# Chapter 10

# Clustering

## 10.1 Slide 10/15ff

**Theorem 10.1.** *The tree growing algorithm finds a connected dominating set of size* $|CDS| \leq 2(1 + H(\Delta)) \cdot |DS_{OPT}|$.

*Proof.* Assume we knew an optimum dominating set $DS_{OPT}$. We partition the graph along the nodes in $DS_{OPT}$, i.e. each node $u \in DS_{OPT}$ together with its neighbors is a cluster $S_u$. If a node is dominated by more than one node in $DS_{OPT}$, we can assign it arbitrarily to any of them.

Now we concentrate on a single cluster $S_u$. Whenever the tree growing algorithm colors some nodes in $S_u$ either black or grey, we call that a step. Whenever we do a step, we count the number of white nodes in $S_u$. Initially, before doing the first step, all nodes in $S_u$ are white; we say that $u_0 = |S_u|$. After the $i^{th}$ step $u_i$ nodes are colored white, with $u_i < u_{i-1}$, for $i > 0$. After $k$ steps all nodes in $S_u$ are either grey or black, i.e. $u_k = 0$.

Coloring a node black costs 1 for the tree growing algorithm. We will share this cost among all the nodes that are newly colored. In the first step we color $u_0 - u_1$ nodes in $S_u$. Since the tree growing algorithm always colors at most 2 nodes black, these nodes share a cost of at most $2/(u_0 - u_1)$.

After the first step, node $u$ is eligible to be colored black by the tree growing algorithm, because at least one node in $S_u$ is grey. By doing so, in step $i$ we can at least newly color $u_i$ nodes. Since the tree growing algorithm always selects nodes with maximum coloring potential, we know that in step $i$ at least $u_i$ nodes are colored. That is, the charge to the newly colored nodes in step $i$ in $S_u$ is at most $2/u_i$.

Now we just need to sum up. The total cost $C$ of the nodes in $S_u$ is bounded from above by

$$
\begin{aligned}
C &\leq \frac{2}{u_0 - u_1}(u_0 - u_1) + \sum_{i=1}^{k-1} \frac{2}{u_i}(u_i - u_{i+1}) \\
&= 2 + 2\sum_{i=1}^{k-1} \frac{u_i - u_{i+1}}{u_i} \\
&\leq 2 + 2\sum_{i=1}^{k-1} (H(u_i) - H(u_{i+1})) \\
&= 2 + 2(H(u_1) - H(u_k)) = 2(1 + H(u_1)) \leq 2(1 + H(\Delta)).
\end{aligned}
$$

The third line follows from the definition of the Harmonic number $H(n) = \sum_{i=1,\ldots,n} 1/i \approx \log n + 0.7$: For two integers $b$, $a$, with $b > a$, we have $H(b) - H(a) = \sum_{i=a+1,\ldots,b} 1/i \geq (b-a)/b$.

So far, we showed that the cost shared by the nodes in $S_u$ is bounded from above by $2(1+H(\Delta))$. The optimum dominating set $DS_{OPT}$ only needed node $u$ in set $S_u$, that is, the cost of the optimum dominating set in $S_u$ is 1. This is the case of all nodes/clusters in the optimum dominating set, therefore the total overhead of the tree growing algorithm is $2(1 + H(\Delta))$ as claimed. □