# On the Precision of Social and Information Networks


Kamesh Munagala (Duke)

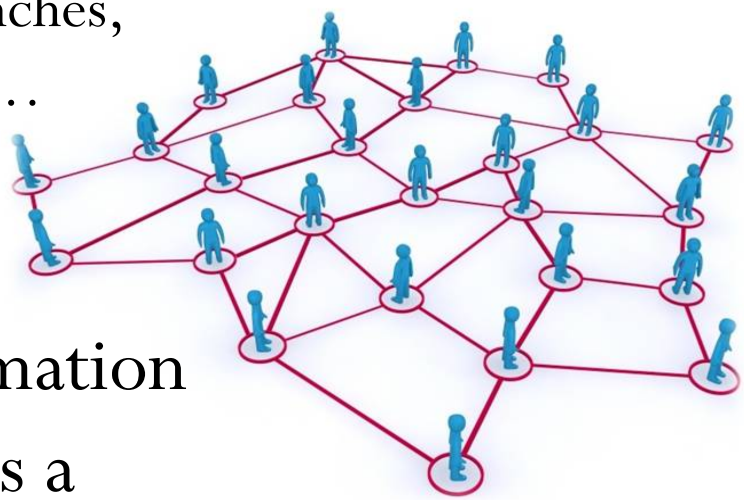
Reza Bosagh Zadeh (Stanford)

Ashish Goel (Stanford)

Aneesh Sharma (Twitter, Inc.)

# Information Networks

- Social Networks play an important role in information dissemination
  - Emergency events, product launches, sports updates, celebrity news,…



- Their effectiveness as information dissemination mechanisms is a source of their popularity

# A Fundamental Tension

Two conflicting characteristics in social networks:

- **Diversity:** Users are interested in diverse content
- **Broadcast:** Users disseminate information via posts/ tweets – these are blunt broadcast mechanisms!

# Running Example

Bob tweets about:
- Christianity
- DC Politics
- Bulls

Charlie tweets about:
- Jay-Z
- Lady Gaga
- Kobe

Adam interested in
- Apple
- Rap music
- Lakers

# Running Example

Bob tweets about:
- Christianity
- DC Politics
- Bulls

Charlie tweets about:
- Jay-Z
- Lady Gaga
- Kobe

**Follow**

**Follow**

Adam interested in
- Apple
- Rap music
- Lakers

# A Fundamental Tension

Two conflicting characteristics in social networks:

- **Diversity:** Users are interested in diverse content
- **Broadcast:** Users disseminate information via posts/tweets – these are blunt broadcast mechanisms!

**Precision:** Do users receive a lot of un-interesting content?

**Recall:** Do users miss a lot of interesting content?

# Question we study

**Can information networks have high precision and recall?**

# Case Study: Twitter

- A random tweet is uninteresting to a random user…
- But users have interests and follow others based on these

**Information networks like Twitter are constructed according to users' interests**

# Revisiting our example…

Bob tweets about:

- Christianity
- DC Politics
- Bulls

Charlie tweets about:

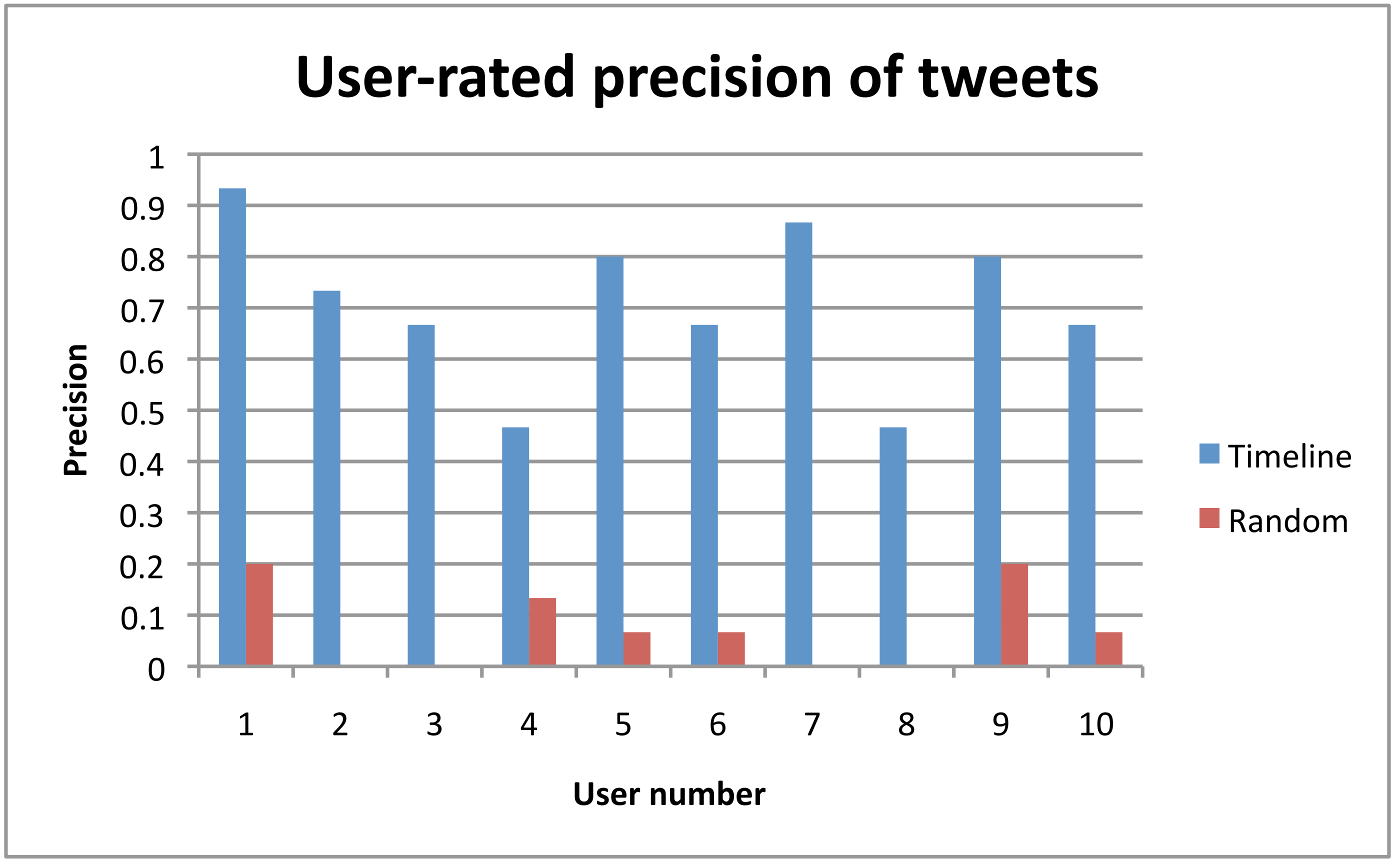
- Jay-Z
- Lady Gaga
- Kobe

**Follow**

Adam interested in

- Apple
- Rap music
- Lakers

# Small User Study on Twitter



# Roadmap

- User Behavior Assumption:
  1. Users have immutable interests (independent of the network)
  2. Choose to connect to other users based on their interests
  3. Step (2) is optimized for precision and recall

# Roadmap

- User Behavior Assumption:
  1. Users have immutable interests (independent of the network)
  2. Choose to connect to other users based on their interests
  3. Step (2) is optimized for precision and recall

- **Question 1:** What conditions on the structure of user interests are necessary for high precision and recall, and small dissemination time?

- **Question 2:** Can we empirically validate these conditions as well as the conclusion on Twitter?

# User-Interest Model

- Set of interests I; Set of users U

- Each interest *i* is associated with two sets of users:
  - **Producers** $P(i)$ = Users who tweet about *i*
  - **Consumers** $C(i)$ = Users who are interested in *i*

- Denote the mapping from users to interests as $Q(I, U)$

- Assume: $P(i) \subseteq C(i)$ for all interests *i*

# Example

User b

$P(b) = \{s, t\}$
$C(b) = \{r, s, t\}$

User c

$\mathbf{P(c) = \{q, t\}}$
$C(c) = \{q, s, t\}$

User a

$P(a) = \{q\}$
$\mathbf{C(a) = \{q, r, s\}}$

# Social (user-user) Graph *G(U,E)*

User b

$P(b) = \{s, t\}$
$C(b) = \{r, s, t\}$

User c

**$P(c) = \{q, t\}$**
$C(c) = \{q, s, t\}$

**User a receives interests**
**$R(a) = \{q, t\}$**

**Social graph**

User a

$P(a) = \{q\}$
**$C(a) = \{q, r, s\}$**

# Precision and Recall

Functions of user-interest map $Q(I,U)$ and social graph $G(U,E)$

$$\text{Precision}(u) = \frac{|R(u) \cap C(u)|}{|R(u)|}$$

Consumption interests

Interests received from followees

$$\text{Recall}(u) = \frac{|R(u) \cap C(u)|}{|C(u)|}$$

# PR Score

$$PR(u) = \frac{|R(u) \cap C(u)|}{|R(u) \cup C(u)|}$$

$$\approx \mathrm{Min}\left(\mathrm{Precision}(u), \mathrm{Recall}(u)\right)$$

# Example Revisited

User b

P(b) = {s, t}
C(b) = {r, s, t}

User c

**P(c) = {q, t}**
C(c) = {q, s, t}

**Social graph**

**R(a) = {q, t}**
**C(a) = {q, r, s}**

**PR(a) = ¼ = 0.25**

User a

P(a) = {q}
**C(a) = {q, r, s}**

# Improved Score

User b

**P(b) = {s, t}**
C(b) = {r, s, t}

User c

**P(c) = {q, t}**
C(c) = {q, s, t}

**Social graph**

R(a) = {q, s, t}
C(a) = {q, r, s}

PR(a) = 2/4 = 0.5

User a

P(a) = {q}
**C(a) = {q, r, s}**

# $\alpha$-PR User-Interest Maps *Q(I,U)*

**A user-interest map Q(I,U) is $\alpha$-PR if:**
**There exists a social graph G(U,E) s.t.**
**all users u have PR-Score $\geq \alpha$**

Special case: 1-PR means that
R(u) = C(u) for all users u

# Necessary Conditions for 1-PR

- **Condition 1:**

  If $Q(I, U)$ *is* "non-trivial" and $G(U,E)$ is (strongly) connected:

  ***Then*** $\boldsymbol{P(i) \subset C(i)}$ **for some interest** $\boldsymbol{i}$

- **Informal implication:**

  Users have broader consumption interests and narrower production interests

# Experimental Setup

- Classify text of tweets using 48 topics
  - Yields "topic distribution" for each user
  - Entropy of distribution lies between 0 and $\log_2(48) = 3.87$

- $P(u)$= Interest distribution in tweets produced by $u$
- $C(u)$ = Interest distribution in URL clicks made by $u$

# Verifying Condition 1

| TYPE OF INTEREST DISTRIBUTION | AVERAGE SUPPORT | AVERAGE ENTROPY |
|---|---|---|
| Consumption Interests | 7.78 | 2.00 |
| Production Interests | 3.96 | 1.24 |

# Can Interests be chosen at Random?

*Different interests can have different "participation levels"*

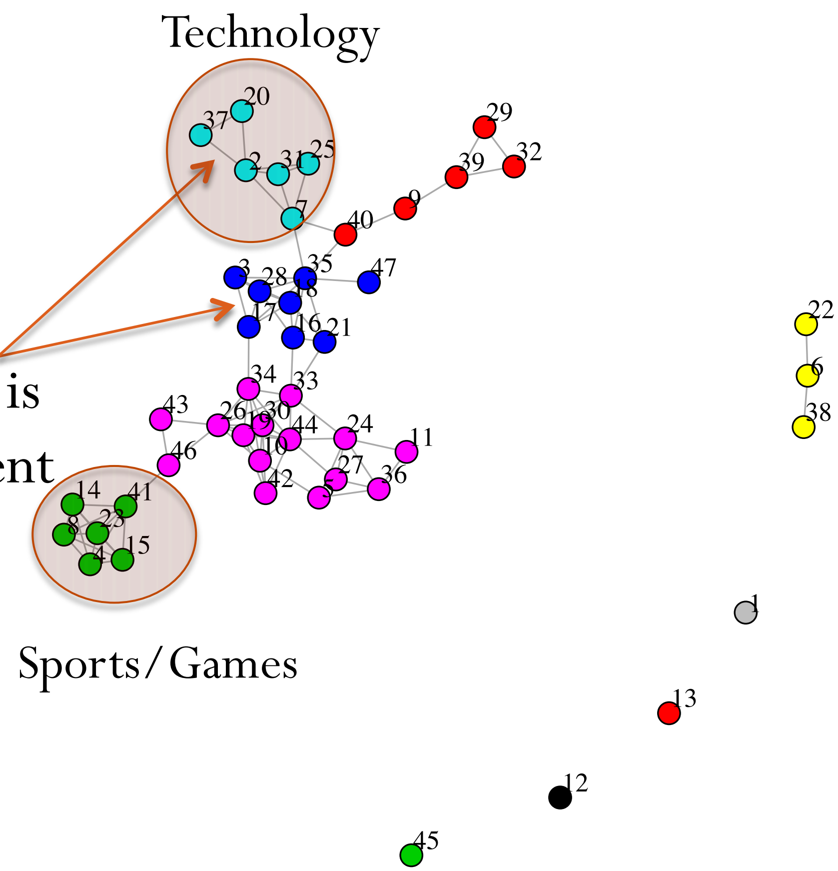
**Theorem:** If users choose production and consumption interests **at random** preserving participation levels of the interests, then (under minor assumptions):

**With high probability the interest structure is not $\alpha$-PR for any constant $\alpha$**

**Key proof idea:** $Q(I, U)$ behaves like an expander graph

# Condition 2: Interests have Structure

**Edge** = Share more users than is predicted by a random assortment



# Interest Structure achieving 1-PR

**Kronecker product model**

Attributes/Dimensions

| | Kobe | Gaga | Lakers | Obama |
|---|---|---|---|---|
| **User u** | Y | N | M | N |

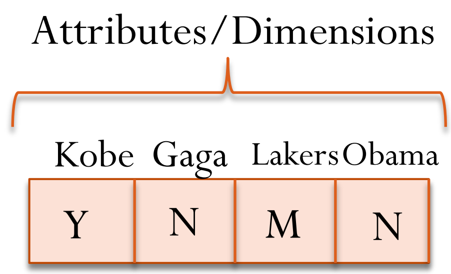
d = O(log n) dimensions

K = O(log n) values

# Interest Structure achieving 1-PR

## Kronecker product model

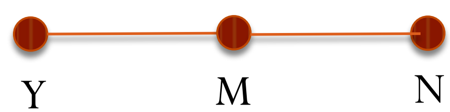
**User u**




d = O(log n) dimensions

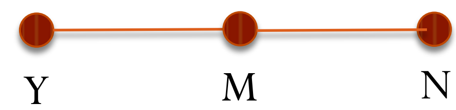
K = O(log n) values

Similarity graph on values



# Interest Structure

Similarity graph on values




Attributes/Dimensions

| | Kobe | Gaga | Lakers | Obama |
|---|---|---|---|---|
| **Interest i** | Y | | M | |
| **Producer** | Y | * | M | * |
| **Consumer** | M | * | N | * |
| **Not interested** | N | * | M | * |

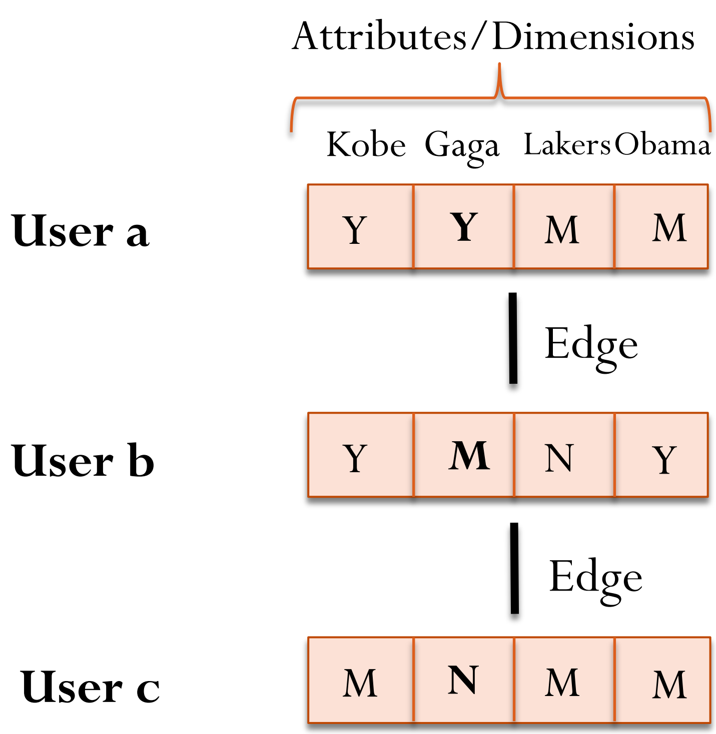
Set of relevant dimensions & their values

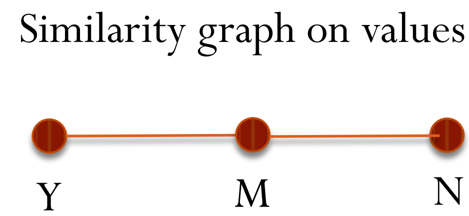
*Agrees exactly* on all relevant dimensions

*Similar* on all relevant dimensions

# User-user Graph

[Leskovec, Chakrabarti, Kleinberg, Faloutsos, Ghahramani '10]




**Undirected Edge between two users iff ALL dimensions are similar**

Such graphs can have:

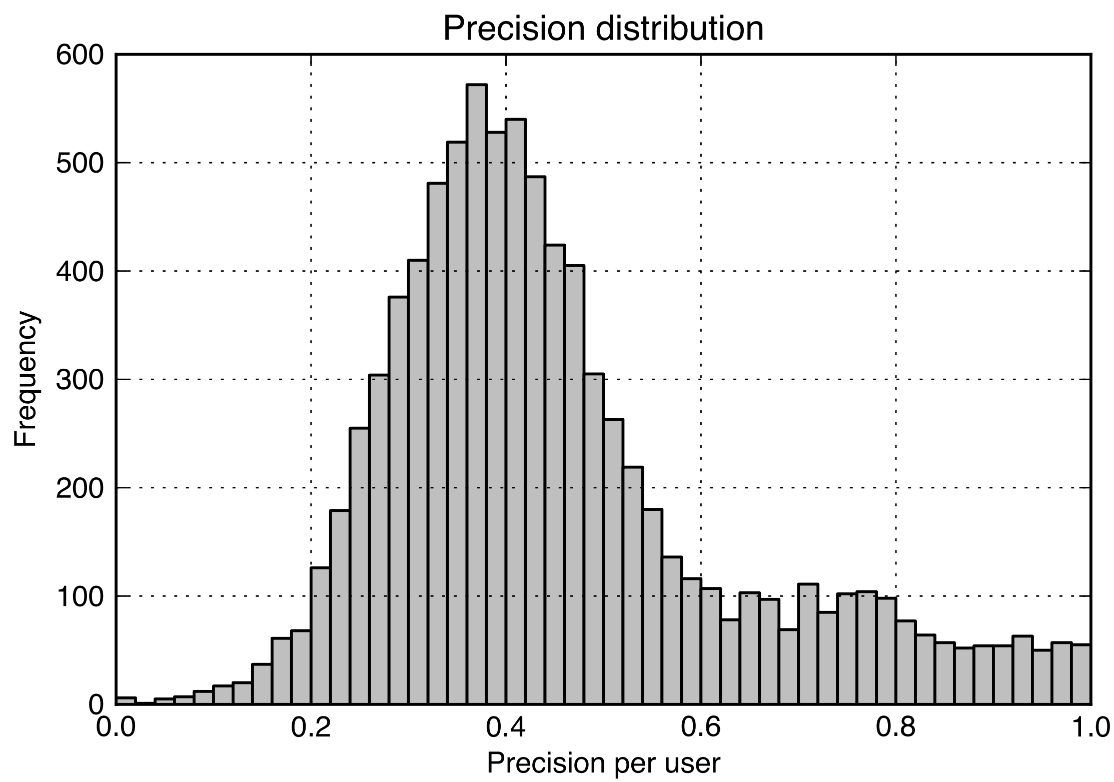
- Super-constant average degree
- Heavy tailed degree distributions
- Constant diameter

# Main Positive Result

- **The Kronecker interest structure has 100% PR!**
- Users only receive interesting information
- Users receive all information they are interested in
- The dissemination time is constant.

# Empirical Study of Precision

$$\text{Precision}(u) = \frac{\sum_{(u,v)\in E} |C(u) \cap P(v)|}{\sum_{(u,v)\in E} |P(v)|}$$




**Median precision = 40%**
**Baseline precision = 17%**

**Interpretation:** *One in 2.5 interests received on any follow edge are interesting*

# Caveat: This is only a first step!

- Measuring interests
  - Used URL clicks as a measure of consumption/relevance
  - Used 48 topics as proxy for interests
  - Not considered quality of tweets in measuring interest
  - Not explored structure of interests in great detail

- Empirical validation
  - User studies are more reliable, but our study is small
  - We did not measure recall or dissemination time

# Open Questions

- Better empirical measures of interests and PR?
  - In-depth analysis of structure of interests
  - How can recall be measured?

- Can high PR information networks arise in a decentralized fashion?
  - How can users discover high PR links?

# Thank You!