Almost All String Graphs are Intersection Graphs of Plane Convex Sets

János Pach[*†‡] *Bruce Reed*[§] *Yelena Yuditsky*[¶]

April 13, 2020

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.

1 Overview

The *intersection graph* of a collection C of sets is the graph 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.* A full-dimensional compact convex set in the plane will be called simply a *convex set*.

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 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].

[*]Rényi Institute, Budapest, *e-mail*: pach@cims.nyu.edu
[†]IST Austria, Vienna, partially supported by Austrian Science Fund (FWF), grant Z 342-N31.
[‡]MIPT, Moscow, partially supported by the Ministry of Education and Science of the Russian Federation in the framework of MegaGrant no 075-15-2019-1926.
[§]School of Computer Science, McGill University, Montreal, Canada, Laboratoire I3S CNRS , and Professor Visitante Especial, IMPA, Rio de Janeiro breed@cs.mcgill.ca.
[¶]School of Computer Science, McGill University, Montreal, Canada; yuditskyL@gmail.com.

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 $V_n = \{1, \ldots, n\}$. In particular, $\textsc{String}_n$ is the collection of all string graphs with the vertex set V_n.

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

Theorem 2 *Every graph G whose vertex set can be partitioned into four parts such that three of them induce a clique in G and the fourth 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 $\{C_1, \ldots, C_n\}$ 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. We can assume without loss of generality that the selected points are in general position. 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 cross 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 $\textsc{Canon}_n \subset \textsc{Conv}_n$, given by the following theorem.

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.*

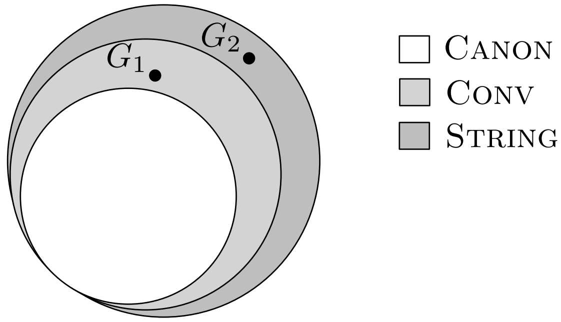

Figure 1: The graph G_1 is a planar graph with more than 20 vertices. The graph G_2 is the graph from the construction of Kratochvíl and Matoušek [KrMa91].

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 discuss previous related results and thereby introduce some needed notation and tools 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 give an outline of the proof of our main result, Theorem 1, and deduce one of its corollaries: an asymptotic formula for the number of string graphs with n vertices (see Theorem 7 below). After some necessary preparation in Section 5, we fill in the details in Sections 6, 7, and 8.

2 The structure of typical graphs in a 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 ($G \in \text{String}$) or that G is an intersection graph of plane convex sets ($G \in \text{Conv}$) are hereditary. The same is true for the properties that G contains no subgraph or 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 $|\mathrm{P}_n|$, 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 previously best known 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 an 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 stable 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*.*

Prömel and Steger proved that for every H, if P is the property of being H-free then P satisfies

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

The work of Prömel and Steger was completed by Alekseev [Al93] and by Bollobás and Thomason [BoT95, BoT97], who proved that this inequality holds for every hereditary graph property P.

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 $r-s$ stable sets, and the number of such graphs is of the order described by 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

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

Theorem 1 allows us to strengthen this result considerably, as shown below it easily implies:

Theorem 7

$$|\textsc{String}_n| = 2^{\frac{3n^2}{8}+\frac{9n}{4}+o(n)}.$$

To prove Theorem 1, we adopt an approach introduced by Prömel and Steger [PrS91]. They observed that a partition of $V(G)$ into a clique and a stable set certifies that G is C_4-free, because no matter which edges between the clique and the stable set are present, there can be no C_4. Considering one such partition they obtained that there are at least $2^{\frac{n^2}{4}-1}$ C_4-free graphs on n vertices. They proved that almost every C_4-free graph permits such a partition and hence the speed of the C_4-free graphs $2^{\frac{n^2}{4}+o(n)}$. They also proved [PrS92a] that 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]. One result in this direction is due to Alon, Balogh, Bollobás, and Morris [AlBBM11]. They proved that almost every graph with a hereditary Property P can be partitioned into $\chi_c(\mathrm{P})$ parts with a simple internal structure.

The first step of our proof is to strengthen the result of Alon et al. [AlBBM11] when P is the string graphs. We show that we can actually find a partition of almost every string graph into four parts such that each part satisfies the properties in their definition of simple structure and furthermore: three of the parts can be made into cliques by deleting $o(n)$ vertices and the fourth can be made into the disjoint union of cliques by deleting $o(n)$ vertices.

The second step of our proof is to show that actually almost every string graph which has such a partition has one in which the union of the four deleted sets is empty.

We give a more detailed outline of the proof of Theorem 1 in Section 4, and fill in the details in Sections 6–8.

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.

Theorem 8 *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 $ij \in E(H)$.*

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.

The graph 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

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

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

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

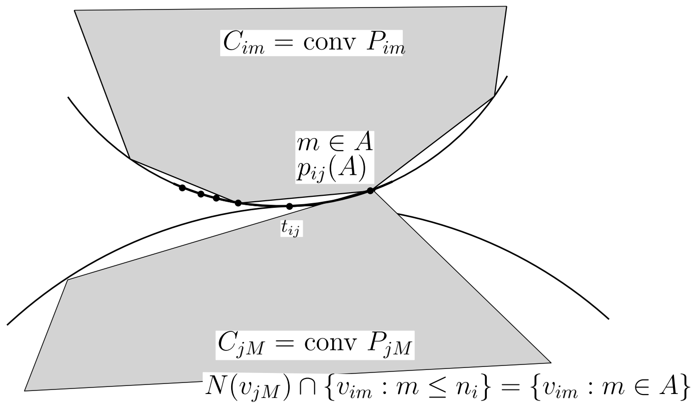

Figure 2: The point $p_{ij}(A)$ is included in P_{jM}.

differs from the intersection graph of

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

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

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

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

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

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 $C_{jM} - D_j$ 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. □

A very similar argument was outlined in [KuKr98].

Applying Theorem 8 to the graph obtained from K_5 by deleting one of its edges, Theorem 4 follows.

4 Outline of the proof of Theorem 1

Definition 9 *By a great partition of a graph G we mean an ordered partition of its vertex set $V(G)$ into X_1, X_2, X_3, X_4 such that (i) for $i \le 3$, $G[X_i]$ is a clique while $G[X_4]$ is the disjoint union of two cliques. We call a graph great if it has a great partition and mediocre otherwise.*

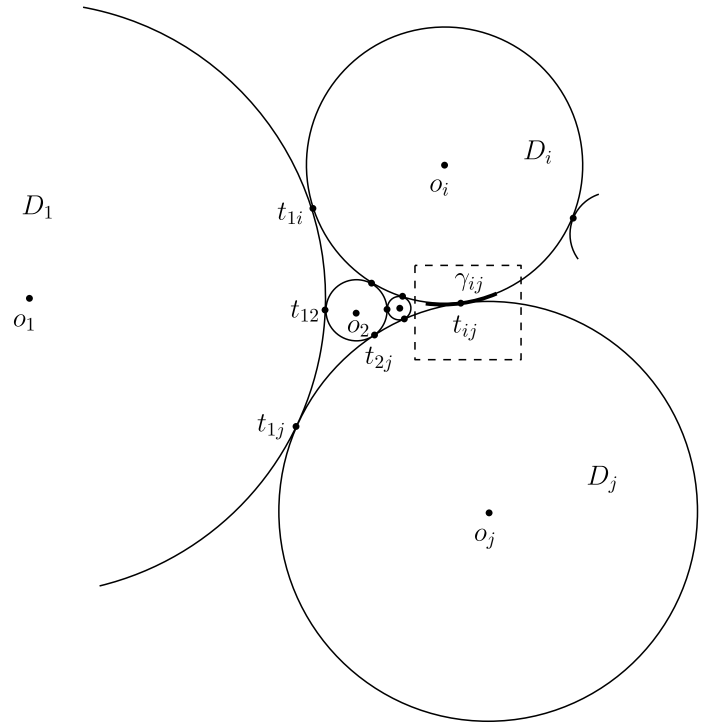

Figure 3: Tangent disks D_i and D_j touching at t_{ij}.

Theorem 1 simply states that the ratio of the number of mediocre string graphs on n vertices over the number of great graphs with n vertices is $o(1)$.

As previously discussed, the proof of Theorem 1 splits into two parts. We first show that for almost every G in STRING_n, we can find a partition of $V(G)$ into four parts each of which has a simple internal structure, including that we can delete an exceptional set of size $o(n)$ so that three of the parts induce cliques and the fourth induces the disjoint union of two cliques. We then show that, we can almost always chose the exceptional set to be empty.

To state the result we obtain in the first step, we need to agree on some notation and terminology. 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 $A \subset V(G)$, let $G[A]$ denote the subgraph of G induced by A.

The result we obtain in the first step is the following.

Theorem 10 *For every sufficiently small δ, there are $\gamma > 0$ and b with the following property. For almost every string graph G on V_n, there is a partition of V_n into $X_1, \dots, X_4, Z_1, \dots, Z_4$ such that there is a set B of at most b vertices for which 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) $|Z_1 \cup Z_2 \cup Z_3 \cup Z_4| \le n^{1-\gamma}$,

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

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

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

We note that our four part partition is $(X_1 \cup Z_1, X_2 \cup Z_2, X_3 \cup Z_3, X_4 \cup Z_4)$. (I) and (II) imply that $Z = Z_1 \cup Z_2 \cup Z_3 \cup Z_4$ is the exceptional set we desire. The other key property is (III), which greatly reduces the choices for the edges within the parts, as we discuss more fully below. On the other hand, (IV) is not surprising. Indeed, having proven the weakening of the theorem where it is deleted, we can obtain the full theorem in just a few lines. Again this is set out below.

See Figure 4 for a pictorial representation of Theorem 10.

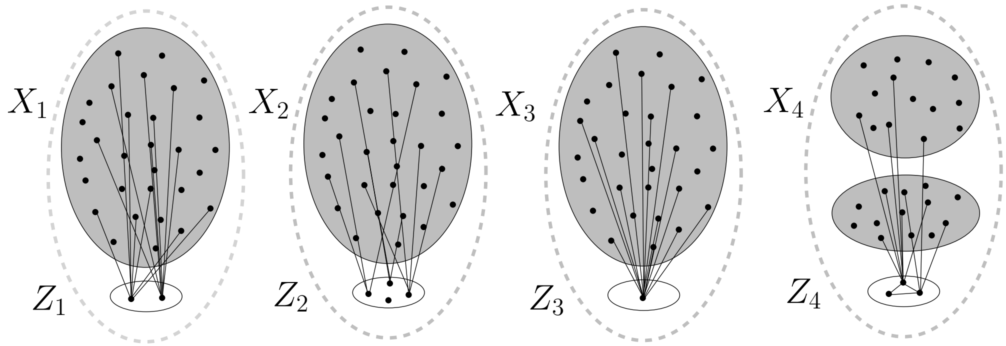

Figure 4: A sketch of a typical string graph as in Theorem 10. The edges between the parts are not drawn. The sets shaded grey are cliques.

Definition 11 *A partition of the vertex set V_n of a graph into 8 parts $X_1, \dots, X_4, Z_1, \dots, Z_4$ is called* good *if it satisfies conditions (I), (II), (III), and (IV).*

*A good partition $X_1, \dots, X_4, Z_1, \dots, Z_4$ is also called $\mathcal{Y}$-*good *for $\mathcal{Y} = (Y_1, Y_2, Y_3, Y_4)$, where $Y_i = X_i \cup Z_i$ for every $i \in [4]$.*

We choose $\delta > 0$ 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, almost every graph in String_n permits a good partition.

By Theorem 10, to complete the proof of Theorem 1. 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. We do that by comparing the number of good partitions of mediocre graphs and the number of great partitions of great graphs. As can be seen from the following claim, we do not over-count much when we consider the number of great partitions in place of the number of great graphs. Obviously, every great graph has at least 6 great partitions, because we can arbitrarily permute the first 3 partition elements. The next statement shows that most great graphs do not permit more than 6 great partitions.

Claim 12 *The ratio between the number of pairs of a great graph together with its great partition and the number of great graphs is $6 + o(1)$.*

Claim 13 *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 number of mediocre string graphs which permit a $\mathcal{Y}$-good partition.*

In order to establish Theorem 1, it is enough to prove Theorem 10, Claim 12, and Claim 13.

After some necessary preparation in the next section, we prove Theorem 10 in Section 6. Claims 12 and 13 are proved in Sections 7 and 8, respectively.

Here we show how Theorem 7 follows from Theorem 1 and Claim 12.

Proof of Theorem 7 Combining Theorem 1 and Claim 12, we see that the ratio of the size of $\textsc{String}_n$ 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 $Y_1,\ldots,Y_4$, and for any such partition there are $O(1)2^{m_{\mathcal{Y}}+|Y_4|}$ graphs for which this is a great partition. This latter term is at most $2^{3n^2/8+6+\frac{n}{4}+o(n)}$, which gives us the claimed upper bound on the speed of string graphs. Furthermore, a simple calculation shows that in a an $\Omega(\frac{1}{n^{3/2}})$ proportion of all 2^{2n} ordered 4-partitions of V_n, no two parts differ in size by more than 1. This gives the desired lower bound. □

Before ending this section, we give a bit of intuition about the proof of Claim 13 which is a key element of our argument. We note that, for a given ordered partition $\mathcal{Y}$, there are fewer than 2^n choices for the 4-tuple $(G[Y_1],G[Y_2],G[Y_3],G[Y_4])$ over G for which $\mathcal{Y}$ is a great partition as the first three of its elements are cliques and the last is the disjoint union of 2 cliques. This is dwarfed by the number of choices for this 4-tuple over mediocre G for which $\mathcal{Y}$ is a good partition.

To counterbalance this fact, we need to consider the edges between the partition elements. If we insist that Y_4 can be partitioned into a specific pair of cliques, and every other Y_i is a clique, then every choice for the edges between the parts yields a great graph for which $\mathcal{Y}$ is a great partition. In contrast, for a choice of $(G[Y_1],G[Y_2],G[Y_3],G[Y_4])$, for some mediocre graph G for which $\mathcal{Y}$ is a good partition, there are much fewer choices for the edges between the parts that yield a mediocre string graph for which $\mathcal{Y}$ is a good partition.

In the proof of Claim 13, we repeatedly exploit information about a quadruple to bound the number of its extensions to mediocre string graphs for which it forms a good partition. This tradeoff between greater choice within the part and less choice between the parts will also be crucially important in the proof of Theorem 10.

5 The starting point

Our starting point is essentially a special case of a result of Alon *et al.*[AlBBM11] which holds for all hereditary properties of graphs. To state it, we need some notations.

For any disjoint subsets $A,B\subset V(G)$, let $G[A,B]$ denote the *bipartite* subgraph of G consisting of all edges of G running between A and B. The *symmetric difference* of two sets, X and Y, is denoted by $X\triangle Y$.

Following Alon *et al.*, for any integer $k>0$, we define $U(k)$ as the bipartite graph with vertex classes $\{1,\ldots,k\}$ and $\{I : I\subset\{1,\ldots,k\}\}$, 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.

Definition 14 *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 $G[A,B]\subseteq G$ 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 the following.

Theorem 15 *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, $S_1, \ldots, S_4, A_1, \ldots, A_4$, such that for some set B of at most b vertices

(a) $G[S_i]$ is $U(k)$-free for every i ($1 \le i \le 4$);

(b) $|A_1 \cup \ldots \cup A_4| \le n^{1-\epsilon}$; 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

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

For those familiar with the paper of Alon, Balogh, Bollobás, and Morris [AlBBM11], we present the details of the minor modifications required for the proof of Theorem 15.

Proof It is sufficient to prove the result for δ sufficiently small. We set $\delta = 3\alpha$ for some $\alpha > 0$ which is required to be sufficiently small. So, in what follows, we can and do replace δ by 3α. (This replacement is essential to readability for those readers who choose to work through [AlBBM11], as δ denotes a different quantity in that paper.)

We essentially follow and repeat the proof of Theorem 1, given in Section 7 of [AlBBM11]. Our Theorem 15 differs from their Theorem 1 in the following ways.

(i) In our case, the hereditary family $\mathcal{P}$ is the family of string graphs. Hence, by Pach and Tóth [PaT06], we have $\chi_c(P) = 4$,

(ii) We allow k to be *any* large enough integer, rather than *one fixed* large integer.

(iii) We allow α to be *arbitrarily small*, as long as it is small enough in terms of k (and $\mathcal{P}$).

(iv) ϵ is chosen as a function of α and k.

(v) There is an integer b which is chosen as a function of α and k such that there exist a choice B of at most b vertices and a partition of A into A_1, A_2, A_3, A_4 for which our property (c) holds – with δ replaced by 3α.

(vi) The sentence beginning with *"Moreover"* has to be deleted.

Let $k \in \mathbb{N}$ be sufficiently large, let $\alpha = 3\delta > 0$ be sufficiently small in terms of k, and choose γ sufficiently small in terms of k and α, and ϵ sufficiently small in terms of all of these parameters. By Lemma 17 of [AlBBM11], almost every graph $G \in \mathcal{P}$ has a BBS-partition P for $(\epsilon, \delta, \gamma)$ into 4 parts. Let B be a maximal (2α)-bad set for (G, P).

By Lemmas 22 and 23 of [AlBBM11], for almost every G, there exists an α-adjustment $P' = (S'_1, \ldots, S'_4)$ of (G, P) with respect to B. Let $A = U(G, P', k)$ be the exceptional set given by the algorithm. By Lemma 24, for almost every G, $|A| \le n^{1-\epsilon}$. Let c stand for $c(\alpha, \mathcal{P})$ of Lemma 19. Lemma 24 can be strengthened so that it is also possible to deduce that $|B| \le c$ for almost every $G \in \mathcal{P}$.

Let $A_i = S'_i \cap A, S_i = S'_i - A$, $i \in [4]$. Now, part(a) of our theorem is the same as Theorem 1(b) in [AlBBM11], and part (b) is the same as Theorem 1(a), where ϵ is $\frac{\alpha}{2}$. Part (c) follows immediately from the fact that $S'_1, \ldots, S'_4$ is an α-adjustment.

Next, we state the necessary strengthening of Lemma 24 from [AlBBM11]. Let B denote the set with $|B| > c(\alpha, \mathcal{P})$, where $c(\alpha, \mathcal{P})$ is the constant in Lemma 19. Let $U'(\mathcal{P}_n, \alpha, k)$ be the set $U(\mathcal{P}_n, \alpha, k)$.

Lemma 16 *Let $r \geq 2$ and let $\mathcal{P}$ be a hereditary property of graphs with $\chi_c(\mathcal{P}) = r$ (in our case, $r = 4$). There exists $k = k(\mathcal{P}) \in \mathbb{N}$ such that for any α sufficiently small in terms of k and $\mathcal{P}$, the following holds.*

Let $\epsilon, \alpha, \gamma > 0$ be sufficiently small, and let $n \in \mathbb{N}$ be sufficiently large. Then we have $|U'(\mathcal{P}_n, \alpha, k)| \leq 2^{(1-\frac{1}{r})\binom{n}{2} - n^{2-2\alpha}}$.

The original proof of Lemma 24 actually proves this stronger result, provided that
(a) in the first paragraph, we set out that c to be $c(\alpha, \mathcal{P})$ from their Lemma 19;
(b) in the definition of $U_n = \mathcal{U}(P_n, \alpha, k) - (\mathcal{B}(\mathcal{P}, \alpha, n^{1-2\alpha}) \cup \mathcal{D}(\mathcal{P}_n, \alpha))$, replace $n^{1-2\alpha}$ by c;
(c) delete the assumption *"if $c = c(\alpha, \mathcal{P})$ is sufficiently large"*. □

6 The proof of Theorem 10

The aim of this section is to deduce Theorem 10 from Theorem 15. We will need the following result from [AlBBM11].

Lemma 17 (Theorem 2 in [AlBBM11]) *For every $k \in \mathbb{N}$, there is $\rho = \rho(k) > 0$ such that the number of $U(k)$-free graphs on $V_t = \{1, 2, \ldots, t\}$ is less than $2^{t^{2-\rho}}$.*

Proof of Theorem 10 We choose k large enough and $0 < \delta < \frac{1}{40}$ sufficiently small in terms of k so that for some b and ϵ, almost every string graph has a partition into $S_1, S_2, S_3, S_4, A_1, A_2, A_3, A_4$ such that for some set B of at most b vertices, (a)-(c) of Theorem 15 hold.We call such a partition an *ABBM partition.*

Let $0 < \alpha < \min\{\epsilon, \frac{\rho}{2}\}$ where ρ is the parameter from Lemma 17 for the above k. Let $\gamma = \frac{\alpha}{40}$ and $l = l(n) = \lceil n^{1-\frac{\alpha}{7}} \rceil$.

In the sequel, we show that almost every string graph G has an ABBM partition with the following properties.

(1) For each $i \in [3]$, $\exists Y_i \subset S_i$ with $|Y_i| \leq \frac{1}{8} \cdot n^{1-\gamma}$ such that $G[S_i - Y_i]$ is a clique.

(2) $\exists Y_4 \subset S_4$ with $|Y_4| \leq \frac{1}{8} \cdot n^{1-\gamma}$ such that $G[S_4 - Y_4]$ is the disjoint union of two cliques.

(3) For every $i \in [4]$, $||S_i \cup A_i| - \frac{n}{4}| \leq n^{1-\gamma}$.

Note that this implies Theorem 10. Indeed, by setting $Z_i = A_i \cup Y_i$ and $X_i = S_i - Y_i$ for each $i \in [4]$, part(I) of the theorem follows from (1) and (2). Furthermore, for sufficiently large n, part (II) also follows from (1), (2), Theorem 15(b), and from the fact that $\gamma = \frac{\alpha}{40} \leq \frac{\epsilon}{40}$. We obtain part (III), because $(S_1, \ldots, S_4, A_1, \ldots, A_4)$ is an ABBM partition and $S_i \cup A_i = X_i \cup Z_i$ for each $i \in [4]$. Finally, part (IV) from property (3) above.

Theorem 15 tells us that almost every string graph belongs to the family

$$\mathcal{G}_1 = \mathcal{G}_1(k, \delta) := \{G \in \textsc{String} \mid G \text{ has an ABBM partition}\}.$$

So we only need to prove that the number of graphs in $\mathcal{G}_1$ for which there is an ABBM partition *not satisfying* at least one of the properties (1), (2), (3), is $o(|(\mathcal{G}_1)_n|)$. Consider three other special families of graphs, and let $l = l(n) = \lceil n^{1-\frac{\alpha}{7}} \rceil$, as was specified at the beginning of the proof.

$$\mathcal{G}_2 := \{G \in (\mathcal{G}_1)_n \mid G \text{ has an ABBM partition for which } \exists i \in [4] \text{ s.t. } ||S_i \cup A_i| - \frac{n}{4}| > n^{1-\gamma}\}.$$

$\mathcal{G}_3 := \{G \in (\mathcal{G}_1)_n - \mathcal{G}_2 \mid G$ has an ABBM partition for which $\exists i \in [4]$ s.t. $G[S_i]$ contains l disjoint sets of size 3, each inducing a path or a stable set$\}$.

$\mathcal{G}_4 := \{G \in (\mathcal{G}_1)_n - \mathcal{G}_2 - \mathcal{G}_3 \mid G$ has an ABBM partition for which $\exists i \neq j \in [4]$ s.t. $G[S_i]$ and $G[S_j]$ contain l disjoint sets of size 4, each inducing the disjoint union of a vertex and a triangle$\}$.

We show the following.

Lemma 18 $|\mathcal{G}_2 \cup \mathcal{G}_3 \cup \mathcal{G}_4| = o(|(\mathcal{G}_1)_n|)$.

Lemma 18 implies that almost every string graph is in $(\mathcal{G}_1)_n - (\mathcal{G}_2 \cup \mathcal{G}_3 \cup \mathcal{G}_4)$. For any such graph G, we consider an ABBM partition $S_1, S_2, S_3, S_4, A_1, A_2, A_3, A_4$. Since G is not in $\mathcal{G}_3$, no $S_i, i \in [4]$, has l disjoint subsets, each od which induces a path or stable set with 3 vertices. Since G is not in $\mathcal{G}_4$, by swapping 4 with some other index if necessary, we can ensure that no $S_i, i \in [3]$, has l disjoint subsets of S_i, each of which induces the disjoint union of a triangle and a vertex. For $i \in [3]$, let Y_i be a *maximum* set of disjoint subsets of S_i, each of which induces a path or stable set of size 3, or is the disjoint union of a triangle and a vertex. Let Y_4 be a *maximum* set of disjoint subsets of S_4, each inducing either a path or a stable set with 3 vertices. Clearly, each Y_i has at most $7l$ elements. Since $l = \lceil n^{1-\frac{\alpha}{7}} \rceil$ and $\gamma = \frac{\alpha}{40}$, this is less than $\frac{n^{1-\gamma}}{8}$, provided n is large enough.

Any graph that has no induced paths on 3 vertices is the disjoint union of cliques. If, in addition, the graph has no stable set of size 3, it is the disjoint union of at most 2 cliques. If, on top of this, the graph contains no subset that induces the disjoint union of a vertex and a triangle, and it has at least 5 vertices, then it is a clique. Thus, for every sufficiently large n, every ABBM partition of G satisfies (1), (2), and (3).

To complete the proof of Theorem 10, it remains to establish Lemma 18.

Proof of Lemma 18 We compute separately a lower bound on the size of $\mathcal{G}_1$ and an upper bound on the size of $\mathcal{G}_2 \cup \mathcal{G}_3 \cup \mathcal{G}_4$, and we show that the latter is of smaller order than the former. Computing a lower bound on the size of $\mathcal{G}_1$ is easy.

Lemma 19 $|(\mathcal{G}_1)_n| \geq 2^{\frac{3n^2}{8}-6}$.

Proof For any $n > 3$, fix a partition S_1, S_2, S_3, S_4 of V_n into 4 parts, where each S_i has either $\lceil n/4 \rceil$ or $\lfloor n/4 \rfloor$ vertices. Let A_i be empty for all i. For any sufficiently large n and for any graph G on V_n, for which each S_i is a clique, this yields an ABBM partition. To see this, let B contain exactly one vertex from each of S_1, S_2, S_3, S_4. Now every choice of edges between the S_i yields a distinct string graph with the given certifying partition. Since there are more than $\frac{3n^2}{8} - 6$ pairs of vertices that lie in distinct partition elements, the statement is true. □

To obtain an upper bound on $|\mathcal{G}_2 \cup \mathcal{G}_3 \cup \mathcal{G}_4|$, consider each of the possible 8^n partitions of V into $(S_1, \ldots, S_4, A_1, \ldots, A_4)$, separately.

For any partition $\mathcal{Y}$ of V_n into 4 parts, define $m(\mathcal{Y})$ to be the number of pairs of vertices not taken from the same part. When we use m in this section, we mean $m(S_1 \cup A_1, S_2 \cup A_2, S_3 \cup A_3, S_4 \cup A_4)$.

Definition 20 *The projection of G onto a partition of its vertex set is the set of subgraphs induced by the partition elements.*

A projection onto $(S_1 \cup A_1, \ldots, S_4 \cup A_4)$ is an ABBM projection for $(S_1, S_2, S_3, S_4, A_1, A_2, A_3, A_4)$ if it is the projection of some graph for which $(S_1, S_2, S_3, S_4, A_1, A_2, A_3, A_4)$ is an ABBM partition.

We now use Lemma 17 to bound the number of different ABBM projections for a given $(S_1, S_2, S_3, S_4, A_1, A_2, A_3, A_4)$.

Lemma 21 *The number of possible ABBM projections for a partition* $(S_1, S_2, S_3, S_4, A_1, A_2, A_3, A_4)$ *of* V_n *is at most* $O(2^{8n^{2-\alpha}})$.

Proof By Lemma 17, the number of choices for a $U(k)$-free graph on each S_i is $O(2^{n^{2-\rho}}) = O(2^{n^{2-\alpha}})$. Let $A = \cup_{i=1}^{4} A_i$. Since, in an ABBM partition, $|A|$ is at most $n^{1-\epsilon}$, there are at most $2^{n^{2-\epsilon}} = O(2^{n^{2-\alpha}})$ choices for the edges with at least one endpoint in the set A. It follows that there are at most $O(2^{8n^{2-\alpha}})$ choices for the ABBM projection for this partition, over all graphs for which it is an ABBM partition. $\square$

Lemma 22 $|\mathcal{G}_2| = o(|(\mathcal{G}_1)_n|)$.

Proof Let $S_1, S_2, S_3, S_4, A_1, A_2, A_3, A_4$ be a partition of V_n such that for some $i \in [4]$, we have $||S_i \cup A_i| - \frac{n}{4}| > n^{1-\gamma}$. Then $m < \frac{3n^2}{8} + 6 - n^{2-\gamma}$. There are at most 2^m graphs that have the same ABBM projection on $S_1, S_2, S_3, S_4, A_1, A_2, A_3, A_4$. Combining this bound with the upper bound in Lemma 21, and comparing it with the lower bound $|(\mathcal{G}_1)_n| \geq 2^{\frac{3n^2}{8}-6}$ from Lemma 19, we get the desired result. $\square$

In order to complete the proof of Theorem 10, we need to exploit the fact that for some specific choice of an ABBM projection for $(S_1, S_2, S_3, S_4, A_1, A_2, A_3,\ A_4)$, it cannot occur that all 2^m graphs extending this projection are string graphs. Specifically, we will use the following result.

Lemma 23 *Let* H *be a non-string graph, and let* L_1, L_2, L_3, L_4 *be a partition of* $V(H)$. *Let* P *be a projection on a partition* (Y_1, Y_2, Y_3, Y_4) *of* V_n. *Assume that for each* $i \in [4]$, *we can choose a family* $\mathcal{W}^i$ *of* q *disjoint subsets of* Y_i, *each inducing a graph isomorphic to* $H[L_i]$.

Then the number of string graphs whose projection on $\mathcal{Y}$ *is* P *is at most* $2^{m(\mathcal{Y})}\left(1 - \frac{1}{2^{\binom{|V(H)|}{2}}}\right)^{\frac{q^2}{4}}$.

Proof It is well known that there is a prime p between $\frac{q}{2}$ and q. For each $i \in [4]$, let $J_0^i, \ldots, J_{p-1}^i$ be p members of $\mathcal{W}^i$. For $1 \leq r, s \leq p$, consider the 4-tuple $J_r^1, J_{r+s}^2, J_{r+2s}^3, J_{r+3s}^4$, where addition is taken modulo p. For each of the $p^2 \geq \frac{q^2}{4}$ such 4-tuples, there is a way to choose edges between pairs of vertices in distinct elements of the 4-tuple so that we get a copy of H. Hence, the resulting graph where such a choice is made is not a string graph. Therefore, among the $c \leq 2^{\binom{|V(H)|}{2}}$ choices for the edges between the elements of the 4-tuple, at most $c-1$ can occur in a string graph with the given projection on $S_1, S_2, S_3, S_4, A_1, A_2, A_3, A_4$. Furthermore, by our choice of the 4-tuples, for any x in S_i and y in S_j with $i \neq j$, there is at most one 4-tuple containing both x and y. The result follows. $\square$

To exploit this lemma, we need to consider the partitions of certain non-string graphs set out in the following result, which is a slight generalization of Lemma 3.2 in [PaT06] with essentially the same proof.

Lemma 24 *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.*

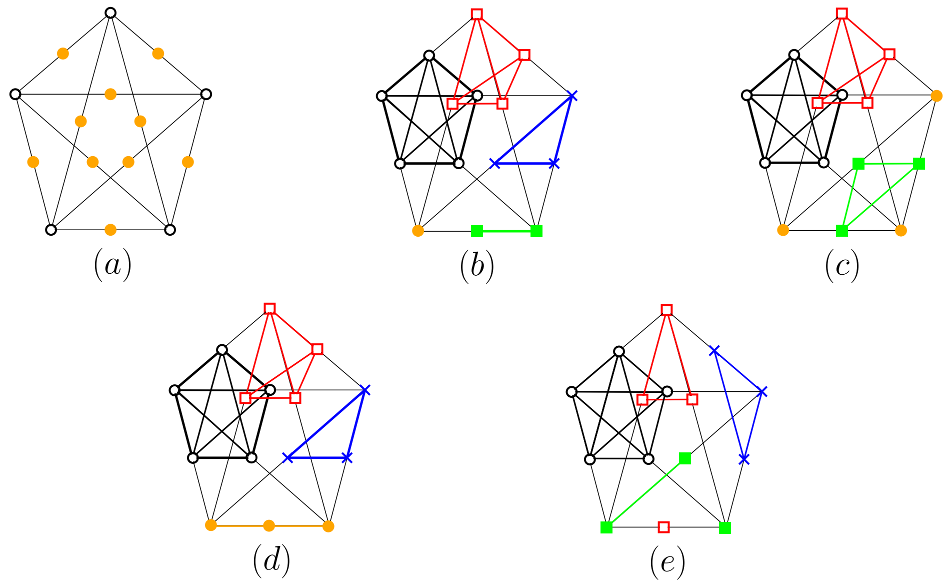

Figure 5: Possible partitions of a non-string graph.

Proof Suppose for contradiction that there is a graph H with the above properties that has a string representation. Choose such a graph with the maximum number of edges. Continuously contract each string (curve) representing v_i ($1 \le i \le 5$) to a point p_i, Note that, by our choice of H, at the end of the process we still have a string representation of H. For every pair $i \neq j$, consider a non-selfintersecting arc of the curve representing v_{ij} with endpoints p_i and p_j. These arcs define a drawing of K_5, in which no two *independent* edges intersect. However, K_5 is not a planar graph, hence, by a well known theorem of Hanani and Tutte [Ch34], [Tu70], no such drawing exists. □

Corollary 25 *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 25.

Remark 26 *Corollary 25 immediately implies that* $\chi_c(\text{STRING}) \le 4$*. Indeed, there exist* $(5, s)$*-colorable non-string graphs for* $s \le 3$ *(by (a)), for* $s = 4$ *and* 5 *(by (b)). In fact, we have* $\chi_c(\text{STRING}) = 4$ *[PaT06].*

Lemma 27 $|\mathcal{G}_3| = o(|(\mathcal{G}_1)_n|)$.

Proof We define two subfamilies of $(\mathcal{G}_1)_n$.

$$\mathcal{H}_1 := \{G \in (\mathcal{G}_1)_n - \mathcal{G}_2 \mid G \text{ has an ABBM partition for which } \exists i \neq j \in [4] \text{ s.t. } G[S_i] \text{ and } G[S_j] \text{ contain } l \text{ disjoint stable sets of size } 10\}.$$

$$\mathcal{H}_2 := \{G \in (\mathcal{G}_1)_n - \mathcal{G}_2 - \mathcal{H}_1 \mid G \text{ has an ABBM partition for which } \exists i \in [4] \text{ s.t. } G[S_i] \text{ does not contain } l \text{ disjoint cliques of size 5}\}.$$

First, we show that $|\mathcal{H}_1| = o(|(\mathcal{G}_1)_n|)$. Consider a partition of V_n into parts $S_1, S_2, S_3, S_4, A_1, A_2, A_3, A_4$. By Lemma 21, we know that the number of ABBM projections for the above partition is $O(2^{8(n^{2-\alpha})})$. By Corollary 13(a), for any $\{i, j\} \subseteq [4]$, there is a non-string graph H whose vertex set can be partitioned into L_1, L_2, L_3, L_4 with the following property: if $k \notin \{i, j\}$, then L_k is non-empty, and if $k \in \{i, j\}$, then L_k is a stable set of size at most 10. So, applying Lemma 23 with $q = l$, the number of string graphs G for which a specific ABBM projection for $S_1, S_2, S_3, S_4, A_1, A_2, A_3, A_4$ shows that G is in $\mathcal{H}_1$, because there are distinct $G[S_i]$ and $G[S_j]$ containing l disjoint stable sets of size 10, is at most $2^m(1 - \frac{1}{2^{200}})^{l^2}$. Hence, the number of graphs in $\mathcal{H}_1$ is at most $2^{8(n^{2-\alpha})} \cdot 2^m \cdot 2^{-cn^{2-\frac{\alpha}{2}}}$ for some constant $c > 0$. Since $m < \frac{3n^2}{8} + 6$, taking into account the lower bound on $|(\mathcal{G}_1)_n|$ in Lemma 19, we get $|\mathcal{H}_1| = o(|(\mathcal{G}_1)_n|)$, as desired.

Secondly, we show that $|\mathcal{H}_2| = o(|(\mathcal{G}_1)_n|)$. For this, we need the following observation. By Ramsey theorem, every set of 2^{15} vertices contains either a clique of size 5 or a stable set of size 10. Therefore, if a graph J does not contain l disjoint stable sets of size 10, then it must contain $(|V(J)| - 10(l-1) - 2^{15})/5$ disjoint cliques of size 5.

Let $G \in \mathcal{H}_2$. By the definition of $\mathcal{H}_2$, there is an ABBM projection for some $S_1, S_2, S_3, S_4, A_1, A_2, A_3, A_4$ such that at least one $G[S_i]$ does not contain l disjoint cliques of size 5. Since $\mathcal{H}_2$ is disjoint from $\mathcal{G}_2$, every S_j contains more than $\frac{n}{5}$ vertices. By the last paragraph, this implies that S_i must contain a set $\mathcal{S}^i$ of l disjoint stable sets $Z_1^i, \ldots, Z_l^i$ of size 10. Since $\mathcal{H}_2$ is disjoint from $\mathcal{H}_1$, we also obtain by the last paragraph that no $G[S_j]$ with $j \neq i$ contains l disjoint stable sets of size 10, and, hence, every such $G[S_j]$ contains a set $\mathcal{C}^j$ of l disjoint cliques $Z_1^j, \ldots, Z_l^j$ of size 5.

By Corollary 25 (c), there is a non-string graph which can be partitioned into $L_1, \ldots, L_4$, where for $j \neq i$, L_i is a clique of size at most 5, and L_i is a stable set of size at most 3. Applying Lemma 23 with $q = l$, the number of ways to extend an ABBM projection of G which shows that G does not belong to $\mathcal{H}_2$ to a string graph is at most $2^{\frac{3n^2}{8}+6}(1 - \frac{1}{2^{\binom{18}{2}}})^{\frac{l^2}{4}}$. Hence, as before, by Lemma 21, the number of graphs in $\mathcal{H}_2$ is at most $2^{8(n^{2-\alpha})} \cdot 2^{\frac{3n^2}{8}+6} 2^{-cn^{2-\frac{\alpha}{2}}}$, for some $c > 0$. Comparing this bound with the lower bound on $|(\mathcal{G}_1)_n|$, we get that $|\mathcal{H}_2| = o(|(\mathcal{G}_1)_n|)$, as desired.

To complete the proof of the lemma, we need to bound the size of $\mathcal{G}_3 - (\mathcal{G}_2 \cup \mathcal{H}_1 \cup \mathcal{H}_2)$. Consider a graph G in this class and an ABBM projection of G for a partition $S_1, S_2, S_3, S_4, A_1, A_2, A_3, A_4$ of V_n which shows that $G \in \mathcal{G}_3 - (\mathcal{G}_2 \cup \mathcal{H}_1 \cup \mathcal{H}_2)$. In particular, for some $i \in [4]$, we can choose a set $\mathcal{S}^i$ of $\frac{l}{2}$ disjoint subsets of size 3 in $G[S_i]$ such that either each subset induces a path or each subset induces a stable set. Further, since G does not belong to $\mathcal{G}_1 \cup \mathcal{H}_1 \cup \mathcal{H}_2$, for every $j \in [4] - \{i\}$ we can choose a collection $\mathcal{C}^j$ of $\frac{l}{2}$ disjoint subsets of size 5 in $G[S_j]$, each of which induces a clique.

By Corollary 25 (c) or (d), we obtain that there is a non-string graph H which can be partitioned into $L_1, .., L_4$ such that for $j \neq i$, L_j is a clique of size at most 5 while for each S in $\mathcal{S}^i$, $G[S]$ induces $H[L_i]$. Applying Lemma 23 with $q = \frac{l}{2}$, the number of different ways how to extend the ABBM projection of G which shows that $G \in \mathcal{G}_3 - (\mathcal{G}_2 \cup \mathcal{H}_1 \cup \mathcal{H}_2)$ to a string graph, is at most $2^{\frac{3n^2}{8}+6}(1 - \frac{1}{2^{153}})^{\frac{l^2}{16}}$. Hence, as before, by Lemma 21, the number of graphs in $\mathcal{G}_3 - (\mathcal{G}_2 \cup \mathcal{H}_1 \cup \mathcal{H}_2)$ is at most $2^{8(n^{2-\alpha})} \cdot 2^{\frac{3n^2}{8}+6} 2^{-cn^{2-\frac{\alpha}{2}}}$, for some $c > 0$. Again, comparing this bound with the lower bound on $|(\mathcal{G}_1)_n|$, we obtain $|\mathcal{G}_3 - (\mathcal{G}_2 \cup \mathcal{H}_1 \cup \mathcal{H}_2)| = o(|(\mathcal{G}_1)_n|)$. □

Using similar ideas and Corollary 25 (e), we can also derive the following.

Lemma 28 $|\mathcal{G}_4| = o(|(\mathcal{G}_1)_n|)$.

Proof Our string graph G is not in $\mathcal{G}_2 \cup \mathcal{G}_3$, thus Ramsey theory tells us that each $G[S_i]$ contains a set $\mathcal{C}^i$ of l disjoint cliques of size 5. Since $G \in \mathcal{G}_4$, we can find distinct $G[S_i]$ and $G[S_j]$ containing collections $\mathcal{T}^i$ and $\mathcal{T}^j$, resp., of l disjoint sets, each of which induce the disjoint union of a vertex and a triangle. Now, Corollary 25 (e) implies that there is a non-string graph H whose vertex set can be partitioned into $L_1, \ldots, L_4$ with the property that for $k \notin \{i,j\}$, L_k is a clique of size at most 5, while L_i and L_j can be obtained as the disjoint union of a clique and a triangle. Again, applying Lemma 23 and Lemma 21, we obtain the desired result. □

This completes the proof Lemma 18 and, thus, of Theorem 10. □

□

7 The proof of Claim 12

We will exploit the fact that if a string graph has a great partition and we fix the subgraphs 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 which permits the same great partition. This fact implies that the edge patterns between different parts of a particular great partition are chosen uniformly at random and it is very unlikely that they define a graph which also permits some other great partition. This allows us to prove Claim 12.

Proof of Claim 12 To prove our claim, we focus on ordered pairs of a graph and a corresponding great partition $(G, (X_1, X_2, X_3, X_4))$, that is, (X_1, X_2, X_3, X_4) is a great partition of G with the following property which we denote by (P*):

(a) any two vertices of G in a part X_i that induces a clique have at least $\frac{13n}{32}$ common neighbours;

(b) any two vertices in different parts have fewer than $\frac{13n}{32}$ common neighbours;

(c) for every part X_i and every vertex $v \notin X_i$, v forms a P_3 with two vertices of X_i; and

(d) X_4 does not induce a clique.

Clearly, every great graph has *at least* six great partitions obtained by permuting the indices of the partition elements. We show that
(i) every graph on V_n has *at most* six great partitions satisfying (P*), and
(ii) almost every great partition of a great graph on V_n satisfies property (P*).
These two statements together prove our claim.

To prove (i), we assume that (X_1, X_2, X_3, X_4) and (X'_1, X'_2, X'_3, X'_4) are two great partitions of a graph G, both of which satisfy property (P*). Clearly, (a) and (b) tell us that for $i \leq 3$, X_i is contained in some X'_j. It follows from property (c) that each such X_i is, in fact, of size at least 2 and equal to some X'_j. Hence, the two sets of partition elements are the same. By property (d), $X'_4 = X_4$. This proves (i).

Next, we prove (ii). For any (ordered) partition $\mathcal{X} = (X_1, X_2, X_3, X_4)$ of V_n, let $I = I(\mathcal{X})$ be all choices of edges within the partition elements which result in this partition being great. As before, let $m = m(\mathcal{X})$ denote the number of pairs of vertices not lying in the same partition element.

There are $|I|2^m$ graphs for which this partition is great, as we can pair any choice from I with any choice of edges between the partition elements. Furthermore, I can be chosen by specifying a partition of X_4 into two disjoint cliques. Thus, there is at least one and at most 2^{n-1} choices for I. If there is $i \in [4]$ such that $|X_i| \geq \frac{n}{4} + cn^{\frac{2}{3}}$ for some $c > 0$, then $m \leq \frac{3n^2}{8} - \frac{c^2 n^{\frac{4}{3}}}{2}$, which accounts only for $o(1)$ proportion of great graphs. So, we can assume that, for almost every great partition of a great graph, we have that $|X_i| = \frac{n}{4} + o(n^{\frac{2}{3}})$ for every $i \in [4]$. It remains to show that the for any partition $\mathcal{X}$ for which $|X_i| = \frac{n}{4} + o(n^{\frac{2}{3}})$ for every $i \in [4]$, the number of great graphs G for which $\mathcal{X}$ is a great partition failing to satisfy property (P*), is $o(|I|2^m)$.

As $|X_4| = \frac{n}{4} + o(n^{\frac{2}{3}})$, almost every graph on $|X_4|$ vertices which is the disjoint union of two cliques is not a clique. So, there are $o(|I|)2^m = o(|I|2^m)$ graphs for which X is a partition for which (d) fails to hold.

Choose a great graph from amongst the 2^m graphs extending a choice of I uniformly at random, by adding each edge between two vertices in different parts independently with probability $\frac{1}{2}$.

Observe that, given any three vertices u, v, w not contained in the same X_i, the probability that w is a common neighbour of both u and v is at most $\frac{1}{2}$ if w lies in the same partition element as one of u or v, and is exactly $\frac{1}{4}$ otherwise. Taking into account the restriction on the size of the X_i, we obtain that the expected number of common neighbours of two vertices is at most $\frac{1}{4} \cdot \frac{2n}{4} + \frac{1}{2} \cdot \frac{2n}{4} + o(n) = \frac{3n}{8} + o(n)$ if they are in different partition elements, and at least $\frac{n}{4} + \frac{1}{4} \cdot \frac{3n}{4} + o(n) = \frac{7n}{16} + o(n)$ if they are in the same partition element that induces a clique.

Furthermore, given the partition, the (random) number of common neighbours of two vertices which lie together in some X_i that forms a clique is the sum of $|X_i| - 2$ and $n - |X_i|$ independent random variables, each of which is 1 with probability $\frac{1}{4}$ and 0 with probability $\frac{3}{4}$. In the same vein, if u lies in X_i and v lies in X_j for distinct i and j, then their number of common neighbours is the sum of $n - |X_i - N(u)| - |X_j - N(v)|$ independent random variables, $|X_i \cap N(u)| + |X_j \cap N(v)|$ of which are equally likely to be 0 or 1, and the rest of which are 0 with probability $\frac{3}{4}$ and 1 with probability $\frac{1}{4}$. Thus, for every choice of I, $\binom{n}{2}$ applications of the Chernoff bound, one for each pair of vertices, show that the number of great graphs extending this choice, for which either (P*)(a) or (P*)(b) fails is $o(2^m)$.

Consider now an X_i and a vertex v outside of X_i. Partition X_i into $\frac{|X_i|}{2}$ disjoint pairs of vertices. (Assume for simplicity that X_i is even, the other case can be treated in exactly the same way.) For each pair, there is at least one choice out of the 4 possibilities for the edges between this pair and v, for which these 3 vertices induce a path. Thus, when we randomly construct a great graph extending I, the probability that none of these sets of 3 vertices induces a path is less than $(\frac{3}{4})^{\frac{|X_i|}{2}} \leq (\frac{3}{4})^{\frac{n}{9}}$. Since there are fewer than n choices for v and only 4 choices for X_i, it follows that (c) holds for almost all great graphs extending I. This completes the proof of (ii) and our claim. $\square$

8 The proof of Claim 13

At the end of Section 4, we have already given some intuition about the proof.

Proof of Claim 13 Let $\mathcal{Y} = (Y_1, Y_2, Y_3, Y_4)$ be a partition of V_n such that $\left||Y_i| - \frac{n}{4}\right| \leq n^{1-\gamma}$, for every $i \in [4]$.

As before, let m be the number of pairs of vertices not contained in the same partition element, and note that there are exactly $2^{|Y_4|-1}$ choices for $G[Y_4]$ for a graph G 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.

As in Section 6, our approach is to show that, while there may be more choices for $G[Y_i]$ for mediocre graphs for which $\mathcal{Y}$ is a good partition, for each such choice we have much fewer than 2^m choices for mediocre string graphs extending it. However, we will have to sharpen our results, with respect to both the number of projections we need to consider and to how many string graphs extend each projection.

Let $\mathcal{F} = \mathcal{F}(\mathcal{Y})$ denote the set of mediocre string graphs which permit a $\mathcal{Y}$-good partition. For any $G \in \mathcal{F}$, let $P(G)$ denote the projection of G on the sets (Y_1, Y_2, Y_3, Y_4). That is, $P(G)$ is the disjoint union of the subgraphs $G[Y_1], G[Y_2], G[Y_3]$, and $G[Y_4]$.

We begin by exploiting the existence of the special set B of vertices to bound the number of choices for the edges of a good projection onto $\mathcal{Y}$, leaving a specified set $W \subseteq V_n$ of vertices.

Lemma 29 *Let $W \subseteq V_n$. The number of possibilities for the set of edges incident to the vertices in W in a projection $P(G)$, over all $G \in \mathcal{F}$, is $o(2^{\sqrt{\delta}n|W|+n(b+1)})$.*

Proof We can specify the edges of a projection of a graph in $\mathcal{F}$ incident to the vertices in W by first specifying the vertices in B and the edges out of each vertex of B. Next, for each $i \in [4]$ and each vertex $w \in W \cap Y_i$, we specify a vertex $v_w \in B$ for which the symmetric difference of $N(v_w) \cap Y_i$ and $N(w) \cap Y_i$ has at most δn elements, and we also specify the elements of this symmetric difference. So, there are at most $\binom{n}{b} 2^{nb} b^{|W|} \binom{n}{|W|} \binom{n}{\delta n}^{|W|}$ choices for the set of edges of $P(G)$ leaving W, over all $G \in \mathcal{F}$. We note that if δ is sufficiently small, then this is $o(2^{\sqrt{\delta}n|W|+n(b+1)})$. □

This immediately implies the following.

Corollary 30 *The number of projections on (Y_1, Y_2, Y_3, Y_4) of all graphs in $\mathcal{F}$ is $o(2^{\sqrt{\delta}n^{2-\gamma}+n(b+3)})$.*

Proof We can specify a projection $P(G)$, $G \in \mathcal{F}$, by specifying the vertices of $Z = Z_1 \cup \ldots \cup Z_4$ and the edges out of them into their corresponding parts, along with the partition of X_4 into two cliques. Applying Lemma 29, there are $o(2^n 2^{|X_4|-1} 2^{\sqrt{\delta}n|Z|+n(b+1)})$ choices for $P(G)$, over all $G \in \mathcal{F}$. Note that $o(2^{\sqrt{\delta}n|Z|}) = o(2^{\sqrt{\delta}n^{2-\gamma}})$, because $|Z| \leq n^{1-\gamma}$, by part (III) of Theorem 10. □

Our next step is to strengthen Lemma 23 by considering the situation where we fix not just the projection, but also the edges out of some small set, and bound the number of choices for the remaining edges between the partition elements which yield a string graph.

Lemma 31 *Let H be a non-string graph and let L_0, L_1, L_2, L_3, L_4 be a partition of $V(H)$, where some L_i may be empty. Let P be a projection on $\mathcal{Y}$. Fix a set W_0 of $|L_0|$ vertices of G, and a mapping f from W_0 to L_0.*

Then the number of string graphs G whose projection P on $\mathcal{Y}$ has the property that

> *(*) for every $j \in [4]$, there is a collection $\mathcal{W}_j$ of q disjoint sets of vertices of $Y_j - W_0$ such that, for every $W \in \mathcal{W}_j$, the mapping f extends to an isomorphism from $G[W_0 \cup W]$ to $H[L_0 \cup L_j]$,*

is $2^m \left(1 - \frac{1}{2^{\binom{|V(H)|}{2}}}\right)^{\frac{q^2}{4}}$.

Proof Let p be a prime between $\frac{q}{2}$ and q. We first choose the edges out of W_0. For those choices for which (*) holds, for every $j \in [4]$, let $J_0^j, \ldots, J_{p-1}^j$ be p elements of $\mathcal{W}_j$, whose existence is guaranteed by condition (*).

For $1 \le r, s \le p$, consider the 4-tuple $J_r^1, J_{r+s}^2, J_{r+2s}^3, J_{r+3s}^4$, where addition is modulo p. For each of the $p^2 \ge \frac{q^2}{4}$ such 4-tuples, there is a way of choosing edges between its elements so that the resulting extension is not a string graph. Thus, of the $c \le 2^{\binom{|V(H)|}{2}}$ choices for the edges between the elements of the 4-tuple, at most $c-1$ can occur in a string graph with the given projection and given choice of edges out of W_0. Furthermore, by the way in which we chose our 4-tuples, for any $x \in Y_i$ and $y \in Y_j$ with $i \ne j \in [4]$, there is at most one 4-tuple containing both x and y. The result follows. □

The next result illustrates the power of this lemma.

For a mediocre graph $G \in \mathcal{F}$, we call a set $T \subset V(G)$ *versatile* if for each $i \in [4]$ with $Y_i \cap T = \emptyset$, there is clique C_i in $G[Y_i]$ such that for all subsets $T' \subseteq T$, there are $\frac{n}{\log n}$ vertices of C_i that are adjacent to all elements of T' and to none of $T - T'$. We denote by P_3 a path of 3 vertices and by S_3 a stable set on 3 vertices. Let

$$\begin{aligned}\mathcal{F}_1 := \{G \in \mathcal{F} \mid\ &\text{there is } i \in [4], \text{ and versatile set } T_i \subset Y_i \text{ such that}\\ &|T_i| = 3 \text{ and } G[T_i] \text{ is isomorphic to } P_3 \text{ or } S_3\}.\end{aligned}$$

Lemma 32 $|\mathcal{F}_1| = o(2^m)$.

Proof For each choice of a good projection of a graph G onto $\mathcal{Y}$, each choice of a set T_i of 3 vertices contained in Y_i which induce an S_3 or P_3, and for each choice of the set of edges incident to T_i which make T_i versatile, we count the number of string graphs which have this projection and for which the set of edges incident to T_i is the specified set.

By Corollary 25 (c) or (d), there is a non-string graph H whose vertex set can be partitioned into 3 cliques of size at most 5, and a set L_0 such that $H[L_0]$ is isomorphic to $G[T_i]$. We label these 3 cliques as L_j for $j \in [4] - \{i\}$. We set $L_i = \emptyset$. Let f be an isomorphism from $G[T_i]$ to $H[L_0]$. We claim that for each $j \in [4]$, we can choose a family $\mathcal{W}_j$ of $\frac{n}{10\log n}$ disjoint cliques in Y_j of size $|L_j|$ with the property that for each $W \in \mathcal{W}_j$, the mapping f extends to an isomorphism from $G[T_i \cup W]$ to $H[L_0 \cup L_j]$. If $i = j$, then each of these cliques is an empty set. Otherwise, each element of $\mathcal{W}_j$ will be contained in C_j. We choose the vertices of the cliques in $\mathcal{W}_j$ one at a time, avoiding the vertices of C_j in cliques which have already been chosen. Since L_j and C_j are both cliques, to ensure that f extends to an isomorphism, we just need to make sure that our choice for the image of each vertex of L_j has the correct neighbourhood in T_i. By the definition of versatility, there are at least $\frac{n}{\log n}$ vertices of C_j with the desired neighbourhood, and, since we choose at most $\frac{5n}{10\log n}$ vertices from this set, one will not yet be chosen.

Applying Lemma 31 with $W_0 = T_i$, the number of choices for a string graph extending the projection, for which the set of edges incident to T_i have been specified, is at most $2^m\left(1 - \frac{1}{2^{\binom{18}{2}}}\right)^{\frac{n^2}{400(\log n)^2}}$. By Corollary 30, there are $2^{o\left(\frac{n^2}{(\log n)^2}\right)}$ choices for our projection. There are only 4 choices for i, at most n^3 choices for the vertices of T_i, and at most 2^{3n} choices for the edges incident to the vertices T_i. The desired result follows. □

We can prove an analogous result for sets of size at most 8 that intersect 2 parts of the partition. To state this result, we need a definition. A graph J' is called *extendible* if there is some non-string graph whose vertex set can be partitioned into 2 cliques of size at most 5 and a set inducing J'. Let

$$\mathcal{F}_2 := \{G \in \mathcal{F} \mid \text{ there are } i \neq j \in [4], T_i \subset Y_i \text{ and } T_j \subset Y_j \text{ such that } |T_i|, |T_j| \leq 4,\ T_i \cup T_j \text{ is versatile, and } G[T_i \cup T_j] \text{ is extendible}\}.$$

Lemma 33 $|\mathcal{F}_2| = o(2^m)$.

Proof For each choice of a partition $\mathcal{Y} = (Y_1, Y_2, Y_3, Y_4)$, distinct i and j in $[4]$, a projection of a graph G onto $\mathcal{Y}$, a set T_i of at most 4 vertices contained in Y_i, a set T_j of at most 4 vertices in Y_j, and the set of edges incident to $T_i \cup T_j$ which make $G[T_i \cup T_j]$ extendible and $T_i \cup T_j$ versatile, we count the number of string graphs which have this projection and for which the set of edges incident to $T_i \cup T_j$ is the specified set.

Since $G[T_i \cup T_j]$ is extendible, there is a non-string graph H whose vertex set can be partitioned into 2 cliques of size at most 5, and a set L_0 such that $H[L_0]$ is isomorphic to $G[T_i \cup T_j]$. We label these 2 cliques as L_k for $k \in [4] - \{i, j\}$. We set $L_i = L_j = \emptyset$. Let f be an isomorphism from $G[T_i \cup T_j]$ to $H[L_0]$. We claim that for each $k \in [4]$, we can choose a family $\mathcal{W}_k$ of $\frac{n}{10 \log n}$ cliques of size $|L_k|$ such that for each W in $\mathcal{W}_k$, f extends to an isomorphism from $G[T_i \cup T_j \cup W]$ to $H[L_0 \cup L_k]$. For $k \in \{i, j\}$, each of these cliques is the empty set. For $k \notin \{i, j\}$, each element of $\mathcal{W}_k$ will be contained in C_k. Since L_k and C_k are both cliques, to ensure that f extends to an isomorphism, we just need to ensure that our choice for the image of each vertex of L_k has the correct neighbourhood in $T_i \cup T_j$. By the definition of versatility, there are at least $\frac{n}{\log n}$ vertices of C_k with the desired neighbourhoods, and, since we choose at most $\frac{5n}{10 \log n}$ vertices from this set, one will not yet be chosen.

By Lemma 31, the number of choices for a string graph extending the projection for which the set of edges incident to $T_i \cup T_j$ have been specified, is at most $2^m \left(1 - \frac{1}{2^{\binom{18}{2}}}\right)^{\frac{n^2}{400(\log n)^2}}$. By Corollary 30, there are $2^{o\left(\frac{n^2}{(\log n)^2}\right)}$ choices for our projection. There are at most 6 choices for $\{i, j\}$, at most n^8 choices for the vertices of $T_i \cup T_j$, and at most 2^{8n} choices for the edges incident to the vertices $T_i \cup T_j$. The desired result follows. □

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 either

(a) is contained in some Y_i and induces S_3 or P_3, or

(b) is of size 8, contains exactly 4 vertices from each of 2 distinct partition elements, and induces an extendible subgraph.

Note that every element of $\mathcal{W}$ must intersect $Z = Z_1 \cup \ldots \cup Z_4$, hence $|\mathcal{W}| \leq |Z|$. Set $W^* = \cup_{W \in \mathcal{W}} W$, and let $Y_i' = Y_i - W^*$. Note that $|W^*| \leq 8|\mathcal{W}| \leq 8|Z|$ and that for every i, Y_i' has more than $\frac{n}{5}$ vertices and $G[Y_i']$ is the disjoint union of two cliques, by the maximality of $\mathcal{W}$. In what follows, we focus on graphs in $\mathcal{F} - (\mathcal{F}_1 \cup \mathcal{F}_2)$. Hence, the edges of $G - P(G)$ must be chosen in such a way that no set $S \in \mathcal{W}$ is versatile. Let

$$\mathcal{F}_3 := \{G \in \mathcal{F} - (\mathcal{F}_1 \cup \mathcal{F}_2) \mid |\mathcal{W}| \geq C \text{ for } C = 10^6 b\}.$$

Lemma 34 $|\mathcal{F}_3| = o(2^m)$.

Proof For every choice of a projection onto $\mathcal{Y}$ and a collection $\mathcal{W}$ of at most $n^{1-\gamma}$ disjoint sets of vertices of size at most 8, we count the number of graphs G in $\mathcal{F}_3$, for which this projection is $P(G)$ such that we can choose $\mathcal{W}_G$ to be $\mathcal{W}$.

Since each Y'_k is the disjoint union of 2 cliques, each Y'_k, $k \in [4]$, contains a clique C_k with at least $\frac{n}{10}$ vertices. The graph G was chosen to be outside $\mathcal{F}_1 \cup \mathcal{F}_2$, for any set $D \in \mathcal{W}$, there is a subset $D' \subseteq D$ and a $j \in [4]$ with $Y_j \cap D = \emptyset$ such that there are fewer than $\frac{n}{\log n}$ vertices of C_j which are adjacent to all of D' and none of $D - D'$. This implies that the number of choices for the edges of $E(G) - E(P(G))$ with one endpoint in D is $o(2^{\frac{3n|D|}{4} - \frac{n}{10000}})$. Indeed, if there were no restrictions, the number of choices for the edges of $E(G) - E(P(G))$ would be at most $2^{\frac{3n|D|}{4} + o(n)}$. On the other hand, for any choice $D' \subseteq D$, in an unrestricted choice we expect at least $|C_j|/2^{|D|} > n/2^8$ vertices of C_j to have neighbourhood D' on D. Applying the Chernoff bounds to the probability that we only get $\frac{n}{\log n}$ such vertices in C_j yields the claimed bound on the number of choices for the edges from D.

Given a choice of $\mathcal{W}$, the number of choices for graphs on $Y'_1, \ldots, Y'_4$ is less than 2^n. Applying Lemma 29 to W^*, we obtain that the number of choices for the edges of $P(G)$ which have exactly one endpoint in W^*, is $O(2^{n(b+1)+\sqrt{\delta}|W^*|n})$. There are fewer than $2^n|W^*|^{|W^*|}2^{|W^*|^2}$ choices for W^*, a partition of it yielding $\mathcal{W}$, and the edges with both endpoints in W^*. Combining these facts with the result of the previous paragraph, we get that the number of graphs in $\mathcal{F} - (\mathcal{F}_1 \cup \mathcal{F}_2)$ is at most $O(2^{n(b+1)+(\sqrt{\delta}n+\log|W^*|+|W^*|)|W^*|} \cdot 2^m \cdot 2^{-\frac{|\mathcal{W}|n}{10000}})$. We can and do choose δ small enough so that if $|W^*| \geq C$, then the above is $o(2^m)$. □

It remains to count the number of graphs in $\mathcal{F} - (\mathcal{F}_1 \cup \mathcal{F}_2 \cup \mathcal{F}_3)$. We begin with the following.

Lemma 35 *The number of projections onto $\mathcal{Y}$ which extend to a string graph in $\mathcal{F} - (\mathcal{F}_1 \cup \mathcal{F}_2 \cup \mathcal{F}_3)$ is $2^{O(n)}$*

Proof We are counting the number of choices of the good projection $P(G)$ onto $\mathcal{Y}$, over all mediocre string graphs $G \in \mathcal{F} - (\mathcal{F}_1 \cup \mathcal{F}_2 \cup \mathcal{F}_3)$. That is, over all G for which we can choose $\mathcal{W}$ with $|\mathcal{W}| \leq C$ such that no element of $\mathcal{W}$ is versatile. Recall that, by the definition of $\mathcal{W}$, each element in $\mathcal{W}$ is of size at most 8 and, hence, $|W^*| \leq 8C$. Also recall that, for $Y'_i = Y_i - W^*$, by the maximality of $\mathcal{W}$, each $G[Y'_i]$, $i \in [4]$, is the disjoint union of 2 cliques.

We *claim* that the number of projections of this type is at most $\binom{n}{8C}(8C)^{8C}2^{8Cn+1} = 2^{O(n)}$. Indeed, there are at most $\binom{n}{8C}(8C)^{8C}$ ways to choose the vertices in $\mathcal{W}$ and partition them into sets. There are at most 2^{8Cn} ways to choose the neighborhoods of the vertices in W^*. Finally, there are at most 2^n ways to partition each Y'_i into 2 cliques. Thus, our claim is true and the lemma holds. □

Let

$$\mathcal{F}_4 := \{G \in \mathcal{F} - (\mathcal{F}_1 \cup \mathcal{F}_2 \cup \mathcal{F}_3) \mid \text{ there are } i \neq j \in [4] \text{ such that both } Y'_i \text{ and } Y'_j \text{ contain two components larger than } n^{2/3}\}.$$

Lemma 36 $|\mathcal{F}_4| = o(2^m)$.

Proof To prove the lemma, we consider one of the $2^{O(n)}$ projections which extends to a graph in $\mathcal{F}_4$ and count how many string graphs it extends to.

By Corollary 25 (e), there is a non-string graph H whose vertex set can be partitioned into (L_1, L_2, L_3, L_4) so that for $k \in [4] - \{i,j\}$, L_k is a clique of size at most 5 and $G[L_i]$ and $G[L_j]$ are the disjoint union of a vertex and a clique of size 3.

For $k \in \{i,j\}$, we can find a family $\mathcal{T}^k$ of $\frac{n^{2/3}}{3}$ disjoint sets in Y'_i, each inducing the disjoint union of a triangle and a vertex. For $k \in [4] - \{i,j\}$, we can find a family $\mathcal{T}^k$ of $\frac{n^{2/3}}{3}$ disjoint sets in Y'_i each of which is a clique of size $|L_k|$. Applying Lemma 31 with L_0 empty, we see that the number of choices for the edges between the partition elements which extend this projection to a string graph is at most $2^m(1 - \frac{1}{2^{\binom{18}{2}}})^{\frac{n^{4/3}}{36}}$. Since, by Lemma 35, the number of choices for the projection is $2^{O(n)}$, the desired result follows. □

For each vertex $v \in W^*$, we define the *rank of v with respect to a partition element* Y_i as $\max\{\min(|N(v) \cap K|, |K - N(v)|) \mid K \text{ is a component of } Y'_i\}$. We use rank$(v)$ to denote the minimum of these ranks over the partition elements. We say that v is *extreme* on Y_i if its rank with respect to Y_i is less than $n^{\frac{2}{3}}$. Let

$$\mathcal{F}_5 := \{G \in \mathcal{F} - (\mathcal{F}_1 \cup \mathcal{F}_2 \cup \mathcal{F}_3 \cup \mathcal{F}_4) \mid \text{ there is } v \in W^* \text{ such that}$$
$$v \text{ is not extreme on any partition element}\}.$$

Lemma 37 *$|\mathcal{F}_5| = o(2^m)$.*

Proof For each projection of a graph $G \in \mathcal{F}_5$ onto $\mathcal{Y}$, we count the number of choices for $E(G) - E(P(G))$ over all string graphs G with this projection. Consider a vertex v in W^* which is not extreme on any partition element. By Corollary 25 (b), there is a non-string graph H such that for some vertex $w \in V(H)$, $H - w$ can be partitioned into four cliques (L_1, L_2, L_3, L_4) each of size at most 5. By our choice of v, for $k \in [4]$ we can find a family $\mathcal{W}^k$ of $\frac{n^{\frac{2}{3}}}{5}$ disjoint cliques in Y'_k, each of which contains $|N(w) \cap L_k|$ neighbours of v. For any 4-tuple consisting of an element from each $\mathcal{T}_k$, there is a choice of edges between the elements of the 4-tuple which implies that H is a subgraph of G. Applying Lemma 31 with $L_0 = \{w\}$, we see that the number of choices for the edges between the partition elements which extend this projection to a string graph is at most $2^m(1 - \frac{1}{2^{\binom{21}{2}}})^{\frac{n^{4/3}}{100}}$. By Lemma 35, the number of choices for the projection is $2^{O(n)}$ and the desired result follows. □

It remains to analyze the case when every vertex of W^* is extreme on some partition element. We consider a new partition $\mathcal{Y}^* = (Y_1^*, \ldots, Y_4^*)$ obtained from $\mathcal{Y}$ by moving each element of W^* to a set Y_i with respect to which its rank is equal to rank(v). Let W_i^*, $i \in [4]$, be the set of vertices in $W^* \cap Y_i^*$. Let

$$\mathcal{F}_6 := \{G \in \mathcal{F} - \cup_{i=1}^5 \mathcal{F}_i \mid \text{ there is } v \in W^* \text{ such that for some } j \text{ for which}$$
$$v \notin Y_j^* \text{ the rank of } v \text{ with respect to } Y_k \text{ is less than } \frac{n}{\log n}\}.$$

Lemma 38 *$|\mathcal{F}_6| = o(2^m)$.*

Proof First, we specify the number of choices for $\mathcal{Y}^*, \mathcal{W}^*$, and a projection onto $\mathcal{Y}^*$ which can be extended to a graph in $\mathcal{F}_6$.

We note that we can specify the new partition $\mathcal{Y}^*$ and the set W^* by fixing a choice for W^*, and a choice of for each element u of W^* of a partition element with respect to which u has rank rank(u). This is $O(4^{8C} \cdot \binom{n}{|W^*|}) = O(\binom{n}{|W^*|})$ choices.

Since $Y_i^* - W^* = Y_i - W^*$ and we are not considering graphs in $\mathcal{F}_4$, there are at most $2^{\max\{|Y_i|, 1\le i\le 4\}}\binom{n}{n^{2/3}}^3 = 2^{|Y_4|+O(n^{1-\gamma}+n^{3/4})}$ choices for the edges of such a mediocre string graph which lie within the $Y_i^* - W^*$. For every vertex $u \in Y_i^* \cap W^*$, to specify the edges from u to $Y_i^* - W^*$, we need to specify for each of the at most 2 components K of $Y_i^* - W^*$, the smaller of the sets $K \cap N(u)$ and $K - N(u)$ (with ties broken arbitrarily) and then whether this set is $K \cap N(u)$ or $K - N(u)$. There are at most $4\binom{n}{\text{rank}(u)}^2 \le 4\binom{n}{n^{2/3}}^2$ such choices. Since there are $O(1)$ choices for the edges within W^*, we see that there are $2^{|Y_4|+o(n)}$ choices for the projection of such a G on one of $2^{o(n)}$ possible $\mathcal{Y}^*$.

Now, we count the number of graphs G, the projection of which onto one of these Y^* is one of the given projections, for which there is a vertex v of W^* such that for some Y_j with $v \notin Y_j^*$, the rank of v on Y_j is less than $\frac{n}{\log n}$. To specify the edges from v to $Y_j^* - W^*$, we need to specify for each of the at most 2 components K of $Y_j^* - W^*$, the smaller of the sets $K \cap N(v)$ and $K - N(v)$ (with ties broken arbitrarily) and then whether this set is $K \cap N(v)$ or $K - N(v)$. Hence, there are at most $4\binom{n}{n/\sqrt{\log n}}^2 = 2^{o(n)}$ choices for the edges from v to Y_j^* which make v extreme on Y_j. Hence, letting m' denote the number of pairs of vertices lying in different elements of $\mathcal{Y}^*$, we have that the number of such G is $2^{m'+|Y_4|+o(n)-n/4}$.

As the size of each Y_i differs from $\frac{n}{4}$ by at most $n^{1-\gamma}$, and we move only a constant number of vertices, the difference between m and m' is $O(n^{1-\gamma})$. So, the number of choices for G in $\mathcal{F}_6$ is $o(2^{m+|Y_4|})$. $\square$

Next, we focus on graphs $G \in \mathcal{F} - \cup_{i=1}^{6}\mathcal{F}_i$. Let $P(G)$ be a projection on $\mathcal{Y}$. Let v be a vertex of W^* maximizing rank(v) and let i be the integer for which $v \in Y_i^*$. We define rank$'(v)$ as rank(v), unless rank$(v) = 0$. If rank$(v) = 0$ and we can choose v to be in a P_3 or S_3 of $G[Y_i^*]$, then we set rank$'(v) = 1$. Let

$$\mathcal{F}_7 := \{G \in \mathcal{F} - \cup_{i=1}^{6}\mathcal{F}_i \mid \text{there is a vertex } v \in W^* \text{ with rank}'(v) > 0\}.$$

Lemma 39 $|\mathcal{F}_7| = o(2^{m+|Y_4|})$.

Proof Let $G \in \mathcal{F}_7$ and let $P(G)$ be its projection on $\mathcal{Y}$. Let v be a vertex of W^* maximizing rank$'(v)$ and let $i \in [4]$ be such that $v \in Y_i^*$. Assume that rank$'(v) > 0$. If rank$'(v) > 1$, then we choose a set of rank'(v) different P_3s, all containing v, but otherwise disjoint and contained in Y_i. Let $\mathcal{T}^i$ be this set of P_3s, and denote its elements by $T_1^i, \ldots, T^i_{\text{rank}'(v)}$. If rank$'(v) = 1$, we choose an S_3 or P_3 containing v to be T_1^i.

By Corollary 25 (c) or (d), there is a non-string graph H whose vertex set can be partitioned into (L_1, L_2, L_3, L_4), where $H[L_i] = G[T_1^i]$ and for $k \in [4] - \{i\}$, L_k is a clique of size at most 5. For each $1 \le q \le \text{rank}'(v)$, let f_q be an isomorphism from T_q^i to $H[L_i]$ such that the image of v is the same under all f_q. Let v' be this image. For each $j \in [4] - i$, let n_j be the number of vertices of L_j adjacent to v'. As G is not in $\mathcal{F}_6$, for every $j \in [4] - i$, the rank of v is at least $\frac{n}{\log n}$ on Y_j. Thus, we can choose a set $\mathcal{C}^j$ of $q = \frac{n}{10 \log n}$ disjoint cliques $C_1^j, \ldots, C_q^j$ in $G[Y_j - W^*]$, each of size $|L_j|$, such that v has exactly n_j neighbours in each of them. In other words, there is an isomorphism f from $G[\{v\} \cup C^j]$ to $H[\{v'\} \cup L_j]$ which maps v' to v, for each $j \ge 2$ and $C^j \in \mathcal{C}^j$.

We count first the extensions of such a projection onto $\mathcal{Y}$ to a string graph for which there is some T_q^i with the property that for all j in $[4] - \{i\}$, there are more than $\frac{n}{(\log n)^2}$ values of k for which f_q extends to an isomorphism from $G[T_q^i \cup C_k^j]$ to $H[L_i \cup L_j]$. In this case, by Lemma 31, there are at most $2^{m'}(1 - \frac{1}{2^{\binom{18}{2}}})^{\Omega(n^2/(\log n)^4)}$ such string graphs extending a given projection.

We count next the number of extensions for which there is no T_q^i with the above property. The probability that f_q extends to an isomorphism from $H[L_i \cup L_j]$ to $G[T_q^i \cup C_k^j]$ for some $j \in [4] - \{i\}$ if we choose the edges between the $Y_i^* - W^*$ randomly, is at least 2^{-15} and these probabilities are all independent. So, an application of the Chernoff bounds tells us that the probability that for one specific T_q^i there is some j for which there are no $n/(\log n)^2$ different values of k for which f_q extends to an isomorphism from $G[T_q^i \cup C_k^j]$ to $H[L_i \cup L_j]$, is $2^{-\Omega(n/\log n)}$. Thus, the number of extensions for which there is no such T_q^i is $2^{m' - \Omega(\frac{rank'(v)n}{\log n})}$.

The number of choices for W^* and Y^* is at most $n^{|W^*|}4^{|W^*|} = 2^{O(\log n)}$. The number of choices for the edges of our projection is $O(2^{|Y_4| + n^{1-\gamma}} \binom{n}{n^{2/3}}^3 \binom{n}{rank'(v)}^{2|W^*|})$. Using that W^* is a constant and $\text{rank}'(v)$ is at most $n^{2/3}$, this is $O(2^{|Y_4| + n^{1-\gamma} + n^{3/4}})$. Since $m' = m + O(n^{1-\gamma})$, we obtain that the number of string graphs for which $\text{rank}'(v) \neq 0$ is $o(2^{m+|Y_4|})$. □

To complete the proof of Theorem 1, we need to show the following.

Lemma 40 *$|\mathcal{F} - \cup_{i=1}^7 \mathcal{F}_i| = o(2^{m+|Y_4|})$.*

Proof Consider $G \in \mathcal{F} - \cup_{i=1}^7 \mathcal{F}_i$ and let $P(G)$ be its projection on $\mathcal{Y}$. As noted before, there are $2^{O(\log n)}$ choices for W^* and $\mathcal{Y}^*$, and $O(1)$ choices for the edges within W^*. Let v be a vertex of W^* maximizing $\text{rank}'(v)$. Now, we have $\text{rank}'(v) = 0$. This means that every Y_i^*, $i \in [4]$, is the disjoint union of 2 cliques. Since $G \notin \mathcal{F}_4$, and $Y_i^* - W^* = Y_i - W^*$, letting max denote the maximum over all Y_i^* of the size of the smallest component of Y_i^*, there are fewer than $\binom{n}{max}\binom{n}{n^{2/3}}^3$ choices for the partition of every $Y_i^* - W^*$ into 2 cliques. Given such a partition for each i, there are $4^{|W^*|} = O(1)$ choices for the edges out of W^* in the projection of G onto Y^*. Hence, the number of choices for such a projection with $max \leq \frac{n}{1000}$ is $2^{O(\log n)}\binom{n}{\frac{n}{1000}}\binom{n}{n^{2/3}}^3 = o(2^{|Y_4| - \frac{n}{\log n}})$. There are at most $2^{m'} = 2^{m+O(n^{1-\gamma})}$ string graphs extending each such projection and, hence, $o(2^{m+|Y_4|})$ such string graphs in total.

Therefore, we need only count the number of graphs G in $\mathcal{F} - \cup_{i=1}^7 \mathcal{F}_i$ for which there is some i such that Y_i^* has two components of size exceeding $\frac{n}{1000}$. Since $G \notin \mathcal{F}_4$, for all $i \neq j$, the smaller component of Y_i^* has at most $n^{2/3}$ vertices. The number of choices for $\mathcal{Y}^*, W^*$ and for the projection of such a G onto $\mathcal{Y}^*$ is $O(2^{O(\log n) + |Y_4| + O(n^{1-\gamma}) + O(n^{3/4})})$. If for all $j \neq i$, Y_j^* is a clique, then G is a great graph which has a great partition obtained by re-indexing the elements of $\mathcal{Y}$. So, we can assume that this is not the case and find a subgraph D which induces the disjoint union of a vertex and a clique of size 3 contained in Y_j^*, for some $j \neq i$.

By Corollary 25 (e), there is a non-string graph H whose vertex set can be partitioned into (L_1, L_2, L_3, L_4), where $H[L_i]$ and $H[L_j]$ are disjoint unions of a vertex and a clique of size at most 3, and for $k \in [4] - \{i, j\}$, L_k is a clique of size at most 5. We can choose a subgraph $D' \subseteq D$ such that there is an isomorphism f from $G[D']$ to $H[L_j]$. For every $k \in [4] - \{j\}$, we can choose a set $\mathcal{C}^k$ of $\frac{n}{4000}$ disjoint sets $C_1^k, \ldots, C_p^k$ in $G[Y_j - W^*]$, each of which induces a subgraph isomorphic to $H[L_j]$ (for $k = i$, we need to exploit our lower bound on the size of the smaller component of Y_i^*).

We count first the extensions of such a projection onto $\mathcal{Y}$, where for every $k \in [4] - \{j\}$, there are more than $\frac{n}{\log n}$ values of ℓ for which f extends to an isomorphism from $G[D' \cup C_\ell^k]$ to $H[L_i \cup L_k]$. By Lemma 31, there are at most $2^{m'}(1 - \frac{1}{2^{\binom{18}{2}}})^{\Omega(n^2/\log^2 n)}$ string graphs extending a given such projection.

We count next the number of extensions with the property that for some $k \in [4] - \{i\}$, there are fewer than $\frac{n}{\log n}$ values of ℓ for which f extends to an isomorphism from $G[D' \cup C_\ell^k]$ to $H[L_i \cup L_k]$. The probability that f extends to an isomorphism from $G[D' \cup C^j_{\ell'}]$ to $H[L_i \cup L_j]$ for some $j \in [4] - \{i\}$, if we choose the edges between the Y^*_i randomly, is at least 2^{-15}, and these probabilities are all independent. So, applying the Chernoff bounds, we obtain that the probability that there is some j such that there are no $\frac{n}{\log n}$ values of ℓ' for which f extends to an isomorphism from $H[L_i \cup L_j]$ to $G[D' \cup C^j_{\ell'}]$, is $2^{-\Omega(n)}$.

Since the total number of projections we are considering, over all choices of $\mathcal{Y}^*$ and W^*, is $O(2^{O(\log n)+|Y_4|+O(n^{1-\gamma})+n^{3/4}})$ and $m' = m + O(n^{1-\gamma})$, we conclude that the total number of string graphs extending these projections is $o(2^{m+|Y_4|})$, and we are done. □

This completes the proof of Theorem 1. □

Acknowledgement

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, and all three referees for their careful work and valuable remarks.

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. Kühn, D. Osthus, T. Townsend. Forbidding induced even cycles in a graph: typical structure and counting, J. Comb. Theory, Ser. B **131** 170-219 (2018)

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

[KuKr98] J. Kratochvíl and A. Kuběna. On intersection representations of co-planar graphs, *Discrete Math.* **178** (1998) 251–255.

[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.

[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.