Artificial Intelligence and Network Theory to understand fake news and forecasting elections

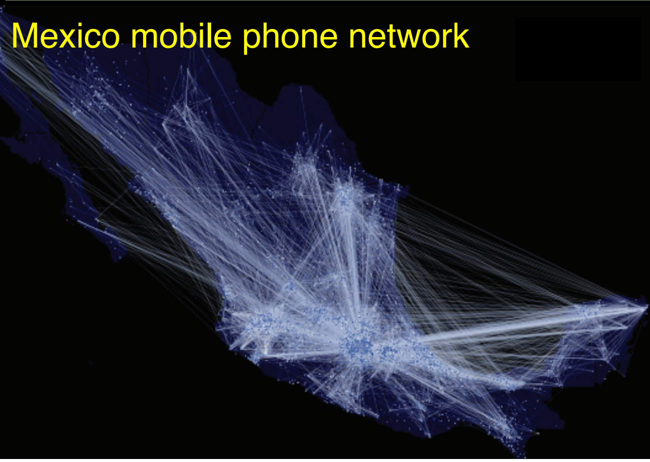
Using Twitter to predict society trends and finding influencers

- Chaoming Song, mini-CUSPEA, Fudan U
- Ping Wang, mini-CUSPEA, Fudan U
- Lin Bo, mini-CUSPEA, Fudan U
- Shaojun Luo, mini-CUSPEA, Peking U
- Kun Wang, mini-CUSPEA, Fudan U
- Yuliang Jin, mini-CUSPEA, Peking U
- Zhuo Yin, mini-CUSPEA, Fudan U
- Sen Pei, Beihang, CSC
- Zhenkun Zhuo, Beihang, CSC
- Flaviano Morone
- Francesca Lucini-Arese
- Alex Bovet
- Hernan Makse

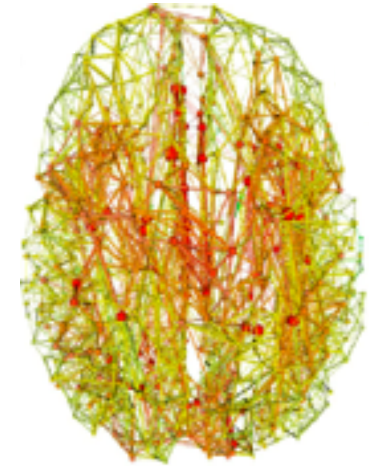
Physics Department
City College of New York

With gratitude to the mini-CUSPEA students from Fudan and Peking University

Information processing networks



Makse, Morone, Nature (2015)
Shaojun Luo, et al. Nature Comm (2017)
Bovet, Makse, Nature Comm (2018)



Society
7 billion
1000 friends

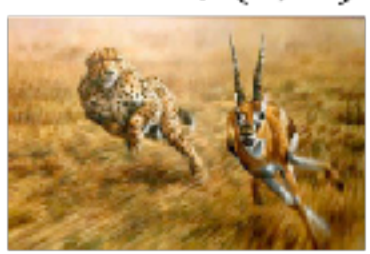
MAKSE LAB
Network theory
Artificial Intelligence
Big-Data Science

The brain
80 billion
10000 axons

Morone, Makse,
Nat. Comm (2019)

Ecosystems and markets

Predator-Prey (- , +)



Morone, Del Ferraro, Makse,
Nature Phys (2019)

The key is to understand the network

THE SOLUTION: REDUCE BIG DATA TO A FEW INFLUENCERS

-Kitsak, Gallos, et al Makse, Nature Phys (2010)
-Morone, Makse Nature (2015)
-Morone, Min, Roth, Makse, PNAS (2017)
-Luo, Morone, Makse, Nature Comm (2017)
-Ferraro, Canals, Makse et al, Nat Comm (2018)
-Morone, Ferraro, Makse, Nature Phys (2018)
-Bovet, Makse, Nature Comm (2018)

INFLUENCERS

Your influence is not yours,

it depends on everyone else

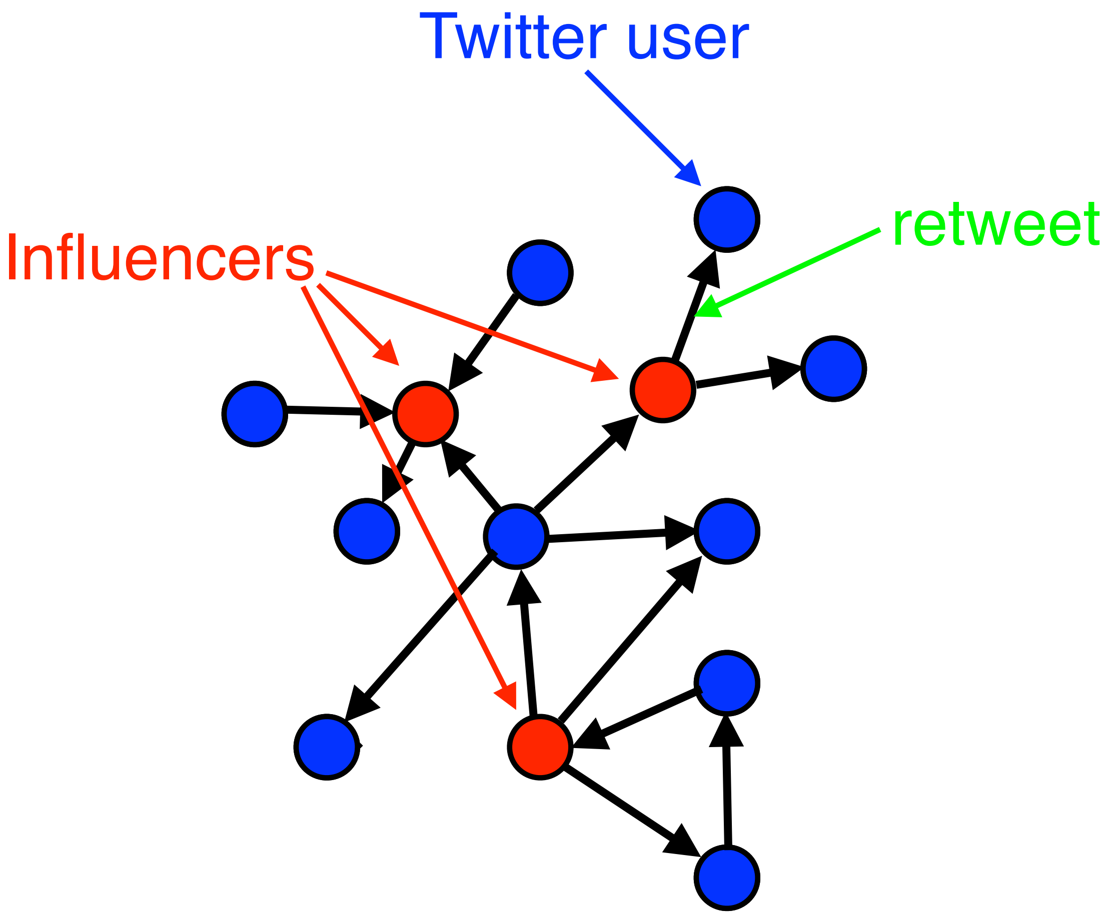
BIG DATA → MACHINE LEARNING → DIMENSIONAL REDUCTION → COLLECTIVE INFLUENCERS

BIG DATA → NETWORK → DIMENSIONAL REDUCTION

Algorithm to identify influencers in a Twitter network

Influencers = if they leave the network, most users become isolated

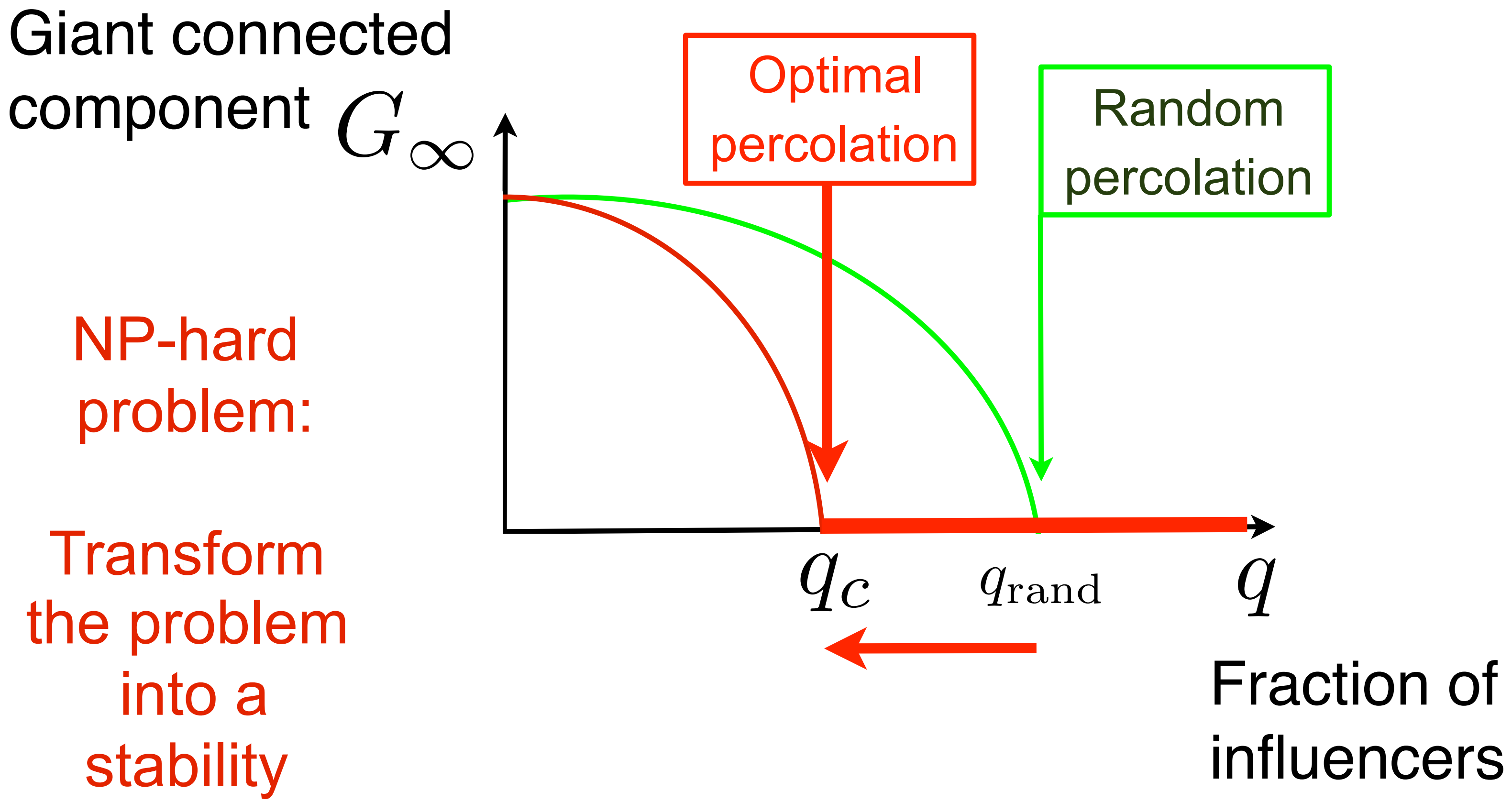
Map the problem to optimal percolation



Optimal Percolation = minimize q_c

Maximization of Influence, Kempe, Tardos, Kleinberg (2003) (NP-hard, computer science, sociology)

Morone, Makse, Nature (2015)



NP-hard problem:

Transform the problem into a stability problem

Finding influencers with the Collective Influence (CI) Algorithm

Network theory identifies a new class of influencer neglected by hubs and Google PageRank

PR: $A_{ij} \rightarrow \lambda_{\max} < k_{\max}$

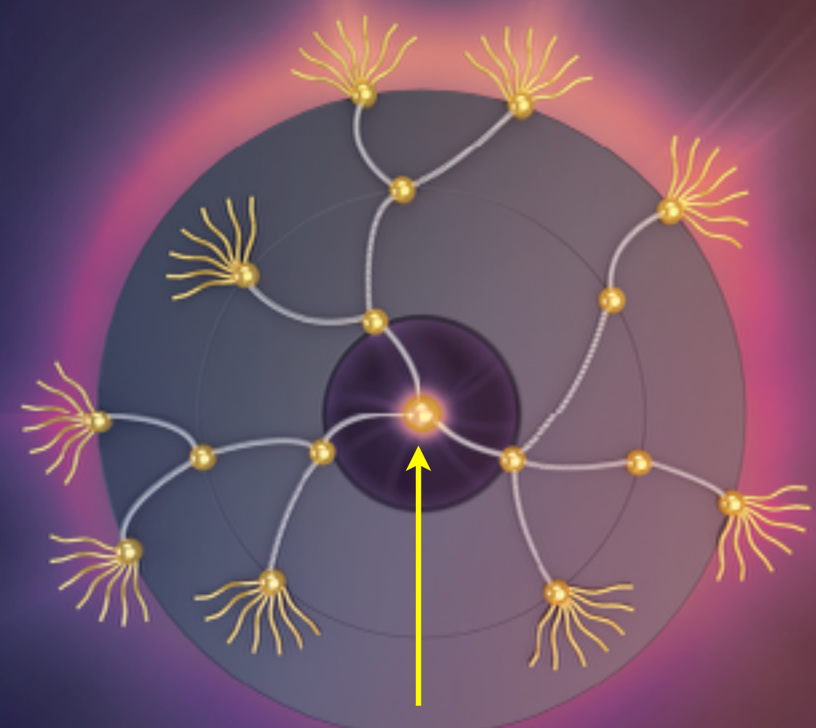
CI: $\mathcal{B}_{ij} \rightarrow \lambda_{\max} = \kappa - 1$

Top influencer: a low degree node surrounded by hierarchical coronas of hubs

Related to Granovetter theory "Strength of weak ties" (1973) in Sociology

$$O(N) \sim N \log N$$

$$\mathrm{CI}_{\ell} = (k_i - 1) \sum_{j \in \partial Ball(\ell)} (k_j - 1)$$



Top influencer = Weak node

Application of Influencer Algorithm

1. Machine learning to Predict elections: Trump vs Clinton
2. Understanding spreading of fake news
3. Tipping point of ecosystem collapse by global warming and stock market crashes
4. Understanding disease from malfunction of influential genes

1. kcore-analytics.com: An influencer platform to track opinion trends in Twitter

KCORE ANALYTICS

LLC

Start-up: track influencers and predict elections with AI

Search Engine for Influencers

Twitter Influencers

Collective Influence Algorithm

Layer 2: machine learning to understand the US 2016 Presidential election.
Bovet, Morone, Makse
Sci Rep. (2018)

Layer 3: Granger causality to uncover the influence of Fake News in the US 2016 Presidential election.
Bovet, Makse, Nat Comm (2018)

Fake News: From the New York Times

The New York Times | U.S. Share

The Plot to Subvert an Election

Unraveling the Russia Story So Far

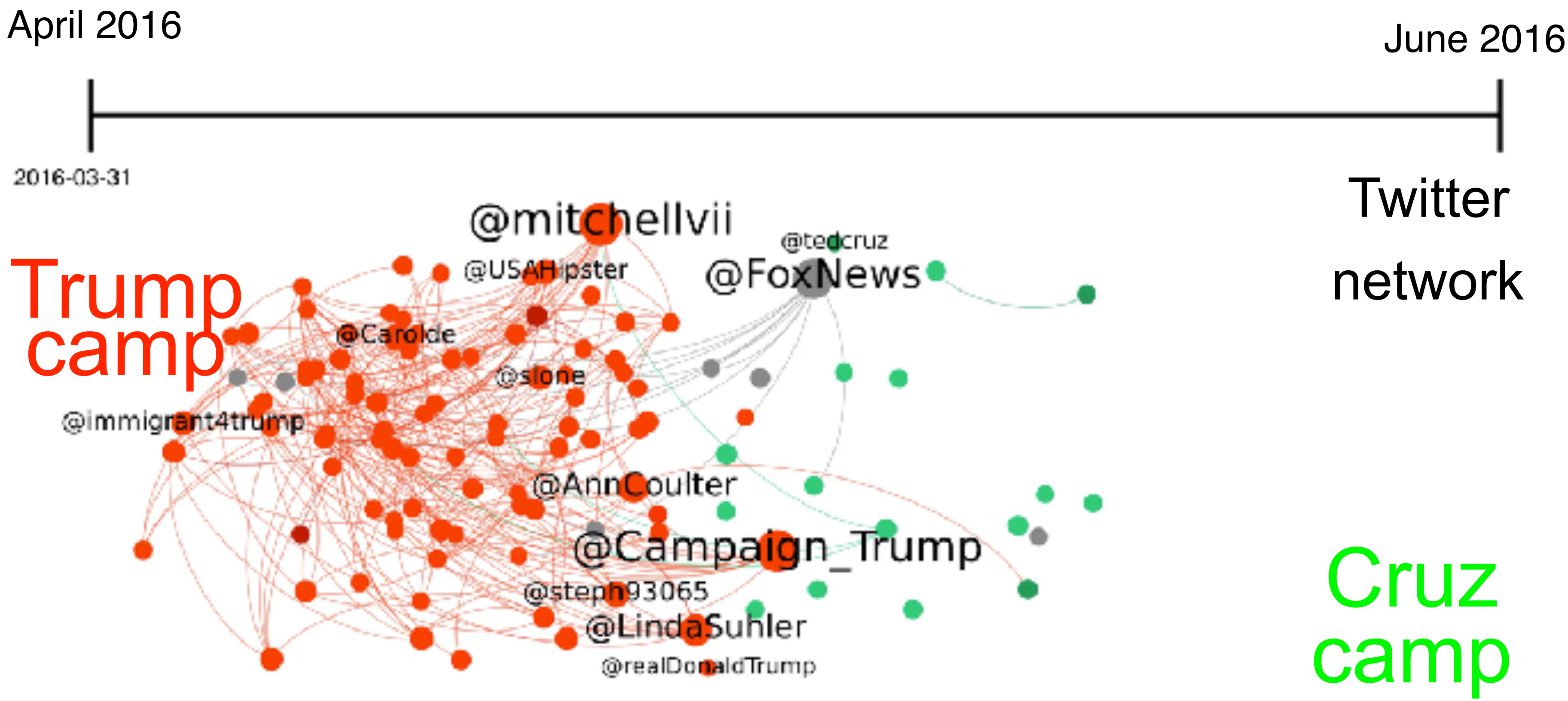
For two years, Americans have tried to absorb the details of the 2016 attack — hacked emails, social media fraud, suspected spies — and President Trump's claims that it's all a hoax. The Times explores what we know and what it means.

By SCOTT SHANE and MARK MAZZETTI

SEPT. 20, 2018

NYT: you should have read our papers….

Twitter influencers in the Trump campaign



2016 US Primary election

Twitter influencers

Who is Linda Suhler, PhD?

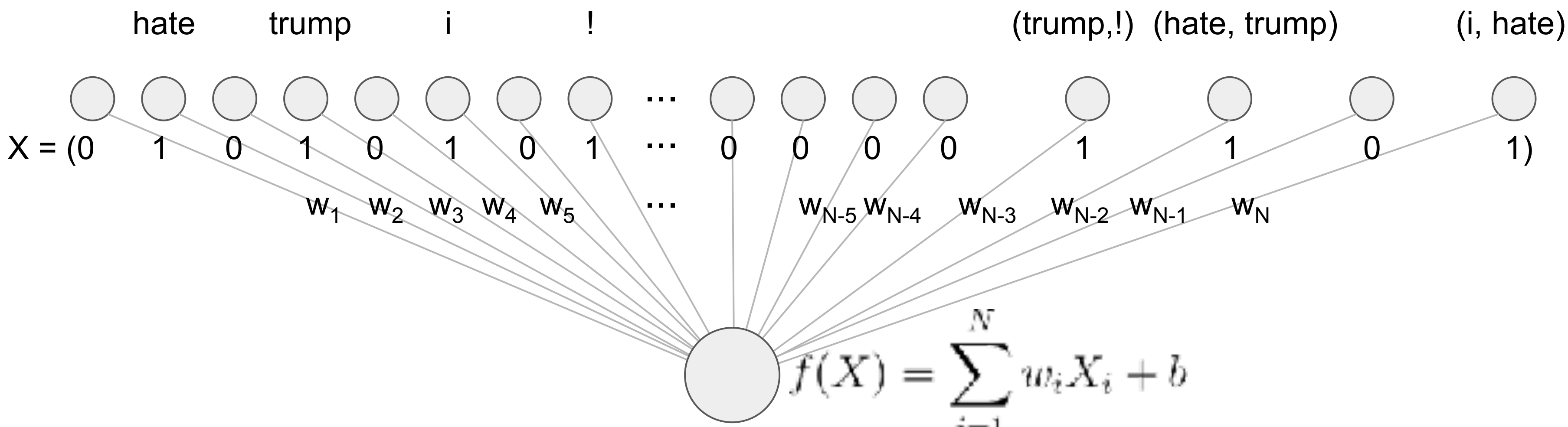
1 tweet every 10 minutes for the last 7 years

2. Simple classifier for Opinion

How to classify a tweet:

I love Trump!

I love Trump! #MAGA



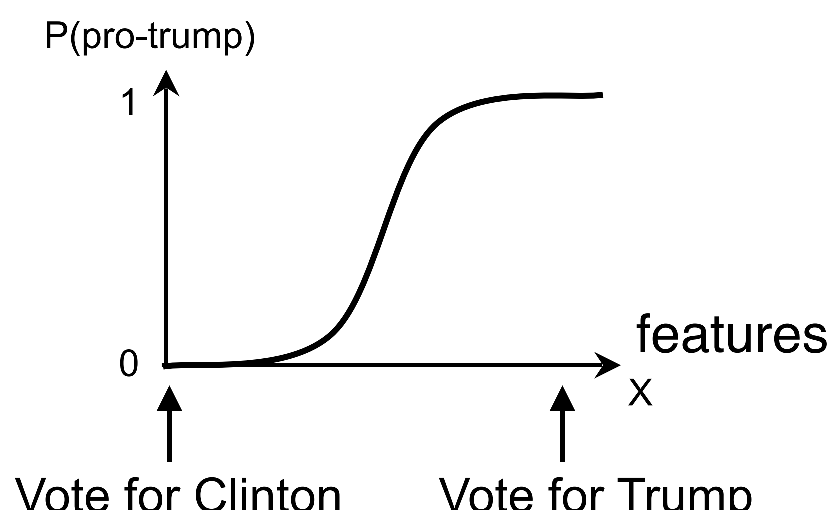
$$f(X) = \sum_{i=1}^{N} w_i X_i + b$$

IMPORTANT NUMBERS:

- 3.5M features = 3.5M parameters to fit
Whole English Language = 1M

- 1M training examples from easily identified hashtags

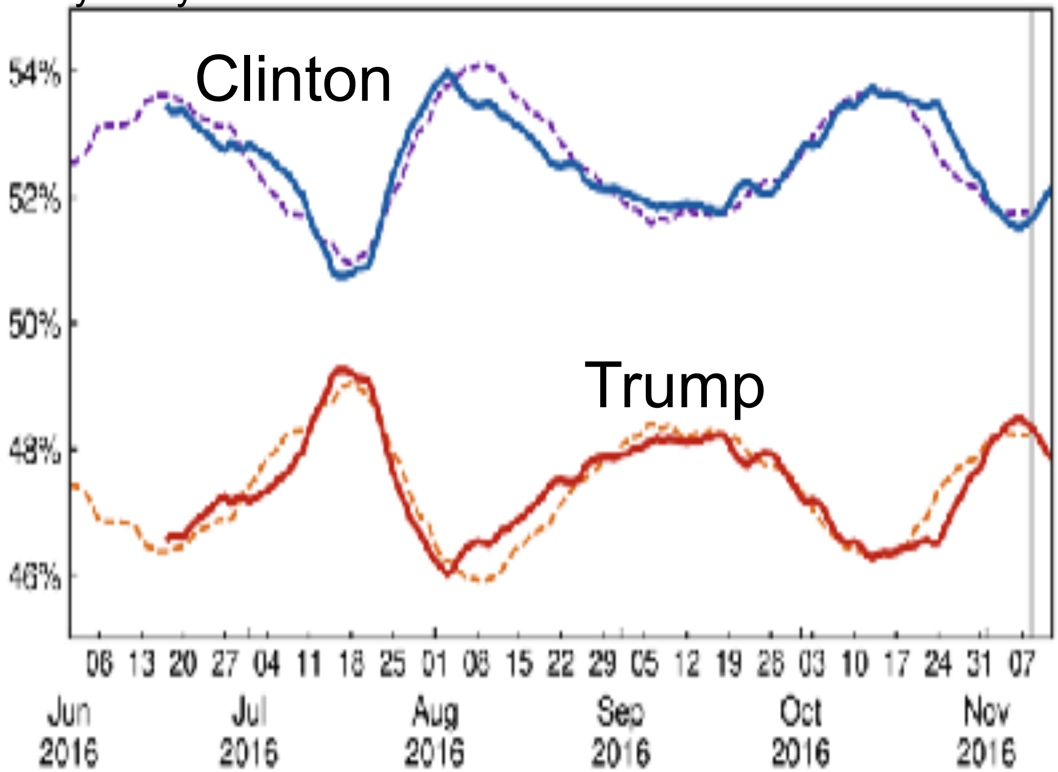
- We infer: 170M total number of tweets



Machine Learning replaces the election polls from the New York Times

Classify every tweet

Bovet, Morone, Makse, Sci Rep (2018)



51.2%

48.8%

The Upshot at New York Times Polls = $19billion

Our Twitter Machine Learning = FREE

(almost free = $30k grad student salary)

Training set of 1 million tweets

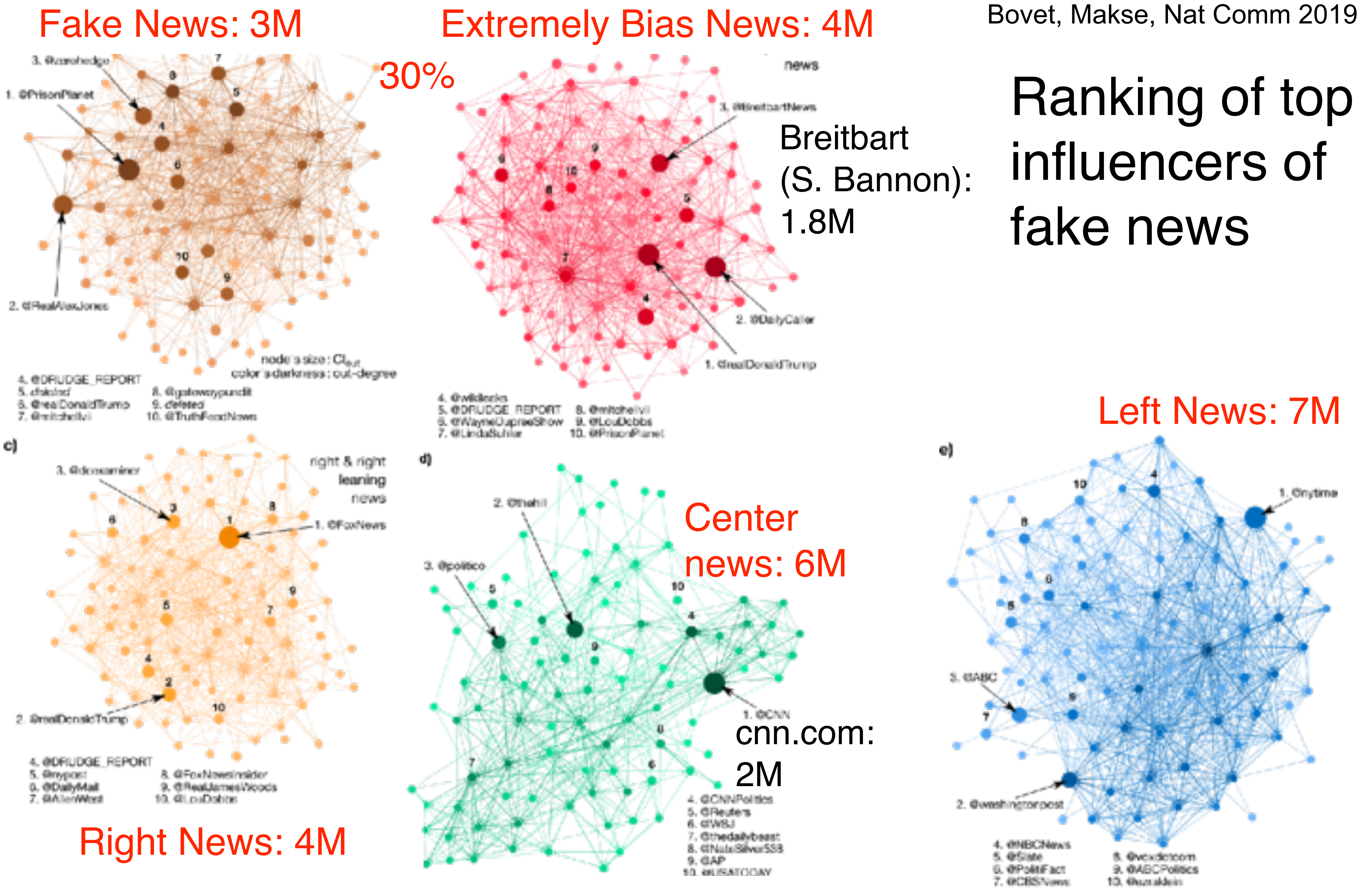
Predicting the opinion of 170 million tweets

of 11 million people

A complete characterization of Twitter Fake News and their Influencers in 2016 US Presidential Election

Bovet, Makse, Nat Comm 2019

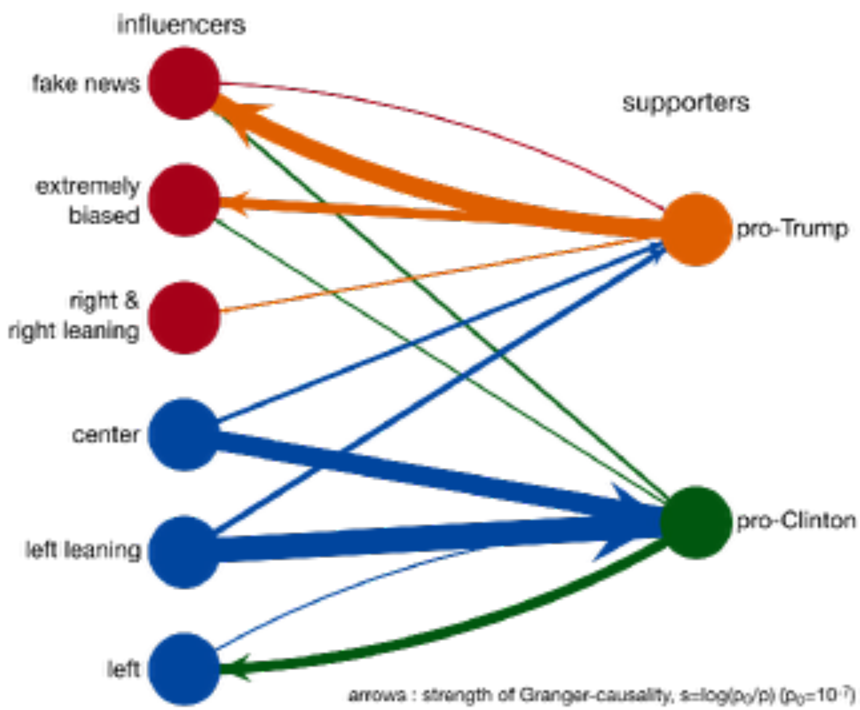
Ranking of top influencers of fake news



3. Who influences who?
Did fake news influence voters in the US presidential election?

Activity correlations of influencers

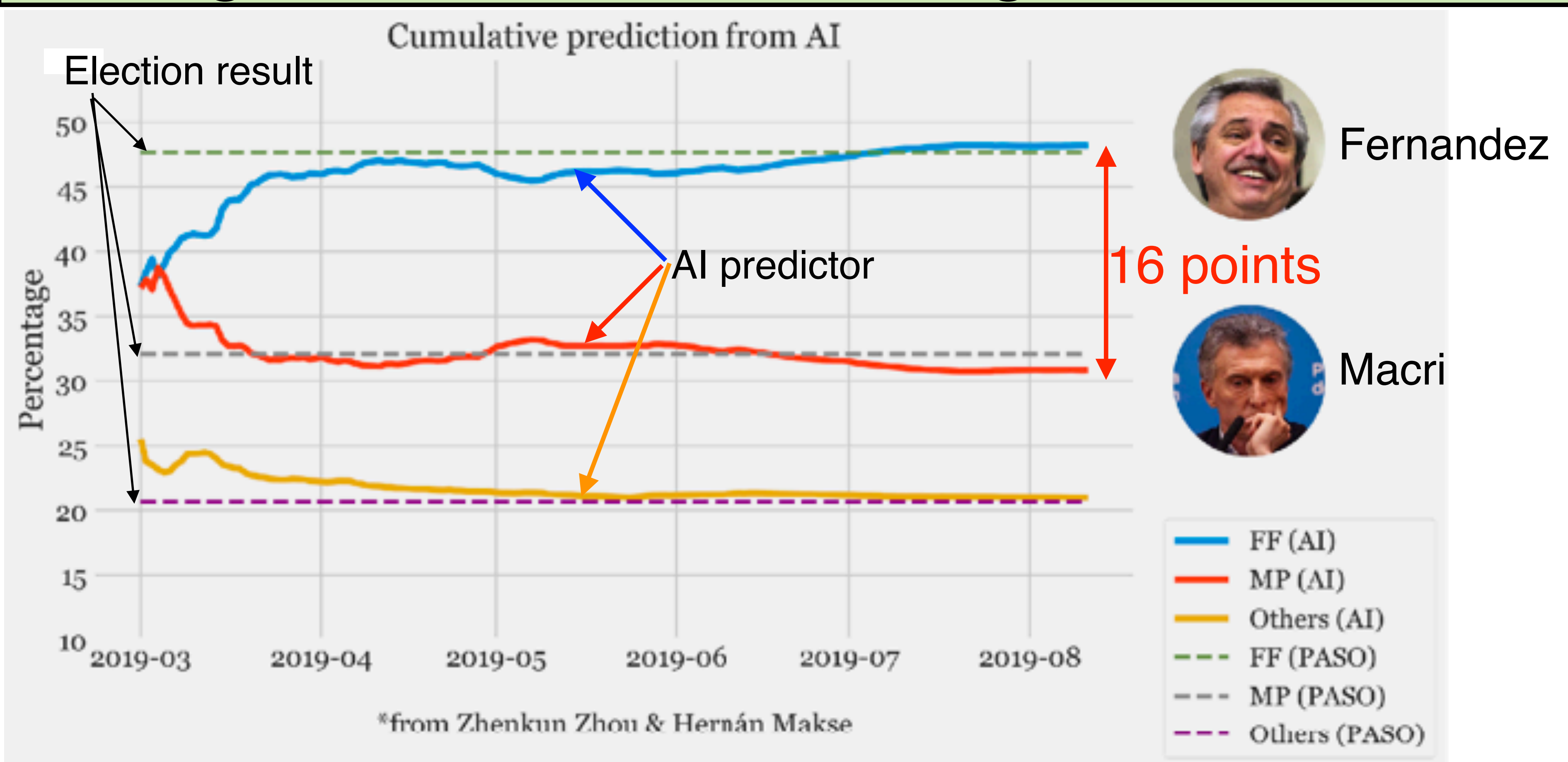
Granger causality: surprising



Trump supporters granger-cause influencers of fake and extremely bias news

Clinton influencers and liberal media granger-cause Clinton supporters

We recently predicted the unpredictable: Argentina election on August 11, 2019



All traditional pollsters (Elypsis) predicted a win for right-candidate Macri.
Left-candidate Fernandez won by 16%.
The bond market crashed by 40%.

Why the polls fail?

Summary

1. Dimensional reduction: Machine learning and Network Science

2. Applications to elections, fake news, brain and genetic networks and ecosystems

3. Brain and genetic networks need a different network paradigm for information processing