



# Information Seeking and Responding Networks in Physical Gatherings: A Case Study of Academic Conferences in Twitter

Xidao Wen and Yu-Ru Lin  
University of Pittsburgh



**Jonathan Deber** @jdeber · Apr 28

3D printed teddy bears with a felt printer. They're like normal teddy bears, but CADier. And cooler. #cadierisawordright? #chi2014



**ITS 2014** @ITS\_2014 · Apr 28

@hilfeheiko Sounds interesting. Care to comment or expound? #chi2014



[View conversation](#)



**Wendy Moncur** @wendymoncur · Apr 28

Fitz at #chi2014 - regular FitBit users often put their devices through the wash & have to buy new ones. Washable wearables the next step?



**Tia Shelley** @TiaTalksHCI · Apr 28

Compatibility versus natural interactions? How much does your interaction resemble the action you want to do. #chi2014



**Ben Kirman** @benki · Apr 28

The #chi2014 robot has a human slave following it around. Did they not learn from the robots paper last year? [dl.acm.org/citation.cfm?i...](http://dl.acm.org/citation.cfm?i...)



**CHI2014** @chi2014 · Apr 28

@barneyc excellent question about #CHI2014 proceedings and mobile devices. Anyone out there have a clever way to solve that?



[View conversation](#)



**Tia Shelley** @TiaTalksHCI · Apr 28

Is there a substantial difference between immersion and flow? Is flow just the max attainable level of immersion. #chi2014



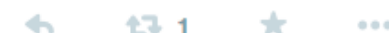
**Dr Simone Stumpf** @DrSimoneStumpf · Apr 28

"What does this mean for HCI?" Theoretical understanding of Making culture is all fine but maybe also look to supporting practice? #chi2014



**Tia Shelley** @TiaTalksHCI · Apr 28

When moving to a mobile device, how do controls change experience of gameplay? #chi2014



# Challenges

- Participants may seek and disseminate information while rushing to keep up with the event.
- Participants often find themselves exposed to new places, new people, and new topics.

# Research Questions

- RQ1: How do we meaningfully capture the distinctive information needs on Twitter during academic conferences?

# Research Questions

- RQ2: Can communications of the information needs be inferred from users' prior tweeting patterns and network positions?
- Who are the users that tend to post certain types of information needs?
- Can we identify potential responders to these information needs?

# Data Collection

- 190k tweets posted in 66 CS conferences over five years (2009-2013).
- ~12k information-seeking tweets (tweets that contain at least one “?” [Paul 2011]).

# Information Seeking Annotation & Labels

- We developed a category scheme (*Fleiss Kappa* = 0.58) of the information seeking tweets based on previous studies [Morris 2011 & Efron 2010].
- We obtain annotations of the type labels for ~2k information-seeking tweets through Amazon Mechanical Turk.
- We trained n-gram multi-class SVM classifiers (*accuracy* = .721) to automatically label the rest of the information-seeking corpus.

**Coordinate  
Events/Ask  
Favors**

**3.2%**

*Getting set to head to LA for #siggraph2012 with a stable of #makerbot replicators. Anyone wanna join me there?*

**Express  
Opinions**

**19.7%**

*"Since we couldn't do either qual or quant research, we decided on mixed methods." huh?! #chi2012*

**Promote  
Information**

**32.7%**

*Can't make it to #FAST13 this week? Consider attending via live stream instead: <https://t.co/mNyA4lGe>*

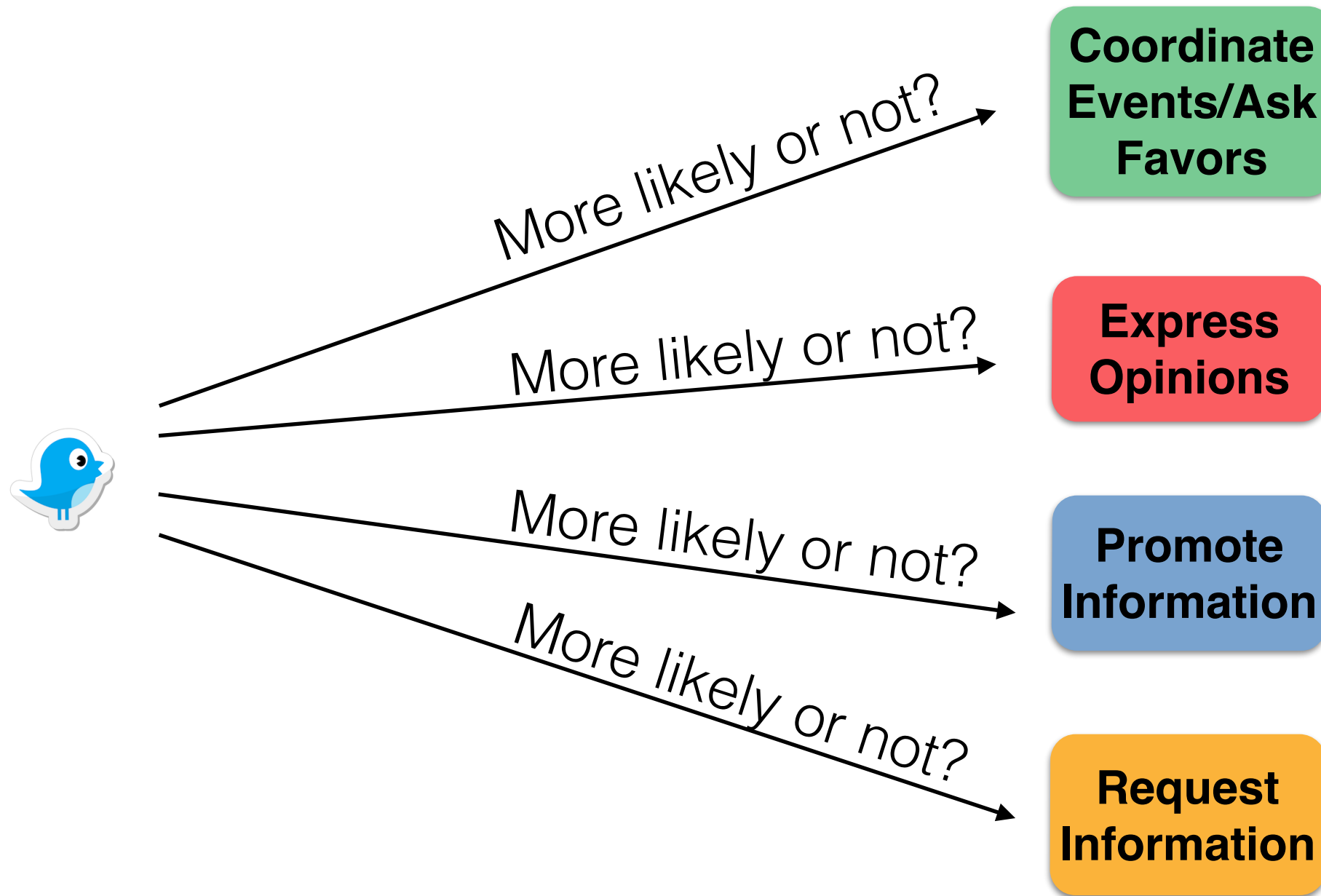
**Request  
Information**

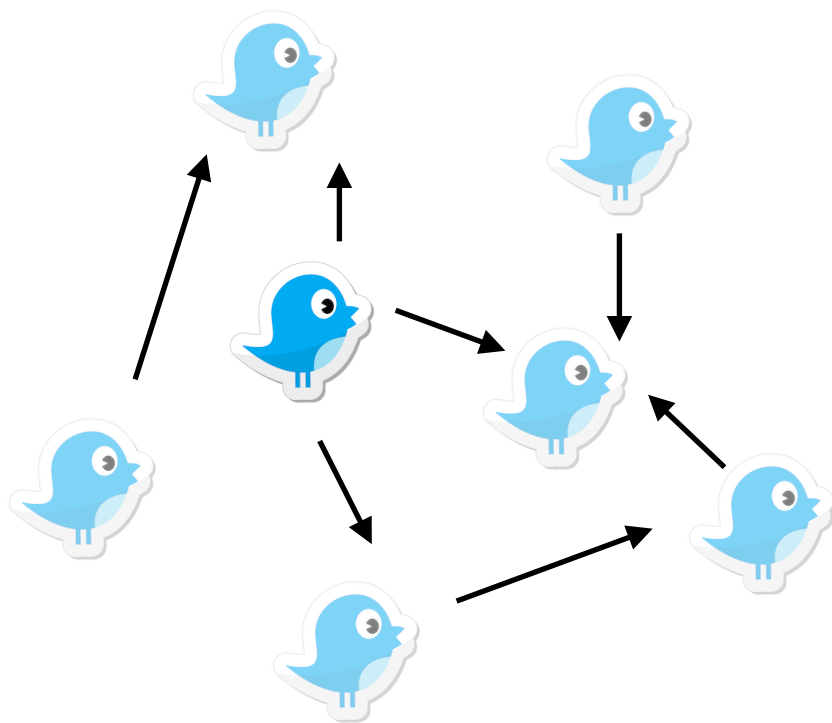
**44.4%**

*I wish I was at #CHI2010. I wish I got to listen to Genevieve Bell. Does anyone know if I can listen to the keynote online?*

RQ2A:

Who are the users that tend to post certain types of information needs?



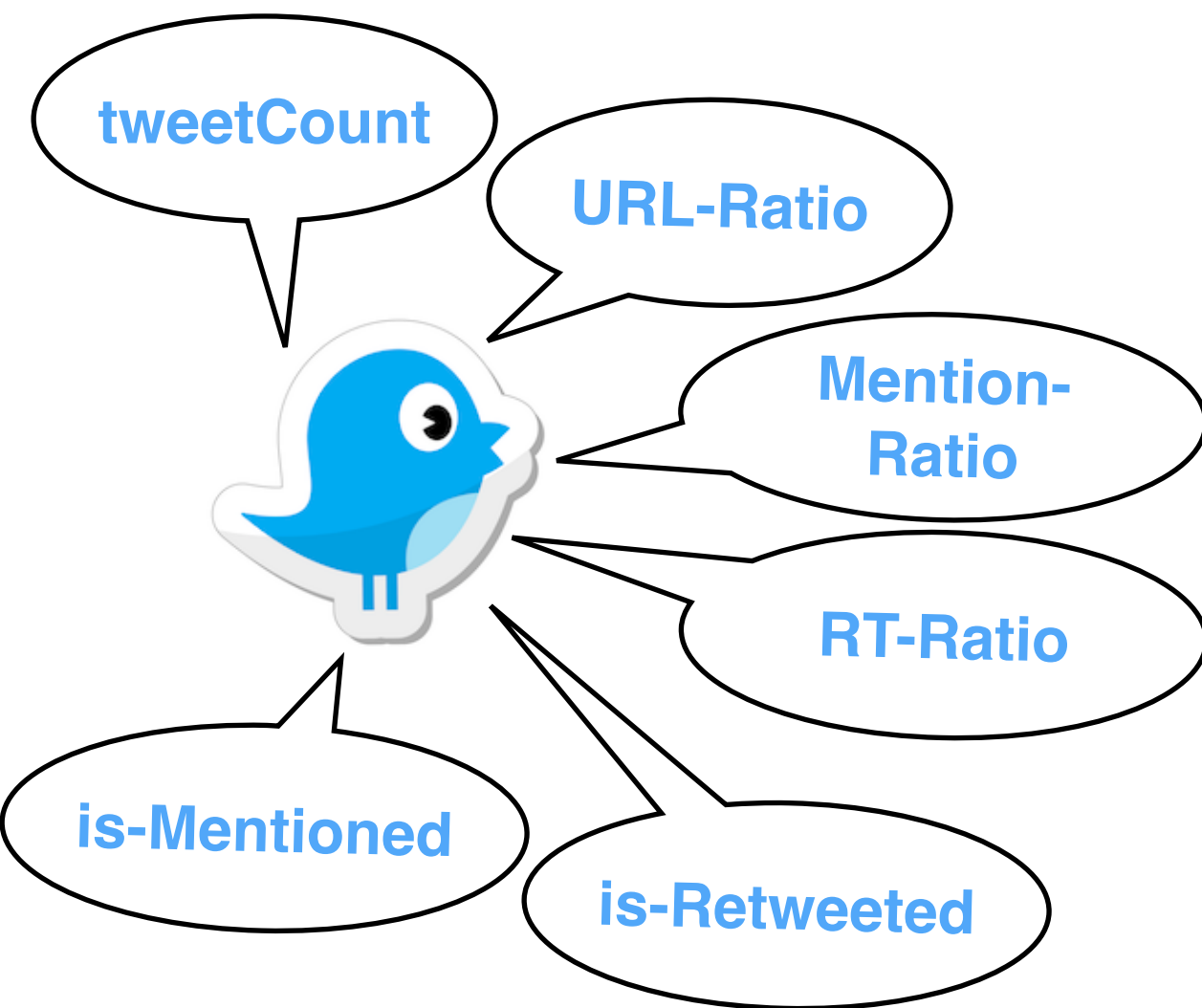


**Coordinate  
Events/Ask  
Favors**

**Express  
Opinions**

**Promote  
Information**

**Request  
Information**

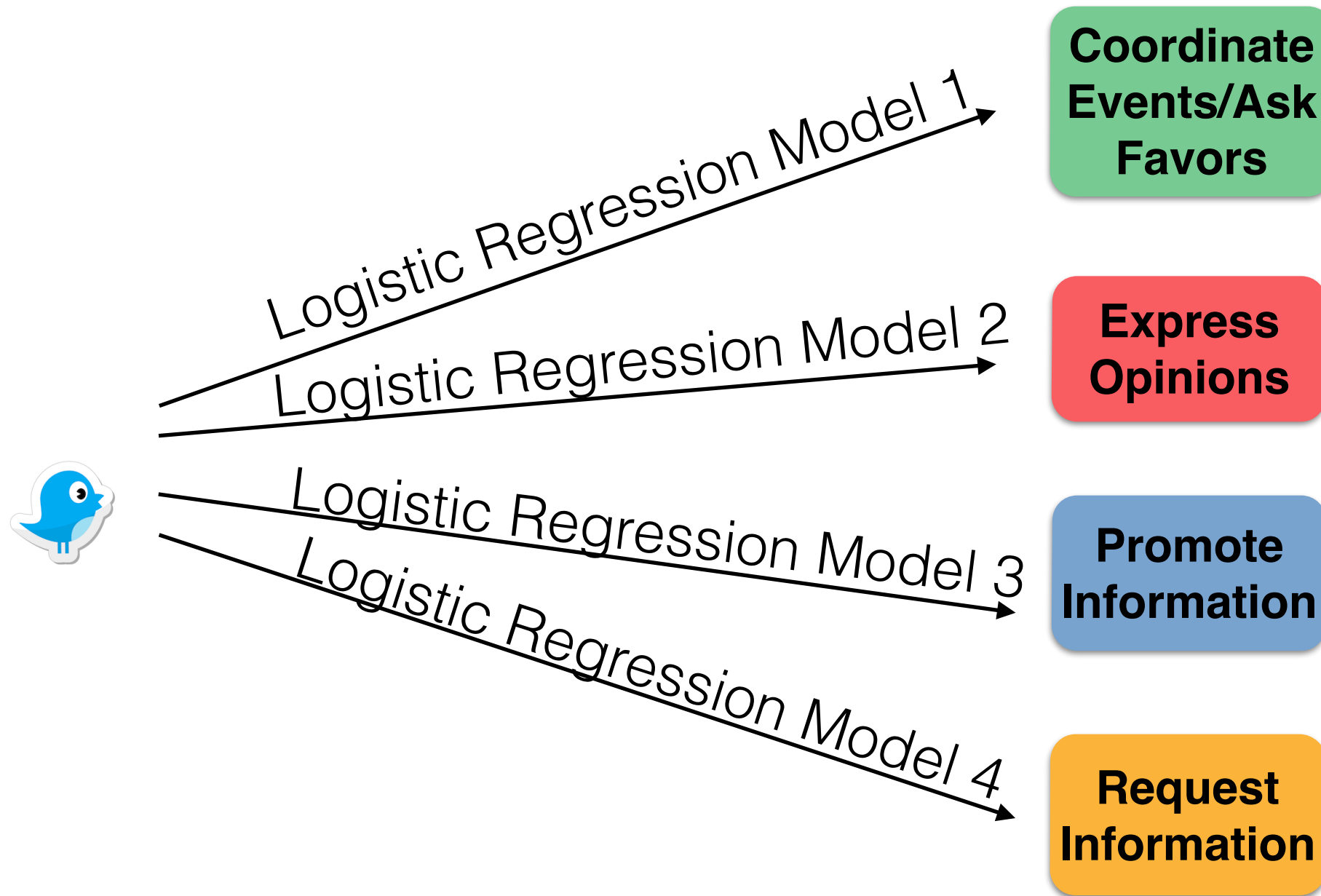


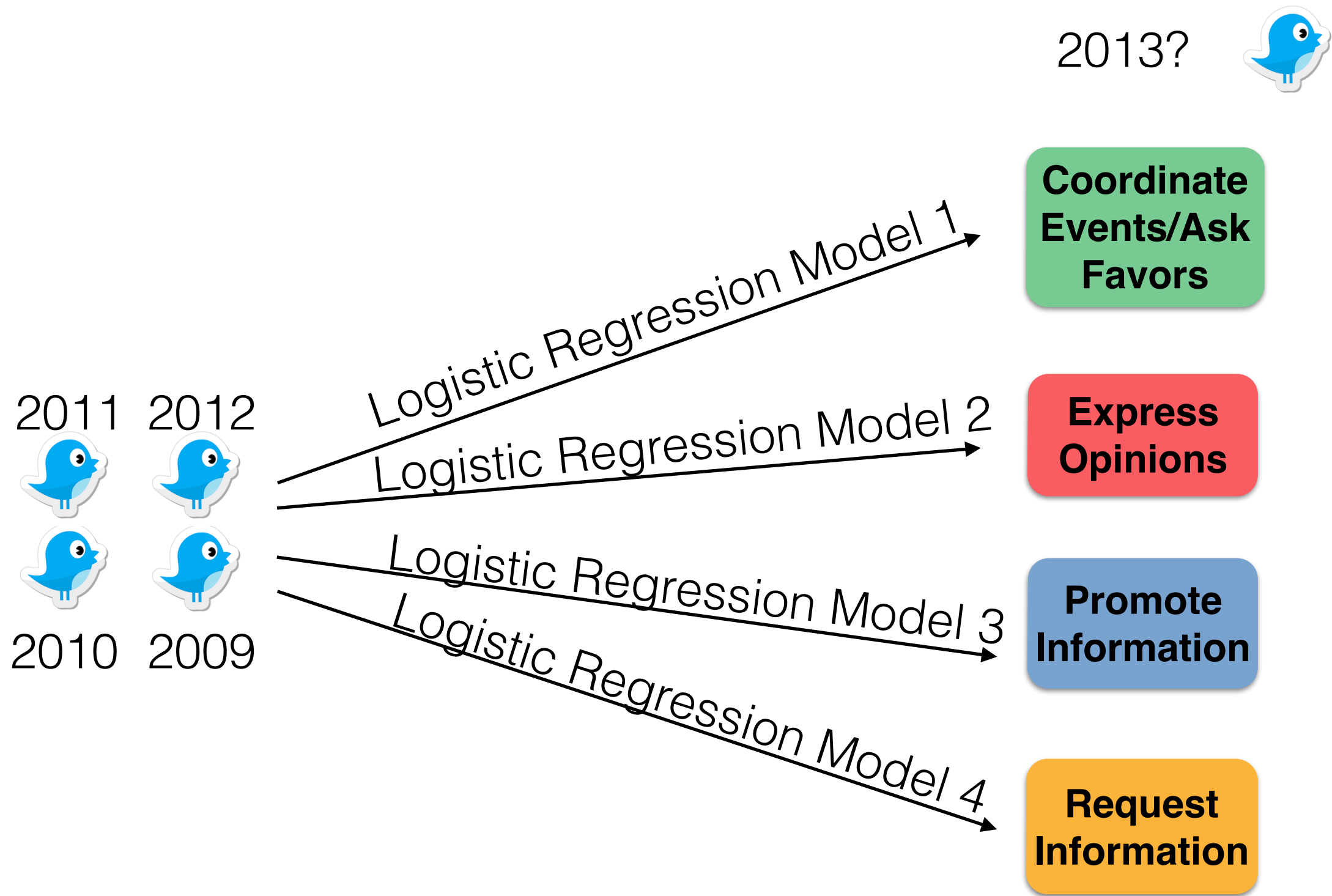
**Coordinate  
Events/Ask  
Favors**

**Express  
Opinions**

**Promote  
Information**

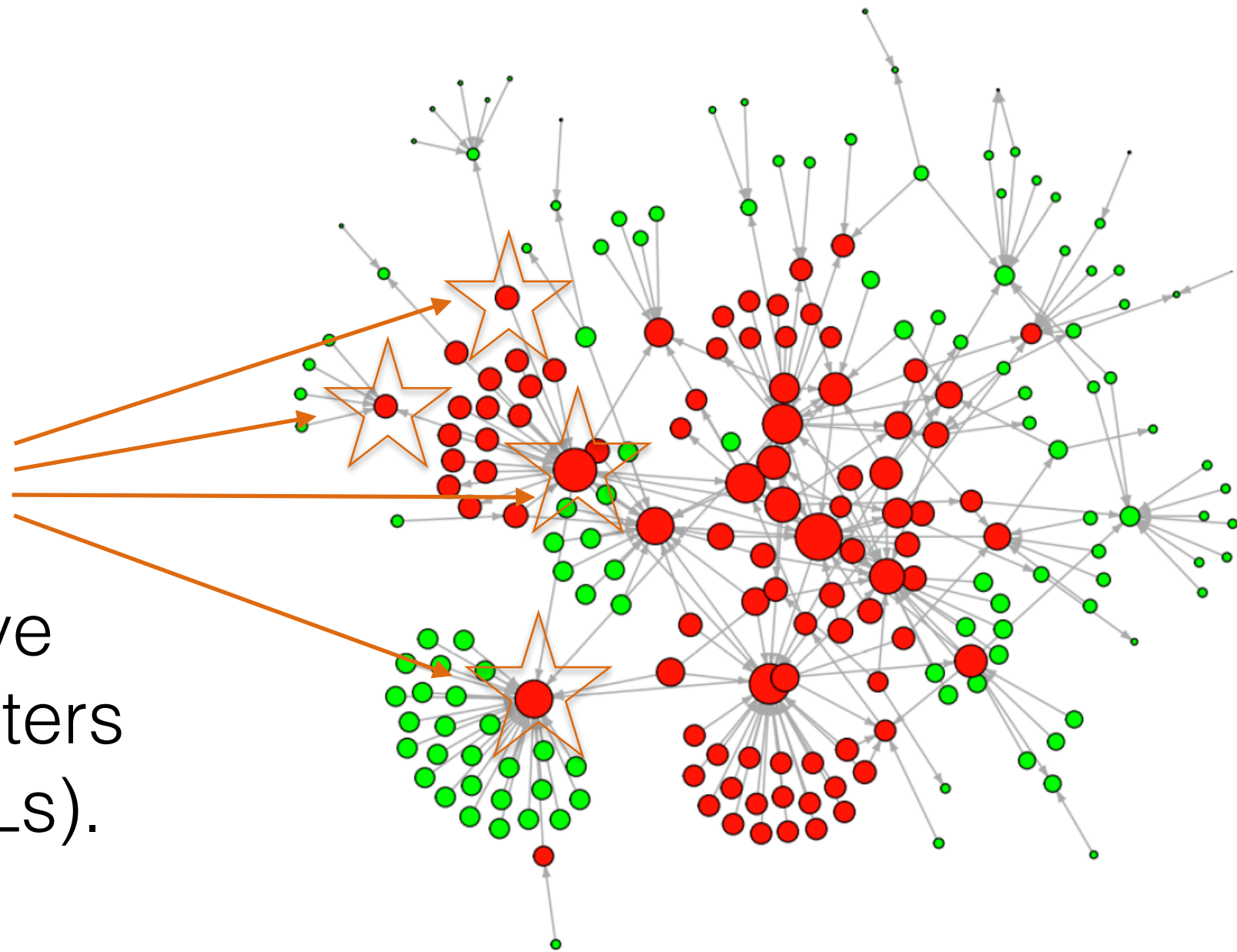
**Request  
Information**





## Coordinators

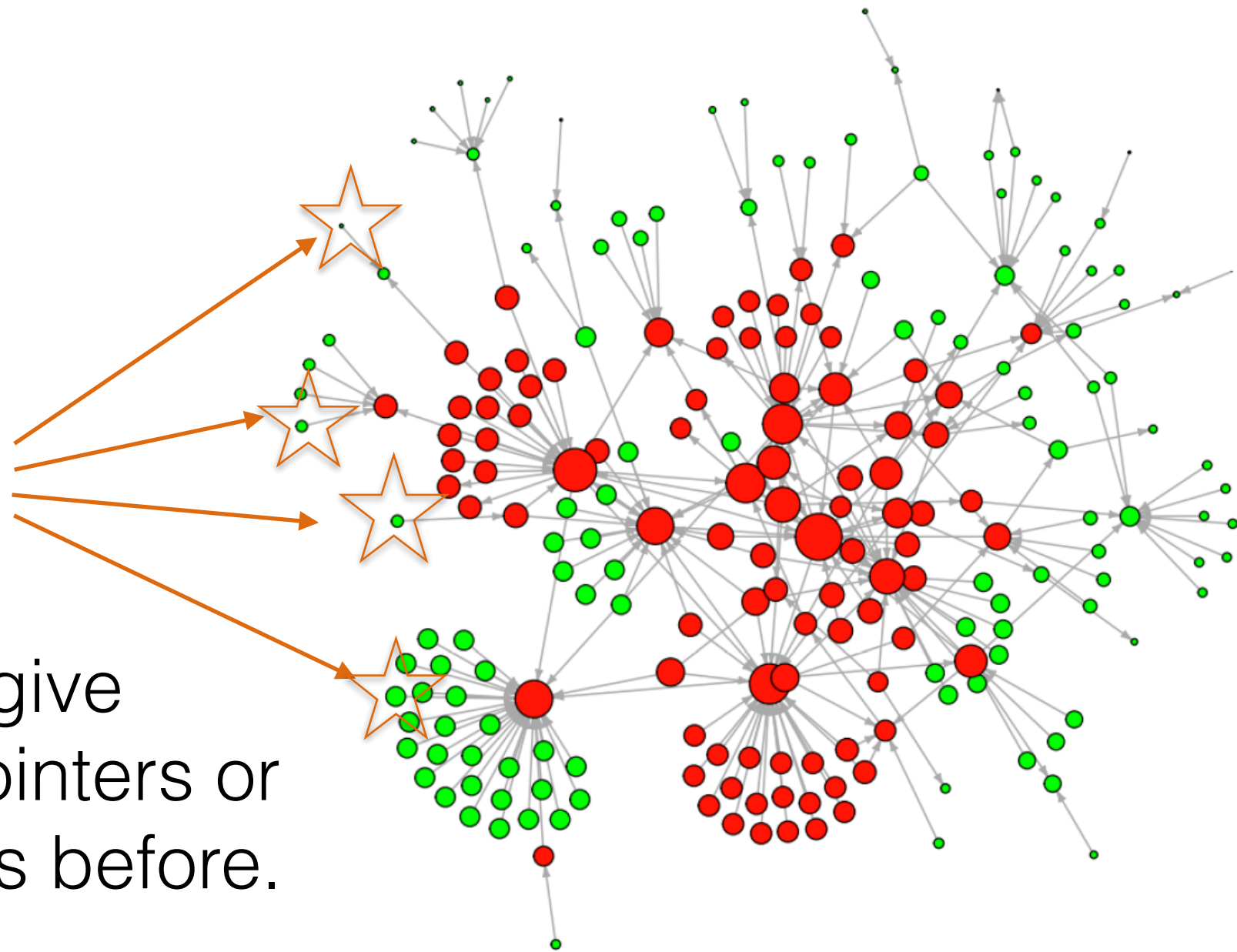
Less likely to give  
information pointers  
(via posting URLs).



SIGKDD retweet network (Size of  
the nodes are proportional to the closeness  
centrality)

**Requestors**

Less likely to give  
information pointers or  
mention others before.



SIGKDD retweet network (Size of  
the nodes are proportional to the closeness  
centrality)

## Promoters

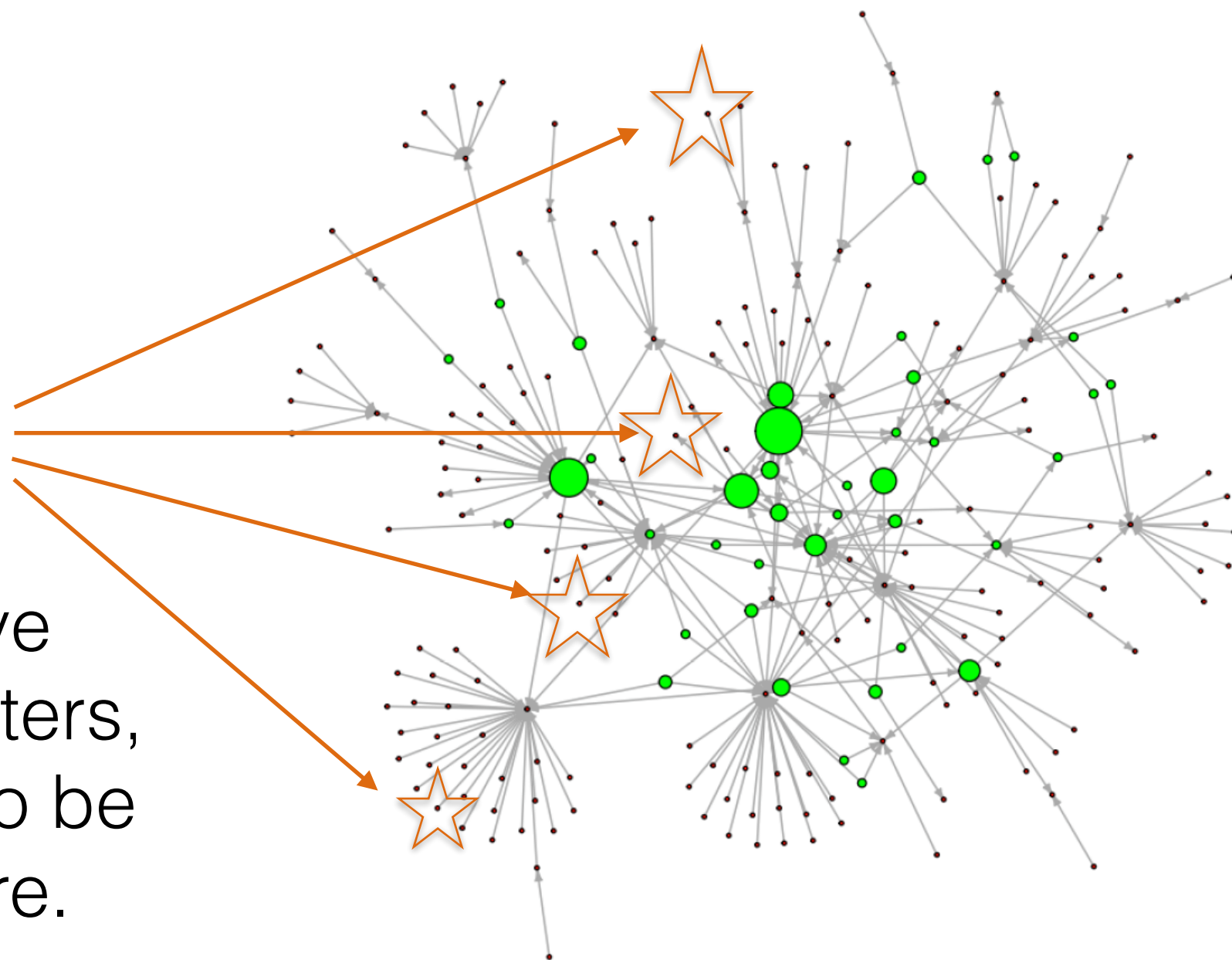
More likely to give  
information pointers,  
be mentioned, or be  
retweeted.



SIGKDD retweet network (Size  
of the nodes are proportional to the out-  
degrees)

## Opinionators

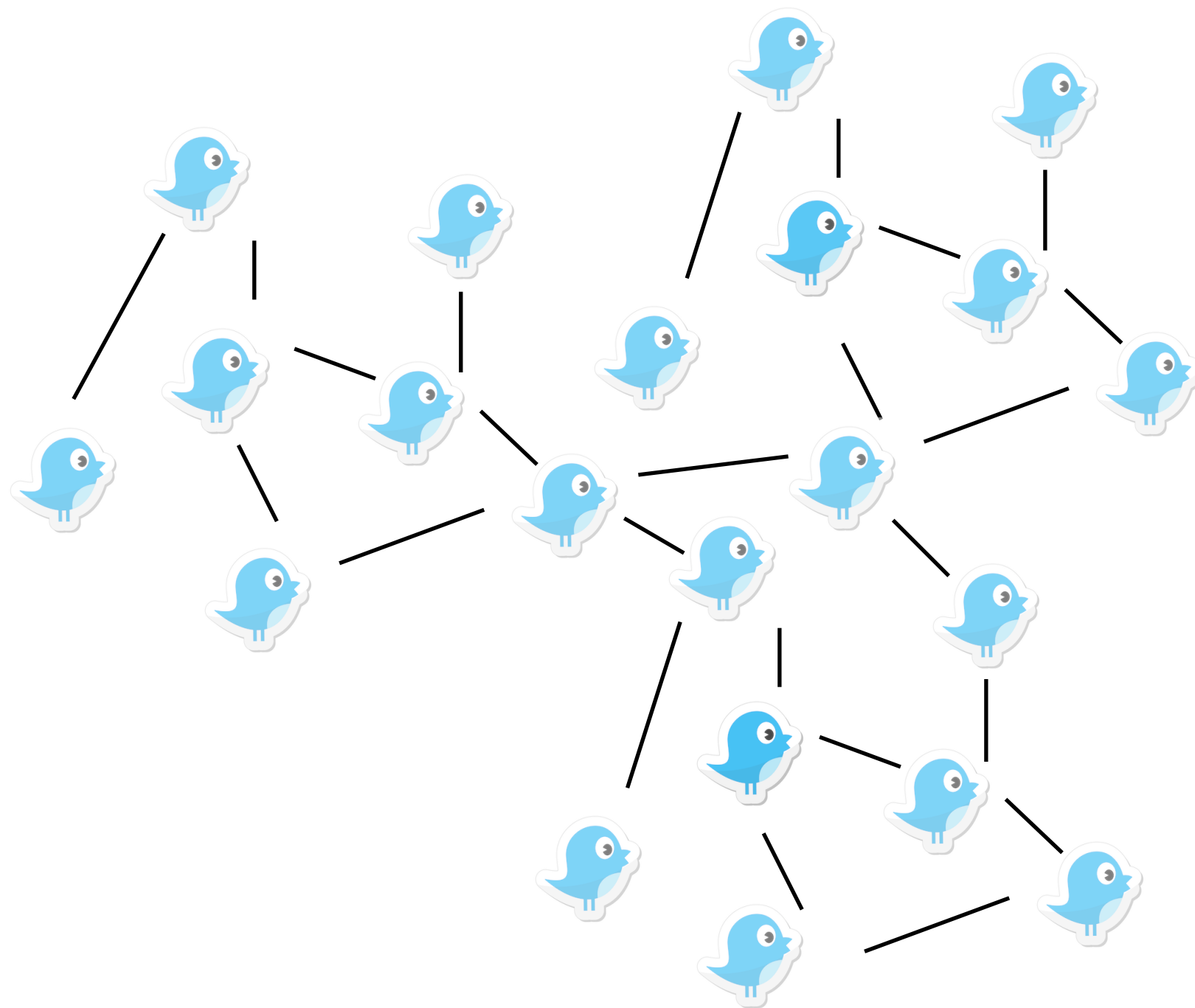
Less likely to give information pointers, but more likely to be mentioned before.

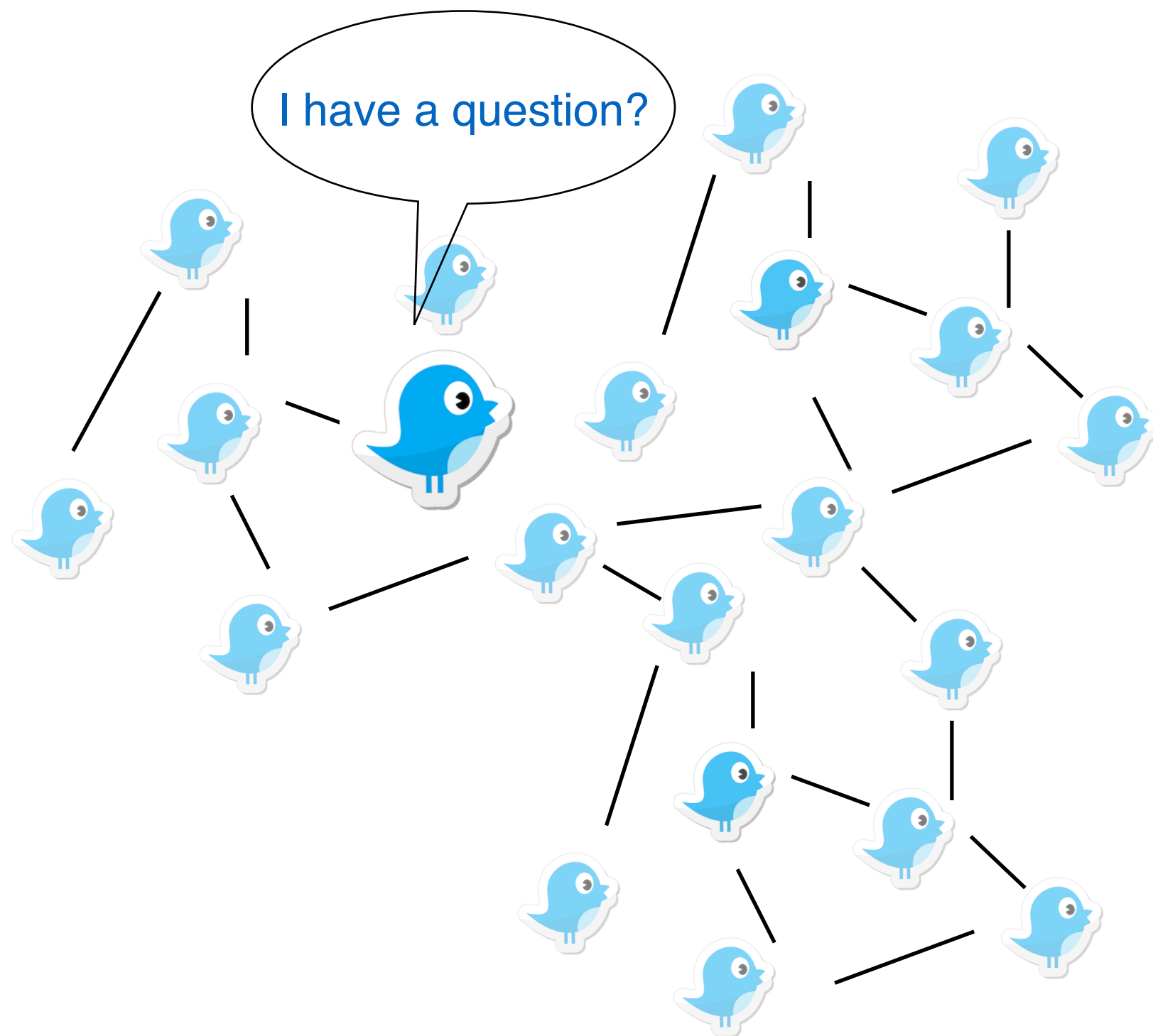


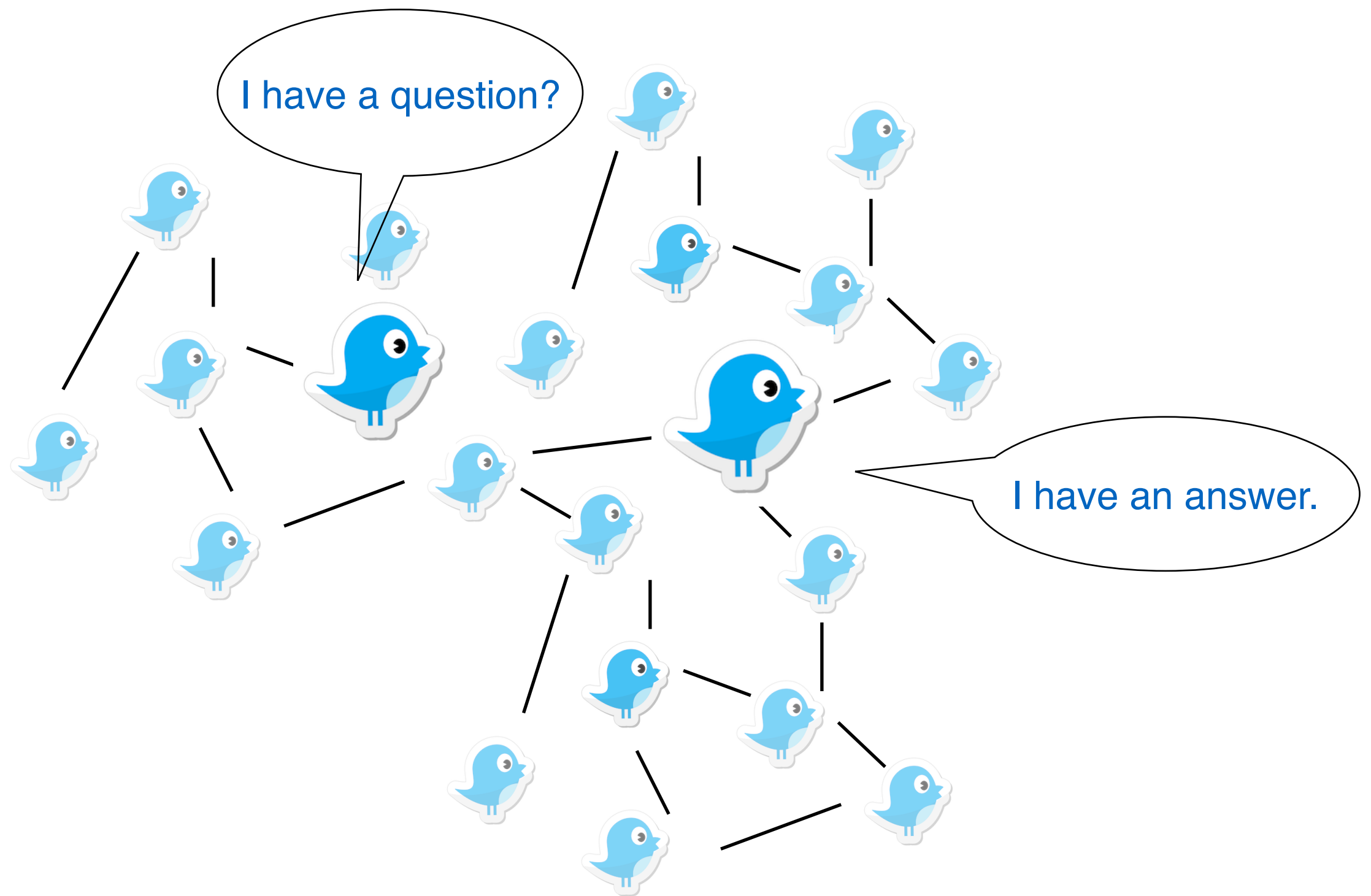
SIGKDD retweet network (Size of the nodes are proportional to the out-degrees)

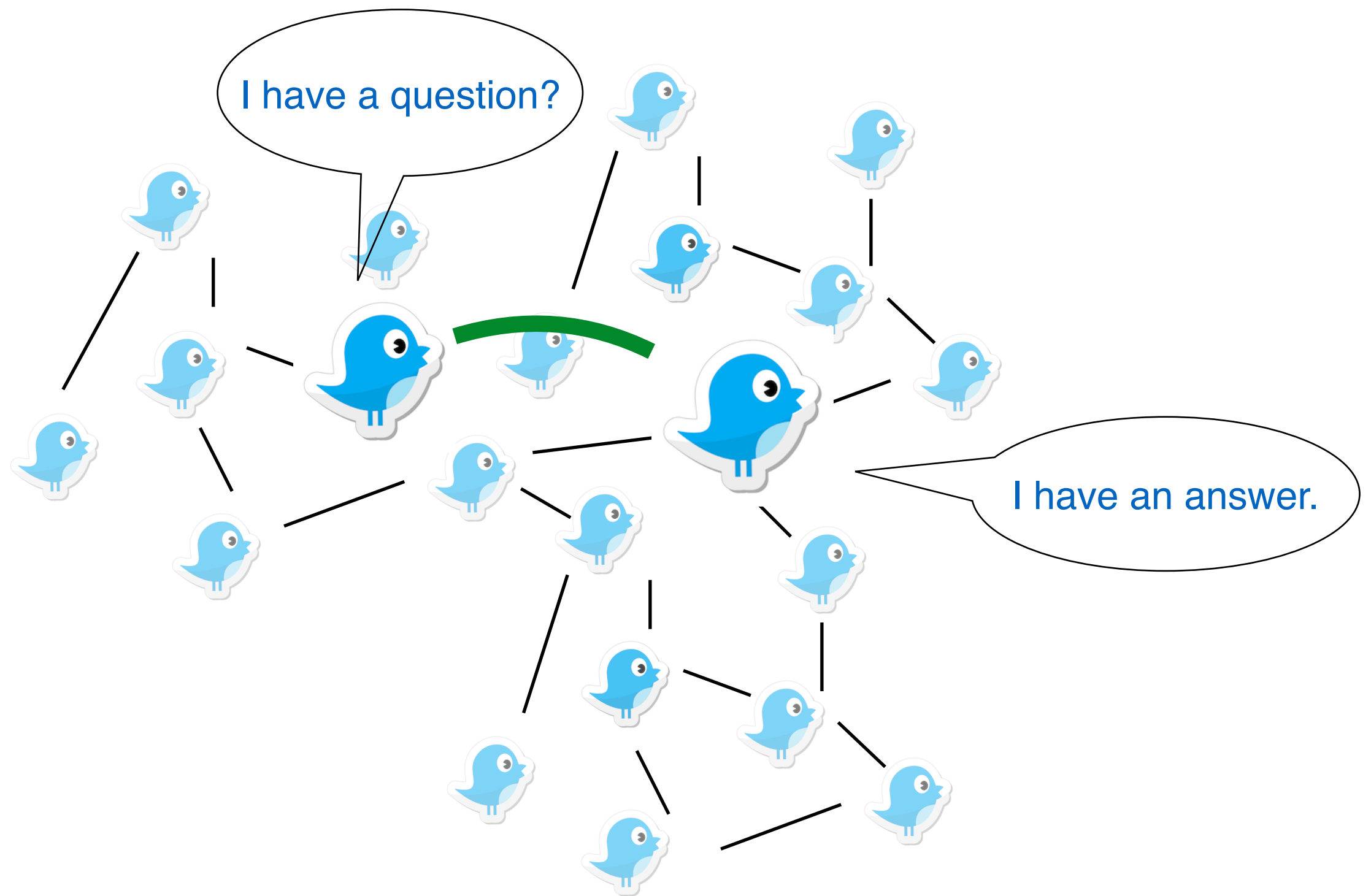
RQ2B:

Can we identify potential responders  
to the information needs?

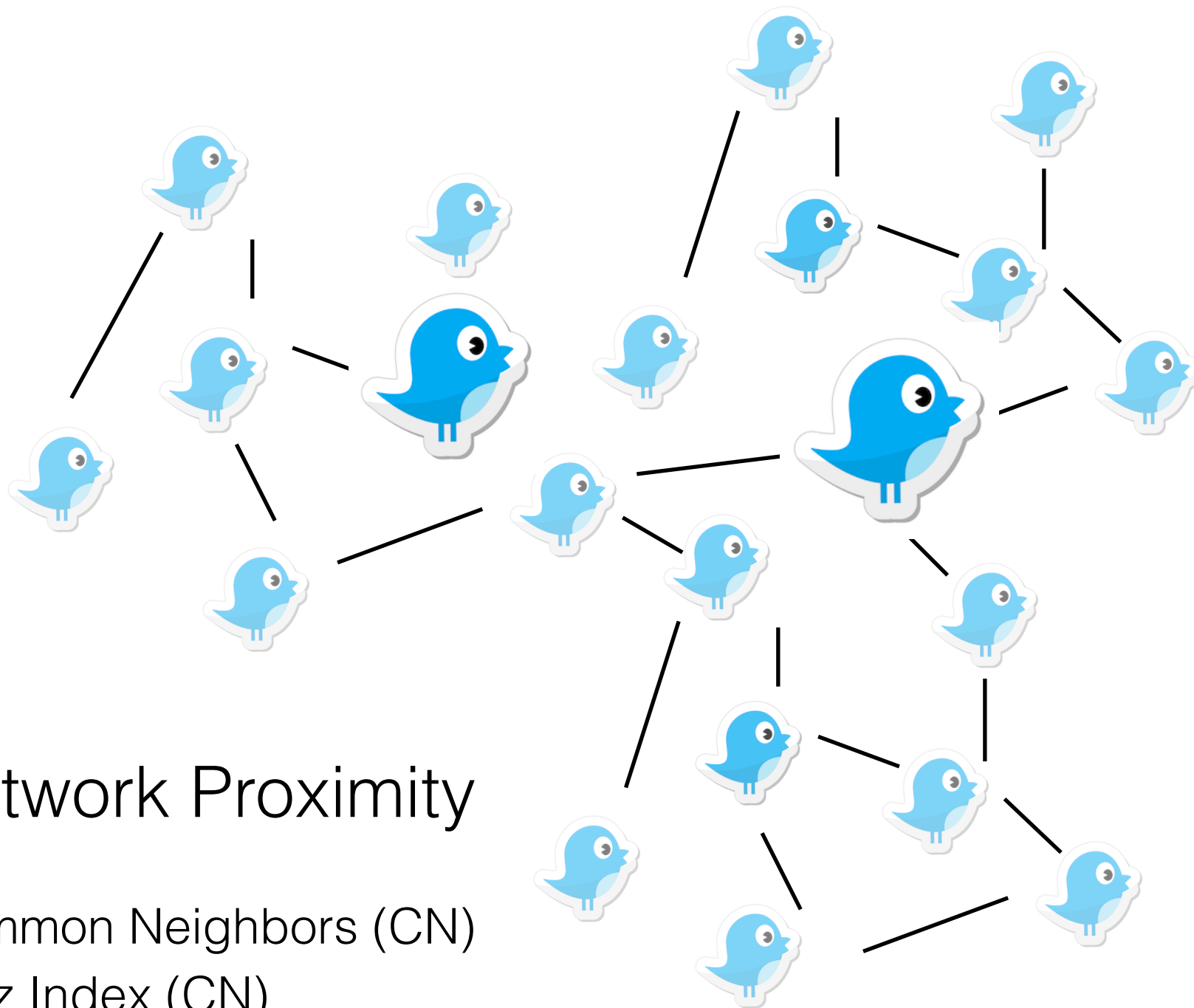








# Link Prediction Task



## Network Proximity

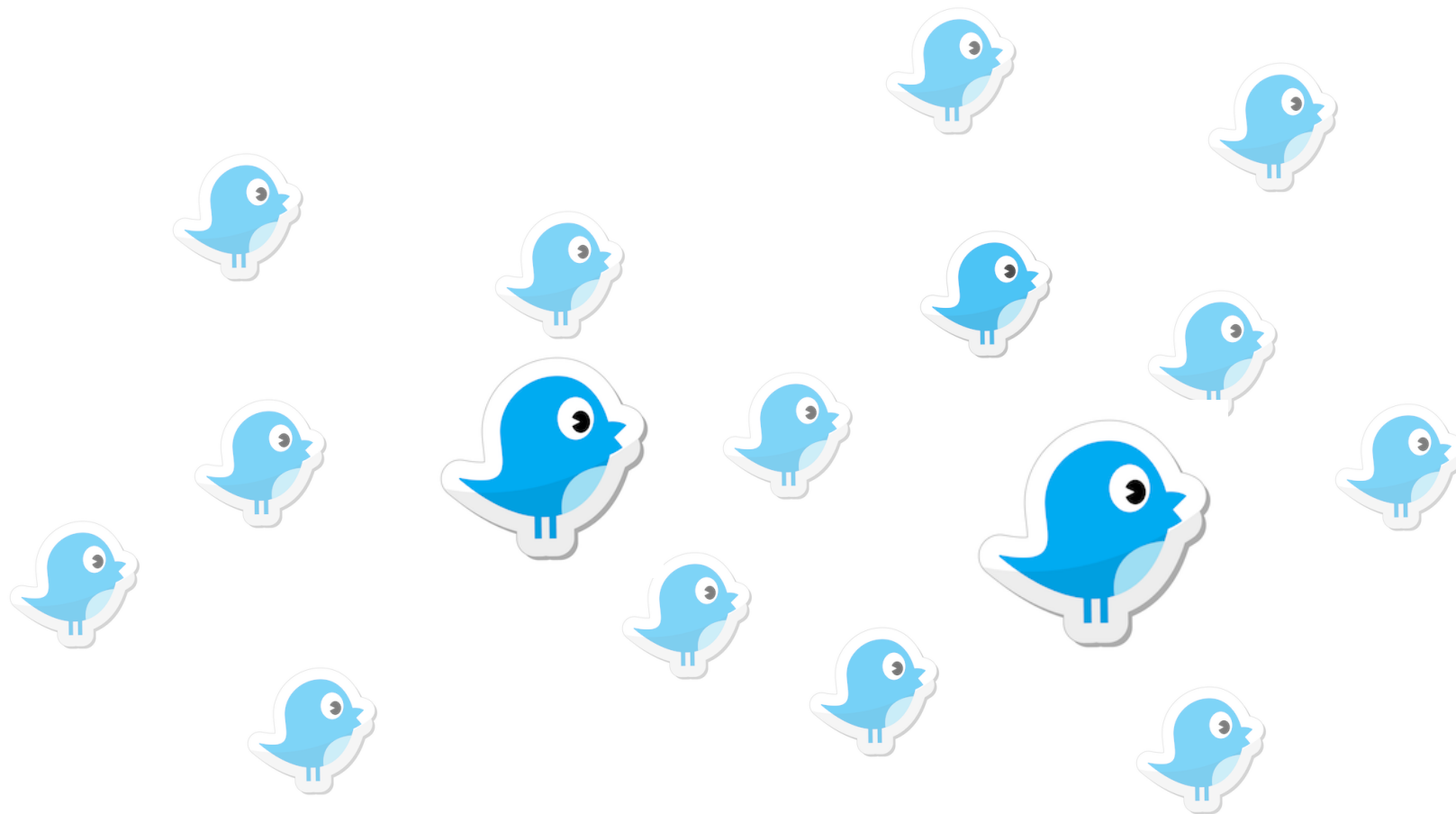
Common Neighbors (CN)

Katz Index (CN)

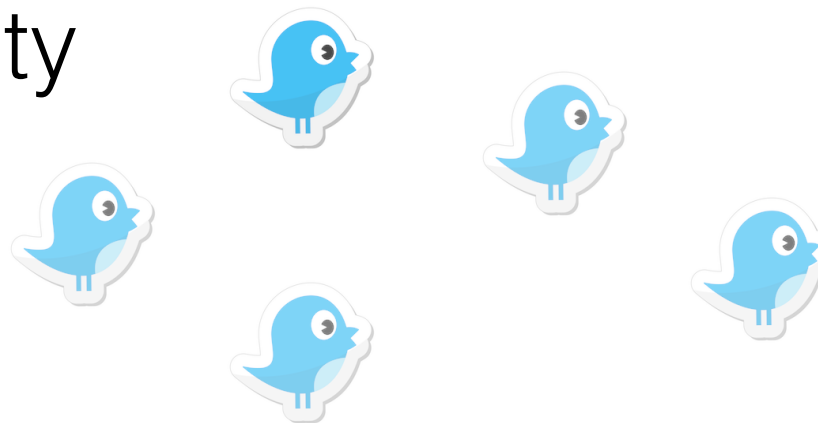
Jaccard Index (JI)

Preferential Attachment (PA)

# Link Prediction Task

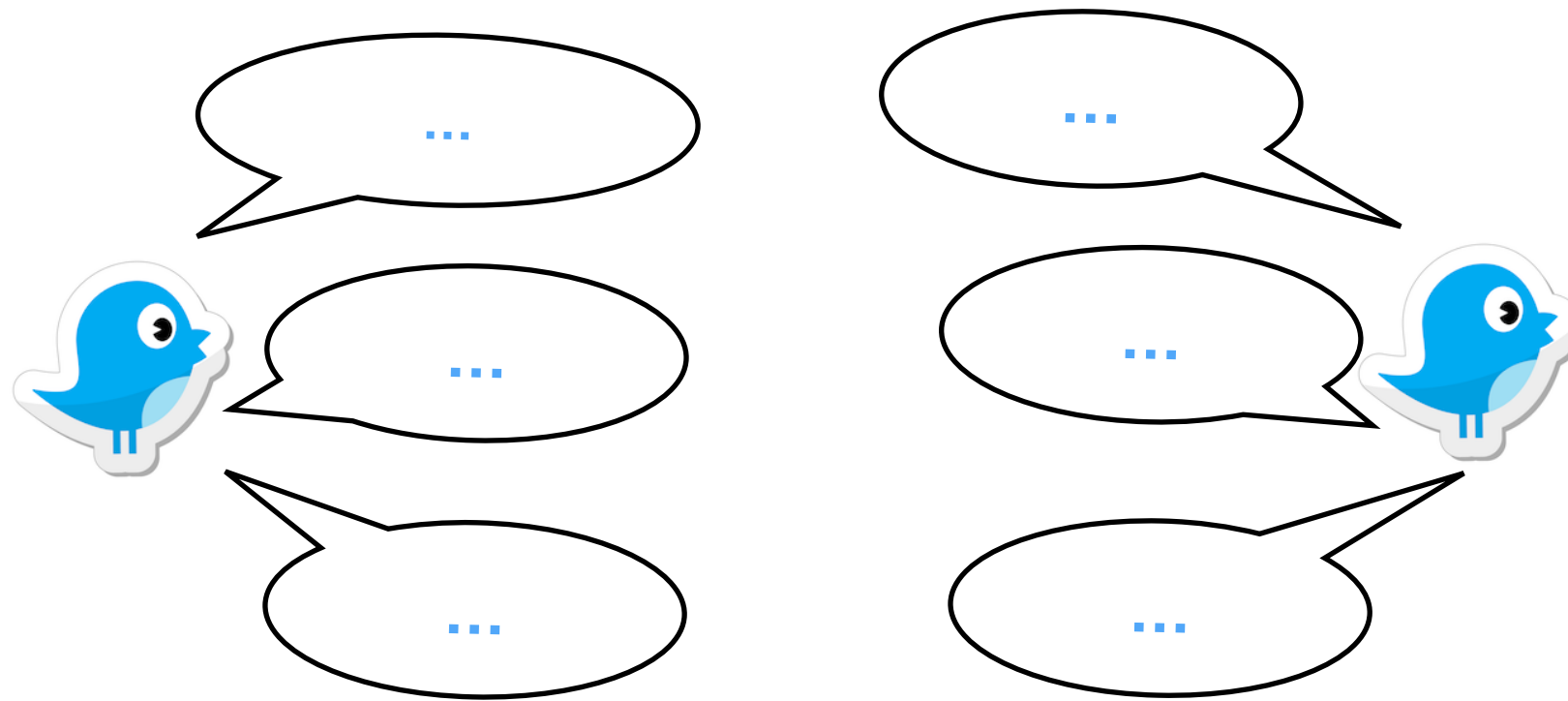


Content-based Similarity



# Link Prediction Task

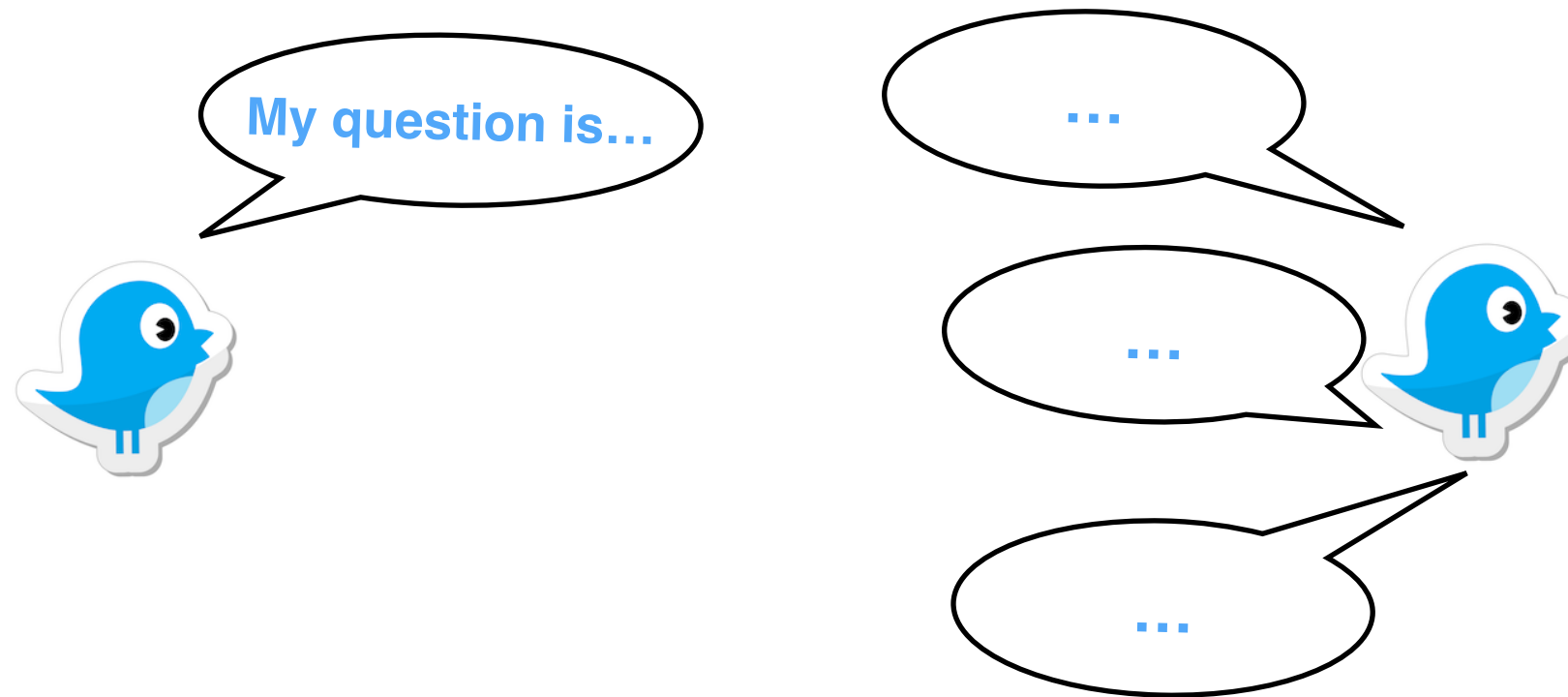
## Content-based Similarity



- User Similarity
  - Similarity between users' tweets in the past.

# Link Prediction Task

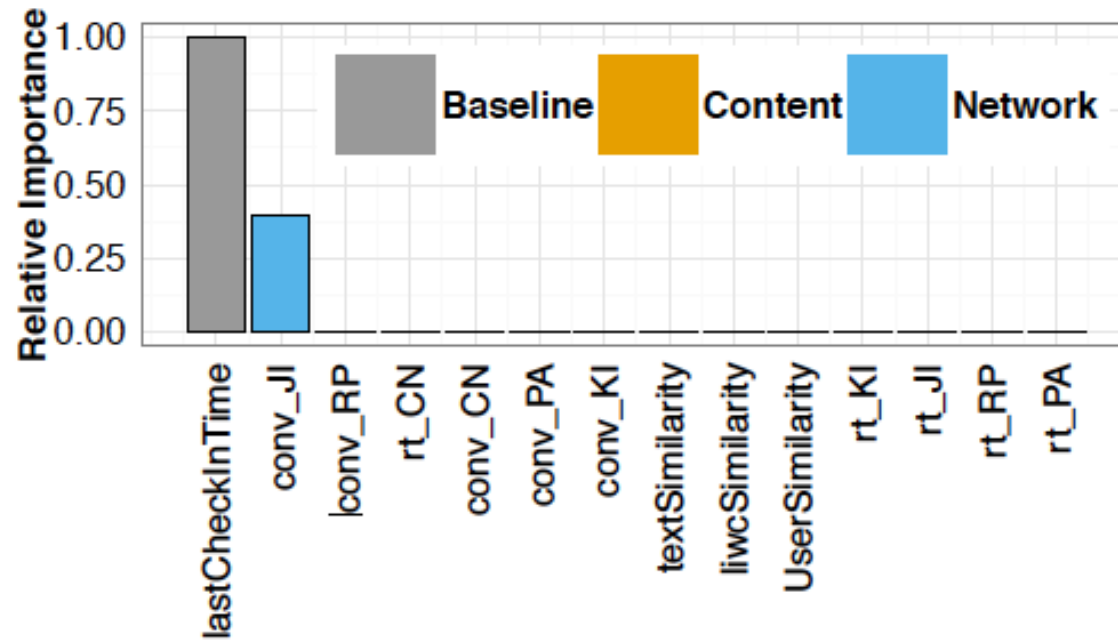
## Content-based Similarity



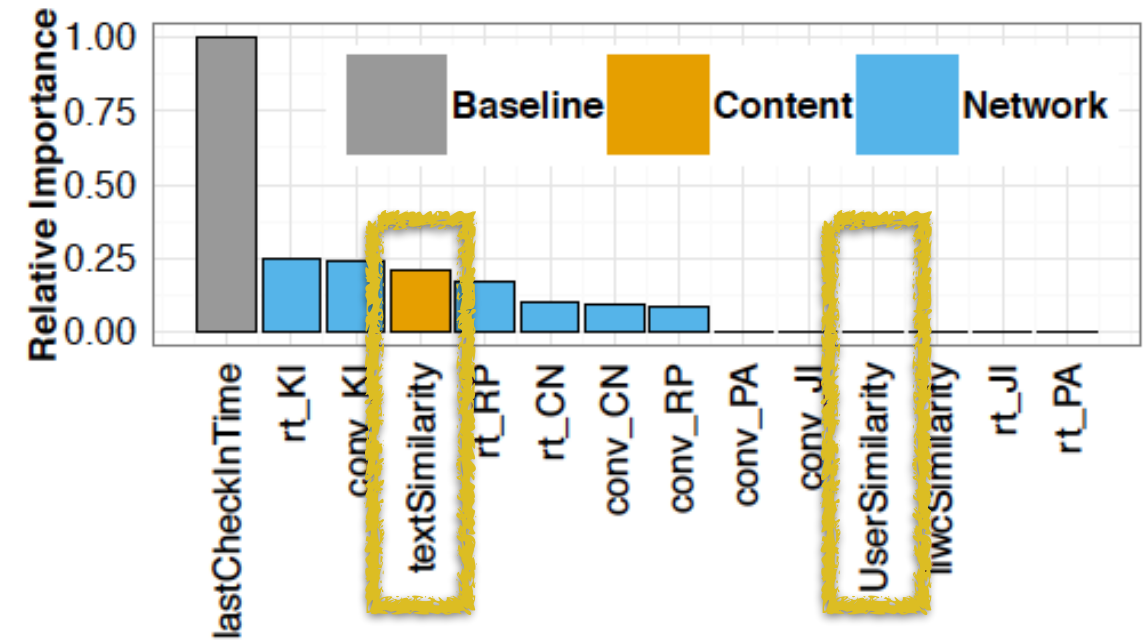
- Text Similarity
  - Similarity between question and users' tweets in the past.

- Baseline
  - The time difference between the question's posted time and the potential respondent's last seen activity.
- Network Proximity
- Content Similarity
  - User Similarity, Text Similarity, LIWC Similarity
- Balanced dataset

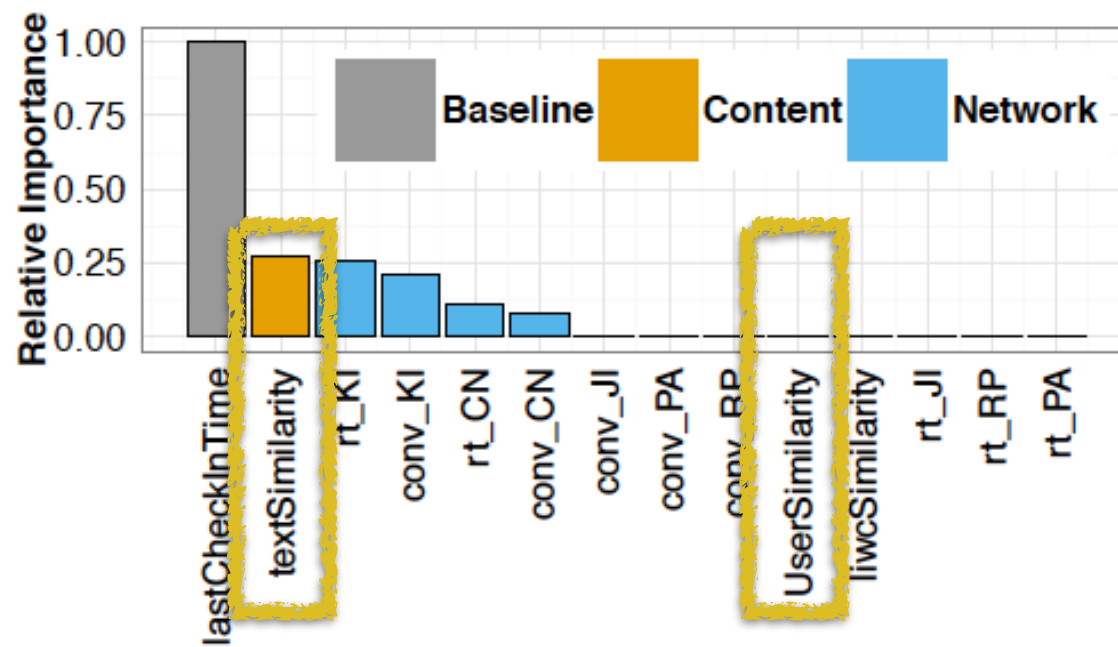
| Feature Sets     | RF    | ADA          | Bagging      | SVM   |
|------------------|-------|--------------|--------------|-------|
| Baseline         | 0.73  | <b>0.774</b> | 0.77         | 0.669 |
| Baseline+Network | 0.824 | 0.774        | <b>0.834</b> | 0.689 |
| Baseline+Content | 0.784 | 0.783        | <b>0.81</b>  | 0.688 |
| Combined         | 0.843 | 0.783        | <b>0.859</b> | 0.714 |



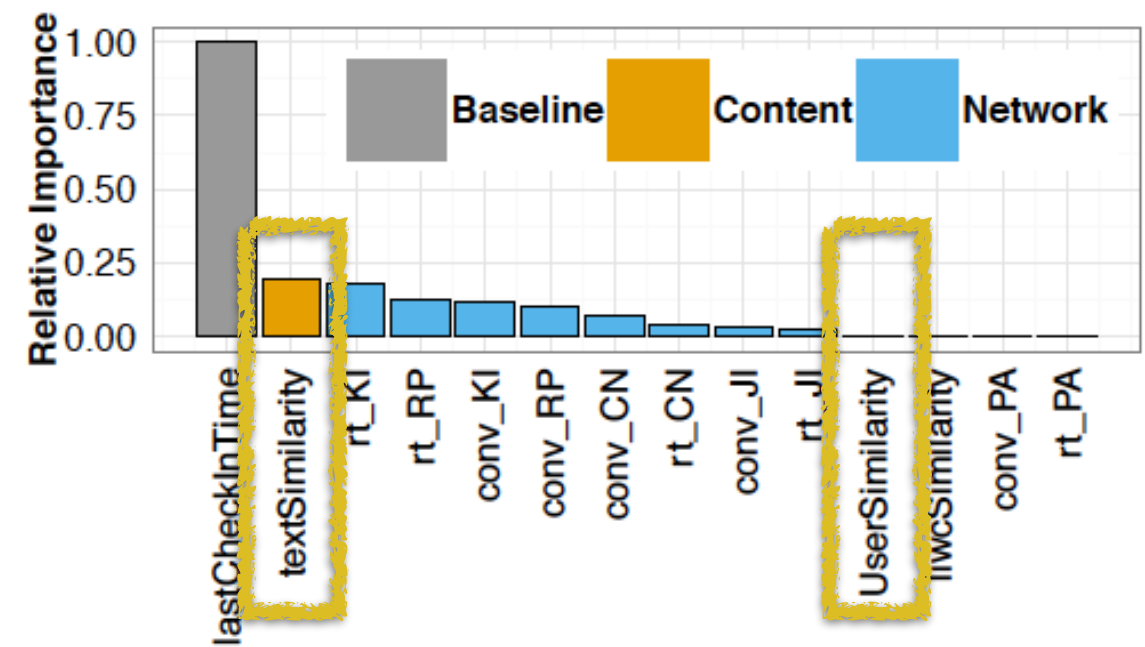
(a) Coordinate Events



(b) Promote Information



(c) Express Opinions



(d) Request Information

Similarity between question and users past tweets is more important!

# Summary

- We provided the first large-scale empirical study on the information seeking and responding networks in physical gatherings, using Twitter in Academic conferences as a case study.
- We profiled the four prototypical information seekers in the conferences.
  - Opinionators, Coordinators, Promoters, Requesters.
- With a set of similarity measures, we are able to predict the replying action to the questions with high AUC.

# Potential Applications & Implications

- For the Individuals:
  - We can route the less noted questions to the probable respondents.
- For the communities:
  - Event organizers could consider constructing question leader list, so that people with various interests could selectively join the discussions more quickly.
  - Information Seeking and Responding dynamics relates to the sustainability of the community.

Thank you!  
Any Questions?