Crowd-sensing with Polarized Sources

Tanvir Amin, Tarek Abdelzaher, Dong Wang
University of Illinois at Urbana Champaign
{maamin2, zaher, dwang24}@illinois.edu

Bolek Szymanski
Rensselaer Polytechnic Institute
szymab@rpi.edu

Presented by
Tanvir Amin

IEEE DCOSS 2014

Humans as Sensors

Egypt Unrest

Fukushima Disaster

Sandy Gas Outage

Crimea Annexation

Syria Chemical Attack

Twitter | Facebook | Google+ | Instagram | Flickr

Credibility Estimation | Anomaly Detection | Timeline Reconstruction | …

Binary Sensor Model

- Assume that each observation is either True or False
  - True means independently observable events.

> Flights halted at airport in east Ukrainian city of Donetsk after armed separatists demand withdrawal of troops bbc.in/1nKCMIB
>
> Reply Retweet Favorite ••• More

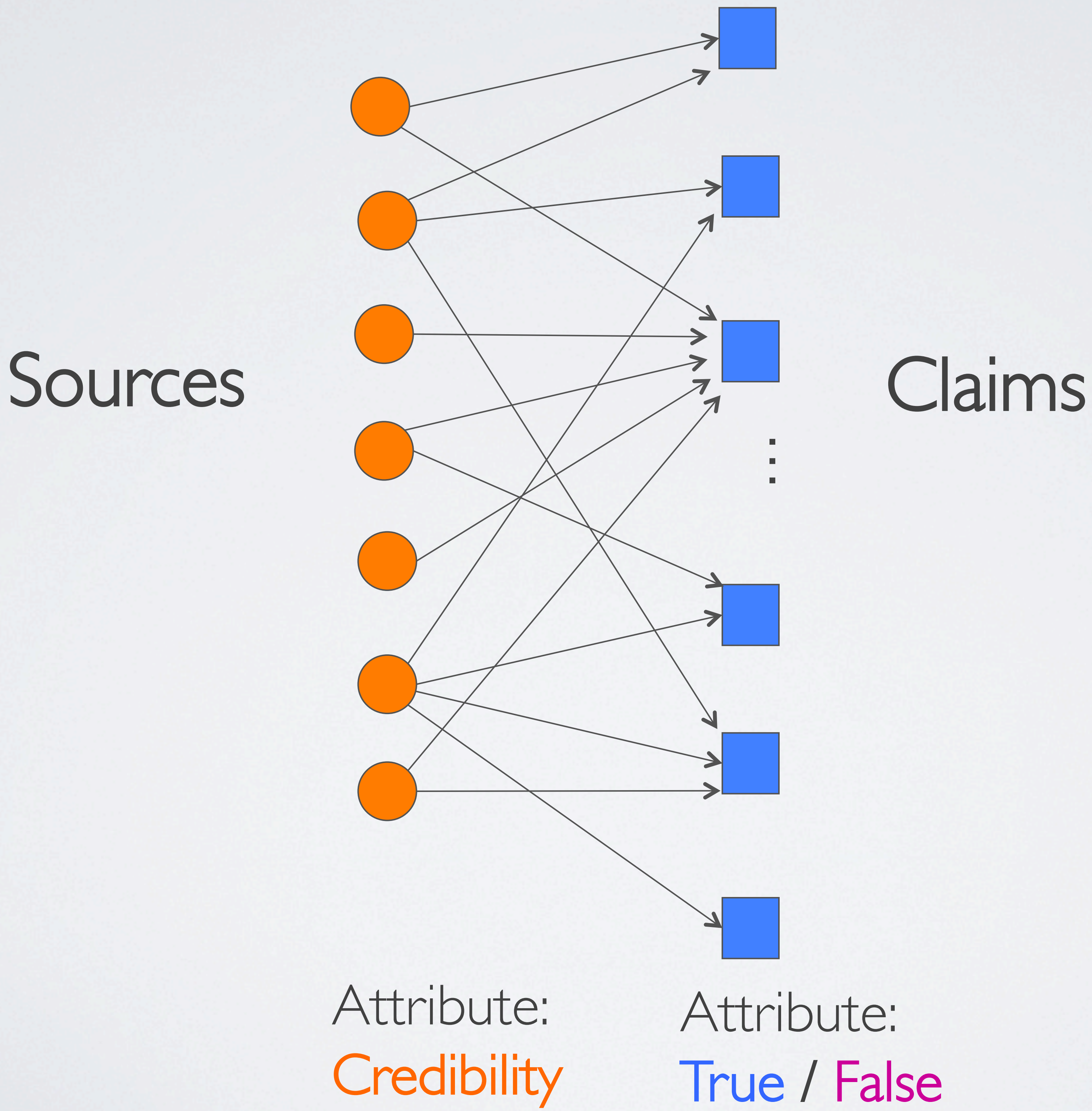
> After Hurricane Irene hit Puerto Rico, the streets were so flooded that a shark managed to swim in a street. yfrog.com/kly3ljjj
>
> Reply Retweet Favorite ••• More
>
> 12:04 AM - 25 Aug 2011

- Some observations are neither True nor False, representing Non-Factual claims.
  - May be slogans or emotions or opinions.

> America doesn't learn. Another mass shooting. This time in California. A person should never be able to buy that many bullets #massShooting
>
> Reply Retweet Favorite ••• More

> Anyone else think that Steve jobs missed a trick by not calling the battery charge on an iPhone "apple juice" ???
>
> Reply Retweet Favorite ••• More

Source Claim Network



State of the Art

- Independent Sources, Independent Claims IPSN 2012
  - On Truth Discovery in Social Sensing
- Confidence Bounds SECON 2012
  - On scalability and robustness limitations of real and asymptotic confidence bounds in social sensing
- Admission Control INSS 2012
  - On Diversifying Source Selection in Social Sensing
- Conflicting Claims RTSS 2013
  - Exploitation of Physical Constraints for Reliable Social Sensing
- Non-independent Sources IPSN 2014
  - Using Humans as Sensors: An Estimation Theoretic Perspective
- Polarized Sources DCOSS 2014
  - This paper

Case Study: Egypt 2013

- Event: 2013 Uprising regarding former Egyptian President Mohamed Morsi
- Crawler starting July 2013, and continued for more than four months.
- 17 GB of tweets collected. 600K were "English" containing the word "Morsi", which belonged to 173K cascades of different claims / observations.
- The largest 1000 cascades were manually annotated as being Pro-Morsi or Anti-Morsi or Neither
  - Accounted for 44K sources and 95K tweets

Pro Claims

Mahmoud M. Abuelnass
@MAbuelnass

Follow

#EGYPT
#Morsi supporters denied right amid reports of arrests and beating
#Military_Coup
facebook.com/photo.php?fbid…

Reply Retweet Favorite More

RETWEETS
2

2:44 AM - 19 Jul 2013

BBC Breaking News
@bbcnews_ticker

Follow

Egyptian police fire tear gas to disperse supporters of ousted President Mohammed Morsi, reports say bbc.in/17mafk3

Reply Retweet Favorite More

RETWEET
1

8:10 AM - 13 Aug 2013

Anti Claims

ArabidolNews
@Arabidolnews

#RT " @DrBassemYoussef Helicopter footage of Anti Morsi protests, very impressive youtube.com/watch?v=Vux_-v… … "

Reply Retweet Favorite ••• More

RETWEETS 2 FAVORITES 5

5:48 AM - 2 Jul 2013

Khaled Abol Naga
@kalnaga

Amnesty International | Egypt: Evidence points to torture carried out by Morsi supporters amnesty.org/en/for-media/p…

Reply Retweet Favorite ••• More

RETWEETS 60 FAVORITES 21

2:09 PM - 18 Aug 2013

Neither

BBC Breaking News @BBCBreaking

Follow

PHOTO: Supporters of ousted President Morsi (left) and supporters of the army (right) in Cairo bbc.in/16jb0K5 pic.twitter.com/j9GnkBfoJH

Reply Retweet Favorite More

RETWEETS 370 FAVORITES 73

11:56 AM - 26 Jul 2013

Flag media

TavernKeepers @TavernKeepers

Follow

Our two part series on #Benghazi. Was it a botched kidnapping? Were #Morsi #Obama involved? The Blind Sheik? tavernkeepers.com/benghazi-murde… #LNYHBT

Reply Retweet Favorite More

RETWEET 1 FAVORITE 1

8:01 PM - 5 Nov 2013

semprecontro @semprecontro

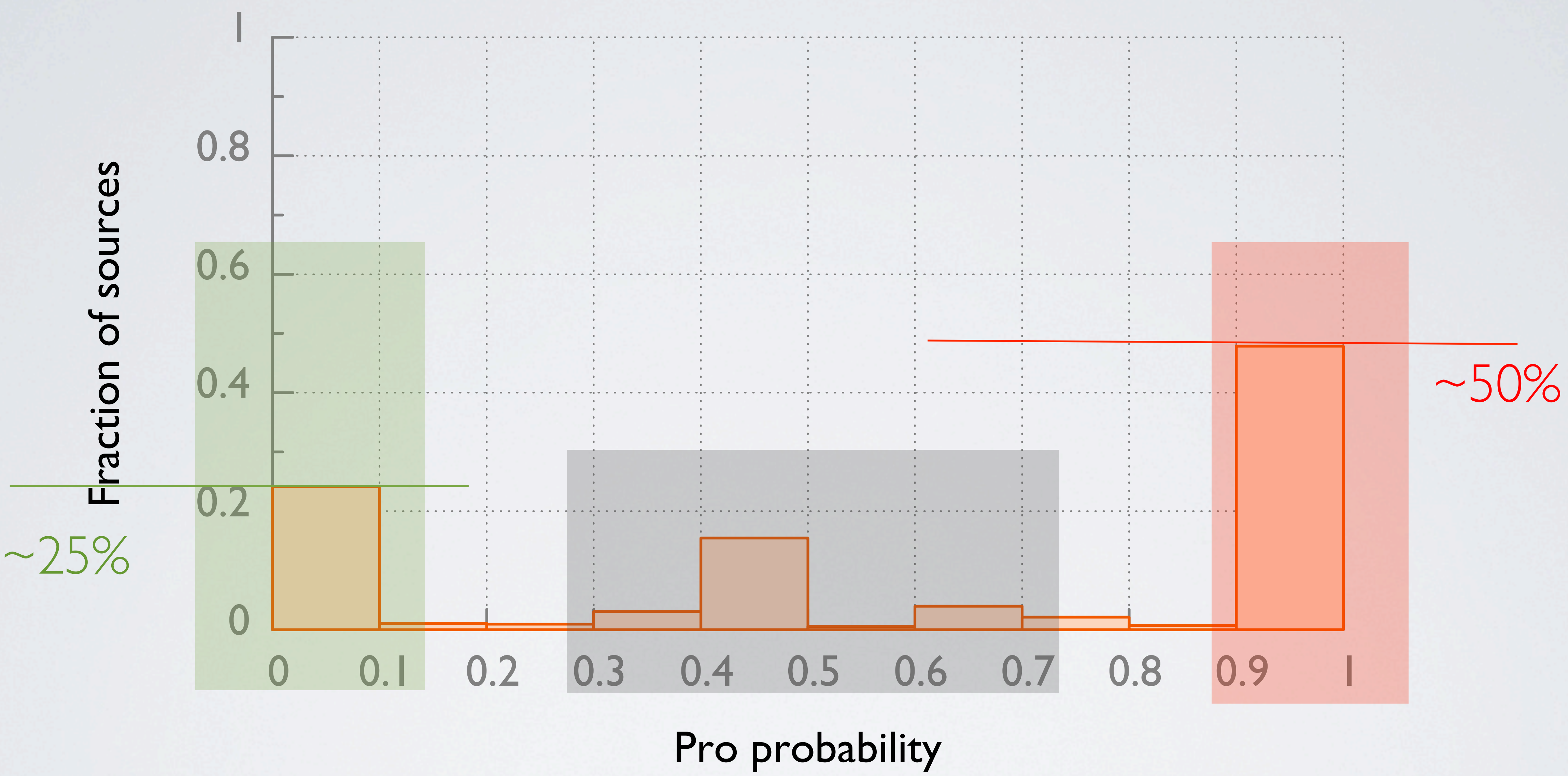
Follow

#Syrian refugees living in Egypt get swept up by turmoil following Morsi's ouster | Fox News" (fxn.ws/18fYHQS)

Reply Retweet Favorite More

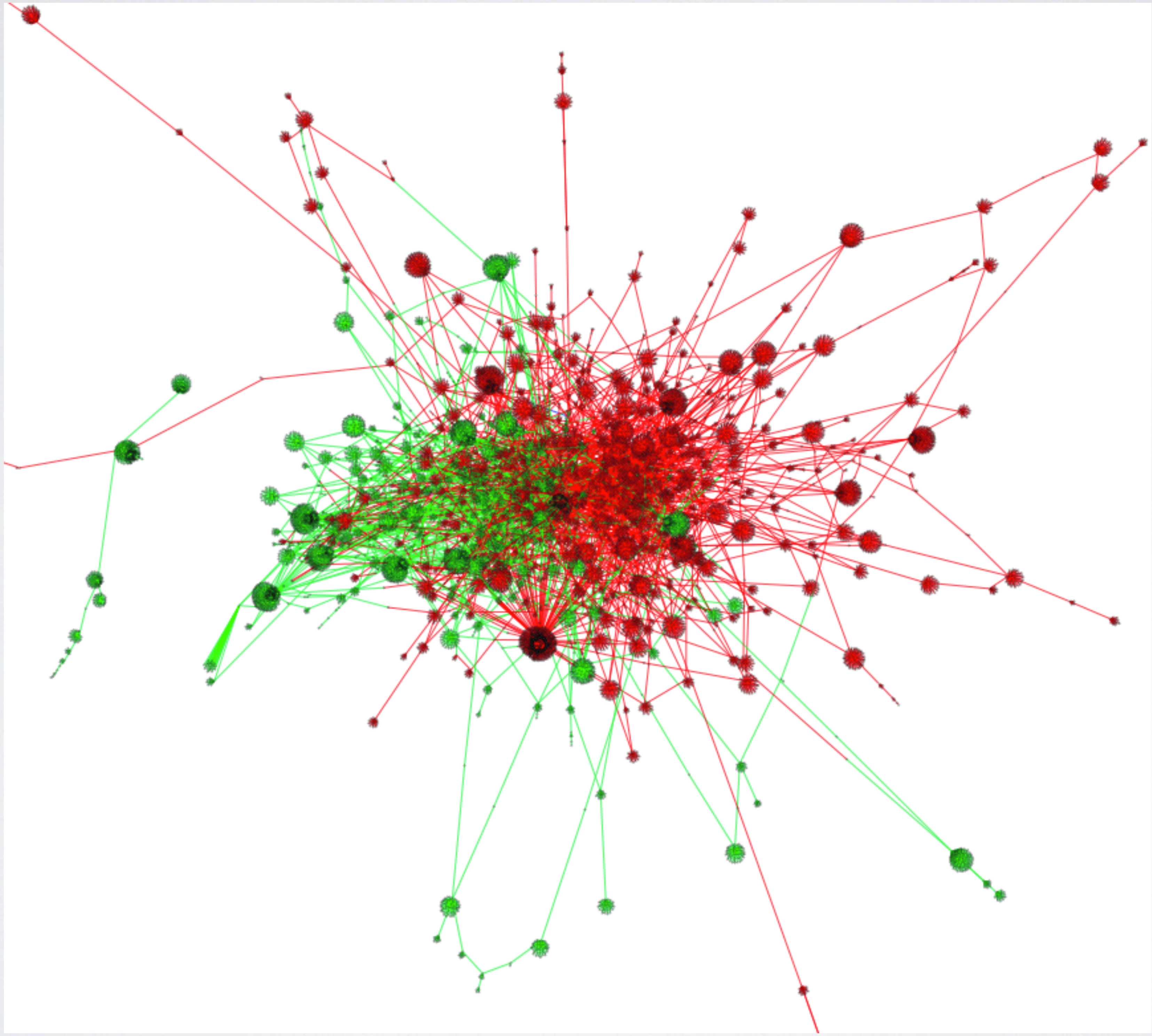
3:02 AM - 15 Jul 2013

Degree of Polarization



- Some users are highly polarized, and mostly forwards tweets favoring camp they belong to (Pro or Anti)
- Some users are neutral

Social Propagation Network



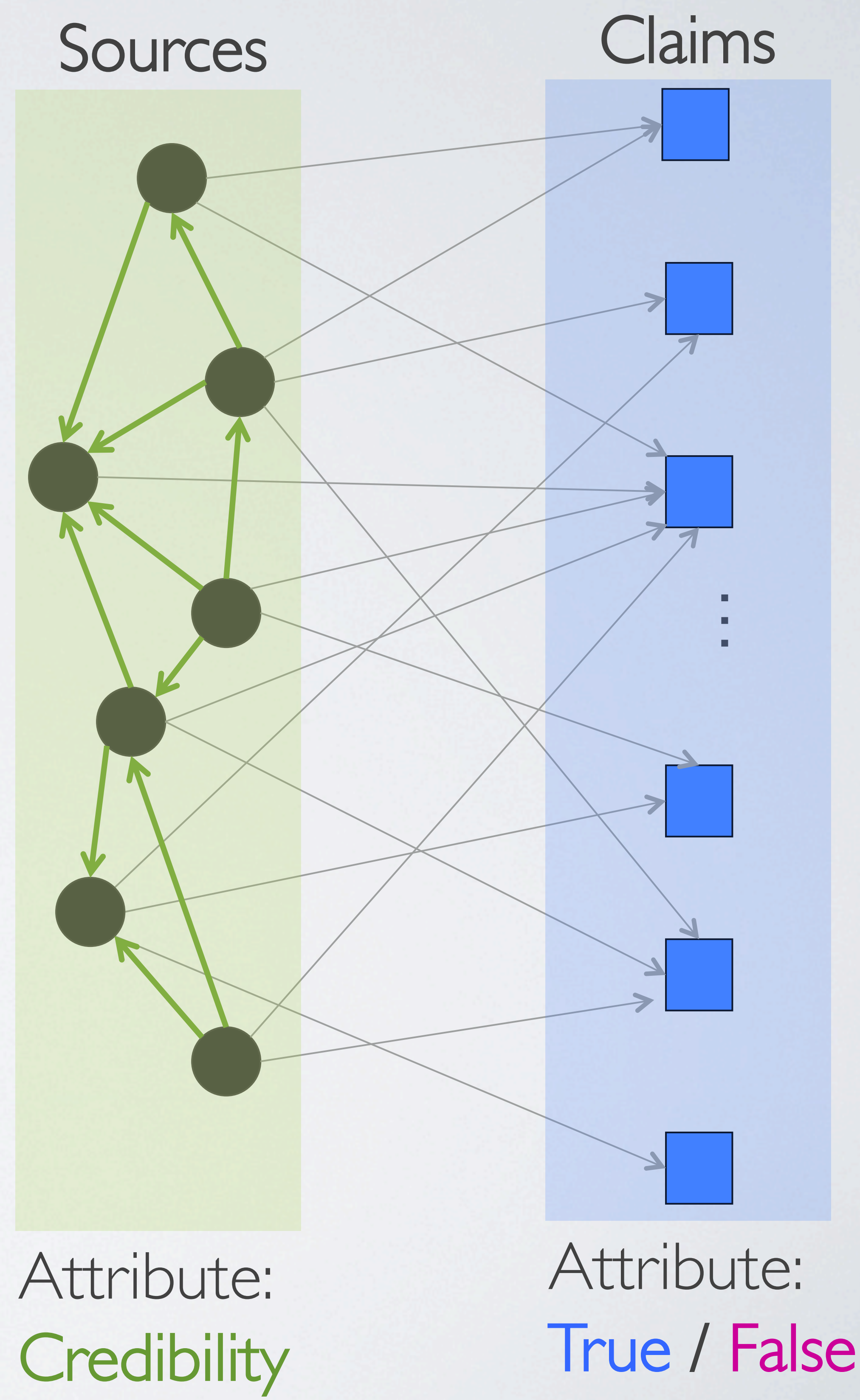
Anti network — Pro network

Effect of Polarization

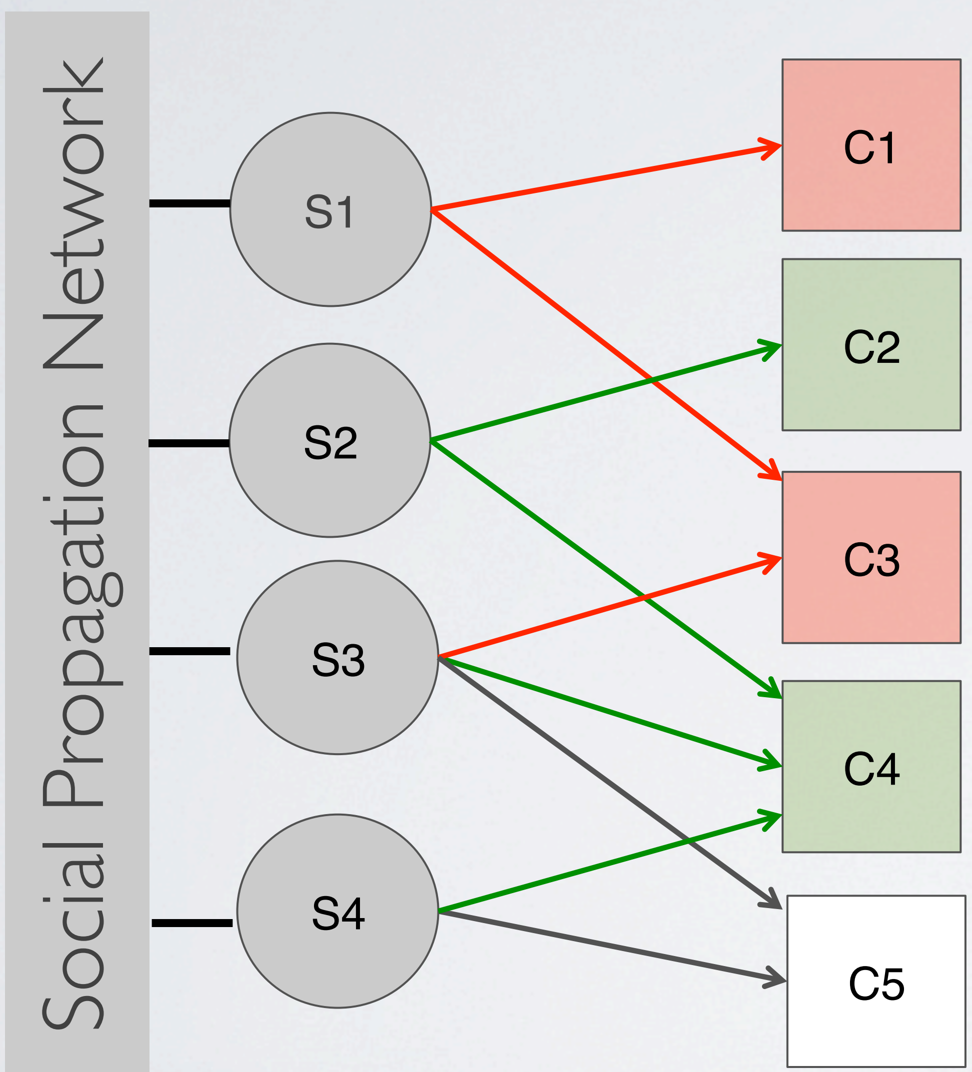
- When sources are biased towards a topic, their observation errors on that topic are more correlated.
  - When they do not share a bias, errors are independent.
  - Corroboration among correlated sources carry less statistical weight than when they are independent.
- Polarity unaware algorithm improperly computes the correlation between sources.

Polarity-aware Fact Finder

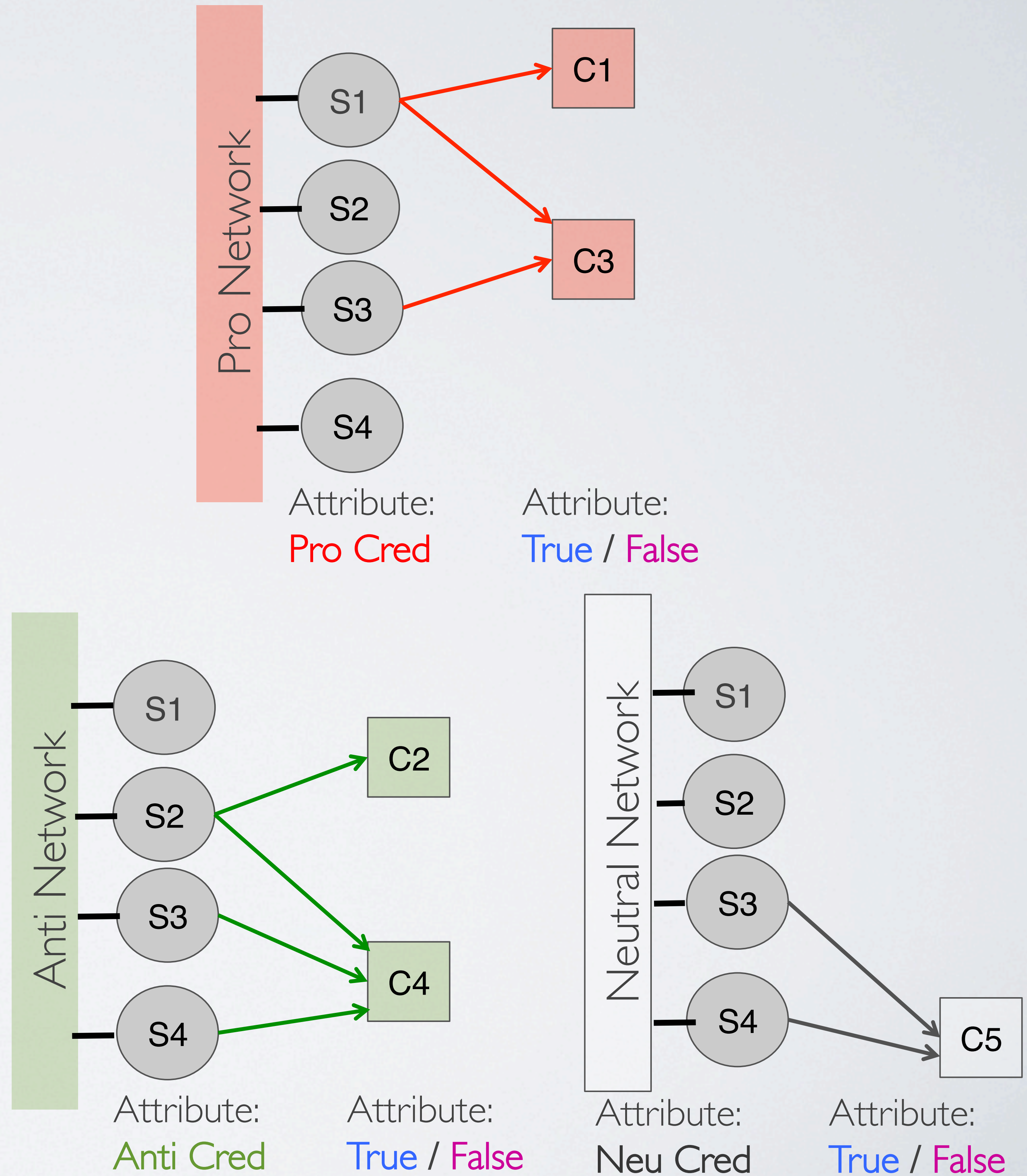
- Computes the latent networks from pro, anti, and neutral claims (SIGMETRICS 2012)
- Uses each network to estimate correlated errors in a manner that depends on content type.
- Accounts for correlation in credibility analysis.



Polarity-aware Fact Finder



Original Problem



Evaluation

	Input	Combined	Polarized
Pro Claims	199	147	128
Anti Claims	109	88	76
Neutral Claims	692	543	496

Evaluation

	Set A	Set B
Definition	Polarized Exclusive	Combined Exclusive
Total	38	116
Factual	26	82
Non-factual (0)	12	34
True (1)	25	72
False (-1)	1	10
False Claims	2.6 %	8.6 %
Factual True	96 %	88 %

Conclusion

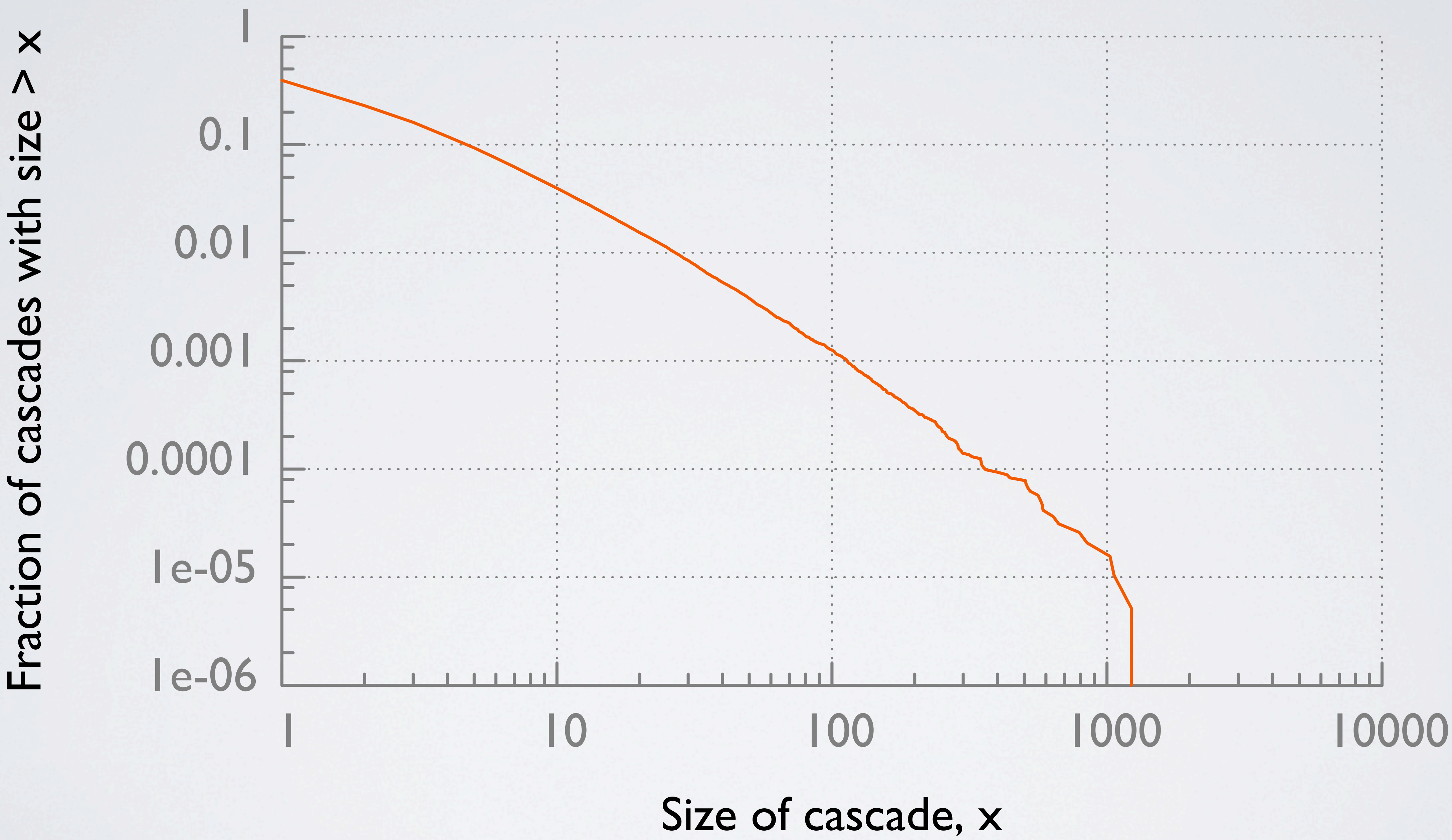
- Separation of claims by polarity prevents estimation of false dependencies between neutral sources.
- Probability of error reduced by factor of three for the factual claims.
- More than 18% improvement in overall Quality of Information.
- Easily extensible to incorporate ML or NLP analysis which may improve the fact-finding performance.
- Idea of polarities can be extended to "topics" with arbitrary relations and hierarchy.
- Did not consider adversarial sources.

otázky سوالات Domande प्रश्न
прашања
问题 soalan 質問 pitanja Sorular
კითხვები Ceisteanna tanong Spørgsmål
întrebări Kérdések Հարցեր ຄຳຖາມ

Questions

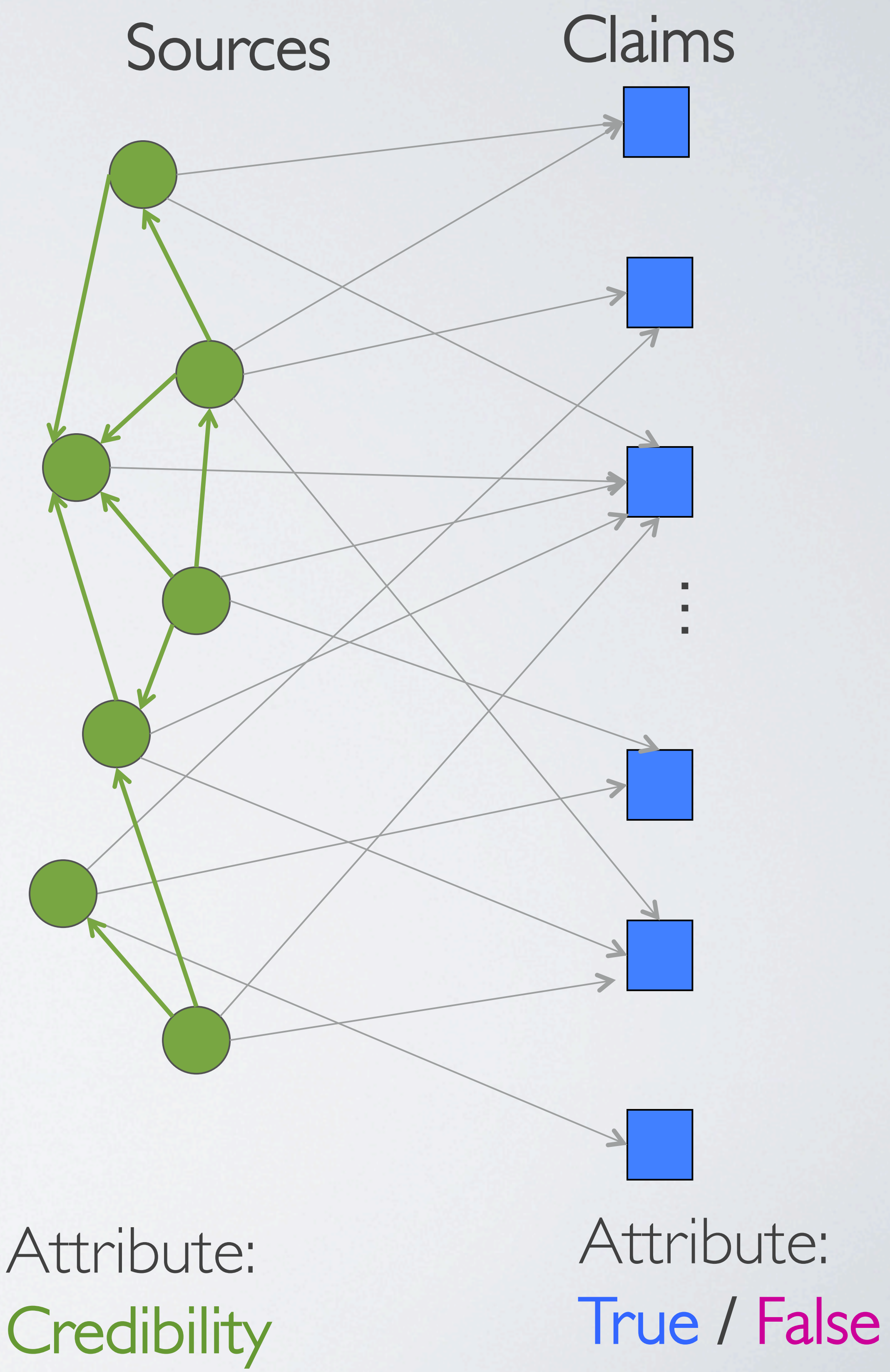
galderak cwestiynau
질문 الأسئلة Küsimused প্রশ্ন Fragen
પ્રશ્નો
คำถามที่ Ερωτήσεις Preguntas
vragen
dotazy Вопросы ਸਵਾਲ שאלות ప్రశ్నలు
פֿראַגן கள்விகள் pertanyaan

Why Largest Cascades?

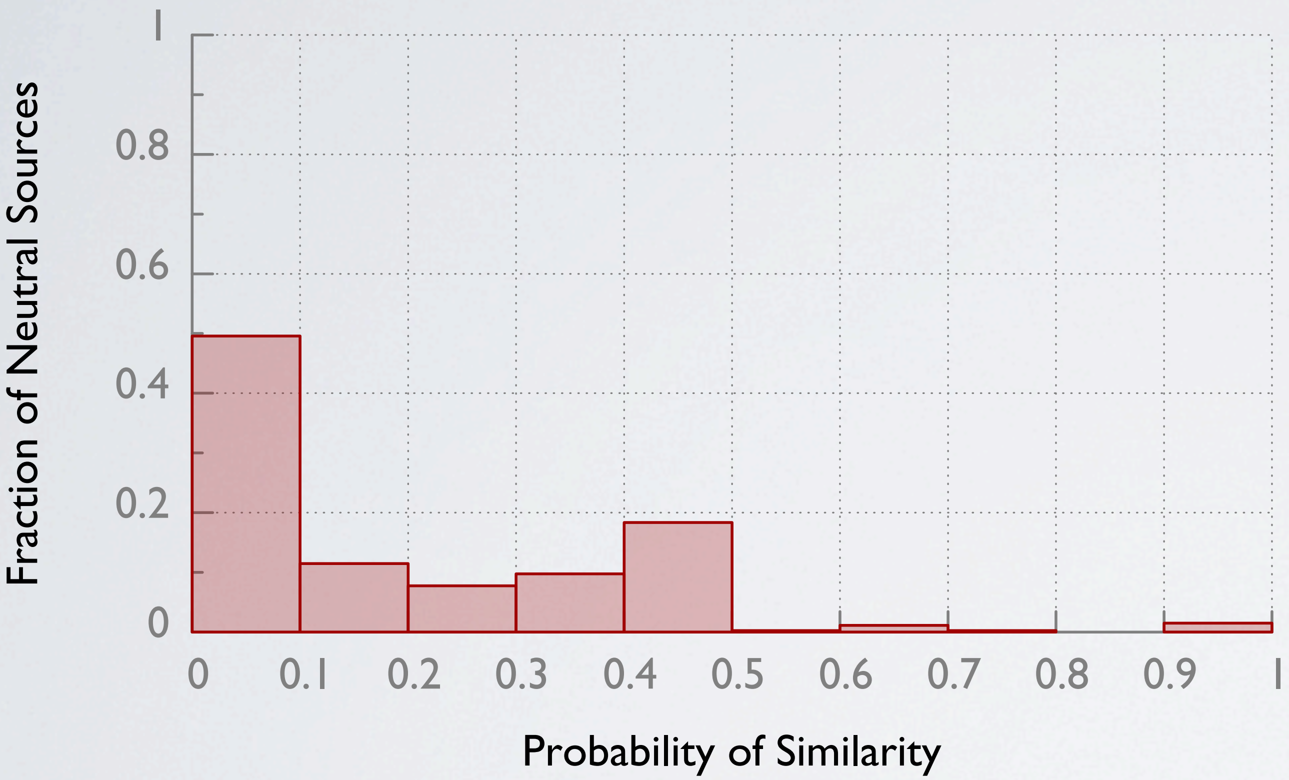


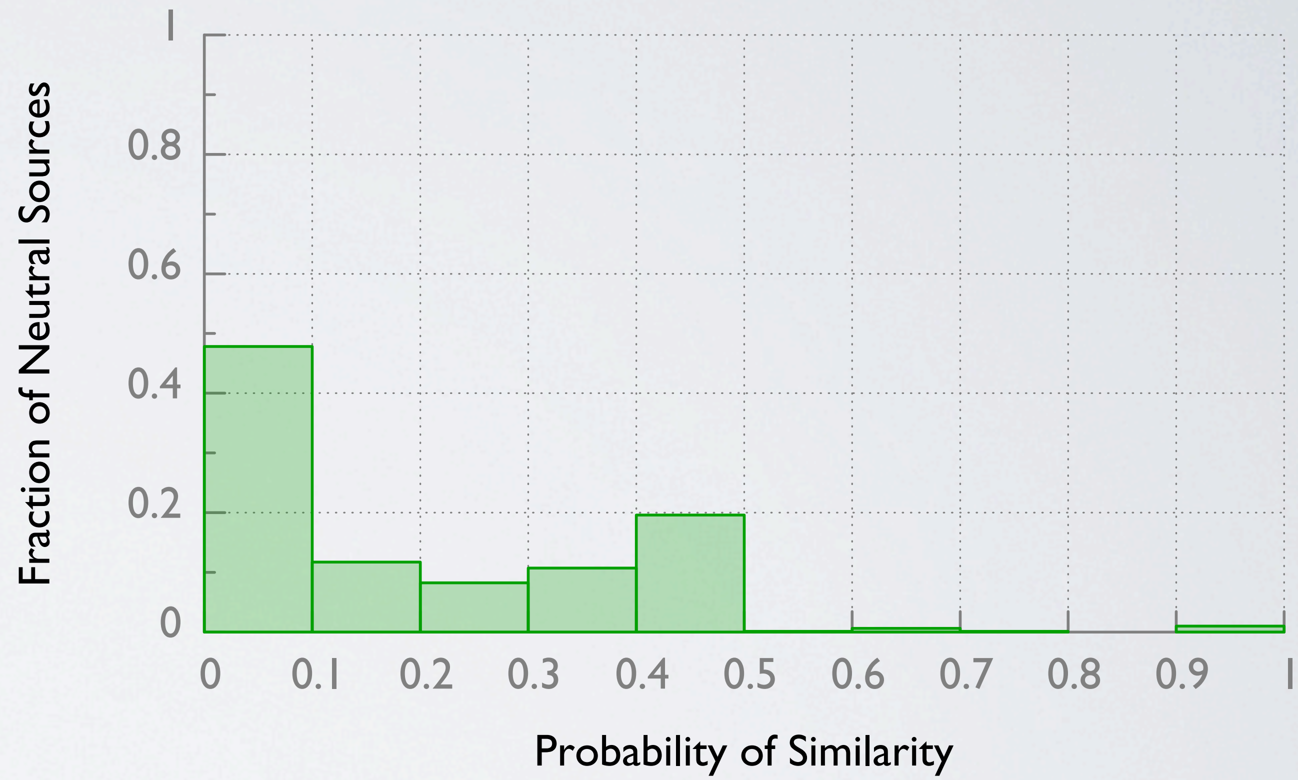
Polarity-aware Fact Finder

- Computes the latent networks from pro, anti, and neutral claims (SIGMETRICS 2012)
- Uses each network to estimate correlated errors in a manner that depends on content type.
- Accounts for correlation in credibility analysis.



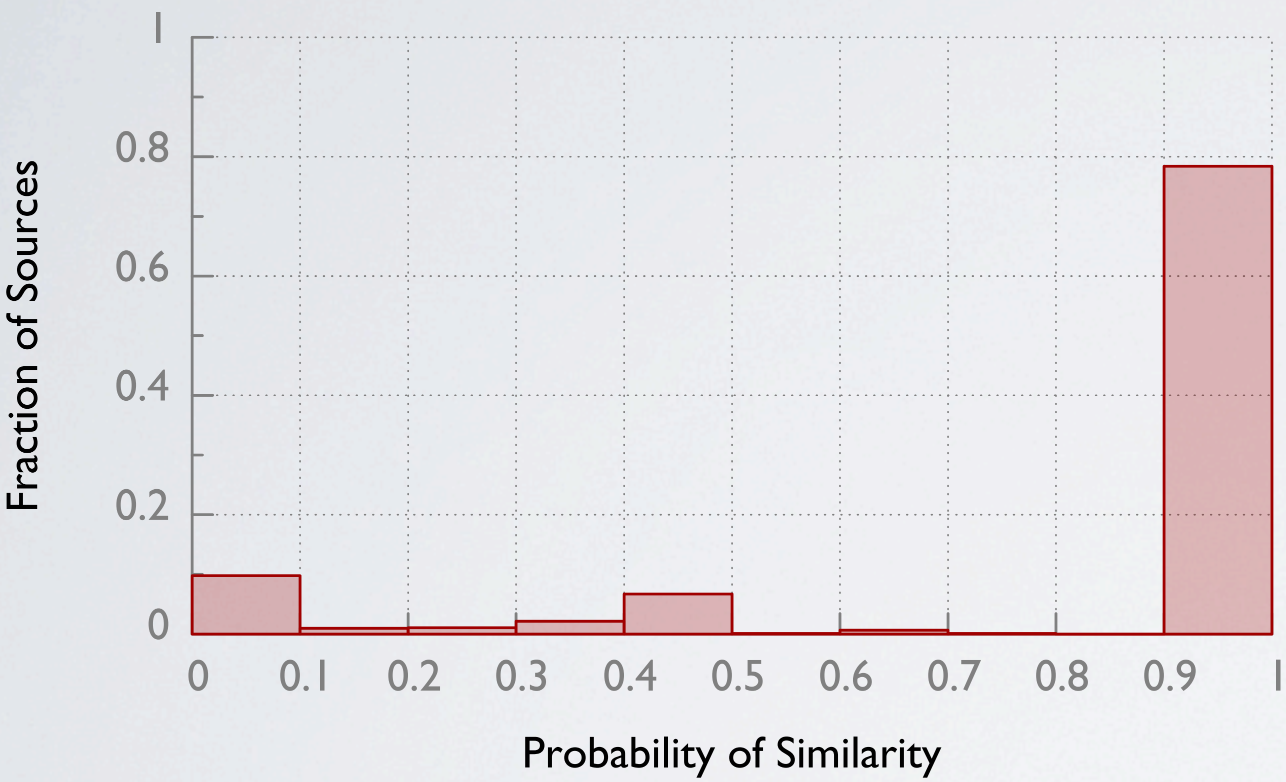
Effect of Polarization

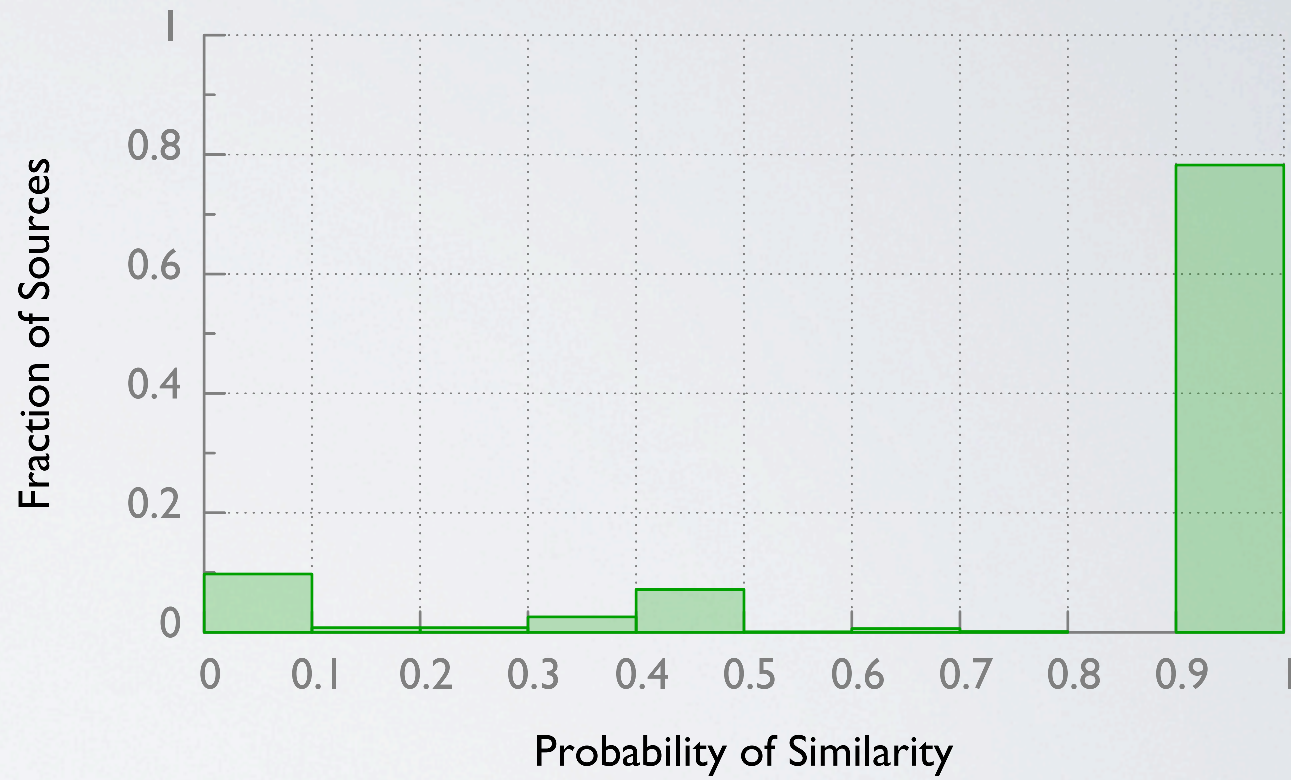




The non polarized algorithm confuses the dependency of the neutral sources

Effect of Polarization





Dependency of the strongly polarized sources are correctly determined by both polarized and non-polarized algorithms

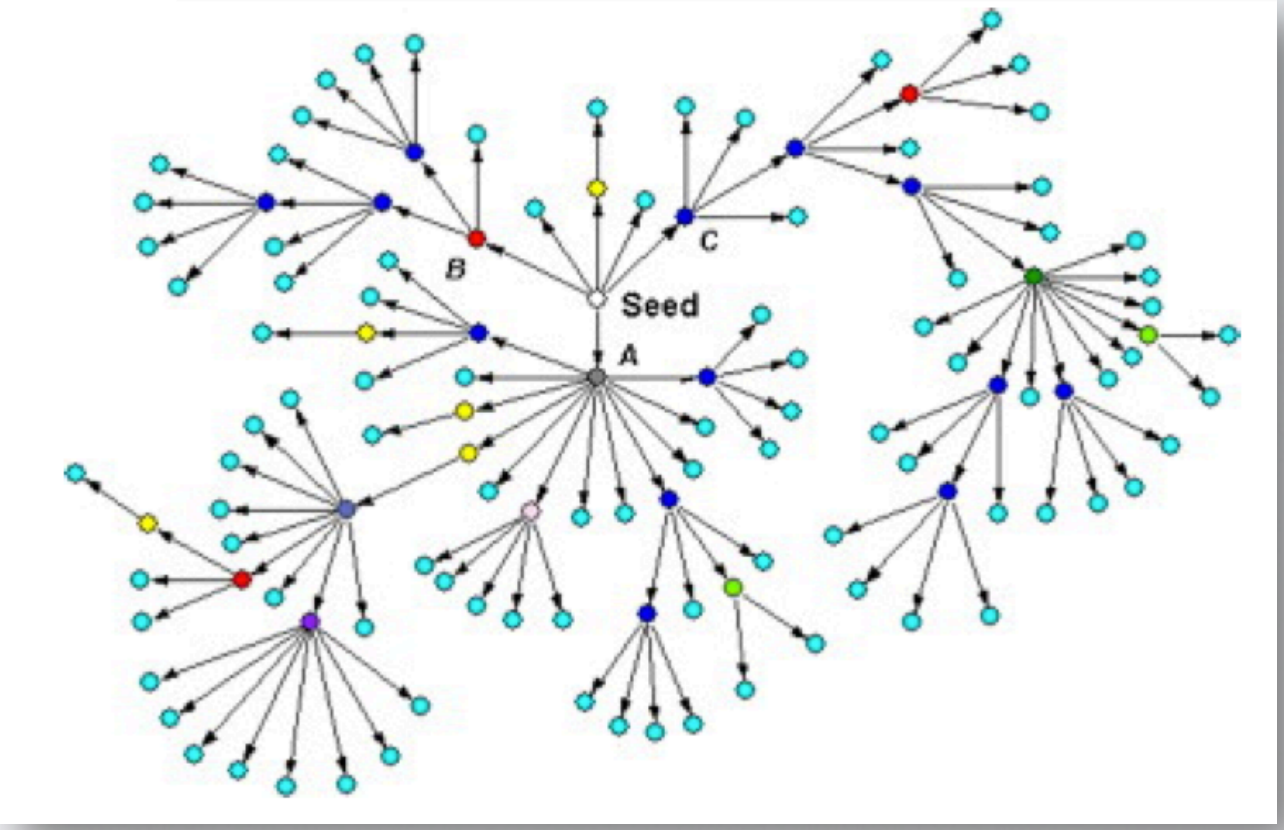
Evaluation

	Input	Combined	Polarized
Pro Claims	199	147	128
Anti Claims	109	88	76
Neutral Claims	692	543	496
Total	1000	778	700

- 662 output claims were common to both algorithms

Epidemic Propagation

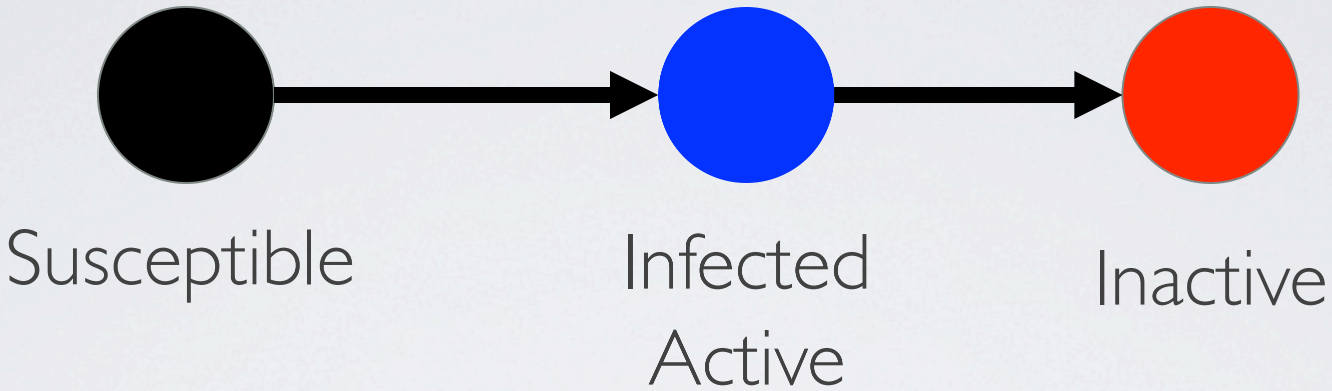
- Nodes get infected, infect other nodes, and the process continues resulting in a propagation graph.

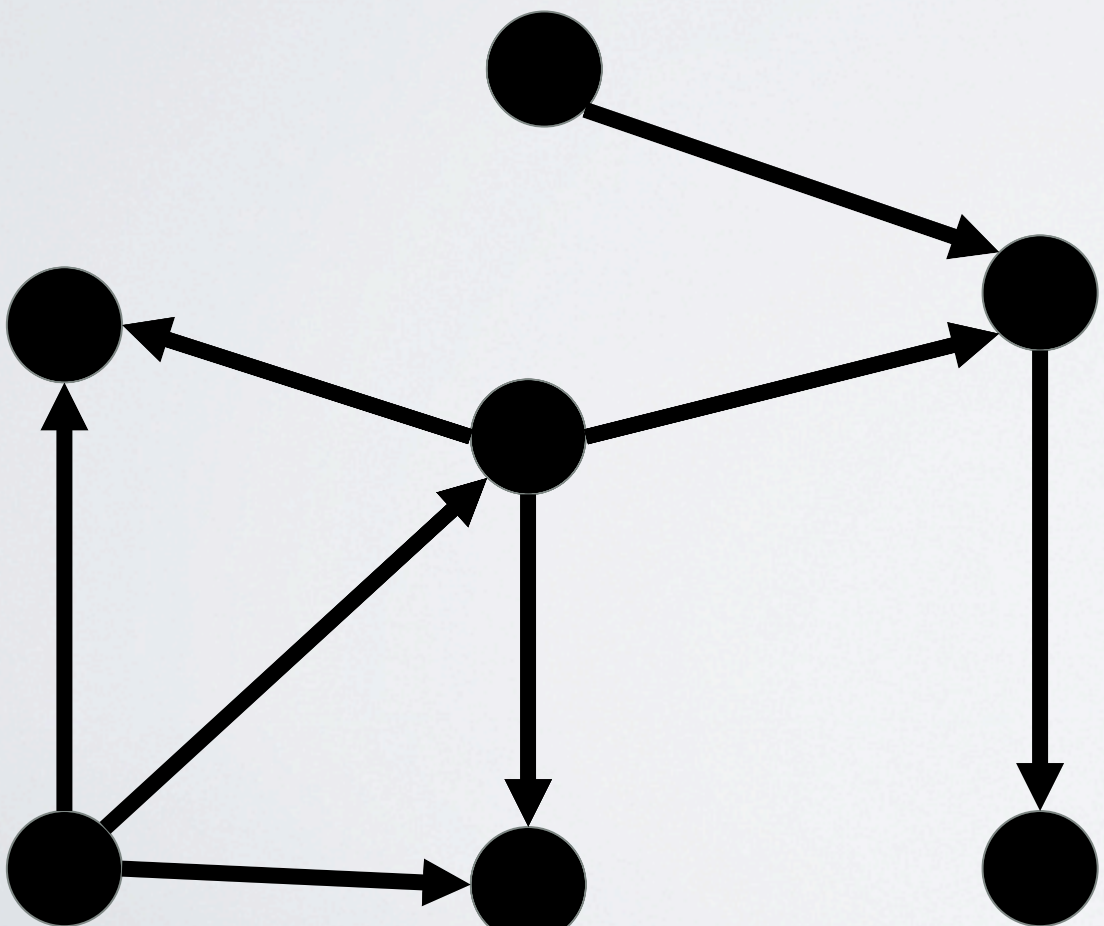


- Inverse Problem*: **Given the observation of infections, find the structure.**

* Praneeth Netrapalli, Sujay Sanghavi, Learning the Graph of Epidemic Cascades. SIGMETRICS 2012

Standard Independent Cascade Model*



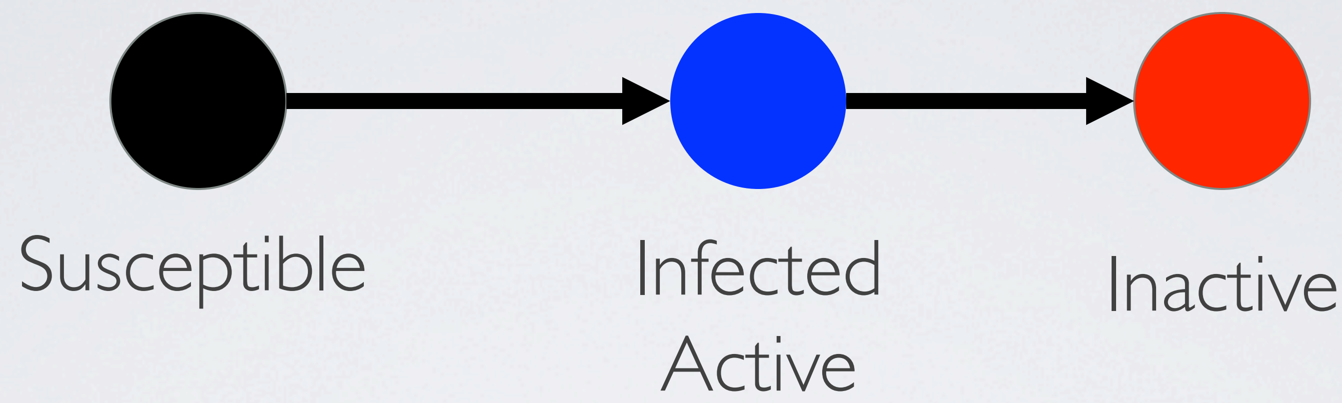


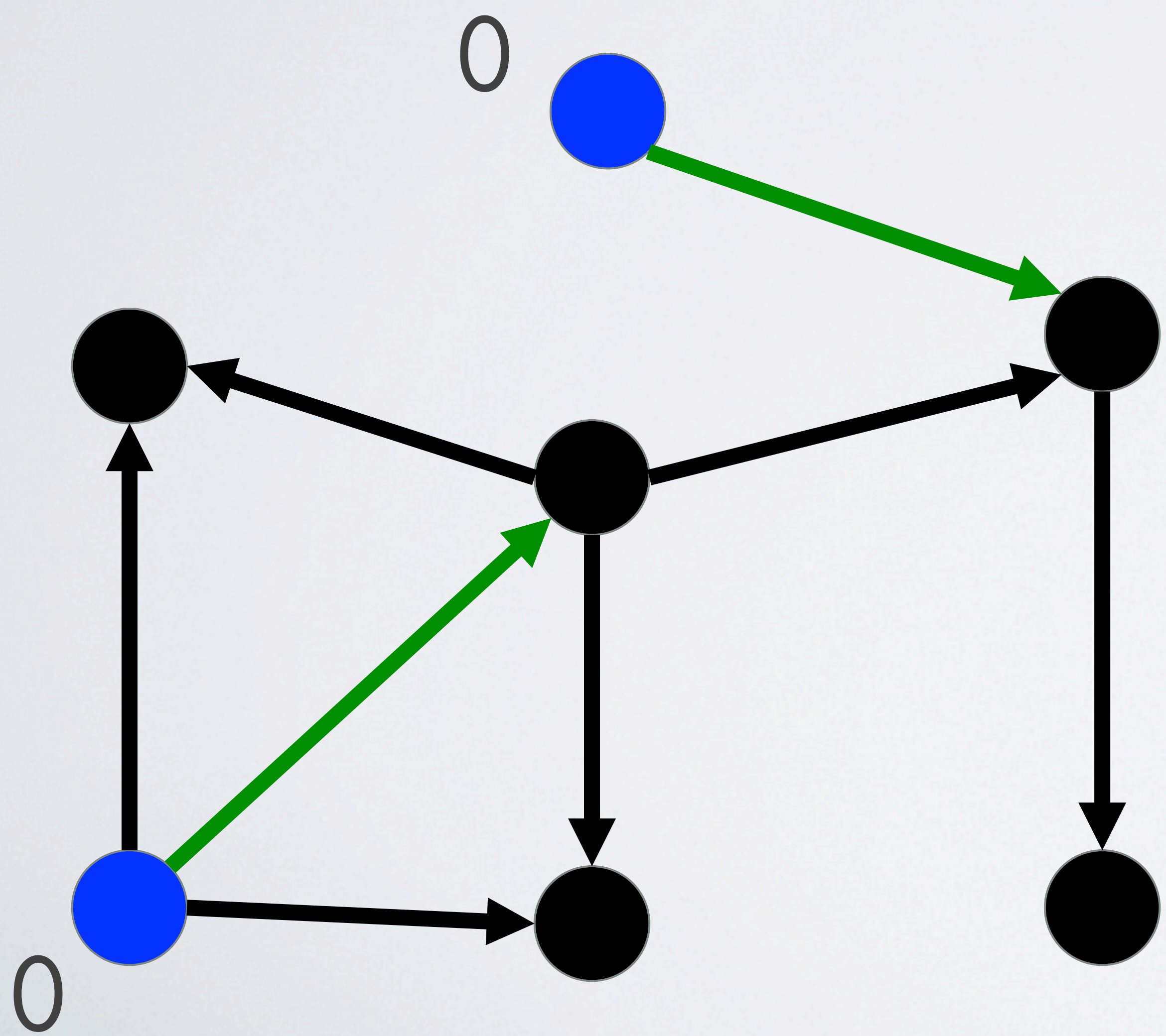
- Proceed in discrete step
- Initial seeds random with probability p_{init}
- i infect j with p_{ij}
- Active node inactive after one step

* Proposed by Goldenberg, Libai, and Muller, "Talk of the Network: A complex systems look at the underlying process of word-of-mouth. Marketing Letters 2001. Also appears in Kempe, Kleinberg, and Tardos "Maximizing the spread of influence through a social network, KDD 03

- Illustrative example taken from author's presentation

Standard Independent Cascade Model

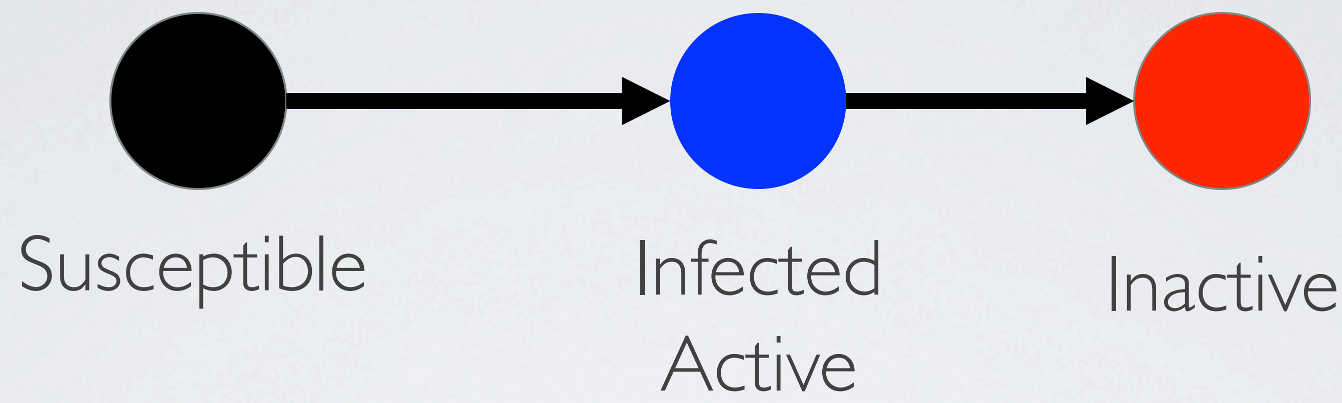


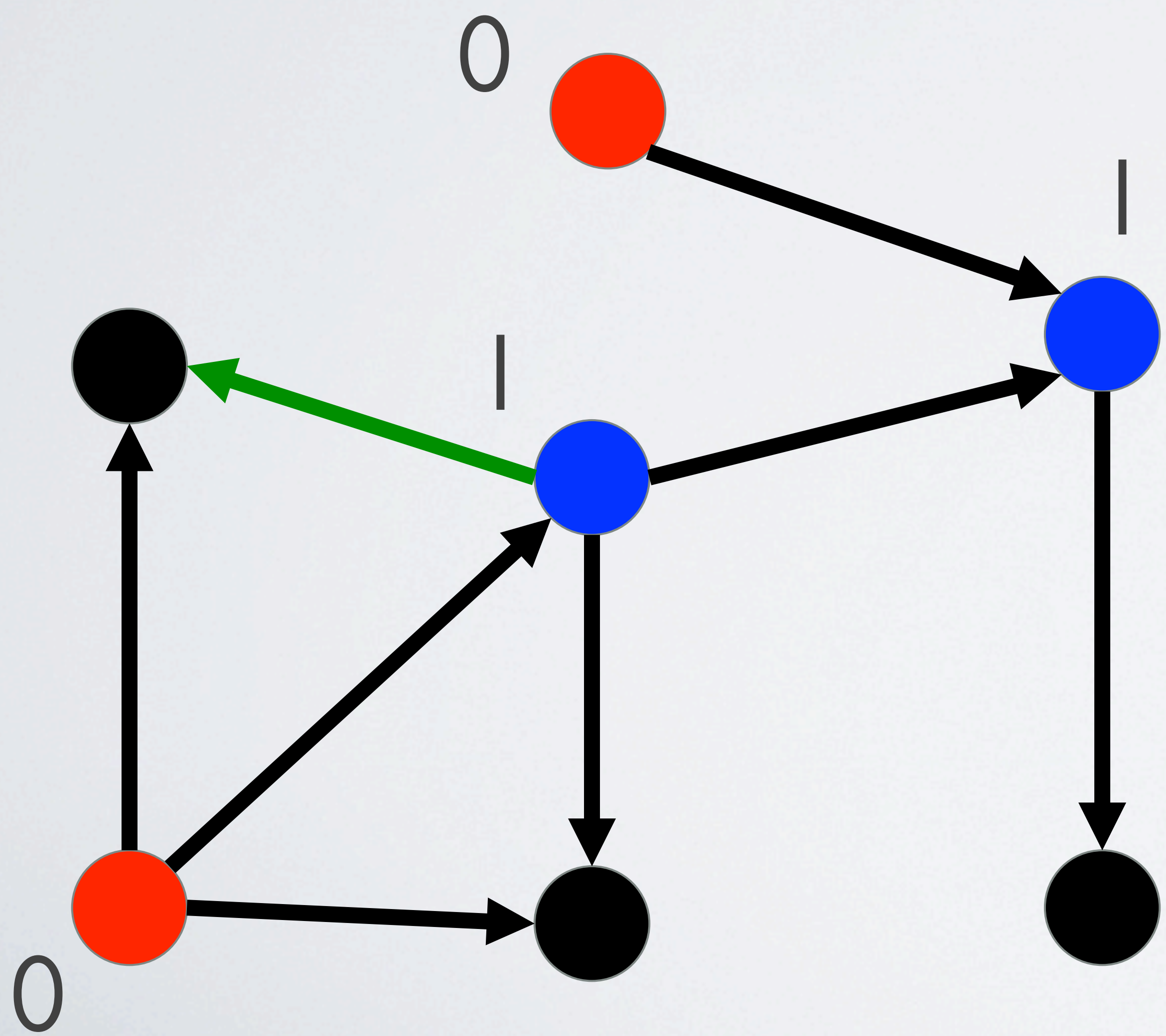


- Proceed in discrete step
- Initial seeds random with probability p_{init}
- i infect j with p_{ij}
- Active node inactive after one step

* Praneeth Netrapalli, Sujay Sanghavi, Learning the Graph of Epidemic Cascades. SIGMETRICS 2012

Standard Independent Cascade Model

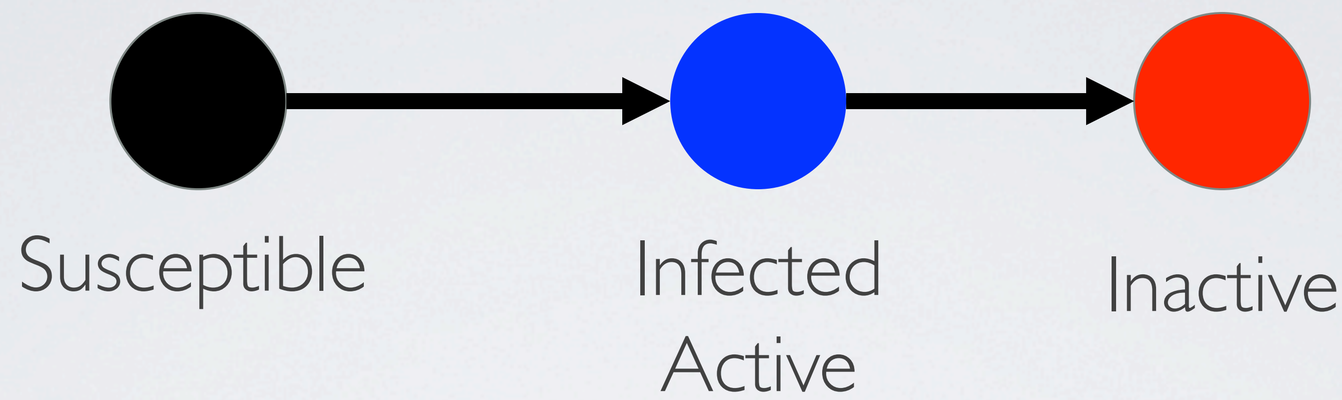


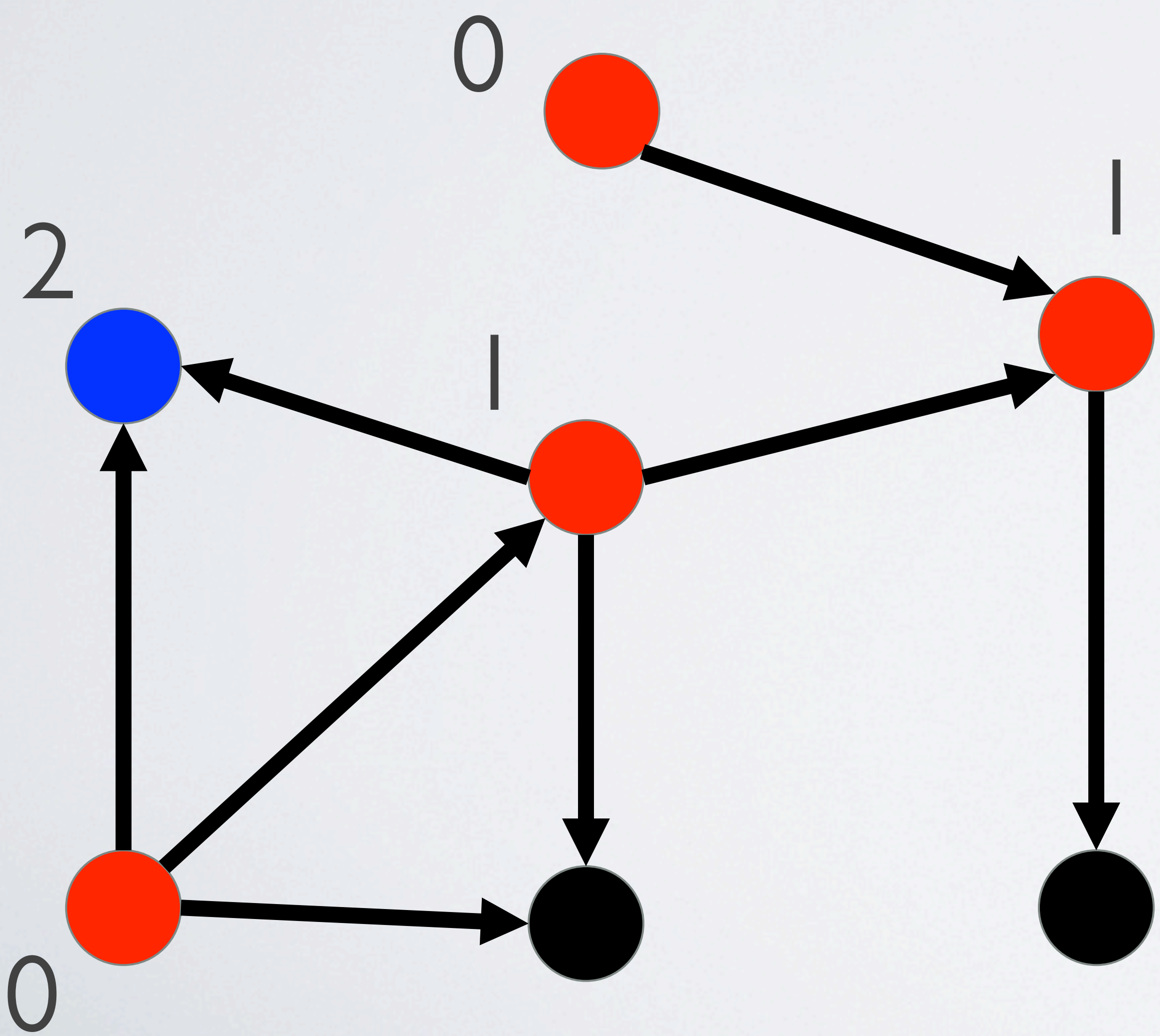


- Proceed in discrete step
- Initial seeds random with probability p_{init}
- i infect j with p_{ij}
- Active node inactive after one step

* Praneeth Netrapalli, Sujay Sanghavi, Learning the Graph of Epidemic Cascades. SIGMETRICS 2012

Standard Independent Cascade Model

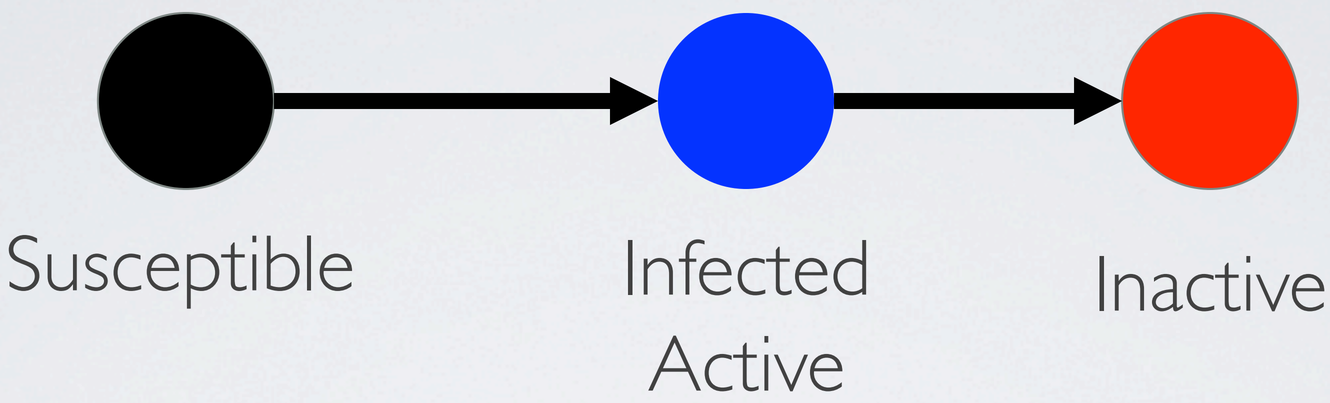


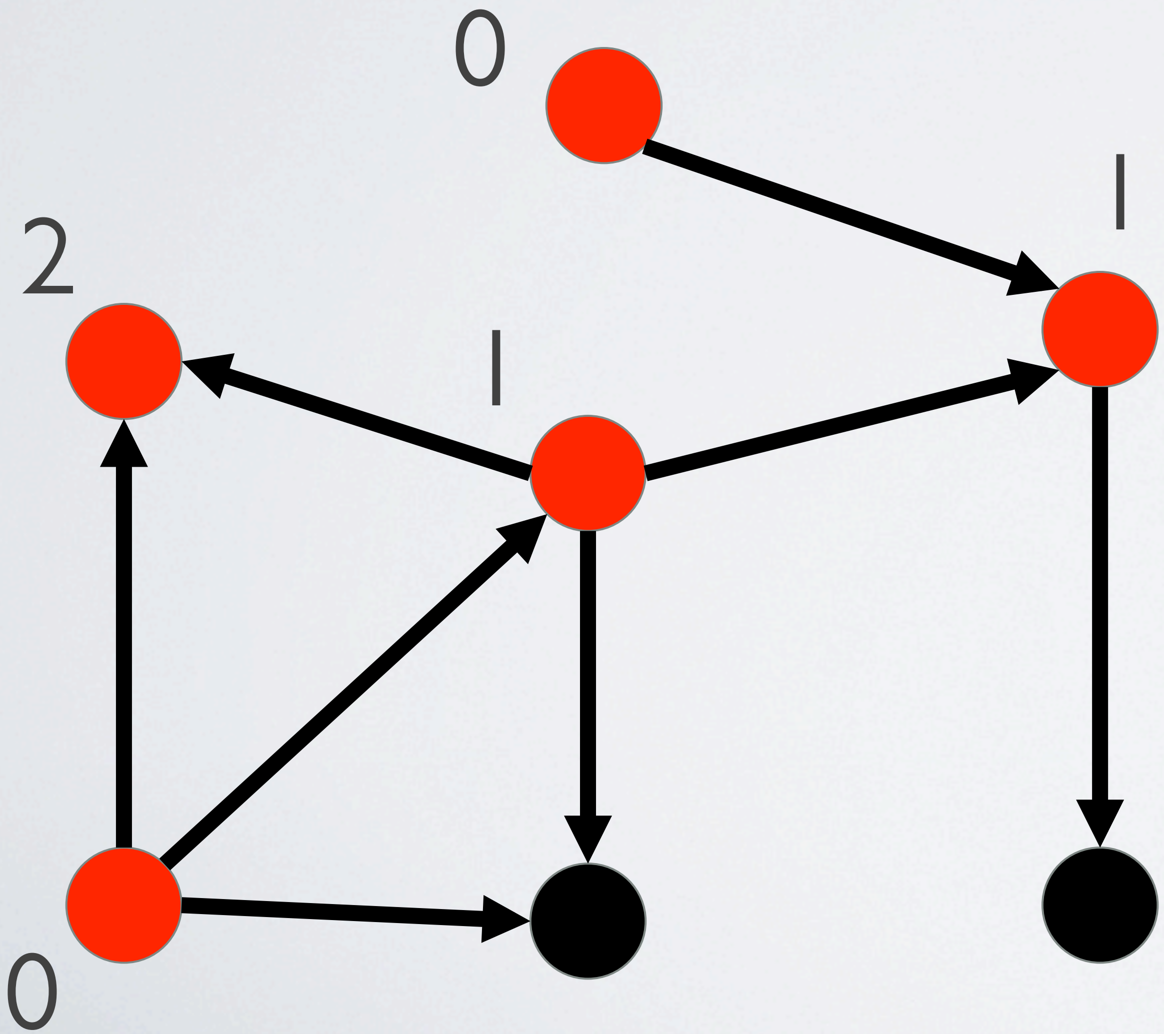


- Proceed in discrete step
- Initial seeds random with probability p_{init}
- i infect j with p_{ij}
- Active node inactive after one step

* Praneeth Netrapalli, Sujay Sanghavi, Learning the Graph of Epidemic Cascades. SIGMETRICS 2012

Standard Independent Cascade Model



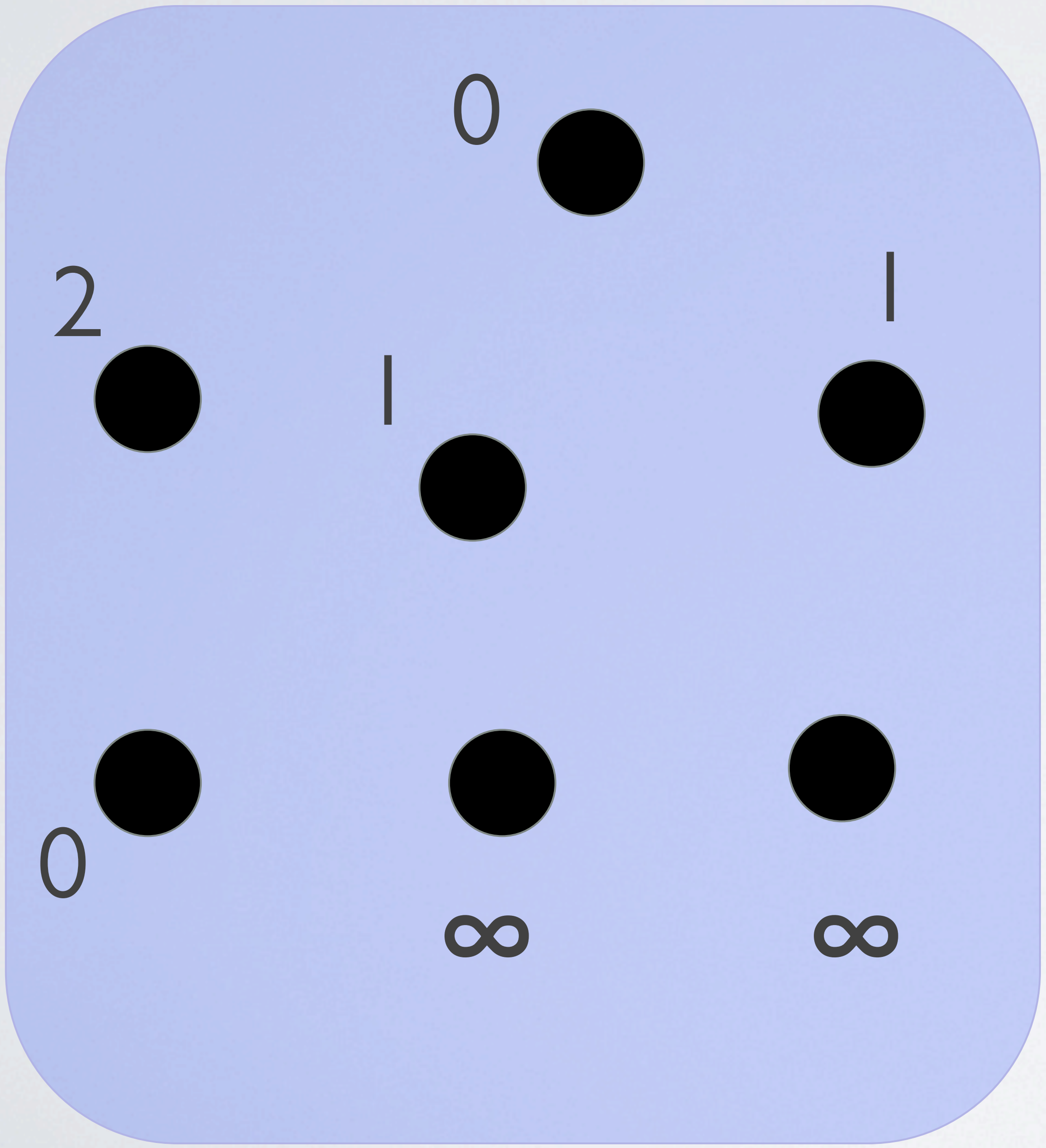


- Proceed in discrete step
- Initial seeds random with probability p_{init}
- i infect j with p_{ij}
- Active node inactive after one step

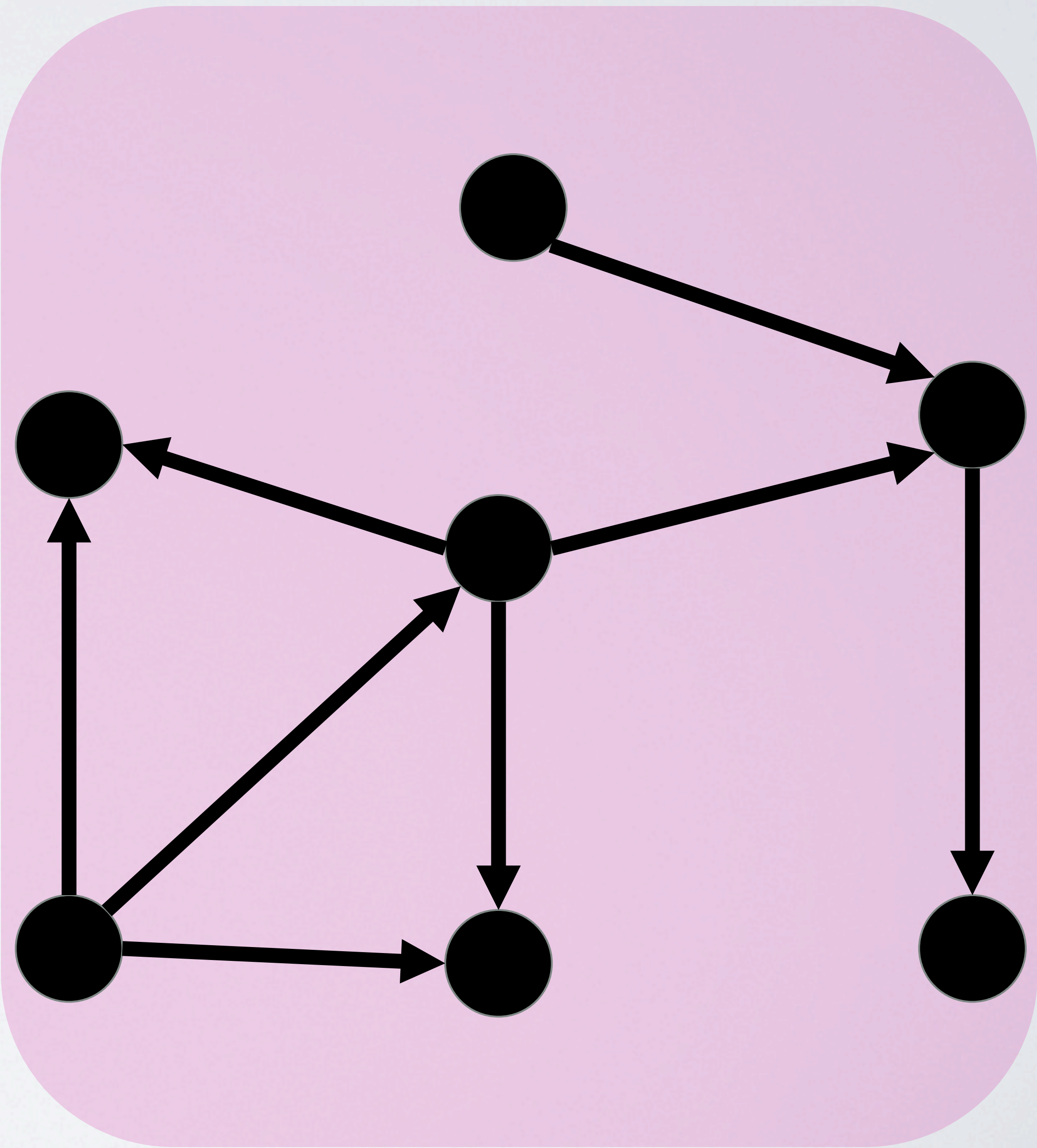
* Praneeth Netrapalli, Sujay Sanghavi, Learning the Graph of Epidemic Cascades. SIGMETRICS 2012

Structure Learning Problem

Given node activation times

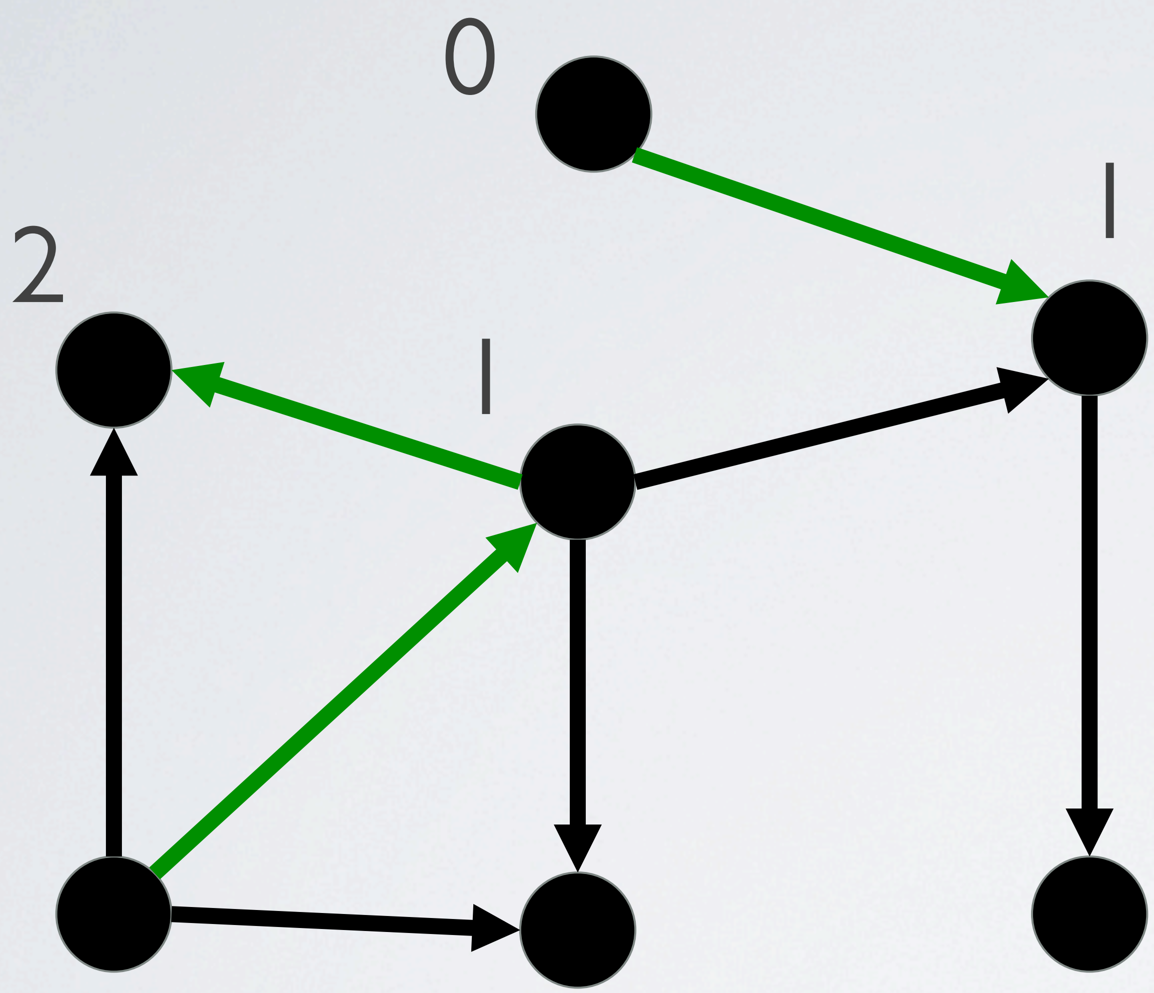


Find network structure



* Praneeth Netrapalli, Sujay Sanghavi, Learning the Graph of Epidemic Cascades. SIGMETRICS 2012

Structure Learning Problem



The example cascade had no evidence of the black edges. Hence, one cascade is not sufficient to learn the structure.

Q: How many cascades necessary?

Q: Given the cascades, how to find the structure?

* Praneeth Netrapalli, Sujay Sanghavi, Learning the Graph of Epidemic Cascades. SIGMETRICS 2012