Geometric Graphs: The Crossing Lemma and Two Applications

These notes borrow from Chap 4 of the wonderful book by J. Matoušek on Discrete Geometry (http://www.ms.mff.cuni.cz/acad/kam/matousek).

Let $G = (V, E)$ be a graph with n labeled vertices and m edges. A drawing, or planar embedding of G assigns each vertex $v_i \in V$ to a distinct point $P_i \in R^2$. An edge $e = \overline{v_i v_j} \in E$ is drawn as a continuous, simple (non-self-intersecting) curve in the plane joining P_i and P_j. The edge is the open subsetset of the curve excluding P_i and P_j; it is not allowed to contain any other point P_k. The crossing number *of the drawing* is the number of pairs of edges that meet (however many times). The crossing number *of the graph* - $CR(G)$ - is the minimum possible crossing number in any drawing of G. If G is planar $CR(G) = 0$.

The following result (called the crossing lemma) was originally proved in '82 by Ajtai, Chvatal, Newborn, and Szemerédi [2], and independently by T. Leighton [6] in '84. It is an extremely useful result (several applications will be mentioned) and the following transparently simple proof is a shining example of the possible power and beauty of the probabilistic method.

Theorem 1 *Let $G = (V, E)$ be a graph with n vertices and m edges. If $m \geq 9n/2$*

$$CR(G) \geq c\frac{m^3}{n^2};$$

we may take $c = 4/243$, as shown in the proof (though it is not the best constant).

Proof: Take a graph $G = (V, E)$ with n vertices, m edges and $CR(G) = x$. There must be some drawing of G with x crossing pairs of edges, and we will take such a best-possible drawing.

Euler's relation (see for example [7]) implies that if $m \geq 3n - 5$ then $CR(G) > 0$. This implies the trivial bound:

$$CR(G) \geq m - 3n; \tag{1}$$

if G is a planar graph on n vertices that has a maximal number of edges ($m < 3n - 5$), every edge you add creates at least one new crossing pair of edges. This simple fact will be enough for the proof.

We construct a random subgraph $G' = (V', E') \subseteq G = (V, E)$ by randomly pruning vertices and edges as follows: Initially set $V' = V$ and $E' = E$. Do n Bernoulli trials with success probability p. If the i^{th} trial is failure then prune v_i from V'; otherwise v_i remains in V'. Delete any edge $e = \overline{uv} \in E'$ if either u or v (or both) was pruned. Finally, keep the old drawing of the vertices and edges of G that survived pruning (the drawing of G' is the drawing of G with pruned edges and vertices erased). Let n', m', and x' be random variables counting respectively the number of vertices, edges, crossing pairs, in the inherited drawing of G'. By (1), $x' \geq m' - 3n'$, since $x' \geq CR(G')$. Therefore

$$E(x') \geq E(m') - 3E(n').$$

Clearly $E(n') = np$, $E(m') = mp^2$, and $E(x') = xp^4$, and we have

$$x \geq \frac{m}{p^2} - 3\frac{n}{p^3}. \tag{2}$$

The right hand side is a function of p that has its maximum when $p = p^* = 9n/(2m)$. The constraint $p^* \leq 1$ forces $m \geq 9n/2$. If we evaluate the right hand side of (2) at p^* we get the asserted bound on $x = CR(G)$●

Remarks:

1. This proof uses a weak, elementary fact, (1), magically combined with probability theory, which somehow magnifies its impact. The original proofs had more meat.

2. The lower bound is nearly optimal: there are graphs G with n vertices and $m \geq 9n/2$ edges that have drawings with about $.0595m^3/n^2$ crossings. Therefore $CR(G) \leq .0595m^3/n^2$ for these graphs.

3. The constant $c = 4/243$ (about .016..) was recently improved to about .0296.., using something a little stronger than Euler's relation. Therefore the upper and lower bounds for $CR(G)$ only differ by a factor of about 2.01 (see [9]).

4. If you subtract an from the crossing number lower bound in Theorem 1, it will slightly weaken the bound. However this will now hold for all $m > 3n$. In particular setting $a = 1.5$ makes the bound zero when $m = 4.5n$ and negative for smaller m.

5. Its not clear where this proof originated. It appears in "Proofs from the Book" [1].

In addition to its own inherent interest, Theorem 1 has many applications. Two will be mentioned, the first of which is a simple proof of a famous theorem that has many applications of its own. The second is another interesting and important result in combinatorial geometry that fundamentally depends on Theorem 1. Both are presented as evidence that Theorem 1 - which we proved in such a simple and pretty way using the probabilistic method - is very useful.

Application 1 - point/line incidences: Let $\mathcal{L} = \{\ell_1, \ldots, \ell_m\}$ be a set of m given lines in the plane and $\mathcal{P} = \{P_1, \ldots, P_n\}$, a set of n given points. We want to count

$$I(\mathcal{P}, \mathcal{L}) = |\{(i, j) : P_i \in \ell_j\}|,$$

the number of point/line incidences for the given collections ($P = (u, v) \in \ell = \{(x, y) : ax + by = c\}$ if and only if $au + bv = c$). If you take $\mathcal{L}$ to be three lines with different slopes and $\mathcal{P}$ to be the three intersection points you see $I(\mathcal{P}, \mathcal{L}) = 6$ (each point of $\mathcal{P}$ is in 2 lines of $\mathcal{L}$). Write

$$I(n, m) = \max\left(I(\mathcal{P}, \mathcal{L}) : |\mathcal{P}| = n, |\mathcal{L}| = m\right). \tag{3}$$

The example shows $I(3, 3) \geq 6$ (in fact it is 6).

$I(n, m)$ expresses a fundamental combinatorial property about Euclidean geometry and the relation between points and lines. The salient fact is

Theorem 2 *(Szemerédi-Trotter '83 [11])* $I(n, m) = O(n^{2/3}m^{2/3} + m + n)$.

Proof: Take sets $\mathcal{P}$ of n points and $\mathcal{L}$ of m lines for which $I(\mathcal{P}, \mathcal{L}) = I(n, m)$. Consider the graph $G = (V, E)$ whose vertices are the points of $\mathcal{P}$. The edges in E join P_i and P_j if and only if they are on the same line $\ell \in \mathcal{L}$ AND they are adjacent (no other point of $\mathcal{P}$ is on ℓ and "between" these points. For any line $\ell \in \mathcal{L}$ the number of points of $\mathcal{P}$ incident with ℓ is 1 plus the number of edges ℓ contributes to E. Adding up over all lines in $\mathcal{L}$ we get

$$I(\mathcal{P}, \mathcal{L}) = |E| + m = I(n, m). \tag{4}$$

By the crossing number lower bound of Theorem 1 (using one of the remarks)

$$CR(G) \geq c\frac{|E|^3}{n^2} - an.$$

On the other hand suppose lines $\ell_1 \in \mathcal{L}$ and $\ell_2 \in \mathcal{L}$ intersect. That common point can be in at most one edge of G in ℓ_1 and at most one edge of G in ℓ_2. Therefore

$$CR(G) \leq \binom{m}{2}.$$

Combining the two inequalities gives $c|E|^3 \leq \binom{m}{2}n^2 + an^3$ and since the cube root is a concave function,

$$|E| = O(n^{2/3}m^{2/3} + n).$$

Applying (4) gives the desired conclusion.●

Remarks:

1. Taking $n = m$, the theorem says that n points and n lines have $O(n^{4/3})$ incidences. This is the correct order of magnitude because of the following example which has n points, n lines, and $cn^{4/3}$ incidences, $c = 2^{-1/3} = .7937....$ The points will be $\mathcal{P} = \{(i, j) : i = 1, \ldots, k;\ j = 1, \ldots, 2k^2\}$. There are $n = 2k^3$ points in all. The lines are $\mathcal{L} = \{\ell_{i,j}, i = 1, \ldots, k;\ j = 1, \ldots, 2k^2\}$, a set of size n, where line $\ell_{i,j}$ has equation $y = ix + j - 1$. It is clear that each line in $\mathcal{L}$ meets exactly k points in $\mathcal{P}$, so $I(\mathcal{P}, \mathcal{L}) = nk = n(n/2)^{1/3}$.

2. There is another construction that attains the bound in Theorem 2 for the cases where $m \neq n$.

3. The present proof using crossing number is very simple. The original proof is complicated and difficult. Check it out [11]!

Application 2 - Halving Sets: Given a set $S = \{P_1, \ldots, P_n\}$ of $n = 2k$ points in the plane, every line ℓ partitions S into three sets: points which are ON ℓ, points which are on one particular side of ℓ (we call this set L, for left) and those on the other side (which we call this set R, for right). It is interesting to know how many different partitions there can be with $|L| = |R|$ (the sets L and R in such a partition are halving sets). The answer has relevance in several application areas, for example machine-learning and is still one of the main open problems in combinatorial geometry.

We form the geometric graph $G = (V, E)$ whose vertices are the points of S. For $i < j$ there will be an undirected edge $e = P_iP_j \in E$ if and only if the line through P_i and P_j has $(n - 2)/2$ points of S on *both* sides (so P_i and P_j are on the line and the rest are evenly split between the two open halfspaces). The vertices in such an edge are a *"halving pair"*. The graph is the halving graph for S. Write $h(S) = |E|$, the number of halving pairs in S, and

$$h(n) = \max\left(h(S) : |S| = n\right);$$

this is the maximal number of edges that are possible in a halving graph G with n vertices. Results by Erdös ('71) and by Erdös, Lóvasz, Simmons, Strauss ('73) [4] showed that there are constants c and d for which

$$cn \log n \leq h(n) \leq dn^{3/2}. \tag{5}$$

About 20 years later the upper bound was reduced to $dn^{3/2}/\log^* n$ [8]. Though the improvement is negligible, it showed that the upper bound in (5) is not the correct order for $h(n)$. The proof is long and intricate, and very much more difficult than the proof of (5) itself.

In '98 T. Dey [3] used the crossing number to shatter the the existing upper bound. Using a simple geometric argument (which we omit) he showed that for a halving graph G with n vertices,

$$CR(G) \leq an^2;$$

(note that with n vertices, G potentially has $O(n^2)$ edges and any pair might cross). Thus Dey's bound above improves on the naive $O(n^4)$ bound on the crossing number of any graph.

On the other hand by Theorem 1,

$$CR(G) \geq b\frac{|E|^3|}{n^2}.$$

Combining the inequalities we see that $|E| \leq dn^{4/3}$ (much smaller than the bound in (5)).

Remarks:

1. Here is a situation where it seems difficult to precisely upper bound the number of edges in a graph. On the other hand for our halving pair graph, Dey was able to crudely upper bound a *function* of the number of edges, namely the crossing number of the graph. Combined with Theorem 1 this translated into a relatively simple method for getting a good upper bound on $|E|$.

2. In '01 [12], the lower bound for $h(n)$ in (5) was (somewhat surprisingly) raised to $\Omega(n2^{\sqrt{\log n}})$ (note that $2^{\sqrt{\log n}}$) is much larger than $\log n$).

References

[1] M. Aigner and G. Ziegler. *Proofs from the Book, third edition*, Springer-Verlag, Berlin, (2003).

[2] M. Ajtai, V. Chvátal, M. Newborn, and E. Szemerédi. "Crossing-free subgraphs". *Theory and Practice of Combinatorics, North-Holland Mathematics Studies, vol. 60.*, 9-12 (1982).

[3] T. Dey. "Improved bounds for planar k-sets and related problems". *Discrete and Computational Geometry 19 (3)* 373-382 (1998).

[4] P. Erdös, L. Lovász, A. Simmons, and E. Strauss. Dissection graphs of planar point sets. In J. Srivastava and et.al., editors, *A Survey of Combinatorial Theory*, North-Holland, Amsterdam, 139-149 (1973).

[5] L. Lovász. On the number of halving lines. *Ann. Univ. Sci. Budapest, Eötvös, Sect. Math.*, 14: 107-108, (1971).

[6] T. Leighton. *Complexity Issues in VLSI. Foundations of Computing Series* MIT Press, Cambridge, MA, (1983).

[7] J. O.'Rourke. *Computational Geometry in C, second ed.* Cambridge U. Press, New York (1998).

[8] J. Pach, W. Steiger, and E. Szemerédi. "An upper bound on the number of planar k-sets". *Discrete and Computational Geometry 7, 109-123 (1992).*

[9] J. Pach and G. Tóth. "Graphs drawn with few crossings per edge". Combinatorica 17*, 427-439 (1997).*

[10] L. Székely. "Crossing numbers and hard Erdös problems in discrete geometry". *Combinatorics, Probability and Computing 6*, 353-358 (1997).

[11] E. Szemerédi and T. Trotter. "Extremal problems in discrete geometry". *Combinatorica 3*, 381-392 (1983).

[12] G. Tóth. "Point sets with many k-sets". *Discrete and Computational Geometry 26 (2)*, 187-194 (2001).