On the "Familiar Stranger" Phenomenon in a Large-scale VoD System

Chen Zhang*, Yuedong Xu*, Yipeng Zhou†, and Xiaoming Fu‡

*Research Center of Smart Networks and Systems, Fudan University, China
†School of Computer Science, University of South Australia, Australia
‡Institute of Computer Science, University of Goettingen, Germany
{zhang_chen12, ydxu}@fudan.edu.cn, yipeng.job@gmail.com, fu@cs.uni-goettingen.de

***Abstract*—Video-on-Demand (VoD) services are having more and more impact on people's daily life. The huge number of view records generated during the services reveal lots of information about video popularity and user behavior. In this work, we seek to understand user behavior from a network approach. Users who overlapped view records are defined as "familiar strangers" of each other. With network science concepts, dynamic familiar stranger networks are established based on view records from one of China's largest VoD service providers. Analysis shows that users can be separated into light users and heavy users. Community detection results show that there is a threshold of familiarity that drastically deteriorates the community structure of the familiar stranger network. Further investigation on the evolution of the communities suggests that there exists considerate number of dynamic communities with low members turnover. Our findings about the interest-based familiar stranger network unveils interesting networked user behavior patterns and provides meaningful guidelines for Internet applications like online user communities and recommendation systems.**

I. Introduction

VoD services are having more and more impact on people's daily life. According to Cisco, IP video traffic will be 82 percent of all consumer Internet traffic by 2020, up from 70 percent in 2015 [1]. They also predicted that Consumer VoD traffic will nearly double by 2020. With the enormous traffic, problems with regard to videos naturally draw both industrial and academical interest.

Some previous works paid much attention to video user behavior and experience. The authors of [2] presented one of the first measurement studies of large-scale VoD systems. They found that users access content in a pattern that follows a Poisson distribution. Authors of [3] did similar measurements on YouTube and found that video viewcount distribution exhibits a power-law pattern with truncated tails. They further proposed that video viewcount is predictable. In [4], the authors modelled the quality of experience (QoE) of wireless networks of a large-scale VoD system. They also observed that the viewing time of streaming users fits a hyper-exponential distribution quite well, implying that all viewers of longer videos and shorter ones behave differently.

Previous works have also been dedicated to video popularity prediction. In [5], the authors examined the relation between viewcount and other video-related metrics such as likes and comments. They discovered significant positive correlation between viewcount and metrics like # of comments, # of favors and # of ratings. Many linear-regression-based models have then been introduced to predict video popularity. Social network data have also been brought in to help better understand the variation in video popularity. In [6], researchers merged data from YouTube and Twitter and obtained effective cross-network predictors on sudden bursts of Twitter-driven YouTube views and steadily growing ones. Authors of [7] pointed out that popularity dynamics of OSN-driven views are non-linear. Therefore, instead of linear models, they applied an epidemic model based on data from Renren, the largest OSN in China at that time, and achieved amazingly accurate prediction results.

Generally speaking, the interaction between users and videos can be depicted as a bipartite graph. Users establish links to videos by uploading or interacting with existing videos, such as watching, liking, marking, and making comments. The two perspectives mentioned in the two previous paragraphs only counts local information of video nodes and user nodes. A different way to look at the bipartite graph is to project the graph onto one of the two disjoint sets. There have been, however, very few discussions on the projections of the bipartite graph. The authors of [8] discussed the video response features of YouTube. They regarded YouTube as a social network consisting of videos and corresponding uploaders. They altered that opportunistic behavior among video uploaders may jeopardise user experience.

To the best of our knowledge, there is no published work on the relationship among video service users. In fact, the video service users themselves form an Online Social Network. The volumes as well as other features of the overlap of a pair of users' view history marks the similarity of their preference for videos. The similarity indicates a virtual connection between the pair of users. We define users with adequate similarity as the "*familiar strangers*" of each other. The notion of "familiar stranger" is originally a social psychological term, referring to an individual who is recognized from regular activities, but with whom one does not interact [9]. The original definition emphasize the spacial similarity. In the context of the virtual environment of the cyber space, users meet each other at videos they all watched. Therefore, the spacial similarity is replaced by the similarity of content. The similarity of content is less intuitive, but more profound, since it is a much better indicator of the similarity in interest, hobby, personality, etc. Therefore, the relation according to the similarity of content and preference is worthier of the name of "familiar strangers."

In this study, we aim to understand user behavior from a network approach. We construct a familiar stranger network that evolves within time using our dataset provided by one of the largest VoD service providers in China. We find that there is an obvious difference between heavy users and light users. Cluster detection results show that the network has good community structure under certain conditions. Long-lived dynamic communities are discovered, indicating that there are many groups of people whose preference for videos remain similar for long. Our findings about the interest-based familiar stranger network provides meaningful guidelines for the maintenance of online user communities, which may lead to better recommendation systems.

The rest of the paper is organized as follows. In Section II, the dataset used in the study is introduced and washed. The dynamic familiar stranger network is constructed in Section III. In Section IV, community detection algorithm and dynamic community tracing algorithm are applied to the familiar stranger network, and the outcomes are discussed in detail. We conclude in Section V.

II. Dataset

A. Overview

Unlike those used in most of the other previous works, which were crawled by the researcher themselves, our dataset is provided by one of the largest Video-on-Demand (VoD) service providers in China. We have full access to all records of video access from mobile devices running Android, Apple OS and Symbian between September/1/2014 and September/14/2014. There are more than 200 million of view records everyday. Each entry of the view records contains a set of information as follows:

$<$*Timestamp, VideoID, VideoType, UserID*, $\ldots>$.

Here, *Timestamp* marks the time instant that a streaming session ends. *VideoType* identifies the content type, i.e. TV episode, music, news, animation, variety show, and amusement. *VideoID* and *UserID* are encrypted so that we cannot recover the exact meta information of watched videos. After invalid view records such as repeated watching in a very short interval have been cleansed, the dataset contains more than 2.5 billion entries, covering around 100 million individual users and 1.4 million video clips.

We focus only on the user graph of TV episodes in this study. In this study, the term of "video" only refers to a TV episode unless explicitly mentioned. We do so mainly because of the fact that, TV episodes generates a dense user network. (In our dataset, TV episodes, making up only around 5% of the total volume of the whole video set, astonishingly attracted 35% of view counts.) The commercial value and social impact of TV content are also good reasons for us to focus on TV episodes. Fig. 1 shows the access frequency of TV-typed videos in our dataset.

B. Distribution of Viewcounts

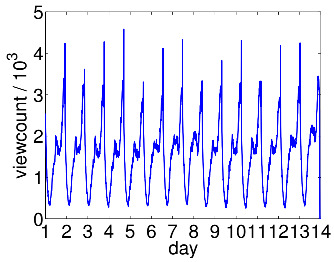


Fig. 1. An overview of frequency of access to TV-typed videos. The frequency of access is cumulated into 10-minute intervals over the dataset.

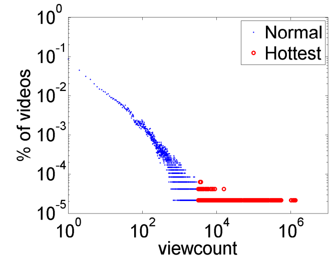
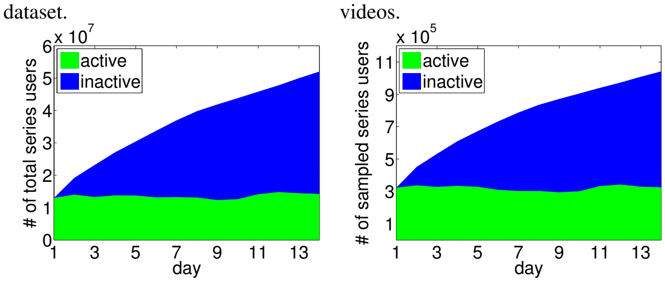


Fig. 2. View count distribution on Day 1. The blue dots stands for videos that fit well with power law distributions, while the red circles refer to the hottest videos.



(a) (b)

Fig. 3. The number of TV users and its development in the user graph. Fig. 3a involves all users, while Fig. 3b depicts the sampled users.

Popularity, mainly measured by viewcount, varies drastically among TV episodes. The distribution of view counts on day 1 (September/1/2014) of all TV-typed videos is shown in Fig. 2. Empirically, human activity usually follows a power-law distribution. For the videos with relatively smaller view counts, the popularity curve can be well captured by a power-law distribution, as illustrated by blue dots. However, for videos with extremely high view counts, the popularity distribution exhibits a heavy tail, as illustrated by red circles. We assume that extremely popular videos cannot serve well as an indicator of each individual's personal interest, since almost everyone watched them. We therefore consider excluding these videos. We used a fitting approach. When fitting the distribution of view counts, we compute coefficient of determination (R^2) to measure the goodness of fitting. An intuitive observation is that after the most popular videos have been removed, the view counts of the remaining videos can be better captured by the power-law distribution, and hence yielding a larger R^2.

After the pruning, around 2000 TV clips are excluded. We end up with more than 50 thousand clips.

C. Volume of Users

There are more than 50 million users involved in the complete user graph of TV-typed videos. Because of some technical reasons, in this article, we base our discussion on a user graph of 2.5% of all users randomly sampled (more than 1 million users). The sampling has been carried out multiple times, and the features of the user graphs are almost the same.

Fig. 3 demonstrates the number of daily active users and the cumulative number of users in the dataset.

III. The Familiar Stranger Network

In this section, we focus on user behavior from a network approach. We will define the user graph, and construct the familiar stranger network based on the user graph.

A. Definition

Real-world social networks are usually evolving. Edges in real-world social networks change within time. They can be established, maintained, or destroyed at a certain time. In order to bring in the evolution mechanism to the familiar stranger network, we introduce a sliding window. For the familiar stranger network on day T, we only consider the view records within the last W days from T, which are covered by the sliding window that starts from day $\max(1, T - W + 1)$ and ends on T. The edges caused by view records outside the window are "forgotten".

Let $\mathcal{U}$ be the set of users, and $\mathcal{V}(u_i, t)$ be the set of videos that user $u_i \in \mathcal{U}$ watch on day t. The familiarity between users u_i and u_j on day t, denoted by $w(u_i, u_j, t)$, can be measured by the volume of the intersection of the set of videos u_i and u_j watched within W days from day t.

$$w(u_i, u_j, t) = \begin{cases} \sum_{\alpha,\beta\in[t',t]} |\mathcal{V}(u_i,\alpha) \cap \mathcal{V}(u_j,\beta)| & , \quad i \neq j \\ 0 & , \quad i = j \end{cases} , \tag{1}$$

where $t' = \max(1, t - W + 1)$. $w(u_i, u_j, t)$ is equivalent to the weight of the edge between u_i and u_j in **user graph**, which is the projection of the video-users bipartite graph onto the user nodes.

In order to move on to the **familiar stranger network**, we further define $isFamiliar(u_i, u_j)$ as a boolean indicator marking whether u_i and u_j are familiar strangers of each other.

$$isFamiliar(u_i, u_j, t) = \begin{cases} 1 & , \quad w(u_i, u_j, t) > \eta \\ 0 & , \quad w(u_i, u_j, t) \leq \eta \end{cases} , \tag{2}$$

where η is the threshold. η has great impact on the structure of the graph, which will be shown in Section IV. But when looking at the degree and strength distributions, η doesn't make a large difference. In this section, we choose $\eta = 1$.

The familiarity of u_i and u_j in the familiar stranger network, which is equivalent to the weight of the edge between u_i and u_j in the familiar stranger network, is defined as

$$w_e(u_i, u_j, t) = \max[0, w(u_i, u_j, t) - \eta]. \tag{3}$$

In weighed networks, the degree and strength of nodes are two basic metrics. In this work, the degree of u_i in the familiar stranger network is denoted as

$$dgr(u_i, t) = \sum_{\forall u_j \in \mathcal{U}} isFamiliar(u_i, u_j, t). \tag{4}$$

Similarly, the strength of u_i in the familiar stranger network is denoted as

$$str(u_i, t) = \sum_{\forall u_j \in \mathcal{U}} w_e(u_i, u_j, t). \tag{5}$$

B. The Dynamic Familiar Stranger Network

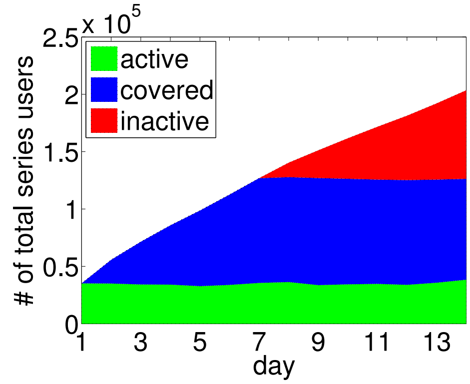


Fig. 4. An illustration of the number of users covered by the sliding window.

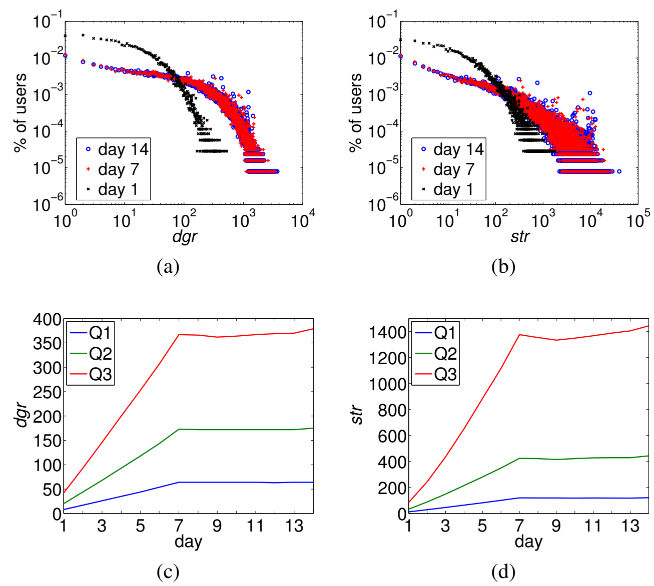


Fig. 5. An illustration of the dgr and str distributions under a sliding window. Fig. 5a and Fig. 5b show the distributions on Day 1, 7, and 14. Fig. 5c and Fig. 5d showed the evolution of the $Q1$, $Q2$ and $Q3$ of the dgr and str distributions respectively.

For a dynamic familiar stranger network, it will take as long as the length of the sliding window for the network to get fully established. Prior to its establishment, the network undergoes expansion. Since our dataset spans over 14 days, we find $W = 7$ to be a reasonable choice to trace the change in the network during both stages before and after establishment.

The volume of the familiar stranger network is shown in Fig. 4. Everyday, about 35 thousand users in the network accessed videos. When the network is fully established, there are about 130 thousand users covered by the sliding window. A total number of more than 200 thousand users had shown up in the familiar stranger network. This shows that the network has a considerate overturn.

Fig. 5 demonstrates the evolution of the dgr and str of the dynamic familiar stranger network. Fig. 5a and Fig. 5b show the dgr and str distributions on Day 1, Day 7, and Day 14, which are respectively the first day, the day when the familiar stranger network is fully established for the first time, and the last day. All distributions appear to be synthesized by two different power-law distributions. This feature will be discussed in Section IV. In the figures, an obvious expansion can be seen from Day 1 to Day 7. However, the distributions of Day 7 perfectly overlap with those of Day 14.

In order to trace the evolution among the entire 14-day, especially from day 7 to 14, we use the percentiles to depict the general distribution features of the graph. Fig. 5c and Fig. 5d demonstrates the evolution of dgr and str respectively, using the the $25th$, $50th$ and $75th$ percentile, corresponding to $Q1$, $Q2$ and $Q3$ respectively. Both dgr and str experience a linear-like growth during the establishment of the familiar stranger network. After the graph is fully established, both dgr and str remain stable.

It is reasonable to say that, once established, the distribution and density of the familiar stranger network is quite stable. The stability of the graph lays the foundation of potential applications, such as community detection, which will be discussed in Section IV.

IV. Community Detection

In This section, the community structures in the dynamic familiar stranger network established in Section III-B is discussed. Community structures are very common in such scale-free networks and networks related to human behavior [10]. We have reasons to believe that in the familiar stranger network, significant community structures also exists. As a first step into the "familiar stranger" phenomenon in the user graph, we treat the graph as unweighed.

A. The Impact of the Threshold Variable η

In Section III-A, it is mentioned that η has a large impact on the community structure of the familiar stranger network. In this section, we use the popular modularity-based community detection algorithms [12], [14] to discover the community structures and their evolution. Aside from modularity, other metrics such as the size of the graph and its connectivity are also important aspects. For those familiar stranger networks with poor connectivity, the familiar stranger networks do not make much sense, since the communities are broken into disjoint subgraphs, there is very limited diversity in users' taste. This is certainly not what happens in the real world. On the other hand, if the graph has a big connected component, the users may have wide interest in videos, and the user of a video may have very different backgrounds. In this case, the communities interact well with others.

Fig. 6 shows the modularity, the coverage (the percentage of users in the largest connected component, abbr. *LCP*), the size of the *LCP*, and the number of communities detected in the *LCP* in the dynamic familiar stranger network on Day 14, which is the last day in the dataset. Generally, the coverage and size of the *LCP* declines with the growth of threshold variable η. In Fig. 6a, however, the coverage experience an abnormal increase around $\eta = 40$ and $\eta = 50$, but the coverage is already below 20%, therefore the communities are too small to draw any conclusion. Interestingly, there seems to be an inflection point around $\eta = 35$. At around $\eta = 30$, the modularity is maximized. However, on surpassing $\eta = 35$, the modularity deteriorates drastically. Similar pattern is observed in the coverage of the *LCP*. The coverage declines smoothly when $\eta < 30$, while when $\eta > 30$, the coverage drops hastily to below 20%.

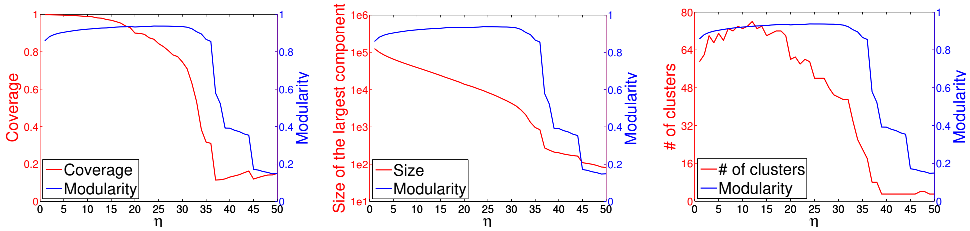
Similarly, when $\eta < 30$, the number of users in the *LCP* (Fig. 6b) decreases exponentially with a stable exponent. On surpassing $\eta = 30$, the size of the *LCP* experience a much more drastic drop. It only recovers the smooth decline after $\eta = 36$. The number of clusters detected in the *LCP* (Fig. 6c) remains relatively stable on a high level when η is low ($[1, 20]$). As η grows ($[20, 30]$), the number smoothly declined a little. When η grows larger than 30, the number of clusters hastily drops by more than 90%.

Given that η is used to measure if a pair of users are familiar enough to be familiar strangers (See the definition in (2)), the magic threshold around $\eta = 35$ in Fig. 6 leads to a conclusion that in the familiar stranger network, edges with weight larger than 35 are topologically very different from the rest. This observation of the inflection point accords with the distribution patterns previously observed in the users' dgr (Fig. 5a) and str (Fig. 5b). In order to verify that, the weight distribution in the dynamic familiar stranger network on Day 14, which is the same day as in Fig. 6, is shown in Fig. 8.

In Fig. 8, it is obvious that the link weight distribution curve is synthesized by two power-law distribution curves with very different exponents. Together, Fig. 6 and Fig. 8 suggest that very few people has connections with weight larger than 35. By purging the weak links hence eliminating the light users, the fact unveils itself that users heavily involved in the familiar stranger network are not tightly organized. This indicates that they enjoy a high degree of diversity in the means of preference for videos. (Note that when $\eta > 40$, only less than 5 communities can be detected from the largest connected component. In this case, members of a community must have considerate number of cross-community connections, so that the modularity remains at a low level.) More or less, these users have some interest in common with each other. But their distinctive preference for video is diversified enough to fall into, or even represent different clusters.

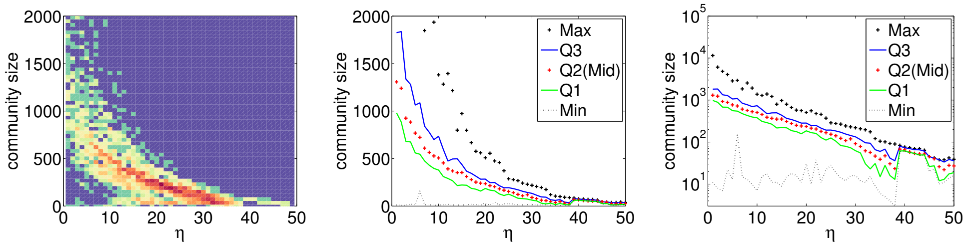
The size of the communities detected in the graph is also an important feature. Fig. 7 shows the evolution of the size distribution of the communities when η changes. Fig. 7a is the heat map of $(\eta, community\ size)$ pairs. Each small square in the figure represents the number of communities that fall into the corresponding size interval detected in the graph generated with the corresponding threshold η. A color closer to red represents a higher number. Generally, the distribution shrinks exponentially with the growth of η. For a fixed η, the size distribution seems to be following a normal distribution, being symmetric with regard to the centroid, which is the point with the darkest color.

In Fig. 7b, the $100th$, $75th$, $50th$, $25th$ and $0th$ percentiles are used to depict the evolution of the size distribution. Every



(a) modularity v.s. coverage of *LCP* (b) modularity v.s. size of *LCP* (c) modularity v.s. # of clusters

Fig. 6. The Newman-Girvan modularity v.s. the coverage (Fig. 6a), size (Fig. 6b), and the number of detected clusters (Fig. 6c) of the largest connect component of the dynamic familiar stranger network on Day 14 under different η.



(a) (b) (c)

Fig. 7. The distribution of community size in the largest connected component of the dynamic familiar stranger network of Day 14. Fig. 7a is the heat map of $(\eta, community\ size)$ pairs. Fig. 7b shows the evolution of the $100th$, $75th$, $50th$, $25th$ and $0th$ percentiles under different η in linear scale. Fig. 7c is the log scale version of Fig. 7b.

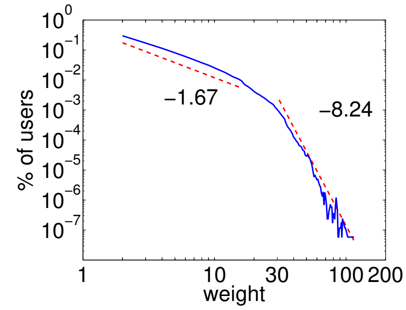


Fig. 8. The weight distribution of the edges in the dynamic familiar stranger network on Day 14. It is obvious from the figure that the distribution curve is synthesized by two power-law distribution curves of different exponent. The exponents are mark near the shifted asymptote.

one of the five measurements declines exponentially, and the $75th$ and the $25th$ percentiles are symmetric with regard to the median ($50th$ percentile). The symmetry of the distribution suggests that the size distribution is very likely to be following a normal distribution, instead of the power-law distribution suggested by many other works. [11], [13] This is, however, a lovely feature. Tiny communities are harder to make use of. And normal distribution indicates much less tiny communities than power-law distribution does.

In order to better understand the decline pattern of the size distribution parameters, Fig. 7c is presented in log scale. Interestingly, the $100th$, $75th$, $50th$, and $25th$ percentiles share the same slope, indicating the four parameters decay at a unanimous rate.

B. The Evolution of Communities

Typically, a dynamic community undergoes stages such as birth, merging, splitting, expansion, contraction, and death [14]. Dynamic communities active for long indicate that their members remain highly familiar with each other for a long period of time. The existence of long-lived dynamic communities is essential to applications such as online user communities, recommendation systems, and content prefetching.

In this study, we apply the dynamic community tracing method in [14]. The Jaccard Coefficient is used to measure the similarity of the communities. $sim(\mathcal{A},\mathcal{B}) = \frac{|\mathcal{A}\cap\mathcal{B}|}{|\mathcal{A}\cup\mathcal{B}|}$. A threshold θ is used to determine the relationship between two clusters from two consecutive days. In [14], the authors suggested that $\theta = 0.3$ be a moderate threshold value for Jaccard Coefficient. However, in [14], the dataset did not have a significant overturn. While in our study, the user set have a considerate overturn. Fig. 9 shows the Jaccard Coefficient of the *LCP* in the dynamic familiar stranger network between consecutive days. It can be told that the overturn of the familiar stranger network is relatively big. Therefore, $\theta = 0.3$ would be too strict in our dataset. In this study, $\theta \in \{0.1, 0.2, 0.3\}$ is tried. Also, because we are more interested in long-lived communities, the dynamic communities are categorized by their lifetime. In our dataset, the maximum lifetime is 8 days, so the dynamic communities that have been active for more than or equals 7 days are considered long-lived. Those with lifetime shorter or equals 2 days are considered short-lived.

Fig. 10 shows the absolute number and proportion of the dynamic communities of the three categories under different η. When $\theta = 0.1$, long-lived dynamic communities comprises of more than 80% of all the detected dynamic communities when $\eta < 30$, as is shown in Fig. 10d. However, it is noticed that the total number of communities in each day's familiar stranger network is less than 80 (Fig. 6c), which is astonishingly less than the number of dynamic communities in Fig. 10a. In this case, the dynamic communities are too small to make sense.

When $\theta = 0.3$, which is the recommended value in [14], not much long-lived dynamic communities are detected (Fig. 10c). And the long-lived communities do not make up of a majority of all detected communities (Fig. 10f). We draw the conclusion that due to the high overturn in our dataset, $\theta = 0.3$ is a relatively strict threshold.

When $\theta = 0.2$, the number of long-lived dynamic communities (Fig. 10b) accords well with the total number of communities in each day's familiar stranger network (Fig. 6c). And the proportion of the long-lived dynamic communities are relatively high. Long-lived dynamic communities have low overturn, which is defined by the tracing algorithm. That being said, users in long-lived communities maintain the familiar

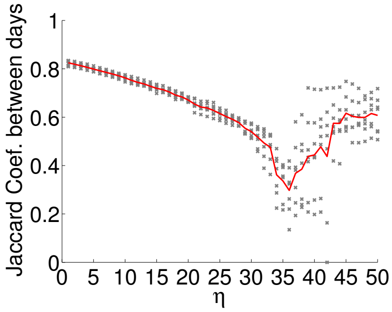


Fig. 9. The Jaccard Coefficient of the largest connected component in the dynamic familiar stranger network. The grey dots with x coordinate η marks the 7 Jaccard Coefficients between the 8 snapshots of the network with familiarity threshold η. The red line depicts the average of the Jaccard Coefficients of networks with different η.

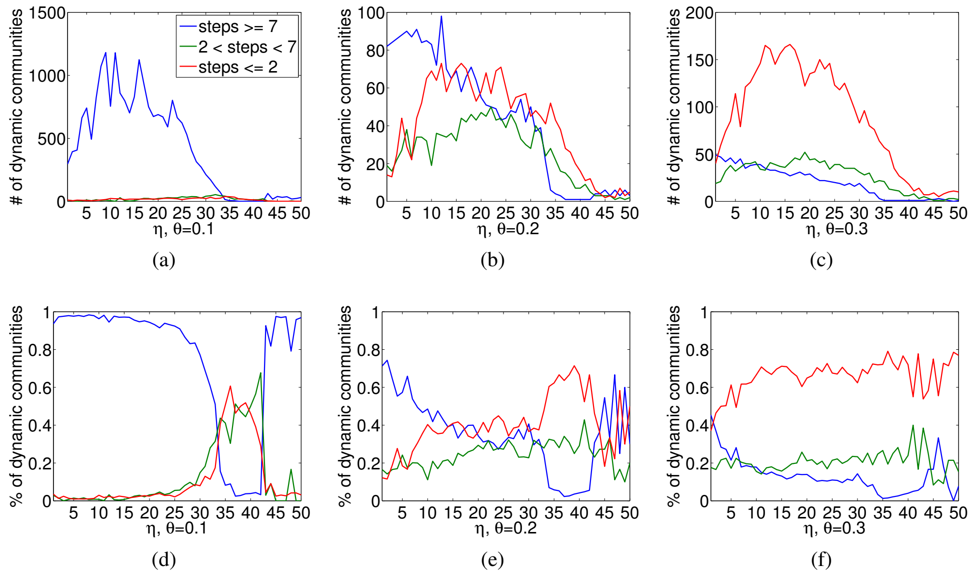


Fig. 10. The absolute number (Fig. 10a-10c) and proportion (Fig. 10d-10f) of the dynamic communities of the three categories under different η. The θ's are marked in the $xlabel$. The blue lines correspond to long-lived dynamic communities, while the red lines mark the short-lived ones. The green lines represent those with medium lifetime.

stranger relations with others for a long time, which means their preference for videos remain similar over time.

We conclude that under appropriate criteria, we can detect dynamic communities that evolves for a long time.

V. Conclusion

In this work, we make the first attempt to understand the user behavior of a VoD system from a network approach. Based on the user graph extracted from one of the largest VoD systems in China, a familiar stranger network that depicts the users' similarity in preference for TV episodes is constructed. The familiar stranger network resembles the combination of two distinct scale-free networks, suggesting that there be very few heavy users and quite a lot of light users. Community detection results with varying familiarity threshold η unveils the structure of the familiar stranger network. More concretely, the light users are centered around the heavy users, most of whom have different preference for videos from each other. It is also discovered that under appropriate criteria, we can detect dynamic communities that remain active for a long period of time, indicating stable familiar stranger relationship between community members. The features of the familiar stranger network could be very useful for Internet applications like online user community, recommendation systems, and content prefetching.

Acknowledgement: This work is partly supported by Natural Science Foundation of China (No. 61402114).

References

[1] Cisco. Cisco visual networking index: Forecast and methodology, 2015-2020. http://www.cisco.com/c/dam/en/us/solutions/collateral/service-provider/visual-networking-index-vni/complete-white-paper-c11-481360.pdf, 2016.

[2] H. Yu, D. Zheng, B. Y. Zhao, and W. Zheng. Understanding user behavior in large-scale video-on-demand systems. In *Proceedings of ACM SIGOPS/EuroSys 2006*, EuroSys '06, pages 333–344, New York, NY, USA, 2006. ACM.

[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 ACM SIGCOMM IMC'07*, pages 1–14. ACM, 2007.

[4] Z. Xiao, Y. Xu, H. Feng, T. Yang, B. Hu, and Y. Zhou. Modeling streaming qoe in wireless networks with large-scale measurement of user behavior. In *IEEE GLOBECOM'15*, pages 1–6. IEEE, 2015.

[5] G. Chatzopoulou, C. Sheng, and M. Faloutsos. A first step towards understanding popularity in youtube. In *IEEE INFOCOM Workshop'10*, pages 1–6. IEEE, 2010.

[6] H. Yu, L. Xie, and S. Sanner. Twitter-driven youtube views: Beyond individual influencers. In *Proceedings of ACM MM'14*, pages 869–872. ACM, 2014.

[7] H. Li, X. Ma, F. Wang, J. Liu, and K. Xu. On popularity prediction of videos shared in online social networks. In *Proceedings of ACM CIKM'13*, pages 169–178. ACM, 2013.

[8] F. Benevenuto, T. Rodrigues, V. Almeida, J. Almeida, and K. Ross. Video interactions in online video social networks. *ACM Transactions on Multimedia Computing, Communications, and Applications (TOMM)*, 5(4):30, 2009.

[9] S. Milgram, J. E. Sabini, and M. E. Silver. *The individual in a social world: Essays and experiments*. Mcgraw-Hill Book Company, 1992.

[10] S. Fortunato. Community detection in graphs. *Physics reports*, 486(3):75–174, 2010.

[11] A. Arenas, L. Danon, A. Diaz-Guilera, P. M. Gleiser, and R. Guimera. Community analysis in social networks. *The European Physical Journal B-Condensed Matter and Complex Systems*, 38(2):373–380, 2004.

[12] V. D. Blondel, J.-L. Guillaume, R. Lambiotte, and E. Lefebvre. Fast unfolding of communities in large networks. *Journal of statistical mechanics: theory and experiment*, 2008(10):P10008, 2008.

[13] M. E. Newman and M. Girvan. Finding and evaluating community structure in networks. *Physical review E*, 69(2):026113, 2004.

[14] D. Greene, D. Doyle, and P. Cunningham. Tracking the evolution of communities in dynamic social networks. In *Advances in social networks analysis and mining (ASONAM), 2010 international conference on*, pages 176–183. IEEE, 2010.