

# History by Gmail

Sanby Lee | CS448b

# Motivation

- Search for video → finding completely random old threads with friends
- Different time in life → topics, writing style, frequency
- How did things change so quickly? What else has changed?
- Communication patterns over email create a record of who we talked to and what we did. *May not match with our own perception.*

**Create a tool that provides insight into our social interactions through email & maps the changes over time**

# Questions to Answer

What words do you use?

common words

↳ word cloud

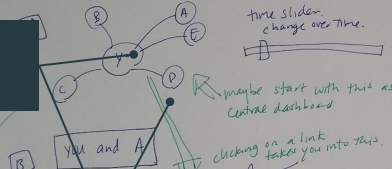
↳ normalize against common words across all

★ all of these stats crunched on the backend offline

only the final data at user rolled up level is loaded into JSON

⑤ change over time

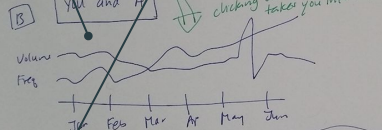
how do I show this??



time slider change over time.

maybe start with this as central dashboard

clicking on a link takes you into this.



⑥ clusters = user selects which attributes matter  
get Vectors.  
CA are automatically shown on dashboard.  
user - user can click and drag circle to user - user can click and drag circle to capture a group of people - non analyze them all at once. → avg stats across that group.

Who do you talk to?

① Who do you talk to?

a basic ranking

A	35 times total	→ define Time =
B	20 times total	
C	10 times total	

can we find correlations with metadata?

② How frequently?

= FREQ

- take # <sup>days</sup> ~~times~~ between each <sup>thread</sup> communication
- discard top + bottom quartile.
- average the spans

A	every 3 days
B	every 5 days
C	every 6 days
D	every 30 days

③ How much?

= VOL

- avg word count
- avg length of thread

A	300 word responses	25 msgs per thread
B	250 word responses	10 msgs per thread
C		
D		

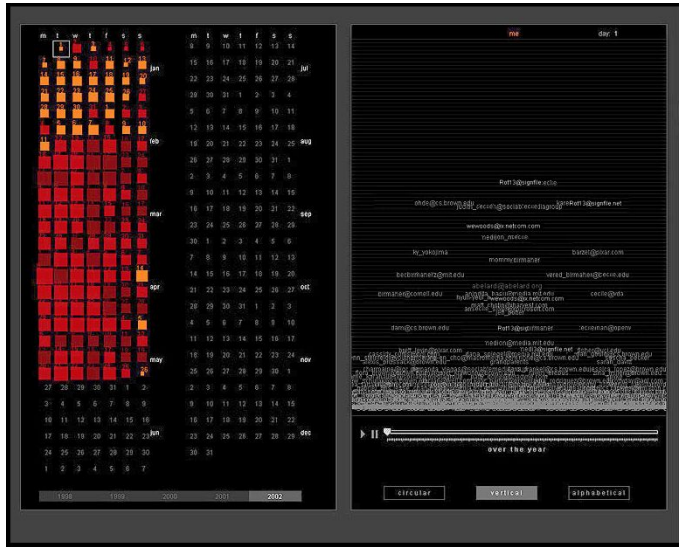
Change over time?

Are there clusters?  
By communication pattern

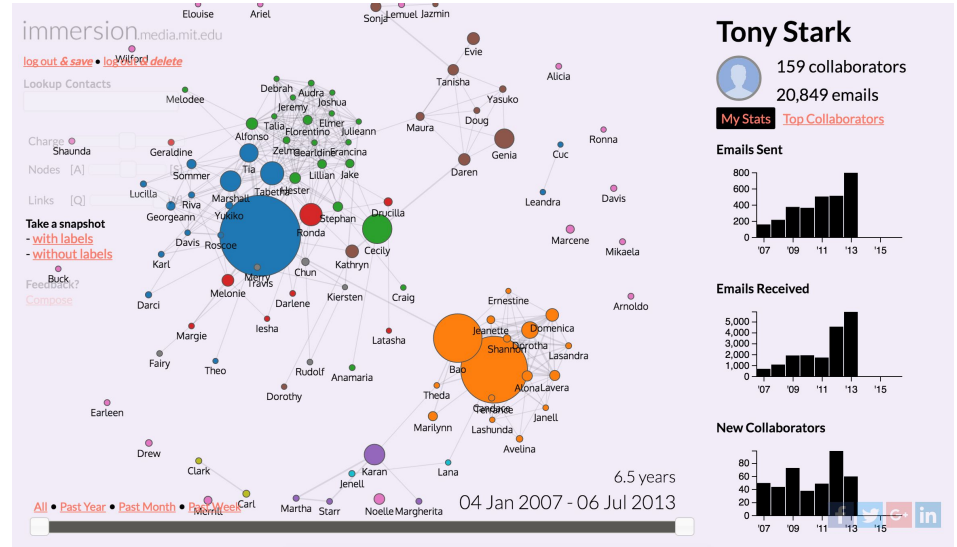
How frequently?  
# of msgs per time period

How much?  
Size or length of msgs

# Prior Work



“Digital Artifacts for Remembering and Storytelling: *PostHistory* and *Social Network Fragments*” (Viegas et. al 2004)

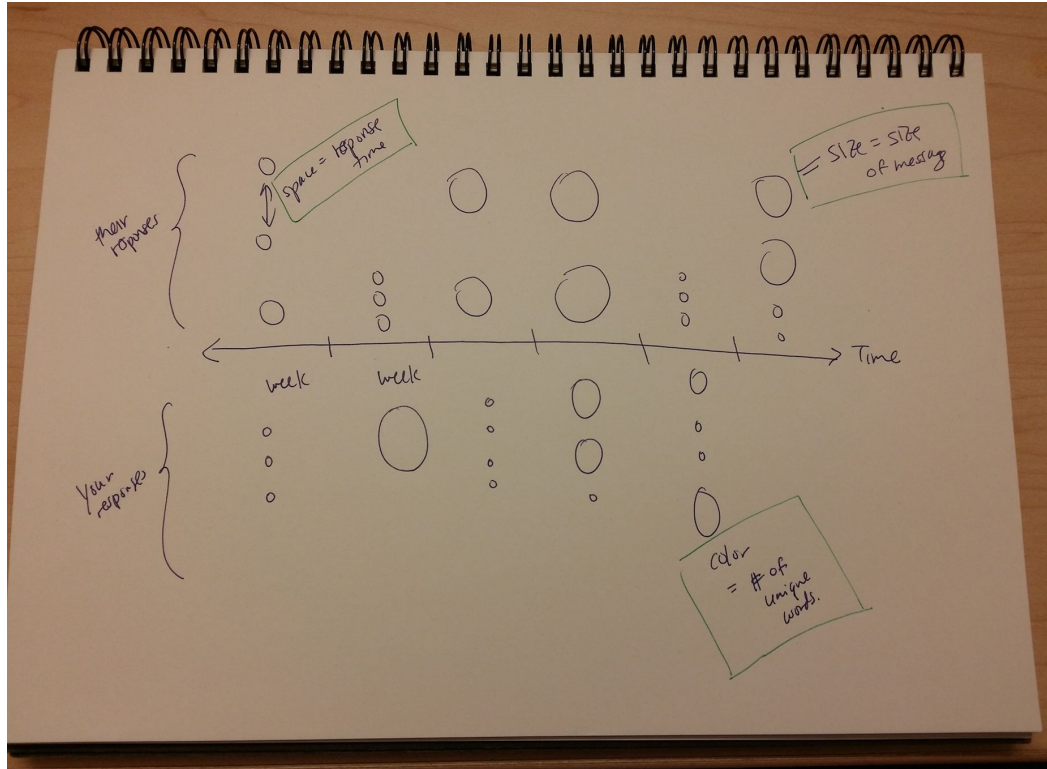


Immersion Project @ MIT Media Lab (Smilkov, Jagdish & Hidalgo)

# New Direction

- Move away from “dashboard” concept → targeted visualization
- Prior work uses volume of email to establish “rhythms” of communication
  - Can we incorporate other attributes to determine rhythm?
    - Speed of response time
    - Length of messages
    - # of unique words in communication
  - Find different patterns of communication
    - Quick back-and-forth exchanges
    - Lengthy and deep discussions
- Sentiment analysis of your relationship with a person

# Draft Interface





# Progress Report

- Downloaded 2 years of email data in structured format from Gmail API
  - Data is at the message level
  - For each message: to, from, date, message body + lots of other data
  - Preliminary analysis: Who do I communicate with most frequently?
- Next step - refocus on specific visualization
  - Pick one targeted question to answer
- After that
  - Parse data to aggregate at the user-user level → Format
  - Code final design in d3/js

# Questions for Audience

What bigger questions would you like to answer by looking at your email data?

- “Has my circle of friends grown bigger or smaller over time?”
- “Who have I gotten closer to? Who have I fallen out of touch with?”
- “Have my priorities changed over time?”
- “What are my ‘patterns’ of communication with different people?”