

## Overview

As a [Lean 4](#) project for the ETHZ course [Formalizing Analysis of Algorithms \(HS25\)](#), we are to formalize an algorithm, or (extremal combinatorics) concept of our choice.

We propose to attempt formalizing a proof of the following theorem, which concerns the existence of (simple) graphs with high girth<sup>1</sup> and high chromatic number<sup>2</sup>:

### Theorem (Erdős 1959)

For any  $k, l \in \mathbb{N}$ , there exists a graph  $G$  with  $\gamma(G) > l$  and  $\chi(G) > k$ .

Such a formalization would follow a variation of Erdős’ original proof (Erdős 1959), akin to how it is commonly found in modern literature (Aigner and Ziegler 1999; Jukna 2011; Zhao 2019; Alon and Spencer 2016; Matoušek and Vondrák 2001).

The proof is often taken as an important example to demonstrate the [probabilistic method](#) (and method of alterations) and touches on the general topics of [graph theory](#), [probability theory](#) and some [\(asymptotic\) analysis](#).

## Planning

From analyzing the proof from various sources in literature, we are able to propose a more detailed outline for this project. Specifically, we list various concepts or hurdles that will likely be faced during formalization, including an informal estimate of the relative effort they might take us at this stage. Note that the listed efforts are biased to start at ‘noteworthy difficulty’, as those will be the most interesting but also increasingly tricky parts of our formalization.

*Notation:*  $(\dagger) \simeq$  “likely of noteworthy difficulty”,  $(\ddagger) \simeq$  “possibly of key difficulty”.

### Varia

- Effective splitting into subproofs and formulation of lemmas.
- Many algebraic transformations of moderate difficulty (sums, inequalities).
- Asymptotic analyses. ( $\ddagger$ )

### Probability theory

- Probability theory in Lean 4 using measure-based definitions (expected value, Markov inequality, random variables, etc.) ( $\dagger$ )
- Defining and using [random graphs](#) à la Erdős–Rényi. ( $\dagger$ )
- Application of the probabilistic method. ( $\ddagger$ )

<sup>1</sup>For a graph  $G$ , the girth  $\gamma(G)$  is the length of the smallest cycle.

<sup>2</sup>For a graph  $G$ , the chromatic number  $\chi(G)$  is the number of colors needed to color the nodes such that no two adjacent nodes share the same color.

### Graph structures and -properties

- Counting the number of cycles in a graph of some length. ( $\dagger$ )
- [Girth](#) of a graph ( $\gamma(G)$ ). ( $\dagger$ )
- [Chromatic number](#) of a graph ( $\chi(G)$ ). ( $\dagger$ )
- Size of the largest [independent set](#) of a graph ( $\alpha(G)$ ). ( $\dagger$ )
- Proving properties of induced subgraphs. ( $\dagger$ )

### On using Mathlib

For probability, most libraries appear to be geared towards continuous probability distributions, in contrast to discrete ones, which is what is needed here. We thus decide to build the needed probability theory from scratch using the Mathlib [Measure Theory library](#) and combining them (from scratch) with the graph concepts mentioned above.

Mathlib4 ostensibly supports [simple graphs](#), [graph coloring](#), [graph girth](#), and [sub-graphs](#), which we will try to make use of. There appears to be no support for independent sets, which we will have to formalize from scratch.

[Limits](#) seem to be supported by Mathlib and we will make use of such.

### References

- Aigner, Martin and Günter M Ziegler (1999). “Proofs from the Book”. In: *Berlin. Germany* 1.2, p. 7.
- Alon, Noga and Joel H Spencer (2016). *The probabilistic method*. John Wiley & Sons.
- Erdős, Paul (1959). “Graph theory and probability”. In: *Canadian Journal of Mathematics* 11, pp. 34–38.
- Jukna, Stasys (2011). *Extremal combinatorics: with applications in computer science*. Vol. 571. Springer.
- Matoušek, Jiří and Jan Vondrák (2001). “The probabilistic method”. In: *Lecture Notes, Department of Applied Mathematics, Charles University, Prague*.
- Zhao, Yufei (2019). “18.218 Probabilistic Method in Combinatorics, Spring 2019”. In.