Assignment Week - 1

Social Networks

July 12, 2024

Question 1. What will be the output of the following snippet of code?

```
print(" a + 'b' + c + ' ' + d ")
```

1.) a + "b" + c + ' ' + d

2.) "a + 'b' + c + ' ' + d"

3.) a + 'b' + c + ' ' + d

4.) a + "b" + c + ' ' + d

Correct Answer: Option 3

Explanation 1. Different types of Quotes: If you want to print single quotes, you will have to put them inside double quotes

---

Question 2. Which of the following are valid names for variables in Python?

1.) python_

2.) python

3.) 1python

4.) I python

5.) I python

Correct Answer: Option 1, 2 & 5

Explanation 2. Variable names cannot begin with an integer (Although integers are allowed in variable names, but not in the beginning) and space is not a valid character for a variable name; instead one can use underscore.

Question 3. Farhat wishes to populate a list of 1000 integers, where each integer is drawn at random from the range 10 to 50 (both end points are included). Help Farhat in choosing the right snippet of code.

1.) import random as rand
Integers = [rand.randint(10,51) for i in range(1000)]

2.) import random as rand
Integers = [rand.randint(10,50) for i in range(1001)]

3.) import random as rand
Integers = [rand.random(10,50) for i in range(1001)]

4.) import random as rand
Integers = [rand.random(10,51) for i in range(1000)]

5.) import random as rand
Integers = [rand.randint(10,50) for i in range(1000)]

6.) import random as rand

```
Integers = [] for i in
range(1000):
    Integers.append(rand.randint(10,50))
```

Correct Answer: Option 5 & 6

Explanation 3. randint is the function used to generate random integers. The function takes 2 arguments: (start of the range, end of the range) with both end points included. The option 5 uses list comprehension which produces the same result as option 6.

Question 4. Code (a):

```
def black_box(O):       #Argument is a dictionary return([(i,O[i]) for i in O])
```

Code (b):

```
def blackest_box(O):  #Argument is a list of tuples of two values return({i[0]:i[1] for i in O})
```

1.) Code (a) returns a tuple with (value, key) pair of the dictionary provided.
Code (b) returns a dict with (Second value of tuple as key : First value of the tuple as value) for the provided list of tuples.

2.) Code (a) returns a tuple with (value, key) pair of the dictionary provided.
Code (b) returns a dict with (First value of tuple as key : Second value of the tuple as value) for the provided list of tuples.

3.) Code (a) returns a tuple with (key, value) pair of the dictionary provided.
Code (b) returns a dict with (Second value of tuple as key : First value of the tuple as value) for the provided list of tuples.

4.) Code (a) returns a tuple with (key, value) pair of the dictionary provided.
Code (b) returns a dict with (First value of tuple as key : Second value of the tuple as value) for the provided list of tuples.

Correct Answer: Option 4

Explanation 4. function black box accepts a dictionary as an argument and returns a list of tuples with key, value pairs of the dictionary and uses comprehension for returning the value. function blackest box accepts a list of tuples with two values as argument and returns a dictionary with key being the first value of the tuples inside the list and value being the second value of the tuples.

Question 5. Predict the output of the following code:

```
L = [1, 3, -1, 4, -2, 5, 3]

try:
    n=10 for i in range(n):
        if L[i] < 0:
            L[i] = 0 except
IndexError:
    for i in range(n - len(L)):
        L.append(0)

finally:
    print(L)
```

1.) Output = [1, 3, -1, 4, -2, 5, 3]

2.) Output = [1, 3, 0, 4, 0, 5, 3]

3.) Output = [1, 3, -1, 4, -2, 5, 3, 0, 0, 0]

4.) Output = [1, 0, -1, 0, -2, 0, 3]

5.) Output = [0, 3, 0, 4, 0, 5, 0]

6.) This code is wrong

Correct Answer: ~~Option 3~~ **[1, 3, 0, 4, 0, 5, 3, 0, 0, 0]**

Correct Answer is missing in the given options.

Explanation 5. The function replaces all the integer values inside the list L with zeroes if they are less than zero and it appends zero to the list if the code gives an IndexError, since the length of the list is less than range used in the code. In the end, finally always gets executed and hence, the code prints the list.

Question 6. Predict the output of the following code:

```
D = {'mom' : True, 'malayalam' : True, 'dad' : True, 'work' : False} L = ['mom', 'dad', 'non',
'work']
```

```
for word in L: try:
        if D[word]:
            print('This is a Palindrome')
        else: print('This is not a Palindrome')
    except: print('This key is not present in the dict')
```

1.) This is a Palindrome This is a Palindrome This is a Palindrome

2.) This key is not present in the dict This key is not present in the dict
This key is not present in the dict

3.) This is a Palindrome This is a Palindrome
This key is not present in the dict

4.) This is a Palindrome This is a Palindrome
This key is not present in the dict This is not a Palindrome

5.) This is a Palindrome This is a Palindrome
This is not a Palindrome

Correct Answer: Option 4

Explanation 6. Both 'mom' and 'dad' are palindromes and are present in the dictionary as keys. While 'non' is not present in the dictionary as key, despite being a palindrome and 'work' is not a palindrome despite being a key inside the dictionary.

Question 7. Farhat has a list of topper students and he wants to select a student randomly from the list to enlist in a top secret project he is working on. Help Farhat choose the function which can help him do it.

1.) random.sample(students,1)

2.) random.choice(students)

3.) random.choices(students)

4.) random.randint(0, len(students) - 1)

5.) None of the above

Correct Answer: Option 2

Explanation 7. random.choices is the function in the random library which helps one select a random element from a data structure.

Question 8. Gogo has created a list of connections in his classroom. Each connection is represented as a tuple of two individuals who are friends. Gogo needs to create an undirected graph and find the number of nodes and edges in the graph.

```
Connections = [('Jagirat','Jatin'), ('Jagirat', 'Ashutosh'), ('Jatin', 'Gitansh'),
('Ashutosh','Gitansh'), ('Gitansh','Nishit')]

import networkx as nx def
class_network(Connections):
    G = nx.graph()
    G.add_edges_from(Connections) num_nodes =
G.number_of_nodes() num_edges =
G.number_of_edges() return num_nodes,
num_edges print(class_network(Connections))
```

1.) (4,5)

2.) (5,4)

3.) (5,6)

4.) (6,5)

5.) (5,5)

6.) (4,4)

7.) (6,6)

8.) (4,6)

Correct Answer: Option 5

Explanation 8. There are 5 nodes in the graph: Jagirat, Jatin, Ashutosh, Gitansh, Nishit and there are 5 edges between these 5 nodes: ('Jagirat','Jatin'), ('Jagirat', 'Ashutosh'), ('Jatin', 'Gitansh'), ('Ashutosh','Gitansh'), ('Gitansh','Nishit')

Question 9. Aashik wants to create a random graph with 10 nodes with the probability of an edge being present as 0.35. Which function can he choose?

1.) nx.random graph(10,0.35)

2.) nx.gnp _random graph(10,0.35)

3.) nx.erdos renyi graph(10,0.35)

4.) Both (b) and (c)

Correct Answer: Option 4

Explanation 9. Both nx.gnp random graph(n, p) and nx.erdos _renyi graph(n, p) functions of the networkx library can be used to create a graph with some 'n' nodes and probability of an edge being present being 'p'.

Question 10. Aashik now wants to create a directed graph and add edges to it. Help Aashik with choosing the right function.

1. nx.graph() & graph.add edge()

2. nx.Digraph() & graph.add edge()
3. nx.Multigraph() & graph.add edge from()
4. nx.MultiDigraph() & graph.add edges from()

Correct Answer: Option 2

Explanation 10. You will have first import networkx. To create a graph, you will have to use the function nx.graph() and then to add edges to the graph, you will have to use graph.add edge() function.

Question 11. Now, Aashik wants to see the graph he has created, Which of the folllowing options can help him with it?

1.) nx.draw(graph) & plt.show()
2.) nx.plot(graph) & plt.show()
3.) nx.display(graph)
4.) nx.draw(graph) & plt.visualize()

Correct Answer: Option 1

Explanation 11. You will have first import 2 libraries: matplotlib and networkx. To draw a graph, you can use the function nx.draw(graph) and plt.show() to show the said graph.

Question 12. Aashik wants to check if the first random graph he created was connected or not. Which function will he use now?

1.) nx.is connected(graph)
2.) nx.connected(graph)
3.) nx.isconnected(graph)
4.) nx.is fully connected(graph)

5.) nx.is complete(graph)

Correct Answer: Option 1

Explanation 12. nx.is connected(Graph) function of the networkx library is the function that will help you check if your graph is a connected graph or not.

Question 13. Ashutosh created a social network graph of his class and wants to find out who is the person he can be friends with very easily. Which algorithmic concept can he use to do this?

1.) Page Ranking

2.) BFS

3.) DFS

4.) Link Prediction

Correct Answer: Option 4

Explanation 13. Link Prediction is the concept that helps you recommend a potential friend to a person based on the friends network they have. This is the same concept that is used in friend recommendations in social network sites such as facebook, linkedin, instagram etc.

Question 14. Sundari has 55 students in his class and he wants to create a network graph for the class. What is the maximum number of graphs possible for the given class.

1.) 2^{55}

2.) ${}_{55}C_2$

3.) $2_{{}_{55}C_2}$

4.) 55^2

Correct Answer: Option 3

Explanation 14. We have 55 nodes in the graph. To make a edge we require 2 nodes, therefore, we can have ${}^{55}C_2$ possible edges in the graph. But either an edge can be present or it can be absent, hence we have 2 choices for each edge and we have ${}^{55}C_2$ such decisions to make, whether to keep that edge or not. Therefore, the answer is $2^{{}^{55}C_2}$.

SOCIAL NETWORKS - NPTEL July 2024

Week 2

1. Which of the following statements is/are True?

   Statement I - Web graph is a directed network.
   Statement II - Facebook friendship network is an undirected network.

   (a) I only
   (b) II only
   (c) Both
   (d) None

   Answer: (c)

   Explanation: Statement I: The web graph is a directed network because hyperlinks between web pages have direction (from one page to another).

   Statement II: The Facebook friendship network is an undirected network because friendships are mutual (if A is friends with B, then B is also friends with A).

2. What is the clustering coefficient of a node that has 6 neighbors and 3 connections between those neighbors?

   (a) 0.2
   (b) 0.5
   (c) 0.75
   (d) 0.9

   Answer: (a)

   Solution: For a node with 6 neighbors, there can be a maximum of (6 choose 2) = 15 possible connections among those neighbors. If 3 out of those 15 connections actually exist, the clustering coefficient is calculated as the ratio of the actual connections to the maximum possible connections, which is $3/15 = 0.2$.

3. Name the method used to read dataset in 'txt' format.

   (a) $read\_gml()$
   (b) $read\_edgelist()$
   (c) $read\_txt()$
   (d) $read\_gexf()$

   Answer: (b)

4. Given a complete graph with 120 vertices, what is the diameter of the Graph?

   (a) 0
   (b) 1
   (c) 2
   (d) 3

   Answer: (b)

   Explanation: In a complete graph, every vertex is directly connected to every other vertex by a single edge. Therefore, the shortest path between any two vertices is always 1, making the diameter of the graph 1.

5. Which statement accurately reflects the characteristics of node degrees according to Power law?

   (a) every individual in a social network has an equal number of connections
   (b) exhibit a uniform distribution of connections among all users
   (c) a small number of individuals have a substantially higher number of connections compared to the majority
   (d) each node has an identical degree, promoting equality in connectivity.

   Answer: (c)
   Lecture: 21

6. Given is a graph G with $|V| = n$ number of nodes and $|E|$ number of edges. In which of the following cases, we can guarantee that G is connected?

   (a) $|E| = n$
   (b) $|E| = n - 1$
   (c) $|E| = n(n-1)/2$
   (d) $|E| = n \log_2 n$

   Answer:(c)
   Lecture 25-28:58
   Solution: A graph with n(n-1)/2 edges is a complete graph and hence connected. With log n edges, we can't always guarantee connectedness, though the graph will be connected with a high probability.

7. Which of the following statements is True for GML format of networks?
   Statement I: Labels and attributes can be added
   Statement II: Weights can be added

   (a) I only
   (b) II only
   (c) Both
   (d) None

   Answer:(c)
   Lecture: 19

8. What is the reason for a path between words like "love" and "hatred" in the synonymy network?

   (a) faulty edges
   (b) contradictory paths to find antonyms
   (c) The network algorithm identifies unrelated words as synonyms
   (d) Words can undergo semantic shifts, acquiring new meanings or evolving to represent opposite concepts

   Answer: (d)
   Lecture: 16

9. Given a graph with 5 nodes and 8 edges, find the density of the graph. Hint: Answer with two digits precision
   Answer: 0.80
   Solution: Density of the graph = Number of edges/ Maximum possible edges for the graph =8/5C2 = 8/10=0.80

10. For any vertex v in an undirected (without loop, multiple edges) , the clustering coefficient of v ranges from:
    A. -1 to +1
    B. 0 to 1
    C. $-\infty$ to $+\infty$
    D. 0 to $+\infty$

    Answer: (b)
    Solution: If there are no edges between the neighbors of the vertex v, then cc(v) = 0. On the other hand, if all its neighbors are connected to each other, then cc(v) = 1.

SOCIAL NETWORKS - NPTEL JULY 2024

Week 3

1. Why do some individuals have lower clustering coefficient?

   (a) They prefer solitude

   (b) They have no friends at all

   (c) Their friends are from different circles

   (d) They are too busy to socialize

   Answer: (c)

   Solution: Some individuals have lower clustering coefficients in a friendship network because they connect friends from different social circles who do not know each other, have fewer overall connections, or have more random and dispersed friendship patterns, leading to fewer mutual friends

2. Which of the following best describes betweenness centrality of a node?

   (a) All the paths between the given node and the highest degree node.

   (b) All the shortest paths that pass through the given node.

   (c) All the paths that pass through the given node.

   (d) All the shortest paths between the given node and the highest degree node.

   Answer: (b)

   Solution: Betweenness centrality is a measure of centrality in a graph based on shortest paths. For every pair of nodes in a connected graph, there exists at least one shortest path between the nodes. The betweenness centrality for each node is the number of these shortest paths that pass through the node.

3. What is the neighbourhood overlap of an edge connecting B and C from the below given graph?

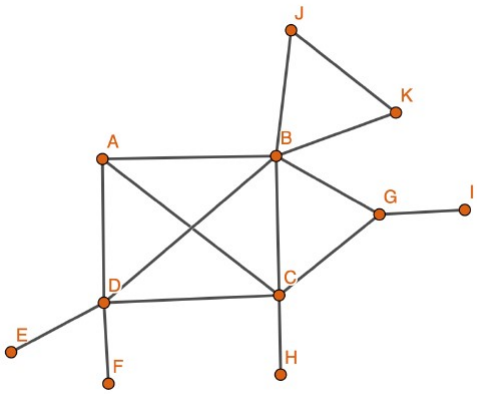


Figure 1: Graph H

   (a) 1/2

   (b) 2/5

   (c) 1/6

   (d) 1/4

Answer: (a)

Solution: Neighbourhood overlap of an edge connecting A and B is defined as the ratio of number of nodes who are neighbors of both A and B to the number of nodes who are neighbors of at least one of A and B. In the given graph, the ratio is, evidently, 1/2 for an edge connecting B and C.

4. Compute the embeddedness of the edge AB in the graph H in Figure 1.

   (a) 0
   (b) 1
   (c) 2
   (d) 3

   Answer: (c)

   Solution: Embeddedness of an edge in a network to be the number of common neighbors the two endpoints have

5. Assume the neighborhood overlap of an edge 'e' is 0, then 'e' is

   (a) local bridge
   (b) strong tie
   (c) an edge with high betweenness
   (d) edge to be removed first in Girvan Newmann algorithm

   Answer: (a)

   Solution: A local bridge is an edge connecting two nodes with no common neighbors, meaning its removal would increase the shortest path length between those nodes. Hence, low neighborhood overlap typically signifies that an edge plays a crucial role in connecting different parts of the network, acting as a local bridge.

6. Consider a large social network where we have two communities C1 and C2 that are connected by only through two nodes P and Q. They exhibit a property called

   (a) Neighborhood overlap
   (b) Triads
   (c) structural holes
   (d) Embeddedness

   Answer: (c)

   Solution: Structural holes refer to gaps or "holes" in a network where certain nodes bridge or connect otherwise disconnected parts of the network. In this case, nodes P and Q act as bridges between the two communities C1 and C2, highlighting the concept of structural holes.

7. Which of the following conditions is typically used to identify an ideal community within a network?

   (a) High edge density within the community and low edge density with nodes outside the community
   (b) Low clustering coefficient within the community and high clustering coefficient with nodes outside the community
   (c) High degree centrality within the community and low degree centrality with nodes outside the community
   (d) High neighborhood overlap within the community and high degree centrality with nodes outside the community

   Answer:(a)

   Solution: An ideal community in a social network is often characterized by a high density of connections among its members (high edge density within the community) and relatively few connections between its members and those outside the community (low edge density with nodes outside the community). This ensures that the community is tightly-knit internally while being isolated from other communities.

8. Alex starts a new job and establishes connections with both his new colleagues and an external network of industry professionals. What role/property of social networks is exemplified by Alex in this situation?

(a) Structural hole
(b) Brokerage
(c) Both
(d) None

Answer:(c)

Solution: Alex connects two different groups (new colleagues and external industry professionals) that may otherwise be disconnected, thereby filling a structural hole. Alex's role in linking these two groups demonstrates brokerage, as he facilitates connections between otherwise separate networks. Thus, Alex's role exemplifies both structural holes and brokerage.

9. According to Girvan Newman method, the most vital edge for connecting different regions of the network is the node with ——— betweenness.

(a) zero
(b) low
(c) high
(d) medium

Answer: (c)

Solution: The Girvan-Newman method is used for community detection in networks by iteratively removing edges with the highest betweenness centrality. Betweenness centrality measures how often an edge lies on the shortest path between pairs of nodes. Edges with high betweenness are considered crucial for connecting different parts of the network. By removing these edges, the method aims to reveal the underlying community structure of the network.

10. For each pair of nodes A and B in the graph, the flow between A and B divides itself evenly along all the possible shortest paths from A to B: so if there are k shortest paths from A and B, then 1/k units of flow pass along each one. Here we define the betweenness of an edge to be the total amount of flow it carries, counting flow between all pairs of nodes using this edge. Which edge will be removed first based on Girvan Newman method?

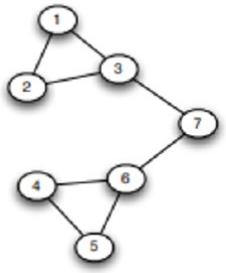


Figure 2: Graph H

A. 6-7
B. 1-3
C. 4-6
D. 6-5

Answer: (a)

Solution: The 6-7 edge carries the full unit of flow from each node among 4, 5, and 6 to each of the other nodes 1,2,3, and 7. Thus the betweenness is 3.4=12 which is the highest.

SOCIAL NETWORKS - NPTEL JULY 2024

Week 4

1. Which of the following is an example of homophily?

   (a) A diverse team working together to solve a problem

   (b) People of the same ethnicity forming a community group

   (c) Randomly assigned roommates becoming friends

   (d) A mentor-mentee relationship between individuals from different backgrounds

   Answer: (b)

   Explanation: Homophily refers to the tendency of individuals to associate and bond with others who are similar to themselves in various ways, such as ethnicity, interests, values, or social status. Individuals are forming a group based on a shared characteristic—ethnicity. This exemplifies how people tend to connect with others who are similar to themselves.

2. Which of the following best describes the concept of selection in homophily?

   (a) The tendency of people to change their beliefs to fit those of their social group

   (b) The process by which individuals choose to associate with others who are similar to themselves

   (c) The influence of group norms on individual behaviors

   (d) The random formation of social ties without regard to similarity

   Answer: (b)

   Explanation: Selection in homophily refers to the process where individuals actively choose to form connections with others who share similar characteristics, such as interests, values, or backgrounds.

3. suppose we have a network in which a p fraction of all individuals are male, and a q fraction of all individuals are female. If the fraction of cross-gender edges is significantly less than 2pq, then there is evidence for homophily. For a network to have a fraction of cross-gender edges that is significantly more than 2pq, the network exhibits

   (a) inverse homophily

   (b) triadic closure

   (c) membership closure

   (d) foci closure

   Answer: (a)

   Explanation: The network of romantic relationships is a clear example of this where most of the edges are cross-gender.

4. Suppose Ram and Chander have k common friends. Given that each common friend gives Ram and Chander an independent probability p of forming a link, what is the probability of not forming a friendship between Ram and Chander because of k common friends?

   (a) p^k

   (b) $(1-p)$

   (c) $(1-p)^k$

   (d) $1-(1-p)^k$

Answer: (c)

Explanation:
Probability of forming a link due to one of the common friends = p
Probability of a link not forming due to one of the common friends = $1 - p$
Probability of not forming a link due to all the 'k' common friends =$(1 - p)^k$
probability of forming a link due to atleast one of the 'k' common friends =$1 - (1 - p)^k$

5. Two friends Simi and Niraj have taken different courses according to their interests. Simi has completed 10 courses and Niraj has completed 5 courses in all. There are 3 courses that Simi and Niraj have taken up in common. What is the similarity measure for Simi and Niraj?

   (a) $\frac{1}{2}$
   (b) $\frac{3}{4}$
   (c) $\frac{1}{4}$
   (d) $\frac{2}{3}$

   Answer: (c)

   Solution: If A is the set of courses completed by Simi and B is the set of courses completed by Niraj, then similarity measure for Simi and Niraj is given by
   $SimilarityMeasure = \frac{A \cap B}{A \cup B} = \frac{3}{12} = \frac{1}{4}$

6. Identify the mechanism in play for the following context: 'Karate club introduces Anna to Daniel'

   (a) Triadic closure
   (b) Focal closure
   (c) Membership closure
   (d) Neighborhood overlap

   Answer: (b)

   Solution: Focal Closure is the tendency of two people to form a link when they have a focus in common. Here, Karate club is the focus that is common to Anna and Daniel which leads to a friendship.

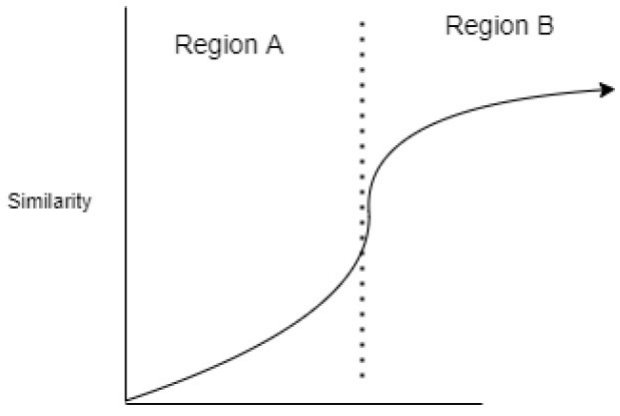
7. Identify the mechanism in play for the following context: 'Anna introduces Bob to Sam'

   (a) Triadic closure
   (b) Focal closure
   (c) Membership closure
   (d) Neighborhood overlap

   Answer: (a)

   Explanation: Triadic closure is a social network phenomenon where if two individuals (Bob and Sam) are introduced to each other by a mutual friend (Anna), a triad is formed in the network.

8. Given the average similarity of two editors on Wikipedia, relative to the time (0) represented by the dotted line at which they first communicated. Time, on the x-axis, is measured in discrete units, where each unit corresponds to a single Wikipedia action taken by either of the two editors. The rapid increase in similarity before first contact is due to



(a) Selection

(b) Social Influence

(c) Tradic closure

(d) Focal closure

Answer:(a)

Explanation: This increase is due to selection, where editors who are already similar are more likely to start communicating with each other.

9. In a scenario, if there were two friends and one person is a part of a Dance Crew and he influences the other to take part in it this will lead to

(a) Triadic closure

(b) Focal closure

(c) Membership closure

(d) Selection

Answer: (c)

Explanation: membership closure occurs because the existing member (one of the two friends) influences the other (the second friend) to also join the Dance Crew, leading to a situation where both friends are part of the same group. This strengthens the ties within the group and aligns the memberships of individuals who are already connected through friendship.

10. Consider an evolving network; evolving in the following two ways.
A: With time, new edges are being formed in the network.
B: Also, with time, nodes in the network are changing their attributes.
A and B represent

(a) Homophily and social influence

(b) Triadic closure and social influence

(c) Homophily and focal closure

(d) Triadic closure and membership closure

Answer: (a)

Explanation: A: With time, new edges are being formed in the network. This evolution could represent homophily. In the context of homophily, new edges or connections in the network are often formed between nodes that are similar to each other. As nodes with similar attributes or behaviors connect over time, new edges are more likely to form between them.

B: Also, with time, nodes in the network are changing their attributes. This evolution could represent social influence. Social influence occurs when nodes (individuals) change their attributes or behaviors in response to interactions with other nodes. As individuals are influenced by their connections, they may modify their attributes to align more closely with their peers.

SOCIAL NETWORKS - NPTEL JULY 2024

Week 5

1. Let - denote friendship and + denotes enmity relationship in a social network. Select all the options that represent stable triangular relationship structures?

   (a) $+++$

   (b) $++-$

   (c) $+--$

   (d) $---$

   Answer: (c) and (d)

   Explanation:

   In this context, "-" denotes friendship and "+" denotes enmity. According to structural balance theory, a triangular relationship is stable if the relationships result in a positive product. The stable structures are +–, where two friends share a common enemy (a balanced situation), and —, where all three are friends. The structures +++ (all enmity) and ++- (two enemies and a friend) are unstable because they create internal tensions that violate balance theory.

2. It is observed that a friend's friend often becomes a friend, and an enemy's enemy often becomes an ally. What are the reasons behind these phenomena?

   (a) Social influence and clustering, respectively

   (b) Social influence and structural balance, respectively

   (c) Triadic closure and structural balance, respectively

   (d) Triadic closure and clustering, respectively

   Answer: (c)

   Explanation:

   Triadic closure: This concept explains why a friend's friend tends to become a friend. It suggests that if two people have a common friend, they are more likely to become friends themselves to close the triad and form a stronger, more interconnected network.

   Structural balance: This concept explains why an enemy's enemy tends to become an ally. Structural balance theory posits that social networks evolve towards a state where relationships are balanced, meaning that the enemy of my enemy is likely to be my ally to maintain consistency and balance in social relationships.

3. Which of the following four node Network is balanced?

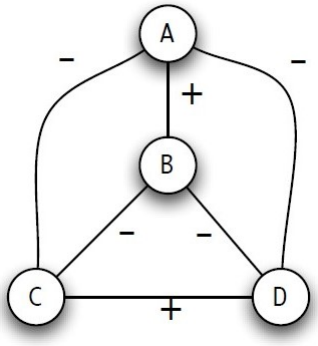


Figure 1: Network A

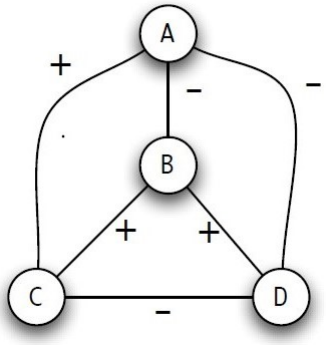


Figure 2: Network B

(a) A only
(b) B only
(c) Both A and B
(d) Neither A nor B

Answer: (a)

Explanation: The Network A is balanced because each group of three nodes satisfies the Structural Balance Property described earlier. Conversely, Network B is not balanced, as the three nodes A, B, and C have exactly two positive edges, which violates the Structural Balance Property.

4. Given the two dimensional grid in Figure 3, comment if the node in the center of grid A and grid B are satisfied respectively, given threshold t=5.

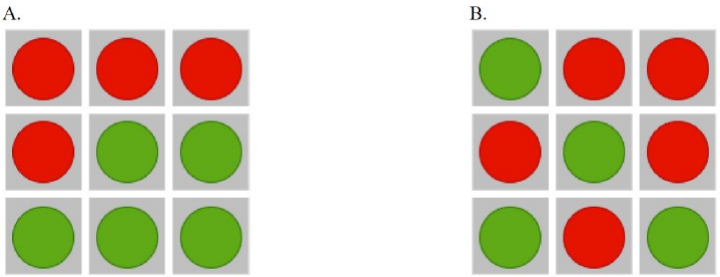


Figure 3: Schelling's Model

(a) satisfied, satisfied
(b) unsatisfied, satisfied

(c) satisfied, unsatisfied

(d) unsatisfied, unsatisfied

Answer: (d)

Explanation: Both central nodes in grids A and B are unsatisfied because, in each case, the number of neighboring nodes that are similar to the central node is less than the threshold $t = 5$

5. For a graph with seven nodes, what is the maximum number of possible triangles in the graph?

   (a) 20

   (b) 30

   (c) 35

   (d) 50

   Answer: (c)

   Solution: In a graph having n, nodes, there can be nchoose3 triangles, which is equal to $n(n-1)(n-2)/6$

$$\binom{7}{3} = \frac{7\,!}{3\,!(7-3)\,!} = 35$$

6. Which of the following is an invalid threshold value for the two dimensional Schelling model grid?

   (a) 1

   (b) 2

   (c) 4

   (d) 9

   Answer: (d)

   Solution: In a 2 dimensional grid, each individual has a maximum of 8 neighbours. This means that the tolerance can vary between 0 to 8.

7. The number of internal, corner and boundary nodes in a Schelling grid of dimension 10 × 10 are

   (a) 64, 4 and 36 respectively

   (b) 81, 0 and 36 respectively

   (c) 81, 4 and 32 respectively

   (d) 64, 4 and 15 respectively

   Answer: (a)

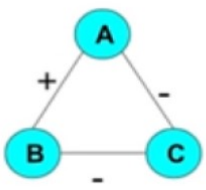
   Explanation:

   These are nodes that are not on the boundary or corner of the grid. Number of internal nodes=$(10-2)(10-2) = 88 = 64$

   These are nodes located at the four corners of the grid. For any grid, there are always exactly 4 corner nodes.

   These nodes are located on the edges of the grid but not at the corners.
   Number of boundary nodes=$4(10-2)+4 = 36$

8. Which social belief does the following signed triad represent?



(a) A friend's friend is an enemy

(b) A friend's friend is a friend

(c) An enemy's enemy is a friend

(d) An enemy's enemy is an enemy

Answer:(c)

Explanation: An enemy's enemy is a friend: This belief is a direct application of balance theory, which states that if two people are enemies, their mutual enemy should be a friend, leading to a balanced network.

9. If a complete graph can be divided into two sets of mutual friends, with complete mutual antagonism between the two sets, then it is said to be

(a) balanced

(b) unbalanced

(c) connected

(d) symmetric

Answer: (a)

10. Which of the following statements are True for a graph to exhibit structural balance?

Statement I: A labelled graph where all the nodes are friends with each other.

Statement II: A labelled graph where the nodes can be divided into two groups, X and Y , such that every pair of nodes in X like each other, every pair of nodes in Y like each other, and everyone in X is the enemy of everyone in Y

(a) I only

(b) II only

(c) Both I and II

(d) None

Answer: (c)

Explanation: Both statements I and II are True for a graph to exhibit structural balance. Here's why:

Statement I: A labelled graph where all the nodes are friends with each other.

This statement describes a completely positive cycle. In terms of structural balance, this is a balanced scenario because every relationship is friendly, and there are no negative ties to cause imbalance. Therefore, this scenario satisfies the condition for structural balance.

Statement II: A labelled graph where the nodes can be divided into two groups, X and Y, such that every pair of nodes in X like each other, every pair of nodes in Y like each other, and everyone in X is the enemy of everyone in Y.

This describes a graph that can be partitioned into two sets of nodes where intra-group relationships are positive (friends), and inter-group relationships are negative (enemies).

SOCIAL NETWORKS - NPTEL July 2024

Week 6

1. In the context of a web graph, which of the following statements best describes the significance of directed links between web pages?

   (a) Directed links indicate the physical location of web pages on a server.

   (b) Directed links represent the number of visits each web page receives.

   (c) Directed links show the navigation paths from one web page to another.

   (d) Directed links determine the color and layout of web pages.

   Answer: (c)

2. Consider the implementation of Page Rank using Random Walk. Which of the following is TRUE?

   Statement I - Random walk points is higher for the nodes with higher number of inlinks
   Statement II - Random walk point is decreased by one when we reach a node

   (a) I only

   (b) II only

   (c) Both I and II

   (d) Neither I nor II

   Answer: (a)

   Explanation:

   Statement I: Random walk points are higher for the nodes with higher number of inlinks. This is true. In the PageRank algorithm, a node's rank is influenced by the number of inlinks (incoming links) it has. Nodes with more inlinks are more likely to be visited during a random walk, thereby accumulating more "points" or a higher rank.

   Statement II: Random walk point is decreased by one when we reach a node.

   This is not true. When a random walk reaches a node, the point for that node is not decreased by one. Instead, when a random walk reaches a node, the random walk point for that node is increased because it indicates another visit to that node, contributing to its overall rank.

3. In a game where participants Alex, Sam, and Jamie distribute gold coins equally, it is observed that the game converges. At the convergence state, which of the following statements is most accurate?

   (a) Alex, Sam, and Jamie always end up with an equal number of gold coins.

   (b) Alex, Sam, and Jamie may or may not have an equal number of gold coins.

   (c) The total number of gold coins with Alex, Sam, and Jamie increases as the game converges.

   (d) The game never converges if Alex, Sam, and Jamie distribute the gold coins equally.

   Answer: (b)

   Explanation: When people distribute gold coins equally, the game converges. At the convergence state, everybody might/might not have the same number of gold coins.

4. Which type of node could potentially cause an issue in the random walk (drop) gold coins distribution game?

   (a) node with highest indegree

   (b) node with least indegree

   (c) node with highest outdegree

(d) node with zero outdegree

Answer: (d)

Explanation: A node with zero outdegree has no outgoing edges, meaning it cannot transfer gold coins to any other node. This creates a significant problem in the random walk (drop) gold coins distribution game. When gold coins reach a node with zero outdegree, they become trapped there since the node has no way to pass them on to other nodes. This accumulation of coins at a single node disrupts the intended continuous flow and redistribution of coins among all nodes. As a result, the game may fail to achieve a balanced or stable distribution of coins, preventing proper convergence.

5. Examine the graph depicted in the following figure. Each circle contains a number representing the amount of gold coins the corresponding node holds. Determine the number of gold coins each node will have after one iteration of the equal sharing gold coins game.

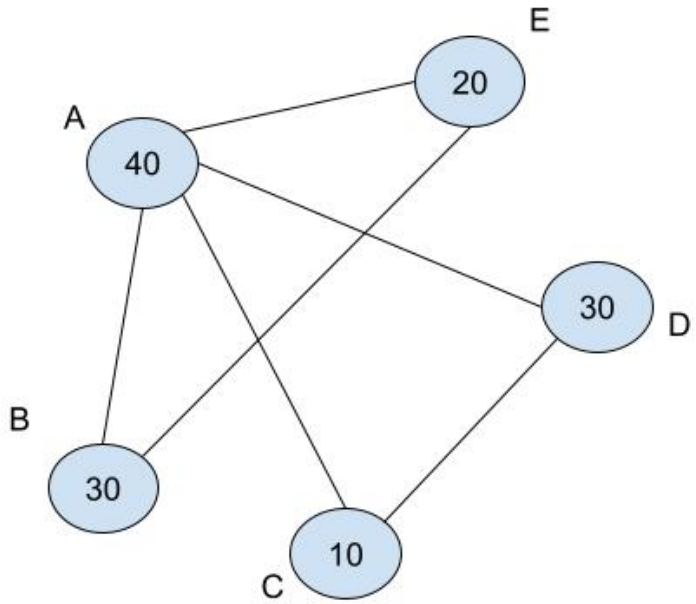


Figure 1: Network H

(a) A: 45, B: 20, C: 25, D: 15, E: 25

(b) A: 45, B: 20, C: 25, D: 15, E: 5

(c) A: 45, B: 25, C: 25, D: 15, E: 25

(d) A: 45, B: 20, C: 25, D: 25, E: 25

(e) A: 45, B: 20, C: 25, D: 15, E: 20

Answer: (a)

Explanation:

P'(A) = P(B)/2 + P(C)/2 + P(D)/2 + P(E)/2 = 15 + 5 + 15 + 10 = 45

P'(B) = P(A)/4 + P(E)/2 = 10 + 10 = 20

P'(c) = P(A)/4 + P(D)/2 = 10 + 15 = 25

P'(D) = P(C)/2 + P(A)/4 = 5 + 10 = 15

P'(E) = P(A)/4 + P(B)/2 = 10 + 15 = 25

6. Which of the following best describes the primary concept behind the Google PageRank algorithm?

   (a) The algorithm ranks pages based solely on the relevance of their content to the search query, using keywords and semantic analysis

   (b) The algorithm ranks pages by evaluating the quantity and quality of backlinks, distributing rank scores iteratively based on link structure and transition probabilities

   (c) The algorithm ranks pages by analyzing user engagement metrics such as click-through rates, time spent on page, and bounce rates

   (d) The algorithm ranks pages based on their presence and influence on social media platforms, considering factors like likes, shares, and comments

   Answer: (b)

   Explanation:

   Option (a) is incorrect because PageRank does not rank pages solely based on content relevance.

   Option (b) is correct as it accurately describes the primary concept behind PageRank, which involves analyzing backlinks and iteratively distributing rank scores based on link structure and transition probabilities.

   Option (c) is incorrect because PageRank does not incorporate user engagement metrics as its primary ranking factor.

   Option (d) is incorrect because PageRank does not rank pages based on social media influence

7. What is the relationship between indegree and page rank values for the mentioned citation network?

   (a) Indegree is not correlated with page rank values

   (b) indegree is directly proportional to page rank

   (c) indegree is inversely proportional to page rank values

   (d) None of the above

   Answer: (a)

   Explanation: While indegree (the number of incoming links to a node) can influence the PageRank of a node, it is not the sole determinant of PageRank. PageRank values are calculated based on the probability distribution of a random walker visiting nodes, which depends not only on the number of incoming links but also on the quality and rank of the linking nodes. Additionally, PageRank considers the distribution of rank scores through outgoing links.

   Thus, even though there can be a correlation between indegree and PageRank (nodes with higher indegree often tend to have higher PageRank), this correlation is not strict. High indegree does not guarantee high PageRank if the incoming links are from low-ranked nodes, and low indegree does not necessarily mean low PageRank if the incoming links are from highly ranked nodes. Therefore, indegree and PageRank values are not perfectly correlated.

8. Which of the following correctly depicts teleportation?

   (a) Jumping from the current node to its neighbor's neighbor.

   (b) Going back to the previous node which was explored.

   (c) Jumping to any random node in the network

   (d) Jumping to the node in the network which has maximum outdegree.

   Answer: (c)

   Explanation: It has been shown in the lecture videos that once the random walk algorithm gets stuck/trapped at a node, it jumps to a randomly chosen node in the network. This concept is known as Teleportation.

9. What happens if one or more nodes have no outlink while implementing the page rank algorithm?

   (a) Points in every node in the network keeps increasing without bounds

   (b) Points get accumulated in such nodes with no outlinks

(c) Points get to reach zero in the nodes with zero outlinks

(d) Points in every node in the network converges in exactly one iteration

Answer: (b) Explanation: In the PageRank algorithm, nodes with no outlinks are referred to as "dangling nodes." When a random walker reaches a dangling node, it has nowhere to go.This causes accumulation of points in nodes with no outlinks.

10. Identify the correct sequence of implementing page rank using points distribution method repeating this process until convergence.

I. Assign 100 points to each of the n nodes

II. Get ranking based on the points accumulated

III. Create a directed graph with n nodes

IV. Distribute points

(a) I, II, IV, III

(b) II, I, IV, III

(c) III, I, IV, II

(d) III, II, I, IV

Answer: (c)

Explanation: The correct sequence for implementing the PageRank algorithm using the points distribution method involves:

Creating the directed graph (Step III): This is the initial step where the structure of the network is defined, including the nodes and the directed edges between them.

Assigning initial points (Step I): Each node is given an initial value of 100 points (or any arbitrary value) to start the process.

Distributing points (Step IV): Points are distributed from each node to its outlinks according to the algorithm, simulating the random walk process. This step is repeated iteratively.

Getting the ranking (Step II): After points are distributed and the system reaches convergence (where the points distribution stabilizes), the final ranking of the nodes is determined based on the accumulated points.

SOCIAL NETWORKS - NPTEL July 2024

Week 7

1. Assume that actions X and Y yield every player a payoff of 5 and 10 respectively. Further assume that there are two friends, Alex and Blake. Alex decides to adopt action X while Blake decides to adopt action Y. What are the payoffs that they get?

   (a) 5,10
   (b) 5,0
   (c) 10,5
   (d) 0,10

   Answer: (a) Explanation: In the scenario provided: Action X yields a payoff of 5 to the player who chooses it. Action Y yields a payoff of 10 to the player who chooses it. Given that: Alex chooses Action X, Alex will receive the payoff associated with Action X, which is 5. Blake chooses Action Y, Blake will receive the payoff associated with Action Y, which is 10.

2. In a company, a new software is being introduced. Employees will only adopt the software if at least a fraction q of their colleagues have started using it. If a few tightly-knit departments of employees work closely with each other, what could prevent the software from being adopted company-wide?

   (a) The department has fewer than q of the employees
   (b) The department has fewer than 1 - q of the employees
   (c) The department has more than q of the employees
   (d) The department has more than 1 - q of the employees

   Answer: (d)

   Explanation: In this scenario, the adoption threshold is q. For the software to be adopted company-wide, enough employees in every department must adopt it. However, suppose there is a department where more than 1-q of the employees are closely connected. In that case, they may resist the software adoption because they are more influenced by each other than by employees outside their department. Thus, a tightly-knit department with more than 1-q of the employees could prevent the software from being adopted company-wide.

3. In the context of viral marketing, a company aims to promote a new product by targeting key individuals in a social network. These individuals are selected based on their ability to influence others within the network. Given that the network has varying degrees of connectivity among its members, which of the following strategies is most effective for maximizing the spread of the product?

   (a) Target individuals with the highest number of direct connections, regardless of their influence on others
   (b) Target individuals who are central to the network, as they act as bridges between different groups
   (c) Target the most isolated individuals, as they are less likely to have been influenced by others
   (d) Target individuals who are closely connected within a small, dense cluster, as they can influence others within their immediate group

   Answer: (b)

   Explanation: Individuals who are central to a network, often referred to as "hubs" or "bridges," play a crucial role in connecting different parts of the network. By targeting these individuals, the product or message can quickly spread across the network, reaching various groups that are otherwise less connected. Option B is the most effective strategy for maximizing the spread of a product in viral marketing, as it leverages the central position of key individuals who can influence multiple groups within the network.

4. Given a node X having 10 friends/neighbors. 3 of its neighbors have decided to adopt the behavior/action A having a payoff of 10 while 7 of its friends have adopted the action B yielding a payoff of 4. What is the payoff that X gets from its friends who have adopted action A?

(a) 30

(b) 70

(c) 12

(d) 40

Answer: (a)

Explanation: To find the total payoff that node X gets from its friends who have adopted action A: Number of friends adopting action A: 3 Payoff from each friend adopting action A: 10 Total payoff = Number of friends adopting action A × Payoff per friend So, the calculation is: Total Payoff = 3 × 10 = 30

5. For the node X mentioned in question 4, What is the payoff that X gets from its friends who have adopted B?

(a) 30

(b) 40

(c) 28

(d) 38

Answer: (c)

Explanation: To find the total payoff that node X gets from its friends who have adopted action B: Number of friends adopting action B: 7 Payoff from each friend adopting action B: 4 Total payoff = Number of friends adopting action A × Payoff per friend So, the calculation is: Total Payoff = 7 × 4 = 28

6. For node X mentioned in question 4, What is the final action/behavior that X would adopt?

(a) A

(b) B

(c) neither A nor B

(d) cannot say

Answer: (a)

Explanation: Node X would adopt action A because the total payoff from friends who have adopted action A (30) is higher than the total payoff from friends who have adopted action B (28).

7. For the given figure, what is the density of community formed by nodes A, B, C and D?

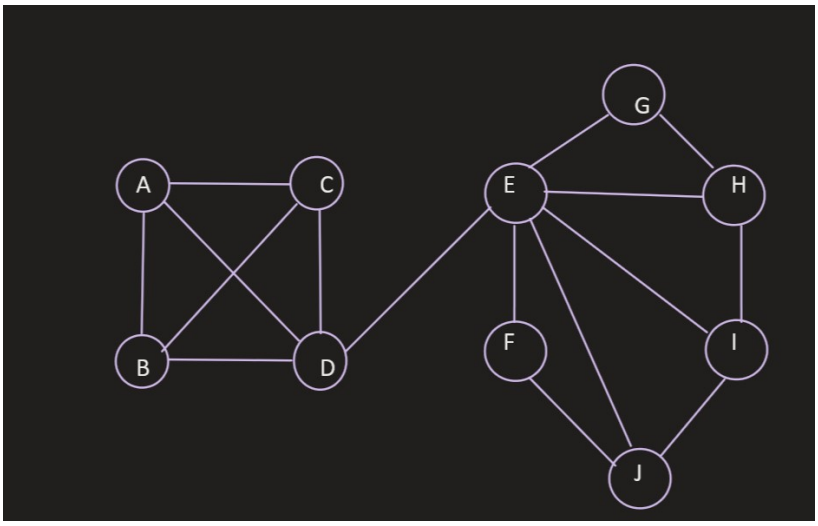


Figure 1: Network G

(a) 0

(b) 1

(c) 2

(d) 3

Answer: (b)

Explanation:

The density of a community is given by the ratio of the actual number of edges to the total number of possible edges in that community. Density = 6/6C2 = 6/6 = 1.

8. Given the network below, if the value above the node is its threshold value to join the protest, is the protest likely to happen?

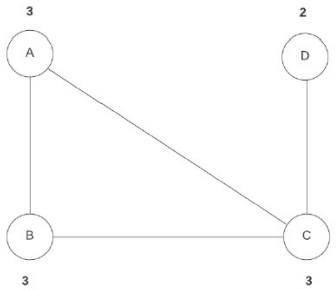


Figure 2: Network H

(a) Yes

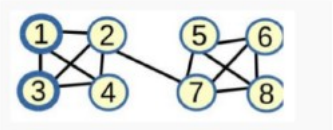
(b) No

(c) Cannot say

Answer: (a)

Explanation: The threshold value for each node is satisfied so protest is likely to happen

9. Given a network as shown in the following Figure, assume that initially, every node in this network has adopted behavior B. Next, a new behavior A is introduced in the network and nodes 1 and 3 are the initial adopters of this behavior A, i.e., nodes 1 and 3 now have adopted behavior A and the rest of the nodes have adopted behavior B. The payoff associated with A is a = 3 and the payoff associated with B is b = 2. After the introduction of this new behavior A in the network, all the nodes will start weighing their options and might change their behavior. This leads to a cascade in the network. When the cascade ends, identify the nodes that will adopt the behavior A.



(a) 1,3,2

(b) 1,3,2,4

(c) 1,3,2,7

(d) 1,3,2,7,5

Answer: (b) Explanation: For Node 2: Payoff to adopt behavior A is Neighbors who adopted A x payoff for A = 2 x 3 = 6 Payoff to adopt B = Neighbors who adopted B x payoff for B = 2 x 2 = 4 As payoff(A)¿payoff(B), Node 2 adopts A

For Node 4: Payoff to adopt behavior A is Neighbors who adopted A x payoff for A = 2 x 3 = 6 Payoff to adopt B = Neighbors who adopted B x payoff for B = 1 x 2 = 2 As payoff(A)¿payoff(B), Node 4 also adopts A

For all the other nodes, there is no neighbor who has adopted behavior B, hence they do not adopt B.

10. Let v be a node in a graph. Suppose that a fraction α of the neighbors of v has adopted behavior X, and a fraction $1 - \alpha$ has adopted behavior Y. In other words, if v has k neighbors, then αk neighbors adopt X, and $(1 - \alpha)k$ neighbors adopt Y. Behavior X yields a payoff of x, and behavior Y yields a payoff of y. Behavior X is a more advantageous choice for v if:

    (a) $\alpha \geq x/y$

    (b) $\alpha \geq y/x$

    (c) $\alpha \geq x/(x + y)$

    (d) $\alpha \geq y/(x + y)$

    Answer: (d)

SOCIAL NETWORKS - NPTEL July 2024

Week 8

1. In an academic research network, where researchers cite other researchers' work and are cited by others, which of the following is correct?

   (a) Researchers who cite other researchers are represented by hubs, and researchers whose work is cited are represented by authorities

   (b) Researchers who are cited by others are represented by hubs, and researchers who cite others are represented by authorities

   (c) Both researchers who cite others and researchers who are cited act as hubs

   (d) Both researchers who are cited and researchers who cite others act as authorities

   Answer: (a)

   Explanation: In an academic research network: Researchers who cite other researchers act as hubs because they are creating links to other researchers' work. Researchers whose work is cited act as authorities because their work is being referenced by many other researchers.

2. In the graph G shown in the following figure, assume that the current pagerank values of A, B and C are 0.5, 0.4 and 0.1 respectively. What will be their pagerank values after one iteration?

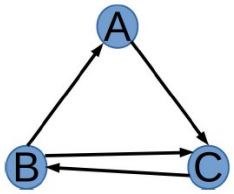


Figure 1: Graph G

   (a) A: 0.2,B: 0.4,C: 0.4

   (b) A: 0.2,B: 0.1,C: 0.7

   (c) A: 0.1,B: 0.2,C: 0.7

   (d) A: 0.7,B: 0.2,C: 0.1

   Answer: (b)

   Explanation: A gets half of points of B, B gets all points of C and C gets half points of B + all points of A

3. Given a network P with 8 nodes with an initial page rank value of 1/8 each. What is the page rank value at the end of 2 iterations?

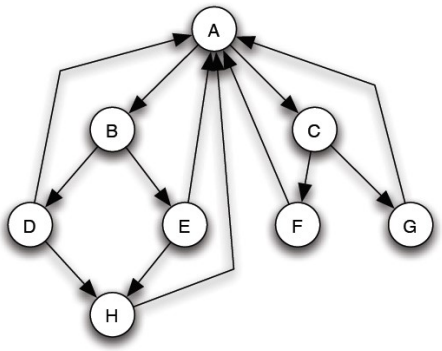


Figure 2: Network P

(a) 1/8
(b) 1/4
(c) 1/32
(d) 1/64

Answer: (b)

Explanation:

In the first iteration, B' = A/2 = 1/16. A' = 1/16 + 1/16 + 1/8 + 1/8 + 1/8 = 1/2 In the second iteration, B" = A'/2 = 1/4

4. Assume we perform a sequence of k hub-authority updates. Each update works as follows: First apply the Authority Update Rule to the current set of scores.Then apply the Hub Update Rule to the resulting set of scores. At the end, the hub and authority scores may involve numbers that are very large. Select all that is TRUE.

(a) We care only about their relative sizes
(b) we can normalise them
(c) we can simply reduce each value by a fixed value in every iteration
(d) we care only about their absolute values

Answer: (a) and (b)

Explanation: a. In the HITS algorithm, the exact values of hub and authority scores are less important than their relative sizes. The algorithm aims to identify the relative importance of each node, and scaling the scores or normalizing them does not affect the ranking of nodes. The focus is on the relative ranking rather than the absolute magnitude of the scores. b. Normalization is a common practice to ensure that the scores remain manageable and comparable. By normalizing the hub and authority scores (e.g., dividing by the maximum score or ensuring that the sum of scores is 1), we can avoid issues with very large numbers and maintain the relative rankings of the nodes. c. This approach might not preserve the relative importance and could lead to inaccurate results. d. The HITS algorithm is designed to identify the most influential hubs and authorities based on their relative scores, not their absolute magnitudes.

5. Assume the hub and authority score for each node in the given graph J as 1. What is the hub and authority score of node 3 in the given graph J in iteration k=1?

(a) a(1)= 1,h(1)= 2
(b) a(1)= 2,h(1)= 1
(c) a(1)= 3,h(1)= 2
(d) a(1)= 2,h(1)= 3

Answer: (a) Explanation: Authority Update Rule: For each page p, update auth(p) to be the sum of the hub scores of all pages that point to it. Hub Update Rule: For each page p, update hub(p) to be the sum of the authority scores of all pages that it points to.

For node 3: only node 1 points to node 3, so a(1) =1. node 3 points to nodes 4 and 5 which implies that h(1) = 2.

6. Consider the following bipartite graph, which of the following is/are correct for repeated improvement?

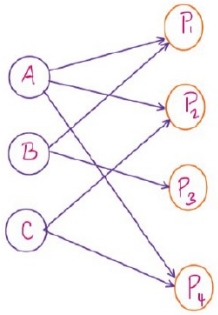


Figure 3: Graph B

(a) B = P1 + P3
(b) C = P3 + P4
(c) P2 = A + B
(d) P3 = B

Answer: (a), (d)

7. Which of the following is TRUE for a Markov Matrix?

Statement I - The sum of elements in every column is 1.
Statement II - Highest eigenvalue of a Markov matrix is 1.

(a) I only
(b) II only
(c) Both
(d) None

Answer: (c)

8. For the given network H, what are the values after say 10 iterations of Page rank updates for the initial value of 1/3 for every node in the network?

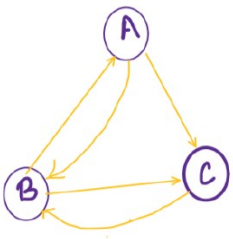


Figure 4: Network H

(a) 0.17,0.50,0.33
(b) 0.33,0.33,0.33
(c) 0.50,0.17,0.33
(d) 0.22,0.44,0.33

Answer: (d)

Explanation:

Try this in any Excel-like application. The update values for each node are given as follows:
A'= B/2, B'= A/2 + C, C'= A/2 + B/2

Iteration	A	B	C
1	0.33	0.33	0.33
2	0.17	0.50	0.33
3	0.25	0.42	0.33
4	0.21	0.46	0.33
5	0.23	0.44	0.33
6	0.22	0.45	0.33
7	0.22	0.44	0.33
8	0.22	0.45	0.33
9	0.22	0.44	0.33
10	0.22	0.44	0.33

9. Which of the following best describes the process for updating the authority scores of web pages in the Hubs and Authorities Algorithm?

   (a) Update the authority score of a page p by aggregating the hub scores of all pages with hyperlinks pointing to p.

   (b) Update the authority score of a page p by summing the authority scores of all pages to which p has outbound hyperlinks.

   (c) Update the authority score of a page p by aggregating the hub scores of all pages to which p has outbound hyperlinks.

   (d) Update the authority score of a page p by summing the authority scores of all pages that p points to.

   Answer: (a)

   Explanation:

   In the Hubs and Authorities algorithm (HITS algorithm):

   1. Authority Update Rule: The authority score of a page p is updated to be the sum of the hub scores of all pages that link to it.

   2. Hub Update Rule: The hub score of a page p is updated to be the sum of the authority scores of all pages it links to.

10. Let C be a unit circle with center at (0,0) in the XY-plane. Then H, the point at which the vector(4,3) intersects C is

    (a) (0,0)

    (b) (3,4)

    (c) (0.6,0.8)

    (d) (0.8,0.6)

    Answer: (d)

    Solution: The magnitude $\|(4,3)\|$ is calculated as:

$$\|(4,3)\| = \sqrt{4^2 + 3^2} = \sqrt{16+9} = \sqrt{25} = 5$$

    To normalize the vector, we divide each component by the magnitude:

$$\mathbf{v}_{\text{normalized}} = \left(\frac{4}{5}, \frac{3}{5}\right) = (0.8, 0.6)$$

SOCIAL NETWORKS - NPTEL July 2024

Week 9

1. Which of the following mechanisms is most likely responsible for the emergence of a power law degree distribution in complex networks?

   (a) Nodes are added with equal probability to all existing nodes.

   (b) Nodes connect randomly without preference to other nodes.

   (c) New nodes preferentially attach to existing nodes with higher degrees.

   (d) All nodes have an equal chance of forming connections.

   Answer: (c)

   Explanation: New nodes preferentially attaching to existing nodes with higher degrees is the process described by Barabási and Albert, leading to the power law degree distribution observed in many real-world networks.

2. Given a network G being generated by the 'rich get richer' phenomenon. Figure 1 shows the snapshot of the network at time t. A new node 'u' enters the network at time t + 1 and makes an edge with one of the existing nodes. What is the probability that node 'u' will make an edge with node 'B'?

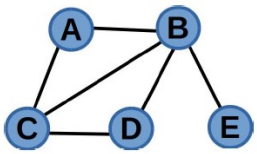


Figure 1: Network G

   (a) 1/2

   (b) 1/3

   (c) 1/4

   (d) 1/5

   Answer: (b)

   Explanation: The sum of degrees of all the nodes in the network is 12. The degree of the node C is 1. Hence, P(link formation = 4/12 = 1/3)

3. Consider the set $E = 2, 4, 6, \ldots, 28, 30$ A value a_1 is selected uniformly at random from this set. Then, a second value a_2 is chosen uniformly at random from the same set. This process is repeated until 10 values, $a_1, a_2, .., a_{10}$ are selected. The sum $S = a_1 + a_2 + .. + a_{10}$ is then calculated. Which of the following ranges represents all possible values that the sum can take?

   (a) $[20, 24, .., 300]$

   (b) $[20, 24, .., 200]$

   (c) $[10, 24, .., 240]$

   (d) $[2, 4, .., 300]$

   Answer: (a)

   Explanation:

4. In a random graph with 400 nodes and edges between any two nodes with a probability of 0.5, where can one expect the peak of the degree-distribution graph?

(a) 100

(b) 200

(c) 300

(d) 250

Answer: (b)

Explanation: In a normal distribution, the peak of the degree-distribution graph is seen around the middle of the domain. Thus here we can say the peak is expected to be around 400 x 0.5 = 200

5. Which of the following is an example of a real-world phenomenon that follows a power law distribution?

   (a) The heights of individuals in a population

   (b) The number of citations academic papers receive

   (c) The distribution of shoe sizes among adults

   (d) The daily temperature in a specific city

   Answer: (b) Explanation: The number of citations academic papers receive often follows a power law distribution, where a small number of papers receive a large number of citations, and most papers receive relatively few.

6. According to the power law, the frequency (shown on the y-axis) is inversely proportional to k, where k represents the values plotted on the x-axis. In this relationship, the exponent a is ideally between which values?

   (a) 0,1

   (b) -1, 1

   (c) 1,5

   (d) 2,3

   Answer: (d)

   Explanation: In many real-world cases, the exponent a in a power law distribution often falls between 2 and 3. This range is particularly common in networks like social networks, citation networks, and many others.

7. Which of the following statements is true about the Barabási-Albert (BA) model for scale-free networks?

   (a) Nodes are added randomly to the network

   (b) The BA model follows a preferential attachment mechanism

   (c) All nodes in the BA model have equal probabilities of forming connections

   (d) New nodes in the BA model are equally likely to connect to any existing node

   Answer: (b)

   Explanation: (a) Nodes are added randomly to the network—This is incorrect because, in the Barabási-Albert (BA) model, nodes are not added randomly. New nodes are added in a way that prefers connecting to existing nodes with higher degrees (more connections).
   (b) The BA model follows a preferential attachment mechanism—This is correct. The BA model is based on the principle of preferential attachment, meaning that new nodes are more likely to connect to nodes that already have a large number of connections, leading to the formation of hubs.
   (c) All nodes in the BA model have equal probabilities of forming connections—This is incorrect because, in the BA model, nodes with more connections (higher degrees) have a higher probability of attracting new connections.
   (d) New nodes in the BA model are equally likely to connect to any existing node—This is incorrect because the BA model's preferential attachment mechanism ensures that new nodes are more likely to connect to nodes with more connections, not randomly.

8. In the Erdos-Renyi $G(n, p)$ model for random graphs, which of the following statements accurately describes the model?

   (a) The model involves adding a fixed number of edges randomly to a complete graph with n nodes

(b) Each pair of nodes in a graph with n nodes has a probability p of being connected by an edge, independently of other pairs

(c) The model constructs a graph by preferentially attaching new nodes to existing nodes with higher degrees

(d) In this model, the probability p determines the number of nodes added at each step

Answer: (b)

Explanation: In the Erdos-Renyi $G(n, p)$ model, every possible edge between n nodes is included with a probability p and this connection is made independently of other edges.

9. Assume a network G is created having 100 nodes using Networkx. For each node, 'weight' should take a value uniformly at random from 100 to 200. Identify the code snippet to associate an attribute 'weight' with each of the nodes.

(a)
```
for each in G.nodes ( ) :
    G.node[each]['weight']=random.randint(100,200)
```

(b)
```
for each in G.nodes():
    G.node[each]['weight']=random.randint(99,201)
```

(c)
```
for each in G.nodes():
    G.node['weight']=random.randint(99,201)
```

(d)
```
for each in G.nodes():
    G.node['weight']=random.randint(100,200)
```

Answer: (a)

10. Consider the given network K, which of the following nodes does a new node 'X' entering the network choose to form an edge according to preferential attachment?

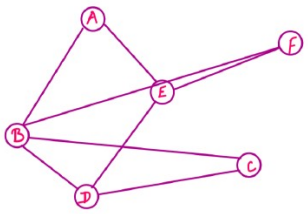


Figure 2: Network K

(a) A

(b) B

(c) C

(d) D

(e) E

(f) F

Answer: (b)

Solution: According to preferential attachment a new node chooses a node with maximum degree as compared to the other nodes.

SOCIAL NETWORKS - NPTEL July 2024

Week 10

1. In the percolation model of the SIR epidemic model, a node in the network becomes infected if:

   (a) There is a direct edge between the node and the initially infected node, regardless of its state

   (b) The infection rate is greater than 0.5 for all edges in the network

   (c) There exists a path of open edges from the initially infected node to the node in question

   (d) The node has at least one neighbor that is not infected

   Answer: (c)

   Explanation: According to the percolation model, a node becomes infected only if there is a path of "open" edges connecting it to the initially infected node.

2. Which of the following correctly describes the behavior of a disease in a population based on the basic reproductive number, R_0?

   (a) When $R_0 < 1$, every infected person infects less than one secondary person, and the disease persists with a high probability

   (b) When $R_0 > 1$, every infected person infects more than one secondary person, and the disease always dies away

   (c) When $R_0 < 1$, the disease dies out with probability 1, since each infected person causes less than one secondary infection

   (d) When $R_0 > 1$, the disease persists in the network with a probability of 1

   Answer: (c)

   Explanation: The basic reproductive number R_0 represents the average number of secondary infections caused by a single infected individual. When $R_0 < 1$, each infected person infects fewer than one other person on average, meaning the infection rate is insufficient to sustain an outbreak. Over time, the disease will die out, and thus the disease dies out with probability 1. When $R_0 > 1$, each infected person is causing more than one secondary infection, so the disease has the potential to spread and persist within the population. However, there remains a small probability that the disease may still die out (e.g., due to random fluctuations in transmission), but overall, the disease persists with positive probability. Thus, the correct statement is that when $R_0 < 1$, the disease will die out with probability 1

3. Which of the following best represents the SIS (Susceptible-Infected-Susceptible) disease model

   (a) Chickenpox, where people who recover gain lifelong immunity

   (b) Common cold, where individuals can recover but remain susceptible to reinfection

   (c) Measles, where recovery typically leads to permanent immunity

   (d) Rabies, where infection usually leads to death without the possibility of reinfection

   Answer: (b)

   Explanation: The SIS model describes diseases where individuals do not gain permanent immunity after recovery and can become susceptible to the infection again. The common cold is a classic example of such a disease, as people can catch the cold multiple times throughout their lives.

4. If there exist a graph G, in which there are 10 nodes in level 1 and every node has 10 children then, number of nodes in 5^{th} level will be

   (a) 50

   (b) 5^{10}

(c) 10^5

(d) 10^{10}

Answer: (c)

Explanation: Each node in 1st layer has 10 children so in 2nd layer there will be 10 x 10 Each node in 2nd layer has 10 children so in 3rd layer there will be 10 x 10 x 10 .. Each node in 5th layer has 10 children so in 5th layer there will be 10^5

5. Given the probability of spreading disease is 0.3, what is the probability that the same person is not infected for the second time?

   (a) 0.3

   (b) 0.7

   (c) 1

   (d) 0

   Answer: (b)

6. In an SIS model, if the probability of transmitting the disease is 0.4, what is the likelihood that an individual who has recovered will be reinfected?

   (a) 0

   (b) 0.4

   (c) 0.6

   (d) 1

   Answer: (b)

   Explanation: In SIS model, person who recovers remains susceptible to the disease with the same probability.

7. Consider the example of digital music streaming services, where we plot a graph between a song's popularity and its number of streams, following a power law distribution. The most popular songs receive a significantly high number of streams, while lesser-known songs receive far fewer streams. Assuming that top-charting songs account for 30% of total streams, which of the following is true?

   (a) Streaming only top-charting songs is profitable

   (b) Streaming niche songs will lead to a loss

   (c) Streaming both top-charting and niche songs is profitable

   (d) None of the above

   Answer: (c)

   Explanation: In a power law distribution scenario, focusing only on top-charting songs can yield profits due to their popularity. However, combining both top-charting and niche songs is more profitable, as it maximizes revenue by catering to both mainstream and niche markets. Streaming niche songs alone does not guarantee a loss, as they can still contribute significantly to overall profitability through the long-tail effect.

8. If the basic reproductive number (R_0) of contagion is greater than 1, meaning each infected individual, on average, infects more than one other person, which of the following strategies would be most effective in reducing the spread and curbing the contagion?

   (a) Implementing widespread vaccination to increase immunity in the population

   (b) Allowing large public gatherings to promote herd immunity through natural infection

   (c) Reducing healthcare services to manage the number of new infections

   (d) Ignoring the infection and focusing on treatment after symptoms appear

   Answer: (a)

   Explanation: When $R_0 > 1$, the disease spreads quickly. To control this, vaccination is an effective way to reduce the number of susceptible individuals, thereby lowering the reproductive number and curbing the spread. Options B, C, and D would lead to an increase in the spread and are not effective in controlling the contagion..

9. According to Zipf's law, the frequency of the nth most common word in a language is proportional to

   (a) $1/n$
   (b) $log(n)$
   (c) $1/log(n)$
   (d) $nlog(n)$

   Answer: (a)

   Explanation: Zipf's Law states that in natural languages, the frequency of a word is inversely proportional to its rank. This means that the frequency of the nth most common word is approximately proportional to $1/n$, where n is the word's rank in terms of frequency.

10. In the given network K, each infected node has edges to two nodes in the next layer; since it infects each with probability 2/3, the expected number of new cases caused by a node is

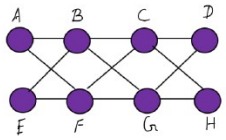


Figure 1: Network K

   (a) $\frac{2}{3}$
   (b) $\frac{4}{3}$
   (c) $\frac{1}{3}$
   (d) $\frac{1}{4}$

   Answer: (b)

SOCIAL NETWORKS - NPTEL July 2024

Week 11

1. In Stanley Milgram's famous small-world experiment, which was designed to investigate the "six degrees of separation" concept, participants were asked to send a letter to a target individual using only acquaintances. What was a key finding of this experiment?

   (a) Most letters reached the target individual through a direct connection within three steps

   (b) The average number of steps required to reach the target individual was around 6, demonstrating the small-world phenomenon

   (c) Participants found that there was no measurable average number of steps to reach the target

   (d) The letters typically reached the target individual through a network of over 100 steps

   Answer: (b)

   Explanation: Milgram's small-world experiment aimed to understand the average number of steps required to connect two random individuals in a network. The finding demonstrated that despite the vast number of people, the average number of intermediary steps needed to connect any two people in the world was about 6, illustrating the small-world phenomenon.

2. Suppose each of your 50 friends has 30 friends (excluding you). Each of those friends also has 30 friends (excluding them), and this pattern of each having 30 friends continues. How many people can you reach at level i (where Level 1 refers to your direct friends, Level 2 refers to your friends' friends, and so on)?

   (a) 30^i

   (b) $50 + 30^i$

   (c) $50 + 30^{i-1}$

   (d) $50 \times 30^{i-1}$

   Answer: (d)

   Explanation:

   - At level 1, you can reach 50 friends.
   - At level 2, each of your 50 friends has 30 unique friends, giving $50 \times 30 = 50 = 50 \times 30^1$ people
   - At level 3, the same pattern continues, resulting in 50×30^2
   - For i levels, you can reach $50 \times 30^{i-1}$

3. Imagine you work in a company where most of your close connections are with colleagues from your own department. However, you occasionally attend industry conferences where you meet people from different companies and countries, building new, distant connections. According to Watts and Strogatz, which of the two phenomena are responsible for creating small-world networks?

   (a) Your close colleagues frequently form connections among themselves (triadic closure), while your occasional contacts at conferences represent weak ties.

   (b) Your colleagues form a tightly-knit group (community structure), while new connections are created within smaller subgroups (triadic closure).

   (c) You tend to connect with nearby colleagues due to shared work (homophily), while meeting distant professionals at conferences forms weak ties.

   (d) Your work department (a focus of shared activity) brings people together (foci closure), while you tend to form connections with similar colleagues (homophily).

Answer: (c)

Explanation: Homophily describes how people connect with those close to them or who share similar environments, like your colleagues. Weak ties are the connections you build with distant or unfamiliar people, such as those you meet at conferences, expanding your network.

4. Alex creates a social network graph on NetworkX by forming a 2D grid where each person is connected to their immediate neighbors. After constructing the grid, Alex randomly reconnects some of these relationships with a probability of 0.02. What is the expected average path length between any two people in this network?

   (a) 1

   (b) 2

   (c) 6

   (d) 20

   Answer: (c)

   Explanation: In a 2D grid network, the average path length between nodes (people) is relatively large due to the local connections. With a rewiring probability of 0.02, some edges are randomly replaced, introducing shortcuts and reducing the average path length. For small-world networks with a low rewiring probability, the average path length is typically reduced but not drastically. Given typical results, the average path length is often around 6 in such scenarios.

5. In a network of friends, the distance between Jordan and Taylor is 25. If the network is built using the Watts-Strogatz model with k=2, what is the probability that Jordan and Taylor are connected through a weak tie?

   (a) 1

   (b) 0.01

   (c) 0.04

   (d) 0.0016

   Answer: (d)

   Explanation: In the Watts-Strogatz model: k = 2 indicates that each node initially connects to two neighbors on either side. The probability of two nodes being connected through a weak tie can be approximated by: $Probability = \frac{1}{d^k}$ where d is the distance between the nodes and k is the number of neighbors.

6. In a professional network, suppose Ravi wants to connect with Priya through the shortest path. Ravi has four colleagues: Aarti, Manoj, Sunil, and Meera. The distances from each colleague to Priya are as follows: Aarti is 8 steps away, Manoj is 3 steps away, Sunil is 7 steps away, and Meera is 5 steps away. If Ravi uses a decentralized approach to find the shortest path to Priya, which immediate colleague should he choose?

   (a) Aarti

   (b) Manoj

   (c) Sunil

   (d) Meera

   Answer: (b) Explanation: In a decentralized search, Ravi should choose the colleague who has the shortest distance to Priya. Here, Manoj is the closest to Priya, with a distance of 3 steps.

7. Consider a social network where initially, each person is connected to their immediate neighbors in a 1-D grid. To introduce some randomness, the network applies a random rewiring process with a probability of 0.03. This means that each connection between two neighbors has a 3% chance of being rewired to connect with a random person in the network. What impact does this random rewiring have on the network?

   (a) It increases the average path length by creating more local connections

   (b) It has no effect on the average path length as the network remains a regular grid

   (c) It decreases the average path length by introducing shortcuts and creating a small-world effect

   (d) It decreases the number of connections in the network, making it less connected

Answer: (c)

Explanation: Random rewiring in a generative model, like the Watts-Strogatz model, introduces shortcuts by randomly replacing some of the existing connections. This process tends to reduce the average path length between nodes, creating a small-world effect where nodes are more interconnected despite the network's overall sparseness.

8. In a 2-D Watts-Strogatz model, which of the following is the ideal value of clustering exponent for an efficient decentralized search?

   (a) -1
   (b) 0
   (c) 1
   (d) 2

   Answer: (d)

   Explanation: In a 2-D Watts-Strogatz model, for an efficient decentralised search, the ideal value of clustering exponent is 2 so that random links follow an inverse-square distribution.

9. Consider a large online marketplace where each seller is connected to local buyers and a few random buyers across the platform to facilitate quicker connections. If this marketplace represents a small-world network, what is the typical average distance between any two random sellers?

   (a) n
   (b) $\log n$
   (c) $n \log n$
   (d) $\log n^2$

   Answer: (b)

   Explanation: In a small-world network, the average distance between any two random sellers scales as log n. This reflects how the random connections in the network significantly reduce the path length, even in large networks.

10. Imagine you are part of a local sports club (Cluster 'A') with many teammates. You also discover another club (Cluster 'B') that has a lot of team members who know each other well. When you meet someone from Cluster 'B' and become friends with them, this new friendship illustrates:

    (a) Social influence
    (b) Selection
    (c) Triadic closure
    (d) Weak tie

    Answer: (d)

    Explanation: In this example, befriending someone from a different sports club connects two previously separate groups. This new connection is an example of a weak tie, as it links different parts of the social network.

SOCIAL NETWORKS - NPTEL July 2024

Week 12

1. Which of the following is True in a hierarchical network structure?

   (a) Low-ranking employees form the densely connected core, while high-ranking employees disperse around the core as the periphery

   (b) The core and the periphery can swap roles within the network

   (c) Whether an individual is in the core or the periphery is independent of their rank or position

   (d) High-ranking employees form the densely connected core, while low-ranking employees disperse around the core as the periphery

   Answer: (d)

   Explanation: In a hierarchical network structure, high-ranking employees are part of the densely connected core, while low-ranking employees disperse around this core as the periphery.

2. A node in a network belongs to exactly 3 cliques (fully connected subgraphs) of sizes 3, 4, and 6. Which of the following cannot be the core number of this node?

   (a) 1

   (b) 3

   (c) 4

   (d) 6

   Answer: (a)

   Explanation: While 1 can be a core number, given that the node participates in higher cliques with higher degrees, the implication is that the lowest core number consistent with its connections should be greater than 1.

3. In a social network of a popular online community, the "core" refers to the group of users who are highly interconnected and influential. The "periphery" refers to users who are less connected. Which of the following best describes a "pseudo-core" in this context?

   (a) Users who are part of the highly connected core group

   (b) Users who are part of the less connected periphery group

   (c) Users who are not in the innermost core group but have a similar influence or spreading power as the innermost core

   (d) Users who are not in the outermost periphery group but have a similar influence or spreading power as the outermost periphery

   Answer: (c)

   Explanation: Pseudo-cores are users who do not belong to the innermost core of the network but have equal spreading power or influence as those in the innermost core.

4. Choose the most relevant option that represents an Internet meme:

   (a) A widely shared video of a cat doing something humorous, circulated across social media

   (b) A text message from your friend asking for help with an assignment

   (c) An official email from your boss about a company meeting

   (d) A news article discussing global economic trends

Answer: (a)

Explanation: An Internet meme is typically a piece of content, such as an image, video, or phrase, that spreads rapidly across the internet, often in a humorous or relatable way. Option A describes a humorous video of a cat that is shared widely on social media, which fits the definition of an Internet meme. The other options refer to messages or content that are personal, formal, or informative, and not intended for viral, humorous sharing.

5. What defines a k-core in a sub-graph of a graph?

   (a) The maximum degree of nodes in the sub-graph is exactly k

   (b) Each node in the sub-graph has a degree less than k

   (c) Each node in the sub-graph has a degree greater than k

   (d) Each node in the sub-graph has a degree greater than or equal to k

   Answer: (d)

   Explanation: A k-core is a sub-graph of a larger graph where each node has a degree (the number of connections it has) that is greater than or equal to k. The k-core structure helps in identifying the most connected parts of a network.

6. In a corporate organization, which factor is most correlated with an employee's ability to influence decision-making?

   (a) The number of direct reports the employee has (degree of the node)

   (b) How quickly the employee can access other departments (closeness)

   (c) How often the employee acts as a key connector between departments (betweenness)

   (d) The employee's position within the central management team (coreness)

   Answer: (d)

   Explanation: Coreness refers to the employee's position within the central management team, which correlates strongly with their ability to influence decision-making due to their central role in the organization.

7. Which of the following statements is true about myopic search algorithms?

   (a) Myopic search algorithms always perform as well as optimal search algorithms

   (b) The time complexity of myopic search algorithms is generally high

   (c) Myopic search algorithms are guaranteed to find the optimal solution every time

   (d) Optimal search algorithms typically have lower time complexity than myopic search algorithms

   Answer: (b)

   Explanation: Myopic search algorithms can have high time complexity. They may require exploring many potential solutions locally and can be less efficient in finding a solution compared to more globally informed algorithms.

8. In a network of n social media users, where each user has a degree (number of direct connections) of x, for which value(s) of x will the following statement hold true: "For every user with degree x, their shell number will also be x?

   (a) 1

   (b) n-1

   (c) n

   (d) n+1

   Answer: (a)

   Explanation: The shell number of a node in a graph represents the node's position relative to the core structure of the graph. For a node with degree 1, which is connected to only one other node, its shell number will also be 1 because it belongs to the outermost shell in a shell decomposition. For nodes with degrees 2 or higher, the relationship between degree and shell number becomes more complex and does not necessarily equate to the degree. Thus, the statement specifically holds true for nodes with degree 1.

9. In a social network analysis using the k-shell decomposition algorithm, what does the i^{th} iteration involve?

(a) Removing all users with exactly i connections from the network.

(b) Recursively removing users with exactly i connections until no such users remain in the network

(c) Recursively removing users with i or fewer connections until no such users remain in the network

(d) Recursively removing users with i or more connections until no such users remain in the network

Answer: (c)

Explanation: In the i^{th} iteration of the k-shell decomposition algorithm, nodes with degree i or less are recursively removed from the graph until none remain. This helps in identifying the structure of the network in terms of its k-shell layers.

10. What is the number of nodes in the 4-core of the given graph H?

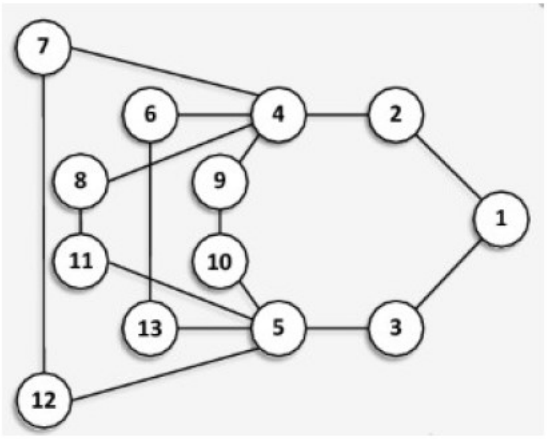


Figure 1: Graph H

(a) 0

(b) 1

(c) 2

(d) 3

Answer: (a)

Explanation: First all the nodes with degree 2 are removed. Then, nodes 4 and 5 become isolated nodes, they are also removed in the same iteration. Hence, entire network falls in a single shell/core.