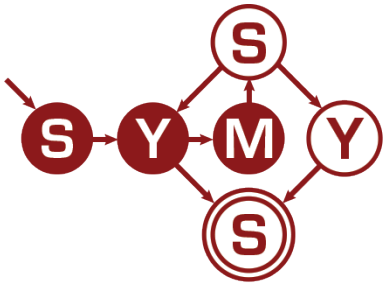
The Message or the Messenger? Inferring Virality and Diffusion Structure from Online Petition Signature Data

Chi Ling Chan, Justin Lai, Bryan Hooi*,
Todd Davies

Stanford University

* Carnegie Mellon University



STANFORD UNIVERSITY
SYMBOLIC SYSTEMS PROGRAM

SocInfo 2017, Oxford, UK

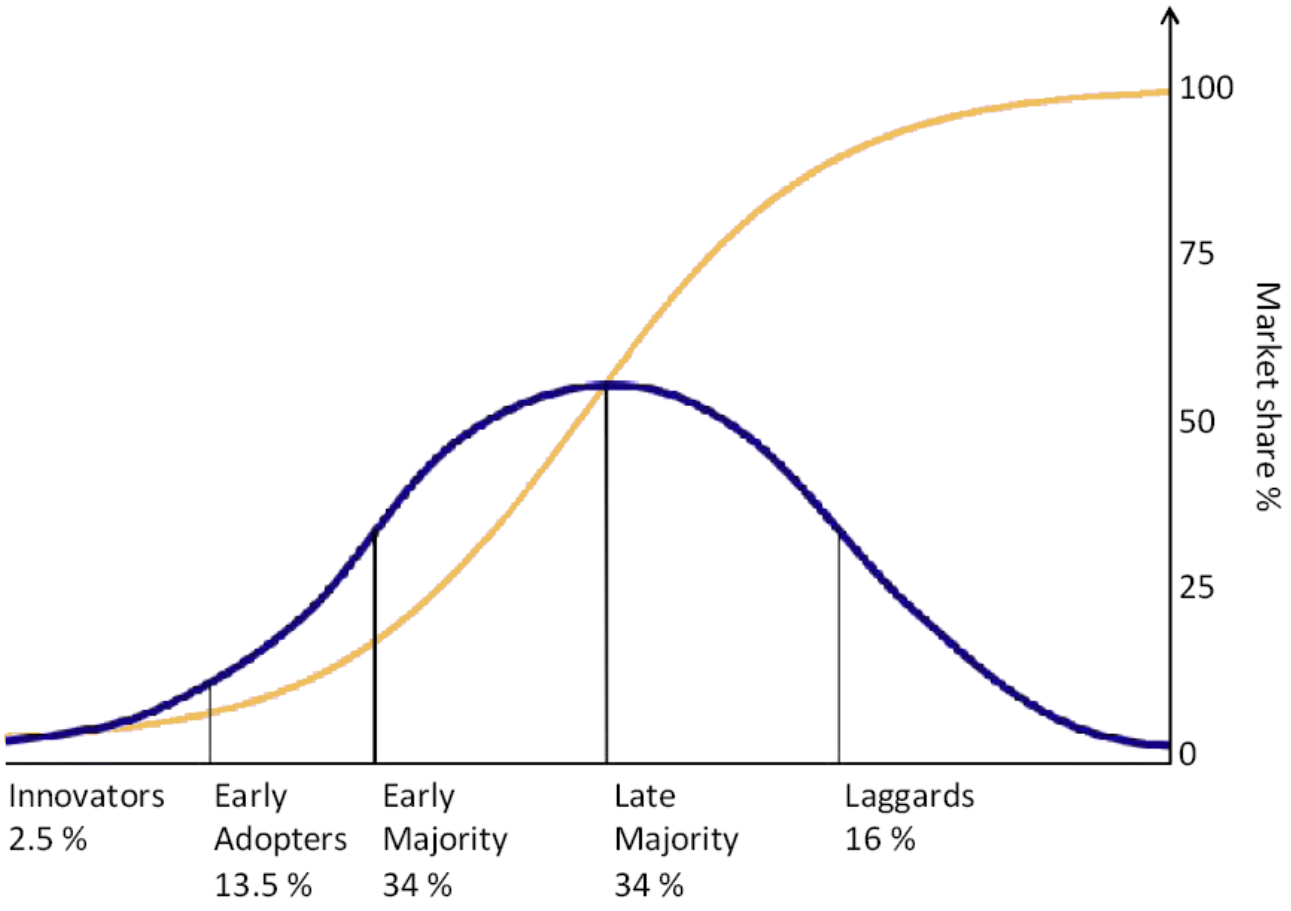
September 15, 2017

CSLI

From https://upload.wikimedia.org/wikipedia/commons/7/72/Duncan_Watts.jpg

Viral marketing
(Faberge shampoo ad, 1982)

The classical adoption pattern



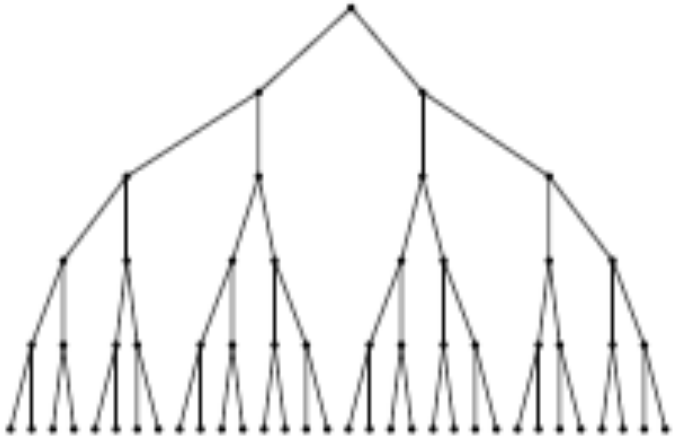
Graph from
https://commons.wikimedia.org/wiki/File:Diffusionofideas.PNG

Schematic diffusion patterns

Broadcast

Messenger is important

Viral



Message is important (?)

Fig. 1 from Goel et al. (2016), "The Structural Virality of Online Diffusion"
(https://cs.stanford.edu/people/ashton/pubs/twiral.pdf)

Structural virality as the Wiener index

(Goel, Anderson, Hofman, & Watts 2016)

$$\nu(T) = \frac{1}{n(n-1)} \sum_{i=1}^{n} \sum_{j=1}^{n} d_{ij}$$

$v(T)$ = the average distance between all pairs of nodes in a diffusion tree T (or, equivalently, the average depth of all nodes as roots)

for $n > 1$ nodes

d_{ij} = the shortest distance between nodes i and j

Random Twitter cascades ordered by structural virality

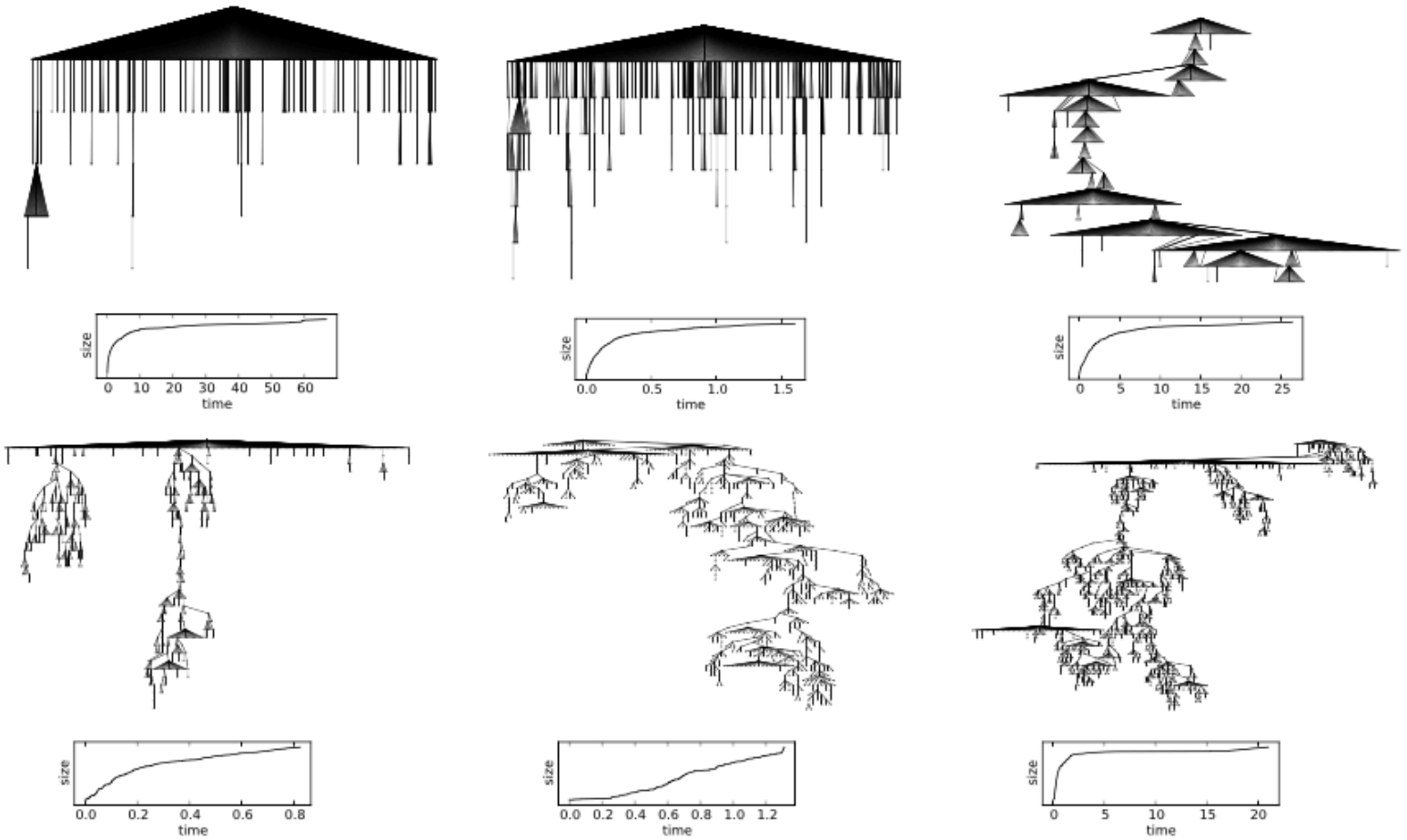


Fig. 3 from "The Structural Virality of Online Diffusion"
(https://cs.stanford.edu/people/ashton/pubs/twiral.pdf)

Structural virality by cascade size/popularity on Twitter, per domain

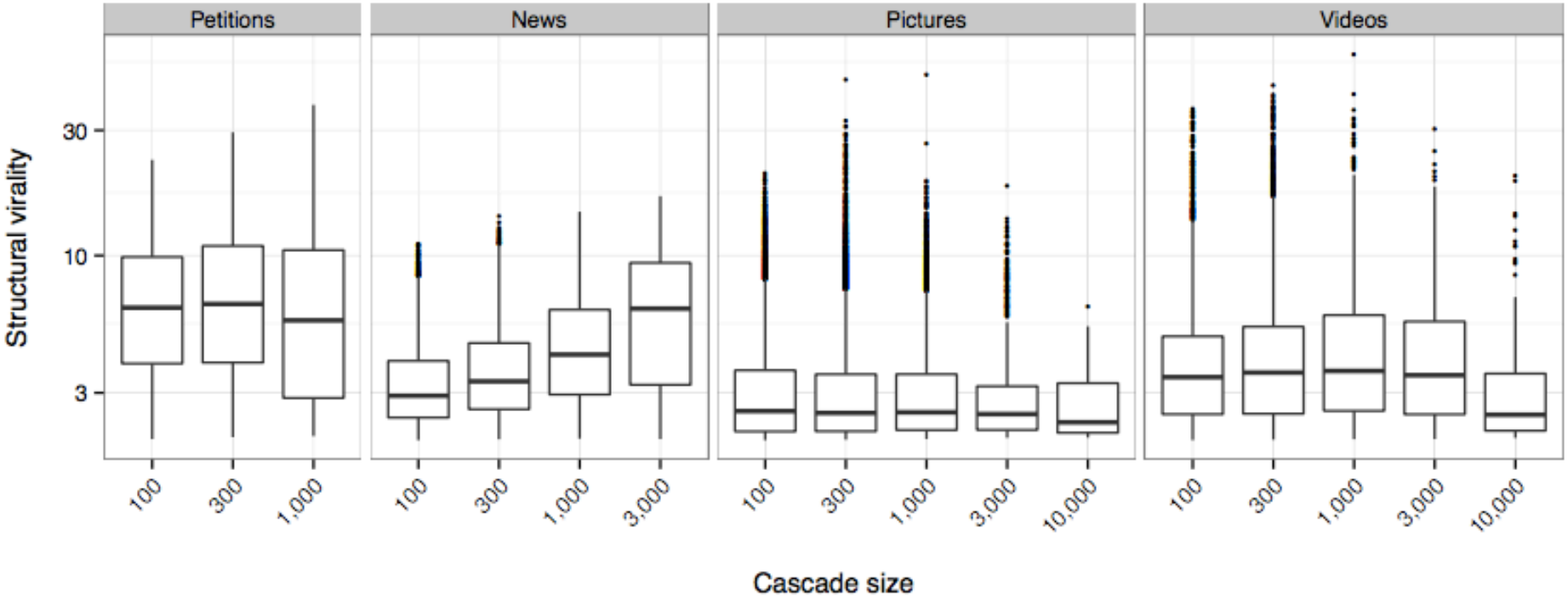


Fig. 5 from "The Structural Virality of Online Diffusion" (https://cs.stanford.edu/people/ashton/pubs/twiral.pdf)

Correlation between popularity and structural virality for 4 domains

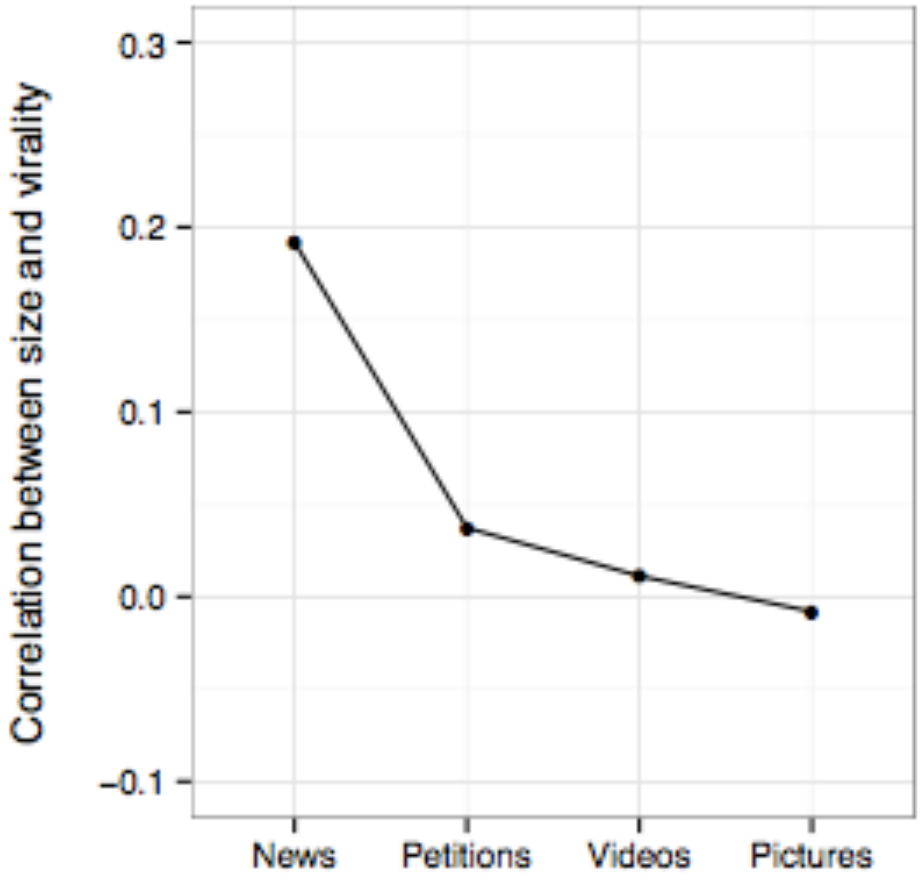


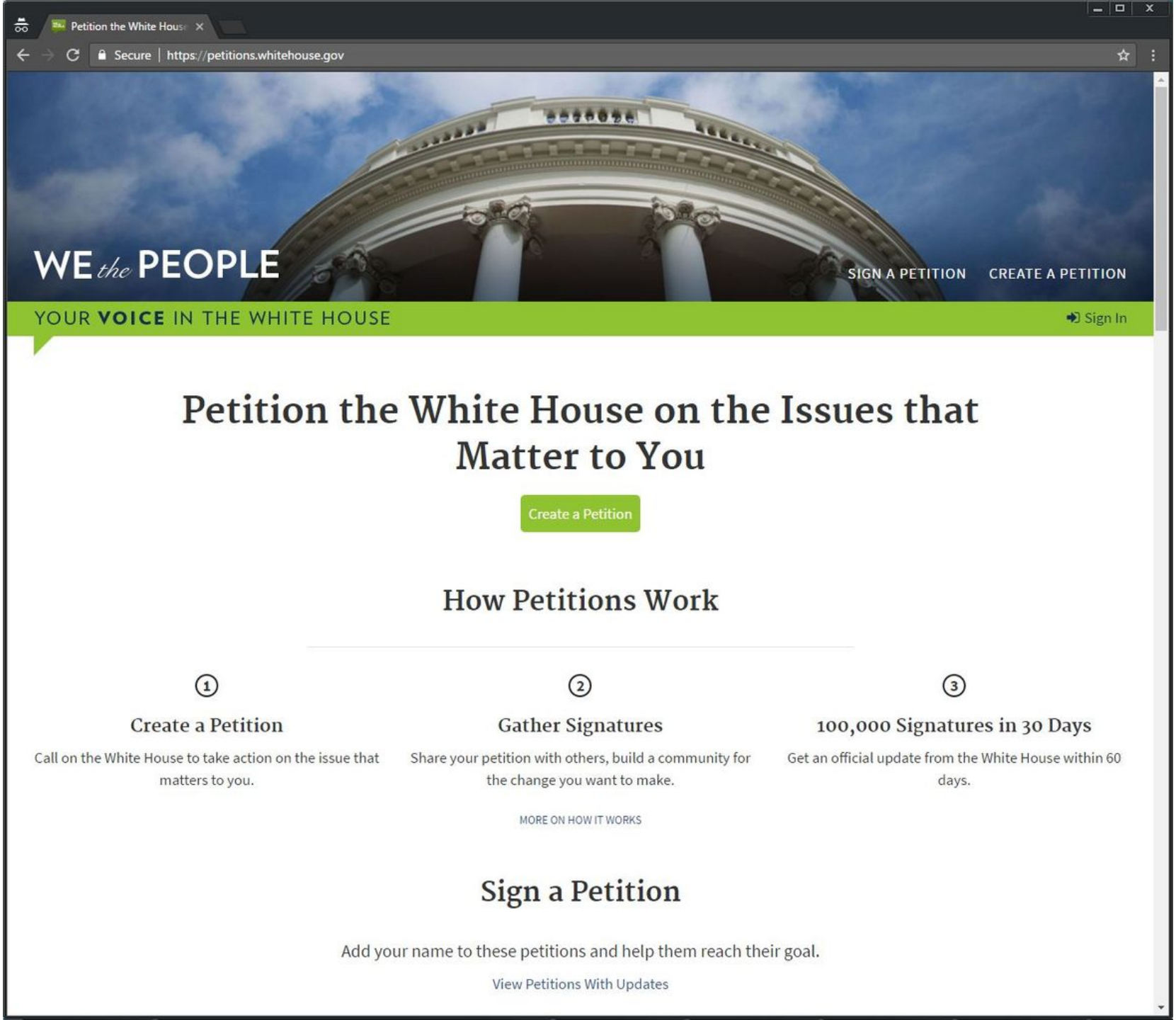
Fig. 6 from "The Structural Virality of Online Diffusion" (https://cs.stanford.edu/people/ashton/pubs/twiral.pdf)

Structural virality versus intrinsic virality ('infectiousness')

Main model in Goel et al. (2016) assumes constant infectiousness (intrinsic appeal).

They say: "In other words, taking infectiousness as a proxy for quality, in our simulations the largest and most viral cascades are not inherently better than those that fail to gain traction, but are simply more fortunate (Watts 2002)."

So structural virality does not imply intrinsic virality/infectiousness.


Petition the White House
Secure | https://petitions.whitehouse.gov
WE the PEOPLE
SIGN A PETITION
CREATE A PETITION
YOUR VOICE IN THE WHITE HOUSE
Sign In
Petition the White House on the Issues that Matter to You
Create a Petition
How Petitions Work
1
Create a Petition
Call on the White House to take action on the issue that matters to you.
2
Gather Signatures
Share your petition with others, build a community for the change you want to make.
3
100,000 Signatures in 30 Days
Get an official update from the White House within 60 days.
MORE ON HOW IT WORKS
Sign a Petition
Add your name to these petitions and help them reach their goal.
View Petitions With Updates

Questions about petitions

Can we infer structural virality (or “broadcastness”) just from time-stamped signature data?

Are successful petitions on We The People more structurally viral than failed ones?

Is petition success predicted by infectiousness/intrinsic virality?

Do actual petition signature data show patterns at odds with what research using Twitter cascades would suggest?

A few other previous findings

First day signature total is very predictive of petition popularity/success on the No. 10 Downing Street petition site (Hale, Margetts, & Yasseri 2013)

Successful petitions on The Petition Site gather a large fraction of their signatures early on (Proskurnia et al. 2017)

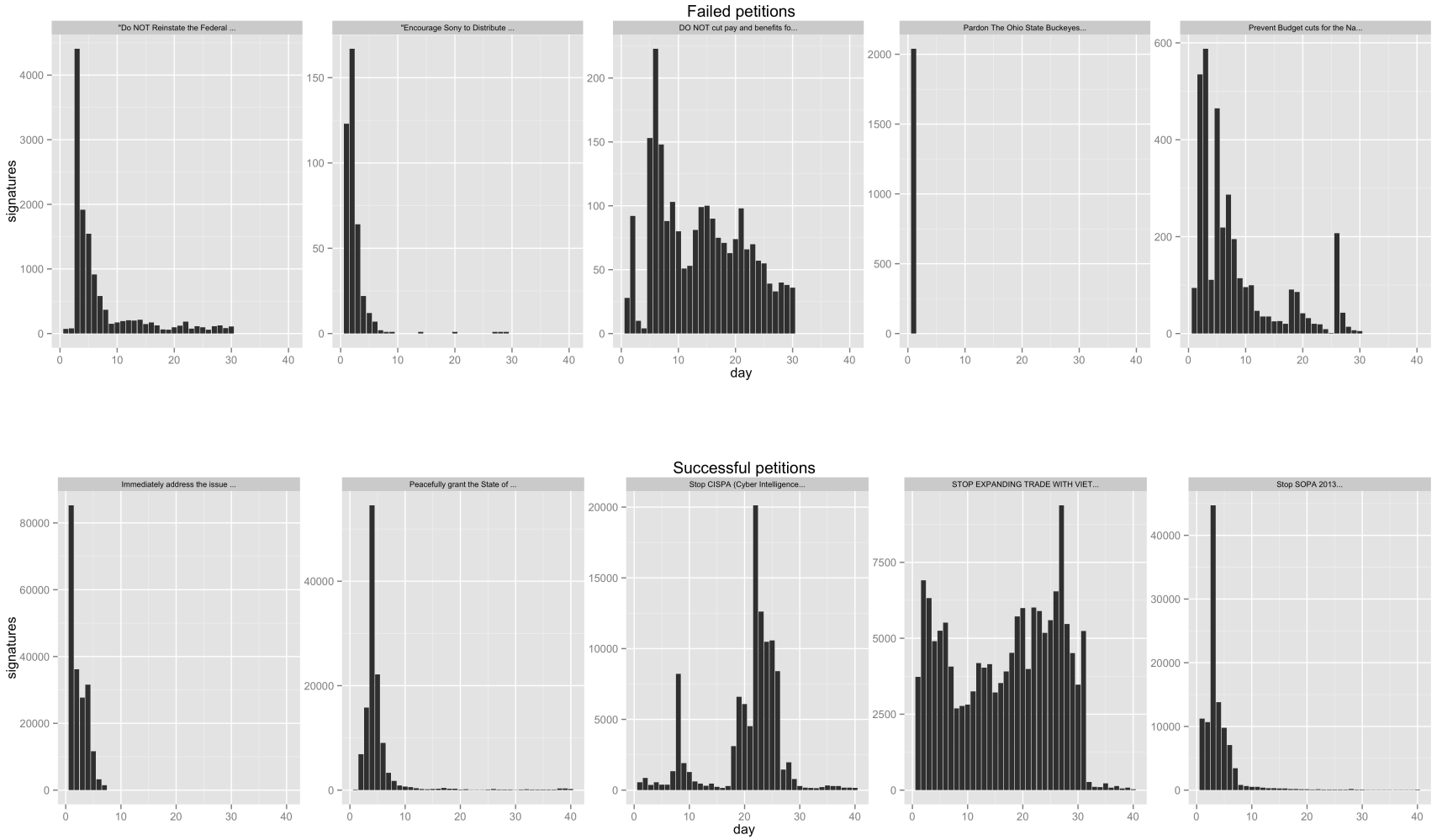
Successful/popular petitions are rare (many studies)

Data characterization

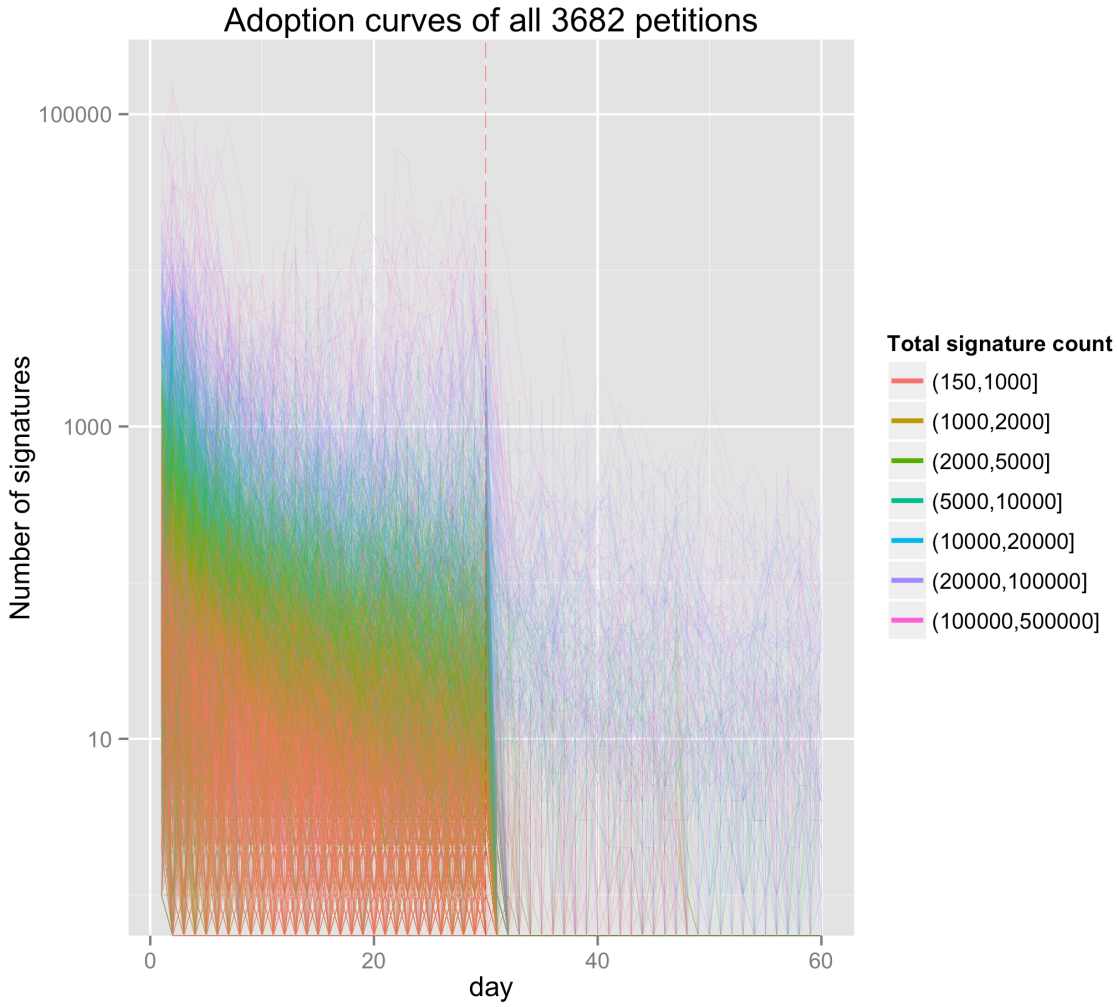
3682 WTP petitions collected between Sept. 20, 2011 and March 30, 2015

59 (1.6%) reached the signature threshold for a White House response

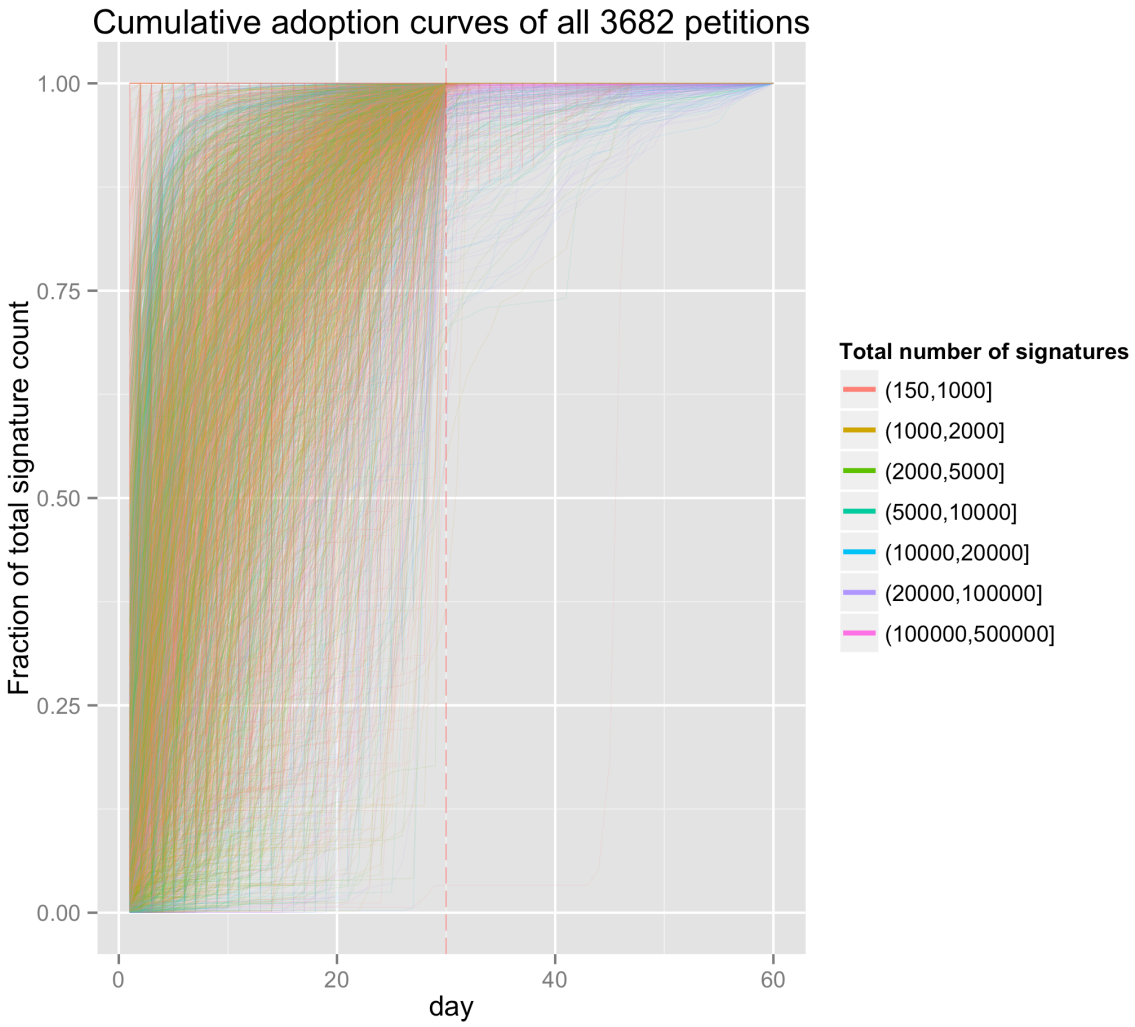
Signature graphs for randomly chosen failed vs. successful petitions



Day-by-day signature counts for petitions of different final popularities



Cumulative adoption curves for petitions of different popularities



Exceed ratios: inverse indicators of ***structural*** virality

- *Total exceed ratio* (an inverse measure of structural virality)

$$E_{Tot} = \frac{\Sigma_{i \in L}(S(i) - \max[S(i-1), S(i+1)])}{\Sigma_{i=1}^{T} S(i)}$$

for a given petition over *T* time periods, in which *S(i)* signatures are obtained in period *i*, and *L* is the set of all peak periods within *T*

- *Global-peak-only exceed ratio* E_{GPO} = adjacent-periods signature difference for just the global peak period divided by total signatures (an indicator of the largest broadcast event)

First day/second day (FDSD) ratio: an indicator of ***intrinsic*** virality

Assumptions:

- Most petitions are launched by some kind of broadcast event on the first day
- Therefore, petitions that achieve more signatures on the second day than on the first day will be, on average, higher in intrinsic appeal than those with higher FDSD ratios

Average total exceed ratio E_{Tot} for all petitions: successful versus unsuccessful

	Successful (N = 59)	Unsuccessful (N = 3623)
Daily	0.152 (sd = .13)	0.224 (sd = .04)
Hourly	0.148 (sd = .09)	0.230 (sd = .03)

Failed petitions were 47.4% higher for daily total exceed ratio, and 55.4% higher for hourly ($p < .0001$ for both)

Daily global-peak-only exceed ratio E_{GPO} was 0.105 (sd=.11) for successful and 0.155 (sd=.19) for unsuccessful petitions ($p = .042$).

> Cf. Goel et al., 2016: "If popularity is consistently related to any one feature, it is the size of the largest broadcast event."

FDSD Ratio: Testing for intrinsic virality

Percentage of petitions with more signatures on the second day than on the first day

- Successful: 68% (N=59)
- Unsuccessful: 38% (N=3623)

($p < .00001$ by Chi-square)

Measures of shape [with type of virality measured]

Measure of Shape	Interpretation
Skewness	Whether distribution has larger 'tails' extending to right (positive) [IV]
Kurtosis	How peaked a distribution is [SV]
Location of global peak	The day on which the petition received the most signatures [IV]
Number of local peaks	Number of days on which the petition received more signatures than on adjacent days [SV]

All these measures indicate higher structural and intrinsic virality for more popular petitions in the WTP data set.

Theoretical model: highlights

First broadcast event on day 1

Variable infectiousness for each petition (basic reproduction number R_0 = average number of signers in next period for each signer in present period): message strength

Constant average broadcast size X for all petitions after first broadcast: messenger strength

Simulation over 5000 petitions replicates qualitative patterns observed for regression of signature totals on measures of shape

Summary

Analysis of We the People temporal signature data suggests more popular/successful petitions are higher in both structural and intrinsic virality than less popular/unsuccessful petitions, on all the measures chosen as indicators for SV and IV.

Our measure E_{GPO} indicates that successful petitions are less likely to depend on a single large broadcast event than unsuccessful ones for their signature totals.

Simulations support a model of petition signing in which intrinsic virality/infectiousness varies across petitions.

Further work…

More refined model of individual petition decisions to produce exceed ratio and FDSD results

Looking at spatial data/location stamps

Finally…

Thanks to
- Marek Hlavac
- Lee Ross
- Howard Rheingold

Data and code are available at
https://github.com/justinlai/petitiondata

Comments welcome: davies@stanford.edu