Predicting the Popularity of Web 2.0 Items based on User Comments

Xiangnan He, Ming Gao, Min-Yen Kan, Yiqun Liu and Kazunari Sugiyama

National University of Singapore
Singapore Management University
Tsinghua University

User Generated Content: A driving force of Web 2.0

Challenges:

- ➢ Information overload [1]
- ➢ Dynamic, temporally evolving Web
- ➢ Rich but noisy UGC

Daily growth of UGC:

- Twitter: 500+ million tweets
- Flickr: 1+ million images
- YouTube: 360,000+ hours of videos

[1] Xiangnan He et al. Comment-based Multi-view Clustering of Web 2.0 Items. In *Proc. of WWW 2014*.

Dynamic, temporally evolving Web – Challenges in Web search ranking

- Illustrative Example:

Querying "The Voice of China" on 2013/7/24

(A Chinese reality talent show started in 2012 – 1 season/year)

[Top 3 results of Google constrained in YouTube domain]

9.5 K

7.4 K

72

Top results are all old popular videos of the last season, only attract **less than 10k views** in future 3 days.

First result of the new season → Ranked 16th, but extremely popular (more than 100k views)

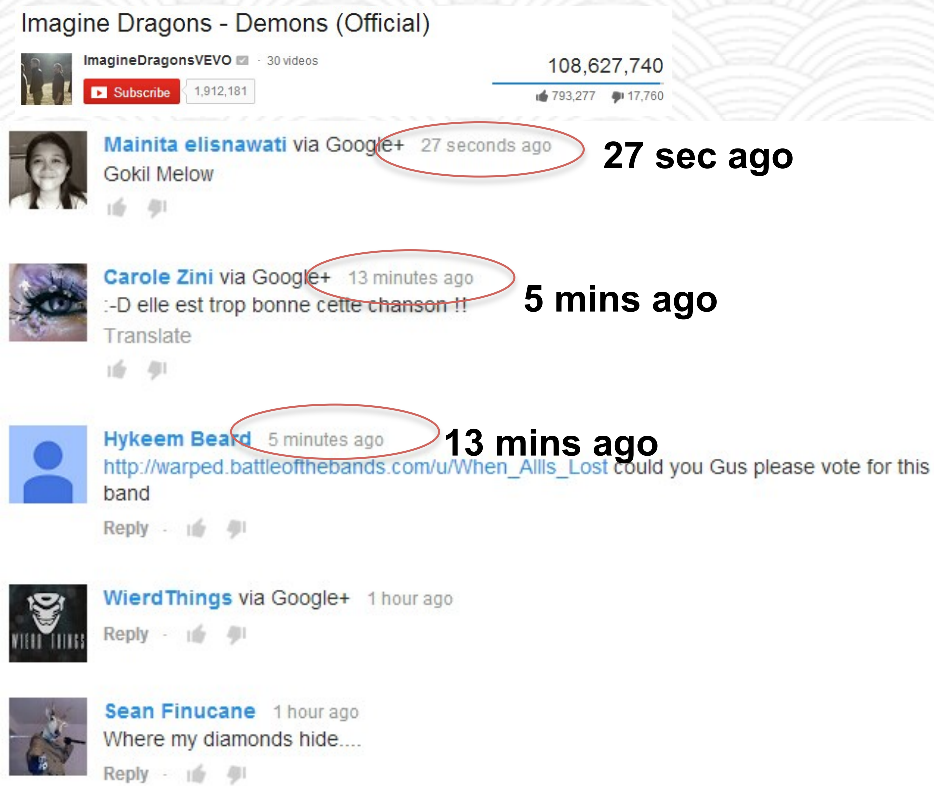
Why Popularity Prediction?

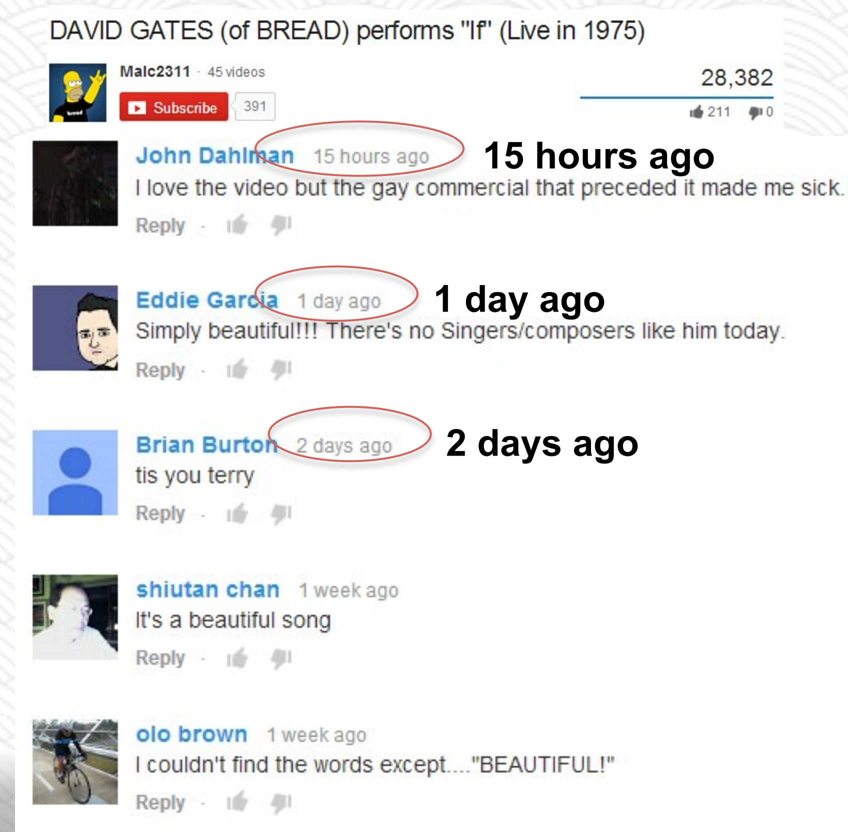
Why Popularity Prediction?

- Traditional solutions - mining the view histories of items.
- However, it is not easy to perform prediction when one is not the content providers:
  - View histories are cost to build (need repeated crawling)
- Our proposal -- predicting popularity (view # as metric) based on user comments, which are more easily accessible than views.

Why user comments?

- Comments contain signal of item's future popularity:
  - Commenting timestamps.
  - Commenting users.
  - Textual comments.





Comments Vs. Views

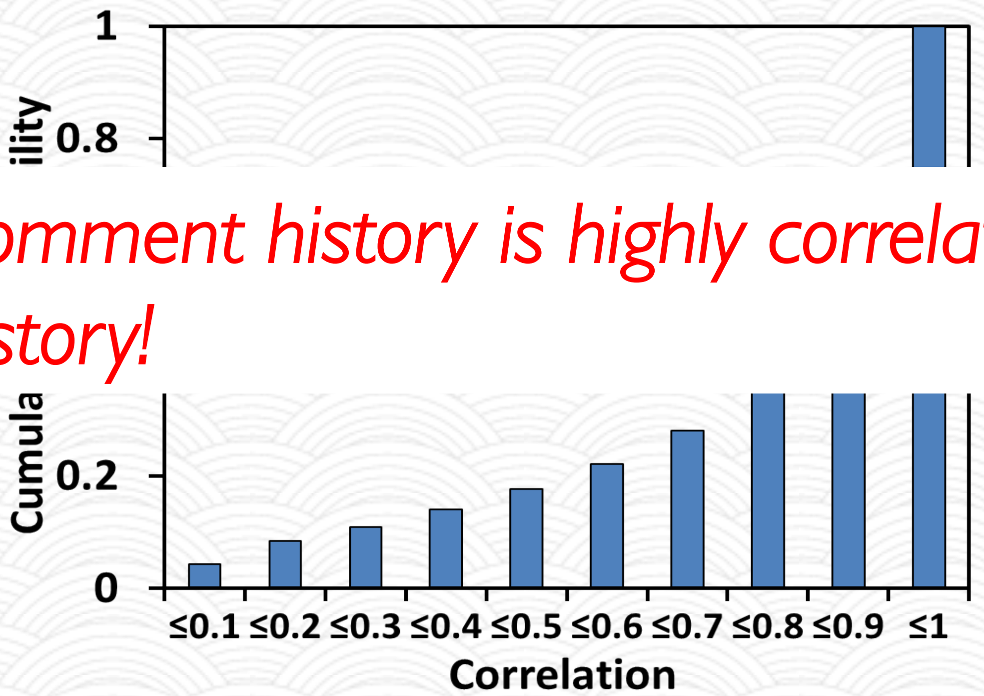
- Intuitively, comment series should have correlation with view series.

A sample video's statistics in YouTube

- Q1: Can *comment series* be used to replace *view series* for prediction?
- Q2: How the past user comments contribute to future popularity?

Correlation of Comments and Views

- Q1: Can *comment series* be used to replace *view series* for prediction?



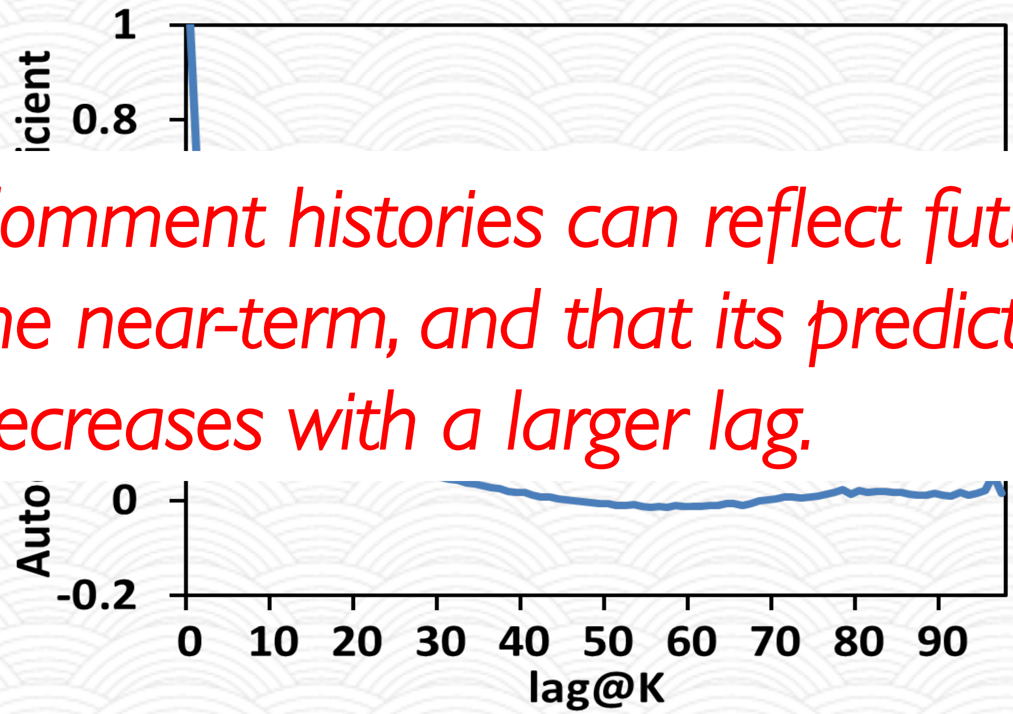
Mean = 0.76

Comment history is highly correlated with view history!

CDF of videos with respect to their comments-views correlation.

Comment Series Autocorrelation

- Q2: How past user comments contribute to future popularity?



Comment histories can reflect future popularity in the near-term, and that its predictive ability decreases with a larger lag.

Autocorrelation of comment series

Prediction Based on Comment Series

- Intuitive Solution: adopt time series prediction methods (e.g. regression) on comment series.

- Problem: Sparsity!!
  - Many items have no comments at particular time unit.

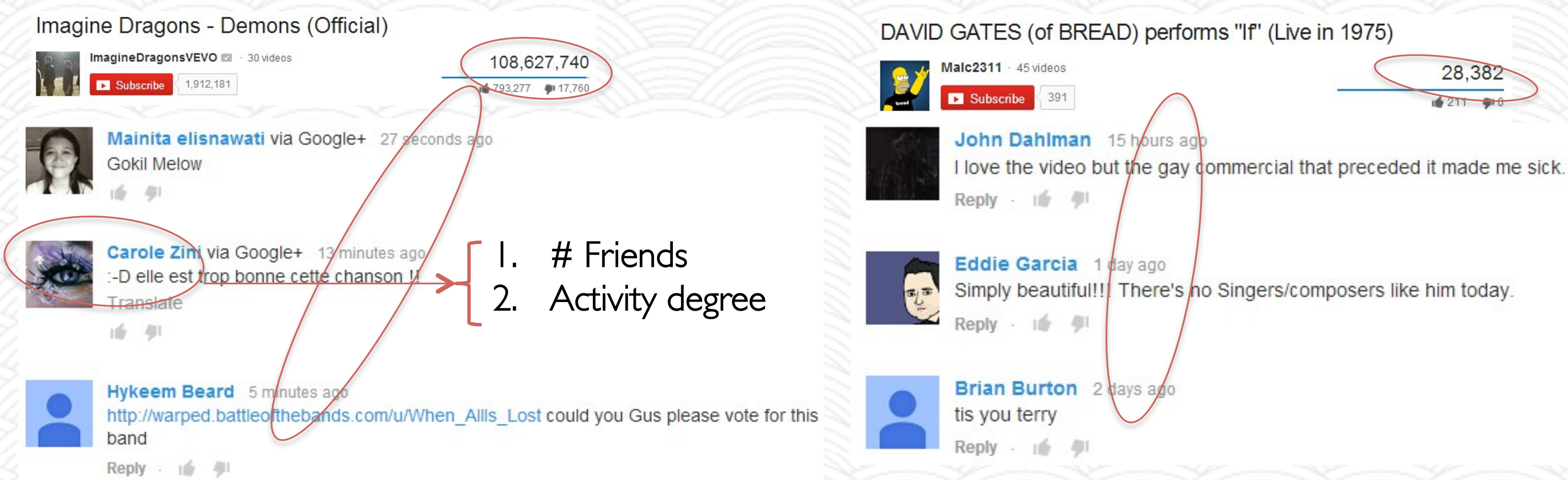
- We need to incorporate more SIGNALs for quality prediction!

Outline

- Goal and Motivation
- Preliminary analysis
  - Correlation analysis of comments and views
  - Autocorrelation analysis of comment series
- Proposed Method
  - Hypotheses on comment-based prediction
  - Bipartite User-Item Ranking (BUIR)
- Experiments
- Conclusion

Hypotheses on Comment-based Prediction

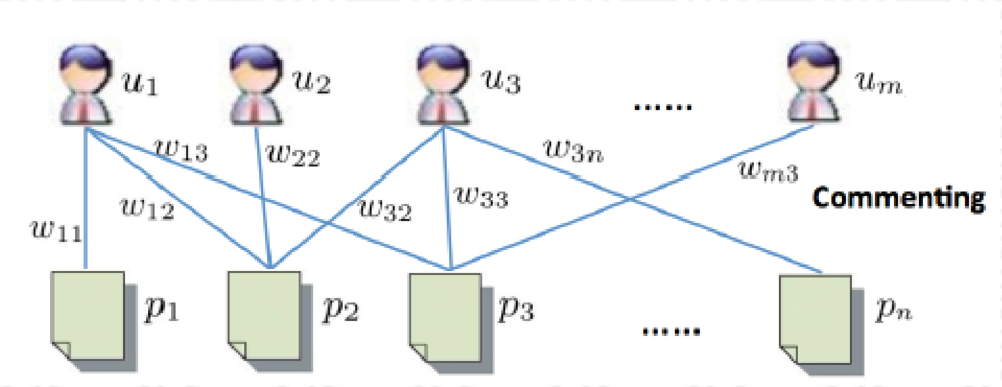
- **H1. Temporal factor:** More recent comments -> More likely to be popular;
- **H2. Social Influence factor:** More influential the commented users -> More likely to be popular [4];
- **H3. Current Popularity factor:** More current popularity is -> More likely to be popular ("rich-get-richer" effect).



[4] K. Lerman and T. Hogg. Using a model of social dynamics to predict popularity of news. In *Proc. of WWW 2010.*

Proposed Solution – *BUIR*

- Bipartite User-Item Ranking:
  - Modeling user comments as a bipartite graph;
  - Ranking items by capturing the three hypotheses (i.e. ranking by predicted popularity [2]).



Example: Bipartite User-Item Structure

Edge weight:

$$w_{ij} = \delta^{a(t_0 - t_{ij}) + b}$$

[2] Peifeng Yin et al. A straw shows which way the wind blows: ranking potentially popular items from early votes. In *Proc. of WSDM 2012*.

BUIR – *Regularization framework*

- Devising regularizers for three hypotheses:
  - H1. Temporal factor (more users commented on recently)
  - H2. Social influence factor (more influential users)
  - H3. Current popularity factor (more popular now)
- Capturing H1 & H2:
  - If an item is *recently* commented by *many influential* users, it should be ranked high.

$$\frac{1}{2}\eta \sum_{j=1}^{|P|} \sum_{i=1}^{|U|} \boxed{w_{ij}} \left(\frac{1}{\sqrt{d_j^p}} p_j - \boxed{\frac{1}{\sqrt{d_i^u}} u_i} \right)^2$$

BUIR – Regularization framework

- Devising regularizers for three hypotheses:
  - H1. Temporal factor (more users commented on recently)
  - H2. Social influence factor (more influential users)
  - H3. Current popularity factor (more popular now)
- Capturing H2 & H3:

$$\alpha \sum_{j=1}^{|P|} (p_j - \boxed{p_j^0})^2 + \beta \sum_{i=1}^{|U|} (u_i - \boxed{u_i^0})^2$$

Item's initial score

$$p_j^0 = \frac{\log v_j}{\sum_{k=1}^{|P|} \log v_k}$$

User's initial score

$$u_i^0 = \frac{\log(1+g_i)}{\sum_{k=1}^{|U|} \log(1+g_k)}$$

BUIR – Iterative solution

- Regularization function to minimize:

$$R = \frac{1}{2}\eta \sum_{j=1}^{|P|}\sum_{i=1}^{|U|} w_{ij}\left(\frac{1}{\sqrt{d_j^p}}p_j - \frac{1}{\sqrt{d_i^u}}u_i\right)^2 + \alpha\sum_{j=1}^{|P|}(p_j - p_j^0)^2 + \beta\sum_{i=1}^{|U|}(u_i - u_i^0)^2$$

- Alternating optimization:
  - Iterative updating rules:

$$p_j = \frac{2\alpha}{\eta + 2\alpha}p_j^0 + \frac{\eta}{\eta + 2\alpha}\sum_{i=1}^{|U|}\frac{w_{ij}u_i}{\sqrt{d_j^p}\sqrt{d_i^u}}$$

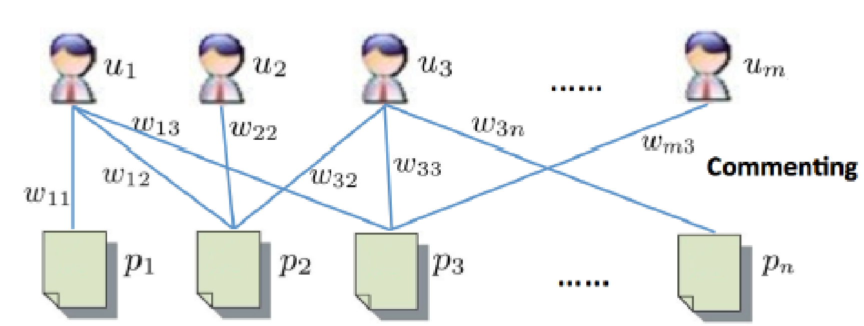
$$u_i = \frac{2\beta}{\eta + 2\beta}u_i^0 + \frac{\eta}{\eta + 2\beta}\sum_{j=1}^{|P|}\frac{w_{ij}p_j}{\sqrt{d_j^p}\sqrt{d_i^u}}$$

  - Guarantee to find the global minima (the Hessian is positive semi-definite).

Interpretation of BUIR

- Matrix form of the iterative solution:

$$\mathbf{p} = \frac{1}{1+2\alpha}\mathbf{S}_w^T\mathbf{u} + \frac{2\alpha}{1+2\alpha}\mathbf{p}_0,$$
$$\mathbf{u} = \frac{1}{1+2\beta}\mathbf{S}_w\mathbf{p} + \frac{2\beta}{1+2\beta}\mathbf{u}_0.$$



  - where $S_w = [\frac{w_{ij}}{\sqrt{d_j^p}\sqrt{d_i^u}}]_{m\times n}$
- Mutual reinforcement between users and items:
  - Comment by a user increases the target item's score;
  - The item increases the user's score (n.b. activity degree).
- Random walk in the bipartite graph
  - Can be seen as a variant of PageRank

Outline

- Goal and Motivation
- Preliminary analysis
- Proposed Method
- Experiments
  - Overall Evaluation
  - Query-specific Evaluation
  - Tiered Popularity Evaluation
- Conclusion

Experiments - Settings

- Datasets:
  - Search results of 10 queries.

Dataset	# Item	# Comment	# User	Avg C:I
YouTube	21,653	7,246,287	3,620,487	334.7
Flickr	26,815	169,150	37,690	6.3
Last.fm	16,284	530,237	77,996	32.6

  - 10%: Parameter tuning in regularization, 90%: Testing.
- Crawled on two dates:
  - Initial date (t_0) and Evaluation date (t_0 + 3)
  - Ground-truth is the #view received between the two dates.
- Evaluation metrics:
  - Spearman coefficient and NDCG@10 (query-specific evaluation)

Dataset will be available soon in my homepage: http://www.comp.nus.edu.sg/~xiangnan/

Experiments - Baselines

- Compare with 5 methods:
  - VC: Rank based on current View Count (corresponds to H3).
  - CCP: Comment Count in the Past 3 days (corresponds to H1).
  - CCF: Comment Count in the Future 3 days (oracular method with access to future comments).
  - ML: Multivariate Linear regression model proposed by Pinto et al. 2013 [3] (current state-of-the-art method).
  - PR: PageRank (with personalized vectors) in the user-item graph.

[3] Henrique Pinto et al. Using Early View Patterns to Predict the Popularity of YouTube Videos. In *Proc. of WSDM 2013*.

Overall Evaluation

Spearman coefficient (%) of ranking all items

	YouTube	Flickr	Last.fm
VC	73.39	58.42	67.31
CCP	83.35	59.43	67.21
CCF	84.53	59.41	67.20
ML	78.24	58.00	38.09
PR	80.72	28.15	10.24
BUIR	**87.72****	**64.60****	**70.43****

1. BUIR performs best in all datasets ($p < 0.01$).

2. VC obtains good performance, indicating effectiveness of H3

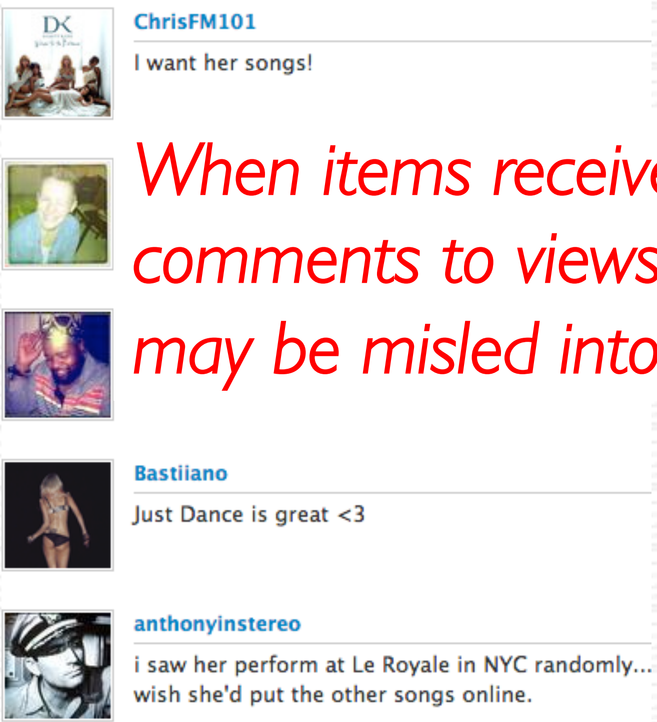
3. Difference between CCF and CCP are insignificant.

4. ML does not perform well:
   - ➢ Short-term prediction;
   - ➢ Optimization criterion (mRSE VS. Ranking)

5. Separately handling two vertex types in bipartite graph is important!

Case Study of Top Rankings

- Abnormal items in top rankings:
  - "Lady Gaga" and "Madonna", ranked at 4th and 7th by BUIR, but their true rank is 170th and 178th, respectively.



When items receive uneven high ratio of comments to views, our comment-based method may be misled into incorrect rankings.

Comments of Lady Gaga in Last.fm

Query-specific Evaluation I

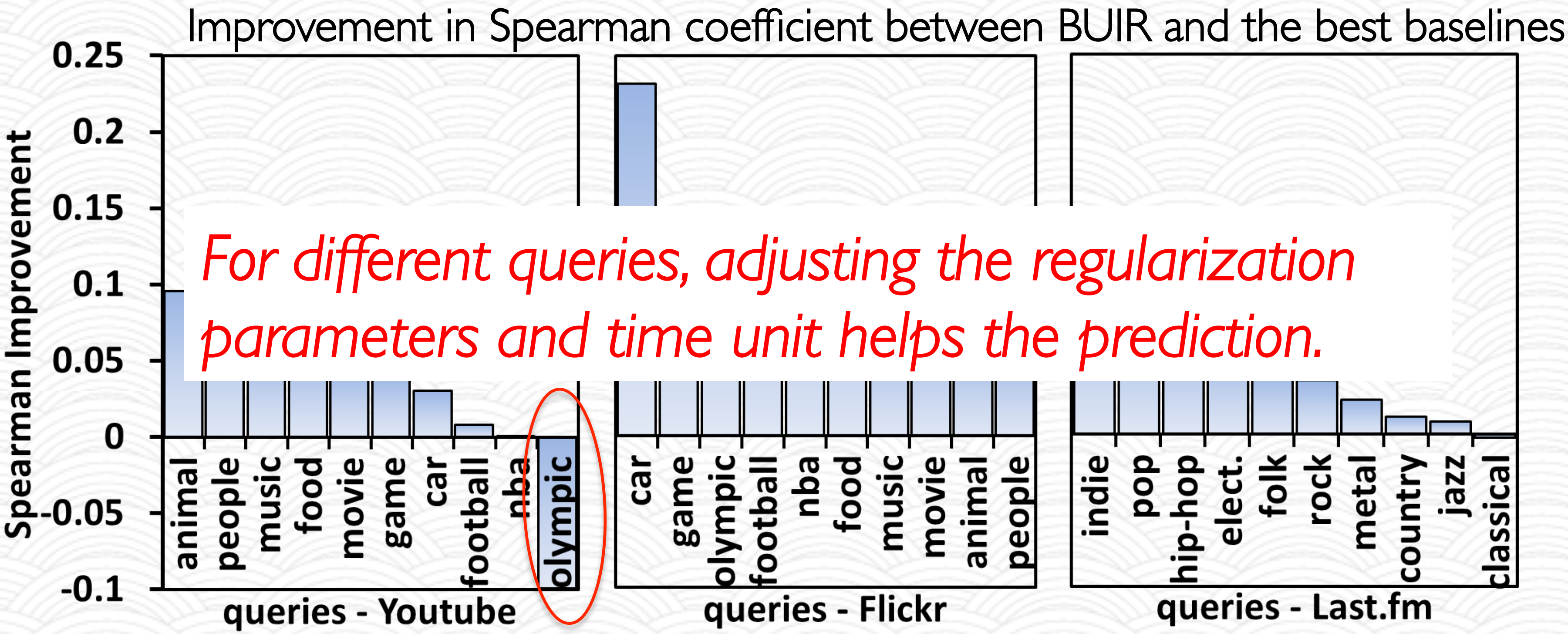
NDCG@10 (mean ± standard deviation) of 10 queries

	YouTube	Flickr	Last.fm
VC	**64.70±22.23***	**67.19±15.75***	**90.25±4.96***
ML	27.85±30.76	50.74±18.64	74.30±11.15
PR	61.10±21.92	54.53±22.62	81.16±10.07
BUIR	**76.13±12.29***	**74.19±15.70***	**88.19±4.68***

Current View Count is a good prediction indicator for most popular items!

* denotes the statistical significance for $p < 0.05$

Query-specific Evaluation II



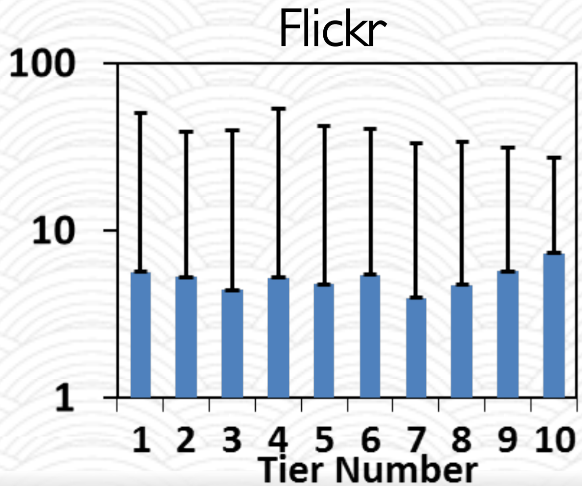
For different queries, adjusting the regularization parameters and time unit helps the prediction.

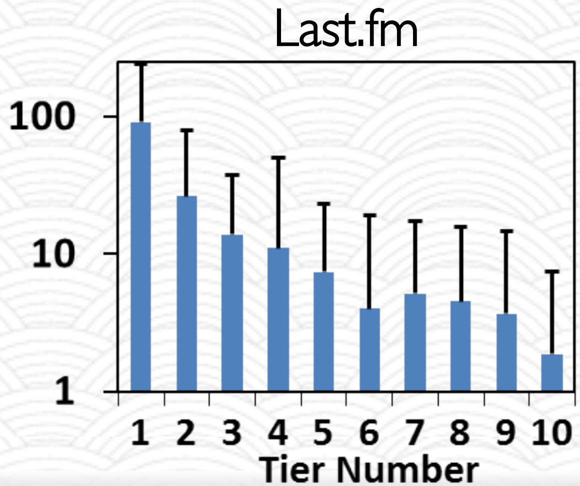
Reasons:

1. London Olympic event – users commented according to their country's medaling – H2 (social influence factor) does not hold.
2. Freshness – for these new videos, when we change the time unit to hourly basis, our method improves.

Tiered Popularity Evaluation

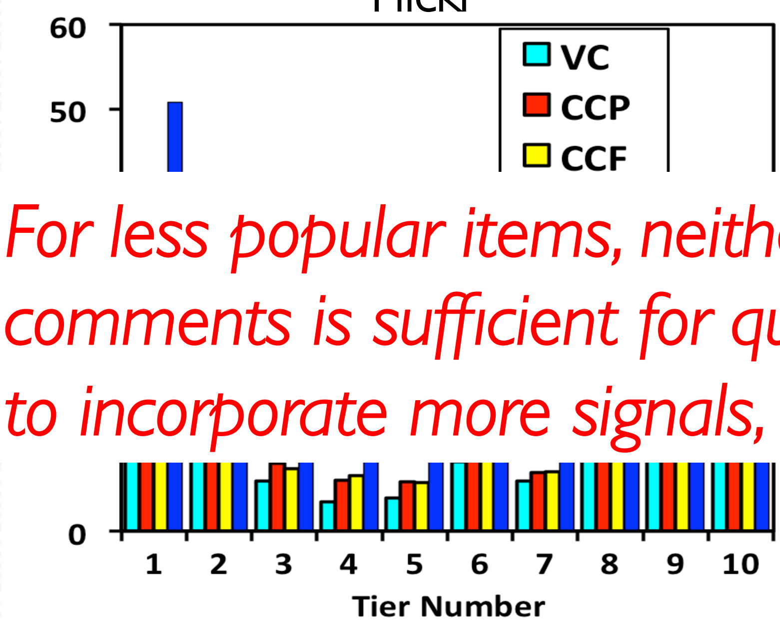
- Experimental Settings
  - Step 1: Sort the items by descending view count at the ranking time;
  - Step 2: Split items into ten equal-sized subsets: Tier-1 (most popular) to Tier-10 (least popular).
- Comment statistics of the ten popularity tiers:



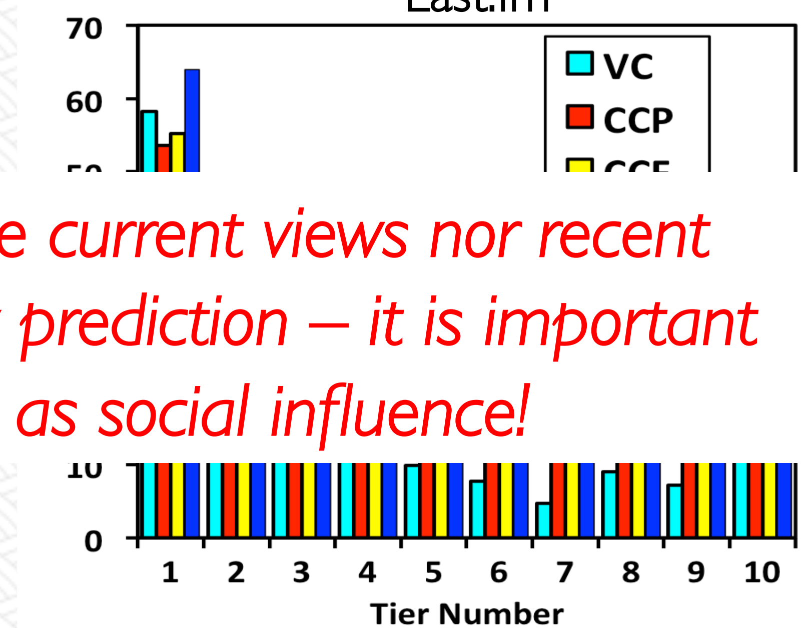


Tiered Popularity Evaluation

Flickr



Last.fm



For less popular items, neither the current views nor recent comments is sufficient for quality prediction – it is important to incorporate more signals, such as social influence!

1. BUIR consistently performs better, and the improvement over CCP and CCF are more noticeable for high tiers (less popular items);
2. VC predicts well for popular items, but suffers a lot for less popular items.
3. CCF does not always outperform CCP, although CCF utilizes future knowledge, indicating the limitation of simply using comment count for prediction.

Hypotheses Study

Performance decrease of different parameter settings

	YouTube	Flickr	Last.fm
[illegible]	[illegible]	[illegible]	[illegible]
$\alpha, \beta = 0$	51.24 (-42 %)	53.77 (-17 %)	47.22 (-33 %)

Every factor captured in BUIR — H1, H2 and H3 — is necessary for high-quality popularity prediction based on user comments.

Conclusion and Future Work

- Systematically studied how to best utilize user comments for predicting popularity of Web 2.0 Items.
  - ✓ H1. Temporal factor (fundamental assumption)
  - ✓ H2. Social Influence factor (good signal for less popular items)
  - ✓ H3. Current popularity factor (good signal for popular items)
- Proposed BUIR ranking algorithms for bipartite graphs:
  - ✓ Convergence and global optimum guaranteed.
  - ✓ Easily extended to incorporate more hypotheses.
- Future work:
  - Can comment content (relevance and sentiment) aid prediction?
  - Operationalize our comment-based prediction and clustering (see my WWW'14 work) into contextual advertising and recommender system.

ADDITIONAL SLIDES

Query-specific Evaluation I

Spearman coefficient (mean ± standard deviation) of 10 queries

	YouTube	Flickr	Last.fm
VC	71.98±14.14	46.72±7.82	67.86±5.76
CCP	82.41± 2.50	48.06±7.90	66.97±4.70
CCF	**83.42±2.7***	48.12±7.80	67.27±4.45
ML	76.95± 5.50	50.00±6.50	39.15±4.04
PR	79.66± 4.72	27.80±14.87	9.22 ±11.66
BUIR	**85.98±5.92***	**55.22± 6.10***	**70.42±4.43***

"*" denotes the statistical significance for $p < 0.05$.

References

- [1] Xiangnan He et al. Comment-based Multi-view Clustering of Web 2.0 Items. In *Proc. of WWW 2014*.
- [2] Peifeng Yin et al. A straw shows which way the wind blows: ranking potentially popular items from early votes. In *Proc. of WSDM 2012.*
- [3] Henrique Pinto et al. Using Early View Patterns to Predict the Popularity of YouTube Videos. In *Proc. of WSDM 2013*.
- [4] K. Lerman and T. Hogg. Using a model of social dynamics to predict popularity of news. In *Proc. of WWW 2010.*