SocialVoD: a Social Feature-based P2P System

Wei Chang, and Jie Wu

Presenter: En Wang

Temple University, PA, USA

Outline

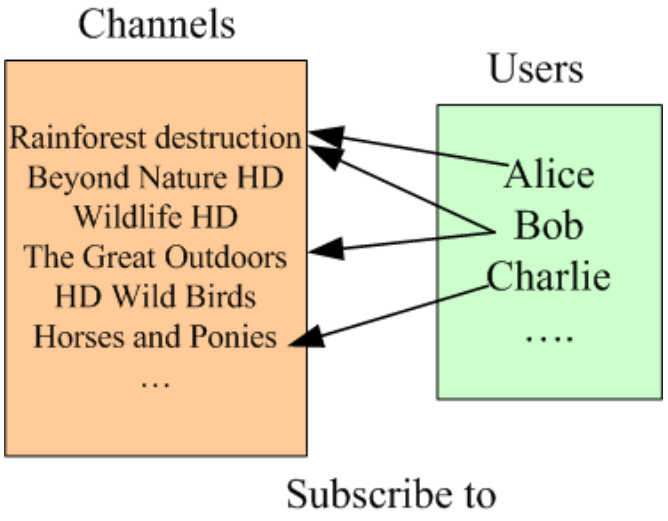
1. Introduction
2. Background and Related Work
3. Vimeo Category and User Interests
4. Overlay Design and File Searching
5. Evaluation
6. Conclusion

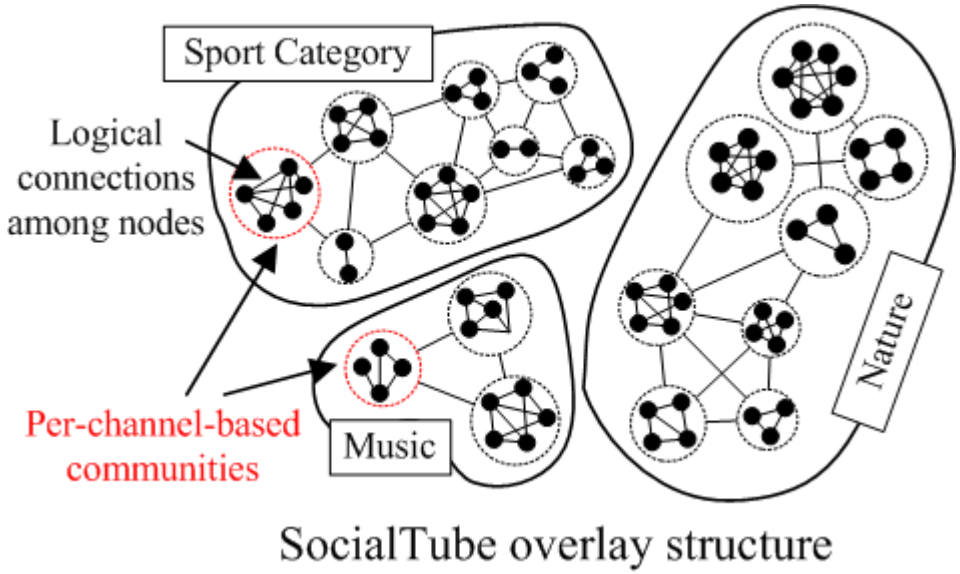
Introduction

- Video-on-demand (VoD) service has been explosively growing since its first appearance. For maintaining an acceptable buffering delay, the bandwidth costs have become a huge burden for the service providers.
- Complementing the conventional client-server architecture (CS) with a peer-to-peer system (P2P) can significantly reduce the central server's bandwidth demands.
- Previous works focus on establishing a P2P overlay for each video, which not only generates prohibitive costs for maintaining the overlay, but also create plenty of redundant links between a pair of nodes on different overlays.

Introduction (cont.)

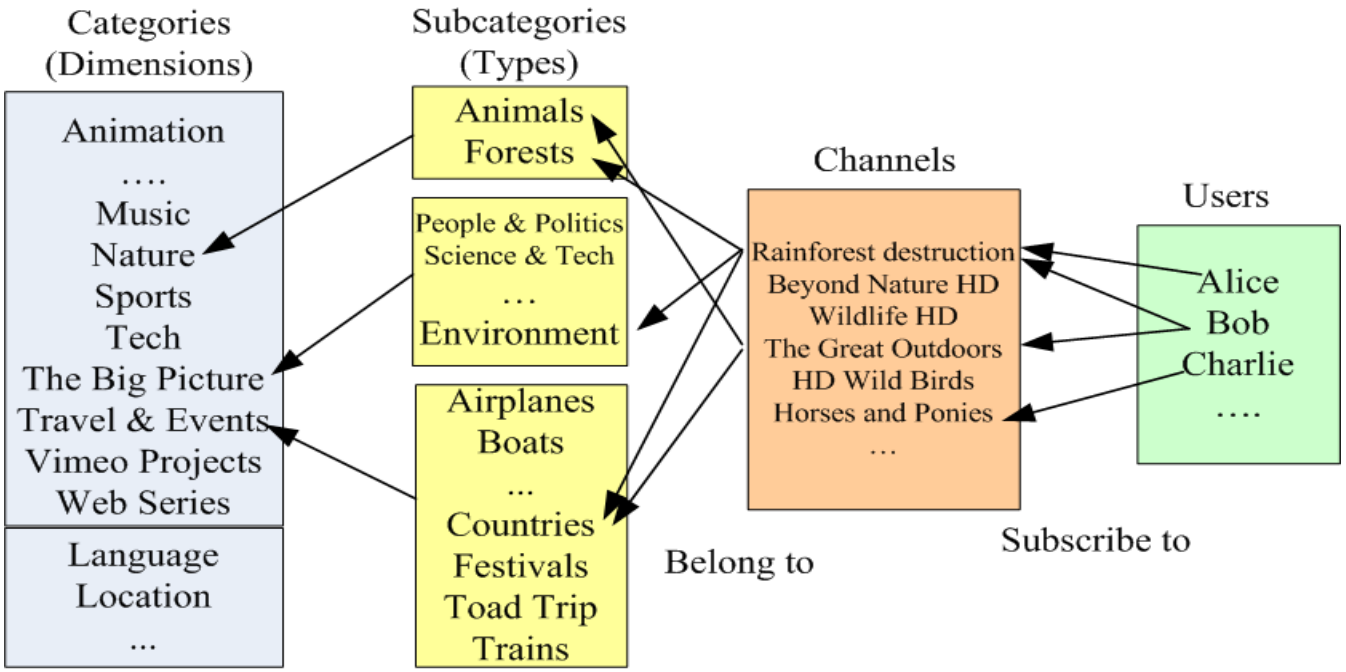
- Many VoD sites provide channel subscription functions to users.
- Considering that the subscribers of a channel are likely to watch the same videos from the channel, SocialTube proposes an interest-based per-channel hierarchical overlay.





Introduction (cont.)

- SocialTube has two problems.
  - It is not efficient for subscribers to explore (unsubscribed channels') videos belonging to other categories.
  - It is not applicable for the VoD sites, such as Vimeo, who are rich in category information.



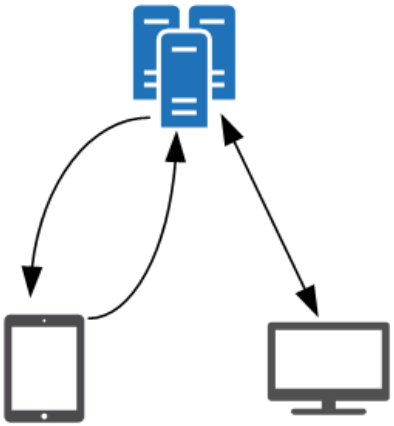
Introduction (cont.)

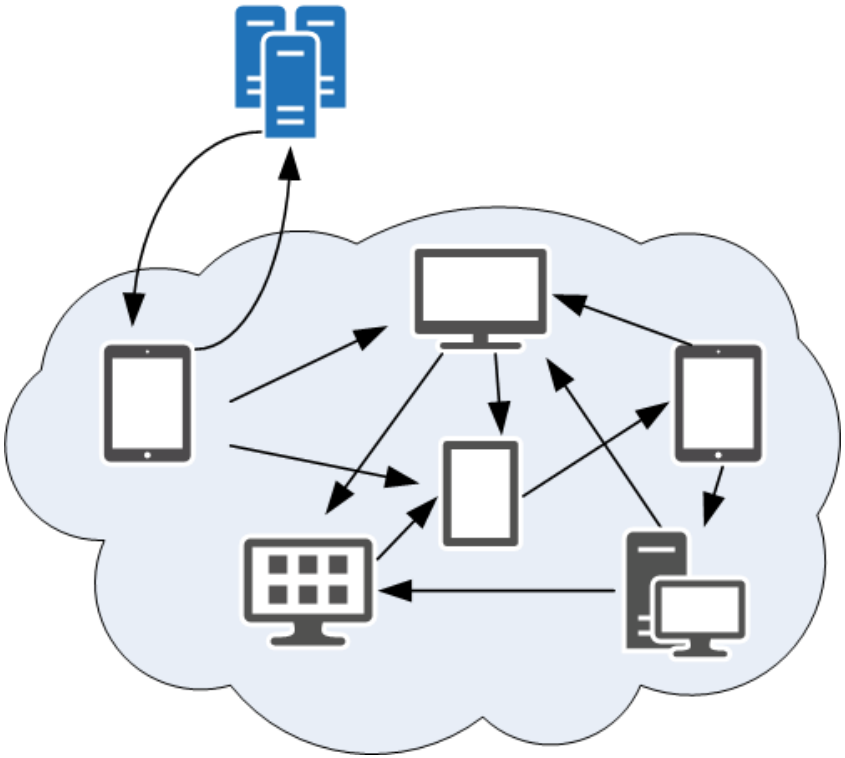
- In this paper, we propose a new hierarchical P2P overlay structure, called SocialVoD, by exploring both the subscription relationships and channels' similarities on Vimeo.
- The key design of SocialVoD is the utilization of channel subscribers' social closeness: the closer the watching preferences of users are, the better the overlay connectivity should be between them.
- We not only build interest-based per-channel overlays, but also, in a high-level, we organize these overlays according to channels' existing category information: by neighboring the channel overlays, whose category information differs in one dimension, the resulting graph contains enough routing "hints" for seeking other unsubscribed channels' videos.

Outline

1. Introduction
2. Background and Related Work
3. Vimeo Category and User Interests
4. Overlay Design and File Searching
5. Evaluation
6. Conclusion

Background and Related Work





P2P overlay design:

- Without social information
  - Hypercup and et al.
- With social information
  - NetTube: same watching history
  - SocialTube: same channel subscription
  - SocialVoD: neighboring nodes with similar watching preferences

Outline

1. Introduction
2. Background and Related Work
3. Vimeo Category and User Interests
4. Overlay Design and File Searching
5. Evaluation
6. Conclusion

Vimeo Category and User Interests

- It has been shown that most videos on VoD systems are short, and there are strong clustering behaviors among users.
- We further investigate the category features on Vimeo, and the change of users' interests over years.
- There are 58.06% popular channels (>300 subscribers) associated with more than one category feature.
- Users' watching preferences do change over the years.

TABLE I. THE DISTRIBUTION OF VIMEO CHANNELS' ASSOCIATED CATEGORY NUMBERS.

Number of Categories	1	2	3	4	5	...
Percentage (%)	41.94	38.11	8.44	7.54	3.96	0

TABLE II. THE CHANGE OF CATEGORY INTERESTS OVER 5 YEARS (%)

Year	Narrative	Arts	Music	Video School	Personal	Big Picture	Sports	Animation	Travel	Tech	Experimental	Comedy	Nature
2011	21.7	16.6	41.7	0	3.3	0	10.0	3.3	0	0	0	3.4	0
2010	41.2	4.8	4.1	4.5	11.4	4.6	0	9.4	13.6	0	1.8	4.6	0
2009	35.7	1.8	12.5	0	0	7.1	14.3	5.4	0	14.3	8.9	0	0
2008	13.0	6.2	22.5	24.5	4.1	4.1	4.1	6.2	0	2.9	4.2	8.2	0
2007	8.8	9.3	8.1	16.8	7.0	0	0	12.5	0	7.6	17.9	6.0	6.0

Watching Pattern of Users

- We summarize users' watching pattern of as follows:
  - Users frequently look for the videos from their subscribed channels;
  - Users are likely to seek other unsubscribed channels' videos, which satisfy their watching habits;
  - The watching habits may change. Users may access certain videos to explore new preferences.
- Based on these features, we design the new P2P overlay structure, SocialVoD.

Outline

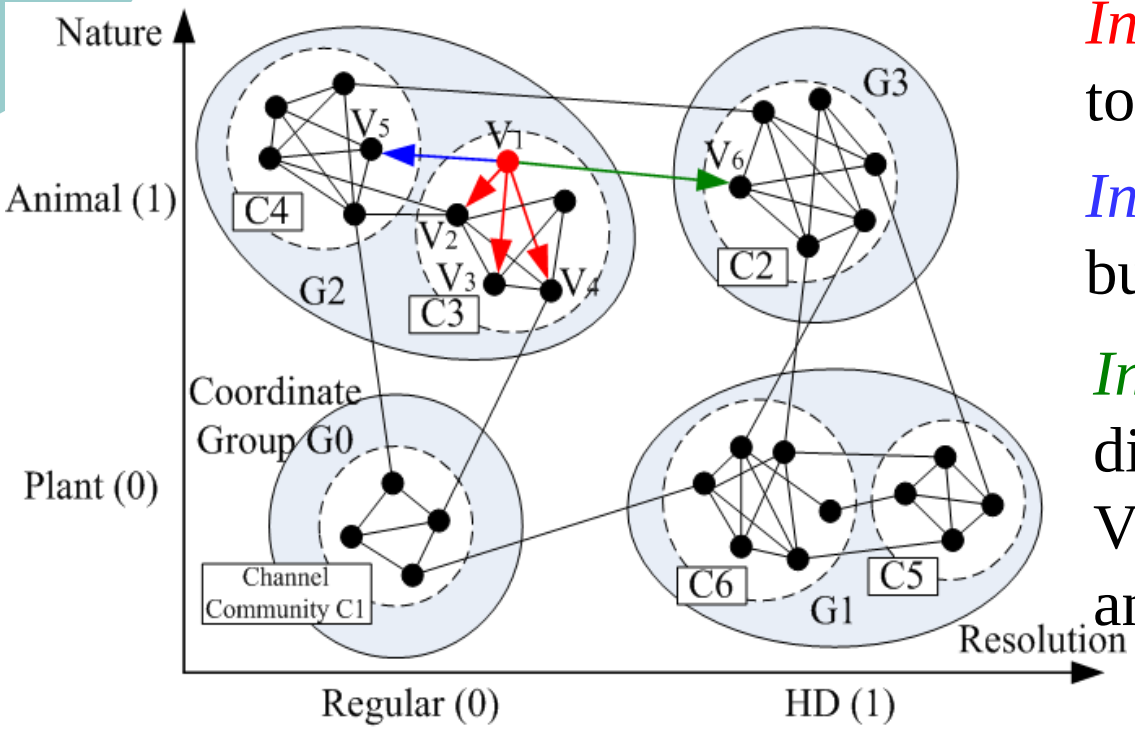
1. Introduction
2. Background and Related Work
3. Vimeo Category and User Interests
4. Overlay Design and File Searching
5. Evaluation
6. Conclusion

Overlay Design: basic model

- Feature coordinate of a video: $(f_1, f_2, \ldots, f_m), 1 \leq f_i \leq r_i$
- e.x. , in "Language" category dimension, $f_i = 0, 1, 2, \ldots$ may indicate English, Chinese, Spanish, and et al.
- Based on the feature coordinates, the category-feature space is partitioned into $\Pi_{i=1}^{m} r_i$ groups, and each channel belongs to one and only one group.
- Mapping: real-world users → nodes on P2P
- Establishing a logical connection in P2P: storing the IP address of another user.

Overlay Design: overlay construction

- There are three types of connections in SocialVoD, and the number of connections for each type following: $p : (1-p)q : (1-p)(1-q)$, $p, q \in (0.5, 1)$



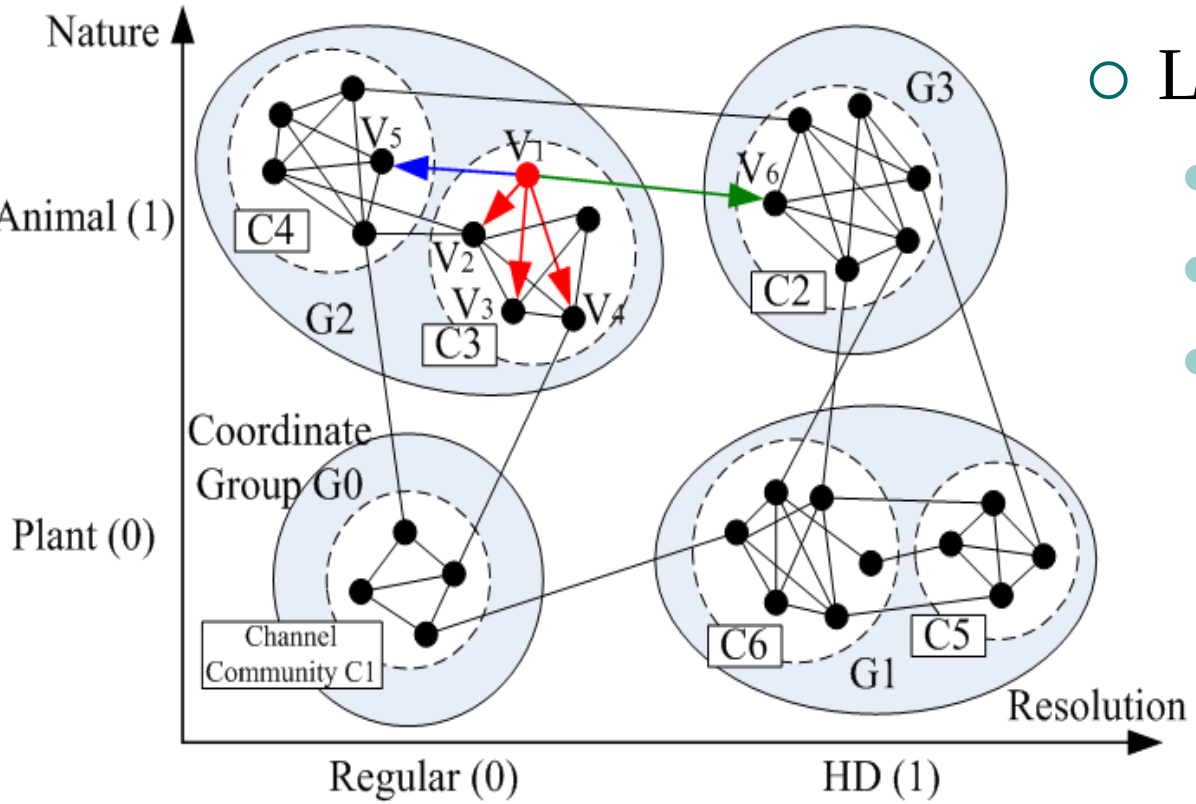
Intra-channel: V_i,V_j subscribes to the same channel, i.e. V_1V_2.

Intra-group: different channel but same coordinate, i.e. V_1V_5.

Inter-group: feature coordinates differ in one dimension, i.e. V_1V_6, where coordinates of V_1 and V_6 are (0,1) and (1,1).

Overlay Construction (cont.)

- SocialVoD creates a two-tier community structure and three levels of overlays. Different types of queries are conducted by different overlays.



- Link management
  - New node joining
  - Link updating
  - SocialVoD gives more emphasis on how to reach a destination, instead of where the destination is.

Routing in SocialVoD

- A query is conducted by two steps in SocialVoD:
  - Feature-spaced inter-group routing: use the highest-level overlay to locate the group with the specific feature coordinate.
  - Intra-group/channel searching: apply the conventional random walk-based approach to search the target file within the specific group or channel.
- In order to control the network traffic, each query has a limited number of query copies, assuming h. During inter-group routing, one should avoid the situations where several query messages search the same space.

Coordinate Routing Sequence

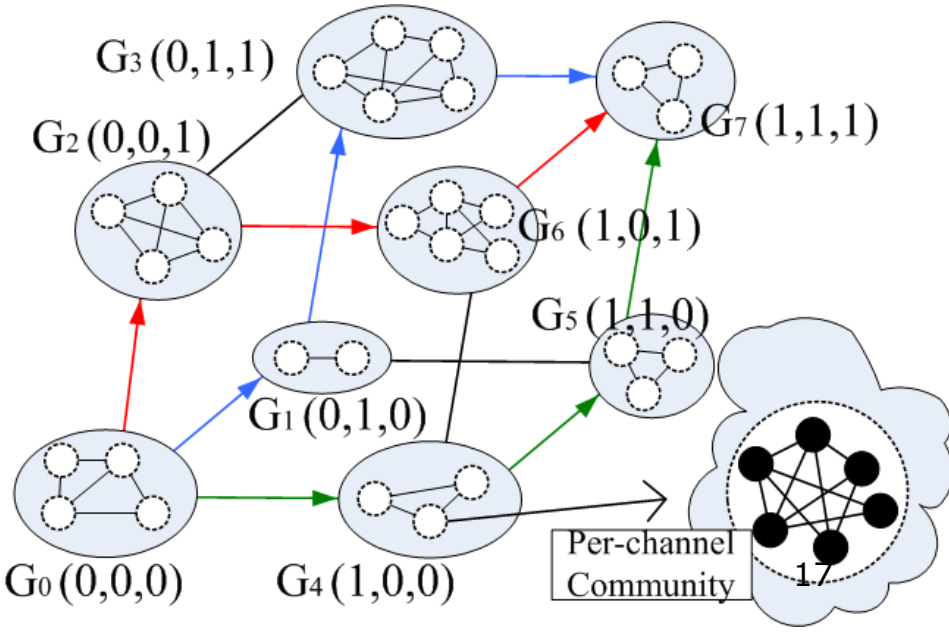
- Feature difference set $H(S, D)$: a set of different dimensions between source S and destination D.
- Coordinate routing sequences: suppose S differs D in k dimensions, $H(S, D) = \{1, 2, ..., k\}$; there are k coordinate sequences forming k node-disjoint paths.

$$R_1 = \langle 1, 2, ..., k-1, k \rangle$$
$$R_2 = \langle 2, 3, ..., k, 1 \rangle$$
$$\cdots\cdots$$
$$R_k = \langle k, 1, ..., k-2, k-1 \rangle$$

- For example:

$R_2 = \langle 2, 3, 1 \rangle$: blue path

$R_3 = \langle 3, 1, 2 \rangle$: red path



The Selection of Coordinate Routing Sequences

- When h<k, the source S needs to determine which path (or coordinate routing sequence) has more of a chance to reach the destination group.
- We approximate a coordinate routing sequence's success rate by subcategories' popularity:

$$T_i' = (|H|-1)! \prod_{j \in H \setminus \{i\}} T_j, \text{ where } \quad T_i = \frac{|\{v_j | F_i(v_j) = d_i, v_j \in V\}|}{|\{v_j | F_i(v_j) \neq s_i, v_j \in V\}|}$$

- Source node always selects the top-h sequences with the highest success rate. This method essentially explores the rare intergroup links first.

Feature Space Routing

Algorithm 2 Inter-group Routing from Source S

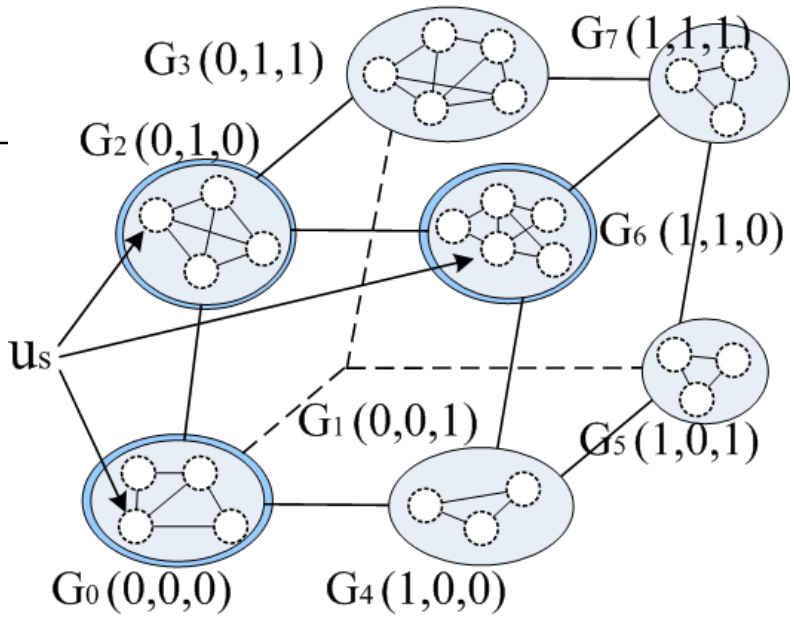
while $h > 0$ **do**
  if $R \neq \phi$ **then**
    /*shortest path routing*/
    Pick up a sequence R_i from R.
    Find $G_j \in N(G_s)$ s.t. $F_i(G_j) \neq s_i, F_i(G_j) = d_i$.
    $R \leftarrow R \backslash \{R_i\}$, $mode \leftarrow 0$, send $(R_i, mode)$ to G_j.
  else
    /*non-shortest path routing*/
    Find $G_j \in N(G_s)$ s.t. $F_i(G_j) \neq s_i, i \in \overline{H}$.
    $\overline{H} \leftarrow \overline{H} \backslash \{i\}$, $mode \leftarrow 1$, send $(H \bigcup \{i\}, mode)$ to G_j
  $h \leftarrow h - 1$

Algorithm 3 Inter-group Routing for a Relay Node in G_i

if $F(G_i) = F(D)$ **then**
  Use intra-group routing within G_i.
else
  if $mode = 0$ **then**
    $R \leftarrow R \backslash R[1]$, the next resolving dimension $k \leftarrow R[1]$.
    Find $G'_i \in N(G_i)$ s.t. $F_k(G_i) \neq d_k, F_k(G'_i) = d_k$.
    Send $(R, mode)$ to G'_i.
  if $mode = 1$ **then**
    $H \leftarrow H \backslash \{j\}$.
    Find $G'_i \in N(G_i)$ s.t. $\exists k \in H, F_k(G_i) \neq d_k, F_k(G'_i) = d_k$. Send $(H, mode)$ to G'_i.

- Source node:
  - Are there redundant searching paths?
  - If not, which coordinate sequences should be used?
- Relay node:
  - Am I in the destination feature group?
  - For a query message, does it use shortest-path mode or non-shortest path mode?
  - Who should I forward the message to?

Extension

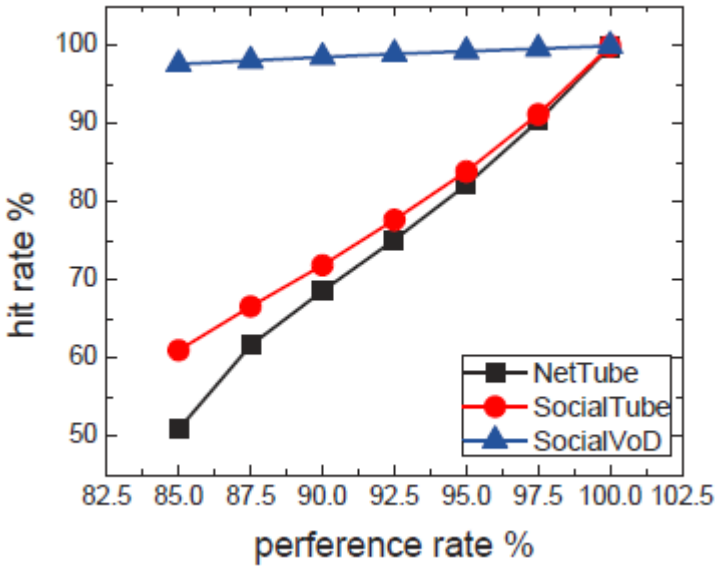


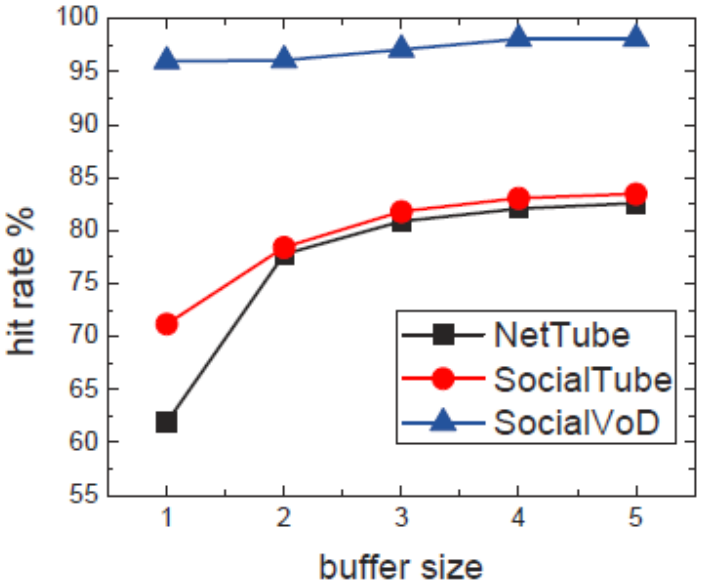
- Multi-channel subscription
  - Initial node selection
  - Feature matching shortcut
- Pre-fetch for new videos
  - Random walk-based propagation
- Routing for non-subscribers
  - Most frequently searched subcategory features
  - Local overlay keeps updating while developing his own watching preference

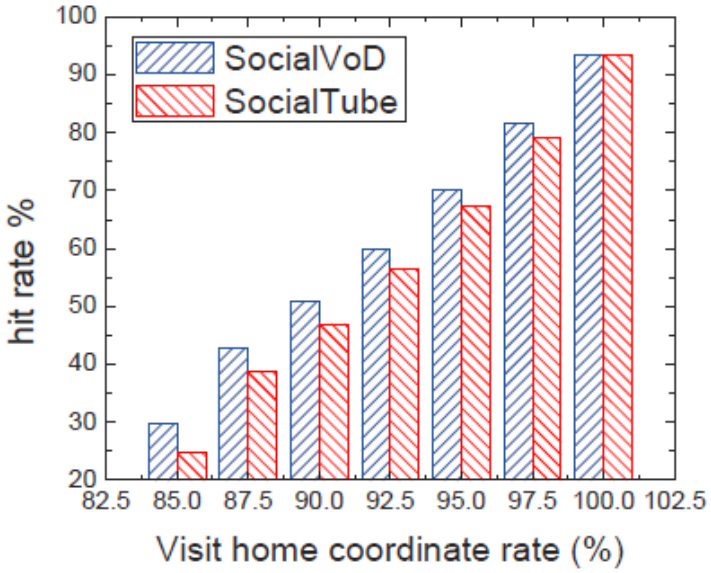
Outline

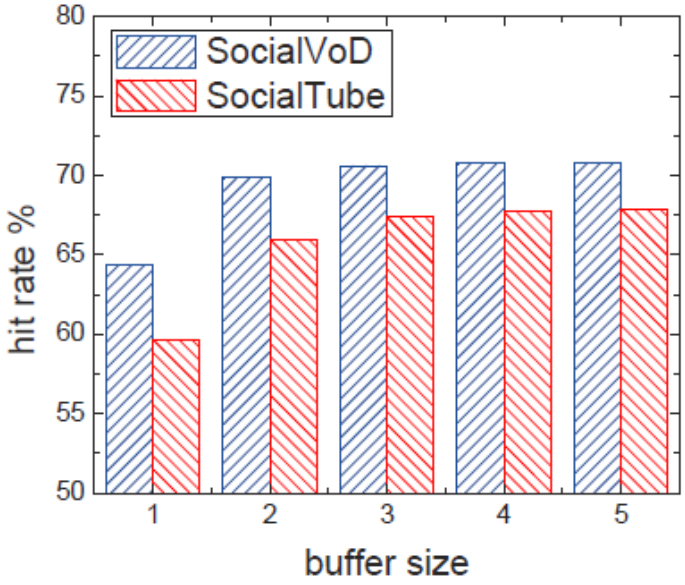
1. Introduction
2. Background and Related Work
3. Vimeo Category and User Interests
4. Overlay Design and File Searching
5. Evaluation
6. Conclusion

Evaluation


hit rate %
perference rate %
NetTube
SocialTube
SocialVoD


hit rate %
buffer size
NetTube
SocialTube
SocialVoD


hit rate %
Visit home coordinate rate (%)
SocialVoD
SocialTube


hit rate %
buffer size
SocialVoD
SocialTube

Outline

1. Introduction
2. Background and Related Work
3. Vimeo Category and User Interests
4. Overlay Design and File Searching
5. Evaluation
6. Conclusion

Conclusion

- We find that the existing P2P systems do not fully utilize the social properties behind the users' watching patterns.
- Based on the existing channel subscription knowledge and category structure on the Vimeo platform, we introduce a new hierarchical P2P overlay structure.
- Our new system explores the existing category information as an overlay's construction "guide" and routing "hints", which significantly reduces the searching space of videos.
- We provide extensive simulations to show that our new system can efficiently locate the files and is applicable in a large scale.

Thank you!