Problem 4: Prove the following color version of Szemerédi's Regularity Lemma: For every $\varepsilon > 0$ and $r \in \mathbb{N}$ there exists $M \in \mathbb{N}$ such that for every edge-colored graph G with r colors, there exists a partition of the vertex set $V(G) = V_1 \cup \ldots \cup V_k$ with $k \leq M$ such that all but at most εk^2 pairs (V_i, V_j) are ε-regular in each color class.

Solution (by Ferran Espuña): The notion of ε-regularity stated in the problem is slightly different from the one we saw in class. We will for now ignore this and address it later. To start, we will prove:

Proposition. *For every $\varepsilon > 0$ and $r \in \mathbb{N}$ there exists $M \in \mathbb{N}$ such that for every graph G on n vertices, edge-colored with r colors, there exists a partition of the vertex set $V(G) = V_1 \cup \ldots \cup V_k$ with $k \leq M$ such that, in each color class,*

$$\sum_{\substack{V_i, V_j \text{ not} \\ \varepsilon-\text{regular}}} |V_i||V_j| \leq \varepsilon n^2$$

That is, the partition is ε-regular in each color class.

Proof. The proof is analogous to the Szemerédi's Regularity Lemma proof we saw in class. We start with the trivial partition and keep refining it. For a partition π and a color c, let

$$d_2(\pi, c) = \sum_{V_i, V_j \in \pi} \frac{|V_i||V_j|}{n^2} d_c(V_i, V_j)^2$$

where the density $d_c(V_i, V_j)$ is the proportion of pairs (v_i, v_j) that are colored c. We know from class that this quantity does not decrease when refining the partition. Also, we know that if π is not ε-regular, in color class c, then there exists a refinement π' such that $|\pi'| \leq |\pi| 2^{2|\pi|}$ and

$$d_2(\pi', c) \geq d_2(\pi, c) + \varepsilon^5$$

Finally, we also know that this quantity is between 0 and 1. We can apply the same density increment argument as in class to the vector

$$(d_2(\pi, 1), \ldots, d_2(\pi, r))$$

In each step, if there is still a color class i such that π is not ε-regular, in the color class i, we can refine the partition to increase the ith coordinate of the vector by at least ε^5, while the other coordinates don't decrease. Therefore, this process must stop after at most $\frac{r}{\varepsilon^5}$ steps, at which point the partition is ε-regular in each color class and the number of parts is bounded by a tower of 2s of height at most $\frac{2r}{\varepsilon^5}$. □

The question now is how we can adapt this proof to the different notion of ε-regularity stated in the problem. The two would be equivalent if it were not for the fact that the sizes of the parts of a partition can be very different. We need to find a way to refine a partition at the end of each step such that the sizes of the parts at the end are more balanced. Given a partition with k parts, we will aim for about k^2 parts. Suppose that

$$(a-1)\frac{n}{k^2} < |V_i| \leq a\frac{n}{k^2}$$

Then, we will split V_i into a parts of size at most $\frac{n}{k^2}$. This will give us a partition with number of parts k' satisfying

$$k^2 = \sum_i \frac{|V_i|k^2}{n} \leq k' \leq \sum_i 1 + \frac{|V_i|k^2}{n} = k^2 + k$$

Note that the upper bound is just a function of k. This means that the argument above works exactly the same but with a higher bound on the number of parts. However, at each step, we may assume that any part V_i' satisfies

$$k'|V_i'| \leq \frac{n}{k^2}(k^2+k) = n\left(1+\frac{1}{k}\right) \leq n\left(1+\hat{\varepsilon}\right) \tag{1}$$

Where we can make $\hat{\varepsilon} := \frac{1}{k}$ as small as we'd like depending on ε by starting the process with an arbitrary partition with a large number of parts.

Now, suppose that, for some color class c, we have more than εk^2 pairs (V_i, V_j) that are not ε-regular. We can write this as

$$\sum_{\substack{V_i,V_j \text{ not} \\ \varepsilon-regular}} 1 \geq \varepsilon k^2 \tag{2}$$

We will do the same density increment argument as before, but now with $\varepsilon' = \frac{\varepsilon}{2}$. Suppose, by way of contradiction, that

$$\sum_{\substack{V_i,V_j \text{ not} \\ \varepsilon'-regular}} |V_i||V_j| < \varepsilon' n^2 = \frac{\varepsilon n^2}{2} \tag{3}$$

Multiplying (2) by $\left(\frac{n}{k}\right)^2$ and substracing (3) from it, we get

$$\sum_{\substack{V_i,V_j \text{ not} \\ \varepsilon'-regular}} \left(\frac{n}{k}\right)^2 - |V_i||V_j| > \frac{\varepsilon n^2}{2} \tag{4}$$

Note, however, that if we extend the sum to all pairs (V_i, V_j), (including the same part twice), we get $-\mathrm{Var}(X) < 0$, where X is the size of a part of the partition chosen uniformly at random. This means that there is a set of pairs $I \subset \pi^2$ satisfying

$$\sum_{(V_i,V_j)\in I} \left(\frac{n}{k}\right)^2 - |V_i||V_j| < -\frac{\varepsilon n^2}{2} \implies \sum_{(V_i,V_j)\in I} |V_i||V_j| > \frac{\varepsilon n^2}{2} + |I|\left(\frac{n}{k}\right)^2$$

Multiplying by k^2, and applying the bound on the relative size of the parts we obtained in (1), we get

$$n^2\left(\frac{\varepsilon k^2}{2}+|I|\right)<|I|\,n^2\,(1+\hat{\varepsilon}) \implies \frac{\varepsilon k^2}{2|I|}+1<1+\hat{\varepsilon} \implies \frac{\varepsilon k^2}{2|I|}<\hat{\varepsilon}$$

But obviously $|I|\leq k^2$ so $\frac{\varepsilon}{2}<\hat{\varepsilon}$, which is a contradiction if we choose $\hat{\varepsilon}$ small enough. This means that (3) cannot hold, so the partition is not ε'-regular in color class c, allowing us to refine the partition further. The rest of the argument follows as before, but the density increment is now $(\frac{\varepsilon}{2})^5$ and $|\pi'|\leq f\left(|\pi|2^{2|\pi|}\right)$, where $f(k)=k(k+1)$.