Mathematical Diversity and Elegance in Proofs: Who will be the next Renaissance man?

Brendan W. Sullivan
Carnegie Mellon University
Math Grad Student Seminar

April 18, 2012

Abstract

It has been claimed by some mathematicians that Bernhard Riemann was the last great "Renaissance man" of mathematics, in that he was quite knowledgeable of–and, indeed, proved major theorems in–several seemingly disjoint branches of the subject. One could argue for other notable exceptions from the 20th century–perhaps John von Neumann, Bertrand Russell, even Douglas Hofstadter–but the fact remains that it has become genuinely more and more unlikely (converging towards impossible, even) to hold this polymathic position. Consider this hour-long seminar my humble attempt to help us all along the path towards this esteemed and noble title. By the end of the hour, we will have proven some interesting and fundamental results in number theory, graph theory, geometry and combinatorics, using techniques from topology, probability theory, linear algebra, and analysis! Nothing will be particularly difficult or advanced; rather, we seek to celebrate the diversity of mathematics and the beauty and elegance inherent to some theorems and their proofs.

"If only I had the theorems! Then I should find the proofs easily enough."
—Bernhard Riemann

"I have tried to avoid long numerical computations, thereby following Riemann's postulate that proofs should be given through ideas and not voluminous computations."
—David Hilbert

1 Infinitude of Primes: Number Theory via Topology

This ingenious proof (and the original motivation to construct this talk, in fact!) is due to Israeli-American mathematician Hillel (Harry) Fürstenberg, and was published in 1955 [1] when he was a 20-year old undergraduate at Yeshiva University. He has since gone on to be known for his applications of probability and ergodic theory to number theory and Lie groups. In particular, he proved Szemerédi's Theorem, concerning minimal density to guarantee arithmetic progressions, using ergodic theory; Szemerédi's original, ingenious proof was a direct generalization of previous combinatorial arguments for small cases of the statement.

First, recall the statement of this theorem–one of the most fundamental ideas in all of mathematics, really–and Euclid's original proof, as translated from Book IX, Proposition 20 of his *Elements*.

Theorem 1. *There are infinitely many prime numbers.*

Proof. Consider any finite list of prime numbers, say $p_1, p_2, \ldots, p_k$. Define $P = p_1 p_2 \cdots p_k$ to be the product of those primes, and let $q = P + 1$. We claim that we have now identified at least one more prime not contained in the finite list first considered. We have two cases:

- If q is prime, then this is a new prime, since $q > p_i$ for any i.
- If q is composite, then some prime p divides it. We claim p cannot be any of the p_i primes in the original list. If it were, then p divides both P (since p is in the product) and q (by assumption), so p divides their difference, which is 1. This is clearly not possible. Thus, p is a new prime not in the original list.

In either case, we have shown there is at least one more prime. Since this holds for an arbitrary list of finite numbers, there are infinitely many primes. □

Fairly ingenious, right? Notice that this is *not* a proof by contradiction, but it does contain a *reductio ad absurdum* argument within one of the cases.

Now, let's go through Fürstenberg's proof. You can view his original paper here.

Proof. We will define a topology τ on the set of integers $\mathbb{Z}$ known as the *evenly spaced integer topology*. Specifically, we define a *base* for the topology. Recall that a topology is a set of subsets of a set $\mathcal{S}$ that are defined to be *open*; it must satisfy:

1. $\varnothing$ and $\mathcal{S}$ are open
2. any (arbitrary) union of open sets is open
3. any finite intersection of opens sets is open

A base is a set of subsets of the larger space, and it is used to create a topology by specifying that open sets are unions of sets in the base. We say that a base *generates* a topology.

In this specific instance, our base is the set of all *arithmetic progressions*. An arithmetic progression is a set of integers of the form

$$S(a, \lambda) = \{a + k\lambda \mid \lambda \in \mathbb{Z}\}$$

for any $a, \lambda \in \mathbb{Z}$. We then define

$$B = \varnothing \cup \bigcup_{a,\lambda\in\mathbb{Z}} S(a, \lambda)$$

and let τ be the topology generated by this base. (Notice that we include $\varnothing \in B$ so that $\varnothing$ is open by definition.)

For illustration, let's consider some open sets in this topology. Any arithmetic progression is, itself, an open set, as is any union of arithmetic progressions For instance, the following are all open sets in τ:

$$\begin{aligned}
\{\dots, 1, 6, 11, 16, 21, 26, 31\} &= S(1, 5) = \mathbb{Z} \setminus \bigcup_{b\in\{2,3,4,5\}} S(b, 5) \\
\{\dots, 2, 5, 8, 11, 14, 17, \dots\} &= S(2, 3) = Z \setminus (S(1, 3) \cup S(3, 3)) \\
\{\dots, 1, 2, 4, 5, 7, 8, 10, 11, \dots\} &= S(1, 3) \cup S(2, 3) = \mathbb{Z} \setminus S(3, 3) \\
\{\dots, 1, 3, 5, 6, 7, 9, 11, 13, 15, 16, \dots\} &= S(1, 2) \cup S(1, 5) \\
\{\dots, 3, 4, 8, 12, 13, 16, 18, 20, 23, 24, 28 \dots\} &= S(3, 5) \cup S(4, 4) \\
\mathbb{Z} &= S(0, 1)
\end{aligned}$$

The last line (plus $\varnothing \in B$) verifies property (1) of a topology. The fact that B is a base guarantees property (2) holds. We now prove that property (3) holds. For illustration, we keep the following example in mind, which is constructed from the examples above:

$$S(1, 5) \cap S(2, 3) = \{-4, 11, 26, 41, \dots\} = S(11, 15) \qquad =$$

Another interesting example, which is much harder to think about but very much true, is presented below:

$$\begin{aligned}
[S(1, 3) \cup S(2, 3)] \cap [S(1, 2) \cup S(1, 5)] &= \{\dots, 1, 5, 7, 11, 13, 16, 17, 19, 23, \dots\} \\
&= S(1, 6) \cup S(5, 6)
\end{aligned}$$

This sub-proof relies on the following important property of topological bases:

$$\forall B_1, B_2 \in B.\ \forall x \in B_1 \cap B_2\ .\exists B_3 \in B. \quad (x \in B_3\ \wedge\ B_3 \subseteq B_1 \cap B_2)$$

When this property holds, we can define $B_1 \cap B_2$ as an arbitrary union of basic open sets by just unioning over all $x \in B_1 \cap B_2$, and for each such x, finding a basic open set containing x that "fits inside" the intersection. (For an example, think about the fact that the set of all open intervals is a base for the standard topology on the real line. In that case, we can always squeeze an open interval around any point contained in the intersection of two open intervals; namely, that intersection, itself!) Considering general open sets, we can just write them as a union of basic open sets, and then consider the individual intersections of those basic sets. Note: this indicates the benefit of defining toplogies in terms of bases; we only have to work with these simple, basic sets.

We now can easily verify this property holds for our base B. Let $B_1, B_2 \in B$ be two basic open sets, say $B_1 = S(a_1, \lambda_1)$ and $B_2 = (a_2, \lambda_2)$, and let $x \in B_1 \cap B_2$. We seek an arithmetic progression containing x that is also contained in B_1 and B_2. Since x is in both of those arithmetic progressions, we just need to adjust the "step size" to incorporate only elements from those progressions. A moment's thought reveals that definining $\lambda = \mathrm{lcm}(\lambda_1, \lambda_2)$ will work:

$$x \in B_3 := S(x, \lambda) \subseteq S(x, \lambda_i) = B_i \quad \text{for each } i \in \{1, 2\}$$

Look back at the two examples given above to see how this works. Also, note that sometimes $B_1 \cap B_2 = B_3$ (as in the first, easier example presented) whereas sometimes $B_1 \cap B_2$ must be written as a union of basic open sets (as in the second example). (Also, note that the quantification is over all $x \in B_1 \cap B_2$, so this allows for vacuous quantification in the case where $B_1 \cap B_2 = \varnothing$.)

Now that we know τ is a topology, we can explore its properties! Specifically, we will start relating τ to the primes. Consider the following equality

$$\mathbb{Z} \setminus \{-1, +1\} = \bigcup_{p \text{ prime}} S(0, p)$$

This follows from the Fundamental Theorem of Arithmetic (every integer, except -1 and $+1$, is a multiple of some prime). We want to somehow argue that the union on the right-hand side *must* be an infinite union. Somewhat surprisingly, we will accomplish this via contradiction, using the notion of *open* and *closed* sets in this topology! Recall that a *closed* set is the complement of some open set.

Can the set on the left- or right-hand side of the above equality be open or closed? First, notice that $\{-1, +1\}$ is not open because it is finite (and every open set is infinite, since it contains an arithmetic progression), so $\mathbb{Z} \setminus \{-1, +1\}$ is *not closed*. FWIW it *is* open, though, since

$$\mathbb{Z} \setminus \{-1, +1\} = \bigcup_{a \in \{2,3,4,\dots\}} S(a, a)$$

Anyway, what about the set on the right? To answer this, we notice that any arithmetic progression is actually clopen. We saw some examples above, and we can generalize those ideas to the following:

$$S(a, \lambda) = \mathbb{Z} \setminus \bigcup_{j=1}^{\lambda-1} S(a + j, \lambda)$$

That is, an arithmetic progression consists of all integers *except* those that belong to any arithmetic progression whose starting point has been shifted from that of the original progression by an amount less than the step size. This shows that, in particular, $S(0,p)$ is closed for all primes p. By the basic properties of a topology, any finite union of closed sets is closed (since the complement is a finite intersection of open sets).

Thus, if there were only finitely many primes, then the union representation on the right-hand side would be closed while the complement representation on the left-hand side would be not closed. This is a contradiction and, therefore, there are infinitely many primes! $\square$

There is a fantastic (and fantastically concise) paper by Idris Mercer [2] that removes the topological language of this proof and presents only the *essence* of the construction. Specifically, there are two claims about intersections and unions of arithmetic progressions (namely that finite unions of finite intersections can also be written as finite intersections of finite unions, and that any intersections thereof are empty or infinite) and then the final equality considered in the proof above. In a way, this is "better" because it can be completed in half a page and doesn't require anything from the reader beyond a willingness to think about integers; however, in a way, this is "worse" because it doesn't present any further properties of the integers that this topology uncovers, nor does it actually make the connection to a seemingly disjoint branch of mathematics that makes Fürstenberg's proof so noteworthy.

See page 5 in [3] and [4] and [5] for more information.

2 Turán's Theorem: Graph Theory via Probability and Analysis

This is a fundamental result of graph theory, and one of the earliest examples of the burgeoning field of *extremal combinatorics*, which is closely related to *Ramsey Theory*. Heuristically speaking, Ramsey Theory answers questions of the form *"How big does a structure have to be to guarantee a certain substructure can be found?"*, whereas extremal combinatorics answers questions of the form *"How big can a structure be before a certain substructure pops up, and how can we construct a largest counterexample?"*. They don't seem so different now, actually ...

Anyway, Turán's Theorem concerns how large a simple graph can be without creating a large *clique*. Recall that a simple graph is just a finite set of vertices and edges between them, and a clique is a subset of the vertices such that any pair of vertices is joined by an edge. (That is, a clique is a subgraph that is, itself, a complete graph.) Let $p \in \mathbb{N}$ and try to construct a graph G on n vertices that does not contain a p-clique (i.e. the largest clique is of size $\omega(G) \leq p-1$).

We can maximize the number of edges by partitioning the vertices into $p-1$ pairwise disjoint subsets $V_1, \dots, V_{p-1}$, each of size $\frac{n}{p-1}$, and adding an edge between $x \in V_i$ and $x \in V_j$ if and only if $i \neq j$ (i.e. x and y belong to different vertex subsets). We call this the *Turán graph on p and n*. With this construction, the number of edges is

$$|E| = \binom{p-1}{2}\left(\frac{n}{p-1}\right)^2 = \left(1 - \frac{1}{p-1}\right)\frac{n^2}{2}$$

since, for every pair of vertex subsets, all possible edges are added between the $\frac{n}{p-1}$ vertices in each subset. (There are constructions of a very similar form when $p-1 \nmid n$ which are also called *Turán graphs*, but we will ignore those for sake of simplicity. It only affects the sharpness of the bound presented in the result below.) The theorem in question tells us that this construction is optimal, that it "packs the most edges possible" into such a graph.

Theorem 2. *Let $G = (V, E)$ be a simple, finite graph on n vertices. Let $p \geq 2$ and assume G has no p-clique. Then*

$$|E| \leq \left(1 - \frac{1}{p-1}\right)\frac{n^2}{2}$$

The case $p = 2$ is trivial, and proofs of the case $p = 3$ (wherein triangle-free graphs have at most $n^2/4$ edges) were known prior to Turán's proof (and there are some amazingly elegant proofs, now known, that use the Cauchy and Arithmetic/Geometric Mean inequalities!). What we present here is a proof of the general theorem using probability theory and the Cauchy-Schwarz inequality.

Proof. Label the vertices $V = \{v_1, \dots, v_n\}$ and let d_i be the degree of v_i (that is, the number of edges incident to v_i). Let $\omega(G)$ denote the size of the largest clique. First, we prove the following claim:

$$\omega(G) \geq \sum_{i=1}^{n} \frac{1}{n - d_i}$$

(This is the part of the proof that is probabilistic.) Choose a random permutation of the vertices, denoted by $\pi = v_1 v_2 \cdots v_n$, chosen uniformly at random, so each permutation is chosen with probability $\frac{1}{n!}$. We now construct a clique of G from this permutation, and we will call it C_π.

For the i-th vertex v_i of the permutation π, include $v_i \in C_\pi$ if and only if v_i is adjacent to all the vertices v_j with $j < i$ that precede it in π. By definition, $v_1 \in C_\pi$, and C_π is a clique. (Think about ordering the elements of C_π according to the permutation π.) Define $X = |C_\pi|$ to be a random variable on the space of permutations Π. To deduce the claimed bound, we find $E[X]$.

Specifically, we write $X = \sum_{i=1}^{n} X_i$ as a sum of indicator random variables, where $X_i = 1$ or $X_i = 0$, depending on whether $v_i \in C_\pi$ or $v_i \notin C_\pi$. Now, let's think about the permutations π that yield $v_i \in C_\pi$. To have this, v_i must appear *before* all of its non-neighbors in the permutation. The vertex v_i has $n - 1 - d_i$ non-neighbors, and thus there are $n - d_i$ vertices (including v_i itself) that we want to consider. With equal probability, any one of them will be the first to appear in the ordering of the permutation, so the probability that v_i is the first one is $\frac{1}{n-d_i}$. Thus, $E[X_i] = \frac{1}{n-d_i}$. Then, by linearity of expectation,

$$E[X] = \sum_{i=1}^{n} E[X_i] = \sum_{i=1}^{n} \frac{1}{n - d_i}$$

Since this essentially tells us the "average size" of a clique, there must be a clique with *at least* this size. This proves the lower bound on $\omega(G)$ that we claimed above.

Now, we move into the analytics part of the proof, which is actually quite short. Recall the Cauchy-Schwarz inequality, which states that for any two sets of real numbers, $\{a_1, \ldots, a_n\}$ and $\{b_1, \ldots, b_n\}$,

$$\left(\sum_{i=1}^{n} a_i b_i\right)^2 \leq \left(\sum_{i=1}^{n} a_i^2\right)\left(\sum_{i=1}^{n} b_i^2\right)$$

Using $a_i = \frac{1}{\sqrt{n-d_i}}$ and $b_i = \frac{1}{a_i}$, we have $a_i b_i = 1$ and thus

$$n^2 \leq \omega(G) \sum_{i=1}^{n} (n - d_i)$$

after applying the previous claim. We assumed that $\omega(G) \leq p - 1$, and we can simplify the sum on the far right, yielding

$$\begin{aligned} n^2 \leq \omega(G) \sum_{i=1}^{n} (n - d_i) &= \omega(G) \left(\sum_{i=1}^{n} n - \sum_{i=1}^{n} d_i\right) \\ &= \omega(G)\left(n^2 - 2|E|\right) \leq (p-1)(n^2 - 2|E|) \end{aligned}$$

Simplifying this inequality yields the original result! □

This technique of doing "something" random to a graph–ordering the vertices or edges, coloring the vertices or edges, or what have you–is actually fairly common in Ramsey Theory and extremal combinatorics and falls under the umbrella of the Probabilistic Method, pioneered by Erdös. It is a striking example of how the techniques and results of one branch of mathematics apply immediately to another, with some ingenuity, of course. Unfortunately, one of the main drawbacks is that this method is highly *non-constructive*, in that it can guarantee the *existence* of some object or property without giving any indication as to how to find it, in practical terms. Still, this method and its concepts is mathematically beautiful, and this particular proof is surprising and elegant enough to warrant exposition.

The standard proof of Turán's Theorem assumes a graph G with the maximum number of edges and shows that there cannot exist vertices u, v, w such that edge uv is present but neither uw nor vw is present. This provides an equivalence relation on the vertices of G where $u \sim v$ if and only if they are non-neighbors. The next claim argues that this number of edges is maximized when the sizes of the equivalence classes differ by at most one. See pages 207-210 in [3] and [6] for more information.

3 Cayley's Formula: Combinatorics via Linear Algebra

Cayley's Formula–which says that the number of trees on n labeled vertices is n^{n-2}–should probably be known as Borchardt's Formula, since it was the Prussian/German mathematician Carl Wilhelm Borchardt who first proved the formula in 1860. Cayley cited Borchardt's paper in his 1889 note that generalized some of the notions contained therein, but perhaps it was his introduction of some of the corresponding graph theoretic terminology that made the name "Cayley's Formula" stick. Interestingly enough, though, the proof we present below is probably vastly more similar to Borchardt's original proof than Cayley's generalization. (Not having been able to track down Borchardt's paper, I can't confirm this claim; I am merely relying on the fact that Wikipedia asserts that Borchardt's proof was "via a determinant" [9].) Cayley's proof relies on the aptly-named Prüfer Sequences that uniquely identify trees; establishing this bijection between trees and sequences allows one to combinatorially count the sequences and prove the formula. The proof we present here is rooted in linear algebra; specifically, we apply Kirchhoff's Matrix Tree Theorem, a more general claim about the number of spanning trees in a connected graph.

First, we must introduce some terminology and preliminary results. Our ultimate goal here is to count the number of trees on n labeled vertices. Recall that a tree on n vertices is a simple, connected graph with no cycles; equivalently, it has $n-1$ edges, its vertices are connected by unique paths, it is maximally acyclic, or it is minimally connected. Given any simple, connected graph $G = (V, E)$ a *spanning tree* is a subgraph of G that is, itself, a tree and includes all the vertices of G. (This is why we specify G is connected, otherwise this is not possible.) Kirchhoff's theorem states a fact about the number of spanning trees on a general graph. Later on, we will apply the theorem to a specific case–namely, the complete graph K_n on n vertices–to derive Cayley's Formula.

To state Kirchhoff's theorem, we need to define some matrices. To any simple, connected graph $G = (V, E)$ where $V = [n]$ (i.e. the n vertices are labeled), we associate three $n \times n$ matrices: the *degree matrix* D, the *adjacency matrix* A, and the *Laplacian matrix* L. The first two matrices are defined by

$$D_{ij} = \begin{cases} \deg(i) & \text{if } i = j \\ 0 & \text{otherwise} \end{cases}$$

$$A_{ij} = \begin{cases} 1 & \text{if } ij \in E \\ 0 & \text{otherwise} \end{cases}$$

and then we define $L = D - A$.

One important observation is that all of the rows of L sum to 0; this is because the diagonal entry L_{ii} is the degree of vertex i, and the off-diagonal entries L_{ij} (with $i \neq j$) on that row are either 0 or -1, depending on whether vertices i and j are adjacent. Since this holds for *every* row, it follows that $\lambda = 0$ is an eigenvalue of L, corresponding to the eigenvector $\underline{v} = [1, \ldots, 1]^T$. There will be $n-1$ other eigenvalues, as well. This fact relies on some properties of the Laplacian matrix that we will not fully discuss here. (If you are looking for a little convincing, in the paragraph below, we point out that we can express $L = CC^T$, where C is an $n \times m$ (where $E = [m]$) edge incidence matrix arising from an arbitrary assignment of directions to all edges. This shows that L is positive semi-definite, and therefore has n real eigenvalues.)

Furthermore, it is always possible to factor the Laplacian matrix as $L = CC^T$, and this is the form of the matrix to which we will apply the Binet-Cauchy formula. Specifically, we first define the *edge incidence matrix* B, where the rows are indexed by the vertex set, $[n]$, and the columns are indexed by the vertex set $[m]$ (where $|E| = m$ and $m \geq n-1$ since G is connected), and

$$B_{ij} = \begin{cases} 1 & \text{if vertex } i \text{ is incident to edge } j \\ 0 & \text{otherwise} \end{cases}$$

Next, we assign an *arbitrary* direction to all the edges and construct a new incidence matrix C that switches some of the 1s to -1s based on whether that directed edge is going “into” or “out of” that vertex. It is important to note that the direction really can be arbitrary, and we only do this to make all of the columns (corresponding to edges) have a sum of 0. This allows us to take advantage of some useful properties of matrices, linearly dependent vectors, and determinants. You’ll see what we mean as we go along! ☺

We claim $L = CC^T$. Notice that L is an $n \times n$ symmetric matrix (because $L^T = (CC^T)^T = CC^T$), the diagonal entries M_{ii} are the degrees of vertex i (because CC^T takes every row of C and “dot products” it with itself, producing a $+1$ any time a -1 or $+1$ appears in the row), and the off-diagonal entries L_{ij} (with $i \neq j$) are either -1 (if $ij \in E$, because then row i and row j will have a -1 and $+1$ appear in one column and in all other columns, one of the rows will have a 0 entry) or 0 (if $ij \notin E$, because then every column will show a 0 entry in one of the two rows). This corresponds with our previous definition $L = D - A$.

Now, we can state Kirchhoff’s Theorem:

Theorem 3. *Let G be a simple, connected graph with $V = [n]$, and let λ_i, for $1 \leq i \leq n-1$, be the nonzero eigenvalues of L. Then the number of spanning trees $t(G)$ on G is given by*

$$t(G) = \frac{1}{n} \prod_{i=1}^{n-1} \lambda_i$$

Equivalently $t(G) = C_{ij}$, where C_{ij} is any *cofactor of the matrix L. (Recall that the cofactor is the signed minor, $(-1)^{i+j} \det(L_{\not{i}\not{j}})$, where $L_{\not{i}\not{j}}$ is the matrix produced by removing row i and column j from L.)*

It is actually the second part of the theorem that we will use to obtain Cayley’s Formula. For now, we will prove a modified subcase of this theorem, ignoring the statement about eigenvalues and only considering signed minors along the diagonal, i.e. where $i = j$ so $(-1)^{i+j} = 1$. (This will make the complicated argument much easier to follow.) The first step of the proof, though, requires a powerful linear algebraic result known as the Binet-Cauchy Formula [10], so we state that first.

Lemma 1. *Let A be an $r \times s$ matrix and B be an $s \times r$ matrix. Let $\binom{[s]}{r}$ represent the set of subsets of $[s]$ of size r. For any $S \in \binom{[s]}{r}$, let $A_{[r],S}$ represent the $r \times r$ submatrix of A whose columns are the columns of A whose indices belong to S; similarly, let $B_{S,[r]}$ be the $r \times r$ submatrix of B whose rows are the rows of B whose indices belong to S. Then,*

$$\det(AB) = \sum_{S \in \binom{[s]}{r}} \det\left(A_{[r],S}\right) \det\left(B_{S,[r]}\right)$$

Later on, we will be considering a special situation where $B = A^T$. In that case, notice that

$$B_{S,[r]} = A^T_{[r],S} \Rightarrow \det\left(B_{S,[r]}\right) = \det\left(A^T_{[r],S}\right) = \det\left(A_{[r],S}\right)$$

and so the Binet-Cauchy formula simplifies to a sum of squares of determinants.

A proof of the Binet-Cauchy Formula falls outside the scope of this talk, so we will skip it. There are several known proofs, though, the most "direct" being a formal manipulation of definitions of determinants and a careful following of notation. Interestingly enough, though, there is a proof (see pages 170-172 in [3]) of the result via graph theoretic considerations, where the matrices represent weighted bipartite graphs and the formula represents the weights of disjoint path systems.

Consider the following example and keep it in mind as we prove Kirchhoff's Theorem. We will use it to construct all of the relevant matrices and make some arguments about the situation, in general.

Example 1. Let G be the kite graph on [4].

******** Find B and C and L. Consider removing row 1, so $i = 1$ in the proof.

In first case (NOT a tree) M by columns 2,3,5.

In second case (IS a tree) M by columns 1,3,5

Finally, we are ready to prove Kirchhoff's Theorem!

Proof (of Theorem 3). Let $G = (V, E)$ be a simple, connected graph with $V = [n]$ and $E = [m]$. Construct the incidence matrix B, then assign the edges E an arbitrary direction to construct the $n \times m$ directed incidence matrix C, and define $L = CC^T$. Let $i \in [n]$ and remove row i and column i from L to obtain $L_{\not{i}\not{i}}$. Notice that

$$L_{\not{i}\not{i}} = C_{\not{i}} C_{\not{i}}^T$$

where $C_{\not i}$ is the matrix C with row i removed. We want to find $\det(L_{\not i \not i})$, and we will do this by applying the Binet-Cauchy Formula,

$$\det L_{\not i \not i} = \sum_{M} (\det M)^2$$

where the summation is over all matrices M that are $(n-1) \times (n-1)$ submatrices of C, identified by choosing $n-1$ columns (i.e. edges) from the m total columns. (Note: this is always possible since $m \geq n-1$ because G is connected, and we removed row i from C, so there are m columns and $n-1$ rows.)

We now consider what happens for different choices of M in this summation. First, we claim that M always corresponds to a subgraph of G with $n-1$ edges and n vertices. (Note: the subgraph may or may not be connected! That consideration comes next!) The $n-1$ columns identify edges, so certainly there are $n-1$ of them, and this encodes information about vertex i still (this is represented by columns whose sum is nonzero, implying that there is a corresponding ± 1 in row i). Now, the important part: we claim that

$$\det M = \begin{cases} \pm 1 & \text{if the edges identified by } M \text{ span a tree in } G \\ 0 & \text{otherwise} \end{cases}$$

Suppose the $n-1$ edges identified by M do *not* span a tree. This means the subgraph of G that M represents contains at least two connected components, and thus one such component does *not* contain vertex i. Consider the rows of M corresponding to the vertices of that component. All edges incident to those vertices in the corresponding subgraph are contained within this component, so every column has both a -1 and $+1$. Thus, if we sum these rows, we obtain a row vector of 0s. This means those row vectors are *linearly dependent* and thus $\det M = 0$ (because we could perform some row operations to yield a row of 0s in the matrix).

Now, suppose the $n-1$ edges identified by M *do* span a tree. This means the subgraph of G with the edges identified by the columns of M is a tree, and can therefore be "deconstructed" by removing, one at a time (breaking ties arbitrarily), a leaf (i.e. a vertex of degree 1). Furthermore, there is always at least *two* leaves in any tree, so we can always identify a leaf that does *not* correspond to vertex i (whose corresponding row was removed from C). We now describe how this "deconstruction" process corresponds to finding $\det M$. A leaf corresponds to a row of M with only a -1 or $+1$ in it (i.e. only one edge, either coming in or going out) and the other entries 0, so we use that row in the standard definition of finding a determinant. Crossing out that row and the column corresponding to that ± 1 (i.e. that *edge* in the tree subgraph), we now need to find the determinant of an $(n-2) \times (n-2)$ submatrix of M that *also* corresponds to a tree! Thus, the same argument applies: we can find a row with only a single ± 1 and other entries 0, and use that row in the determinant calculation. By standard properties of trees, we can continue this (breaking ties arbitrarily if there is more than one leaf at any step) and eventually completely "deconstruct" the tree and reduce the situation to a 1×1 matrix that is ± 1. At every step in this process, we are multiplying by ± 1, and therefore $\det M = \pm 1$.

This completes the proof, because the summation stated above counts any spanning tree as $(\pm 1)^2 = 1$ and any other subgraph as 0. Therefore, we have shown that $t(G) = \det(L_{\not{i}\not{i}})$ for any $i \in [n]$. □

Where were we? Oh right, Cayley's Formula! Let's apply Kirchhoff's Theorem to the particular case where $G = K_n$, the complete graph on n vertices. The Laplacian of K_n is an $n \times n$ matrix whose diagonal entries are all $n - 1$ (the degree of all vertices) and whose off-diagonal entries are all $n - 1$ (every vertex is adjacent to every other vertex). Let's use $i = 1$ and remove that row and column, leaving $L_{\not{1}\not{1}}$ which is an $(n-1) \times (n-1)$ matrix with the same structure as L. We need to find $\det(L_{\not{1}\not{1}})$. Our reference [3] claims this is "an easy computation" and we will attempt to describe the process here. It is a series of row operations on the matrix to convert it to an upper triangular matrix without altering the determinant.

1. In $L_{\not{1}\not{1}}$, subtract the last row from row j, for every $2 \leq j \leq n-2$, and place that in row j. This yields ns on those diagonals, a column of $-n$s in those far-right columns, and 0s elsewhere.

2. Add $(n-1)$ times the last row to the first row and place that in the first row. This yields 0 in the top-left entry, $n(n-2)$ in the top right entry, and 0s elsewhere in the top row.

3. Successively add row j, for every $2 \leq j \leq n-2$, to the first row and place that in the first row. This yields all 0s in the top row except for the top-right entry, which is n.

4. Swap rows 1 and $(n-1)$, while also scaling the new top row by -1. This double-step does not alter the determinant $(-1 \times -1 = 1)$.

5. We now have an upper triangular matrix with 1 in the top-left entry, and n as every other diagonal entry. Thus,

$$\det L_{\not{1}\not{1}} = 1 \cdot \underbrace{n \cdot n \cdots n}_{n-2 \text{ times}} = n^{n-2}$$

See pages 173-178 in [3] and [8] and [9] for more information.

4 Bibliography

References

[1] Fürstenberg, Harry (1955). "On the infinitude of primes". *American Mathematical Monthly* **62** (5): 353.

[2] Mercer, Idris D. (2009). "On Furstenberg's Proof of the Infinitude of Primes". *American Mathematical Monthly* **116**: 355-356.

[3] Aigner, Martin and Ziegler, Günter M. *Proofs from the Book*. Springer: Berlin (3rd ed, 2000).

[4] Wikipedia: Fürstenberg's proof of the infinitude of primes
`http://en.wikipedia.org/wiki/F%C3%BCrstenberg%27s_proof_of_the_infinitude_of_primes`

[5] Wikipedia: Evenly spaced integer topology
`http://en.wikipedia.org/wiki/Evenly_spaced_integer_topology`

[6] Wikipedia: Turán's theorem `http://en.wikipedia.org/wiki/Tur%C3%A1n%27s_theorem`

[7] Alsina, Claudi and Nelsen, Roger B. *Charming Proofs: A Journey Into Elegant Mathematics*. Mathematical Association of America: USA (2010).

[8] Wikipedia: Kirchhoff's theorem
`http://en.wikipedia.org/wiki/Kirchhoff%27s_matrix_tree_theorem`

[9] Wikipedia: Cayley's formula
`http://en.wikipedia.org/wiki/Cayley%27s_formula`

[10] Wikipedia: Cauchy-Binet formula
`http://en.wikipedia.org/wiki/Binet-Cauchy_formula`