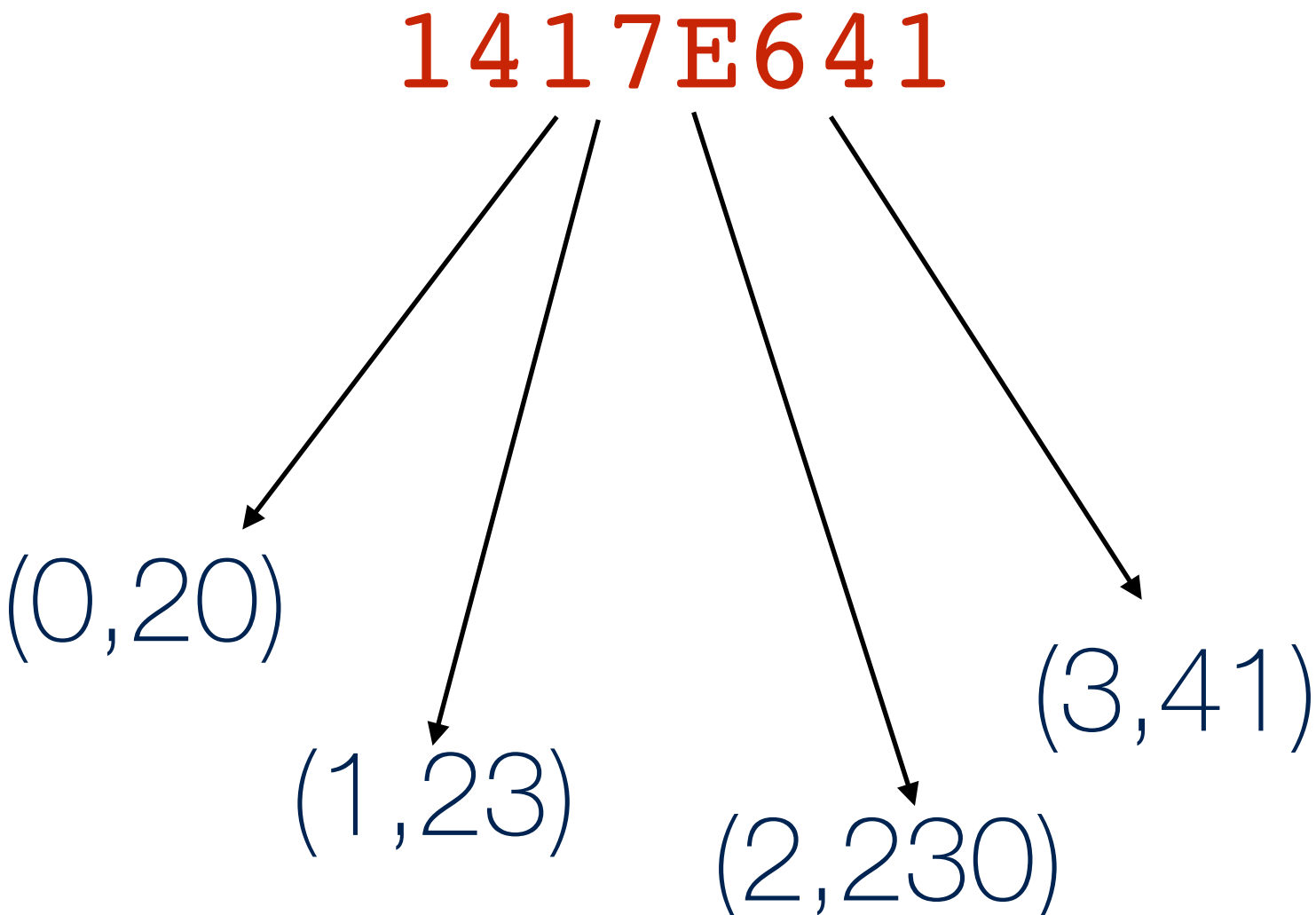


ACACTGCAC

AC
CA
AC
CT
TG
GC
CA
AC

1417E641

Map Reads
to Signature



Use LSH to FlatMap
Reads to Buckets

R1:1417E641
R2:0117E641

Predicted
Similarity:
 $(S_1 \cap S_2) / (S_1 \cup S_2)$

Compare Reads
Within Buckets