Memes propagation as a complex system

Francesco Malferrari 193103

General definition

"s. m. Singolo elemento di una cultura o di un sistema di comportamento, replicabile e trasmissibile per imitazione da un individuo a un altro o da uno strumento di comunicazione ed espressione a un altro (giornale, libro, pellicola cinematografica, sito internet, ecc.)." Treccani

"A cultural feature or a type of behaviour that is passed from one generation to another, without the influence of genes: A meme is the cultural equivalent of the unit of physical heredity, the gene." Cambridge Dictionary

How does it change on the Internet? Sounds and videos

“Rickrolling”

“The Nyan Cat”

How does it change on the Internet? Images

Variable text with fixed image

They often tell funny events from cultural elements or everyday life

Cultural propagation: example

Spider-Man (1967 TV series) episode 19 "*Double identity*" season 1

Spider-Man: Into the Spider-Verse (2018); post-credits scene

Spider-Man: No Way Home (2021); scene before final battle

Spider-Man: Across the Spider-verse (2023); chasing scene

Marketing meme

Kalyug
@kalyugmemes

Me & my boys when Nykaa's Pink Summer Sale goes live...

mynykaa • Follow

mynykaa LOVE THE MEMES y'all have been making about Nykaa's Pink Summer Sale... Keep 'em coming!

Here is an EXTRA surprise!
Use "EXTRA15" and get 15% on orders above ₹4000/-
*Max coupon discount of ₹700/- | Applicable on app orders only!

#Nykaa #repost #meme #memeseveryday #memesdaily #NykaaPinkSummerSale #PaintTheSummerPink #PinkSummer #Summer #SummerVibes #SummerMood #SummerStateOfMind #SummerSale #SaleIsLive

Edited · 1w

rainbow_butterfly_unicorn Love the last one so relatable those influencers r forcing me to delete my social media account bcoz self control naam ki koi cheez nhi hai

1w 2 likes Reply

View replies (1)

annesha_banerjee_mukhopadhyay It was on my wedding

9,203 likes
MAY 10

Add a comment... Post

Memes evolve

The Selfish Gene (1976)

Dawkins' hypothesis

1. Replication
2. Variation
3. Selection

Visual memes

1. To take (and eventually modify) an image and add some text to represent a situation
2. New memes are continuously created to target funny situations or jokes and compete for users' attention flowing across online communities
3. A meme that cannot attract human attention nor adapt to transmit new contents disappears adapting to the fast online environment

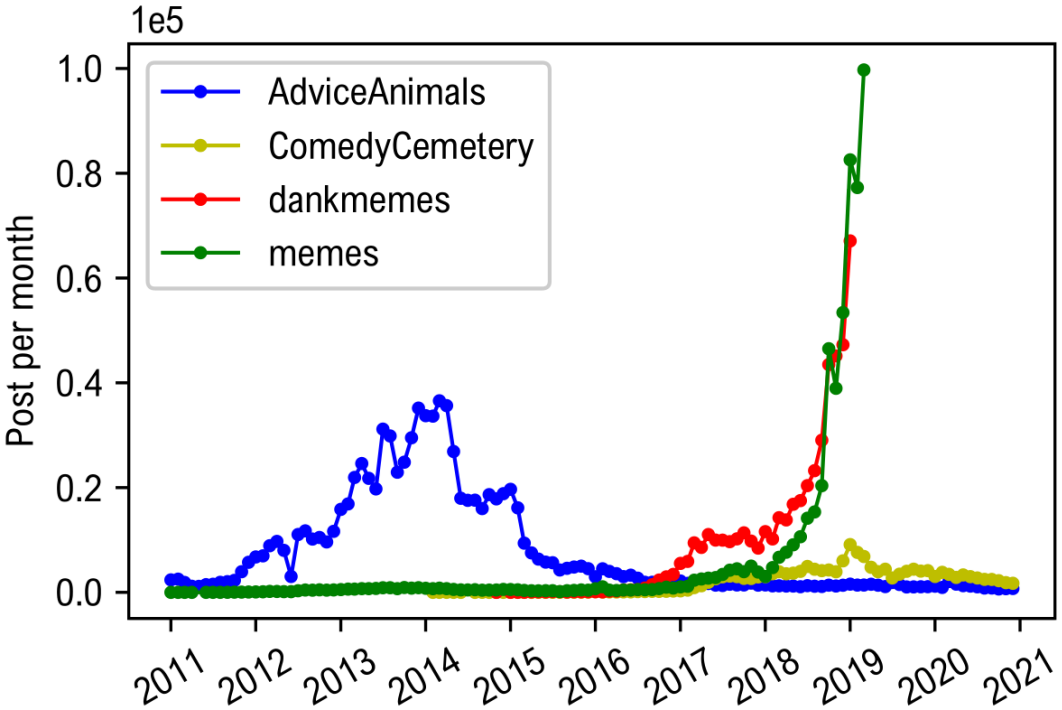
"Memetics is the system and art of importing genetics to social sciences." [3]

First research in this talk

- Title: “Entropy and complexity unveil the landscape of memes evolution” [1]
- Analysis: 2 million memes posted across four subreddits
  - r/memes
  - r/dankmemes
  - r/AdviceAnimals
  - r/ComedyCemetery
- Period of time: from 2011 to 2020

Dataset used

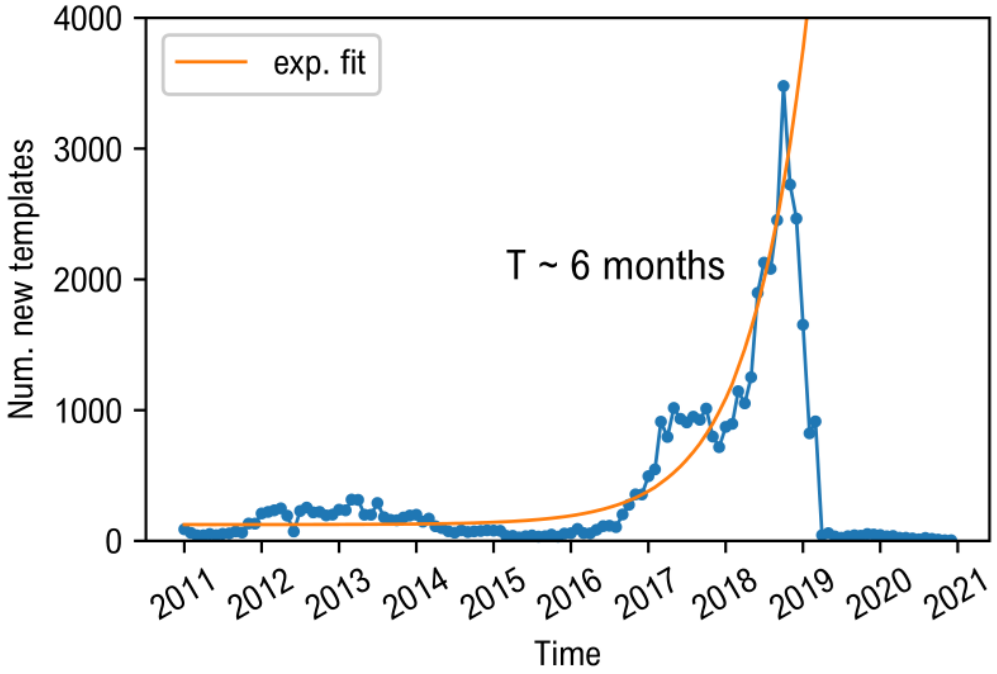
2 million of memes from Reddit

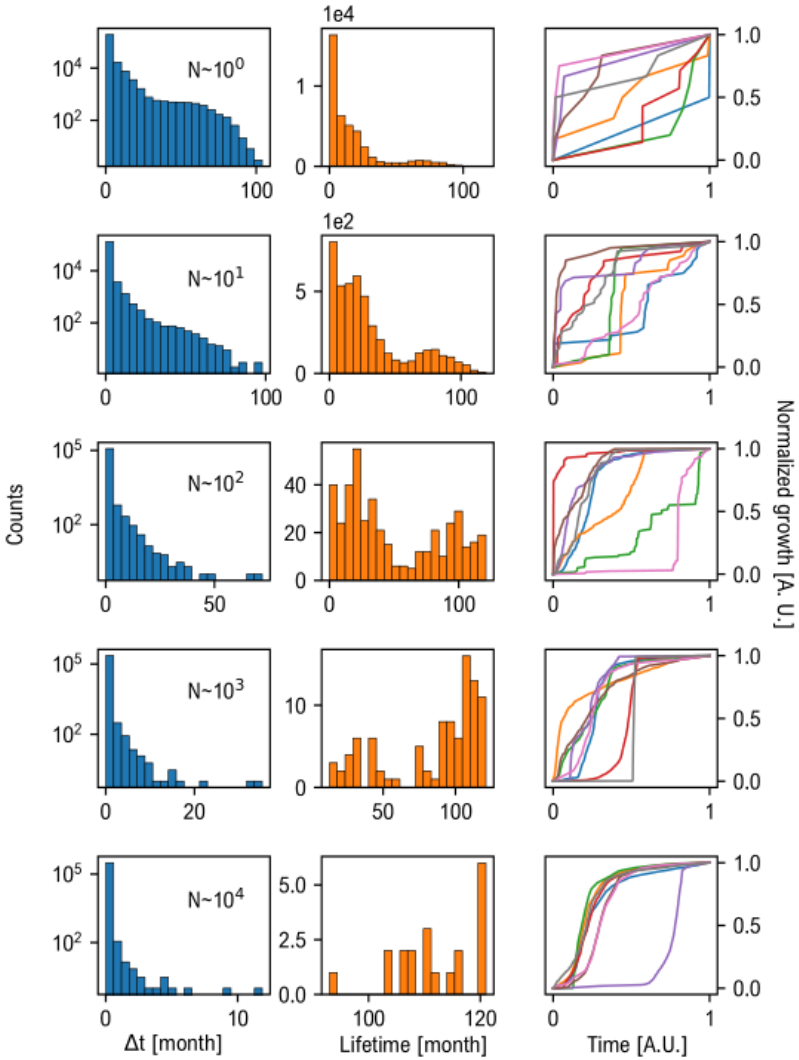


For subreddits r/memes and r/dankmemes it wasn't possible to analyse the data over 2019, because of the exponential trend in the number of memes

Evolution rate

“The growth rate is estimated through an exponential fit, with a doubling time of T ~ 6 months.” [1]



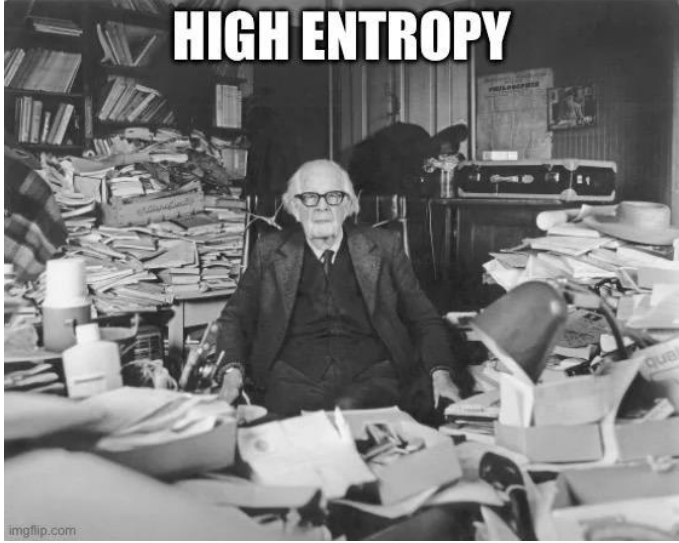


Mutation rate

- "Small clusters show a heterogeneous distribution of instances' inter-arrival times, comprising both small-bursty clusters and small-slowly paced ones." [1]
- "As the cluster size grows, we observe a shift of the inter-times distribution towards low values, unveiling faster dynamics, while the lifetime distribution is concentrated around larger values." [1]
- Long lasting memes tend to a strong early adoption

N.B. This doesn't does not accurately represent those cases where a meme takes time to gain popularity. "This aspect may be due to non-trivial popularity dynamics, calling for further research." [1]

Entropy



- “Permutation entropy H and statistical complexity C are two quantities that can be used to synthesize general properties of images, based on the value and relative disposition of their pixels.” [1]
- H measures the degree of disorder in the pixel arrangement.
- To compute H and C all colored images were converted to grayscale.

$$P = \{p_1, \dots, p_n\}\ with \sum_{i=1}^{N} p_i = 1 \longrightarrow \text{Probability distribution}$$

$$H(P) = \frac{S(P)}{\log(\mathrm{N})} = \frac{1}{\log(N)} \sum_{i=1}^{N} p_i \log\left(\frac{1}{p_i}\right) \longrightarrow \text{Permutation entropy. } S \text{ is the Shannon Entropy}$$

Complexity

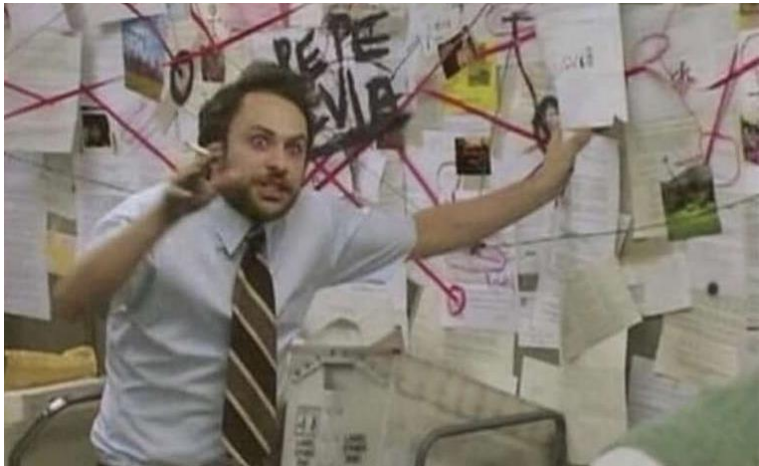
$$U = \{u_1, \dots, u_n\}\ with \sum_{i=1}^{N} u_i = 1 \longrightarrow$$ Uniform distribution

$$D(P,U) = S\left(\frac{P+U}{2}\right) - \frac{S(P)}{2} - \frac{S(U)}{2} \longrightarrow$$ Jensen-Shannon divergence

$$C(P) = \frac{D(P,U)H(P)}{D^*}$$

Statistical complexity C measures the amount of "structural" complexity in the pixel arrangement

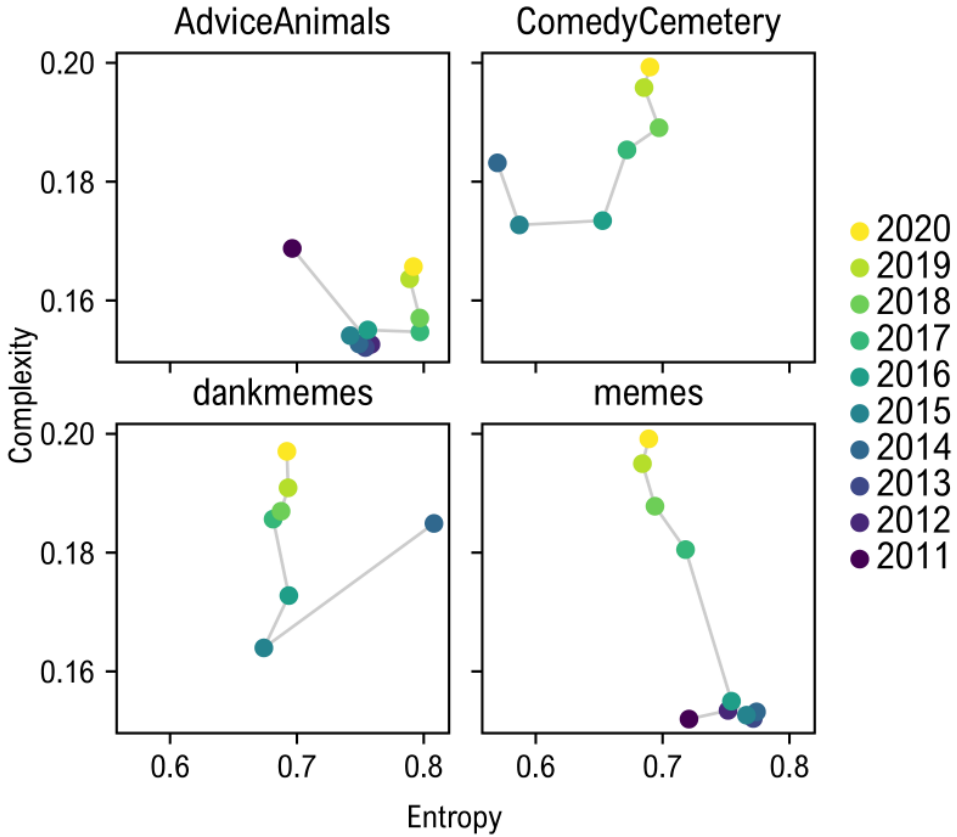
D^* is the normalizing factor



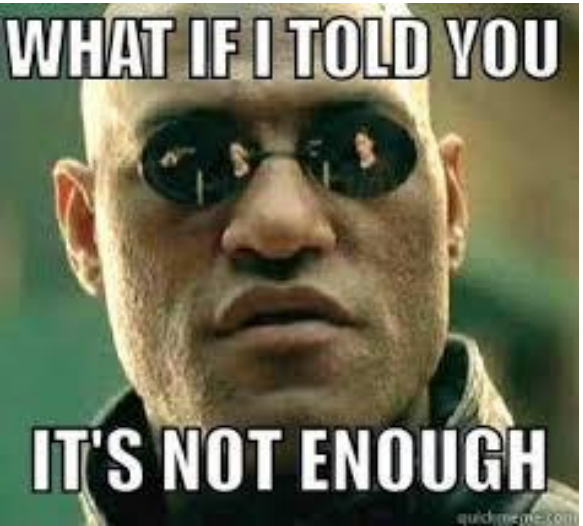
Observations obtainable from complexity and entropy

Evolution towards more complex structures with a stable entropy.
One exception is r/AdviceAnimals, where the evolution of produced memes is limited by posting rules. In addition it could be linked to an overall decrease in memes production.

“The tendency of memes to evolve towards more complex structures can be explained considering this object as part of the emerging internet meta-language. […].This may lead to a segregation effect, with a specific dialect depending on the community in which a meme is shared. […]. This aspect leads to the use of more complex and specific patterns.” [1]



Studying a single community is not enough

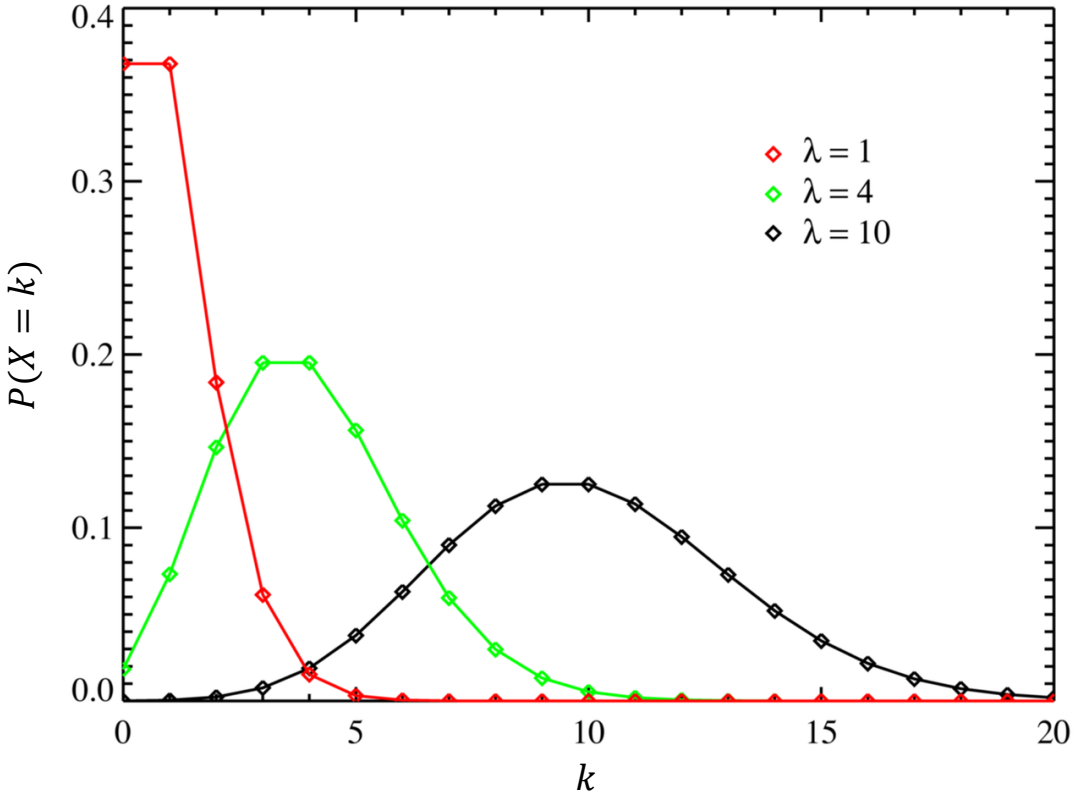


Web-scale analysis

A second paper

- Title: “A web-scale analysis of the community origins of image memes” [2]
- 906481 image memes never been posted before the analysis
- The memes were collected in one month in different communities around the web (Facebook, Reddit, 4chan, Tumblr, etc...)
- Reposts in the originated communities weren’t included because the focus was in external influence
- Period of time: one month
- Removed collaborative communities, as they do not represent a genuine cultural diffusion
- The goal was to answer the organizational sociologists that thought that cultural innovation comes from the periphery and is then adapted by intermediate communities to be diffused in the cores (or mainstream)
- The system that has been created is so robust that it would take more than 30 peripherical communities not considered as prolific as the cores to unbalance it

Wich distribution to use? Poisson



$$\mathrm{P}\{X = k\} = \frac{\lambda^k}{k!} e^{-\lambda} \qquad [4]$$

where

- $\lambda > 0$ is the average number of events per time interval
- X is a discrete random variable
- k is a non-negative integer value

"The Poisson distribution can be considered to be a special case of the negative binomial distribution." [5]

Wich distribution to use? Negative binomial distribution

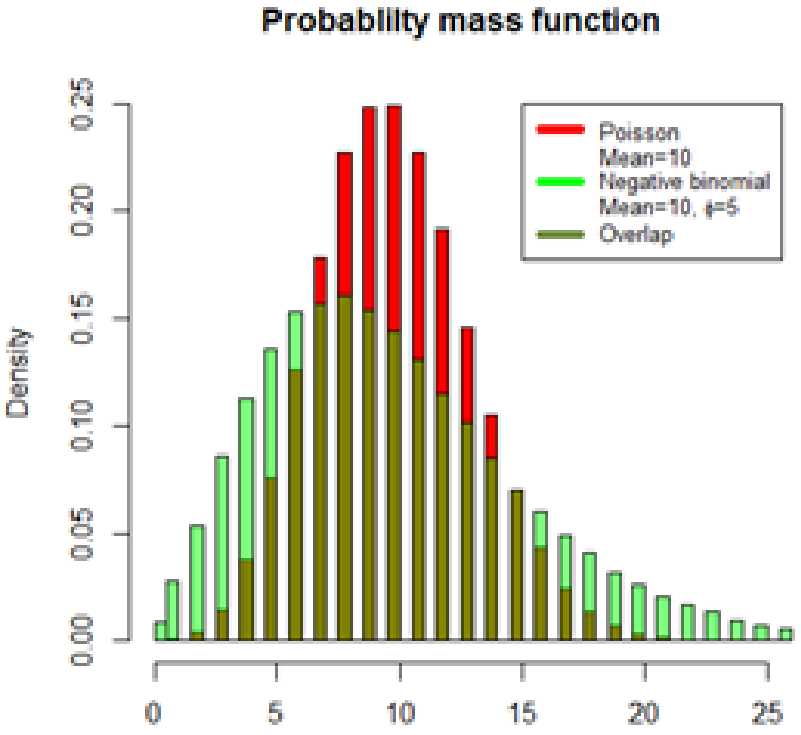


Image taken from [5]

This distribution tends to be wider and shallower than the Poisson one

$$\Pr(Y = y_i|\mu_i, \alpha) = \frac{\Gamma(y_i+\alpha^{-1})}{\Gamma(\alpha^{-1})\Gamma(y_i+1)}\left(\frac{1}{1+\alpha\mu_i}\right)^{\alpha^{-1}}\left(\frac{\alpha\mu_i}{1+\alpha\mu_i}\right)^{y_i} \quad [6]$$

- "The negative binomial considers the results of a series of trials that can be considered either a success or failure" [5]
- This represents the discrete probability function of a given number of successes before a certain number of failures

An important variable for this talk:

$$\mu_i = \exp(\ln(t_i) + \beta_1 x_{1i} + \beta_2 x_{2i} + \cdots + \beta_k x_{ki})$$

Wich distribution to use? Zero-inflated negative binomial distribution

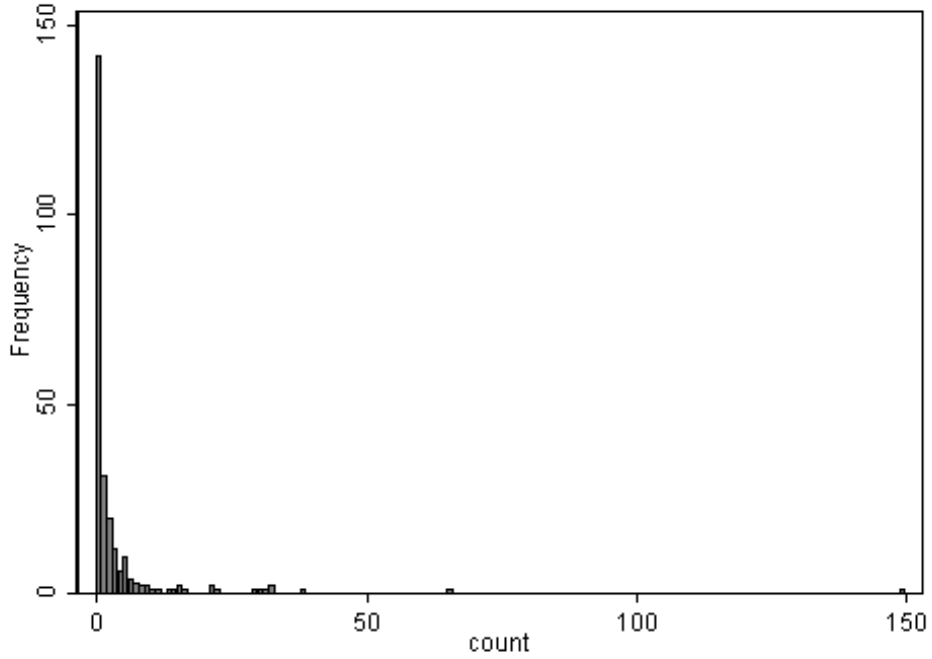


Image taken from [7]

$$\Pr(y_i = j) = \begin{cases} \pi_i + (1-\pi_i)g(y_i = 0) & \text{if j} = 0 \\ (1-\pi_i)g(y_i) & \text{if j} > 0 \end{cases}$$

- It is appropriate to model data with excessive zeros (overdispersion)
- It is a mixture of two statistical processes:
  - Always generates zero counts with probability π_i
  - Generates both zero and nonzero counts according to negative binomial distribution $g(y_i)$ with probability $1 - \pi_i$

An important variable for this talk: $\pi_i = \frac{\lambda_i}{1+\lambda_i}$

Measures used

- Harmonic centrality
- Zero-inflated negative binomial regression
  - The Poisson distribution better models counts of independent events; not suitable for memes
  - The negative binomial regression doesn’t model at its best model data with overdispersed zeros such as this
  - Separates the regression into two steps and estimates them at the same time:
    - One models the probability of the data point to be a zero (never diffused)
    - The other is the negative binomial component that estimates the number of diffusion events of a certain meme

Zero-inflated negative binomial regression

	Diffusions by meme	
	Negative Binomial positive coefficient = more diffusion events	**Zero-Inflated** negative coefficient = more likely to diffuse
Harmonic Centrality (standardized)	-0.125^{***}	-0.585^{***}
	(0.027)	(0.019)
Total Posts (log)	-0.105^{***}	0.134^{***}
	(0.011)	(0.007)
Active Members (log)	0.028^{***}	0.052^{***}
	(0.006)	(0.004)
(Intercept)	2.486^{***}	1.410^{***}
	(0.091)	(0.064)
Observations	906,481	
Log Likelihood	−105,898.500	

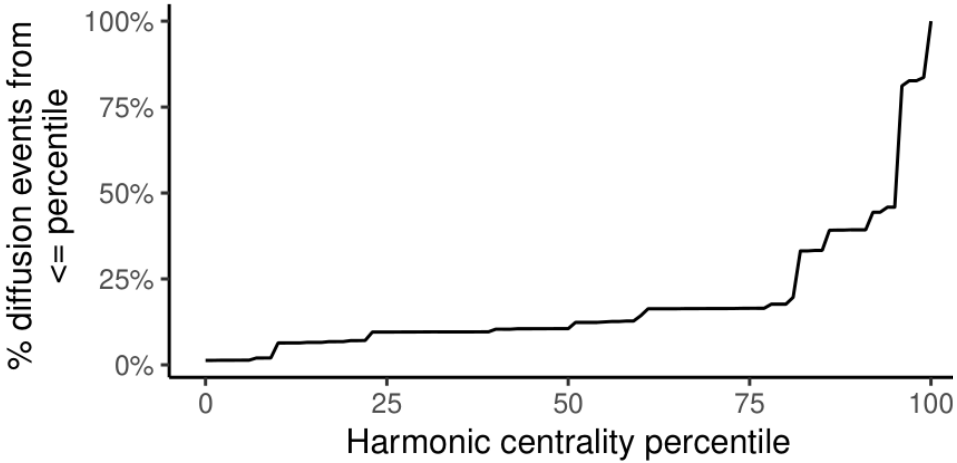
$^{*}p<0.05$; $^{**}p<0.01$; $^{***}p<0.001$

For a meme:

- Larger values for zero-inflated factor (to simplify π_i) more likely it predicts zero values (so less likely to diffuse)
- Larger coefficients for negative binomial factor (to simplify μ_i) means it predicts more diffusion events (so it diffuses farther)

"Memes originating from more popular communities are more likely to diffuse but diffuse less far." [2]

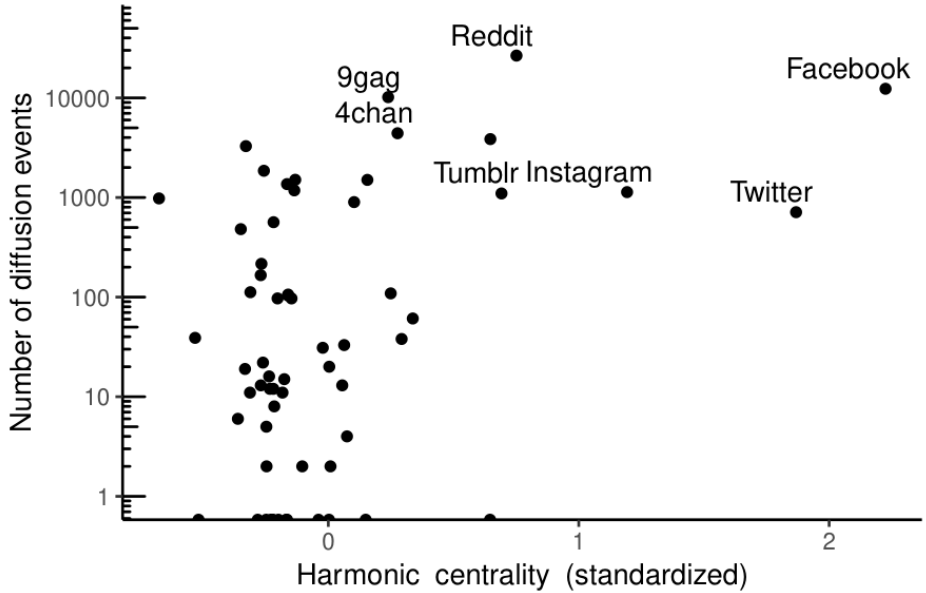
Harmonic centrality



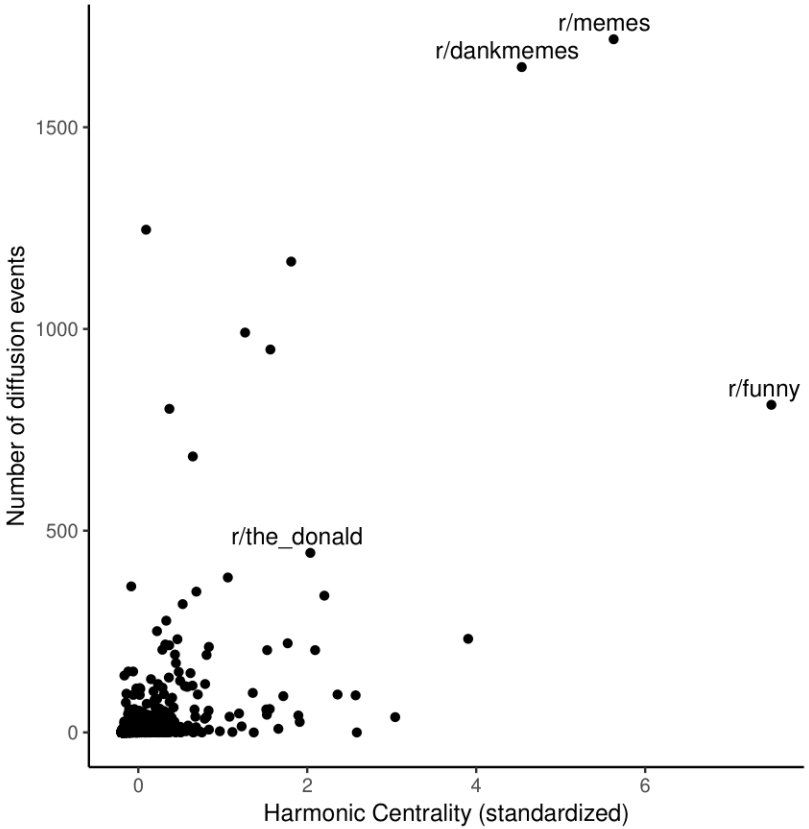
- 62% of memes originated from the top 10% communities by centrality.
- About 20% originate from the top 20%.
- About 18% originate from the lowest 70%.

"Communities with the highest harmonic centrality - including Reddit, Facebook, Twitter, and Instagram - were also among the most productive in terms of originating the most diffusion events." [2]

Reddit and Facebook are the principal cores with the most diffused memes



Analysis inside Reddit



- Same analysis but inside the Reddit subnetwork
- “Similar to our cross-community results, the majority of memes that diffused between sub-communities on Reddit originated in the core of the Reddit network, rather than in intermediate or peripheral subcommunities.” [2]

- 69% of memes originated from the top 10% communities by centrality.
- About 21% originate from the top 20%.
- About 10% originate from the lowest 70%.

Zero-inflated negative binomial regression for Reddit

Diffusions by meme on Reddit

	Negative Binomial positive coefficient = more diffusion events	**Zero-Inflated** negative coefficient = more likely to diffuse
Harmonic Centrality (standardized)	0.364^{***}	0.288^{***}
	(0.081)	(0.053)
log(OC posts)	-1.473^{***}	3.071^{***}
	(0.358)	(0.198)
log(total posts)	1.820^{***}	-2.856^{***}
	(0.345)	(0.190)
log(users)	-0.339^{**}	-0.313^{***}
	(0.041)	(0.027)
(Intercept)	1.081^{***}	5.166^{***}
	(0.291)	(0.195)
Observations	371,113	
Log Likelihood	−37,593.440	

$^{*}p<0.05$; $^{**}p<0.01$; $^{***}p<0.001$

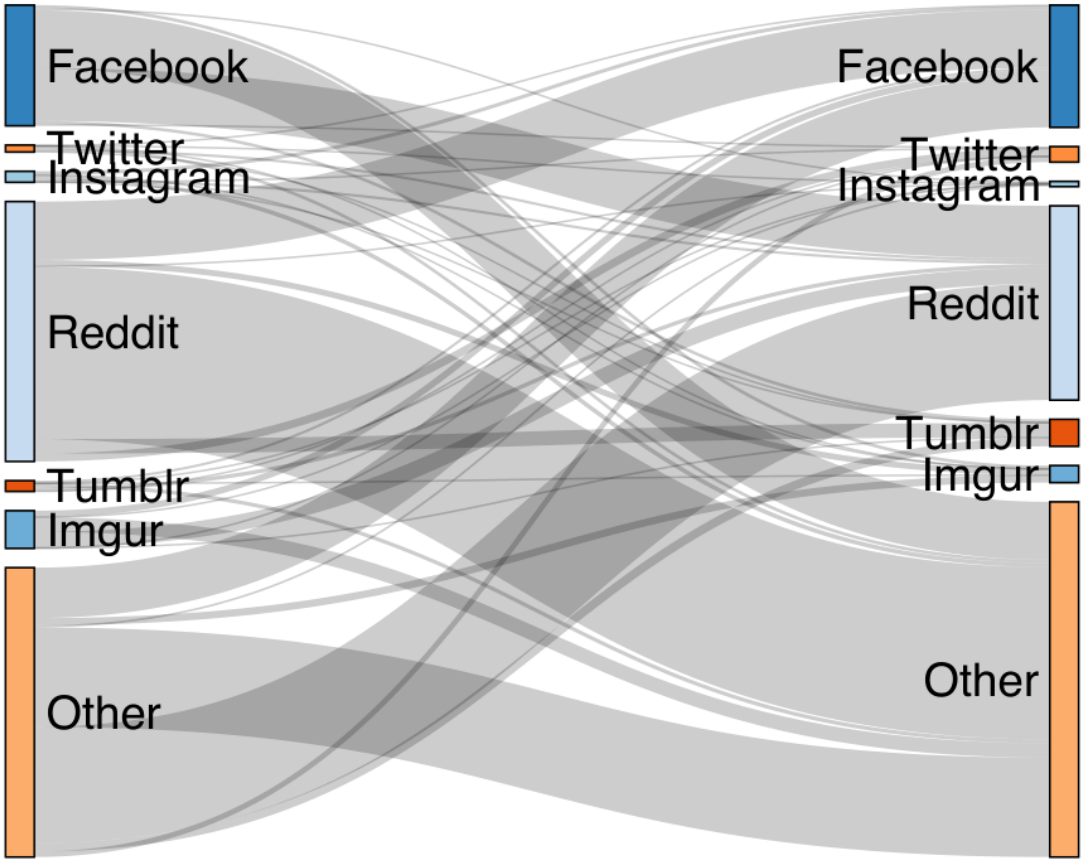
"Memes originating from more central subcommunities on Reddit are less likely to diffuse, but diffuse farther." [2]

Recap

Negative Binomial		**Zero-Inflated**	
positive coefficient = more diffusion events		negative coefficient = more likely to diffuse	
Web-scale analysis	Single-community analysis	Web-scale analysis	Single-community analysis
Negative	Positive	Negative	Positive

For this reason it is important to make a web-scale analysis rather than settling for a single community

Flows of original content

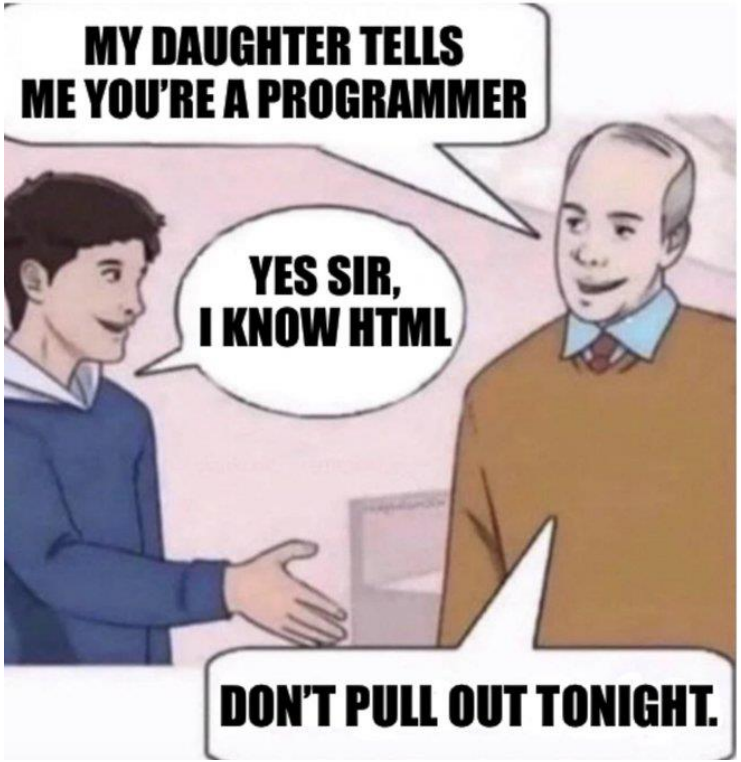


- About 2/3 of the non-original memes posted in intermediate and peripheral communities were originated in Facebook and Reddit.
- About 1/2 of the non-original memes posted on Facebook and Reddit originates in non-core communities.
- “The core communities do post a non-trivial amount of content that was originated in intermediate and peripheral communities, but other content originated in other core communities is still the largest representation.” [2]

Limitations of this study

- For privacy reasons it wasn’t possible to index private accounts and groups, such as WeChat, Facebook groups and Discord channels (unless some content has been archived, like 4chan).
- Not majority memes communities weren’t tracked, otherwise the analysis could include images spreading that are not memes (ex. News). The exceptions are r/the_donald, /b/ and /pol/ wich didn’t originate memes widely diffused.
- Textless and non-English memes were removed for the reason at the previous point
- It is not possible to find the cause-effect correlation between the size of a community and the ability to spread its memes
- The researchers can further investigate the role of the communities that are peripheral within popular platforms and in the web-scale. Both did not originate a large proportion of the overall image meme diffusions.

Personal considerations: evolution of texts



Positive figure

They are not
the same meme

Negative figure


"That's all Folks!"

References

[1] Valensise, C.M., Serra, A., Galeazzi, A. *et al.* Entropy and complexity unveil the landscape of memes evolution. *Sci Rep* **11**, 20022 (2021). https://doi.org/10.1038/s41598-021-99468-6

[2] Morina, D. and Bernstein, M.S. (2022) 'A web-scale analysis of the community origins of image memes', *Proceedings of the ACM on Human-Computer Interaction*, 6(CSCW1), pp. 1–25. doi:10.1145/3512921.

[3] Hokky Situngkir, 2004. "**On Selfish Memes: culture as complex adaptive system**," Computational Economics 0405001, University Library of Munich, Germany.

[4] Poisson distribution - Encyclopedia of Mathematics. (n.d.). https://encyclopediaofmath.org/wiki/Poisson_distribution

[5] View of The Negative Binomial regression | The Southwest Respiratory and Critical Care Chronicles. (n.d.). https://pulmonarychronicles.com/index.php/pulmonarychronicles/article/view/200/491

[6] https://www.ncss.com/wp-content/themes/ncss/pdf/Procedures/NCSS/Zero-Inflated_Negative_Binomial_Regression.pdf

[7] https://stats.oarc.ucla.edu/stata/dae/zero-inflated-negative-binomial-regression/

Thank you for
your attention