Graphs and the Probabilistic Method

Bailee Zacovic

Probabilistic Combinatorics Seminar
Prof. Dr. Lisa Sauermann
Universität Bonn

Oct 20, 2023

Contents

1 Introduction

"There is a difference between a thing and talking about a thing."
Kurt Gödel

"That we understand Being is not just actual; it is also necessary. Without such an opening up of Being, we could not be 'human' in the first place."
Martin Heidegger

There is a beautiful interface between combinatorics and philosophy witnessed by problems of existence. Mathematicians could say everything and anything about the features of a mathematical object, such as sheafs of modules or the Riesz representation of a linear functional, but such properties are meaningless without the *existence* of the object itself, upon which whole theories are built. From a metaphysical perspective, existence theorems–which beget being from non-being–are among the most impressive mathematical accomplishments.

How do we resolve existence problems? One naive approach might be explicit construction. Unfortunately, constructive methods can prove exceptionally difficult and/or expensive. Can we bypass direct methods? This is precisely the project of probabilistic combinatorics: *to prove mathematical objects boasting special properties exist by situating oneself in an appropriate probability space of similar objects, and showing the desired properties hold with nonzero probability.*

2 Graph Notations & Preliminaries

Definition 2.0.1. A *graph* is a pair $G = (V, E)$, where $V := V(G) = \{v_1, v_2, \ldots\}$ prescribes the vertex set and $E := E(G) = \{\{v_i, v_j\} : \text{there is an edge joining } v_i \text{ and } v_j\}$ prescribes the edge set.

We offer a few important families of graphs in Figure 1.

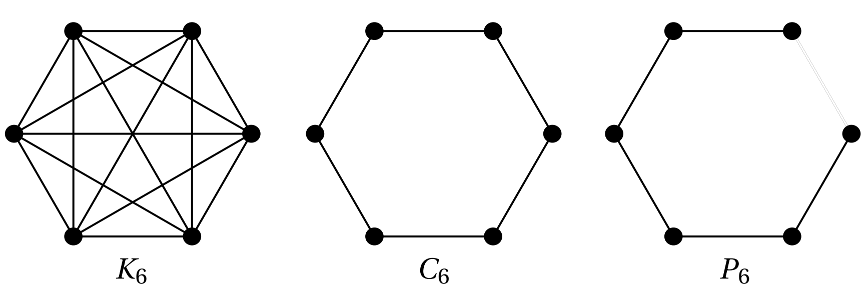

Figure 1: The complete graph K_n, cycle graph C_n, and path graph P_n when $n = 6$.

Another common family of graphs is the *random graph*, which we cannot illustrate here. It is the graph $G = G(n, p)$ on n vertices where the probability of any edge $\{v, w\}$ appearing in the graph is p.

Definition 2.0.2. We call two vertices v and w *adjacent* if $\{v, w\} \in E$.

Definition 2.0.3. The *degree* of $v \in V(G)$, denoted by $\deg(v)$, is given by the number of adjacent vertices.

We'll denote a coloring of a graph G by n colors via a function $c : V(G) \to \{1, 2, \ldots, n\}$. We proceed to a few measurements of sparsity in a graph.

Definition 2.0.4. A *stable set* $S \subseteq V$ is a set of pair-wise non-adjacent vertices.

Definition 2.0.5. The *stability number* of a graph G denoted by $\alpha(G)$ is the cardinality of the largest stable set $S \subseteq V$. That is,

$$\alpha(G) := \max_{S \subseteq V \text{ stable}} |S|.$$

Recall Figure 1. One can check that that,

$$\alpha(K_n) = 1, \ \ \alpha(C_n) = \left\lfloor \frac{n}{2} \right\rfloor, \ \ \text{and } \alpha(P_n) = \left\lceil \frac{n}{2} \right\rceil.$$

Definition 2.0.6. The *chromatic number* of a graph G denoted by $\chi(G)$ is the minimum number of colors needed to color the vertices of G such that $c(v) \neq c(w)$ whenever $\{v, w\} \in E$.

Returning to our guiding examples of K_n, C_n, and P_n, one can similarly check that,

$$\chi(K_n) = n, \ \ \chi(C_n) = \begin{cases} 2 & \text{when } n \text{ is even} \\ 3 & \text{when } n \text{ is odd} \end{cases}, \ \ \text{and } \chi(P_n) = \begin{cases} 1 & \text{when } n = 1 \\ 2 & \text{when } n \geq 2 \end{cases}.$$

Definition 2.0.7. The *girth* of a graph G is the length of the shortest cycle in G.

Whenever G contains no cycles, it is convention to say $\text{girth}(G) = \infty$.
We can again check the following:

$$\text{girth}(K_n) = 3, \ \ \text{girth}(C_n) = n, \ \ \text{and girth}(P_n) = \infty.$$

One might naturally expect a *high* stability number to correspond to a *low* chromatic number and also a *high* girth. In Section 4, we will explore one particularly elegant application of the probabilistic method due to Erdős which will shock this intuition. Before doing so, we will introduce a bit of language and technology necessary for the proof.

3 Asymptotics

We define the following asympototic notation, which is commonly used in computer science to discuss the efficiency and running time of algorithms.

Let $(a_n)_{n\in\mathbb{N}}$ and $(b_n)_{n\in\mathbb{N}}$ be sequences.

Definition 3.0.1. ("Big O") We say a_n is "big O" of b_n, denoted $a_n = O(b_n)$, whenever there exists some constant C and $n_0 \in \mathbb{N}$ such that $|a_n| \leq Cb_n$ for all $n > n_0$. That is, the ratio of their respective growth is bounded.

Definition 3.0.2. ("Little O") We say a_n is "little o" of b_n, denoted $a_n = o(b_n)$, if for all $\epsilon > 0$, there exists n_ϵ such that $|a_n| < \epsilon b_n$ for all $n > n_\epsilon$. That is, relative to b_n, a_n decays to zero.

Two less common notations include "Omega" and "Theta".

Definition 3.0.3. ("Omega") We say a_n is "omega" of b_n, denoted $a_n = \Omega(b_n)$, whenever there exists a constant $c > 0$ and n_0 such that $a \geq cb_n$ for all $n > n_0$. That is, the growth of a_n is bounded below by a positive constant factor of b_n.

Definition 3.0.4. ("Theta") We say a_n is "theta" of b_n, denoted $a_n = \Theta(b_n)$, whenever there exist constants C and c,and some n_0 such that $cb_n \leq a_n \leq Cb_n$ for all $n > n_0$. That is, the growth of a_n is trapped within constant factors from above and below of b_n.

4 Applications of the Probabilistic Method: A story

Recall the notions of chromatic number $\chi(G)$ and girth(G). It is not so hard to generate examples of graphs with *high* chromatic number and *low* girth: for example, K_n. It is also not so hard to think of examples of graphs with *low* chromatic number and *high* girth: for example, C_n or the empty graph.

Question 4.0.1. (A thought experiment) Should there exist graphs with arbitrarily large chromatic number *and* girth? That is, if I feed a machine any pair of positive integers (a, b), will there always exist an output G satisfying $\chi(G) = a$ and girth$(G) = b$?

This question is highly non-trivial, since chromatic number and girth seem almost antithetical measures. In fact, the conviction of the mathematical community pre-1959 was that this certainly was not true. Until Paul Erdős proved otherwise.

Theorem 4.0.2. *(Erdős, 1959) For all k, ℓ there exists a graph G satisfying girth$(G) > \ell$ and $\chi(G) > k$.*

Proof. Fix $\theta < \frac{1}{\ell}$, and let $G = G(n,p)$, $p = n^{\theta-1}$. Let X denote the number of cycles of size at most ℓ. Then $X := \sum_{j=1}^{\ell} X_j$, where X_j denotes the number of cycles of size j. Now, letting $(n)_j := \binom{n}{j} j!$, we compute

$$\mathbb{E}[X_j] = \frac{(n)_j p^j}{2j},$$

which encodes summing over all possible ordered tuples of j vertices $(v_{i_1}, \ldots, v_{i_j})$ and multiplying by the probability that each $\{v_{i_k}, v_{i_{k+1}}\}$ appears in the edge set, where we then divide by $2j$ to disregard the starting point and orientation of the cycle. Next, we observe that,

$$\mathbb{E}[X_j] = \frac{(n)_j p^j}{2j} \le \frac{n^j p^j}{2j} = \frac{n^j n^{\theta j - j}}{2j} = n^{\theta j}/2j.$$

This then implies that

$$\mathbb{E}[X] = \sum_{j=1}^{\ell} \frac{(n)_j p^j}{2j} \le \sum_{j=1}^{\ell} \frac{n^{\theta j}}{2j} = o(n).$$

By the Markov inequality 5.0.6,

$$\mathbb{P}\left[X \ge \frac{n}{2}\right] \le \frac{2\mathbb{E}[X]}{n} = o(1).$$

Setting $X = \left\lceil \frac{3}{p} \ln(n) \right\rceil$, we deduce

$$\mathbb{P}[\alpha(G) \ge X] \le \binom{n}{X} (1-p)^{\binom{X}{2}} < \left(ne^{-p(X-1)/2}\right)^X = o(1),$$

recalling $\binom{n}{X} \le \frac{nX}{X!}$, $(1-p) < e^{-p}$, and the union bound. Taking n sufficiently large such that $\mathbb{P}\left[X \ge \frac{n}{2}\right]$, $\mathbb{P}[\alpha(G) \ge X] < \frac{1}{2}$ ensures the existence of some graph G with fewer than $\frac{n}{2}$ cycles of size at most ℓ and $\alpha(G) < 2n^{1-\theta} \ln(n)$. What remains is to take this object which we *know* exists and modify it as needed to obtain the desired object.

To this end, we break each cycle of size at most ℓ by removing a vertex from each. Call this new graph G^*. Automatically, $\mathrm{girth}(G^*) > \ell$. Also, $|V(G^*)| \ge \frac{n}{2}$, and $\alpha(G^*) \le \alpha(G)$. Hence,

$$\chi(G^*) \ge \frac{|G^*|}{\alpha(G^*)} \ge \frac{n}{2} \cdot \frac{1}{3n^{1-\theta} \ln(n)} = \frac{n^\theta}{6 \ln(n)} \to \infty$$

as $n \to \infty$. This completes the proof.

□

Question 4.0.3. What do these constructed graph G^* actually look like?

Since 1959, many explicit constructions have been discovered. We include the following example for a bit of culture and concreteness.

Example 4.0.4. (The Mycielskian) Consider $\ell = k = 3$. We can construct an explicit graph satisfying the conclusion of Theorem 4.0.2 by forming *the Mycielskian* of C_5 as see in Figure 2.

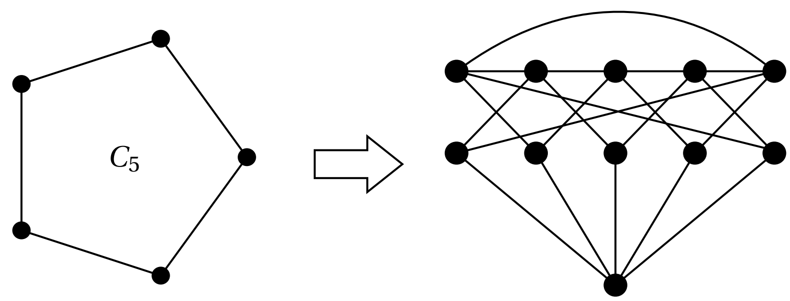

Figure 2: The Mycielskian graph associated to C_5. Its girth and chromatic number exceed $\ell = k = 3$.

5 Probabilistic Bounds

We briefly review a bit of classical probability theory.

Definition 5.0.1. A *random variable*, traditionally denoted X, is a variable whose (numerical) value depends on a random experiment.

Random variables come in two flavors: discrete and continuous.

Example 5.0.2. (Binomial Random Variable) The binomial random variable, denoted $\mathrm{Bi}(n,p)$ counts the number of successes of an experiment (whose probability of success is p and whose probability of failure is $1-p$) performed n times.

Example 5.0.3. (Bernoulli Random Variable) The Bernoulli random variable, denoted $\mathrm{Be}(p) = \mathrm{Bi}(1,p)$ is a special case of the binomial random variable, namely when only a single trial occurs.

Definition 5.0.4. For E some event, the *indicator function* of E (also known as a *zero-one event*) is given by $\mathbb{1}[E] = \begin{cases} 1 & \text{if } E \text{ occurs} \\ 0 & \text{otherwise.} \end{cases}$

As usual, we denote $\mathbb{E}[X]$ the *expected value* and $\mathrm{var}[X]$ the *variance* of a random variable X. We proceed to a few fundemental bounds in probability theory.

Theorem 5.0.5. *(Chebyshev's Inequality, 1867) If var(X) exists, then*

$$\mathbb{P}\left[|X - \mathbb{E}[X]| \geq t\right] \leq \frac{var(X)}{t^2}$$

where $t > 0$.

Chebyshev's inequality effectively bounds the concentratedness of a random variable around its mean from below.

Theorem 5.0.6. *(Markov's Inequality, 1935) If $X \geq 0$ almost surely, then*

$$\mathbb{P}[X \geq t] \leq \frac{\mathbb{E}[X]}{t}$$

where $t > 0$.

Markov's inequality similarly bounds a non-negative random variable's positive deviation from its mean from above.

Definition 5.0.7. The *conditional expectation of a random variable X given some even E* is the value denoted $\mathbb{E}[X|E]$.

Definition 5.0.8. The *conditional expectation of a random variable X given some random variables $Y_1, \ldots, Y_k$* is a function denoted $\mathbb{E}[X|Y_1, \ldots, Y_k]$.

Crucially, while $\mathbb{E}[X|E]$ is a revised *value* in light of the occurrence of E, $\mathbb{E}[X|Y_1, \ldots, Y_k]$ is a revised *function* in light of information accessible in $Y_1, \ldots, Y_k$.

Often, we cannot compute expected values directly. What we can do, in the setting of a discrete probability space, is partition the space into distinct events–the probabilities of which are known–to elucidate an expected value.

Definition 5.0.9. For a probability space $\Omega = E_1 \cup E_2 \cup \ldots$, the *law of total probability* yields the following decomposition of $\mathbb{E}[X]$:

$$\mathbb{E}[X] = \sum_i \mathbb{E}[X|E_i] \cdot \mathbb{P}[E_i].$$

In the special case $X = \mathbb{1}[E]$, we have $\mathbb{E}[X] = \mathbb{P}[E] = \sum_i \mathbb{P}[E|E_i] \cdot \mathbb{P}[E_i]$.

Question 5.0.10. When is a random variable X close to its mean $\mathbb{E}[X]$? Formally, for all $t \geq 0$, how do we bound $\mathbb{P}[X \geq \mathbb{E}[X] + t]$?

By Markov's inequality 5.0.6, we deduce for all $u \geq 0$,

$$\mathbb{P}[X \geq \mathbb{E}[X] + t] = \mathbb{P}\left[e^{uX} \geq e^{u(\mathbb{E}[X]+t)}\right] \leq e^{-u(\mathbb{E}[X]+t)} \cdot \mathbb{E}[uX].$$

When $X = \sum_{i=1}^n X_i$ for $\{X_i\}_{1 \leq i \leq n}$ a collection of independent random variables, we further obtain

$$\mathbb{P}[X \geq \mathbb{E}[X] + t] \leq e^{-u(\mathbb{E}[X]+t)} \cdot \prod_{i=1}^n \mathbb{E}\left[e^{uX_i}\right].$$

Supposing $X = \mathrm{Bi}(n, p)$, we deduce for $\lambda = np = \mathbb{E}[X]$,

$$\mathbb{P}[X \geq \lambda + t] \leq e^{-u(\lambda+t)} \cdot (1 - p + pe^u)^n, \tag{1}$$

where one can check by expanding that $1-p+pe^u = \mathbb{E}[e^{u\mathrm{Ber}(p)}]$. How can we understand the right-hand side of 1? Is it "large"? How does it grow in t?

Fixing $t < n-\lambda$ (which is a sensible restriction, since $X \le n$, i.e. $X \le \mathbb{E}[X]-(n-\lambda)$), we might wonder where the minimum is achieved with respect to e^u. We could check via differentiation:

$$\frac{d}{de^u}\frac{(1-p+pe^u)^n}{e^{u(\lambda+t)}} = 0 \implies e^u = \frac{(\lambda+t)(1-p)}{p(n-\lambda-t)} \text{ is critical,}$$

hence a minima as can be checked. It follows that, for $0 \le t \le n-\lambda$,

$$\begin{aligned}
\mathbb{P}[X \ge \lambda+t] &\le \left(\frac{p^{n-\lambda-t}}{(\lambda+t)(1-p)}\right)^{\lambda+t} \cdot \left(\frac{n(1-p)}{n-\lambda-t}\right)^n \\
&= \left(\frac{1-p}{n-\lambda-t}\right)^{n-\lambda-t} \cdot \left(\frac{\lambda}{\lambda+t}\right)^{\lambda+t} \cdot n^{n-\lambda-t} \\
&= \left(\frac{n-\lambda}{n-\lambda-t}\right)^{n-\lambda-t} \cdot \left(\frac{\lambda}{\lambda+t}\right)^{\lambda+t}
\end{aligned}$$

To recap, we obtain

$$\mathbb{P}[X \ge \lambda+t] \le \left(\frac{n-\lambda}{n-\lambda-t}\right)^{n-\lambda-t} \cdot \left(\frac{\lambda}{\lambda+t}\right)^{\lambda+t}. \tag{2}$$

This is what is called a *Chernoff bound*, and was established in 1952. A Chernoff bound is an exponentially decaying upper bound on the tail of a random variable. It is sharper than Markov and Chebyshev, which are mere power law bounds on decay. The above formulation is not unique; Chernoff bounds come in different versions. Importantly, it furnishes a tight concentration bound of a binomial random variable around its mean.

In practice, however, the bound given in 2 is unwieldy; the behavior of the bound is not manifestly clear. We present the following cleaner formulation, which makes exponential decay at $t \to \infty$ manifest.

Theorem 5.0.11. *For $X = Bi(n,p)$, $\lambda = np$, and*

$$\varphi(z) = \begin{cases} (1+z)\ln(1+z) - z & z \ge -1, \\ \infty & z < -1 \end{cases},$$

then

$$\mathbb{P}\left[X \ge \mathbb{E}[X]+t\right] \le e^{-\lambda\varphi(t/\lambda)} \le e^{-\frac{t^2}{2(\lambda+t/3)}} \quad t \ge 0, \tag{3}$$

$$\mathbb{P}\left[X \le \mathbb{E}[X]-t\right] \le e^{-\lambda\varphi(-t/\lambda)} \le e^{-\frac{t^2}{2\lambda}} \quad t \ge 0. \tag{4}$$

Proof. We have, by 2 for $0 \le t \le n-\lambda$,

$$\begin{aligned}
\mathbb{P}[X \geq \lambda + t] &\leq \left(\frac{n-\lambda}{n-\lambda-t}\right)^{n-\lambda-t} \cdot \left(\frac{\lambda}{\lambda+t}\right)^{\lambda+t} \\
&= e^{\ln\left(\left(\frac{n-\lambda}{n-\lambda-t}\right)^{n-\lambda-t} \cdot \left(\frac{\lambda}{\lambda+t}\right)^{\lambda+t}\right)} \\
&= e^{(\lambda+t)\ln\left(\frac{\lambda}{\lambda+t}\right)+(n-\lambda+t)\ln\left(\frac{n-\lambda}{n-\lambda+t}\right)} \\
&= e^{-(\lambda+t)\ln\left(1+\frac{t}{\lambda}\right)+(n-\lambda+t)\ln\left(1-\frac{t}{n-\lambda}\right)} \\
&= e^{-\lambda\varphi(t/\lambda)-(n-\lambda)\varphi\left(-\frac{t}{n-\lambda}\right)} \\
&\leq e^{-\lambda\varphi(t/\lambda)} \leq e^{-\frac{t^2}{2(\lambda+t/3)}}.
\end{aligned}$$

One can similarly derive Equation 4.

□

Corollary 5.0.12. $\mathbb{P}[X \geq \mathbb{E}[X] + t] \leq e^{-\frac{t^2}{2\lambda}+\frac{t^3}{6\lambda^2}}$.

The following corollary is a multiplicative version of a Chernoff bound, as it bounds relative error rather than absolute error, and is often more practical.

Corollary 5.0.13. *For $X = Bi(n,p)$, $\epsilon > 0$,*

$$\mathbb{P}[|X - \mathbb{E}[X]| \geq \epsilon\mathbb{E}[X]] \leq 2e^{\varphi(\epsilon)\mathbb{E}[X]}, \tag{5}$$

where again $\varphi(\epsilon) = (1+\epsilon)\ln(1+\epsilon) - \epsilon$, and

$$\mathbb{P}[|X - \mathbb{E}[X]| \geq \epsilon\mathbb{E}[X]] \leq 2e^{2\frac{\epsilon^2}{3}\mathbb{E}[X]} \quad \textit{when } \epsilon \leq \frac{3}{2}. \tag{6}$$

One natural question arises:

Question 5.0.14. Which bounds are best?

The answer to this question of course depends upon the given parameters. At least when $p < \frac{1}{2}$, the bounds in 3 and 4 are best.