Econ 1033	**Problem set 3**
Prof. Ben Golub	Due: **Wed., Sep 23, 2020 11:50pm**

4. (40 points)

e. (10 points) Choose two distinct nodes uniformly at random. Prove that, if n is large enough, then with probability arbitrarily close to 1 the distance between them in G' is less than $n^{1/3}$. *Hint: chop up the graph into disjoint, connected, square subgraphs of size $k(n)$-by-$k(n)$, where $k(n)$ is wisely chosen so that you can prove any pair of subgraphs is very likely to be linked by a random edge.* (This will not quite prove the statement you want, but should give you an idea for how to proceed.)
Also, explain why your proof could be extended to show the same statement replacing $1/3$ with any positive number α.

Fix $a \in (0, 1)$. Partition G as suggested into connected, square regions of size $k(n)$-by-$k(n)$, where $k(n)$ is n^a rounded to the nearest integer. For simplicity, we will assume that this partitioning is exact (with no leftover nodes).[1]

Let H be a random graph whose nodes are regions and which has a link if and only if two regions share a random edge in G'. Note that the number of nodes in H is $m = n^2/n^{2a} = n^{2(1-a)}$. The expected number of random edges in G' touching any region v is

$$.02n^{2a}.$$

By a simple calculation we can see that the probability of any two random edges connecting the same two regions is very small; by removing a small constant fraction of edges, we can get rid of all "parallel" edges in H.[2] Passing to the subgraph of H made up only of non-duplicated edges we can assume we have a random graph where any pair of nodes is linked with probability

$$q = \frac{.01n^{2a}}{m} = \frac{.01n^{2a}}{n^2/n^{2a}} = .01n^{2(2a-1)}.$$

Writing n in terms of m, we can see that the probability that two nodes share an edge is $.01m^{\frac{2a-1}{1-a}}$. In other words, the expected degree of a node in H is $.01m^{\frac{a}{1-a}}$. We will show that:

Lemma 1. *Let $a = 0.3$. Fixing any $v, w \in H$, with high probability there is a path of length 3 or less between them.*

Assume we have this lemma. Fix any nodes $x, y \in G'$. Let v be the node in H corresponding to x's region and w be the node in H corresponding to y's region. Here is how we will get from x to y in G':

1. Walk from x to x', where x' is in region v and has an edge in G' to some other region $v_1 \in H$.

[1] It is only a bit of extra bookkeeping to show this assumption does not make any difference: essentially the same proof will work with regions of slightly uneven size.

[2] This can be shown via Chebyshev's inequality. Alternatively, you can explicitly write the probability of two regions having *at least* one edge between them and use simple calculus to show this is very well-approximated by the probability of having *exactly* one edge.

2. Within region v_1, walk to another node that has an edge in G' to a different region $v_2 \in H$.
3. Within region v_2, walk to another node that has an edge in G' to region w.
4. Within region w, arriving at some node, walk to y.

Since we know there is with high probability a path of length 3 in H from v to w, we know this is possible. Note that within-region parts of the above path cost at most $2n^{0.3}$ steps; getting from region to region costs 1 step. Thus we have a path of length at most $4 \cdot 2n^{0.3} + 3$ in total, which is less than $n^{1/3}$ for large n.

Proof of lemma. Now fix nodes v and w in H. By reasoning very similar to Problem 3(b), we can show that with arbitrarily high probability, both nodes have realized degrees between $\underline{d} = .015m^{\frac{a}{1-a}}$ and $\overline{d} = .005m^{\frac{a}{1-a}}$. Let $N(v)$ be the random neighborhood of v and $N(w)$ be the random neighborhood of W. We will show that with high probability, there is a path of length at most 3 from v to w. If $N(v)$ and $N(w)$ have a nonempty intersection, then there is such a path. So condition on the event where $N(v)$ and $N(w)$ are disjoint. We can write the number of edges between $N(v)$ and $N(w)$ as a sum of indicator variables

$$S = \sum_{v' \in N(v), w' \in N(w)} 1_{\{v', w'\} \text{ is an edge}}.$$

These are essentially independent and each has a probability $.01m^{\frac{2a-1}{1-a}}$, while the number of random variables here is at least $\underline{d}^2$. Multiplying these together you see that the expectation is growing in n. We can see by Chebyshev's inequality, very similarly to 2(b), that with high probability the sum is at least 1, which means there is a path of length at least 3. The only part here that was a little quick "essentially independent"—the edge indicators are actually very slightly *negatively* correlated.[3] But if you think about it, that doesn't mess up our argument: your bound on the variance of S gets even tighter, and you can use Chebyshev's inequality just the same way to give an upper bound on the probability that S deviates too far below its expectation. □

To show that there are paths of length n^{α} for any $\alpha > 0$, just use smaller regions. (This part of your answer can be quite informal.) We can now think H as a random graph with expected degrees m^{β} for some $\beta \in (0, 1)$ and edges that, again, are essentially uncorrelated, so the graph is very close to Erdos-Renyi. It suffices to say (a fact you can find on the Internet) that if the expected degree of nodes in H grows much faster than $\log(n)$, which is always the case in your constructions, then the diameter of H will, with high probability, be less $C \log(n)$ for some large C, and then you can use the same proof as above for showing the total path length is bounded.

[3] If region v' has a link to region w' then it is very slightly less likely to have a link to region w'' because one of its potential outgoing edges has been used up.