



Algorithmic Crowdsourcing

(and Applications in Social Networking)

Jie Wu

Dept. of Computer and Info. Sciences
Temple University

Road Map

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





What

Why

Basic Components

INTRODUCTION

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)



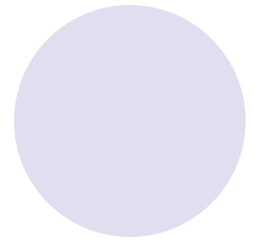
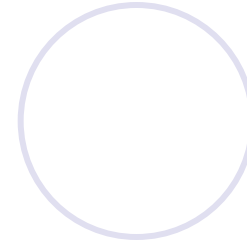
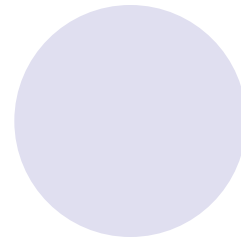
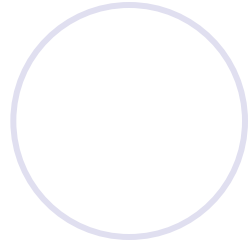
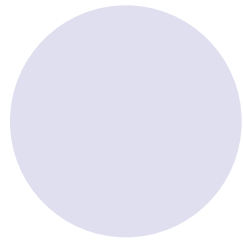
Requester



HotPOST 2013



Worker Pool



Worker

HIT

Dashboard

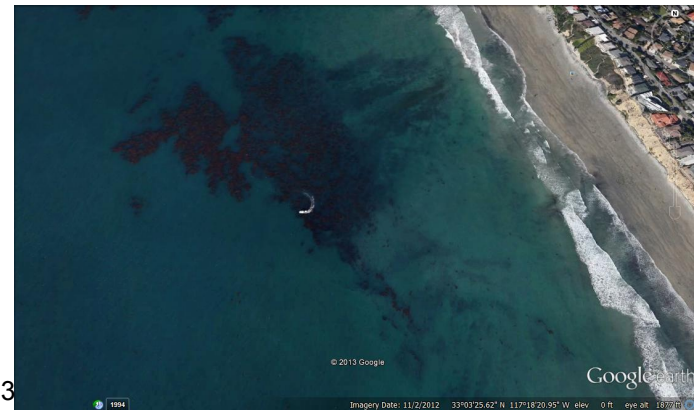
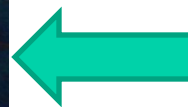
MECHANICAL TURK

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.



HotPOST 2013

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



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



- 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

The Worker

All HITS

1-10 of 1982 Results

Sort by:

[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
 - By creation date, payment amount, time allotment

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)

During a HIT

Timer: 00:02:27 of 10 minutes

Finished with this HIT? Let someone else do it?

Submit HIT

Return HIT

Total Earned: \$4.72
Total HITs Submitted: 7

Automatically accept the next HIT

Copy Text from Business Card

Requester: Oscar Smith

Qualifications Required: None

Reward: \$0.02 per HIT

HITs Available: 1

Duration: 10 minutes

Instructions

You are in

Please Copy Text from Business Card:
Training Mode. You will still be paid for completing this HIT.



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

Michele Howard ?

You are Correct!

Information Designer Company

You are Correct!

You are Correct!

mhoward@lhouse.com www.lhouse.com

You are Correct!

You are Correct!

Address: ?

444 Liberty Avenue

Address Line 1 is **Four Gateway Center**

Note: This address has 3 address lines. To add another Address Line click on the blue 'add line' link below.

Please try again.

[add line](#)

- Shows duration of time
- Gives worker the option to "Return" the HIT

HotPOST 2013

Completing a HIT

Timer: 00:00:00 of 10 minutes

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

Accept HIT

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

Sample Dashboard



Your Account | **HITs** | Qualifications

292,650 HITs available now

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

Find **HITs** containing _____ that pay at least \$ **0.00** for which you are qualified require Master Qualifications

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

Your Worker ID: **XXXXXXXXXXXX**

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...](#)

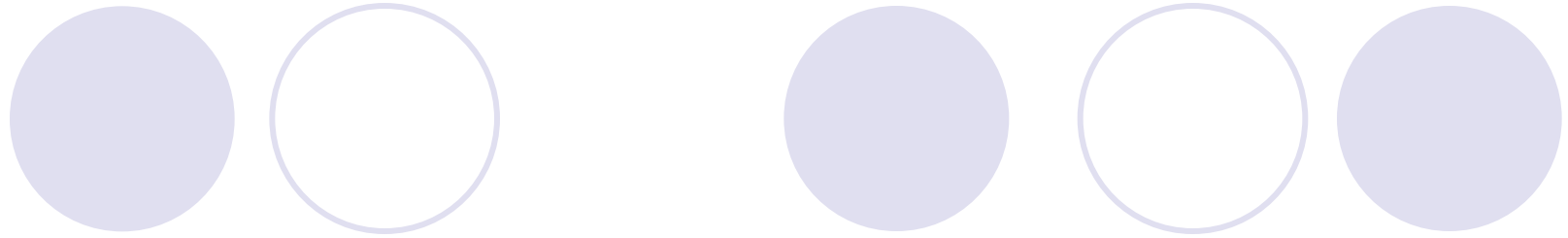
HIT Totals [\(What's this?\)](#)

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	—

HotPOST 2013

[FAQ](#) | [Contact Us](#) | [Careers at Amazon](#) | [Developers](#) | [Press](#) | [Policies](#) | [Blog](#)

©2005-2013 Amazon.com, Inc. or its Affiliates



EteRNA


APPLICATIONS: BIOLOGY

EteRNA: CMU, Stanford

- Choose difficulty level
- Solve puzzles:
A to U, C to G
- Aim: to gain mastery over the way RNA folds

The screenshot shows the EteRNA website interface. At the top, there is a navigation bar with the EteRNA logo and the tagline "Make Molecules Advance Science". The main content area is titled "Challenges" and features a "Go To Player Puzzles" button. Below this, there are several puzzle cards, each with a title, a difficulty level, a number of people cleared, and a reward. The puzzles include "Simple Hairpin", "Hairpin Series", "Hairpin RNA", "Loop Strategy - Learn about the 1-1 Loop", "Y RNA - Difficulty Level 0", "Synthetic RNA - Difficulty Level 0", "Oryza sativa 4 - Difficulty Level 0", "Oryza sativa 5 - Difficulty Level 0", and "Chlamydomonas reinhardtii - Difficulty Level 0". Each puzzle card also includes a small diagram of the RNA structure. On the right side of the interface, there is a chat window titled "Anonymou.." showing a conversation between users "Matt Indykiewicz" and "Spanda". The chat window also includes a search bar and a list of recent news items.

EteRNA: CMU, Stanford



username password [Log in](#) [Register](#) [Lost password?](#) [Facebook connect](#)

[Roadmap](#) [Puzzle](#) [RNA Lab](#) [Community](#) [About](#)

Loop Strategy - Learn about the 1-1 Loop

 21728 55 01 Feb 2011

[Play Now](#)

Structure Notation

Puzzle Description

Welcome to the Strategy Puzzles Series. This series will help you learn basic strategies to solve more advanced puzzles in the Challenges and Lab sections of the game.

""Loops""

HotPOST 2013

A loop is a series of unpaired bases between pairs. One of the loop is shown in the picture below. This loop has 2

Anonymou..

Chat Players Online (36)

Matt Indykiewicz: I didnt use the same cuz this is linked with fb [10:43 AM]
Matt Indykiewicz: RollingIndian2212 [10:43 AM]
Matt Indykiewicz: thats me ^.^ [10:43 AM]
TomoeUzumaki: hey giys [10:44 AM]
Spanda: hey [10:44 AM]
Matt Indykiewicz: hey gurls [10:44 AM]
Spanda: im doing some puzzles so that i dont get deleted :D [10:44 AM]
TomoeUzumaki: *guys [10:45 AM]
Spanda: is rank 25882 enough? [10:45 AM]

Please log in to chat

[Open Chat Window](#)




Latest News

EteRNA: CMU, Stanford

The screenshot shows the EteRNA website interface. At the top left, there is a link "Strategy - Learn about the 1-1 Loop". At the top right, there is a "Register" link. Below these are navigation links: "Roadmap", "Puzzle", "RNA Lab", "Community", and "About EteRNA".

A large, tilted "MISSION!" stamp is on the left. In the center, a white outline of an RNA structure is shown against a dark background with faint circular patterns.

Three mission instructions are listed on the left:

-  **Your RNA must fold into the structure in white outline.**
-  **You must have 1 or fewer G-C pairs.**
-  **You must have 1 or more G-U pairs.**

At the bottom, there is a control bar with various icons and a sequence of nucleotides: A, U, G, C.

On the right side, there is a "Chat" window titled "Chat" with "Players Online (36)". The chat log shows the following messages:

- deleted :D [10:44 AM]
- TomoeUzumaki: *guys [10:45 AM]
- Spanda: is rank 25882 enough? [10:46 AM]
- Matt Indykiewicz: Nope [10:46 AM]
- Spanda: wat is enough? [10:46 AM]
- TomoeUzumaki: start solving [10:47 AM]
- Spanda: IM WORKING ON IT [10:47 AM]

Below the chat log, it says "Please log in to chat".

EteRNA: CMU, Stanford

Loop Strategy - Learn about the 1-1 Loop

Total:
5.6 kcal
Loop
5.6kcal

0/1 0/1

Roadmap Puzzle R

HotPOST 2013

A U G C



GWAP.com
reCAPTCHA
OnToGalaxy

COMMONSENSE KNOWLEDGE

GWAP.com: CMU (no longer available)

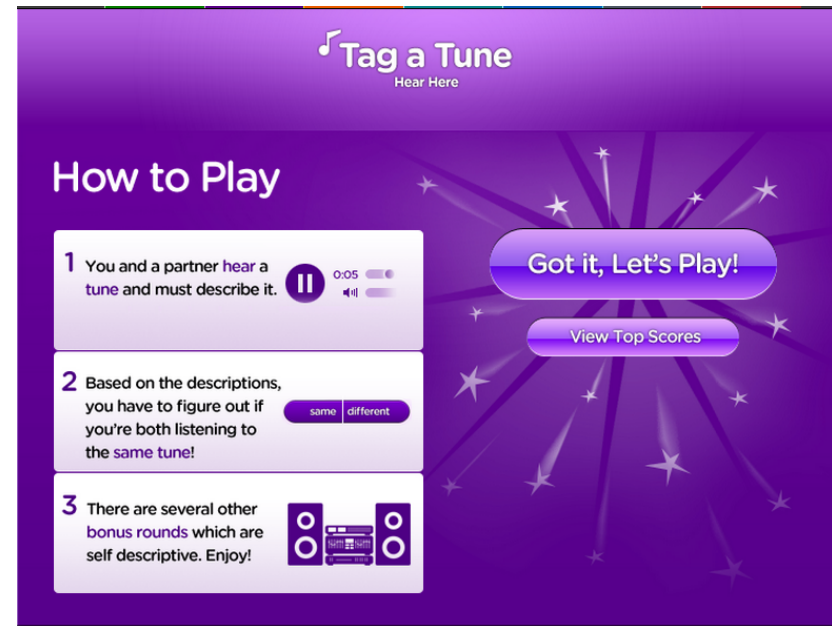
ESP Game

- Labeling images



Tag a Tune

- Labeling tunes



reCAPTCHA: CMU



reCAPTCHA™

- 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!

HotPOST 2013

[Blog](#) | [About Us](#) | [Contact](#) | [FAQs](#) | [Terms](#) | [Privacy](#)
© 2009, all rights reserved.

OnToGalaxy: University of Bremen

Commander: `guest_1370454894191` :: Rank: Ensign :: Shields: 100 - 100 :: Hull: 100 -

Menu

Remember this hint:
The keyword 'Grass' is related to 'green',
and it is also related to itself, to 'Grass'

Free Repair

Press ENTER to continue.


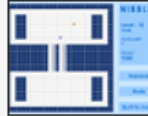


Caution, Ensign!
My scans show the next
convoy will be accompanied
by some basic attack drones.

HotPOST 2013
Press ENTER to continue

Ensign
Score: 6000

CHAT GAME MORE GAMES

Try these recommended games

-  [shining galactica](#)
SHOOTER
★★★★☆
-  [Nibbler](#)
ACTION
★★★★☆
-  [Coherent Beam](#)
ACTION
★★★★☆
-  [Versatility](#)
MUSIC
★★★★☆

Load More Games

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





Galaxy Zoo



Fine-grained Recognition

IMAGE PROCESSING

Galaxy Zoo: Zooniverse




The screenshot displays the Galaxy Zoo Zooniverse interface. At the top, a navigation bar includes links for CLASSIFY, SCIENCE, STORY, GALAXY ZOO (highlighted in a yellow banner), ASTRONOMERS, DISCUSS, and PROFILE. The main content area features a large image of a galaxy on the left. To the right of the image, the word "Classify" is displayed, along with "SDSS" and "Invert" icons. Below this, a question is posed: "SHAPE Is the galaxy simply smooth and rounded, with no sign of a disk?". Three options are provided: "Smooth" (represented by a white circle), "Features or disk" (represented by a spiral galaxy), and "Star or artifact" (represented by a starburst).

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

Classify   [Help](#) [Restart](#)

SDSS **Invert**

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




  

Smooth Features or disk Star or artifact

GalaxyZoo: Zooniverse



- Zooniverse
 - A website dedicated to citizen science projects
- A platform to label the different galaxies
- Step-by-step instructions and visual guidelines

ROUND
How rounded is it?





Completely round In between Cigar shaped

ODD
Is there anything odd?

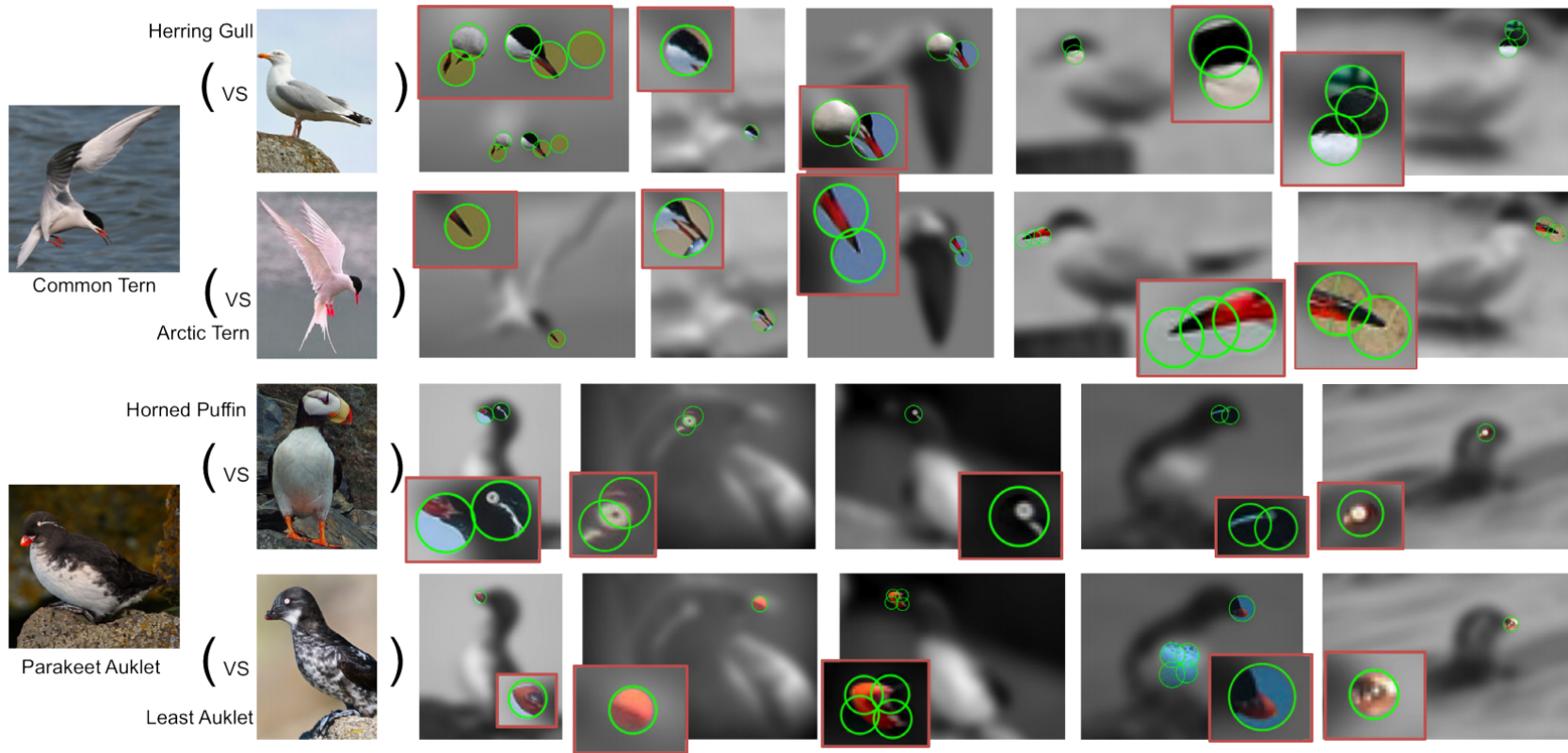
Yes No

BULGE
How prominent is the central bulge, compared with the rest of the galaxy?

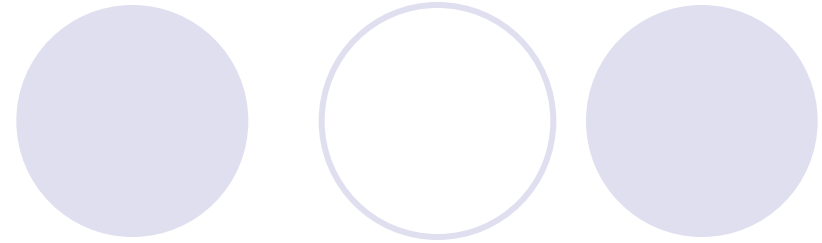
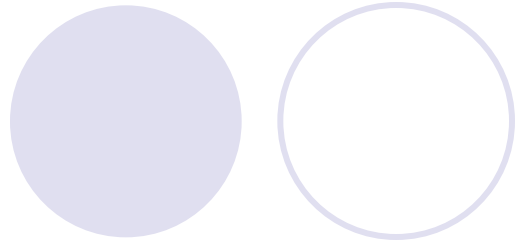
   

No bulge Just noticeable Obvious Dominant

Fine-Grained Recognition



J. Deng et al, "Fine-Grained Crowdsourcing for Fine-Grained Recognition," CVPR 2013



Sequential

Iterative and Parallel

Divide-and-Conquer

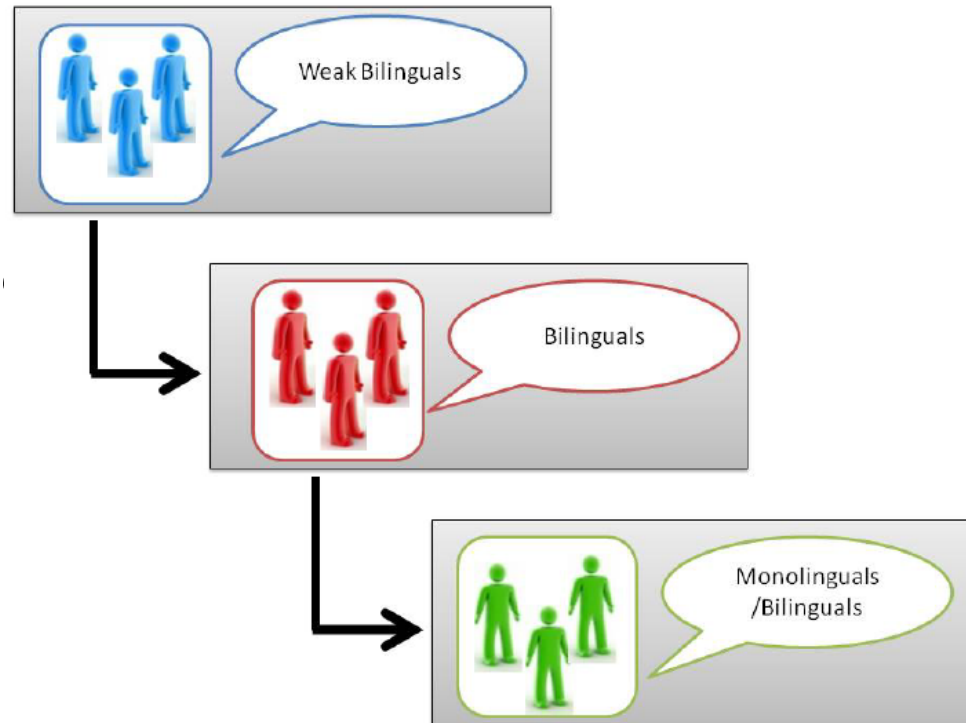
Divide-and-Conquer and Aggregate

Map and Reduce: a Special Case

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

Handwriting Recognition Task - Mozilla Firefox

- Please improve the transcription of this handwriting.
- People will vote whether to approve your changes.

You (?) (?) (?) (work). (?) (?) (?) work (not) (time). I (?) (?) a few grammatical mistakes. Overall your writing style is a bit too (phony). You do (?) have good (points), but they got lost amidst the (writing). (signature)

Submit

MTurk Task - Mozilla Firefox

- Please select the better transcription for this handwriting.
- Differences are highlighted in yellow.

You (misspelled) (several) (words) (work). (?) (?) (?) work next (time). I also notice a few grammatical mistakes. Overall your writing style is a bit too (phony). You do (?) have good (points), but they got lost amidst the (writing). (signature)

You (?) (?) (?) (work). (?) (?) (?) work (not) (time). I (?) (?) a few grammatical mistakes. Overall your writing style is a bit too (phony). You do (?) have good (points), but they got lost amidst the (writing). (signature)

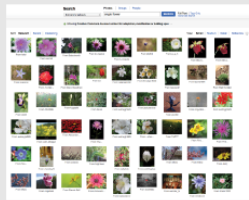
improvement \$0.05

3 votes @ \$0.01

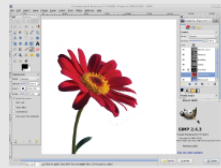
G. Little et al, "Exploring Iterative and Parallel Human Computation Processes," HCOMP 2010

Divide-and-Conquer

Step 1: Find
"Find a close-up image of a flower on flickr.com and load it in GIMP."



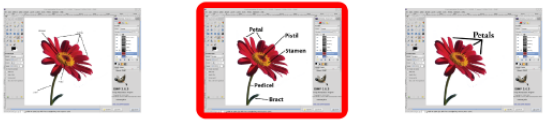
Step 5: Label
"Consult Wikipedia to label the petal, pistil, stamen, pedicel, and bract."




Step 2: Vote



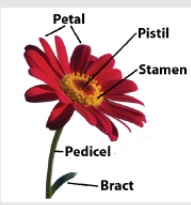
Step 6: Vote




Step 3: Crop
"Crop the image so that only the petals and top of the stem are visible."



Final



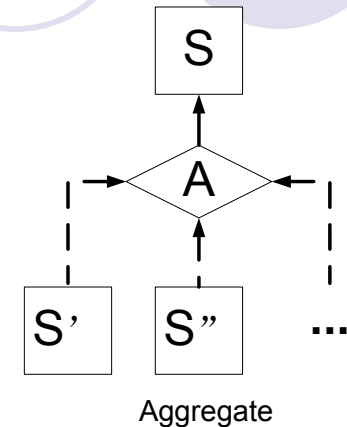
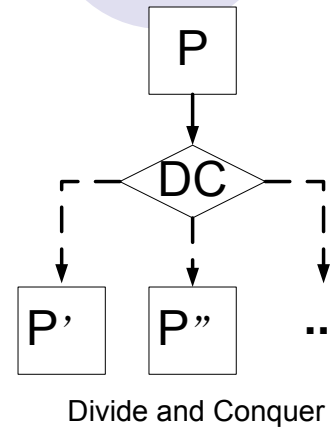
Step 4: Vote



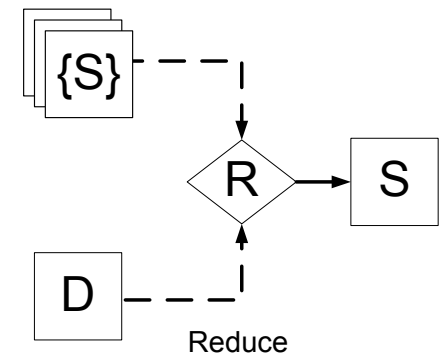
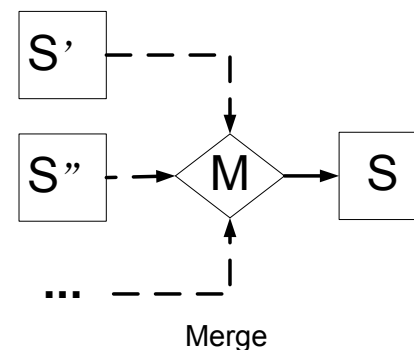
K. I. Murray et al, "Multiverse: Crowd Algorithms on Existing Interfaces," CHI 2013

Divide-and-Conquer and Aggregate

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

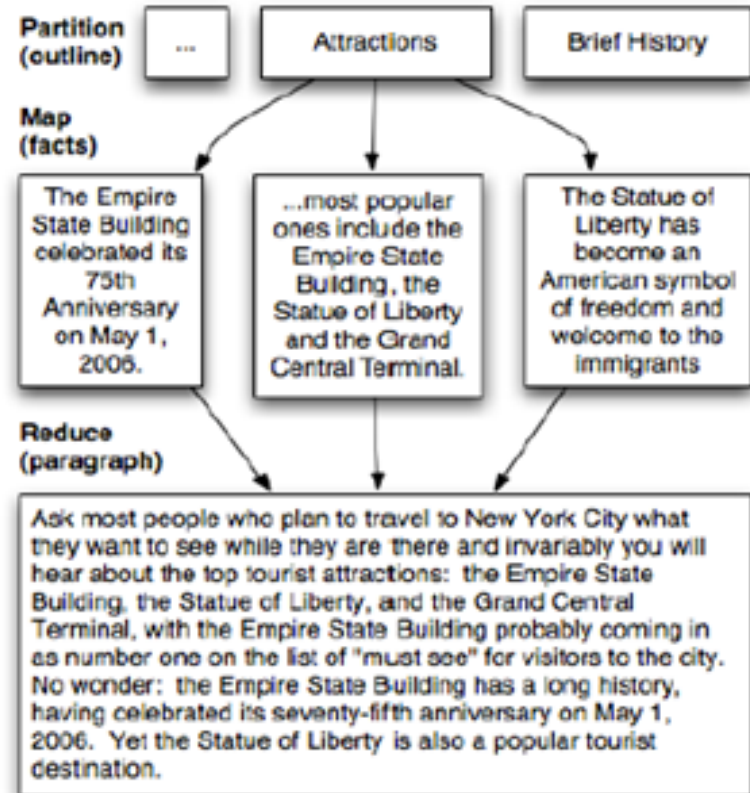
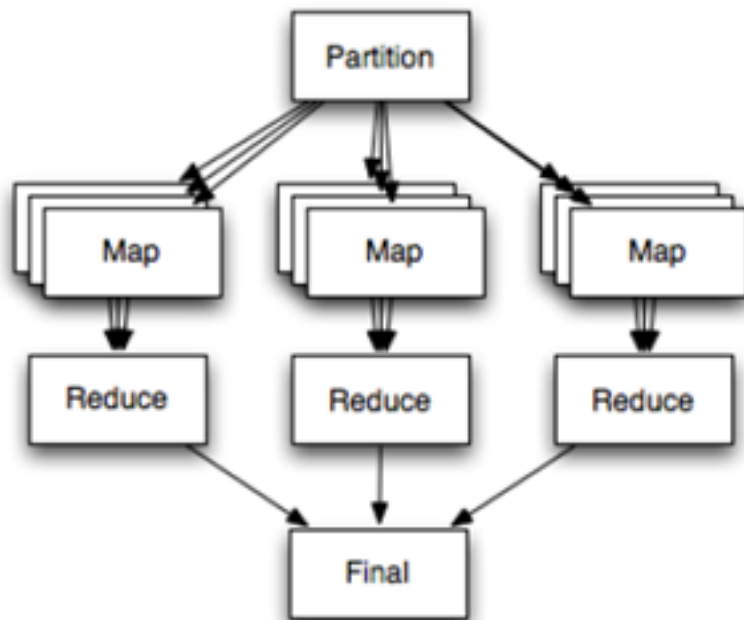


- Two special aggregates
 - Merge
 - Reduce



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



Challenges

Opportunities

CHALLENGES AND OPPORTUNITIES

Challenges



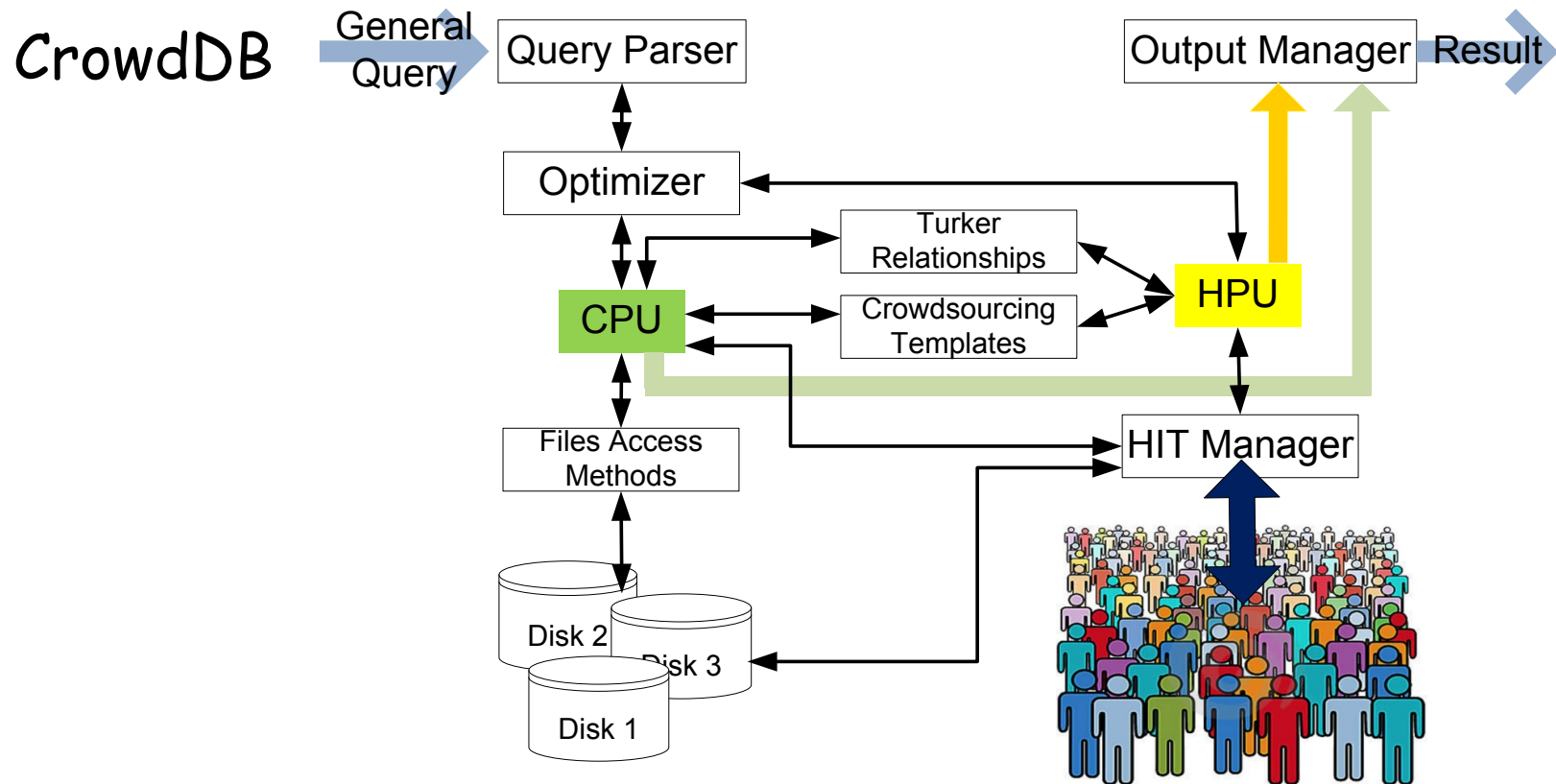
- Trade-offs: time, cost, and quality
 - Max algorithm with human error (with a probability)
 - Maximize quality (via redundancy) subject to cost and time

P. Venetis et al, "Max Algorithms in Crowdsourcing Environments,"
WWW 2012

- 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



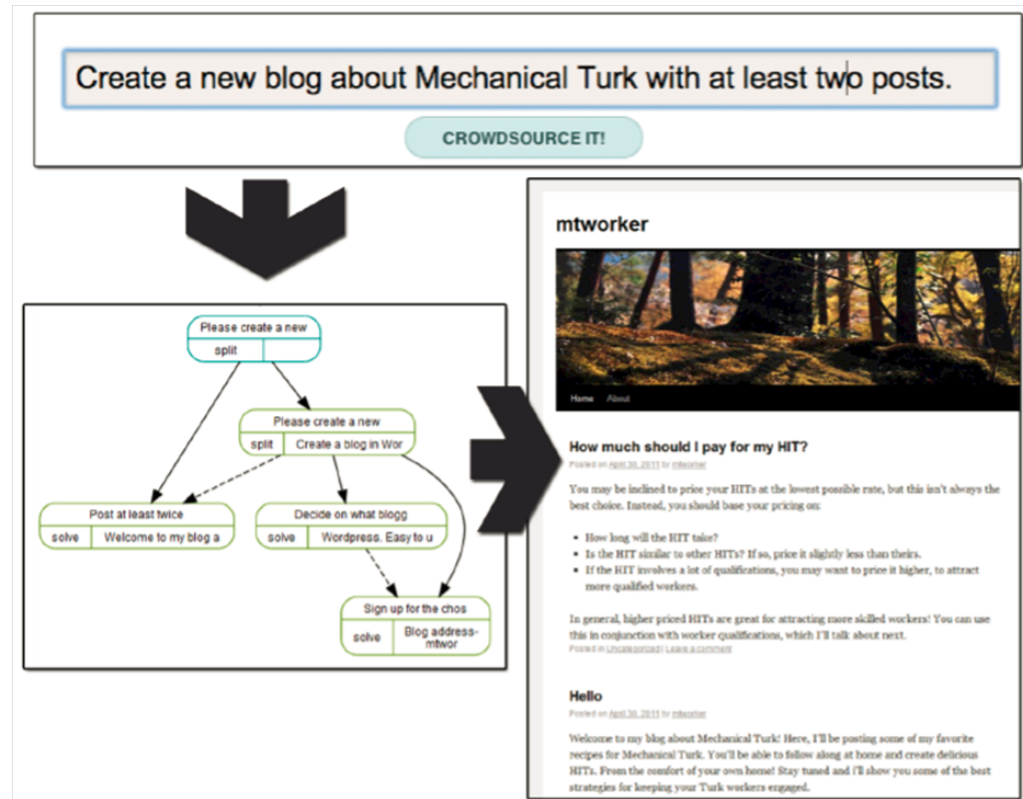
M. Franklin et al, "Crowddb: Answering Queries with Crowdsourcing," SIGMOD 2011

HotPOST 2013

Challenges: Collaborative Workflows

Turkomatic

- Complex works require careful and accurate design workflow
- Problems:
 - Loop subtasks
 - Task starvation



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

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



Beyond independent workers

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



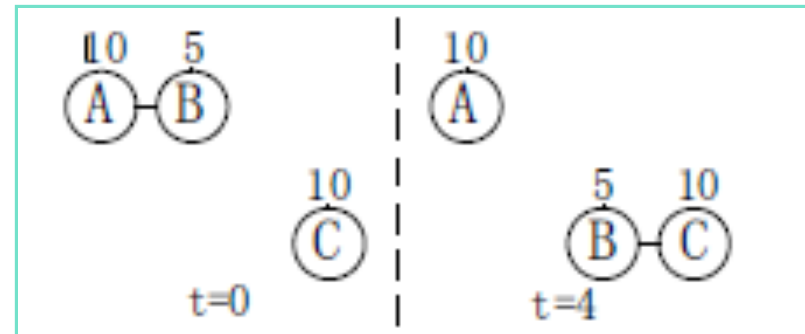


Computational Surplus Around
QQ Example

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

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

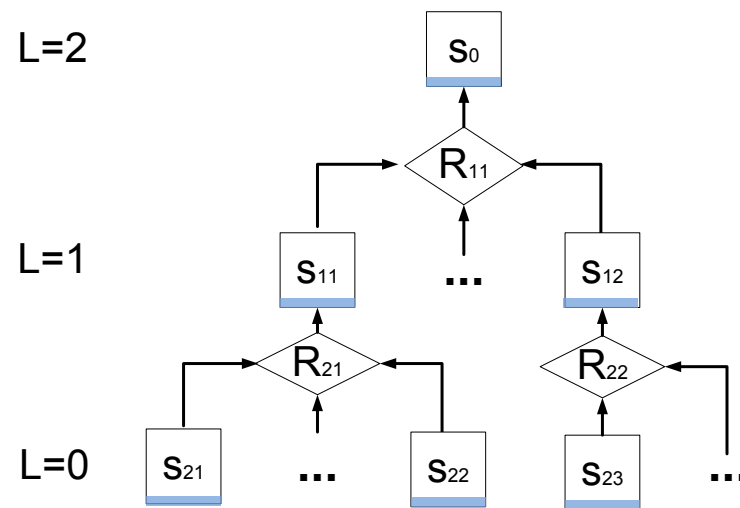


Iterative Request/Reply (reduce)

- 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



Iterative Request/Reply (merge)

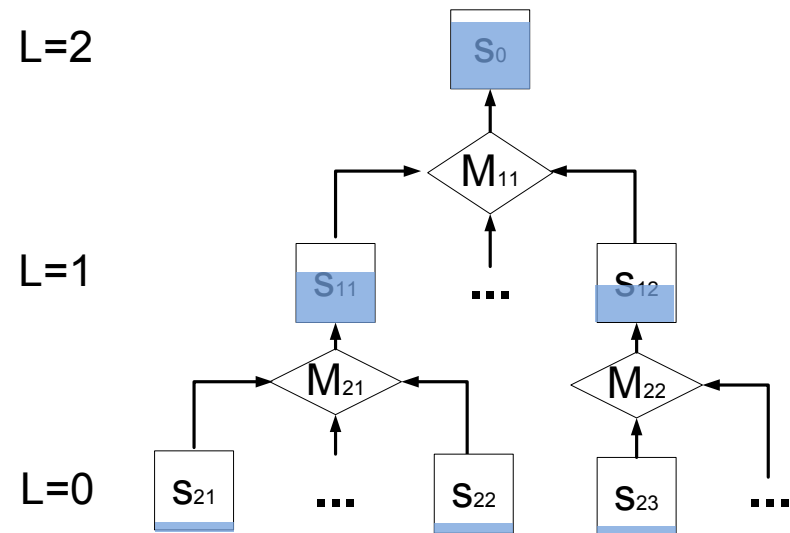
- Initial label is $L = "2"$ (subtract L by 1 when forwarding this request to QQ friends)

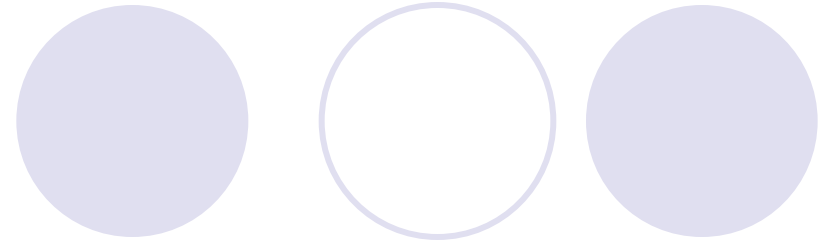
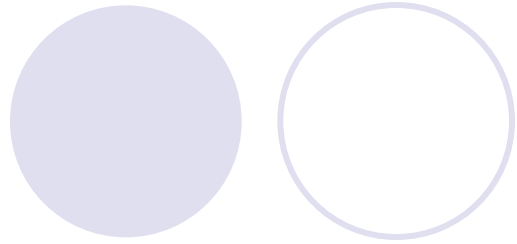
- When $L = 0$, return the following:

- 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
- Many unexplored algorithmic problems
 - Worker selection
 - Social connections of workers
 - Workflow design
 - Cost-time-quality trade-offs
 - Incentive mechanisms

Acknowledgements

○ Wei Chang
Temple Univ.

○ Grace Ju
Carnegie Mellon Univ.

