

The Wisdom of Crowds: Network effects, and the Importance of Experts

Aris Anagnostopoulos
Sapienza University of Rome

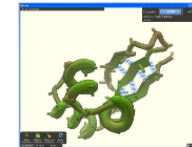
Online collaboration systems

Systems creating knowledge by massive online collaboration:

- Tagging/geotagging systems:



- Games with a purpose:



- Content creation systems:



- Crowdsourcing:



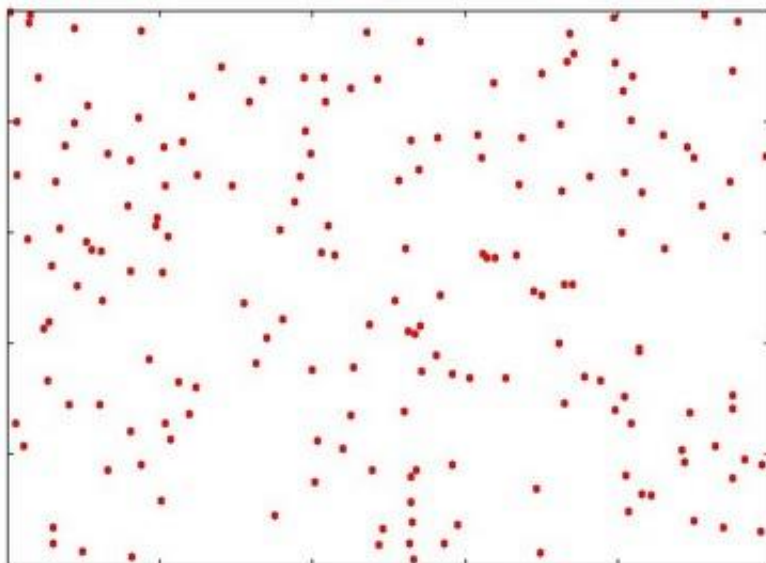
- Open source community:



- Polymath project:

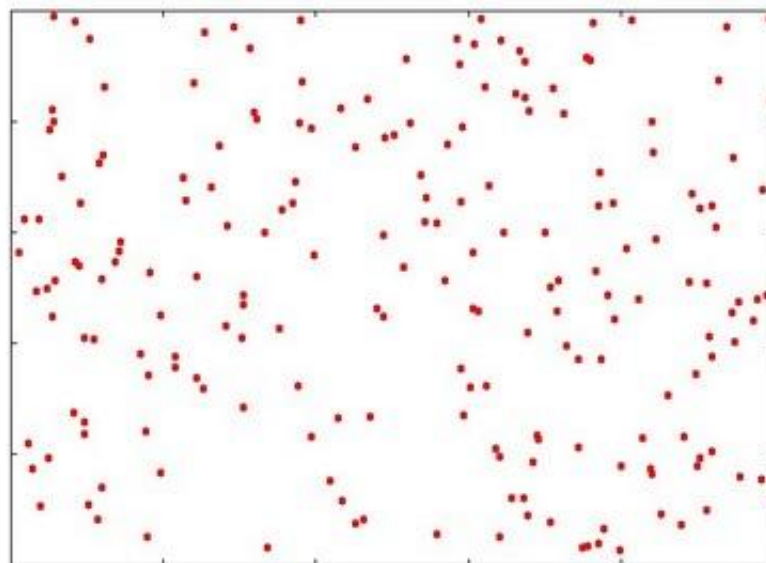


Which photo has more dots?



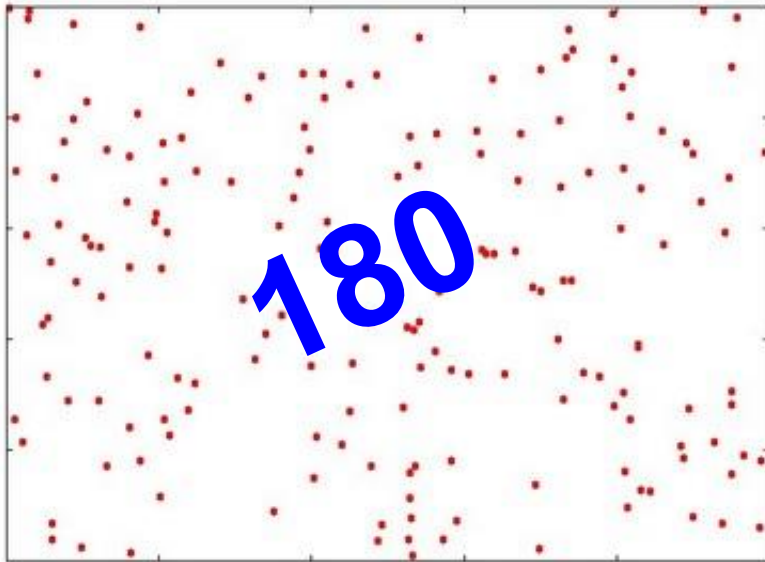
FIRST PICTURE

VS.



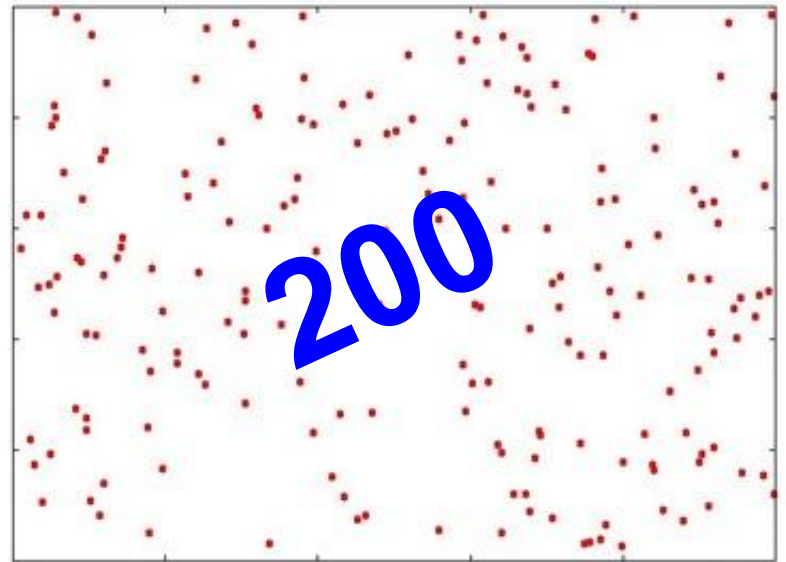
SECOND PICTURE

Which photo has more dots?



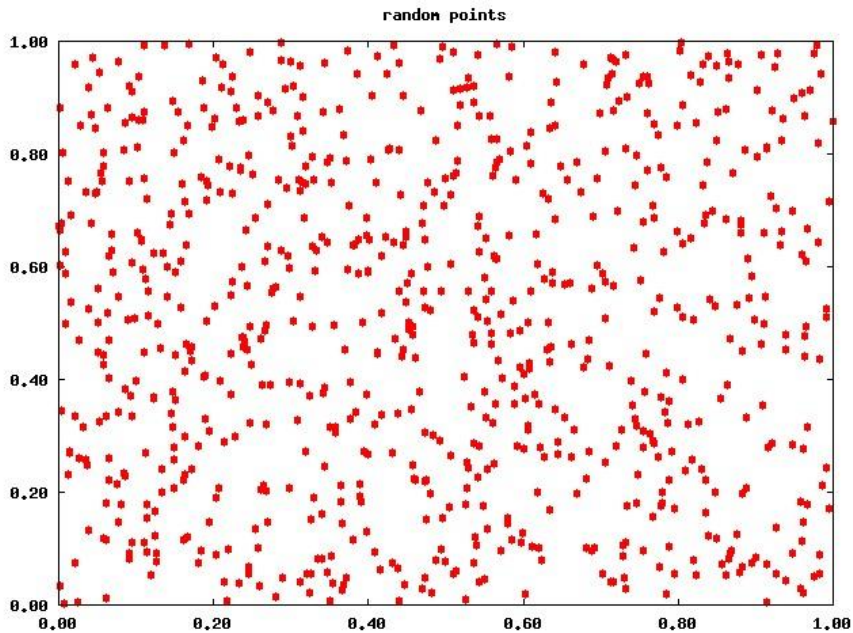
FIRST PICTURE

vs.



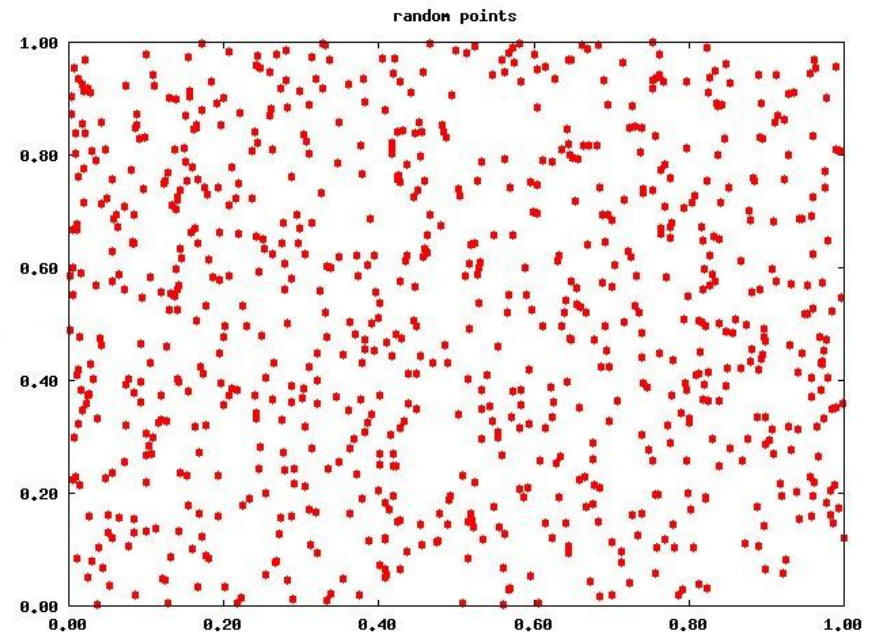
SECOND PICTURE

Which photo has more dots?



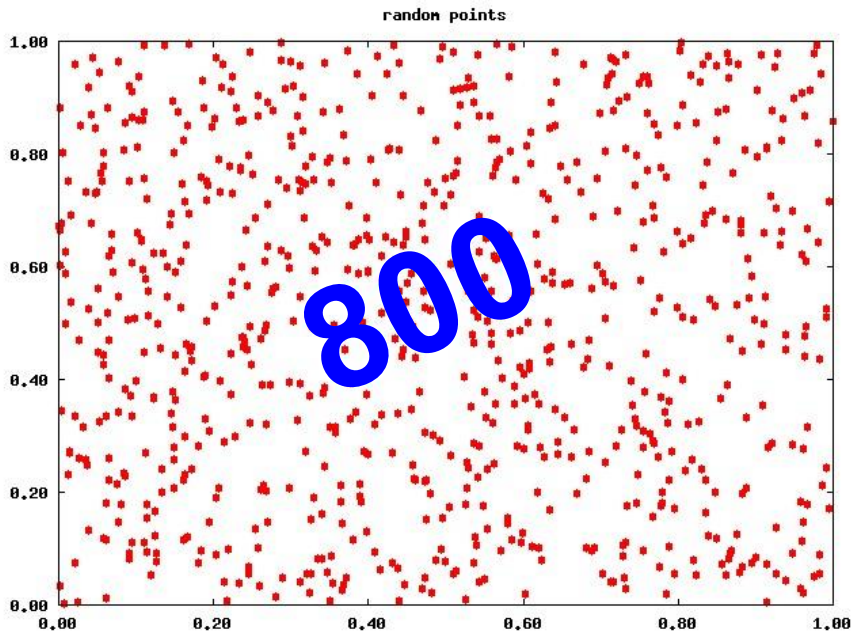
FIRST PICTURE

VS.



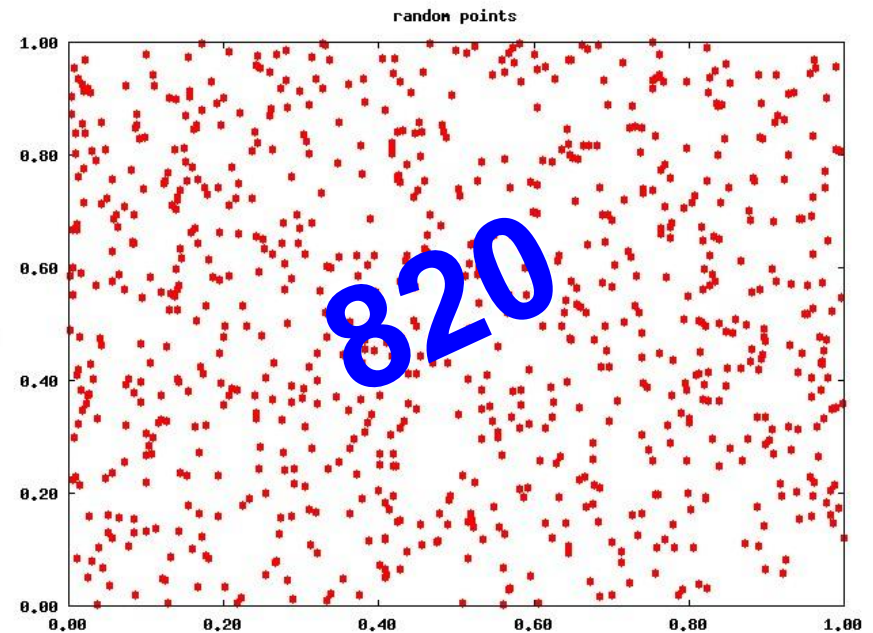
SECOND PICTURE

Which photo has more dots?



FIRST PICTURE

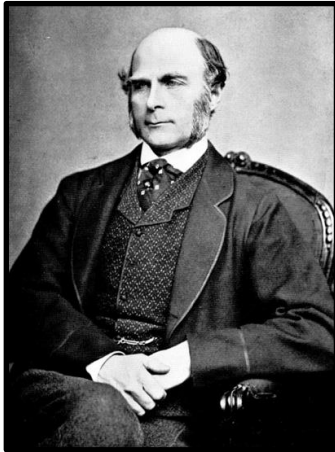
VS.



SECOND PICTURE

Wisdom of crowds – First experiment

At a 1906 country fair in Plymouth, UK, **Sir Francis Galton** made an experiment, asking people to estimate the weight of a slaughtered ox.



What does the ox weigh?
(1198 pounds)

He asked 800 participants.

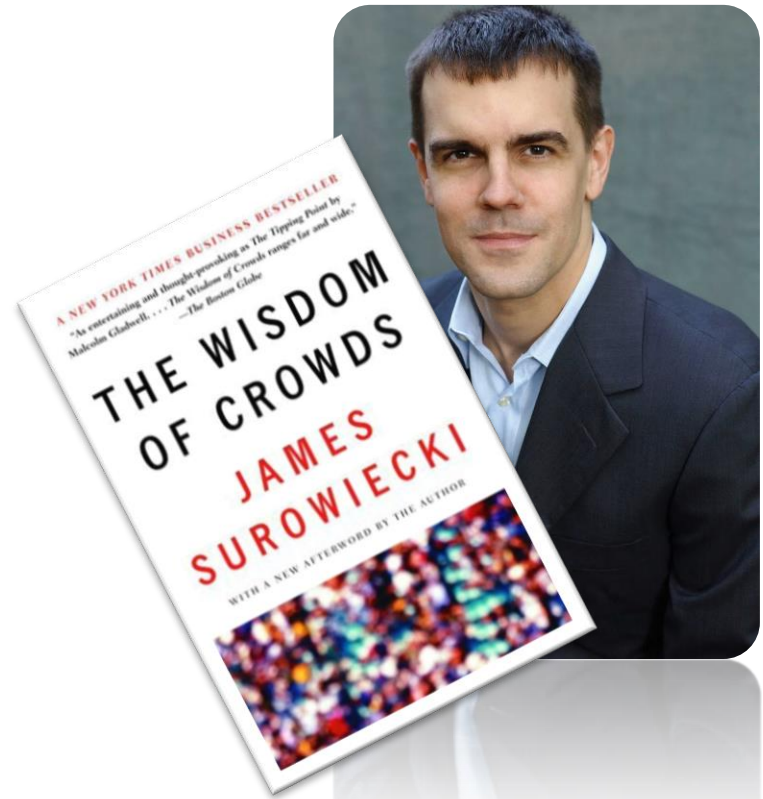
The answers' median was **1207 pounds** (1% error)

The wisdom of crowds

The premise of the wisdom of crowds is that averaging the opinion of many individuals on a topic can give accurate answers.

Examples and applications:

- Francis Galton experiment
- Who wants to be a millionaire
- Recommendation systems
- Prediction markets
- Twitter
- Democracy
- The book of James Surowiecki has many examples



This talk

We will look at three dimensions of the problem:

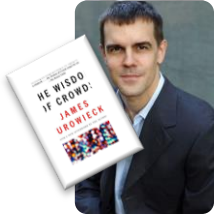
- Network effect on the **wisdom of crowds**
- The role of homophily and polarization in the spreading of (mis)information
- How to schedule **experts** in crowdsourcing

This talk

We will look at three dimensions of the problem:

- **Network effect on the wisdom of crowds**
- The role of homophily and polarization in the spreading of (mis)information
- How to schedule experts in crowdsourcing

The wisdom of crowds



Main requirement:

Independence of opinions and diversity

What happens when we talk and influence each other?

Answer: Often bad things

- Think about democracy:
 - Italy, USA, Greece, have voters that keep/kept bringing terrible governments
- GroupThink
- Spread of conspiracy theories

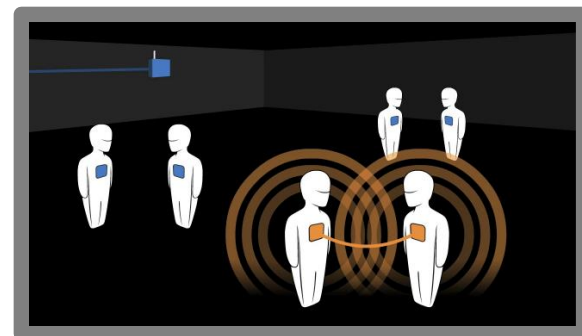
We want to study the **network effect** on the wisdom of crowds in a **natural** setting

Instructions to participants

Instructions:

Phase 1:

- Answer 4 simple questions (5 min)
- Return the answers
- Take and wear an RFID tag

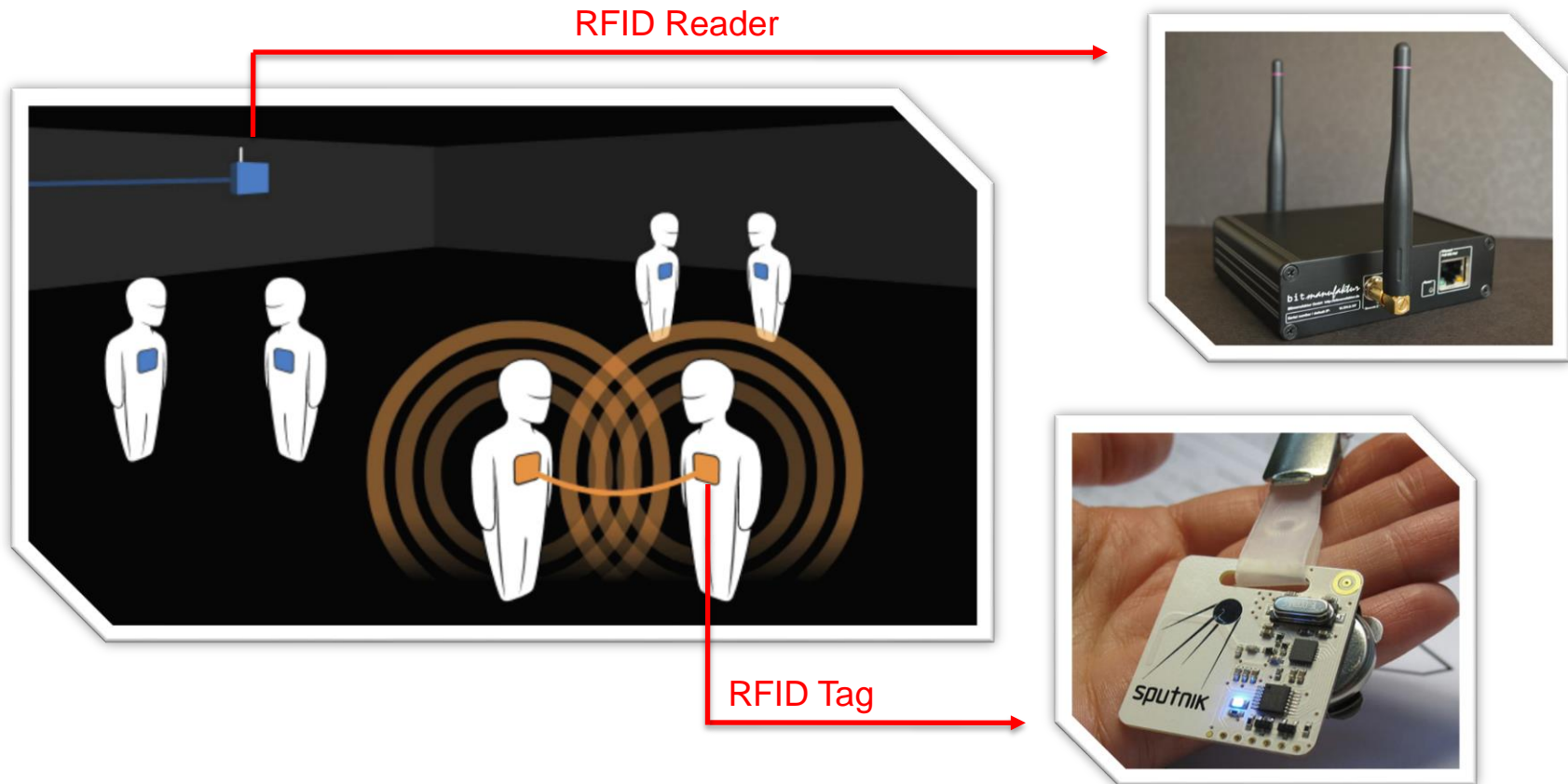


Phase 2

- Discuss the questions with others (20 min)
- At the end answer the questions again and return the tags

Tracking individual interactions

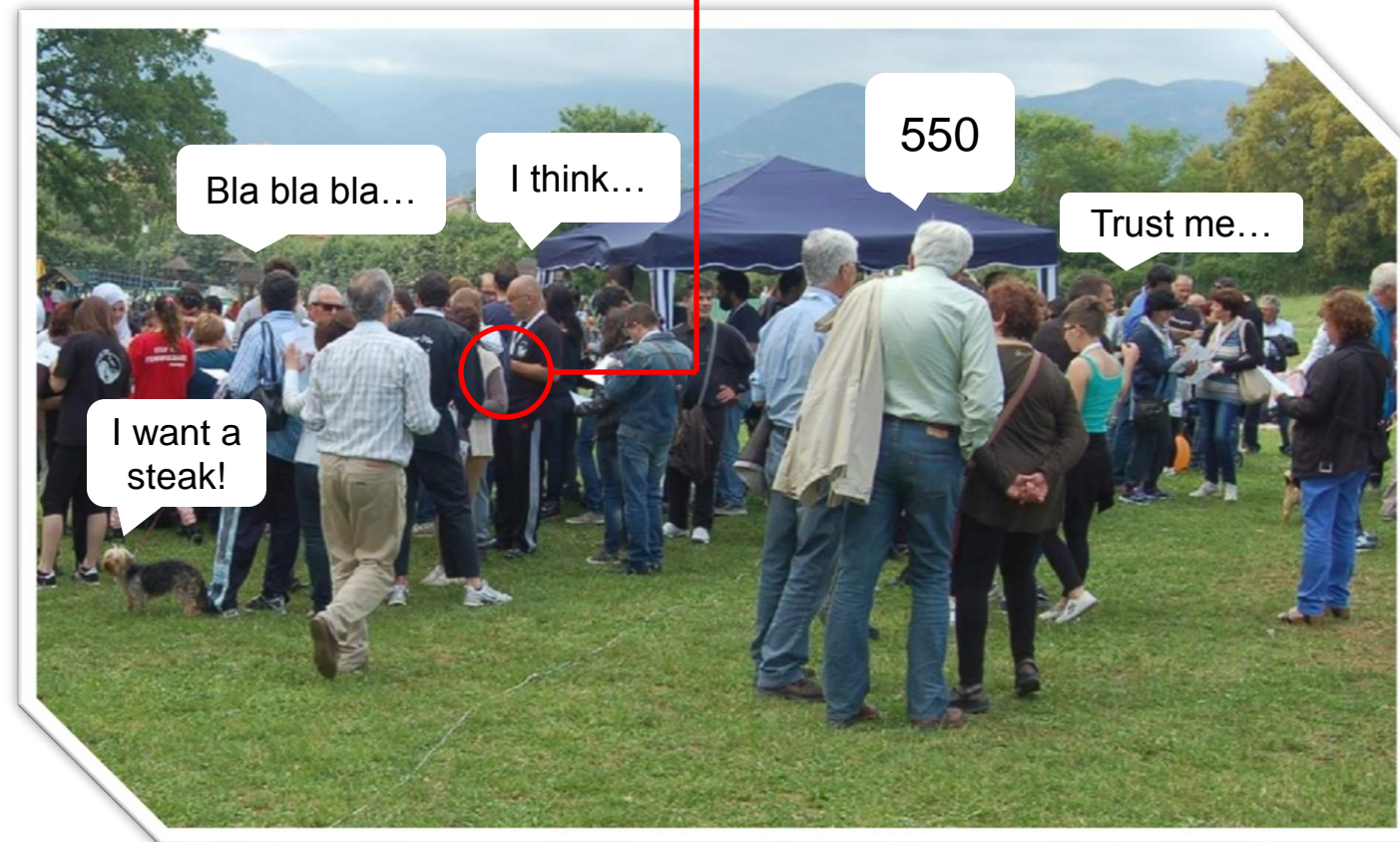
We can use RFID tags to track sustained face-to-face proximity among people.



Collection of F2F interactions

A typical scenario...

Each participant wears an RFID tag



Examples of questions

Innate/Learnt Ability (Class 1)

- How many spaghetti are in the pack?
- How many points are there in the following picture?



Knowledge and Reasoning (Class 2)

- What was the average female population of Italy over the years 1960–1970?
- What is the value in EUR of the coins thrown into the Trevi fountain in 2012?



Prediction (Class 3)

- How many goals in total will the following teams score in the first round (3 games each) of the 2014 Mundial? Brazil, Spain, Greece, Italy, France, Argentina, Germany, Russia
(asked before the mundial... 😊)



Experiments deployed so far

1. WSDM 2013 Conference,
Feb 2013 (69 attendees)



2. My 2013 data mining class,
May 2013 (37 attendees)



3. Priverno's town yearly fair,
May 2014 (60 attendees)

4. My 2014 data mining class,
May 2014 (25 attendees)

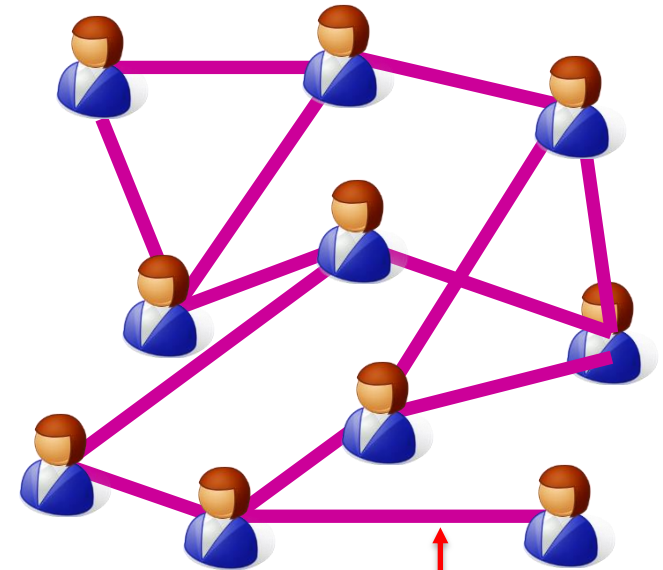


Interaction graph



Representation

An **interaction graph** $G = (V, E)$ represents the interactions between the people.



node $\in V$

edge $\in E$
(interaction)

Interaction graphs

Priverno fair

Undirected graph

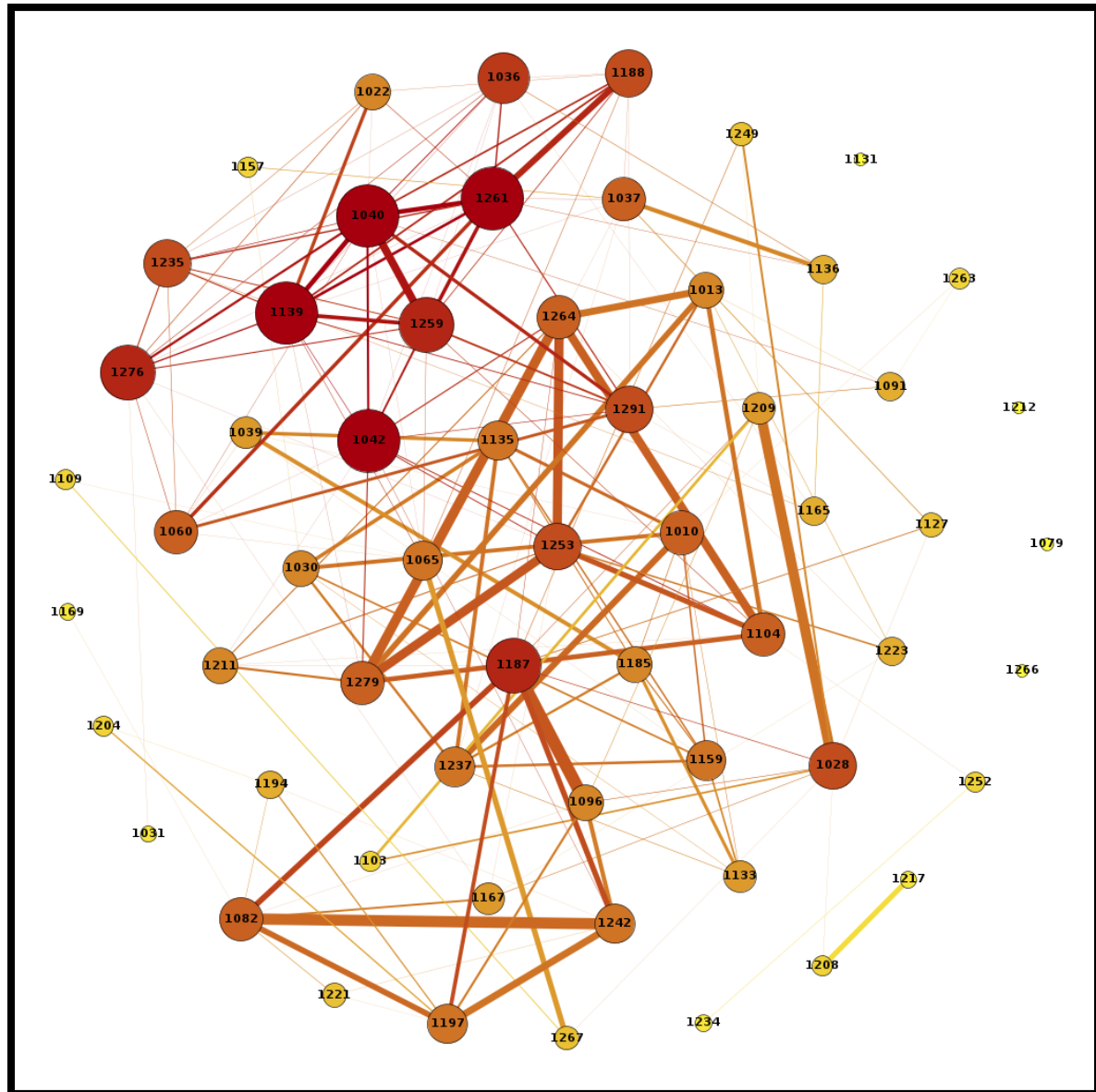
Nodes: 60

Edges: 128

Density: 0.072

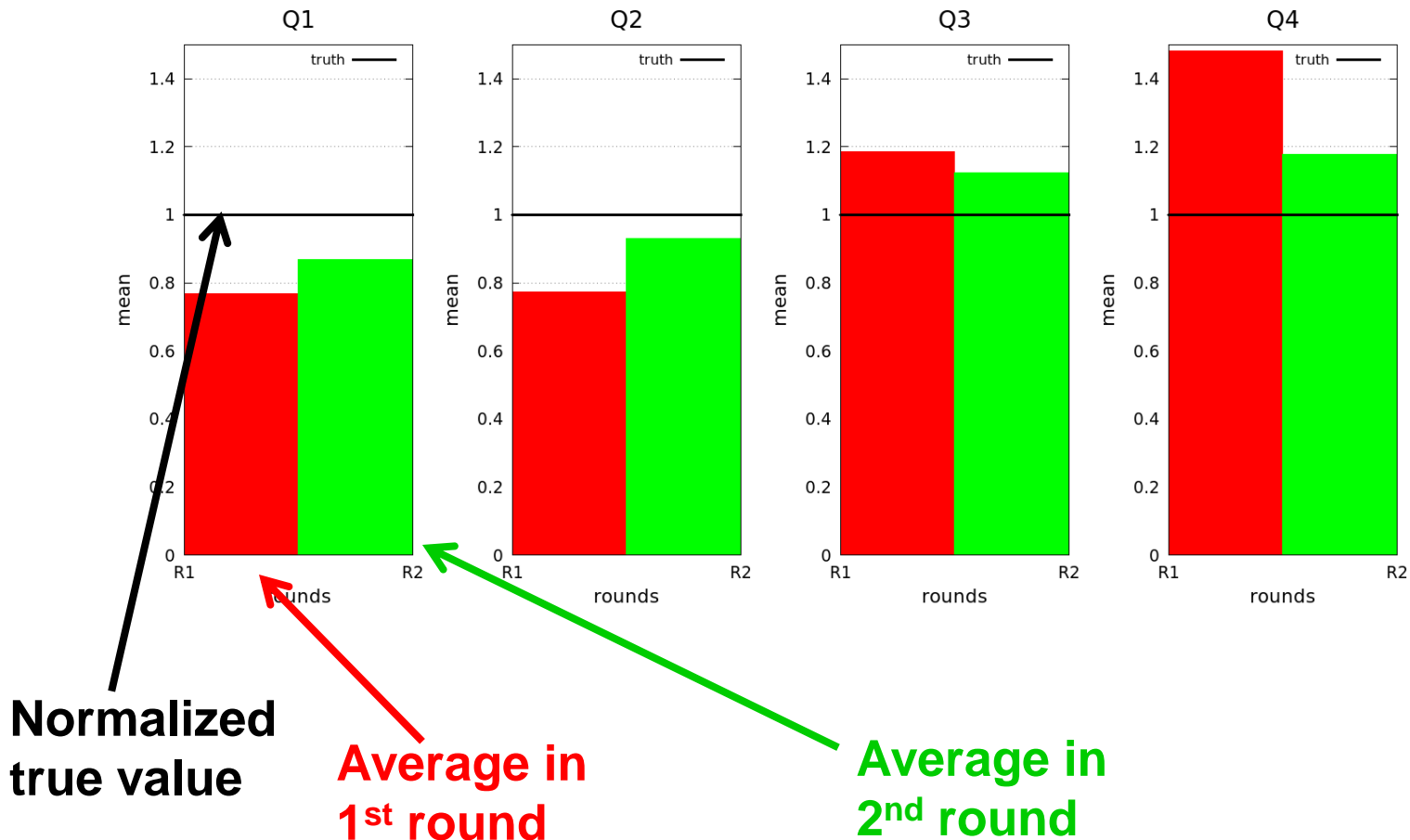
Network Diameter: 9

Communities: 15



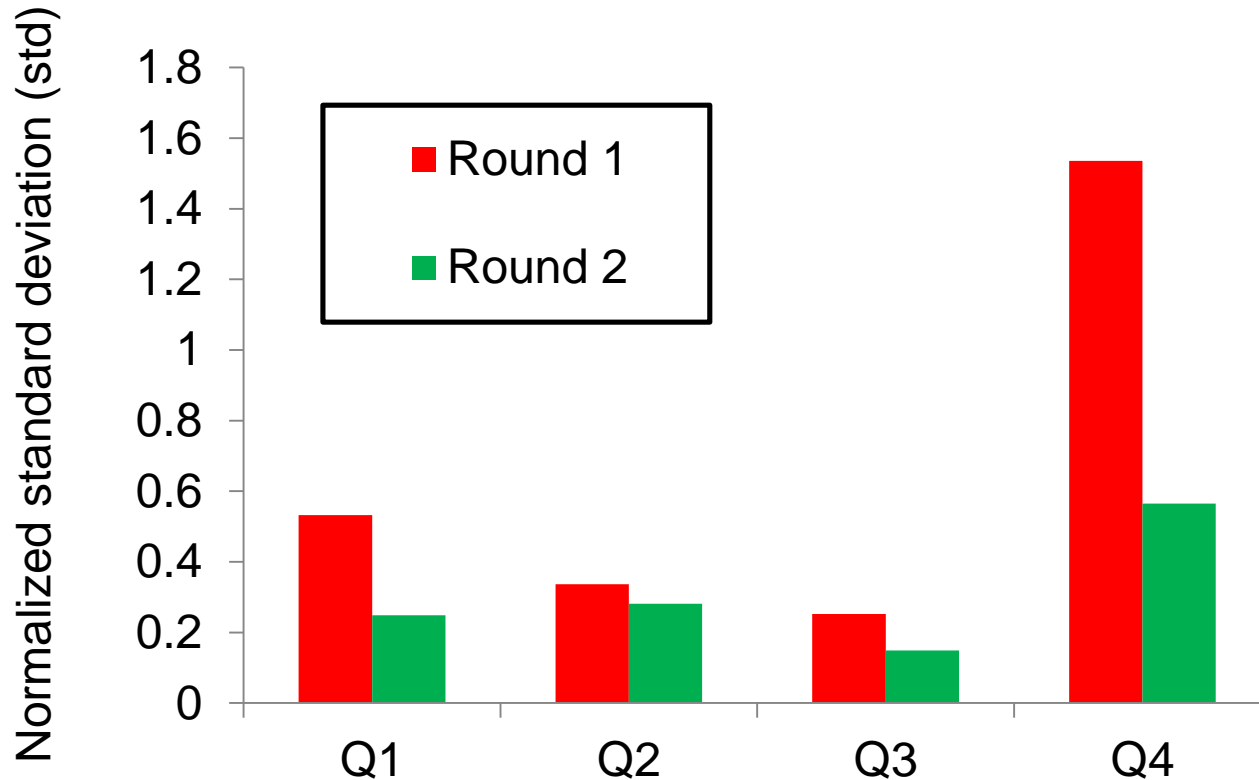
Main findings: average improves

Priverno fair (the others are similar):



Main findings: std decreases

Priverno fair (the others are similar):



Modeling user interactions

Having all these data we want to design **models** for opinion formation

Why?

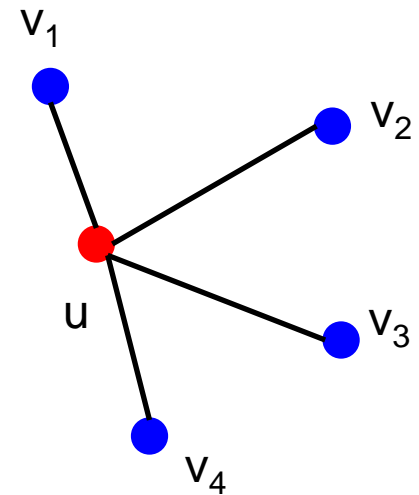
- Understand the opinion-formation process
- Understand effect of peer pressure
- Explain how interaction can lead to **improved** results

Hard: different people, lots of noise, missing info

Modeling user interactions

DeGroot model:

$$A'(u) = \frac{A(u) + A(v_1) + A(v_2) + A(v_3) + A(v_4)}{1 + 4}$$



Generalized DeGroot model:

$$A'(u) = \frac{\alpha A(u) + A(v_1) + A(v_2) + A(v_3) + A(v_4)}{\alpha + 4}$$

$A(u)$: answer of u at R1 $A'(u)$: answer of u at R2

But how can we explain the improvement?

Some reflection

- Peer interaction can lead to a more accurate crowd
- ... in contrast to previous studies in artificial settings where interaction was imposed
- How can we explain it?
- When does interaction improves and when does it harm?
- Models...

This talk

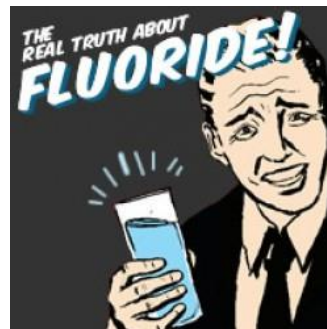
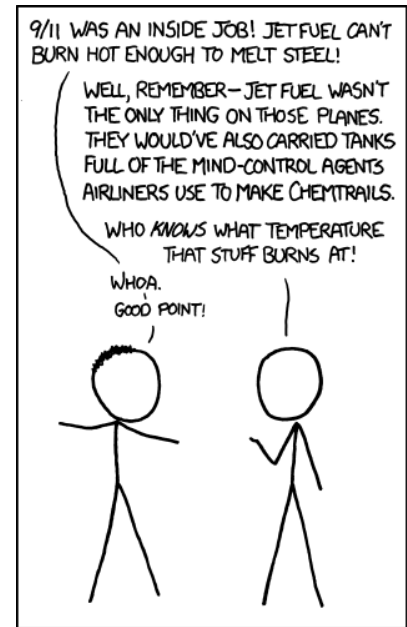
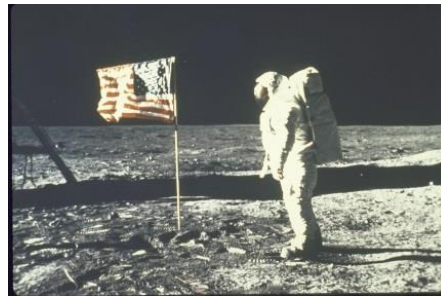
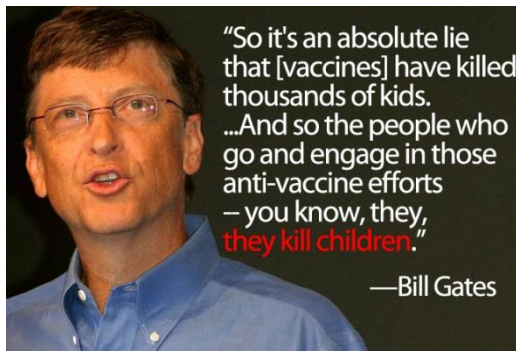
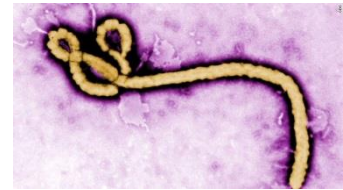
We will look at three dimensions of the problem:

- Network effect on the **wisdom of crowds**
- **The role of homophily and polarization in the spreading of (mis)information**
- How to schedule **experts** in crowdsourcing

Can we always trust the crowd?

Numerous examples where large part of the population believes false info:

- Does democracy always work?
- Conspiracy theories
- Unsubstantiated science (e.g., homeopathy)
- **How does such info become popular?**



Facebook study

Posts from 79 italian facebook group pages:

- 34 **science** group pages
- 65K posts
- 2.5M likes, 1.5M shares

- 39 **conspiracy** group pages
- 200K posts
- 6.5M likes, 16M shares

Crawled the network of likers and found their connections:

- 1.2M nodes
- 35M edges

A facebook post

180K likes



26K shares

User polarization

We have 1.2M users who have liked **science**/**conspiracy** posts.

Are they consistent with the content they like?

For each user u define user **polarization** $\rho(u)$:

$$\rho(u) = \frac{\textit{consp}}{\textit{consp} + \textit{sci}}$$

consp: # **conspiracy** posts u liked

sci: # **science** posts u liked

User polarization

We have 1.2M users who have liked **science**/**conspiracy** posts.

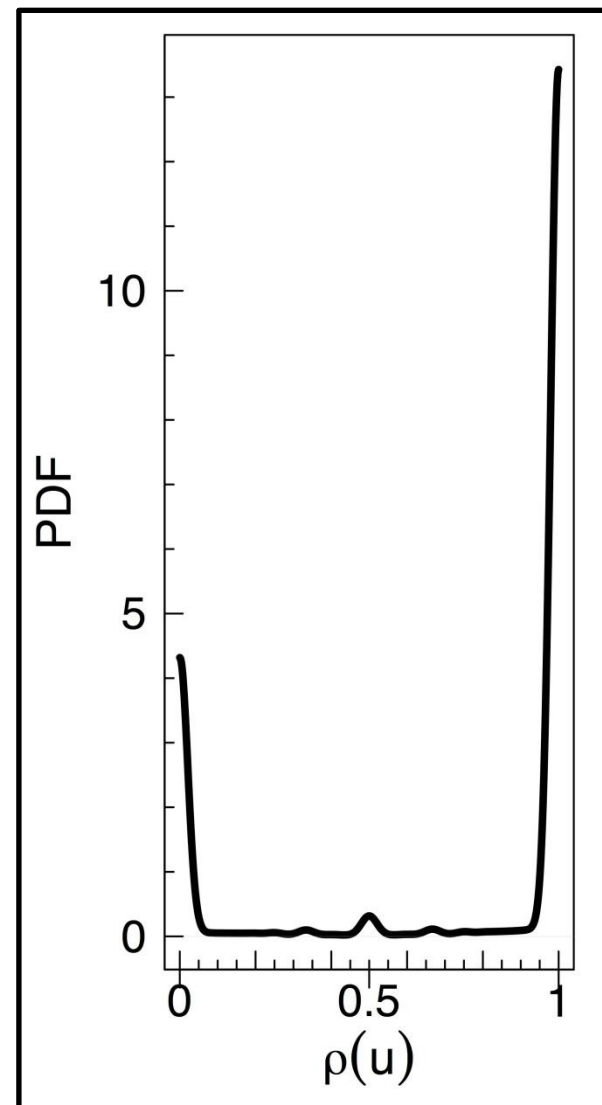
Are they consistent with the content they like?

For each user u define user **polarization** $\rho(u)$:

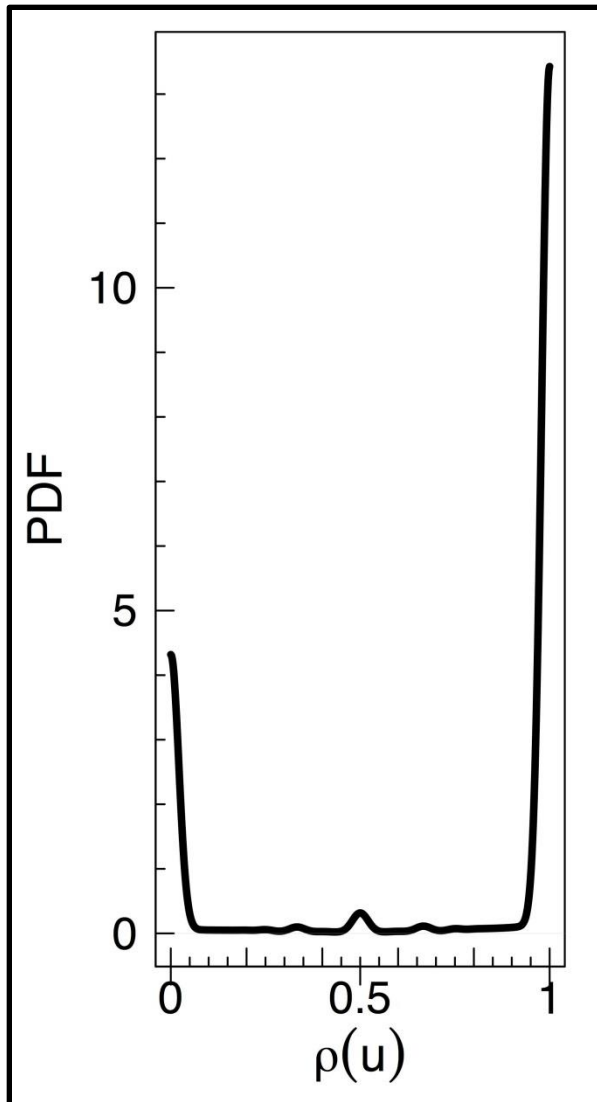
$$\rho(u) = \frac{\text{consp}}{\text{consp} + \text{sci}}$$

consp: # **conspiracy** posts u liked

sci: # **science** posts u liked



User polarization



We can select two subsets of users:

Science users: $\{u: \rho(u) \leq 5\%\}$

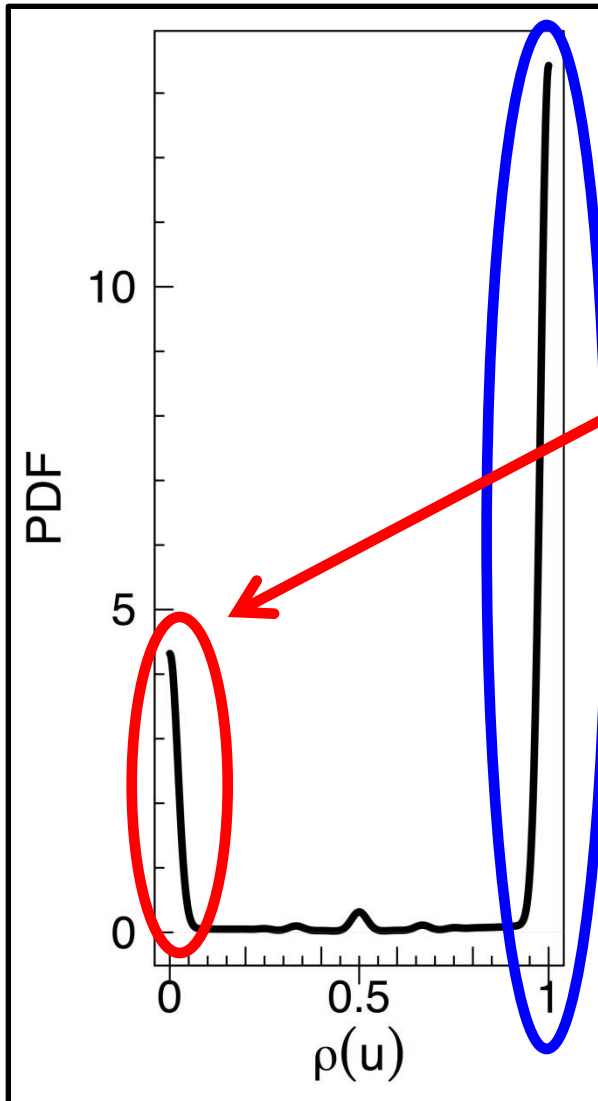
Conspiracy users: $\{u: \rho(u) \geq 95\%\}$

User polarization

We can select two subsets of users:

Science users: $\{u: \rho(u) \leq 5\%\}$

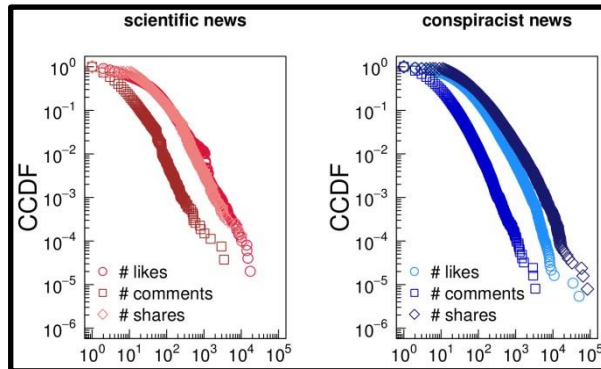
Conspiracy users: $\{u: \rho(u) \geq 95\%\}$



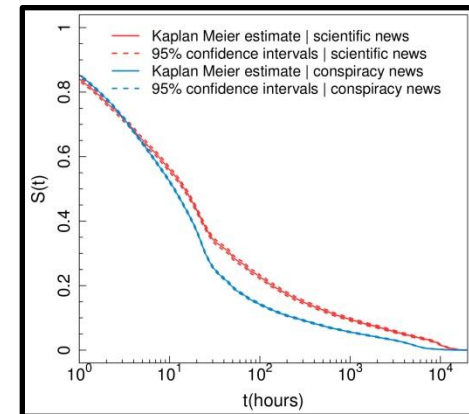
Science vs. conspiracy

Science and conspiracy posts and users show very similar behavior:

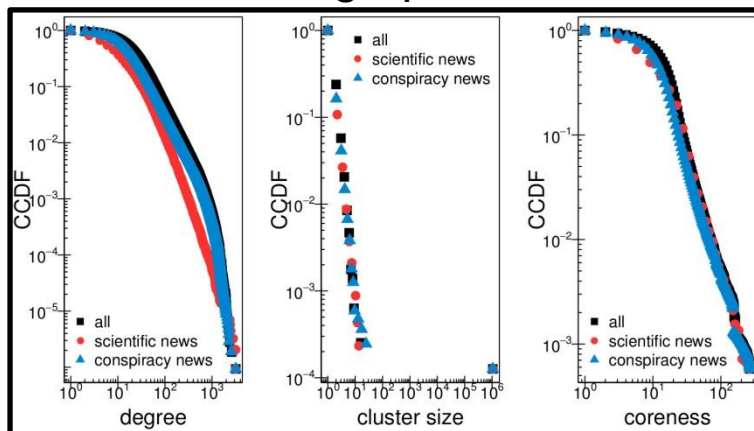
Post statistics



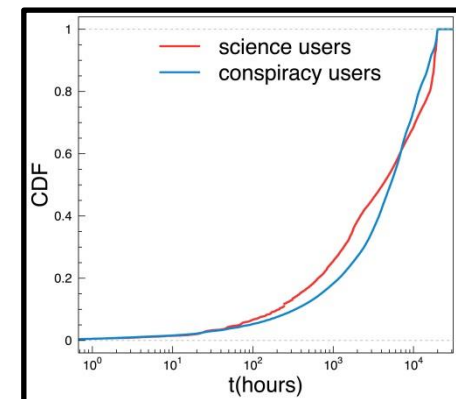
Post lifetime



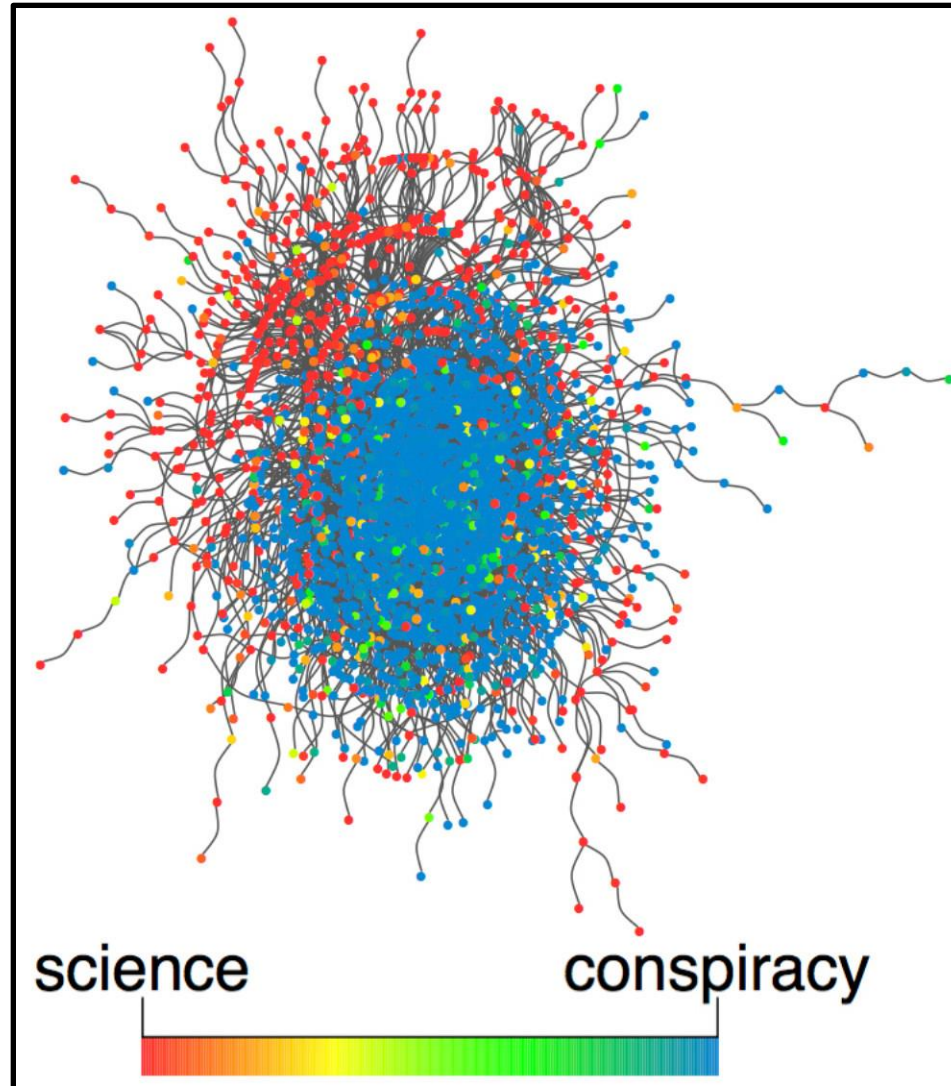
User subgraph statistics



User lifetime

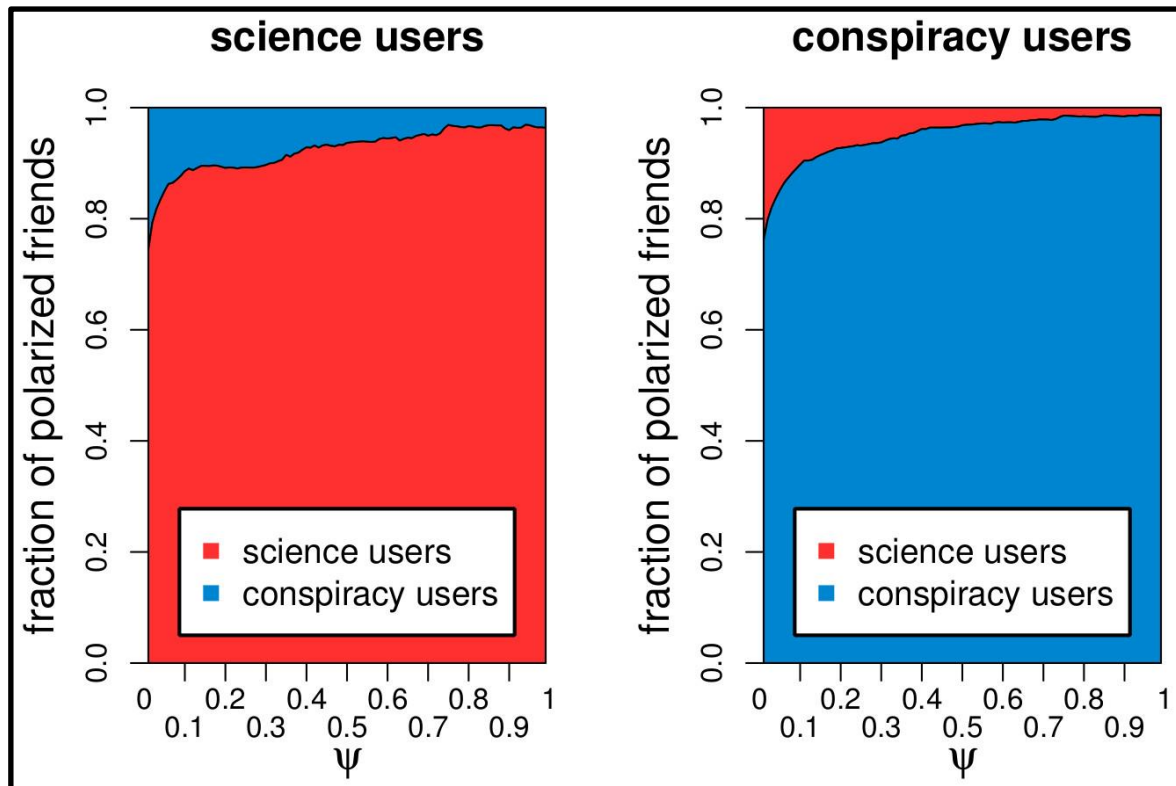


Largest connected component



Homophily

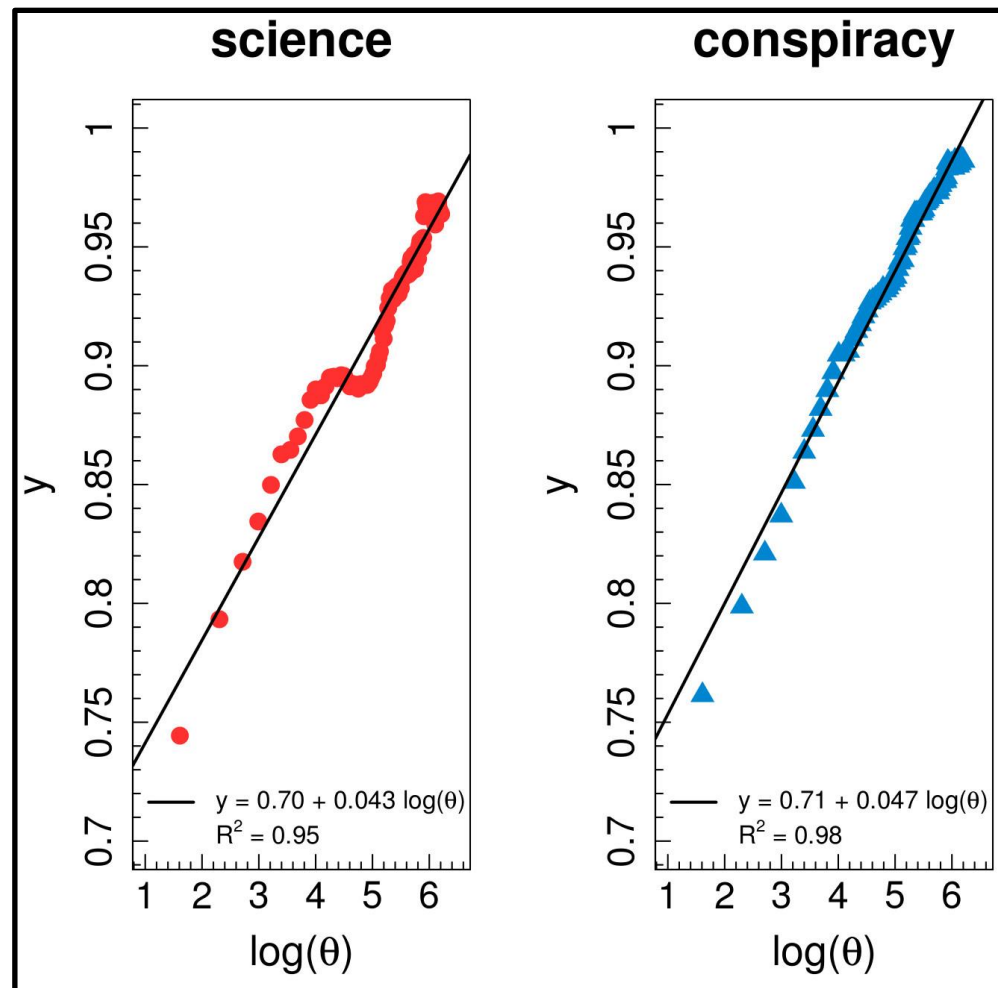
Homophily: tendency of individuals to associate with similar others



$\psi(u) \propto \# \text{ likes}$: Normalized liking activity of u

Prediction of polarized friends

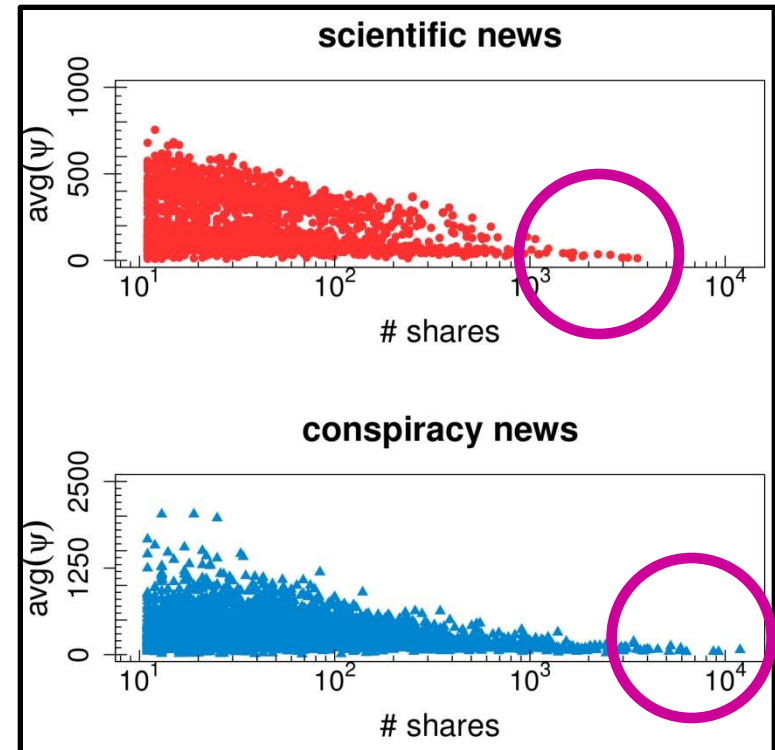
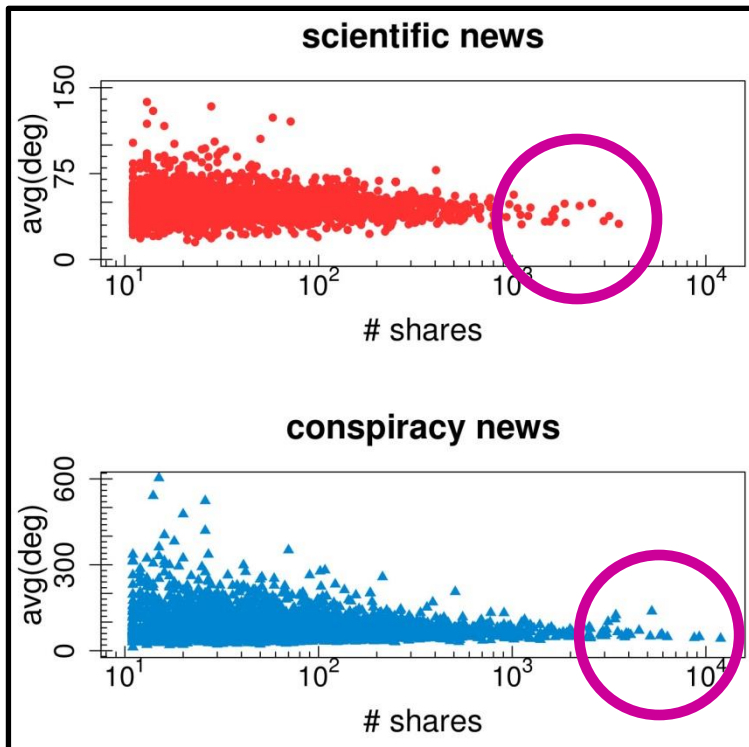
We can predict the ratio of u 's friends who have the same polarization with u as a function of u 's #likes:



$\theta(u) = \#likes$: Liking activity of u

How do posts become viral?

How does the average user of a **viral** post look?



$\text{deg}(u)$: # friends of node u

$\psi(u) \propto \# \text{ likes}$: Normalized liking activity of u

Troll posts

We also downloaded info about 4.7K **troll posts**: posts with clearly useless or wrong information:



“The Italian Senate voted and accepted (257 in favor and 165 abstentions) a law proposed by Senator Cirenga aimed at funding with 134 billion Euro the policy makers to find a job in case of defeat in the political competition.”

1.1K likes

36K shares

Troll posts

We also downloaded info about 4.7K **troll posts**: posts with clearly useless or wrong information:



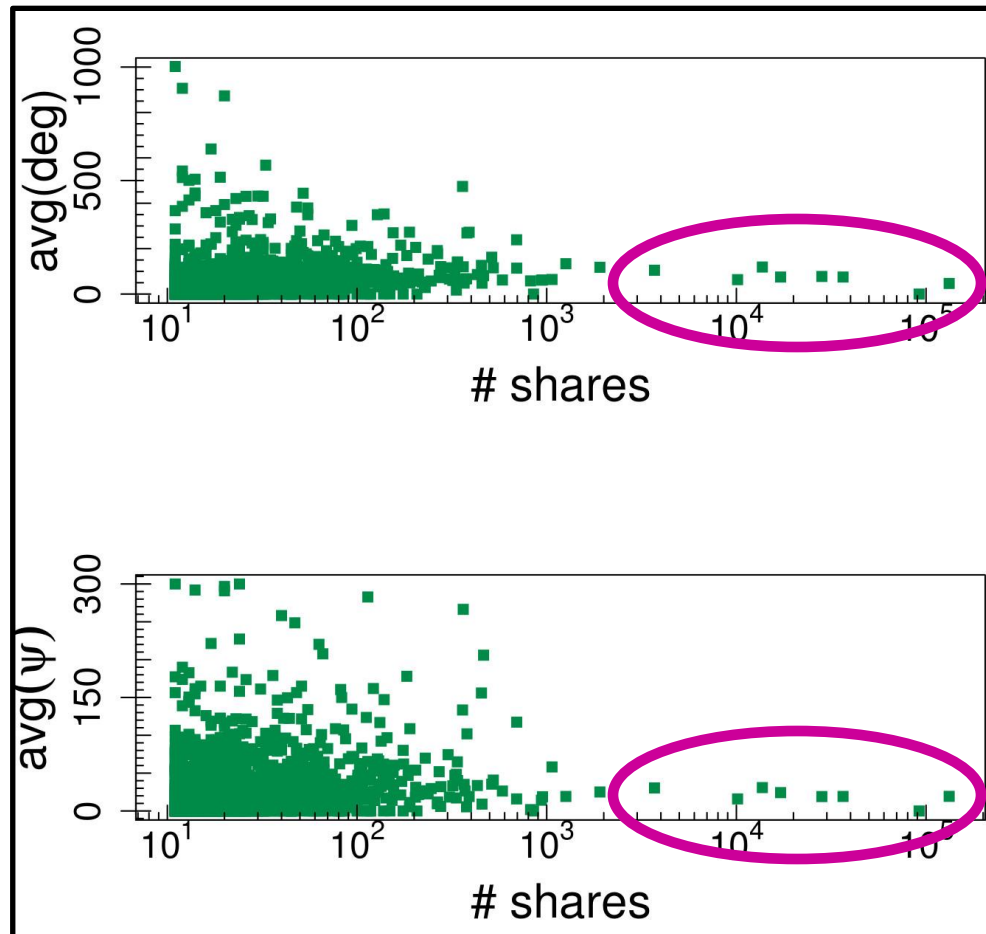
“The Italian Senate voted and accepted (257 in favor and 165 abstentions) a law proposed by Senator Cirenga aimed at funding with 134 billion Euro the policy makers to find a job in case of defeat in the political competition.”

1.1K likes

36K shares

- 315+5 members in Italian senate!
- Cirenga does not exist!
- 134B EUR > 1/20 of French GDP!

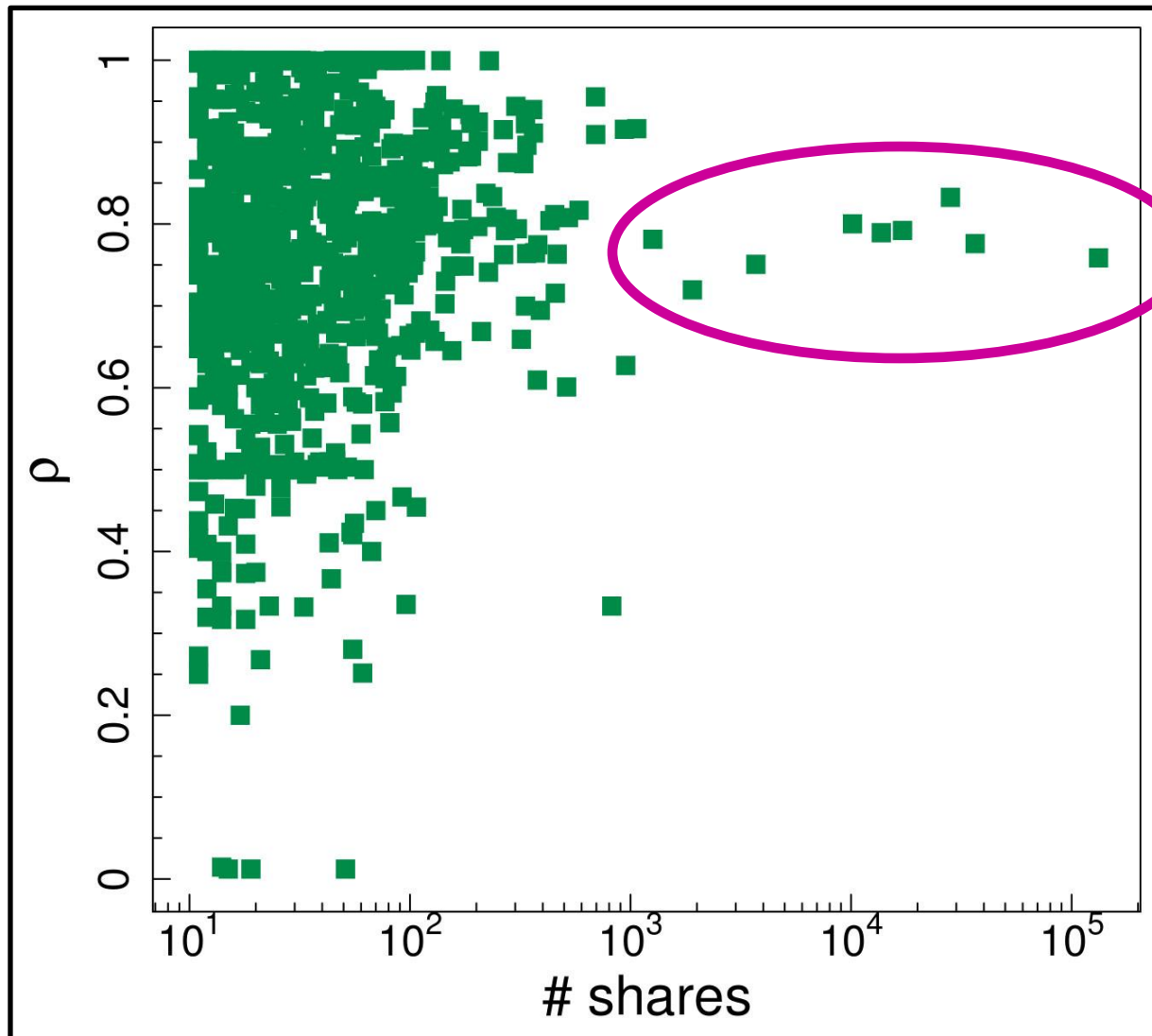
Troll posts: degree and activity



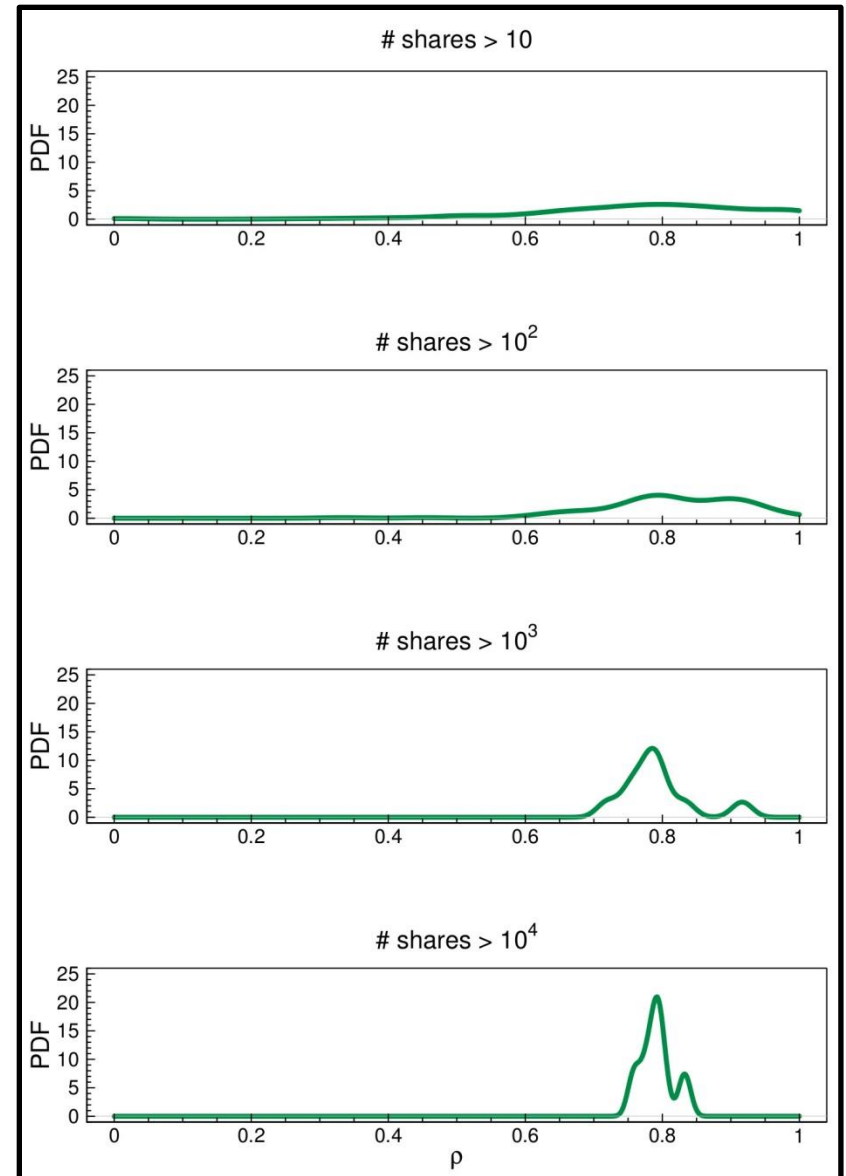
$\text{deg}(u)$: # friends of node u

$\psi(u) \propto \# \text{ likes}$: Normalized liking activity of u

Troll posts: polarization



Troll posts: polarization at different virality levels



Some reflection

- Peer influence can reinforce ones ideas
- ... to the extent that people might believe clearly false info
- Clear evidence of psychological phenomena such as
 - **Cognitive closure:** the human desire to eliminate ambiguity and arrive at definite conclusions (sometimes irrationally)
 - **Confirmation bias:** tendency to search for, believe, and remember info in a way that is aligned with ones beliefs

This talk

We will look at three dimensions of the problem:

- Network effect on the **wisdom of crowds**
- The role of homophily and polarization in the spreading of (mis)information
- **How to schedule experts in crowdsourcing**

Rest of the talk

Wisdom of **crowds** and wisdom of **experts**:

- We saw that in some cases the crowd cannot be trusted
- For some problems experts are indispensable!
- But experts are scarce and expensive
- What can we do with **(lots)** of nonexperts?

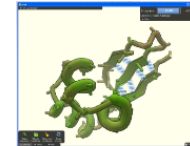
Online collaboration systems

Systems creating knowledge by massive online collaboration:

- Tagging/geotagging systems:



- Games with a purpose:



- Content creation systems:



- Crowdsourcing:



- Open source community:



- Polymath project:



Online collaboration systems

Systems creating knowledge by massive online collaboration:

- Tagging/geotagging systems:



- Games with a purpose:



- Content creation systems:



- Crowdsourcing:



- Open source community:



- Polymath project:



What is crowdsourcing

Crowdsourcing: is the process of obtaining information by using contributions from a large group of people.

There are tasks hard for computers but easy for humans (**human tasks**):

- Compare 2 photos (to select the best one that represents the Colosseum)
- Translate a sentence
- Choose the best search result to a query
- ...




Crowdsourcing platforms: Online services that allow, through APIs, to get answers from humans at a low cost

- Amazon Mechanical Turk
- CrowdFlower



Crowdsourcing – Amazon Mechanical Turk



[Your Account](#)[HITs](#)[Qualifications](#)

[Introduction](#) | [Dashboard](#) | [Status](#) | [Account Settings](#)

Mechanical Turk is a marketplace for work.

We give businesses and developers access to an on-demand, scalable workforce. Workers select from thousands of tasks and work whenever it's convenient.

311,326 HITs available. [View them now.](#)

Make Money by working on HITs

HITs - *Human Intelligence Tasks* - are individual tasks that you work on. [Find HITs now.](#)


As a Mechanical Turk Worker you:

- Can work from home
- Choose your own work hours
- Get paid for doing good work

Find an interesting task

Work

Earn money



[Find HITs Now](#)

or [learn more about being a Worker](#)

Get Results from Mechanical Turk Workers

Ask workers to complete HITs - *Human Intelligence Tasks* - and get results using Mechanical Turk. [Register Now](#)


As a Mechanical Turk Requester you:

- Have access to a global, on-demand, 24 x 7 workforce
- Get thousands of HITs completed in minutes
- Pay only when you're satisfied with the results

Fund your account

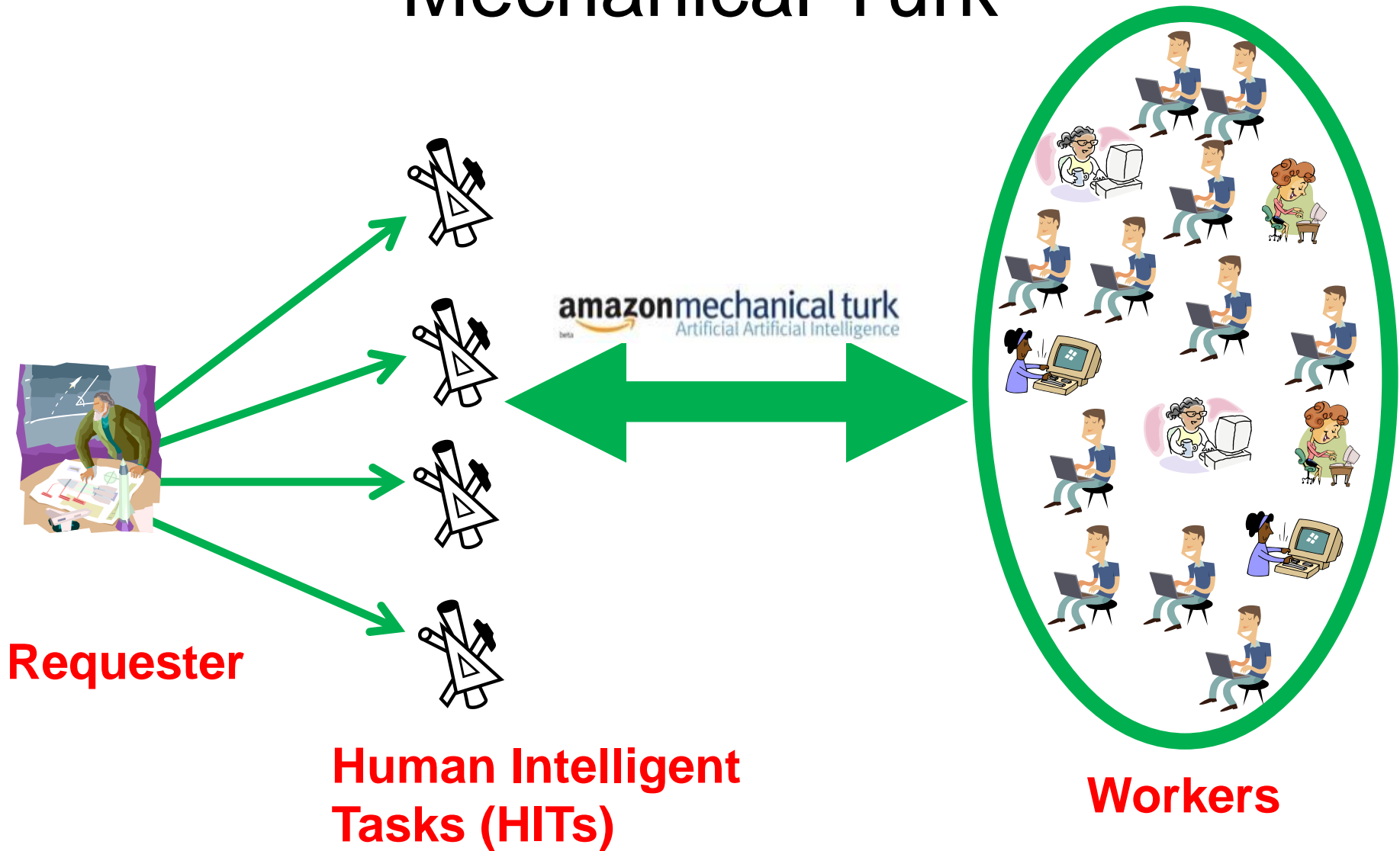
Load your tasks

Get results

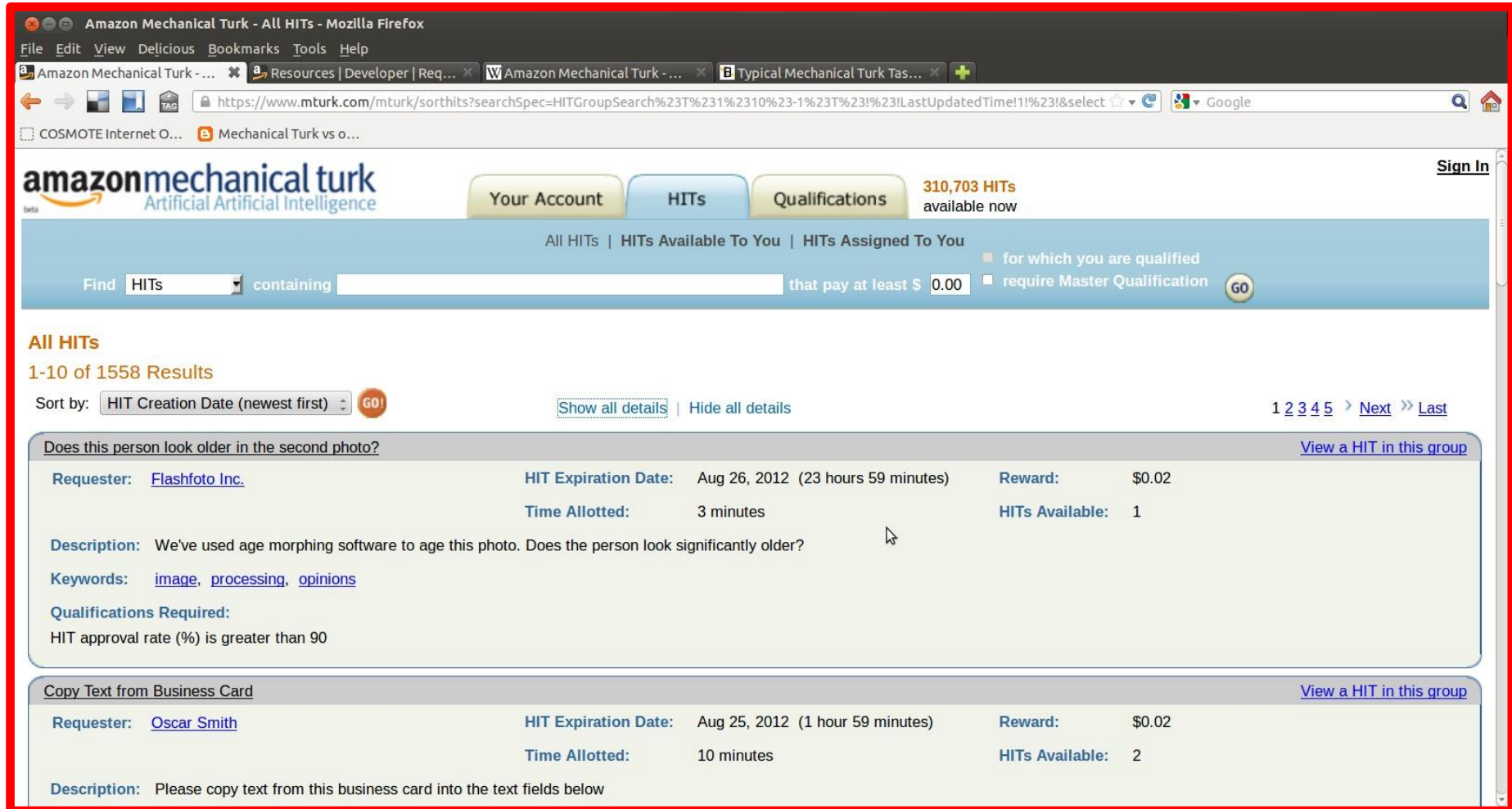


[Get Started](#)

Crowdsourcing – Amazon Mechanical Turk



Crowdsourcing – Amazon Mechanical Turk



The screenshot shows the Amazon Mechanical Turk interface in a Mozilla Firefox browser. The page displays search results for HITs (Human Intelligence Tasks). The top navigation bar includes links for 'Your Account', 'HITS', and 'Qualifications', along with a status '310,703 HITS available now'. A search bar shows 'Find HITS containing' followed by a search filter 'that pay at least \$ 0.00'. Below the search bar, the results are sorted by 'HIT Creation Date (newest first)'. Two HITs are visible:

HIT Title	Requester	HIT Expiration Date	Time Allotted	Reward	HITs Available
Does this person look older in the second photo?	Flashfoto Inc.	Aug 26, 2012 (23 hours 59 minutes)	3 minutes	\$0.02	1
Copy Text from Business Card	Oscar Smith	Aug 25, 2012 (1 hour 59 minutes)	10 minutes	\$0.02	2

Each HIT entry includes a description, keywords, and qualifications required. The first HIT description is: 'We've used age morphing software to age this photo. Does the person look significantly older?'. The second HIT description is: 'Please copy text from this business card into the text fields below'.

Crowdsourcing – Amazon Mechanical Turk

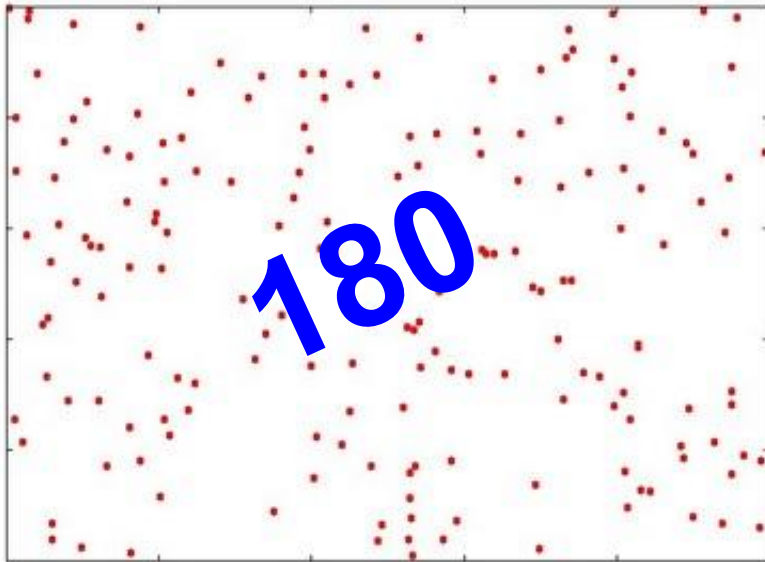
The screenshot shows the Amazon Mechanical Turk website in a Mozilla Firefox browser. The page displays a list of HITs (Human Intelligence Tasks). The first HIT is highlighted with a pink border and contains the following details:

- Requester:** [Flashfoto Inc.](#)
- HIT Expiration Date:** Aug 26, 2012 (23 hours 59 minutes)
- Reward:** \$0.02
- Time Allotted:** 3 minutes
- HITs Available:** 1000
- Description:** We've used age morphing software to age this photo. Does the person look significantly older?
- Keywords:** [image](#), [processing](#), [opinions](#)
- Qualifications Required:** HIT approval rate (%) is greater than 90

Below this HIT, there is another HIT for copying text from a business card:

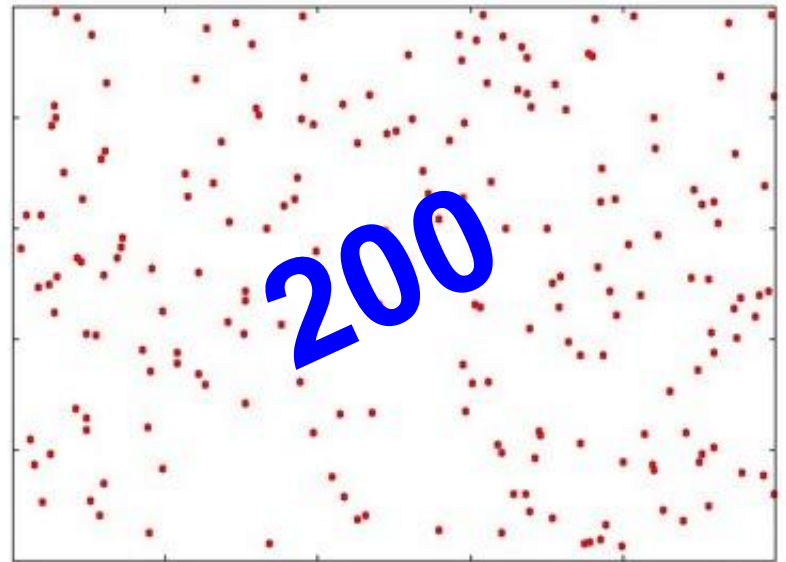
- Requester:** [Oscar Smith](#)
- HIT Expiration Date:** Aug 25, 2012 (1 hour 59 minutes)
- Reward:** \$0.02
- Time Allotted:** 10 minutes
- HITs Available:** 2
- Description:** Please copy text from this business card into the text fields below

Which photo has more dots?



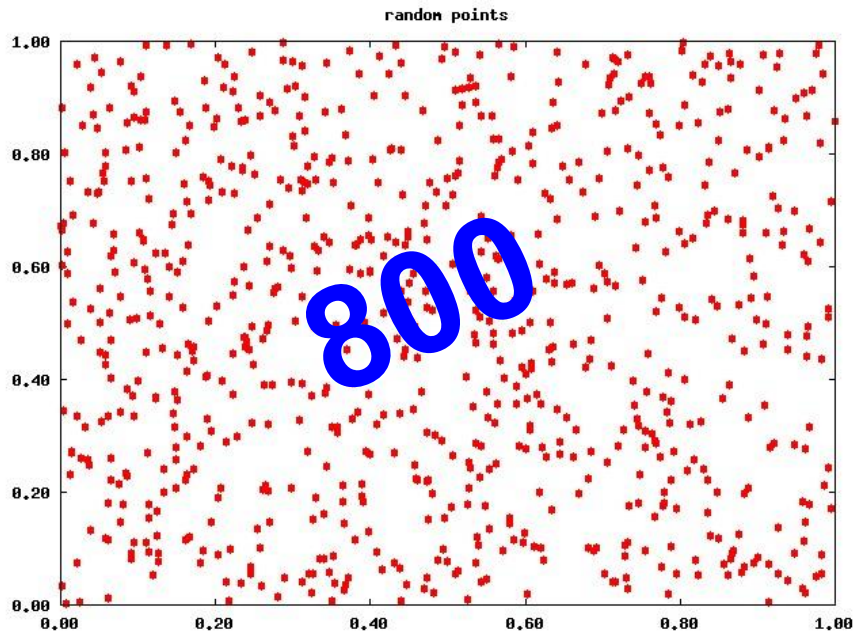
FIRST PICTURE

vs.



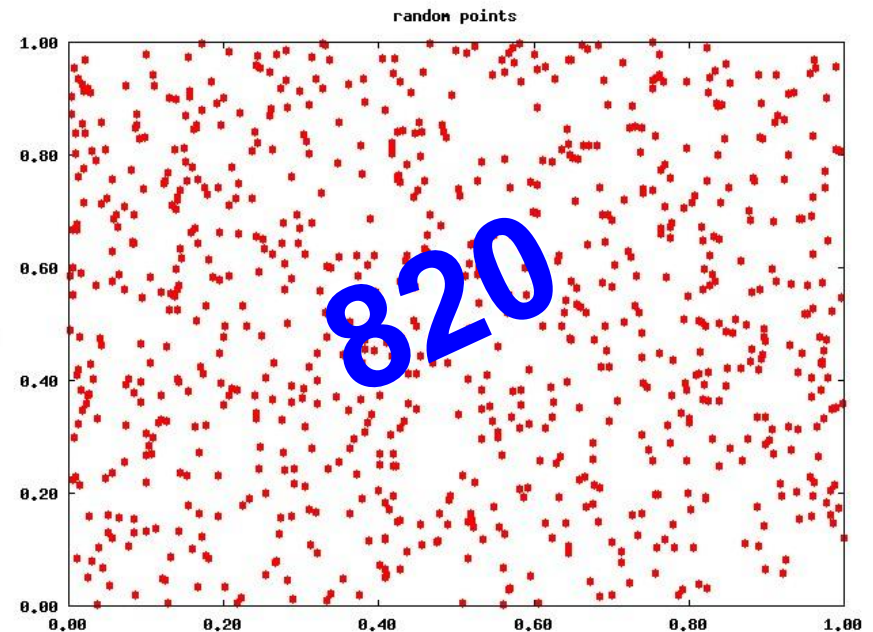
SECOND PICTURE

Which photo has more dots?



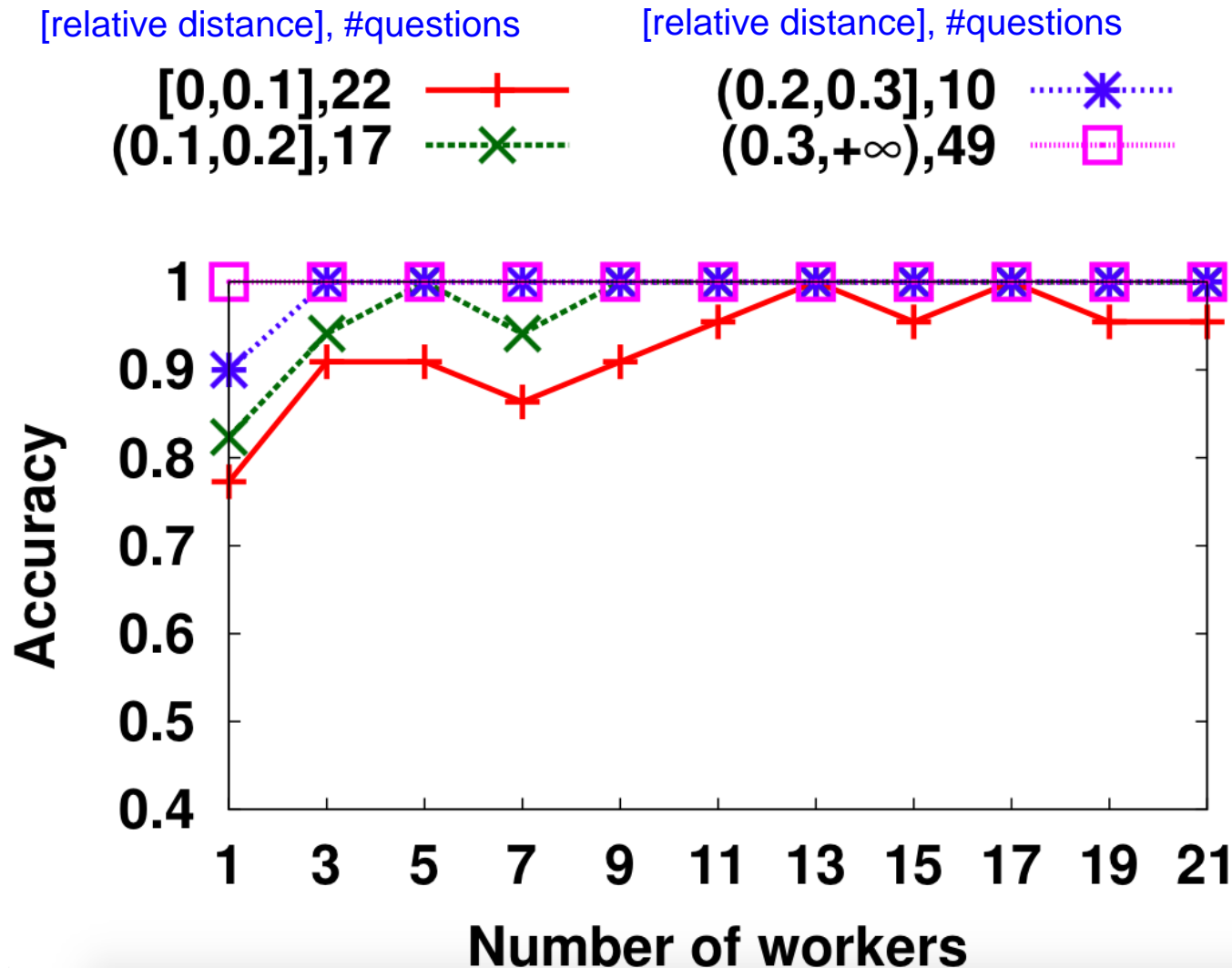
FIRST PICTURE

vs.



SECOND PICTURE

Accuracy vs. number of responses



Which car is more expensive?



FIRST CAR

- 2013 BMW 650 i xDrive -
body style: Coupe
doors: 2
engine: 4.4L V8 32V GDI DOHC
Twin Turbo

VS.



SECOND CAR

- 2013 Mercedes-Benz SL550 -
body style: Convertible
doors: 2
engine: 4.6L V8 32V GDI DOHC
Twin Turbo

Which car is more expensive?



vs.



FIRST CAR

- 2013 BMW 650 i xDrive -
body style: Coupe
doors: 2
engine: 4.4L V8 32V GDI DOHC
Twin Turbo

SECOND CAR

- 2013 Mercedes-Benz SL550 -
body style: Convertible
doors: 2
engine: 4.6L V8 32V GDI DOHC
Twin Turbo

Which car is more expensive?



VS.



FIRST CAR

- 2013 Mercedes-Benz SL550 -
body style: Convertible
doors: 2
engine: 4.6L V8 32V GDI DOHC
Twin Turbo

SECOND CAR

- 2013 Audi S8 4.0T quattro -
body style: Sedan
doors: 4
engine: 4.0L V8 32V GDI DOHC
Twin Turbo

Which car is more expensive?



VS.



FIRST CAR

- 2013 Mercedes-Benz SL550 -
body style: Convertible
doors: 2
engine: 4.6L V8 32V GDI DOHC
Twin Turbo

SECOND CAR

- 2013 Audi S8 4.0T quattro -
body style: Sedan
doors: 4
engine: 4.0L V8 32V GDI DOHC
Twin Turbo

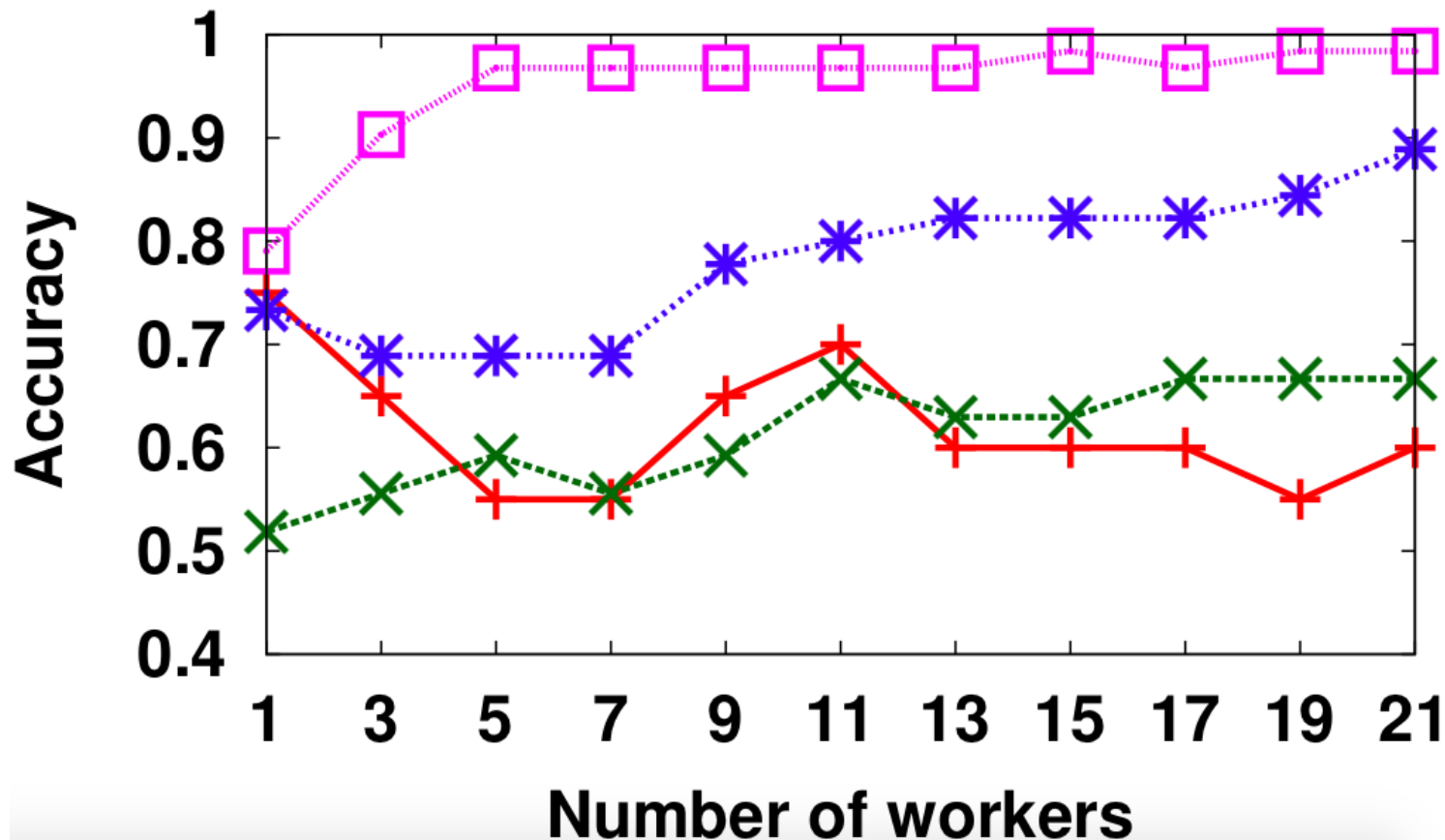
Accuracy vs. number of responses

[relative distance], #questions

[0,0.1],20
(0.1,0.2],27

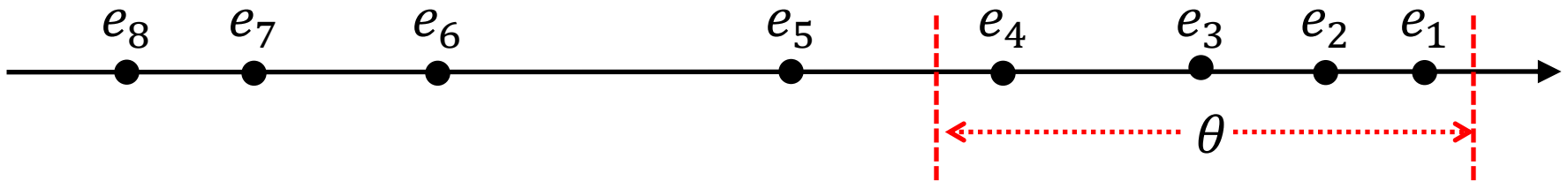
[relative distance], #questions

(0.2,0.5],45
(0.5,+ ∞),62



Modeling the error

- Consider a set of elements with different values



- Threshold error model:**
- We present to a worker a pair (e_i, e_j)
 - If $|e_i - e_j| \geq \theta$ worker returns **correct** answer
 - If $|e_i - e_j| < \theta$ worker returns **arbitrary** answer

Note that if the difference is $< \theta$ no matter how many workers we ask, we cannot obtain a more accurate response

Using expert workers

Usually workers are untrained

An **expert** is a more capable worker:

- May have been trained
- More scarce
- More expensive

Experts have started being offered by crowdsourcing systems

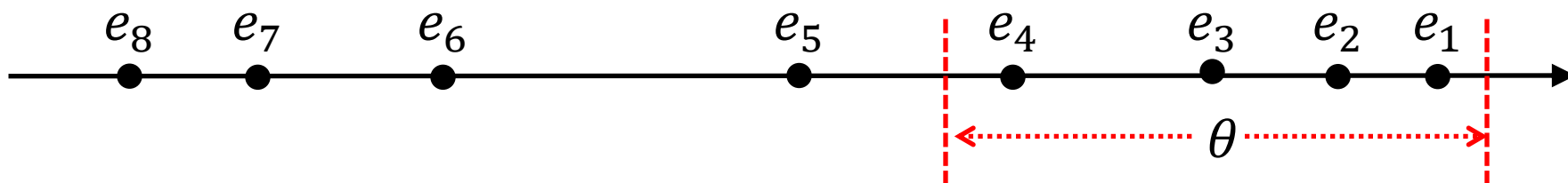
- “Masters,” “skilled,” ...

When should we use regular workers and when experts?

Think of ‘Who wants to be a millionaire’

Modeling the error

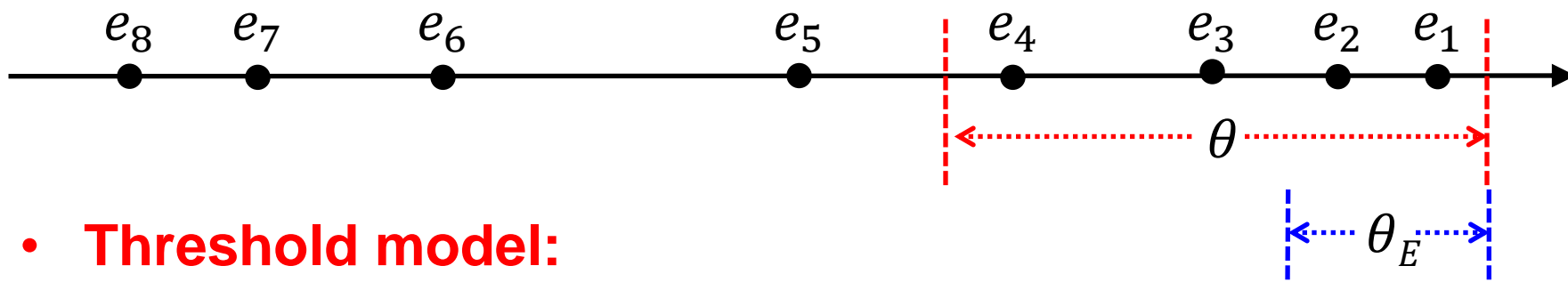
- Consider a set of elements with different values



- Threshold model:**
- We present to a worker a pair (e_i, e_j)
 - If $|e_i - e_j| \geq \theta$ worker returns **correct** answer
 - If $|e_i - e_j| < \theta$ worker returns **arbitrary** answer

Modeling the error

- Consider a set of elements with different values



- Threshold model:**

- We present to a worker a pair (e_i, e_j)
 - If $|e_i - e_j| \geq \theta$ worker returns **correct** answer
 - If $|e_i - e_j| < \theta$ worker returns **arbitrary** answer

Experts have a lower error threshold θ_E

Simple task: compute the MAX

A model allows us to formalize and analyze the problem

- We provide an algorithm that finds an element as close to the max as possible
- We prove that it makes as few expert comparisons as possible

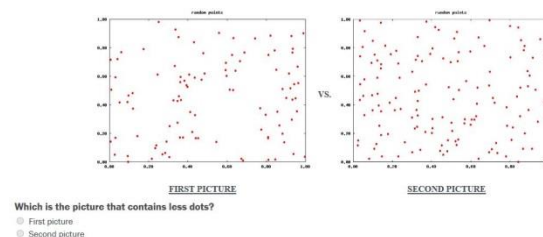
Feel free to ask for details after the talk.

Experiments using the Crowd

Tested on Crowdsourcing platform
with 3 datasets:  **CrowdFlower**

1. $n = 50$ pictures with **DOTS**

Goal: find more dots



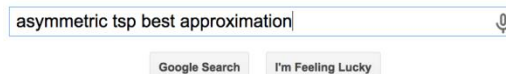
2. $n = 50$ **CARS**

Goal: find most expensive



3. $n = 50$ **QUERY RESULTS**

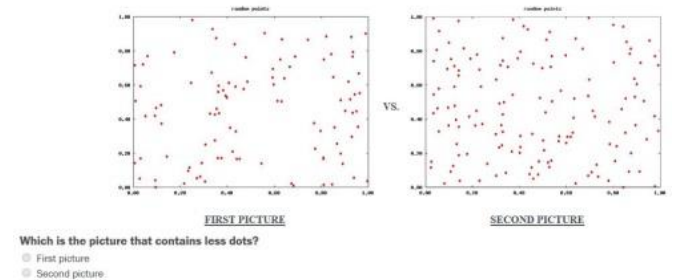
Goal: find most relevant result
for a given query



Results

In all our 3 sets of experiments:

The combination of nonexpert and expert users finds the best results with a low cost.



FIRST CAR

- 2013 BMW 650i xDrive -
body style: Coupe
doors: 2
engine: 4.4L V8 32V GDI DOHC
Twin Turbo

VS.



SECOND CAR

- 2013 Mercedes-Benz SL550 -
body style: Convertible
doors: 2
engine: 4.6L V8 32V GDI DOHC
Twin Turbo



asymmetric tsp best approximation



Google Search

I'm Feeling Lucky

Future directions

Understand better when we have wisdom or ignorance of the crowds

- Experiments in more controlled environments
- Large-scale experiments (twitter)
- Models
- Algorithms
- More detailed analysis of misinformation

Thanks!

Questions, comments, etc.: <http://aris.me>

