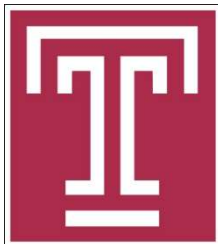# Crowdsourcing as a Future Collaborative Computing Paradigm
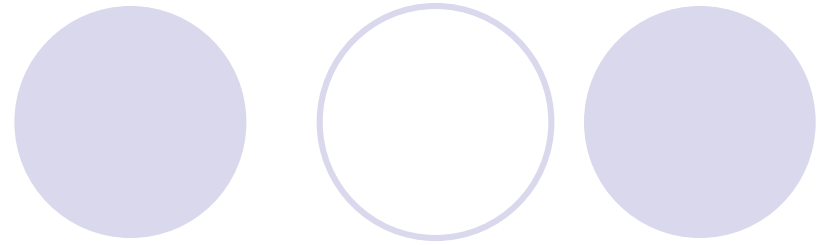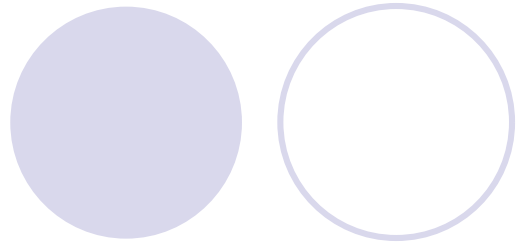
Jie Wu

Dept. of Computer and Info. Sciences

Temple University

Road Map

- Introduction
- Mechanical Turk
- Sample Applications
- Algorithmic Paradigms
- Challenges and Opportunities
- Conclusion

What and why

Basic components

Motivation examples

INTRODUCTION

What is Crowdsourcing?

- Coordinating a crowd (a large group of people online) to do microwork (small jobs) that solves problems (that software or one user cannot easily do)

- Crowdsourcing: crowd + outsourcing (through Internet)

- Amazon's Mechanical Turk and CrowdFlower

The Benefits of Crowdsourcing

- Performance
  - Inexpensive and fast
  - The whole is greater than the sum of its parts

- Human Processing Unit (HPU)
  - More effective than CPU (for some apps)
    - Verification and validation: Image labeling
    - Interpretation and analysis: Language translation
    - Surveys: Social network survey

- High adoption in business (85% of the top global brands) based on eYeka

Basic Components

- Requester
  - People submit jobs (microwork)
  - Human Intelligence Tasks (HITs)
- Worker
  - People work on jobs
- Platform
  - Job management

Amazon Mechanical Turk (MTurk): 18th century chess playing robot with a human inside


Requester

Worker Pool

History

- **1714:** Longitude Prize: tried to find a way to measure a ship's longitudinal position

- **1884:** 800 volunteers catalogued words to create *Oxford English Dictionary*

- **2001:** Wikipedia: free-access, free content Internet encyclopedia

- **2005**: Threadless.com, with members creating own design

- **2006:** Howe and Robinson introduced the term in Wired

- **2008**: Brabham published first scholarly work using the term
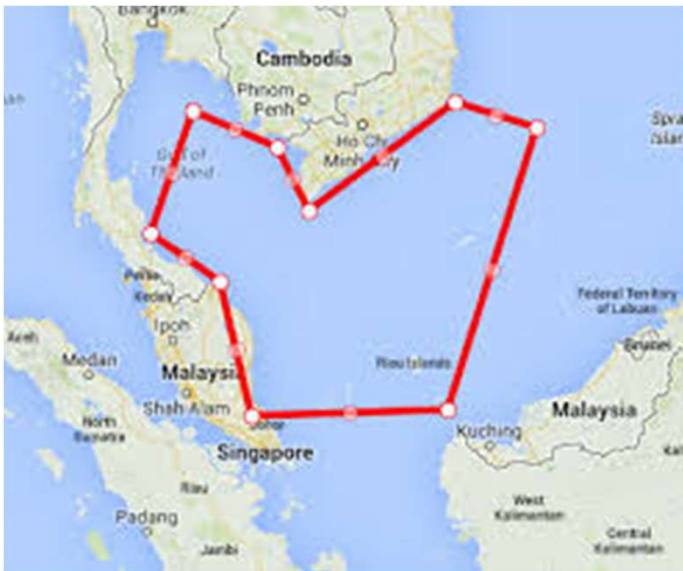
Help Find Jim Gray

- Jim Gray, Turing Award winner, went missing with his sailboat outside San Francisco Bay in January 2007.

- Use satellite image to search for his sailboat.
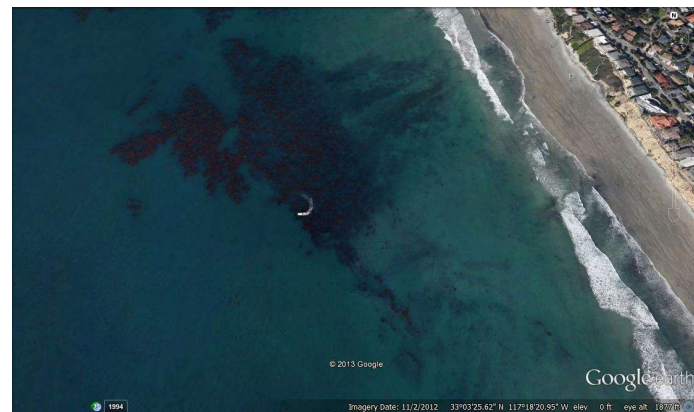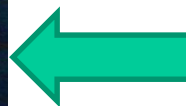
Malaysia Airlines Flight MH 370

- **DigitalGlobe**
  - Crowdsourcing volunteers comb satellite photos for Malaysia Airlines jet

- **March 11, 2014** (from CSU prof. email)

I just saw on our local Denver Fox news (KDVR.com) that a local company, DigitalGlobe, has reoriented their satellites to take high-res images in the area where the plane may have crashed. Crowdsourcing efforts are on to have people scan these images and find signs of debris. I was reminded of Jie Wu's talk earlier this month.

DARPA Network Challenges

WE HAVE A WINNER!

MIT RED BALLOON CHALLENGE TEAM

Read about the winner of the DARPA Network Challenge

- Problem (2009): $40,000 challenge award for the first team to find 10 balloons.

- MIT team won under 9 hours.

- Winning strategy
  - $2,000 per balloon to the first person to send the correct location
  - $1,000 to the person who invited the winner
  - $500 to whoever invited the inviter
  - ... (or to charity) ...

Tag Challenges

- **Problem (March 31, 2012):** Find five suspects in Washington, D.C., New York, London, Stockholm, and Bratislava.

- Winner from UCSD CrowdScanner: located 3 of the 5 suspects.

- Winning strategy: same as MIT. Also, recruiters of the first 2,000 get $1.
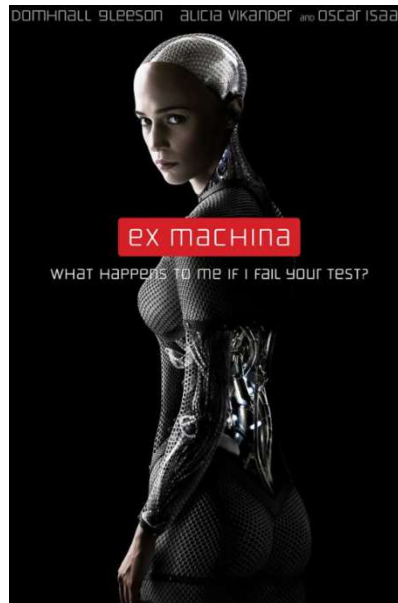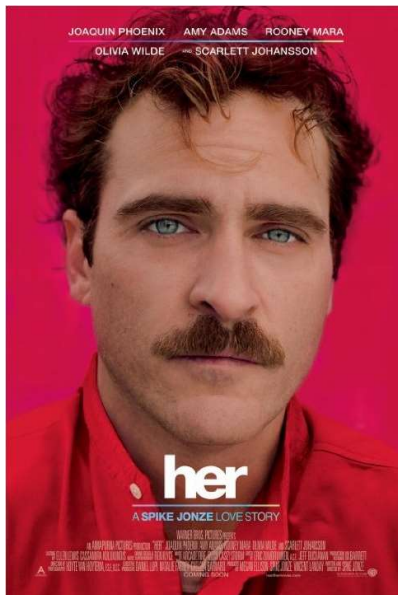
Washington DC

New York City

Bratislava

AI Could End Human Race (Stephen Hawking)

- ## Stephen Hawking
  - "Humans, who are limited by slow biological evolution, couldn't compete, and would be superseded."

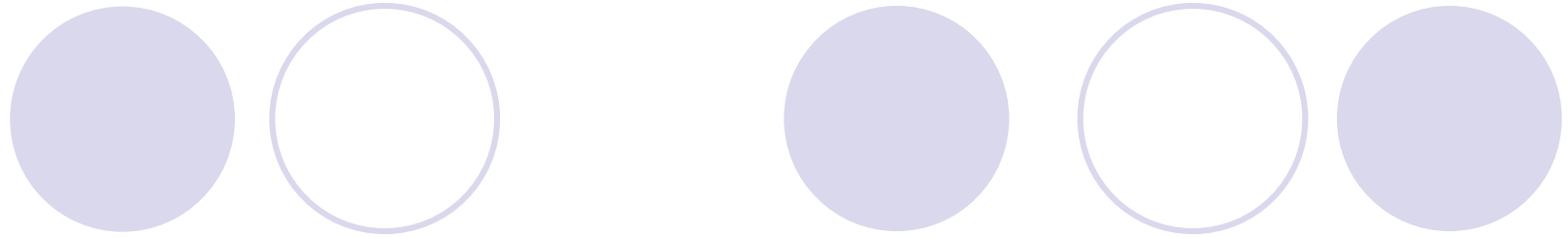- ## Recent movies
  - Her (2014) & Ex-Machina (2015)

Smarter Than You Think

- **Who is smarter**
  - Human or computer?
- AI will redefine
  - What it means to be human
- Our Machine Masters
  NY Times, Oct. 31, 2014

- 1997 (Chess)
  - Kasparov vs. Deep Blue
- 1998
  - Kasparov vs. Topalov: 4:0
  - Kasparov + machine vs.
    Topalov + machine: 3:3
- 2005 (freestyle tournament)
  - Grand-master (>2,500)
  - Machine (Hydra)
  - Grand-master + machine
  - Amateurs (>1,500) + machine *
- 2016 (Go game)
  - AlphaGo vs. Lee Sedol: 4:1
  - AlphaGo vs. Jie Ke: 3:0 (May 2017)

Platform

HIT

Worker

MECHANICAL TURK

amazon mechanical turk
Artificial Artificial Intelligence

Name | Account Settings | Sign Out | Help

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.
293,089 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**

TASKS

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

- As a worker, make an average of $0.05 per task

- Paid directly to Amazon account

- 130 M tasks posted (2009-2014)

- As requester, set up simple tasks for workers to complete

- Quality control is possible through MTurk services

Worker: Contract for a HIT

amazonmechanical turk
Artificial Artificial Intelligence

Your Account | HITs | Qualifications | 293,115 HITs available now

All HITs | HITs Available To You | HITs Assigned To You

for which you are qualified
require Master Qualification

Find HITs containing that pay at least $ 0.00 GO

All HITs
1-10 of 1982 Results

Sort by: HIT Creation Date (newest first) GO!   Show all details | Hide all details   1 2 3 4 5 > Next >> Last

Copy Text from Coupon Image			Not Qualified to work on this HIT (Why?)	View a HIT in this group
Requester: Coupon Vision	HIT Expiration Date: Jun 21, 2014 (51 weeks 2 days)	Reward: $0.08		
	Time Allotted: 10 minutes	HITs Available: 14		

Proofread OCR Data			Take Qualification test (Why?)	View a HIT in this group
Requester: Brian Robertson	HIT Expiration Date: Jul 3, 2013 (6 days 23 hours)	Reward: $0.30		
	Time Allotted: 2 hours	HITs Available: 2		

Get product codes and prices from receipt image (get bonuses for long receipts)			Request Qualification (Why?)	View a HIT in this group
Requester: Shopping	HIT Expiration Date: Jul 1, 2013 (4 days 23 hours)	Reward: $0.03		
	Time Allotted: 45 minutes	HITs Available: 2		

Click and provide fast feedback B-US RHL-003			Not Qualified to work on this HIT (Why?)	View a HIT in this group
Requester: CrowdFlower	HIT Expiration Date: Jul 3, 2013 (6 days 23 hours)	Reward: $0.01		
	Time Allotted: 30 minutes	HITs Available: 219		

Basic Caption Requirements				View a HIT in this group
Requester: Redwood	HIT Expiration Date: Jun 26, 2014 (52 weeks)	Reward: $0.02		
	Time Allotted: 15 minutes	HITs Available: 11		

Identify company/publication from a photo   Not Qualified to work on this HIT (Why?) | View a HIT in this group

- Select a HIT (human intelligence task)
  - By creation date, payment amount, time allotment

Worker: Reviewing a HIT

Want to work on this HIT?   Want to see other HITs?

Accept HIT   **Skip HIT**

Total Earned: $4.72
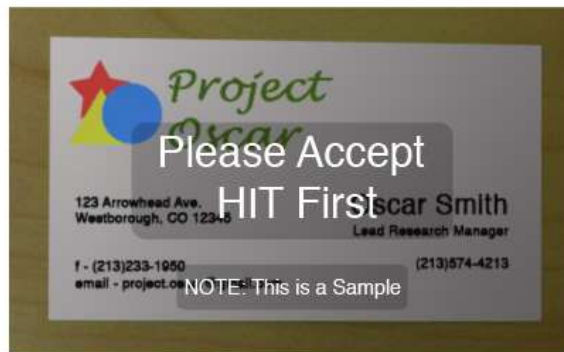**Total HITs Submitted:** 7

Copy Text from Business Card
Requester: Oscar Smith
Qualifications Required: None

Reward: $0.02 per HIT   **HITs Available:** 39   **Duration:** 10 minutes

---

Please Copy Text from Business Card:

Your Current Quality Score is: --
If you have a high enough score, you will be   ?
considered for promotion to a Trusted Worker.

Please Accept
HIT First

*Project
Oscar*

123 Arrowhead Ave.
Westborough, CO 12345

Oscar Smith
Lead Research Manager

f - (213)233-1950   (213)574-4213
email - project.o

NOTE: This is a Sample

Name		?
Title	Company
Email	Website

Address: ?

Address Line 1

add line

| City | State | Zip Code |

Phone: click here if not a U.S. phone number ?

| Work | Ext. |
| Mobile |
| Fax |

add phone

Please **select/crop** company logo or
image from the business card above.
Click + Drag to select the company logo.

- Review the HIT before accepting
  - Shown full task, allotted time (10 minutes), reward amount ($0.02)

Types of HITs

- Information Finding
  - Searching the web to answer a question
- Interpretation and Analysis
  - Interpreting web content
- Verification and Validation
  - Verifying and validating certain information
- Content Creation
  - Generating new content
- Content Access
  - Accessing web content
- Surveys
  - Taking appropriate action based on the survey result

Threadless

SHOP MEN'S BY POPULARITY / NEWEST

DEPARTMENT   VIEW ALL

Men's
Women's
Kids
Babies
Toddlers
Phone Cases
Wall Art
Accessories
Home

STYLE   VIEW ALL

Tees
Triblend Tees
V-Necks
Longsleeves
Hoodies
Sweatshirts
Tanks
Sublimated Triblend Tees

SIZE

X Small
Small
Medium
Large
X Large
2X Large

24% OFF
Let's Summon Demons
Design by Steven Rhodes
$25 $19 and up

20% OFF
Can Knot
Design by Haasbroek
$30 $24

24% OFF
Many Lands Under One Sun
Design by Rick Crane
$25 $19

24% OFF
The Cat and the Koi
Design by roni saptoni
$25 $19

iStock Photo

TopCoder

Major Types of Crowdsourcing

- **Virtual labor markets (VLM)** – Platforms where users can complete work for monetary compensation (e.g., Amazon's MTurk)

- **Tournament Crowdsourcing (TC)** – Also known as ideas competitions, where only the best solution is compensated (e.g., Crowdflower and TopCoder)

- **Open collaboration (OC)** - Typically do not offer monetary compensation, people are often prompted through social media with the opportunity to help out; like an open call to anyone (e.g., Wikipedia)
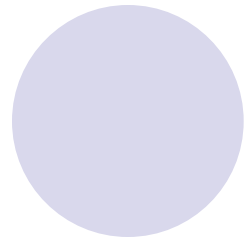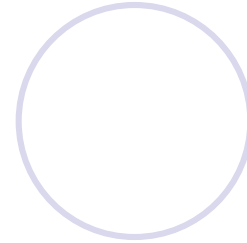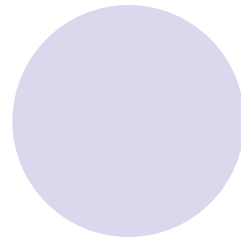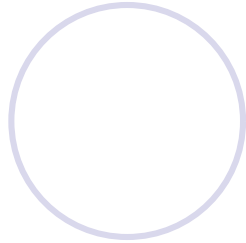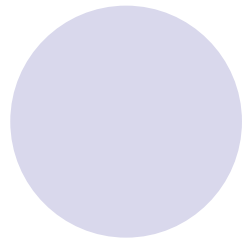
Galaxy Zoo

Fine-grained Recognition

APPS: IMAGE PROCESSING

GalaxyZoo: Zooniverse

GalaxyZoo: Zooniverse

Fine-Grained Recognition: Tohme

(a) Raw Google Street View (GSV) image

(b) Results of computer vision curb ramp detection (lighter red is higher confidence) — TP=8; FP=10; FN=0

(c) Results after crowdsourced verification — TP=8; FP=0; FN=0

K. Hara et al, "Tohme: Detecting Curb Ramps in Google Street View Using Crowdsourcing, Computer Vision, and Machine Learning," UIST 2014
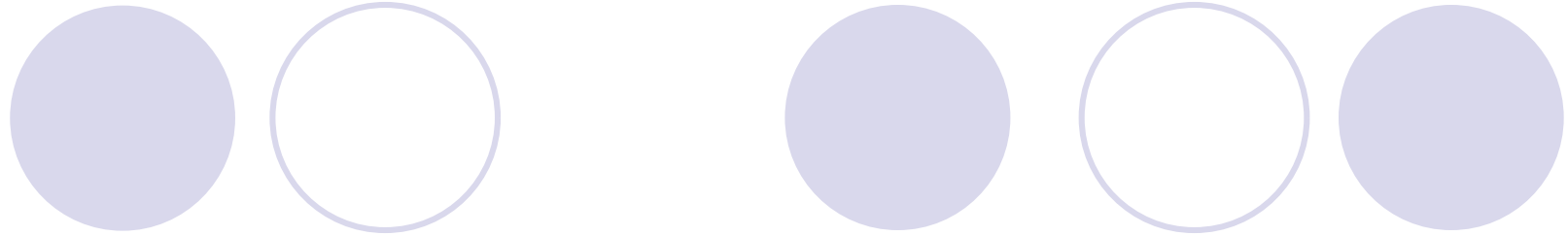
GWAP.com

reCAPTCHA

Crowdvoting (Crowdfunding and Crowdsearching)

APPS: COMMONSENSE KNOWLEDGE

GWAP.com: CMU
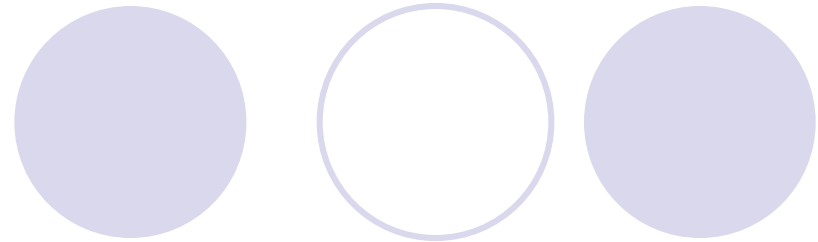
ESP Game
- Labeling images

Tag a Tune
- Labeling tunes

reCAPTCHA: CMU

Crowdvoting in Social Networks

- Online voting
- Example
  - WeChat group
  - Social recognition as incentive
- Others
  - Crowdfunding
  - Crowdsearching
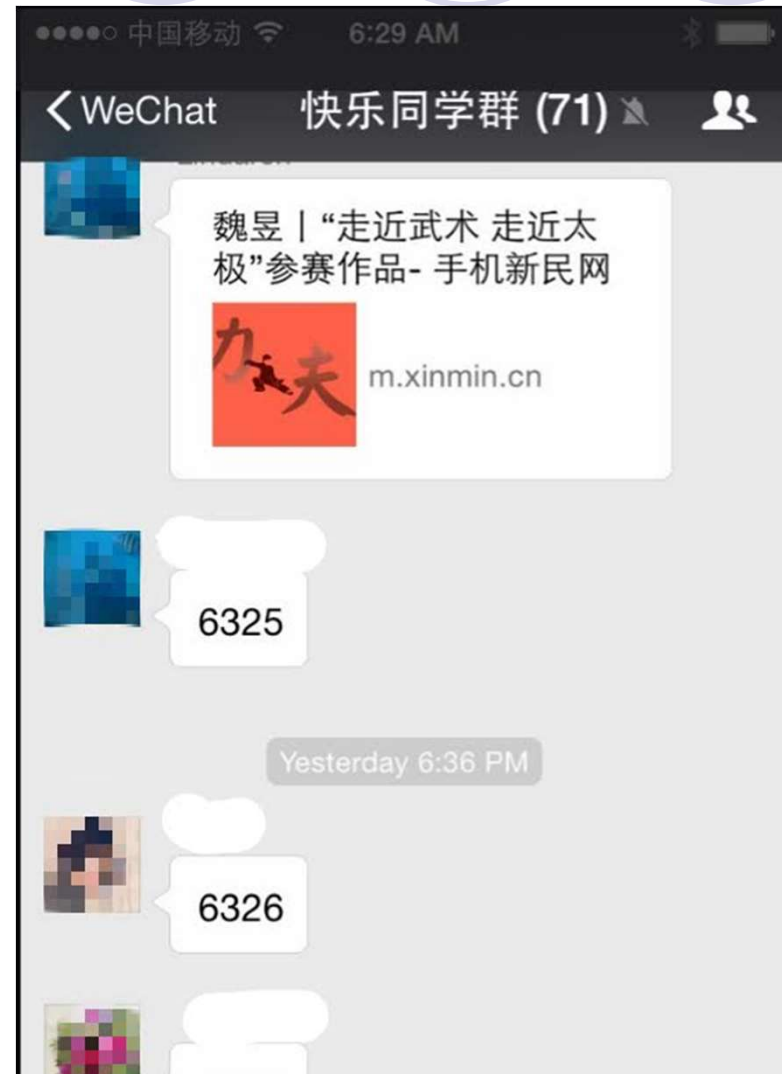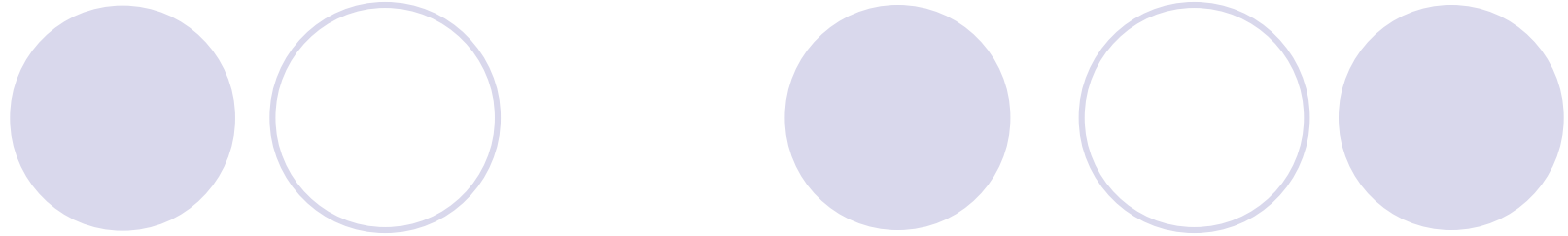
Software Engineering

Online feedback

Waze

Sparse Crowdsensing (Smart city)

APPS: LARGE PROJECTS
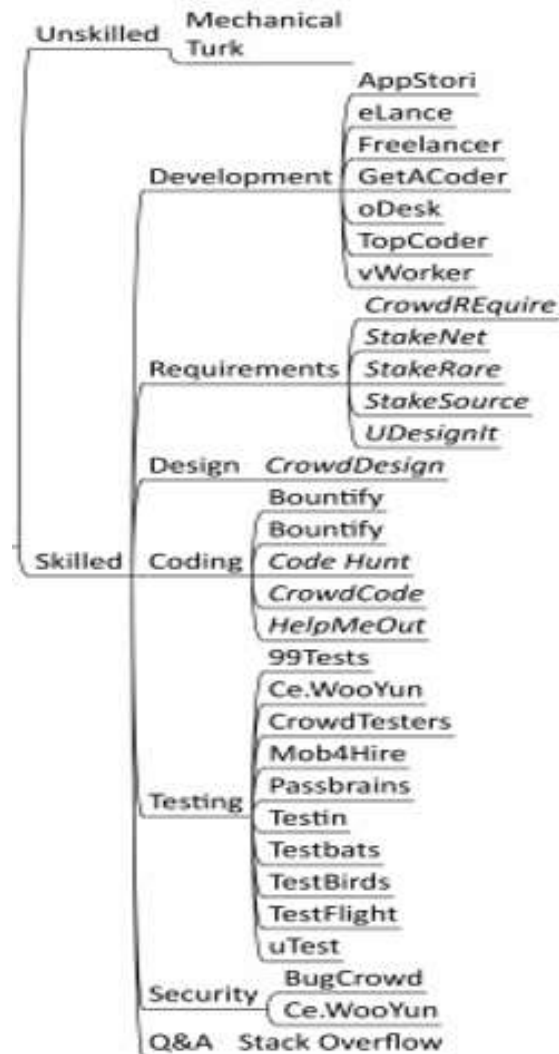
Software Engineering

- ## Issues/concerns

  - Planning, scheduling, coordination, motivation, intellectual property

- ## Some solutions

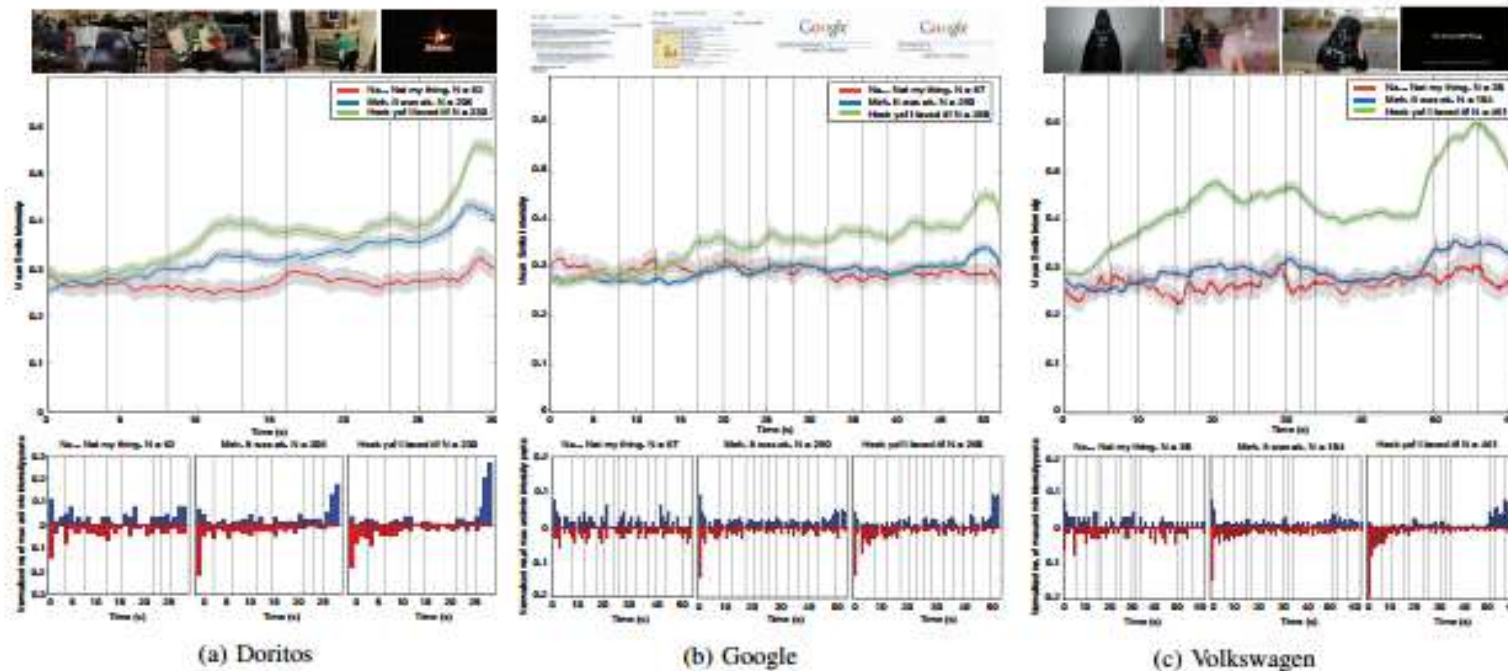 - Mturk and TopCoder use monetary rewards

Ke Mao et al., "A survey of the use of of crowdsourcing in software engineering," Journal of Systems and Software, 2017
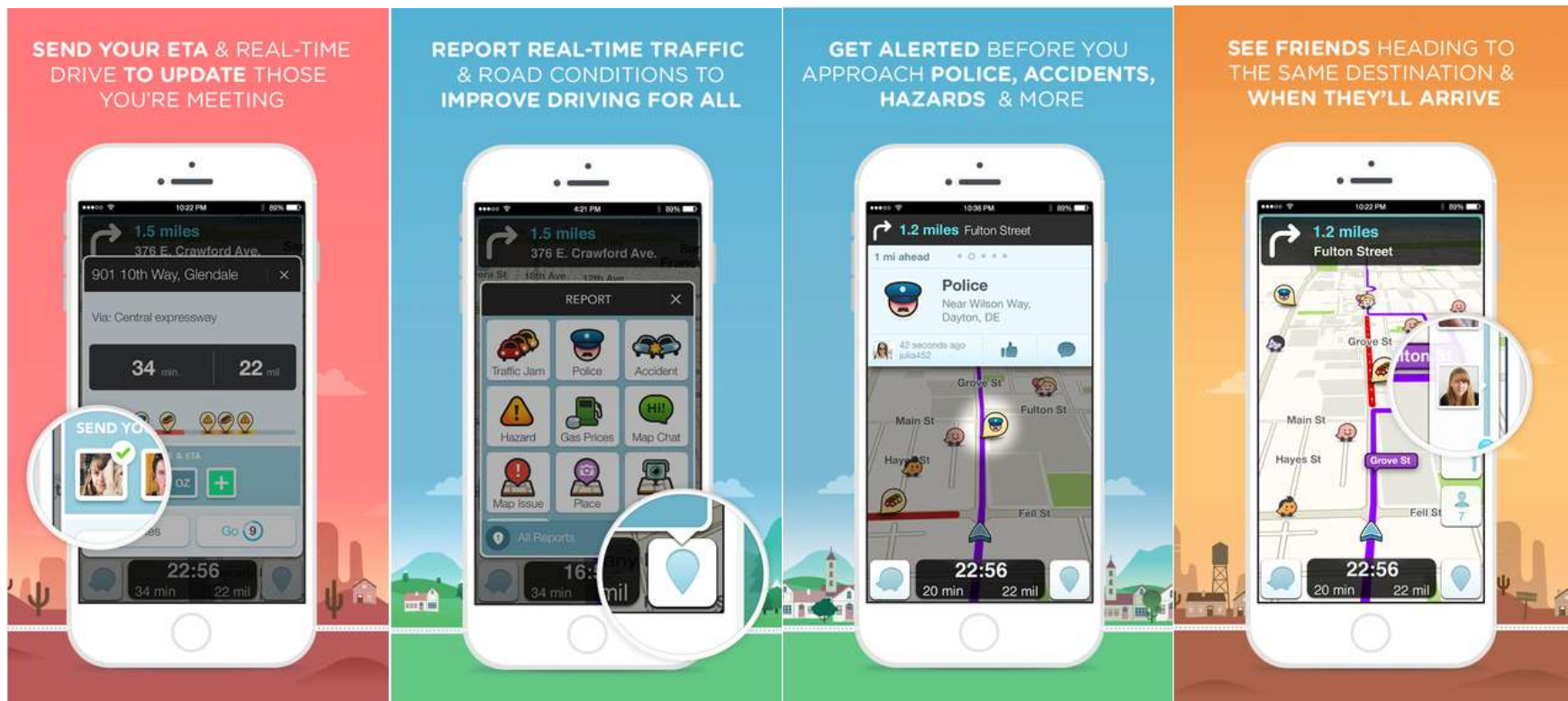
Online Feedback

- Crowdsourcing facial responses to online videos

(a) Doritos   (b) Google   (c) Volkswagen

D. McDuff et al., "Crowdsourcing facial responses to online videos," *Proc. of Int'l Conf. on Affective Computing and Intelligent Interaction (ACII)*, 2015.

Waze as Mobile Crowdsourcing

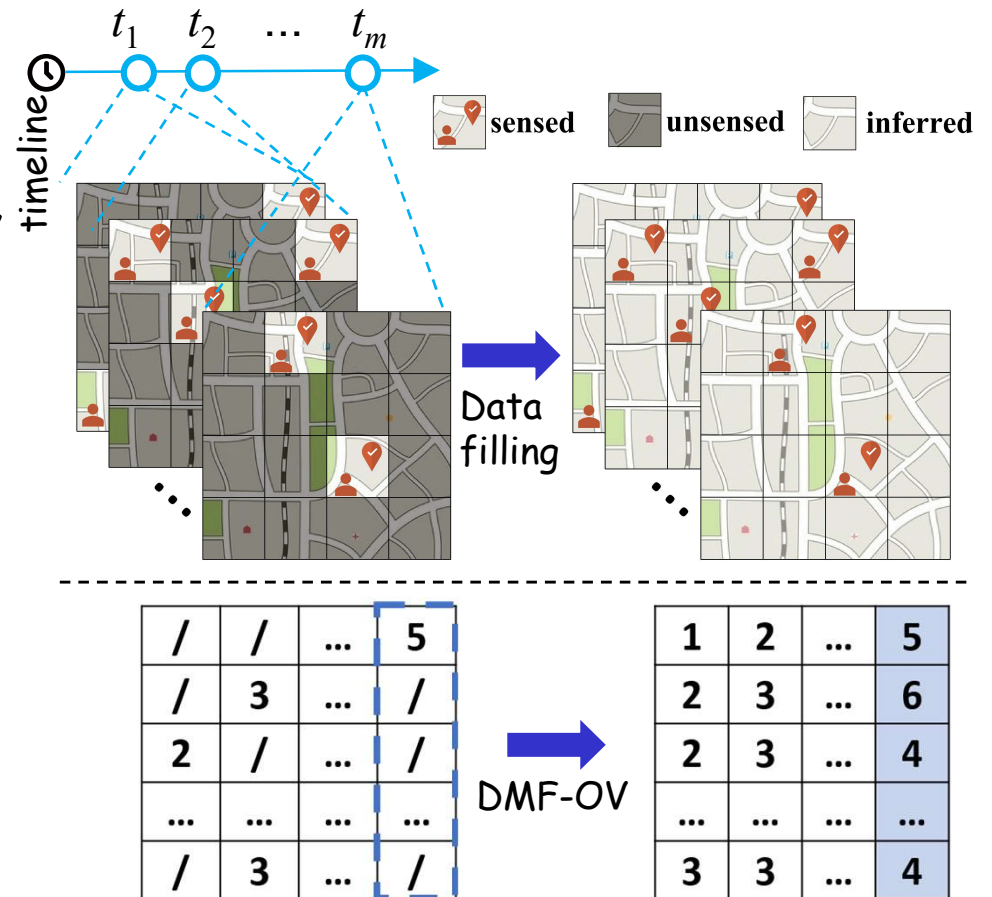- Waze is the world's largest community-based traffic and navigation app.

Data filling in Sparse Crowdsourcing
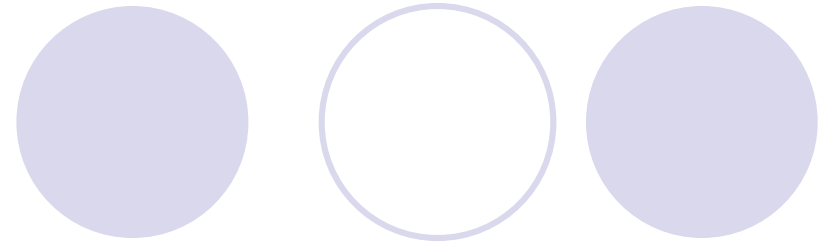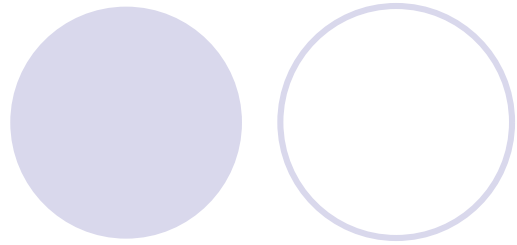
Data characteristics

- Non-linear and mobile
- Spatio-temporal similarity

Mathematical model

- Basic method

  Deep Matrix Factorization

- Refine the loss function

  Outlier Value Loss

- Add a memory network

  Attention Mechanism

E. Wang, M. Zhang, Y. Yang, Y. Xu, and J. Wu, "Exploiting Outlier Value Effects in Sparse Urban CrowdSensing," *IWQoS* 2021.
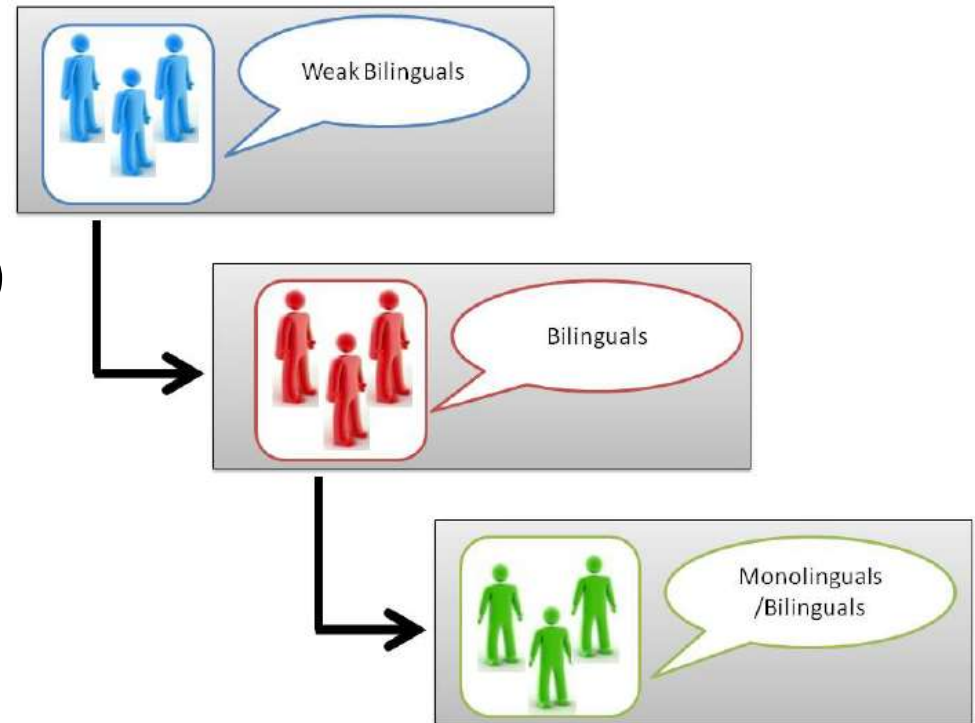
Sequential

Iterative and Parallel

Divide-and-Conquer and Aggregate

PARADIGMS

Sequential: Collaborative Workflow

- Lexical translation

  (weak bilinguals or machine)

- Assistive translation

  (strong bilinguals)

- Refine sentence

  (monolinguals)

V. Ambati et al., "Collaborative Workflow for Crowdsourcing Translation," CSCW 2012

Iterative and Parallel
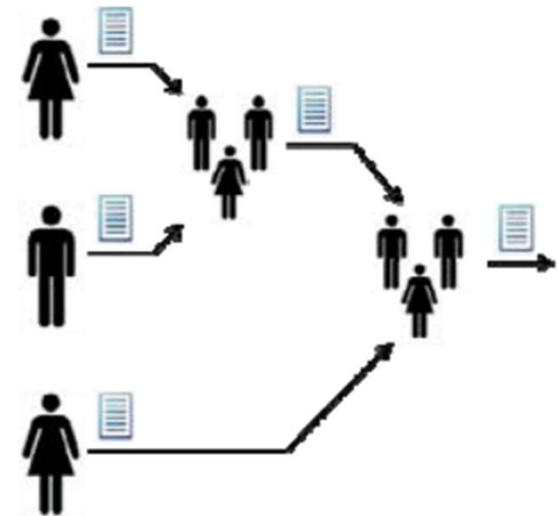
- Iterative improve and vote

G. Little et al, "Exploring Iterative and Parallel Human Computation Processes," HCOMP 2010

Divide-and-Conquer and Aggregate

- ## Divide-and-Conquer and Aggregate
  - Decompose a problem statement and aggregate the results

P → DC → P', P'' ...

Divide and Conquer

S ← A ← S', S'' ...
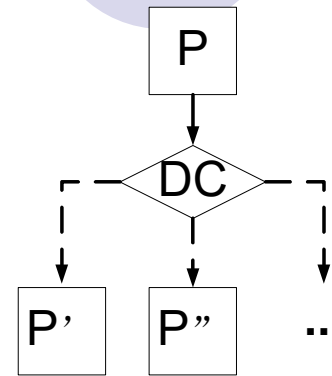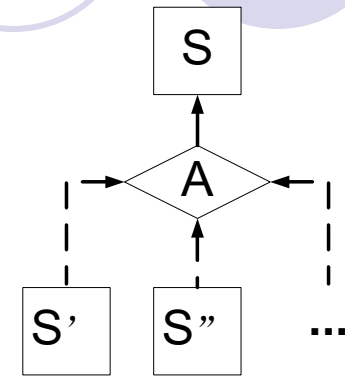
Aggregate

- ## Two special aggregates
  - Merge
  - Reduce

S', S'' ... → M → S

Merge

{S}, D → R → S
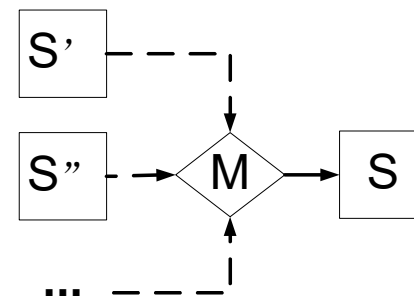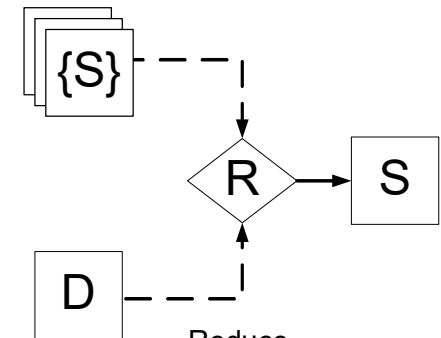
Reduce

P. Minder et al., "Crowdlang - First Steps Towards Programmable Human Computers for General Computation," AAAI 2011.

Challenges

Opportunities

CHALLENGES AND OPPORTUNITIES

Challenges

Each set has S/2 items

r workers | r

Each set has S/10 items

| r | r | r | r | r | r | r | r | r | r |

- **Trade-offs: time, cost, and quality**
  - Max algorithm with human error (with a probability)
  - Maximize quality (via redundancy) subject to cost and time

P. Venetis et al., "Max Algorithms in Crowdsourcing Environments," WWW 2012
R. Kawajiri et al., "Steered Crowdsensing: Incentive Design Towards Quality-oriented Place-centric Crowdsensing", UBICOMP 2014
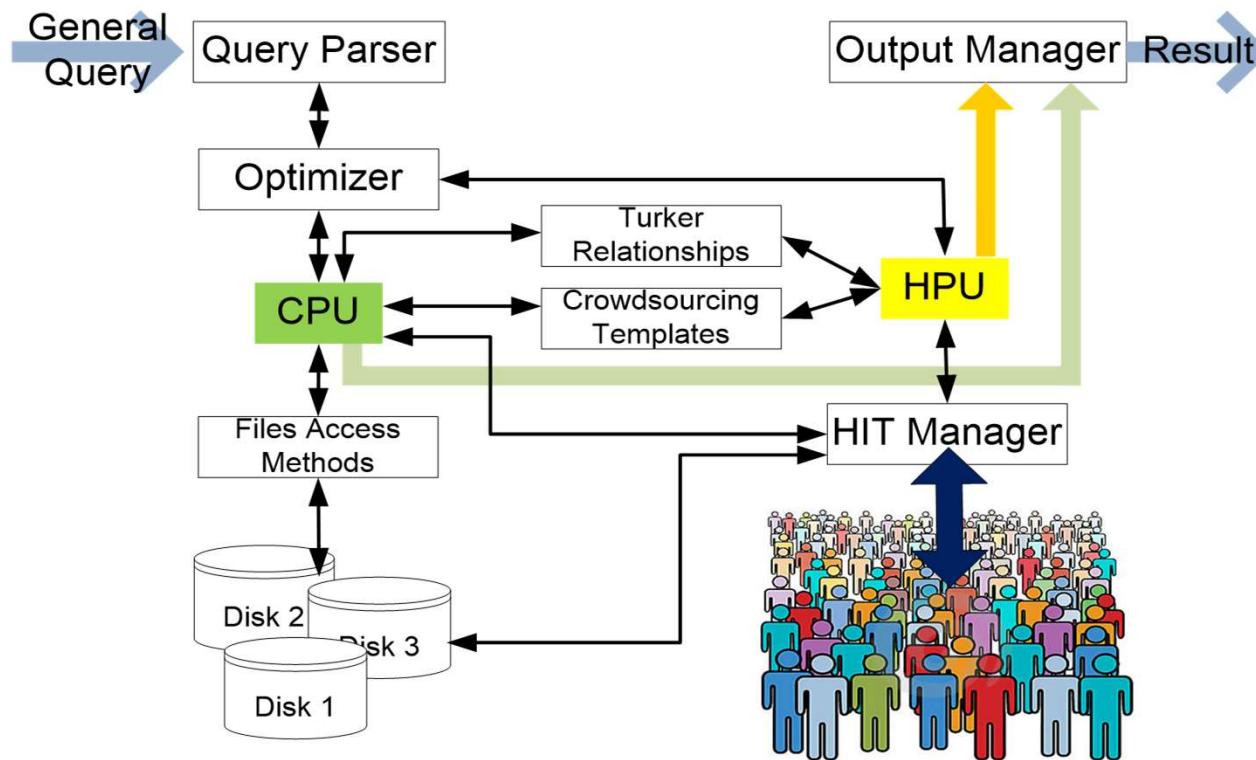
- **Incentive:** money, social recognition, and self esteem
  - Platform-centric: a Stackelberg game
  - User-centric: auction-based incentive mechanism

D. Yang et al., "Crowdsourcing to Smartphones: Incentive Mechanism Design for Mobile Phone Sensing," *MobiCom* 2012

Challenges: HPU + CPU

- CrowdDB:

M. Franklin et al., "CrowdDB: Answering Queries with Crowdsourcing," SIGMOD 2011

CPU-assisted HPU

- **Entity Matching**

  e.g. JHU matches

  Expert-free

  Rule pruning

  Apply Rules

- **Radiology**
  - Filter

    HPU:

    CPU:
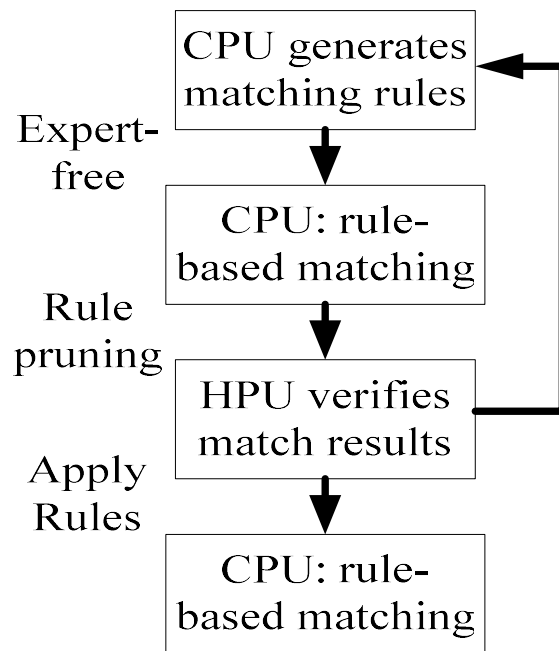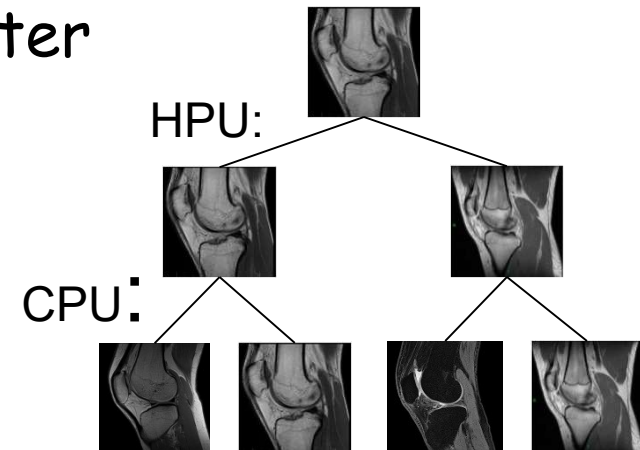
  - Diagnose

C. Gokhale et al., "Corleone: Hands-off Crowdsourcing for Entity Matching," *SIGMOD* 2014

Opportunities

- **Beyond simple workflows**
  - Graph search
  - Graph match

- **Beyond simple worker selection**
  - Dynamic procurement

- **Beyond independent workers**
  - Social networks

Beyond Simple Workflows

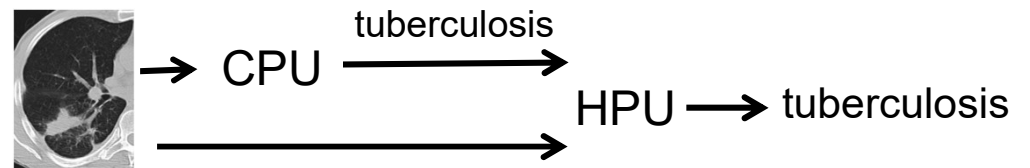Blend of bottom-up and open process with top-down organization goals

- Graph search (for macrowork)
  - Human-assisted graph search
  - Best sequence of questions with simple Y/N answers

  A. Parameswaran et al., "Human-Assisted Graph Search: It's Okay to Ask Questions," *VLDB 2010*

- Graph match
  - People graph (who knows and/or communicates with whom)
  - Puzzle graph (ideas are compatible and can merge)
  - Natural dynamic for people to merge their compatible ideas

  C. Brummitt et al., "Jigsaw Percolation: What Social Networks Can Collaboratively Solve a Puzzle," *The Annals of Applied Probability*, 2012

Beyond Simple Worker Selection

## Dynamic Procurement (multi-armed bandit)

- A gambler facing a row of slot machines

- Which one to play, how many times, and in which order

- Each machine having a random reward from a fixed distribution

- Objective: maximizing the sum of rewards earned through a sequence of lever pulls

Beyond independent workers

A. Badanidiyuru et al., "Bandits with Knapsacks: Dynamic Procurement for Crowdsourcing," JACM 2013

Crowdsourcing without Prior Knowledge

- Requesters are sensitive to the completion time

- Platform generally has a limited budget

- Platform does not know users' reliability in prior

- Mobile users are strategic about their cost
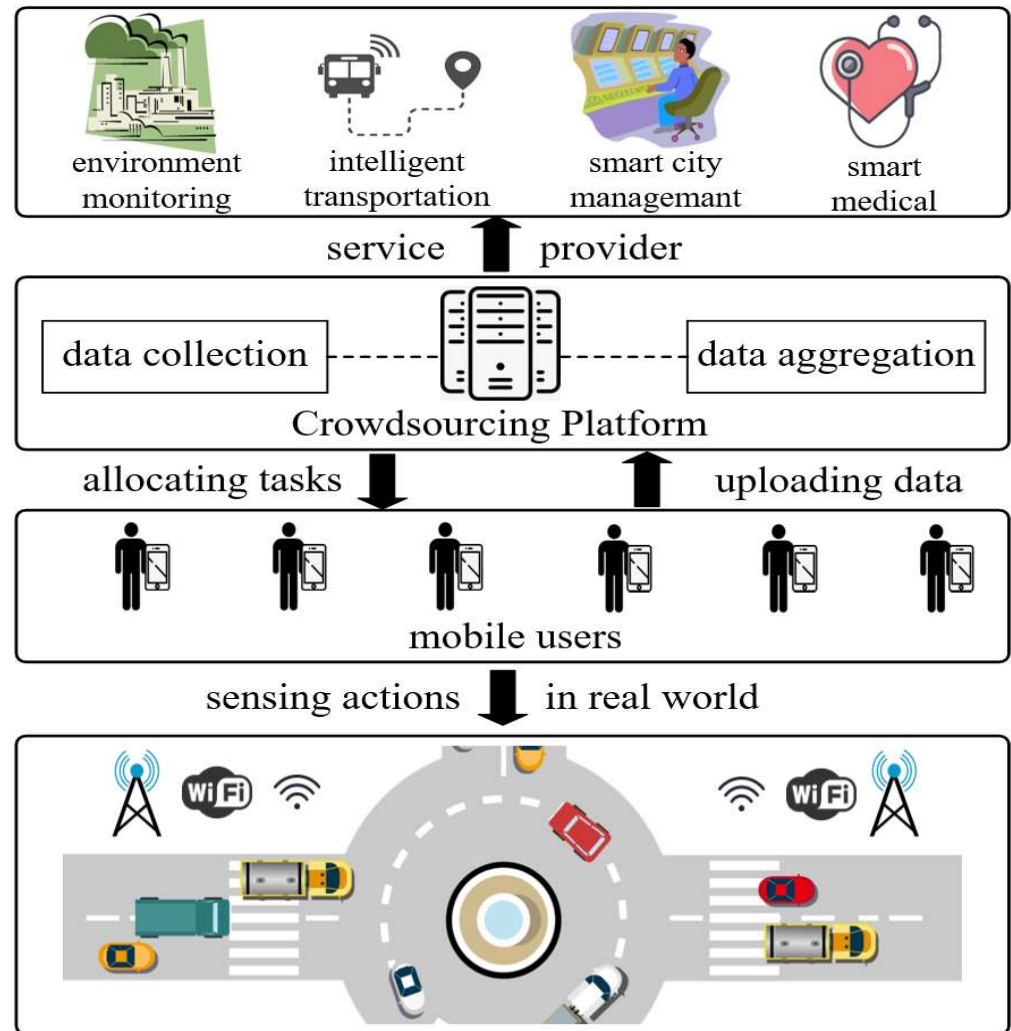
G. Gao, H. Huang, M. Xiao, J. Wu, Y. Sun, and S. Zhang, "Auction-Based Combinatorial Multi-Armed Bandit Mechanisms with Strategic Arms," *INFOCOM* 2021

Beyond Independent Workers
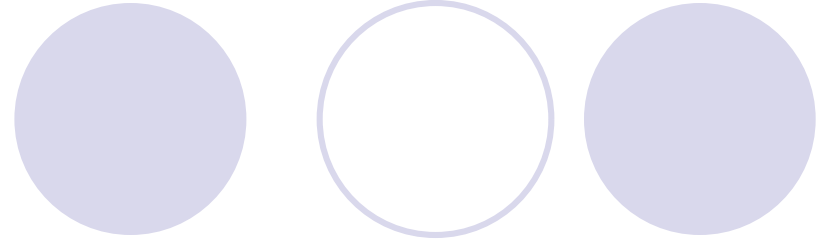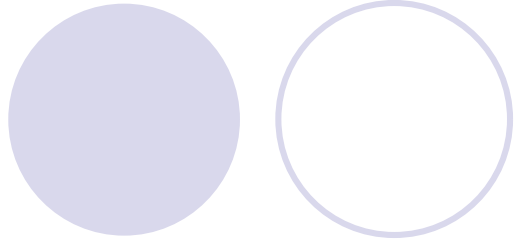
Social network of workers

Recruitment of workers through social ties

- ## Friends help friends
    - Response delay
    - Computation (by a friend)
    - Reply delay

- ## Water-filling

- d_i: response + reply
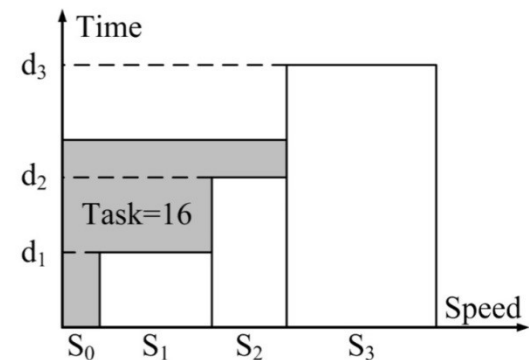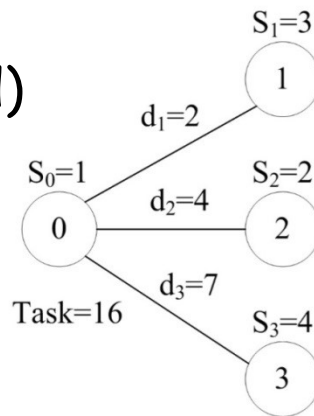
M. Xiao, J. Wu, L. Huang, Y. Wang, and C. Liu, "Multi-task Assignment for Crowdsensing in Mobile Social Networks," *INFOCOM* 2015

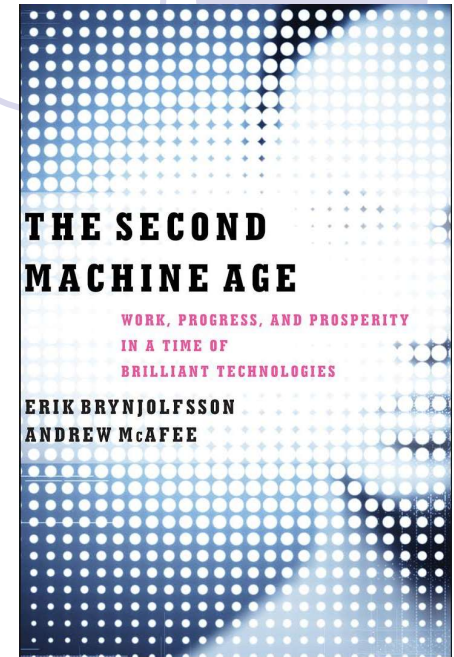Summary

Acknowledgements

CONCLUSION

Summary

- HPU as a new paradigm to compliment the traditional CPU-based computing for big data

- Many un(der)explored algorithmic problems
  - Social connections and proper training of workers
  - Workflow design
  - Cost-time-quality-uncertainty trade-offs
  - Incentive, gamification, and satisfaction mechanisms
  - Mobile crowdsourcing: energy consumption, communication cost, truthfulness, and privacy

Summary

Collective intelligence = networked brain + massively connected & intelligent machine

- Finding ways to race *with* the machine rather than racing *against* it

- MTurk as *"artificial AI"* as a way for people to race with machine

THE SECOND MACHINE AGE

WORK, PROGRESS, AND PROSPERITY IN A TIME OF BRILLIANT TECHNOLOGIES

ERIK BRYNJOLFSSON
ANDREW McAFEE

Acknowledgements

- Wei Chang

  Saint Joseph's University

- Grace Ju

  Carnegie Mellon University

- Wenjun Jiang

  Hunan University

- Ying Dai

  Linkedin