

SOCIAL SOFTWARE ENGINEERING

Swapneel Sheth
@swapneel

COMS W4156: Advanced Software Engineering
Fall 2013

DEVELOPER'S PROBLEM SOLVING
SEQUENCE: 1. GOOGLE 2.
COWORKERS 3. #STACKOVERFLOW 4.
RTFM. 5. THINK

SOFTWARE ENGINEERING

- Not a **stand-alone** activity
- People use forums, Stack Overflow, etc. to debug code and find solutions

SOCIAL SOFTWARE ENGINEERING

- Software Engineering has always been **social** - people work in large teams and collaborate with each other
- Make Software Engineering **easier** by using the Social Aspects of software development

CODEBOOK (MICROSOFT RESEARCH)

- **Social Networking over Code**

- Use Source Code artifacts to provide useful information

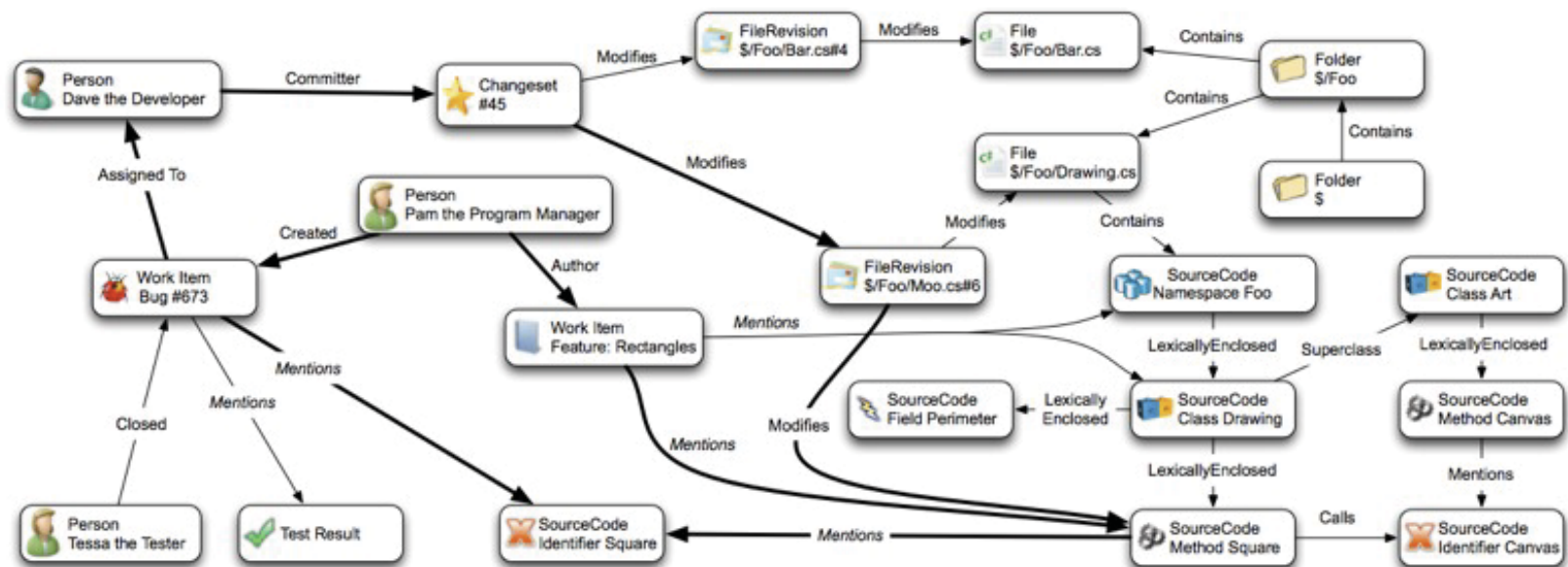
- Papers

- Begel and Zimmermann, Keeping up with your Friends: Function Foo, Library Bar.DLL, and Work Item 24, Web2SE 2010
- Begel, Phang, and Zimmermann, Codebook: Discovering and Exploiting Relationships in Software Repositories, ICSE 2010

CODEBOOK

1. Find out who the **most relevant** engineers (developers, testers, program managers, operations, leads, etc.) are in order to contact them
2. Find an **expert** to talk to who knows a lot about a feature, API, product or service
3. Get a list of servers, directories and repositories where the **related** code, bug reports, work items, specifications, etc. are located
- 4....

CODEBOOK



CODEBOOK

CODEBOOK FEEDS

NEWS COMPONENTS BUGS TESTS ANALYSES SEARCH



LIBRARY SORT.DLL

Checkins Building Tests People Projects

New [bug report 8724](#) "Quicksort does not work anymore after upgrade of sort.dll to latest [build 2354](#)". 8 hours ago by [Xia](#).

Performance of [Build 2356](#) increased by 25% compared to [Build 2353](#). 7 days ago by [TFS-PerfTest](#).

[Build 2356](#) successfully completed. 3 days ago by [TFS-Build](#).

Checkin: "[Reimplemented Quicksort. Much faster now](#)". 1 days ago by [Chuck](#).

Performance of [Build 2353](#) decreased by 15% compared to [Build 2352](#). 4 days ago by [TFS-PerfTest](#).

[Sort.dll](#) in [Build 2353](#) passed all test cases. 5 days ago by [TFS-Test](#).

FxCop reported [6 design warnings](#) and [7 performance warnings](#) in [Build 2353](#). 6 days ago by [FxCop](#).

[Build 2353](#) successfully completed. 6 days ago by [TFS-Build](#).

Fixed [bug report 8723](#) "NullPointerExpection in QuickSort". 7 days ago by [Caroline](#).

Checkin: "[Quicksort now sorts collections](#)". December 31st, 2009 by [Chuck](#).

Checkin: "[New shuffle method](#)". December 29th, 2009 by [Chuck](#).

Checkin: "[MergeSort completed](#)". December 21st, 2009 by [Caroline](#).

[Sort.dll](#) in [Build 2352](#) passed all test cases. December 20th, 2009 by [TFS-Test](#).

[Build 2352](#) successfully completed. December 20th, 2009 by [TFS-Build](#).

Checkin: "[Fixed Null Pointer Exception](#)". December 19th, 2009 by [Chuck](#).

[Build 2351](#) failed with errors. December 19th, 2009 by [TFS-Build](#).

Checkin: "[New data structures for sort algorithms](#)". December 19th, 2009 by [Chuck](#).

[Edward](#) moved to the [SQL Server/Java](#) team and is now reporting to [Karl Heine](#). December 10th, 2009 by [Active Directory](#).

New [bug report 8723](#) "Quicksort does sort large arrays incorrectly". December 4th, 2009 by [Richard](#).

[Caroline](#) joined the team as a developer. December 3rd, 2009 by [Caroline](#).

Checkin: "[Improved the performance of MergeSort](#)". November 23rd, 2009 by [Chuck](#).

Checkin: "[New reverse function](#)". November 19th, 2009 by [Edward](#).

Checkin: "[First implementation of MergeSort](#)". November 12th, 2009 by [Edward](#).

Other news...

[Subscribe to RSS feed of SORT.DLL](#)

▼ DETAILS

Primary contact: [Chuck](#)
Specification: [http://charlieps/](#)
Documentation: [http://codebook/](#)
TFS Instance: [http://sort.dll:8080/](#)
Builds: [View builds](#)

▼ TEAM (6)



▼ INTERNAL CUSTOMERS (33)



▼ EXTERNAL CUSTOMERS (725)

▼ SORT.DLL DEPENDS ON (2)

-  [Google.dll](#)
243 calls, 93 references
Team: Visual Studio
-  [system32.dll](#)
42 calls, 37 references
Team: Windows/System32

▼ DEPENDS ON SORT.DLL (180)

-  [excel.dll](#)
141 calls, 28 references
Team: Office/Excel
-  [internetexplorer.dll](#)
85 calls, 33 references
Team: MSN Services/IE
-  [outlook.dll](#)
45 calls, 37 references
Team: Office/Outlook

[More...](#)

GAIL MURPHY'S GROUP - UNIVERSITY OF BRITISH COLUMBIA

- Fritz and Murphy, Using information fragments to answer the questions developers ask, ICSE 2010

DEVELOPERS' QUESTIONS

Who is working on what (people specific)

1. Who is working on what?*
2. What are they [coworkers] working on right now?*
3. How much work [have] people done?*
4. Who changed this [code], focused on person?*
5. Who to assign a code review to? / Who has the knowledge to do the code review?
6. What have people been working on?*
7. Which code reviews have been assigned to which person?*

DEVELOPERS' QUESTIONS

Changes to the code (code specific)

8. What is the evolution of the code?
9. Why were they [these changes] introduced?*
10. Who made a particular change and why?
11. What classes has my team been working on?*
12. What are the changes on newly resolved work items related to me?
13. Who is working on the same classes as I am and for which work item?
14. What has changed between two builds [and] who has changed it?*
15. Who has made changes to my classes?
16. Who is using that API [that I am about to change]?*
17. Who created the API [that I am about to change]?*
18. Who owns this piece of code? / Who modified it the latest?*
19. Who owns this piece of code? / Who modified it most?*
20. Who to talk to if you have to work with packages you haven't worked with?
21. What classes have been changed?*
22. [Which] API has changed (to see which methods are not supported any more)?*
23. What's the most popular class? [Which class has been changed most]?*
24. Which other code that I worked on uses this code pattern / utility function?
25. Which code has recently changed that is related to me?
26. How do recently delivered changes affect changes that I am working on?*
27. What code is related to a change?

DEVELOPERS' QUESTIONS

Broken builds

- 32. What caused this build to break? (Which change caused the stack trace?)
- 33. Who caused this build to break? (Who owns the broken tests?)
- 34. Who changed the test case most recently that caused the build to fail?
- 35. Which changes caused the tests to fail and thus the build to break?

DEVELOPERS' QUESTIONS

Other Questions

- 42. How is the team organized?*
- 43. Who has made changes to [a] defect?*
- 44. Who has made comments in defect?*
- 45. [What is] the collaboration tree around a feature?*
- 46. Which conversations in work items have I been mentioned?*

ANSWERING SUCH QUESTIONS

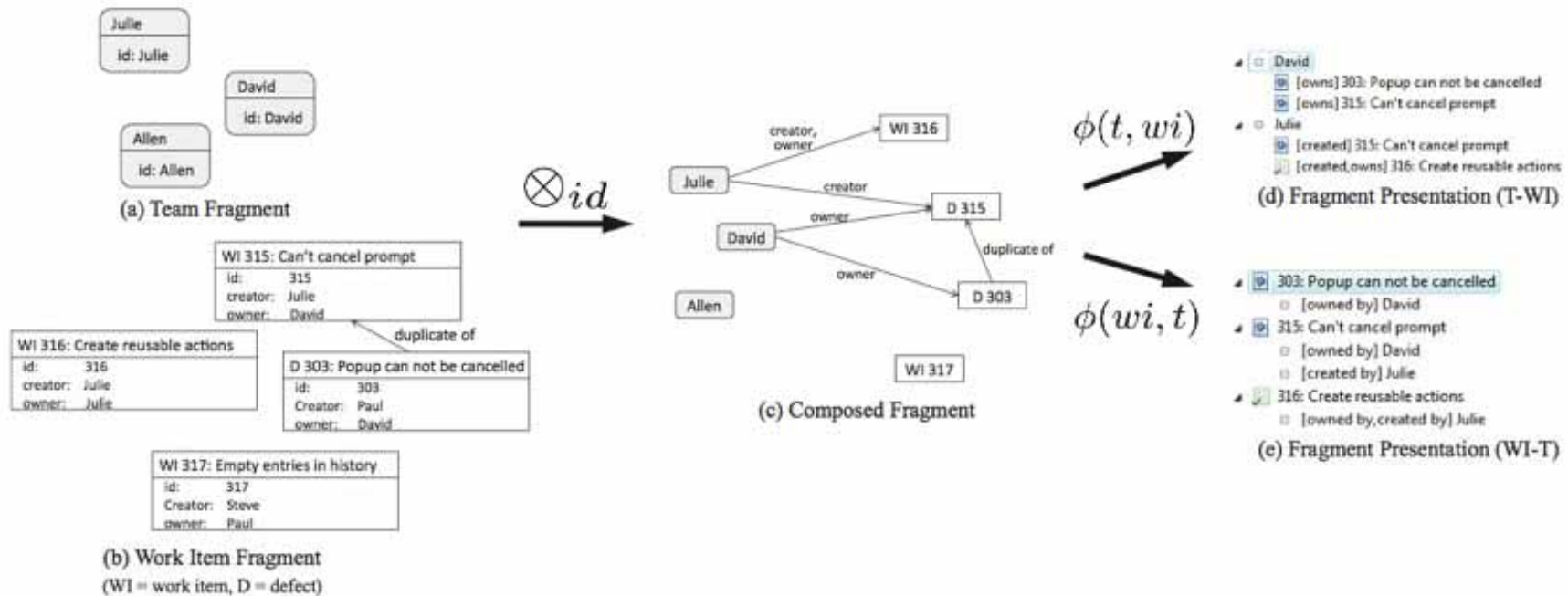


Figure 1: Approach to Answer the Question "What have people been working on?"

PUZZLE

- 29 pairs of spouses
- At least 1 man cheated
- Each woman must determine if her husband was faithful
- 2 nights, 1 question per night per woman
- 2 possible gun shots
- What strategy can the women come up with?

SOCIAL SOFTWARE ENGINEERING

- Using Software Engineering for **social solutions** to problems



...WITH A LITTLE HELP FROM MY FRIENDS

WISDOM OF THE CROWDS

- Wikis
 - Wikipedia, Wikitravel, Wiktionary, ...
- Social bookmarking
 - Reddit, Digg, Delicious, ...
- Social Games
 - MMOGs, Farmville, ...
- ...

RECOMMENDATIONS, RECOMMENDATIONS, RECOMMENDATIONS!!!

- Things to buy (Amazon)
- Movies to watch (IMDb, Netflix)
- Friends you may know (Facebook)
- Music you may like (Last.FM)
- Jokes you may like (Jester)
- ...

COLLABORATIVE FILTERING

- Filter information based on interests

COLLABORATIVE FILTERING

- Why filter?
 - To make it more relevant, useful
- What information?
 - Likes, dislikes, preferences
- What interests?
 - Movies, Music, Friends, ...

COLLABORATIVE FILTERING

- How?
 - Memory-Based
 - Find similarity between users or items
 - Cosine Similarity, Pearson Correlation, Nearest Neighbor, ...
 - Model-Based
 - Use data mining or machine learning algorithms
 - Bayesian networks, Clustering, Latent Semantic Models, ...

NETFLIX & NETFLIX PRIZE

- Movie and TV show rentals
- 10 million subscribers
- 100,000 titles on DVD in 2009
- Announced the Netflix Prize in 2006

NETFLIX PRIZE

- Improve the recommendations used by Netflix by **10%**
- Grand Prize of **\$1,000,000**
- Released movie rating data
 - ~100,000,000 movie ratings
 - ~500,000 users
 - ~20,000 movies

NETFLIX PRIZE

- Techniques used from varying fields: computer science, statistics, machine learning, psychology
- Won by “BellKor's Pragmatic Chaos” in June 2009
- Team Members - 2 from AT&T Labs, 1 from Yahoo!, 1 from Commendo Research and Consulting GmbH, 2 from Pragmatic Theory

PRIVACY TESTING

PRIVACY TESTING

- Code/API Bugs
- Policy Changes

SOCIAL TESTING

- Leverage one's friends on these websites to detect privacy violations
- Don't need access to source code
- Sheth and Kaiser, End-User Regression Testing for Privacy, Columbia University Technical Report, 2012

APPROACH

- Ask a friend to see what's visible
- Monitor this periodically
- Notify user if something changes
- Can detect both:
 - Over-sharing - This piece of information should have been private, but it can be viewed by others.
 - Under-sharing - This piece of information should have been public, but it cannot be viewed by others.

POSSIBLE REASONS

- Added more information manually
- Changed privacy settings either deliberately or accidentally
- Didn't do anything – there is a bug in the code or API, possibly due to code changes
- Didn't do anything – the social system made a wide policy change where this information for many or all of its users is now visible

CHANGING PRIVACY SETTINGS

====Apr 24 2012 and Apr 25 2012====

feed – Old: 2, 19, New: 1, 0

photos – Old: 2, 25, New: 1, 0

posts – Old: 2, 19, New: 1, 0

tagged – Old: 2, 5, New: 1, 0

videos – Old: 2, 1, New: 1, 0

locations – Old: 2, 1, New: 1, 0

(a) Turning On the Privacy Settings

====Apr 25 2012 and Apr 27 2012====

feed – Old: 1, 0, New: 2, 19

photos – Old: 1, 0, New: 2, 25

posts – Old: 1, 0, New: 2, 20

tagged – Old: 1, 0, New: 2, 4

videos – Old: 1, 0, New: 2, 1

locations – Old: 1, 0, New: 2, 1

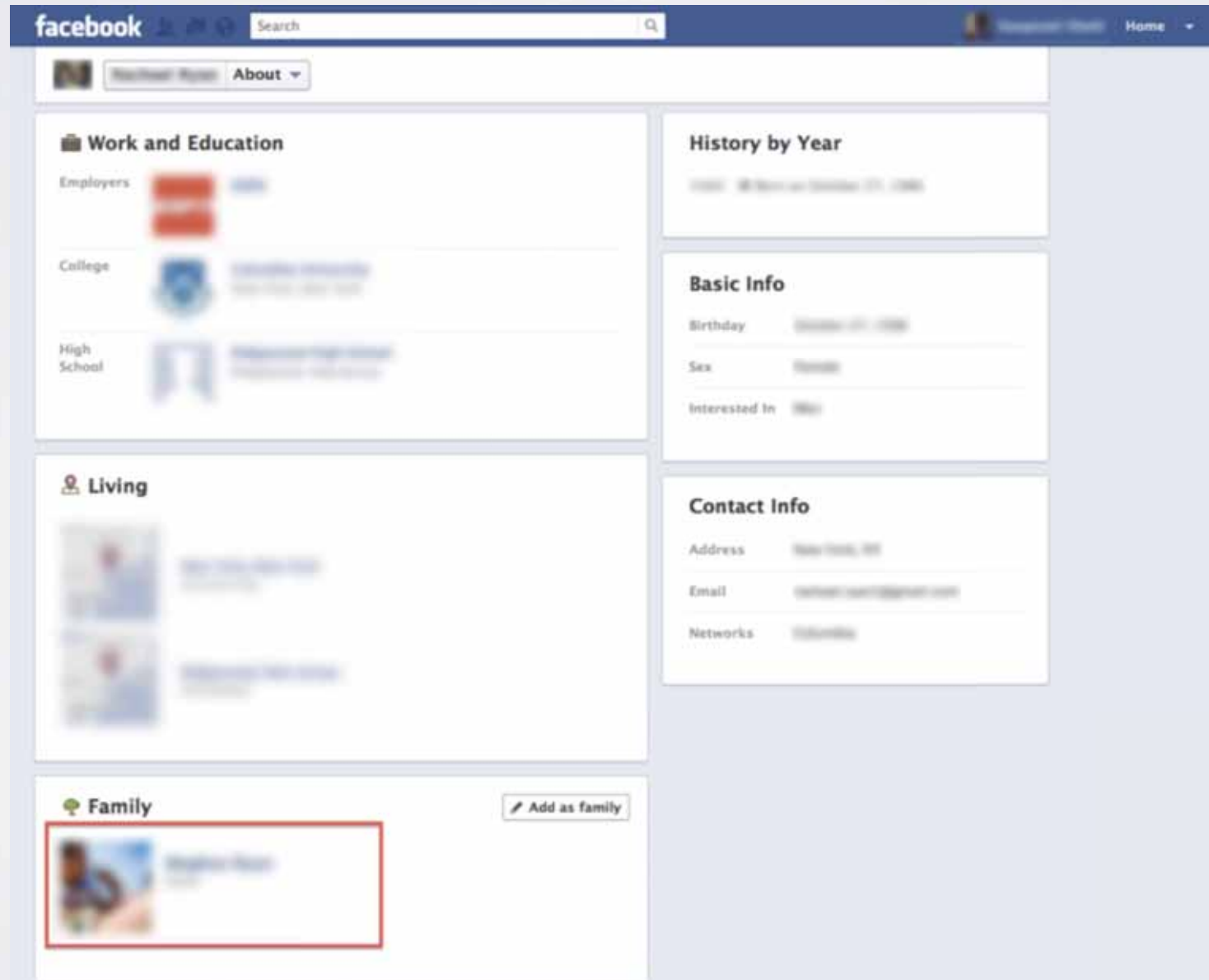
(b) Turning Off the Privacy Settings

FACEBOOK BUG

```
=====May 04 2012 and May 07 2012=====
feed - Old: 2, 21, New: 2, 20
posts - Old: 2, 15, New: 2, 14
=====May 11 2012 and May 14 2012=====
feed - Old: 2, 20, New: 2, 19
tagged - Old: 2, 11, New: 2, 10
=====May 18 2012 and May 21 2012=====
posts - Old: 2, 14, New: 2, 15
tagged - Old: 2, 10, New: 2, 8
=====May 22 2012 and May 23 2012=====
albums - Old: 2, 3, New: 1, 0
feed - Old: 2, 19, New: 1, 0
likes - Old: 2, 2, New: 1, 0
photos - Old: 2, 25, New: 1, 0
posts - Old: 2, 15, New: 1, 0
tagged - Old: 2, 8, New: 1, 0
family - Old: 2, 1, New: 1, 0
groups - Old: 2, 2, New: 1, 0
locations - Old: 2, 4, New: 1, 0
```

Listing 4: Facebook Family Bug — Output as seen from our tool

EXAMPLE BUGS FOUND



EXAMPLE BUGS FOUND

facebook DEVELOPERS Search Documentation Support Blog Apps

Graph API Explorer

Home » Tools » Graph API Explorer

Application: [X] Graph API Explorer

Access Token: X [Get Access Token](#)

[GET](#) [Submit](#)

```
{
  "data": [
  ]
}
```

[Like](#) [Send](#) 29,092 people like this. Be the first of your friends.

EXAMPLE BUGS FOUND

=====Jul 02 2012 and Jul 06 2012=====

feed	–	Old: 2, 14,	New: 2, 15
links	–	Old: 1, 0,	New: 2, 23
posts	–	Old: 2, 3,	New: 2, 5
tagged	–	Old: 2, 14,	New: 2, 15

PUBLIC LINKS

- 73 out of 516 users (11.05%)
- Beginning of July when links became public

INTERVIEW

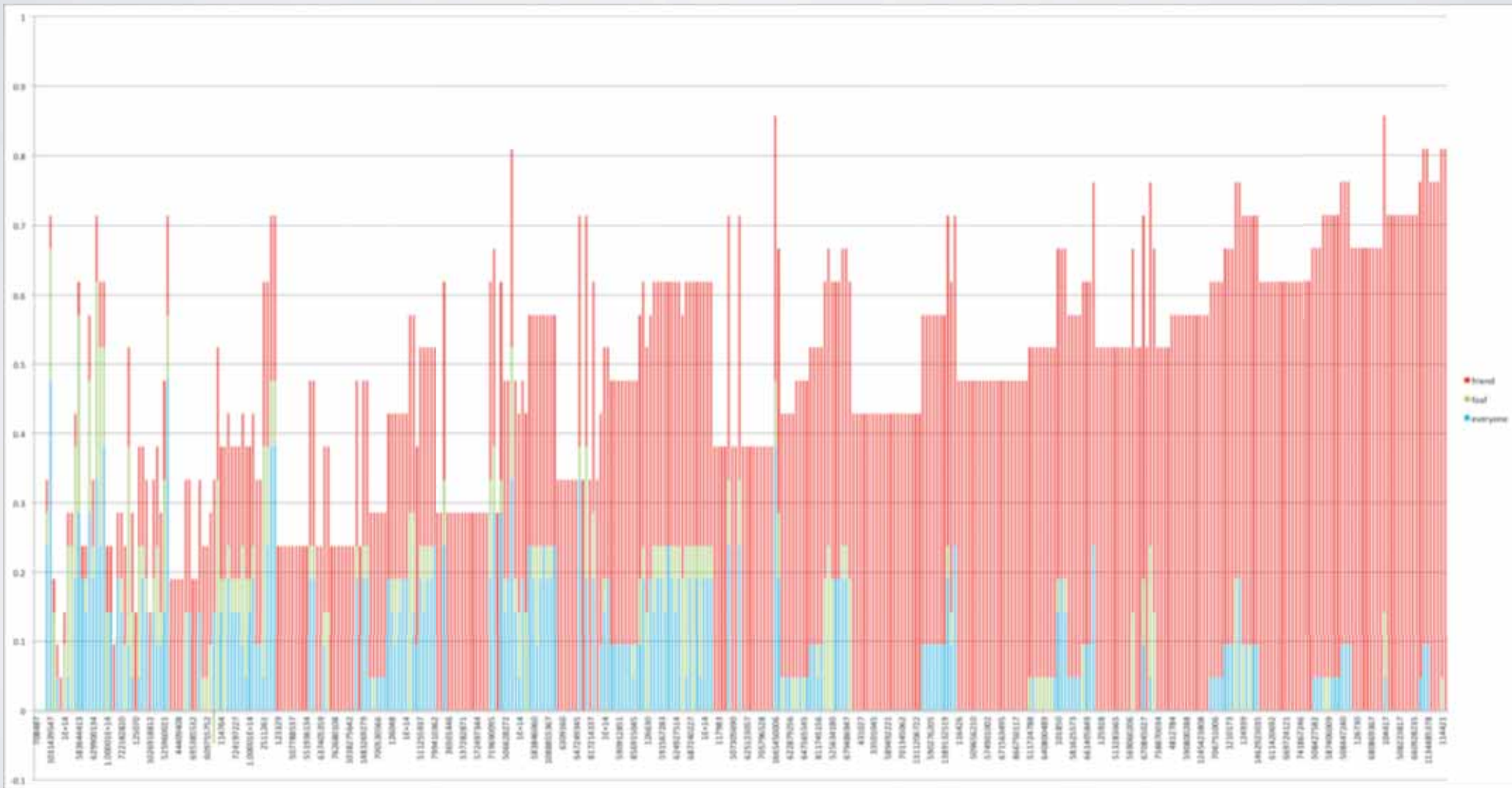
- “whoa [...] ok..... that is weird. ok, all of these links are valid, but, am surprised you can see them [...] I, as a developer, opened my account for developer access. it's the only way possible and I just thought I was authorizing that one app. they must have their permissions f***ed up [...] it's either that, or facebook **changed my settings automatically**”

MAKING PRIVACY UNDERSTANDABLE

N HEADS ARE BETTER THAN ONE

- Using crowdsourcing to improve user understanding of privacy settings
- Hopkins, Castaneda, Sheth, and Kaiser, N Heads are Better than One, Columbia University Technical Report, 2013

SAMPLE FACEBOOK DATA



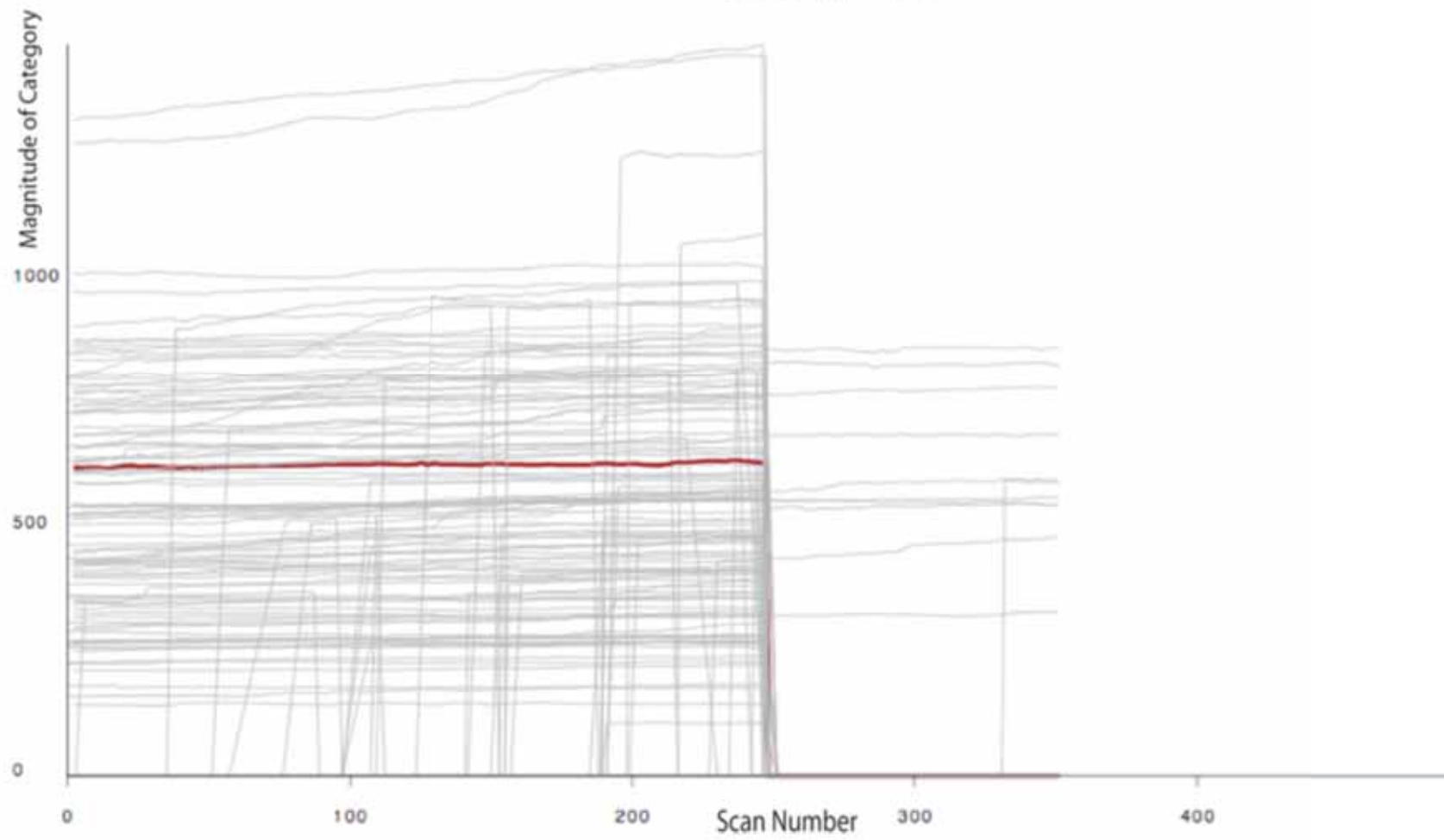
PRIVACY SETTINGS

- List of all privacy settings (current default, for most systems)
- Coarse Grained (Easy/Medium/Hard)
- Short survey
- Crowdsourced

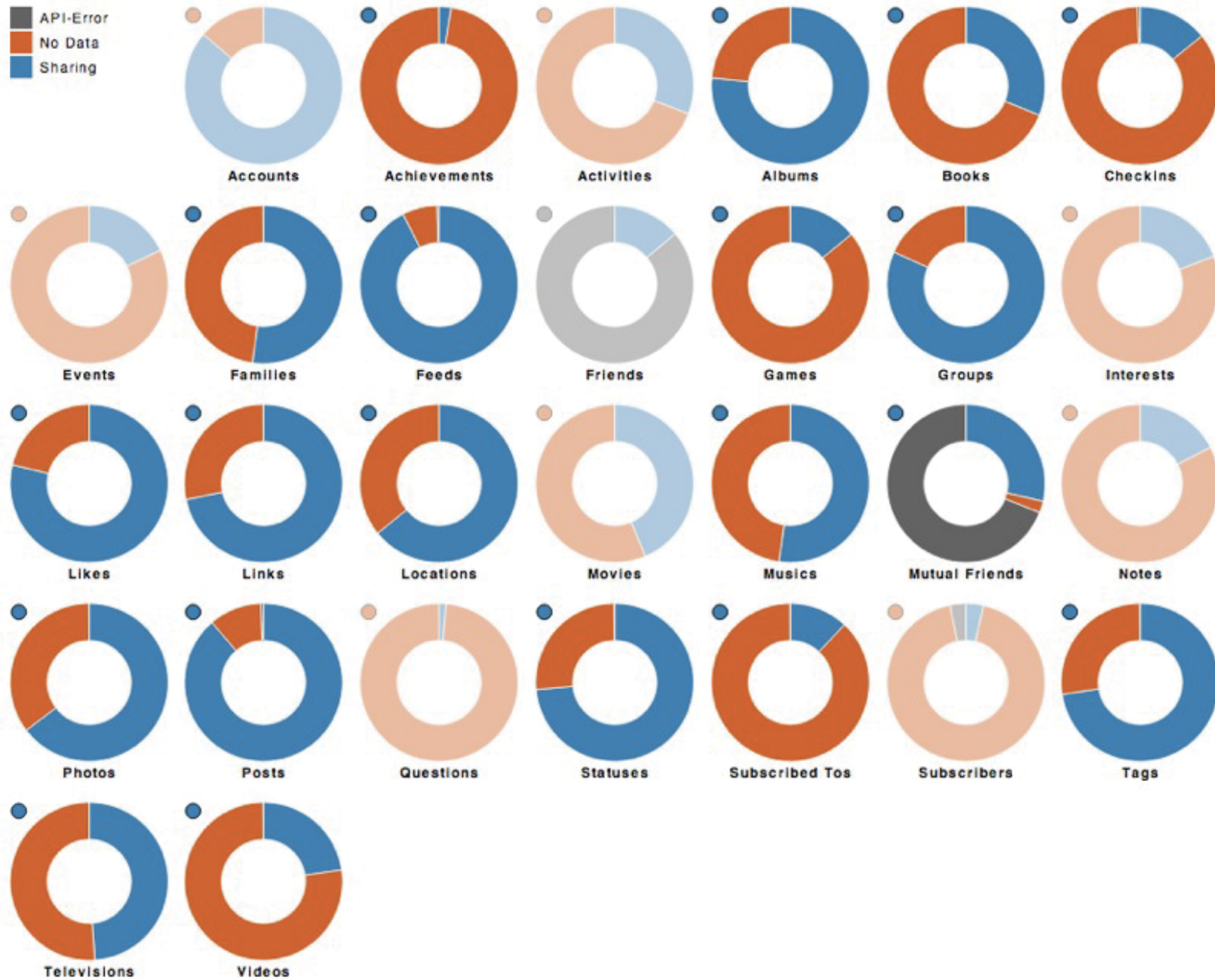
CROWDSOURCED SETTINGS

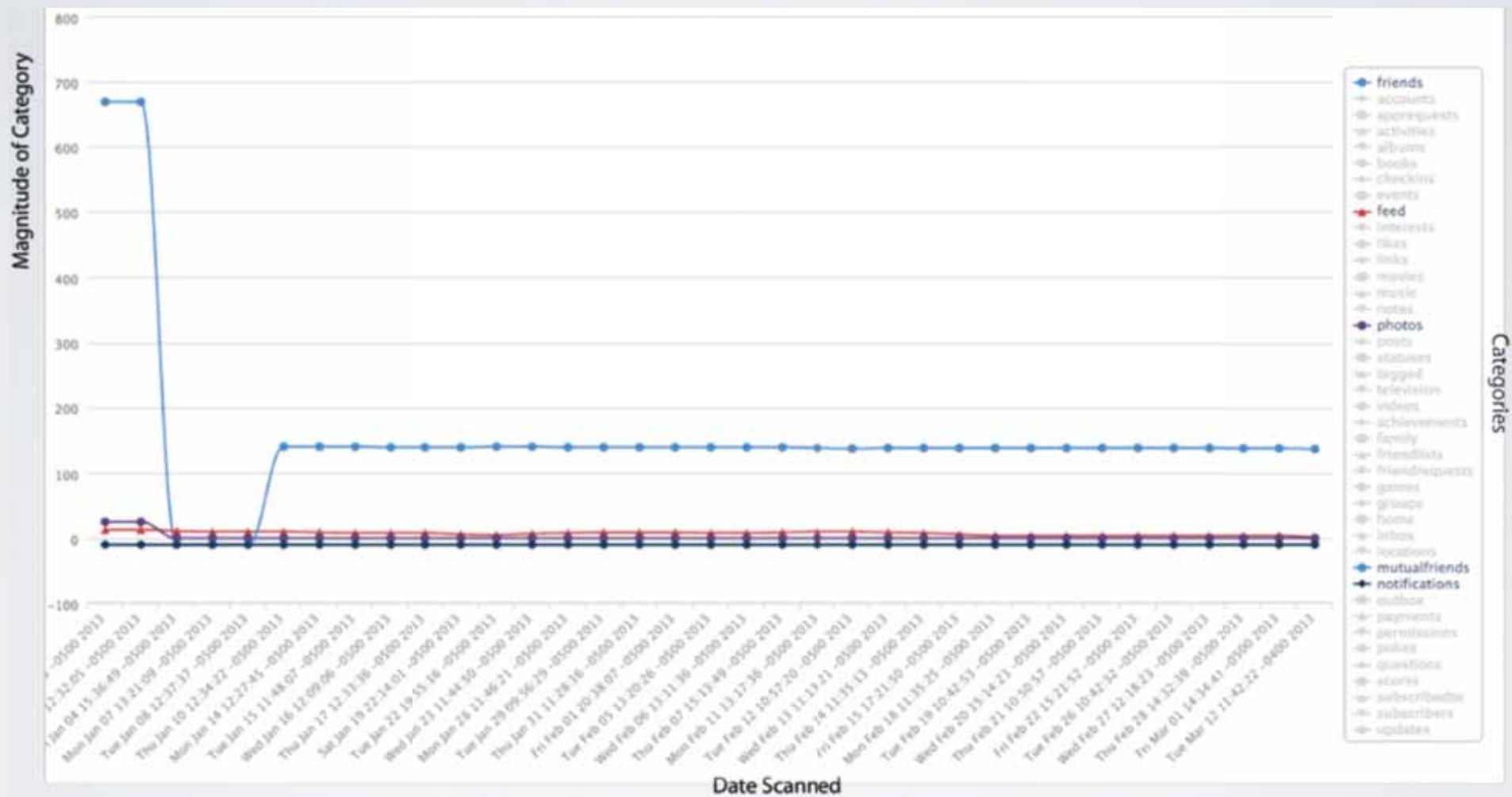
- Compare your settings to “others”
- Mimic “others”
- “Others” - trusted friends, certain networks, all friends, etc.

Friends Data



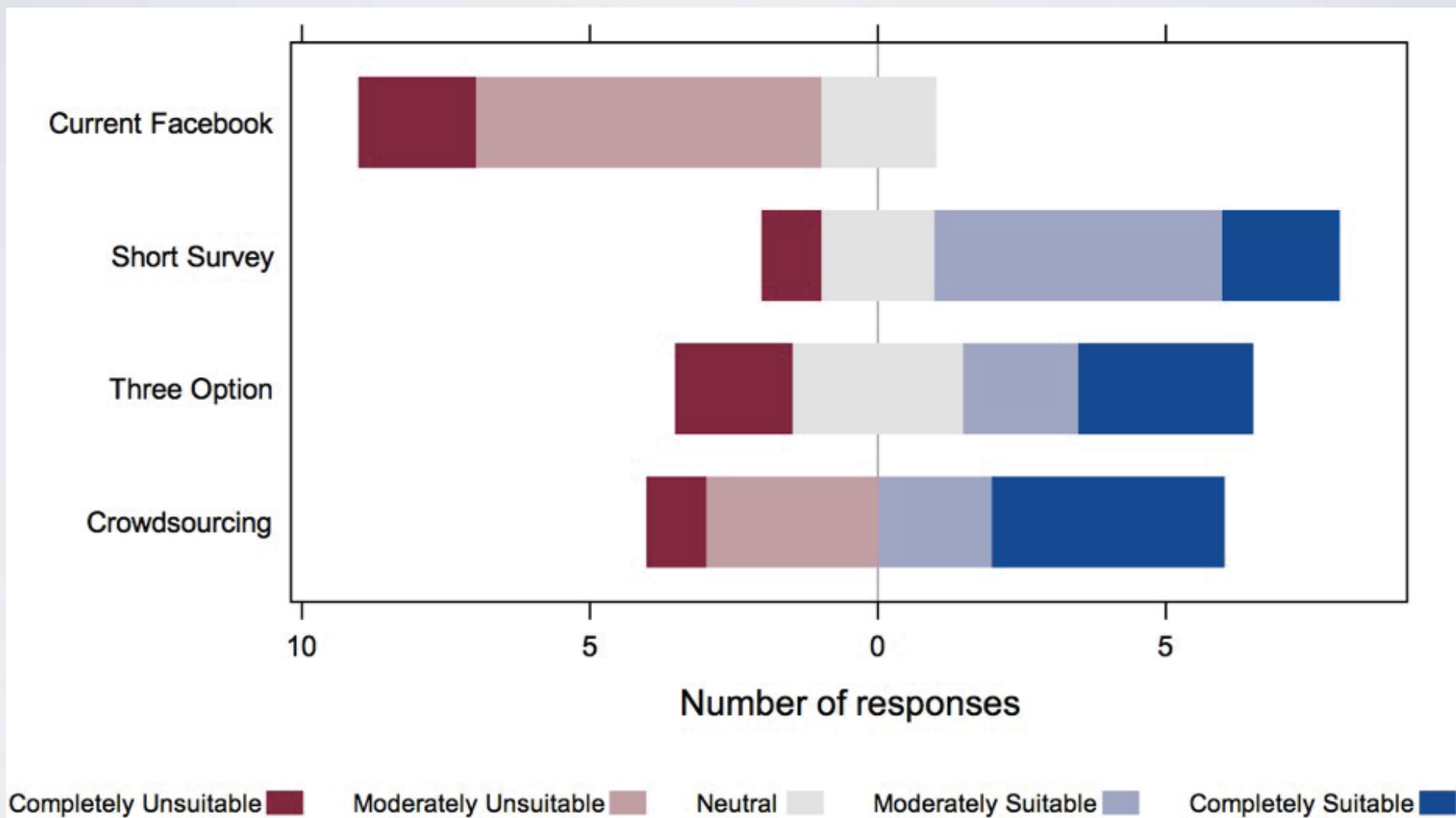
Per Category Sharing





A word cloud featuring various terms related to social media and digital interactions. The words are arranged in a dense, overlapping manner, with some words being significantly larger than others. The colors of the words are diverse, including shades of red, orange, green, blue, purple, and brown. The words are oriented horizontally, though some are slightly tilted. The background is a plain, light gray.

achievement
inbox
television
album
status
group
interest
home
notification
event
like
note
video
activity
payment
question
movie
checkin
outbox
feed
friendlist
permission
account
poke
music
post
game
score
book
tagged
link
family
photo
mutualfriend
friend
subscribedto
subscriber
update
friendrequest



SOCIAL SOFTWARE ENGINEERING

Swapneel Sheth
@swapneel

swapneel@cs.columbia.edu