KIM'S LOWER BOUND FOR $R(3,k)$

JULIAN SAHASRABUDHE

Abstract. Course notes for the PCMI summer program on extremal and probabilistic combinatorics. This course focuses almost exclusively on the mathematics that surrounds the numbers $R(3,k)$.

1. Introduction

In the next two lectures I will hope to give a reasonably complete proof of the breakthrough result of Kim, which shows that the bound of Shearer/ Ajtai-Komlós Szemerédi is sharp up to constant factors.

Theorem 1.1. *We have*

$$R(3,k) = \Theta\left(\frac{k^2}{\log k}\right).$$

The way that we are going to prove this is use a variant of the so called triangle-free process.

The triangle-free process is a way of building a random triangle-free graph. We start with the empty graph G_0 on $[n]$ vertices and then, add edges one at a time as follows. Given G_i, we let

$$O_i = \big\{ e \in K_n : G_i + e \not\supset K_3 \big\}.$$

Then we define $G_{i+1} = G_i + e_{i+1}$ where $e_{i+1} \sim O_i$ is chosen uniformly at random among all pairs in O_i. We will actually introduce a modified version of this process, that we will work with and which turns out to be much easier to prove things about.

But, for the moment, we will let us just work at a very high level to understand how this process should behave and in particular why the terminating graph G_τ satisfies

$$\alpha(G_\tau) = O(\sqrt{n \log n}).$$

1.1. **Heuristic for the terminating density of the triangle-free process.** In the discussion that follows, we make one large (and non-rigorous) assumption: that the triangle-free process, at density p, "behaves as $G(n,p)$, apart from the fact that it has no triangles". Let G_p be the triangle-free process when it has density p and let $\mathbb{P}_p$ denote the corresponding probability measure. We first ask: how large could we reasonably expect p to be? For this, let us say a pair $e \in K_n \setminus G_p$ is[1] *open* if we can add it to G_p without forming a triangle. Now

[1]Here and throughout the paper we identify a graph with its edge set. So a graph is a set of unordered pairs.

note that if

$$\binom{n}{2}\mathbb{P}\big(e \text{ is open}\big) = \mathbb{E}_p\left|\big\{e \in K_n : e \text{ is open}\big\}\right| \ll p\binom{n}{2}, \tag{1}$$

then the number of edges that one could possibly add after this point (since we can only add open edges) is dwarfed by the number of edges that have already been added, and therefore is negligible. Thus to find out where the process stops we want to solve

$$\mathbb{P}(e \text{ is open}) \approx p$$

To calculate the probability a pair $e = xy$ is open (and thus the expectation in (1)), we note that if e is open then for all $v \in [n] \setminus \{x, y\}$, either xv or yv is not an edge of G_p. Using this and our hypothesis that G_p behaves like $G(n,p)$, we have

$$\mathbb{P}_p\big(xy \text{ is open}\big) \approx \prod_v \mathbb{P}_p\big(\{xv, yv\} \not\subset G_p\big) = \big(1-p^2\big)^{n-2} = e^{-(1+o(1))p^2 n}. \tag{2}$$

Therefore, noting that

$$e^{-(1+o(1))p^2 n} = p \qquad \Longrightarrow \qquad p = (1+o(1))\sqrt{\frac{\log n}{2n}},$$

Which gives a natural guess for the terminating density of the triangle-free process. In fact this is correct. Note that this is a $\sqrt{\log n}$ factor *denser* than the graph that gave our earlier lower bound for $R(3,k)$ - which is good news.

1.2. **Heuristic for the independence number.** Now that we have the maximum density in hand, we can easily get a heurstic for the size of the largest independent set - again by using our maximum "G_p looks like $G(n,p)$ apart from the fact it has no triangles".

So if p to be this maximum density of the triangle-free process it is natural to guess that

$$\alpha(G_p) = (1+o(1))\alpha(G(n,p)) = (1+o(1))(\log n)/p.$$

While this heuristic is in fact correct, we are brushing up against a weak-spot in our heuristic: in the triangle-free graph G_p neighbourhoods are also independent sets. So a better heuristic might be

$$\alpha(G_p) = (1+o(1))\max\big\{pn, (\log n)/p\big\} = (\sqrt{2}+o(1))\sqrt{n\log n}\,. \tag{3}$$

While the maximum degree condition is actually superfluous it points to a difficulty in the proof.

2. The triangle-free nibble

We now introduce the modified version of the triangle-free process that we will study, which we call the triangle-free nibble.

Now given n, we define (somewhat arbitrarily) the parameter

$$\gamma = (\log n)^{-10}, \tag{4}$$

and describe a random process which defines a random triangle-free graph G. At stage i of the process we will have produced graphs $G_1, \ldots, G_i$ so that

$$G_{\leqslant i} = G_1 \cup \cdots \cup G_i,$$

is triangle free. Similar to the situation that we saw in our sketch of Shearer's theorem by a nibble we will ensure that

$$e(G_i) \approx \gamma n^{3/2} \qquad \text{and} \qquad e(G_{\leqslant i}) \approx i\gamma n^{3/2}.$$

Thus we will want to run this process for T rounds where

$$T = \frac{\delta}{\gamma}\sqrt{\log n}$$

steps, for some $c > 0$, to obtain the right density. Thus $G_{\leqslant T}$ will be our final graph.

At each stage of the process we will have a set of open pairs O_i. These are the edges that can be added to $G_{\leqslant i}$ without forming a triangle. We then form G_{i+1} by sampling a random graph G'_{i+1} with a fixed probability p from these pairs and then defining $G_{i+1} \subset G'_{i+1}$ to be a maximal triangle-free subgraph.

It is useful to define

$$G'_{\leqslant i} = G'_1 \cup \cdots \cup G'_i. \tag{5}$$

As we will see in a second it will be useful to shrink O_i faster than strictly nessisary to maintain some regularity properties of the process.

2.1. **Formal definition of process.** We initialize the process with $O_0 = K_n$.

We now define the process for steps $i \geqslant 1$. Given

$$G'_1, \ldots, G'_i, \quad G_1, \ldots, G_i, \quad \text{and} \quad O_1, \ldots, O_i,$$

we define $G'_{i+1}, G_{i+1}, O_{i+1}$ as follows.

1. (Nibble step) We first sample G'_{i+1} randomly from the open pairs O_i. Write

$$e(O_i) = \theta_i \binom{n}{2} \qquad \text{define} \qquad p_{i+1} = \frac{\gamma}{\theta_i\sqrt{n}}, \tag{6}$$

   and then sample

$$G'_{i+1} \sim G(O_i, p_{i+1}),$$

   if $p_{i+1} \leqslant \gamma^3$. Otherwise, stop and set $\tau = i$.

2. (Cleaning step) Let $G_{i+1} \subset G'_{i+1}$ be maximal such that $G_{\leqslant i} \cup G_{i+1}$ is triangle free.

3. (Regularization step) To define O_{i+1}, we first define,

$$O'_{i+1} = \left\{ e \in (K_n \setminus G'_{\leqslant i+1}) \cap O_i : e \text{ does not form a triangle with } G'_{\leqslant i+1} \right\}. \tag{7}$$

   Then for each $e \in O_i$, we define the quantity

$$q_e = \frac{\min_{f \in O_i} \mathbb{P}\big(f \in O'_{i+1}\big)}{\mathbb{P}\big(e \in O'_{i+1}\big)} \in [0,1]\,, \tag{8}$$

   where the above probabilities are with respect to the randomness of $G'_{i+1} \sim G(O_i, p_{i+1})$. We now define $Q_{i+1} \subset O_i$ to be a random subset where each pair $e \in O_i$ is independently included with probability q_e. Finally we define

$$O_{i+1} = Q_{i+1} \cap O'_{i+1}. \tag{9}$$

This completes a step in the process.

3. Well-running of the process

We first show that the process runs for long enough. Which amounts to studying how the graphs $G'_{\leqslant i}$ and O_i evolve. Let us write

$$e(G'_{\leqslant i}) = \psi_i \binom{n}{2} \qquad \text{and} \qquad e(O_i) = \theta_i \binom{n}{2} . \tag{10}$$

Here ψ_i is actually easy to understand, we have $\psi_i = (1+o(1))i\gamma/\sqrt{n}$ with high probability just by Chernoff. Understanding θ_i on the other hand is much more subtle.

Instead of directly studying the running time of this process - we instead look to show that the graphs $G'_{\leqslant i}$ and O_i satisfy some nice properties as the process runs, with high probability. In particular define the event $\mathcal{A}$ by defining the events $\mathcal{A}_i = \mathcal{A}_i(O_{i-1}, G'_{\leqslant i-1})$ which only depend on the randomness in the ith step. We then define the "global" event

$$\mathcal{A} = \bigcap_{0 \leqslant i \leqslant T} \mathcal{A}_i .$$

We shall show that

Lemma 3.1. *For all $i \leqslant T$ we have*

$$\mathbb{P}\big(\mathcal{A}_i \,\big|\, \mathcal{A}_{i-1}\big) \geqslant 1 - \gamma^{10}$$

and therefore $\mathbb{P}(\mathcal{A}) = 1 - o(1)$.

We now define the events $\mathcal{A}_i$. For this we define an error term

$$\varepsilon_i = i\gamma^3$$

Indeed, define $\mathcal{A}_i$ to be the intersection of the following properties of $O_i, G'_{\leqslant i}$. For all $i \geqslant 0$ we require that on the event $\mathcal{A}_i$ we have

$$\big|\theta_i - \mathbb{E}_i\, \theta_i\big| \leqslant \gamma^4 \cdot \mathbb{E}_i\, \theta_i \tag{11}$$

We then include the degree and codegree conditions into $\mathcal{A}_i$

$$\Delta(O_i) \leqslant (1+\varepsilon_i)\theta_i n \qquad \text{and} \qquad |N_i^{\circ}(x) \cap N_i^{\circ}(y)| \leqslant (1+\varepsilon_i)\theta_i^2 n, \tag{12}$$

for all $x \neq y$. We also require that on $\mathcal{A}_i$ we have the events

$$\Delta(G'_i) \leqslant (1+\varepsilon_i)\gamma\sqrt{n}\,, \tag{13}$$

for all $i \geqslant 1$. We also maintain, for all $i \geqslant 0$,

$$\Delta(G'_{\leqslant i}) \leqslant (1+\varepsilon_i)\psi_i n \quad \text{and} \quad |N'_{\leqslant i}(x) \cap N'_{\leqslant i}(y)| \leqslant i(\log n)^2\,, \tag{14}$$

whenever $x \neq y$. Finally, we require the cross-degree condition: for all $x, y \in [n]$ we have

$$|N'_{\leqslant i}(x) \cap N_i^{\circ}(y)| \leqslant (1+\varepsilon_i)\theta_i\psi_i n, \tag{15}$$

on the event $\mathcal{A}_i$

This concludes the definition of the event $\mathcal{A}_i$.