# 18 Social Media Sharing

Social media, a group of Internet-based applications that build on the ideological and technological foundations of Web 2.0, allow the creation and exchange of user-generated contents [1]. These services combine rich Graphical User Interfaces (GUI) and multimedia contents, and enable ubiquitous content generation and sharing and scalable communication. They have substantially changed the way organizations, communities, and individuals communicate. Two distinct characteristics of Web 2.0 are considered as key factors to the success of the new generation of social media:

**Collective intelligence.** The base knowledge of general users of the Web is contributing to the content of the Web. The users have become the content providers and therefore the contents are richer and more dynamic.

**Rich connections and activities.** The users and their contents are linked to each other, creating an organic growth of connections and activities. A social network built out of the connections enables strong ties amongst the users and broad and rapid content propagation.

There are many types of social media services, including user-generated content sharing (e.g., YouTube), online social networking (e.g., Facebook), question-and-answer (e.g., Ask), and collaboratively edited encyclopedia (e.g., Wikipedia), to name but a few. In these social media services, the contents, as well as the users, have become interconnected, enabling convenient information *sharing* for feelings, activities, and location information, as well as resources, blogs, photos, and videos.

With the pervasive penetration of wireless mobile networks, the advanced development of smartphones and tablets, and the massive market of mobile applications, social media contents can now be easily generated and accessed at any time and anywhere. YouTube's own statistics have reported that YouTube mobile gets over 600 million views a day, making up almost 40 % of YouTube's global watch time. The new trends on social media creation, deployment, and spreading, far beyond conventional media, have brought up numerous well-known Internet memes and celebrities that have no doubt been changing our daily life.

In this chapter, we present an overview of this rapidly evolving field, particularly for social media services for multimedia content sharing. We first overview two important social media services, namely user-generated media content sharing and

Z.-N. Li et al., *Fundamentals of Multimedia*,
Texts in Computer Science, DOI: 10.1007/978-3-319-05290-8_18,

online social networking. We closely examine the YouTube video sharing service. We then discuss media object propagation in social networks. Finally, we discuss user behaviors in sharing and the associated optimization.

## 18.1 Representative Social Media Services

We now present the background for two important social media services and their representative implementations.

### 18.1.1 User-Generated Content Sharing

Having arisen in web publishing and new media content production circles, *User-generated content* (UGC) plays a key role in today's social media services. It is used for a wide range of applications with different types of media, e.g., text, music, picture, and video, as well as a combination of open source, free software, and flexible licensing or related agreements to further reduce the barriers to collaboration, skill-building, and discovery. For content generation and sharing, video data are arguably more difficult than such other types of media as text and pictures, given their large size, high bandwidth demand, and long playback duration.

In traditional video on-demand and live streaming services, videos are offered by enterprise content providers, stored in servers, and then streamed to users. A number of new generation video sharing websites, represented here by YouTube, offer users opportunities to make their videos directly accessible by others, by such simple operations in Web 2.0 as embedding and sharing.

Established in 2005, YouTube is so far the most significant and successful video sharing website. It allows registered users to upload videos, mostly short videos. The users can watch, embed, share, and engage with videos easily. As one of the fastest growing websites in the Internet, YouTube had served 100 million videos per day in 2006; yet by December 2013, more than 1 billion unique users visit YouTube each month, over 6 billion h of video are watched each month—that is almost an hour for every person on Earth, and 100 h of new videos are uploaded every minute. YouTube is also highly globalized—it is localized in 61 countries and across 61 languages, and 80 % of YouTube traffic comes from outside the US. The success of similar sites (e.g., Vimeo, Youku, and Tudou) further confirms the mass market interest in UGC video sharing services.

### 18.1.2 Online Social Networking

Online social networking services provide an Internet-based platform to connect people with social relations, e.g., friends, classmates, and colleagues in the real world, or people who simply share common interests.

Facebook, founded in 2004, is one of the dominating online social networking services on the Internet. As of November 2013, Facebook had 1.19 billion active users worldwide and 728 million of them log onto Facebook on a daily basis. It provides users with a platform to connect with friends, by updating status, uploading photos, commenting and "liking" other's posts, etc. Currently, there are 4.5 billion likes and comments generated each day, together with 300 million photos uploaded. Facebook opens its API for developers to build thousands of applications and games, which makes it more enjoyable.

Another important social networking website, Twitter, is a representative of microblog, a simpler but much faster version of blog. It allows users to send text-based posts, called `tweets`, of up to 140 characters. Although short, the tweets can link to richer contents such as images and videos. By following friends or interested accounts, such as news providers, celebrities, brands, and organizations, Twitter users can obtain real-time notifications, and spread the posts by a `retweet` mechanism. Like Facebook, Twitter also opens its APIs, and there is a large collection of registered Twitter applications available, particularly for mobile users, making Twitter easier to access.

Both Facebook and Twitter support the sharing and propagation of such media objects as pictures, music, and video among friends, although the media content may be hosted by external sites. Recently, Twitter has also begun offering the Vine service, which, available exclusively for mobile users, enables them to create and post video clips. A Vine user can create a short video clip up to six seconds long while recording through Vine's in-app camera. The clip can then be shared through Twitter, Facebook, or other social networking services.

## 18.2 User-Generated Media Content Sharing

Social media has greatly changed mechanisms of content generation and access, and also brings unprecedented challenges to server and network management. Understanding the features of social media services is thus crucial to traffic engineering and to the sustainable development of these new generation of services. We now provide an overview of their unique features, using YouTube as a representative.

### 18.2.1 YouTube Video Format and Meta-data

YouTube's video playback technology is based on Adobe's Flash Player, which allows YouTube to display videos with quality comparable to well-established video playback technologies (such as Windows Media Player, QuickTime, and Realplayer). YouTube accepts uploaded videos in many formats, which are converted to the .FLV (Adobe Flash Video) format after uploading. It is well recognized that the use of a uniform and easily playable format is critical to the success of YouTube. YouTube used the H.263 video codec earlier, and introduced "high quality" format with the

**Table 18.1** An example of the meta-data of a YouTube video

| ID | YiQu4gpoa6k |
|---|---|
| Uploader | NewAgeEnlightenment |
| Date added | August 08, 2008 |
| Category | Sports |
| Video length | 270 s |
| Number of views | 924,691 |
| Number of ratings | 1,039 |
| Number of comments | 212 |
| Related videos | ri1h2_jrVjU, 0JdQlaQpOuU, … |

H.264 codec for better viewing quality in late 2008 (See Fig. 16.2 in Chap. 16 for a complete list of the audio/video formats offered by YouTube to date).

YouTube assigns each video a distinct 11-digit ID composed of the characters 0–9, a–z, A–Z, -, and _. Each video contains the following intuitive meta-data: $video\ ID$, $uploader$, $date\ added$, $category$, $length$, $number\ of\ views$, $number\ of\ ratings$, $number\ of\ comments$, and a list of $related\ videos$. The related videos are linked to other videos that have similar titles, descriptions, or tags, all of which are chosen by the uploader. A YouTube page only shows at most 20 related videos at once, but more can be displayed by scrolling down the list. A typical example of the meta-data is shown in Table 18.1.

### 18.2.2 Characteristics of YouTube Video

There have been significant research efforts aimed at understanding the workloads of traditional media servers, for example, video popularity and access locality [2,4,5]. While sharing similar characteristics, many of the video statistics of these traditional media servers are quite different from YouTube-like sites, e.g., the video length distribution and user access pattern. More importantly, these videos are generally movies and TV programs that are not generated by ordinary users, nor are they connected by social relations.

**Video Category**

In YouTube, one of 15 categories is selected by a user when uploading a video. Table 18.2 lists the number and percentage of all the categories, from a dataset of 5 million videos crawled over a 1.5-year span [3]. We can see that the distribution is highly skewed: the most popular category is "Entertainment," at about 25.4 %, and the second is "Music," at about 24.8 %. These two categories of videos constitute half of the entire YouTube video collection, suggesting that YouTube is mainly an entertainment-like site. "Unavailable" are videos set to private, or videos that have been flagged as inappropriate content, for which the crawler can only get meta information from the YouTube API. "Removed" are videos that have been deleted

**Table 18.2** List of YouTube video categories

| Rank | Category | Count | Percentage (%) |
|---|---|---|---|
| 1 | Entertainment | 1,304,724 | 25.4 |
| 2 | Music | 1,274,825 | 24.8 |
| 3 | Comedy | 449,652 | 8.7 |
| 4 | People and blogs | 447,581 | 8.7 |
| 5 | Film and animation | 442,109 | 8.6 |
| 6 | Sports | 390, 619 | 7.6 |
| 7 | News and politics | 186,753 | 3.6 |
| 8 | Autos and vehicles | 169,883 | 3.3 |
| 9 | Howto and style | 124,885 | 2.4 |
| 10 | Pets and animals | 86,444 | 1.7 |
| 11 | Travel and events | 82,068 | 1.6 |
| 12 | Education | 54,133 | 1.1 |
| 13 | Science and echnology | 50,925 | 1.0 |
| 14 | Unavailable | 42,928 | 0.8 |
| 15 | Nonprofits and activism | 16,925 | 0.3 |
| 16 | Gaming | 10,182 | 0.2 |
| 17 | Removed | 9,131 | 0.2 |

by the uploader, or by a YouTube moderator (due to violation of the terms of use), but are still linked to by other videos.

**Video Length**

The length of YouTube videos is the most distinguishing difference from traditional video contents. Whereas most traditional servers contain a significant portion of long videos, typically 1–2 hour movies (e.g., HPLabs Media Server [4] and OnlineTVRecorder [5]), YouTube mostly comprises short video clips, and 98.0 % of the videos' lengths are within 600 s. Although YouTube has increased its initial 10 min length limit to 15 min and allows certain users to upload videos of unlimited length, most of the user-generated videos remain quite short in nature.

Figure 18.1 shows the histogram and cumulative distribution function (CDF) of YouTube videos' lengths within 700 s, which exhibits three peaks. The first peak is within 1 min, and contains 20.0 % of the videos, which shows that YouTube is primarily a site for very short videos. The second peak is between 3 and 4 min, and contains about 17.4 % of the videos. As shown in Fig. 18.2, this peak corresponds to the videos in the "Music" category, which is the second most popular category for YouTube. The third peak is near the maximum of 10 min, which is the earlier limit on the length of uploaded videos.

Figure 18.2 shows the video length distributions for the top four most popular categories. "Entertainment" videos have a distribution similar to the entire videos. "Music" videos have a high peak between 3 and 4 min (29.1 %), which is the

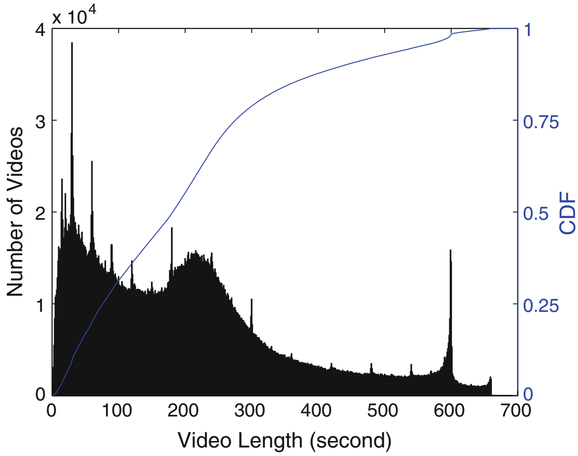


**Fig. 18.1** Histogram and cumulative distribution (CDF, the *solid line*) of YouTube video length

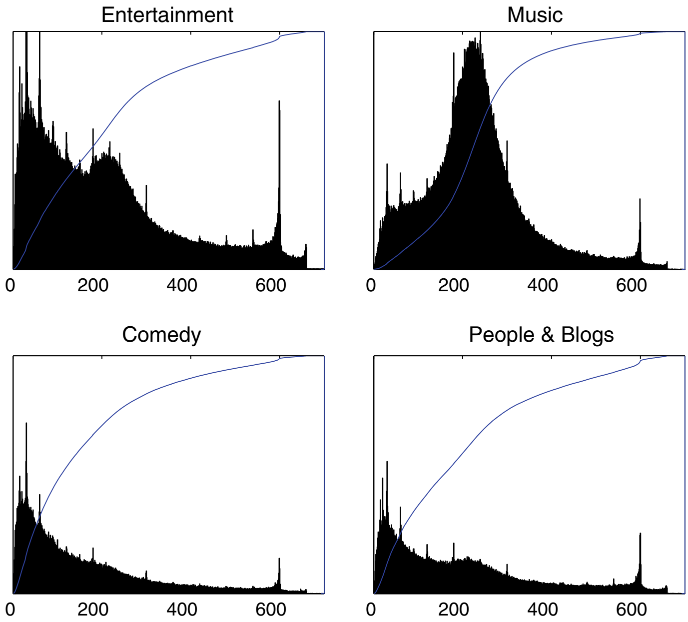


**Fig. 18.2** Length histograms and cumulative distributions for the four *top* categories

typical length range for music TVs. "Comedy" and "People & Blogs' videos have more videos within 2 min (53.1 % and 41.7 % respectively), likely corresponding to "highlight" type of clips.

### Access Patterns

Given that UGC video length is shorter by two orders of magnitude as compared to traditional movies or TV shows, YouTube's video content production is significantly faster with less effort [6]. It has been found that the 10 % top popular videos account for nearly 80 % of views, indicating that YouTube is highly skewed toward popular videos. This also implies that proxy caching can have high hit ratios since only a small portion of the videos will be requested frequently.

Yet YouTube users tend to abort the playback very soon, with 60 % of videos being watched for less than 20 % of their duration, which is particularly true for mobile users [7]. Furthermore, only 10 % of the videos are watched again on the following day [8]. A closer examination that classifies YouTube videos into top videos, removed pirate videos, and random videos has shown that copyrighted videos tend to get most of the views earlier, while videos in the top lists tend to experience sudden bursts of popularity [9,10]. In a campus network, however, the top popular videos do not contribute much to the total videos viewed on a daily basis, probably because the users are of closer relations in sharing video [8]. All these points suggest that YouTube users' viewing behaviors are highly diversified, affected by both the video quality (as in traditional video sharing) as well as their social relations (unique to social media).

## 18.2.3 Small-World in YouTube Videos

YouTube is a prominent social media service: there are communities and groups in YouTube, and thus videos are no longer independent of each other. Such social networking is unique to this new generation of video sharing services. It has been shown that, besides web searching, tracing related video is another top view source in YouTube-like UGC sites. There is a strong correlation between the number of views of a video and that of its top related videos [10], and this also provides more diversity on video views, helping users discover more videos of their own interest rather than the popular videos only.

The small-world network phenomenon is probably the most interesting characteristic for social networks. Milgram [11] initiated the study of small-world networks when investigating the phenomenon that people are linked by short chains of acquaintance (a.k.a., six degrees of separation). Such networks possess characteristics of both random graphs [1] and regular graphs[2] [13]. More formally, given the network as a graph $G = (V, E)$, the *clustering coefficient* $C_i$ of a node $i \in V$ is the proportion of all the possible edges between neighbors of the node that actually exist in the graph, and the clustering coefficient of the graph, $C(G)$, is the average of the clustering coefficients of all nodes. The *characteristic path length* $d_i$ of a node

[1] A random graph is generated based on certain probability distributions. Purely random graphs, built according to the *Erdös-Rényi* (ER) model, exhibit a small characteristic path length (varying typically as the logarithm of the number of nodes) along with a small clustering coefficient [12].
[2] A regular graph is a graph where each vertex has the same degree.

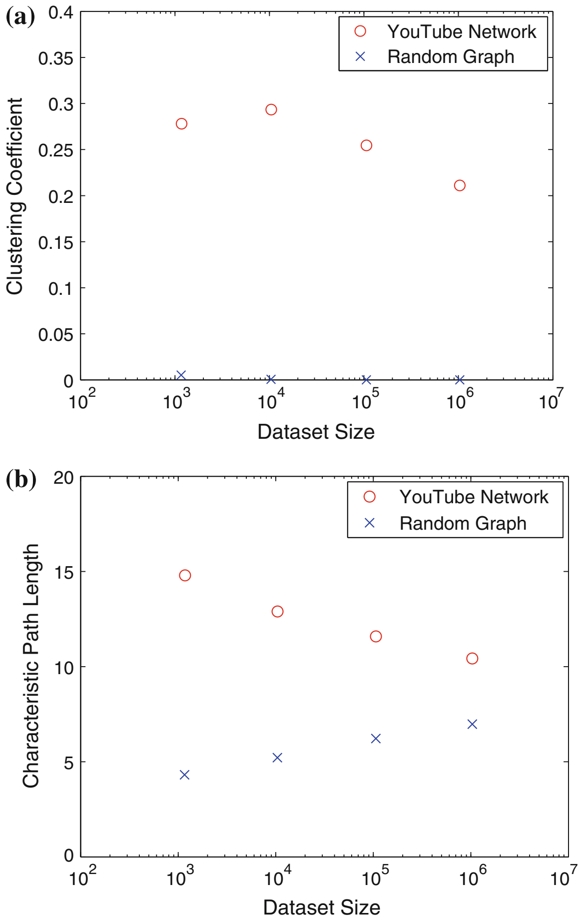


**Fig. 18.3** Small-world characteristic of YouTube videos

$i \in V$ is the average of the minimum number of hops it takes to reach all other nodes in $V$ and the characteristic path length of the graph, $D(G)$, is then the average of the characteristic path lengths of all nodes. A small-world network has a large clustering coefficient like a regular graph, but it also has a small characteristic path length like a random graph.

The graph topology for the network of YouTube videos can be measured by using the related links in the YouTube dataset to form directed edges in a video graph. Figure 18.3a shows the clustering coefficient for the graph, as a function of the size of the dataset. The clustering coefficient is quite high (between 0.2 and 0.3), especially in comparison to random graphs (nearly 0). There is a slow decreasing trend in the clustering coefficient, showing that there is some inverse dependence on the graph size, which is common for small-world networks [14]. Figure 18.3b shows

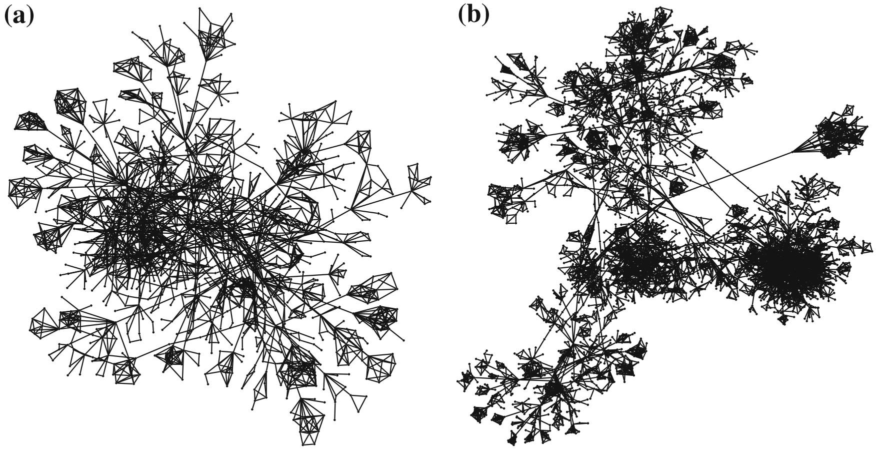


**Fig. 18.4** Two sample graphs of YouTube videos and their links

the characteristic path length for the graphs. It can be seen that the average diameter (between 10 and 15) is only slightly larger than the diameter of a random graph (between 4 and 8), which is quite good considering the still large clustering coefficient of these datasets. Moreover, as the size of graph increases, the characteristic path length decreases for the YouTube video graph, but increases for random graphs with the same number of nodes and average node degrees. This phenomena further verifies that the YouTube graph is a small-world network.

The small-world characteristics of the video graph can also be observed from their visual illustrations (see Fig. 18.4 for two representatives of 1000 and 4000 nodes). The clustering behavior is very obvious in these two graphs, due to the user-generated nature of the tags, titles, and descriptions of the videos that are used by YouTube to find related ones. The results are similar to other real-world user-generated graphs that exist, yet their parameters can be quite different. For example, the graph formed by URL links in the World Wide Web exhibits a much longer characteristic path length of 18.59 [15]. This is likely due to the larger number of nodes ($8 \times 10^8$ in the Web), but it also indicates that the YouTube network of videos is a much closer group.

### 18.2.4 YouTube from a Partner's View

YouTube displays advertisements on the webpages to monetize videos, and this has been the main source of YouTube's revenue. Besides user-generated videos, such companies and organizations as Electronic Arts, ESPN, and Warner Brothers are also providing their premium videos on YouTube now. To accommodate these content owners with copyrighted videos and popular channels, YouTube has introduced a *YouTube Partner Program*, which has largely improved the quality of YouTube videos, and has further increased YouTube's revenue.

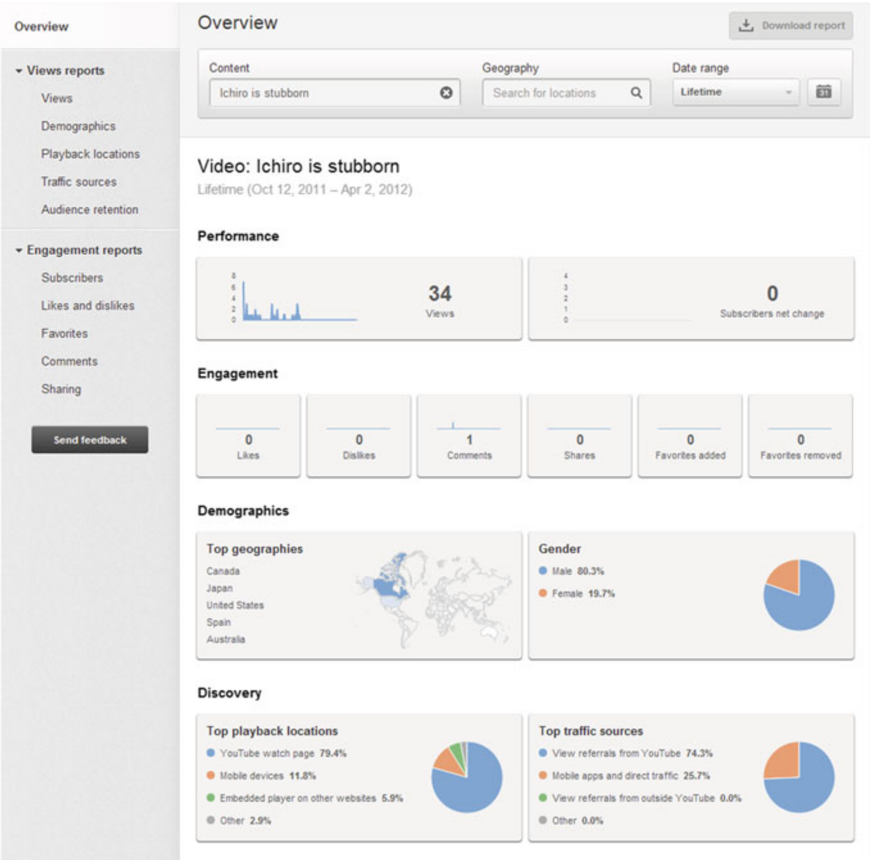


**Fig. 18.5** YouTube insight *dashboard*

The statistics of videos are of great potential value to the YouTube partners. For example, which videos are popular? And which external websites are referring more views? The partners can leverage these statistics to adapt their content deployment and user engagement strategies. To help the YouTube partners with this goal, YouTube introduced the *Insight Analytics* to provide various basic statistics on videos and channels. Figure 18.5 gives a snapshot of the Web-based Insight Analytics dashboard.

YouTube users have various means to reach YouTube videos. The last webpages where the viewers come from is called *referral sources*. Understanding referrals is essential for YouTube partners to adapt their user engagement strategy. We can classify the referral sources into four categories [16]:

`Suggestion` The referral comes from YouTube's related video links;

`Video Search` The referral comes from YouTube or search results, e.g., from Google;

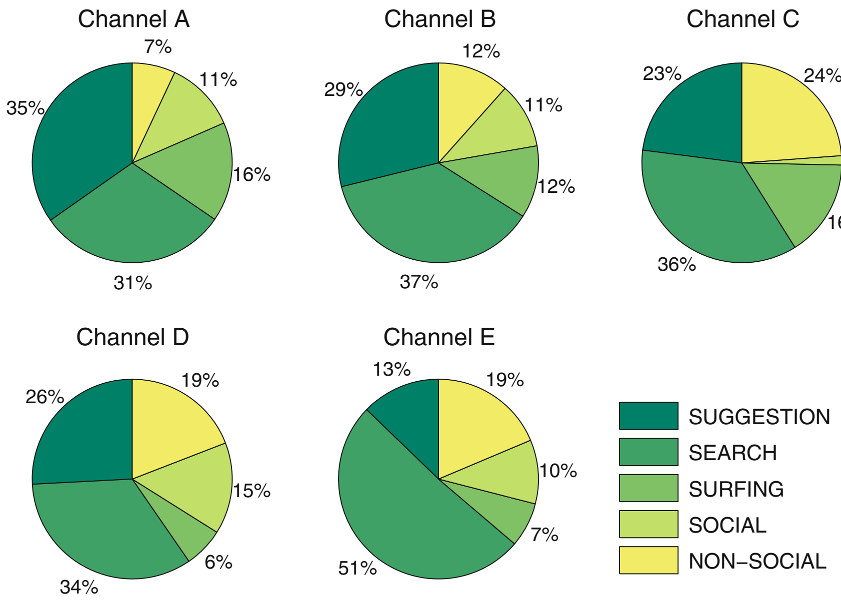


**Fig. 18.6** Breakdown of the referral source

`YouTube Surfing` The referral comes from any YouTube pages (except for related video links and search results), including annotation links, YouTube channel pages, subscriber links, paid and unpaid YouTube promotion, and etc.;

`Social Referral` The referral source is a link on an external webpage, or the video was embedded on an external webpage;

`Non-Social Direct` YouTube analytics does not identify a referral source, indicating that the viewer navigated directly to the video, e.g., by copying and pasting the URL.

Figure 18.6 shows the breakdown of the above five categories for four sample channels. It is clear that the breakdown percentages are channel-dependent. For example, one-third of the users reach Channel A videos from suggested videos, and one-third reach from search results; Channel B and Channel D is similar to Channel A; very few users reach Channel C videos from external sources; half of the users reach Channel E videos from search results.

This observation confirms that search results and related videos (`Suggestion`) are the top sources of views [9,10]. Although `Social Referral` is not the top view source, the impact of external website referral cannot simply be ignored. There is a great chance that a user, attracted by an external referral, will watch more videos from the related video list. In other words, `Social Referral` can be considered as an introductive referral.

**Table 18.3** Summary of top external websites referrers

| | Channel A | Channel B | Channel C |
|---|---|---|---|
| 1st | 9.0 % downloading site | 16.2 % Facebook | 31.9 % gaming wiki |
| 2nd | 4.4 % Facebook | 2.2 % n/a | 7.6 % Facebook |
| 3rd | 2.6 % forum | 1.5 % downloading site | 5.3 % gaming blog |
| 4th | 1.7 % gaming site | 1.2 % n/a | 5.1 % gaming site |
| 5th | 1.5 % gaming site | 0.9 % downloading site | 3.7 % Internet video site |
| | Channel D | Channel E | |
| 1st | 41.2 % Reddit | 62.4 % Facebook | |
| 2nd | 9.9 % Facebook | 2.4 % music streaming | |
| 3rd | 4.7 % Twitter | 2.0 % music blog | |
| 4th | 2.0 % blog | 2.0 % Twitter | |
| 5th | 1.7 % entertainment site | 1.6 % music blog | |

Table 18.3 lists the top-five external website sources for each channel. It does not disclose the specific names of the websites except such notable social networking service as Facebook, Twitter, and Reddit, and simply uses general descriptions. It again can be seen from the table that there is channel-dependency. No external website dominates the external referrals for Channel A, and the small percentage indicates that there is a great number of sources. In Channel E, over 60 % of the referrals are from Facebook, over 20 times greater than the second one. Facebook also dominates in Channel B, yet the percentage is not as high as in Channel E. Facebook is the second in both Channel C and Channel D, while the first ones have high percentage. In summary, YouTube videos have been sharing in many different portals and we will further study the propagation structures in online social networks later.

### 18.2.5 Enhancing UGC Video Sharing

The size of user-generated YouTube videos (most being less than 25 MB) is much smaller than a traditional video (a typical MPEG movie of 700 MB). Yet the number of these videos (in the billion order now) is orders of magnitude higher than that of traditional video services (e.g., only 412 in HPLabs' Media Server [4]), and this number is rapidly increasing with new user contributions. As such, the scalability challenges faced by YouTube-like social media services is indeed more significant. This is further complicated by the social networking amongst YouTube users and videos, which however also opens new opportunities. We next discuss the implications of these unique characteristics of YouTube videos toward improving its service, in terms of latency, bandwidth, storage, and scale.

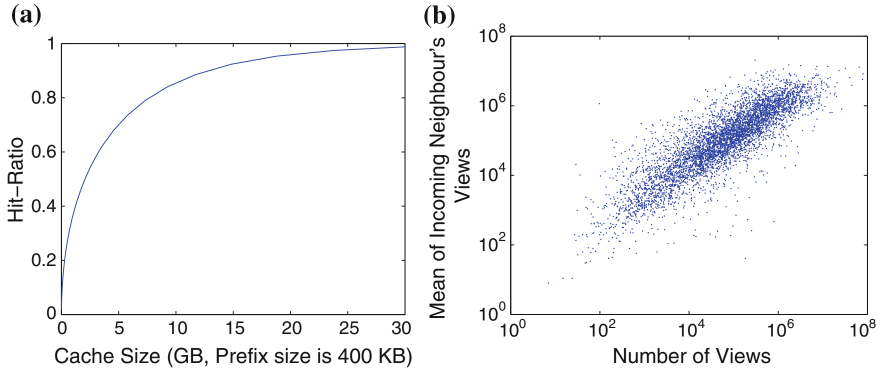


**Fig. 18.7** Proxy caching for YouTube videos. **a** Prefix caching hit-ratio as a function of cache size **b** Mean of neighbors' views against the number of views for a video

## Proxy Caching

As mentioned earlier, YouTube is highly skewed toward popular videos: 10 % top popular videos account for nearly 80 % of the views [6]. On the other hand, YouTube users tend to abort the playback very soon, with 60 % of videos being watched for less than 20 % of their duration, which is particularly true for mobile users [7]. As such, proxy caching, in particular, prefix caching [17] (see Sect. 16.1.2), can be quite effective. Assume for each video, the proxy will cache a 10 second initial prefix, i.e., about 400 KB of the video. Based on existing statistics, Fig. 18.7a plots the hit-ratio as a function of the cache size, assuming that the cache space is devoted only to the most popular videos. To achieve an 80 % hit-ratio, the proxy would require less than 8 GB of disk space for the current YouTube video repository, which is acceptable for today's proxy servers.

The cache efficiency can be further improved by exploring the small-world characteristic of the related video links. That is, if a group of videos have a tight relation, then a user is likely to watch another video in the group after finishing the first one. This expectation is confirmed by Fig. 18.7b, which shows a clear correlation (correlation coefficient being 0.749) between the number of views for a videos and the mean views of its related videos. Once a video is played and cached, the prefixes of its directly related videos can also be pre-fetched and cached, if the cache space allows.

## Storage in Content Distribution Networks

Given the sheer data volume of user-generated content and the locality of users' interests, when storing the UGC videos in a CDN it is necessary to partition the contents and assign them into different geo-distributed CDN servers. Different from traditional web contents or standalone movies that are isolated, social media contents have connections among each other, and thus the partition is nontrivial.

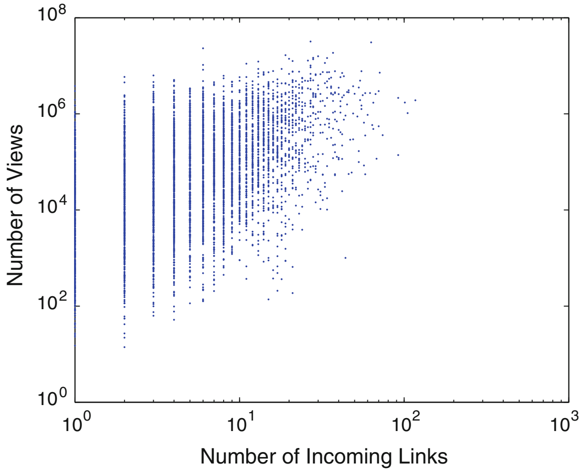


**Fig. 18.8** Popularity against incoming links

It is intuitive to preserve the social relationship when partitioning the social graph, because contents with social relationship are likely be accessed together or within a short period. If the two related contents are located in two different servers, it might increase the lookup time and the communication overhead [18]. Meanwhile, the user access pattern should also be considered, so as to balance the workloads of the servers.

Consider YouTube's video graph, which is a directed graph with videos related to others. Figure 18.8 shows the scatter plot of the YouTube video's popularity against the number of incoming links. There is a clear trend that videos with more incoming links are more popular, because these videos have more chances to be accessed through related videos. As shown in Fig. 18.7b, there is also strong correlation between the video views of a node against its neighbor's. Most of the videos have a comparable number of views as their neighbors'.

In summary, a popular content's social neighbors are probably also popular, and they are likely to be clustered based on the social relationships. As such, partitioning the content entirely based on social relationship would lead to unbalanced partitions, that is, some CDN servers will have high or even overwhelming workload while some servers can be nearly idle.

Figure 18.9 gives a simple example to make this phenomenon clearer: 8 nodes, each with a weight in terms of popularity as shown, constitute a social graph, which is going to be divided into two parts. As a result, the number of interconnections is 2 of the left graph, but the standard deviation of the total weight in each part is as great as 348; for the right graph that takes popularity into account, although the number of interconnections increases to 4, the standard deviation is reduced to as small as 23, and thus the two parts will be accessed more evenly.

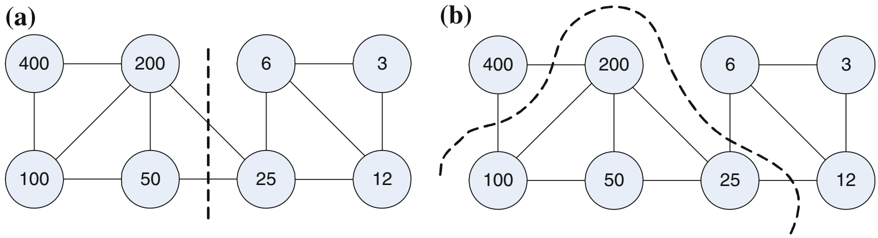


**Fig. 18.9** Example of different partitions based on **a** social relationship only, **b** both social relationship and popularity

## Peer-to-Peer Sharing

A shift to the peer-to-peer paradigm has also been suggested to overcome the scalability challenge. However, since most YouTube videos are short, a peer-to-peer overlay can be highly dynamic and unstable with frequent joins and leaves. On the other hand, the social relations can be used again to foster user collaboration and peer partner search. One example that explores the user interest correlation for peer-assisted short video sharing is NetTube [19].

In NetTube, the server stores all the videos and supplies them to clients. The clients also share the videos through peer-to-peer communications—each client (a peer) caches all its previously played videos, and makes them available for redistributing. As such, for a client interested in a particular video, all the peers that have previously downloaded this video can serve as potential suppliers, forming an overlay for this video, together with the peers that are downloading this video.

A mesh-based overlay is formed for each video in NetTube, where the peers pull expected data from a set of partners (other peers or the server) through a sliding-window-based scheduling algorithm. Yet, given the shorter video length, the startup and playback delay would be amplified from the perception of users. Given the users are less patient in waiting for short videos, more trials of joining/leaving would occur, leading to even higher churn rates. To address these challenges, a novel delay-aware scheduling that is customized for the short videos is developed. It implements an intelligent indicator in the downloading buffer to tell whether the peer is about to encounter delay. If yes, it will utilize an aggressive strategy for transmitting the data to mitigate delay. That is, the senders will prioritize such requests, even if they have to suspend some other transmissions.

Since for such short video sharing, a client in general will watch a series of videos, it is necessary to quickly locate the potential suppliers for the next video and enable a smooth transition. To this end, NetTube introduces an upper layer overlay on top of the overlays of individual videos. In the upper layer overlay, given a peer, neighborhood relations are established among all the overlays that contain this peer. This is a conceptual relation that will not be used for data delivery; instead it enables quick search for video suppliers in the social network context, i.e., clustering clients with similar interests. To achieve fast and smooth transition, NetTube further introduces

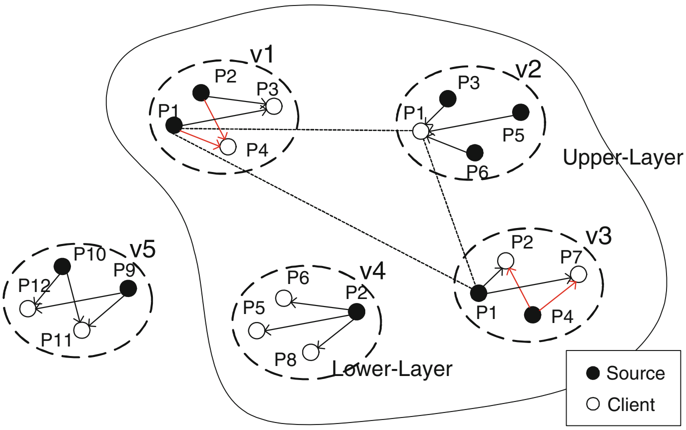


**Fig. 18.10** Illustration of a bi-layer overlay in NetTube

a cluster-aware pre-fetching, where the system pre-fetches *video prefixes* during the playback of the current video. With the existence of video interest correlation, the hit-rate of pre-fetching can be very high after a client plays back multiple videos, as discussed earlier. Figure 18.10 shows this bi-layer overlay.

## 18.3 Media Propagation in Online Social Networks

The new generation of online social network services, such as Facebook or Twitter, directly connect people through cascaded relations, and information thus spreads much faster and more extensively than through conventional web portals or news-group services, not to mention cumbersome emails [20]. As an example, Twitter first reported Tiger Woods' car crash 30 min before CNN, inverting the conventional 2.5-h delay of online blogging after mainstream news report [21].

With the development in broadband access and data compression, video has become an important type of object spreading over social networks, beyond earlier simple text or image object sharing [20,22]. Yet video objects, as richer media, possess quite different characteristics. From a data volume perspective, video objects are generally of much larger size than other types of objects; hence, most videos are fed from external hosting sites, e.g., YouTube, and then spread as URL links (together with titles and/or thumbnails). As a matter of fact, today's video sharing services and social networking services have become highly integrated. YouTube enables automatic posting on Facebook and Twitter based on users' options, and the users can also share interesting videos on their social networking webpages. YouTube's own statistics reveal that 500 years of YouTube video are watched every day on Facebook, and over 700 YouTube videos are shared on Twitter each minute.

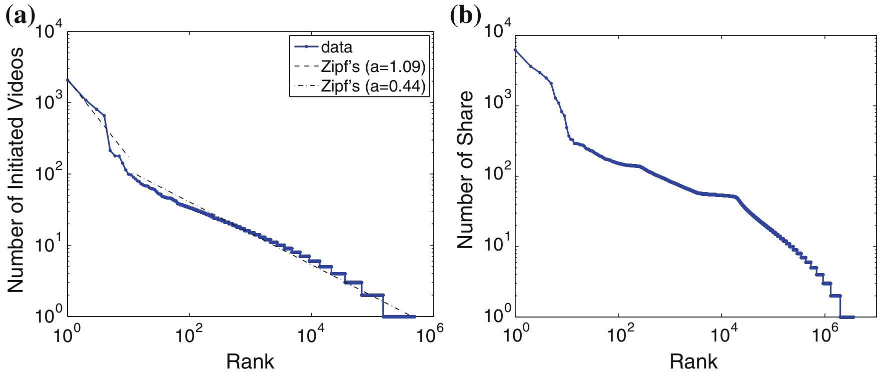


**Fig. 18.11** Rank distributions of initiated videos and shared videos. **a** Initiated videos **b** Shared videos

From a social perspective, text diaries and photos often possess personal information, while videos are generally more "public." Together with shorter links, videos often spread more broadly than texts and images. Yet the sheer and ever-increasing data volume, the broader coverage, and the longer access durations of video objects also present significant challenges compared to other types of objects, not only to social networking service management, but also to network traffic engineering and to the resource provisioning of external video sites.

### 18.3.1 Sharing Patterns of Individual Users

Since video object sharing involves both propagation over the online social network and accesses to the external video site, there are two critical questions to answer.

1. How often do users initiate video sharing?
2. How often do users further share a video upon receiving it?

Each initiator triggers the first share of a video. It has been found that, in a one-week dataset of 12.8 million video sharing and 115 million viewing events, 827 thousand initiating records can be extracted [23]. While this number is not small, it is only 6.5 % out of the 12.8 million sharing records. This indeed reflects the pervasiveness and power of video spreading in social network.

The rank distribution of the initiators (in terms of the number of initiated videos) is plotted in Fig. 18.11a. It suggests that most users initiate few videos, but a few *active users* have initiated a remarkable number of videos. The most active user indeed has initiated over 2,000 videos in one week.

Zipf's law [24] is usually used to describe a skewed distribution, which is a straight line in logarithmic scale. However, the data in Fig. 18.11a cannot be simply fitted by one Zipf line: the data after top-10 appear to be a straight line, but the top-10 data clearly differ from the rest. Yet they can be roughly fitted by another Zipf line. The distinction suggests the existence of two possible types of users with different initiating behaviors: (1) most of the initiators (over 99 %) initiate only a few videos; and (2) a set of active initiators have much more friends and also initiate a much larger number of videos. The change at threshold 90 is actually quite sharp, showing clear distinction in the two types of initiators. These active initiators serve as hubs that draw much more attention than the general users and are worthy of particular attention in system optimization.

The distribution of the number of each user's shares is shown in Fig. 18.11b, which again indicates that there are some extremely active users sharing a great number of videos, although most of the users only share a small number of videos. There are also users who have watched more than 1,000 videos without sharing any, like *free-riders* in peer-to-peer systems.

The above observations suggest that the users have diverse activeness, and we can roughly distinguish three types of users.

- *Spreaders* (SU), a small number of users who initiate a lot of videos, and also have many friends, being hub-like. Some spreaders are non-personal accounts specifically interested in collecting and spreading interesting, funny, attractive contents, including videos; it is also possible that spreaders are bots, spreading videos in a spam manner;
- *Free-riders* (FU), who watch many videos without sharing any, which noticeably hinders video spreading;
- *Ordinary users* (OU), who sometimes initiate a few videos, watch some shared videos, and share some videos they watched.

### 18.3.2 Video Propagation Structure and Model

Figure 18.12 shows visual examples of two typical propagation structures: one type has a moderate depth, but limited branching—most of the branches are directly from the source, with no further branching, as shown in Figure 18.12a, b; the other type branches frequently at different levels, and some branches can be very long, as shown in Fig. 18.12c and Fig. 18.12d (the root node is enlarged for better visualization). Table 18.4 further lists the descriptions from each video's URL, revealing their content. We can see that the video propagation path and coverage are highly diverse, depending on both the video content itself and the user watching and sharing behavior.

There have been many studies on the propagation structure and model of message sharing through online social networks [20,25]. A widely used model that is aware of the users' status is the *epidemic model*, which describes the transmission of communicable disease through individuals [26]. As well as in epidemiology, it has also been recently used to model computer virus infections and information propagations such as news and rumors [27].

One classical epidemic model is the *SIR model* (Susceptible-Infectious-Recovered), first proposed by Kermack and McKendrick [28]. It considers a fixed population with three compartments: Susceptible (S), Infectious (I), and Recovered (R). The initial letters also represent the number of people in each compartment at a

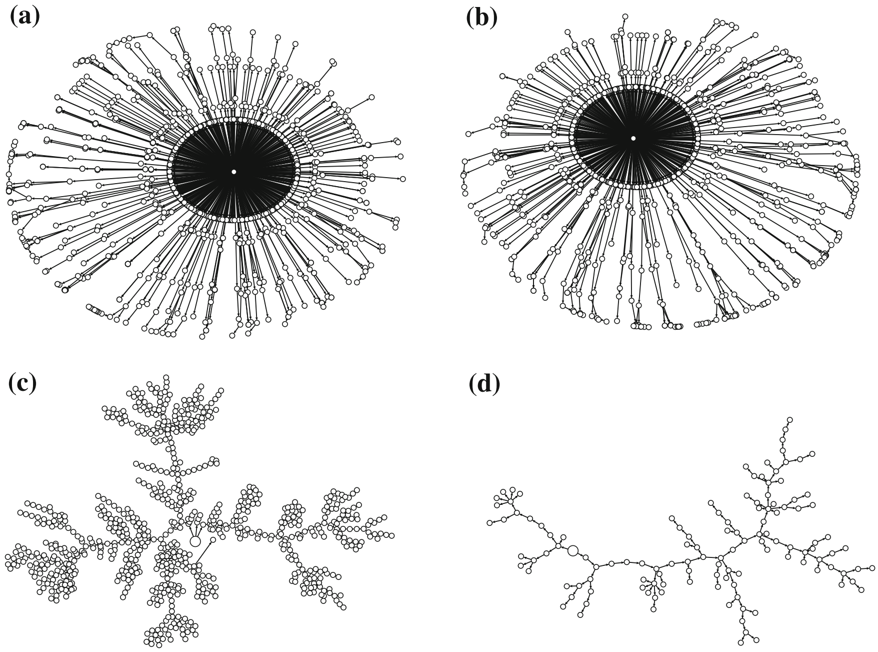


**Fig. 18.12** Illustration of spreading trees for popular videos. **a** size = 1093, height = 9 **b** size = 951, height = 8 **c** size = 805, height = 30 **d** size = 126, height = 23

**Table 18.4** Statistics and descriptions of the four videos (see Figure 18.12)

| Size | Height | Views | Length (s) | Category | Description |
|---|---|---|---|---|---|
| 1093 | 9 | 34,531 | 123 | news | a father picked up daughter from school by helicopter |
| 951 | 8 | 14,281 | 60 | advt. | earth hour promotion video |
| 805 | 30 | 12,658 | 306 | music | charity single “Children” by Chinese stars |
| 126 | 23 | 1,431 | 235 | comedy | funny lip sync video |

particular time $t$, that is, $S(t)$ represents the number of individuals not yet infected; $I(t)$ represents the number of individuals who have been infected and are capable of spreading the disease to those in the susceptible category; $R(t)$ represents the number of individuals who have been infected and then recovered. Given transition rate $\beta$ from S to I and $\gamma$ from I to R, the following equations can be derived:

$$\frac{dS}{dt} = -\beta SI, \quad \frac{dI}{dt} = \beta SI - \gamma I, \quad \frac{dR}{dt} = \gamma I.$$

There is a natural mapping between conventional object sharing propagation in social networks and the compartments of the SIR model. For a particular object, all

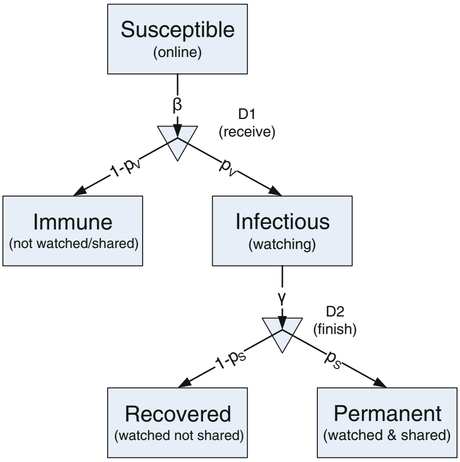


**Fig. 18.13** The Susceptible-Immune-Infectious-Recovered-Permanent (SI$^2$RP) model

the users in the social network are Susceptible at the beginning; at a certain time, the users accessing the object are Infectious, indicating that they are able to infect others by sharing the object. They can be Recovered if they choose not to share. Yet for video spreading, the mapping is not complete:

1. A user can choose not to watch the received video, and likely not participate in the spreading as well. To differentiate these users and the users in R who have watched or directly shared the video, we categorize these users to a new compartment, Immune (Im).
2. In the classical SIR model, the transition is time-dependent, i.e., at any time, there is a chance that the stage transits to the next one. While for video spreading in social networks, the transition of the stages depends on decisions at a certain time. For example, the user needs to choose watch or not, and share or not share. To address this problem, two temporary decision stage, D1 and D2 can be introduced.
3. It is necessary to differentiate the users who have shared the video and those who have not after watching the video. A new compartment, Permanent (P), can be introduced, indicating users who have shared the video, and otherwise Recovered.

The enhanced $SI^2RP$ (Susceptible-Immune-Infectious-Recovered-Permanent) model is illustrated in Fig. 18.13. For each video propagation process, the initiator is Infectious at the beginning.

The transition rate from S to D1 is $\beta$, and thus a Susceptible user will spend $1/\beta$ unit time to receive a shared video from a friend. The user then makes a decision whether or not to watch the video. If the user is not interested in it and decides not to watch or share, she/he is considered as Immune. The probability of the user watching or directly sharing the video can be denoted as $p_V$.

If the user decides to watch the video, she/he becomes Infectious. The transition rate from I to D2 is $\gamma$, indicating that the user will spend $1/\gamma$ time to finish watching the video. The user then makes the second decision, whether or not to share the video. If the user decides not to share, she/he becomes Recovered or Permanent otherwise. The probability of a user deciding to share the video is denoted by $p_S$. The transition rate $\beta$ and $\gamma$ can be inferred from measurement results, so are $D_I$, $p_V$, and $p_S$ in the model. These four probability distribution or probability characterize the behavior of different types of users, namely spreaders (SU), ordinary users (OU), and free-riders (FU).

- An SU initiates video shares according to distribution $D_I$ (An OU or FU does not initiate).
- A user watches videos shared by friends with probability $p_V$, which is based on the reception rate.
- After watching, an SU or OU shares the video with probability $p_S$, which is based on the share rate.

### 18.3.3 Video Watching and Sharing Behaviors

It has been found that, in social network sharing, compared to strangers, friends have relatively higher probability of reciprocal visits. When a content is uploaded by a friend, a user is more likely to browse. The users are also more active in viewing profiles than leaving comments, and consequentially, latent interactions cover a wider range of friends than visible interactions [29]. More importantly, most of the users are willing to share their resources to assist others with close relations, which naturally leads to collaborative delivery [30].

Different from text or images that can be instantly viewed, a posted video will not be really watched until the recipient clicks the link. Upon receiving the video post, the recipient (friend or follower) has three options:

1. Watches the video, and thus the requirement of streaming quality, such as startup latency and playback continuity, should be satisfied.
2. Not to watch the live video, but download the video and expect to watch it later.
3. Shows no interest in the video. If the user does not want to watch the video now or later, she/he may not share the resources with other uploader's friends, either.

The coexistence clearly makes a system design more complicated. More specifically, there exist two types of friends interested in the posted video, namely *streaming users* and *storage users*. The streaming users expect to watch the video immediately, and the storage users will download and watch the video at a different time, due to the presence of other concurrent events.

The streaming users might stop watching after a while if they find the video is out of their interest, even though the video is posted by friends. Such dynamics will affect the data delivery if they serve as relays for other users. On the other hand, the storage users who are downloading the video asynchronously do not have the concern of interest nor playback quality, until they start to watch the video. Hence such users are considered relatively stable.

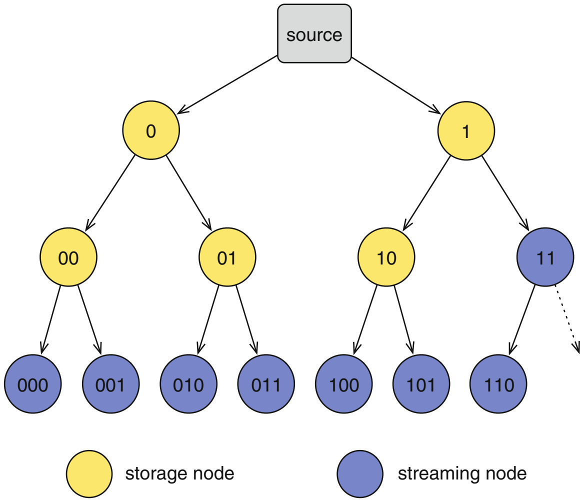


**Fig. 18.14** Example of a labeled overlay tree with IDs

### 18.3.4 Coordinating Live Streaming and Online Storage

The COOLS (Coordinated Live Streaming and Storage Sharing) system [30] utilizes the stable storage users to improve the QoS for streaming users in social video sharing. COOLS advocates a peer-to-peer tree overlay design for video posting and sharing. It is known that a tree overlay with data push is more efficient than a mesh overlay with data pull, but maintaining the tree with node churns is a daunting task. Fortunately, the existence of storage users implies that their churns are much less frequent than the traditional live streaming, which can thus be strategically placed to improve the robustness of a tree overlay.

To efficiently coordinate the two types of users, COOLS uses a labeled tree that embeds node locations in the overlay. An example of a binary labeled tree is given in Fig. 18.14. The two children of the root node (the source) have ID 0 and 1, respectively. For a given node, its left child's ID is the node's ID appended by a 0, and the right child's ID is that appended by a 1. As such, the ID embeds the location of a node and also that of its all ancestors. The number of digits (*length*) indicates its depth in the tree.

COOLS defines a partial order of the ID: if two IDs are of identical length, the one with a greater value is considered greater (for example, 010 is greater than 001); otherwise, the longer ID is greater (for example, 000 is greater than 11).

It also defines an *increment* operation of the ID: if not all the bits of the ID are 1s, an increment operation will increase the ID value by 1; otherwise, the length of ID will be increased by 1 and all bits are set to 0. This gives the *next value* of the ID. The operation of *decrement* can be defined accordingly by decreasing the ID.

Since the storage nodes are relatively stable, they should be placed at more critical locations of the tree, that is, close to the source. In other words, the storage nodes' IDs should be smaller than that of streaming nodes after the tree is stabilized. Figure 18.14 shows the organization of two types of nodes in the overlay tree.

For streaming nodes, it is necessary to guarantee short startup latencies, which requires them to be close to the source as well. Fortunately, since the storage users are delay-tolerant, the dilemma can be eliminated by prioritizing the streaming nodes in the initial stage.

Specifically, COOLS first constructs two trees, one contains all the streaming nodes, referred to as the *streaming tree*, and the other contains all the storage nodes, referred to as the *storage tree*. The source only delivers data in the streaming tree at the beginning. After the streaming nodes have buffered enough data to avoid outage, the two trees will be merged to one final overlay tree.

The source records the current maximum ID of each tree. To construct the two trees, the source adds nodes to the corresponding trees sequentially. A newly added node will be assigned an ID as the next value of the maximum ID. The node thus knows its parent's location by checking the prefix of its own ID. If the source has enough children (2 in this case), it will provide the address of one of its children whose ID is the same as the first digit of the new node's ID; that is, which branch should the new node go. Otherwise, the new node becomes one of the source's children.

Figure 18.15 shows the procedure of merging the two trees. Denote the next value of the current maximum ID of the storage tree as *firstID*, as it will be the first ID of a streaming node after merging the two trees. The new ID of the left streaming child node will be $firstID$, and the right streaming child node is assigned with the next value of $firstID$. The source also computes a potential maximum ID based on the two original maximum IDs, denoted as *finalID*, e.g., 0000 in Figure 18.15. The source then disseminates this value throughout the tree.

After the two trees have been merged, the overlay is probably not a complete tree, as some streaming nodes may locate deeper than expected, based on the $finalID$. These nodes are in an *non-steady state*, e.g., node 0000, 0001, 1100, 1101 and 1110 in the second step of Fig. 18.15. Some leaf storage nodes are also non-steady if they should have children but do not have yet, also based on $finalID$, e.g., nodes 00, 01 and 10. Other nodes are in a *steady state*. The non-steady streaming nodes should be promoted upwards.

The non-steady nodes send control messages toward the source. If the node finds out that its ID is no smaller than $finalID$, it will send a *promotion message*; if its potential children's ID is smaller than $finalID$ but does not have any child yet, the node will send a *child requiring message*. A rendezvous node (not necessary the source) receiving such messages matches them, and notifies the two senders to connect with each other. For example, in Fig. 18.15, node 00 matches itself with node 0000, node 0 matches node 01 with node 0001, node 1 matches node 10 with nodes 1100 and 1101, and the source matches node 01 with node 1110.

Suppose the heights of the original two trees are $H_l$ and $H_s$, respectively. To merge and promote, in the worst case, all the promotion and child requiring messages are

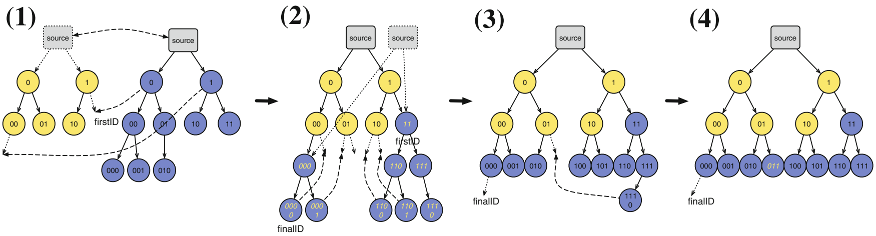

**Fig. 18.15** An example of COOLS overlay construction: creating, merging, and promotion

matched at the source. Thus in each round, the nodes in the lowest depth send promotion message and get matched, which takes $(H_s + H_l')$ unit time, where $H_l'$ is initially $H_l$ and decreased by 1 in each round. The tree's height will eventually become $H$, and all the nodes between depth $H_s$ and depth $H$ are streaming node. Hence there will be $(H_l + H_s - H)$ rounds. For a complete tree, all the three heights are bounded by $O(\log N)$, and the time to complete the promotion is therefore bounded by $O((\log N)^2)$, where $N$ is the total number of nodes in the system.

## 18.4 Further Exploration

Research on the social relations and social graphs in the human society has a long history, so does that on disease propagation in epidemiology [26,28]. Online social media and social networking however appeared only in very recent years and are still undergoing rapid changes. Research in this field remains in an early stage and many exciting topics are to be explored [31].

## 18.5 Exercises

1. Find out a typical Web 1.0 application and a typical Web 2.0 application, and discuss their key differences.
2. Discuss the key differences between YouTube videos and the traditional movies and TV shows. How would they affect content distribution?
3. YouTube publishes statistics about its videos online. As of the end of 2013, we have the following statistics:
   - More than 1 billion unique users visit YouTube each month
   - Over 6 billion h of video are watched each month on YouTube
   - 100 hours of video are uploaded to YouTube every minute
   - 80 % of YouTube traffic comes from outside the US
   - YouTube is localized in 61 countries and across 61 languages
   - Mobile makes up almost 40 % of YouTube's global watch time

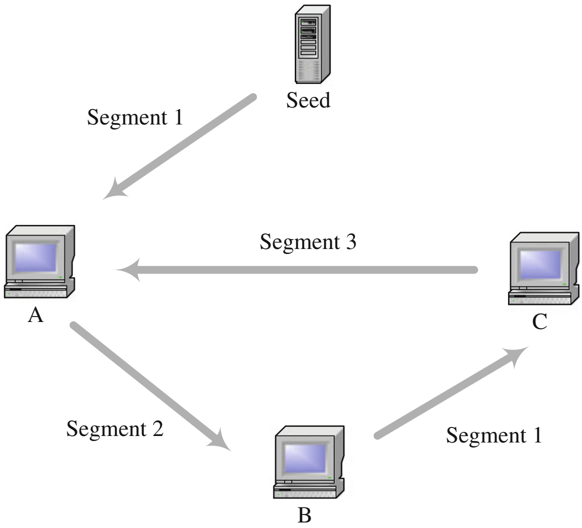


**Fig. 18.16** An example of the tit-for-tat strategy

Check the recent statistics and estimate the monthly growth speed of YouTube. Suggest some reasons that make YouTube-like services expand so quickly and the potential challenges therein.

4. Is it beneficial to place all the content from a social media service in one server? If not, what are the challenges to place the content in multiple servers?
5. Discuss the propagation and consumption patterns of multimedia content in a social networking tool that you are familiar with.
6. Given a positive integer $n$ and a probability value $0 \leq p \leq 1$, an *Erdös-Rényi* (ER) random graph $G(n, p)$ is with $n$ vertices where each possible edge has probability $p$ of existing. This is the most important class of random graphs.
   (a) Write a simple program to generate ER random graphs, and calculate their characteristic path lengths and clustering coefficients. Compare them with the YouTube video graph we have discussed earlier.
   (b) Discuss whether the graph formed by an online social network, say the graph of Facebook user accounts, is such a random graph or not. Hint: Think about the way that the edges are formed.
7. A simple model for information propagation is *gossip*. With gossip, a network node, upon receiving a message, will randomly forward it to the neighboring nodes with probability $p$.
   (a) Write a simple program to simulate the gossip algorithm in randomly generated networks. A node may delay $t$ time before forwarding. Discuss the impact of $p$ and $t$ on the coverage and propagation speed of a message.
   (b) Is it beneficial if the nodes can have different values of $p$? If so, provide some guidelines in the selection of $p$ for each node.

   (c) Is gossip suitable for modeling the propagation process of a picture shared in a realworld social network, say Facebook? How about video ?
8. In an online social network, a free rider only consumes videos but does not share videos. Free riders also exist in peer-to-peer file sharing: they download data segments from others, but never upload. BitTorrent adopts a *tit-for-tat* strategy to solve the incentive problem, i.e., you get paid only if you contribute. As depicted in Fig. 18.16, peers A, B, and C download different segments from each other. This forms a feedback loop; for example, uploading segment 2 from A to B will be feedback to A by the upload of segment 3 from C to A, which stimulates peer A to cooperate.
   (a) Discuss whether the tit-for-tat strategy works for video propagation with free riders.
   (b) For live video streaming with delay constraints, with tit-for-tat work?
9. The basic binary tree in COOLS can be quite high. For example, when there are 1000 nodes, the tree height can easily reach to 10.
   (a) What are the potential problems with a tall tree.
   (b) One simple solution to reduce the height of the tree to is increase the number of children for each node. Will this solution work for COOLS?
   (c) Suggest a possible solution that practically works and analyze its effectiveness.

## References

1. A. M. Kaplan, M. Haenlein, Users of the world, unite! the challenges and opportunities of social media. Bus. Horiz. **53**(1), 59–68, (2010)
2. S. Acharya, B. Smith, P. Parnes, Characterizing user access to videos on the World Wide Web. In *Proceedings of ACM/SPIE Multimedia Computing and Networking (MMCN)*, 2000
3. X. Cheng, J. Liu, C. Dale, Understanding the characteristics of internet short video sharing: a YouTube-based measurement study. IEEE Trans. on. Multimedia. **15**(5), 1184–1194 (2013)
4. W. Tang, F. Yun, L. Cherkasova, Amin Vahdat, *Long-term Streaming Media Server Workload Analysis and Modeling* (Technical report, HP Labs, 2003)
5. T. Hoßfeld K. Leibnitz, A qualitative measurement survey of popular internet-based IPTV systems. In *Proceedings of International Conference on Communications and Electronics (ICCE)*, June 2008, pp. 156–161
6. M. Cha, H. Kwak, P. Rodriguez, Y-Y Ahn, 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)*, October 2007, pp. 1–14
7. A. Finamore, M. Mellia, M. M. Munafò, R. Torres, S. G. Rao, YouTube Everywhere: impact of device and infrastructure synergies on user experience. In *Proceedings of the 2011 ACM SIGCOMM conference on Internet measurement (IMC '11)*, November 2011, pp. 345–360
8. P. Gill, M. Arlitt, Z. Li, A. Mahanti, YouTube Traffic Characterization: a view from the edge. In *Proceedings of the 7th ACM SIGCOMM Conference on Internet Measurement (IMC '07)*, October 2007, pp. 15–28

9. F. Figueiredo, B. Fabrício, J. M. Almeida, The tube over time: characterizing popularity growth of YouTube videos. In *Proceedings of the fourth ACM International Conference on Web Search and Data Mining (WSDM '11)*, February 2011, pp. 745–754
10. R. Zhou, S. Khemmarat, L. Gao, The impact of YouTube recommendation system on video views. In *Proceedings of the 10th Annual Conference on Internet Measurement (IMC '10)*, October 2010, pp. 404–410
11. Stanley Milgram, The Small World Problem. Psychol. Today **2**(1), 60–67 (1967)
12. D. B. West, *Introduction to Graph Theory*, (2nd Edition). (Oxford University Press, 2001)
13. Duncan J. Watts, Steven H. Strogatz, Collective dynamics of "Small-World" networks. Nature **393**(6684), 440–442 (1998)
14. Erzsébet Ravasz, Albert-László Barabási, Hierarchical organization in complex networks. Physical Review E **67**(2), 026112 (2003)
15. Réka Albert, Hawoong Jeong, Albert-László Barabási, The diameter of the World Wide Web. Nature **401**, 130–131 (1999)
16. X. Cheng, M. Fatourechi, X. Ma, J. Liu, Insight data of YouTube: from a partner's view. In Proceedings of ACM NOSSDAV, March 2014
17. S. Sen, J. Rexford, D. F. Towsley, Proxy prefix caching for multimedia streams. In *Proceedings of IEEE INFOCOM*, 1999
18. M. E.J. Newman, M. Girvan, Finding and evaluating community structure in networks. Phy. Rev. E, **69**(2), 026113 (2004)
19. X. Cheng and J. Liu. NetTube: exploring social networks for peer-to-peer short video sharing. In *Proceedings of IEEE INFOCOM*, April 2009, pp. 1152–1160
20. D. Wang, Z. Wen, H. Tong, C. -Y Lin, C. Song, A. -L Barabasi, Information spreading in context. In *Proceedings of the 20th International Conference on World Wide Web (WWW '11)*, April 2011, pp. 735–744,
21. J. Leskovec, L. Backstrom, J. Kleinberg Meme-tracking and the dynamics of the news cycle. In *Proceedings of the 15th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, June 2009, pp. 497–506
22. K. Dyagilev, S. Mannor, Y-T. Elad, Generative models for rapid information propagation. In *Proceedings of the First Workshop on Social Media Analytics (SOMA '10)*, July 2010, pp. 35–43
23. X. Cheng, H. Li, J. Liu, Video sharing propagation in social networks: measurement, modeling, and analysis. In *Proceedings of IEEE INFOCOM Mini-Conference*, April 2013
24. G. K. Zipf, Human Behavior and the Principle of Least Effort. (Addison-Wesley, Boston, 1949)
25. H. Kwak, C. Lee, H. Park, S. Moon, What is twitter, a social network or a news media? In *Proceedings of the 19th International World Wide Web Conference (WWW '10)*, April 2010, pp. 591–600
26. J. D. Daley, J. Gani, J. M. Gani, Epidemic modelling: an introduction. Cambridge Studies in Mathematical Biology. (Cambridge University Press, Cambridge, 2001)
27. Z. Liu, Y. -C. Lai, N. Ye, Propagation and immunization of infection on general networks with both homogeneous and heterogeneous components. Phy. Rev. E **67**(1), 031911 (2003)
28. W. O. Kermack, A. G. McKendrick, A Contribution to the mathematical theory of epidemics. Proc. R. Soc. Lond. Series A, 115(772), 700–721, (1927)
29. J. Jiang, C. Wilson, X. Wang, P. Huang, W. Sha, Y. Dai, B. Y. Zhao, Understanding latent interactions in online social networks. In *Proceedings of the 10th Annual Conference on Internet measurement (IMC '10)*, November 2010, pp. 369–382
30. Xu Cheng and Jiangchuan Liu. Tweeting videos: Coordinate live streaming and storage sharing. In *Proceedings of the 20th International Workshop on Network and Operating Systems Support for Digital Audio and Video, NOSSDAV '10*, pp. 15–20, 2010
31. C. Kadushin, *Understanding Social Networks: Theories, Concepts, and Findings.* (Oxford University Press, New York, 2012)