Network Analysis - Quiz 1

Q1. What is the difference between *in-degree* and *out-degree* in a directed graph?
The difference between in-degree and out-degree in a directed graph is:

- **In-degree** refers to the number of edges coming into a node.
- **Out-degree** refers to the number of edges going out from a node.

Q2. Define *betweenness centrality* and explain what it reveals about a node's role in a network.
Betweenness centrality measures the extent to which a node lies on the shortest paths between other nodes in the network. A high betweenness centrality indicates that the node serves as an essential bridge or intermediary, controlling communication flow and connecting different parts of the network.

Q3. What is meant by the *diameter* of a graph? How is it different from a *shortest path*?
The **diameter** of a graph is the longest shortest path between any two nodes in the graph. It represents the maximum distance one must traverse to connect two nodes optimally. In contrast, a **shortest path** specifically refers to the minimum number of steps required to reach one particular node from another.

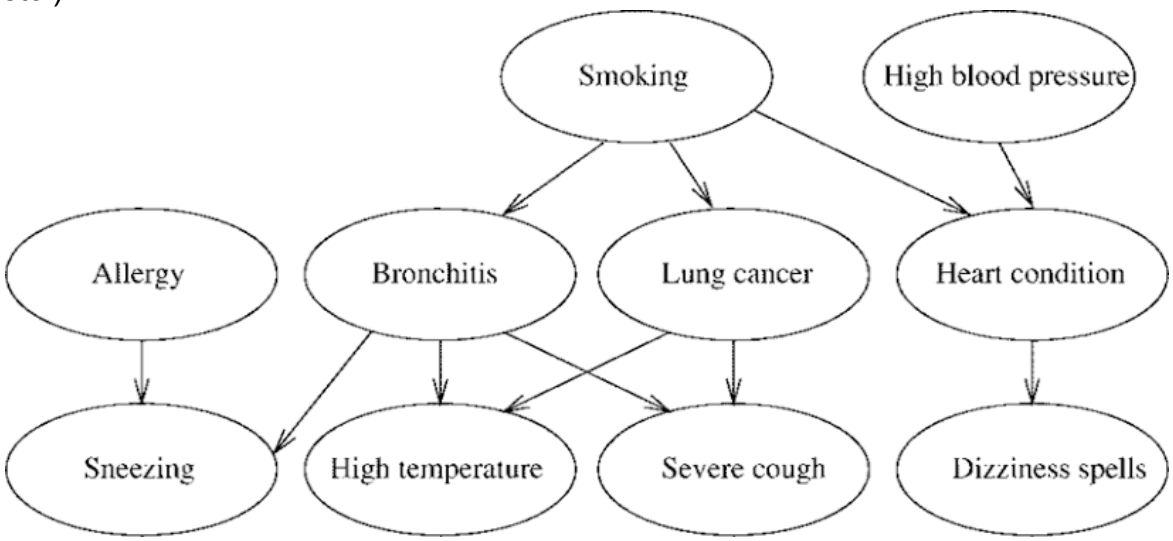
Q4. What is meant by community detection in network analysis? Briefly describe.
Community detection involves identifying groups or clusters of nodes within a network that have a higher density of connections internally compared to externally. It reveals structures, subgroups, or clusters within networks, allowing better understanding of interactions and relationships among nodes.

Q5. **True or False:** If two directed acyclic graphs (DAGs) represent the same set of conditional independencies, then their Bayesian Networks define the same set of joint probability distributions.
True.

Q6. What are the two forms of "learning" required for a Bayesian Network? Briefly describe.
i. Structure Learning: Identifying the network structure (i.e., relationships and dependencies among variables).
ii. Parameter Learning: Estimating conditional probability distributions for each node based on observed data.

For Q7 - Q10

The following network is a Bayesian Network inspired by Santos et al. (1998). **Use abbreviated variables as needed**, and define all symbols used in your responses to Questions 7-10 in a key (e.g., HT = High Temperature, BP = Blood Pressure, etc.).



Abbr.	Variable	Parents (read straight off the arrows in the figure)
SM	Smoking	—
HBP	High Blood Pressure	—
AL	Allergy	—
BR	Bronchitis	SM
LC	Lung Cancer	SM
HC	Heart Condition	SM, HBP
SN	Sneezing	AL, BR
HT	High Temperature	BR, LC
SC	Severe Cough	BR, LC
DS	Dizziness Spells	HC

Q7. Express the joint probability distribution of the network in a product of conditional probabilities (factored form). Ensure all parent-child relationships are captured.
Using the Bayesian-network chain rule, $P(X) = \prod_i P(X_i \mid Parents(X_i))$
the full joint over the ten variables factorises as:
P(SM, HBP, AL, BR, LC, HC, SN, HT, SC, DS) =
P(SM) * P(HBP) * P(AL) *
P(BR | SM) * P(LC | SM) * P(HC | SM, HBP) *
P(SN | AL, BR) * P(HT | BR, LC) * P(SC | BR, LC) *
P(DS | HC)

Q8. List all v-structures in the Bayesian Network. Recall, a v-structure is a configuration of the form X → Z ← Y where X and Y are not directly connected.

Child (collider) Z	Parents X,Y	Why it qualifies
HC	SM, HBP	SM → HC ← HBP and SM is **not** connected to HBP
SN	AL, BR	AL → SN ← BR and AL not connected to BR
HT	BR, LC	BR → HT ← LC and BR not connected to LC
SC	BR, LC	BR → SC ← LC and BR not connected to LC

So there are **four v-structures** in total.

Q9. Assume each variable in the Medical Diagnosis BN is binary. Specify the size of the conditional probability table (CPT) for each of the following variables:
1) High Temperature 2) Severe Cough 3) High Blood Pressure

For a binary child with k binary parents
- **rows** = 2^k parent configurations
- **table entries** = rows × 2 outcomes
- **independent parameters** = rows (because each row sums to 1).

Variable	Parents	k	Rows 2^k	Table entries	Free parameters
High Temperature (HT)	BR, LC	2	4	8	4
Severe Cough (SC)	BR, LC	2	4	8	4
High Blood Pressure (HBP)	—	0	1	2	1

*Note: Column highlighted in green gives the size of CPT for respective variables.

Q10. Based on the structure of the Medical Diagnosis BN:
a) Are *Smoking* and *Dizziness Spells* d-separated given *Lung Cancer*?

For reference of d-separation queries:
• A path is **blocked** if it contains
– a non-collider that is **conditioned**, or
– a collider none of whose descendants are conditioned.
• Otherwise the path is **active**.
• Two nodes are **d-separated** if **every** path between them is blocked.

Sol-10a) Smoking ⊥? Dizziness Spells | Lung Cancer
Active path: **SM → HC → DS**
HC is a non-collider and *not* conditioned; LC is irrelevant to this path.
Hence at least one active path exists ⇒ **Not d-separated**.
Answer: No. Smoking and Dizziness Spells are **not** d-separated given LC.

b) Are *Smoking* and *Dizziness Spells* d-separated given *Heart Condition*?
Sol-10b) Smoking ⊥? Dizziness Spells | Heart Condition
Condition on HC blocks the only simple path SM → HC → DS (HC is a conditioned non-collider).
No alternative route reaches DS without passing through HC.
Answer: Yes. They **are** d-separated given HC.

c) Are *Allergy* and *Severe Cough* d-separated given *High Temperature*?
Sol-10c) Allergy ⊥? Severe Cough | High Temperature
Most obvious route: AL → SN ← BR → SC
• SN is a collider on this path and neither SN nor its descendants are in the conditioning set (HT is *not* a descendant of SN), so the path is blocked.
• No other path avoids that same collider.
Answer: Yes. Allergy and Severe Cough **are** d-separated once we condition on HT (they were already d-separated even without HT).

d) Is there an active path between *Heart Condition* and *Sneezing*?
If yes, provide one such path.
Sol-10d) Active path between Heart Condition and Sneezing?
Yes. One such path is HC ← SM → BR → SN
• SM and BR are **non-colliders** that are not conditioned, so the path is open.
• There is no collider along this route, therefore it remains active.
Answer: Yes. The path HC ← SM → BR → SN is active (no conditioning).