

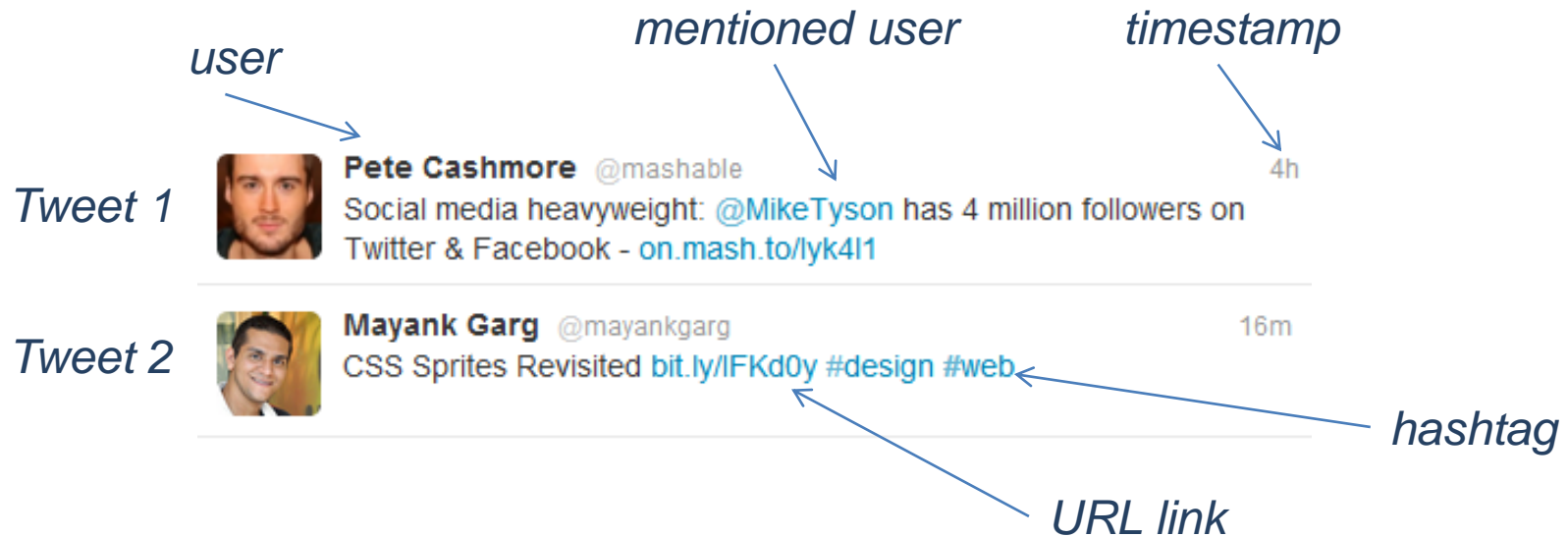


Searching Microblogs – Big Topics on Big Data

Wilfred Ng

Lightning Research Introduction Seminar
Oct-12-2015

Microblogs



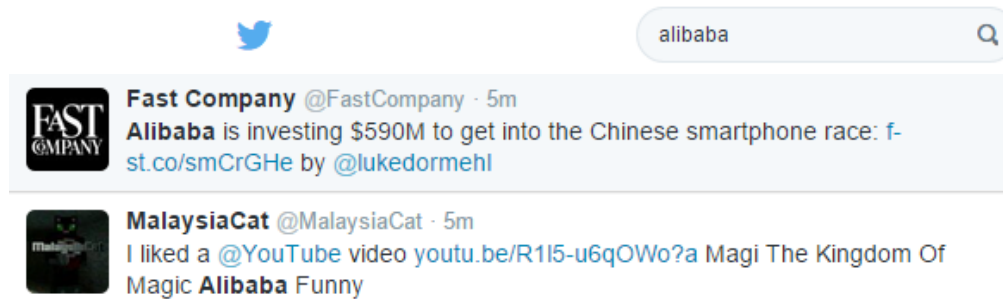
- Both social network and social media
 - Users post updates (tweets)
 - Links between users (follow, mention, re-tweet)

Who Uses Twitter?

- As of 2015,
 - Over 300 million monthly active users
 - Nearly 600 million tweets sent per day
 - In the US, 23% of online adults use Twitter
 - User base skewed toward younger users (<30 years old)
 - 43% of users use mobile



Twitter Search



- Over 2 billion search queries per day (2015)
 - **Timely information:** news, trends, events
 - E.g., weather, breaking news, road accidents
 - **Social information:** celebrities, public opinions
 - E.g. Public opinions on president election
 - **Topical information**
 - E.g. new products or technologies of Apple

Problems with Twitter Search

Around 50 tweets mentioning "iPad" posted within 1-minute

Results for **ipad**

Tip: use [operators](#) for advanced search.

Tweets

Tweets with links

Tweets near you

People



iPad Station Dennis Waymire

Supplies Of iPad 2 Starting To Improve <http://tinyurl.com/3mxd6ly>

27 seconds ago



KatiaNikitina Katia Nikitina

I'm playing #TouchPetsDogs2 on iPad! Find me on plus+, my username is 'nikitinakatia'. <http://ngmo.co/fYR8xD>

25 seconds ago



vanzeeben Ben Van Zee

does anyone with an iPad ever use it for work? if so, do you ever wish you had 3g?

25 seconds ago



captaintweet97 Captain Tweet

The Week in iPad Cases: Mix and match <http://goo.gl/fb/OqcZL>

25 seconds ago



robert_larsen Robert Larsen

Designing for iPad and listening to Iggy Pop.

25 seconds ago

Issue **A**

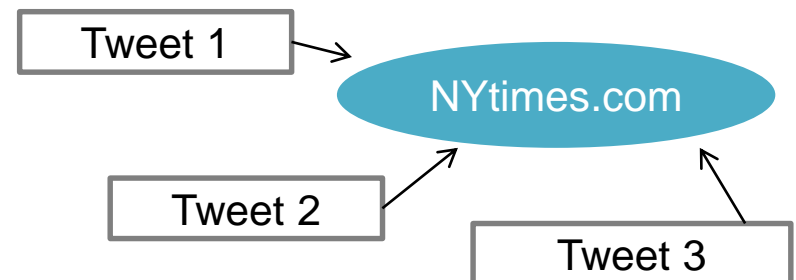
Analysis in Microblogs

Feature categories for Analysing Quality Content

- Punctuation and spelling
- Syntactic and semantic complexity
- Grammaticality
- Timestamp
- Link-based features (**Important!**)

Link-based features

- Links to other items
 - Re-tweet, reply tweet, mention of users
 - URLs, hashtags
- Link target's reputation
 - URL domain
 - Hashtag
 - User (tweet's author)



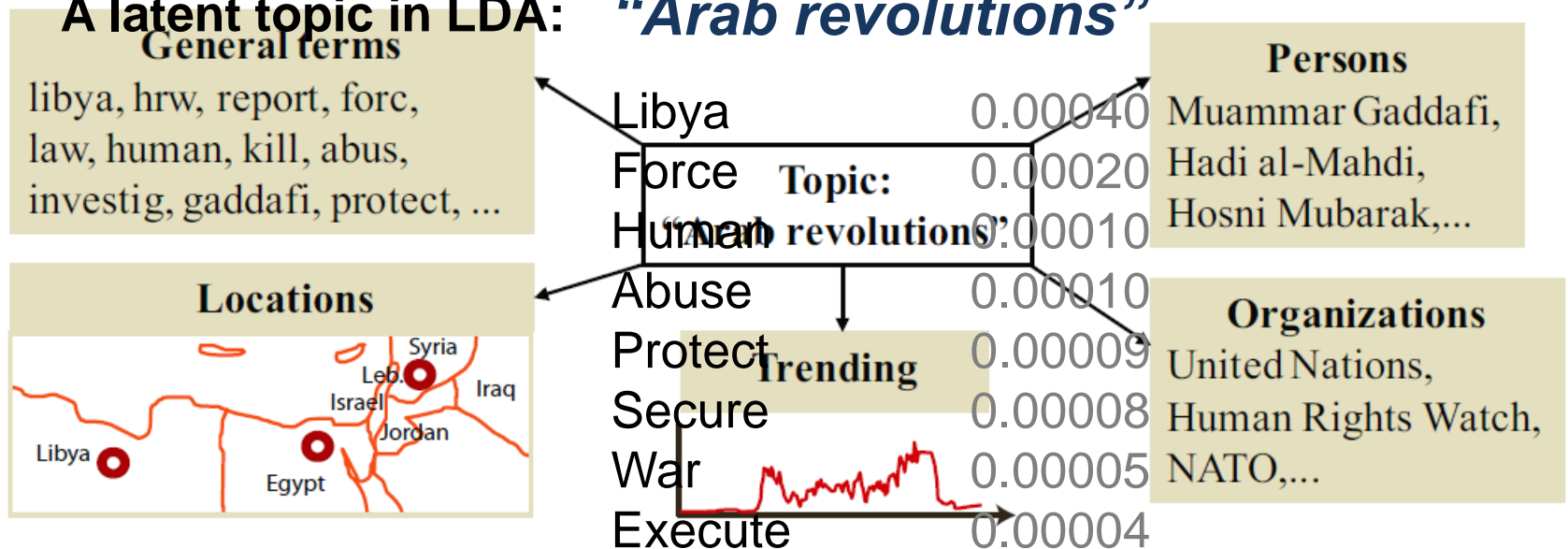
Issue **B**

Broadening Data Semantics

Challenge of Semantic Issues

- Tweet shortness
 - Multi-faceted Topic Modeling
 - Geographic Topic Modeling
- Fail to capture rich semantics

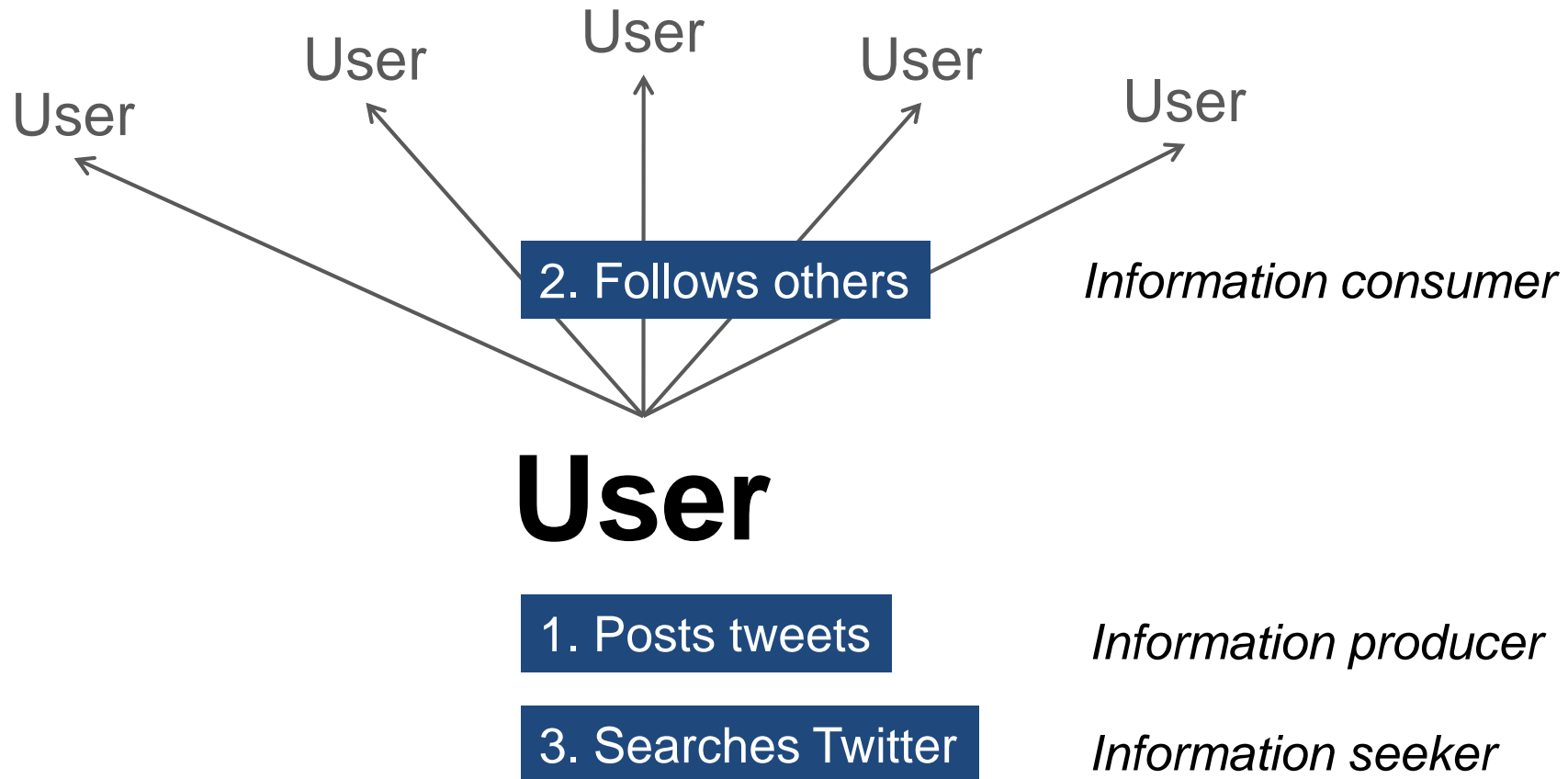
A latent topic in LDA: *“Arab revolutions”*



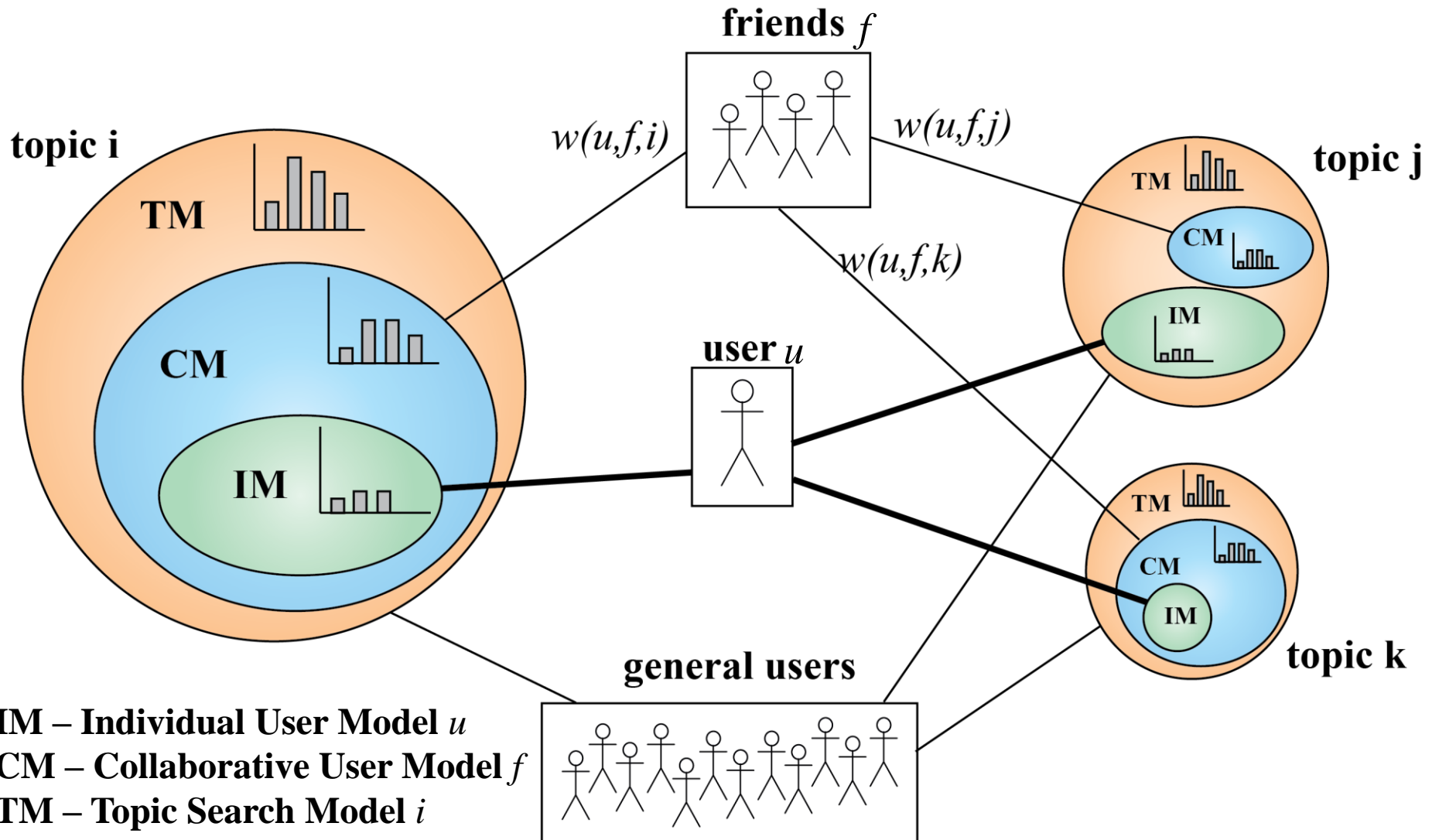
Issue C

Collaborative Personalized Microblog Search

Personalization challenge



Collaborative User Model



Bring Home Messages

- Microblog search → Searching on Big Data
- **A**nalysis of Content Quality
- **B**road Data Semantics Modeling
 - Multi-faceted Topic Modeling
 - Geographic Topic Modeling
- **C**ollaborative Personalized Microblog Search

Big Thanks!