Chapter 26

Strong(er) Concentration Bounds II: Talagrand's Inequality

We now see another concentration bound, *Talagrand's inequality*, alongside some more advanced techniques in probabilistic analysis.

26.1 Coloring Locally Sparse Graphs

As we have seen multiple times so far, every graph $G = (V, E)$ with maximum degree Δ can be properly colored with $(\Delta + 1)$ colors. However, this bound in general is quite pessimistic and in fact, there are only two (families of) connected graphs that need $(\Delta + 1)$ colors: odd cycles and cliques (this is known as *Brooks' theorem* and you will prove this in Exercise 26.4). In general, there are various results in graph theory showing that the "further" a graph gets from a $(\Delta + 1)$-clique, the "fewer" colors it needs for proper coloring. In this chapter, we examine a classical problem in this context, coloring *locally sparse* graphs.

Problem 26.1. Let $\varepsilon \in (0, 1)$ be a parameter and $\Delta_0 = \Delta_0(\varepsilon)$ be an integer, sufficiently large as a function of ε. Suppose $G = (V, E)$ is a Δ-regular graph for some $\Delta \geqslant \Delta_0$ such that for any vertex $v \in V$, at least $\varepsilon\Delta^2$ edges are missing between vertices in $N(v)$; in other words, at least $\varepsilon\Delta^2$ pairs among neighbors of v are *not* adjacent to each other.

Can we always color G with $\Delta - \Omega_\varepsilon(\Delta)$ colors?

In words, Problem 26.1 asks that if the neighborhood of every vertex is not too dense—so the graph is locally sparse—can we color it with considerably fewer than Δ colors?[1] The answer to this question turns out to be *yes* due to a classical result of [MR97] in graph theory.

Theorem 26.2 ([MR97])**.** *There is an absolute constant $\gamma \in (0, 1)$ such that for every $\varepsilon \in (0, 1)$, every Δ-regular graph G in Problem 26.1 can be properly colored with $\Delta - \gamma \cdot \varepsilon\Delta$ colors.*

We prove this theorem in the rest of this chapter. The proof consists of two steps. The first and main step uses the probabilistic method to find a coloring of a fraction of vertices with the property that *all* vertices see "many" pairs of same-colored neighbors. This is a highly useful property when it comes to coloring: for a vertex $v \in V$, a same-colored pair of neighbors the

[1]We note that the requirement for G to be Δ-regular, as opposed to having maximum degree Δ, is benign and unnecessary (see Exercise 26.3); however, to avoid distraction from the main ideas, we have assumed Δ-regularity in Problem 26.1 explicitly.

degree of v by two at the cost of using only one color originally available to v. Thus, if v has "enough" such pairs, its remaining degree (among uncolored vertices) will become smaller than the number of colors available to it. The second part of the argument then greedily colors each remaining vertex using this guarantee.

For the rest of the proof, define the number of available colors as:

$$C := \Delta - \gamma \cdot \varepsilon \cdot \Delta \tag{26.1}$$

for some $\gamma \in (0, 1)$ to be defined later.

Consider the following algorithm for partially coloring vertices of G:

Algorithm 26.3.

For every vertex $v \in V$:

1. sample a tentative color $c(v) \in [C]$ uniformly at random and independently;
2. if the tentative color $c(v) \neq c(u)$ for every neighbor $u \in N(v)$, **color** v with $c(v)$.

Let $\varphi : V \to [C] \cup \{\perp\}$ denote the partial coloring returned by Algorithm 26.3 where $\varphi(v) = \perp$ means the vertex v was not assigned a color. Notice that in the partial coloring φ, there are no edges $(u, v) \in E$ such that $\varphi(u) = \varphi(v) = c$ for $c \in [C]$ by construction. In other words, φ is a proper coloring on vertices it does not leave blank.

To continue, we need some definitions first. For any vertex $v \in V$, define:

- $\texttt{Avail}(v)$: the colors in $[C]$ that are not used to color any neighbor of v in φ in Algorithm 26.3; these colors are *available* to v for the next step of the coloring. We further denote $\texttt{avail}(v) := |\texttt{Avail}(v)|$.
- $N_{\perp}(v)$: the set of uncolored neighbors of v in φ in Algorithm 26.3; these are neighbors of v that need to be colored in the next step. We further denote $\texttt{deg}_{\perp}(v) := |N_{\perp}(v)|$.

The following lemma is the main part of the argument.

Lemma 26.4. *For any fixed $v \in V$, after running Algorithm 26.3,*

$$\Pr\Big(\texttt{avail}(v) < \texttt{deg}_{\perp}(v) + 1\Big) \leqslant \exp\left(-\gamma \cdot \varepsilon^2 \cdot \Delta\right).$$

The proof of Lemma 26.4 is where we introduce and use Talagrand's inequality.

26.1.1 Proof of Lemma 26.4

Throughout this proof, fix a vertex $v \in V$. Define $F(v) \subseteq \binom{N(v)}{2}$ as an arbitrary set of $\varepsilon\Delta^2$ pairs of vertices in $N(v)$ that are not adjacent to each other (such a set exists by the premise of Theorem 26.2). We refer to each $(u, w) \in F(v)$ as an *anti-edge*. Define X as the number of colors that both endpoints of some anti-edge in $F(v)$ sampled and retained. All these colors are used by φ in the neighborhood of v at least twice. We thus have,

$$\texttt{avail}(v) = C - X - \sum_{c \in [C]} \mathbb{I}(\exists u \in N(v) : \varphi(u) = c \text{ but } c \text{ is not counted toward } X)$$

$$\texttt{deg}_{\perp}(v) \leqslant \Delta - 2X - \sum_{c \in [C]} \mathbb{I}(\exists u \in N(v) : \varphi(u) = c \text{ but } c \text{ is not counted toward } X);$$

the first line is because all colors counted toward X are used to color vertices in $N(v)$, and the second line is because each such color removes at least two neighbors of v (by the not-sampling requirement in X) and any other color used in φ removes at least one neighbor of v.

Putting these together we have

$$\texttt{avail}(v) \geqslant \texttt{deg}_{\perp}(v) + C - \Delta + X.$$

Thus, to prove Lemma 26.4, we can focus on proving that

$$\Pr\Big(X \leqslant \Delta - C(= \gamma\varepsilon\Delta)\Big) \leqslant \exp\left(-\gamma \cdot \varepsilon^2 \Delta\right), \tag{26.2}$$

which immediately implies the lemma.

We have gotten closer to the familiar territory of concentration bounds. We first lower bound the expectation of X (we will be very loose with the constants in the calculations to avoid distracting from the main message).

Claim 26.5. $\mathbb{E}[X] \geqslant e^{-12} \cdot \varepsilon \cdot \Delta$.

Proof. To prove the lower bound on X, it will be easier to work with a proxy random variable $\tilde{X}$. Firstly, define the random variables $\left\{\tilde{X}_{uw}\right\}_{(u,w)\in F(v)}$ where:

> $\tilde{X}_{uw} = 1$ iff $c(u) = c(w) = c$ for some $c \in [C]$ and c is not sampled by any other $z \in N(v) \cup N(u) \cup N(w)$.

Define $\tilde{X} := \sum_{(u,w)\in F(v)} X_{uw}$. Notice that $X \geqslant \tilde{X}$ point-wise because $\tilde{X}$ counts the number of colors that are retained by a pair in $F(v)$ and were never sampled by any other vertex in $N(v)$ (the first requirement is the same as X and the second requirement makes $\tilde{X}$ more strict).

By linearity of expectation,

$$\mathbb{E}\left[\tilde{X}\right] = \sum_{(u,w)\in F(v)} \mathbb{E}\left[\tilde{X}_{uw}\right] = \sum_{(u,w)\in F(v)} \Pr\left(\tilde{X}_{uw} = 1\right).$$

Fix any $(u,w) \in F(v)$. For $\tilde{X}_{uw}$ to be 1, there should be a color $c \in [C]$ such that $c(u) = c(w) = c$ and c is not sampled by $N(v) \cap N(u) \cap N(w)$. Given that these events are disjoint for different choices of c, we have

$$\begin{aligned}
\Pr\left(\tilde{X}_{uw} = 1\right) &= \sum_{c\in[C]} \Pr\left(c(u) = c(w) = c\right) \cdot \prod_{z\in N(v)\cup N(u)\cup N(w)\setminus\{u,w\}} \Pr\left(c(z) \neq c\right) \\
&\qquad \text{(because choice of colors by vertices are independent)} \\
&\geqslant C \cdot \frac{1}{C^2} \cdot \left(1 - \frac{1}{C}\right)^{3\Delta} \\
&\qquad \text{(as } c(\cdot) \text{ is uniform over } [C] \text{ and } |N(v) \cup N(u) \cup N(w)| \leqslant 3\Delta) \\
&\geqslant \frac{1}{C} \cdot \exp\left(-\frac{6\Delta}{C}\right) \qquad \text{(as } 1 - x \geqslant e^{-2x} \text{ for } x \in (0, 1/2)) \\
&\geqslant \frac{1}{C} \cdot e^{-12}. \qquad \text{(as } C \geqslant \Delta/2)
\end{aligned}$$

Thus, we have,

$$\mathbb{E}\left[\tilde{X}\right] \geqslant |F(v)| \cdot \frac{1}{C} \cdot e^{-12} \geqslant e^{-12} \cdot \varepsilon \cdot \Delta,$$

given $|F(v)| = \varepsilon\Delta^2$ and $\Delta \geqslant C$. Since $\mathbb{E}[X] \geqslant \mathbb{E}\left[\tilde{X}\right]$, we are done. □

At this point, given Claim 26.5, proving Eq (26.2) becomes precisely a *concentration bound* by setting $\gamma \approx e^{-12}$. The question then is what type of concentration bound we can use here,

The variables in $\left\{\tilde{X}_{uw} : (u, w) \in F(v)\right\}$ used in the proof of Claim 26.5 are clearly *not* independent of each other, so we cannot use the Chernoff bound. We could (in principle) set up a proper martingale and attempt to use Azuma's inequality (Proposition 25.7); however, it is not that clear how to define such martingales (you are encouraged to try doing this; also try McDiarmid's inequality of Exercise 25.2 to see if it works here). Instead, we use this opportunity to introduce the *Talagrand's inequality* that provides a quite strong concentration bound which is also relatively easy to use (certainly easier, at least here, than setting up proper martingales and applying Azuma's inequality).

Talagrand's inequality. A function $f(x_1, \ldots, x_n)$ is called ***c*-Lipschitz** if changing any x_i can affect the value of f by at most c. Additionally, f is called ***r*-certifiable** iff whenever $f(x_1, \ldots, x_n) \geqslant s$, there exist at most $r \cdot s$ variables $x_{i_1}, \ldots, x_{i_{r \cdot s}}$ so that knowing the values of these variables certifies $f \geqslant s$.

Proposition 26.6 (Talagrand's inequality; cf. [MR02])**.** *Let $X_1, \ldots, X_n$ be n independent random variables and $f(X_1, \ldots, X_n)$ be a c-Lipschitz, r-certifiable function. For any $1 \leqslant t \leqslant \mathbb{E}[f]$,*

$$\Pr\left(|f - \mathbb{E}[f]\,| > t + 60c\sqrt{r \cdot \mathbb{E}[f]}\right) \leqslant 4\exp\left(-\frac{t^2}{8c^2 r\,\mathbb{E}[f]}\right).$$

Remark. Talagrand's inequality is natively a concentration bound around the *median* of f not the *expectation* (unlike all other concentration inequalities we have seen so far in the book). For instance, even by setting $t \to 0$, one cannot use Proposition 26.6 to argue the variable will have a value arbitrary close to its expectation with non-zero probability (again, unlike all other concentration inequalities we have seen so far).

Despite that, Proposition 26.6 can till state the bound around the expectation using the fact that for a c-Lipschitz, r-certifiable function f, $|\mathbb{E}[f] - \text{med}(f)| = O(c\sqrt{r\,\mathbb{E}[f]})$.

Remark. Perhaps the most common concentration inequality used in probabilistic analysis is the Chernoff bound. Despite its simplicity, Chernoff bound is applicable to a surprisingly large number of situations. Yet, there are still some limits to the power of this basic tool. In such cases, one may want to consider stronger tools such as Talagrand's, McDiarmid's, or Azuma's inequalities. At a very high level, the common theme of all these concentration inequalities is the following: if we have a random variable T which is a function of n independent trials $X_1, \ldots, X_n$ and T is not "very sensitive" to the outcome of each trial[a], then T is "more or less" concentrated around its expectation.

[a] This is captured by the notions of Lipschitz constant and certifiability in Talagrand's inequality.

Back to the proof of Lemma 26.4. Let us now see how we can use Proposition 26.6 to prove Eq (26.2). Recall that our goal is basically to upper bound the following probability:

$$\Pr\left(X \leqslant \frac{1}{2} \cdot \mathbb{E}[X]\right).$$

Here X is a function of the colors sampled in Algorithm 26.3 for vertices in the (closed) two-hop neighborhood of v, denoted by $N_2(v)$. Let $\{\xi_u\}_{u \in N_2(v)}$ denote the random choices of these vertices and note that these are all independent and we can write X as a deterministic function of them. Thus, we have the right ingredient for applying Proposition 26.6.

To apply Proposition 26.6 we need to check Lipschitzness and certifiability. It is easy to see that X is $O(1)$-Lipschitz: if we change the value of some ξ_u, meaning the color sampled for a vertex u, then at most two colors in $[C]$ can be affected in terms of whether or not they were counted in X (the two choices for u) and each color can only change the value of X by one. Unfortunately however, X is not easily certifiable: to certify a color c is counted in X, we need to reveal the color of $\approx \Delta$ vertices to ensure the color can be retained also by the endpoints of the anti-edge. It is not hard to see that applying Proposition 26.6 for bounding X as an $O(1)$-Lipschitz, $O(\Delta)$-certifiable function is not going to give us a strong concentration bound.

In the following we are going to show a simple but highly useful trick when it comes to probabilistic analysis, by bounding X indirectly. Define the following two additional variables with respect to Algorithm 26.3:

- W: number of colors that are sampled by endpoints of *at least* one anti-edge in $F(v)$, regardless of whether or not the color is assigned to them.
- Z: number of colors that are sampled by endpoints of *at least* one anti-edge in $F(v)$, but is *not* assigned to *at least one* of the endpoints.

Both these variables are also functions of $\{\xi_u\}_{u \in N_2(v)}$. Notice that $X = W - Z$ this way. The nice thing is that both W and Z are $\Theta(1)$-certifiable (for W point to endpoints of an anti-edge in $F(v)$ that sampled the color; for Z additionally point to one of the neighbors of this anti-edge that also sampled the color, hence not allowing one of them to retain it). They are also both $\Theta(1)$-Lipschitz for the same reason X is: changing the choice of one color for a vertex can only affect the two colors involved (the original one and the changed one). As such, we can apply Talagrand's inequality (Proposition 26.6) to obtain that (we only write the bound for W; the same exact argument works also for Z):

$$\begin{aligned}\Pr\left(|W - \mathbb{E}[W]| \geqslant \mathbb{E}[X]/10\right) &\leqslant \exp\left(-\Theta(1) \cdot \frac{(\mathbb{E}[X] - \Theta(1)\sqrt{\Delta})^2}{\Delta}\right) && (W \leqslant \Delta/2 \text{ always}) \\ &\leqslant \exp\left(-\Theta(\varepsilon^2) \cdot \Delta\right).\end{aligned}$$

(by Claim 26.5 on expected value of X, and since Δ is sufficiently large so $\varepsilon\Delta \gg \sqrt{\Delta}$)

As such, we obtain that w.h.p. both W and Z are concentrated and thus also w.h.p.,

$$X = W - Z \geqslant \mathbb{E}[W] - \mathbb{E}[X]/10 - (\mathbb{E}[Z] + \mathbb{E}[X]/10) = \mathbb{E}[X] - \mathbb{E}[X]/5 = (4/5) \cdot \mathbb{E}[X].$$

Thus, by a union bound,

$$\Pr\left(X \leqslant \frac{1}{2} \cdot \mathbb{E}[X]\right) \leqslant \Pr\left(X \leqslant \frac{4}{5} \cdot \mathbb{E}[X]\right) \leqslant \exp\left(-\Theta(\varepsilon^2) \cdot \Delta\right).$$

Finally, since $\mathbb{E}[X] \geqslant e^{-12} \cdot \varepsilon \cdot \Delta$, taking γ to be the minimum of $e^{-12}/2$ and the unspecified constant in the $\Theta(\varepsilon^2)$-term above, concludes the proof of Lemma 26.4.

26.1.2 Concluding the Proof of Theorem 26.2

There is only one more step needed at this point. By Lemma 26.4, we know that for any fixed $v \in V$, after running Algorithm 26.3,

$$\Pr\left(\texttt{avail}(v) < \texttt{deg}_{\perp}(v) + 1\right) \leqslant \exp\left(-\gamma \cdot \varepsilon^2 \cdot \Delta\right).$$

If Δ was large enough, say, $\Delta \gtrsim \varepsilon^{-2} \log n$, then the RHS above would have been $1/\text{poly}(n)$, enough to do a union bound over all vertices. But, we only know that Δ is large enough as a function of ε in Theorem 26.2. However, recall that our goal is not to design an algorithm for finding the coloring, but only to show its existence, i.e., we have a probabilistic-method proof here. The solution thus is easy: apply LLL (Theorem 23.3) instead.

For any vertex v, define a bad event:

$$B_v \quad \equiv \quad (\texttt{avail}(v) < \texttt{deg}_{\perp}(v) + 1)\,.$$

Notice that B_v depends solely on the colors sampled by vertices in the (closed) two-hop neighborhood of v, namely, $N_2(v)$. As such, B_v is independent of all $\{B_u\}_u$ as long as $N_2(v) \cap N_2(u) = \emptyset$. This means that the degree of the dependency graph between the events B_v's is at most Δ^4. Since for large enough Δ (as a function of ε),

$$\exp\left(-\gamma \cdot \varepsilon^2 \cdot \Delta\right) \cdot \left(\Delta^4 + 1\right) \ll \frac{1}{e},$$

we can apply LLL (Theorem 23.3) to prove the *existence* of a partial coloring $\varphi : V \to [C] \cup \{\perp\}$ wherein the event B_v does *not* happen for any $v \in V$.

The last step of the proof is to extend the coloring φ to a proper C-coloring of the entire graph greedily: go over the vertices in some arbitrary order and for any uncolored vertex $v \in V$ (under φ), pick a color from $\texttt{Avail}(v)$ that has not yet been assigned to a neighbor of v in this greedy pass; since $\texttt{avail}(v) \geqslant \texttt{deg}_{\perp}(v) + 1$, such a color always exists.

This way, we find a proper C-coloring of G, concluding the proof of Theorem 26.2.

Exercises

Exercise 26.1. Let $X_1, \ldots, X_n$ be i.i.d. uniform on $[0, 1]$ (so the X_i are distinct with probability 1), and let

$$L \;=\; L(X_1, \ldots, X_n)$$

denote the length of the longest *strictly increasing* subsequence: the largest k for which there are indices $i_1 < \cdots < i_k$ with $X_{i_1} < \cdots < X_{i_k}$. This exercise applies McDiarmid's inequality (Exercise 25.2) to L.

1. Show that L satisfies the bounded-differences hypothesis of Exercise 25.2 with $c = 1$: for every $i \in [n]$ and all inputs that differ only in the i-th coordinate,

$$|L(x_1, \ldots, x_i, \ldots, x_n) - L(x_1, \ldots, x_i', \ldots, x_n)| \leqslant 1.$$

2. Conclude that for every $b > 0$,

$$\Pr\left(|L - \mathbb{E}\left[L\right]| \geqslant b\right) \leqslant 2 \exp\left(-\frac{b^2}{2n}\right),$$

so that L lies within $O(\sqrt{n})$ of its expectation with any desired constant probability.

Exercise 26.2. Let $G \sim \mathbb{G}(n, 1/2)$ and let $\mu(G)$ be the size of a maximum matching of G. As in the vertex-exposure analysis of the chromatic number in the previous chapter, reveal G through the independent coordinates $Y_1, \ldots, Y_n$, where Y_i records which of the vertices $1, \ldots, i-1$ are neighbors of vertex i; together the Y_i's determine G, and $\mu(G) = f(Y_1, \ldots, Y_n)$ where f outputs the matching number.

1. Prove that f is 1-Lipschitz.

2. Apply McDiarmid's inequality of Exercise 25.2 with the n coordinates $Y_1, \ldots, Y_n$ to conclude a concentration bound for $\mu(G)$.

Exercise 26.3. Prove that every graph of maximum degree Δ whose neighborhoods each miss at least $\varepsilon \cdot \Delta^2$ pairs is a subgraph of a Δ-regular graph with the same property

Exercise 26.4 (♦). Brooks' theorem states that every connected graph of maximum degree Δ which is neither an odd cycle nor a $(\Delta+1)$-clique can be properly colored with Δ colors. Prove Brooks' theorem.

Exercise 26.5 (♦). Use the algorithmic LLL and ideas behind it in Chapter 24 to turn the proof of Theorem 26.2 algorithmic. In particular, design a randomized algorithm with running time $O(n\Delta^2)$ for finding a $(1-\Omega(\varepsilon))\Delta$-coloring of locally-sparse graphs with parameter $\varepsilon \in (0, 1)$ and sufficiently large Δ as a function of ε.