Assignment 2: Due Thursday 21st March at 5pm

Late assignments will not be accepted except by prior arrangement (for a good reason)

Please include your student number in your handed up work, as Canvas doesn't give this to me automatically.

1. Use your existing code to read in (the 6 node) directed graph specified as an edge list in TGF (Trivial Graph Format). At

   `https://roughan.info/notes/Network_Modelling/10data.html`

   you will find a second data file `a1010101_graph_A2.tgf` (in the same format) for you to apply this to. Note that the input is a *multi-graph* in that edges can appear more than once, and self-loops are allowed.

   (a) Write a function to calculate the in- and out-degree of each node.

   (b) Write a function that performs a recursive depth-first search to find a node with out-degree 0, or returns Inf (or something similar), if it can't find one. Start at node 1, and where there are alternative "next-hop" nodes, always search them in order from lowest to highest.

   Ensure to output the results of your code on your test input file.

   [4 marks]

2. (a) Derive the computational complexity of the Eulerian cycle algorithm described in lectures.

   (b) Write code to find an Eulerian walk in a directed graph using the files given above. Note that the input is a *multi-graph* in that edges can appear more than once, and self-loops are allowed.

   [3 marks]

3. Assemble the following "reads" into a complete genome. The number gives the multiplicity of each read (*i.e.,* how often it occurred).

read	multiplicity
ACT	2
ATG	1
CTG	1
CTT	1
GAC	2
GGA	1
TGA	1
TGG	1
TTC	1

   [2 marks]

4. The originator of the de Brujin graph was interested in a topic other than genome sequencing. He was interested in universal sequences.

   Take a set of symbols Ω, for example for a binary sequence we would take the symbols $\Omega = \{0, 1\}$. A k-universal string is a string in which each possible k-mer appears exactly once as a sub-string. In fact, de Brujin wanted "circular" strings, which come back to their start as well.

   So for the binary 3-universal case, he would have the string

   $$00011101$$

   or any cyclic permutation of this string (note that we imagine the end joining up with the start.

   (a) For a set Ω with $|\Omega| = n$ symbols, if it exists, how long would a k-universal circular string be?

   (b) Given $\Omega = \{a, b, c\}$, use a de Brujin graph to construct a 2-universal circular string.

   [1 marks]