AdHeat, An Influence-based Social Ads Model & its Tera-scale Algorithms

Edward Y. Chang

Google Research

Comparison between Parallel Computing Frameworks

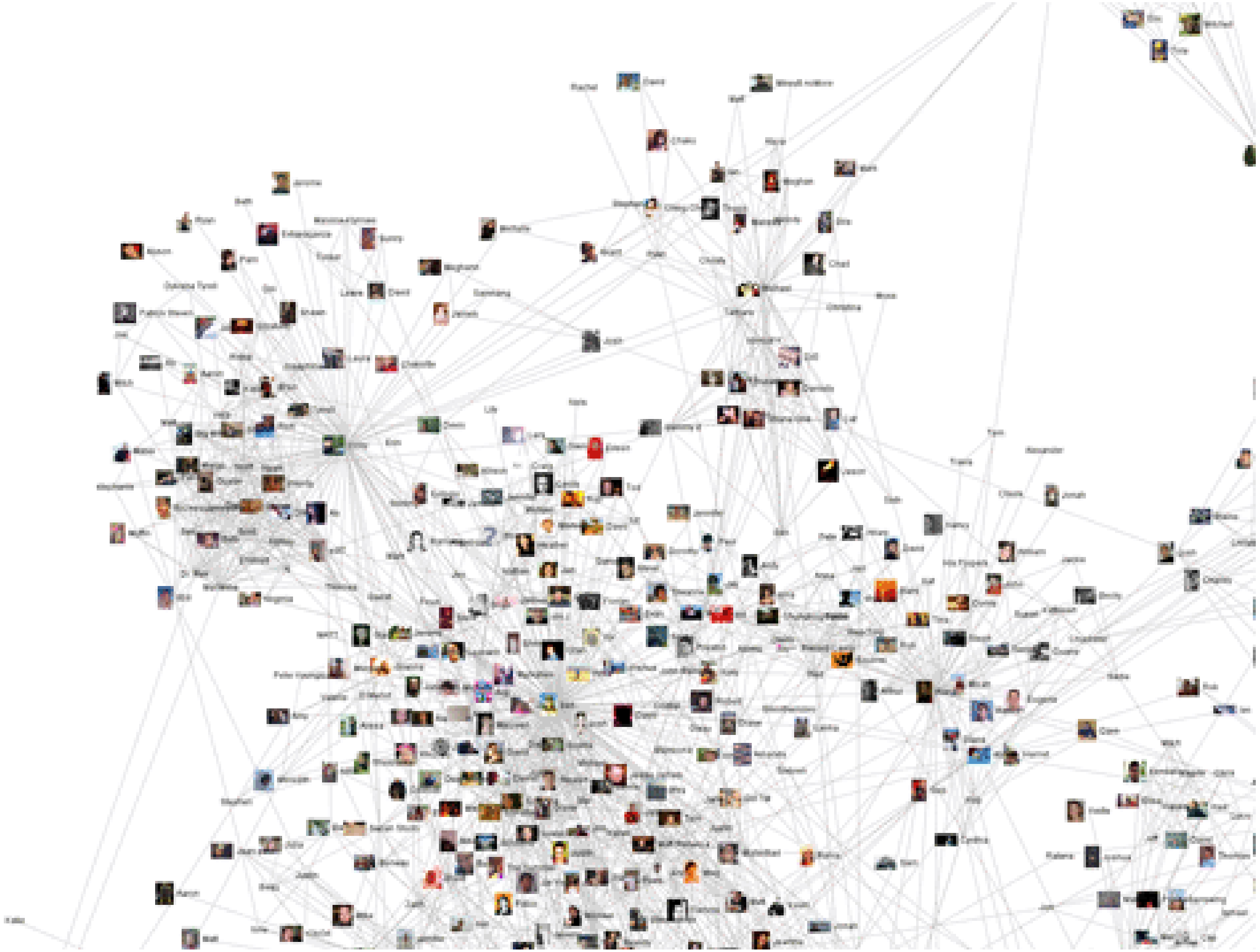
- **Parallel LDA** (ACM Transactions on Internet Technology, 2010)
- **Parallel Spectral Clustering** (PAMI, 2010)
- **Parallel SVMs** (NIPS, 2007)

	MapReduce	**Pregel**	**MPI**
GFS/IO and task rescheduling overhead between iterations	Yes	No +1	No +1
Flexibility of computation model	AllReduce only +0.5	AllReduce only +0.5	Flexible +1
Efficient AllReduce	Yes +1	Yes +1	Yes +1
Recover from faults between iterations	Yes +1	Yes +1	Apps
Recover from faults within each iteration	Yes +1	Yes +1	Apps
Final Score for scalable machine learning	3.5	4.5	5

Outline

- Social Network Ad Model
  - Relevance Model
  - Influence Model
- Key Algorithms
  - UserRank
  - Hint Word Generation
  - Diffusion

Social Networks [Jeff Heer, visualization]



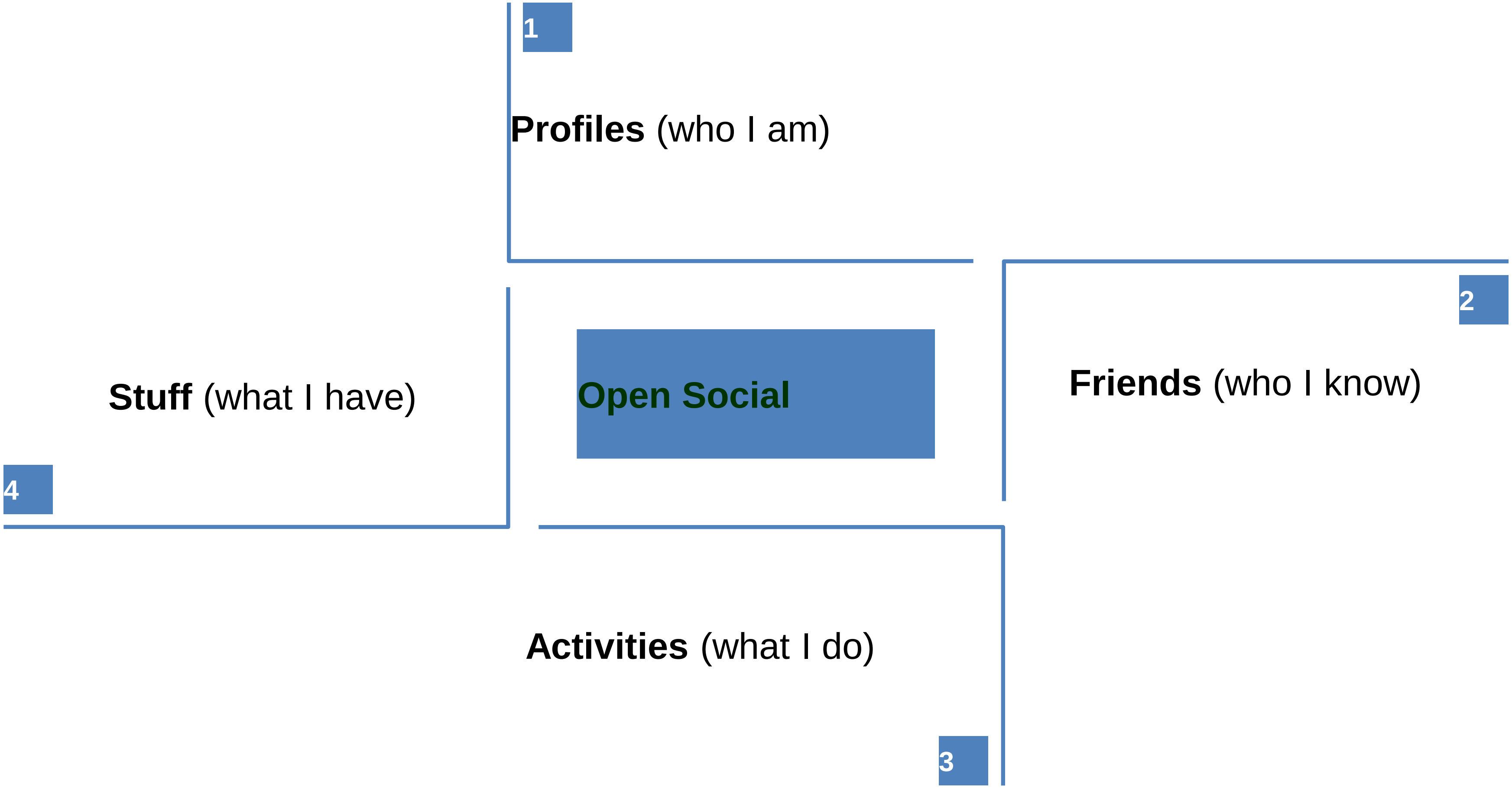
Task: Targeting Ads at SNS Users

Users

Ads

Mining Profiles, Friends & Activities for Relevance

Open Social APIs

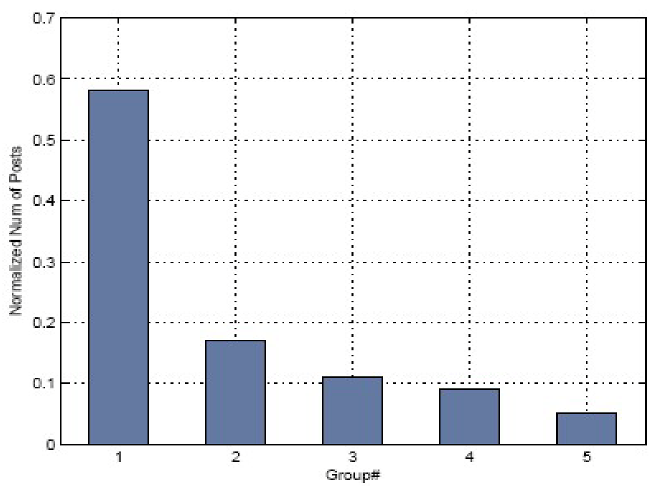


Relevance Model

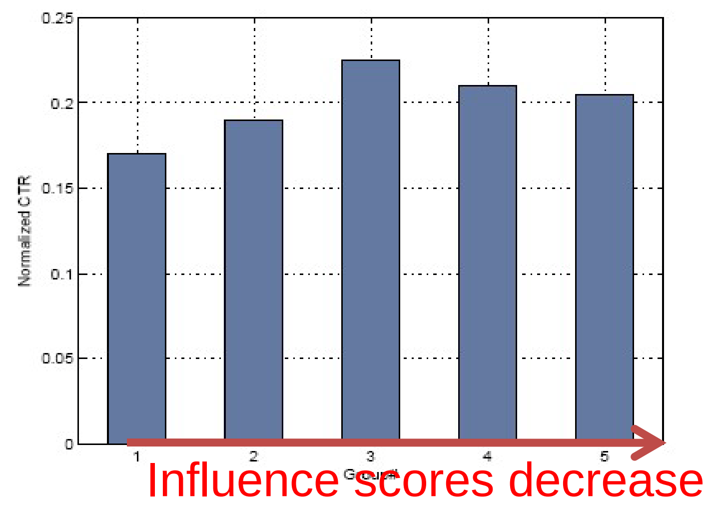
Limitation #1

Relevance $\nrightarrow$ High CTR

- Correlation between users' Influence and Performance
  - Rank users by their content contributions
  - Evaluate *relevance* vs. *CTR*



(a) Content



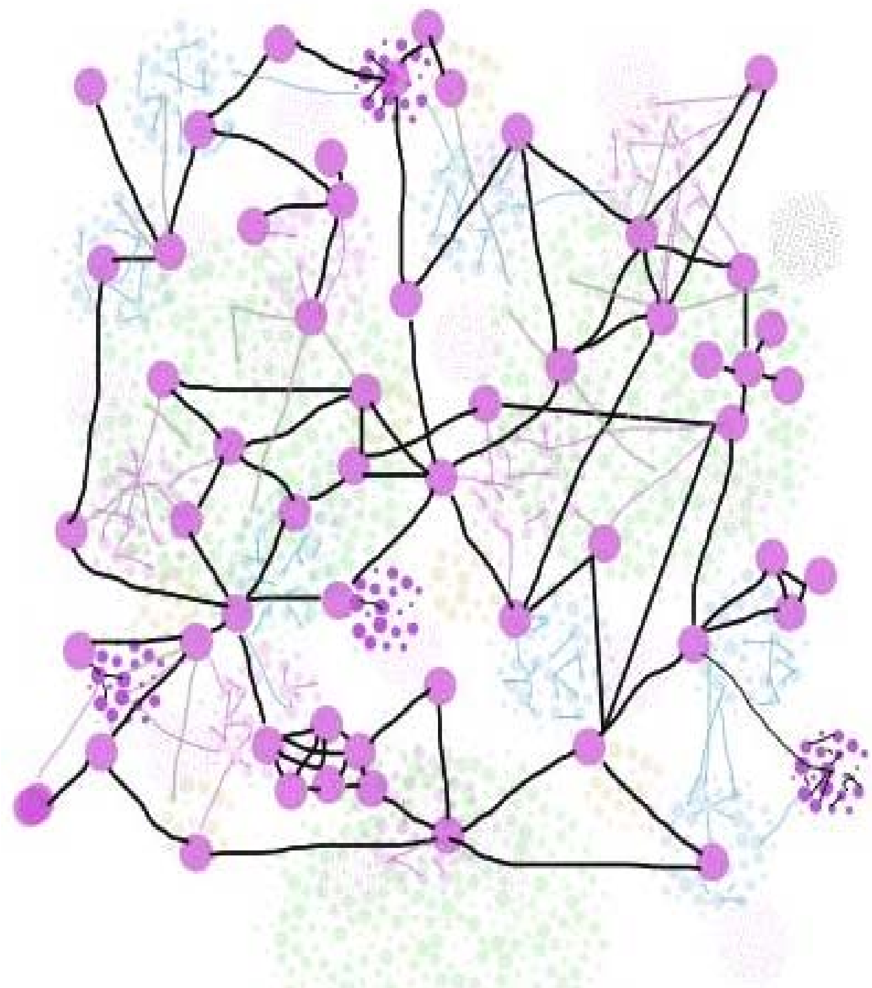
(b) CTR

Summary of Relevance

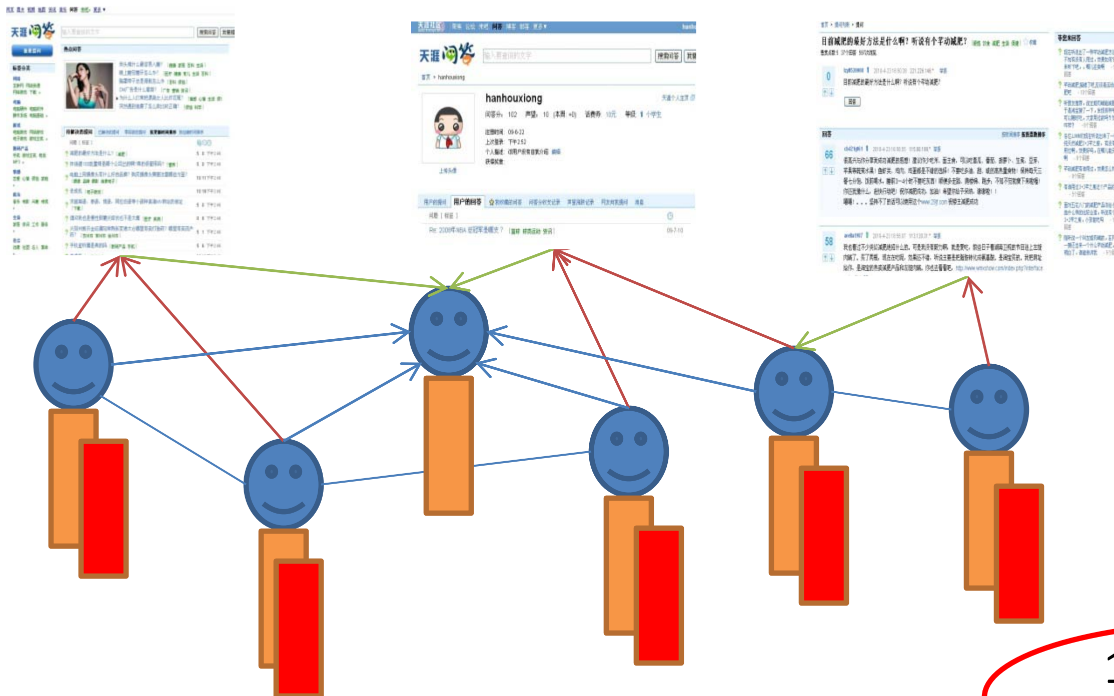
- Relevance analysis based on
  - User profile/friends/activities/stuff
- Active users
  - Sufficient data to conduct relevance analysis
  - Do not click on relevant ads
- Inactive users
  - Data too sparse to conduct relevance analysis

AdHeat: Consider also User *Influence*

- Advertisers compete for users who are
  - relevant
  - *influential*
- SNS Influence Analysis
  - Centrality
  - Expertise
  - Activeness
  - Heat Diffusion Rate



AdHeat



1. Relevance analysis

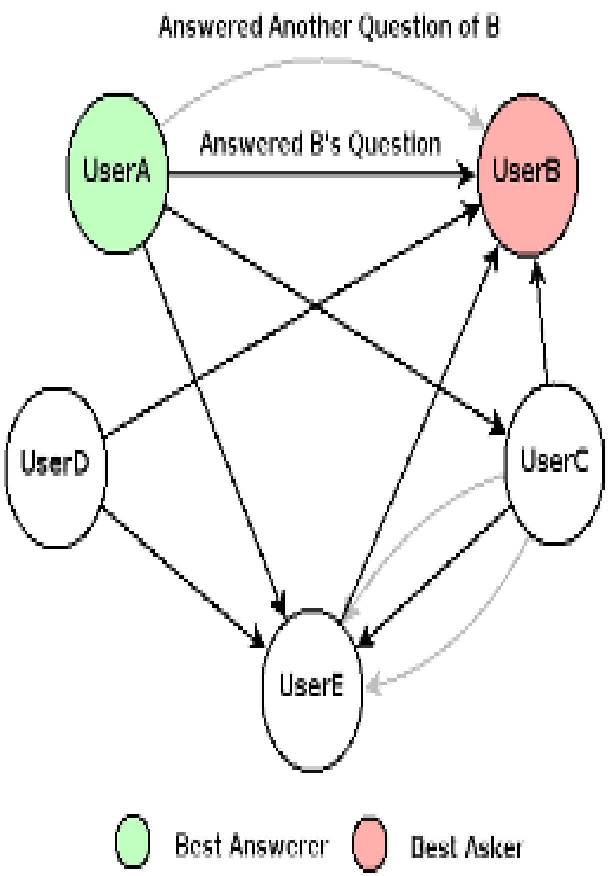
- AdHeat model
  - mines the Individuals' characteristics/interests based on their contributions;
  - quantifies mutual influence between users based on their interactions, constructs social network graph, and ranks the users by their influence;
  - propagates the interests of the influential users to those who are influenced by them.

2. Influential user ranking
3. Relevance propagation

Outline

- Social Network Ad Model
  - Relevance Model
  - Influence Model
- Key Algorithms
  - UserRank
  - Hint Word Generation
  - Diffusion

UserRank [VLDB 2010]



- Rank users by quantity **(number of links)** and quality **(weights on links)** of contributions

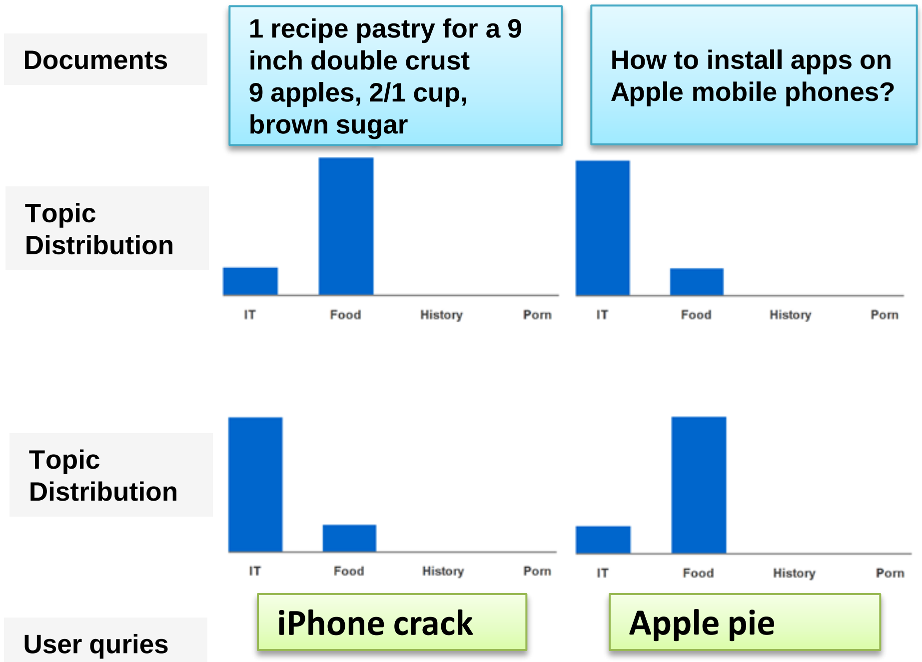
  Quality include:
  - **Relevance.** Is an answer relevant to the Q? Measured by KL divergence between *latent-topic vectors* of A and Q
  - **Originality.** Detect potential plagiarism and spam
  - **Topic-dependent Factors.**

Outline

- Social Network Ad Model
  - Relevance Model
  - Influence Model
- Key Algorithms
  - UserRank
  - Hint Word Generation
  - Diffusion

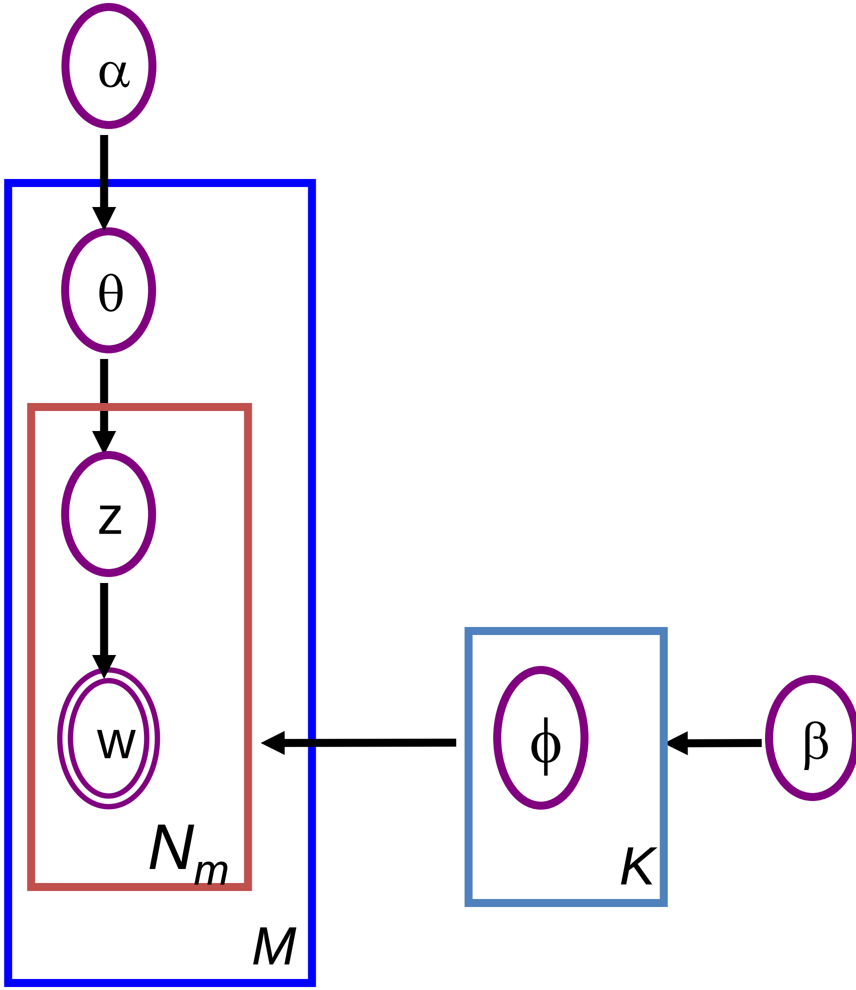
Latent Semantic Analysis

- Construct a latent layer for better for semantic matching
- Example:
  - iPhone crack
  - Apple pie



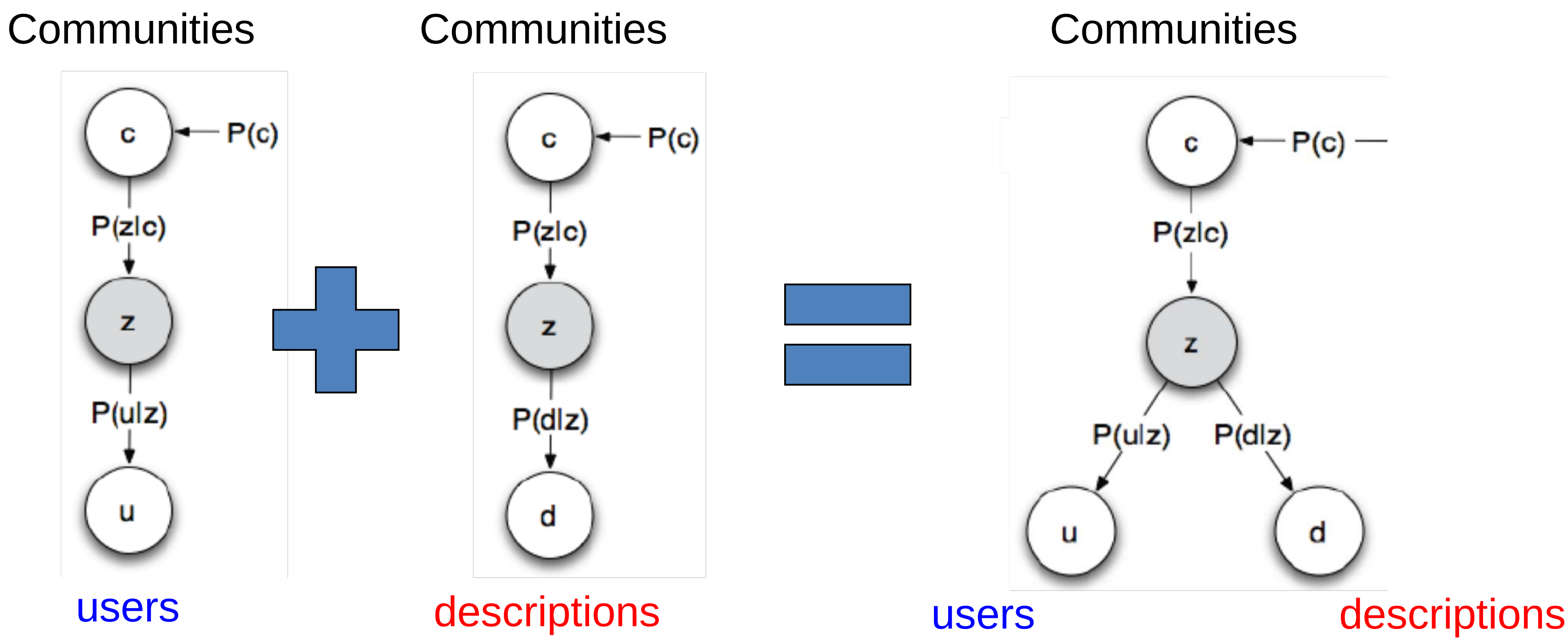
Latent Dirichlet Allocation [D. Blei, M. Jordan 04]

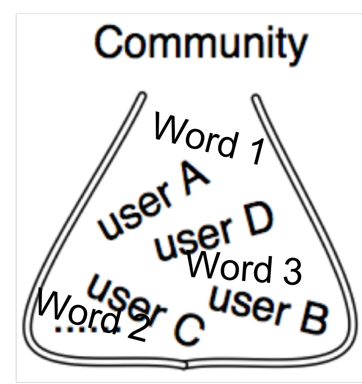
- α: uniform Dirichlet ϕ prior for per document d topic distribution (corpus level parameter)
- β: uniform Dirichlet ϕ prior for per topic z word distribution (corpus level parameter)
- θ_d is the topic distribution of document d (document level)
- z_{dj} the topic if the j^{th} word in d, w_{dj} the specific word (word level)



Combinational Collaborative Filtering Model (CCF)

[KDD2008]





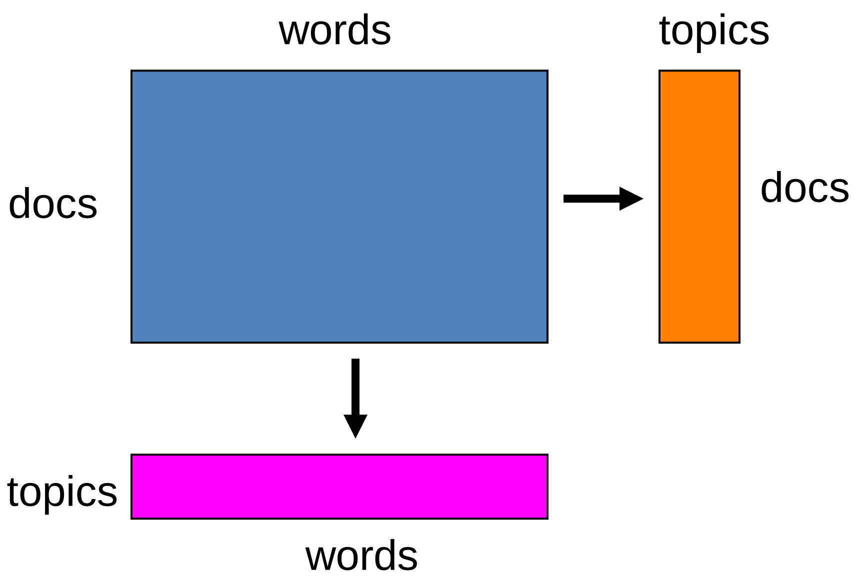
LDA Gibbs Sampling: Inputs & Outputs

Inputs:

1. training data: documents as bags of words
2. parameter: the number of topics

Outputs:

1. model parameters: a co-occurrence matrix of topics and words.
2. by-product: a co-occurrence matrix of topics and documents.



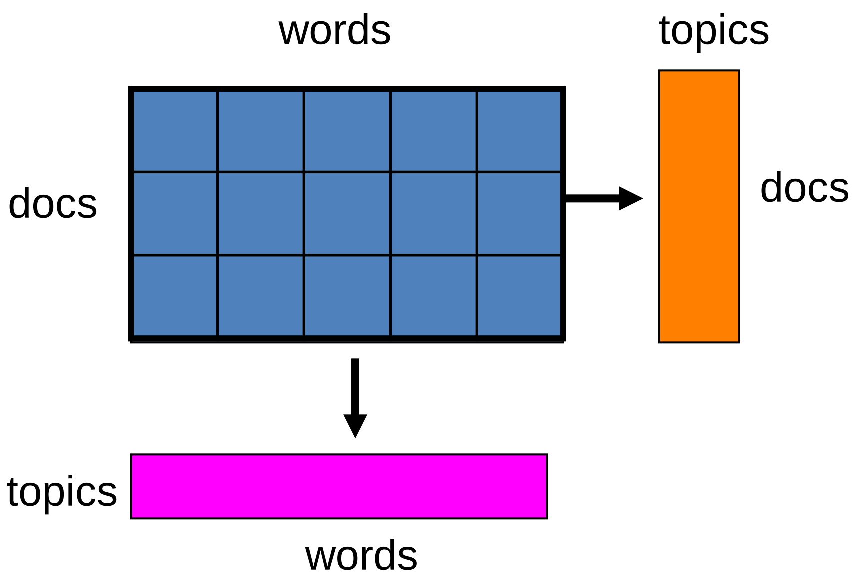
Parallel Gibbs Sampling

Inputs:

1. training data: documents as bags of words
2. parameter: the number of topics

Outputs:

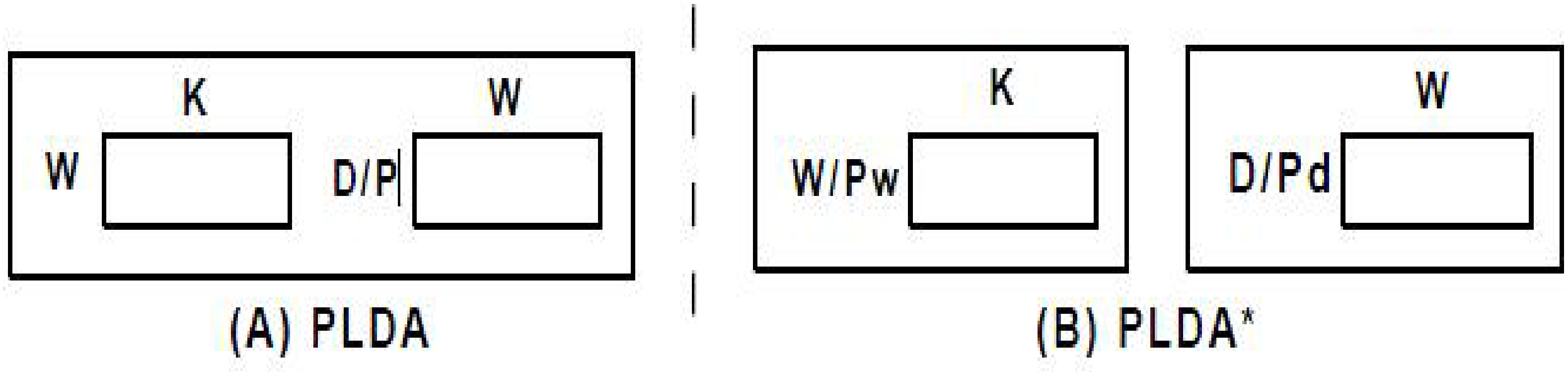
1. model parameters: a co-occurrence matrix of topics and words.
2. by-product: a co-occurrence matrix of topics and documents.



PLDA* -- enhanced parallel LDA

[ACM Transactions on IT]

- PLDA is restricted by memory: Topic-word matrix has to fit into memory
- Restricted by Amdahl's Law: communication costs too high



PLDA* -- enhanced parallel LDA

- Take advantage of bag of words modeling: each Pw machine processes vocabulary in a word order
- Pipelining: fetching the updated topic distribution matrix while doing Gibbs sampling

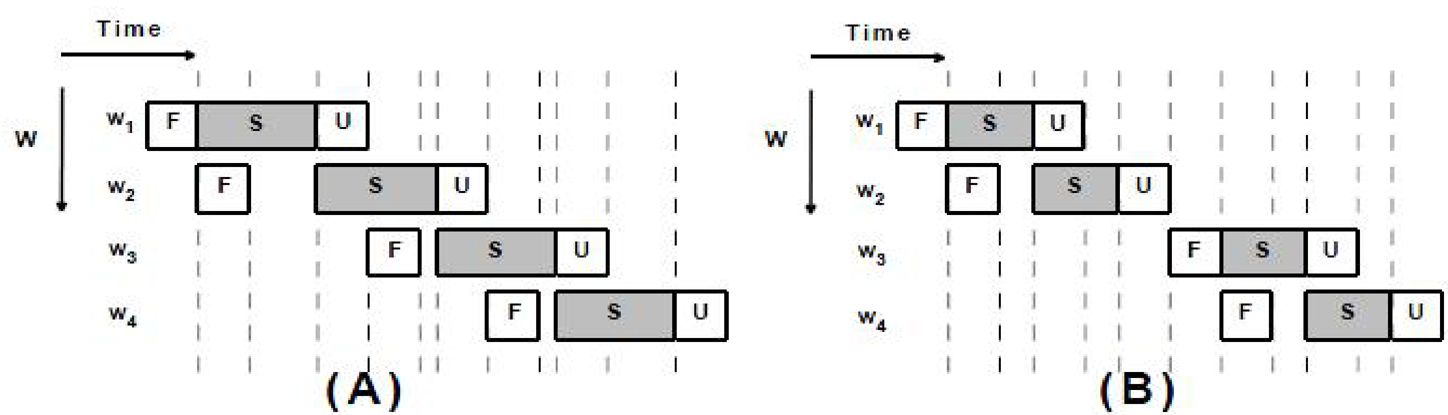
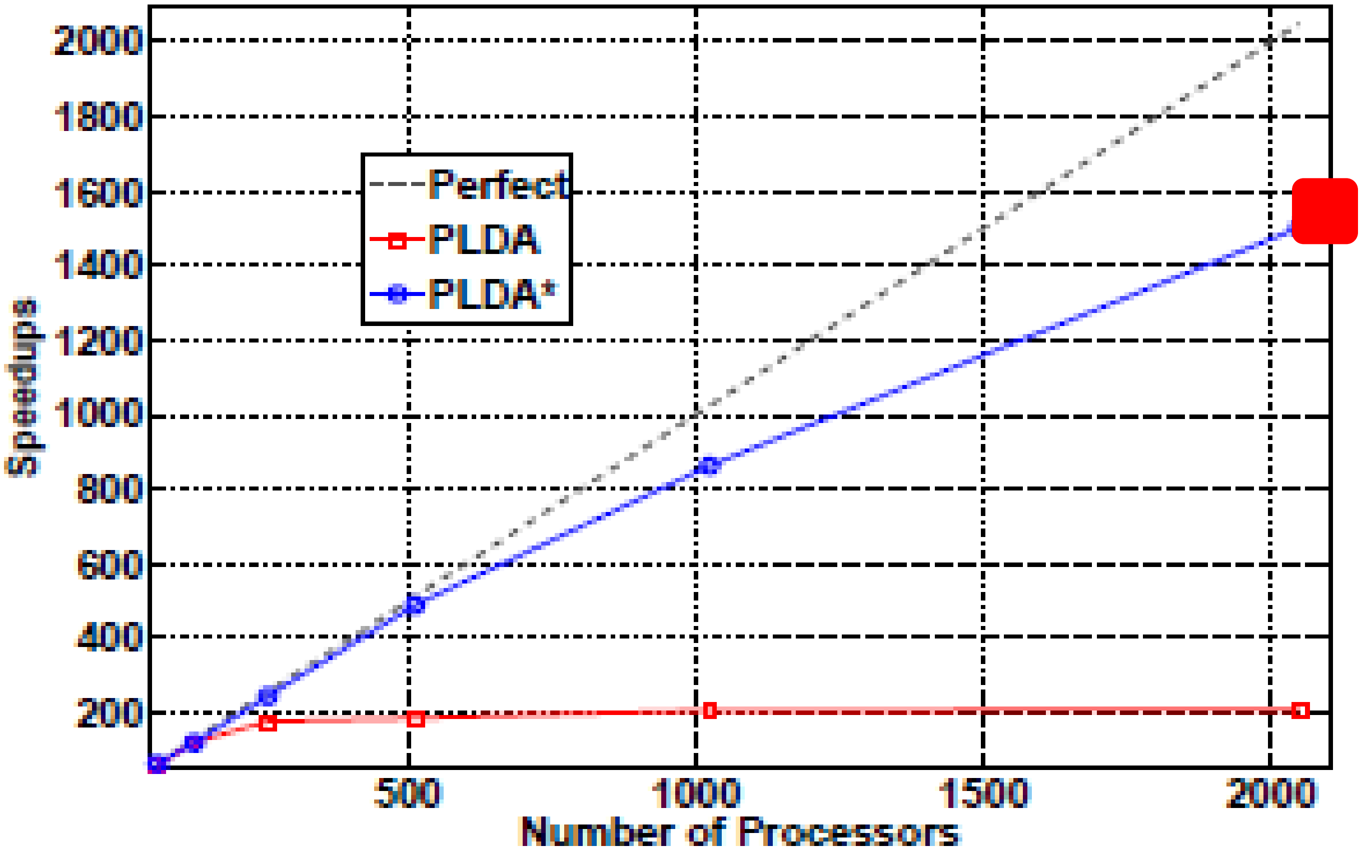


Fig. 4: Pipeline-based Gibbs Sampling in PLDA* . (A): $t_s \geq t_f + t_u$. (B): $t_s < t_f + t_u$.

Speedup

3.2B word occurrences

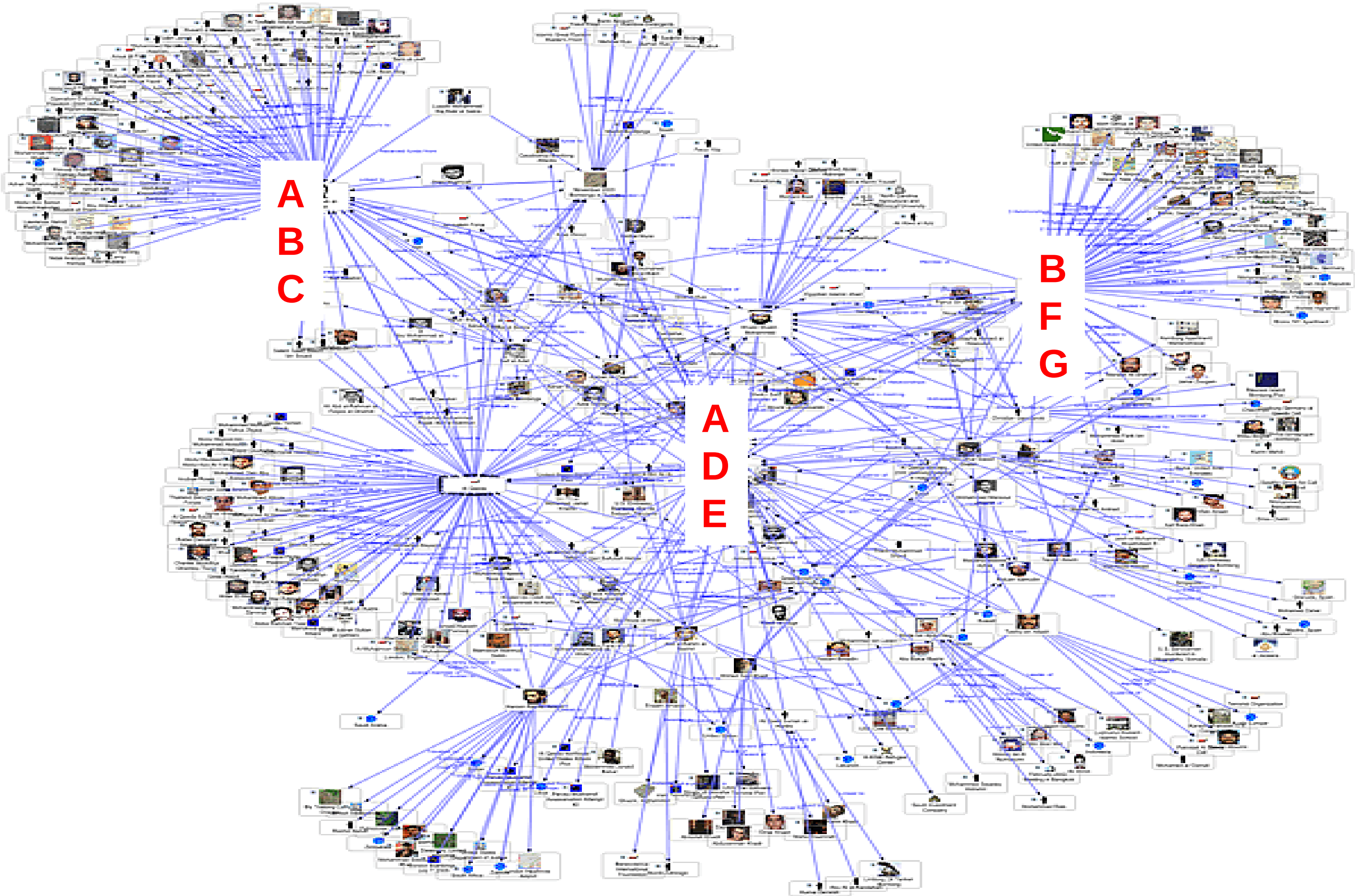
1,500x using 2,000 machines



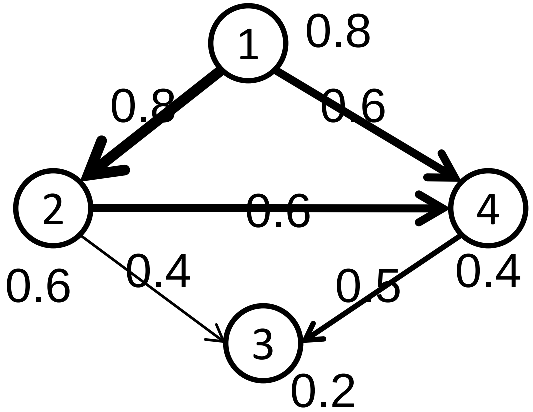
Outline

- Social Network Ad Model
  - Relevance Model
  - Influence Model
- Key Algorithms
  - UserRank
  - Hint Word Generation
  - Diffusion

Influence Analysis, Relevance Analysis, Influence-based Relevance Propagation



Illustrative Example

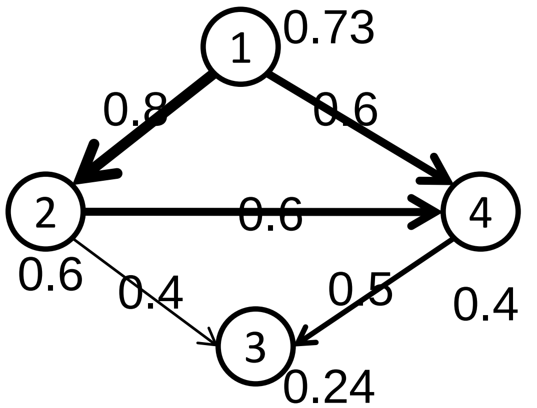


$$h^1 = \begin{bmatrix} 0.8 & 0.6 & 0.2 & 0.4 \end{bmatrix}^T$$

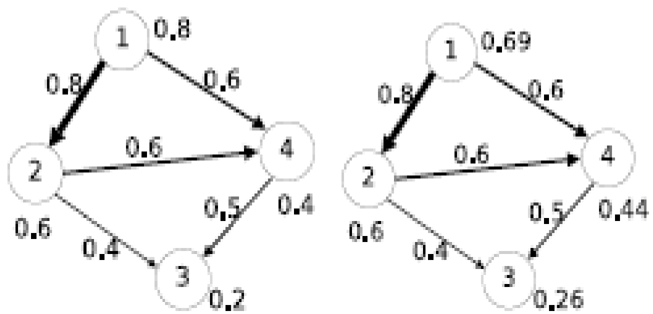
Hint words:
#1: (a, 0.6) (b, 0.4)
#2: (c, 0.8) (b, 0.2)
#3: (e, 0.5) (f, 0.5)
#4: (d, 0.9) (b, 0.1)

Word Propagation:
#1: (a, 0.6) (b, 0.4)
#2: (c, 0.69) (b, 0.23) (a, 0.08)
#3: (e, 0.4) (f, 0.4) (c, 0.1) (d, 0.07) (b, 0.03)
#4: (d, 0.66) (b, 0.16) (a, 0.11) (c, 0.07)

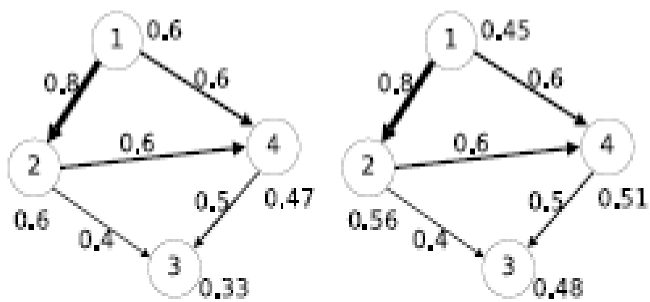
$$h^1 = (1 + \frac{\Gamma \circ \mathrm{A}}{M}) h^0 = [0.73 \quad 0.6 \quad 0.24 \quad 0.4]^T$$



Influence Propagation



(a) $n = 0$ (b) $n = 2$

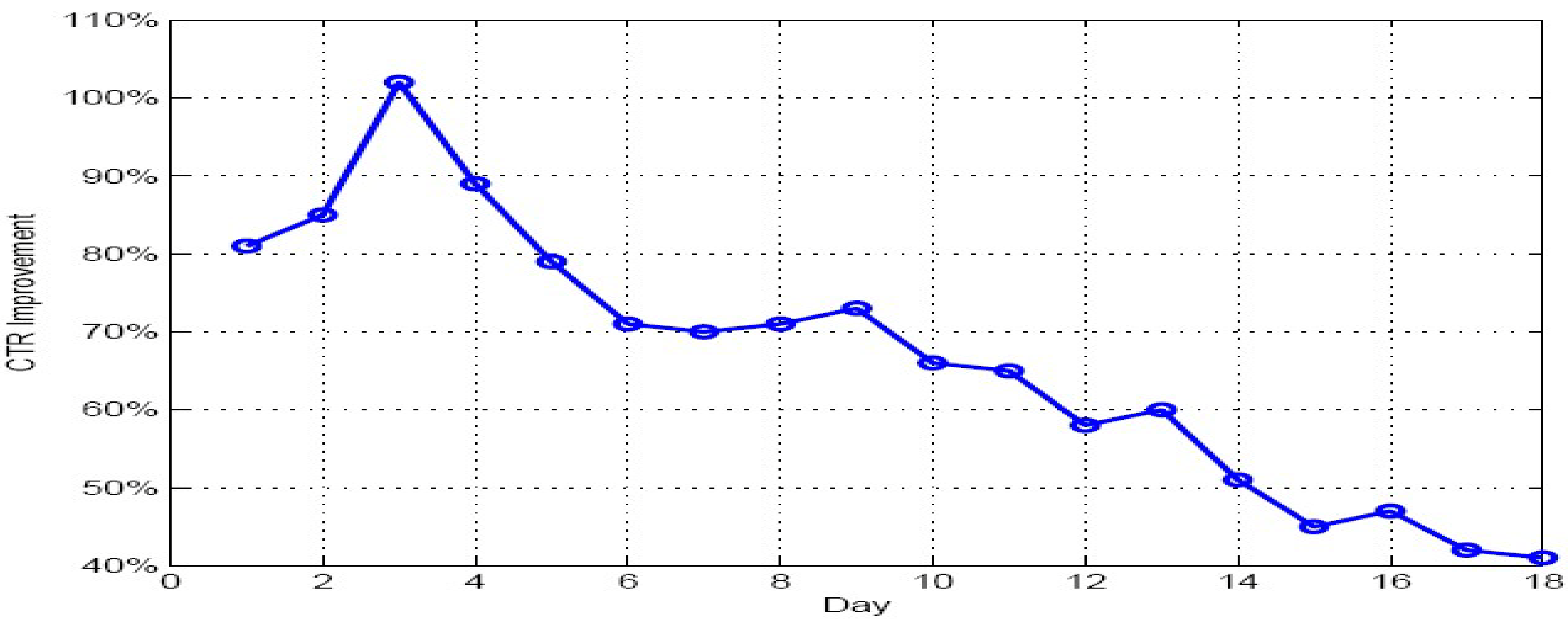


(c) $n = 4$ (d) $n = 8$

n^{th}	User	Hint Words
0	#1	(a, 0.6) (b, 0.4)
	#2	(c, 0.8) (b, 0.2)
	#3	(e, 0.5) (f, 0.5)
	#4	(d, 0.9) (b, 0.1)
1	#1	(a, 0.6) (b, 0.4)
	#2	(c, 0.69) (b, 0.23) (a, 0.08)
	#3	(e, 0.4) (f, 0.4) (c, 0.1) (d, 0.07) (b, 0.03) (a, 0.01)
	#4	(d, 0.66) (a, 0.18) (c, 0.07) (b, 0.01)
2	#1	(a, 0.6) (b, 0.4)
	#2	(c, 0.65) (b 0.24) (a, 0.11)
	#3	(e, 0.32) (f, 0.32) (c, 0.18) (d, 0.11) (b, 0.06) (a, 0.03)
	#4	(d, 0.5) (a, 0.25) (b, 0.15) (c, 0.12)
4	#1	(a, 0.6) (b, 0.4)
	#2	(c, 0.59) (b, 0.25) (a, 0.16)
	#3	(c, 0.26) (e, 0.21) (f, 0.21) (d, 0.13) (b, 0.11) (a, 0.08)
	#4	(d, 0.34) (a, 0.29) (b, 0.21) (c, 0.17)
8	#1	(a, 0.6) (b, 0.4)
	#2	(c, 0.59) (b, 0.25) (a, 0.16)
	#3	(c, 0.33) (b, 0.16) (e, 0.13) (f, 0.13) (a, 0.13) (d, 0.12)
	#4	(a, 0.29) (d, 0.26) (b, 0.23) (c, 0.22)

Influence Model with Propagation

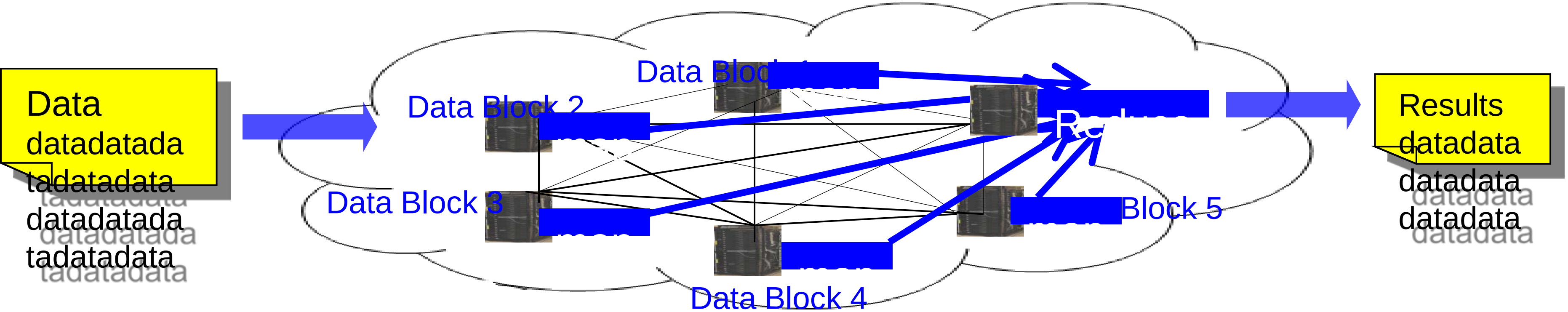
- For two groups of users to be shown ads, G1 and G2
  - G1: AdHeat with propagation (M3)
  - G2: AdHeat without propagation (M2)

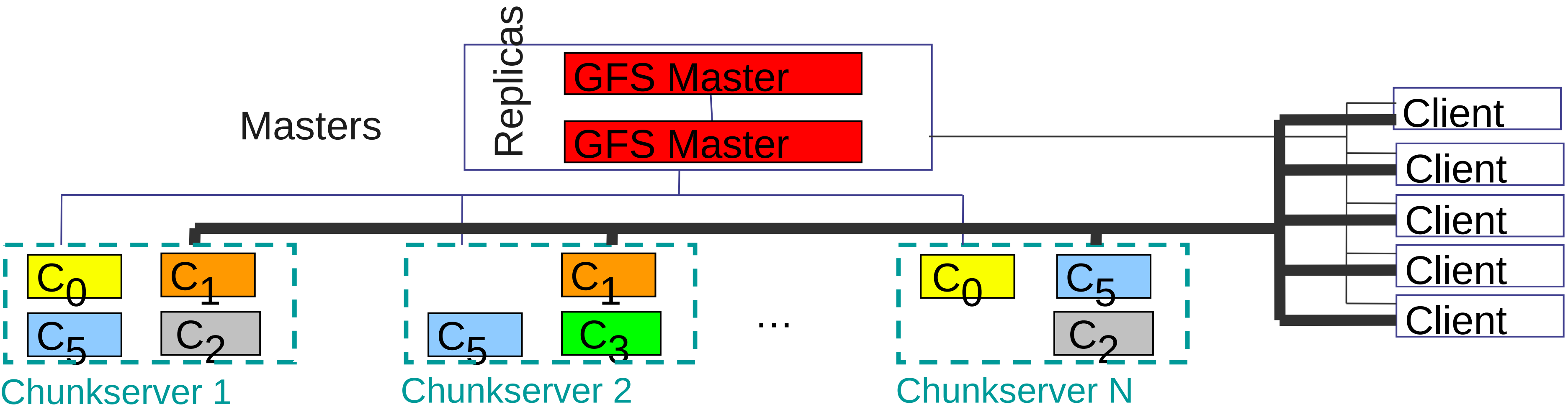


Improvement of Accumulative CTR (M3 vs. M2)

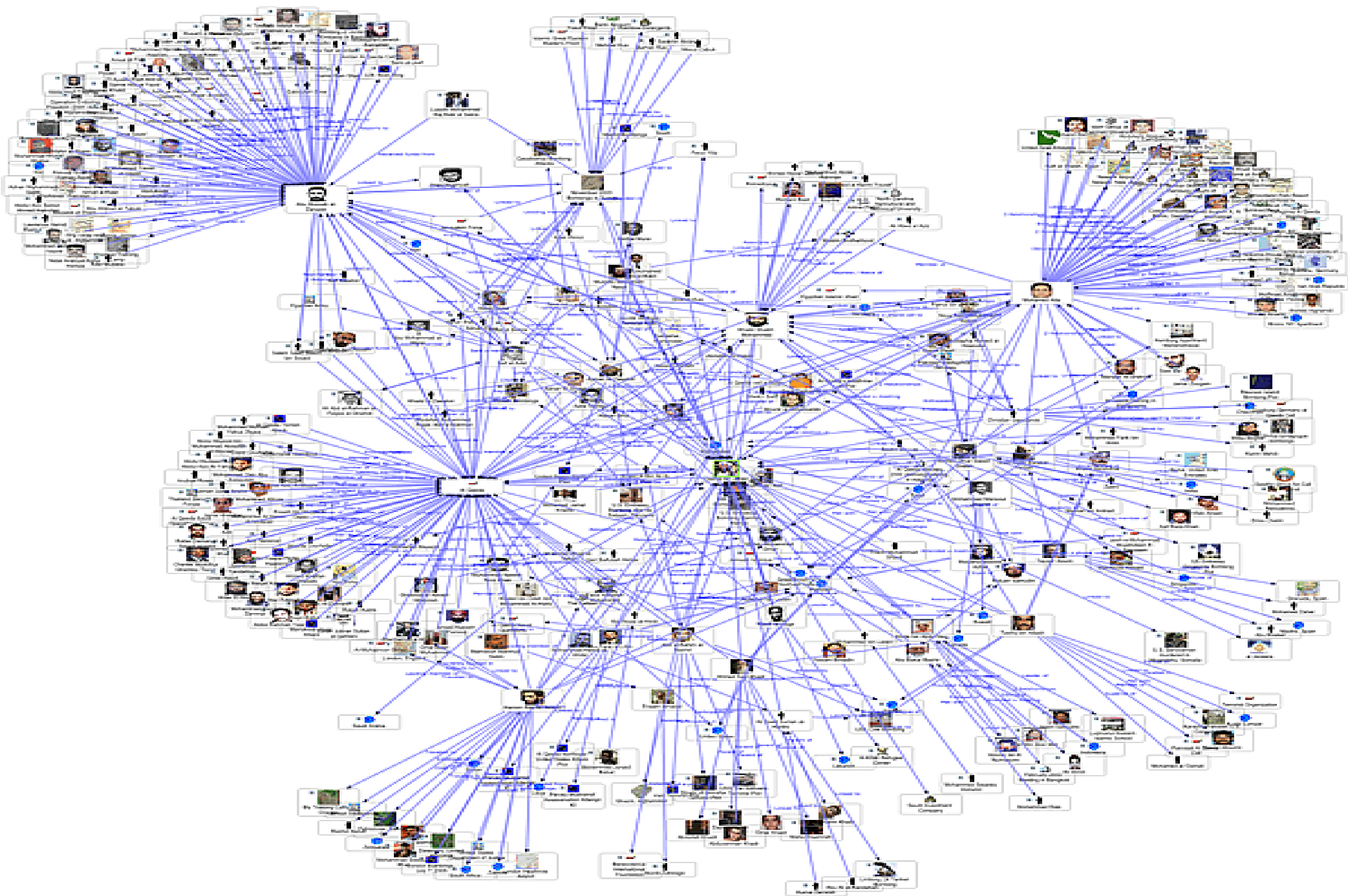
AdWords, AdSense, AdHeat

	Target	Interaction	Propagation	Page	Bid
AdWords	Query	X	X	Google pages	Key words
AdSense	Content	X	X	Web pages	Key words
AdHeat	User	√	√	User Home page	Users


Data
datadatadata
tadatadata
datadatada
tadatadata
Data Block 2
Data Block 3
Data Block 4
Block 5
Results
datadata
datadata
datadata


Masters
Replicas
GFS Master
GFS Master
Client
Client
Client
Client
Client
C_0
C_1
C_5
C_2
C_1
C_5
C_3
…
C_0
C_5
C_2
Chunkserver 1
Chunkserver 2
Chunkserver N

Social Network Analysis



References

- **AdHeat (Social Ads)**:
  - AdHeat: An Influence-based Diffusion Model for Propagating Hints to Match Ads, H.J. Bao and E. Y. Chang, WWW 2010, North Carolina, April 2010.
  - Parallel Spectral Clustering in Distributed Systems, Wen-Yen Chen, Yangqiu Song, Hongjie Bai, Chih-Jen Lin, and E. Y. Chang, IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI), 2010.
- **UserRank:**
  - Confucius and its Intelligent Disciples, X. Si, E. Y. Chang, Z. Gyongyi, VLDB, September 2010 .
  - Topic-dependent User Rank, Xiance Si, Z. Gyongyi, E. Y. Chang, and M.S. Sun, Google Technical Report.
- **Large-scale Collaborative Filtering**:
  - PLDA*: Parallel Latent Dirichlet Allocation for Large-Scale Applications, ACM Transactions on Internet Technology, 2010.
  - Collaborative Filtering for Orkut Communities: Discovery of User Latent Behavior, W.-Y. Chen, J. Chu, E. Y. Chang, WWW 2009: 681-690.
  - Combinational Collaborative Filtering for Personalized Community Recommendation, W.-Y. Chen, E. Y. Chang, KDD 2008: 115-123.
  - Parallel SVMs, E. Y. Chang, et al., NIPS 2007.