Supplementary Material

A Proof of Theorem 2.1

We will need to Hoeffding-Chernoff bound for negative dependence:

Theorem A.1. *Assume $X_i \in [0,1]$ are negatively dependent variables and define $X = \sum_{i=1}^{n} X_i$ then*

$$P\left(|X - \mathbb{E}[X]| \geq \epsilon \mathbb{E}[X]\right) \leq 2\exp\left(-\frac{\epsilon^2 \mathbb{E}[X]}{3}\right)$$

See [7] 1.6 and 3.1 for details.

First it is important to note a change in notation from [19] in order to be consistent with notation used in [14]. A ϵ-spectral approximation in our paper is weaker then a $(1+\epsilon)$-spectral approximation in [19].

We will now go over the main changes needed to prove Theorem 6.1 in [19] (disregarding S.2) with negatively dependent sampling of edges and weights $w_{ij} \in [0,1]$.

The proof of Claim 6.5 is quite straightforward. The claim of Lemma 6.6 needs to be changed to $\mathbb{E}[\Delta^k_{r,t}\Delta^l_{t,r}] \leq \frac{w_{r,t}}{\gamma^{k+l-1}d_r}$ instead of $\mathbb{E}[\Delta^k_{r,t}\Delta^l_{t,r}] \leq \frac{1}{\gamma^{k+l-1}d_r}$. The changes to the proof are again straightforward (remembering $w_{ij} \in [0,1]$).

The main change is to Lemma 6.4. Using the modified Lemma 6.6 and substituting negative dependents for independence one can prove

$$\sum_{\sigma\ valid\ for\ T,\tau} \prod_{s\in T} \left[\Delta_{v_{s-1},v_s} \prod_{i:\tau(i)=s} \Delta_{v_{i-1},v_i}\right] \leq \frac{1}{\gamma^{k-|T|}} \sum_{\sigma\ valid\ for\ T,\tau} \prod_{s\in T} \frac{w_{v_{s-1},v_s}}{d_{v_{s-1}}}.$$

instead of equation 10 in the paper. The last change is to pick $\sigma(s)$ proportional to w_{v_{s-1},v_s} instead of uniformly to prove that

$$\sum_{\sigma\ valid\ for\ T,\tau} \prod_{s\in T} \frac{w_{v_{s-1},v_s}}{d_{v_{s-1}}} \leq 1$$

instead of equation 11. From there on all changes are straightforward.

B Proof of Theorem 3.1

Let $\tilde{L} = \tilde{L}^{in} + \tilde{L}^{out}$. Let P be the zeros eigenspace of $\tilde{L}^{in}$, which is the same as the zero eigenspace of L^{in}, if all the $\tilde{W}^i$ are connected. Let Q be the space spanned by the first k eigenvectors of $\tilde{L}$. According to the Sin-Theta theorem [6], $||\sin(\Theta(P,Q))|| \leq \frac{||\tilde{L}^{out}||}{\mu_2^{in}}$ where $||\cdot||$ is the spectral norm of $\tilde{L}^{out}$ and μ_2^{in} is the second smallest unnormalized eigenvalue of $\tilde{L}^{in}$. To prove the theorem we will show that $\mu_2^{in} = \Omega(n^\alpha)$ and that $||\tilde{L}^{out}|| = \mathcal{O}(n^\beta + n^\gamma)$.

The first claim is through using the first two assumptions and the following lemma

Lemma B.1. *Let λ_2 and μ_2 be the second smallest normalized and unnormalized eigenvalues of L, and $d = \min_i D_{ii}$ then $\mu_2 \geq \lambda_2 \cdot d$.*

Proof. *From the min-max theorem we have*

$$\mu_2 = \min_{U:\,\dim(U)=2}\left\{\max_{x\in U\setminus\{0\}} \frac{x^T L x}{||x||^2}\right\}$$

$$\lambda_2 = \min_{U:\,\dim(U)=2}\left\{\max_{x\in U\setminus\{0\}} \frac{x^T L x}{x^T D x}\right\}$$

and the lemma follows from the fact that $x^T D x \geq d||x||^2$. □

Let m_1 be the number of edges needed in order to have an ϵ-spectral approximation of each inner-cluster matrix W^i for $\epsilon = 3/4$ with probability $\delta/2$, then by theorem 2.1 we have that $m_1 = \tilde{\mathcal{O}}(n^{2-\alpha})$. Using this fact, the first two assumptions and lemma B.1, it is easy to see that $\mu_2^{in} = \Omega(n^\alpha)$.

We now need to show that $||\tilde{L}^{out}|| = \mathcal{O}(n^\beta + n^\gamma)$. The main tool would be the matrix Chernoff inequlity for sampling matrixes without replacements.

Theorem B.1. *Consider a finite sequence of Hermitian matrices* $\mathbf{X_1}, ..., \mathbf{X_k}$ *sampled uniformly without replacements from a finite set of matrices of dimension* n. *Assume that*

$$\mathbf{X_k} \succeq \mathbf{0} \qquad ||\mathbf{X_i}|| \leq \mathbf{R}.$$

Define $\mathbf{Y} = \sum_{\mathbf{i=1}}^{\mathbf{k}} \mathbf{X_i}$ *then*

$$P(||\mathbf{Y}|| \geq (\mathbf{1}+\epsilon)||\mathbb{E}[\mathbf{Y}]||) \leq n \cdot \left(\frac{e^\epsilon}{(1+\epsilon)^{1+\epsilon}}\right)^{||\mathbb{E}[Y]||/R}$$

Proof. *This is an adaptation of theorem 5.1.1 from [20] replacing the independence requirement to sampling without replacements. In order to adapt the proof we notice that the only place where independence is used in in lemma 3.5.1 (subadditivity of the matrix cumulant generating functions) where we need to prove that*

$$\forall\theta\in\mathbb{R} \qquad \mathbb{E}\left[Tr\left(\exp\left(\sum\theta\mathbf{X_i}\right)\right)\right] \leq Tr\left(\exp\left(\sum\log\mathbb{E}e^{\theta\mathbf{X_i}}\right)\right) \tag{2}$$

Using the result of [12], if $\mathbf{X_i}$ *are sampled uniformly at random without replacements for a finite set, and* $\mathbf{Y_i}$ *are sampled with the same probability with replacements then*

$$\mathbb{E}\left[Tr\left(\exp\left(\sum\theta\mathbf{X_i}\right)\right)\right] \leq \mathbb{E}\left[Tr\left(\exp\left(\sum\theta\mathbf{Y_i}\right)\right)\right] \tag{3}$$

so we can conclude that

$$\mathbb{E}\left[Tr\left(\exp\left(\sum\theta\mathbf{X_i}\right)\right)\right] \leq \mathbb{E}\left[Tr\left(\exp\left(\sum\theta\mathbf{Y_i}\right)\right)\right] \leq$$
$$\leq Tr\left(\exp\left(\sum\log\mathbb{E}e^{\theta\mathbf{Y_i}}\right)\right) = Tr\left(\exp\left(\sum\log\mathbb{E}e^{\theta\mathbf{X_i}}\right)\right).$$

where the second inequality is from 2 as $\mathbf{Y_i}$ *are independent.*

We define for each edge e connecting nodes in different clusters the matrix $\mathbf{X_e}$ that is equal to zero with probability $1-p$ and is equal to $\frac{1}{p}L_e$ with probability p, where L_e is the Laplacian of a single edge graph with weight w_e. Then $\tilde{L}^{out} = \sum_{e\in S_{out}} \mathbf{X_e}$, $\mathbb{E}[\tilde{L}^{out}] = L^{out}$, $\mathbf{X_e} \succeq \mathbf{0}$ and $||\mathbf{X_e}|| \leq 1/p$.

If we use the matrix Chernoff inequality with $1+\epsilon = 2e\cdot n^{\gamma-\beta}$ then

$$P(||\tilde{L}^{out}|| \geq 2en^\gamma) \leq n\left(\frac{1}{2n^{\gamma-\beta}}\right)^{2en^\gamma p}$$

So if $p = m/\binom{n}{2} = \mathcal{O}\left(\frac{log(n)}{n^\gamma}\right)$ we get that $P(||\tilde{L}^{out}|| \geq 2en^\gamma) < \delta/2$ for large enough n.

C Cut Approximation

We will start by proving an analog of theorem 2.1 in the paper. We will use the following lemma from [13]:

Lemma C.1. *Let G be an undirected graph with n vertices and minimal cut $c > 0$. For all $\alpha \geq 1$ the number of cuts with weight smaller of equal to αc is less then $n^{2\alpha}$.*

The lemma is proven in [13] for graphs with integer weights, but the extension to any positive weights is trivial by scaling and rounding. We can now state and prove the theorem guaranteeing good cut approximations.

Theorem C.1. *Let G be a graph with weights $w_{ij} \in [0,1]$, with minimal cut $c > 0$, and $\tilde{G}$ its approximation after sampling m edges uniformly. If $m \geq \binom{n}{2}\frac{3(2\ln(n)+\ln(\frac{1}{\delta})+k)}{\epsilon^2 c}$ where $k = \ln(2+4\ln(n))$, then the probability that $\tilde{G}$ is not an ϵ-cut approximation is smaller then δ.*

Proof. *This is an adaptation of the proof in [14] - consider a cut with weight αc. Let $p = m/\binom{n}{2}$ the probability to sample a single edge. Let $Y_e = X_e \cdot w_e$ where X_e is an indicator whether edge e on the cut was sampled and w_e its weight. Define Y the sum of Y_e on all the edges along the cut, then by the fact that edges are negatively dependent and theorem A.1, the probability that the cut is not an ϵ approximation is smaller then*

$$2\exp\left(-\frac{\epsilon^2\mathbb{E}[Y]}{3}\right) = 2\exp\left(-\frac{\epsilon^2\alpha cp}{3}\right) \leq$$
$$\leq 2\exp\left(-\left(\ln\left(\frac{1}{\delta}\right)+k\right)\alpha\right)\cdot n^{-2\alpha}$$

Define $P(\alpha) = 2\exp\left(-\left(\ln\left(\frac{1}{\delta}\right)+k\right)\alpha\right)\cdot n^{-2\alpha}$ and let $f(\alpha)$ the number of cuts with value αc in the original graph. By the union bound the probability that some cut is not an ϵ approximation is less then $\sum_{\alpha\geq 1} f(\alpha)P(\alpha)$ (notice that this sum is well defined since $f(\alpha)$ is non zero only in a finite number of α values). Defining $F(\alpha) = \sum_{\alpha\geq x} f(x)$ then by the previous lemma $F(\alpha) \leq n^{2\alpha}$. Let g be any measure on $[1,\infty)$ such that $G(\alpha) = \int_1^\alpha dg \leq n^{2\alpha}$, then the integral $\int_1^\infty P(x)dg$ is maximized when $G(\alpha) = n^{2\alpha}$. This is due to the fact that P is a monotonically decreasing function, so if the inequality is not tight at some point x_1 we could increase the value by picking $\tilde{g}(x) = g(x) + \tilde{\epsilon}\delta(x-x_1) - \tilde{\epsilon}\delta(x-x_2)$ for some appropriate $x_2 > x_1$ and $\tilde{\epsilon}$ (where δ is the Dirac delta function) . From this we can conclude that the probability of some cut not being an ϵ-approximation is bounded by

$$n^2P(1) + \int_1^\infty P(\alpha)\frac{dn^{2\alpha}}{d\alpha}d\alpha = 2\delta e^{-k} + \frac{4\delta\ln(n)e^{-k}}{\ln(\frac{1}{\delta})+k}$$
$$\leq \delta\left(2+4\ln(n)\right)\exp(-k) = \delta$$

□

A drawback is that the theorem gives a bound that depends on the minimal cut, which we do not know, and unlike the situation in [14] we cannot approximate it using the full graph. We can prove a bound that uses only known data about the graph. The following theorem shows we can lower bound c.

Lemma C.2. *Let G be a graph with weights $w_{ij} \in [0,1]$, with minimal cut c, and $\tilde{G}$ its approximation after sampling m edges with minimal cut $\tilde{c} > 0$. Define $p = m/\binom{n}{2}$ the probability to sample a single edge. Also define $l = \frac{3\ln(\frac{1}{\delta})}{4}$ and $\beta = \sqrt{1+\frac{l}{p\tilde{c}}} - \sqrt{\frac{l}{p\tilde{c}}}$. With probability greater then $1-\delta$ the following inequality holds - $c \geq \tilde{c}\cdot\beta^2$.*

Proof. *Let S be a subset of vertices such that $|\partial_G S| = c$ then from the Chernoff-Hoeffding inequality (the* *one-sided version)*

$$P\left(|\partial_{\tilde{G}}S| \geq (1+\epsilon)|\partial_G S|\right) = P\left(p|\partial_{\tilde{G}}S| \geq (1+\epsilon)p|\partial_G S|\right)$$
$$\leq \exp\left(-\frac{\epsilon^2 pc}{3}\right)$$

Where we multiply by p to have all the elements bounded by 1. Setting $\epsilon = \sqrt{\frac{3\ln(\frac{1}{\delta})}{pc}}$ *we get that with probability greater then* $1-\delta$ *that*
$pc\left(1+\sqrt{\frac{3\ln(\frac{1}{\delta})}{pc}}\right) = pc + \sqrt{3\ln(\frac{1}{\delta})}\sqrt{pc} \geq |\partial_{\tilde{G}}S| \geq p\tilde{c}$. *By completing the square we get that*

$$\left(\sqrt{cp}+\sqrt{\frac{3\ln(\frac{1}{\delta})}{4}}\right)^2 = \left(\sqrt{cp}+\sqrt{l}\right)^2 \geq p\tilde{c}+l$$

which means (after some simple algebraic manipulation) that

$$c \geq \tilde{c}\beta^2$$

□

We can combine these to theorems and get

Theorem C.2. *Let G be a graph with weights* $w_{ij} \in [0,1]$ *and* $\tilde{G}$ *its approximation after sampling m edges with minimal cut* $\tilde{c} > 0$. *Define* β *and k as in previous theorems. If* $m \geq \binom{n}{2}\frac{3(2\ln(n)+\ln(\frac{2}{\delta})+k)}{\epsilon^2\beta^2\tilde{c}}$ *then the probability that* $\tilde{G}$ *is not an* ϵ*-cut approximation is smaller then* δ.

Proof. *This is just using lemma C.2 with error probability* $\frac{\delta}{2}$ *and using that c for theorem C.1 with the same error probability and the union bound.* □

This theorem gives a high probability bound that depends only on observable quantities. While the notation is a bit cumbersome, it is easy to see that if $p\tilde{c} \gg \ln(\frac{1}{\delta})$, i.e. the unscaled weight of the smallest cut is not too small, then $\beta \approx 1$ and we have a bound that is almost as good as if we knew the real c.

We will now prove theorem 3.2 in the paper.

Theorem (3.2)**.** *Let G be a graph with weights* $w_{ij} \in [0,1]$ *and* $\tilde{G}$ *its approximation after observing m edges. Assume G is partitioned into* ℓ *clusters each has minimal cut greater or equal to* c_{in}, *and the cuts separating clusters from the others is smaller then* c_{out}. *Furthermore assume* $c_{in} > 4c_{out}$. *If* $m \geq \frac{12n^2}{c_{in}}\left(2\ln(n)+\ell\ln(\frac{2}{\delta})+k\right)$ *then the cuts separating the clusters are smaller then any cut that cuts into one of the clusters.*

Proof. *After seeing m edges, the probability for sampling any edge inside any cluster is* $p = m/\binom{n}{2}$. *By theorem C.1 we have that if* $m \geq \frac{12n^2}{c_{in}}\left(2\ln(n)+\ln(\frac{2^\ell}{\delta})+k\right)$ *then the probability of any cut in a single cluster being smaller then* $\frac{c_{in}}{2}$ *is smaller then* $\frac{\delta}{2\ell}$, *with the union bound we have that with probability greater then* $1-\frac{\delta}{2}$ *all cuts in any cluster (and therefore any cut in* $\tilde{G}$ *that cuts some cluster) have weights greater or equal to* $\frac{c_{in}}{2}$.

We now need to show that the cuts separating the clusters are not too large. Consider a cut separating some clusters from the others. If the weight of this cut is c we need to show that with probability greater then $1-\frac{\delta}{2^{\ell+1}}$ *we have* $\tilde{c} < \frac{c_{in}}{2}$. *This means that we want to show that* $\tilde{c} < (1+\tilde{\epsilon})c \leq (1+\tilde{\epsilon})c_{out} = \frac{c_{in}}{2}$, *i.e. we can use the negatively dependent Chernoff-Hoeffding inequality (theorem A.1) with* $\tilde{\epsilon} = \frac{c_{in}}{2c_{out}} - 1 > \frac{c_{in}}{4c_{out}}$ *(using the fact that* $c_{in} > 4c_{out}$*) and get that the* $P(p\tilde{c} - pc > (1+\tilde{\epsilon})pc_{out}) \leq \exp\left(-\frac{\tilde{\epsilon}^2 pc_{out}}{3}\right) \leq \exp\left(-\frac{pc_{in}}{12}\right)$. *As* $m \geq \frac{12n^2}{c_{in}}\ln(\frac{2^\ell}{\delta})$ *we can finish the proof.* □

D Adaptive Unbiased Sampling

For adaptive unbiased sampling, one needs to reweigh the sampled edge weights according to their sampling probabilities. This re-scaling is not easy to compute in general when sampling without replacement, as the probability of sampling an edge is a marginal distribution over all the algorithm's possible trajectories. Sampling with replacement is much easier, since it only depends on the sampling probability in the current iteration. Moreover, as long as we sample only a small part of all edges, the risk of re-sampling an already-sampled edge is negligible. Therefore we will show bounds concerning on adaptive sampling with replacements. We will present here a proof that it has the same theoretical guarantees as uniform sampling for cut approximation.

Let $\tilde{G}_i$ be the graph build at step i, an adaptive sampling algorithm is an algorithm who picks an edge at step $i+1$ with probability $p(e;\tilde{G}_i)$ that depends on $\tilde{G}_i$. In order to prove that with high probability $\tilde{G} = \tilde{G}_m$ is a ϵ-approximation of G for $m = o(n^2)$ we need that $p(e;\tilde{G}_i)$ isn't too small on any edge. This can be easily done by sampling according to a modified distribution - with probability 0.5 pick an edge uniformly, and with probability 0.5 pick it according to $p(e; H_i)$. The new distribution satisfies $\tilde{p}(e;\tilde{G}_i) = \frac{1}{2}p(e;\tilde{G}_i) + \frac{1}{n(n-1)} > \frac{1}{n^2}$.

The graphs $\tilde{G}_i$ are by no means independent. Although one can view (after subtracting the mean) them as a martingale process, using the method of bounded differences [7] will not suffice, as it depends on the square of the bounding constant, so we will have a n^4 factor that only gives a trivial bound. We will next show that a high probability bound does exists.

Consider a cut with weight c that contains the edges $e_1, ..., e_l$ and consider any bounded adaptive sampling algorithm with replacements with m steps. Define X_{ik} with $1 \leq i \leq l$ and $1 \leq k \leq m$ to be the random variable that has value $\frac{w(e_i)}{\tilde{p}(e_i)}$ if the edge e_i was chosen at step k and zero otherwise. Define $Y_k = \sum_{i=1}^{l} X_{ik}$, Y_k is the weight added to the cut at step k and its expectation is c.

Lemma D.1. *If* $\forall i, l : \tilde{p}(e_i) \geq \rho$ *and* $w(e_i) \leq 1$ *then*

$$\mathbb{E}[\exp(t\rho Y_k)|\tilde{G}_{k-1}] \leq \exp(c\rho(e^t - 1))$$

Proof. *Since at most one of the positive variables* X_{ik} *is nonzero for a constant* k *then they are negatively dependent when conditioned by* $\tilde{G}_{k-1}$. *This implies that* $\mathbb{E}[\exp(t\rho Y_k)|\tilde{G}_{k-1}] \leq \prod_{i=1}^{l} \mathbb{E}[\exp(t\rho X_{ik})|\tilde{G}_{k-1}]$. *By definition of* X_{ik} *we get that*

$$\mathbb{E}[\exp(t\rho X_{ik})|\tilde{G}_{k-1}] = \tilde{p}(e_i) \cdot \exp\left(\frac{t\rho w(e_i)}{\tilde{p}(e_i)}\right) + (1 - \tilde{p}(e_i)) \tag{4}$$

One can easily verify that the right hand side of equation 4 decreases monotonically with $\tilde{p}(e_i)$, *so the fact that* $\rho < \tilde{p}(e_i)$ *and* $w(e_i) \leq 1$ *implies that*

$$\mathbb{E}[\exp(t\rho X_{ik})|\tilde{G}_{k-1}] \leq \rho w(e_i)e^t + (1 - \rho w(e_i)) =$$

$$= \rho w(e_i)(e^t - 1) + 1 \leq \exp(\rho w(e_i)(e^t - 1))$$

Where the last inequality is due to the fact that for $1 + x < e^x$. *We can finish the proof since*

$$\mathbb{E}[\exp(t\rho Y_k)|\tilde{G}_{k-1}] \leq \prod_{i=1}^{l} \mathbb{E}[\exp(t\rho X_{ik})|\tilde{G}_{k-1}]$$
$$\leq \exp(\rho c(e^t - 1)).$$

as $\sum w(e_i) = c$. ☐

We can now prove the concentration of measure bound for a single cut

Theorem D.1. *Let G be a graph such that $w(e_i) \leq 1$ and $\tilde{G} = \tilde{G}_m$ the output of a bounded adaptive sampling algorithm with replacements such that $\tilde{p}(e_i) \geq \rho$ then the probability that a cut with weight c in $\tilde{G}_m$ is not a ϵ-approximation is bounded by $2\exp\left(-\frac{\epsilon^2 \rho mc}{3}\right)$.*

Proof. *We need to show that*

$$P\left(\left|\sum_{k=1}^{m} Y_k - mc\right| > \epsilon mc\right) \leq 2\exp\left(-\frac{\epsilon^2 \rho mc}{3}\right)$$

The proof is similar to the proof of the Chernoff bound, replacing independence with lemma D.1. First look at $P\left(\sum_{k=1}^{m} Y_k > (1+\epsilon)mc\right)$. Using the standard trick for all $t > 0$

$$P\left(\sum_{k=1}^{m} Y_k > (1+\epsilon)mc\right) =$$
$$P\left(\exp\left(t\rho \sum_{k=1}^{m} Y_k\right) > \exp(t(1+\epsilon)\rho mc)\right)$$

By the Markov inequality this is bounded by $\frac{\mathbb{E}\left[\exp\left(t\rho \sum_{k=1}^{m} Y_k\right)\right]}{\exp(t(1+\epsilon)\rho mc)}$. The law of total expectation states that $\mathbb{E}\left[\exp\left(t\rho \sum_{k=1}^{m} Y_k\right)\right] = \mathbb{E}\left[\mathbb{E}\left[\exp\left(t\rho \sum_{k=1}^{m} Y_k\right) | \tilde{G}_{m-1}\right]\right]$. As $\sum_{k=1}^{m-1} Y_k$ is a deterministic function of $\tilde{G}_{m-1}$ this is equal to

$$\mathbb{E}\left[\mathbb{E}\left[\exp(t\rho Y_m) | \tilde{G}_{m-1}\right] \exp\left(t\rho \sum_{k=1}^{m-1} Y_k\right)\right]$$
$$\leq \mathbb{E}\left[\exp\left(t\rho \sum_{k=1}^{m-1} Y_k\right)\right] \exp(\rho c(e^t - 1)).$$

using lemma D.1. By induction we can conclude that the expectation is smaller then $\exp(\rho mc(e^t - 1))$. We have shown that

$$P\left(\sum_{k=1}^{m} Y_k > (1+\epsilon)mc\right) \leq \frac{\exp(\rho mc(e^t - 1))}{\exp(t(1+\epsilon)\rho mc)}$$

Following the steps as in the standard Chernoff bound proof one can show that this is smaller (for the right t) then $\exp\left(-\frac{\epsilon^2 \rho mc}{3}\right)$. The proof for this bound on $P\left(\sum_{k=1}^{m} Y_k < (1-\epsilon)mc\right)$ is done in a similar fashion, and using the union bound we finish our proof.

Using $\rho = \frac{1}{n^2}$ one can now show similar theorems to what we shown in the previous section with this theorem replacing the (negatively dependent) Chernoff bound.