On a quest to identify online conversations

Davide Vega

davide.vega@it.uu.se

@dvladek @uuinfolab

SOME MOTIVATION AND BACKGROUND

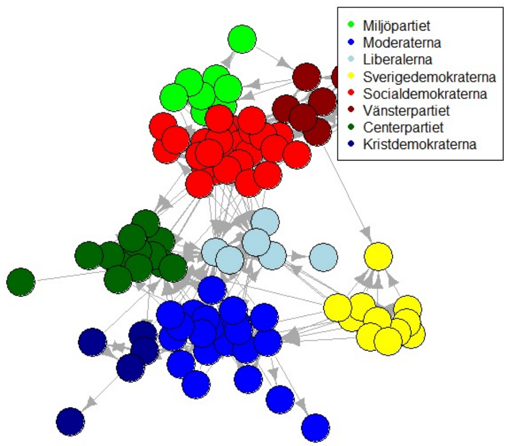
What is a conversation?

Image: https://www.peoplematters.in/blog/watercooler/fool-proof-conversation-starters-to-use-at-work-events-14755

What is NOT an online conversation?

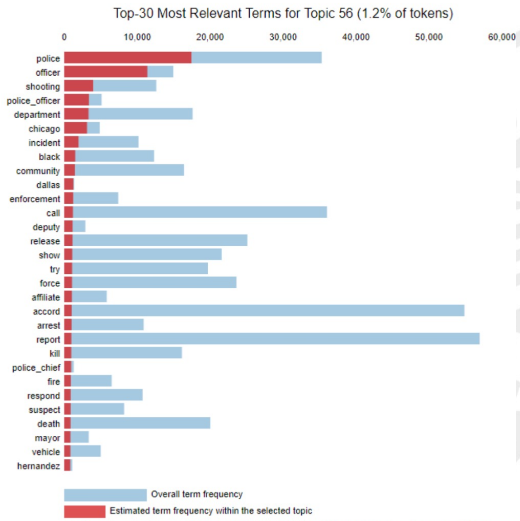
- Network Science perspective
  - Information cascade / diffusion
  - Threads in an online social network
  - A community structure (clustering)

Example Fig. RT network. Swedish political parties after 2018 elections



- Language perspective
  - Polarization/sentiment (clustering)
  - Topics (clustering)

Example Fig. Tweets related with news agencies, topics about law enforcement



1. saliency(term w) = frequency(w) * [sum_t p(t | w) * log(p(t | w)/p(t))] for topics t; see Chuang et. al (2012)
2. relevance(term w | topic t) = λ * p(w | t) + (1 - λ) * p(w | t)/p(w); see Sievert & Shirley (2014)

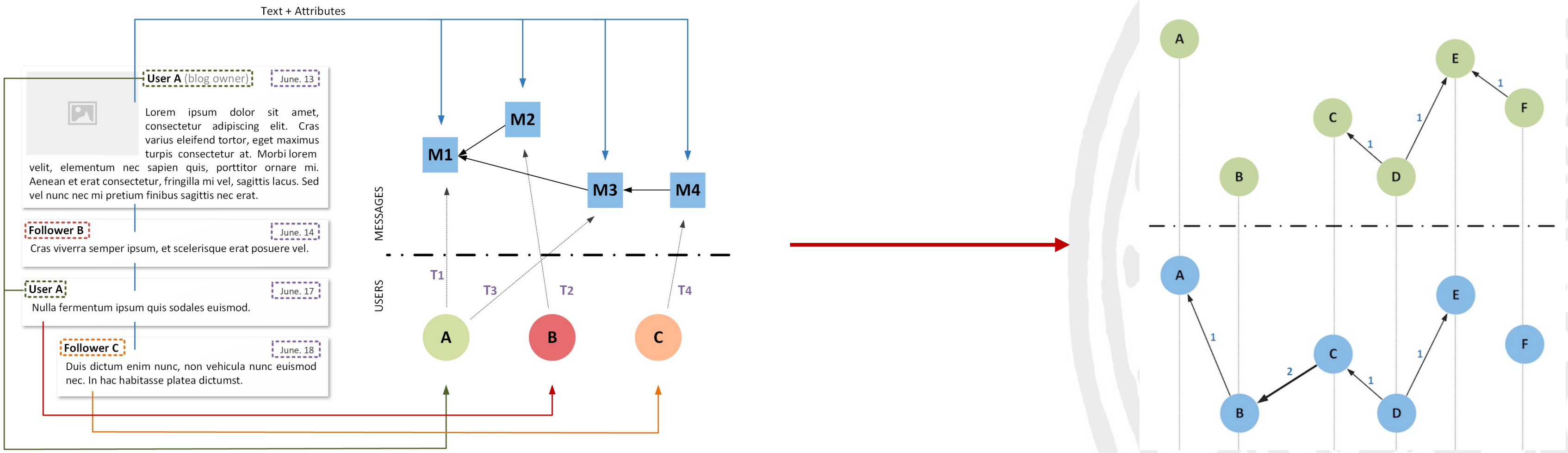
- **Some combination?**

SOME QUEST ATTEMPTS

Identifying conversations. Attempt 1

- Temporal Text Network data model
  - 2-mode network (actors / messages)
  - Original text and time preserved

- Discrete analysis
  - Create a k-multilayer network (Layers represent topics + time)
  - Community detection using a clique percolation mechanism



User A (blog owner) June. 13

Lorem ipsum dolor sit amet, consectetur adipiscing elit. Cras varius eleifend tortor, eget maximus turpis consectetur at. Morbi lorem velit, elementum nec sapien quis, porttitor ornare mi. Aenean et erat consectetur, fringilla mi vel, sagittis lacus. Sed vel nunc nec mi pretium finibus sagittis nec erat.

Follower B June. 14

Cras viverra semper ipsum, et scelerisque erat posuere vel.

User A June. 17

Nulla fermentum ipsum quis sodales euismod.

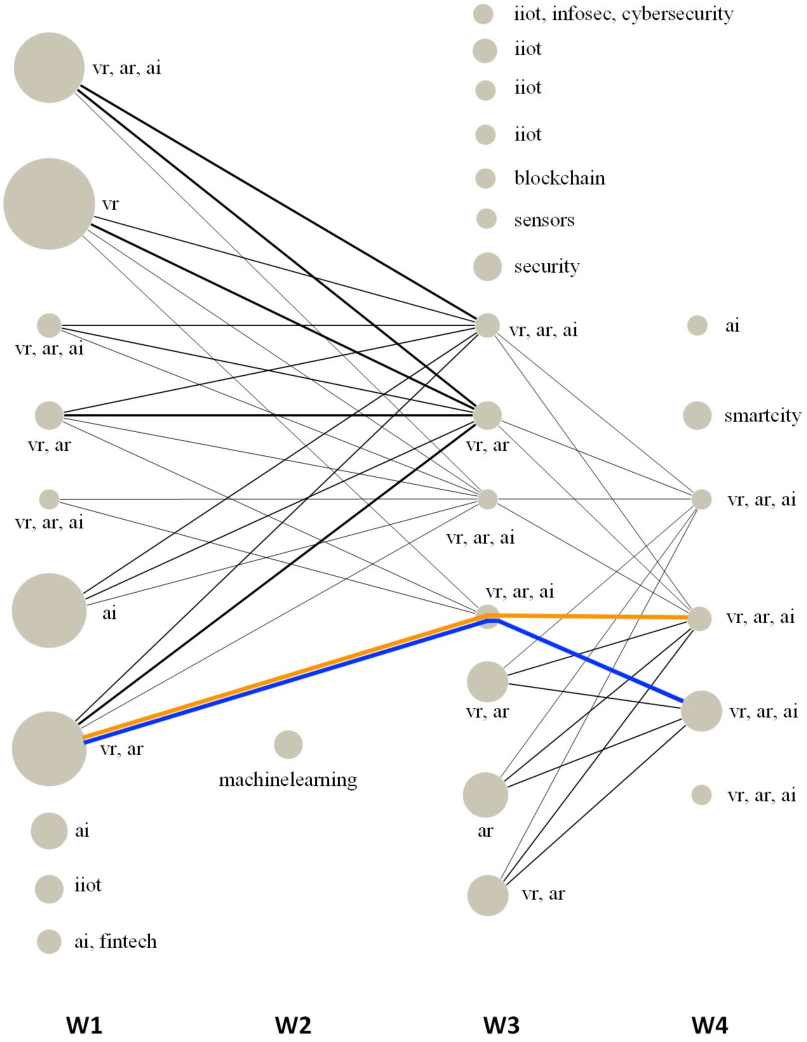
Follower C June. 18

Duis dictum enim nunc, non vehicula nunc euismod nec. In hac habitasse platea dictumst.

Identifying conversations. Attempt 1

- Pros:
  - Flexible
  - Interpretable
  - Time *coherent*

- Cons
  - Too many decisions
    - e.g., time-division granularity
  - Cliques do not imply conversation



Evolution of communities in the IoT space (Twitter dataset)

Identifying conversations. *Current* attempt

- Temporal Text Network data model
  - 2-mode network (actors / messages)
  - Original text and time preserved

- Rewrite the discrete analysis task as inference problem
  - Using the stochastic block model (SBM) as network prior:
  - Reconstruct the 2-mode network structure M, constrained to
    - b_i^A group membership of actor (node) i
    - b_i^M group membership of message (node) i
    - λ_{rs}^t edge probability from group r to group s after time t
    - ζ_{rq} message (node) probability from group r to topic q

- Current problem:
  - Too many parameters -> danger of overfitting
  - No clear choice of t

Open questions

- Still unclear what online conversations actually are
  - Does the quest make sense?
  - Is a definition / modeling / measuring problem?

- Can the inference task be solved?
  - If not, what should be the strategy?

On a quest to identify online conversations

Davide Vega

davide.vega@it.uu.se

@dvladek @uuinfolab