Notes: 02/14/2017

1 Problem: Average Degree of a Graph

Problem 1 (Avg. Degree of a Graph)**.** *Given a simple graph $G = (V, E)$ with $|V| = n$ and $\Omega(n)$ edges, we want to estimate the average degree of the graph.*

(*simple* just means no parallel edges or loops)

The graph G is given as an array of vertices, where each vertex lists both its degree and its neighbors in an array, thus allowing the following operations:

- **Degree query:** given vertex v, returns $\deg(v)$
- **Neighbor query:** given vertex v and index $j \in [\deg(v)]$, returns jth neighbor of v

In this lecture, we're only going to use the first one to get an algorithm for getting the average degree of G within a factor of (almost) 2 (with high probability). The second one will be used next time to get a better approximation factor (up to $1 + \epsilon$).

Generally, finding the average of a list of numbers is impossible in sublinear time, but the structure of simple graphs can be used. First, degree lists by themselves already have a special structure - for example, an n-vertex graph can have a degree list of $(1, \ldots, 1, n-1)$ but not $(0, \ldots, 0, n-1)$. So we can actually achieve (almost) a factor-of-2 approximation using only degree queries. Second, a graph permits us to look at the neighbors of vertices. The worst-case scenario is having a small number of very high-degree vertices and never seeing them - but by having high degrees, these vertices have lots of neighbors and we can therefore find them in a sense by checking random neighbors of the vertices we sample. This lets us achieve $(1 + \epsilon)$-approximation with the use of neighbor queries.

Question: *Why can't we just sample vertices, average the degrees, and apply Chernoff-Hoeffding?* Because Chernoff-Hoeffding requires samples from $[0, 1]$, and scaling down from $[0, n-1]$ (requiring precision to a factor of $\approx \epsilon/n$) would make the number of necessary samples too big. Basically, the problem is that there might be a small number of large-degree vertices (up to $n-1$) which make up a significant proportion of the total degree, but sampling won't find them.

2 Idea: Bucketing

Can divide vertices into *buckets* based on their degree. By definition, bucket i is defined as:

$$B_i := \{v : (1+\beta)^{i-1} < \deg(v) \leq (1+\beta)^i\}$$

where $\beta = \epsilon/c$ (where c is some constant, to be defined later). We define buckets for $i = 0, 1, \ldots, t$, where

$$t = O\Big(\frac{\log n}{\epsilon}\Big)$$

This ensures that all vertices with positive degree go in a bucket[1] (degree-0 vertices don't get a bucket, but we can ignore them since they don't contribute to the total degree, basically).

Defining $d_{B_i} := \sum_{v \in B_i} \deg(v)$, we get by definition that:

$$|B_i|(1+\beta)^{i-1} \leq d_{B_i} \leq |B_i|(1+\beta)^i$$

Since the buckets contain all (positive-degree) vertices and are mutually exclusive, if we define $d_{\text{tot}} := \sum_{v \in V} \deg(v) = \sum_{i=1}^{t} d_{B_i}$, we get

$$\sum_{i=1}^{t} |B_i|(1+\beta)^{i-1} \leq d_{\text{tot}} \leq \sum_{i=1}^{t} |B_i|(1+\beta)^i$$

The idea is now to estimate the size of each B_i by sampling.

3 The Algorithm

The basic algorithm (sampling is uniform unless otherwise specified) is:

1. Take a sample S of nodes where $|S| = \Theta(\sqrt{n} \cdot \text{polylog}(n) \cdot \text{poly}(1/\epsilon))$ (details in appendix)
2. Defining $S_i = S \cap B_i$, define as follows:

$$\rho_i = \begin{cases} \frac{|S_i|}{|S|}, & \text{if } |S_i| \geq \sqrt{\frac{\epsilon}{n}} \frac{|S|}{ct} \\ 0, & \text{otherwise} \end{cases}$$

3. Return $\sum_i \rho_i (1+\beta)^{i-1}$ as the estimated average degree.

The intuition behind this is, if sufficiently many samples are in S_i, then ρ_i is a good estimate of $\frac{|B_i|}{n}$ (which is what we wanted to estimate). Otherwise, we just ignore it (deliberately under-counting) and hope this doesn't screw it up too much. The reason for the magnitude of $|S|$ is to allow us to control the threshold for $|S_i|$ very precisely, e.g. the $\sqrt{n}$ cancels out the $1/\sqrt{n}$ term in the threshold. The exact properties of $|S|$ are given in the appendix.

One thing to keep in mind is, if we ignore the condition on the size of S_i for ρ_i to 'count', we get $\mathbb{E}\rho_i = \frac{|B_i|}{n}$ by linearity of expectations.

Finally, we get:

Theorem 1. *With probability at least $1 - 3t/n$, the algorithm's estimate of the average degree is not off by a factor of more than $2 + \epsilon$ from the actual average degree of G.*

[1] see appendix for exactly why

4 Analysis

We really need two pieces of analysis: (1) showing that the estimate is not too large, and (2) showing that it's not too small. This analysis is dependent on dividing buckets into 'big' and 'small': a bucket B_i is *big* if it satisfies:

$$|B_i| \geq \frac{2\sqrt{\epsilon n}}{ct}$$

and is *small* otherwise. We will also use the classification of a *very small* bucket for those that satisfy:

$$|B_i| < \frac{\sqrt{\epsilon n}}{2ct}$$

The idea is as follows: all "big" buckets B_i is very likely to be sampled sufficiently so that $\rho_i > 0$, and we are likely to get a reasonable estimate of their sizes. Thus, we will 'see' endpoints in "big" buckets reliably. Smallness is defined so that the number of edges which can have *both* endpoints in small buckets can only make up a small proportion of the total number of edges (in particular, this is why we have to assume $\Omega(n)$ edges).

The definition of "very small" buckets is meant to be used alongside the rule for when $\rho_i > 0$ – essentially, this entire analysis is built on Chernoff, which doesn't work well when the expected number of samples from a bucket is too small. In particular, if very small buckets represent a significant proportion of the total degree, there is a risk of oversampling them. Thus, any bucket which returns too few samples might throw the estimate off, and so we throw such samples away.

Proof: it's not too large. Consider what happens if the sampling is perfect, i.e. if $\rho_i = \frac{|B_i|}{n}$ for all i. Then we get

$$\sum_i \rho_i(1+\beta)^{i-1} = \sum_i \frac{|B_i|}{n}(1+\beta)^{i-1} = \frac{1}{n}\sum_i |B_i|(1+\beta)^{i-1} \leq \frac{d_{\text{tot}}}{n} = d_{\text{avg}}$$

because $(1+\beta)^{i-1} \leq \deg(v)$ for all $v \in B_i$.

We also know that $\mathbb{E}\rho_i \leq \frac{|B_i|}{n}$ ('$\leq$' because of the chance it's set to 0), so we consider the possibility that ρ_i is a little larger than expected, but not by much. Suppose for all i that $\rho_i \leq \frac{|B_i|}{n}(1+\gamma)$ where γ is some small constant. Then, we get (by the same logic) that

$$\sum_i \rho_i(1+\beta)^{i-1} \leq d_{\text{avg}}(1+\gamma)$$

We can use Chernoff-Hoeffding to show that this is in fact overwhelmingly likely to happen (for an appropriate value of γ, which luckily can be set to $\gamma = 1$) - see appendix for precise details (for the details, we need to know $|S|$ anyway). □

Proof: it's not too small. As before, a "big" bucket B_i has sufficiently many vertices that with high probability $|S_i|$ will pass the threshold and ρ_i will be nonzero; a "small" bucket B_j has a nontrivial chance of having $\rho_j = 0$.

Then we can note that edges come in three types, based on what kind of buckets their endpoints are in:

1. big-big
2. big-small
3. small-small

The first kind of edges are approximated well, the second kind are at least half-counted (they are counted from the big-bucket side), and we can't make any guarantees about the third kind. We get our approximation factor of 2 by showing that small-small edges have an upper limit on how common they can be.

This is achieved by noting that a small bucket has at most $\frac{2\sqrt{\epsilon n}}{ct}$ vertices, and there are at most t such buckets, so at most $\frac{2\sqrt{\epsilon n}}{c}$ vertices in total belong to small buckets. Thus, the number of small-small edges is at most

$$\left(\frac{2\sqrt{\epsilon n}}{c}\right)^2 = O(\epsilon n)$$

which, because there are $\Omega(n)$ edges in total, represents no more than an ϵ proportion of edges (controlled by adjusting the constant c). $\square$

A Number of Buckets

The buckets are designed to contain vertices with degrees in $((1+\beta)^{i-1}, (1+\beta)^i]$. So to cover all vertices, we just want t such that $(1+\beta)^t \geq n$. Since $\beta = \epsilon/c$, we get

$$(1+\epsilon/c)^t \geq n \iff t\log(1+\epsilon/c) \geq \log n$$

But taking the Taylor expansion of $\log(1+x)$ to one term (since ϵ is small) gives

$$\log(1+\epsilon/c) \approx \epsilon/c$$

so we get that we just need the smallest t such that

$$t \geq \frac{2c\log n}{\epsilon}, \text{ i.e. we only need } t = O\Big(\frac{\log n}{\epsilon}\Big)$$

(the '2' is there as a fudge factor to account for the '$\approx$', but that's just another constant).

B All the Gory Details

Here, we're just going to go over everything with all the gory details included. First, we need to define exactly what it means that "the number of edges is $\Omega(n)$". For now, we're just going to assume there are at least n edges; generalizing to "at least λn edges" (for some $\lambda > 0$) is fairly easy to do by adjusting the constant c appropriately.

B.1 Clarifying $|S|$

We need to decide on the number of vertices that our algorithm is going to sample. For now it will seem like we pulled this out of the air, but the following theorems should make it clear why we picked these values. We will set:

$$|S| = 6tc\log(n)\sqrt{n/\epsilon}$$

Note that since $t = O\Big(\frac{\log n}{e}\Big)$, we have $|S| = O(\sqrt{n}\log^2(n)(1/\epsilon)^{3/2})$. Note also that the threshold for $\rho_i > 0$ is then

$$\sqrt{\frac{\epsilon}{n}}\frac{|S|}{ct} = 6\log n$$

B.2 On Big and Small Buckets

We now consider what it means for a bucket to be "big". There are two perspectives here:

- a "big" bucket is any B_i for which $\rho_i > 0$ when the algorithm is run;
- a "big" bucket is any B_i containing sufficiently many vertices.

The first depends on the actual running of the algorithm (and is randomized to some degree) while the other is fixed, but not known to the algorithm.

We'll proceed with "big" as the second definition (it's more useful for showing a limit on the number of small-small edges) and we'll define buckets for which $\rho_i > 0$ as "found". We'll first show that with high probability, all big buckets are found. Formally:

Definition 1 (Big, Very Small, and Found Buckets)**.** *A bucket B_i is* big *if $|B_i| \geq \frac{2\sqrt{\epsilon n}}{ct}$; it is* very small *if $|B_i| < \frac{\sqrt{\epsilon n}}{2ct}$; finally, a bucket is* found *if $|S_i| \geq \sqrt{\frac{\epsilon}{n}} \frac{|S|}{ct}$ (i.e. if satisfies the condition so that $\rho_i > 0$).*

The intuition behind these definitions is as follows. The expected number of samples obtained from B_i can be computed as

$$|S_i| = \frac{|B_i|}{n}|S|$$

Therefore, if $|B_i| = \frac{\sqrt{\epsilon n}}{ct}$, we get that

$$|S_i| = \frac{|B_i|}{n}|S| = \sqrt{\frac{\epsilon}{n}} \frac{|S|}{ct}$$

i.e. the expected number of samples is exactly the number needed for bucket i to be found. If $|B_i|$ is double this (or more), the Chernoff Bound implies a very high probability that B_i is found; if $|B_i|$ is less than half this size, the Chernoff Bound implies a very low probability that B_i is found.

B.3 Important Lemmas

We now prove that big buckets are likely to be found, and very small buckets likely not to be.

Lemma 1 (Big Buckets)**.** *The probability that a big bucket B_i is* not *found is at most $1/n$*

Proof. We begin by expressing $|S_i|$ as a sum of independent Bernoulli random variables. Let $u_1, u_2, \ldots, u_{|S|}$ be the vertices sampled; they are uniform across V and independent. Then, we define:

$$X_{i,j} := \mathbf{1}\{u_j \in B_i\}$$

This happens with probability

$$\frac{|B_i|}{n} \geq \sqrt{\frac{\epsilon}{n}} \frac{2}{ct}$$

and clearly $|S_i| = \sum_{j=1}^{|S|} X_{i,j}$. Furthermore, by linearity of expectations,

$$\mathbb{E}|S_i| = \frac{|B_i|}{n}|S| \geq \sqrt{\frac{\epsilon}{n}} \frac{2|S|}{ct}$$

Note that this is double the threshold. Thus, we can apply the Chernoff Bound:

$$\mathbb{P}[\rho_i = 0] = \mathbb{P}\Big[|S_i| < \sqrt{\frac{\epsilon}{n}} \frac{|S|}{ct}\Big] \leq \mathbb{P}\big[|S_i| < (1 - 1/2)\mathbb{E}|S_i|\big] \leq \exp\Big(-\frac{1}{8}\mathbb{E}|S_i|\Big) \leq \exp\Big(-\sqrt{\frac{\epsilon}{n}} \frac{|S|}{4ct}\Big)$$

Setting $|S| = 6tc\log(n)\sqrt{n/\epsilon}$, we can get:

$$\mathbb{P}[\rho_i > 0] \leq \exp(-\log(n)) = 1/n$$

thus completing the proof. □

Lemma 2 (Very Small Buckets)**.** *The probability that a very small bucket B_i is not found is at most $1/n$*

Proof. Again, $|S_i|$ is a sum of independent Bernoulli random variables, as in the above proof. We also recall that

$$\mathbb{E}|S_i| = \frac{|B_i|}{n}|S| \leq \sqrt{\frac{\epsilon}{n}}\frac{|S|}{2ct}$$

i.e. half the threshold. Thus, we can apply the Chernoff Bound:

$$\mathbb{P}[\rho_i > 0] = \mathbb{P}\Big[|S_i| \geq \sqrt{\frac{\epsilon}{n}}\frac{|S|}{ct}\Big] \leq \mathbb{P}\big[|S_i| \geq (1+1)\mathbb{E}|S_i|\big] \leq \exp\Big(-\frac{1}{3}\mathbb{E}|S_i|\Big) \leq \exp\Big(-\sqrt{\frac{\epsilon}{n}}\frac{|S|}{6ct}\Big)$$

Setting $|S| = 6tc\log(n)\sqrt{n/\epsilon}$, we get:

$$\mathbb{P}[\rho_i > 0] \leq \exp(-\log(n)) = 1/n$$

thus completing the proof. □

These two lemmas are used to show the following key corollary:

Corollary 1. *With probability at least $1-t/n$ (which for large n is close to 1 since $t = O(\log n)$), all big buckets are found and all very small buckets are not found.*

Proof. This follows from the previous lemmas and a simple application of the Union Bound. In particular, there are at most t big or very small buckets combined. Each big bucket has at most a $1/n$ probability of not being found, and each very small bucket has at most a $1/n$ probability of being found. Thus, by the Union Bound, the probability that *any* big bucket is not found, or any very small bucket is found, is at most $1 - t/n$. □

Another reason behind the definition of a "big" bucket was that it imposes a strict limit on how many edges can exist between vertices of "small" buckets. This is important because such "small-small" edges might be basically impossible to find.

Lemma 3 (Number of Small-Small Edges)**.** *There are $O(\epsilon n)$ small-small edges.*

Proof. A 'small' bucket has at most $\frac{2\sqrt{\epsilon n}}{ct}$ vertices, and there are at most t small buckets. So at most $\frac{2\sqrt{\epsilon n}}{c}$ vertices are in small buckets, so there are at most

$$\Big(\frac{2\sqrt{\epsilon n}}{c}\Big)^2 = \frac{4}{c^2}\epsilon n$$

small-small edges. □

If there are at least λn edges in total (as we assumed), we can let $c = 4\lambda^{-1/2}$ to get at most $(\lambda/4)\epsilon n$ small-small edges.

B.4 Putting Everything Together

So now we need to put everything together. What are the ways that the algorithm could fail?

1. not all big buckets are found or some very small buckets are not found;
2. the estimates are significantly overstated;
3. the estimates are significantly understated.

Note that even if one of these events happens, it doesn't necessarily mean the output was wrong - but we will count it as 'failure' anyway. The important thing is that if *none* of these happen, the output will be approximately correct, so this is a lower bound on the probability of success.

Definition 2. *Let d_{tot} and d_{avg} be the total and average degree of G, respectively, and let $\hat{d}_{\text{tot}}$ and $\hat{d}_{\text{avg}}$ be the estimated total and average degree of G (i.e. the output of the algorithm, which is a random variable).*

By Corollary 1, we know that the probability that error (1) happens is at most t/n. Keeping that in mind, *we proceed in the assumption that this didn't happen.*

We start with the assertion that the number of elements in any given bucket is likely to be significantly too large.

Theorem 2. *$\hat{d}_{\text{avg}} \leq 2d_{\text{avg}}$ with probability at least $1 - t/n$.*

Proof. First, we note that by linearity of expectation, $\mathbb{E}\rho_i \leq \frac{|B_i|}{n}$ (it would be equal, but the rule of setting $\rho_i = 0$ if $|S_i|$ doesn't hit the threshold brings the expected value down a little). This gives us:

$$\mathbb{E}\hat{d}_{\text{avg}} = \sum_i \big(\mathbb{E}\rho_i\big)(1+\beta)^{i-1} \leq \sum_i \frac{|B_i|}{n}(1+\beta)^{i-1} = \frac{1}{n}\sum_i |B_i|(1+\beta)^{i-1} \leq \frac{d_{\text{tot}}}{n} = d_{\text{avg}}$$

But now we represent $|S_i|$ as the sum of $|S|$ independent random variables $X_{i,j}$, where $X_{i,j}$ is 1 if the jth sampled vertex is in B_i, and 0 otherwise. We can then apply the Chernoff Bound:

$$\mathbb{P}\big[|S_i| > (1+1)\mathbb{E}|S_i|\big] \leq \exp\Big(-\frac{1}{3}\mathbb{E}|S_i|\Big) = \exp\Big(-\frac{|B_i|}{3n}|S|\Big)$$

Plugging in $|S| = 6tc\log(n)\sqrt{n/\epsilon}$, we get an upper bound of:

$$\exp\Big(-\frac{|B_i|}{3n}6tc\log(n)\sqrt{n/\epsilon}\Big) \leq \exp\Big(-2|B_i|\,n^{-1/2}\epsilon^{-1/2}tc\log(n)\Big)$$

Luckily, because we assume that all very small buckets are not found, we can ignore them and assume B_i is not very small, i.e. $|B_i| \geq \frac{\sqrt{\epsilon n}}{2ct}$. Thus, cancelling everything out, we get

$$\mathbb{P}\big[|S_i| > 2\mathbb{E}|S_i|\big] \leq \exp(-\log n) = 1/n$$

Union-bounding over the at most t different buckets gives that $|S_i| < 2\mathbb{E}|S_i|$ for *all* (not-very-small) buckets with probability at least $1 - t/n$.

This finishes the proof since this implies that for all i,

$$\rho_i \le \frac{|S_i|}{|S|} \le \frac{2\mathbb{E}|S_i|}{|S|} = \frac{2|B_i|}{n}$$

so we can finally conclude that

$$\hat{d}_{\text{avg}} = \sum_i \rho_i (1+\beta)^{i-1} \le \sum_i \frac{2|B_i|}{n}(1+\beta)^{i-1} = \frac{2}{n}\sum_i |B_i|(1+\beta)^{i-1} \le 2\frac{d_{\text{tot}}}{n} = 2d_{\text{avg}}$$

with probability at least $1 - t/n$. □

Remark: this proof explains why we set $\rho_i = 0$ when insufficiently many samples are taken - this decisively prevents error from oversampling of very small buckets.

We now handle the probability of *undersampling* big buckets:

Theorem 3. $\hat{d}_{\text{avg}} \le \frac{1}{2+\epsilon} d_{\text{avg}}$ *with probability at least* $1 - t/n$.

Proof. This proof only depends on the sampling of big buckets; the worst-case scenario for small buckets B_i is that $\rho_i = 0$, and so we ignore them.

We now consider a big bucket B_i, and we want to know the probability that our estimate is too small by a factor of $1 + \epsilon^*$ (where we define ϵ^* later), i.e. $\rho_i < \frac{|B_i|}{(1+\epsilon^*)n}$. We assume that ϵ^* is sufficiently small that $\frac{1}{1+\epsilon^*} \le 1 - 2\epsilon^*$.

We claim that this probability is at most $1/n$. This proof is completely analogous to the proof that a big bucket is found with probability at least $1 - 1/n$, but we'll give it here anyway. First, we know that for a big bucket B_i,

$$\rho_i < \frac{|B_i|}{(1+\epsilon^*)n} \iff |S_i| < \frac{|B_i|}{(1+\epsilon^*)n}|S| = \frac{1}{1+\epsilon^*}\mathbb{E}|S_i|$$

We use this to note that

$$\rho_i < \frac{|B_i|}{(1+\epsilon^*)n} \implies |S_i| < \frac{1}{1+\epsilon^*}\mathbb{E}|S_i| \implies |S_i| < (1-2\epsilon^*)\mathbb{E}|S_i|$$

so we get that

$$\mathbb{P}\Big[\rho_i < \frac{|B_i|}{(1+\epsilon^*)n}\Big] \le \mathbb{P}\Big[|S_i| < (1-2\epsilon^*)\mathbb{E}|S_i|\Big]$$

Then we simply plug into the Chernoff Bound to get:

$$\mathbb{P}\Big[|S_i| < (1-2\epsilon^*)\mathbb{E}|S_i|\Big] \le \exp\Big(-2(\epsilon^*)^2\mathbb{E}|S_i|\Big) \le \exp\Big(-(\epsilon^*)^2\sqrt{\frac{\epsilon}{n}}\frac{4|S|}{ct}\Big)$$

and, plugging in $|S| = 6ct\log(n)\sqrt{n/\epsilon}$ gives

$$\mathbb{P}\Big[\rho_i < \frac{|B_i|}{(1+\epsilon^*)n}\Big] \le \exp(-(\epsilon^*)^2 \log n)$$

We can then simply apply the Union Bound to get the following:

$$\mathbb{P}[\rho_i \geq |B_i|/(2n), \text{ for all big buckets } B_i] \geq 1 - t/n$$

We now employ the trick of dividing the edges into the three categories: (i) big-big, (ii) big-small, (iii) small-small. Let $E_{\text{b-b}}$, $E_{\text{b-s}}$, $E_{\text{s-s}}$ be the sets of these edges (respectively). Suppose that for all big B_i, we have $\rho_i \geq |B_i|/(2n)$, and let

$$I := \{i : B_i \text{ is big}\}$$

Then we get

$$\hat{d}_{\text{avg}} = \sum_i \rho_i (1+\beta)^{i-1} \geq \sum_{i \in I} \rho_i (1+\beta)^{i-1} = \frac{1}{1+\beta} \sum_{i \in I} \rho_i (1+\beta)^i$$
$$\geq \frac{1}{1+\beta} \sum_{i \in I} \frac{|B_i|}{(1+\epsilon^*)n} (1+\beta)^i = \frac{1}{n(1+\epsilon^*)(1+\beta)} \sum_{i \in I} |B_i|(1+\beta)^i$$

But we note that since any vertex in bucket B_i has degree $\leq (1+\beta)^i$, we can get:

$$\sum_{i \in I} |B_i|(1+\beta)^i \geq \sum_{v \in B_i \,:\, i \in I} \deg(v) = 2|E_{\text{b-b}}| + |E_{\text{b-s}}| = d_{\text{tot}} - |E_{\text{b-s}}| - 2|E_{\text{s-s}}|$$

(since by definition $d_{\text{tot}} = 2|E| = 2|E_{\text{b-b}}| + 2|E_{\text{b-s}}| + 2|E_{\text{s-s}}|$). Furthermore, we note that:

$$\frac{d_{\text{tot}} - |E_{\text{b-s}}| - 2|E_{\text{s-s}}|}{d_{\text{tot}}} = \frac{d_{\text{tot}} - |E_{\text{b-s}}|}{d_{\text{tot}}} - \frac{2|E_{\text{s-s}}|}{d_{\text{tot}}}$$

Taking this one term at a time, we get:

$$\frac{d_{\text{tot}} - |E_{\text{b-s}}|}{d_{\text{tot}}} \geq \frac{2|E_{\text{b-b}}| + |E_{\text{b-s}}|}{2|E_{\text{b-b}}| + 2|E_{\text{b-s}}|} \geq \frac{1}{2}$$

(because we first subtract $2|E_{\text{s-s}}|$ from both top and bottom, and then subtract $2|E_{\text{b-b}}|$ from both top and bottom). Furthermore, since there are at least λn edges, $d_{\text{tot}} \geq 2\lambda n$ and so by Lemma 3:

$$\frac{2|E_{\text{s-s}}|}{d_{\text{tot}}} \leq \frac{|E_{\text{s-s}}|}{\lambda n} \leq \frac{(4/c^2)\epsilon n}{\lambda n} = \frac{4\epsilon}{c^2 \lambda}$$

Hence, we can put all of this together and get

$$\frac{2n(1+\beta)\hat{d}_{\text{avg}}}{d_{\text{tot}}} \geq \frac{d_{\text{tot}} - |E_{\text{b-s}}| - 2|E_{\text{s-s}}|}{d_{\text{tot}}} \geq \frac{1}{2} - \frac{4\epsilon}{c^2\lambda}$$

By definition, $d_{\text{avg}} = d_{\text{tot}}/n$ so we can simplify to

$$(1+\epsilon^*)(1+\beta)\frac{\hat{d}_{\text{avg}}}{d_{\text{avg}}} \geq \frac{1}{2} - \frac{4\epsilon}{c^2\lambda} \implies \frac{\hat{d}_{\text{avg}}}{d_{\text{avg}}} \geq \frac{\frac{1}{2} - \frac{4\epsilon}{c^2\lambda}}{(1+\epsilon^*)(1+\beta)}$$

□

Finally, we can put all of this together to get the final theorem:

Theorem 4. *The algorithm successfully approximates d_{avg} to a factor of $2+\epsilon$ with probability at least $1 - 3t/n = 1 - O\left(\frac{\log n}{\epsilon n}\right)$.*