Problem Set 5

Please hand in your solutions for this problem set via email (roesner@cs.uni-bonn.de) or personally at Room 2.060 until Tuesday, 13th of November.

Problem 1
Let us look at the following algorithm inspired by the streaming algorithm for \( k \)-center.

1. Set \( C_{\lfloor P \rfloor} = P \) and \( \ell = \min_{p \neq q \in P} d(p, q) \).
2. For \( k = \lfloor P \rfloor - 1 \) to 1 do
   3. Set \( C_k = C_{k+1} \)
   4. While \( |C_k| > k \) do
   5. Compute \( I \subseteq C_{k+1} \) by calling \texttt{maximal-independent-set}(G^2_\ell(C_{k+1}))
   6. Set \( C_k := I \) and \( \ell = 2 \cdot \ell \)

It computes for every \( k \in \{1, \ldots, \lfloor P \rfloor\} \) a set \( C_k \) of at most \( k \) centers. Additionally we recursively define a cluster \( C_{p,k} \) for each point \( p \in C_k \) as follows.

1. For each \( p \in P \) set \( C_{p,\lfloor P \rfloor} = \{p\} \).
2. For \( k = \lfloor P \rfloor - 1 \) to 1 do
   3. For each \( p \in C_k \) Set \( C_{p,k} = C_{p,k+1} \).
   4. For each \( p \in C_{k+1} \setminus C_k \)
   5. Let \( n_p = \arg \min_{q \in C_k} d(p, q) \).
   6. Set \( C_{n_p,k} = C_{n_p,k} \cup C_{p,k+1} \).

- Show that this algorithm induces a hierarchical clustering.
- Show an upper bound for its approximation factor.

Problem 2
We look at clustering with outliers. As before we would like to adjust an algorithm for the \( k \)-center problem with outliers to the \( k \)-supplier problem with outliers.

- Show how bad the approximation factor can become when we again start with approximating the \( k \)-center version and then replace every center with its closest location. That is, assume that we ignore all knowledge about the possible center locations and solve the \( k \)-center problem with outliers, but then use this solution for the \( k \)-supplier problem with outliers by moving the chosen centers to the closest possible location.
• What changes when we instead use an approximation algorithm for the \(k\)-center problem with outliers and forbidden centers? This means that we are allowed to define a set \(F \subseteq P\) which we can not use as a center. In this case we are allowed to define up to \(|P|\) different sets of forbidden centers and can compute a solution for each of them.

**Problem 3**

We look at the problem of assigning children to a kindergarten in a big city. Let \(P\) be the set which contains the kindergarten teachers and all children signed up to go to a kindergarten. Let \(a : P \to \{T, C\}\) be a mapping that tells us whether \(p \in P\) is a teacher \((a(p) = T)\) or a child \((a(p) = C)\). In addition we know the set \(L\) of the places, where the different kindergartens are located. An assignment \(f : P \to L\) of children and teachers to kindergartens is called **fair** if all kindergartens have the same number of children per teacher assigned to them, i.e., for all \(\ell \in L\) we have

\[
\text{ratio}(\ell, C) := \frac{|\{x \in f^{-1}(\ell) \mid a(x) = C\}|}{|f^{-1}(\ell)|} = \frac{|\{x \in P \mid a(x) = C\}|}{|P|} = : \text{ratio}(P, C).
\]

(Notice that this implies that \(\text{ratio}(\ell, T) = \text{ratio}(P, T)\)). Assume that the number of children in \(P\) is equal to \(t\) times the number of teachers in \(P\) for an integer \(t \in \mathbb{N}\) \((\text{ratio}(P, T) = 1/(t + 1))\). If we assume that each of the teachers is already employed by one of the kindergartens show how a fair assignment of the children to the kindergartens, which minimizes the maximum distance a parent has to go to bring their child to its kindergarten, can be computed.