Moore bound for irregular graphs

Recall that we want to prove the following theorem:

Theorem 1. *Let G be an n-vertex graph with $\delta(g) \geq 2$, and with girth $g(G) \geq 2k+1$ and average degree $\bar{d} = \frac{2m}{n}$. Then*

$$n \geq 1 + \bar{d} \sum_{i=0}^{k-1} (\bar{d}-1)^i.$$

We saw last time that this implies:

Corollary 2. $ex(n, \{C_3, C_4, ..., C_{2k}\}) \leq \frac{1}{2}\left(n^{1+\frac{1}{k}} + n\right)$

Let us now describe the idea of the proof:

Take a random walk, starting at a random vertex:

a) Choose an initial vertex v_0 at random. To avoid getting stuck in sparse parts of G, choose v_0 according to the degree of the vertices, i.e.

$$\mathbb{P}(v_0 = v) = \frac{d(v)}{\sum_u d(u)} =: \pi(v);$$

b) Choose an edge $\overrightarrow{e_1}$ uniformly at random from edges at v_0 and get a new vertex v_1;

c) Choose a different edge $\overrightarrow{e_2}$ from v_1 uniformly at random from among all edges at v_1, except for $\overrightarrow{e_1}$ and get a new vertex v_2;

d) Repeat until we get a walk of the desired length.

Claim 3. *If v_0 is distributed as π, then for any $j \geq 1$, v_j is also distributed as π and $\overrightarrow{e_j}$ is uniformly distributed over $\overrightarrow{E}$.*

Proof. We prove the distribution of e_j by induction on j:

$j = 1$: Fix $(u, v) \in \overrightarrow{E}$. We have to show that $\mathbb{P}(\overrightarrow{e_1} = (u, v)) = \frac{1}{n\bar{d}}$. We have

$$\mathbb{P}(\overrightarrow{e_1} = (u,v)) = \mathbb{P}(v_0 = u \wedge \overrightarrow{e_1} = (u,v)) = \mathbb{P}(v_0 = u)\mathbb{P}(\overrightarrow{e_1} = (u,v)|v_0 = u) = \frac{d(u)}{n\bar{d}}\frac{1}{d(u)} = \frac{1}{n\bar{d}}.$$

$j \to j+1$: We have to show that $\mathbb{P}(\overrightarrow{e_{j+1}} = (u, v)) = \frac{1}{n\bar{d}}$.

$$\begin{aligned}\mathbb{P}(\overrightarrow{e_{j+1}} = (u,v)) &= \sum_{\vec{e}} \mathbb{P}(\overrightarrow{e_{j+1}} = (u,v)|\overrightarrow{e_j} = \vec{e})\mathbb{P}(\overrightarrow{e_j} = \vec{e}) \overset{I.H.}{=} \frac{1}{n\bar{d}} \sum_{\vec{e}} \mathbb{P}(\overrightarrow{e_{j+1}} = (u,v)|\overrightarrow{e_j} = \vec{e}) \\ &= \frac{1}{n\bar{d}} \sum_{x \in N(u)\setminus\{v\}} \mathbb{P}(\overrightarrow{e_{j+1}} = (u,v)|\overrightarrow{e_j} = (x,u)) = \frac{1}{n\bar{d}}(d(u)-1)\frac{1}{d(u)-1} \\ &= \frac{1}{n\bar{d}}.\end{aligned}$$

The distribution of v_j follows from that of e_j: $\mathbb{P}(v_j = v) = \sum_{u \in N(v)} \mathbb{P}(\overrightarrow{e_j} = (u,v)) = \frac{d(v)}{n\bar{d}}$.

□

Lemma 4. *For an arbitrary graph G with minimum degree at least 2, we have*

$$\mathbb{E}_{v\sim\pi}[n_i(v)] \geq \bar{d}(\bar{d}-1)^{i-1}$$

Observe how the previous lemma implies the theorem. In an n-vertex graph with girth at least $2k+1$, $n \geq \sum_{i=0}^{k} n_i(v)$ for any vertex v. Now if we choose v randomly according to π and use linearity of expectation, we can conclude that *there exists* a vertex v_0 with

$$\sum_{i=0}^{k} n_i(v_0) \geq \sum_{i=0}^{k} \mathbb{E}\left[n_i(v)\right] \geq 1 + \bar{d} \sum_{i=0}^{k-1} (\bar{d}-1)^i.$$

In order to prove the Lemma, we will introduce the concept of *Entropy* and see some basic properties.

Entropy

Definition 5. Let X be a discrete random variable, and let p be the probability distribution function, i.e. $p(x) = \mathbb{P}(X = x)$. Then the *entropy* of X is

$$H(X) = \mathbb{E}_x[-\log_2(p(x))] = -\sum_x p(x) \log_2 p(x)$$

Remark 6.

1. For $z = 0$, we set $z \log_2 z := \lim_{x \to 0} x \log_2 x = 0$.
2. The entropy depends only on the distribution of the random variable, not on its values.
3. The entropy can be thought of as a measure of uncertainty of the random variable.
4. All logarithms considered in this section are assumed to be base 2.

Examples.

a) If $X \sim \text{Bernoulli}(p)$, then $H(X) = -p \log p - (1-p) \log(1-p)$.
This is the so called *binary entropy function.*

b) If X is the uniform distribution on an n-element set, then $H(X) = -\sum_{i=1}^{n} \frac{1}{n} \log \frac{1}{n} = \log n$.

We have the following upper bound on the entropy of a random variable.

Propositions 7. *If X is a random variable taking on n values, then $H(X) \leq \log n$.*

Definition 8. Given two random variables X, Y, the *joint entropy* of X and Y is

$$H(X,Y) = \mathbb{E}_{x,y}[-\log(p(x,y))]$$

and the *conditional entropy* of X given Y is

$$H(X|Y) := \mathbb{E}_y[H(X|\{Y=y\})] = \mathbb{E}_y[-\sum_x \mathbb{P}(X=x|Y=y) \log \mathbb{P}(X=x|Y=y)].$$

The link between the joint entropy and the conditional entropy is given by the following proposition.

Propositions 9. *If X, Y are two random variables, then*

$$H(X,Y) = H(Y) + H(X|Y).$$

Proof. Homework. □

We are now ready to proof the Lemma 4

Proof. (of Lemma 4) Note that $n_i(v)$ is the size of the range of the random variable $(\overrightarrow{e}_1, \overrightarrow{e}_2, \ldots, \overrightarrow{e}_i | v_0 = v)$, for every vertex $v \in V(G)$. Hence by the concavity of the log-function and Proposition 7, we have

$$\log(\mathbb{E}_{v\sim\pi}[n_i(v)]) \geq \mathbb{E}_{v\sim\pi}[\log n_i(v)] \geq \mathbb{E}_{v\sim\pi}\left[H(\overrightarrow{e_1}\overrightarrow{e_2}\ldots\overrightarrow{e_i}|v_0 = v)\right].$$

By Proposition 9 and linearity of expectation, this is further equal to

$$\mathbb{E}_{v\sim\pi}[H(\overrightarrow{e_1}|v_0 = v)] + \sum_{j=2}^{i} \mathbb{E}_{v\sim\pi}[H(\overrightarrow{e_j}|\overrightarrow{e_{j-1}}, ..., \overrightarrow{e_0}, v_0 = v)]$$

The edge e_1 is chosen uniformly out of $d(v)$ edges, hence by Example (b) we have

$$\begin{aligned}
&= \mathbb{E}_{v\sim\pi}(\log d(v)) + \sum_{j=2}^{i} \mathbb{E}_{v\sim\pi}[H(\overrightarrow{e_j}|\overrightarrow{e_{j-1}}, v_{j-1}]\\
&= \mathbb{E}_{v\sim\pi}(\log d(v)) + \sum_{j=2}^{i} \mathbb{E}_{v_{j-1}\sim\pi}[\log(d(v_{j-1}) - 1)]\\
&= \mathbb{E}_{v\sim\pi}[\log(d(v)(d(v) - 1)^{i-1})]\\
&\overset{\text{by definition of }\pi}{=} \frac{1}{n\bar{d}} \sum_{v} d(v) \log(d(v)(d(v) - 1)^{i-1})\\
&\overset{x\mapsto x\log x(x-1)^{i}-1\text{is convex on }x\geq 2}{\geq} \frac{1}{n\bar{d}} n\bar{d} \log(\bar{d}(\bar{d} - 1)^{i-1})\\
&= \log(\bar{d}(\bar{d} - 1)^{i-1}).
\end{aligned}$$

Hence $\mathbb{E}_{x\sim\pi}[n_i(v)] \geq \bar{d}(\bar{d} - 1)^{i-1}$ as the logarithm function is increasing. □

Upper bound on the Turán number of C_{2k}

Instead of forbidding *every* cycle of length up to $2k$, we now only disallow C_{2k}. We will prove that the $\Theta(n^{1+\frac{1}{k}})$ can be saved.

Theorem 10. *Let $k \geq 2$. Then $ex(n, C_{2k}) \leq 8(k - 1)n^{1+\frac{1}{k}}$.*

Remark 11. An upper bound of the order $n^{1+\frac{1}{k}}$ was first proved by Bondy and Simonovits. The Proof we give here is based on Verstraete's and gives a better constant factor. Later Pikhurko removed the 8 factor and replaced it with $(1 + o(1))$ and recently Bukh and... gave a constant factor sublinear in k. However, it is still not known, whether one could remove the $k-$dependence of the constant factor.

Proof. We may assume $k \geq 3$ (for $k = 2$, that is for $\text{ex}(n, C_4)$, we already have proved an upper bound with a better constant factor).
Given a graph G with $8(k-1)n^{1+\frac{1}{k}}$ edges, we can pass to a bipartite subgraph $H \subseteq G$ with minimum degree greater than $4(k - 1)n^{\frac{1}{k}}$.
Pick an arbitrary vertex $v_0 \in V$, and define

$$V_i := \{v \in V : d_H(v_0, v) = i\}$$

Note that V_i is an independet set (H is bipartite) and the neighbourhood of V_i is $N(V_i) = V_{i-1} \cup V_{i+1}$. Let

$$l = \min\{i : e(V_i, V_{i+1}) \geq 2(k - 1)|V_{i+1}|\},$$

the smallest index such that the vertex set of V_{l+1} has a "large" average degree "backwards". Intuitively, if the vertices of some V_i have a small average degree "backwards", then they must have many edges "forward", towards V_{i+1}. Now if even $i + 1 \leq l$, then V_{i+1} still send only a small number of edges "backwards". These two things: V_i sending many edges towards V_{i+1} and V_{i+1} sending few edges back towards V_i are only possible if the vertex sets "expand", that is, if V_{i+1} is much larger than V_i. This cannot happen many times, since the size is eventually bounded by the number of vertices n. We formalise this in the next Claim.

Claim 12. $l \leq k-1$.

Proof. For $i < l$ on the one hand we have $e(V_i, V_{i+1}) < 2(k-1)|V_{i+1}|$. On the other hand, since $e(V_{i-1}, V_i) < 2(k-1)|V_i|$, using $\delta(H) > 4(k-1)n^{\frac{1}{k}}$ we also have that $e(V_i, V_{i+1}) > 4(k-1)n^{\frac{1}{k}}|V_i| - 2(k-1)|V_i| \geq 2(k-1)n^{\frac{1}{k}}|V_i|$. These two estimates imply that for every $i < l$, we have $2(k-1)|V_{i+1}| > e(V_i, V_{i+1}) > 2(k-1)n^{\frac{1}{k}}|V_i|$ and hence

$$|V_{i+1}| > n^{\frac{1}{k}}|V_i| > n^{\frac{2}{k}}|V_{i-1}| > ... > n^{\frac{i+1}{k}}|V_0| = n^{\frac{i+1}{k}}.$$

For $i = k-1$ this would give $|V_k| > n$, a contradiction. So $k-1$ must be at least l. □

Now recall that $E(V_l, V_{l+1}) \geq 2(k-1)|V_{l+1}|$. Counting from the other side, $l-1 < l$ and $\delta(H) \geq 4(k-1)n^{1+\frac{1}{k}}$ still implies that $e(V_l, V_{l+1}) > 4(k-1)n^{\frac{1}{k}}|V_l| - 2(k-1)|V_l| \geq 2(k-1)n^{1+\frac{1}{k}}|V_l|$. So the average degree of $H[V_l \cup V_{l+1}]$ is at least $2(k-1)$ and consequently there exists a subgraph $H_0 \subseteq H[V_l \cup V_{l+1}]$ with minimum degree $\delta(H_0) > k-1$.
We will build a cycle of length at least $2k$ in H_0. To do so, let P be a longest path in H_0, and let w be an endpoint. By maximality of P, $N(w) \subseteq V(P)$. Since $|N(w)| \geq \delta(H_0) \geq k$, and the neighbours of w are not adjacent on P, the distance of w to its farthest neighbour on P is at least $2k-1$. Adding this edge to the path closes a cycle of length at least $2k$. This cycle then also has a chord, since $|N(w)| \geq \delta(H_0) \geq 3$. Call this cycle with a chord K.
We need the following lemma:

Lemma 13. *Let K consists of a cycle of length m and a chord, and let $c : V(K) \to \{R, B\}$ be an improper colouring of its vertices. Then for every $l < m$, there is a path of length l in K whose endpoints have different colours.*

Proof. Homework. □

Consider now a minimal subtree $T' \subseteq H[V_0 \cup V_1 \cup \ldots \cup V_l]$ that contains the vertex set $V(K) \cap V_l$. Let q be the height of T, where $1 \leq q \leq l$. Note that since $\delta(H) \geq 3$, the root of this tree has degree at least two.
Choose a branch of the root and colour all its descendants in $V(K) \cap V_l$ `red` and all others `blue`. This gives rise to an improper colouring of K and so, by the above lemma, there is an $R-B$ path of length $2(k-q)$. The `red` vertices are all in V_l and hence this path must start in V_l. Since its length is even, it will also end in V_l. Now append the two q-paths from the endpoints to the root of T'. Due to the minimality of T' these paths are disjoint and hence we obtained a cycle of length $2(k-q)+q+q = 2k$, finishing the proof.

□

Remark 14. The proof in fact gives cycles of many different even length: just consider all $R-B$ paths of different length.