Q. 5] How to discover communities in Social-Network Graph directly?

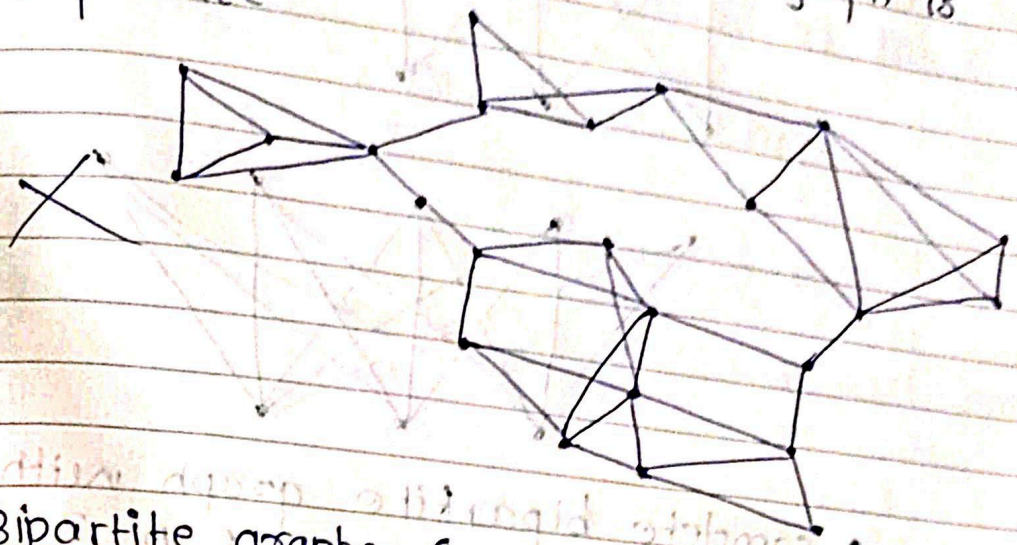
→ (1) It is not possible to place an indivisual in two different communities. so it need a technique for discovering community directly.

(2) A] clique (graph theory):

In the mathematical area of graph theory a clique is a subset of vertices of an undirected graph such that every two distinct vertices in clique are

adjacent ; i.e. its induced subgraph is complete.

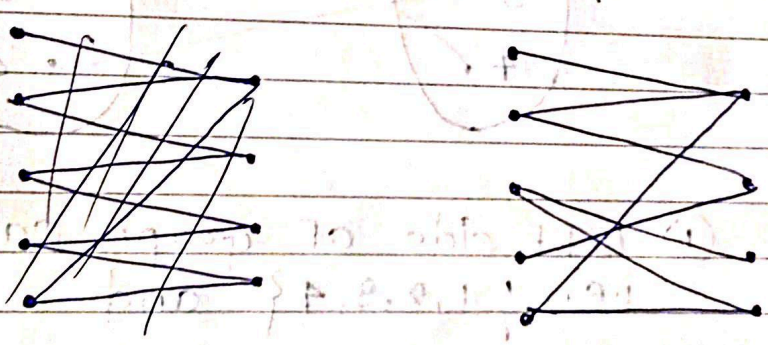
[B] ~~Bipartite~~



[B] Bipartite graphs: (graph theory):

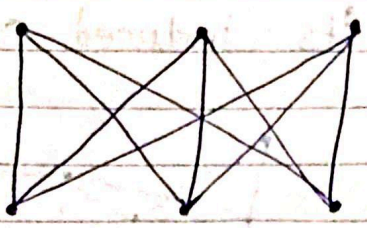
① In the mathematical field of graph theory a bipartite graph is graph ~~ph~~ whose vertices can be divided into two disjoint & independant sets U & V such that every edge connects a vertex in U to one in V.

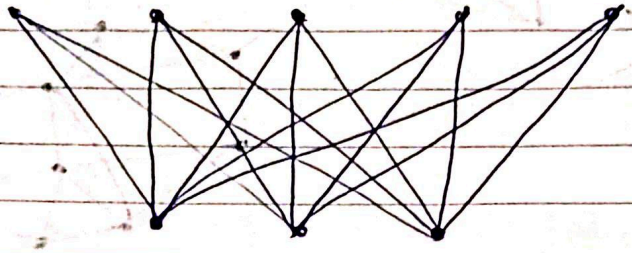
② vertex sets U and V are called parts of graph



[C] compete Bipartite graphs

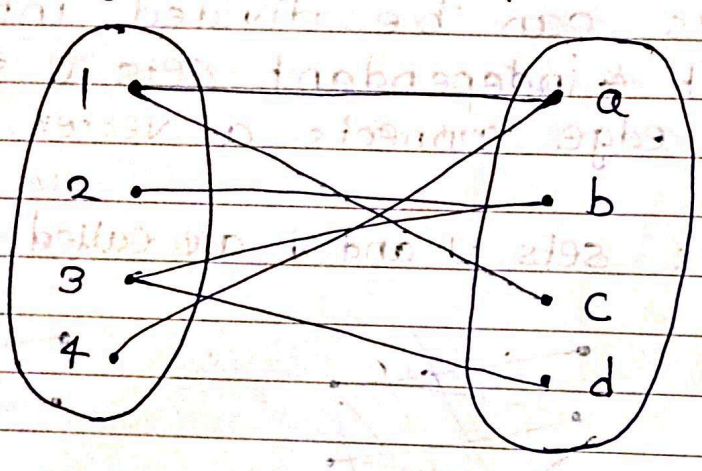
In math. field of graph theory, a complete bipartite graph or biclique is special kind of bipartite graph where every vertex of first set is connected to every vertex of second set.





A complete bipartite graph with
$m = 5$ & $n = 3$

D] Finding complete bipartite subgraphs:-



① Left side of graph called as items i.e. {1, 2, 3, 4} and

right side of graph called as basket i.e. {a, b, c, d}

② basket a consist of item 1 and 4 like that;
$a = \{1, 4\}$, $b = \{2, 3\}$, $c = \{1\}$ and $d = \{3\}$

(B) If s=2 and t=1 must find item-sets of size 1 that appear in at least two baskets.

(4) {1} is one such item-set; {3} is another.

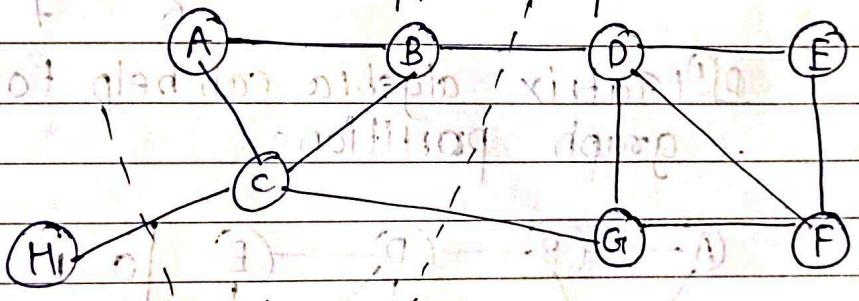
Q.6] How social Network graph can be partitioned to identify - communities?

→ A] To partitioned a graph, we will use important tools from matrix theory.

B] (1) objective:- (1) minimize number of edges that connect different components.

(2) Divide nodes into two sets so that cut or set of edges that connect nodes in different sets is minimized.

(3) Two sets are approx. equal in size.



smallest cut Best cut

C] (1) Normalized cuts:-

~~(1) vol(S)~~ -

(1) volume of set s (vol(S)) - to be no. of edges with at least one end in S.

(2) Let ~~cut~~ cut(S,T) be number of edges that connect a node in S to a node in

Then normalized cut value for S and T is

$$\frac{cut(S,T)}{vol(S)} + \frac{cut(S,T)}{vol(T)}$$

ex. ⇒ (Refer above graph)

① If we choose $S = \{H\}$ & $T = \{A, B, C, D, E, F, G\}$ then

$cut(S,T) = 1$

$vol(S) = 1$ & $vol(T) = 11$

normalized cut $= \frac{1}{1} + \frac{1}{11} = 1.09$

② ~~othe~~ consider other cut, $S = \{A, B, C, H\}$

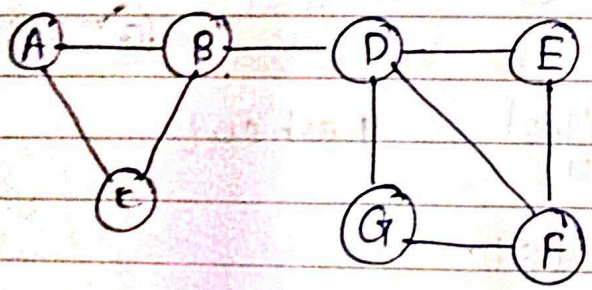
$T = \{D, E, F, G\}$

$cut(S,T) = 2$ & $vol(S) = 6$

$vol(T) = 7$

normalized cut $= \frac{2}{6} + \frac{2}{7} = 0.62$

D① Matrix algebra can help to find good graph partitions



$$\begin{bmatrix} 0 & 1 & 1 & 0 & 0 & 0 & 0 \\ 1 & 0 & 1 & 1 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 1 & 1 & 1 \\ 0 & 0 & 0 & 1 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 & 1 & 0 & 1 \\ 0 & 0 & 0 & 1 & 0 & 1 & 0 \end{bmatrix}$$

adjacency matrix

② The second matrix we need is degree matrix for graph.
This graph has nonzero entries only on diagonal.

$$\begin{bmatrix} 2 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 3 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 2 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 4 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 2 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 3 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 2 \end{bmatrix}$$ ← degree matrix

③ adjacency matrix → A &
degree matrix → D then
Laplacian matrix $L = D - A$.

$$\begin{bmatrix} 0 & 1 & 1 & 0 & 0 & 0 & 0 \\ 1 & 0 & 1 & 1 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 1 & 1 & 1 \\ 0 & 0 & 0 & 1 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 & 1 & 0 & 1 \\ 0 & 0 & 0 & 1 & 0 & 1 & 0 \end{bmatrix} - \begin{bmatrix} 2 & & & & & & \\ & 3 & & & & & \\ & & 2 & & & & \\ & & & 4 & & & \\ & & & & 2 & & \\ & & & & & 3 & \\ & & & & & & 2 \end{bmatrix} = \begin{bmatrix} 2 & -1 & -1 & 0 & 0 & 0 & 0 \\ -1 & 3 & -1 & -1 & 0 & 0 & 0 \\ -1 & -1 & 2 & 0 & 0 & 0 & 0 \\ 0 & -1 & 0 & 4 & -1 & -1 & -1 \\ 0 & 0 & 0 & -1 & 2 & -1 & 0 \\ 0 & 0 & 0 & -1 & -1 & 3 & -1 \\ 0 & 0 & 0 & -1 & 0 & -1 & 2 \end{bmatrix}$$

adjacency L

④ Best way to Laplacian matrix can obtain a partition by taking one set to be the node i whose corresponding vector x_i is +ve & other whose component -ve.

so Laplacian matrix →

$$\begin{bmatrix} 3 & -1 & -1 & -1 & 0 & 0 \\ -1 & 2 & -1 & 0 & 0 & 0 \\ -1 & -1 & 3 & 0 & 0 & -1 \\ -1 & 0 & 0 & 3 & -1 & -1 \\ 0 & 0 & 0 & -1 & 2 & -1 \\ 0 & 0 & -1 & -1 & -1 & 3 \end{bmatrix}$$

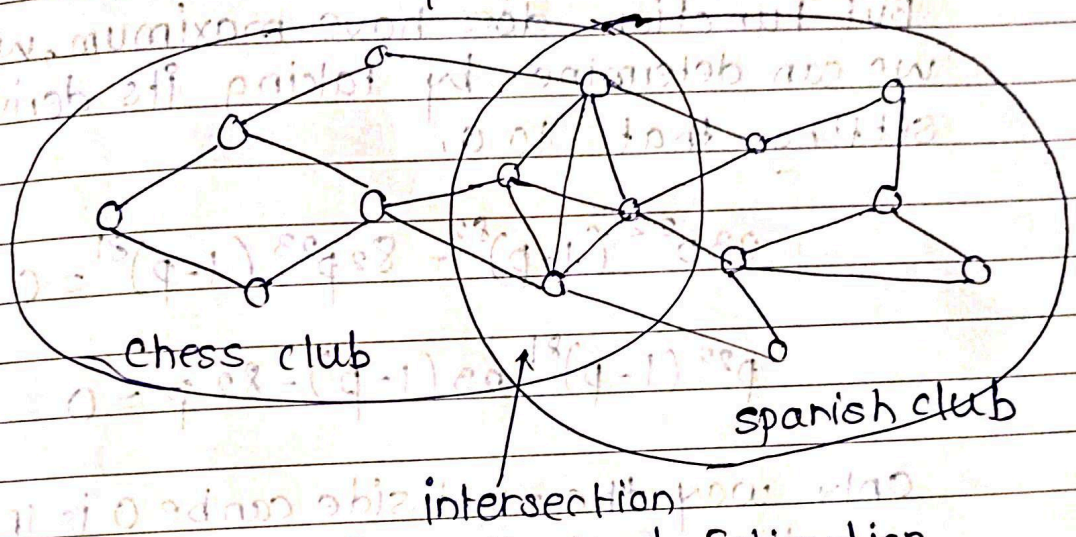
E] Alternative partitioning method :-

Q.1] How to find overlapping communities in Social Network Graph?

→ Finding overlapping communities :-

1] This assumes that probability that two Indivisuals are connected by an ~~ang~~ edge ("friends") increases as they become members of more communities in common.

2] Nodes are people & there is an edge bet^n two nodes if people are "friends".



3] MLE :- Maximum-Likelihood Estimation

(i) Model has parameter that determine probability of generating any particular instance of artifact.

(ii) This probability is called likelihood of those parameter values.

(iii) We assume that value of parameters that gives largest value of likelihood is correct model for observed artifact.

(iv) In graph, each edge is present with probability p & not not present with probability 1-P.

(v) The only parameter we can adjust is p.

(vi) Following MLE principle, true value of p is the one for which the probability of generating the observed graph is the highest.

(vii) In above graph, nodes = 15
edges = 23
105 pairs of 15 nodes.
probability of generating exactly the graph is given by function $p^{23}(1-p)^{82}$

But function does have maximum, which we can determine by taking its derivatives setting that to 0.

$$23p^{22}(1-p)^{82} - 82p^{23}(1-p)^{81} = 0$$

$$p^{22}(1-p)^{81}\,23(1-p) - 82p = 0$$

only way the right side can be 0 is if p is 0 or last factor $23(1-p) - 82p = 0$

Likelihood of generating graph is maximized when $23 - 23p - 82p = 0$ or

$$p = \frac{23}{105}$$

4] AFFiliation - Graph model :-
once we see how parameter of model influence likelihood of seeing a given graph, we can address how one would solve for values of parameter that give maximum likelihood.

Q.8] Explain Affiliation-Graph model to find overlapping communities in social-Network Graph.

Q.9] Why triangles in social-Network Graph are counted? Explain algorithm for finding triangles in social Network graph

A]

→ ~~① consider Graph with n nodes & m edges~~

A] ① counting the number of triangles helps us to measure extent to which a graph looks like a social network.

B] Algorithm:-

① consider a graph of n nodes & m≥n nodes & nodes are integers 1, 2, - - n

② call a node a heavy hitter if its degree is at least $\sqrt{m}$

③ heavy-hitter triangle is triangle whose all three nodes are heavy-hitter.

④ since each node contribute to degree of only two nodes, there would then have to be more than m edges.

Assuming graph is represented by its edges, pre-process graph as follows:

1. compute the degree of each node. The total time required is O(m).
2. create an index on edges, with the pair of nodes at its ends as the key. A hash table suffices. It can be constructed in O(m) tim

3. create another index of edges, this one with key equal to single node.

We shall order the node as follows.

① First, order nodes by degree.

② If v & u have same degree, recall the both v & u are integers, so order them numerically

③ That is, we say $v < u$ if & only if either

1. The degree of v is less than degree of u or
2. The degree of u & v are same & $v < u$.

Heavy-Hitter Δleg :-

① There are only $O(\sqrt{m})$ heavy-hitter nodes, can consider all sets of three of these nodes.

② There are $O(m^{3/2})$ possible heavy hitter Δles, & using the index on edge we can check if all three edges exist in $O(1)$ time.

③ $\therefore$ $O(m^{3/2})$ time is needed to find all heavy-hitter Δles.