

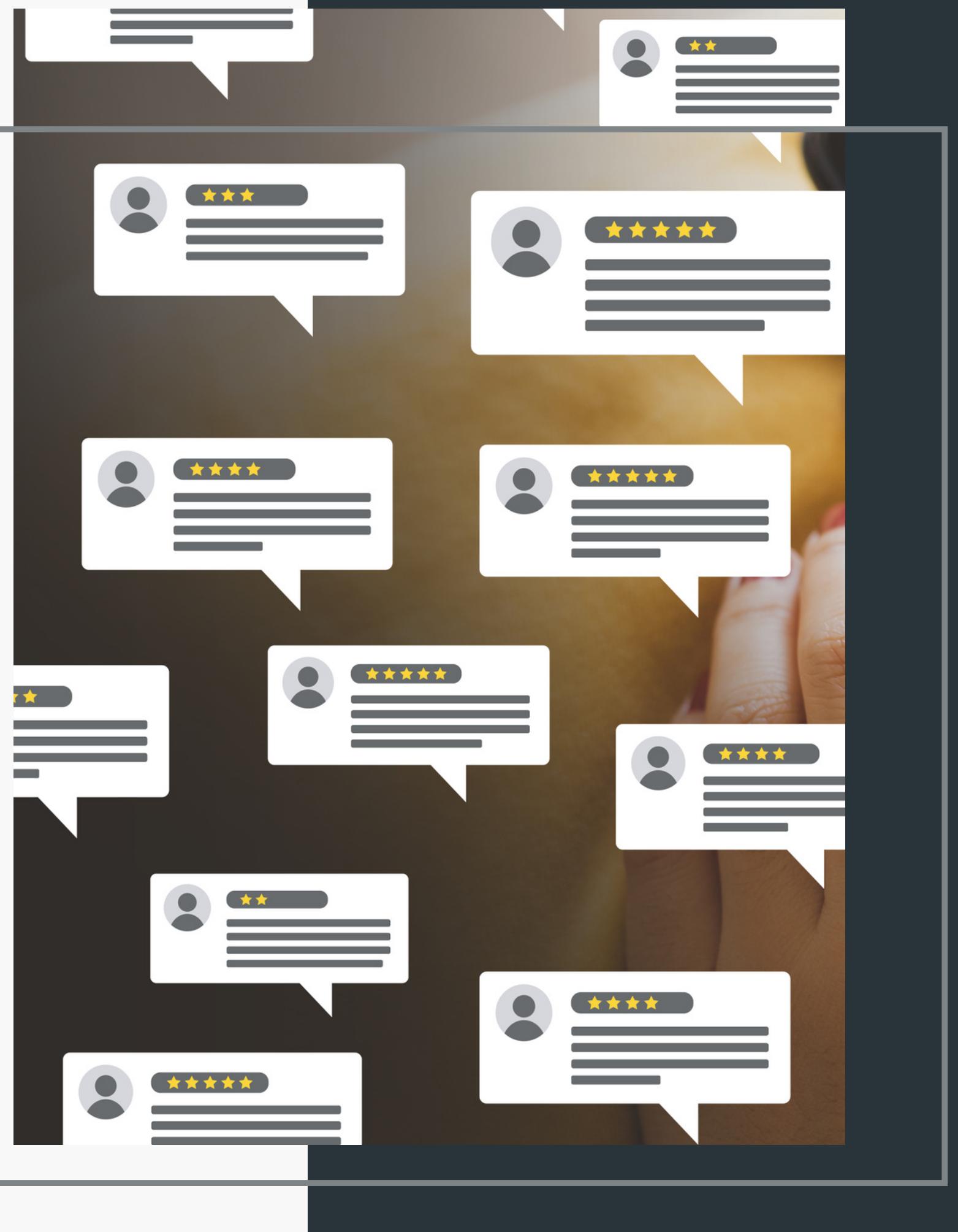


12TH CONFERENCE ON EXCELLENCE IN RESEARCH AND EDUCATION

EASYCHAIR PAPER SUBMISSION : 245
**ANALYSING INTERACTION OF FAKE
REVIEW ON E-COMMERCE
PLATFORM USING TEXT- BASED
NETWORK ANALYTICS**

A research paper on the analysis of product reviews that are user generated and computer generated using network analysis

By Kartik Gupta and Isha Bansal



RESEARCH PROBLEM

Online information in the form of reviews or news has gained the attention of the research and the industry in recent years. Detecting such information becomes crucial when we are dependent on information due to lack of resources. Fake informational content degrades information found on sites like Amazon and TripAdvisor. Phoney negative reviews on high-quality items are incredibly damaging to businesses, while fake positive ratings on low-quality products are equally damaging to customers. Fake positive ratings on low-quality items are also detrimental to rivals that sell average or good-quality products with fewer reviews. The research aims to examine and observe the key characteristics of the reviews based on a semantic network based approach as well as analyse the features of false reviews.

INTRODUCTION

Companies and researchers have shown much growth in developing better models to help predict whether reviews are user generated or computer generated in an eCommerce setting.

Consumers need information before purchasing any form of product, and in a eCommerce platform online reviews carry important value for consumers.

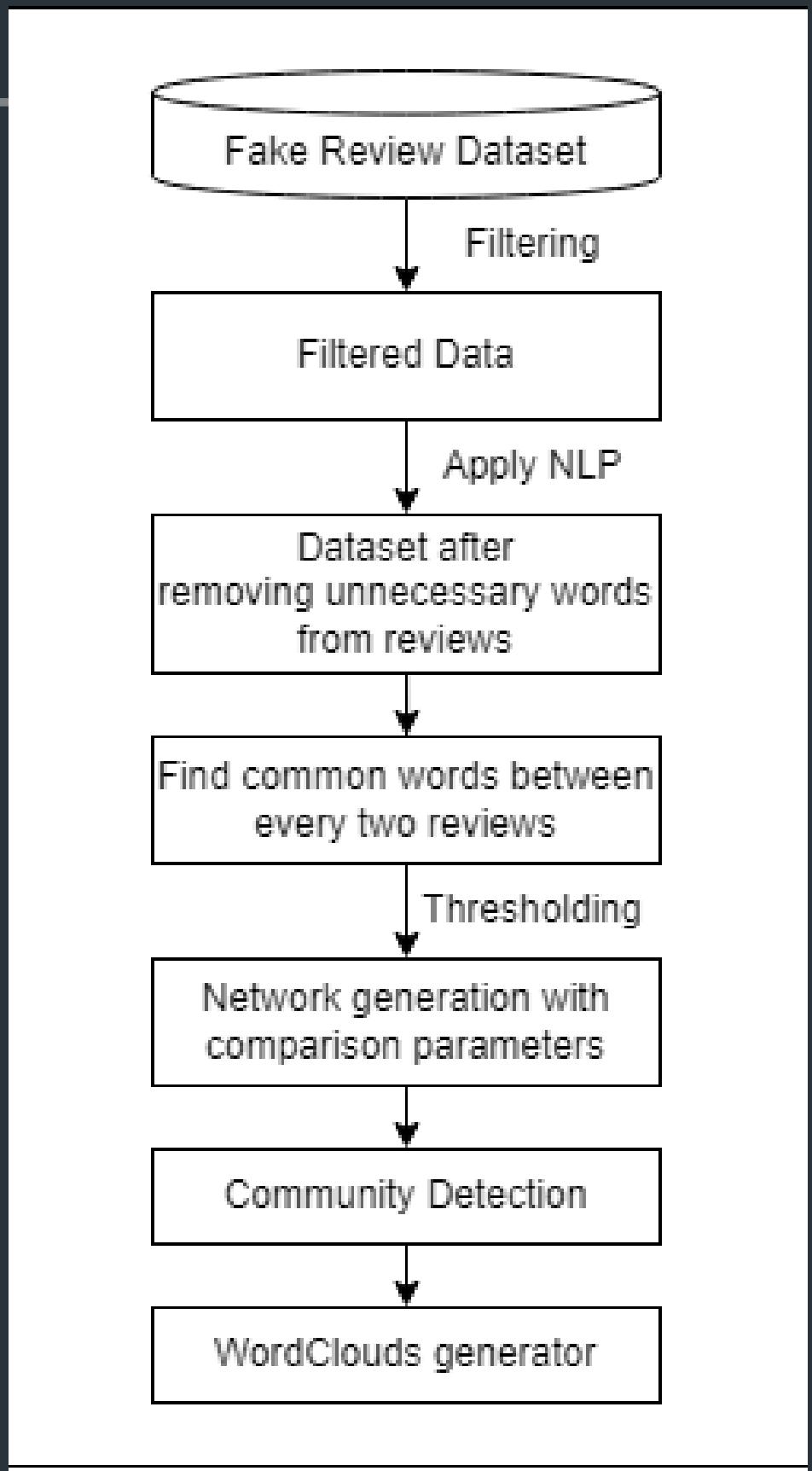
We propose a method of studying the properties of the reviews by creating a text-based network and analysed the properties of the graphs. Understanding the properties of reviews and doing differential analysis of the computer generated reviews from the original reviews.

METHODOLOGY

Network Analysis is performed on the Kindle_store_5 category because it contains the highest number of reviews in comparison to others. Later, Natural Language Processing (NLP) is performed on every review of the filtered dataset to remove all unnecessary words and bring text to a standard form.

The reviews are then interconnected based on commonness between the cleaned text of the reviews based on the number of common words. Based on commonness, social network graphs are created along with limiting the number of connections based on a threshold value.

Community Detection is used to create clusters inside the network which will help to discover reviews with shared interests. The same clusters are visualized using WordClouds which assists in finding the most common words in the network.



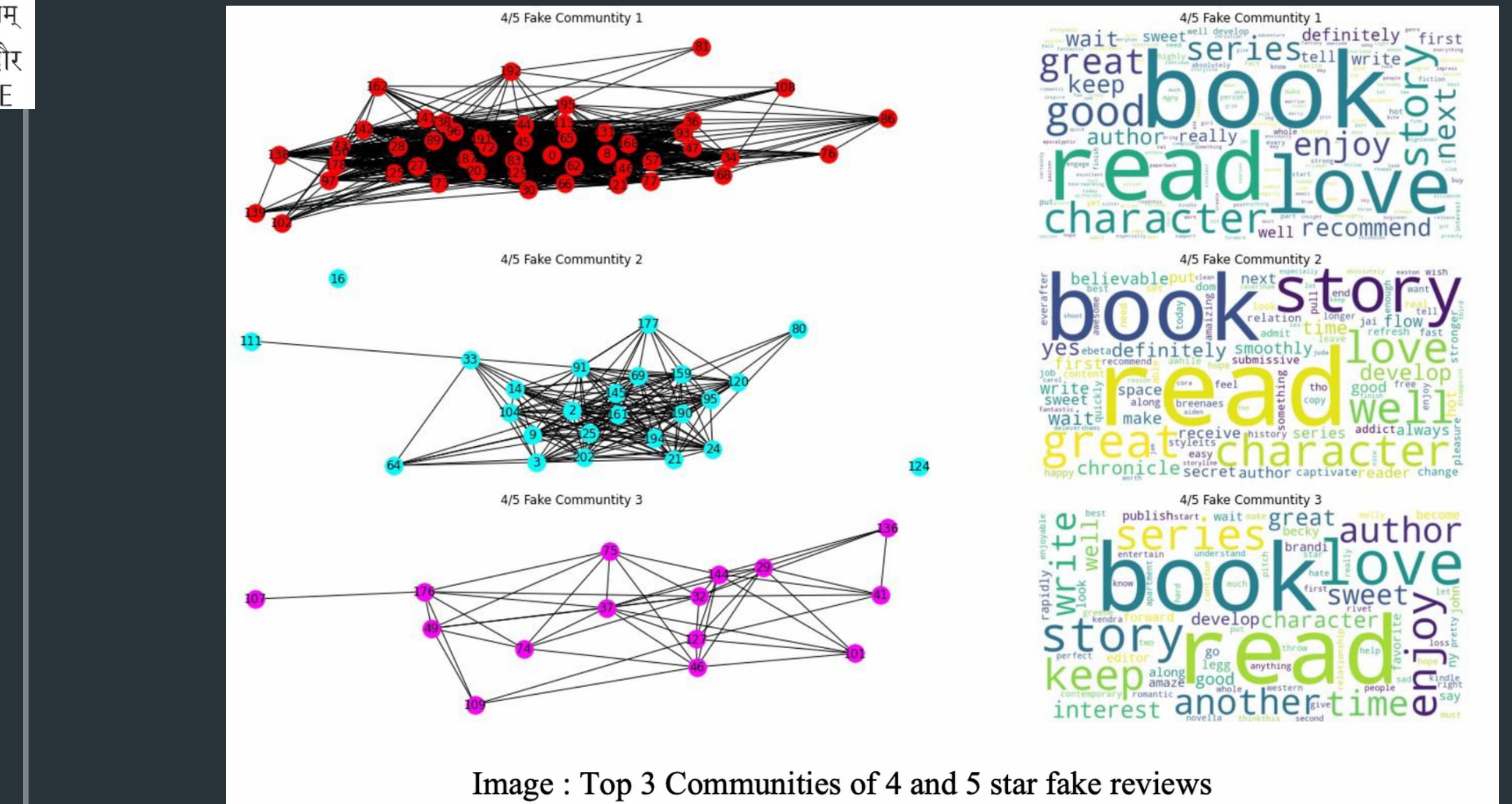


सिद्धिमूलं प्रबन्धनम्

भा. प्र. सं. इन्दौर
IIM INDORE

12TH CONFERENCE ON EXCELLENCE IN RESEARCH AND EDUCATION

RESULTS PART - 1



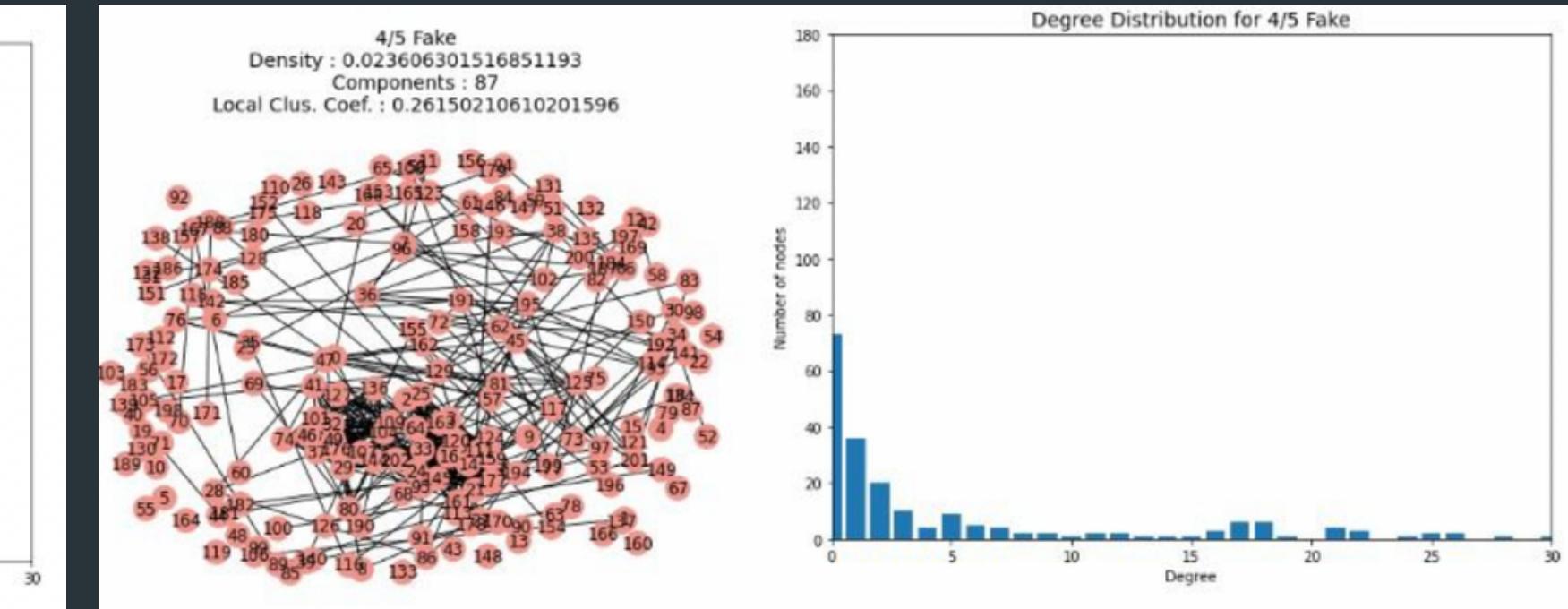
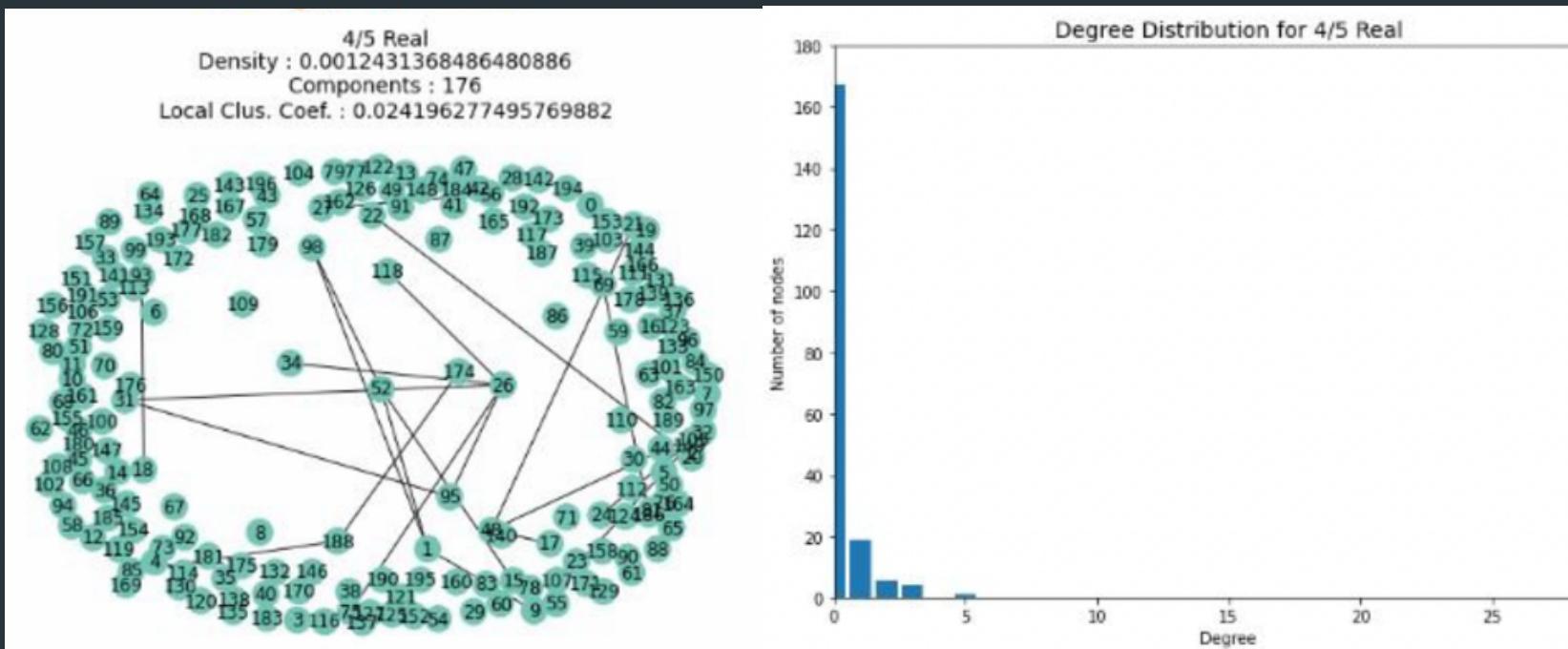
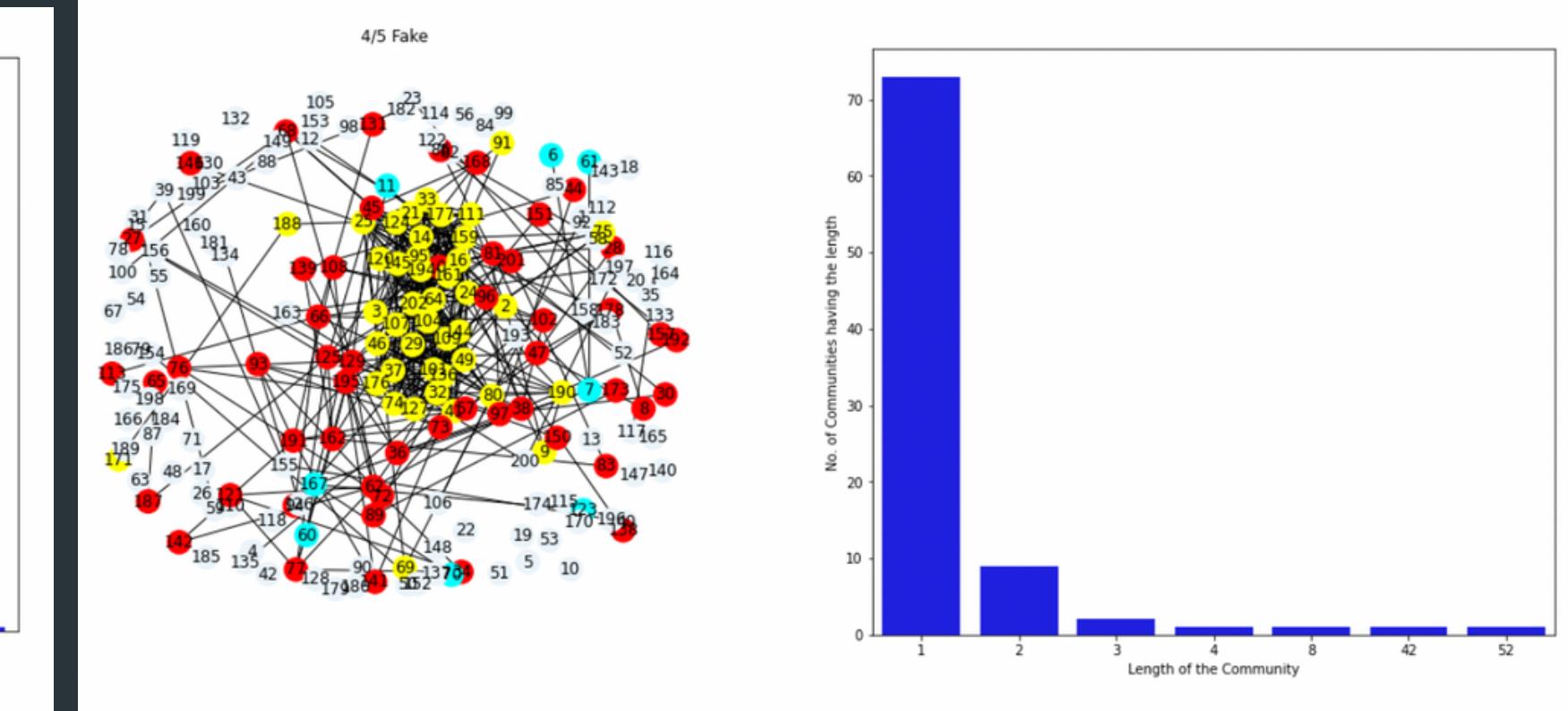
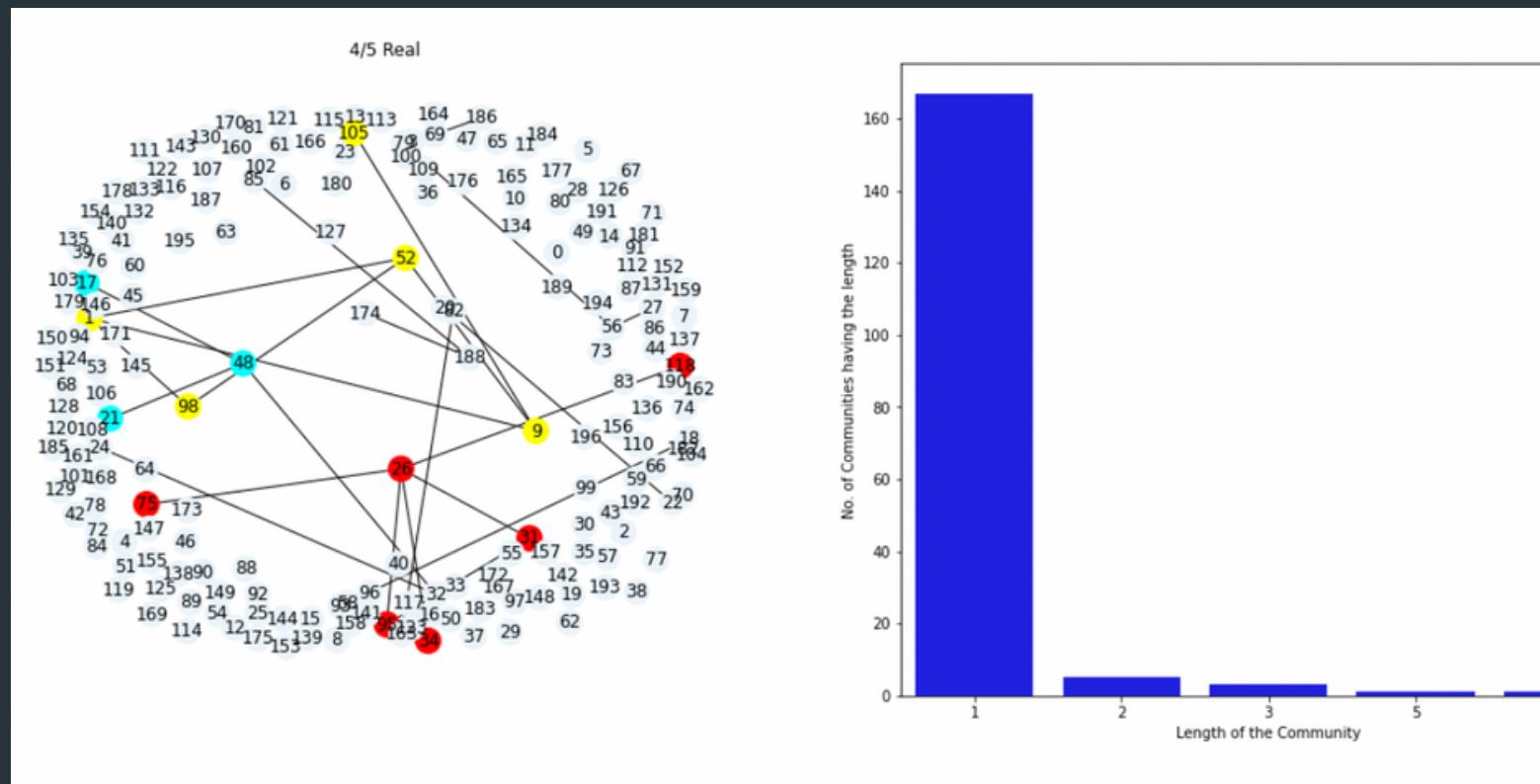
RESULTS PART - 2



Image : Top 3 Communities of 4 and 5 star real reviews



RESULTS PART - 3





RESULTS PART - 4

Networks graphs	Density	No of components	Local clustering coeff
1-2 star real	0	123	0.0
1-2 star fake	0.0029	110	0.0551
4-5 star real	0.0012	176	0.0241
4-5 star fake	0.0236	87	0.2615

CONCLUSION

From the results obtained, we conclude the following:

- Overall, computer-generated reviews have more connections with other nodes in comparison to the original or human created reviews.
- The 4 and 5 star computer generated have more connections which show that the product is being highly praised and the overall review rating being manipulated more.
- Degree distribution in 4 and 5 star CG reviews is higher when compared to their corresponding 4 and 5 star real reviews.
- The generated communities of CG reviews have higher amount of similarities.



सिद्धिमूलं प्रबन्धनम्

भा. प्र. सं. इन्डौर
IIM INDORE

12TH CONFERENCE ON EXCELLENCE IN RESEARCH AND EDUCATION

EASYCHAIR PAPER SUBMISSION : 245

THANK YOU