On smoothed analysis in dense graphs and formulas

Michael Krivelevich* Benny Sudakov † Prasad Tetali ‡

Abstract

We study a model of random graphs, where a random instance is obtained by adding random edges to a large graph of a given density. The research on this model has been started by Bohman et al. in [3], [4]. Here we obtain a sharp threshold for the appearance of a fixed subgraph, and for certain Ramsey properties. We also consider a related model of random k-SAT formulas, where an instance is obtained by adding random k-clauses to a fixed formula with a given number of clauses, and derive tight bounds for the non-satisfiability of thus obtained random formula.

1 Introduction

In this paper we will be mostly concerned with the following model of random graphs: a fixed graph $H = (V, E)$ on n vertices is given and a set R of m edges is chosen uniformly at random from the set of all $\binom{n}{2} - |E|$ non-edges of H to form a random graph

$$G_{H,m} = (V, E \cup R)\,.$$

This rather novel model of random graphs was introduced by Bohman, Frieze and Martin in [3], its further properties have been investigated by Bohman, Frieze, Krivelevich and Martin in [4]. It can be viewed as a natural generalization of the classical Erdős-Rényi model $G_{n,m}$ of random graphs [8]. Indeed, when the graph H is taken to be the empty graph on n vertices, the model $G_{H,m}$ reduces to $G_{n,m}$. Another close relative is what is sometimes called the network reliability model, where a random graph is formed by taking a random subset of edges of a fixed graph H; in fact, those two models are easily seen to be equivalent by taking complements. An important piece of motivation, explaining the title of the paper, comes from a relatively new and rapidly developing concept of smoothed analysis introduced by Spielman and Teng [14]. In the smoothed analysis model, a random instance is usually generated by performing a small perturbation of a given instance. An

*Department of Mathematics, Raymond and Beverly Sackler Faculty of Exact Sciences, Tel Aviv University, Tel Aviv 69978, Israel. E-mail: krivelev@post.tau.ac.il. Research supported in part by a USA-Israel BSF Grant and by a grant from the Israel Science Foundation.

†Department of Mathematics, Princeton University, Princeton, NJ 08544, USA. Email address: bsudakov@math.princeton.edu. Research supported in part by NSF grants DMS-0355497, DMS-0106589, and by an Alfred P. Sloan fellowship.

‡School of Mathematics and College of Computing, Georgia Institute of Technology, Atlanta, GA 30332, USA. Email address: tetali@math.gatech.edu. Research supported in part by NSF grants DMS-0100298, DMS-0401239.

analogous concept for graphs would be to add/delete a small set of random edges. However, as it frequently turns out, deletion of a small set of edges normally does not affect the existence of a desired property; we can thus restrict a small perturbation to addition of random edges alone.

A typical question one usually considers in the model $G_{H,m}$ is as follows. Let $\mathcal{A}$ be a monotone property of graphs on n vertices (where as usually a monotone property is a family of graphs on n vertices closed under adding edges). How many random edges need to be added to H so that the resulting random graph $G_{H,m}$ will almost surely have property $\mathcal{A}$? As customary in the theory of random graphs, we take an asymptotic viewpoint here, assuming that the number of vertices n tends to infinity. Also, an event holds *almost surely*, or a.s. for brevity, if its probability tends to 1 as n tends to infinity. Bohman, Frieze and Martin considered Hamiltonicity in [3], further properties such as appearance of a fixed sized clique, diameter and connectivity have been treated in [4].

As it appears, the model $G_{H,m}$ is too general to get concrete and meaningful results, valid for all graphs H. It is thus desirable to restrict H to belong a certain class of graphs. One natural candidate is the class $\mathcal{G}(n,d)$, composed of all graphs on vertex set $[n] = \{1,\ldots,n\}$ with minimum degree at least dn, where $d > 0$ is a fixed constant. This is the class considered in [3], [4]. A more general class is $\mathcal{G}_a(n,d)$, composed of all graphs with vertex set $[n]$ and *average* degree at least dn, i.e. with at least $\frac{dn^2}{2}$ edges, this is the class we treat here. Obviously, $\mathcal{G}(n,d) \subset \mathcal{G}_a(n,d)$, so all results of this paper are valid for the smaller class of graphs on n vertices with minimum degree at least dn. We wish to stress here that, as hinted before, a small perturbation of the graph $G \in \mathcal{G}_a(n,d)$ amounts – from our point of view – to the addition of few random edges, as deleting few random edges leaves the average degree of G essentially unchanged.

In this paper we consider two properties of the model $G_{H,m}$, for $H \in \mathcal{G}_a(n,d)$. The first property is the appearance of a copy of a fixed graph Γ, discussed in Section 2. We are able to solve this problem more or less completely by establishing the threshold for the appearance of Γ in $G_{H,m}$ as a function of Γ and parameter d. In Section 3 we discuss Ramsey-type properties of $G_{H,m}$ – a potentially rich class of questions for which we are able to get first results. More details, including background, our results and their proofs can be found in the corresponding sections.

We also consider a very similar model of random formulas. In this model, we start with a fixed instance F of a k-SAT formula on n variables. A random formula G is then obtained by adding to F a set R of m randomly chosen k-clauses, the goal is to reach almost surely a non-satisfiable formula. A formal definition of the model together with our result and its proof appear in Section 4. Section 5, the last section of the paper, contains some concluding remarks.

Remark 1.1 All the properties we consider here are monotone. We therefore have an alternative way of adding random edges. We can take each non-edge of H and add it with probability p, creating the random graph $G_{H,p}$ where $p\left(\binom{n}{2} - |E(H)|\right) = (1+o(1))m$. It then follows as in Bollobas [6], Theorem 2.2 that for monotone property $\mathcal{P}$, if $G_{H,p} \in \mathcal{P}$ almost surely then also does $G_{H,m}$. This observation can sometimes simplify our calculations. Similarly, instead of adding a set of m randomly chosen k-clauses to some fixed k-SAT formula F on n variables we can consider the case when each of k-clauses which is not in F is picked randomly and independently with appropriate probability

$p = \Theta(m/n^k)$.

Throughout the paper we will systematically omit floor and ceiling signs for the sake of clarity of presentation. Also we use the notations $a_n = \Theta(b_n)$, $a_n = O(b_n)$ or $a_n = \Omega(b_n)$ for $a_n, b_n > 0$ and $n \to \infty$ if there are absolute constants C_1 and C_2 such that $C_1 b_n < a_n < C_2 b_n$, $a_n < C_2 b_n$ or $a_n > C_1 b_n$ respectively. The notations $a_n = o(b_n)$ means that $a_n/b_n \to 0$ as $n \to \infty$ and if $a_n/b_n \to \infty$, we write $a_n = \omega(b_n)$.

2 Fixed subgraph

In order to formulate our result we need to introduce some notation. Let Γ be a fixed graph and denote by v_Γ the number of vertices of Γ and by e_Γ the number of its edges. Let

$$m(\Gamma) = \max\left\{\frac{e_{\Gamma'}}{v_{\Gamma'}} \mid \Gamma' \subseteq \Gamma, v_{\Gamma'} > 0\right\},$$

and for every positive integer r define

$$m_r(\Gamma) = \min_{V(\Gamma)=\cup_i V_i} \max_i m(\Gamma[V_i]),$$

where the minimum is over all partitions of $V(\Gamma)$ into at most r parts. Note that, by definition, $m_r(\Gamma) < 1/2$ if and only if Γ is r-colorable. For readers convenience we present few examples that illustrate this parameter.

- Let $\Gamma = K_t$ be a complete graph on t vertices, and let $2 \le r \le t$. Then an optimal partition is obtained by partitioning the vertices of Γ into r parts of sizes $\lfloor t/r \rfloor$ and $\lceil t/r \rceil$, yielding:

$$m_r(\Gamma) = \frac{\binom{\lceil t/r \rceil}{2}}{\lceil t/r \rceil} = \frac{\lceil t/r \rceil - 1}{2}.$$

- Let $\Gamma = K_{2,t,t}$ be a complete 3-partite graph with parts of size $2, t, t$, and let $r = 2$. Then we can partition the vertices of Γ into two parts such that each part induces a star of size t. It is easy to see that this partition is optimal, thus $m_2(K_{2,t,t}) = \frac{t}{t+1}$.

- Let $\Gamma = K_{t,t,t,t}$, then one can show that $m_2(K_{t,t,t,t}) = t/2$ by partitioning the vertices into two parts, each spanning a $K_{t,t}$. To show that this is optimal note that in every partition into two parts one part contains at least $2t$ vertices which form a complete multipartite graph Γ' with parts of size at most t. It is easy to verify that in this case $m(\Gamma')$ is smallest for a bipartite graph.

We are now ready to formulate the main result of this section.

Theorem 2.1 *Let $0 < d < 1$ be a fixed constant and let $r \ge 2$ be a unique integer satisfying: $d \in \left(\frac{r-2}{r-1}, \frac{r-1}{r}\right]$. Let Γ be a fixed graph.*

(i) If $H \in \mathcal{G}_a(n,d)$ and $m = \omega\left(n^{2-\frac{1}{m_r(\Gamma)}}\right)$, then almost surely $G_{H,m}$ contains a copy of Γ.
(ii) There exists a graph $H_0 \in \mathcal{G}_a(n,d)$ with $d = (1+o(1))\frac{r-1}{r}$ such that if $m = o\left(n^{2-\frac{1}{m_r(\Gamma)}}\right)$, then a.s. $G_{H_0,m}$ fails to contain a copy of Γ.

Theorem 2.1 generalizes a result of Bohman et al. [4], who proved it for the special case $\Gamma = K_t$. The main probabilistic ingredient of the proof is the following classical result of Bollobás [5].

Theorem 2.2 *If Γ is a fixed graph then*

$$\lim_{n\to\infty} \Pr\left(G_{n,p} \supseteq \Gamma\right) = \begin{cases} 1, & \text{if } p = \omega\left(n^{-1/m(\Gamma)}\right); \\ 0, & \text{if } p = o\left(n^{-1/m(\Gamma)}\right). \end{cases}$$

As is typical with quite a few problems involving dense graphs, the core tool used here is the celebrated Regularity Lemma of Szemerédi [15]. We will formulate and utilize two results, the first being a rather standard application of the Regularity Lemma, while the second usually serving as its frequent companion. Both theorems have been stated and proved, explicitly or implicitly, in quite a few papers applying the Regularity Lemma, including, inter alia, [4]. We thus omit their proof. First, we need a definition of an ϵ-regular pair. For two disjoint vertex sets A and B in a graph G, let $e(A,B)$ denote the number of edges with one vertex in A and the other in B. Also, let $d(A,B) = \frac{e(A,B)}{|A||B|}$ be the *density* of the pair (A,B).

Definition 2.3 *Let $\epsilon > 0$. Given a graph G and two disjoint vertex sets $A \subset V(G)$, $B \subset V(G)$, we say that the pair (A,B) is ϵ-regular if for every $X \subset A$ and $Y \subset B$ satisfying*

$$|X| > \epsilon|A| \qquad \text{and} \qquad |Y| > \epsilon|B|$$

we have

$$|d(X,Y) - d(A,B)| < \epsilon.$$

Theorem 2.4 *Let $r \geq 2$ be an integer and let $d > \frac{r-2}{r-1}$ be a fixed constant. Then there exist real constants ϵ, γ and integer constants K, n_0 such that for all $n \geq n_0$ every graph G on n vertices with average degree at least dn contains r disjoint vertex sets $A_1, \ldots, A_r$ of cardinality $|A_1| = \ldots = |A_r| \geq n/K$ such that for each $1 \leq i \neq j \leq r$, the pair (A_i, A_j) is ϵ-regular of density at least γ.*

Theorem 2.5 *For all real $\epsilon, \gamma > 0$ and integers $r, T > 0$ there exist an integer n_0 and a real δ so that the following holds for all $n \geq n_0$. Let G be an r-partite graph with parts $A_1, \ldots, A_r$ of cardinality n. Assume that for each $1 \leq i \neq j \leq r$ the pair (A_i, A_j) is ϵ-regular of density at least γ. For each $1 \leq i \leq r$ choose a random subset U_i of cardinality T in A_i. Then with probability at least δ the r-partite subgraph of G with parts $U_1, \ldots, U_r$ is complete.*

To rephrase somewhat informally, Theorems 2.4, 2.5 combined assert that every sufficiently large graph G on n vertices with average degree at least dn, $d > \frac{r-2}{r-1}$ contains $\Theta(n^{rT})$ copies of a complete r-partite graph on rT vertices whose parts are all of size T.

Proof of Theorem 2.1. Following Remark 1.1 it is enough to prove our theorem for the model $G_{H,p}$ with $p = (1 - o(1))\frac{m}{\binom{n}{2} - \frac{dn}{2}}$. We first deal with Part (ii).

Let H_0 be a complete r-partite graph on n vertices with nearly equal parts $A_1, \ldots, A_r$. It is easy to verify that $|E(H_0)| = (1 + o(1))\frac{r-1}{r}\frac{n^2}{2}$, implying that $H_0 \in \mathcal{G}_a(n, d)$ with $d = (1 + o(1))\frac{r-1}{r}$. Let $\Gamma_1, \Gamma_2, \ldots$ be a family of subgraphs of Γ such that $m(\Gamma_i) \geq m_r(\Gamma)$ for all i. Clearly, this family has constant size. Also note that for all $1 \leq j \leq r$, the edges inside A_j form a copy of the random graph $G_{n/r,p}$ with $p = o\left(n^{-\frac{1}{m_r(\Gamma)}}\right)$ and these copies are independent. Therefore, by Theorem 2.2, almost surely none of the parts A_j contains any of the finitely many subgraphs Γ_i of Γ with $m(\Gamma_i) \geq m_r(\Gamma)$. Since by the definition of $m_r(\Gamma)$, for any partition of the vertex set of Γ into r parts one part should induce such a graph Γ_i, we conclude that a.s. there is no copy of Γ in $G_{H,p}$.

The proof of Part (i) of Theorem 2.1 is somewhat more involved and relies on Theorems 2.2, 2.4, 2.5. Fix a partition of the vertex set of graph Γ into r parts $W_1, \ldots, W_r$ so that $m(\Gamma[W_j]) \leq m_r(\Gamma)$. Denote $\Gamma_j = \Gamma[W_j]$, $j = 1, \ldots, r$. Let $T = \max_{1 \leq j \leq r} |W_i|$.

Let now $H \in \mathcal{G}_a(n, d)$. Applying Theorem 2.4 to H, we get r disjoint vertex sets $A_1, \ldots, A_r \subset V(H)$ of linear size such that all pairs (A_i, A_j), $1 \leq i \neq j \leq r$, are ϵ-regular of positive density $\gamma > 0$, for some ϵ, γ depending only on d.

Recall that the set R of random edges added to H contains every non-edge of H independently with probability $\omega\left(n^{-\frac{1}{m_r(\Gamma)}}\right)$. Thus, for an appropriately chosen function $\phi = \phi(n) \to \infty$, we can represent the set R as a union of ϕ independent sets of random edges R_i, where every non-edge of H is in R_i randomly and independently with probability which is still $\omega\left(n^{-\frac{1}{m_r(\Gamma)}}\right)$. Consider what happens when we add set R_i.

By definition, for all i and j, the edges of R_i inside A_j form a copy of the random graph $G_{\Theta(n),p}$ with $p = \omega\left(n^{-\frac{1}{m_r(\Gamma)}}\right)$ and all these copies are independent. Thus, by Theorem 2.2, almost surely the random set R_i puts a copy of Γ_j inside A_j, for each $j = 1, \ldots, r$. Assume this is the case indeed, and let $U_j, j = 1, \ldots, r$, be the vertex set of such a copy. Obviously, the sets U_j are mutually independent and thus can be considered as random sets of size $|U_j|$ inside A_j. Recall that $|U_j| \leq T$ for all $j = 1, \ldots, r$. Therefore, by Theorem 2.5, with probability at least δ, the r-partite graph $H[U_1 \cup \ldots \cup U_r]$ is complete, in which case the graph $H \cup R_i$ spans a copy of Γ on $U_1 \cup \ldots \cup U_r$. Thus with probability at least $(1 - o(1))\delta > \delta/2$ there is a copy of Γ in $H \cup R_i$. Observe that $\delta > 0$ is a constant depending only on d and Γ. As the random sets R_i are independent, it follows that the probability that $H \cup R$ does not have a copy of Γ is at most $\left(1 - \frac{\delta}{2}\right)^{\phi} = o(1)$. $\square$

3 Ramsey-type properties

We start by recalling a commonly used notation in Ramsey theory. Let G, G_1 and G_2 be three graphs. We write $G \to (G_1, G_2)$ if every Red-Blue coloring of the edges of G contains either a Red copy of G_1 or a Blue copy of G_2.

The basic question studied in Ramsey theory is, given two fixed graphs G_1 and G_2, to decide when graph G is "rich enough" for $G \to (G_1, G_2)$. Here richness can be interpreted as the size of

G or the edge density, i.e., the ratio of edges to vertices. When studying random graphs a natural question is, given fixed graphs G_1 and G_2, to determine a threshold function for the edge probability p such that $G_{n,p} \to (G_1, G_2)$ almost surely. This question and its variants were studied extensively, see, e.g., [13] and its references.

In this section we discuss Ramsey-type properties in the model $G_{H,m}$, with $H \in \mathcal{G}_a(n,d)$, for fixed $0 < d < 1$. Our goal is to determine the threshold function for the number of random edges m which one needs to add to H to guarantee that a.s. $G_{H,m} \to (G_1, G_2)$. Here we are able to solve the first interesting case when $G_1 = K_3$ and $G_2 = K_t, t \geq 3$.

Theorem 3.1 *(i) For every $t \geq 3$ there exists a graph H on n vertices and $n^2/4$ edges such that adding $m = o(n^{2-2/(t-1)})$ random edges to H produces a graph G that almost surely has a Red-Blue edge coloring with no Red copy of K_3 and no Blue copy of K_t.*
(ii) For $0 < d < 1$, let H be a graph on n vertices with average degree at least dn. Then adding $\omega(n^{2-2/(t-1)})$ random edges to H produces almost surely a graph G such that $G \to (K_3, K_t)$.

Note that the case $t = 3$ in the second part of the above statement follows easily from Theorem 2.1 of the preceding section. Indeed, the well known fact that $K_6 \to (K_3, K_3)$ suggests that it suffices to show that for $m = \omega(n)$, $G_{H,m}$ almost surely contains a copy of K_6. The latter follows from Part (i) of Theorem 2.1 if we substitute there $\Gamma = K_6$, $0 < d < 1$ and $r = 2$. For such choices of parameters $m_r(\Gamma) = m_2(K_6) = 1$ and the assertion of the theorem guarantees the almost sure existence of K_6. The following definition plays an important role in our proof.

Definition 3.2 Let $G = (V, E)$ be a graph and let U be a subset of V. Denote by

$$N^*(U) = \{v \in V \mid (v,u) \in E, \text{ for every } u \in U\}$$

the *common neighborhood* of U in G. A subset $V_0 \subset V$ of vertices is called *typical* if its common neighborhood has linear size, i.e., $|N^*(V_0)| \geq cn$, for some $c > 0$.

Remark 3.3 Although we are studying properties of $G_{H,m}$, for H which has average degree at least dn (for a fixed constant $0 < d < 1$), throughout the rest of this section we will assume for convenience that we are in fact dealing with a graph whose minimum degree is at least $dn/2$. Since we will only be using estimates on the number of edges in linear-sized (in n) subsets of vertices this results in no loss of generality – indeed, every graph H on n vertices with average degree dn contains a subgraph H' on $n' \geq dn/2$ vertices with minimum degree at least $dn/2$.

We need the following lemma.

Lemma 3.4 *Let $0 < d < 1$ be a fixed constant and let $H \in \mathcal{G}(n, d/2)$. Let $t \geq 4$ and let R be the set of m random edges with $m = \omega(n^{2-2/(t-1)})$. Finally let $G = G_{H,m}$. Then as $n \to \infty$ we have*
(i) $$\Pr[G \text{ contains a typical copy of } K_t] \to 1.$$
(ii) Let $0 < \alpha < 1$ be a fixed constant. Then

$$\Pr[\text{ Every set } U \text{ of size } |U| \geq \alpha n \text{ contains a typical copy of } K_{t-1} \text{ in } G] \to 1.$$

Proof. We prove Part (ii) first and indicate the changes necessary for the proof of the simpler statement in Part (i). For $v \in V(H)$, let $d(v)$ denote the degree of v in H. For a subset $U \subseteq V(H)$ we denote by $d(v,U)$ the number of neighbors of v in U. First observe that every collection of a linear number of vertices of H (and hence of $G_{H,m}$) contains lots of subsets of size $(t-1)$ (and also t), which are typical. Indeed, let $U \subset V(H)$ be of size $\geq \alpha n$ (for $0 < \alpha < 1$). Since the minimum degree of H is at least $dn/2$ we have that

$$\sum_{v \in V(H)} d(v,U) = \sum_{u \in U} d(u) \geq \frac{dn}{2}|U| \geq \frac{\alpha d}{2} n^2.$$

Therefore

$$\begin{aligned}\sum_{\substack{V_0 \subset U \\ |V_0| = t-1}} |N^*(V_0)| &= \sum_{v \in V(H)} \binom{d(v,U)}{t-1} \geq n \binom{\sum_{v \in V(H)} \frac{d(v,U)}{|n|}}{t-1} \\ &= n\binom{\alpha dn/2}{t-1} \geq \left(\frac{\alpha d}{2(t-1)}\right)^{t-1} n^t = \Theta(n^t).\end{aligned}$$

As every $|N^*(V_0)| \leq n$, we conclude that there is a constant $c > 0$ such that $|N^*(V_0)| \geq cn$ for at least $\Theta(n^{t-1})$ subsets $V_0 \subset U$ of size $t-1$.

The rest of the proof is straightforward using the so-called Janson Inequality (see [10] or [2]). Following Remark 1.1 we can assume that we add edges not in H randomly with probability $p = \omega\big(n^{-2/(t-1)}\big)$. Due to monotonicity we may also assume that $p(n)$ does not exceed $n^{-2/(t-1)}$ by much, say, $p(n) \leq n^{-2/(t-1)} \log n$.

Let μ_U denote the expected number of typical copies of K_{t-1} in the (random) graph on U, and let Δ_U denote the correlation term:

$$\Delta_U = \sum_{\substack{T,T' \subset U \\ |T|=|T'|=t-1 \\ |T \cap T'| \geq 2}} \Pr[T \text{ and } T' \text{ each induce a typical } K_{t-1}].$$

It is easy to check that

$$\mu_U = \Theta\left(n^{t-1} p^{\binom{t-1}{2}}\right) = \omega(n),$$

and that

$$\Delta_U = O\left(n^{2t-4} p^{(t-1)(t-2)-1}\right) = O\left(n^{2/(t-1)} (\log n)^{(t-1)(t-2)-1}\right) = o(\mu_U).$$

Using the Janson inequality it now follows that

$$\Pr[\text{ number of typical copies of } K_{t-1} \text{ in } U = 0] \leq e^{-\mu_U + \frac{\Delta_U}{2}} = e^{-\omega(n)}.$$

Therefore, with probability $1 - 2^n e^{-\omega(n)} = 1 - o(1)$ every set U of size at least αn contains a typical copy of K_{t-1}.

The proof of part (i) is obtained along the same lines. Observe that the expected number μ of typical copies of K_t in G is $\Theta\big(n^t p^{\binom{t}{2}}\big) = \omega(1)$, and the correlation term is

$$\Delta = \sum_{\substack{T,T' \subset V \\ |T|=|T'|=t \\ |T \cap T'| \geq 2}} \Pr[T \text{ and } T' \text{ each induce a } K_t \text{ in } G] = O\left(n^{2t-2} p^{2\binom{t}{2}-1}\right) = o(1).$$

Once again the assertion follows using the Janson inequality. $\square$

Since t is fixed, we may apply Lemma 3.4 greedily, and obtain that in fact, almost surely every set of a linear number of vertices contains a *linear-sized* family of vertex disjoint typical copies of K_{t-1}. Thus the following corollary is immediate, with $G = G_{H,m}$ as in Lemma 3.4.

Corollary 3.5 *Let $0 < \alpha < 1$ be a fixed constant, let $U \subset V(G)$ with $|U| \geq \alpha n$. There exists a constant $0 < \beta = \beta(\alpha, d, t) < 1$ such that when $n \to \infty$,*

$$\Pr[\textit{Every set } U \textit{ of size } |U| \geq \alpha n \textit{ contains a family of } \beta n \textit{ vertex disjoint typical copies of } K_{t-1}] \to 1.$$

We also make use of the following two theorems which establish Ramsey-type properties of random graphs. The first result that we use was obtained by Rödl and Ruciński [12].

Theorem 3.6 *For every $a > 0$, there exist C and N_0 such that for $n > N_0$ and $p > Cn^{-1/2}$, we have*

$$\Pr[G_{n,p} \to (K_3, K_3)] > 1 - \exp(-an^{3/2}).$$

Before stating the second theorem, we need a couple of definitions from [11]. Let C_3 denote a cycle of length 3. For a graph $H = (V(H), E(H))$, with at least three vertices, the *maximum 2-density* $d_2(H)$ of H is defined to be

$$d_2(H) = \max\left\{\frac{|E(J)| - 1}{|V(J)| - 2} : J \subseteq H \text{ with } |V(J)| \geq 3\right\}.$$

Finally, a graph H is called 2-balanced if the maximum above is attained by H. The following theorem was proved by Kohayakawa and Kreuter [11].

Theorem 3.7 *Let H be a 2-balanced graph with $d_2(H) > d_2(C_3) = 2$. Then for every constant $a > 0$, there exist constants B and N_0 such that for $n > N_0$ and $p > Bn^{-\beta}$ with*

$$\beta = \beta(C_3, H) = \frac{|V(H)| - 2 + 1/d_2(C_3)}{|E(H)|},$$

we have

$$\Pr[G_{n,p} \to (C_3, H)] > 1 - \exp(-an^{1+1/2}).$$

We note that the result in [12] is considerably more general in that it is true for $r \geq 2$ colors, and for a random subgraph G_p of a reasonably regular graph G (for precise statements, see Theorems 1 and 2 in [12]). Similarly, the results in [11] are also more general, and in particular deal with $r \geq 2$ colors and with r monochromatic cycles C_l of varying lengths $l \geq 3$. Note that the results in [11] are only stated with probabilities $1 - o(1)$ without explicit calculation of the exponentially small error terms in $o(1)$. However, one can check that the above bound we report (for the special case of C_3) is in fact valid. For our purposes, the following immediate corollary of the above theorems suffices.

Corollary 3.8 *Let $t \geq 3$ be a fixed integer. For every $a > 0$, there exist C and N_0 such that for $n > N_0$ and $p > Cn^{-(2t-3)/t(t-1)}$, we have*

$$\Pr[G_{n,p} \to (K_3, K_t)] > 1 - \exp(-an^{3/2}).$$

Proof. The case of $t = 3$ is precisely Theorem 3.6. For $t \geq 4$, observe that $d_2(K_t) = \frac{\binom{t}{2}-1}{t-2} > 2 = d_2(C_3)$, and thus Theorem 3.7 is applicable with $\beta(C_3, K_t) = (2t-3)/(t(t-1))$. □

Having finished all the necessary preparations we are now ready to prove the main result of this section.

Proof of Theorem 3.1. To prove (i), take $H = K_{\frac{n}{2},\frac{n}{2}}$ and color the edges of H by Red, and the random edges by Blue. As $G(n,m)$, with $m = o(n^{2-2/(t-1)})$, a.s. does not contain a copy of K_t by Theorem 2.2, the claim follows.

We now proceed with the proof of Part (ii) of the theorem for $t \geq 4$. Based on Lemma 3.4, Corollary 3.5 and a trivial combination of the union bound with Corollary 3.8 we may assume that G has the following properties:

1. G has a typical K_t;

2. Every set U of $|U| = \Theta(n)$ vertices of G contains a typical K_{t-1} in G;

3. For $t \geq 5$, every set U of $|U| = \Theta(n)$ vertices of G satisfies: $G[U] \to (K_3, K_{t-2})$.

Observe that the last statement follows as $n^{-\frac{2}{t-1}} \geq n^{-\frac{2(t-2)-3}{(t-2)(t-3)}}$ for $t \geq 5$. We will be able to take the hidden constant in 2)–3) above as small as will be required in the proof. So assume that indeed G has Properties 1)–3) above, and consider an arbitrary Red-Blue coloring of the edges of G. Assume by the way of contradiction that this coloring contains neither a Red K_3 nor a Blue K_t (fixed $t \geq 4$).

Claim. There is a vertex whose Blue degree is cn, for some constant $0 < c < 1$.

Proof. By the first property, G contains a typical copy of K_t. Denoting the set of the vertices of the typical copy of K_t by T, we have $|N^*(T)| = c_1 n$, for some constant $0 < c_1 < 1$. Since there is no Blue K_t, at least one of the edges inside T must be colored Red, and let $(u,v) \in E(H)$ be such an edge. For every $w \in N^*(u,v)$, at least one of (u,w) or (v,w) is Blue - otherwise we get a Red K_3. Therefore the Blue degree of u or v is at least $c_1 n/2$.

Let the vertex with the linear-sized Blue degree be v, and let $N_B(v)$ be the Blue neighborhood of v. By Property 2 of G, $N_B(v)$ contains a family of βn pairwise disjoint, typical copies of K_3, for some constant $0 < \beta < 1$. Each of them must have a Blue edge – otherwise we get a Red K_3 and are done. Let those Blue edges be $(x_1,y_1),\ldots,(x_s,y_s)$, where $s = \beta n$. Denote by $N_i = N^*(x_i,y_i)$, the common neighborhood of x_i, y_i, for $i = 1,\ldots,s$. Note that each N_i is linear in size, since $\{x_i,y_i\}$ is a pair from a typical K_3. Let $\min_i |N_i| = c_2 n$ for some constant $0 < c_2 < 1$. Now we consider two cases.

Case (a). Suppose that for some $i \in \{1,\ldots,s\}$, and some constant $c > 0$,

$$\left|\Big\{w \in N_i : (x_i,w),(y_i,w) \text{ are both Blue}\Big\}\right| = cn. \tag{1}$$

Then we get for that choice of i, a Blue edge (x_i, y_i) with common Blue neighborhood $N_{\text{Blue}} \subseteq N_i$ of size cn. The case of $t = 4$ can easily be dispensed with at this point: Property 2 guarantees that N_{Blue} contains a copy of $K_{t-1} = K_3$. If at least one of the edges of this K_3 is Blue, then together with the pair (x_i, y_i) we obtain a Blue K_4. If not, then we obtain a Red K_3.

For $t \geq 5$, we invoke Property 3) of G. Since $|N_{Blue}| = cn$, we have $G[N_{Blue}] \to (K_3, K_{t-2})$. Thus together with the Blue edge (x_i, y_i) and the fact that N_{Blue} is the Blue neighborhood of (x_i, y_i), we have the Ramsey property $G \to (K_3, K_t)$. Thus, in both cases we have arrived at a contradiction with our assumption on the coloring of G.

Case (b). Suppose that for none of the $i \in \{1, \ldots, s\}$ Equation (1) holds. Then for each i, the edge (x_i, y_i) is incident to $(1-o(1))|N_i|$ Red edges. Moreover, since each Red edge is counted at most twice as we vary i, we infer that $N_B(v)$ is incident to at least $(1/2)(1-o(1)) \sum_{i=1}^{s} |N_i| \geq (1/2 - o(1))\beta c_2 n^2$ Red edges. Hence there exists a vertex $w \in V$ whose Red degree to $N_B(v)$ is at least $(1/2 - o(1))\beta c_2 n$. Let $\hat{N}$ be the intersection of the Blue neighborhood of v and the Red neighborhood of w, so that we have $|\hat{N}| = (1/2 - o(1))\beta c_2 n = \Theta(n)$. Once more according to Property 2, $\hat{N}$ contains a copy of K_{t-1}. If this copy has a Red edge, then this Red edge forms a Red triangle with w. Otherwise, all of the edges of the above K_{t-1} are Blue, forming a Blue K_t together with v, yielding a contradiction and thus concluding the proof of the theorem. □

4 Getting (un)satisfied with a little use of randomness

Given n boolean variables $x_1, \ldots, x_n$, a *literal* is a variable x_i or its negation $\overline{x_i}$ for all $1 \leq i \leq n$. For a set of literals Y denote by $\overline{Y}$ the set of literals $\{\overline{y} \mid y \in Y\}$. A disjunction of k literals is called a k-*clause*. Let $\mathcal{C}_k$ be the set of all k-*clauses* over $x_1, \ldots, x_n$, i.e., the set of $2^k \binom{n}{k}$ possible disjunctions of k distinct, non-complementary literals. A k-SAT formula F is a conjunction of clauses from $\mathcal{C}_k$.

Given a k-SAT formula F, the Satisfiability problem asks whether there exists a truth assignment to the variables of F under which it evaluates to true. Satisfiability has been a central problem of computational complexity since 1971 when Cook proved that it is complete for the class NP, i.e., it is as hard as any other decision problem for which positive answer can be verified in polynomial time.

A random k-SAT formula $F_k(n, m)$ is formed by selecting uniformly at random m clauses from $\mathcal{C}_k$ and taking their conjunction. The study of properties of random k-SAT formulas has become an extremely active area of research in the last fifteen years. Here the main goal is to obtain precise estimate of the threshold value of m which makes the k-SAT formula almost surely unsatisfiable (see, e.g., [1] and its references).

In this section we consider another model of random formulas. In this model, we start with a fixed (satisfiable) instance F of a k-SAT formula on n variables. A random formula G is then obtained by adding to F a set R of m clauses that are chosen uniformly at random from the set of all k-clauses which are not in F. The goal is to reach almost surely a non-satisfiable formula. If F is the empty formula, i.e., contains no clauses, we obtain the original random k-SAT model $F_k(n, m)$. Here we are interested in another regime when the formula F is quite "dense". In this case we are able to obtain an asymptotic order of magnitude of m that guarantees almost sure non-satisfiability. Specifically, we prove the following result.

Theorem 4.1 *(i) Let $c > 0$, $0 \leq \epsilon < 1/k$, and let F be a satisfiable k-SAT formula on n variables with at least $cn^{k-\epsilon}$ clauses. Then almost surely a conjunction of F with $m = \omega(n^{k\epsilon})$ random clauses from $\mathcal{C}_k$ is not satisfiable.*
(ii) For every $0 \leq \epsilon \leq 1/k$, there exists a k-SAT formula on n variables with $\Omega(n^{k-\epsilon})$ clauses such that a conjunction of F with $m = o(n^{k\epsilon})$ random clauses from $\mathcal{C}_k$ is almost surely satisfiable.

Proof. Following Remark 1.1 we can assume that instead of adding m random clauses we will add every clause in $\mathcal{C}_k$ with probability $p = \Theta(m/n^k) = \omega(n^{k\epsilon - k})$.

To show Part (i) of the theorem we will prove the following more general statement. Let $1 \leq s \leq k$ and let F be an s-SAT formula on n variables with at least $\Omega(n^{s-\epsilon})$ clauses. For every clause in $\mathcal{C}_k$ choose it randomly and independently with probability $p = \omega\left(n^{k\epsilon - k}\right)$ and denote this set of random clauses by R. Then almost surely a conjunction of F with the clauses in R is not satisfiable. Note that the case $s = k$ of this statement gives the assertion of Part (i). We prove our claim by induction on s. If $s = 1$, then F is simply a conjunction of $\Omega(n^{1-\epsilon})$ literals which all should get true value to satisfy F. Denote the set of these literals by Y and pick every clause in $\mathcal{C}_k$ with probability $p = \omega\left(n^{k\epsilon - k}\right)$. Since $|\overline{Y}| = |Y| = \Omega(n^{1-\epsilon})$ we have at least $\binom{|\overline{Y}|}{k} = \Omega(n^{k-k\epsilon})$ k-clauses whose all variables are in $\overline{Y}$. So the probability that we choose one of them is at least

$$1 - (1-p)^{\binom{|\overline{Y}|}{k}} = 1 - (1-p)^{\Omega(n^{k-k\epsilon})} = 1 - e^{-\Omega(p\, n^{k-k\epsilon})} = 1 - e^{-\omega(1)} = 1 - o(1).$$

Thus a.s. at least one clause in R has all its literals in $\overline{Y}$ and therefore is not satisfied by any assignment that satisfies F.

Now suppose that F is an $(s+1)$-SAT formula on n variables with at least $\Omega(n^{s+1-\epsilon})$ clauses. Denote by Y the set of all literals contained in at least $\Omega(n^{s-\epsilon})$ clauses of F. Since the total number of clauses in F is $\Omega(n^{s+1-\epsilon})$ and every literal is contained in at most $O(n^s)$ such clauses one can easily see that Y has size $\Omega(n^{1-\epsilon})$.

Recall that the set R of random k-clauses added to F contains every clause in $\mathcal{C}_k$ independently with probability $\omega\left(n^{k\epsilon - k}\right)$. Thus we can represent the set R as a union of $k+1$ independent sets of random clauses R_i, where R_i is a set of k-clauses that are chosen from $\mathcal{C}_k$ randomly and independently with probability which is still $\omega\left(n^{k\epsilon - k}\right)$.

Consider what happens when we first add the set R_{k+1}. As we already explained above, since $|\overline{Y}| = |Y| = \Omega(n^{1-\epsilon})$, we obtain that almost surely there is a clause $\overline{y_1} \vee \overline{y_2} \vee \cdots \vee \overline{y_k}$ in R_{k+1} such that that $y_i \in Y$ for all i. Next, for every $1 \leq i \leq k$ consider all the clauses of F which contain y_i, and delete y_i from them. This gives a set of s-clauses whose conjunction will be denoted by F_i. Now add k independent sets of random k-clauses $R_1, \ldots, R_k$. By definition each F_i has at least $\Omega(n^{s-\epsilon})$ s-clauses so we can use induction. By induction hypothesis, almost surely there is no satisfying assignment for conjunction of F_i and R_i for every $1 \leq i \leq k$. Therefore a.s. every assignment that satisfies the conjunction of F with $R_1, \ldots, R_k$ should set the value of all literals $y_i, 1 \leq i \leq k$, to be true. Then the clause $\overline{y_1} \vee \overline{y_2} \vee \cdots \vee \overline{y_k}$ is not satisfied. This implies that a.s. there is no assignment that satisfies the conjunction of F with all clauses in $R_1 \cup \cdots \cup R_k \cup R_{k+1}$. This completes the proof of the induction step and the proof of Part (i).

The following example proves Part (ii) of the theorem. Partition the set of variables $\{x_1, \ldots, x_n\}$ as $A \cup B$ where $|A| = n^{1-\epsilon}$. Let F be a conjunction of all the clauses from $\mathcal{C}_k$ which have exactly one literal in A and $k-1$ literals in B. By definition the number of clauses in F is $\Theta(n^{k-\epsilon})$. Set all variables of A to be true (this will satisfy F) and choose every clause in $\mathcal{C}_k$ randomly and independently with probability $p = o\left(n^{k\epsilon - k}\right)$. Denote this set of random k-clauses by R. We claim that almost surely there is an assignment for remaining variables that satisfies R and therefore there is a satisfying assignment for the conjunction of F and R.

To prove this claim we transform R into another random formula R' that contains only literals and 2-clauses. First note that almost surely there is no clause in R whose all literals are in $\overline{A}$. Indeed, since $|\overline{A}| = |A| = n^{1-\epsilon}$, the probability that there is such a clause equals $\binom{|\overline{A}|}{k} p \le n^{k(1-\epsilon)} p = o(1)$. Also every clause in R which has a literal in A is satisfied so we discard it. Fix an ordering $x_1 < \overline{x_1} < x_2 < \overline{x_2} < \ldots$ of our literals. For every literal $z \in B \cup \overline{B}$ we add it to R' if there are $k-1$ literals $z_1, \ldots, z_{k-1} \in \overline{A}$ such that $z \vee z_1 \vee \cdots \vee z_{k-1} \in R$. Since $\epsilon \le 1/k \le 1/2$, the probability of this event is

$$p_1 = \binom{|\overline{A}|}{k-1} p = \binom{n^{1-\epsilon}}{k-1} p \le n^{(k-1)(1-\epsilon)} p = o\left(n^{-1+\epsilon}\right) = o\left(n^{-1/2}\right).$$

For every pair of literals $z, z' \in B \cup \overline{B}$ we add a 2-clause $z \vee z'$ to R' if there is a clause in R which contains both z and z' and in which these two literals are the smallest among the all its literals from $B \cup \overline{B}$. Since $\epsilon \le 1/k$, the probability of this event is

$$p_2 \le \sum_{i=0}^{k-2} \binom{|A|}{i} \binom{|B|}{k-i} 2^{k-2} p \le O\left(n^{k-2} p\right) = o\left(n^{k\epsilon - 2}\right) = o(n^{-1}).$$

Also note that by definition all these events are independent since they depend on the disjoint sets of clauses from $\mathcal{C}_k$. By our construction, any satisfying assignment for R' is also a satisfying assignment for R. Therefore, to complete the proof of Part (ii) of the theorem, it is enough to prove the following simple statement.

Proposition 4.2 *Let R' be a random formula of $1-$ and 2-clauses on n variables such that every 1-clause is present in R' independently with probability $p_1 = o\left(n^{-1/2}\right)$ and every 2-clause is present in R' independently with probability $p_2 = o\left(n^{-1}\right)$. Then almost surely R' is satisfiable.*

Proof. To prove this proposition we apply the standard method of search for a satisfying assignment of boolean formulas called the *pure literal rule*. Our proof is an adaptation of the arguments used in [7] and [9] to establish a threshold for satisfiability of the random 2-SAT.

A literal is called *pure* if its negation does not appear in the formula. Consider the procedure where we repeatedly set a pure literal in R' to be true and delete all clauses from R' that contain this literal. We continue until there are no more pure literals. We claim that a.s. we are left with the empty formula.

In order to prove this claim, associate a (random) multigraph $G = G(R')$ with the random formula R', where G may contain multiple loops and multiple edges. G is defined as follows: the vertices of G are variables $z_1, \dots, z_n$; a 1-clause z_i or $\overline{z_i}$ corresponds to a loop at z_i; a 2-clause $y_i \vee y_j$ with $y_i \in \{z_i, \overline{z_i}\}$, $y_j \in \{z_j, \overline{z_j}\}$ translates to an edge (z_i, z_j). It is immediate that a loop appears in G with probability at most $2p_1$, while an edge appears in G with probability at most $4p_2$.

Suppose we are left with the set of clauses in which every literal appears together with its negation. Then G contains a subgraph with all degrees at least 2 (where a loop contributes 1 to the degree of its vertex). It is easy to check that then there is a sequence of s distinct variables $z_1, \dots, z_s$ such that one of the following four cases holds:

1. G contains loops at z_1, z_s and edges $(z_1, z_2), \dots, (z_{s-1}, z_s)$. This happens with probability at most
$$\sum_{s=2}^{n} n^s (2p_1)^2 (4p_2)^{s-1} = o(np_1^2) = o(1) \ ;$$
2. $s \geq 3$ and G contains edges $(z_1, z_2), \dots, (z_{s-1}, z_s), (z_s, z_1)$. This happens with probability at most
$$\sum_{s=3}^{n} n^s (4p_2)^s = o(1) \ ;$$
3. $s = 1$ and G contains two loops at z_1. This happens with probability at most $n(2p_1)^2 = o(1)$;
4. $s = 2$ and G contains two edges between z_1 and z_2. This happens with probability at most $n^2(4p_2)^2 = o(1)$.

Therefore almost surely none of the four events happens and hence R' is satisfiable. □

5 Concluding remarks

We strongly feel that the results of this paper and its predecessors [3], [4] are just a beginning of a prolific line of research on this very interesting model. Undoubtedly many more interesting questions in this setting can be posed.

Going back to our paper, it would be nice to be able to prove (in a more elementary and self-contained fashion) the Ramsey-type result from Section 3 without appealing to the powerful results of [12] and [11]. Another related Ramsey-type question is to establish a threshold for Ramsey property $G \to (G_1, G_2)$ for pairs (G_1, G_2) other than $G_1 = K_3$, $G_2 = K_t$, treated here.

A related problem to that considered in Section 4 is to find a threshold for non-2-colorability of an r-uniform hypergraph G, obtained by adding random edges to a hypergraph H on n vertices with a given number of edges. Our preliminary estimates show that, despite apparent similarity of this problem to that of Section 4, they are in fact very different, and the value of the threshold function is different as well.

Acknowledgment. We would like to thank Yoshiharu Kohayakawa for helpful tips on wading through [11].

References

[1] D. Achlioptas and Y. Peres, The Threshold for Random k-SAT is $2^k \log 2 - O(k)$, *Proceedings of the 35th Annual ACM Symposium on Theory of Computing* (2003), 223–231.

[2] N. Alon and J. Spencer, **The Probabilistic Method**, 2nd ed., Wiley, New York, 2000.

[3] T. Bohman, A. Frieze and R. Martin, How many random edges make a dense graph Hamiltonian?, *Random Structures and Algorithms* **22** (2003), 33–42.

[4] T. Bohman, A. Frieze, M. Krivelevich and R. Martin, Adding random edges to dense graphs, *Random Structures and Algorithms* **24** (2004), 105–117.

[5] B. Bollobás, Random Graphs, In *Combinatorics*, Proceedings, Swansea 1981, London Math. Soc. Lecture Note Ser. 52, Cambridge Univ. Press, Cambridge (1981) 80-102.

[6] B. Bollobás, **Random graphs**, 2^{nd} ed. Cambridge Studies in Advanced Mathematics, 73. Cambridge University Press, Cambridge, 2001.

[7] V. Chvátal and B. Reed, Mick gets some (the odds are on his side), *Proceedings of the 33rd Annual IEEE Symposium on Foundations of Computer Science* (1992), 620–627.

[8] P. Erdős and A. Rényi, On the evolution of random graphs, *Publ. Math. Inst. Hungar. Acad. Sci.* **5** (1960) 17-61.

[9] A. Goerdt, A threshold for unsatisfiability, *Proceedings of the 17th International Symposium of Mathematical Foundations of Computer Science*, Springer LNCS 629 (1992), 264–275.

[10] S. Janson, T. Łuczak, and A. Ruciński, **Random Graphs**, Wiley, New York, 2000.

[11] Y. Kohayakawa and B. Kreuter, Threshold functions for asymmetric Ramsey properties involving cycles, *Random Structures and Algorithms* **11** (1997), 245–276.

[12] V. Rödl and A. Ruciński, Random graphs with monochromatic triangles in every edge coloring, *Random Structures and Algorithms* **5** (1994), 253–270.

[13] V. Rödl and A. Ruciński, Threshold functions for Ramsey properties, *J. Amer. Math. Soc.* **8** (1995), 917–942.

[14] D. A. Spielman and S.-H. Teng, Why the simplex algorithm usually takes polynomial time, *Proceedings of the 33rd Annual ACM Symposium on Theory of Computing* (2001), 296–305.

[15] E. Szemerédi, Regular partitions of graphs, *Colloques Internationaux C.N.R.S.* N° *260 – Problèmes Combinatoires et Théorie des Graphes, Orsay* (1976), 399-401.