

Humans are “Humans”: Understanding and Modeling Humans

Chien-Ju (CJ) Ho

Logistics: Assignment 2

- Due: October 4 (Friday)
- Some tricky details in implementing EM:
 - How to choose the weights during the step of weighted majority voting?
 - The MLE $\log(p/(1-p))$ weight is very sensitive to noise when p approaches 0 or 1
 - You can do either one of the following:
 - Crop the probability using ideas in Section 4 of “Who Moderates the Moderators”
 - Use the $(2p-1)$ weight for workers with correct probability p
 - technically not leading to MLE, but it give the smallest theoretical error bounds (see lecture 3)

Logistics: Project Proposal

- Due this Friday.
- If you want to chat this week, you can schedule a meeting:
 - <https://calendly.com/chienjuho/30min>
- I'll also reserve time after the proposal deadline.

Crowdsourcing (Requester's Perspective)



Input



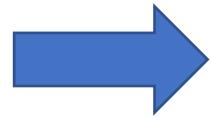
Black Box



Flower
Dog
Cute
...

Output

The Crowd is Made of People



Flower
Dog
Cute
...

Who are these people?

Why should we care?

- So we know how to better utilize and interact with the crowd
 - CS tools let us reason about programs run on machines (runtime, scalability, correctness, ...)
 - Need to develop **models of human behavior** when humans are in the loop
 - Most studies so far make strong assumptions about human behavior
- Remind us crowd workers are humans like us (sounds obvious, but...)
 - Human-centered research has been receiving great attention recently
 - Ethical-related issues (fairness, transparency, and privacy) are important
- **Relatively under-explored.** Potential interesting topics for your projects...
 - They might be quite challenging though

Who are these workers?

- Demographic information
- Are workers working independently?
- How many workers are there?

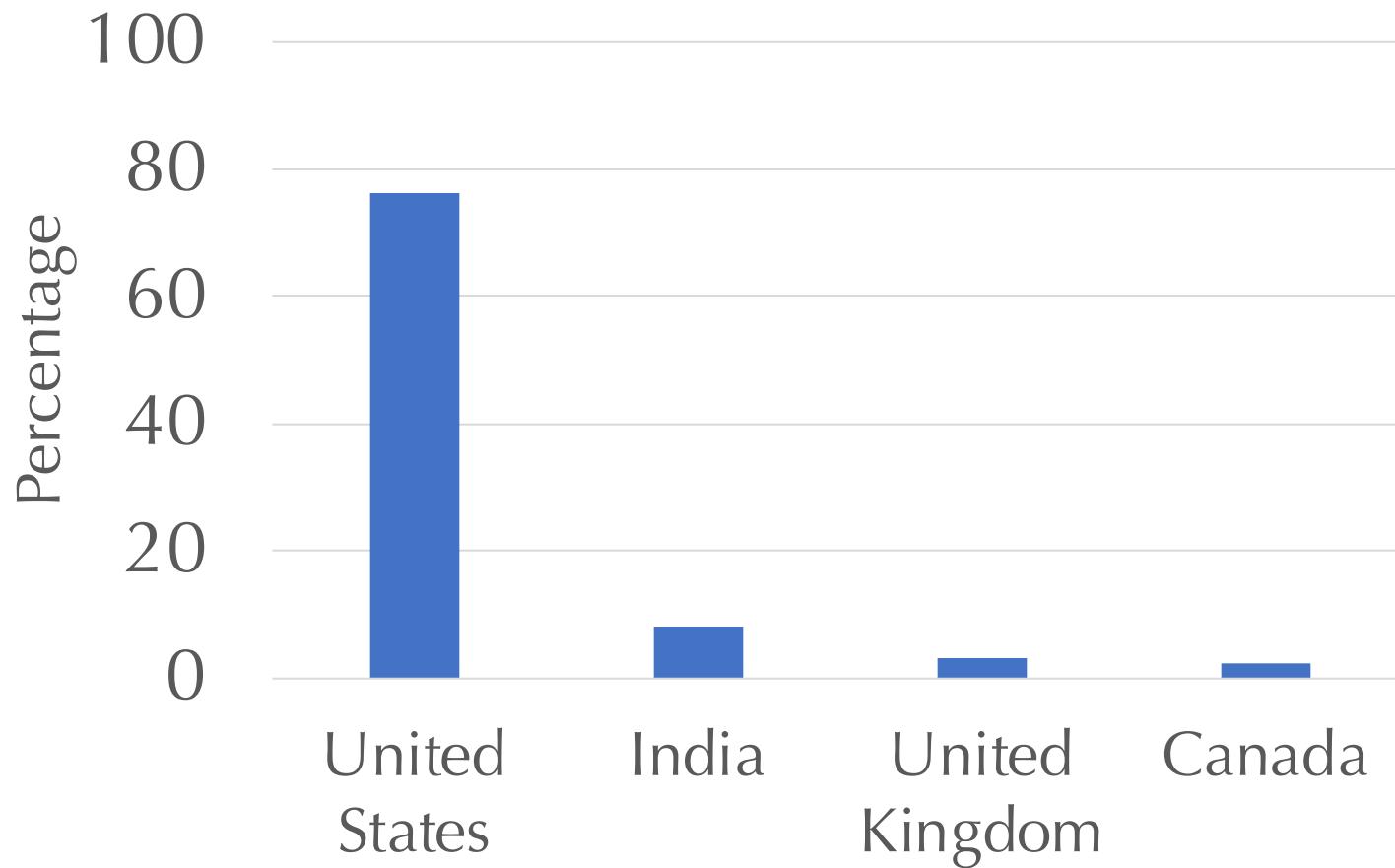
The information is not trivial to get

- There is no official census data about crowd workers
- Potential solutions:
 - Conducting surveys
 - Ethnographic analysis of digital trace
 - What else?

The Demographic of MTurk

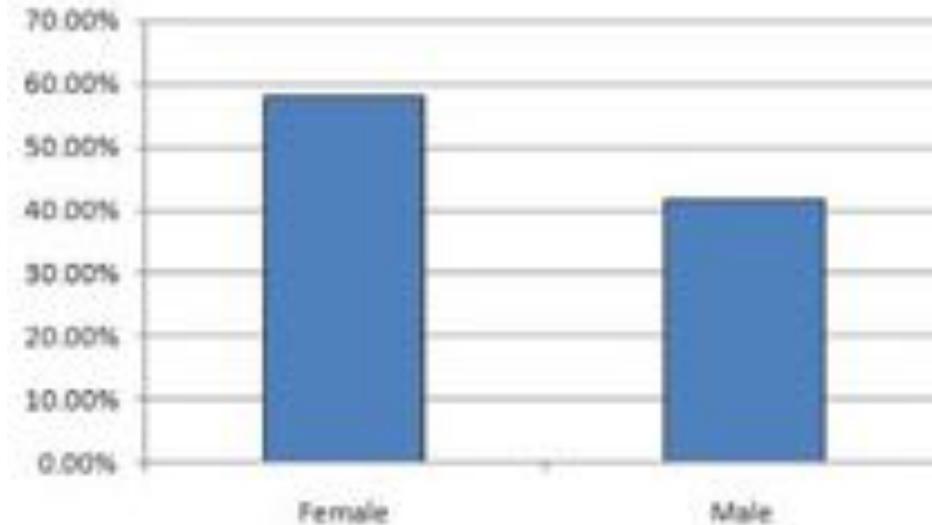
- Survey on 1,000 Turkers
 - Conduct the survey twice (Dec. 2008 and Oct. 2008)
 - Consistent statistics
 - Blog Post:
 - [A Computer Scientist in a Business School](#)
- One of the early attempts to understand MTurk demographics

MTurk Worker Demographics (2008): Country

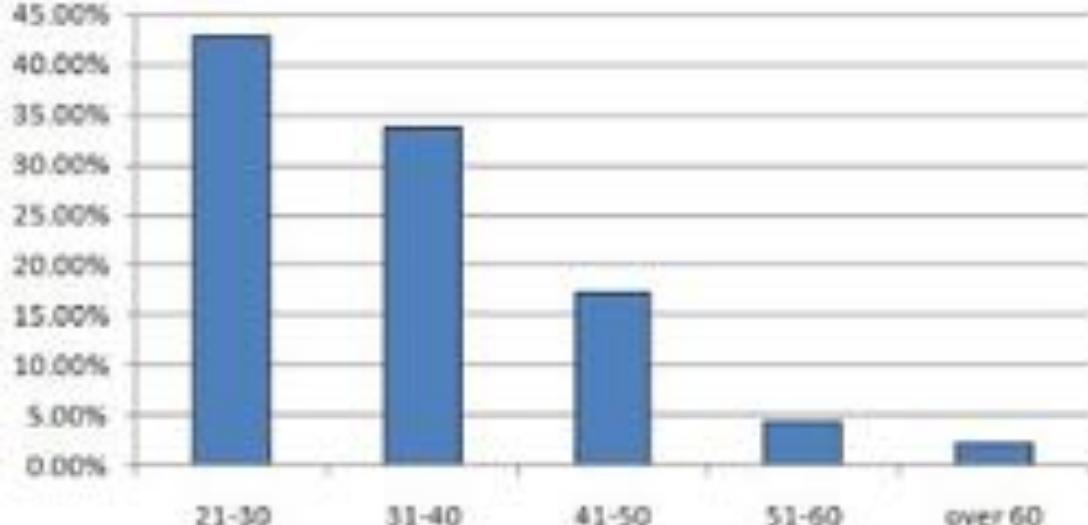


The majority of crowd workers on MTurk come from US

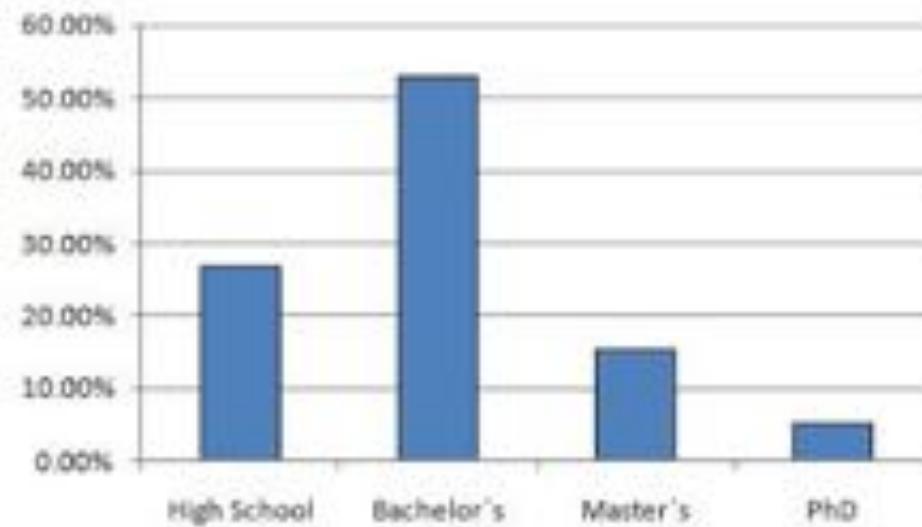
Gender



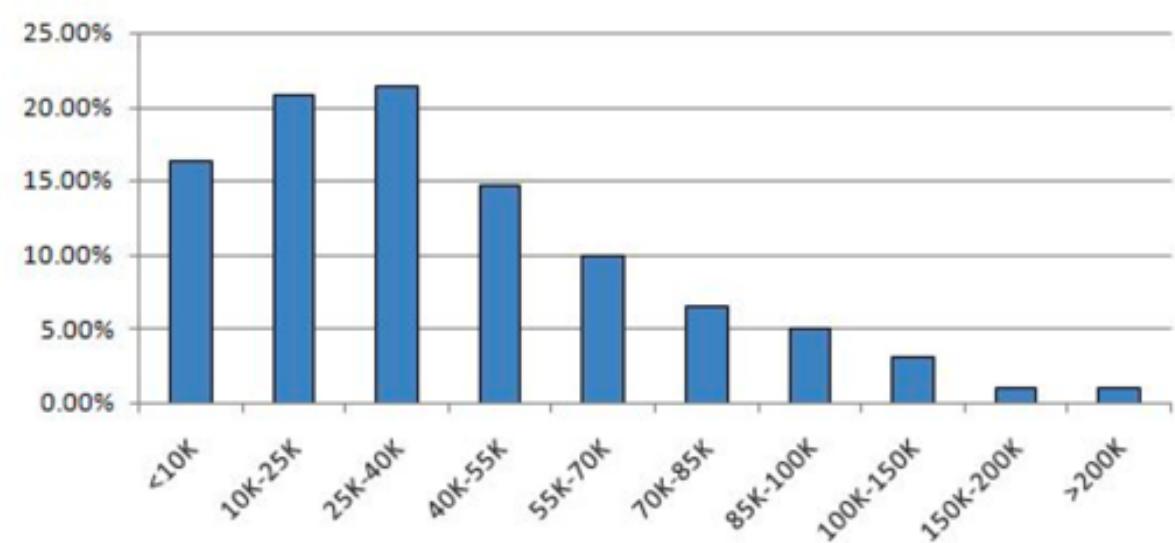
Age



Education



Income



MTurk Worker



Younger 54%

More Female 70%

Lower Income 65%

Smaller Family 45%

Internet Users (2008, US)



Age: 21-35 22%

Female 50%

Household income < 60K 45%

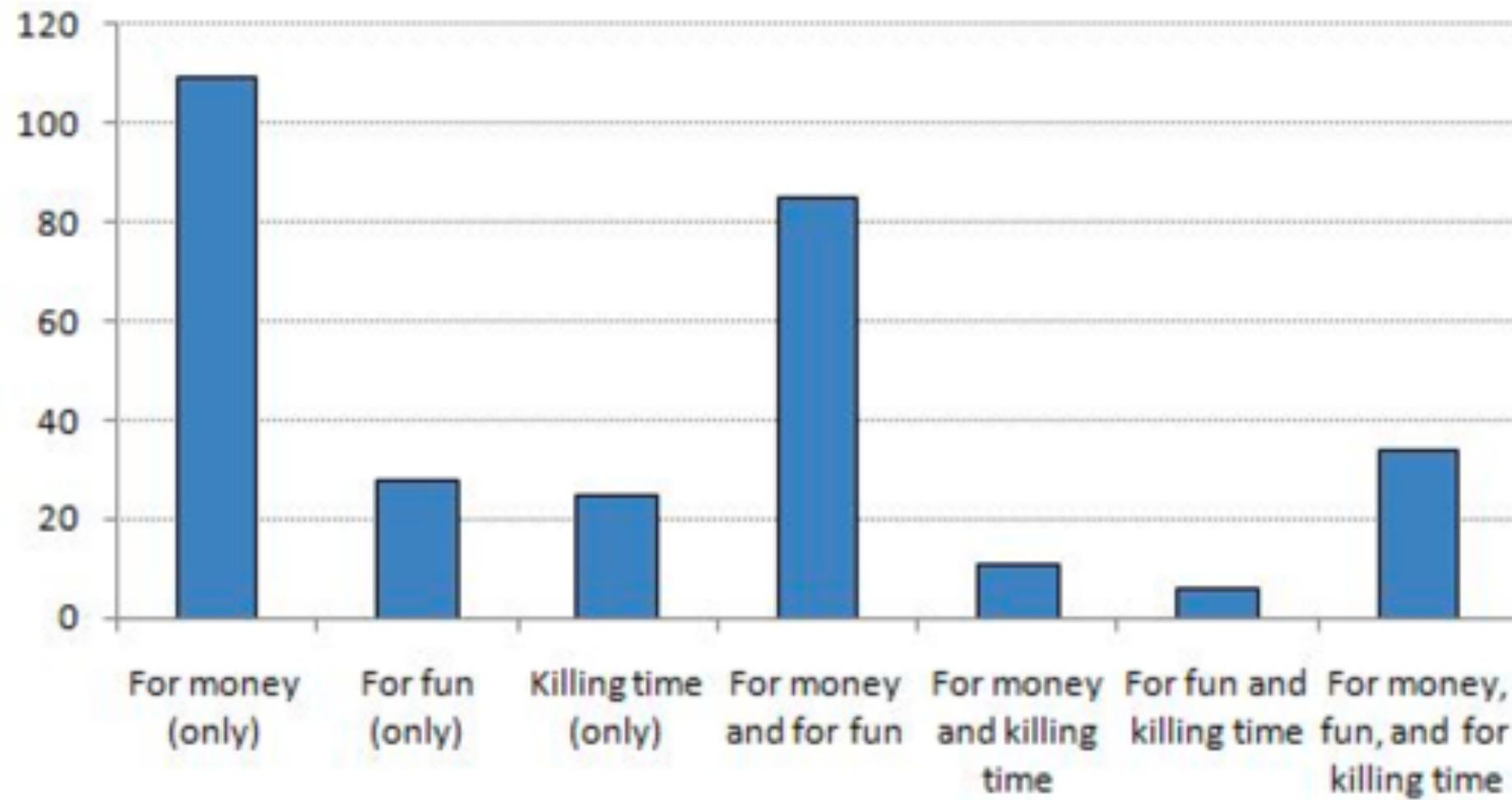
Household <= 2 people 28%

MTurk Worker Demographics (2008): Payment received



<http://www.behind-the-enemy-lines.com/2008/09/how-much-turking-pays.html>

MTurk Worker Demographics (2008): Purpose

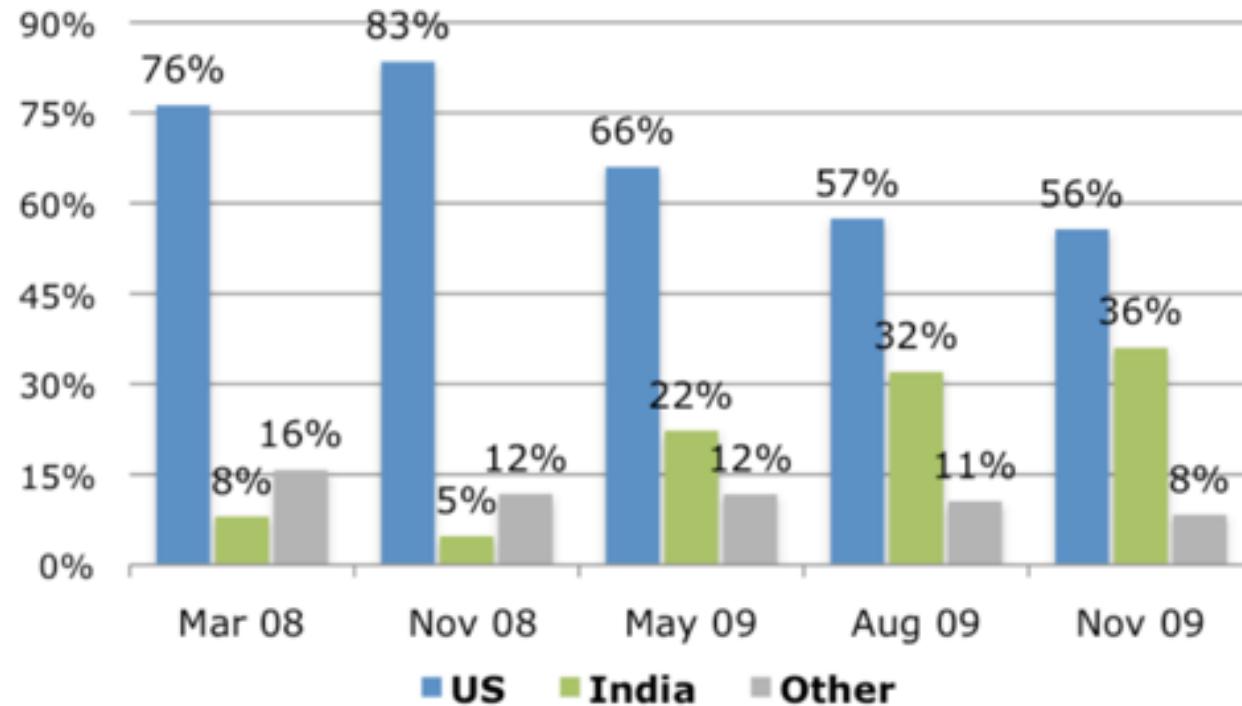


Faces of MTurk workers



https://waxy.org/2008/11/the_faces_of_mechanical_turk/

But Demographics Also Changes Over Time...



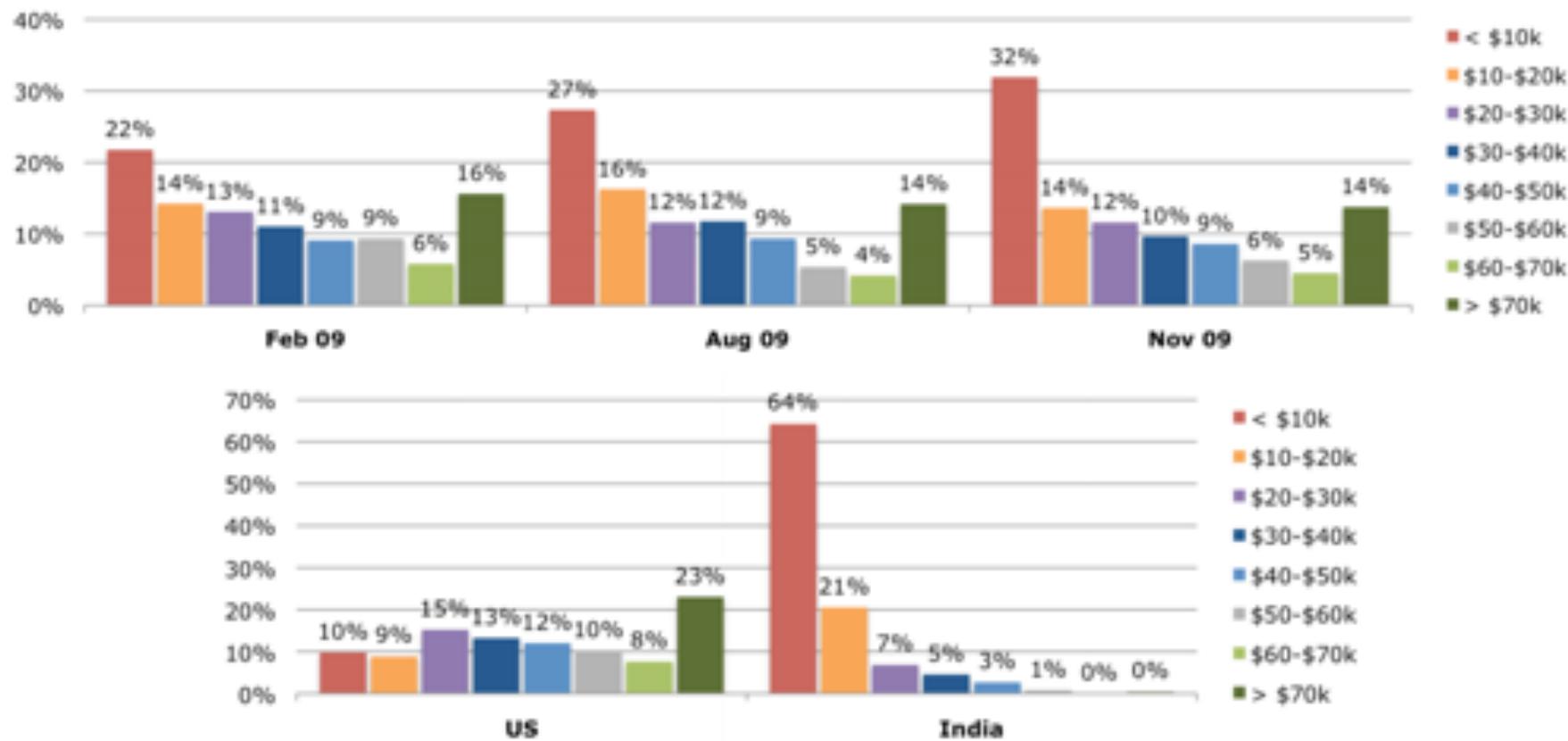
From a primary US-based workforce to an increasingly international group of workers.

But Demographics Also Changes Over Time...

		Nov 08	May 09	Aug 09	Nov 09
Average Age	<i>US</i>	33.6	34.3	33.2	35.4
	<i>India</i>	28.5	28.8	27.6	26.4
Gender	<i>US</i>	28% male, 72% female	34% male, 66% female	31% male, 69% female	37% male, 63% female
	<i>India</i>	75% male, 25% female	61% male, 39% female	69% male, 31% female	66% male, 34% female
Education	<i>US</i>	32% Bachelors, 11% Graduate	34% Bachelors, 14% Graduate	34% Bachelors, 19% Graduate	38% Bachelors, 17% Graduate
	<i>India</i>	69% Bachelors, 29% Graduate	56% Bachelors, 18% Graduate	56% Bachelors, 13% Graduate	45% Bachelors, 21% Graduate

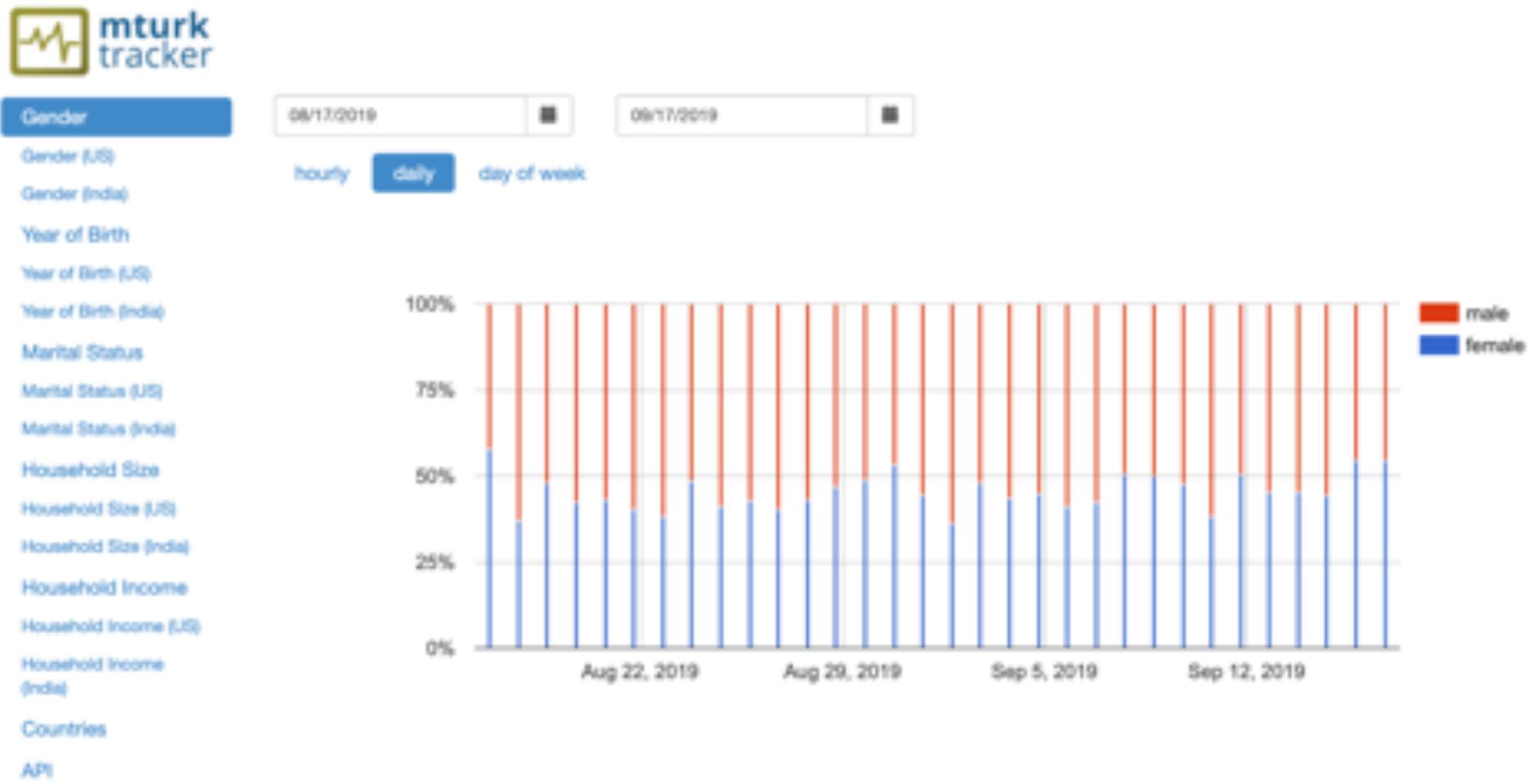
Ross, Joel, et al. "Who are the crowdworkers?: shifting demographics in mechanical turk." *CHI'10 EA*

But Demographics Also Changes Over Time...



Ross, Joel, et al. "Who are the crowdworkers?: shifting demographics in mechanical turk." *CHI'10 EA*

MTurk-Tracker: A Long-Term Demographic Survey



Each worker can take at most one survey in 30 days <http://demographics.mturk-tracker.com/#/gender/all>

MTurk-Tracker: A Long-Term Demographic Survey



Gender

Gender (US)

Gender (India)

Year of Birth

Year of Birth (US)

Year of Birth (India)

Marital Status

Marital Status (US)

Marital Status (India)

Household Size

Household Size (US)

Household Size (India)

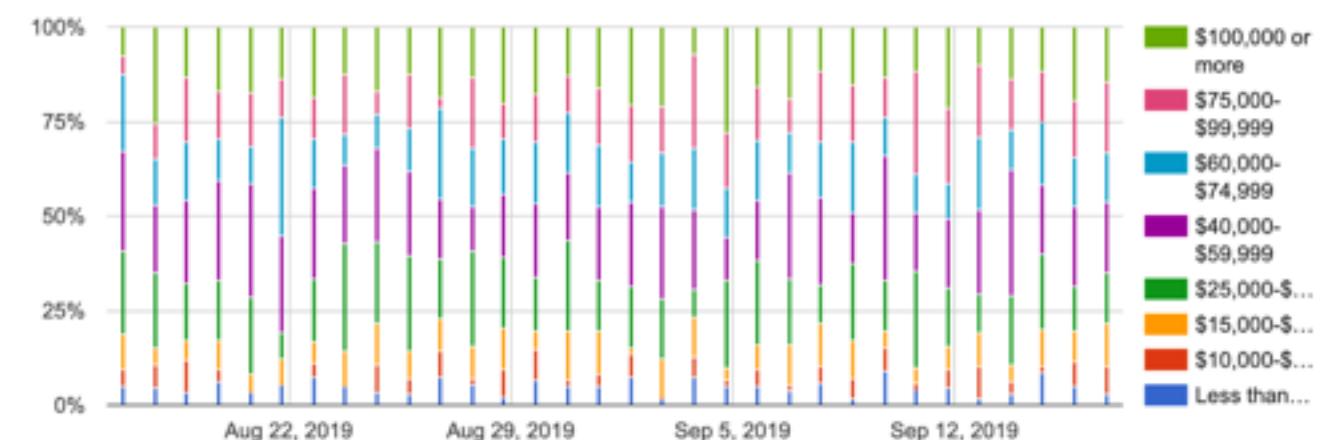
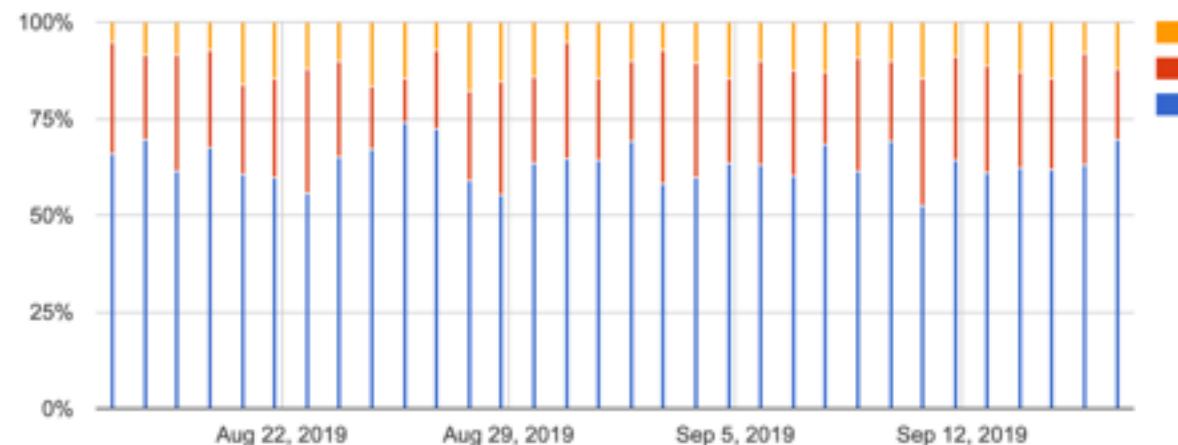
Household Income

Household Income (US)

Household Income (India)

Countries

API



Warm-up Discussion

Does the demographic information collected from self-reported survey responses suffer from any problems?

Can you come up with potential fixes for the problems?

What about ethnographic analysis of digital trace in the required reading?

- Potential issue 1: Data bias (sampling bias, human bias)
- Potential issue 2: Untruthful self-reports

Sampling Bias

Ideally, we want to select participants from the population “uniformly at random”.

However, it’s not always that easy...

1948 US Presidential Election

- Truman vs. Dewey
 - Chicago Daily Tribune decided to run a phone poll of how people voted



Truman



What happened?

One explanation: we cannot claim anything for certain.

“The Inescapability of Uncertainty: AI, Uncertainty, and Why You Should Vote No Matter What Predictions Say”. Vaughan and Wallach. 2016.

However, there are bigger issues here...

- Phones are expensive in 1948...
- Dewey was more favored in rich populations
- Imagine you are polling from people in DC/Texas/NY to predict who will win the presidential election...

Sampling Bias

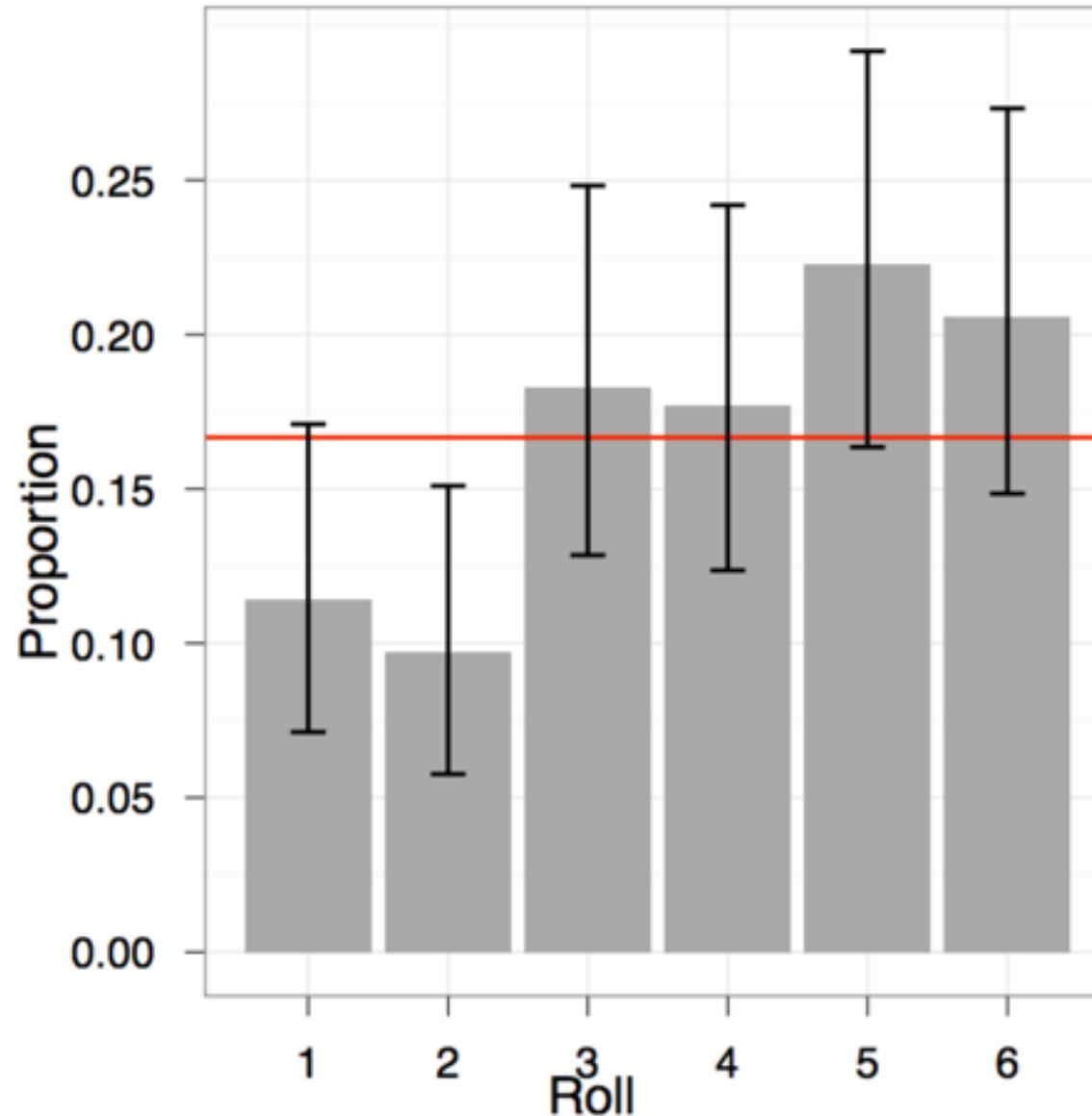
- No simple solutions: Be aware of the issue and think about how much impacts it might bring in your results.
- You might be able to “correct” the bias if you have extra knowledge of the population.
 - Importance sampling
 - Potentially interesting/relevant paper
 - [Conducting Truthful Surveys, Cheaply](#). Roth and Schoenebeck. EC 2012.

Are workers honest?

- Workers are asked to answer demographics questions [Suri et al. 2011]
 - Sex, Age, Location, Income, Education
- Ask workers to privately roll a die and report the outcome.

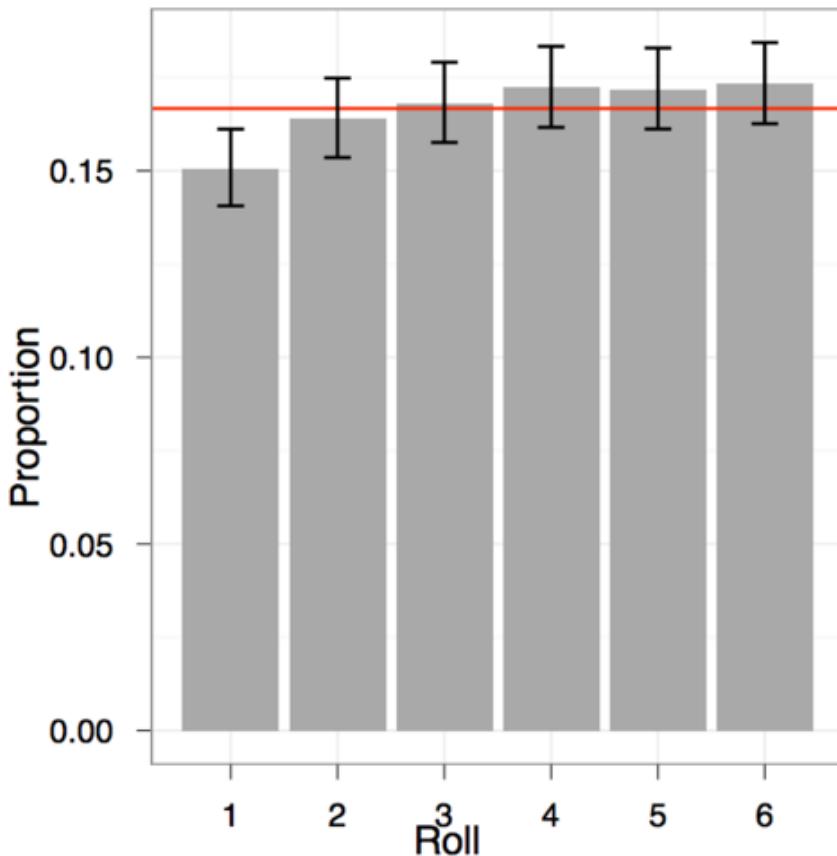
$$\text{Payment} = \$0.25 + \$0.25 * \text{roll}$$

- If all workers are honest, mean report: 3.5
- What do you think the mean was?



Mean: 3.91

Ask workers to report 30 rolls



- Not conclusive evidence, but workers are more honest than we think.
- However, some workers are not. We should be careful to avoid attacks.
 - E.g., designing consistency check questions

Methods for Information Collection

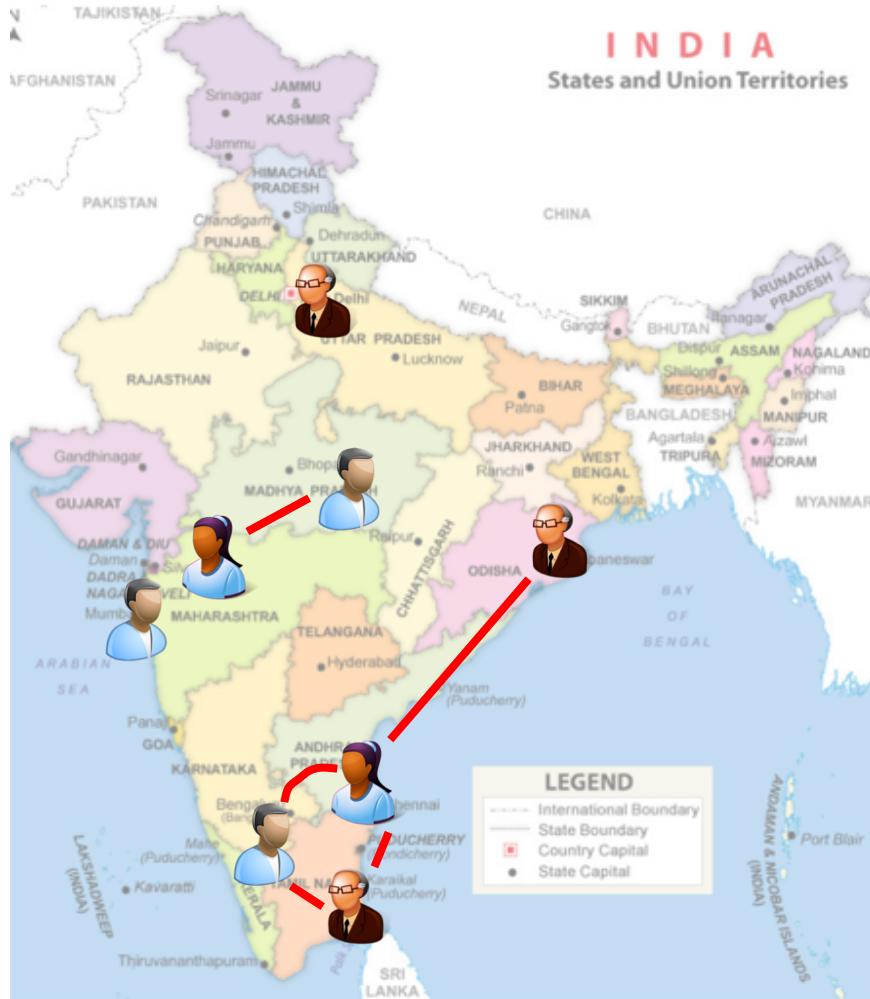
- Direct surveys
- Ethnographic analysis (as in the required reading)
 - Rely on researchers' judgements. Not quantitative.
 - Slowly increasingly, ML has come to help.
- More intricate methods
 - Example 1: How to measure the communication network of the crowd
 - Example 2: How to measure the size of the crowd

Common Assumption: Workers are Independent



[Yin et al., 2016]

Some Evidence From the Field



Workers talk to each other to...
(based on ~100 interviews)



Help with
administrative
overhead



Share useful
information



Recreate social
connections

Can we Map the Network?



What is the scale?

What is the structure?

How is it used?

[Yin et al., 2016]

Why is it Challenging?



Not accessible
through API



Not on the
MTurk platform



Nowhere to
download



Can't just crawl
from the web

The goal...

- ◆ Elicit “true” connections only
- ◆ Elicit as many true connections as possible
- ◆ Preserve workers’ privacy

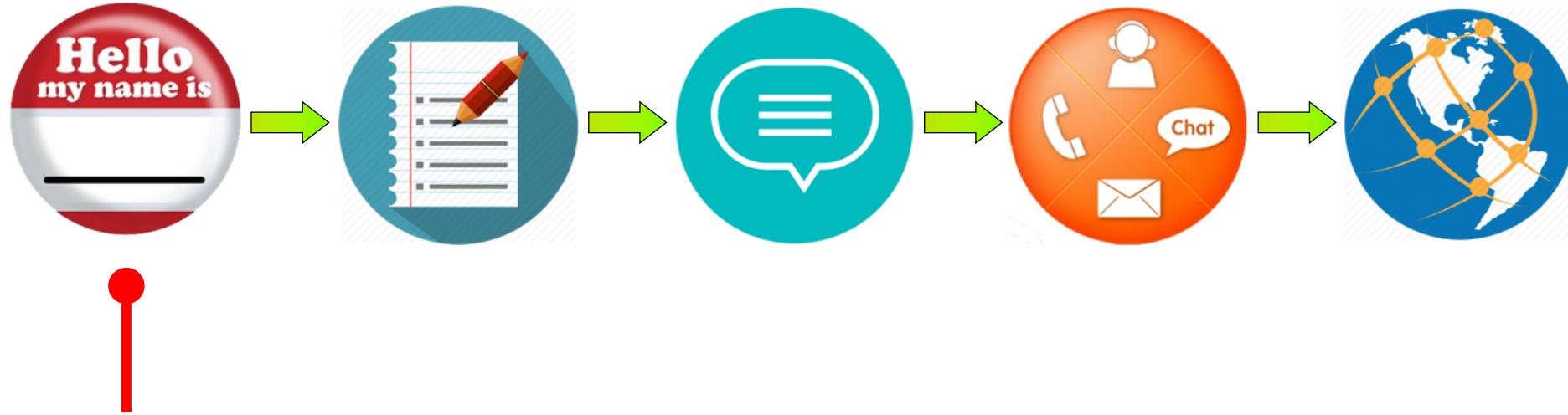
Can't directly ask for names!

Can't pay by connections!

A Web App

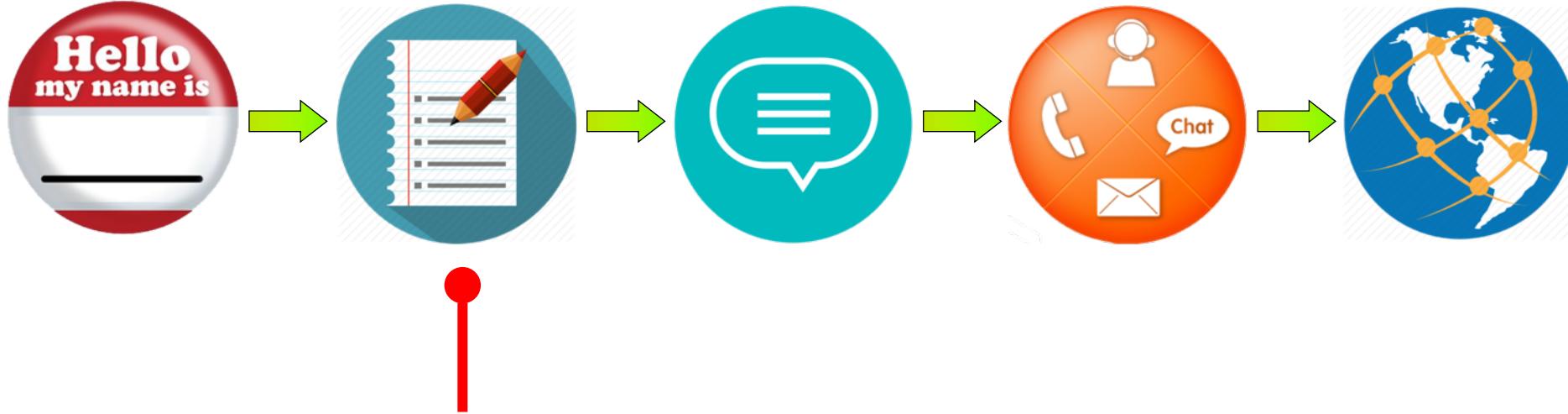
- Workers **self-report** their connections
- Provides some **value back** to the workers so it's their best interest to report as many true connections as possible

The Network Mapping App



Step 1: Create a nickname for yourself

The Network Mapping App



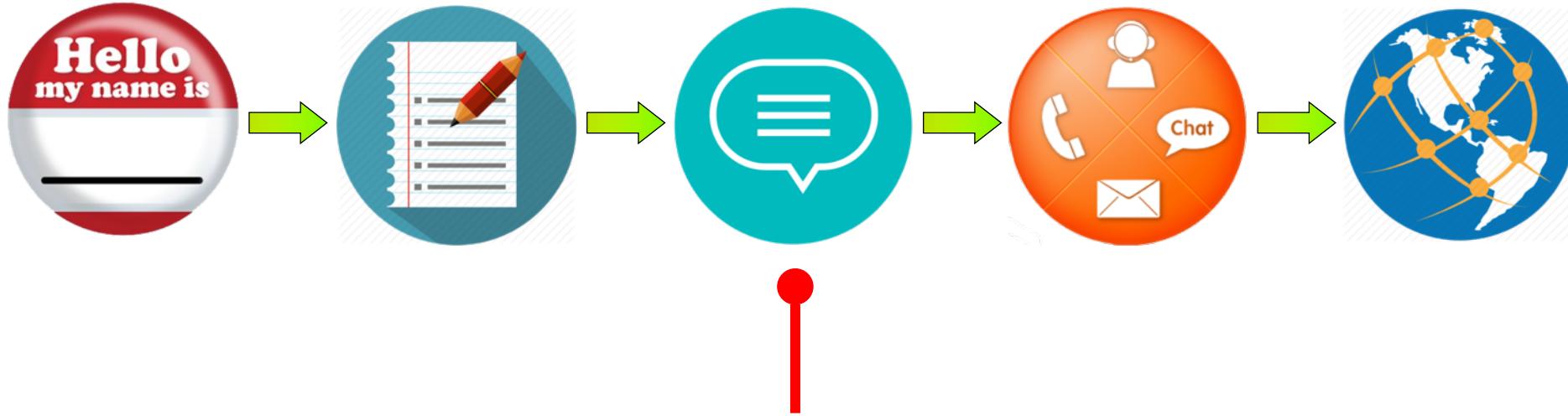
Step 2: Complete a brief demographic survey

Age
Gender
Country
Approval rate
...

- Share with all other workers
- Share with workers connected to me
- Not share with anyone

Value back: Who are the other workers?

The Network Mapping App



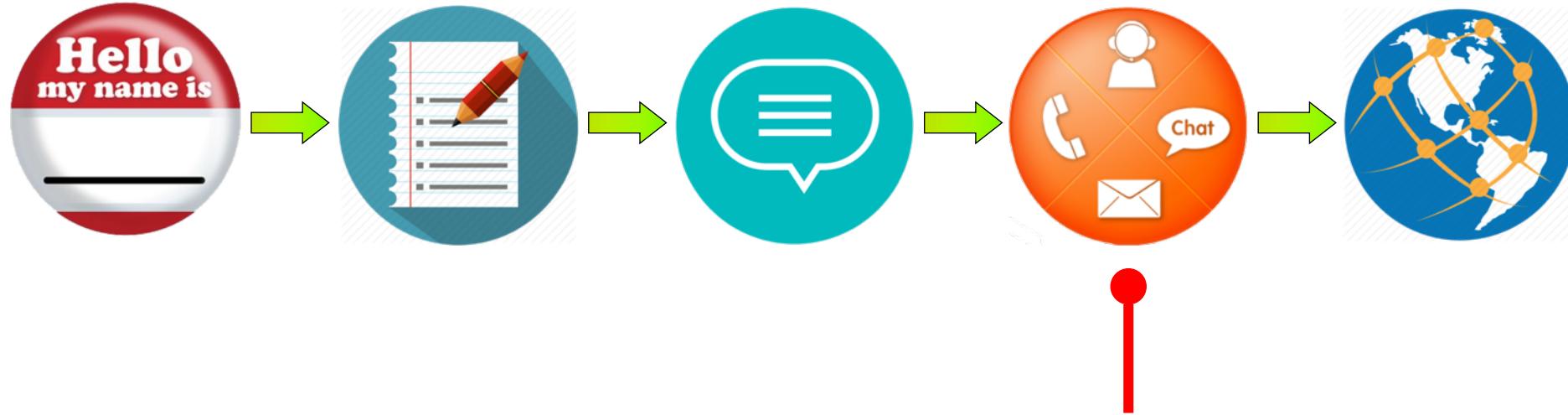
Step 3: Tell us about your personal experience with MTurk

Why did you start Turkling?
What motivates you to keep Turkling?

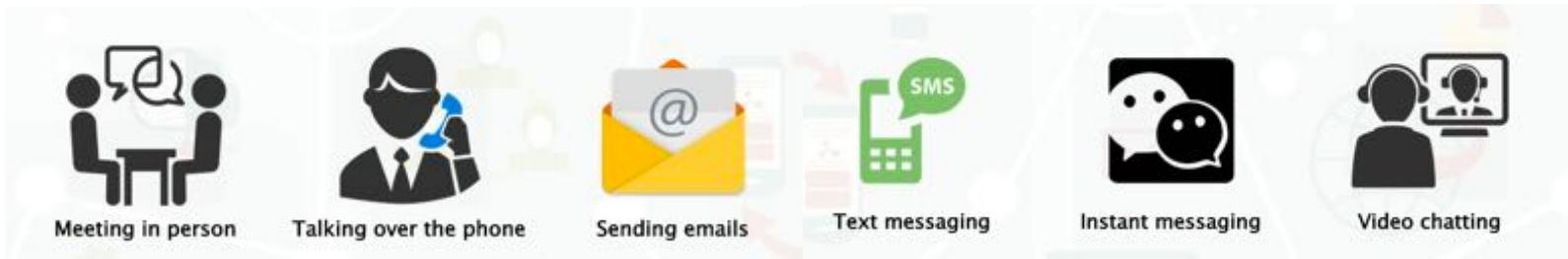
(Selected based on a pilot study!)

Value back: What are other workers' Turkling stories?

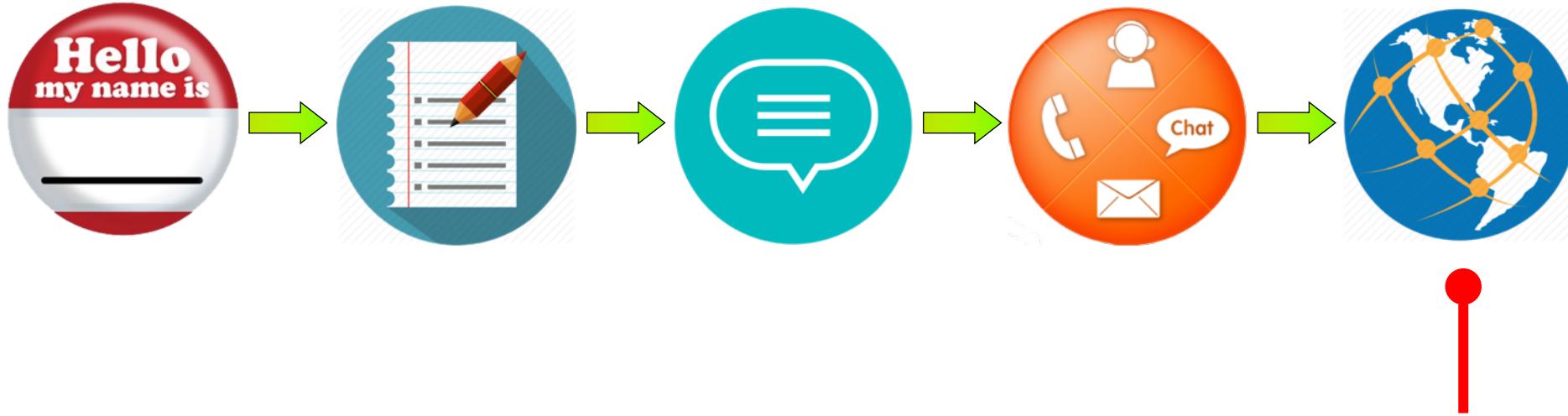
The Network Mapping App



Step 4: Swap nicknames with Turkers you know

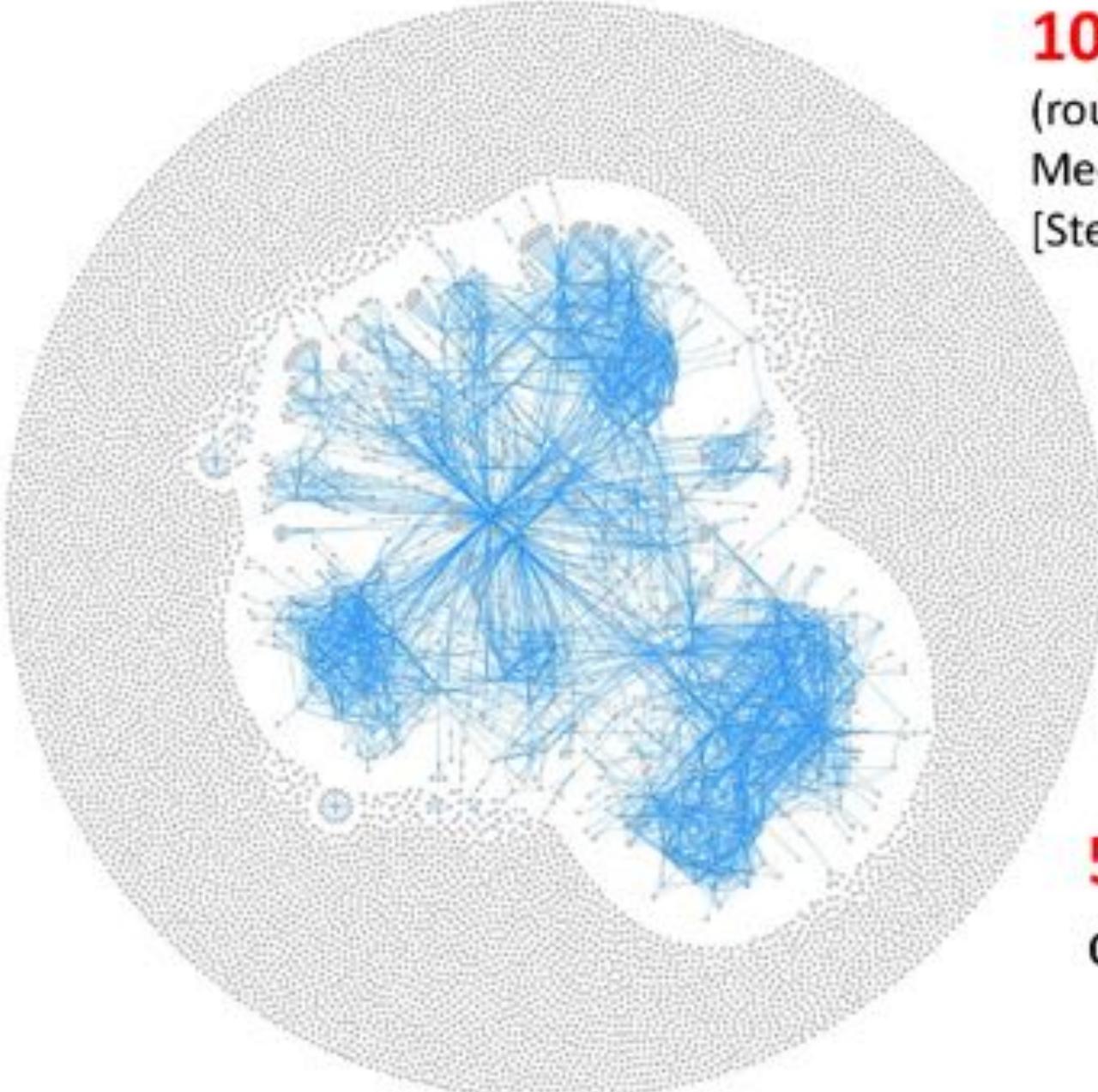


The Network Mapping App



Step 5: Explore the Turker network

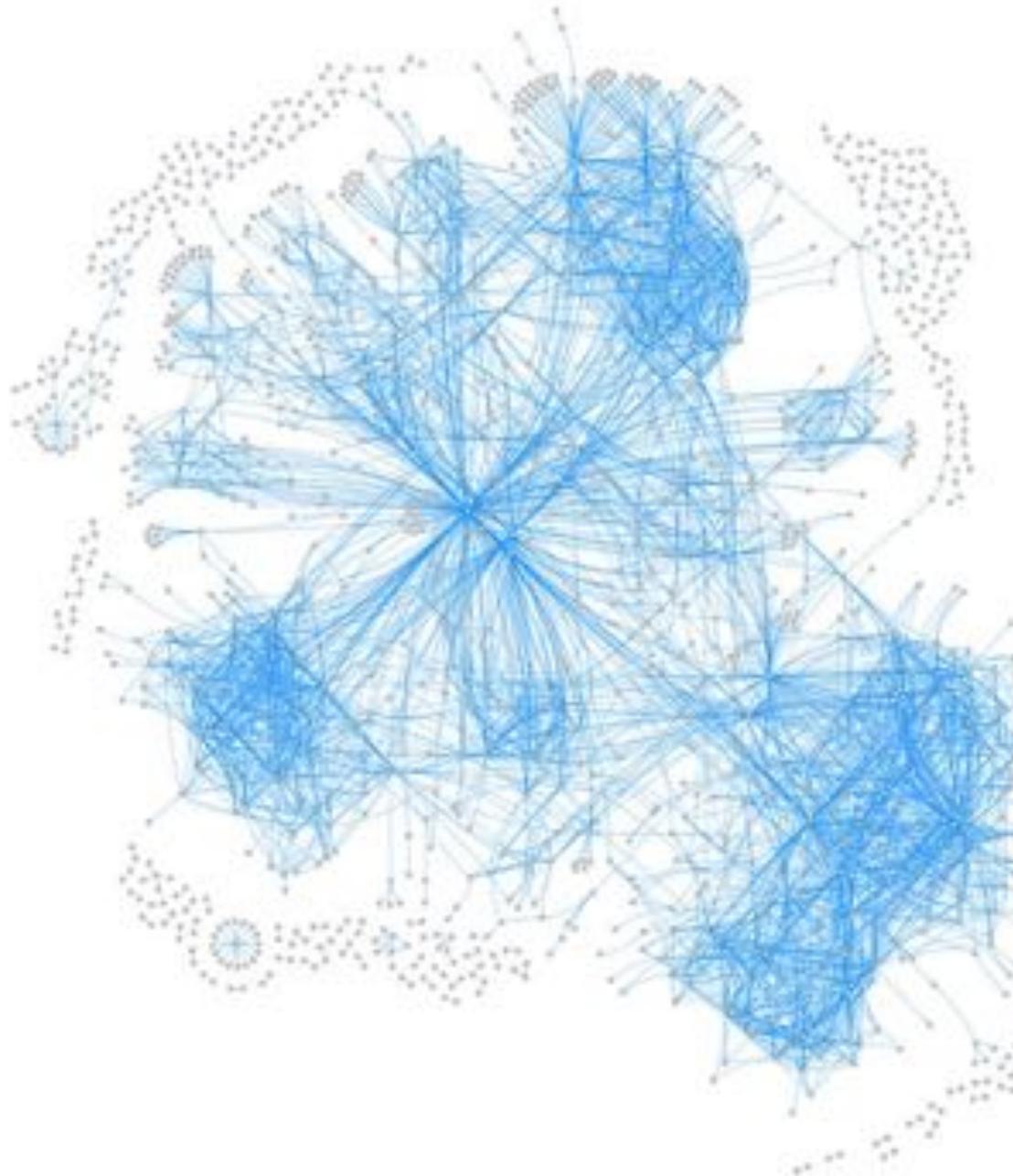




10,354 workers
(roughly a census of
Mechanical Turk
[Stewart et al. 2015])

5268
connections

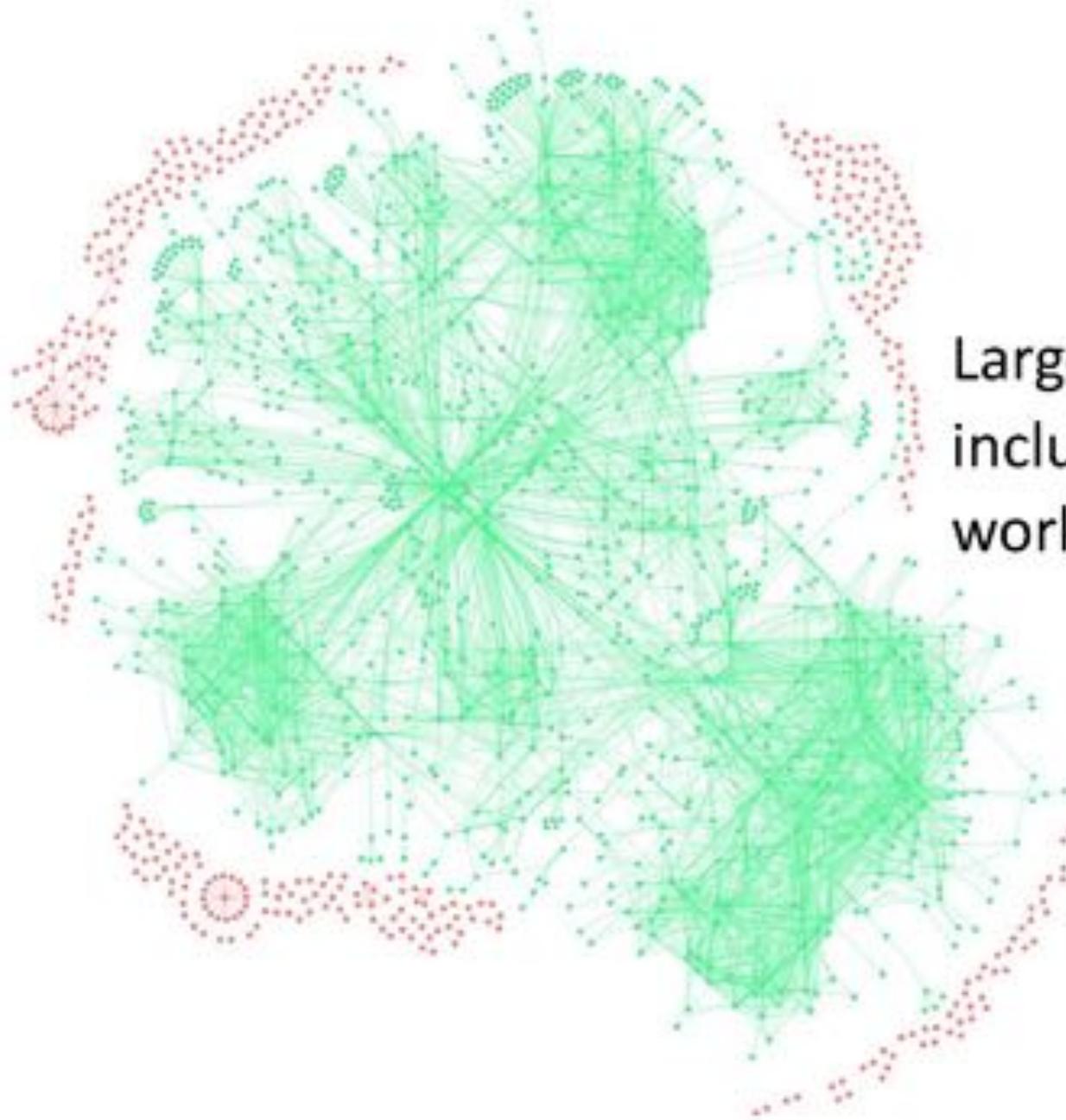
[Yin et al., 2016]



1,389 (13%)
connected
workers

On average,
workers
communicate
with **7.6** others

Max degree
is **321**
[Yin et al., 2016]



Largest component
includes **994 (72%)**
workers

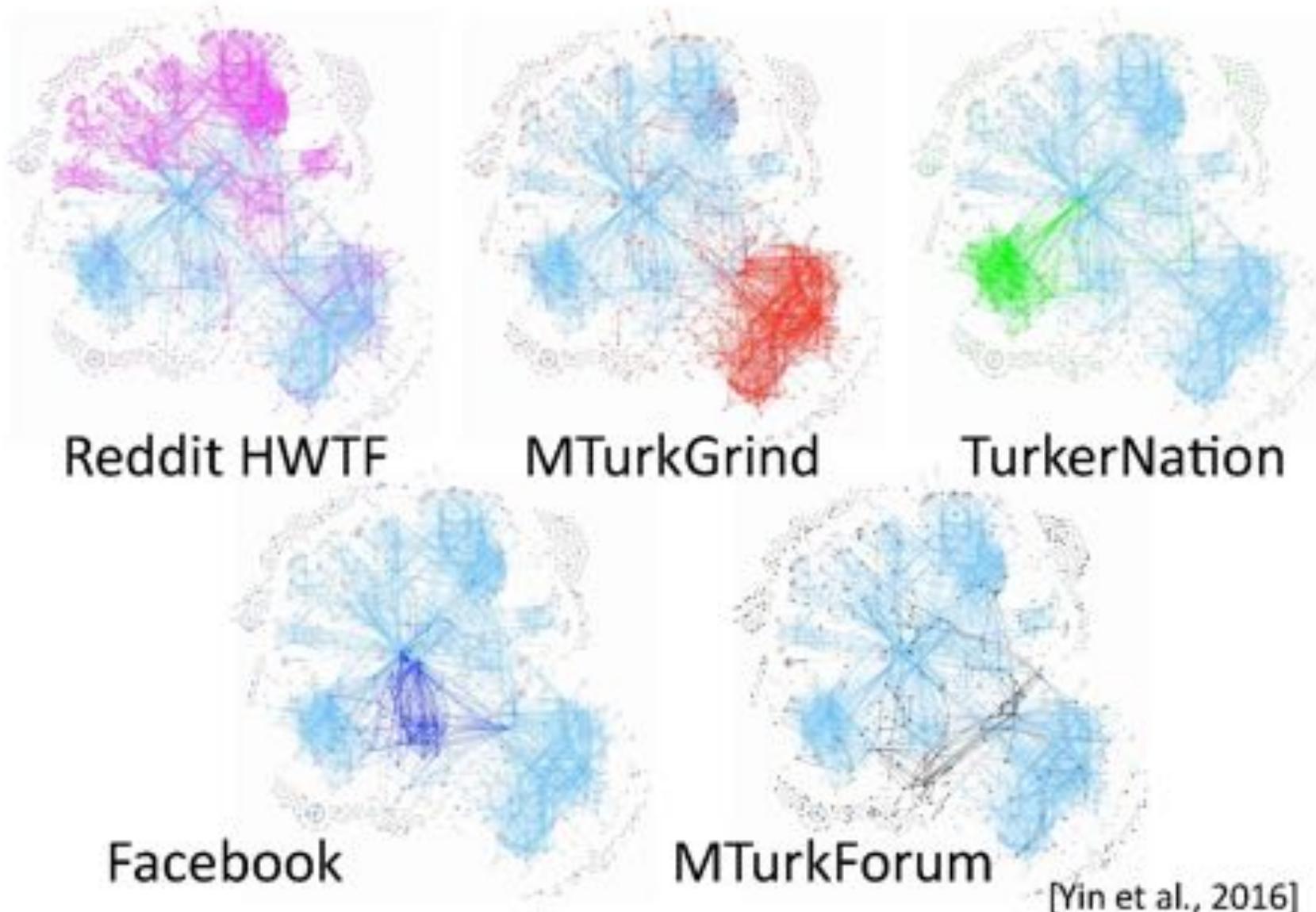
[Yin et al., 2016]

A Network Enabled By Forums

- 59% of all workers and 83% of connected workers reported using at least one forum.
- 90% of all edges are between pairs of workers who communicate via forums, and 86% are between pairs who communicate *exclusively* through forums.

[Yin et al., 2016]

Forums Create Subcommunities



Subcommunities Are Different



Topological Structure: How tightly connected is each subcommunity?



Temporal Dynamics: Do relationships endure over time?



Communication Content: Is communication social or strictly business?

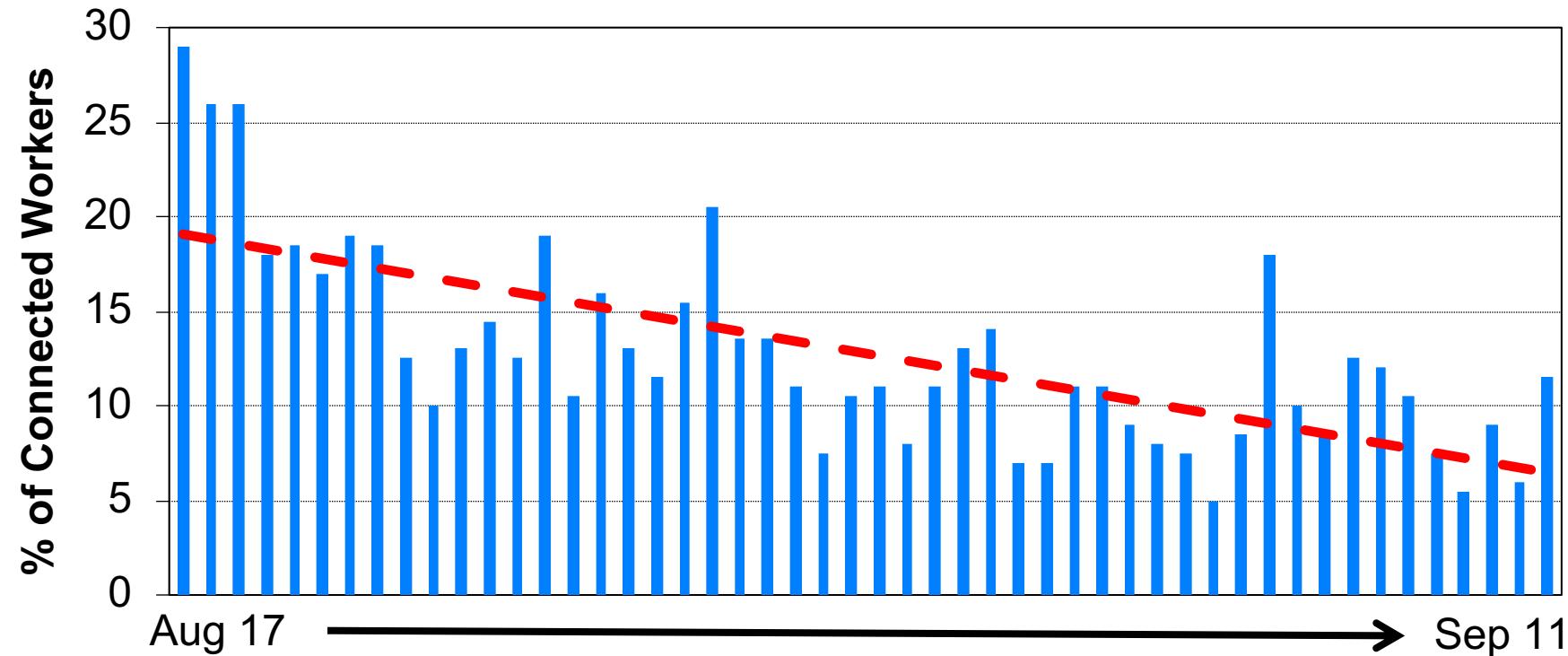
[Yin et al., 2016]

Measures of Success

Property	Connected	Unconnected
Be active > 1 year	55%	46%
Use forums	83%	56%
Master	11%	7%
Approval rate	98.6%	97.4%

Connected workers were also **more likely** than unconnected workers to find our task **early**.

Connected Workers Find HITs Earlier



Discussion

- What are the implications of this study?
- What do you think the fact that there are communication networks would affect the way we think of crowdsourcing?

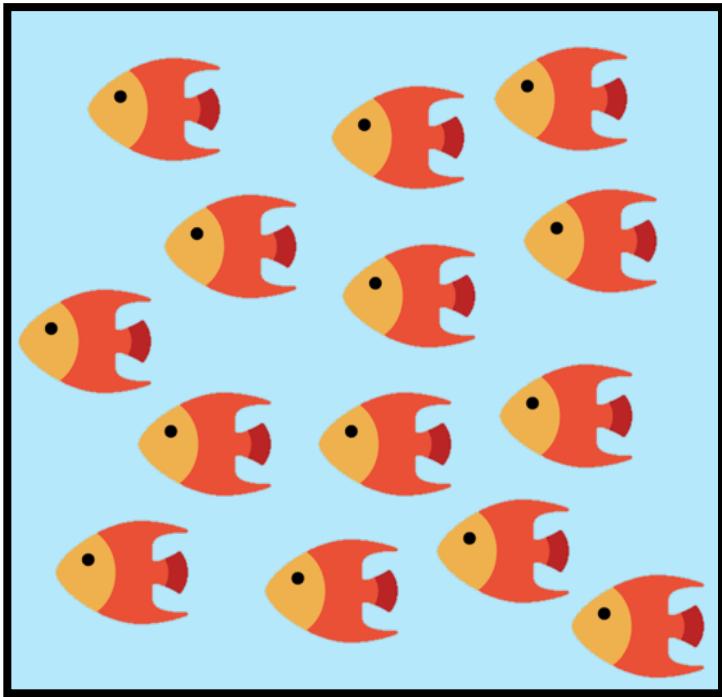
Oct 16 Practical Issues: Non-Independent Work
Presenter: Guanghui, Zifan, and Yanni

Required
[MicroTalk: Using Argumentation to Improve Crowdsourcing Accuracy.](#)
Drapeau et al. HCOMP 2016.

Optional
[Revolt: Collaborative Crowdsourcing for Labeling Machine Learning Datasets.](#) Chang et al. CHI 2017.
[Atelier: Repurposing Expert Crowdsourcing Tasks as Micro-Internships.](#)
Suzuki et al. CHI 2016.
[Cicero: Multi-Turn, Contextual Argumentation for Accurate Crowdsourcing.](#)
Chen et al. CHI 2019
[Leveraging Peer Communication to Enhance Crowdsourcing.](#) Tang, Ho, and Yin. WWW 2019

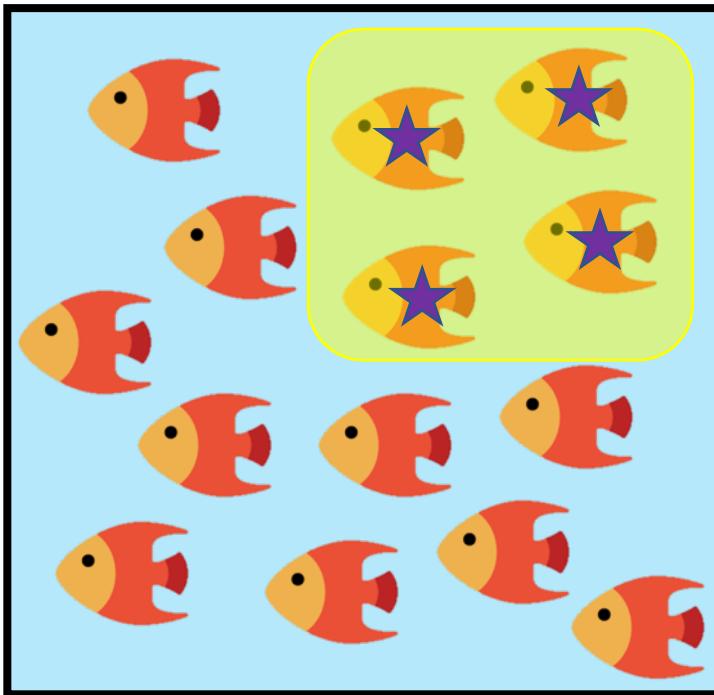
How Many Workers are There?

How Many Fishes are in the Lake?

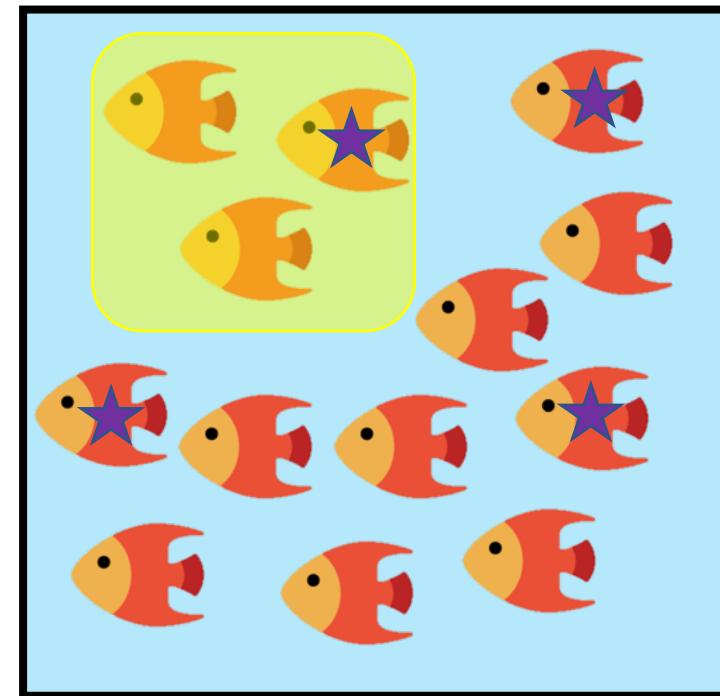


How Many Fishes are in the Lake?

The capture-recapture analysis



M fish, N_1 captured (& marked)



N_2 recaptured, n marked

$$n = \frac{N_1}{M} \times N_2$$

$$M = \frac{N_1 N_2}{n}$$

How Large is the Crowd?

- Applying the capture-recapture analysis to workers who respond to demographic surveys, we concluded that $M= 13,410$.
- What assumptions are made?
 - A1: Closed population model: No workers will leave the worker pool, and no new workers will join the pool
 - A2: Equal propensity of participation: The probability to participate in demographic survey is equal across all workers.

How Large is the Crowd?

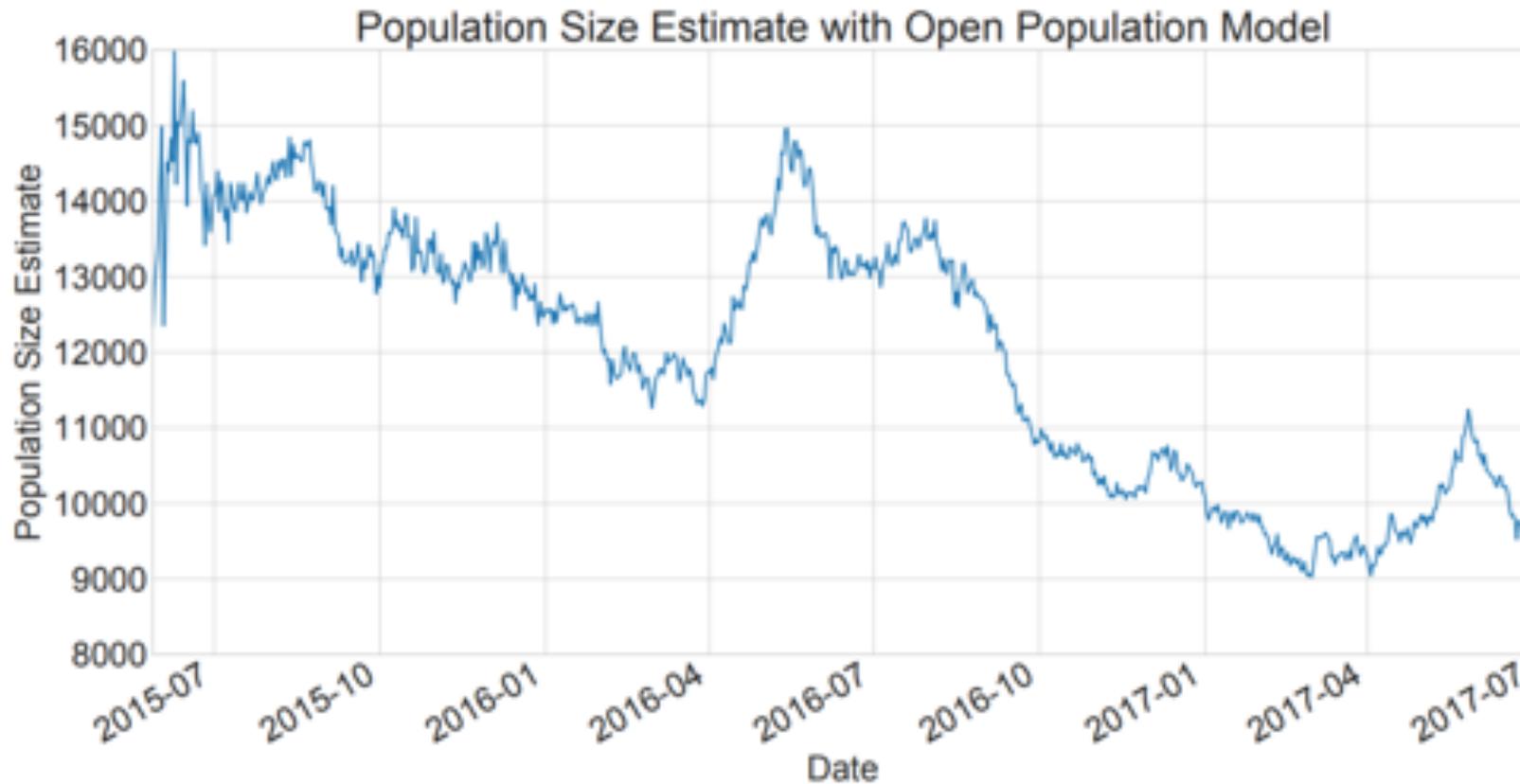
Relaxing A1: Open population model

- Survival probability: $S(t) = \exp(-\lambda t)$
- At time $d - t$, N_{d-t} workers participated in the survey
- At time d , assume that there are M_d workers in total. Among N_d workers that participated in the survey, $n_{d,t}$ workers also participated in the survey at time $d - t$.

$$n_{d,t} = \frac{N_{d-t} \exp(-\lambda t)}{M_d} \times N_d$$

$$M_d = \frac{N_{d-t} N_d \exp(-\lambda t)}{n_{d,t}}$$

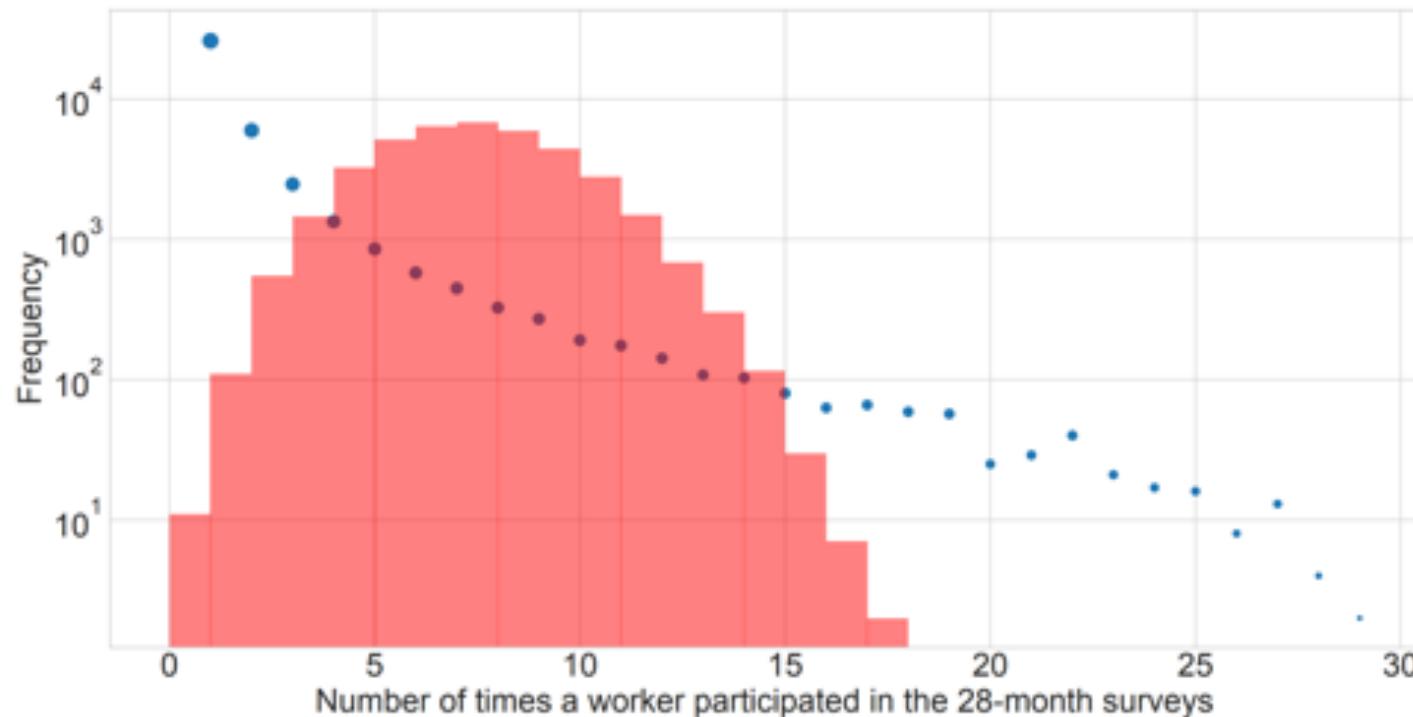
How Large is the Crowd?



Half-life of the MTurk worker population is 404 days

How Large is the Crowd?

- Is “A2: Equal propensity of participation” reasonable?
 - If every worker participate the survey with the same probability, we should see a binomial distribution over # participations with many surveys.



How Large is the Crowd?

Relaxing A2: Accounting for propensity of participating surveys

- Assume there are N workers in total, each worker gets a propensity score $a_i \sim p(a)$
- Capture n_1 workers, $P(\text{capture}|a_i) = 1 - (1 - \frac{a_i}{\sum_{j=1}^N a_j})^{n_1} \approx \frac{n_1 a_i}{N E[a]}$
- Recapture n_2 workers, $P(\text{capture, recapture}|a_i) = \frac{n_1 n_2 a_i^2}{N^2 E[a]^2}$
- The expected number of workers participating in both surveys:

$$m = N \int \frac{n_1 n_2 a_i^2}{N^2 E[a]^2} p(a) da = \frac{n_1 n_2}{N} \left(1 + \frac{\text{Var}[a]}{E[a]^2}\right)$$

- Following this method, the estimate for the number of Mturk workers is 178,000!

Data-Driven Approaches

- Want to collect some information.
- It's challenging to obtain enough data points to estimate the information.
- Propose models (i.e., making assumptions) of the process. So you only need to estimate a small number of model parameters.
- Using ML approaches to learn those parameters from data.
- Your estimate is only as accurate as the best your model can do.

Discussion

- What additional information do you want to know about Turkers?
- Can you come up with methods to collect that information?
- Does that information help us design better crowdsourcing platforms? How?