

Algorithms in Genome Research
Winter 2013/2014

Exercises

Number 6, Discussion: 2013 December 13

1. Draw the full de-Bruijn graph for the alphabet $\mathcal{A} := \{1, 2, 3\}$ with $k = 2$.
2. A basic data structure used for short-read sequence assembly is the de-Bruijn graph.
 - (a) While it is conceptually easy, there are several challenges when you want to implement it in practice – name a few.
 - (b) Give an efficient implementation of a de-Bruijn graph as a Java class.
3. Draw the 4-dimensional de-Bruijn graph (i.e. where vertices correspond to 4-mers) for the following set of reads. Can you assemble the data set into a single contig?
GTTAAT, AGACG, ACGTT, CACGG, ACTAGG, CTGACG, TACTAG, GACCAGA, TAATG, AATGC, TGCAC, GCACG, ATGCA, GTTAATG, AAATG, TGCAC, GCACG, CACGG, TAATGA, AATGAC, CAGAC, AGACG, ACCAGA, ATAATG, TAATG, AATGA, GCACGG, ACTAG, TTAATG, TAATG, TGACC, ATAAT, CCAGA, ATGCA, ATAAT, ACCTGA, ATGCAC, TGCAC, CGTTA, CGTTA, TTAATG, GACCA, ACCAG, CCAGA, CAGAC, ATGAC, GACGTT, ATGGA, ACGTT.
4. Which type of errors or ambiguities can we find in a de-Bruijn graph during an assembly process? How can we deal with them?