

Algorithmic Crowdsourcing and Applications in Big Data

Jie Wu

Dept. of Computer and Info. Sciences

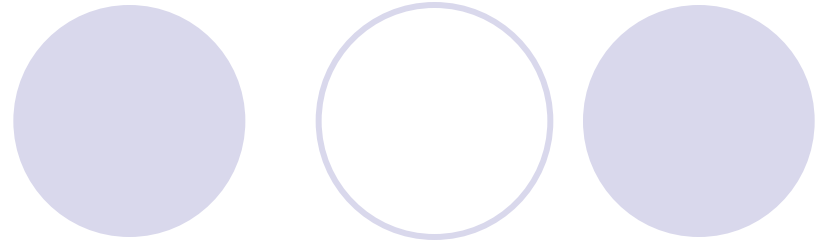
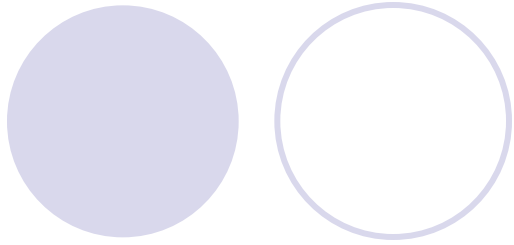
Temple University



Road Map

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





What

Why

Basic Components

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.

From Wikipedia

- “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



Big Data

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



Transparent Computing and Big Data



Worker Pool

Malaysia Airlines Flight 370



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

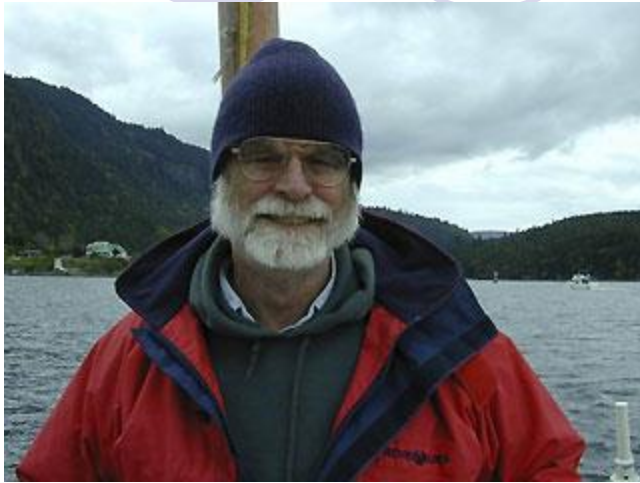
● March 11

I just saw on our local Denver Fox news (KDVR.com) that a local company, Digital Globe, 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.**

As you already know, Digital Globe has participated in the ISTE C IAC meetings (Jay Smith for example).

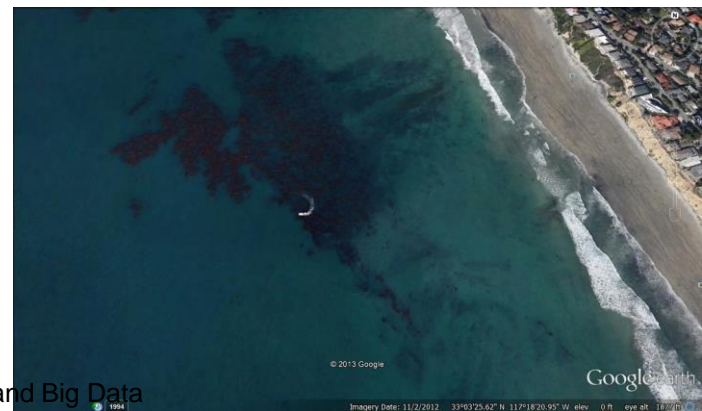
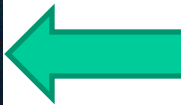
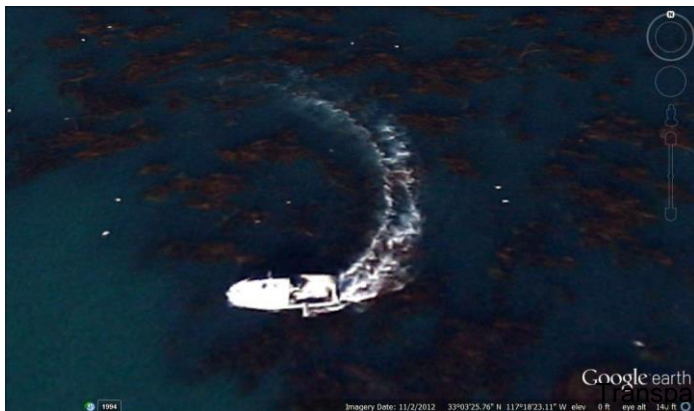


Example 1: Help Find Jim Gray



- 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.



Example 2: 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) ...

Example 3: 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

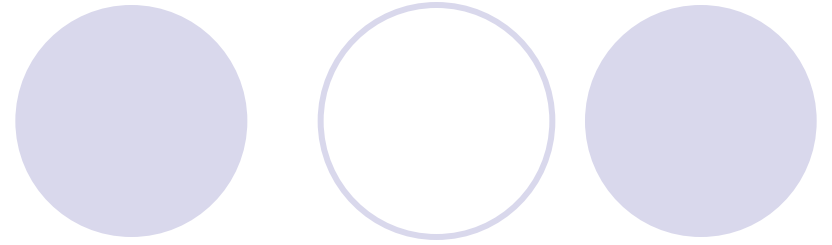
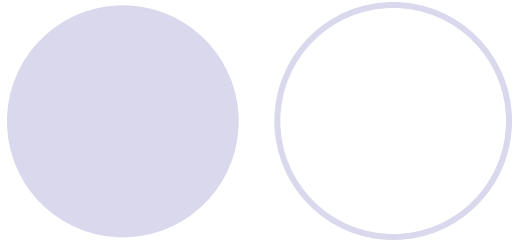


New York City



Bratislava





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



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

Worker: Contract for a HIT

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
 - By creation date, payment amount, time allotment

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

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: 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

Transparent Computing and Big Data

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	—

Transparent Computing and Big Data

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

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

Avoid Shady Requester

How Turkopticon works:

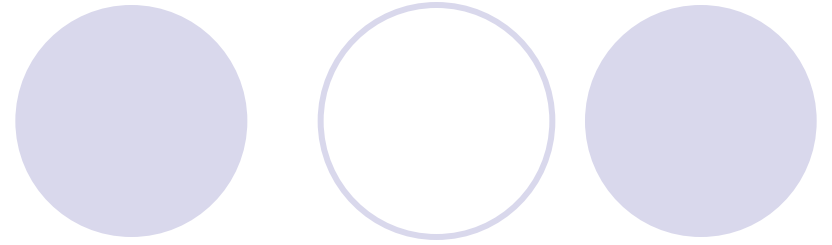
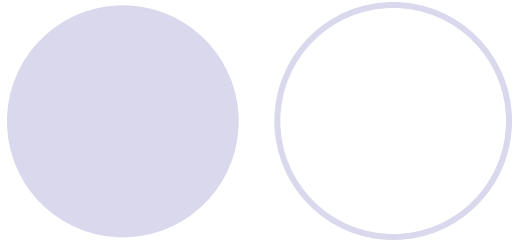
Turkopticon adds functionality to Amazon Mechanical Turk as you browse for HITs and review status of work you've done. As you browse HITs, Turkopticon places a button next to each requester and highlights requesters for whom there are reviews from other workers. Bad reviews let you avoid shady employers 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 displaying review metrics. The requester is identified as 'Product Search' and has a 'HIT Expiration Date' field. The dropdown menu lists 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.

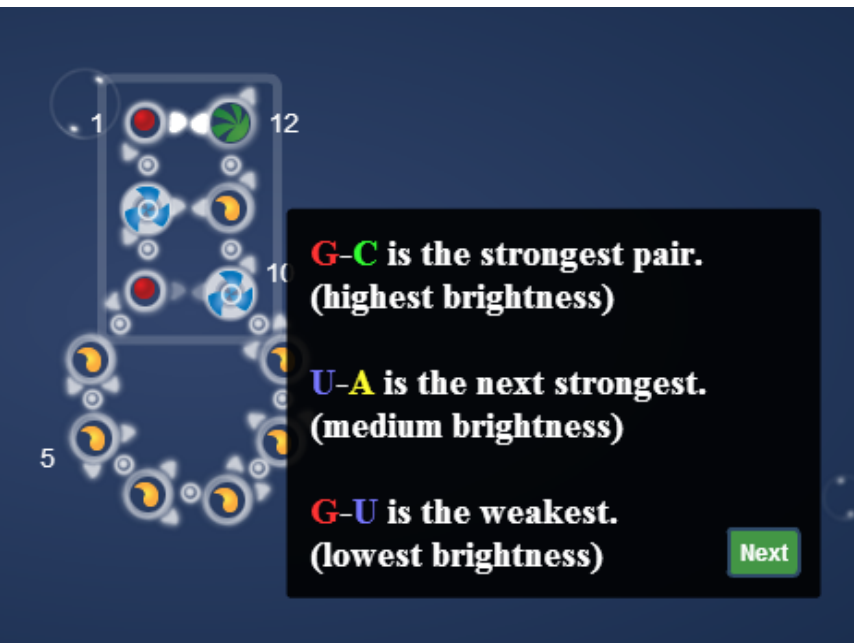


EteRNA

APPS: BIOLOGY

EteRNA: CMU, Stanford

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



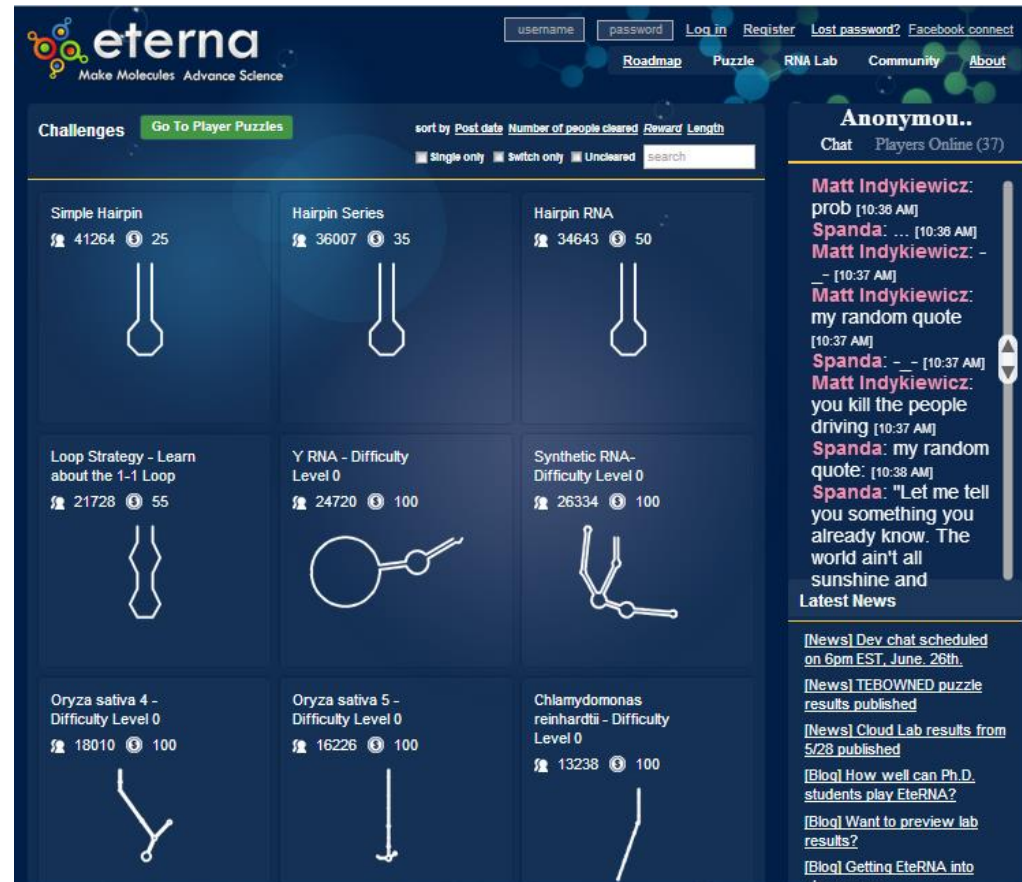
1 12
10
5

**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

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

Anonymou..
Chat Players Online (37)

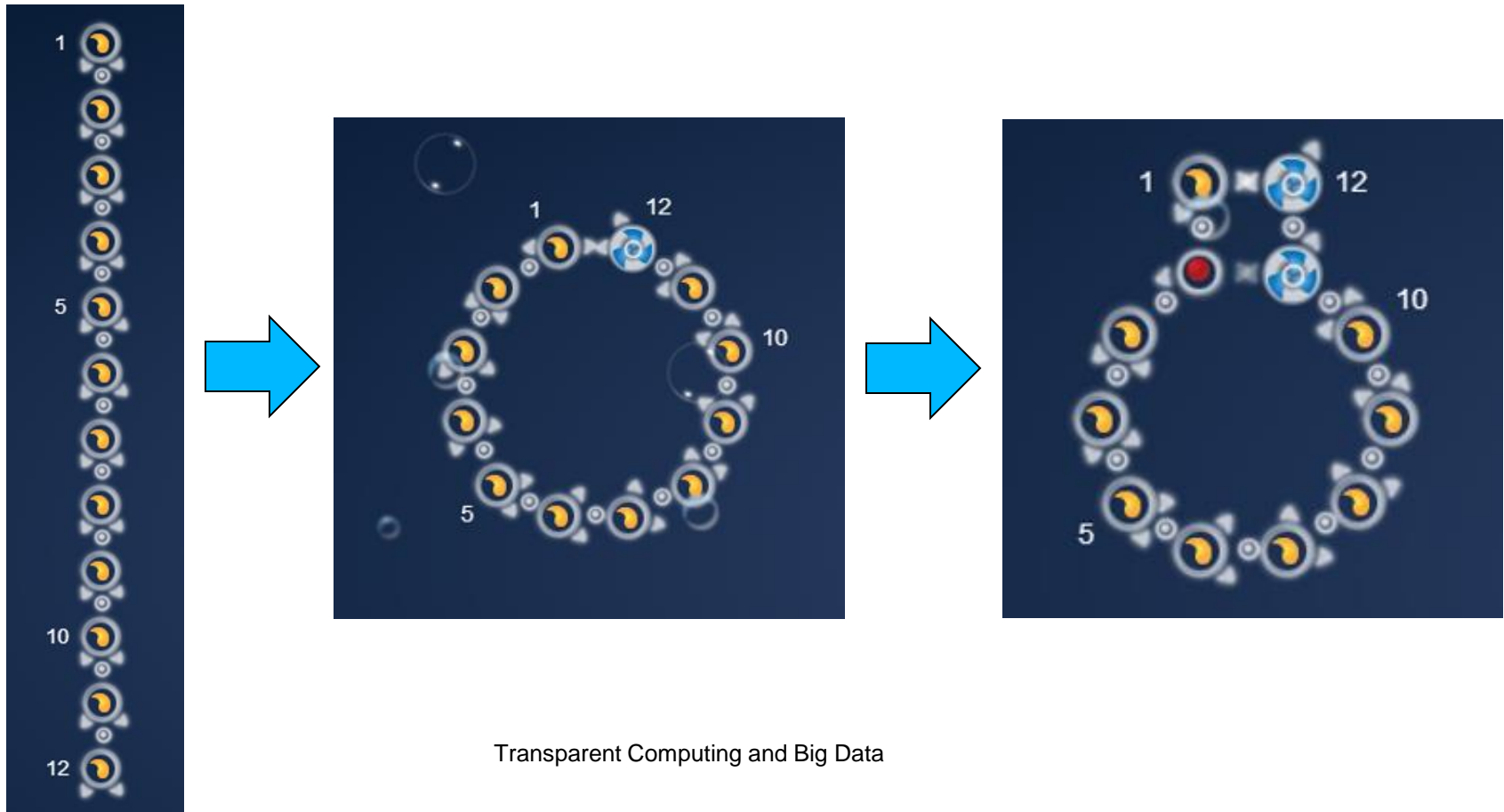
Matt Indykiewicz: prob [10:36 AM]
Spanda: ... [10:36 AM]
Matt Indykiewicz: - [10:37 AM]
Matt Indykiewicz: my random quote [10:37 AM]
Spanda: -- [10:37 AM]
Matt Indykiewicz: you kill the people driving [10:37 AM]
Spanda: my random quote: [10:38 AM]
Spanda: "Let me tell you something you already know. 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 classrooms

EteRNA: CMU, Stanford

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



EteRNA: CMU, Stanford

The screenshot shows the EteRNA website interface. At the top, the logo "eterna" is displayed with the tagline "Make Molecules Advance Science". Navigation links include "username", "password", "Log in", "Register", "Lost password?", "Facebook connect", "Roadmap", "Puzzle", "RNA Lab", "Community", and "About".

The main content area is titled "Simple Hairpin" and features a "Follow" button. It displays a puzzle progress indicator: a person icon, "47875", a coin icon, "25", and the date "16 Nov 2010". A large green "Play Now" button is prominent. Below this is the "Structure Notation" section, which shows the sequence "((((.....)))))" and a grid of dots representing the RNA structure, with minus and plus buttons for zooming.

The "Puzzle Description" section contains the following text:

""Difficulty Level""
Welcome to the Hairpin RNA Series.

""Science""
A small hairpin RNA or short hairpin RNA (shRNA) is a sequence of RNA that makes a tight hairpin turn that can be used to silence gene expression via RNA interference. shRNA uses a vector introduced into cells and utilizes the U6 or H1 promoter to ensure that the shRNA is always expressed. This vector is usually passed on to daughter cells, allowing the gene silencing to be inherited. The shRNA hairpin structure is cleaved by the cellular machinery into siRNA, which is then bound to the RNA-induced silencing complex (RISC). This complex binds to and cleaves mRNAs which match the siRNA that is bound to it.

On the right side, there is a chat window titled "Anonymou.." with "Chat" and "Players Online (51)". The chat log shows messages from users like "hoglahoo", "TomoeUzumaki", "Rayrane", "Kaze Tachinu", "cohenf", and "Tachinu*". Below the chat is a "Please log in to chat" prompt and an "Open Chat Window" link.

Below the chat is the "Latest News" section, which lists several news items with links, such as "Cloud Lab 11, 12, and 13 results published", "Eterna results published in PNAS", and "Crowdsourcing errors in any puzzle description".

At the bottom right, there is a "Latest Wiki" section with a link to "[Wiki]What is RNA?".

EteRNA: CMU, Stanford

MISSION!

 **Your RNA must fold into the structure in white outline.**

 **You must have 2 or more G-U pairs.**

1 16
15
5 10

Transparent Computing and Big Data

A U G C

The screenshot displays the EteRNA game interface. At the top left, a tilted white box with a black border contains the word "MISSION!". Below it, two red icons with white text provide instructions: a red exclamation mark icon followed by "Your RNA must fold into the structure in white outline.", and a "Click here to start!" icon followed by "You must have 2 or more G-U pairs.". In the center, a secondary structure diagram of an RNA sequence is shown, consisting of 16 nucleotides arranged in a stem-loop structure. The 5' and 3' ends are labeled with the numbers 5 and 10, respectively. The top of the stem is labeled with 1 and 16, and the middle of the stem with 15. The bottom of the stem is labeled with 5 and 10. At the bottom of the screen, there is a navigation bar with a "Transparent Computing and Big Data" logo and a sequence of nucleotides: A, U, G, C.

EteRNA: CMU, Stanford

The screenshot displays the EteRNA web interface. At the top left, it says "Hairpin RNA" with a star icon. The top right has a "Register" link. Below the title bar are navigation tabs: "Roadmap", "Puzzle", "RNA Lab", "Community", and "About EteRNA".

On the left side, there is a "Total: -1.3 kcal" display and two red buttons with icons: one with a blue and red dot and the number "1/2", and another with a red hairpin icon.

The central area shows a large RNA hairpin structure with numbered nucleotides (1, 5, 10, 15, 16) and various colored icons (yellow, red, green, blue) representing different nucleotides or mutations.

On the right side, there is a "Chat" window titled "Chat" with "Players Online (51)". The chat history includes:

- TomoeUzumaki: it's oscar nominated [5:02 PM]
- hoglahoo: is it a japanese animated historical fantasy adventure film? [5:02 PM]
- TomoeUzumaki: yes [5:03 PM]
- hoglahoo: got it [5:03 PM]
- steven123505: sounds perfect for you [5:05 PM]
- jedg: Hi Rachel and Franki [5:07 PM]
- RL: Hey jed its Ryan [5:07 PM]
- jedg: sup Ryan [5:08 PM]

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

At the bottom of the interface, there is a toolbar with various icons for navigation and editing, including a target icon, a power icon, and a settings icon. Below the toolbar, the sequence "A 0 U 1 G 2 C" is visible, with the "2" highlighted in a green box.

EteRNA: CMU, Stanford

The screenshot displays the EteRNA game interface. At the top left, the puzzle is titled "Hairpin RNA" with a star icon. The top right contains navigation links: "Roadmap", "Puzzle", "RNA Lab", "Community", "About EteRNA", and "Register".

On the left side, a "Total: -3.1 kcal" box is visible. Next to it are two icons: a pair of colored dots (red and blue) with "2/2" below, and a beaker icon.

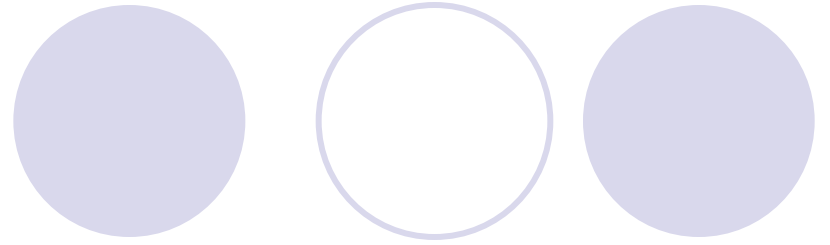
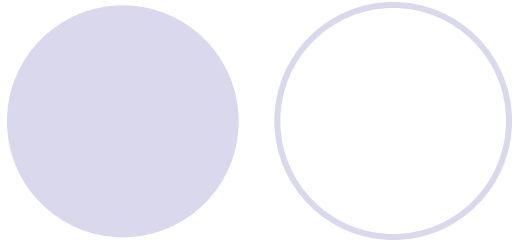
The central area features a vertical RNA hairpin structure composed of 16 nucleotides, numbered 1 to 16. The structure is partially colored with various icons: a yellow crescent moon, a blue gear, a red circle, a green leaf, and a white square. The bottom of the structure is a loop containing several yellow crescent moons.

On the right side, there is a "Chat" window titled "Chat" with "Players Online (50)". The chat log shows the following messages:

- grade or something? [5:09 PM]
- TomoeUzumaki: nope [5:09 PM]
- RL: most of them are [5:09 PM]
- jedg: kk who is tho [5:09 PM]
- hoglahoo: I was in your grade 20 years ago [5:10 PM]
- hoglahoo: does that count? [5:10 PM]
- TomoeUzumaki: haha [5:10 PM]
- jedg: haha so ur 77 [5:10 PM]
- hoglahoo: yes, I'm 77 [5:10 PM]
- RL: You me franki and rachel [5:10 PM]

Below the chat window, a message reads "Please log in to chat".

At the bottom of the interface, there is a toolbar with various icons for navigation and puzzle manipulation. On the right side of the toolbar, a sequence of nucleotides is displayed: A 1, U 2, G 2, C.



Galaxy Zoo

Fine-grained Recognition

APPS: IMAGE PROCESSING

Galaxy Zoo: Zooniverse

CLASSIFY

STORY

SCIENCE



DISCUSS

PROFILE

LANGUAGE



Classify



UKIDSS



Invert

Examples

Restart

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



DISCUSS

PROFILE

LANGUAGE



Classify



UKIDSS



Invert

Examples

Restart

ROUND

How rounded is it?



Completely round



In between



Cigar shaped

GalaxyZoo: Zooniverse

CLASSIFY

STORY

SCIENCE

GALAXY ZOO

DISCUSS

PROFILE

LANGUAGE

Classify



UKIDSS



Invert

Examples

Restart

SPIRAL

How many spiral arms are there?



1



2



3



4

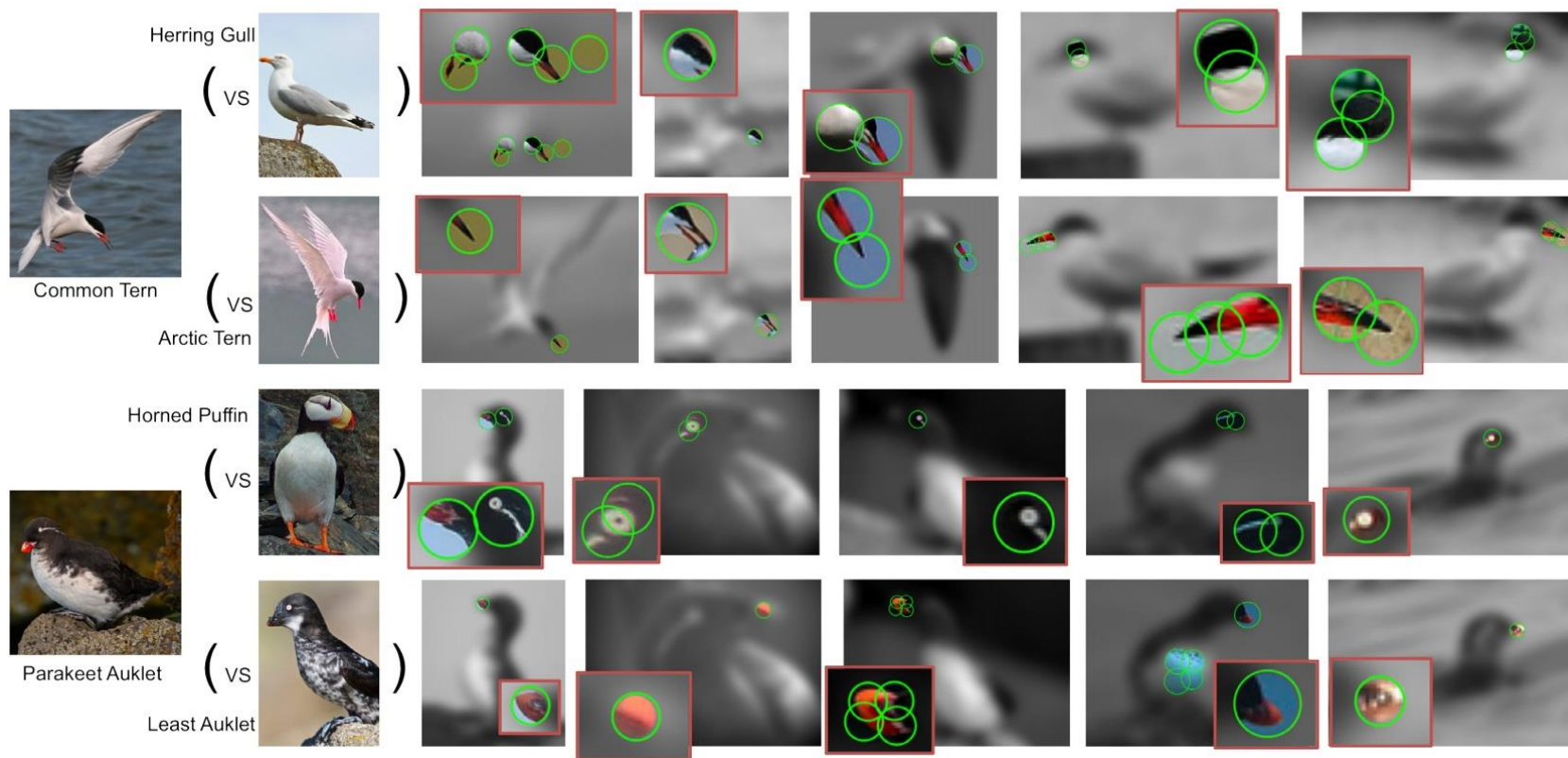


More than 4

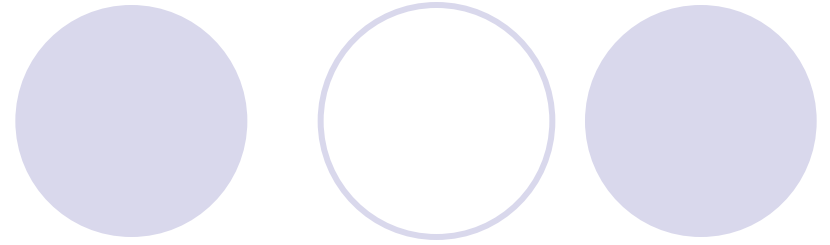
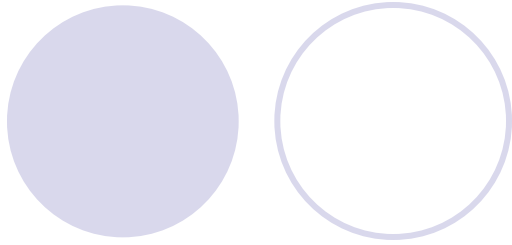


Can't tell

Fine-Grained Recognition



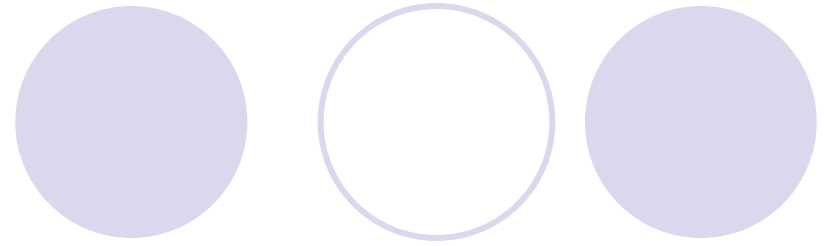
J. Deng et al, "Fine-Grained Crowdsourcing for Fine-Grained Recognition," CVPR 2013



GWAP.com
reCAPTCHA
OnToGalaxy

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.

make a new game

Tree

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

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.
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

Transparent Computing and Big Data

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

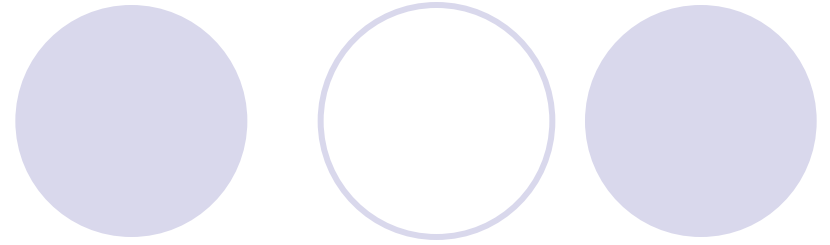
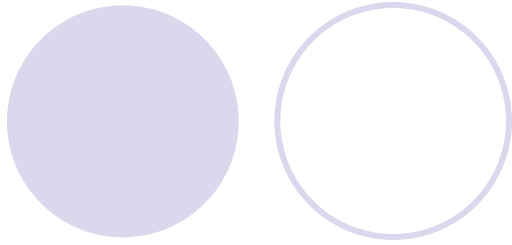


→ LEARN HOW reCAPTCHA WORKS

USE reCAPTCHA ON YOUR SITE

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

NEW See how accurate reCAPTCHA is at digitizing content!



Sequential

Iterative and Parallel

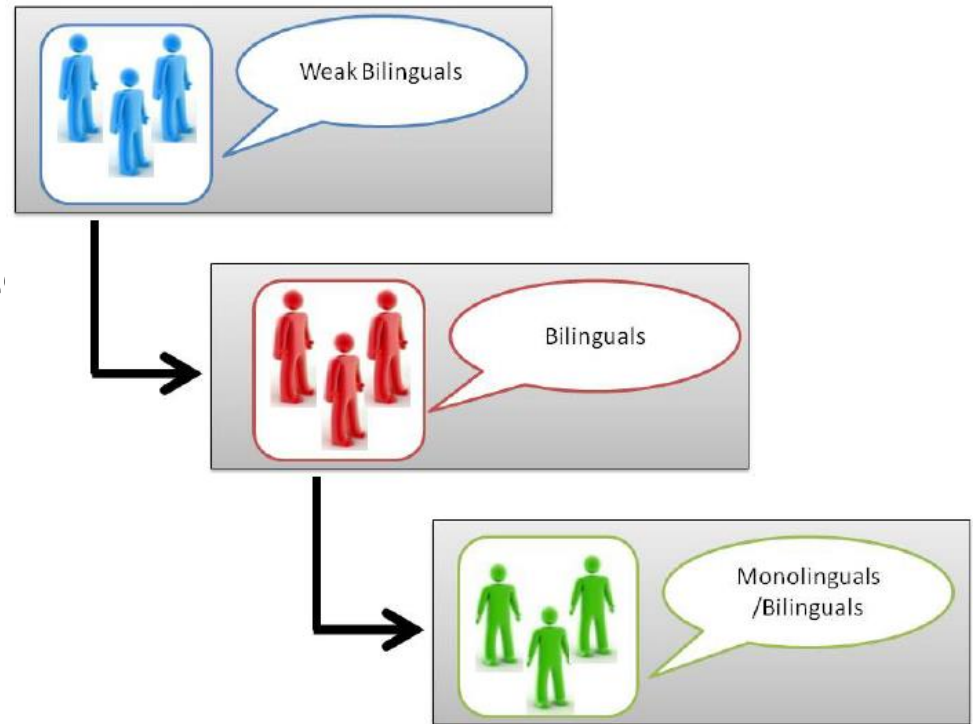
Divide-and-Conquer and Aggregate

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

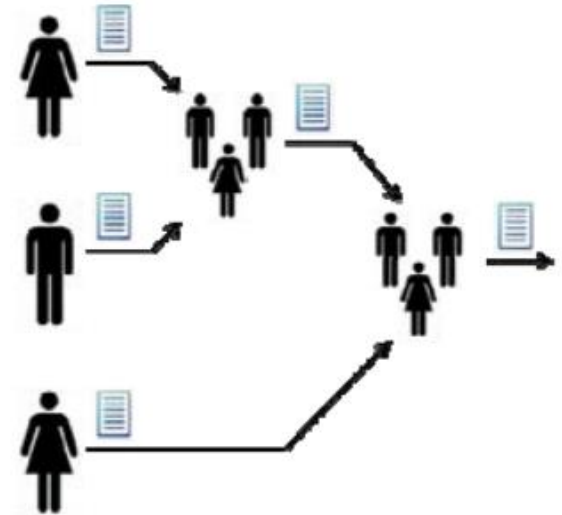
Iterative and Parallel

- Iterative improve and vote

The screenshot shows two windows from a Mozilla Firefox browser. The left window, titled "Handwriting Recognition Task", displays a handwritten note and instructions: "Please improve the transcription of this handwriting. People will vote whether to approve your changes." Below the text, a transcription is shown with several errors in parentheses: "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 (phony). You do (7) have good (points), but they got lost amidst the (writing). (signature)". A "Submit" button is at the bottom.

The right window, titled "MTask", shows the same handwritten note and instructions: "Please select the better transcription for this handwriting. Differences are highlighted in yellow." It presents two alternative transcriptions for the user to choose from. The first alternative has "several" highlighted in yellow, and the second has "not" and "a few" highlighted in yellow.

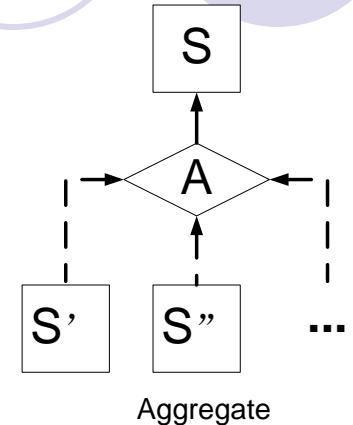
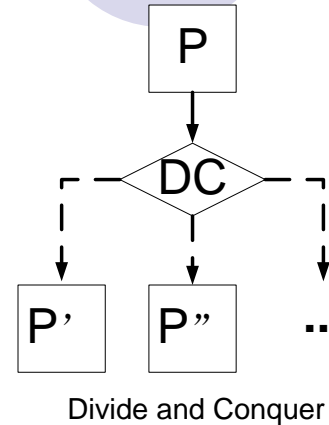
Below the screenshots, a diagram illustrates the iterative process. It shows a sequence of human icons (stick figures) connected by arrows, representing the flow of work. The first icon is labeled "improvement \$0.05". The second icon is labeled "3 votes @ \$0.01". The process continues with more icons and arrows, ending with an ellipsis "...".



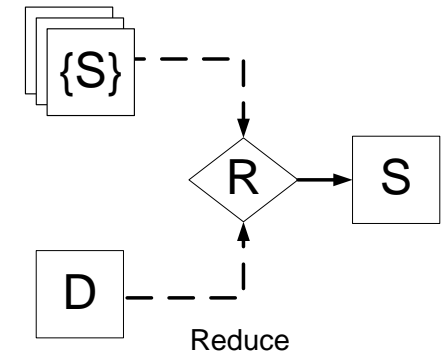
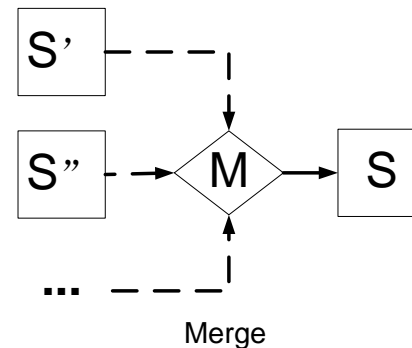
G. Little et al, "Exploring Iterative and Parallel Human Computation Processes," HCOMP 2010

Divide-and-Conquer and Aggregate

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

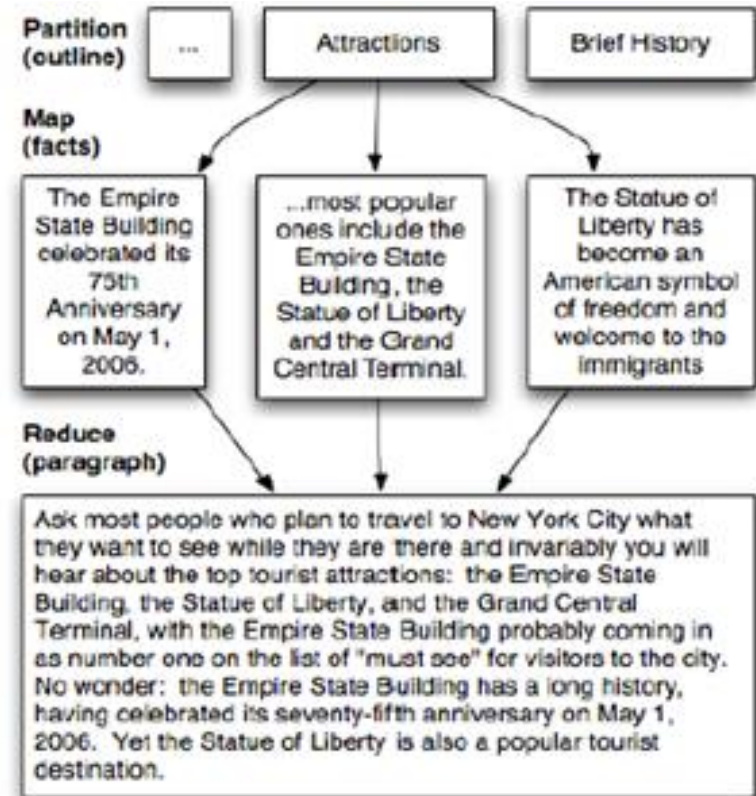
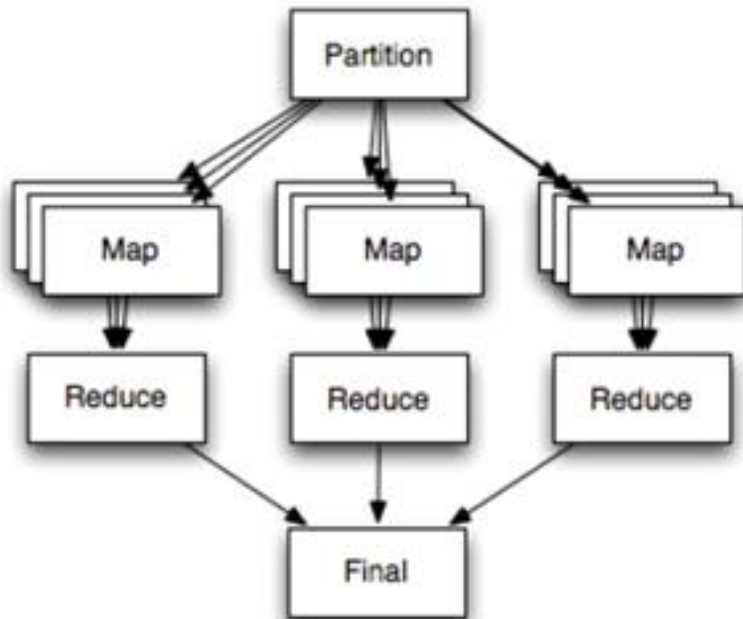


- Two special aggregates
 - Merge
 - Reduce

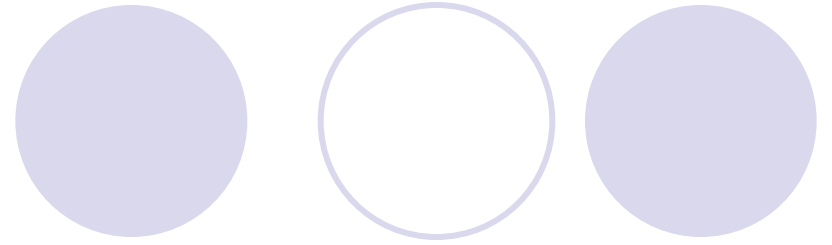
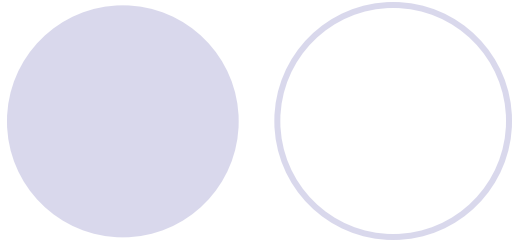


P. Minder et al, "Crowdlang - First Steps Towards Programmable Human Computers for General Computation," AAI 2011.

Map and Reduce: A Special Case



A. Kittur et al, "Crowdforge: Crowdsourcing complex work," UIST 2011



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

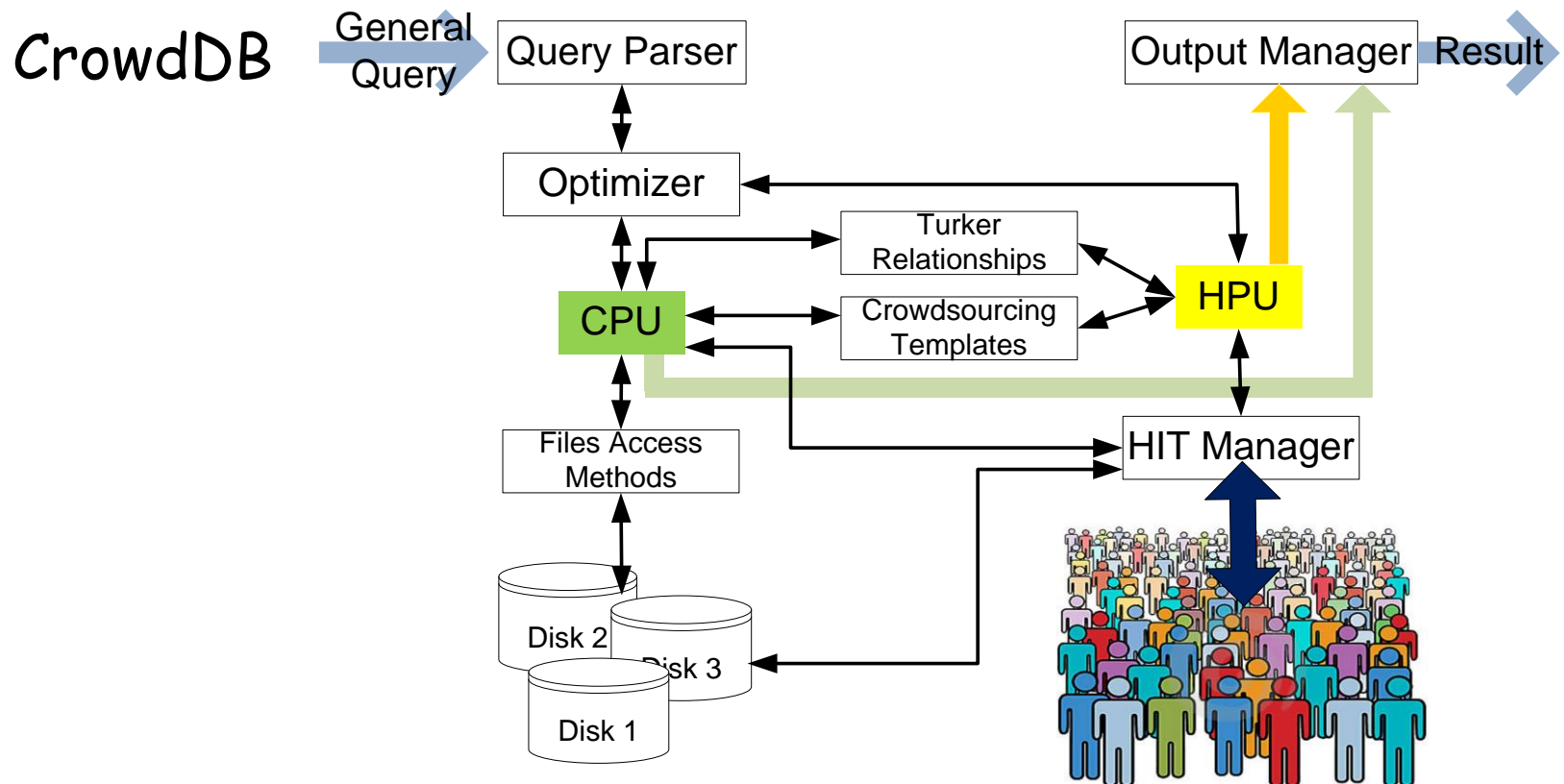
- Trade-offs: time, cost, and quality
 - Max algorithm with human error (with a probability)
 - Maximize quality (via redundancy) subject to cost and time

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

- 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

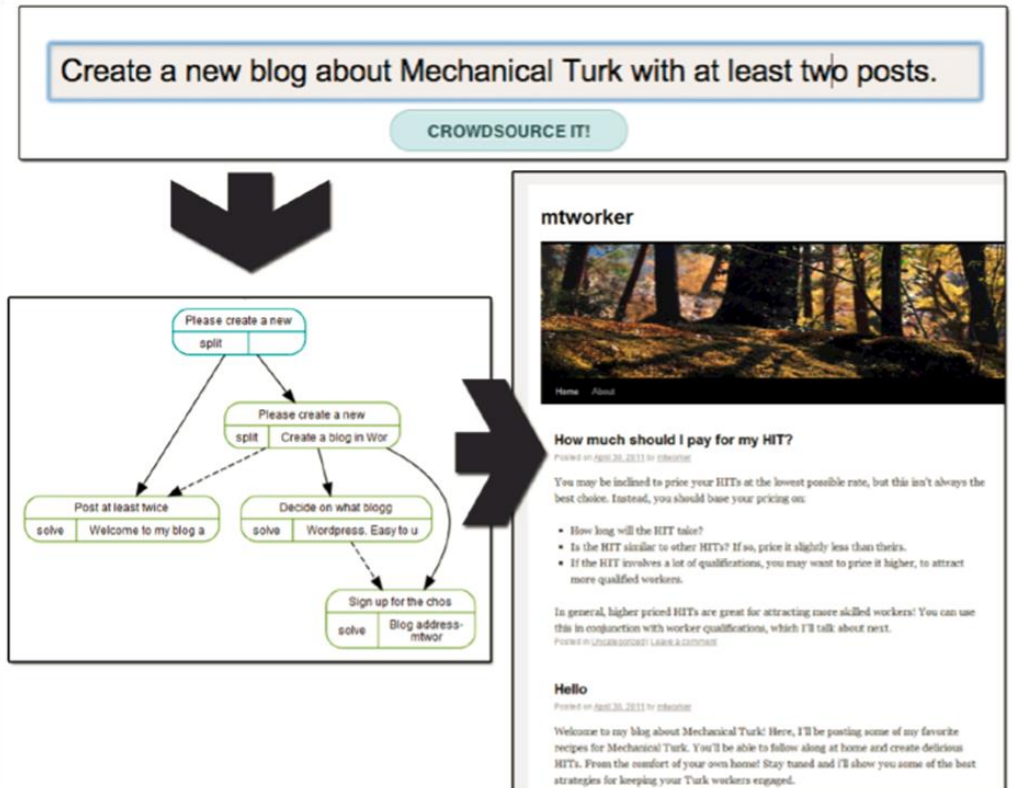


M. Franklin et al, "Crowddb: Answering Queries with Crowdsourcing," SIGMOD 2011

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

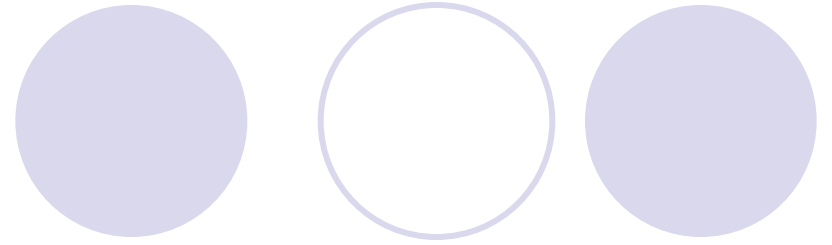
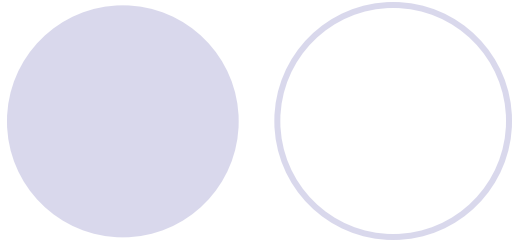


A. Badanidiyuru et al,
"Bandits with Knapsacks:
Dynamic Procurement for
Crowdsourcing," 2013

Beyond Independent Workers

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



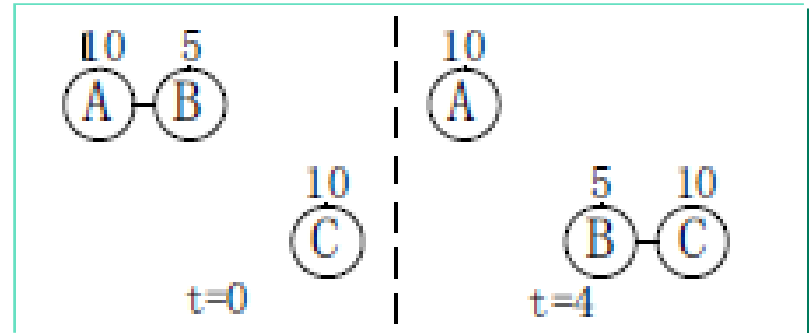


Computational Surplus Around
QQ Example

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

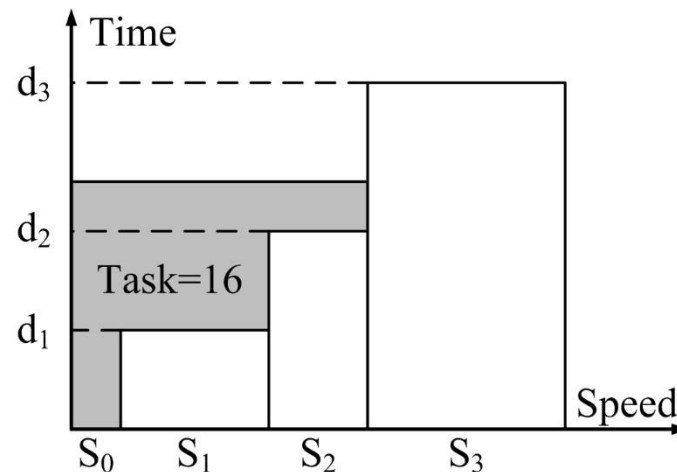
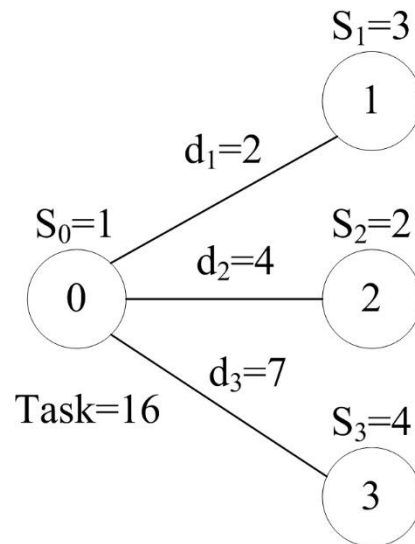


S. Zhang et al, "Minimum Makespan Workload Dissemination in DTNs: Making Full Utilization of Computational Surplus Around," MobiHoc 2013

Water Filling Schedule

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

- d_i : response + reply



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

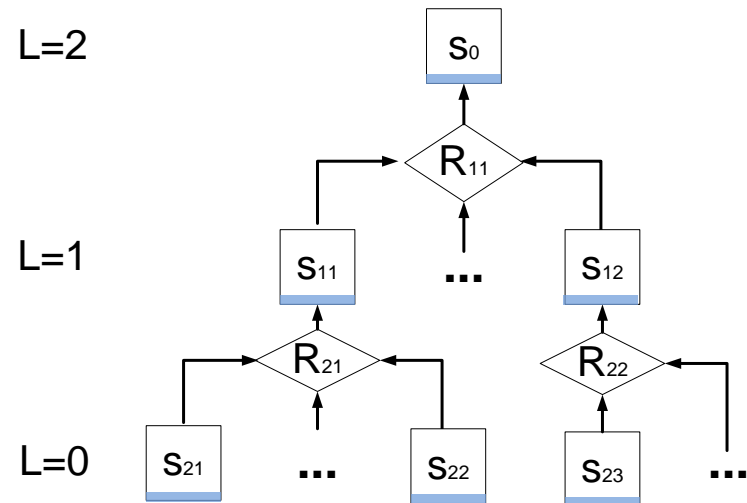


Iterative Request/Reply (reduce)

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

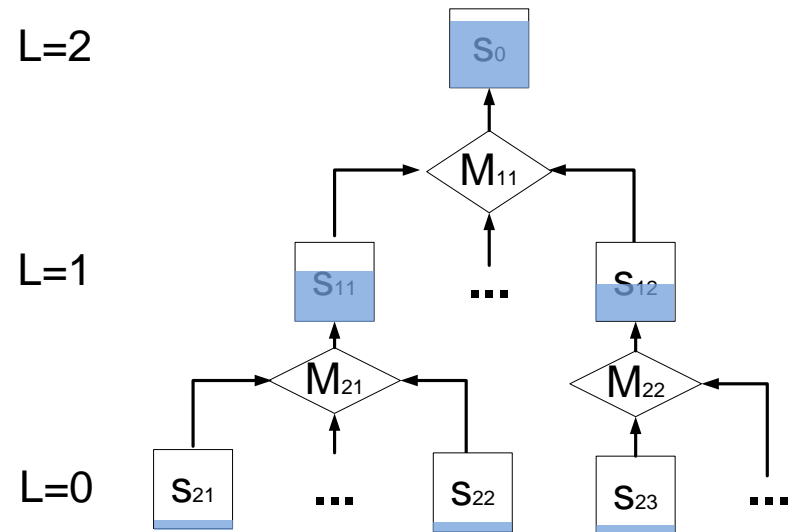
- When $L = 0$, return the total number of QQ friends

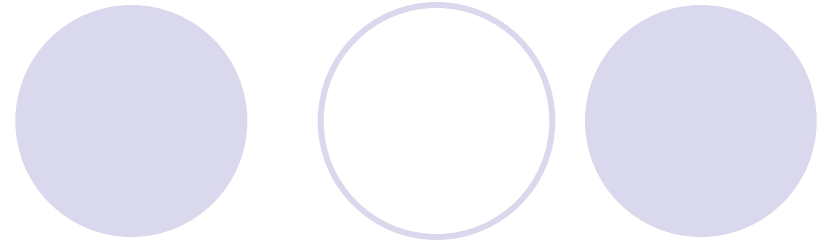
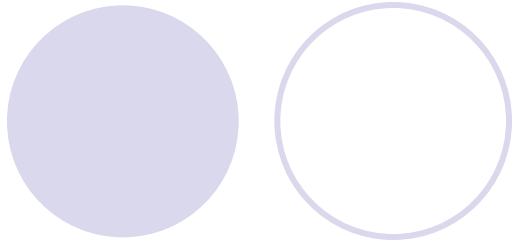
- When $L > 0$, do the following:
 - Forward this request to all QQ friends
 - After receiving the first 10 replies, compute the average number of friends, and send them back to me



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 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

