# An Improved Approximation Algorithm for Multiway Cut
## Gruia Calinescu Howard Karlo Yuval Rabani
## Interim report for CS 6150

Gurupragaash Annasamy Mani Praveen Thiraviya Rathinam

December 16, 2015

## 1 TODO

- What is Multiway cut
- Why is multiway cut MAXSNP hard ?
- what is MAXSNP/SNP ?
- Related work - Dalhaus - the below one . check if we should add some more ?
- Ford fulkerson for max flow ????
- Chopra and Rao [6] and Cunningham [7] develop a polyhedral approach to Multiway Cut ?
- Bertisimas non-linear formulation
- The case k=2 is not the only polynomially solvable instance of the multiway cut problem. Lovdsz and Cherkasskij show that when c(e) = 1 and G is Eulerian, then the multiwaycut problem is polynomially solvable.
- write about simplex, polytope

## 2 LP1

The problem the authors are looking into is a Multiway cut. What is a multiway cut ? Let G = (V, E) be an undirected graph on V = 1, 2, . . . , n in which each edge uv E has a non-negative cost c(u, v) = c(v, u), and let T = 1, 2, . . . , k V be a set of terminals. Multiway Cut is the problem of finding a minimum cost set C E such that in (V, E C), each of the terminals 1, 2, . . . , k is in a different component. The LP stated in the paper is

$$\min \quad \sum_{uv \in E} c(u,v)d(u,v) \tag{1}$$

$$\text{such that} \quad (V,d) \text{ is a semimetric} \tag{2}$$

$$d(t_1,t_2) = 1 \quad \forall t_1, t_2 \in T, t_1 \neq t_2 \tag{3}$$

$$d(u,v) \in (0,1) \quad \forall u, v \in V \tag{4}$$

To understand how this works, we need to know what a semimetric is. Semimetric is a pair $(V, d)$, where $V$ is a set and $d$ is a function which operator on $V$. $d : V \times V \to \mathbb{R}$, such that, $\forall u, v \in V$

$$d(u,v) = d(v,u) \geq 0 \tag{5}$$

$$d(u,u) = 0 \tag{6}$$

$$d(u,v) \leq d(u,w) + d(w,v) \tag{7}$$

Consider the graph in Figure. Let nodes A and E be terminals and we want to find a cut. In the lp, all the $d(u, v)$ where $uv$ is an valid edges is the decision variable and the value of it can be 0 or 1. From equation 3, we can say that $d(A, E) = 1$ and we dont know the other values. But based on equation 7, we can write $d(A, E) \leq d(A, B) + d(B, E)$, $d(A, E) \leq d(A, C) + d(C, E)$ and $d(A, E) \leq d(A, D) + d(B, D)$. These contribute to the additional contraints which are derived from the contraint $(V, d)$ is a semimetric. Now the LP can should assign atleast value 1 to one of the pairs in $(d(A, B), d(B, E))$, $d(A, C), d(C, E)$ and $d(A, D) + d(B, D)$. Assign the d(u, v) with a value 1 means, selecting the edge for the cut. Since our objective function is to minimize it, the LP will find which edge costs minimum and that will be the optimal solution. The authors want to come up with much more stricter contraints to reduce the number of valid solution. Lets look into the contraint $\sum_{t \in T} d(u, t) = k - 1 \quad \forall u \in V$ using the example from the Figure. In this example, the nodes A, B and C are terminals. If we run our LP without the new contraints, there are four valid solutions, which represent the cuts (AD, DB), (AD, DC), (DB, DC) and (AD, DB, DC). In this set of values, (AD, DB, DC) is not a optimal one but still is a valid solution. In a large graph, there could be many such solution, which could increase the number of valid solution. But this one can be removed easily with the new contraint, $\sum_{t \in T} d(u, t) = k - 1 \quad \forall u \in V$. The new contraint says, if there is a node u, as long as its not connected to more than one terminal it is fine. So if there are k nodes, then $\sum_{t \in T} d(u, t) = k - 1$, which means only $K - 1$ cuts are required and not $K$ cuts

# 3 Understanding

Isolation Heuristic - In isolation heuristic, if there are K terminals, in each iteration we attach one terminal to the source and all the other terminals to the sink and then run a max-flow and and find the min-cut. Let the edge set after each iteration be $E_i$. Now the lowest K-1 cuts are taken and the union of them gives the multi-way cut. This algorithm gives an approximation of $2(1 - \frac{1}{k})$. Proof:

- Run the isolation heuristic and get the set of edges. Let them be A.
- Lets say $E^*$ is the optimal edge set for the multiway cut with K terminals. Then it means if $E^*$ is removed then there will be $K$ disjoint connected graphs$(V_1, V_2, \cdots, V_k)$, each having one terminal respectively
- Lets say $E^* = \sum i = 1^k E_i^*$ and each $E_i^*$ represents the edges removed to disconnect $V_i$ from the rest.
- Lets say $\delta(V_i)$ gives the set of all outgoing edges from $V_i$.
- Now we can say that, $w(E_i) \leq w(\delta(V_i))$ because, both of isolates the terminal $i$ from the rest and we know $E_i$ is the mincut. So $w(E_i)$ cannot be greater than $w(\delta(V_i))$
- Lets says there is $m$ edges between $V_i$ and $V_{i+1}$. Then both $\delta(V_i)$ and $\delta(V_{i+1})$ will include those $m$ edges.
- $2w(E^*) = \sum_{i=1} k\delta(V_i)$. Its 2 times since each edges is double counted.
- We know $w(A) \leq (1 - \frac{1}{k})(\sum_{i=1}(E_i^*))$. This is because, A is the union of first K-1 smallest set. In this expression, we are adding up all the $K$ values. Since only $k - 1$ values are taken, we are multiplying it with $1 - \frac{1}{k}$ and since it the union of all these set, it will be equal to or less than the summation of indiviual sets.

$$
\begin{aligned}
w(A) &\leq (1 - \frac{1}{k})(\sum_{i=1}(E_i^*)) \\
&\leq (1 - \frac{1}{k})(\sum_{i=1} \delta(V_i) \\
&\leq 2(1 - \frac{1}{k})w(E^*) \\
&\leq 2(1 - \frac{1}{k})OPT
\end{aligned}
$$

How did the authors comeup with the Linear program which they say is an SLP.

- As discussed earlier, any valid solution will split the Graphi $G(V, E)$ into $K$ graphs$(C_1, C_2, \cdots, C_k)$, with each $C_i$ having a terminal $S_i$

- Let $\delta(C_i)$ represent the set of edges which goes out of $C_i$
- Each vertex $v \in V$ is represented as $x_i^j$ and this value is set to 1 if the $i^{th}$ vertex is part of the componenet $C_i$, else its set to 0. Since a vertex can be a part of only one component, we can say that $\sum_{j=1}^{k} x^j = 1, \forall x \in V$,
- We define another notation $z_e^i$ which operates on all the edges in E and is set to 1 if the edge $e$ is in $\delta(C_i)$ else its set to 0
- If $e$ is in the $\delta(C_i)$, then it connect $(u, v)$ where $u \in C_i, v \notin C_i$. Thus we can write $z_e^i$ as $z_e^i = x_u^i - x_v^i$, since $x_u^i$ will be 1 and $x_v^i$ will be 0, as per the definition of $x$. If $e$ is not in the $\delta(C_i)$, then both u and v belong to a same component, then both $x_u^i$ and $x_v^i$ will be 1 if both belong to component $C_i$ else both will be 0.
- Now for the objective function is $\frac{1}{2}\sum\limits_{e\in E} c_e * \sum\limits_{i=1}^{k} z_e^i$, where $c_e$ is the cost/weight of the edge. We are multiplying it with $\frac{1}{2}$ because of the double counting which we discussed in the previous explanation.
- Thus the LP is

$$\min \quad \frac{1}{2}\sum_{e\in E} c_e * \sum_{i=1}^{k} z_e^i$$

$$\text{such that} \quad z_e^i = |x_u^i - x_v^i|, \quad \forall e \in E$$

$$\sum_{i=1}^{k} x_v^i = 1, \quad \forall v \in V$$

$$x_{s_i}^i = 1, \quad \forall s_i \in T(\text{Set of Terminals})$$

$$x_v^j \geq 0, \quad \forall v \in V \quad (\text{Got by adding LP relaxation to } x_v^j \in (0,1), \quad \forall v \in V)$$