Answering Ad Hoc Causal Questions in Web Search

Emre Kıcıman

emrek@microsoft.com

Microsoft Research

should i **buy a fixer upper or new home?**

should i **rent or buy a home in 2014**

should i **get pre**
before house h

should i **refinan**

should i **get a d**

should i **break u**
boyfriend

should i **pay off**

should i **file bar**

should i **get a fl**

should i **lease o**

should i **join a g**

should i **eat bef**
working out

should i **text him**

should i **buy a chromebook**

should i **quit my job**

should i **pop a blister**

should i **get bangs**

should i **get married**

should i **pop a burn blister**

should i **see a doctor**

should i **consolidate my student loans**

should i **do cardio before or**

too

with a cold

army

box one

florida

hird child

atine

g

yday

cher

a real estate

should i **quit drinking coffee**

should i **shave my head**

These queries are asking …

“what happens after _____”

Expectation Exploration Searches

Everyday, people find themselves in unfamiliar situations and consider the potential outcomes of actions.

- Individuals ask themselves (and Bing!) “What’ll happen if I *do that*?”
- Policy-makers ask “What happens when someone *does that*?”

Goal is to inform.

- Understand consequences of a considered action
- Seeking validation / support for ongoing consequences
- Explore hypotheticals

Databases answer other questions

But, no database or knowledge base answers expectation exploration questions

- Individuals rely on articles, coaches, friends' advice, and gut instinct.
- Policy-makers rely on experiments, data analyses, and gut instinct.

Capturing outcomes from longitudinal social media data

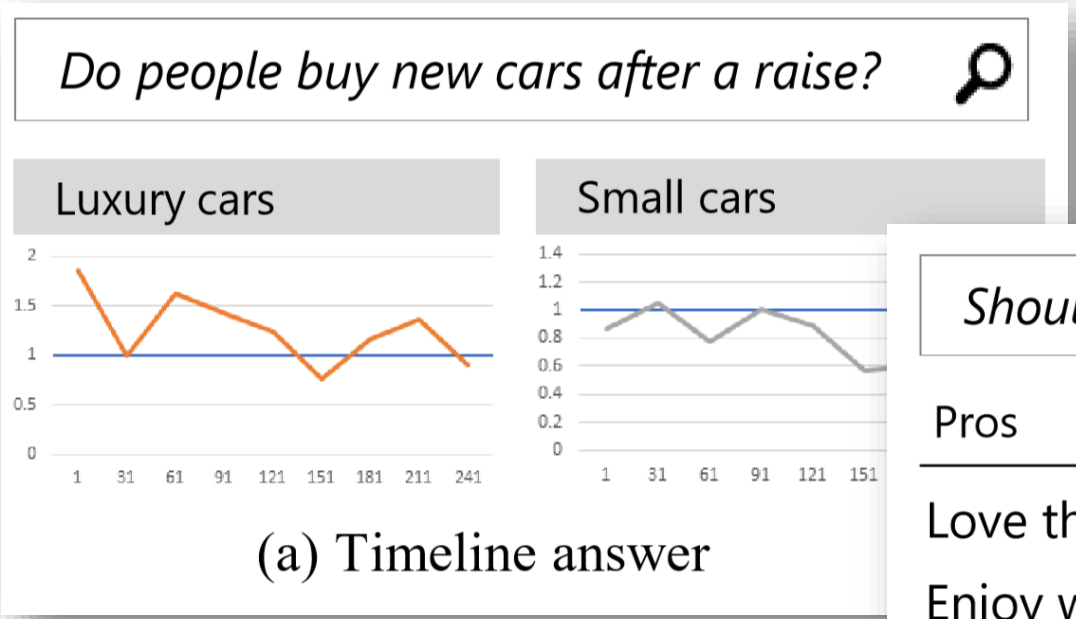
The information we need is already being recorded

- Millions of people frequently and publicly report the actions they take and, over time, the outcomes they experience
- … and they have been doing this for years

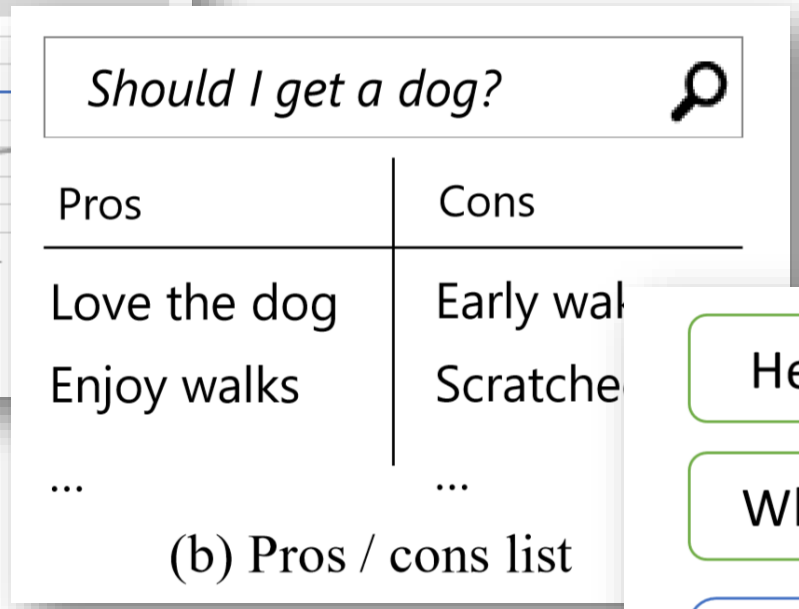
We can *compute* answers to people's questions about an experience

- Use causal inference methods to compare timelines of people who reported the experience to timelines of people who did not.
- Applies to varied data: Twitter, LinkedIn, search logs, financial, …

Building block for IR experiences



(a) Timeline answer



(b) Pros / cons list

(c) Conversational agent

[KDD'15, DESIRES'18]

A new analysis task for IR

To answer "what if", "should I", and other expectation exploration tasks

Corpus: **Rich, Individual-Level Longitudinal Data**

- Today: Social media
- Tomorrow: Many data sources

Query: **What happens after *experience T*?**

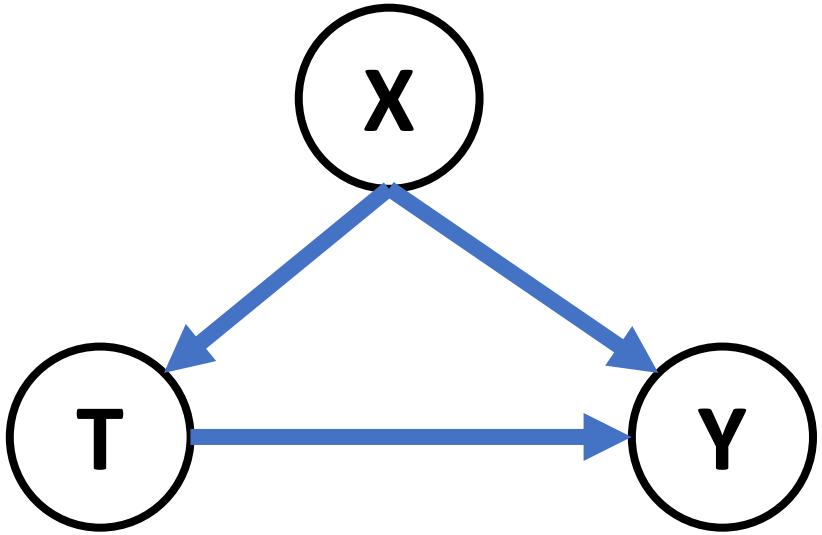
- Need enough information to recognize "*people who did **T***"

Output: ***Expectation Map* captures changes in outcomes over time**

- Building block for many user experiences

What happens after …

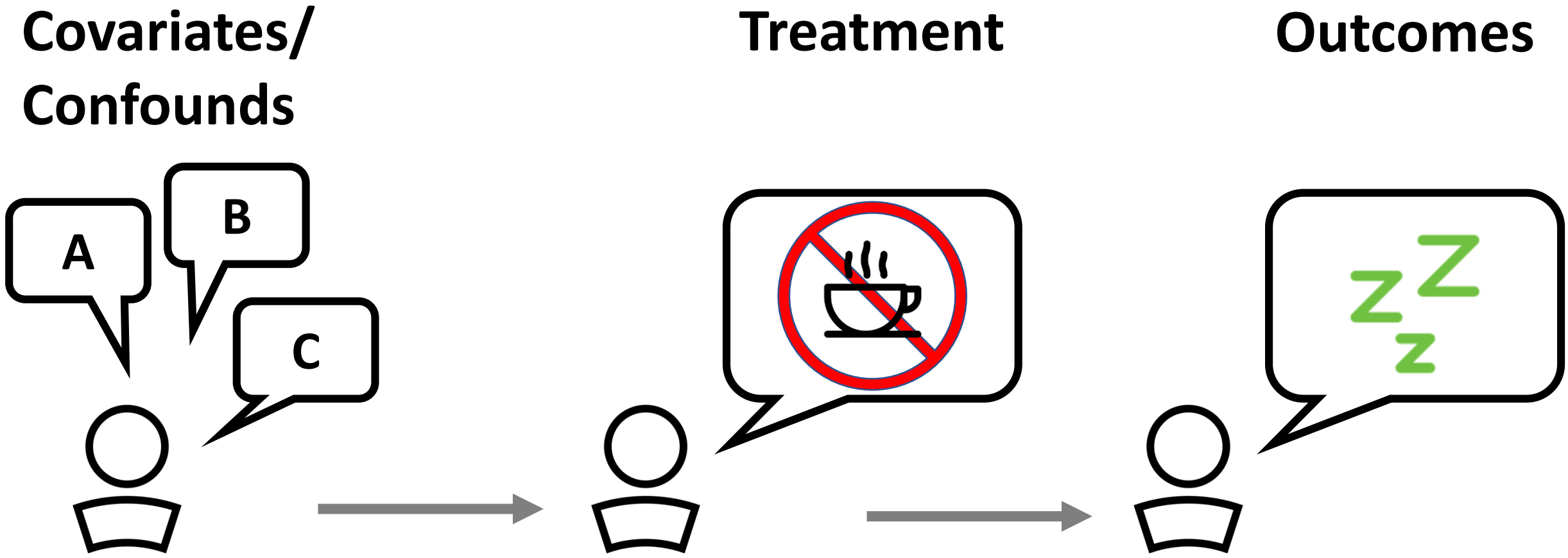

A
B
C



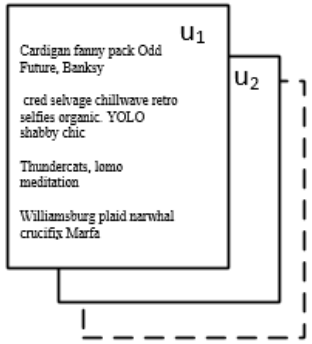
- T is a binary treatment status, identified by a search query
- X is all social media messages that occurred before T, represented as bag of n-grams
- Y is all social media messages that occur after T. E.g., separate effect per word or phrase

Rest of this talk

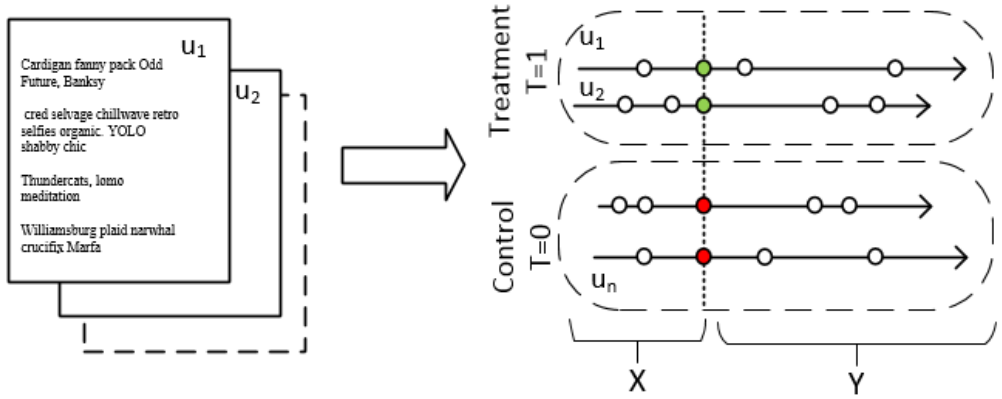
- Causal inference over social media data
- Experiences applying to social sciences problems
- Looking forward



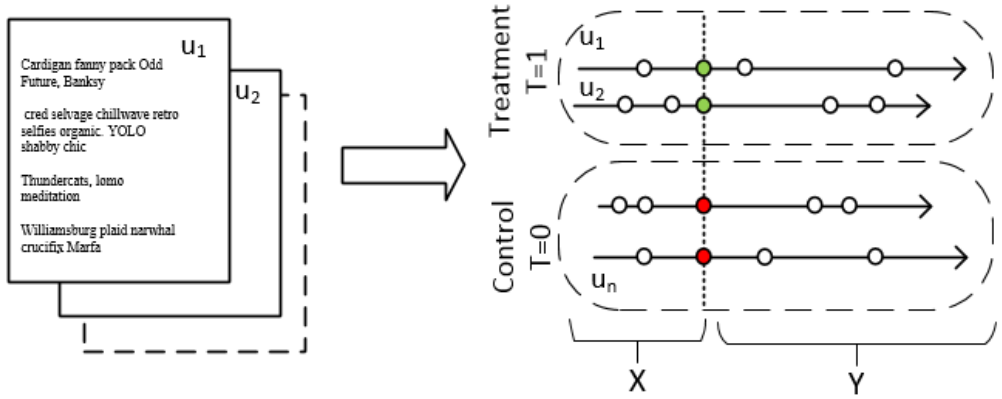
- Treatment identification: ~query-matching (IR / NLP)
- Inference of treatment effects: causal inference
- Interpretability of effects: HCI / IR


u1
Cardigan fanny pack Odd Future, Banksy
cred selvage chillwave retro selfies organic. YOLO shabby chic
Thundercats, lomo meditation
Williamsburg plaid narwhal crucifix Marfa
u2

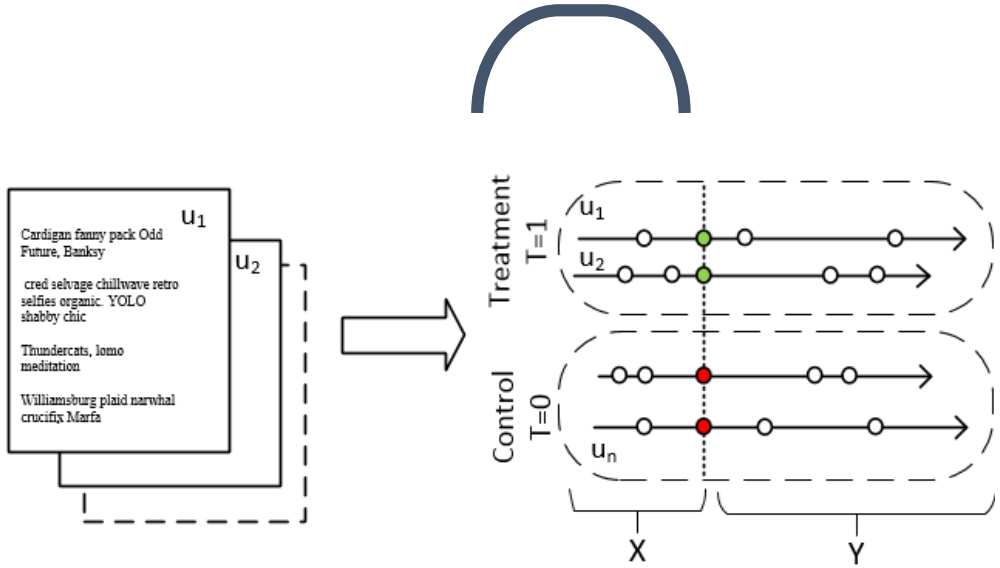
1. Build individual timelines of user activities



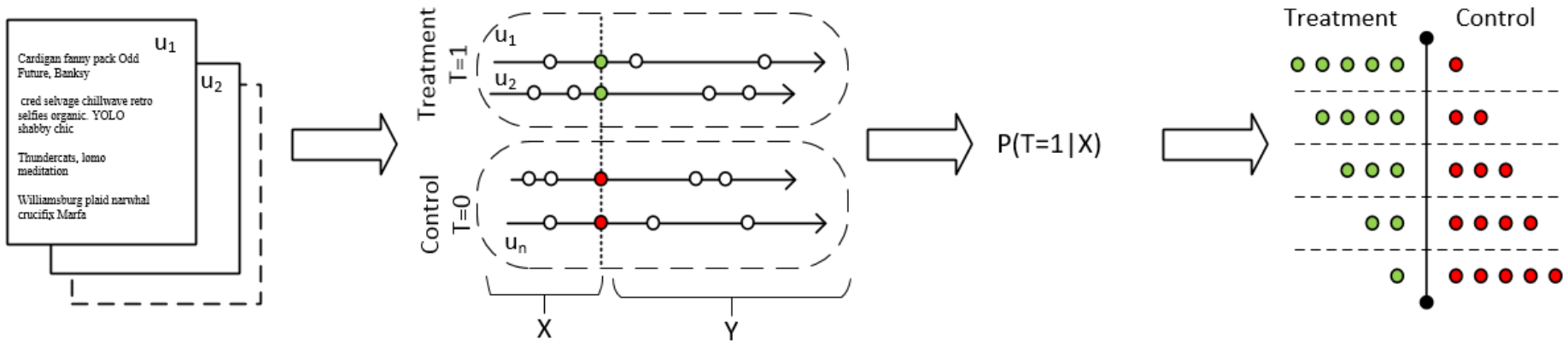
2. Treatment identification



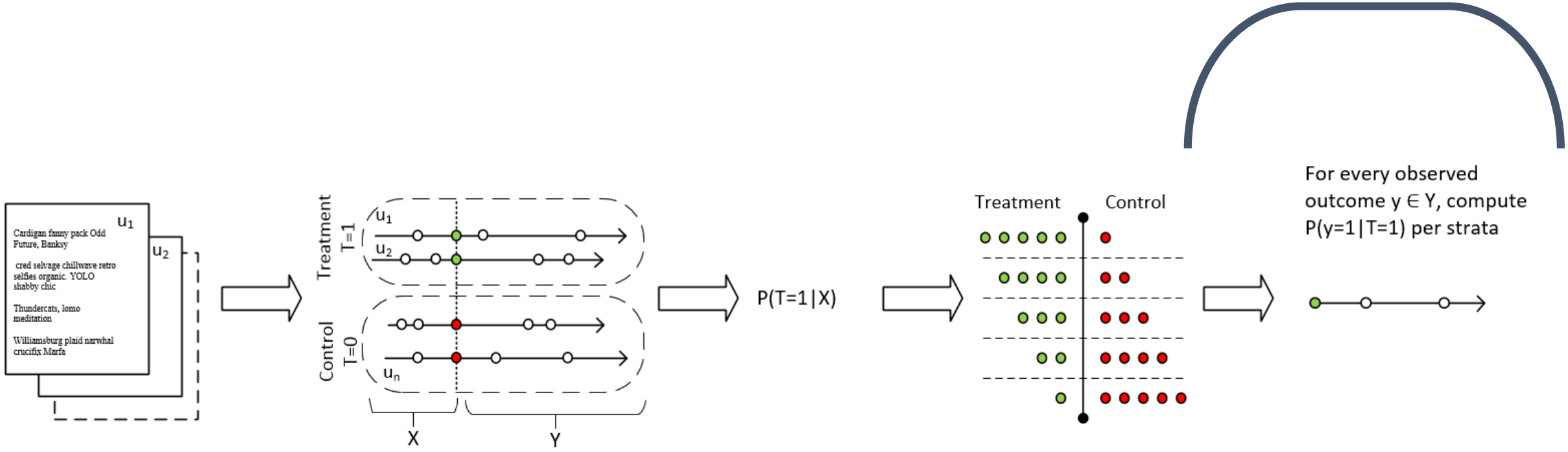
3. Extract covariates



4. Identify similar groups of users



5. Calculate population average outcomes



Output: Expectation Map

Example treatment "ankle sprained"

Elapsed Time	Outcome Y	Effect
Day 1	crutches	50x
Day 1	play soccer	0.10x
…		
Day 42	crutches	1x

What words do people mention more once they say they have **gout**?

Outcome	Count	Absolute Increase	Z-Score
Flare_up	35	4.1%	12.33
Uric_acid	27	2.9%	10.36
Uric	28	2.9%	10.11
Flare	81	4.9%	9.92
Big_toe	38	2.9%	9.86
Joint	301	7.2%	7.22
Aged	32	1.7%	6.51
Correlation	45	2.8%	6.11
Bollock	53	2.5%	5.96
Shite	108	3.4%	5.93

What words do people mention more once they say they have high **triglyceride** levels?

Outcome	Count	Absolute Increase	Z-Score
Your_risk	46	24.8%	18.12
Statin	48	23.1%	17.69
Lower	120	35.9%	17.18
Cardiovascular	54	23.0%	16.72
Healthy_diet	55	19.3%	16.54
Fatty_acid	29	18.3%	16.37
Help_prevent	73	26.9%	16.01
Risk_factor	33	18.3%	15.55
Fish_oil	48	24.4%	15.42
inflammation	78	25.1%	15.30

Rest of this talk

- Causal inference over social media data
- **Experiences applying to social sciences problems**
- Looking forward

Applications in computational social science

- Transitions from mental health discussions to suicidal ideation
  - Explores heterogeneous effects [CHI16], and human assessments of balance [ICWSM17]
  - w/De Choudhury (GATech) et al.
- Daily Behaviors and Sleep and Exercise Quality
  - Combines query logs, location traces and MS Band data [JDSA 18]
  - w/Farajtabar, Nathan, and White
- Conjunction of factors triggering waves of seasonal influenza
  - Explores geographic flow, juxtaposition of multiple large-scale datasets, county-level matching [eLife 2018]
  - w/Chattopadhyay (UChicago), Elliott (UChicago), Shaman (Columbia), Rzhetsky (U Chicago)

College

- Success in college predicts individual career success, career happiness and economic achievement.
- Drives macro-economic growth too.
- But 1 in 3 college students leaves without earning a degree
- Many factors jeopardize college success: family responsibilities, financial pressures, individual behaviors

Alcohol in College

- College students drink more than non-college peers
- Persistent public health issue: hangovers, lowered academic performance, DUI arrests, risky sexual behavior, sexual and other assaults.
- Excessive alcohol consumption is negatively associated with college success
- Studies of college behavior through self-report surveys and interviews; in-person or deliberate recruiting.

What happens to students who mention alcohol more?

Corpus: 5 years of tweets by students starting college in Fall 2010

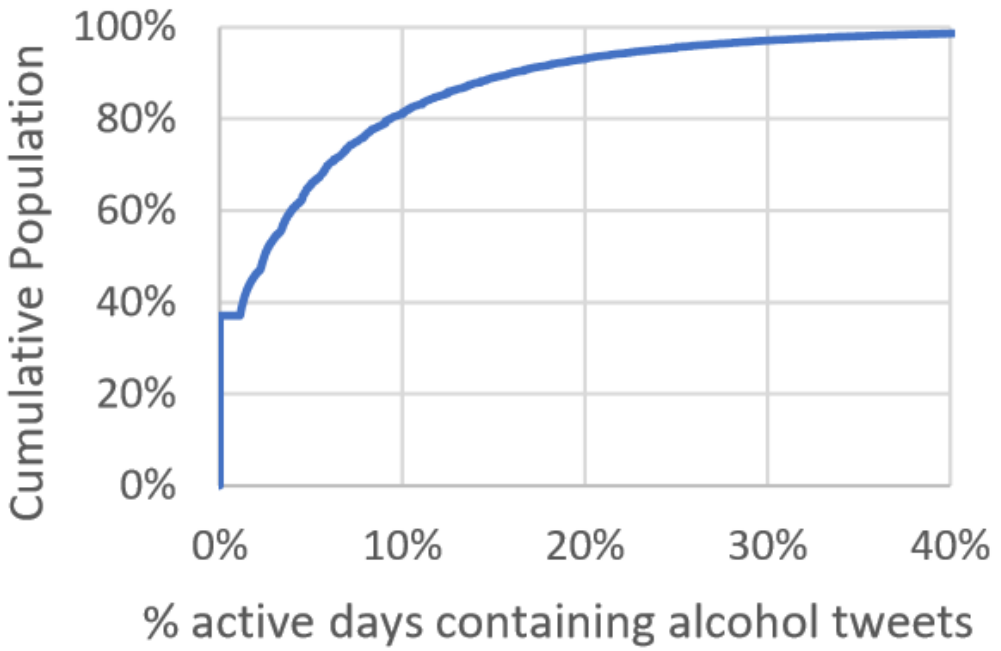
- 63k users, Aug 2010 – May 2015
- ~650M total tweets
- ~12 tweets/day per active person

Treatment: Mention alcohol keywords often during first semester

Outcome: Focus on topics known to be relevant to college success

Table 3. Example tweets matching alcohol keywords

@username hey! getting drunk is not the answer! but I will happily be by your side chipping in.
shouldn't have started drinking this wine :-)
beer in the fridge n I'm ready to go
I hope the club gets snowed in and I'm left with all this vodka
need to do some serious pre drinking so my beer coat warms me up tonight



Identify alcohol keywords in Fall 2010 (9/15-12/15/2010)

Compare top 1/3 of alcohol mentioners (by % active days w/alcohol keywords)

… to bottom 1/3 of alcohol mentioners (no alcohol mentions)

Outcome Topics

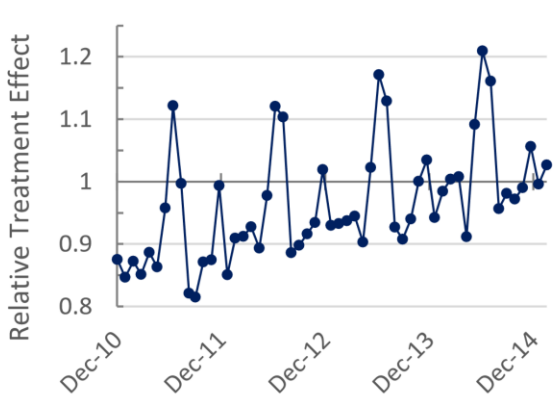
Table 2: Topics linked to college success

Concept	Seed words	Final topic words	Example Tweets
Peer group interaction	friend, boyfriend, girlfriend	boyfriend, buddy, roommate, bandmate, fiance, *+20 more*	Yup buddy I found my bandmate!!
Family responsibilities	mother, father, brother, sister	mother, father, brother, stepdad, grandmother, *+21 more*	Thankful for my little bro and mom I have a sister #fact
Study habits	study, library, homework	study, library, math, tutor, textbooks, worksheets, *+49 more*	anyone that wants to study for history we're in the library but anyways ima off to study
Financial pressures	debt, student loans, loans	wages, afford, utilities, tuition, evicted, fees, *+28 more*	finally my wages wooo @anon its all about money. Im in debt. dont want more loans
Legal/criminal challenges	police, cops, jail, parole	cops, police, restraining, probation, rehab, *+15 more*	meeting my parole officer cops pulling out breathalyzer f***k we drunk

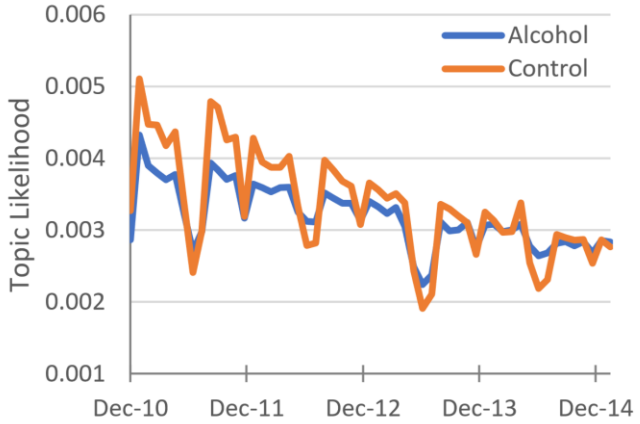
In addition, 195+ validated empath topics

Academic effects

People in the Alcohol group were much less likely to mention studying over the next several years, and somewhat less likely over the entire time period



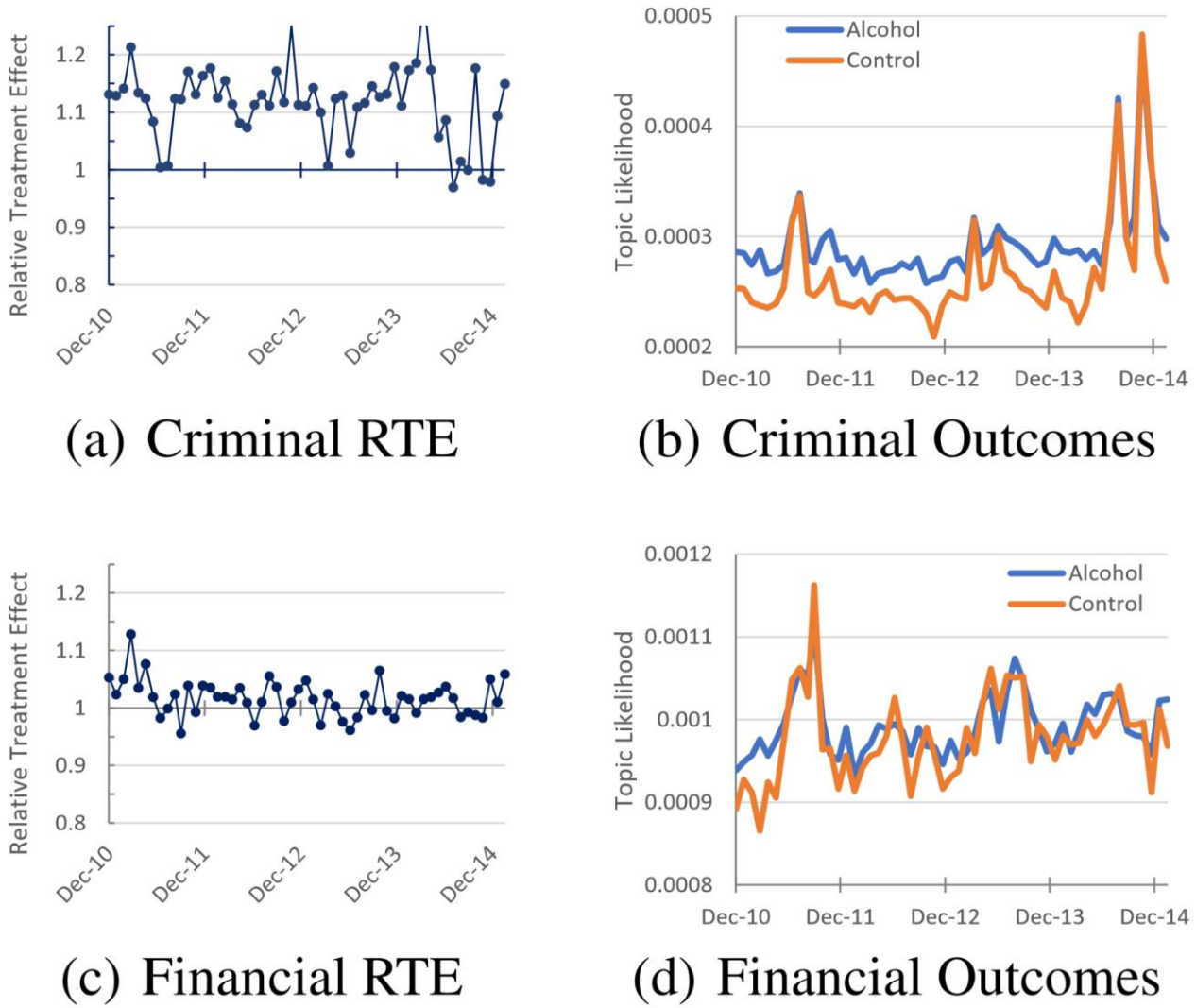
(a) Study Habits RTE



(b) Study Habits Outcomes

Criminal and Financial effects

People in the Alcohol group were more likely to mention legal and criminal challenges through most of our study; and slightly more likely to mention financial pressures over time.



Methodological Limitations (1)

- Population biases require measurement
- Analysis misses language meaning beyond tokens
  - Requires qualitative reading
- Studying changes in conversational behavior only
  - E.g., reporting biases
  - E.g., inverted causal relationships

Methodological Limitations (2)

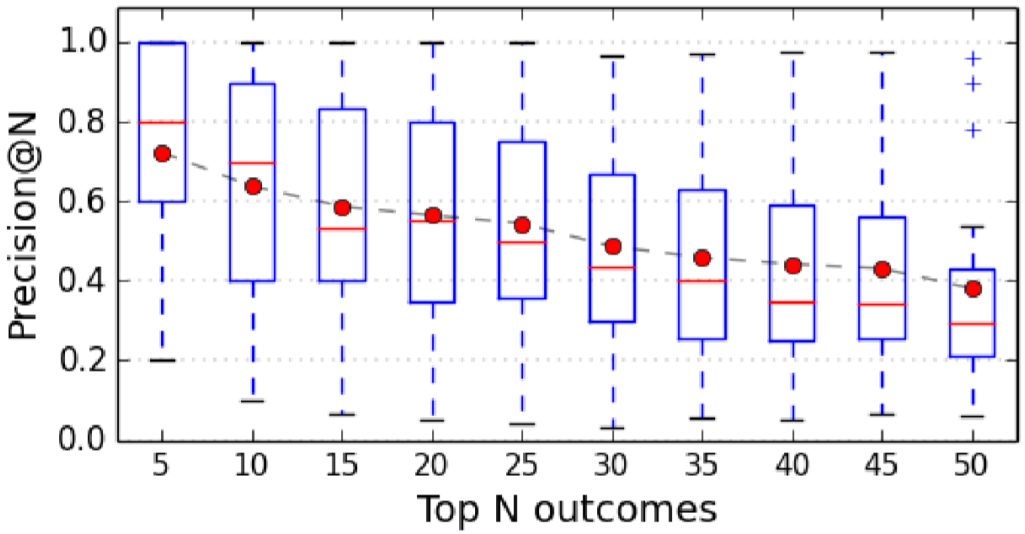
- Assumes ignorability of unobserved confounds
- Assumes SUTVA (no network effects)
- Heterogeneous effects
- Survivor bias.
- When treatment is complex (e.g., when not a point treatment) can be confounded by other experiences

Rest of this talk

- Causal inference over longitudinal data
- Experiences applying to social sciences problems
- **Looking forward**

Does this generalize?

- Calculate consequences of 39 experiences from business, health and society questions
  - E.g., people taking Prozac, getting a divorce, buying life insurance, investing money
  - Experiences selected from Bing search queries
- MTurk workers evaluate surface validity based on supporting social media messages, and online reference material



More broadly, key evaluation criteria must go beyond ***Correctness*** *to* include ***Interpretability*** *and* ***Usefulness***

[CSCW'17]

Opportunities: Core methods

- Improved automation of longitudinal analyses
  - Feature representation for longitudinal
  - Mitigation of known social data biases
  - Under scaling and performance constraints

- Personalization and heterogeneity: Identifying effect modifiers

- Integration of domain knowledge as available

Opportunities: Treatment identification

- Improving query interface

- Expanding classes of treatments beyond point-treatments
  - E.g., Multi-dose treatments

Opportunities: End-to-end task integration

- Interaction and iteration
- Interpretability of results and supporting evidence
- Evaluation methods and criteria for end-to-end task fulfillment

Summary

Everyone asks “what happens after …”

- For decision support, for expectation setting, for exploration, …

Causal inference over longitudinal, individual-level data can give answers

- A building block for new IR answers and experiences

Developed/validated methods in computational social science tasks

- Mental health; sleep and exercise quality; alcohol and college trajectories; …

Now: Scale, interaction, interpretability to make broadly accessible

- Many exciting research opportunities

Questions?

@emrek; emrek@microsoft.com

Selected papers mentioned in the talk (all at http://kiciman.org/)

- Kıcıman, Richardson, Towards Decision Support and Goal Achievement: Identifying Action-Outcome Relationships from Social Media. KDD'15
- De Choudhury, Kıcıman, Dredze, Coppersmith, Kumar, Shifts to Suicidal Ideation from Mental Health Content in Social Media. CHI'16 *honorable mention*
- Olteanu, Varol, Kıcıman, Distilling the Outcomes of Personal Experiences: A Propensity-scored Analysis of Social Media. CSCW'17
- Kıcıman, Counts, Gasser, Using Longitudinal Social Media Analysis to Understand the Effects of Early College Alcohol Use. ICWSM'18
- **Kıcıman, Thelin, Answering *What If, Should I* and other Expectation Exploration Queries Using Causal Inference over Longitudinal Data. DESIRES'18**