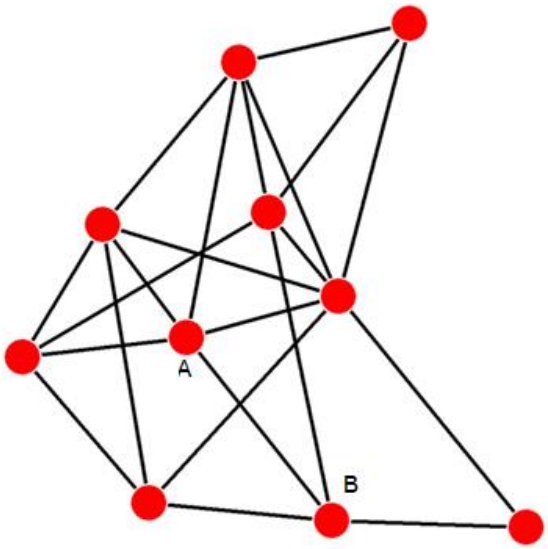
Problem set 7

This problem set covers the content from week 9: biological networks and chromatin organization

Tips and rules:

- You can answer in English or in Thai.
- There can be more than one correct answer. What I am looking for from you is not just the correct answer but the rationale for your answer.
- Please provide evidence of how you think and what sources of information you used.
- AI such as ChatGPT may be used. You can also work together with friends. But you must write the answer in your own words.
- Any incidence of plagiarism and copying of another student's work will be reported to the Graduate Affairs.

Biological networks



Given the above network, calculate the following topological characteristics for nodes A and B

Q1: Degree

Q2: Clustering coefficient

Q3: Closeness centrality (Use the first definition in https://en.wikipedia.org/wiki/Closeness_centrality)

Q4: Based on your results for **Q1-Q3**, compare the importance of nodes A and B on this network's connectivity. Would the removal of node A or B be more detrimental to the connectivity of the network? *Hint: There is no right or wrong answer. Rely on your reasoning and interpretation of those scores.*

Q5: Provide an example of a biological network **that has not been mentioned in class**. Define what the nodes and edges are. What could be the weights on the edges?

In class, we saw an example that each gene on a network may possess different topological properties, such as having high betweenness but low degree and vice versa, and that these combinations of topological properties can have specific meanings in the context of biology. For example, a gene with low degree but high betweenness maybe a gene that connect multiple function pathways. Interpret the possible biological roles of following genes:

Q6: Gene A has high degree but low clustering coefficients

Q7: Gene B has low degree but high clustering coefficients

In class, we saw an example (https://www.nature.com/articles/ncomms10331) that multiple biological networks can be combined to predict new relationships.

Q9: Explain how new target diseases for an existing drug can be discovered by such method. *Hint: Study the mentioned article.*

Q10: Provide another example of how multiple biological networks can be combined to derive new knowledge. Define the biological networks involved and explain what knowledge they can tell us.

Q11: When clustering nodes in a network into modules, most algorithms will ask you to set a parameter named **resolution**. Explain what it is and how it can impact the clustering results.

Chromatin organization

Q12: In class, we saw multiple experimental evidence showing that the folding of chromatins inside nucleus is not random. For example, https://pubmed.ncbi.nlm.nih.gov/18978785/ shows that certain genes were consistently placed in a certain area inside the nucleus. Also, inactive genes on heterochromatin are attached to the nuclear lamina.

Explain biological benefits from such precise positioning of genes inside the nucleus.

Briefly explain how the following experimental techniques work and how to interpret their results together with transcriptomics data to identify the underlying regulatory mechanisms.

	How does the technique work?	How to interpret the results together with transcriptomics data?
Q13: Bisulfite sequencing		
Q14: ATAC-seq		
Q15: ChIP-seq for transcription factor		

Q16: When interpreting the effect of a peak from ChIP-seq or ATAC-seq data, why shouldn't we simply assign each detected peak to the nearest gene? What approach would you use to

Q17: It takes very high sequencing depth to study 3D structural information of the chromatin (because we need $>> N^2$ reads to quantify all pairwise interactions between N genomic regions). When performing such experiment, how can we test whether we have obtained enough sequencing depths?