

# Efficient Algorithms for Personalization in Social Networks




Ashish Goel

  
**Pankaj Gupta**  
 @pankaj

TWEETS 5,520    FOLLOWERS 949    FOLLOWING 9,329

Compose new Tweet...

**Who to follow** Refresh View all

-  **Safeway** @Safeway  
Followed by Waven Renewal...  
Follow Promoted
-  **Discovery News** @Discovery  
Follow
-  **Michael** @Michael  
Followed by Wavel Kire et al...  
Follow

Popular accounts Find friends

**Trends** Change

- #DinnerDay
- L Wavel South
- #CDA
- Venezuela
- Jim Hay
- DNC
- Brwn
- #KurtHumb
- KateLynn
- Miss

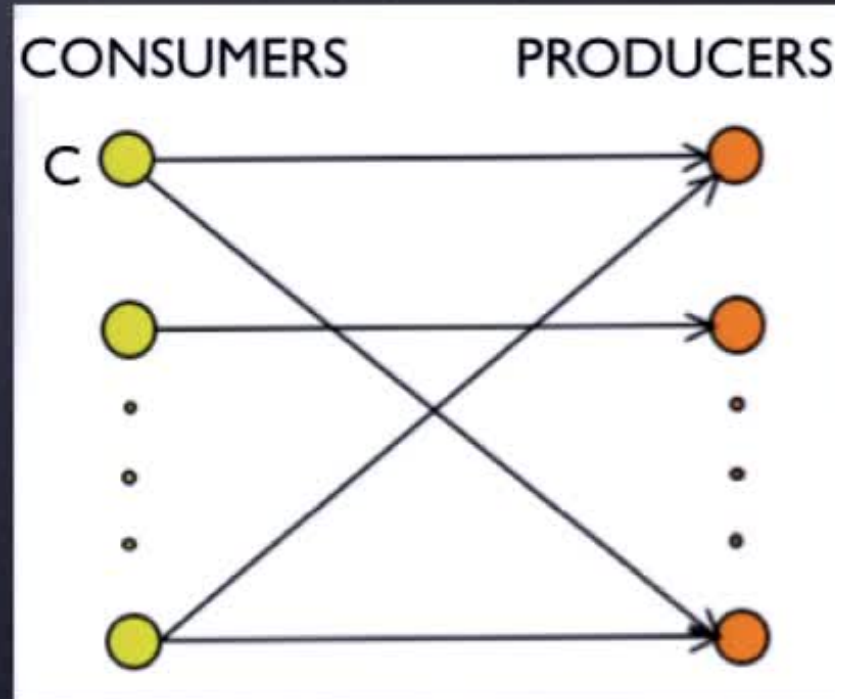
### Tweets

- 3 new Tweets
-  **Aapo Kyrola** @kyrola 1m  
GraphChi used for computational biology: homes di.unim.it/valentini/page...  
Result: As fast as in-memory computation, way faster than Neo4j.  
Expand
  -  **Rohit Ghosh** @Rohit\_G... 3m  
This is big, theguardian.com/science/2014/m... Primordial Gravitational waves  
'seen' in polarization of early light. @Einstein  
View summary
  -  **D-Lab @ MIT** @dlab\_mit 2m  
March 26: Follow-up @harvest\_fuel webinar to March 30th "Charcoal Briquette  
Enterprise Development" @dlab\_mit @harvest\_fuel...  
Expand
  -  **Mark McBride** @mcbri 2m  
today I'm responding to all Messages with "congratulations".  
Expand
  - 1 more reply
  -  **Mark McBride** @mcbri 4m  
@matassar you know that's like me clowning you for the links in DMs debacle  
Expand
  -  **Ben Matassar** @matassar 3m  
@mcbri So?  
Expand
  -  **Eater SF** @eatersf 5m  
Check out the scene at @fortmasonsf's Fort Mason opening night, see what's  
new this year. eater.co/11wewwR po.twitter.com/qjXGMdOdn  


# Collaborative Filtering

To get recommendations for C, compute similarity scores for all consumers, and relevance scores for all producers, with respect to C

1. Start with  $\text{sim}(C) = 1$
2. Propagate similarity scores along graph edges to compute relevance scores, and vice-versa



Many propagation methods; Often, a linear system of equations

# Collaborative Filter: Love or Money

How should we do this propagation? Two extremes:

**LOVE:** All the similarity score of a consumer  $X$  gets transferred to each producer that  $X$  follows, and the same in the reverse direction

→ Analogous to Singular Value Decompositions in the dense graph limit (HITS)

**MONEY:** If  $X$  follows  $d$  producers, then a fraction  $1/d$  of the similarity score of  $X$  gets transferred to each producer that  $X$  follows (SALSA)



# Personalized PageRank

Given a **consumer C**, perform a random walk on the Follow graph. If the walk is at node **v**, then the walk:

- Jumps back to node **C** with probability  $\alpha$
- Follows a random edge out of **v** with probability  $1 - \alpha$

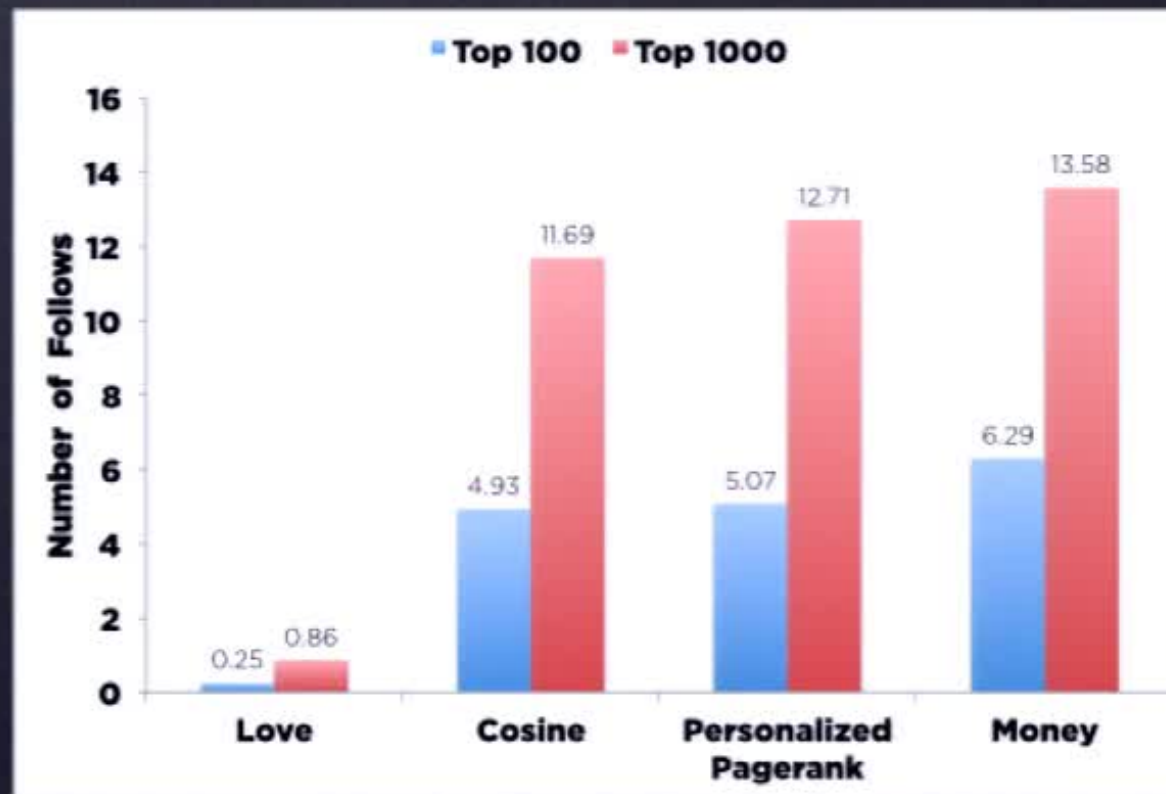
The Personalized PageRank of node **Y** is the weight of **Y** in the stationary distribution of this random walk

SALSA/Money is just Personalized PageRank run on the undirected consumer–producer graph

# A Dark Test

Run various algorithms to predict follows, but don't display the results. Instead, just observe how many of the top **predictions** get followed organically

[Bahmani, Chowdhury, Goel; 2010]



# Personalized PageRank

Given a **consumer C**, perform a random walk on the Follow graph. If the walk is at node **v**, then the walk:

- Jumps back to node **C** with probability  $\alpha$
- Follows a random edge out of **v** with probability  $1 - \alpha$

The Personalized PageRank of node **Y** is the weight of **Y** in the stationary distribution of this random walk

SALSA/Money is just Personalized PageRank run on the undirected consumer–producer graph







# Strategic Impact

Creates billions of new follows every year

- More than 1/8 of new follows are directly via the Who-to-Follow module
- More than 15% of active users (> 36 Million users) make at least one follow every month via this module

# Promoted Tweets and Promoted Accounts

The screenshot displays the Twitter interface with two main sections highlighted by purple ovals. The left section, titled "Who to follow", lists three accounts: "CKM Advisors" (a promoted account), "girish sastry", and "Shiv Ramamurthi". The right section, titled "Tweets", shows a list of tweets, with the bottom tweet from "NewRelic" (a promoted tweet) highlighted.

**Who to follow section:**

- CKM Advisors** (@CKMAdvL...) - Promoted account. Includes a "Follow" button and a "Promoted" badge.
- girish sastry** (@gishsastry) - Includes a "Follow" button.
- Shiv Ramamurthi** (@mogro...) - Includes a "Follow" button.

**Tweets section:**

- Aneesh Sharma** (@aneeshs) - 4m. Tweet: "Feeling lucky to be at #analytics2014 with @ashishgoel @johnsirois @pankaj @sgurumur for our #edelmanaward presentation. Go #teammwitter!"
- John Sirois** (@johnsirois) - 5m. Tweet: "Hanging out with @ashishgoel @sgurumur @pankaj @aneeshs #analytics2014. Special thanks to our #edelmanaward coaches John Birge & Carrie Beam"
- NewRelic** (@newrelic) - Mar 11. Tweet: "4 Essential Tips from the Coding CEO. How New Relic CEO Lew Cime still builds product: [blog.newrelic.com/2014/03/11/sxs...](http://blog.newrelic.com/2014/03/11/sxs...)" - Promoted by NewRelic.

# Impact on Revenue

“The Who-To-Follow system was crucial, in a fundamental way, for the Promoted Accounts product, and the Promoted Tweets product also initially used the Who-To-Follow system’s targeting”

– Alex Roetter (VP of Engineering, Revenue)

# Need for Efficient Algorithms

1. Fast Cosine Similarity
2. Fast Incremental PageRank
3. Fast Personalized PageRank
4. Computing via Intersections



# Promoted Tweets and Promoted Accounts

The screenshot displays the Twitter mobile application interface. At the top, navigation icons for Home, Notifications, Search, Profile, and the Twitter bird logo are visible. The left sidebar is titled "Who to follow" and lists three accounts: "CKM Advisors" (with a "Promoted" badge), "girish sastry", and "Shiv Ramamurthi". The right sidebar is titled "Tweets" and shows a "1 new Tweet" notification. It features three tweets: one from Aneesh Sharma, one from John Sirois, and one from NewRelic (which is also marked as "Promoted by NewRelic").

**Who to follow** · Refresh · View all

- CKM Advisors** @CKMAdvi... ✕  
Follow Promoted
- girish sastry** @girishsastry ✕  
Followed by Utkarsh Srivast...  
Follow
- Shiv Ramamurthi** @mogro... ✕  
Followed by Stanford Alumn...  
Follow

Popular accounts · Find friends

**Tweets**

1 new Tweet

- Aneesh Sharma** @aneeshs · 4m  
Feeling lucky to be at #analytics2014 with @ashishgoel @johnsirois @pankaj @sgurumur for our #edelmanaward presentation. Go #teamtwitter!  
Expand Reply Retweet Favorite More
- John Sirois** @johnsirois · 5m  
Hanging out with @ashishgoel @sgurumur @pankaj @aneeshs #analytics2014. Special thanks to our #edelmanaward coaches John Birge & Carrie Beam  
Expand Reply Retweet Favorite More
- NewRelic** @newrelic · Mar 11  
4 Essential Tips from the Coding CEO. How New Relic CEO Lew Cime still builds product: [blog.newrelic.com/2014/03/11/sxs...](http://blog.newrelic.com/2014/03/11/sxs...)  
Promoted by NewRelic  
Expand Reply Retweet Favorite More

# Strategic Impact

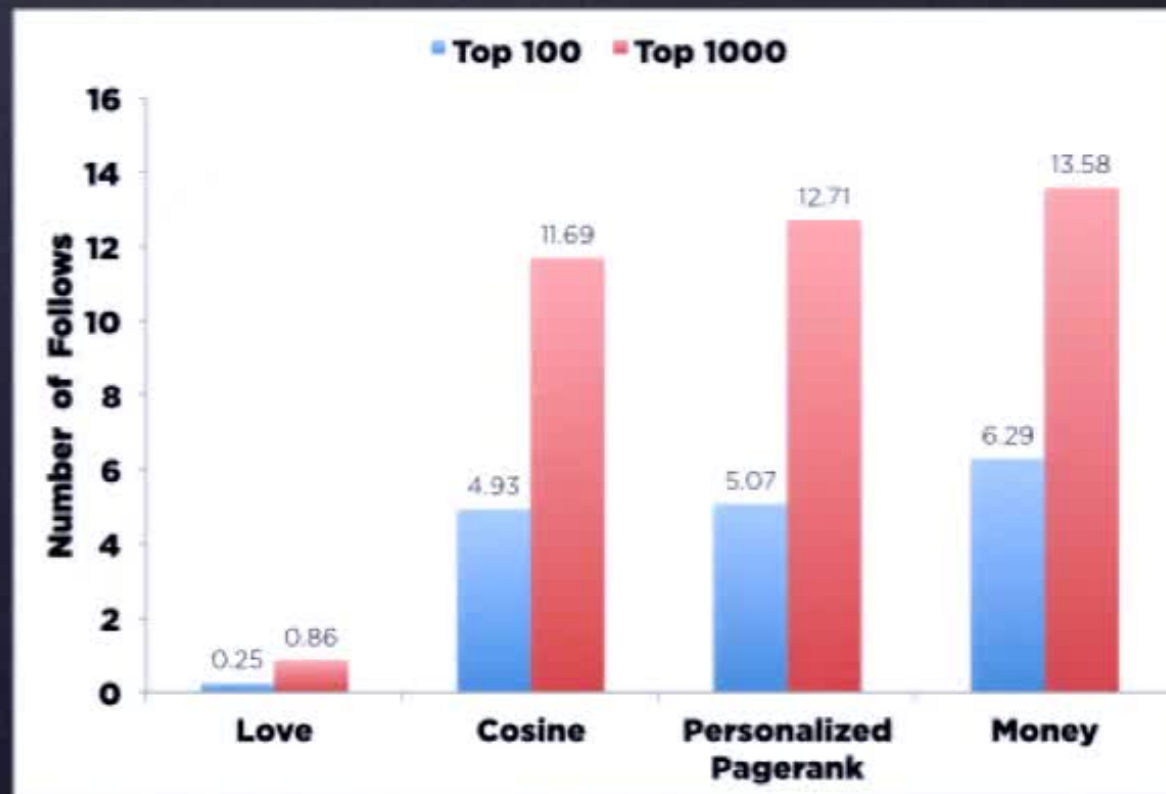
Creates billions of new follows every year

- More than 1/8 of new follows are directly via the Who-to-Follow module
- More than 15% of active users (> 36 Million users) make at least one follow every month via this module

# A Dark Test

Run various algorithms to predict follows, but don't display the results. Instead, just observe how many of the top **predictions** get followed organically

[Bahmani, Chowdhury, Goel; 2010]



# Personalized PageRank

Given a **consumer C**, perform a random walk on the Follow graph. If the walk is at node **v**, then the walk:

- Jumps back to node **C** with probability  $\alpha$
- Follows a random edge out of **v** with probability  $1 - \alpha$

The Personalized PageRank of node **Y** is the weight of **Y** in the stationary distribution of this random walk

SALSA/Money is just Personalized PageRank run on the undirected consumer–producer graph



# Collaborative Filter: Love or Money

How should we do this propagation? Two extremes:

**LOVE:** All the similarity score of a consumer  $X$  gets transferred to **each** producer that  $X$  follows, and the same in the reverse direction

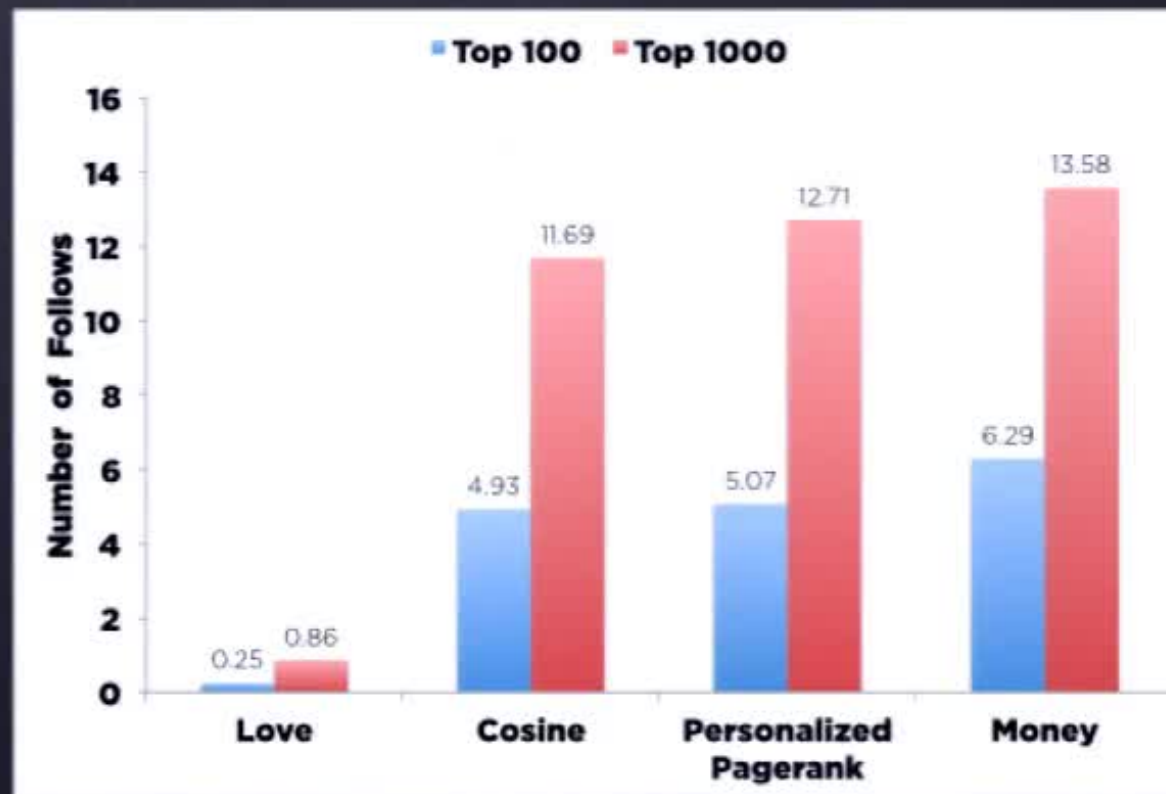
→ Analogous to Singular Value Decompositions in the dense graph limit (HITS)

**MONEY:** If  $X$  follows  $d$  producers, then a fraction  $1/d$  of the similarity score of  $X$  gets transferred to each producer that  $X$  follows (SALSA)

# A Dark Test

Run various algorithms to predict follows, but don't display the results. Instead, just observe how many of the top **predictions** get followed organically

[Bahmani, Chowdhury, Goel; 2010]



# Strategic Impact

Creates billions of new follows every year

- More than 1/8 of new follows are directly via the Who-to-Follow module
- More than 15% of active users (> 36 Million users) make at least one follow every month via this module

# Promoted Tweets and Promoted Accounts

The screenshot shows the Twitter mobile app interface. On the left, the 'Who to follow' section is highlighted with a blue border. It lists three accounts: CKM Advisors (@CKMAdvL...), girish sastry (@girishsastry), and Shiv Ramamurthi (@mogro...). The CKM Advisors account is marked as 'Promoted'. On the right, the 'Tweets' section shows a list of tweets. The tweet by NewRelic (@newrelic) is circled in purple. It is a promoted tweet with a yellow 'Promoted' label and the text 'Promoted by NewRelic'.

Who to follow · Refresh · View all

- CKM Advisors** @CKMAdvL... Follow Promoted
- girish sastry** @girishsastry Follow  
Followed by Utkarsh Srivast...
- Shiv Ramamurthi** @mogro... Follow  
Followed by Stanford Alumn...

Popular accounts · Find friends

Tweets

1 new Tweet

- Aneesh Sharma** @aneeshs · 4m  
Feeling lucky to be at #analytics2014 with @ashishgoel @johnsirois @pankaj @sgurumur for our #edelmanaward presentation. Go #teambtwitter!
- John Sirois** @johnsirois · 5m  
Hanging out with @ashishgoel @sgurumur @pankaj @aneeshs #analytics2014. Special thanks to our #edelmanaward coaches John Birge & Carrie Beam
- NewRelic** @newrelic · Mar 11  
4 Essential Tips from the Coding CEO. How New Relic CEO Lew Cime still builds product: [blog.newrelic.com/2014/03/11/sxs...](http://blog.newrelic.com/2014/03/11/sxs...)  
Promoted by NewRelic



# Impact on Revenue

“The Who-To-Follow system was crucial, in a fundamental way, for the Promoted Accounts product, and the Promoted Tweets product also initially used the Who-To-Follow system’s targeting”

– Alex Roetter (VP of Engineering, Revenue)

# Need for Efficient Algorithms

1. Fast Cosine Similarity
2. Fast Incremental PageRank
3. Fast Personalized PageRank
4. Computing via Intersections

PowerPoint File Edit View Insert Format Arra... Tools Slide... Wind... TexP... Help

disco.pptx

Search in Presentation

Home Themes Tables Charts SmartArt Transitions Animations Slide Show

Slides Font Paragraph Insert

# Keyword similarity in Tweets

- Consider a corpus of  $N$  documents (eg. Tweets) of length  $L$  each, and a dictionary of  $D$  words
  - Typical values:  $N = 1B$ ,  $L = 10$ ,  $D = 100K$
- Goal: Find cosine similarity of all pairs of words

$$\text{CosineSimilarity}(w_1, w_2) = \frac{\text{Count}\{w_1, w_2\}}{\sqrt{\text{Count}\{w_1\} * \text{Count}\{w_2\}}}$$

- Critically important problem; also used to identify similar nodes in networks
- We only need cosine similarity for nodes/words that are

Click to add notes

Normal View Slide 9 of 13 75%



PowerPoint File Edit View Insert Format Arra... Tools Slide... Wind... TexP... Help

disco.pptx

Search in Presentation

Home Themes Tables Charts SmartArt Transitions Animations Slide Show

Slides Font Paragraph Insert

# Data Model: Map Reduce

- An immensely successful idea which transformed offline analytics and bulk-data processing. Hadoop (initially from Yahoo!) is the most popular implementation.
- **MAP:** Transforms a (key, value) pair into other (key, value) pairs using a UDF (User Defined Function) called Map. Many mappers can run in parallel on vast amounts of data in a distributed file system
- **SHUFFLE:** The infrastructure then transfers data from the mapper nodes to the "reducer" nodes so that all the (key, value) pairs with the same key go to the same reducer and get grouped into a single large (key, <val<sub>1</sub>, val<sub>2</sub>, ..>) pair

Click to add notes

Normal View Slide 9 of 13 75%



# Data Model: Map Reduce

- An immensely successful idea which transformed offline analytics and bulk-data processing. Hadoop (initially from Yahoo!) is the most popular implementation.
- **MAP:** Transforms a (key, value) pair into other (key, value) pairs using a UDF (User Defined Function) called Map. Many mappers can run in parallel on vast amounts of data in a distributed file system
- **SHUFFLE:** The infrastructure then transfers data from the mapper nodes to the “reducer” nodes so that all the (key, value) pairs with the same key go to the same reducer and get grouped into a single large (key, <val<sub>1</sub>, val<sub>2</sub>, ..>) pair
- **REDUCE:** A UDF that processes this grouped (key, <val<sub>1</sub>, val<sub>2</sub>, ..>) pair for a single key. Many reducers can run in parallel.

# Complexity Measures

- Key-Complexity:
  - The maximum size of a key-value pair
  - The amount of time taken to process each key
  - The memory required to process each key
- Sequential Complexity:
  - The total time needed by all the mappers and reducers together
  - The total output produced by all the mappers and reducers together
- Number of MapReduce phases

[Goel, Munagala; 2012]

Complexity

THE AMOUNT OF WORK DONE PER COMPUTER IF WE HAD INFINITELY MANY COMPUTERS

- Key-Complexity:
  - The maximum size of a key-value pair
  - The amount of time taken to process each key
  - The memory required to process each key
- Sequential Complexity:
  - The total time needed by all the mappers and reducers together
  - The total output produced by all the mappers and reducers together

[Goel, Munagala; 2012]

# Complexity

OFTEN, IMPORTANT ONLY FOR REDUCERS


- Key-Complexity:
  - The maximum size of a key-value pair
  - The amount of time taken to process each key
  - The memory required to process each key
- Sequential Complexity:
  - The total time needed by all the mappers and reducers together
  - The total output produced by all the mappers and reducers together

[Goel, Munagala; 2012]



# Complexity Measures

- Key-Complexity:
  - The maximum size of a key
  - The amount of time to process each key
  - The memory required to process each key
- Sequential Complexity:
  - The total time needed by all the mappers and reducers together
  - The total output produced by all the mappers and reducers together



THE AMOUNT OF WORK  
DONE ON A SINGLE  
COMPUTER

# Complexity Measures

- Key-Complexity:
  - The maximum size of a key-value pair
  - The amount of time taken to process each key
  - The memory required to process each key
- Sequential Complexity:
  - The total time needed by all the mappers together
  - The total output produced by all the mappers and reducers together



SHUFFLE SIZE

# Complexity Measures

- Key-Complexity:
  - The maximum number of values for a single key
  - The amount of time needed to sort the values for a single key
  - The memory required to process each key
- Sequential Complexity:
  - The total time needed by all the mappers and reducers together
  - The total output produced by all the mappers and reducers together

THE AMOUNT OF WORK  
DONE TO AGGREGATE ALL  
THE VALUES FOR A SINGLE  
KEY (SORTING) IS NOT A  
COMPLEXITY MEASURE

# Keyword similarity in Tweets

- Consider a corpus of  $N$  documents (eg. Tweets) of length  $L$  each, and a dictionary of  $D$  words
  - Typical values:  $N = 1B$ ,  $L = 10$ ,  $D = 100K$
- Goal: Find cosine similarity of all pairs of words

$$\text{CosineSimilarity}(w1, w2) = \frac{\text{Count}\{w1, w2\}}{\sqrt{\text{Count}\{w1\} * \text{Count}\{w2\}}}$$

- Critically important problem; also used to identify similar nodes in networks
- We only need cosine similarity for nodes/words that are indeed very similar (say  $\text{CosineSimilarity} > \epsilon$ , for  $\epsilon=0.1$ )



# Example Application

USER-USER  
SIMILARITY

The image shows a screenshot of a Twitter profile page for Jason Goldman (@goldman). The page is divided into several sections:

- Header:** Displays the user's name "Jason Goldman" and handle "@goldman".
- Tweet:** A tweet by Jason Goldman mentioning "Tosca, Act I. Many similarities" with an "Expand" link.
- Similar to Jason Goldman:** A section titled "Similar to Jason Goldman" with a right-pointing arrow. It lists three suggested accounts:
  - BestAt: The Best @'s** (@BestAt) with a "Follow" button.
  - Truebe** (@Truebe) with a "Follow" button.
  - Rose** (@rose) with a "Follow" button.
- Tweets:** A list of tweets by Jason Goldman:
  - Tweet 1: "@lukester I ate 5lbs of sashimi" with a "View conversation" link.
  - Tweet 2: "@joshelman @peterpham ca" with a "View conversation" link.
  - Tweet 3: "I can feel fantasy sports lurking... obsession that will one day..." with an "Expand" link.
  - Tweet 4: "@iano it was like your Golden... being turned into a blueberry..."
- Footer:** Copyright notice: "© 2012 Twitter" and links for "About", "Help", "Terms", "Privacy", "Blog", "Status", "Apps", "Resources", "Jobs", "Advertisers", "Businesses", "Media", "Developers".

# Cosine Similarity [Brute Force]

- Assumption: The tweet corpus is annotated with word counts, so every occurrence of every word is tagged with the frequency of that word as well
  - Need this for other reasons anyway (eg. trends, search)
- Count all co-occurrences between all words
  - `MAP(tweet)`: for every pair of words  $w_1, w_2$  in the tweet, `EMIT({w1, w2}, 1)`
  - `REDUCE({w1, w2}, <1, 1, ...>)`: If the size of the value vector (i.e. number of 1's) is larger than  $\epsilon \sqrt{\text{Count}\{w_1\} * \text{Count}\{w_2\}}$  then `EMIT({w1, w2})`
- Sequential complexity: Requires shuffling  $N * L * L$  data across in Hadoop,  $\approx$  100 Billion records
- Reduce-key complexity: Could be as large as  $N$ , or with combining,  $K$  (the number of mappers)
- Observation: most of the data shuffled by brute force algorithm was being wasted

# Cosine Similarity [Random Sampling]

$$\text{CosineSimilarity}(w1, w2) = \frac{\text{Count}\{w1, w2\}}{\sqrt{\text{Count}\{w1\} * \text{Count}\{w2\}}}$$

- **MAP(tweet):** for every pair of words  $w1, w2$  in the tweet, **EMIT**( $\{w1, w2\}, 1$ ) with probability

$$R / (\sqrt{\text{Count}\{w1\} * \text{Count}\{w2\}}),$$

where  $R = (\log D) / \epsilon \approx 100$

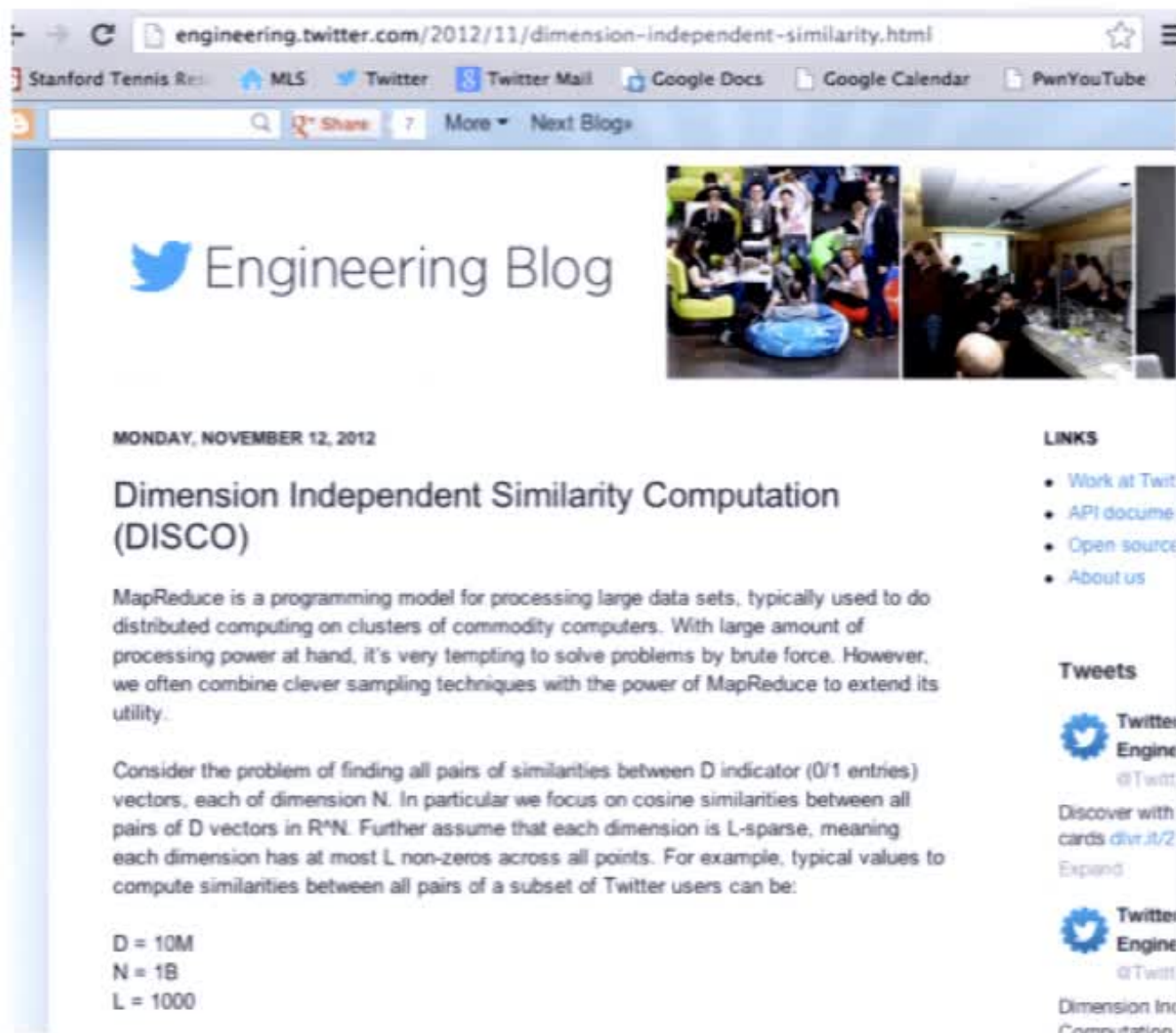
- **REDUCE**( $\{w1, w2\}, \langle 1, 1, \dots \rangle$ ): **EMIT**( $\{w1, w2\}, (\text{size of value-vector}) / R$ )
- Unbiased estimate of CosineSimilarity, and accurate whp when  $\text{CosineSimilarity} > \epsilon$
- Expected Reduce-Key complexity: At most  $R$
- Sequential complexity: Shuffle size goes down from  $NL^2$  to around  $DRL$ . ( $\approx 100B \rightarrow \approx 100M$ )



# Cosine Similarity [contd]

- In production at Twitter
- Described in recent twitter engineering blog post

Bogah-Zadeh and Goel: Dimension Independent Similarity Computation






The screenshot shows a web browser displaying a blog post from the Twitter Engineering Blog. The URL is `engineering.twitter.com/2012/11/dimension-independent-similarity.html`. The page features the Twitter Engineering Blog logo and a header image showing people at a conference. The main content is a blog post dated Monday, November 12, 2012, titled "Dimension Independent Similarity Computation (DISCO)". The post discusses MapReduce and cosine similarity. It includes a list of parameters:  $D = 10M$ ,  $N = 1B$ , and  $L = 1000$ . On the right side, there are sections for "LINKS" and "Tweets".

engineering.twitter.com/2012/11/dimension-independent-similarity.html

Stanford Tennis Re: MLS Twitter Twitter Mail Google Docs Google Calendar PwnYouTube

Share 7 More Next Blogs

 Engineering Blog

MONDAY, NOVEMBER 12, 2012

## Dimension Independent Similarity Computation (DISCO)

MapReduce is a programming model for processing large data sets, typically used to do distributed computing on clusters of commodity computers. With large amount of processing power at hand, it's very tempting to solve problems by brute force. However, we often combine clever sampling techniques with the power of MapReduce to extend its utility.


Consider the problem of finding all pairs of similarities between  $D$  indicator (0/1 entries) vectors, each of dimension  $N$ . In particular we focus on cosine similarities between all pairs of  $D$  vectors in  $R^N$ . Further assume that each dimension is  $L$ -sparse, meaning each dimension has at most  $L$  non-zeros across all points. For example, typical values to compute similarities between all pairs of a subset of Twitter users can be:


$D = 10M$   
 $N = 1B$   
 $L = 1000$

**LINKS**

- [Work at Twitter](#)
- [API documents](#)
- [Open source](#)
- [About us](#)

**Tweets**

 **Twitter Engine**  
@Twitter  
Discover with cards [div.it/2](#)  
Expand

 **Twitter Engine**  
@Twitter  
Dimension Inc  
Computation

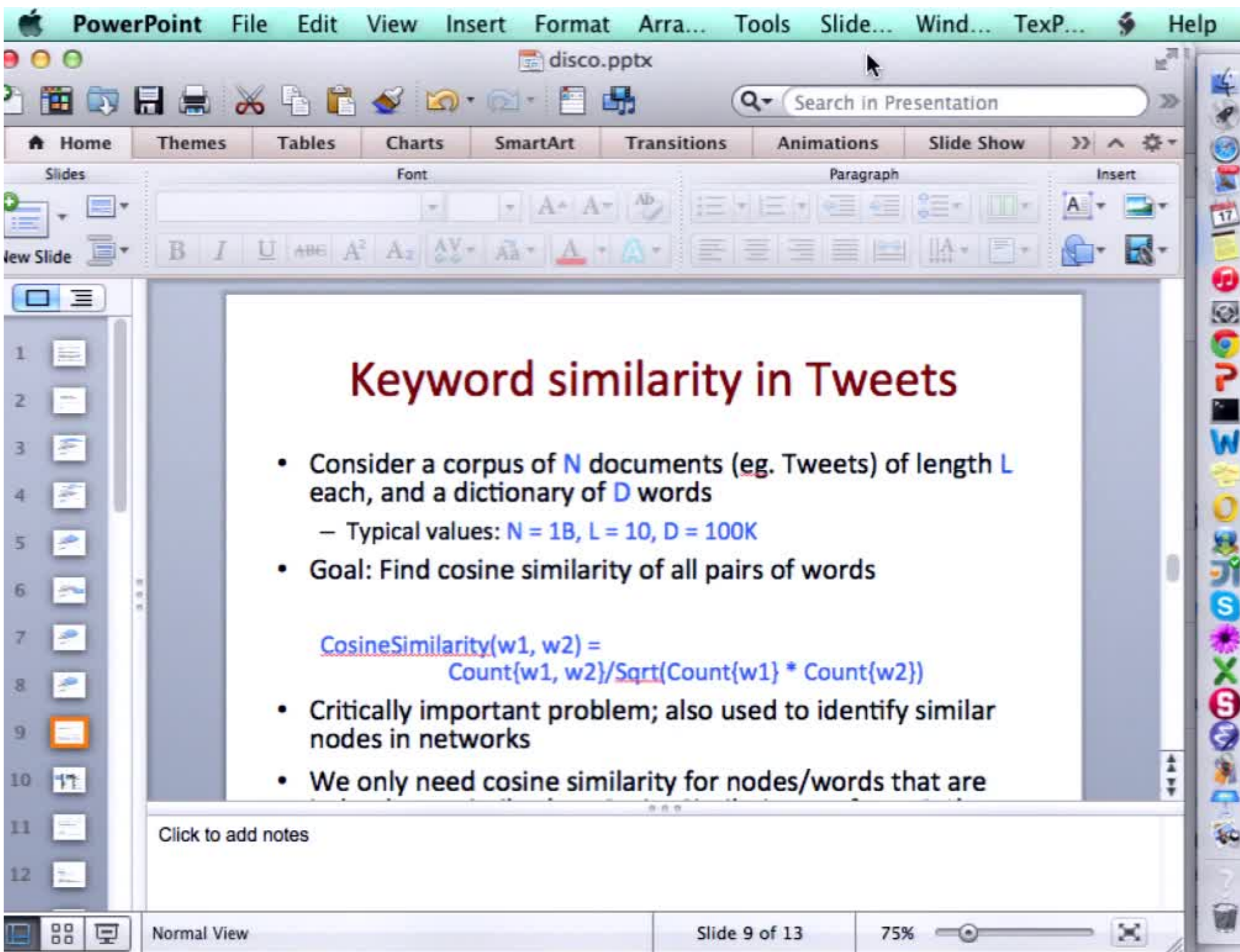


# Keyword similarity in Tweets

- Consider a corpus of  $N$  documents (eg. Tweets) of length  $L$  each, and a dictionary of  $D$  words
  - Typical values:  $N = 1B$ ,  $L = 10$ ,  $D = 100K$
- Goal: Find cosine similarity of all pairs of words

$$\text{CosineSimilarity}(w1, w2) = \frac{\text{Count}\{w1, w2\}}{\text{Sqrt}(\text{Count}\{w1\} * \text{Count}\{w2\})}$$

- Critically important problem; also used to identify similar nodes in networks
- We only need cosine similarity for nodes/words that are indeed very similar (say  $\text{CosineSimilarity} > \epsilon$ , for  $\epsilon=0.1$ )



# Need for Efficient Algorithms

1. Fast Cosine Similarity
2. Fast Incremental PageRank
3. Fast Personalized PageRank
4. Computing via Intersections



# Incremental PageRank

Updates to social graph are made in real-time

- As opposed to a batched crawl process for web search
- Real-time updates to PageRank are important to capture trending events

Goal: Design an algorithm to update PageRank incrementally (i.e. upon an edge arrival)

- $t$ -th edge arrival: Let  $(u_t, v_t)$  denote the arriving edge,  $d_t(v)$  denote the out-degree of node  $v$ , and  $\pi_t(v)$  its PageRank



# Incremental PageRank via Monte Carlo

Start with  $R = O(\log N)$  random walks from every node

At time  $t$ , for every random walk through node  $u_t$ , re-route it to use the new edge  $(u_t, v_t)$  with probability  $1/d_t(u_t)$

→ Time/number of network-calls for each re-routing:  $O(1/\alpha)$

Claim: This faithfully maintains  $R$  random walks after arbitrary edge arrivals

Need the graph and the stored random walks in fast distributed memory

# Promoted Tweets and Promoted Accounts

The screenshot displays the Twitter interface with two main sections highlighted by purple ovals. The left section, titled "Who to follow", lists three accounts: "CKM Advisors" (a promoted account), "girish sastry", and "Shiv Ramamurthi". The right section, titled "Tweets", shows a tweet from "NewRelic" which is also a promoted tweet. The "Promoted" labels are highlighted in purple ovals.

**Who to follow** · Refresh · View all

- CKM Advisors** @CKMAdvi... **Follow** **Promoted**
- girish sastry** @girsisastry **Follow**  
Followed by Utkarsh Srivast...
- Shiv Ramamurthi** @mogro... **Follow**  
Followed by Stanford Alumn...

**Tweets**

1 new Tweet

- Aneesh Sharma** @aneeshs · 4m  
Feeling lucky to be at #analytics2014 with @ashishgoel @johnsirois @pankaj @sgurumur for our #edelmanaward presentation. Go #teamtwitter!  
Expand
- John Sirois** @johnsirois · 5m  
Hanging out with @ashishgoel @sgurumur @pankaj @aneeshs #analytics2014. Special thanks to our #edelmanaward coaches John Birge & Carrie Beam  
Expand
- NewRelic** @newrelic · Mar 11  
4 Essential Tips from the Coding CEO. How New Relic CEO Lew Cirne still builds product: [blog.newrelic.com/2014/03/11/sxs...](http://blog.newrelic.com/2014/03/11/sxs...)  
**Promoted by NewRelic**  
Expand

# Incremental PageRank

Updates to social graph are made in real-time

- As opposed to a batched crawl process for web search
- Real-time updates to PageRank are important to capture trending events

Goal: Design an algorithm to update PageRank incrementally (i.e. upon an edge arrival)

- $t$ -th edge arrival: Let  $(u_t, v_t)$  denote the arriving edge,  $d_t(v)$  denote the out-degree of node  $v$ , and  $\pi_t(v)$  its PageRank



# Incremental PageRank via Monte Carlo

Start with  $R = O(\log N)$  random walks from every node

At time  $t$ , for every random walk through node  $u_t$ , re-route it to use the new edge  $(u_t, v_t)$  with probability  $1/d_t(u_t)$

→ Time/number of network-calls for each re-routing:  $O(1/\alpha)$

Claim: This faithfully maintains  $R$  random walks after arbitrary edge arrivals

Need the graph and the stored random walks in fast distributed memory



# Incremental PageRank Time

Assume that the edges of the graph are chosen by an adversary, but then presented in random order

Theorem: # of **Power iteration takes time  $M R/\alpha$**  per arrival goes to 0

→ t-th arrival: # of **Power iteration takes time  $M R/\alpha$**   $N R/(\alpha t)$

**$N \log N/\alpha$  vs  $M$**

→ Total time over  $M$  arrivals =  $O((N R \log N)/\alpha^2)$

→ Comparable to doing **power iteration**/Monte Carlo just once!

[Bahmani, Goel, Chowdhury, VLDB 2010]

# Personalized PageRank

Network-based Personalized Search is not yet mature

Missing technical piece: Efficient algorithms for Personalized PageRank Queries

→ Given source  $s$  and target  $t$ , estimate the Personalized PageRank of  $t$  for  $s$  with high accuracy, if it is greater than  $\delta$

# Incremental PageRank Time

Assume that the edges of the graph are chosen by an adversary, but then presented in random order

Theorem: # of **Power iteration takes time  $M R/\alpha$**  per arrival goes to 0

→ t-th arrival: # of **Power iteration takes time  $M R/\alpha$**   $N R/(\alpha t)$

→ Total time over M arrivals =  $O((N R \log N)/\alpha^2)$

→ Comparable to doing **power iteration**/Monte Carlo just once!

[Bahmani, Goel, Chowdhury, VLDB 2010]

# Personalized PageRank

Network-based Personalized Search is not yet mature

Missing technical piece: Efficient algorithms for Personalized PageRank Queries

→ Given source  $s$  and target  $t$ , estimate the Personalized PageRank of  $t$  for  $s$  with high accuracy, if it is greater than  $\delta$



Web News Images Videos Maps More Search tools

About 838,000 results (0.50 seconds)

### Nils Lofgren - Wikipedia, the free encyclopedia

en.wikipedia.org/wiki/Nils\_Lofgren

Nils Hilmer Lofgren (born June 21, 1951) is an American rock musician, recording artist, songwriter, and multi-instrumentalist. Along with his work as a solo artist, ...

Nils Lofgren (album) - Cry Tough (Nils Lofgren album) - Night After Night

### Löfgren syndrome - Wikipedia, the free encyclopedia

en.wikipedia.org/wiki/Löfgren\_syndrome

Löfgren syndrome is a type of acute sarcoidosis that is frequent in Scandinavian, Irish, African and Puerto Rican women. It was characterized in 1953 by Sven ...

Presentation - Prognosis - Genetics - Treatment Options

### In the news



### Nils Lofgren ready to cross the Jersey state line

Asbury Park Press - 19 hours ago

Ni

Musi

Nils Lofgren American recording artist and writer Wikipedia

Born in Illinois

Heig

Spot

Musi (Since)

Com Silen



About 838,000 results (0.50 seconds)

### Nils Lofgren - Wikipedia, the free encyclopedia

[en.wikipedia.org/wiki/Nils\\_Lofgren](http://en.wikipedia.org/wiki/Nils_Lofgren)

Nils Hilmer Lofgren (born June 21, 1951) is an American rock musician, recording artist, songwriter, and multi-instrumentalist. Along with his work as a solo artist, ...

[Nils Lofgren \(album\) - Cry Tough \(Nils Lofgren album\) - Night After Night](#)

### Löfgren syndrome - Wikipedia, the free encyclopedia

[en.wikipedia.org/wiki/Löfgren\\_syndrome](http://en.wikipedia.org/wiki/Löfgren_syndrome)

Löfgren syndrome is a type of acute sarcoidosis that is frequent in Scandinavian, Irish, African and Puerto Rican women. It was characterized in 1953 by Sven ...

[Presentation - Prognosis - Genetics - Treatment Options](#)

### In the news



### Nils Lofgren ready to cross the Jersey state line

Asbury Park Press - 19 hours ago

Just how Jersey is Maryland-raised rocker Nils Lofgren? He met his wife at the Stone Pony ...

[E Street Band Guitarist Nils Lofgren Feeling "Excited" and "Refreshed" as He Launches New Solo Tour](#)

Ni

Musi

Nils L

Amer

recor

and r

Wikip

Born

Illinoi

Heig

Spot

Musi

(Sinc

Com

Silen

Sibli





Löfgren syndrome is a type of acute sarcoidosis that is frequent in Scandinavian, Irish, African and Puerto Rican women. It was characterized in 1953 by Sven ...  
Presentation - Prognosis - Genetics - Treatment Options

In the news



Nils Lofgren ready to cross the Jersey state line

Asbury Park Press - 19 hours ago

Just how Jersey is Maryland-raised rocker Nils Lofgren? He met his wife at the Stone Pony ...

E Street Band Guitarist Nils Lofgren Feeling "Excited" and "Refreshed" as He Launches New Solo Tour

WFJA Classic Hits and Oldies 105.5 FM - 4 hours ago

More news for lofgren

Nils Lofgren, Musician, Songwriter

www.nilslofgren.com/

Nils Lofgren is a virtuoso rock guitarist who has recorded and performed with Bruce Springsteen as a member of the E Street Band, Ringo Starr and Neil Young.

U.S. Congresswoman Zoe Lofgren

- Wikip
- Born
- Illinoi
- Heig
- Spot
- Musi
- (Sinc
- Com
- Silen
- Sibli
- Pro
- Tv
- Pec



### Nils Lofgren ready to cross the Jersey state line

Asbury Park Press - 19 hours ago

Just how Jersey is Maryland-raised rocker Nils Lofgren? He met his wife at the Stone Pony ...

E Street Band Guitarist Nils Lofgren Feeling "Excited" and "Refreshed" as He Launches New Solo Tour

WFJA Classic Hits and Oldies 105.5 FM - 4 hours ago

More news for lofgren

### Nils Lofgren, Musician, Songwriter

[www.nilslofgren.com/](http://www.nilslofgren.com/)

Nils Lofgren is a virtuoso rock guitarist who has recorded and performed with Bruce Springsteen as a member of the E Street Band, Ringo Starr and Neil Young.

### U.S. Congresswoman Zoe Lofgren

<https://lofgren.house.gov/>

Zoe Lofgren (D-Calif.), the top Democrat on the House Judiciary Subcommittee on Immigration and Border Security, lamented today's Judiciary subcommittee ...

### Löfgren's Syndrome | Doctor | Patient.co.uk

[www.patient.co.uk](http://www.patient.co.uk) > Professional Reference

Löfgren's syndrome is a subtype of acute sarcoidosis involving: [ 38484 · Ryan CW

Navigation sidebar with icons for: MUST (Sinc), Com Silen, Sibli, Pro, Tv, Pec, and a profile picture for Steve Zanc.





met his wife at the Stone Pony ...

E Street Band Guitarist Nils Lofgren Feeling "Excited" and "Refreshed" as He Launches New Solo Tour

WFJA Classic Hits and Oldies 105.5 FM - 4 hours ago

More news for lofgren

### Nils Lofgren, Musician, Songwriter

[www.nilslofgren.com/](http://www.nilslofgren.com/)

Nils Lofgren is a virtuoso rock guitarist who has recorded and performed with Bruce Springsteen as a member of the E Street Band, Ringo Starr and Neil Young.

### U.S. Congresswoman Zoe Lofgren

<https://lofgren.house.gov/>

Zoe Lofgren (D-Calif.), the top Democrat on the House Judiciary Subcommittee on Immigration and Border Security, lamented today's Judiciary subcommittee ...

### Löfgren's Syndrome | Doctor | Patient.co.uk

[www.patient.co.uk](http://www.patient.co.uk) > Professional Reference

Löfgren's syndrome is a subtype of acute sarcoidosis involving: [ 38484 : Byun CW, Yang SN, Yoon JS, et al ] Hilar lymphadenopathy Erythema nodosum...

### Nils Lofgren | Music Biography, Streaming Radio and ...

[www.allmusic.com/artist/nils-lofgren-mn0000414745](http://www.allmusic.com/artist/nils-lofgren-mn0000414745)



Pro



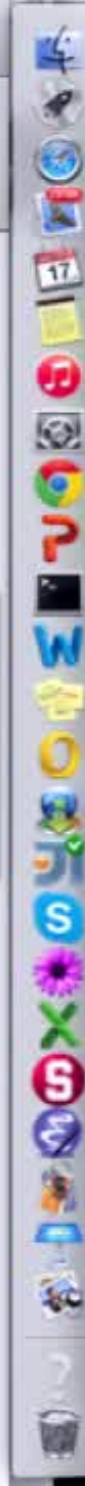
Pec



Steve Zand

See

Zoe l Born



More news for lofgren

Nils Lofgren, Musician, Songwriter

www.nilslofgren.com/

Nils Lofgren is a virtuoso rock guitarist who has recorded and performed with Bruce Springsteen as a member of the E Street Band, Ringo Starr and Neil Young.

U.S. Congresswoman Zoe Lofgren

https://lofgren.house.gov/

Zoe Lofgren (D-Calif.), the top Democrat on the House Judiciary Subcommittee on Immigration and Border Security, lamented today's Judiciary subcommittee ...

Löfgren's Syndrome | Doctor | Patient.co.uk

www.patient.co.uk > Professional Reference

Löfgren's syndrome is a subtype of acute sarcoidosis involving: [ 38484 : Byun CW, Yang SN, Yoon JS, et al ] Hilar lymphadenopathy Erythema nodosum...

Nils Lofgren | Music Biography, Streaming Radio and ...

www.allmusic.com/artist/nils-lofgren-mn0000414745

Find Nils Lofgren bio, music, credits, awards, & streaming radio on AllMusic - Guitarist/singer best known for his work with...

Nils Lofaren (@nilslofaren) | Twitter

Pro



Pec



Steve Zanc

See

Zoe l  
Born  
Offic







lofgren



+Asi

Web News Images Videos Maps More Search tools

About 838,000 results (0.30 seconds)

### Nils Lofgren - Wikipedia, the free encyclopedia

en.wikipedia.org/wiki/Nils\_Lofgren

Nils Hilmer Lofgren (born June 21, 1951) is an American rock musician, recording artist, songwriter, and multi-instrumentalist. Along with his work as a solo artist, ...

Nils Lofgren (album) - Cry Tough (Nils Lofgren album) - Night After Night

### Löfgren syndrome - Wikipedia, the free encyclopedia

en.wikipedia.org/wiki/Löfgren\_syndrome

Löfgren syndrome is a type of acute sarcoidosis that is frequent in Scandinavian, Irish, African and Puerto Rican women. It was characterized in 1953 by Sven ...

Presentation - Prognosis - Genetics - Treatment Options

### In the news



Nils Lofgren ready to cross the Jersey state

Ni

Musi

Nils L

Amer

recor

and r

Wikip

Born

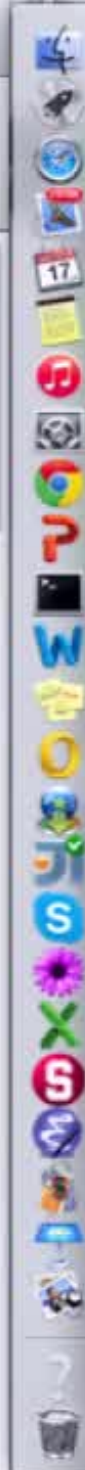
Illinoi

Heig

Spot

Musi

(Sinc





Search

714 results for mohamed zaki

Some search results have been filtered to improve relevance. Show all results

- All
- People
- Jobs
- Companies
- Groups
- Universities
- Posts
- Inbox

- Relationship
- All
  - 1st Connections (0)
  - 2nd Connections (0)
  - Group Members (0)
  - 3rd + Everyone Else (714)

 **mohd zaki hj mohamed zaki mohamed**  
 --  
 Malaysia  
 Similar [Connect](#)

 **Mohamed Zaki**  
 CEO at Super Jet Limo  
 San Francisco Bay Area · Retail  
 Similar [Connect](#)

 **Mohamed Abbas Zaki Abas Zaki Zaki**  
 --  
 Egypt  
 Similar [Connect](#)

 **Mohamed Zaki Zaki**  
 DR PROFESSOR at HORTICULTURE RES. INSTITUTE [Connect](#)



- Search**
- Advanced >
- All**
- People
- Jobs
- Companies
- Groups
- Universities
- Posts
- Inbox

- Relationship**
- All
- 1st Connections (0)
- 2nd Connections (0)
- Group Members (0)
- 3rd + Everyone Else (164)

164 results for mohammad zaki

Did you mean *mohamed zaki*?

Some search results have been filtered to improve relevance. [Show all results](#)

 **Mohammad Zaki**  
 Operations Manager at RZ Livescan Services  
 Greater Los Angeles Area • Airlines/Aviation  
[Similar](#) [Connect](#)

 **Mohammad Zaki** <sup>3rd</sup>  
 Senior Linguist and Interpreter  
 Alexandria, Louisiana Area • Translation and Localization  
[Similar](#) [Connect](#)

 **Mohammad Shah Zaki**  
 --  
 United States  
[Similar](#) [Connect](#)



- Search
- Advanced >
- All
- People
- Jobs
- Companies
- Groups
- Universities
- Posts
- Inbox

714 results for mohamed zaki

Some search results have been filtered to improve relevance. Show all results

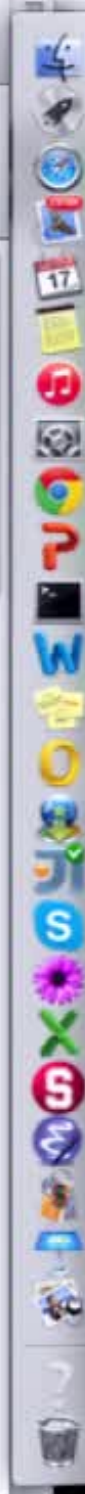
 **mohd zaki hj mohamed zaki mohamed**  
 --  
 Malaysia  
 Similar Connect

 **Mohamed Zaki**  
 CEO at Super Jet Limo  
 San Francisco Bay Area · Retail  
 Similar Connect

 **Mohamed Abbas Zaki Abas Zaki Zaki**  
 --  
 Egypt  
 Similar Connect

 **Mohamed Zaki Zaki**  
 DR PROFESSOR at HORTICULTURE RES. INSTITUTE Connect

- Relationship
- All
  - 1st Connections (0)
  - 2nd Connections (0)
  - Group Members (0)
  - 3rd + Everyone Else (714)





Search

19,798 results for zaki

Some search results have been filtered to improve relevance. Show all results

- All
- People
- Jobs
- Companies
- Groups
- Universities
- Posts
- Inbox

- Relationship
- All
  - 1st Connections (0)
  - 2nd Connections (36)
  - Group Members (0)
  - 3rd + Everyone Else (19762)



Zaki Hasan in 2nd

BU PM at Cisco Systems, WebEx San Francisco Bay Area · Computer Software 1 shared connection · Similar

Connect

Current: BU PM Offer Development CCTG Offers & Pricing at Cisco Syste... Past: Business Systems Analyst - Digital Supply Chain, Software Subscr... Solution Architect - Renewals at Cisco Systems Business Solution Manager - Regulatory Affairs at Genentech

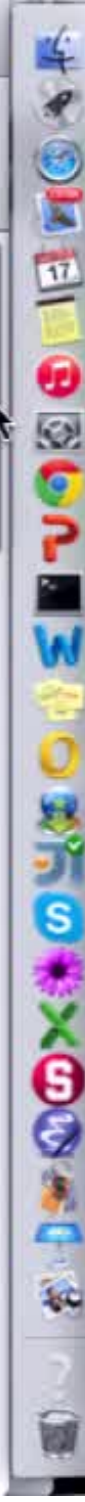


Zaki Fasihuddin in 2nd

Digital Executive, Advisor, Investor, Entrepreneur San Francisco Bay Area · Internet 1 shared connection · Similar

Connect

Current: Managing Director, SF Digital at McDonald's Corporation Co-founder & Advisor at Metrical Mentor at Momentum Accelerator Co-founder & Board Member at Real Liquidity



- 1st Connections (0)
- 2nd Connections (36)
- Group Members (0)
- 3rd + Everyone Else (19762)

Current: Managing Director, SF Digital at McDonald's Corporation  
 Co-founder & Advisor at Metrical  
 Mentor at Momentum Accelerator  
 Co-founder & Board Member at Real Liquidity

- Location**
- All
  - Egypt (4244)
  - Malaysia (1926)
  - United States (1738)
  - Saudi Arabia (1282)
  - India (1135)
  - + Add

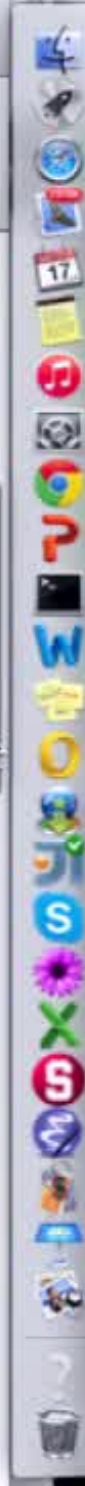
53 Companies results for zaki

-  **The Zaki Law Firm**  
Law Practice • Myself Only
-  **Zaki Design**  
Internet • 1-10 employees
-  **Raine & Horne Penang**  
Real Estate • 11-50 employees

 **Zaki Khanzada** 2nd  
 Highly Experienced Travel Professional @ Tri City Travel, Inc  
 San Francisco Bay Area • Airlines/Aviation **Connect**  
 ▶ 1 shared connection • Similar

 **Zaki Manian** 2nd  
 Founder at Skuchain  
 San Francisco Bay Area • Biotechnology **Connect**  
 ▶ 1 shared connection • Similar

- Current Company**
- All
  - Saudi Aramco (59)
  - PETRONAS (26)
  - Schlumberger (23)
  - IBM (22)





- Company Size
- Fortune
- When Joined
- Search like a Pro**  
Find key people in half the time with Premium Filters  
[Learn more](#)



**Omer Zaki** 2nd  
Engineering at Amazon Web Services  
San Francisco Bay Area • Internet  
2 shared connections • Similar

Connect

Current: Senior Manager, Engineering, Product Management at Amazon W...  
Past: Manager, Engineering, Product Management at Amazon Web Ser...  
Senior Product Manager at Amazon Web Services  
Senior Product Manager at Dell

- 1
- 2
- 3
- 4
- 5
- 6
- 7
- 8
- 9
- 10
- Next >

Ads



**Hair Transplantation!**  
Undetectable. No knife. No sutures. Less pain. Work next day. Call now.



**2016 ASCPT Annual Meet**  
Save the date and make yo plans to attend San Diego in 2016!



- All
- 1st Connections (0)
- 2nd Connections (36)
- Group Members (0)
- 3rd + Everyone Else (19762)

- Location**
- All
  - Egypt (4244)
  - Malaysia (1926)
  - United States (1738)
  - Saudi Arabia (1282)
  - India (1135)
  - + Add

- Current Company**
- All
  - Saudi Aramco (59)
  - PETRONAS (26)
  - Schlumberger (23)



**Zaki Fasihuddin** <sup>2nd</sup>  
 Digital Executive, Advisor, Investor, Entrepreneur  
 San Francisco Bay Area · Internet

Connect

1 shared connection · Similar



**Ali Diab** <sup>1st</sup>  
 CEO at Collective Health

Current: Managing Director, SF Digital at McDonald's Corporation  
 Co-founder & Advisor at Metrical  
 Mentor at Momentum Accelerator  
 Co-founder & Board Member at Real Liquidity



53 Companies results for zaki

**The Zaki Law Firm**  
 Law Practice · Myself Only

**Zaki Design**  
 Internet · 1-10 employees

**Raine & Horne Penang**  
 Real Estate · 11-50 employees



**Zaki Khanzada** <sup>2nd</sup>  
 Highly Experienced Travel Professional @ Tri City Travel, Inc  
 San Francisco Bay Area · Airlines/Aviation

Connect

- Location**
- All
  - Egypt (4244)
  - Malaysia (1926)
  - United States (1738)
  - Saudi Arabia (1282)
  - India (1135)
  -


- Current Company**
- All
  - Saudi Aramco (59)
  - PETRONAS (26)
  - Schlumberger (23)
  - IBM (22)
  - Hewlett-Packard (21)
  -

**Industry**

Current: Managing Director, SF Digital at McDonald's Corporation  
 Co-founder & Advisor at Metrical  
 Mentor at Momentum Accelerator  
 Co-founder & Board Member at Real Liquidity

### 53 Companies results for zaki

-  **The Zaki Law Firm**  
Law Practice • Myself Only
-  **Zaki Design**  
Internet • 1-10 employees
-  **Raine & Horne Penang**  
Real Estate • 11-50 employees

 **Zaki Khanzada** 2nd  
 Highly Experienced Travel Professional @ Tri City Travel, Inc  
 San Francisco Bay Area • Airlines/Aviation **Connect**  
 ▶ 1 shared connection • Similar

 **Zaki Manian** 2nd  
 Founder at Skuchain  
 San Francisco Bay Area • Biotechnology **Connect**  
 ▶ 1 shared connection • Similar





Search

283 results for mohammed zaki

Did you mean mohamed zaki?

Some search results have been filtered to improve relevance. Show all results

- All
- People
- Jobs
- Companies
- Groups
- Universities
- Posts
- Inbox

- Relationship
- All
  - 1st Connections (0)
  - 2nd Connections (2)
  - Group Members (0)
  - 3rd + Everyone Else (281)

**Mohammed Zaki** 2nd  
 Prof at RPI  
 Albany, New York Area · Higher Education  
 6 shared connections · Similar [Connect](#)

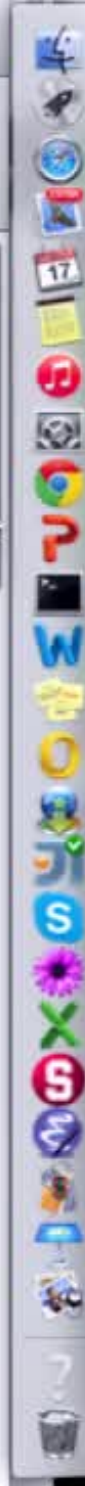
**Ravi Kumar** 1st  
 Senior Staff Research Scientist at Google

**Joydeep Ghosh** 1st  
 Professor at Univ of Texas at Austin

**Jure Leskovec** 1st  
 Assistant Professor at Stanford University

[View all 6 shared >](#)

**Mohammed Naseem Ul Haq Zaki** 2nd






- Advanced >
- All
- People**
- More...

- Relationship**
- All
  - 1st Connections (6)
  - 2nd Connections (0)
  - Group Members (0)
  - 3rd + Everyone Else (0)


- Location**
- < All
  - United States (6)
  - San Francisco Bay Area (3)
  - Greater Chicago Area (1)
  - Austin, Texas Area (1)
  - Greater New York City A... (1)
  - + Add

**Connections of Mohammed Zaki**

1st Connections x Reset

 **Jure Leskovec** 1st  
 Assistant Professor at Stanford University  
 San Francisco Bay Area • Research  
 ▶ 93 shared connections • Similar • 500+

[Message](#)

 **Ravi Kumar** 1st  
 Senior Staff Research Scientist at Google  
 San Francisco Bay Area • Computer Software  
 ▶ 123 shared connections • Similar • 500+

[Message](#)

 **Panos Ipeirotis** in 1st  
 Professor at New York University  
 Greater New York City Area • Computer Software  
 ▶ 44 shared connections • Similar • 500+

Current: Consultant at Citi  
 Consultant at The World Bank  
 Associate Professor at NYU Stern School of Business  
 Past: Visiting Scientist at Google

[Message](#)



- Group Members (0)
- 3rd + Everyone Else (0)

- Location**
- All
  - United States (6)
  - San Francisco Bay Area (3)
  - Greater Chicago Area (1)
  - Austin, Texas Area (1)
  - Greater New York City A... (1)
  - + Add**

- Current Company**
- All
  - Yahoo (1)
  - Google (1)
  - Stanford University (1)
  - University of Illinois at Ch... (1)
  - NYU Stern School of Bus... (1)
  - + Add**



**Ravi Kumar** 1st  
 Senior Staff Research Scientist at Google  
 San Francisco Bay Area • Computer Software  
 123 shared connections • Similar • 500+

Message



**Panos Ipeirotis** 1st  
 Professor at New York University  
 Greater New York City Area • Computer Software  
 44 shared connections • Similar • 500+

Current: Consultant at Citi  
 Consultant at The World Bank  
 Associate Professor at NYU Stern School of Business

Past: Visiting Scientist at Google

Message



**Tanya Berger-Wolf** 1st  
 Associate Professor at University of Illinois at Chicago  
 Greater Chicago Area • Research  
 21 shared connections • Similar • 500+

Message



**Joydeep Ghosh** 1st  
 Professor at Univ of Texas at Austin  
 Austin, Texas Area • Higher Education  
 17 shared connections • Similar • 500+

Message





# Personalized PageRank

Network-based Personalized Search is not yet mature

Missing technical piece: Efficient algorithms for Personalized PageRank Queries

→ Given source  $s$  and target  $t$ , estimate the Personalized PageRank of  $t$  for  $s$  with high accuracy, if it is greater than  $\delta$



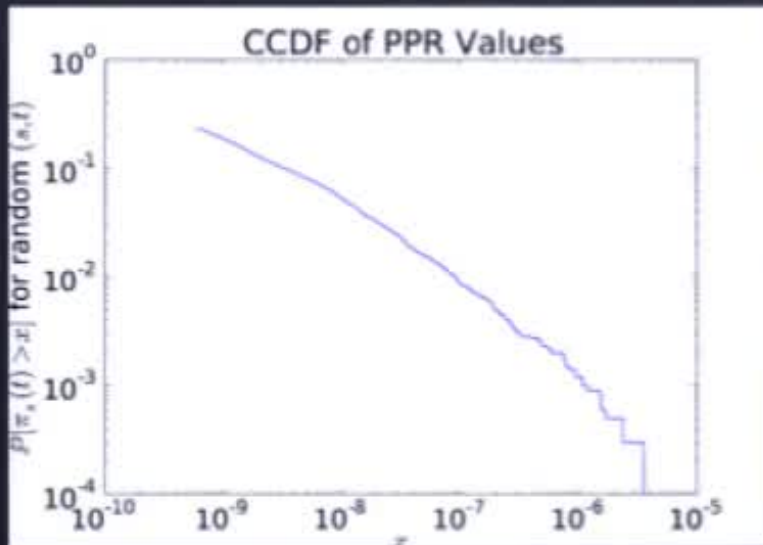
# Personalized PageRank

Given a **consumer C**, perform a random walk on the Follow graph. If the walk is at node  $v$ , then the walk:

- Jumps back to node  $C$  with probability  $\alpha$
- Follows a random edge out of  $v$  with probability  $1 - \alpha$

The Personalized PageRank of node  $Y$  is the weight of  $Y$  in the stationary distribution of this random walk

# Existing Methods for PPR Queries



Monte Carlo uses time  $> 1/\delta$

“Local Update” uses time  $d/\delta$

[ $d = M/N$  is the average degree]

On Twitter-2010, if  $\delta = \frac{4}{n} \approx 10^{-7}$ , then

$$\Pr[\pi(s, t) > \delta] = 1\%$$

# Personalized PageRank

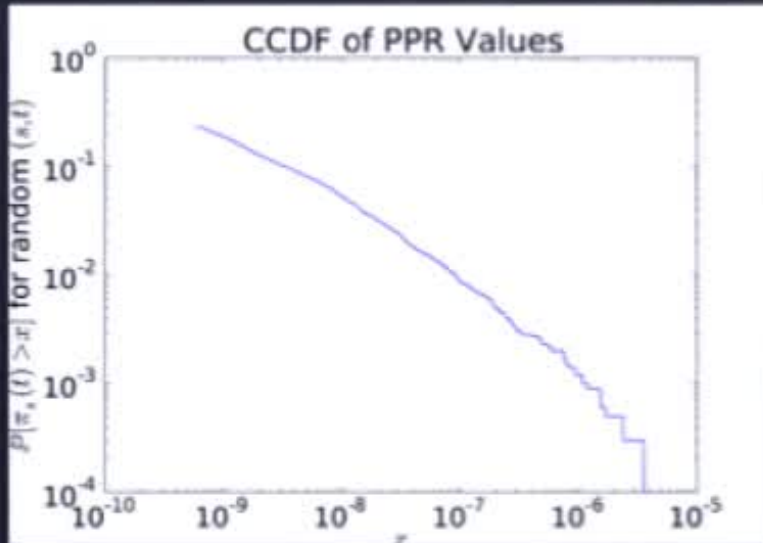
Given a **consumer C**, perform a random walk on the Follow graph. If the walk is at node  $v$ , then the walk:

- Jumps back to node C with probability  $\alpha$
- Follows a random edge out of  $v$  with probability  $1 - \alpha$

The Personalized PageRank of node Y is the weight of Y in the stationary distribution of this random walk



# Existing Methods for PPR Queries



Monte Carlo uses time  $> 1/\delta$

“Local Update” uses time  $d/\delta$

[ $d = M/N$  is the average degree]

On Twitter-2010, if  $\delta = \frac{4}{n} \approx 10^{-7}$ , then

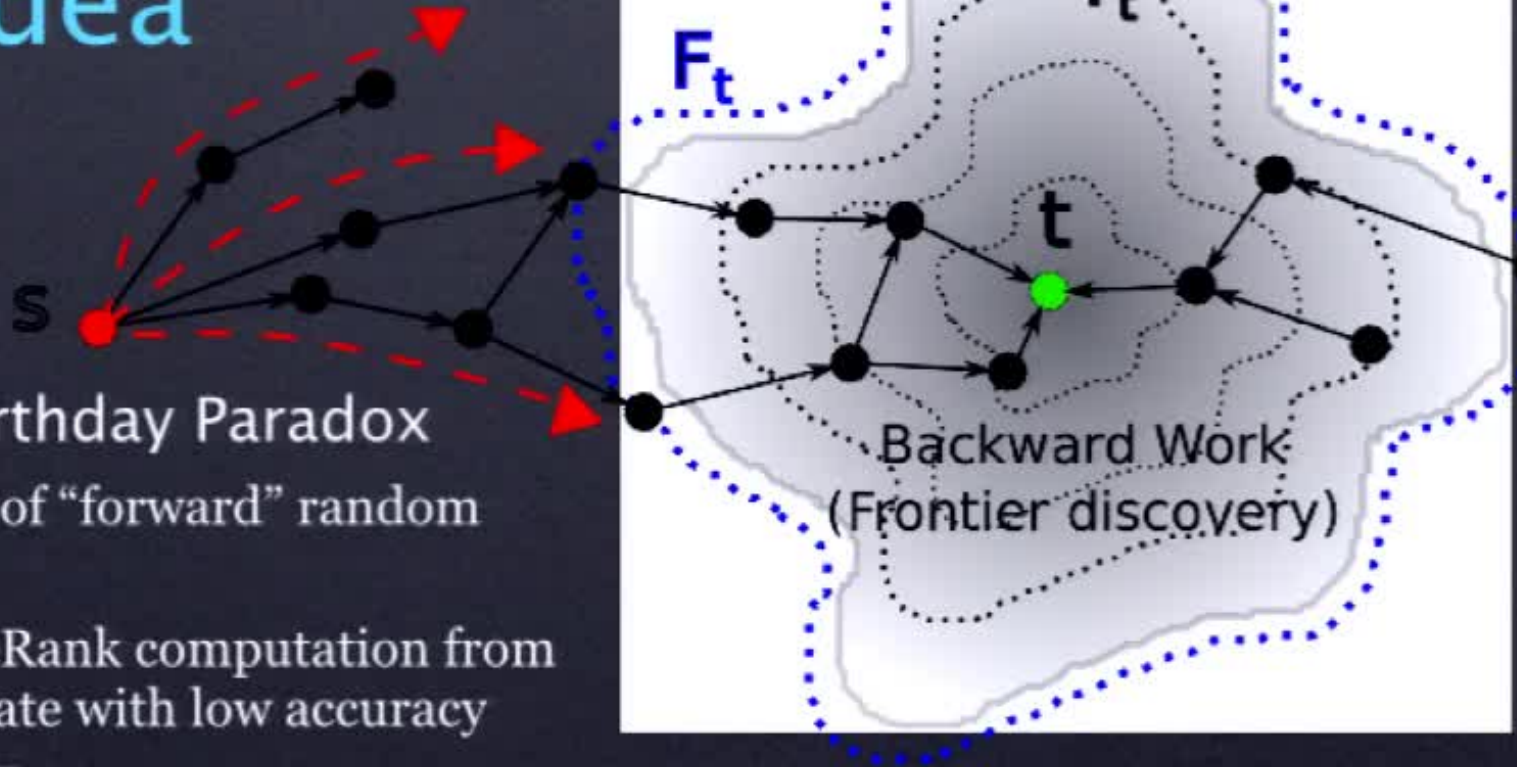
$$\Pr[\pi(s, t) > \delta] = 1\%$$

# FAST PPR

We can answer PPR queries in either

- Average time  $\tilde{O}(\sqrt{d/\delta})$
- Worst case time  $\tilde{O}(\sqrt{d/\delta})$  with  $\tilde{O}(\sqrt{d/\delta})$  storage and pre-processing time per node
- Typical values:  $\delta \sim 10^{-8}$ ,  $d \sim 100$ ; results in a  $> 100$ -fold decrease

# Basic Idea



## Intuition: The Birthday Paradox

- Do small number of “forward” random walks from  $s$
- Do “reverse” PageRank computation from  $t$  using Local Update with low accuracy
- Use number of collisions as an estimator
- Need to “catch” a collision just before it happens



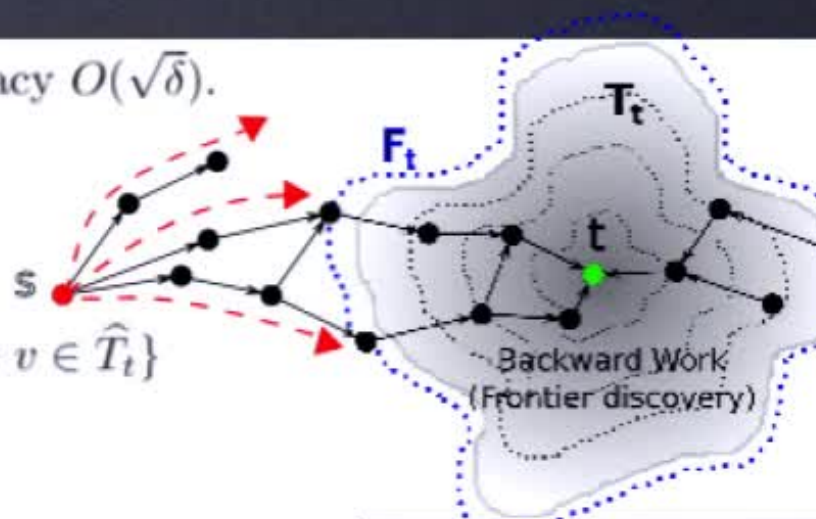
# Simple Version of FAST PPR

1. Use Local Update to compute estimates  $\hat{\pi}(v, t)$  to accuracy  $O(\sqrt{\delta})$ .

2. Define

$$\text{Target Set } \hat{T}_t = \{v \in V : \hat{\pi}(v, t) > \sqrt{\delta}\}$$

$$\text{Frontier } \hat{F}_t = \{u \in V \setminus \hat{T}_t : (u, v) \in E \text{ for some } v \in \hat{T}_t\}$$



3. Take  $O\left(\frac{\log(n)}{\sqrt{\delta}}\right)$  Random Walks  $\{W_i\}$ , terminating each early if it hits  $\hat{F}_t$ .

Define

$$X_i = \begin{cases} \hat{\pi}(u, t), & W_i \text{ hits } u \in \hat{F}_t \\ 0, & W_i \text{ does not hit } \hat{F}_t \end{cases}$$

4. Return empirical mean  $\{X_i\}$ .

# Running Time for Simple Version

For a uniformly random target node  $t$ , the average per-query running time is

$$O\left(\frac{1}{\sqrt{\delta}} (\bar{d} + \log(n))\right).$$

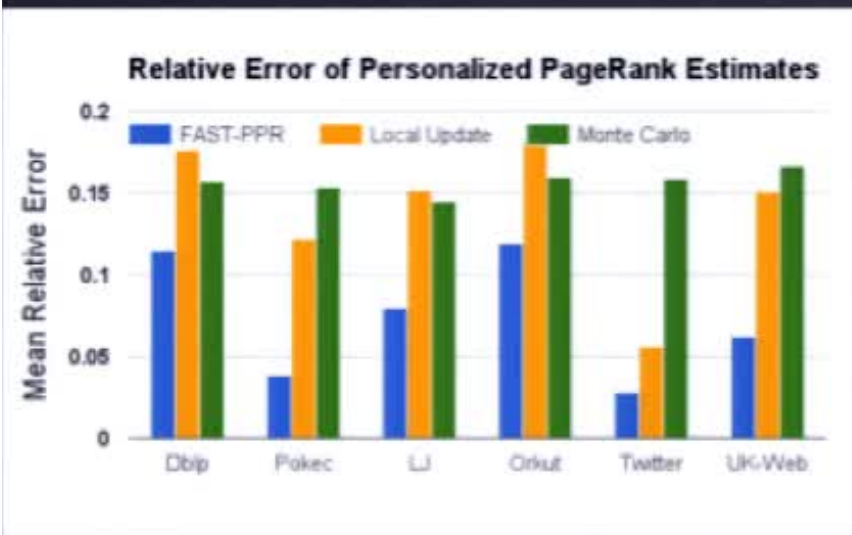
Reverse work  
(Local Update)

Forward work  
(Monte Carlo)

We get final running time of  $\tilde{O}(\sqrt{d/\delta})$  by using different accuracies in forward and reverse computation

We use  $\tilde{O}(\sqrt{d/\delta})$  pre-processing/space to go from average to worst case running time

# Experiments



- Admits Distributed Implementation (much faster)
- Works when source is a set of nodes
- Lower bound of  $1/\sqrt{\delta}$
- Open problem: do we need the  $\sqrt{d}$ ?

[Lofgren, Banerjee, Goel, Seshadhri, KDD, 2014]



# Computing via Intersections

Consider “sharding” a graph among  $K$  computers, i.e. the vertices of the graph get randomly partitioned into  $K$  sets,  $V_1, V_2, \dots, V_K$ , one for each compute node. Vertex  $v \rightarrow$  Shard  $s(v)$

Typical: For each vertex  $v$ , store all its edges in Shard  $s(v)$  as a key-value pair  $(v, \text{ADJ}(v))$

Additional trick: also store the reverse map, i.e. for each node  $w$ , store  $(w, \text{ADJ}(w) \cap V_j)$  on the  $j$ -th compute node

Advantage: Intersection of the neighborhood of  $u$  and  $v$  can be computed efficiently with one scatter-gather query

[Gupta, Satuluri et al., 2014]



#### Low Battery

Your Mac will sleep soon unless plugged into a power outlet.

Close

# Computing via Intersections

Both PPR and All Pairs Cosine Similarity can be efficiently reduced to graph intersections

Also, approximate shortest path reduces to graph intersections

[Lofgren, Goel, Gupta; manuscript]



#### Low Battery

Close

Your Mac will sleep soon unless plugged into a power outlet.

# Summary

Personalization algorithms can lead to both growth and monetization in social networks

Random walks and Cosine similarity are particularly efficient

Careful consideration of algorithms and architecture together can lead to efficiency

Next frontier: personalized search



Associate Professor at NYU Stern School of Business  
Past: Visiting Scientist at Google

- Current Company**
  - All
  - Yahoo (1)
  - Google (1)
  - Stanford University (1)
  - University of Illinois at Ch... (1)
  - NYU Stern School of Bus... (1)
  - + Add



**Tanya Berger-Wolf** 1st  
Associate Professor at University of Illinois at Chicago  
Greater Chicago Area • Research  
▶ 21 shared connections • Similar • 500+

[Message](#)



**Joydeep Ghosh** 1st  
Professor at Univ of Texas at Austin  
Austin, Texas Area • Higher Education  
▶ 17 shared connections • Similar • 500+

[Message](#)



**Ricardo Baeza-Yates** 1st  
VP of Research at Yahoo!  
San Francisco Bay Area • Internet  
▶ 54 shared connections • Similar • 500+

**Posts by Ricardo Baeza-Yates**  
[IEEE Computer special issue on Big Data Management \(Mar 2015\)](#)  
 April 6, 2015 • 532 Views

- Industry**
- Past Company**
- School**
- Profile Language**
- Nonprofit Interests**

Ads



Everything

People

Photos

Videos

News

Advanced Search

All people

People you follow

Everywhere

Near you

Who to follow · Refresh · View all

Andreas Krause @arkrause

Results for #SIAMSDM15

Top

3 new



Liangyue Li @Liangyue\_Li · Apr 29 flying to Vancouver tomorrow #SIAMSDM15

Interaction icons (reply, retweet, like, share)



IARPA @IARPAnews · Apr 20 Can we predict rare disease outbreaks? (#SIAMSDM15 best paper winner): bioRxiv preprint doi: https://doi.org/10.1101/011111; this version posted April 20, 2015. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Interaction icons (reply, retweet, like, share)



LINQS Group @LINQS\_edu · Apr 17 Congrats @thodrek et al. for best #SIAMSDM15 paper (Disease Outbreaks Using Multiple Data Sources). linqs.cs.umd.edu/basilic/web/Pu...

Interaction icons (reply, retweet, like, share)





- Everything
- People
- Photos
- Videos
- News
- Advanced Search

- All people
- People you follow

- Everywhere
- Near you

Who to follow · Refresh · View all

# Results for #SIAMSDM15

Top / All

3 new results



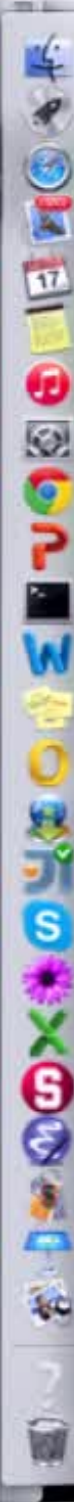
**Liangyue Li** @Liangyue\_Li · Apr 29  
flying to Vancouver tomorrow #SIAMSDM15



**IARPA** @IARPAnews · Apr 20  
Can we predict rare disease outbreaks? See the OSI progri  
#SIAMSDM15 best paper winner): bit.ly/1O87cWf #BigDa



**LINQS Group** @LINQS\_edu · Apr 17  
Congrats @thodrek et al. for best #SIAMSDM15 paper on  
Disease Outbreaks Using Multiple Data Sources  
linqs.cs.umd.edu/basilic/web/Pu...





- Videos
- News
- Advanced Search

- ✓ All people
- People you follow

- ✓ Everywhere
- Near you

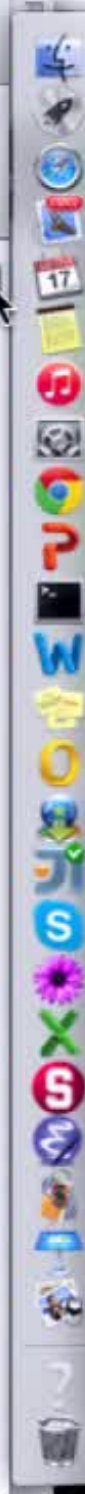
Who to follow · Refresh · View all

- 
**eytan bakshy** @eytan
   
 Follow
- 
**Dean Eckles** @deaneckles
   
 Follow

Photos · View all



**มหกรรมคืนความสวยงาม เมืองโบราณ**  
 (๕) พฤษภาคม 2558 เป็นต้นไป  
**ฟรี** ตกลงเล่นมหากุศลชบก่อนเสร็จ  
**พิเศษบัตรเที่ยวเมืองโบราณ 10**  
 ...หมายเหตุ คำสรรณียบบริการฉบับ 5!  
 สอบถามเพิ่มเติมได้ที่ โทร. 02 323 4094



**Drew Conway** @drewconway

Follow

Find friends

Trends · Change

**#GalaxyTabA**  
Get a #GalaxyTabA by switching to a TD chequing acct: bit.ly/110iZ3Z  
Promoted by TD (Canada)

**#DraftDay**  
Your Sports Talk Florida 2015 NFL Draft Day Primer  
18.8K Tweets about this trend

**#clickbaitbooks**  
Gareth L Powell on Twitter  
Just started trending

**#5WordFilmPlots**  
12.4K Tweets about this trend

**Florian Bellanger** @FlorianBelanger · 7h

@Siam\_lady - then they should speak when necessary - #

Reply Retweet Favorite More

**syu** @reosyu\_BASS · 7h

#Nowplaying LOVESICK ~You Don't Know~ - SIAM SHA



Vertical sidebar containing various application icons such as calendar, mail, and social media apps.



#BRVS

Just started trending

Iraq

Stewart Lays Into Judith Miller: You Pushed Us Into 'Devastating'...

32.8K Tweets about this trend

Gregory Mertens

Lokeren defender Gregory Mertens passes away

31.2K Tweets about this trend

Carson

4km ENE of Carson, California

15.2K Tweets about this trend



Retweet Like Reply More

People · View all



Ed Sheerul @shahrul\_fikri97 · 7h  
Diam siam

Retweet Like Reply More



LUNA SEA\_SG\_team @LS\_SGteam · 7h  
Artiste line up for LUNATIC FEST announced!

[Alexandros] KA F KA 9mm Parabellum Bullet GLA



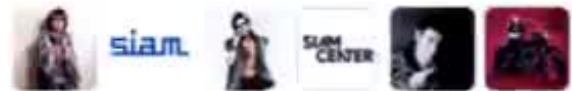
- to use iPads  
272K Tweets about this trend
- #BRVS  
Just started trending
- Iraq  
Stewart Lays Into Judith Miller: You Pushed Us Into 'Devastating'...  
32.8K Tweets about this trend
- Gregory Mertens  
Lokeren defender Gregory Mertens passes away  
31.2K Tweets about this trend
- Carson  
4km ENE of Carson, California  
15.2K Tweets about this trend

 Daily Deals : [ift.tt/1GqQ58G](https://ift.tt/1GqQ58G) #2910 Vintage Sterling Silver Siam Brooch ~ 12.4grams ~...



Reply Retweet Favorite Share More

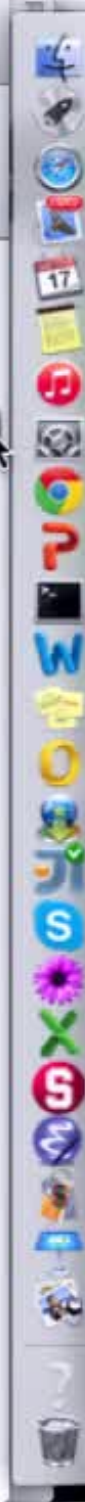
People · [View all](#)



 **Ed Sheerul** @shahrul\_fikri97 · 7h  
Diam siam

Reply Retweet Favorite Share More

 **LUNA SEA\_SG\_team** @LS\_SGteam · 7h  
Artiste line up for LUNATIC FEST announced!



- Everything
- People
- Photos
- Videos
- News
- Advanced Search

- All people
- People you follow

- Everywhere
- Near you

Who to follow · Refresh · View all

### Results for #SIAMSDM15

Top / All

3 new results



**Liangyue Li** @Liangyue\_Li · Apr 29  
flying to Vancouver tomorrow #SIAMSDM15

[Reply] [Retweet] [Like] [Share] [More]



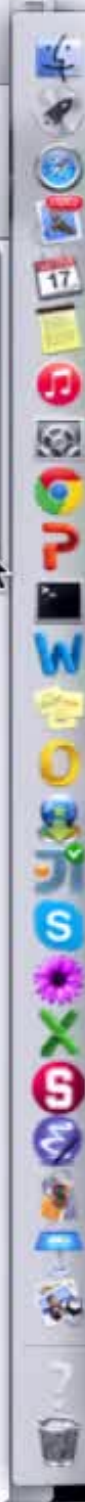
**IARPA** @IARPAnews · Apr 20  
Can we predict rare disease outbreaks? See the OSI prog...  
#SIAMSDM15 best paper winner): bit.ly/1O87cWf #BigDa

[Reply] [Retweet] 3 [Like] 2 [Share] [More]



**LINQS Group** @LINQS\_edu · Apr 17  
Congrats @thodrek et al. for best #SIAMSDM15 paper on  
Disease Outbreaks Using Multiple Data Sources  
linqs.cs.umd.edu/basilic/web/Pu...

[Reply] [Retweet] 2 [Like] 3 [Share] [More]





- Videos
- News
- Advanced Search

- All people
- People you follow


- Everywhere
- Near you

Who to follow · Refresh · View all

- 
**Andreas Krause** @arkrause  
 Followed by Aneesh Sharma...
- 
**Michael Dinitz** @mdinitz  
 Followed by Griqory Yarosla...

 **Liangyue Li** @Liangyue\_Li · Apr 29  
 flying to Vancouver tomorrow #SIAMSDM15

← ↻ ☆ 👤 ⋮

 **IARPA** @IARPAnews · Apr 20  
 Can we predict rare disease outbreaks? See the OSI prog...  
 #SIAMSDM15 best paper winner): bit.ly/1O87cWf #BigDa

← ↻ 3 ☆ 2 👤 ⋮

 **LINQS Group** @LINQS\_edu · Apr 17  
 Congrats @thodrek et al. for best #SIAMSDM15 paper on  
 Disease Outbreaks Using Multiple Data Sources  
[linqs.cs.umd.edu/basilic/web/Pu...](http://linqs.cs.umd.edu/basilic/web/Pu...)

← ↻ 2 ☆ 3 👤 ⋮

 **Amol Deshpande** @amolumd · Apr 17  
 Theo @thodrek, Lise Getoor, and co-authors receive Best  
 2015 for [linqs.cs.umd.edu/basilic/web/Pu....](http://linqs.cs.umd.edu/basilic/web/Pu....) Congratulatic

← ↻ 2 ☆ 2 👤 ⋮

