Almost All String Graphs are Intersection Graphs of Plane Convex Sets

János Pach[1]
Ecole Polytechnique Fédérale de Lausanne and Rényi Institute, Hungarian Academy of Sciences
Lausanne, Switzerland and Budapest, Hungary
pach@cims.nyu.edu

Bruce Reed
School of Computer Science, McGill University, Laboratoire I3S CNRS, and Professor Visitante Especial, IMPA
Montreal, Canada and Rio de Janeiro, Brazil
breed@cs.mcgill.ca

Yelena Yuditsky
School of Computer Science, McGill University
Montreal, Canada
yuditskyL@gmail.com

Abstract

A *string graph* is the intersection graph of a family of continuous arcs in the plane. The intersection graph of a family of plane convex sets is a string graph, but not all string graphs can be obtained in this way. We prove the following structure theorem conjectured by Janson and Uzzell: The vertex set of *almost all* string graphs on n vertices can be partitioned into *five* cliques such that some pair of them is not connected by any edge ($n \to \infty$). We also show that every graph with the above property is an intersection graph of plane convex sets. As a corollary, we obtain that *almost all* string graphs on n vertices are intersection graphs of plane convex sets.

2012 ACM Subject Classification Mathematics of computing → Graph theory, Theory of computation → Randomness, geometry and discrete structures

Keywords and phrases String graph, intersection graph, plane convex set

Digital Object Identifier 10.4230/LIPIcs.SoCG.2018.68

Related Version A full version of the paper is available at [PRY18], `http://arxiv.org/abs/1803.06710`.

Acknowledgements This research was carried out while all three authors were visiting IMPA in Rio de Janeiro. They would like to thank the institute for its generous support.

1 Overview

The *intersection graph* of a collection C of sets is a graphs whose vertex set is C and in which two sets in C are connected by an edge if and only if they have nonempty intersection. A *curve* is a subset of the plane which is homeomorphic to the interval $[0, 1]$. The intersection graph of a finite collection of curves ("strings") is called a *string graph*.

Ever since Benzer [Be59] introduced the notion in 1959, to explore the topology of genetic structures, string graphs have been intensively studied both for practical applications and

[1] Supported by Swiss National Science Foundation Grants 200021-165977 and 200020-162884.

theoretical interest. In 1966, studying electrical networks realizable by printed circuits, Sinden [Si66] considered the same constructs at Bell Labs. He proved that not every graph is a string graph, and raised the question whether the recognition of string graphs is decidable. The affirmative answer was given by Schaefer and Štefankovič [ScSt04] 38 years later. The difficulty of the problem is illustrated by an elegant construction of Kratochvíl and Matoušek [KrMa91], according to which there exists a string graph on n vertices such that no matter how we realize it by curves, there are two curves that intersect at least 2^{cn} times, for some $c > 0$. On the other hand, it was proved in [ScSt04] that every string graph on n vertices and m edges can be realized by polygonal curves, any pair of which intersect at most $2^{c'm}$ times, for some other constant c'. The problem of recognizing string graphs is NP-complete [Kr91, ScSeSt03].

In spite of the fact that there is a wealth of results for various special classes of string graphs, understanding the structure of general string graphs has remained an elusive task. The aim of this paper is to show that *almost all* string graphs have a very simple structure. That is, the proportion of string graphs that possess this structure tends to 1 as n tends to infinity.

Given any graph property P and any $n \in \mathbb{N}$, we denote by P_n the set of all graphs with property P on the (labeled) vertex set $V_n = \{1, \dots, n\}$. In particular, $\textsc{String}_n$ is the collection of all string graphs with the vertex set V_n. We say that an n-element set is partitioned into parts of *almost equal size* if the sizes of any two parts differ by at most $n^{1-\epsilon}$ for some $\epsilon > 0$, provided that n is sufficiently large.

▶ **Theorem 1.** *As $n \to \infty$, the vertex set of almost every string graph $G \in \textsc{String}_n$ can be partitioned into 4 parts of almost equal size such that 3 of them induce a clique in G and the 4th one splits into two cliques with no edge running between them.*

▶ **Theorem 2.** *Every graph G whose vertex set can be partitioned into 4 parts such that 3 of them induce a clique in G and the 4th one splits into two cliques with no edge running between them, is a string graph.*

Theorem 1 settles a conjecture of Janson and Uzzell from [JaU17], where a related weaker result was proved in terms of graphons.

We also prove that a typical string graph can be realized using relatively simple strings.

Let $\textsc{Conv}_n$ denote the set of all intersection graphs of families of n labeled convex sets $\{C_1, \dots, C_n\}$ in the plane. For every pair $\{C_i, C_j\}$, select a point in $C_i \cap C_j$, provided that such a point exists. Replace each convex set C_i by the polygonal curve obtained by connecting all points selected from C_i by segments, in the order of increasing x-coordinate. Observe that any two such curves belonging to different C_is intersect at most $2n$ times. The intersection graph of these curves (strings) is the same as the intersection graph of the original convex sets, showing that $\textsc{Conv}_n \subseteq \textsc{String}_n$. Taking into account the construction of Kratochvíl and Matoušek [KrMa91] mentioned above, it easily follows that the sets $\textsc{Conv}_n$ and $\textsc{String}_n$ are not the same, provided that n is sufficiently large.

▶ **Theorem 3.** *There exist string graphs that cannot be obtained as intersection graphs of convex sets in the plane.*

We call a graph G *canonical* if its vertex set can be partitioned into 4 parts such that 3 of them induce a clique in G and the 4th one splits into two cliques with no edge running between them. The set of canonical graphs on n vertices is denoted by $\textsc{Canon}_n$. Theorem 2 states $\textsc{Canon}_n \subset \textsc{String}_n$. In fact, this is an immediate corollary of $\textsc{Conv}_n \subset \textsc{String}_n$ and the relation $\textsc{Canon}_n \subset \textsc{Conv}_n$, formulated as

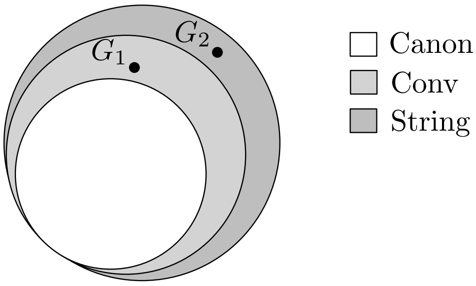

Figure 1 The graph G_1 is the any planar graph with more than 20 vertices. The graph G_2 is the graph from the construction of Kratochvíl and Matoušek [KrMa91].

▶ **Theorem 4.** *The vertices of every canonical graph G can be represented by convex sets in the plane such that their intersection graph is G.*

The converse is not true. Every planar graph can be represented as the intersection graph of convex sets in the plane (Koebe [Ko36]). Since no planar graph contains a clique of size exceeding four, for $n > 20$ no planar graph with n vertices is canonical.

Combining Theorems 1 and 4, we obtain the following.

▶ **Corollary 5.** *Almost all string graphs on n labeled vertices are intersection graphs of convex sets in the plane.*

See Figure 1 for a sketch of the containment relation of the families of graphs discussed above.

The rest of this paper is organized as follows. In Section 2, we recall the necessary tools from extremal graph theory, and adapt a partitioning technique of Alon, Balogh, Bollobás, and Morris [AlBBM11] to analyze string graphs; see Theorem 8. In Section 3, we collect some simple facts about string graphs and intersection graphs of plane convex sets, and combine them to prove Theorem 4. In Section 4, we strengthen Theorem 8 in two different ways and, hence, prove Theorem 1 modulo a small number of exceptional vertices. We wrap up the proof of Theorem 1 in Section 5.

2 The structure of typical graphs in an hereditary family

A *graph property* P is called *hereditary* if every induced subgraph of a graph G with property P has property P, too. With no danger of confusion, we use the same notation P to denote a (hereditary) graph property and the family of all graphs that satisfy this property. Clearly, the properties that a graph G is a string graph ($G \in \textsc{String}$) or that G is an intersection graph of plane convex sets ($G \in \textsc{Conv}$) are hereditary. The same is true for the properties that G contains no subgraph, resp., no induced subgraph isomorphic to a fixed graph H.

It is a classic topic in extremal graph theory to investigate the typical structure of graphs in a specific hereditary family. This involves proving that almost all graphs in the family have a certain structural decomposition. This research is inextricably linked to the study of the growth rate of the function $|P_n|$, also known as the *speed* of P, in two ways. Firstly, structural decompositions may give us bounds on the growth rate. Secondly, lower bounds on the growth rate help us to prove that the size of the exceptional family of graphs which fail to have a specific structural decomposition is negligible. In particular, we will both use a preliminary bound on the speed in proving our structural result about string graphs, and apply our theorem to improve the best known current bounds on the speed of the string graphs.

In a pioneering paper, Erdős, Kleitman, and Rothschild [ErKR76] approximately determined for every t the speed of the property that the graph contains no clique of size t. Erdős, Frankl, and Rödl [ErFR86] generalized this result as follows. Let H be a fixed graph with chromatic number $\chi(H)$. Then every graph of n vertices that does not contain H as a (not necessarily induced) subgraph can be made $(\chi(H)-1)$-partite by the deletion of $o(n^2)$ edges. This implies that the speed of the property that the graph contains no subgraph isomorphic to H is

$$2^{\left(1-\frac{1}{\chi(H)-1}+o(1)\right)\binom{n}{2}}. \tag{1}$$

Prömel and Steger [PrS92a, PrS92b, PrS93] established an analogous theorem for graphs containing no *induced subgraph* isomorphic to H. Throughout this paper, these graphs will be called *H-free*. To state their result, Prömel and Steger introduced the following key notion.

▶ **Definition 6.** A graph G is *(r,s)-colorable* for some $0 \le s \le r$ if there is a r-coloring of the vertex set $V(G)$, in which the first s color classes are cliques and the remaining $r-s$ color classes are independent sets. The *coloring number* $\chi_c(\mathrm{P})$ of a hereditary graph property P is the largest integer r for which there is an s such that all (r,s)-colorable graphs have property P. Consequently, for any $0 \le s \le \chi_c(\mathrm{P})+1$, there exists a $(\chi_c(\mathrm{P})+1, s)$-colorable graph that does not have property P.

The work of Prömel and Steger was completed by Alekseev [Al93] and by Bollobás and Thomason [BoT95, BoT97], who proved that the speed of any hereditary graph property P satisfies

$$|\mathrm{P}_n| = 2^{\left(1-\frac{1}{\chi_c(\mathrm{P})}+o(1)\right)\binom{n}{2}}. \tag{2}$$

The lower bound follows from the observation that for $\chi_c(\mathrm{P}) = r$, there exists $s \le r$ such that all (r,s)-colorable graphs have property P. In particular, P_n contains all graphs whose vertex sets can be partitioned into s cliques and $r-s$ independent sets, and the number of such graphs is equal to the right-hand side of (2).

As for string graphs, Pach and Tóth [PaT06] proved that

$$\chi_c(\textsc{String}) = 4. \tag{3}$$

Hence, (2) immediately implies

$$|\textsc{String}_n| = 2^{\left(\frac{3}{4}+o(1)\right)\binom{n}{2}}. \tag{4}$$

If we want to tighten the above estimates, another idea of Prömel and Steger [PrS91] is instructive. They noticed that the vertex set of almost every C_4-free graph can be partitioned into a clique and an independent set, and no matter how we choose the edges between these two parts, we always obtain a C_4-free graph. Therefore, the speed of C_4-freeness is at most $(1+o(1))2^n 2^{\frac{1}{2}\binom{n}{2}}$, which is much better than the general bound $2^{\left(\frac{1}{2}+o(1)\right)\binom{n}{2}}$ that follows from (2). Almost all C_5-free graphs permit similar *"certifying partitions"*. It is an interesting open problem to decide which hereditary families permit such partitions and what can be said about the inner structure of the subgraphs induced by the parts. This line of research was continued by Balogh, Bollobás, and Simonovits [BaBS04, BaBS09, BaBS11]. The strongest result in this direction was proved by Alon, Balogh, Bollobás, and Morris [AlBBM11], who proved that for almost every graph with a hereditary property P, one can delete a small fraction of the vertices in such a way that the rest can be partitioned into $\chi_c(\mathrm{P})$ parts with a very simple inner structure. This allowed them to replace the bound (2) by a better one:

$$|\mathrm{P}_n| = 2^{\left(1-\frac{1}{\chi_c(\mathrm{P})}\right)\binom{n}{2}+O(n^{2-\epsilon})}.$$

This will be the starting point of our analysis of string graphs. As we shall see, in the case of string graphs, our results allow us to replace the $2^{O(n^{2-\epsilon})}$ in this bound by $2^{\frac{9n}{4}+o(n)}$. See [BB11, KKOT15, RY17, ReSc17], for related results.

We need some notation. Following Alon *et al.*, for any integer $k > 0$, define $U(k)$ as a bipartite graph with vertex classes $\{1, \ldots, k\}$ and $\{I : I \subset \{1, ..., k\}\}$, where a vertex i in the first class is connected to a vertex I in the second if and only if $i \in I$. We think of $U(k)$ as a "universal" bipartite graph on $k + 2^k$ vertices, because for every subset of the first class there is a vertex in the second class whose neighborhood is precisely this subset.

As usual, the *neighborhood* of a vertex v of a graph G is denoted by $N_G(v)$ or, if there is no danger of confusion, simply by $N(v)$. For any disjoint subsets $A, B \subset V(G)$, let $G[A]$ and $G[A, B]$ denote the subgraph of G induced by A and the *bipartite* subgraph of G consisting of all edges of G running between A and B, respectively. The *symmetric difference* of two sets, X and Y, is denoted by $X \triangle Y$.

▶ **Definition 7.** Let k be a positive integer. A graph G is said to *contain* $U(k)$ if there are two disjoint subsets $A, B \subset V(G)$ such that the bipartite subgraph $G[A, B] \subseteq G$ induced by them is isomorphic to $U(k)$. Otherwise, with a slight abuse of terminology, we say that G is $U(k)$-free.

By slightly modifying the proof of the main result (Theorem 1) in [AlBBM11] and adapting it to string graphs, we obtain

▶ **Theorem 8.** *For any sufficiently large positive integer k and for any $\delta > 0$ which is sufficiently small in terms of k, there exist $\epsilon > 0$ and a positive integer b with the following properties.*

The vertex set V_n ($|V_n| = n$) of almost every string graph G can be partitioned into eight sets, $S_1, ...S_4, A_1, ..., A_4$, and a set B of at most b vertices such that

(a) *$G[S_i]$ is $U(k)$-free for every i ($1 \le i \le 4$);*
(b) *$|A_1 \cup A_2 ... \cup A_4| \le n^{1-\epsilon}$; and*
(c) *for every i ($1 \le i \le 4$) and $v \in S_i \cup A_i$ there is $a \in B$ such that*

$$|(N(v) \triangle N(a)) \cap (S_i \cup A_i)| \le \delta n.$$

In other words, for the right choice of parameters, almost all string graphs have a partition into 4 parts satisfying the following conditions. There is a set of sub-linear size in the number of vertices such that deleting its elements, the subgraphs induced by the parts are U(k)-free. Moreover, there is another set B of at most constantly many vertices such that the neighborhood of every vertex with respect to the part it belongs to is similar to the neighbourhood of some vertex in B. In the full version of the paper [PRY18], we sketch the proof of this result, indicating the places where we slightly deviate from the original argument in [AlBBM11].

3 String graphs vs. intersection graphs of convex sets – proof of Theorem 4

Instead of proving Theorem 4, we establish a somewhat more general result.

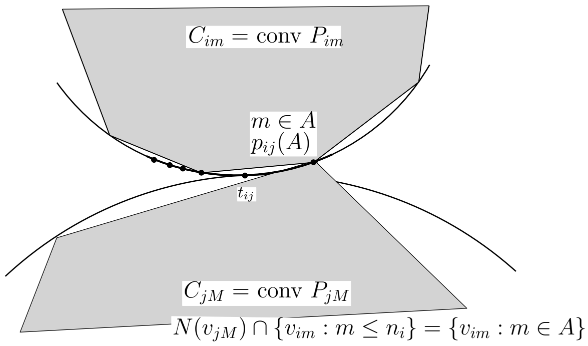

Figure 2 The point $p_{ij}(A)$ is included in P_{jM}.

▶ **Theorem 9.** *Given a planar graph H with labeled vertices $\{1, \ldots, k\}$ and positive integers $n_1, \ldots, n_k$, let $\mathrm{H}(n_1, \ldots, n_k)$ denote the class of all graphs with $n_1 + \ldots + n_k$ vertices that can be obtained from H by replacing every vertex $i \in V(H)$ with a clique of size n_i, and adding any number of further edges between pairs of cliques that correspond to pairs of vertices $i \neq j$ with $ij \in E(G)$.*

Then every element of $\mathrm{H}(n_1, \ldots, n_k)$ is the intersection graph of a family of plane convex sets.

Proof. Fix any graph $G \in \mathrm{H}(n_1, \ldots, n_k)$. The vertices of H can be represented by closed disks $D_1, \ldots, D_k$ with disjoint interiors such that D_i and D_j are tangent to each other for some $i < j$ if and only if $ij \in E(H)$ (Koebe, [Ko36]). In this case, let $t_{ij} = t_{ji}$ denote the point at which D_i and D_j touch each other. For any i $(1 \leq i \leq k)$, let o_i be the center of D_i. Assume without loss of generality that the radius of every disk D_i is at least 1.

G has $n_1 + \ldots + n_k$ vertices denoted by v_{im}, where $1 \leq i \leq k$ and $1 \leq m \leq n_i$. In what follows, we assign to each vertex $v_{im} \in V(G)$ a finite set of points P_{im}, and define C_{im} to be the convex hull of P_{im}. For every $i, 1 \leq i \leq k$, we include o_i in all sets P_{im} with $1 \leq m \leq n_i$, to make sure that for each i, all sets $C_{im}, 1 \leq m \leq n_i$ have a point in common, therefore, the vertices that correspond to these sets induce a clique.

Let $\varepsilon < 1$ be the *minimum* of all angles $\measuredangle t_{ij} o_i t_{il} > 0$ at which the arc between two consecutive touching points t_{ij} and t_{il} on the boundary of the same disc D_i can be seen from its center, over all $i, 1 \leq i \leq k$ and over all j and l. Fix a small $\delta > 0$ satisfying $\delta < \varepsilon^2/100$.

For every $i < j$ with $ij \in E(H)$, let γ_{ij} be a circular arc of length δ on the boundary of D_i, centered at the point $t_{ij} \in D_i \cap D_j$. We select 2^{n_i} distinct points $p_{ij}(A) \in \gamma_{ij}$, each representing a different subset $A \subseteq \{1, \ldots, n_i\}$. A point $p_{ij}(A)$ will belong to the set P_{im} if and only if $m \in A$. (Warning: Note that the roles of i and j are not interchangeable!)

If for some $i < j$ with $ij \in E(H)$, the intersection of the neighborhood of a vertex $v_{jM} \in V(G)$ for any $1 \leq M \leq n_j$ with the set $\{v_{im} : 1 \leq m \leq n_i\}$ is equal to $\{v_{im} : m \in A\}$, then we include the point $p_{ij}(A)$ in the set P_{jM} assigned to v_{jM}, see Figure 2 for a sketch. Hence, for every $m \leq n_i$ and $M \leq n_j$, we have

$$v_{im} v_{jM} \in E(G) \iff P_{im} \cap P_{jM} \neq \emptyset.$$

In other words, the intersection graph of the sets assigned to the vertices of G is isomorphic to G.

It remains to verify that

$$v_{im} v_{jM} \in E(G) \iff C_{im} \cap C_{jM} \neq \emptyset.$$

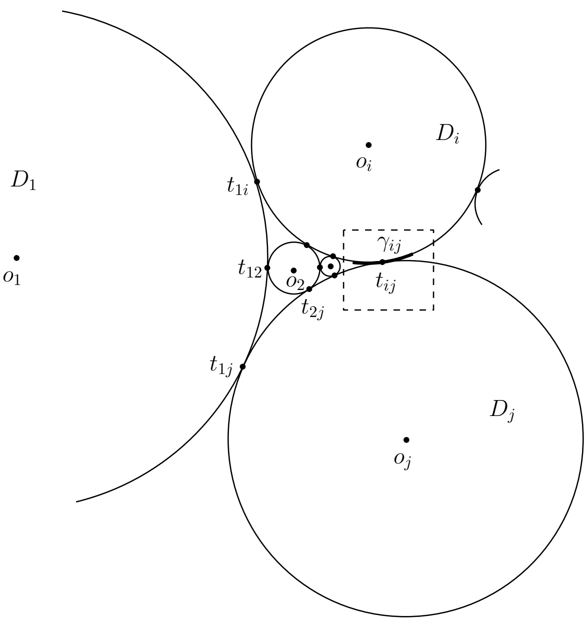

Figure 3 Tangent disks D_i and D_j touching at t_{ij}.

Suppose that the intersection graph of the set of convex polygonal regions

$$\{C_{im} : 1 \leq i \leq k \text{ and } 1 \leq m \leq n_i\}$$

differs from the intersection graph of

$$\{P_{im} : 1 \leq i \leq k \text{ and } 1 \leq m \leq n_i\}.$$

Assume first, for contradiction, that there exist i, m, j, M with $i < j$ such that D_i and D_j are tangent to each other and C_{jM} contains a point $p_{ij}(B)$ for which

$$B \neq N_{jM} \cap \{v_{im} : 1 \leq m \leq n_i\}. \tag{5}$$

Consider the unique point $p = p_{ij}(A) \in \gamma_{ij}$ that belongs to P_{jM}, that is, we have

$$A = N_{jM} \cap \{v_{im} : 1 \leq m \leq n_i\}.$$

Draw a tangent line ℓ to the arc γ_{ij} at point p. See Figure 3. The polygon C_{jM} has two sides meeting at p; denote the infinite rays emanating from p and containing these sides by r_1 and r_2. These rays either pass through o_j or intersect the boundary of D_j in a small neighborhood of the point of tangency of D_j with some other disk $D_{j'}$. Since δ was chosen to be much smaller than ε, we conclude that r_1 and r_2 lie entirely on the same side of ℓ where o_j, the center of D_j, is. On the other hand, all other points of γ_{ij}, including the point $p_{ij}(B)$ satisfying (5) lie on the opposite side of ℓ, which is a contradiction.

Essentially the same argument and a little trigonometric computation show that for every j and M, the set $C_{jM} \setminus D_j$ is covered by the union of some small neighborhoods (of radius $< \varepsilon/10$) of the touching points t_{ij} between D_j and the other disks D_i. This, together with the assumption that the radius of every disk D_i is at least 1 (and, hence, is much larger than ε and δ) implies that C_{jM} cannot intersect any polygon C_{im} with $i \neq j$, for which D_i and D_j are not tangent to each other. ◀

Applying Theorem 9 to the graph obtained from K_5 by deleting one of its edges, Theorem 4 follows.

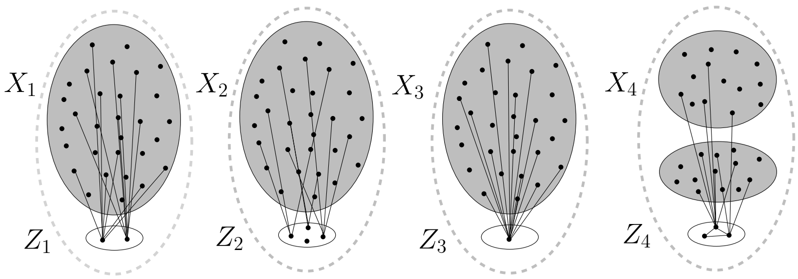

Figure 4 A sketch of a typical string graph as in Theorem 10. The edges between the parts are not drawn. The sets shaded grey are cliques.

4 Strengthening Theorem 8

In this section, we strengthen Theorem 8 in two different ways. To avoid confusion, in the formulation of our new theorem, we use X_i in place of S_i and Z_i in place of A_i. We will see that we can insist that the four parts of the partition have approximately the same size. Secondly, we can guarantee that X_1, X_2, and X_3 are cliques and X_4 induces the disjoint union of two cliques. More precisely, setting $Z = Z_1 \cup Z_2 ... \cup Z_4$, we prove the following result, which is similar in flavour to a result in [ReSc17].

▶ **Theorem 10.** *For every sufficiently small δ, there are $\gamma > 0, b > 4 + \frac{2}{\delta}$ with the following property. For almost every string graph G on V_n, there is a partition of V_n into $X_1, ..., X_4, Z_1, ..., Z_4$ such that for some set B of at most b vertices the following conditions are satisfied:*

(I) *$G[X_1]$, $G[X_2]$, and $G[X_3]$ are cliques and $G[X_4]$ induces the disjoint union of two cliques.*

(II) $|Z_1 \cup Z_2 \cup Z_3 \cup Z_4| \leq n^{1-\gamma}$,

(III) *for every i $(1 \leq i \leq 4)$ and every $v \in X_i \cup Z_i$, there exists $a \in B$ such that*

$$|(N(v) \triangle N(a)) \cap (X_i \cup Z_i)| \leq \delta n,$$

(IV) *for every i $(1 \leq i \leq 4)$, we have $\left||Z_i \cup X_i| - \frac{n}{4}\right| \leq n^{1-\gamma}$.*

See Figure 4 for an illustration of Theorem 10.

For the proof of Theorem 10 we need the following statement which is a slight generalization of Lemma 3.2 in [PaT06], and it can be established in precisely the same way, details are given in the full version of the paper [PRY18].

▶ **Lemma 11.** *Let H be a graph on the vertex set $\{v_1, \ldots, v_5\} \cup \{v_{ij} : 1 \leq i \neq j \leq 5\}$, where $v_{ij} = v_{ji}$ and every v_{ij} is connected by an edge to v_i and v_j. The graph H may have some further edges connecting pairs of vertices (v_{ij}, v_{ik}) with $j \neq k$. Then H is not a string graph.*

▶ **Corollary 12.** *For each of the following types of partition, there exists a non-string graph whose vertex set can be partitioned in the specified way:*

(a) 2 *stable (that is, independent) sets each of size at most 10;*
(b) 4 *cliques each of size at most five and a vertex;*
(c) 3 *cliques each of size at most five and a stable set of size* 3*;*
(d) 3 *cliques each of size at most five and a path with three vertices;*
(e) 2 *cliques both of size at most five and* 2 *graphs that can be obtained as the disjoint union of a point and a clique of size at most* 3.

See Figure 5 for an illustration of Corollary 12.

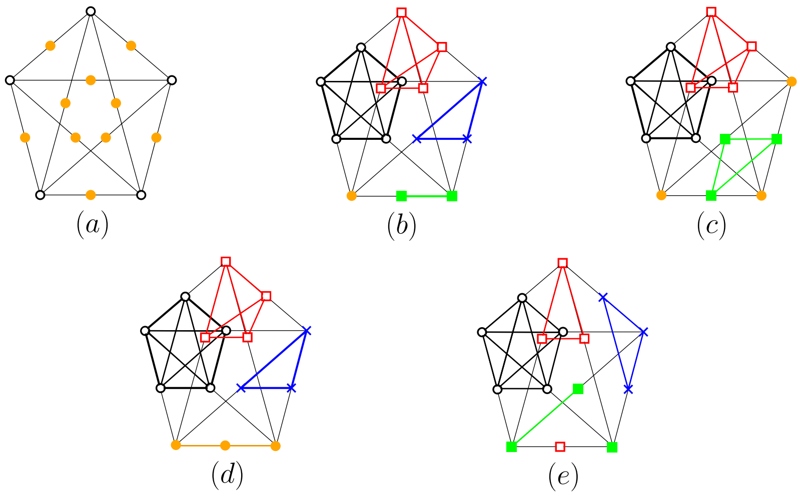

Figure 5 Possible partitions of a non-string graph.

Proof of Theorem 10. We choose k sufficiently large and then $\delta < \frac{1}{40}$ sufficiently small in terms of k. We choose $\epsilon, b > 0$ such that Theorem 8 holds for this choice of k and δ and so that ϵ is less than the ρ of Lemma 14 for this choice of k. We set $\gamma = \frac{\epsilon}{10}$ and consider n large enough to satisfy certain implicit inequalities below. We know that the subset $\mathcal{S}(k, \delta)_n$ of $\textsc{String}_n$, consisting of those graphs for which there is a set B of at most b vertices and a partition into S_i and A_i satisfying (a),(b), and (c) set out in Theorem 8, contains almost every string graph. We call such a partition, *certifying*. We need to show that almost every graph in $\mathcal{S}(k, \delta)_n$ has a certifying partition for which we can repartition $S_i \cup A_i$ into $X_i \cup Z_i$ so that (I),(II), and (IV) all hold (that (III) holds, is simply Theorem 8 (c) and $S_i \cup A_i = X_i \cup Z_i$).

We prove this fact via a sequence of lemmas. In doing so, for a specific partition, we let $m = m(A_1 \cup S_1, A_2 \cup S_2, A_3 \cup S_3, A_4 \cup S_4)$ be the number of pairs of vertices not lying together in some $A_i \cup S_i$. The first lemma gives us a lower bound on $|\mathcal{S}(k, \delta)_n|$, obtained by simply counting the number of graphs which permits a partition into four cliques all of size within one of $\frac{n}{4}$. Its proof is given in the full version of the paper [PRY18].

▶ **Lemma 13.** $|\mathcal{S}(k, \delta)_n| \geq 2^{\frac{3\binom{n}{2}}{4}}$.

The second gives us an upper bound on the number of choices for $G[S_i]$ for graphs G in $\mathcal{S}(k, \delta)_n$ for which $S_1, S_2, S_3, S_4, A_1, A_2, A_3, A_4$ is a certifying partition. It is Corollary 8 in [AlBBM11].

▶ **Lemma 14.** *For every k, there is a positive ρ such that for every sufficiently large l, the number of $U(k)$-free graphs with l vertices is less than $2^{l^{2-\rho}}$.*

Next we prove:

▶ **Lemma 15.** *The number of graphs in $\mathcal{S}(k, \delta)_n$ which have a certifying partition such that for some i, $||A_i \cup S_i| - \frac{n}{4}| > n^{1-\gamma}$ is $o(|\mathcal{S}(k, \delta)_n|)$.*

Proof. The number of choices for a partition of V_n into $S_1, S_2, S_3, S_4, A_1, A_2, A_3, A_4$ is at most 8^n. If this partition demonstrates that S_i is $U(k)$-free and n is large, Lemma 14 tells us that there are only $2^{n^{2-\epsilon}}$ choices for $G[S_i]$. The number of choices for the edges out of each vertex of A_i is 2^{n-1}. So, since $|A_i|$ is at most $n^{1-\epsilon}$, we know there are at most $2^{n^{2-\epsilon}}$ choices for the edges out of A_i. It follows that there are at most $2^{11(n^{2-\epsilon})}$ choices for our partition and the graphs $G[S_1 \cup A_1], ..., G[S_4 \cup A_4]$ over all G in $\mathcal{S}(k,\delta)_n$ which can be certified using this partition. Furthermore,the number of graphs in $\mathcal{S}(k,\delta)_n$ permitting such a certifying choice is at most 2^m. Since, $|\mathcal{S}(k,\delta)_n| \geq 2^{\frac{3\binom{n}{2}}{4}}$, it follows that almost every graph G in $\mathcal{S}(k,\delta)_n$ has no certifying partition for which $m < \frac{3\binom{n}{2}}{4} - 12(n^{2-\epsilon})$. The desired result follows. ◀

Setting $l = l_n = \lceil n^{1-\frac{\epsilon}{7}} \rceil$, we have the following.

▶ **Lemma 16.** *The number of graphs in $\mathcal{S}(k,\delta)_n$ which have a certifying partition for which there are distinct i and j such that both S_i and S_j contain l disjoint independent sets of size 10 is $o(|\mathcal{S}(k,\delta)_n|)$.*

Proof. Consider a choice of certifying partition and induced subgraphs H_1, H_2, H_3, H_4 where $V(H_i) = A_i \cup S_i$. By Corollary 12(a), for any pair of independent sets of size 10, at least one of the 2^{100} choices of edges between the sets yields a bipartite non-string graph. Thus, the number of choices for edges between the partitions which extend our choice to yield a graph in $String_n$ is at most $2^m(1 - \frac{1}{2^{100}})^{l^2}$. Since $m < \frac{3\binom{n}{2}}{4}$ and $l^2 = \omega(n^{2-\frac{\epsilon}{2}})$, it follows that for almost every graph in $\mathcal{S}(k,\delta)_n$, almost every certifying partition does not contain two distinct such i and j. ◀

Ramsey theory tells us that if a graph J does not contain l disjoint stable sets of size 10, it contains $|V(J)| - 10(l-1) - 2^{15}$ disjoint cliques of size 5. Combining applications of this fact to three of the $G[S_i]$, Corollary 11(c), and an argument similar to that used in the proof of Lemma 16 allows us to prove the following lemma. Details can be found in the full version of the paper [PRY18].

▶ **Lemma 17.** *The number of graphs G in $\mathcal{S}(k,\delta)_n$ which have a certifying partition for which there is an $i = i(G)$ such that S_i does not contain l disjoint cliques of size 5 is $o(|\mathcal{S}(k,\delta)_n|)$*

With this lemma in hand, we can mimic the argument used in its proof to obtain the following two lemmas. In doing so, we apply Corollary 11 (c),(d), and (e).

▶ **Lemma 18.** *The number of graphs G in $\mathcal{S}(k,\delta)_n$ which have a certifying partition for which there is an $i = i(G)$ such that S_i contains l disjoint sets of size three each inducing a stable set or a path is $o(|\mathcal{S}(k,\delta)_n|)$.*

▶ **Lemma 19.** *The number of graphs G in $\mathcal{S}(k,\delta)_n$ which have a certifying partition for which there are two distinct i such that S_i contains l disjoint sets of size four each inducing the disjoint union of a vertex and a triangle is $o(|\mathcal{S}(k,\delta)_n|)$.*

Combining these lemmas, and possibly permuting indices, we see that almost every graph in $\mathcal{S}(k,\delta)_n$ has a certifying partition for which for every $i \leq 4$ we have $||Z_i \cup X_i| - \frac{n}{4}| \leq n^{1-\gamma}$, no S_i contains more than l sets inducing a path of length three or a stable set of size three, and for every $k \leq 3$, S_k does not contain l disjoint sets inducing the disjoint union of a vertex and a triangle. For each such graph, we consider such a partition. For all $i < 4$, we let Z_i be the union of A_i and a maximum family of disjoint sets in X_i each inducing a path of length 3, a stable set of size three, or the disjoint union of a triangle and a vertex. We let Z_4 be the union of A_4 and a maximum family of disjoint sets in X_4 each inducing a path of length three or a stable set of size three. We set $X_i = S_i - Z_i$. ◀

5 Completing the proof of Theorem 1

In this section, we prove our main result. By a *great* partition of G we mean a partition of its vertex set into X_1, X_2, X_3, X_4 such that for $i \leq 3$, X_i is a clique and X_4 is the disjoint union of two cliques. We call a graph *great* if it has a great partition and *mediocre* otherwise. Theorem 1 simply states that almost every string graph G on V_n is great.

Thus, we are trying to show that almost every string graph has a partition into sets $X_1, X_2, X_3, X_4, Z_1, Z_2, Z_3, Z_4$ satisfying Theorem 10 (I) with the sets Z_i empty. We choose δ so small that Theorem 10 holds and δ also satisfies certain inequalities implicitly given below. We apply Theorem 10 and obtain that for some positive γ and b, for almost every graph in String_n there is a partition of V_n into $X_1, ..., X_4, Z_1, .., Z_4$ satisfying (I), (II), (III), and (IV). Note that if we reduce γ the theorem remains true. We insist that γ is at most $\frac{1}{64000000}$. We call such partitions *good*. We need to show that the number of mediocre string graphs on V_n with a good partition is of smaller order than the number of great graphs on V_n.

The following result tells us that the number of great graphs on V_n is of the same order as the number of great partitions of graphs on V_n.

▶ **Claim 20.** *The ratio between the number of great partitions of graphs on* V_n *and the number of graphs which permit such partitions is* $6 + o(1)$.

So, it is sufficient to show that the number of mediocre string graphs with a good partition on V_n is of smaller order than the number of graphs with a great partition on V_n. In doing so, we consider each partition separately. For every partition $\mathcal{Y} = (Y_1, Y_2, Y_3, Y_4)$ of V_n we say that a good partition satisfying (I)-(IV) with $Y_i = X_i \cup Z_i$ for every i is $\mathcal{Y}$-good. We prove:

▶ **Claim 21.** *For every partition* $\mathcal{Y} = (Y_1, Y_2, Y_3, Y_4)$ *of* V_n*, the number of graphs which permit a great partition with* $X_i = Y_i$ *for every* i *is of larger order then the size of the set* $\mathcal{F} = \mathcal{F}_{\mathcal{Y}}$ *of mediocre string graphs which permit a* $\mathcal{Y}$*-good partition.*

To complete the proof of Theorem 1 we need to show that our two claims hold.

Before doing so, we deviate momentarily and discuss the speed of the string graphs. Combining Theorem 1 and Claim 20, we see that the ratio of the size of $|\text{String}_n|$ over the number of ordered great partitions of graphs on V_n is $\frac{1}{6} + o(1)$, so we need only count the latter. There are 2^{2n} ordered partitions of V_n into $Y_1, ..., Y_4$, and there are $2^{m+|Y_4|}$ graphs for which this is a great partition, where, as before, $m = m(Y_1, Y_2, Y_3, Y_4)$ is the number of pairs of vertices not lying together in some Y_i. This latter term is at most $2^{\frac{3\binom{n}{2}}{4} + \frac{n}{4}}$, which gives us the claimed upper bound on the speed of string graphs. Furthermore, a simple calculation of the 2^{2n} ordered 4-partitions of V_n shows that there is an $\Omega(\frac{1}{n^{\frac{3}{2}}})$ proportion where no two parts differ in size by more than one. This gives us the claimed lower bound.

We now prove our two claims. In proving both, we exploit the fact that if a string graph has a great partition and we fix the subgraph induced by the parts of the partition, then any choice we make for the edges between the sets X_i will yield another string graph permitting the same great partition.

This fact implies that the edge arrangements between the partition elements of a graph permitting a particular great partition are chosen uniformly at random and, hence, are unlikely to lead to a graph permitting some other great partition. This allows us to prove Claim 20, which we do in the full version of the paper [PRY18].

Proof of Claim 21. Let m be the number of pairs of vertices not contained in a partition element and note that there are exactly $(2^{|Y_4|-1})$ choices for $G[Y_4]$ for a graph for which $\mathcal{Y}$ is a great partition, and hence $2^m(2^{|Y_4|-1})$ graphs for which $\mathcal{Y}$ is a great partition.

Our approach is to show that while there may be more choices for the $G[Y_i]$ for mediocre graphs for which $\mathcal{Y}$ is a good partition, for each such choice we have many fewer than 2^m choices for mediocre string graphs extending these subgraphs.

We note that by the definition of good, we need only consider partitions such that each Y_i has size $\frac{n}{4} + o(n)$.

Let $G \in \mathcal{F}$ and let $P(G)$ be the projection of G on the sets (Y_1, Y_2, Y_3, Y_4), that is, the disjoint union of the sets $G[Y_1], G[Y_2], G[Y_3]$, and $G[Y_4]$.

Now, (I) of Theorem 10 bounds the number of choices for $G[Y_i]$ by 1 if $i < 3$ and $2^{|Y_4|}$ if $i = 4$. Furthermore, (III) bounds the number of edges out of Z_i in terms of its size and (II) bounds its size. Putting this all together we obtain the following lemma. Its proof can be found in the full version of the paper [PRY18].

▶ **Lemma 22.** *Let (Y_1, Y_2, Y_3, Y_4) be a partition of V_n, the number of possible projections on (Y_1, Y_2, Y_3, Y_4) of graphs in $\mathcal{F}$ is $o(2^{nb+1+\sqrt{\delta}n|Z|}) = o(2^{|Y_4|-1} \cdot 2^{\sqrt{\delta}n^{2-\gamma}})$.*

For a mediocre graph G in $\mathcal{F}$, we call a set D *versatile* if for each $i \in [4]$ with $Y_i \cap D = \emptyset$, there is clique C_i in Y_i such that for all subsets D' of D there are $\frac{n}{\log n}$ vertices of C_i which are adjacent to all elements of D' and to none of $D \setminus D'$.

▶ **Lemma 23.** *The number of mediocre string graphs in $\mathcal{F}$ such that for some i there is a versatile subset T_i of 3 vertices of Y_i inducing a path or a stable set of size three,is $o(2^m)$.*

Proof. To begin, we count the number of mediocre graphs which extend a given projection on (Y_1, Y_2, Y_3, Y_4) where T_i induces such a graph. We first expose the edges from Y_i to determine if T_i is versatile and then count the number of choices for the remaining edges between the partition elements. If T_i is versatile we choose cliques C_k which show this is the case.

By Corollary 12 (c) or (d), there is a non-string graph J whose vertex set can be partitioned into 3 cliques of size at most five, and a graph J_i isomorphic to the subgraph of the projection induced by T_i. We label these three cliques as J_k for $k \in \{1, 2, 3, 4\} - \{i\}$ and let f be an isomorphism from J_i to T_i. For each vertex $v \in V(J_k)$, let $N(v) = f(N_J(v) \cap V(J_i))$ and Z_v be those vertices of C_k whose neighbourhhod on T_i is $N(v)$. Now, since $|Z_v| \geq \frac{n}{\log n}$ for all v in each $V(J_k)$, for each $k \neq i$, we can choose $n' = \lceil \frac{n}{10 \log n} \rceil$ cliques of size at most five $C_1^k, ..., C_{n'}^k$ such that there is bijection $h_{k,l}$ from J_k to C_l^k with $h_{k,l}(v) \in Z_v$ for every $v \in J_k$.

If we choose our cliques in this way then for any set of three cliques $\{C_{i(k)}^k | k \neq i\}$ there is a choice of edges between the cliques which would make the union of these three cliques with T_i induce J. Thus, there is one choice of edges between the cliques which cannot be used in any extension of H to a string graph. Mimicking an earlier argument, this implies that the number of choices for edges between the partition elements which extend H to a string graph is at most $2^{m - \frac{n^2}{\log^3 n}}$. By the bound in Lemma 22 on the number of possible projections, the desired result follows. ◀

Using Corollary 12 (e) in places of (c) & (d), we can (and do in the in the full version of the paper [PRY18]) prove an analogous result for sets of size 8 intersecting two partition elements. To state it we need a definition. A graph J is *extendible* if there is some non-string graph whose vertex set can be partitioned into two cliques of size five and a set inducing J.

▶ **Lemma 24.** *The number of mediocre string graphs in $\mathcal{F}$ such that for some distinct i and k there are subsets T_i of Y_i and T_k of Y_k, both of size four, whose union is both versatile and induces an extendible graph is $o(2^m)$.*

For every mediocre string graph G in $\mathcal{F}$, we choose a maximum family $\mathcal{W} = \mathcal{W}_G$ of disjoint sets each of which is either (a) contained in some Y_i and induces one of a stable set of size three or a path of length three, or (b) contains exactly four vertices from each of two distinct partition elements and is extendible. For every such choice we count the number of elements of $\mathcal{F}$ whose projection yields the given choice of $\mathcal{W}$.

Now, by the definition of a good partition, each Y_k contains a clique C_k containing half the vertices of X_k and hence at least $\frac{n}{10}$ vertices. Lemmas 23 and 24 imply that we can restrict our attention to graphs for which for any subset T in $\mathcal{W}$, there is a subset N of T and a j with Y_j disjoint from T such that there are fewer than $\frac{n}{logn}$ vertices of C_k which are adjacent to all of N and none of $T - N$. This implies that the number of choices for the edges from T to other partition elements is $o(2^{\frac{3n\lceil T\rceil}{4} - \frac{n}{10000}})$.

Every element of $\mathcal{W}$ must intersect Z, so that $|\mathcal{W}| \leq |Z|$. Set $W^* = \cup_{W \in \mathcal{W}} W$, and let $Y_i' = Y_i - W^*$. Note that for every i, Y_i' has more than $\frac{n}{5}$ vertices and $G[Y_i']$ is the disjoint union of two cliques. Given a choice of $\mathcal{W}$, the number of choices for projections on $V_n \setminus W^*$ is less than 2^n. Mimicking the proof of Lemma 22, the number of choices for the vertices of W^*, and the edges of G from the vertices in W^* which remain within the partition elements of $\mathcal{Y}$ is $O(2^{bn+\sqrt{\delta}|W^*|n})$. Combining this with the result of the last paragraph yields:

▶ **Lemma 25.** *There is a constant C such that the number of mediocre string graphs in $\mathcal{F}$ for which $|\mathcal{W}| > C$ is $o(2^{m+|Y_4|})$.*

So, we can restrict our attention to mediocre graphs which have a partition for which $|\mathcal{W}| \leq C$. Similar tradeoffs allow us to handle them. Full details are found in the full version of the paper [PRY18]. ◀

References

Al93 V. E. Alekseev. On the entropy values of hereditary classes of graphs, *Discrete Math. Appl.* **3** (1993), 191–199.

AlBBM11 N. Alon, J. Balogh, B. Bollobás, and R. Morris. The structure of almost all graphs in a hereditary property, *J. Combin. Theory Ser. B* **101(2)** (2011), 85–110.

BaBS04 J. Balogh, B. Bollobás, and M. Simonovits. On the number of graphs without forbidden subgraph, *J. Combin. Theory Ser. B* **91** (2004), 1–24.

BaBS09 J. Balogh, B. Bollobás, and M. Simonovits. The typical structure of graphs without given excluded subgraphs, *Random Structures Algorithms* **34** (2009), 305–318.

BaBS11 J. Balogh, B. Bollobás, and M. Simonovits. The fine structure of octahedron-free graphs, *J. Combin. Theory Ser. B* **101(2)** (2011), 67–84.

BB11 J. Balogh and J. Butterfield. Excluding induced subgraphs: critical graphs, *Random Structures and Algorithms* **38** (2011), 100–120.

Be59 S. Benzer. On the topology of the genetic fine structure, *Proc. Nat. Acad. Sci.* **45** (1959), 1607–1620.

BoT95 B. Bollobás and A. Thomason. Projections of bodies and hereditary properties of hypergraphs, *Bull. Lond. Math. Soc.* **27** (1995), 417–424.

BoT97 B. Bollobás and A. Thomason. Hereditary and monotone properties of graphs, *The Mathematics of Paul Erdős, Vol. II, R. L. Graham and J. Nešetřil (Eds.)* **14** (1997), 70–78.

Ch34 Ch. Chojnacki (A. Hanani). Über wesentlich unplättbare Kurven im dreidimensionalen Raume, *Fund. Math.* **23** (1934), 135–142.

ErFR86 P. Erdős, P. Frankl, and V. Rödl. The asymptotic number of graphs not containing a fixed subgraph and a problem for hypergraphs having no exponent, *Graphs Combin.* **2** (1986), 113–121.

ErKR76 P. Erdős, D. J. Kleitman, B. L. Rothschild. Asymptotic enumeration of K_n-free graphs, *International Colloquium on Combinatorial Theory, Atti dei Convegni Lincei* **17** (1976), 19–27.

JaU17 S. Janson and A. J. Uzzell. On string graph limits and the structure of a typical string graph, *J. Graph Theory* **84** (2017), 386–407.

KKOT15 J. Kim, D. Kuhn, D. Osthus, T. Townsend. Forbidding induced even cycles in a graph: typical structure and counting, `http://arxiv.org/abs/1507.04944` (2015).

Ko36 P. Koebe. Kontaktprobleme der Konformen Abbildung, *Ber. Sachs. Akad. Wiss. Leipzig, Math.–Phys. Kl.*, **88** (1936), 141–164.

Kr91 J. Kratochvíl. String graphs II: recognizing string graphs is NP-hard, *J. Combin. Theory Ser. B* **52** (1991), 67–78.

KrMa91 J. Kratochvíl and J. Matoušek: String graphs requiring exponential representations, *J. Combin. Theory Ser. B* **53** (1991), 1–4.

PaT06 J. Pach and G. Tóth. How many ways can one draw a graph?, *Combinatorica* **26** (2006), 559–576.

PRY18 J. Pach, B. Reed and Y. Yuditsky. Almost all string graphs are intersection graphs of plane convex sets, `http://arxiv.org/abs/1803.06710`.

PrS91 H. J. Prömel and A. Steger. Excluding Induced Subgraphs I: Quadrilaterals, *Random Structures and Algorithms.* **2** (1991), 53–79.

PrS92a H. J. Prömel and A. Steger. Almost all Berge graphs are perfect, *Combin., Probab. & Comp.* **1** (1992), 53–79.

PrS92b H. J. Prömel and A. Steger. Excluding induced subgraphs. III. A general asymptotic, *Random Structures Algorithms 3(1)* (1992), 19–31.

PrS93 H. J. Prömel and A. Steger. Excluding induced subgraphs II: Extremal graphs, *Discrete Appl. Math.* **44** (1993) 283–294.

ReSc17 B. Reed and A. Scott. The typical structure of an H-free graph when H is a cycle, *manuscript.*

RY17 B. Reed and Y. Yuditsky. The typical structure of H-free graphs for H a tree, *manuscript.*

ScSeSt03 M. Schaefer, E. Sedgwick, and D. Štefankovič. Recognizing string graphs in NP. *Special issue on STOC 2002 (Montreal, QC), J. Comput. System Sci.* **67** (2003), 365–380.

ScSt04 M. Schaefer and D. Štefankovič. Decidability of string graphs, *J. Comput. System Sci.* **68** (2004), 319–334.

Si66 F. W. Sinden. Topology of thin film RC-circuits, *Bell System Technological Journal* (1966), 1639–1662.

Tu70 W. T. Tutte. Toward a theory of crossing numbers, *J. Combinatorial Theory* **8** (1970), 45–53.