

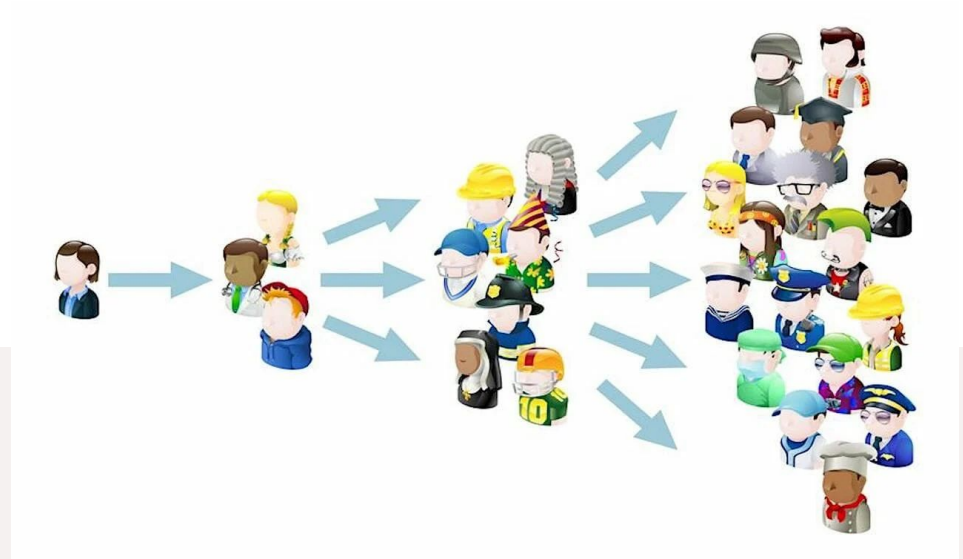
Social Network Analysis: WOM Marketing in Rural India

TEAM MEMBERS

Asmita Mukherjee (MT21115)

Simran (MT21146)

GROUP 1



UNDER THE GUIDANCE OF

Dr. Tanmoy Chakraborty (Professor)

Shivani Kumar (TA)

Table of Contents

- Problem definition: Motivation + Introduction
- Literature review
- Dataset description
- Baselines: Overview + results
- Proposed solution sketch
- Results
- Results analysis and limitations
- Future scope and conclusion

Problem Definition

Motivation and Introduction



Motivation and Introduction

Motivation

- Government Schemes
 - Education, Employment, Immunization, Ration etc.
- Rural people → Unawareness

Essential to find individuals that are

"highly central" → diffuse information faster and far.

Introduction

- traditional methods of degree centrality, page rank
- Community detection jointly with influential nodes

Literature Review

Related Work 1 : Using Gossips to spread Information

Related Work 2 : Node embedding based influential spreaders identification

Related Work 3 : Modelling Information Diffusion Dynamics over Social Networks

Related Work 4 : Word of Mouth, Digital Media, and Open Innovation at the Agricultural SMEs



Related Work



Using Gossips to spread Information

- 1) Can highly central or influential nodes be identified, when we do not have the exact network data, but only via asking individuals to identify the most influential members of their community.
- 2) Are the members identified via the above method more effective in transmitting information as compared to randomly chosen individuals or even highly respected individuals in the community.
- Diffusion Metric to identify the most influential individuals.
-

A node embedding based influential spreaders identification

- 1) Modified deepwalk to better represent the spread of information in the network
- 2) K-means on the latent representation to obtain clusters
- 3) Selection of core nodes from the cluster by finding the node with most similarity with all the other nodes in the cluster. Found by taking difference of the vector representation of two nodes

Related Work



Modelling Information Diffusion Dynamics over Social Networks

Shortcomings of present ML models for analysis of diffusion of information.

- 1) The results learned by the models is specific to only one particular network that it is trained on and might not generalize well with other networks.
- 2) The ML models often ignores the "actions & the decisions" of the users.

Solution : Game Theoretic approach.

Players have 2 choices, to forward info or to not.

Factors affecting WOM marketing in digital era

- **Age**
- Gender
- **Education**
- Mobile phone ownership
- **Direct social interaction**
- Social media ownership
- Distance
- **Nationality**

Binary Logistic Model - accept/ reject hypothesis.

Dataset Description

Basic information about the dataset

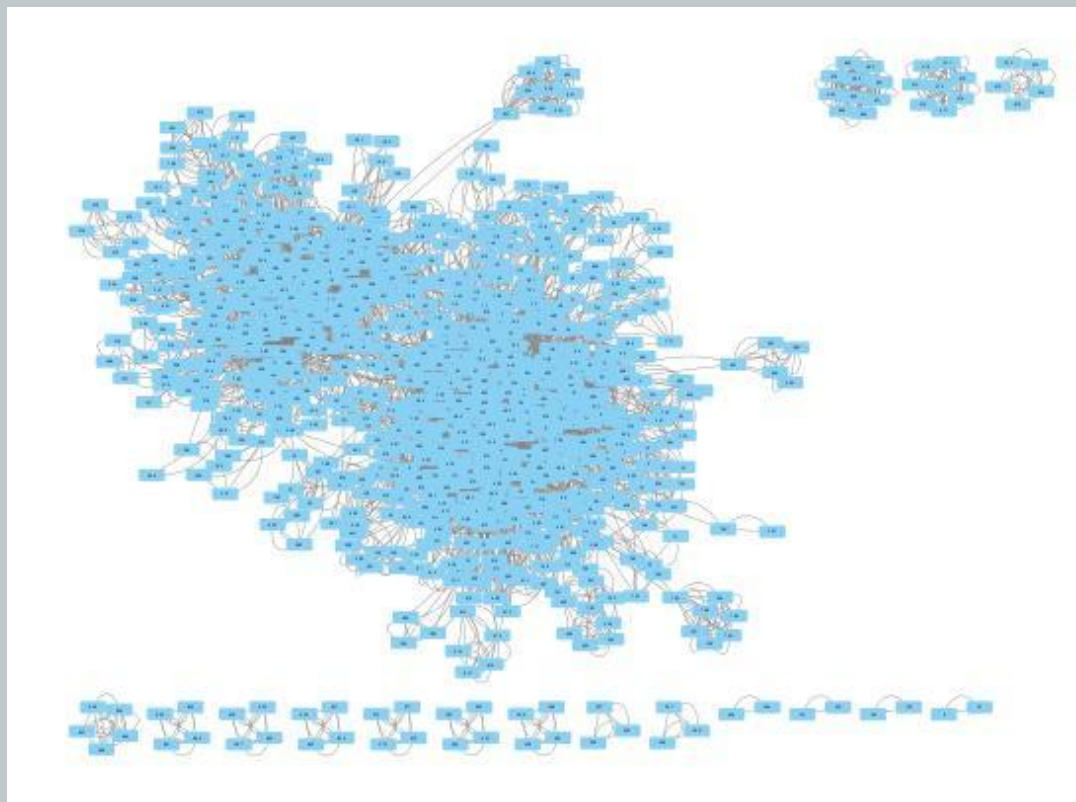


Dataset Description

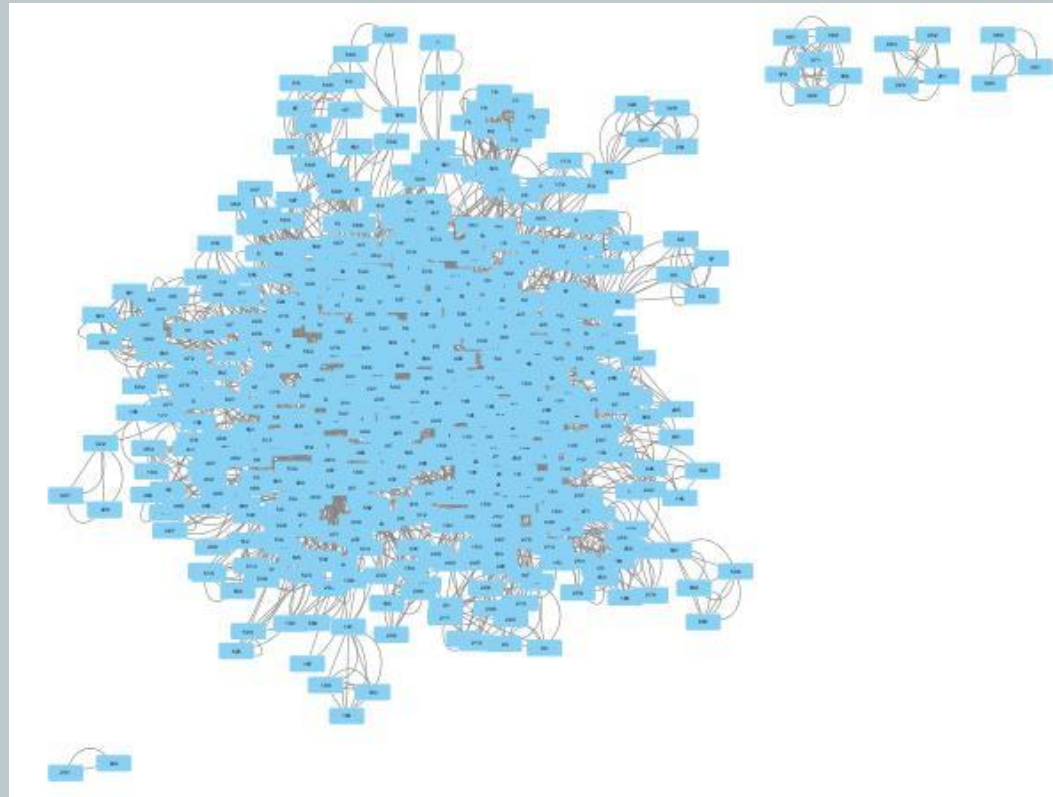
About Dataset

- Microfinance company
- Questions: (who they)
 - borrow money from
 - give advice to
 - help with a decision
 - borrow kerosene or rice from
 - lend kerosene or rice to
 - lend money to
 - obtain medical advice from
 - engage socially with
 - are related to
 - go to temple with
 - invite to one's home
 - visit in another's home.

Village 1



Village 2



Baselines

Overview and Results



Baselines

- Degree Centrality
 - person who deals with a lot of people.
- Closeness Centrality
 - spread information faster.
- Betweenness Centrality
 - spread information as far as possible
- PageRank
 - influential nodes
- Katz Centrality
 - Takes into consideration other connected nodes with the current node and not only the immediate neighbors.



Proposed Solution

A sketch of proposed solution



Proposed Solution Sketch



Main Idea

Community detection and influential nodes finding (jointly)

Step 1

- Deep Walk on the graph → latent representation of the nodes



Step 2

- K-Means clustering to detect communities
K = no. of influential nodes
= no. of clusters



Step 3

- Find Central Nodes from community and report them as initial seed nodes



Step 4

- Model information flow through general threshold model.

<u>Active Node set</u>	<u>% of nodes informed</u>
Betweenness	19.63%
VoteRank	19.41%
Degree	19.41%
Closeness	19.41%
Page rank	20.09%
Eigenvector	17%
Load	19.63%

Experiments with different measures to obtain the seed nodes from the cluster.

Results

Final Result and comparison with Baselines



	Village 1 (in %)	Village 2 (in %)
Degree Centrality	20.03	18.38
Closeness Centrality	19.21	20.09
Betweenness Centrality	19.87	22.95
Page Rank	Failed to converge (50K)	8.68
<u>Our Model</u>	22.85 ↑	25.75 ↑

Percentage Of Informed nodes at the end of iteration in general threshold model

Future Scope and Conclusion



Future scope And Conclusion

Future Scope

- Better latent representation to represent information flow
- Modelling information flow via deep learning methods

Conclusion

- Community detection and influential nodes
- DeepWalk
- K-Means
- Central nodes from pagerank as influential nodes
- Beat the traditional methods for influential nodes

Thanks!

Any questions?

