Hang Lung Mathematics Awards
Vol. 6 (2014)

Hang Lung Mathematics Awards 2014

Bronze Award

Probability, Matrices, Colouring and Hypergraphs

Team Members: Hok Kan Yu, Dave Lei, Ka Chun Wong, Sin Cheung Tang
Teacher: Mr. Yan Ching Chan
School: Po Leung Kuk Centenary Li Shiu Chung Memorial College

Hang Lung Mathematics Awards
Vol. 6 (2014), pp. 103–130

PROBABILITY, MATRICES, COLOURING AND HYPERGRAPHS

TEAM MEMBERS
Hok Kan Yu, Dave Lei,
Ka Chun Wong, Sin Cheung Tang

TEACHER
Mr. Yan Ching Chan

SCHOOL
Po Leung Kuk Centenary Li Shiu Chung Memorial College

Abstract. In this project, we have achieved various results using Probabilistic methods. By exploiting the concept of probability and expected value, we manage to achieve three results: on distribution of entries on a cube, on colouring of vertices of a hypergraph and a lower bound of a maximal independent set on a hypergraph.

Preface

This project is mainly about the application of probability on graph theory. We are inspired and motivated to delve into this field by the amazing usefulness of probability - we can claim a certain graph exist by proving the probability of its existence larger than 0, without even constructing it. We also appreciate the works of some celebrated Mathematicians, such as Paul Erdős.

In Chapter 1, we go over some basic concepts concerning Probability Theory and Graph Theory, which is reorganized from the course notes on ”Probability and Discrete Mathematics” written by Professor Roman Kotecký from University of Warwick. This basic introduction attempts to make our following works more ”eatable”. After the introduction of Probability Theory, we have shown a piece of our original work: a upper bound of $\mathbb{P}(X-\mu \geq \lambda)$ and $\mathbb{P}(X=0)$.

Going through the basics, we attempted to demonstrate the versatility of this probabilistic method in the field of graph theory in Chapter 2. With reference to the exercises in [1], we successfully showed how such a tool can be utilized to investigate orthonormal vectors and distribution of entries on a matrix. For the part of graph theory, we extracted 3 prominent results done by Erdős, Szele and Turán. S_k

property of a tournament was investigated by Erdős in 1947 while the existence of a tournament with large Hamiltonian paths was proved by Szele in 1943. Last but not least, Turán's theorem, as suggested by its name, was Turán's work in 1941. With the aid of exercises in [1], we discovered a corollary of Turán's theorem, which is related to triangle-free graphs.

In Chapter 3, motivated by the problems we solved in the previous chapter, we presented several pieces of our original work. We are able to extend the problem of matrices to the world of cube. Also, we have a generalization on the problems of matrices and cubes, which is our original work. Up till Chapter 2, the entire discussion was based on ordinary graphs. Thanks to the probabilistic method, we were capable of achieving some interesting results in hypergraphs, which are featured in Chapter 3. 4-colouring problem of hypergraphs is actually a problem created by Niranjan Balachandran. The section following, which generalized it to k-colouring problem, is our original work.

With reference to the proof of weak Turán's theorem on ordinary graphs, included in [1], we generalized it by ourselves to a hypergraph version. This result is our pride and joy in this project.

To round off this preface, we want to note down here one of the most intricate yet beautiful discovery of this project. With the use of mean (or expected value), we can prove the existence of objects with a property greater or less than the mean. A straightforward method to evaluate the mean is to convert it into indicator functions, which can in turn be replaced by probability function. By assuming uniform distribution, the probability space constructed allow us to evaluate the probability very easily. Thus, expected value, indicator function and probability function form a miraculous triangle in proving existence. Not only is the concept of probability triple is intriguing, but also is this powerful triangle.

Probability and expected value, though being mocked as useless in college mathematics, in fact allow us to see a wondrous world, which was probably unexpected by us before this project.

Acknowledgement

We would like to express our greatest gratitude to the following people, since they construct the probability space for this project to happen.

Throughout this project, our teacher advisor Mr Chan has led us through the materials concerning probability and graph theory, giving us advice to us at the expense of his own spare time. Being our most humourous Mathematics teacher, he always includes myriads of anecdotes of Mathematics in his lessons. His unique style of teaching is worth mentioning, so is his patience.

Equally notable is the irreplaceable role our alumni Mr Tsoi Kwok Wing has played. We wholehearted thank him for offering us materials regarding probability and graph theory. Despite being our number theory teacher, he inspired us to do an investigation into probabilistic methods. We also thank him for his LaTeX support; without his consistent instruction, our project would be presented in a far less elegant manner.

Our parents, who bring us to this world, allow us to relish the excitement of living. Had it not been for their unremitting effort and unfailing love, we would not be able to reach that far. We also apologize for diffuse attention to them due to immersion in this project.

Supports from our families and teachers are admittedly indispensible, but we shall never forget those from our friends; they are like stars strewn in the sky. Their light are meagre, yet more than enough to sooth our daunted hearts. We thank our talented friend Cheng Wai Ching, who has written a tailor-made programme for us to investigate different graphs. Chan Tsun Ying, Chan Wai Yi, Cheung Chin, Cheng Yi Man, Chu Tik Ming, Chung Rena, Fung Wai San, Lai Kin Ngai, Li Pui Wun, Lo Ki Lok, Mak Yick Tim, Tang Tsz Ying, Tsui Suet Ying Gloria, and So Chun Kit are definitely our intimate friends; without them, we are merely vertices in the graph ”lonely hearts party”[4]. Thank you.

Last but not least, availing ourselves of this opportunity, we also want to express our gratitude towards our English teachers, Miss Siu and Miss Kwan. Without them, the last piece of this project, this thesis, would not be born.

1. The Notion of Probability and Graph

In high school mathematics, probability is about ”When 2 cards are drawn at random from a deck of 52 playing cards, find the probability that there are 1 King and 1 Queen” or ”If the letters of the word ‘WOMAN’ is rearranged randomly, find the probability that the first letter is a vowel” [3], and we barely know what a graph is. However, it is only a layman's perspective and this so-called ‘probability’ barely contribute anything to human kind. In essence, probability methods are a versatile tool. In order to appreciate the beauty of this method, we need to understand subtlety behind - Probability Theory.

1.1. Probability Theory

To start with, we shall introduce the probability triple $(\Omega, \mathscr{F}, \mathbb{P})$. Ω is called the **sample space** which is simply a non-empty set.

Definition 1. *Let Ω be a non-empty set. A σ-algebra $\mathscr{F}$ is a non-empty collection of subsets of Ω such that*

1. *If $E \in \mathscr{F}$, then we have its complement $\Omega \backslash E \in \mathscr{F}$.*
2. *If $E_1, E_2, \cdots \in \mathscr{F}$, then we have $\bigcup_{n=1}^{\infty} E_n \in \mathscr{F}$.*

For example, we have that the power set $\mathscr{P}(\Omega)$ is an example of σ-algebra of Ω. Throughout this report, we are going to make the following assumption.

Agreement. In this report, we are going to take Ω to be finite and $\mathscr{F} = \mathscr{P}(\Omega)$.

[See reviewer's comment (2)]

Because of this assumption, we are allowed to simplify the definition of a probability measure $\mathbb{P}$ as follows.

Definition 2. *Let Ω be finite and $\mathscr{F} = \mathscr{P}(\Omega)$. A probability measure is a map $\mathbb{P} : \mathscr{F} \to [0,1]$ such that*

1. *$\mathbb{P}(\Omega) = 1$.*
2. *If $A, B \in \mathscr{F}$ are disjoint ($A \cap B$ is empty), then we have*

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

[See reviewer's comment (3)]

Example 3. *Suppose we are tossing a fair coin once. The sample space is given by $\Omega = \{H, T\}$, where H and T denotes the event when a head and a tail is shown respectively. Therefore, $\mathscr{F} = \mathscr{P}(\Omega) = \{\emptyset, \{H\}, \{T\}, \{H, T\}\}$. Since the coin is fair, the probability measure is defined by fixing $\mathbb{P}(\{H\}) = \frac{1}{2}$ and $\mathbb{P}(\{T\}) = \frac{1}{2}$. (This is usually called the uniform probability.)*

We call $(\Omega, \mathscr{F}, \mathbb{P})$ a probability triple. Again, due to our **Agreement**, we can simplify the definition of a random variable, as follows.

Definition 4. *A random variable is a function $X : \Omega \to \mathbb{R}$.*

Example 5. *Suppose we are tossing a fair dice. Then we obtain the obvious sample space by*

$$\Omega = \{\mathit{ONE},\ \mathit{TWO},\ \mathit{THREE},\ \mathit{FOUR},\ \mathit{FIVE},\ \mathit{SIX}\}$$

and $\mathscr{F} = \mathscr{P}(\Omega)$ *as agreed. Again, as the dice is fair, we have* $\mathbb{P}(\{x\}) = \frac{1}{6}$ *for each* $x \in \Omega$*. Now we can define an obvious random variable* $X : \Omega \to \mathbb{R}$ *by*

$$\begin{aligned} ONE &\mapsto 1 \\ TWO &\mapsto 2 \\ THREE &\mapsto 3 \\ FOUR &\mapsto 4 \\ FIVE &\mapsto 5 \\ SIX &\mapsto 6 \end{aligned}$$

By convention (and an abuse of notations), we shall denote $\mathbb{P}(X^{-1}\{x\})$ simply by $\mathbb{P}(X = x)$. In the previous example, we have

$$\mathbb{P}(X = 1) = \mathbb{P}(X^{-1}\{1\}) = \mathbb{P}\{ONE\} = \frac{1}{6}$$

Definition 6. *Let* $(\Omega, \mathscr{F}, \mathbb{P})$ *be a probability triple and* $X : \Omega \to \mathbb{R}$ *be a random variable. We define its expectation (also called mean value) by*

$$\mathbb{E}(X) = \mathbb{E}_{\mathbb{P}}(X) = \sum_{x \in X(\Omega)} x\mathbb{P}(X = x)$$

and its variance by

$$var(X) = var_{\mathbb{P}}(X) = \mathbb{E}((X - \mathbb{E}(X))^2).$$

[See reviewer's comment (4)]

There are some properties of expectations worth mentioning, since they are handy in our works throughout the project.

Theorem 7. *Let* $(\Omega, \mathscr{F}, \mathbb{P})$ *be a probability triple and* X, Y *be random variables. Then we have*

1. *If* $X \leq Y$*, then* $\mathbb{E}(X) \leq \mathbb{E}(Y)$*.*
2. $\mathbb{E}(X + Y) = \mathbb{E}(X) + \mathbb{E}(Y)$ *and for any* $c \in \mathbb{R}$, $\mathbb{E}(cX) = c\mathbb{E}(X)$*.*
3. *If* X *and* Y *are independent and uncorrelated, then* $\mathbb{E}(XY) = \mathbb{E}(X)\mathbb{E}(Y)$*.*

[See reviewer's comment (5)]

Proof. See [6]. [See reviewer's comment (6)] □

Example 8. *We continue exploiting Example 5. The expectation of* X *is*

$$\mathbb{E}(X) = \sum_{n=1}^{6} n\mathbb{P}(X = n) = \frac{1}{6}(1 + 2 + 3 + \cdots + 6) = \frac{7}{2}$$

and its variance is

$$\sum_{n=1}^{6} \frac{1}{6}\left(n-\frac{7}{2}\right)^2 = \frac{35}{12}$$

Tossing a fair dice is an easy yet important example that familiarises us with the formal notation of probability. Now let's put expectation and variance into somewhat a little more practical use.

Theorem 9. *Let X be a random variable with $\mathbb{E}(X) = 0$ and variance σ^2. For all $\lambda > 0$, we have that*

$$\mathbb{P}(X \geq \lambda) \leq \frac{\sigma^2}{\sigma^2 + \lambda^2}$$

Proof. Since $\mathbb{E}(X) = 0$, we have the variance $\sigma^2 = \mathbb{E}(X^2) - (\mathbb{E}(X)^2) = \mathbb{E}(X^2)$. Now consider the random variable $X + a$. We have

$$\begin{aligned}
\mathbb{P}(X \geq \lambda) &= \mathbb{P}(X + a \geq \lambda + a) \\
&\leq \mathbb{P}((X + a)^2 \geq (\lambda + a)^2) \\
&\leq \frac{\mathbb{E}(X^2 + 2aX + a^2)}{(\lambda + a)^2} \\
&= \frac{\sigma^2 + a^2}{(\lambda + a)^2}
\end{aligned}$$

For optimum $a = \dfrac{\sigma^2}{\lambda}$, we have $\mathbb{P}(X \geq \lambda) \leq \dfrac{\sigma^2}{\sigma^2 + \lambda^2}$, as claimed.

[See reviewer's comment (7)] □

For Chebyshev's inequality (see [6]), we have

$$\mathbb{P}(X - \mu \geq \lambda) \leq \mathbb{P}(|X - \mu| \geq \lambda) \leq \frac{\sigma^2}{\lambda^2}.$$

Can we attain a better bound in certain conditions? Yes, we achieved it!

Theorem 10. *Let X be a random variable with $\mathbb{E}(X) = \mu$ and variance σ^2. For all $\lambda > 0$, we have*

$$\mathbb{P}(X - \mu \geq \lambda) \leq \frac{\sigma^2}{\sigma^2 + \lambda^2}.$$

[See reviewer's comment (8)]

Proof.

$$\begin{aligned}\mathbb{P}(X-\mu\geq\lambda) &= \mathbb{P}(X+a-\mu\geq\lambda+a)\\ &\leq \mathbb{P}((X+a-\mu)^2\geq(\lambda+a)^2)\\ &\leq \frac{\mathbb{E}(X-\mu+a)^2}{(\lambda+a)^2}\\ &= \frac{\sigma^2+a^2-2a\mathbb{E}(X-\mu)}{(\lambda+a)^2}\\ &= \frac{\sigma^2+a^2}{(\lambda+a)^2}\end{aligned}$$

For optimum $a=\dfrac{\sigma^2}{\lambda}$, we have $\mathbb{P}(X-\mu\geq\lambda)\leq\dfrac{\sigma^2}{\sigma^2+\lambda^2}$. □

Notice that if we replace $(X-\lambda)$ with $(\lambda-X)$, the proof is still valid. In other words, we have $\mathbb{P}(\mu-X\geq\lambda)\leq\frac{\sigma^2}{\sigma^2+\lambda^2}$.

Corollary 11. *Let X be a random variable, taking non-negative integral values. For $\mathbb{E}(X^2)<\infty$, then*

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

[See reviewer's comment (9)]

Proof. We can see that

$$\mathbb{P}(X=0)=\mathbb{P}(\mu-X\geq\mu)\leq\frac{\sigma^2}{\sigma^2+\mu^2}=\frac{\sigma^2}{\mathbb{E}(X^2)-\lambda^2+\lambda^2}=\frac{\sigma^2}{\mathbb{E}(X^2)}.$$

We are done. □

Note that here is a comparision with Chebyshev's inequality:

$$\mathbb{P}(X=0)=\mathbb{P}(|X-\mu|\geq\mu)\leq\frac{\sigma^2}{\mu^2}.$$

Thus we observe that $\dfrac{\sigma^2}{\mathbb{E}(X^2)}$ is a better bound.

To commence the journey of probabilistic methods, the introduction of the indicator $\mathbb{1}_A$ of an event $A\in\mathscr{F}$ is indispensable.

Definition 12. *Define $\mathbb{1}_A(\omega)=1$ for $\omega\in A$ and $\mathbb{1}_A(\omega)=0$ for $\omega\notin A$.*

Then we have the following equality

$$\mathbb{P}(A)=\mathbb{E}(\mathbb{1}_A),$$

by observing that $\mathbb{E}(\mathbb{1}_A)=0\cdot\mathbb{P}(\mathbb{1}_A=0)+1\cdot\mathbb{P}(\mathbb{1}_A=1)=\mathbb{P}(A)$.

1.2. Graph Theory

Graph theory is the study of graphs, and the origin of graph theory is the paper written by Leonhard Euler on the Seven Bridges of Kőnigsberg in 1736. [See reviewer's comment (10)]

But before our introduction, we shall take the set of natural number $\mathbb{N} = \{1, 2, 3, \dots\}$ and the set of integers $\mathbb{Z} = \{\dots, -1, 0, 1, 2, \dots\}$. For convenience, we also introduce the following notations:

$\{x, y\}$	an unordered pair
(x, y)	an ordered pair
$[n]$	$\{1, 2, 3, \dots, n\}$
$\lvert V\rvert$	number of elements in a finite set V
$\lvert V\rvert^k$	the set of k-element subsets of V

[See reviewer's comment (11)]

A graph G is made up of two elements: vertices and edges, usually denoted by a pair (V, E), where V is a non-empty set of vertices and E is a subset of $|V|^2$. [See reviewer's comment (12)]

We shall then introduce the idea of a subgraph, a clique and an independent set.

Definition 13. *Let $G = (V, E)$ be a graph.*

1. *A subgraph is a graph $G' = (V', E')$ of $G = (V, E)$ where $V' \subset V$ and $E' \subset E$.*

2. *A subgraph $G = (V', E')$ is induced by G on V' if all edges in E connecting vertices in V' are in E'.*

3. *A k-clique in G is a complete subgraph (where vertices are pair-wisely connected) induced by G on k vertices. Define $\omega(G)$ as the size of a biggest clique. i.e.*

$$\omega(G) = \max\{|K| : K \subset V,\ \textit{induced graph } G(K) \textit{ is complete}\}.$$

4. *An independent set in G is a subgraph without any edges induced by G. Define $\alpha(G)$ as the size of a biggest independent set. i.e.*

$$\alpha(G) = \max\{|K| : K \subset V,\ \textit{induced graph } G(K) \textit{ is independent set}\}.$$

For example, for a graph $G = (V, E)$ with $|V| = n$ vertices and $|E| = 0$, we have that $\alpha(G) = n$. We shall show a more colourful and vivid example.

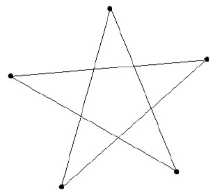


Figure 1. Graph G

Consider the above graph $G = (V, E)$ in Figure 1, with $|V| = 5$ and $|E| = 5$. The black dots are vertices V and the black lines are edges E.

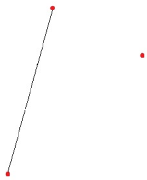


Figure 2. Subgraph G'

This is a subgraph $G' = (V', E')$ of G. We then have the red dots as V', which is elements of V; similarly, the black line is E' as elements of E. Note that a subgraph does not necessarily contain the exact edges that V' are connected to in G.

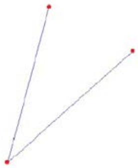


Figure 3. Induced subgraph G''

Compared to a typical subgraph, an induced subgraph contains the exact edges that connect V' in G. In this subgraph $G'' = (V'', E'')$, the blue lines E'' connects exactly all red dot V'', which is a partial mimic of graph G in Figure 1.

FIGURE 4. a 2-clique of graph G

Above we have a 2-clique of G, which is also the case of $\omega(G)$ because we cannot find three certain vertices in G such that they are all connected, i.e. a 3-clique. Thus we have $\omega(G) = 2$.

FIGURE 5. an independent set of graph G

These two purple dots are not connected in G, so they are an indepentdent set of G. Owing to the fact that we cannot find any three unconnected vertices, i.e. an independent set of 3 vertices, we have $\alpha(G) = 2$.

2. Examples of Probabilistic Methods

Paul Erdős' is undoutedly the pioneer of the application of probabilistic methods. He is, in layman's perspective, regarded as a werido. For example, he pharsed "children" as "epsilons", "people who stopped doing mathematics" as "dead", and "to give a mathematical lecture" as "to preach". Despite his interesting personality, his contribution to probabilistic methods is profound. In this chapter, we are going to demonstrate some of the powerful use of probabilistic methods. It is intriguing to see that probabilistic methods can be applied to fields other than probability itself, such as linear algebra and graph theory - our main course.

2.1. Unit Vectors in $\mathbb{R}^n$

The first example is extracted from an exercise of [1] on investigating unit vectors in $\mathbb{R}^n$. We will give our solutions to this exercise, using probabilistic methods.

Theorem 14. *Let $v_1, v_2, \ldots, v_n \in \mathbb{R}^n$ where $|v_i| = 1$ for all $1 \le i \le n$. Then there exist $\epsilon_1, \epsilon_2, \ldots, \epsilon_n \in \{-1, 1\}$ such that*

$$|\epsilon_1 v_1 + \epsilon_2 v_2 + \cdots + \epsilon_n v_n| \le \sqrt{n};$$

and also $\epsilon'_1, \epsilon'_2, \ldots, \epsilon'_n \in \{-1, 1\}$ such that

$$|\epsilon'_1 v_1 + \epsilon'_2 v_2 + \cdots + \epsilon'_n v_n| \ge \sqrt{n}$$

[See reviewer's comment (13)]

Proof. [See reviewer's comment (14)]

Let $v_j = (i_{j_1}, i_{j_2}, i_{j_3}, ..., i_{j_n})$, where $\sum_{k=1}^{n} i_{j_k}^2 = 1$. We want to show that

$$\mathbb{E}\left(\left|\sum_{j=1}^{n} \epsilon_j v_j\right|\right) = \sqrt{n}.$$

[See reviewer's comment (15)]

Note that $\mathbb{E}\left(\left|\sum_{j=1}^{n} \epsilon_j v_j\right|\right) = \mathbb{E}\left(\sum_{k=1}^{n}\sum_{j=1}^{n} \epsilon_j i_{jk}\right)$. By expanding the right hand side, we get

$$\left(\sum_{j=1}^{n} \epsilon_j i_{jk}\right)^2 = \epsilon_1 i_{1k}^2 + \epsilon_1 \epsilon_2 i_{1k} i_{2k} + \cdots + \epsilon_n^2 i_{nk}^2.$$

Since each ϵ_j is independent and $E(\epsilon_j) = 0$, we have

$$\mathbb{E}(\epsilon_{j1}\epsilon_{j2} i_{j_1 k} i_{j_2 k}) = \mathbb{E}(\epsilon_{j1})\mathbb{E}(\epsilon_{j2})\mathbb{E}(i_{j_1 k})\mathbb{E}(i_{j_2 k}) = 0,$$

where j_1 is not equal to j_2. Therefore, by summing everything, we have

$$\mathbb{E}\left[\sum_{k=1}^{n}\left(\sum_{j=1}^{n} \epsilon_j i_{jk}\right)\right]^2 = \mathbb{E}\left[\sum_{k=1}^{n}\left(\sum_{j=1}^{n} \epsilon_j^2 i_{jk}^2\right)\right].$$

Since $\epsilon_j^2 = 1$, we thus have

$$\begin{aligned}
\mathbb{E}\left[\sum_{k=1}^{n}\left(\sum_{j=1}^{n}\epsilon_j i_{jk}\right)\right]^2 &= \mathbb{E}\left[\sum_{k=1}^{n}\left(\sum_{j=1}^{n} i_{jk}^2\right)\right] \\
&= \mathbb{E}\left[\sum_{j=1}^{n}\left(\sum_{k=1}^{n} i_{jk}^2\right)\right] \\
&= \mathbb{E}\left[\sum_{j=1}^{n} 1\right] \\
&= n.
\end{aligned}$$

Thus there exist $\epsilon_1, \epsilon_2, \ldots, \epsilon_n \in \{-1, 1\}$ such that

$$\left|\sum_{j=1}^{n}\epsilon_j v_j\right|^2 \geq n$$

and $\epsilon_1', \epsilon_2', \ldots, \epsilon_n' \in \{-1, 1\}$ such that

$$\left|\sum_{j=1}^{n}\epsilon_j' v_j\right|^2 \leq n.$$

By taking square root from the both sides, the initial claim is agreed. $\square$

2.2. Distribution of Integers on a Matrix

The following exercise is from [1] and motivate us to generalize this problem in Chapter 3. But first we shall present our own solution to this problem.

Lemma 15. *If a number k appears in a matrix for b times, there are in total at least $2\sqrt{b}$ rows and columns that have k in it.*

The proof is quite obvious. [See reviewer's comment (16)] The number of rows and columns with k will be the least if the 'k's are arranged in a square like shape. By direct counting, there are at least $2\sqrt{b}$ rows and columns with k.

Theorem 16. *In an $n \times n$ matrix which each of the numbers $1, 2, \ldots, n$ appears exactly n times, there is a row or column in the matrix with at least $\sqrt{n}$ distinct numbers.*

Proof. Let X be the number of distinct numbers appearing in a row/column and A_k is the event that $k \in \{1, 2, \ldots, n\}$ appears in a row or column, we have that

$$\mathbb{E}(X) = \sum_{k=1}^{n}\mathbb{E}(\mathbb{1}_{A_k}) = n\mathbb{E}(\mathbb{1}_{A_k}) = n\mathbb{P}(A_k)$$

By Lemma 15, there are in total at least $2\sqrt{n}$ rows and columns that have k in it, which means $\mathbb{P}(A_k) \geq \dfrac{2\sqrt{n}}{2n}$. We thus have that

$$\mathbb{E}(X) = n\mathbb{P}(A_k) \geq n\frac{2\sqrt{n}}{2n} = \sqrt{n}$$

We are done. [See reviewer's comment (17)] □

2.3. Tournament

Back to the origin, probabilistic methods are used to solve problems of graph theory. Below we extract two of the earliest theorems proved. [See reviewer's comment (18)] The first one is about them existence of a graph with a special property due to Erdő in 1947, while the second one is done by Szele in 1943 regarding the existence of tournaments containing a large number of Hamiltonian paths. The theorems in this section are well-known and their proofs can be found in [1]. Yet, we rewrite their proofs in our own words and include them in this section because these proofs are the major motivations for achieving our results in Chapter 3 by probabilistic methods.

Definition 17.

1. *A tournament on V is an oriented graph $\overrightarrow{T} = (V, \overrightarrow{E})$ with orientation on all edges of the complete graph on V; for any $x, y \in V$, $x \neq y$, either $(x, y) \in \overrightarrow{E}$ or $(y, x) \in E$, but not both. If $(x, y) \in \overrightarrow{E}$, we say that x is bearing y.*
2. *A tournament $\overrightarrow{T}$ has property S_k, if for every set $K \subset V$ of k vertices, $|K| = k$, there exists $v \in V$ that is beating all $x \in K$, i.e., $(v, x) \in \overrightarrow{E}$ for every $x \in K$.*

Theorem 18. *(Erdős). Let $k, n \in \mathbb{N}, k \geq 2$, be such that*

$$\binom{n}{k}(1 - 2^{-k})^{n-k} < 1.$$

Then there exists a tournament of n vertices that has the property S_k.

Proof. Consider a random tournament on $V = [n] \equiv \{1, 2, \ldots, n\}$ (each edge oriented independently with probability 1/2 (a toss of a fair coin); i.e., all $2^{\binom{n}{2}}$ tournaments occur equally likely). Fix $K \subset V$ of size k; A_k is the event (set of tournaments), for which there is no vertex beating all vertex in K. For any fixed $v \in V \backslash K$, the number of tournaments which v beats all vertices in K is $2^{\binom{n}{2}-k}$ (there are two possible directions for all edges except those linking v and all vertices in K). Therefore, the probability that v beats all vertices in K is 2^{-k} and that v does not beat all vertices in K is $1 - 2^{-k}$. There are $(n - k)$ vertices in $V \backslash K$ and the event that any of them does not beat all vertices in K are mutually independent (they involve different sets of edges). Thus, we have

$$\mathbb{P}(A_K) = (1 - 2^{-k})^{n-k}$$

and thus,

$$\mathbb{P}(S_k \text{ not true}) = \mathbb{P}(\cup_{K \subset V_{|K|=k}} A_K) \leq \sum_{K \subset V, |K|=k} \mathbb{P}(A_K) = \binom{n}{k}(1 - 2^{-k})^{n-k} < 1$$

$\square$

Except the problem concerning S_k property in a tournament, we are going to investigate another tournament problem: Hamiltonian paths in a tournament. We shall first give the definition.

Definition 19.

1. *A path is a non-empty graph $P = (V, E)$ of the form $V = \{x_1, x_2, \ldots, x_k\}$ and $E = \{\{x_1, x_2\}, \{x_2, x_3\}, \ldots, \{x_{k-1}, x_k\}\}$, where all vertices x_i are distinct.*
2. *Let $G = (V, E)$ be a graph. A path is a Hamiltonian path if it contains all vertices of G.*
3. *Let $\overrightarrow{T} = (V, \overrightarrow{E})$ be a tournament. We say that a Hamiltonian path in a complete graph on V is according with the tournament $\overrightarrow{T}$ if each edge of the path is passed in accordance with the orientation given by $\overrightarrow{T}$.*

Theorem 20. *(Szele). There is a tournament $\overrightarrow{T}$ on n vertices that has at least $\frac{n!}{2^{n-1}}$ Hamiltonian paths according with $\overrightarrow{T}$.*

Proof. Consider random tournament $\overrightarrow{T}$ with uniform distribution on all orientations. For a given permutation π on $\{1, 2, \ldots, n\}$, X_π is the indicator of the event that the Hamiltonian path given by π is according to $\overrightarrow{T}$, i.e. the event that all oriented edges $(\pi(i), \pi(i+1))$, $i = 1, \ldots, n-1$, appear in $\overrightarrow{T}$. As the orientation of different edges is chosen independently,

$$\mathbb{E}(X_\pi) = \mathbb{P}(\pi(i), \pi(i+1)) \in \overrightarrow{T} \quad \text{for} \quad i = 1, 2, \ldots, n-1 = \frac{1}{2^{n-1}}$$

Let $X(\overrightarrow{T})$ be the total number of Hamiltonian paths according with $\overrightarrow{T}$. Its expectation is $\mathbb{E}(X) = \sum_\pi \mathbb{E}(X) = \dfrac{n!}{2^{n-1}}$. So, there must be a tournament with at least $\dfrac{n!}{2^{n-1}}$ Hamiltonian paths. $\square$

2.4. Turán's Theorem

The following Turán's Theorem has inspired us to work on the lower bound on $\alpha(H)$ for hypergraphs. Before we plunge into the hypergraph version of Turan's theorem, we shall admire the simple graph version of it.

Theorem 21. *(Turán). For any graph $G = (V, E)$, the size $\alpha(G)$ of the largest independent set satisfies the lower bound*

$$\alpha(G) \geq \frac{|V|^2}{2|E| + |V|}$$

Proof. See [1]. □

The following is a direct consequence of the Turán's Theorem that we extracted from an exercise of [1].

Corollary 22. *Any graph on n vertices with no triangles has at most $n^2/4$ edges.*

[See reviewer's comment (19)]

Proof. Let $G = (V, E)$ and $E^c = \{e \in [V]^2 : e \notin E\}$. By Turan's Theorem, we have

$$\alpha(G) \geq \frac{|V|^2}{2|E| + |V|}.$$

For any graph G, we consider G' such that E^c and $E \notin G'$. Thus, we have

$$\omega(G') \geq \frac{|V|^2}{|V| + 2(\binom{n}{2} - |E|)}.$$

If G' is triangle-free, then we have $\omega(G') \leq 2$. Thus, for G' with n vertices and trianglefree, we have $2 \geq \dfrac{n^2}{n + n(n-1) - 2|E|}$. Rearranging this yields $|E| \leq \dfrac{n^2}{4}$. We are done. [See reviewer's comment (20)] □

3. Our Results

Finally, here comes the long expected part of original works! These results do not stem from randomness or by sheer luck. We have distribution of entries on a cube, colouring on hypergraphs, and independent set on hypergraphs. In essence, Probability and expected value have led us something probably unexpected.

3.1. First Result: On the Distribution of Entries on a Cube

Having solved the problem concerning a matrix in Theorem 16 of Section 2.2, how about entries on a cube? We achieve a result by mimicking the proof of Theorem 17.

Lemma 23. *If a number k appears in a cube for b times, there are in total at least $3b^{\frac{2}{3}}$ rows, columns and aisles that have k in it.*

[See reviewer's comment (21)]

Again, the proof is quite clear. The number of rows, columns and aisles with k will be the least if the "k"s are arranged in a cube like shape. By direct counting, there are at least $3b^{\frac{3}{2}}$ rows, columns and aisles with k.

Theorem 24. *In an $n \times n \times n$ cube which each of the numbers $1, 2, \ldots, n$ appears exactly n^2 times, there is a row/column/aisle in the cube with at least $n^{\frac{1}{3}}$ distinct numbers.*

Proof. [See reviewer's comment (22)] Fix a row/column/aisle. Let X be the number of distinct numbers appearing in this row/column/aisle and A_k is the event that the number $k \in \{1, 2, \ldots, n\}$ appears in this row/column/aisle, then we have that

$$\mathbb{E}(X) = \sum_{k=1}^{n} \mathbb{E}(\mathbb{1}_A) = n\mathbb{E}(\mathbb{1}_{A_k}) = n\mathbb{P}(A_k).$$

By Lemma 23, there are at least $3n^{\frac{4}{3}}$ rows, columns and aisles in the cube that contains the number k, which implies $\mathbb{P}(A_k) \geq \frac{3n^{\frac{4}{3}}}{3n^2} = \frac{n^{\frac{1}{3}}}{n}$. Then we have

$$\mathbb{E}(X) = n\mathbb{P}(A_k) \geq n\frac{n^{\frac{1}{3}}}{n} = n^{\frac{1}{3}}$$

and the result follows. □

Using the idea of Theorem 16, we are able to achieve the following corollary, by considering a matrix with numbers filled in for a not fixed number of times.

Corollary 25. *In an $n \times n$ matrix in which each of the numbers $\{1, 2, ..., a\}$ appears at least b times, there is a row/column in the matrix with at least $\frac{ab^{\frac{1}{2}}}{n}$ distinct numbers.*

Proof. Fix a row/column. Let X be the number of distinct numbers appearing in this row/column and N_i is the event that $i \in \{1, 2, \ldots, a\}$ appears in this row/column. We have that

$$\begin{aligned}\mathbb{E}(X) &= \sum_{i=1}^{a} \mathbb{E}(\mathbb{1}_{N_i}) \\ &\geq a\mathbb{E}(\mathbb{1}_{N_i}) = a\mathbb{P}(A_k), \text{k is the number that appears with the least frequency.}\end{aligned}$$

By using Lemma 23, the total number of rows and columns that contains the number $k \geq 2b^{\frac{1}{2}}$, which means $\mathbb{P}(A_k) \geq \dfrac{2b^{\frac{1}{2}}}{2n} = \dfrac{b^{\frac{1}{2}}}{n}$. We then have

$$\mathbb{E}(X) \geq a\mathbb{P}(A_k) \geq a\frac{b^{\frac{1}{2}}}{n}$$

and the results follows. □

In particular, when we put $a = n$ and $b = n$,we get that there is a row/column in the matrix with at least $\frac{n(n^{\frac{1}{2}})}{n} = \sqrt{n}$ distinct numbers, which is exactly what Theorem 16 says. The idea can also be applied when it comes to a cube, and we get the following corollary.

Corollary 26. *In an $n \times n \times n$ cube which each of the numbers $\{1, 2, \ldots, a\}$ appears at least b times, there is a row/column/aisle in the cube with at least $\frac{ab^{\frac{2}{3}}}{n^2}$ distinct numbers.*

Proof. Fix a row/column/aisle. Let X be the number of distinct numbers appearing in this row/column/aisle and N_i is the event that the number $i \in \{1, 2, \ldots, a\}$ appears in this row/column/aisle. We have

$$\begin{aligned}\mathbb{E}(X) &= \sum_{i=1}^{a} \mathbb{E}(\mathbb{1}_{N_i}) \\ &\geq a\mathbb{E}(\mathbb{1}_{N_i}) = a\mathbb{P}(A_k), \text{k is the number that appears with the least frequency.}\end{aligned}$$

Using Lemma 23, the total number of rows,columns and aisles with $k \geq 3b^{\frac{2}{3}}$, which implies that $\mathbb{P}(A_k) \geq \frac{3b^{\frac{2}{3}}}{3n^2} = \frac{b^{\frac{2}{3}}}{n^2}$. We thus have

$$\mathbb{E}(X) \geq a\mathbb{P}(A_k) \geq a\frac{b^{\frac{2}{3}}}{n^2}$$

and the results follows. □

When we put $a = n$ and $b = n^2$, we get that there is a row/column/aisle in the cube with at least $\frac{n(n^2)^{\frac{2}{3}}}{n^2} = n^{\frac{1}{3}}$ distinct numbers, which is exactly what we have shown in Theorem 24.

3.2. Second Result: Colouring on Hypergraphs

We shall begin with the definition of a n-uniform hypergraph.

Definition 27. *Let V be a non-empty set of vertices.*

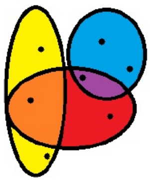
1. *A n-uniform hypergraph is a graph $H = (V, E)$ where $E \subset [V]^n$. Elements in E and V are called the edges and vertices of the graph respectively.*
2. *A graph $H' = (V', E')$ is a subgraph of H if $V' \subset V$ and $E' \subset E$.*
3. *A subgraph $H' = (V', E')$ is induced by H on V' if all edges satisfy the following property: e is in E' if all vertices in e are in V'.*
4. *An independent set in H is a subgraph without any edges induced by H.*

Roughly speaking, a n-uniform hypergraph is just a graph with its edges associated to n vertices. In particular, a 2-uniform hypergraph is nothing else but a conventional graph.

Definition 28. *The average degree of a n-uniform hypergraph H is given by*

$$d(H) = \frac{1}{V(H)} \sum_{x \in V(H)} |\{e : x \in e \text{ and } e \in H\}| = \frac{n|E(H)|}{|V(H)|}.$$

Let's have a brief example of hypergraph. For clear presentation, we use a circle to include all vertices of an edge of n-uniform hypergraph. In other words, n vertices are included in each circle.



Above we have a 3-uniform hypergraph H. Note that a vertex can be included in more than 1 circle; meanwhile, it is possible that a vertex is not included in any circle. The average degree of the above graph is $d(H) = \frac{3|E(G)|}{|V(G)|} = \frac{3(3)}{7} = \frac{9}{7}$.

The following theory concerning 4-colouring problem of hypergraphs is actually a problem created by Niranjan Balachandran when he taught Probabilistic methods in Combinatorics at Caltech in 2010 [2]. We shall rewrite the proof in our own words because it is integral to our work: generalization of k-colouring hypergraph.

Theorem 29. *If $n \geq 4$ and $H = (V, E)$ is a n-uniform hypergraph with $|E| \leq \frac{4^{n-1}}{3^n}$, then it is possible to colour all vertices with 4 colours such that every edge has at least one vertex of each colour.*

Proof. [See reviewer's comment (23)] Let a, b, c and d be four distinct colours. Let X be the number of edges consisting of vertices with less than four differently colours and A be the event that there is an edge consisting of vertices with less than four colours. Denote A_a as the event that there is an edge such that the colour a is missing, and so on for A_b, A_c, and A_d. Then we have that

$$\begin{aligned}
\mathbb{E}(X) &= |E|\mathbb{E}(\mathbb{1}_A) \\
&\leq |E|\mathbb{E}(\mathbb{1}_{A_a} + \mathbb{1}_{A_b} + \mathbb{1}_{A_c} + \mathbb{1}_{A_d}) \\
&= 4|E|\left(\frac{3}{4}\right)^n, \text{ because } \mathbb{E}(\mathbb{1}_{A_a}) = \mathbb{P}(A_a) = (3/4)^n \text{ (resp. } A_b, A_c, A_d.) \\
&= |E|\frac{3^n}{4^{n-1}}
\end{aligned}$$

Thus, if $|E| \leq \dfrac{4^{n-1}}{3^n}$, then we have $\mathbb{E}(X) \leq 1$. Hence, there exists a graph with $X < 1$ or equivalently $X = 0$ as X is non-negative. □

We are able to generalize the above problem, which is our original work.

Theorem 30. *If $n \geq k$, where k is an integer ≥ 2 and H is a n-uniform hypergraph with $|E| \leq \dfrac{k^{n-1}}{(k-1)^n}$, then it is possible to colour all vertices with k colours such that every edge has at least one vertex of each colour.*

Proof. Let $C_1, C_2, C_3, \ldots, C_k$ be k different colours. Let X be the number of edges consisting of vertices with less than k different colours and A be the event that there exists an edge consisting of vertices with less than k different colours. Denote A_{C_1} as the event that there exists an edge without a C_1-coloured vertex, and so on for $A_{C_2}, A_{C_3}, \ldots, A_{C_k}$ respectively. Then we have

$$\begin{aligned}
\mathbb{E}(X) &= |E|\mathbb{E}(\mathbb{1}_A) \\
&\leq |E|\mathbb{E}(\mathbb{1}_{A_{C_1}} + \mathbb{1}_{A_{C_2}} + \mathbb{1}_{A_{C_3}} + \cdots + \mathbb{1}_{A_{C_k}}) \\
&= k|E|\left(\frac{k-1}{k}\right)^n, \text{ because } \mathbb{E}(\mathbb{1}_{A_{C_l}}) = \mathbb{P}(A_{C_l}) = \left(\frac{k-1}{k}\right)^n. \\
&= |E|\frac{(k-1)^n}{k^{n-1}}
\end{aligned}$$

Thus, if $|E| \leq \dfrac{k^{n-1}}{(k-1)^n}$, then we have $\mathbb{E}(X) \leq 1$. Hence, there exists a graph with $X < 1$, i.e. $X = 0$. □

In particular, if we put $k = 2$, it is exactly what Paul Erdős had done (See [2]).

3.3. Third Result: Size of Maximal Independent Set on Hypergraphs

We shall denote the size of a maximal independent set of H by $\alpha(H)$.

Theorem 31. *For a n-uniform hypergraph H of which $d(H) \geq 1$, we have*

$$\alpha(H) \geq |V(H)|d(H)^{-\frac{1}{n-1}}\left(1-\frac{1}{n}\right)$$

Proof. Let S be an induced subgraph of H such that the vertices of S are chosen with a probability $p \in (0,1)$. Then we have

$$\mathbb{E}(|V(S)|) = \sum_{v \in V(H)} \mathbb{E}(\mathbb{1}_v) = |V(H)|p$$

and

$$\mathbb{E}(|E(S)|) = \sum_{e \in V(H)} \mathbb{E}(\mathbb{1}_e) = |E(H)|p^n = \frac{d(H)}{n}|V(H)|p^n.$$

Subtracting these, we have

$$\mathbb{E}(|V(S)|) - \mathbb{E}(|E(S)|) = |V(H)|p\left(1-\frac{d(H)}{n}p^{n-1}\right).$$

Therefore, there exists $S \subset H$ such that the difference between the number of vertices and edges is at least

$$A(p) = |V(H)|p\left(1-\frac{d(H)}{n}p^{n-1}\right). \tag{3.1}$$

Then we remove one vertex from each edge of $E(S)$. By removing at most $|E(S)|$ vertices and the associated edges, we obtain an independent set with at least $A(p)$ vertices, i.e. we have

$$|V(S)| - |E(S)| \geq A(p).$$

We demonstrate the process of removing vertices with an example in Figure 6 on the next page.

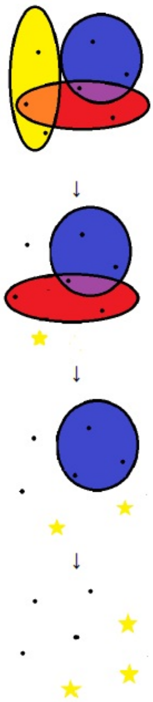


FIGURE 6. The Process of Removing Vertices (Yellow stars refer to the removed vertices)

By differentiating (3.1) with respect to p, we know that $A(p)$ attains its maximum when we choose $p = d(H)^{-\frac{1}{n-1}}$. In this case, we have

$$A(p) = |V(H)|d(H)^{-\frac{1}{n-1}}\left(1 - \frac{1}{n}\right).$$

Hence, we have $\alpha(H) \geq |V(H)|d(H)^{-\frac{1}{n-1}}\left(1 - \frac{1}{n}\right)$, as claimed. □

For an integer $n \geq 2$, consider a n-uniform hypergraph $H = (V, E)$ with k disjoint edges. In other words, we have $e_i \cap e_j$ is empty whenever $e_i, e_j \in E$ and $i \neq j$. Therefore, we have $|V| = nk$ and the size of the maximal independent set $\alpha(H)$ is given by $(n-1)k$. Applying our freshly baked bound in **Theorem 31**, we get

$$\begin{aligned} |V(H)|d(H)^{-\frac{1}{n-1}}\left(1 - \frac{1}{n}\right) &= nk\left(\frac{n|E|}{|V|}\right)^{-\frac{1}{n-1}}\left(1 - \frac{1}{n}\right) \\ &= nk(1)^{-\frac{1}{n-1}}\left(1 - \frac{n-1}{n}\right) \\ &= (n-1)k \end{aligned}$$

Therefore, the bound form our original **Theorem 31** is indeed *sharp*. In particular, we demonstrate it with an example when $n = 3$; let us call it the Happy Graph.

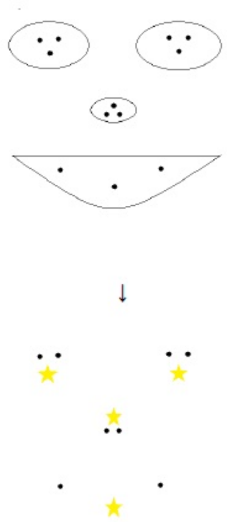


FIGURE 7. The Happy Graph. (Removing vertices from a 3-uniform graph with 4 disjoint edges; yellow stars denote the removed vertices.)

In the Happy Graph (Figure 7), we have a 3-uniform hypergrpah $H = (V, E)$ where $|V| = 12$ and $|E| = 4$. We remove one vertex from each edge in order to obtain the largest independent set. As demonstrated in Figure 7, we need to remove 4 vertices in total. Thus we have $\alpha(G) = 8$ and this agrees with

$$|V(H)|d(H)^{-\frac{1}{n-1}}\left(1-\frac{1}{n}\right) = 12\left(\frac{3(4)}{12}\right)^{-\frac{1}{3-1}}\left(1-\frac{1}{3}\right) = 8.$$

3.4. Bad Graphs: when is our bound not effective?

Despite the sharp example we have demonstrated in the Happy Graph (Figure 7), there are some hypergraphs whose size of maximal independent set is significiantly greater than the bound we obtained in Theorem 31. We are calling these graphs as "bad graphs". In this section, we are going to investigate what characteristics or properties do these "bad graphs" possess.

To begin with this investigation, we try to motivate ourselves by some examples that we interpret as "bad". One such example is given by Figure 8. Of course, we shall begin with the usual graphs (or 2-uniform hypergraph) that we are familiar with.

FIGURE 8. An Example of "Bad Graphs". (Yellow stars refer to the removed vertices)

Let $H_1 = (V_1, E_1)$ be the graph in Figure 8. It is glaring that by removing one vertex, we can obtain an independent set. Therefore, we have $\alpha(H_1) = 8$. According to our bound in Theorem 31, we have

$$\begin{aligned} |V(H_1)|d(H_1)^{-\frac{1}{n-1}}\left(1-\frac{1}{n}\right) &= |V(H_1)|\left(\frac{2|E|}{|V_1|}\right)^{-\frac{1}{n-1}}\left(1-\frac{1}{n}\right) \\ &= 9\left(\frac{2(8)}{9}\right)^{-\frac{1}{2-1}}\left(1-\frac{1}{2}\right) \\ &= 2.53125 \end{aligned}$$

which is much smaller than 8.

For another example, consider the following 3-uniform hypergraph (Figure 9).

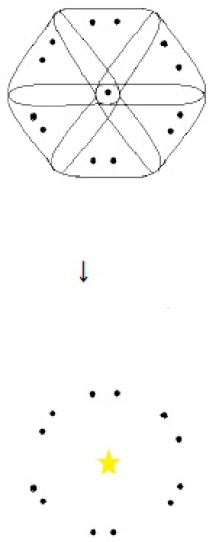


FIGURE 9. Another Example of "Bad Graphs" (Yellow stars refer to the removed vertices)

Let $H_2 = (V_2, E_2)$ be the graph in Figure 9. Then we have $\alpha(H_2) = 12$. However, the lower bound in Theorem 31 yields

$$\begin{aligned} |V(H_2)|d(H_2)^{-\frac{1}{n-1}}\left(1-\frac{1}{n}\right) &= |V(H_2)|\left(\frac{3|E|}{|V_2|}\right)^{-\frac{1}{n-1}}\left(1-\frac{1}{n}\right) \\ &= 13\left(\frac{3(6)}{13}\right)^{-\frac{1}{3-1}}\left(1-\frac{1}{3}\right) \\ &\approx 7.365 \end{aligned}$$

which is significantly less than 12.

From the above two examples (Figure 8 and Figure 9), we observed that bad graphs arises when we remove excessive vertices. In particular, in the first example H_1 (Figure 8), we only have to remove one vertex, whereas in the proof of Theorem 31 we removed

$$\mathbb{E}(|E_1|) = 2.53125 \text{ vertices.}$$

Similarly, in the second example H_2 (Figure 9), practically we only have to remove one single vertex, whereas in the proof of Theorem 31 we removed

$$\mathbb{E}(|E_2|) = 3.6826 \text{ vertices.}$$

To put this observation into a more technical tone, bad graphs arise when some of the vertices have a very high degree (for example, the middle vertices in Figure 8 and Figure 9 respectively) while some, a very low degree. In other words, bad graphs have a great variance among the degree of vertices.

REFERENCES

[1] N.Alon, J.H.Spencer, *The Probabilistic Method*, Third edition, Wiley-Interscience Series in Discrete Mathematics and Optimization, John Wiley & Sons, Inc., Hoboken, NJ, 2008

[2] Niranjan Balachandran, *Lectures Notes on The Probabilistic Method in Combinatorics*, http://www.its.caltech.edu/ñbalacha/Teaching/2009-10/The_Probabilistic_method_Combinatorics.pdf

[3] PF Man, KY Tsui, CM Yeung & KH Yeung, *Mathematics in Action 5B*, Pearson Education Asias Limited, Hong Kong, 2009

[4] R. Kotecký, *Lectures Notes on Probability and Discrete Mathematics at University of Warwick*, unpublished

[5] J.S.Rosenthal, *A First Look at Rigorous Probability Theory*, Second edition, World Scientific Publishing Co. Pte. Ltd., Hackensack, NJ, 2006

[6] S. M. Ross, *A First Course in Probability*, Eighth edition, Pearson Prentice Hall, Upper Saddle River, N.J., 2011

[7] S. M. Ross, *Introduction to Probability Models*, Ninth edition, Academic Press, Amsterdam, 2007

Reviewer's Comments

We organize our comments in two parts: mathematical and expositional.

Comments on the mathematical content

1. The reviewer has comments on the wordings, which have been amended in this paper.
2. The reviewer thinks it is better to call it an assumption rather than an agreement.
3. After the definition the authors may add a sentence like this:
   > Since the sample space Ω is finite, a probability measure can be constructed by defining the probability $\mathbb{P}(\{\omega\})$ for each $\omega \in \Omega$. Then for any $A \subset \Omega$ we have $\mathbb{P}(A) = \sum_{w \in A} \mathbb{P}(\{\omega\})$.
4. The reviewer thinks it is better to use the following definition:
   $$\mathbb{E}(X) = \sum_{\omega \in \Omega} X(\omega)\mathbb{P}(\{\omega\}).$$
   This formulation is more fundamental (as it is an integral on Ω) and avoids writing explicitly the range $X(\Omega)$ of X.
5. Independence has not been defined yet. Also, if X and Y are independent, they are automatically uncorrelated. So it is redundant to say that X and Y are independent and uncorrelated.
6. State precisely the result used (e.g. Theorem 2.5.1 of [ZZ99]). Same for other citations in the paper.
7. State that $a > 0$. (Otherwise, the first inequality in the display may not hold. For example, $1 \geq -5$ but $1^2 < (-5)^2 = 25$.) Also, say that the second inequality follows from Markov's inequality. Finally, before the last sentence ('For *the* optimum...'), add a sentence like this one: 'Since $a > 0$ is arbitrary, we may find the best inequality by optimizing over a.'
8. This follows directly from Theorem 9: $X - \mu$ has mean 0 and variance σ^2.
9. It is not required that X takes integral values. Also, give a label to the bound. Then, we may write: 'Since $\mathbb{E}(X^2) \geq \mu^2$', we see that the bound $(*)$ is better than the one given by Chebyshev's inequality. (The authors may also note that the bound is tight: consider the case where $\mathbb{P}(X = 0) = \mathbb{P}(X = 1) = \frac{1}{2}$.)
10. Cite Euler's paper.
11. $|V|^k$ is not a good notation for the set of subsets of V that have size k. It is rather confusing as $|V|$, the size of V, is a number.
12. It may be helpful to clarify that the graph is undirected and simple.
13. Say that $|\cdot|$ is the Euclidean norm.
14. First mention that now you treat $\epsilon_1, \ldots, \epsilon_n$ as independent random variables that are uniformly distributed on $\{-1, 1\}$ (i.e., Rademacher random variables). Also, the use of $i_{j1}, \ldots, i_{jn}$ for the components of v_j is rather confusing (because i is usually an integer). A better one may be $v_j = (a_{j1}, \ldots, a_{jn})$. Another choice is $\mathbf{v}_j = (v_{j1}, \ldots, v_{jn})$.

15. The displayed equation should be

$$\mathbb{E}\left[\left(\sum_{j=1}^{n} \epsilon_j v_j\right)^2\right] = n.$$

16. While the main idea is clear, the proof is not completely trivial. For example, b may be not a perfect square. Please give the details and give a bound which is an integer.
17. The authors have to explain what is random here. 'Consider a random variable X constructed as follows. First, pick a row or a column randomly. Since the matrix is a square matrix, there are $2n$ choices and each row/column will be picked with probability $\frac{1}{2n}$. Then, we let X be the number of distinct numbers in that row/column...'
18. Probabilistic methods are used in many areas – not only in graph theory. The authors may say 'In this section we present two early results that were established by probabilistic methods'.
19. Corollary 22. Define 'triangle'.
20. The reviewer thinks the authors mean $E^c \in G'$ and $E \notin G'$. Please explain how to use $\omega(G')$ in Turán's theorem.
21. The reviewer believes the bound should be $3b^{1/3}$ instead of $3b^{2/3}$. For example, consider a cube with side 2, and suppose all entries are the same. The total number of entries (k) is 8, and the total number of rows, columns and aisles is 6. Then $6 = 3 \cdot 8^{1/3}$. The statements of the subsequent results should be modified accordingly.
22. Please explain what is random in the proof (same for the proofs of the next two corollaries).
23. Again, what is random here? Note that the hypergraph is fixed and is thus non-random. To use probabilistic methods, one must specify a random object (e.g. a random colouring). (This comment will not be repeated for the remaining results.)

Comments on exposition

The overall structure of the paper is quite easy to follow. After providing in Chapter 1–2 the necessary background and motivations, the main (original) results are stated and proved in Chapter 3.

Chapter 1. The brief review of basic probability is clear. The only comment of the reviewer is that the examples (mainly about throwing a die) are all rather trivial: from them it is not easy to appreciate the need and depth of the rigorous formulation (σ-algebra, countable additivity, etc). The section on graph theory is well written (a brief discussion of modern applications would be interesting for the reader).

Chapter 2. It would be great to explain why probabilistic methods are useful (especially in graph theory). For example, many combinatorial objects (with prescribed properties) are difficult to construct explicitly, and probability provides a relatively easy way to prove existence of these objects.

Chapter 3. The authors have to be more careful when presenting probabilistic proofs. As mentioned above, before talking about things like expectation, the way things are randomized must be described precisely.

There are some grammatical mistakes and misuses of language that could be avoided. Also, there are several 'emotional' sentences which, in the reviewer's opinion, are not appropriate in an academic paper. To give an example, let us discuss one paragraph of the paper in some detail (see the beginning of Chapter 3):

> Finally, here comes the long expected part of original works! These results do not stem from randomness or by sheer luck. We have distribution of entries on a cube, colouring on hypergraphs, and independent set on hypergraphs. In essence, Probability and expected value have led us something probably unexpected.

The first two sentences are somewhat too emotional. The third sentence is grammatically incorrect. The last sentence looks better (lead 'to'), but it is not discussed in the remainder of the paper why the results are 'probably unexpected'. Let us rewrite the paragraph in a way which is more appropriate in this context:

> In this chapter we present our original results achieved only after a lot of effort. They are about the distribution of entries in a cube as well as colourings and independent sets of hypergraphs. All these results will be proved by probabilistic methods, especially the use of expected values. Indeed, a main theme of this paper is that probabilistic methods are powerful tools that allow us to obtain surprising proofs of surprising results.