PREPARED BY: RAJARAJESWARI VAIDYANATHAN **CWID**: A20362220 **COURSE**: CS422 – DATA MINING

TABLE OF CONTENTS

TWITTER MESSAGES:

TWITTER MESSAGES:

user : Tim, Tweet : "@Tom Howdy"

user : Mike, tweet : "Welcome @Tom and @Anne!"

user : Tom, tweet : "Hi @Mike and @Anne!"

user : Anne, tweet : "Howdy!"

A twitter user graph is constructed from the above twitter messages.

TWITTER USER GRAPH:

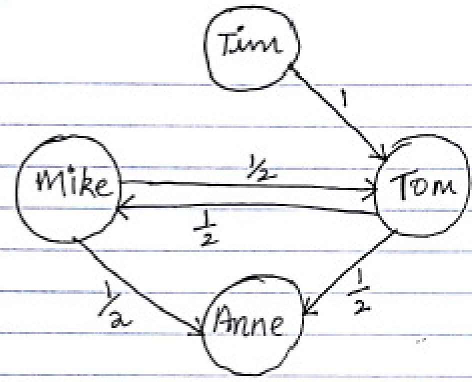
Each user is a node. There exists an edge between users, if the user denotes or points to another user in the twitter message.

Twitter Graph $G = (V, E)$

where,

V = Vertices of the graph, which are the users

E = Edge in the graph, an edge exists if an user mentions '@' another user.



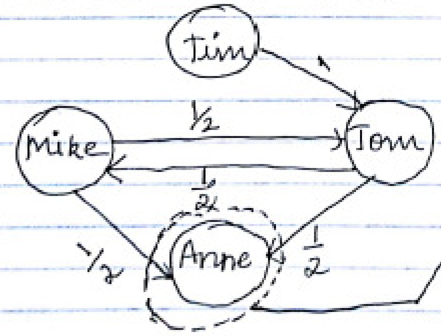
Twitter User Graph

Page Rank Approach in Twitter User Graph:

We use the Page Rank approach to find the rank (or) influence of each user in the graph.

⇒ Higher the Rank of the user more the influence on them.

From the above graph, we get a dead end at the node "ANNE"



'Dead end' i.e., No outgoing edges.

From the above graph, the node "ANNE" has no outgoing edges. Hence it is a 'dead end' and the surfer now needs to make a random jump.

Hence we use the method 'Taxation' which in turn uses the 'Teleport" value in the page rank calculation.

Page Rank Vector Estimate $= \beta M V + \frac{(1-\beta)e}{n}$
V'

where, β = ~~Transport~~ Teleport Value

M = Transition Matrix

V = Previous / Initial Page Rank Vector.

e = Vector of all 1's.

N = Number of nodes in the graph.

TRANSITION MATRIX FROM TWITTER USER GRAPH:

$$
M = \begin{array}{c c}
 & \begin{array}{cccc} \text{ANNE} & \text{MIKE} & \text{TIM} & \text{TOM} \end{array} \\
\begin{array}{c} \text{ANNE} \\ \text{MIKE} \\ \text{TIM} \\ \text{TOM} \end{array} &
\left[\begin{array}{cccc} 0 & \frac{1}{2} & 0 & \frac{1}{2} \\ 0 & 0 & 0 & \frac{1}{2} \\ 0 & 0 & 0 & 0 \\ 0 & \frac{1}{2} & 1 & 0 \end{array} \right]
\end{array}
$$

The transition matrix is constructed from the twitter user graph. For example, There exists a single edge from Tim to Tom. Hence the value "1" is placed in the utility matrix. There exists 2 outer edges from Mike to Tom and Anne which interprets that there is a probability to jump to either Tom or Anne from Mike. Hence the outer edges share the probability of $\frac{1}{2}$ each from Mike.

Iterations using Page Rank over the Twitter user Graph.

AT ITERATION 1:

$$v' = \beta M v + \frac{(1-\beta)e}{N}$$

where $V = \frac{1}{4}$, $e = \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix}$, $N = 4$

Teleportation value is given as 0.1

$\therefore \quad 1 - \beta = 0.1$

$\beta = 0.9$

$$\therefore \quad V' = 0.9 \begin{bmatrix} 0 & 0.5 & 0 & 0.5 \\ 0 & 0 & 0 & 0.5 \\ 0 & 0 & 0 & 0 \\ 0 & 0.5 & 1 & 0 \end{bmatrix} \begin{bmatrix} 0.25 \\ 0.25 \\ 0.25 \\ 0.25 \end{bmatrix} + 0.1 \begin{bmatrix} 0.25 \\ 0.25 \\ 0.25 \\ 0.25 \end{bmatrix}$$

$$= 0.9 \begin{bmatrix} 0 + 0.125 + 0 + 0.125 \\ 0 + 0 + 0 + 0.125 \\ 0 + 0 + 0 + 0 \\ 0 + 0.125 + 0.25 + 0 \end{bmatrix} + \begin{bmatrix} 0.025 \\ 0.025 \\ 0.025 \\ 0.025 \end{bmatrix}$$

$$v' = 0.9 \begin{bmatrix} 0.25 \\ 0.125 \\ 0 \\ 0.375 \end{bmatrix} + \begin{bmatrix} 0.025 \\ 0.025 \\ 0.025 \\ 0.025 \end{bmatrix}$$

$$v' = \begin{bmatrix} 0.225 \\ 0.1125 \\ 0 \\ 0.3375 \end{bmatrix} + \begin{bmatrix} 0.025 \\ 0.025 \\ 0.025 \\ 0.025 \end{bmatrix}$$

$$v' = \begin{bmatrix} 0.25 \\ 0.1375 \\ 0.025 \\ 0.3625 \end{bmatrix} \longrightarrow \text{End of Iteration 1}$$

Second Iteration.

$$v' = \beta M v + (1-\beta) e/N$$

Where $e = \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix}$ $N = 4$ $v' = \begin{bmatrix} 0.25 \\ 0.1375 \\ 0.025 \\ 0.3625 \end{bmatrix}$

$\beta = 0.9$

$\boxed{1-\beta = 0.1}$

$$v' = 0.9 \begin{bmatrix} 0 & 0.5 & 0 & 0.5 \\ 0 & 0 & 0 & 0.5 \\ 0 & 0 & 0 & 0 \\ 0 & 0.5 & 1 & 0 \end{bmatrix} \begin{bmatrix} 0.25 \\ 0.1375 \\ 0.025 \\ 0.3625 \end{bmatrix} + 0.1 \begin{bmatrix} 0.25 \\ 0.25 \\ 0.25 \\ 0.25 \end{bmatrix}$$

$$v' = 0.9 \begin{bmatrix} 0 + (0.5)(0.1375) + 0 + (0.5)(0.3625) \\ 0 + 0 + 0 + 0.5(0.3625) \\ 0 + 0 + 0 + 0 \\ 0 + (0.5)(0.1375) + 1(0.025) + 0 \end{bmatrix} + \begin{bmatrix} 0.025 \\ 0.025 \\ 0.025 \\ 0.025 \end{bmatrix}$$

$$\therefore \quad v^1 = 0.9\begin{bmatrix} 0.25 \\ 0.18125 \\ 0 \\ 0.09375 \end{bmatrix} + \begin{bmatrix} 0.025 \\ 0.025 \\ 0.025 \\ 0.025 \end{bmatrix}$$

$$v^1 = \begin{bmatrix} 0.25 \\ 0.18812 \\ 0.025 \\ 0.10937 \end{bmatrix} \Rightarrow \text{End of Iteration } 2.$$

THIRD ITERATION:

$$v^1 = \beta M v + (1-\beta) e/N$$

$$V = \begin{bmatrix} 0.25 \\ 0.18812 \\ 0.025 \\ 0.10937 \end{bmatrix}, \quad e = \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix}, \quad N = 4, \quad \beta = 0.9 \quad \boxed{1-\beta = 0.1}$$

$$v' = 0.9\begin{bmatrix} 0 & 0.5 & 0 & 0.5 \\ 0 & 0 & 0 & 0.5 \\ 0 & 0 & 0 & 0 \\ 0 & 0.5 & 1 & 0 \end{bmatrix}\begin{bmatrix} 0.25 \\ 0.18812 \\ 0.025 \\ 0.10937 \end{bmatrix} + 0.1\begin{bmatrix} 0.25 \\ 0.25 \\ 0.25 \\ 0.25 \end{bmatrix}$$

$$V' = 0.9 \begin{bmatrix} 0 + 0.5(0.18812) + 0 + 0.5(0.10937) \\ 0 + 0 + 0 + 0.5(0.1093) \\ 0 + 0 + 0 + 0 \\ 0 + 0.5(0.18812) + 1(0.025) + 0 \end{bmatrix} + \begin{bmatrix} 0.025 \\ 0.025 \\ 0.025 \\ 0.025 \end{bmatrix}$$

$$V' = 0.9 \begin{bmatrix} 0.14874 \\ 0.05468 \\ 0 \\ 0.11906 \end{bmatrix} + \begin{bmatrix} 0.025 \\ 0.025 \\ 0.025 \\ 0.025 \end{bmatrix}$$

$$V' = \begin{bmatrix} 0.158866 \\ 0.07421 \\ 0.025 \\ 0.13215 \end{bmatrix} \Rightarrow \text{End of Iteration 3//.}$$

Hence, at the end of 3^{rd} Iteration

$$V' = \begin{bmatrix} 0.15886 \\ 0.0742 \\ 0.025 \\ 0.1321 \end{bmatrix} \begin{matrix} \Rightarrow \text{ANNE} \\ \Rightarrow \text{MIKE} \\ \Rightarrow \text{TIM} \\ \Rightarrow \text{TOM} \end{matrix} \left.\vphantom{\begin{matrix} a \\ a \\ a \\ a \end{matrix}}\right\} \text{RANKS OF VARIOUS USERS}$$

Inference and Conclusion.

ANNE is found to have the maximum rank and she is the most Influential Person.

$\boxed{\text{Higher the rank of the user, more the influence on them.}}$

Ans: **Rank Order of Users**

1. ANNE → 0.158
2. TOM → 0.132
3. MIKE → 0.074
4. TIM → 0.025