CMG

November 6, 2018

1 The standard CMG of KRV

1. Start with G as empty graph
2. Find a b-balanced cut (both sides $\geq bn$ vertices for some $b = \Omega(1)$) with expansion $< 1/4$. Let T be smaller side. (This differs from KRV, which always finds *sparsest* b-balanced cut.)
3. Add arbitrary vertices until T has size $n/2$
4. Add arbitrary matching between $T, \overline{T}$ into G
5. Repeat while there exists b-balanced cut with expansion $< 1/4$
6. Do extra stuff to ensure no unbalanced sparse cuts too (irrelevant for our case)

1.1 Proof strategy

Define $p_{u,v}(t)$ as follows. We have matchings $M_1, \ldots, M_t$ in G so far. Consider a random walk starting at u. For steps i from 1 to t in that order, with probability 1/2, walk along the edge in M_i incident to the walk's current location. Let $p_{u,v}(t)$ be the probability that the walk (starting from u) ends at v. We define the potential function

$$\Phi(t) := \sum_u \text{entropy}(p_{u,\cdot}(t)) = \sum_u \left(-\sum_v p_{u,v}(t) \log p_{u,v}(t) \right).$$

Observe that

1. $\Phi(t) \geq 0$, since entropy is always nonnegative.
2. $\Phi(t) \leq n \log n$, since each entropy$(p_{u,\cdot}(t))$ is at most $\log n$.
3. $\Phi(t) \geq \Phi(t-1)$. We show the stronger statement entropy$(p_{u,\cdot}(t)) \geq$ entropy$(p_{u,\cdot}(t-1))$ for all $u \in V$. This is intuitively because (1) the walk at step t averages, for each edge in matching M_i, the values of $p_{u,v}(t-1)$ at the two matched vertices, and (2) entropy function is concave.

We'll show that $\Phi(t)$ increases by $\Omega(n)$ whenever we find a b-balanced cut with expansion $< 1/4$.

From now on, let's re-define b so that $|T| = bn$, i.e., the smaller side of the cut has exactly bn vertices. Define $q_u := \sum_{v \in \overline{T}} p_{u,v}(t)$, the probability that a walk starting from u ends up in $\overline{T}$.

Claim 1. *$\sum_{u \in T} q_u(t) < bn/8$. That is, on average, a random walk starting at a vertex in T should have a small chance (1/8) of "escaping" to $\overline{T}$.*

Proof. There are less than $bn/4$ edges crossing T. We want to show that each edge is "responsible" for a total of $1/2$ probability inside $\sum_{u\in T} q_u(t)$. This means that $\sum_{u\in T} q_u(t) \le (1/2)\cdot |E(T,\overline{T})| < bn/8$.

Imagine running each random walk starting from each $u \in T$ in parallel. We can couple the walks so that at any given time, there is a single random walk currently at each vertex in V. This means that for all $v \in V$, $\sum_u p_{u,v}(t) = 1$. Now, for each edge inside $E(T,\overline{T})$ (say, (u,v) where $u \in T, v \in \overline{T}$), consider the matching M_i that it belongs to. The edge is responsible for mixing $1/2$ of the total of 1 mass at u into v at step i of the random walks. Note that the inequality $\sum_{u\in T} q_u(t) \le (1/2)\cdot |E(T,\overline{T})|$ is because there may be some chance that a random walk goes out of T (and therefore charged to some edge in $E(T,\overline{T})$ but then returns back to T. $\square$

Fix a vertex u such that $q_u(t) \le 1/4$, i.e., a random walk starting from u escapes with probability at most twice the average of $1/8$. By Claim 1 and an averaging argument, there are $\ge bn/2$ vertices $u \in T$ satisfying this. Our next goal is to show that for each such u, the vertices $v \in T$ with $p_{u,v}(t) \ge 2p_{u,\pi(v)}(t)$, where $\pi(v)$ is the matched edge of v in M_t (and therefore $\pi(v) \in \overline{T}$ by construction), make up at least a constant fraction of the total probability by $p_{u,v}(t)$. It is easy to see (algebra) that if we mix two probabilities p, q with $p \ge 2q$, then the entropy increase

$$-\left(2\cdot\left(\frac{p+q}{2}\right)\log\left(\frac{p+q}{2}\right)\right) - (-p\log p - q\log q)$$

is $\Omega(p)$. So it suffices to find a constant fraction of vertices $v \in T$ satisfying $p_{u,v}(t) \ge 2p_{u,\pi(v)}(t)$.

Recall that every vertex in T is matched to a vertex in $\overline{T}$. Call a vertex $v \in T$ "good" if $p_{u,v}(t) \ge 2p_{u,\pi(v)}(t)$, and "bad" otherwise. We have $q_u(t) = \sum_{v\in\overline{T}} p_{u,v}(t) \le 1/4$, so the total probability $p_{u,v}(t)$ of all bad $v \in T$ is at most $1/2$. Since this vertex u has probability $\ge 3/4$ inside T, the probability that it's inside T and good is $\ge 3/4 - 1/2 = 1/4$. In other words, $\sum_{v \text{ good}} p_{u,v}(t) \ge 1/4$. So we're done.

2 Modifications

Let's make the following two changes to the algorithm:

1. We find approximate b-balanced sparse cuts. In particular, if the graph has expansion $< \frac{1}{\text{polylog}(n)}$, then we can find a b-balanced cut of expansion $< 1/4$.

2. We don't augment T into a set of size $n/2$. So we find a matching in $(T,\overline{T})$ that saturates T but doesn't necessarily saturate $\overline{T}$.

I think both of these changes are fine. Note that for (1), the potential function proof of CMG makes no use of the conductance of G. That is, no matter the current conductance of G, as long as we can find a b-balanced cut of sparsity $< 1/4$, then $\Phi(t)$ increases by $\Omega(n)$. For (2), some vertices in each M_i are now no longer matched, but we can re-define the random walks so that if a walk is currently at an unmatched vertex at step i, then stay there with probability 1. This allows us to again couple the random walks starting from each $u \in V$, so that every edge in $E(T,\overline{T})$ is responsible for $1/2$ probability out.