

Article

An Information Recommendation Technique Based on Influence and Activeness of Users in Social Networks

Minsoo Lee * and Soyeon Oh 

Department of Computer Science and Engineering, Ewha Womans University, Seoul 03760, Korea; soyeon.oh@ewhain.net

* Correspondence: mlee@ewha.ac.kr; Tel.: +82-2-3277-3401

Featured Application: The technique proposed in this paper could be used to provide smarter recommendations via analyzing the semantics of the social information gathered from various sources in a socially connected and networked environment. Such applications could be performing recommendations on things such as products, restaurants, travel services, medical services, insurance service and so on to provide problem solutions or enable smart decisions or recommendations based on the gathered social information.

Abstract: Over the past few years, the number of users of social network services has been exponentially increasing and it is now a natural source of data that can be used by recommendation systems to provide important services to humans by analyzing applicable data and providing personalized information to users. In this paper, we propose an information recommendation technique that enables smart recommendations based on two specific types of analysis on user behaviors, such as the user influence and user activity. The components to measure the user influence and user activity are identified. The accuracy of the information recommendation is verified using Yelp data and shows significantly promising results that could create smarter information recommendation systems.

Keywords: social network service; recommendation technique; user influence; user activity



Citation: Lee, M.; Oh, S. An Information Recommendation Technique Based on Influence and Activeness of Users in Social Networks. *Appl. Sci.* **2021**, *11*, 2530. <https://doi.org/10.3390/app11062530>

Academic Editors: Barbara Guidi, Carlos A. Iglesias, Giulio Rossetti and Kevin Koidl

Received: 1 February 2021
Accepted: 9 March 2021
Published: 12 March 2021

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2021 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

The explosive growth in social network services has now made it a common place for communication among various communities and a platform to build relationships through interactions. It also supports mutual exchange of information based on common interests and could be a great source of data for information recommendation [1].

Although it may be greatly desired to provide such information recommendation services based on vast amounts of social network information, there are several difficulties in making such an information recommendation to appear to be intelligent for the user. There are vast amounts of social relationships in a social network and it is difficult to filter the essential information that meets the users' needs directly from the massive amounts of produced data. The demand from various users to find the information that they need has increased greatly. To obtain the desired information, many users search the information produced by others with social relationships that meet their specific requirements and needs [2–4].

Information recommendation has been mostly based on a collaborative filtering method that uses a matrix of users and items and contains user ratings on items. This matrix is used to identify similar users and similar items of interest to recommend items to target users. However, this method has a cold start problem, which occurs when there is not enough rating information in the matrix. With social network services becoming available, information recommendation approaches tried to use the social network information to solve the cold start problem by using neighbors of users in the social network to identify users with similar taste [5]. Although this may work in some situations, we find that

rather than just using simple neighbor information it is more effective to find users who are influencing the social network and consider those influencer's opinions with more weight in the information recommendation [6,7]. In addition to this, compared to the static matrix of users and items used by collaborative filtering approaches, the social networks can capture the dynamic timelines of users and thus contain historical information of users' activities. Such activity logs containing the users' interaction time and frequencies can enhance the credibility of the information recommendation by giving more weight to users who provide opinions more actively with newer information.

In this paper, we propose an information recommendation technique that reflects two specific types of analysis based on the social relations among the users participating in a social network. The two types of analysis are as follows. The first is user influence, which represents the analysis of factors affecting other users' behaviors. By analyzing the social network factors that affect user behavior, we can understand why users make decisions in specific situations and thus provide a better recommendation for the issue suggested. The second is user activity, which shows the level of activeness of the user who is eligible for creating the relevant information. The user activity in a social network is important for information recommendation because users who show specific activity patterns may be possessing valid and recent information useful to the recommendation. We have applied these proposed concepts on real Yelp [8] social network data to verify that the user influence and user activity concepts in social networks are useful for information recommendation.

The research methodology for the proposed information recommendation techniques based on social network data is as follows. First, we establish the concepts of user influence and user activity in the context of social network data and explain their relationship with the information recommendation. We investigate the core components that can define the concepts of user influence and user activity based on social network data. Formulas including these core components, to calculate the user influence and user activity, are devised. Based on these concepts, we can map these core components to real individual social network data components. Second, we use Yelp data to calculate the influence and activity of the users in a social network. The Yelp data components are mapped to the individual components in the formulas to calculate the user influence and user activity. Third, the most influential users and the most active users are identified, and we verify that these users can closely represent the overall user ratings. This means that user influence and user activity can be used to efficiently estimate credible ratings given by influential users and active users in a social network.

This paper is organized as follows. Section 2 discusses the techniques and data that are used in this work. Section 3 describes two major techniques used in our analysis along with their components and mathematic formula. In Section 4 we tested the applicability of this technique by calculating the user influence and user activity based on each factor suggested in Section 3, using real-world data. Finally, Section 5 provides the summary and conclusion.

2. Related Work

2.1. Recommendation Based on Social Networks

Recommendation techniques are being actively researched with additional social network data becoming increasingly available. Most approaches try to integrate collaborative filtering with additional social network information [9]. The friend relationship is mostly the considered factor in social networks for the recommendation techniques [1]. This is due to the significance of the influence of friends on the users. Most of the information recommendation approaches have focused solely on the friend relationship in social networks and rather put more emphasis on providing an efficient way of calculating the friends of users to nicely integrate the results into the information recommendation in a timely manner [10].

However, the social network has evolved and possesses much more information than just friend relationships and new kinds of patterns of influence are occurring among users in terms of marketing and purchasing behaviors led by so called influencers in the social network [7]. The notion of influence can be defined as a flow that brings change into the attitude or behavior of others. The concept of influence in many existing approaches just consider the quantitative aspects, such as the number of followers in a social network service [11,12]. However, the influence concept should also deal with the qualitative aspects, such as reviews, interest, and others that cause the change in attitude or behavior of users. Our work considers several such components that could cause the change in user attitude and behavior.

When considering influence in social networks, it is also very useful and efficient to put emphasis on the considerable minorities who contribute to influence on multiple users and induce the active behaviors from them [13–15]. Such an influence can be captured by incorporating experts existing in the social network.

In networked structured environments, such as the Web, traditionally important pages or nodes in a network were identified by using PageRank algorithms [16], which are specific examples of general random walk algorithms that the rank importance of nodes only based on the static link structure [17–20]. These algorithms were also applied to identify friends in a static social network for the information recommendation. In our approach, we identify influential users based on the characteristics of the dynamic interactions among users rather than just using static link information. As multimedia data are increasingly available on social networks these days, research issues on using the dynamic interaction among users and the multimedia contents for information recommendation is gaining interest [21].

Information recommendation can be enhanced by using historical data from the social network, especially for applications in mobile environments. The mobile environment enables historical location and time data to be efficiently managed and can be combined with the social network data when performing information recommendation [22,23]. By analyzing such historical information, the credibility of the information provided by a user could be more accurately estimated. Reference [24] uses a time weight in information recommendation to gradually decrease the rating provided by users based on time. The user activities in the social network may be monitored while historic information regarding the activities are maintained and become very useful for information recommendation. Social networks accumulate information about the user's past evaluation activities, making it possible to analyze the user's tendency of evaluations. Such historical information can include things such as the time that the evaluations are provided for, the content of the interactions, the location of the user, and the time that was spent in the social network. This historical activity information is used in our work to carefully estimate the activeness of a user, which can represent the validness or credibility of the information provided by the user.

2.2. Yelp Data

Yelp provides a local information service as a multinational corporation headquartered in San Francisco, California. It intends to take an important role between local business and users. It was founded by Jeremy Stoppelman and Russel Simmons, which started the service in 2004.

Yelp provides a crowd source review on the Web and its application. It also runs on-line reservation applications, food delivery services, open tables, etc. It has more than 86 million mobile users per month and owns more than 95 million reviews [25].

Yelp publicly offers a sample set of data that consist of three files: the business file, user file, and review file. Each file also includes the Json Objects [26]. Table 1 shows an example of a user object in Yelp data, including the attribute, meaning, and example data of the Yelp data.

Table 1. User object in the Yelp data.

Attribute	Meaning	Example Data
type	type info (ex.user)	"type": "user"
user_id	assigned id for user	"user_id": "4duCDxDMiRJJbc2CmnziAg"
name	name of user	"name": "Douglas"
review_count	Total number of reviews from a certain user	"review_count": 19
average_stars	average stars from a certain user	"average_stars": 4.299999999999998
votes	votes from other users	"votes": {"funny": 1, "useful": 0, "cool": 0}
friends	id of user's friend	"friends": ["Cg4CUfihhK4mXKo1RYhVow", "sHGpxxBcP59Tzdd696nj_A",]
elite	the year when a certain user got "elite"	"elite": [2008, 2009, 2010, 2011]}
yelping_since	Signed in Yelp from when	"yelping_since": "2010-08"
compliments	messages from other users	"compliments": {"funny": 3, "cute": 1, "plain": 2, "writer": 1, "note": 1, " ... }
fans	the number of fans for a certain user	"fans": 3

The Yelp data is used for verifying the applicability of the proposed concepts. The reason that Yelp has been used is because it provides a public data set based on real social network data, provides a considerably large amount of data adequate for the experiments, and consists of a variety of social network data enabling various kinds of components to be used in many combinations compared to other social network data sources that only consist of ratings and friend data.

3. User Influence and User Activity

Social networks provide a large amount of various information, and influential people on such social networks usually are a source of information that many people recognize as very useful. Users that are very active on social networks can provide timely information in addition to providing abundant information. In this paper, we have tried to formalize the concept of influence and activity of users in a social network, so that the concept can be applied to information recommendation. User influence means the level of people having similar behaviors, as derived from the information that was generated by a specific user. User activity represents the activeness of users on a specific category from the moment when they were eligible to produce information. The user influence and activity can be applied to the evaluation given by the user. The total recommendation of an item will be dependent upon the integration of all such users' item evaluations that reflect the user influence and user activity considerations. Yelp data were used to analyze the user influence and user activity in experiments to investigate the accuracy of the proposed recommendation technique.

3.1. User Influence

User influence represents the amount of impact that a certain user has. The amount of impact here can be defined as the degree of influence that causes the change in other users' behaviors.

To analyze the level of influence, four different factors were considered. These major factors are shown below in Figure 1, as the number of generated useful information on a certain category, expertness, interest, and compliments that were given by other users.

The rationale for these four factors of user influence are as follows. First, an influential user is one who usually provides useful information to others in a specific category. For example, an influential user may write some message on a board and many other users may give many positive votes to the message. Second, an influential user is one who is generally acknowledged as an expert in the social network community with an

exceptional and formal status. An example is being selected as an elite class user or leader of the community. Third, an influential user is one who receives a lot of attention and interest from other users and has large number of followers or subscribers. For example, an influential user may have a lot of fans or subscribers of published content. Fourth, an influential user is one who usually gets positive compliments from others. For instance, the influential user may receive many messages containing many positive comments from other users.

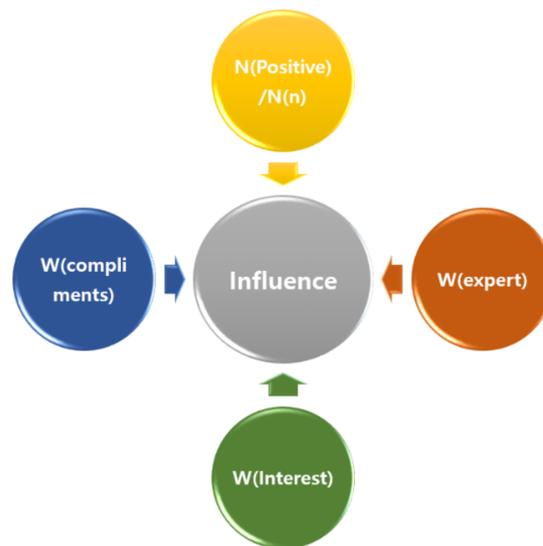


Figure 1. Elements of “influence”.

The *Influence* (P, C) that formulates the user P 's influence over category C is measured by the sum of the weights of four different factor variables, as shown in Equation (1).

$$Influence(Pn, Ci) = \alpha_1\beta_1 \frac{N(Reviews_{positive_{C_i}})}{N(Reviews_{all_{C_i}})} + \alpha_2\beta_2 W_{expert} + \alpha_3\beta_3 W_{interest} + \alpha_4\beta_4 W_{compliments} \quad (1)$$

In this formula, the four components each have two kinds of configurable parameters, α and β , which are multiplied with the components. The α is a parameter that is used to assign different weights to the components so that the different application requirements can emphasize on specific components. If the same value for α is given to all four components, the components are considered as of equivalent importance. The β is a parameter that is used to normalize the component values so that the ranges of the component values may be adjusted to be in identical or similar ranges of values. The β for each component could be the inverse value of the maximum of the component value to make the components fit into the range of 0 to 1. For example, if the maximum value of the third component is 1000, then β_3 may be assigned 0.001. If the β are all assigned as 1, no normalization is done, and the various component ranges will affect the outcome.

In this formula, the first component measures the amount of useful information regarding category C_i that is generated by the user Pn . The first component, $N(Reviews_{positive_{C_i}})/N(Reviews_{all_{C_i}})$, means the rate of positive reviews from other users compared to the total number of reviews regarding the category. The professionalism of a user, which is implemented as W_{expert} , is a factor that estimates whether a user is an expert or not related to the specific category. $W_{interest}$ represents the number of users who shows interest in the user by subscribing to information produced by the user. The number of positive messages from other users, $W_{compliments}$, considers personal messages that were sent by other users and especially the content of the messages including positive information. When considering the use of Yelp data, the data fields that could be used for the calculation

of the components in the proposed formula are as follows. The vote attribute of the review object could be used for the first component, the elite attribute in the user object could be used for the *Wexpert*, the fan attribute in the user object could be used for the *Winterest*, and the compliments attribute in the user object could be used for the *Wcompliments*.

In this formula, we have assigned configurable parameters to adjust and give a weight to each component. In different situations, the weights could be assigned as different values, and this could be provided as a configuration parameter for the information recommendation module. Some users may want to depend more on formally recognized experts, or some may want to put more emphasis on users that have more compliments. The weight parameter can support such needs by being applied before each component in the formula. However, the tuning of this configurable parameter for the weight may need some expertise and understanding of the data characteristics as well. In our experiments, we did not give any biased weight to any of the components because nobody really knows the intention of all users and we just wanted to see how the result comes out with a general case of giving equal weight to all components. This customization feature may enhance the usability as personalization would be possible for various information recommendation modes.

3.2. User Activity

User activity represents how active the user has been on a specific category since the user was eligible to create information. We consider the three factors shown in Figure 2 to analyze the user's activity.



Figure 2. Elements of “activity”.

To assess the quality of a user's evaluation activities, the three factors for the user activity were decided as follows. First, active users that provide meaningful and important recommendation data will generate evaluations that differentiate between good and bad items, and therefore will have a well distributed evaluation value range pattern. For example, a user who seriously considers evaluations will have some amount of variation in the evaluations, resulting in giving some high points and some low points to items. On the other hand, users who do not really care about giving meaningful evaluations on items will usually provide the same or simple alternating points for all items, which creates an abnormal distribution of the evaluation values. Second, active users will usually generate large numbers of opinion contents, and such users should be considered more dependable because they provide more content that people can consider and make decisions from. For example, users who have a lot of experience and useful information may provide a lot of reviews on various sites, giving lots of useful information. On the other hand, users who are not very active will generate few opinion contents and make it hard to decide if the user has enough credibility. Third, the timeliness of information is very important and usually active users would have recently generated opinion contents. For example, if the time that a review was written by a user is quite old, then we can consider the content of the information not very useful and the user is probably not very active recently. Active users that have written more recent reviews should have a higher significance.

In Figure 2, the coefficient of variation digitizes the distribution of the evaluations that were made by the users. It shows if either a user made a meaningful evaluation considering the characteristics of each item thoroughly, or just meaninglessly marked the same constant rating. The total amount of activity and degree of recent activity are the factors that calculate the activeness of the user from the point that the user was eligible for creating information till the recent period. They enable analysis of the user interest on specific categories and estimation of the reliability of the generated information.

$$Activity(P_n, C_i) = \alpha_1 \beta_1 \frac{\sqrt{\sum_{i=1}^N \frac{(EV_i - \overline{EV})^2}{N}}}{\overline{EV}} + \alpha_2 \beta_2 \sum_{D=1}^k activity_D \times \sqrt{D - T_{register}} + \alpha_3 \beta_3 \frac{1}{e^{(0.05T_{dormant})}} \quad (2)$$

Equation (2) shows the mathematical formula for user activity based on the three factors, which are the coefficient of variation, total amount of activity, and degree of recent activity that were explained above in Figure 2. The user activity denoted as $Activity(P_n, C_i)$ for user P_n regarding category C_i can be formulated by the sum of three weighted components each enclosed in separate boxes in Equation (2). In this formula, the three components also have two kinds of configurable parameters, α and β , which are multiplied with the components. The functions of these α and β are identical to the ones in Equation (1). The α is a parameter that is used to assign different weights to the components so that the different application requirements may be supported to put emphasis on specific components. The β is a parameter that is used to normalize the component values for adjusting the ranges of the component values to be in identical or similar ranges of values.

The first component represents the variation in the recorded evaluations given by the user and is calculated by the degree of variation in the evaluations as a coefficient. The variable EV_i denotes the evaluation value by user P_n regarding category C_i , and the overlined EV is the average of the EV . N is the number of evaluations made by the user. The second component provides the amount of the user's activity that is calculated by accumulating the multiplication of the activity level and activity weight variable for each day. The activity level measures how much activity the user has carried out each day such as the number of reviews written each day, and the activity weight uses a square root function to gradually put more weight on recent activity levels. The $activity_D$ is the number of activities or reviews created by the user on day D . The $T_{register}$ is the data that the user has registered to or created an account as a user for the social network. Lastly, the third part is for figuring out the level of recent activity and $T_{dormant}$ here means the time duration that the user was recently inactive and is calculated by the difference between the present time and the latest activity time.

In this formula, the configurable parameters to adjust and give weights have been assigned to each component as well. The weights could be assigned differently to reflect various user preferences on the components of information recommendation. This could be a configuration parameter for the information recommendation module. Some users may want to depend more on active users that give a very strict rating for evaluations or some may want to consider more recent information providing users. In our case, we again did not give any biased weight to any of the components because the real intention of users could vary, and we experimented with a general case of giving equal weight to all components.

The user influence and user activity are not directly associated with each other. The reason is that there are cases where a high user influence does not necessarily mean high user activity or low user activity. A user that has a lot of expertise and many fans with high user influence could have high user activity if the user has a lot of recent reviews, but alternatively could have low user activity if the user did not give reviews for a long time. The reverse situation also applies. High user activity does not necessarily mean high user influence or low user influence. A user that writes recent reviews with high user activity may get a lot of positive responses, resulting in high user influence, or may get a lot of negative responses, resulting in low user influence.

4. Performance Evaluation Results and Discussion

4.1. Experimental Setup and Method for Performance Evaluation

Experiments were carried out to verify the performance of the proposed information recommendation methods using social network data. The experiments were based on Yelp data. The category of the information recommendation is restaurants. The number of users considered were 55,452, the number of restaurants were 5556, and the total number of reviews considered were 233,718. The components of the user influence and user activity were mapped to Yelp data attributes, as described in the Sections 3.1 and 3.2. The reviews could obtain 11 types of different feedback responses from users and 7 of the types were considered as positive feedback responses.

The user influence and user activity were calculated for all 55,452 users. The users were then sorted in the order of high to low user influence and user activity, respectively. We suggest that the users showing high user influence or high user activity can provide a rating that would be acceptable to most users. This means that the rating provided by the user with high user influence or high activity should mostly be in line with the overall ratings of all users that provided ratings for the restaurant. To verify this, we first chose the top 10 users with the highest user influence and the highest user activity, respectively. The most recently visited 10 restaurants by each user were selected. For each of these restaurants, we then compare the rating provided by the user in the top 10 with the average rating of all users. This will show how much the user with high influence or high activity can represent the users. If this difference is small, it means that the user with high influence or high activity provides a representative rating. Otherwise, if the difference is large, it means that the rating provided is not a very representative one.

In most information recommendation approaches that use social network data, the friend information is used. We compared our proposed approach against the approach where only the friend information is used. We first identify the top 10 users that have the largest number of friends. The ratings of the 10 most recently visited restaurants by each of these users are compared against the overall ratings by all users that have given a rating. If the difference is small, the user with many friends provides a representative rating. Otherwise, the rating provided is not a very representative one.

Some additional analyses were performed to investigate the relationship between the user influence or user activity and some interesting attribute variables in Yelp. The Pearson correlation method was used to calculate the correlation.

4.2. Performance Evaluation for User Influence

For this test, 55,452 users who were evaluated in the restaurant category were extracted to verify the performance of the user influence. For all 55,452 users, the user influence was calculated based on the Yelp data attributes that map to the components of the user influence formula. The α and β configurable parameters were all set to 1, which means that the weights for each component were equally set to 1 and the components were not normalized, to see how the component data characteristics would be affecting the influence.

Among these users, the top 10 influential users were selected. Table 2 shows these top 10 influential users as well as some of the lowest 5 influential users. The user ID, calculated user influence, calculated user activity, number of elites, number of fans, number of reviews, and the recent review date are shown.

These top 10 influential users' evaluations of restaurants were then compared against the average evaluation of all users on the same restaurants. If the evaluations of the top 10 influential users are similar to the average of all user evaluations, the top 10 influential users may be considered to provide credible evaluations.

Figure 3 shows a comparison graph for each of the top 10 influential users. Each of the top 10 influential user evaluations were compared against the average evaluation of all users for each of the 10 or less restaurants most recently reviewed by each of the top 10 influential users. The restaurants are denoted as items in the graphs. The graphs show that the evaluation given by the top 10 influential users are very similar to the average

evaluations of all users. This shows that the top 10 influential users are providing an evaluation that has a very close pattern to the overall users.

Table 2. Top 10 and lowest 5 influential user data.

Rank	User ID	Influence	Activity	Num Elite	Num Fans	Num Friends	Num Reviews	Recent Date
Top 1	w6Vv-kldGpmvSGqXvTbAdQ	690.19616	220.88925	6	1172	459	5	22 February 2011
Top 2	8E0DGec8LNn6oDmPHmj-mg	633.01554	73.350444	7	1072	422	2	8 March 2011
Top 3	spJUPXI7QaIctU0FO5c42w	564.30388	72.313040	6	956	505	4	18 February 2011
Top 4	qbfQRHLvZk5WSkKY0l_IMw	557.45278	170.21839	4	948	306	7	14 December 2007
Top 5	rpOyqD_893cqmDAfjLbdog	522.11761	161.54219	10	877	527	4	15 March 2010
Top 6	DrKQzBFAvxhyjLgbPSW2Qw	400.63439	108.59097	8	672	368	4	20 January 2009
Top 7	mFOZOspQOacWIMVsyXbEbg	344.03795	187.33503	8	575	256	6	1 November 2012
Top 8	nrOCJCQUgXwdUIwg8QHirw	340.25130	235.14995	6	572	271	7	26 January 2013
Top 9	LbgQK5B_5IkN77FgRjHhrg	339.59576	131.90147	7	569	185	6	30 October 2013
Top 10	vyfsQo-estP8EfiIFMsL6g	338.85608	252.19245	8	566	98	6	22 May 2013
Low 5	wUXSmpXrGdztyKwz5b_Ng	0.0	2.83789×10^{-36}	0	0	0	2	8 August 2009
Low 4	itUTDvrHwmxU_C0P8x9sdw	0.0	4.48535×10^{-39}	0	0	3	1	1 April 2009
Low 3	6j4Oh-Lq2loLV5apkFjwTg	0.0	4.48535×10^{-39}	0	0	1	1	1 April 2009
Low 2	hyJ87UjROEtL-nbKSew_Ow	0.0	1.05758×10^{-41}	0	0	0	1	1 December 2008
Low 1	t5Xb5GY1QLj7ly7vugO4bg	0.0	1.50334×10^{-61}	0	0	0	1	1 June 2006

Figure 4 shows the difference between the evaluations by the top 10 influential users and the average of the evaluations by all users. Figure 4a shows the evaluation difference for each of the restaurants. The evaluation difference is very close to 0 for all the restaurants. This means that even for different restaurants the top 10 influential users do not display a big fluctuation and maintain consistent credibility. Figure 4b shows that the average of the difference for each top 10 influential user is within the +1 or −1 range. This shows that the top 10 influential users can provide reasonably credible evaluations for restaurants.

4.3. Performance Evaluation for User Activity

The 55,452 users who were evaluated in the restaurant category were used to verify the performance of the user activity. For all 55,452 users, the user activity was calculated based on the Yelp data attributes that map to the components of the user activity formula.

The α and β configurable parameters were again all set to 1, which means that the weights for each component were equally set to 1 and the components were not normalized, to see how the component data characteristics would be affecting the activity.

Among these users, the top 10 active users were selected. Table 3 shows these top 10 active users as well as some of the lowest 5 active users. The user ID, calculated user activity, calculated user influence, number of elites, number of fans, number of reviews, and the recent review date are shown.

These top 10 active users' evaluations of restaurants were then compared against the average evaluation of all users on the same restaurants. If the evaluations of the top 10 active users are similar with the average of all user evaluations, the top 10 active users may be considered to provide credible evaluations.

Figure 5 shows a comparison graph for each of the top 10 active users. Each of the top 10 active user evaluations were compared against the average evaluation of all users for each of the 10 or less restaurants most recently reviewed by each of the top 10 active users. The restaurants are denoted as items in the graphs. The graphs show that the evaluation given by the top 10 active users are generally similar to the average evaluations of all users in terms of the increase and decrease pattern trends. However, the evaluations by the top 10 active users show some strong opinions in the form of spikes occurring at some restaurants by giving very high evaluations for some restaurants that are moderately high evaluated by all the users, or by giving very low evaluations for some restaurants

that are moderately low evaluated by all the users. This could be understood from the characteristics of an active user, who will actively provide reviews that suggest strong opinions in the form of very good or very bad to gain constant interest from the social network community. This kind of spike in the data pattern creates a little more difference for some points compared to the user influence graphs. This shows that the top 10 active users are providing an evaluation that has a generally close pattern to the overall users but may show stronger opinionated evaluations in some cases.

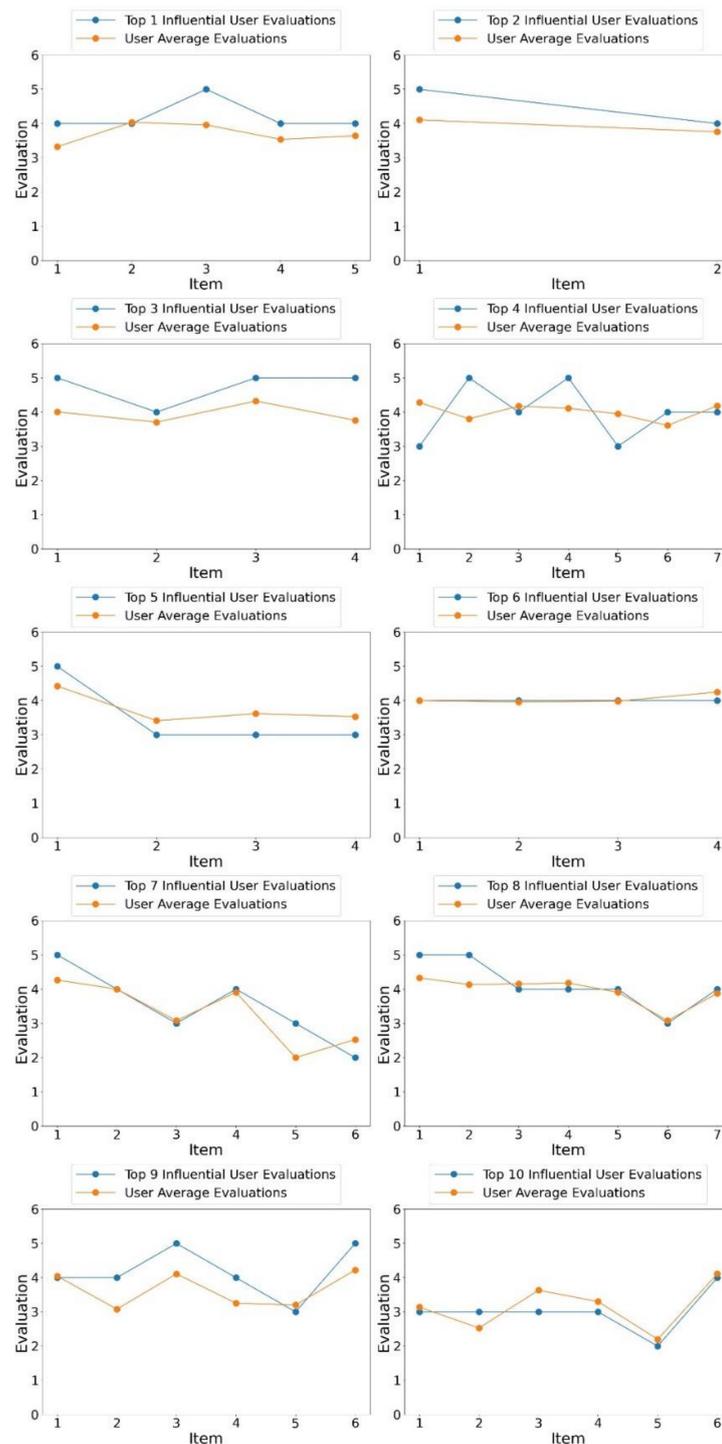


Figure 3. Comparison of the evaluations by the top 10 influential users with the average evaluations by all users.

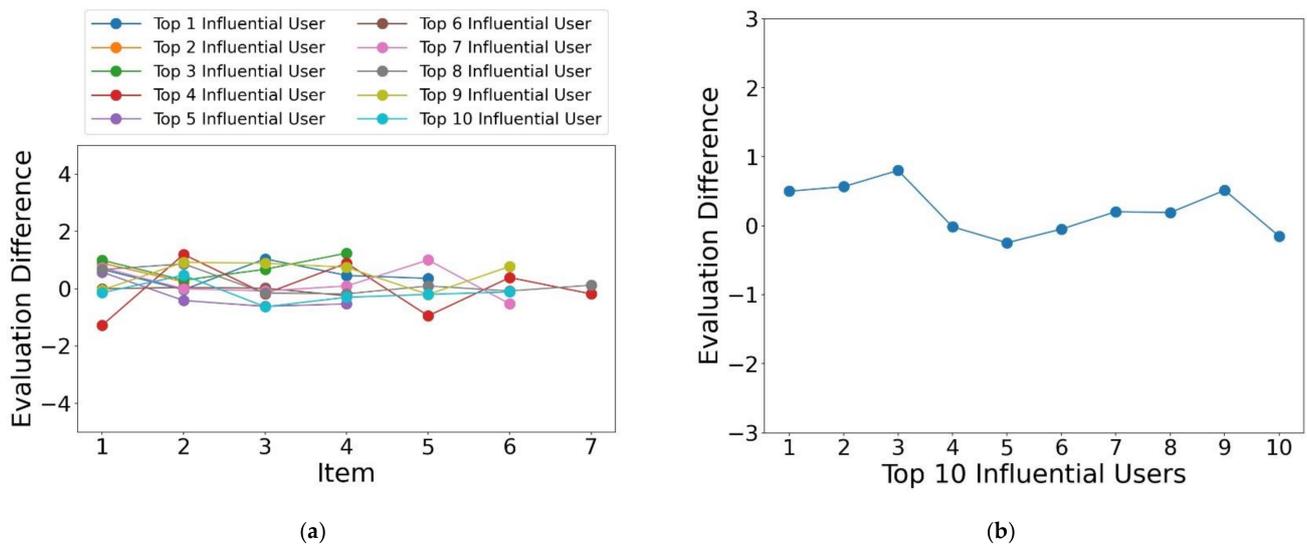


Figure 4. Evaluation difference between top 10 influential users and the average of all users: (a) the difference is shown for each restaurant; (b) the average evaluation difference is shown for each of the top 10 influential users.

Table 3. Top 10 and lowest 5 active user data.

Rank	User ID	Activity	Influence	Num Elite	Num Fans	Num Friends	Num Reviews	Recent Date
Top 1	kGgAARL2UmvCcTRfiscjug	10,832.10917	85.14843	1	143	632	301	11 December 2013
Top 2	DrWLhrK8WMZf7Jb-Oqc7ww	10,337.21580	7.18467	0	11	20	323	27 January 2014
Top 3	0bNXP9quoJEgyVZu9ipGgQ	7282.89702	74.09587	7	114	344	417	27 November 2013
Top 4	C6lOtaaYdLIT5fWd7ZYIuA	4794.69461	36.28776	7	49	709	339	5 January 2014
Top 5	pEVf8GRshP9HUKSpizc9LA	4733.53982	34.25725	5	49	46	287	1 December 2013
Top 6	q9XgOylNsSbqZqF_SO3-OQ	4429.29002	36.26023	7	49	403	269	22 January 2014
Top 7	H0lel3jz1MLNUJ6cc1x0Pw	4227.67528	22.42461	6	27	252	147	21 November 2013
Top 8	exefpuK6O1ctUUqTxq5XLg	3954.32938	4.79183	0	7	13	158	24 January 2014
Top 9	wHg1YkCzdZq9WBJOTRgxHQ	3522.54059	30.31668	7	39	168	212	8 January 2014
Top 10	kJyR4gT1pfCcNjEY9-YMoQ	3288.75576	2.16633	0	2	14	168	21 January 2014
Low 5	vCbrHCnLgTEccWpMcgFiRQ	5.03457×10^{-45}	0.5	0	0	1	1	1 July 2008
Low 4	wqPaSfr7teGzs-w3N1CO7g	2.38432×10^{-46}	0.4375	0	0	0	1	1 May 2008
Low 3	p5FcpR2d8u58rbTCDDQ1nw	7.95725×10^{-48}	0.16666	0	0	0	2	23 February 2008
Low 2	M1JCxPUKpIK8j09AfiLcfg	6.70276×10^{-59}	1.58316	0	1	1	1	1 October 2006
Low 1	t5Xb5GY1QLj7Iy7vugO4bg	1.50334×10^{-61}	0.0	0	0	0	1	1 June 2006

Figure 6 shows the difference between the evaluations by the top 10 active users and the average of evaluations by all users. Figure 6a shows the evaluation difference for each of the restaurants. The evaluation difference shows a little bit more fluctuation than the user influence case but is still reasonably close to 0 for all the restaurants. This means that, even for different restaurants, the top 10 active users do not display significantly big fluctuations and maintain a consistent credibility. Figure 6b shows that the average of the difference for each top 10 active users is within the +1 or -1 range. The minor fluctuations seem to be disappearing when calculating the average difference among the restaurants. In other words, the strongly opinioned evaluations by active users are sometimes very high but other times very low and are cancelled out when calculating the average of the differences. This shows that the top 10 active users can provide reasonably credible evaluations for restaurants with occasional stronger opinioned evaluations.

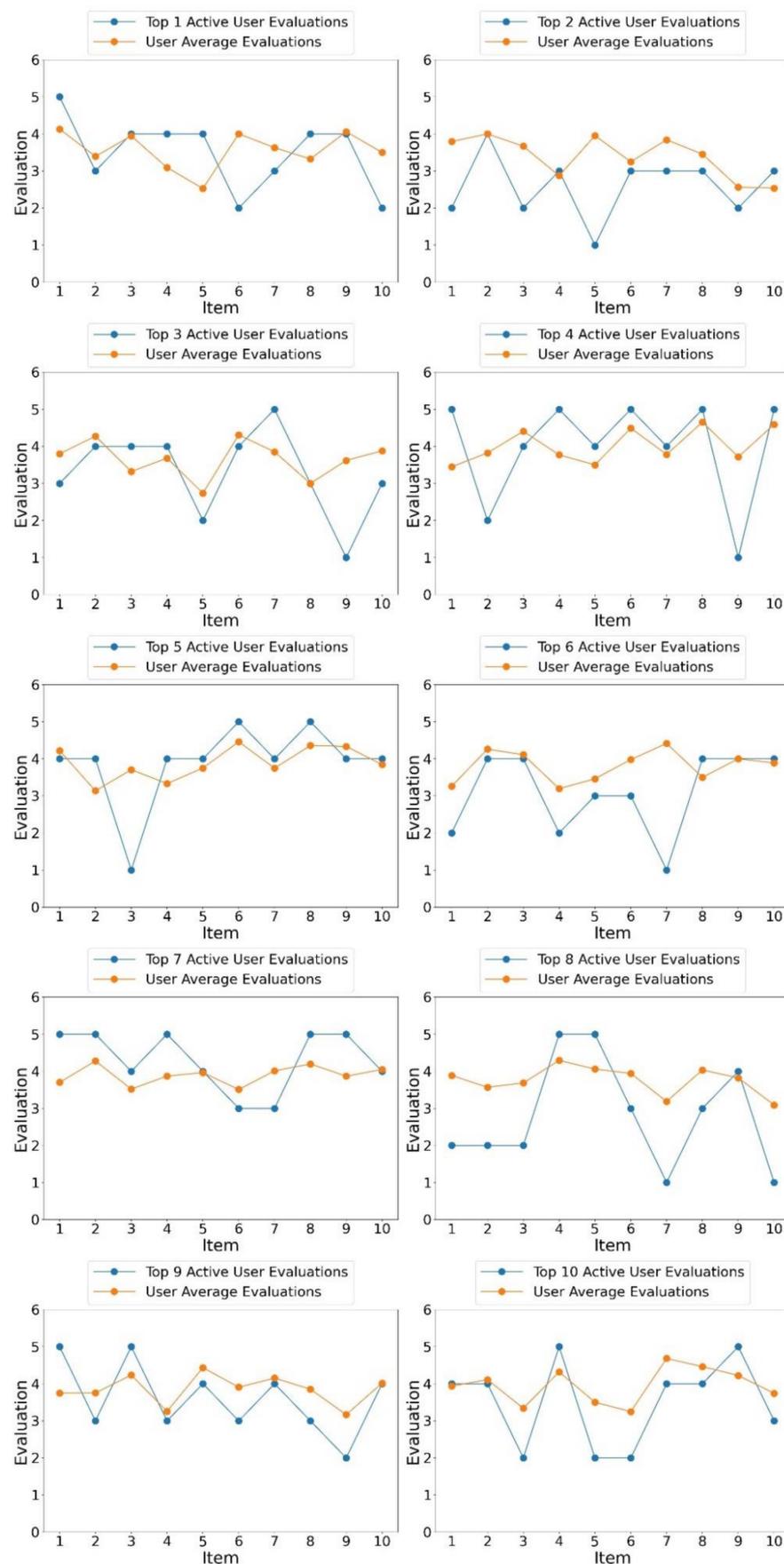


Figure 5. Comparison of the evaluations by the top 10 active users to the average evaluations by all users.

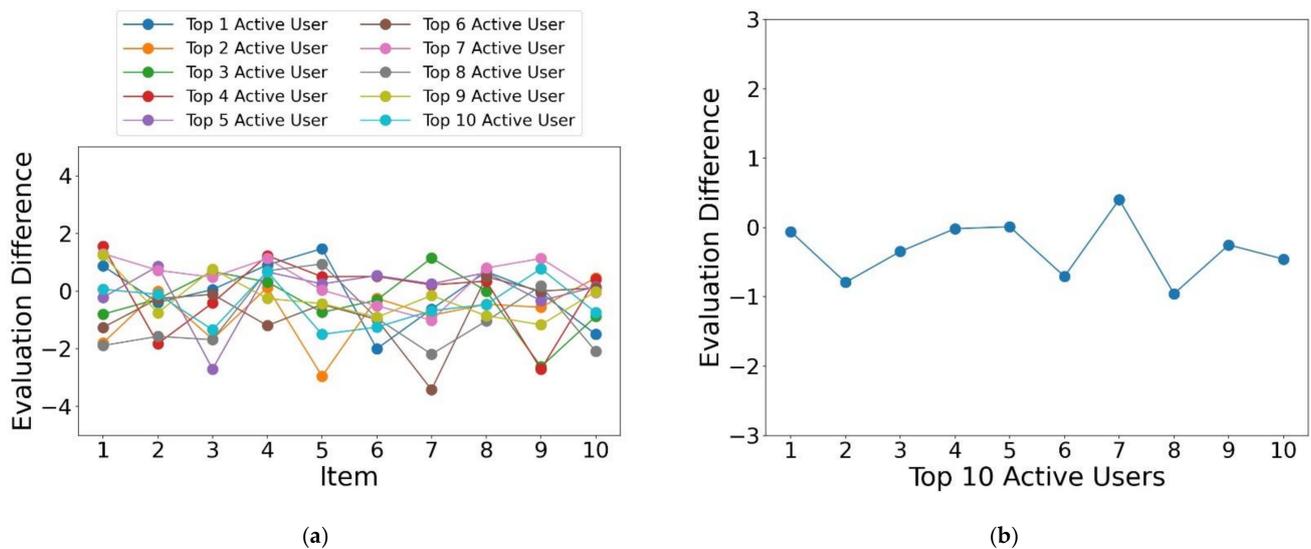


Figure 6. Evaluation difference between the top 10 active users and the average of all users: (a) the difference is shown for each restaurant; (b) the average evaluation difference is shown for each of the top 10 active users.

4.4. Comparison with User Friends

The most widely used approach for incorporating social network data into information recommendation is using the friend relationships in the social network. We investigate the performance of this approach that considers users with large number of friends so that it could be compared with our proposed concepts.

Among the 55,452 users who were evaluated in the restaurant category, the users with a large number of friends were used to compare the performance. The α and β configurable parameters were again all set to 1. The top 10 users with a large number of friends were identified and the evaluations of these users were compared with the average evaluations of all users for each of the 10 or less restaurants most recently reviewed by each of the top 10 users with a large number of friends.

Figure 7 shows a comparison graph for each of the top 10 users with a large number of friends. Each of the evaluations of the top 10 users with a large number of friends are compared against the average evaluation of all users for each of the 10 or less restaurants most recently reviewed by each of the top 10 users with a large number of friends. The restaurants are denoted as items in the graphs. The graphs show that the evaluation given by the top 10 users with a large number of friends show quite different pattern trends from the average evaluations of all users in terms of the increase and decrease pattern trends. Several restaurants were evaluated quite differently. For example, some of the restaurants with low evaluations by the top 10 users with a large number of friends were evaluated as high by all users, or vice versa.

In contrast, our proposed top influential and active users can provide evaluations that are more consistent with the pattern trend shown by the average evaluation of all users for each of the restaurants.

Figure 8 shows the difference between the evaluations by the top 10 users with a large number of friends and the average of evaluations by all users. Figure 8a shows the evaluation difference for each of the restaurants. The evaluation difference shows more fluctuation than the user influence case and is similar with the ranges of the user active case. It may be considered reasonably close to 0 for all the restaurants. Figure 8b shows that the average of the difference for each of the top 10 users with a large number of friends is within the +1 or -1 range. The minor fluctuations seem to be disappearing when calculating the average difference among the restaurants. Although this may look like the top 10 users with a large number of friends do not display significantly big fluctuations and maintain a consistent credibility, the observation from Figure 7 should also be considered

and understood that the difference here includes cases where the top 10 users with a large number of friends make the opposite evaluations as all the users. Making the opposite evaluation is more serious than having a strong opinion in an evaluation. The opposite evaluation cases make this approach less credible when considering the evaluations of users with a large number of friends.

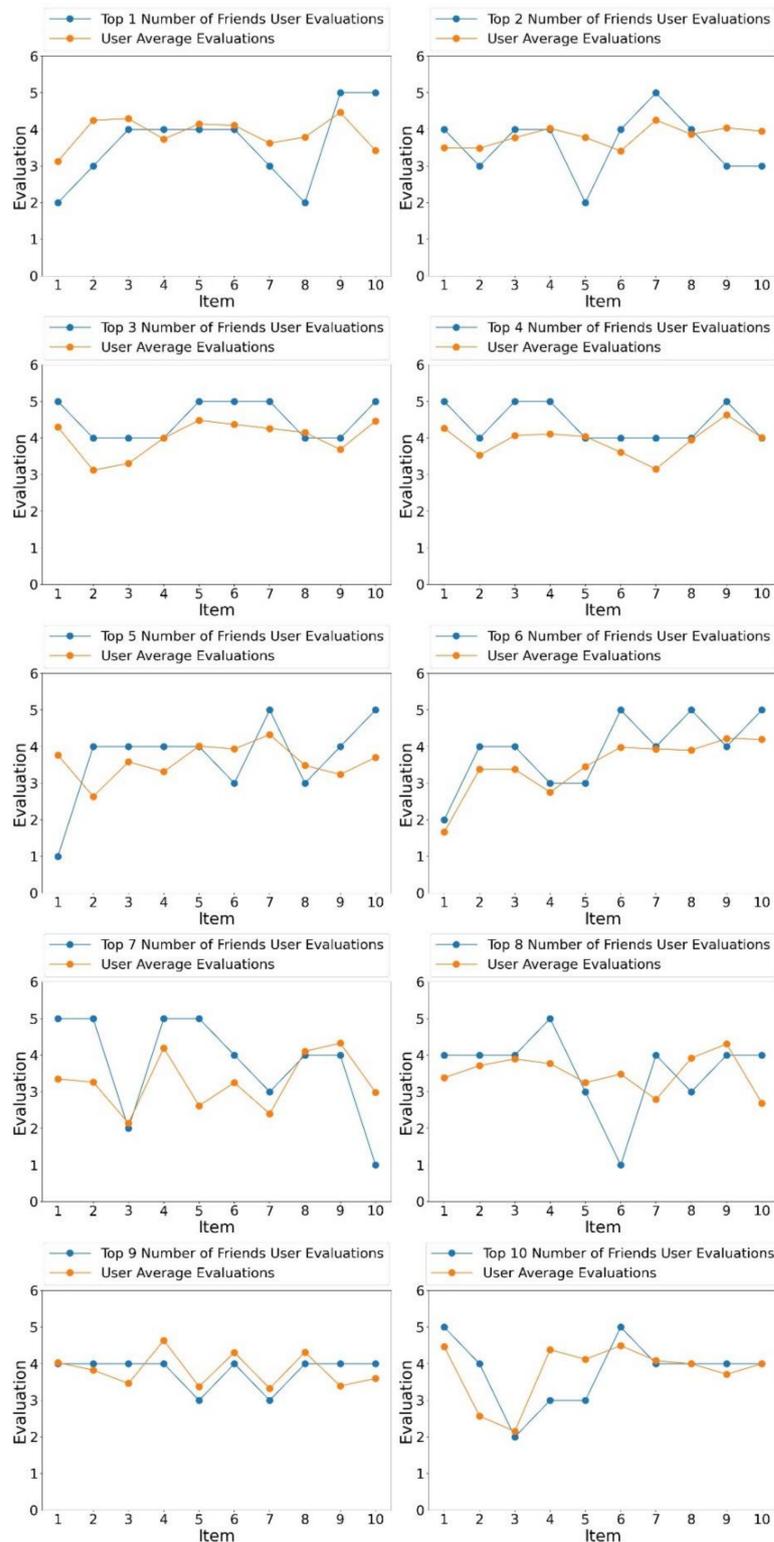


Figure 7. Comparison of the evaluations by the top 10 users with a large number of friends with the average evaluations by all users.

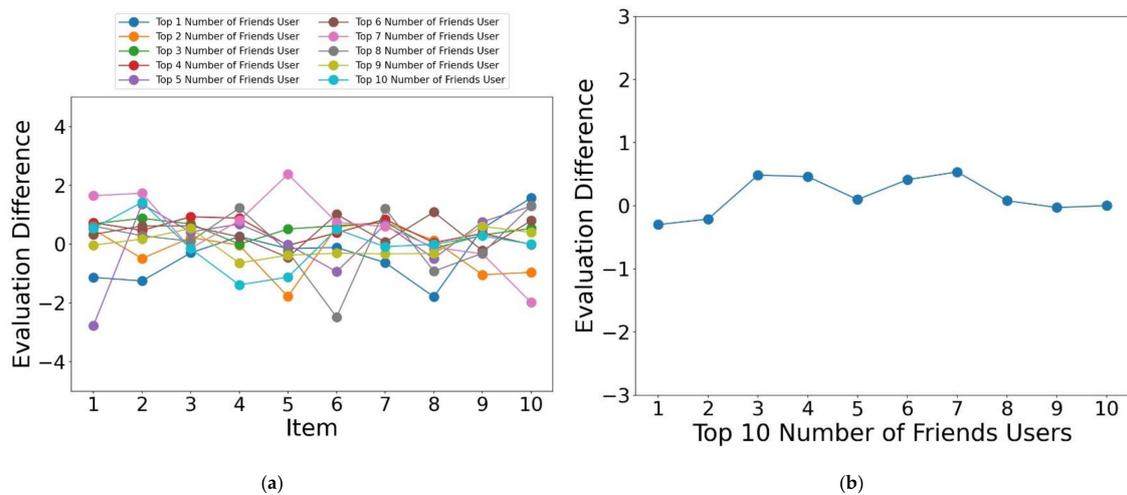


Figure 8. Evaluation difference between the top 10 users with a large number of friends and the average of all users: (a) the difference is shown for each restaurant; (b) the average evaluation difference is shown for each top 10 users with a large number of friends.

4.5. Analysis of the Correlations between User Influence, User Activity, and Yelp Variables

The user influence and user activity are dependent upon several variables in the Yelp data set. One could provide better insight into these concepts if the correlations between these concepts and the Yelp variables are analyzed. The correlations with the user influence and the correlations with the user activity are shown in Table 4. The correlations between user influence, user activity, number of elites, number of fans, number of friends, number of reviews, and recent review date were analyzed.

Table 4. Correlations between user influence, user activity, and the Yelp variables.

Concept	Influence	Activity	Num Elite	Num Fans	Num Friends	Num Reviews	Recent Date
Influence	1	0.141479	0.485973	0.993308	0.4809687	0.14866026	-0.09194966
Activity	0.141479	1	0.24435	0.119344	0.35697629	0.70084475	0.14612793

The user influence and user activity do not show a strong correlation. The user influence shows a strong correlation with number of fans and a moderate correlation with number of friends. This means that users that have a strong relationship with a large fan base would have much higher user influence. The user influence may be used as a measure to consider when marketing for image improvement is more important. The user activity shows a strong correlation with number of reviews and a moderate relationship with number of friends. This means that the users who write many reviews and have strong contents in reviews would have a much higher user activity. The user activity may be used as a measure to consider when the spreading of quality content is more important.

4.6. Considering Normalization of Components in User Influence, User Activity

Each of the different components of the user influence and user activity are dependent upon the social network data characteristics and could have a significantly different range of values. For instance, some components may be in the range of 0 to 1 while others may be in the range of 1000 to 3000. This significant difference in the range of values for the components could impact the calculation results of the user influence and user activity. To provide a way to adjust the range of values for the components, a configurable parameter β that normalizes the component value is provided. Additional experiments have been carried out to analyze the effect of normalizing the components of the user influence and user activity. The configurable parameter β is set to the inverse of the maximum value of each component, respectively. This will allow the component value to be scaled to a value

that is within the range of 0 to 1. For the user influence, the parameter values were set as $\beta_1 = 1.0$, $\beta_2 = 0.1$, $\beta_3 = 1.0$, and $\beta_4 = 1.125$. For the user activity, the parameter values were set as $\beta_1 = 1.125$, $\beta_2 = 0.0001$, and $\beta_3 = 1.1618$. The configurable parameters α that were used to assign weights were all set to 1. The user influence and user activity for all the 55,452 users were calculated and the top 10 influential users and top 10 active users were selected.

Figure 9 shows each of the top 10 influential user evaluations compared against the average evaluation of all users for each of the 10 or less restaurants most recently reviewed by the top 10 influential users.

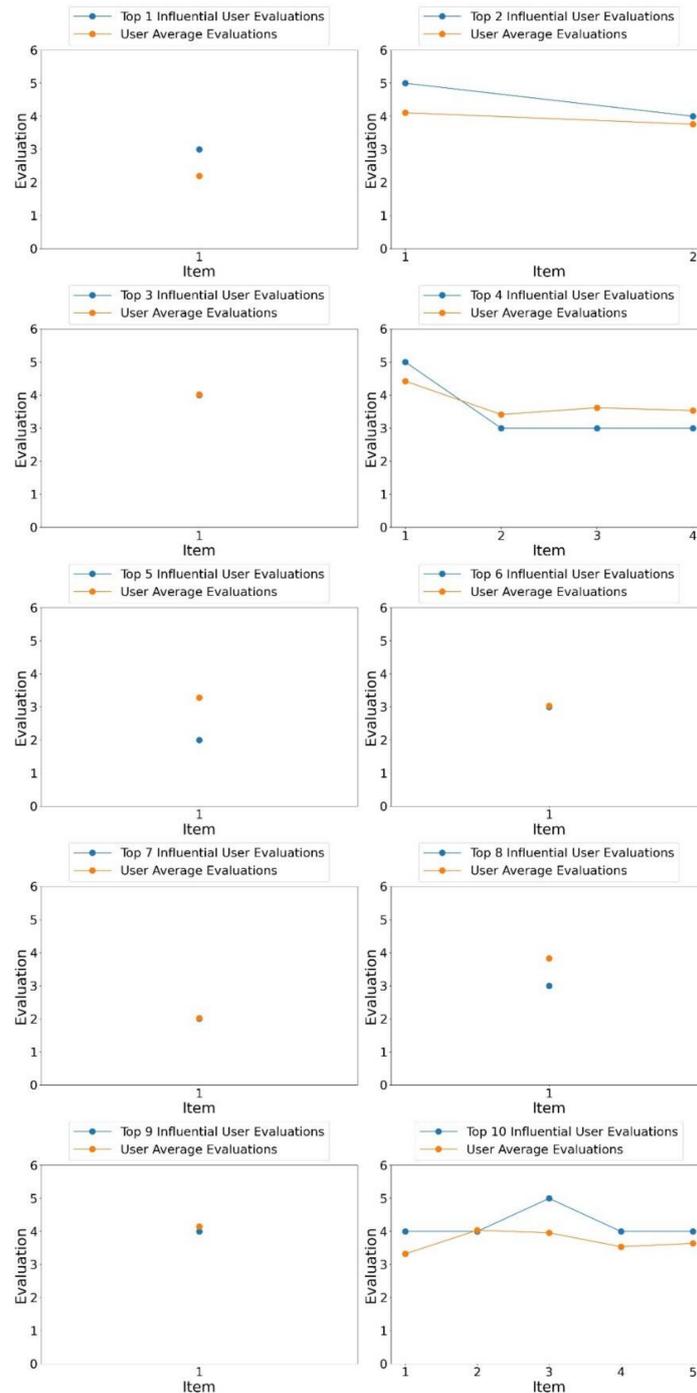


Figure 9. Comparison of the evaluations by the top 10 influential users with the average evaluations by all users.

The results show that the top 10 influential users make evaluations quite close to the average of all the users' evaluations, within a difference of -1 to $+1$. However, the selected users show some limitations, such as the number of reviews written by the users being very small, so in many cases only one restaurant could be included in the comparison. This makes it very hard to decisively conclude that these top 10 influential users' evaluations are credible.

Figure 10 shows a comparison graph for each of the top 10 active users. Each of the evaluations of the top 10 active users are compared against the average evaluation of all users for each of the 10 or less restaurants most recently reviewed by each of the top 10 active users. The results show that the evaluations by the top 10 active users are significantly different from the average evaluations by all users. The results verify that the pure normalization may not be sufficient for the selection of credible active users.

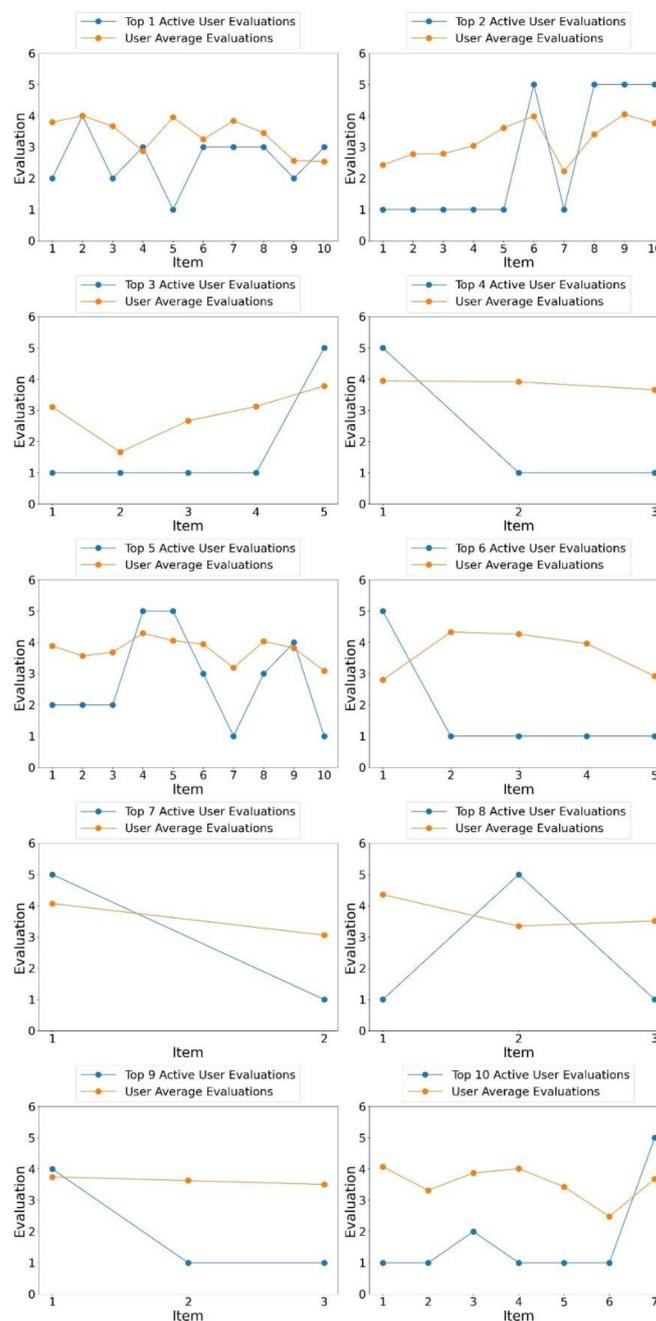


Figure 10. Comparison of the evaluations by the top 10 active users with the average evaluations by all users.

The analysis results from Figures 9 and 10 show that only normalizing and giving the same weight to the components may have unexpected results due to the data characteristics. Therefore, the data characteristics should be carefully understood, and the parameters need to be configured appropriately. In addition to normalizing the components, the weights of the components may be adjusted to reflect the emphasized part of the user influence and user activity.

The correlations with user influence and the correlations with user activity, with the normalized and equal weight components, are shown in Table 5. The correlations between user influence, user activity, number of elites, number of fans, number of friends, number of reviews, and recent review date were analyzed.

Table 5. Correlations between user influence, user activity, and the Yelp variables.

Concept	Influence	Activity	Num Elite	Num Fans	Num Friends	Num Reviews	Recent Date
Influence	1	−0.0389	0.3599	0.1643	0.1481	0.1039	−0.2211
Activity	−0.0389	1	0.0472	0.0132	0.1057	0.3048	0.1390

The user influence and user activity no longer show a strong correlation with any of the other variables. This could be because now that the components are normalized, a user that has an average value for each component could become better than users with very high values in only one of the components and low values in other components. The correlation among the user influence, user activity, and Yelp variables could become weak in these cases.

As an observation of our whole experiments, in the initial setting where no normalization was done and the same weight for all components were used, the emphasis naturally shifted to the component that had a large range value. The analysis results using this approach was considerably good. This provides insight into which component weight should have a larger weight assigned. In our second setting, with the normalization and equivalent weights for all components, this kind of characteristic was ignored, resulting in unexpected results. The weights of the components could be additionally adjusted to enhance the usefulness of the user influence and user activity concepts.

5. Conclusions

In this paper, we proposed an information recommendation technique that could use information generated from a social network environment with dynamic interactions among participants in the network. The social information recommendation approach especially considers concepts based on data provided in the Yelp data. This technique was shaped by two major components, namely, user influence and user activity. User influence includes four different factors: the number of generated useful information on a certain category, expertness, interest, and compliments that were given by other users. User activity is composed of three factors: the coefficient of variation, total amount of activity, and the degree of recent activity. Each of these components were experimented with using Yelp data and showed that they provide reasonable results in information recommendation using social networks.

The contributions of this work are as follows. First, the concepts of user influence and user activity based on social network data were introduced as a new direction on how to view and integrate social data into information recommendation. Second, while most approaches focus only on friend data when integrating social network data into information recommendation, our approach includes various data created in the social network to offer a guideline towards a more flexible and expandable model of information recommendation using social network data. Third, the methodology on how to apply the theoretical formulas for user influence and user activity on real social network data cases was shown by giving an example on applying the concepts to real Yelp data attributes. Fourth, the results of

our analysis on how user influence and user activity perform are given to provide useful insights and possible room for improvements on these concepts.

The concepts proposed can be applied to different social media platforms or different domains and topics. Different social media platforms will have a different organization of data and different types of data with different characteristics. This means that some of the components in the user influence or user activity may be missing or other more useful data may be available. For the missing components, they would just be ignored, making their component weight become zero. This means that the results may be different from the results in our experiments due to the difference in the social network data. If needed, adjustments could be done on the weights of the components and new kinds of components could be added as well. When applying the proposed concepts to different domains or topics, other than the restaurant item used in our experiment using Yelp data, the result is mostly expected to have a similar result because the components are designed to be domain or topic independent as much as possible. Nevertheless, the domain or topic could still be very special and yield different results if the domain or topic is an area where few reviews exist or very few experts, such as only a single person, exist. Social network data is very dynamic, and many unknown factors may still play an important role in various kinds of social network data. Therefore, more experiments may be carried out on different kinds of social network data as future work.

There could be more interesting concepts other than user influence and user activity for the information recommendation using social network data. Some concepts not dealt with in this work may be related to newer environments, such as mobile data providing locations that are frequently visited by a user. As new environments and more social network data are becoming available, more interesting concepts could be researched and added to our information recommendation using social network data, as further improvements.

Author Contributions: Methodology and software, S.O.; conceptualization and writing—review and editing, M.L. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported by the Ewha Womans University Research Grant of 2019.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: No new data were created or analyzed in this study. The program codes to perform the experiments for recommendation using Yelp data are available online at dwlabs.ewha.ac.kr/mlee/codes/sourcecode.zip. The programs are written in Python.

Conflicts of Interest: The authors declare no conflict of interest. The funders had no role in the design of the study; in the collection, analysis, or interpretation of data; in the writing of the manuscript, and in the decision to publish the results.

References

1. Li, Y.; Liu, J.; Ren, J. Social Recommendation Model Based on User Interaction in Complex Social Networks. *PLoS ONE* **2019**, *14*, e0218957. [[CrossRef](#)] [[PubMed](#)]
2. Chen, S.; Owusu, S.; Zhou, L. Social Network Based Recommendation Systems: A Short Survey. In Proceedings of the 2013 International Conference on Social Computing, Alexandria, VA, USA, 8–14 September 2013; pp. 882–885. [[CrossRef](#)]
3. Anandhan, A.; Shuib, L.; Ismail, M.A.; Mujtaba, G. Social Media Recommender Systems: Review and Open Research Issues. *IEEE Access* **2018**, *6*, 15608–15628. [[CrossRef](#)]
4. King, I.; Lyu, M.R.; Ma, H. Introduction to Social Recommendation. In *Proceedings of the 19th International Conference on World Wide Web—WWW '10*; ACM Press: New York, NY, USA, 2010; p. 1355. [[CrossRef](#)]
5. Castillejo, E.; Almeida, A.; López-de-Ipiña, D. Social Network Analysis Applied to Recommendation Systems: Alleviating the Cold-User Problem. In *Ubiquitous Computing and Ambient Intelligence*; Bravo, J., López-de-Ipiña, D., Moya, F., Eds.; Springer: Berlin/Heidelberg, Germany, 2012; Volume 7656, pp. 306–313, ISBN 978-3-642-35376-5. [[CrossRef](#)]
6. Burgess, E. Recommendations from Influencers Rival That of Friends. Influencer Orchestration Network (ION). Available online: <https://www.ion.co/twitter-has-released-a-report-showing-consumers-seek-product-recommendations-from-influencers-almost-as-much-as-they-do-from-friends> (accessed on 19 February 2021).

7. Jiménez-Castillo, D.; Sánchez-Fernández, R. The Role of Digital Influencers in Brand Recommendation: Examining Their Impact on Engagement, Expected Value and Purchase Intention. *Int. J. Inf. Manag.* **2019**, *49*, 366–376. [[CrossRef](#)]
8. Yelp. Available online: <https://www.yelp.com/> (accessed on 1 December 2020).
9. Margaris, D.; Spiliotopoulos, D.; Vassilakis, C. Social Relations versus near Neighbours: Reliable Recommenders in Limited Information Social Network Collaborative Filtering for Online Advertising. In Proceedings of the 2019 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining, Vancouver, BC, Canada, 27–30 August 2019; pp. 1160–1167. [[CrossRef](#)]
10. Tang, H.X.; Qian, X. Research on Recommendation Algorithm in Social Networks. *Appl. Mech. Mater.* **2014**, *496*, 1865–1868. [[CrossRef](#)]
11. Mukamakuza, C.P.; Sacharidis, D.; Werthner, H. The Impact of Social Connections in Personalization. In *Proceedings of the Adjunct Publication of the 27th Conference on User Modeling, Adaptation and Personalization*; ACM: New York, NY, USA, 2019; pp. 337–342. [[CrossRef](#)]
12. Jiang, M.; Cui, P.; Liu, R.; Yang, Q.; Wang, F.; Zhu, W.; Yang, S. Social Contextual Recommendation. In *Proceedings of the 21st ACM International Conference on Information and Knowledge Management—CIKM '12*; ACM Press: New York, NY, USA, 2012; p. 45. [[CrossRef](#)]
13. Berkani, L. A Semantic and Social-based Collaborative Recommendation of Friends in Social Networks. *Softw. Pract. Exp.* **2020**, *50*, 1498–1519. [[CrossRef](#)]
14. Davoodi, E.; Afsharchi, M.; Kianmehr, K. A Social Network-Based Approach to Expert Recommendation System. In *Hybrid Artificial Intelligent Systems*; Corchado, E., Snášel, V., Abraham, A., Woźniak, M., Graña, M., Cho, S.-B., Eds.; Springer: Berlin/Heidelberg, Germany, 2012; Volume 7208, pp. 91–102, ISBN 978-3-642-28941-5. [[CrossRef](#)]
15. Berkani, L.; Belkacem, S.; Ouafi, M.; Guessoum, A. Recommendation of Users in Social Networks: A Semantic and Social Based Classification Approach. *Expert Syst.* **2020**, e12634. [[CrossRef](#)]
16. Fujiwara, Y.; Nakatsuji, M.; Yamamuro, T.; Shiokawa, H.; Onizuka, M. Efficient personalized pagerank with accuracy assurance. In *Proceedings of the 18th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining—KDD '12*; ACM Press: New York, NY, USA, 2012; p. 15. [[CrossRef](#)]
17. Gupta, P.; Goel, A.; Lin, J.; Sharma, A.; Wang, D.; Zadeh, R. WTF: The who to follow service at Twitter. In *Proceedings of the 22nd International Conference on World Wide Web—WWW '13*; ACM Press: New York, NY, USA, 2013; pp. 505–514. [[CrossRef](#)]
18. Lempel, R.; Moran, S. SALSA: The stochastic approach for link-structure analysis. *ACM Trans. Inf. Syst.* **2001**, *19*, 131–160. [[CrossRef](#)]
19. Fujiwara, Y.; Nakatsuji, M.; Onizuka, M.; Kitsuregawa, M. Fast and exact top-k search for random walk with restart. *Proc. VLDB Endow.* **2012**, *5*, 442–453. [[CrossRef](#)]
20. Shahriari, M.; Jalili, M. Ranking Nodes in Signed Social Networks. *Social Netw. Anal. Min.* **2014**, *4*. [[CrossRef](#)]
21. Amato, F.; Moscato, V.; Picariello, A.; Sperli, G. Recommendation in Social Media Networks. In Proceedings of the 2017 IEEE Third International Conference on Multimedia Big Data (BigMM), Laguna Hills, CA, USA, 19–21 April 2017; pp. 213–216. [[CrossRef](#)]
22. Bao, J.; Zheng, Y.; Wilkie, D.; Mokbel, M. Recommendations in Location-Based Social Networks: A Survey. *GeoInformatica* **2015**, *19*, 525–565. [[CrossRef](#)]
23. Naik, P.; Desai, P.V.; Pati, S. Location Based Place Recommendation Using Social Network. In Proceedings of the 2019 IEEE 5th International Conference for Convergence in Technology (I2CT), Bombay, India, 29–31 March 2019; pp. 1–5. [[CrossRef](#)]
24. Dai, Y. A Collaborative Filtering Recommendation Algorithm Based on Time Weight. *Adv. Mater. Res.* **2010**, *159*, 667–670. [[CrossRef](#)]
25. About Yelp. Available online: <http://www.yelp.com/about> (accessed on 1 December 2020).
26. Yelp Open Dataset. Available online: <https://www.yelp.com/dataset> (accessed on 1 December 2020).