**Social Networking Notes**

G = nx.read\_edgelist('data.txt')

print(nx.info(G)) -> Graph type, no. of edges, no.of node, everage degree

nx.read\_pajek('data.net')

nx.read\_graphml('data.net')

nx.read\_gexf('data.net')

=============================

Degree distribution

-> tells us no. of nodes having particular degree

nx.degree(G) -> { 'node': degree}

nx.degree(G).values() -> list of degrees

if a network follows complete power law then , log log plot will be straight line.

plt.loglog(unique\_Degree, count)

==============================

Density = no. of edges present/total possible edges

nx.desnity(G)

total possible dges = nC2

=======================

clustering coefficient = for a given node CC tells us, no. of links present among the neighbours of this node with respect to the total possible links

= actual no. of friendship/total possible friendship

nx.clustering(G) -> {'node': clustering\_Coef}

nx.average\_clustering(G)

=======================================

Diameter of a network = maximum shortest path to travel to go from one node to another or we can say it is the shortest path between two most distant nodes in the network

nx.diameter(G)

========================================

gephi -> open source tool to visualize data written in java

=========================================

Probability of any node ‘v’ being isolated when an edge is included = 2/n

Probability of ‘v’ not being isolated = 1-2/n

Pr(‘v’ not being isolated after putting ‘k’ edges) = (1-2/n)k

When we include ‘n’ edges then,

Pr(‘v’ not being isolated) = (1-2/n)n = ((1-1/(n/2))n/2)2) ------------ (i)

Equivalent to (1-1/n)n = 1/e

So (i) = 1/e2

e = 2.17  
So probability ‘v’ not isolated when we put ‘n’ edges is 1/e2

Note:   
if we have ‘n’ nodes then then we just need nlogn number of edges to make the graph connected  
  
=================================================

random.choice(x) -> x is a list, this will give a random list element

4. Key tag in GraphML is used to assign

(a) Node only

(b) Edge only

(c) Both node and edge

(d) loop

Answer: (c)

8. Choose the data set format which starts with the keyword “graph”?

(a) GML

(b) Graph Exchange XML

(c) txt

(d) GEXF

Answer: (a)

9. The degree distribution of most real-world networks follows which law?

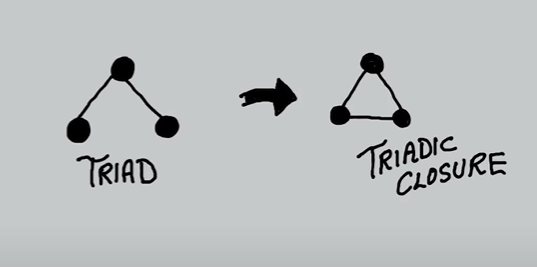
(a) Zipf’s Law

(b) Benford’s Law

(c) Power Law

(d) Difficult to say; can follow any distribution

Answer: (c)

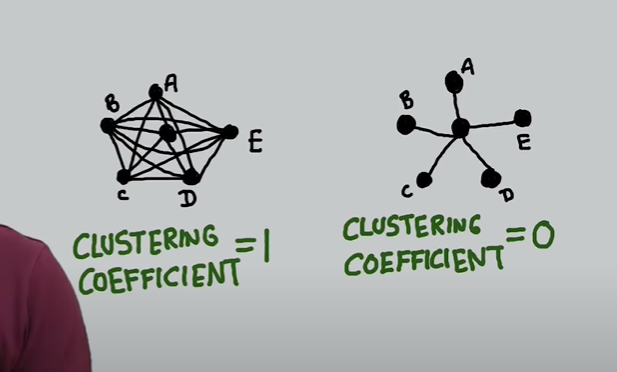


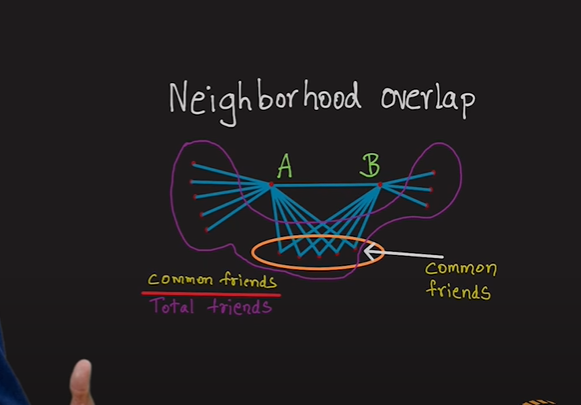
C

B

A

Given that , A is good friend with both B and C, then B and C will also eventually become good friend this conversion is basically triadic closure  
but, if so is not happening then they are considered as weak ties.





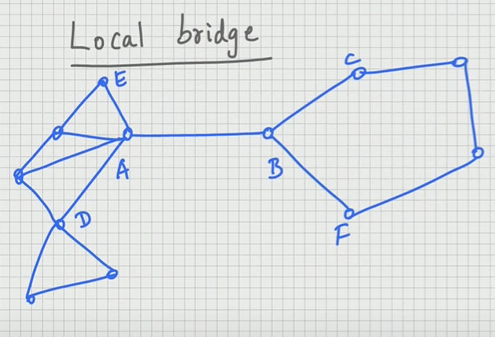
Mathematically, neighbourhood overlap

=========================================================

Local Bridge

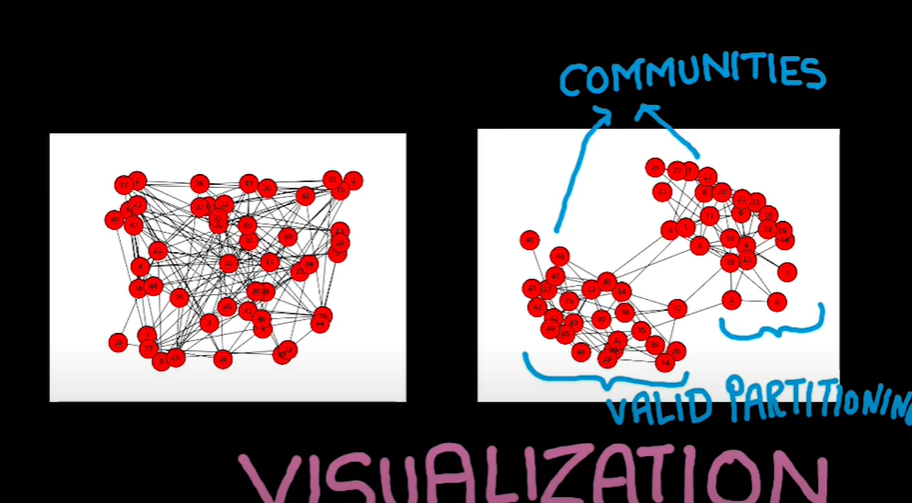
A and B is called local bridge if there is no triad between them (no common friends)

Every bridge is a local bridge



Such, local bridge is also called weak ties as we can see, none of B’ friend A knows and vice versa.

**Embeddedness**  
Embeddedness of relationship between A and B is defines as the no. of common friends they have.



Between two communities, it’s valid partitioning if there is more intracommunity edges(within) and less intercommunity edges(between)

**Betweenness**

Betweenness of an edge (X-Y) is no. of shortest path between A and B that passes through edge (X-Y) / total shortest path between vertices A and B

Summation of all those fractions with possible two vertices A and B with edge (X-Y) coming in the path  
Higher the betweenness higher is the chance that (X-Y) is a bridge connecting communities, removing these edges will give the communities

===========================================================

Questions:

2. Choose the correct statement.

(a) Clustering Coefficient denotes the probability of two nodes to become friends with each other

(b) Triadic Closure phenomenon is rare in all kinds of networks

(c) There is no likelihood for a person to become friend in future if he is not a friend in the present

(d) Acquaintances lead to strong ties

Answer: (a)

3. In Girvan Newman Algorithm, Edge G-H get removed after D-E what does this imply?

(a) G-H has high betweenness than D-E

(b) D-E has high betweenness than G-H

(c) G-H has more shortest paths

(d) Both edge are of same betweenness

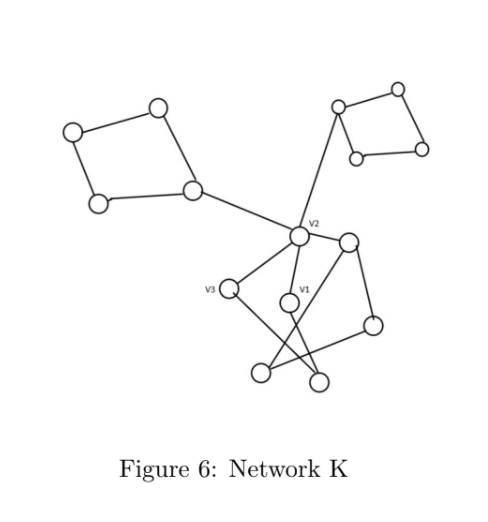
Answer: (b)

7. What role/property of social networks is at display here in network K?

I. Structural hole

II. V2 monopolises

III. Brokerage



(a) I and II only

(b) II and III only

(c) I and III only

(d) I, II and III

Answer: (d) [[notes]](https://chat.openai.com/share/97a7100d-445a-402e-b8d3-d5ec99a6d41b)

9. Given that neighbourhood overlap of an edge m is equal to 0. Then m is

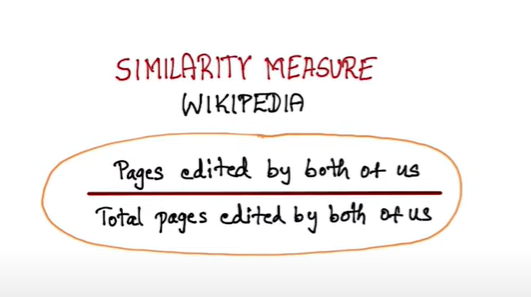
(a) a strong tie

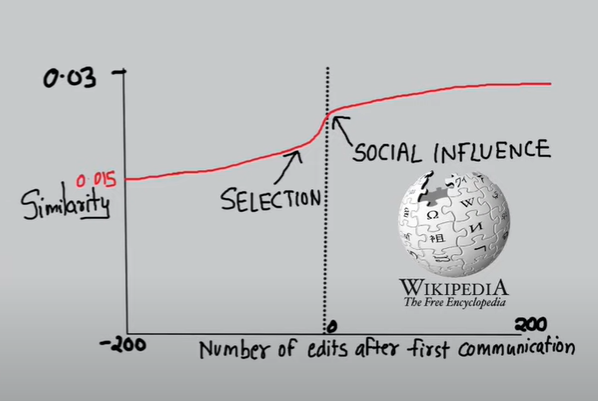
(b) a local bridge

(c) edge with high betweenness

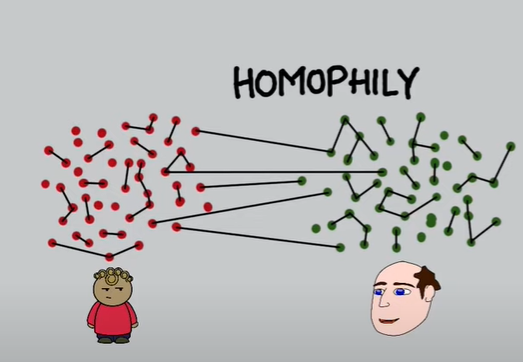
(d) a triad

Answer: (b)

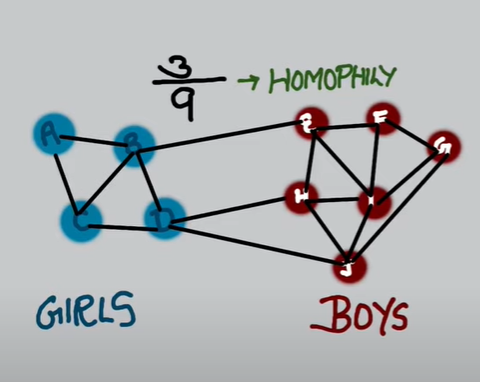
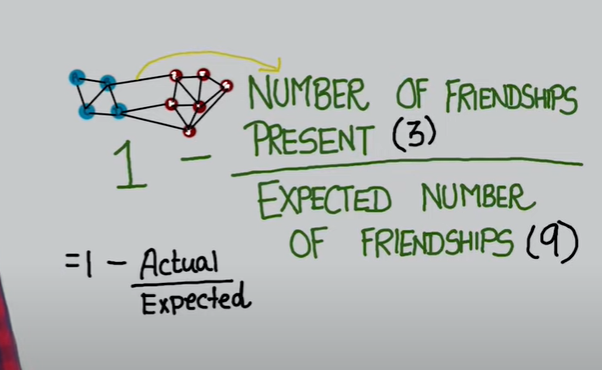




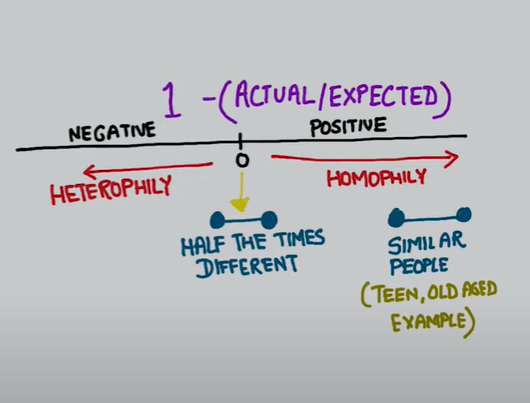
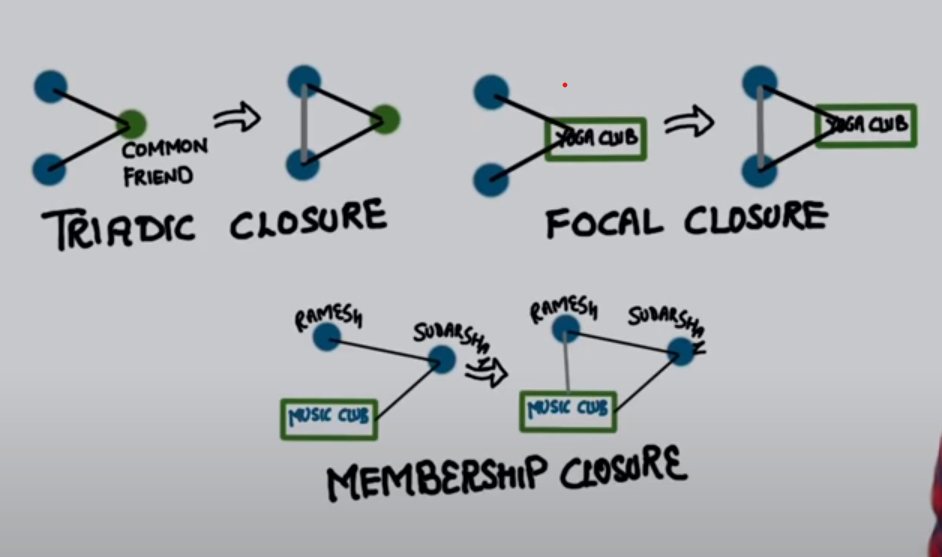
**Homophily**Picking any random edge from a graph if most of the time the edge is between between same kind of nodes then graph is homophily.



Here if there is a group of teenagers and adults, then most f the friendship will be seen within the groups i.e., teens will be talking with teens and adult with adults.  
So this kind of graph demonstrates homophily.

 no. of friendship between girl and boys(expected) = 18/2 18->total friendship

no. of friendship between girl and boys(actual) = 3

**Focal closure** -> two people going to same club tends to become friends.

**Membership closure** -> If A and B are friends and A goes to a club, then B also starts going to that club.

Probability of two people becoming friends increases when the number of common friends increases.

P -> probability A and B are friends when they have 1 common friend

Probability A and B are not friends when they have 1 comm. Friend = 1-P

Probability A and B are not friend when they have k common friend = (1-P)k

Probability A and B are friends when they have k common friend = 1-(1-P)k

T(k) = 1 – (1 – P)k = Probability for 2 people being friends when they have k common friends.

**Questions:**

Similarity measure can be defined as

(a) Total number of work done by number of work in common

(b) Number of work they do in common by Total Number of work done by both

(c) Total number of friends by number of friends in common

(d) Number of friends in common by Total Number of friends

Answer: (b)

Social affiliation network is

(a) Complete and bipartite

(b) Complete and not bipartite

(c) Bipartite and not Complete

(d) Neither complete nor Bipartite

Answer: (c)

In general, when we calculate the value for homophily it represents value between 0 to 1, what value will we

get if it shows heterogeneity?

(a) 0

(b) 1

(c) < 1

Answer: (c)

Which situation doesn’t occur when homophily in a network of 2 Classes is greater than 0.5?

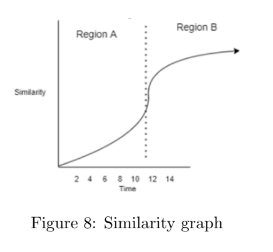
(a) Across edges are low in proportion

(b) People tend to make friends within their group

(c) The probability of selecting across edge is low

(d) People make friends outside the group

Answer: (d)



Which phenomenon is represented in region A of the graph shown below?

(a) Social Influence

(b) Foci closure

(c) Membership closure

(d) Selection

Answer: (d)

10. Which phenomenon is represented in region B of the similarity graph below?

(a) Social Influence

(b) Foci closure

(c) Membership closure

(d) Selection

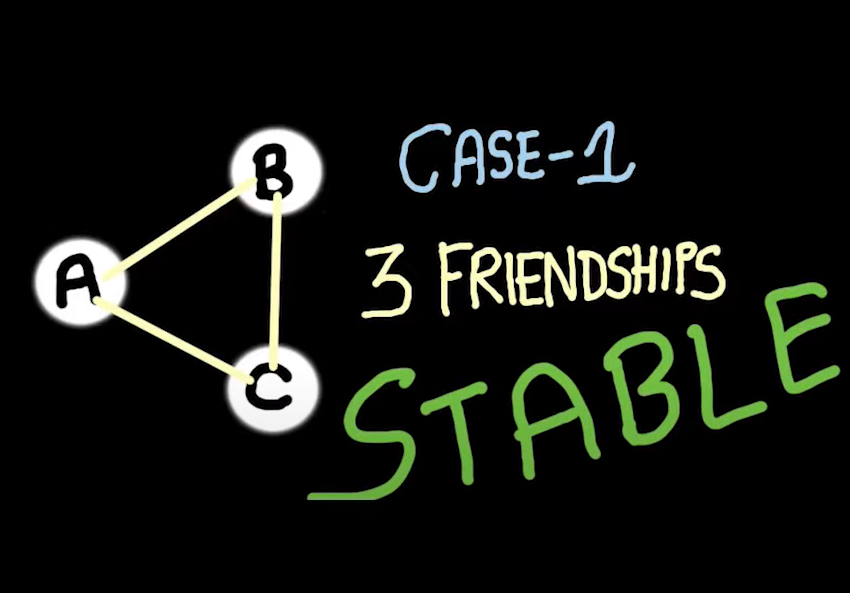
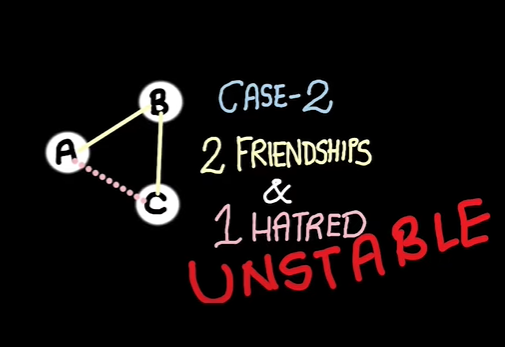
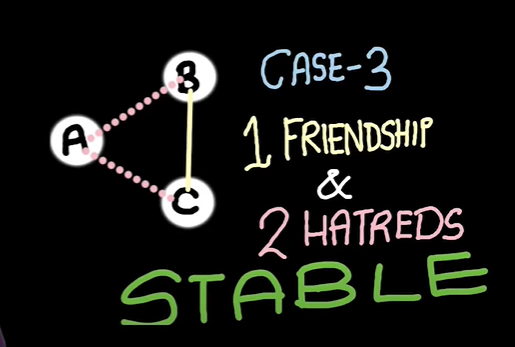
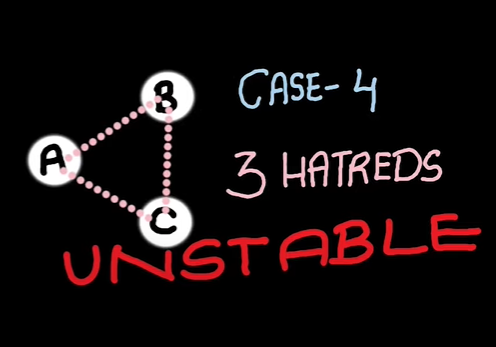
Answer: (a)

**Spatial Segregation:**

Spatial Segregation refers to the distribution of social groups or any other elements in space. In Spatial Segregation, people tend to migrate to other places where they have more neighbours who are like them.

**Schelling model:**

model of segregation showed that even when individuals (or "agents") didn't mind being surrounded or living by agents of a different race or economic background, they would still choose to segregate themselves from other agents over time!

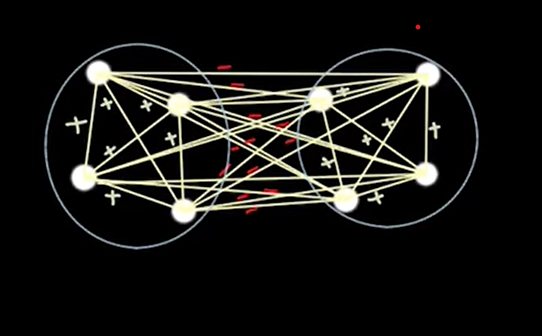
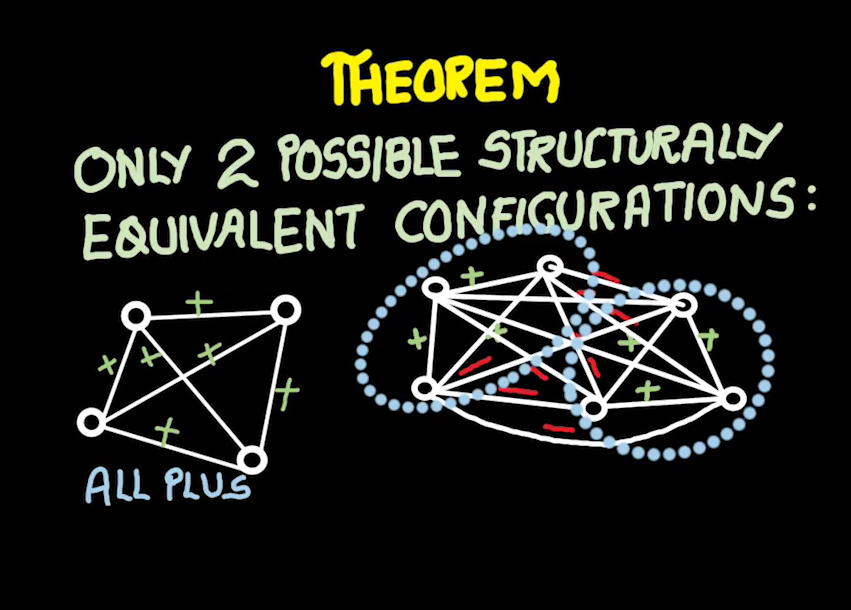
**** ****  

Case 2 converts to case 3 to get a stable triangle

Case 4 converts to case 3 to get a stable triangle ( enemy’s enemy is friend)

A graph is structurally balanced if all the edges are positive (friendship among every node)

A graph with negative edges can also be structurally balanced, in such graph their emerges two components or groups.

**Questions:**

Two of my close friends hate each other. In how many ways can the structure evolve to a stable configuration?

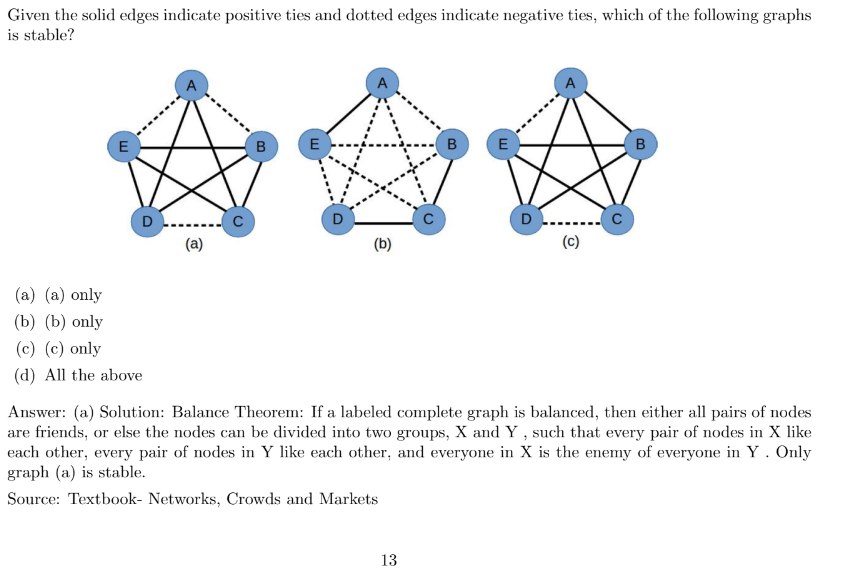
(a) 1

(b) 2

(c) 3

(d) It is already stable

Answer: (c)



**Doubt**

When is a signed triangle said to be unstable?

(a) zero positive edges

(b) 3 positive edges

(c) even number of positive edges

(d) odd number of positive edges

Answer: (c)

Can we have a complete signed graph on 4 nodes (K4) and 5 nodes (K5) respectively, each having exactly one

unstable triangle respectively?

(a) No, No

(b) No, Yes

(c) Yes, No

(d) Yes, Yes

Answer: (a)

Unstable triangles are a result of having odd numbers of negative edges, and in complete signed graphs, all nodes have an equal number of negative and positive edges, so the number of negative edges connecting any set of nodes is always even.

Therefore, the correct answer is (a) "No, No." In complete signed graphs on 4 nodes (K4) and 5 nodes (K5), there cannot be exactly one unstable triangle.

**Web graph:**

Collection of web pages(nodes) linked to each other though directed edges(hyperlinks)

A node accumulates a lot of gold coins if the nodes pointing to it accumulates a lot of coins,

Meaning a web page is ranked high if the pages pointing to it are ranked high

**Getting the web graph:** just take random walk i.e., visit each node randomly and it will give the graph.

**Ranking:**

1. **Using point distribution**

Initially all the nodes will have some coins.

A node will share it’s point among its neighbours, and will also receive from it’s neighbours, after doing this for a while there will come a time where the points of each node doesn’t change even after sharing. The node having highest point is ranked high.

*(doesn’t matter if the nodes have different coins at start, still will give right result)*

1. **Random walk**

At the beginning, a node is randomly and uniformly selected. From this node, one of its neighbours is chosen at random, and that neighbour is assigned a point. This process is then repeated from the selected neighbour.

In some cases, a node may not have any outgoing edges. In such situations, a random node is selected from the entire graph, and the process of walking and accumulating points is resumed. This process is referred to as teleportation.

Over time, a point of convergence will be reached when the coin assignments no longer change. The node with the highest point total will be ranked the highest.

*Indegree and Page rank are not co-related.*

*Both algorithms will give same result and in decreasing order of page rank.*

**Questions:**

Which of the following is the most efficient way of obtaining the big web graph containing billions of nodes?

(a) Breadth First Search

(b) Depth First Search

(c) Linear search

(d) Binary search

(e) Random walk

Answer: (e)

What is the problem in link analysis that teleportation solves?

(a) differentiate an important website from all websites

(b) It prevents web crawlers from getting stuck in infinite loops

(c) It helps web crawler to visit all nodes during analysis

(d) none of the above

Answer: (c)

Choose the correct option corresponding to the gold coins’ distribution game.

(a) The game might not converge.

(b) The game converges even with people having an equal or unequal number of gold coins.

(c) The game converges only when people have an unequal number of gold coins.

(d) The game converges only when people have an equal number of gold coins.

Answer: (b)

Consider algorithm 1 to be equal sharing coin distribution game and algorithm 2 to be random dropping coin

distribution game. Which of the following is true?

(a) Algorithm 1 ranks the nodes in ascending order of their importance while algorithm 2 ranks the nodes in

descending order of importance.

(b) Both the algorithms rank the nodes in descending order of their importance but give different results.

(c) Algorithm 1 ranks the nodes in descending order of their importance while algorithm 2 ranks the nodes

in ascending order of importance.

(d) Both the algorithms rank the nodes in descending order of their importance and give same result.

Answer: (d)

Choose the correct statement with respect to the pagerank matrix.

(a) Pagerank matrix is symmetric

(b) Sum of elements in each column is 1.

(c) Pagerank matrix is same as the adjacency matrix of a graph.

(d) Sum of elements in each row is 1.

Answer: (b)

Given a complete network having 4 nodes. We take a random walk of length 1 million on this network. Every

time we arrive on a node, we gift it a gold coin. The approximate number of gold coins each node collects at

the end of this experiment is

(a) 1 million each

(b) Two nodes collect half of a million gold coins and two nodes remain empty handed

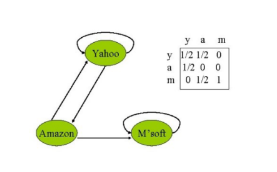
(c) Quarter a million each

(d) Since the experiment is probabilistic, nothing can be said

Answer: (c)

Solution: The network is complete so every node is equally important. So, we roughly give an equal number of coins to all the nodes. So, its 1/4th of a million each.

A gold coin distribution game is played on the following network. When the game converges,



(a) Yahoo, Amazon and M’soft, each collects one third of the coins

(b) M’soft collects all the coins

(c) Yahoo and M’soft together gets all the coins and Amazon gets none

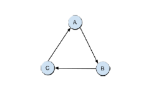
(d) All the coins are lost

Answer: (b)

Solution: M’soft redistributes all coins to itself. Hence, it eventually ends up having all the coins in the system.

In the graph G shown in following figure, assume that the current pagerank values of A, B and C are 0.3, 0.3

and 0.4 respectively. What will be their pagerank values after one iteration?



(a) A : 0.4,B : 0.3,C : 0.3

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

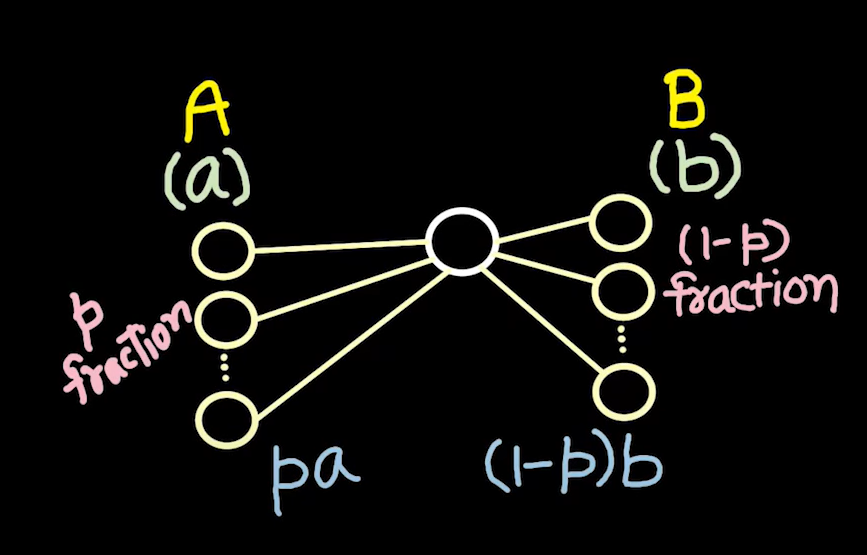
(c) A : 0.3,B : 0.3,C : 0.4

(d) A : 0.3,B : 0.4,C : 0.3

Answer: (a) A gets the point of C, B gets the point of A and C gets the points

of B.





A and B are certain actions, like A (going to library), B (going for shopping)

And a, b are the payoffs received after doing A and B respectively.

Now, if p fraction of my friends, do A, then the payoff = pa

And (1-p) fraction payoff doing B = pb   
Now, I will choose among A and B based on the two payoffs.   
If I do A, it means:

pa >= (1-p)b  
p >= b/b+a  
So, p is the threshold. Meaning, p fraction of my friend has to do A, for me to do A  
  
**Cluster Density:**

Density of a cluster is d, if for every node in a cluster it has at least d fraction of friends inside that cluster.

**Cascade:**

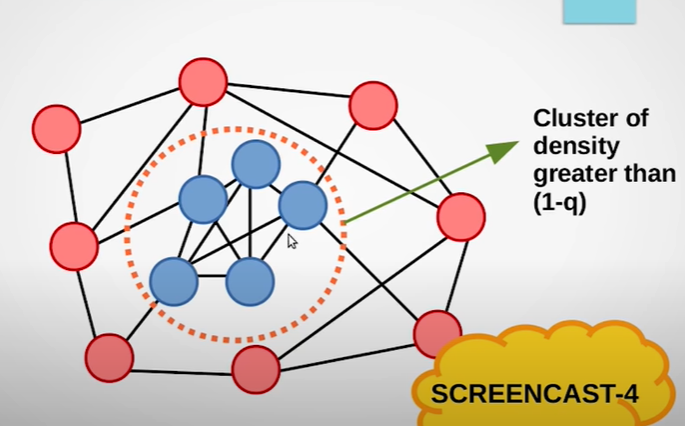
It is the process in which when a node adopts certain action, then others also tends to do same and this spreads across the network.

q-> threshold of adoption

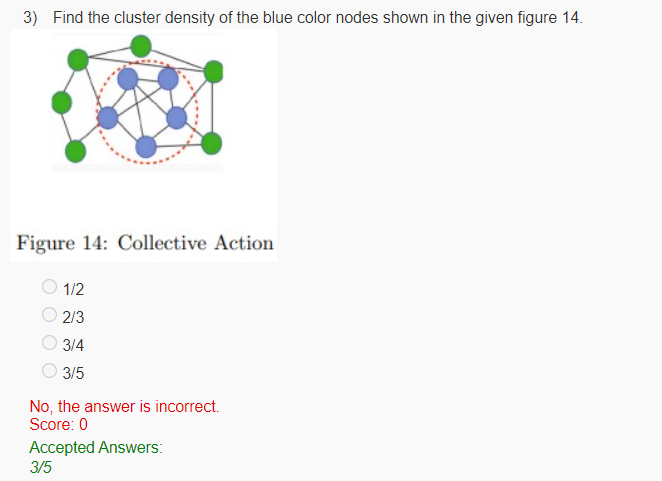
***If q fraction of my friend adopt an action, I will also do that action***

***cascade will not complete itself i.e., it cannot spread through entire network if there exists a cluster of density > 1-q in the network.***

***It also means,  
if a cascade is incomplete => there exists a cluster of density > 1-q***

******

**Questions:**

****

Which of the following are valid factors for influence diffusion?

 Payoff

 Key people

 Both Payoff and key people

 Neither Payoff nor key people

No, the answer is incorrect.  
Score: 0

Accepted Answers:

*Both Payoff and key people*

Consider the basic coordination based game theoretical model for cascading behavior. The payoffs associated with behaviours ‘A’ and ‘B’ are a and b respectively. Consider two friends x and y adopting behaviours ‘A’ and ‘B’ respectively. Assuming that x and y have no other friends than each other, the total payoff which they receive is

 ab

 a + b

 a/b

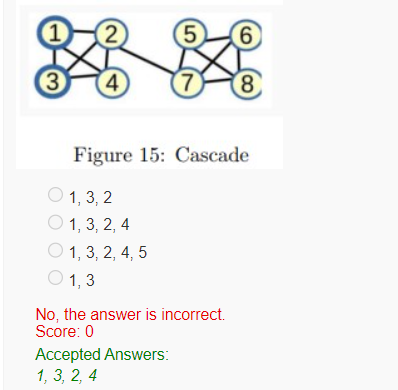
 b/(a + b)

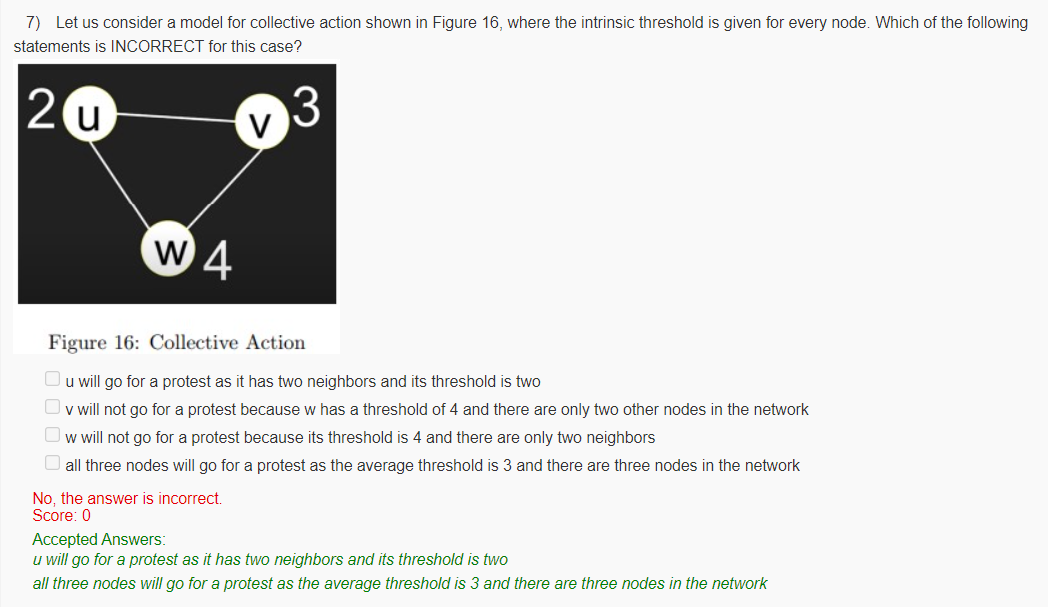
No, the answer is incorrect.  
Score: 0

Accepted Answers:

*a + b*

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 the 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, which all are the nodes who have adopted the behavior A.

**

******

***Here, intrinsic threshold basically means for u let’s say, it will go to protest only if two people including itself will go.***

**Doubt:**

******

The threshold for any node to switch from B to A in the A-B coordination game is the fraction of its neighbors that must be using A in order for it to switch. This threshold is calculated as follows:

threshold = b / (a + b)

where b is the payoff for using B and a is the payoff for using A.

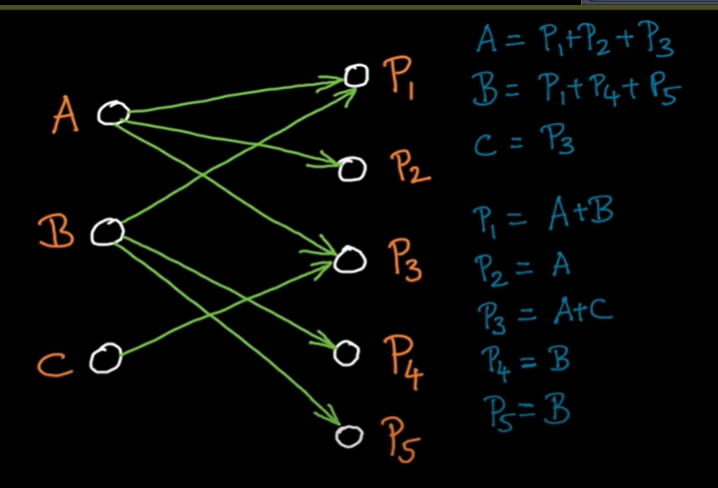
In the case of the social network in Figure 17, with a payoff of a = 3 and b = 2, the threshold for any node to switch from B to A is:

threshold = 2 / (3 + 2) = 2/5

Therefore, any node in the network will switch to A if and only if at least two of its neighbours are already using A.

**Principle of repeated improvement**

refers to the iterative process of enhancing and optimizing various aspects of a social network platform to provide a better user experience, foster user engagement, and ensure the long-term success of the network.



Here, let’s say A, B and C are people

P1, P2, P3, P4, P5 are the places

A recommending P1, P2, P3 and so on  
So, A gets point = P1 + P2 + P3  
and P1 gets the point of people who recommend = A + B

**Hubs and Authorities:**

Hubs -> are the nodes where we get information (A, B and C)

Authorities -> Hubs point to these authorities (P1, P2 , P3, P4 and P5)

A node can be hub or authority. How important a node is measured by

What does it point to and second is who are pointing to it.

With repeated improvement , the point accumulated in the node converges that is the proportion of point a node takes with respect to other nodes becomes same.

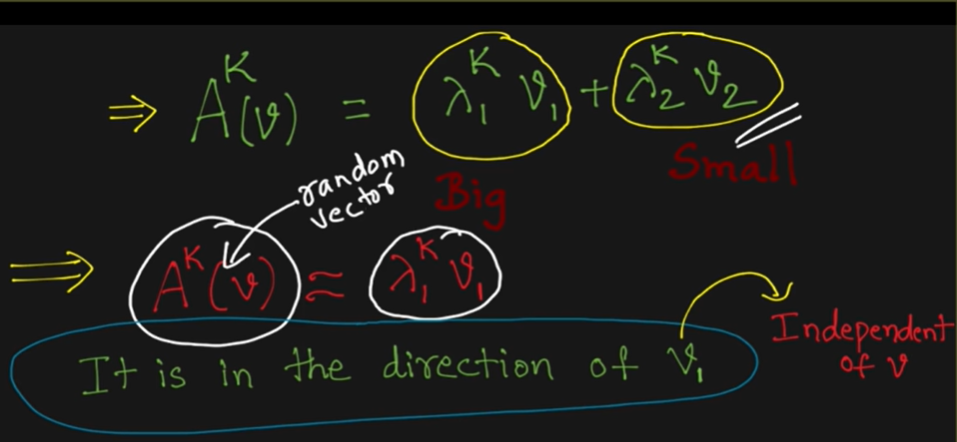
**Convergence in matrix multiplication:**

Let there be meatrix A and B  
A x B = C  
now if we multipy A x C = D   
Now if keep on multiply the product with A, then the there will come a point of convergence, when the value will not change.

No matter what vector we choose, the resultant multiplication of the matrix will always result in the same vector (direction).

**Addition of Two Vectors:**

On adding small and big vector, the resultant is close to the big vector, both in terms of direction and magnitude.

****

**Markov matrix:**

largest eigen value is 1

sum of any coumn is 1

Identify whether the statements are True.  
  
Statement I - When a matrix is applied on its eigenvectors, the direction of the eigenvector only changes  
Statement II - Eigenvectors are linearly independent of each other

 I only

 II only

 II only

 None

No, the answer is incorrect.  
Score: 0

Accepted Answers:

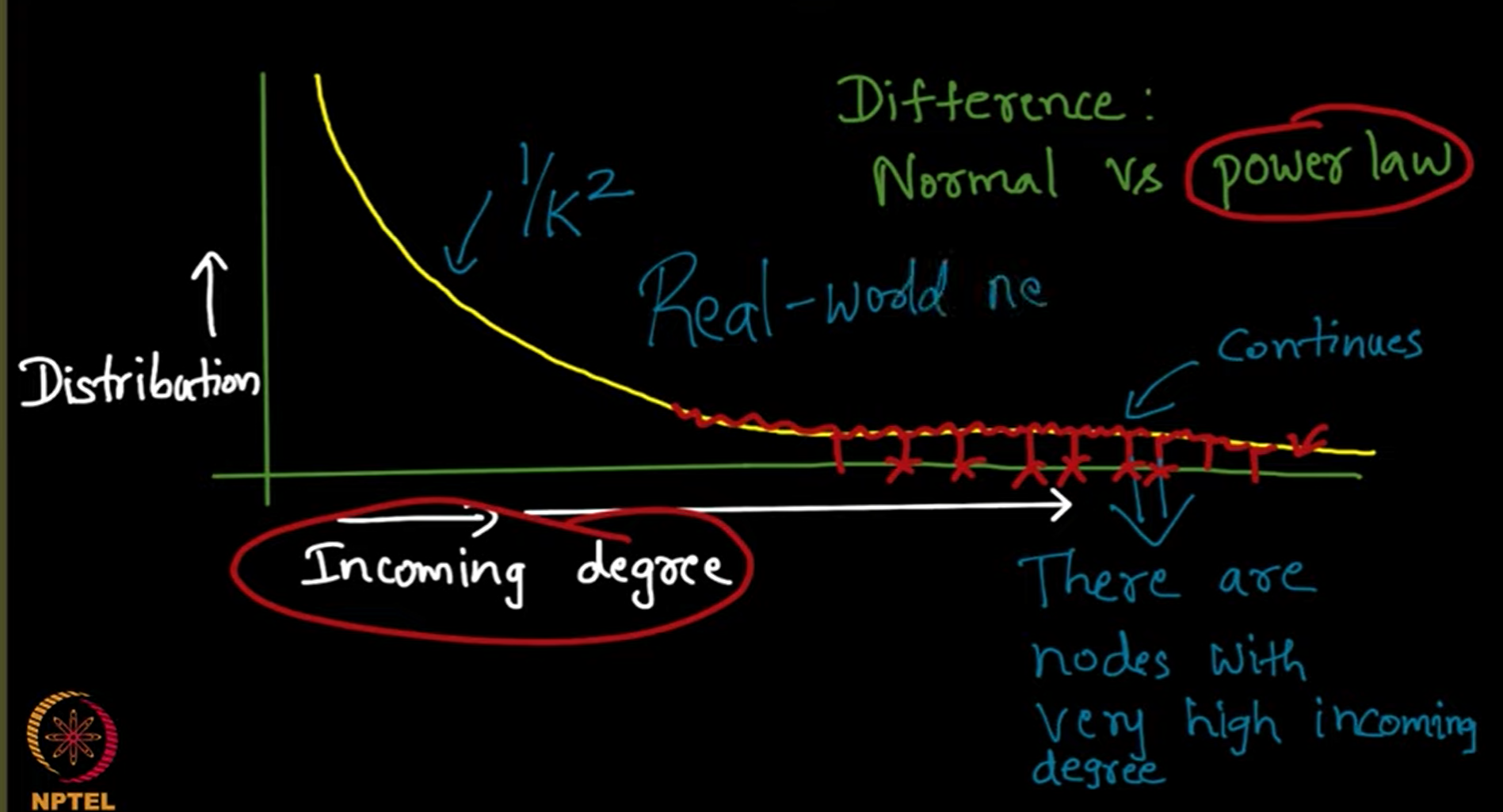
*II only*

Sum of few random variables gives Normal Distribution ( Bell curve graph)

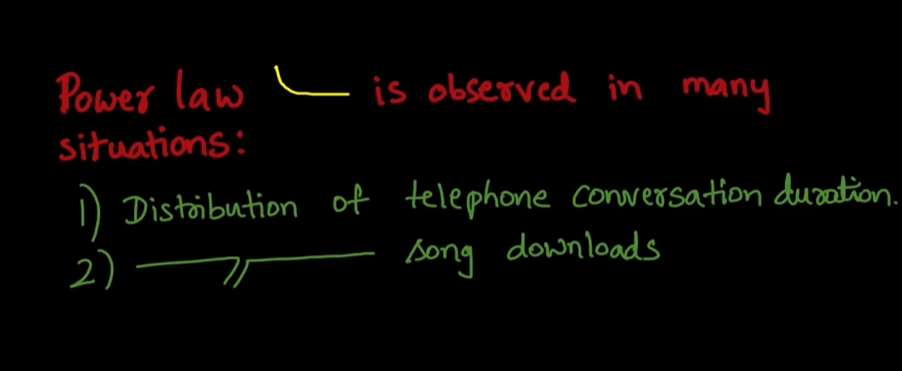
It’s observed in examples:

1. Height of town

This is called central limit theorem.



Other than web graph, power law distribution is seen in:



**Detecting the presence of power law:**

If there is a plot between log f(k) -> y axis and logk -> x axis, and there is straight line, ie of the form y=mx then that graph exhibits power law.

i.e, take the loglog plot of the given graph , if it’s straight line, then it’s power law.

e.g., f(k) = 1/ka

take log both side, gives

logf(k) = -a logk

of form, y=mx so, it’s a power law

Power law happens because of one of the hypotheses called **Preferential Attachment**

A node with higher node will attract more nodes to it.

This process is called rich get richer phenomenon.

**Implementing a Random Graph (Erdos- Renyi Model) ( Normal Distribution)**

Each edge has fixed probability of whether it will be added to the network or not.

Removing nodes which are more connected, makes the graph disconnected quickly.

**Questions:**

Consider the degree distribution graph plotted with degree of nodes along x axis and percentage of nodes with that degree along y axis. Given a graph with 500 nodes and edges with probability 0.5, where does this graph peak?

 250

 10

 100

 400

Yes, the answer is correct.  
Score: 1

Accepted Answers:

*250*

The degree distribution graph peaks at n\*p => 500\*0.5 = 250

What is the resultant distribution for Erdos Renyi model and Barabasi Albert model respectively?

 normal, normal

 normal, power law

 power law, normal

 power law, power law

Yes, the answer is correct.  
Score: 1

Accepted Answers:

*normal, power law*

Barabasi Albert model -> Rich gets richer phenomenon ( Power law graph)

In a random network, whether the nodes were removed randomly or selectively, the number of nodes to be removed to make the graph disconnected is similar. What are the possible reasons for this behavior?

 Network has hubs

 edges to this network were added preferentially

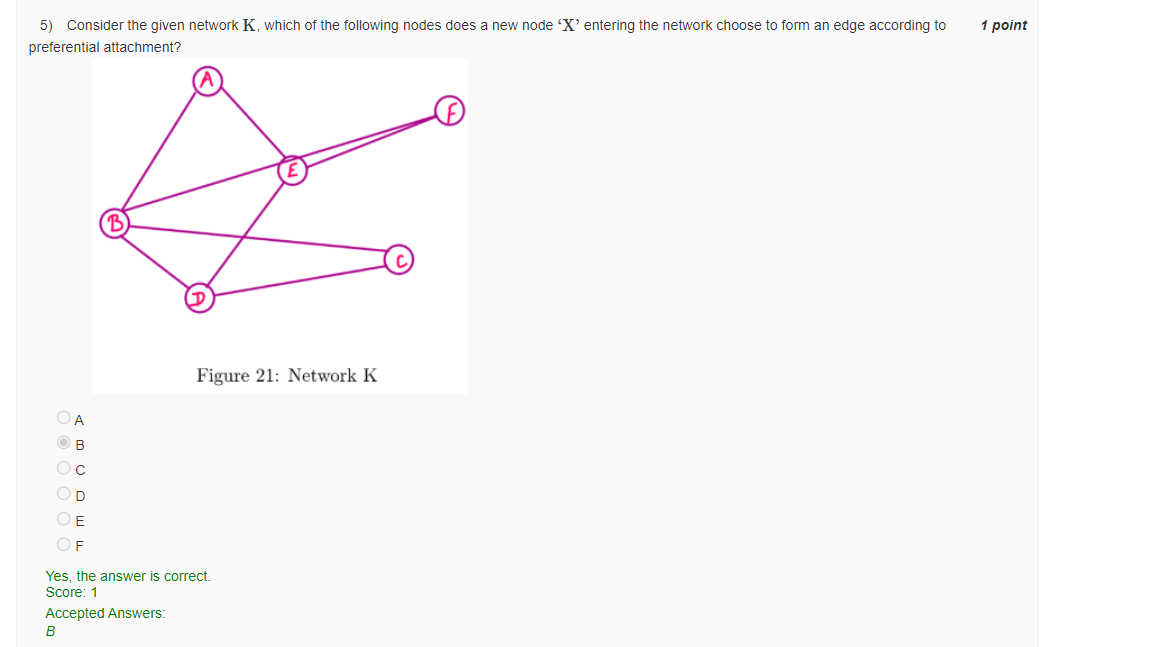
 edges were added randomly

 Network is dense

Yes, the answer is correct.  
Score: 1

Accepted Answers:

*edges were added randomly*

****

The node will get attached to the one having highest indegree.

What is the probability PP for the Erdos Renyi model?

 probability of a node to have self loop

 probability that two nodes are connected by an edge.

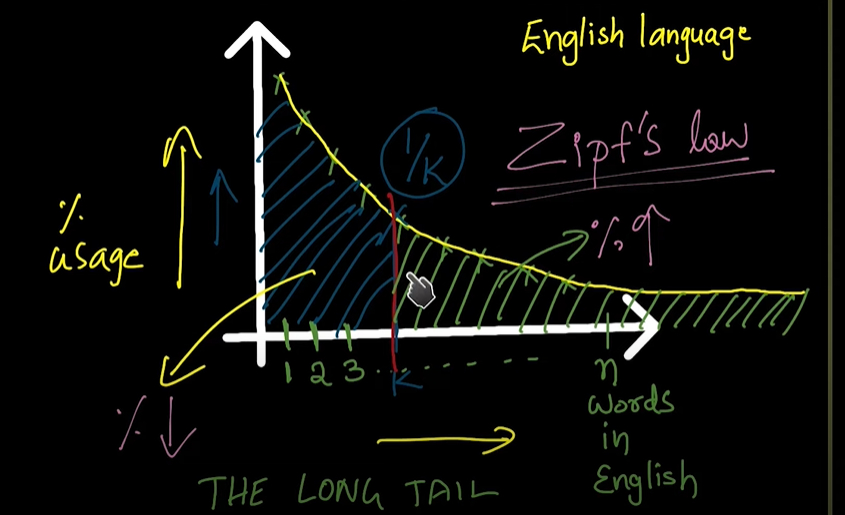
 probability that a node has a different attribute.

 probability that a node belongs to a specific community.

Yes, the answer is correct.  
Score: 1

Accepted Answers:

*probability that two nodes are connected by an edge.*

****

Long tail - Even though most of words are used less still they amount to more than the popular ones

**Epidemics**

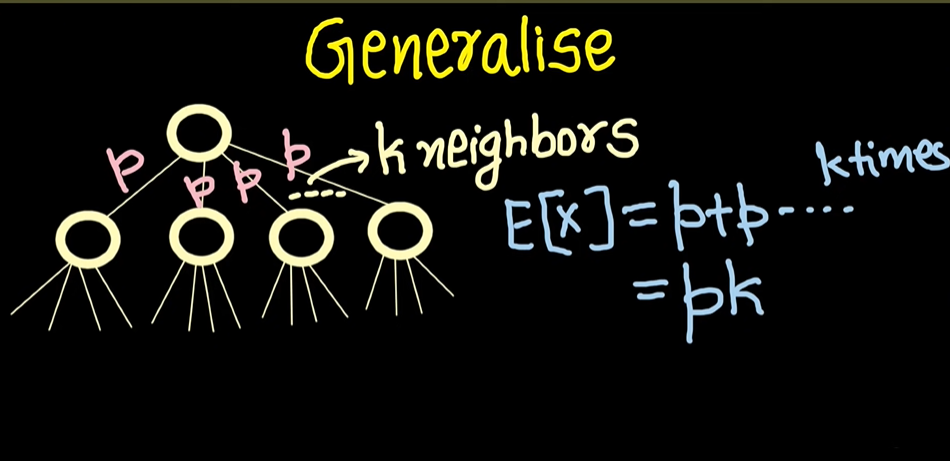
The type network depends on the type of pathogen involved.  
For, flu -> network will be dense

HIV -> network will be sparse

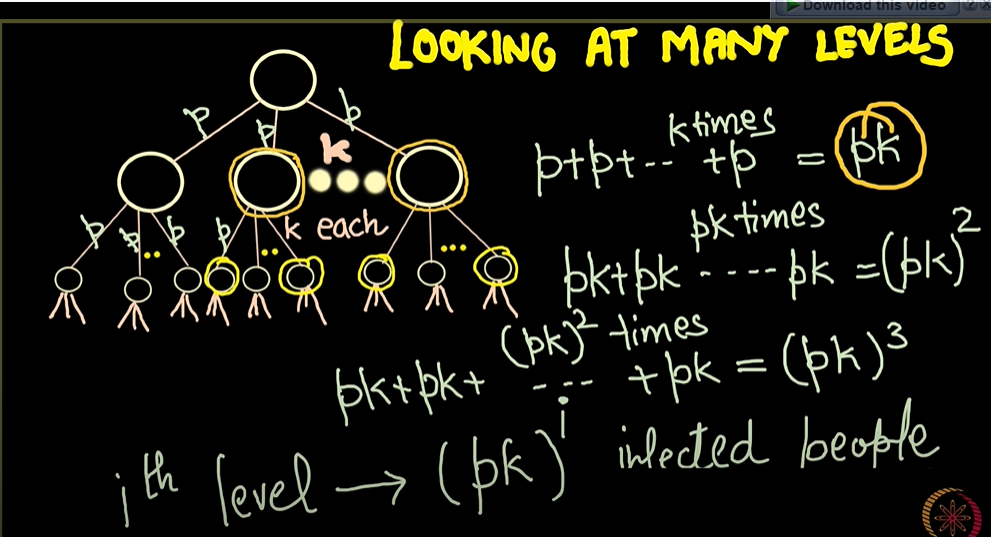
For modelling the spread of disease, two things are required:

1. Pathogen involved ( How contagious it is)
2. Network structure

**Branching Model**



If the root has the probability p, by which it can affect each neighbour, then the expected number of neighbours getting affected will be = pk -> secondary infection



pk = **Basic Reproductive Number (R0)**

R0 < 1 : Infection dies away with probability = 1

R0 > 1 : Infection persists in the network with positive probability > 0   
 Here too infection can die away

This property of R0 is called **knife edge property**

**SIR Model**

It talks about 3 life cycle of an infection.

1. Susceptible
2. Infected
3. Recovered/Removed

The nodes in the network can be in one state at a time.

**SIS Model**

1. Susceptible
2. Infected
3. Susceptible

It happens in case of infection like common cold, where it can happen to a person many times even after getting recovered

***SIR vs SIS***

SIR -> the process will come to an end

SIS -> process keep running forever

**Percolation Model:**

* **Static view of SIR model**

Here, in the start itself we get all the nodes that could be infected by certain probability.

**Questions:**

In the percolation model (static view of the SIR model), assume that TI=1, For every edge Eu,v in the network, we toss a biased coin which shows head with a probability of p, which is the infection rate of the disease, i.e., the probability that v will become infected in the next iteration, given that u is infected. If head turns up, we assume an edge to be open, else blocked. According to this percolation model, a node w in the network will become infected

iff there is a path consisting of blocked edges from any of the initially infected nodes to w..

iff there is a path consisting of open edges from any of the initially infected nodes to w..

iff there is a path from any of the initially infected nodes to w.. The path may consist of any edgesopen/ blocked.

iff there does not exist any path from any of the initially infected nodes to w..

Yes, the answer is correct.  
Score: 1

Accepted Answers:

*iff there is a path consisting of open edges from any of the initially infected nodes to*w.

Consider a branching model of network where every node has 5 children and the probability of transmission of an infection is 0.4, what is the probable number of people infected at level 4?

 24

 16

 20

 5

Yes, the answer is correct.  
Score: 1

Accepted Answers:

*16*

Answer = (pk)level

In a branching model, Basic reproductive number helps to identify if a disease will be an epidemic or not. Which of the following statements is/are true?  
Statement I - If R0<1, then the disease dies out from the network with a probability 1  
Statement II - If R0>1, then the disease persists in the network with a probability greater than 0.

 I only

 II only

 Both I and II

 None

Yes, the answer is correct.  
Score: 1

Accepted Answers:

*Both I and II*

Suppose the basic reproductive number is estimated to be R0=1.2.R0=1.2. If a vaccine giving 100% immunity is available next time and a fraction v = 0.4 of randomly selected individuals were vaccinated, an estimate of the new reproductive number would be

 3

 0.72

 0.48

 1

Yes, the answer is correct.  
Score: 1

Accepted Answers:

*0.72*

**Explanation**  
The basic reproductive number, R₀, represents the average number of secondary infections produced by a single infected individual in a completely susceptible population. When a fraction of the population is vaccinated, we can calculate the effective reproductive number, Rₑ, which represents the average number of secondary infections produced by a single infected individual in a partially immune population.

**The formula to calculate Rₑ when a fraction v is vaccinated and the vaccine provides 100% immunity is:**

**Rₑ = R₀ \* (1 - v)**

In this case, R₀ is 1.2, and v is 0.4 (40% of randomly selected individuals were vaccinated):

Rₑ = 1.2 \* (1 - 0.4) Rₑ = 1.2 \* 0.6 Rₑ = 0.72

**Small World Theory**

**Milgram’s Experiment**

Any two people in the world are separated on average by 6 people

**Watts – Strogats model:**

Small – world phenomenon is observed due to:

1. Homophily
2. Weak ties

In this model, graph is generated in the following way: every node is connected to their immediate or geographically closer neighbour, and after this random rewiring of some of the edges is done (Delete a random edge and add a new random edge in a network).

Search on Small World Network (SWN) is decentralised. This decentralised search is called **Myopic Search** as we are only concerned with the nodes at the nearest.

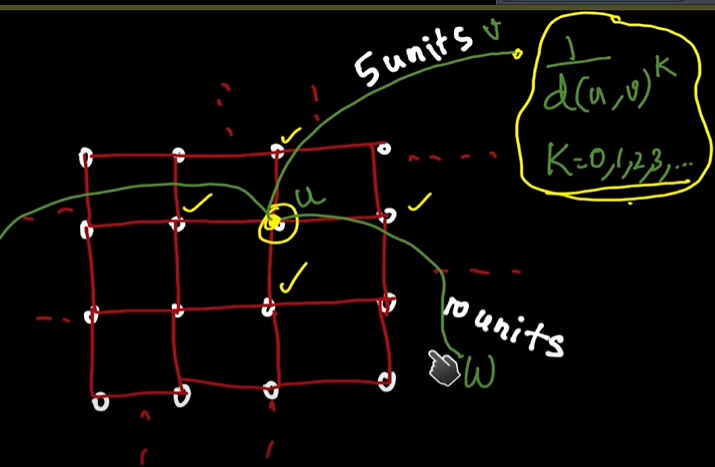
In Myopic Search, while searching, we assume:

* We don’t know about the long-range contact/weak ties of a node

And since we are not aware of these weak ties, therefore it’s possible that their could exists a path between a source node and target node that is shorter than what we got from myopic search.

*Myopic Search time increase logarithmic with increase in nodes.*

The search is done using *Local Method,* it means a person only know its immediate friend, and that immediate in turn also know it’s immediate friend and that’s how the end person or destination is found.

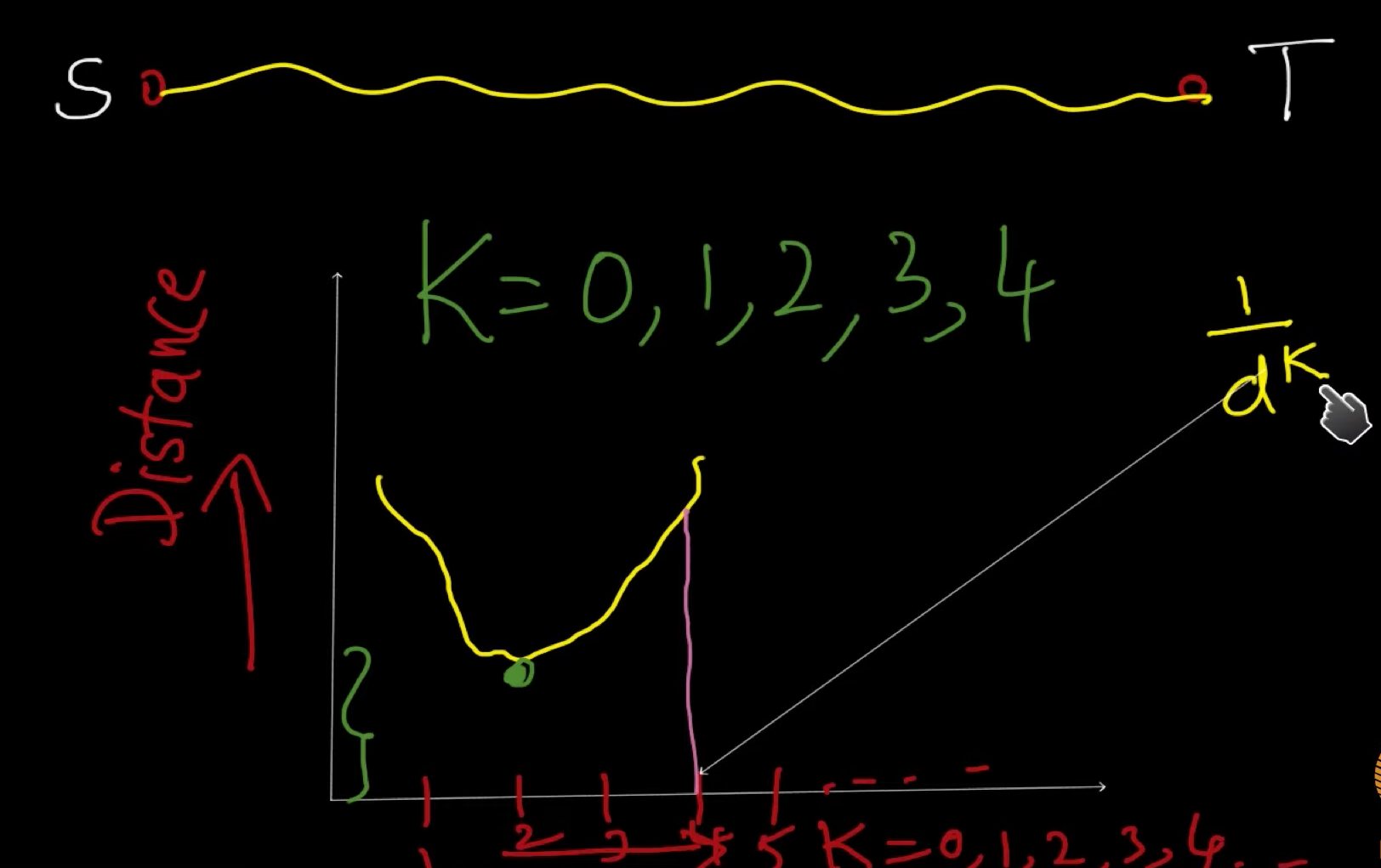


A node ‘u’ will make connection/friends with and node ‘v’ is base on the probability of distance between them, it’s given by

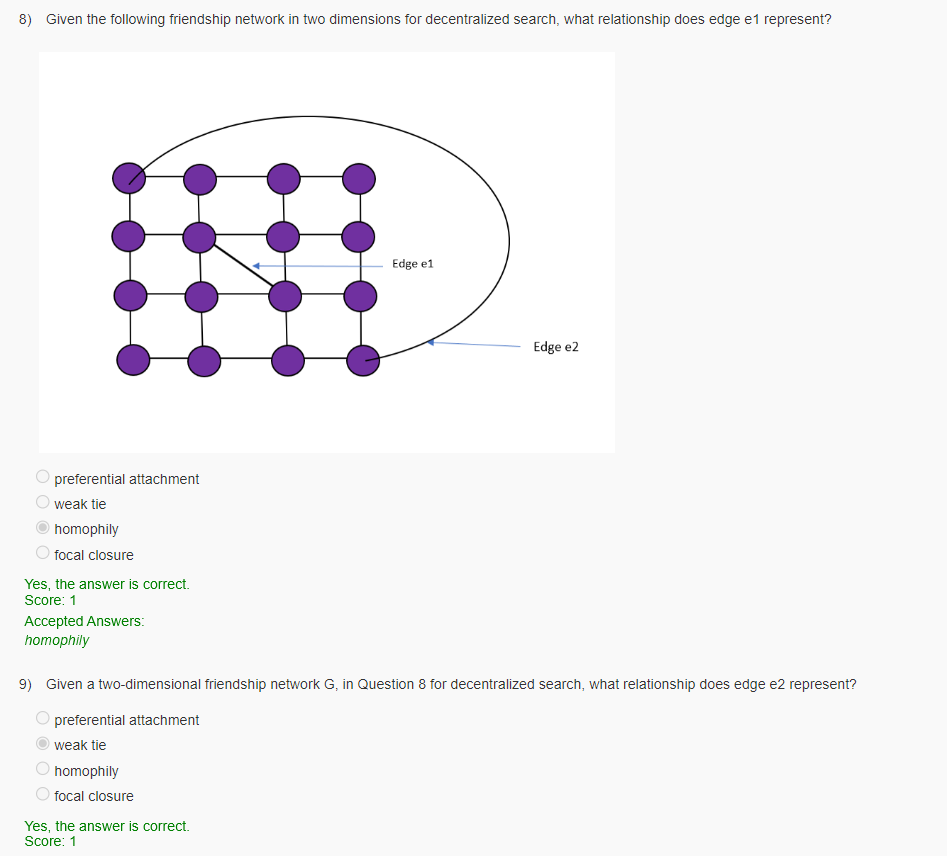
K = 0, 1, 2, …

k=2, distance is the shortest, in 2d graph

k=1, distance is the shortest, in 1d graph



**Questions:**

****

In a decentralised search, what is the ideal value for kk in a one dimensional network and two dimensional grid repectively to obtain the shortest distance between two nodes?

 1, 2

 1, 1

 2, 2

 2, 1

Yes, the answer is correct.

Which of the following ties are required to perform a decentralized search?

 Only strong ties

 Only weak ties

 Both strong and weak ties

 neither strong nor weak tie

Yes, the answer is correct.

Assume that each of your friends has 20 friends other than you. Similarly, each of their friends has 20 friends other than them and so on. Then, how many people can you reach in i levels (Level one refers to your friends, level 2 refers to your friends’ friends and so on)?

20i20

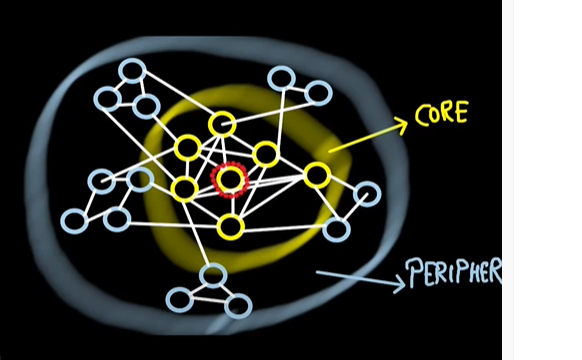
20i20

i2020

20202020

Yes, the answer is correct.

**Core-periphery structure**

****

**If the core node is infected then the content will get viral, as the core nodes are connected to all the nodes(pheriphery)**

1 – core -> degree of each node in subgraph >= 1

2 – core -> degree of each node in subgraph >= 2

Finding the key node or the core node is one of the ways to spread an idea in a network other way is the novelty/quality of the idea

**K-shell decomposition algorithm**

It is the way of finding 1-cores, 2-cores,… in the network.

To find 1-core:

In the network remove the nodes which are of degree 1, and keep them in bucket 1- core, and keep doing this till no 1 degree node is left. The left part of network will now have only the nodes of degree >= 2.

Now, we do the same process for degree 2 nodes, and put these 2 degree nodes in bucket labelled 2-core.

After these process, when all the nodes are removed, then the last core bucket is the the core of the network.

Let’s say on removing 3 degree nodes the network get empty. Then the core is 3

At the end,

1-core = 1-core bucket union 2-core bucker union 3-core bucket

2-core = 2-core bucker union 3-core bucket

3-core = 3-core bucket

k-core of network = Bucket (k) union Bucket (k+1) + ….

Cascading capacity of a shell

We choose a shell that is any core, then the number of nodes it affect is its cascading capacity.

It’s observed that, if we plot a graph between shell number (x-axix) and number of nodes they infect(y-axix), then the graph first increases and then after some particular shell number it becomes constant.

So, it means that to spread any idea in a network we don’t really need the innermost core to infect the network.

Other than these innermost cores, there exists many outermost cores which can be used to infect the network.

These are called **pseudo cores**.

**Questions:**

Which of the following is INCORRECTINCORRECT when creating edges that represent homophily in the network?

 add edges to one node on the left and one on the left of every node in a ring network

 add edges to two nodes on the left and two on the left of every node in a ring network

 add random edges between any two nodes in the network

 add edges between nodes that are geographically close

Yes, the answer is correct.  
Score: 1

Accepted Answers:

*add random edges between any two nodes in the network*

Considering myopic search, which of the following propositions is true?

 Globally optimal solutions are always identified by myopic search.

 Locally optimal solutions are always discovered through myopic search, but not always the most ideal ones worldwide.

 A depth-first search algorithm is a type of myopic search.

 Myopic search is a form of heuristic search algorithm that calculates the distance to the target using a heuristic function.

Yes, the answer is correct.  
Score: 1

Accepted Answers:

*Locally optimal solutions are always discovered through myopic search, but not always the most ideal ones worldwide.*

Which of the following statements is true?  
Statement I - Myopic search requires more computational resources as compared to optimal search  
Satement II - Myopic search may not find the optimal solution

 I only

 II only

 Both

 none

No, the answer is incorrect.  
Score: 0

Accepted Answers:

*II only*

Select all the characteristics of the Decentralised search.

 Collaborative effort in searching

 There is a central node which exactly performs the search

 Google is an example of a Decentralised search engine

 effort is distributed across multiple nodes

Yes, the answer is correct.  
Score: 1

Accepted Answers:

*Collaborative effort in searching*

*effort is distributed across multiple nodes*

Which of the following contains the most influential nodes according to the KK-shell algorithm?

 Higher shell number

 Lower shell number

 Shell number 1

 None of the above

Yes, the answer is correct.  
Score: 1

Accepted Answers:

*Higher shell number*

Which of the following statements are true?  
Statement I - The influential power of a set of nodes is the probability with which they cannot spread disease if they get injected with it first.  
Statement II - The influential power of a set of nodes XX is the number of nodes that get infected with the disease if the disease starts with X.X.

 I only

 II only

 Both

 None

Yes, the answer is correct.  
Score: 1

Accepted Answers:

*II only*