Class 17: Going viral

Matthew J. Salganik

Sociology 204: Social Networks
Princeton University

1/2 What does viral look like?

What does it mean to go viral?

grilledcheesesocial grilled cheese social · 1-28

Baked feta pasta viral recipe! Inspired by **#uunifeta** via **@liemessa** & **@tiiupiret** **#learnontiktok** **#foodtiktok** **#foodie**

♫ **She Share Story (for Vlog) - 山口夕依**

279.8K

1495

78.7K

`https://www.tiktok.com/@grilledcheesesocial/video/6922946148760030469`

Similarities between the papers

- Both papers deal with a similar empirical phenomena and both struggle to figure out what is the **right question**

Similarities between the papers

- Both papers deal with a similar empirical phenomena and both struggle to figure out what is the **right question**
- Both papers are in Pasteur's quadrant (motivated by use and seeking fundamental understanding)

Quest for fundamental understanding?	Consideration of use? No	Consideration of use? Yes
Yes	Pure basic research (Bohr)	Use-inspired basic research (Pasteur)
No		Pure applied research (Edison)

For more information, see Salganik (2018): `https://www.bitbybitbook.com/en/1st-ed/running-experiments/making/partner/`

- Both papers deal with a similar empirical phenomena and both struggle to figure out what is the **right question**
- Both papers are in Pasteur's quadrant (motivated by use and seeking fundamental understanding)
- Both papers include small and big cascades offering a systematic approach

- Both papers deal with a similar empirical phenomena and both struggle to figure out what is the **right question**
- Both papers are in Pasteur's quadrant (motivated by use and seeking fundamental understanding)
- Both papers include small and big cascades offering a systematic approach
- Both papers require data that was not possible until recently

- Both papers deal with a similar empirical phenomena and both struggle to figure out what is the **right question**
- Both papers are in Pasteur's quadrant (motivated by use and seeking fundamental understanding)
- Both papers include small and big cascades offering a systematic approach
- Both papers require data that was not possible until recently
- The papers end up with different ways of approaching the problem: descriptive vs predictive

The Structural Virality of Online Diffusion

Sharad Goel, Ashton Anderson
Stanford University, Stanford, California, 94305 {scgoel@stanford.edu, ashton@cs.stanford.edu}

Jake Hofman, Duncan J. Watts
Microsoft Research, New York, New York 10016 {jmh@microsoft.com, duncan@microsoft.com}

What is virality?

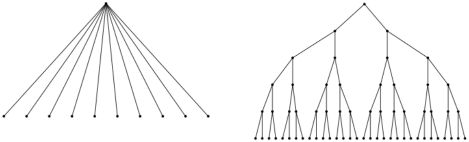


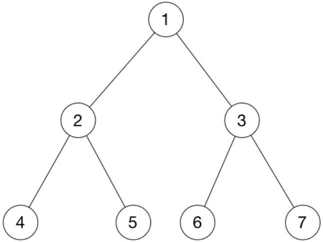
Figure 1 **A schematic depiction of broadcast versus viral diffusion, where nodes represent individual adoptions and edges indicate who adopted from whom.**

Wiener index (from chemistry):

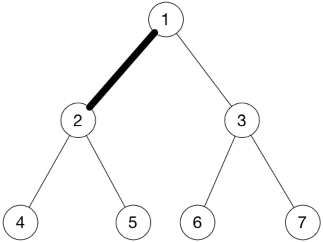
$$\nu(T) = \frac{1}{n(n-1)} \sum_{i=1}^{n} \sum_{j=1}^{n} d_{ij}$$

where $d_{i,j}$ is the length of the shortest path between i and j

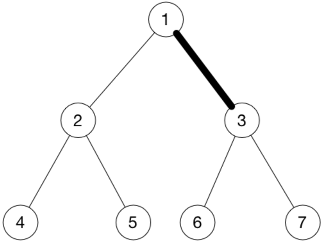
In other words, expected path length between two randomly chosen points



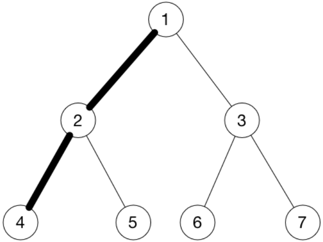
	1	2	3	4	5	6	7
1	0						
2							
3							
4							
5							
6							
7							



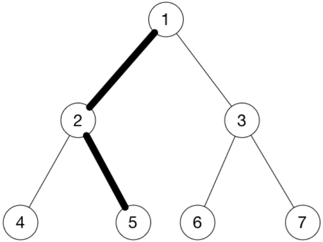
	1	2	3	4	5	6	7
1	0	1					
2							
3							
4							
5							
6							
7							



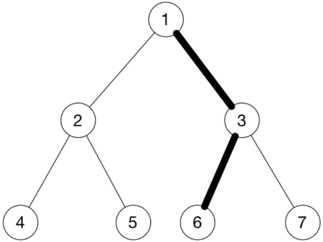
	1	2	3	4	5	6	7
1	0	1	1				
2							
3							
4							
5							
6							
7							



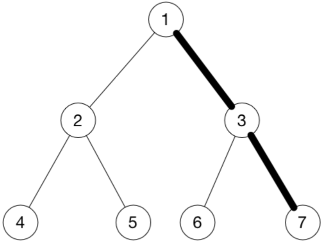
	1	2	3	4	5	6	7
1	0	1	1	2			
2							
3							
4							
5							
6							
7							



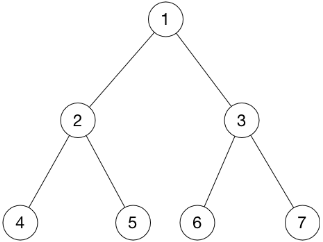
	1	2	3	4	5	6	7
1	0	1	1	2	2		
2							
3							
4							
5							
6							
7							



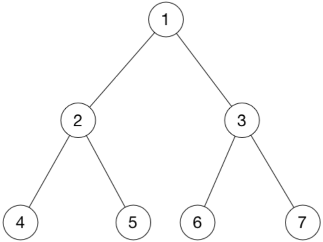
	1	2	3	4	5	6	7
1	0	1	1	2	2	2	
2							
3							
4							
5							
6							
7							



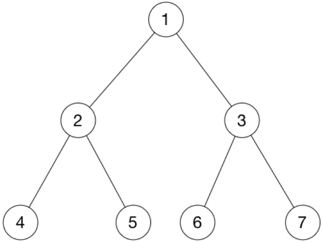
	1	2	3	4	5	6	7
1	0	1	1	2	2	2	2
2							
3							
4							
5							
6							
7							



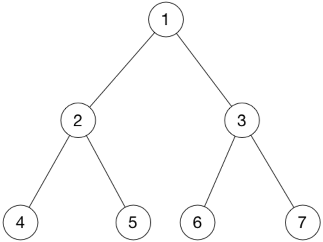
	1	2	3	4	5	6	7
1	0	1	1	2	2	2	2
2	1	0	2	1	1	3	3
3							
4							
5							
6							
7							



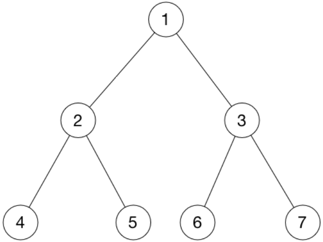
	1	2	3	4	5	6	7
1	0	1	1	2	2	2	2
2	1	0	2	1	1	3	3
3	1	2	0	3	3	1	1
4							
5							
6							
7							



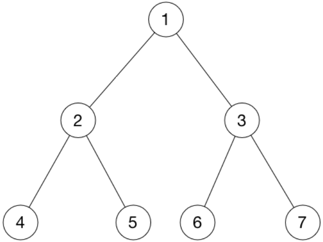
	1	2	3	4	5	6	7
1	0	1	1	2	2	2	2
2	1	0	2	1	1	3	3
3	1	2	0	3	3	1	1
4	2	1	3	0	2	4	4
5							
6							
7							



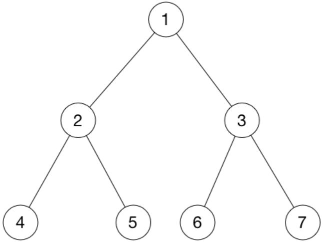
	1	2	3	4	5	6	7
1	0	1	1	2	2	2	2
2	1	0	2	1	1	3	3
3	1	2	0	3	3	1	1
4	2	1	3	0	2	4	4
5	2	1	3	2	0	4	4
6							
7							



	1	2	3	4	5	6	7
1	0	1	1	2	2	2	2
2	1	0	2	1	1	3	3
3	1	2	0	3	3	1	1
4	2	1	3	0	2	4	4
5	2	1	3	2	0	4	4
6	2	3	1	4	4	0	2
7							

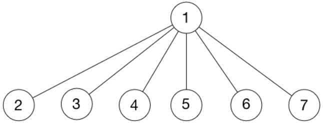


	1	2	3	4	5	6	7
1	0	1	1	2	2	2	2
2	1	0	2	1	1	3	3
3	1	2	0	3	3	1	1
4	2	1	3	0	2	4	4
5	2	1	3	2	0	4	4
6	2	3	1	4	4	0	2
7	2	3	1	4	4	2	0

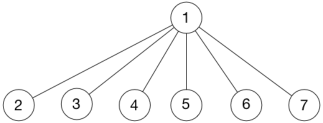


	1	2	3	4	5	6	7
1	0	1	1	2	2	2	2
2	1	0	2	1	1	3	3
3	1	2	0	3	3	1	1
4	2	1	3	0	2	4	4
5	2	1	3	2	0	4	4
6	2	3	1	4	4	0	2
7	2	3	1	4	4	2	0

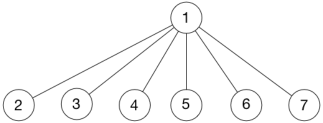
$$\nu(T) \approx 2.29$$



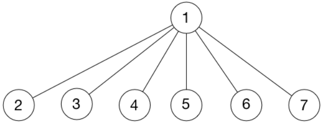
	1	2	3	4	5	6	7
1	0						
2							
3							
4							
5							
6							
7							



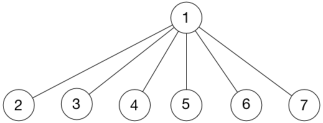
	1	2	3	4	5	6	7
1	0	1					
2							
3							
4							
5							
6							
7							



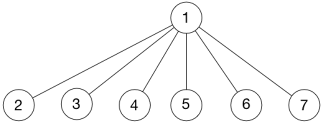
	1	2	3	4	5	6	7
1	0	1	1				
2							
3							
4							
5							
6							
7							



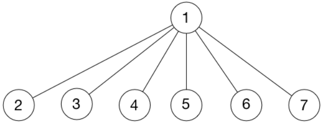
	1	2	3	4	5	6	7
1	0	1	1	1			
2							
3							
4							
5							
6							
7							



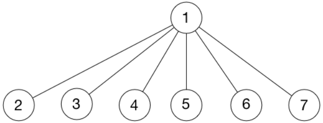
	1	2	3	4	5	6	7
1	0	1	1	1	1		
2							
3							
4							
5							
6							
7							



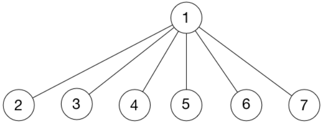
	1	2	3	4	5	6	7
1	0	1	1	1	1	1	
2							
3							
4							
5							
6							
7							



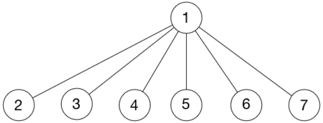
	1	2	3	4	5	6	7
1	0	1	1	1	1	1	1
2							
3							
4							
5							
6							
7							



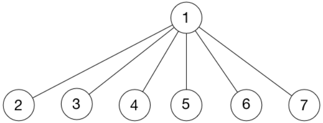
	1	2	3	4	5	6	7
1	0	1	1	1	1	1	1
2	1	0	2	2	2	2	2
3							
4							
5							
6							
7							



	1	2	3	4	5	6	7
1	0	1	1	1	1	1	1
2	1	0	2	2	2	2	2
3	1	2	0	2	2	2	2
4							
5							
6							
7							

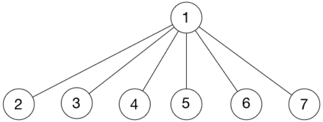


	1	2	3	4	5	6	7
1	0	1	1	1	1	1	1
2	1	0	2	2	2	2	2
3	1	2	0	2	2	2	2
4	1	2	2	0	2	2	2
5	1	2	2	2	0	2	2
6	1	2	2	2	2	0	2
7	1	2	2	2	2	2	0

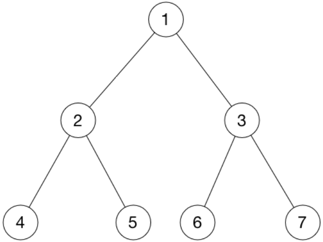


	1	2	3	4	5	6	7
1	0	1	1	1	1	1	1
2	1	0	2	2	2	2	2
3	1	2	0	2	2	2	2
4	1	2	2	0	2	2	2
5	1	2	2	2	0	2	2
6	1	2	2	2	2	0	2
7	1	2	2	2	2	2	0

$$\nu(T) \approx 1.71$$



$\nu(T) \approx 1.71$



$\nu(T) \approx 2.29$

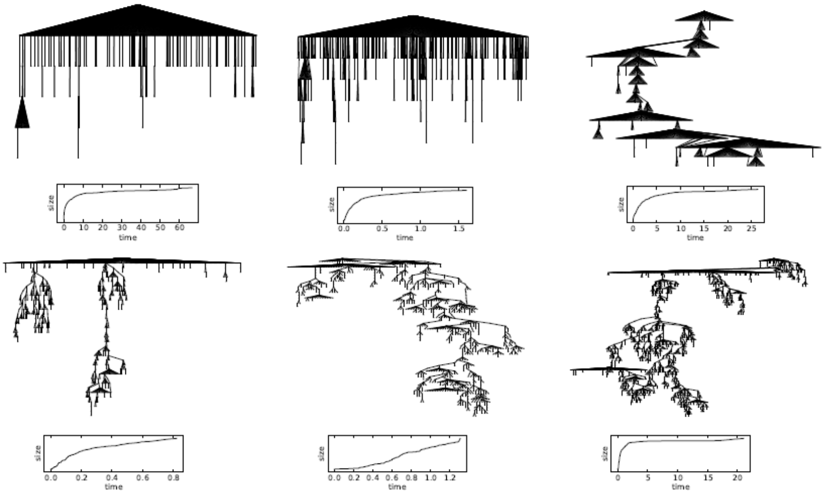


Figure 3 A random sample of cascades stratified and ordered by increasing structural virality, ranging from 2 to 50. For ease of visualization, cascades were restricted to having between 100 and 1000 adopters. Cumulative adoption curves (i.e., total cascade size over time) are shown below each cascade, with time indicated in hours.

describing outcomes vs describing generative process

What do viral cascades look like?

- 622 million unique pieces of content (links) shared via Twitter
- 1.2 billion adoptions (posting of content)
- videos, images, news stories, and petitions

"Big data" is needed because large cascades are very, very rare.

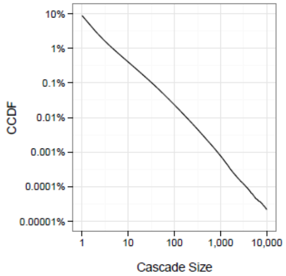


Figure 2 Distribution of cascade sizes on a log-log scale, aggregated across the four domains we study: videos, news, pictures, and petitions.

- Most things don't grow (99% of adoptions are accounted for by the root node and the immediate followers of the root node)

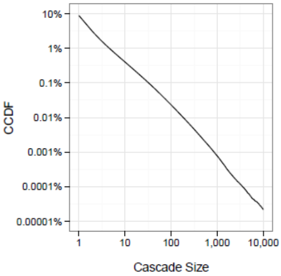


Figure 2 **Distribution of cascade sizes on a log-log scale, aggregated across the four domains we study: videos, news, pictures, and petitions.**

- Most things don't grow (99% of adoptions are accounted for by the root node and the immediate followers of the root node)
- They focus on the cascades that include at least 100 nodes (1 in 4,000 events).

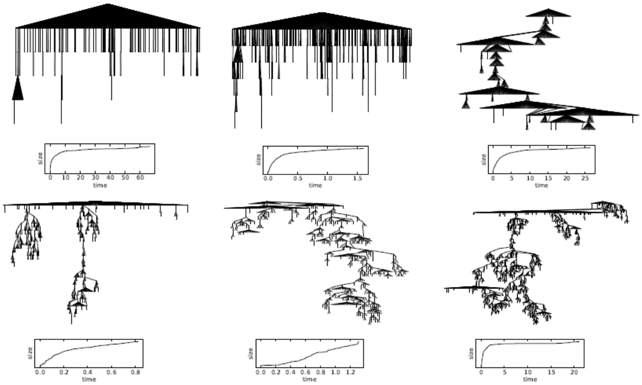


Figure 3 A random sample of cascades stratified and ordered by increasing structural virality, ranging from 2 to 50. For ease of visualization, cascades were restricted to having between 100 and 1000 adopters. Cumulative adoption curves (i.e., total cascade size over time) are shown below each cascade, with time indicated in hours.

- Examples of different structural virality

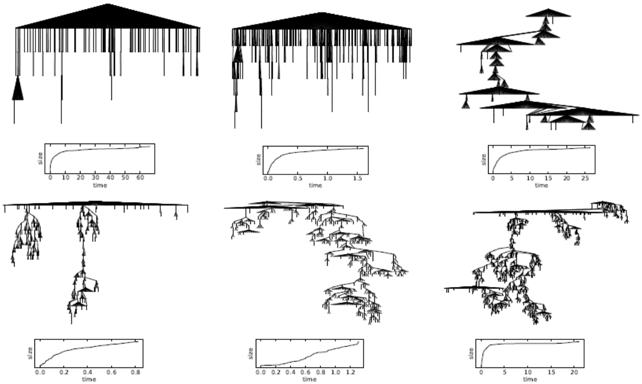


Figure 3 A random sample of cascades stratified and ordered by increasing structural virality, ranging from 2 to 50. For ease of visualization, cascades were restricted to having between 100 and 1000 adopters. Cumulative adoption curves (i.e., total cascade size over time) are shown below each cascade, with time indicated in hours.

- Examples of different structural virality
- Structural virality captures something different from speed of adoption and diffusion curves

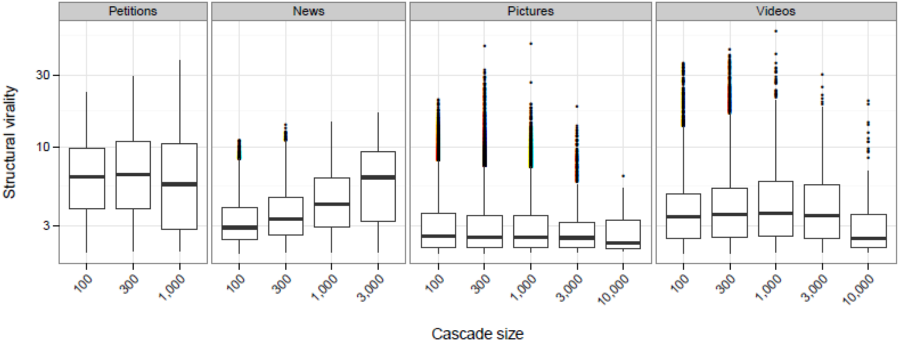


Figure 5 **Boxplot of structural virality by size on a log-log scale, separated by domain. Lines inside the boxes indicate median structural virality, while the boxes themselves show interquartile ranges.**

Knowing the size of a cascades reveals little about structural virality. This is true for all 4 types of content (but a bit less true for news).

What combination of spreading process and network structure is consistent with these results?

What combination of spreading process and network structure is consistent with these results?
SIR model on network with power law degree distribution

How might the ideas in this paper be used?

https://www.youtube.com/watch?v=wSwOszoHuoI

Now we have a sense of what cascades can look like, but can they be predicted?