DiVA
http://www.diva-portal.org

This is the published version of a paper published in *The Electronic Journal of Combinatorics*.

Citation for the original published paper (version of record):

Falgas-Ravry, V., Markström, K. (2016)
Random subcube intersection graphs I: cliques and covering.
The Electronic Journal of Combinatorics, 23(3): P3.43

Access to the published version may require subscription.

N.B. When citing this work, cite the original published paper.

Permanent link to this version:
http://urn.kb.se/resolve?urn=urn:nbn:se:umu:diva-127244

Random subcube intersection graphs I: cliques and covering

Victor Falgas-Ravry* Klas Markström†

Institutionen för matematik och matematisk statistik
Umeå Universitet
Umeå, Sweden

{victor.falgas-ravry,klas.markstrom}@umu.se

Submitted: Aug 7, 2015; Accepted: Aug 26, 2016; Published: Sep 2, 2016
Mathematics Subject Classifications: 05C80, 05C62

Abstract

We study *random subcube intersection graphs*, that is, graphs obtained by selecting a random collection of subcubes of a fixed hypercube Q_d to serve as the vertices of the graph, and setting an edge between a pair of subcubes if their intersection is non-empty. Our motivation for considering such graphs is to model 'random compatibility' between vertices in a large network.

For both of the models considered in this paper, we determine the thresholds for covering the underlying hypercube Q_d and for the appearance of s-cliques. In addition we pose a number of open problems.

1 Introduction

In this paper we introduce and study two models of *random subcube intersection graphs*. These are random graph models obtained by (i) selecting a random collection of subcubes of a fixed hypercube Q_d, to serve as the vertices of the graph, and (ii) setting an edge between a pair of subcubes if their intersection is non-empty. Our basic motivation for studying these random graphs is that they give a model for 'random compatibility' between vertices. Before we make our models mathematically precise let us consider some examples of the applications we have in mind.

A first example is the random k-SAT problem, which has attracted the attention of both physicists [27] and mathematicians [1] for many decades. In this problem we have a set of n Boolean variables and some number m of Boolean clauses on k variables are chosen at random. Each clause forbids exactly one of the 2^k possible assignments to the k

*Research supported by a grant from the Kempe foundation.
†Research supported by a grant from the Swedish Research Council.

variables in the clause and the question of interest is whether there exists an assignment to the n variables which is compatible with all the clauses. It is known that there is a sharp threshold [22] for satisfiability with respect to the number m of clauses chosen and that for large k [15] this threshold is located at approximately $m = n2^k \ln(k)$. It is conjectured that for all fixed k there exists a constant c_k such that the satisfiability threshold is asymptotically $c_k n$ (a proof of this conjecture for all sufficiently large values of k has been announced recently [17]).

The random k-SAT problem is a problem about random subcubes. Given a clause C the assignments which are incompatible with C are given by the subcube where the k variables in C are assigned the values forbidden by C. This is an $(n-k)$-dimensional subcube of the n-dimensional cube of all possible assignments, and a collection of clauses is unsatisfiable if and only if the union of the corresponding subcubes contains all vertices of the n-cube. The random k-SAT problem is thus equivalent to finding the threshold for covering all the vertices of a cube by a collection of random subcubes. This example also suggests that a mathematical analysis of the covering problem will be harder for subcubes than for e.g. the usual independent random intersection graph models (where it is analogous to the classical coupon collector problem).

A second example of applications we have in mind comes from social choice theory. Suppose we have a society V which is faced with d political issues, each which can be resolved in a binary fashion. We represent the two policies possible on each issue by $0, 1$, and the family of all possible sets of policies by a d-dimensional hypercube Q_d.

Individual members of the society may have fixed views on some issues, but may be undecided or indifferent on others. We can thus associate to each citizen $v \in V$ a subcube of acceptable policies $f(v)$ in a natural way. The subcube intersection graph G arising from (V, Q_d, f) then represents political agreement within the society: uv is an edge of G if and only if the citizens u and v can agree on a mutually acceptable set of policies.

A key characteristic of subcubes is that they possess the *Helly property*: if we have s subcubes $f(v_1), f(v_2), \dots f(v_s)$ of Q_d which are pairwise intersecting, then their total intersection $\bigcap_{i=1}^{s} f(v_i)$ is non-empty (this is an easy observation, already made in [25]). A consequence of this fact is that in the model for political agreement described above, s-cliques represent s-sets of citizens able to agree on a mutually acceptable set of policies and, say, unite their forces to promote a common political platform. This motivates our study of the clique number in (random) subcube intersection graphs. For a more detailed discussion of the role of subcube intersection graphs in the context of social choice theory and voting problems, see [31].

There are many examples of compatibility graphs naturally modeled by subcube intersection graphs. Some closely resemble the one above: the work of matrimonial agencies or the assignment of room-mates in the first year at university, for instance, naturally lead to the study of such compatibility graphs. Another class of examples can be found in the medical sciences. For kidney or blood donations, several parameters must be taken into account to determine whether a potential donor–receiver pair is compatible. Large random subcube intersection graphs provide a way of modeling these compatibility relations over a large pool of donors and receiver, and of identifying efficient matching schemes.

1.1 The models

Let us now describe our models more precisely. We begin with some basic definitions and notation.

Definition 1.1 (Intersection graphs)**.** A *feature system* is a triple (V, Ω, f), where V is a set of vertices, Ω is a set of *features*, and f is a function mapping vertices in V to subsets of Ω. Given a vertex $v \in V$, we call $f(v) \subseteq \Omega$ its *feature set*. We construct a graph G on the vertex-set V from a feature system (V, Ω, f) by placing an edge between $u, v \in V$ if their feature sets $f(u), f(v)$ have non-empty intersection. We call G the *intersection graph* of the feature system (V, Ω, f).

Intersection graphs are well-studied objects with many applications— see the monograph of McKee and McMorris [30]. In this paper we shall study intersection graphs where Ω and the feature sets $\{f(v)\colon v \in V\}$ have some additional structure. Namely, Ω shall be a high-dimensional hypercube Q_d and the feature sets will consist of subcubes of Q_d.

Definition 1.2 (Hypercubes and subcubes)**.** The *d-dimensional hypercube* is the set $Q_d = \{0,1\}^d$. A *k-dimensional subcube* of Q_d is a subset obtained by fixing $(d-k)$-coordinates and letting the remaining k vary freely. We may regard subcubes of Q_d as elements of $\{0,1,\star\}^d$, where $\star$ coordinates are free and the $0,1$ coordinates are fixed.

We shall define two models of random subcube intersection graphs. Both of these are obtained by randomly assigning to each vertex $v \in V$ a feature subcube $f(v)$ of Q_d and then building the resulting intersection graph.

Definition 1.3 (Uniform model)**.** Let V be a set of vertices. Fix $k, d \in \mathbb{N}$ with $k \leqslant d$. For each $v \in V$ independently select a k-dimensional subcube $f(v)$ of Q_d uniformly at random, and set an edge between $u, v \in V$ if $f(u) \cap f(v) \neq \varnothing$. Denote the resulting random subcube intersection graph by $G_{V,d,k}$.

Definition 1.4 (Binomial model)**.** Let V be a set of vertices. Fix $d \in \mathbb{N}$ and $p \in [0,1]$. For each $v \in V$ independently select a subcube $f(v) \in \{0,1,\star\}^d$ at random by setting $(f(v))_i = \star$ with probability p and $(f(v))_i = 0,1$ each with probability $\frac{1-p}{2}$ independently for each coordinate $i \in \{1, \dots d\}$ (we refer to such a subcube as a *binomial random subcube*). Denote the resulting random subcube intersection graph by $G_{V,d,p}$.

Remark 1.5. We may view $G_{V,d,p}$ as the intersection of d independent copies of $G_{V,1,p}$ on a common vertex-set V. Indeed an edge uv of $G_{V,d,p}$ is present if and only if for each of the d dimensions of Q_d we have that $f(u)$ and $f(v)$, viewed as vectors, are identical or at least one of them is $\star$. The graph $G_{V,1,p}$ is itself rather easy to visualise: we first randomly colour the vertices in V with colours from $\{0,1,\star\}$, and then remove from the complete graph on V all edges between vertices in colour 0 and vertices in colour 1.

1.2 Degree distribution, edge-density and relation to other models of random graphs

Our two models of random subcube intersection graphs bear some resemblance to previous random graph models. To give the reader some early intuition into the nature of random subcube intersection graphs, we invite her to consider the degree distributions and edge-densities found in them, and to contrast them with models of random graphs with similar degree distributions and edge-densities.

Let us first note that in order to get an random model which is both structurally interesting and amenable to asymptotic analysis we typically consider the case where $d \to \infty$, and the other parameters are functions of d.

The degree of a given vertex in the uniform model $G_{V,d,k}$ is a binomial random variable with parameters $|V|-1$ and q, where q is the probability that two uniformly chosen k-dimensional subcubes of Q_d meet. If $k = k(d) = \lfloor \alpha d \rfloor$ for some fixed $\alpha \in (0,1)$, then one can show $q = q(\alpha) = e^{-g(\alpha)d + o(d)}$, where

$$\begin{aligned} g(\alpha) = &2\log\left(\alpha^{\alpha}(1-\alpha)^{1-\alpha}\right) - \left(\sqrt{(1-\alpha)^2+\alpha^2} - 1 + \alpha\right)\log\left(\sqrt{(1-\alpha)^2+\alpha^2} - 1 + \alpha\right) \\ &- 2\left(1 - \sqrt{(1-\alpha)^2+\alpha^2}\right)\log\left(1 - \sqrt{(1-\alpha)^2+\alpha^2}\right) \\ &- \left(\sqrt{(1-\alpha)^2+\alpha^2} - \alpha\right)\log\left(2\sqrt{(1-\alpha)^2+\alpha^2} - 2\alpha\right). \end{aligned}$$

This expression is not, however, terribly instructive.

The quantity q is also the edge-density of $G_{V,d,k}$. When $|V| = n$, the appropriate random graph to compare and contrast it with is thus an Erdős–Rényi random graph $G(n,q)$ with edge probability q. However $G_{V,d,k}$ displays some significant *clustering*: our results can be used to show for instance that dependencies between the edges cause triangles to appear well before we see a linear number of edges, in contrast to the Erdős–Rényi model $G(n,q)$.

The edge-density of the binomial model $G_{V,d,p}$ is easy to compute: it is exactly $\left(1 - \frac{(1-p)^2}{2}\right)^d = e^{-d\log\left(\frac{2}{1+2p-p^2}\right)}$. The degree distribution of $G_{V,d,p}$ is more complicated, however. Increasing the dimension of a subcube by 1 doubles its volume inside Q_d, so that larger subcubes expect much larger degrees. The number of feature subcubes from our graph met by a fixed subcube of dimension αd is a binomial random variable with parameters $|V|-1$ and $\left(\frac{1+p}{2}\right)^{(1-\alpha)d}$. The number of vertices in V whose feature subcubes have dimension αd is itself a binomial random variable with parameters $|V|$ and $\binom{d}{\alpha d} p^{\alpha d}(1-p)^{(1-\alpha)d}$. As in this paper we will typically be interested in the case where d is large and V has size exponential in d, we will expect to see some feature subcubes with dimension much larger or much smaller than pd. This will have a noticeable effect on the properties of the graph $G_{V,d,p}$.

Among the random graph models studied in the literature, $G_{V,d,p}$ in many ways resemble the multi-type inhomogeneous random graphs studied in [10], which also have vertices of several different types and differing edge probabilities, though we should point out there are significant differences. First of all some 'types' corresponding to vertices

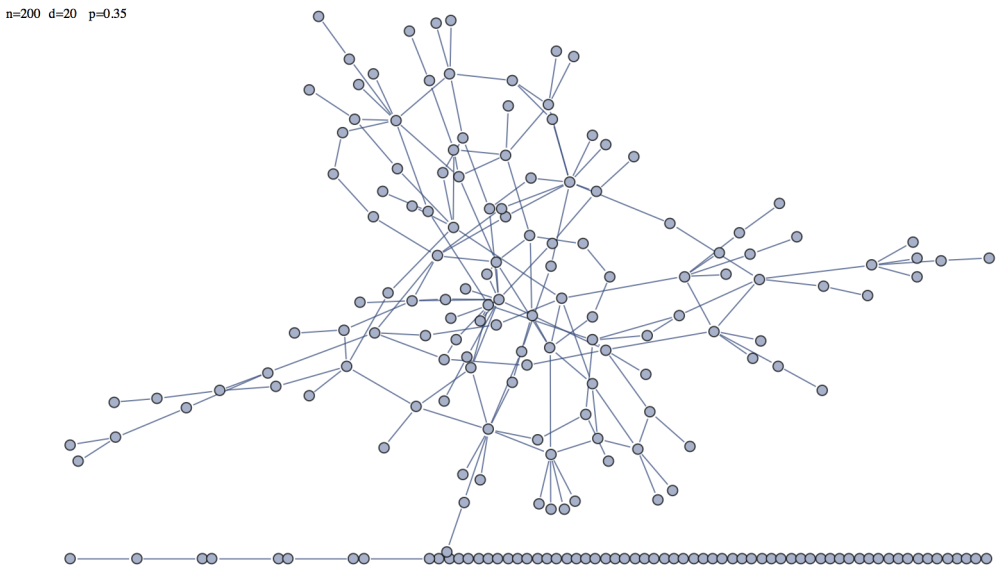

Figure 1: An example of the binomial random subcube intersection graph model $G_{V,d,p}$ with $|V| = 200$, $d = 20$ and $p = 0.35$. The row of vertices at the bottom right are all isolated.

with unusually large or unusually small feature subcubes will have only a sublinear (and random) number of representatives. Secondly, the binomial model shares the clustering behaviour of the uniform model (see Remark 2.3), differentiating it from the models considered in [10]. We note that a further general model for inhomogeneous random graphs with clustering was introduced by Bollobás, Janson and Riordan in [11], for which this second point does not apply.

Finally, let us mention the standard models of *random intersection graphs*. Write $[m]$ for the discrete interval $\{1, 2, \dots m\}$. In the *binomial random intersection graph model* $\mathcal{G}(V, [m], p)$, each vertex $v \in V$ is independently assigned a random feature set $f(v) \subseteq [m]$. This feature set is obtained by including $j \in [m]$ into $f(v)$ with probability p and leaving it out otherwise independently at random for each feature $j \in [m]$. Edges are then added between all pairs of vertices $u, v \in V$ with $f(u) \cap f(v) \neq \varnothing$ to obtain a random intersection graph on V. A variant on this model is to choose feature sets $f(v)$ uniformly at random from the k-subsets of $[m]$; this yields the *uniform random intersection graph model* $\mathcal{G}(V, [m], k)$.

While these two random intersection graph models bear some resemblance (in terms of clustering, for example) to our random subcube intersection graph models, there are also some significant differences due to the underlying structure of our feature sets. Let us note amongst other things that subsets of $[m]$ do not have the Helly property, and that the effects on the degree of increasing the size of a feature set by 1 in a binomial random intersection graph are far less dramatic than the effects of increasing the dimension of a feature subcube by 1 in a binomial random subcube intersection graph. In particular, the binomial random subcube intersection graph model $G_{V,d,p}$ has a much more dramatic

variation of degrees than its non-structured counterpart $\mathcal{G}(V, [m], p)$.

We end this section by noting that there has been some interest in another model of 'structured' random intersection graphs, namely *random interval graphs*. The idea here is to associate to each vertex $v \in V$ a feature interval $f(v) = I_v = [a_v, b_v] \subseteq [0, 1]$ at random and to set an edge between $u, v \in V$ whenever $I_u \cap I_v \neq \varnothing$. Here 'at random' means the intervals are generated by independent pairs of uniform $U(0, 1)$ random variables, which serve as the endpoints. A d-dimensional version of this model also exists, where we associate to each vertex a d-dimensional box lying inside $[0, 1]^d$. This gives rise to *(random) d-box graphs*.

In the setting of intervals and d-boxes, we do have the Helly property. The random interval and random d-box graph models are however quite different from the random subcube intersection graphs we study in this paper.

1.3 Previous work on random intersection graphs and subcube intersection graphs

Subcube intersection graphs were introduced by Johnson and Markström [25], with motivation coming from the example in social choice theory we discussed above. They studied cliques in subcube intersection graphs from an extremal perspective, obtaining a number of results on Ramsey- and Turán-type problems and providing a counterpoint to the probabilistic perspective of the work undertaken in this paper. Subcube intersection graphs have also appeared in connection with biclique covers — the old and well-studied problem of covering the edges of a graph with as few complete bipartite subgraphs as possible — see the work of Pinto [36] on the subject.

The random intersection graph models $\mathcal{G}(V, [m], p)$ and $\mathcal{G}(V, [m], k)$ we presented in the previous subsection have for their part received extensive attention from the research community since they were introduced by Karoński, Scheinerman and Singer-Cohen [26] and Singer-Cohen [42]. By now, many results are known on their connectivity [6, 24, 37, 42], hamiltonicity [9, 18], component evolution [3, 7, 37], clique number [8, 26, 39], independence number [33], chromatic number [4, 34], degree distribution [43] and near-equivalence to the Erdős–Rényi model $\mathcal{G}_{n,p}$ for some range of the parameters [21, 38], amongst other properties. Even more recently, there has been interest in obtaining versions of the results cited above for inhomogeneous random intersection graph models.

Finally there has been some work on random intersection graphs and d-box graphs that runs somewhat parallel to the work of Johnson and Markström and of this paper. From an extremal perspective, sufficient conditions for the existence of large cliques in d-box graphs were investigated by Berg, Norine, Su, Thomas and Wollan [5] in the context of models for social agreement and approval voting, while random interval graphs were introduced by Scheinerman [40], and have been extensively studied [16, 23, 35, 41].

1.4 Results of this paper

In this paper we study the behaviour of the binomial and uniform subcube intersection models when d is large (see Remark 1.8 below for a discussion of the constant d case).

We study two main properties, that of containing a clique of size $s = s(d)$, and that of covering the entirety of the underlying hypercube Q_d with the union $\bigcup_{v\in V} f(v)$ of the feature subcubes.

Both of these properties are closed under the addition of vertices to V (or, equivalently, of subcubes $f(v)$ to the family of feature subcubes). The question is then how large V needs to be for these properties to hold *with high probability* (whp), that it to say with probability tending to 1 as $d \to \infty$.

In the case of covering, this question can be thought of as a structured variant of the classical coupon collector problem, Problem 2.15, which had not been considered before. This problem is discussed in greater detail in Section 2.3, and further investigated in the preprint [20], but is still at present far from settled.

Returning to our random subcube intersection graphs, we formally take a dynamic view of our models: for fixed $p, \alpha \in [0, 1]$ we consider a nested sequence of vertex sets $V_1 \subset V_2 \subset \ldots$, with $|V_n| = n$, and corresponding nested sequences of binomial random subcube intersection graphs $B_n = G_{V_n,d,p}$ and uniform random subcube intersection graphs $U_n = G_{V_n,d,\lfloor \alpha d \rfloor}$.

Definition 1.6. Let $\mathcal{P}$ be a property of subcube intersection graphs that is closed with respect to the addition of vertices. The *hitting time* $N^b_{\mathcal{P}} = N^b_{\mathcal{P}}(d, p)$ for $\mathcal{P}$ for the binomial sequence $(B_n)_{n\in\mathbb{N}}$ is

$$N^b_{\mathcal{P}} := \min \{n \in \mathbb{N} \colon B_n \in \mathcal{P}\}.$$

Further, the *hitting time* $N^u_{\mathcal{P}} = N^u_{\mathcal{P}}(d, \alpha)$ for $\mathcal{P}$ for the uniform sequence $(U_n)_{n\in\mathbb{N}}$ is

$$N^u_{\mathcal{P}} := \min \{n \in \mathbb{N} \colon U_n \in \mathcal{P}\}.$$

In this paper, we restrict our attention to the binomial model $G_{V,d,p}$ with $p \in (0, 1)$ fixed, and to the uniform model $G_{V,d,k}$ with $k = k(d) = \lfloor \alpha d \rfloor$ for $\alpha \in (0, 1)$ fixed. In both cases, the interesting behaviour occurs when $|V| = e^{xd}$ for x bounded away from 0. We thus typically use this number x as a parameter, rather than the actual number $n = |V|$ of vertices in the graph. Our aim is to establish concentration of the exponent of the hitting time. We thus make the following definitions:

Definition 1.7. Let $\mathcal{P}$ be a property of subcube intersection graphs that is closed with respect to the addition of vertices. A real number $t \geqslant 0$ is a *threshold* for $\mathcal{P}$ in the binomial model (with parameter p) if

$$\frac{\log\left(N^b_{\mathcal{P}}(d, p)\right)}{d} \to t$$

in probability as $d \to \infty$.

Further,a real number $t \geqslant 0$ is a *threshold* for $\mathcal{P}$ in the uniform model (with parameter $k = \lfloor \alpha d \rfloor$) if

$$\frac{\log\left(N^u_{\mathcal{P}}(d, \alpha)\right)}{d} \to t$$

in probability as $d \to \infty$.

In other words, $t > 0$ is a threshold for the property $\mathcal{P}$ in the binomial model if for any sequence of vertex sets $V = V(d)$

$$\lim_{d\to\infty} \mathbb{P}(G_{V,d,p} \in \mathcal{P}) = \begin{cases} 0 & \text{if } |V(d)| \leqslant e^{xd} \text{ for some } x < t, \\ 1 & \text{if } |V(d)| \geqslant e^{xd} \text{ for some } x > t, \end{cases}$$

Similarly, $t > 0$ is a threshold for $\mathcal{P}$ in the uniform model if for any sequence of vertex sets $V = V(d)$,

$$\lim_{d\to\infty} \mathbb{P}(G_{V,d,k} \in \mathcal{P}) = \begin{cases} 0 & \text{if } |V(d)| \leqslant e^{xd} \text{ for some } x < t, \\ 1 & \text{if } |V(d)| \geqslant e^{xd} \text{ for some } x > t. \end{cases}$$

Our main results are determining the thresholds for the appearance of cliques and for covering the ambient hypercube in both the binomial and the uniform model. In most cases we also give some slightly more precise information about the thresholds, going into the lower order terms. We show in particular that around the covering threshold, the clique number of our models undergoes a transition: below the covering threshold, the clique number is whp of order $O(1)$; close to the covering threshold, it is whp of order a power of d; finally above the covering threshold, it is whp of order exponential in d.

Our paper is structured as follows. In Section 2, we state and prove our results for the binomial model. In Section 3 we use these to obtain our results for the uniform model. Finally in Section 4 we discuss small p and large p behaviour, and end with a number of open problems and conjectures concerning random subcube intersection graphs.

Remark 1.8. In this paper, as we have said, we are focussing on our models in the case where $d \to \infty$. What happens when d is fixed and the number of vertices goes to infinity? In some applications, this may be a more relevant choice of parameters. The asymptotic behaviour in this case is however much simpler. Indeed, let d be fixed and let U be the family of all subcubes of Q_d. We may define a subcube intersection graph G^d on U by setting an edge between two subcubes if their intersection is non-empty. The binomial model $G_{|V|,d,p}$ is then just a *random weighted blow-up* of G^d: each vertex v of G^d is replaced by a clique with a random size c_v, where $\sum_v c_v = |V|$, and by standard Chernoff bounds $c_v = (1 + o(1))p_v|V|$ for every v, where p_v is the probability that a binomial random cube is equal to v. Thus knowledge of the finite graph G^d will give us essentially all the information we could require concerning the graph $G_{V,d,p}$ as $|V| \to \infty$.

Similarly, the asymptotic behaviour of $G_{V,d,k}$ for d fixed can be inferred from the properties of the intersection graph G^d_k of the k-dimensional subcubes of Q_d. We note that this latter graph G^d_k may be thought of as a subcube analogue of (the complement of) a Kneser graph, and is an interesting graph theoretic object in its own right. Along with related constructions, it appears for example in the aforementioned work of Pinto [36] on biclique covers.

1.5 A note on approximations and notation

Throughout this paper we shall need some standard approximations. In particular we shall often use $\binom{m}{\beta m} = e^{-m\log\left(\beta^\beta(1-\beta)^{1-\beta}\right)+O(\log m)}$ (for $\beta \in (0,1)$ fixed) and $(1-\eta)^m =$

$e^{-\eta m+O(\eta^2 m)}$ (for $\eta = o(1)$). We will also use the notation $f(n) \ll g(n)$ to denote that $f(n) = o(g(n))$, and $f(n) \gg g(n)$ to denote that $g(n) = o(f(n))$.

2 The binomial model

2.1 Summary

In this section, we prove our results for the binomial model. Denote by K_s the complete graph on s vertices. Recall that the *clique number* $\omega(G)$ of a graph G is the largest s such that G contains a copy of K_s as a subgraph.

Theorem 2.1. *Let $p \in (0,1)$ and $\varepsilon > 0$ fixed. Let $s = s(d)$ be a sequence of non-negative integers with $s(d) = o\left(\frac{d}{\log d}\right)$. Set*

$$t_s(p) = -\frac{1}{s}\log\left(2\left(\frac{1+p}{2}\right)^s - p^s\right).$$

Then for every sequence of vertex sets $V(d)$ with $x(d) = \frac{1}{d}\log|V(d)|$,

$$\lim_{d\to\infty}\mathbb{P}\left(G_{V,d,p} \text{ contains a } K_s\right) = \begin{cases} 0 & \text{if } x(d) \leqslant t_s(p) + \frac{\log s}{d} - \varepsilon\frac{\log d}{d} \\ 1 & \text{if } x(d) \geqslant t_s(p) + \frac{2\log s}{d} + \varepsilon\frac{\log d}{d}. \end{cases}$$

Corollary 2.2. *Let $p \in (0,1)$ and $s \in \mathbb{N}$ be fixed. The threshold for the appearance of s-cliques in $G_{V,d,p}$ is*

$$t_s(p) = \log\frac{2}{1+p} - \frac{1}{s}\log\left(2 - \left(\frac{2p}{1+p}\right)^s\right).$$

Remark 2.3. As we shall see in the proof of Theorem 2.1, from the moment it becomes non-zero, the number of edges in $G_{V,d,p}$ remains concentrated about its expectation $e^{2(x-t_2)d+o(d)}$. If there was no clustering in $G_{V,d,p}$, that is, if cliques appeared no earlier than they would in the Erdős–Rényi model with parameter $e^{-2t_2 d}$, then we would expect s-cliques to appear roughly when $x = (s-1)t_2$.

However, it is the case that $t_s < (s-1)t_2$ for all $p \in [0,1)$ and all $s \geqslant 3$. This is an exercise in elementary calculus. In particular, s-cliques appear much earlier than we would expect them to given the edge-density of our binomial random subcube intersection graphs. Indeed, letting $p \to 0$, we have by Corollary 2.2 that for $s \geqslant 3$

$$t_s(p) = \left(1 - \frac{1}{s}\right)\log 2 - p + O(p^2),$$

while the threshold for having a linear number of edges is $2t_2 = \log 2 - 2p + O(p^2)$, which is strictly larger provided p is chosen sufficiently small. Thus for every $s \in \mathbb{N}$, there exists $p_s \in [0,1]$ such that for all fixed $p \in [0, p_s]$, whp we see s-cliques appear in $G_{V,d,p}$ before we have a linear number of edges. This stands in stark contrast to the situation for the Erdős–Rényi model.

Theorem 2.4. *Let $p \in (0,1)$ and $\varepsilon > 0$ be fixed. Let $V = V(d)$ be a sequence of vertex sets with $x(d) = \frac{1}{d}\log|V(d)|$. Then, for the binomial model $G_{V,d,p}$,*

$$\lim_{d\to\infty} \mathbb{P}\left(\bigcup_{v\in V} f(v) = Q_d\right) = \begin{cases} 0 & \text{if } x(d) \leqslant \log\frac{2}{1+p} + \frac{\log d}{d} + \frac{\log(\log 2-\varepsilon)}{d} \\ 1 & \text{if } x(d) \geqslant \log\frac{2}{1+p} + \frac{\log d}{d} + \frac{\log(\log 2+\varepsilon)}{d}. \end{cases}$$

Corollary 2.5. *Let $p \in (0,1)$ be fixed. Then the threshold for covering the ambient hypercube Q_d with the feature subcubes from $G_{V,d,p}$ is*

$$t_{cover}(p) = \log\frac{2}{1+p}.$$

Remark 2.6. $\lim_{s\to\infty} t_s(p) = t_{\text{cover}}(p)$.

Theorem 2.7. *Let $p \in (0,1)$ and $\varepsilon > 0$ be fixed, and let $s = s(d)$ be a sequence of integers with $s \gg d/\log d$. Then for every sequence of vertex sets $V(d)$ with $x(d) = \frac{1}{d}\log|V(d)|$,*

$$\lim_{d\to\infty} \mathbb{P}\left(G_{V,d,p} \text{ contains a } K_s\right) = \begin{cases} 0 & \text{if } x(d) \leqslant \log\frac{2}{1+p} + \frac{\log s}{d} - \frac{\varepsilon\log d}{d} \\ 1 & \text{if } x(d) \geqslant \log\frac{2}{1+p} + \frac{\log s}{d} + \frac{\varepsilon}{d}. \end{cases}$$

Further, if $\frac{s(d)}{d} \to \infty$ as $d \to \infty$, then we may improve the lower bound on the appearance of s-cliques to $x(d) \leqslant \log\frac{2}{1+p} + \frac{\log s}{d} - \frac{\varepsilon}{d}$.

Theorem 2.1 is proved in Section 2.2, where in addition we prove some key results on the dimension of the feature subcubes of the vertices in the first s-clique to appear in our graph. These will be needed in Section 3 when we study the uniform model. Theorem 2.4 is proved in Section 2.3, while Theorem 2.7 is proved in Section 2.4. Our results give whp lower and upper bounds on certain hitting times, and their proofs are split accordingly into two parts, one for each direction.

Before we proceed to the proofs, let us remark that our results imply that the clique number $\omega(G_{V,d,p})$ undergoes a transition around the covering threshold.

Corollary 2.8. *Let $p \in (0,1)$. Let $V = V(d)$ be a sequence of vertex-sets and $x(d) = \frac{1}{d}\log|V(d)|$. The following hold:*

- *if there is $s \in \mathbb{N}$ and $\varepsilon > 0$ such that $t_s + \varepsilon < x < t_{s+1}$, then whp $\omega(G_{V,d,p}) = s$;*
- *if there is $s \in \mathbb{N}$ such that $x = t_s + o(1)$, then whp $\omega(G_{V,d,p}) \in \{s, s-1\}$;*
- *if there is $\gamma > 0$ such that $x = x(d) = t_{cover} + \gamma\frac{\log d}{d} + o\left(\frac{\log d}{d}\right)$, then whp $\omega(G_{V,d,p})$ has order $d^{\gamma+o(1)}$;*
- *if there is $c > 0$ such that $x = x(d) = t_{cover} + c + o(1)$, then whp $\omega(G_{V,d,p})$ has order $e^{cd+o(d)}$.*

2.2 Below the covering threshold

Proof of Theorem 2.1. Without loss of generality we may assume $V = [n]$. Set $x = \frac{1}{d}\log n$, and let $\varepsilon > 0$ be fixed. Let $s = s(d)$ be a sequence of non-negative integers with $s(d) = o(d/\log d)$.

Let $q(s,d)$ denote the probability that a given s-set of vertices induces an s-clique in $G_{[n],d,p}$. By Remark 1.5, we have that

$$q(s,d) = q(s,1)^d = \left(2\left(\frac{1+p}{2}\right)^s - p^s\right)^d = \exp\left(-sdt_s(p)\right).$$

Let $X = X(d)$ be the random variable denoting the number of copies of K_s in $G_{[n],d,p}$.
Lower bound: suppose $x \leqslant t_s(p) + \frac{\log s}{d} - \varepsilon\frac{\log d}{d}$. Then

$$\begin{aligned}\mathbb{E}X = \binom{n}{s} q(s,d) &= \exp\left(sdx - s\log s - sdt_s(p) + O(s)\right)\\ &\leqslant \exp\left(-\varepsilon s\log d + O(s)\right) = o(1).\end{aligned}$$

It follows by Markov's inequality that whp $X = 0$ and $G_{[n],d,p}$ contains no s-clique, proving the first part of the theorem.
Upper bound: suppose $x \geqslant t_s(p) + \frac{2\log s}{d} + \varepsilon\frac{\log d}{d}$. We have

$$\mathbb{E}X = \binom{n}{s} q(s,d) \geqslant \left(\frac{n}{s}\right)^s q(s,d) \geqslant \exp\left(\varepsilon s\log d\right) \gg 1.$$

We use Chebyshev's inequality to show that X is concentrated about this value (and hence that whp $G_{[n],d,p}$ contains an s-clique).

Fix $i:\ 0 \leqslant i \leqslant s$. Let A, B be two s-sets of vertices meeting in exactly i vertices. Using Remark 1.5 and the inclusion-exclusion principle, we compute the probability b_i that both A and B induce a copy of K_s in $G_{[n],d,p}$:

$$\begin{aligned}b_i &= \left(2\left(\frac{1+p}{2}\right)^{2s-i} + 2\left(\frac{1+p}{2}\right)^{2s-2i} p^i - 4\left(\frac{1+p}{2}\right)^{s-i} p^s + p^{2s-i}\right)^d\\ &= \left(\frac{1+p}{2}\right)^{(2s-i)d}\left(2 + 2\left(\frac{2p}{1+p}\right)^i - 4\left(\frac{2p}{1+p}\right)^s + \left(\frac{2p}{1+p}\right)^{2s-i}\right)^d.\end{aligned}$$

(Note $b_0 = q(s,d)^2$ and $b_s = q(s,d)$.) Now,

$$\mathbb{E}X^2 = \binom{n}{s}\sum_{i=0}^{s}\binom{n-s}{s-i}\binom{s}{i} b_i.$$

We claim that the dominating contribution to this sum comes from the $i = 0$ term. Indeed, for $s(d) = o\left(\frac{d}{\log d}\right)$, d large and $x \geqslant t_s + \frac{2\log s}{d} + \varepsilon\frac{\log d}{d}$,

$$\begin{aligned}\frac{\binom{n-s}{s-i}\binom{s}{i}}{\binom{n-s}{s}}\frac{b_i}{b_0} &\leqslant \frac{2s^{2i}}{n^i}\frac{b_i}{b_0} \qquad \text{(provided } d \text{ is sufficiently large)}\\ &= 2s^{2i}\exp\Big(d\Big[-ix - (2s-i)\log\frac{2}{1+p}\\ &\quad + \log\left(2 + 2\left(\frac{2p}{1+p}\right)^i - 4\left(\frac{2p}{1+p}\right)^s + \left(\frac{2p}{1+p}\right)^{2s-i}\right)\\ &\quad + 2s\log\frac{2}{1+p} - \log\left(\left(2 - \left(\frac{2p}{1+p}\right)^s\right)^2\right)\Big]\Big)\end{aligned}$$

(the second and third term in the exponent coming from $\log b_i$ and the last two terms coming from $\log b_0 = \log\left(q(s,d)^2\right)$)

$$\begin{aligned}&\leqslant \frac{2}{d^{i\varepsilon}}\exp\Big(d\Big[\log\left(2 + 2\left(\frac{2p}{1+p}\right)^i - 4\left(\frac{2p}{1+p}\right)^s + \left(\frac{2p}{1+p}\right)^{2s-i}\right)\\ &\quad - \left(2 - \frac{i}{s}\right)\log\left(2 - \left(\frac{2p}{1+p}\right)^s\right)\Big]\Big),\end{aligned} \tag{2.1}$$

with the inequality in the last line coming from substituting $t_s(p) + \frac{2\log s}{d} + \varepsilon\frac{\log d}{d}$ for x, and rearranging terms. We now resort to the following technical lemma.

Lemma 2.9. *For all $y \in [0,1]$ and all integers $0 \leqslant i \leqslant s$, the following inequality holds:*

$$\left(2 + 2y^i - 4y^s + y^{2s-i}\right)^s \leqslant \left(2 - y^s\right)^{2s-i}.$$

We defer the proof of Lemma 2.9 (which is a simple albeit lengthy exercise) to Appendix A. Set $y = \left(\frac{2p}{1+p}\right)$. As $0 < p < 1$, we have $y \in (0,1)$. Applying Lemma 2.9, we have

$$\log\left(2 + 2y^i - 4y^s + y^{2s-i}\right) - \left(2 - \frac{i}{s}\right)\log\left(2 - y^s\right) < 0.$$

Substituting this into the expression inside the exponential in (2.1), we get

$$\frac{\binom{n-s}{s-i}\binom{s}{i}}{\binom{n-s}{s}}\frac{b_i}{b_0} \leqslant \frac{2}{d^{i\varepsilon}}.$$

Thus

$$\begin{aligned}\mathbb{E}\left(X^2\right) &\leqslant \binom{n}{s}\binom{n-s}{s}b_0\left(1 + \frac{2}{d^\varepsilon}\left(1 + \frac{1}{d^\varepsilon} + \frac{1}{d^{2\varepsilon}} + \cdots\right)\right)\\ &= \left(\mathbb{E}X\right)^2(1 + o(1)).\end{aligned}$$

In particular, $\mathrm{Var}(X) = o\left(\left(\mathbb{E}X\right)^2\right)$, and by Chebyshev's inequality whp X is at least $\frac{1}{2}\mathbb{E}X > 0$. Thus whp $G_{[n],d,p}$ contains (many) s-cliques. □

Remark 2.10. We have shown that the transition between whp no s-cliques and whp many s-cliques in $G_{[n],d,p}$ takes place inside a window of width (with respect to x) of order $O\left(\frac{\log d}{d}\right)$. In the case when s is bounded, $s(d) = O(1)$, it is easy to run through the proof of Theorem 2.1 again and show that in fact we may replace $\varepsilon \log d$ in the statement of the Theorem by any function $h = h(d)$ tending to infinity with d, so that the width of the window may be reduced to $O\left(\frac{h}{d}\right)$.

Having proven Theorem 2.1, we now turn our attention to the following question. Let $s \in \mathbb{N}$ be fixed. What is the dimension of the features subcubes in the first s-clique to appear in $G_{[n],d,p}$?

As always, write x for $\frac{1}{d}\log n$. Let S_α be a subcube of dimension αd. Suppose S_α is the feature subcube of some $v \in [n]$, $f(v) = S_\alpha$. Then the expected number of s-cliques involving v is

$$\mathbb{E}\#\{K_{s-1} \text{ meeting } S_\alpha\} = \binom{n-1}{s-1}\left(\frac{1+p}{2}\right)^{(s-1)(1-\alpha)d}\left(2\left(\frac{1+p}{2}\right)^{s-1} - p^{s-1}\right)^{\alpha d}$$

An application of Wald's equation yields that the expected number E^s_α of pairs $(v, \mathcal{S})$ for which (i) $v \in [n]$ is a vertex with a feature subcube $f(v)$ of dimension αd, and (ii) $\mathcal{S}$ is an s-set of vertices from $[n]$ containing v and inducing an s-clique in $G_{[n],d,p}$, is:

$$\begin{aligned}
E^s_\alpha &= \mathbb{E}\#\{\alpha d\text{-dimensional feature subcubes}\} \times \mathbb{E}\#\{K_{s-1} \text{ meeting } S_\alpha\} \\
&= n\binom{d}{\alpha d} p^{\alpha d}(1-p)^{(1-\alpha)d}\binom{n-1}{s-1}\left(\frac{1+p}{2}\right)^{(s-1)(1-\alpha)d}\left(2\left(\frac{1+p}{2}\right)^{s-1} - p^{s-1}\right)^{\alpha d} \\
&= \exp\Big(d\Big[sx + \alpha\log\left(\left(\frac{p}{\alpha}\right)\left(2\left(\frac{1+p}{2}\right)^{s-1} - p^{s-1}\right)\right) \\
&\qquad + (1-\alpha)\log\left(\left(\frac{1-p}{1-\alpha}\right)\left(\frac{1+p}{2}\right)^{s-1}\right)\Big] + o(d)\Big).
\end{aligned}$$

Define

$$\begin{aligned}
t^\alpha_s := -\frac{1}{s}\Big(&\alpha\log\left(\frac{p}{\alpha}\cdot\left(2\left(\frac{1+p}{2}\right)^{s-1} - p^{s-1}\right)\right) \\
&+ (1-\alpha)\log\left(\left(\frac{1-p}{1-\alpha}\right)\cdot\left(\frac{1+p}{2}\right)^{s-1}\right)\Big).
\end{aligned}$$

The expression above can then be rewritten as $E^s_\alpha = e^{sd(x - t^\alpha_s + o(1))}$. Set

$$\alpha_s = \alpha_s(p) := p\left(\frac{2\left(\frac{1+p}{2}\right)^{s-1} - p^{s-1}}{2\left(\frac{1+p}{2}\right)^{s} - p^{s}}\right).$$

Remark 2.11. The quantity α_s is exactly the probability that a given vertex receives colour $\star$ in $G_{[n],1,p}$ conditional on it forming an s-clique with a fixed $(s-1)$-set of vertices. In particular it follows from Remark 1.5 that $\alpha_s d$ is the expected dimension of feature subcubes in an s-clique in $G_{[n],d,p}$.

Remark 2.12. For $0 < p < 1$ fixed, the sequence $(\alpha_s)_{s\in\mathbb{N}}$ is strictly increasing and tends to $\frac{2p}{1+p}$ as $s \to \infty$. Note in particular that for all $s > 1$, $\alpha_s(p) > \alpha_1(p) = p$.

Proposition 2.13. *Let $p \in (0,1)$ be fixed. Then for every $s \in \mathbb{N}$ the following equality holds:*

$$t_s = t_s^{\alpha_s}.$$

Moreover, as a function of α, t_s^α is strictly decreasing for $\alpha \in [0, \alpha_s)$ and strictly increasing for $\alpha \in (\alpha_s, 1]$. In particular, α_s is the unique minimum of t_s^α over all $\alpha \in [0,1]$.

Proof. The first part of our proposition is a simple calculation. Recall from the proof of Theorem 2.1 that $q(s,1) = 2\left(\frac{1+p}{2}\right)^s - p^s$ is the probability that a given s-set of vertices forms an s-clique in $G_{[n],d,1}$. Note that

$$t_s = -\frac{1}{s}\log q(s,1), \qquad \alpha_s = \frac{pq(s-1,1)}{q(s,1)} \text{ and}$$
$$q(s,1) - pq(s-1,1) = (1-p)\left(\frac{1+p}{2}\right)^{s-1}.$$

Thus,

$$\begin{aligned}
t_s^{\alpha_s} &= -\frac{1}{s}\Bigg(\alpha_s \log\left(\left(\frac{p}{\alpha_s}\right)\left(2\left(\frac{1+p}{2}\right)^{s-1} - p^{s-1}\right)\right) \\
&\qquad + (1-\alpha_s)\log\left(\left(\frac{1-p}{1-\alpha_s}\right)\left(\frac{1+p}{2}\right)^{s-1}\right)\Bigg) \\
&= -\frac{1}{s}\Bigg(\alpha_s \log\left(\frac{q(s,1)}{q(s-1,1)}q(s-1,1)\right) \\
&\qquad + (1-\alpha_s)\log\left(\frac{(1-p)q(s,1)}{q(s,1)-pq(s-1,1)}\left(\frac{1+p}{2}\right)^{s-1}\right)\Bigg) \\
&= -\frac{1}{s}\left(\alpha_s \log q(s,1) + (1-\alpha_s)\log q(s,1)\right) \\
&= -\frac{1}{s}\log q(s,1) = t_s,
\end{aligned}$$

as required.

Now, let us show that $t_s^{\alpha_s}$ is in fact the unique minimum of t_s^α over $\alpha \in [0,1]$. Making use of our observations above, we may write st_s^α as

$$st_s^\alpha = \alpha \log\left(\frac{\alpha}{q(s,1)\alpha_s}\right) + (1-\alpha)\log\left(\frac{1-\alpha}{q(s,1)(1-\alpha_s)}\right).$$

The derivative with respect to α is

$$s\frac{d}{d\alpha}\left(t_s^\alpha\right) = \log\left(\frac{\alpha}{q(s,1)\alpha_s}\right) - \log\left(\frac{1-\alpha}{q(s,1)(1-\alpha_s)}\right),$$

which is strictly negative for $0 \leqslant \alpha < \alpha_s$, zero for $\alpha = \alpha_s$ and strictly positive for $1 \geqslant \alpha > \alpha_s$, establishing our claim. □

Used in conjunction with Theorem 2.1 (or more precisely Corollary 2.2), Proposition 2.13 enables us to identify with quite some precision the dimension of the feature subcubes of the vertices which witness the emergence of s-cliques in $G_{[n],d,p}$. Formally we return to our dynamic view of the model, and we consider the graph $B_n = G_{[n],d,p}$ at the hitting time $n = N_s^b$ for the property of containing a clique on s vertices. By definition of the hitting time, $G_{[n],d,p}$ contains at least one K_s-subgraph. Set $W_s = W_s(d,p)$ to be the set of all vertices in $[n]$ which are contained in such a K_s-subgraph.

Proposition 2.14. *Whp, all feature subcubes of vertices contained in $W_s(d,p)$ have dimension $(\alpha_s + o(1))\, d$.*

Proof. Fix $\varepsilon > 0$. By Proposition 2.13, there exists $\delta > 0$ such that if $t_s^\alpha \leqslant t_s + \delta = t_s^{\alpha_s} + \delta$, then $|\alpha - \alpha_s| < \varepsilon$. By Corollary 2.2, whp the hitting time N_s^b for containing an s-clique satisfies $e^{t_s d - \frac{\delta}{2}d} \leqslant N_s^b \leqslant e^{t_s d + \frac{\delta}{2}d}$. We show that for $|V| = e^{xd}$ and $|x - t_s| < \frac{\delta}{2}$ whp no vertex in $G_{V,d,p}$ with a feature subcube of dimension αd with $|\alpha_s - \alpha| \geqslant \varepsilon$ is contained in a copy of K_s. Since $\varepsilon > 0$ was arbitrary, this is enough to establish the proposition.

Set $I_\varepsilon = \left\{\frac{i}{d} \colon i \in \{0, 1, \dots d\}\right\} \setminus (\alpha_s - \varepsilon, \alpha_s + \varepsilon)$. The expected number of pairs (v, S) where $v \in V$ has a feature subcube of dimension αd for some $\alpha \in I_\varepsilon$, $v \in S$ and $S \subseteq V$ induces a copy of K_s in $G_{V,d,p}$ is:

$$\sum_{\alpha \in I_\varepsilon} E_\alpha^s = \sum_{\alpha \in I_\varepsilon} e^{sd(x - t_s^\alpha)} \leqslant d e^{sd(x - t_s - \delta + o(1))} \leqslant e^{-\frac{s\delta}{2}d + o(d)} = o(1).$$

Markov's inequality thus implies that whp no such pair (v, S) exists in $G_{V,d,p}$. In particular all vertices of $G_{V,d,p}$ which are contained in a copy of K_s must have dimension αd for some $\alpha \colon |\alpha - \alpha_s| < \varepsilon$, as claimed. □

2.3 The covering threshold

We may view the question of covering the hypercube Q_d with randomly selected subcubes as an instance of the following problem.

Problem 2.15 (Generalised Coupon Collector Problem)**.** Let Ω be a (large) finite set, and let X be a random variable taking values in the subsets of Ω. Suppose we are given a sequence of independent random variables $X_1, X_2, \dots, X_n$ with distribution given by X. When (for which values of n) do we have $\bigcup_{i=1}^n X_i = \Omega$ holding whp?

When X is obtained by selecting a singleton from Ω uniformly at random, Problem 2.15 is the classical coupon collector problem (see [19] and [32] for early incarnations of the problem). Much is known about the distribution of the *covering time* $T = \min\{n\colon \bigcup_{i\leqslant n} X_i = V\}$ in the case of k-uniform coupon collectors, when X is a k-subset of V selected uniformly at random, for some fixed set-size $k = o(n)$. Problem 2.15 asks us to determine the distribution of T for more general distributions X. In particular, we would like to understand how the eventual structure in the distribution of X may affect the location of T and allow for behaviour deviating from that typically seen in k-uniform coupon collectors. Problem 2.15 is investigated in much more depth in [20]. For the purposes of the present paper, we therefore restrict ourselves to making the following simple observation, before specializing to the case of subcubes.

Proposition 2.16. *Set $|\Omega| = m$. Suppose X is such that for every $\omega, \omega' \in \Omega$, $\mathbb{P}(\omega \in X) = \mathbb{P}(\omega' \in X)$. Then, for every fixed $\varepsilon > 0$,*

$$\lim_{m\to\infty} \mathbb{P}\left(\bigcup_{i=1}^{n} X_i = \Omega\right) = \begin{cases} 0 & \text{if } n \ll \frac{m}{\mathbb{E}|X|} \\ 1 & \text{if } n \geqslant (1+\varepsilon)\frac{m\log m}{\mathbb{E}|X|} \end{cases}$$

Proof. For any fixed $\omega \in \Omega$,

$$\mathbb{P}\left(\omega \notin \bigcup_{i=1}^{n} X_i\right) = (1 - \mathbb{P}(\omega \in X))^n = \left(1 - \frac{\mathbb{E}|X|}{m}\right)^n.$$

Thus if $n = o\left(\frac{m}{\mathbb{E}|X|}\right)$, the probability that $\omega \notin \bigcup_{i=1}^{n} X_i$ is $e^{-o(1)} = 1 - o(1)$, whence whp $\bigcup_{i=1}^{n} X_i \neq \Omega$. On the other hand, if $n \geqslant (1+\varepsilon)\frac{m\log m}{\mathbb{E}|X|}$, then

$$\mathbb{E}\left|\Omega \setminus \bigcup_{i=1}^{m} X_i\right| = \sum_{\omega\in\Omega} \mathbb{P}\left(\omega \notin \bigcup_{i=1}^{n} X_i\right) = m\left(1 - \frac{\mathbb{E}|X|}{m}\right)^n = o(1)$$

so that by Markov's inequality whp there are no uncovered elements and $\bigcup_{i=1}^{n} X_i = \Omega$. □

The bounds we give in Proposition 2.16 are crude, but also essentially sharp (see [2, 20]). In our setting, we have $\Omega = Q_d$, and the $(X_i)_{i=1}^n$ are the feature subcubes of vertices in the binomial random subcube intersection graph $G_{[n],d,p}$. Note that the expected volume of a feature subcube $f(v)$ is:

$$\mathbb{E}|f(v)| = \sum_{i=0}^{d} \mathbb{P}(f(v) \text{ has dimension } i)2^i = \sum_{i=0}^{d} \binom{d}{i}(1-p)^{d-i}(2p)^i = (1+p)^d,$$

while on the other hand typical feature subcubes have dimension $pd + o(d)$ and thus volume $2^{pd+o(d)}$. Since $2^p < 1+p$ for all $p \in (0,1)$, typical feature subcubes have a volume much smaller than the expected volume. In particular, the variance of the volume of a feature subcube is large, and our covering problem differs significantly from the classical coupon collector problem.

We need to make one more definition before proceeding to the proof of Theorem 2.4.

Definition 2.17. The *Hamming distance* $\mathrm{dist}(y, y')$ between two elements y, y' of Q_d is the number of coordinates in which they differ.

Proof of Theorem 2.4. Without loss of generality, we may assume that $V = [n]$, and that we are working with the binomial random subcube intersection graph $G_{[n],d,p}$. Let $\mathbf{0}$ denote the all zero element $(0, 0, \ldots 0)$ from Q_d. The expected number of elements of $Q_d = \{0,1\}^d$ not covered by the union of the feature subcubes $\bigcup_{v=1}^n f(v)$ is

$$\begin{aligned}\mathbb{E}|Q_d \setminus \bigcup_{v=1}^n f(v)| &= |Q_d|\mathbb{P}\left(\mathbf{0} \notin \bigcup_{v=1}^n f(v)\right) = 2^d\left(1-\left(\frac{1+p}{2}\right)^d\right)^n \\ &= \exp\left(d\log 2 - n\left(\frac{1+p}{2}\right)^d\left(1+O\left(\frac{1+p}{2}\right)^d\right)\right)\end{aligned}$$

Now let ε be fixed with $0 < \varepsilon < \log 2$.

Upper bound: suppose $n = e^{xd} \geqslant \left(\frac{2}{1+p}\right)^d d\,(\log 2 + \varepsilon)$. Then the expected number of uncovered elements of Q_d is at most $e^{-\varepsilon d + o(d)} = o(1)$, whence we deduce from Markov's inequality that whp $\bigcup_{v=1}^n f(v) = Q_d$.

Lower bound: suppose $n = e^{xd} = \lfloor\left(\frac{2}{1+p}\right)^d d\,(\log 2 - \varepsilon)\rfloor$. Then the expected number of uncovered elements of Q_d is $e^{\varepsilon d + o(d)}$, which is large. We use Chebyshev's inequality to show the actual number of uncovered elements is concentrated about this value.

For $1 \leqslant i \leqslant d$, let $\mathbf{e}_{[i]}$ denote the element of $Q_d = \{0,1\}^d$ with its first i coordinates equal to 1 and its last $d-i$ coordinates equal to 0. Clearly we have $\mathrm{dist}(\mathbf{0}, \mathbf{e}_{[i]}) = i$. The probability that neither of $\mathbf{0}, \mathbf{e}_{[i]}$ is covered is:

$$\begin{aligned}\mathbb{P}\left(\mathbf{0}, \mathbf{e}_{[i]} \notin \bigcup_{v=1}^n f(v)\right) &= \left(1 - 2\left(\frac{1+p}{2}\right)^d + p^i\left(\frac{1+p}{2}\right)^{d-i}\right)^n \\ &= \exp\left(-n\left[2\left(\frac{1+p}{2}\right)^d - p^i\left(\frac{1+p}{2}\right)^{d-i}\right] + O\left(n\left(\frac{1+p}{2}\right)^{2d}\right)\right) \\ &= \exp\left(-(\log 2 - \varepsilon)\,d\left(2 - \left(\frac{2p}{1+p}\right)^i\right) + o(1)\right).\end{aligned}$$

Thus

$$\begin{aligned}\mathbb{E}\left(\left|Q_d \setminus \bigcup_{v=1}^n f(v)\right|^2\right) &= 2^d \sum_{i=0}^d \binom{d}{i}\mathbb{P}\left(\mathbf{0}, \mathbf{e}_{[i]} \notin \bigcup_{v=1}^n f(v)\right) \\ &= e^{2\varepsilon d}\sum_{i=0}^d \frac{\binom{d}{i}}{2^d}\exp\left(\left(\frac{2p}{1+p}\right)^i(\log 2 - \varepsilon)\,d + o(1)\right). \qquad (2.2)\end{aligned}$$

Pick $\eta\colon 0<\eta<1/2$ sufficiently small such that

$$\varepsilon > \eta\log\frac{1}{\eta}+(1-\eta)\log\frac{1}{1-\eta}$$

is satisfied. Then

$$\sum_{0\leqslant i\leqslant \eta d}\frac{\binom{d}{i}}{2^d}\exp\left(\left(\frac{2p}{1+p}\right)^i(\log 2-\varepsilon)\,d\right) < \eta d\frac{\binom{d}{\eta d}}{2^d}\exp\left((\log 2-\varepsilon)\,d\right)$$
$$= \eta d\binom{d}{\eta d}e^{-\varepsilon d} = o(1). \tag{2.3}$$

On the other hand, for $i>\eta d$ we have $\left(\frac{2p}{1+p}\right)^i(\log 2-\varepsilon)\,d = o(1)$, since $\frac{2p}{1+p}<1$, so that

$$\sum_{i\geqslant\eta d}\frac{\binom{d}{i}}{2^d}\exp\left(\left(\frac{2p}{1+p}\right)^i(\log 2-\varepsilon)\,d+o(1)\right)$$
$$\leqslant \sum_{i\geqslant\eta d}\frac{\binom{d}{i}}{2^d}e^{o(1)} \leqslant 1+o(1). \tag{2.4}$$

Substituting the bounds (2.3) and (2.4) into (2.2), we get

$$\mathbb{E}\left(\left|Q_d\setminus\bigcup_{v=1}^n f(v)\right|^2\right)\leqslant e^{2\varepsilon d}(1+o(1)) = (1+o(1))\left(\mathbb{E}\left|Q_d\setminus\bigcup_{v=1}^n f(v)\right|\right)^2,$$

whence $\mathrm{Var}|Q_d\setminus\bigcup_{v=1}^n f(v)| = o\left(\mathbb{E}\,|Q_d\setminus\bigcup_{v=1}^n f(v)|^2\right)$. It follows by Chebyshev's inequality that whp $\bigcup_{v=1}^n f(v)$ leaves $(1+o(1))e^{\varepsilon d}$ elements of Q_d uncovered when $n\leqslant\left(\frac{2}{1+p}\right)^d d\,(\log 2-\varepsilon)$, as claimed. □

2.4 Above the covering threshold

Proof of Theorem 2.7. Without loss of generality, we may assume that $V=[n]$. Fix $\varepsilon>0$, and let $s=s(d)$ be a sequence of natural numbers with $\frac{s\log d}{d}\to\infty$ as $d\to\infty$.
Upper bound: Here, unlike in the proof of Theorem 2.1, we eschew estimates of the total number of s-cliques present in $G_{v,d,p}$, but proceed instead via a covering argument. Indeed, by the Helly property of subcubes of Q_d, $G_{[n],d,p}$ contains an s-clique if and only if some element of the ambient hypercube Q_d is contained in at least s feature subcubes. Denote by

$$\mathrm{Vol}[n] := \sum_{v=1}^n |f(v)|$$

the sum of the sizes of the feature subcubes. By linearity of expectation,

$$\mathbb{E}\mathrm{Vol}[n] = n\mathbb{E}|f(1)| = n(1+p)^d.$$

Set $x = \log \frac{2}{1+p} + \frac{\log s}{d} + \frac{\varepsilon}{d}$. For $n \geqslant \lceil e^{xd} \rceil$, we have $\mathbb{E}\mathrm{Vol}[n] \geqslant e^{\varepsilon} s 2^d$, which means that elements of the ambient hypercube are expected to be contained in $e^{\varepsilon} s > s$ feature subcubes. Thus, to show that $G_{[n],d,p}$ whp contains (many) s-cliques for this value of n, it is enough to show that $\mathrm{Vol}[n]$ is concentrated about its mean. Again, we use the second-moment method to do this. By linearity of variance we have

$$\begin{aligned}\mathrm{VarVol}[n] = n\mathrm{Var}(f(1)) &= n\left(\left(\sum_{i=1}^{d}\binom{d}{i}(1-p)^{d-i}p^i 2^{2i}\right) - (1+p)^{2d}\right) \\ &= n\left((1+3p)^d - (1+2p+p^2)^d\right).\end{aligned}$$

Applying Chebyshev's inequality,

$$\begin{aligned}\mathbb{P}\left(\mathrm{Vol}[n] < s2^d\right) &= \mathbb{P}\left(\mathrm{Vol}[n] < e^{-\varepsilon}\mathbb{E}\mathrm{Vol}[n]\right) \\ &\leqslant \frac{\mathrm{VarVol}[n]}{\left(1-e^{-\varepsilon}\right)^2\left(\mathbb{E}\mathrm{Vol}[n]\right)^2} \\ &< \frac{(1+3p)^d}{\left(1-e^{-\varepsilon}\right)^2 n(1+p)^{2d}} \\ &\leqslant \frac{1}{\left(1-e^{-\varepsilon}\right)^2}\frac{(1+3p)^d}{2^d(1+p)^d} \qquad \text{(substituting in the value of } n\text{)} \\ &= \frac{1}{\left(1-e^{-\varepsilon}\right)^2}\left(1-\left(\frac{1-p}{2(1+p)}\right)\right)^d = o(1).\end{aligned}$$

In particular,

$$\mathbb{P}\left(G_{[n,d,p]} \text{ contains an } s\text{-clique}\right) \geqslant 1 - \mathbb{P}\left(\mathrm{Vol}[n] < s2^d\right) = 1 - o(1),$$

proving the claimed upper bound on the threshold for the emergence of s-cliques.

Remark 2.18. The proof above actually shows a little more: for $x > \log\left(\frac{1+3p}{(1+p)^2}\right)$, we have $\mathrm{VarVol}[n] = o\left(\left(\mathbb{E}\mathrm{Vol}[n]\right)^2\right)$ and thus by Chebyshev's inequality whp $\mathrm{Vol}[n] = (1 + o(1))n(1+p)^d$. In other words there are sufficiently many feature subcubes at this point that the large variance of their individual volumes ceases to matter. Note that this occurs before the covering threshold, since $\log\left(\frac{1+3p}{(1+p)^2}\right) < \log\frac{2}{1+p}$.

Lower bound when $s = O(d)$: in this case we use Markov's inequality just as in the proof of Theorem 2.1. Set $x = \log\frac{2}{1+p} + \frac{\log s}{d} - \varepsilon\frac{\log d}{d}$, and let $n = \lfloor e^{xd} \rfloor$. Let $X = X(d)$ be the number of s-cliques in the graph. Then

$$\begin{aligned}\mathbb{E}X &= \binom{n}{s}\left(2\left(\frac{1+p}{2}\right)^s - p^s\right)^d \\ &= \exp\left(sxd - s\log s - ds\log\frac{2}{1+p} + d\log\left(2 - \left(\frac{2p}{1+p}\right)^s\right) + O(s)\right) \\ &= \exp\left(-s\varepsilon\log d + O\left(\max(s,d)\right)\right) = o(1) \qquad (\text{since } s\log d \gg \max(s,d))\end{aligned}$$

so that whp $X = 0$ and $G_{[n],d,p}$ contains no s-clique.

Lower bound when $s \gg d$: here we use a covering idea. Suppose $n = e^{-\varepsilon} s \left(\frac{2}{1+p}\right)^d$. The number $C_{\mathbf{0}}$ of feature subcubes containing the element $\mathbf{0} = (0, 0 \dots 0)$ is the sum of n independent identically distributed Bernoulli random variables with parameter $\left(\frac{1+p}{2}\right)^d$. We have $\mathbb{E}C_{\mathbf{0}} = n\left(\frac{1+p}{2}\right)^d \leqslant e^{-\varepsilon} s$. Applying a Chernoff bound, we deduce that

$$\mathbb{P}(C_{\mathbf{0}} \geqslant s) \leqslant e^{-\frac{\varepsilon^2}{3}s}.$$

In particular the expected number of elements of Q_d contained in at least s feature subcubes is $|Q_d|\mathbb{P}(C_{\mathbf{0}} \geqslant s) \leqslant 2^d e^{-\frac{\varepsilon^2}{3}s}$, which is $o(1)$ for $s \gg d$. It follows by Markov's inequality that whp there is no such element, and thus, by the Helly property for subcube intersections, that whp $G_{[n],d,p}$ contains no copy of K_s. Further by monotonicity of the property of containing an s-clique, whp $G_{[n'],d,p}$ fails to contain a K_s for any $n' \leqslant n$. □

3 The uniform model

In this section, we prove our results for the uniform model. We note that these are generally less precise than those we obtained for the binomial model, owing to the greater difficulty of performing clique computations.

3.1 Summary

Fix $s \in \mathbb{N}$. We established in Section 2 (Proposition 2.14) that in $G_{V,d,p}$, whp the feature subcubes of the vertices in the first s-cliques to appear as we increase $|V|$ all have dimension $(\alpha_s + o(1))d$, where α_s is the function

$$\alpha_s \colon p \mapsto \frac{p\left(2\left(\frac{1+p}{2}\right)^{s-1} - p^{s-1}\right)}{\left(2\left(\frac{1+p}{2}\right)^{s} - p^{s}\right)}.$$

We show in Proposition 3.9 that α_s is a bijection from $(0,1)$ to itself. This will allow us to determine the threshold for the appearance of s-cliques in the uniform model.

Theorem 3.1. *Let $\alpha \in (0,1)$ and $s \in \mathbb{N}$ be fixed, and let $k(d) = \lfloor \alpha d \rfloor$. Set $p = {\alpha_s}^{-1}(\alpha)$. Then, the threshold for the appearance of s-cliques in $G_{V,d,k}$ is*

$$T_s(\alpha) = t_s(p) + \alpha \log \frac{p}{\alpha} + (1-\alpha)\log\frac{1-p}{1-\alpha}.$$

Theorem 3.2. *Let $\alpha \in (0,1)$ and $\varepsilon > 0$ be fixed, and let $k(d) = \lfloor \alpha d \rfloor$. Let $V = V(d)$ be a sequence of vertex sets with $|V(d)| = e^{xd}$. Then, for the uniform model $G_{V,d,k}$,*

$$\lim_{d\to\infty} \mathbb{P}\left(\bigcup_{v\in V} f(v) = Q_d\right) = \begin{cases} 0 & \text{if } x(d) \leqslant (1-\alpha)\log 2 + \frac{\log d}{d} + \frac{\log(\log 2 - \varepsilon)}{d} \\ 1 & \text{if } x(d) \geqslant (1-\alpha)\log 2 + \frac{\log d}{d} + \frac{\log(\log 2 + \varepsilon)}{d}. \end{cases}$$

Corollary 3.3. *Let $\alpha \in (0,1)$ be fixed, and let $k(d) = \lfloor \alpha d \rfloor$. Then the threshold for covering the ambient hypercube Q_d with the feature subcubes from $G_{V,d,k}$ is*

$$T_{cover}(\alpha) = (1-\alpha)\log 2.$$

Remark 3.4. As we observed in Remark 2.12, we have $\lim_{s\to\infty} \alpha_s(p) = \frac{2p}{1+p}$. From this we deduce that for large s, we have $\alpha_s^{-1}(\alpha) = \frac{\alpha}{2-\alpha} + o(1)$. Substituting this into $T_s(\alpha)$, we see that

$$T_s(\alpha) \to T_{\text{cover}}(\alpha)$$

as $s \to \infty$, mirroring our observation in Remark 2.6 for the binomial model.

Theorem 3.5. *Let $\alpha \in (0,1)$ and $\varepsilon > 0$ be fixed, and let $k(d) = \lfloor \alpha d \rfloor$. Let $s = s(d)$ be a sequence of natural numbers with $\frac{s}{d} \to \infty$ as $d \to \infty$. Suppose $V = V(d)$ is a sequence of vertex sets. Then,*

$$\lim_{d\to\infty} \mathbb{P}\left(G_{V,d,k} \text{ contains an } s\text{-clique}\right) = \begin{cases} 0 & \text{if } |V(d)| \leqslant (1-\varepsilon)s2^{d-k} \\ 1 & \text{if } |V(d)| \geqslant (s-1)2^{d-k}+1. \end{cases}$$

Remark 3.6. Theorem 3.1 and Corollary 3.3 show how significant 'outliers' (subcubes with unusually high dimension) are for the behaviour of the binomial model. Indeed, Proposition 2.14 tells us that for $0 < p < 1$ fixed and $s \geqslant 3$, the vertices in the first s-clique to appear in $G_{V,d,p}$ have feature subcubes of dimension $(\alpha_s(p) + o(1))\, d$. Since $\alpha_s(p) > p$ it shall follow straightforwardly from the proof of Theorem 3.1 that $t_s(p) < T_s(p)$. Similarly, by Corollaries 2.4 and 3.3, we have for $0 < p < 1$ fixed that

$$t_{\text{cover}}(p) = \log \frac{2}{1+p} < (1-p)\log 2 = T_{\text{cover}}(p).$$

From the covering threshold upwards, Corollary 3.3 and Theorem 3.5 suggest that, when considering questions about cliques and covering, the right instance of the binomial model to compare $G_{V,d,\lfloor \alpha d \rfloor}$ with is $G_{V,d,p}$ with $p = 2^\alpha - 1$ (rather than $p = \alpha$ as we might have expected). For these two models, the covering threshold and the thresholds for higher order cliques coincide. Since both models have the same expected volume of feature subcubes, this vindicates the use of volume/covering arguments for determining the thresholds for higher order cliques. Note however that $G_{V,d,\lfloor \alpha d \rfloor}$ and $G_{V,d,2^\alpha - 1}$ have different thresholds for lower order cliques. Our binomial model and uniform model thus behave differently, and there is no good coupling between them below the covering threshold.

Finally, let us add that, just as in the binomial model, the clique number $\omega(G_{V,d,k})$ in the uniform model undergoes a transition around the covering threshold.

Corollary 3.7. *Let $\alpha \in (0,1)$ be fixed and let $k = k(d) = \lfloor \alpha d \rfloor$. Let $V(d)$ be a sequence of vertex-sets and $x(d) = \frac{1}{d}\log|V(d)|$ as usual. The following hold:*

- *if there is $s \in \mathbb{N}$ and $\varepsilon > 0$ such that $T_s + \varepsilon < x < T_{s+1}$, then whp $\omega(G_{V,d,k}) = s$;*

- *if there is $s \in \mathbb{N}$ such that $x = T_s + o(1)$, then whp $\omega(G_{V,d,k}) \in \{s, s-1\}$;*
- *if there is $\gamma > 1$ such that $x = x(d) = T_{cover} + \gamma\frac{\log d}{d} + o\left(\frac{\log d}{d}\right)$, then whp $\omega(G_{V,d,k})$ has order $d^{\gamma+o(1)}$;*
- *if there is $c > 0$ such that $x = x(d) = T_{cover} + c + o(1)$, then whp $\omega(G_{V,d,k})$ has order $e^{cd+o(d)}$.*

Remark 3.8. There is a gap here: we do not know what the order of the clique number is when $x = x(d) = T_{\text{cover}} + \gamma\frac{\log d}{d} + o(\frac{\log d}{d})$ for a fixed real γ with $0 < \gamma \leqslant 1$. We make the natural conjecture that for this value of $x(d)$, we should have $\omega(G_{V,d,k}) = d^{\gamma+o(1)}$, similarly to the binomial model.

Theorems 3.1, 3.2 and 3.5 are proved in Sections 3.2, 3.3 and 3.4 respectively. Our results give whp lower and upper bounds on certain hitting times for the uniform model, and we often split their proofs accordingly into two parts.

3.2 Below the covering threshold

Proposition 3.9. *The function α_s is a bijection from $[0,1]$ to itself, and has a continuous inverse over its domain.*

Proof. Since $\alpha_s(0) = 0$ and $\alpha_s(1) = 1$, all we have to do is show that the derivative of α_s with respect to p is strictly positive in $[0,1]$, whence we are done by the inverse function theorem.

Setting $y = \frac{2p}{1+p}$, we can rewrite $\alpha_s(p)$ as $\alpha_s(y) = \frac{2y-y^s}{2-y^s}$. By the chain rule,

$$\frac{d\alpha_s}{dp}(p) = \left(\frac{dy}{dp}(p)\right)\left(\frac{d\alpha_s}{dy}(y(p))\right) = \frac{2}{(1+p)^2}\frac{(4 - 2sy^{s-1} + 2(s-1)y^s)}{(2-y^s)^2}.$$

The derivative, with respect to y, of the numerator in the expression above is

$$2\left(-2s(s-1)(1-y)y^{s-2}\right) \leqslant 0 \qquad (\text{since } y \in [0,1]).$$

Thus the minimum of the numerator is attained when $y(p) = 1$. In particular,

$$\frac{d\alpha_s(p)}{dp}(p) \geqslant \frac{2}{(1+p)^2}\frac{2}{(2-y^s)^2} > 0,$$

as required. □

In general, computing an explicit closed-form expression for the inverse of α_s is difficult, reflecting the fact that computing the probability that the intersection of an s-set of k-dimensional subcubes chosen uniformly at random is non-empty is difficult (or at least unpleasant). It is for this reason that in Theorem 3.1 we give the thresholds for the uniform model in terms of the thresholds for the binomial model.

Proof of Theorem 3.1. The key observation is that we can view the binomial model as the result of a two stage random process. In the first stage we randomly partition the set of vertices V into sets $V_0, V_1 \dots V_d$, where $v \in V$ is included in V_i with probability $\binom{d}{i} p^i (1-p)^{d-i}$ independently at random for each v, i. In the second stage for each k we associate independently to each vertex in V_k a feature subcube of dimension k chosen uniformly at random, and then build the subcube intersection graph as normal. In particular, the restriction of $G_{V,d,p}$ to the (random) subset V_k is exactly (an instance of) the uniform model $G_{V_k,d,k}$. We shall use this to pull results back from the binomial model to the uniform model.

Let $\varepsilon > 0$ and $\alpha \in (0,1)$ be fixed, and let $k(d) = \lfloor \alpha d \rfloor$. Let $p = (\alpha_s)^{-1}(\alpha)$.

Upper bound: Pick η with $0 < \eta < \alpha$. Let $p' = p'(\alpha - \eta) = (\alpha_s)^{-1}(\alpha - \eta)$. Consider the binomial random subcube intersection graph $G_{[N],d,p'}$, and let $X = \frac{\log N}{d}$.

Suppose $X = t_s(p') + \varepsilon + o(1)$. By Corollary 2.2, $G_{[N],d,p'}$ then whp contains an s-clique. Further, by Proposition 2.14, whp there exists such a clique in which all subcubes have dimension at least $k_- = \lceil (\alpha - 2\eta)d \rceil$ and at most $k_+ = \lfloor \alpha d \rfloor$.

Let V' denote the set of vertices in $G_{[N],d,p}$ whose feature subcubes have dimension in the range $[k_-, k_+]$. Since the number of vertices with a feature subcube of a given dimension is a binomial random variable, a standard Chernoff bound shows that, for d large enough, the size of V' is whp at most

$$N' \leqslant 3\eta d N \binom{d}{\lfloor (\alpha-\eta)d \rfloor} p'^{\lfloor (\alpha-\eta)d \rfloor} (1-p')^{d - \lfloor (\alpha-\eta)d \rfloor}$$

For each $v \in V'$ with a feature subcube of dimension k' with $k_- \leqslant k' \leqslant k_+$, select a $(k_+ - k')$-subset of its fixed coordinates uniformly at random from all possibilities, and change those coordinates to wildcards $\star$ (i.e. to free coordinates). This gives a new feature subcube $f'(v)$ with dimension exactly k_+.

We now restrict our attention to the subcube intersection graph G defined by V' and the 'lifted' feature subcubes $(f'(v))_{v \in V'}$. Observe that the distribution on k_+-dimensional subcubes given by f' is exactly the uniform distribution. Thus G is in fact an instance of the uniform model G_{V',d,k_+}. Furthermore the 'lifting' procedure we performed on the feature subcubes $(f(v))_{v \in V'}$ has not destroyed any edge — indeed increasing the dimension of feature subcubes can only add edges — so that whp G contains an s-clique.

It follows that the threshold for the whp appearance of s-cliques in the uniform model with parameter $k_+ = \lfloor \alpha d \rfloor$ is at most

$$\begin{aligned} \frac{\log N'}{d} &\leqslant \frac{\log N}{d} + (\alpha - \eta) \log \frac{p'}{\alpha - \eta} + (1 - \alpha + \eta) \log \frac{1-p'}{1-\alpha+\eta} + o\left(\frac{\log d}{d}\right) \\ &= t_s(p') + (\alpha - \eta) \log \frac{p'}{\alpha - \eta} + (1 - \alpha + \eta) \log \frac{1-p'}{1-\alpha+\eta} + \varepsilon + o(1). \end{aligned}$$

Since $\varepsilon, \eta > 0$ were arbitrary, and since both p' and t_s are continuous functions (of $\alpha - \eta$ and $p' = (\alpha_s)^{-1}(\alpha - \eta)$ respectively), the threshold for the appearance of s-cliques in the

k_+-uniform model is at most

$$\lim_{\varepsilon,\eta\to 0^+} t_s(p') + (\alpha-\eta)\log\frac{p'}{\alpha-\eta} + (1-\alpha+\eta)\log\frac{1-p'}{1-\alpha+\eta} + \varepsilon$$
$$= t_s(p) + \alpha\log\frac{p}{\alpha} + (1-\alpha)\log\frac{1-p}{1-\alpha},$$

proving the claimed upper bound on $T_s(\alpha)$. (Recall that $p = (\alpha_s)^{-1}(\alpha) = \lim_{\eta\to 0} p'$.)
Lower bound: consider the binomial random subcube intersection graph $G_{[N],d,p}$, and let $X = \frac{\log N}{d}$.

Suppose $X = t_s(p) - \varepsilon + o(1)$. By Corollary 2.2, whp $G_{[N],d,p}$ contains no s-clique. In particular the subgraph of $G_{[N],d,p}$ induced by the set of vertices V' whose feature subcube have dimension $\lfloor \alpha d \rfloor$ is also K_s-free. As we observed, this random subgraph is identical in distribution to the random uniform subcube intersection graph $G_{V',d,k}$. Let $N' = |V'|$ be the number of vertices it contains.

By a standard Chernoff bound, whp

$$N' \geqslant \frac{1}{2} N \binom{d}{\lfloor \alpha d \rfloor} p^{\lfloor \alpha d \rfloor} (1-p)^{d-\lfloor \alpha d \rfloor}.$$

It follows that the threshold for the whp appearance of s-cliques in the uniform model $G_{V,d,k}$ is at least

$$t_s(p) - \varepsilon + \alpha\log\frac{p}{\alpha} + (1-\alpha)\log\frac{1-p}{1-\alpha} + o(1).$$

Since $\varepsilon > 0$ was arbitrary, the claimed lower bound on $T_s(\alpha)$ follows. □

3.3 The covering threshold

Proof of Theorem 3.2. This is very similar to the proof of Theorem 2.3. Assume without loss of generality that $V = [n]$. We let $\alpha \in (0,1)$ be fixed, set $k = k(d) = \lfloor \alpha d \rfloor$ and consider the uniform random subcube intersection graph $G_{[n],d,k}$.

Let $\mathbf{0}$ denote the all zero element $(0,0,\ldots,0)$ from Q_d. The expected number of elements of $Q_d = \{0,1\}^d$ not covered by the union of the feature subcubes $\bigcup_{v=1}^n f(v)$ is

$$\mathbb{E}\left|Q_d \setminus \bigcup_{v=1}^n f(v)\right| = |Q_d|\mathbb{P}\left(\mathbf{0} \notin \bigcup_{v=1}^n f(v)\right) = 2^d\left(1 - \frac{1}{2^{d-k}}\right)^n$$
$$= \exp\left(d\log 2 - \frac{n}{2^{d-k}}\left(1+o(1)\right)\right).$$

Now let ε be fixed with $0 < \varepsilon < \log 2$.
Upper bound: suppose $n = e^{xd} \geqslant 2^{d-k} d(\log 2 + \varepsilon)$. Then the expected number of uncovered elements of Q_d is $e^{-\varepsilon d + o(d)} = o(1)$, whence by Markov's inequality we have that whp $\bigcup_{v=1}^n f(v) = Q_d$, as desired.

Lower bound: suppose $n = e^{xd} = \lfloor 2^{d-k} d(\log 2 - \varepsilon) \rfloor$. Then the expected number of uncovered elements of Q_d is $e^{\varepsilon d + o(d)}$, which is large, and we use Chebyshev's inequality to show the actual number of uncovered elements is concentrated about this value. As before we compute the expectation of the square of the number of uncovered elements by considering pairs of points lying at Hamming distance i from one another. Let $\mathbf{e}_{[i]}$ denote the element of Q_d with 1 in the first i coordinates and 0 otherwise. In the following we take binomial coefficients of negative values to be 0.

$$\begin{aligned}\mathbb{E}\left(\left|Q_d \setminus \bigcup_{v=1}^{n} f(v)\right|^2\right) &= 2^d \sum_{i=0}^{d} \binom{d}{i} \mathbb{P}\left(\mathbf{0}, \mathbf{e}_{[i]} \notin \bigcup_{v=1}^{n} f(v)\right) \\ &= 2^d \sum_{i=0}^{d} \binom{d}{i} \left(1 - \frac{2}{2^{d-k}} + \frac{\binom{d-i}{k-i}}{\binom{d}{k}} \frac{1}{2^{d-k}}\right)^n \\ &= e^{2\varepsilon d} \sum_{i=1}^{d} \frac{\binom{d}{i}}{2^d} \exp\left(\frac{\binom{d-i}{k-i}}{\binom{d}{k}} d(\log 2 - \varepsilon) + o(1)\right).\end{aligned}$$

We now bound the sum above just as we did in the proof of Theorem 2.3, to show

$$\mathbb{E}\left(\left|Q_d \setminus \bigcup_{v=1}^{n} f(v)\right|^2\right) = (1 + o(1)) e^{2\varepsilon d}.$$

Since the details are similar, we omit them. We deduce just as in Theorem 2.3 that $\operatorname{Var} |Q_d \setminus \bigcup_{v=1}^{n} f(v)| = o\left(\mathbb{E}\,|Q_d \setminus \bigcup_{v=1}^{n} f(v)|^2\right)$. By Chebyshev's inequality whp at least $e^{\varepsilon d + o(d)} \gg 1$ elements of Q_d are not covered by $\bigcup_{v=1}^{n} f(v)$, as required. □

3.4 Above the covering threshold

Proof of Theorem 3.5. This is similar to the proof of Theorem 2.7. Without loss of generality, we may assume that $V = [n]$. Fix $\varepsilon > 0$ and $\alpha \in (0, 1)$. Let $k = k(d) = \lfloor \alpha d \rfloor$, and consider the random subcube intersection graph $G_{[n],d,k}$.

Upper bound: this case is in fact easier than for the binomial model. By the Helly property, $G_{[n],d,k}$ contains an s-clique if and only if some element x of the ambient hypercube Q_d is contained in at least s feature subcubes. Now in $G_{[n],d,k}$ every feature subcube has dimension k, thus certainly if $n \geqslant \frac{(s-1)|Q_d|}{2^k} + 1 = (s-1)2^{d-k} + 1$, we have by the pigeon-hole principle that some $x \in Q_d$ is contained in at least s feature subcubes, and thus $G_{[n],d,k}$ contains a copy of K_s.

Lower bound: suppose $n \leqslant (1 - \varepsilon) s 2^{d-k}$. Let $\mathbf{0}$ be the all zero element from Q_d. The number $C_{\mathbf{0}}$ of feature subcubes containing $\mathbf{0}$ is the sum of n independent identically distributed Bernoulli random variables with parameter $2^{-(d-k)}$. We have $\mathbb{E}C_{\mathbf{0}} = n2^{-(d-k)} \leqslant (1 - \varepsilon)s$. Applying a Chernoff bound, we deduce that

$$\mathbb{P}(C_{\mathbf{0}} \geqslant s) \leqslant e^{-\frac{\varepsilon^2}{3} s}.$$

In particular the expected number of elements of Q_d contained in at least s feature subcubes is at most $2^d e^{-\frac{\varepsilon^2}{3}s}$, which is $o(1)$ for $s \gg d$. It follows by Markov's inequality and the Helly property for subcube intersections that whp $G_{[n],d,k}$ contains no copy of K_s. □

4 Concluding remarks

4.1 Small p and large p

In this paper we focussed on the case where $p \in (0,1)$ is fixed (in the binomial model). Let us make here a few remarks about the small p and large p case.

Small p: note first of all that the proofs of Theorems 2.1, 2.4 and 2.7 all extend to the case when $p = p(d) \to 0$ as $d \to \infty$. Similarly, Theorems 3.2 and 3.5 for the uniform model also hold when $\alpha = \alpha(d) \to 0$ as $d \to \infty$. The proof of Theorem 3.1 does not, however, go through as it is stated in the paper — it needs stronger concentration for the dimension of the feature subcubes in Proposition 2.14.

There are two further remarks worth making concerning the small p case. First of all, as p (or α) tends to 0, the covering results Theorems 2.4 and 3.2 'converge' to the classical coupon collector problem. Taking $\Omega = Q_d$ and independently drawing random elements from Ω, we expect to make roughly $|\Omega| \log |\Omega| = 2^d d \log 2$ draws before we cover Ω, and this is the limit of $\left(\frac{2}{1+p}\right)^d d \log 2$ as $p \to 0$ (binomial model) and of $2^{(1-\alpha)d} d \log 2$ as $\alpha \to 0$ (uniform model).

Secondly, for the uniform model with constant parameter $k = 1$, the uniform model is closely related to bond percolation on the hypercube, which is a well-studied random graph model in its own right (see e.g. [12, 13]). On the other hand, the binomial model with parameter $p = \frac{1}{d}$ is different: the dimension of its feature subcubes have an approximatively Poisson distribution, and one does see feature subcubes of large bounded dimension. These will have an impact on the thresholds for lower-order cliques — indeed, a quick calculation shows that for s fixed, the expected dimension of feature subcubes in s-cliques of $G_{V,d,\frac{1}{d}}$ is $2 - \frac{1}{d} + O(\frac{1}{d^2})$.

Large p: in this case, we expect quasirandom behaviour from $G_{V,d,p}$ (behaviour 'like an Erdős–Rényi random graph'). We establish it below in the special case when $p = 1 - \varepsilon(d)$, with $\varepsilon(d)$ of order $\frac{1}{\sqrt{d}}$, when the edge-density is of constant order.

Proposition 4.1. *Let $\varepsilon(d)$ be a sequence of reals from the interval $[0,1]$ with $\varepsilon^2 d$ bounded away from both 0 and $+\infty$. Then for $p = 1 - \varepsilon$, with probability tending to 1 as $n \to \infty$ the graph $G_{[n],d,p}$ is quasirandom with parameter $e^{-\frac{\varepsilon^2 d}{2}}$.*

Proof. We shall use the celebrated quasirandomness theorem of Chung, Graham and Wilson [14], which states (amongst other things) that if the number of K_2 (*edges*) and the number of C_4 ($4-$*cycles*) contained in a graph are 'what you would expect if the graph was a typical Erdős–Rényi random graph with parameter q', then G is quasirandom with parameter q (we refer the reader to [14] for a formal definition of quasirandomness).

Let $\varepsilon(d)$ be sequence of positive real numbers as in the statement of the proposition, and let $q = e^{-\frac{\varepsilon^2 d}{2}}$. Set $p = 1 - \varepsilon$, and consider a labelled 4-set $\{v_1, v_2, v_3, v_4\}$ of vertices from $G_{[n],d,p}$. The probability that v_1v_2 forms an edge of the graph is

$$\mathbb{P}(\text{edge}) = \left(1 - \frac{(1-p)^2}{2}\right)^d = e^{-\frac{\varepsilon^2}{2}d + O(\varepsilon^3 d)} = q(1 + o(1)),$$

while the probability that all of the edges v_1v_2, v_2v_3, v_3v_4 and v_4v_1 are present in the graph is

$$\mathbb{P}(\text{4-cycle}) = \left(2\left(\frac{1+p}{2}\right)^4 - p^4 + p^2(1-p)^2\right)^d = e^{-2\varepsilon^2 d + O(\varepsilon^3 d)} = q^4(1 + o(1)).$$

We now verify that the numbers of edges $\#\{K_2\}$ and of 4-cycles $\#\{C_4\}$ are concentrated about their respective expectations. We appeal to the second moment method once more. For the edge K_2 we already established that $\mathrm{Var}\#\{K_2\} = o\left((\mathbb{E}\#\{K_2\})^2\right)$ in the proof of Theorem 2.1. Thus by Chebyshev's inequality, we have the required concentration: whp $G_{[n],d,p}$ contains $(1 + o(1))\mathbb{E}\#\{K_2\} = (1 + o(1))\binom{n}{2}q$ edges. Regarding 4-cycles, we have

$$\begin{aligned}\mathbb{E}\left(\#\{C_4\}\right)^2 &= \binom{n}{4}3\left(\binom{n-4}{4}3\mathbb{P}\left(\text{4-cycle}\right)^2 + O(n^3)\right)\\ &= \left(\mathbb{E}\#\{C_4\}\right)^2(1 + o(1)),\end{aligned}$$

by using in the second line the fact that $\mathbb{P}\left(\text{4-cycle}\right)^2 = e^{-2\varepsilon^2 d} \gg \frac{1}{n}$. We deduce that $\mathrm{Var}\#\{C_4\} = o\left((\mathbb{E}\#\{C_4\})^2\right)$ and that the number of 4-cycles in $G_{[n],d,p}$ is concentrated about its mean: whp

$$\#\{C_4\} = (1 + o(1))\frac{n(n-1)(n-2)(n-3)}{8}q^4.$$

Our proposition then follows from the quasirandomness theorem of Chung, Graham and Wilson [14]. □

Thus in this case the binomial model behaves like an Erdős–Rényi random graph. It is not hard to use this to show that the uniform model with parameter $k = d - O(\sqrt{d})$ is also quasirandom. As such, there is nothing very novel about our models in this range.

4.2 Further questions

There are a number of natural questions to ask about our models.

Connectivity: when do random subcube intersection graphs become whp connected? In the binomial model with $p \leqslant \frac{1}{3}$, it can be shown that $t_{\text{connected}} = \log\frac{2}{1+p}$, coinciding with the covering threshold. For the upper bound, the rough idea is to show that above the covering threshold, whp every 1-dimensional subcube of Q_d is covered by a feature

subcube, which implies connectivity of the intersection graph. For the lower bound, on the other hand, one uses standard probabilistic techniques to establish the whp existence of isolated feature subcubes of small dimension.

On the other hand for $p > \frac{1}{3}$ the situation becomes more complicated, and one will have to analyse the structure of $G_{V,d,p}$ in greater detail. Indeed, the lower bound ceases to be valid, as 'small' feature subcubes do not appear fast enough to provide us with isolated vertices. We believe that in this case the connectivity threshold of $G_{V,d,p}$ will be determined by the disappearance of isolated vertices from the subgraph $G_{V_k,d,k}$ induced by the set V_k of vertices with feature subcubes of dimension $k = (2p/(1+2p-p^2) + o(1))\, d$.

The heuristic behind this belief is that the binomial model $G_{V,d,p}$ has a kind of 'jellyfish' structure, with strands of vertices with 'small' feature subcubes hanging off a 'head' of vertices with 'large' feature subcubes. More precisely, consider an αd-dimensional subcube S, and let S' be a binomial random subcube of Q_d with parameter p. A quick calculation tells us that conditional on $S \cap S' \neq \varnothing$, the expected dimension of S' is in fact

$$p\alpha d + \frac{2p}{1+p}(1-\alpha)d = \left(p + \frac{p(1-p)}{1+p}(1-\alpha)\right) d := (p + h_p(\alpha))d.$$

Thus we have a 'dimensional bonus' in $G_{V,d,p}$: the expected dimension of the feature subcube of a neighbour of a vertex v is significantly larger than the average dimension pd. This 'bonus' $h_p(\alpha)d$ decreases monotonically as the dimension αd of $f(v)$ increases, with $\lim_{\alpha\to 1} h_p(\alpha) = 0$. The unique solution to $\alpha = p + h_p(\alpha)$ is attained at $\alpha = \alpha_\star(p)$ given by

$$\alpha_\star := \frac{2p}{1+2p-p^2}.$$

The picture we have then of $G_{V,d,p}$ is of feature subcubes of dimension less than $\alpha_\star d$ hanging off feature subcubes of dimension greater than $\alpha_\star d$. We expect therefore the connectivity of the vertices with roughly $\alpha_\star d$-dimensional feature subcubes to determine the connectivity of the entire graph.

In particular, to understand the connectivity of the binomial model with $p > 1/3$ we shall first need to understand the connectivity of the uniform model with parameter $k = \alpha_\star d$, for which our best guess is that connectivity occurs as soon as the last isolated vertex has acquired a neighbour.

Component evolution: in our random subcube intersection graph models, how does the order of a largest connected component evolve as we add vertices? when does a giant component (i.e. a component containing a fixed proportion $\varepsilon > 0$ of the vertices) emerge? For the binomial model, we expect that 'jellyfish' structure of our random graphs shall play an important role here again. Most feature subcubes in $G_{V,d,p}$ have dimension $(p+o(1))d$, and most of their neighbours have feature subcubes of dimension $(\beta+o(1))d$, where $\beta = p + p(1-p)^2/(1+p)$. Thus we would expect that the component evolution inside the uniform model with parameters d and $k = \beta d$ will essentially determine the component evolution of the binomial model $G_{V,d,p}$.

Regarding the uniform model, a reasonable guess would be that, following the Erdős–Rényi random graph paradigm, a giant component emerges around the same time as

the average degree in the graph hits 1. Such a phenomenon is known to hold for site percolation on many families of graphs; see the recent work of Krivelevich [28] on the subject, whose tools may prove relevant to our problem.

Motivation for considering the connectivity and the component evolution of random intersection graphs comes from the study of the geometry of the solution space in random k-SAT.

Independence and chromatic numbers: it remains a completely open problem to understand graph-theoretic independence in random subcube intersection graphs. We do not know how to track the independence number (the size of a largest independent set of vertices) in our models, and more generally we do not know how to perform anything but the most basic computations involving non-edges. Perhaps recent development in the theory of containers together with an analysis of the graph-theoretic properties of G_d^k, the intersection graph of all k-dimensional subcubes of Q_d, could lead to progress on this question. Further afield, we have a lower bound on the chromatic number of random subcube intersection graphs coming from the clique number, but no non-trivial upper bound. It would be nice to have a better understanding of the behaviour of these standard graph parameters for our models.

Bias: given our motivation for studying subcube intersection graphs, it would be desirable to allow some bias in the distribution of the feature subcubes. For instance, in a polarised society with two-party politics it is likely citizens will have either mostly zeroes ('left-wing opinions') or mostly ones ('right-wing opinions') amongst their opinions. It would be interesting to study the change in the behaviour of our models as the polarisation becomes stronger. Some progress on the covering problem in the presence of bias was recently made by Larsson and Markström [29].

Generalised coupon collector: our generalised version of the coupon collector, Problem 2.15, has been investigated in [20], but many questions remain open. How does the structure of the distribution of the coupon sets X determine the distribution of the covering time T? While the paper [20] gives several criteria for behaviour like/unlike that of a uniform coupon collector, we are still far from a complete answer to this question.

Acknowledgements

The authors would like to thank Joel Larsson and two anonymous referees for their helpful comments which have led to great improvements in the presentation of this paper.

References

[1] D. Achlioptas, A. Naor, and Y. Peres. Rigorous location of phase transitions in hard optimization problems. *Nature*, 435(7043):759–764, 2005.

[2] D.J. Aldous. Threshold limits for cover times. *Journal of Theoretical Probability*, 4(1):197–211, 1991.

[3] M. Behrisch. Component evolution in random intersection graphs. *Electronic Journal of Combinatorics*, 14:R17, 2007.

[4] M. Behrisch, A. Taraz, and M. Ueckerdt. Coloring random intersection graphs and complex networks. *SIAM Journal on Discrete Mathematics*, 23(1):288–299, 2009.

[5] D.E. Berg, S. Norine, F.E. Su, R. Thomas, and P. Wollan. Voting in agreeable societies. *American Mathematical Monthly*, 117(1):27–39, 2010.

[6] S.R. Blackburn and S. Gerke. Connectivity of the uniform random intersection graph. *Discrete Mathematics*, 309(16):5130–5140, 2009.

[7] M. Bloznelis, J. Jaworski, and K. Rybarczyk. Component evolution in a secure wireless sensor network. *Networks*, 53(1):19–26, 2009.

[8] M. Bloznelis and V. Kurauskas. Large cliques in sparse random intersection graphs. *ArXiv preprint*, arXiv:1302.4627, 2013.

[9] M. Bloznelis and I. Radavičius. On hamiltonicity of uniform random intersection graphs. *Liet. Mat. Rink*, 51:443–447, 2010.

[10] B. Bollobás, S. Janson, and O. Riordan. The phase transition in inhomogeneous random graphs. *Random Structures & Algorithms*, 31(1):3–122, 2007.

[11] B. Bollobás, S. Janson, and O. Riordan. Sparse random graphs with clustering. *Random Structures & Algorithms*, 38(3):269–323, 2011.

[12] B. Bollobás, Y. Kohayakawa, and T. Łuczak. The evolution of random subgraphs of the cube. *Random Structures & Algorithms*, 3(1):55–90, 1992.

[13] C. Borgs, J.T. Chayes, R. Van der Hofstad, G. Slade, and J. Spencer. Random subgraphs of finite graphs: III. the phase transition for the n-cube. *Combinatorica*, 26(4):395–410, 2006.

[14] F.R. Chung, R.L. Graham, and R.M. Wilson. Quasi-random graphs. *Combinatorica*, 9(4):345–362, 1989.

[15] A. Coja-Oghlan. The asymptotic k-sat threshold. In *Proceedings of the 46th Annual ACM Symposium on Theory of Computing*, STOC '14, pages 804–813, New York, NY, USA, 2014. ACM.

[16] P. Diaconis, S. Holmes, and S. Janson. Interval graph limits. *Annals of combinatorics* 17(1): 27–52, 2013.

[17] J. Ding, A. Sly, and N. Sun. Proof of the satisfiability conjecture for large k. *ArXiv preprint*, arXiv:1411.0650, November 2014.

[18] C. Efthymiou and P. Spirakis. On the existence of Hamiltonian cycles in random intersection graphs. *Automata, Languages and Programming*, pp 99–99, 2005.

[19] P. Erdős and A. Rényi. On a classical problem of probability theory. *Magyar Tud. Akad. Mat. Kutató Int. Közl*, 6(1-2):215-?220, 1961.

[20] V. Falgas-Ravry, J. Larsson, and K. Markström. Speed and concentration of the covering time for structured coupon collectors. *ArXiv preprint*, arXiv:1601.04455, 2016.

[21] J.A. Fill, E.R. Scheinerman, and K.B. Singer-Cohen. Random intersection graphs when $m = \omega(n)$: an equivalence theorem relating the evolution of the G(n,m,p) and G(n,p) models. *Random structures and Algorithms*, 16(2):156–176, 2000.

[22] E. Friedgut. Sharp thresholds of graph properties, and the k-sat problem. *J. Amer. Math. Soc.*, 12(4):1017–1054, 1999. With an appendix by Jean Bourgain.

[23] E. Godehardt and J. Jaworski. On the connectivity of a random interval graph. *Random Structures & Algorithms*, 9(1-2):137–161, 1998.

[24] E. Godehardt, J. Jaworski, and K. Rybarczyk. Random intersection graphs and classification. *Advances in data analysis*, pages 67–74, 2007.

[25] J.R. Johnson and K. Markström. Turán and Ramsey properties of subcube intersection graphs. *Combinatorics, Probability and Computing*, 22:55–70, 0 2013.

[26] M. Karoński, E.R. Scheinerman, and K.B. Singer-Cohen. On random intersection graphs: the subgraph problem. *Combinatorics, Probability and Computing*, 8(1-2):131–159, 1999.

[27] S. Kirkpatrick and B. Selman. Critical behavior in the satisfiability of random boolean expressions. *Science*, 264(5163):1297–1301, 1994.

[28] M. Krivelevich. The phase transition in site percolation on pseudo-random Graphs. *The Electronic Journal of Combinatorics* 23(1): 1–12, 2016.

[29] J. Larsson. On random cover and matching problems. *Licentiate Thesis*, Umeå Universitet, 2016.

[30] T.A. McKee and F.R. McMorris. Topics in intersection graph theory. *SIAM monographs on Discrete Mathematics and Applications*, 1999.

[31] K. Nehring and C. Puppe. The structure of strategy-proof social choice — Part I: General characterization and possibility results on median spaces. *Journal of Economic Theory* 135(1):269–305, 2007.

[32] D.J. Newman. The double dixie cup problem. *The American Mathematical Monthly*, 67(1):58–61, 1960.

[33] S. Nikoletseas, C. Raptopoulos, and P. Spirakis. Large independent sets in general random intersection graphs. *Theoretical Computer Science*, 406(3):215–224, 2008.

[34] S. Nikoletseas, C. Raptopoulos, and P. Spirakis. Colouring non-sparse random intersection graphs. In *Mathematical Foundations of Computer Science 2009*, pages 600–611. Springer, 2009.

[35] N. Pippenger. Random interval graphs. *Random Structures and Algorithms*, 12(4):361–380, 1998.

[36] T. Pinto Biclique Covers and Partitions. *The Electronic Journal of Combinatorics*, 21(1): 1–19, 2014.

[37] K. Rybarczyk. Diameter, connectivity, and phase transition of the uniform random intersection graph. *Discrete Mathematics*, 311(17):1998–2019, 2011.

[38] K. Rybarczyk. Equivalence of a random intersection graph and G(n, p). *Random Structures & Algorithms*, 38(1-2):205–234, 2011.

[39] K. Rybarczyk and D. Stark. Poisson approximation of the number of cliques in random intersection graphs. *Journal of Applied Probability*, 47(3):826–840, 2010.

[40] E.R. Scheinerman. Random interval graphs. *Combinatorica*, 8(4):357–371, 1988.

[41] E.R. Scheinerman. An evolution of interval graphs. *Discrete Mathematics*, 82(3):287–302, 1990.

[42] K.B. Singer-Cohen. *Random intersection graphs*. PhD thesis, Johns Hopkins University, 1995.

[43] D. Stark. The vertex degree distribution of random intersection graphs. *Random Structures & Algorithms*, 24(3):249–258, 2004.

Appendix A: proof of Lemma 2.9

Lemma 2.9. *For all $y \in [0,1]$ and all integers $0 \leqslant i \leqslant s$, the following inequality holds:*

$$\left(2 + 2y^i - 4y^s + y^{2s-i}\right)^s \leqslant \left(2 - y^s\right)^{2s-i}.$$

Proof. This is a simple (albeit lengthy) exercise in calculus. Let $h_1(y) = 2 + 2y^i - 4y^s + y^{2s-i}$. Since we may write $h_1(y)$ as

$$h_1(y) = 2(1 - y^s) + 2y^i(1 - y^{s-i}) + y^{2s-i},$$

we have $h_1(y) > 0$ for $y \in [0,1]$. The function

$$g(y) = \frac{\left(2 - y^s\right)^{2s-i}}{\left(2 + 2y^i - 4y^s + y^{2s-i}\right)^s}$$

is thus well-defined and differentiable in the interval $[0,1]$.

We want to show that $g(y) \geqslant 1$ for all $y \in [0,1]$. If $i = s$ or if $i = 0$ the function g is identically 1 on the interval $[0,1]$, in which case we have nothing to prove. Assume therefore in what follows that $0 < i < s$. We have $g(0) = 2^{s-i} > 1$ and $g(1) = 1$, so we will be done if we can show that the function g is monotone decreasing in the interval $[0,1]$. We compute the derivative of g:

$$\begin{aligned}
g'(y) &= \frac{1}{(h_1(y))^{s+1}}\Big(-s(h_1(y))'(2 - y^s)^{2s-i} + h_1(y)((2 - y^s)^{2s-i})'\Big) \\
&= \frac{-sy^{i-1}(2 - y^s)^{2s-i-1}}{(h_1(y))^{s+1}}\Big(4i - (4s + 2i)y^{s-i} + (4s - 2i)y^{2s-2i} \\
&\qquad\qquad + (4s - 4i)y^s - (4s - 4i)y^{2s-i}\Big) \\
&= -\frac{sy^{i-1}(2 - y^s)^{2s-i-1}(1 - y^{s-i})}{(h_1(y))^{s+1}}\Big(4i - (4s - 2i)y^{s-i} + (4s - 4i)y^s\Big).
\end{aligned}$$

We claim $g'(y) \leqslant 0$ for all $y \in [0,1]$. Clearly

$$-\frac{sy^{i-1}(2-y^s)^{2s-i-1}(1-y^{s-i})}{(h_1(y))^{s+1}} \leqslant 0$$

for all $y \in [0,1]$. Thus the only factor we have left to consider is

$$h_2(y) = 4i - (4s-2i)y^{s-i} + (4s-4i)y^s.$$

Claim. *$h_2(y) > 0$ for all $y \in [0,1]$.*

The claim above implies that $g'(y) \leqslant 0$ for all $y \in [0,1]$, whence $g(y) \geqslant g(1) = 1$ for all $y \in [0,1]$, as desired.

Proof of Claim. We have $h_2(0) = 4i$ and $h(1) = 2i$, both of which are strictly positive. We differentiate h_2 to check for other extrema inside the interval $[0,1]$.

$$h_2'(y) = (s-i)y^{s-i-1}\left(-(4s-2i) + 4sy^i\right).$$

In addition to $y = 0$, h_2' has one root in the interval $[0,1]$, namely

$$y_\star = \left(1 - \frac{i}{2s}\right)^{\frac{1}{i}}.$$

At $y_\star$, we have

$$h_2(y_\star) = 4i\left(1 - \left(1 - \frac{i}{2s}\right)^{\frac{s}{i}}\right) > 0.$$

Thus for all $y \in [0,1]$,

$$h_2(y) \geqslant \min\Big(h_2(0), h_2(y_\star), h_2(1)\Big) > 0,$$

establishing the claim. □

This completes the proof of Lemma 2.9. □