1. Construct a De Bruijn Graph for the following set of reads with $k = 4$:
   $R = \{ATGAT, GATTA, ATTAT, CATGA\}$. Make sure to clearly indicate edge multiplicities.

   What do you think the original sequence was?

2. Sequencing coverage (the number of times each base was sequenced) can affect what a de Bruijn graph looks like. Suppose you have "perfect sequencing" data (no errors and reads uniformly cover the genome) and have an average sequencing coverage of c. What effect does this have on the resulting de Bruijn graph?

3. Sequencing coverage (the number of times each base was sequenced) can affect what a de Bruijn graph looks like. Suppose you have gaps in your sequencing - that is portions of the genome that are not covered. What effect does this have on the resulting de Bruijn graph?

4. Eulerization is the process of turning a graph (or a multi-graph) into a Eulerian graph (one that contains an Eulerian Path). Construct the De Bruijn graph for the following set of sequences with $k = 3$. $R = \{ATGGC, GTGCA, GGCGTG\}$. (The fact that this set of sequences has different lengths should not be of a concern to you in this problem).

   What is the fewest number of edge additions or removals to make the resulting graph Eulerian?

5. What effects can choosing smaller or larger values of k have upon the resulting de Bruijn graph?

6. Can you think of any graph simplification that would be helpful for being able to find Eulerian paths (or a set of such paths) in a de Bruijn graph?