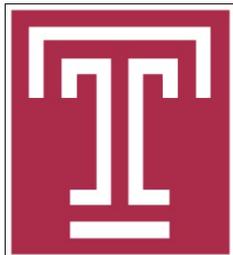


Algorithmic Crowdsourcing: Current State and Future Perspective

Jie Wu

Dept. of Computer and Info. Sciences

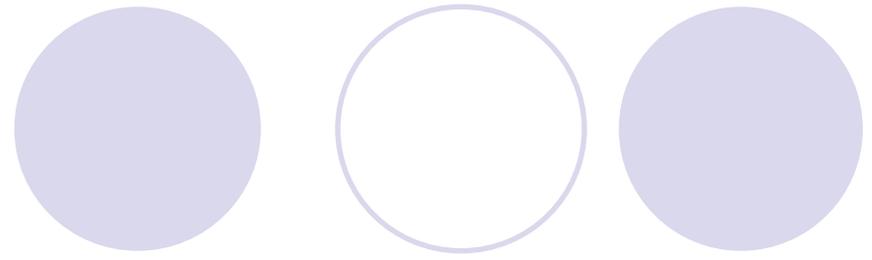
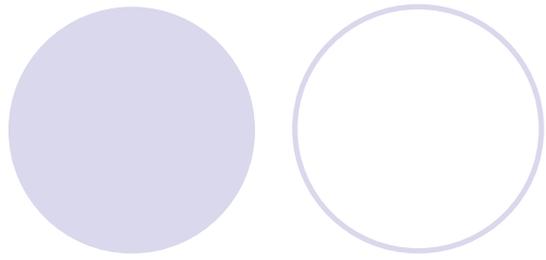
Temple University



Road Map

- Introduction and Motivation
- Mechanical Turk
- Applications
- Paradigms
- Challenges and Opportunities
- Social Crowdsourcing
- Conclusion





What

Why

Basic components

Motivation examples

INTRODUCTION

Big Data is Everywhere!

- Lots of data is being collected:

Volume, Variety, Velocity

- Web data, e-commerce
- Purchases
- Bank/credit card transactions
- Video and images
- Social networks



How Much Data?

- Google processes 100 PB a day
- Wayback Machine has 3 PB + 100 TB/month (3/2009)
- WeChat has 600 M users and 20 B message per day
- Facebook has 2.5 PB of user data + 15 TB/day (4/2009)
- eBay has 6.5 PB of user data + 50 TB/day (5/2009)



640K ought to be enough for anybody.

Big Data Era

- “In information technology, big data consists of datasets that grow so large that they become awkward to work with using on-hand database management tools.”
- Computers are not efficient in processing certain data (e.g., image processing)



What is Crowdsourcing?

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



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



Get Results from Mechanical Turk Workers

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

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



The Benefits of Crowdsourcing

- Performance
 - Inexpensive
 - Fast
- Human Processing Unit (HPU)
 - More effective than CPU (for some apps)
 - Image labeling
 - Language translation
 - Social network survey
 - ...

Basic Components

- Requester
 - People submit jobs
 - Human Intelligence Tasks (HITs)
- Worker
 - People work on jobs
- Platform
 - Job management
 - Amazon Mechanical Turk (MTurk)



Requester

amazon
mechanical turk
beta



Worker Pool

Malaysia Airlines Flight MH 370

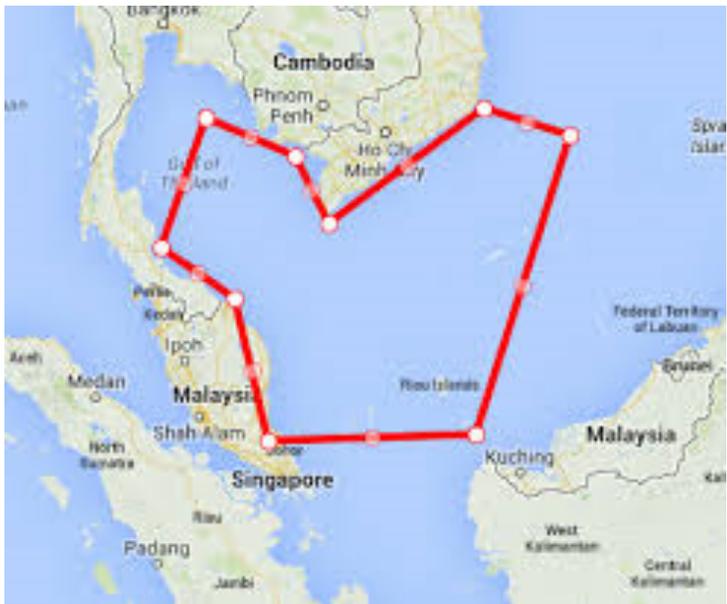


- DigitalGlobe

- Crowdsourcing volunteers comb satellite photos for Malaysia Airlines jet

- March 11 (from a 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.

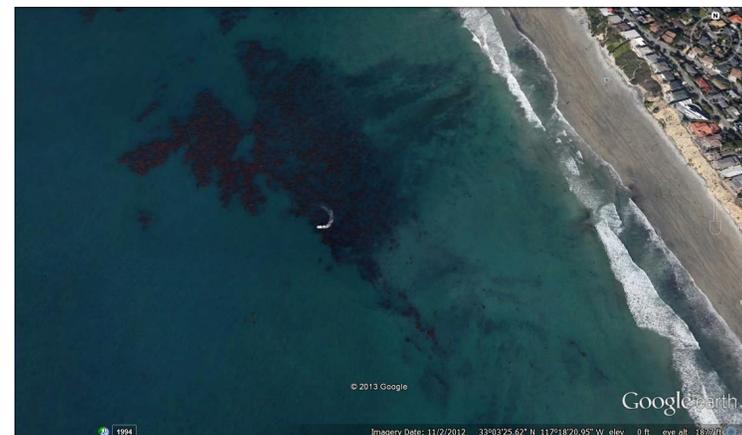
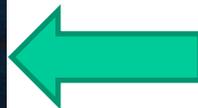


Help Find Jim Gray



- 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.



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

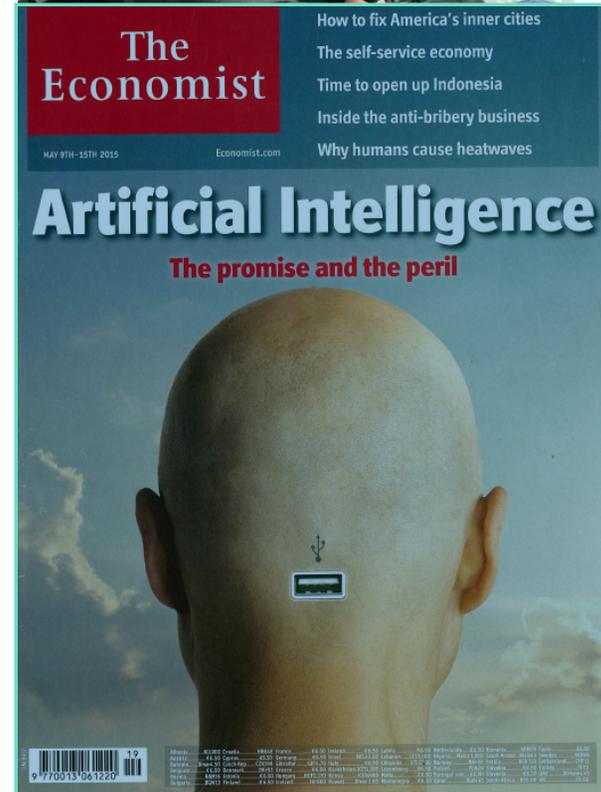
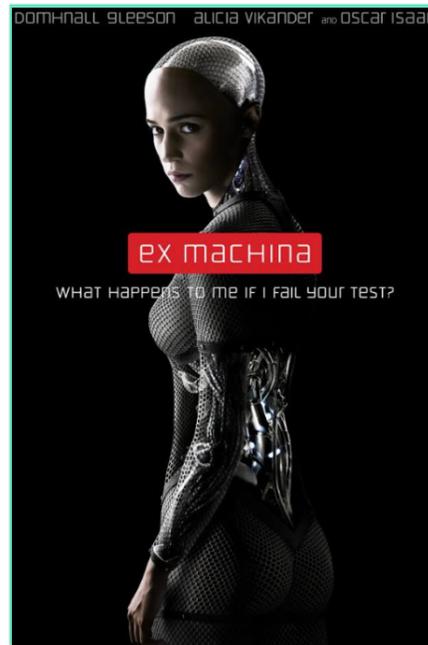
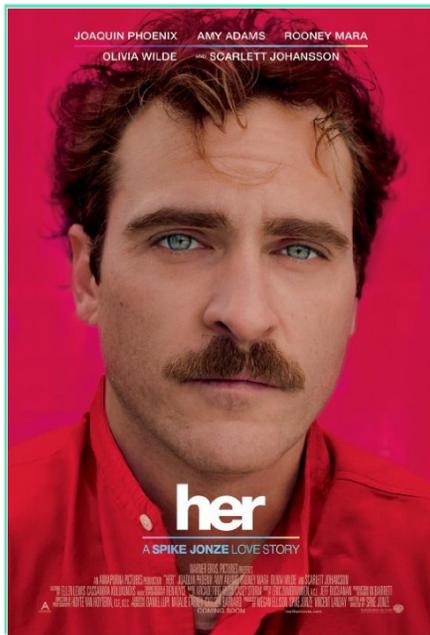


Bratislava



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)

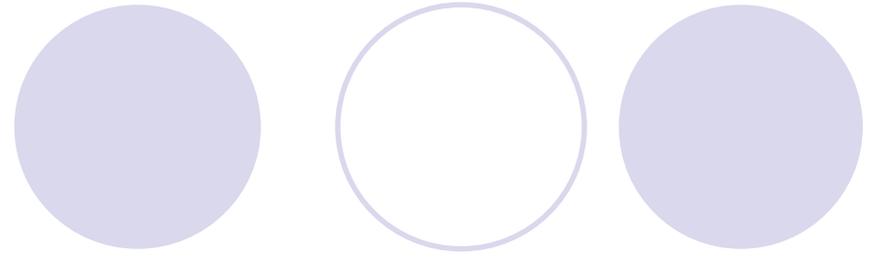
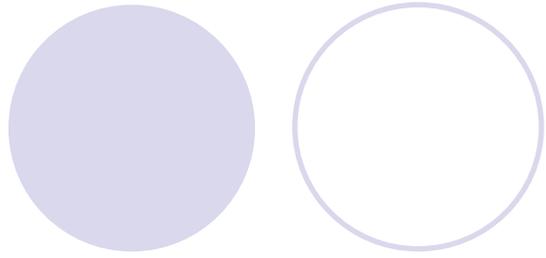


Smarter Than You Think



- 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



Worker

HIT

Dashboard

Requester

MECHANICAL TURK

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.

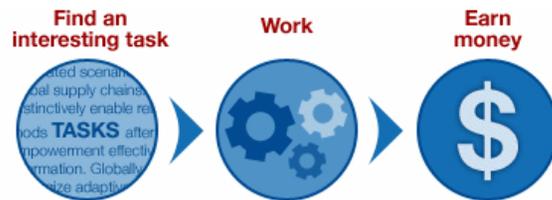
293,089 HITS available. [View them now.](#)

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 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



[Get Started](#)

- As a worker, make an average of \$0.03 per task
- Paid directly to Amazon account

- As requester, set up simple tasks for workers to complete
- Quality control is possible through MTurk services

Worker: Contract for a HIT

All HITS

1-10 of 1982 Results

Sort by:

Show all details | Hide all details

1 2 3 4 5 > Next >> Last

Requester	HIT Expiration Date	Time Allotted	Reward	HITS Available	Actions
Coupon Vision	Jun 21, 2014 (51 weeks 2 days)	10 minutes	\$0.08	14	View a HIT in this group
Brian Robertson	Jul 3, 2013 (6 days 23 hours)	2 hours	\$0.30	2	View a HIT in this group
Shopping	Jul 1, 2013 (4 days 23 hours)	45 minutes	\$0.03	2	View a HIT in this group
CrowdFlower	Jul 3, 2013 (6 days 23 hours)	30 minutes	\$0.01	219	View a HIT in this group
Redwood	Jun 26, 2014 (52 weeks)	15 minutes	\$0.02	11	View a HIT in this group

- Select a HIT
 - By creation date, payment amount, time allotment

Worker: Reviewing a HIT

Timer: 00:00:00 of 10 minutes

Want to work on this HIT?

Accept HIT

Want to see other HITs?

Skip HIT

Total Earned: \$4.72
Total HITs Submitted: 7

Copy Text from Business Card

Requester: Oscar Smith

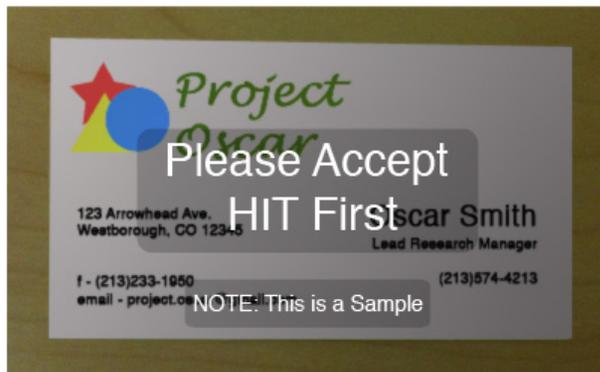
Qualifications Required: None

Reward: \$0.02 per HIT

HITs Available: 39

Duration: 10 minutes

Please Copy Text from Business Card:



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

Your Current Quality Score is:

If you have a high enough score, you will be [?](#) **--** considered for promotion to a Trusted Worker.

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](#)

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

Worker: Completing a HIT

Timer: 00:00:00 of 10 minutes

Want to work on this HIT?

Accept HIT

Want to see other HITs?

Skip HIT

Total Earned: \$4.72
Total HITs Submitted: 8

✔ Your results have been submitted to Oscar Smith and will be approved or rejected shortly.
You can work on this new HIT by clicking the "Accept HIT" button.

Copy Text from Business Card

Requester: Oscar Smith

Qualifications Required: None

Reward: \$0.02 per HIT

HITs Available: 3

Duration: 10 minutes

- Confirmation message in green
- Automatically shows the next HIT submitted by the same requester
- Check Dashboard to see if HIT is accepted

Worker: Sample Dashboard



Your Account

HITS

Qualifications

292,650 HITS available now

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

Find

containing

that pay at least \$

for which you are qualified
 require Master Qualification

Dashboard - (If you're not , [click here.](#))

Your Worker ID:

Total Earnings [\(What's this?\)](#)

Rewards You Have Earned

	Value
Approved HITS	\$4.72
Bonuses	\$0.00
Total Earnings	\$4.72

Your HIT Status [\(What's this?\)](#)

Date	Submitted	Approved	Rejected	Pending	Earnings
Today	1	0	0	1	\$0.00
Jun 3, 2013	7	7	0	0	\$4.72

[View more...](#)

HIT Totals [\(What's this?\)](#)

HITS You Have Accepted	Value	Rate	HITS You Have Submitted	Value	Rate
HITS Accepted	9	—	HITS Submitted	8	—
... Submitted	8	88.9%	... Approved	7	100.0%
... Returned	1	11.1%	... Rejected	0	0.0%
... Abandoned	0	0.0%	... Pending	1	—

Avoid Shady Requester

How Turkopticon works:

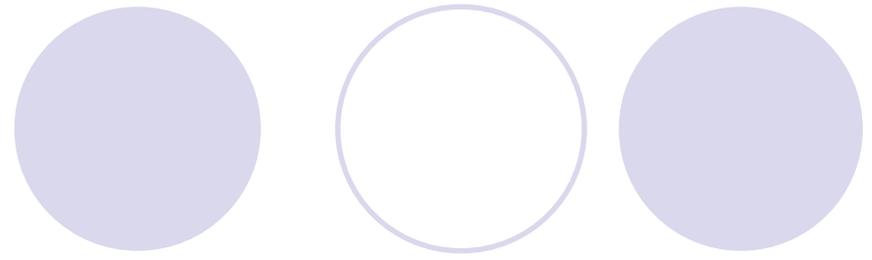
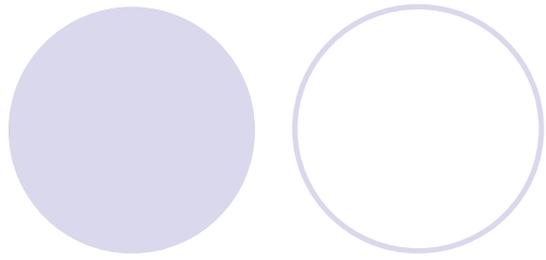
Turkopticon adds functionality to Amazon Mechanical Turk as you browse for HITs and review status of work you've done. As you browse HITs, Turkopticon places a button next to each requester and highlights requesters for whom there are reviews from other workers. Bad reviews let you avoid shady employers and good reviews help you find fair ones. You can view reports made against requesters with a quick click.

The screenshot shows a requester's profile with a dropdown menu open. The requester is 'Product Search' and the HIT is 'Product Search'. The dropdown menu displays the following metrics:

Metric	Score	Max Score
communicativity	1.00	5
generosity	2.57	5
fairness	2.86	5
promptness	2.00	5

Below the metrics, there is a link 'What do these scores mean?' and a note 'Scores based on 7 reviews'. A button 'Report your experience with this requester »' is also visible.

As you review HITs you've completed, are there HITs you weren't fairly paid for? Turkopticon adds a button that lets you review requesters from your "Status Detail" page.



EteRNA

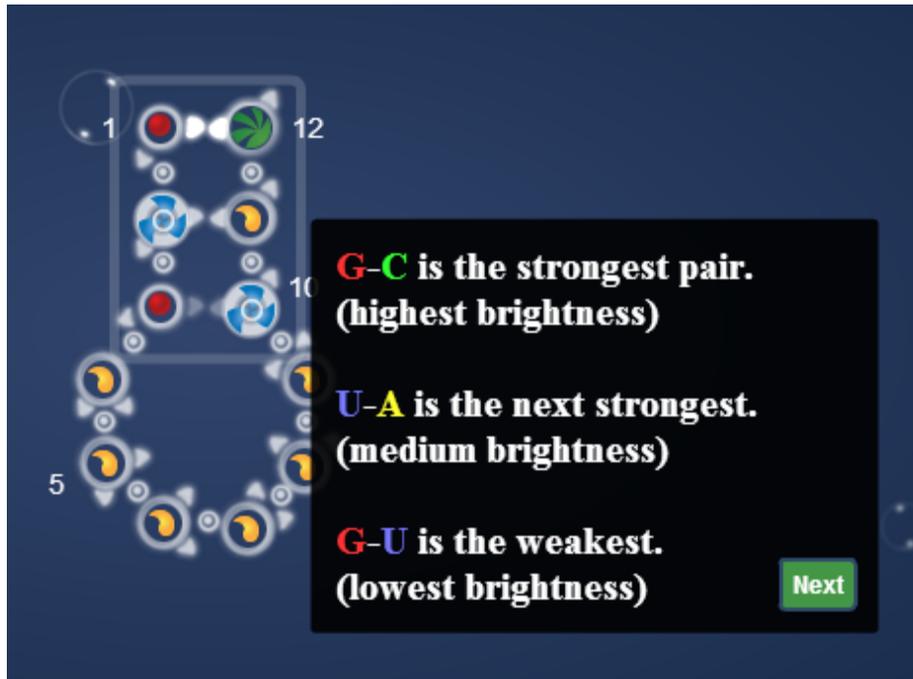
Galaxy Zoo

Fine-grained Recognition

APPS: IMAGE PROCESSING

Biology: EteRNA: CMU, Stanford

- Aim: to gain mastery over the way RNA molecules folds.

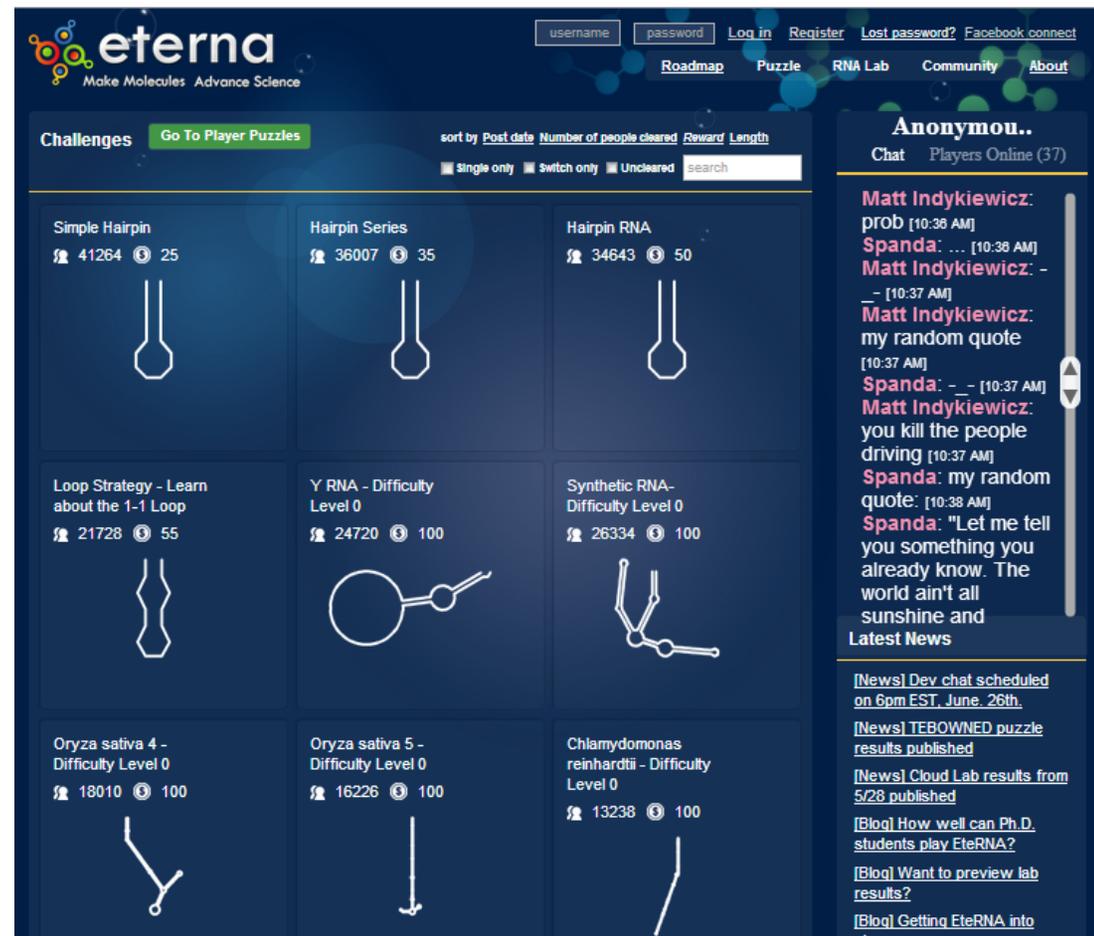


G-C is the strongest pair. (highest brightness)

U-A is the next strongest. (medium brightness)

G-U is the weakest. (lowest brightness)

Next



eterna
Make Molecules Advance Science

username password Log in Register Lost password? Facebook connect

Roadmap Puzzle RNA Lab Community About

Challenges Go To Player Puzzles

sort by Post date Number of people cleared Reward Length

Single only Switch only Uncleared search

Challenge Name	Difficulty	Number of people cleared	Reward
Simple Hairpin	25	41264	25
Hairpin Series	35	36007	35
Hairpin RNA	50	34643	50
Loop Strategy - Learn about the 1-1 Loop	55	21728	55
Y RNA - Difficulty Level 0	100	24720	100
Synthetic RNA- Difficulty Level 0	100	26334	100
Oryza sativa 4 - Difficulty Level 0	100	18010	100
Oryza sativa 5 - Difficulty Level 0	100	16226	100
Chlamydomonas reinhardtii - Difficulty Level 0	100	13238	100

Anonymou..
Chat Players Online (37)

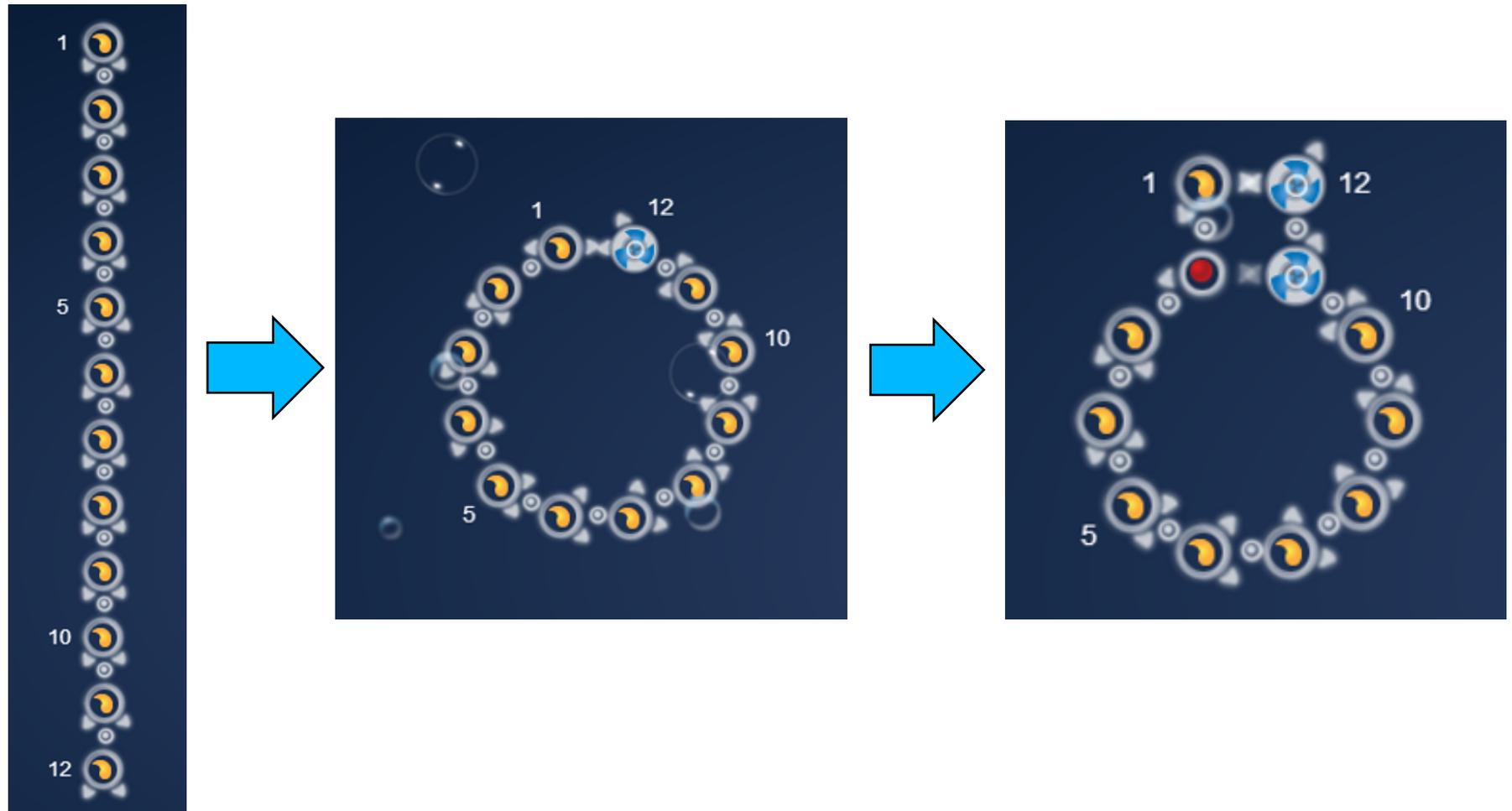
Matt Indykiewicz: prob [10:38 AM]
Spanda: ... [10:38 AM]
Matt Indykiewicz: - - [10:37 AM]
Matt Indykiewicz: my random quote [10:37 AM]
Spanda: - - [10:37 AM]
Spanda: my random quote: [10:38 AM]
Spanda: "Let me tell you something you already know. The world ain't all sunshine and

Latest News

[News] Dev chat scheduled on 6pm EST, June, 26th.
[News] TEBOWNED puzzle results published
[News] Cloud Lab results from 5/28 published
[Blog] How well can Ph.D. students play EteRNA?
[Blog] Want to preview lab results?
[Blog] Getting EteRNA into classroom

EteRNA: CMU, Stanford

- By assigning different colors (RNA nucleotides), a RNA chain will fold into different structure



Galaxy Zoo: Zooniverse

The screenshot displays the Galaxy Zoo Zooniverse interface. At the top, a navigation bar includes links for CLASSIFY, STORY, SCIENCE, DISCUSS, PROFILE, and LANGUAGE. The 'GALAXY ZOO' logo is prominently displayed in the center of the navigation bar. Below the navigation bar, the main content area is divided into two sections. On the left, a large image shows a galaxy with a bright blue core and a diffuse, rounded structure. On the right, the 'Classify' section contains several interactive elements: a 'UKIDSS' icon, an 'Invert' icon, and two buttons labeled 'Examples' and 'Restart'. Below these elements, the 'SHAPE' section asks the user: 'Is the galaxy simply smooth and rounded, with no sign of a disk?'. At the bottom of the interface, three options are presented: 'Smooth' (represented by a white circle), 'Features or disk' (represented by a spiral galaxy icon), and 'Star or artifact' (represented by a star icon).

CLASSIFY STORY SCIENCE **GALAXY ZOO** DISCUSS PROFILE LANGUAGE

Classify   [Examples](#) [Restart](#)

UKIDSS Invert

SHAPE
Is the galaxy simply smooth and rounded, with no sign of a disk?

Smooth Features or disk Star or artifact

Galaxy Zoo: Zooniverse

CLASSIFY STORY SCIENCE **GALAXY ZOO** DISCUSS PROFILE LANGUAGE

Classify  
UKIDSS Invert

Examples Restart

ROUND
How rounded is it?

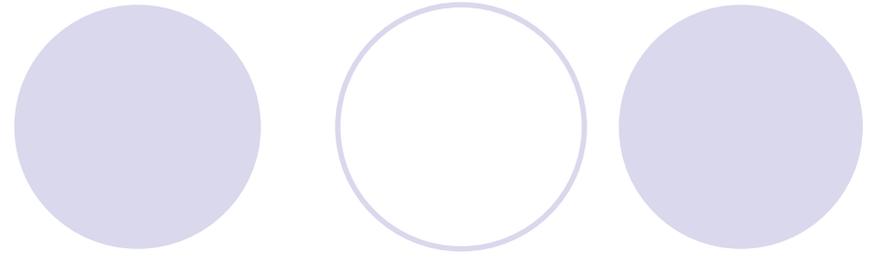
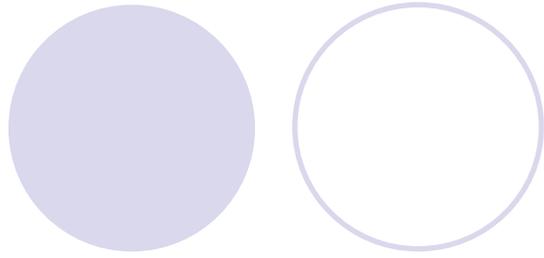
  

Completely round In between Cigar shaped

Fine-Grained Recognition: Tohme



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



GWAP.com

OnToGalaxy

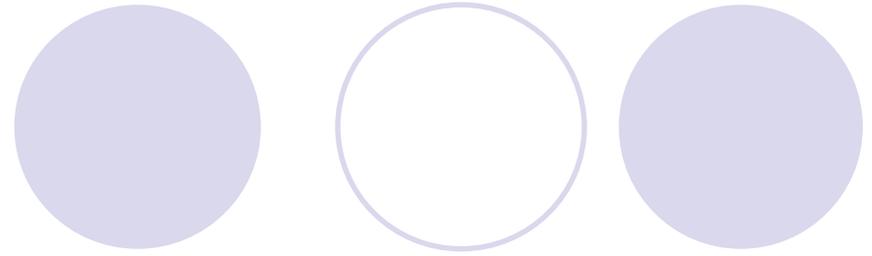
reCAPTCHA

ChaCha

Crowdsourcing

APPS: COMMONSENSE KNOWLEDGE

GWAP.com: CMU



ESP Game

- Labeling images

ESP Game
Concentrate...

How to Play

- 1 You and a partner see the same image.
- 2 Each of you must guess what words your partner is typing.

Got it, Let's Play!

View Top Scores

Tag a Tune

- Labeling tunes

Tag a Tune
Hear Here

How to Play

- 1 You and a partner hear a tune and must describe it.
- 2 Based on the descriptions, you have to figure out if you're both listening to the same tune!
- 3 There are several other bonus rounds which are self descriptive. Enjoy!

Got it, Let's Play!

View Top Scores

OnToGalaxy: University of Bremen

- Given a keyword
 - e.g., "tourism"
- Collect pods with words related to keyword
 - e.g., "voyage"
- Shoot down pods with unrelated words
 - e.g., "resist"
- An experimental game platform



reCAPTCHA: CMU



- WHAT IS reCAPTCHA
- GET reCAPTCHA
- PROTECT YOUR EMAIL
- MY ACCOUNT
- RESOURCES: DOCS & PLUGINS

reCAPTCHA IS A FREE ANTI-BOT SERVICE THAT HELPS DIGITIZE BOOKS.

steamboat train, from New
this **morning** ran off the track
New-London. Four cars plunge



→ [LEARN HOW reCAPTCHA WORKS](#)

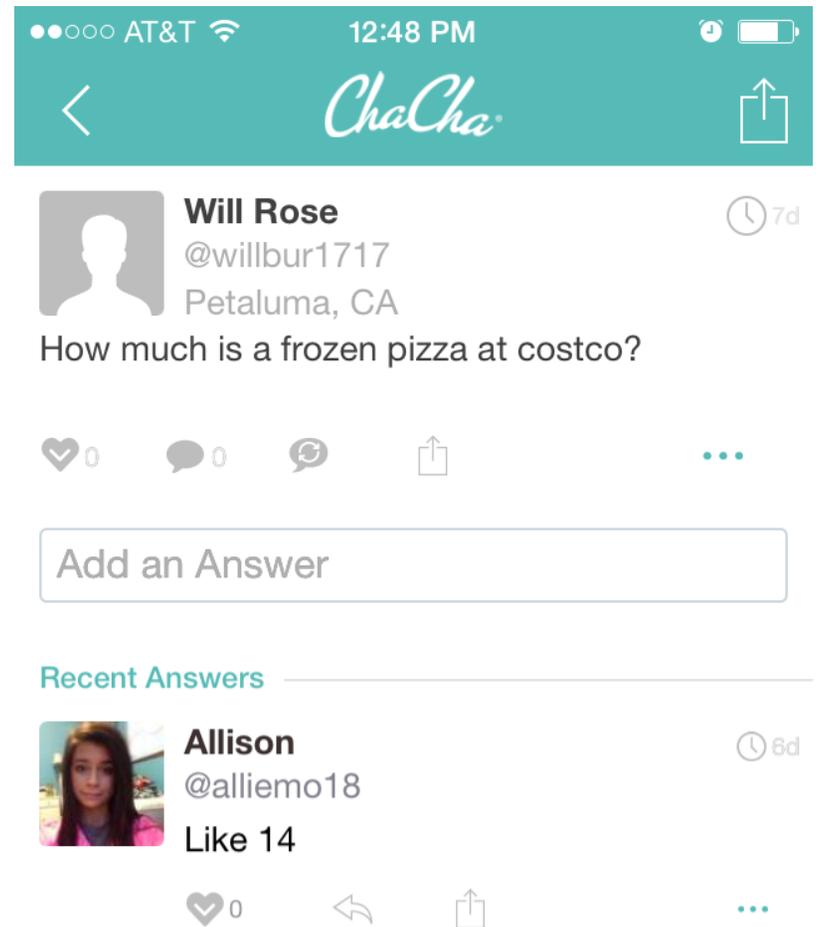
[USE reCAPTCHA ON YOUR SITE](#)

- STRONG SECURITY**
- ACCESSIBLE TO BLIND USERS**
- 30+ MILLION SERVED DAILY**

NEW See how accurate reCAPTCHA is at digitizing content!

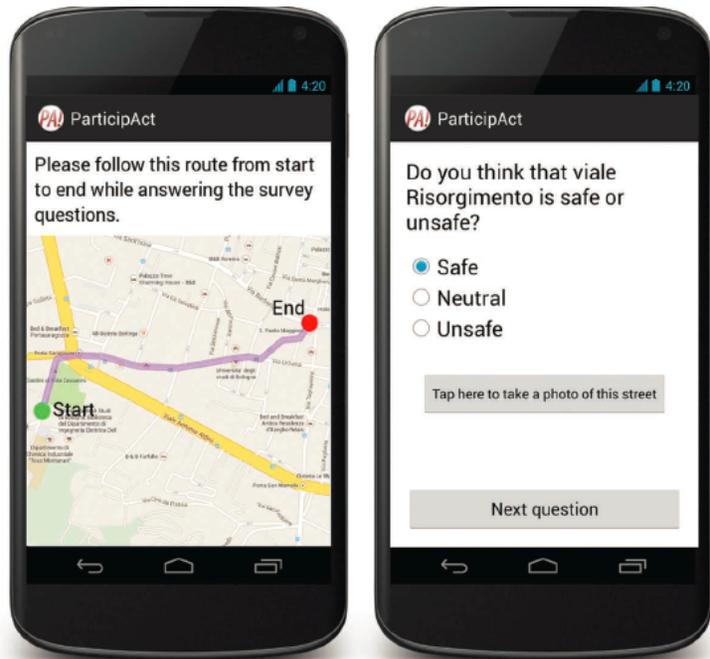
ChaCha Search Engine: powered by people

- ChaCha is a human-guided search engine.
- Questions are answered real-time in an “ask a smart friend” format.
- One can access ChaCha via its website, text message, or mobile apps.



Crowdsensing: Smart City

Traveling safety



G. Cardone et al, "The ParticipAct Mobile Crowd Sensing Living lab: The Testbed for Smart Cities," IEEE Communications Magazine 2014

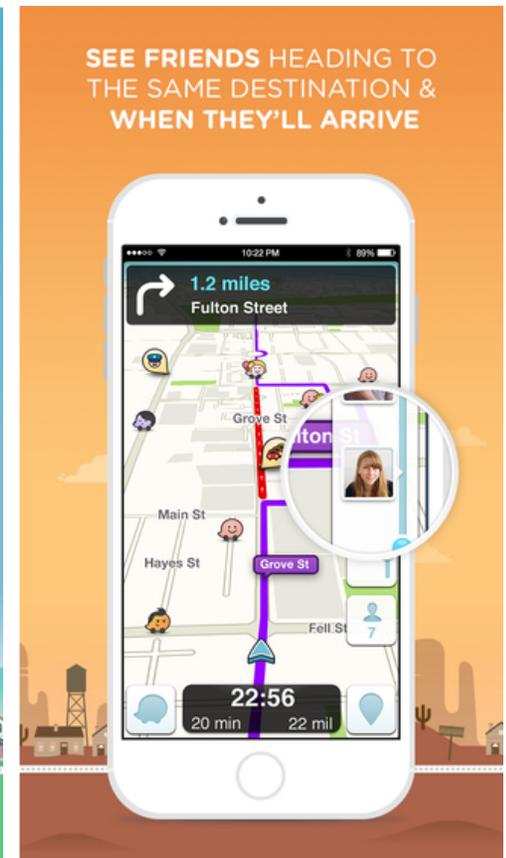
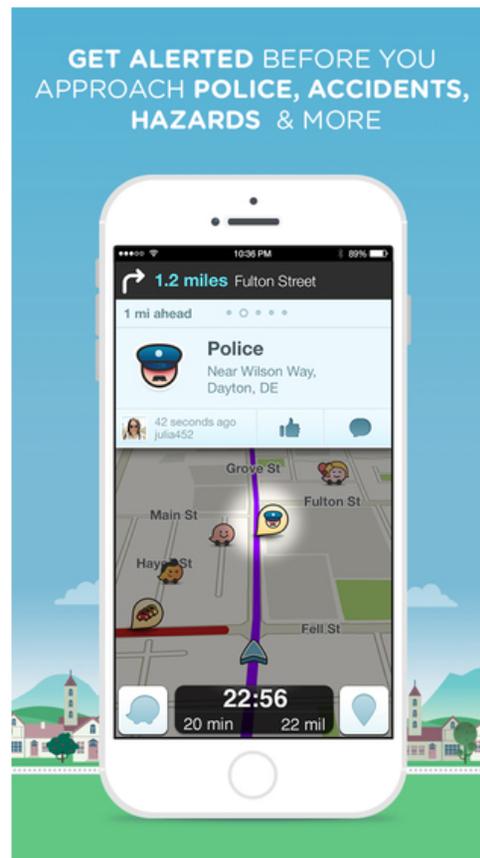
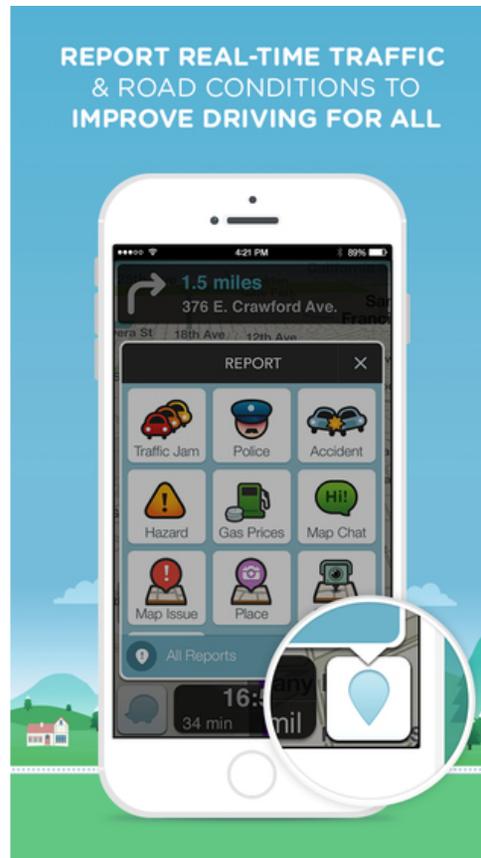
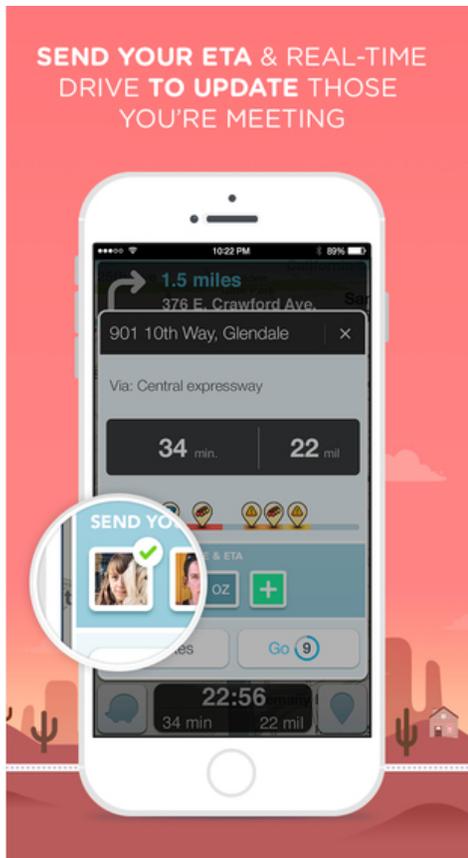
Activity census



R. Vaish et al, "Twitch Crowdsourcing: Crowd Contributions in Short Bursts of Time," CHI 2014

Waze Social GPS, Maps & Traffic

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



Smartphone-based Crowdsensing

● Smart city

○ Personalized recommendation

P. Organisciak et al, "A Crowd of Your Own: Crowdsourcing for On-demand Personalization," AAAI 2014

○ Public transportation

Z. He et al, "High Quality Participant Recruitment in Vehicle-based Crowdsourcing using Predicable Mobility," INFOCOM 2015

○ Indoor map construction

X. Guo et al, "ShopProfiler: Profiling Shops with Crowdsourcing Data," INFOCOM 2014

○ Speaker counting

C. Xu et al, "Crowdsensing the Speaker Count in The Wild: Implications and Applications," IEEE communications magazine 2014

● Some challenges

○ GPS-less (energy efficient)

X. Sheng et al, "Leveraging GPS-less sensing Scheduling for Green Mobile Crowd Sensing," JIoT 2014

○ Trustfulness & Game

Z. Feng et al, "TRAC: Truthful Auction for Location-aware Collaborative Sensing in Mobile Crowdsourcing," INFOCOM 2014

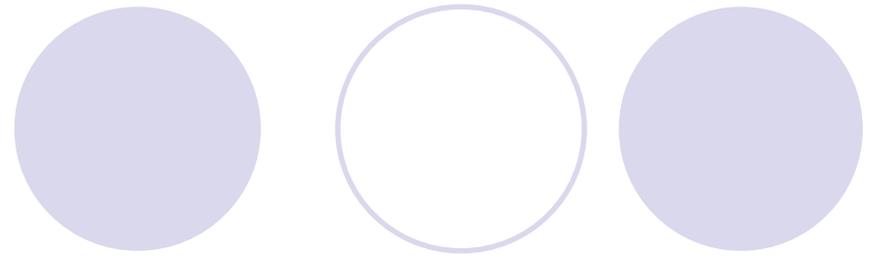
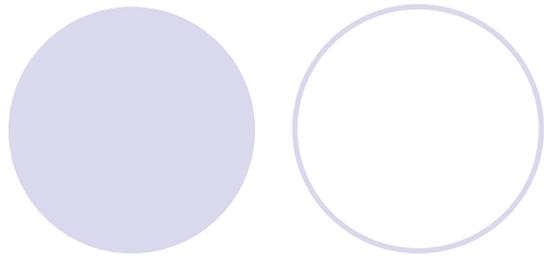
T. Luo et al, "Crowdsourcing with Tullock Contests: a New Perspective," INFOCOM 2015

○ Coverage

M. Karliopoulos et al, "User Recruitment for Mobile Crowdsensing over Opportunistic Networks," INFOCOM 2015

○ Privacy

L. Pournajaf et al, "Spatial Task Assignment for Crowd Sensing with Cloaked Locations," MDM 2014



Sequential

Iterative and Parallel

Divide-and-Conquer and Aggregate

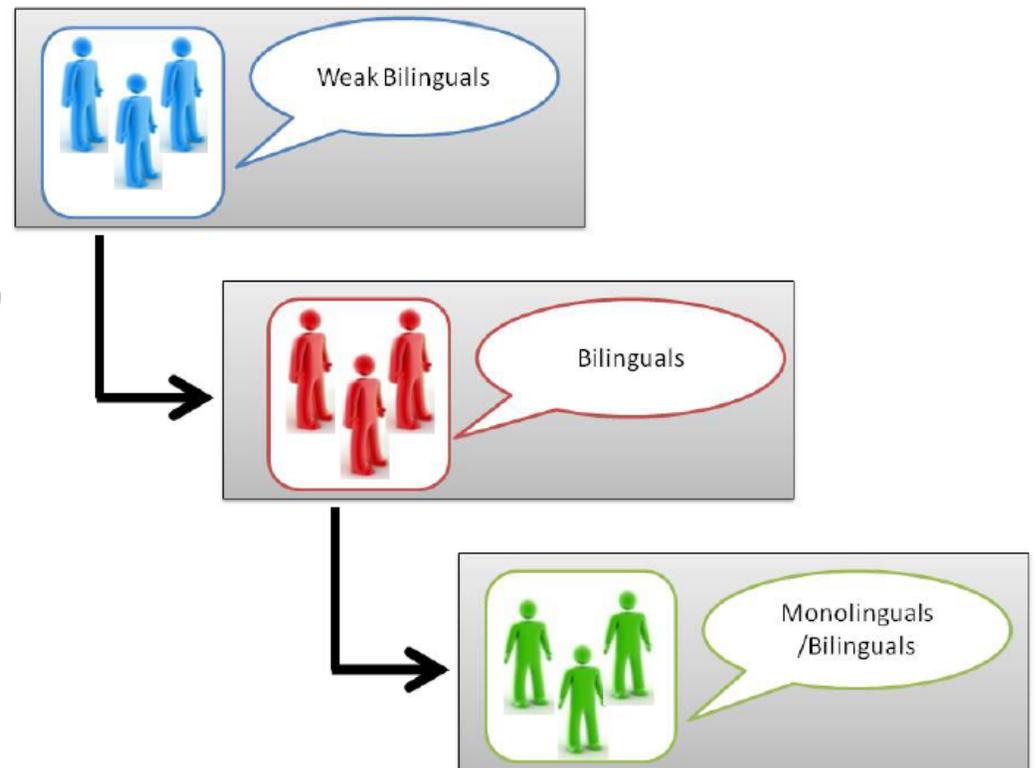
Map and Reduce: a Special Case

Publish/Subscribe

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

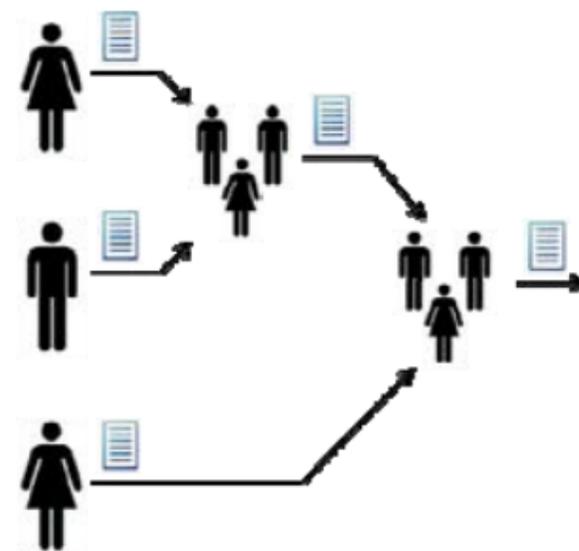
- Iterative improve and vote

The screenshot shows two browser windows side-by-side. The left window, titled 'Handwriting Recognition Task - Mozilla Firefox', displays a handwritten document and instructions: 'Please improve the transcription of this handwriting. People will vote whether to approve your changes.' Below the document, a feedback message reads: 'You (7) (7) (7) (work). (7) (7) (7) work (not) (time). I (7) (7) a few grammatical mistakes. Overall your writing style is a bit too (phoney). You do (7) have good (points), but they got lost amidst the (writing). (signature)'. A 'Submit' button is at the bottom right.

The right window, titled 'MTurk Task - Mozilla Firefox', displays the same handwritten document and instructions: 'Please select the better transcription for this handwriting. Differences are highlighted in yellow.' Below the document, two options are presented for selection:

- > You (misspelled) (several) (words) (work). (7) (7) (7) work (not) (time). I also notice a few grammatical mistakes. Overall your writing style is a bit too (phoney). You do (7) have good (points), but they got lost amidst the (writing). (signature)
- > You (7) (7) (7) (work). (7) (7) (7) work (not) (time). I (7) (7) a few grammatical mistakes. Overall your writing style is a bit too (phoney). You do (7) have good (points), but they got lost amidst the (writing). (signature)

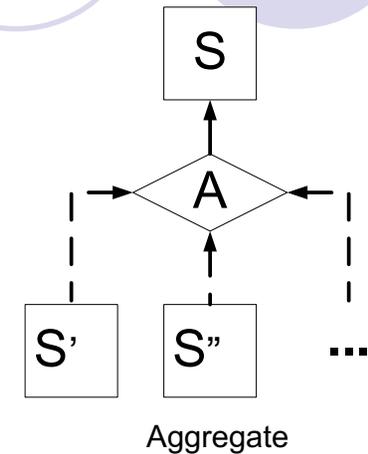
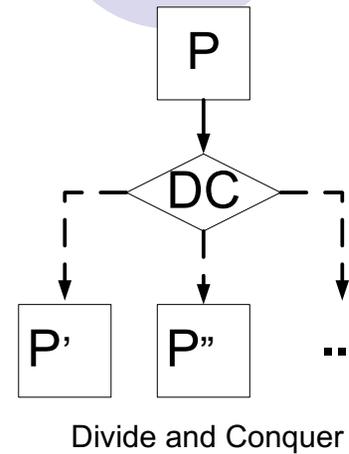
Below the screenshots, a diagram illustrates the iterative process. It shows a sequence of human icons connected by arrows. The first icon is labeled 'improvement \$0.05'. The second icon is labeled '3 votes @ \$0.01'. The process continues with more icons and arrows, ending with an ellipsis '...'. The diagram shows how a single person's improvement leads to a group of people voting, which then leads to further improvements and votes.



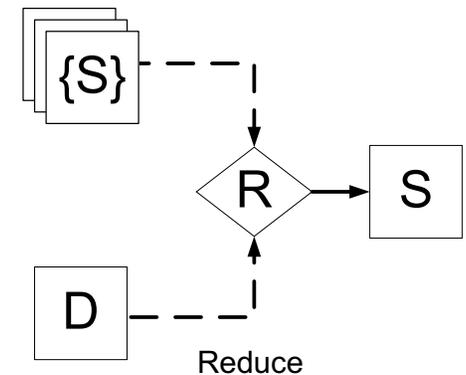
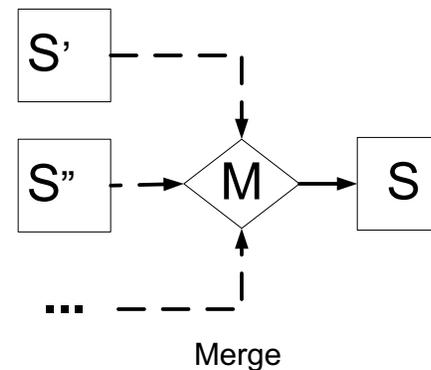
G. Little et al, "Exploring Iterative and Parallel Human Computation Processes," HCOMP 2010

Divide-and-Conquer and Aggregate

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

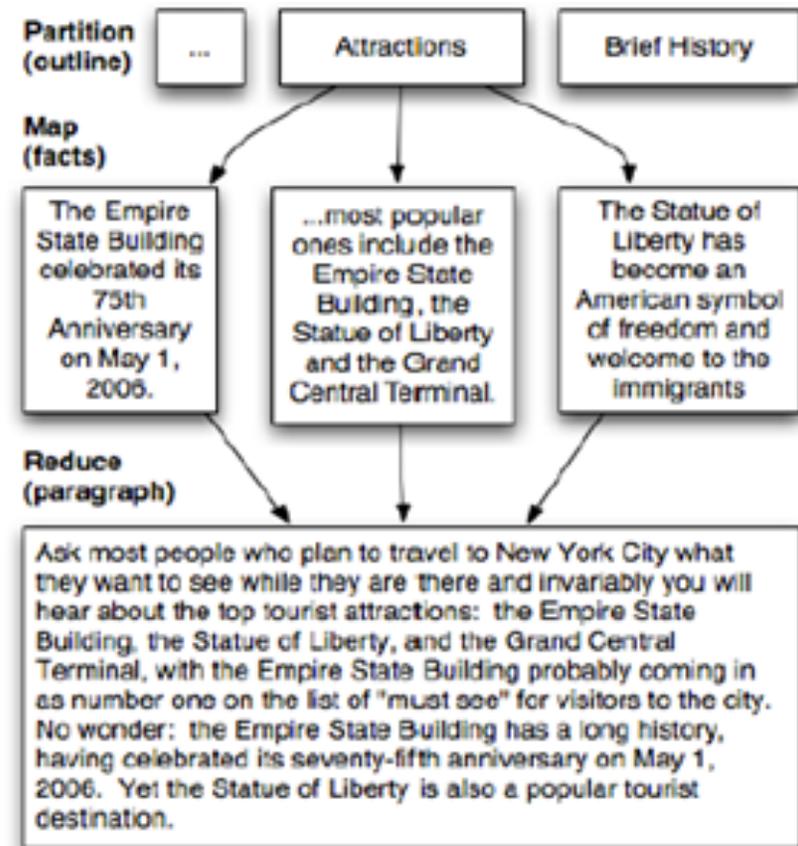
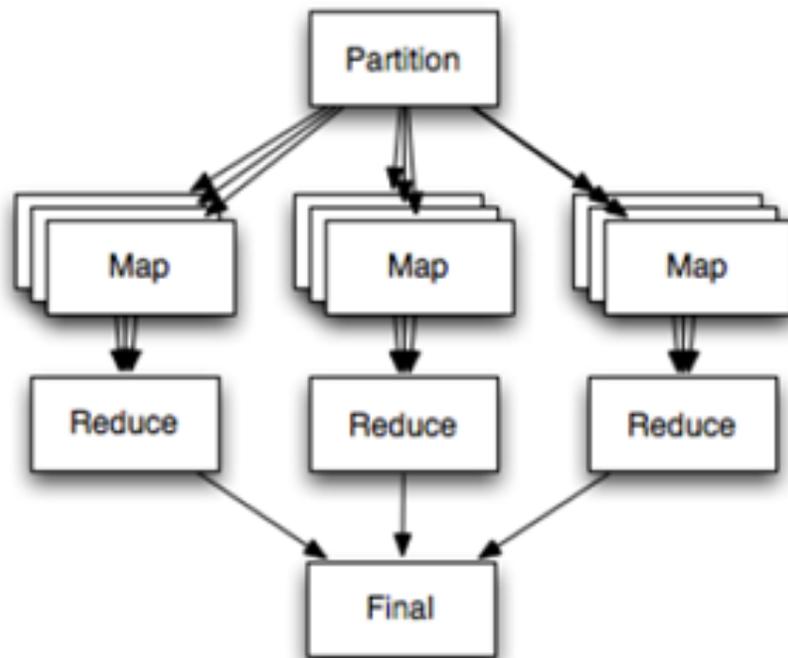


- Two special aggregates
 - Merge
 - Reduce



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

Map and Reduce: A Special Case



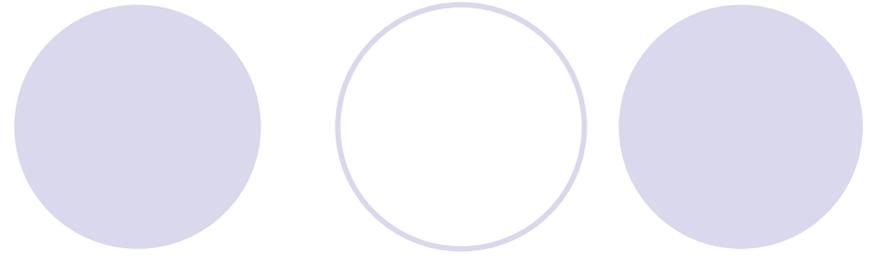
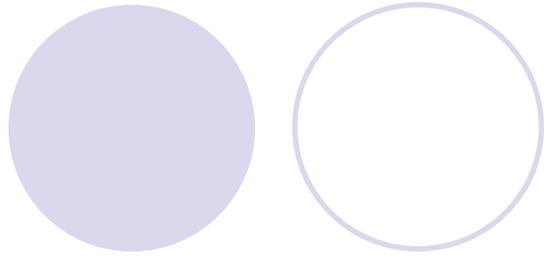
A. Kittur et al, "Crowdforge: Crowdsourcing Complex Work," UIST 2011

Publish and Subscribe

- Pub/Sub middleware-based task assignments



I. Zarko et al, "IoT Data Management Methods and Optimization Algorithms for Mobile Publish/Subscribe Services in Cloud Environments," EU FP 7 OpenIoT Project 2013



Challenges

Opportunities

CHALLENGES AND OPPORTUNITIES

Challenges

Each set has $S/2$ items

r workers r

Each set has $S/10$ items

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

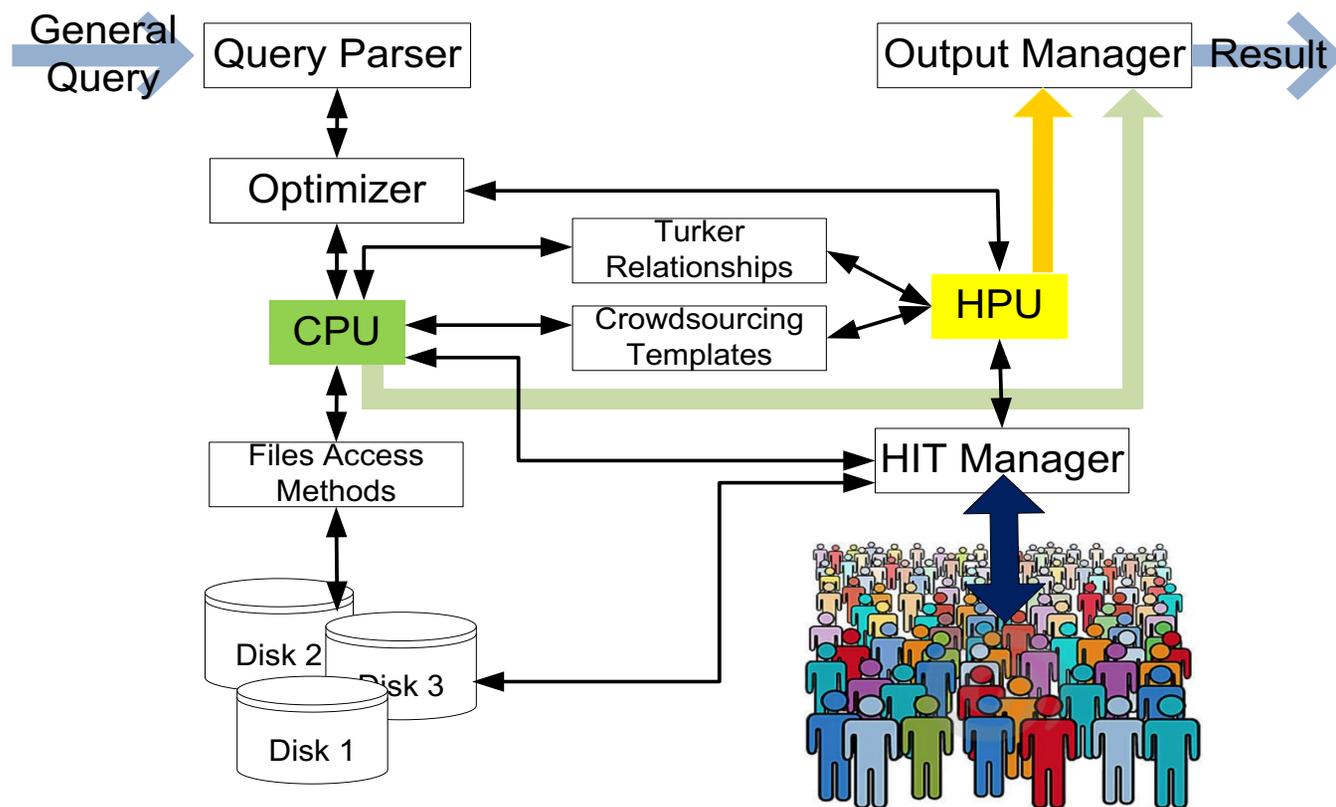
- Incentive: money, glory, and love

- 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:

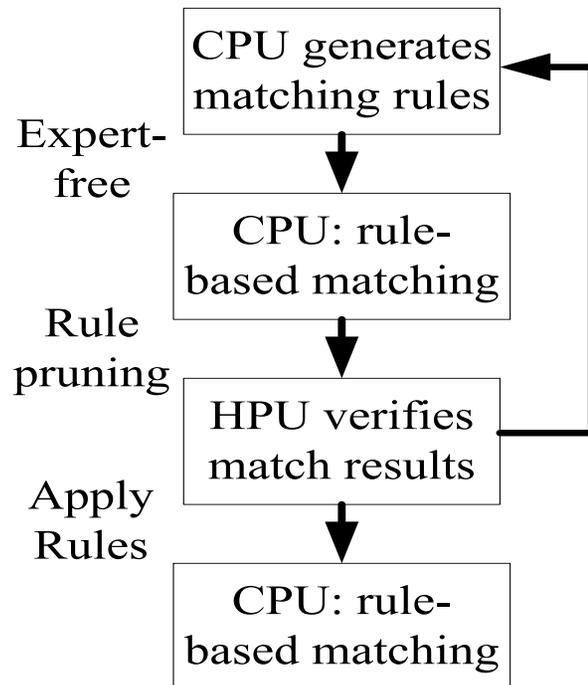


M. Franklin et al, "CrowdDB: Answering Queries with Crowdsourcing,"
SIGMOD 2011

CPU-assisted HPU

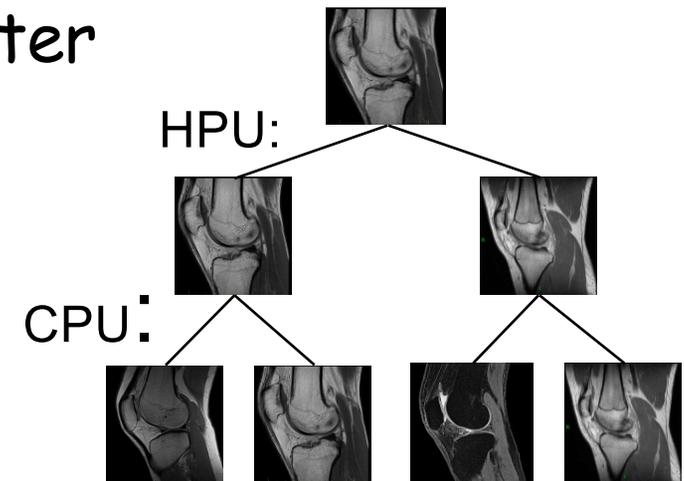
- Entity Matching

e.g. JHU matches John Hopkins Univ.

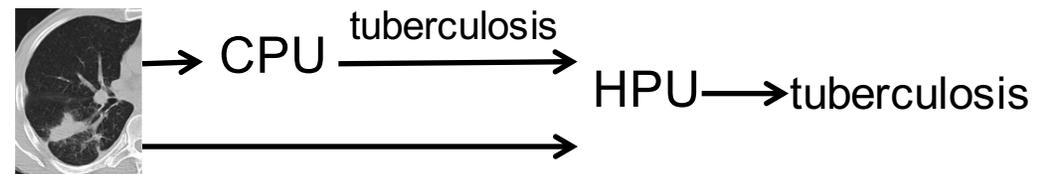


- Radiology

- Filter



- Diagnose

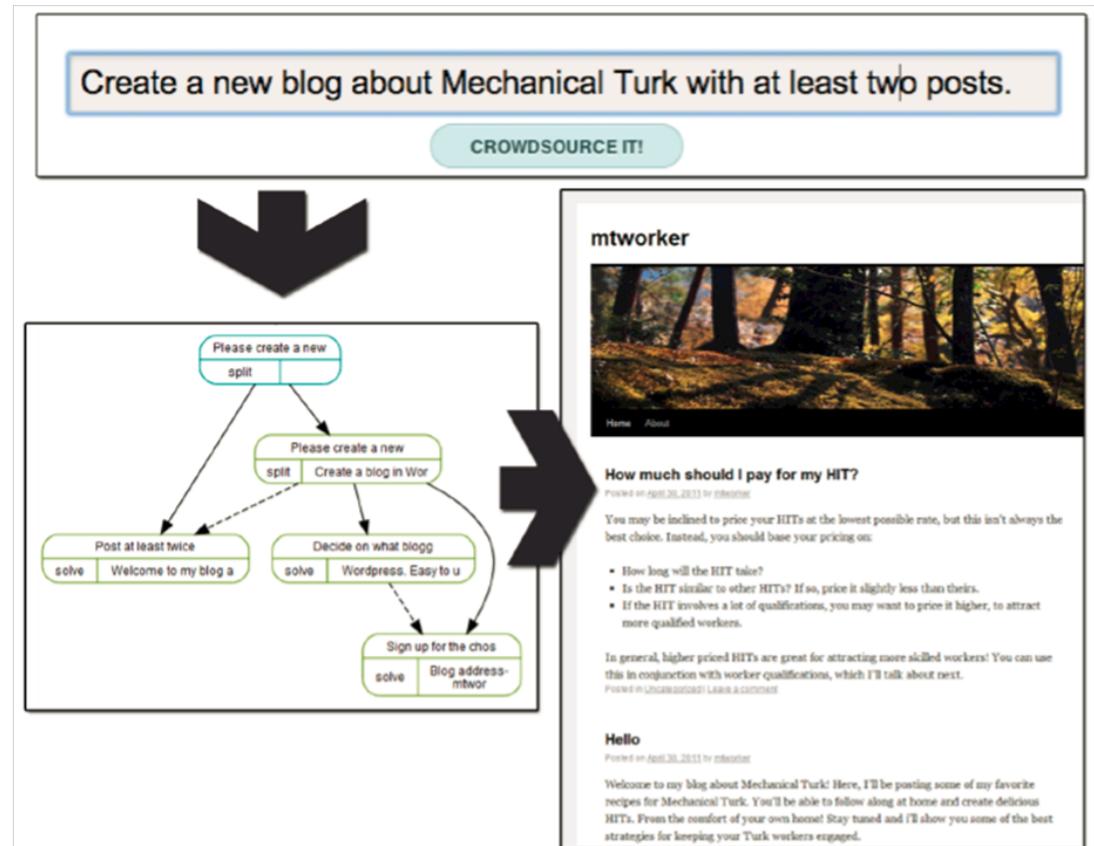


C. Gokhale et al, "Corleone: Hands-off Crowdsourcing for Entity Matching," SIGMOD 2014

Challenges: Collaborative Workflows

Turkomatic

- Complex works require careful and accurate design workflow
- Problems:
 - Loop subtasks
 - Task starvation
 - Multi-stage task with limited budget



Kulkarni et al, "Collaboratively Crowdsourcing Workflows with Turkomatic," CSCW 2012

C. Fofi et al, "Design Patterns for Hybrid Algorithmic-Crowdsourcing Workflows," CBI 2014

Challenges: Multi-dimensional Data

Multi-dimension

- Personal activity data:

- Eating hobby
- Shopping preferences
- Incomes
- Emotional state

- Social data:

- Close friends
- Similar users

- Environmental data:

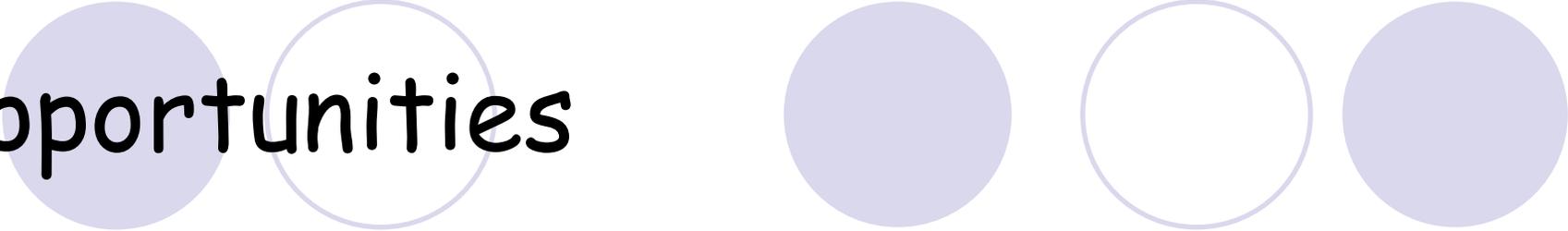
- Locations
- Climates

Integrate all context



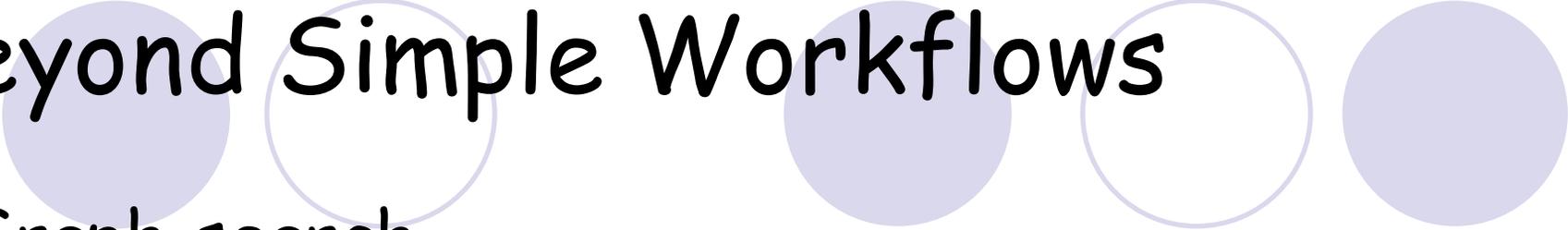
X. Hu et al, "Multidimensional Context-aware Social Network Architecture for Mobile Crowdsensing," IEEE Communications magazine 2014

Opportunities



- Beyond simple workflows
 - Graph search
 - Graph match
- Beyond simple worker selection
 - Dynamic procurement
- Beyond independent workers
 - Social networks

Beyond Simple Workflows



- Graph search

- 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," 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

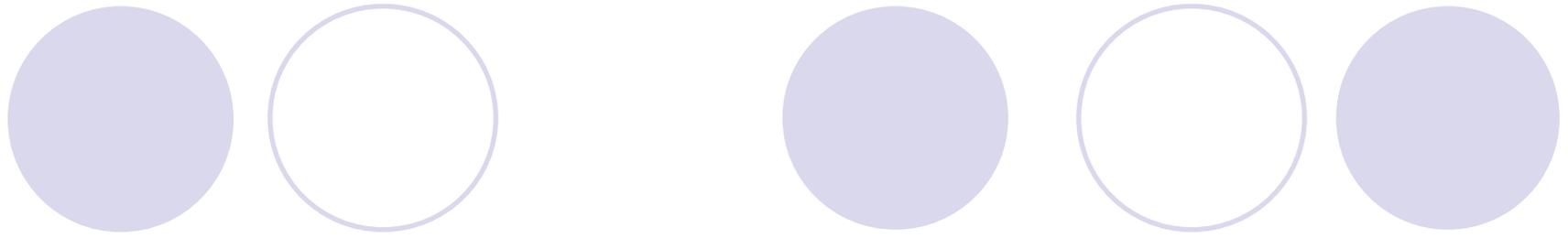


A. Badanidiyuru et al,
"Bandits with Knapsacks:
Dynamic Procurement for
Crowdsourcing," 2013

Beyond Independent Workers

- Social network of workers
- Iterative recruitment of workers through **social ties**
- Challenges
 - Graph searching
 - Timeliness of responses
 - Stoppage condition



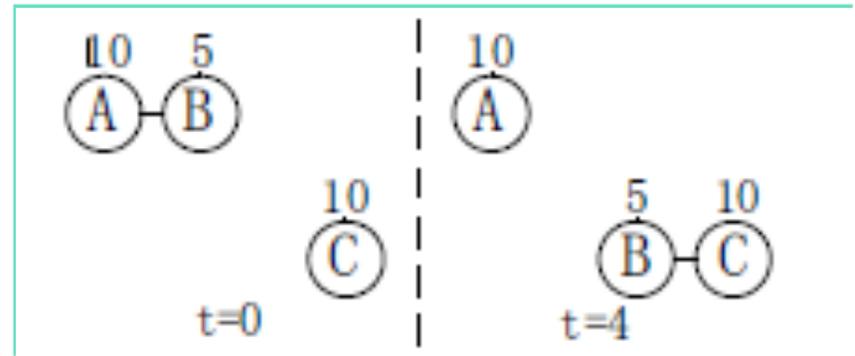


Computational Surplus Around
QQ Example

SOCIAL CROWDSOURCING

Computational Surplus Around

- Friends help friends
 - Fixed individual capability
 - Probabilistic friends' capability
- Makes dissemination decisions
 - Based on the estimations of the fixed and potential computational capacities

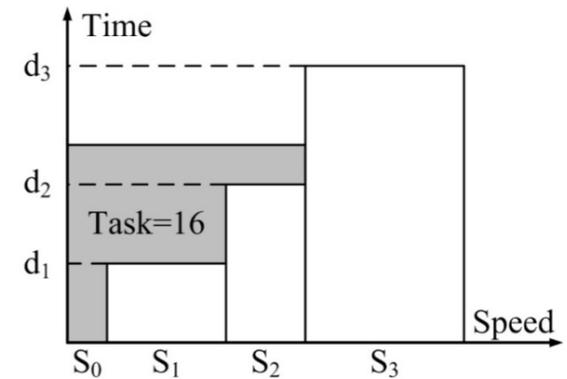
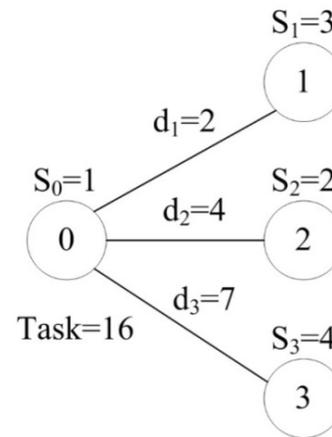


S. Zhang et al, "Minimum Makespan Workload Dissemination in DTNs: Making Full Utilization of Computational Surplus Around," MobiHoc 2013

Water Filling Schedule

- Response delay
- Computation (by a friend)
- Reply delay

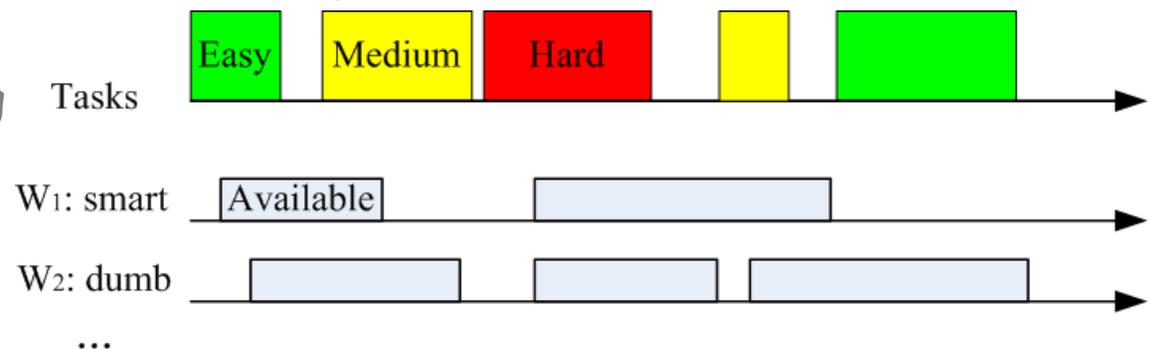
- d_i : response + reply



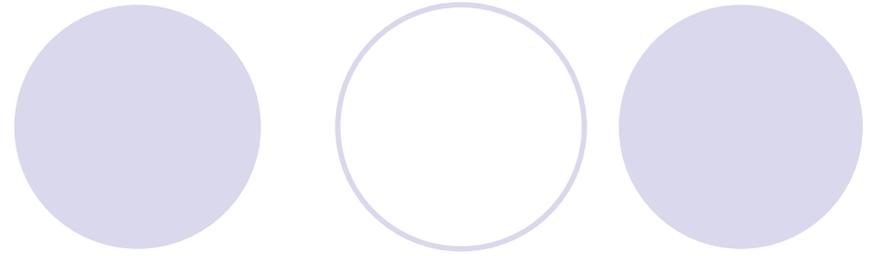
M. Xiao et al, "Multi-task Assignment for Crowdsensing in Mobile Social Networks," INFOCOM 2015

- Scheduling across time: assign jobs to workers

J. Bragg et al, "Parallel Task Routing for Crowdsourcing," AAAI 2014



QQ Example



- Tencent QQ, or QQ
 - Instant messaging
- As of March 2013
 - 798.2 million active QQ accounts
 - Peak of 176.4 million simultaneous online users
- QQ experiment
 - Exploring social status of QQ users by responses

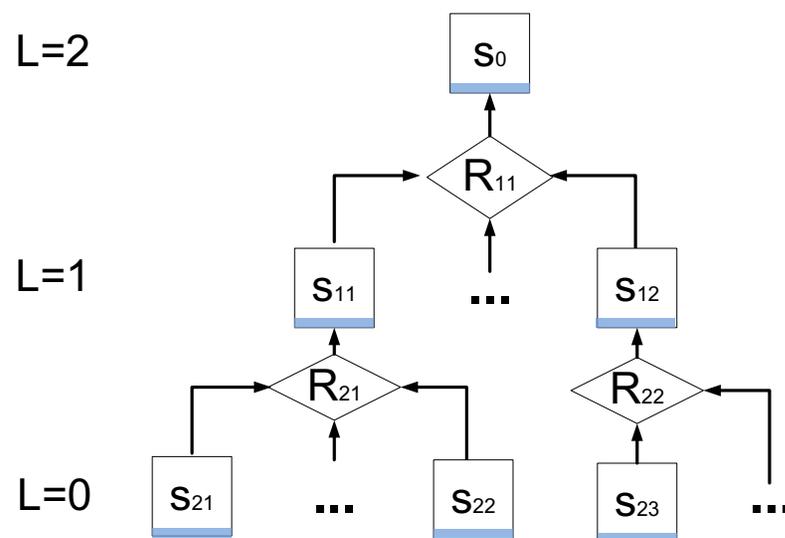


Recursive Doubling (reduce)

- Initial label is $L = "2"$ (subtract L by 1 when forwarding this request to QQ friends)

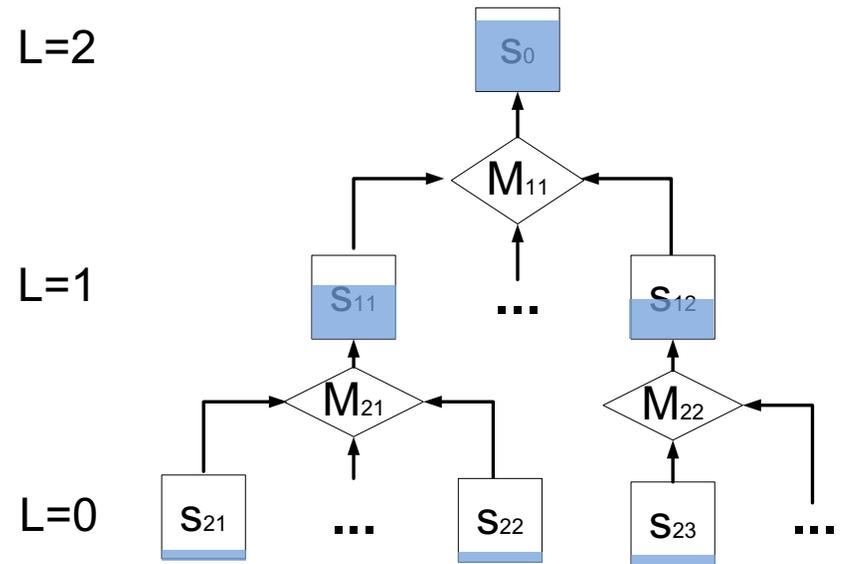
- When $L = 0$, return the total number of QQ friends

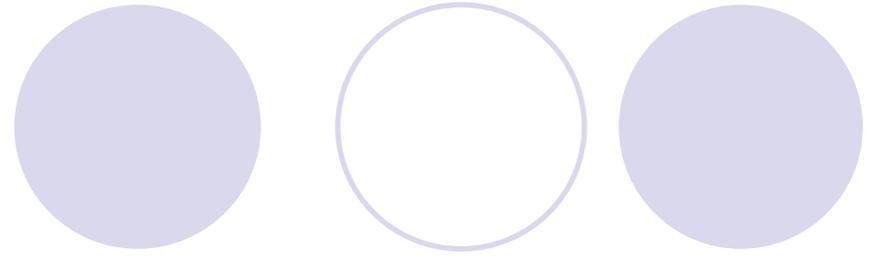
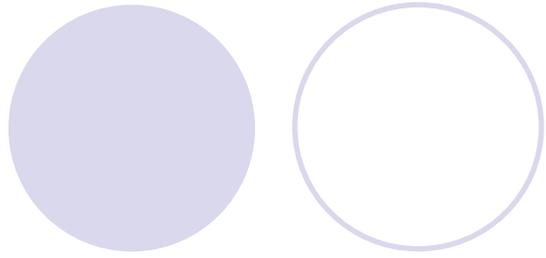
- When $L > 0$, do the following:
 - Forward this request to all QQ friends
 - After receiving the first 10 replies, compute the average number of friends, and send them back to me



Recursive Doubling (merge)

- Initial label is $L = "2"$ (subtract L by 1 when forwarding this request to QQ friends)
- When $L = 0$, return the following:
 - Basic information (B)
 - Number of friends (N)
 - Timestamps (T)
- When $L > 0$, do the following:
 - Forward this request to all QQ friends
 - **Pack** the first 10 replies, together with your own information (B, N, T), and send them back to me

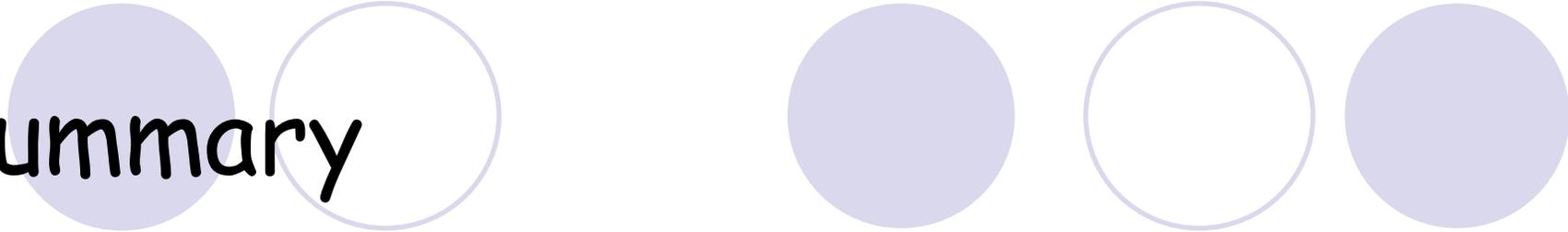




Summary

Acknowledgements

CONCLUSION



Summary

- HPU as a new paradigm to compliment the traditional CPU-based computing for big data
- Many unexplored algorithmic problems
 - Worker selection
 - Social connections of workers
 - Workflow design
 - Cost-time-quality trade-offs
 - Incentive mechanisms

Acknowledgements

○ Wei Chang

Temple University

○ Grace Ju

Carnegie Mellon University

○ Wenjun Jiang and Ying Dai

Hunan University and LinkedIn

