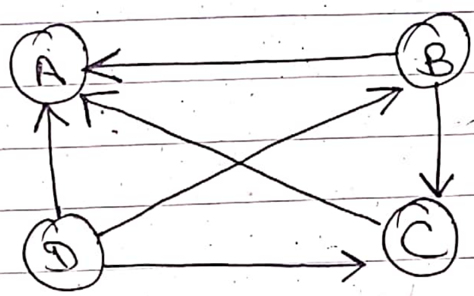
# Web System AND Algorithm:

## Unit: 2 Searching:

* Link Analysis:
  1. Page Rank Algorithm
  2. HITS (Hypertext Induced Topic Search) Algorithm.

① Page Rank Algorithm: $\left[ PR(u) = \sum_{v \in B_u} \frac{PR(v)}{L(v)} \right]$

Example: Assume the following graph & use page rank algorithm



= Sol

Initially, page rank of each page, $PR(A) = PR(B) = PR(C) = PR(D)$

$= \frac{1}{N}$

$= \frac{1}{4}$

$= 0.25$

① First Iteration:

$PR(A) = \frac{PR(v)}{L(v)}$

$= \frac{PR(B)}{L(B)} + \frac{PR(C)}{L(C)} + \frac{PR(D)}{L(D)}$

$= \frac{0.25}{2} + \frac{0.25}{1} + \frac{0.25}{3}$

$= 0.458$

$$PR(B) = \frac{PR(D)}{L(D)}$$

$$= \frac{0.25}{3}$$

$$= 0.083$$

$$PR(C) = \frac{PR(B)}{L(B)} + \frac{PR(D)}{L(D)}$$

$$= \frac{0.25}{2} + \frac{0.25}{3}$$

$$= 0.208$$

$$PR(D) = 0$$

<u>2nd Iteration:</u>

new rank: $PR(A) = 0.458$

$PR(B) = 0.083$

$PR(C) = 0.208$

$PR(D) = 0$

now,

$$PR(A) = \frac{PR(B)}{L(B)} + \frac{PR(C)}{L(C)} + \frac{PR(D)}{L(D)}$$

$$= \frac{0.083}{2} + \frac{0.208}{1} + \frac{0}{3}$$

$$= 0.2495$$

$$PR(B) = \frac{PR(D)}{L(D)}$$

$$= \frac{0}{3} = 0$$

$$PR(C) = \frac{PR(B)}{L(B)} + \frac{PR(D)}{L(D)}$$

$$= \frac{0.083}{2} + \frac{0}{3}$$

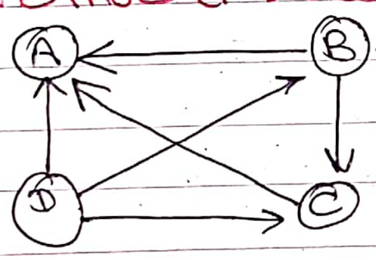
$$= 0.0415$$

$PR(D) = 0$

now,

| pages | Iteration 0 | iteration 1 | Iteration 2 | Page Rank |
|---|---|---|---|---|
| A | 0.25 | 0.458 | 0.2495 | 1 |
| B | 0.25 | 0.083 | 0 | 3 |
| C | 0.25 | 0.208 | 0.0415 | 2 |
| D | 0.25 | 0 | 0 | 4 |

Another method (matrix Representation): $[PR_{t+1} = H\ PR_t]$




Sol

$$H = \begin{array}{c|cccc} & A & B & C & D \\ \hline A & 0 & 1/2 & 1 & 1/3 \\ B & 0 & 0 & 0 & 1/3 \\ C & 0 & 1/2 & 0 & 1/3 \\ D & 0 & 0 & 0 & 0 \end{array}$$

[∵ Look for outlinks in each nodes]

Initial Page rank of all nodes $PR_0 = \begin{bmatrix} 0.25 \\ 0.25 \\ 0.25 \\ 0.25 \end{bmatrix}$

$PR_1 = H\,PR_0$

$$= \begin{bmatrix} 0 & 0.5 & 1 & 0.33 \\ 0 & 0 & 0 & 0.33 \\ 0 & 0.5 & 0 & 0.33 \\ 0 & 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} 0.25 \\ 0.25 \\ 0.25 \\ 0.25 \end{bmatrix}$$

$$= \begin{bmatrix} 0 + (0.5 \times 0.25) + (1 \times 0.25) + (0.33 \times 0.25) \\ 0 + 0 + 0 + (0.33 \times 0.25) \\ 0 + (0.5 \times 0.25) + 0 + (0.33 \times 0.25) \\ 0 + 0 + 0 + 0 \end{bmatrix}$$

$$= \begin{bmatrix} 0.4575 \\ 0.0825 \\ 0.2075 \\ 0 \end{bmatrix}$$

$PR_2 = H\,PR_1$

$$= \begin{bmatrix} 0 & 0.5 & 1 & 0.33 \\ 0 & 0 & 0 & 0.33 \\ 0 & 0.5 & 0 & 0.33 \\ 0 & 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} 0.4575 \\ 0.0825 \\ 0.2075 \\ 0 \end{bmatrix}$$

$$= \begin{bmatrix} 0+(0.5\times0.0825)+(1\times0.2075)+0 \\ 0+0+0+(0.33\times0) \\ 0+(0.5\times0.0825)+0+0 \\ 0+0+0+0 \end{bmatrix}$$

$$= \begin{bmatrix} 0.24875 \\ 0. \\ 0.04125 \\ 0 \end{bmatrix}$$

So, The page rank after $2^{nd}$ iteration is $A = 0.24875$

$B = 0$

$C = 0.04125$

$D = 0$

also, $|PR_2 - PR_1|$

$$= \begin{bmatrix} |0.24875 - 0.4575| \\ |0 - 0.0825| \\ |0.04125 - 0.2075| \\ |0| \end{bmatrix}$$

$$= \begin{bmatrix} 0.20875 \\ 0.0825 \\ 0.16625 \\ 0 \end{bmatrix}$$

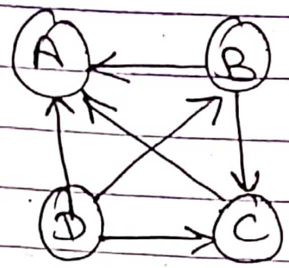
$= (0.20875 + 0.0825 + 0.16625 + 0)$

$= 0.4575 \rightarrow$ (Convergence criteria or threshold.)

Note: Last point of note.

# * HITS Algorithm:

Example: Consider the graph with base set as given below:



Authnew = Sum of Hub score that points to it.
Hubnew = Sum of Auth score that it points to

= Sol

Initially, the Authority & Hub score of all nodes are consider as 1 i.e.

| | Iteration 0 | | Iteration 1 | | Iteration 2 | |
|---|---|---|---|---|---|---|
| | Auth | Hub | Auth | Hub | Auth | Hub |
| A | 1 | 1 | 3/6 | 0/6 | 1 = 6/14 | 0 = 0 |
| B | 1 | 1 | 1/6 | 2/6 | 3/6 = 3/14 | 5/6 = 5/14 |
| C | 1 | 1 | 2/6 | 1/6 | 5/6 = 5/14 | 3/6 = 3/14 |
| D | 1 | 1 | 0/6 | 3/6 | 0 = 0 | 6/6 = 6/14 |

After certain iterations, authority & hub score of each node will converge to unique value and the node with highest authority will have highest priority.

In this case, Page A has the high authority value.

## * Another method (matrices Representation):

i.e,

$$\boxed{\begin{aligned} a_i &= M^T h_{i-1} \\ h_i &= M a_{i-1} \end{aligned}}$$

where,

$a_i$ = Vector of authority score.
$h_i$ = Vector of hub's score.

$$M = \begin{array}{c|cccc} & A & B & C & D \\ \hline A & 0 & 0 & 0 & 0 \\ B & 1 & 0 & 1 & 0 \\ C & 1 & 0 & 0 & 0 \\ D & 1 & 1 & 1 & 0 \end{array}$$

# Connection from any particular node to other nodes.

Initially,

$$a_0 = \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix}, \quad h_0 = \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix}$$

now, <u>First iteration</u>:

$$a_1 = M^T h_{i-1}$$

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

$$= \begin{bmatrix} 0+1+1+1 \\ 0+0+0+1 \\ 0+1+0+1 \\ 0+0+0+0 \end{bmatrix}$$

$$= \begin{bmatrix} 3 \\ 1 \\ 2 \\ 0 \end{bmatrix} \xrightarrow{\text{After Normalization}} \begin{bmatrix} 3/6 \\ 1/6 \\ 2/6 \\ 0 \end{bmatrix}$$

$h_1 = M a_{i-1}$

$$h_1 = \begin{bmatrix} 0 & 0 & 0 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 0 & 0 & 0 \\ 1 & 1 & 1 & 0 \end{bmatrix} \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix}$$

$$= \begin{bmatrix} 0 \\ 1+0+1+0 \\ 1+0+0+0 \\ 1+1+1+0 \end{bmatrix}$$

$$= \begin{bmatrix} 0 \\ 2 \\ 1 \\ 3 \end{bmatrix} \xrightarrow[\text{normalization}]{\text{After}} \begin{bmatrix} 0 \\ 2/6 \\ 1/6 \\ 3/6 \end{bmatrix}$$

<u>2nd iteration:</u>

$a_2 = M^T H_{i-1}$

$$= \begin{bmatrix} 0 & 1 & 1 & 1 \\ 0 & 0 & 0 & 1 \\ 0 & 1 & 0 & 1 \\ 0 & 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} 0 \\ 2/6 \\ 1/6 \\ 3/6 \end{bmatrix}$$

$$= \begin{bmatrix} 0 + (1 \times \frac{2}{6}) + (1 \times \frac{1}{6}) + (1 \times \frac{3}{6}) \\ 0+0+0+(1 \times \frac{3}{6}) \\ 0 + (1 \times 2/6) + 0 + (1 \times 3/6) \\ 0 \end{bmatrix}$$

$$= \begin{bmatrix} 1 \\ \frac{3}{6} \\ \frac{5}{6} \\ 0 \end{bmatrix} \xrightarrow[\text{Normalization}]{\text{After}} \begin{bmatrix} \frac{6}{14} \\ \frac{3}{14} \\ \frac{5}{14} \\ 0 \end{bmatrix}$$

Again,

$h_2 = M a_{i-1}$

$$= \begin{bmatrix} 0 & 0 & 0 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 0 & 0 & 0 \\ 1 & 1 & 1 & 0 \end{bmatrix} \begin{bmatrix} 3/6 \\ 1/6 \\ 2/6 \\ 0 \end{bmatrix}$$

$$= \begin{bmatrix} 0 \\ (1\times\frac{3}{6}) + 0 + (1\times\frac{2}{6}) + 0 \\ (1\times\frac{3}{6}) + 0 + 0 + 0 \\ (1\times\frac{3}{6}) + (1\times\frac{1}{6}) + (1\times\frac{2}{6}) + 0 \end{bmatrix}$$

$$= \begin{bmatrix} 0 \\ \frac{5}{6} \\ \frac{3}{6} \\ \frac{6}{6} \end{bmatrix} \xrightarrow[\text{Normalization}]{\text{After}} \begin{bmatrix} 0 \\ \frac{5}{14} \\ \frac{3}{14} \\ \frac{6}{14} \end{bmatrix}$$

Scanned with CamScanner

# Unit 3:
## Creating Suggestion & Recommendation:

① Association rule mining:
  ① Apriori Algorithm.
  ② FP-Growth.

① Apriori Algorithm:

example:

| TID | Items |
|---|---|
| T1 | I1, I2, I3 |
| T2 | I2, I3, I4 |
| T3 | I4, I5 |
| T4 | I1, I2, I4 |
| T5 | I1, I2, I3, I5 |
| T6 | I1, I2, I3, I4, I5 |

min_support = 40%

min_Confidence = 60%

= Sol

$C_1$ =

| Item set | Support |
|---|---|
| I1 | 66.67% |
| I2 | 83.33% |
| I3 | 66.67% |
| I4 | 66.67% |
| I5 | 50% |

$L_1$ =

| Item set | Support |
|---|---|
| I1 | 66.67% |
| I2 | 83.33% |
| I3 | 66.67% |
| I4 | 66.67% |
| I5 | 50% |

$C_2$ =

| Itemset | Support |
|---|---|
| (I1, I2) | 66.67% |
| (I1, I3) | 50% |
| (I1, I4) | 33.33% |
| (I1, I5) | 33.33% |
| (I2, I3) | 66.67% |
| (I2, I4) | 50% |
| (I2, I5) | 33.33% |
| (I3, I4) | 33.33% |
| (I3, I5) | 33.33% |
| (I4, I5) | 33.33% |

$L_2$ =

| Itemset | Support |
|---|---|
| (I1, I2) | 66.67% |
| (I1, I3) | 50% |
| (I2, I3) | 66.67% |
| (I2, I4) | 50% |

Association rule for L2 :

I1 → I2 confidence = $\frac{4}{4} \times 100\% = 100\%$

I2 → I1 confidence = $\frac{4}{5} \times 100\% = 80\%$

I1 → I3 confidence = $\frac{3}{4} \times 100\% = 75\%$

I3 → I1 confidence = $\frac{3}{4} \times 100\% = 75\%$

I2 → I3 confidence = $\frac{4}{5} \times 100\% = 80\%$

I3 → I2 confidence = $\frac{4}{4} \times 100\% = 100\%$

I2 → I4 confidence = $\frac{3}{5} \times 100\% = 60\%$

I4 → I2 confidence = $\frac{3}{4} \times 100\% = 75\%$

$C_3$ =

| Itemset | Support |
|---|---|
| (I1, I2, I3) | 50% |

$L_3$ =

| Itemset | Support |
|---|---|
| (I1, I2, I3) | 50% |

Association rule for L3 :

I1 → {I2, I3} confidence = $\frac{3}{4} \times 100\% = 75\%$

I2 → {I1, I3} confidence = $\frac{3}{5} \times 100\% = 60\%$

I3 → {I1, I2} confidence = $\frac{3}{4} \times 100\% = 75\%$

{I2, I3} → I1 confidence = $\frac{3}{4} \times 100\% = 75\%$

$\{I1, I3\} \rightarrow I2$ Confidence $= \frac{3}{3} \times 100\% = 100\%$

$\{I1, I2\} \rightarrow I3$ Confidence $= \frac{3}{4} \times 100\% = 75\%$

## ② FP-Growth:

| TID | List of item_IDs |
|---|---|
| T100 | I1, I2, I5 |
| T200 | I2, I4 |
| T300 | I2, I3 |
| T400 | I1, I2, I4 |
| T500 | I1, I3 |
| T600 | I2, I3 |
| T700 | I1, I3 |
| T800 | I1, I2, I3, I5 |
| T900 | I1, I2, I3 |

= Same example already done in DWDM in Initial page.

## Unit 4:
## Clustering: Grouping things together

### ① The K-means clustering:

example 1:

| | | | | | | |
|---|---|---|---|---|---|---|
| x | 1 | 2 | 2 | 3 | 4 | 5 |
| y | 1 | 1 | 3 | 2 | 3 | 5 |

=Sol

Let $p_1 = (1,1)$, $p_2 = (2,1)$, $p_3 = (2,3)$, $p_4 = (3,2)$, $p_5 = (4,3)$, $p_6 = (5,5)$

Let, $C_1 = (2,1)$ & $C_2 = (2,3)$

Iteration 1:

Calculate the distance between clusters centers & each data points. i.e.

$d(C_1, p_1) = \sqrt{(2-1)^2+(1-1)^2} = 1$

$d(C_1, p_2) = \sqrt{(2-2)^2+(1-1)^2} = 0$

$d(C_1, p_3) = \sqrt{(2-2)^2+(1-3)^2} = 2$

$d(C_1, p_4) = \sqrt{(2-3)^2+(1-2)^2} = 1.41$

$d(C_1, p_5) = \sqrt{(2-4)^2+(1-3)^2} = 2.83$

$d(C_1, p_6) = \sqrt{(2-5)^2+(1-5)^2} = 5$

$d(C_2, P_1) = \sqrt{(2-1)^2+(3-1)^2} = 2.23$

$d(C_2, P_2) = \sqrt{(2-2)^2+(3-1)^2} = 2$

$d(C_2, P_3) = \sqrt{(2-2)^2+(3-3)^2} = 0$

$d(C_2, p_4) = \sqrt{(2-3)^2+(3-2)^2} = 1.41$

$d(C_2, p_5) = \sqrt{(2-4)^2+(3-3)^2} = 2$

$d(C_2, p_6) = \sqrt{(2-5)^2+(3-5)^2} = 3.61$

Thus, after first iteration.

cluster 1 = $\{P_1, P_2, P_4\}$

cluster 2 = $\{P_3, P_5, P_6\}$

## Iteration 2:

new cluster centers are

$$C_1 = (1,1),(2,1),(3,2) = \left(\frac{1+2+3}{3}, \frac{1+1+2}{3}\right)$$

$$= \left(\frac{6}{3}, \frac{4}{3}\right)$$

$$= (2, 1.33)$$

$$C_2 = (2,3),(4,3),(5,5) = \left(\frac{2+4+5}{3}, \frac{3+3+5}{3}\right)$$

$$= \left(\frac{11}{3}, \frac{11}{3}\right)$$

$$= (3.67, 3.67)$$

Again, Calculate the distance between each points & new cluster centers.

$d(C_1, P_1) = \sqrt{(2-1)^2+(1.33-1)^2} = 1.05$

$d(C_1, P_2) = \sqrt{(2-2)^2+(1.33-1)^2} = 0.33$

$d(C_1, P_3) = \sqrt{(2-2)^2+(1.33-3)^2} = 1.67$

$d(C_1, P_4) = \sqrt{(2-3)^2+(1.33-2)^2} = 1.20$

$d(C_1, P_5) = \sqrt{(2-4)^2+(1.33-3)^2} = 2.61$

$d(C_1, P_6) = \sqrt{(2-5)^2+(1.33-5)^2} = 4.74$

$d(C_2, P_1) = \sqrt{(3.67-1)^2+(3.67-1)^2} = 3.77$

$d(C_2, P_2) = \sqrt{(3.67-2)^2+(3.67-1)^2} = 3.15$

$d(C_2, P_3) = \sqrt{(3.67-2)^2+(3.67-3)^2} = 1.79$

$d(C_2, P_4) = \sqrt{(3.67-3)^2+(3.67-2)^2} = 1.79$

$d(C_2, P_5) = \sqrt{(3.67-4)^2+(3.67-3)^2} = 0.75$

$d(C_2, P_6) = \sqrt{(3.67-5)^2+(3.67-5)^2} = 1.88$

After 2nd Iteration:

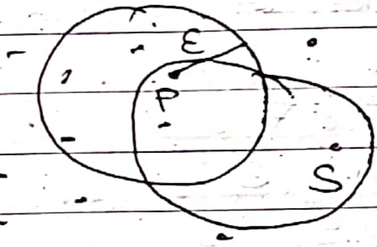
$C_1 = \{P_1, P_2, P_3, P_4\}$

$C_2 = \{P_5, P_6\}$

Note : Repeat the process until:
i) The values of updated cluster center doesn't changes or
ii) The same clusters forms in iterative iterations.

* Elbow method :- Determines the suitable number of clusters that can be formed for the given datasets.

## * DBSCAN: (Density-based spatial clustering of Applications with Noise)




$\varepsilon = 1.5$

minpoint = 3 i.e. if cluster have min 3 points it is corepoint & forms a valid cluster.

e.g:-

| X | 1 | 2 | 2 | 3 | 4 | 5 |
|---|---|---|---|---|---|---|
| Y | 1 | 1 | 3 | 2 | 3 | 5 |

= Sol

\# Distance

A = (1,1), B = (2,1), C = (2,3), D = (3,2), E = (4,3), F = (5,5)

| | B/N A ⇒B | B c | ⇒B C B/N | D c | ⇒B E B/N | N F B/N |
|---|---|---|---|---|---|---|
| A | 0 | 1 | 2.24 | 2.24 | 3.61 | 5.66 |
| B | 1 | 0 | 2 | 1.41 | 2.82 | 5 |
| C | 2.24 | 2 | 0 | 1.41 | 2 | 3.61 |
| D | 2.24 | 1.41 | 1.41 | 0 | 1.41 | 3.61 |
| E | 3.61 | 2.82 | 2 | 1.41 | 0 | 2.24 |
| F | 5.66 | 5 | 3.61 | 3.61 | 2.24 | 0 |

A: B ↓ b    B: A,D    C: D ↓ b    D: B,C,E    E: D ↓ b    F:

= Since, B & D has 3 points included they are the corepoints so,

The points in B & D are boundary points.

Since, F is Noise & also, D belongs to B which is a corepoint & B belongs to D which is also a core point only one cluster is form i.e.

$C_1 = \{B, A, D, C, E\}$.

Example 2: Min point = 3 & $\varepsilon = 1.5$

| | B/N A →N | B/N B →N | B/N C →B | D c | E c | F c |
|---|---|---|---|---|---|---|
| → A | 0 ✓ | ✓ 0.7 | 5.7 | 3.6 | 4.2 | 3.2 |
| B | 0.7 ✓ | ✓ 0 | 4.9 | 2.9 | 3.5 | 2.5 |
| C | 5.7 | 4.9 | 0 ✓ | 2.2 | 1.4 ✓ | 2.5 |
| D | 3.6 | 2.9 | 2.2 | 0 ✓ | 1 ✓ | 0.5 ✓ |
| → E | 4.2 | 3.5 | ✓ 1.4 | 1 ✓ | ✓ 0 | 1.1 ✓ |
| F | 3.2 | 2.5 | 2.5 | 0.5 ✓ | ✓ 1.1 | 0 ✓ |

A : B   B : A   C : E   D : E, F   E : C, D, F   F : D, E

C : Neighborhood points = E which is a core point thus C is a border point.

A : Neighborhood points = B which is a B/N.
B : Neighborhood points = A which is a B/N.

Thus, A and B are Noise or outlier.

## * Robust clustering Using Links (ROCK):

- used for categorical data (Market basket datas).
- When data are not numerical data but itemset.

① $sim(P_i, P_j) = \frac{|P_i \cap P_j|}{|P_i \cup P_j|}$

② $g(P_i, P_j) = \frac{link[P_i, P_j]}{(n+m)^{1+2f(\theta)} - n^{1+2f(\theta)} - m^{1+2f(\theta)}}$

③ $f(\theta) = \frac{1-\theta}{1+\theta}$ where, $\theta$ = threshold given similarity

Example:

p1 = {Judgement, faith, prayer, fair}
p2 = {fasting, faith, prayer}
p3 = {fair, fasting, faith}
p4 = {fasting, prayer, pilgrimage}
Similarity threshold = 0.3
number of required clusters = 2

= Sol $Sim(P_i, P_j) = \frac{|P_i \cap P_j|}{|P_i \cup P_j|}$

| | $P_1$ | $P_2$ | $P_3$ | $P_4$ |
|---|---|---|---|---|
| p1 | 1 | 2/5 = 0.4 | 2/5 = 0.4 | 1/6 = 0.167 |
| p2 | 0.4 | 1 | 2/4 = 0.5 | 2/4 = 0.5 |
| p3 | 0.4 | 0.5 | 1 | 1/5 = 0.2 |
| p4 | 0.167 | 0.5 | 0.2 | 1 |

Similarity table

Scanned with CamScanner

checking each cell and replace cell with 1 if threshold is greater than 0.3 or with 0 if threshold is less than 0.3

now,

| | P1 | P2 | P3 | P4 |
|---|---|---|---|---|
| P1 | 1 | 1 | 1 | 0 |
| P2 | 1 | 1 | 1 | 1 |
| P3 | 1 | 1 | 1 | 0 |
| P4 | 0 | 1 | 0 | 1 |

Adjacency table.

now, calculate Adjacency table multiplication i.e.

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

$$= \begin{bmatrix} 3 & 3^{*} & 3 & 1 \\ 3 & 4 & 3^{*} & 2 \\ 3 & 3 & 3 & 1 \\ 1 & 2 & 1 & 2 \end{bmatrix}$$

now,

| | P1 | P2 | P3 | P4 |
|---|---|---|---|---|
| P1 | — | 3* | 3 | 1 |
| P2 | 3 | — | 3* | 2 |
| P3 | 3 | 3 | — | 1 |
| P4 | 1 | 2 | 1 | — |

Common neighbors tables.

now,

$$g(P_i, P_j) = \frac{\text{Link}[P_i, P_j]}{(n+m)^{1+2f(\theta)} - n^{1+2f(\theta)} - m^{1+2f(\theta)}}$$

$$f(\theta) = \frac{1-\theta}{1+\theta} = \frac{1-0.3}{1+0.3} = \frac{0.7}{1.3} = 0.538$$

now,

| pair | Goodness measure |
|---|---|
| P1, P2 | 1.35 |
| P1, P3 | 1.35 |
| P1, P4 | 0.45 |
| P2, P3 | 1.35 |
| P2, P4 | 0.90 |
| P3, P4 | 0.45 |

$$P_1, P_2 = \frac{\text{Link}[P_i, P_j]}{(n+m)^{1+2f(\theta)} - n^{1+2f(\theta)} - m^{1+2f(\theta)}}$$

$$= \frac{3}{2^{1+2\times0.538} - 1 - 1}$$

initially, the values of n & m is 1 as we consider each points as a single cluster.

$$= \frac{3}{2.216}$$

$$= 1.35$$

$$P_1, P_3 = \frac{3}{2^{1+2\times0.538} - 1 - 1} = 1.35$$

$$P_1, P_4 = \frac{1}{2.216}$$ similarly for all pairs.

now, The points with the highest Goodness measures value forms a clusters we have, ① $p_1, p_2$
② $p_1, p_3$
③ $p_2, p_3$
with highest values so, take any from these

we take, $p_1$ $p_2$ as a single cluster now, update the common neighbor table.

| | $\{p_1, p_2\}$ | $p_3$ | $p_4$ |
|---|---|---|---|
| $\{p_1, p_2\}$ | — | $3+3=6$ | $1+2=3$ |
| $p_3$ | | — | 1 |
| $p_4$ | | | — |

updated neighbor table.

now, calculate goodness measure :

| pair. | Goodness measure. |
|---|---|
| $\{p_1, p_2\}, p_3$ | 1.31 |
| $\{p_1, p_2\}, p_4$ | 0.66 |
| $p_3, p_4$ | 0.22 |

now,

$$\{p_1, p_2\}, p_3 = \frac{\text{link}[p_i, p_j]}{(n+m)^{1+2f(\theta)} - n^{1+2f(\theta)} - m^{1+2f(\theta)}}$$

$$= \frac{6}{(2+1)^{1+2\times0.538} - 2^{1+2\times0.538} - 1}$$

$$= \frac{6}{9.783 - 4.216 - 1}$$

$$= \frac{6}{4.567}$$

$$= 1.31$$

Similarly for all

now, highest values of Goodness measure is in $\{P_1, P_2\}, P_3$
it form a cluster & we get

cluster 1 = $\{P_1, P_2, P_3\}$
cluster 2 = $\{P_4\}$

which is our requirement.