Tumblr Blog Recommendation with Boosted Inductive Matrix Completion

Donghyuk Shin
Dept of Computer Science
UT-Austin

ACM CIKM'15
Melbourne, Australia
Oct 20, 2015

Joint work with S. Cetintas, K. Lee and I. S. Dhillon

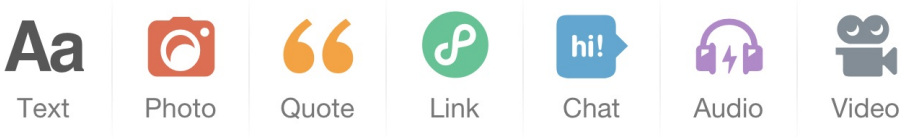
Thanks to the SIGIR Student Travel Grant!

Tumblr Blog Recommendation

Tumblr

Microblogging service:

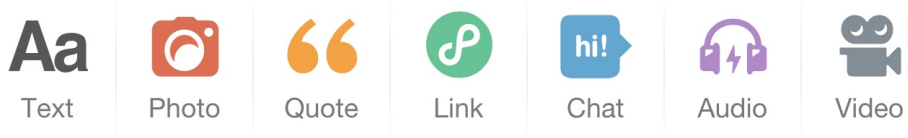
- No limit on length of posts.
- Supports various types of posts:



Tumblr

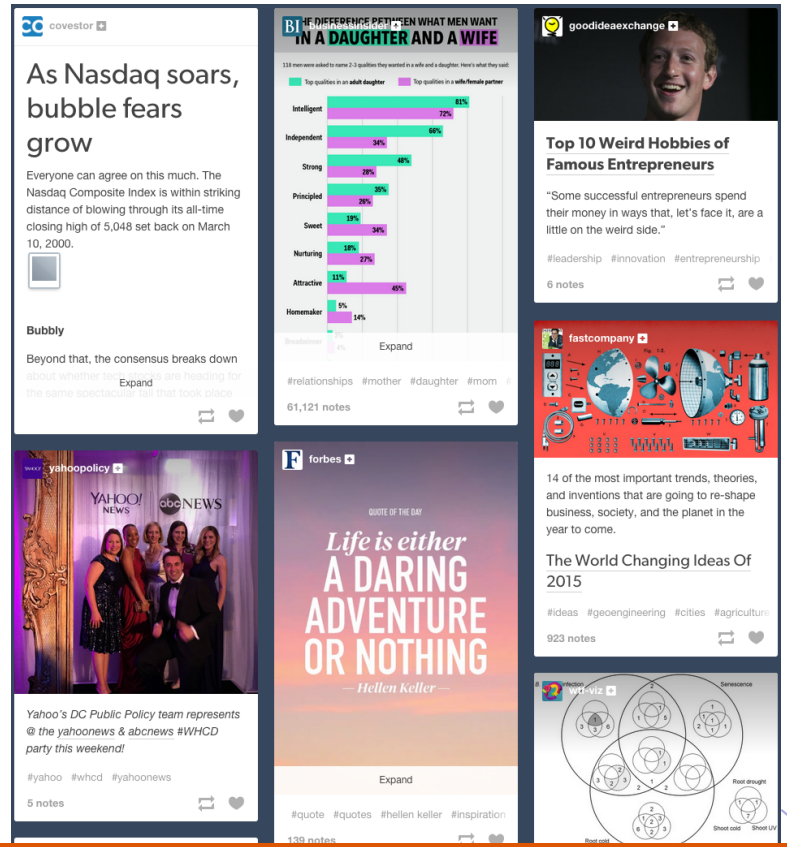
Microblogging service:

- No limit on length of posts.
- Supports various types of posts:



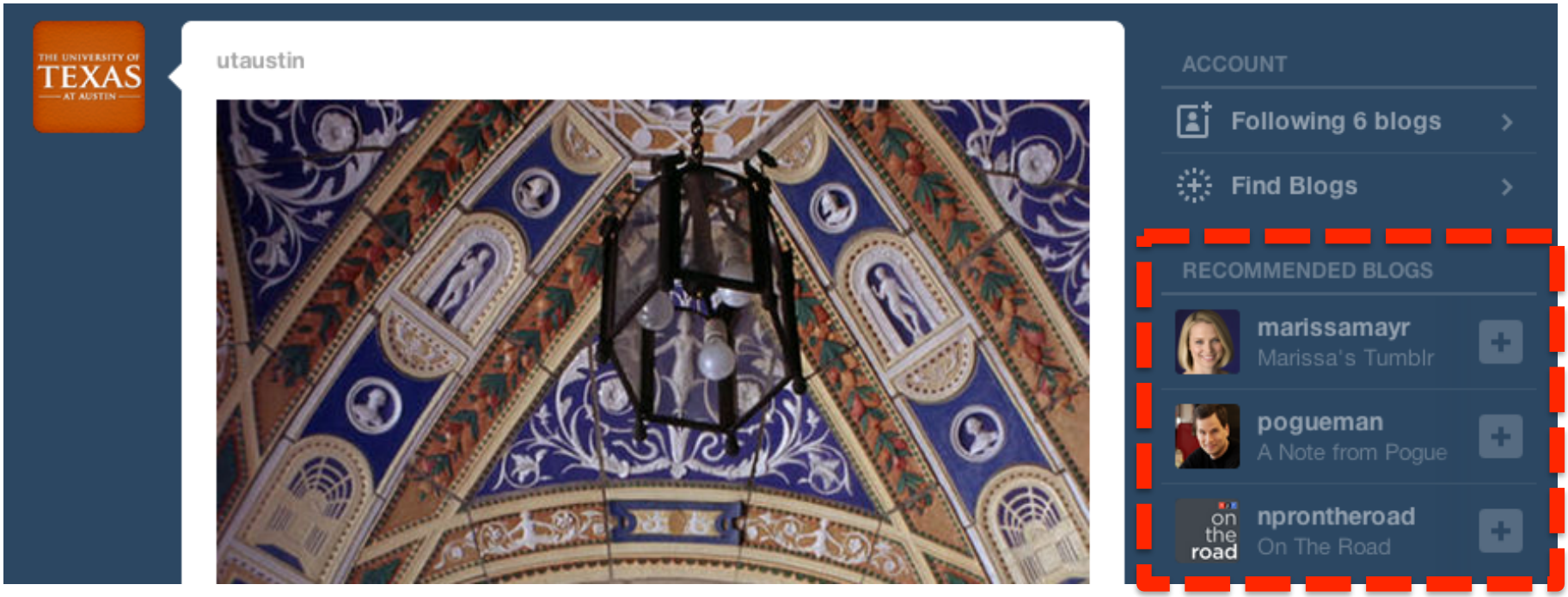
Social network service:

- *Follow* other blogs – 260M blogs
- Like or reblog other posts – 122B posts



Blog Recommendation

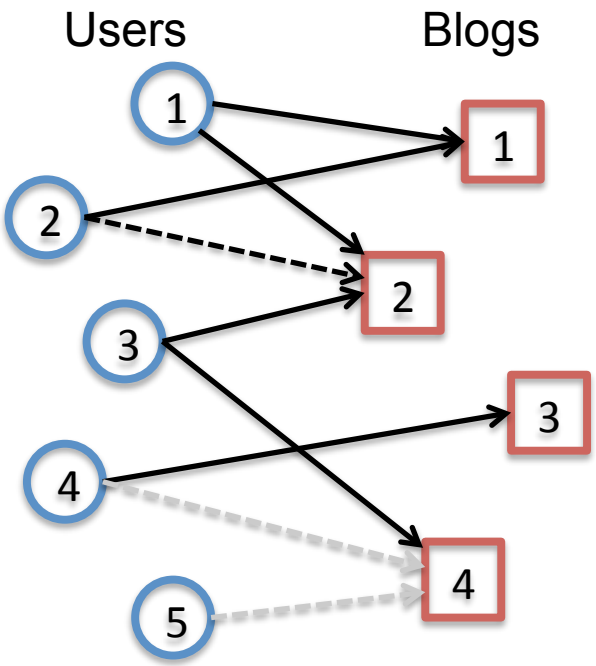
Who to follow?

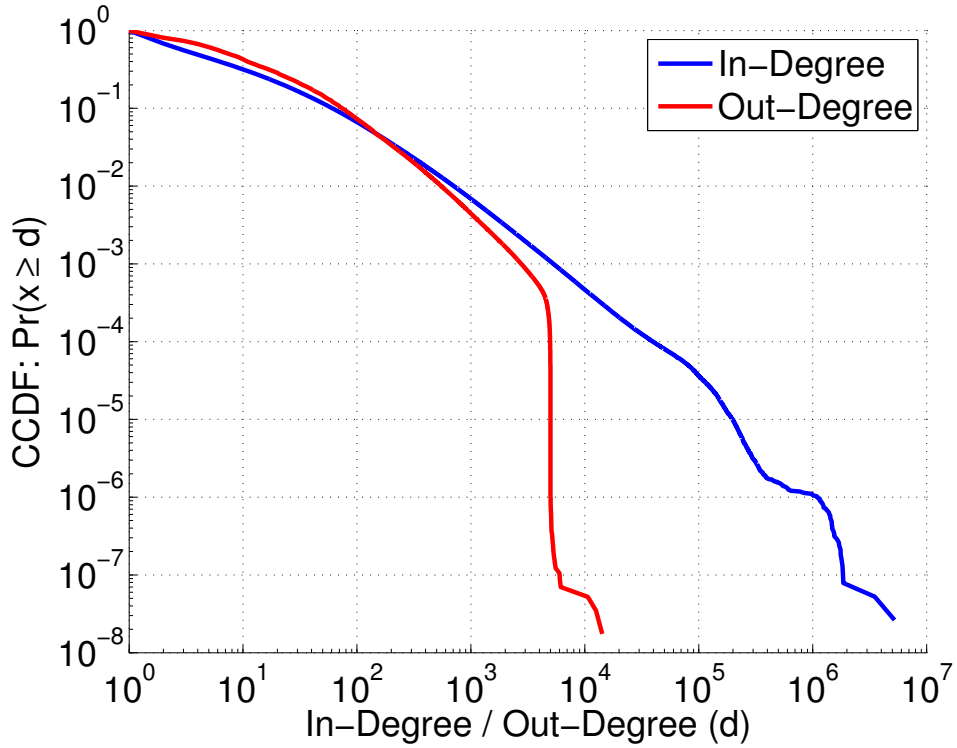


- Help users find content they want $\Rightarrow$ enhance user experience and engagement
- Help increase followers for sponsored or advertiser blogs

Follower Graph

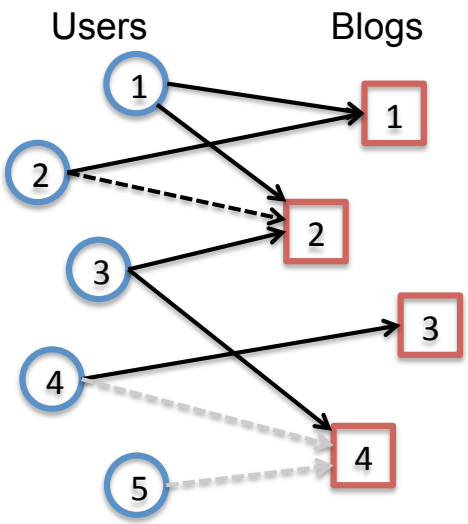
- Snapshot from Jun. 2014
- Directed graph: 76.86M nodes (users/blogs) and 2.27B edges (follows) with timestamps
- 50% have 0 in-degree, 25% have 0 out-degree

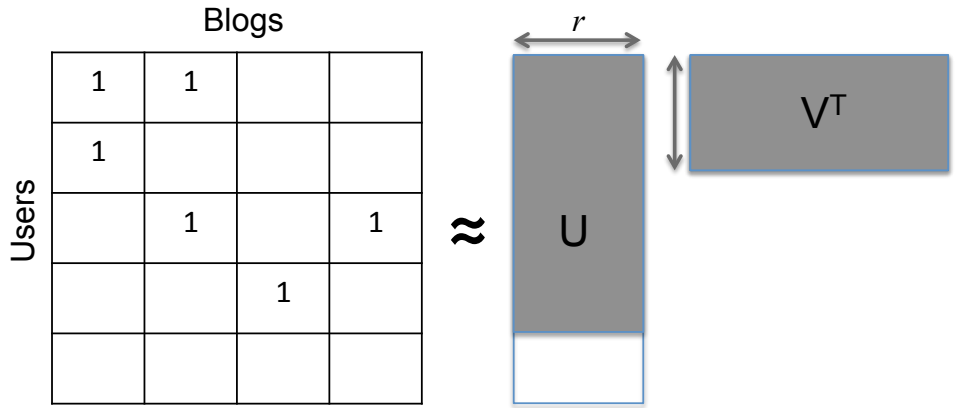




Existing Methods

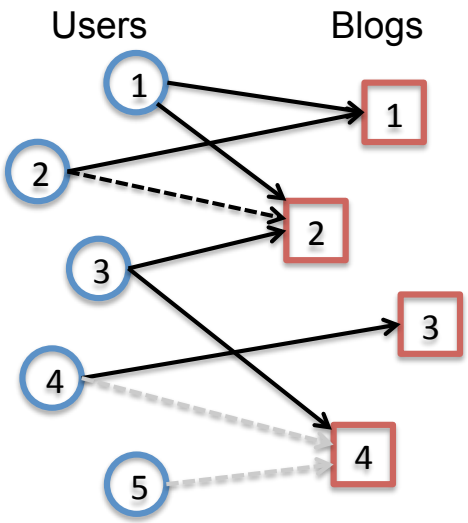
- Global ranking: recommend most popular blogs
- Graph proximity: Common Neighbors, Katz, etc
- Standard matrix completion
- ...

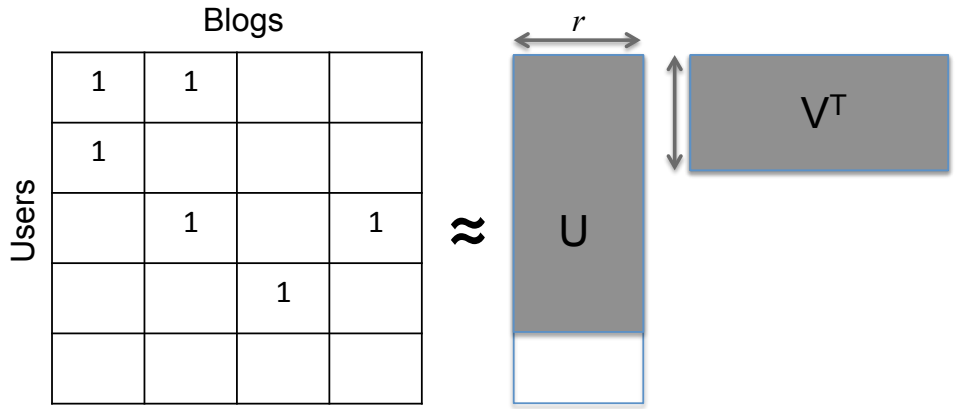




Existing Methods

- Global ranking: recommend most popular blogs
- Graph proximity: Common Neighbors, Katz, etc
- Standard matrix completion
- ...





Can NOT make meaningful recommendations for users or blogs with no follow information (e.g., user 5)

Additional Information from Posts

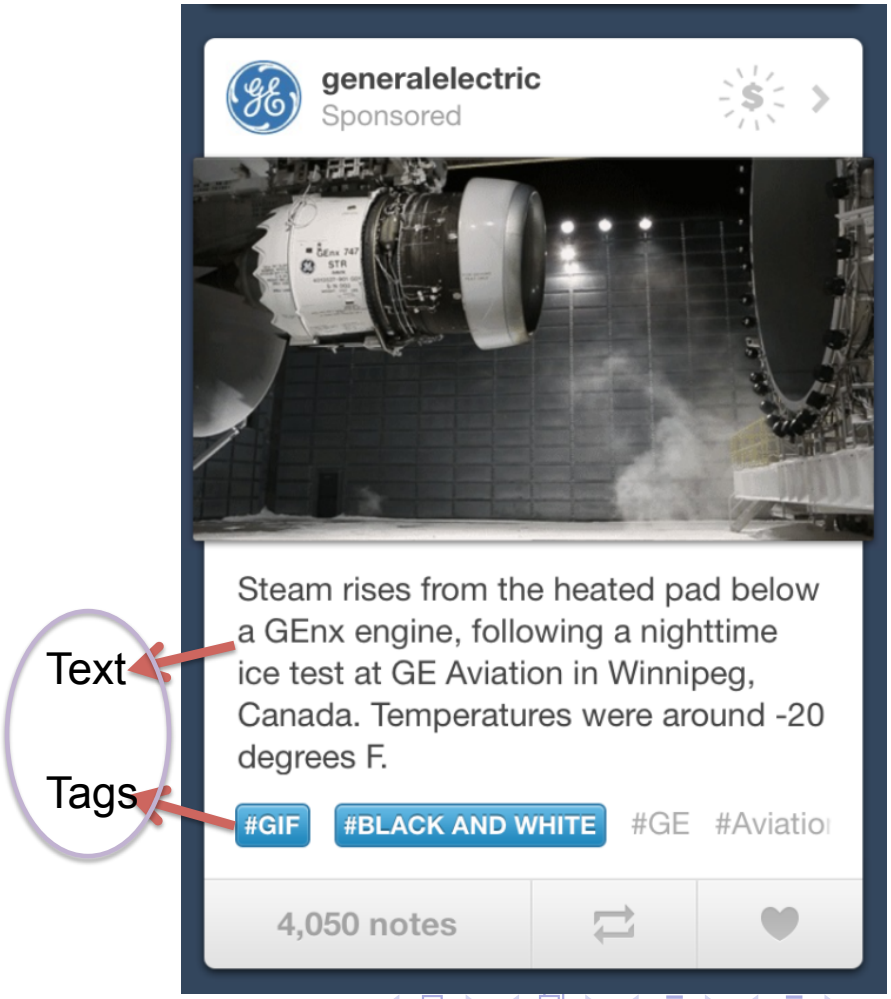
Tumblr Posts

Various sources of additional information:

Tumblr Posts

Various sources of additional information:

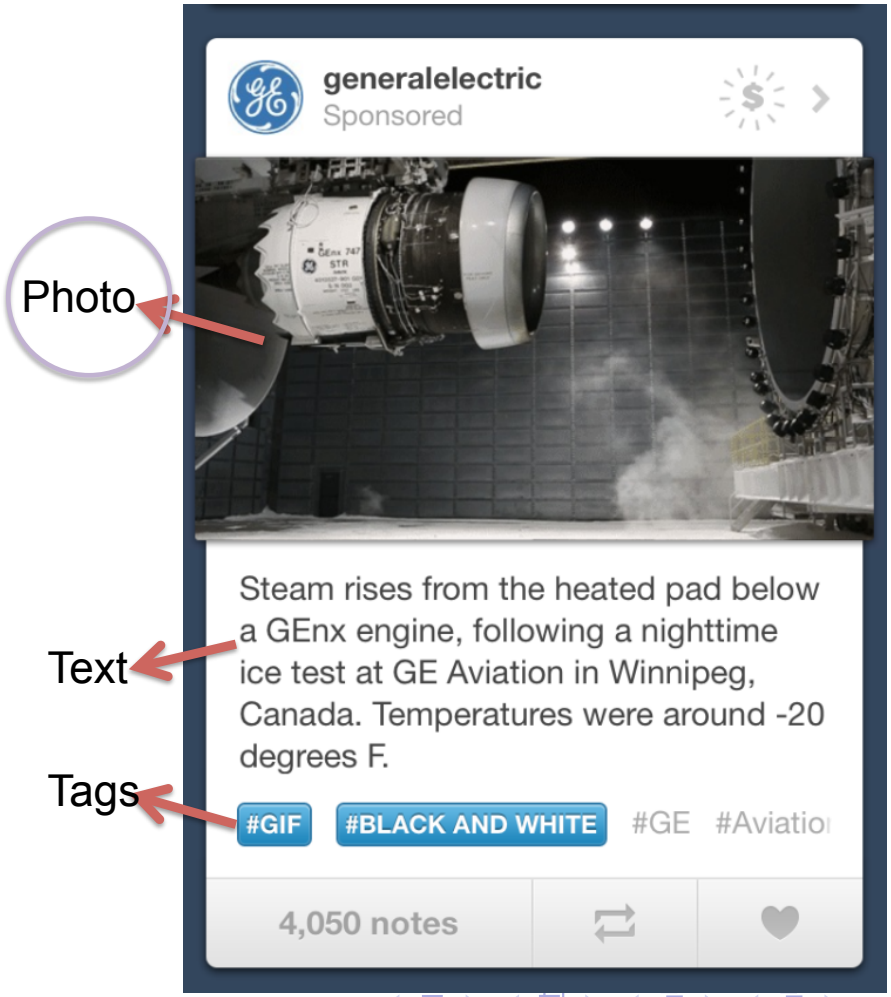
- Textual: text and tags



Tumblr Posts

Various sources of additional information:

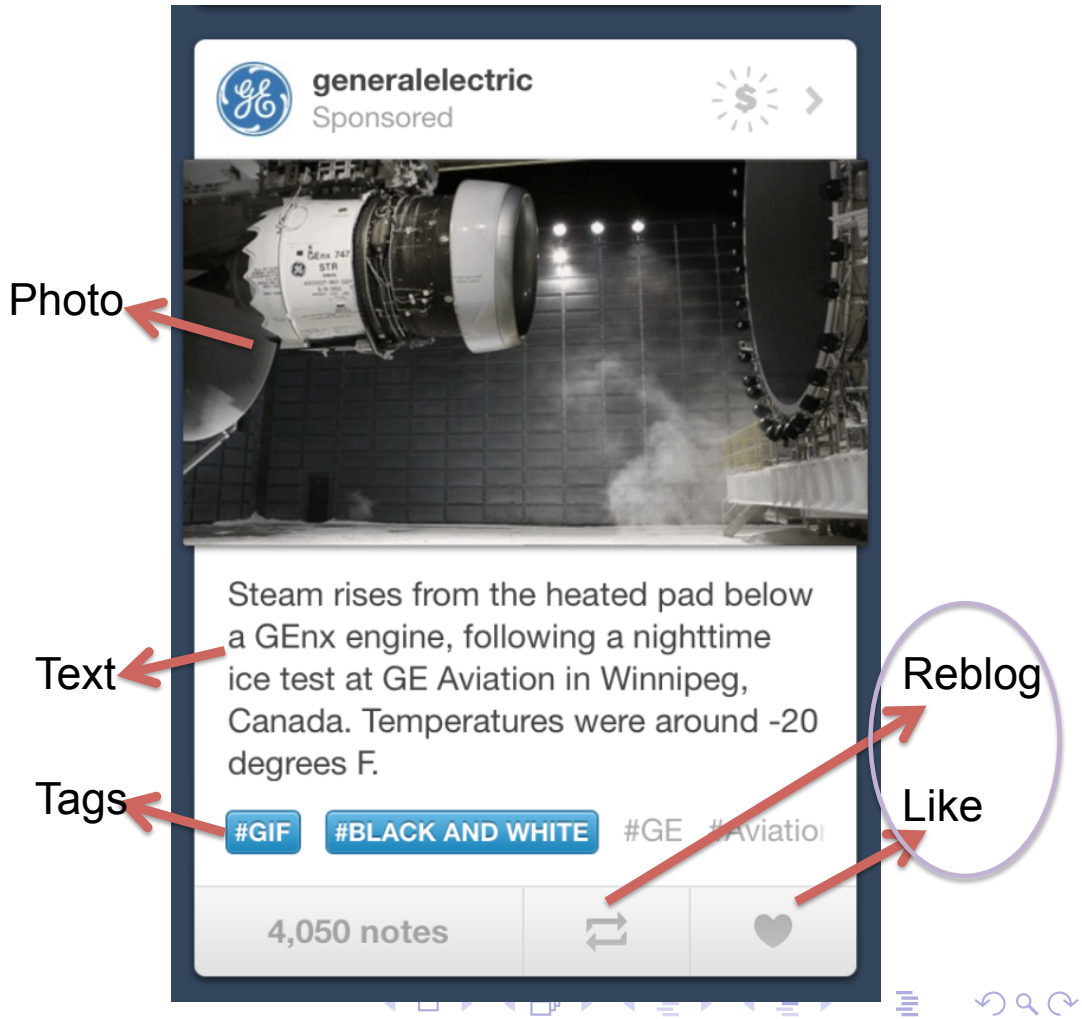
- Textual: text and tags
- Visual: photos (and videos)



Tumblr Posts

Various sources of additional information:

- Textual: text and tags
- Visual: photos (and videos)
- Activity: reblog and like



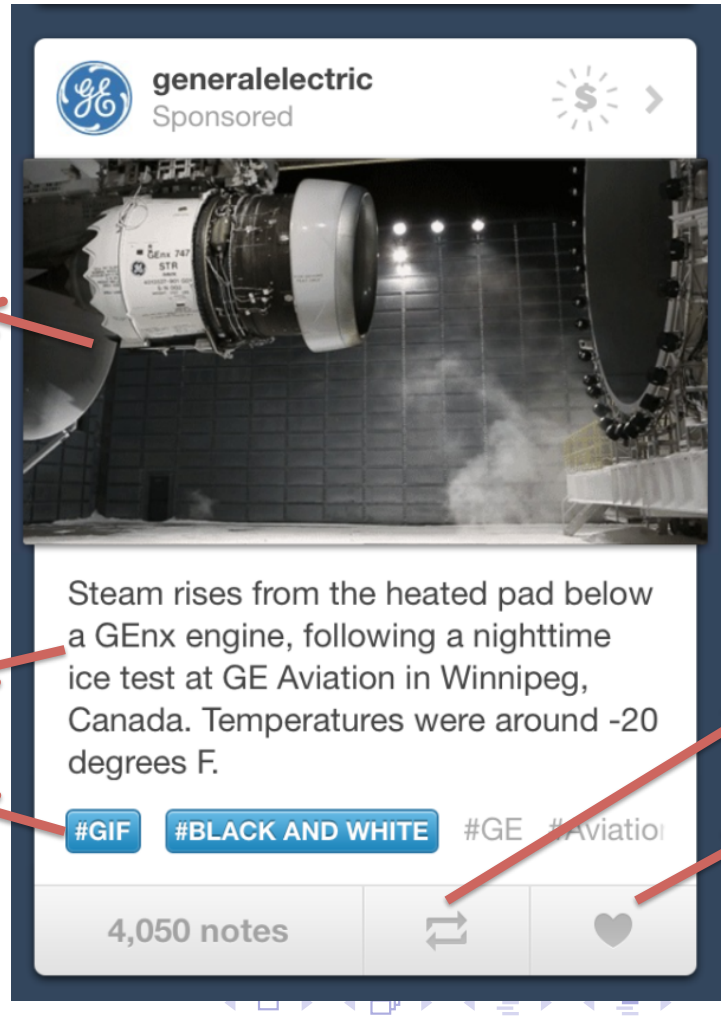
Tumblr Posts

Various sources of additional information:

- Textual: text and tags
- Visual: photos (and videos)
- Activity: reblog and like

User and blog features are generated from each source.

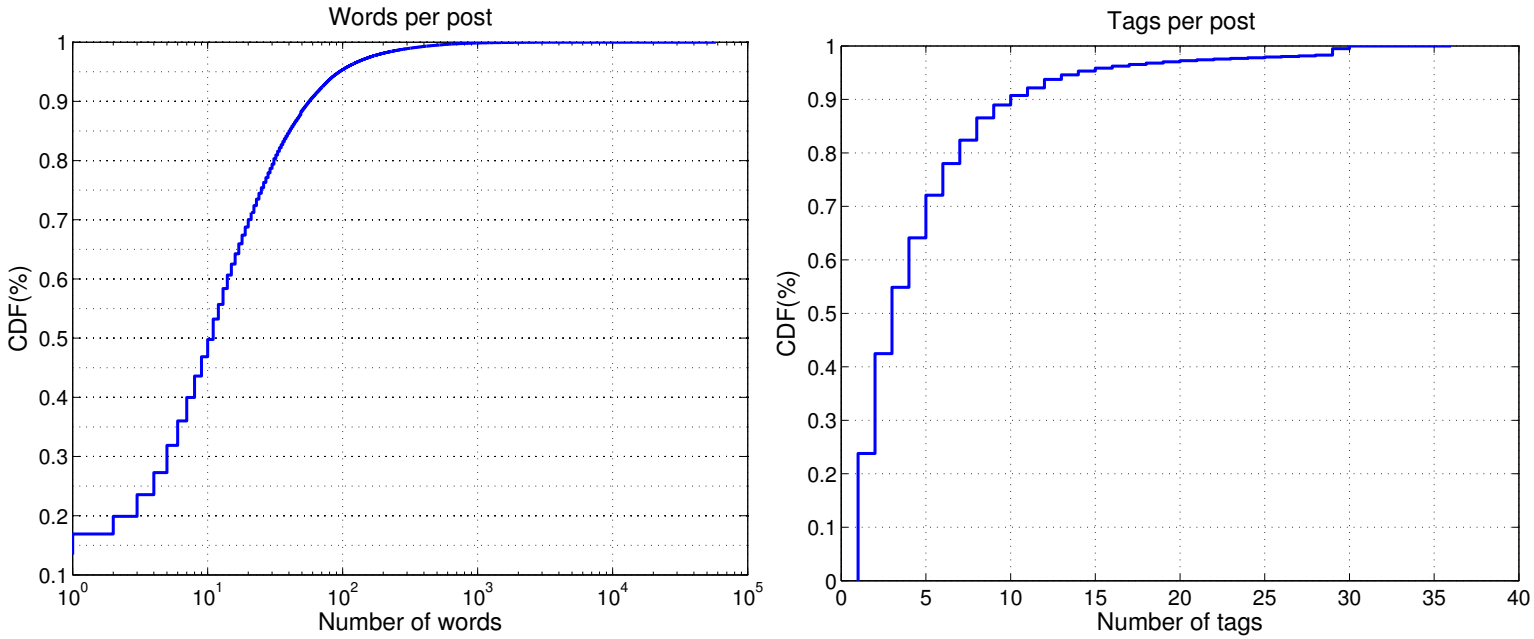
- 150M new posts per month
- Collected over a 5 month period: more than 7.5 TB of data



User and Blog Features – Textual

Text and Tags:

- Extremely sparse and noisy – 28.7 words / 4.8 tags per post on average



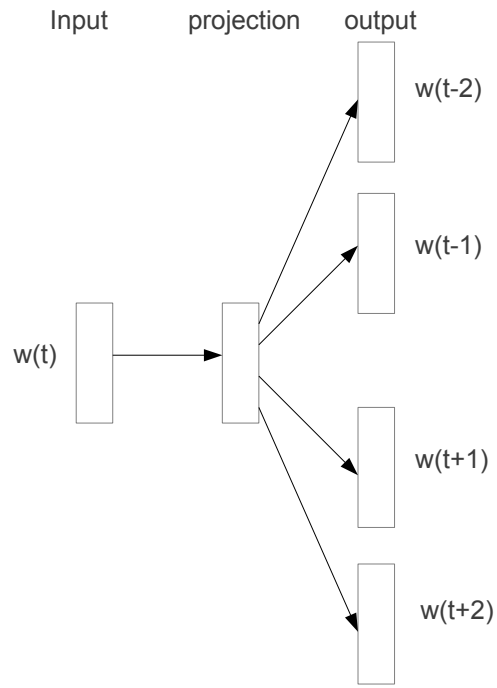
Problematic for existing bag-of-words models (LSA, LDA, etc).

User and Blog Features – Textual

Text and Tags: extract features using word2vec

[Mikolov et.al. 2013]

- Learns a vector representation of each word, such that words in similar context are close to each other



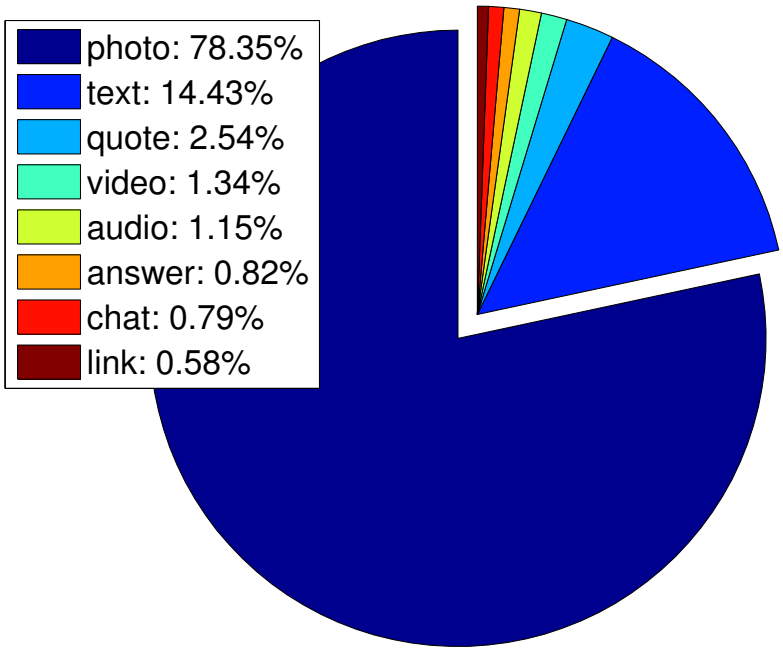
Textual features:

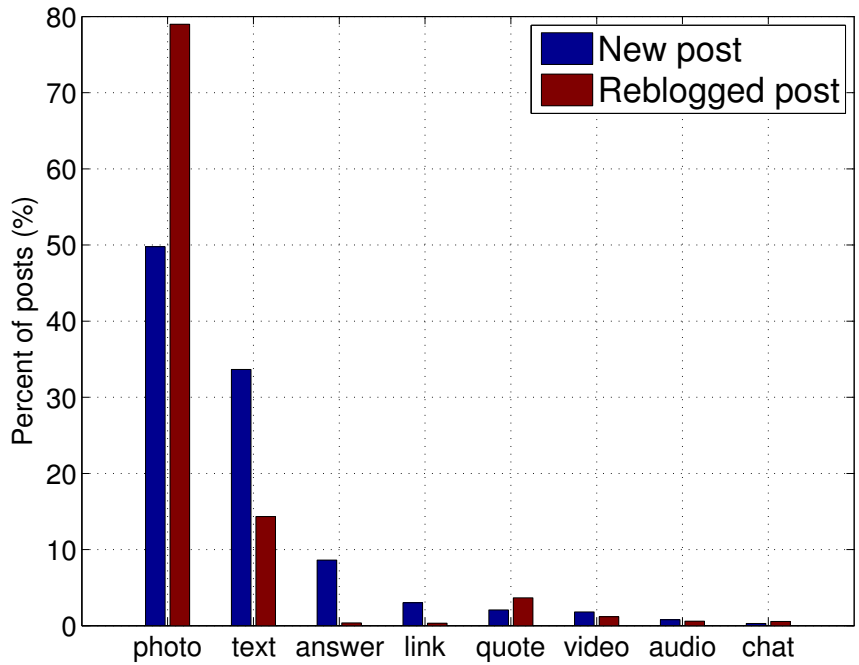
1. Learn d-dimensional vector representations of each word
2. Cluster words into c clusters via k-means
3. Compute histogram of word clusters for each post
4. Take average of histograms as features of users and blogs

User and Blog Features – Visual

Photos:

- About 78% of total posts are photo posts.
- Majority of reblogged posts are also photo posts.

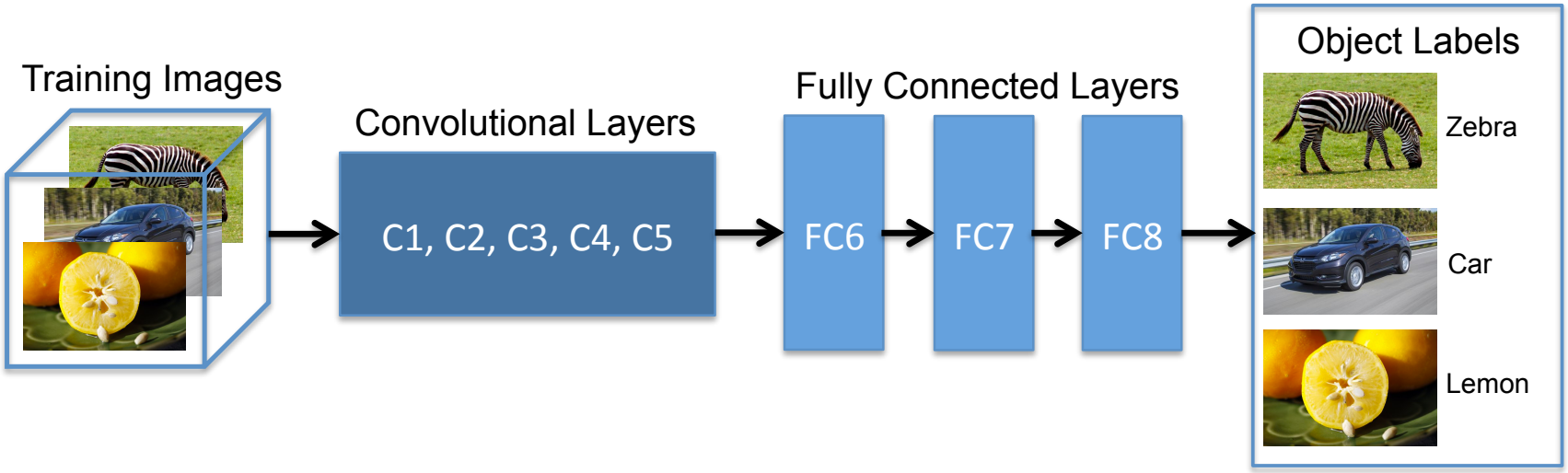




User and Blog Features – Visual

Photos: extract features using deep learning

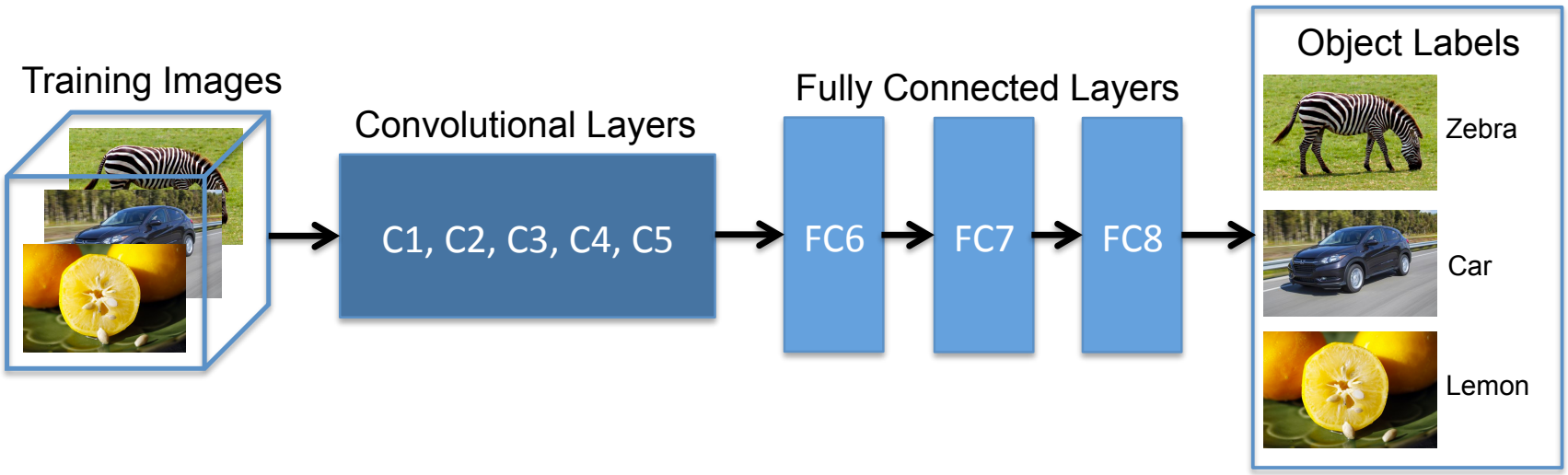
- Used a convolutional neural network with a total of 8 layers – 7 hidden layers and a soft-max layer [Donahue et.al. 2014]
  - Trained on 1.5M Flickr images (labels unavailable for Tumblr images)



User and Blog Features – Visual

Photos: extract features using deep learning

- Used a convolutional neural network with a total of 8 layers – 7 hidden layers and a soft-max layer [Donahue et.al. 2014]
  - Trained on 1.5M Flickr images (labels unavailable for Tumblr images)

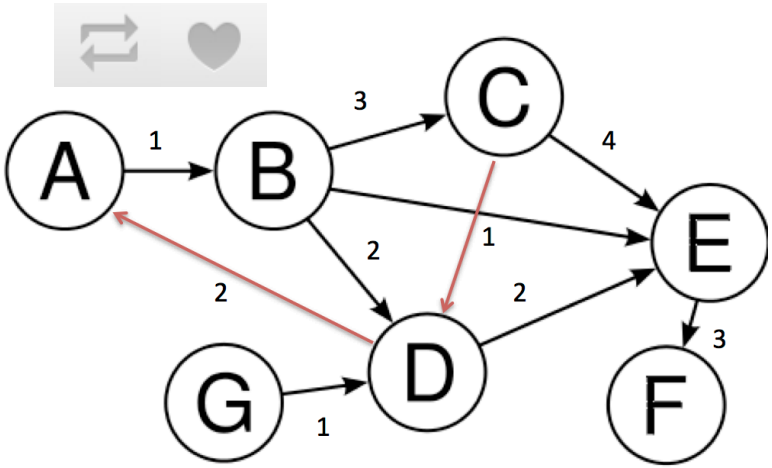


- Final output is a 958-dimensional vector of confidence scores over pre-defined categories
  - Averaged over photos that were posted/liked/reblogged

User and Blog Features – Activity

Both like and reblog activity can be represented as a graph W:

- W_{ij}: number of reblogged/liked posts of blog j by user i.

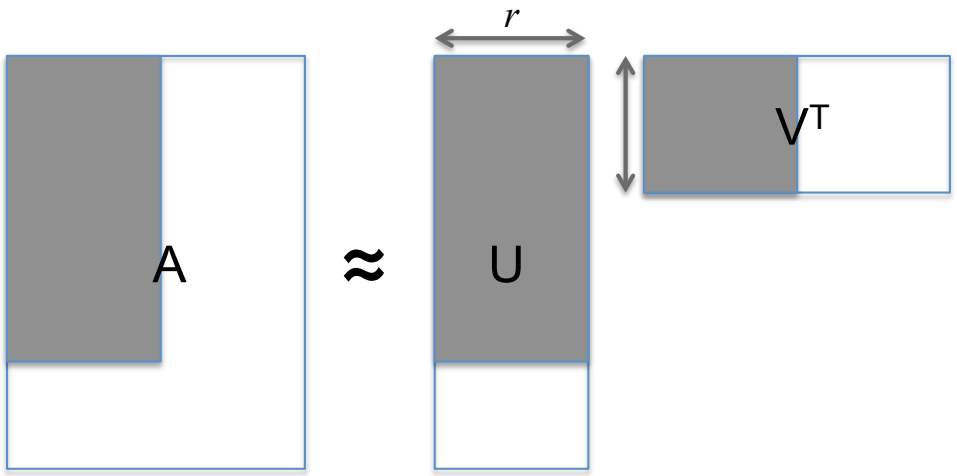


We extract principal components of the reblog/like graph.

Proposed Method:
Boosted Inductive Matrix Completion

Standard Matrix Completion

Model A_{ij} with low-rank matrix: $A_{ij} = \mathbf{u}_i \mathbf{v}_j^T$

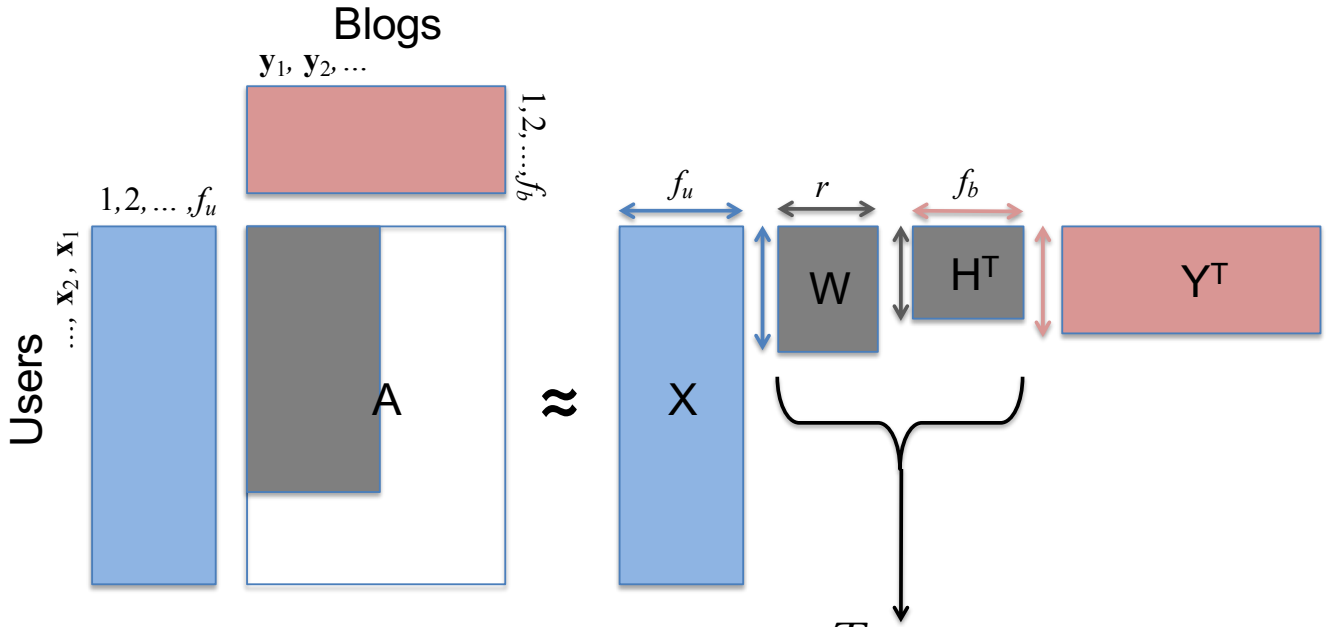


$$\min_{U,V} \sum_{(i,j)\in\Omega} (A_{ij} - \mathbf{u}_i \mathbf{v}_j^T)^2 + \frac{\lambda}{2}(\|U\|_F^2 + \|V\|_F^2),$$

- Theoretically well-understood + Fast solvers [Jain et.al. 2013, Yu et.al. 2014a]
- Restricted to transductive setting: predictions can only be made for existing users/items
- Suffers performance with extreme sparsity in data

Inductive Matrix Completion

Model A_{ij} with user and item feature vectors: $A_{ij} = \mathbf{x}_i^T Z \mathbf{y}_j$



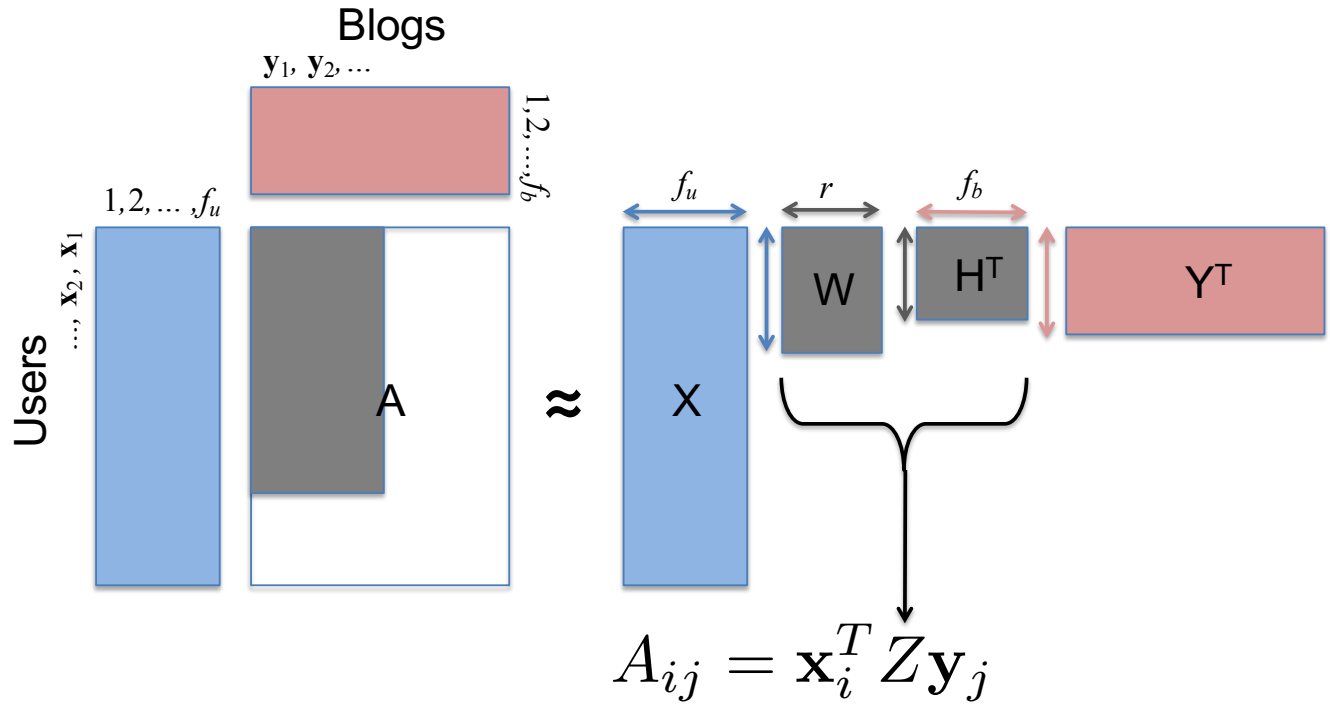
$$A_{ij} = \mathbf{x}_i^T Z \mathbf{y}_j$$

$$\min_{W,H} \sum_{(i,j)\in\Omega} (A_{ij} - \mathbf{x}_i^T W H^T \mathbf{y}_j)^2 + \frac{\lambda}{2}(\|W\|_F^2 + \|H\|_F^2)$$

- Utilizes side information – alleviate data sparsity issues
- Inductive setting: can make predictions for new users and items
- Theoretical guarantees + Fast solver [Yu et.al. 2014b, Zhong et.al. 2015]

Inductive Matrix Completion

Model A_{ij} with user and item feature vectors: $A_{ij} = \mathbf{x}_i^T Z \mathbf{y}_j$

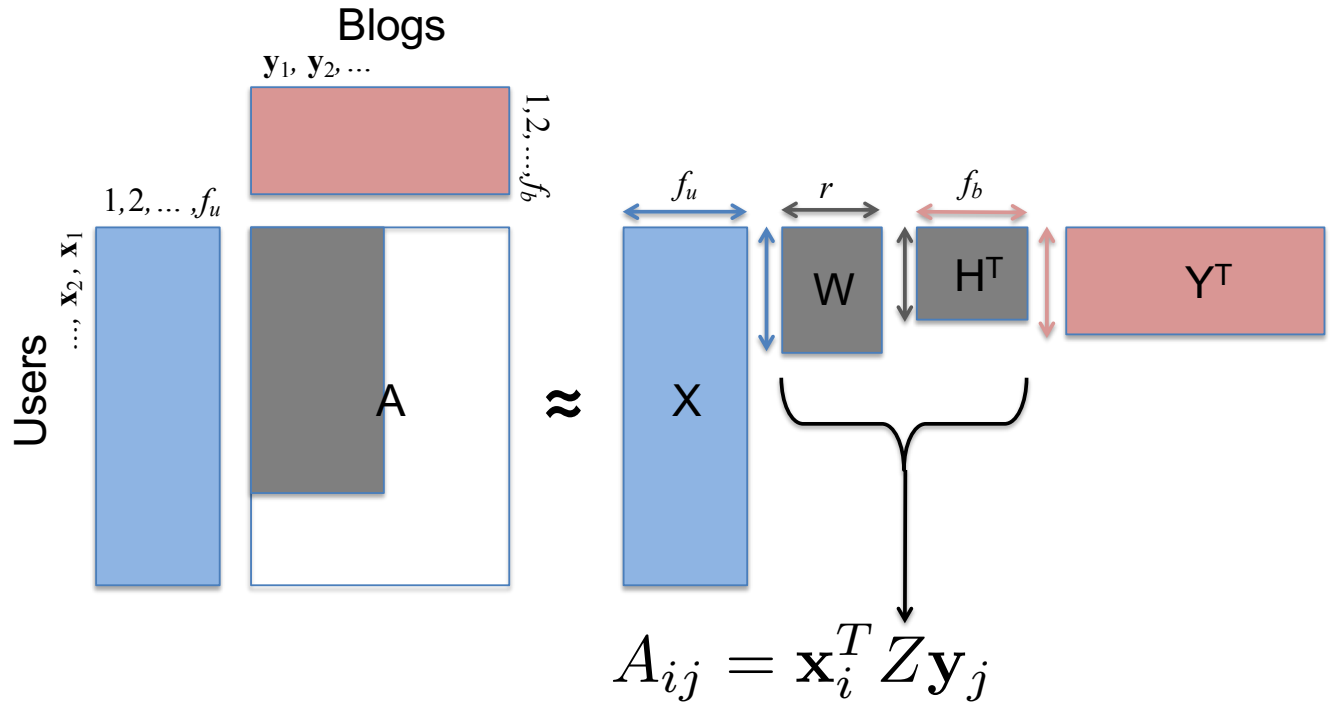


- IMC model is too rigid as it heavily depends on user/item features:

$$A = XZY^T \quad \Rightarrow \quad col(A) \subseteq col(X), \quad col(A^T) \subseteq col(Y)$$

Inductive Matrix Completion

Model A_{ij} with user and item feature vectors: $A_{ij} = \mathbf{x}_i^T Z \mathbf{y}_j$



- IMC model is too rigid as it heavily depends on user/item features:

$$A = XZY^T \quad \Rightarrow \quad col(A) \subseteq col(X), \quad col(A^T) \subseteq col(Y)$$

  - Noisy or weakly informative features that do not support A
  - Users or items without any features (e.g., users without any posts)

Boosted Inductive Matrix Completion

Combine MC and IMC and utilize the power of both methods:

$$A_{ij} = \mathbf{u}_i \mathbf{v}_j^T + \alpha \mathbf{x}_i^T Z \mathbf{y}_j$$

- Exploits both observation and feature information – α controls the contribution of features
- Can handle sparsity issues in either source

Boosted Inductive Matrix Completion

Combine MC and IMC and utilize the power of both methods:

$$A_{ij} = \mathbf{u}_i \mathbf{v}_j^T + \alpha \mathbf{x}_i^T Z \mathbf{y}_j$$

- Exploits both observation and feature information – α controls the contribution of features
- Can handle sparsity issues in either source

Caveats:

- Tuning for a good α is difficult
- Jointly optimizing for all factor matrices can be inefficient

Boosted Inductive Matrix Completion

Our approach: focus on the residual in an additive manner

$$R_{ij} = A_{ij} - \mathbf{u}_i \mathbf{v}_j^T = \mathbf{x}_i^T Z \mathbf{y}_j$$

1. First find the support of A with U and V
2. Then focus on the part that cannot be modeled by MC

Boosted Inductive Matrix Completion

Our approach: focus on the residual in an additive manner

$$R_{ij} = A_{ij} - \mathbf{u}_i \mathbf{v}_j^T = \mathbf{x}_i^T Z \mathbf{y}_j$$

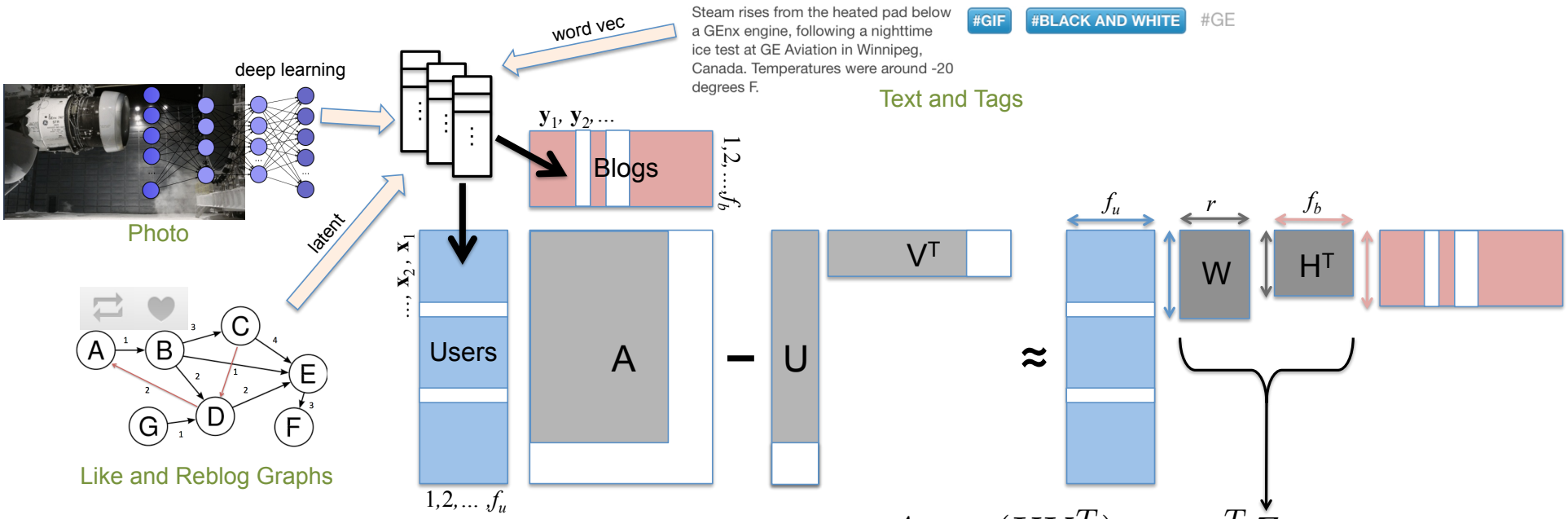
1. First find the support of A with U and V
2. Then focus on the part that cannot be modeled by MC

Combine effectiveness of both MC and IMC models:

- Captures low-rank structure of A + latent structure using features
- Captures entries in A where MC fails to learn – especially useful when $\|R\|$ is large
- No need to fine tune α
- Can utilize fast MC/IMC solvers

Boosted Inductive Matrix Completion

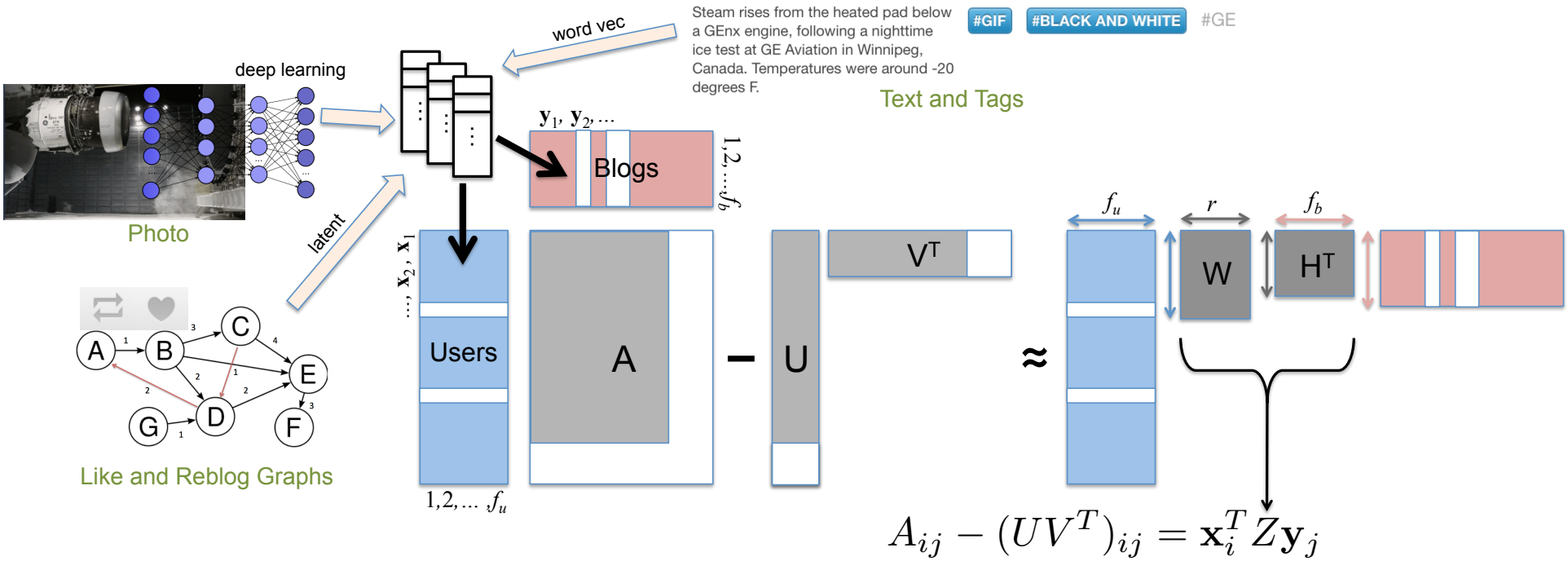
1. Learn U and V via standard MC
2. Train IMC on the residual $R = A - UV^T$



$$A_{ij} - (UV^T)_{ij} = \mathbf{x}_i^T Z \mathbf{y}_j$$

Boosted Inductive Matrix Completion

1. Learn U and V via standard MC
2. Train IMC on the residual $R = A - UV^T$



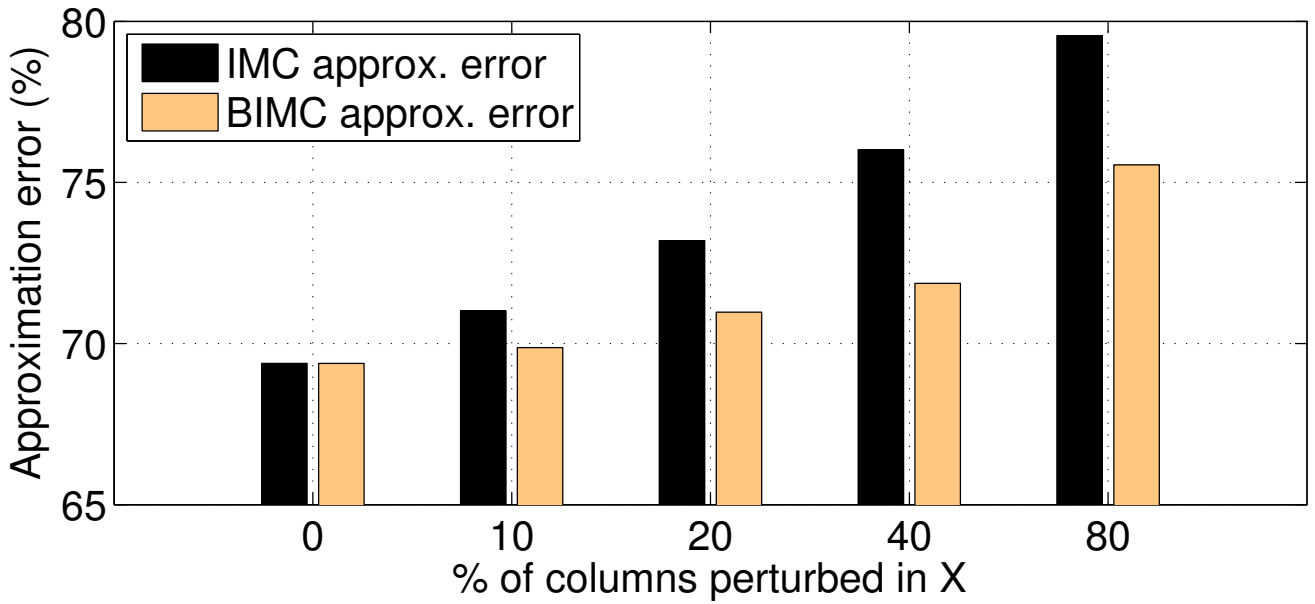
$$A_{ij} - (UV^T)_{ij} = \mathbf{x}_i^T Z \mathbf{y}_j$$

- Converse approach (IMC⇒MC) gives comparable results.

Robustness to Noisy Features

Dataset: MovieLens-100K (100K ratings from 1,000 users on 1,700 movies)

- Compute the rank-20 SVD of $A \approx U_A \Sigma_A V_A$.
- Set $X = U_A$ and $Y = V_A$ for both IMC and BIMC.
- Perturb columns of X by adding random noise.



Experimental Results

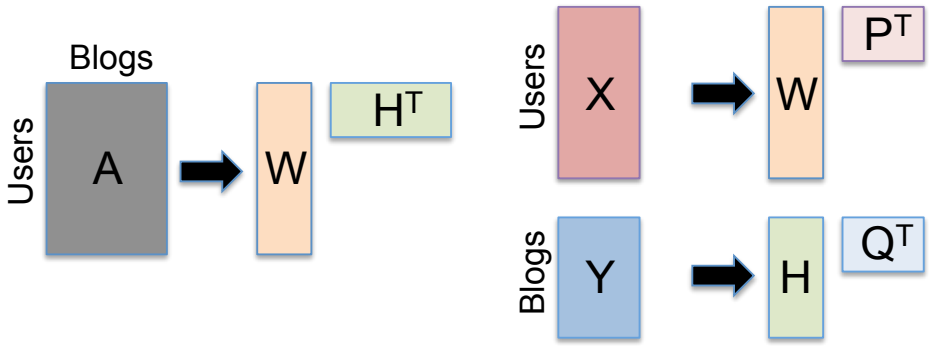
Comparison Methods

Baselines:

- Global ranking, SVD, Matrix Completion (MC)

Methods incorporating side-information:

- Inductive Matrix Completion (IMC)
- Collective Matrix Completion (CMC) [Gunasekar et.al. 2015]
  - Jointly recover a collection of matrices with shared low rank structure



- Graph-based proximity measure between users and blogs (Katz):
  - Compute Katz scores using $C = \begin{bmatrix} U & A \\ A^T & B \end{bmatrix}$, where U and B are similarity matrices computed from user and blog features, respectively.

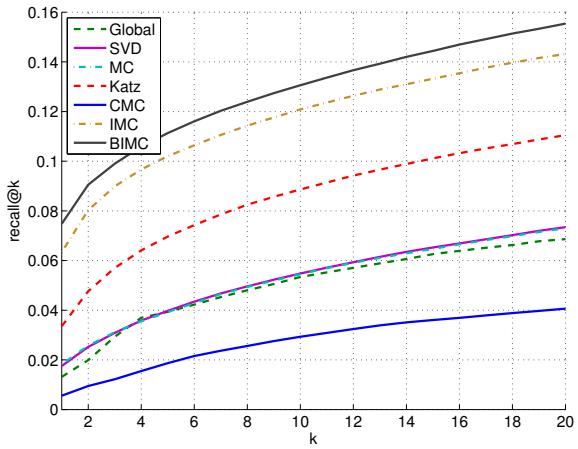
Experimental Results

- 1M sampled users and blogs
- 10-fold cross-validation (similar results with temporal evaluation)

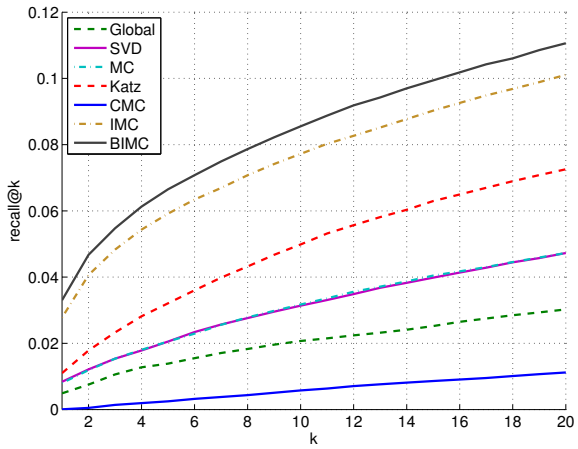
Method	PRC@10	RCL@10	AUC
Global	1.03%	4.80%	0.8687
SVD	1.28%	5.10%	0.8530
MC	1.28%	5.07%	0.8515
Katz	1.90%	8.15%	0.9209
CMC	0.49%	2.41%	0.8996
IMC	2.93%	11.33%	0.9075
BIMC	**3.21%**	**12.28%**	**0.9221**

User and Blog Groups

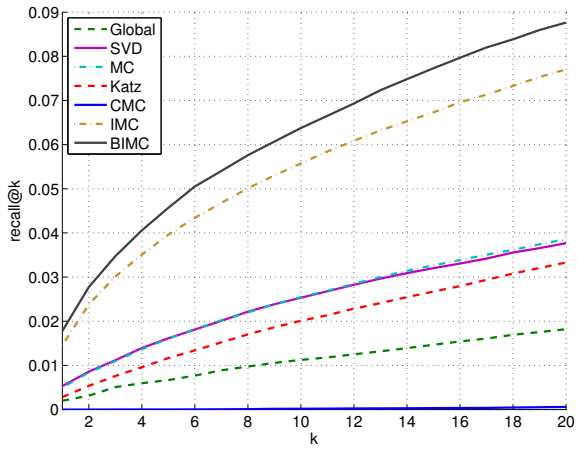
User groups with different number of followees ($n_f \leq 40$, $40 < n_f \leq 100$, $100 < n_f$)



(a) Low (89.4%)

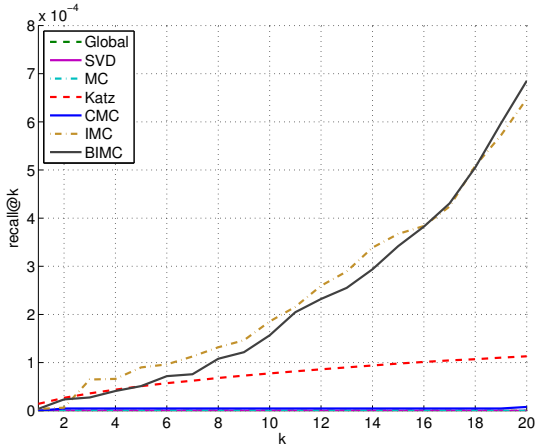


(b) Medium (7.8%)

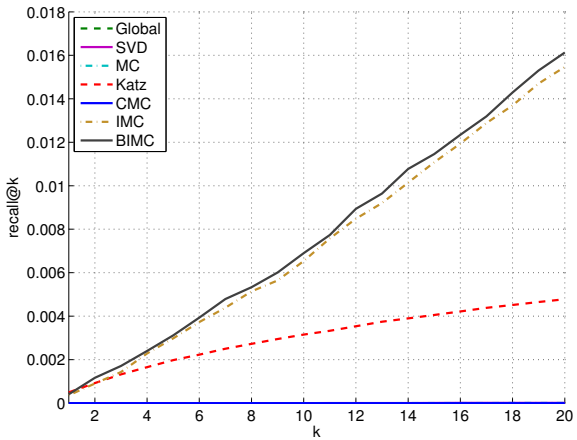


(c) High (2.8%)

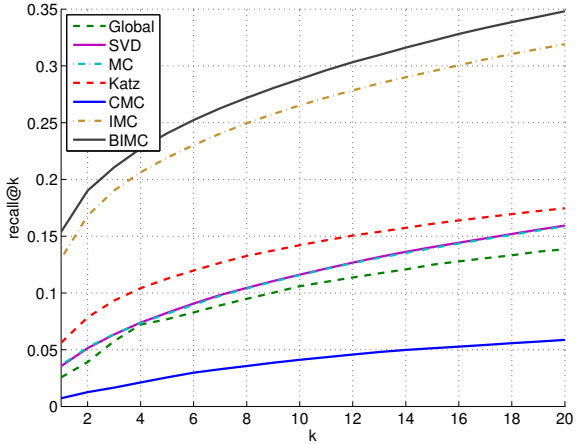
Blog groups with different number of followers ($n_f \leq 40$, $40 < n_f \leq 100$, $100 < n_f$)



(d) Low (95.1%)

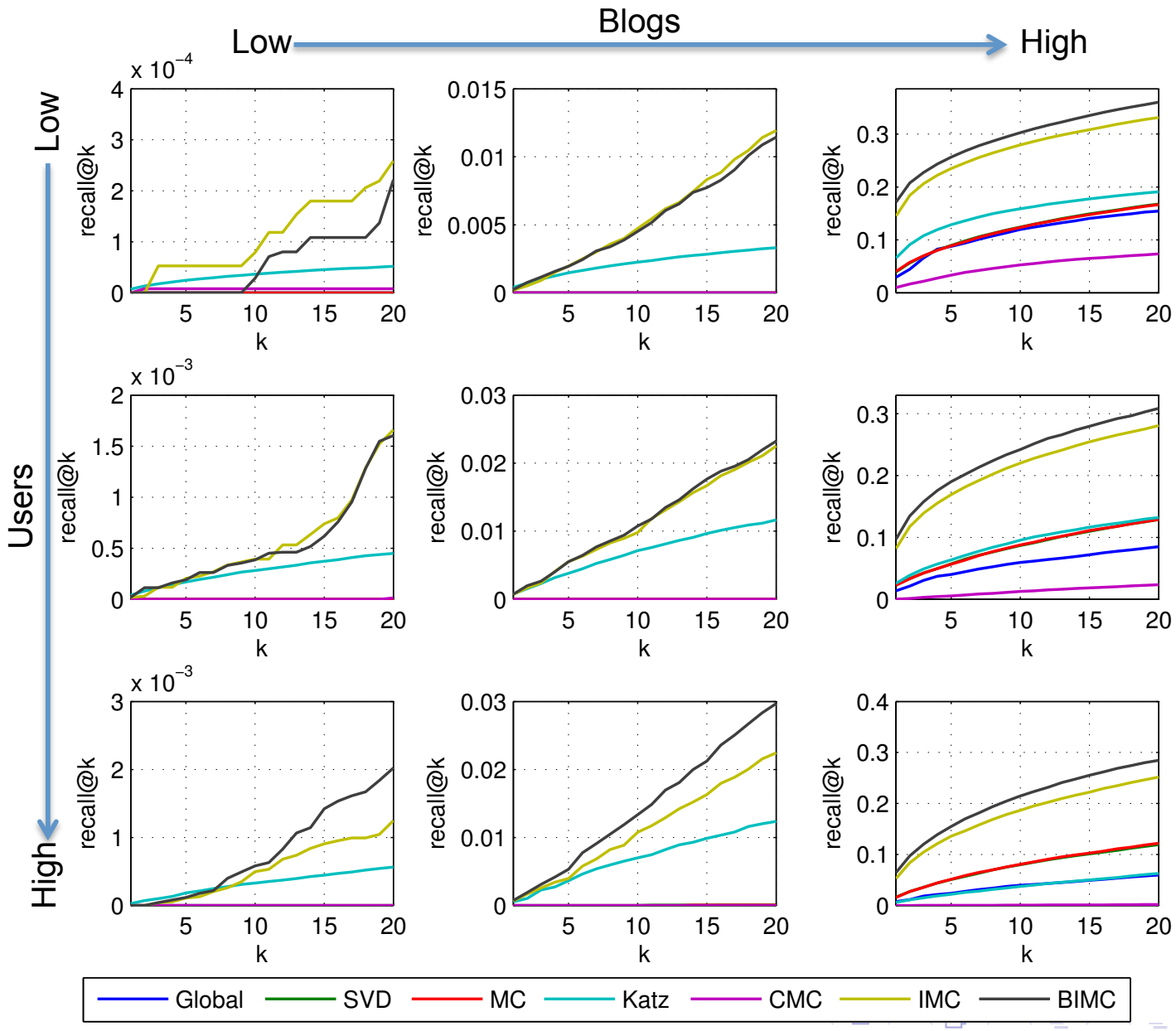


(e) Medium (3.3%)



(f) High (1.6%)

Recall@k


Blogs
Low
High
Users
Low
High
x 10^-4
x 10^-3
x 10^-3
recall@k
k
Global
SVD
MC
Katz
CMC
IMC
BIMC

Conclusion

Tumblr Blog Recommendation

- Incorporate features extracted from various side information (text, image and activity) with state-of-the-art deep learning approaches.
- BIMC: effectively combining the power of both MC and IMC
  - Exploits both observation and feature information
  - Can deal with sparsity issues in either source
  - Can make predictions for new users and items
  - Scalable due to fast MC/IMC solvers
  - Yields the best recommendation performance

Thanks!

References

[1] T. Mikolov, I. Sutskever, K. Chen, G. S. Corrado, and J. Dean. *Distributed Representations of Words and Phrases and their Compositionality*. NIPS 2013.

[2] J. Donahue, Y. Jia, O. Vinyals, J. Hoffman, N. Zhang, E. Tzeng, and T. Darrell. *DeCAF: A Deep Convolutional Activation Feature for Generic Visual Recognition.* ICML 2014.

[3] P. Jain, P. Netrapalli, and S. Sanghavi. *Low-rank matrix completion using alternating minimization.* STOC 2013.

[4] K. Zhong, P. Jain, and I. S. Dhillon. *Efficient Matrix Sensing Using Rank-1 Gaussian Measurements*. ALT 2015.

[5] H. Yu, C. Hsieh, S. Si, I. S. Dhillon. *Parallel Matrix Factorization for Recommender Systems.* KAIS 2014a.

[6] H. Yu, P. Jain, P. Kar, and I. S. Dhillon. *Large-scale Multi-label Learning with Missing Labels*. ICML 2014b.

[7] S. Gunasekar, M. Yamada, D. Yin, and Y. Chang. *Consistent Collective Matrix Completion under Joint Low Rank Structure.* AISTATS 2015.