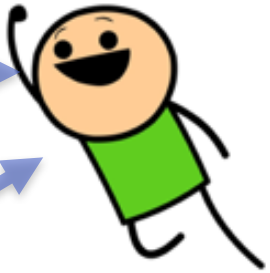
Lost in *propagation*? Unfolding News Cycles from the Source

Chenhao Tan, Adrien Friggeri, Lada Adamic
Cornell University, Facebook

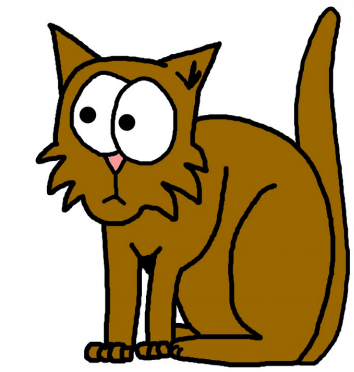
News reaches us via different ways

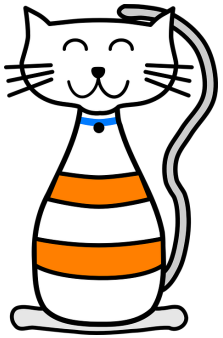
The New York Times

The Washington Post

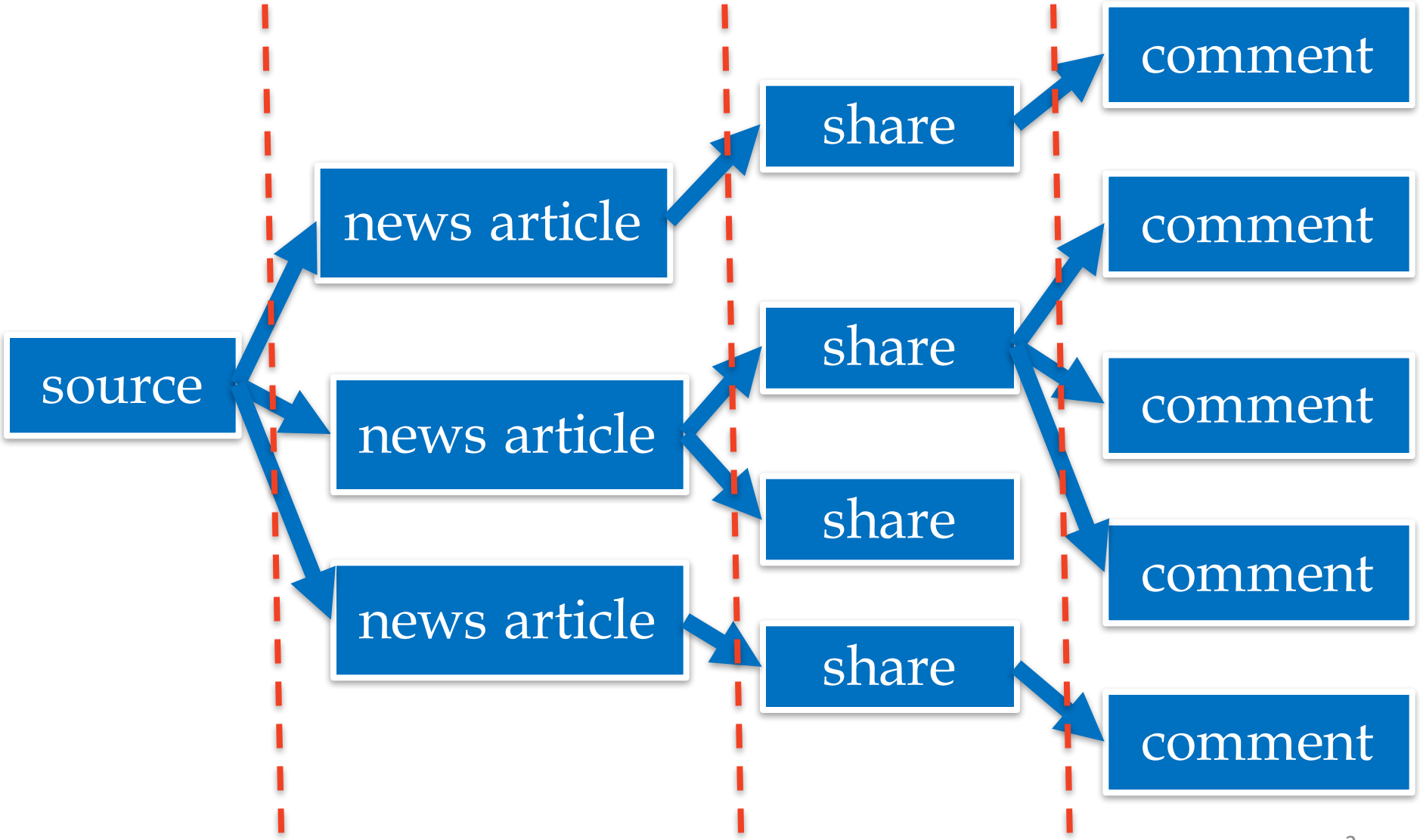


Telephone game

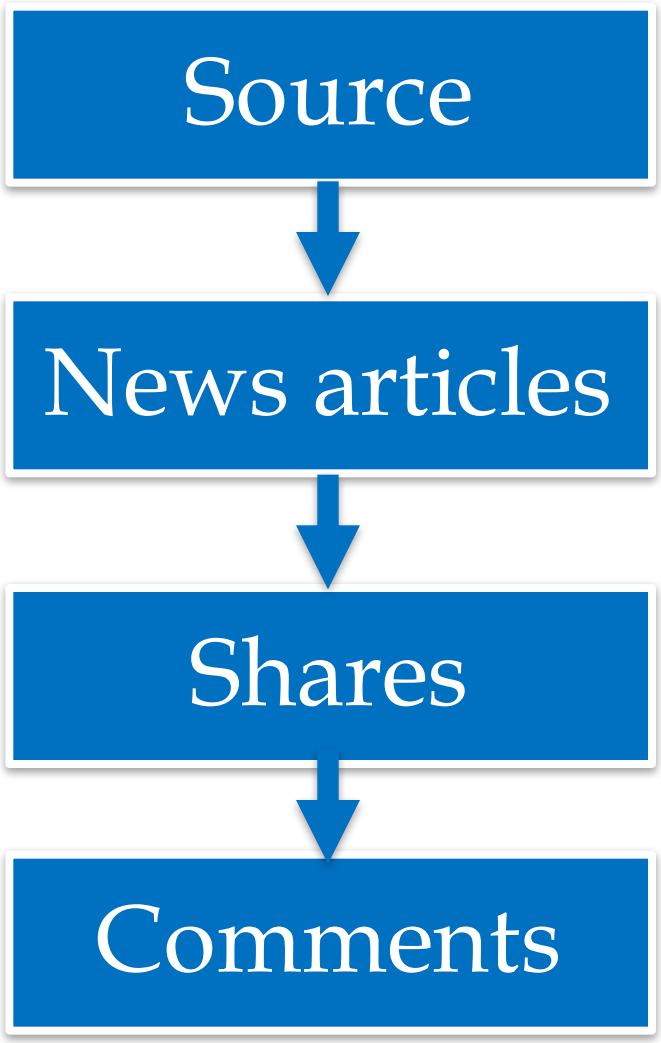




Propagation on Facebook



Propagation on Facebook



- Information sources can be dynamic and complex
- Relatively static sources:

✓ presidential speeches (**politics**)
✓ university press releases (**science**)
✓ press releases from Facebook, Google, etc (**technology**)
✓ statements from the Fed (**finance**)

Obama's speech on deaths of Warren Weinstein and Giovanni Lo Porto

The White House
Office of the Press Secretary

For Immediate Release April 23, 2015

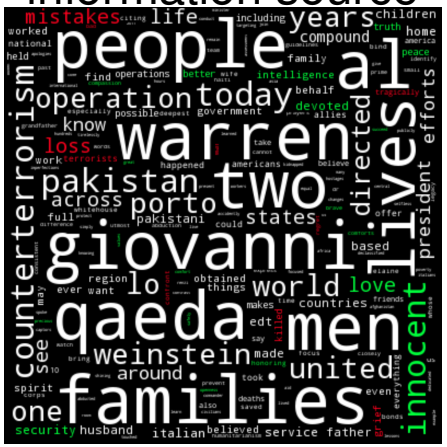
Statement by the President on the Deaths of Warren Weinstein and Giovanni Lo Porto

James S. Brady Press Briefing Room

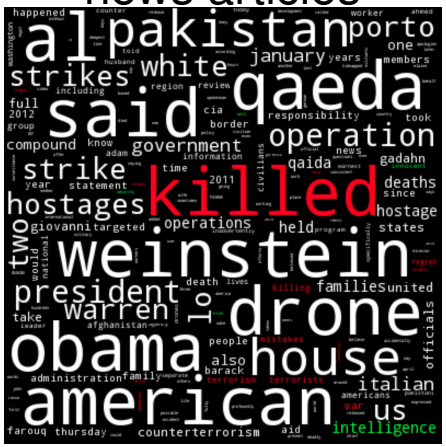
10:05 A.M. EDT

Obama's speech on deaths of Warren Weinstein and Giovanni Lo Porto

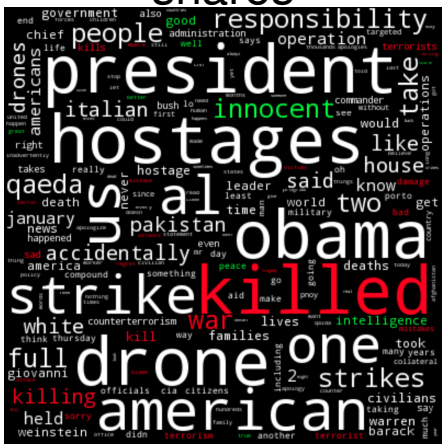
information source



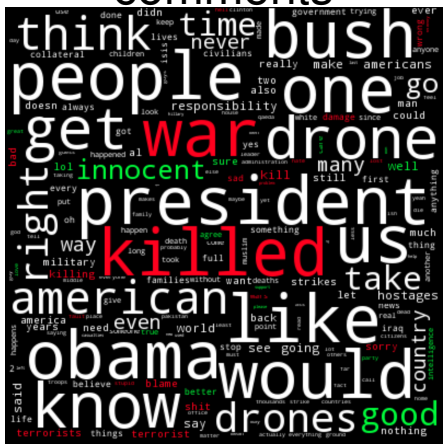
news articles



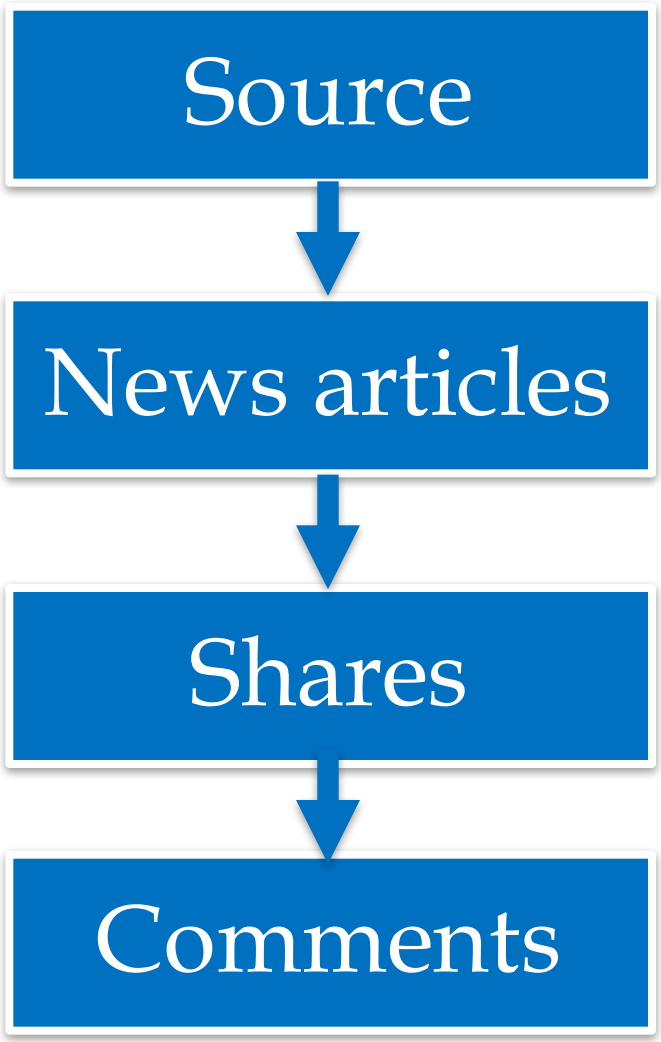
shares



comments



Propagation on Facebook

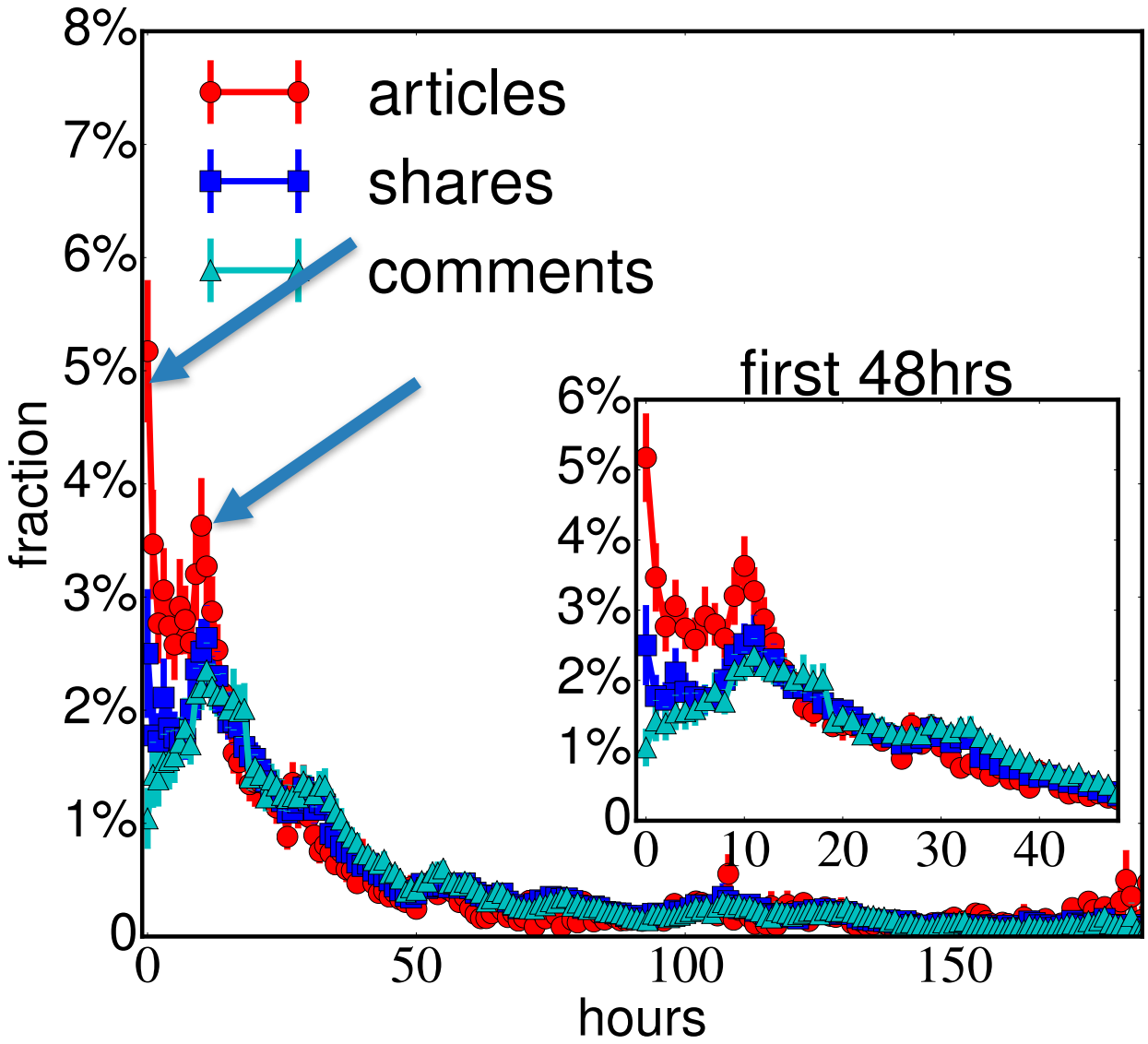


News cycle of a source

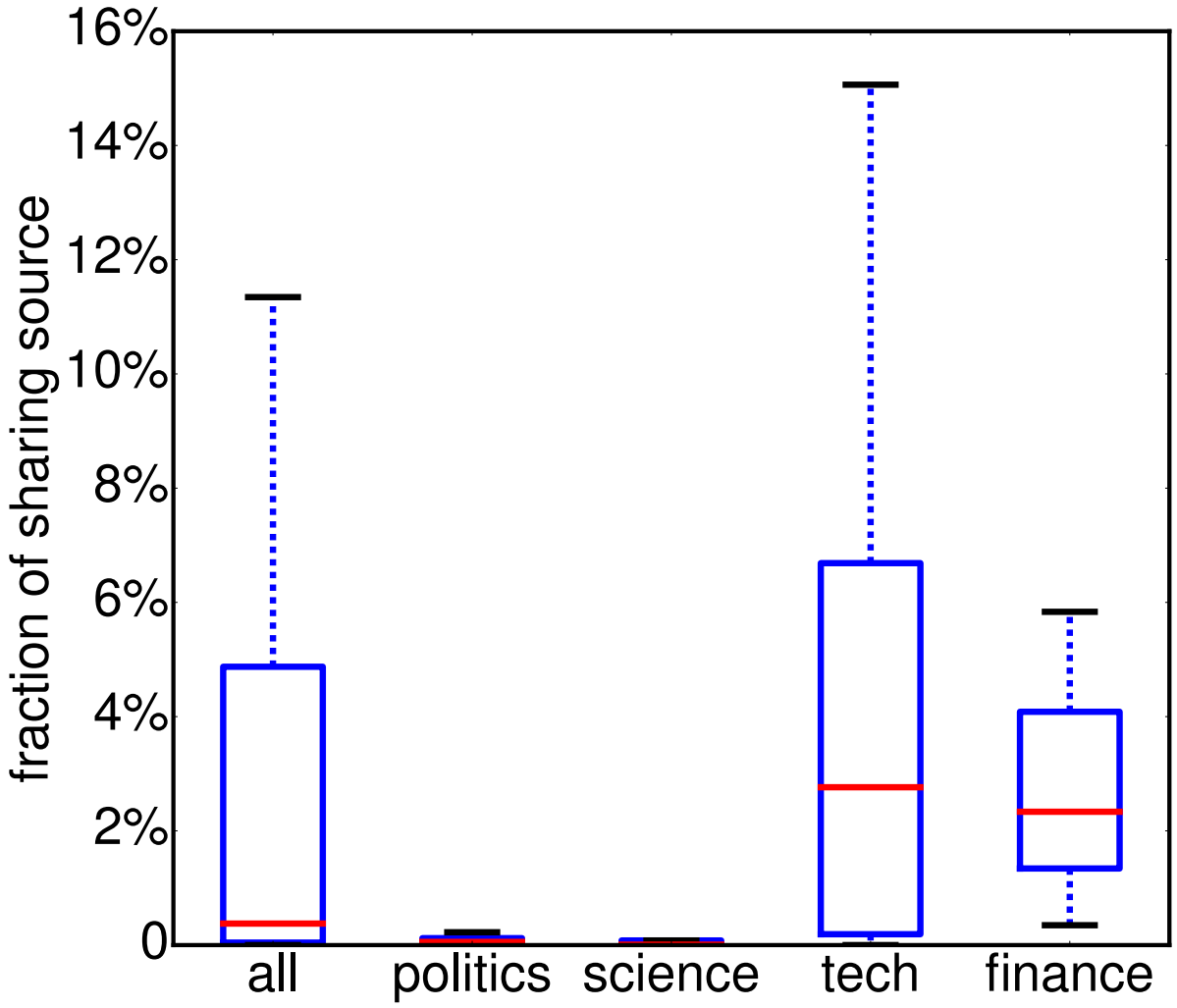
News media coverage

Evolution in propagation

Length of news cycle

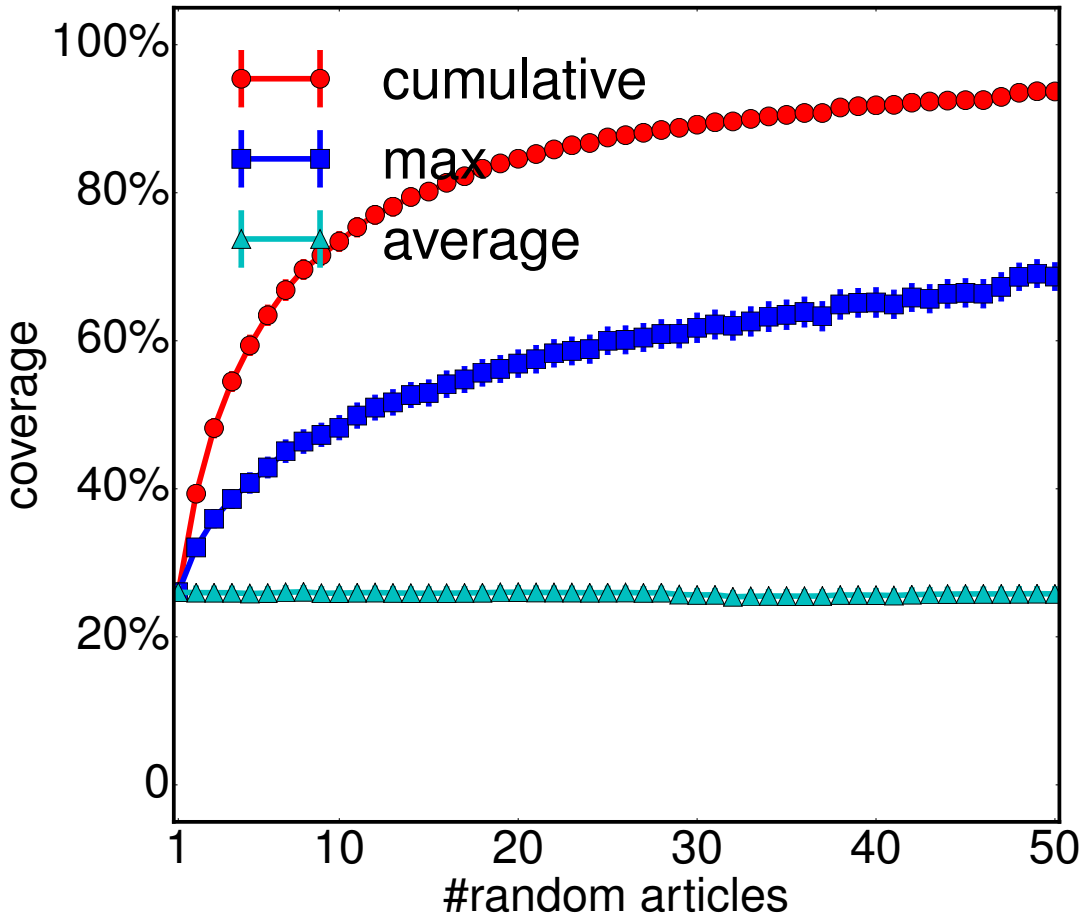


News articles dominate information sources: sources are rarely shared

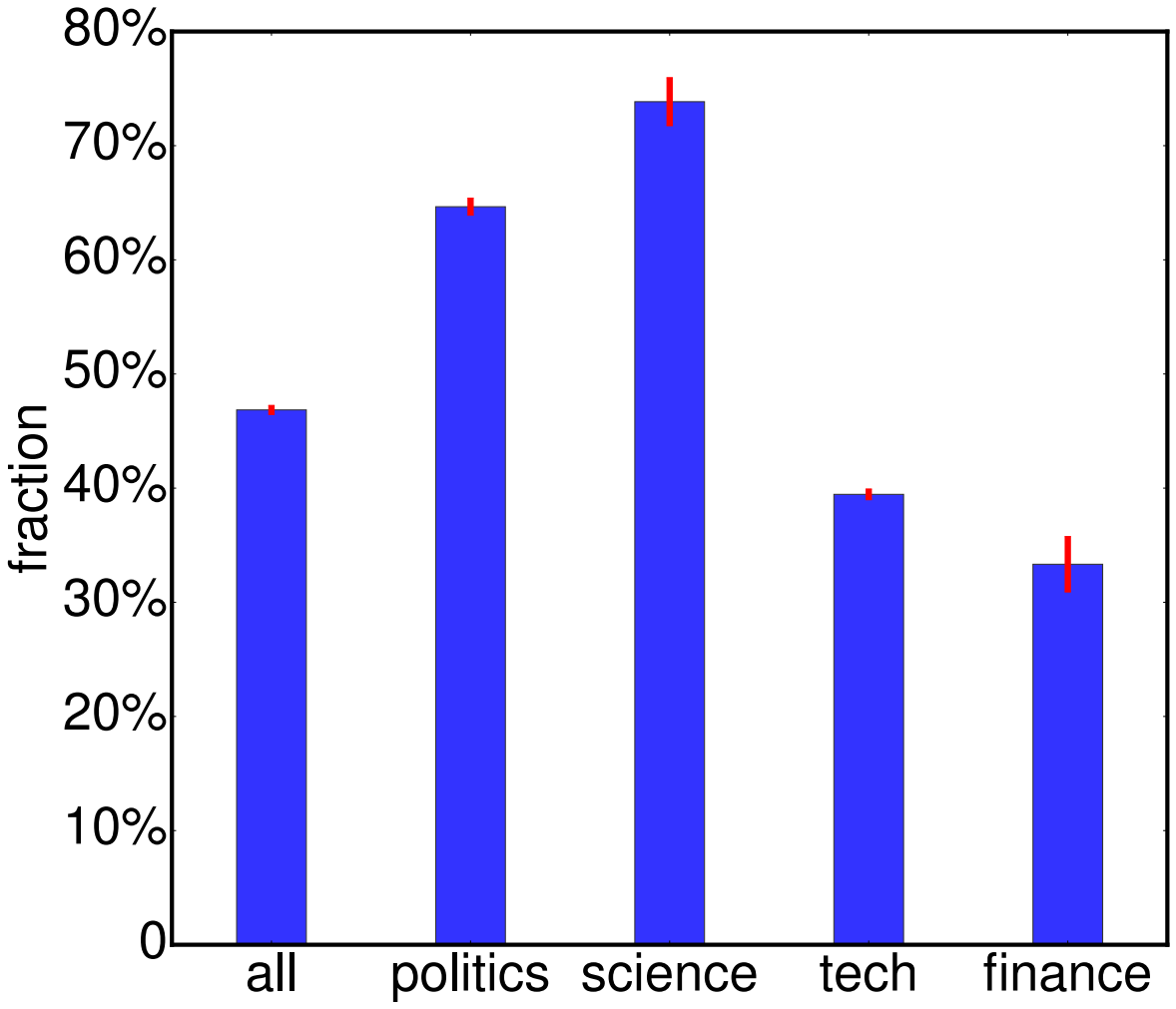


News media coverage of the source

- Stopwords were excluded
- An average article covers 20% of the words in the source
- Several articles combined can cover most of the source

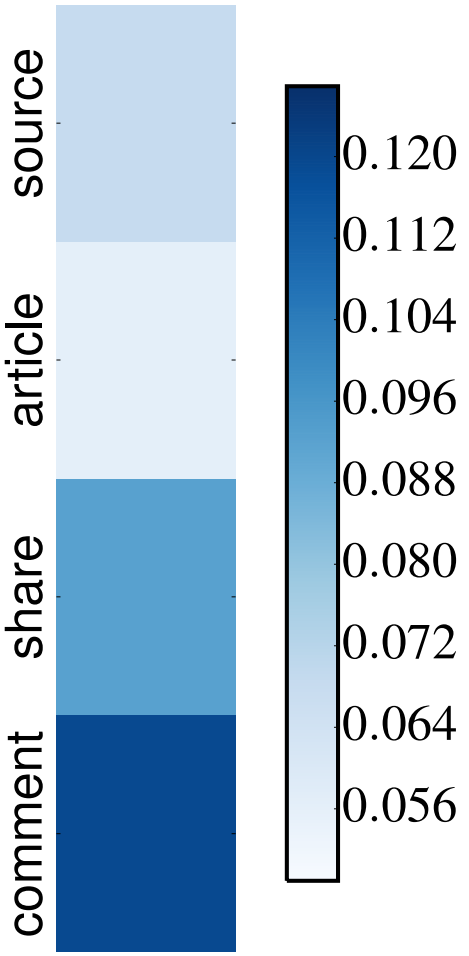


Quoting the source



Sentiment evolution

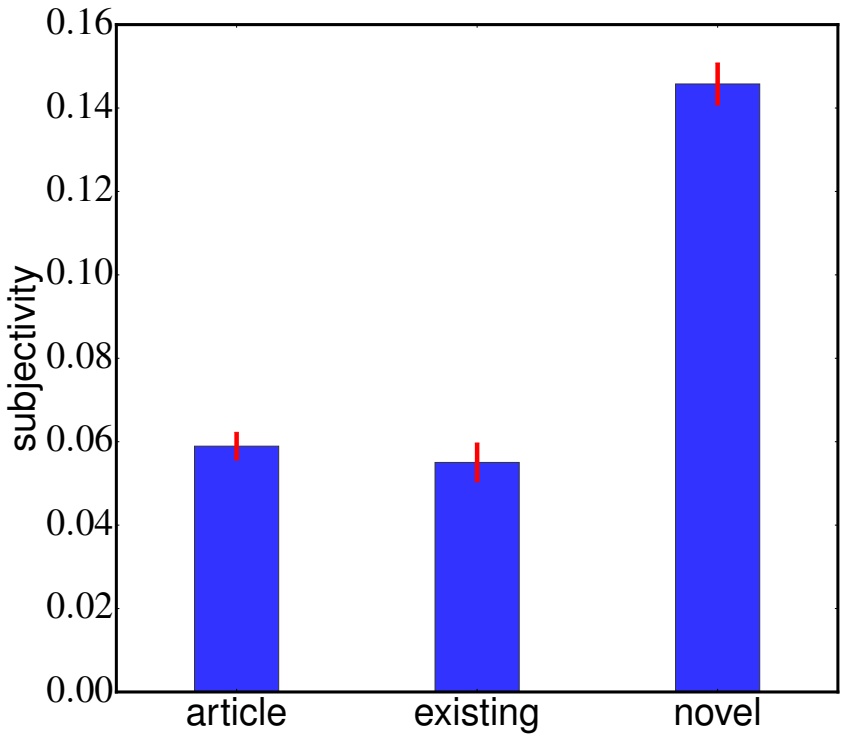
- Subjectivity defined as the fraction of emotional words
- News articles are the least subjective
- Shares and comments get more and more subjective



Blue: more subjective

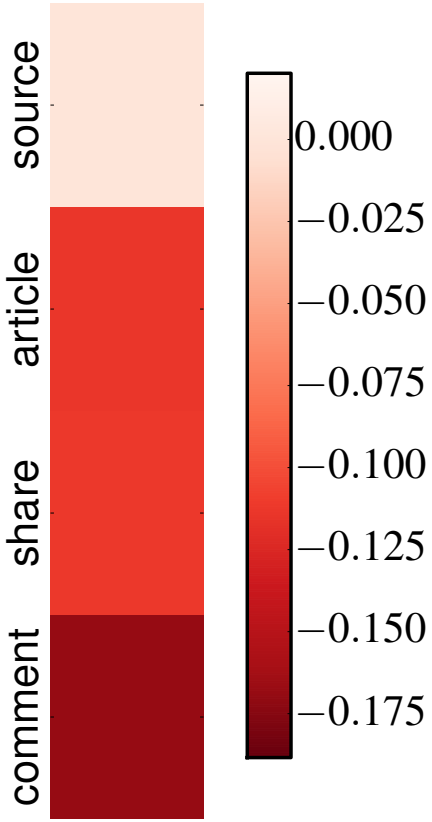
Why does subjectivity increase?

an article → shares → { words already in news articles *(magnifier)*; novel words added by individuals *(creator)* }

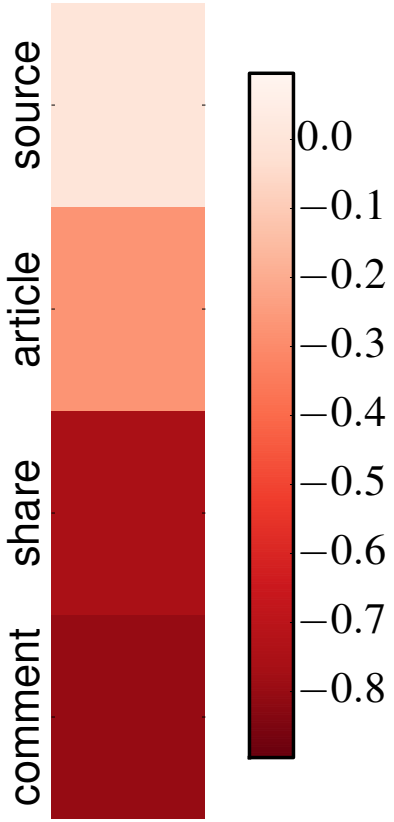


More and more negative in the propagation

LIWC



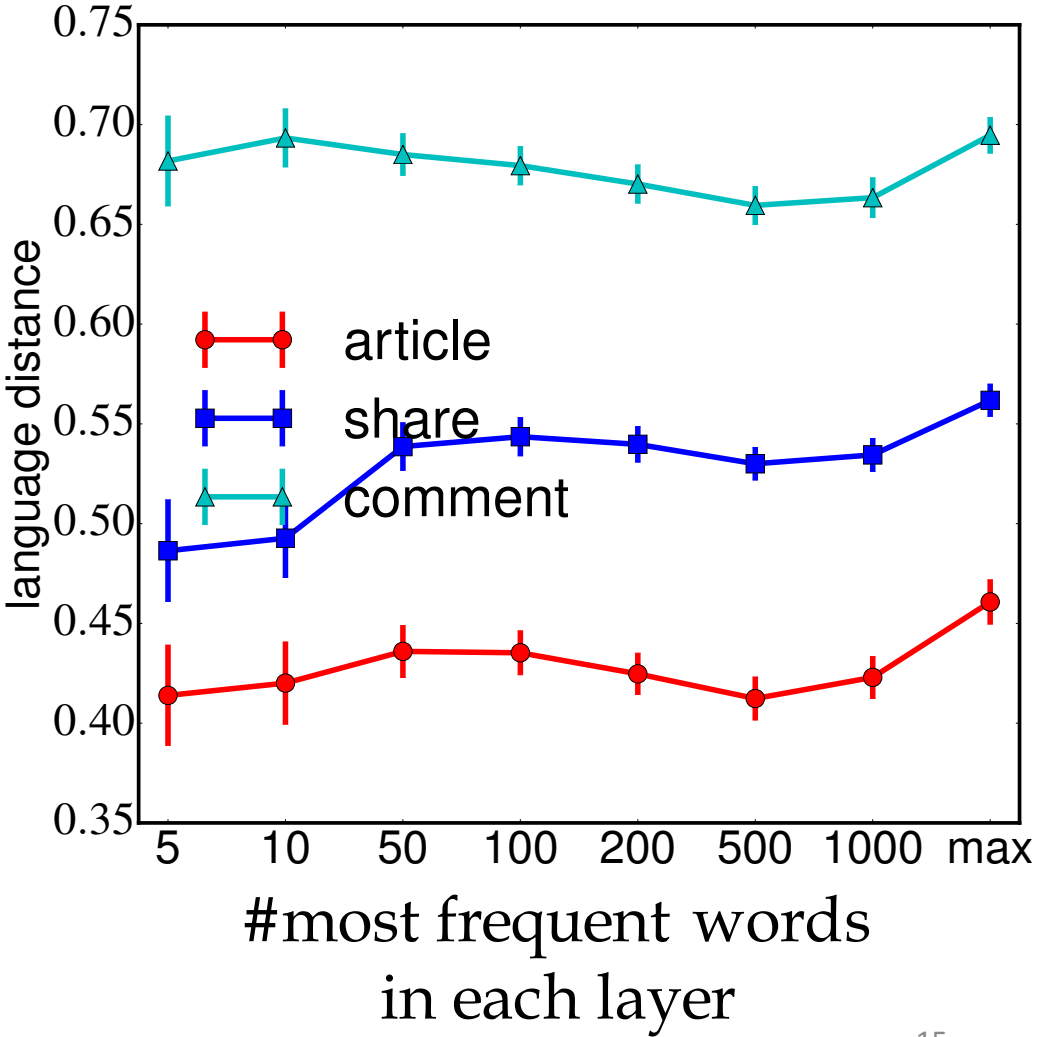
Vader [Hutto and Gilbert 2014]



Red: more negative

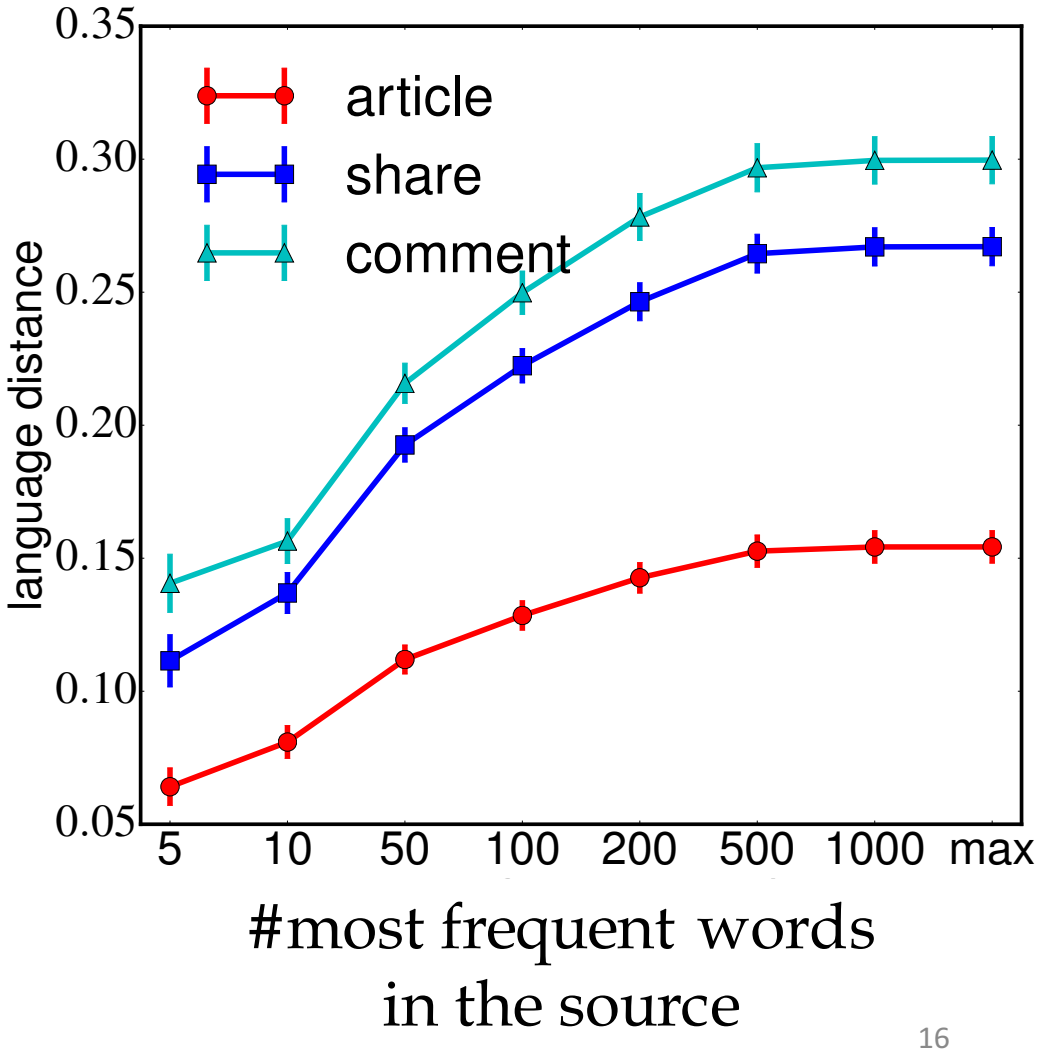
Language model differences

- Jensen-Shannon distance from the source between unigram language models
- Not much difference as the vocabulary size changes



Projected language model differences

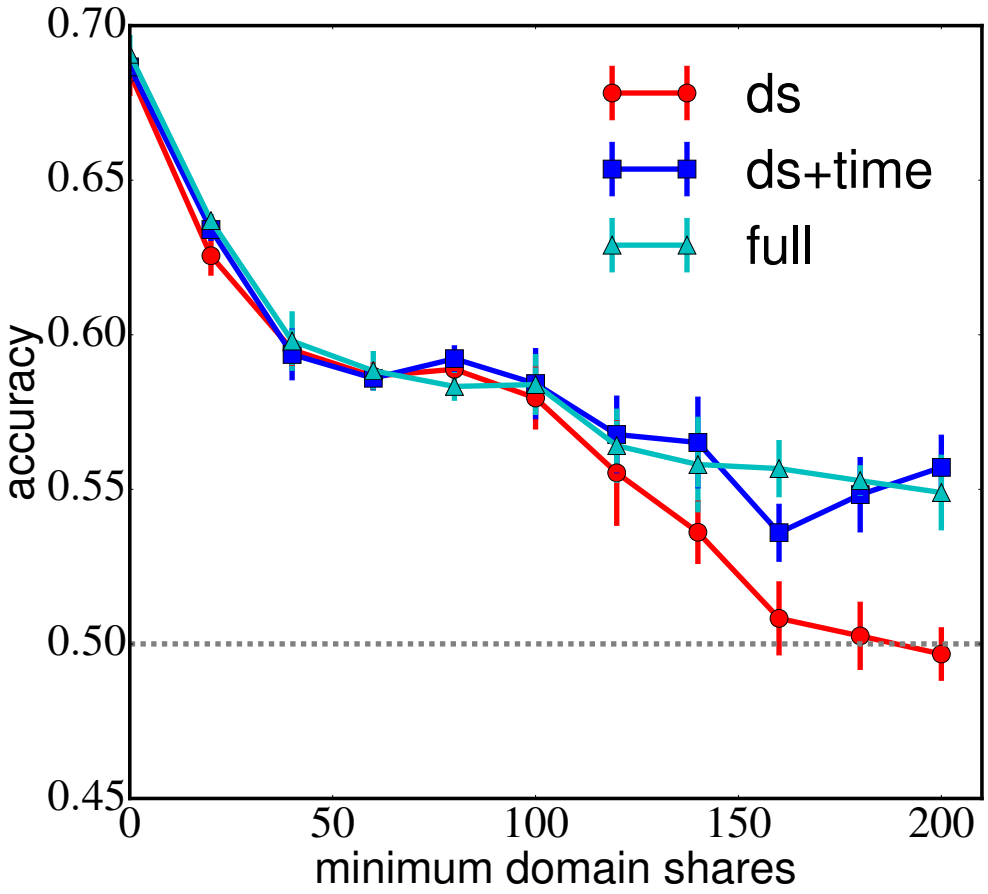
- Jensen-Shannon distance between unigram language models projected on the most frequent word in the source
- Most frequent words in the sources are better preserved



Predicting popular news articles

No difference between these lines:

distances from the source in language and sentiment do not improve prediction!



Take aways

- While information gets more and more different, news articles combined "provide" a comprehensive coverage
- Subjectivity increases as individuals introduce novel content
- More understanding of this process may inform building better tools for the public to understand news