VIETNAM NATIONAL UNIVERSITY, HANOI
UNIVERSITY OF ENGINEERING AND TECHNOLOGY

Dat Mai-Cong

THE ROLE OF SOCIAL TIES IN SOCIAL RECOMMENDATION SYSTEMS

Major: **Computer Science**

HA NOI - 2015

VIETNAM NATIONAL UNIVERSITY, HANOI
UNIVERSITY OF ENGINEERING AND TECHNOLOGY

Dat Mai-Cong

THE ROLE OF SOCIAL TIES IN SOCIAL RECOMMENDATION SYSTEMS

Major: Computer Science

Supervisor: Assoc. Prof. Dr. Thuy Ha-Quang
Co-Supervisor: MSc. Le Luong-Thai

HA NOI - 2015

AUTHORSHIP

"I hereby declare that the work contained in this thesis is of my own and has not been *previously submitted for a degree or diploma at this or any other higher education institution. To the best of my knowledge and belief, the thesis contains no materials previously published or written by another person except where due reference or acknowledgement is made."*

Signature:………………………………………………

SUPERVISOR'S APPROVAL

"I hereby approve that the thesis in its current form is ready for committee examination as a requirement for the Bachelor of Computer Science degree at the University of Engineering and Technology."

Signature:………………………………………………

ACKNOWLEDGEMENT

First of all, I would like to express my sincere thanks to my advisors Assoc. Prof. Dr. Thuy Ha-Quang for his support and guidance throughout this thesis work.

I am grateful to MSc. Le Luong-Thai to her support and reviews for this thesis.

I would like to give a big thank to brother and sister in Knowledge and Technology Laboratory (KT-lab) who have supported me to complete this research.

I would also like to give gratitude to University of Engineering and Technology that are provided the environment and conditions for my learning.

I am greatly indebted to my family for their encouragements, unconditional support and patience.

Because time is limited and the condition of this thesis is inevitable shortcomings, I look forward to the comments of the teacher and the concern you have with this issue.

ABSTRACT

Social Recommendation Systems have received increasing attention of scientists in recent years. Many researches are published in this field such as Jiliang Tang et al (2013) [1], Jiliang Tang, Jie Tang, HuanLiu (2014) [2]. The increasing grown of social network also brings many opportunities to improve Recommendation Systems [3] [4]. Social theories, models for Social Recommendation Systems are developed to explain and prove the positive effect of social relation to quality of Social Recommendation Systems [4]. In which, Social tie strength is also used to improve quality of Recommendation Systems.

This thesis focuses on exploiting the effect of Social Tie to the performance of Recommendation Systems based on some researches in [3] [5] [6]. Based on these researches, the thesis has proposed a model for mining the social tie strength to enhance quality of Recommendation Systems in two dimensions of tie strength: *Appearances together in photos, Number of friends in common*. Simultaneously, the thesis also implements this model as experiment and collects data by using a survey of rating for 99 movies to 80 Facebook users. Experimental results show that the exploitation of tie strength was initially effective in improving the social recommendation.

***Keywords*: *Social Recommendation Systems, Recommendation Systems, Social Ties, Tie Strength, Collaborative filtering, Social Theory, Social media*.

.

TÓM TẮT

Trong những năm gần đây, hệ tư vấn xã hội ngày càng nhận được sự quan tâm từ các nhà khoa học, có nhiều nghiên cứu về hệ tư vấn xã hội được công bố như các nghiên cứu của Jiliang Tang và cộng sự (2013) [1], Jiliang Tang và Jie Tang, HuanLiu (2014) [2]. Sự phát triển của mạng xã hội cũng mang lại nhiều cơ hội cho việc cải thiện chất lượng hệ tư vấn [3] [4]. Các lý thuyết xã hội và một số mô hình tư vấn cũng được phát triển để giải thích và chứng minh cho vai trò của qua hệ xã hội trong các hệ tư vấn [4]. Trong đó, độ mạnh liên kết giữa các người dùng trong mạng xã hội cũng được sử dụng để tang chất lượng tư vấn.

Khóa luận tập trung vào việc khai thác độ mạnh liên kết của các người dùng trong mạng xã hội dựa trên các nghiên cứu trong [3] [5] [6]. Dựa trên các cơ sở nghiên cứu đó, khóa luận đã đề nghị một mô hình khai thác liên kết xã hội để tăng cường tư vẫn xã hội dựa trên độ mạnh liên kết tính theo hai tham số là "số bạn chung", và "số ảnh chung". Khóa luận cũng đã xây dựng, cài đặt mô hình trên và thu thập dữ liệu dựa trên một khảo sát đánh giá 99 bộ phim của 80 người dùng trên mạng xã hội Facebook. Kết quả thực nghiệm cho thấy việc khai thác độ mạnh liên kết đã có tác dụng bước đầu trong việc cải thiện chất lượng tư vấn.

Từ khóa: *Social Recommendation Systems, Recommendation Systems, Social Ties, Tie Strength, Collaborative filtering, Social Theory, Social media*.

.

TABLE OF CONTENTS

List of Figures

List of Tables

ABBREVATIONS

CF	Collaborative filtering
TS	Ties Strength
TF-IDF	Term frequency–inverse document frequency
TF	Term frequency
IDF	Inverse document frequency
SVD	Singular value decomposition
MAE	Mean absolute error
NMAE	Normalized mean absolute error
RMSE	Root mean squared error

Chapter 1

INTRODUCTION

1.1. Motivation

Nowadays, people are always faced with the making decision such as what to wear? What movie to see? What something to buy? What book to read? What game to play? And so on. Recommendation Systems are developed to help online users solving these tasks. Using Recommendation Systems means that use the wisdom of the crown [3], to support making a choice process. Recommendation Systems are used in many online systems and they are very important in the success of online websites such as Amazon.com, Epinions.com, Netflix, and MovieLens.org [5]. In the techniques of Recommendation Systems, the highlight is collaborative filtering. Collaborative filtering is introduced in 1990s, that technique predicts the user's interest based on ratings information from other similar users or other similar items.

The quality of Recommendation Systems is very important, so, how to improve this quality is also necessary. Nowadays, the development of social network brings the opportunity to improve the quality of Recommendation Systems. For example, it can be used diversity of relationship with the communities (such as "trust" on Epinions.com, "reputation" on eBay

…). In the thesis, the role of Social Ties Strength is focused to improve Recommendation Systems.

1.1.1. Social Network with Tie Strength

Social network is a network model has social nature. It consists of nodes and edges where nodes are linked together by edges as a relationship. Each node is an entity in the network. Each entity can be a person, a community, a company, or movie… and the entity interacts by an edge, each edge can be friend relation, partner relation, enemy relation … Figure 1.1 shows an example about social network with nodes and edges.

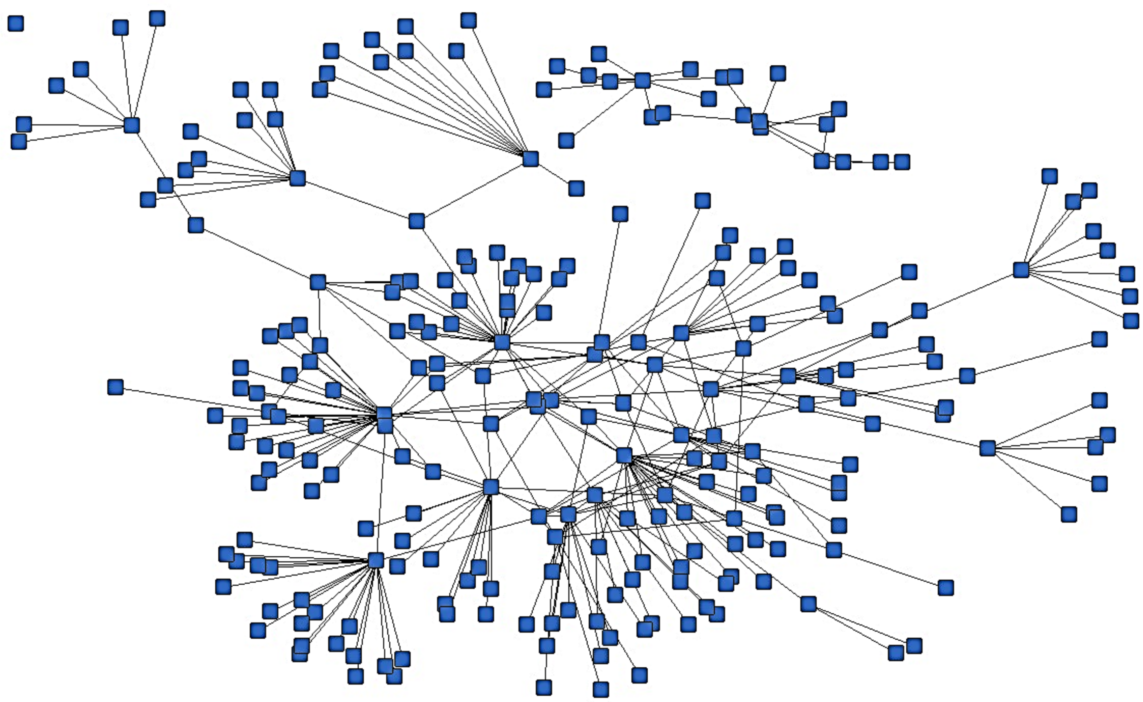

Figure 1.1: An example of social network diagram.

As a mentioned before, each node plays one role in social network and each edge also plays one role too, which means, edges play different role. For convenience, the concept **tie strength** is in use. In other words, tie strength quantifies the characteristics of two notes. Tie strength can divide into **strong tie** and **weak tie** [7]. The relations between the family, close friend are also known as strong ties, and the relations of acquaintances are called weak tie. In chapter 2, Tie Strength and their characteristics are presented in detail.

1.2. Contributions and thesis overview

The purpose of this thesis is to investigate about Social Ties and their dimension, how to use the Social Ties to improve Recommendation Systems. Secondly, thesis implements some algorithms about Recommendation Systems as collaborative filtering and integrates the collaborative and tie strength.

The rest of this thesis is organized as follows.

Chapter 2 provides theoretical background, focus on Recommendation Systems and Social Tie strength theory. At first, Recommendation Systems are introduced by presenting about Recommendation Systems techniques as Content-based filtering, Collaborative filtering, Hybrid Recommendation Systems in details. Then, the thesis presents the way to evaluate a Recommendation Systems and some common problems of Recommendation Systems. At second, the thesis presents Social Recommendation and effects of social factor to make the difference between Social Recommendation and traditional Recommendation Systems. The last of this chapter, thesis will concentrate on Social Tie, Tie Strength and their characteristics. In this section, features and dimensions of social ties are represented.

In chapter 3, firstly, the positive effect of Social Tie Strength to the quality of Recommendation Systems are determined by giving exists researches of Koroleva and Štimac in [8], Li et al in [9], Oliver Oechslein and Thomas Hess in [5]. Secondly, a model is proposed to illustrate the positive influence of Tie Strength to Recommendation Systems rather than traditional Recommendation Systems based on experiments of Arazy O et al in [6]. In this model, four phrases are constructed that consist of **Data preprocessing** for raw data preprocessing, **Collaborative filtering system** and **Social Collaborative filtering system** to implement the Collaborative filtering algorithm and Collaborative filtering combined with Tie strength, and **Evaluation** for making a comparison between two algorithms.

In chapter 4, the model in the chapter 3 was implement, then, results are evaluated. Results obtained are positive to prove that the positive effect of Social Tie strength to Recommendation Systems.

Lastly, chapter 5 is conclusions and future works. In this chapter, we conclude all what we did in this thesis, also its strength and weakness; then we show some work we need to do in future.

LITERATURE REVIEW

2.1. Traditional Recommendation Systems

Recommender Systems are a subclass of Information Filtering system that use to predict the preference or interest of user to item [10] [11]. User is a person who uses internet services (e.g. user on MovieLens.org, user on Yahoo.com …). Item is a something that user interest. It is also a product that user want to receive advice or want to make recommendations (e.g. movies, books, music, news, Web page, images …). The level of preference that user evaluates to an item is called a *rating*. These ratings can take many forms, it depends on the system in question [12]. The rating value can be real or integer number, such as the rating value might be from 1 to 5 stars. Some Recommendation Systems use the binary scale as like/dislike, trust/distrust. A person can rate for one or more items. Each item can receive evaluation from one or more people.

The set of all value of triple (*User*, *Item, Rating*) refers to *ratings matrix*. (*User, Item*) pairs that user do not rate for item are unknown values in the ratings matrix [12]. Moreover, the task of Recommendation Systems is filled the unknown value in ratings matrix. The below figure shows the example about the ratings matrix. In the Figure 2.1, there are four movies (*Batman Begins, Alice in Wonderland, Dumb and Dumber, Equilibrium*) and three users (User A, User B, User C) in a movie Recommendation Systems. Ratings value is in 5-star scale.

	Batman Begins	*Alice in Wonderland*	*Dumb and Dumber*	*Equilibrium*
User A	4	?	3	5
User B	?	5	4	?
User C	5	4	2	?

Figure 2.1: Example about ratings matrix in 5-stars scale.

The cell with marking by "?" symbol shows the not rated value (*unknown value* in rating matrix). That means, user A does not rate *Alice in Wonderland* movie. User B does not rate for *Batman Begins* and *Equilibrium* movies, user C does not rate for *Equilibrium* movie.

In this thesis, some notations in Recommendation Systems are denoted for the later chapters. Definition that:

- $U = \{u_1, u_2, \dots, u_n\}$ is set of n users. $I = \{i_1, i_2, \dots, i_m\}$ is set of m items.
- I_u is set of items rating by user u, U_i is set of users who rating for item i.
- $\boldsymbol{R}$ is ratings matrix, $r_{u,i}$ is the rating between user u and item i.
- r_u is ratings vector of user u, r_i is the ratings vector for item i.
- $\bar{r}_u, \bar{r}_i$ is the average rating value of user u or item $\boldsymbol{i}$.
- $p_{u,i}$ is the prediction value between user u and item i.
- $\pi_{u,i}$ is the preference between user u and item i. (Note that preference is differed from rating value, but we can assume that $r_{u,i} \approx \pi_{u,i}$)

There are some kinds of Recommendation Systems, by [10] [11], Recommendation Systems can classify in three types:

- Content-based filtering: this approach is based on the characteristics and content of an item and the preferences of a user (or user profile).

- Collaborative filtering: this approach is based on the amount of information from collaborative users or the similar items.

- Hybrid Recommendation Systems: integration of Content-based filtering and Collaborative filtering.

The Figure 2.2 shows an example about three types of Recommendation Systems.

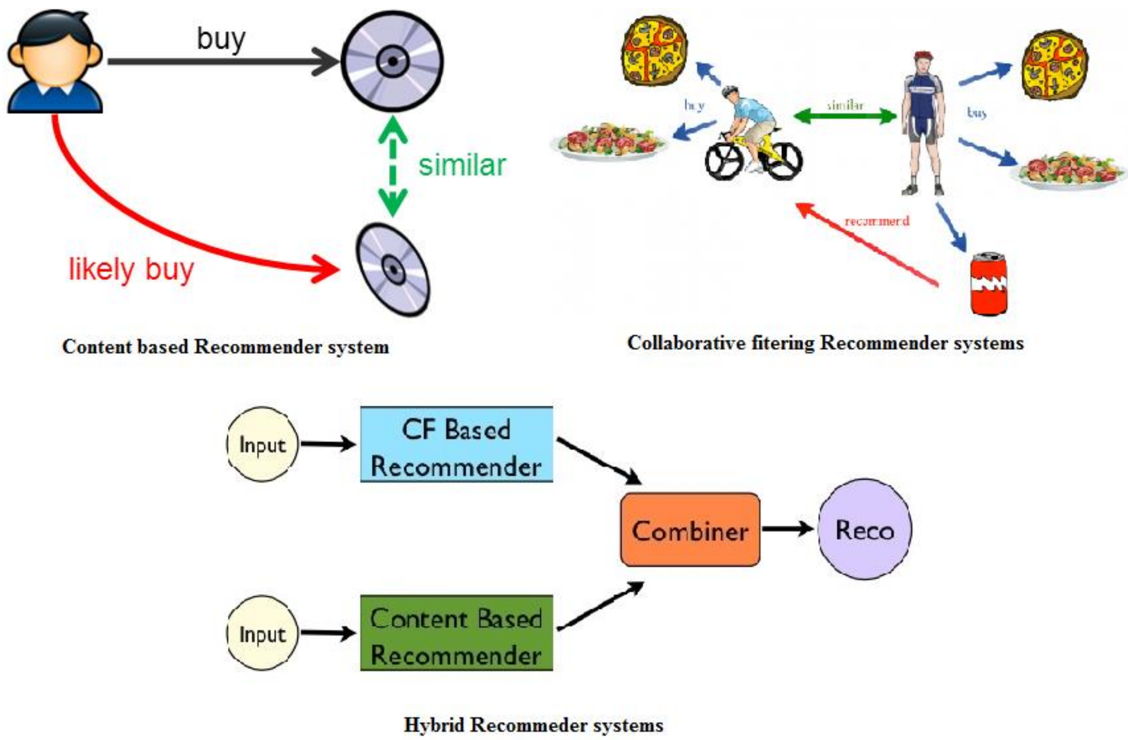

Figure 2.2: An example about Content-based filtering Recommendation Systems , Collaborative filtering Recommendation Systems , Hybrid Recommendation Systems.

2.1.1. Content-based filtering approach

Content-based filtering approach is based on the correlation between items content and user profile (or user preferences) [13]. The content of each item is described by a set of keywords, besides that, the user's profile is built on the type of item that user likes. The Recommendation Systems use content-based filtering approach recommend items that similar to items which user liked in the past. For example, if a user were rated for a book in love novel, Recommendation Systems would learn and make recommendation other books in this type (love novels).

To present features of the items, the "**TF-IDF**" (term frequency–inverse document frequency) algorithm is in use. TF (or term frequency) weight of a key word is a frequency of this word in a document. IDF (or inverse document frequency) of a key word is an inverse of this word frequency in the document.

To make a user profile, there are two type of information is focused on:

- A model of the user's preference
- A history of user's interaction with Recommendation Systems

In [14], users and items are presented in vectors. $i_{j,k}$ is a weight of keyword k in content v_j. v_j is presented by set $I_j = \{i_{j1}, i_{j2}, \dots, i_{j,k}\}$. $u_{j,k}$ is profile of a user with keyword k that user u_i used to rate an item in the past. This can be rewritten the user u_i by a set of profile as below: $U_i = \{u_{i1}, u_{i2}, \dots, u_{i,k}\}$. To calculate the correlation between user i and item j, it can be used cosine correlation of two vector U_i and I_j :

$$sim\left(U_i, I_j\right) = \cos\left(U_i, I_j\right) = \frac{\sum_{l=1}^{k} u_{i,l} i_{j,l}}{\sqrt{\sum_{l=1}^{k} u_{i,l}^2 \cdot \sum_{l=1}^{k} i_{j,l}^2}} \qquad (2.1)$$

In addition, Recommendation Systems based on content-based approach are also using Bayes classification, decision tree, neutron network…

2.1.2. Collaborative filtering approach

Collaborative Filtering is a popular algorithm that automatically predicts the interest of an active user by collecting rating information from other similar users or items. The underlying assumption of Collaborative Filtering is that the active user will prefer those items which the similar users prefer [15]. Collaborative Filtering can be divided into two approaches: Memory-based and Model-based.

The Memory-based approaches (It is also known as Nearest Neighbor Collaborative Filtering) are very popular algorithm in the commercial Collaborative Filtering system [16] [17]. It was based on the interaction history of users in the past to make a recommendation.

The Model-based approaches is algorithm that built a model of user rating by computing the expected value of user's prediction. This algorithm uses the data-mining, machine learning to find pattern based on training dataset.

The Figure 2.3 demonstrates the common process of collaborative filtering systems.

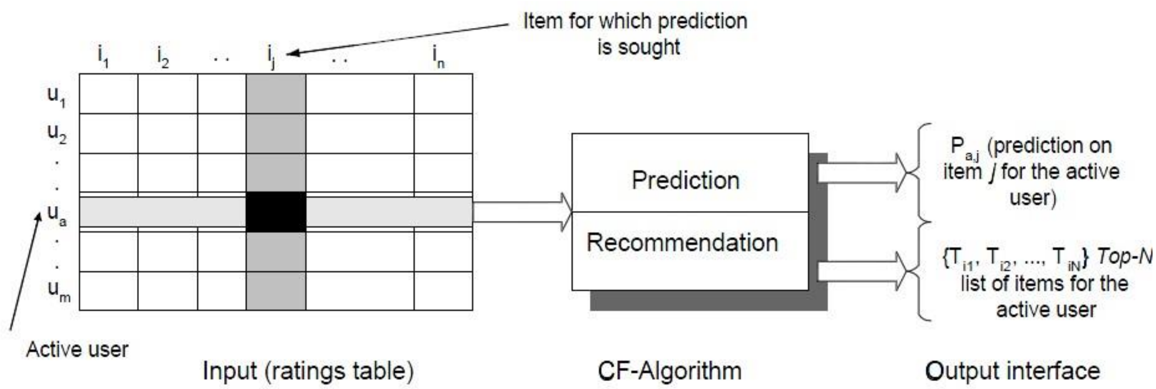

Figure 2.3: Collaborative filtering process.

Collaborative Filtering algorithms represent the entire $m \times n$ user-item data as a ratings matrix A. Each entry $a_{i,j}$ in A represent the preference score (ratings) of the ith user on the jth item. Each individual ratings are within a numerical scale and it can as well be zero indicating that the user has not yet rated that item.

2.1.2.1. Memory based approach

Memory based methods use user-item matrix or sample to predict the unknown value [1]. It can be divided into User-based methods and Item-based methods.

2.1.2.1.1. User-based methods

User-based collaborative filtering (also known as k-NN collaborative filtering) was introduced in the article [17]. This method finds the similar users to the current user, that similar users and current user must have both rated on the same items. For example, to predict Nam's interest for item A he does not rate, this method finds the users that have high agreement with Nam on the items they have both rated (for example Nguyen, Dung, Thanh). Then, the rating of Nguyen, Thanh, Dung to item A are weighted by level agreement with Nam to predict the interest of Nam to item A.

User-based CF system requires three components: rating matrix $\boldsymbol{R}$, similarity function $s: U \times U \rightarrow \mathbb{R}$ to compute the similarity between two users and a method to predict the user preferences [12].

Rating matrix $\boldsymbol{R}$ is defined in the previous section, now, we go to compute the prediction method and compute similar user's method.

a. Computing prediction

To calculate the prediction for a user u, user-based CF uses similar function $s: U \times U \rightarrow \mathbb{R}$ to find the set of neighborhood $N \subseteq U$ of u's neighbors. Then, the system combines the user's rating in N to calculate the interest of user u to item i. The weight of user in N is the similarity of them to the current user. The following equation is used to generate the predictions:

$$p_{u,i} = \bar{r}_u + \frac{\sum_{u' \in N} s(u,u')(r_{u',i} - \bar{r}_{u'})}{\sum_{u' \in N} |s(u,u')|} \qquad (2.2)$$

Subtracting the user mean rating in equation 2.2 to avoid the case some users has tended to give higher rating or lower rating to an item than other ones.

The important problem is how many neighbors to select. In some Recommendation Systems system, such as Grouplens, all users are considered as neighbors [17]. In some others, the size of the set N is depended on similarity threshold [12]. If the size of neighbors set is large, the prediction value will be more accurate. However, the complexity of computing is large too. Therefore, it is balanced between the accuracy of prediction and the complexity.

b. Computing user similarity

Computing user's similarity plays important role in implementation User-based CF, considering some similarity function as Cosine similarity, Pearson correlation, Constrained Pearson correlation.

Cosine similarity

In this algorithm, users are presented as $|I|$-dimension vectors (I is set of items). User similar is cosine distance between two ratings vectors:

$$s(u,v) = cosin(\overrightarrow{r_u}, \overrightarrow{r_v}) = \frac{r_u \cdot r_v}{\|r_u\|.\|r_v\|} = \frac{\sum_i r_{u,i} r_{v,i}}{\sqrt{\sum_i r_{u,i}^2}\sqrt{\sum_i r_{v,i}^2}} \qquad (2.3)$$

If the value of similarity is 1, two vectors are the same orientation, if that value is 0, two vectors is crossed, user u and v are distinct. In addition, if this value is -1, two is not similar.

Pearson correlation

This algorithm calculates the similarity between two users by computing the statistical correlation of two users that have the common rating [12]. Pearson correlation allows to compute high similarity of users that have few common ratings. The correlation is calculated as follow equation:

$$s(u,v) = \frac{\sum_{i \in I_u \cap I_v} (r_{u,i} - \bar{r}_u)(r_{v,i} - \bar{r}_v)}{\sqrt{\sum_{i \in I_u \cap I_v} (r_{u,i} - \bar{r}_u)^2} . \sqrt{\sum_{i \in I_u \cap I_v} (r_{v,i} - \bar{r}_v)^2}} \qquad (2.4)$$

In this algorithm, threshold for number of co-rated items for correlation can be set to reduce the complexity of computation.

Constrained Pearson correlation

The Constrained Pearson Correlation is computed by the following equation:

$$s(u,v) = \frac{\sum_{i \in I_u \cap I_v}(r_{u,i} - r_z)(r_{v,i} - r_z)}{\sqrt{\sum_{i \in I_u \cap I_v}(r_{u,i} - r_z)^2} . \sqrt{\sum_{i \in I_u \cap I_v}(r_{v,i} - r_z)^2}} \quad (2.5)$$

Where r_z is the neutral value (neither like nor dislike). For example, Ringo system is rating in 7-scale, and, 4 is neutral value.

Others Correlation

There are some others correlation such as Spearman rank correlation, mean-squared difference… Nevertheless, in this thesis, they are not mentioned.

c. Example

Considering one example to deeply understand User-based method, this example is available in [12]. However, all calculation are represented.

	Batman Begins	*Alice in Wonderland*	*Dumb and Dumber*	*Equilibrium*
User A	4	?	3	5
User B	?	5	4	?
User C	5	4	2	?
User D	2	4	?	3
User E	3	4	5	?

Figure 2.4: Example ratings matrix

Observing the ratings matrix in Figure 2.4, the task is that finding the prediction of User C for movie *Equilibrium*. Using bellow configurations:

- Pearson correlation.
- Neighborhood size of 2.
- Weighted average with mean offset (Using equation 2.1)

C's mean is: $r_C = \frac{5+4+2}{3} = 3.667$

There are two users rating for *Equilibrium* so A, D are two neighbors, so,$\bar{r}_A = \frac{4+3+5}{3} = 4$, $\bar{r}_D = \frac{2+4+3}{3} = 3$, since, the similar between A C, D C are:

$$s(A,C) = \frac{(4-4)(5-3.667)+(3-4)(2-3.667)}{\sqrt{(4-4)^2+\ (3-4)^2}\sqrt{(5-3.667)^2+\ (2-3.667)^2}} = 0.781$$

$$s(D,C) = \frac{(2-3)(5-3.667)+(4-3)(4-3.667)}{\sqrt{(2-3)^2+\ (4-3)^2}\sqrt{(5-3.667)^2+\ (4-3.667)^2}} = -0.515$$

From above, the prediction of C for *Equilibrium* is:

$$p_{C,Equilibrium} = 3.667 + \frac{0.781.(5-4)+\ -0.515.(2-3)}{0.781+0.515} = 4.667$$

Therefore, the prediction between user C and *Equilibrium* movie is 4.667.

2.1.2.1.2. Item-based methods

From above section, it is clear that user-based CF uses the similarity between two users to compute the prediction, similar to user-based, item-based CF uses the similarity between the rating patterns of items. If two items are received the same behavior from users (like or dislike, trust or distrust …) then they are similar. Users have tended to receive recommendation for similar items. This method is similar to user-based method, but items correlation is deduced from user's interest patterns rather than selected from items data [12]. Item-based methods find the most similarity of items to make predictions. In almost systems, the number of users is larger than items; it allows neighbors finding that is simpler than user-based CF.

Finding the similar users is more complicated than before because when user rates or re-rates items, their rating vectors are changed, which means, the neighborhood's determine belong to other users. Since, the results of predictions will be changed. For this reason, almost user-based CF systems find the neighbors set at the prediction time are needed [18].

Item-based CF systems use the user's rating for items and item's similarity to generate predictions or recommendations. It has required some components: similarity function $s: I \times I \rightarrow \mathbb{R}$ and method to calculate the predictions (or recommendations) from ratings and similarities [12] [18].

a. Computing Prediction

Similar to user-based CF procedure, in the item-based CF procedure, the neighbors of items set (similar item set) S are found, In S set, k items have the most similar to current item i and have the rating by user u are chosen. In the [18], Sarwar et al found $k = 14$ is good for MovieLens dataset.

After collecting S, if choosing the similar score as weight, the predictions as follows equation:

$$p_{u,i} = \frac{\sum_{j \in S} s(i,j).r_{u,j}}{\sum_{j \in S} |s(i,j)|} \quad (2.6)$$

In the equation 2.5, the rating value is nonnegative, but the similarity score can be negative, so the prediction can be negative, this is not important. To correct this problem, the threshold similar is created to make sure that only non-negative similar is in used.

There is a new equation to generate predictions from origin equation:

$$p_{u,i} = \frac{\sum_{j \in S} s(i,j).(r_{u,j} - b_{u,i})}{\sum_{j \in S} |s(i,j)|} + b_{u,i} \quad (2.7)$$

It can be used others weight to find the prediction. In the article [19], Bell et al are proposed another way to choose weight. In details, for each user u and item i, weight value w is the solution of the equation $A.w = b$. w_j is optimal weight of user u and item j. A and b is calculated as follow equations:

$$a_{j,k} = \sum_{v \neq u} \pi_{v,j} \pi_{v,k} \quad (2.8)$$

$$b_j = \sum_{v \neq u} \pi_{v,j} \pi_{v,i} \qquad (2.9)$$

Then, the prediction is:

$$p_{u,i} = \sum_{j \in S} w_j r_{u,j} \qquad (2.10)$$

b. Computing Item similarity

As mentioned above, calling S is an item's similarity matrix, the unknown value in S is filled by zero (0 – no similarity). It is different from rating matrix [12]. We have some methods to calculate the item similarity: Cosine similarity, Conditional probability, Pearson correlation…

Cosine Similarity

Cosine similarity is the most popular in similarity metric; it is simple and fast for implementation. In addition, the result is good for accuracy. Using cosine similarity, the similarity score between two items i and j is:

$$s(i,j) = \frac{r_i.r_j}{\|r_i\|\|r_j\|} \qquad (2.11)$$

Conditional Probability

Conditional probabilities is similarity function for unary rating (such as shopping purchase histories) $s(i,j) = Pr_B[j \in B | i \in B]$ where B is purchase histories of user. It can be formulized this equation by scaling with α to balance for frequently occurring items [12]:

$$s(i,j) = \frac{Freq(i \wedge j)}{Freq(i).(Freq(j))^{\alpha}} \qquad (2.12)$$

In equation (2.12), α is damping factor to reduce the effect if j is rated by many users. For example, item j is purchased by many users, so it is similar with many items, α is in using to reduce the effect of j to similarity value.

In this case, note that $s(i,j)$ is differed from $s(j,i)$.

Pearson Correlation

Pearson correlation of item-based CF is similar with equation 2.4 in user-based CF, but it does not work well as cosine similar [19]. Therefore, it is not mention in details for this thesis.

c. Example

For practice, using again the data from Figure 5. The task is that computes the prediction of user C for movie *Equilibrium.* In this example, item-based CF with cosine similarity are used. The length of movie vector is calculated in $L_2 - Norm$.

Now, similarity between *Equilibrium* and others are computed:

$$s(Eq, Batman) = \frac{4\,.\,5 + 3\,.\,2}{\sqrt{4^2 + 5^2 + 2^2 + 3^2}\sqrt{5^2 + 3^2}} = 0.607$$

$$s(Eq, AliceIW) = \frac{4\,.\,3}{\sqrt{5^2 + 4^2 + 4^2 + 4^2}\sqrt{5^2 + 3^2}} = 0.241$$

$$s(Eq, Dumber) = \frac{3\,.\,5}{\sqrt{4^2 + 5^2 + 2^2 + 3^2}\sqrt{5^2 + 3^2}} = 0.35$$

User C has rated for three movies, but, for this example, two similar items for generating prediction are in use, so, it is *Batman Begins, Dumb and Dumber:*

$$p_{C,Eq} = \frac{s(Eq, Batman).\,r_{C,Batman} + s(Eq, Dumber).\,r_{C,Dumber}}{|s(Eq, Batman)||s(Eq, Dumber)|} =$$

$$= \frac{0.607\,.\;5 + 0.35\,.\,2}{0.607 + 0.35} = 3.903$$

Therefore, the prediction of rating between user C and *Equilibrium* is 3.903

2.1.2.2. Model based approach

Model-based method differs from memory-based method; it assumes that there is a model to generate ratings and using technique in machine learning, data mining from the training dataset to generate prediction [1]. Model-based method groups different user in training dataset into some small class by using rating patterns [20]. This approach uses machine learning and probabilistic algorithms: Bayesian networks, clustering, rule based approaches [20], neuron networks, Markov decision processes, random wall based method.... In additional, the dimension reduction technique as SVD is also used in general.

In the thesis's domain, model-based CF is not mention in details.

2.1.3. Hybrid Recommendation Systems

Hybrid Recommendation Systems are combined collaborative filtering systems and content-based filtering systems to avoid their limitations. In other words, Hybrid Recommendation Systems combine the advantage of collaborative filtering systems and content-based filtering systems. There are some methods to classify Hybrid Recommendation Systems.

In [1], Jiliang Tang et al divide Hybrid Recommendation Systems into three type:

- ***Combining different recommenders***: for this oriented, Recommendation Systems are implemented in separate content-based algorithm and collaborative filtering algorithm, and then, results are combined to generate the last recommendation.

- ***Adding content based characteristics to CF models***: as the name, in this method, the system combines the user's profile and uncommonly rated items to compute the user similarity. This approach overcomes the sparsity problem.

- ***Adding CF based characteristics to content based models***: This approach combines the dimensionality reduction technique and user profile to make a recommendation.

In [21], Burke et al group them in seven classes:

- ***Weighted Recommenders***: this system uses some recommenders and combines them to generate predictions.

- ***Switching Recommenders***: this system is combined many recommendation algorithms, and switch between them in the specify context to make the best result.

- ***Mixed Recommenders***: this approach presents the results of some Recommendation Systems together, but does not combine them in a list as Weighted Recommendations Systems.

- ***Feature-combining recommenders***: the system uses many recommendations data sources as inputs.

- ***Cascading recommenders***: this method uses the outputs of a Recommendation Systems as an input of another system.

- ***Feature-augmenting recommenders***: this system uses the output of an algorithm as one of the input features for another algorithm.

- ***Meta-level recommenders***: this system uses a model to train one algorithm, then uses this model is as input of other Recommendation Systems.

2.1.4. Evaluation Recommendation Systems

The evaluation Recommendation Systems are necessary to estimate the accuracy of algorithms. For evaluation Recommendation Systems, there are some parameters: prediction accuracy, accuracy over time, ranking accuracy … But in the thesis area, prediction accuracy is focused.

For evaluation the accuracy of Recommendation Systems, there are some measurements can use:

Mean absolute error (MAE): this method is also known as absolute deviation; it is the mean of different in absolute between each prediction and rating pair value for all cases in the test set. Equation 2.13 shows the formula to compute MAE value:

$$MAE = \frac{1}{n} \sum_{u,i} \left| p_{u,i} - r_{u,i} \right| \quad (2.13)$$

Normalized mean absolute error (NMAE): This measurement is normalized of MAE by dividing the range of possible ratings. Equation 2.14 shows the formula to compute NMAE value:

$$NMAE = \frac{1}{n(r_{high} - r_{low})} \sum_{u,i} \left| p_{u,i} - r_{u,i} \right| \qquad (2.14)$$

Root mean squared error (RMSE): this error usually uses for large errors. It is computed same as MAE. Equation 2.15 shows the formula to compute RMSE value:

$$RMSE = \sqrt{\frac{1}{n} \sum_{u,i} (p_{u,i} - r_{u,i})^2} \quad (2.15)$$

2.1.5. Some problem in Recommendation Systems

Recommendation Systems have many challenges and problems such as Cold-start problem, Scalability of the approach, Recommending the items in the Long tail, Accuracy of the prediction, Novelty and diversity of recommendation Sparse ,Missing, Erroneous and Malicious data, Conflict resolution while using ensemble/ hybrid approaches, Ranking of the recommendations, Impact of context-awareness, Impact of mobility and pervasiveness, Big-data, Privacy concerns [22] …. In the thesis area, some popular problem in Recommendation Systems are mentioned: Cold-start problem, data sparsity problem, attack problem, privacy concerns, and explanation problem.

2.1.5.1. Cold-start problem

The cold-start problem usually happens on to Collaborative Filtering Systems that users or items information are missing to induce obstacle to Recommendation Systems. In other words, Recommendation Systems do not have information about the user or item to generate recommendations. It takes two flavors:

- ***Item cold-start***: it happens when a new item has been added to the database, Recommendation Systems are not enough rating to prediction.

- ***User cold-start***: it happens when a new user has joined, the system is not had history information of users.

Cold-start is very popular because Recommendation System not only need data, but it also needs high quality data. When item cold-start and user cold-start are happened in concurrence, it is called ***bootstrap problem*** [23].

2.1.5.2. Data sparsity problem

Similar to cold-start problems, data sparsity is usually in the Collaborative Filtering Systems [22]. It is phenomenal that the ratings of users to items is limited. Different from cold-start problems, data sparsity is the system's problem.

2.1.5.3. Attacks problem

Hackers with other aims also can attack recommendation Systems. For example, the attackers can make the virtual rating for items. The consequence is that users receive imprecise recommendation.

2.1.5.4. Privacy concerns

In order to generate good recommendation, Recommendation Systems need more details information from the user's profile, but many users do not approve this. This is a challenge.

There are many users approve to provide privacy information, and how Recommendation Systems protect their information are very important.

2.1.5.5. Explanation problem

Recommendation Systems provide advises but they do not explain for this recommendation. For example, when some items are bought in together by many users, Recommendation Systems will recommend them in together for recommendation.

2.2. Social Recommendation

2.2.1. Social media and Social theories.

2.2.1.1. Social media

Social media is computer-mediated tools that allow people to create, share or exchange information, and pictures, videos in virtual communities and networks in everywhere and every time. It is based on Web 2.0 foundation. The Figure 2.5 shows some example about social media services as facebook.com, youtube.com, and twitter.com ... Social media data significantly differ from the traditional data. It is big, noisy, incomplete, unstructured and linked with social relations [4].

Figure 2.5: Some famous social media services

To explain and understand deeply about social media data and social phenomenon, considering about social theories. There are three social theories: Social Correlation theory, Balance theory, Status theory as Figure 2.6.

2.2.1.2. Social Theories

Social theories from social sciences are useful to explain various types of social phenomena. Social theories also use to predict tie strength with Social Media. The research in [24] shows that the role and properties of social relations are different. For example, when we are sick, our family and close friend (has the higher degree of relations) always care us more than others. In social media, it is increasingly possible for us to observe social data from hundreds of millions of individuals [4]. The thesis is concentrated on three important theories.

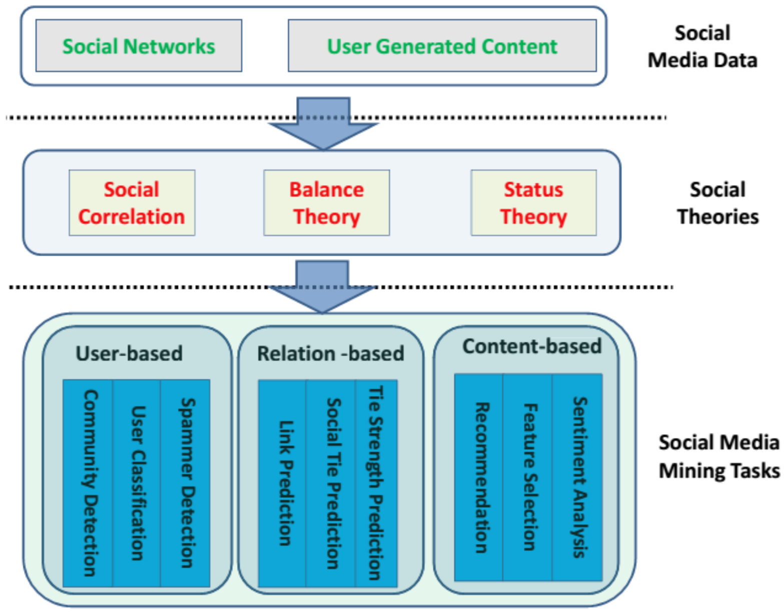

Figure 2.6: Social theories in Social Media Mining

Before going to details, considering some notations:

- $U = \{u_1, u_2, \dots, u_n\}$ is the set of n users
- $I = \{i_1, i_2, \dots, i_m\}$ is the set of m items
- $S \in \boldsymbol{R}^{n*n}$ is the relation of user and user
- $R \in \boldsymbol{R}^{n*m}$ is the interaction of user and content
- $C \in \boldsymbol{R}^{m*K}$ is content – feature matrix where K is the number of feature extracted from content set

2.2.1.2.1. Social Correlation Theories

As the name, Social Correlation theory shows the correlation between behaviors or attributes of adjacent users in a social network. It includes three major social process: homophily, influence and confounding.

- **Homophily** shows the tendency from users of others that share something similar. For example, people have the same interest that often takes part in the same group…
- **Influence** says that the people tend to follow the behaviors of their friends or people around them. For example, if the most of one's friend is laborious, he could be influenced of them.
- **Confounding** is the correlation between users under the influence of environment, for example: two people in the same schools are more likely become friend than two people in difference schools.

The Figure 2.7 show more clearly about three major social forces:

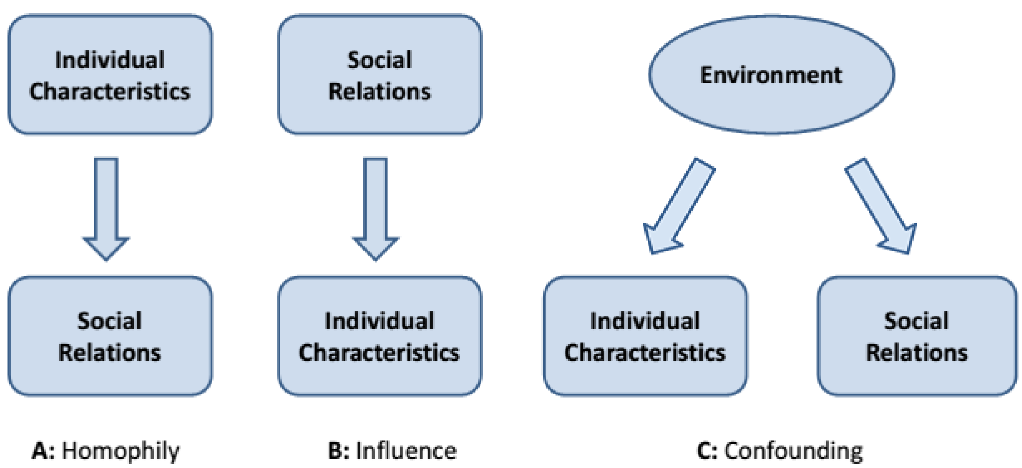

Figure 2.7: Major social forces of Social Correlation theory

In order to consider functional of social correlation to social media data, which means, there is needed to answer "*are users with social relations more similar than these without?*" [4] [25]. For interpretation, this question, for each relation from u_i to u_j, call that:

- s_{ij} is the similar between u_i and u_j
- r_{ik} is the similar between u_i and a random chosen user u_k
- S is the set of similarities of pair of connected users
- R is the set of pair of randomly chosen users

If $S \geq R$, users with social relation that are more similar than without. In fact, there are many researches about this problem, the real experiment in Twitter users, Epinions users, Digg users, Blog-Category users, foursquare users to prove this problem.

2.2.1.2.2. Balance Theories

A The Balance Theory is proposed by Fritz Heider. It illustrates how people develop their relationships with other or something in their environment. The ideal of this theory is: users see a system or something is in balance, if it is out of balance, then users are motivated to restore a position of balance. Based on this ideal, in 1958, Fritz Heider states the balance theory as four clauses [26]:

- My friend's friend is my friend
- My friend's enemy is my enemy
- My enemy's friend is my enemy
- My enemy's enemy is my friend

To formalize this theory, denote that s_{ij} a sign of the relation of two user u_i and u_j, $s_{ij} = 1$ if observed positive relation of two user, $s_{ij} = -1$ if observed negative relation of two user. Considering triad users $< u_i, u_j, u_k >$ is balanced if:

- $s_{ij} = 1$ and $s_{jk} = 1$, then $s_{ik} = 1$; or
- $s_{ij} = -1$ and $s_{jk} = -1$, then $s_{ik} = 1$;

The following figure show the illustration of Balance Theory:

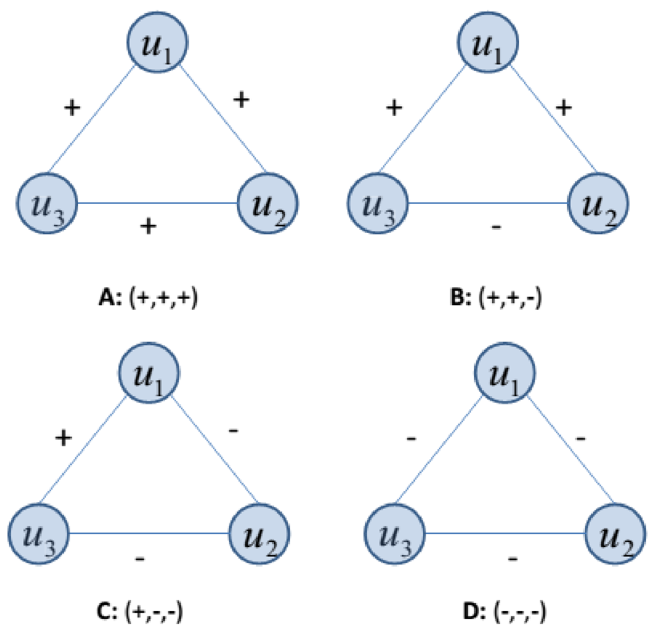

Figure 2.8: An Illustration of Balance Theory

The Figure 2.8 shows four possible combination $A(+,+,+)$, $B(+,+,-)$, $C(+,-,-)$ and $D(-,-,-)$. In the four this combination, there are only $A(+,+,+)$ and $C(+,-,-)$ allowing Balance theory. It is clear that the Balance theory is developed for undirected social networks, the direction is ignored when using this theory for directed social networks [4].

Balance theory is universal in the real world and social networks, it usually uses to predict the relation of users (friend or not) in social networks.

2.2.1.2.3. Status Theories

Status Theory is different from Balance Theory; it can be apply to directed social networks. Status theory refers to the degree or rank of users in social networks [25]. Status Theory defines that if there is link from user u_i to user u_j then user u_i is higher status than user u_j. Considering the triad user $< u_i, u_j, u_k >$ that u_i links to u_j, $u_k$links to u_i, u_j. There are total 16 cases of sign of relations. Considering 4 over 16 cases as this Figure 2.9:

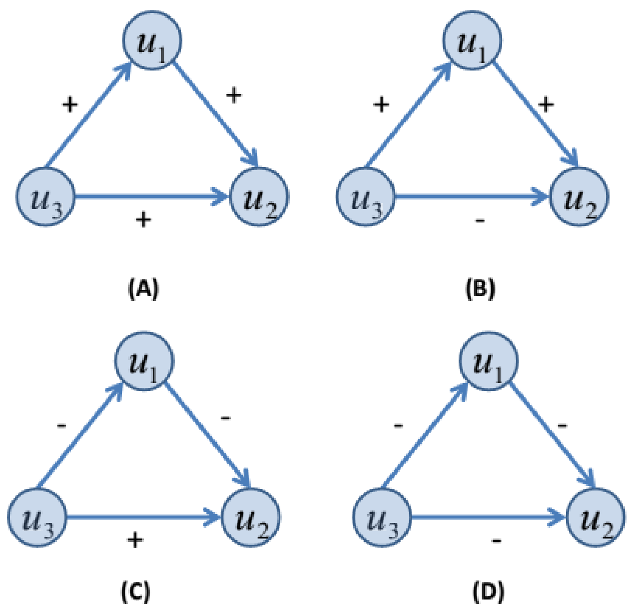

Figure 2.9: An illustration for four out of sixteen type of contextualized links for Status Theory

Considering Figure 2.9, there are only case A, and case D is allowing Status Theory. In case A, status of u_3 is larger than of u_1, status of $u_1$is larger than of u_2, so status of u_3 is larger than u_2. Similar to case D. Consider case B, status of u_3 is larger than ofvu_1, status of $u_1$is larger than of $u_2$but the status of u_3 is smaller than u_2. It is a contradiction, than similar to case C.

The development of social media encourages the development of data, information. It brings many opportunities for developing Recommendation Systems. The next section presents about Social Recommendation.

2.2.2. Social Recommendation

Social Recommendation has first introduced in 1997 by Kautz et al [1] [27]. There are many researches in Social Recommendation on later but there is no common definition for Social Recommendation. In article [1], Jiliang Tang et al were given definition of Social Recommendation in narrow and broad.

In the narrow meaning, Social Recommendation is a traditional Recommendation System that using social relation as additional input. Social relation can be *trust, friendship, reputation, following, follower* [1] … In this definition, Social Recommendation Systems assume that users are correlated if they have social relation [1]. Which means, users can be effected from their friend in decision-making.

In the broad definition, Social Recommendation is any Recommendation System in the social media domains. In this definition, Social Recommendation focus on items, tag, people, communities, behaviors [1].

Because of the limitation of thesis, the thesis will consider only narrow definition of Social Recommendation systems.

2.2.2.1. Special feature of Social Recommendation

Social Recommendation systems have some difference from traditional Recommendation Systems. Firstly, users in traditional Recommendation Systems are assumed that independence, but in the Social Recommendation, users are linked by social relations (trust, friendship…). Social relations and social network construct induce influence to user's correlation. It is clear that users in connecting are correlated rather than users in independence (homophily in social theory). In other words, beside the ratings matrix, Social Recommendation systems also need a social networks structure and social relations matrix.

Considering the example (this example is taken from [1]): There are five users connected and five items as Figure 2.10. Figure 2.11 illustrates the rating matrix when using traditional Recommendation Systems. Figure 2.12 demonstrates two data representations:

rating matrix and social relation matrix. In the Social relation matrix, the cell of pair $< u_i, u_j >= 1$ mean u_i, u_j are connected, the cell of pair $< u_i, u_j >= 0$ for otherwise.

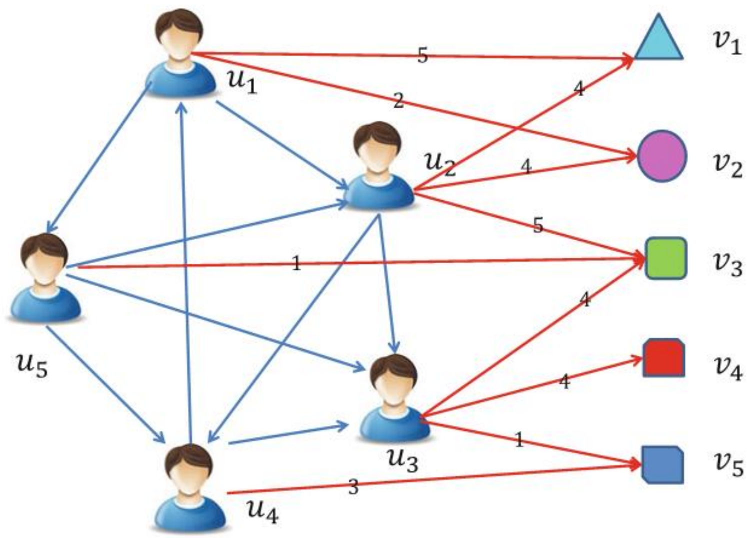

Figure 2.10: Connected user

	v_1	v_2	v_3	v_4	v_5
u_1	5	?	2	?	?
u_2	4	4	5	?	?
u_3	?	?	4	4	1
u_4	?	?	?	?	3
u_5	?	?	1	?	?

Figure 2.11: Using Traditional Recommendation Systems

	v_1	v_2	v_3	v_4	v_5
u_1	5	?	2	?	?
u_2	4	4	5	?	?
u_3	?	?	4	4	1
u_4	?	?	?	?	3
u_5	?	?	1	?	?

	u_1	u_2	u_3	u_4	u_5
u_1	0	1	0	0	1
u_2	0	0	1	1	0
u_3	0	0	0	0	0
u_4	1	0	1	0	0
u_5	0	1	1	1	0

Figure 2.12: Using Social Recommendation Systems.

Secondly, Social Recommendation is not only developing about recommendation, but it is also involves the other field research *Social Network Analysis* (SNA) [1]. By Jiliang Tang et al in [1], Social Network Analysis is the methodical analysis of social networks has emerged as a key technique in modern sociology. It is concerned with many fields: economics, geography, communication and so on. Social Recommendation systems use the knowledge of their field such as Social Theory (Social Correlation), Ties prediction, community detection… to improve the performance of recommender results.

2.2.2.2. Social Recommendation systems

In this section, the thesis is concentrated in Collaborative Filtering for Social Recommendation systems. Social Recommendation is required two inputs: rating matrix information and social relation information. Similar to traditional Recommendation Systems, in Social Recommendation, Collaborative Filtering also divide into two classes: memory based and model based. For convenience, define $T \in \mathbb{R}^{n \times n}$ is users relation where $T(i,j) = 1$ if and only if user u_i connect to u_i.

2.2.2.2.1. Memories based Social Recommendation systems

Similar to traditional approaches, Memory based social recommender systems also include two steps:

- Find the neighbors set

- Prediction the missing rating value

But the difference in social approach is in step one. In social approach, the neighborhood set is computed by aggregating from the rating information and social information (from correlated user). Denote the set of neighborhood of user in this orient is N^+. Now, considering some method to compute set N^+.

a. Social based Weight Mean

This method is proposed by Victor et al in [28] [29]. They define N^+ in equation 2.16:

$$N^+(i) = \{u_j | T(i,j) = 1\} \qquad (2.16)$$

b. Trust Walker

This method is proposed by Jamali and Ester in [30]. They define the similar between items i, j in quation 2.17:

$$sim(i,j) = \frac{1}{1+e^{-\frac{N_{ij}}{2}}} \times PCC(i,j) \qquad (2.17)$$

Where N_{ij} is the number of user that have both rate for two items i, j. $PCC(i,j)$ is Pearson Correlation Coefficient of two items i, j.

There are some other model but it is not considered in thesis as TidalTrust of Golbeck et al in 2006 [31], MoleTrust of Massa and Avesani in 1994 [32].

2.2.2.2.2. Model based Social Recommendation systems

In the model based Social Recommendation systems, the most popular method is Matrix factorization. For social approach of Matrix factorization technique, Jliang Tang et al [1] add a term to weight to the traditional matrix factorization technique for social relation as follows equation (Equation 2.18):

$$min_{U,V,\Omega} \|W \odot (R - U^T V)\|_F^2 + \lambda(\|U\|_F^2 + \|V\|_F^2 + \|\Omega\|_F^2) + \alpha Social(T, S, \Omega) \quad (2.18)$$

In equation 2.18, first and second term are modelled in the traditional matrix factorization, it is not mentioned in here. The third term $\alpha Social(T, S, \Omega)$ is focused - introduced to capture social information. Ω is the set of parameters learned from social information. α controls the contributions from social information. W controls the weights of known ratings in the learning process [1].

2.3. Social Tie Theories

2.3.1. Introduction

Social ties (also known as interpersonal ties), was first introduced by Granovetter in [7], are the connection that carries information between people in social networks. In other words, social ties defines the characteristic between two nodes. It can classify in some type: *strong ties*, *weak ties*, and *absent ties*. Each type of social ties has its role in the process of information transmission and propagation in the social network.

Strong ties are the connections of users in closed relationships. Which means, strong ties refer to people (or users) in the family, trusted friends, close friends relations. Those users usually share the same things or have similar preferences in some fields or have tended to similar behavior. When investigating about the role of social ties to Recommendation Systems, Oliver Oechslein et al [5] indicate that strong ties are more influential at an individual level. Koroleva and Štimac [8] show that the users have tended to receive advises from their friend in closed (users with strong ties).

Weak ties are the connections of people that are acquainted. Researches in [24] [33] illustrates that acquaintances usually provide fresh information and often present heterogeneous behaviors. In [7], Granovetter show that weak ties also play very important role in propagation new content information. In [34], Steffes and Burgee also show that weak ties have more affect to individual decision making.

In the social network, strong ties usually connect between users in community and weak ties use to link these community. In [35], David Easley and Jon Kleinberg show that if one node with two neighbors have strong ties, there is at least one weak tie between these neighbors. The below figure (Figure 2.13) shows an example:

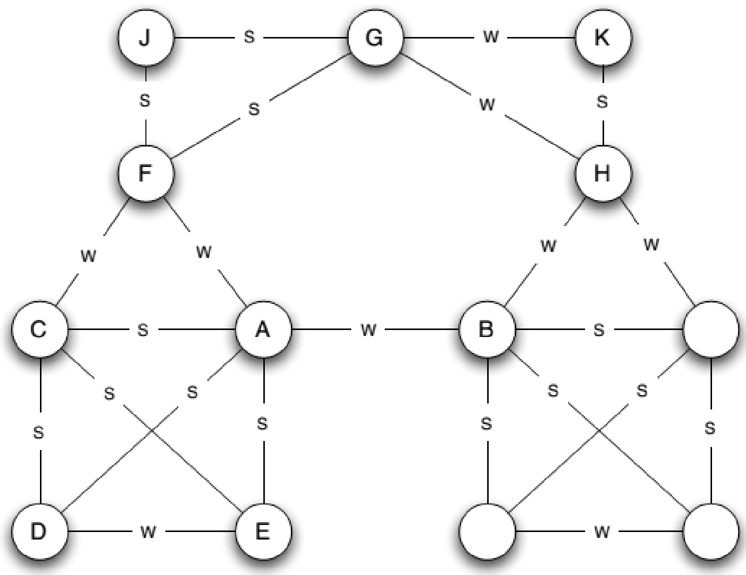

Figure 2.13: An example about weak ties and strong ties.

By Granovetter in [7], absent ties are those relationships (or ties) without substantial significance. For example, the people are living in the same street, but they are very litter interacted with others. These exists do not affect to others in network, and they can be ignored.

Next, Social ties strength is mentioned to known that how social ties are weak or strong.

2.3.2. Social Tie Strength

Social tie strength (or interpersonal ties strength) is a probably linear combination of the amount of time, intimacy, emotional intensity and reciprocal services that are characteristic the tie [5] [24]. From this definition, tie strength has four dimensions: *amount of time, intimacy, emotional intensity and reciprocal services*. Those later researches are expanding the list of dimensions of tie strength and there are at least seven dimensions: *Intensity, Intimacy, Duration, Reciprocal Services, Structural, Emotional Support and Social Distance*.

Intensity

For intensity, there are some manifest: Inbox messages exchanged, mail exchanged, Friend's status updates, Friend's photo comments …

Intimacy

Intimacy variable includes *Days since last communication*, *Appearances together in photo*, *Distance between hometowns, Friend's relationship* status.

Duration

Duration variable usually are in user is *Days since fist communication.*

Reciprocal Services

For reciprocal services, *Applications in common* is usually in use.

Structural

Structural variable includes *Number of mutual friends, Groups in common, Norm. TF-IDF of interests and about*

Emotional Support

Emotional Support includes *Wall and inbox positive emotion words, Wall and inbox negative emotion words*

Social Distance

Social distance variable includes *Number of occupations difference*, *Educational difference (degrees), Overlapping words in religion*

These above are seven dimensions and their manifest that I have found. There are many dimensions and manifest but seven parameter above are popular.

2.4. Summary

In this chapter, the thesis has shown the knowledge about types of traditional Recommendation Systems: content based filtering, collaborative filtering (memory based and model based), hybrid Recommendation Systems, social media, social theories, social ties, and social ties strength.

For the next chapter, the thesis will present about the role of social ties strength in Recommendation Systems and the model of combination social ties into traditional Recommendation Systems (Social Recommendation).

THE METHOD

3.1. The role of Social Tie Strength

In the chapter 2, the thesis has presented about the Recommendation Systems and Social Tie. In this section of chapter 3, the thesis is going to illustrate the role of tie strength in Recommendation Systems.

Researches in [5] [8] show that strong ties play important role in the recommendation. These users (or people) in strong tie relationships usually have high trust with others in this relation. Since, exploiting strong tie of users will help in making coherent recommendations. Koroleva and Štimac [8] indicate that Facebook users have tended to get information from their strong ties. In [9], Li et al reveal that recommendation accuracy reflects personalization quality. In [33], Marsden et al show that the accuracy of Recommendation Systems in strong ties rather than weak ties, but in [34], Steffes et al note that weak ties have more influence than strong ties to decision making procedure. Therefore, whether strong tie or weak tie, they have a positive influence on quality of Recommendation Systems. Ties strength also has a positive influence on the recommendation's credibility and accuracy [5]. Since above reasons, Social Tie Strength can be used to improve Recommendation Systems.

3.2. A model to indicate the effect of Social Tie strength to Recommendation Systems

3.2.1. General Idea

The objective of this study has investigated the positive influence of social ties to the Recommendation Systems. Specially, in this thesis, two types of social ties are researched: Intimacy Variables (*Number of appearances together in photo*), Structure variable (*Number of friends in commons* or *Number of mutual friends*) to Recommendation Systems from Facebook users.

Why the thesis chooses Facebook users for this research. The main reason is that Facebook is large and famous social network, the number of Facebook users are very huge. One extra reason is the convenience of collecting data.

Why the thesis chooses *Number of appearances together in photo* and *Number of friends in commons*. First reason, in the Facebook, photo is very useful to share the emotion and the highlight of the user's activities. Almost Facebook users will use photos to save memorable moments. Since, *Number of appearances together in photo* as known as a highlight dimension of Facebook social tie strength and use it to research. *Number of friends in commons* is also remarkable dimension of social tie in Facebook. It is clear that if two users that have many friends in common are having strong tie rather than two users that have fewer mutual friends. The second reason, *Number of appearances together in photo* and *Number of friends in commons* are two in the litter of social tie dimensions of Facebook that in free. The last reason, in our understanding of improving Social Recommendation Systems, there is no research about the effect of *Number of appearances together in photo* and *Number of friends in commons* of Facebook to Social Recommendation Systems. From above reasons, above two types can be chosen to study.

The thesis chooses research in movie recommendation field because it is very popular (MovieLens, Netflix, IMDB …) and convenient for getting data.

In order to study tie strength, a survey of 80 users in social networks Facebook are made, That Facebook users are called candidates. These candidates will rate for 99 movies from 2005 to 2014 in 5-scale (very bad, bad, normal, good, very good). Since, a list of users (candidates), list of items (movies), rating data (candidates to movies) are obtained. Each candidate is requested to choose three people (trusted sources) in candidates list that he or she wants to receive advice on choosing moves. Then, candidates are asked about *Appearances together in photo, Number of friend in commons* with these sources. After that, these information are checked by using the URL request to Facebook: https://www.facebook.com/ + “userID” + “?and=” + “sourceID” to request again this information.

In order to highlight the effect of Tie Strength to Recommendation Systems, two algorithm: collaborative filtering and collaborative filtering combined with tie strength (*Appearances together in photo, Number of friend in commons)* information are implemented. Then results are evaluated by using MAE measurement and comparison.

All above ideas are formalized in a model in the next section.

3.2.2. A model to indicate the effect of Tie strength to Recommendation Systems.

A model is proposed to compare efficiency between two algorithms: traditional CF and Combination of CF and Social Tie. This model can divide in four phrases:

- Data preprocessing.
- Collaborative filtering systems.
- Social Collaborative filtering systems (Combination: CF, Tie strength).
- Evaluation.

The Figure 3.1 shows the model in details:

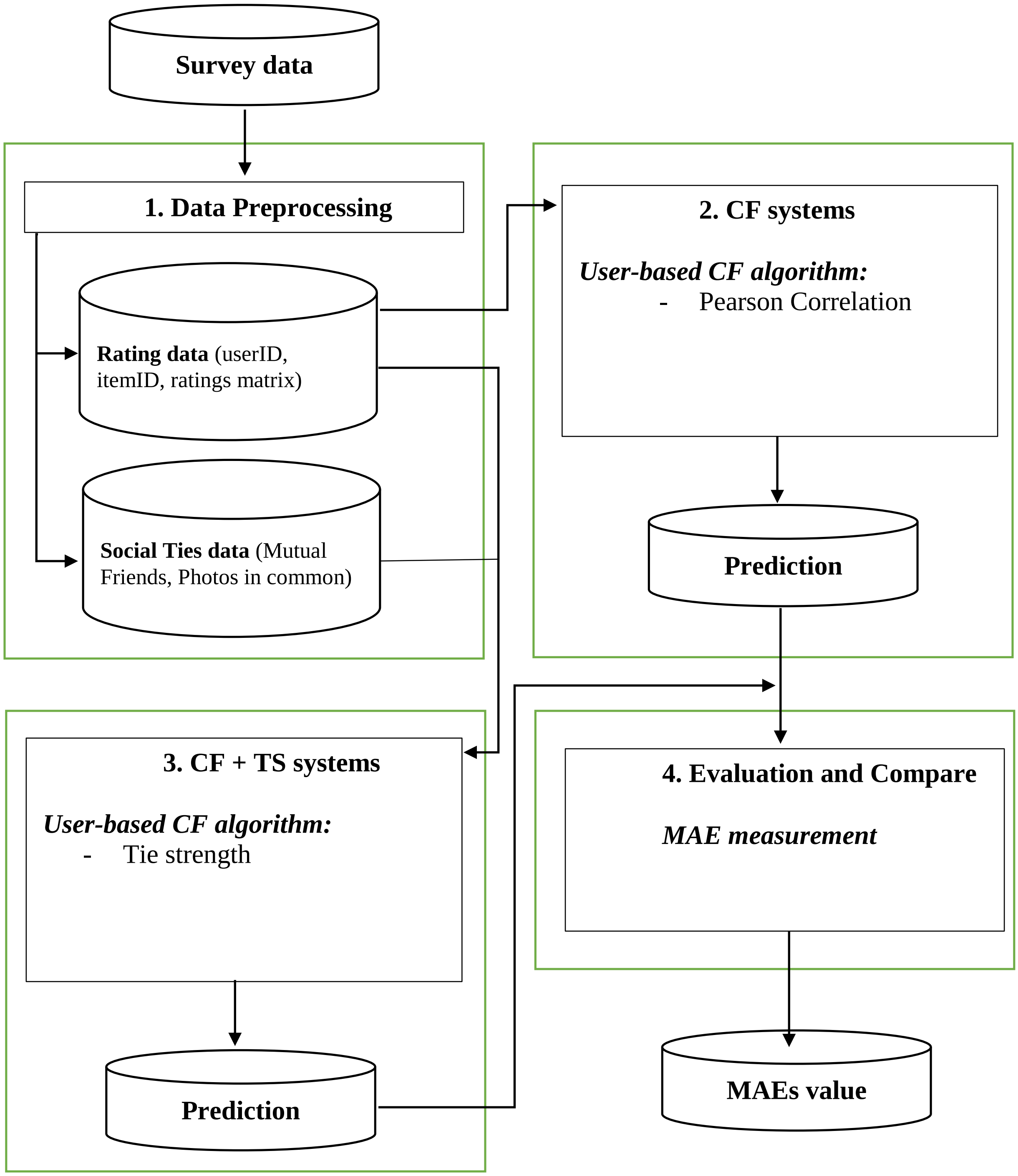

Figure 3.1: A model to evaluate the role of Tie strength to Recommendation Systems.

3.2.2.1. Data preprocessing.

Purpose: Raw data preprocessing.

Input: *Survey data.*

Output: *Ratings data* (users ID, items ID, ratings matrix) and *Social data* (Number of mutual friends between users and their source, number of photos in common).

Method:

From input is survey data, two tables are obtained, the first table is rating of users to movies, called **rating table**, the second table is the three source choosing, called **source table**.

For this phrase, firstly, first table is analyzed into list of users, list of items, and rating matrix as *rating data*. For second table about three source, An URL request of Facebook is used to obtain Mutual Friend and Photos in common of users and their source as *Social data*.

Steps:

1. Analysis **rating table** into *rating data* by hand and saves to file.
2. Analysis **source table**:
    a. Read current user ID and source ID from **source table**
    b. Request URL: https://facebook.com/ + “currentUserID” + “?and=” + “sourceID” to browser
    c. Read number of mutual friends and number of photos in common of current user and his sources from browser.
    d. Saving these data to file, this is *social data*.

3.2.2.2. Collaborative filtering systems

Purpose: Implement Collaborative filtering based on user-based approach with Pearson Correlation.

Input: *Rating data* (users ID, items ID, ratings matrix).

Output: *Predictions matrix.*

Method:

The second phrase, user-based collaborative filtering systems with Pearson correlation is in use to implement CF system. User-based CF system and Pearson correlation are illustrated in chapter 2.

Steps:

1. Read all user ID from *rating data* file into array.
2. Loop each user ID in array
    a. Calculate the Pearson correlation of current user to all user remaining by equation 2.4
    b. Find three maximum value of Pearson correlation to obtain neighbor set
    c. Calculate the prediction using equation 2.2.
3. Save all prediction value into a *Predictions matrix.*

3.2.2.3. Collaborative filtering combine Tie strength

Purpose: Implement Collaborative filtering using Tie Strength as weight instead of Pearson Correlation.

Input: *Rating data* (users ID, items ID, ratings matrix), Social data (Number of mutual friends between users and their source, number of photos in common).

Output: *Predictions matrix.*

Method:

For this phrase, tie strength is applied to collaborative filtering. The equation that represent by Ofer Arazy et al in [3] is in use to generate prediction as equation 3.1:

$$p_{u,i} = \bar{r}_u + \frac{\sum_{u' \in N} s(u,u')(r_{u',i} - \bar{r}_{u'})}{\sum_{u' \in N} |s(u,u')|} \quad (3.1)$$

Where: $s(u, u')$ is tie strength between two users u , u'.

Steps:

1. Read all user ID from *rating data.*
2. Loop for each user ID
    a. Read three source of this user ID from *social data*
    b. Calculate the prediction by using equation 3.1
3. Save all prediction into *Prediction matrix.*

3.2.2.4. Evaluation

Purpose: Comparison the prediction results from algorithms in phrase 2 and phrase 3.

Input: *Predictions matrix* (from phrase 2 and 3).

Output: *MAE value* (using to evaluate).

Method:

Using MAE measurement to calculate the MAE value.

Steps:

1. Read two *Prediction matrix* from phrase 2 and 3.

2. Calculate *MAE value* by using equation 2.13.

3.2.3. Summary

For this chapter, the effect of Tie Strength to Recommendation Systems is presented and a model to evaluate the effect of Tie strength to Recommendation Systems is introduced.

For next chapter, the thesis will represent the result when implement from chapter 3.

EXPERIMENTS AND DISCUSSIONS

4.1. Overview

For this experiment, the method of Arazy O et al in the article [6] is used. In this article, Arazy O et al implement algorithms of Recommendation Systems: uses traditional CF and CF combined with social relation, social ties strength. After implementation, they make a comparison to see the effect of CF when combining to social relations. In thesis, this method is used but parameters are changed for suitable with thesis. In detail, social tie strength is used in two dimension: *Appearances together in photo (or Number of photos in common), Number of friends in commons (or Number mutual friends)* to combine with traditional Recommendation Systems.

The aim of the experiment is competition two algorithms: traditional collaborative filtering and collaborative filtering combined with mutual friends and photos in common to highlight the positive effect of Social Tie Strength to Recommendation Systems.

The model in the chapter 3 is implemented for the experiment. As mentioned in chapter 3, six modules are constructed:

- **com.data**: uses to *rating data* process as: file process, format input…
- **com**.**TSprocess**: uses to process data for *social ties data* as: file process, format input…

- **com.similarity**: to calculate similarity of users.
- **com.prediction**: uses to compute the prediction, it include algorithm in phrase 2 and 3.
- **com.evaluation**: implements the MAE measurement.
- **com.math**: implements some basic math function as average calculation …

4.2. Tools in use

The Table 4.1 and Table 4.2 show the configuration of hardware and list of software in use:

Configuration of hardware:

Hardware component	Information
Processor	Intel(R) core(TM) i3-2350M, CPU 2.30GHz
RAM	4GB
Operating System	Windows 8 64bit
Hard Disk Drive	500GB

Table 4.1: Systems configuration information

List of software:

Index	Software	Author	Source
1	Eclipse IDE for Java Developers ,Version: Luna Release (4.4.0)		https://www.eclipse.org
2	Commons-math library version 3.5 release	Open source software	http://commons.apache.org/proper/commons-math
3	Microsoft Excel 2013	Microsoft	https://store.office.com

Table 4.2: List of tools in use

4.3. Data

In order to study tie strength, a survey of 80 users in social networks Facebook is made, that users as known as candidates. These candidates have rated for 99 movies from 2005 to 2014 in 5-scale (very bad, bad, normal, good, very good). Since, a list of users (candidates), a list of items (movies), rating data (candidates to movies) are collected. Each candidate is request to choose three people (trusted sources) in the candidates list that he or she wants to receive advice on choosing moves. Then, candidates are asking about *Appearances together in photo, Number of friend in commons* with these sources. After that we check it by using the URL request from Facebook: https://www.facebook.com/ + "user ID" + "?and=" + "source ID". 80 candidates almost us friend in Facebook, so, it can believe that the rating value is trusted. The component of candidate are various as Table 4.3:

	University friends	High School friends	Family	Unknown	Others friends	Total
Number	43	22	3	8	4	80
Percent(%)	53.75	27.5	3.75	10	5	100

Table 4.3: The component of candidates.

For data collection, data is completed from 19/4/2015 to 1/5/2015. Because of difficulties in collecting the social ties data, there are also 80 users that complete the survey, but, in some research about social ties as [5] [24] [6], these author also get data from surveys, and the number of participants is not much. For example, in [5], Oliver Oechslein et al used 193 participants, in [24], Eric Gilbert and Karrie Karahalios used 35 participants, in [6], Ofer Arazy et al used 99 participants.

Figure 4.1, Figure 4.2 and Figure 4.3 bellows illustrate data that are proceeded:

1. Interstellar (2014)
2. Guardians of the Galaxy (2014)
3. X-Men: Days of Future Past (2014)
4. Gone Girl (2014)
5. Captain America: The Winter Soldier (2014)
6. The Grand Budapest Hotel (2014)
7. Edge of Tomorrow (2014)
8. The Amazing Spider-Man 2 (2014)
9. Divergent (2014)
10. Lucy (I) (2014)
11. Birdman: Or (The Unexpected Virtue of Ignorance)
12. Fury (2014)
13. Whiplash (2014)
14. 300: Rise of an Empire
15. How to Train Your Dragon 2 (2014)
16. Need for Speed (2014)
17. Kingsman: The Secret Service (2014)
18. The Wolf of Wall Street (2013)
19. Hansel & Gretel: Witch Hunters (2013)
20. After Earth (2013)
21. Oz the Great and Powerful (2013)
22. Kick-Ass 2 (2013)
23. Monsters University (2013)
24. Despicable Me 2 (2013)
25. Fast & Furious 6 (2013)

Figure 4.1: Example about items list.

45 Nguyễn Minh Thuận
46 Vượng
47 Jun
48 Lã Thị Thu hiền
49 Nguyễn Bảo Khoa
50 Lê Đức Minh
51 Vũ Thị Thanh Hải
52 tạ minh thăng
53 zoro
54 Hai
55 Mai Huu Phuc
56 Bùi Mạnh Chiến
57 mai huy quý
58 Vũ Thu Hiền
59 Nguyễn Thị Hương
60 Trung
61 Đào Thị Vân Mây
62 Trần Thị Mỹ Linh
63 Huong Nv
64 Minh Tuấn
65 Nguyễn thanh nguyệt
66 fabio
67 Le thi huyen trang
68 Kẹo Ngọt
69 Mai Văn Hoan
70 Ngô Khắc Hoàng

Figure 4.2: Example about users list

Họ và tên	Facebook	1. Interstellar (2014)	2. Guardians of the Galax	3. X-Men: Days of Future	4. Gone Girl (2014)	5.
Trần Đức Mười	https://www.facebook.con	3	4	4	4	
Nam Duong	https://www.facebook.con	5	5	5	3	
Nguyen	https://www.facebook.con	5	1	4	4	
AliaNgo	https://www.facebook.com/galaxies.03091992			4	4	
Mai Công Đạt	https://www.facebook.com/lenhhoxung.dnth		3	4		
mai thùy dung	https://www.facebook.con	5	5	5	3	
Quỳnh	https://www.facebook.com/quynh.mai.39			4		
Mai Kim Tài	https://www.facebook.con	5	4	4	3	
Chu Chí Quang	https://www.facebook.con	4	5	5	3	
Chu Văn Tạo	https://www.facebook.com/tao.chu.752					
giangpth	https://www.facebook.com/giangphamtranhuong					
Vương Thị Hồng	https://www.facebook.con	3	3	5	3	
Nguyễn Trường Thịnh	https://www.facebook.con	5		5		
Hoàng Thu Thủy	https://www.facebook.con	3	1	3	4	
tranh	https://www.facebook.con	5	3	3		
Miên Nguyễn	https://www.facebook.con	3	3	4	2	

Figure 4.3: Example about the rating matrix collected from survey.

4.4. Result and Discussion

In the experiment, data are divided and compared by using 10-fold method, each fold are generated at random. In which, 80% data is used for training and 20% data for test.

The Table 4.4 below shows the MAE measurement for each fold, notate that low MAE value is better than high MAE value.

	Fold 1	**Fold 2**	**Fold 3**	**Fold 4**	**Fold 5**	**Fold 6**	**Fold 7**	**Fold 8**	**Fold 9**	**Fold 10**
CF	1.4612	1.4448	1.4346	1.4407	1.3807	1.4151	1.3912	1.3565	1.4117	1.4490
CF + Mutual Friend	1.4109	1.4178	1.392	1.3719	1.3119	1.3164	1.3474	1.2837	1.3392	1.3629
CF + Photo in common	1.4119	1.4185	1.3949	1.3735	1.314	1.3182	1.3495	1.2862	1.3418	1.3646

Table 4.4: The MAE value of CF method and CF + tie strength method.

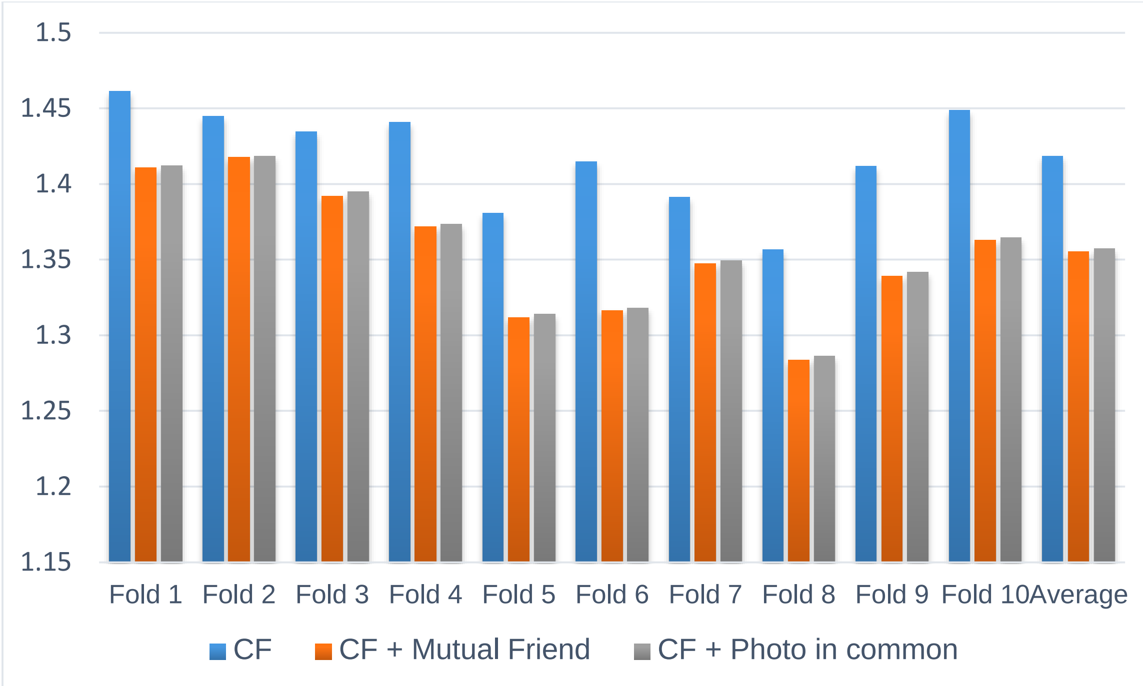

Figure 4.4: MAE value over 10 fold in graph.

From the results of the experiment, in ten folds, it is clearly that the results are positive. In all fold, the method CF + Mutual friend always gives best results, CF + Photo in common give results that approximate CF + Mutual Friend. Both methods are better results than traditional CF method. To have a clearly view, considering Figure 4.4, in fold 1, fold 2 and fold 3, MAE values of CF + Mutual friend and CF + photo in common are approximate and smaller than CF method, but distances of CF line to two others are not large. In seven remained folds, these distances become larger. Noticeably, in fold 6, the MAE of CF method (1.4151) is clearly larger than two method remaining (1.3182 and 1.3164). MAE value of fold 8 is the best. The MAE values of fold 1, fold 2, and fold 10 are quite high. And it can be seen that the MAE of CF + Mutual friends is the smallest in ten folds, which means, mutual friend factor is slightly better than photo in common factor for Recommendation Systems in this data.

CONCLUSIONS

5.1. Conclusions

In conclusion, this thesis shows the influence of Social Ties in Recommendation Systems. To do this task, firstly, traditional Recommendation Systems and Social Recommendation and these algorithms (chapter 2) were introduced. Next, the effect of social media to users in Social Recommendation through social theories was investigated. Secondly, the dimension of Social Ties (chapter 2) and how they can affect to Recommendation Systems (chapter 3) were studied. Finally, Collaborative filtering algorithm with Pearson correlation and Collaborative filtering combined with Social Ties were implemented to compare the result, and the results are positive to show that the effect of Social Ties to Recommendation Systems (chapter 3 and chapter 4).

In order to complete implementation, a survey to collect data (has social ties strength factor) from Facebook users was completed. This work is taken from us much effort. However, it also weak point of the thesis because of the limitation of on the number of users take part in the survey.

5.2. Future Works

In the future, firstly, other dimensions of Social Ties to Recommendation Systems will be expanding in research such as:

- *Duration* variable : Days since from first communication

- *Intimacy* variable: Days since from last communication, Inbox intimacy words …

- *Emotional Support* Variables: Wall & inbox positive emotion words

- *Predictive Intensity* Variables: Wall words exchanged, Inbox messages exchanged …

Secondly, more data will be collecting in order to make data more objective than in the thesis.

REFERENCES

[1] Tang, Jiliang and Hu, Xia and Liu, Huan, "Social recommendation: a review," *Social Network Analysis and Mining,* vol. 3, pp. 1113-1133, 2013.

[2] Tang, Jiliang and Tang, Jie and Liu, Huan, "Recommendation in social media: recent advances and new frontiers," in *Proceedings of the 20th ACM SIGKDD international conference on Knowledge discovery and data mining*, ACM, 2014, pp. 1977-1977.

[3] Arazy, Ofer and Kumar, Nanda and Shapira, Bracha, "Improving Social Recommender Systems," *IT professional,* vol. 11, no. 4, pp. 38-44, 2009.

[4] Tang, Jiliang and Chang, Yi and Liu, Huan, "Mining social media with social theories: A survey," *ACM SIGKDD Explorations Newsletter,* vol. 15, pp. 20-29, 2014.

[5] Oechslein, Oliver and Hess, Thomas, "The Value of a Recommendation: The Role of Social Ties in Social Recommender Systems," in *System Sciences (HICSS), 2014 47th Hawaii International Conference on*, IEEE, 2014, pp. 1864-1873.

[6] Arazy, Ofer and Elsane, I and Shapira, Bracha and Kumar, Nanda, "Social relationships in recommender systems," in *Proc. of the 17th Workshop on Information Technologies & Systems*, 2007.

[7] Granovetter, Mark S, "The strength of weak ties," *American journal of sociology,* pp. 1360-1380, 1973.

[8] Koroleva, Ksenia and timac, Vid, "Tie Strength vs. Network Overlap: Why Information from Lovers is More Valuable than from Close Friends on Social Network Sites?," *Proceedings of the 33rd International Conference on Information Systems (ICIS),* pp. 1-17, 2012.

[9] Li, Seth and Karahanna, Elena, "Peer-based recommendations in online B2C e-commerce: comparing collaborative personalization and social network-based personalization," in *System Science (HICSS), 2012 45th Hawaii International Conference on*, IEEE, 2012, pp. 733-742.

[10] Konstan, Joseph A and Riedl, John, "Recommender systems: from algorithms to user experience," *User Modeling and User-Adapted Interaction,* vol. 22, pp. 101-123, 2012.

[11] Ricci, Francesco and Rokach, Lior and Shapira, Bracha, "Introduction to recommender systems handbook," in *Recommender systems handbook*, Springer, 2011, pp. 1-35.

[12] Ekstrand, Michael D and Riedl, John T and Konstan, Joseph A, "Collaborative filtering recommender systems," *Foundations and Trends in Human-Computer Interaction,* vol. 4, pp. 81-173, 2011.

[13] Pazzani, Michael J and Billsus, Daniel, "Content-based recommendation systems," in *The adaptive web*, Springer, 2007, pp. 325-341.

[14] Zafarani, Reza and Abbasi, Mohammad Ali and Liu, Huan, Social media mining: an introduction, Cambridge University Press, 2014.

[15] Ma, Hao and King, Irwin and Lyu, Michael R, "Effective missing data prediction for collaborative filtering," in *Proceedings of the 30th annual international ACM SIGIR conference on Research and development in information retrieval*, ACM, 2007, pp. 39-46.

[16] Linden, Greg and Smith, Brent and York, Jeremy, "Amazon. com recommendations: Item-to-item collaborative filtering," *Internet Computing, IEEE*, vol. 7, pp. 76-80, 2003.

[17] Resnick, Paul and Iacovou, Neophytos and Suchak, Mitesh and Bergstrom, Peter and Riedl, John, "GroupLens: an open architecture for collaborative filtering of netnews," in *Proceedings of the 1994 ACM conference on Computer supported cooperative work*, ACM, 1994, pp. 175-186.

[18] Sarwar, Badrul and Karypis, George and Konstan, Joseph and Riedl, John, "Item-based collaborative filtering recommendation algorithms," in *Proceedings of the 10th international conference on World Wide Web*, ACM, 2001, pp. 285-295.

[19] Bell, Robert M and Koren, Yehuda, "Scalable collaborative filtering with jointly derived neighborhood interpolation weights," in *Data Mining, 2007. ICDM 2007. Seventh IEEE International Conference on*, IEEE, 2007, pp. 43-52.

[20] Gong, SongJie and Ye, HongWu and Tan, HengSong, "Combining memory-based and model-based collaborative filtering in recommender system," in *Circuits, Communications and Systems, 2009. PACCS'09. Pacific-Asia Conference on*, IEEE, 2009, pp. 690-693.

[21] Burke, Robin, "Hybrid recommender systems: Survey and experiments," *User modeling and user-adapted interaction*, vol. 12, pp. 331-370, 2002.

[22] Lakshmi, Soanpet Sree and Lakshmi, T Adi, "Recommendation Systems: Issues and challenges," *(IJCSIT) International Journal of Computer Science and Information Technologies*, vol. 5, no. 4, pp. 5771-5772, 2014.

[23] Guo, Guibing, "Resolving data sparsity and cold start in recommender systems," in *User Modeling, Adaptation, and Personalization*, Springer, 2012, pp. 361-364.

[24] Gilbert, Eric and Karahalios, Karrie, "Predicting tie strength with social media," in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, 2009, pp. 211--220.

[25] Leskovec, Jure and Huttenlocher, Daniel and Kleinberg, Jon, "Signed networks in social media," in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, ACM, 2010, pp. 1361-1370.

[26] Khanafiah, Deni and Situngkir, Hokky, "Social balance theory: revisiting Heider's balance theory for many agents," 2004.

[27] Kautz, Henry and Selman, Bart and Shah, Mehul, "Referral Web: combining social networks and collaborative filtering," *Communications of the ACM,* vol. 40, pp. 63-65, 1997.

[28] Victor, Patricia and Cornelis, Chris and De Cock, Martine and Teredesai, Ankur, "A Comparative Analysis of Trust-Enhanced Recommenders for Controversial Items," in *ICWSM*, 2009.

[29] Victor, Patricia and De Cock, Martine and Cornelis, Chris, "Trust and recommendations," in *Recommender systems handbook*, Springer, 2011, pp. 645-675.

[30] Jamali, Mohsen and Ester, Martin, "Trustwalker: a random walk model for combining trust-based and item-based recommendation," in *Proceedings of the 15th ACM SIGKDD international conference on Knowledge discovery and data mining*, ACM, 2009, pp. 397-406.

[31] Golbeck, Jennifer, Generating predictive movie recommendations from trust in social networks, Springer, 2006.

[32] Massa, Paolo and Avesani, Paolo, "Trust-aware recommender systems," in *Proceedings of the 2007 ACM conference on Recommender systems*, ACM, 2007, pp. 17-24.

[33] Marsden, Peter V and Campbell, Karen E, "Measuring tie strength," *Social forces,* vol. 63, no. 2, pp. 482-501, 1984.

[34] Steffes, Erin M and Burgee, Lawrence E, "Social ties and online word of mouth," *Internet research,* vol. 19, pp. 42-59, 2009.

[35] Easley, David and Kleinberg, Jon, Networks, crowds, and markets: Reasoning about a highly connected world, Cambridge University Press, 2010.