1 The crossing number lemma

1.1 Basic notation and terminology for graphs

- graph, vertex, edge, adjacent, incident, neighborhood, degree
- (induced) subgraph, spanning, complete, independent
- walk, path, cycle, isomorphic, connected, tree, leaf

A **graph** $G(V, E)$ (sometimes we just write G if the context is clear) is defined by a pair of sets: $V(G)$, the set of **vertices**, and $E(G)$, a set of two-element subsets of $V(G)$ called **edges**. Again, when context is clear, we just write V and E. Given an edge containing two vertices, those vertices are said to be **adjacent**, while the edge is said to be **incident** to the vertices. Given a vertex v in a graph, the set of vertices adjacent to v is called the **neighborhood** of v. The number of edges that contain a vertex v is called the **degree** of v, sometimes denoted $\deg_G(v)$

Suppose we have a graph $G(V, E)$, and we consider a graph $H(W, F)$, where $W \subset V$ and $F \subset E$. Then H is called a **subgraph** of G. If additionally, every edge from E whose vertices are contained in W is present in H, we call H an **induced** subgraph of G. If every vertex of V is in some edge of F, then we say that H **spans** G. If every pair of vertices is present in the edge set, the graph is called **complete**. The complete graph on n vertices is often denoted K_n. A set of vertices is called **independent** if none of them adjacent to one another.

A sequence of vertices $(v_1, v_2, \ldots, v_n)$ in a graph $G(V, E)$ is called a **walk** when v_i is adjacent to v_{i+1} for all $i \in [1..(n-1)]$. If the vertices are distinct, a walk is also called a **path**. A path on n vertices is sometimes denoted P_n. If $\{x_n, x_1\} \in E$ as well, the path is called a **cycle**. A cycle on n vertices is sometimes denoted C_n.

If $G(V, E)$ and $H(W, F)$ are graphs, and there exists a bijection $f : V \to W$ such that two vertices in V are adjacent in G if and only if their images under f are adjacent in H. We sometimes write $G \cong H$. A graph G is called **connected** if there is a path in G connecting every possible pair of vertices. Notice that this is distinct from completeness! A connected graph with no cycles is called a **tree**. A vertex of degree one in a tree is called a **leaf**. A tree that spans a graph is called a **spanning tree**. A graph on multiple trees is called a **forest**.

1.2 Planar graphs

- planar, crossing, face, Euler's formula, crossing number

In a drawing of a graph, an instance of two edges crossing each other is called a **crossing**. A graph is **planar** if it can be drawn without any two edges crossing. A **face** is a region (bounded or unbounded) delineated by the edges in a drawing of a planar graph.

Proposition 1 (Euler's formula)**.** *Suppose G is a simple, connected, planar graph with v vertices, $e \geq 1$ edges, and f faces, then $v + f = e + 2$.*

Proof. Proceed by induction on the number of edges. Start with a graph consisting of only one edge. Notice that it has two vertices and one face, satisfying the formula. For our induction hypothesis, suppose that the formula holds for all simple, connected, planar graphs with n edges. Now, to show that the formula will hold for any simple, connected, planar graph with $n+1$ edges, we watch what happens whenever we add an edge. Notice that any time you add an edge, you either add a vertex or a face. Therefore, we are done, by induction. □

Proposition 2. *Suppose G is a simple, connected, planar graph with $e \geq 2$ edges, and f faces, then $3f \leq 2e$.*

Proof. Draw the graph. For each edge in the graph, draw a little dog ear on each side, so that each edge looks like a wiener dog. Observe that there are two dog ears on each edge, so the number of dog ears is exactly $2e$. Notice that each face needs **at least** three dog ears, so the number of dog ears is $\geq 3f$. □

Combining Propositions 1 and 2 yields the following corollary.

Corollary 3. *Suppose G is a planar graph with $e \geq 2$ edges, and v vertices, then*

$$e - 3v + 6 \leq 0. \tag{1}$$

Note that the complete graph on four vertices is planar, as it can be drawn without any edges crossing, but the complete graph on five vertices is NOT planar, as it has five vertices and ten edges, and therefore fails to obey the inequality in Corollary 3.

Proposition 4. *Suppose G is a graph (possibly nonplanar or disconnected) with $e \geq 2$ edges, and f faces, then the minimum number of crossings under any redrawing of G is*

$$cr(G) \geq e - 3v + 6. \tag{2}$$

Proof. Take our graph G, and if it satisfies (1), then we know it is planar, and we trivially satisfy (2). If not, then we consider a drawing of G that exhibits the minimal number of crossings, and select one of the crossing edges, e_1. We now consider a new graph, G_1, which is just G with the edge e_1 removed. Now, e_1 contributed at least one crossing to $cr(G)$, so by removing it, there should be at least one crossing fewer in G_1. Of course, it is possible that e_1 contributed many crossings, so we just have the inequality $cr(G_1) \leq cr(G) - 1$. If G_1 satisfies (1), then we stop. Otherwise, we repeat the above procedure, removing edges until we get to some j for which G_j is planar. We then have

$$0 = cr(G_j) \leq cr(G) - j,$$

which yields the desired result. □

1.3 A pinch of probability

- probability, expected value

We'll prove a nice graph theory result by applying some simple probability. First, we define a **probability** as a number $p \in [0, 1]$. Next, we define the **expected value**, $\mathbb{E}$, of a variable to be its probabilistically weighted sum. That is,

$$\mathbb{E}(x) = \sum_j p_j x_j,$$

where x takes the value x_j with probability p_j, and $\sum_j p_j = 1$.

Example 1. *Suppose I flip a fair coin* 100 *times. What is the expected number of heads?*
Solution: *In the event I flip heads, I add one to my count, and in the event I flip tails, I add zero to my count.*

$$\mathbb{E}(\textit{heads in 100 flips}) = \sum_{\textit{all flips}} (p_{heads} \cdot 1 + p_{tails} \cdot 0)$$

$$= 100 \cdot \frac{1}{2} \cdot 1 + 100 \cdot \frac{1}{2} \cdot 0 = 50.$$

We are greatly glossing over this, but here we will say that expectation is *linear*, meaning that $\mathbb{E}(X) + \mathbb{E}(Y) = \mathbb{E}(X + Y)$. The following example is a very superficial illustration of this fairly deep fact.

Example 2. *Suppose I flip a fair coin 60 times, and record the number of heads. Suppose I flip the same coin 40 more times and record the number of heads. The expected number of heads from the first sequence of flips was 30, and the expected number from the second sequence was 20. It should come as no surprise that* $30 + 20 = 50$, *which is the expected number of flips we got from Example 1.*

1.4 The crossing number lemma

The next proof is from Székely's *Crossing numbers and hard Erdős problems in discrete geometry.* It was initially shown by other people. See the references in Székely's paper.

Lemma 5 (Crossing Number Lemma)**.** *Suppose G is a graph with e edges, v vertices, and $e \geq 4v$, then*

$$cr(G) \gtrsim \frac{e^3}{v^2}.$$

Proof. Let H be a random induced subgraph of G, where each vertex is chosen with probability $p = \frac{4v}{e}$. Now, write (2) for H, and take expected values of both sides.

$$\mathbb{E}(cr(H)) \leq \mathbb{E}(e_H) - 3\mathbb{E}(v_H) + 6 \tag{3}$$

Verify that $\mathbb{E}(v_H) = pv$, $\mathbb{E}(e_H) = p^2 e$, and $\mathbb{E}(cr(H)) \leq p^4(cr(G))$, and (3) becomes

$$p^4 cr(G) \gtrsim p^2 e - 3pv + 6,$$

which yields the claimed estimate. □

1.5 The return of incidence theory

- Szemerédi-Trotter

Again, this proof is due to Székely, but the result was initially shown by Szemerédi and Trotter.

Theorem 6 (Szemerédi-Trotter, *Extremal problems in discrete geometry*)**.** *Given a collection of n points and m lines in the plane, the number of incidences of points and lines is*

$$I \lesssim (nm)^{\frac{2}{3}} + n + m.$$

Proof. Construct a graph, G, by letting the vertices be the n points, and the edges be the $I - m$ segments connecting the consecutive points on a given line. So $v = n$, and $e = I - m$. Notice that the two distinct lines can cross each other only once, so number of crossings in G must be less than $\binom{m}{2} \approx m^2$. If $e \leq 4v$, then $I \leq 4n + m$. If $e > 4v$, then we can apply Lemma 5, and we get that $I \lesssim (nm)^{\frac{2}{3}} + m$, and we are done. □

1.6 Crossings and repeated distances

- unit distance problem

When Erdős posed his distinct distances problem, he also raised the **unit distance problem**, which is to determine how often a single distance could occur in any large finite set of points in the plane. It is so called because we can scale the set to make whatever the most popular distance is 1, and vice versa. Here is the best known result. It is due to Spencer, Szemereédi, and Trotter, but we, again, give a proof by Székely.

Theorem 7. *In any large finite set of n points in the plane, no distance can occur more often than $n^{\frac{4}{3}}$ times.*

Proof. Draw unit circles centered at each point in our set. Construct a multigraph, G, by letting the vertices be the n points, so $v = n$. Now let the edges be the arcs of circles between consecutive points on any given circle. Notice that each edge corresponds to an incidence of a point and a unit circle. For each pair of points determining a unit distance, there will be exactly two such incidences (each point will lie on the other point's circle). So $e = I$, the number of incidences, which is exactly two times the number of unit distances.

If there are fewer than $10v$ edges, then we have no more than $10v = 10n$ incidences, so we have no more than $\lesssim n$ unit distances. So we proceed by assuming $e > 10v$. We will now need to prune our multigraph so that it becomes a graph, and we can apply Lemma 5. We do this by removing any loops, or duplicate edges between pairs of points. To do this without destroying our structure, we remove any circle with fewer than three points. Notice that we lose at most $2n$ edges in doing this (no more than two points per circle, and no more than n circles), meaning that we still have $> 8n$ edges remaining. Now, it is possible that a pair of points could be on two distinct circles (but no more than two), each that have several other points. In this case, we remove one of the two edges. In this way, we lose no more than half of our edges, meaning that we still have $e > 4v$ after pruning.

Notice that the two distinct circles can cross each other only once, so number of crossings in G must be less than $\binom{n}{2} \approx n^2$. Because we have that $e > 4v$, we can apply Lemma 5, we get

$$\frac{I^3}{n^2} = \frac{e^3}{v^2} \lesssim cr(G) \lesssim n^2$$

which gives us that $I \lesssim n^{\frac{4}{3}}$, as claimed. □

We now apply the pigeonhole principle backwards along with the previous result to get a better bound on the distinct distances problem.

Theorem 8 ($n^{\frac{2}{3}}$ estimate)**.** *Any large finite set of n points in the plane determines at least $n^{\frac{2}{3}}$ distinct distances.*

Proof. There are $\binom{n}{2} \approx n^2$ point pairs, each of which determines some distance. By Theorem 7, we know that no distance can occur more than $n^{\frac{4}{3}}$ times. So by the pigeonhole principle, there must be at least $n^{\frac{2}{3}}$ distinct distances. □

1.7 Some additive number theory

- sum set, product set, sums and products problem

Given sets $A, B \subset \mathbb{R}$, define the **sum set**

$$A + B := \{a + b : a \in A, b \in B\},$$

and define the **product set**

$$AB := \{ab : a \in A, b \in B\}.$$

The **sums and products problem** is to determine for any large finite $A \subseteq \mathbb{R}$ what is $\max\{|A + A|, |AA|\}$ in terms of $|A|$? Erdős has conjectured that it should be at least $|A|^{2-\epsilon}$, for any $\epsilon > 0$. Here is a nontrivial estimate.

Theorem 9 (Elekes *On the number of sums and products*)**.** *Given a set, A, of n real numbers,*

$$\max\{|A \cdot A|, |A + A|\} \gtrsim n^{\frac{5}{4}}.$$

Proof. We will construct a set of points and lines. The points will be

$$\mathcal{P} = \{(a_i + a_j, a_k \cdot a_l) : a_i, a_j, a_k, a_l \in A\},$$

and the lines will be

$$\mathcal{L} = \{y = a_s \cdot (x - a_t) : a_s, a_t \in A\}.$$

There are $|A + A| \cdot |A \cdot A|$ points, and n^2 lines. Notice that each line is coincident to at least n points. Apply Theorem 6, and see that

$$n^3 = n \cdot |\mathcal{L}| \lesssim I \lesssim \left(|A + A| \cdot |A \cdot A| \cdot n^2\right)^{\frac{2}{3}} + |A + A| \cdot |A \cdot A| + n^2.$$

So $|A + A| \cdot |A \cdot A| \gtrsim n^{\frac{5}{2}}$, and the result follows. □