

small data

from mobile health to immersive recommendation

Deborah Estrin

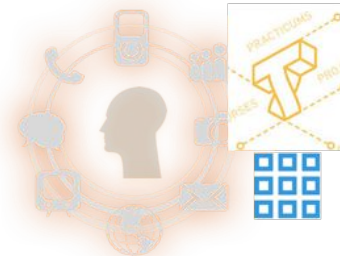
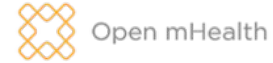
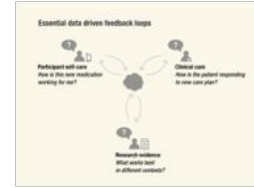
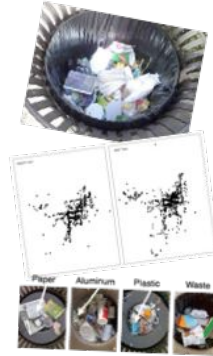
[Cornell Tech](#): Professor of Computer Science, Founder of Jacobs Health Tech Hub

[Weill Cornell Medical College](#): Professor of Healthcare Policy and Research

[Open mHealth](#): Co-Founder

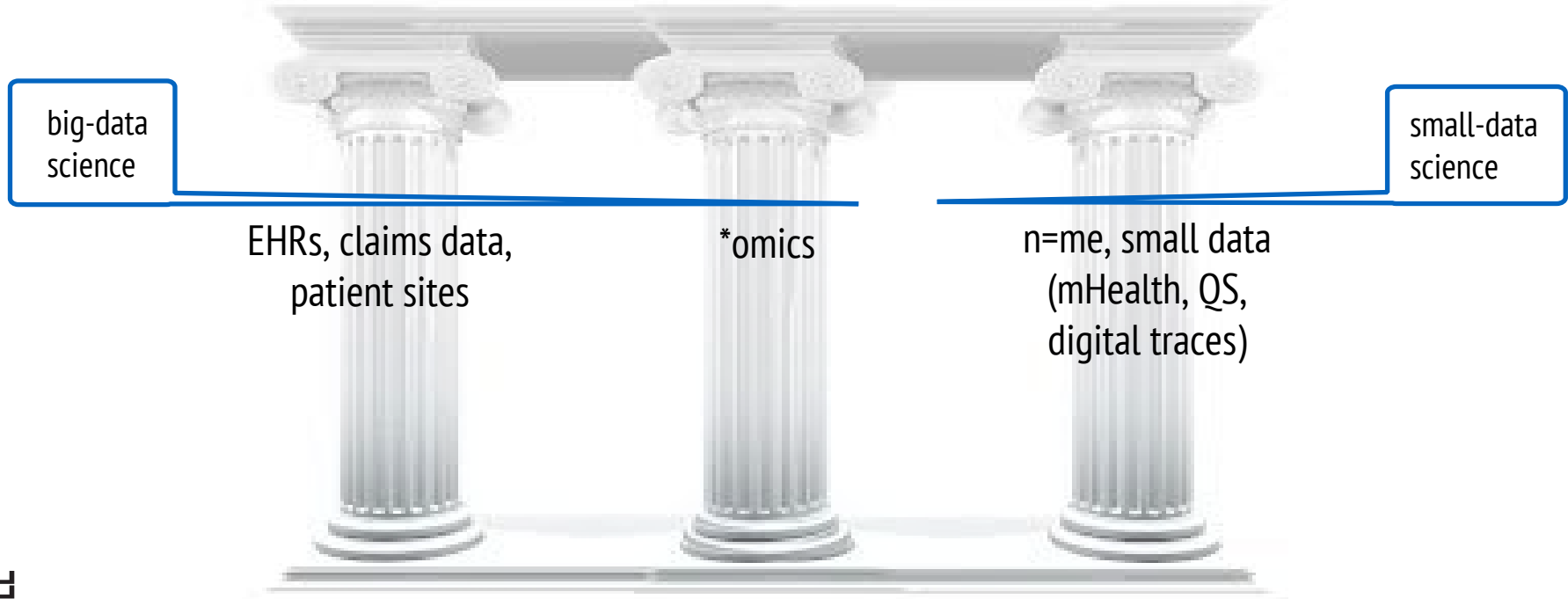
destrin@cs.cornell.edu, [@deborahestrin](#), <http://destrin.smalldata.io>

$n=me$



X

Data fuel Predictive, Preventive, Personalized, Patient-Centric Medicine ~~Health~~



harness previously-unmeasured function and behavior
to fuel personalized and evidence-producing care and behavior

The Reality

Chronic diseases -- 7 of 10 deaths and 86% healthcare costs (US,CDC)

Precision medicine, advanced genomics alone will **not** address the most preventable mortalities whose **primary determinants are behavioral**.

Research objective

Develop **small-data driven techniques for user modeling**, phenotyping, preference profiling, social-support, and behavioral incentives to help us feel and be 'better'

...to fill the gap between what we are born with and what we die from...



Participant self-care

How is this new medication working for me?



Clinical care

How is the patient responding to new care plan?

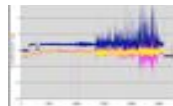


Research evidence

What works best in different contexts?

small data: diverse data sources for the individual, $n=1$

Passively-recorded activity, location



“Real” Sensors, wearables



Novel, Visual, self-report

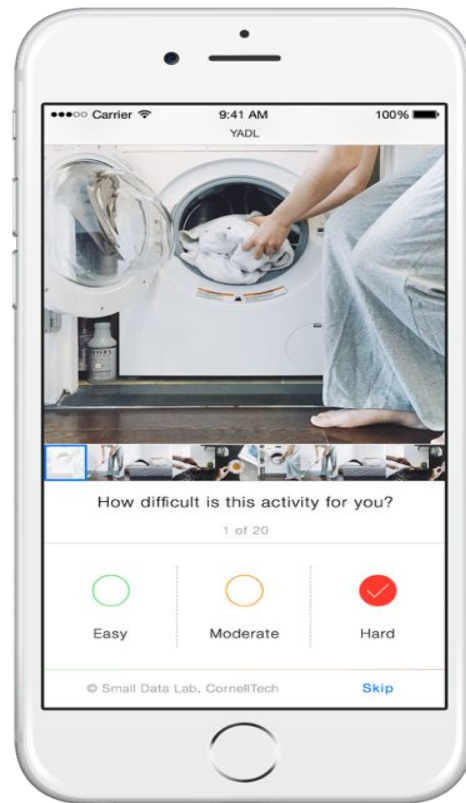
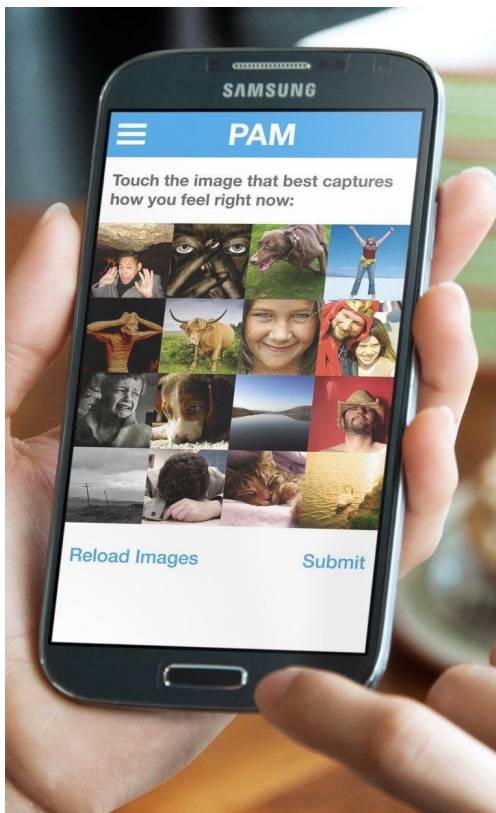


Digital traces: purchases, media



Smart(er) self report:

Photographic Affect Meter (PAM), Your Activities of Daily Living (YADL)



Pollak, et al

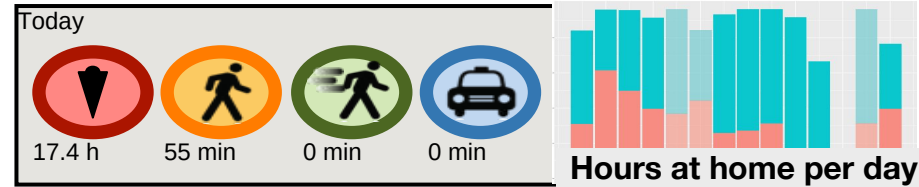
small data lab

Challenge: moving up the information food chain

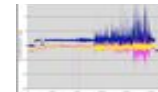
**patient function
(behavioral biomarkers)**



summarization, fusion



raw measurements



Industry leaders: Ginger.io Check Engine Light

Continuous & Passive
Ginger.io fills in data gaps that are often missed when using self-report measures. This reduces patient burden and results in improved data quality.

Data Collection



Behavior Patterns



Check Engine Light
Identify at-risk patients based on objective patient data, and alert a provider or caregiver.

Health Status



Assess intervention
Track outcomes more explicitly, and determine efficacy of outreach approaches

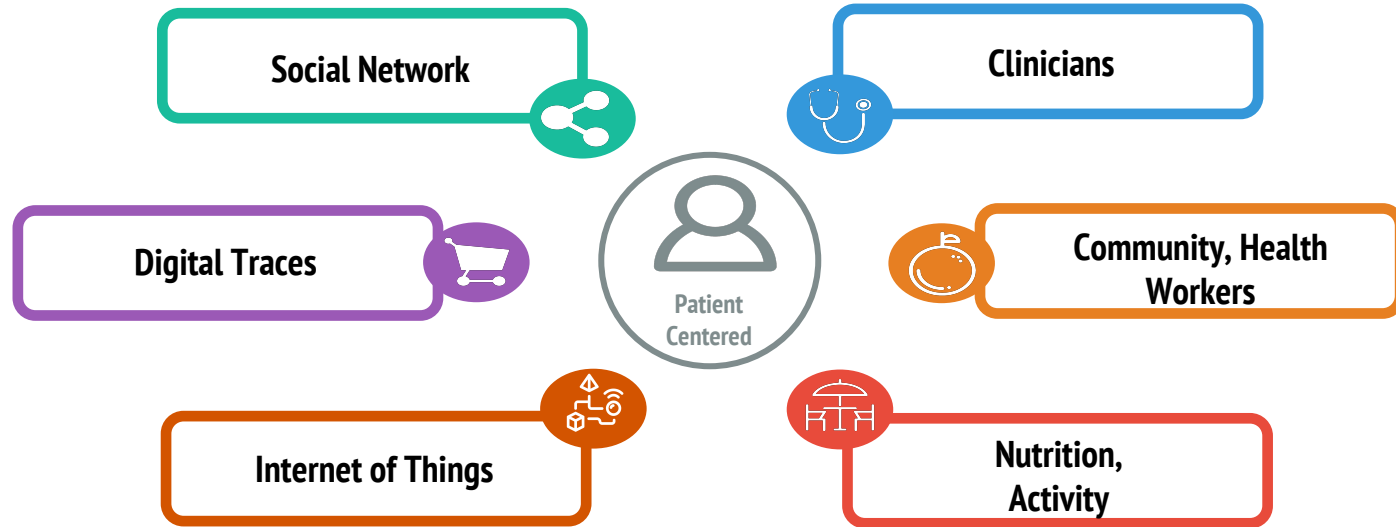


Protecting Privacy
Ginger.io is HIPPA compliant, and collects statistics, not specific contacts, locations, or content.

<http://ginger.io>

Confidential & proprietary. All content © Ginger.io, Inc.

Actionable by whom?



Solving patient's problems
using data-driven social facilitation

MoodRhythm(TM) provides tools for Bipolar patients to measure and manage their own daily routines



Tanzeem Choudhury, HealthRhythms

small-data fueled applications

nutrition
(e.g., household shopping)



pain management
(e.g., lower back exercises)

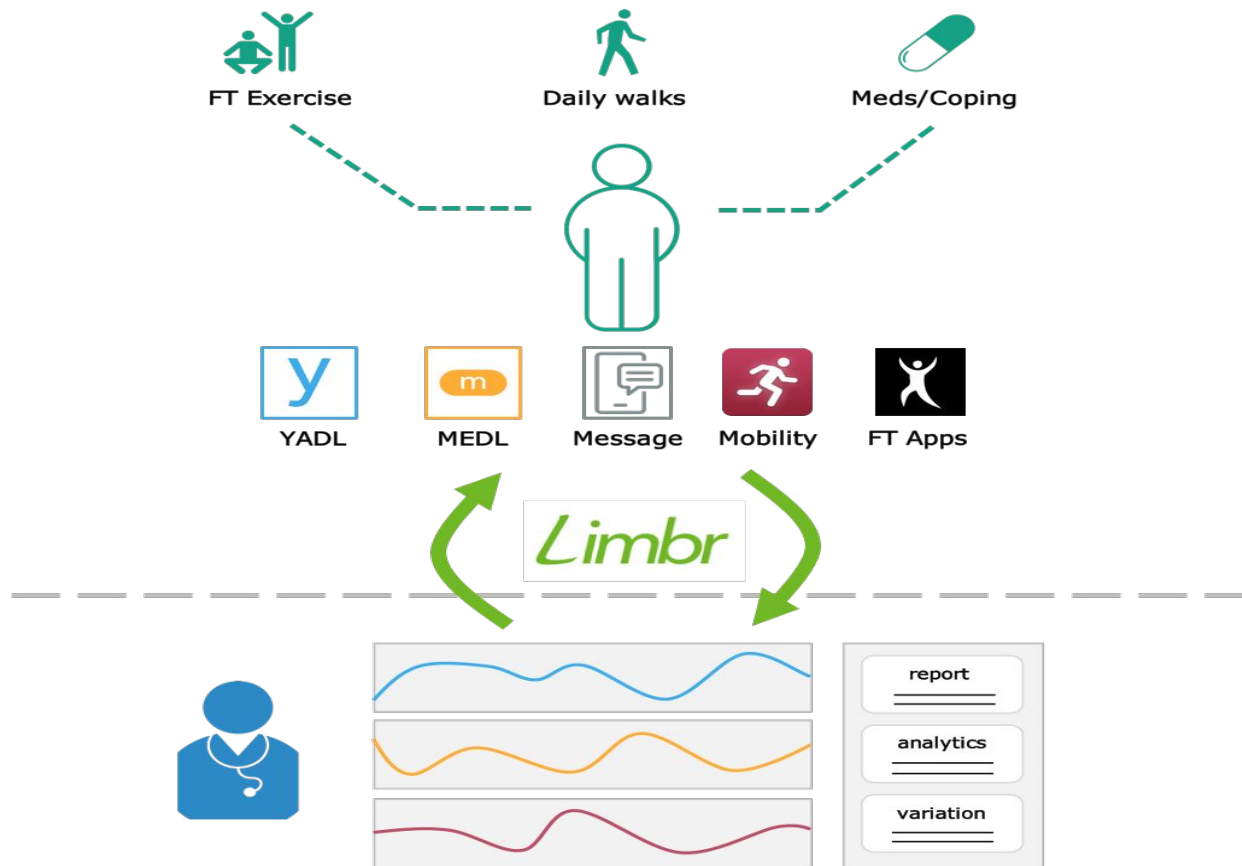


student health
(e.g., time management)



measure, motivate, modify, maintain aspirational health behaviors

Limbr: an intervention for chronic pain



YADL Monthly Assessment Design Process

Key items identification from standard forms

Key items identification from online resources

Key items aggregation

Image matching and iterative refinement



Walking

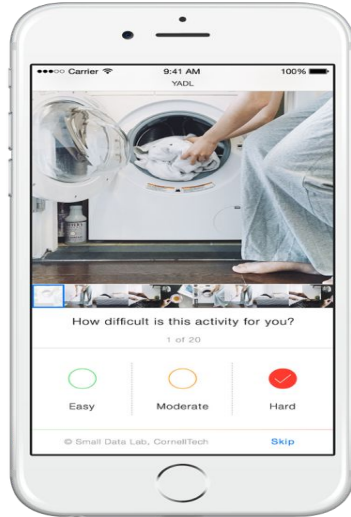


Climbing stairs

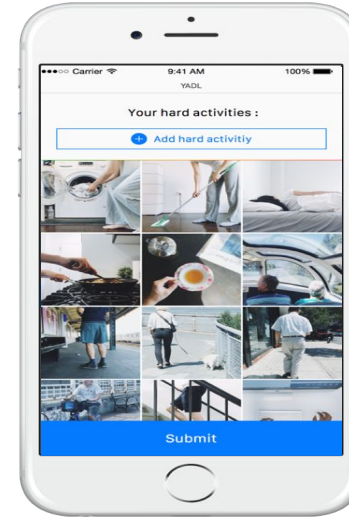
Activity Phrases List:

- Bathing
- Feeding
- Dressing independently
- Walking
- Climbing stairs
- Rising from chair
- Yard work

Your Activities of Daily Living (YADL)



Monthly Identification of
Painful/Difficult/Interfered-with ADLs
<http://bit.ly/1KoqUIP>



Daily Assessment of ADLs that were
Painful/Difficult/Interfered-with that day

Why isn't there more evidence???

It takes time to generate evidence; but not this much time!!

have the marketplace and current academic processes underperformed?

Lets learn from all (other) things digital

rapid iteration (continuous improvement using analytics)

modularity (generalizable learning)

There is hope

[Sage Bionetworks Bridge and Synapse projects](#): collaborative research studies and data commons

[Open mHealth](#): standardized, clinical representation and open source components

Emerging “evidence engines”: Mobile Research Study Frameworks



ResearchKit™



ResearchStack



Open mHealth



touchlab

Bits

DIGITAL HEALTH

Bringing iPhone-style Medical Research to the Android World

By STEVE LOHR

NOVEMBER 12, 2015 11:00 AM

6 Comments

Email

Share

Tweet

Save



<http://researchstack.org>
destrin@cornell.edu,
michael.carroll@cornell.edu

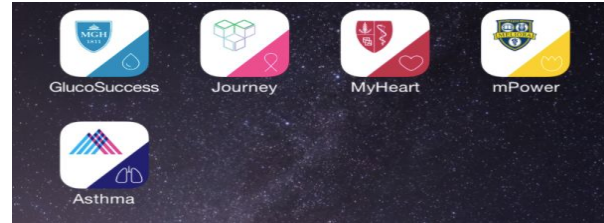
JACOBS
INSTITUTE

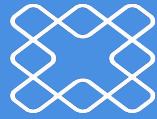
AT
CORNELL
TECH



Mobile Consent

- Sage IRB approved eConsent for PD and BCS
- Standardization across all 5 Research Kit apps (approved at Mt Sinai, Stanford, MGH)
- Released as open source toolkit for adaptation
- Established principle that participant can dictate degree of sharing
 - ~75% of participants overall share broadly





Open mHealth

Open source framework that helps developers integrate health app & device data using a common language



@openmhealth

<http://openmhealth.org>



<http://crowdsignals.io>

HOME

ABOUT

COMMUNITY

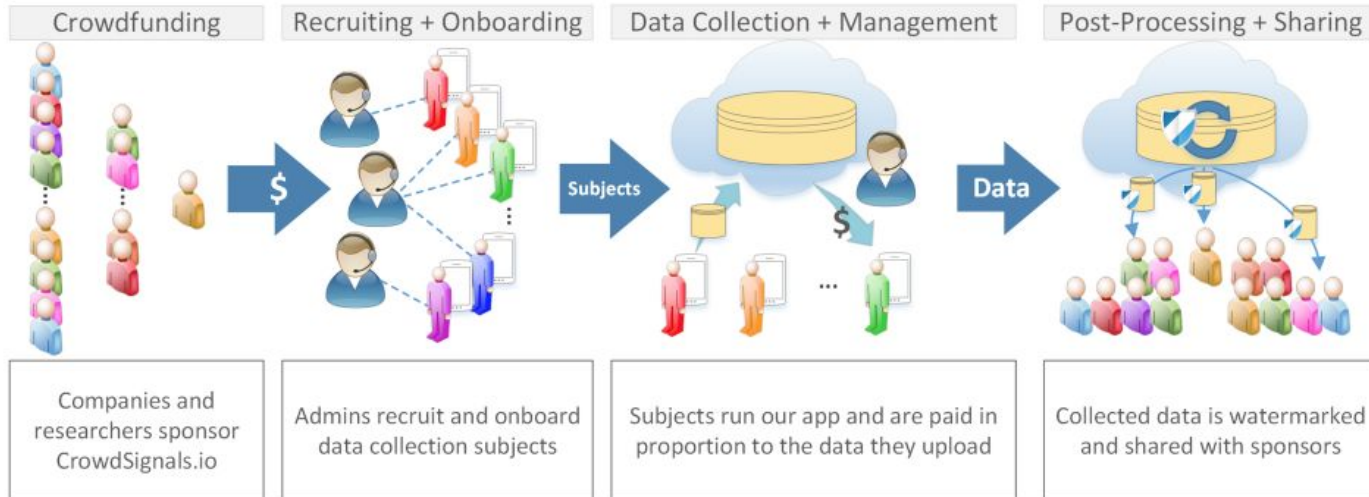
DATA

ORGANIZERS

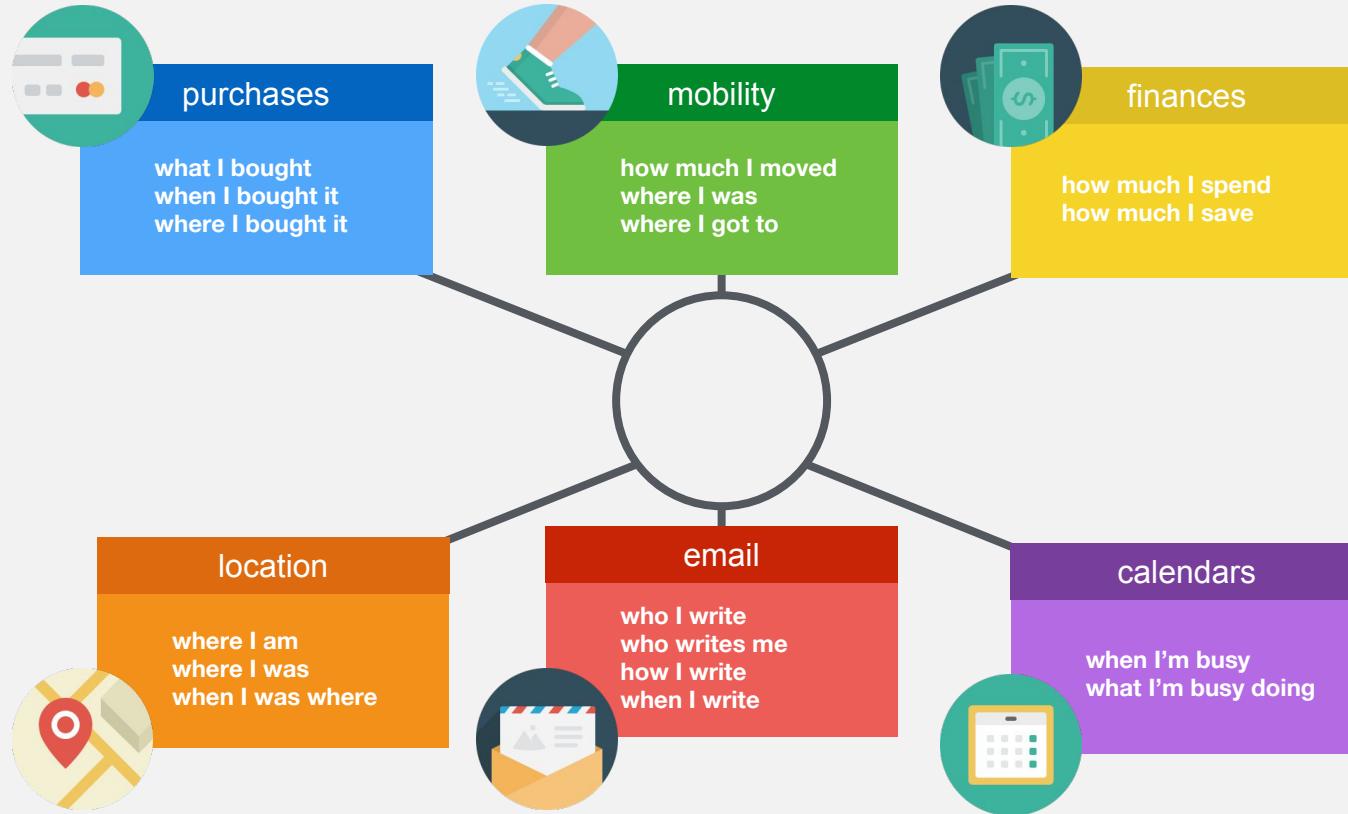
FAQ

CONTACT

HOW IT WORKS

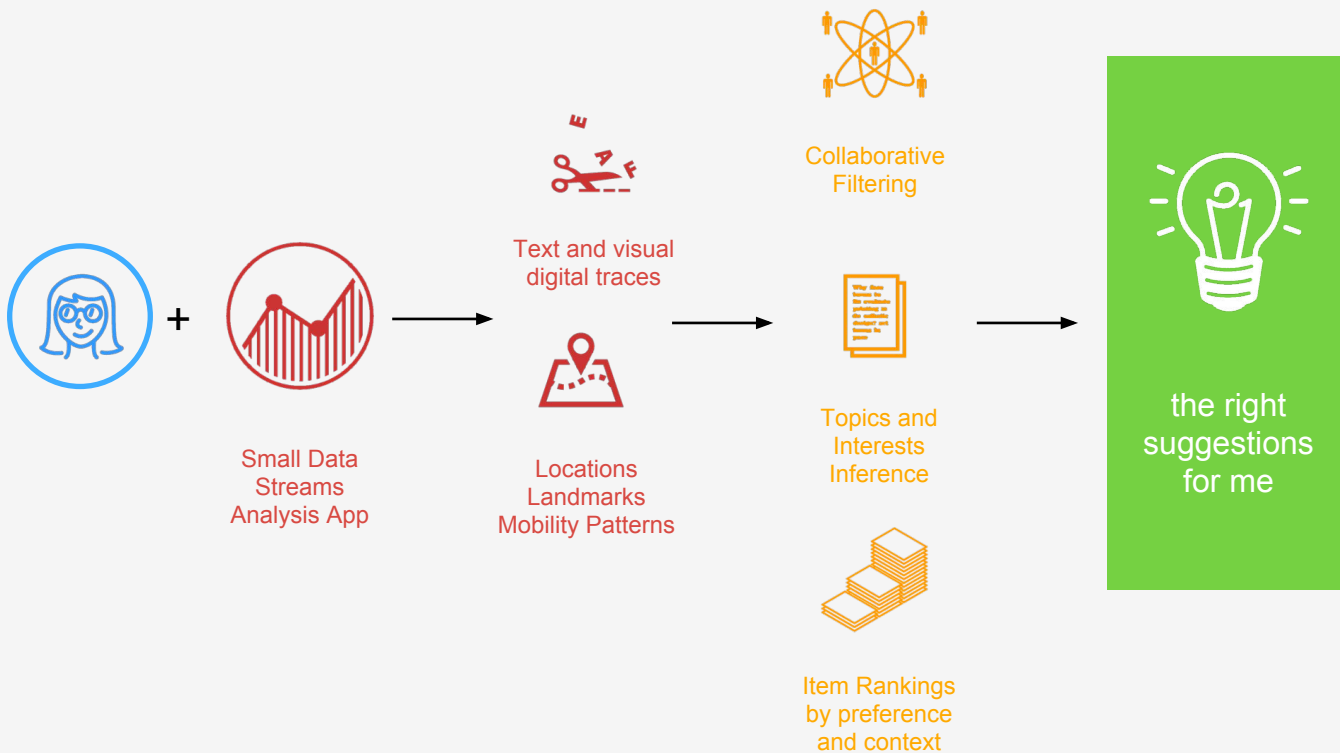


beyond mobile and beyond health

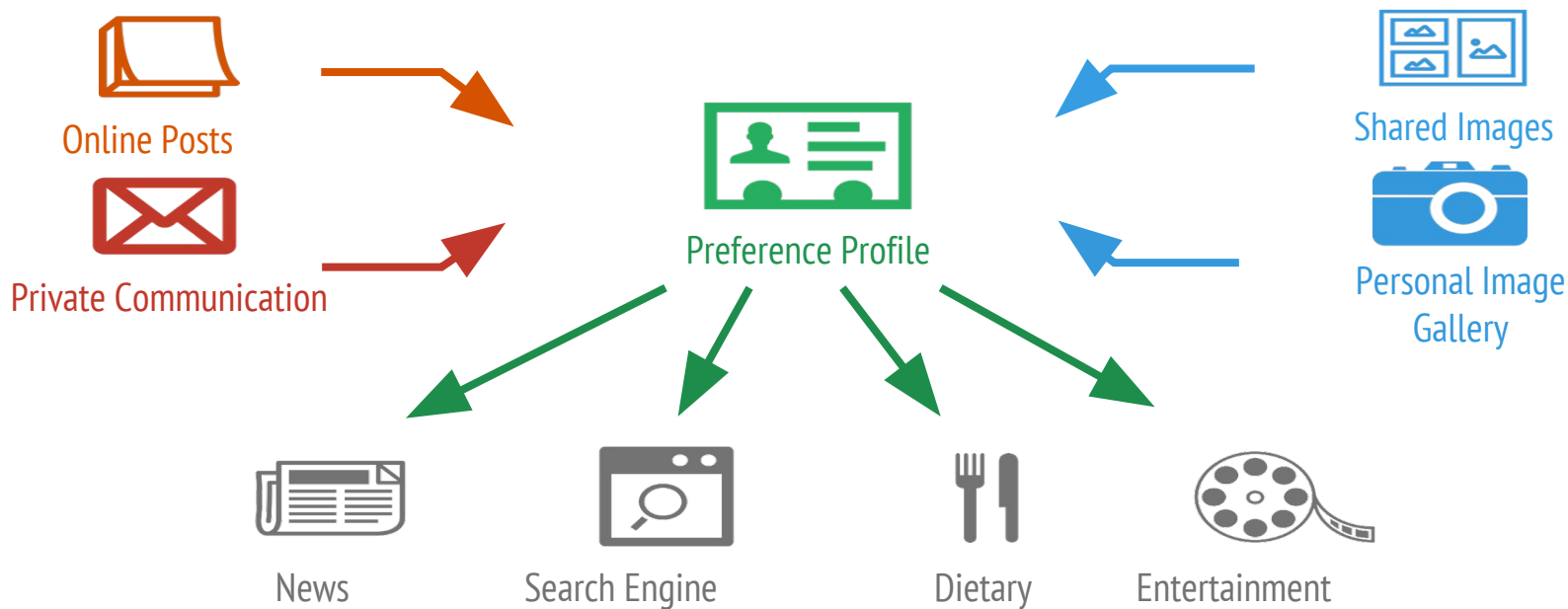


Fueling personalized recommendation

model preferences w/ location, language, activity, photos; selective sharing



User-centric Preference-learning using small data

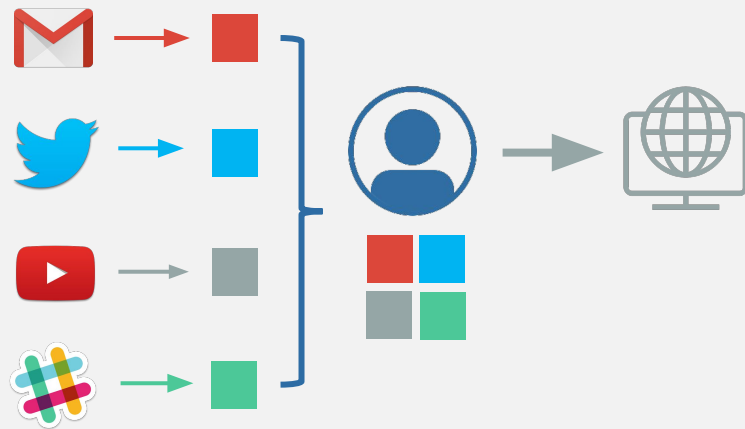


Tried it out in News and Meetup Recommendation

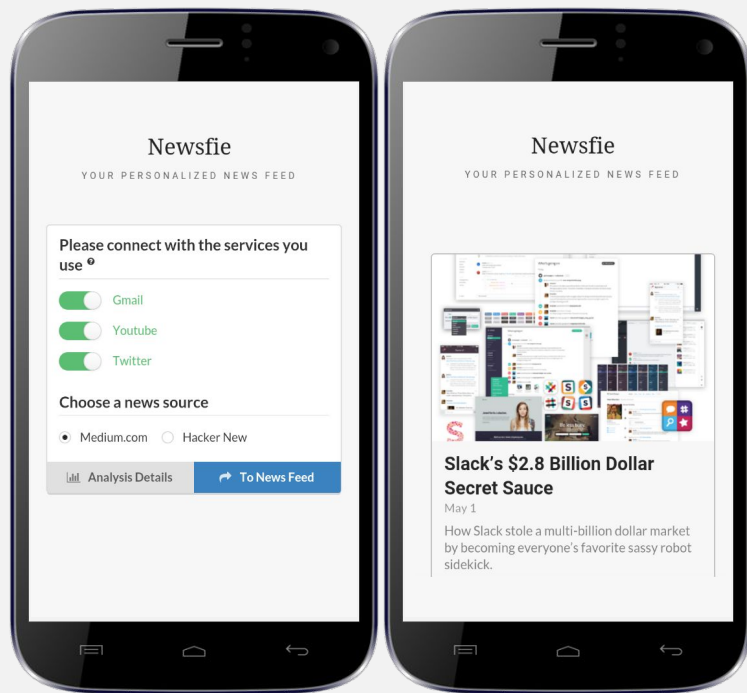
Cold start problem



Immersive rec.



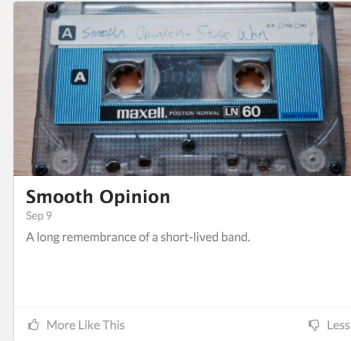
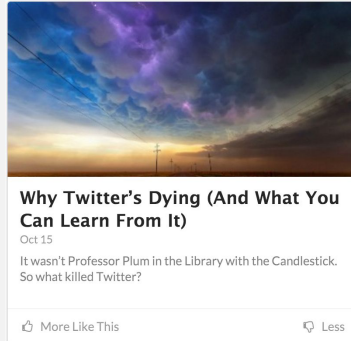
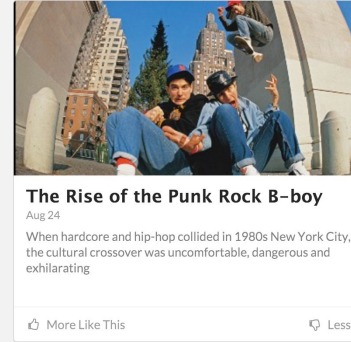
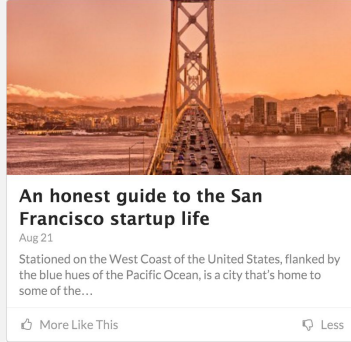
News recommendation using Medium, Twitter



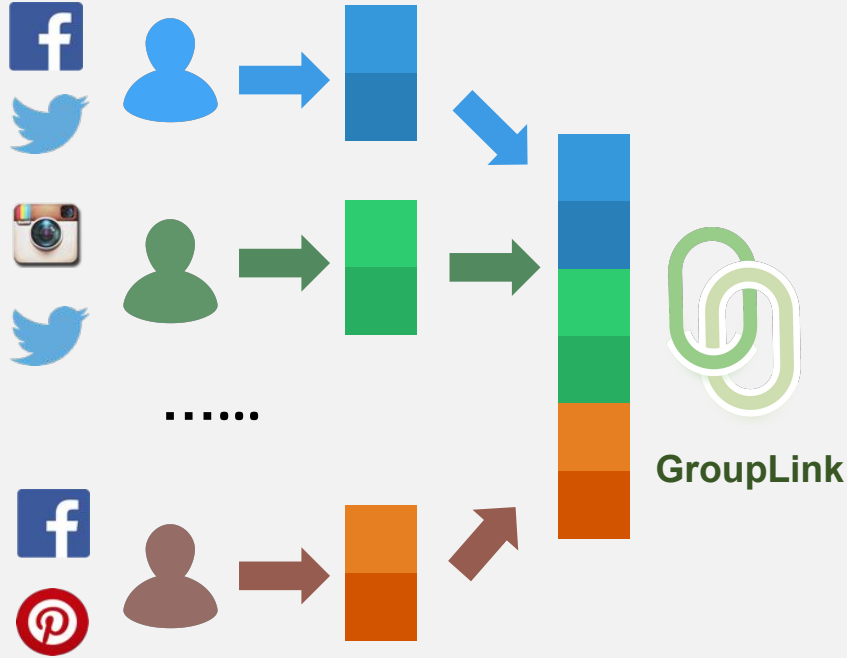
Publicly available version running to support better user studies in future

Modular—available to try out improved algorithms, new application contexts

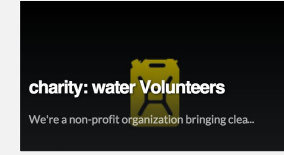
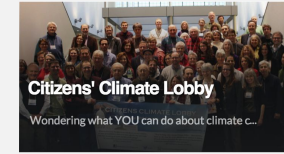
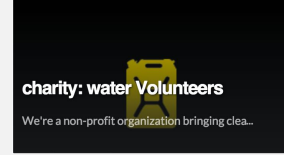
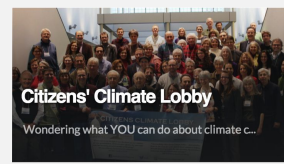
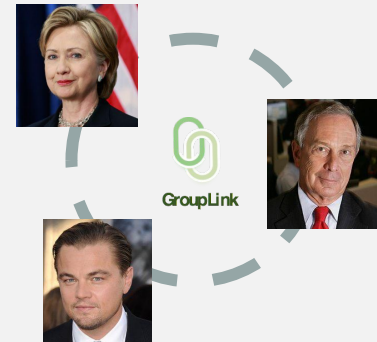
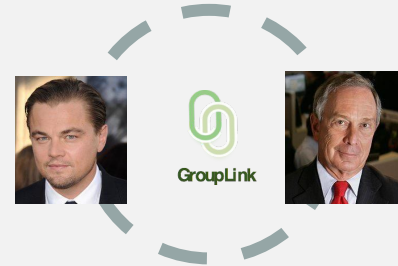




Meetup Recommendation



<https://lifestreams.smalldata.io/hn/grouplink/>



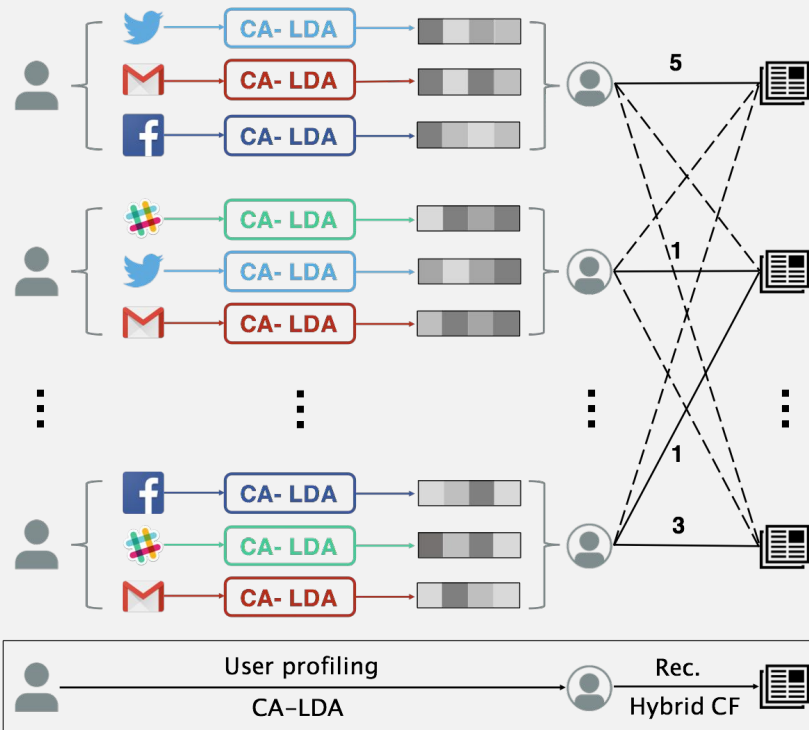
Overview

- **Phase 1 User Profiling**

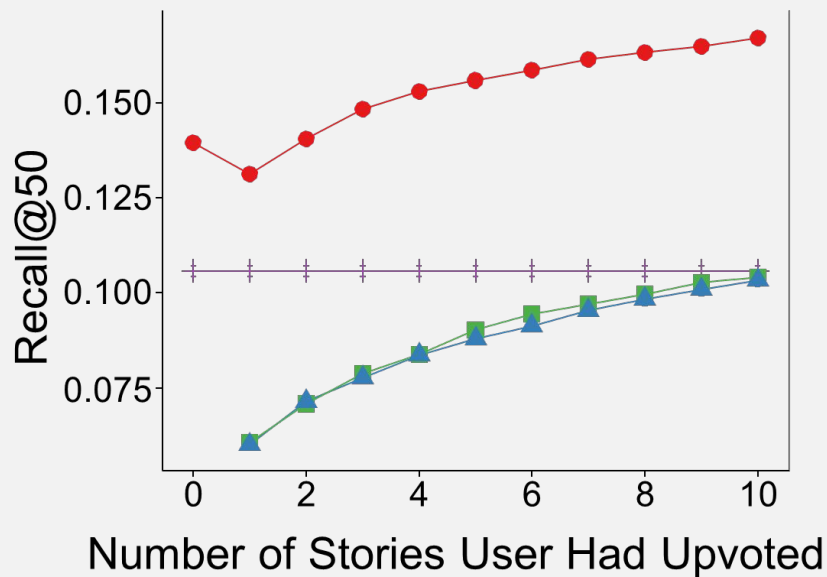
Create user profile from users' digital traces

- **Phase 2 Recommendation**

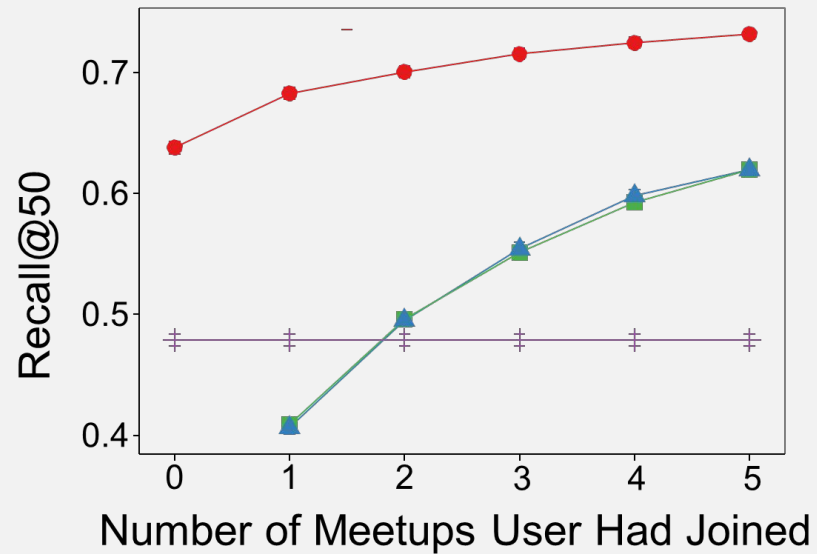
Hybrid collaborative filtering algorithm combines user/item profiles and ratings



● ImmRec
▲ CTM
■ PMF
✦ POPULAR

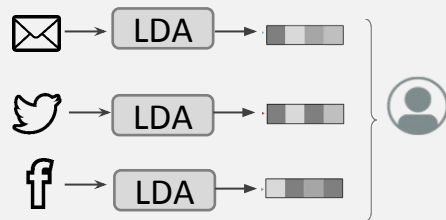


● ImmRec
▲ CTM
■ PMF
✦ POPULAR

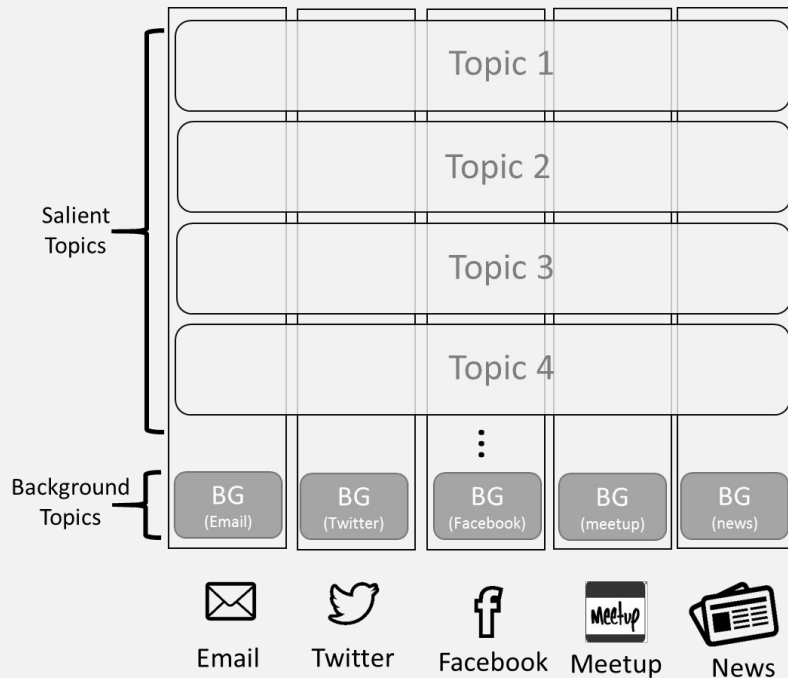


Topic modeling: Context Aware LDA

LDA topic model trained with Item Corpus did not work well



- Co-train multiple contexts for generalizability
- Two kinds of topics
 - **Salient topics** shared across contexts
 - **Background topics** specific to individual contexts



Food Purchase recommendation: Pushcart



1. Member Household
Signs Up



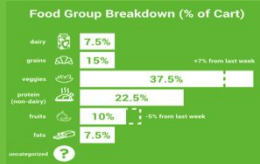
2. Member Sets
A Goal



3. Online Grocery
Purchases
Auto-Collected
Via Email



4. Groceries
Auto-Analyzed



5. Member Receives
Swap Suggestions
From Coach

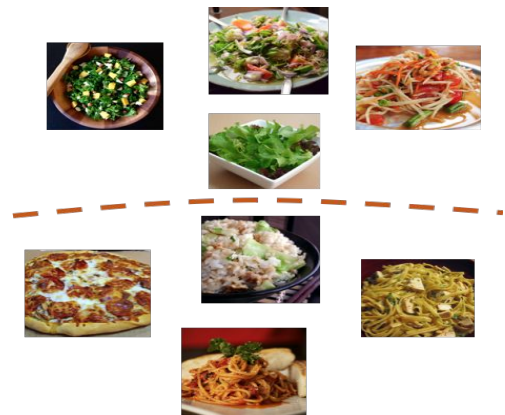
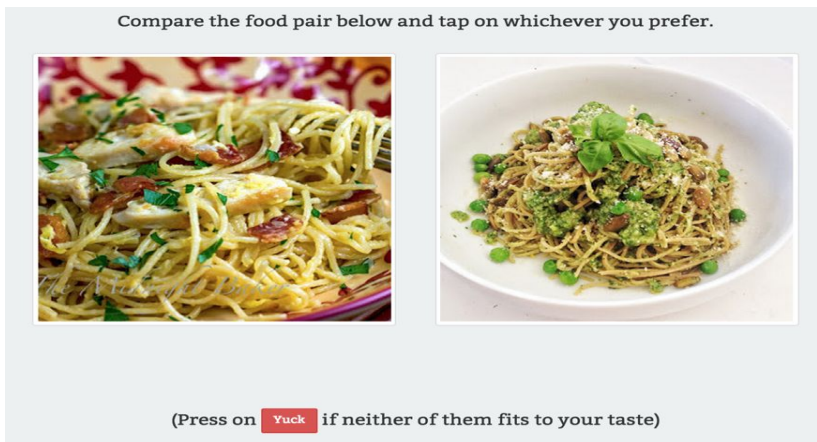


6. Member Receives
Optimized
Shopping Lists



<http://gopushcart.com/>

PlateClick: Bootstrapping your food preferences



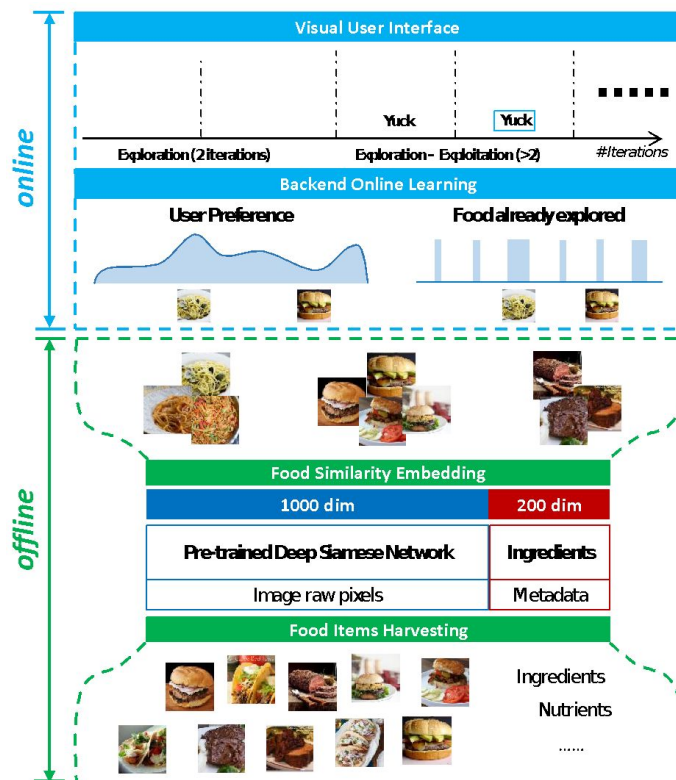
Visual pair-wise comparisons
image and metadata analysis, online
learning

- low cognitive load, personalized and legible.
- Preference Elicitation, no history required, no ratings
- Completed within a minute.

Personal diet-preference profile engine for food
recommendation and health applications

bit.ly/plateclick

PlateClick: System



Online Learning

Online Learning framework

- What images to present to the user?
- How to update users' preferences?

Food Similarity Embedding

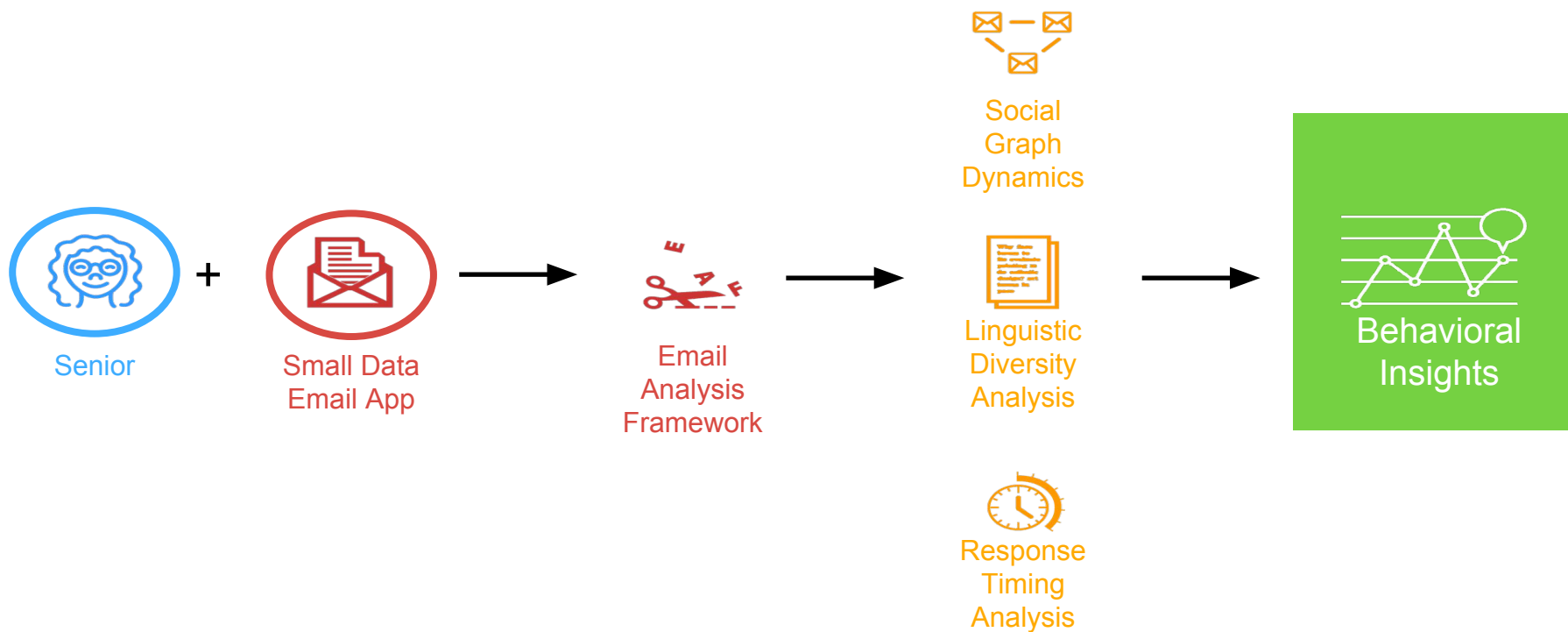
Users have **close** preferences for **similar** items

- Feature representation that reflects similarities

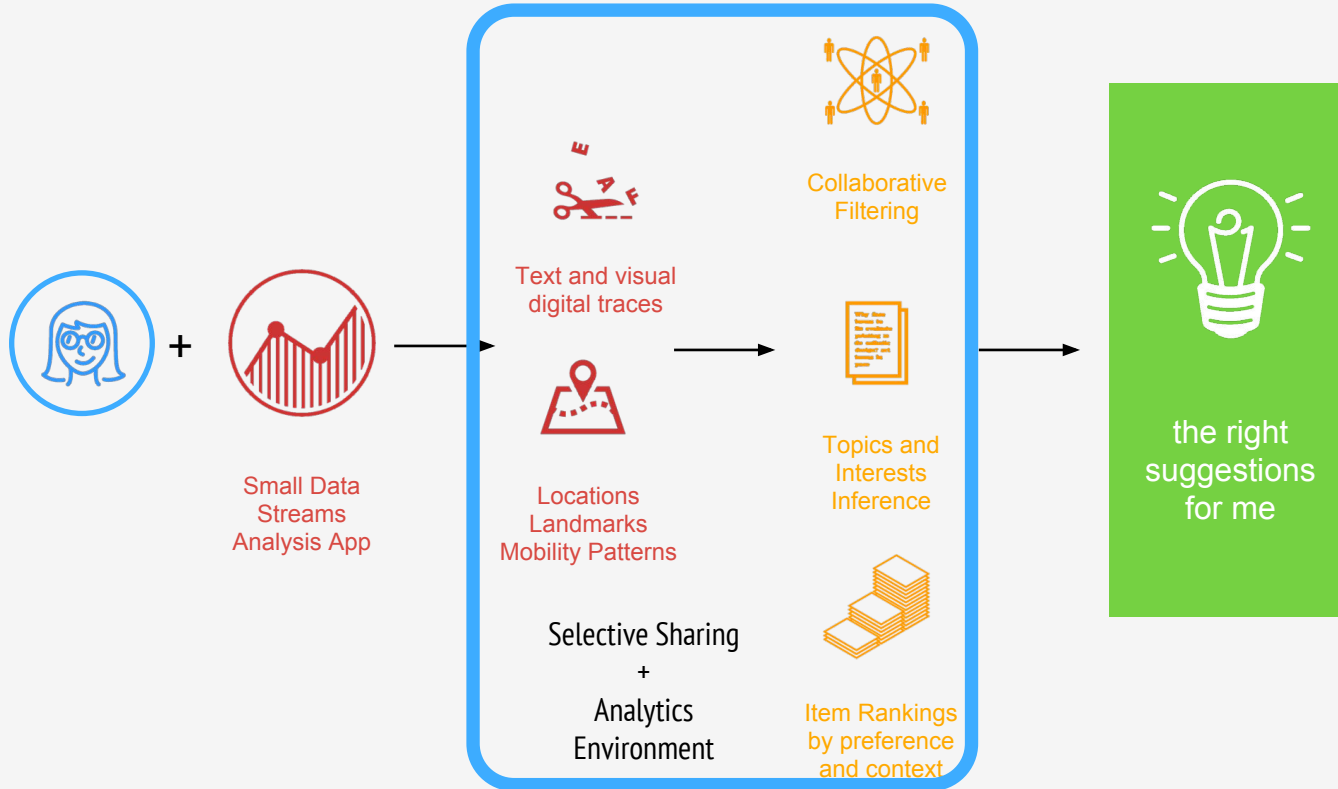
Food Items Harvesting

- Food images and metadata.

Future directions: modeling user-state based on personal language patterns: sentiment, cognition



Future directions: Creating selective-sharing analytics environment



Creating an ecosystem around small data



**1. modular APIs
shims for
accessing data
from popular
apps and services**



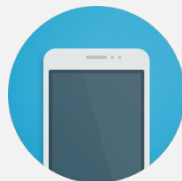
**2. simple storage
patterns for
ephemeral raw
data processing**



**3. reusable data
processing and
analytics modules
on raw data
primitives**



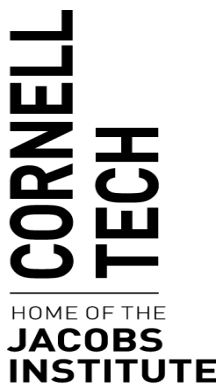
**4. plug-and-play
visualizations and
UIs on different
outputs of linked
small data analysis**



**5. specific applications
tailored to key use cases
and patient care models**



**6. consumer, developer and
research communities
built around small data
building blocks and
applications**



For more information

<http://smalldata.tech.cornell.edu>

<http://destrin.smalldata.io>

<http://tech.cornell.edu/programs/masters-programs/ms-in-is-health-tech>

@deborahestrin, @openmhealth

Faculty: Deborah Estrin, JP Pollak,

Clinical Advisors in Residence: Daniel Stein, Thomas Tsang

Technologists-in-Residence: Aaron Baum, Michael Carroll, Diana Freed

PhD Students: Faisal Alquadoomi, Andy Hsieh, Fabian Okeke, Longqi Yang

Funding: NSF, NIH, AOL, Google, UHG, Pfizer, RWJF, Intel, Cornell Tech

Co-authors and Collaborators at small data lab, Cornell Tech

PhD students



Faisal Alquadoomi



Andy Hsieh



Fabian Okeke



Longqi Yang

Researchers and Technologists in Residence



JP Pollak,
Senior Researcher-in-Residence

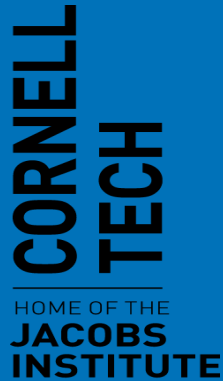


Michael Carroll



Diana Freed

Faculty Collaborators: Yoav Artzi, Serge Belongie, Tanzeem Choudhury, Ari Juels, Mor Naaman, Arnaud Sahuguet



2 Year MS in Health Tech @ Jacobs Institute

Admissions open now for AY16-17

- Focus on tech innovation for individual and population health
- Technologies and applications for health status monitoring and prediction, chronic disease management, consumer health behavior
- Commercial engagement in course projects and student internships
- Developed in collaboration with leaders from relevant healthcare sectors: Business, Clinical, Research, Entrepreneurial, and with faculty.

Join us at Cornell Tech to create Patient-Centered Health Tech

- **Internet of Things** that sources signals from heart beats to traffic to serve the Internet of People
- **Personalization** that promotes healthy eating and other preventive and protective activities
- **Social media engagement** and exposure used to understand and improve patient state of mind
- **High Touch care giving** made smart and scalable by high tech that informs human caregivers
- **Security** that balances individual privacy and sharing for science

