

Using Chernoff Bound in the Proof of Lemma 6 of Approximate Similarity Search Under Edit Distance using Locality-Sensitive Hashing

Lemma 6. For any string x of length d , $Pr_\rho[\tau(x, \rho) \text{ is complete}] \geq 1 - 1/n^2$.

A transcript $\tau(x, \rho)$ is complete if $|\tau(x, \rho)| < 8d/(1 - p_a) + 6 \log n$.
Let the transcript contain l insert operations, as the maximum length of a string is d , $|\tau(x, \rho)| \leq d + l$

We check the bounds for the probability of $l > 7d/(1 - p_a) + 6 \log n$.
Consider success to be when we do not have hash-insert operations. This behaves like a geometric progression with probability $= (1 - p_a)$. The Expected number of hash-inserts we need to get an operation which is not a hash-insert would be $\frac{1}{(1 - p_a)}$.
Hence, $E[\text{hash inserts for each character}] = 1/(1 - p_a)$.
 $\therefore E[l] = d/(1 - p_a)$.

One of the Chernoff bounds is:
 $Pr(X \geq (1 + \delta)E[X]) \leq (\frac{e^\delta}{(1 + \delta)^{(1 + \delta)}})^{E[X]}$
 $\implies Pr(X \geq (1 + \delta)E[X]) \leq e^{(\delta - (1 + \delta) \ln(1 + \delta)) * E[X]} \text{ --- (1)}$
To calculate $Pr(l > 7d/(1 - p_a) + 6 \log n)$ with $E[X] = d/(1 - p_a)$, $\delta = 6 + \frac{6(1 - p_a) \log n}{d}$

Referring to "A guided tour of Chernoff bounds":
 $\epsilon - (1 + \epsilon) \ln(1 + \epsilon) \leq -\epsilon^2/3$, when $0 \leq \epsilon \leq 1$

We have the value of $\delta = 6 + \frac{6(1 - p_a) \log n}{d} > 6$, hence we cannot use the above bound, instead we use:
 $\delta - (1 + \delta) \ln(1 + \delta) \leq -\delta/3$, when $\delta > 1$

Substituting in (1), we get:
 $Pr(l \geq (1 + \delta)E[X]) \leq e^{-\delta E[X]/3}$
 $Pr(l > 7d/(1 - p_a) + 6 \log n) < e^{-(6d/(1 - p_a) + 6 \log n)/3}$

As $\frac{6d}{(1 - p_a)} > 0$,

$$\begin{aligned}
&\implies \Pr(l > 7d/(1 - p_a) + 6 \log n) < e^{-(6 \log n)/3} \\
&\implies \Pr(l > 7d/(1 - p_a) + 6 \log n) < 1/n^2 \\
&\text{Hence, } \Pr(l < 7d/(1 - p_a) + 6 \log n) > 1 - 1/n^2
\end{aligned}$$