# User Modeling

You are what you post!

Represent users by documents a user = a document including all she said

### You are what you post!

- Modeling User Personality (Computationa) (Social Science)
  - Personality traits in psychology
  - Big Five: extraversion, emotional stability, agreeableness, conscientiousness, and openness to experience
- Modeling User Health Profile (Computational Epidemiology)
  - Privacy of the user, Ethical principles
- Modeling Gender and Ethnicity
  - First names → gender; Last names → ethnicity
- Modeling User Location

#### **Predicting Personal Life Events from Streaming Social Content**

Maryam Khodabakhsh Ferdowsi University of Mashhad maryamkhodabakhsh@stu.mail.ac.ir

> Fattane Zarrinkalam Ryerson University fzarrinkalam@ryerson.ca

> > (a) sliding:  $w_s=2$

 $S_{\mathbf{n}} = [|\mathbf{c}|\mathbf{c}|\mathbf{w}|\mathbf{m}|\mathbf{m}|\mathbf{t}|...]$ 

c wm

w m m

mm t

c c w

Hossein Fani University of New Brunswick hfani@unb.ca

> Ebrahim Bagheri Ryerson University bagheri@ryerson.ca

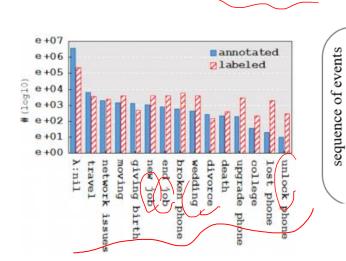
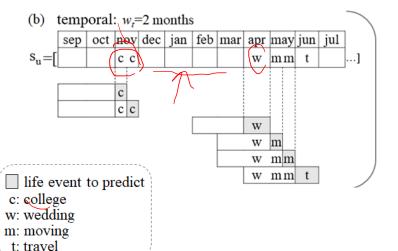


Figure 1: Distribution of personal life events by event class.



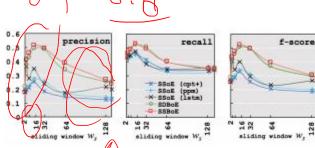


Figure 3: Comparative results of the sliding strategy.

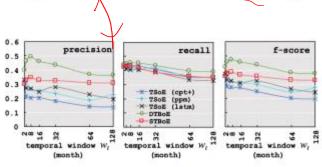
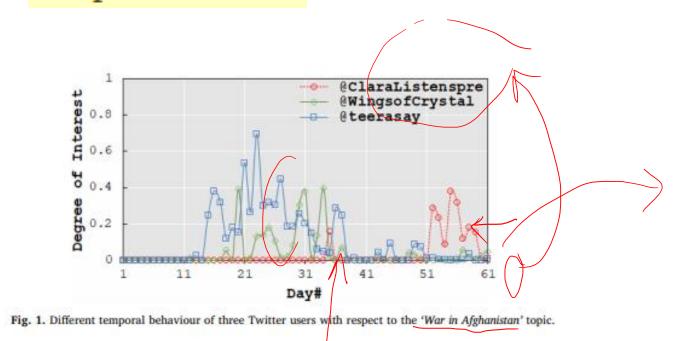


Figure 4: Comparative results of the temporal strategy.

User community detection via embedding of social network structure and temporal content\*



All users are interested in  $z_{44}$ ='War in Afghanistan'

User community detection via embedding of social network structure and temporal content\*

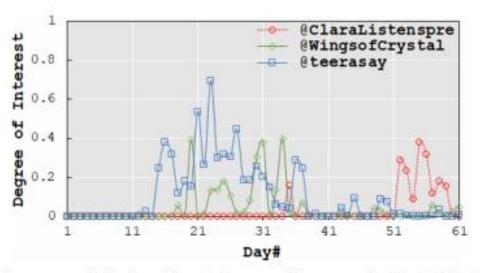
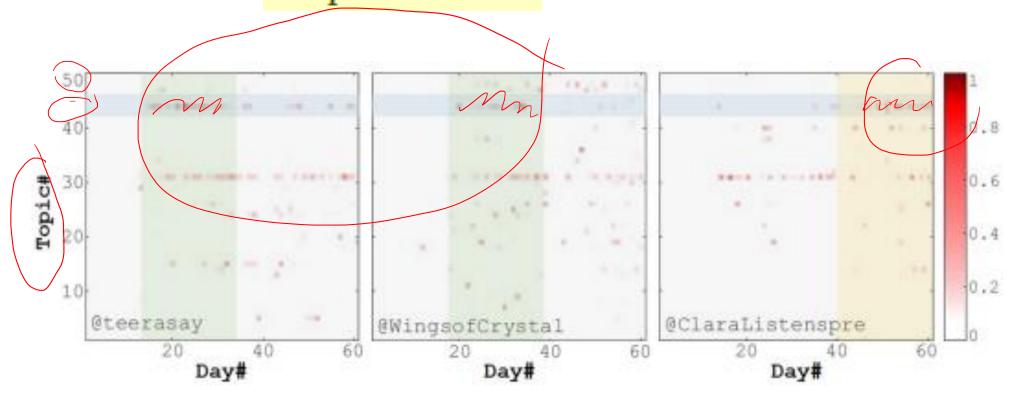


Fig. 1. Different temporal behaviour of three Twitter users with respect to the 'War in Afghanistan' topic.

All users are interested in z<sub>44</sub>='War in Afghanistan' but not aligned in time!

User community detection via embedding of social network structure and temporal content\*

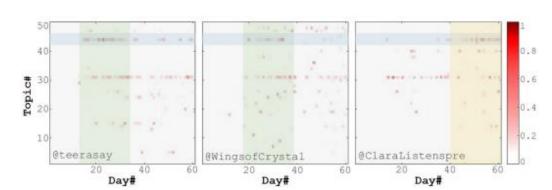


All users are interested in  $z_{44}$ ='War in Afghanistan' but not aligned in time!

- User Clustering
  - Timeseries (Image) Clustering

User ↔ Documents → User Vector ↔ Document Vector

- How to include time?

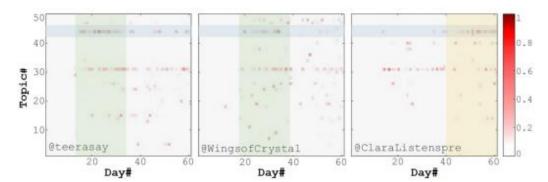


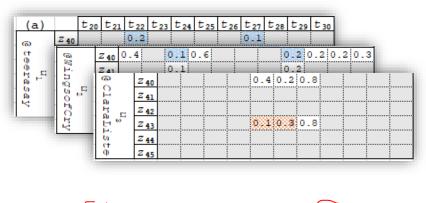
- User Clustering
  - User Vector Representation

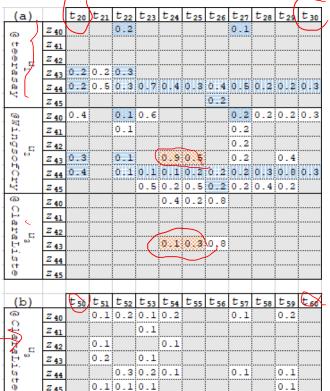
User ↔ Documents → User Vector ↔ Document Vector

- How to include time?

User at time  $t \leftrightarrow A$  document that has all she said at time t User = [Doc<sub>0</sub> Doc<sub>1</sub>, ..., Doc<sub>T</sub>]



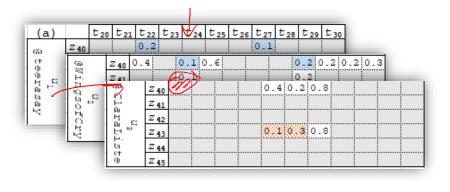


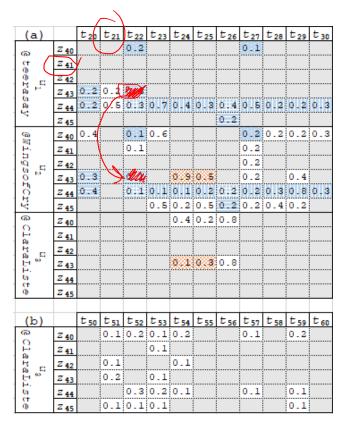


User = 
$$[Doc_0, Doc_1, ..., Doc_T]$$

LDA

User =  $[[z^{(0)}_{1:K}], [z^{(1)}_{1:K}], ..., [z^{(T)}_{1:K}]]$ 





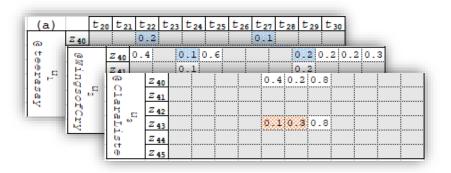
User = 
$$[Doc_0, Doc_1, ..., Doc_T]$$

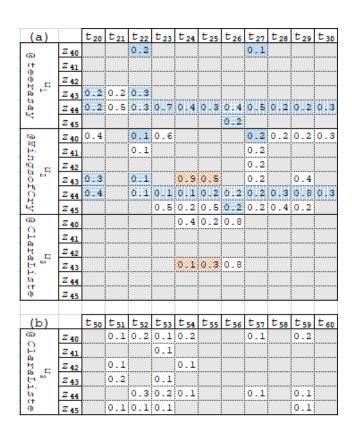
LDA

User =  $[[z^{(0)}_{1:K}], [z^{(1)}_{1:K}], ..., [z^{(T)}_{1:K}]]$ 

Two users are similar if they share more cells! each cell =  $1 \times 1 \times 1$  cube =  $\{u_i\} \times \{z_j\} \times \{t_k\}$ Shared cell =  $n \times m \times k$  cube

e.g., 
$$\{u_1u_2\} \times \{z_{44}\} \times \{t_{22} \ t_{23} \ ... \ t_{30}\}$$



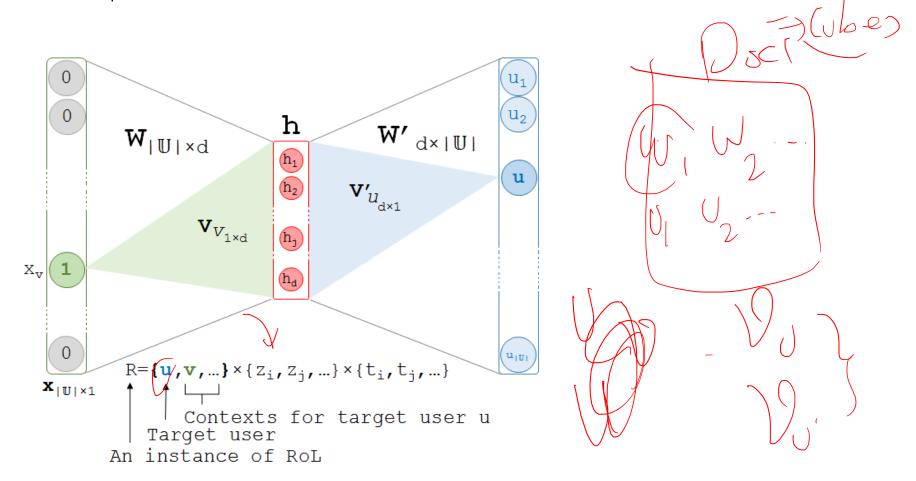


Region of Like-mindedness (RoL) iff  $y_t^u[z] \approx y_t^v[z]$ 

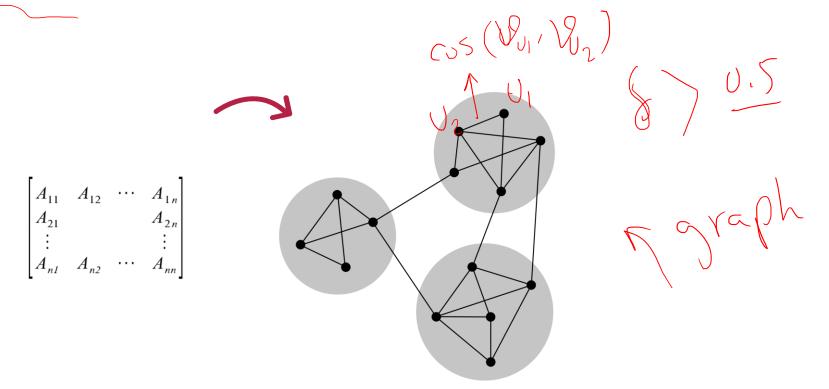
Two users are similar if they share more cells! each cell =  $1 \times 1 \times 1$  cube =  $\{u_i\} \times \{z_j\} \times \{t_k\}$ Shared cell =  $n \times m \times k$  cube

e.g., 
$$\{u_1u_2\} \times \{z_{44}\} \times \{t_{22} \ t_{23} \ ... \ t_{30}\}$$

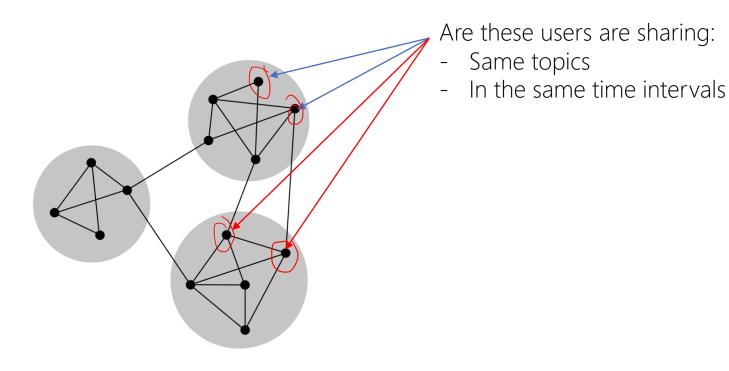
- User Clustering
  - Timeseries (Image) Clustering
  - User2Vec: User Vector Representation: Two Similar Users → Similar Vectors



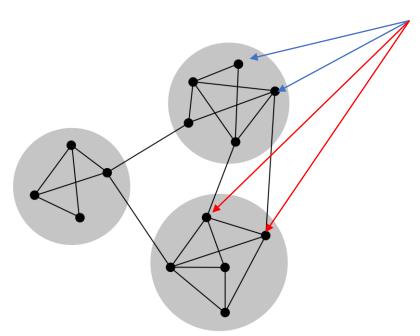
Louvain Method (Blondel et al. JSTAT 2008)



Evaluation: how accurate are the communities?



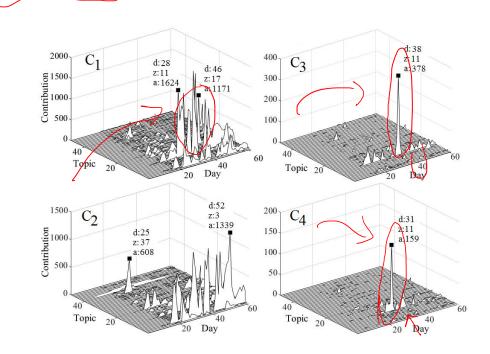
#### Evaluation: how accurate are the communities?



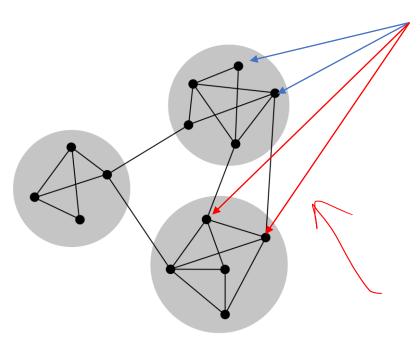
Are these users are sharing:

- Same topics
- In the same time intervals

#### Qualitative:



#### Evaluation: how accurate are the communities?



Are these users are sharing:

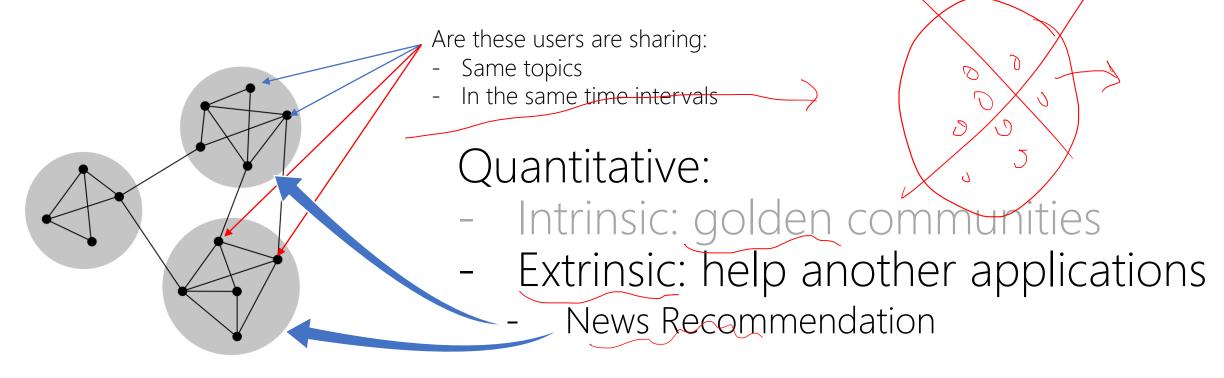
- Same topics
- In the same time intervals

#### Quantitative:

- Intrinsic: golden communities

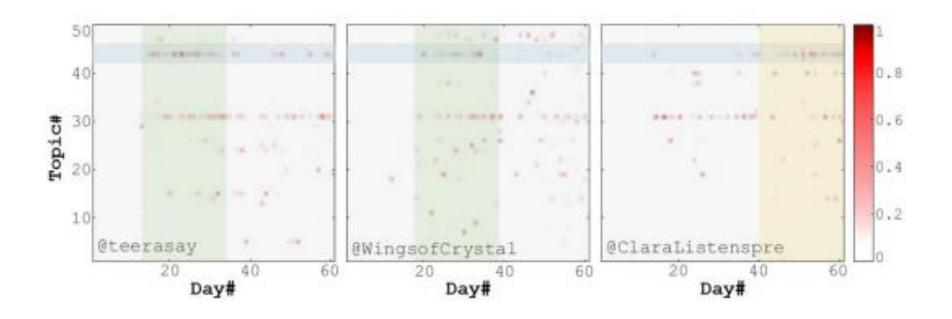
Rand

Evaluation: how accurate are the communities?



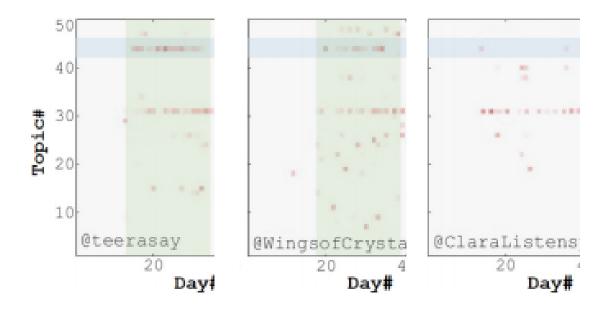
Recommend news articles to users to read at today, tomorrow, next week. Instead of per user recommendation, we recommend to the communities!

Evaluation: how accurate are the communities?

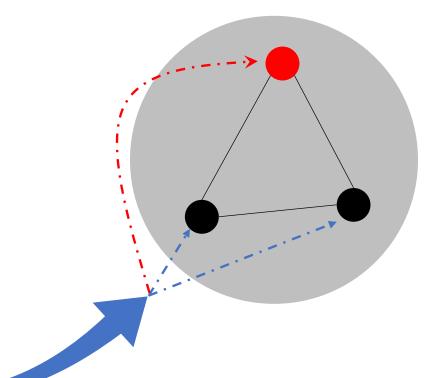


Recommend news articles to users to read at today, tomorrow, next week. Instead of per user recommendation, we recommend to the communities!

Evaluation: how accurate are the communities?



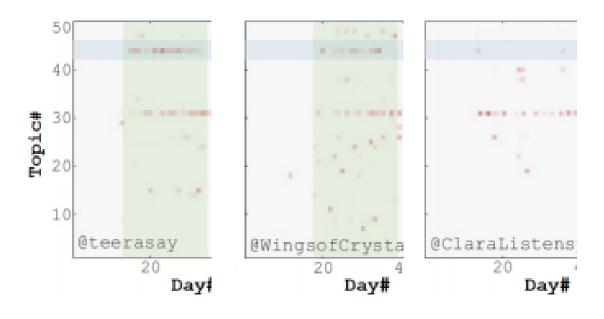
all three users in same community



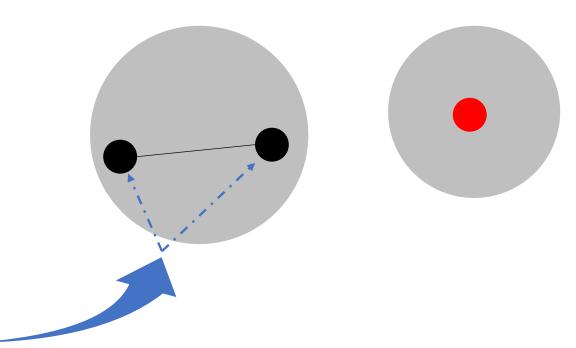
Recommend news articles about

- Z<sub>44</sub>: "War in Afghanistan"
- at day = 40

#### Evaluation: how accurate are the communities?



the first two users in same community the last user in another community

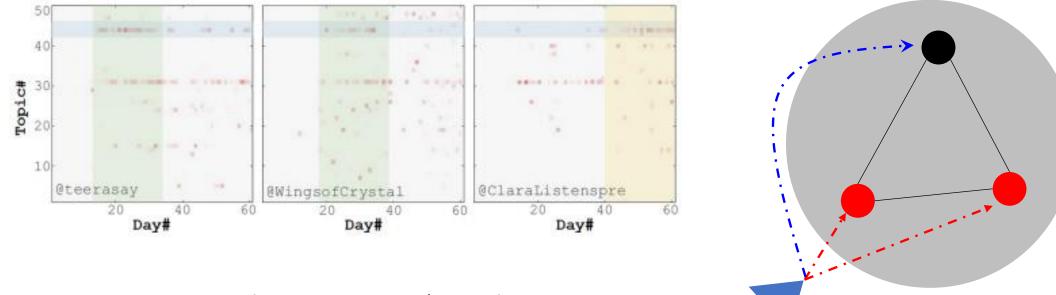


Recommend news articles about

- Z<sub>44</sub>: "War in Afghanistan"
- at day = 40

Evaluation: how accurate are the communities?

all three users in same community

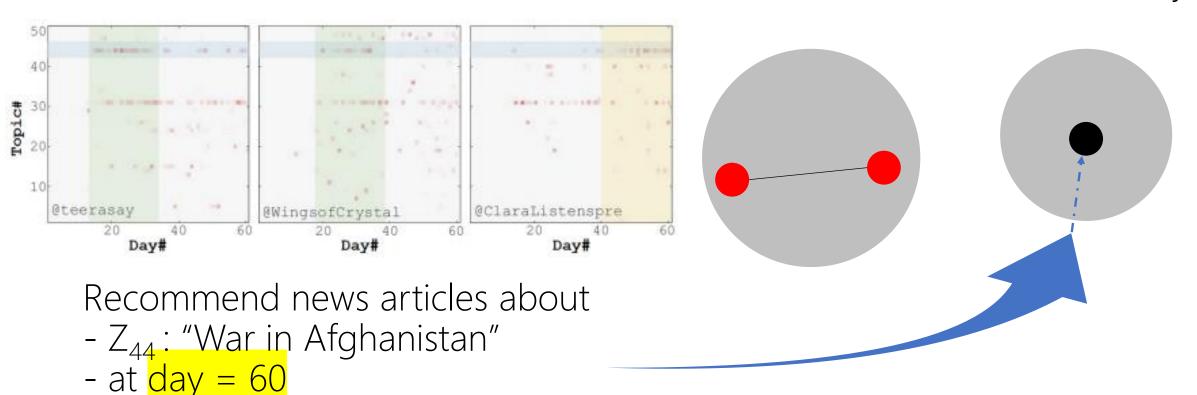


Recommend news articles about

- Z<sub>44</sub>: "War in Afghanistan"
- at day = 60

Evaluation: how accurate are the communities?

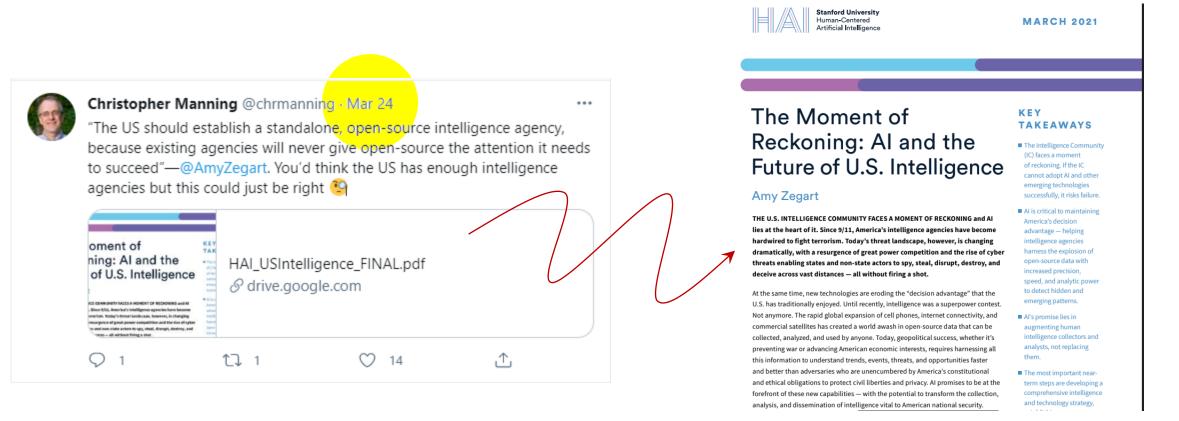
the first two users in same community the last user in another community



#### **Evaluation Strategy**

#### Assumption:

users are interested in the topics of the news article about which they have posted



#### **Evaluation Strategy**

Golden Dataset Curation:

news articles to which a user has explicitly linked in her tweets

mentions = {(user, news article, timestamp)}

#### **Evaluation Strategy**

News Recommendation:

- We recommend news article n
  - About topic z
- At timestamp t
- To a community that shows overall burst at time t about z

**Evaluation Strategy** 

News Recommendation:

mentions = {(user, ?, timestamp)}

We hope the community

- Read news article *n*
- Tweet news article *n*

#### **Evaluation Strategy**

News Recommendation:

```
mentions = {(user, ?, timestamp)}

We evaluate

- Our hope: (user, news article, timestamp)}

- The reality: (user, news article, timestamp)
```

#### **Evaluation Strategy**

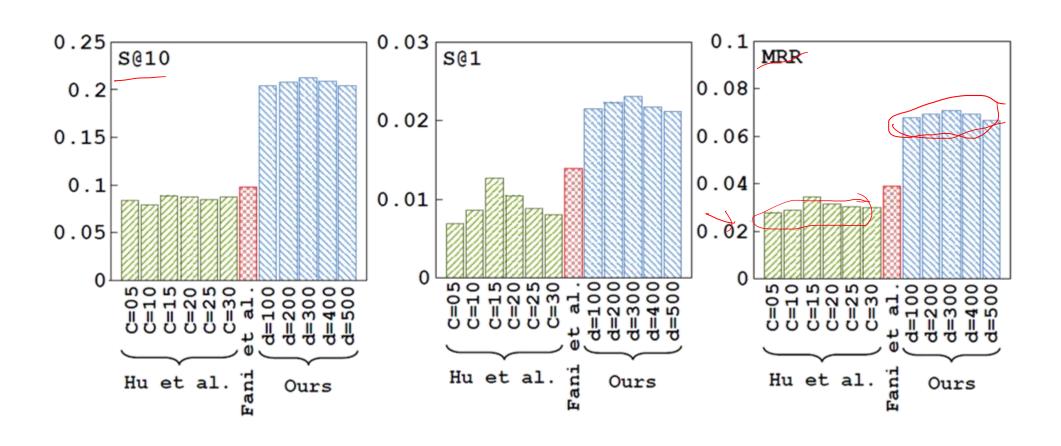
#### Dataset:

- Abel et al.: Twitter,
  - 3M tweets
  - Posted by 135K users
  - Between Nov. 1 and Dec. 31, 2010.

#### Golden Entries:

- 25,756 triples extracted from 3,468 distinct news articles posted by 1,922 users

#### **Evaluation Strategy**



#### Evaluation Strategy (Second)

#### User Prediction:

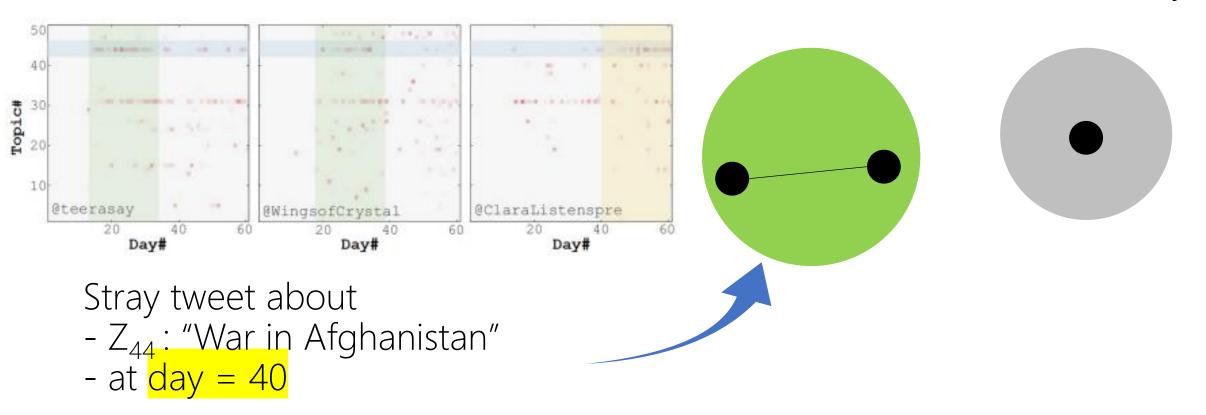
Seen a stray tweet, who is the author?

Probably from community  $m{C}$  because this community talked a lot about the topics of the mentioned news article n at timestamp t

mentions = {(?, news article, timestamp)}

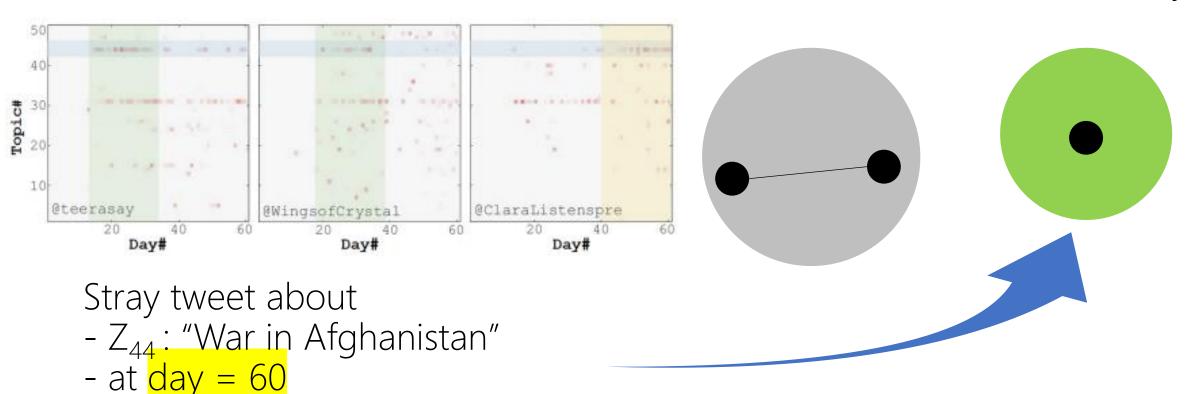
Evaluation: how accurate are the communities?

the first two users in same community the last user in another community



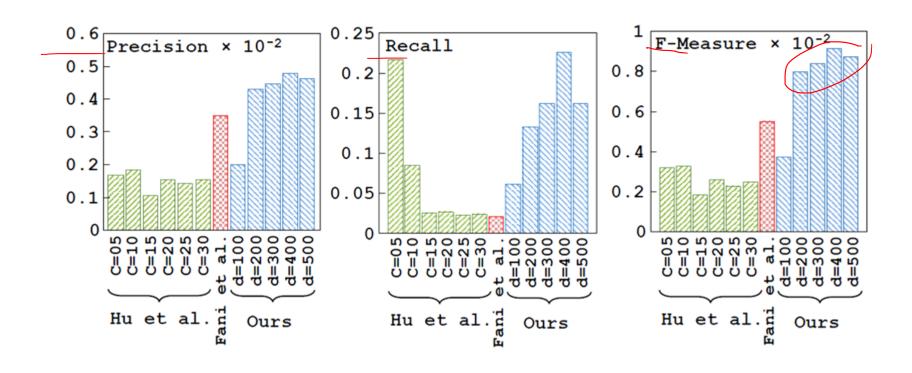
Evaluation: how accurate are the communities?

the first two users in same community the last user in another community



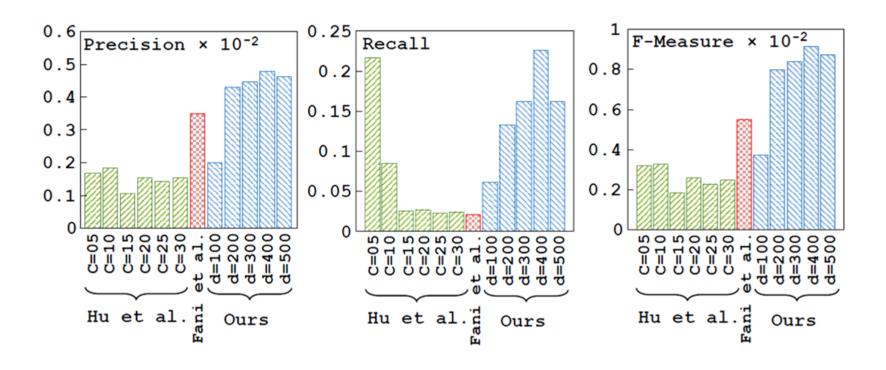
#### Evaluation Strategy (Second)

#### User Prediction:



#### Evaluation Strategy (Second)

#### User Prediction:



### User Community Prediction

### User Community Detection in Future!

Temporal Latent Space Modeling for Community Prediction

Hossein Fani<sup>1,2(⊠)</sup>, Ebrahim Bagheri<sup>2</sup>, and Weichang Du<sup>1</sup>

Faculty of Computer Science, University of New Brunswick, Fredericton, Canada {hfani,wdu}@unb.ca
Laboratory for Systems, Software and Semantics (LS3), Ryerson University, Toronto, Canada {hossein.fani,bagheri}@ryerson.ca

### User Community Prediction

