

DH-500 Paper Presentation

**Big Questions for Social Media Big Data:  
Representativeness, Validity and Other Methodological  
Pitfalls**  
by Zeynep Tufekci

Junhong Li

2021.04.15



Social media



Large-scale  
databases of human  
activity

Methodological

**Challenges?**

Conceptual

# Main Content

- Methodological Issues
  1. The model organism problem
  2. Hashtag analyses (Dependent variables selecting)
  3. The denominator problem
  4. Overlook the ecology for platform
- Interpretation Issues
  1. Complex social interactions meanings
  2. Unintelligible to algorithms
  3. Network methods
  4. “node-to-node” & “field” effects
  5. Self-awareness

# The model organism problem

ICWSM



&



# The model organism problem

Limit



&



# Hashtag analyses (Dependent variables selecting)



Figure 1

The frequency of top 20 hashtags associated with Gezi Protests. (Banko and Babacan, 2013)

# The denominator problem



Viewing Sprout Social's History esc to go back

**History**

- Direct Messages
- Tweets
- Mentions
- Followers
- Following

**Tweet 1:** Chloe West @ChloeWest28 (2k+) Aug 23, 2019  
I love a good #Twitter audit. Check out my latest @SproutSocial article to learn how to do a quick 20-minute refresh of your account  
[sproutsocial.com/insights/twitt...](https://sproutsocial.com/insights/twitt...)

**Tweet 2:** Chloe West @ChloeWest28 (2k+) Aug 5, 2019  
@TheFigCo @SproutSocial Thanks for sharing! ✨

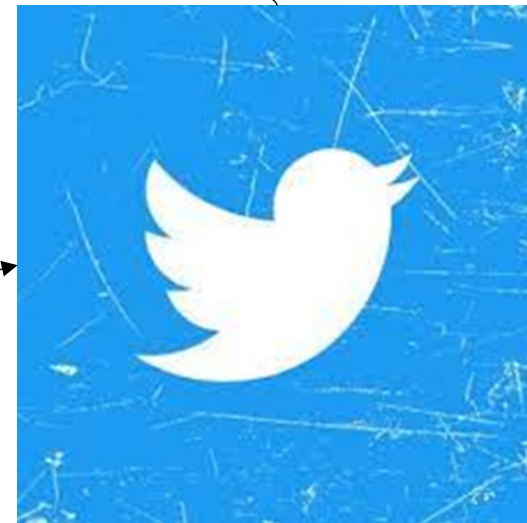
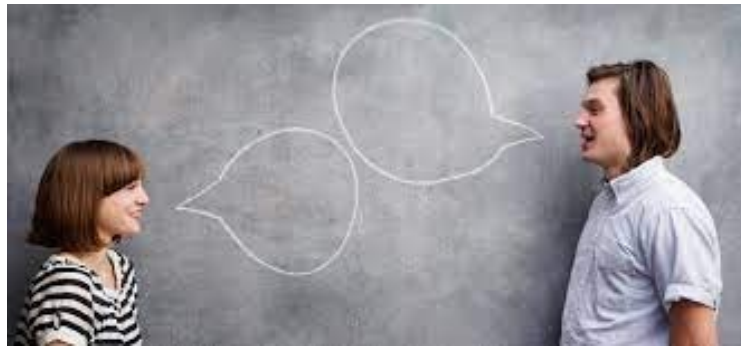
**Tweet 3:** Chloe West @ChloeWest28 (2k+) Jul 1, 2019  
Does your Twitter bio need some love? Check out my article for @SproutSocial on Twitter bio ideas

**Sprout Social Profile:**  
@SproutSocial  
Social media management platform for businesses. Give the 30-day trial a spin: <https://t.co/wSxSchVdH3> Support: @SproutSupport  
Chicago, IL, USA  
<http://t.co/ei9ii8DwcW>

Followers	Following
104.5k	28.7k

Contact Info  
Follow status  
Twitter Lists

# Overlook the ecology for platform





# Complex social interactions meanings



Retweet



# Unintelligible to algorithms

- Subtweets

making a tweet referring to a person algorithmically invisible to that person.

- Hate-Links

Occurs when a user links to another user's link rather than quoting or mentioning the user

- Screen Captures

Using Screen captures rather than quotes is also invisible to algorithms.

- Other methods

Blind to machine algorithms!!!

## Network methods

- Social network & airlines network
- Information & germs
  1. Adjacency in social media is multi-faceted and it is hard to map to physical proximity; human ‘nodes’ have wide range of information sources
  2. Whether there is a straightforward relationship between information exposure and rate of influence need to be investigated.

## “node-to-node” & “field” effects

- Human social information flows do not occur only through node-to-node networks but also through field effects, large-scale societal events which impact a large group of actors contemporaneously.  
e.g. Weather occurrence has field effects and often do not diffuse solely through interpersonal interaction.

# Self-awareness



Figure 2

Ankara Mayor leads a hashtag campaign that will eventually trend worldwide

# Conclusion

Social media big data is a powerful addition to the scientific toolkit. However, this emergent field needs to be placed on firmer methodological and conceptual footing. There are some practical steps for the challenges discussed above:

- Target non- social dependent variables
- Qualitative pull-outs
- Baseline Panels
- Industry outreach
- Convergent answers and complimentary methods
- Multidisciplinary teams
- Methodological awareness in review

**Thanks !**