Supersaturation for hereditary properties

David Saxton *

January 20, 2013

Abstract

Let $F$ be a collection of $r$-uniform hypergraphs, and let $0 < p < 1$. It is known that there exists $c = c(p, F)$ such that the probability of a random $r$-graph in $G(n, p)$ not containing an induced subgraph from $F$ is $2^{c+o(1)/n}$. Let each graph in $F$ have at least $t$ vertices. We show that in fact for every $\epsilon > 0$, there exists $\delta = \delta(\epsilon, p, F) > 0$ such that the probability of a random $r$-graph in $G(n, p)$ containing less than $\delta nt$ induced subgraphs each lying in $F$ is at most $2^{c+\epsilon/\epsilon(n)}$. This statement is an analogue for hereditary properties of the supersaturation theorem of Erdős and Simonovits. In our applications we answer a question of Bollobás and Nikiforov.

1 Hereditary properties

Let $F$ be a collection of $r$-uniform hypergraphs (which we abbreviate to $r$-graphs). Let $P = \text{Forb}(F)$ be the collection of all $r$-graphs not containing an induced subgraph from $F$. $P$ is a hereditary property: it is a collection of graphs closed under graph isomorphism and under taking induced subgraphs. Let $P^n \subset P$ be the set of these graphs on $n$ vertices. Let $G(n, p)$ be a random $r$-graph on $n$ vertices where each edge is included uniformly and independently with probability $p$.

Proposition 1 (Alekseev [1], Bollobás and Thomason [4]). Let $P$ be a hereditary property for $r$-graphs. Define $c_n$ via

$$Pr[G(n, p) \in P^n] = 2^{-c_n}.$$ 

Then the limit $\lim_{n \to \infty} c_n$ exists.

(Strictly speaking, Alekseev only proved the case $r = 2$ and $p = 1/2$, but his argument shows that Proposition 1 follows from the Erdős-Hanani conjecture as proved by Rödl [10]. Bollobás and Thomason [4] show that in fact $c_n$ is increasing in $n$.)

For the property $P = \text{Forb}(F)$, let $c(p, F)$ be the limit in Proposition 1. For $p = 1/2$, the above probability is exactly proportional to the number of graphs without an induced subgraph from $F$.

The case $r = 2$ has been studied extensively. Prömel and Steger [9] showed that $c(1/2, \{F\}) = 1/t$, where $t$ is the maximum integer for which there exists

*Supported by the Engineering and Physical Sciences Research Council.
than Theorem 2.

a more direct analogue of the supersaturation theorem of Erdős and Simonovits have the following immediate corollary of Theorem 2, which can be considered ∗

Write \( \text{ex}(\text{G}(n, p)) \in P \) such that the number of generated graphs is 2

\( G \) there exist graphs determined by the number of subgraphs of a single graph. More specifically, and Nikiforov \([3]\), using results relying on Szemerédi’s Regularity Lemma. They be done by Theorem 2 together with the proof of c

\( \text{ex}(\text{G}(n, p)) \in P \) such that any \( r \)

Then for every \( \epsilon > 0 \) there exist \( n_0 \) and \( \delta > 0 \) (depending only on \( \epsilon, p, \mathcal{F} \)) such that if \( \mathcal{A} \) is a collection of \( r \)-graphs on \( n > n_0 \) vertices with \( \text{Pr}[\text{G}(n, p) \in \mathcal{A}] > 2^{(-c+\epsilon)(r)} \),

then some graph in \( \mathcal{A} \) contains at least \( \delta n^r \) induced subgraphs each lying in \( \mathcal{F} \).

2 Supersaturation

Our main theorem is that, loosely speaking, the probability of containing a positive density of induced subgraphs each lying in \( \mathcal{F} \) is not that much less than the probability of containing a single induced subgraph lying in \( \mathcal{F} \).

Theorem 2. Let \( \mathcal{F} \) be a collection of \( r \)-graphs each with at least \( t \) vertices, let \( 0 < p < 1 \) and let \( c = c(p, \mathcal{F}) \) be as above, i.e., letting \( \mathcal{P} \) be the set of \( r \)-graphs not containing an induced subgraph lying in \( \mathcal{F} \),

\[
\text{Pr}[\text{G}(n, p) \in \mathcal{P}] = 2^{(-c+\epsilon)(1)}(r).
\]

Then for every \( \epsilon > 0 \) there exist \( n_0 \) and \( \delta > 0 \) (depending only on \( \epsilon, p, \mathcal{F} \)) such that if \( \mathcal{A} \) is a collection of \( r \)-graphs on \( n > n_0 \) vertices with \( \text{Pr}[\text{G}(n, p) \in \mathcal{A}] > 2^{(-c+\epsilon)(r)} \),

we wish to draw a parallel with the supersaturation theorem of Erdős and Simonovits \([7]\) which says the following. For a collection of forbidden \( r \)-graphs \( \mathcal{F} \) each on at least \( t \) vertices, let \( \text{ex}(n, \mathcal{F}) \) be the maximum number of edges of an \( r \)-graph on \( n \) vertices containing no copy of any graph in \( \mathcal{F} \) (not necessarily induced), and let \( \gamma = \lim_{n \to \infty} \text{ex}(n, \mathcal{F})(r)^{-1} \) be the extremal density limit. Then for every \( \epsilon > 0 \) there exists \( \delta > 0 \) and \( n_0 \) (depending only on \( \epsilon \) and \( \mathcal{F} \)) such that any \( r \)-graph with more than \( (\gamma + \epsilon)(r) \) edges on \( n > n_0 \) vertices contains at least \( \delta n^r \) subgraphs each lying in \( \mathcal{F} \).

The case \( p = 1/2, r = 2, \mathcal{F} = \{F\} \) of Theorem 2 was proved by Bollobás and Nikiforov \([3]\), using results relying on Szemerédi’s Regularity Lemma. They ask whether a proof could be given avoiding the regularity lemma. We show that this is indeed possible. (More specifically, for the result they state this can be done by Theorem 2 together with the proof of \( c = 1/t \) by Alekseev \([2]\).)

Theorem 2 also generalizes a theorem of Erdős, Rothschild and Kleitman \([6]\), where they prove the case \( p = 1/2 \) and \( \mathcal{F} = \{K_1\} \), a complete graph on \( t \) vertices.

In \([9]\) Prömel and Steger prove that in fact the number of graphs on a vertex set \( V \) of size \( |V| = n \) not containing an induced \( F \) subgraph is essentially determined by the number of subgraphs of a single graph. More specifically, there exist graphs \( G = (V, E) \) and \( G_0 = (V, E_0) \) with \( E \cap E_0 = \emptyset \) such that every graph \( (V, E_0 \cup X) \) with \( X \subset E \) does not contain an induced \( F \) subgraph, and the number of generated graphs is \( 2^{\left| E \right|} = 2^{(1-c)(r)+o(n^2)} \) with \( c = c(1/2, \{F\}) \). Write \( \text{ex}^*(n, F) \) for the maximum number of edges \( |E| \) of such a graph \( G \). We have the following immediate corollary of Theorem 2 which can be considered a more direct analogue of the supersaturation theorem of Erdős and Simonovits than Theorem 2.
Corollary 3. Let $F$ be a 2-graph and let $\epsilon > 0$. Then there exist $n_0$ and $\delta > 0$ (depending only on $\epsilon, F$) such that for a vertex set $V$ of size $n = |V| > n_0$, if $G = (V, E)$ is a graph on $ex'(n, F) + \epsilon \binom{n}{2}$ edges then for every set of edges $E_0 \subset V^{(2)} \setminus E$ there exists a subset $X \subset E$ such that the graph $(V, E_0 \cup X)$ contains at least $\delta n|F|$ induced copies of $F$.

3 Proof of Theorem 2

A partial Steiner system with parameters $(r, m, n)$ for a vertex set $V$ of size $n$ is a collection of sets $D \subset V^{(m)}$ such that every $r$-element subset of $V$ appears at most once as a subset of a set in $D$. Observe that

$$|D| \leq \binom{n}{r} \binom{m}{r}^{-1}.$$ 

As proved by Rödl [10], there exist partial Steiner systems which cover almost all $r$-element subsets.

Proposition 4 (Rödl). For $r < m$, $\lambda > 0$, there exists $n_0$ such that for every $n > n_0$, there exists a partial Steiner system $D$ with parameters $(r, m, n)$ such that

$$|D| \geq (1 - \lambda) \binom{n}{r} \binom{m}{r}^{-1}.$$ 

Let $F, p, \epsilon, c, A, t$ be as in Theorem 2, and let the common vertex set of the graphs in $A$ be $V$. For a graph $G$ and a subset $D$ of the vertices of $G$, write $G[D]$ for the induced subgraph on the vertex set $D$. Write also $F < G$ to denote that $G$ contains an induced subgraph lying in $F$. For a collection of graphs $C$ on a common vertex set of size $k$, write

$$\mu_n(C) = \Pr[G(k, p) \in C]$$

for the measure of the set $C$ in the space $G(k, p)$. Thus $\mu_n(A) > 2^{(-c+\epsilon)(\cdot)}$.

Lemma 5. There exist $\eta, \gamma, \lambda > 0$ and an integer $m$ (depending only on $F, p, \epsilon$) such that the following is true. Let $D = \{D_1, \ldots, D_d\}$ be a partial Steiner system with parameters $(r, m, n)$ on vertex set $V$ with $d \geq (1 - \lambda) \binom{n}{r} \binom{m}{r}^{-1}$. Let $I = \{i \in [d] : \mu_n(\{G \in A : F < G[D_i]\}) \geq \gamma \mu_n(A)\}$. Then $|I| \geq \eta d$.

Proof. For $m \geq 1$, let $B$ be the set of graphs on vertex set $[m]$ that do not contain an induced subgraph lying in $F$. Then $\mu_m(B) = 2^{(-c+\epsilon')(\cdot)}$ for some $\epsilon' \to 0$ as $m \to \infty$. Fix $m$ sufficiently large such that $\epsilon'$ is sufficiently small (to be determined later).

We will choose $\lambda > 0$ later. Partition $A$ as $A = \cup_{S \subset [d]} A_S$, where

$A_S = \{G \in A : \{i : F < G[D_i]\} = S\}$.

Let $\theta_i$ be the measure of the set of graphs $G \in A$ such that $F < G[D_i]$, so

$$\theta_i = \sum_{S \supseteq i} \mu_n(A_S).$$
In this notation, \( I = \{ i \in [d] : \theta_i \geq \gamma \mu_n(A) \} \). Let \( \eta = |I|/d \). We aim to show that we can take \( \eta > 0 \) (independent of \( n \)). Observe that
\[
\sum_{S \subseteq [d]} \Delta_{\mu_n(A_S)} = \sum_{i \in [d]} \theta_i \leq (\eta d) \mu_n(A) + (1 - \eta) \gamma d \mu_n(A).
\] (1)

Observe also that \( \mu_n(A_S) \leq \mu_n(B)^{d - |S|} \) (since the projection of \( A_S \) onto any \( D_i, i \in S \) is contained inside a copy of \( B \) on \( D_i \)). Hence
\[
\sum_{S: |S| < \nu d} \mu_n(A_S) \leq \sum_{i=0}^{\nu d} \binom{d}{i} \mu_n(B)^{d-i}
\leq \nu d \binom{d}{\nu d} \mu_n(B)^{(1-\nu)d}
\leq 2^{O(\nu d + (c + \epsilon')(1-\nu)(1-\lambda)(\nu)}
\] (2)

where \( O(\nu) \to 0 \) as \( \nu \to 0 \). Since \( \mu_n(A) = 2^{(-c + \epsilon')(\nu)} \), we may pick \( \epsilon', \nu, \lambda > 0 \) sufficiently small such that the quantity in (2) is at most \( \mu_n(A)/2 \). Thus by (1),
\[
(\eta d) \mu_n(A) + ((1 - \eta) \gamma d) \mu_n(A) \geq \sum_{S: |S| \geq \nu d} |S| \mu_n(A_S) \geq (\nu d) \mu_n(A)/2,
\]
i.e., \( \nu/2 \leq \eta + (1 - \eta) \gamma \). Set \( \gamma = \nu/4 \); this gives \( \eta \geq (\nu/4)/(1 - \nu/4) > 0 \) as required. \( \square \)

We are now ready to prove Theorem 2.

**Proof.** Let \( \eta, \gamma, \lambda, m \) be as in Lemma 5. Let \( n \) be sufficiently large for the existence of an \( (r, m, n) \) partial Steiner system \( \mathcal{D} \) covering a proportion of at least \( 1 - \lambda \) of the \( r \)-subsets of \( V \). Let
\[
X = \{ D \in V^{(m)} : \mu_n(\{ G \in \mathcal{A} : F < G[D]\}) \geq \gamma \mu_n(A) \}.
\]

Let \( \sigma \) be a randomly and uniformly chosen permutation of \( V \), and let \( \mathcal{D}_\sigma \) be the partial Steiner system generated from \( \mathcal{D} \) by permuting the vertex set \( V \) by \( \sigma \). Applying Lemma 5 with \( \mathcal{D}_\sigma \) and taking expectations shows that \( |X| \geq \eta \gamma \binom{n}{m} \).

In particular some graph \( G \in \mathcal{A} \) contains at least \( \eta \gamma \binom{n}{m} \) \( m \)-sets containing an induced subgraph lying in \( F \). Each fixed copy of an \( F \in \mathcal{F} \) is included in at most \( \binom{n-t}{m-t} \) \( m \)-sets. Hence \( G \) contains at least
\[
\gamma \eta \binom{n}{m} \binom{n-t}{m-t}^{-1} \geq \gamma \eta (2m)^{-t} n^t
\]
distinct induced subgraphs each lying in \( F \) (provided \( n \geq 2t \)). We can therefore take \( \delta = \gamma \eta (2m)^{-t} \), independent of \( n \), as required. \( \square \)

**Acknowledgements** The author would like to thank Andrew Thomason for his many helpful suggestions.
References

[1] V.E. Alekseev, Hereditary classes and coding of graphs (in Russian), *Probl. Cybern.* 39 (1982), 151–164.

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

[3] B. Bollobás and V. Nikiforov, The number of graphs with large forbidden subgraphs, *European J. Combin.* 31 (2010), 1964–1968.

[4] B. Bollobás and A. Thomason, Projections of bodies and hereditary classes of hypergraphs, *Bull. London Math. Soc.* 27 (1995), 416–424.

[5] B. Bollobás and A. Thomason, The structure of hereditary properties and colourings of random graphs, *Combinatorica* 20 (2000), 173–202.

[6] P. Erdős, D.J. Kleitman, and B.L. Rothschild, Asymptotic enumeration of $K_n$-free graphs, in *Colloquio Internazionale sulle Teorie Combinatorie*. Accad, Naz. Lincei, Rome (1976) 19–27.

[7] P. Erdős and M. Simonovits, Supersaturated graphs and hypergraphs, *Combinatorica* 3 (1983), 181–192.

[8] E. Marchant and A. Thomason, The structure of hereditary properties and 2-coloured multigraphs, *Combinatorica*.

[9] H.J. Prömel and A. Steger, Excluding induced subgraphs III: a general asymptotic, *Rand. Struct. Alg.* 3 (1992), 19–31.

[10] V. Rödl, On a packing and covering problem, *European J. Combin.* 6 (1985), 69–78.