# Advanced Discrete Mathematics 2013 – Problem Set 10 – Solutions

1. *Consider a bipartite graph $(A, B)$ and numbers $p \in [0, 1]$ and $\varepsilon > 0$ such that the following property holds: for any $X \subseteq A$ and $Y \subseteq B$ such that $|X| \geq \varepsilon|A|$ and $|Y| \geq \varepsilon|B|$ we have $|d(X, Y) - p| \leq \varepsilon$. Prove that $(A, B)$ is $2\varepsilon$-regular.*

   By the triangle inequality, for any $X \subseteq A$ and $Y \subseteq B$ such that $|X| \geq \varepsilon|A|$ and $|Y| \geq \varepsilon|B|$ we have $|d(X, Y) - d(A, B)| \leq |d(X, Y) - p| + |d(A, B) - p| \leq 2\varepsilon$.

2. *Let $(A, B)$ be an $\varepsilon$-regular bipartite graph. Let $A' \subseteq A$ and $B' \subseteq B$ be such that $|A'| \geq \alpha|A|$ and $|B'| \geq \alpha|B|$, where $\alpha > \varepsilon$. Prove that $(A', B')$ is $\varepsilon'$-regular with $\varepsilon' = \max\{2\varepsilon, \varepsilon/\alpha\}$.*

   By the $\varepsilon$-regularity of $(A, B)$, we have $|d(A', B') - d(A, B)| \leq \varepsilon$. Let $X \subseteq A'$ and $Y \subseteq B'$ be such that $|X| \geq \varepsilon'|A'| \geq \varepsilon|A|$ and $|Y| \geq \varepsilon'|B'| \geq \varepsilon|B|$. Again by the $\varepsilon$-regularity of $(A, B)$, we have $|d(X, Y) - d(A, B)| \leq \varepsilon$. Therefore, by the triangle inequality, we have $|d(X, Y) - d(A', B')| \leq |d(X, Y) - d(A, B)| + |d(A', B') - d(A, B)| \leq 2\varepsilon \leq \varepsilon'$.

3. *Prove that Szemerédi's regularity lemma is equivalent to the following statement:*
   *For every $\varepsilon > 0$ there is $M \in \mathbb{N}$ such that every graph $G$ with at least $1/\varepsilon$ vertices admits a partition $V(G) = V_1 \cup \ldots \cup V_k$ such that*

   - $1/\varepsilon \leq k \leq M$,
   - $|V_1| \leq \ldots \leq |V_k| \leq |V_1| + 1$,
   - *all but at most $\varepsilon k^2$ pairs $(V_i, V_j)$ with $1 \leq i < j \leq k$ are $\varepsilon$-regular.*

   The general idea is to apply one statement with a much smaller $\varepsilon$ and then to modify the obtained partition to get the other statement. Unfortunately, the actual computation is quite technical.

   First, we deduce the statement in the problem from Szemerédi's regularity lemma. Let $G$ be a graph with at least $n_0$ vertices, with $n_0$ to be chosen later. Apply the regularity lemma to $G$ and $\varepsilon'$ with $\varepsilon'$ to be chosen later (for this we need $n_0 \geq 1/\varepsilon'$). This gives a partition $V(G) = V_0' \cup V_1' \cup \ldots \cup V_k'$ with $|V_0'| \leq \varepsilon' n$. Distribute the elements $V_0'$ into $V_1', \ldots, V_k'$ as equally as possible, obtaining sets $V_1, \ldots, V_k$. For a set $X \subseteq V_i$ and $X' = X \cap V_i'$, if $|X| \geq \varepsilon|V_i|$, then we have

   $$|X \setminus X'| \leq \frac{|V_0|}{k} + \alpha \leq \frac{\varepsilon' n}{k} + \alpha \leq \varepsilon'\left(\frac{n}{k} + \alpha\right) + 1 = \varepsilon'|V_i| + 1 \leq \beta|V_i| \leq \frac{\beta}{\varepsilon}|X|,$$

   where

   $$\alpha \in \left\{ -\frac{|V_0| \bmod k}{k}, 1 - \frac{|V_0| \bmod k}{k} \right\} \quad \text{and} \quad \beta = \varepsilon' + \frac{M(\varepsilon')}{n_0 - M(\varepsilon')}.$$

   Furthermore, we have

   $$|X'| = |X| - |X \setminus X'| \geq |X| - |V_i \setminus V_i'| \geq (\varepsilon - \beta)|V_i| \geq (\varepsilon - \beta)|V_i'| \geq \varepsilon'|V_i'|,$$

   where the last inequality holds if $\varepsilon' + \beta \leq \varepsilon$. Now, suppose that $(V_i', V_j')$ with $1 \leq i < j \leq k$ is $\varepsilon'$-regular. Let $X \subseteq V_i$ and $Y \subseteq V_j$ be such that $|X| \geq \varepsilon|V_i|$ and $|Y| \geq \varepsilon|V_j|$,

and let $X' = X \cap V_i'$ and $Y' = Y \cap V_j'$. By the $\varepsilon'$-regularity of $(V_i', V_j')$, we have $|d(X', Y') - d(V_i', V_j')| \le \varepsilon'$. We also have

$$\begin{aligned}
\frac{e(X,Y)}{|X|\cdot|Y|} - \frac{e(X',Y')}{|X'|\cdot|Y'|} &\le \frac{e(X',Y') + |X\setminus X'|\cdot|Y| + |X|\cdot|Y\setminus Y'|}{|X|\cdot|Y|} - \frac{e(X',Y')}{|X|\cdot|Y|} \\
&= \frac{|X\setminus X'|}{|X|} + \frac{|Y\setminus Y'|}{|Y|} \le \frac{2\beta}{\varepsilon}, \\
\frac{e(X',Y')}{|X'|\cdot|Y'|} - \frac{e(X,Y)}{|X|\cdot|Y|} &\le e(X',Y')\left(\frac{1}{|X'|\cdot|Y'|} - \frac{1}{|X|\cdot|Y|}\right) \le 1 - \frac{|X'|\cdot|Y'|}{|X|\cdot|Y|} \\
&\le 1 - \left(1 - \frac{\beta}{\varepsilon}\right)^2 \le \frac{2\beta}{\varepsilon}, \\
|d(X,Y) - d(X',Y')| &= \left|\frac{e(X,Y)}{|X|\cdot|Y|} - \frac{e(X',Y')}{|X'|\cdot|Y'|}\right| \le \frac{2\beta}{\varepsilon}, \\
|d(X,Y) - d(V_i,V_j)| &\le |d(X,Y) - d(X',Y')| + |d(X',Y') - d(V_i',V_j')| \\
&\quad + |d(V_i',V_j') - d(V_i,V_j)| \le \frac{4\beta}{\varepsilon} + \varepsilon' \le \varepsilon,
\end{aligned}$$

where the last inequality holds if $\varepsilon'$ and $\beta$ are small enough. To ensure the latter we should first choose $\varepsilon'$ small enough, and then choose $n_0$ large enough so that $\beta$ become small enough. Then, for the partition $V(G) = V_1 \cup \ldots \cup V_k$ to satisfy the conditions in the problem, we should choose $M = \max\{M(\varepsilon', n_0\}$. For graphs with fewer than $M$ vertices, the statement in the problem holds trivially.

Now, we deduce Szemerédi's regularity lemma from the statement in the problem. Let $G$ be a graph with at least $n_0$ vertices, with $n_0$ to be chosen later. Apply the statement in the problem with $\varepsilon' = \varepsilon/2$ to obtain a partition $V(G) = V_1' \cup \ldots \cup V_k'$. Let $V_0$ contain one vertex per each set $V_i'$ with $|V_i'| = |V_1'| + 1$, and let $V_i = V_i' \setminus V_0$ for $1 \le i \le k$. We have $|V_0| \le k \le \varepsilon n_0$ if $n_0 \ge 1/\varepsilon'$. Moreover, by the result in problem 2, $(V_i, V_j)$ is $\varepsilon$-regular whenever $(V_i', V_j')$ is $\varepsilon'$-regular. Therefore, the partition $V(G) = V_0 \cup V_1 \cup \ldots \cup V_k$ satisfies the conclusion of Szemerédi's regularity lemma. We only need to set $n_0 = 1/\varepsilon'$ and $M = \max\{M(\varepsilon'), n_0\}$, so that graphs with fewer than $n_0$ vertices satisfy the conclusion trivially.

4. *Prove that for every $\varepsilon > 0$ there is $n \in \mathbb{N}$ such that every set $A \subseteq \{1, \ldots, n\}^2$ with $|A| \ge \varepsilon n^2$ contains three points of the form $(x, y)$, $(x + a, y)$ and $(x, y + a)$ with $a \ne 0$.*

   Consider the horizontal lines, vertical lines, and lines with slope $-45°$ passing through at least one point of $\{1, \ldots, n\}^2$. Let $H$, $V$ and $S$ be the sets of those horizontal, vertical and skew lines, respectively. Thus $|H| = |V| = n$ and $|S| = 2n - 1$. Define the edges of a tripartite graph $(H, V, S)$ so that there is an edge between two vertices-lines if the intersection point of these lines belongs to $A$. Each triangle in $(H, V, S)$ corresponds to a triple of points of the form $(x, y)$, $(x + a, y)$ and $(x, y + a)$, all belonging to $A$, with possibly $a = 0$. Conversely, any triple of points $(x, y)$, $(x + a, y)$ and $(x, y + a)$ in $A$ gives rise to a triangle on vertices representing the horizontal line passing trough $(x, y)$ and $(x + a, y)$, the vertical line passing through $(x, y)$ and $(x, y + a)$, and the skew line passing through $(x + a, y)$ and $(x, y + a)$. The number of triangles corresponding to $a = 0$ is exactly $|A| \ge \varepsilon n^2$, and they are all edge-disjoint. Therefore, they cannot be all destroyed by removing fewer than $\varepsilon n^2$ edges. By the triangle removal lemma applied to $(H, V, S)$ with $\alpha = \varepsilon/16$, there are at least $\delta(4n - 1)^3$ triangles in $(H, V, S)$. If we choose $n$ big enough so that $\delta(4n - 1)^3 > n^2$, then we can find at least one triangle corresponding to a triple of points $(x, y)$, $(x + a, y)$ and $(x, y + a)$ in $A$ with $a \ne 0$.

5. *Let $p \in [0,1]$ and $\varepsilon > 0$. Prove that a random bipartite graph $(A, B)$ with $|A| = |B| = n$ and edge probability $p$ satisfies the property in problem 1 (and thus is $2\varepsilon$-regular) with probability approaching 1 as $n \to \infty$. Use the following Chernoff's bound:*
*For every $\varepsilon > 0$ there is $c_\varepsilon > 0$ such that if $X_1, \ldots, X_n$ are independent random variables with values in $\{0,1\}$ and $\mu = E(X_1 + \ldots + X_n)$, then*

$$P(|X_1 + \ldots + X_n - \mu| > \varepsilon\mu) < 2e^{-c_\varepsilon \mu}.$$

For any vertices $a \in A$ and $b \in B$, let $Z_{a,b}$ be an indicator random variable equal to 1 if $ab$ is an edge or 0 otherwise. Thus $E(Z_{a,b}) = p$. For any sets $X \subseteq A$ and $Y \subseteq B$ with $|X| \geq \varepsilon n$ and $|Y| \geq \varepsilon n$, let $Z_{X,Y}$ be a random variable equal to the number of edges between $X$ and $Y$. Thus,

$$Z_{X,Y} = \sum_{a \in X} \sum_{b \in Y} Z_{a,b}.$$

By the linearity of expectation, we have

$$E(Z_{X,Y}) = \sum_{a \in X} \sum_{b \in Y} E(Z_{a,b}) = p|X| \cdot |Y|.$$

By Chernoff's bound, we have

$$P\left(\left|\frac{Z_{X,Y}}{|X| \cdot |Y|} - p\right| > \varepsilon\right) \leq P\left(\left|\frac{Z_{X,Y}}{|X| \cdot |Y|} - p\right| > \varepsilon p\right) < 2e^{-c_\varepsilon p|X| \cdot |Y|} \leq 2e^{-c_\varepsilon \varepsilon^2 p n^2}.$$

Therefore, the probability that there are any $X \subseteq A$ and $Y \subseteq B$ with $|X| \geq \varepsilon n$, $|Y| \geq \varepsilon n$ and $|d(X,Y) - p| > \varepsilon$ is less than

$$2^{2n} \cdot 2e^{-c_\varepsilon \varepsilon^2 p n^2},$$

which approaches 0 as $n \to \infty$.

---