Characteristics of Online Social Services Sharing Long Duration Content

Hyewon Lim[1], Taewhi Lee[2], Namyoon Kim[1], Sang-goo Lee[1], and Hyoung-Joo Kim[1]
[1]Computer Science and Engineering, Seoul National University, Seoul, Republic of Korea
[2]BigData Software Platform Research Department, Electronics and Telecommunications Research Institute, Daejeon, Republic of Korea

Abstract – *Recent advances in web technologies have brought a boom in online social services. Online social services encompass a broad spectrum of how to cater to varying web users' interests, social content sharing being one of them. Studies show that users of social content sharing services tend to focus on the items themselves, rather than communicating with others at all. Unlike social networking services, which generally exhibit a healthy degree of user interaction, user interaction within social content sharing services shows much variation, from nonexistent to very active. Consumption time of target content has also shown to be an important factor in determining the various features of content sharing services. In this paper, we analyze social cataloging services and compare them to the YouTube network in* [1], *to figure out the characteristics of networks which contain content that requires a considerable amount of time to fully digest, such as films or books. We find that our dataset shows a higher level of homophily, reciprocity, and a lower level of assortativity than those of the YouTube network. Furthermore, we study features that affect users' selection of* new items. *Our results show that interest similarity matter, the genre of books in particular.*

Keywords: social content sharing service, content consumption time, social cataloging service, influential features

1 Introduction

Over the past several years, online social services have grown at an unprecedented rate, and have served as a catalyst in the explosion of online media content. Unlike the past when only a small number of people were capable of creating media for the public to consume, we are already in an era of a user created content deluge. Online social services have played an important role in allowing creative users to share their content and find an audience. By supporting user activities, they contribute to enhancing interconnectivity, self-expression and information sharing [2]. Online social services can be classified by their differing aims, system components, and the behavioral patterns of users. Among them, social content sharing services deal with target content such as videos for YouTube, books for LibraryThing, and images for Flickr. The main purpose of social content service users is to share their items and to find other items that match their interests, rather than to make a relationship or communicate with others.

Social content sharing services look very similar to each other, but each of them has different characteristics depending on its target content. Music, video clips, or photos have short consumption times and users can usually access them within the service; however, movies or books, which we will refer to as long content, require relatively longer consumption times, spanning from a few hours to several days, and services dealing with long content usually provide users with only the meta-data and perhaps a fraction of actual content, such as a sample teaser or chapter. Because of the fundamental difference in how short and long content are provided in social content sharing services, content consumption time can affect user's content selection and communication. For long content, users will select items carefully based on personal preference; they also tend to express their opinions through reviews or scores rather than casually conversing with others. That is, long content users maintain their relationships by reading news feeds about their friends, without any interactions.

Although communication in social content sharing services is less active compared to that in more general social services, a user's item selection is not independent of others' behaviors and must be influenced by exposure to social opinion and sentiment, even by a miniscule amount, because of the various social features typical of such services. Social Influence is defined as change in an individual's thoughts, feelings, attitudes, or behaviors that results from interaction with another individual or a group [3]. In an online social service, users influence is based on the trust in users or contents, either through relationships or sharing items. Understanding the features that propagates influence is important in analyzing, recommending, or advertising items and users.

In this paper, we analyze long content services LibraryThing, a social book cataloging service, and Userstory Book, a similar service primarily used in Korea. As the names suggest, the target items are books. We compare said services' data to a YouTube analysis [1] in regard to assortative linking, reciprocity, and homophily, with varying content consumption times. We find that the assortativity of social content sharing services with long content is low, but the level of reciprocity and homophily is high. We also analyze social features such as interest similarity and user behavior in order to figure out which ones affect users' item selection. The most influential feature is the interest similarity among users; in our dataset the genre of a book plays an important role in making relationships and selecting new items. We believe that our work is the first study to analyze social content sharing services in terms of varying content consumption times.

The remainder of this paper is structured as follows: We briefly introduce some related work in section 2. Our data set and experiment results to compare with [1] are presented in Sections 3 and 4. Section 5 shows which features play an important role in users' item selection, followed by the conclusion and future work in Section 6.

2 Related Work

As online social services become more popular, researchers try to analyze them to understand their key characteristics. Some researches try to categorize the services; its result differs based on various perspectives. Kaplan *et al*. [4] sort social media services into six categories by intimacy and immediacy. In this research, social content sharing services are classified as "Content Community", with medium social presence and low self-presentation. In [5], online social services are categorized into four groups by formality and interaction. They put social content sharing services in the "Cooperation" group, which has low formality and high interaction.

Much effort has been made for analyzing the static and dynamic features of social content sharing such as network structure and user behavior. User behavior is an especially important feature in understanding phenomena present in social networks, such as social influence and user similarity. Mislove *et al*. [6] analyze the structural characteristics of four popular sites: LiveJournal, YouTube, Flickr, and Orkut. Wattenhofer *et al*. [1] study the social network in YouTube. They show the differences between the YouTube network and traditional online social networks using three features: assortative linking, reciprocity, and user homophily. They also show the dichotomy of 'social' and 'content' activities and examine said activities' popularity in YouTube. The analysis of social cataloging services performed by [7], [8] both use an aNobii network. In [7], they investigate structural and evolutionary features and mine geographical information. Tang *et al*. [8] study the reading diversity of users using five similarity measures. Calculation of the interest similarity of users by using relationships is performed in [9]. They show that the similarity decreases with the weakening of connection strength between users. Crandall *et al*. [10] study the role of user interactions between similarity and social influence. They find that social interaction is both a cause and effect of similarity and social influence. According to their research, users show a sharp increase in similarity immediately before their first interaction; after the interaction, their similarity is increases slowly.

There has also been much effort to model social network phenomena. In [5]–[8], the researchers try to model social influence using user behavior. Yeung et al. [11] propose a probabilistic model for user adoption behavior to capture implicit influence in social content sharing services. In [12], [13], they research topic-based social influence.

3 Dataset

Social cataloging services are suitable for figuring out the characteristics of long content services because it naturally takes longer to read a book than to view a photo or watch a short video clip.

Users make their own reading lists, rate the books, and write reviews. Naturally, they also make relationships, comment on other's pages, and join communities much like when using online social network services. In this paper, we use LibraryThing and Userstory Book for the dataset. Table 1 shows the summary of our dataset.

Table 1. Data Summary

	Library Thing	**Userstory Book**
The number of users	108,221	12,933
The number of relationships (unilateral)	302,728	13,591
The number of relationships (reciprocal)	225,783	7,582
The number of books	13,285,867	100,168
The number of comments	161,340	2,181

3.1 LibraryThing

LibraryThing is one of the most famous social cataloging services launched in 2005. It has almost 1.8 million users and over 80 million book information so far. The user relationship is unilateral; they do not need to get consent to be connected. In addition to the functions described above, it is possible to tag the book in the list of their own.

In a social cataloging service, user behavior related to books is more important than user relationships; therefore, it is difficult to crawl data using some users as a seed like in social network services. We choose users who have one of three books as a seed, the books being "The Casual Vacancy" and "The Hunger Games", which are the most popular books in our crawling period and "Wuthering Heights", which has been loved for a long time. The reason we select "Wuthering Heights" is to avoid collecting users in a limited age group. We collected the data from January 23rd, 2013 to January 30th, 2013 using breadth first search. The dataset contains 108,221 users, 302,728 unilateral relationships, 13,285,867 book entries, and 161,340 comments.

3.2 Userstory Book

Userstory Book is a social cataloging service in Korea launched in 2009. It has almost 20,000 users and over 180,000 book entries so far. The relationship is unilateral like that of LibraryThing. The users cannot tag their books on their own lists, but they can sort the status of the books into three groups: “plans to read”, “currently reading”, and “already read.” We collected the entire data until May 8th, 2012. The dataset contains 12,933 users, 13,591 unilateral relationships, 100,168 books, and 2,181 comments. The size of the dataset is remarkably small compared to the LibraryThing dataset; however, it is important to understand what the characteristics of the entire network of the social cataloging service imply. We will also show that the Userstory Book is suitable for analyzing because it has a similar structural tendency to the LibraryThing dataset despite the smaller size.

3.3 Analysis of network structure

In both services, user interactions occur if a user leaves messages on others’ pages (or wall in Userstory Book) or replies to reviews. In fact, most users rarely comment on others’ reviews; hence, we make an interaction graph based on the comments on personal pages. If there is an edge from user *u* to user *v*, it means *u* leaves a message on *v*’s page. We ignore the replies of the messages and the messages written to oneself. The interaction graph of LibraryThing consists of 29,989 users and 94,209 interaction edges, and 11,524 users have one or more reciprocal edges. In the case of Userstory Book, the interaction graph consists of 827 users and 1,287 interaction edges and 321 users have one or more reciprocal edges. It shows that 23% and 6.7% of the total users have interactions with others. Of these, only 16,534 and 420 users, about 50% of users from both interaction graphs, interact with their friends. These results mean that, like other social sharing services, most users usually use the service not for communicating with others, but for making their reading lists and leaving their own impressions. Figure 1 represents the degree distribution of the entire graph and the interaction graph. We omit the extremely large value, which represents the user who is an author and has over 5,000 degrees, in the entire graph of LibraryThing (the graph on left top). In the entire graph, there are extremely popular users who have lots of friends, but most of the users, about 90% have one or zero friends; in the interaction graph, there are also extremely popular users, but the difference from the entire graph is that most users have more than one friend.

The number of users who have one or more friends is 60,317 at LibraryThing and 2,699 at Userstory Book; about 56% and 21% of users make a friend. However, 25,972 users and 1,258 users have only one friend. It implies that the number of users who actively make social relationships is very low, and most users are indifferent to social behavior.

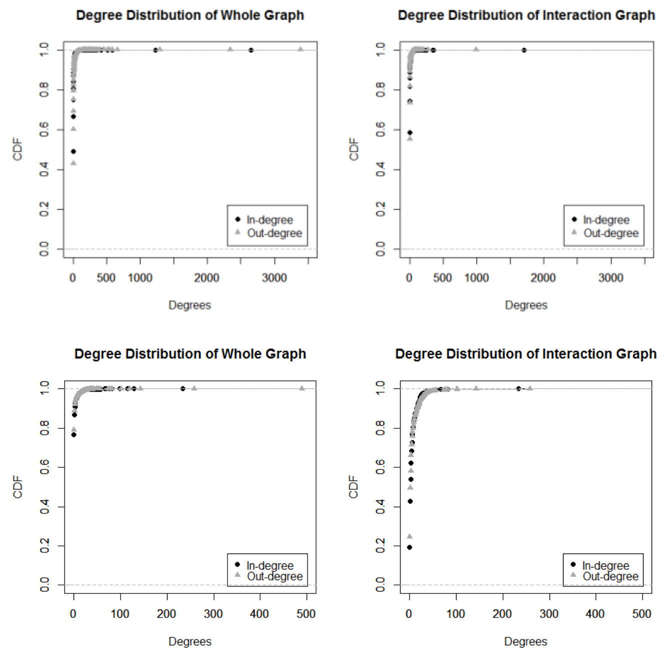


Figure 1. The degree distribution of the entire graph and the interaction graph for LibraryThing (top) and Userstory Book (bottom)

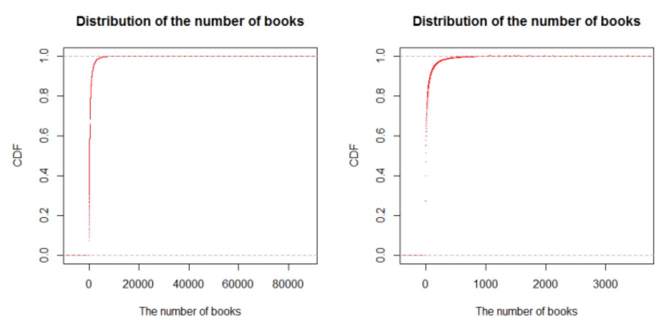


Figure 2. A cumulative distribution of books by how many a user has, for LibraryThing (left) and Userstory Book (right)

Figure 2 shows the distribution of the number of books each user has. There are also extremely active users who have many books in their list, but, in the case of Userstory book, most of the books’ status is set to “plans to read” and genre distribution is not skewed. It means the extremely active users are better than the rest in expressing their interest in all genres. According to our estimation, on average, each user has 5.31 friends and 388.25 books at LibraryThing and 1.05 friends and 31.65 books at Userstory Book, including books planned to be read. We also examine the relationships between the number of books and user popularity; the result is in Figure 3. Figure 3 shows that the most active users who have many books are not more popular than general users. The Pearson’s

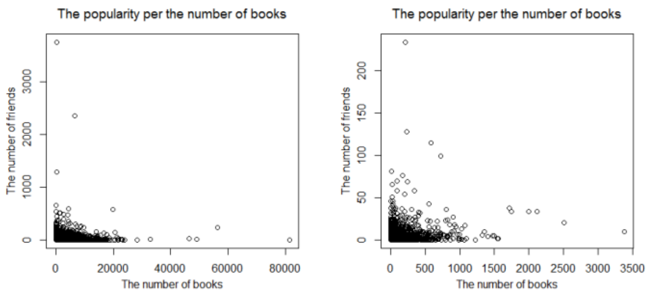


Figure 3. The relation between a user's popularity and the number of books he or she has, for LibraryThing (left) and Userstory Book (right)

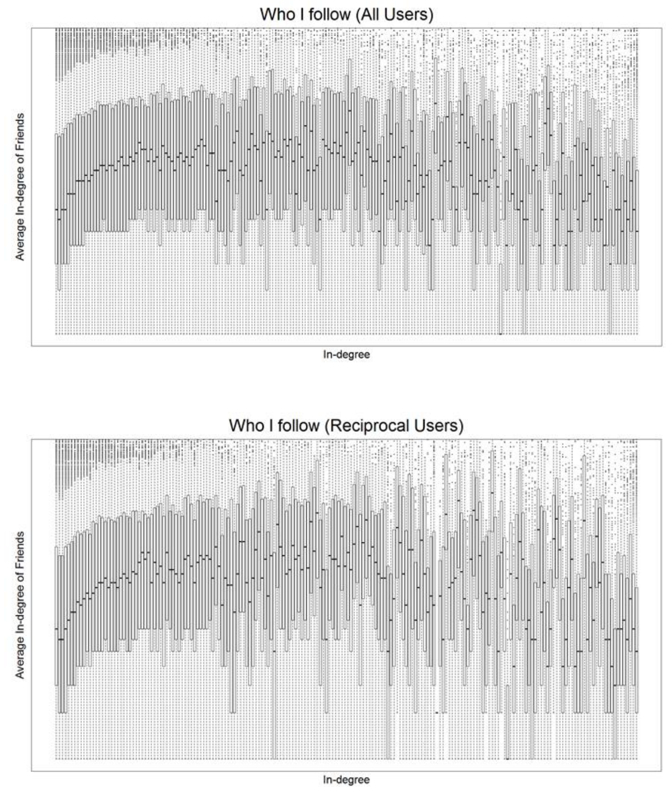
correlation coefficient is 0.154 and 0.276 for LibraryThing and Userstory Book, respectively. Active behaviors like reading many books and expressing one's interest does not affect making relationships and vice versa. This indicates that users present their impressions not for giving information to others, but for their own needs and self-satisfaction. Also, it shows that a social cataloging service is different from a social networking service in that active users have lots of friends.

4 Characteristics of Long Content Service

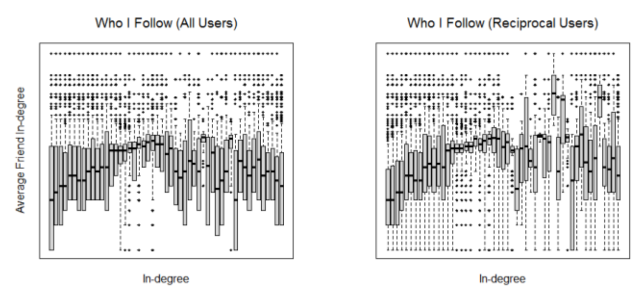
Wattenhofer *et al*. [1] find the differences between a YouTube network and a traditional online social network in terms of assortativity, reciprocity, and homophily. We analyze these features using our dataset and compare them to the result of [1] in order to figure out whether distinctions exist depending on content consumption time.

4.1 Assortativity

Assortativity is the tendency of nodes to connect with other nodes with similar degrees of a certain unit. We examine assortative links based on user popularity. The results are shown in figure 4. The x-axis represents the in-degree of the users, and the y-axis represents the average in-degree of their friends. Figure 4(a) is the result of the LibraryThing dataset and figure 4(b) is the result of the Userstory Book dataset. The plot shows that users from social relationships with others who have a certain amount of in-degrees regardless of the number of in-degrees of themselves, in both datasets. This tendency has nothing to do with the type of the links. The assortativity measurement of the subscription network in YouTube [1] shows that most users in a subscription graph subscribe to a publisher whose popularity is above a certain threshold, and significant differences depend on the type of links; reciprocal users are more assortative than the entire userbase.



(a)



(b)

Figure 4. Similarities between users and their friends' popularities using (a) the entire users and the reciprocal users in the LibraryThing dataset and (b) the entire users of the and the reciprocal users in the Userstory Book dataset. There is no difference in assoratívity between the entire users and the reciprocal users in both datasets.

4.2 Reciprocity

In our dataset, most links are reciprocal; 95.93% of links in the LibraryThing dataset and 73.32% of the Userstory Book dataset are bidirectional. That is, the number of the users who join in interacting with each other is less, but the level of reciprocity is very high. This implies that the relationships are usually superficial and tend to be passive engagements [14]. Most of the users keep up with friends by reading news, reviews, ratings, and such without actual communications. They do not try to develop deeper connections.

Our results are quite large compared to the reciprocities of other directed social networks such as YouTube at 25.42% [1], Filckr at 68% [15], and the sampling data of CiteULike which have only 93 reciprocal links out of 11,295 unilateral links [9]. The details of the results are in Table 2.

Table 2. The proportion of reciprocity

	Library Thing	**Userstory Book**
The number of links	60,317	2,699
Reciprocal links	57,864	1,979
	95.93%	73.32%

4.3 Homophily

We use genres of the books a user has to measure homophily among users. For this analysis, we only use the Userstory Book dataset because LibraryThing does not provide any information on genres. The books in a user's list cover a wide range of genres. Because there is more than one genre, we assume that the most read genre in the user list represents the user's primary interest. Prior to the analysis, we classified 29 different genres based on how online bookstores generally categorize their genres, because the genre classification in Userstory Book is too specific.

We measure the level of homophily in two graphs; the entire graph and the interaction graph. Regardless of the type of graph, about 50% of users are interested in the same genre as their friends: 49.24% from the entire graph and 52.25% from the interaction graph. This indicates that the social cataloging services are more homophilous than the YouTube network, which is on average 26.58% and 27.46% respectively [1]. We expect to get better results if the books are classified into fewer genres.

5 Influential Features in Long Content Services

Although the social content sharing services that provide longer content have low communication ratios, a user's content selection is not independent from that of others; users cannot be completely isolated from social influence because of the various social features provided by the services. We measure features that affect users' item selection.

We begin by examining the relational features such as friend and interaction networks. We assume that a user's friend influences the user the most because of the high reciprocity ratio. However, table 3 shows that 48,674 users and 1,759 users of both services share the same books with their friends, and about 20% of the books in a user's reading list show up in the user's friends' reading lists. In the case of interactional relationships, the percentage of user-friend reading list overlap is less than the results found in friend relationships. This is because half of the users who leave messages on others' pages have no relationships with the recipient, and the messages have little to do with books in particularly. The diversity of the books also affects this results.

Table 3. The result of social influence from friends: each column denotes (a) how many books a user has in common with his or her friends, (b) how many friends have one or more common books with each user, (c) the percentage of overlap between a user's book list and that of his or her friends, and (d) the number of users who have the same books with the list of their friends.

		(a)	**(b)**	**(c)**	**(d)**
Library Thing	AVG	127.36	4.94	24.33%	48,674
	SD	377.82	17.91	22.70%	
Userstory Book	AVG	17.31	4.44	22.80%	1,759
	SD	49.05	9.55	22.33%	

Table 4. Social influence of users who leave messages in others' pages: columns are the same as in table 3.

		(a)	**(b)**	**(c)**	**(d)**
Library Thing	AVG	108.07	2.96	16.60%	24,506
	SD	321.55	7.64	18.46%	
Userstory Book	AVG	10.05	1.86	13.83%	453
	SD	20.75	2.46	16.74%	

Next, we measure how similar a user is to his or her friends in terms of the genre composition of their reading lists. For this analysis, we only use the Userstory Book dataset and the 29 genres we classified in the homophily experiment because of the absence of genre information in LibraryThing. We consider that each user has an n-dimensional vector $\vec{v}(u_i)$,

$$\vec{v}(u_i) = (g_1, g_2, \cdots, g_n) \tag{1}$$

where n is the number of genres and g_k is the number of books belonging to the k^{th} genre the user u_i has. We set n = 29. For more meaningful results, we only focus on users who have ten or more books. The genre similarity between users are measured using cosine similarity [16],

$$GenreSim(\vec{u},\vec{v}) = Cosine(\vec{u},\vec{v}) = \frac{\vec{u}\cdot\vec{v}}{\|\vec{u}\|\|\vec{v}\|} \qquad (2)$$

Figure 5 shows the genre similarity. We observe that the genre similarity among users with social relationships is higher than that of all possible pairs of users. Among social relationships, the similarity of those with reciprocal relationships is a little higher. That is, users may consider common interests when choosing friends, and these relationships in turn can influence their item selection. As in the homophily experiment, we expect to get better results if we deal with fewer genres.

Shared recent activities of others' are helpful for users in choosing new books. Books can be registered in the reading list either before or after the reading. In the latter case, there is a time difference between beginning the book and registering the books after having completely read it. We assume that the average period of reading a book is within three days[1], and examine whether there are other users who have already registered the book before within the three days of another user's registration of the same book. If such cases happen quite often, we can consider that a user may be influenced by others' recent activities; we guess that there is a high possibility the user looks at the other's recently registered book in the news feed. Figure 6 shows results from the entire Userstory Book dataset. We also only use the Userstory Book dataset because they provide all users' activities. We find that about 80% of the books are not overlapped, a user's registration of a certain book does not happen during the three days after that same book is registered by any other user. This means people may see a book in other users' recent activities, but they do not select that book as their next item. We conclude that recent activity is not an influential feature in passive engagement.

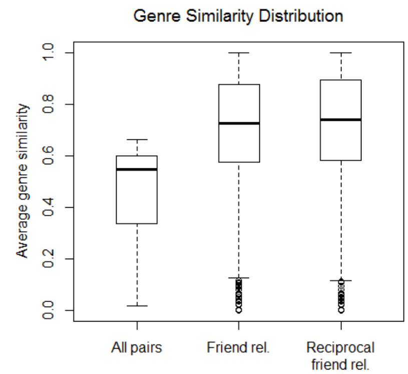


Figure 5. The distribution of genre similarity

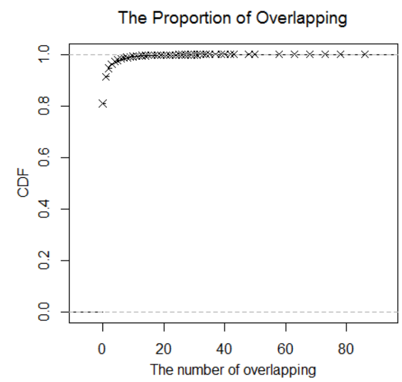


Figure 6. The proportion of the book overlapping between a user's reading list and others' recent activities

6 Conclusion

In this paper, we classified social content sharing services based on the consumption time of the target content. To figure out the characteristics of networks that have long content, we used two social cataloging services, LibraryThing and Userstory Book. The dataset of Userstory Book is quite smaller than the dataset of LibraryThing, but it also meant we were able to work on a comprehensive dataset of a social cataloging service. This is good for figuring out the general characteristics of long content services. We analyzed and compared the datasets to the YouTube network [1], a representative short content service, in terms of assortativity, reciprocity, and homophily. According to results, our datasets have low assortativity, but the level of reciprocity and homophily is higher than those of the YouTube network. We also study the social features in both datasets that affect users' item selection. From our results, we find that interest similarity is an important factor in determining users' item selection. We found that the genre of books also plays an important role in forming user relationships and selecting new items. For future work, we hope to focus deeper into social features of networks with long content such as the effect of communities and more detailed statistical information.

7 Acknowledgments

This work was supported by the National Research Foundation of Korea (NRF) grant funded by the Korea Government (MSIP) (No. 20110030812), by ETRI R&D Program ("Development of Big Data Platform for Dual Mode Batch-Query Analytics, 14ZS1400") funded by the Government of Korea, and by the National Research Foundation of Korea (NRF) grant funded by the Korea Government (MSIP) (No. 20110017480)

[1] We searched polls about the average time it takes to read a book. In these polls, more than half of people answered that they usually take 1 to 3 days to read a book. The sources are as follows: http://www.goodreads.com/poll/show/45995-how-long -does-it-take-you-to-finish-an-average-size-book-approx and http://dearauthor.com/features/poll-misc/poll-how-long-does-it-take-to-read-a-book/#ViewPollResults

8 References

[1] M. Wattenhofer, R. Wattenhofer, and Z. Zhu, “The YouTube Social Network,” in *Proceedings of International AAAI Conference on Weblogs and Social Media*, 2012, pp. 354–361.

[2] M. Papagelis and V. Murdock, “Individual Behavior and Social Influence in Online Social Systems,” *Proc. 22nd ACM Conf. Hypertext hypermedia - HT ’11*, pp. 241–250, 2011.

[3] L. Rashotte, “Social Influence,” *blackwell Encycl. Soc. Psychol.*, vol. 26, no. 3, pp. 562–563, 2007.

[4] A. M. Kaplan and M. Haenlein, “Users of the world, unite! The challenges and opportunities of Social Media,” *Bus. Horiz.*, vol. 53, no. 1, pp. 59–68, Jan. 2010.

[5] N. Cook and D. Tapscott, *Enterprise 2.0*. 2008, p. 180.

[6] A. Mislove, M. Marcon, K. P. Gummadi, P. Druschel, and B. Bhattacharjee, “Measurement and analysis of online social networks,” *Proc. 7th ACM SIGCOMM Conf. Internet Meas. - IMC ’07*, p. 29, 2007.

[7] L.-M. Aiello, A. Barrat, C. Cattuto, G. Ruffo, and R. Schifanella, “Link Creation and Profile Alignment in the aNobii Social Network,” *Soc. Comput. (SocialCom), 2010 IEEE Second Int. Conf.*, 2010.

[8] M. Tang, Y. Ke, and Y. Sie, “The Estimation of aNobii Users ’ Reading Diversity Using Book Co-ownership Data : A Social Analytical Approach,” in *The 5th International Conference on Social Informatics (SocInfo)*, 2013, pp. 274–283.

[9] D. H. Lee and P. Brusilovsky, “Social Networks and Interest Similarity: The Case of CiteULike,” in *Proceedings of the 21st ACM conference on Hypertext and hypermedia HT 10*, 2010, pp. 151–155.

[10] D. Crandall, D. Cosley, D. Huttenlocher, J. Kleinberg, and S. Suri, “Feedback effects between similarity and social influence in online communities,” *Proceeding 14th ACM SIGKDD Int. Conf. Knowl. Discov. data Min. KDD 08*, vol. 363, no. 3, p. 160, 2008.

[11] C. A. Yeung, “Capturing Implicit User Influence in Online Social Sharing Categories and Subject Descriptors,” *Proc. 21st ACM Conf. Hypertext hypermedia HT 10*, pp. 245–254, 2010.

[12] J. Tang, J. Sun, C. Wang, and Z. Yang, “Social influence analysis in large-scale networks,” *Proc. 15th ACM SIGKDD Int. Conf. Knowl. Discov. data Min. - KDD ’09*, p. 807, 2009.

[13] Y. Caiv and Y. Chen, “Mining influential bloggers: From general to domain specific,” in *KnowledgeBased and Intelligent Information and Engineering Systems*, 2009, vol. 5712 LNAI, pp. 447–454.

[14] C. Marlow, L. Byron, T. Lento, and I. Rosenn, “Maintained Relationships on Facebook,” 2009. [Online]. Available: http://overstated.net/2009/03/09/maintained-relationships-on-facebook.

[15] M. Cha, A. Mislove, and K. P. Gummadi, “A measurement-driven analysis of information propagation in the flickr social network,” *Proc. 18th Int. Conf. World wide web - WWW ’09*, p. 721, 2009.

[16] P.-N. Tan, M. Steinbach, and V. Kumar, *Introduction to Data Mining*. Addison-Wesley, 2005.