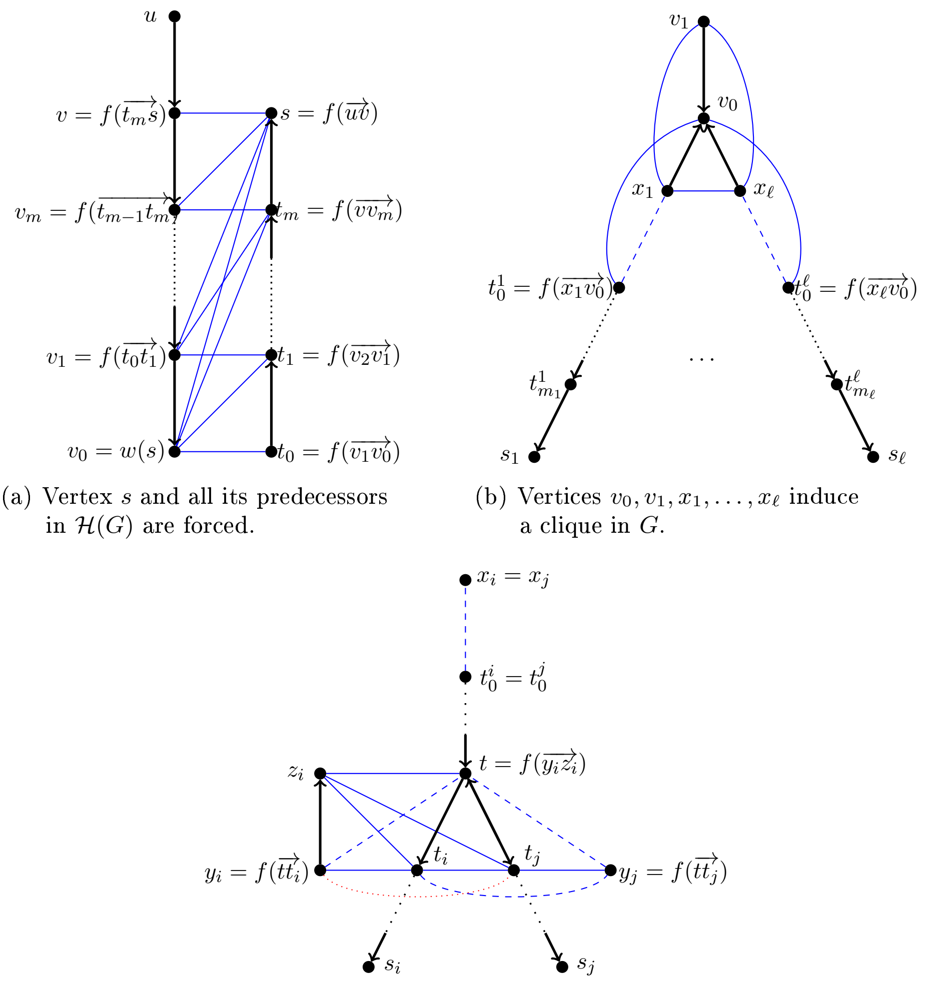
For the last part, suppose there exists a set of ℓ additional forced sinks $\{s_1, \dots, s_\ell\}$ which are nonisolated in $\mathcal{H}(G)$ and such that all their predecessors in $\mathcal{H}(G)$ are forced with $w(s_i) = v_0$ for $1 \leq i \leq \ell$ (for an illustration, see Figure 4.6(b)). For each such sink s_i, by the previous paragraph, the vertices of $F(s_i)$ induce a directed path $\{t_0^i, \dots, t_{m_i}^i, s_i\}$ in $\mathcal{H}(G)$. Moreover we know that there is a vertex x_i such that t_0^i is $x_i v_0$–forced. We claim that the set of vertices $X = \{x_1, \dots, x_\ell\}$ together with v_0 and v_1, form a clique in G of $\ell + 2$ vertices.

We first claim that for all i, j in $\{1, \dots, \ell\}$, $x_i \neq t_0^j$. If $i = j$, this is clear by our assumptions. Otherwise, suppose by contradiction, that $x_i = t_0^j$ for some $i \neq j$ in $\{1, \dots, \ell\}$. Then we claim that $x_j = t_0^i$. Indeed, by the previous part of the proof, we know that $f(\overrightarrow{x_j v_0}) = t_0^j = x_i$ — hence $x_j \not\sim x_i$. But since $\overrightarrow{x_i v_0}$ is an arc in $\mathcal{H}(G)$, we must have $f(\overrightarrow{x_i v_0}) = x_j$. Again, we know that $f(\overrightarrow{x_i v_0}) = t_0^i$, hence $x_j = t_0^i$. Let t_1^i denote the successor of t_0^i in the directed path from t_0^i to s_i in $\mathcal{H}(G)$. We know from the previous part of the proof that $f(\overrightarrow{t_0^i t_1^i}) = x_i = t_0^j$. However since $t_0^i = x_j$ we also know that $f(\overrightarrow{t_0^i v_0}) = x_i$. This implies that $N_G[v_0] = N_G[t_1^i]$, a contradiction since these two vertices are distinct and G is twin–free.

Now, observe that the vertices of X must all be pairwise adjacent. All vertices of X are adjacent to v_0, and for each x_i, $N[v_0] = N[x_i] \cup \{t_0^i\}$, hence x_i is adjacent to all neighbours of v_0 except t_0^i. But by the previous paragraph, we know that $t_0^i \neq x_j$ for all $j \in \{1, \dots, \ell\}$, hence x_i is adjacent to all $x_j \neq x_i$, $j \in \{1, \dots, \ell\}$. For the same reason, each x_i is adjacent to v_1. Hence, the vertices of X form a clique together with v_0 and v_1.

Finally, let us show that all the vertices of X are distinct: by contradiction, suppose that $x_i = x_j$ for some $i \neq j$, $1 \leq i, j \leq \ell$. Since t_0^i is $x_i v_0$–forced and t_0^j is $x_j v_0$–forced, we have $t_0^i = t_0^j$. Since s_i and s_j are distinct, this means that s_i and s_j have one predecessor in common. Hence their common predecessor which is nearest to s_i and s_j, say t, has two out-neighbours. Let t_i (respectively t_j) be the out-neighbour of t which is a predecessor of s_i (respectively s_j) — see Figure 4.6(c) for an illustration. We know that there are two vertices y_i, y_j such that $y_i = f(\overrightarrow{t t_i})$ and $y_j = f(\overrightarrow{t t_j})$. First note that y_i and y_j are distinct: otherwise, we would have $N[t_i] = N[t] \cup \{y_i\} = N[t] \cup \{y_j\} = N[t_j]$ and then t_i, t_j would be twins in G. Observe that since $t \not\sim y_i$ and $y_i \neq f(\overrightarrow{t t_j})$, we have $t_j \not\sim y_i$. We know that t is forced, in fact by the first part of this proof, we also know that $t = f(\overrightarrow{y_i z_i})$ for some vertex z_i. Hence $z_i \sim t$, and since $N[t] \subseteq N[t_j]$,

$z_i \sim t_j$. But since $t_j \neq f(\overrightarrow{y_i z_i})$, $t_j \sim y_i$, a contradiction. Hence x_i and x_j are distinct, which completes the proof. □



(a) Vertex s and all its predecessors in $\mathcal{H}(G)$ are forced.

(b) Vertices $v_0, v_1, x_1, \dots, x_\ell$ induce a clique in G.

(c) If $x_i = x_j$, the dotted edge $y_i t_j$ is both an edge and a nonedge of G.

Figure 4.6: Three situations in the proof of Lemma 4.17. Arcs belong to $\mathcal{H}(G)$. Full thin edges belong to G only, dashed edges are non–edges in G.

Finally, let us recall and prove Proposition 4.10.

Proposition. *Let G be a graph having no k-clique. Then there exists a constant $\gamma(k)$ depending only on k, such that $f(G) \geq \frac{1}{\gamma(k)}$.*

Proof. To prove the result, we use $\mathcal{H}(G)$ to construct a set $X = \{x_1, \dots, x_\ell\}$ of non–forced vertices such that $\bigcup_{i=1}^{\ell} A(x_i) = V(G)$, where $A(x_i)$ is a set of at most $\gamma(k)$ vertices. Then we have $\ell \geq \frac{n}{\gamma(k)}$ vertices in X and the claim of the proposition follows.

We now describe a procedure to build set X while considering each non–isolated sink of $\mathcal{H}(G)$. We denote by s the currently considered sink.

Case 1: Sink s is non–forced. Then we set $A(s)$ to be $F(s)$ together with all the vertices which are forced by a pair u, v of vertices of $F(s)$. Note that by Lemma 4.13, $|F(s)| \leq \beta(k)$, where $\beta(k)$ only depends on k. Hence, $|A(s)| \leq \beta(k) + \binom{\beta(k)}{2}$.

Case 2: Sink s is forced. By Lemma 4.17, either s has a non–forced predecessor t such that $F(s) \subseteq F(t)$, or there exists a non–forced vertex $w(s)$ such that $F(s) \subseteq N_G[w]$.

In the first case, we choose t as our non–forced vertex, and we set $A(t)$ to be $F(t)$ together with all the vertices which are forced by a pair u, v of vertices of $F(t)$. Again we have $|A(t)| \leq \beta(k)+\binom{\beta(k)}{2}$.

In the second case, we choose $w = w(s)$ as our non–forced vertex. Now, let $S = \{s, s_1, \ldots, s_\ell\}$ be the set of forced sinks having no non–forced predecessor and such that $w(s) = w(s_1) = \ldots w(s_\ell)$. By Lemma 4.17 we know that there are $\ell + 1$ distinct vertices inducing a clique together with w, hence $\ell + 2 < k$. We set $A(w)$ to be $F(w) \cup F(s) \cup F(s_1) \cup \ldots \cup F(s_\ell)$ together with all the vertices which are forced by a pair u, v of vertices of this set. We have $|A(w)| \leq k\beta(k) + \binom{k\beta(k)}{2}$.

We have now covered all the vertices which are not isolated in $\mathcal{H}(G)$, since for each nonisolated sink s of $\mathcal{H}(G)$, $F(s)$ is a subset of $A(x)$ for some $x \in X$. Besides, all isolated vertices of $\mathcal{H}(G)$ which are forced, have also been put into some set $A(x)$. Hence only non–forced isolated vertices of $\mathcal{H}(G)$ need to be covered. For each such vertex v, we add v to X and set $A(v) = \{v\}$.

Finally, all vertices belong to some set $A(x)$, $x \in X$, and the size of each set $A(x)$ is at most $\gamma(k) = k\beta(k) + \binom{k\beta(k)}{2}$, which completes the proof. □

4.3 Upper bounds for graphs with girth at least 5

This section is devoted to the study of graphs that have girth at least 5. We will use these results in Section 4.4, to deal with random regular graphs.

One can check that for graphs of girth 5, applying the Local Lemma does not lead to meaningful results. However, by using the Alteration method, a better bound can be given.

We start by defining an auxiliary notion that will be used in this section. A subset $D \subseteq V(G)$ is called a 2–*dominating set* if for each vertex v of $V(G) \setminus D$, $|N(v) \cap D| \geq 2$ (see [57]). The next lemma shows that we can use a 2–dominating set to construct an identifying code.

Lemma 4.18. *Let G be a twin–free graph on n vertices having girth at least 5. Let D be a 2–dominating set of G. If the subgraph induced by D, $G[D]$, has no isolated edge, D is an identifying code of G.*

Proof. First observe that D is dominating since it is 2–dominating. Let us check that D is also separating.

Note that all the vertices that do not belong to D are separated because they are dominated at least twice each and $g(G) > 4$.

Similarly, a vertex $x \in D$ and a vertex $y \in V(G) \setminus D$ are separated since y has two vertices which dominate it, but they cannot both dominate x (otherwise there would be a triangle or a 4-cycle in G).

Finally, consider two vertices of D. If they are not adjacent they are separated by themselves. Otherwise, by the assumption that $G[D]$ has no isolated edge and that G has no triangles, we know that at least one of them has a neighbour in D, which separates them since it is not a neighbour of the other. □

The following theorem makes use of Lemma 4.18. The idea of the proof is inspired by a probabilistic proof of a result on dominating sets which can be found for instance in [8, Theorem 1.2.2].

Theorem 4.19. *Let G be a graph on n vertices with minimum degree d and girth at least 5. Then*

$$\gamma^{ID}(G) \leq (1 + o_d(1))\frac{3\log d}{2d}n \; .$$

Moreover, if G has average degree $\overline{d} = O_d(d(\log d)^2)$, then,

$$\gamma^{ID}(G) \leq \frac{\log d + \log\log d + O_d(1)}{d}n \; .$$

Proof. Let $S \subseteq V(G)$ be a random subset of vertices, where each vertex $v \in V(G)$ is added to S uniformly at random with probability p (where p will be determined later). For every vertex $v \in V(G)$, we define the random variable X_v as follows,

$$X_v = \begin{cases} 0 & \text{if } |N[v] \cap S| \geq 2 \\ 1 & \text{otherwise.} \end{cases} \; .$$

Let $T = \{v \mid X_v = 1\}$. This set contains, in particular, the subset of vertices which are not 2–dominated by S. Note that $|T| = \sum X_v$. Let us estimate the size of T. Observing that $|N[v] \cap S| \sim \text{Bin}(\deg(v) + 1, p)$ and $\deg(v) \geq d$, we obtain

$$\begin{aligned} \mathbb{E}(|T|) &= \sum_{v \in V(G)} \mathbb{E}(X_v) \\ &\leq n\left((1-p)^{d+1} + (d+1)p(1-p)^d\right) \\ &= n(1-p)^d((1-p) + (d+1)p) \\ &\leq n(1+dp)e^{-dp} \; , \end{aligned}$$

where we have used the fact that $1 - x \leq e^{-x}$. Now, note that the set $D = S \cup T$ is a 2–dominating set of G. We have $|D| \leq |S| + |T|$. Hence,

$$\begin{aligned} \mathbb{E}(|D|) &\leq \mathbb{E}(|S|) + \mathbb{E}(|T|) \\ &\leq np + n(1+dp)e^{-dp} \; . \end{aligned} \tag{4.5}$$

Let us set $p = \frac{\log d + \log\log d}{d}$. Plugging this into Equation (4.5), we obtain

$$\mathbb{E}(|D|) \leq \frac{\log d + \log\log d}{d}n + \frac{1 + \log d + \log\log d}{d \log d}n = \frac{\log d + \log\log d + O_d(1)}{d}n \; .$$

This shows that there exists at least one 2–dominating set D having this size.

Case 1: (general case) Note that we can use Lemma 4.18 by considering all pairs u, v of vertices of D forming an isolated edge in $G[D]$, and add an arbitrary neighbour of either one of them to D. Observe that such a vertex exists, otherwise u and v would be twins in G. Since there are at most $\frac{|D|}{2}$ such pairs, we obtain a 2–dominating set of size at most $|D| + \frac{|D|}{2} = (1 + o_d(1))\frac{3\log d}{2d}n$ having the desired property. Now applying Lemma 4.18 completes Case 1.

Case 2: (sparse case) Whenever $\overline{d} = O_d(d(\log d)^2)$, we can get a better bound by estimating the number of isolated edges of $G[D]$. For convenience, we define the random variables Y_{uv} for each edge uv of G, as follows,

$$Y_{uv} = \begin{cases} 1 & \text{if } N[u]\Delta N[v] \subseteq V(G) \setminus S \\ 0 & \text{otherwise} \end{cases} .$$

An isolated edge in $G[D]$ might have been created in several ways. First, at the initial construction step of S: if both u, v belong to S, but none of their other neighbours do which happens with probability at most $p^2(1-p)^{2d-2}$. A second possibility is in the step where we add the vertices of T to our solution. This could happen if both u, v were not dominated at all by S, which occurs with probability at most $(1-p)^{2d}$, or if exactly one of u, v was part of S and none of their neighbours were, which has probability at most $2p(1-p)^{2d-1}$. Thus, the total probability of having an isolated edge in $G[D]$ is bounded from above as follows.

$$\Pr(Y_{uv} = 1) \leq p^2(1-p)^{2d-2} + (1-p)^{2d} + 2p(1-p)^{2d-1} = (1-p)^{2d-2} .$$

Using the previous observation together with the facts that $p = \frac{\log d + \log\log d}{d}$ and $1 - x \leq e^{-x}$, let us calculate the expected value of $Y = \sum_{uv\in E(G)} Y_{uv}$.

$$\mathbb{E}(Y) = \sum_{uv\in E(G)} \mathbb{E}(Y_{uv}) \leq \frac{n\overline{d}}{2}(1-p)^{2d-2} \leq \frac{n\overline{d}}{2}e^{-(2d-2)p} = \frac{n\overline{d}e^{-2(\log d+\log\log d)}}{2} = \frac{n\overline{d}}{2d^2(\log d)^2} .$$

We construct U by picking an arbitrary neighbour of either u or v for each edge uv such that $Y_{uv} = 1$. We have $|U| \leq Y$. The final set $\mathcal{C} = S \cup T \cup U$ is an identifying code. Now we have

$$\mathbb{E}(|\mathcal{C}|) \leq \mathbb{E}(|S|) + \mathbb{E}(|T|) + \mathbb{E}(|U|) \leq \frac{\log d + \log\log d + O_d(1)}{d}n + \frac{\overline{d}}{2d^2(\log d)^2}n .$$

Using that $\overline{d} = O_d(d(\log d)^2)$,

$$\mathbb{E}(|\mathcal{C}|) \leq \frac{\log d + \log\log d + O_d(1)}{d}n . \tag{4.6}$$

Then there exists some choice of S such that $|\mathcal{C}|$ has the desired size, and completes the proof. □

In fact, it is showed in the next section (Corollary 4.23) that Theorem 4.19 is asymptotically tight.

4.4 Identifying codes of random regular graphs

From the study of regular graphs arises the question of the value of the identifying code number for most regular graphs. We know some lower and upper bounds for this parameter, but is it concentrated around some value? A good way to study this question is to look at random regular graphs.

We will use the Configuration Model to study random regular graphs for a constant d. Recall that this model has been defined in Section 2.3.1.

The following theorem provides an upper bound on the identifying number that holds with high probability.

Theorem 4.20. *Let $G \in G(n,d)$ then for any $d \geq 3$, with high probability,*

$$\gamma^{ID}(G) \leq \tfrac{\log d + \log\log d + O_d(1)}{d} n \ .$$

Proof. First of all we have to show that almost all random regular graphs are twin–free.

Observe that the number of perfect matchings of K_{2m} is $(2m-1)!! = (2m-1)(2m-3)(2m-5)\dots 1$. Fix a vertex u of G and let $N(u) = \{v_1, \dots, v_d\}$. We bound from above the probability that u and v_1 are twins, i.e. $N[u] = N[v_1]$. The number of perfect matchings of K_{nd} such that in the resulting graph G of $G(n,d)$, v_1 and v_2 are adjacent, is at most $(d-1)(d-1)(nd-2d-3)!!$. Indeed, there must be an edge between v_1 and v_2, which gives $(d-1)(d-1)$ possibilities. Since u has d neighbours, the number of possibilities for the remaining graph is the number of perfect matchings of $K_{nd-2d-2}$.

Analogously the number of perfect matchings with $v_2, v_3 \in N(v_1)$ is at most $(d-1)(d-1)(d-2)(d-1)(nd-2d-5)!!$. Thus, we have

$$\begin{aligned}
\Pr(N[u] = N[v_1]) &\leq \Pr(N[u] \subseteq N[v_1]) \\
&= \frac{(d-1)(d-1)(d-2)(d-1)\dots 2(d-1)1(d-1)(nd-4d+1)!!}{(nd-2d-1)!!} \\
&\leq \frac{d^{d-1}(d-1)!}{(nd-2d-1)\dots(nd-4d+3)} \\
&\leq \left(\frac{d}{n}\right)^{d-1} ,
\end{aligned}$$

for n large enough.

As we have at most $\frac{nd}{2}$ possible pairs of twins (one for each edge), by the union bound and since $d \geq 3$, for sufficiently large n we obtain

$$\Pr(G \text{ has twins}) \leq \frac{nd}{2}\left(\frac{d}{n}\right)^{d-1} ,$$

which tends to 0 when $n \to \infty$. Therefore, random regular graphs are twin–free with high probability.

By (4.6), for any $G \in G(n,d)$, we have a set $\mathcal{C}$ with

$$|\mathcal{C}| \leq \frac{\log d + \log\log d + O_d(1)}{d} n ,$$

that separates any pair of vertices except from the ones where both vertices belong to a triangle or a 4-cycle. We have to add some vertices to $\mathcal{C}$ in order to separate the vertices of these small cycles.

Classical results on random regular graphs (independently showed in [23, Corollary 2.19] and in [128]) state that the random variables that count the number of cycles of length k, X_k, tend in distribution to independent Poisson variables with parameter $\lambda_k = \frac{1}{2k}(d-1)^k$.

Observe that

$$\mathbb{E}(X_3) = \frac{(d-1)^3}{6} \quad \text{and} \quad \mathbb{E}(X_4) = \frac{(d-1)^4}{8} ,$$

i.e. a constant number of triangles and 4-cycles are expected.

Using Markov's inequality we can bound the probability of having too many small cycles,

$$\Pr(X_3 > t) \leq \frac{(d-1)^3}{6t} \quad \text{and} \quad \Pr(X_4 > t) \leq \frac{(d-1)^4}{8t} .$$

Setting $t = \vartheta(n)$, where $\vartheta(n) \to \infty$, the previous probabilities are $o(1)$. Then, with high probability, we have at most $\vartheta(n)$ cycles of length 3 and $\vartheta(n)$ cycles of length 4.

Let $T = \{u_1, u_2, u_3\}$ be a triangle in G. As $d \geq 3$ there exists at least one vertex v_i outside the triangle (moreover, we showed that the graph has no twins *w.h.p.*). Since our graph is twin–free, for each ordered pair (u_i, u_j) there exists some vertex v_{ij}, such that $v_{ij} \in N(u_i) \backslash N(u_j)$. Observe that we can add v_{12}, v_{23} and v_{31} to $\mathcal{C}$ and then any pair of vertices from T will be separated.

If $T = \{u_1, u_2, u_3, u_4\}$ induces a K_4, each pair of vertices of T is contained in some triangle and is separated by the last step. If T induces a 4-cycle, adding T to $\mathcal{C}$ separates all the elements in T. Otherwise, T induces two triangles and adding T to $\mathcal{C}$ separates the two vertices which have not been separated in the last step.

After these two steps, we have added at most $7\vartheta(n)$ vertices to $\mathcal{C}$. Hence, for any $G \in G(n,d)$ *w.h.p.* we obtain

$$\gamma^{\text{ID}}(G) \leq \frac{\log d + \log\log d + O_d(1)}{d} n + 7\vartheta(n) = \frac{\log d + \log\log d + O_d(1)}{d} n .$$

Observe that the $\frac{O_d(1)}{d} n$ term contains the $7\vartheta(n)$ term. □

Theorem 4.20 shows that despite the fact that for any d, we know infinitely many d–regular graphs having a very large identifying code number (e.g. $n - \frac{n}{d}$ for the graphs of Construction 4.25 of Section 4.5), almost all d–regular graphs have a very small identifying code.

Moreover, $\gamma^{\text{ID}}(G)$ is concentrated, as the following theorem and its corollary show. In fact the following result might be already known, since a similar result is stated for independent dominating sets in [76]. However we could not find it in the literature and decided to give a proof for the sake of completeness.

Theorem 4.21. *Let $G \in G(n,d)$, then w.h.p. all the dominating sets of G have size at least*

$$\frac{\log d - 2\log\log d}{d} n .$$

Proof. We will proceed by contradiction. Given a set of vertices D of size m, we will compute the probability that D dominates $Y = V(G) \setminus D$. Recall that G has been obtained from the configuration model by selecting a random perfect matching of K_{nd}. Let $y \in Y$ fixed, then let $A_y = \{N(D) \cap \{y\} \neq \emptyset\}$ be the event that y is dominated by D. Its complementary event corresponds to the situation where none of the edges of the perfect matching of K_{nd} connects the points corresponding to y to the ones corresponding to any vertex of D. Define $W_D = \cup_{v \in D} W_v$ as the set of cells corresponding to D in K_{nd}. Then for any $v \in W_D$, the event B_v corresponds to the fact that v is not connected to any point in W_y. If $W_D = \{v_1, \dots, v_{md}\}$,

$$\begin{aligned}
\Pr(\overline{A_y}) &= \Pr(\cap_{v \in W_D} B_v) \\
&= \Pr(B_{v_1}) \Pr(B_{v_2} \mid B_{v_1}) \dots \Pr(B_{v_{md}} \mid \cap_{i=1}^{md-1} B_{v_i}) \\
&= \left(1 - \frac{d}{nd-1}\right)\left(1 - \frac{d}{nd-3}\right) \dots \left(1 - \frac{d}{nd-(2md-1)}\right) \\
&= \prod_{i=1}^{md} \left(1 - \frac{d}{nd-(2i-1)}\right) \\
&\geq \prod_{i=1}^{md} \left(1 - \frac{1}{n-2m}\right) .
\end{aligned}$$

Since $1 - x = e^{-x + (\log(1-x)+x)}$ (here we take $x = \frac{1}{n-2m}$) and $\log(1-x) + x = O(x^2)$ (by the Taylor expansion of the logarithm in $x = 0$), we obtain

$$\begin{aligned}
\Pr(\overline{A_y}) &\geq \exp\left(-\sum_{i=1}^{md} \frac{1}{n-2m} + O\left(\frac{1}{(n-2m)^2}\right)\right) \\
&= \exp\left(-(1+o(1))\frac{md}{n-2m}\right) .
\end{aligned}$$

The probability that D is dominating all vertices of $Y = \{y_1, \dots, y_{n-m}\}$ is

$$\Pr\left(\cap_{y \in Y} A_y\right) = \Pr\left(A_{y_1}\right) \Pr\left(A_{y_2} \mid A_{y_1}\right) \dots \Pr\left(A_{y_{n-m}} \mid \cap_{j=1}^{n-m-1} A_{y_j}\right) .$$

We claim that $\Pr\left(A_{y_i} \mid \cap_{j=1}^{i-1} A_{y_j}\right) \leq \Pr\left(A_{y_i}\right)$. Suppose that $y_1, \dots, y_{i-1}$ are dominated. This means that the corresponding perfect matching of K_{nd} has an edge between one of the points corresponding to y_j $(1 \leq j \leq i-1)$ and one of the points corresponding to the vertices of D. The probability that y_i is not dominated by D is now the probability that none of the remaining edges of the perfect matching connect any vertex of D with y_i. Hence,

$$\begin{aligned}
\Pr\left(\overline{A_{y_i}} \mid \cap_{j=1}^{i-1} A_{y_j}\right) &= \left(1 - \frac{d}{nd-2i+1}\right)\left(1 - \frac{d}{nd-2i-1}\right) \dots \left(1 - \frac{d}{nd-2md+1}\right) \\
&\geq \left(1 - \frac{d}{nd-1}\right)\left(1 - \frac{d}{nd-3}\right) \dots \left(1 - \frac{d}{nd-2md+1}\right) \\
&= \Pr(\overline{A_{y_i}}) .
\end{aligned}$$

By considering the complementary events, $\Pr\left(A_{y_i} \mid \cap_{j=0}^{i-1} A_{y_j}\right) \leq \Pr\left(A_{y_i}\right)$. Hence these events are negatively correlated, and

$$\Pr\left(\cap_{y \in Y} A_y\right) \leq \prod_{i=1}^{n-m} \Pr(A_{y_i}) \leq \left(1 - e^{-\frac{md}{n-2m}}\right)^{n-m} \leq \exp\left\{-(n-m)e^{-\frac{md}{n-2m}}\right\} .$$

For the sake of contradiction, let $m \leq \frac{\log d - c\log\log d}{d} n$ for some $c > 2$. Then,

$$\begin{aligned}
\Pr\left(\cap_{y\in Y} A_y\right) &\leq \exp\left(-\left(1-\frac{\log d - c\log\log d}{d}\right) n \exp\left\{-\frac{\log d - c\log\log d}{1-2^{\frac{\log d - c\log\log d}{d}}}\right\}\right)\\
&= \exp\left(-\left(1+o_d(1)\right) n \exp\left\{-\frac{\log d - c\log\log d}{1+o_d(1)}\right\}\right)\\
&= (1+o_d(1)) e^{-\frac{(\log d)^c}{d} n}\ .
\end{aligned}$$

Note that if no set of size m dominates Y, neither will do a smaller one. So we have to look just at the sets of size m. The number of these sets can be bounded by

$$\begin{aligned}
\binom{n}{m} \leq \frac{n^m}{m!} &\leq \left(\frac{en}{m}\right)^m\\
&= \left(\frac{de}{\log d - c\log\log d}\right)^{\frac{\log d - c\log\log d}{d} n}\\
&= (1+o_d(1)) \left(\frac{de}{\log d}\right)^{\frac{\log d - c\log\log d}{d} n}\ ,
\end{aligned}$$

where we have used $m! \geq \left(\frac{m}{e}\right)^m$.

Let A_{DS} be the event that G has a dominating set of size m. Applying the union bound, we have

$$\begin{aligned}
\Pr(A_{DS}) &\leq (1+o_d(1)) \left(\frac{de}{\log d}\right)^{\frac{\log d - c\log\log d}{d} n} e^{-\frac{(\log d)^c}{d} n}\\
&= (1+o_d(1)) \exp\left(\frac{\log d - c\log\log d}{d}(\log d + 1 - \log\log d) n - \frac{(\log d)^c}{d} n\right)\\
&= (1+o_d(1)) \exp\left(\left(\frac{(\log d)^2}{d} - \frac{(\log d)^c}{d} + o_d\left(\frac{(\log d)^2}{d}\right)\right) n\right) \longrightarrow 0\ ,
\end{aligned}$$

since $c > 2$. This shows that *w.h.p.* no set of size less than $\frac{\log d - 2\log\log d}{d} n$ can dominate the whole graph and completes the proof. □

Since any identifying code is a dominating set, we obtain the following immediate corollary.

Corollary 4.22. *Let $G \in G(n,d)$, then, with high probability,*

$$\gamma^{ID}(G) \geq \frac{\log d - 2\log\log d}{d} n\ .$$

Plugging together Theorems 4.20 and 4.21, we can provide the following result.

Corollary 4.23. *Let $G \in G(n,d)$, then, with high probability,*

$$\frac{\log d - 2\log\log d}{d} n \leq \gamma^{ID}(G) \leq \frac{\log d + \log\log d + O_d(1)}{d} n\ .$$

4.5 Extremal constructions

This section gathers some constructions which show the tightness of some of our upper bounds. Some of these constructions can be found in [58].

Construction 4.24. *Given any d_H–regular multigraph H (without loops) on n_H vertices, let $\mathcal{C}_1(H)$ be the graph on $n = n_H(d_H+1)$ and maximum degree $\Delta = d_H + 1$ constructed as follows.*

1. *Replace each vertex v of H by a clique $K(v)$ of $d_H + 1$ vertices*
2. *For each vertex v of H, let $N(v) = \{v_1, \ldots, v_{d_H}\}$ and $K(v) = \{k_0(v), \ldots, k_{d_H}(v)\}$. For each $k_i(v)$ but one ($1 \le i \le d_H$), connect it with an edge in $\mathcal{C}_1(H)$, to a unique vertex of $K(v_i)$, denoted $f(k_i(v))$.*

One can see that the graphs $\mathcal{C}_1(H)$ given by Construction 4.24 are twin–free. Moreover, for each vertex v of H and for each $1 \le i \le d_H$, note that $f(k_i(v))$ is $k_0(v)k_i(v)$–forced. Therefore $\mathcal{C}_1(H)$ has $d_H n_H = n - \frac{n}{\Delta}$ forced vertices. In fact these forced vertices form an identifying code, therefore $\gamma^{\text{ID}}(\mathcal{C}_1(H)) = n - \frac{n}{\Delta}$. An example of this construction is given in Figure 4.7, where H is the hypercube of dimension 3, H_3, and the black vertices are those which belong to a minimum identifying code of $\mathcal{C}_1(H_3)$.

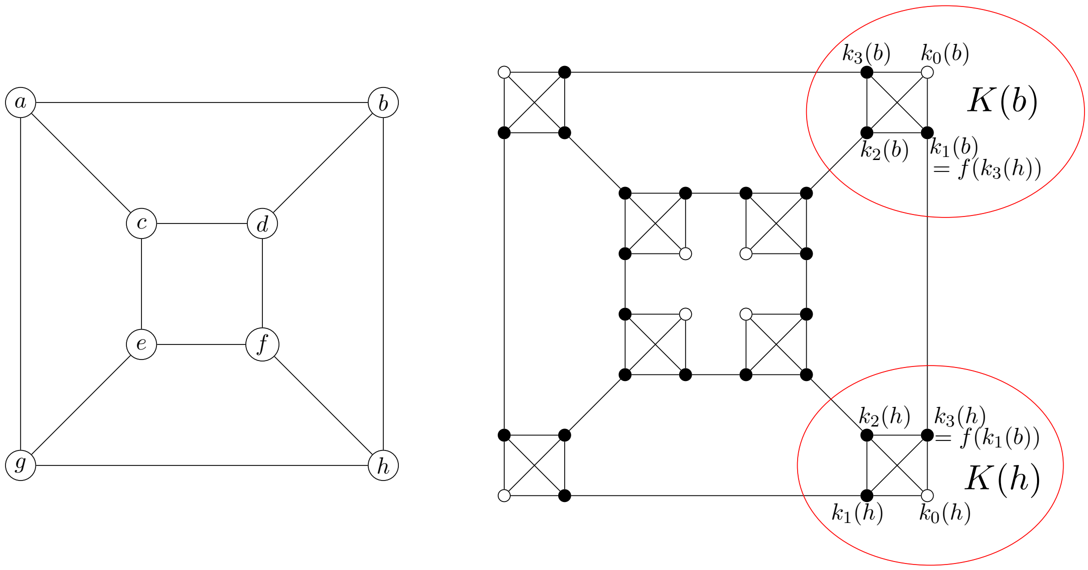


Figure 4.7: The graphs H_3 and $\mathcal{C}_1(H_3)$.

The following construction is very similar, but yields regular graphs.

Construction 4.25. *[58] Given any d_H–regular multigraph H (without loops) on n_H vertices, let $\mathcal{C}_2(H)$ be the d–regular graph on $n = n_H d_H$ vertices (where $d = d_H$) constructed as follows.*

1. *Replace each vertex v of H by a clique $K(v)$ of d_H vertices.*
2. *For each vertex v of H, let $N(v) = \{v_1, \ldots, v_{d_H}\}$ and $K(v) = \{k_1(v), \ldots, k_{d_H}(v)\}$. For each $k_i(v)$ ($1 \le i \le d_H$), connect it with an edge in $\mathcal{C}_2(H)$, to a unique vertex of $K(v_i)$, denoted $f(k_i(v))$.*

Note that for some vertex v of H, in order to separate each pair of vertices $k_i(v), k_j(v)$ of $K(v)$ in $\mathcal{C}_2(H)$, either $f(k_i(v))$ or $f(k_j(v))$ must belong to any identifying code. Repeating this argument for each pair shows that at least $d-1$ such vertices are needed in the code. Since for any two cliques $K(u)$ and $K(v)$, the set of these neighbours are disjoint, this shows that at least $n_H(d-1)$ vertices are needed in an identifying code of $\mathcal{C}_2(H)$. In fact it is easy to construct an identifying code of this size. This shows that despite the fact that $\mathcal{C}_2(H)$ has no forced vertices, $\gamma^{\mathrm{ID}}(\mathcal{C}_2(H)) = n - \frac{n}{d}$. An example of this construction is given in Figure 4.8, where H is the complete graph K_5, and the black vertices form a minimum identifying code of $\mathcal{C}_2(K_5)$.

Construction 4.24 and 4.25 are close to Sierpiński graphs, which were defined in [89]. Recently in [72], it has been showed that Sierpiński graphs are also extremal with respect to Conjecture 4.2, i.e. for any Sierpiński graph G on n vertices with maximum degree Δ, $\gamma^{\mathrm{ID}}(G) = n - \frac{n}{\Delta}$.

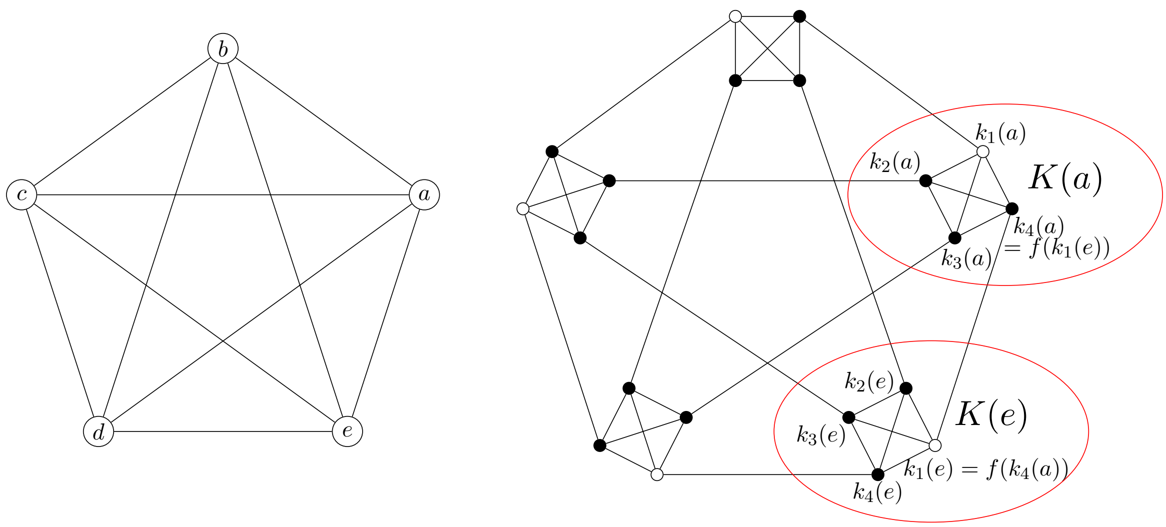


Figure 4.8: The graphs K_5 and $\mathcal{C}_2(K_5)$.

Construction 4.26. *[58] Given an even number $2k$ and an integer $d \geq 3$, we construct a twin–free d–regular triangle–free graph $\mathcal{C}_3(2k, d)$ on $n = 2kd$ vertices as follows.*

1. *Let $\{c_0, \ldots, c_{2k-1}\}$ be a set of $2k$ vertices and add the edges of the perfect matching $\{c_1c_2, \ldots, c_{2k-3}c_{2k-2}, c_{2k-1}c_0\}$.*

2. *For each even i $(0 \leq i \leq 2k-2)$, build a copy $K(i)$ of the complete bipartite graph $K_{d-1,d-1}$. Join vertex c_i to all vertices of one part of the bipartition of $K(i)$, and join vertex c_{i+1} to all other vertices of $K(i)$.*

Consider an identifying code of $\mathcal{C}_3(2k, d)$. Note that in each copy $K(i)$ of $K_{d-1,d-1}$, at least $2d-4$ vertices belong to the code in order to separate the vertices being in the same part of the bipartition of $K(i)$. Now if exactly $2d-4$ vertices of $K(i)$ belong to the code, in order to separate the two remaining vertices, either c_i or c_{i+1} belongs to the code. Hence for each odd i, at most three vertices from $\{c_i, c_{i+1}\} \cup V(K(i))$ do not belong to a code of $\mathcal{C}_3(2k, d)$. On the other hand, taking all vertices c_i such that i is even together with $d-2$ vertices of each part of the bipartition of each copy of $K_{d-1,d-1}$ yields an identifying code of this size. Hence $\gamma^{\mathrm{ID}}(\mathcal{C}_3(2k, d)) = k + 2k(d-2) = n - \frac{n}{2d/3}$. An example of this construction is given in Figure 4.9, where $2k = 8$, $d = 3$, and the black vertices form a minimum identifying code of $\mathcal{C}_3(8, 3)$.

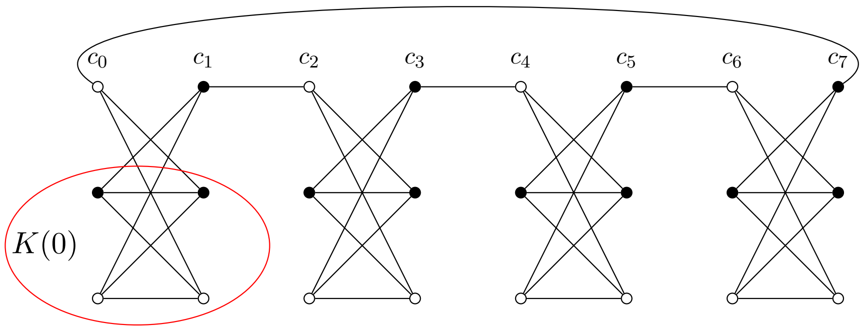


Figure 4.9: The graph $\mathcal{C}_3(8, 3)$.

4.6 Concluding remarks and open questions

1. Motivated by the graph B_k (see Figure 4.3), we ask the following question.

Question 4.27. *Does there exist a value of Δ and an infinite family of graphs with maximum degree Δ having exactly $\frac{n}{\Delta+1}$ non–forced vertices?*

Answering this question in a positive way would provide a counterexample to Conjecture 4.2. Note that for the similar question where we replace $\Delta + 1$ by Δ, the answer is positive by Construction 4.24 of Section 4.5. For any Δ, this construction provides arbitrarily large graphs having exactly $\frac{n}{\Delta}$ non–forced vertices.

2. The main question that still needs to be answered is whether Conjecture 4.2 is true for every graph with maximum degree Δ. We have showed that the conjecture is asymptotically true for every graph with a constant proportion of non–forced vertices (Corollary 4.7), which includes regular graphs or graphs with no clique of constant order as a minor (Corollary 4.11).

An easier question is the following weaker version of Conjecture 4.2.

Question 4.28. *Is it true that for any nontrivial connected twin–free graph G of maximum degree Δ, we have*

$$\gamma^{ID}(G) \leq n - \frac{n}{O(\Delta)} \ ?$$

Recall that we have examples with a large number of forced vertices. Thus, it is not true that any graph has a constant proportion of non–forced vertices. Thus, Theorem 4.6 does not suffice to answer the previous question.

Nonetheless, we believe that the same probabilistic approach could give a better result by understanding which is the role of the forced vertices in the proof of Theorem 4.6. In our proof, we do not use the set of forced vertices to identify the rest of the vertices. It is obvious that any identifying code will contain this set of vertices, but they also help to separate the other pairs. For instance, in graph the B_k (see Figure 4.3), the set of forced vertices $V(A_k)$ is already an identifying code.

3. Consider a k–coloring of G, $\chi : V(G) \to [k]$. For any set $S \subseteq V(G)$, we denote by $\chi(S) =$

$\cup_{v \in S}\chi(v) \subseteq [k]$, the set of colors in S. Then, χ is *locally identifying*, if it is proper and for any edge $uv \in E(G)$, with $N[u] \neq N[v]$, we have

$$\chi(N[u]) \neq \chi(N[v]) \ .$$

The locally identifying chromatic number of G, $\chi_{lid}(G)$, is the minimum integer k such that G admits a locally identifying coloring.

It was conjectured in [55] that for any graph G with maximum degree Δ,

$$\chi_{lid}(G) = O(\Delta^2) \ . \tag{4.7}$$

In the same paper, the authors show that for such graphs, $\chi_{lid}(G) = O(\Delta^3)$.

Using the technique displayed in the proof of Theorem 4.6, one can show that for any graph G with maximum degree Δ, (4.7) holds. In [60], we provide a constructive proof to show that

$$\chi_{lid}(G) \leq 2\Delta^2 - 3\Delta + 3 \ .$$

This bound cannot be improved much. Using the projective plane one can construct a graph H with maximum degree Δ and $\chi_{lid}(H) = \Delta^2 - \Delta + 1$.

4. Note that Theorem 4.19 cannot be extended much in the sense that if we drop the condition on girth 5, we know arbitrarily large d–regular triangle–free graphs having large minimum identifying code number. For instance, Construction 4.26 of Section 4.5 provides a graph G which satisfies $\gamma^{\text{ID}}(G) = n - \frac{n}{d}$. Similarly, we cannot drop the minimum degree condition. Indeed it is known that any $(\Delta - 1)$-ary complete tree $T_{\Delta,h}$ of height h, which is of maximum degree Δ, minimum degree 1 and has infinite girth, also has a large identifying code number (i.e. $\gamma^{\text{ID}}(T_{\Delta,h}) = n - \frac{n}{\Delta - 1 + o_\Delta(1)}$ [17]).

CHAPTER 5

Large spanning subgraphs admitting small identifying codes

5.1 Introduction

Consider any graph parameter that is not monotone with respect to graph inclusion. Given a graph G, a natural problem in this context is to study the minimum value of this parameter over all spanning subgraphs of G. In particular, how many edge deletions are sufficient in order to obtain from G a graph with near–optimal value of the parameter? Herein, we use random methods to study this question with respect to the identifying code number of a graph, a well–studied non–monotone parameter. An identifying code of graph G is a set $\mathcal{C}$ of vertices which is a dominating set, and such that the closed neighborhood within $\mathcal{C}$ of each vertex v uniquely determines v.

The basic notation on graph theory and a formal definition of identifying code can be found in Section 4.1 of the previous chapter. Recall that for every twin–free graph G on n vertices having at least one edge, we have

$$\lceil \log_2(n+1) \rceil \leq \gamma^{\mathrm{ID}}(G) \leq n-1 \; .$$

In view of the above lower bound, we say that an identifying code $\mathcal{C}$ of G is *asymptotically optimal* if

$$|\mathcal{C}| = O(\log n) \; .$$

In this chapter we will deal with graphs that have a large identifying code number or that do not admit an identifying code, this is, they contain twins. Our goal will consist in slightly modifying such graphs in order to decrease their identifying code number and obtain an asymptotically optimal identifying code, unless its domination number prevents us from doing so.

One of the reasons for a graph to have a large identifying code number is that it has a large domination number (this one being a monotone parameter under edge deletion). For instance, we need a linear size set in order to dominate all the vertices of a bounded degree graph. When this is the case, we cannot expect to decrease much the size of a minimum identifying code by deleting edges from G, as the deletion of edges cannot decrease the domination number.

However, there are many graphs with small domination number where the identifying code number is very large [59, 62]. Typically, this phenomenon appears in graphs having a specific "rigid" structure. Supporting this intuition, Frieze et al. [66] have showed that the random graph $G(n,p)$ with $p \in (0,1)$, admits an asymptotically optimal identifying code. In particular, they prove in [66] that

$$\gamma^{\text{ID}}(G(n,p)) = (1+o(1))\frac{2\log n}{\log(1/q)} ,$$

where $q = p^2 + (1-p)^2$. This suggests that the lack of structure in dense graphs implies the existence of a small identifying code.

By selecting at random a small set of edges that can be deleted to "add some randomness" in the graph, we obtain the following result.

Theorem 5.1. *For every graph G on n vertices (n large enough) with maximum degree $\Delta = \omega(1)$ and minimum degree $d \geq 66\log\Delta$, there exists a subset of edges $F \subset E(G)$ of size*

$$|F| = O(n\log\Delta) ,$$

such that

$$\gamma^{ID}(G \setminus F) = O\left(\frac{n\log\Delta}{d}\right) .$$

Observe that when $d = \Theta(n)$, this result is asymptotically equal to the one in [66].

In order to show Theorem 5.1, we define a suitable random spanning subgraph of G: we first we choose a code $\mathcal{C}$ by selecting each vertex independenty at random, and then we randomly delete edges among the edges containing vertices of $\mathcal{C}$. We then analyze the construction by applying concentration inequalities and the use of the local lemma.

A similar approach has been used in the literature when considering random subgraphs of a graph (see Section 2.3.1). Our random subgraph model is adapted to the analysis of identifying codes, and can be seen as a weighted version of G_p, like the one proposed by Alon [3].

Theorem 5.1 is asymptotically best possible in terms of both the number of deleted edges and the size of the final identifying code for every graph with $\Delta = \text{Poly}(d)$ (see Corollary 5.10). For smaller values of the minimum degree, we prove that our result is almost optimal. We also show that the two conditions, $\Delta = \omega(1)$ and $d \geq c\log\Delta$ for some constant c, are necessary.

When considering the case of adding edges to the graph, we get analogous (symmetric) results, showing that every graph is a large spanning subgraph of some graph that admits a small identifying code. This result also turns out to be tight. We also describe an application to the closely related topic of *watching systems*.

This chapter is organized as follows. In Section 5.2, we define our model of random subgraphs and use it to prove Theorem 5.1. In Section 5.3 we construct a family of graphs to show that Theorem 5.1 is almost tight for all values of d and Δ. We present some other consequences of our result in Section 5.4, in particular, we argue about the case where edges are added to the graph. The chapter concludes with some final remarks and open problems (Section 5.5).

5.2 Main theorem

In this section, we prove Theorem 5.1. We will need some tools and lemmas.

5.2.1 Random subgraphs and identification

In what follows, for every set of vertices $B \subseteq V(G)$ and each $v \in V(G)$, we let $N_G^B(v) = N_G(v) \cap B$ be the set of neighbors of v in B. Analogously, $N_G^B[v] = N_G[v] \cap B$. We denote by $d_B(v) = |N_G^B(v)|$, the degree of v within set B.

Definition 5.2. *Given a graph G and $B \subseteq V(G)$, a function $f : V(G) \to \mathbb{R}^+ \cup \{0\}$ is said to be (G, B)–bounded if for each vertex u, $f(u) \le d_B(u)$ and for each pair u, v of vertices with $d_B(u) \ge d_B(v)$, $f(u)/d_B(u) \le f(v)/d_B(v)$. Given a (G, B)–bounded function f, we define the random spanning subgraph $G(B, f)$ of G as follows:*

- *$G(B, f)$ contains all edges of the subgraph $G[V(G) \setminus B]$ induced by $V(G) \setminus B$, and*
- *each edge uv incident with B is independently chosen to be in $G(B, f)$ with probability $1 - p_{uv}$, where*
$$p_{uv} = \frac{1}{4}\left(\frac{f(u)}{d_B(u)} + \frac{f(v)}{d_B(v)}\right) .$$
Observe that, since $f(u) \le d_B(u)$ for each vertex $u \in V(G)$, we have $p_{uv} \le 1/2$.

The next lemma gives an exponential upper bound on the probability that two vertices of $G(B, f)$ are not separated by B. This lemma is crucial in our main proof.

Lemma 5.3. *Let G be a graph, $B \subseteq V(G)$, and f a (G, B)–bounded function. In the random subgraph $G(B, f)$, for every pair u, v of distinct vertices with $d_B(u) \ge d_B(v)$, we have*

$$\Pr\left(N^B_{G(B,f)}[u] = N^B_{G(B,f)}[v]\right) \le e^{-3f(u)/16} .$$

Proof. Consider the following partition of $S = N_G^B[u] \cup N_G^B[v]$ into three parts: S_1, the vertices of B dominating u but not v; S_2, the vertices of B dominating v but not u; and S_3, the vertices of B dominating both u and v.

Let D be the random variable which gives the size of the symmetric difference of $N^B_{G(B,f)}[u]$ and $N^B_{G(B,f)}[v]$. The statement of the lemma is equivalent to $\Pr(D = 0) < e^{-3f(u)/16}$.

The random variable $D = |N^B_{G(B,f)}[u] \oplus N^B_{G(B,f)}[v]|$ can be written as the sum of independent Bernoulli variables
$$D = \sum_{w \in S} D_w ,$$
where $D_w = 1$ if and only if w dominates precisely one of the two vertices u or v in $G(B, f)$.

Therefore, for any $w \notin \{u, v\}$,

$$\Pr(D_w = 1) = \begin{cases} 1 - p_{uw} & \text{if } w \in S_1, \\ 1 - p_{vw} & \text{if } w \in S_2, \\ p_{uw}(1 - p_{vw}) + p_{vw}(1 - p_{uw}) & \text{if } w \in S_3. \end{cases}$$

Since we want to bound from above the probability that $D = 0$, we can always assume that $u, v \notin N^B_{G(B,f)}[u] \oplus N^B_{G(B,f)}[v]$. Recall that $d_B(u) \geq d_B(v)$. By the definition of a (G, B)–bounded function, we have that $p_{uw} \leq p_{vw}$ for each $w \in S_3$. Since $x(1 - x)$ has a unique maximum at $x = 1/2$ and $p_{uw}, p_{vw} \leq 1/2$, we also have:

$$p_{vw}(1 - p_{uw}) \geq p_{uw}(1 - p_{uw}) \geq \frac{f(u)}{4d_B(u)} \left(1 - \frac{f(u)}{4d_B(u)}\right) = g(u) \ , \tag{5.1}$$

for each $w \in S_3$.

For $w \in S$, denote by q_w the parameter of the Bernoulli random variable D_w. Then,

$$\begin{aligned} \mathbb{E}(D) &\geq \sum_{w \in N^B_G(u)} q_w \\ &= \sum_{w \in S_1} q_w + \sum_{w \in S_3} q_w \\ &= \sum_{w \in S_1} (1 - p_{uw}) + \sum_{w \in S_3} (p_{uw}(1 - p_{vw}) + p_{vw}(1 - p_{uw})) \\ &\geq \sum_{w \in S_1} p_{uw}(1 - p_{uw}) + \sum_{w \in S_3} p_{uw}(1 - p_{uw}) \\ &\geq g(u)d_B(u) \\ &= \frac{f(u)}{4} \left(1 - \frac{f(u)}{4d_B(u)}\right) \\ &\geq \frac{3}{16} f(u) \ . \end{aligned} \tag{5.2}$$

Finally, we have that

$$\Pr(D = 0) = \prod_{w \in S} (1 - q_w) \leq e^{-\sum_{w \in S} q_w} = e^{-\mathbb{E}(D)} \leq e^{-3f(u)/16} \ ,$$

and the lemma follows. □

5.2.2 Proof of the main result

We are now ready to prove the main theorem.

Proof of Theorem 5.1. The proof is structured in the following steps:

1. We select a set $\mathcal{C}$ at random, where each vertex is selected independently with probability p. Using the Chernoff inequality, we estimate the probability of the event $A_{\mathcal{C}}$ that $\mathcal{C}$ is small enough for our purposes. From $\mathcal{C}$, we construct the spanning subgraph $G(\mathcal{C}, f)$ of G as given in Definition 5.2, for some suitable function f.

2. We use the local lemma (Lemma 2.12) and Lemma 5.3 to bound from below the probability that the following events (whose conjunction we call A_{LL}) hold jointly: (i) in $G(\mathcal{C}, f)$, each pair of vertices that are at distance at most 2 from each other are separated by $\mathcal{C}$; and (ii) for each such pair and each member of this pair in G, its degree within $\mathcal{C}$ in G is close to its expected value $d(v)p$. We show that with nonzero probability, $A_{\mathcal{C}}$ and A_{LL} hold jointly.

3. We find a dominating set D of G with $|D| = O(|\mathcal{C}|)$; by Observation 4.3, if A_{LL} holds, then $\mathcal{C} \cup D$ is an identifying code.

4. Finally, we show that, subject to $A_{\mathcal{C}}$ and A_{LL}, the expected number of deleted edges is as small as desired.

Step 1. Constructing $\mathcal{C}$ and $G(\mathcal{C}, f)$

Let $\mathcal{C} \subseteq V(G)$ be a subset of vertices, where each vertex v in G is chosen to be in $\mathcal{C}$ independently with probability

$$p = \frac{66 \log \Delta}{d} .$$

Observe that $p \leq 1$ since $d \geq 66 \log \Delta$.

Consider the random variable $|\mathcal{C}|$ and recall that $\mathbb{E}(|\mathcal{C}|) = np$.

Define $A_{\mathcal{C}}$ to be the event that

$$|\mathcal{C}| \leq 2np = \frac{132 n \log \Delta}{d} . \qquad (A_{\mathcal{C}})$$

Since the choices of the elements in $\mathcal{C}$ are done independently, by setting $\varepsilon = 1$ in Lemma 2.2, notice that $c_\varepsilon > 1/3$, we have

$$\Pr(\overline{A_{\mathcal{C}}}) < e^{-\frac{22 n \log \Delta}{d}} . \qquad (5.3)$$

We let

$$f(u) = \min\left(66 \log \Delta, d_{\mathcal{C}}(u)\right) .$$

Observe that f is $(G, \mathcal{C})$–bounded. We construct $G(\mathcal{C}, f)$ as the random spanning subgraph of G given in Definition 5.2, where each edge uv incident to a vertex of $\mathcal{C}$ is deleted with probability p_{uv}.

Step 2. Applying the local lemma

Let u, v be a pair of vertices at distance at most 2 in G. We define the following events:

- B_{uv} is the event that there exists a vertex $w \in \{u, v\}$ such that the degree of w within $\mathcal{C}$ is deviating from its expected value $d(w)p$ by half, i.e. $|d_{\mathcal{C}}(w) - d(w)p| \geq \frac{d(w)p}{2}$;
- C_{uv} is the event that $N^{\mathcal{C}}_{G(\mathcal{C},f)}[u] = N^{\mathcal{C}}_{G(\mathcal{C},f)}[v]$;
- A_{uv} is the event that B_{uv} or C_{uv} occurs ($A_{uv} = B_{uv} \cup C_{uv}$);
- A_{LL} is the event that no event A_{uv} occurs $\left(A_{LL} = \cap_{uv} \overline{A_{uv}}\right)$.

In order to apply the local lemma, we wish to upper bound the probability of A_{uv}. We have:

$$\begin{aligned}\Pr(A_{uv}) &\le \Pr(B_{uv}) + \Pr(C_{uv}) \\ &= \Pr(B_{uv}) + \Pr(C_{uv}|B_{uv}) \cdot \Pr(B_{uv}) + \Pr(C_{uv}|\overline{B_{uv}}) \cdot \Pr(\overline{B_{uv}}) \ .\end{aligned}$$

Let us upper bound $\Pr(B_{uv})$. We use Lemma 2.2 with $\varepsilon = 1/2$. Observe that $c_\varepsilon > \frac{1}{10}$, and thus

$$\begin{aligned}\Pr(B_{uv}) &< \Pr\left(|d_{\mathcal{C}}(u) - d(u)p| \ge \frac{d(u)p}{2}\right) + \Pr\left(|d_{\mathcal{C}}(v) - d(v)p| \ge \frac{d(v)p}{2}\right) \\ &\le 2e^{-\frac{1}{10}d(u)p} + 2e^{-\frac{1}{10}d(v)p} \\ &= 2e^{-\frac{66d(u)\log\Delta}{10d}} + 2e^{-\frac{66d(v)\log\Delta}{10d}} \\ &\le 4e^{-\frac{33\log\Delta}{5}} \\ &\le 4\Delta^{-\frac{33}{5}} \ .\end{aligned}$$

Next, we give an upper bound for $\Pr(C_{uv}|\overline{B_{uv}})$. For such a purpose, we apply Lemma 5.3 with $B = \mathcal{C}$ and $f(u) = \min(66\log\Delta, d_{\mathcal{C}}(u))$. Observe that f is $(G, \mathcal{C})$–bounded. Since B_{uv} does not hold, we know that $d_{\mathcal{C}}(u)$ and $d_{\mathcal{C}}(v)$ are large enough, i.e. for $w \in \{u, v\}$, $d_{\mathcal{C}}(w) \ge \frac{d(w)p}{2} \ge \frac{dp}{2} = 33\log\Delta$; thus $f(u), f(v) \ge 33\log\Delta$. We have:

$$\Pr(C_{uv}|\overline{B_{uv}}) \le e^{-\frac{3\cdot 33\log\Delta}{16}} \le \Delta^{-\frac{99}{16}} \ . \tag{5.4}$$

The probability that the event A_{uv} holds is

$$\begin{aligned}\Pr(A_{uv}) &\le \Pr(B_{uv}) + \Pr(C_{uv}|B_{uv}) \cdot \Pr(B_{uv}) + \Pr(C_{uv}|\overline{B_{uv}}) \cdot \Pr(\overline{B_{uv}}) \\ &\le 4\Delta^{-\frac{33}{5}} + 1 \cdot 4\Delta^{-\frac{33}{5}} + \Delta^{-\frac{99}{16}} \cdot 1 \\ &\le 2\Delta^{-\frac{99}{16}} =: p_{LL} \ ,\end{aligned}$$

where we used $\Delta = \omega(1)$.

We now note that each event A_{uv} is mutually independent of all but at most $2\Delta^6$ events $A_{u'v'}$. Indeed, A_{uv} depends on the random variables determining the existence of the edges incident to u and v. This is given by probabilities p_{uw} and p_{vw} that depend on $d_{\mathcal{C}}(w)$, where w is at distance at most one from either u or v. Thus, A_{uv} depends only on the vertices at distance at most two from either u or v belonging to $\mathcal{C}$. In other words, A_{uv} and $A_{u'v'}$ are mutually independent unless there exist a vertex w at distance at most two from both pairs; in other words, $d(\{u, v\}, \{u', v'\}) \le 4$. Hence, there are at most $2\Delta^4$ choices for the vertex among $\{u', v'\}$ that is closest from $\{u, v\}$ (say u'), and at most Δ^2 additional choices for v', since $d(u', v') \le 2$.

Therefore, we can apply Lemma 2.12 if

$$e \cdot 2\Delta^{-\frac{99}{16}} \cdot (2\Delta^6 + 1) \le 1 \ ,$$

which holds since $\Delta = \omega(1)$.

Now, by Lemma 2.12 and since there are at most $\frac{n\Delta^2}{2}$ events A_{uv} (one for each pair of vertices at distance at most 2 from each other) and $p_{LL} = 2\Delta^{-\frac{99}{16}}$,

$$\Pr(A_{LL}) \ge (1 - e \cdot p_{LL})^M \ge e^{-2e\cdot p_{LL}M} \ge e^{-2en\Delta^{2-\frac{99}{16}}} \ , \tag{5.5}$$

where we have used $(1-x) = e^{-x(1-O(x))} \geq e^{-2x}$, if $x = o(1)$.

Step 3. Revealing the identifying code

Let us lower bound the probability that both $A_{\mathcal{C}}$ and A_{LL} hold, by using (5.3) and (5.5):

$$\begin{aligned}\Pr(A_{\mathcal{C}} \cap A_{LL}) &\geq \Pr(A_{LL}) - \Pr(\overline{A_{\mathcal{C}}}) \\ &\geq e^{-2en\Delta^{2-\frac{99}{16}}} - e^{-\frac{22n\log\Delta}{d}} ,\end{aligned}$$

which is strictly positive if

$$\frac{22\log\Delta}{d} > 2e\Delta^{2-\frac{99}{16}} ,$$

which holds since n is large (and hence $\Delta = \omega(1)$ is large too), and $d \leq \Delta$.

Hence, there exists a set $\mathcal{C}$ of size $132\frac{n\log\Delta}{d}$ such that all vertices at distance 2 from each other are separated by $\mathcal{C}$, and such that the degree in $\mathcal{C}$ of all vertices is large enough.

In order to build an identifying code, we must also make sure that all vertices are dominated. It is well–known that for every graph G, $\gamma(G) \leq (1+o(1))\frac{n\log d}{d}$ (see e.g. [8, Theorem 1.2.2]). Hence, we select a dominating set D of G with size $(1+o(1))\frac{n\log d}{d}$. Then, by Observation 4.3, $\mathcal{C} \cup D$ is an identifying code of size at most

$$(132+1+o(1))\frac{n\log\Delta}{d} \leq 134\frac{n\log\Delta}{d} .$$

Step 4. Estimating the number of deleted edges

Let $Y = |E(G) \setminus E(G(\mathcal{C}, f))|$ be the number of edges we have deleted from G to obtain $G(\mathcal{C}, f)$. Recall that each edge $uv \in E(G)$ is deleted independently from G with probability

$$p_{uv} = \frac{1}{4}\left(\frac{f(u)}{d_{\mathcal{C}}(u)} + \frac{f(v)}{d_{\mathcal{C}}(v)}\right) ,$$

if one of its endpoints is in $\mathcal{C}$.

Since $\Pr(A_{\mathcal{C}} \cap A_{LL}) > 0$, there is a small identifying code of G obtained by deleting at most $\mathbb{E}(Y|A_{\mathcal{C}} \cap A_{LL})$ edges. We next give an upper bound for $\mathbb{E}(Y|A_{\mathcal{C}} \cap A_{LL})$. If both $A_{\mathcal{C}}$ and A_{LL} hold, then

$$p_{uv} \leq \frac{1}{4}\left(\frac{66\log\Delta}{d_{\mathcal{C}}(u)} + \frac{66\log\Delta}{d_{\mathcal{C}}(v)}\right) .$$

The expected number of deleted edges is

$$\mathbb{E}(Y|A_{\mathcal{C}} \cap A_{LL}) = \sum_{\substack{uv\in E(G)\\ (\{u,v\}\cap\mathcal{C})\neq\emptyset}} p_{uv} .$$

Observe that in order to estimate this quantity, we can split the two additive terms in each p_{uv}: for every $u \notin \mathcal{C}$, we sum all the terms $\frac{66\log\Delta}{4d_{\mathcal{C}}(u)}$ for all $v \in \mathcal{C}$ being neighbors of u; for every $u \in \mathcal{C}$, we sum all the terms $\frac{66\log\Delta}{4d_{\mathcal{C}}(u)}$ for all $v \in V(G)$ being neighbors of u.

$$\begin{aligned}
\mathbb{E}(Y|A_{\mathcal{C}} \cap A_{LL}) &\leq \frac{1}{4}\left(\sum_{u\notin\mathcal{C}}\sum_{v\in N_G^{\mathcal{C}}(u)}\frac{66\log\Delta}{d_{\mathcal{C}}(u)} + \sum_{u\in\mathcal{C}}\sum_{v\in N_G(u)}\frac{66\log\Delta}{d_{\mathcal{C}}(u)}\right)\\
&\leq \frac{1}{4}\left(\sum_{u\notin\mathcal{C}} d_{\mathcal{C}}(u)\frac{66\log\Delta}{d_{\mathcal{C}}(u)} + \sum_{u\in\mathcal{C}} d(u)\frac{66\log\Delta}{d_{\mathcal{C}}(u)}\right)\\
&\leq \frac{1}{4}\left(|V(G)\setminus\mathcal{C}|\cdot 66\log\Delta + \sum_{u\in\mathcal{C}} 2\frac{66\log\Delta}{p}\right)\\
&\leq \frac{1}{4}\left(n\cdot 66\log\Delta + 2|\mathcal{C}|d\right)\\
&\leq \frac{66n\log\Delta + 264n\log\Delta}{4}\\
&\leq 83n\log\Delta\ ,
\end{aligned}$$

where we used the fact (implied by A_{LL}) that for each vertex v, $\frac{d(v)p}{2} \leq d_{\mathcal{C}}(v)$ at the second line, and that $A_{\mathcal{C}}$ implies $|\mathcal{C}| \leq 132\frac{n\log\Delta}{d}$ at the fifth line.

Summarizing, we have showed the existence of a small identifying code in a spanning subgraph of G obtained by deleting at most $\mathbb{E}(Y|A_{\mathcal{C}} \cap A_{LL})$ edges from G, which completes the proof. □

5.3 Asymptotic optimality of Theorem 5.1

In this section, we discuss the optimality of Theorem 5.1, first with respect to the size of the constructed code and the number of deleted edges, and then with respect to the hypothesis $\Delta = \omega(1)$ and $d \geq 66\log\Delta$.

5.3.1 On the size of the code and the number of deleted edges

As commented in the Introduction, by removing edges, the dominating number never decreases. It is well–known (see e.g. [8, Theorem 1.2.2]) that the domination number of a graph with minimum degree d satisfies

$$\gamma(G) \leq (1+o(1))\frac{n\log d}{d}\ . \tag{5.6}$$

This bound is sharp. If G is a tight example for (5.6), then for every subgraph H of G,

$$\gamma^{\mathrm{ID}}(H) \geq \gamma(H) \geq \gamma(G) = (1+o(1))\frac{n\log d}{d}\ .$$

This shows that Theorem 5.1 cannot be improve much in terms of identifying code size. In this section we will show that, indeed, Theorem 5.1 is also tight in terms of number of deleted edges.

Charon, Honkala, Hudry and Lobstein showed that deleting an edge from G can decrease by at most 2 the identifying code number of a graph [33]. That is, for every graph G and any edge $uv \in E(G)$,

$$\gamma^{\mathrm{ID}}(G) \leq \gamma^{\mathrm{ID}}(G\setminus uv) + 2\ .$$

This directly implies that for every graph with linear identifying code number, one needs to delete a subset F of at least $\Omega(n)$ edges, to get a graph with $\gamma^{\text{ID}}(G \setminus F) = o(n)$.

We will show that, indeed, one needs to delete at least $\Omega(n \log n)$ edges from the complete graph to get a graph with an asymptotically optimal identifying code. Using this, we will derive a family of graphs with arbitrary minimum degree d, that asymptotically attains the bound of Theorem 5.1, both in number of edges and size of the minimum code, when $\Delta = \text{Poly}(d)$.

Let start by showing that every graph with an asymptotically optimal identifying code cannot contain too few edges.

Lemma 5.4. *For every $M' \geq 0$, there exists a constant $c_0 > 0$ such that any graph G with $\gamma^{ID}(G) \leq M' \log n$ contains at least $c_0 n \log n$ edges.*

Proof. Set α_0 as the smallest positive root of

$$f(\alpha) = \alpha \log\left(\frac{M' + \alpha}{\alpha} e\right) - 1/2 \,. \tag{5.7}$$

Note that $f(\alpha)$ is well–defined since $\lim_{\alpha \to 0} f(\alpha) = -1/2$ and $f(1) = \log(M' + 1) + 1/2 > 0$.

Suppose by contradiction that there exists a graph G containing less than $c_0 n \log n$ edges, with $c_0 = \alpha_0/4$, that admits an identifying code $\mathcal{C}$ of size at most $M' \log n$. Let U be the subset of vertices of degree at least $\alpha_0 \log n$. Notice that

$$|U| \leq \frac{2|E(G)|}{\alpha_0 \log n} \leq \frac{2c_0}{\alpha_0} n = \frac{n}{2} \,.$$

Since $|\mathcal{C}| \leq M' \log n$ and any $v \in V(G) \setminus U$ has degree smaller than $\alpha_0 \log n$, the number of possible nonempty sets $N_G[v] \cap \mathcal{C}$, is smaller than

$$\begin{aligned}
\sum_{i=1}^{\alpha_0 \log n} \binom{|\mathcal{C}|}{i} &\leq \binom{M' \log n + \alpha_0 \log n}{\alpha_0 \log n} \\
&\leq \left(\frac{(M' + \alpha_0)e}{\alpha_0}\right)^{\alpha_0 \log n} \\
&= n^{\alpha_0 \log\left(\frac{M'+\alpha_0}{\alpha_0} e\right)} \\
&= \sqrt{n} \,.
\end{aligned}$$

where we have used that $\binom{a}{b} \leq \left(\frac{ae}{b}\right)^b$ for the second inequality and the fact that α_0 is a root of (5.7) for the last one.

Since $|V(G) \setminus U| \geq n/2$ there must be at least two vertices $v_1, v_2 \in V(G) \setminus U$ such that $N_G[v_1] \cap \mathcal{C} = N_G[v_2] \cap \mathcal{C}$, and thus $\mathcal{C}$ cannot be an identifying code, a contradiction. □

The following lemma relates the identifying code number of a graph G to the one of its complement $\overline{G}$.

Lemma 5.5. *Let G be a twin-free graph. If $\overline{G}$ is twin–free, then*

$$\gamma^{ID}(\overline{G}) \leq 2\gamma^{ID}(G) \,.$$

Proof. Let $\mathcal{C}_0$ be a minimum identifying code of G. We will show that there exists a set $\mathcal{C}_1$ of size at most $\gamma^{\mathrm{ID}}(G)-1$ and a special vertex v, such that $\mathcal{C} = \mathcal{C}_0 \cup \mathcal{C}_1 \cup \{v\}$ is an identifying code of $\overline{G}$.

For the sake of simplicity, we define the following relation. Two vertices $u, v \in V(G)$ are related if and only if $N_G(u) \cap \mathcal{C}_0 = N_G(v) \cap \mathcal{C}_0$ and $u \not\sim v$ (i.e. considering $\mathcal{C}_0$ in G, u, v are separated by one of u, v). This will be denoted as $u \equiv_G v$. It can be checked that this relation is an equivalence relation.

Claim 5.6. *Every pair of distinct vertices $u \not\equiv_G v$ is separated by $\mathcal{C}_0$ in $\overline{G}$.*

Proof. By the definition of $\equiv_G$, either $N_G(u) \cap \mathcal{C}_0 \neq N_G(v) \cap \mathcal{C}_0$ or $u \sim v$.

If $N_G(u) \cap \mathcal{C}_0 \neq N_G(v) \cap \mathcal{C}_0$, there exists $w \in \mathcal{C}_0$ (and $w \notin \{u, v\}$) such that $w \in N_G(u) \oplus N_G(v)$. Then, $w \in N_{\overline{G}}(u) \oplus N_{\overline{G}}(v)$, hence w still separates u, v in $\overline{G}$.

If $N_G(u) \cap \mathcal{C}_0 = N_G(v) \cap \mathcal{C}_0$, then $u \sim v$. If at least one of them belongs to $\mathcal{C}_0$, then this vertex separates u, v in $\overline{G}$. Otherwise, $u, v \notin \mathcal{C}_0$ and we have $N_G(u) \cap \mathcal{C}_0 = N_G[u] \cap \mathcal{C}_0$ and $N_G[v] \cap \mathcal{C}_0 = N_G(v) \cap \mathcal{C}_0$. Hence $N_G[u] \cap \mathcal{C}_0 = N_G[v] \cap \mathcal{C}_0$. But then $\mathcal{C}_0$ does not separate u, v in G, a contradiction. □

In particular, this implies that any vertex in an equivalence class of size one is separated by $\mathcal{C}_0$ from all other vertices in $\overline{G}$.

Claim 5.7. *If $u \equiv_G v$ and both $u, v \notin \mathcal{C}_0$, then $u = v$.*

Proof. Since $u, v \notin \mathcal{C}_0$, $N_G[u] \cap \mathcal{C}_0 = N_G(u) \cap \mathcal{C}_0$ and $N_G[v] \cap \mathcal{C}_0 = N_G(v) \cap \mathcal{C}_0$. Using that they are equivalent, we have that $N_G[u] \cap \mathcal{C}_0 = N_G[v] \cap \mathcal{C}_0$. Since $\mathcal{C}_0$ is an identifying code of G, we must have $u = v$. □

Claim 5.8. *Let $U = \{u_1, \ldots, u_s\}$ be an equivalence class of $\equiv_G$. Then all the pairs in U can be separated in $\overline{G}$ by using $s-1$ vertices.*

Proof. We will prove the claim by induction. For $s = 2$ it is clearly true: since $\overline{G}$ is twin–free, we can select $w \in N_{\overline{G}}[u_1] \oplus N_{\overline{G}}[u_2]$, and w separates u and v in $\overline{G}$.

For every $s > 2$, consider the vertices $u_1, u_2 \in U$ and let $w \in N_{\overline{G}}[u_1] \oplus N_{\overline{G}}[u_2]$. Since U forms a clique in $\overline{G}$, $w \notin U$. Then w splits the set U into U_1, the set of vertices of U adjacent to w in $\overline{G}$, and U_2, the set of vertices in U non–adjacent to w in $\overline{G}$. Let $|U_1| = s_1$ and $|U_2| = s_2$; by construction, $s_1, s_2 < s$.

Now, the pairs of vertices of U with one vertex from U_1 and one vertex from U_2 are separated by w. By induction, the pairs of vertices in U_1 can be separated using $s_1 - 1$ vertices and the ones in U_2 using $s_2 - 1$. Thus we need at most $(s_1 - 1) + (s_2 - 1) + 1 = s - 1$ vertices to separate all the pairs of vertices in U. □

From the previous claims, it is straightforward to deduce that there is a set $\mathcal{C}_1$ of size at most $|\mathcal{C}_0| - 1$ vertices that separates all the pairs in $\overline{G}$ that are not separated by $\mathcal{C}_0$.

Eventually, there might be a unique vertex v such that $N_{\overline{G}}[v] \cap (\mathcal{C}_0 \cup \mathcal{C}_1) = \emptyset$ (if there were two such vertices, they would not be separated by $\mathcal{C}_0 \cup \mathcal{C}_1$, a contradiction). Hence, $\mathcal{C} = \mathcal{C}_0 \cup \mathcal{C}_1 \cup \{v\}$ is an identifying code of $\overline{G}$ of size at most $2|\mathcal{C}_0| = 2\gamma^{\mathrm{ID}}(G)$. □

Proposition 5.9. *For every $M \geq 0$, there exists a constant $c > 0$ such that for every set of edges $F \subset E(K_n)$ satisfying $\gamma^{ID}(K_n \setminus F) \leq M \log n$, $|F| \geq cn \log n$.*

Proof. Set $M' = M/2$ and let $c = c_0$ be the constant given by Lemma 5.4 for this M'. Suppose that there exists a set F of edges, $|F| < cn \log n$ such that $G = K_n \setminus F$ satisfies $\gamma^{\mathrm{ID}}(G) \leq M \log n$. By Lemma 5.5, the graph $\overline{G}$ admits an identifying code of size at most $2M \log n = M' \log n$. By Lemma 5.4, we get a contradiction. □

If $\Delta = \mathrm{Poly}(d)$, the former proposition provides an example of a graph for which the result of Theorem 5.1 is asymptotically tight.

For every $d > 0$, consider the graph H_d to be the disjoint union of cliques of order $d+1$. We may assume that $d+1$ divides n for the sake of simplicity. Denote by $H_d^{(1)}, \ldots, H_d^{(s)}$, $s = \frac{n}{d+1}$, the cliques composing H_d.

Since $H_d^{(i)}$ is a connected component, an asymptotically optimal identifying code for H_d must also be asymptotically optimal for each $H_d^{(i)}$. By Proposition 5.9, we must delete at least $\Omega(d \log d)$ edges from $H_d^{(i)}$ to get an identifying code of size $O(\log d)$.

Thus, one must delete at least $\Omega(sd \log d) = \Omega(n \log d)$ edges from H_d to get an optimal identifying code.

Corollary 5.10. *For every $d = \omega(1)$ and every $M \geq 0$, there exists a constant $c > 0$ such that for every set of edges $F \subset E(H_d)$ satisfying $\gamma^{ID}(H_d \setminus F) \leq M \frac{n \log d}{d}$, we have $|F| \geq cn \log d$.*

We remark that a connected counterexample can also be constructed from H_d by connecting its cliques using few edges, without affecting the above result.

Corollary 5.10 implies that Theorem 5.1 is asymptotically tight when $\Delta = \mathrm{Poly}(d)$, since in that case $\log \Delta = O(\log d)$. In the case where d is sub–polynomial with respect to Δ, there is a gap between the result of Theorem 5.1 and the construction provided here.

5.3.2 On the hypothesis

We conclude this section by discussing the necessity of the conditions $\Delta = \omega(1)$ and $d \geq 66 \log \Delta$ in Theorem 5.1.

First note that, if Δ is bounded by a constant, we need at least $\frac{n}{\Delta+1} = \Theta(n)$ vertices to dominate G. Thus, no code of size smaller than $\Theta(n)$ can be obtained by deleting edges of the graph.

On the other hand, the condition $d \geq 66 \log \Delta$ in Theorem 5.1, is also necessary (up to a constant factor) as can be deduced from the following proposition.

Proposition 5.11. *For arbitrarily large values of* Δ*, there exists a graph* G *with maximum degree* Δ *and minimum degree* $d = \frac{\log_2 \Delta}{2}$ *such that, for every spanning subgraph* $H \subseteq G$*,*

$$\gamma^{ID}(H) = (1 - o(1))n \ .$$

Proof. Consider the bipartite complete graph $G = K_{r,s}$ where $s = 2^{2r}$. Denote by V_1 the stable set of size r and by V_2 the stable set of size s. Observe that $d = r = \frac{\log_2 s}{2} = \frac{\log_2 \Delta}{2}$.

For every twin–free spanning subgraph $H \subseteq G$, let $\mathcal{C} \subseteq V(G)$ be an identifying code of H. Let us show that most of the vertices in V_2 must be in $\mathcal{C}$. Let $S \subseteq V_2$ be the subset of vertices in V_2 that are not in the code. Thus, for every $u \in S$, $N_{\mathcal{C}}[u] = N_{\mathcal{C}}(u)$. Observe that $N_{\mathcal{C}}(u) \subseteq V_1$, and hence, there are at most 2^r possible candidates for such $N_{\mathcal{C}}(u)$. Since $\mathcal{C}$ is dominating and separating all the pairs in S, all the subsets $N_{\mathcal{C}}(u)$ must be nonempty and different, which implies, $|S| < 2^r$. Hence, we have

$$|\mathcal{C}| \geq |V_2 \setminus S| \geq 2^{2r} - 2^r = (1 - o(1))2^{2r} = (1 - o(1))n \ .$$

□

5.4 Consequences of our results

We now describe consequences of our results on the case when we want to *add* edges to a graph to decrease its identifying code number, and to the notion of watching systems.

5.4.1 Adding edges

In the previous sections, we have studied how much can the identifying code number decrease when we delete few edges from the original graph. In this section, we discuss the symmetric question of how much can the addition of edges help to decrease this parameter.

The question of how much can a parameter decrease when deleting/adding edges has been already studied for some monotone parameters. However, if the parameter is monotone, only one of either deleting or adding, can help to decrease it. One of the interesting facts of studying the identifying code number is that, since it is a non–monotone parameter, we can have similar results for both procedures.

As before, let G be a graph with maximum degree Δ and minimum degree d. We aim to find a set of edges F with $F \cap E(G) = \emptyset$ such that $\gamma^{\text{ID}}(G \cup F)$ is small. This set F will be provided by applying Theorem 5.1 to the graph $\overline{G}$, that has maximum degree $\Delta(\overline{G}) = n - 1 - d$ and minimum degree $d(\overline{G}) = n - 1 - \Delta$. Thus, it will have size

$$|F| = O\left(n \log \Delta(\overline{G})\right) ,$$

and

$$\gamma^{\text{ID}}(\overline{G} \setminus F) = O\left(\frac{n \log \Delta(\overline{G})}{d(\overline{G})}\right) \ .$$

Since $\overline{G} \setminus F = \overline{G \cup F}$, we have the following corollary of Theorem 5.1 and Lemma 5.5.

Corollary 5.12. *For every graph G on n vertices with minimum degree $d = n - \omega(1)$ and maximum degree Δ such that $n - \Delta \geq 66\log(n-d)$, there exists a set of edges F with $F \cap E(G) = \emptyset$ of size*

$$|F| = O\left(n \log(n-d)\right) \ ,$$

such that

$$\gamma^{ID}(G \cup F) = O\left(\frac{n\log(n-d)}{n-\Delta}\right) \ .$$

This result is also asymptotically tight. Otherwise, by using again Lemma 5.5, we could translate our case to the case of deleting edges and we would get a contradiction with the optimality of Theorem 5.1.

5.4.2 Watching systems

The result of Theorem 5.1 has a direct application for *watching systems*, which are a generalization of identifying codes [12, 11]. In a watching system, we can place on each vertex v a set of *watchers*. To each watcher w placed on v, we assign a nonempty subset $Z(w) \subseteq N[v]$, its *watching zone*. We now ask each vertex to belong to a unique and nonempty set of watching zones; the minimum number of watchers that need to be placed on the vertices of G to obtain a watching system is the *watching number* $w(G)$ of G.

It is clear from the definition that $\gamma(G) \leq w(G) \leq \gamma^{\mathrm{ID}}(G)$, since the vertices of any identifying code form a watching system (where the watching zones are the closed neighborhoods). In fact, even the following holds:

Observation 5.13. *For every twin–free graph G,*

$$w(G) \leq \min\{\gamma^{ID}(H), \ \textit{where } H \textit{ is a spanning subgraph of } G\} \ .$$

Indeed, consider the spanning subgraph H_0 of G with smallest identifying code number, and define the watching system to be the vertices of an optimal identifying code of H_0, with the watching zones being the closed neighborhoods in H_0.

In [12, Theorems 2 and 3], the authors propose the following upper bound for graphs with given maximum degree:

Theorem 5.14 ([12])**.** *Let G be a graph with maximum degree Δ, then*

$$\lceil \log_2(n+1) \rceil \leq w(G) \leq \gamma(G)\lceil \log_2(\Delta+2) \rceil \ .$$

Note that for every values of parameters γ and Δ, the upper bound from the above theorem is sharp for the graph consisting of γ disjoint copies of a star on $\Delta + 1$ vertices.

Recall that the bound on the dominating number provided in (5.6) is tight. In particular, a d–regular graph chosen uniformly at random is an asymptotically tight example with high probability. Indeed, for almost all d–regular graphs G, the upper bound of Theorem 5.14 gives

$$w(G) \leq \gamma(G)\lceil \log \Delta + 2 \rceil = \Omega\left(\frac{n \log^2 d}{d}\right) \ . \tag{5.8}$$

By Observation 5.13, a direct corollary of Theorem 5.1 is the following:

Corollary 5.15. *For every graph G on n vertices with minimum degree $d \geq 66 \log \Delta$ and maximum degree $\Delta = \omega(1)$, we have:*

$$w(G) \leq O\left(\frac{n \log \Delta}{d}\right) .$$

On the one hand, since $\gamma(G) \geq \frac{n}{d+1}$, Corollary 5.15 is always at least as good as Theorem 5.14. On the other hand, if $\gamma(G)$ is large, it asymptotically improves Theorem 5.14, as in case of a typical d-regular graph (see the bound in (5.8)).

5.5 Concluding remarks and open questions

1. The kind of results we provide in this chapter can be connected to the notion of resilience. Given a graph property $\mathcal{P}$, the *global resilience* of G with respect to $\mathcal{P}$ is the minimum number of edges one has to delete to obtain a graph not satisfying $\mathcal{P}$. The resilience of monotone properties is well studied, in particular, in the context of random graphs [122].

Our result can be interpreted in terms of the resilience of the following (non–monotone) property $\mathcal{P}$: "G has a large identifying code number in terms of its degree parameters, d and Δ". For every graph G satisfying the hypothesis $\Delta = \omega(1)$ and $d \geq 66 \log \Delta$, Theorem 5.1 can be stated as: the resilience of G with respect to $\mathcal{P}$ is $O(n \log \Delta)$. Moreover, Corollary 5.10 shows that there are graphs that attain this value of the resilience.

2. In Theorem 5.1, we show the existence of a small identifying code for a large spanning subgraph of G. However, our proof is not constructive and, besides, the probability that such pair exists is exponentially small, due to the use of the local lemma. The algorithmic version of the local lemma proposed by Moser and Tardos [109], allows to explicitly find a configuration that avoids all the bad events A_{uv}, when these events are determined by a finite set of mutually independent random variables. Unfortunately, this is not the case here, since A_{uv} depends on the random variables determining the existence of certain edges close to uv. These random variables are not independent because of the definition of p_{uv}.

On the other hand, if we do not want to argue in terms of the maximum degree Δ, one can show that by deleting a set of $O(n \log n)$ random edges, any set of size $O\left(\frac{n \log n}{d}\right)$ is an identifying code with probability $1 - o(1)$. In such a case, the proof provides a randomized algorithm which constructs the desired code for almost all subgraphs. It is an open question whether this algorithm can be derandomized.

3. A notion similar to identifying codes, *locating-dominating sets*, has also been extensively studied in the literature (see e.g. [96] for many references). A set $\mathcal{C}$ of vertices of G is a locating-dominating set if $\mathcal{C}$ is a dominating set which separates all pairs of vertices in $V(G) \setminus \mathcal{C}$. It follows that any identifying code is a locating-dominating set, hence Theorem 5.1 also holds for this notion. In fact, the proof of Corollary 5.10 can be adapted for this case too.

4. As further research, it would be very interesting to close the gap between the result in

Theorem 5.1 and the lower bound given by the example in Corollary 5.10. Motivated by this example, we ask the following question:

Question 5.16. *Is it true that for every graph G with minimum degree d, there exists a subset of edges $F \subset E(G)$ of size*

$$|F| = O\left(n \log d\right) ,$$

such that

$$\gamma^{ID}(G \setminus F) = O\left(\frac{n \log d}{d}\right) ?$$

It seems to us that the techniques used in this chapter will not provide an answer to the previous question. The main obstacle is the use of the local lemma, which forces us to take into account the role of the maximum degree of G.

CHAPTER 6

Consecutive pattern avoiding in permutations

6.1 Introduction

A permutation $\pi \in \mathcal{S}_n$ *contains* $\sigma \in \mathcal{S}_m$ *as a consecutive pattern* if there exists $0 \leq i \leq n-m$ such that $\mathrm{st}(\pi_{i+1}, \dots, \pi_{i+m}) = \sigma$, that is, there are m consecutive elements in π that have the relative order prescribed by σ. For instance, if $\sigma = (1, 2, \dots, m)$, then π contains σ as a consecutive pattern if and only if it contains m consecutive increasing elements (a run of length m). A permutation $\pi \in \mathcal{S}_n$ is called *consecutive σ–avoiding* if it does not contain σ as a consecutive pattern. We denote by $\alpha_n(\sigma)$ the number of permutations in $\mathcal{S}_n$ that are σ–avoiding.

The problem of determining $\alpha_n(\sigma)$ is inspired by the problem of finding the number of permutations of length n that avoid a pattern σ non necessarily in consecutive positions. A permutation $\pi \in \mathcal{S}_n$ *contains* σ if there exist $1 \leq i_1 < \dots < i_m \leq n$ such that $\mathrm{st}(\pi_{i_1}, \dots, \pi_{i_m}) = \sigma$. Clearly, if π avoids σ, then π also avoids σ as a consecutive pattern. Knuth [90] introduced the non–consecutive case and exactly determined the number of permutations avoiding some pattern of length 3. There are many interesting results in the area (see e.g. [25, 5]) as well as the famous Stanley–Wilf conjecture which was proved by Marcus and Tardos [99].

For every $\sigma \in S_m$, it is hard to provide an exact formula for $\alpha_n(\sigma)$ when n is large. Asymptotic formulas can be derived for some special patterns as showed by Elizalde and Noy [50]. In particular, the authors give an estimation of $\alpha_n(\sigma)$ for every pattern σ of length 3 and also for some patterns of length 4. However, even for the case of length 4, there are still some patterns for which the asymptotic behavior of $\alpha_n(\sigma)$ is not known.

Elizalde [48] showed that for every $\sigma \in \mathcal{S}_m$, the following limit

$$\rho_\sigma := \lim_{n\to\infty} \left(\frac{\alpha_n(\sigma)}{n!} \right)^{1/n} ,$$

exists and that $0.7839 < \rho_\sigma < 1$ if $m \geq 3$. A stronger result is given in [47], where the authors show that $\alpha_n(\sigma) \sim c_\sigma\, \rho_\sigma^n\, n!$, for some constant c_σ that only depends only on σ.

A pattern of length m is called *monotone* if it is either $(1, 2, \dots, m)$ or $(m, \dots, 2, 1)$. It is clear that $\alpha_n(1, 2, \dots, m) = \alpha_n(m, \dots, 2, 1)$, since $\pi = (\pi_1, \dots, \pi_n)$ is $(1, 2, \dots m)$–avoiding if and only if its reversing $(\pi_n, \dots, \pi_1)$ is $(m, \dots, 2, 1)$–avoiding. It was conjectured in [50] that monotone patterns are the most avoided ones among all patterns of length m, when n is large enough. This is known as the *Consecutive Monotone Pattern (CMP) Conjecture*.

Conjecture 6.1 (CMP conjecture [50]). *For every $\sigma \in \mathcal{S}_m$,*

$$\rho_\sigma \le \rho_{(1,2,\dots,m)} \ .$$

The results in [50], determining ρ_σ for every $\sigma \in \mathcal{S}_3$, settle in the affirmative the CMP conjecture for patterns of length 3. Elizalde and Noy [51] show that the conjecture is true for the large class of non–overlapping patterns. A pattern is non–overlapping if two copies of the pattern in a permutation share at most one position.

Regarding the least avoided pattern among all the patterns of length m, Nakamura [111] posed the following conjecture motivated by some simulations for small values of n and m.

Conjecture 6.2 ([111]). *For every $\sigma \in \mathcal{S}_m$,*

$$\rho_\sigma \ge \rho_{(1,2,\dots,m-2,m,m-1)} \ .$$

Both conjectures have been recently proved by Elizalde [49]. The proofs are based on computing the generating function for the number of σ–avoiding permutations, $P_\sigma(z) = \sum \alpha_n(\sigma) \frac{z^n}{n!}$, combined with the cluster method of Goulden and Jackson [71].

Here we will use a completely different approach to the consecutive pattern avoiding problem through the probabilistic method. While this approach is not as precise as the generating function technique, it provides simpler alternative proofs of some known results, as the CMP conjecture, and allows one to obtain more general results.

All along this chapter, m will be a fixed integer while n will be considered to tend to infinity. For the sake of simplicity, however, we will use asymptotic notation on m. If this is the case, we will consider m to tend to infinity while n will be an arbitrarily large function of m.

Since we are interested in ρ_σ, we will consider m to be a fixed integer and n to be arbitrarily large with respect to n. Our first result bounds ρ_σ from above when the pattern σ is not monotone.

Theorem 6.3. *For every $\sigma \in \mathcal{S}_m \setminus \{(1, 2, \dots, m), (m, \dots, 2, 1)\}$,*

$$\rho_\sigma \le 1 - \frac{1}{m!} + O\left(\frac{1}{m^2 \cdot m!}\right) \ .$$

To prove this theorem we make use of Suen Inequality (see Theorem 2.16), a powerful tool that provides an upper bound on the probability that none of the events of a certain collection happen simultaneously.

By comparing the upper bound given by Theorem 6.3 with the result obtained by Elizalde and Noy [51] for $\rho_{(1,2,\dots,m)}$ we get the following corollary,

Corollary 6.4. *There exists an integer m_0, such that for any $m \geq m_0$ and every pattern $\sigma \in \mathcal{S}_m \setminus \{(1,2,\ldots,m),(m,\ldots,2,1)\}$,*

$$1-\rho_\sigma \geq \left(1+\Omega\left(\frac{1}{m}\right)\right)\left(1-\rho_{(1,2,\ldots,m)}\right).$$

This immediately provides an alternative probabilistic proof for the CMP conjecture for large values of m. Analyzing more carefully the proof of Theorem 6.3 for small values of m, we can show that the CMP conjecture holds for $m \geq 5$. The same approach, however, does not provide meaningful results for the case $m = 4$.

This corollary also gives a lower estimation of the minimum gap between $\rho_{(1,2,\ldots,m)}$ and ρ_σ, for every non monotone $\sigma \in \mathcal{S}_m$.

Theorem 6.3 can be extended to the whole set of patterns, $\mathcal{S}_m$, by weakening the upper bound; for every $\sigma \in \mathcal{S}_m$,

$$\rho_\sigma \leq 1-\frac{1}{m!}+O\left(\frac{1}{m\cdot m!}\right).$$

The second part of the chapter is devoted to the proof of a general lower bound on ρ_σ for $\sigma \in \mathcal{S}_m$.

Theorem 6.5. *For every $\sigma \in \mathcal{S}_m$,*

$$\rho_\sigma \geq 1-\frac{1}{m!}-O\left(\frac{m}{(m!)^2}\right).$$

To prove this lower bound we use the one–sided version of the Lovász Local Lemma (see Lemma 2.14). This bound is asymptotically tight and an extremal example is provided by the pattern $(1,2,\ldots,m-2,m,m-1)$, the least avoided pattern of length m. Unlike in the case of the upper bound and the CMP conjecture, our proof of Theorem 6.5 can not be adapted to extract a proof of Conjecture 6.2.

As Theorem 6.3 and Theorem 6.5 give bounds for the value of ρ_σ in terms of m, a natural question is to determine how most of the patterns behave. In this direction a much stronger upper bound, close to the general lower bound, is showed to hold for most of the patterns.

Theorem 6.6. *Let $\sigma \in \mathcal{S}_m$ be chosen uniformly at random. Then, for each $2 \leq k \leq m/2$,*

$$\rho_\sigma \leq 1-\frac{1}{m!}+O\left(\frac{4^m}{(m-k)!m!}\right),$$

with probability at least $1-\frac{2}{(k+1)!}-m\frac{2^m}{(m/2)!}$.

This theorem shows that when m large enough, for most of the patterns the value of ρ_σ is concentrated close to the lower bound provided by Theorem 6.5. The idea behind the proof of this result is that the number of permutations avoiding a pattern depends on the maximum overlapping position of this pattern. It can be showed that almost all patterns do not have a large overlap and thus, they are far from the upper bound attained by monotone patterns, the ones with maximum overlap.

This chapter is organized as follows. The upper bound on ρ_σ is studied in Section 6.3, where Theorem 6.3 and Corollary 6.4 are proved. Section 6.4 is devoted to the proof of Theorem 6.5. Finally, in Section 6.5 we provide an upper bound for most of the patterns σ by proving Theorem 6.6. We conclude with some remarks and open questions in Section 6.6.

6.2 A probabilistic approach on consecutive pattern avoiding

Our goal in this section is to give a proper set of "bad events" $\mathcal{A}$ for our problem and build a good dependency graph for $\mathcal{A}$. The probability that a random permutation avoids all the events in $\mathcal{A}$ will be the probability that a random permutation avoids $\sigma \in \mathcal{S}_m$ as a consecutive pattern.

Let $\pi \in \mathcal{S}_n$ be chosen uniformly at random, and let $\sigma \in \mathcal{S}_m$ be a fixed pattern. We consider the set of events $\mathcal{A} = \{A_0, \dots, A_{n-m}\}$ where $A_i := \{\mathrm{st}(\pi_{i+1}, \dots, \pi_{i+m}) = \sigma\}$. As before, we let X_i be the indicator random variable of the event A_i and let $X = \sum_{i=0}^{n-m} X_i$ denote the number of events in $\mathcal{A}$ which are realized. Then, π avoids σ as a consecutive pattern if and only if $X = 0$, that is, no copy of the pattern σ appears. We have,

$$\alpha_n(\sigma) = \Pr(X = 0)n!\ ,$$

where the dependency of X on σ will be clear from the context. In particular we will be interested in

$$\rho_\sigma = \lim_{n\to\infty} \Pr(X = 0)^{1/n}\ . \tag{6.1}$$

Bounding from above the number of edges in a dependency graph H is crucial in order to give a proper upper bound on the probability that no event in $\mathcal{A}$ holds. The following lemma shows that we can choose a dependency graph H with few edges.

Lemma 6.7. *Let $S, T \subseteq \{0, 1, \dots, n-m\}$ be two disjoint subsets such that for each $(i, j) \in S \times T$, we have $|i - j| \geq m$. Then, the set of events $\{A_i\}_{i\in S}$ and $\{A_j\}_{j\in T}$ are mutually independent.*

Proof. In order to prove that $\{A_i\}_{i\in S}$ and $\{A_j\}_{j\in T}$ are mutually independent, we use the random permutation model defined in Section 2.3. Recall that a uniform permutation from $\mathcal{S}_n$ can be obtained by considering $Z_1, \dots, Z_n$ independent uniform random variables in $(0, 1)$ and by choosing $\pi = \mathrm{st}(Z_1, \dots, Z_n)$.

Observe that the event A_i depends only on the random variables $Z_{i+1}, \dots, Z_{i+m}$. By the hypothesis of the lemma, we have that for each $i \in S$ and $j \in T$, $|i - j| \geq m$. Then, using the Mutual Independence Principle (see Observation 2.7) with $\mathcal{F}_i = \{i + 1, \dots, i + m\}$, and noting that

$$(\cup_{i\in S}\mathcal{F}_i) \cap (\cup_{j\in T}\mathcal{F}_j) = \emptyset\ ,$$

we have that the set of events $\{A_i\}_{i\in S}$ and $\{A_j\}_{j\in T}$ are mutually independent. □

According to the previous lemma, the graph H with vertex set $V(H) = \{0, 1, \dots, n-m\}$, where $ij \in E(H)$ if and only if $0 < |i - j| < m$, is a dependency graph for $\mathcal{A}$. Throughout the chapter, we will use this graph H as a dependency graph of $\mathcal{A}$.

6.3 An upper bound on ρ_σ and the CMP conjecture.

In this section we show how Suen inequality (see Theorem 2.16) can be used to provide a meaningful upper bound on ρ_σ. Then, we derive an explicit lower bound for $\rho_{(1,2,\dots,m)}$ using a result of Elizalde and Noy [51]. For large values of m, a proof of the CMP conjecture follows from these two previous results. In the last part of the section we prove the conjecture for small values of m.

A simple upper bound follows directly from the construction of the dependency graph H in the previous section. Consider $I = \{km : 0 \le k < n/m\}$, then

$$\Pr(X=0) = \Pr\left(\bigcap_{i=0}^{n-m} \overline{A_i}\right) \le \Pr\left(\bigcap_{i\in I} \overline{A_i}\right) = \prod_{i\in I}\left(1 - \Pr\left(A_i \mid \bigcap_{j\in I, j<i} \overline{A_j}\right)\right) .$$

By using Lemma 6.7 with $S = \{i\}$ and $T = \{j : j \in I, j < i\}$,

$$1 - \Pr\left(A_i \mid \bigcap_{j\in I, j<i} \overline{A_j}\right) = 1 - \Pr(A_i) = 1 - \frac{1}{m!} .$$

Since $|I| \ge n/m - 1$, this implies

$$\rho_\sigma \le \left(1 - \frac{1}{m!}\right)^{1/m} = 1 - O\left(\frac{1}{m \cdot m!}\right) .$$

A better upper bound is given in Theorem 6.3 by taking into account the interaction between pairs of dependent events.

A pattern $\sigma \in \mathcal{S}_m$ has an *overlap at* k, $1 \le k \le m-1$, if $\mathrm{st}(\sigma_1, \dots, \sigma_k) = \mathrm{st}(\sigma_{m-k+1}, \dots, \sigma_m)$, namely, the first and the last k positions have the same relative order. For instance, the permutation $(2,5,8,7,1,3,6,4)$ has an overlap at 4, since $\mathrm{st}(2,5,8,7) = \mathrm{st}(1,3,6,4) = (1,2,4,3)$, at 1, since $\mathrm{st}(2) = \mathrm{st}(4) = (1)$, and does not have an overlap at any other position. Observe that a pattern does not have an overlap at k, if and only if

$$\Pr(A_i \cap A_{i+m-k}) = 0 . \tag{6.2}$$

For every $\sigma \in \mathcal{S}_m$, define the set

$$\mathcal{O}_\sigma = \{k : \Pr(A_i) \cap \Pr(A_{i+m-k}) \neq 0,\ 1 \le k \le m-1\}.$$

Notice that $\mathcal{O}_\sigma$ is the set of positions at which σ has an overlap. For instance, the monotone pattern $(1,2,\dots,m)$ has $\mathcal{O}_{(1,2,\dots,m)} = \{1,2,\dots,m-1\}$ and for the pattern $\sigma = (2,5,8,7,1,3,6,4)$, $\mathcal{O}_\sigma = \{1,4\}$.

The following lemma is one of the crucial facts to prove Theorem 6.3.

Lemma 6.8. *Let $m \ge 2$ and $\sigma \in \mathcal{S}_m$. Then $m-1 \in \mathcal{O}_\sigma$ if and only if σ is a monotone pattern.*

Proof. It is clear that both monotone patterns satisfy $m-1 \in \mathcal{O}_\sigma$. Let us show that for every other pattern, $m-1 \notin \mathcal{O}_\sigma$.

Suppose that $m-1 \in \mathcal{O}_\sigma$. This implies that

$$\mathrm{st}(\sigma_1, \dots, \sigma_{m-1}) = \mathrm{st}(\sigma_2, \dots, \sigma_m) \; . \tag{6.3}$$

Since σ is not a monotone pattern, there exists an index $2 \le i \le m-1$ such that $\sigma_{i-1} > \sigma_i < \sigma_{i+1}$ or $\sigma_{i-1} < \sigma_i > \sigma_{i+1}$. Without loss of generality we assume the latter. Now observe that (6.3) implies that if $\sigma_{i-1} < \sigma_i$, then $\sigma_i < \sigma_{i+1}$, leading a contradiction. □

Thus, we can consider that the maximum overlap of $\sigma \in \mathcal{S}_m \setminus \{(1,2,\dots m),(m,\dots,2,1)\}$ is at most at $m-2$. We observe that there are non monotone patterns that have an overlap at $m-2$. For instance, consider $m = 2t$ and $\sigma = (1, t+1, 2, t+2, \dots, t, 2t)$, or $(2,1,4,3,\dots,m,m-1)$.

The following lemma gives some insight on the structure of the permutations that contain two close occurrences of a pattern σ.

Lemma 6.9. *Let $\sigma \in \mathcal{S}_m$ with $k \in \mathcal{O}_\sigma$ and suppose that $\tau \in \mathcal{S}_{2m-k}$ is such that the events A_0 and A_{m-k} hold. If $\sigma' = \mathrm{st}(\sigma_{m-k+1}, \dots, \sigma_m)$, then, for each $i = 0, 1, \dots, k-1$, we have $\tau_{m-i} = \sigma_{k-i} + \sigma_{m-i} - \sigma'_{k-i}$.*

Proof. Fix some $i < k$. By the event A_0, we know that τ_{m-i} must be larger than $\sigma_{m-i} - 1$ elements and smaller than $m - \sigma_{m-i}$ elements from $(\tau_1, \dots, \tau_{m-i-1}, \tau_{m-i+1}, \dots, \tau_m)$. By the event A_{m-k}, it is also true that τ_{m-i} is larger than $\sigma_{k-i} - 1$ and smaller than $m - \sigma_{k-i}$ elements from $(\tau_{m-k+1}, \dots, \tau_{m-i-1}, \tau_{m-i+1}, \dots, \tau_{2m-k})$.

Consider now the permutation $\sigma' = \mathrm{st}(\sigma_{m-k+1}, \dots, \sigma_m) \in \mathcal{S}_k$. Then there are $\sigma'_{k-i} - 1$ elements that are counted twice when we look at the elements smaller than σ_{m-i} or σ_{k-i}, and $k - \sigma'_{k-i}$ also double counted when we look to the larger ones. Therefore

$$\tau_{m-i} > \sigma_{k-i} + \sigma_{m-i} - 2 - (\sigma'_{k-i} - 1) \; ,$$

and

$$\tau_{m-i} \le 2m - k - \big(m - \sigma_{k-i} + m - \sigma_{m-i} - (k - \sigma'_{k-i})\big) \; .$$

Observing that the first inequality is strict, we get

$$\tau_{m-i} = \sigma_{i+1} + \sigma_{m-i} - \sigma'_{k-i}.$$

□

By using this last lemma, we can provide an upper bound on the probability that two given occurrences of a pattern appear.

Lemma 6.10. *For every $\sigma \in \mathcal{S}_m$ and any $k \in \mathcal{O}_\sigma$,*

$$\Pr(A_i \cap A_{i+m-k}) \le \frac{\binom{2(m-k)}{m-k}}{(2m-k)!} \; .$$

Proof. Set $\tau = \mathrm{st}(\pi_{i+1}, \dots, \pi_{i+2m-k})$. Recall that $\pi \in \mathcal{S}_n$ has been chosen uniformly at random, which implies that τ is also uniformly distributed in $\mathcal{S}_{2m-k}$. Moreover, π satisfies A_i and A_{i+m-k} if and only if τ satisfies A_0 and A_{m-k}.

There are $(2m-k)!$ possible candidates for τ. We will count how many of them are such that the events A_0 and A_{m-k} hold. By Lemma 6.9, we know that the elements $\{\tau_{m-k+1},\dots,\tau_m\}$ are uniquely determined by σ and k. Thus, we select a subset of $m-k$ elements among the $2m-2k$ available ones in order to build $(\tau_1,\dots,\tau_{m-k})$. Since τ satisfies A_0, once these elements have been chosen, there is just one order such that $\mathrm{st}(\tau_1,\dots,\tau_m)=\sigma$, and only one way to set the last $m-k$ elements of τ, in order to satisfy A_{m-k}.

Since $\tau=\mathrm{st}(\pi_{i+1},\dots,\pi_{i+2m-k})$, for every permutation π chosen uniformly at random in $\mathcal{S}_n$,

$$\Pr(\pi \text{ satisfies } A_i\cap A_{i+m-k}) = \Pr(\tau \text{ satisfies } A_0\cap A_{m-k}) \le \frac{\binom{2(m-k)}{m-k}}{(2m-k)!}\ .$$

□

Now we are able to prove Theorem 6.3.

Proof of Theorem 6.3. First of all we compute μ, Δ^* and δ^*, needed to apply Suen inequality. The expected number of occurrences of the pattern σ does not depend on σ and can be computed as

$$\mu=\sum_{i=0}^{n-m}\Pr(A_i)=\frac{n-m+1}{m!}\le\frac{n}{m!}\ .$$

Assume that $i<j$ and $j-i=m-k$. Recall that by the choice of our dependency graph H, two events A_i and A_j are not adjacent if $i-j\ge m$.

By Lemma 6.7, Lemma 6.10 and (6.2), Δ^* can be expressed as

$$\Delta^*=\sum_{ij\in E(H)}\Pr(A_i\cap A_j)=\sum_{i=0}^{n-m}\sum_{k=\max\{1,2m+i-n\}}^{m-1}\Pr(A_i\cap A_{i+m-k})\le n\sum_{k\in\mathcal{O}_\sigma}\frac{\binom{2(m-k)}{m-k}}{(2m-k)!}\ ,\qquad(6.4)$$

where we assume that

Since σ is not monotone, by Lemma 6.8 we have that $m-1\notin\mathcal{O}_\sigma$. Thus, by using that $\binom{2a}{a}\le\frac{4^a}{\sqrt{\pi a}}$, we have

$$\begin{aligned}\Delta^* &\le n\sum_{k=1}^{m-2}\frac{\binom{2(m-k)}{m-k}}{(2m-k)!}\\ &\le n\sum_{k=1}^{m-2}\frac{4^{m-k}}{\sqrt{\pi(m-k)}\cdot(2m-k)!}\\ &\le n\sum_{k=1}^{m-2}\frac{4^{m-k}}{\sqrt{2\pi}(2m-k)!}\\ &=\left(1+\frac{4}{m+3}+O(m^{-2})\right)\frac{16n}{\sqrt{2\pi}(m+2)!}\\ &\le\frac{17n}{\sqrt{2\pi}(m+2)!}\ ,\end{aligned}\qquad(6.5)$$

for any m large enough.

Observe that the degree of a vertex in the dependency graph H is at most $2(m-1)$. Then,

$$\delta^* = \max_{0\le i\le n-m} \sum_{j\colon ij\in E(H)} \Pr(A_j) = 2(m-1)\Pr(A_j) = \frac{2(m-1)}{m!} \le \frac{2}{(m-1)!} \ .$$

Since $e^{2\delta^*} \le e^{4/(m-1)!} \le 2$ if $m \ge 4$ and by using that $e^{-a} \le 1 - \frac{a}{1+a}$, for any $a \ge -1$; both (2.22) and (6.1) imply that

$$\begin{aligned}
\rho_\sigma &\le \exp\left(- \frac{1 - \frac{34}{\sqrt{2\pi}(m+2)(m+1)}}{m!} \right) \\
&\le 1 - \frac{\frac{1}{m!} - \frac{34}{\sqrt{2\pi}(m+2)(m+1)\,m!}}{1 + \frac{1}{m!}} \\
&\le 1 - \left(1 - O\left(\frac{1}{m!}\right)\right)\left(\frac{1}{m!} - \frac{34}{\sqrt{2\pi}(m+2)(m+1)\,m!}\right) \\
&\le 1 - \frac{1}{m!} + \frac{14}{m^2\cdot m!} \ .
\end{aligned}$$

for any large enough m. This completes the proof. □

6.3.1 A lower bound on $\rho_{(1,2,\ldots,m)}$

Next, we proceed to prove Corollary 6.4. This is achieved by obtaining a lower bound on $\rho_{(1,2,\ldots,m)}$ and by showing that this bound is larger than the upper bound given in Theorem 6.3. A recent result of Elizalde and Noy gives an implicit expression for $\rho_{(1,2,\ldots,m)}$.

Theorem 6.11 (Elizalde and Noy [51]). *Let $z_0 = \rho^{-1}_{(1,2,\ldots,m)}$. Then z_0 is the smallest real root of*

$$g(z) = \sum_{i\ge 0} \frac{z^{mi}}{(mi)!} - \sum_{i\ge 0} \frac{z^{mi+1}}{(mi+1)!} \ .$$

From this last theorem we can extract an explicit lower bound on $\rho_{(1,2,\ldots,m)}$.

Lemma 6.12. *For any m large enough,*

$$\rho_{(1,2,\ldots,m)} \ge 1 - \frac{1}{m!} + \frac{1}{m\cdot m!} + O\left(\frac{1}{m^2\cdot m!}\right) \ .$$

Proof. Observe that for nonnegative values of z

$$f(z) = 1 - z + \frac{z^m}{m!} - \frac{z^{m+1}}{(m+1)!} + \frac{z^{2m}}{(2m)!} \ge g(z) \ ,$$

since, for $z \in \mathbb{R}^+$, $g(z)$ can be written as an alternating sum whose terms are strictly decreasing. Since $g(0) = 1$ and z_0 is the smallest real root of $g(z)$, we can conclude that the smallest real root of $f(z)$, z_1, satisfies $z_1 \ge z_0$. Thus $\rho_{(1,2,\ldots,m)} \ge 1/z_1$ and it suffices to compute an upper bound on z_1.

Observe that we can consider $z > 1$. Write $z = (1-\varepsilon)^{-1}$, with $0 < \varepsilon < 1$. Then $z^{-2m}f(z) = 0$ becomes

$$-(1-\varepsilon)^{2m-1}\varepsilon + \frac{(1-\varepsilon)^{m-1}}{(m+1)!}(m-(m+1)\varepsilon) + \frac{1}{(2m)!} = 0\,.$$

By using that $1 - nx \le (1-x)^n \le 1 - nx + n^2x^2$ for any $x > 0$,

$$\begin{aligned} 0 &\le -(1-(2m-1)\varepsilon)\varepsilon + \frac{1-(m-1)\varepsilon + (m-1)^2\varepsilon^2}{(m+1)!}(m-(m+1)\varepsilon) + \frac{1}{(2m)!} \\ &\le a\varepsilon^2 + b\varepsilon + c\,, \end{aligned}$$

where

$$\begin{aligned} a &= \left(2m - 1 + \frac{(m-1)(m^2+1)}{(m+1)!}\right), \\ b &= -\left(1 + \frac{m^2+1}{(m+1)!}\right) \text{ and} \\ c &= \left(\frac{m}{(m+1)!} + \frac{1}{(2m)!}\right)\,. \end{aligned} \tag{6.6}$$

Let ε' be such that $a(\varepsilon')^2 + b\varepsilon' + c = 0$. Then,

$$\rho_{(1,2,\dots,m)} \ge (1-\varepsilon') = 1 - \frac{-b-\sqrt{b^2-4ac}}{2a}\,. \tag{6.7}$$

If m is large enough we can get an asymptotic expression for ε'. Suppose that $b^2 \gg 4ac$, then the smallest root of $ax^2 + bx + c = 0$ can be approximated by

$$x = -\frac{c}{b} - \frac{ac^2}{b^3} + O\left(\frac{a^2c^3}{b^5}\right)\,. \tag{6.8}$$

By putting together (6.6) and (6.8), and by using that $(1+x)^{-1} = 1 - x + O(x^2)$, we have

$$\begin{aligned} \varepsilon' &= \frac{\frac{m}{(m+1)!} + \frac{1}{(2m)!}}{1 + \frac{m^2+1}{(m+1)!}} + O\left(\frac{m^3}{(m+1)!^2}\right) \\ &= \frac{m}{(m+1)!} + O\left(\frac{m^3}{(m+1)!^2}\right) \\ &= \frac{1}{m!\left(1+\frac{1}{m}\right)} + O\left(\frac{m^3}{(m+1)!^2}\right) \\ &= \frac{1}{m!} - \frac{1}{m \cdot m!} + O\left(\frac{1}{m^2 \cdot m!}\right)\,, \end{aligned}$$

for large enough m. This proves the lemma. □

6.3.2 The CMP conjecture for small values of m.

Theorem 6.3 and Corollary 6.4 are stated for sufficiently large values of m to avoid some technicalities. In this subsection we will make a refinement of our former analysis which allows us to derive the CMP conjecture for $m \ge 5$. First we need an auxiliary lemma that will be useful for small values of m.

Lemma 6.13. *Let $\sigma \in \mathcal{S}_m \setminus \{(1,2,\ldots,m),(m,\ldots,2,1)\}$ with $m \geq 5$. If $m-2 \in \mathcal{O}_\sigma$, then $m-3 \notin \mathcal{O}_\sigma$.*

Proof. Suppose that $m-3 \in \mathcal{O}_\sigma$ and $m-2 \in \mathcal{O}_\sigma$, that is, $\mathrm{st}(\sigma_1,\ldots,\sigma_{m-2}) = \mathrm{st}(\sigma_3,\ldots,\sigma_m)$ and $\mathrm{st}(\sigma_1,\ldots,\sigma_{m-3}) = \mathrm{st}(\sigma_4,\ldots,\sigma_m)$. Thus, $\mathrm{st}(\sigma_3,\ldots,\sigma_m)$ is a monotone permutation since for each $i \in \{1,\ldots,m-4\}$, if $\sigma_i > \sigma_{i+1}$ then both $\sigma_{i+2} > \sigma_{i+3}$ and $\sigma_{i+3} > \sigma_{i+4}$. Without loss of generality assume that $\mathrm{st}(\sigma_3,\ldots,\sigma_m) = (1,2,\ldots,m-2)$.

Moreover, we have $\sigma_1 < \sigma_2 < \sigma_3$ since $\mathrm{st}(\sigma_1,\sigma_2,\sigma_3) = \mathrm{st}(\sigma_3,\sigma_4,\sigma_5)$ and $\mathrm{st}(\sigma_3,\sigma_4,\sigma_5) = (1,2,3)$. Thus $\mathrm{st}(\sigma_1,\ldots,\sigma_m) = (1,2,\ldots,m)$ getting a contradiction. ☐

Proposition 6.14. *The CMP conjecture is true for $m \geq 5$.*

Proof. On the one hand, a more precise upper bound on ρ_σ, $\sigma \in \mathcal{S}_m \setminus \{(1,2,\ldots,m),(m,\ldots,2,1)\}$, can be derived from (2.22) by using directly the upper bound on Δ^* provided in (6.4),

$$\rho_\sigma \leq U(\sigma) = \exp\left(-\frac{1}{m!} + e^{-\frac{4}{(m-1)!}} \sum_{k \in \mathcal{O}_\sigma} \frac{\binom{2(m-k)}{m-k}}{(2m-k)!}\right) .$$

On the other hand, let $L(m)$ be the lower bound on $\rho_{(1,2,\ldots,m)}$ that follows from (6.6) and (6.7).

Since σ is not monotone, by Lemma 6.8, we have that $m-1 \notin \mathcal{O}_\sigma$. One can check, by using an algebraic manipulator, that for every $\sigma \in \mathcal{S}_m \setminus \{(1,2,\ldots,m),(m,\ldots,2,1)\}$, $U(\sigma) \leq L(m)$ as long as $m \geq 7$, which implies the CMP conjecture for all these values.

Consider now that $\sigma \in \mathcal{S}_6 \setminus \{(1,2,\ldots,6),(6,\ldots,2,1)\}$. By Lemma 6.8 and Lemma 6.13, $\mathcal{O}_\sigma$ is a subset of either $\{1,2,4\}$ or $\{1,2,3\}$. One can check in both cases that $U(\sigma) \leq L(6)$

To conclude the proof, fix a pattern $\sigma \in \mathcal{S}_5 \setminus \{(1,2,\ldots,5),(5,\ldots,2,1)\}$. By using again Lemma 6.8 and Lemma 6.13, $\mathcal{O}_\sigma$ is a subset of either $\{1,3\}$ or $\{1,2\}$. If $3 \notin \mathcal{O}_\sigma$, then one can check that $U(\sigma) \leq L(5)$. Otherwise we need to improve a particular case of Lemma 6.10.

Assume that $3 \in \mathcal{O}_\sigma$. We claim that

$$\Pr(A_i \cap A_{i+2}) \leq \frac{2}{7!} .$$

We will count the number of $\tau = (\tau_1,\ldots,\tau_7) \in S_7$ that satisfy the events A_0 and A_2. Notice that $\mathrm{st}(\tau_3,\tau_4,\tau_5)$ is not monotone, otherwise, σ would be also monotone. By symmetry we can assume that $\mathrm{st}(\tau_3,\tau_4,\tau_5) = (1,3,2)$, thus $\tau_3 < \tau_5 < \tau_4$. Besides, the events A_0 and A_2 imply that $\mathrm{st}(\tau_3,\tau_4,\tau_5) = \mathrm{st}(\tau_1,\tau_2,\tau_3) = \mathrm{st}(\tau_3,\tau_6,\tau_7)$. As a consequence, we also have $\tau_1 < \tau_3 < \tau_2$ and $\tau_5 < \tau_7 < \tau_6$. Then, $\tau_1 = 1$ and $\tau_3 = 2$.

Let us continue by distinguishing cases depending on σ_2. If $\sigma_2 < \sigma_5$ ($\sigma = (1,3,2,5,4)$), then all the other elements are fixed and $\tau = (1,3,2,5,4,7,6)$. If $\sigma_2 > \sigma_4$ ($\sigma = (1,5,2,4,3)$), then $\tau = (1,7,2,6,3,5,4)$. Finally, if $\sigma_5 < \sigma_2 < \sigma_4$ ($\sigma = (1,4,2,5,3)$), then there are two options to complete τ, $(1,4,2,6,3,7,5)$ and $(1,5,2,6,3,7,4)$.

There are 7! possible permutations for $\mathrm{st}(\pi_{i+1},\ldots,\pi_{i+7})$ each of them appearing with the same probability. From these permutations at most 2 satisfy the events A_0 and A_2 for a given $\sigma \in \mathcal{S}_5 \setminus \{(1,2,\ldots,5),(5,\ldots,2,1)\}$. Thus, the claim follows.

It can be checked that $L(5)$ is larger than the bound we get by setting

$$\Delta^* = (\Pr(A_i \cap A_{i+2}) + \Pr(A_i \cap A_{i+4}))\, n \leq \left(\frac{2}{7!} + \frac{\binom{8}{4}}{9!} \right) n$$

in (2.22), where $\Pr(A_i \cap A_{i+4})$ has been computed by using Lemma 6.10. □

6.4 A lower bound on ρ_σ.

The setting used to give an upper bound on the number of permutations avoiding a given pattern can be also used to provide a lower bound on ρ_σ. Now we need a way to bound from below the probability that $X = 0$ and for this purpose we will use the Lovász Local Lemma.

Usually, the Local Lemma is used to show the existence of a certain configuration that does not satisfy any of the bad events in $\mathcal{A}$. In our problem it is trivial to see that, for any pattern $\sigma \in \mathcal{S}_m$, there exists at least one permutation of length n that avoids σ. We are interested in providing an explicit lower bound on the probability that a permutation selected uniformly at random avoids σ. This can be also attained through the local lemma. Thus, we will use it to derive a lower bound on the number of permutations of length n that avoid σ.

The one–sided version of the Lovász Local Lemma (see Lemma 2.14) is particularly convenient for our approach.

Next, we show how the use this version of the local lemma to prove a lower bound on ρ_σ.

Proof of Theorem 6.5. Let $\mathcal{A} = \{A_0, \ldots, A_{n-m}\}$ and let X be defined as in Section 6.3. Set $m(i) = i - m + 1$. By using Lemma 6.7 with $S = \{i\}$ and $T = \{0, 1 \ldots, i - m\}$, we have

$$\Pr\left(A_i \mid \bigcap_{j \leq i-m} \overline{A_j} \right) = \Pr(A_i) \; .$$

Recall that $\Pr(A_i) = \frac{1}{m!}$. Since all the events have the same probability to appear, we set $x_i = x$, for each $0 \leq i \leq n - m$, in (2.19) from Lemma 2.14, to get

$$\frac{1}{m!} \leq x(1-x)^{m-1} \; .$$

By setting $x = \frac{e^{\frac{m-1}{m!}}}{m!}$, the above inequality is satisfied and the Local Lemma can be applied. In particular, we obtain the following lower bound on the probability that $X = 0$,

$$\Pr(X = 0) = \Pr\left(\bigcap_{i=0}^{n-m} \overline{A_i} \right) \geq \left(1 - \frac{e^{(m-1)/m!}}{m!} \right)^{n-m+1} .$$

and, by using (6.1),

$$\rho_\sigma \geq 1 - \frac{e^{(m-1)/m!}}{m!} = 1 - \frac{1}{m!} - O\left(\frac{m}{(m!^2)} \right) \; .$$

□

The lower bound given by Theorem 6.5 is tight. This can be showed by using a result of Elizalde [49], where the author proved that the least avoided pattern is $(1, 2, \ldots m-2, m, m-1)$. The author also gives an implicit lower bound on $z_0 = \rho^{-1}_{(1,2,\ldots,m-2,m,m-1)}$ as the smallest real root of

$$f(z) = 1 - z + \frac{z^m}{m!} - m\frac{z^{2m+1}}{(2m-1)!} \ .$$

The following explicit upper bound can be derived from the previous equation, as in Lemma 6.12,

$$\rho_{(1,2,\ldots,m-2,m,m-1)} \leq 1 - \frac{1}{m!} - \Omega\left(\frac{m}{(m!)^2}\right) \ ,$$

showing that Theorem 6.5 is tight.

6.5 The typical value of ρ_σ.

The results of the previous sections provide tight upper and lower bounds on ρ_σ for every $\sigma \in \mathcal{S}_m$. In this section we want to show that, for a typical pattern, ρ_σ lies much closer to the lower bound than to the upper bound. That is, the number of σ–avoiding permutations of length n, for $\sigma \in \mathcal{S}_m$ chosen uniformly at random, is typically closer to the number of permutations that avoid $(1, 2, \ldots, m-2, m, m-1)$ than to the number of permutations that avoid $(1, 2, \ldots, m)$.

Define $\mathcal{N}_k \subseteq \mathcal{S}_m$ as the set of patterns of length m that overlap at position k. The following lemma bounds from above the size of these sets.

Lemma 6.15. *Let $\sigma \in \mathcal{S}_m$ be chosen uniformly at random. Then*

1. $\Pr(\sigma \in \mathcal{N}_k) = \frac{1}{k!}$ *if* $2 \leq 2k \leq m$.
2. $\Pr(\sigma \in \mathcal{N}_k) \leq \frac{2^m}{(m/2)!}$ *if* $m < 2k \leq 2(m-1)$.

Proof. Choose $\sigma \in \mathcal{S}_m$ uniformly at random. Recall that the condition for $\sigma \in \mathcal{N}_k$ is that $\tau^{(1)} = \mathrm{st}(\sigma_1, \ldots, \sigma_k)$ and $\tau^{(2)} = \mathrm{st}(\sigma_{m-k+1}, \ldots, \sigma_m)$ are equal to each other.

If $2k \leq m$, then $\tau^{(1)}$ and $\tau^{(2)}$ are independent, by Lemma 6.7, and uniformly distributed in $\mathcal{S}_k$. For every $\tau, \tau' \in \mathcal{S}_k$

$$\Pr(\tau^{(1)} = \tau \mid \tau^{(2)} = \tau') = \Pr(\tau^{(1)} = \tau) \ .$$

Thus, we can compute the exact probability of being in $\mathcal{N}_k$

$$\Pr(\sigma \in \mathcal{N}_k) = \Pr(\tau^{(1)} = \tau^{(2)}) = \sum_{\tau \in \mathcal{S}_k} \Pr(\tau^{(1)} = \tau \cap \tau^{(2)} = \tau) = k! \Pr(\tau^{(1)} = \tau)^2 = \frac{1}{k!} \ .$$

Suppose now that $2k > m$. For any integer $\ell \geq 1$, observe that $\mathcal{N}_{m-k} \subseteq \mathcal{N}_{m-\ell k}$. This in particular implies that $|\mathcal{N}_{m-k}| \leq |\mathcal{N}_{m-\ell k}|$. Thus, for any k such that $2k > m$ there exists an integer $k' \in [m/2, 3m/4]$ such that $|\mathcal{N}_k| \leq |\mathcal{N}_{k'}|$. So we may assume that $k \leq 3m/4$.

Partition the pattern σ in $s = \lfloor \frac{m}{m-k} \rfloor$ parts of length $m-k$ by defining the permutations $\tau^{(i)} = \mathrm{st}(\sigma_{(m-k)(i-1)+1}, \ldots, \sigma_{(m-k)i})$ for each $1 \leq i \leq s$ and one part, $\tau^{(s+1)}$, consisting in the last $m - s(m-k)$ positions. Observe that, in order to have an overlap at k we must have $\tau^{(1)} = \tau^{(i)}$ for each $i \le s$ and $\tau^{(s+1)} = \mathrm{st}(\sigma_{(s-1)(m-k)}, \dots, \sigma_k)$. This condition is clearly necessary but not sufficient for a pattern to overlap at k.

By the choice of σ, the permutations $\tau^{(i)}$ are uniformly distributed, and, by Lemma 6.7, they are mutually independent. This implies,

$$\Pr(\sigma \in \mathcal{N}_k) \le \Pr(\tau^{(s+1)} = \mathrm{st}(\sigma_{(s-1)(m-k)}, \dots, \sigma_k)) \prod_{i \le s} \Pr(\tau^{(i)} = \tau^{(1)})$$
$$= \frac{1}{(m - s(m-k))!} \left(\frac{1}{(m-k)!} \right)^{s-1} .$$

If $k \le 2m/3$, then $s = 2$. If $k = m/2 + t$, then

$$\Pr(\sigma \in \mathcal{N}_k) \le \frac{1}{(2t)!(m/2 - t)!} = \frac{\binom{m/2+t}{2t}}{(m/2+t)!} < \frac{2^m}{(m/2)!} .$$

If $2m/3 \le k \le 3m/4$, then $s = 3$. If $k = 2m/3 + t$, then

$$\Pr(\sigma \in \mathcal{N}_k) \le \frac{1}{(3t)!(m/3 - t)!(m/3-t)!} = \frac{\binom{m/3+2t}{3t}}{(m/3+t)!(m/3-t)!} < \frac{2^m}{(m/2)!} .$$

□

For each $1 \le k \le m - 1$, define the set $\mathcal{M}_k \subseteq \mathcal{S}_m$ as the set of patterns of length m such that $\mathcal{O}_\sigma \subseteq \{1, 2, \dots k\}$. Observe that $\mathcal{M}_{m-1}$ coincides exactly with $\mathcal{S}_m$. The elements in $\mathcal{M}_1$ are called *non–overlapping* patterns. They have been enumerated in [26] and also extensively studied in [49].

We use the previous lemma to give a lower bound on the size of $\mathcal{M}_k$.

Lemma 6.16. *Let $\sigma \in \mathcal{S}_m$ be chosen uniformly at random. Then, for each $1 \le k \le m/2$,*

$$\Pr(\sigma \in \mathcal{M}_k) \ge 1 - \frac{2}{(k+1)!} - m\frac{2^m}{(m/2)!}.$$

Proof. Observe that we can bound from below the size of $\mathcal{M}_k$ by using the sets $\mathcal{N}_k$,

$$|\mathcal{M}_k| = \left| \mathcal{S}_m \setminus \bigcup_{\ell=k+1}^{m-1} \mathcal{N}_\ell \right| \ge m! - \sum_{\ell=k+1}^{m-1} |\mathcal{N}_\ell| . \tag{6.9}$$

By Lemma 6.15, for each k such that $2k \le m$,

$$\sum_{\ell=k+1}^{m-1} \Pr(\sigma \in \mathcal{N}_\ell) \le \frac{1}{(k+1)!} + \frac{1}{(k+2)!} + \cdots + \frac{1}{\lfloor m/2 \rfloor !} + m\frac{2^m}{(m/2)!} .$$

The relation in (6.9) gives

$$\Pr(\sigma \in \mathcal{M}_k) \ge 1 - \sum_{\ell=k+1}^{m-1} \Pr(\sigma \in \mathcal{N}_\ell) \ge 1 - \sum_{\ell=k+1}^{m/2} \frac{1}{\ell!} - m\frac{2^m}{(m/2)!} \ge 1 - \frac{2}{(k+1)!} - m\frac{2^m}{(m/2)!} ,$$

which proves the statement. □

Recall that $\mathcal{M}_1$ corresponds to the set of non–overlapping patterns. The proof of Lemma 6.16 implies that $|\mathcal{M}_1| \geq (3-e)m!$. This bound can be refined. Indeed, Bóna [26] showed that

$$0.364098149 \leq \frac{|\mathcal{M}_1|}{m!} \leq 0.3640992743 .$$

The previous bound on $|\mathcal{M}_k|$ is clearly non sharp. A better estimation of the size of $\mathcal{N}_k$ when $2k > m$, would help to understand the distribution of ρ_σ when $\sigma \in \mathcal{S}_m$ is chosen uniformly at random.

Next lemma gives a better upper bound on Δ^* than the one in (6.5) when the pattern does not have a large overlap.

Lemma 6.17. *For every $\sigma \in \mathcal{M}_k$,*

$$\Delta^* \leq \frac{4^{m-k}}{(2m-k)!} n .$$

Proof. Since $\sigma \in \mathcal{M}_k$ we have $\Pr(A_i \cap A_{i+m-j}) = 0$ for all $k < j \leq m-1$. By Lemma 6.10,

$$\Delta^* \leq \sum_{i=0}^{n-m} \sum_{j=1}^{\max\{k,2m+j-n\}} \Pr(A_i \cap A_{i+m-j}) \leq n \sum_{j=1}^{k} \frac{\binom{2(m-j)}{m-j}}{(2m-j)!} \leq n \sum_{j=1}^{k} \frac{4^{m-j}}{\sqrt{\pi(m-j)}(2m-j)!} \leq \frac{4^{m-k}}{(2m-k)!} n .$$

□

Proof of Theorem 6.6. Assume that $\sigma \in \mathcal{M}_k$. By using the notation for μ, Δ^* and δ^* from Section 6.3, it follows from Lemma 6.17 that

$$\frac{\Delta^*}{\mu} \leq \frac{4^{m-k} m!}{(2m-k)!} = \frac{4^m}{\binom{2m-k}{m}(m-k)!} \leq \frac{4^m}{(m-k)!} .$$

and that $e^{2\delta^*} \leq e^{4/(m-1)!} \leq 2$ for each $m \geq 4$. Analogously to the proof of Theorem 6.3, we can derive the following upper bound,

$$\rho_\sigma \leq 1 - \frac{1}{m!} + O\left(\frac{4^m}{(m-k)!m!}\right) .$$

The above upper bound is satisfied when $\sigma \in \mathcal{M}_k$. This event holds, by Lemma 6.16, with probability at least $1 - \frac{2}{(k+1)!} - m\frac{2^m}{(m/2)!}$. This completes the proof. □

6.6 Concluding remarks and open questions

1. The techniques displayed in this chapter could be applied for the study other type of patterns in permutations [27, 88] or in other combinatorial structures, like matrices [67, 99].

In permutations, the most interesting case appears when considering non–consecutive pattern avoiding. One may think that the ideas introduced here could be useful to study this problem. Unfortunately, due to the strong dependency among a large part of the defined events, our techniques seem useless there.

Generalized patterns have been recently introduced [31] to cover all the different kinds of patterns in permutations. A generalized pattern is a triple $p = (\sigma, S, T)$ with $\sigma \in S_m$ and $S, T \subseteq \{0, 1, \dots, m\}$. An occurrence of p in π is a set of positions $1 \le i_1 < \dots < i_m \le n$ such that $\mathrm{st}(\pi_{i_1}, \dots, \pi_{i_m}) = \sigma$. Moreover, for all $j \in S$, $i_{j+1} = i_j + 1$ and for all $k \in T$, $\pi_{i_k+1} = \pi_{i_k} + 1$. By convention, $i_0 = \pi_{i_0} = 0$ and $i_{m+1} = \pi_{i_{m+1}} = m + 1$.

Observe that consecutive patterns are obtained when $S = \{1, 2, \dots, m-1\}$ and $T = \emptyset$. If $S = T = \emptyset$, then we recover the definition of a non–consecutive pattern.

For a general set S, the number of dependencies created among the events is still too large to apply our techniques. However, if $S = \{1, 2, \dots, m-1\}$ one could apply the same ideas, even if $T \neq \emptyset$. In such a case, it can be useful to consider a lopsidependency graph in order to get rid of the positive correlations.

2. In order to prove Conjecture 6.2 by using probabilistic techniques, one could try to mimic the same strategy we have used for proving the CMP conjecture. First, determine the subset of patterns σ such that $\alpha_n(\sigma) = \alpha_n(1, 2, \dots, m-2, m, m-1)$, the least avoided pattern, and then, improve the lower bound given by Theorem 6.5 for the patterns which are not in the previous subset.

Our approach, however, is hopeless to tackle Conjecture 6.2. Notice that no assumption on the properties of the pattern σ has been used in the proof of Theorem 6.5, unlike in the proof of the upper bound in Theorem 6.3. Unfortunately, the local lemma cannot distinguish the different nature of the dependencies among events. Thus, no better lower bound can be achieved by restricting to a smaller subset of patterns. This prompts us to formulate the following question.

Question 6.18. *Let $\sigma \in \mathcal{S}_m$ be chosen uniformly at random. Is it true that*

$$\rho_\sigma \ge 1 - \frac{1}{m!} + f(m) ,$$

for certain $f(m) \ge 0$, with probability at least $g(m) > 0$?

This is also the main problem to provide a lower bound for $\rho_{(1,2,\dots,m)}$ (see Lemma 6.12) using our probabilistic setting.

Question 6.19. *Is it possible to provide a probabilistic proof for Lemma 6.12?*

To answer this question, one would need to understand the probabilities $\Pr\left(A_i \mid \bigcap_{j<i} \overline{A_j}\right)$ when $\sigma = (1, 2, \dots, m)$. It must be stressed that, in such case, the upper bound $\Pr\left(A_i \mid \bigcap_{j<i} \overline{A_j}\right) \le \Pr(A_i)$ does not suffice to provide a meaningful lower bound on $\rho_{(1,2,\dots,m)}$.

3. One of the crucial steps in proving Theorem 6.6 is to upper bound the size of the sets $\mathcal{N}_k$, for any $k \le m-1$. Lemma 6.15 provides the exact value of $|\mathcal{N}_k|$ when $2k \le m$. Nonetheless, the upper bound given when $m > 2k$ is far from being tight.

If $m > 2k$, not every pattern of length k is a candidate for $\tau^{(1)} = \mathrm{st}(\sigma_1, \dots, \sigma_k)$ and $\tau^{(2)} = \mathrm{st}(\sigma_{m-k+1}, \dots, \sigma_m)$. This observation suggests that, if $2k > m$, then the probability that a pattern σ chosen uniformly at random belongs to $\mathcal{N}_k$ is smaller than in the case when $2k \le m$. This motivates the following conjecture

Conjecture 6.20. *Let $\mathcal{N}_k$ be the set of patterns that overlap at position k. Then,*

$$|\mathcal{N}_k| \leq \frac{1}{k!} \ ,$$

for every $1 \leq k \leq m-1$.

This conjecture is also supported by numerical computations for small values of m.

Notice that the expected number of overlaps of a randomly chosen pattern can be expressed as the sum of $|\mathcal{N}_k|$. An interesting fact is that, if Conjecture 6.20 is true, then this is at most e. Furthermore, if this conjecture holds, then it would provide a stronger version for Theorem 6.6.

CHAPTER 7

On the Lonely Runner Conjecture

7.1 Introduction

Let n be a positive integer and let $v_1, \dots, v_n, v_{n+1}$ be a set of different positive real numbers, also called speeds. For any real number x, denote by $\|x\|$, the distance from x to the closest integer

$$\|x\| = \min\{x - \lfloor x \rfloor, \lceil x \rceil - x\} .$$

For any real number x, denote by $\{x\}$ its fractional part.

$$\{x\} = x - \lfloor x \rfloor .$$

The Lonely Runner Conjecture was posed independently by Wills [127] in 1967 and Cusick [41] in 1982. Suppose that $n+1$ runners are running on the unit circle with different speeds and starting at the origin. Then, for each runner, there is a time where he is far from all the other runners. More formally,

Conjecture 7.1. *For every $n \geq 1$, every set of different speeds $v_1, \dots, v_{n+1}$ and each $k \in [n+1]$, there exists a time t such that*

$$\|t(v_i - v_k)\| \geq \frac{1}{n+1} ,$$

for every $i \in [n+1]$, $i \neq k$.

This conjecture can be restated by assuming that the runner we want to isolate has speed zero. Thus, he stays at the origin all the time and one must show that there is a time where all the runners are far enough from the origin.

Conjecture 7.2 (Lonely Runner Conjecture)**.** *For every $n \geq 1$ and every set of nonzero speeds $v_1, \dots, v_n$, there exists a time t such that*

$$\|tv_i\| \geq \frac{1}{n+1} ,$$

for every $i \in [n]$.

From now on, when we talk about the Lonely Runner Conjecture we will refer to the statement of Conjecture 7.2.

Observe that, if true, the Lonely Runner Conjecture would be best possible. For the set of speeds,

$$v_i = i \quad \text{for every } i \in [n] \;, \tag{7.1}$$

there is no time for which all the runners are further from the origin than $\frac{1}{n+1}$. This example is not unique and an infinite family of extremal sets can be found in [70].

The Lonely Runner Conjecture appears in many different situations. We next describe some known results and some related motivations. Let us first notice that the conjecture is obviously true for $n = 1$, since at some point $\|tv_1\| = 1/2$, and it is also easy to show that it holds for $n = 2$. Many proofs for $n = 3$ are given in the context of diophantine approximation (see [18, 41]). A computer–assisted proof for $n = 4$ was given by Cusick and Pomerance motivated by a view-obstruction problem in geometry [42], and later Biena et al. [19] provided a simpler proof by connecting it to nowhere zero flows in regular matroids. The conjecture was proved for $n = 5$ by Bohmann, Holzmann and Kleitman [21]. Barajas and Serra [13] have showed that the conjecture holds for $n = 6$ by studying the regular chromatic number of distance graphs.

In [21], the authors also showed that the conjecture can be reduced to the case where all speeds are positive integers and in the sequel we will assume this to be the case. In such a case, we also may assume that t takes values on the $(0, 1)$ unit interval, since at $t \in \mathbb{Z}$, $\|tv_i\| = 0$ for all i.

On the other hand, the conjecture can be showed to be true in the case where the set of speeds has a special structure. For instance, Czerwiński [43] showed a strengthening of the conjecture for the case where all the speeds are chosen uniformly at random among all the n-subsets of $[N]$ as $N \to \infty$. In particular, Czerwiński's result implies that, for almost all sets of runners, there exists a time where all the runners are arbitrarily close to $1/2 \in (0, 1)$. The dependence of N with respect to n, for which this result holds, was improved recently by Alon [4].

Dubickas [46] used a result of Peres and Schlag [116] in lacunary integer sequences to prove that the conjecture holds if the sequence of increasing speeds grows fast enough; in particular, if n is large and

$$\frac{v_{i+1}}{v_i} \geq 1 + \frac{22 \log n}{n} \;,$$

for every $1 \leq i < n$. These results introduce the use of the Lovász Local Lemma to deal with the dependencies created among the runners.

Another approach to the conjecture is to reduce the gap of loneliness. That is, show that there exists a $\delta \leq \frac{1}{n+1}$ such that, for any set of nonzero speeds, there exists a time $t \in (0, 1)$ such that

$$\|tv_i\| \geq \delta \quad \text{for every } i \in [n] \;. \tag{7.2}$$

For this approach it is particularly useful to define the following sets,

$$A_i = \{t \in (0, 1) : \; \|tv_i\| < \delta\} \;.$$

For every $t \in A_i$, we will say that the i–th runner is δ–close to the origin at time t. Otherwise, we will say that the runner is δ–far from the origin at time t.

The set A_i can be thought of as an event in the probability space $(0,1)$ with the uniform distribution. In that case, notice that we have $\Pr(A_i) = 2\delta$ independently from the value of v_i.

If the following equation is satisfied

$$\Pr\left(\bigcap_{i=1}^{n} \overline{A_i}\right) > 0 \,, \tag{7.3}$$

then, there exists a time t for which (7.2) holds. Observe that this is not a necessary condition. For instance, if we consider the set of speeds given by (7.1), the Lonely Runner Conjecture is satisfied but it can be checked that $\Pr(\cap_{i=1}^{n} \overline{A_i}) = 0$ when $\delta = \frac{1}{n+1}$.

Here, it is also convenient to consider the indicator random variables X_i for the events A_i. Let $X = \sum_{i=1}^{n} X_i$ count the number of runners which are δ–close from the origin at a time $t \in (0,1)$ chosen uniformly at random. Then, condition (7.3) is equivalent to $\Pr(X = 0) > 0$.

A first straightforward result in this direction is obtained by using the union bound in (7.3). For any $\delta < \frac{1}{2n}$, we have

$$\Pr\left(\bigcap_{i=1}^{n} \overline{A_i}\right) \geq 1 - \sum_{i=1}^{n} \Pr(A_i) = 1 - 2\delta n > 0 \,.$$

This result was improved by Chen [35], who showed that, for any set of n nonzero speeds, there exists a time t such that

$$\|tv_i\| \geq \frac{1}{2n - 1 + \frac{1}{2n-3}} \,, \tag{7.4}$$

for every $i \in [n]$.

If $2n - 3$ is a prime number, then the previous result was extended by Chen and Cusick [36]. They showed that for any set of n speeds, there exists a time t such that

$$\|tv_i\| \geq \frac{1}{2n-3} \,,$$

for every $i \in [n]$.

In order to improve (7.4), we exactly compute the pairwise join probabilities $\Pr(A_i \cap A_j)$, the amount of time that two runners spend close to the origin at the same time. As a corollary, we give the following lower bound on $\mathbb{E}(X^2)$.

Proposition 7.3. *For any $\delta < 1$, we have*

$$\mathbb{E}(X^2) \geq 2\delta n \left(\delta\left(1 + \frac{c}{\log \delta^{-1}}\right) n + 1\right) \,,$$

for some constant $c > 0$.

Using this bound, we are able to improve Chen's result on the gap of loneliness around the origin.

Theorem 7.4. *For every $\varepsilon > 0$, every sufficiently large n and every set of nonzero speeds $v_1, \dots, v_n$, there exists a time $t \in (0,1)$ such that*

$$\|tv_i\| \geq \frac{1}{2n-2+\varepsilon} \, ,$$

for every $i \in [n]$.

The proof of this theorem uses a Bonferroni–type inequality (see Lemma 2.6) that improves the union bound with the knowledge of pairwise intersections.

Another interesting result on the Lonely Runner Conjecture, was given by Czerwiński and Grytczuk [44]. We say that a runner k is *almost alone at time t* if there exists a $j \neq k$ such that

$$\|t(v_i - v_k)\| \geq \frac{1}{n+1} \, ,$$

for every $i \neq j, k$.

In [44], the authors showed that every runner is almost alone at some time. This means that Conjecture 7.2 is true, if we are allowed to to make one runner invisible, that is, there exists a time when all runners but one are far enough from the origin.

Theorem 7.5 ([44])**.** *For every $n \geq 1$ and every set of nonzero speeds $v_1, \dots, v_n$, there exist a time $t \in (0,1)$ and a $j \in [n]$ such that*

$$\|tv_i\| \geq \frac{1}{n+1}$$

for every $i \neq j$.

As a corollary of Proposition 7.3, we get the following result that extends Theorem 7.5 when n is large.

Theorem 7.6. *For every sufficiently large n and every set of nonzero speeds $v_1, \dots, v_n$, there exist $t_1, t_2 \in (0,1)$ and $j_1, j_2 \in [n]$, $j_1 \neq j_2$, such that for any $\ell \in \{1,2\}$,*

$$\|t_\ell v_i\| \geq \frac{1}{n+1} \, ,$$

for any $i \neq j_\ell$.

This theorem extends Theorem 7.5 by showing the existence of not only one but two runners whose deletion leave the origin alone at some point.

A similar result can be derived by using a model of dynamic circular interval graphs. Then, we can show that at least two runners are almost alone at the same time.

Theorem 7.7. *For every sufficiently large n and every set of different speeds $v_1, \dots, v_n$, there exist a time $t \in (0,1)$, $k_1, k_2 \in [n]$, $k_1 \neq k_2$, and $j_1, j_2 \in [n]$ such that for any $\ell \in \{1,2\}$,*

$$\|t(v_i - v_{k_\ell})\| \geq \frac{1}{n} \, ,$$

for any $i \neq k_\ell, j_\ell$.

This chapter is organized as follows. In Section 7.2 we compute the pairwise join probabilities for the events A_i and give a proof for Proposition 7.3. As a corollary of these results, we also show Theorem 7.4 (Subsection 7.2.1) and Theorem 7.6 (Subsection 7.2.2). In Section 7.3 we introduce an approach on the problem based on dynamic interval graphs and prove Theorem 7.7. Finally, in Section 7.4 we provide some conclusions and open questions.

7.2 Correlation among runners

In this section we want to study the pairwise join probabilities $\Pr(A_i \cap A_j)$, for any $i, j \in [n]$. Notice first, that, if A_i and A_j were independent events, then we would have $\Pr(A_i \cap A_j) = 4\delta^2$. This is not true in the general case, but, as we will see later on, these probabilities can be showed to be large enough.

Let us start by studying the case when the speeds v_i and v_j are coprime. For each ordered pair (i, j) with $i, j \in [n]$, we define

$$\varepsilon_{ij} = \left\{ \frac{v_i}{(v_i, v_j)} \delta \right\} , \tag{7.5}$$

where (v_i, v_j) denotes the greatest common divisor of v_i and v_j.

Let us also consider the function $f : (0, 1) \times (0, 1) \to \mathbb{R}$, defined by

$$f(x, y) = \min(x, y) + \max(x + y - 1, 0) - 2xy . \tag{7.6}$$

Proposition 7.8. *Let $v_j < v_i$ be coprime positive integers and $0 < \delta < 1$. Then*

$$\Pr(A_i \cap A_j) = 4\delta^2 + \frac{2f(\varepsilon_{ij}, \varepsilon_{ji})}{v_i v_j} .$$

Proof. By the sake of simplicity, we write $A = A_i$ and $B = A_j$. Observe that A and B can be expressed as

$$A = \bigcup_{k=0}^{v_i - 1} \left(\frac{k}{v_i} - \alpha, \frac{k}{v_i} + \alpha \right) \qquad B = \bigcup_{l=0}^{v_j - 1} \left(\frac{l}{v_j} - \beta, \frac{l}{v_j} + \beta \right)$$

where $\alpha = \delta / v_i$ and $\beta = \delta / v_j$.

If $I = (-\alpha, \alpha)$ and $J = (-\beta, \beta)$, we have

$$\begin{aligned}
\Pr(A \cap B) &= \Pr\left(\bigcup_{k \le v_i, l \le v_j} (I + k/v_i) \cap (J + l/v_j) \right) \\
&= \Pr\left(\bigcup_{k \le v_i, l \le v_j} I \cap (J + l/v_j - k/v_i) \right) \\
&= \sum_{k=0}^{v_i v_j - 1} \Pr\left(I \cap (J + k/v_j v_i) \right) ,
\end{aligned}$$

where in the last equality we use the fact that $(v_i, v_j) = 1$.

For each $-1/2 < x < 1/2$, define $d(x) = \Pr(I \cap (J + x))$. We can write $d(x)$ as follows (see Figure 7.1):

$$d(x) = \begin{cases} \beta + \alpha + x, & x \in [-(\beta+\alpha), -(\beta-\alpha)] \\ 2\alpha, & x \in [-(\beta-\alpha), \beta-\alpha] \\ \beta+\alpha-x, & x \in [\beta-\alpha, \beta+\alpha] \\ 0 & \text{otherwise} \end{cases}$$

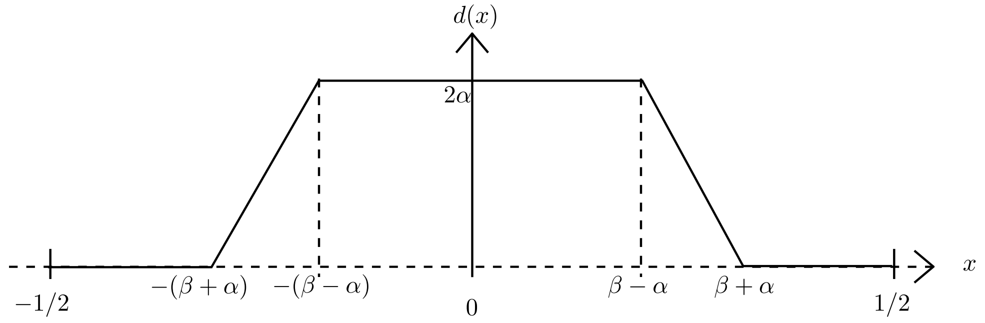


Figure 7.1: Plot of $d(x)$ in $(-1/2, 1/2)$.

By symmetry, we have

$$\frac{\Pr(A \cap B)}{2} = \frac{d(0)}{2} + \sum_{j=1}^{(\beta+\alpha)v_iv_j} d\left(\frac{j}{v_iv_j}\right) = \alpha + \sum_{j=1}^{(\beta+\alpha)v_iv_j} \min\left(2\alpha, \beta+\alpha-\frac{j}{v_iv_j}\right) .$$

Write $\alpha v_i v_j = p + \varepsilon_{ji}$ and $\beta v_i v_j = q + \varepsilon_{ij}$, where p and q are integers and $0 \le \varepsilon_{ji}, \varepsilon_{ij} < 1$.

Observe that

$$d\left(\frac{q-p}{v_iv_j}\right) v_iv_j = \begin{cases} 2(p+\varepsilon_{ji}) & \text{if } \varepsilon_{ji} \le \varepsilon_{ij} \\ 2p+\varepsilon_{ji}+\varepsilon_{ij} & \text{if } \varepsilon_{ji} > \varepsilon_{ij} \end{cases} = 2p + \varepsilon_{ji} + \min(\varepsilon_{ji}, \varepsilon_{ij}) ,$$

and that

$$d\left(\frac{q+p+1}{v_iv_j}\right) v_iv_j = \begin{cases} 0 & \text{if } \varepsilon_{ji}+\varepsilon_{ij} \le 1 \\ \varepsilon_{ji}+\varepsilon_{ij}-1 & \text{if } \varepsilon_{ji}+\varepsilon_{ij} > 1 \end{cases} = \max(0, \varepsilon_{ji}+\varepsilon_{ij}-1) .$$

Then,

$$\begin{aligned} \frac{\Pr(A \cap B)}{2} v_iv_j &= p + \varepsilon_{ji} + \sum_{j=1}^{p+q+\varepsilon_{ji}+\varepsilon_{ij}} \min(2(p+\varepsilon_{ji}), q+p+\varepsilon_{ji}+\varepsilon_{ij}-j) \\ &= p + \varepsilon_{ji} + \sum_{j=1}^{q-p-1} 2(p+\varepsilon_{ji}) + 2p + \varepsilon_{ji} + \min(\varepsilon_{ji}, \varepsilon_{ij}) \\ &\quad + \sum_{j=q-p+1}^{p+q} (q+p+\varepsilon_{ji}+\varepsilon_{ij}-j) + \max(0, \varepsilon_{ji}+\varepsilon_{ij}-1) \\ &= 2(p+\varepsilon_{ji})(q+\varepsilon_{ij}) + f(\varepsilon_{ji}, \varepsilon_{ij}) . \end{aligned}$$

Thus,

$$\Pr(A \cap B) = \frac{2}{v_i v_j}\left(2(p+\varepsilon_{ji})(q+\varepsilon_{ij}) + f(\varepsilon_{ji},\varepsilon_{ij})\right) = 4\delta^2 + \frac{2f(\varepsilon_{ji},\varepsilon_{ij})}{v_i v_j} .$$

□

It is easy to generalize Proposition 7.8 for non coprime numbers.

Proposition 7.9. *Let $v_j < v_i$ be positive integers and $0 < \delta < 1$. Then*

$$\Pr(A_i \cap A_j) \geq 4\delta^2 + \frac{2(v_i,v_j)^2 f(\varepsilon_{ji},\varepsilon_{ij})}{v_i v_j} .$$

Proof. Consider $v'_i = \frac{v_i}{(v_i,v_j)}$ and $v'_j = \frac{v_j}{(v_i,v_j)}$. Define $A'_i = \{t \in (0,1) : \|tv'_i\| < \delta\}$ and $A'_j = \{t \in (0,1) : \|tv'_j\| < \delta\}$. Observe that

$$\Pr(A_i \cap A_j) = \Pr(A'_i \cap A'_j) .$$

The proof follows by applying Proposition 7.8 to v'_i and v'_j, which are coprime. □

The proofs of Propositions 7.8 and 7.9 are based on the proofs of Lemmas 3.4 and 3.5 in a paper of Alon and Ruzsa [7].

Observation 7.10. *The pairwise join probability given by Proposition 7.9 is minimized when $v_j = 1$ and $v_i = \lfloor \delta^{-1} \rfloor$. Thus, for any v_i and v_j we have*

$$\Pr(A_i \cap A_j) \geq 4\delta^2 + \frac{2f(\delta, \delta\lfloor\delta^{-1}\rfloor)}{\lfloor\delta^{-1}\rfloor} \geq 2\delta^2 , \tag{7.7}$$

which follows by noting that $\delta\lfloor\delta^{-1}\rfloor \geq 1-\delta$.

Using the previous inequality, we can provide a first lower bound on the second moment of X,

$$\mathbb{E}(X^2) = \sum_{i\neq j} \Pr(A_i \cap A_j) + \sum_{i=1}^{n} \Pr(A_i) \geq 2\delta^2 n(n-1) + 2\delta n \geq 2\delta n(\delta(n-1)+1) . \tag{7.8}$$

We devote the rest of this section to improve (7.8). Let us first show when f is nonnegative.

Lemma 7.11. *The function $f(x,y)$ is nonegative in $[0,1/2]^2$ and in $[1/2,1)^2$.*

Proof. If $0 \leq x, y \leq 1/2$, then $\min(x,y) \geq 2xy$, which implies $f(x,y) \geq 0$.

Moreover,

$$\begin{aligned} f(1-x,1-y) &= \min(1-x,1-y) + \max(1-x-y,0) - 2(1-x-y+xy) \\ &= \min(y,x) + \max(0,x+y-1) - 2xy \\ &= f(x,y) . \end{aligned}$$

Therefore, we also have $f(x,y) \geq 0$ for all $1/2 \leq x, y \leq 1$. □

The following lemma shows that the error term of $\Pr(A_i \cap A_j)$ provided in Proposition 7.9, cannot be too negative if v_i and v_j are either too close or too far from each other.

Lemma 7.12. *Let $M \geq 2$ be any integer, $\gamma = M^{-1} > 0$ and $v_j < v_i$. If either $(1-\gamma)v_i \leq v_j$ or $\gamma\delta v_i \geq v_j$, then*

$$\frac{(v_i,v_j)^2 f\left(\varepsilon_{ij},\varepsilon_{ji}\right)}{v_i v_j} \geq -\gamma\delta^2 .$$

Proof. For the sake of simplicity, let us write $v_i/(v_i,v_j) = k\delta^{-1} + x$ and $v_j/(v_i,v_j) = l\delta^{-1} + y$ with k and l nonnegative integers and $0 \leq x, y < \delta^{-1}$. In particular, observe that $\varepsilon_{ij} = x\delta$ and $\varepsilon_{ji} = y\delta$. Moreover, we can assume that v_i and v_j are such that $f\left(\varepsilon_{ij},\varepsilon_{ji}\right)$ is negative; otherwise, the lemma is obviously true.

We split the proof in two different cases.

- **Case A:** $\gamma\delta v_i \geq v_j$. Observe that, since $v_j/(v_i,v_j) \geq 1$, we have $v_i/(v_i,v_j) \geq M\delta^{-1}$.
- **Case B:** $(1-\gamma)v_i \leq v_j$. By Lemma 7.11, in such case we can assume that either $k = l$, $y < \delta^{-1}/2$ and $x \geq \delta^{-1}/2$ (Cases B.1 and B.2); or $k = l+1$, $y \geq \delta^{-1}/2$ and $x < \delta^{-1}/2$ (Cases B.3 and B.4).

Figure 7.2 illustrates the situation considered in each subcase below.

Case A.1 ($y \leq x$)**:**

We have,

$$\frac{(v_i,v_j)^2 f\left(\varepsilon_{ij},\varepsilon_{ji}\right)}{v_i v_j} \geq \frac{(v_i,v_j)^2(\varepsilon_{ji} - 2\varepsilon_{ij}\varepsilon_{ji})}{v_i v_j} = \frac{(v_i,v_j)^2(y\delta - 2xy\delta^2)}{v_i v_j} \geq \frac{(v_i,v_j)(1-2x\delta)}{v_i} \cdot \delta ,$$

where the last inequality holds from the fact that $f\left(\varepsilon_{ij},\varepsilon_{ji}\right) < 0$ and $y \leq v_j/(v_i,v_j)$.

We have that $k\delta^{-1} + x = v_i/(v_i,v_j) \geq M\delta^{-1}$ implying $k \geq M$, since M is an integer. Observe also that, since $y \leq x$ and $f(\varepsilon_{ij},\varepsilon_{ji})$ is negative, by Lemma 7.11 we have $\delta^{-1}/2 \leq x < \delta^{-1}$. Then,

$$\frac{(v_i,v_j)^2 f\left(\varepsilon_{ij},\varepsilon_{ji}\right)}{v_i v_j} \geq \frac{(v_i,v_j)(1-2x\delta)}{v_i}\delta \geq \frac{1-2x\delta}{M+x\delta}\delta^2 \geq -\gamma\delta^2 .$$

Case A.2 ($y > x$)**:**

In this case,

$$\frac{(v_i,v_j)^2 f\left(\varepsilon_{ij},\varepsilon_{ji}\right)}{v_i v_j} \geq \frac{(v_i,v_j)^2(\varepsilon_{ij} - 2\varepsilon_{ij}\varepsilon_{ji})}{v_i v_j} = \frac{(v_i,v_j)^2(x\delta - 2xy\delta^2)}{v_i v_j} \geq \frac{(v_i,v_j)(1-2y\delta)}{v_j} \cdot \gamma\delta ,$$

where the last inequality holds from the fact that, in this case, $x \leq Mv_i/(v_i,v_j) = \gamma^{-1}v_i/(v_i,v_j)$.

As before, since $f(\varepsilon_{ij},\varepsilon_{ji})$ is negative, by Lemma 7.11 we have $\delta^{-1}/2 \leq y < \delta^{-1}$ and $v_j/(v_i,v_j) \geq y$. Therefore,

$$\frac{(v_i,v_j)^2 f\left(\varepsilon_{ij},\varepsilon_{ji}\right)}{v_i v_j} \geq \frac{(v_i,v_j)(1-2y\delta)}{v_j} \cdot \gamma\delta \geq \frac{1-2y\delta}{y} \cdot \gamma\delta \geq -\gamma\delta^2 .$$

Case B.1 ($k = l$ and $x + y \leq \delta^{-1}$):

In this case, since $x + y \leq \delta^{-1}$, $\max\{0, \varepsilon_{ij} + \varepsilon_{ji} - 1\} = 0$.

By using $v_i/(v_i, v_j) = k\delta^{-1} + x \geq k\delta^{-1}$ and $v_j/(v_i, v_j) = k\delta^{-1} + y \geq y$, we have

$$\frac{(v_i, v_j)^2 f(\varepsilon_{ij}, \varepsilon_{ji})}{v_i v_j} = \frac{(v_i, v_j)^2(\varepsilon_{ji} - 2\varepsilon_{ij}\varepsilon_{ji})}{v_i v_j} = \frac{(v_i, v_j)^2 y(1 - 2x\delta)}{v_i v_j}\delta \geq \frac{1 - 2x\delta}{k + x\delta}\delta^2 .$$

Since $v_j \geq (1 - \gamma)v_i$, we have $y \geq (1 - \gamma)x - \gamma k\delta^{-1}$. Combined with $x + y \leq \delta^{-1}$, it follows that $x \leq \frac{1+\gamma k}{2-\gamma}\delta^{-1}$. Thus,

$$\frac{(v_i, v_j)^2 f(\varepsilon_{ij}, \varepsilon_{ji})}{v_i v_j} \geq \frac{1 - 2x\delta}{k + x\delta}\delta^2 \geq \frac{1 - \frac{2(1+\gamma k)}{2-\gamma}}{k + \frac{1+\gamma k}{2-\gamma}}\delta^2 = -\gamma\delta^2 ,$$

for each $k \geq 0$.

Case B.2 ($k = l$ and $x + y \geq \delta^{-1}$):

Now, $\max\{0, \varepsilon_{ij} + \varepsilon_{ji} - 1\} = (x + y)\delta - 1$. Then,

$$\frac{(v_i, v_j)^2 f(\varepsilon_{ij}, \varepsilon_{ji})}{v_i v_j} = \frac{(v_i, v_j)^2(y\delta + (x + y)\delta - 1 - 2xy\delta^2)}{v_i v_j} .$$

This expression is minimized in the same point as in the case B.1, $x = \frac{1+\gamma k}{2-\gamma}\delta^{-1}$ and $y = (1 - \gamma)x - \gamma k\delta^{-1}$. Hence, the same computations suffice to show that

$$\frac{(v_i, v_j)^2 f(\varepsilon_{ij}, \varepsilon_{ji})}{v_i v_j} \geq -\gamma\delta^2 ,$$

for any $k \geq 0$.

Case B.3 ($k = l + 1$ and $x + y \leq \delta^{-1}$):

Again, $\max\{0, \varepsilon_{ij} + \varepsilon_{ji} - 1\} = 0$. Since $v_i/(v_i, v_j) = k\delta^{-1} + x \geq k\delta^{-1}$ and $v_j/(v_i, v_j) = (k - 1)\delta^{-1} + y \geq y$, $v_j \geq (1 - \gamma)v_i$ implies that $y \geq (1 - \gamma)x - (\gamma k - 1)\delta^{-1}$.

Then,

$$\frac{(v_i, v_j)^2 f(\varepsilon_{ij}, \varepsilon_{ji})}{v_i v_j} = \frac{(v_i, v_j)^2(\varepsilon_{ji} - 2\varepsilon_{ij}\varepsilon_{ji})}{v_i v_j} = \frac{(v_i, v_j)^2 y(1 - 2x\delta)}{v_i v_j}\delta \geq \frac{1 - 2x\delta}{k + x\delta}\delta^2 .$$

From the equations $y \geq (1-\gamma)x-(\gamma k-1)\delta^{-1}$ and $x+y \leq \delta^{-1}$, one can deduce that $x \leq \frac{\gamma k}{2-\gamma}\delta^{-1}$. Thus,

$$\frac{(v_i, v_j)^2 f(\varepsilon_{ij}, \varepsilon_{ji})}{v_i v_j} \geq \frac{1 - 2x\delta}{k + x\delta}\delta^2 \geq \frac{1 - \frac{2\gamma k}{2-\gamma}}{k + \frac{\gamma k}{2-\gamma}}\delta^2 = \left(\frac{2 - \gamma}{2k} - \gamma\right)\delta^2 \geq -\gamma\delta^2 ,$$

for any $k \geq 0$.

Case B.4 ($k = l + 1$ and $x + y \geq \delta^{-1}$): As in the case B.2, we have $\max\{0, \varepsilon_{ij} + \varepsilon_{ji} - 1\} = (x + y)\delta - 1$. Then,

$$\frac{(v_i, v_j)^2 f(\varepsilon_{ij}, \varepsilon_{ji})}{v_i v_j} = \frac{(v_i, v_j)^2(y\delta + (x + y)\delta - 1 - 2xy\delta^2)}{v_i v_j} .$$

This expression is minimized in the same point as in case B.3, $x = \frac{\gamma k}{2-\gamma}\delta^{-1}$ and $y = (1-\gamma)x - (\gamma k - 1)\delta^{-1}$. Hence, we have

$$\frac{(v_i, v_j)^2 f(\varepsilon_{ij}, \varepsilon_{ji})}{v_i v_j} \geq -\gamma\delta^2 \ ,$$

for any $k \geq 0$. □

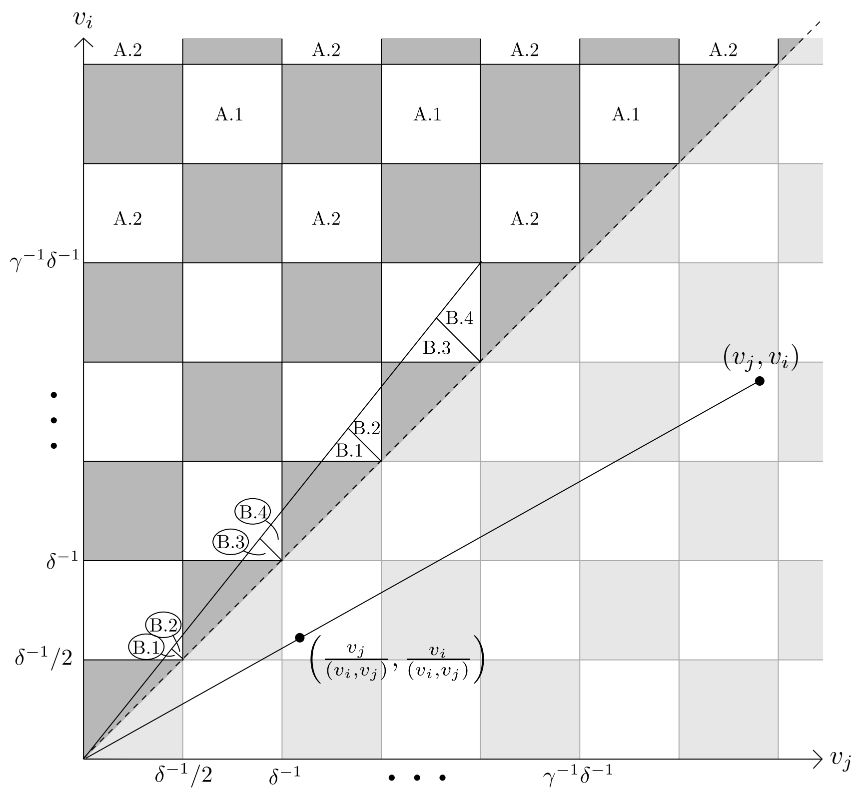


Figure 7.2: Different cases in the proof of Lemma 7.12. Grey areas correspond to positive values of $f(\varepsilon_{ij}, \varepsilon_{ji})$ according to Lemma 7.11.

The following lemma shows that among a large set of positive numbers, there should be a pair satisfying that they are either too close or too far from each other.

Lemma 7.13. *For every $c > 1$, $\alpha > 0$, $\delta < 1$ and every set $x_1 \geq \cdots \geq x_{m+1} > 0$ of nonnegative numbers, with $m \geq \log_c(\alpha\delta^{-1})$, there is a pair $i, j \in [m+1]$ such that*

$$\textit{either} \quad \frac{x_i}{x_j} \leq c \quad \textit{or} \quad \frac{x_i}{x_j} \geq \alpha\delta^{-1}.$$

Proof. Suppose that for each pair $i < j$ we have $x_i > cx_j$. In particular, for each $i \leq m$, we have $x_i > cx_{i+1}$ and $x_1 > c^m x_{m+1} \geq \alpha\delta^{-1}x_{m+1}$. Hence the second possibility holds for $i = 1$ and $j = m+1$. □

For any pair $i, j \in [n]$, we call the pair ε*–good* if $\Pr(A_i \cap A_j) \geq (1-\varepsilon)4\delta^2$. Now we are able to improve the lower bound on the second moment of X given in (7.8).

Proof of Proposition 7.3. Recall that by (7.7), for any pair $i, j \in [n]$, we have $\Pr(A_i \cap A_j) \geq 2\delta^2$. We will show that at least a $\Omega\left(\frac{1}{\log \delta^{-1}}\right)$ fraction of the pairs are ε–good.

Consider the graph H on the vertex set $V(H) = [n]$, where ij is an edge if and only if ij is ε–good. Using Lemma 7.13, we know that there are no independent sets of size larger than $m = \log_c(\alpha\delta^{-1}) = \frac{\log \delta^{-1}}{c'_\varepsilon} + 1$, where c'_ε depends only on ε. Thus, the complement of H, $\overline{H}$, has no clique of size m. By the Erdős–Stone theorem (see [54]), $|E(\overline{H})| \leq \frac{m-2}{m-1}\frac{n^2}{2}$, which implies that there are

$$|E(H)| \geq \frac{n^2}{2(m-1)} ,$$

ε–good unordered pairs.

Now, we are able to give a lower bound on the second moment,

$$\begin{aligned}
\mathbb{E}(X^2) &= \sum_{ij\ \varepsilon\text{–good}} \Pr(A_i \cap A_j) + \sum_{ij \text{ non } \varepsilon\text{–good}} \Pr(A_i \cap A_j) + \sum_{i=1}^{n} \Pr(A_i) \\
&\geq (1-\varepsilon)4\delta^2 \frac{n^2}{\log_c \alpha\delta^{-1} - 1} + 2\delta^2 \left(n(n-1) - \frac{n^2}{\log_c \alpha\delta^{-1} - 1} \right) + 2\delta n \\
&= (1-\varepsilon)4\delta^2 \frac{c'_\varepsilon n^2}{\log \delta^{-1}} + 2\delta^2 \left(1 - \frac{c'_\varepsilon}{\log \delta^{-1}} \right) n^2 + 2\delta n \\
&\geq 2\delta n \left(\delta \left(1 + \frac{c_\varepsilon}{\log \delta^{-1}} \right) n + 1 \right) ,
\end{aligned}$$

for some c_ε that depends only on ε.

□

Next, we show some applications of our bounds, that extend some known results.

7.2.1 First application: Improving the gap of loneliness

In this subsection we show how to use the pairwise join probabilities to prove Theorem 7.4.

For such a purpose, we will use the Bonferroni–type inequality displayed in Lemma 2.6. As we have already mentioned, $\Pr(A_i) = 2\delta$. Thus, it remains to select a tree T that maximizes $\sum_{ij \in E(T)} \Pr(A_i \cap A_j)$.

Lemma 7.14. *For each $\varepsilon' > 0$ and $\delta = o(1)$, there exists a tree T on the set of vertices $[n]$ such that*

$$\sum_{ij \in E(T)} \Pr(A_i \cap A_j) \geq (1-\varepsilon')4\delta^2 n .$$

Proof. By Proposition 7.9 we have

$$\Pr(A_i \cap A_j) = 4\delta^2 + \frac{2(v_i, v_j)^2 f(\varepsilon_{ij}, \varepsilon_{ji})}{v_i v_j}.$$

Set $\gamma < 2\varepsilon'$. We will construct a large forest F on the set of vertices $[n]$, where all the edges $ij \in E(F)$ are ε'–good. That is, they satisfy,

$$\Pr(A_i \cap A_j) \geq (1-\varepsilon')4\delta^2 = (4-2\gamma)\delta^2 \ .$$

Let us show how to select such edges by a procedure. Set $S_0 = [n]$ and $E_0 = \emptyset$. In the k-th step, we select different $i, j \in S_{k-1}$ such that either $v_i/v_j \leq (1-\gamma)^{-1}$ or $v_i/v_j \geq \gamma^{-1}\delta^{-1}$, and set $E_k = E_{k-1} \cup \{ij\}$, $S_k = S_{k-1} \setminus \{i\}$. If no such pair exists, we stop the procedure.

Let τ be the number of steps that the procedure runs before being halted. By Lemma 7.13 with $c = (1-\gamma)^{-1}$ and $\alpha = \gamma^{-1}$ we can always find such an edge ij, provided that the set S_k has size at least $\log_c(\alpha\delta^{-1})$. Thus $\tau \geq n - \log_c(\alpha\delta^{-1})$. Since the size of the sets E_k increases exactly by one at each step, we have $|E_\tau| \geq n - \log_c(\alpha\delta^{-1}) = (1-o(1))n$, since $\delta = o(1)$. Besides, E_τ is an acyclic set of edges. We are never closing a cycle since we always delete one of the endpoints of the selected edge, from the set S_k.

By Lemma 7.12, for each edge in E_τ we have

$$\Pr(A_i \cap A_j) \geq (4-2\gamma)\delta^2 \ .$$

Therefore we can construct a spanning tree T on the vertex set $[n]$ satisfying

$$\sum_{ij \in E(T)} \Pr(A_i \cap A_j) \geq (1-o(1))\,(4-2\gamma)\delta^2 n \geq (1-\varepsilon')4\delta^2 n \ .$$

□

Let us proceed to prove Theorem 7.4.

Proof of Theorem 7.4. By Lemma 2.6 and Lemma 7.14 with $\varepsilon' = \varepsilon/2$, we have

$$\begin{aligned}\Pr\left(\bigcap_{i=1}^{n} \overline{A_i}\right) &\geq 1 - \sum_{i=1}^{n}\Pr(A_i) + \sum_{ij\in E(T)}\Pr(A_i\cap A_j)\\ &\geq 1 - 2\delta n(1-2(1-\varepsilon')\delta) \ .\end{aligned}$$

The above expression is strictly positive for

$$\delta \leq \frac{1}{2n-2+2\varepsilon'} = \frac{1}{2n-2+\varepsilon} \ ,$$

and the theorem follows. □

7.2.2 Second Application: Invisible runners

This subsection is devoted to the proof of Theorem 7.6. In particular, we use the result of Proposition 7.3 to show that there is a large fraction of time where only one runner is δ–close to the origin, for $\delta = \frac{1}{n+1}$. This implies the existence of at least two runners whose delition leave the origin alone at some time.

Proof of Theorem 7.6. For $\delta = \frac{1}{n+1}$, we have $\mathbb{E}(X) = \frac{2n}{n+1} = 2 - (1 + O(n^{-1}))\frac{2}{n}$. Moreover, by Proposition 7.3, for any $\varepsilon > 0$

$$\mathbb{E}(X^2) \geq (1 + O(n^{-1}))\left(4 + \frac{2c}{\log n}\right) ,$$

for some constant $c > 0$.

For every $0 \leq k \leq n$, let $p_k := \Pr(X = k)$. We may assume that $p_0 = 0$ since otherwise, there would exists a time when all the runners are $\frac{1}{n+1}$–far from the origin, which implies Conjecture 7.2. Then we have the following system of linear equations,

$$\begin{array}{ccccccccc} p_1 & + & p_2 & + & \dots & + & p_n & = & 1 \\ p_1 & + & 2p_2 & + & \dots & + & np_n & = & \mathbb{E}(X) \\ p_1 & + & 4p_2 & + & \dots & + & n^2 p_n & = & \mathbb{E}(X^2) . \end{array}$$

From these equations one can deduce that,

$$p_1 = 2 - \mathbb{E}(X) + \sum_{i=1}^{n}(i-2)p_i = (1 + O(n^{-1}))\frac{2}{n} + \sum_{i=1}^{n}(i-2)p_i$$

$$\sum_{i=1}^{n}(i-1)(i-2)p_i = \mathbb{E}(X^2) - 3\mathbb{E}(X) + 2 = (1 + O(n^{-1}))\frac{2c}{\log n} .$$

Then, p_1 is minimized when $p_3 = \cdots = p_{n-1} = 0$ and $p_n = (1 + O(n^{-1}))\frac{2c}{(n-1)(n-2)\log n}$. Thus,

$$p_1 \geq (1 + O(n^{-1}))\frac{2}{n} + (1 + O(n^{-1}))\frac{2c}{n \log n} = (1 + O(n^{-1}))\frac{2}{n - \frac{cn}{\log n}}$$

Since a runner spends no more than $2\delta = \frac{2}{n+1}$ fraction of the time close the origin and c does not depend on n, there should be at least two such runners that make the origin almost alone at some point. □

7.3 Weaker conjectures and interval graphs

In this section we give a proof for Theorem 7.7. The following weaker conjecture has been proposed by Spencer[1].

Conjecture 7.15 (Weak Lonely Runner Conjecture). *For every $n \geq 1$ and every set of different speeds $v_1, \dots, v_n$, there exist a time t and a runner $j \in [n]$, such that*

$$\|t(v_i - v_j)\| \geq \frac{1}{n}$$

for every $i \neq j$.

[1]Transmitted to the author by Jarek Grytczuk.

For every set $S \subseteq [n]$, we say that S is *isolated at time* t if,

$$\|t(v_i - v_j)\| \geq \frac{1}{n} \qquad \text{for each } i \in S,\ j \in V \setminus S. \tag{7.9}$$

Observe that $S = \{i\}$ is isolated at time t, if and only v_i is lonely at time t.

To study the appearance of isolated sets, it is convenient to define a dynamic graph $G(t)$, whose connected components are sets of isolated runners at time t. For each $1 \leq i \leq n$ and $t \in (0, 1)$, define the following dynamic interval of the torus $(0, 1)$ associated to the i-th runner,

$$I_i(t) = \left\{ x \in (0,1) \ : \{x - tv_i\} < \frac{1}{n} \right\} .$$

In other words, $I_i(t)$ is the interval that starts in the position $\{tv_i\}$ of the i-th runner at time t and has length $\frac{1}{n}$.

Now we can define the following dynamic circular interval graph $G(t) = (V(t), E(t))$. The vertex set $V(t)$ is composed by n vertices u_i that correspond to the set of runners, and two vertices u_i and u_j are connected if $I_i(t) \cap I_j(t) \neq \emptyset$ (see Figure 7.3).

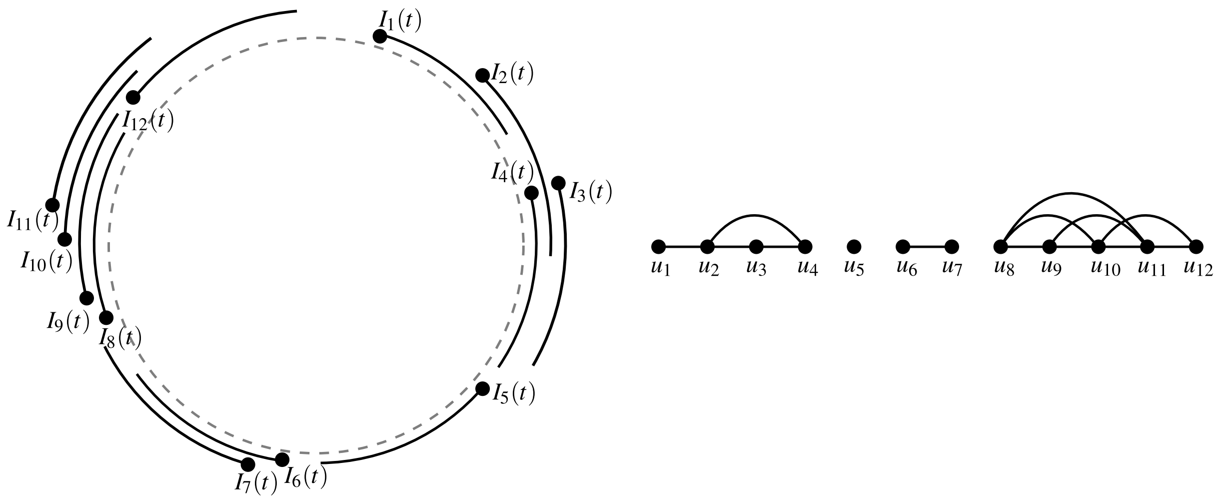


Figure 7.3: An instance of the graph $G(t)$.

Observation 7.16. *The graph $G(t)$ satisfies the following properties,*

1. *$G(0) = K_n$.*
2. *Each connected component of $G(t)$, correspond to an isolated set of runners at time t.*
3. *If u_i is isolated in $G(t)$, then v_i is alone at time t.*
4. *All the intervals have the same size, $|I_i(t)| = 1/n$, and thus, $G(t)$ is a unit circular interval graph.*

We can restate the Lonely Runner Conjecture in terms of the dynamic interval graph $G(t)$.

Conjecture 7.17 (Lonely Runner Conjecture)**.** *For any $i \leq n$ there exists a time t such that u_i is isolated in $G(t)$.*

For every subgraph $H \subseteq G(t)$ we define $\mu(H) = \mu(\cup_{u_i \in V(H)} I_i(t))$, the length of the arc occupied by the intervals corresponding to H. Notice that, if H contains an edge, then

$$\mu(H) < \frac{|V(H)|}{n} , \tag{7.10}$$

since the intervals $I_i(t)$ are closed in one extreme but open in the other one. If H consists of isolated vertices, then (7.10) does not hold.

The dynamic interval graph $G(t)$ allows us to prove a weak version of the conjecture. Let us assume that $v_1 > v_2 > \cdots > v_n$.

Proposition 7.18. *There exist a time t and a nonempty subset $S \subset [n]$ such that S is isolated at time t.*

Proof. Let t be the minimum number for which the equation $tv_1 - 1 = tv_n - 1/n$ holds. This is the first time that v_n is at distance exactly $1/n$ ahead from the fastest runner v_1.

For the sake of contradiction, assume that there is just one connected component of order n. Note that $u_1u_n \notin E(G(t))$ and since $G(t)$ is connected, there exists a path in $G(t)$ connecting u_1 and u_n. By (7.10), we have $\mu(G) < 1$. Thus, there is a point $x \in (0,1)$ such that $x \notin I_i(t)$ for any $i \in [n]$.

Observe that, at time t, all the intervals are sorted in increasing order around $(0,1)$. Let $\ell \in [n]$ be such that $x > \{tv_\ell\}$ and $x < \{tv_{\ell+1}\}$. Then, $\{u_1, \dots, u_\ell\}$ and $\{u_{\ell+1}, \dots, u_n\}$ are in different connected components, since $u_1u_n, u_\ell u_{\ell+1} \notin E(G(t))$. □

We observe that, if one of the parts in Proposition 7.18 consists of a singleton, say $S = \{i\}$, then we would have showed Conjecture 7.15.

Let us show how to apply the idea of the dynamic graph to prove an invisible lonely runner theorem, similar to Theorem 7.5.

Proposition 7.19. *There exists $t \in (0,1)$ such that $G(t)$ has either some isolated vertex or it has at least two vertices of degree one.*

Proof. Define $Y : (0,1) \to \mathbb{N}$ by,

$$Y(t) := |E(G(t))| .$$

Let $t \in (0,1)$ be chosen uniformly at random. Then $Y(t)$ is a random variable over $\left\{0, 1, \dots, \binom{n}{2}\right\}$. We will show that $\mathbb{E}(Y(t)) \leq (n-1)$. If we are able to do so, by a first moment argument, we know that there exists a time t_0 for which $Y(t_0) \leq n-1$. Then, denoting by d_i the degree of u_i, we have

$$\sum_{i=1}^{n} d_i \leq 2(n-1) ,$$

which, if $d_i > 0$ for each i, ensures the existence of at least 2 vertices of degree one, concluding the proof of the proposition.

Now, let us show that $\mathbb{E}(Y(t)) \leq (n-1)$. We can write $Y(t) = \sum_{i<j} Y_{ij}(t)$, where $Y_{ij}(t) = 1$ if u_i and u_j are connected at time t and $Y_{ij}(t) = 0$ otherwise. Then $\mathbb{E}(Y(t)) = \sum_{i<j} \mathbb{E}(Y_{ij}) =$

$\sum_{i<j} \Pr(I_i(t) \cap I_j(t) \neq \emptyset)$. For the sake of simplicity when computing $\Pr(I_i(t) \cap I_j(t) \neq \emptyset)$, we can assume that $v_i = 0$. Since the intervals are half open, half closed, we have $\Pr(I_i(t) \cap I_j(t) \neq \emptyset) = 2/n$, no matter the value of v_j.

Finally,

$$\mathbb{E}(Y(t)) = \sum_{i<j} \frac{2}{n} = \binom{n}{2}\frac{2}{n} = n - 1 \ .$$

□

In the dynamic interval graph setting, an invisible runner is equivalent to a vertex u with a neighbor of degree one, say v. If u is removed, then v becomes isolated in $G(t)$ and thus, alone in the runner setting. Thus, Theorem 7.7 is a direct corollary of Proposition 7.19.

7.4 Concluding remarks and open questions

1. Using the same strategy as the proof of Theorem 7.4 one can show that it holds for some $\varepsilon = \varepsilon(n) \to 0$.

Consider $\gamma = \gamma(n) \to 0$ and $m \leq \gamma n$. We can find a forest containing at least $n - m = (1-\gamma)n$ edges ij such that $\Pr(A_i \cap A_j) \geq (4 - 2\gamma)\delta^2$. In this case,

$$\sum_{ij \in E(T)} \Pr(A_i \cap A_j) \geq (4 - 2\gamma)\delta^2(n - m) = (1 - O(\gamma))4\delta^2 n \ .$$

As in the proof of Theorem 7.4, we may set $c = (1-\gamma)^{-1}$ and $\alpha = \gamma^{-1}$ in Lemma 7.13 to apply Lemma 7.12. Then, the following inequality must be satisfied,

$$\log_c \alpha\delta^{-1} = m \leq \gamma n \ .$$

Some technical but straightforward computations show that this inequality holds if γ is large enough,

$$\gamma = \gamma(n) = \Omega\left(\sqrt{\frac{\log n}{n}}\right) \ .$$

Since $\gamma = \Theta(\varepsilon)$ in the proof of Theorem 7.4, we have that it holds for any $\varepsilon = \Omega\left(\sqrt{\frac{\log n}{n}}\right)$.

2. Proposition 7.3, shows that for $\delta = \frac{1}{n+1}$ we have $\mathbb{E}(X^2) \geq (1 + O(n^{-1}))\left(4 + \frac{2c}{\log n}\right)$. However, we think that the proof of this proposition can be adapted to show that the second moment of X is even larger.

Conjecture 7.20. *For any set of different speeds $v_1, \ldots, v_n$, and $\delta = \frac{1}{n+1}$, we have*

$$\mathbb{E}(X^2) \geq (1 + o(1))6 \ .$$

The proof of this conjecture relies on showing that either most pairs are ε–good or the contribution of the positive error terms is larger than the contribution of the negative ones. On the other hand, notice that $\mathbb{E}(X^2)$ is not bounded from above by any constant. For the set of speeds in (7.1), Cilleruelo showed [38] that

$$\mathbb{E}(X^2) = (1 + o(1))\frac{12}{\pi^2}\delta n \log n \ , \tag{7.11}$$

which is a $\Theta(\log n)$ factor away from the lower bound in Proposition 7.3, when $\delta = \frac{1}{n+1}$. It is an open question whether (7.11) also holds as an upper bound for $\mathbb{E}(X^2)$.

3. Ideally, we would like to estimate the probabilities $\Pr(\cap_{i\in S} A_i)$, for every set $S \subseteq [n]$. In general, it is not easy to compute such probability. As in (7.11), the join probabilities cannot be upper bounded by any constant. However it is reasonable to think, that, for any set S of size s, we have

$$\Pr(\cap_{i\in S} A_i) \geq c_s \delta^s \ ,$$

where c_s depends only on s. Moreover, we know that $c_s \leq 2^s$, since this is the case when the speeds $\{v_i\}_{i\in S}$ are rationally independent. Observation 7.10 shows that $c_2 = 2$.

Bibliography

[1] F. A. Aagesen, C. Anutariya, and V. Wuwongse (eds.), *Intelligence in communication systems, ifip international conference, intellcomm 2004, bangkok, thailand, november 23-26, 2004, proceedings*, Lecture Notes in Computer Science, vol. 3283, Springer, 2004.

[2] R. Aharoni, P. Charbit, and D. Howard, *On a Generalization of the Ryser-Brualdi-Stein Conjecture*, ArXiv e-prints (2013).

[3] N. Alon, *A note on network reliability*, Discrete probability and algorithms (Minneapolis, MN, 1993), IMA Vol. Math. Appl., vol. 72, Springer, New York, 1995, pp. 11–14.

[4] ———, *The chromatic number of random cayley graphs*, to appear in European J. Combin. (2013).

[5] N. Alon and E. Friedgut, *On the number of permutations avoiding a given pattern*, J. Combin. Theory Ser. A **89** (2000), 133–140.

[6] N. Alon, T. Jiang, Z. Miller, and D. Pritikin, *Properly colored subgraphs and rainbow subgraphs in edge-colorings with local constraints*, Random Structures Algorithms **23** (2003), no. 4, 409–433.

[7] N. Alon and I. Z. Ruzsa, *Non-averaging subsets and non-vanishing transversals*, J. Combin. Theory Ser. A **86** (1999), no. 1, 1–13.

[8] N. Alon and J. H. Spencer, *The probabilistic method*, third ed., Wiley-Interscience Series in Discrete Mathematics and Optimization, John Wiley & Sons Inc., Hoboken, NJ, 2008.

[9] D. Angluin and L. G. Valiant, *Fast probabilistic algorithms for hamiltonian circuits and matchings*, Proceedings of the ninth annual ACM symposium on Theory of computing (New York, NY, USA), STOC '77, ACM, 1977, pp. 30–41.

[10] B. Arsovski, *A proof of Snevily's conjecture*, Israel J. Math. **182** (2011), 505–508.

[11] D. Auger, I. Charon, O. Hudry, and A. Lobstein, *Maximum size of a minimum watching system and the graphs achieving the bound*, to appear in Discrete Appl. Math. (2013).

[12] ———, *Watching systems in graphs: An extension of identifying codes*, Discrete Appl. Math. **161** (2013), no. 12, 1674–1685.

[13] J. Barajas and O. Serra, *The lonely runner with seven runners*, Electron. J. Combin. **15** (2008), no. 1, Paper 48, 18 pp.

[14] J. Beck, *An algorithmic approach to the Lovász local lemma. I*, Random Structures Algorithms **2** (1991), no. 4, 343–365.

[15] E. A. Bender and E. R. Canfield, *The asymptotic number of labeled graphs with given degree sequences*, J. Combin. Theory Ser. A **24** (1978), no. 3, 296 – 307.

[16] N. Bertrand, *Codes identifiants et codes localisateurs-dominateurs sur certains graphes*, Master's thesis, ENST, Paris, France, June 2001.

[17] N. Bertrand, I. Charon, O. Hudry, and A. Lobstein, *1-identifying codes on trees*, Australas. J. Combin. **31** (2005), 21–35.

[18] U. Betke and J. M. Wills, *Untere Schranken für zwei diophantische Approximations-Funktionen*, Monatsh. Math. **76** (1972), 214–217.

[19] W. Bienia, L. Goddyn, P. Gvozdjak, A. Sebő, and M. Tarsi, *Flows, view obstructions, and the lonely runner*, J. Combin. Theory Ser. B **72** (1998), no. 1, 1–9.

[20] R. Bissacot, R. Fernández, A. Procacci, and B. Scoppola, *An improvement of the Lovász local lemma via cluster expansion*, Combin. Probab. Comput. **20** (2011), no. 5, 709–719.

[21] T. Bohman, R. Holzman, and D. Kleitman, *Six lonely runners*, Electron. J. Combin. **8** (2001), no. 2, Paper 3, 49 pp.

[22] B. Bollobás, *A probabilistic proof of an asymptotic formula for the number of labelled regular graphs*, European J. Combin. **1** (1980), 311–316.

[23] B. Bollobás, *Random graphs*, Academic Press Inc., London, 1985.

[24] B. Bollobas, D. Mitsche, and P. Pralat, *Metric dimension for random graphs*, ArXiv e-prints (2012).

[25] M. Bóna, *Exact and Asymptotic Enumeration of Permutations with Subsequence Conditions*, ProQuest LLC, Ann Arbor, MI, 1997, Thesis (Ph.D.)–Massachusetts Institute of Technology.

[26] ______, *Non-overlapping permutation patterns*, Pure Math. Appl. (PU.M.A.) **22** (2011), 99–105.

[27] ______, *Combinatorics of permutations*, second ed., Discrete Mathematics and its Applications (Boca Raton), CRC Press, Boca Raton, FL, 2012.

[28] J. A. Bondy, *Induced subsets*, J. Combin. Theory Ser. B **12** (1972), 201–202.

[29] K. M. J. Bontridder, B. V. Halldórsson, M. M. Halldórsson, C. A. J. Hurkens, J. K. Lenstra, R. Ravi, and L. Stougie, *Approximation algorithms for the test cover problem*, Math. Program. **98** (2003), no. 1-3, 477–491.

[30] S. Boucheron, G. Lugosi, and P. Massart, *Concentration inequalities: A nonasymptotic theory of independence.*, Oxford University Press, Oxford, 2013.

[31] M. Bousquet-Mélou, A. Claesson, M. Dukes, and S. Kitaev, $(2+2)$*-free posets, ascent sequences and pattern avoiding permutations*, J. Combin. Theory Ser. A **117** (2010), no. 7, 884–909.

[32] N. J. Cavenagh, C. Greenhill, and I. M. Wanless, *The cycle structure of two rows in a random Latin square*, Random Structures Algorithms **33** (2008), no. 3, 286–309.

[33] I. Charon, I. Honkala, O. Hudry, and A. Lobstein, *Minimum sizes of identifying codes in graphs differing by one vertex*, Cryptogr. Commun. **5** (2013), no. 2, 119–136.

[34] I. Charon, O. Hudry, and A. Lobstein, *Extremal cardinalities for identifying and locating-dominating codes in graphs*, Discrete Math. **307** (2007), no. 3-5, 356–366.

[35] Y. G. Chen, *View-obstruction problems in n-dimensional Euclidean space and a generalization of them*, Acta Math. Sinica **37** (1994), no. 4, 551–562.

[36] Y. G. Chen and T. W. Cusick, *The view-obstruction problem for n-dimensional cubes*, J. Number Theory **74** (1999), no. 1, 126–133.

[37] F. Chung and P. Horn, *The spectral gap of a random subgraph of a graph*, Internet Math. **4** (2007), no. 2-3, 225–244.

[38] J. Cilleruelo, personal communication, 2012.

[39] C. Cooper, A. Frieze, B. Reed, and O. Riordan, *Random regular graphs of non-constant degree: independence and chromatic number*, Combin. Probab. Comput. **11** (2002), no. 4, 323–341.

[40] C. Cooper, R. Gilchrist, I. N. Kovalenko, and D. Novakovic, *Deriving the number of "good" permutations, with applications to cryptography*, Kibernet. Sistem. Anal. (1999), no. 5, 10–16, 187.

[41] T. W. Cusick, *View-obstruction problems. II*, Proc. Amer. Math. Soc. **84** (1982), no. 1, 25–28.

[42] T. W. Cusick and C. Pomerance, *View-obstruction problems. III*, J. Number Theory **19** (1984), no. 2, 131–139.

[43] S. Czerwiński, *Random runners are very lonely*, ArXiv e-prints (2011).

[44] S. Czerwiński and J. Grytczuk, *Invisible runners in finite fields*, Inform. Process. Lett. **108** (2008), no. 2, 64–67.

[45] R. Diestel, *Graph theory, 4th edition*, Graduate texts in mathematics, vol. 173, Springer, 2012.

[46] A. Dubickas, *The lonely runner problem for many runners*, Glas. Mat. Ser. III **46(66)** (2011), no. 1, 25–30.

[47] R. Ehrenborg, S. Kitaev, and P. Perry, *A spectral approach to consecutive pattern-avoiding permutations*, J. Comb. **2** (2011), 305–353.

[48] S. Elizalde, *Asymptotic enumeration of permutations avoiding generalized patterns*, Adv. in Appl. Math. **36** (2006), 138–155.

[49] ———, *The most and the least avoided consecutive patterns*, to appear in Proc. Lond. Math. Soc. (3) (2012).

[50] S. Elizalde and M. Noy, *Consecutive patterns in permutations*, Adv. in Appl. Math. **30** (2003), 110–125.

[51] ———, *Clusters, generating functions and asymptotics for consecutive patterns in permutations*, Adv. in Appl. Math. **49** (2012), 351 – 374.

[52] P. Erdős and L. Lovász, *Problems and results on 3-chromatic hypergraphs and some related questions*, Infinite and finite sets (Colloq., Keszthely, 1973; dedicated to P. Erdős on his 60th birthday), Vol. II, North-Holland, Amsterdam, 1975, pp. 609–627. Colloq. Math. Soc. János Bolyai, Vol. 10.

[53] P. Erdős and J. H. Spencer, *Lopsided Lovász local lemma and Latin transversals*, Discrete Appl. Math. **30** (1991), no. 2-3, 151–154, ARIDAM III (New Brunswick, NJ, 1988).

[54] Paul Erdős and Arthur H Stone, *On the structure of linear graphs*, Bull. Amer. Math. Soc **52** (1946), 1087–1091.

[55] L. Esperet, S. Gravier, M. Montassier, P. Ochem, and A. Parreau, *Locally identifying coloring of graphs*, Electron. J. Combin. **19** (2012), no. 2, Paper 40, 21 pp.

[56] L. Esperet and A. Parreau, *Acyclic edge-coloring using entropy compression*, European J. Combin. **34** (2013), no. 6, 1019–1027.

[57] J. F. Fink and M. S. Jacobson, *n-domination in graphs*, Graph theory with applications to algorithms and computer science (Kalamazoo, Mich., 1984), Wiley-Intersci. Publ., Wiley, New York, 1985, pp. 283–300.

[58] F. Foucaud, *Identifying codes in special graph classes*, Master's thesis, Université Bordeaux 1, Bordeaux, France, June 2009.

[59] F. Foucaud, E. Guerrini, M. Kovše, R. Naserasr, A. Parreau, and P. Valicov, *Extremal graphs for the identifying code problem*, European J. Combin. **32** (2011), no. 4, 628–638.

[60] F. Foucaud, I. Honkala, T. Laihonen, A. Parreau, and G. Perarnau, *Locally identifying colourings for graphs with given maximum degree*, Discrete Math. **312** (2012), no. 10, 1832–1837.

[61] F. Foucaud, R. Klasing, A. Kosowski, and A. Raspaud, *On the size of identifying codes in triangle-free graphs*, Discrete Appl. Math. **160** (2012), no. 10-11, 1532–1546.

[62] F. Foucaud and G. Perarnau, *Bounds for identifying codes in terms of degree parameters*, Electron. J. Combin. **19** (2012), no. 1, Paper 32, 28 pp.

[63] F. Foucaud, G. Perarnau, and O. Serra, *Random subgraphs make identification affordable*, ArXiv e-prints (2013).

[64] A. Frieze and M. Krivelevich, *On rainbow trees and cycles*, Electron. J. Combin. **15** (2008), no. 1, Paper 59, 9 pp.

[65] A. Frieze and M. Krivelevich, *On the non-planarity of a random subgraph*, ArXiv e-prints (2012).

[66] A. Frieze, R. Martin, J. Moncel, M. Ruszinkó, and C. Smyth, *Codes identifying sets of vertices in random networks*, Discrete Math. **307** (2007), no. 9-10, 1094–1107.

[67] Z. Füredi and P. Hajnal, *Davenport-Schinzel theory of matrices*, Discrete Math. **103** (1992), no. 3, 233–251.

[68] J. Galambos and I. Simonelli, *Bonferroni-type inequalities with applications*, Probability and its Applications (New York), Springer-Verlag, New York, 1996.

[69] H. Gebauer, T. Szabó, and G. Tardos, *The local lemma is tight for SAT*, Proceedings of the Twenty-Second Annual ACM-SIAM Symposium on Discrete Algorithms (Philadelphia, PA), SIAM, 2011, pp. 664–674.

[70] L. Goddyn and E. B. Wong, *Tight instances of the lonely runner*, Integers **6** (2006), Paper A38, 14pp.

[71] I. P. Goulden and D. M. Jackson, *Combinatorial Enumeration*, A Wiley-Interscience Publication, John Wiley & Sons Inc., New York, 1983, Wiley-Interscience Series in Discrete Mathematics.

[72] S. Gravier, M. Kovse, M. Mollard, J. Moncel, and A. Parreau, *New results on variants of covering codes in Sierpinski graphs*, ArXiv e-prints (2012).

[73] S. Gravier and J. Moncel, *On graphs having* $V\backslash\{x\}$ *as an identifying code*, Discrete Math. **307** (2007), no. 3-5, 432–434.

[74] S. Gravier, J. Moncel, and A. Semri, *Identifying codes of cycles*, European J. Combin. **27** (2006), no. 5, 767–776.

[75] J. Grytczuk, J. Kozik, and M. Witkowski, *Nonrepetitive sequences on arithmetic progressions*, Electron. J. Combin. **18** (2011), no. 1, Paper 209, 9 pp.

[76] A. Harutyunyan, P. Horn, and J. Verstraete, *Independent dominating sets in graphs of girth five*, to appear in Combin. Probab. Comput. (2009).

[77] P. Hatami and P. W. Shor, *A lower bound for the length of a partial transversal in a Latin square*, J. Combin. Theory Ser. A **115** (2008), no. 7, 1103–1113.

[78] T. W. Haynes, D. J. Knisley, E. Seier, and Y. Zou, *A quantitative analysis of secondary rna structure using domination based parameters on trees*, BMC Bioinformatics **7** (2006), 108.

[79] W. Hoeffding, *Probability inequalities for sums of bounded random variables*, J. Amer. Statist. Assoc. **58** (1963), 13–30.

[80] M. T. Jacobson and P. Matthews, *Generating uniformly distributed random Latin squares*, J. Combin. Des. **4** (1996), no. 6, 405–437.

[81] S. Janson, *Poisson approximation for large deviations*, Random Structures Algorithms **1** (1990), no. 2, 221–229.

[82] ______, *New versions of Suen's correlation inequality*, Random Structures Algorithms **13** (1998), no. 3-4, 467–483.

[83] S. Janson, T. Łuczak, and A. Rucinski, *Random graphs*, Wiley-Interscience Series in Discrete Mathematics and Optimization, Wiley-Interscience, New York, 2000.

[84] S. Janson and N. Wormald, *Rainbow Hamilton cycles in random regular graphs*, Random Structures Algorithms **30** (2007), no. 1-2, 35–49.

[85] M. Kano and X. Li, *Monochromatic and heterochromatic subgraphs in edge-colored graphs—a survey*, Graphs Combin. **24** (2008), no. 4, 237–263.

[86] M. G. Karpovsky, K. Chakrabarty, and L. B. Levitin, *On a new class of codes for identifying vertices in graphs*, IEEE Trans. Inform. Theory **44** (1998), no. 2, 599–611.

[87] J. H. Kim, O. Pikhurko, J. H. Spencer, and O. Verbitsky, *How complex are random graphs in first order logic?*, Random Structures Algorithms **26** (2005), no. 1-2, 119–145.

[88] S. Kitaev, *Patterns in permutations and words*, Monographs in Theoretical Computer Science. An EATCS Series, Springer, Heidelberg, 2011.

[89] S. Klavžar and U. Milutinović, *Graphs $S(n,k)$ and a variant of the Tower of Hanoi problem*, Czechoslovak Math. J. **47(122)** (1997), no. 1, 95–104.

[90] D. E. Knuth, *The Art of Computer Programming. Volume 3*, Addison-Wesley Publishing Co., Reading, Mass.-London-Don Mills, Ont., 1973, Sorting and searching.

[91] M. Krivelevich, C. Lee, and B. Sudakov, *Long paths and cycles in random subgraphs of graphs with large minimum degree*, ArXiv e-prints (2012).

[92] M. Krivelevich and B. Sudakov, *The phase transition in random graphs - a simple proof*, ArXiv e-prints (2012).

[93] M. Krivelevich, B. Sudakov, V.H. Vu, and N.C. Wormald, *Random regular graphs of high degree*, Random Structures Algorithms **18** (2001), no. 4, 346–363.

[94] M. Laifenfeld, A. Trachtenberg, R. Cohen, and D. Starobinski, *Joint monitoring and routing in wireless sensor networks using robust identifying codes*, MONET **14** (2009), no. 4, 415–432.

[95] T. D. LeSaulnier, C. Stocker, P. S. Wenger, and D. B. West, *Rainbow matching in edge-colored graphs*, Electron. J. Combin. **17** (2010), no. 1, Paper 26, 5 pp.

[96] A. Lobstein, *Watching systems, identifying, locating-dominating and discriminating codes in graphs: a bibliography*, `http://www.infres.enst.fr/~lobstein/debutBIBidetlocdom.pdf`.

[97] L. Lu and L. A. Szekely, *Using Lovász Local Lemma in the Space of Random Injections*, Electron. J. Combin. **14** (2007), no. 1, Paper 63, 13 pp.

[98] ———, *A new asymptotic enumeration technique: the Lovász Local Lemma*, ArXiv e-prints (2009).

[99] A. Marcus and G. Tardos, *Excluded permutation matrices and the Stanley-Wilf conjecture*, J. Combin. Theory Ser. A **107** (2004), 153–160.

[100] B. D. McKay, J. C. McLeod, and I. M. Wanless, *The number of transversals in a Latin square*, Des. Codes Cryptogr. **40** (2006), no. 3, 269–284.

[101] B. D. McKay and I. M. Wanless, *Most Latin squares have many subsquares*, J. Combin. Theory Ser. A **86** (1999), no. 2, 322–347.

[102] B. D. McKay and N. C. Wormald, *Asymptotic enumeration by degree sequence of graphs with degrees $o(n^{1/2})$*, Combinatorica **11** (1991), no. 4, 369–382.

[103] ———, *Uniform generation of random Latin rectangles*, J. Combin. Math. Combin. Comput. **9** (1991), 179–186.

[104] M. Molloy and B. Reed, *Further algorithmic aspects of the local lemma*, STOC '98 (Dallas, TX), ACM, New York, 1999, pp. 524–529.

[105] ———, *Graph colouring and the probabilistic method*, Algorithms and Combinatorics, vol. 23, Springer-Verlag, Berlin, 2002.

[106] J. Moncel, *On graphs on n vertices having an identifying code of cardinality $\lceil \log_2(n+1) \rceil$*, Discrete Appl. Math. **154** (2006), no. 14, 2032–2039.

[107] B. M. E. Moret and H. D. Shapiro, *On minimizing a set of tests*, SIAM J. Sci. Statist. Comput. **6** (1985), no. 4, 983–1003.

[108] Robin A. Moser, *A constructive proof of the general Lovász Local Lemma*, Proceedings of the 41st annual ACM symposium on Theory of computing (2009).

[109] Robin A. Moser and Gábor Tardos, *A constructive proof of the general Lovász Local Lemma*, J. ACM **57** (2010), no. 2, Article 11, 15 pp.

[110] T. Müller and J. S. Sereni, *Identifying and locating-dominating codes in (random) geometric networks*, Combin. Probab. Comput. **18** (2009), no. 6, 925–952.

[111] B. Nakamura, *Computational approaches to consecutive pattern avoidance in permutations*, Pure Math. Appl. (PU.M.A.) **22** (2011), 253–268.

[112] G. Perarnau, *A probabilistic approach to consecutive pattern avoiding in permutations*, J. Combin. Theory Ser. A **120** (2013), no. 5, 998–1011.

[113] G. Perarnau and G. Petridis, *Matchings in random biregular bipartite graphs*, Electron. J. Combin. **20** (2013), no. 1, Paper 60, 30 pp.

[114] G. Perarnau and O. Serra, *Correlation among runners and some results on the Lonely Runner Conjecture*, avaiable at `http://www-ma4.upc.edu/~guillem.perarnau/papers/lonely_new.pdf` (2013).

[115] ———, *Rainbow matchings in complete bipartite graphs existence and counting*, to appear in Combin. Probab. Comput. (2013).

[116] Y. Peres and W. Schlag, *Two Erdős problems on lacunary sequences: chromatic number and Diophantine approximation*, Bull. Lond. Math. Soc. **42** (2010), no. 2, 295–300.

[117] D. F. Rall and P. J. Slater, *On location-domination numbers for certain classes of graphs*, Proceedings of the fifteenth Southeastern conference on combinatorics, graph theory and computing (Baton Rouge, La., 1984), vol. 45, 1984, pp. 97–106.

[118] H. J. Ryser, *Neuere probleme der kombinatorik*, Vorträgeüber Kombinatorik, Oberwolfach (1967).

[119] J. B. Shearer, *On a problem of Spencer*, Combinatorica **5** (1985), no. 3, 241–245.

[120] H. S. Snevily, *Unsolved Problems: The Cayley Addition Table of Z_n*, Amer. Math. Monthly **106** (1999), no. 6, 584–585.

[121] S. K. Stein, *Transversals of Latin squares and their generalizations*, Pacific J. Math. **59** (1975), no. 2, 567–575.

[122] B. Sudakov and V. H. Vu, *Local resilience of graphs*, Random Structures Algorithms **33** (2008), no. 4, 409–433.

[123] W. C. S. Suen, *A correlation inequality and a poisson limit theorem for nonoverlapping balanced subgraphs of a random graph*, Random Structures Algorithms **1** (1990), no. 2, 231–242.

[124] S. Thomassé and A. Yeo, *Total domination of graphs and small transversals of hypergraphs*, Combinatorica **27** (2007), no. 4, 473–487.

[125] I. Vardi, *Computational recreations in Mathematica*, Addison-Wesley Publishing Company Advanced Book Program, Redwood City, CA, 1991.

[126] I. M. Wanless, *Transversals in Latin squares*, Quasigroups Related Systems **15** (2007), no. 1, 169–190.

[127] J. M. Wills, *Zwei Sätze über inhomogene diophantische Approximation von Irrationalzahlen*, Monatsh. Math. **71** (1967), 263–269.

[128] N. C. Wormald, *The asymptotic distribution of short cycles in random regular graphs*, J. Combin. Theory Ser. B **31** (1981), no. 2, 168–182.

[129] N. C. Wormald, *Models of random regular graphs*, Surveys in combinatorics, 1999 (Canterbury), London Math. Soc. Lecture Note Ser., vol. 267, Cambridge Univ. Press, Cambridge, 1999, pp. 239–298.