Geometric graphs with no self-intersecting path of length three*

János Pach[†] Rom Pinchasi[‡] Gábor Tardos[§]

Abstract

Let G be a *geometric graph* with n vertices, i.e., a graph drawn in the plane with straight-line edges. It is shown that if G has no self-intersecting path of length 3, then its number of edges is $O(n \log n)$. This result is asymptotically tight. Analogous questions for curvilinear drawings and for arrays are also considered.

1 Introduction

A *geometric graph* is a graph drawn in the plane so that its vertices are points and its edges are possibly crossing straight-line segments. We assume, for simplicity, that the points are in *general position*, i.e., no three points are on a line and no three edges pass through the same point. *Topological graphs* are defined similarly, except that now the edges are not necessarily rectilinear; every edge can be represented by an arbitrary continuous arc which does not pass through any vertex different from its endpoints. Throughout this paper, we also assume that if two edges have an interior point in common, then at this point they properly cross. Clearly, every geometric graph is a topological graph.

Using this terminology, the fact that every planar graph with n vertices has at most $3n-6$ edges can be rephrased as follows: any topological graph with n vertices and more than $3n-6$ edges must have two edges that cross each other. This result is tight even for geometric graphs.

*Work on this paper by János Pach and Rom Pinchasi has been supported by NSF grant CCR-00-98246, by PSC-CUNY Research Award 63382-0032. Work by János Pach and Gábor Tardos has also been supported by Hungarian Science Foundation grant OTKA T-032452.

[†]City College, CUNY and Courant Institute of Mathematical Sciences, New York University, New York, NY 10012, USA; pach@cims.nyu.edu

[‡]Department of Mathematics, Massachusetts Institute of Technology, Cambridge, MA 02139, USA; room@math.mit.edu

[§]Rényi Institute of the Hungarian Academy of Sciences, H-1364 Budapest, P.O.B. 127, Hungary; tardos@renyi.hu

In the mid-sixties Avital and Hanani [AH66], Erdős, and Perles initiated, later Kupitz [K79] and many others continued the systematic study of extremal problems for geometric graphs. In particular, they proposed the following general question. Let H be a so-called *forbidden* geometric configuration or a *class* of forbidden configurations. For example, H may consist of k pairwise crossing edges or may be the class of all configurations of $k+1$ edges, one of which crosses all the others, etc. *What is the maximum number of edges that a geometric graph with n vertices can have without containing any forbidden subconfiguration?* If H consists of $k = 2$ crossing edges, then, according to the previous paragraph, the answer is $3n - 6$. For $k = 3$, this maximum is linear in n (see [AAPPS97]), but for larger values of k the best known bound due to Valtr is only $O(n \log n)$ [V98]. It is an exciting open problem to decide whether one can get rid of the logarithmic factor here. If H is the class of all configurations consisting of $k+1$ edges, one of which crosses all the others, then the maximum number of edges is equal to $(k+2)(n-3)$, provided that $k = 1, 2, 3$, and the maximum is $O(\sqrt{k}n)$ for large values of k (cf. [PT97]). For a survey of many similar results in Geometric Graph Theory, consult [P99].

The above questions can also be regarded as geometric analogues of the fundamental problem of Extremal Graph Theory [B78]: determine the maximum number of edges of all K-*free* graphs on n vertices, i.e., all graphs which do not contain a subgraph isomorphic to a fixed graph K. Denote this maximum by $\mathrm{ex}(n, K)$.

In the present note, we consider the special instance of the above question when H consists of all *self-intersecting* straight-line drawings of a fixed graph K. In other words, what is the maximum number $\mathrm{ex}_{\mathrm{cr}}(n, K)$ of edges that a geometric graph with n vertices can have, if it contains no self-intersecting copy of K? Obviously, we have $\mathrm{ex}_{\mathrm{cr}}(n, K) \geq \mathrm{ex}(n, K)$, because if a graph contains no copy of K, then it cannot contain a self-intersecting copy either. Therefore, if K is not a bipartite graph, then $\mathrm{ex}_{\mathrm{cr}}(n, K)$ is quadratic in n. The question is more exciting for bipartite planar graphs. What happens if $K = P_k$ (or $K = C_k$), a path (or a cycle) of (an even) length k? The case where $K = C_4$ is discussed in [PR02].

We analyze the case when $K = P_3$. The corresponding graph property is a relaxation of planarity: the geometric graphs satisfying the condition are allowed to have two crossing edges, but if this is the case, no endpoint of one of these edges can be joined to an endpoint of the other. Is it still true that the number of edges of such geometric graphs is $O(n)$? The following theorem provides a negative answer to this question.

Theorem 1.1. *The maximum number of edges of a geometric graph with n vertices, containing no self-intersecting path of length* 3, *satisfies*

$$\mathrm{ex}_{\mathrm{cr}}(n, P_3) \leq cn \log n,$$

for a suitable constant c. Apart from the value of the constant, this bound cannot be improved.

The proof of this result (presented in three different versions in the next three sections) applies to a slightly more general situation. Theorem 1.1 remains true for *topological* graphs whose edges are continuous functions defined on subintervals of the x-axis, i.e., every line perpendicular to the x-axis intersects each edge in at most one point. The topological graphs satisfying this condition are usually called *x-monotone*.

What happens if we drop the requirement of x-monotonicity? We do not have any example of a topological graph with n vertices and more than constant times $n \log n$ edges, in which every path of length 3 is *simple*, i.e., non-self-intersecting. The best *upper* bound we have is the following.

Theorem 1.2. *The maximum number of edges of a topological graph with n vertices, containing no self-intersecting path of length 3, is $O(n^{3/2})$.*

As was pointed out by Tutte [T70], *parity* plays an important role in determining the possible crossing patterns between the edges of a topological graph. This may well be a consequence of the Jordan Curve Theorem: every Jordan arc connecting an interior point and an exterior point of a simple closed Jordan curve must cross this curve an *odd* number of times. In particular, Tutte showed that every topological graph with n vertices and more than $3n - 6$ edges has two edges that not only cross each other, but (i) they cross an *odd* number of times, and (ii) they do not share an endpoint. (See also [H34].)

This may suggest that Theorem 1.2 and perhaps any other bound of this type can be sharpened as follows.

Theorem 1.3. *The maximum number of edges of a topological graph with n vertices, containing no path of length 3 whose first and last edges cross an odd number of times, is $O(n^{3/2})$.*

In Section 5 we prove this stronger statement. Somewhat surprisingly (to the authors), it turns out that this last result is asymptotically tight. More precisely, in Section 6 we establish

Theorem 1.4. *Let G be a bipartite graph on n vertices, containing no cycle of length 4. Then G can be drawn in the plane as an x-monotone topological graph with the property that any two edges belonging to a path of length 3 cross an even number of times.*

It is well known that there are C_4-free bipartite graphs of n vertices and at least constant times $n^{3/2}$ edges (see e.g. [B78]).

2 A Davenport-Schinzel bound for double arrays

In this section, we discuss the special case of Theorem 1.1 when G is a *bipartite* geometric (or x-monotone topological) graph, whose vertices are divided by the y-axis into two classes, A and B, and all edges of G run between these classes. We assume, for simplicity, that no two edges of G cross the y-axis at the same point.

Let $a_1b_1, a_2b_2, \ldots, a_mb_m$ be the edges of G listed from top to bottom, in the order of their intersections with the y-axis, where $a_i \in A$ and $b_i \in B$ for every i. Consider the corresponding *double array* ($2 \times m$ matrix)

$$M = \begin{pmatrix} a_1 & a_2 & \ldots & a_m \\ b_1 & b_2 & \ldots & b_m \end{pmatrix}$$

It is easy to verify that if G is a geometric graph (or an x-monotone topological graph) without any self-intersecting path of length three, then the corresponding matrix M does not contain any submatrix of the form $F_1 = \begin{pmatrix} u & v & u & v \\ * & x & x & * \end{pmatrix}$ or $F_2 = \begin{pmatrix} * & u & u & * \\ x & y & x & y \end{pmatrix}$, where $u \neq v$, $x \neq y$ and $*$ stands for an unspecified entry (see Fig. 1(a)).

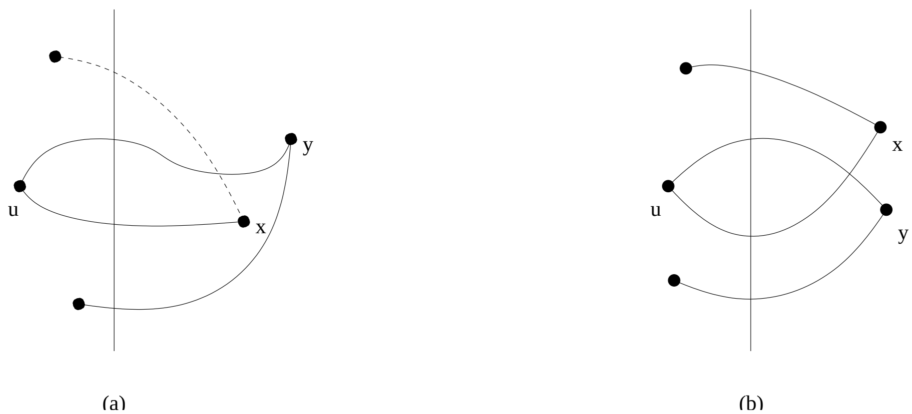

Fig. 1. *(a) F_2 is forbidden, (b) not necessarily forbidden if adjacent edges may cross*

In what follows, we show that if a $2 \times m$ matrix M having at most n distinct entries does not contain any forbidden submatrix of the above two types, then its number of columns is $O(n \log n)$. Therefore, the number of edges of G is at most $O(n \log n)$, as required by Theorem 1.1.

If G is an x-monotone topological graph whose adjacent edges are allowed to cross, and we only require that the first and last edges of every path of length three must be disjoint, then the situation is slightly more complicated, because M may contain submatrices of the above forms (see Fig. 1(b)). However, in this case the following 2×6 submatrices are forbidden:

$$\begin{pmatrix} v & \leftrightarrow & u & v & u & v & \leftrightarrow & u \\ * & & * & x & x & * & & * \end{pmatrix} \tag{1}$$

and

$$\begin{pmatrix} * & & * & u & u & * & & * \\ y & \leftrightarrow & x & y & x & y & \leftrightarrow & x \end{pmatrix}. \tag{2}$$

Here the signs $\leftrightarrow$ indicate that the order of the first two columns and the order of the last two columns are not specified.

Theorem 2.1. *Let M be a $2 \times m$ matrix with at most n distinct entries, all of whose columns are different. If M has no 2×6 submatrix of types (1) or (2), then $m \leq 17n \log_2 n$.*

Proof. We need some definitions. Let

$$M = \begin{pmatrix} a_1 & a_2 & \dots & a_m \\ b_1 & b_2 & \dots & b_m \end{pmatrix}$$

For any $1 \leq i \leq m$, we say that a_i is a *leftmost* (or *rightmost*) entry if $a_k \neq a_i$ for every $k < i$ (or $k > i$, resp.). Accordingly, a_i is called a *second leftmost* (or *second rightmost*) entry if $a_k = a_i$ for precisely one index $k < i$ (or precisely one index $k > i$, resp.). Analogous terms are used for the entries b_i in the second row of M.

A set of consecutive columns of M is called a *block*. A block is said to be *pure* if all elements in the first row of the block are distinct and the same is true for the elements in the second row.

Assume the columns of M are partitioned into l pure blocks. Consider now two consecutive pure blocks, B_1 and B_2, consisting of the columns $i+1, i+2, \dots, j$ and $j+1, j+2, \dots, k$, resp., for some $0 \leq i < j < k \leq n$. Suppose that there is an element which appears in the first row of B_1 as well as in the first row of B_2. That is, $a_p = a_q$ for some $i < p \leq j$ and $j < q \leq k$. We claim that either b_q is a leftmost, second leftmost or rightmost entry, or b_p is a rightmost, second rightmost or leftmost entry. Indeed, otherwise, using the fact that b_q is neither a leftmost nor a second leftmost entry, we obtain that there exists an index $r \leq i$ such that $b_r = b_q$. Since b_q is not a rightmost entry, there is an index $s > k$ such that $b_s = b_q$. Similarly, in view of the fact that b_p is neither a rightmost nor a second rightmost entry, we can conclude that $b_{s'} = b_p$ for some $s' > k$. Since b_p is not leftmost, there is a $r' \leq i$ such that $b_{r'} = b_p$. Observe that now the columns $r, r' < p < q < s, s'$ form a forbidden submatrix of type

$$\begin{pmatrix} * & & * & u & u & * & & * \\ y & \leftrightarrow & x & y & x & y & \leftrightarrow & x \end{pmatrix},$$

a contradiction.

A symmetric argument shows that if $b_p = b_q$ for some $i < p \leq j$ and $j < q \leq k$, then either a_q is a leftmost, second leftmost or rightmost entry, or a_p is a rightmost, second rightmost or leftmost entry. Thus, if we delete from M (and from its block decomposition) every column whose upper or lower element is a leftmost, second leftmost, rightmost, or second rightmost entry, the union of the remainders of any two consecutive blocks becomes pure.

There are at most n distinct entries, each may appear in the first row and in the second row, so the number of deleted columns is at most $8n$. The resulting matrix M'

can be decomposed into $\lceil l/2 \rceil$ pure blocks. Repeating this process at most $\lceil \log_2 l \rceil$ times, we end up with a matrix consisting of at most $m - 8n\lceil \log_2 l \rceil$ columns that form a single pure block. Thus, we have

$$m - 8n\lceil \log_2 l \rceil \leq n.$$

Applying the above procedure to the initial partition of M into $l = m$ pure blocks, each consisting of a single column, the upper bound follows. ■

For many other Davenport-Schinzel type results for matrices, consult [FH92].

As we have pointed out before, the last theorem implies that every geometric or x-monotone topological graph with n vertices and no path of length three whose first and last edges cross each other, has at most constant times $n \log n$ edges, provided that all of its edges can be stabbed by a line. Thus, we immediately obtain

Corollary 2.2. *The maximum number of edges of an x-monotone topological graph with n vertices, containing no path of length 3 whose first and last edges cross, is $O(n \log^2 n)$.*

For geometric graphs, this result is slightly weaker than the bound in Theorem 1.1.

In the rest of this section, we show that the bound in Theorem 2.1 is tight apart from the value of the constant. In fact, we establish a slightly stronger assertion.

Theorem 2.3. *For every n, there exists a $2 \times \lfloor cn \log n \rfloor$ matrix M_n, such that all entries belong to the set $\{1, 2, \ldots, n\}$, no two columns are the same, and it has no submatrix of the form $\begin{pmatrix} u & v & u & v \\ * & x & x & * \end{pmatrix}$ or $\begin{pmatrix} * & u & u & * \\ x & y & x & y \end{pmatrix}$. Here c is a suitable positive constant.*

Proof. Let $M_1 = \binom{1}{1}$. Suppose that we have already constructed a $2 \times m_n$ matrix

$$M_n = \begin{pmatrix} a_1 & a_2 & \ldots & a_m \\ b_1 & b_2 & \ldots & b_m \end{pmatrix}$$

meeting the requirements, in which each number $1 \leq k \leq n$ appears in both rows. For any $1 \leq k \leq n$, let $i(k)$ denote the *smallest* index i for which $a_i = k$ and let $j(k)$ denote the *largest* index j for which $b_j = n + k$. Permuting the numbers if necessary (separately for the top and the bottom row), we can assume that $i(1) < i(2) < \ldots < i(n)$ and $j(1) < j(2) < \ldots < j(n)$.

Consider the matrix

$$M'_n = \begin{pmatrix} a'_1 & a'_2 & \ldots & a'_m \\ b'_1 & b'_2 & \ldots & b'_m \end{pmatrix},$$

where $a'_i = a_i + n$ and $b'_i = b_i + n$ for every i.

Consider the matrix obtained from M_n and M'_n by merging them so that the order of columns within both parts remains unchanged, and, for every $1 \leq k \leq n$, the i_k-th column of M_n follows immediately after the j_k-th column of M'_n. Finally, for every k, insert a so-called *extra* column $\binom{a_{i(k)}}{b'_{j(k)}} = \binom{k}{n+k}$ between $\binom{a'_{j(k)}}{b'_{j(k)}} = \binom{a'_{j(k)}}{n+k}$ and and $\binom{a_{i(k)}}{b_{i(k)}} = \binom{k}{b_{i(k)}}$. Denote the resulting matrix by M_{2n}.

Clearly, M_{2n} has $m_{2n} = 2m_n + n$ columns. It remains to verify that M_{2n} does not have any forbidden submatrix. It is sufficient to prove, by symmetry, that M_{2n} has no submatrix of the form $\begin{pmatrix} u & v & u & v \\ * & w & w & * \end{pmatrix}$

By the induction hypothesis, neither M_n nor M'_n has any forbidden submatrix, so if M_{2n} has a submatrix F of the above type, at least one of its columns must belong to the set of n *extra* columns added in the last step of the construction. Let F be such a submatrix with the *minimum* number of extra columns.

Case 1: The first column of F is an extra column $\binom{a_{i(k)}}{b'_{j(k)}} = \binom{k}{n+k}$, for some $1 \leq k \leq n$. Then the next column of M_{2n}, $\binom{a_{i_k}}{b_{i_k}} = \binom{k}{b_{i_k}}$ cannot belong to F, so we could replace $\binom{a_{i_k}}{b'_{j_k}}$ by this column to obtain a forbidden submatrix with fewer extra columns.

Case 2: The second column of F is an extra column $\binom{a_{i(k)}}{b'_{j(k)}} = \binom{k}{n+k}$, for some $1 \leq k \leq n$. This is impossible, because then in the second row of F (and hence in the second row of M'_n) the entry $n + k$ appears after this place, contradicting the definition of $j(k)$.

The cases when the third or the fourth column of F is an extra column can be treated similarly to Cases 2 and 1, resp. ■

3 Proof of Theorem 1.1

We prove the following more general statement.

Theorem 3.1. *Let G be an x-monotone topological graph of n vertices, which has no self-intersecting path of length 3. Then G has at most constant times $n \log n$ edges.*

We assume without loss of generality that no two edges that share an endpoint cross each other. Otherwise, the two non-common endpoints of these edges must be of degree 1 or 2, because G has no self-intersecting path of length 3. So we can delete these endpoints, and complete the argument by induction on the number of vertices.

It will be convenient to use the following terminology. If a vertex v is the left (resp. right) endpoint of an edge e, then e is said to be a *right* (resp. *left*) edge at v. It follows from our assumption on adjacent edges that the left and the right edges at a given vertex can be ordered from bottom to top.

Let $e_1 = vu_1$ and $e_2 = vu_2$ be two right edges at a vertex v such that the x-coordinate of u_1 is at most as large as the x-coordinate of u_2. We define the *right triangle* determined by them as the bounded *closed* region bounded by e_1, a segment of e_2,, and a segment of the vertical line passing through u_1. The vertex v is called the *apex* of this triangle. Analogously, we can introduce the notion of *left triangle.*

Construct a sequence of subgraphs $G_0, G_1, G_2, \ldots$ of G, as follows. Let $G_0 = G$. If G_i has already been defined for some i, then let G_{i+1} be the topological graph obtained from G_i by deleting at each vertex the bottom 2 and the top 2 left and right edges (if they exist). We delete at most 8 edges per vertex.

Claim 3.2. *For any $k \geq 0$, every triangle determined by two edges of G_k contains at least 2^k pairwise different triangles of G.*

Proof. By induction on k. Obviously, for $k = 0$, the claim is true, because every triangle contains itself. Assume that the claim holds for $k-1$ ($k > 0$). Consider, e.g., a *right* triangle T in G_k, determined by the edges $e_1 = vu_1$ and $e_2 = vu_2$, where the x-coordinate of u_1 is at most as large as the x-coordinate of u_2. Suppose without loss of generality that e_1 lies below e_2. Using the fact that $e_1 \in E(G_k)$, we obtain that at u_1 there are at least two left edges $f_1, f_2 \in E(G_{k-1})$ which lie above e_1. Both of these edges must be entirely contained in T, otherwise we could find a self-intersecting path of length 3. Suppose that f_1 lies below f_2.

Let T_1 and T_2 denote the left triangles with apex u_1, determined by e_1 and f_1, and by f_1 and f_2, resp. Clearly, T_1 and T_2 both belong to G_{k-1}, and they have disjoint interiors. By the induction hypothesis, both T_1 and T_2 have 2^{k-1} pairwise different triangles. It follows that T contains 2^k pairwise different triangles, as required. ■

Now we can easily complete the proof of Theorem 3.1. Since every triangle is specified by a pair of edges meeting at its apex, the total number of different triangles is at most n^3. Hence, for $k > 3\log_2 n$, the graph G_k cannot determine any triangle, and its number of edges is smaller than n. On the other hand, we have that $|E(G_k)| \geq |E(G_0)| - 8kn$. Therefore, $|E(G)| = |E(G_0)| \leq 25n \log_2 n$, completing the proof of Theorem 3.1.

We close this section by showing that, up to the value of the constant c, Theorem 1.1 (and hence Theorem 3.1, too) is best possible. We will recursively construct a sequence of bipartite geometric graphs G_i with 2^i vertices and $(i+1)2^{i-2}$ edges, containing no self-intersecting path of length 3 ($i \geq 1$). Furthermore, we will maintain the following properties for every i.

1. The left and right endpoint of every edge of G_i is black and white, resp.

2. The horizontal edges of G_i are of the same length and form a perfect matching.

3. No two vertices of G_i have the same x-coordinate.

4. For any vertex v of G_i, the order of the edges incident to v according to their slopes coincides with the order according to the lengths of their projections to the x-axis.

Let G_1 consist of a horizontal straight-line segment connecting a black point to a white point of larger x-coordinate. Obviously, this meets the requirements.

Assuming that we have already constructed G_i for some i, we show how to obtain G_{i+1}. Let G'_i be a (color-preserving) translate of G_i such that the leftmost vertex of G'_i is to the right of the rightmost vertex of G_i, every vertex of G'_i lies above all the connecting lines between vertices of G_i, and every vertex of G_i lies below all the connecting lines of G'_i. Let G_{i+1} be the union of G_i and G'_i, together with the following 2^{i-1} edges: connect every black vertex $v \in V(G_i)$ to the white endpoint of the (uniquely determined) horizontal edge in G'_i incident to $T(v)$. Observe that all the edges we added are parallel and they do not pass through any other vertex different from their endpoints. See Fig. 2.

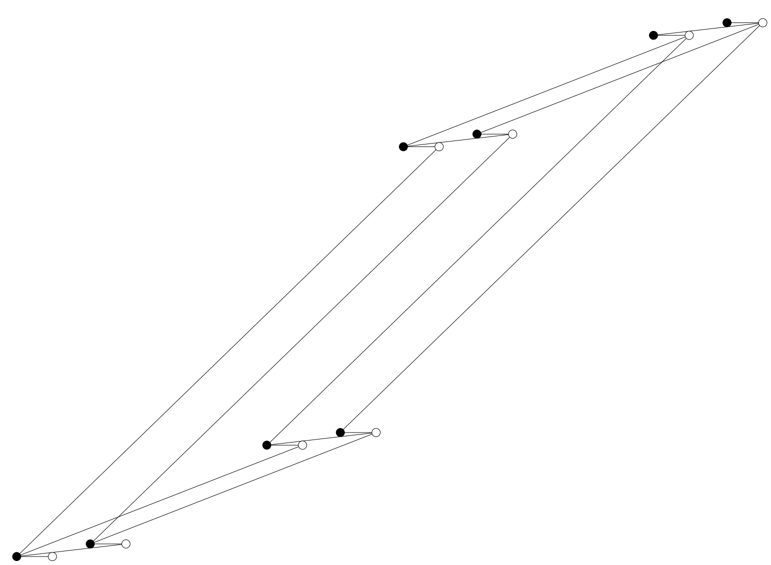

Fig. 2. *The construction of G_k ($k = 4$)*

We have to check that G_{i+1} has the required properties. Clearly, we have $V(G_{i+1})| = 2|V(G)| = 2^{i+1}$ and $|E(G_{i+1})| = 2|E(G_i)| + 2^{i-1} = (i+2)2^{i-1}$. Properties 1, 2, and 3 are all inherited from G_i. To see that condition 4 is satisfied, it is sufficient to observe that every new edge between G_i and G'_i is maximal in both orderings of the edges around both of its endpoints.

It remains to verify that G_{i+1} does not contain a self-intersecting path of length 3. Assume to the contrary that there is such a path P in G_{i+1}, and denote its edges by $e_1 = uv$, $e_2 = vw$, and $e_3 = wz$. Since G_i (and thus G'_i) does not contain a self-intersecting path of length 3, at least one of these edges must be an edge between G_i

and G_i'. Note that there cannot be two such edges, because all edges of G_{i+1} running between G_i and G_i' are parallel. It is also clear that e_2 cannot be such an edge.

Assume, without loss of generality, that e_1 runs between G_i and G_i', and that we have $u \in V(G_i)$ and $v \in V(G_i')$. Thus, e_2 and e_3 belong to G_i'. As v is white, w must be black, and both e_2 and e_3 are to the right of w. Since e_3 crosses e_1, the slope of e_3 must be smaller than the slope of e_2. In view of property 4 for the order of the edges incident to w, we conclude that the x-coordinate of z is smaller than the x-coordinate of v. This implies that the line connecting z and v passes below w, contradicting our assumption that every connecting line of G_i' must pass above all vertices in G_i.

4 A strengthening of Theorem 3.1

The aim of this section is to establish the following stronger form of Theorem 3.1.

Theorem 4.1. *The maximum number of edges of an x-monotone topological graph with n vertices, containing no path of length* 3 *whose first and last edges cross, is* $O(n \log n)$.

Proof. Let G be an x-monotone topological graph with n vertices and m edges, containing no path of length 3 whose first and last edges cross. Our goal is to construct another topological graph G' with $n' \leq 2n$ vertices and $m' \geq m/2 - n$ edges with the property that G' has no path of length 3 whose first and last edges cross, and G' does not have two adjacent edges that cross. Applying Theorem 3.1 to G', the statement follows.

First we assume that G is bipartite, and it has no vertex which is the left endpoint of an edge and the right endpoint of another. This can be achieved by simply splitting each vertex in two.

By the *length* of an edge we mean the length of its projection to the x-axis, and we can use the terms *shorter* and *longer* in this sense. We denote an edge $e \in E(G)$ by $e = uv$, where u is the left and v is the right endpoint of e. We call an edge $e = uv$ *long* if it is the longest either among all edges uv' or among all edges $u'v \in E(G)$. Clearly, there are fewer than $2n$ long edges in G. Let e and e' be two edges of G with e shorter than e' and either $e = uv$ and $e' = uw$, or $e = vu$ and $e' = wu$. We say that e is *above* e' if v is above e'. Similarly, we say e is *below* e' if v is below e' . Note that if e is above or below e' then e is shorter, but e and e' may cross several times.

Let $e = uv$ be an edge of G which is not long. By definition, there exist two edges, $e' = uw$ and $e'' = zv \in E(G)$, such that both of them are longer than e. So e is either above or below e' and the same is true for e''. However, e cannot be above both e' and e''. Indeed, otherwise u is above e'' while v is above e', so e' and e'' cross, contradicting our assumption on G. Similarly, e cannot be below both e' and e''. Thus, each edge $e = uv \in E(G)$ which is not not long either satisfies that e is above every longer edge uw and below every longer edge zv, or it satisfies that e is below every longer edge uw and above every longer edge zv. Without loss of generality, we can assume by symmetry that the former condition holds (which we will refer to as the *monotonicity condition*) for $m' \geq m/2 - n$ edges. Let G'' be the subgraph of G formed by these edges.

We are now in a position to define G'. As an abstract graph, G' will be the same as G". The locations of the vertices will coincide, too. For any edge $e \in E(G'')$, let us denote by $\hat{e}$ the corresponding edge of G'. We draw the edges of G' one by one in decreasing order of length. If e in G'' is neither above nor below another edge, set $\hat{e} = e$. If e is above (below) at least one other edge, let e_- be the shortest edge such that e is above e_- (let e_+ be the shortest edge such that e is above e_+, resp.). Draw $\hat{e}$ in such a way that all of its internal points lie strictly above $\hat{e}_-$ and below $\hat{e}_+$ (if these edges exist). Notice that, if they exist, e_+ and e_- are longer than e, so $\hat{e}_+$ and $\hat{e}_-$ are already defined. We make sure during the construction that if both of these edges exist, than they are disjoint (see property 4 below), with $\hat{e}_-$ passing below $\hat{e}_+$ (see property 2 below). We define $\hat{e}$ to follow e, except for the intervals where $\hat{e}_+$ is below e or $\hat{e}_-$ is above e, because here it should run just below $\hat{e}_+$ or just above $\hat{e}_-$, close enough not to intersect more edges and going on the same side of every vertex. See Fig. 3.

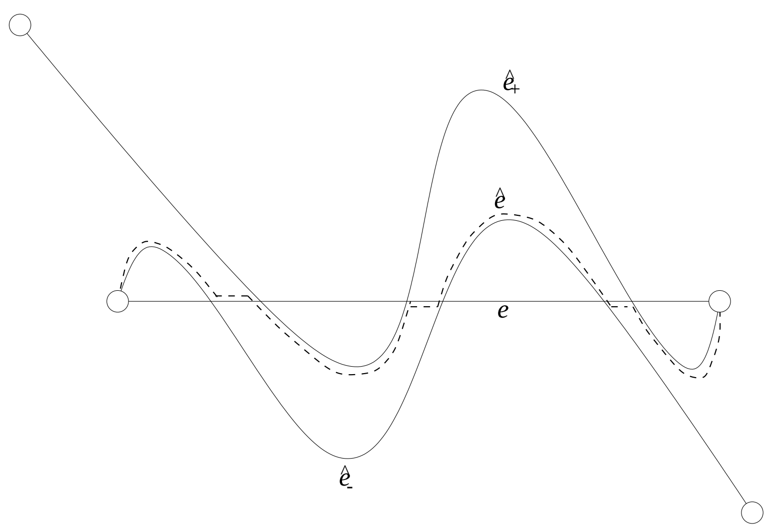

Fig. 3. *The construction of the edge $\hat{e}$ in G'*

We claim that the resulting graph G' has the following properties.

1. If e is below (above) e' in G'', then every interior point of $\hat{e}$ is below (above, resp.) $\hat{e}'$.

2. If e' is below (above) e in G'', then the endpoint of e' which is not an endpoint of e is below (above, resp.) $\hat{e}$.

3. If e, e', and e'' form a path in G'' and e is longer than e', then $\hat{e}$ and e'' do not cross.

4. If e, e', and e'' form a path in G'' then $\hat{e}$ and $\hat{e}''$ do not cross.

We verify these properties by showing that if they hold for the partially drawn graph, they do not get violated when we add an extra edge $\hat{e}$.

(1) By the monotonicity, if there are edges that e is below, the shortest, e_+, is below all others, similarly e_- (if exists) is above all other edges that e is above. So, as property 1 holds so far, it does not get violated provided that $\hat{e}$ is in between $\hat{e}_-$ and $\hat{e}_+$, which is the case.

(2) Let $e = uv$ and assume $e' = uw$ is above e. By definition, w is above e and, by the monotonicity condition, w is above e_-, if the latter exists. As property 2 holds so far, w is above $\hat{e}_-$, so w must be above $\hat{e}$. Similarly, if $e' = zv$ is below e, then z is below $\hat{e}$.

(3) Note that e' is above or below e. By symmetry, we assume without loss of generality, that e' is below e. By monotonicity, this means that they share their right endpoints. Here e and e'' do not cross, as they are first and last edges of a path of length 3, and the left endpoint of e'' is below e. So every point of e'' must be below e or to the right from the right endpoint of e. If e_+ exists, we can apply property 3 to the edges e_+, e', e'' and find that $\hat{e}_+$ does not cross e''. By the construction, wherever $\hat{e}$ runs below e, it follows $\hat{e}_+$, so $\hat{e}$ is disjoint from e''.

(4) We consider two cases.

If both e and e'' are shorter than e', then one of them is below, the other is above e' (monotonicity). Thus, $\hat{e}'$ separates $\hat{e}$ from $\hat{e}''$, by property 1, so they do not cross.

We may assume that e is shorter than e'' (switch them if not), so in the remaining case e'' is longer than e'. So e' is below or above e'', and we can again assume by symmetry, that e' is below e''. Applying property 3 to the path e'', e', e, we find that e is disjoint from $\hat{e}''$. By property 2, the left endpoint of e lies below $\hat{e}''$, so all points of e must be below e'' or to the right of its right endpoint. As $\hat{e}$ follows $\hat{e}_-$, wherever it runs above e, it is enough to show that if e_- exists, $\hat{e}_-$ is disjoint from $\hat{e}''$. If $e_- = e'$, this follows from property 1, otherwise from property 4 of the initial configuration (before drawing $\hat{e}$).

Observe that no two adjacent edges of G' cross each other, by property 1, and no second neighbors, by property 4. So, we can indeed apply Theorem 3.1 to G' and Theorem 4.1 follows. ■

5 Forbidden subgraphs – Proof of Theorem 1.3

For any $k \geq 2$, let F_k denote a graph with vertex set

$$V(F_k) = \{x, y\} \cup \{b_i : 1 \leq i \leq k\} \cup \{c_{ij} : 1 \leq i < j \leq k\}$$

and edge set

$$E(F_k) = \{xb_i, yb_i : 1 \le i \le k\} \cup \{c_{ij}b_i, c_{ij}b_j : 1 \le i < j \le k\}.$$

We need the following theorem, which can be obtained by a straightforward generalization of a result of Füredi [F91].

Theorem 5.1. *For any fixed integer $k \ge 2$, let $\mathrm{ex}(n, F_k)$ denote the maximum number of edges of an F_k-free graph with n vertices. Then we have $\mathrm{ex}(n, F_k) = O(n^{3/2})$.*

Let G be a topological graph with n vertices, containing no path of length 3 whose first and last edges cross an *odd* number of times. To establish Theorem 1.1, it is sufficient to verify that the *abstract* graph obtained from G by disregarding how the edges are drawn does not have a subgraph isomorphic to F_4. We make no notational distinction between the vertices (edges) of our topological graph G and the vertices (edges) of the underlying abstract graph.

Suppose, in order to obtain a contradiction, that G has a forbidden subgraph, and denote its vertices by x, y, b_i, for $1 \le i \le 4$, and c_{ij}, for $1 \le i < j \le 4$. We use the following simple topological fact. If two (possibly self-intersecting) closed curves intersect each other a finite number of times, then an *even* number of these intersections is proper. Consider the following closed curves, each consisting of the edges of a cycle of length 4 in the graph:

$$\gamma_1 = xb_1yb_4x, \quad \delta_1 = xb_2c_{23}b_3x,$$

$$\gamma_2 = xb_2yb_4x, \quad \delta_2 = xb_1c_{13}b_3x,$$

$$\gamma_3 = xb_3yb_4x, \quad \delta_3 = xb_1c_{12}b_2x.$$

Add up the number of times γ_i intersects δ_i for $i = 1, 2, 3$. The intersections between b_i and b_j for $1 \le i < j \le 4$ are all counted twice. By our assumption on G, all other pairs of edges intersect an even number of times. Furthermore, γ_i and δ_i intersect at x too. Note that among these intersections all but the ones at x are proper. At x we have a proper intersection in exactly one of the three pairs. Thus, one of the pairs has an odd number of proper intersections. This contradiction proves that $F_4 \not\subseteq G$, and hence Theorem 1.3.

6 Drawing C_4-free graphs – Proof of Theorem 1.4

Let G be a C_4-free bipartite graph with vertex set $V(G) = A \cup B$, where $A = \{a_1, a_2, \ldots, a_n\}$ and $B = \{b_1, b_2, \ldots, b_n\}$. The edge set of G is denoted by $E(G)$.

We now construct a drawing of G. Pick $2n$ points, $a_1, \ldots, a_n, b_1, \ldots b_n$, on the x-axis, from left to right in this order. These points will be identified with the vertices of G. For every edge $a_ib_j \in E(G)$, draw an x-monotone arc e_{ij} connecting a_i to b_j, according to the following rules:

(i) for any $k > i$, the arc e_{ij} passes above a_k if and only if $a_kb_j \notin E(G)$;

(ii) for any $l < j$, the arc e_{ij} passes above b_l if and only if $a_ib_l \in E(G)$;

(iii) no two distinct arcs "touch" each other (internal crossings are proper).

Notice that, unless two arcs share an endpoint, the *parity* of their number of intersections is determined by these rules.

Take two non-adjacent edges $a_ib_j, a_kb_l \in E(G)$ that belong to a path of length 3. We have to distinguish four different cases:

1. $i < k$, $j < l$, and $a_kb_j \in E(G)$;
2. $i < k$, $j < l$, and $a_ib_l \in E(G)$;
3. $i < k$, $l < j$, and $a_ib_l \in E(G)$;
4. $i < k$, $l < j$, and $a_kb_j \in E(G)$.

Consider the first case. By drawing rule (i), the arc e_{ij} passes below a_k. By rule (ii), e_{kl} passes above b_j. In view of rule (iii), this implies that e_{ij} and e_{kl} cross an even number of times, as required. The second case can be treated similarly and is left to the reader.

In the third case, applying rule (i), we obtain that a_k lies above e_{ij}. It is sufficient to show that the same is true for b_l. At this point, we use that G is C_4-free: since a_ib_j, b_ja_k, $a_kb_l \in E(G)$, we have $a_ib_l \notin E(G)$. By rule (ii), this implies that b_l is above e_{ij}, as required. The last case follows in the same way, by symmetry.

So far we have checked that in our drawing any two non-adjacent edges cross an even number of times. It is not hard to extend the same property to *all* pairs of edges, even if they share endpoints. To this end, we slightly modify the arcs e_{ij} in some very small neighborhoods of their endpoints. Clearly, this will not effect the crossing patterns of non-adjacent pairs.

Fix a vertex a_i. Redraw the arcs e_{ij} incident to a_i so that the counter-clockwise order of their initial pieces in a small neighborhood of a_i will be the same as the order of x-coordinates of their right endpoints. Consider now two arcs, e_{ij}, e_{il}, $(l < j)$, incident to a_i. By rule (ii), b_l lies below e_{ij}. On the other hand, after performing the local change described above, the initial piece of e_{il} will also lie below e_{ij}. This guarantees that e_{ij} and e_{il} cross an even number of times. Repeating this procedure for each vertex a_i, and its symmetric version for each b_j, we obtain a drawing which meets the requirements of Theorem 1.4.

7 Related problems

A. Theorems 1.1 and 3.1 easily imply

Corollary 7.1. *For any tree T other than a star, there exists a constant $c(T)$ such that every geometric (or x-monotone topological) graph G with n vertices and more than $c(T)n \log n$ edges contains a self-intersecting copy of T. That is, we have*

$$\mathrm{ex}_{\mathrm{cr}}(n, T) \leq c(T) n \log n,$$

and, apart from the value of the constant, this bound cannot be improved.

Indeed, deleting one-by-one every vertex of G whose degree is smaller than $|V(T)|$, we end up with a graph G' having at most n vertices and at least $(c(T)\log n - |V(T)|)n$ edges. If $c(T)$ is sufficiently large, then G' has a self-intersecting path of length 3. Using the fact that the degree of every vertex in G' is at least $|V(T)|$, this path can be extended to a copy of T in G' (and hence in G).

B. For any graph G, there may be a big difference between $\mathrm{ex}_{\mathrm{cr}}(n, H)$ and the number of edges guaranteeing the existence of a self-intersecting copy of H with a *specified* crossing pattern. For example, by Corollary 7.1, we have $\mathrm{ex}_{\mathrm{cr}}(n, P_4) = O(n \log n)$, but the best upper bound we have been able establish for the number of edges of a geometric graph with n vertices, containing no path of length 4 whose first and last edges cross each other, is $O(n^{3/2})$ (this follows from Theorem 5.1). For *convex* geometric graphs (i.e., when the vertices are in convex position), this bound can be improved to $O(n \log n)$, which is asymptotically tight.

C. Any drawing of $K_{3,3}$, a complete bipartite graph with 3 vertices in each of its classes, has two non-adjacent edges that cross each other. Clearly, any two edges belong to a cycle of length 4, so

$$\mathrm{ex}_{\mathrm{cr}}(n, C_4) \leq \mathrm{ex}(n, K_{3,3}) = O(n^{5/3}).$$

This bound has been recently improved to $O(n^{8/5})$ by Pinchasi and Radoičić [PR02]. It seems likely that the best possible bound is close to $n^{3/2}$.

It also follows from Theorem 5.1) that $\mathrm{ex}_{\mathrm{cr}}(n, C_6) = O(n^{3/2})$, and it generalizes to topological graphs. On the other hand, we have $\mathrm{ex}_{\mathrm{cr}}(n, C_4) \geq \mathrm{ex}(n, C_6) \geq cn^{4/3}$, for a suitable constant $c > 0$ (see [BS74]).

Theorem 7.2. *Let G be a C_4-free geometric (or x-monotone topological) graph on n vertices, with no self-intersecting cycle of length 6. Then G has at most $O(n^{4/3}\log^{2/3} n)$ edges.*

Proof. Let G have n vertices and $|E(G)| = m > c'n^{4/3}\log^{2/3} n$ edges. For a fixed number $0 < p < 1$, which will be specified later, color randomly and independently with probability p each vertex of G *red*. Let G' be the subgraph of G induced by the red vertices.

Let $i(G')$ denote the number of self-intersecting paths of length 3 in G'. Deleting one edge from each such path, we obtain a graph with no self-intersecting path of length 3. Thus, in view of Theorem 1.1, we have

$$|E(G')| - i(G') < c|V(G')| \log |V(G')|,$$

for some positive c. Taking expected values, this yields

$$p^2|E(G)| - p^4 i(G) < cpn \log(pn).$$

Setting $p = \frac{3cn \log n}{2|E(G)|}$, we obtain $i(G) > \frac{4|E(G)|^3}{27c^2n^2 \log^2 n}$. If c' is large enough, then $i(G) > \binom{n}{2}$, and there must exist two self-intersecting paths of length 3 connecting the same pair of vertices. Putting these two paths together, we get a C_6 which intersects itself at least twice. ■

References

[AAPPS97] P. K. Agarwal, B. Aronov, J. Pach, R. Pollack, and M. Sharir, Quasi-planar graphs have a linear number of edges, *Combinatorica* **17** (1997), 1–9.

[AH66] S. Avital and H. Hanani, Graphs (in Hebrew), *Gilyonot Lematematika* **3** (1966), 2–8.

[B78] B. Bollobás, *Extremal Graph Theory,* Academic Press, New York, 1978.

[BS74] A. Bondy and M. Simonovits, Cycles of even length in graphs, *J. Combinatorial Theory, Ser. B* **16** (1974), 97–105.

[F91] Z. Füredi, On a Turán type problem of Erdős, *Combinatorica 11* **1** (1991), 75–79.

[FH92] Z. Füredi and P. Hajnal, Davenport-Schinzel theory of matrices, *Discrete Mathematics* **103** (1992), 233–251.

[H34] H. Hanani (C. Chojnacki), Über wesentlich unplättbare Kurven in dreidimensionalen Raume, *Fundamenta Mathematicae* **23** (1934), 135–142.

[K79] Y. Kupitz, *Extremal Problems in Combinatorial Geometry, Aarhus University Lecture Notes Series* **53**, Aarhus University, Denmark, 1979.

[P99] J. Pach, Geometric graph theory, in: *Surveys in Combinatorics, 1999 (J. D. Lamb and D. A. Preece, eds.), London Mathematical Society Lecture Notes* **267**, Cambridge University Press, Cambridge, 1999, 167–200.

[PT97] J. Pach and G. Tóth, Graphs drawn with few crossings per edge, *Combinatorica* **17** (1997), 427–439.

[PR02] R. Pinchasi and R. Radoičić, On the number of edges in geometric graphs with no self-intersecting cycle of length 4, to appear.

[T70] W. T. Tutte, Toward a theory of crossing numbers, *J. Combinatorial Theory* **8** (1970), 45–53.

[V98] P. Valtr, On geometric graphs with no k pairwise parallel edges, *Discrete and Computational Geometry* **19** (1998), 461–469.