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 (Computational 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

You are what you post*! within time.*

Predicting Personal Life Events from Streaming Social Content

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

Hossein Fani
University of New Brunswick
hfani@unb.ca

Fattane Zarrinkalam
Ryerson University
fzarrinkalam@ryerson.ca

Ebrahim Bagheri
Ryerson University
bagheri@ryerson.ca

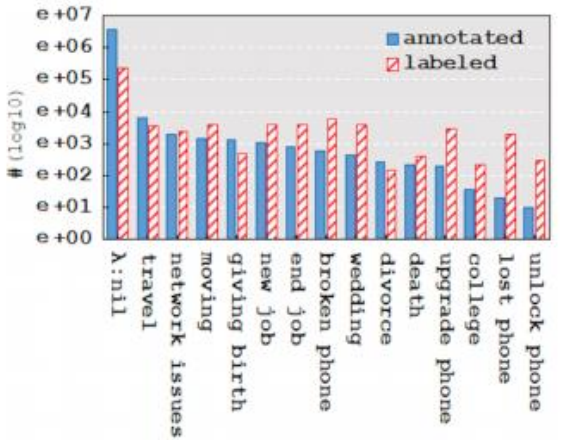
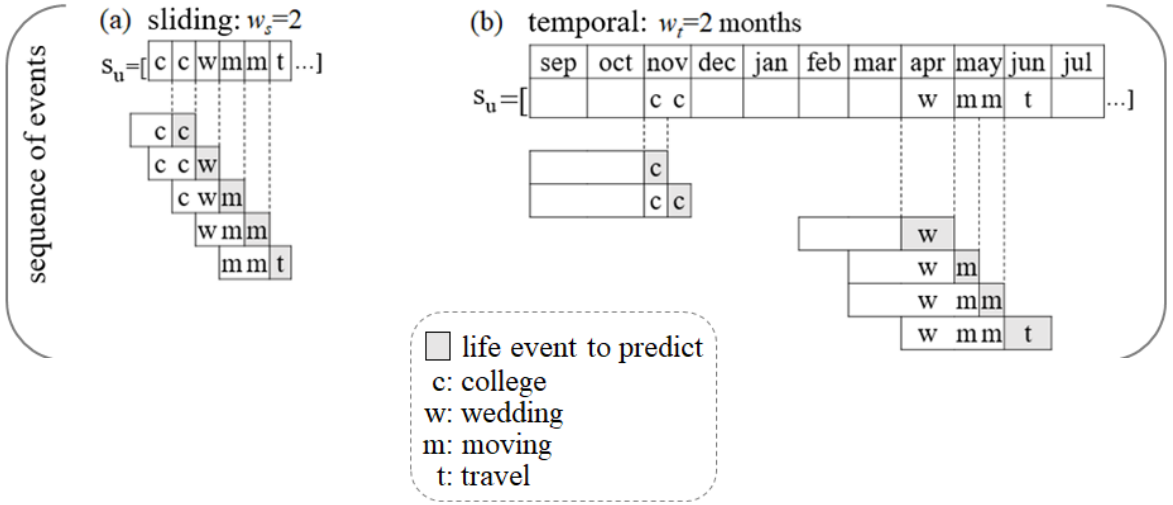


Figure 1: Distribution of personal life events by event class.



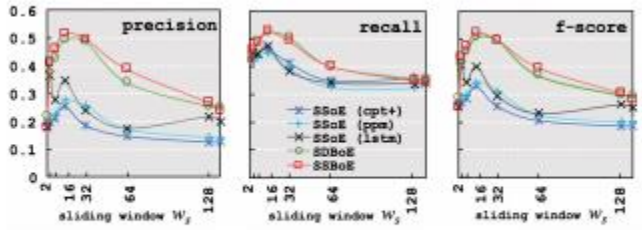


Figure 3: Comparative results of the *sliding* strategy.

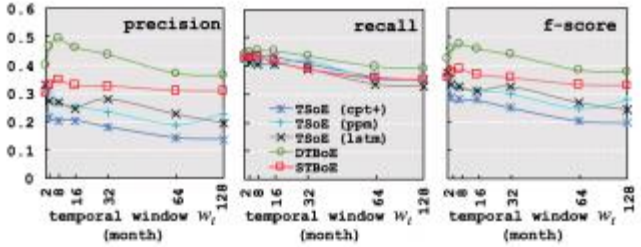


Figure 4: Comparative results of the *temporal* strategy.

You are what you post*! within time.*

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

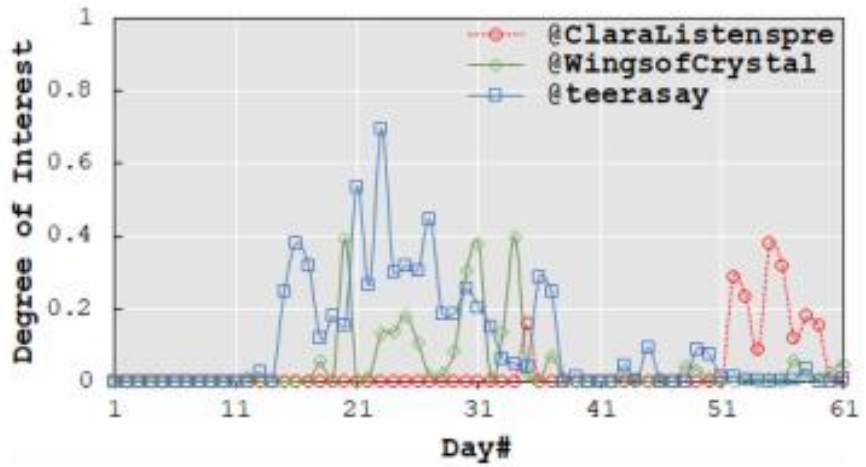


Fig. 1. Different temporal behaviour of three Twitter users with respect to the *'War in Afghanistan'* topic.

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

You are what you post*! within time.*

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

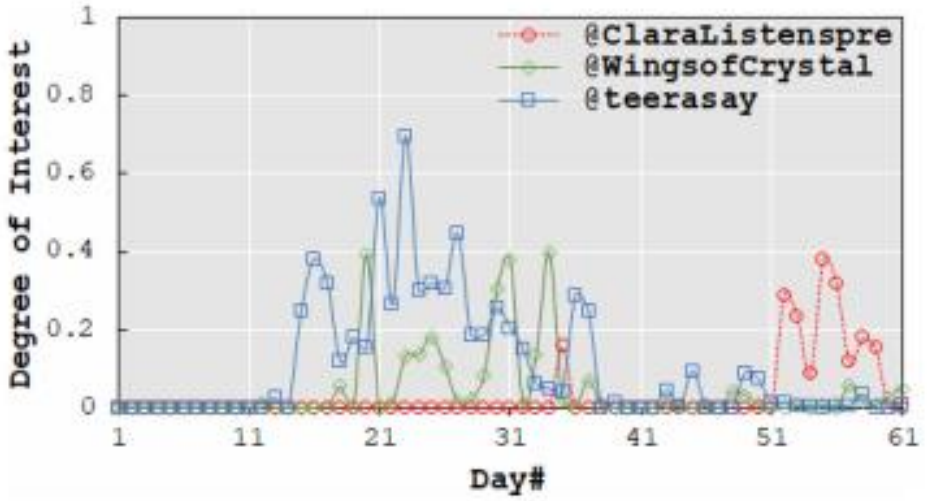
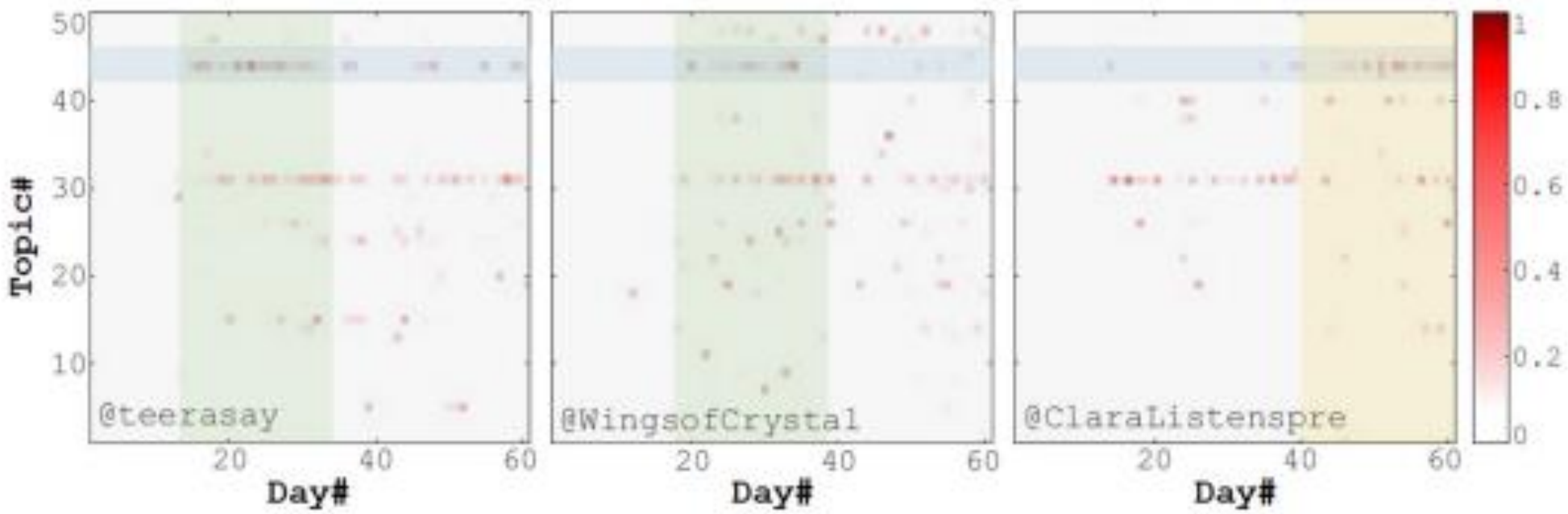


Fig. 1. Different temporal behaviour of three Twitter users with respect to the *'War in Afghanistan'* topic.

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

You are what you post*! within time.*

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



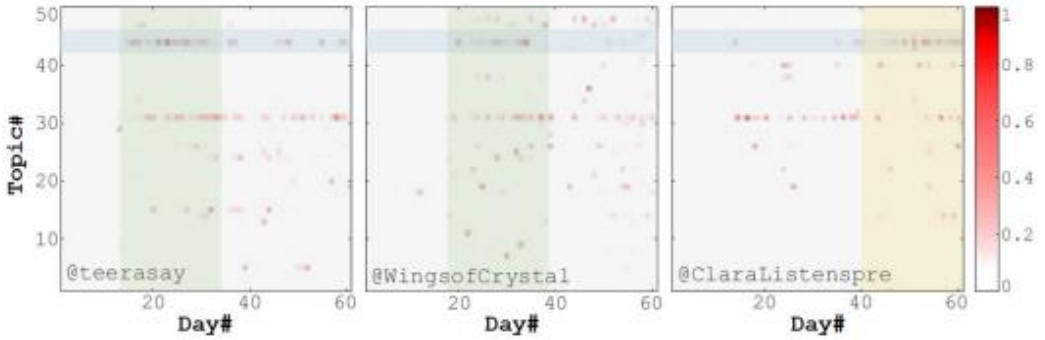
All users are interested in z_{44}='War in Afghanistan'
but not aligned in time*!*

Diachronically Like-minded User Community Detection

- User Clustering
  - Timeseries (Image) Clustering

User ⟷ Documents ⟶ User Vector ⟷ Document Vector

  - How to include time?



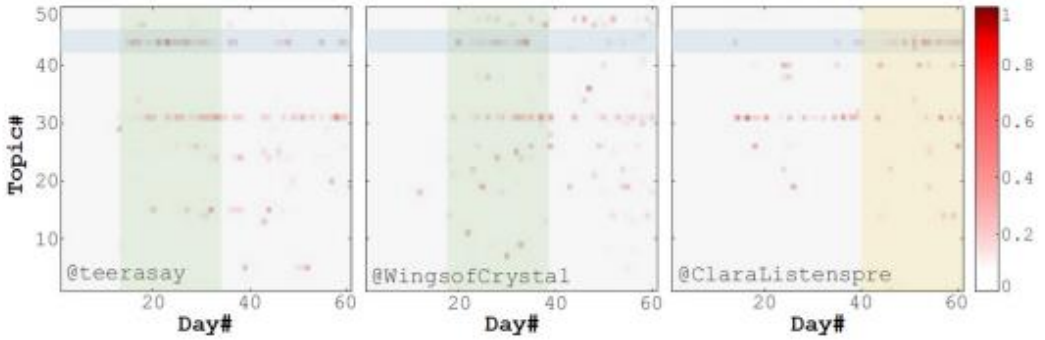
Diachronically Like-minded User Community Detection

- User Clustering
  - User Vector Representation

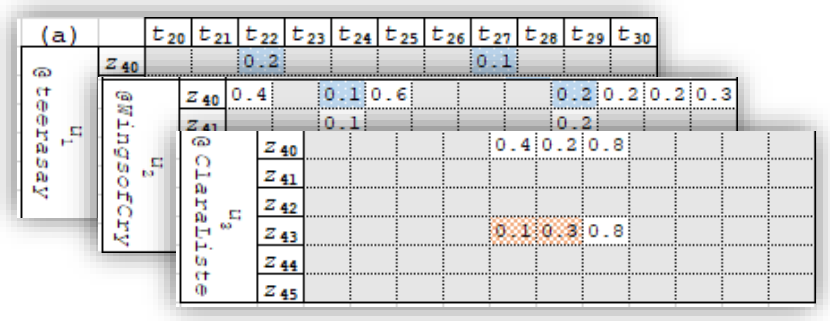
User ↔ Documents ⟶ User Vector ↔ Document Vector

  - How to include time?

User at time t ↔ A document that has all she said at time t
User = [Doc_0, Doc_1, …, Doc_T]



Diachronically Like-minded User Community Detection



$$\text{User} = [\text{Doc}_0, \text{Doc}_1, \ldots, \text{Doc}_T]$$

↓

LDA

↓

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

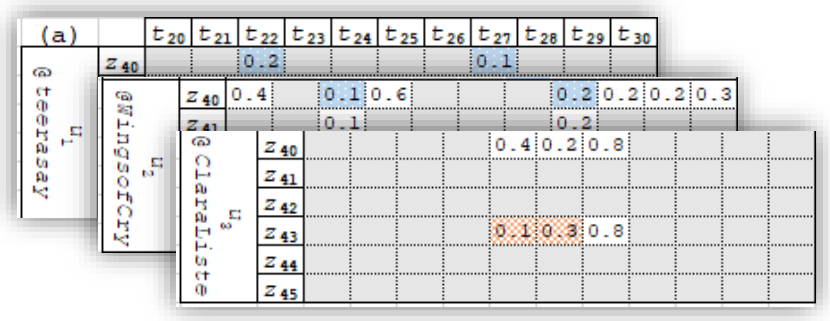
(a)

		t_{20}	t_{21}	t_{22}	t_{23}	t_{24}	t_{25}	t_{26}	t_{27}	t_{28}	t_{29}	t_{30}
@teerasay u_1	z_{40}			0.2					0.1			
	z_{41}											
	z_{42}											
	z_{43}	0.2	0.2	0.3								
	z_{44}	0.2	0.5	0.3	0.7	0.4	0.3	0.4	0.5	0.2	0.2	0.3
	z_{45}							0.2				
@WingsofCry u_2	z_{40}	0.4		0.1	0.6				0.2	0.2	0.2	0.3
	z_{41}			0.1					0.2			
	z_{42}								0.2			
	z_{43}	0.3		0.1		0.9	0.5		0.2		0.4	
	z_{44}	0.4		0.1	0.1	0.1	0.2	0.2	0.2	0.3	0.8	0.3
	z_{45}				0.5	0.2	0.5	0.2	0.2	0.4	0.2	
@ClaraListe u_3	z_{40}					0.4	0.2	0.8				
	z_{41}											
	z_{42}											
	z_{43}					0.1	0.3	0.8				
	z_{44}											
	z_{45}											

(b)

		t_{50}	t_{51}	t_{52}	t_{53}	t_{54}	t_{55}	t_{56}	t_{57}	t_{58}	t_{59}	t_{60}
@ClaraListe u_3	z_{40}		0.1	0.2	0.1	0.2			0.1		0.2	
	z_{41}				0.1							
	z_{42}		0.1			0.1						
	z_{43}		0.2		0.1							
	z_{44}			0.3	0.2	0.1			0.1		0.1	
	z_{45}		0.1	0.1	0.1						0.1	

Diachronically Like-minded User Community Detection



(a)		t_{20}	t_{21}	t_{22}	t_{23}	t_{24}	t_{25}	t_{26}	t_{27}	t_{28}	t_{29}	t_{30}
@teerasay u_1	z_{40}			0.2					0.1			
@WingsofCry u_2	z_{40}	0.4		0.1	0.6				0.2	0.2	0.2	0.3
	z_{41}			0.1					0.2			
@ClaraListe u_3	z_{40}								0.4	0.2	0.8	
	z_{41}											
	z_{42}											
	z_{43}								0.1	0.3	0.8	
	z_{44}											
	z_{45}											

(a)		t_{20}	t_{21}	t_{22}	t_{23}	t_{24}	t_{25}	t_{26}	t_{27}	t_{28}	t_{29}	t_{30}
@teerasay u_1	z_{40}			0.2					0.1			
	z_{41}											
	z_{42}											
	z_{43}	0.2	0.2	0.3								
	z_{44}	0.2	0.5	0.3	0.7	0.4	0.3	0.4	0.5	0.2	0.2	0.3
	z_{45}							0.2				
@WingsofCry u_2	z_{40}	0.4		0.1	0.6				0.2	0.2	0.2	0.3
	z_{41}			0.1					0.2			
	z_{42}								0.2			
	z_{43}	0.3		0.1		0.9	0.5		0.2		0.4	
	z_{44}	0.4		0.1	0.1	0.1	0.2	0.2	0.2	0.3	0.8	0.3
	z_{45}				0.5	0.2	0.5	0.2	0.2	0.4	0.2	
@ClaraListe u_3	z_{40}					0.4	0.2	0.8				
	z_{41}											
	z_{42}											
	z_{43}					0.1	0.3	0.8				
	z_{44}											
	z_{45}											

(b)		t_{50}	t_{51}	t_{52}	t_{53}	t_{54}	t_{55}	t_{56}	t_{57}	t_{58}	t_{59}	t_{60}
@ClaraListe u_3	z_{40}		0.1	0.2	0.1	0.2			0.1		0.2	
	z_{41}				0.1							
	z_{42}		0.1			0.1						
	z_{43}		0.2		0.1							
	z_{44}			0.3	0.2	0.1			0.1		0.1	
	z_{45}		0.1	0.1	0.1						0.1	

$$\text{User} = [\text{Doc}_0, \text{Doc}_1, \ldots, \text{Doc}_T]$$

↓

LDA

↓

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

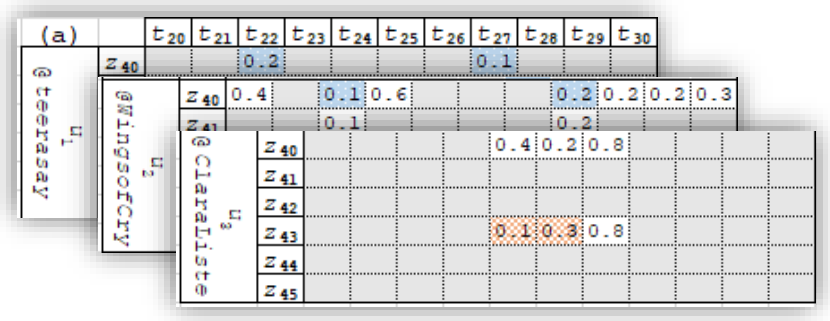
Two users are similar if they share more cells!

each cell = 1×1×1 cube $=\{u_i\}\times\{z_j\}\times\{t_k\}$

Shared cell = n×m×k cube

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

Diachronically Like-minded User Community Detection



(a)

		t_{20}	t_{21}	t_{22}	t_{23}	t_{24}	t_{25}	t_{26}	t_{27}	t_{28}	t_{29}	t_{30}
u_1 @teerasay	z_{40}			0.2					0.1			
	z_{41}											
	z_{42}											
	z_{43}	0.2	0.2	0.3								
	z_{44}	0.2	0.5	0.3	0.7	0.4	0.3	0.4	0.5	0.2	0.2	0.3
	z_{45}							0.2				
u_2 @WingsofCry	z_{40}	0.4		0.1	0.6				0.2	0.2	0.2	0.3
	z_{41}			0.1					0.2			
	z_{42}								0.2			
	z_{43}	0.3		0.1		0.9	0.5		0.2		0.4	
	z_{44}	0.4		0.1	0.1	0.1	0.2	0.2	0.2	0.3	0.8	0.3
	z_{45}				0.5	0.2	0.5	0.2	0.2	0.4	0.2	
u_3 @ClaraListe	z_{40}					0.4	0.2	0.8				
	z_{41}											
	z_{42}											
	z_{43}					0.1	0.3	0.8				
	z_{44}											
	z_{45}											

(b)

		t_{50}	t_{51}	t_{52}	t_{53}	t_{54}	t_{55}	t_{56}	t_{57}	t_{58}	t_{59}	t_{60}
u_3 @ClaraListe	z_{40}		0.1	0.2	0.1	0.2			0.1		0.2	
	z_{41}				0.1							
	z_{42}		0.1			0.1						
	z_{43}		0.2		0.1							
	z_{44}			0.3	0.2	0.1			0.1		0.1	
	z_{45}		0.1	0.1	0.1						0.1	

Region of Like-mindedness (RoL) iff

$$y_t^u[z] \approx y_t^v[z]$$

Two users are similar if they share more cells!

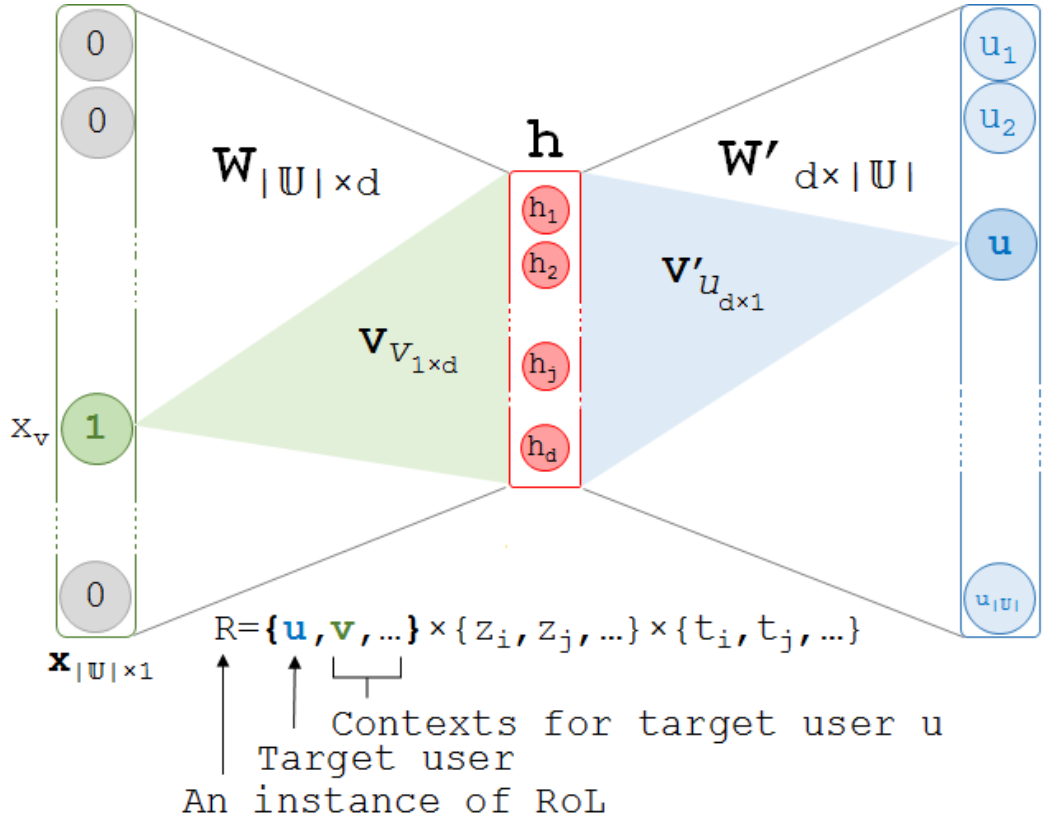
each cell = 1×1×1 cube = $\{u_i\}\times\{z_j\}\times\{t_k\}$

Shared cell = n×m×k cube

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

Diachronically Like-minded User Community Detection

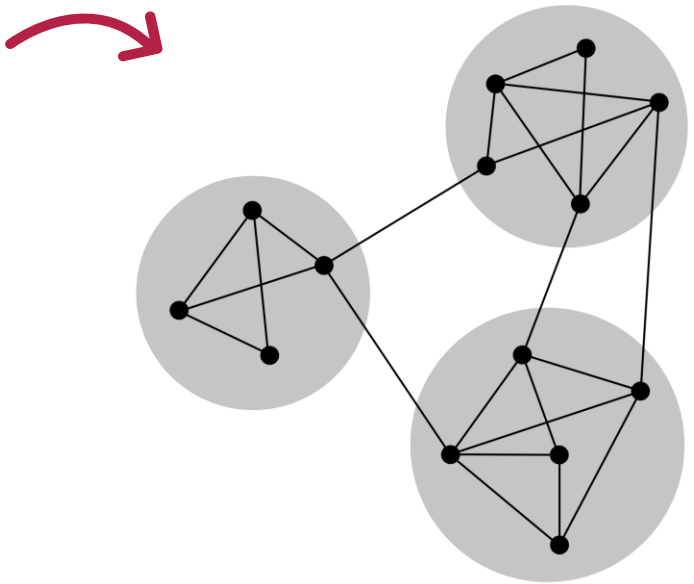
- User Clustering
  - ~~Timeseries (Image) Clustering~~
  - User2Vec: User Vector Representation: Two Similar Users → Similar Vectors



Diachronically Like-minded User Community Detection

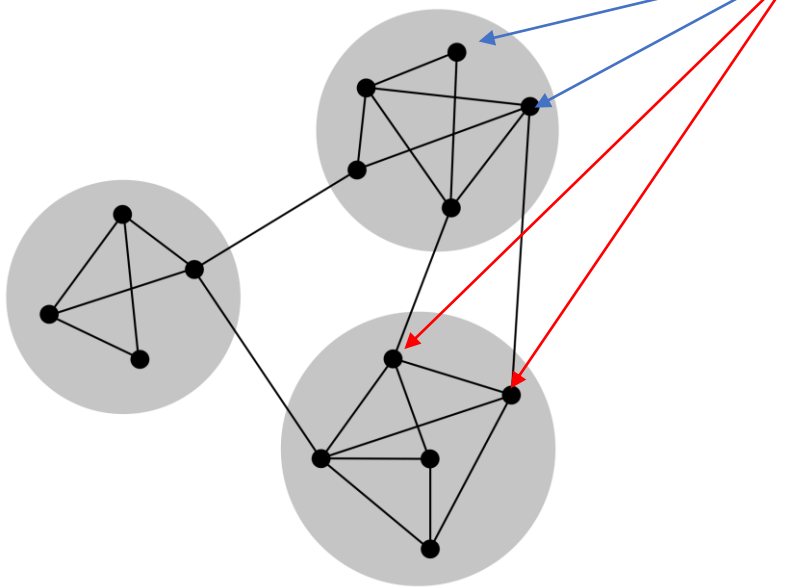
Louvain Method (Blondel et al. JSTAT 2008)

$$\begin{bmatrix} A_{11} & A_{12} & \cdots & A_{1n} \\ A_{21} & & & A_{2n} \\ \vdots & & & \vdots \\ A_{n1} & A_{n2} & \cdots & A_{nn} \end{bmatrix}$$



Diachronically Like-minded User Community Detection

Evaluation: how accurate are the communities?

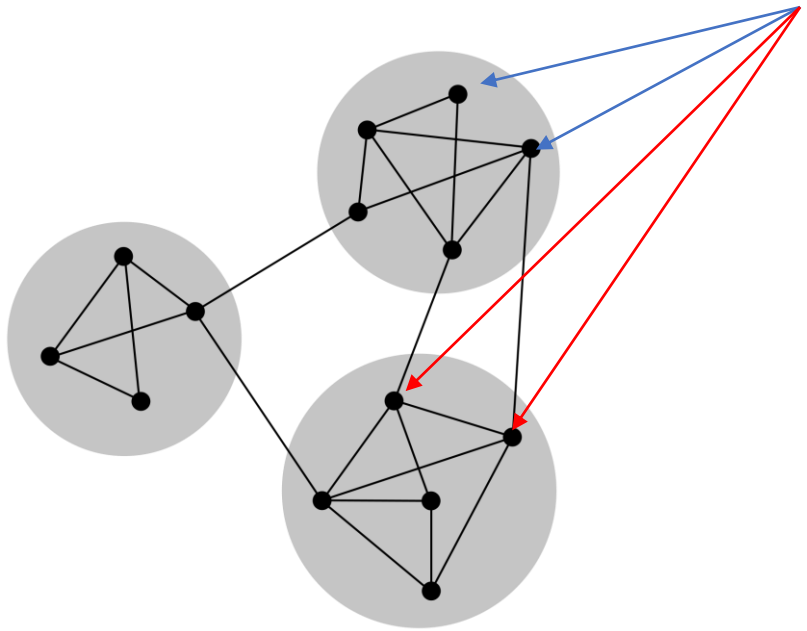


Are these users are sharing:

- Same topics
- In the same time intervals

Diachronically Like-minded User Community Detection

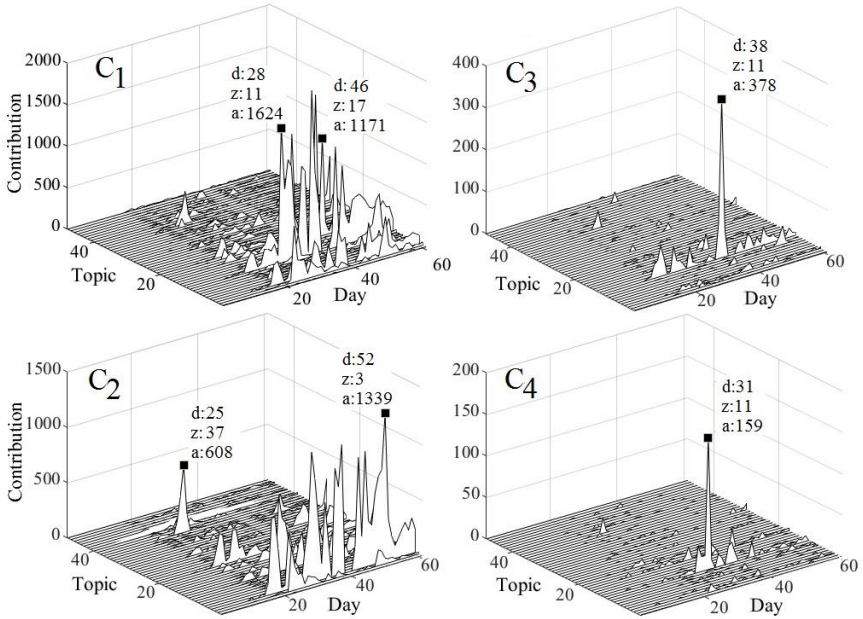
Evaluation: how accurate are the communities?



Are these users are sharing:

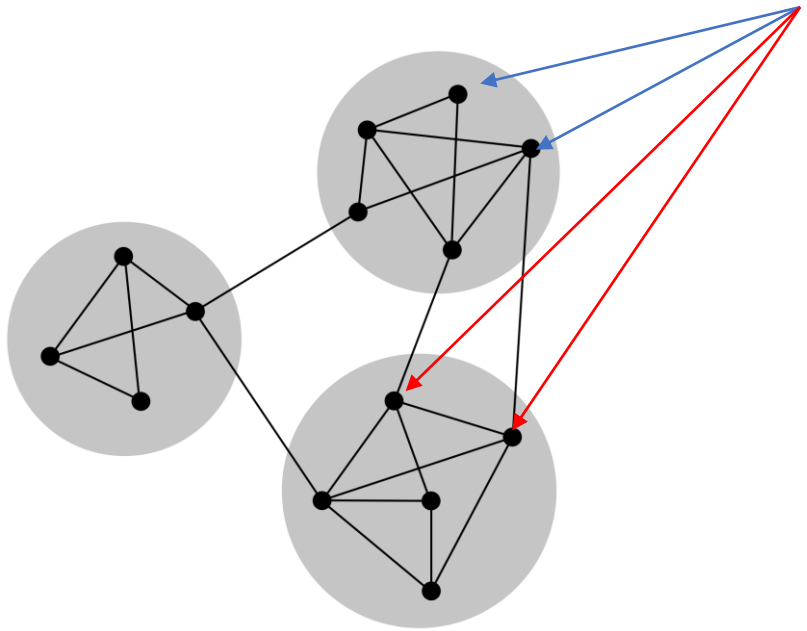
- Same topics
- In the same time intervals

Qualitative:



Diachronically Like-minded User Community Detection

Evaluation: how accurate are the communities?

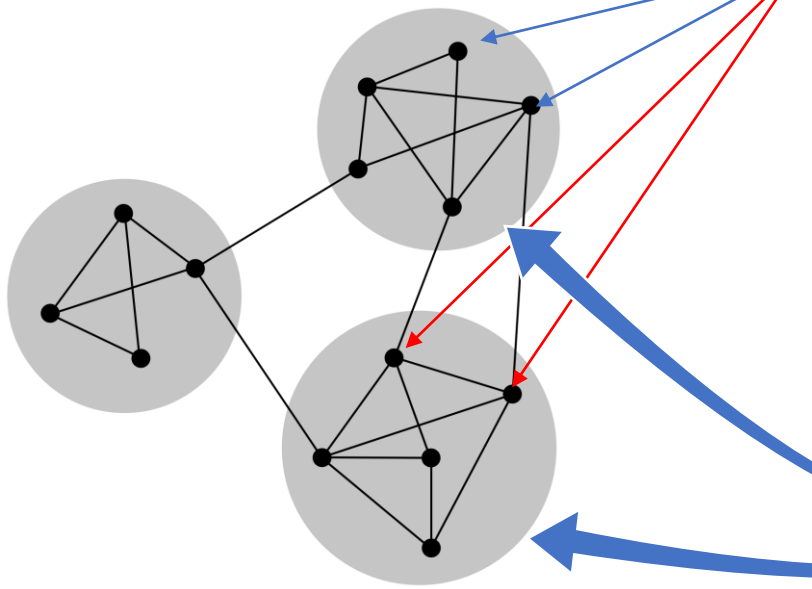


Are these users are sharing:
- Same topics
- In the same time intervals

Quantitative:
- Intrinsic: golden communities

Diachronically Like-minded User Community Detection

Evaluation: how accurate are the communities?



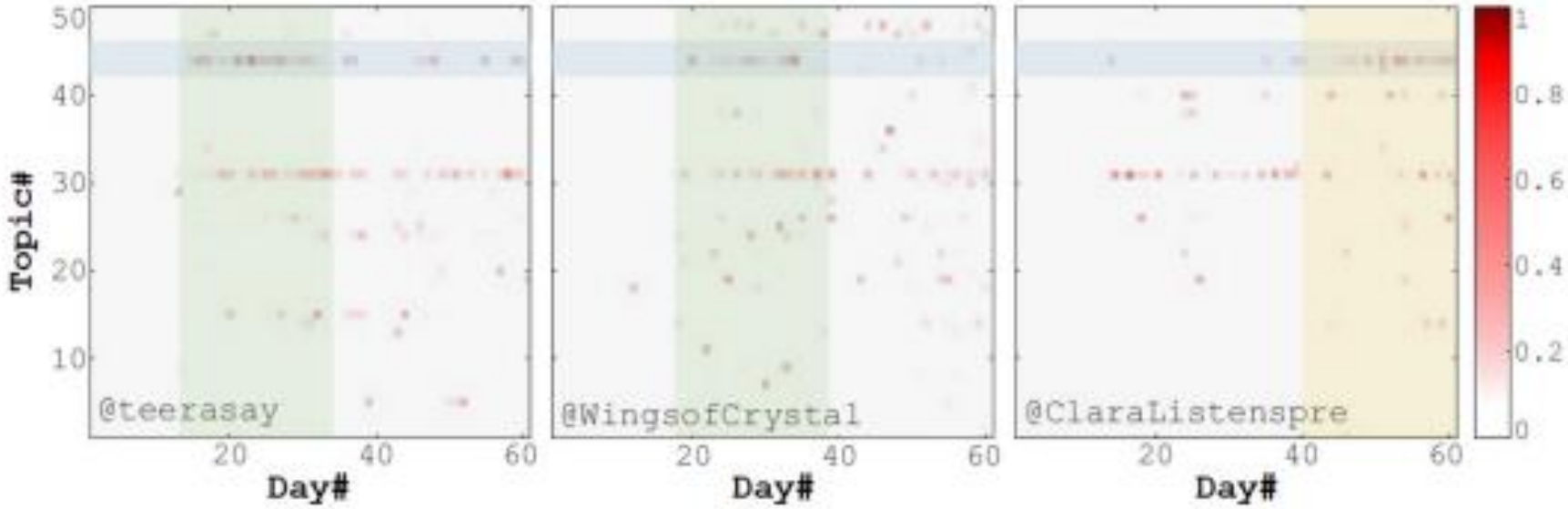
Are these users are sharing:
- Same topics
- In the same time intervals

Quantitative:
- Intrinsic: golden communities
- Extrinsic: help another applications
  - News Recommendation

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

Diachronically Like-minded User Community Detection

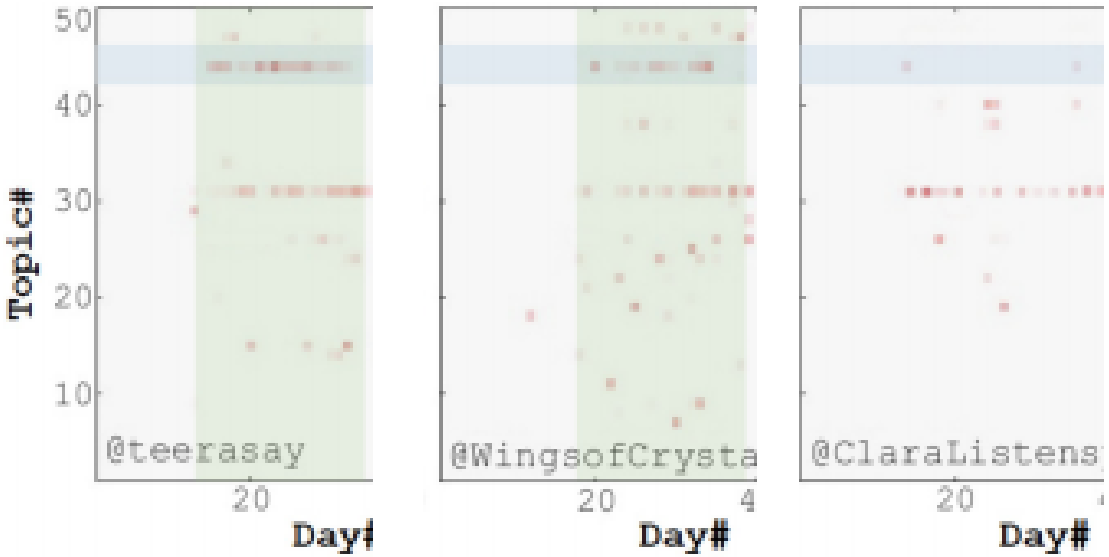
Evaluation: how accurate are the communities?



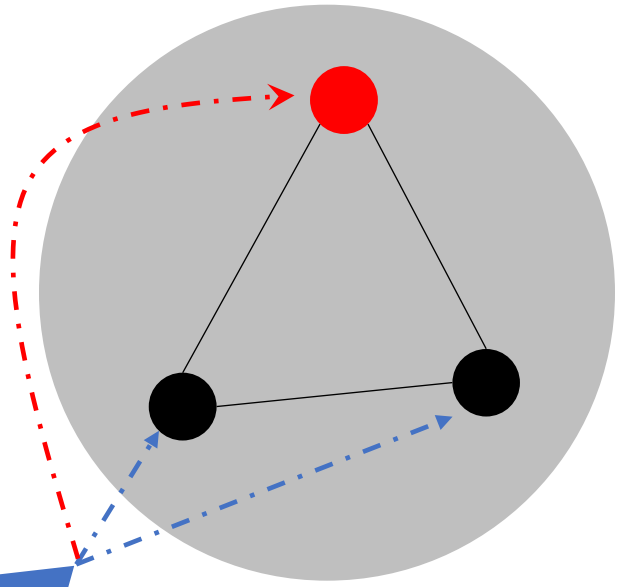
Recommend news articles to users to read at today, tomorrow, next week.
Instead of per user recommendation, we recommend to the communities!

Diachronically Like-minded User Community Detection

Evaluation: how accurate are the communities?



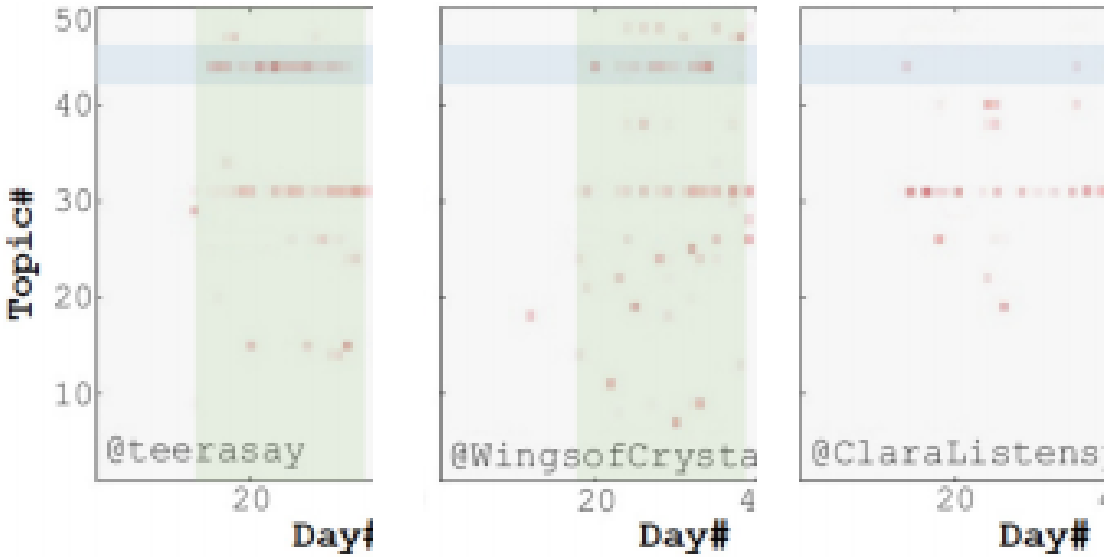
all three users in same community



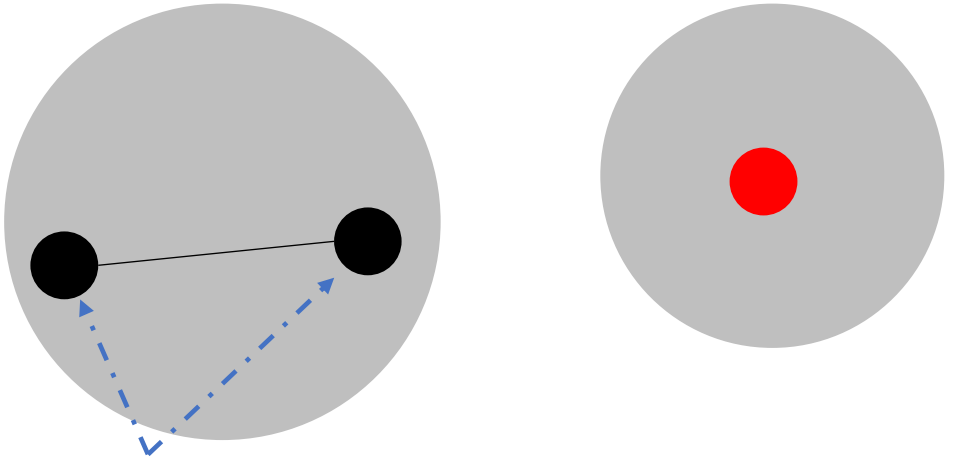
Recommend news articles about
- Z_{44} : "War in Afghanistan"
- at day = 40

Diachronically Like-minded User Community Detection

Evaluation: how accurate are the communities?



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

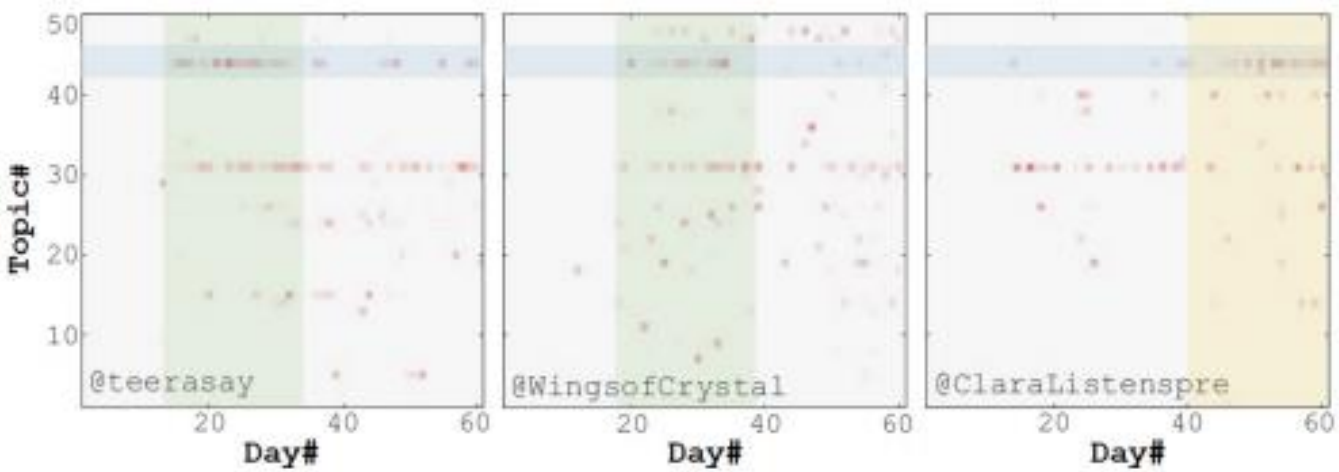


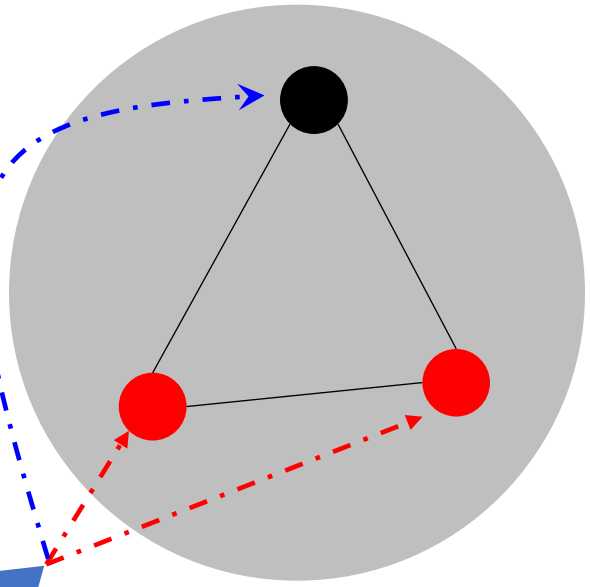
Recommend news articles about
- Z_{44} : "War in Afghanistan"
- at day = 40

Diachronically Like-minded User Community Detection

Evaluation: how accurate are the communities?

all three users in same community



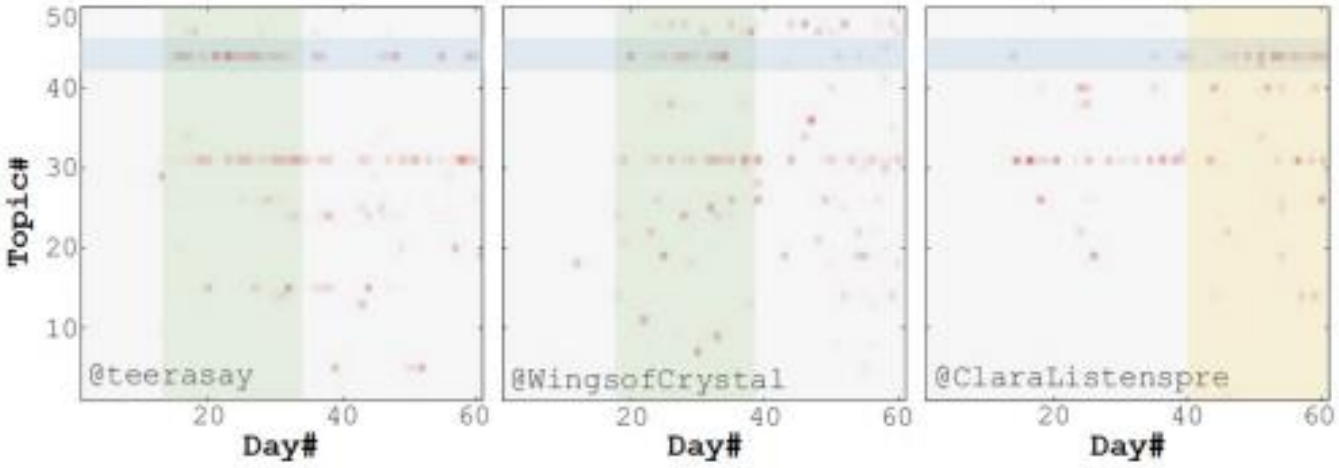


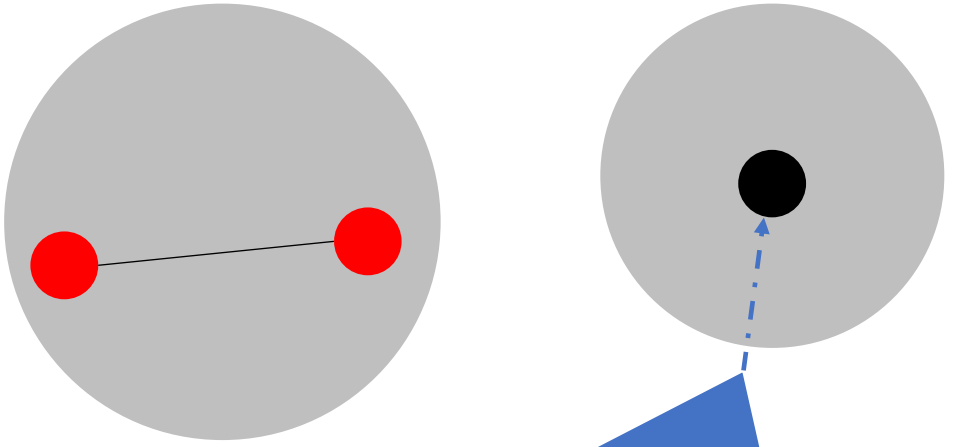
Recommend news articles about
- Z_{44} : "War in Afghanistan"
- at day = 60

Diachronically Like-minded User Community Detection

Evaluation: how accurate are the communities?

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





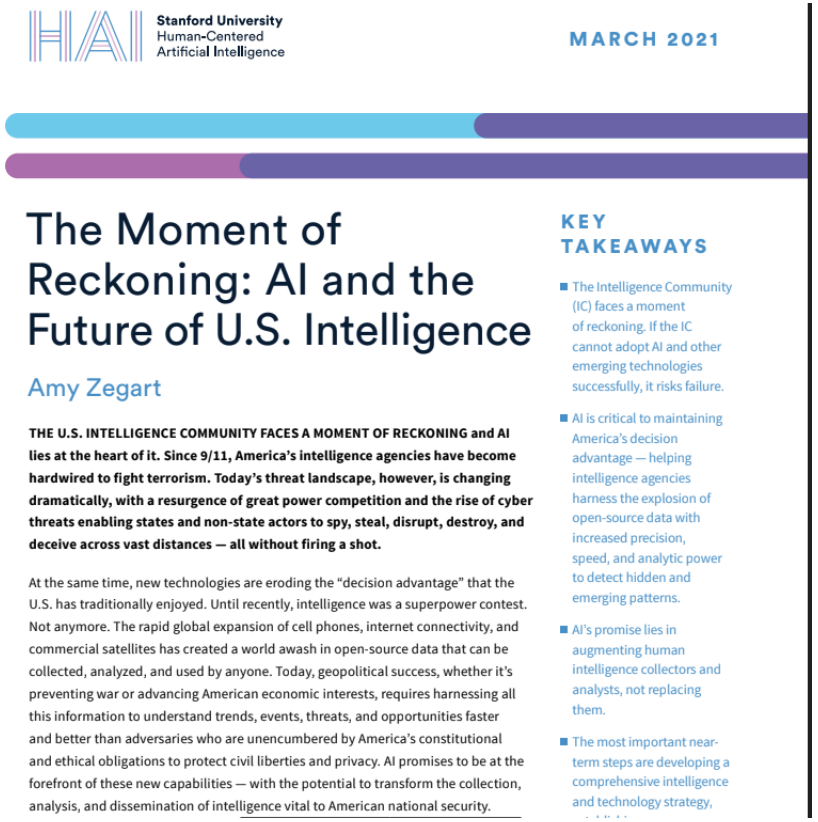
Recommend news articles about
- Z_{44} : "War in Afghanistan"
- at day = 60

Diachronically Like-minded User Community Detection

Evaluation Strategy

Assumption:

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



Diachronically Like-minded User Community Detection

Evaluation Strategy

Golden Dataset Curation:

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

$$\text{mentions} = \{(\text{user}, \text{news article}, \text{timestamp})\}$$

Diachronically Like-minded User Community Detection

Evaluation Strategy

News Recommendation:

$$\text{mentions} = \{(\text{user}, ?, \text{timestamp})\}$$

We recommend news article n

- About topic z
- At timestamp t
- To a community that shows overall burst at time t about z

Diachronically Like-minded User Community Detection

Evaluation Strategy

News Recommendation:

$$\text{mentions} = \{(\text{user}, ?, \text{timestamp})\}$$

We hope the community

- Read news article $\boldsymbol{n}$
- Tweet news article $\boldsymbol{n}$

Diachronically Like-minded User Community Detection

Evaluation Strategy

News Recommendation:

mentions = {(user, ?, timestamp)}

We evaluate

- Our hope: (user, ***n*** , timestamp)}
- The reality: (user, ***news article***, timestamp)

Diachronically Like-minded User Community Detection

Evaluation Strategy

Dataset:

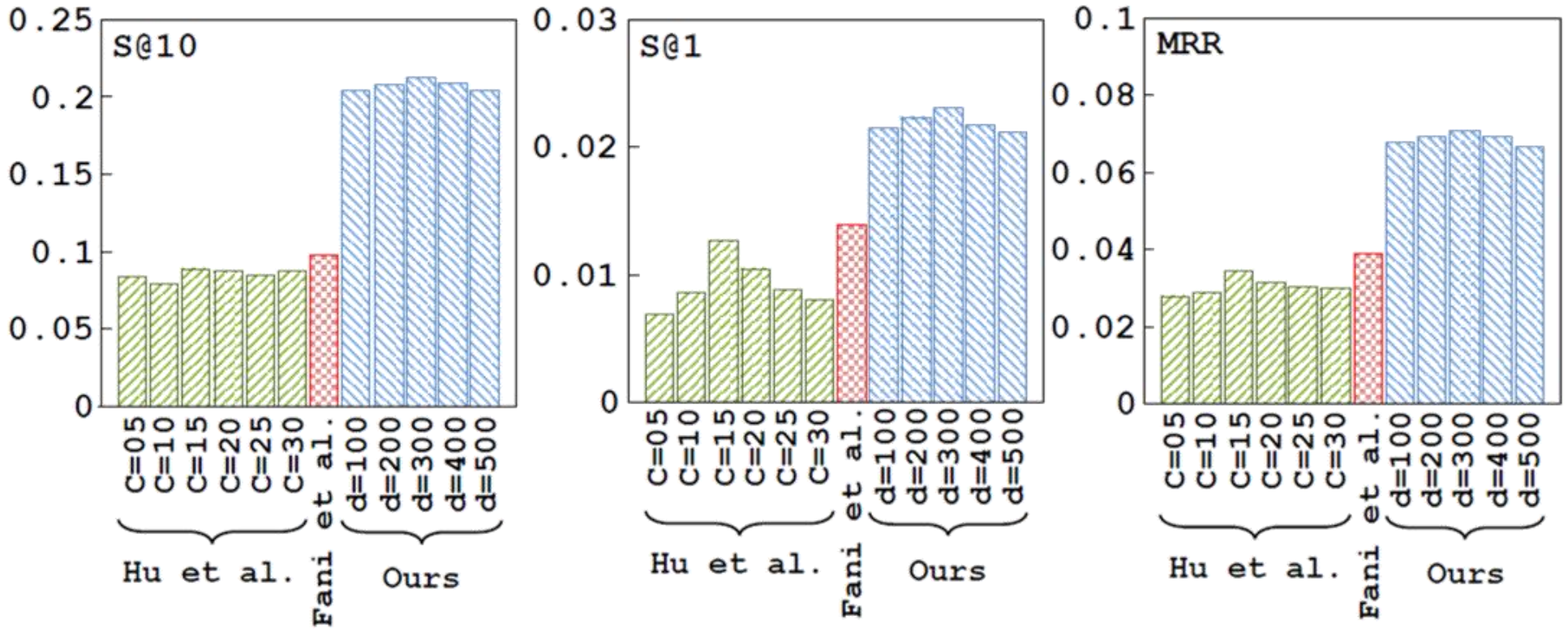
- Abel et al.: Twitter,
  - 3M tweets
  - Posted by 135K users
  - Between Nov. 1 and Dec. 31, 2010.

Golden Entries:

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

Diachronically Like-minded User Community Detection

Evaluation Strategy



Diachronically Like-minded User Community Detection

Evaluation Strategy (Second)

User Prediction:

Seen a stray tweet, who is the author?

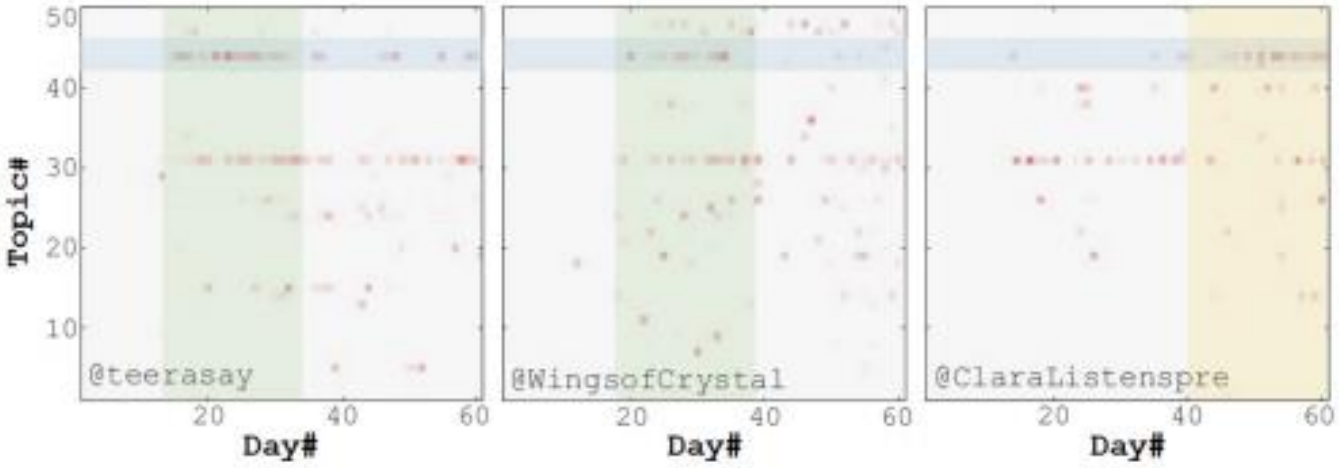
Probably from community C because this community talked a lot about the topics of the mentioned news article n at timestamp t

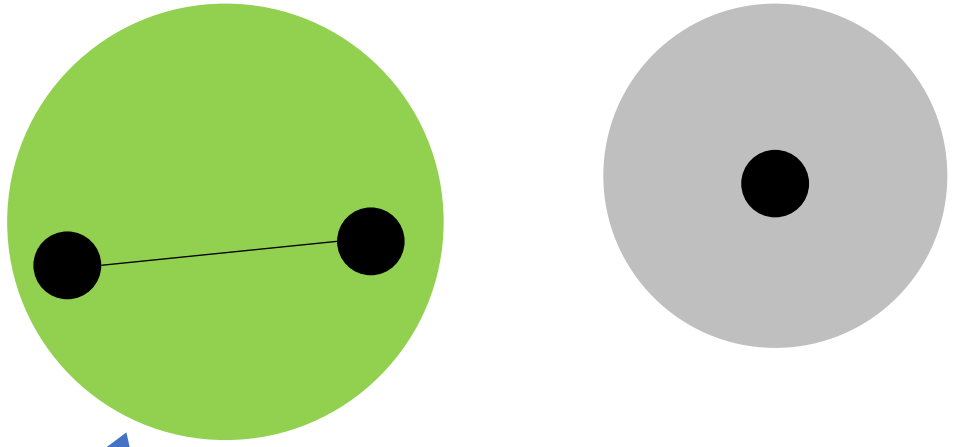
mentions = {(?, news article, timestamp)}

Diachronically Like-minded User Community Detection

Evaluation: how accurate are the communities?

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



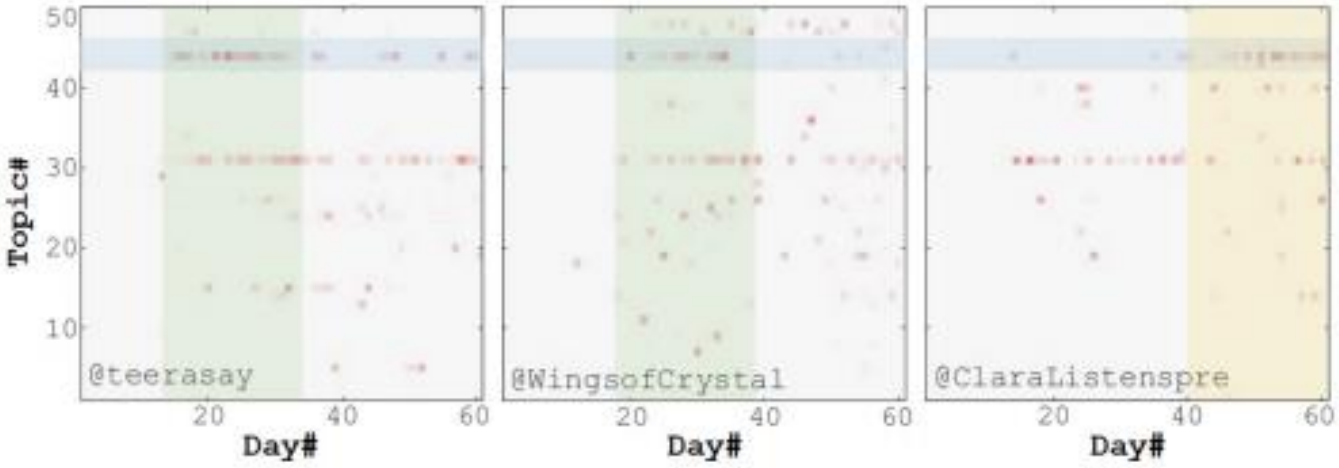


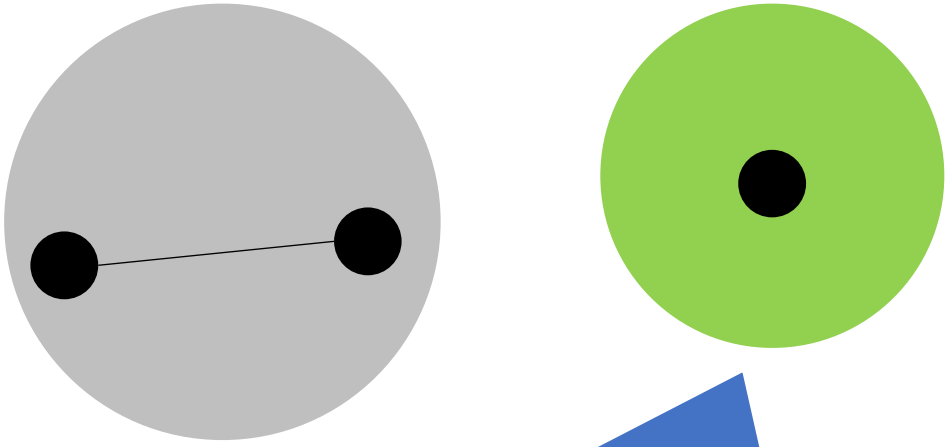
Stray tweet about
- Z_{44} : "War in Afghanistan"
- at day = 40

Diachronically Like-minded User Community Detection

Evaluation: how accurate are the communities?

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



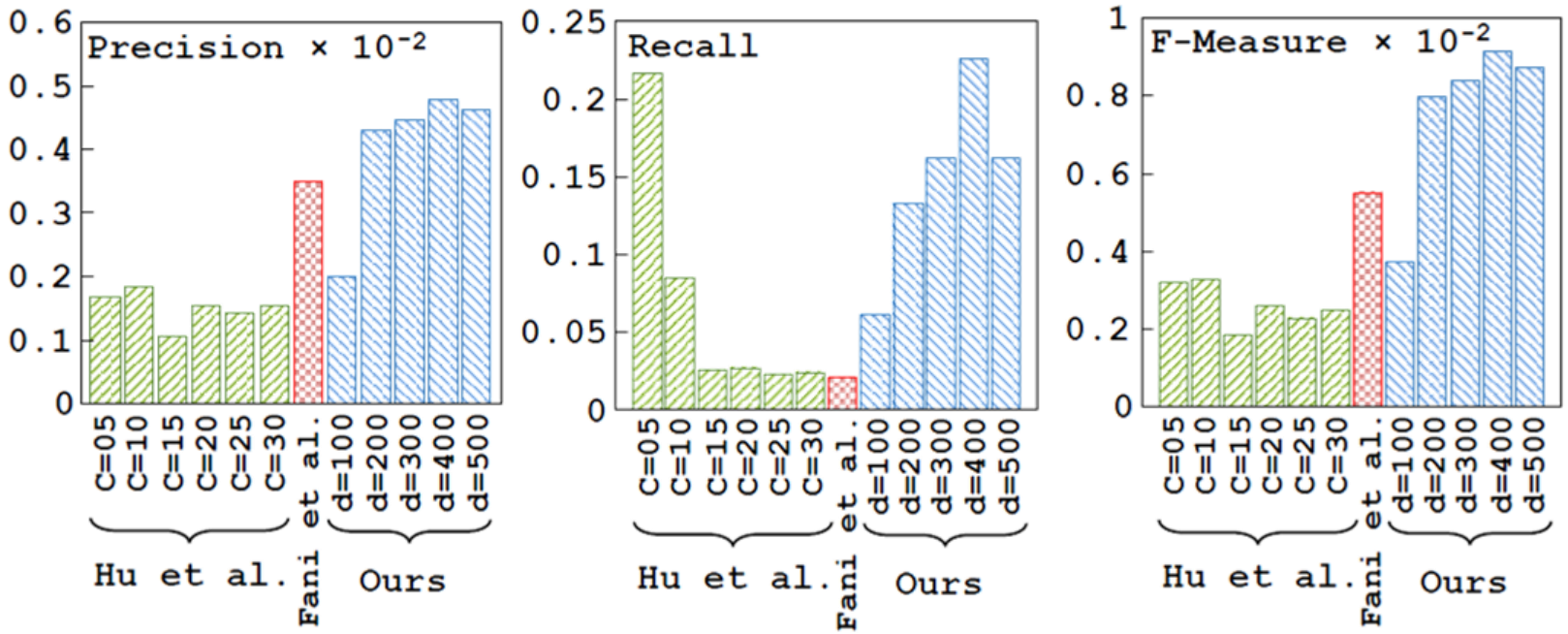


Stray tweet about
- Z_{44} : "War in Afghanistan"
- at day = 60

Diachronically Like-minded User Community Detection

Evaluation Strategy (Second)

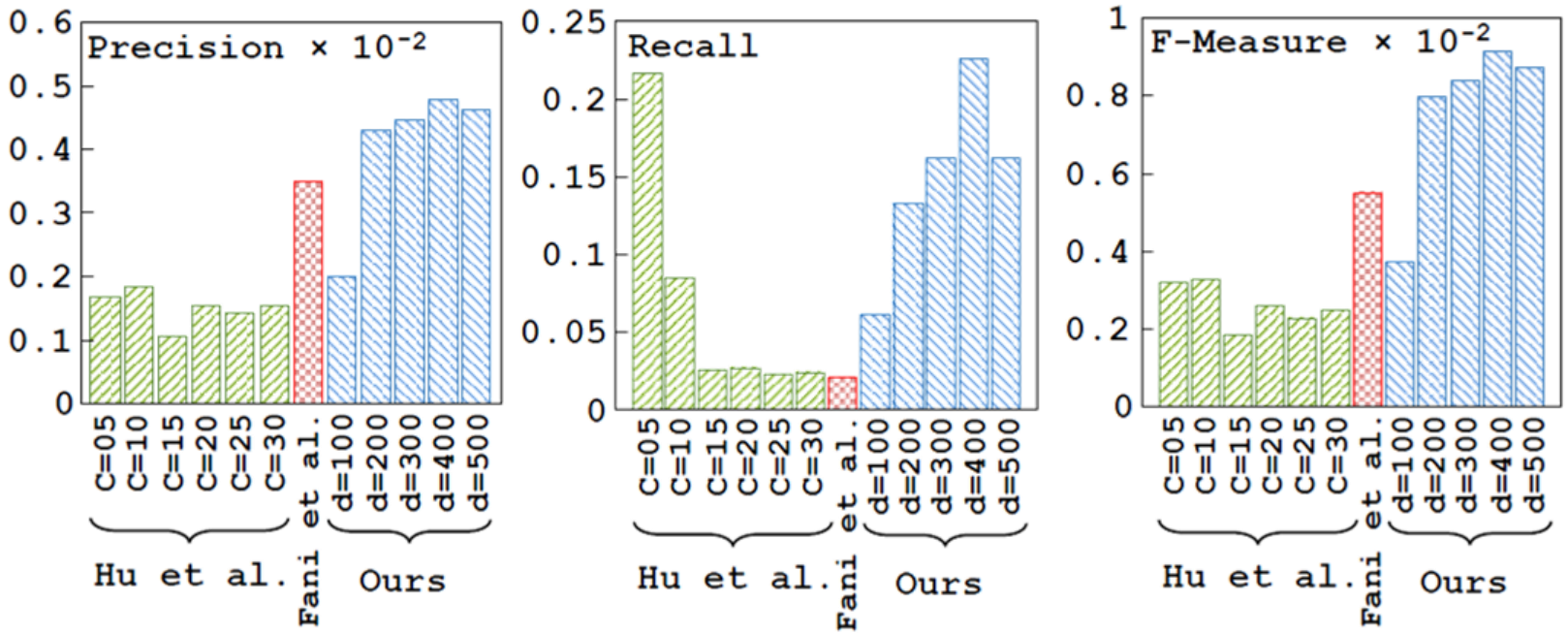
User Prediction:



Diachronically Like-minded User Community Detection

Evaluation Strategy (Second)

User Prediction:



User Community Prediction

User Community Detection in Future!

User Community Prediction