A STUDY ON APPLICABLE AN IMPROVED K-CLIQUE METHOD TO THE RECOMMENDATION SYSTEM

Dr. Phonexay VILAKONE, phonexay_v@fe-nuol.edu.la
Dr. Somphone KANTHAVONG, somphonek@fe-nuol.edu.la
Dr. Khanthanou LUANGXAYSANA, khanthanou@fe-nuol.edu.la

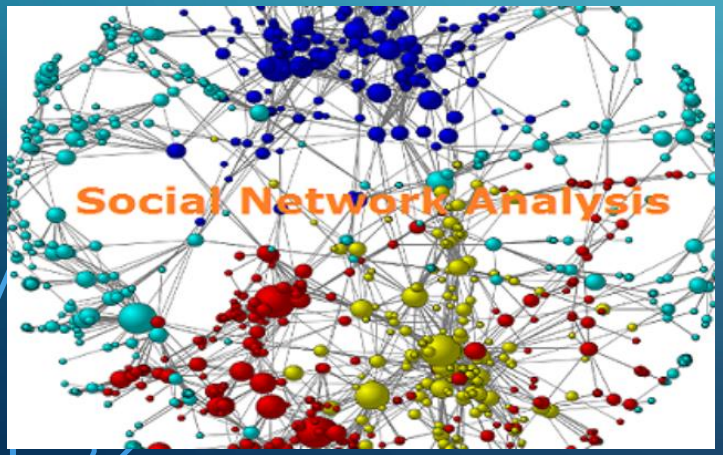
Department of Computer Engineering and IT,
National University of Laos, Laos

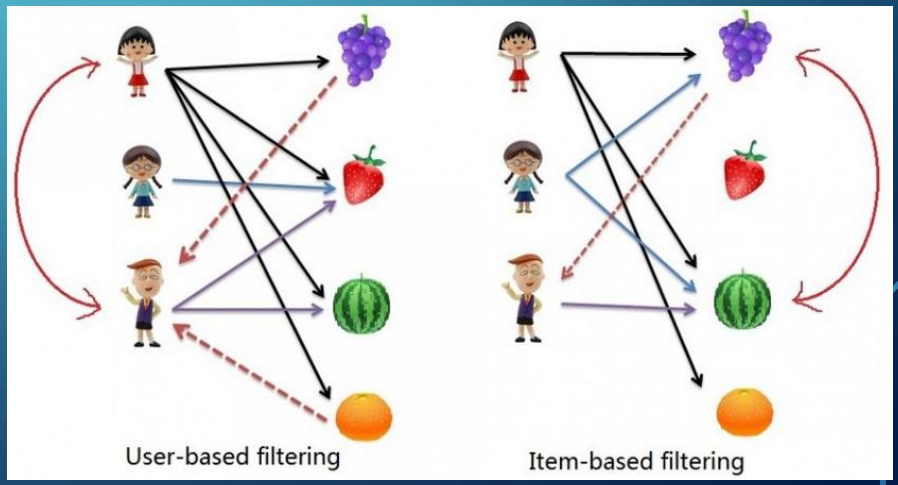
ASEAN IVO Forum 2021
ICT Virtual Organization of ASEAN Institutes and NICT

November 18, 2021, Japan

CONTENT:

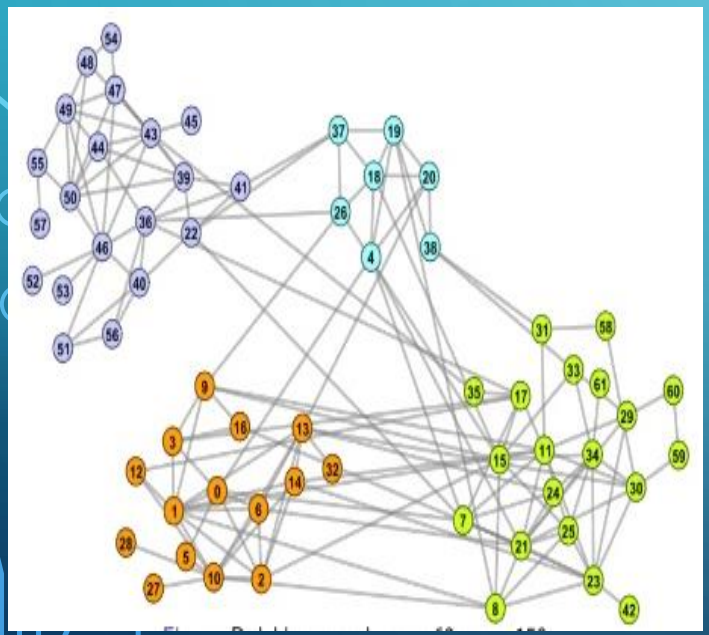
1. Introduction
2. Related Work
3. Proposed Method
4. Evaluation
5. Conclusion & Future Study





INTRODUCTION:

- Today, the users in Online Social Media are increasing rapidly.
- Then, the size of the Social Network community is vast and very intricate.
- The relationship among the users in this Network might be beneficial.



- To classify users from this Network into various communities is a grand challenge from multiple researchers.
- So, we developed an **improved k-Clique method** for finding the sub-community from this Network.
- To prove that, this method is very efficient in community detection from various network graphs. It is a challenge for my research.

INTRODUCTION:

- As we know, the primary strategy of the business over the Internet is a recommendation system.

- And according to the number of items has increased to become more congested.

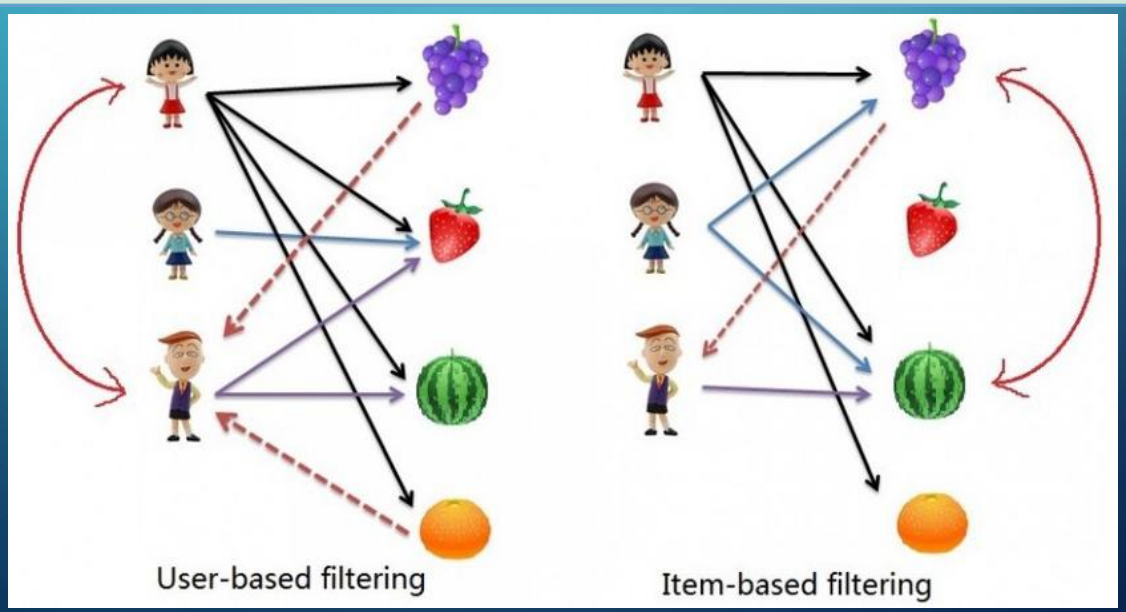
- So, to find items that users are looking for through the existing technologies of a recommendation system is quite a little bit hard.

INTRODUCTION:

The most current recommendation system

For example:

- Collaborative filtering methods.
  - User-based filtering method
  - Item-based filtering method
  - User behavior method
    - **Problem: Cold-Start, Data Sparsity, Scalability**



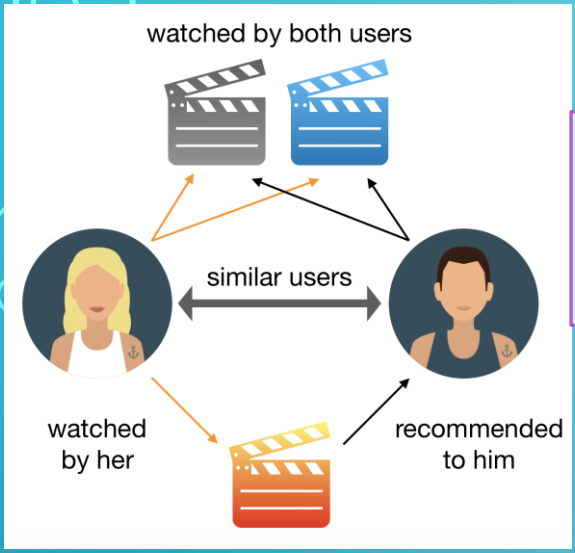
INTRODUCTION: EXAMPLE

NETFLIX

Used Method:

- Viewer ratings, Viewing history, etc.
- Information about the categories, year of release, title, genres, and more.
- Other viewers with similar watching preferences and tastes.
- Time duration of a viewer watching a show
- The device on which a viewer is watching.
- The time of the day a viewer watches.

INTRODUCTION:



- Base on the fundamental knowledge about users who have similar feature information, they might like the same items.

Therefore, the main motivation underlying:

- Create an efficiency network graph from the similarity among user personalized information.
- Find an efficiency method on community detection in various network graph (**improved k-Clique method**).
- Prove the knowledge of users who has similar personalized information might like the same item in the recommendation system.
- Find an algorithm on increasing accuracy in the recommendation system.

RELATED WORK: FORMAL CONCEPT ANALYSIS (FCA) METHOD & FCA NETWORK GRAPH

FCA is a technique of the data analysis that describes the relationship between a particular set of attributes and a specific set of objects.

- There are two types of output.
  - Concept Lattice, which is a collection of formal concepts in data.
  - Attribute implications, which is describing dependencies between attributes of objects.

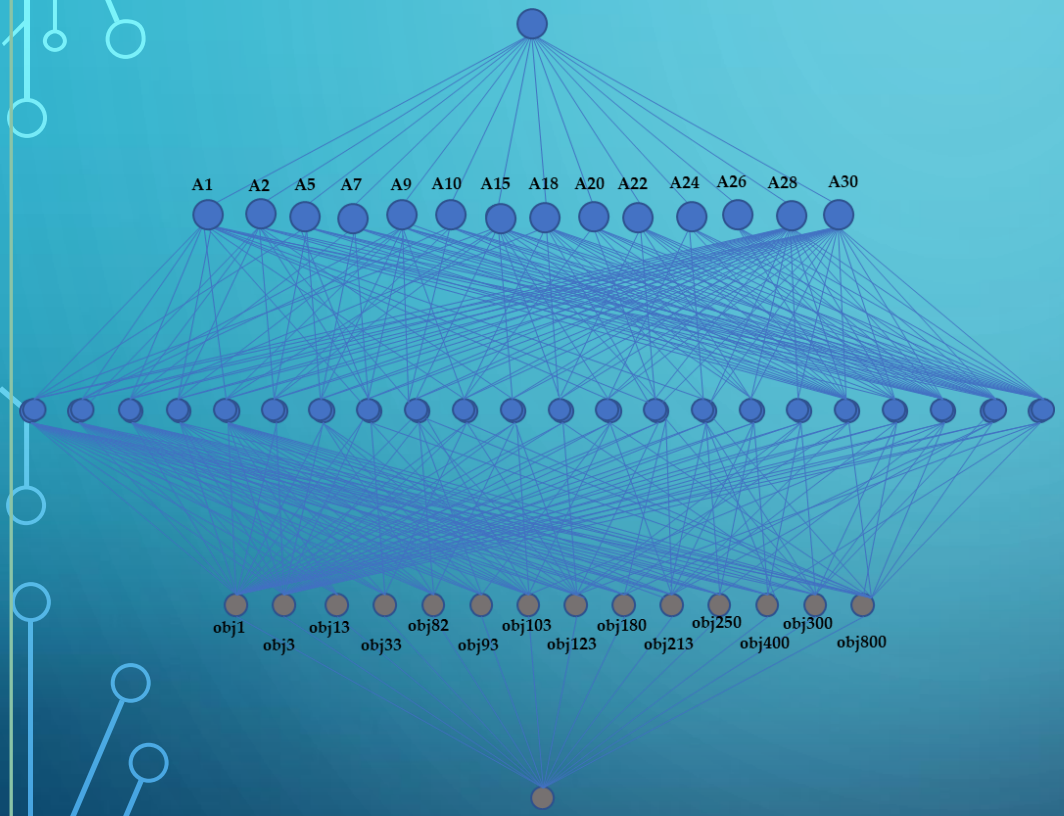
How to create a FCA Network Graph:

- First, created a Formal Context from user's personalized information training dataset.

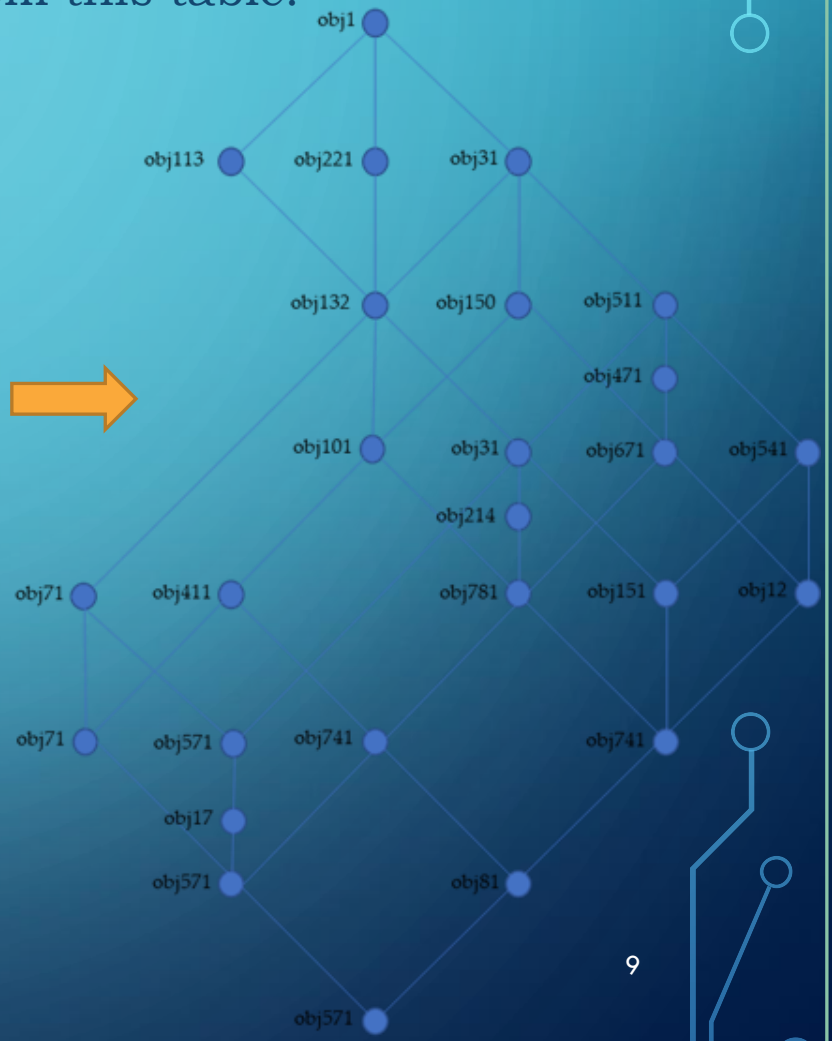
	A: < 20	B: < 30	C: < 40	.....	F: > 80	G: female	H: maile	I: admin	J: artist	H: doctor	.....	DD: other
1	x					x				x		
2		x					x	x				
3			x			x						x
4		x					x		x			
5			x				x			x		
…			x			x		x				
800					x		x		x			

Formal context of a Fundamentals of User's personalized characteristic

- Then, created a concept lattice from the formal context.
- After that, the attribute implications function or association rules function used to create a relationship table that describes as the relationship among the user.
- Finally, a network graph was created from this table.



Concept Lattice



FCA Network Graph

RELATED WORK: K-CLIQUE & IMPROVED K-CLIQUE METHOD

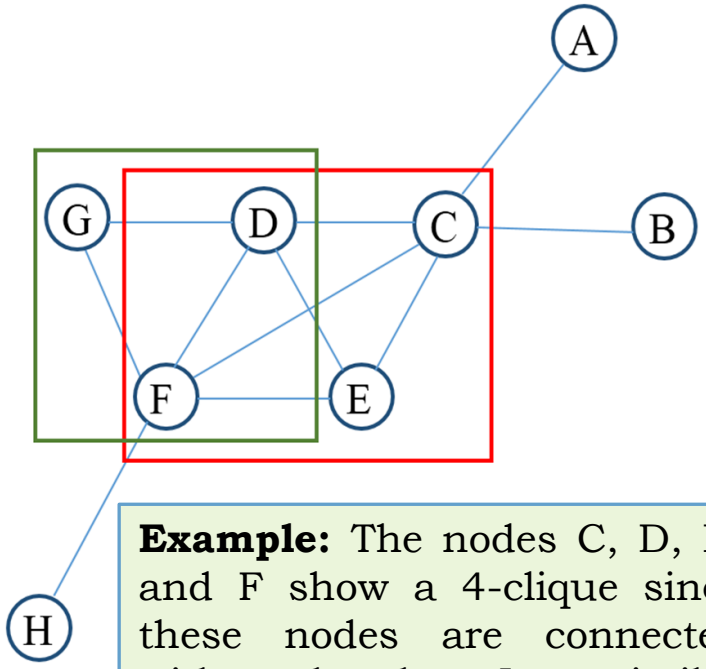
k-Clique:

- It is a technique in the social network analysis
- Used for analyzing data in the complicated community.
- Picks up the communities from network graphs that have a value of *k* cliques.

Improved k-Clique:

- To develop an improved k-clique method:
  - First, a training dataset is created from a Movie Lens Dataset. Second, this dataset used to created a network graph. After that, applied k-Clique to cluster this network graph into several communities by using a value of k = 3 to 14. Later on, a suitable community will find to a new user. Finally, list of 5, 10, and15 movies from suitable community will be recommended to the new users.
- Afterward, we performed an experiment in 10 randomly for each dataset.
  - Rated at least 20 movies = 10 dataset
  - Rated at least 50 movies = 10 dataset
  - Rated at least 100 movies = 10 dataset
  - Rated at least 200 movies = 10 dataset

40 ds



Example: The nodes C, D, E, and F show a 4-clique since these nodes are connected with each other. In a similar manner, the set of nodes D, F, and G show a 3-clique.

PROPOSED METHOD:

Method 1:

An Efficient Movie Recommendation Algorithm Based on Improved k-Clique

Studied, Developed, Applied on various:

- Social Network Analysis Method
- Network Graph Method
- Detecting the Community from the Network Graph (k-Clique)
- Developed an improved k-Clique Method
- Similarity Measure Method
- Collaborative Filtering Method

Method 2:

Personalized Movie Recommendation System Combines Data Mining with the k-Clique

Studied, Developed, Applied on various:

- Network Graph Technologies
- Data Mining Technologies
- Improving algorithm of the proposed work

Method 3:

Movie Recommendation System based on User's Personal Information and Movies' Rated using the method of k-Clique and Normalized Discounted Cumulative Gain

Studied, Developed, Applied on various:

- Social Network Analysis Method like Betweenness Centrality + kNN
- Normal Discounted Cumulative Gain Method
- Improving algorithm of the proposed work

Method 4:

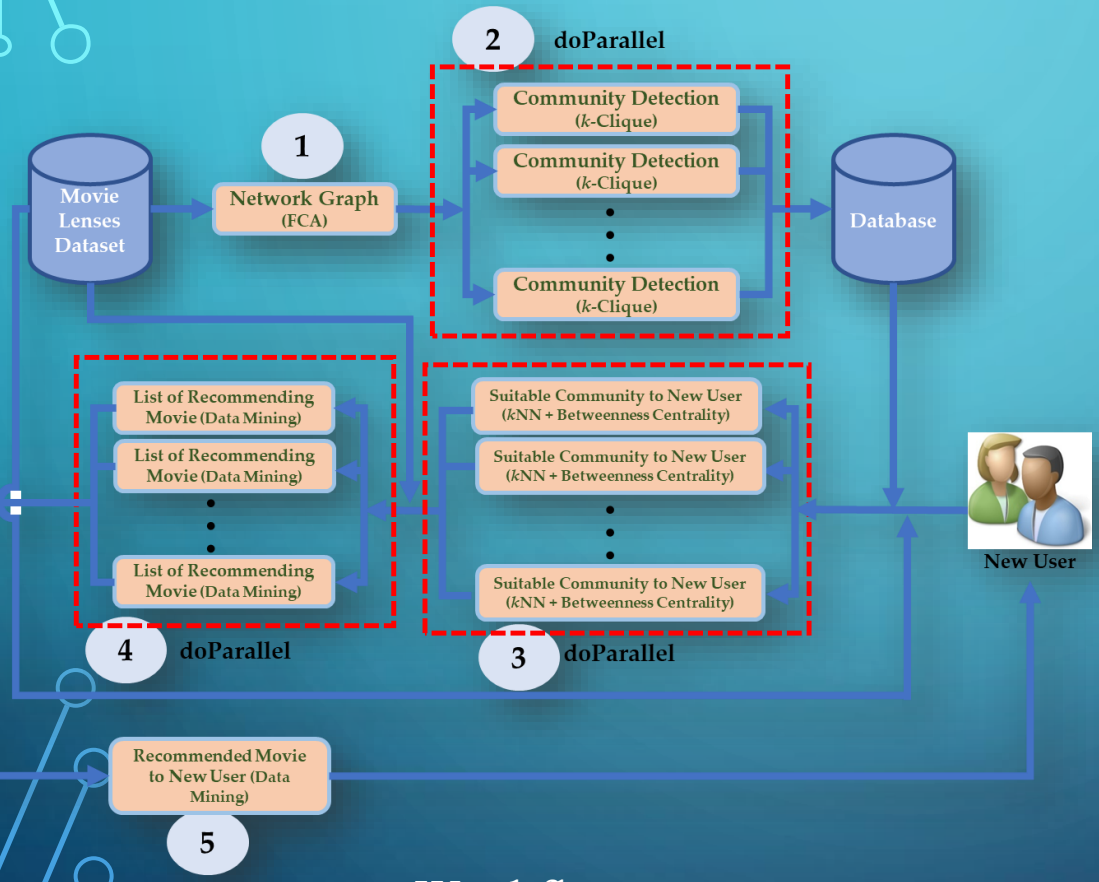
An Efficiency of a DoParallel Algorithm and an FCA Network Graph Applied to Recommendation System

Studied, Developed, Applied on various :

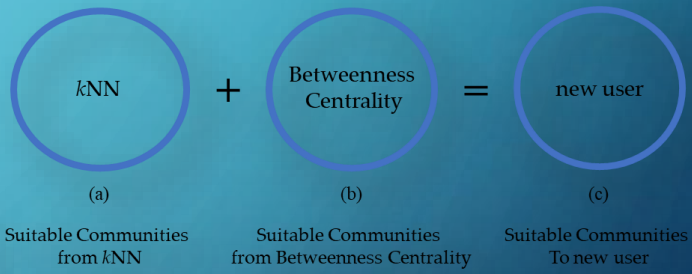
- Formal Concept Analysis(FCA) Method
- Creating a Network Graph Based on FCA
- Various R Parallel Processing Technique

PROPOSED METHOD: METHOD 4

An Efficiency of a DoParallel Algorithm and an FCA Network Graph Applied to Recommendation System



Workflow



Finding Suitable Community

EVALUATION:

- **Experimental Set-Up:**

- **Dataset:**
  - Movie Lens, 100,000 rating data,
  - 193 users, 1,684 movies,
  - 10th random training 80%,
  - 10th random testing 20%,
  - rated 20, 50, 100, 200 movies.
  - Total dataset is 40.

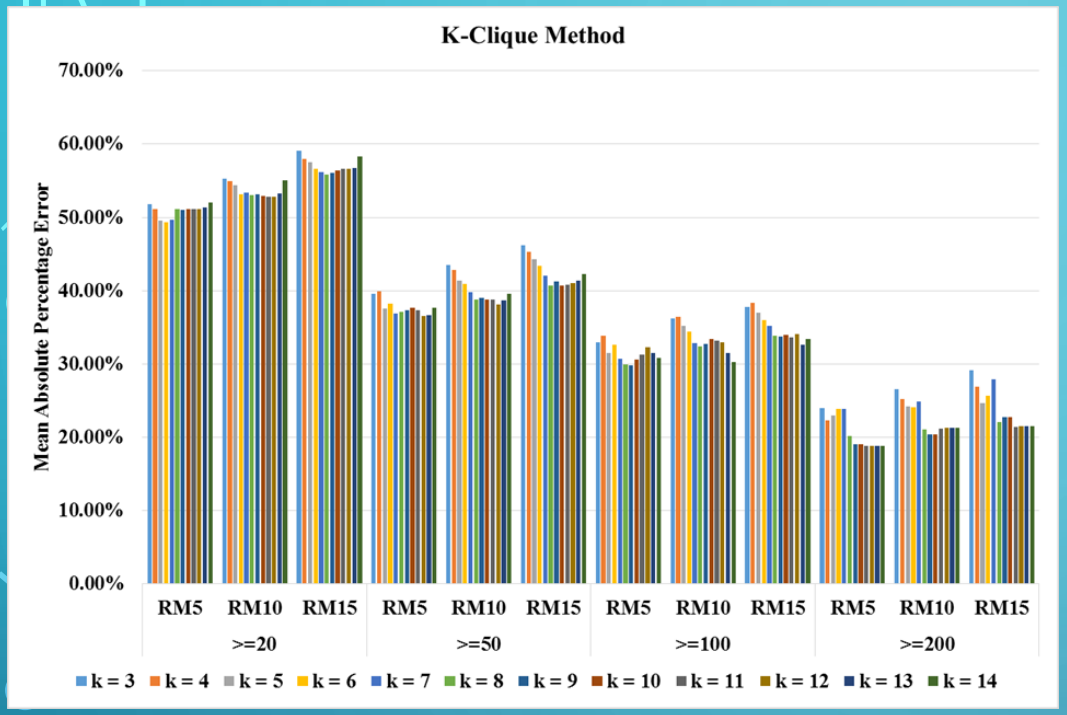
Descriptions	Specification
Operating System	Windows 10 Pro Version 1809
Processor	Intel® Core™ i5 750 @2.67GHz
RAM	24 GB
System type	64 bit Operating System
Hard Disk	500 GB
Data analysis software	R Studio i386 Version 3.4.0, Lattice Minor 2.0

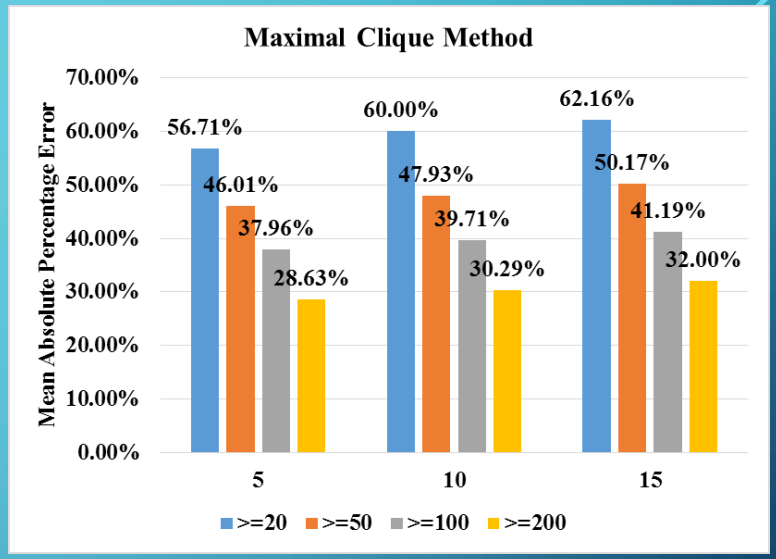
- **Accuracy Measurement Method:**
  - Mean Absolute Percentage Error(MAPE)

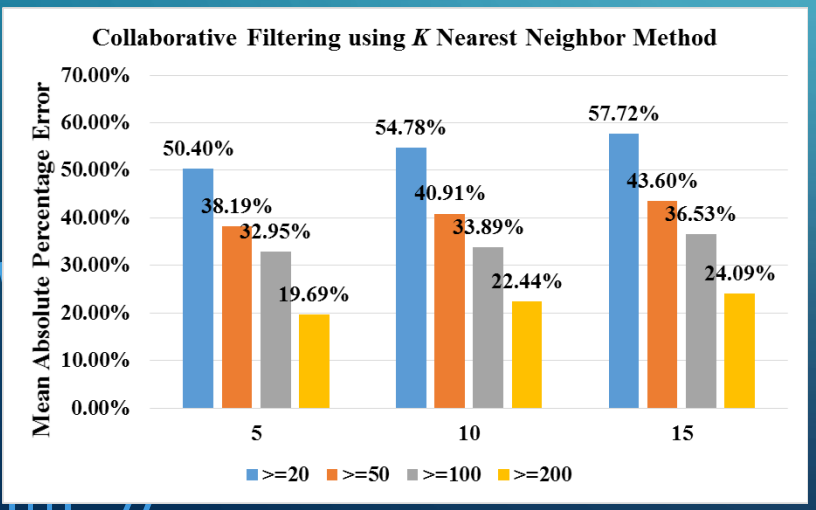
$$\mathbf{MAPE} = \frac{100\%}{n}\sum_{t}^{n}\left|\frac{A_t - F_t}{A_t}\right|$$

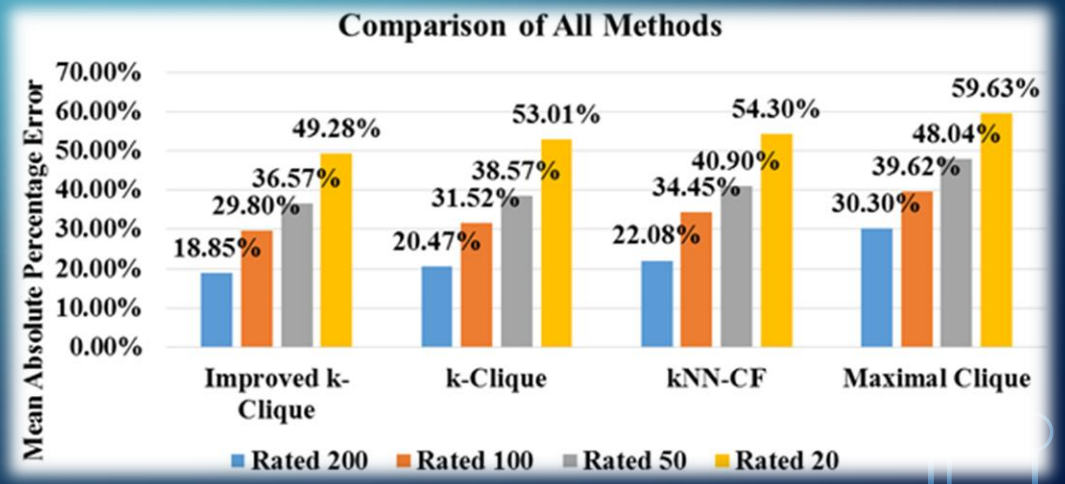
- **Existing Method:**
  - Maximal Clique Method
  - Collaborative Filtering + kNN Method
  - Original Collaborative Filtering Method
  - K-Clique
  - Improved k-Clique
  - K-Clique + Data Mining Method
  - K-Clique + Normal Discounted Cumulative Gain Method
  - Collaborative Filtering + kNN+ NDCG Method

EVALUATION: METHOD 1 RESULT

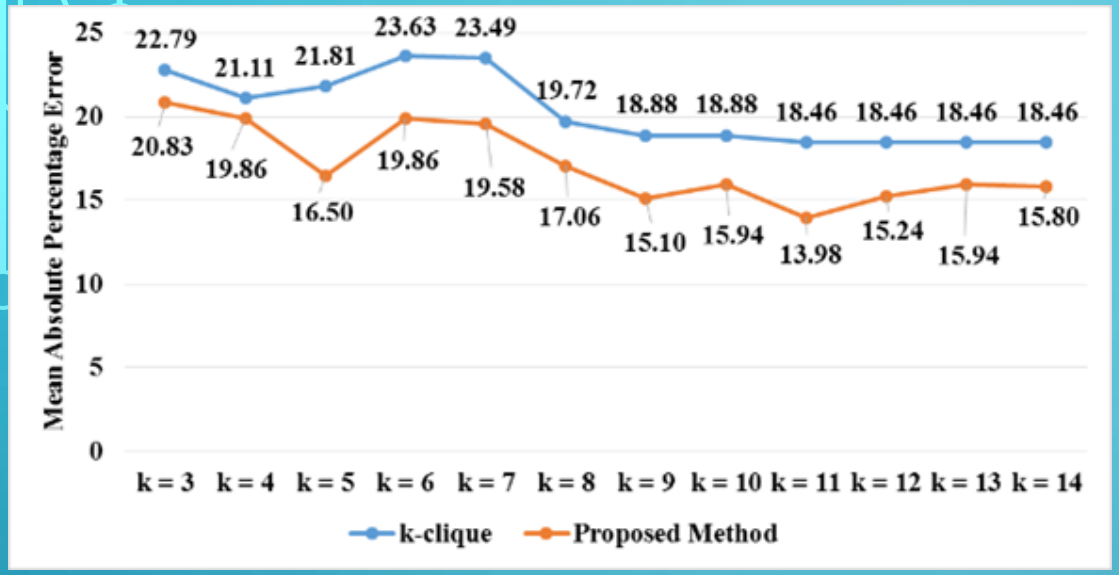






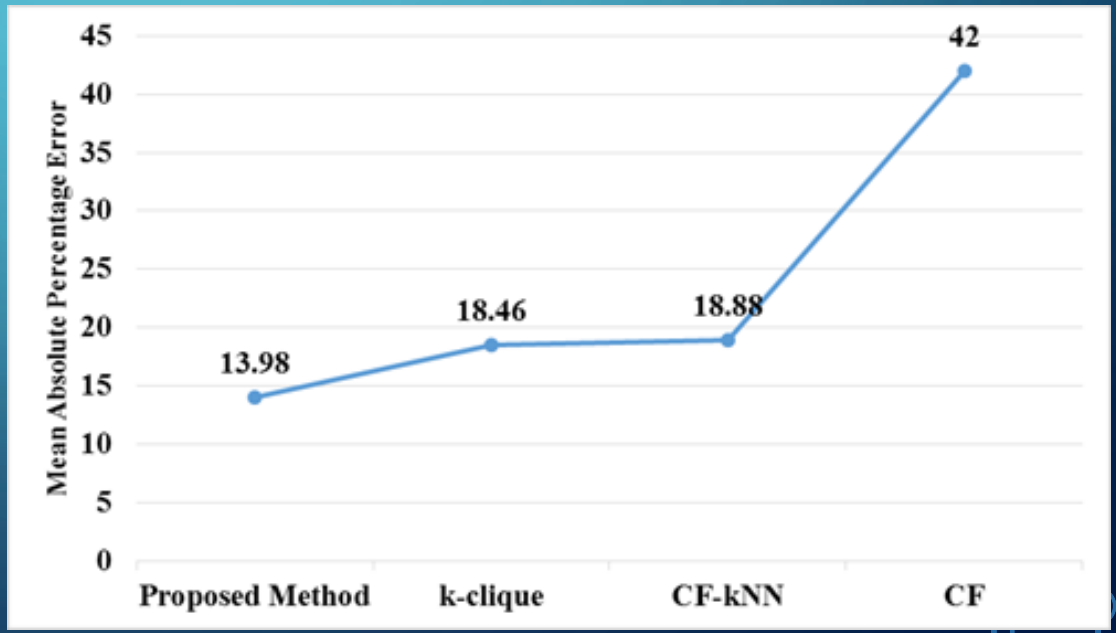


EVALUATION: METHOD 2 RESULT

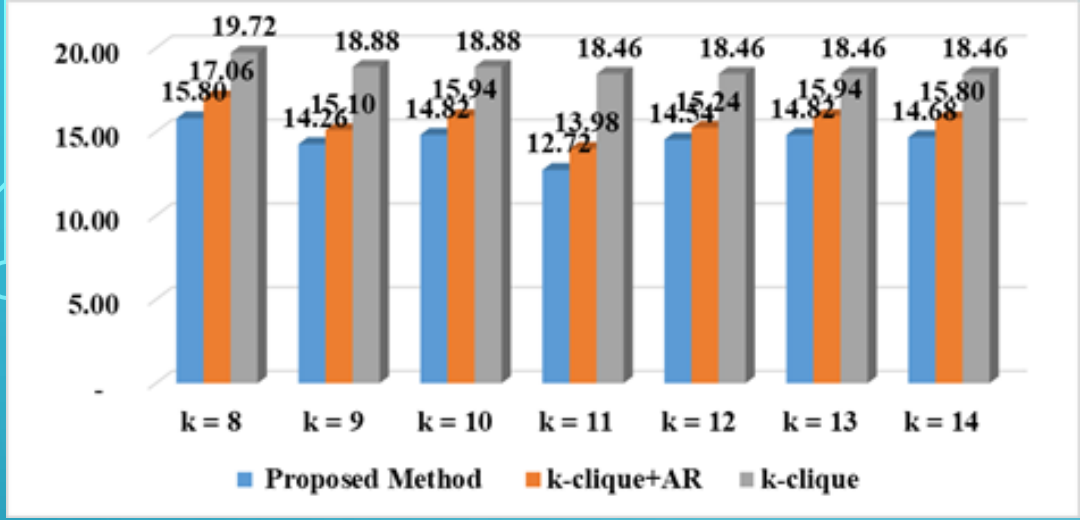


Comparison result of k-clique and proposed work

Comparison result of proposed work and existing method

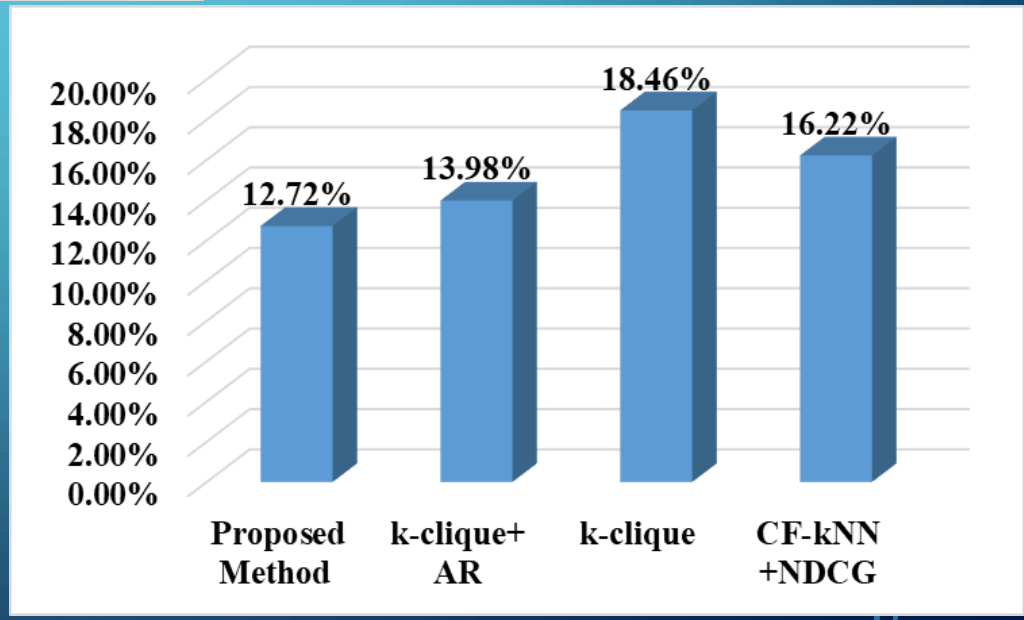


EVALUATION: METHOD 3 RESULT

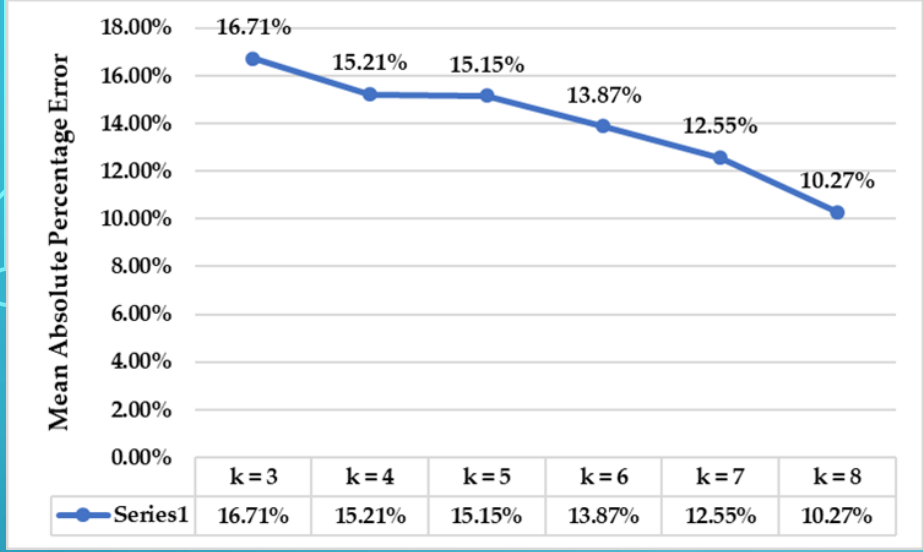


Result Comparison

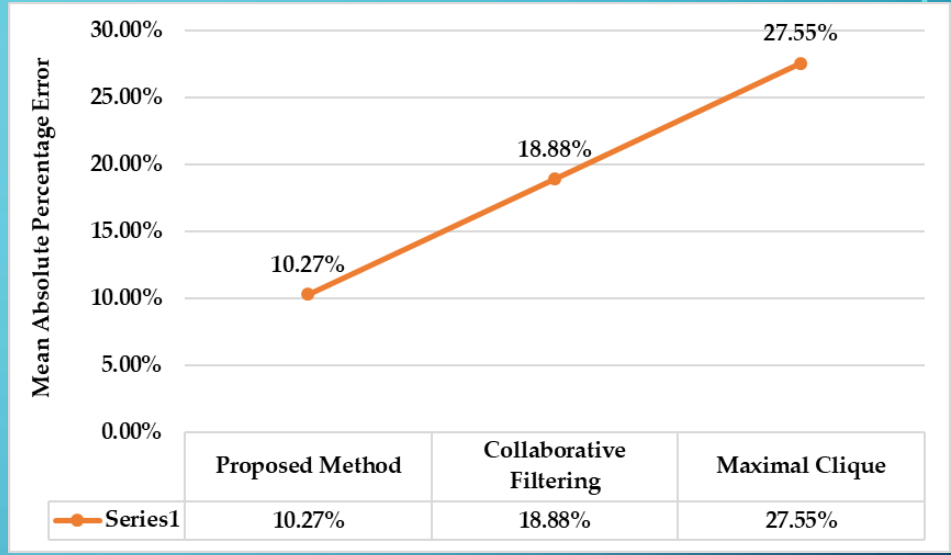
Result Comparison



EVALUATION: METHOD 4 RESULT



K-Clique Result



Comparison Result

#user	Number of Community
k=3	176
k=4	135
k=5	110
k=6	101
k=7	83
k=8	75

Number of Community

	Sequential	doParallel
Proposed Method	48 hours	12 hours

Comparison of Time

CONCLUSION:

- **The major significant findings:**

  We are the first on developed and applied various methods:

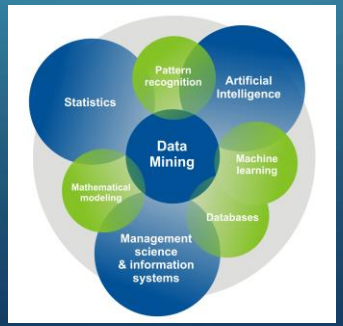
  - Developed a social network graph from the feature similarity among the users.
  - Applied k-clique method to classifies communities in the social network graph.
  - Developed improved k-Clique method.
  - Applied improved k-clique to the recommendation system.
  - Developed a network graph from a Formal Concept Analysis(FCA) method.
  - Applied an FCA network graph to the recommendation system.
  - Applied a “doParallel” algorithm to divide the proposed work algorithm into several processes, then executed them as parallel processing using the multi-core CPU.

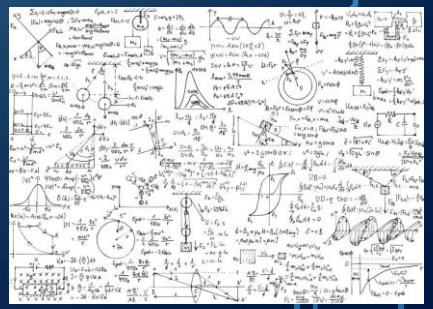
- **The result of the experiment was proved that an improved k-clique was:**
  - Efficient in detecting the community from various network graphs.
  - Very significantly increasing accuracy in the recommendation system.
  - Solved a Cold-Start problem.

FUTURE STUDY:

- Refer to the limitation of the experimental set-up,
- Limitation of the dataset which provides a few features information of users
- and the complexity of the proposed algorithm.

- **In future work:**
  - The various type of datasets and various feature information of users will be applied to the proposed work.
  - The proposed method algorithm would be a modification for easy to apply in real work.





ຂອບໃຈ

Thank You