1. *Given trees T_1 and T_2 and matching threshold r, to ensure a pq-Gram distance d such that $0 \leq d \leq r \leq 1$, then we must have $|I_1| \geq \frac{1-r}{1+r}|I_2|$.*

Proof. Without loss of generality, assume $|T_1| \leq |T_2|$. Let I_1 be the pq-Gram profile corresponding to T_1, and let I_2 be the pq-Gram profile corresponding to T_2.

Recall that the pq-Gram distance is defined as

$$dist(T_1, T_2) = 1 - 2\frac{|I_1 \cap I_2|}{|I_1 \uplus I_2|}.$$

Note that in any case $|I_1 \uplus I_2| = |I_1| + |I_2|$, and in the best case scenario, to maximize $|I_1 \cap I_2|$, $I_1 \subseteq I_2$. Then $|I_1 \cap I_2| = |I_1|$, so we can make the simplification

$$dist(T_1, T_2) = 1 - 2\frac{|I_1 \cap I_2|}{|I_1 \uplus I_2|} = 1 - 2\frac{|I_1|}{|I_1| + |I_2|}.$$

Then if r is the threshold such that $0 \leq r \leq 1$, we want distance d to be such that $0 \leq d \leq r$, so we must have

$$\begin{array}{rrcl}
 & r & \geq & 1 - 2\frac{|I_1|}{|I_1|+|I_2|} \\
\Rightarrow & 2\frac{|I_1|}{|I_1|+|I_2|} & \geq & 1 - r \\
\Rightarrow & \frac{2}{1-r}|I_1| & \geq & |I_1| + |I_2| \\
\Rightarrow & \frac{2}{1-r}|I_1| - |I_1| & \geq & |I_2| \\
\Rightarrow & \frac{2-(1-r)}{1-r}|I_1| & \geq & |I_2| \\
\Rightarrow & \frac{1+r}{1-r}|I_1| & \geq & |I_2| \\
\Rightarrow & |I_1| & \geq & \frac{1-r}{1+r}|I_2|.
\end{array}$$

□