Analysis of the Information Value of User Connections for Video Recommendations in a Social Network

Toon De Pessemier *
WiCa-IBBT-Ghent University
G. Crommonlaan 8
9050 Ghent,Belgium
tdpessem@intec.ugent.be

Simon Dooms
WiCa-IBBT-Ghent University
G. Crommonlaan 8
9050 Ghent,Belgium
sdooms@intec.ugent.be

Joost Roelandts
NETLOG NV
Emile Braunplein 18
9000 Ghent,Belgium
joost@netlog.com

Luc Martens
WiCa-IBBT-Ghent University
G. Crommonlaan 8
9050 Ghent,Belgium
luc.martens@intec.ugent.be

ABSTRACT

The abundance of information and the related difficulty to discover interesting (video) content has complicated the selection process for end-users. Recommender systems try to assist in this content-selection process by using intelligent personalisation techniques which filter the information. However, most commonly-used recommendation algorithms, like collaborative filtering, are not optimized for social networks which contain valuable information about the user's friend connections and the structure of personal relationship networks. Therefore, this paper analyses the data set of a commercially-deployed social network and investigates the information value of user-to-user relations and video interaction behaviour in the user's friend network. The results prove that video selection in a social network is significantly influenced by the consumption behaviour in the personal network of the user. This information might be incorporated as an additional knowledge source into recommender systems, thereby improving the accuracy of the video suggestions. Moreover, the size of the user's social network has a significant positive correlation with the popularity of the user's uploaded videos. As a result, users having a large social network, i.e. be connected to a huge number of people, act as "hubs" of information. Video content uploaded or distributed by these users has a high visibility and acceptance rate on social networks.

Categories and Subject Descriptors

H.3.3 [**Information Storage and Retrieval**]: Information Search and Retrieval

*T. De Pessemier is a Fellow of the Fund for Scientific Research, Flanders (Belgium).

General Terms

Algorithms, Experimentation, Human Factors

Keywords

Social network, recommendation, personalized video content

1. INTRODUCTION

Fast-growing Web 2.0 sites (like YouTube, Flickr, Digg, etc.) have an overwhelming bulk of user-generated content available for their online consumers. Although this exploding content offer can be seen as a way to meet the specific demands and expectations of users, it has complicated the content selection process to the extent that users are overloaded with content and risk to "get lost": though an abundance of information is available, obtaining useful and relevant content is often difficult. Traditional filtering tools, like keyword-based or filtered searches, are not capable to filter out irrelevant content or provide too much search results. An additional filtering based on the overall popularity (expressed by user ratings or consumption patterns) can assist, but requires a broad basis of user feedback before it can make reasonable suggestions. Moreover, rankings based on the overall popularity do not consider personal preferences and individual consumption behaviour, thereby suggesting only the most popular content. This situation reinforces the role of collaborative filtering tools and stimulates the development of recommender systems that assist users in finding the most relevant content.

Besides the traditional photo- or video-sharing websites, people tend to use social networks (like Facebook) to share and distribute their personal pictures and videos. This entails a convergence of user-generated content providers and social networks, thereby collecting not only content interaction behaviour (like rating and viewing behaviour) but also social network data (like friend relationships). These additional social network data might be a valuable information source for recommenders to refine the personal profile of the end-user.

Therefore, this paper analyses the video interaction behaviour of users on such a popular social network together with the potential information value for recommender systems. This study was based on a data set of a large social network called Netlog.com. The remainder of this paper is organized as follows: Section 2 provides an overview of related work regarding recommender systems and social networks. Section 3 gives more insight into the structure and use of Netlog. Characteristic properties of the uploader that might effect the popularity of a video are analysed and discussed in Section 4. Section 5 elaborates on the influence of the user's friend network on the user's selection and interaction behaviour. Finally, we offer a brief conclusion on our research results and point out interesting future work in Section 6.

2. RELATED WORK

Traditionally, recommender systems have been categorized into two main classes: content-based methods and collaborative filtering techniques. Content-based or information filtering methods generate recommendations by matching a user's profile, or other user information, to descriptive product information [10]. These techniques construct a model of underlying user preferences from which personal recommendations are inferred. Examples include keyword filtering approaches and Bayesian network models [2].

In contrast to content-based methods, collaborative filtering techniques do not rely on descriptive information about the content. These techniques are based on the assumption that a good method to find interesting content is to search for other people who have similar interests, and then recommend items that those similar users liked in the past [2]. Early research about collaborative filtering systems has been conducted by GroupLens [13]. More advanced solutions like clustering models [15] and dependency network models [9] have been studied to improve the accuracy of the personal suggestions. In this context, Sarwar et al. proposed Singular Value Decomposition (SVD) to improve scalability of collaborative filtering systems by dimensionality reduction [14].

Content-based techniques do not consider the community knowledge [12] . In contrast, collaborative filtering tend to fail if little information is available about the user or the item (cold start problem), or if the user has uncommon interests. Therefore, hybrid content-based and collaborative recommenders have been explored to smooth out the disadvantages of each. These hybrid combinations have been studied in various domains like movie recommenders [7] and online newspapers [4].

Last years, various studies have been conducted to increase the accuracy of the recommendations which are calculated by user or item similarities based on implicit and explicit feedback. Conversely, O'Donovan and Smyth suggest that this traditional emphasis on similarity may be overstated [11]. They argue that additional factors, like the trustworthiness of users and network relations, have an important role to play in guiding recommendations. The underlying social network of the user has an added value to traditional feedback for user profiling and recommender systems. These network relations can be utilized to deduce trust inferences, transitive associations between users that denote the confidence of one user in another. Moreover, trust inferences can alleviate sparsity and cold-start problems for new users in the network [16]. Such trust relationships have been used by Golbeck and Hendler to personalize the user experience of FilmTrust, a website for movie recommendations [6]. Trust took on the role of a recommender system forming the core of an algorithm to create predictive rating recommendations for movies. The accuracy of their trust-based predicted ratings was found to be significantly better than the accuracy of a traditional recommender system.

Halvey and Keane have studied the social interactions and dynamics in the YouTube website [8]. They concluded that a large number of users do not use the facilities for social interaction available to them in media sharing services. However, people who do use the available tools have much a greater tendency to form social connections. As a result, media sharing services can also exploited user interactions in order to aid the user experience within these services.

Also Bonhard and Sasse suggest that recommender systems can be improved by combining the benefits of social networking applications with the matching capabilities of recommender systems [1]. They conducted several semi-structured interviews and focus groups to elicit concepts and priorities that are important in the decision making process when seeking advice. While it might seem common sense that people would consult their friends for recommendations for movies or music, participants clearly pointed out that the relation to the recommender alone is not sufficient. In taste domains, such as books, movies and TV-programs, people prefer recommendations for content that is consumed (and liked) by people they know. Nevertheless in these previous studies, these assumptions have not been verified on logged data records of an actual social network containing a large amount of user-generated content. Therefore, we investigated if the video interaction behaviour of users is influenced by the video interaction behaviour of their friends on the social network, based on a large data set with consumption records. Moreover, we studied the correlations between the popularity of a video and the characteristics of the uploader and her social network, such as her level of activity and the size of her social network.

3. NETLOG

The statistical analysis of this paper is based on the data of Netlog, a youth community where users can keep in touch with and extend their social network. Users can create their own profile page, upload pictures and videos, add friends to a personal social network, find events and play games. This research is focussed on the user behaviour regarding the videos and the interaction of the user's social network friends with these videos. Users can explore the videos by keyword-based searching, and browsing the lists of featured, newest, most viewed, most commented and top rated videos. Based on their personal social network on Netlog, users can also check the videos uploaded by their friends. Figure 1 shows a screenshot of the video page on the Netlog website.

The data set used for this analysis, a subset of the entire Netlog database, contains approximately 4.3 million registered users who have created a personal profile on the social network. These users have access to more than 2.8 million videos which are available to view and interact with. The types of interaction that are investigated in this study are watching a video, providing a rating, posting a comment, and tagging the video as "favourite", thereby adding it to a personal collection of preferred videos. For each video, the data set contains details about these different types of interaction. In total, 2.2 million comments, 1.3 million ratings, and 4.7 million favourite tags are used in the analysis. To

Table 1: General statistics about the user participation on the Netlog website.

Average Number of Friends per User	41.0557
Average Number of Comments per User	1.0713
Average Number of Ratings per User	0.6357
Average Number of Uploads per User	1.3635
Average Number of Favourite Videos per User	1.9217

correlate these video interactions to the video interactions of the user's friends, the bidirectional friend relationships of the network are used (85.5 million relationships). This large amount of user-to-user links emphasizes the high user connectivity of the social network.

Despite the Web 2.0 features in social networks and video sharing websites to encourage active user participation, the number of user interactions, like explicit ratings, remains very low. E.g. on YouTube, only 54% of all videos are rated and the aggregate ratings only account for 0.22% of the total views. Comments, a more active form of participation, account for mere 0.16% of total views [3]. Other video-sharing web sites have reported similar trends on relatively low user involvements [5]. In case of Netlog, this deficit of user participation regarding videos is strengthened by the fact that the main activity of users on a social network site is to connect to other users and create a friend network, rather than rating or commenting videos. As a result, only 10% of all videos on the Netlog website are rated, 12% of the videos received one or more comments, and 39% of the videos are at least once tagged as favourite. This limited interaction with videos is confirmed by Table 1, which provides some statistics about the user participation on the Netlog website. These data are obtained after filtering out dummy user profiles, i.e. user profiles that are never actually used.

Since recommender systems are typically based on rating behaviour of users to create user profiles, learn user preferences and calculate personal recommendations [13], this limited user participation may undermine the proper functioning of the recommender. Traditional recommendation algorithms are unable to produce accurate recommendations based on merely one comment, rating, upload, or favourite video per user. Implicit feedback, like viewing behaviour, might be used as alternative input for the recommender, but implies an additional uncertainty.

Therefore, the lack of sufficient explicit feedback on videos is an additional reason to use extra information sources, like the video interaction behaviour of friends, as input for recommender systems. Indeed, the number of friend relationships of a user is generally significantly higher than the amount of explicit feedback originating from that user, as indicated in Table 1. If the video interaction behaviour of the user is correlated to the video interaction behaviour of the user's friends, these video interactions of the user's friends might give information about the user's video preferences. Using this information, the input data for recommenders might be enlarged significantly, thereby producing more accurate recommendations. E.g. if an average user has 41 friends, who rated, commented or uploaded each one video, this might be enough data for a recommender to generate accurate recommendations for that user.

4. THE UPLOADER'S CHARACTERISTICS CORRELATED WITH THE VIDEO'S POPULARITY

Today, video sharing websites and social networks may contain an enormous amount of video content. Some of these uploaded videos become very popular in a very short period of time. However, the big majority of these videos will never reach a big mass of people. Obvious features, like the content and the audio-visual quality, mainly determine the popularity of online videos. Though in a social network, additional factors may influence the visibility of newly uploaded videos.

The influence of the social network on the distribution of a video is investigated by calculating the correlations between typical characteristics of the user who uploaded the video and the degree to which a video is "picked up" by the community. Distinctive characteristics of the uploader are the size of her personal social network (i.e. her number of friends) and her level of activity on the social network (i.e. the number of videos she has already uploaded on the social network). The degree to which a video is visible in the social network is measured by some general popularity characteristics (i.e. the number of ratings and comments that the video received by the community, the average of the ratings that the video received, the number of times a video is viewed, and the number of users that have marked the video as favourite). In addition, the number of video interactions of the uploader's friends are measured i.e. the number of friends who rated, commented or marked the video as favourite. Since ratings can be positive as well as negative, the positive rating behaviour of the uploader's friends is also considered separately. (Also comments can be positive as well as negative, but determining the connotation of comments requires linguistic text processing.) The correlations between these values, as shown in Table 2, provide the following interesting insights.

The size of the uploader's social network (i.e. the number of friends) has a significant positive correlation with all the popularity measures of the uploaded video. As a result, the videos uploaded by users with a large social network receive in general more attention (more views, more comments, and more favourites) and a better appreciation (more and higher ratings) than videos uploaded by users with a small network of friends. So a larger personal network of friends might increase the probability that uploaded videos become popular in the community. Indeed, videos of users who have more friends have more possibilities to receive ratings, comments, etc. from these first-order relationships. This is confirmed by the significant positive correlation between the uploader's number of friends, and the number of interactions of these friends on the uploaded video, i.e. the number of ratings (positive and negative), positive ratings, comments and favourite tags.

The second characteristic of the uploader that was investigated is the level of activity regarding video publishing (i.e. the number of videos she has already uploaded on the social network in the past). This number of uploads is negatively correlated with the number of friends that the uploader is linked to. In addition, the uploader's level of activity has a significant negative correlation with the general popularity of the uploaded video (as indicated in the last column of Table 2). In other words, videos of users who occasionally

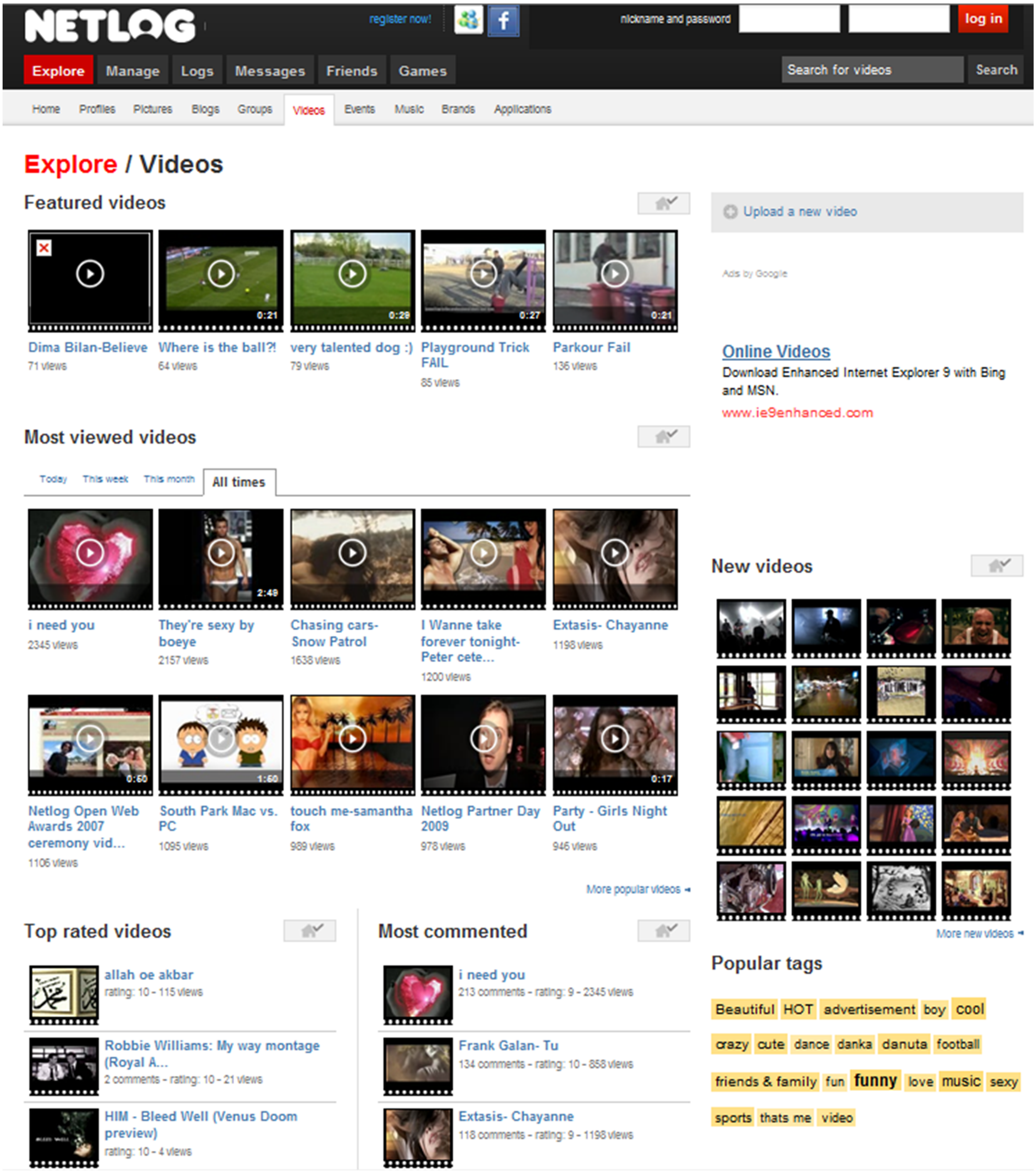

Figure 1: Screenshot of the video overview page on the website of Netlog

Table 2: Correlation between the general popularity of a video and the characteristics of the uploader

Correlation	Number of Friends	Number of Uploads
Number of Ratings of the Video	0.0238	-0.0510
Average Rating of the Video	0.0707	-0.1956
Number of Comments of the Video	0.0319	-0.0712
Number of Favourites of the Video	0.0240	-0.0782
Number of Views of the Video	0.0333	-0.0546
Number of Friends of the Uploader	1.0000	-0.2443
Number of Friends who Rated the Video	0.1393	-0.1423
Number of Friends who Positively Rated the Video	0.1360	-0.1406
Number of Friends who Commented the Video	0.1371	-0.1962
Number of Friends who Marked the Video as Favourite	0.1470	-0.1192

upload a video normally receive more attention than videos of uploaders who are very active in publishing new content. In addition, Table 2 shows a significant negative correlation (-0.2) between the average rating of the uploaded video and the number of videos that the user has already uploaded in the past. Thus, videos originating from active uploaders typically receive a lower appreciation from the community than videos originating from users with a limited number of uploaded videos. The reason for this might be a dilemma between quantity and quality: If users choose to upload more videos, the quality of these videos might decrease. The correlation between the user's number of uploads and the amount of video interactions of the uploader's friends is negative as well. Thus, users are less inclined to select, rate, or comment a video of a friend who constantly publishes new videos.

These findings may be used as extra knowledge for recommender systems to overcome the cold start problem (e.g. if only a limited number of ratings is available for a new video). The uploader's number of friends and number of past uploads might help to predict the future popularity of a newly uploaded video. Moreover, these characteristics of the uploader might be used as an indicator for predicting the interests and interaction behaviour of the uploader's friends, since these are significantly correlated.

5. STATISTICAL EVALUATION OF VIDEO INTERACTION BEHAVIOUR

To investigate the information value of social network relations for the recommendation of (audio-visual) content, we analysed the logging records of the user behaviour in the data set. First, the data considering the video interactions of each user were associated to the friend relationships of the user and the video interactions of these friends. Next, we investigated if video interaction behaviour of the user was preceded by an interaction of one of her friends on the same video.

The columns of Table 3 show the four types of video interaction behaviour that are analysed: providing a (positive or negative) rating for a video, providing a positive rating for a video, commenting a video, and adding the video to the personal collection of 'favourite videos'. The rows of Table 3 show the possible influence sources which might have triggered the user to interact with the video. The user might have encountered the video on one of the pages with popular videos: most rated videos, top rated videos, most commented videos, most favourite videos or most viewed videos. On the other hand, users might have selected a video to watch and interact with because of a link with a friend. This "influence of friends" is analysed based on the video interactions of the user's friends in the period before the user's video interaction. Therefore, we investigated if any of the user's friends has interacted on the same video, earlier.

The numbers in the upper five rows of Table 3 show the fraction of interactions that was committed on videos originating from the popular video lists. For example, the upper-left cell shows that 2.5% of the ratings provided by end-users evaluate a video from the "most rated" list. The bottom five rows of Table 3 indicate the fraction of the user's interactions that was proceeded by an interaction of one of the user's friends on the same video. For example, the bottom-right cell specifies that 13.1% of the videos marked as favourite by an end-user, were already tagged as a favourite video by one of her friends.

The upper five rows of Table 3 show that a considerable amount of the user's interactions (approximately 2%) happen on the lists of popular videos, which is a very small subset of the total set of videos (2.8 million). The bottom five rows of Table 3 indicate that even a greater amount of interactions (ranging from 3% till 20%) are preceded by interactions of the user's friends on the same video. The first column shows that approximately 9% of the user's ratings of a video are preceded by a rating of a friend for that video. Almost 13% of these rated videos received a comment from one of these friends earlier. In addition, more than 8% of the rated videos are uploaded by a friend of the user who provided the rating. And almost 15% of the rated videos are in the list of favourites of one of these friends.

This resemblance between the video interaction of the user and the video interaction of her friends is even more remarkable for providing comments. More than 20% of the user's comments is preceded by a comment of a friend on the same video. Besides, many comments happen on a video that is rated or uploaded by a friend, or on a friend's favourite video. Finally, we witness a link between favourite videos and interactions of friends on these videos: before a video is tagged as a favourite, it was in many cases rated, commented or uploaded by one of the user's friends. E.g., in 13% of the cases that a video is tagged as favourite by the user, it was already a favourite video of one of her friends.

Considering the number of videos in the data set (2.8 million), videos that experienced interaction of the user's friends have a much higher probability to be rated, commented or marked as favourite by the user than videos without interaction of these friends. This user behaviour con-

Table 3: The fraction of the user's ratings, positive ratings, comments and favourites on popular videos (first 5 rows) and the fraction of these interactions that are preceded by an interaction of a friend (last 5 rows)

	Rating	Positive Rating	Comment	Favourite
Most Rated List	0.0250	0.0253	0.0205	0.0141
Top Rated List	0.0249	0.0258	0.0182	0.0158
Most Commented List	0.0226	0.0225	0.0231	0.0114
Most Favourite List	0.0200	0.0208	0.0128	0.0194
Most Viewed List	0.0215	0.0217	0.0162	0.0121
Preceded by a Rating of a Friend	0.0909	0.0923	0.0866	0.0352
Preceded by a Positive Rating of a Friend	0.0855	0.0886	0.0814	0.0342
Preceded by a Comment of a Friend	0.1262	0.1275	0.2069	0.0501
Uploaded by a Friend	0.0850	0.0884	0.1377	0.0208
Marked as Favourite by a Friend	0.1492	0.1554	0.1269	0.1310

firms that end-users are interested in consuming content that is popular in their personal social network. So, besides personal preferences for the content itself, the content selection process might be driven by "interests in friends" or "curiosity". As a result, a user might like to have the videos that are popular with her friends incorporated in her personal video suggestions, thereby even further increasing the resemblance between the user's behaviour and her friends' behaviour. This way, media recommendations become a combination of both content related to the user's personal interests (according to the user's profile) and content related to the activities of the user's friends (according to the user's social network).

6. CONCLUSIONS

Analysis of a data set of interaction behaviour in a social network showed a significant positive correlation between the user's number of friends and the popularity of the user's uploaded videos. This correlation indicates that social network relations increase the visibility of the user's published content. As a result, highly-connected users may have a significant influence on the consumption behaviour of the community and may function as "hubs" of information on the social network: videos of highly-connected people have a high distribution potential via many directly-connected friends.

Moreover, this study investigated if users are influenced by their friends and the activities of these friends, while selecting, consuming and interacting with content. The resemblance between the user's video interactions and the video interactions of her friends indicates that a user is inclined to watch and interact with a video if one of her friend did the same earlier. This influence of the consumption behaviour of friends on the consumption behaviour of users may even cause a cascade of interactions on popular videos, thereby creating "viral videos" on a social network. Since users are attracted by content that is popular in their personal social network, video interactions of the user's friends may be used as an extra information source for recommender systems, thereby making the personal suggestions more social.

In future work we are planning to actually incorporate this extra knowledge in a video recommendation system for social networks. This way the user experience within these services can be improved based on the social interactions of the users. Moreover, additional characteristics of the user that might have an influence on their media-related behaviour, like age or gender, will be investigated.

7. REFERENCES

[1] P. Bonhard and M. Sasse. Knowing me, knowing you - using profiles and social networking to improve recommender systems. *BT Technology Journal*, 24:84–98, 2006.

[2] J. Breese, D. Heckerman, and C. Kadie. Empirical Analysis of Predictive Algorithms for Collaborative Filtering. In *Proceedings of the 14th Conference on Uncertainty in Artificial Intelligence*, pages 43–52, Madison, USA, July 1998.

[3] M. Cha, H. Kwak, P. Rodriguez, Y.-Y. Ahn, and S. Moon. I tube, you tube, everybody tubes: analyzing the world's largest user generated content video system. In *Proceedings of the 7th ACM SIGCOMM conference on Internet measurement*, IMC '07, pages 1–14, New York, NY, USA, 2007. ACM.

[4] M. Claypool, A. Gokhale, T. Miranda, P. Murnikov, D. Netes, and M. Sartin. Combining content-based and collaborative filters in an online newspaper. In *Proceedings of ACM SIGIR Workshop on Recommender Systems*, 1999.

[5] E. Auchard. Participation on Web 2.0 Sites Remains Weak, Available at http://www.reuters.com/article/internetNews /idUSN1743638820070418, 2007.

[6] J. Golbeck and J. Hendler. Filmtrust: movie recommendations using trust in web-based social networks. In *Consumer Communications and Networking Conference, 2006. CCNC 2006. 3rd IEEE*, volume 1, pages 282 – 286, jan. 2006.

[7] N. Good, J. B. Schafer, J. A. Konstan, A. Borchers, B. Sarwar, J. Herlocker, and J. Riedl. Combining collaborative filtering with personal agents for better recommendations. In *Proceedings of the Sixteenth National Conference on Artificial Intelligence*, pages 439–446, 1999.

[8] M. J. Halvey and M. T. Keane. Exploring social dynamics in online media sharing. In *Proceedings of the 16th international conference on World Wide Web*, WWW '07, pages 1273–1274, New York, NY, USA, 2007. ACM.

[9] D. Heckerman, D. M. Chickering, C. Meek, R. Rounthwaite, and C. Kadie. Dependency networks for inference, collaborative filtering, and data visualization. *J. Mach. Learn. Res.*, 1:49–75, September 2001.

[10] R. J. Mooney and L. Roy. Content-based book recommending using learning for text categorization. In *Proceedings of the fifth ACM conference on Digital libraries*, DL '00, pages 195–204, New York, NY, USA, 2000. ACM.

[11] J. O'Donovan and B. Smyth. Trust in recommender systems. In *Proceedings of the 10th international conference on Intelligent User Interfaces*, IUI '05, pages 167–174, New York, NY, USA, 2005. ACM.

[12] A. Popescul, L. H. Ungar, D. M. Pennock, and S. Lawrence. Probabilistic models for unified collaborative and content-based recommendation in sparse-data environments. In *Proceedings of the 17th Conference in Uncertainty in Artificial Intelligence*, UAI '01, pages 437–444, San Francisco, CA, USA, 2001. Morgan Kaufmann Publishers Inc.

[13] P. Resnick and H. R. Varian. Recommender systems. *Commun. ACM*, 40:56–58, March 1997.

[14] B. M. Sarwar, G. Karypis, J. A. Konstan, and J. T. Riedl. Application of dimensionality reduction in recommender system - a case study. 2000.

[15] L. Ungar and D. Foster. Clustering Methods For Collaborative Filtering. In *Proceedings of the Workshop on Recommendation Systems*, pages 114–129. AAAI Press, Menlo Park California, 1998.

[16] J. Weng, C. Miao, A. Goh, Z. Shen, and R. Gay. Trust-based agent community for collaborative recommendation. In *Proceedings of the 5th International Joint Conference on Autonomous Agents and Multiagent Systems*, pages 1260–1262, Hakodate, Japan, May 2006.