Università degli Studi di Padova

Università degli studi di Padova
Dipartimento di Matematica "Tullio Levi-Civita"
Corso di laurea in Matematica

Tesi di Laurea

Applications of the Probabilistic Method in Graph Theory

Relatore:
Prof. Manuel F. Aprile

Laureanda:
Antonella Doria
Matricola 2003943

25 luglio 2025
Anno Accademico 2024/2025

Abstract

The probabilistic method is a powerful, non-constructive tool with applications in Combinatorics and Discrete Mathematics. It is used for proving the existence of structures with certain properties, without actually constructing them. It follows this structure: in order to prove the existence of a specific object, one chooses an appropriate probability space and estimates the probability that one of its objects has the wanted property. Then, if this probability is greater than 0, one concludes that such a structure exists. In Graph Theory, this method can be applied to a variety of different problems. This thesis aims to show two of these applications. The first one will prove the existence of tournaments with many Hamiltonian paths, whereas the second will concern the Lovász Local Lemma and the non-repetitive colouring of the edges of graphs.

Contents

Introduction

The probabilistic method is a powerful technique for proving the existence of combinatorial objects with desired properties, without requiring their explicit construction. Although based on Probability Theory, it can be used for proving theorems that have nothing to do with probability. Indeed, the reason for its rapid development is that, by applying elementary notions of Probability Theory, it makes accessible the solution to many problems otherwise out of reach.

The applications of this technique in Discrete Mathematics are many, and can be classified in two groups. The first is the study of random combinatorial objects, such as random graphs or random matrices. This area mainly involves results in Probability Theory, although many results are motivated by problems in Combinatorics [1]. Basically, a random graph $G(n, p)$ is a probability space over the set of graphs on the vertex set $\{1, 2, \dots, n\}$ determined by

$$\mathbb{P}[(i, j) \;\text{ is an edge of } G] = p.$$

That is, the edge connecting the vertices i and j is present in G with independent probability p [2]. For instance, the random graph $G(10, 1/2)$ is the probability space over all graphs G whose vertex set is $V(G) = \{1, 2, \dots, 10\}$, and for each couple of vertices i and j, a fair coin is thrown to choose whether the edge (i, j) is present in G. In particular, the subject of random graphs began in 1960 with the paper *On the Evolution of Random Graphs* by Paul Erdős and Alfred Rényi who proved that for many monotone increasing properties of graphs, i.e. properties that are preserved under the addition of edges, there is a *sharp threshold* $f(n)$. That is, graphs with n vertices with slightly less edges than a function $f(n)$ are very unlikely to have the property, whereas graphs with slightly more edges than $f(n)$ almost certainly have the property. The model of random graphs is often used in the probabilistic method for proving the existence of specific graphs, although it has developed into an independent branch of Discrete Mathematics, with many applications to Theoretical Computer Science as well as Discrete Mathematics itself [2].

The aim of this work is to further study the second application of the probabilistic method: proving the existence of objects with desired properties by applying probabilistic tools in proofs of deterministic statements. In its basic form, it consists of creating a probability space whose points are combinatorial objects. Then, by showing that the probability of a random object having a desired property is positive, it is possible to conclude that at least one object with that property exists.

It is a non-constructive method, i.e. it does not give an explicit construction of those sought objects. For instance, in Graph Theory a desired property could be having many different paths that touch all the vertices of a graph only once. By applying the probabilistic method, it is possible to prove that a graph with many paths exists, although a standard constructive method could fail at the attempt of finding it.

Before the discovery of the probabilistic method, graph theory was studied mostly using explicit constructions and counting arguments. Although powerful in some cases, the progress made was fairly limited. As stated above, this technique was developed around the late 1950s, mostly by Paul Erdős and Alfred Rényi, although the first application is considered to be a theorem by Tibor Szele, on the maximum number of Hamiltonian paths in tournaments in 1943, which will be the main focus of Chapter 2. This new technique paved the way for many discoveries concerning different problems of graph theory.

Notice that, since most of the probability spaces considered in the study of combinatorial problems are finite, it could be argued that probability is not essential in these proofs. Theoretically, this is the case, as all the finite objects of the probability space could be searched until the right one is found. However, in practice it would be impossible to replace the applications of the many tools of the probabilistic method by counting arguments; moreover, the number of possibilities could be exponential in size, making it impossible to go through all of them in order to find the right one [2].

For instance, suppose there is a party and randomly pick six people from the crowd. Then surely three of them know each other or three of them are mutual strangers. This is easily proved: denote by A, B, C, D, E, F six people randomly selected in a crowd. Then, if A does not know three people, the assumption is proven. Suppose now that A knows B, C and D. If B, C, D are strangers, the assumption is proven, whereas if two of them know each other, say B and C, then A, B and C are a group of three people who know each other and the proof is complete. The question is: "what is the minimum number of people that need to be picked in order to be sure that k people know each other or l people are strangers?". The search for this number, called the *Ramsey number* $R(k, l)$, is an important topic Graph Theory, and the example above proved that $R(3, 3) = 6$. It was proved by constructive methods that $R(4, 4) = 18$, and that $R(5, 5)$ is between 43 and 49, but it has not been found yet. For $k \geq 6$, constructive methods cannot give useful bounds for the Ramsey number. In 1947 Paul Erdős proved by probabilistic method that, for any $k \geq 3$

$$R(k, k) \geq 2^{k/2-1}$$

i.e. in a party of at least $2^{k/2-1}$ individuals, there are k who know each other or k who are mutual strangers. Many consider this theorem the inauguration of the probabilistic method for the elegance of its proof, which appears in almost every text on the probabilistic method (for instance [2, 3, 4]). Still, there is no known explicit construction that could prove, or improve, the lower bound obtained by Erdős.

On the other hand, sometimes it is easier to prove that the probability that none of the objects have "bad" properties is less than unity, hence concluding that, with positive probability, some “good” object exists [5]. This tool was discovered by Paul Erdős and László Lovász in 1975 and published in *Problems and results on 3-chromatic Hypergraphs and some related questions*. In this paper, they introduced for the first time the *Lovász Local Lemma*, which states that given some events such that most of them are independent, the probability of none of them happening is positive, although possibly very small. This lemma and its applications will be the focus of Chapter 3.

The thesis is structured as follows:

Chapter 1: Preliminary Notions Some preliminary notions of graph theory and probability theory are introduced. Some definitions specific to each section will be introduced later. The last section briefly gives the definition of random graphs. Even though this thesis does not specifically explore this topic, it is a useful tool in all the probabilistic proofs that will be shown in the next chapters.

Chapter 2: Maximum Number of Hamiltonian Paths in Tournaments In this chapter a theorem by Szele is discussed and proved: it gives a lower bound of the maximum number of Hamiltonian paths in tournaments. It is considered the first application of the probabilistic method, and its simple probabilistic proof is ideal for showing how this method is implemented. The second section aims to show an improvement on this lower bound due to Adler, Alon and Ross. The proof has loosely the same structure as the first one, although, in order to guarantee a better result, it requires a clever use of known probabilistic tools. The last section will give a brief discussion on the upper bound.

Chapter 3: Lovász Local Lemma In this chapter, a powerful tool for the probabilistic method is discussed: the Lovász Local Lemma. It proves that the probability of none of possible “bad” events happening is positive if the events are almost independent. Three similar versions are discussed and proved, each useful in a different setting. In the second section, the problem of the non-repetitive colouring is discussed, which can be solved using the lemma described above. This application requires the knowledge of sequences and non-repetitive sequences, which will be introduced at the beginning, as well as an important theorem by Thue regarding the non-repetitive colouring of sequences. The last section will give a couple of examples of known constructive proofs of this type of colouring for two different classes of graphs, showing that, in some cases, the constructive proof actually gives the best bound.

Chapter 1

Preliminary Notions

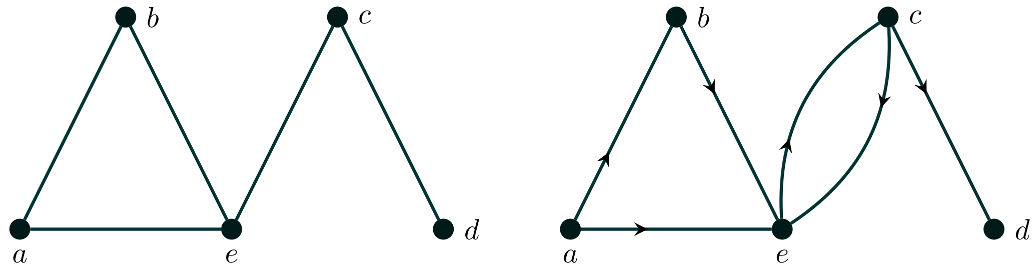
1.1 Graph Theory

In this section, some definitions of graph theory are recalled, and the notation that will be used in the following chapters is fixed. The definitions and notation are taken from [6]. Some definitions relevant only to a specific chapter or section will be introduced later, when necessary.

A *graph* is a couple $G = (V(G),\ E(G))$ such that $V(G)$ is a non-empty finite set, called the *vertex set*, and $E(G)$ is a finite set of unordered pairs of elements of $V(G)$, called the *edge set*. The elements of the vertex set are called *vertices*, or *nodes*, while the pairs of the edge set are called *edges*, with the condition that each edge connects distinct elements of $V(G)$, i.e. $u \neq v$ for any $(u, v) \in E(G)$. When clear from context, the dependency on G is omitted and the notation $G = (V,\ E)$ is adopted. This definition can be extended to *general graphs*: a general graph is one that can contain *loops*, i.e. edges joining a vertex to itself, and multiple edges joining the same two vertices.

An edge $(u, v) = uv \in E(G)$ is said to *join* the vertices u and v, and the vertices u and v are said to be *adjacent*. Additionally, v and u are *incident* with such edge. Similarly, two edges are *adjacent* if they have a vertex in common. The *degree* of a vertex v, $d(v)$, is the number of edges that are connected to v. The *degree of a graph* $G = (V,\ E)$ is $\Delta(G) = \max_{v \in V} d(v)$. For instance, the graph in Figure 1.1(a) has degree $\Delta = 3$.

A *directed graph*, or *digraph*, $D = (V,\ E)$ is a graph in which each edge $(u, v) \in E$ has an orientation. It is important to note that the edge (u, v) is different from the edge (v, u), whereas in undirected graphs this distinction does not appear. This definition can intuitively be extended to *general digraphs* by including parallel edges and loops. For any $v \in V$, the *indegree* $d^-(v)$ is the number of edges going into v, while the *outdegree* $d^+(v)$ is the number of edges going out of v. The *degree* of v, $d(v)$, is the sum of its indegree and outdegree. For instance, in Figure 1.1(b), the indegree and outdegree of each of the vertices are: $d^-(a) = 0,\ d^+(a) = 2,\ d^-(b) = 1,\ d^+(b) = 1,\ d^-(c) = 1,\ d^+(c) = 2,\ d^-(d) = 1,\ d^+(d) = 0$ and $d^-(e) = 3,\ d^+(e) = 1$.



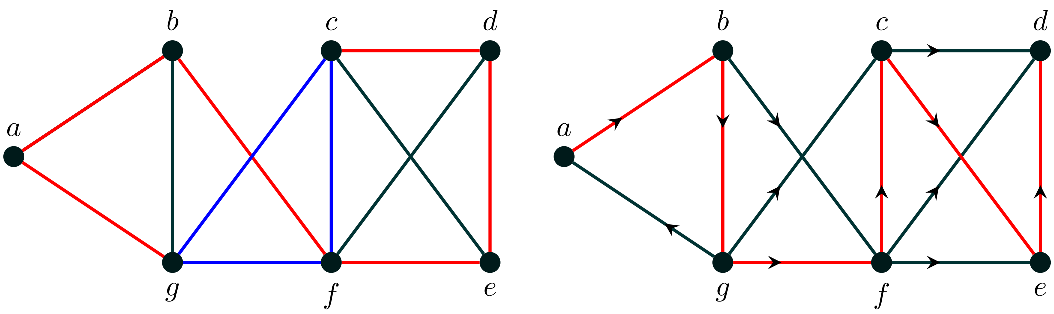
(a) Graph on 5 vertices. **(b)** Digraph on 5 vertices.

Figure 1.1

A *subgraph* of $G = (V,\ E)$ is a graph $G' = (V',\ E')$ such that $V' \subseteq V$ and $E' \subseteq E$. A *spanning subgraph* of $G = (V,\ E)$ is a graph $G' = (V',\ E')$ whose vertex set contains all the vertices of G, that is, $V' = V$. For instance, a subgraph of the graph in Figure 1.1(a) may be $(V',\ E')$ such that $V' = \{a,\ b,\ c,\ d\}$ and $E' = \{ab,\ cd\}$, whereas a spanning subgraph is any subgraph of the form $(V,\ E'')$ such that $V = \{a,\ b,\ c,\ d,\ e\}$ and $E'' \subseteq E = \{ab,\ ae,\ be,\ ec,\ cd\}$. These definitions hold for directed graphs as well.

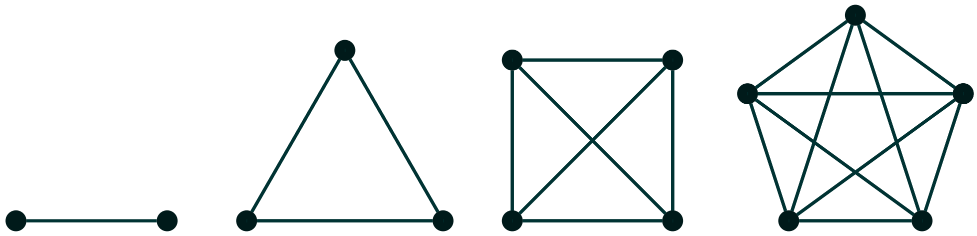
A *path from* v_0 *to* v_m in a graph G is a finite sequence of vertices $v_0 - v_1 - \cdots - v_m$, or $\{v_0,\ v_1, \ldots,\ v_m\}$, such that $(v_{i-1},\ v_i) \in E(G)$ for $i = 1, \ldots,\ m$ and $v_i \neq v_j$ for any $i \neq j$. v_0 is called the *initial vertex* and v_m the *final vertex*. If $v_0 = v_m$ and there is at least one edge, the path is called a *cycle*. The *length* of a path, or cycle, is the number of edges that it contains. A cycle of length 3 is called a *triangle*. For instance, the blue cycle $\{g, c, f, g\}$ in Figure 1.2(a) is a triangle. The same definitions hold for digraphs as well.

A *Hamiltonian path* in a graph G is a path that visits every vertex of G exactly once. A *Hamiltonian cycle* is a cycle that visits every vertex of G. A graph is *Hamiltonian* if it has a Hamiltonian cycle. Similarly, a digraph D is *Hamiltonian* if there is a cycle that contains every vertex of D. For instance, the graph and the digraph in Figure 1.2 both have a Hamiltonian path, shown in red.



(a) In red, a Hamitonian path in a graph. **(b)** In red, a Hamiltonian path in a digraph.

Figure 1.2



(a) K_2 **(b)** K_3 **(c)** K_4 **(d)** K_5

Figure 1.3: Complete graphs on 2, 3, 4, and 5 vertices.

A graph, or digraph, is said to be *connected* if it cannot be expressed as the union of two distinct graphs, or digraphs. Otherwise, it is *disconnected*. It is clear that a graph is connected if, for any two vertices, there is a path between them.

A *complete graph* is a graph $K_n = (V, E)$ whose n vertices are all adjacent, i.e. for any $u,\ v \in V$, $(u, v) \in E$. It is easy to show that

Lemma 1.1. *K_n has $\frac{n(n-1)}{2}$ edges.*

Proof. Indeed, given that an edge is a combination of two elements of $V(K_n)$, the number of edges in K_n is equal to all the possible combination of two elements in the set of the vertices $\{v_1,\ v_2, \ldots,\ v_n\}$. That is,

$$\binom{n}{2} = \frac{n!}{2(n-2)!} = \frac{n(n-1)}{2}.$$

□

A *tournament* $T = T_n$ is an orientation of a complete graph K_n, i.e. for any two vertices $u, v \in V(K_n)$, exactly one of the directed edges (u, v) or (v, u) is present in $E(T)$. Likewise, a tournament T is defined as a directed complete graph. For a fixed vertex set $V = \{v_1,\ v_2, \ldots,\ v_n\}$, there are $2^{\binom{n}{2}}$ possible tournaments with the same vertex set: indeed, for any of the $\binom{n}{2}$ edges there are two possible orientations, ergo $2^{\binom{n}{2}}$ possible tournaments.

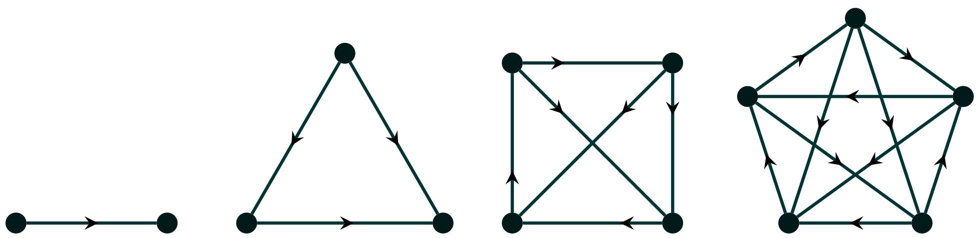


Figure 1.4: Tournaments on 2, 3, 4, and 5 vertices.

In this work, the focus is only on simple graphs and digraphs. It is worth mentioning that many results of simple graphs can be extended to general graphs, hence this assumption is not very restrictive.

1.2 Probability Spaces

In this section, some definitions and basic proofs of probability theory are introduced. The notation and the proofs of this section are taken from [3], unless otherwise specified.

Definition 1.1. Given a set Ω, a *σ-algebra* Σ on Ω is a non-empty collection of subsets of Ω such that $\Omega \in \Sigma$ and:

i) if $E \in \Sigma$, then the complementary event $\overline{E} = \Omega \setminus E \in \Sigma$;

ii) if $E_i \in \Sigma$ for $i = 1,\ 2, \dots$ then $\bigcup\limits_{i \geq 1} E_i \in \Sigma$.

Definition 1.2. Given a σ-algebra Σ, $\mathbb{P} :\ \Sigma\ \longrightarrow [0, 1]$ is a *probability measure* if $\mathbb{P}[\Omega] = 1$ and for any sequence $E_1,\ E_2, \dots \in \Sigma$ such that $E_i \cap E_j = \emptyset$ if $i \neq j$, then

$$\mathbb{P}\Big[\bigcup_{i \geq 1} E_i\Big] = \sum_{i=1} \mathbb{P}[E_i].$$

Definition 1.3. A *probability space* is a triplet $(\Omega,\ \Sigma,\ \mathbb{P})$ where $\Omega \neq \emptyset$ is the set of all possible outcomes and is called a *sample space*, Σ is a σ-algebra on Ω and $\mathbb{P} :\ \Sigma\ \rightarrow [0, 1]$ is a probability measure.

Any subset of Ω is called an *event*. The probability of an event is the measure of the likelihood that the event will occur. If $E_1,\ E_2, \dots$ are events in Ω, then the *union* of these events $\bigcup_{n=1}^{\infty} E_n$ is the event consisting of all outcomes that are in at least one E_n for $n = 1,\ 2, \dots$. The *intersection* of the events $E_1,\ E_2, \dots$ is the one consisting in the outcomes that are in all of the events $E_n,\ n = 1,\ 2, \dots$ and is denoted by $\bigcap_{n=1}^{\infty} E_n$. A sequence of events $E_1,\ E_2, \dots \in \Sigma$ such that $E_i \cap E_j = \emptyset$ if $i \neq j$ are said to be *mutually exclusive*.

Definition 1.4. A *real random variable* on a probability space $(\Omega,\ \Sigma,\ \mathbb{P})$ is a $\mathbb{P}$-measurable function $X : \Omega \longrightarrow \mathbb{R}$. That is, for any $a \in \mathbb{R}$, $\{\omega \in \Omega \mid X(\omega) \leq a\} \in \Sigma$. A random variable is *discrete* if its support Ω has either a finite or an infinite but countable number of elements.

Definition 1.5. The *expectation*, or *expected value*, of a random variable X is

$$\mathbb{E}[X] = \int_{\Omega} X(\omega) d\mathbb{P}.$$

The expectation of a discrete random variable is

$$\mathbb{E}[X] = \sum_{\omega \in \Omega} p(\omega)\ X(\omega) = \sum_{\omega \in \Omega} \mathbb{P}[X = \omega]\ X(\omega).$$

where $p(\omega) = \mathbb{P}[X = \omega]$.

Observation 1.1. Observe that, given a random variable X on a probability space $(\Omega, \Sigma, \mathbb{P})$,

$$\mathbb{E}[X(\omega)] \geq \min_{\omega\in\Omega} X(\omega).$$

Therefore, there exists a point $\omega \in \Omega$ such that $\mathbb{E}[X(\omega)] \geq X(\omega)$.

Definition 1.6. For an event A, the *indicator variable* I_A is

$$I_A(\omega) = \begin{cases} 1 & \text{if } \omega \in A \\ 0 & \text{if } \omega \notin A \end{cases}$$

Definition 1.7. Let A and B be two events, with $\mathbb{P}[B] > 0$. The *conditional probability* of A, given that B occurs, is

$$\mathbb{P}[A \mid B] = \frac{\mathbb{P}[A \cap B]}{\mathbb{P}[B]}.$$

Definition 1.8. Two real random variables $X,\ Y$ are said to be *independent* if, for any two measurable sets $A,\ B \subseteq \mathbb{R}$,

$$\mathbb{P}[X \in A \text{ and } Y \in B] = \mathbb{P}[X \in A]\mathbb{P}[Y \in B]\ ^1.$$

Two discrete random variables X and Y are *independent* if

$$\mathbb{P}[X = x \text{ and } Y = y] = \mathbb{P}[X = x]\ \mathbb{P}[\ Y = y].$$

Observe that if X and Y are independent random variables, then $\mathbb{P}[X \mid Y] = \mathbb{P}[X]$.

Lemma 1.2. *For any $X,\ Y$ variables and $\alpha,\ \beta \in \mathbb{R}$ constants:*

$$\mathbb{E}[\alpha X + \beta Y] = \alpha\mathbb{E}[X] + \beta\mathbb{E}[Y] \tag{1.1}$$

i.e. the expectation is a linear operator.

Proof. By definition

$$\mathbb{E}[\alpha X + \beta Y] = \int_\Omega (\alpha X + \beta Y) d\mathbb{P}.$$

By linearity of the integral, given that αX and βY are continuous functions in Ω

$$\int_\Omega (\alpha X + \beta Y) d\mathbb{P} = \alpha \int_\Omega X d\mathbb{P} + \beta \int_\Omega Y d\mathbb{P} = \alpha\mathbb{E}[X] + \beta\mathbb{E}[Y].$$

□

Lemma 1.3. *If X and Y are independent random variables, then*

$$\mathbb{E}[XY] = \mathbb{E}[X] \cdot \mathbb{E}[Y]. \tag{1.2}$$

[1] $\mathbb{P}[X \in A] = \mathbb{P}[\{\omega \in \Omega \mid X(\omega) \in A\}]$.

The following proof is given only for finite probability spaces. For infinite probability spaces, the proof is similar but more complicated.

Proof (for finite probability spaces). Let $V_X,\ V_Y$ be the finite sets of values attained by X and Y, respectively. Since X and Y are independent, then for any $a \in V_X$ and $b \in V_Y$, $\mathbb{P}[X = a \text{ and } Y = b] = \mathbb{P}[X = a]\ \mathbb{P}[Y = b]$. Therefore,

$$\begin{aligned}\mathbb{E}[XY] &= \sum_{a\in V_X,\ b\in V_Y} ab \cdot \mathbb{P}[X = a \text{ and } Y] \\ &= \sum_{a\in V_X,\ b\in V_Y} ab \cdot \mathbb{P}[X = a]\mathbb{P}[Y = B] \\ &= \Big(\sum_{a\in V_X} a \cdot \mathbb{P}[X = a]\Big)\Big(\sum_{b\in V_Y} b \cdot \mathbb{P}[Y = b]\Big) = \mathbb{E}[X]\mathbb{E}[Y].\end{aligned}$$

□

Lemma 1.4. *For any event A,*

$$\mathbb{E}[I_A] = \mathbb{P}[A]. \tag{1.3}$$

Proof. The proof is a simple computation,

$$\mathbb{E}[I_A] = \int_\Omega I_A(\omega)d\mathbb{P} = \int_A d\mathbb{P} = \mathbb{P}(A).$$

□

Proposition 1.5 (Inclusion-exclusion principle for two events, [7])**.** *For any two events $A,\ B \in \Sigma$, then*

$$\mathbb{P}[A \cup B] = \mathbb{P}[A] + \mathbb{P}[B] - \mathbb{P}[A \cap B].$$

Proof. Using the fact that A and $B - A$ are exclusive events and $A \cup B = A \cup (B - A)$

$$\mathbb{P}[A \cup B] = \mathbb{P}[A \cup (B - A)] = \mathbb{P}[A] + \mathbb{P}[B - A].$$

Then again, $B - A$ and $B \cap A$ are also exclusive events and $B = (B - A) \cup (B \cap A)$, so

$$\mathbb{P}[B] = \mathbb{P}[(B - A) \cup (B \cap A)] = \mathbb{P}[B - A] + \mathbb{P}[B \cap A].$$

□

Theorem 1.6 (Inclusion-exclusion principle, [7])**.** *Let $E_1,\ E_2, \ldots,\ E_n$ be any events, then*

$$\begin{aligned}\mathbb{P}[E_1 \cup \cdots \cup E_n] = \sum_{1\leq i\leq n} \mathbb{P}[E_1] - \sum_{1\leq i_1\leq i_2\leq n} \mathbb{P}[E_{i_1} \cap E_{i_2}] + \\ + \sum_{1\leq i_1\leq i_2\leq i_3\leq n} \mathbb{P}[E_{i_1} \cap E_{i_2} \cap E_{i_3}] - \cdots + (-1)^{n+1}\mathbb{P}[E_{i_1} \cap E_{i_2} \cap \cdots \cap E_n].\end{aligned} \tag{1.4}$$

Proposition 1.7 (Chain rule, [8]). *For $n \in \mathbb{N}$, $n \geq 2$, let $A_1, \dots, A_n$ be events such that $\mathbb{P}[A_1 \cap A_2 \cap \dots \cap A_n] > 0$. Then*

$$\begin{aligned} \mathbb{P}\Big[\bigcap_{i=1}^{n} A_i\Big] &= \mathbb{P}[A_1]\ \mathbb{P}[A_2 \mid A_1]\ \mathbb{P}[A_3 \mid A_1 \cap A_2] \cdots \mathbb{P}[A_n \mid A_1 \cap \dots \cap A_{n-1}] \\ &= \mathbb{P}[A_1] \prod_{i=2}^{n} \mathbb{P}[A_i \mid A_1 \cap \dots \cap A_{i-1}]. \end{aligned} \tag{1.5}$$

1.3 Randomly Generated Graphs

Definition 1.9 ([2]). A *random graph* $G(n, p)$ is a probability space on the set of vertices $V(G) = \{1, 2, \dots, n\}$ determined by

$$\mathbb{P}\big[(i, j) \text{ is an edge of } G\big] = p.$$

with these events mutually independent.

That is, each edge of the graph is present with probability p. If $p = 1/2$, that is the same as tossing a fair coin for each pair of vertices u, v and drawing the edge uv if the outcome is heads. The term *random graph* is a misnomer; in fact, $G(n, p)$ is a probability space over graphs [2].

Similarly, it is possible to define this probability space specifying the precise number of edges that must be present in the graph.

Definition 1.10 ([3]). The *probability space of random graphs* $G(n, m)$ is a finite probability space whose events are all graphs on a fixed set of n vertices. The *probability of a graph* with m edges is

$$p(G) = p^m (1 - p)^{\binom{n}{2} - m}. \tag{1.6}$$

The same definition holds for digraphs.

A *random tournament* is a tournament with a random orientation, i.e. each edge has a orientation chosen randomly and independently with probability $p = 1/2$.

Chapter 2

Maximum Number of Hamiltonian Paths in Tournaments

The study of Hamiltonian paths represents a fascinating problem in Graph Theory. Generally, proving the existence of such paths is notoriously difficult: most of the theorems known on this topic prove the existence of Hamiltonian paths only for specific classes of graphs [9]. Recall that a *Hamiltonian path* in a graph G is a path that visits every vertex of G exactly once, whereas a *Hamiltonian cycle* is a cycle that visits every vertex of G. A graph is *Hamiltonian* if it has a Hamiltonian cycle. These paths and cycles are useful tools in practical applications, such as planning of delivery routes and network design.

Among these special classes of graphs, tournaments – i.e. complete directed graphs – present a particularly rich structure, as every tournament always admits at least one Hamiltonian path. This property justifies the shift from a problem of existence to an enumeration problem, which leads to the question: "if every tournament contains at least one Hamiltonian path, what is the maximum number of Hamiltonian paths it can contain?"

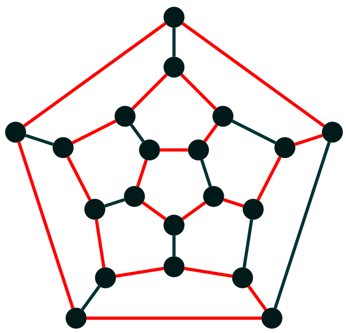


Figure 2.1: A Hamiltonian cycle (shown in red) in the graph of the dodecahedron.

As stated in the Introduction, the probabilistic method is particularly powerful in these type of studies, therefore its involvement in the search for an answer is only natural.

The aim of this chapter is to showcase an application of the probabilistic method by answering the question posed above. Note that the following theorems will only give an upper and lower bound to the maximum number of Hamiltonian paths in tournaments, and not a precise estimate of it. Besides, since these proofs are non-constructive, they do not give an explicit construction of these desired objects.

In order to develop the correct approach, it is best to introduce the following definition, given that, as stated in Observation 2.1, there is a strict link between permutations and Hamiltonian paths in complete graphs.

Definition 2.1 ([9])**.** A *permutation* of n distinct objects is an arrangement of those objects into an ordered line. The notation $\sigma(i)$ denotes the $i-$th element of the permutation $\sigma = \{\sigma(1), \ldots, \sigma(n)\}$.

Denote as S_n the set of all the permutations on the set $\{1, 2, \ldots, n\}$.

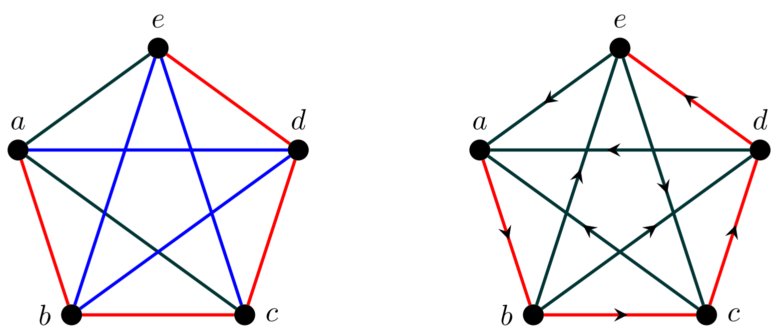
Lemma 2.1 ([9])**.** *There are* $n!$ *possible permutations in a set of* n *objects, i.e.*

$$|S_n| = n! = n(n-1)(n-2)\cdot\dots\cdot 2\cdot 1.$$

Proof. There are n choices for the first element $\sigma(1)$, and for each of these choices, there remain $n-1$ choices for $\sigma(2)$ and so on until there remains only 1 choice for the last element $\sigma(n)$.

□

Observation 2.1. In Chapter 1, a path from v_0 to v_m was defined as a sequence $\{v_0,\ v_1, \ldots,\ v_{m-1},\ v_m\}$ of adjacent vertices. Observe that, given a complete graph $K_n = (V,\ E)$ with $V = \{1,\ 2, \ldots,\ n\}$, a permutation on the set of vertices V is a Hamiltonian path in K_n. On the other hand, if $T = T_n$ is a tournament on n vertices, then a permutation $\sigma = \{\sigma(1), \ldots,\ \sigma(n)\}$ is a Hamiltonian path in T if each $(\sigma(i),\ \sigma(i+1))$, for $i = 1, \ldots,\ n-1$, is an edge of T, with the right orientation.



(a) A complete graph K_5. **(b)** A tournament T on K_5.

Figure 2.2: The permutation $\{a,\ b,\ c,\ d,\ e\}$, in red, is a Hamiltonian path both in K_5 and T, whereas the permutation $\{a,\ d,\ b,\ e,\ c\}$, in blue, is a Hamiltonian path in K_5 but not T.

2.1 The Lower Bound

The first proof on this topic was discovered by T. Szele in 1943 and published in the paper *Kombinatorikai vizsgalatok az iranyitott teljes graffal kapcsolatban* (*Combinatorial studies on complete directed graphs*). It is considered the first application of the probabilistic method, and his proof has the following structure: given a complete graph K_n, orient each edge with equal probability, and compute the probability of the event *"σ is a Hamiltonian path in the tournament"*. After that, by computing the average number of Hamiltonian paths in a tournament, the proof is completed, as it is always true that at least one tournament has more paths than average.

Theorem 2.2 (Szele, 1943). *For every $n \in \mathbb{N}$, there is a tournament on n vertices that has at least $\dfrac{n!}{2^{n-1}}$ Hamiltonian paths.*

That is, denote $P(T)$ as the number of Hamiltonian paths in a tournament T, and for $n \geq 2$, $P(n) = \max\{P(T) \mid T \text{ is a tournament with } n \text{ vertices}\}$ as the maximum number of directed Hamiltonian paths in a tournament on n vertices, then Theorem 2.2 proves that

$$P(n) \geq \frac{n!}{2^{n-1}}.$$

The following proof is taken from [3], although it is discussed in almost every paper on the topic of probabilistic method.

Proof. Let $T = (V,\ E)$ be a random tournament with vertex set $V = \{1,\ 2, \ldots,\ n\}$, and probability $p = 1/2$, i.e. given a complete graph K_n orient each edge $(u, v) \in E(K_n)$ randomly and independently with probability $p = 1/2$, either (u, v) or (v, u). For a given permutation $\sigma = \{\sigma(1),\ \sigma(2), \ldots,\ \sigma(n)\}$ on V, denote by X_σ the indicator random variable of the event that σ is a Hamiltonian path in T, i.e. for any $i = 1, \ldots, n-1$ the edge $(\sigma(i),\ \sigma(i+1)) \in E(T)$. Using Lemma 1.2 and Lemma 1.3, given that the orientation of each edge is chosen independently, then

$$\begin{aligned} \mathbb{E}[X_\sigma] = \mathbb{P}[X_\sigma] =& \mathbb{P}[(\sigma(i), \sigma(i+1)) \in T \text{ for } i = 1,\ 2, \ldots,\ n-1] \\ &= \prod_{i=1}^{n-1} \mathbb{P}[(\sigma(i), \sigma(i+1)) \in T] = \prod_{i=1}^{n-1} \frac{1}{2} = \frac{1}{2^{n-1}}. \end{aligned} \tag{2.1}$$

Now, let $X = \sum_{\sigma \in S_n} X_\sigma$ be the sum over all the permutations of $\{1,\ 2, \ldots,\ n\}$, i.e X is the random variable counting how many Hamiltonian paths there are in T. By Lemma 2.1, there are $n!$ permutations on V, therefore, the average number of Hamiltonian paths in a tournament is

$$\mathbb{E}[X] = E\Big[\sum_{\sigma \in S_n} X_\sigma\Big] = \sum_{\sigma \in S_n} \mathbb{E}[X_\sigma] = \frac{n!}{2^{n-1}}. \tag{2.2}$$

By Observation 1.1, there is at least a tournament T for which $X(T) \geq \mathbb{E}[X(T)]$. Therefore, there exists a tournament with at least $\frac{n!}{2^{n-1}}$ Hamiltonian paths. $\square$

2.2 Improvement of the Lower Bound

In this section, an improvement of the lower bound on P(n) is discussed. This result is a clever application of the probabilistic method based of the proof by Szele, i.e. on computing the average number of Hamiltonian paths in randomly generated tournaments. The main difference is that not every edge in T will be orientated randomly and independently from the others.

Unless otherwise specified, the theorems and proofs of this section are taken from [10].

Theorem 2.3 (Adler, Alon, Ross, 2001)**.** *For every $n \in \mathbb{N}$,*

$$P(n) \geq (e - o(1)) \frac{n!}{2^{n-1}} \tag{2.3}$$

where the term $o(1)$ tends to 0 as n tends to infinity.

First, an overview of the proof is given. In Step 1, we will consider random tournaments such that the edges along certain triangles are orientated as directed cycles, and we will compute the expectation of the random variable X that counts the number of Hamiltonian paths in each tournament. This gives the identity

$$\mathbb{E}[X] = n!/(2^{n-1})\ \mathbb{E}[2^{Y(\sigma)}]$$

where $Y(\sigma)$ is the random variable counting the number of triangles occurring in each permutation.

In Step 2, we will prove that for large n, $Y(\sigma)$ has almost a Poisson distribution. Recall that a *Poisson distribution*, also known as *distribution of rare events*, gives the probability of an event happening a certain number of times. That is, if a random variable X has a Poisson distribution, then

$$\mathbb{P}[X = k] = e^{-\lambda} \frac{\lambda^k}{k!}$$

for $k \in \mathbb{N}^0 = \{0, 1, 2, \dots\}$ and mean $\lambda = \mathbb{E}[X]$, i.e. the average number of times an event occurs.

This allows us to compute $\mathbb{E}[2^{Y(\sigma)}]$ in Step 3. Finally, in step 4, we will prove that for large n, in K_n there exists a collection of many pairwise edge-disjoint triangles. This ensures that the result found in Step 3 can be applied to the identity found in step 1.

Step 1 To improve $P(n)$, the first step is to consider an orientation of the edges of K_n generated by selecting some triangles in K_n and orienting their edges as directed cycles.

First, introduce the notation (u, v, k) for the triangle $u - v - k - u$ of a graph G. Two triangles are *edge-disjoint* if they do not share any edges. A triangle (u, v, k) is *present in a permutation* $\sigma = \{\sigma(1), \dots, \sigma(n)\}$ if u, v, k are three consecutive elements of σ.

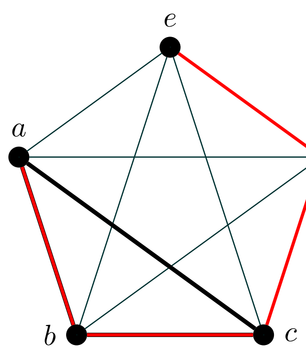


Figure 2.3: The triangle (a, b, c) is present in the permutation $\{a, b, c, d, e\}$, in red, which is a Hamiltonian path of K_5.

Given a complete graph K_n, let $\mathscr{T}$ be a collection of t pairwise edge-disjoint triangles in K_n, and consider a random tournament $T_n = (V, E)$ on K_n obtained by orienting each triangle $(i, j, k) \in \mathscr{T}$ randomly and independently with probability $p = 1/2$, either (i, j), (j, k), (k, i), as in Figure 2.4(a), or (i, k), (k, j), (j, i), as in Figure 2.4(b). Any other edge of T_n is orientated randomly and independently with probability $p = 1/2$.

(a) Orientation (i, j), (j, k), (k, i). **(b)** Orientation (i, k), (k, j), (j, i).

Figure 2.4: The two possible orientations for edge-disjoint triangles in $\mathscr{T}$ in the random tournament T_n.

Let $Y(\sigma)$ be the random variable that counts the number of triangles in $\mathscr{T}$ that are present in a permutation σ. Let X_σ be the indicator random variable of the event that σ is a Hamiltonian path in T_n, i.e. each $(\sigma(i), \sigma(i+1))$ is an edge of T_n with this orientation.

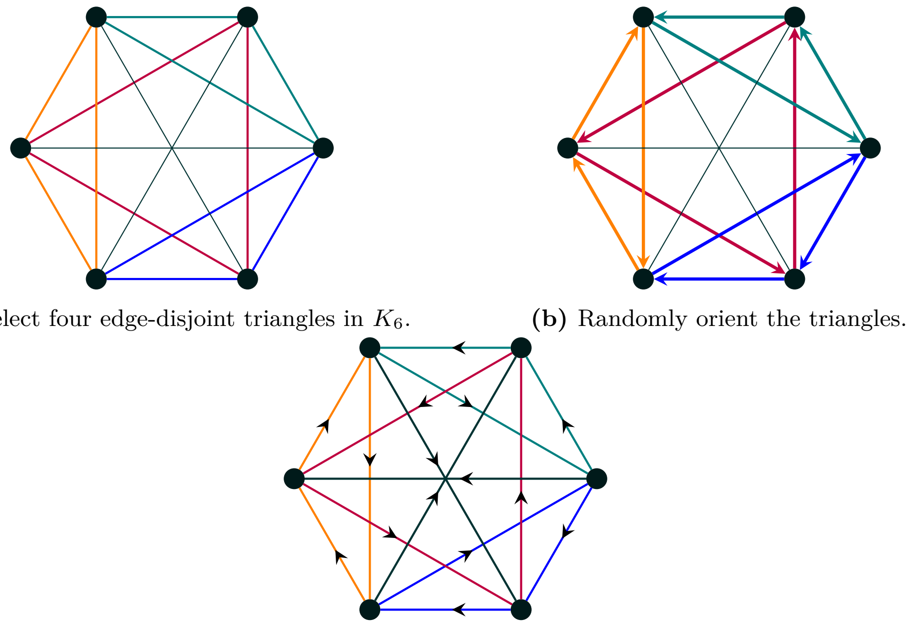
Considering that, if $(i, j, k) \in \mathscr{T}$, then $\mathbb{P}[(i, j) \in E(T_n) \text{ and } (j, k) \in E(T_n)] = \frac{1}{2}$, the expectation of X_σ is:

$$\begin{aligned} \mathbb{E}[X_\sigma] =& \mathbb{P}[X_\sigma] \\ =& \mathbb{P}[(\sigma(i),\ \sigma(i+1)) \in E(T_n) \text{ for } i = 1, 2, \ldots, n-1] \\ =& \frac{1}{2^{n-1}} 2^{Y(\sigma)}. \end{aligned} \tag{2.4}$$

If $Y(\sigma) = 0$, i.e. if there are no triangles in the permutation, then the probability that X_σ is a Hamiltonian path is exactly the same as the one in (2.1). Let $X =$

$\sum_{\sigma \in S_n} X_\sigma$ be the sum over all the permutations of $\{1, 2, \ldots, n\}$, i.e X is the random variable counting how many Hamiltonian paths there are in T_n. Then [1],

$$\begin{aligned} \mathbb{E}[X] = \sum_\sigma \mathbb{E}[X_\sigma] &= \sum_\sigma \frac{2^{Y(\sigma)}}{2^{n-1}} \\ &= \frac{1}{2^{n-1}} \sum_\sigma 2^{Y(\sigma)} \sum_k \mathbb{P}[Y(\sigma) = k] \\ &= \frac{1}{2^{n-1}} \sum_\sigma \sum_k 2^k \mathbb{P}[Y(\sigma) = k] \\ &= \frac{1}{2^{n-1}} \sum_\sigma \mathbb{E}[2^{Y(\sigma)}] = \frac{n!}{2^{n-1}} \mathbb{E}[2^{Y(\sigma)}]. \end{aligned} \tag{2.5}$$



(a) Select four edge-disjoint triangles in K_6.

(b) Randomly orient the triangles.

(c) Randomly orient the remaining edges.

Figure 2.5: An example on K_6 of the random orientation of its edges.

Step 2 Suppose now that $|\mathscr{T}| \geq \frac{n^2}{6} - O(n)$. In order to prove that Y is close to a Poisson distribution with mean 1 for large n, it is necessary to introduce the Poisson Paradigm and Brun's sieve, as discussed in [2].

Suppose that X is the sum of many rare indicator random variables that are "mostly independent", then the *Poisson Paradigm* states that X is close to a Poisson distribution with mean $\mu = \mathbb{E}[X]$ and in particular that $\mathbb{P}[X = 0]$ is nearly $e^{-\mu}$. A possible approach to solving this paradigm is *Brun's sieve*.

[1] Recall that $\sum_{\omega \in \Omega} \mathbb{P}[Y(\sigma) = \omega] = 1$.

Let $B_1, B_2, \ldots, B_m$ be events, and X_i the indicator variable for B_i. Let $X = X_1 + \cdots + X_m$ the number of events that occur and n be a hidden parameter, so that actually $m = m(n),\ B_i = B_i(n),\ X = X(n)$. Define the sum over all subsets $\{i_1, \ldots, i_r\} \subseteq \{1, \ldots, m\}$

$$S^{(r)} = \sum_{\{i_1,\ldots,i_r\}\subseteq\{1,\ldots,m\}} \mathbb{P}[B_{i_1} \cap \cdots \cap B_{i_r}].$$

and $S^{(m)} = \mathbb{P}[B_1 \cap \cdots \cap B_m]$. Notice that $\mathbb{E}[X] = S^{(1)}$ and $\mathbb{E}\left[\binom{X}{r}\right] = S^{(r)}$.

Then, by De Morgan's law [2] and Theorem 1.6

$$\begin{aligned}
\mathbb{P}[X=0] &= \mathbb{P}\left[\overline{B_1} \cap \cdots \cap \overline{B_m}\right] \\
&= \mathbb{P}\left[\overline{B_1 \cup \cdots \cup B_m}\right] \\
&= 1 - \mathbb{P}[B_1 \cup \cdots \cup B_m] \\
&= 1 - S^{(1)} + S^{(2)} - \cdots + (-)^r S^{(r)} \cdots + (-1)^m S^{(m)}.
\end{aligned}$$

Theorem 2.4 (Brun's Sieve, [4])**.** *Suppose that there is a constant μ such that $\mathbb{E}[X] = S^{(1)} \xrightarrow{n\to\infty} \mu$ and, for every fixed r, $E[\binom{X}{r}] = S^{(r)} \xrightarrow{n\to\infty} \mu^r/r!$. Then, for every t:*

$$\begin{aligned}
&\mathbb{P}[X=0] \xrightarrow{n\to\infty} e^{-\mu} \\
&\mathbb{P}[X=t] \xrightarrow{n\to\infty} \frac{\mu^t}{t!} e^{-\mu}.
\end{aligned}$$

The proof of this theorem is discussed in [2].

Now it is possible to prove that, if there are enough triangles in $\mathscr{T}$, then $Y(\sigma)$ has almost a Poisson distribution:

Proposition 2.5. *For every integers $C > 0$, $\epsilon > 0$, $p \geq 0$ and $n_0 = n_0(C, p, \epsilon)$ such that, for any $n \geq n_0$ $\mathscr{T}$ is a collection of $t \geq \frac{n^2}{6} - Cn$ pairwise edge-disjoint triangles in K_n, if $Y = Y(\sigma)$ is the random variable counting the number or triangles of $\mathscr{T}$ present in σ, then*

$$\left|\mathbb{P}(Y = p) - \frac{1}{ep!}\right| < \epsilon. \tag{2.6}$$

Proof. Let's apply Brun's sieve in the form described above. For each triangle $R \in \mathscr{T}$, let X_R be the indicator random variable such that $X_R = 1$ if and only if R is present in σ, and B_R be the event $X_R = 1$. Then the number of triangles present in σ is $Y = \sum_{R\in\mathscr{T}} X_R$. Denote

$$S^{(r)} = \sum_{\{R_1,R_2,\ldots,R_r\}\subset\mathscr{T}} \mathbb{P}[B_{R_1} \cap \cdots \cap B_{R_r}]$$

[2] For any A, B sets, $\overline{A \cup B} = \overline{A} \cap \overline{B}$.

the sum over all subsets $\{R_1, R_2, \ldots, R_r\} \subset \mathscr{T}$ of cardinality r. It is sufficient to show that if $t \geq \frac{n^2}{6} - Cn$ and $n = n(C, r, \epsilon)$ is sufficiently large, then

$$\left| S^{(r)} - \frac{1}{r!} \right| \leq \epsilon. \tag{2.7}$$

Then, by Brun's Method 2.4, with $\mu = 1$ and $X = Y$, the assertion (2.6) is proven.

Fix a collection of r distinct triangles $\{R_1, R_2, \ldots, R_r\} \subset \mathscr{T}$. Now, if no vertex lies in more than one R_i, $i = 1, \ldots, r$, then the number of permutations in which all $R_1, \ldots, R_r$ are present is $6^r(n - 2r)!$.

Then

$$\mathbb{P}[B_{R_1} \cap \cdots \cap B_{R_r}] = \frac{6^r(n-2r)!}{n!}.$$

In every other case, the number of permutations in which all R_i are present is smaller. Then, given that $t \leq n(n-1) \leq \frac{n^2}{6}$,

$$\begin{aligned} S^{(r)} &\leq \binom{t}{r} \frac{6^r(n-2r)!}{n!} \\ &\leq \frac{t!}{r!\ (t-r)!} \frac{6^r(n-2r)!}{n!} \\ &\leq \frac{t\ (t-1)\cdots(t-r+1)}{r!} \frac{6^r(n-2r)!}{n!} \\ &\leq \frac{t^r}{r!} \frac{6^r}{n(n-1)(n-2)\cdots(n-2r+1)} \\ &\leq \frac{n^{2r}}{6^r\ r!} \frac{6^r}{n(n-1)(n-2)\cdots(n-2r+1)} \\ &\leq \left(1 + O\left(\frac{1}{n}\right)\right) \frac{1}{r!}. \end{aligned}$$

On the other hand, considering that $|\mathscr{T}| = t \geq (n^2/6) - Cn$ and the fact that each vertex can lay in at most $(n-1)/2 \leq n/2$ elements of $\mathscr{T}$ [3], the number of collections $\{R_1, R_2, \ldots, R_r\} \subset \mathscr{T}$ such that all the triangles R_i do not share any vertices is

$$\frac{t(t - 3\frac{n}{2})(t - 6\frac{n}{2})\ldots(t - (r-1)\frac{n}{2})}{r!} \geq \frac{t^r}{r!}\left(1 - O\left(\frac{1}{n}\right)\right).$$

Therefore,

$$S^{(r)} \geq \frac{t^r}{r!}\left(1 - O\left(\frac{1}{n}\right)\right) \frac{6^r(n-2r)!}{n!} = \frac{1}{r!}\left(1 - O\left(\frac{1}{n}\right)\right).$$

This concludes the proof of (2.7), which proves the assertion of the proposition. □

In conclusion, the random variable $Y = Y(\sigma)$ counting the number of triangles of $\mathscr{T}$ present in a permutation σ, has approximately a Poisson distribution with mean 1 for large n.

[3] Recall that in a complete graph K_n, each vertex is incident with exactly $n-1$ edges. Therefore, each one can be in at most $(n-1)/2$ triangles that do not share any edges.

Step 3 It is finally possible to compute $\mathbb{E}[2^{Y(\sigma)}]$.

Corollary 2.6. *For every $C > 0$ and $\delta > 0$, there is an $n_0 = n_0(\delta, C)$ such that, for every $n \geq n_0$ and for every collection $\mathscr{T}$ of $t \geq (n^2/6) - Cn$ pairwise edge-disjoint triangles in K_n, the random variable $Y(\sigma)$ defined in step 1, i.e. the number of triangles of $\mathscr{T}$ present in σ, satisfies*

$$\mathbb{E}[2^{Y(\sigma)}] \geq e - \delta. \tag{2.8}$$

Proof. Given that

$$\sum_{p=0}^{\infty} \frac{2^p}{ep!} = \frac{1}{e} \sum_{p=0}^{\infty} \frac{2^p}{p!} = \frac{e^2}{e} = e$$

then, for a fixed $r = r(\delta)$

$$\sum_{p=0}^{r} \frac{2^p}{ep!} \geq e - \frac{\delta}{2}.$$

By Proposition 2.5, if $n \geq n_0 = n_0(\delta, C, r)$, then for all $p \leq r$,

$$\mathbb{P}[Y = p] > \frac{1}{ep!} - \frac{\delta}{2(r+1)2^p}.$$

Therefore, for such n,

$$\begin{aligned}
\mathbb{E}[2^{Y(\sigma)}] &= \sum_{p=0}^{\infty} 2^p \; \mathbb{P}[Y(\sigma) = p] \\
&> \sum_{p=0}^{r} 2^p \left(\frac{1}{ep!} - \frac{\delta}{2(r+1)2^p} \right) \\
&= \sum_{p=0}^{r} \left(e - \frac{\delta}{2} \right) - \sum_{p=0}^{r} \frac{\delta}{2(r+1)} \\
&= \left(e - \frac{\delta}{2} \right)(r+1) - \frac{\delta}{2(r+1)}(r+1) \\
&\geq e - \frac{\delta}{2} - \frac{\delta}{2} = e - \delta.
\end{aligned}$$

□

Step 4 The final step consists in proving that, for any n, there exists a collection $\mathscr{T}$ of $t \geq (n^2/6) - O(n)$ pairwise edge-disjoint triangles in K_n. The original paper [10] proved the following lemma using some results about the existence of Steiner Triple Systems. The proof presented below does not follow this track and proves the lemma directly.

Lemma 2.7. *For any $n \in \mathbb{N}$, there exists a collection $\mathscr{T}$ of $t \geq \frac{(n-2)(n-1)}{6}$ pairwise edge-disjoint triangles in K_n.*

Proof. Using the same notation for triangles as before, i.e. (a,b,c) denote the triangle on the vertices $a,b,c \in V(K_n) = \{1,2,\dots,n\}$ in the complete graph K_n, consider

$$\begin{aligned}\mathcal{F} &= \Big\{(a,b,c) \in \binom{[n]}{3} \mid a+b+c \equiv 0 \pmod{n}\Big\} \\ &= \Big\{(a,b,c) \in \binom{[n]}{3} \mid n \text{ divides } a+b+c\Big\}\end{aligned}$$

where $\binom{[n]}{3}$ is the set of non-ordered triplets chosen from the set $[n] = \{1,2,\dots,n\}$. The first step consists in showing that the triangles in $\mathcal{F}$ are pairwise edge-disjoint, i.e. there are no two triplets with two equal elements. Suppose that (a,b,c) and (a,b,c') are both elements of $\mathcal{F}$. Then, since $a+b+c$ and $a+b+c'$ are divisible by n, so is their difference $c-c'$. That is, $c-c' \equiv 0 \pmod{n}$. Given that $1 \le c, c' \le n$, then it must be $c = c'$.

Now, in order to find the size of $\mathcal{F}$, it is easier to count the number of ordered triplets (a,b,c) first, and then $|\mathcal{F}|$. Denote

$$\mathcal{X} = \{(a,b,c) \in \{1,2,\dots,n\}^3 \mid a+b+c \equiv 0 \pmod{n}\}$$

$$\begin{aligned}A &= \{(a,b,c) \in \mathcal{X} \mid a = b\} \\ B &= \{(a,b,c) \in \mathcal{X} \mid a = c\} \\ C &= \{(a,b,c) \in \mathcal{X} \mid b = c\}\end{aligned}$$

and observe that $|\mathcal{X}| = n^2$, because a and b can be chosen independently, while c is unique by $a+b+c \equiv 0 \pmod{n}$, $|A| = |B| = |C| = n$ because only one between a, b or c can be chosen independently, and the sets $A\cap B$, $A\cap C$, $B\cap C$, $A\cap B\cap C$ are all equal to the set $E = \{(a,b,c) \in \mathcal{X} \mid a = b = c\}$, $|E| \ge 1$. Then, by the inclusion-exclusion principle for sets [4], the number of ordered triplets (a,b,c), with a,b,c all distinct, is

$$|\mathcal{X}| - |A\cup B\cup C| = n^2 - 3n + 2|E| \ge n^2 - 3n + 2 = (n-1)(n-2).$$

The size of $\mathcal{F}$, which contains unordered triplets, is

$$|\mathcal{F}| = \frac{|\mathcal{X}| - |A\cup B\cup C|}{3!} \ge \frac{(n-2)(n-1)}{6}.$$

□

This last step finally concludes the proof of Theorem 2.3, because this lemma ensures that the result found in Corollary 2.6 can be applied in the identity (2.5) which becomes

$$\mathbb{E}[X] = \frac{n!}{2^{n-1}}\mathbb{E}[2^{Y(\sigma)}] \ge \frac{n!}{2^{n-1}}\big(e - o(1)\big). \tag{2.9}$$

[4]Recall that the formula of the inclusion-exclusion principle for three sets is: $|A\cup B\cup C| = |A| + |B| + |C| - |A\cap B| - |A\cap C| - |B\cap C| + |A\cap B\cap C|$.

2.3 The Upper Bound

As for the upper bound of $P(n)$, in the same paper Szele in 1943 proved the following

Theorem 2.8 (Szele, 1943). *There exists a constant c_1 such that*

$$P(n) \leq c_1 \frac{n!}{2^{3n/4}}.$$

Notice that the gap between the upper and lower bound in 1943 was exponential in n. However, Szele conjectured a tighter bound which, if proven, could guarantee that P(n) does not exceed the average number of Hamiltonian paths in a tournament on n vertices by more than a small polynomial factor in n [11].

Conjecture 2.1 (Szele's). *There exists a constant c_2 such that*

$$P(n) \leq c_2 n^{3/2} \frac{n!}{2^{n-1}}.$$

In the 1990s, Alon proved this conjecture in [11]. In this section a brief explanation of his proof is given, omitting some technical details. For the complete proof see [11].

Proof. The proof revolves around the permanent of $(0,1)-$matrices, i.e. matrices whose entries are only 0 or 1. Recall that the *permanent* of a square matrix A of order n is

$$\text{perm}(A) = \sum_{\sigma \in S_n} \prod_{i=1}^{n} a_{i\sigma(i)}$$

In 1963, Minc conjectured an upper bound for the permanent of $(0,1)-$matrices, which was proved by Bregman in 1973. It states that

Theorem 2.9 (Bregman, 1973). *Let A be a square $(0,1)-$matrix of order n, and let R_i denote the number of ones in the $i-$th row for $i = 1, 2, \ldots, n$. Then*

$$\textit{perm}(A) \leq \prod_{i=1}^{n} (R_i!)^{1/R_i}. \tag{2.10}$$

A proof of this theorem can be found in [2].

The number of Hamiltonian cycles in a tournament can be bounded using the permanent of its adjacency matrix. Recall that the *adjacency matrix of a graph G* on n vertices is a square $(0,1)-$matrix $A = (a_{ij})_{i,j=1,\ldots,n}$ of order n, such that $a_{ij} = 1$ if $(i,j) \in E(G)$ and $a_{ij} = 0$ otherwise. Recall that a *1-factor of a tournament T* is a spanning subgraph of T such that every indegree and every outdegree is 1, i.e. a spanning union of vertex-disjoint directed cycles. Denote by $F(T)$ the number of 1-factors of T, and $F(n) = \max\{F(T) \mid T \text{ is a tournament on } n \text{ vertices}\}$.

Let A_T be the adjacency matrix of a tournament T, then

$$\text{perm}(A_T) = F(T). \tag{2.11}$$

Now, denote by $C(T)$ the number of Hamiltonian cycles in the tournament T, and by $C(n) = \max\{C(T) \mid T \text{ is a tournament on } n \text{ vertices}\}$. Since every Hamiltonian cycle is also a 1-factor, the following inequality holds

$$C(T) \leq F(T). \tag{2.12}$$

Therefore, by Bregman's Theorem 2.9, the number of Hamiltonian cycles in a tournament T is bounded by the permanent of its adjacency matrix. Notice that $C(n) \leq F(n)$ holds as well. The following theorem guarantees that, under certain conditions, the inequality (2.10) is limited.

Theorem 2.10. *Define $g(x) = (x!)^{1/x}$. For every integer $S \geq n$, the maximum of the function $\prod_{i=1}^{n} g(x_i)$ subject to the constraints $\sum_{i=1}^{n} x_i = S$ and $x_i \geq 1$ are integers, is obtained if and only if the variables x_i are as equal as possible, (i.e. if and only if each x_i is either $\lfloor S/n \rfloor$ or $\lceil S/n \rceil$).*

Therefore, given that $\sum_{i=1}^{n} R_i = \binom{n}{2} \geq n$ and the R_i are as equal as possible,

$$\prod_{i=1}^{n} (R_i!)^{1/R_i} \leq (1 + o(1)) \frac{\sqrt{\pi}}{\sqrt{2}e} n^{3/2} \frac{(n-1)!}{2^n} \tag{2.13}$$

it is possible to obtain an upper bound on the number of Hamiltonian cycles in a random tournament. Indeed, this explains the following theorem.

Theorem 2.11. *For every tournament T on n vertices,*

$$C(T) \leq F(T) \leq (1 + o(1)) \frac{\sqrt{\pi}}{\sqrt{2}e} n^{3/2} \frac{(n-1)!}{2^n}.$$

Notice that $(n-1)/2^n$ is the expected number of Hamiltonian cycles in a random tournament on n vertices. Therefore, this result shows that the maximum possible number of Hamiltonian cycles is close to the average one.

Proposition 2.12. *For every tournament $T = (V, E)$ on n vertices there is a tournament T' on $n + 1$ vertices such that $C(T') \geq P(T)/4$. Then,*

$$P(n) \leq 4C(n+1) \leq c_2 \; n^{3/2} \frac{n!}{2^{n-1}}.$$

The proof of the conjecture follows easily from Theorem 2.11 and the previous proposition. □

Chapter 3

Lovász Local Lemma

3.1 The Lemma

The probabilistic method is based on proving that the probability of a "good" event happening is positive. Therefore, in order to find a desired configuration, a random one is selected, and if it has a positive probability of being the desired one, then it is possible to conclude that one exists. Instead, in some cases, it is more useful to show that the probability of "bad" events not happening is positive. These "bad" events are usually the possible ways in which a probabilistic experiment could fail. For instance, suppose that $A_1,\ A_2, \dots,\ A_n$ is a collection of events: if $\sum \mathbb{P}[A_i] < 1$, then there is a positive probability that none of them occur. Indeed,

$$\sum_{i=1}^{n} \mathbb{P}[A_i] \geq \mathbb{P}\Big[\bigcup_{i=1}^{n} A_i\Big] = 1 - \mathbb{P}\Big[\bigcap_{i=1}^{n} \overline{A_i}\Big] \Longrightarrow \mathbb{P}\Big[\bigcap_{i=1}^{n} \overline{A_i}\Big] > 0.$$

Even better, if the n events are independent, then their complements are also independent, and

$$\mathbb{P}\Big[\bigcap_{i=1}^{n} \overline{A_i}\Big] = \mathbb{P}\Big[\overline{A_1}\Big]\mathbb{P}\Big[\overline{A_2}\Big]\dots\mathbb{P}\Big[\overline{A_n}\Big] > 0.$$

For instance, if the events are independent with probability less than $1/2$, then $\mathbb{P}\big[\bigcap_{i=1}^{n} \overline{A_i}\big] > 2^{-n}$, which is greater than zero, although very small.

The aim of this chapter is to prove that something similar holds even if the events are independent of all but a small number of other events; this is guaranteed by the Lovász Local Lemma by Erdős and Lovász. A powerful feature of this lemma is that there is no condition on the number of events n: if n is large, $\mathbb{P}\big[\bigcap_{i=1}^{n} \overline{A_i}\big]$ can be extremely small, and the Lovász Local Lemma still guarantees the existence of a configuration that avoids all of these events [5].

Before introducing the Lemma, some definitions will be recalled. After that, three different versions of the Lovász Local Lemma will be discussed and proven: they all provide conditions under which it is possible to avoid "bad" events, with minimal differences of hypothesis. This section follows the notation and the proofs of [3], unless otherwise specified.

Definition 3.1. An event A is *independent of events* $B_1, B_2, \ldots, B_k$, if for any nonempty $J \subseteq \{1, \ldots, k\}$,

$$\mathbb{P}[A \cap \bigcap_{j \in J} B_j] = \mathbb{P}[A]\ \mathbb{P}\Big[\bigcap_{j \in J} B_j\Big].$$

Definition 3.2. Let $A_1, A_2, \ldots, A_n$ be events in a probability space. A directed graph $D = (V, E)$ on n vertices is a *dependency digraph* for $A_1, A_2, \ldots, A_n$ if the event A_i in independent of all the events A_j such that $(i, j) \notin E$.

For instance, if all the events are independent, then their dependency digraph has no edges, i.e. $E(D) = \emptyset$. Asking for *limited dependences* between events is equivalent to requesting few edges in the dependency digraph.

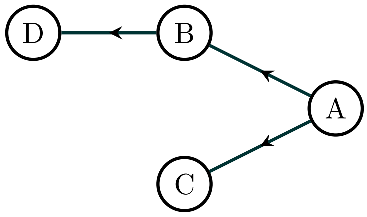


Figure 3.1: The dependency digraph for the events A, B, C, D. The event A depends on B and C. The event B depends on D.

The Local Lemma states that, given a collection of events whose probabilities are small and whose dependency is somehow limited, there is a positive probability that none of these events happen. The following is the general version, or *asymmetric version*, which is generally used when the probability of some events is much larger than the probability of others.

Lemma 3.1 (Lovász Local Lemma, 1975)**.** *Let $A_1, A_2, \ldots, A_n$ be events with $D = (V, E)$ their dependency digraph, and $x_i \in [0, 1)$ real numbers assigned to the events in such way that*

$$\mathbb{P}[A_i] \leq x_i \prod_{(i, j) \in E} (1 - x_j).$$

Then,

$$\mathbb{P}\Big[\bigcap_{i=1}^{n} \overline{A_i}\Big] \geq \prod_{i=1}^{n} (1 - x_i) > 0. \tag{3.1}$$

Proof. The complementary events $\overline{A_i}$ have positive probabilities, and the aim is to ensure that they happen simultaneously. This could not happen if the occurrence of a combination of some $\overline{A_j}$ forced some other A_i to hold. Therefore, it is necessary to bound the probabilities of the events A_i on the condition that the other events do not occur. That is to say, given that a combination of $\overline{A_j}$ occurs, then the probability of A_i also happening is small. This can be accomplished by introducing the parameters x_i. It is useful to prove that, for any subset $S \subset \{1, 2, \ldots, n\}$ and $i \notin S$,

$$\mathbb{P}\Big[A_i \;\Big|\; \bigcap_{j \in S} \overline{A_j}\Big] \leq x_i. \tag{3.2}$$

Proceed by induction on the size of S. If $S = \emptyset$, then the statement follows directly from the hypothesis of the lemma:

$$\mathbb{P}[A_i] \leq x_i \prod_{(i,\ j) \in E} (1 - x_j) \leq x_i.$$

Suppose now that (3.2) holds for any S' such that $|S'| < |S|$, and set $S_1 = \{j \in S \mid (i,\ j) \in E\}$, $S_2 = S \setminus S_1$. If $S_1 = \emptyset$, then A_i is independent of $\bigcap_{j \in S} \overline{A_j}$ and therefore

$$\mathbb{P}\Big[A_i \mid \bigcap_{j \in S} \overline{A_j}\Big] = \mathbb{P}[A_i] \leq x_i.$$

Assume $S_1 \neq \emptyset$. Then, by Definition 1.7 of conditional probability, and the fact that $\bigcap\limits_{j \in S} \overline{A_j} = \bigcap\limits_{j \in S_1} \overline{A_j} \cap \bigcap\limits_{t \in S_2} \overline{A_t}$,

$$\begin{aligned}
\mathbb{P}\Big[A_i \Big| \bigcap_{j \in S} \overline{A_j}\Big] &= \frac{\mathbb{P}\Big[A_i \cap \bigcap_{j \in S} \overline{A_j}\Big]}{\mathbb{P}\Big[\bigcap_{j \in S} \overline{A_j}\Big]} \\
&= \frac{\mathbb{P}\Big[A_i \cap \bigcap_{j \in S_1} \overline{A_j} \Big| \bigcap_{t \in S_2} \overline{A_t}\Big] \cdot \mathbb{P}\Big[\bigcap_{t \in S_2} \overline{A_t}\Big]}{\mathbb{P}\Big[\bigcap_{j \in S_1} \overline{A_j} \Big| \bigcap_{t \in S_2} \overline{A_t}\Big] \cdot \mathbb{P}\Big[\bigcap_{t \in S_2} \overline{A_t}\Big]} \qquad (3.3) \\
&= \frac{\mathbb{P}\Big[A_i \cap \bigcap_{j \in S_1} \overline{A_j} \Big| \bigcap_{t \in S_2} \overline{A_t}\Big]}{\mathbb{P}\Big[\bigcap_{j \in S_1} \overline{A_j} \Big| \bigcap_{t \in S_2} \overline{A_t}\Big]}.
\end{aligned}$$

Given that A_i is independent of the events $\{A_t \mid t \in S_2\}$, the numerator can be bounded as such:

$$\mathbb{P}\Big[A_i \cap \bigcap_{j \in S_1} \overline{A_j} \Big| \bigcap_{t \in S_2} \overline{A_t}\Big] \leq \mathbb{P}\Big[A_i \Big| \bigcap_{t \in S_2} \overline{A_t}\Big] = \mathbb{P}[A_i] \leq x_i \prod_{(i,\ j) \in E} (1 - x_j).$$

In order to bound the denominator, suppose $S_1 = \{j_1,\ j_2, \ldots,\ j_r\}$, then, by the chain rule (1.5), and the induction hypothesis:

$$\begin{aligned}
&\mathbb{P}\Big[\overline{A_{j_1}} \cap \cdots \cap \overline{A_{j_r}} \Bigm| \bigcap_{t \in S_2} \overline{A_t}\Big] = \mathbb{P}\Big[\bigcap_{t \in S_2} \overline{A_t} \cap \overline{A_{j_1}} \cap \cdots \cap \overline{A_{j_r}}\Big] \Big/ \mathbb{P}\Big[\bigcap_{t \in S_2} \overline{A_t}\Big] \\
&= \mathbb{P}\Big[\overline{A_{j_1}} \Bigm| \bigcap_{t \in S_2} \overline{A_t}\Big]\, \mathbb{P}\Big[\overline{A_{j_2}} \Bigm| \overline{A_{j_1}} \cap \bigcap_{t \in S_2} \overline{A_t}\Big] \cdots \mathbb{P}\Big[\overline{A_{j_r}} \Bigm| \overline{A_{j_1}} \cap \cdots \cap \overline{A_{j_{r-1}}} \cap \bigcap_{t \in S_2} \overline{A_t}\Big] \\
&\geq (1 - x_{j_1})(1 - x_{j_2}) \cdots (1 - x_{j_r}) \\
&\geq \prod_{(i,\ j) \in E} (1 - x_j).
\end{aligned}$$

Using these bounds, then (3.3) becomes

$$\mathbb{P}\Big[A_i \Big| \bigcap_{j \in S} \overline{A_j}\Big] \leq \frac{x_i \prod_{(i,\ j) \in E}(1 - x_j)}{\prod_{(i,\ j) \in E}(1 - x_j)} \leq x_i.$$

Hence, concluding the proof of the inequality (3.2). Now, the thesis of the lemma follows easily, using (3.2) and the chain rule (1.5):

$$\begin{aligned}
\mathbb{P}\Big[\bigcap_{i=1}^{n} \overline{A_i}\Big] &= \mathbb{P}\Big[\overline{A_1}\Big]\mathbb{P}\Big[\overline{A_2} \;\Big|\; \overline{A_1}\Big] \dots \mathbb{P}\Big[\overline{A_n} \;\Big|\; \overline{A_1} \cap \dots \cap \overline{A_{n-1}}\Big] \\
&= \Big(1 - \mathbb{P}[A_1]\Big)\Big(1 - \mathbb{P}\Big[A_2 \;\Big|\; \overline{A_1}\Big]\Big) \dots \Big(1 - \mathbb{P}\Big[A_n \;\Big|\; \overline{A_1} \cap \dots \cap \overline{A_{n-1}}\Big]\Big) \\
&\geq (1 - x_1)(1 - x_2) \dots (1 - x_n) \\
&\geq \prod_{(i,\ j) \in E} (1 - x_j).
\end{aligned}$$

□

However, the most popular version of the Lovász Local Lemma is the *symmetric version*, which can be applied if the events have bounded probabilities.

Lemma 3.2 (The Local Lemma; Symmetric Version)**.** *Let $A_1,\ A_2, \dots,\ A_n$ be events with $D = (V, E)$ their dependency digraph, $\mathbb{P}[A_i] \leq p$ for all $i = 1, \dots,\ n$, and $d^+(i) \leq d$ [1] for all i, i.e. each event A_i is independent of all but at most d of the other A_j. If $ep(d+1) \leq 1$, then*

$$\mathbb{P}\Big[\bigcap_{i=1}^{n} \overline{A_i}\Big] > 0.$$

Proof. If $d = 0$, then the events A_i are all independent and the result is trivial. Otherwise, suppose that $d \neq 0$ and set $x_i = \frac{1}{d+1} \leq 1$. Since the outdegree of any vertex in the dependency digraph D is at most d, and $ep(d+1) \leq 1$, then

$$x_i \prod_{(i,\ j) \in E} (1 - x_j) \geq \frac{1}{d+1}\Big(1 - \frac{1}{d+1}\Big)^d \geq \frac{1}{d+1}\frac{1}{e} \geq p.$$

Given that by assumption $p \geq \mathbb{P}[A_i]$, the Lovász Local Lemma 3.1 can be applied to obtain the desired conclusion. □

There is another version of the Local Lemma, whose notation makes it possible to apply it to the study of non-repetitive colouring of graphs that is discussed in the next section.

Lemma 3.3 (The Local Lemma, Multiple Version, [12])**.** *Let $\mathscr{A} = \mathcal{A}_1 \cup \mathcal{A}_2 \cup \dots \cup \mathcal{A}_r$ be a partition of a finite set of events, with $\mathbb{P}[A] = p_i$ for every $A \in \mathcal{A}_i$, $i = 1, 2, \dots, r$. Suppose that there are $0 \leq a_1,\ a_2, \dots,\ a_r < 1$ real numbers, and $d_{ij} \geq 0$ for $i,\ j = 1,\ 2, \dots,\ r$ such that*

1. *for any event $A \in \mathcal{A}_i$, there exists a set $\mathcal{D}_A \subseteq \mathscr{A}$, with $|\mathcal{D}_A \cap \mathcal{A}_j| \leq d_{ij}$ for all $j = 1,\ 2, \dots,\ r$, such that A is independent of $\mathscr{A} \setminus (\mathcal{D}_A \cup \{A\})$,*

[1] Recall that $d^+(v)$ is the outdegree of v, i.e. the number of edges going out of v.

2. $p_i \le a_i \prod_{j=1}^{r}(1-a_j)^{d_{ij}}$ *for all* $i = 1,\ 2, \dots,\ r.$

Then:

$$\mathbb{P}\Big[\bigcap_{A\in\mathscr{A}} \overline{A}\Big] > 0.$$

Again, the proof is simply an application of the asymmetric version of the Local Lemma 3.1.

Proof. By hypothesis, for any event $A \in \mathcal{A}_i$ there exists a set $\mathcal{D}_A = \{B \in \mathscr{A} \mid \text{ the event } A \text{ depends on the event } B,\ B \neq A\}$.

Therefore $\mathcal{D}_A \cap \mathcal{A}_j = \{B \in \mathcal{A}_j \mid \text{ the event } A \text{ depends on the event } B\}$, and the parameter d_{ij} gives an estimate on the size of this set.

For every $i = 1, 2, \dots, r$, set $x_A = a_i \in [0, 1)$ for every $A \in \mathcal{A}_i$. The key is proving that these parameters x_A satisfy the hypothesis of the Lemma 3.1. Indeed,

$$\begin{aligned} p_i \le & a_i \prod_{j=1}^{r}(1-a_j)^{d_{ij}} \\ \le & a_i \prod_{j=1}^{r} \prod_{\{B\in\mathcal{A}_j \mid A \text{ depends on } B\}} (1-a_j) \\ = & x_A \prod_{\{B\in\mathscr{A} \mid A \text{ depends on } B\}} (1-x_B). \end{aligned}$$

Therefore, by the asymmetric version of the Local Lemma, it is possible to conclude that

$$\mathbb{P}\Big[\bigcap_{A\in\mathscr{A}} \overline{A}\Big] > 0.$$

□

3.2 Non-Repetitive Colouring

The best way to truly appreciate the power of the Lovász Local lemma is to show an example of its application. In this section, this is accomplished by discussing the non-repetitive edge colouring of a graph. The aim is to find an answer to the following problem: "What is the minimum amount of colours needed to paint the edges of a graph in such a way that the sequence of colours of each path is non-repetitive?".

To address this question, first it is necessary to introduce the definitions of edge colouring, non-repetitive sequence and $\pi(G)$, the minimum number of colours needed to colour non-repetitively the edges of a graph G. Finally, a result by Alon *et al.* in [12], links $\pi(G)$ to the degree of G. Unless stated otherwise, all the following statements and results are taken from [12].

Definition 3.3 ([9])**.** A *proper k-edge colouring* of a graph G is a function that assigns one of k colours of each edge of G, such that any two adjacent edges must be assigned different colours.

Recall that two edges are adjacent if they have a vertex in common.

Definition 3.4 ([9])**.** A graph is *k-edge colourable*, or simply *k-colourable*, if it admits a proper k-edge colouring. The smallest integer k for which G is k-edge colourable is called the chromatic index of G and is denoted by $\chi'(G)$, or simply χ'.

Definition 3.5. A *finite sequence* $a = a_1 a_2 \dots a_n$ of symbols from a set S is called *non-repetitive* if it does not contain a sequence of the form $xx = x_1 x_2 \dots x_m x_1 x_2 \dots x_m$ $x_i \in S$ as a subsequence of consecutive terms.

For instance, given the set $S = \{1, 2, 3\}$, the sequence $a = 123132123213$ is non-repetitive, whereas $b = 1\mathbf{2323}21$ is not.

Definition 3.6. A *colouring* of the set of edges E of a graph $G = (V, E)$ is *non-repetitive* if the sequence of colours of any path in G is non-repetitive. The minimum number of colours needed is called the *Thue number*, and is denoted by $\pi(G)$.

Note that repetitions forming cycles are allowed.

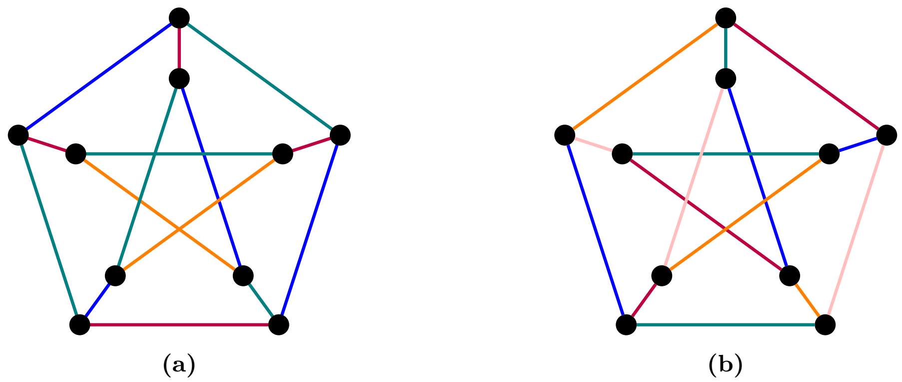


Figure 3.2: A 4-edge colouring and a non-repetitive 5-colouring of the Petersen graph.

For instance, the chromatic index of the Petersen graph in Figure 3.2 is $\chi'(G) = 4$. It is possible to prove that the Thue number is indeed $\pi(G) = 5$.

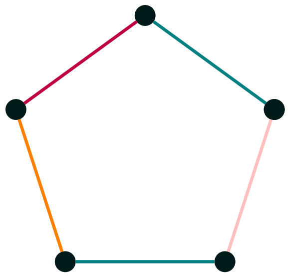
Theorem 3.4 (Thue, 1906)**.** *There exist arbitrarily long non-repetitive sequences built of only three different symbols.*

The idea of the proof is the following: using a constructive method consisting of substitutions over a given set of symbols. For instance, the substitution

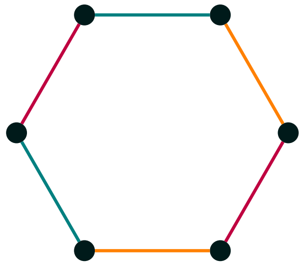
$$\begin{aligned} 1 &\longrightarrow 12312 \\ 2 &\longrightarrow 131232 \\ 3 &\longrightarrow 1323132 \end{aligned} \tag{3.4}$$

over a non-repetitive sequence of the set $\{1, 2, 3\}$, generate a non-repetitive sequence, i.e. the substitution preserves the non-repetitiveness property of a sequence. That is, applying (3.4) to the non-repetitive sequence 123, then 123121312321323132 is still non-repetitive.

This theorem of Thue proves also that, for any path P_n on $n \geq 4$ vertices, $\pi(P_n) = 3$. As a consequence, for any cycle of length $n \geq 3$, $\pi(C_n) \leq 4$.



(a) $\pi(C_5) = 4$.



(b) $\pi(C_6) = 3$.

Figure 3.3

The result for cycles implies that for any graph G of degree $\Delta \leq 2$ then $\pi(G) \leq 4$. Since this is true for χ' [2], it is natural to wonder if $\pi(G)$ is bounded on the class of graphs with $\Delta \leq k$ for each $k \geq 3$. The following theorem assures that it is.

Theorem 3.5 (Alon *et al.*, 2002)**.** *There exists an absolute constant c such that, for any graph G with degree at most Δ,*

$$\pi(G) \leq c\Delta^2. \tag{3.5}$$

The proof of this theorem is an application of the Local Lemma 3.3, where the "bad" events that are to be avoided are of the form "P is a path of length $2i$ whose sequence of colours form a repetition".

Proof. Let $G = (V,\ E)$ be a graph with degree $\Delta = \max_{v \in V} d(v)$. Consider a random colouring of the edges of G with C colours, i.e. assign randomly and independently one of C colours to each edge $(u, v) \in E(G)$, with equal probability $1/C$. The value of C will be determined later.

In order to apply the Multiple Version of the Local Lemma 3.3 it is necessary to determine p_i, a_i and d_{ij}. First, start by defining the events $\mathcal{A}_i$.

For a path P_{2i} of even length in G, let $A(P)$ be the event that the second half of P is coloured the same as the first, i.e. the sequence of colours of P forms a repetition. Set $\mathcal{A}_i = \{A(P) \mid P \text{ is a path of length } 2i \text{ in } G\}$, so that

$$\mathscr{A} = \mathcal{A}_1 \cup \mathcal{A}_2 \cup \cdots \cup \mathcal{A}_r = \{A(Q) \mid Q \text{ is a path in } G\}$$

and $\mathbb{P}[A(P)] = p_i = C^{-i}$.

[2]Recall that a theorem by Vizing states that, for a graph of degree Δ, its chromatic index is either Δ or $\Delta + 1$.

Given that each path P_{2i} of length $2i$ share an edge with at most $4ij\Delta^{2j}$ paths of length $2j$, $d_{ij} = 4ij\Delta^{2j}$ appears to be a good choice[3]. Indeed, given an event $A(P) \in \mathcal{A}_i$, there exists

$$\mathcal{D}_{A(P)} = \{A(Q) \mid Q \text{ is a path that shares an edge with } P\} \subset \mathscr{A}$$

such that

$$\left|\mathcal{D}_{A(P)} \cap \mathcal{A}_j\right| = \left|\{A(Q) \mid Q \text{ is a path of length } 2j \text{ that shares an edge with P}\}\right| \leq 4ij\Delta^{2j}$$

for all j, and $A(P)$ is independent of

$$\mathscr{A} \setminus (\mathcal{D}_{A(P)} \cup \{A\}) = \{A(Q) \mid Q \text{ is a path that has no edge in common with } P\}.$$

Hence, $d_{ij} = 4ij\Delta^{2j}$ is a good choice as it satisfies the first condition of Lemma 3.3.

Now, let $a_i = a^{-i}$, with $a = 2\Delta^2$. As $a_i \leq 1/2$, then

$$(1 - a_i) \geq e^{-2a_i}.$$

The second condition of the lemma applies provided that

$$p_i \leq a_i \prod_j (1 - a_j)^{d_{ij}}.$$

Indeed, for any i,

$$\begin{aligned} a_i \prod_j (1 - a_j)^{d_{ij}} &\leq a_i \prod_j e^{-2a_j d_{ij}} \\ &= a^{-i} \prod_j e^{-2a^{-j} 4ij\Delta^{2j}} \\ &= a^{-i} \prod_j \exp(-8ija^{-j}\Delta^{2j}). \end{aligned}$$

Therefore,

$$\begin{aligned} p_i \leq a_i \prod_j (1 - a_j)^{d_{ij}} &\iff C^{-i} \leq a^{-i} \prod_j \exp\left(-8ija^{-j}\Delta^{2j} \right) \\ &\iff C \geq a \prod_j \exp\left(8ja^{-j}\Delta^{2j}\right) = a \ \exp\left(8 \sum_j j \left(\frac{\Delta^2}{a}\right)^j \right) \\ &\iff C \geq 2\Delta^2 \exp\left(8 \sum_j j \left(\frac{1}{2}\right)^j \right) \\ &\iff C \geq 2\Delta^2 \exp(8 \cdot 2) = 2\Delta^2 e^{16} \end{aligned}$$

[3]That is because there are $2i2j$ possible combinations for the common edge, and at most Δ^{2j} paths of length $2j$ in a graph of degree Δ. Therefore, for each of the Δ^{2j} paths of length $2j$ there are $4ij$ possibilities for the shared edge.

where the last inequality holds because of the fact that $\sum_{j=1}^{\infty} j(1/2)^j = 2$. Hence, for $C \geq 2\Delta^2 e^{16}$ the Multiple Version of the Local Lemma 3.3 guarantees the existence of a non-repetitive C-colouring of G, i.e.

$$\mathbb{P}[\bigcap_{A(P)\in\mathscr{A}} \overline{A(P)}] > 0.$$

This proves the theorem with $c = 2e^{16} + 1$. □

An important aspect of the study of this problem is the asymptotic shape of the function

$$\pi(n) = \max\{\pi(G) \mid G \text{ is a graph of degree } \Delta(G) \leq n\}.$$

The quadratic upper bound on $\pi(n)$ shown in Theorem 3.5 is the best known at this moment. On the other hand, some results suggest that a constructive method could give better results than the probabilistic one. The following conjecture suggest an improvement on the quadratic bound of $\pi(n)$:

Conjecture 3.1. *There is an absolute constant c such that, for all $n \geq 1$,*

$$\pi(n) \leq cn.$$

Non-trivial lower bounds for $\pi(n)$ are desired too. To find one, it is possible to search for bad graphs among members of some special families, for instance complete graphs on $2^k + 1$ vertices. In fact, $\pi(K_5) = 7$ is quite a jump from $\pi(K_4) = 3$.

Theorem 3.5 holds for *non-repetitive vertex colouring* as well, although for some classes of graphs the *vertex Thue number* may be bounded even if the degree Δ is arbitrarily large. For instance, for every tree 4 colours suffice. The following theorem shows that, for vertex colouring, the quadratic dependence of the number of colours is nearly tight.

Theorem 3.6. *There exists an absolute constant $c > 0$ with the following property: for every integer $\Delta > 1$, there exists a graph G with maximum degree Δ such that every non-repetitive vertex colouring of G uses at least $c\frac{\Delta^2}{\log \Delta}$ colours.*

3.3 Explicit Non-Repetitive Colourings

There are some special classes of graphs for which the bounds on the Thue number can be found by giving explicit non-repetitive colourings, for example complete graphs and trees. Recall that a *tree* is a connected graph with no cycles, i.e. with only one path from one vertex to another, and a *star* is a tree on n vertices, such that one has degree $n-1$ and all the others have degree 1. Intuitively, a star is a graph in which all the edges have the same vertex in common. These results suggest that a constructive method is possible and could give better results than the probabilistic one.

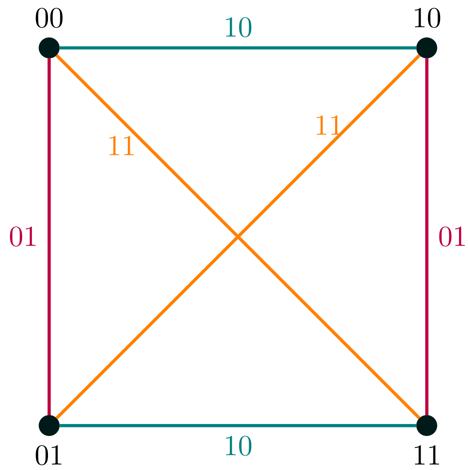


Figure 3.4: A non-repetitive colouring of K_4 with $\pi(K_4) = 3$ colours.

Proposition 3.7. *$\pi(K_{2^k}) = 2^k - 1$ for all $k \geq 1$. Therefore, for any complete graph K_n, $\pi(K_n) \leq 2n - 3$.*

Proof. First, label the vertices of K_{2^k} by distinct elements of the additive group $\mathbb{Z}_2^k = \{0,1\} \times \{0,1\} \times \cdots \times \{0,1\}$, i.e. the direct product of k copies of $\mathbb{Z}_2 = \{0,1\}$. Then, colour the edges by non-zero elements of $\mathbb{Z}_2^k$, so that an edge (x, y) gets the colour $x + y$. In this way, the only repetitions appear in cycles, since $x + x = 0$. □

An example of this construction is shown on the complete graph K_4 in Figure 3.4. The vertices of K_4 are labelled with distinct elements of $\mathbb{Z}_2^2 = \{0,1\} \times \{0,1\}$, and the edges are coloured according to the labels of its vertices. The only repetitions in this colouring appear in cycles, which is allowed. As stated in the proposition, $\pi(K_4) = 4 - 1 = 3$, which is exactly the number of colours needed in the example.

For the next example, a new definition is needed. A sequence $a = a_1 a_2 \ldots a_n$ is *palindrome* if it is equal to its own reflection $\tilde{a} = a_n a_{n-1} \ldots a_1$. For instance, 1231321 is palindrome. It is not possible to obtain a long non-repetitive and palindrome-free sequence of only 3 symbols, whereas it is possible on 4 symbols. Indeed, by taking a long non-repetitive sequence of 3 symbols and adding the fourth one between consecutive blocks of length 2, then the sequence is also palindrome-free. For instance, the sequene 123132123213 becomes 124314324124324134.

Proposition 3.8. *Let T be any tree with $\Delta(T) \geq 2$. Then $\pi(T) \leq (\Delta(T) - 1)$.*

Proof. Let T be a tree with maximum degree $\Delta \geq 2$. As the root of T choose a vertex of degree strictly less than Δ, and arrange the rest of vertices by their distance from the root. The edges of T can be partitioned into levels, $L_1, L_2, \ldots$, each of which consists of disjoint stars.

Let $b = b_1 b_2 b_3 \ldots$ be non-repetitive and palindrome-free sequence over the set $\{1, 2, 3, 4\}$. Each b_i will be assigned to the level L_i of T.

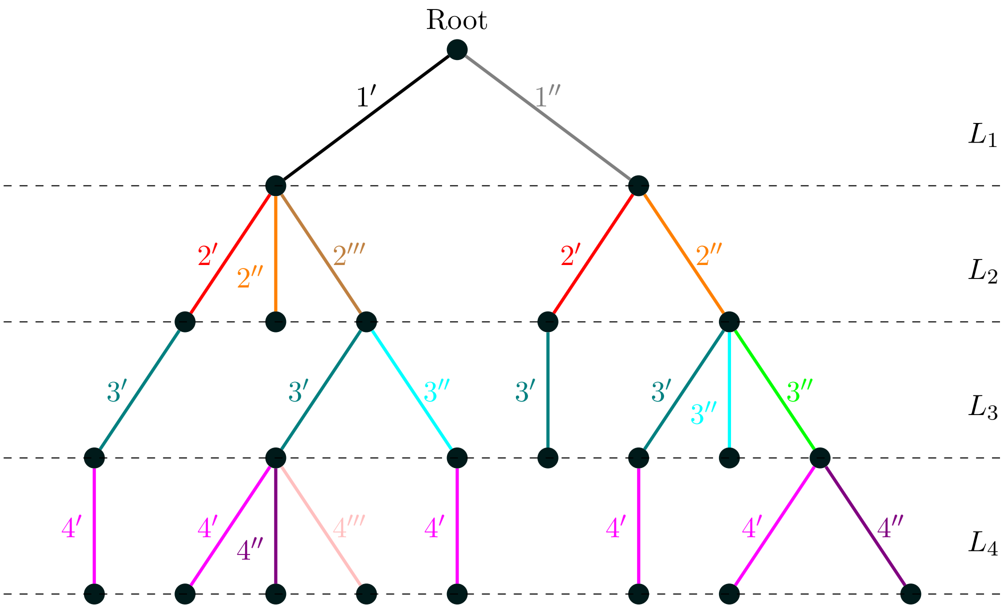


Figure 3.5: A non-repetitive colouring of a tree T with $\Delta(T) = 4$ and 11 colours.

Take 4 disjoint sets of colours $A_i = \{i', i'', \ldots, i^{\Delta-1}\}$ for $i = 1, 2, 3, 4$ and colour each star on level j with distinct colours from the set A_{b_j}.

For the proof that this is indeed a non-repetitive colouring see [12]. □

As an example of this construction, see Figure 3.5. In this tree T, there are four levels $L_1, \ldots, L_4$, and the non-repetitive sequence chosen is $b = 1234$. Therefore, in each level L_i the stars are coloured with elements from the set $A_i = \{i', i'', i'''\}$. The proposition assured that $\pi(T) \leq 4(\Delta(T) - 1) = 12$, but in the example only 11 colours are needed, so $\pi(T) \leq 11$.

Conclusions

The probabilistic method has emerged as a powerful non-constructive tool in Combinatorics, Discrete Mathematics, and many other areas of mathematics and computer science. It was developed in the 1950s, mainly by Paul Erdős who is considered to be the first mathematician who methodically applied probabilistic arguments to solve discrete problems.

The use of the probabilistic method in Graph Theory made it possible to find the solution to many problems that could not be solved via a constructive method. In many cases, the probabilistic proof is still the only one known or the one that guarantees the best result. For instance, the bound on the *Ramsey numbers*, mentioned in the Introduction, can only be proven by probabilistic method. Interestingly, the problem of determining these numbers has grown into such an extensive field that it has developed into its own subject, the *Ramsey Theory*, and is still an active area of research.

The study of the number of Hamiltonian paths in tournaments in Chapter 2 was the first example of the probabilistic method shown in this thesis. The proof of Szele on the lower bound is a clever yet intuitive proof: it counts the average number of Hamiltonian paths in a random tournament and concludes the existence of at least one tournament with more of such paths. An interesting aspect of this theorem is its historical importance. Indeed, it almost certainly was the first application of a probabilistic argument in a problem of Discrete Mathematics. These results, discovered in 1943, proved the following bounds on the maximum number of Hamiltonian paths in tournaments on n vertices $P(n)$:

$$\frac{n!}{2^{n-1}} \leq P(n) \leq c_1 \frac{n!}{2^{3n/4}}.$$

In the 1990s, the following improvements were discovered on the lower and upper bound:

$$(e - o(1)) \frac{n!}{2^{n-1}} \leq P(n) \leq c_2 n^{3/2} \frac{n!}{2^{n-1}}.$$

The gap between these bounds changed from exponential in n to polynomial in n. It would be interesting to close the gap and determine $P(n)$ up to a constant factor [11].

Then, the Lovász Local Lemma was introduced as one of the most useful tools in the probabilistic method. This lemma was proved by Erdős and Lovász in 1975, and guarantees that, under certain conditions, all the "bad events" can be avoided, see

Chapter 3. For many years, it remained entirely non-constructive. That is because, while the Lemma guarantees that $\mathbb{P}[\cap \overline{A_i}]$ is positive, this probability could be exponentially small. Therefore, sampling the probability space to find an object with the desired property could take an exponentially long time.

In 1990, a quest to find constructive proofs of this Lemma began. This is not surprising because a constructive proof is desirable even if a theorem has been proven by non-constructive methods. In 2010 Moser and Tardos showed that the Lovász Local Lemma gives more than pure existence results; in fact, there exists a randomized algorithm that can be used to find a desired object. A *randomized algorithm* works by generating a random number within a specific range and making decisions based on its value. Moser and Tardos discovered that in a special framework of the Lovász Local Lemma, there exists a simple Las Vegas algorithm with expected polynomial runtime, that searches the probability space for a point that avoids all bad events. A *Las Vegas algorithm* is a randomized algorithm that always provides the correct result, but whose running time is subject to randomization. In other words, it guarantees the accuracy of the output while the time it takes to obtain this result may vary. In the following years, the algorithm was extended to different situations by other researchers.

There are many application of the Local Lemma in Graph Theory. In a random graph, the "bad" events could be those configurations that do not satisfy a certain condition. For instance, in the proof of Theorem 3.5, the condition needed was a non-repetitive colouring. Recall that this theorem proved that if a graph G has bounded degree $\Delta(G) \leq \Delta$, then the Thue number $\pi(G)$, i.e. the number of colours needed to allow a non-repetitive colouring, is bounded by

$$\pi(G) \leq c\Delta^2.$$

This is the best upper bound known at the moment, although a conjecture suggests that $\pi(n)$, i.e. the maximum Thue number between graphs on n vertices, could be bounded by $\pi(n) \leq cn$. An interesting topic on this subject is the search for constructive non-repetitive colourings. For instance, there are constructive methods for colouring complete graphs and trees. In fact, the results known for now suggest that constructive proofs could give better estimates on the Thue number, showing that there is still a lot to discover about this topic.

Bibliography

[1] Noga Alon. Paul erdős and probabilistic reasoning. In *Erdős Centennial*, pages 11–33. Springer, 2013.

[2] Noga Alon and Joel H Spencer. *The probabilistic method*. John Wiley & Sons, 2016.

[3] Jiří Matoušek and Jan Vondrák. The probabilistic method. *Lecture Notes, Department of Applied Mathematics, Charles University, Prague*, 2001.

[4] Joel Spencer. *Ten lectures on the probabilistic method*. SIAM, 1994.

[5] Joel Spencer. Probabilistic methods. *Graphs and Combinatorics*, 1:357–382, 1985.

[6] Robin J. Wilson. *Introduction to Graph Theory*. Addison Wesley Longman Limited, 1996.

[7] Balasz M. and Toth B. Inclusion-exclusion principle. *Lecture notes, University of Bristol / Budapest University of Technology and Economics*, 2014.

[8] Francesco Caravenna and Paolo Dai Pra. *Probabilità: Un'introduzione attraverso modelli e applicazioni*, volume 67. Springer Science & Business Media, 2013.

[9] Joy Morris. *Combinatorics*. Joy Morris, Lethbridge, Alberta, version 2.1 edition, 2022.

[10] Ilan Adler, Noga Alon, and Sheldon M Ross. On the maximum number of hamiltonian paths in tournaments. *Random Structures & Algorithms*, 18(3):291–296, 2001.

[11] Noga Alon. The maximum number of hamiltonian paths in tournaments. *Combinatorica*, 10(4):319–324, 1990.

[12] Noga Alon, Jarosław Grytczuk, Mariusz Hałuszczak, and Oliver Riordan. Nonrepetitive colorings of graphs. *Random Structures & Algorithms*, 21(3-4):336–346, 2002.