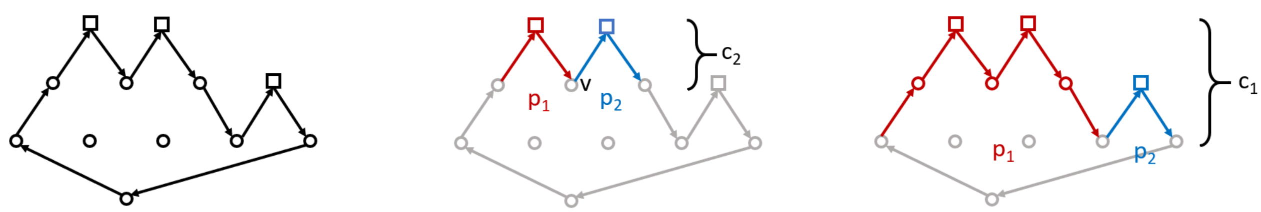
Lemma 1 *Let $GHA(\mathcal{S})$ return a superstring corresponding to a permutation $\sigma = (s_{i_1}, s_{i_2}, \dots, s_{i_n})$. Then an algorithm A that merges adjacent strings in σ in the descending order of the lengths its overlaps is greedy.*

Proof. Let c be a cycle through solution D of $GHA(\mathcal{S})$ in the hierarchical graph and let us denote by c_k a part of c above level k for any k. c_k can be represented as disjoint union $\mathcal{P}_k$ of paths that begin and end on level k, but all its inner vertices are above level k (fig. 1a and 1b).

By construction of c every such path p contains at least one string from $\mathcal{S}$ and if it contains more than one string, then in σ they go sequentially and in the same order in which p visits them. This allows us to treat such paths as already constructed superstrings of the corresponding subsets of $\mathcal{S}$. From this point of view, we can naturally identify every merge of some superstrings s and t with $|\operatorname{overlap}(s,t)| = k$ performed by A with situation when for the corresponding paths $p_s, p_t \in \mathcal{P}_k$ there is a path $p \in \mathcal{P}_{k-1}$ containing $p_s \circ p_t$ as subpath (fig. 1c). It is easy to see, that in general s and t can be merged only if the corresponding paths *touch*, i.e. a head vertex of p_s is also a tail vertex of p_t (of course, this fact doesn't imply that this paths will be merged, as there, for example, may be another path $p_r \in \mathcal{P}_k$ such that p_r and p_t touch and there is $p \in \mathcal{P}_{k-1}$ which contains $p_r \circ p_t$). It is also clear, that if A at some step doesn't merge a pair of superstrings s, t with strictly maximal overlap of the length k, then corresponding paths $p_s, p_t \in \mathcal{P}_k$ touch at some vertex v, but there are paths $p'_s, p'_t \in \mathcal{P}_{k-1}$ such that p'_s ends with $p_s \circ (v, \operatorname{suff}(v))$ and p'_t starts with $(\operatorname{pref}(v), v) \circ p_t$. In other words, p_s isn't merged with any other path with start in v, p_t isn't merged with any other path with end in v and both are contained in the different paths in $\mathcal{P}_{k-1}$.

In this way, it is sufficient to show that such situations don't happen, to prove that A is greedy. Let us suppose the opposite and consider the mentioned paths p_s, p_t, p'_s, p'_t and the vertex v. By definition of v it has income down-arc and outgoing up-arc and also income up-arc and outgoing down-arc, which by construction of D can be only in the case when v is the last chance of the corresponding component $\mathcal{C} \ni v$ to be connected to the rest arcs in D. It immediately follows, that all component $\mathcal{C}$ (and hence the paths p_s and p_t) lies in some path $p_{\mathcal{C}} \in \mathcal{P}_{k-1}$, which contradicts the definition of p'_s and p'_t. This contradiction completes the proof. □



(a) Cycle c for $\mathcal{S}$ containing 2 strings of the length 3 and one string of the length 2

(b) Colored arcs is $\mathcal{P}_2$, which contains two paths p_1 and p_2 touched at v and painted in red and blue, respectively

(c) Colored arcs is $\mathcal{P}_1$. We see, that p_1 and p_2 from previous figure now merged and are contained in new p_1. This means that corresponding strings are also merged.

Figure 1