Szemerédi's Regularity Lemma

Lalla Mouatadid*

Abstract

Szemerédi's Regularity Lemma is an important result in extremal graph theory. Roughly speaking, the lemma states that every graph can be approximated by random graphs; that is, the vertex set of every graph can be split into equal size subsets such that the distribution of the edges between almost any two of these subsets is pseudorandom. The Regularity Lemma has already proved to be a powerful tool in graph theory and additive combinatorics. In this essay, we discuss the full proof to Szemerédi's Regularity Lemma, as presented in [2], and briefly discuss the co-NP-completeness of its decision problem. That is, given a graph $G(V, E)$ and a partition $\mathcal{P} = \{V_0, V_1, ..., V_k\}$ of V, checking whether $\mathcal{P}$ is an ϵ-regular partition of G is co-NP-complete [1].

1 Introduction

Story time! I was first exposed to the Regularity Lemma in Lluis' lecture, where he presented the proof to Roth's theorem about the existence of non trivial 3-term arithmetic progressions in certain subsets of integers. The proof involved the Triangle Removal Lemma, which follows from the Regularity Lemma, but we covered a neat(ter?) combinatorial proof. Here we give the proof of the Regularity Lemma, and in section 3, we discuss the co-NP-completeness of regularity testing [1].

Before presenting Szemerédi's Regularity Lemma, we first cover some necessary definitions. Let $G(V, E)$ be a graph and $A, B \subseteq V$ two disjoint sets. We use $e(A, B)$ to denote the number of edges going between A and B, and $d(A, B)$ to denote the *density* of the pair (A, B), where density is defined as follows:

$$d(A, B) = \frac{e(A, B)}{|A||B|} \tag{1}$$

Note that the density is always a positive value between 0 and 1. We call a pair (A, B) ϵ-regular, for a given $\epsilon > 0$, if for all $X \subseteq A, Y \subseteq B$ where $|X| \geq \epsilon|A|$ and $|Y| \geq \epsilon|B|$, we have:

$$|d(X, Y) - d(A, B)| \leq \epsilon \tag{2}$$

*Department of Computer Science, University of Toronto
Lalla@cs.toronto.edu

We call a partition $\mathcal{P} = \{V_0, V_1, V_2, ..., V_k\}$ of V ϵ-regular, if the following three conditions hold:

1. $|V_0| \leq \epsilon |V|$
2. $|V_1| = |V_2| = ... = |V_k|$
3. At most ϵk^2 pairs (V_i, V_j) with $1 \leq i < j \leq k$ are **not** ϵ-regular.

The set V_0 is sometimes called the *exceptional set*, and is used to "collect" the leftover vertices in order for the rest of the subsets to be of equal size.

The Regularity Lemma simply states that every large graph admits an ϵ-regular partition. More precisely:

Lemma 1 (The Regularity Lemma). *For every $\epsilon > 0$ and every integer $m \geq 1$, there exists an integer M such that every graph of order at least m admits an ϵ-regular partition $\mathcal{P} = \{V_0, V_1, ..., V_k\}$ with $m \leq k \leq M$.*

Clearly having singletons as partition classes form a trivial partition; thus the upper bound M ensures that the V_i's are large. The main idea of the proof is to start with a partition $\mathcal{P}$ of V and to keep refining the partition classes V_i into smaller sets until all the properties of the ϵ-regularity are satisfied.

We follow the proof presented in [2] which goes as such: We start by creating a random partition $\mathcal{P} = \{V_0, V_1, ..., V_k\}$ of V where the $V_{i>0}$ are all of equal size. We define a function q to measure the regularity of the partition $\mathcal{P}$ as a whole, but also $q(V_i, V_j)$ to measure the regularity of the pair (V_i, V_j) in particular. If this pair is irregular, then we refine $\mathcal{P}$ by splitting V_i and V_j into subsets that are more regular. That is, we refine $\mathcal{P}$ to create a more regular partition $\mathcal{P}'$ such that $q(\mathcal{P}) < q(\mathcal{P}')$. The refinement (splitting) process is repeated until we obtain a partition that satisfies the properties mentioned above. We will show that only a bounded number of refinement steps is needed in order to obtain a satisfying partition; but first we present the rest of the definitions.

For any two disjoint sets $A, B \subseteq V$, we have:

$$q(A, B) = \frac{e(A, B)^2}{n^2 |A||B|} \tag{3}$$

$$= d^2(A, B) \frac{|A||B|}{n^2} \tag{4}$$

Where $n = |V|$. If $\mathcal{A}, \mathcal{B}$ are partitions of A and B respectively, $q(\mathcal{A}, \mathcal{B})$ is defined as:

$$q(\mathcal{A}, \mathcal{B}) = \sum_{A' \in \mathcal{A}, B' \in \mathcal{B}} q(A', B') \tag{5}$$

And if $\mathcal{P} = \{V_1, ..., V_k\}$ is a partition of V that does not include the exceptional set V_0, then:

$$q(\mathcal{P}) = \sum_{i<j} q(V_i, V_j) \tag{6}$$

If $\mathcal{P}$ includes V_0, then V_0 is treated as a set of singletons, and $q(\mathcal{P}) = q(\tilde{\mathcal{P}})$, where $\tilde{\mathcal{P}}$ is:

$$\tilde{\mathcal{P}} = \{V_1, V_2, ..., V_k\} \cup \{\{v\} : v \in V_0\} \tag{7}$$

We will show that if we start with a partition $\mathcal{P}$ of V which does not satisfy the regularity conditions, we only need a bounded number of refinements (iterations) before we end up with an ϵ-regular partition. More particular, every new refined $\mathcal{P}'$ has now $q(\mathcal{P}) \leq q(\mathcal{P}')$. We later show (54) that for any partition $\mathcal{P}$ of V, $q(\mathcal{P}) \leq 1$; thus we are guaranteed to obtain a regular partition after a bounded number of refinements.

We will also need the following inequality. For any real numbers $s_1, ..., s_k > 0$ and $t_1, ..., t_k > 0$:

$$\sum \frac{t_i^2}{s_i} \geq \frac{(\sum t_i)^2}{\sum s_i} \tag{8}$$

Let $a_i = \sqrt{s_i}, b_i = \frac{t_i}{\sqrt{s_i}}$, then (8) follows from the Cauchy-Schwarz inequality $\sum a_i^2 \sum b_i^2 \geq (\sum a_i b_i)^2$.

The Proof

Claim 1. *Let $A, B \subseteq V$ be two disjoint sets. If $\mathcal{A} = \{A_1, A_2, ..., A_h\}$ and $\mathcal{B} = \{B_1, B_2, ..., B_{h'}\}$ are partitions of A and B respectively, then:*

$$q(\mathcal{A}, \mathcal{B}) \geq q(A, B) \tag{9}$$

Proof.

$$q(\mathcal{A}, \mathcal{B}) = \sum_{i,j} q(A_i, B_j) \tag{10}$$

$$= \sum_{i,j} \frac{e(A_i, B_i)^2}{n^2 |A_i||B_i|} \tag{11}$$

$$\geq \frac{(\sum\limits_{i,j} e(A_i, B_i))^2}{n^2 \sum\limits_{i,j} |A_i||B_j|} \quad \text{(using (8))} \tag{12}$$

$$= \frac{e(A, B)^2}{n^2 |A||B|} \tag{13}$$

$$= q(A, B) \tag{14}$$

□

Claim 2. *Let $\mathcal{P}$ be a partition of V. If $\mathcal{P}'$ is a refinement of $\mathcal{P}$ then:*

$$q(\mathcal{P}) \leq q(\mathcal{P}') \tag{15}$$

In other words, when we refine a partition we increase its "regularity index", i.e. the q value.

Proof. Let $\mathcal{P} = \bigcup_{i=1}^{k} V_i$ and $\mathcal{P}' = \bigcup_{i=1}^{k} \mathcal{V}_i$, where each $\mathcal{V}_i$ is the refinement of V_i in $\mathcal{P}$ (if any refinement took place). We thus have:

$$
\begin{aligned}
q(\mathcal{P}) &= \sum_{i<j} q(V_i, V_j) && (16)\\
&\leq \sum_{i<j} q(\mathcal{V}_i, \mathcal{V}_j) && (17)\\
&\leq q(\mathcal{P}') && (18)
\end{aligned}
$$

Where the first inequality follows from Claim 1, and the latter inequality follows from the fact that $q(\mathcal{P}') = \sum_{i<j} q(\mathcal{V}_i, \mathcal{V}_j) + \sum_{i} q(\mathcal{V}_i)$. $\square$

We next show that refining an irregular pair will only increase the q value by a little, i.e., the increase is bounded above by a constant.

Claim 3. *Let $A, B \subseteq V$ be two disjoint sets such that (A, B) is not ϵ-regular, for a given $\epsilon > 0$, then there exists partitions $\mathcal{A} = (A_1, A_2), \mathcal{B} = (B_1, B_2)$ of A and B respectively such that:*

$$q(\mathcal{A}, \mathcal{B}) \geq q(A, B) + \epsilon^4 \frac{|A||B|}{n^2} \tag{19}$$

Proof. Let (A, B) be an irregular pair, then there exists subsets $X \subseteq A, Y \subseteq B$ that violate (2). Let $A_1 = X, A_2 = A \backslash X, B_1 = Y, B_2 = B \backslash Y$, and let μ denote:

$$\mu = d(A_1, B_1) - d(A, B) \tag{20}$$

And notice that by definition of μ, we have:

$$e(A_1, B_1) = \frac{|A_1||B_1|e(A, B)}{|A||B|} + \mu|A_1||B_1| \tag{21}$$

Now consider the following:

$$
\begin{aligned}
n^2 q(\mathcal{A}, \mathcal{B}) &= \sum_{i,j} \frac{e(A_i, B_j)^2}{|A_i||B_j|} && (22)\\
&= \frac{e(A_1, B_1)^2}{|A_1||B_1|} + \sum_{i+j>2} \frac{e(A_i, B_j)^2}{|A_i||B_j|} && (23)\\
&\geq \frac{e(A_1, B_1)^2}{|A_1||B_1|} + \frac{\sum\limits_{i+j>2} e(A_i, B_j)^2}{\sum\limits_{i+j>2} |A_i||B_j|} && (24)\\
&= \left(\frac{e(A_1, B_1)^2}{|A_1||B_1|} + \frac{(e(A, B) - e(A_1, B_1))^2}{|A||B| - |A_1||B_1|} \right) && (25)\\
&= \left[\frac{1}{|A_1||B_1|} \left(\frac{|A_1||B_1| e(A, B)}{|A||B|} + \mu |A_1||B_1| \right)^2 \right] && (26)\\
&+ \left[\frac{1}{|A||B| - |A_1||B_1|} \left(\frac{|A||B| - |A_1||B_1|}{|A||B|} e(A, B) - \mu |A_1||B_1| \right)^2 \right] && (27)\\
&= \left[\frac{|A_1||B_1| e(A, B)^2}{(|A||B|)^2} + \frac{2e(A, B)\mu |A_1||B_1|}{|A||B|} + \mu^2 |A_1||B_1| \right] && (28)\\
&+ \left[\frac{e(A, B)^2 (|A||B| - |A_1||B_1|)}{(|A||B|)^2} + \frac{\mu^2 (|A_1||B_1|)^2}{|A||B| - |A_1||B_1|} - \frac{2e(A, B)\mu |A_1||B_1|}{|A||B|} \right] && (29)\\
&= \frac{e(A, B)^2}{|A||B|} + \mu^2 |A_1||B_1| + \frac{\mu^2 (|A_1||B_1|)^2}{|A||B| - |A_1||B_1|} && (30)\\
&\geq \frac{e(A, B)^2}{|A||B|} + \mu^2 |A_1||B_1| && (31)\\
&\geq n^2 q(A, B) + \epsilon^4 |A||B| && (32)
\end{aligned}
$$

Therefore

$$
q(\mathcal{A}, \mathcal{B}) \geq q(A, B) + \epsilon^4 \frac{|A||B|}{n^2} \tag{33}
$$

□

Thus partitioning an irregular pair (A, B) increases q slightly by an increment less than a constant. This is however for a single irregular pair. We next show that if a partition $\mathcal{P}$ is not ϵ-regular, then subpartitioning all the bad pairs in $\mathcal{P}$ will only increase q by a constant.

Claim 4. *Let $0 < \epsilon \leq \frac{1}{4}$, and let $\mathcal{P} = \{V_0, V_1, ..., V_k\}$ be a partition of V where V_0 is the exceptional set of size $|V_0| \leq \epsilon n$ and all other sets are of equal size $|V_1| = ... = |V_k| = c$. If $\mathcal{P}$ is not ϵ-regular, then there exists a partition $\mathcal{P}' = \{V_0', V_1', ..., V_{k'}'\}$ of V where*

$k \leq k' \leq k4^k$, V_0' is the exceptional set with size $|V_0'| \leq |V_0| + \frac{n}{2^k}$, all other sets $V_{i>0}'$ are of equal size and:

$$q(\mathcal{P}') \geq q(\mathcal{P}) + \frac{\epsilon^5}{2}. \tag{34}$$

Proof. For all $1 \leq i < j \leq k$, we define partitions $\mathcal{V}_{ij}, \mathcal{V}_{ji}$ for V_i and V_j respectively as follows: If (V_i, V_j) is ϵ-regular then $\mathcal{V}_{ij} = \{V_i\}, \mathcal{V}_{ji} = \{V_j\}$; otherwise, by Claim 3, we know there exists partitions $\mathcal{V}_{ij}$ of V_i, $\mathcal{V}_{ji}$ of V_j where $|\mathcal{V}_{ij}| = |\mathcal{V}_{ji}| = 2$ and:

$$q(\mathcal{V}_{ij}, \mathcal{V}_{ji}) \geq q(V_i, V_j) + \epsilon^4 \frac{|V_i||V_j|}{n^2} \tag{35}$$

$$= q(V_i, V_j) + \frac{\epsilon^4 c^2}{n^2} \tag{36}$$

Let $\mathcal{V}_i$ be a set of equivalence classes where two vertices $u, v \in V_i$ are equivalent if u and v belong to the same partition class $\mathcal{V}_{ij}$ for all $i \neq j$. Therefore $\mathcal{V}_i$ refines every partition of $\mathcal{V}_{ij}$. And since V_i can be partitioned at most $k-1$ times, we thus have at most 2^{k-1} equivalent classes in $\mathcal{V}_i$:

$$|\mathcal{V}_i| \leq 2^{k-1} \tag{37}$$

Let $\mathcal{V}$ be the partition of V defined as:

$$\mathcal{V} = \{V_0\} \cup \bigcup_{i=1}^{k} \mathcal{V}_i \tag{38}$$

where V_0 is the exceptional set. Thus $\mathcal{V}$ refines $\mathcal{P}$ and:

$$k \leq |\mathcal{V}| \leq k2^k. \tag{39}$$

We next compute $q(\mathcal{V})$, but first we define $\mathcal{V}_0$ to be the set of singletons $\mathcal{V}_0 = \{\{v\} : v \in V_0\}$. And if $\mathcal{P}$ is not ϵ-regular, then there exists more than ϵk^2 pairs (V_i, V_i) for which the partitions $\mathcal{V}_{ij}, \mathcal{V}_{ji}$ are not just $\{V_i\}, \{V_j\}$ respectively (i.e., not trivial). Therefore we get:

$$q(\mathcal{V}) = \sum_{1<j} q(\mathcal{V}_i, \mathcal{V}_j) + \sum_{1 \leq i} q(\mathcal{V}_0, \mathcal{V}_i) + \sum_{0 \leq i} q(\mathcal{V}_i) \tag{40}$$

$$\geq \sum_{1 \leq i < j} q(\mathcal{V}_{ij}, \mathcal{V}_{ji}) + \sum_{1 \leq i} q(\mathcal{V}_0, \{V_i\}) + q(\mathcal{V}_0) \tag{41}$$

$$\geq \sum_{1 \leq i < j} q(V_i, V_j) + \epsilon k^2 \frac{\epsilon^4 c^2}{n^2} + \sum_{1} q(\mathcal{V}_0, \{V_i\}) + q(\mathcal{V}_0) \tag{42}$$

$$= q(\mathcal{P}) + \epsilon^5 \left(\frac{kc}{n}\right)^2 \tag{43}$$

$$\geq q(\mathcal{P}) + \frac{\epsilon^5}{2} \tag{44}$$

Where (40) follows from the definition of the function $q(\mathcal{P})$ when the partition $\mathcal{P}$ includes the exceptional set; which is the case indeed for $\mathcal{V}$. The first inequality follows from Claim 1; the second inequality follows from using (36); and the last inequality follows from the fact that $|V_1| = ...|V_k| = c$ and $\epsilon \leq \frac{1}{4}$, thus $|V_0| \leq \epsilon n \leq \frac{1}{4}n \implies kc \geq \frac{3}{4}n \implies \left(\frac{kc}{n}\right)^2 \geq \frac{9}{16} \geq \frac{1}{2}$.

What remains to do now is clean up the sets of $\mathcal{V}$ in order to turn $\mathcal{V}$ into the desired partition $\mathcal{P}'$. This means refining the sets $\mathcal{V}_i \in \mathcal{V}$ into equal size, and moving all the remaining elements into $\mathcal{V}_0$ without making $\mathcal{V}_0$ too large.

Let $V_1', V_2', ..., V_{k'}'$ be the maximal collection of disjoint sets where $c' = \left\lfloor \frac{c}{4^k} \right\rfloor$ and every $V_i' \subset (V' \in \mathcal{V} \backslash \{\mathcal{V}_0\})$, and let $V_0' = V \backslash \bigcup V_i'$, then the partition $\mathcal{P}' = \{V_0', V_1', ..., V_{k'}'\}$ is indeed the partition of V that satisfies the condition of the Lemma. Clearly $\mathcal{P}'$ refines $\mathcal{C}$, thus by Claim 2 and (44):

$$q(\mathcal{P}') \geq q(\mathcal{C}) \geq q(\mathcal{P}) + \frac{\epsilon^5}{2} \tag{45}$$

By the choice of c', no more than 4^k V_i' sets lie inside the same V_j for some j, thus:

$$k \leq k' \leq k4^k \tag{46}$$

Moreover

$$|V_0'| \leq |V_0| + c'|\mathcal{V}| \tag{47}$$
$$\leq |V_0| + \frac{c}{4^k}k2^k \tag{48}$$
$$= |V_0| + \frac{ck}{2^k} \tag{49}$$
$$\leq |V_0| + \frac{n}{2^k} \tag{50}$$

Where the first inequality follows from the fact that all the $V_{i>0}'$'s use at most c' elements from each $V' \neq V_0$ in $\mathcal{V}$; and the second inequality follows from (39). $\square$

Now we are ready to present the proof of the Regularity Lemma, which simply follows from applying Claim 4 until we obtain an ϵ-regular partition that satisfies our conditions. Before presenting the final proof, we quickly show that for every partition $\mathcal{P}$ of V, $q(\mathcal{P}) \leq 1$.

Proof. Let $\mathcal{P}$ be a partition of V.

$$q(\mathcal{P}) = \sum_{i<j} q(V_i, V_j) \tag{51}$$
$$= \sum_{i<j} \frac{|V_i||V_j|}{n^2} d(V_i, V_j)^2 \tag{52}$$
$$\leq \frac{1}{n^2} \sum_{i<j} |V_i||V_j| \tag{53}$$
$$\leq 1 \tag{54}$$

□

Which means the number of times $q(\mathcal{P})$ will be increased is also bounded above by a constant; thus after after a bounded number of refinements, the final partition $\mathcal{P}'$ is ϵ-regular.

Proof of the Regularity Lemma. Let $\epsilon > 0, m \geq 1$ be given values; and without loss of generality, suppose $\epsilon \leq \frac{1}{4}$. Let $\mathcal{P}$ be a partition of V, and let s denote an upper bound on the number of iterations of Claim 4 to $\mathcal{P}$ before $\mathcal{P}$ becomes ϵ-regular. In particular, let $s = \frac{2}{\epsilon^5}$. In order to satisfy the ϵ-regularity, we need to (a) maintain the bound on the size of the exceptional set, and (b) choose an appropriate bound on M as well.

(a) First recall that the exceptional set must have size $|V_0| \leq \epsilon n$. Therefore before applying Claim 4 (one or multiple times), we need to make sure the current partition $\{V_0, V_1, ..., V_k\}$ satisfies $|V_1| = |V_2| = ... = |V_k|$ and $|V_0| \leq \epsilon n$. From (50) however we know that the size of $|V_0|$ increases by at most $\frac{n}{2^k}$ at every iteration. We thus want to pick a k large enough so that s increments of $\frac{n}{2^k}$ each to the size of the exceptional set do not exceed $\frac{1}{2}\epsilon n$. In addition, we want an n large enough so that for any initial $|V_0| < k$ we still maintain $|V_0| \leq \frac{1}{2}\epsilon n$. Note that if the initial partition has size k, $\{V_1, ..., V_k\}$, then we should allow up to k elements in the exceptional set in order to satisfy $|V_1| = ... = |V_k|$, thus $|V_0| < k$ initially.

Therefore pick $k \geq m$ to be large enough such that:

$$2^{k-1} \geq \frac{s}{\epsilon} \implies \frac{s}{2^k} \leq \frac{\epsilon}{2} \tag{55}$$
$$\implies k + \frac{s}{2^k} n \leq \epsilon n \quad \text{(for all } n \geq \tfrac{2k}{\epsilon}\text{, thus for } \tfrac{k}{n} \leq \tfrac{\epsilon}{2}\text{)} \tag{56}$$

(b) Recall that M is an upper bound on the number of non-exceptional sets in the final partition. After every refinement iteration, the number of non-exceptional sets in the partition can increase from x sets up to $x4^x$ sets; this follows from (46). Consider the function $f : x \longmapsto x4^x$, and chose M to be:

$$M = \max\{f^s(k), \frac{2k}{\epsilon}\} \tag{57}$$
$$= \max\{f^{\frac{2}{\epsilon^5}}(k), \frac{2k}{\epsilon}\} \tag{58}$$

The second term in the max is just to satisfy (56) for any $n \geq M$.

NOW! We are finally ready to construct an ϵ-regular partition $\{V_0, V_1, ..., V_k\}$ with $m \leq k \leq M$ for any graph $G(V, E)$ with $n \geq m$. If n is small, i.e., $n \leq M$, then the partition is trivial. Just partition V into $k = n$ singletons where each $V_{i>0}$ is a single vertex, and let $V_0 = \emptyset$. So suppose $n > M$. Chose V_0 to be the smallest subset of V such that k divides $|V \backslash V_0|$, and let $\{V_1, ..., V_k\}$ be any partition of $V \backslash V_0$ into k sets of equal size. By the choice of V_0, we clearly satisfy $|V_0| < k$, and thus $|V_0| \leq \epsilon n$ (follows from (56)). Let the initial partition be $\mathcal{P} = \{V_0, V_1, .., V_k\}$ and now it suffices to apply Claim 4 over and over again until we obtain an ϵ-regular partition of $G(V, E)$. We perform at most s iterations, and by (56) these iterations will always maintain an exceptional set of size of ϵn even when we perform s iterations. □

Et voilà!

2 Regularity Testing

Notice that the proof above has an algorithmic taste; you start with a partition and keep refining until you obtain the desired partition. We just need to check if a given partition $\mathcal{P}$ is ϵ-regular, and if not, perform another refinement. Surprisingly, testing if a given partition $\mathcal{P}$ is ϵ-regular is not easy! In fact, the problem is co-NP-complete:

Theorem 1. *[1] Given a graph $G(V, E)$, $\epsilon > 0, k \geq 1$ and $\mathcal{P} = \{V_0, V_1, ..., V_k\}$ a partition of G into $k+1$ sets; deciding whether $\mathcal{P}$ is ϵ-regular is co-NP-complete.*

This could be surprising, because in the same paper Alon et al. [1] gave a constructive version of the Regularity Lemma. Meaning, constructing $\mathcal{P}$ can be done in polynomial time, but testing for regularity is co-NP-complete!

In [1], they prove an even stronger theorem; namely that problem is co-NP-complete even for the case of $\mathcal{P} = \{\emptyset, A, B\}$, i.e. $k = 2$, and $\epsilon = \frac{1}{2}$. In other words:

Theorem 2. *[1] Given $\epsilon > 0$ and $G(V, E)$ a bipartite graph with vertex classes A, B such that $|A| = |B| = n$; determining if $\mathcal{P} = \{\emptyset, A, B\}$ is an ϵ-regular partition of G is co-NP-complete.*

Recall that a problem is in co-NP if its complement is in NP; it thus suffices to show that the complement of the problem is NP-complete. We refer the reader to [1] for the complete proof, but will sketch the main idea below, which relies on the following NP-complete problem:

Lemma 2. *Given a bipartite graph $G(A \cup B, E)$ with $|A| = |B| = n$ and $|E| = \frac{n^2}{2} - 1$, deciding whether G contains a subgraph isomorphic to $K_{\frac{n}{2}, \frac{n}{2}}$ is NP-complete.*

Lemma 2 states that G contains a $K_{\frac{n}{2}, \frac{n}{2}}$ if and only if $\{\emptyset, A, B\}$ is NOT ϵ-regular for

$\epsilon = \frac{1}{2}$. So assume G contains $H(A' \cup B', E') = K_{\frac{n}{2},\frac{n}{2}}$, then clearly:

$$|A'| = \frac{1}{2}|A| \tag{59}$$
$$\geq \epsilon |A| \tag{60}$$
$$|B'| = \frac{1}{2}|B| \tag{61}$$
$$\geq \epsilon |B| \tag{62}$$

but $|d(A', B') - d(A, B)| = \frac{1}{2} + \frac{1}{n^2} > \epsilon$, thus (A, B) is not ϵ-regular. For the other direction, suppose $\{\emptyset, A, B\}$ is not ϵ-regular and pick a bad pair (A', B'), $A' \subset A, B' \subset B$, that violates regularity. Thus:

$$|d(A', B') - d(A, B)| = |d(A', B') - \frac{1}{2} + \frac{1}{n^2}| > \epsilon \tag{63}$$
$$= \frac{1}{2} \tag{64}$$

but this is only possible when $d(A', B') = 1$. And by the choice of A', B', we have $|A'| \geq \frac{1}{2}n$ and $|B'| \geq \frac{1}{2}n$; therefore the subgraph of G induced by $A' \cup B'$ contains $K_{\frac{n}{2},\frac{n}{2}}$. And checking that A', B' are bad can be done in linear time, thereby completing the proof.

Acknowledgements. I wish to thank Lluís Vena for his helpful suggestions.

References

[1] Noga Alon, Richard A Duke, Hanno Lefmann, Vojtech Rodl, and Raphael Yuster. The algorithmic aspects of the regularity lemma. *Journal of Algorithms*, 16(1):80–109, 1994.

[2] Reinhard Diestel. Graph theory 2nd edition, 2000.