# BIO231 - Computational Biology I
Assignment 3

Assignment Due: April 16, 2019 5pm

Instructions

- Read the section on plagiarism in your student handbook. Plagiarised assignments will be referred to the DC.

- Assignments must be submitted in the "Assignment" section on LMS as a zip file containing a single word document for sections 1 & 2, and individual files for Python codes. Code must be properly commented. Hard copy of the word document **MUST** also be submitted to the TAs by the deadline.

- Assignments submitted on Dropbox, emailed assignments or late submissions will not be accepted.

- This is an individual assignment. Discussions are not allowed.

- Late submission is not allowed.

1 Biological Networks

1. Write the adjacency matrices for the following networks. [2]
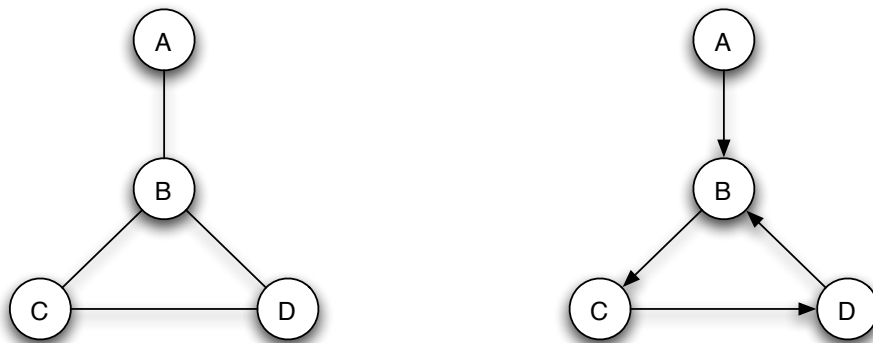
2. Consider the following set of reactions.

$$R_1: \quad A + B \longrightarrow C$$

$$R_2: \quad B + D \longrightarrow E + F$$

$$R_3: \quad B \longrightarrow A + E$$

$$R_4: \quad A \longrightarrow G$$
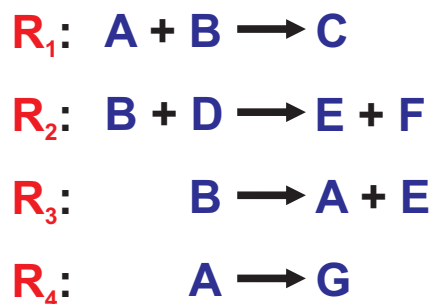
Draw the above reactions in the form of [6]

   (a) An ordinary graph
   (b) A hypergraph
   (c) A bipartite graph

3. List the nodes infected under SI model in the long run in the following network if the infection starts from node I. What will happen if it were an undirected network. [3]
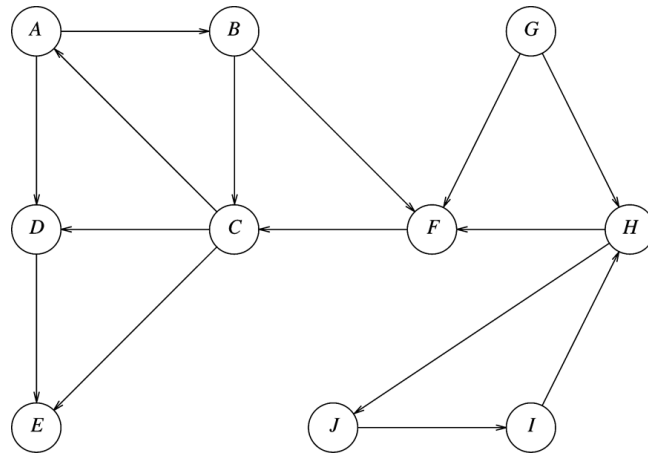
4. Consider the following boolean network. A state is represented by a string XXXX corresponding to the value of each node.
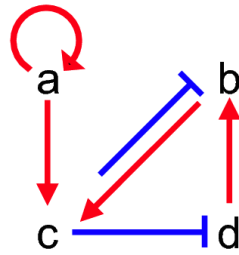
(a) Write down the boolean rules. Multiple edges incident on a node means that all these edges are acting together to produce the required behaviour. [4]

(b) Write down the successive states for the above network. [4]

(c) Write down the system trajectory from the initial state of
    i. a = 0 [1 + 1]
    ii. a = 1. [1 + 1]

Explain the system behaviour in both cases.

2 Proteins

1. Download PDB file (PDB ID: 5CHB) from Protein Data Bank (http://rcsb.org) and answer following questions. [5]

(a) How many amino acids are present in this protein structure?

(b) Which method is used to determine this structure?

(c) What is the resolution of the structure? Is it good or bad?

(d) Use an online tool to generate Ramachandran plot for this protein.

(e) Discuss the output of Ramachandran plot.

2. Find two structures for each of the following domain classes from PDB and generate publication quality images of each structure using PyMol (https://pymol.org/2/). [5+5]

(a) α Domain

(b) β Domain

(c) $\alpha+\beta$ Domain

(d) Multi-domain

(e) Membrane protein

3 Python

1. Write a program that takes DNA sequence as input and output six reading frames, and their amino acid sequence. [6]

2. Write a program that takes protein sequence as input, and predicts the helices in the sequence using Chou-Fasman algorithm.[4]

SECOND BASE

FIRST BASE	U	C	A	G	THIRD BASE
U	UUU, UUC — Phenylalanine (Phe); UUA, UUG — Leucine (Leu)	UCU, UCC, UCA, UCG — Serine (Ser)	UAU, UAC — Tyrosine (Tyr); UAA — Stop codon; UAG — Stop codon	UGU, UGC — Cysteine (Cys); UGA — Stop codon; UGG — Tryptophan (Trp)	U C A G
C	CUU, CUC, CUA, CUG — Leucine (Leu)	CCU, CCC, CCA, CCG — Proline (Pro)	CAU, CAC — Histidine (His); CAA, CAG — Glutamine (Glu)	CGU, CGC, CGA, CGG — Arginine (Arg)	U C A G
A	AUU, AUC, AUA — Isoleucine (Ile); AUG — Methionine (Met) Start codon	ACU, ACC, ACA, ACG — Threonine (Thr)	AAU, AAC — Asparagine (Asn); AAA, AAG — Lysine (Lys)	AGU, AGC — Serine (Ser); AGA, AGG — Arginine (Arg)	U C A G
G	GUU, GUC, GUA, GUG — Valine (Val)	GCU, GCC, GCA, GCG — Alanine (Ala)	GAU, GAC — Aspartic acid (Asp); GAA, GAG — Glutamic acid (Glu)	GGU, GGC, GGA, GGG — Glycine (Gly)	U C A G

Figure 1: Codon table

Name	P(a)	P(b)	P(turn)	f(i)	f(i+1)	f(i+2)	f(i+3)
Alanine	142	83	66	0.06	0.076	0.035	0.058
Arginine	98	93	95	0.070	0.106	0.099	0.085
Aspartic Acid	101	54	146	0.147	0.110	0.179	0.081
Asparagine	67	89	156	0.161	0.083	0.191	0.091
Cysteine	70	119	119	0.149	0.050	0.117	0.128
Glutamic Acid	151	037	74	0.056	0.060	0.077	0.064
Glutamine	111	110	98	0.074	0.098	0.037	0.098
Glycine	57	75	156	0.102	0.085	0.190	0.152
Histidine	100	87	95	0.140	0.047	0.093	0.054
Isoleucine	108	160	47	0.043	0.034	0.013	0.056
Leucine	121	130	59	0.061	0.025	0.036	0.070
Lysine	114	74	101	0.055	0.115	0.072	0.095
Methionine	145	105	60	0.068	0.082	0.014	0.055
Phenylalanine	113	138	60	0.059	0.041	0.065	0.065
Proline	57	55	152	0.102	0.301	0.034	0.068
Serine	77	75	143	0.120	0.139	0.125	0.106
Threonine	83	119	96	0.086	0.108	0.065	0.079
Tryptophan	108	137	96	0.077	0.013	0.064	0.167
Tyrosine	69	147	114	0.082	0.065	0.114	0.125
Valine	106	170	50	0.062	0.048	0.028	0.053

Figure 2: Chou Fasman parameters