NETWORK SCIENCE OF ONLINE INTERACTIONS

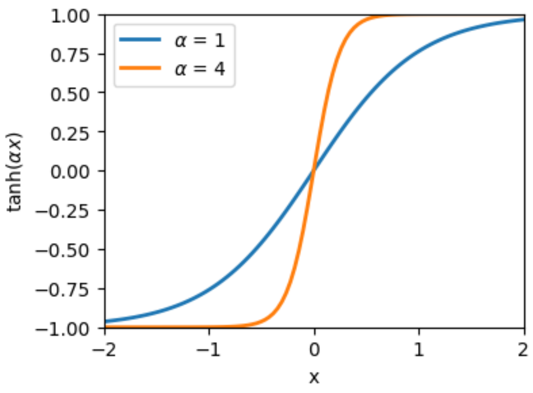
Scaling laws and Inequality across social media

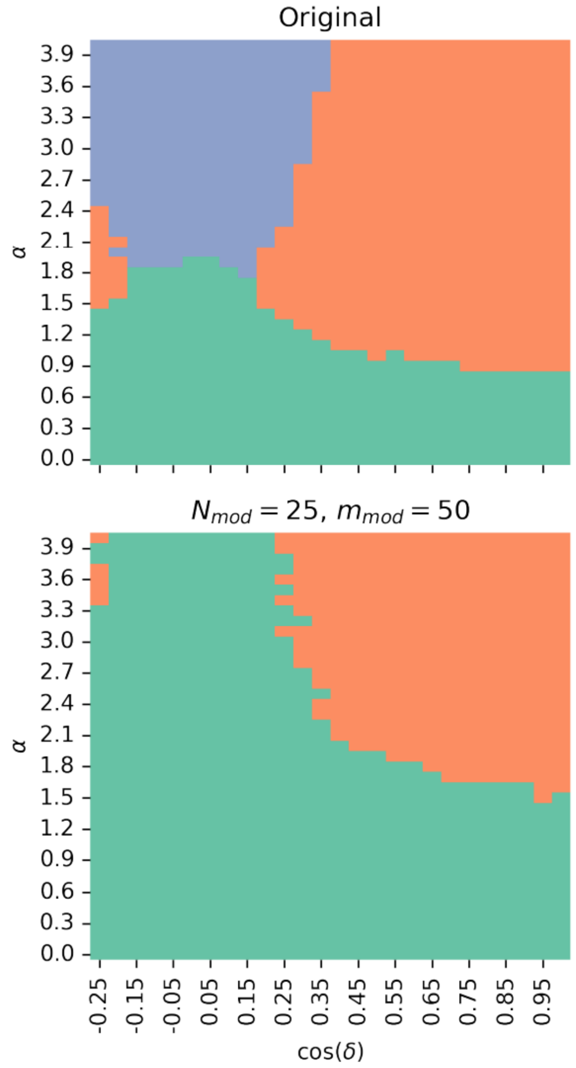
Joao Neto

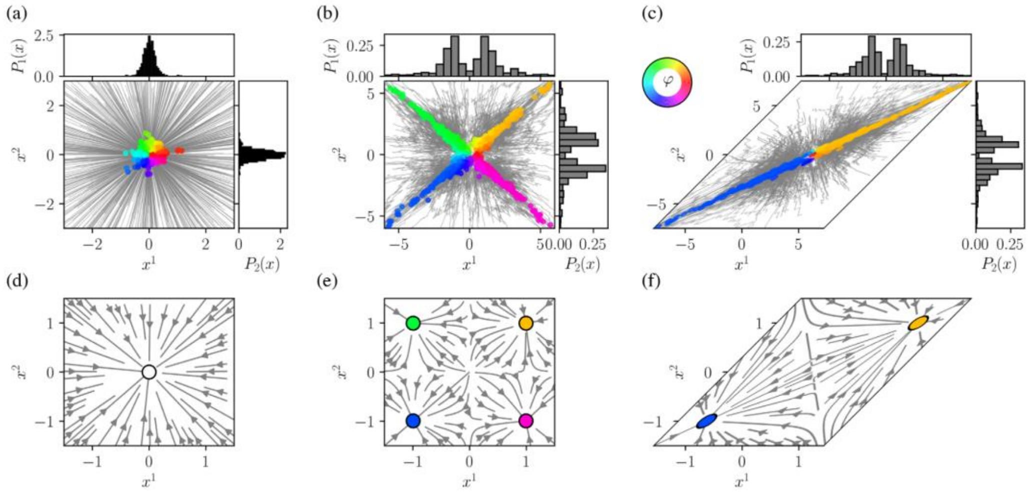
07/Jul/2023

LAST SUMMARY

- Polarization depends heavily on
  - Homophily
  - Relative impact of extreme opinions
- Opinion correlation may result in ideology
- Moderation can counter-balance polarization





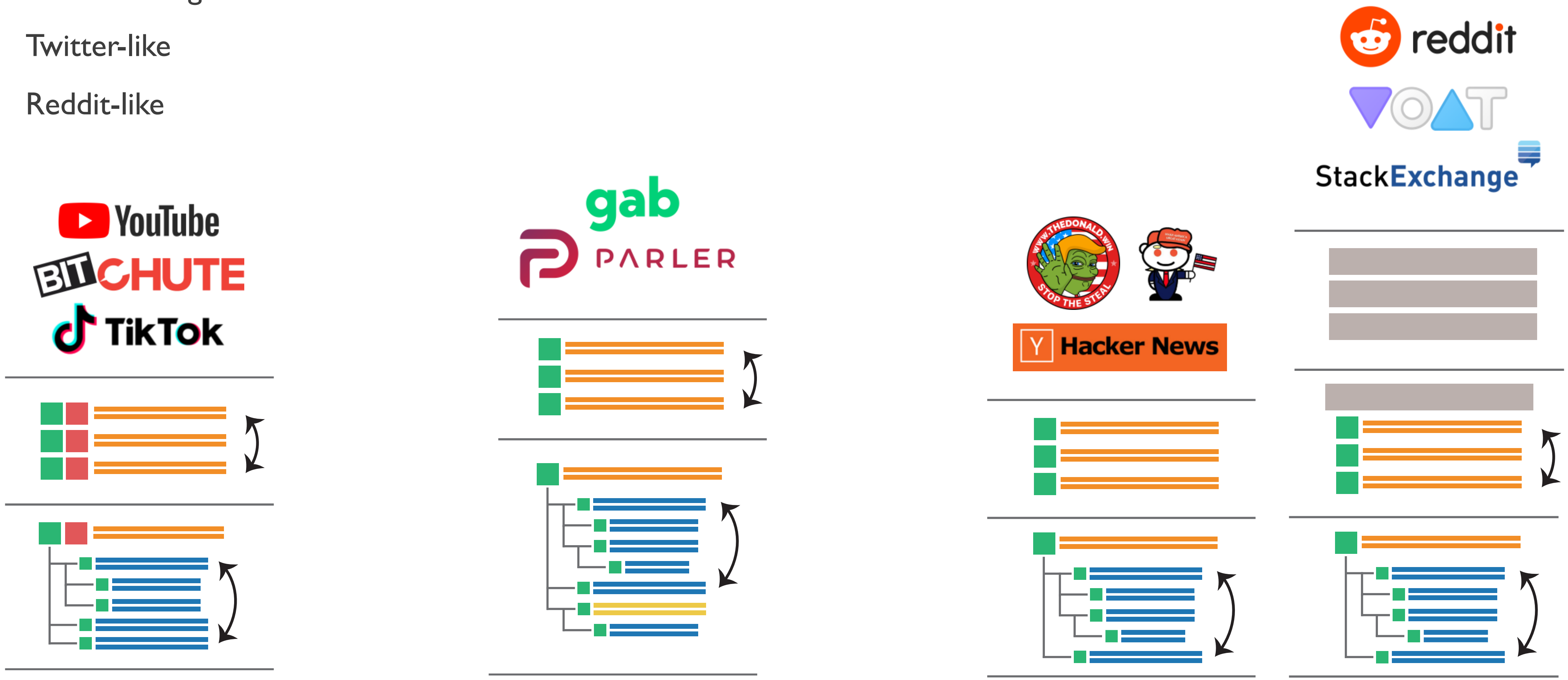


INTRODUCTION

- Many things in social media scale as $y \sim x^{\gamma}$
- So far, mostly probability distributions $P(x) \sim x^{-\alpha}$
- What about other types of scaling?
- How do they vary across social media platforms?

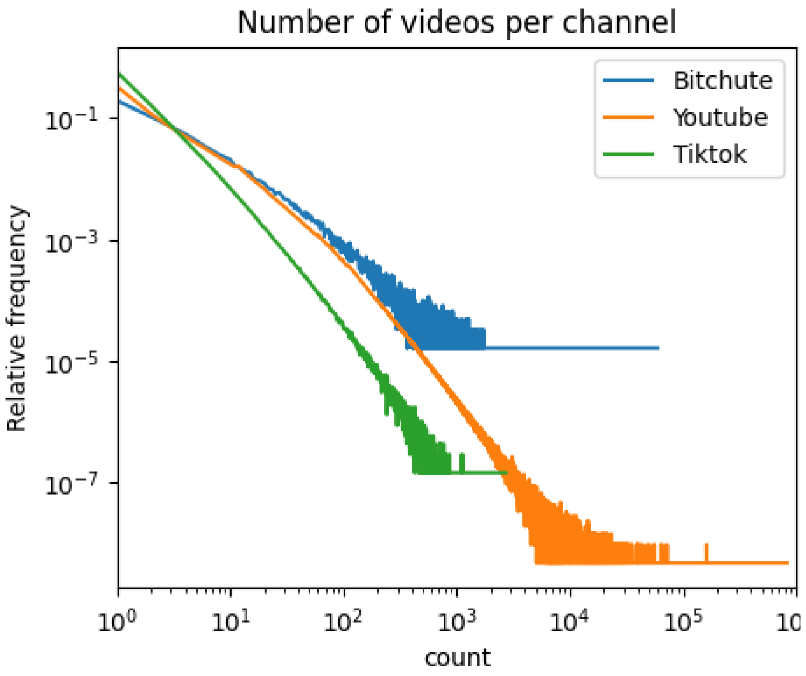
PROBABILITY DISTRIBUTIONS

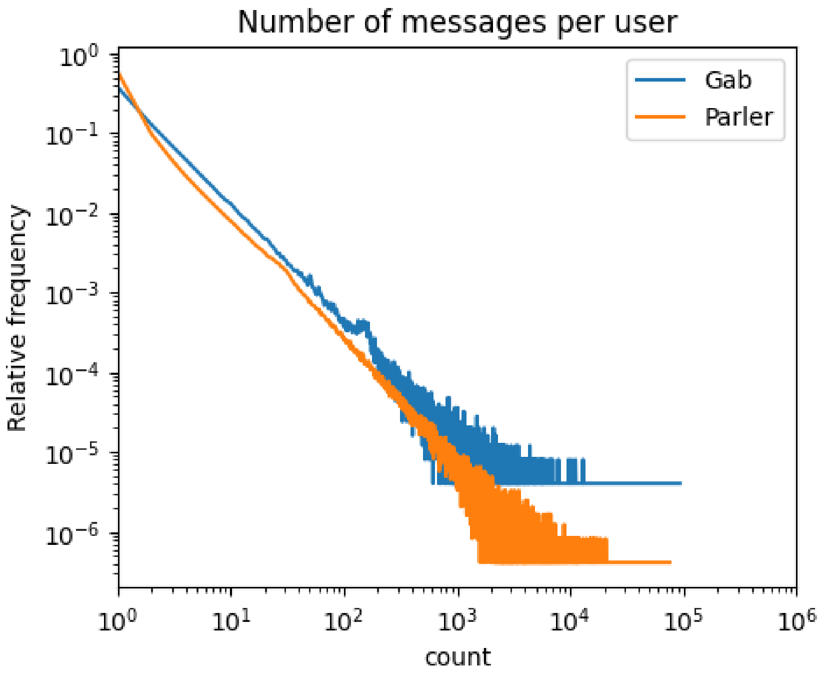
- Categorizing platforms
  - Video sharing
  - Twitter-like
  - Reddit-like

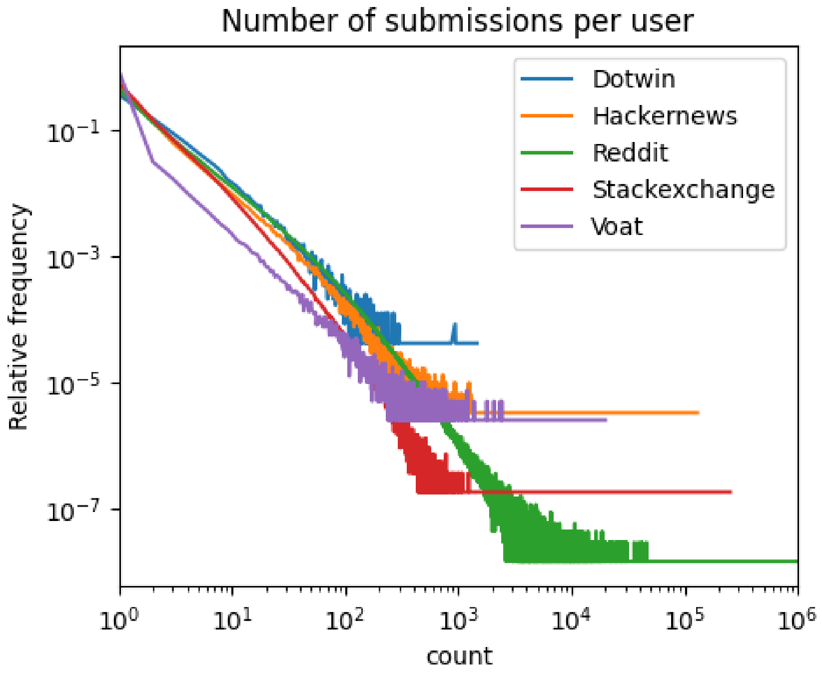


DISTRIBUTION OF CONTENT CREATION PER USER

- Video sharing: Youtube and Bitchute lognormal, Tiktok powerlaw
- Twitter-like: consistent
- Reddit-like: consistent, Voat outliner

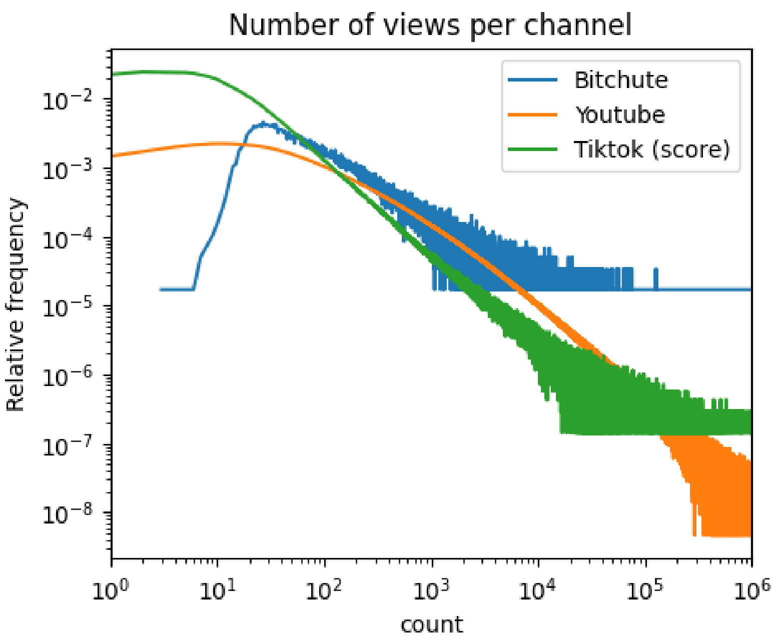


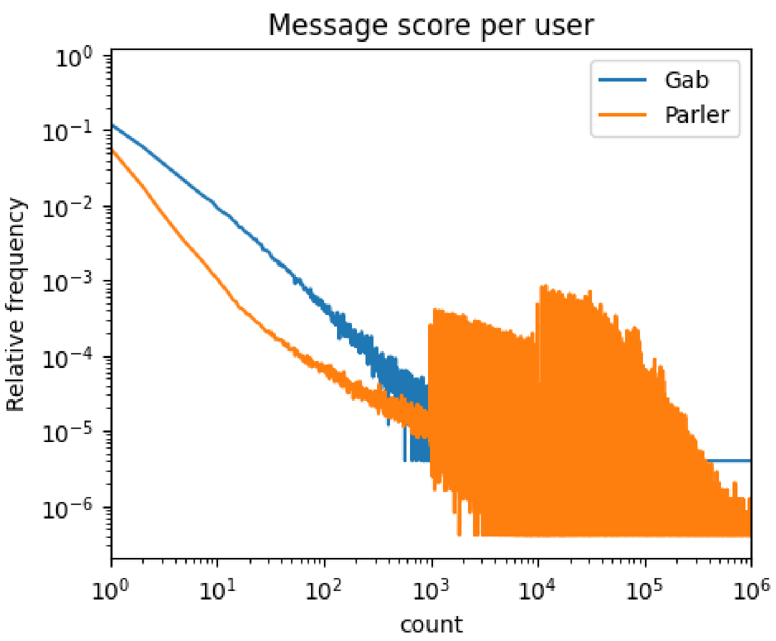


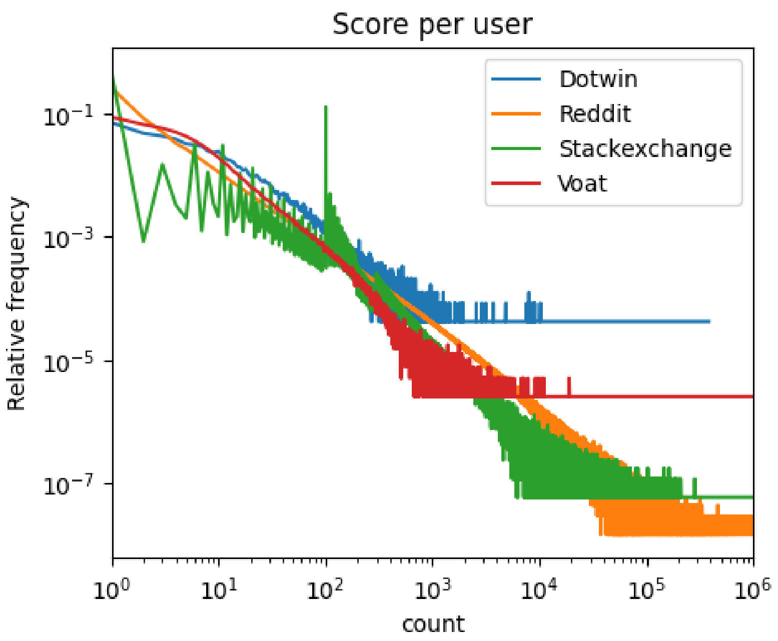


DISTRIBUTION OF CONTENT REACTION PER USER

- Video sharing: Youtube lognormal, Bitchute and Tiktok powerlaw
- Twitter-like: Gab powerlaw, Parler bad data
- Reddit-like: all powerlaw, Stackexchange inconsistent data

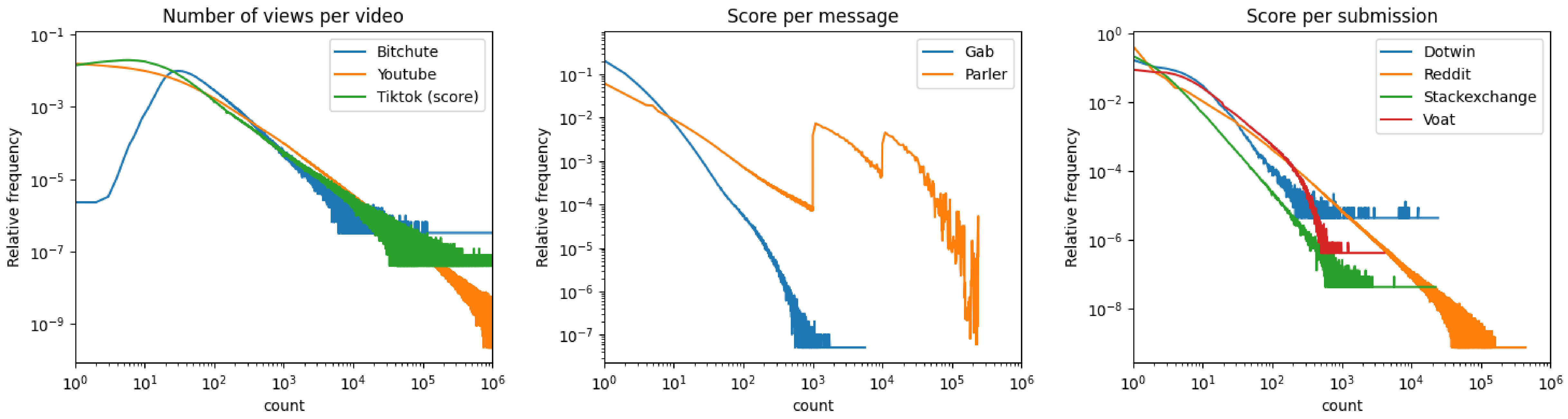






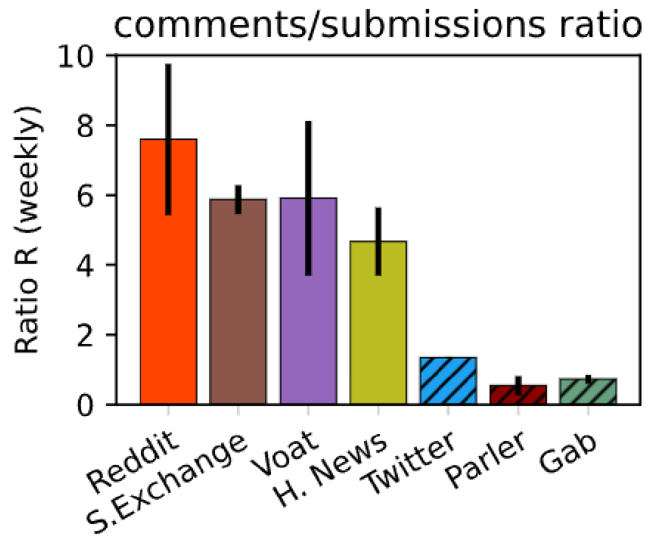
DISTRIBUTION OF REACTIONS PER SUBMISSION

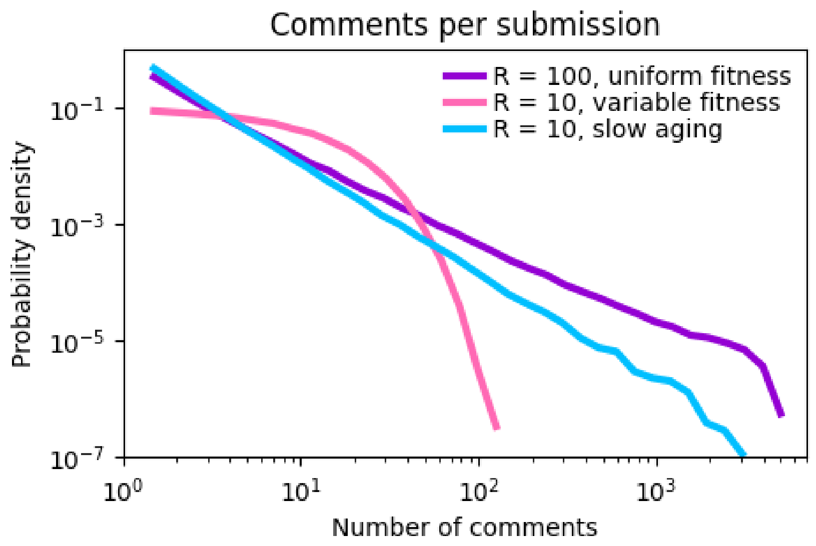
- Video-sharing: consistent
- Twitter-like: very different, Parler bad data
- Reddit-like: powerlaws, different exponents

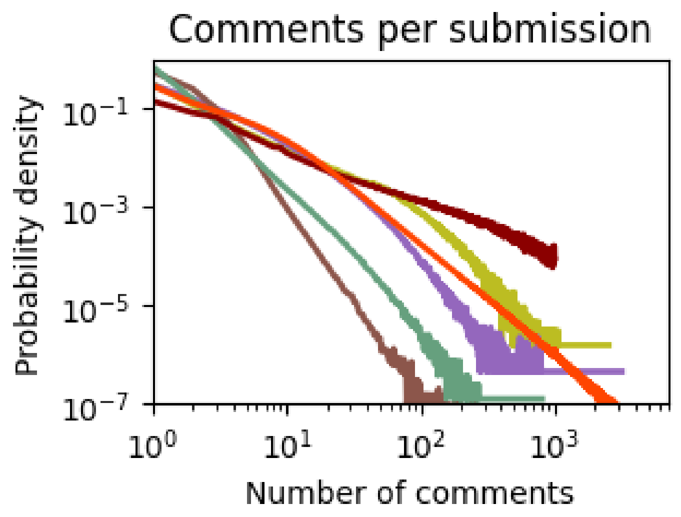


SUMMARY

- Some distributions are consistent between platforms
  - Comment/submission ratio
- Some are very variable
  - Can be due to design differences (e.g. disallowing submission downvote)
  - Can be due to content differences
    - Model with variable fitness and aging effects can reproduce it

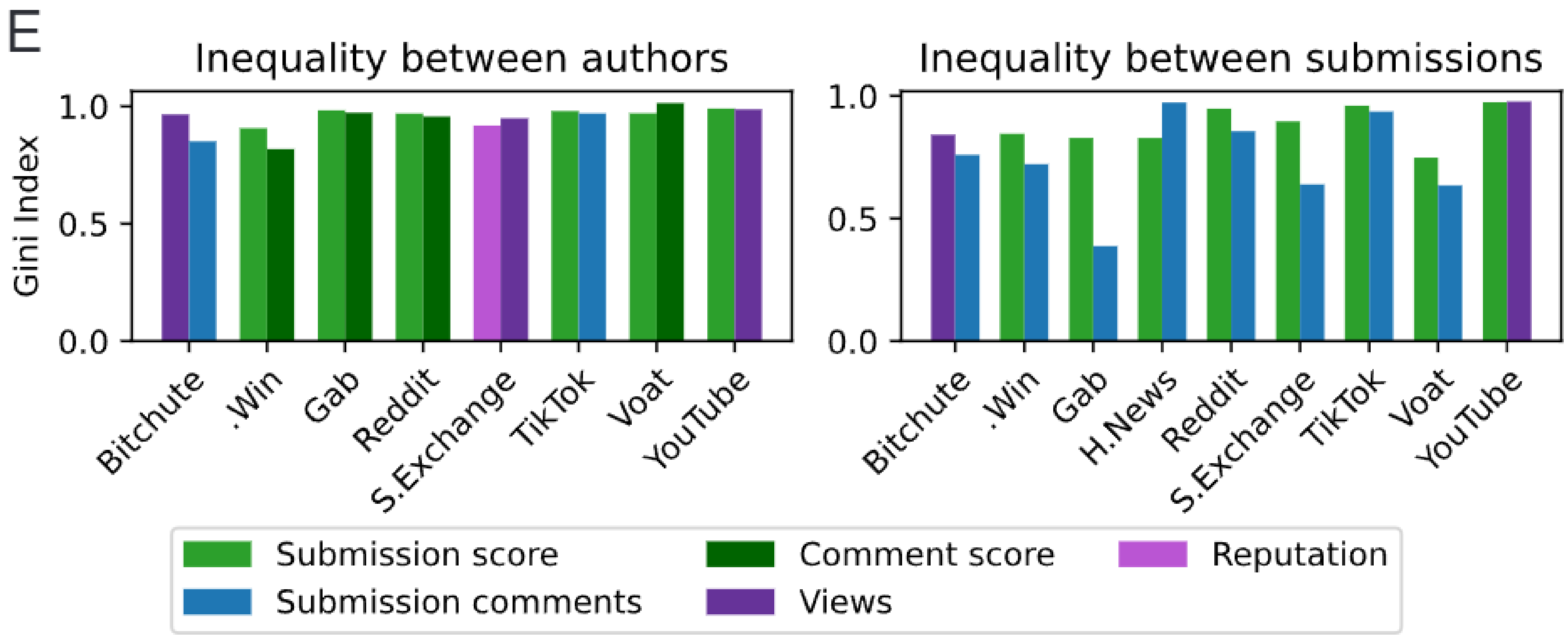






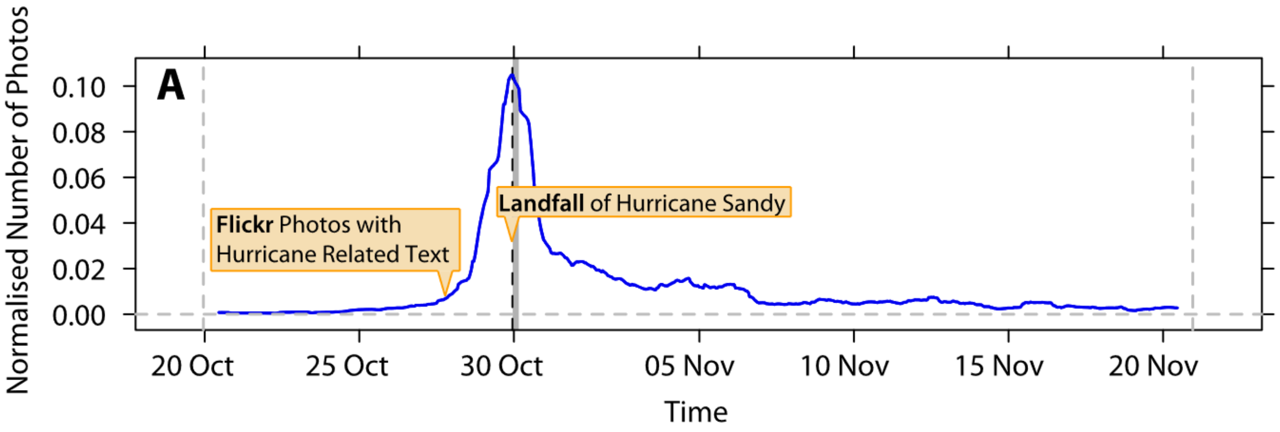
SUMMARY

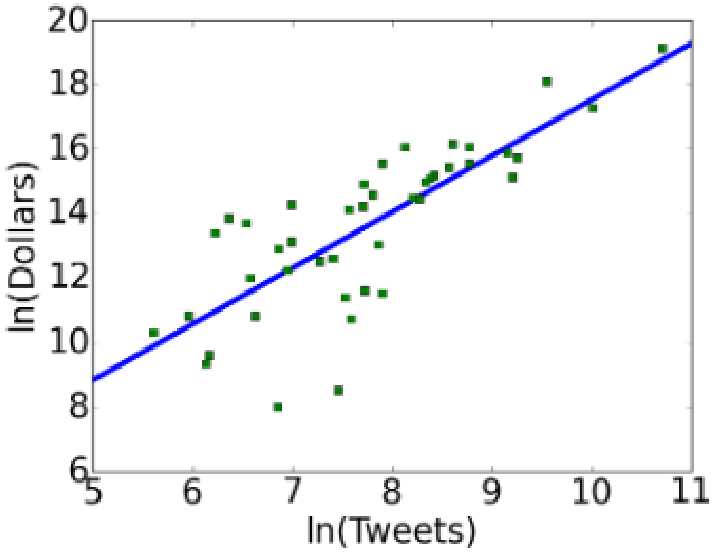
- Consequence: extreme inequality



REAL-WORLD SCALING

- Relationship between chatter and real-world effects
- 2012 Hurricane Sandy
  - largest Atlantic hurricane ever
  - $70B damages
  - 233 killed
- Activity on Flickr peaked at 10% of all photos [1]
- Amount of money donated (per US state) correlated with social media activity [2]
  - $money \sim tweets^{\gamma}, \gamma \approx 1.74$

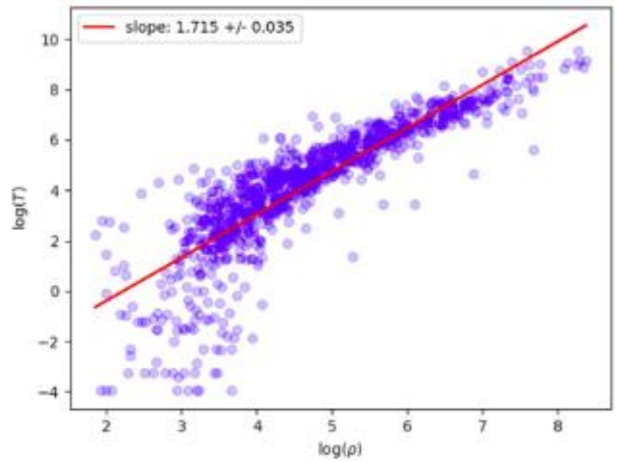


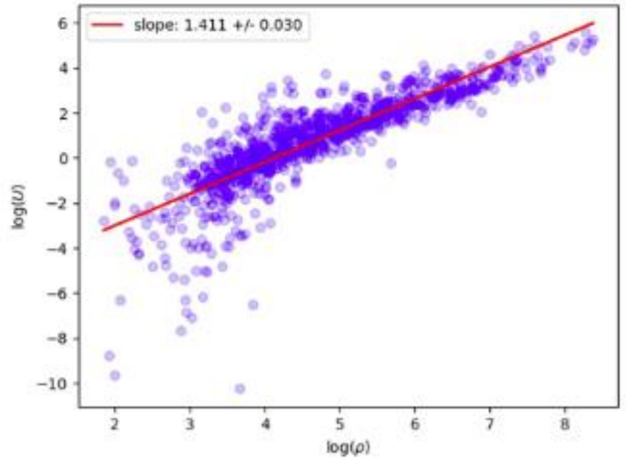


[1] Preis, T. et al, "Quantifying the Digital Traces of Hurricane Sandy on Flickr", Scientific Reports, vol. 3, 2013. doi:10.1038/srep03141.
[2] R. Korolov, et al. "Actions are louder than words in social media," 2015 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM)

REAL-WORLD SCALING

- Maybe $money \sim tweets^{\gamma}$ because $tweets^{\gamma} \sim people$?
- The idea: use geotagged tweets to test this
  - South-West UK, 04/2016-04/2018
- Quantities
  - Grid size
  - Tweet density T, population density ρ, user density U
- Both T and U scale as ρ^{γ}
  - Tweets: $\alpha = 1.715$
  - Users: $\beta = 1.411$
  - $T \sim U^{\gamma}, \gamma = 1.144$

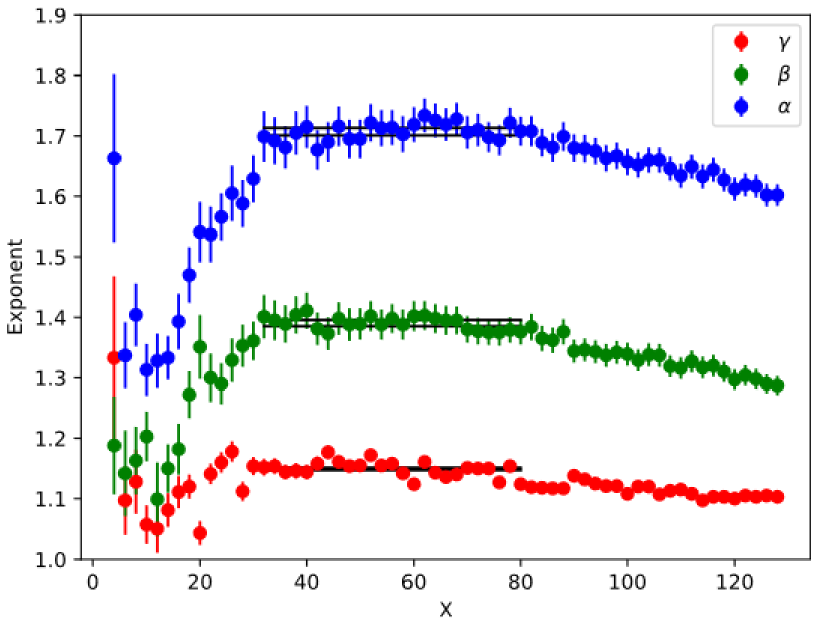


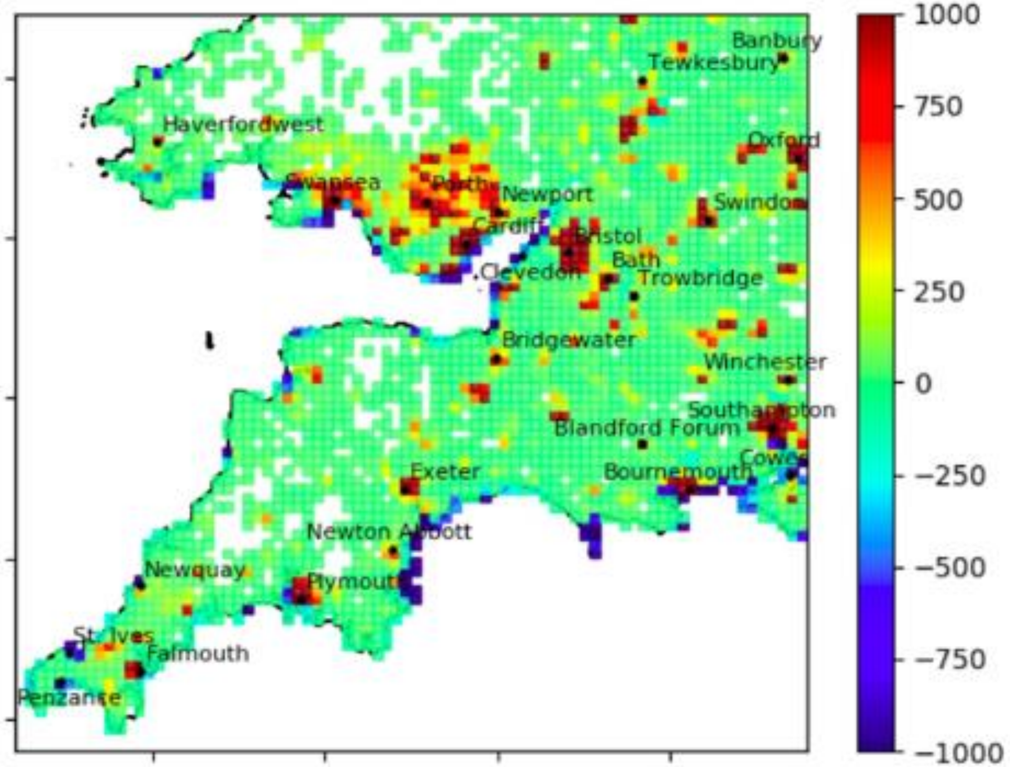


Arthur R, Williams HTP(2019) Scaling laws in geo-located Twitter data. PLoS ONE14(7): e0218454. https://doi.org/10.1371/journal. pone.0218454

REAL-WORLD SCALING

- Grid size effect?
  - Grid too large: few data points
  - Grid too small: noise from e.g. communting
- Fits are robust across a sizeable grid variation
- For scale: Young user density on a 80x80 grid

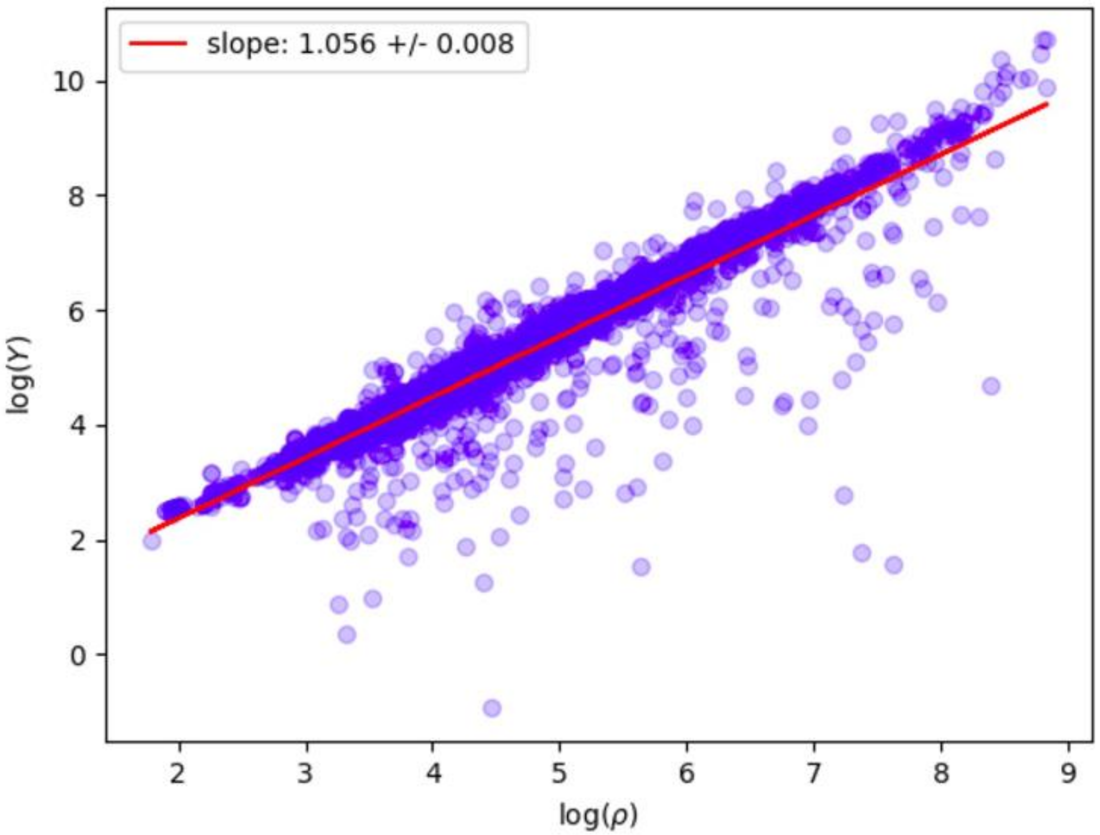




REAL-WORLD SCALING

- Maybe it is just young users?
  - Tend to migrate to bigger cities
- Scaling of young twitter users Y and ρ
  - $\gamma = 1.056$
  - Doesn't explain the effect

- Twitter usage appears to be very concentrated on cities



SOCIAL MEDIA SCALING

- We focused on statistical analysis and spreading from the POV of users and content

- Different spreading object: **hashtags**
  - Akin to Reddit communities, but spread in the wild instead of concentrated
  - They have a life-death cycle: branching processes?

nature COMMUNICATIONS

ARTICLE

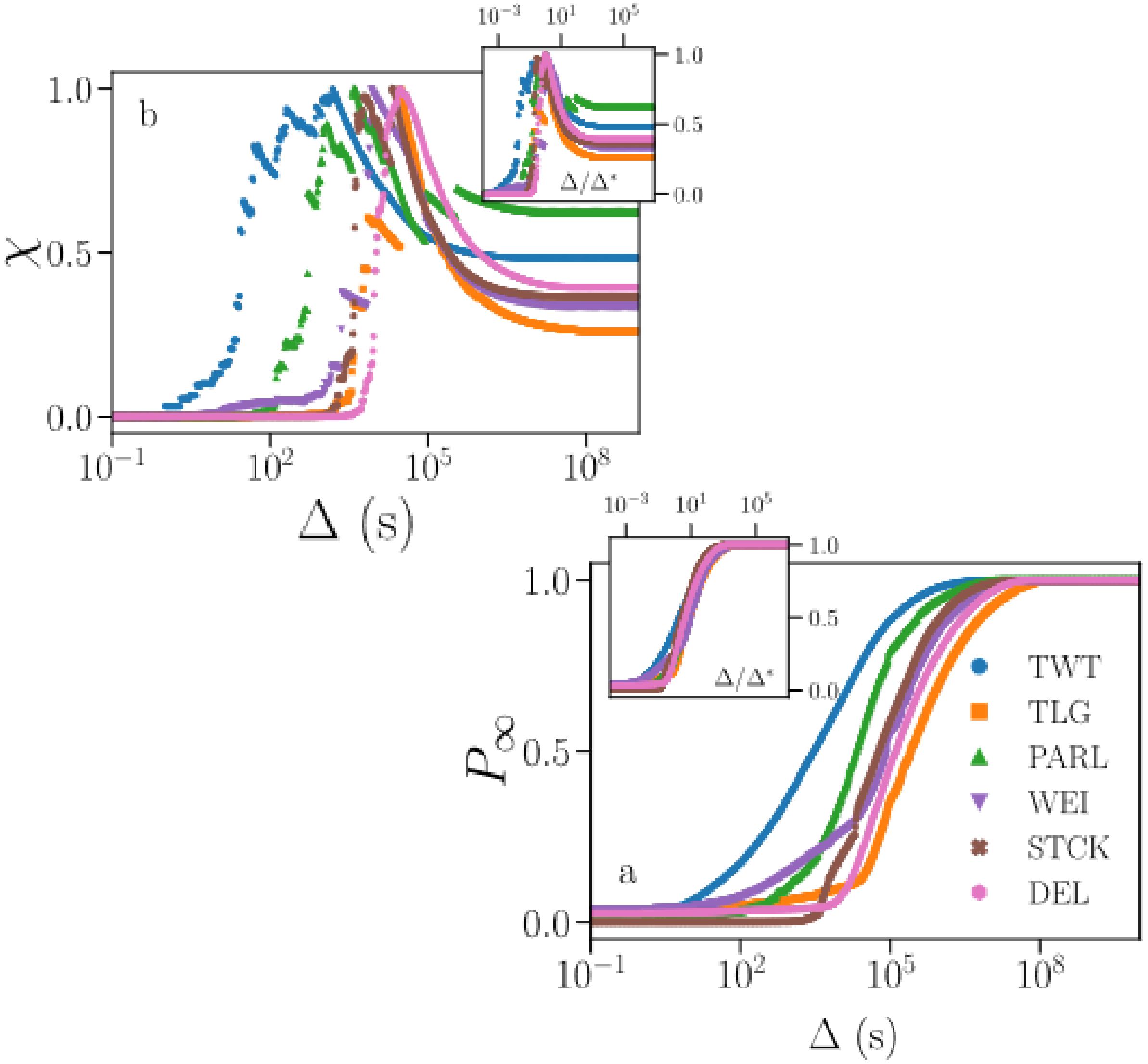
https://doi.org/10.1038/s41467-022-28964-8 OPEN

Universality, criticality and complexity of information propagation in social media

Daniele Notarmuzi[1], Claudio Castellano[2,3], Alessandro Flammini[1], Dario Mazzilli[1,3] & Filippo Radicchi[1]✉

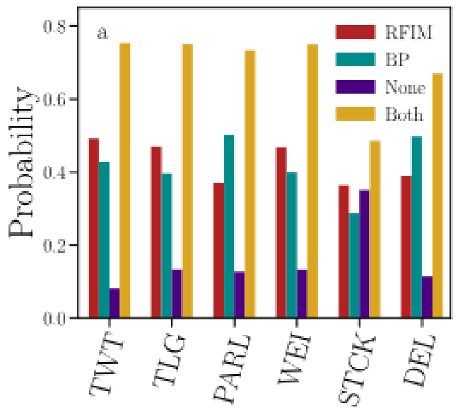
SOCIAL MEDIA SCALING

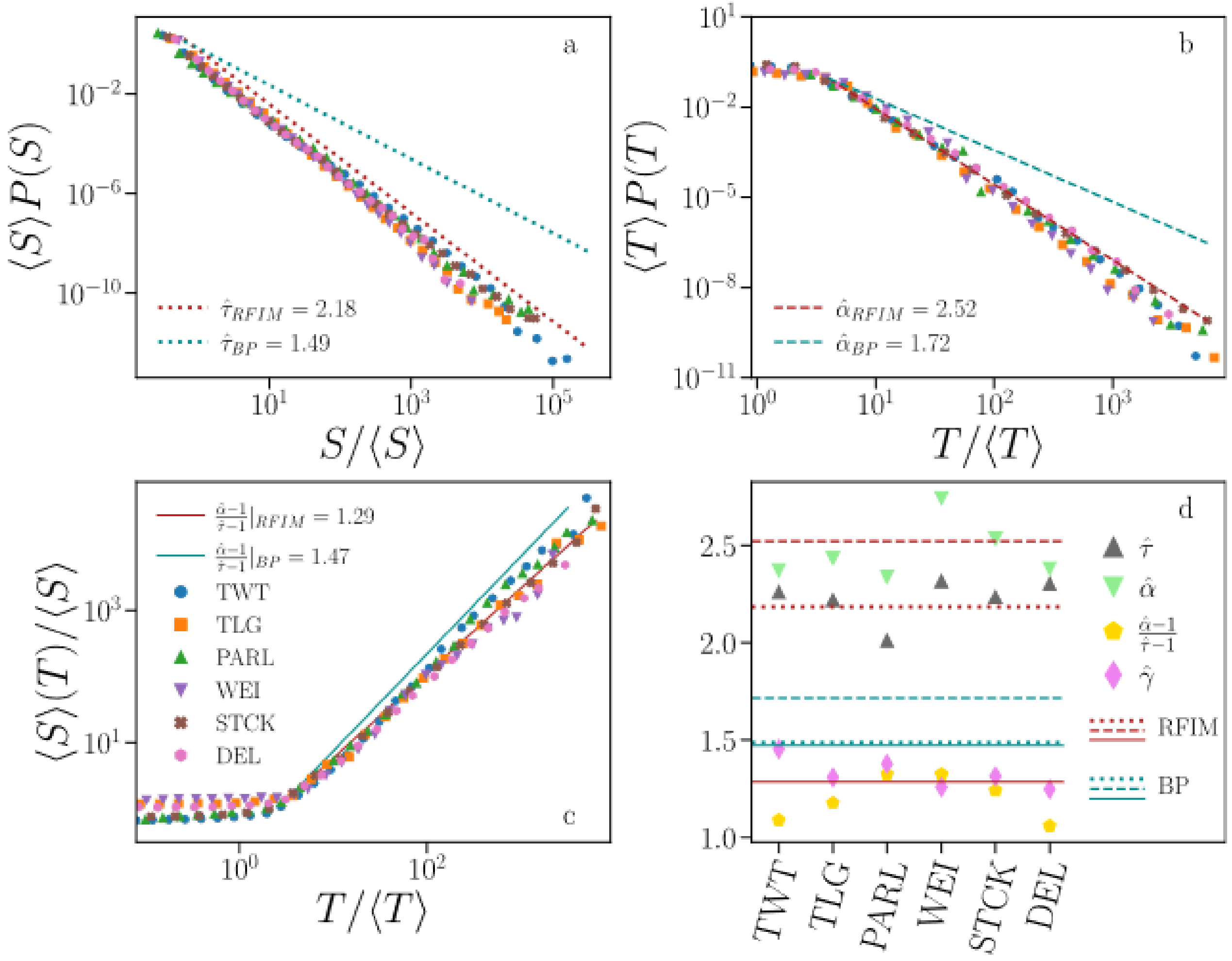
- The idea: treat hashtag propagation as a percolation problem
- The data
  - 206M timeseries (hashtags)
  - From 6 platforms (Twitter, Telegram, Parler, Weibo, Stackoverflow, Delicious
- The issue: define a bin size Δ to determine when a hashtag is "off"
- Solution:
  - For each time series, find $\Delta^\star$ that maximizes the variance of the largest avalanche size s_M
  - Call the variance the susceptibility χ
- Percolation threshold collapses on $\Delta^\star$



SOCIAL MEDIA SCALING

- When using $\Delta^{\star}$, power-law exponents collapse
  - But **not** on a branching process (BP)
  - Random Field Ising Model (RFIM) is a better fit
- Suggests complex contagion (RFIM) instead of simple contagion (BP) for most processes
- Re-analysis using $S > 10$
  - Both are good models





CONCLUSION

- Social media scaling varies by platform
- Can be explained by design differences
- Creates inequality
- Is related to real-world scaling
- May have some universal components