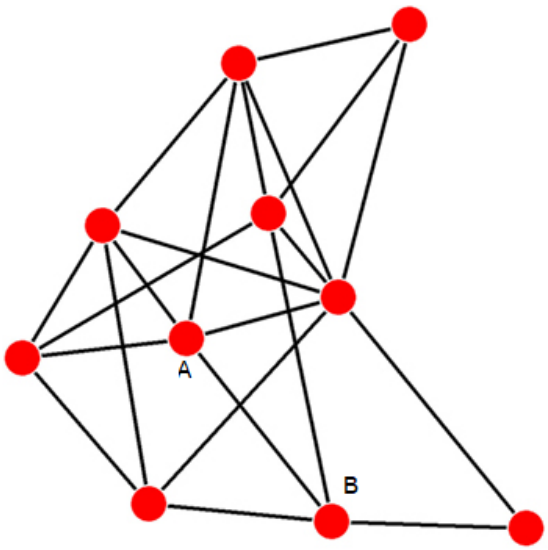
Problem set 6

Biological networks



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

Q1: Degree

Q2: Clustering coefficient

Q3: Closeness centrality (https://en.wikipedia.org/wiki/Closeness_centrality)

Q4: Betweenness centrality (https://en.wikipedia.org/wiki/Betweenness_centrality)

Q5: Based on your results for Q1-4, systematically 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 network? ***Hint****: There is no right or wrong answer. Rely on your reasoning and interpretation of those scores.*

Q6: Provide an example of a biological network **that has not been mentioned in class**. Explain what the nodes and edges represent. Provide an example of what could be the score or weight on each edge (such as expression correlation for gene co-expression network, or binding strength for protein-protein interaction network).

Q7: While some networks, such as transcription factor binding to gene loci, have a clear direction of interaction (pointing from transcription factor toward other genes), gene co-expression network has no concept of direction. This is because saying that the expression of gene A correlates with gene B means the same as saying that the expression of gene B correlates with gene A. However, in reality, we know that some of the co-expression relationships are due to directional mechanism (for example, one gene is upstream of another in a signaling pathway). Can you suggest an experimental and/or computational approach for determining the direction of relationship for interactions in gene co-expression network? ***Hint****: You can try searching literature for ideas.*

We learned that biological networks, like other real-world networks, have a certain set of characteristics, including scale-free, small-world, and power law of degree distribution. Explain what they mean in the context of biology.

Q8: Scale-free.

Q9: Small-world.

In class, we saw an example (https://www.nature.com/articles/ncomms10331) of how combination of multiple networks can be used to predict new relationship, such as drug screening by combining a drug-gene network, a gene-gene network, and a gene-disease network.

Q10: Provide a different example of how multiple biological networks can be combined to derive new knowledge.

Chromatin organization

Q11: Provide two mechanisms by which changes in chromatin conformation regulate gene expression.

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. Explain the biological importance of these phenomena.

Briefly explain (i) how the following experimental techniques work and (ii) how to interpret their data together with transcriptomics information to identify the biological mechanism(s) that are responsible for the observed changes in gene expression.

Q13: Bisulfite sequencing.

Q14: ATAC-seq.

Q15: ChIP-seq for transcription factor.

Q16: When performing functional enrichment analysis on ChIP-seq data, there are many ways to assign each ChIP-seq peak to nearby gene(s). Why can't we just assign each ChIP-seq peak to the nearest gene? What do you think is a good algorithm to do so?

Q17: Both Hi-C and ChIA-PET captures 3D structural information of the chromatin. How do they differ? What biological information can you learn from one technique but not the other?

Q18: It takes very high sequencing depth to probe 3D structural information of the chromatin (because we need $>> N^2$ reads to quantify all pairwise interactions between N genomic regions). When performing Hi-C or ChIA-PET technique, how can we analyze the data to check that we have obtained enough reads? *Hint: This is a similar situation with metagenomics sequencing.*