Saarland Informatics Campus

Assignment 8, Approximation Algorithms
Summer term 2017

Tobias Mömke, Hang Zhou
http://www-cc.cs.uni-saarland.de/course/61/

Due: 20 June 2017

Exercise 8.1 (10 Points)
Complete the proof of Lemma 1 in the lecture: show that there is a polynomial-time deterministic algorithm that can find a value $r \in [0, 1/2)$ such that $c(r) \leq (2\ln(k+1))V(r)$.

In the following, we study the *correlation clustering* problem. The goal is to design an approximation algorithm for this problem using the region growing technique (see lecture).

In the correlation clustering problem, we are given a graph $G = (V, E)$, where every edge $e \in E$ has a weight c_e as well as a label of either "+" or "−". (We note E^+ as the set of positive edges and E^- as the set of negative edges.) The goal is to find a partition of the vertices into components, such that the total weight of "+" edges whose endpoints are in different components plus the total weight of "−" edges whose endpoints are in the same component is minimized.

Exercise 8.2 (8 Points)
Show that the following LP is a linear programming relaxation of the correlation clustering problem:

$$\begin{aligned} \text{minimize } & \sum_{e\in E^-}(1-x_e)c_e + \sum_{e\in E^+} x_e c_e \\ \text{subject to } & 0 \leq x_{uv} \leq 1 && \forall u, v \\ & x_{uv} + x_{vw} \geq x_{uw} && \forall u, v, w \\ & x_{uv} = x_{vu} && \forall u, v. \end{aligned}$$

Let x be an optimal solution to the above LP. We use the same notation of balls $B(u, r)$ as in the lecture. For a set of vertices S, we define $\delta(S)$ slightly differently: it is the set of all *positive* edges that have exactly one endpoint in the set S. Similarly, the definition of the volume $V(u, r)$ differs: it is now with respect to positive edges only.

Consider the following algorithm:

Algorithm 1 Algorithm for correlation clustering

```
x ← an optimal solution to the LP
repeat
    Pick an arbitrary node u ∈ G
    Choose a radius r around u such that c(δ(B(u, r)) ≤ (3 ln(n + 1))V(u, r)
    Output the vertices in B(u, r) as a cluster
    Remove from G the vertices in B(u, r) and their incident edges
until G is empty
```

Exercise 8.3 (8 Points)
Show that for any vertex u one can find in polynomial time a radius $r \leq 1/3$ such that

$$c(\delta(B(u, r))) \leq (3\ln(n+1))V(u, r).$$

Exercise 8.4 (8 Points)
Show that in the solution of the algorithm, the total weight of positive edges between clusters is at most $(3\ln(n+1))\sum_{e\in E^+} x_e c_e$.

Exercise 8.5 (8 Points)
Show that in the solution of the algorithm, the total weight of negative edges inside the same cluster is at most $3\sum_{e\in E^-}(1-x_e)c_e$.

Exercise 8.6 (8 Points)
Conclude that the above algorithm gives an $O(\ln n)$-approximation for the correlation clustering problem.