Network Science

– Homework 1 –

The Importance, Complexity and Fragility of Mathematics

November 23, 2018

Instructor:

Tommaso Erseghe

Leonardo Badia

Student:

Ettore Mariotti

Figure 1: Euclid's Elements: Each vertex is a proposition and is connected to the propositions needed to prove it

1 Introduction

Mathematics is the field on which every other quantitative discipline stand on and, in this sense, it can be thought as the foundation of science. But what is the foundation of Math? There exist different schools of thought about it but one of the most popular links it to logic, where starting from postulates and axioms (true by the very definition) one can construct other propositions provably true (Theorems and Lemmas). In this sense Mathematics can be viewed as a network, in which nodes are propositions and node B is connected with a direct edge to node A (B→A) if B uses A in its proof.

With these ideas in mind, one can think of analyzing math with the eyes of a network scientist, looking for patterns in the science of discovering true patterns. Now, while it is non-trivial to build a graph of modern mathematics (due to its vastness), one can start with what have been its founding core: Euclid's Elements book.

Euclid's Elements (dating back to ancient Greeks) is the first known collection of rigorous logical inference in terms of definition-lemma-theorem. It lays down the foundation for number theory, planar and solid (Euclidean) geometry, as well as other branches of math. Dated to 300BC, it is a systematic compendium of the results of Greeks philosophers that lasted the test of time: for millenniums it has been the main resource of reference and influenced people like Nicolaus Copernicus, Johannes Kepler, Galileo Galilei and Sir Isaac Newton all of which used its ideas in their ground-breaking works.

It consists of 13 books, with a total of 465 propositions, 5 "Common Notions" (which in modern terms we now refer to as axioms) and 5 postulates[1]

[1]Under careful examination it can be seen that there exist another Common Notion not explicitly listed but used in various proofs, namely if $x > y$ then $x + z > y + z$.

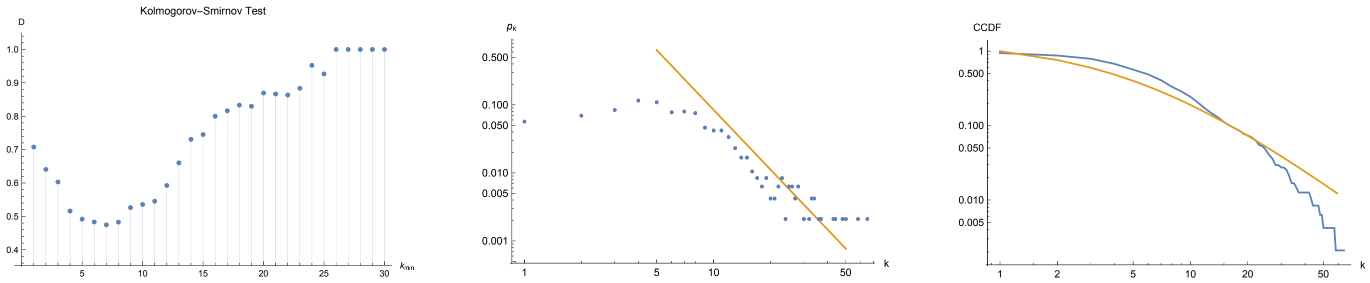

Figure 2: Power-law exponent estimation with Kolmogorov-Smirnov Test (left) and plot of the model with p_k (center) and the CCDF (right)

2 Network Analytics

The network consists of 476 nodes and 2032 edges. Being the graph constructed such that if B→A means that B uses A in its proof we can interpret the in-degree as a measure of how many immediate theorems the proposition allows to prove (a rough measure of usefulness in the global view). On the other hand, the out-degree represent how many theorems you need to invoke in order to prove the proposition (a first approximation of the proof difficulty).

2.1 Degrees Distribution

If we consider the undirected graph, we can see from the scatter plot of the degree distribution p_k that it can be modeled with a power-law function. This manifest a scale-free structure. A maximum-likelihood estimate of the exponent coefficient can be computed with the formula

$$\gamma_{ML} = 1 + \frac{\sum_i 1}{\sum_i \log\left(k_i/k_{min}\right)}$$

where k_{min} is the small degree cutoff. In other word it is the parameter that take in account for the initial plateau effect of p_k. k_{min} have been chosen as the one that minimized the Kolmogorov-Smirnov Test distance between the empirical CCDF and the modeled one defined as

$$CCDF_k = 1 - \frac{\zeta(\gamma, k)}{\zeta(\gamma, k_{min})}$$

Where ζ represent the Riemann Zeta function. The results of the procedure where $k_{min} = 7$ and $\gamma_{ML} = 2.919$, testifying a scale-free regime of the network (figure 2).

Similar analysis have been performed for the in- and out-degree distribution. The results were $k_{min}^{in} = 1$, $\gamma_{ML}^{in} = 1.89$ for the in-degree (Kolmogorov test didn't manifest a nice minimum) and $k_{min}^{out} = 6$, $\gamma_{ML}^{out} = 4.70$ for the out-degree. So we could assert that the "in" network doesn't nicely follow a power law while the "out" directed network is in random network regime ($\gamma > 3$).

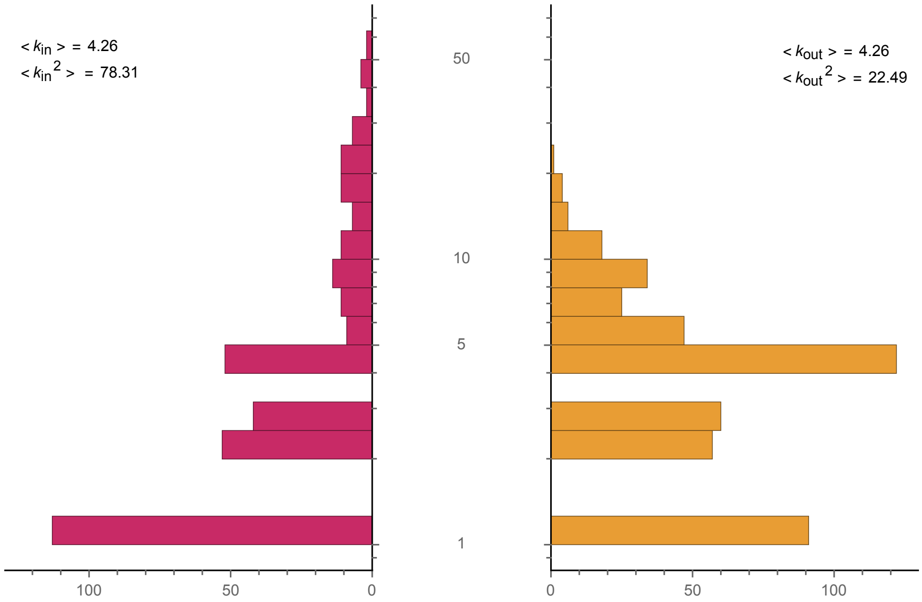

Figure 3: Paired log-binned histograms for in-degree (left) and out-degree (right) distributions

It is not trivial to interpret these result, but it is nonetheless interesting to have a compared look at in- and out-degrees distribution as in figure 3. While the first moment is the same, the second moment is really different. The out-degree might manifest this “boundedness” because it is humanly difficult to come up with an extremely long proof. It is easier to construct intermediate shorter lemmas. On the other hand, once a Theorem is given, its usefulness (out-degree) can occasionally be extremely high.

2.2 Assortativity

A graph is defined to be assortative or disassortative if there is either a positive or a negative correlation between the nodes degrees and the degrees of their neighbors. In order to estimate this, we can try to fit an exponential line on the scatterplot of node degrees k vs. node neighbor degrees $k_{nn}(k)$.

As shown in figure 4, with a $\mu = -0.05$ the network is in a neutral regime. So it is neither assortative nor disassortative. This is in line with the analytical prediction that for a graph with $\gamma = 3$:

$$k_{nn}(k) \sim \frac{m}{2} \cdot \log(N)$$

Making it independent from k and thus neutral. It is also possible to inspect in- and out-degree assortativity, and the plots of figure 4 report a $\mu_{in} = 0.107$ and a $\mu_{out} = 0.217$, displaying a weakly assortative behavior.

So even though with this approach the network as a whole seems neutral, the weakly assortative correlations of both the in and out degrees suggest us that there might actually be assortativity in math: useful theorems slightly tend to open the door for other useful theorems and theorems difficult to prove tend to be proven by other difficult theorems (in line with the fact that a very difficult proof is usually broken in

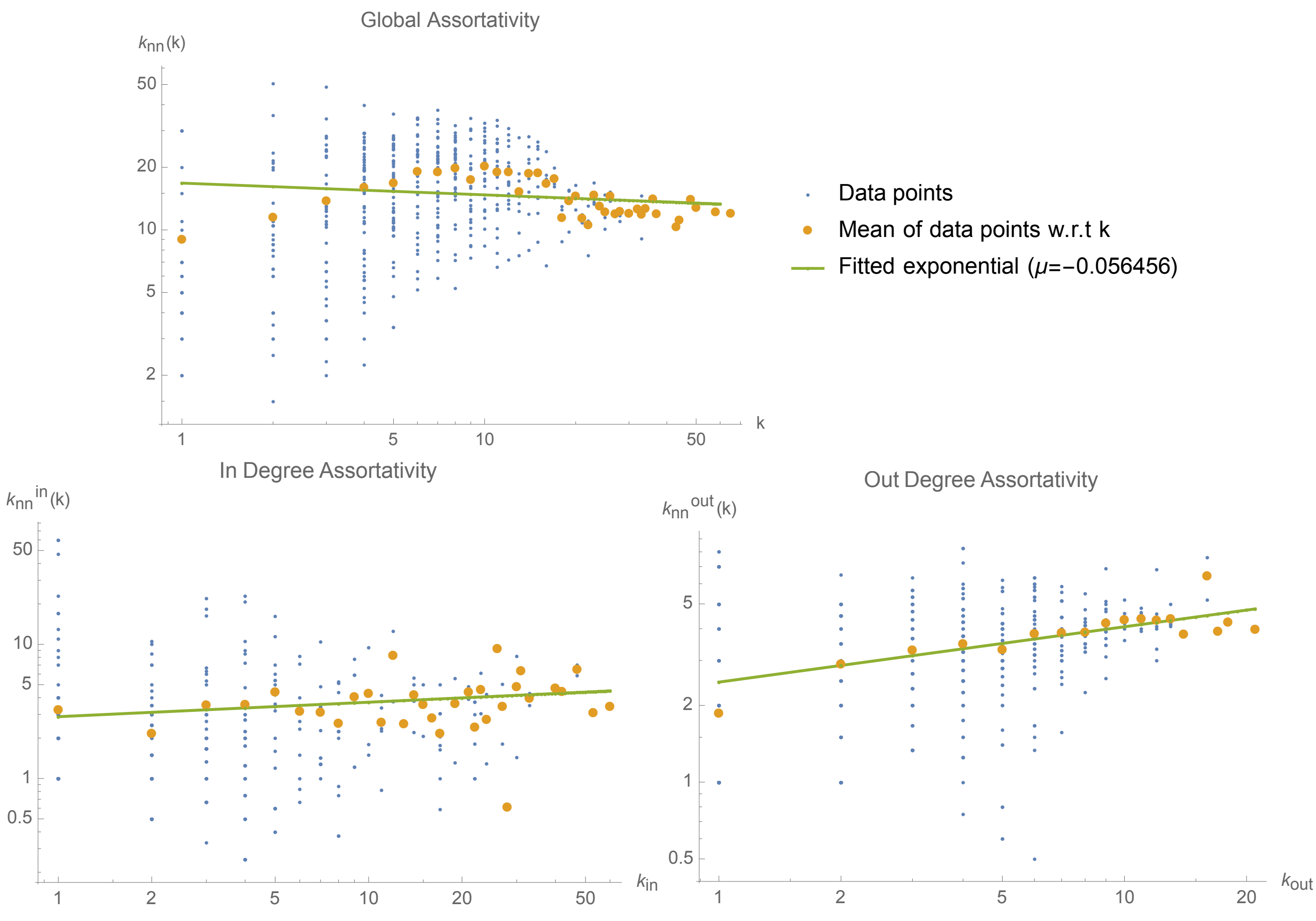

Figure 4: Assortativity of the global degree (up), in degree (left) and out degree (right). $\mu = -0.05$, $\mu_{in} = 0.107$, $\mu_{out} = 0.217$

smaller and more manageable lemmas). It should be noted nonetheless that these values are not extremely significative and they indicate only a slight tendency.

3 Theorems' Attractiveness

In the context of evolving networks, Bianconi-Barabasi proposed a model in which the probability that a new node j connected a node i is proportional to k_i (the degree of i) and to a "fitness factor" η_i, an intrinsic attractive property of node i.

Even though the book is a collection of results whose historical date is unclear, we can think of Euclid's Element as a growing network by sorting the theorems with the order of their appearance in the opera. Given the Axioms and Postulates at time t=0, we can think of Math as a growing network in which every node connects to the already proved ones.

This temporal evolution view allows us to investigate the fitness η_i of each node. The model state that:

$$k_i(t) = m \cdot (t/i)^{\eta_i/C}$$

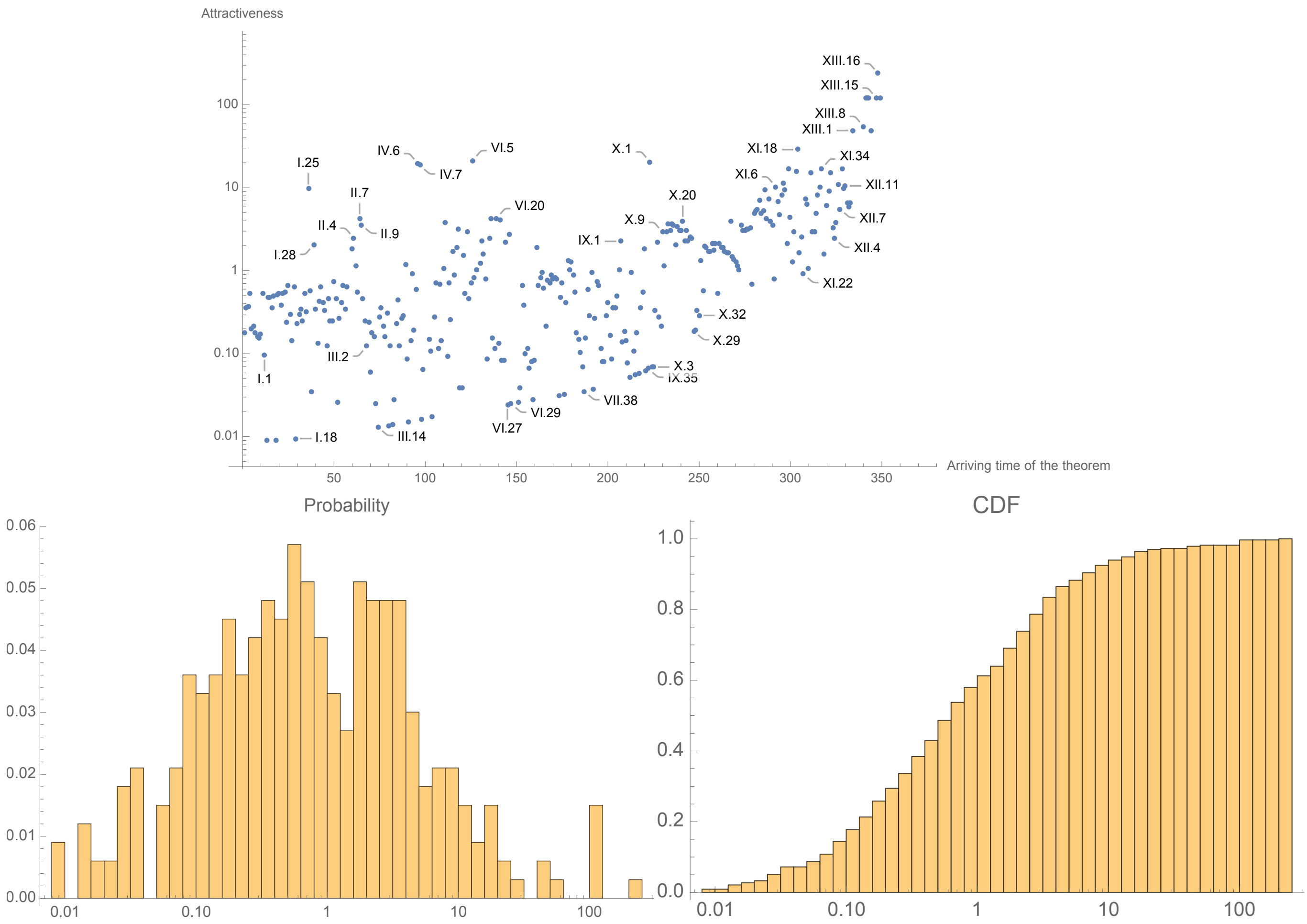

Figure 5: "Growing Attractiveness" of each theorem. Probability distribution (left) and cumulative distribution function (right) of the η_i/C values (using log-binning) for the in-degree

Where k_i is the degree of node i, t is the current time instant, m is the number of nodes added at each time step and C is a constant. We can thus estimate the ratio η_i/C and use that as our measure of "attractiveness". Given our particular case, we are interested in the evolution of the in-degree, as it represents the effective usefulness of a proposition in proving later ones, a kind of "Growing Attractiveness". For this reason we will going to replace $k_i(t)$ with $k_i^{in}(t)$ in the previous formula. For numerical reasons, we will try to estimate the fitness of those nodes that have an in-degree bigger than zero. If from one perspective the existence of zero in-degree vertexes is somehow inevitable in a finite network of this kind, the fact that they are 127 ($\approx 26.6\%$ of the total graph) provokingly suggests that many propositions could be looked at as "curiosities".

In figure 5 we can find a visual representation of the result. Taking into account the fact that the estimated attractiveness is noisier for the later theorems (for the fewer available statistics) it is possible to see that its logarithm follow a normal distribution. Cramér-von Mises test statistic was run to validate this intuition and confirmed it with a p-value of 0.0589.

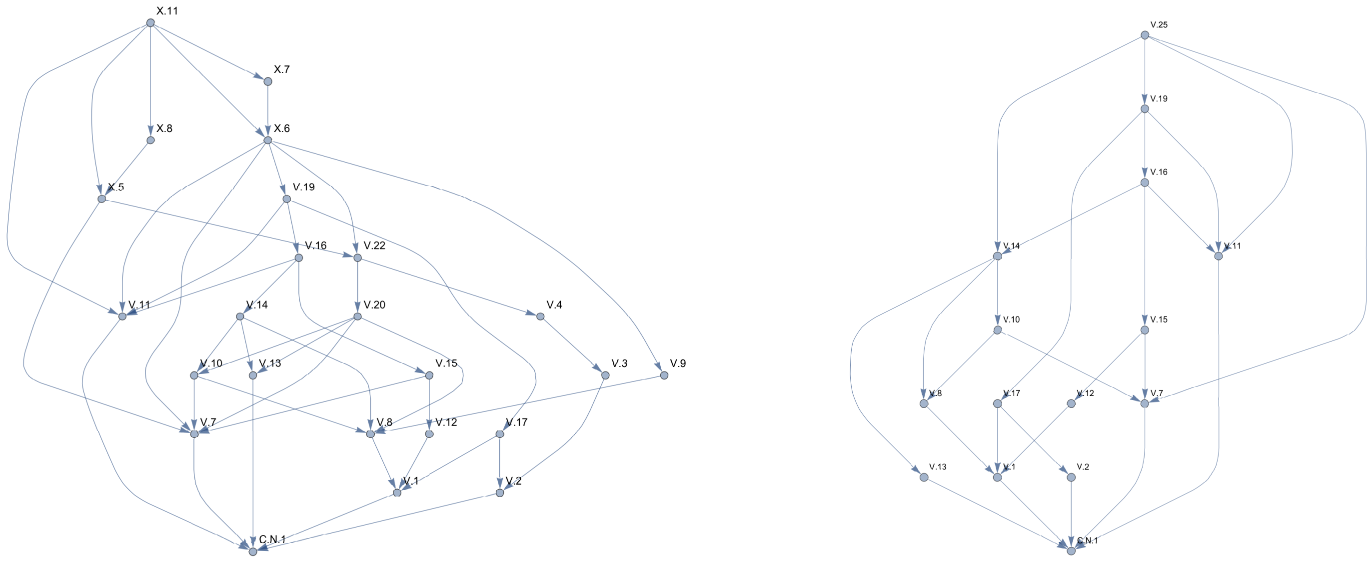

Figure 6: Examples of the complete proof graph for theorem 11 of book X counting 45 edges (left) and theorem 25 of book V counting 26 edges (right))

4 Difficult Theorems (to prove)

Up until now, we have considered the out-degree of a node as a measure of the difficulty of its proof, but that's a limited perspective. At a system-wise level, the difficulty for proving a theorem is better captured by observing the path that from a proposition descend down to the axioms and postulates. This multi-path can be computed by executing a breadth search first algorithm starting from the node of interest and exploring the graph until an endpoint is reached (a non-empty set of postulates or axioms in our case). A better measure of the complexity of a proof would then be the edge count of this "proof graph". Some examples are shown in figure 6.

The proposition with the longest proof (with 867 edges) is Theorem 18 of book XIII which demonstrates that there exist only five platonic solids:

"I say next that no other figure, besides the said five figures, can be constructed which is contained by equilateral and equiangular figures equal to one another."

Or in other words, all the 3D regular polyhedra that can be inscribed in a sphere are the cube, octahedron, tetrahedron, icosahedron, and dodecahedron.

4.1 Most difficult books

Since the opera is organized in books that roughly capture branches of mathematics, an interesting thing is observing the distribution of proof-complexity book-wise. This is reported in figure 7.

The most complex books are the:

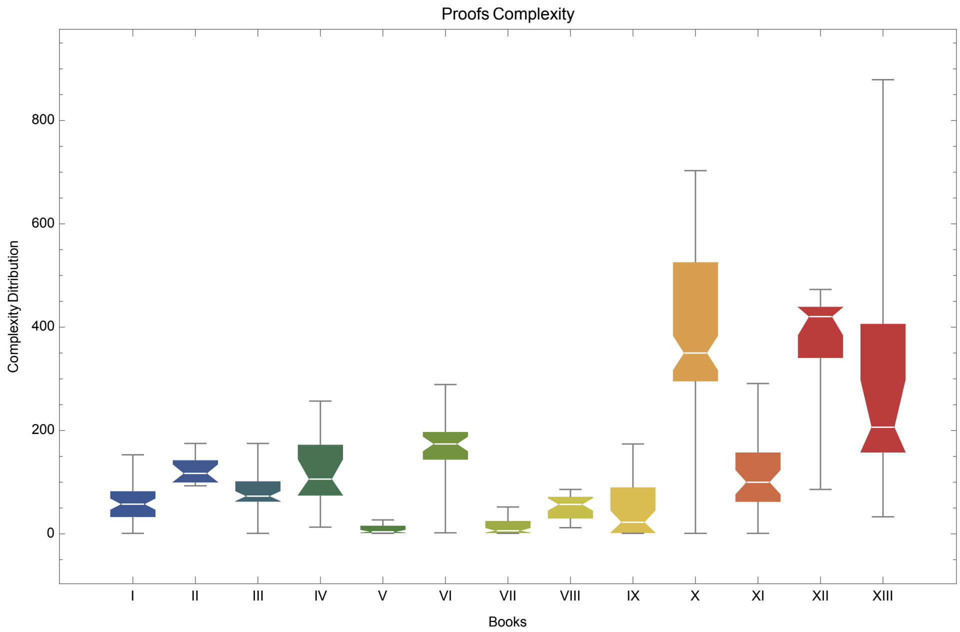

Figure 7: The distribution of proof complexity of each book

- **Book X** where among others Euclides proves the irrationality of the square roots of non-square integers, like $\sqrt{2}$, and classifies the square roots of incommensurable lines into thirteen disjoint categories. He also gives a formula to produce Pythagorean triples.
- **Book XIII** where constructs the five regular Platonic solids inscribed in a sphere and compares the ratios of their edges to the radius of the sphere.

It is reasonable to say that these are the most difficult books because they include a set of very abstract propositions far away from the simple concepts of his axiomatic system.

5 Breaking Math

All the nodes of this network are true propositions, either by the presence of a formal proof (Theorems and Lemmas) or by the fact that we faithfully accept them as true. So our network is apparently fault-resistant for its special construction. But is it really?

5.1 The Fifth Postulate

Of all the accepted-by-definition propositions given in Euclid's Elements, the fifth postulate is the least self-evident:

That, if a straight line falling on two straight lines makes the interior angles on the same side less than two right angles, the two straight lines, if produced indefinitely, meet on that side on which are the angles less than the two right angles.

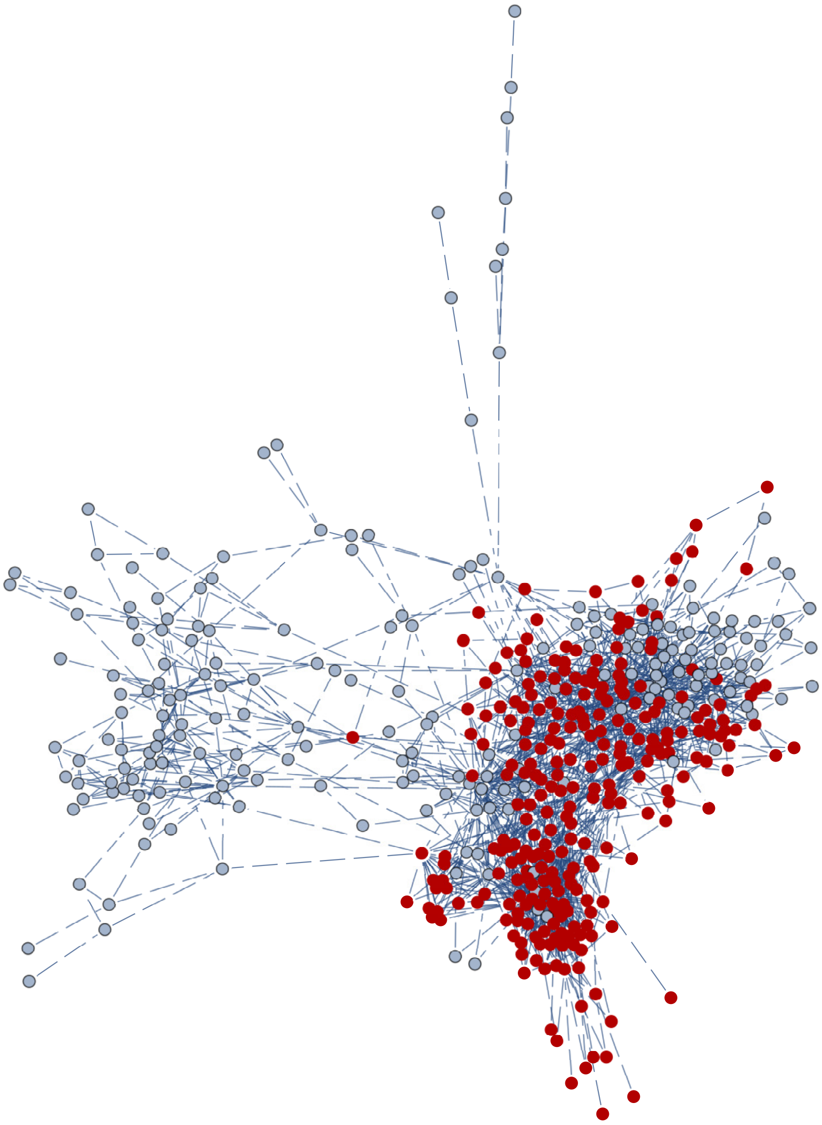

Figure 8: The subgraph highlighted in red represent are all the proposition that need the fifth postulate

Or, equivalently, it claims that there only exists one line parallel to another that passes through an external point. If the order the postulates were listed in the Elements is significant, it indicates that Euclid included this postulate only when he realized he could not prove it or proceed without it. For two thousand years, many attempts were made to prove the fifth postulate from the other four, many of them being accepted as proofs for long periods until the mistake was found.

This is of real interest because if one decides not to accept "parallel postulate", all the propositions that rely upon it cannot be considered as true anymore. In order to quantitatively investigate its impact, it is possible to identify this subgraph by computing a breath search first algorithm starting from the postulate.

As shown in figure 8, the fifth postulate is responsible for the veracity of more than 55% of the entire system.

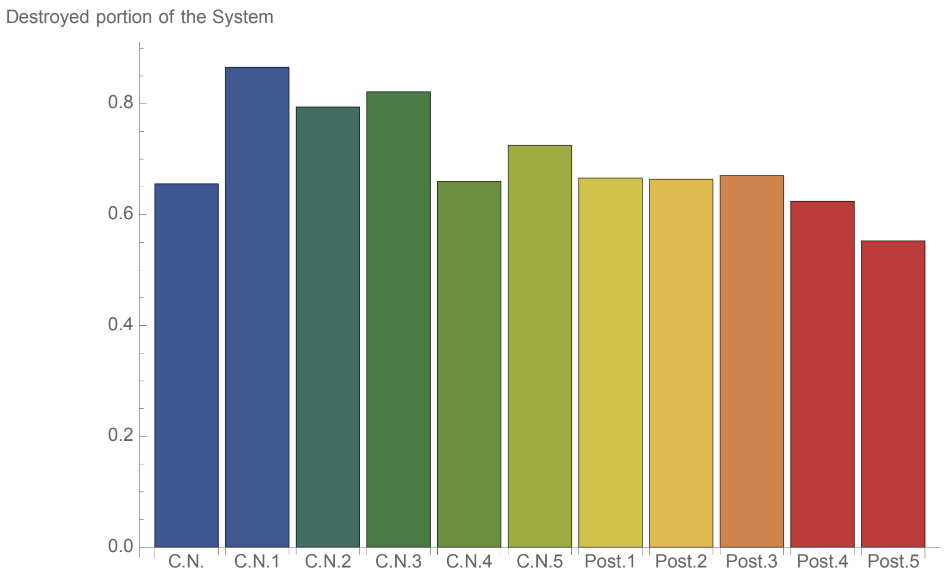

Figure 9: Axioms and Postulates with their relative "apocalypse rank"

5.2 Apocalypse Ranking

So, did Euclid really sorted the postulates according to their importance? The computation done on the fifth postulates can easily be extended to all the others. In this way we can formulate a kind of "apocalypse ranking", which represent the importance of an assumption based on the fraction of theorems that rely on it. As can be seen in figure 9, Euclid roughly captured the right order of importance of his postulates (even though the third have a marginally larger score), but not the order of his common notions.

5.3 Non-Euclidean Geometries

If the non-demonstrability of the fifth postulate could be looked at as a weakness, it gives instead a great opportunity to look at math from a different perspective: one of the many possible self-consistent axiomatic systems.

If by not accepting the fifth postulate we implicitly discard more than half of Euclidean geometry, on the other hand we can open the door to new extremely rich and powerful systems like Elliptic and Hyperbolic geometry. These theories were of key importance for developing new understandings of the world (the theory of General Relativity) and for many other real case applications (for example the design of routes of planes and ships).

If not accepting the least "apocalyptically important" lead to these results, we can only wonder what could be discovered if we decided to discard some of the others...

References

[1] Albert-Laszlo Barabasi, Marton Posfai, *Network Science*. Cambridge: Cambridge University Press, 2016.

[2] Stephen Wolfram, *A New Kind of Science*. Wolfram Media, 2002

[3] Wolfram Research, "Theorem Network from Euclid's Elements" from the Wolfram Data Repository (2017)
https://doi.org/10.24097/wolfram.61787.data