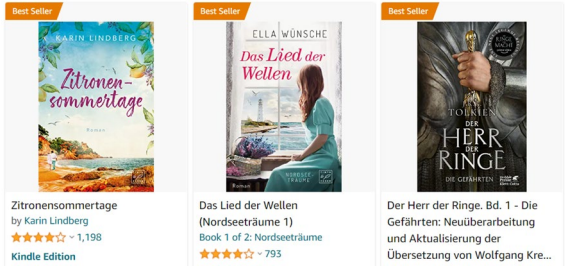
Causal Inference: Why We Should Care

Wolfgang Nejdl, L3S Research Center, Hannov
(with the help of Sandipan Sikdar, Aparup Khatua, L3S)

Predictive Systems on the Web

Recommendation



Mental wellbeing

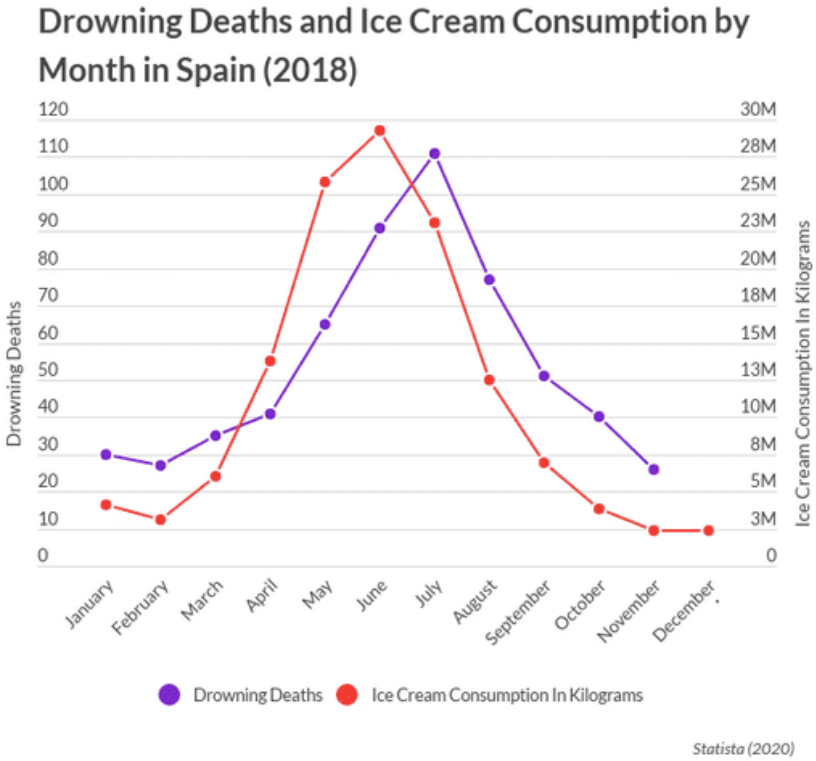
Fake News

https://blog.avast.com/banish-fake-news-think-before-you-share

https://www.businessinsider.com/guides/tech/how-to-pin-a-tweet-on-twitter

Predictive Systems on the Web

- These predictive systems are mostly "*associational*"
  - Leverage large amount data
  - Find correlation patterns
- But correlation is not causation



Correlation is not Causation

- Fake news detection: determine whether a news article is fake based on content
- Simple logistic regression model achieves an accuracy of 78%
- But

FakeNewsNet	
Positive Features	Negative Features
season	trump
at	brad
2018	pitt
the	jenner
awards	jennifier

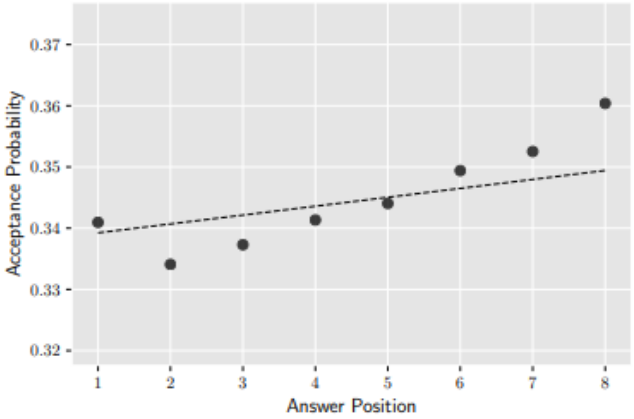
The articles are often clickbaits with celebrity names (Selection bias)

Zhou et. al EACL 2021

Social Media Data Analysis: Simpson's Paradox

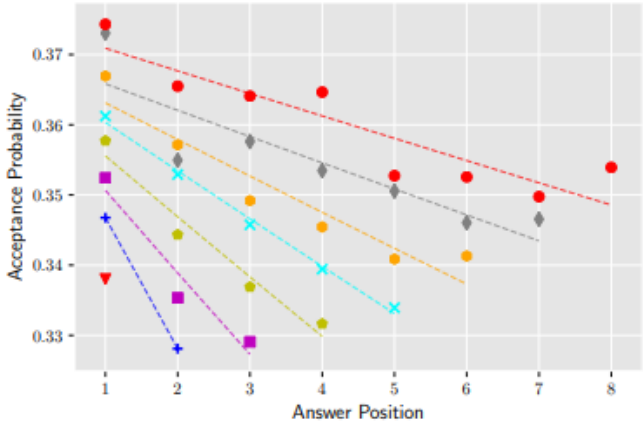
Alipourfard et. al WSDM'2018

- Stack Exchange



(a) Aggregated Data

Answers written later in a session are more likely to be accepted as best answers.



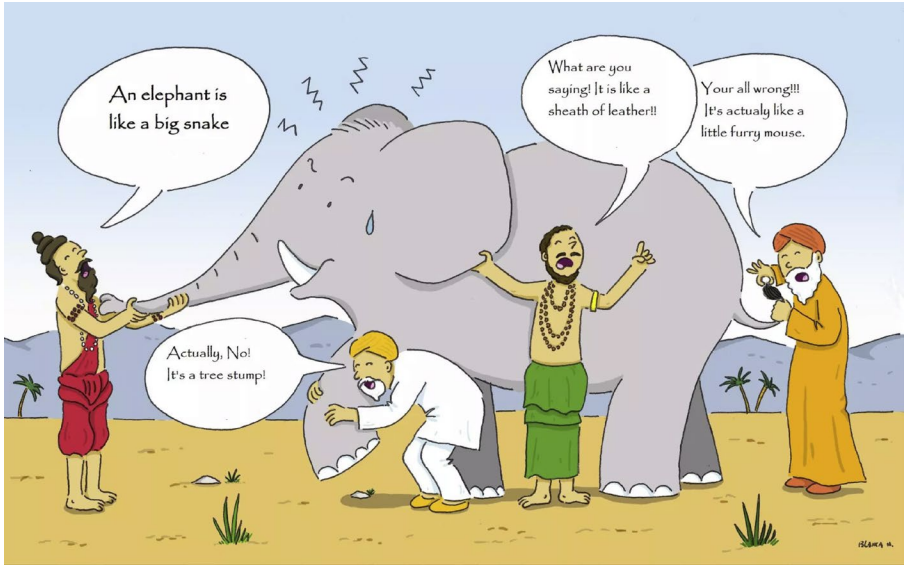
(b) Disaggregated Data

When data is disaggregated by session length, the trend reverses.

Which data do we analyse? Selection Bias …

“What machines are picking up on are not facts about the world. They are facts about the dataset.”

http://nautil.us/issue/40/learning/is-artificial-intelligence-permanently-inscrutable

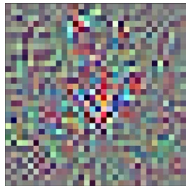


- . . . .

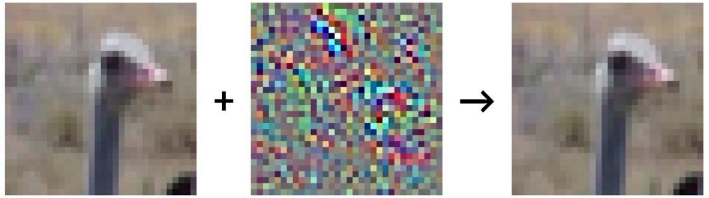
Neural networks find the „best“ features (Madry 2019)

- *Adversarial examples are not bugs, they are features.* Ilyas et al, NIPS 2019.
- https://gradientscience.org/adv/

- A tale about the planet ERM, inhabited by an alien race known as Nets.
- Each individual's place in the social hierarchy is determined by their ability to classify bizarre 32 by-32 pixel images (meaningless to the Nets) into ten completely arbitrary categories.
- These images are drawn from a top -secret dataset, See-Far—outside of looking at those curious pixelated images, the Nets live their lives totally blind.



A TOOGIT, highly indicative of a "1" image. Nets are extremely sensitive to TOOGITs.

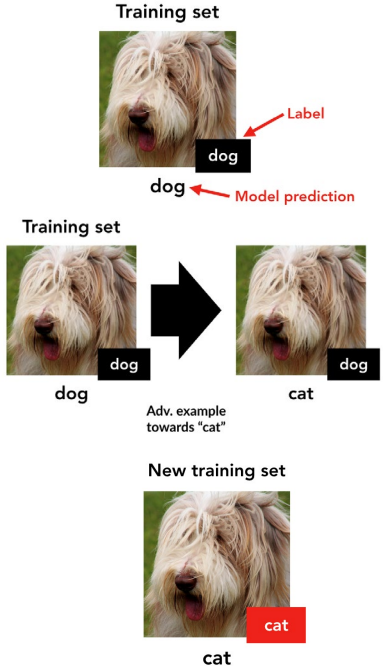


On the left is a "2", in the middle there is a GAB pattern, which is known to indicate "4"—unsurprisingly, adding a GAB to the image on the left results in a new image, *which looks exactly like an image corresponding to the "4" category*.

Neural networks find the „best“ features

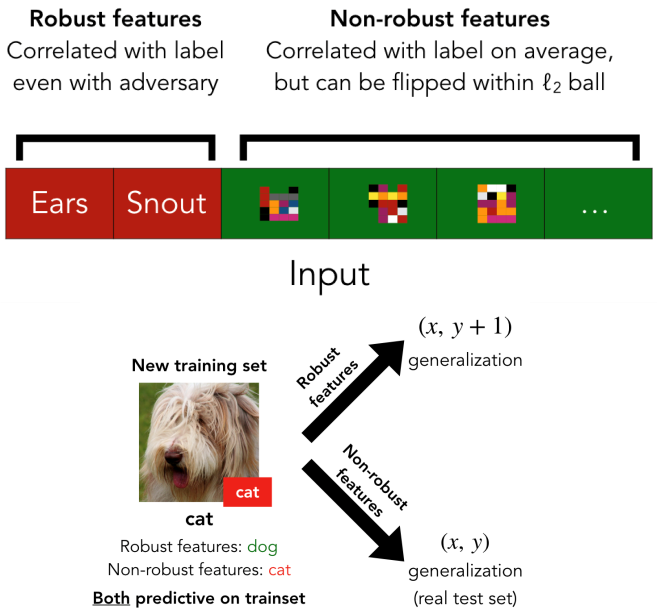
- Start with a CIFAR-10 training set
- Change each example with a targeted adversarial example.
- Construct a new training set based on these (re-labeled) adversarial examples
- Test on the original CIFAR-10 training set

- Remarkably, we find that the resulting classifier actually has moderate accuracy (e.g. 44% for CIFAR)! This is despite the fact that training inputs are associated with their “true” labels *solely through imperceptible perturbations*, and are associated with a different (now incorrect) label matching through *all* visible features.



Neural networks find the „best“ features

- Every training set includes „robust features“ (usually used by humans) and „non-robust features“ (which are brittle and can be disturbed easily)
- Adversarial training tries to disturb these non-robust features to make them useless as discriminators
- Interpretability and causality considerations have to be included already in the training phase
- post-hoc explanation of standard models (which might use these non-robust features) is less useful (as we cannot explain these non-robust features to a human)

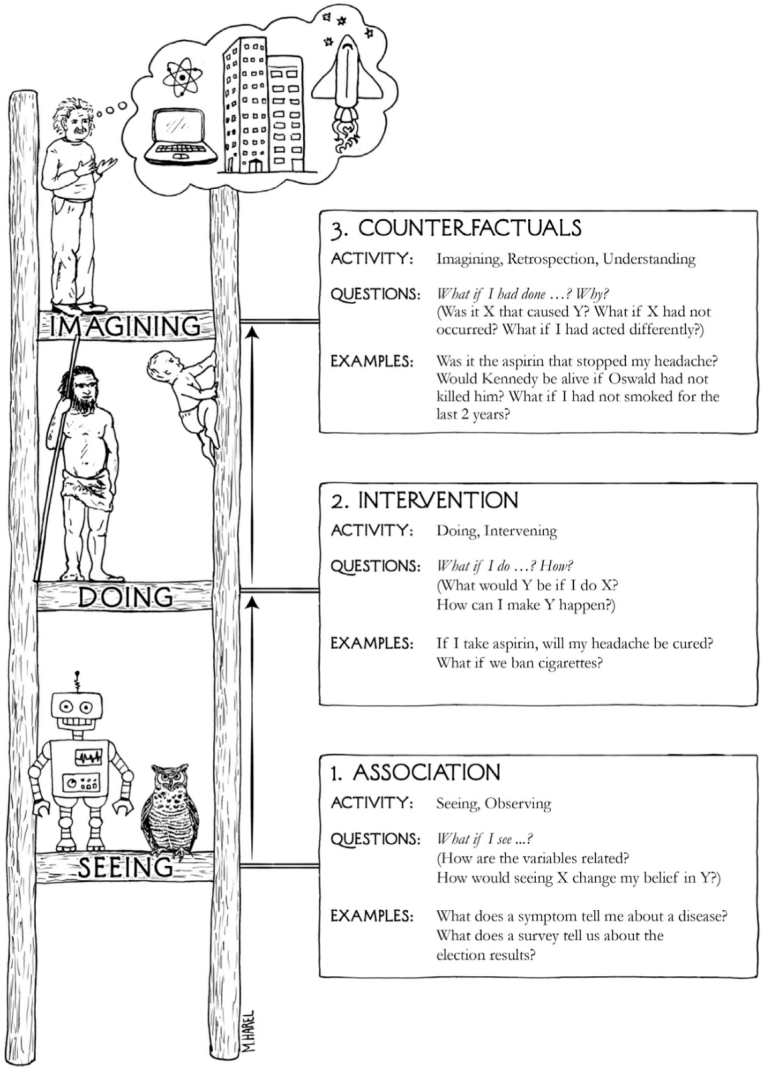


How can causal models help?

- Designing reliable models
  - Models focussing on causal features
  - Getting rid of selection bias
- Reliable data analysis
  - Selection bias
  - Confounders

Pearl's Ladder of Causality

Level (Symbol)	Typical Activity	Typical Questions	Examples
1. Association $P(y\|x)$	Seeing	What is? How would seeing X change my belief in$Y$?	What does a symptom tell me about a disease? What does a survey tell us about the election results?
2. Intervention $P(y\|do(x), z)$	Doing	What if? What if I do X?	What if I take aspirin, will my headache be cured? What if we ban cigarettes?
3. Counterfactuals $P(y_x\|x', y')$	Imagining, Retrospection	Why? Was it X that caused Y? What if I had acted differently?	Was it the aspirin that stopped my headache? Would Kennedy be alive had Oswald not shot him? What if I had not been smoking the past 2 years?



3. COUNTERFACTUALS

ACTIVITY: Imagining, Retrospection, Understanding

QUESTIONS: *What if I had done …? Why?*
(Was it X that caused Y? What if X had not occurred? What if I had acted differently?)

EXAMPLES: Was it the aspirin that stopped my headache? Would Kennedy be alive if Oswald had not killed him? What if I had not smoked for the last 2 years?

2. INTERVENTION

ACTIVITY: Doing, Intervening

QUESTIONS: *What if I do …? How?*
(What would Y be if I do X?
How can I make Y happen?)

EXAMPLES: If I take aspirin, will my headache be cured? What if we ban cigarettes?

1. ASSOCIATION

ACTIVITY: Seeing, Observing

QUESTIONS: *What if I see ...?*
(How are the variables related?
How would seeing X change my belief in Y?)

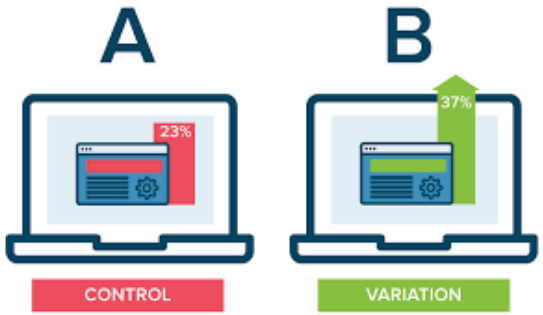
EXAMPLES: What does a symptom tell me about a disease? What does a survey tell us about the election results?

Pearl, Mackenzie: The Book of Why, Basic Books, 2020.

Primer on Causality

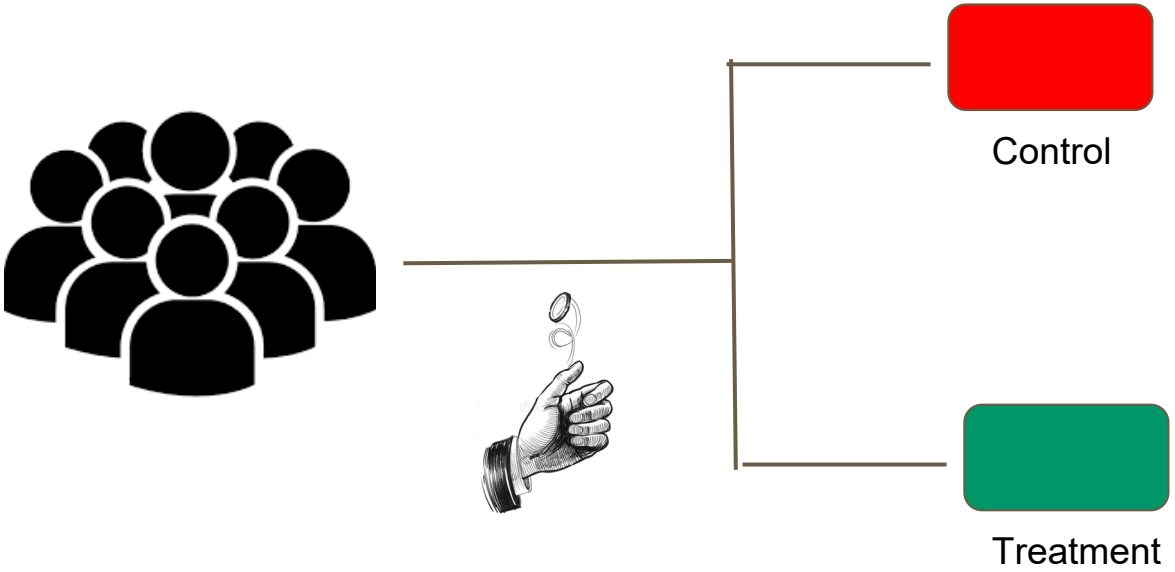
What is causality?

- Science of cause and effect – we can experiment …
  - Random controlled trials in medicine (Is the drug effective?), Web A/B testing (Will changing the interface or algorithm lead to more clicks?)



Randomized Control Trials

- We want to understand if X causes Y (e.g., whether changing the appearance of the website (X) increases number of clicks (Y))



Experimental vs. Observational Data

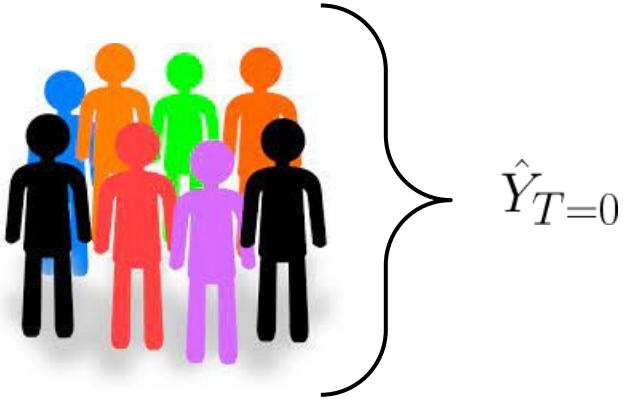
- Performing experiments is often not possible
- Only observed data is available
  - Experiments might not have been performed perfectly
  - Selection bias when deciding control/treatment individuals
  - Much easier to collect data on the Social Web than to do experiments
- How can we measure causal effect with observed data?
- Two models
  - Potential outcome framework
  - Structural causal models

Potential Outcome Framework

- We want to find the effect of a “treatment”, i.e. which effect does Covid -19 misinformation (T=1) have on anxiety?
- “Missing data”: We only observe either $Y_{T=1}$ or $Y_{T=0}$
- We only have observed data and cannot perform experiments

Potential Outcome Framework

- "Missing data": We only observe either $Y_{T=1}$ or $Y_{T=0}$
- Estimate missing data using various methods
- $Y_{T=0}$ now becomes an estimated quantity based on other people who did not receive the treatment



$\hat{Y}_{T=0}$

Potential Outcome Framework

- A person's treatment affects only her outcome and does not affect others' outcome
  - Stable unit treatment value assumption (SUTVA)
  - This is in general NOT true in social networks!
- Nothing systematically different between treated and untreated people
  - Only plausible if the treatment was randomized
  - But we are dealing with observed data, hence selection bias etc …
  - Which part of the population is using Twitter, TikTok, etc?
  - We need to take demographic information into account. Advertisement networks have been doing this for a long time …

Matching

- Barring randomization, treated and untreated people can be different
- Not clear if the estimated effect is due to the treatment or differences in people's characteristics
- Compare people with same characteristics: Matching
- Develop a conditional estimator

Estimated causal effect

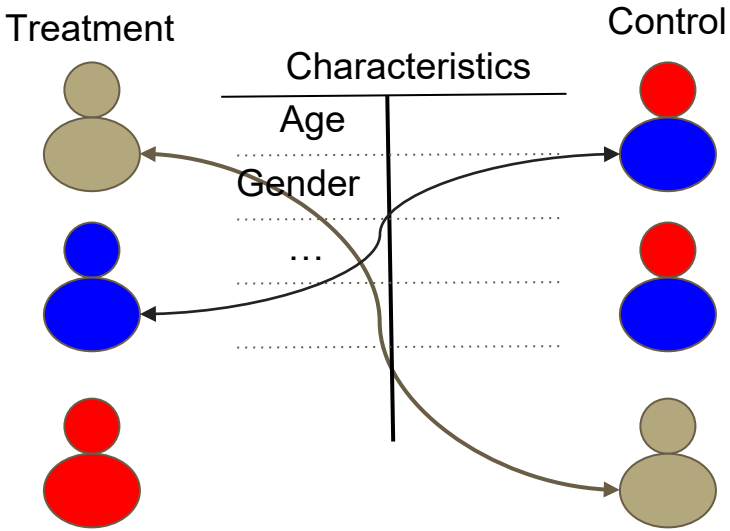
$$= E[Y_{T=1} - \hat{Y}_{T=0}]$$

$$= E[Y|X, T = 1 - Y|X, T = 0]$$

People's characteristics

Matching

- Match individuals on the control and treatment group based on the observed characteristics



Exact matching
(often not possible)

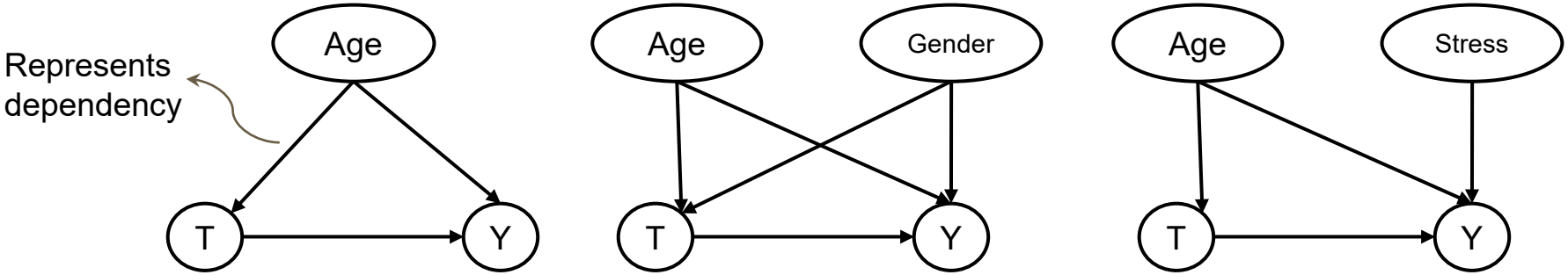
Similarity measured through metrics -

- Mahalanobis distance
- Propensity score based on various attributes

Modeling Dependencies

- People may have interrelated characteristics
  - How are these characteristics inter -related?
- Other factors can influence the observed outcome
  - How do they influence treatment and outcome?
  - Which factors should be included?
- When is it possible to find a causal effect?
  - Graphical models framework

Graphical Models

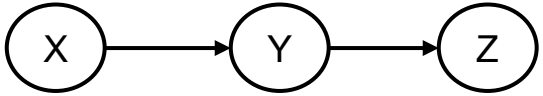


X = {Age}

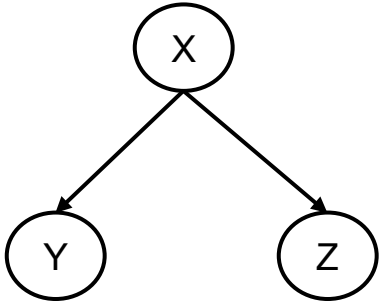
X = {Age, Gender}

X = {Age}

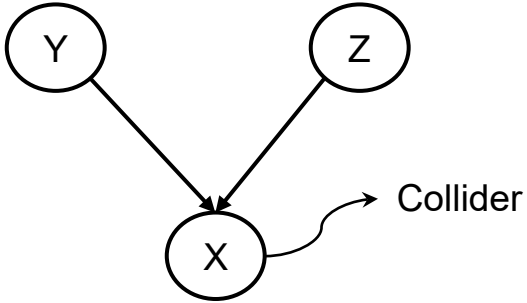
Graphical Models



Chains

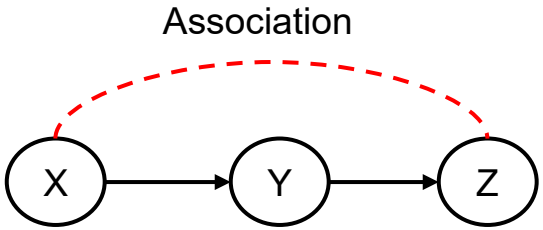


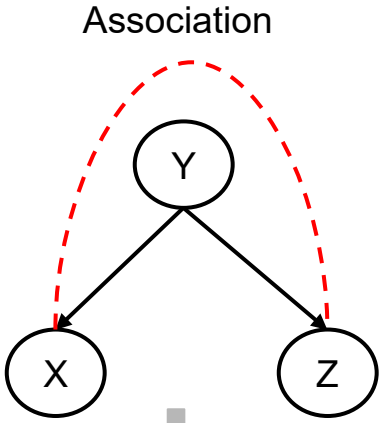
Forks



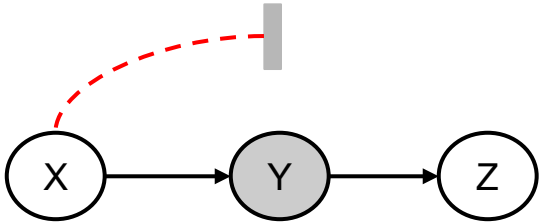
(Immorality)

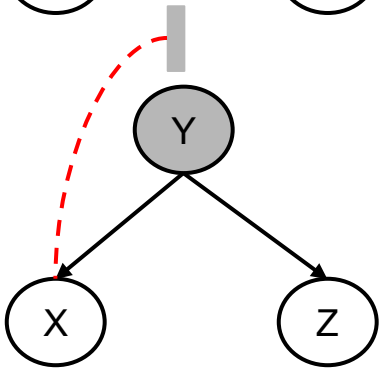
Graphical Models: (In)dependence





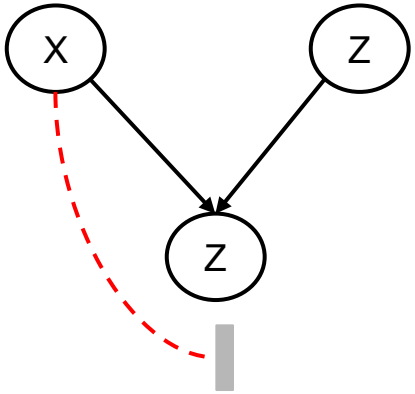
Unblocked path (X and Z are dependent)



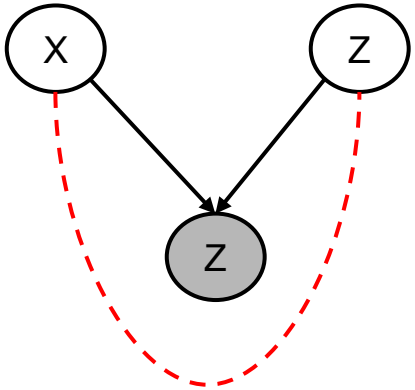


Blocked path (X and Z are independent)

Graphical Models: (In)dependence



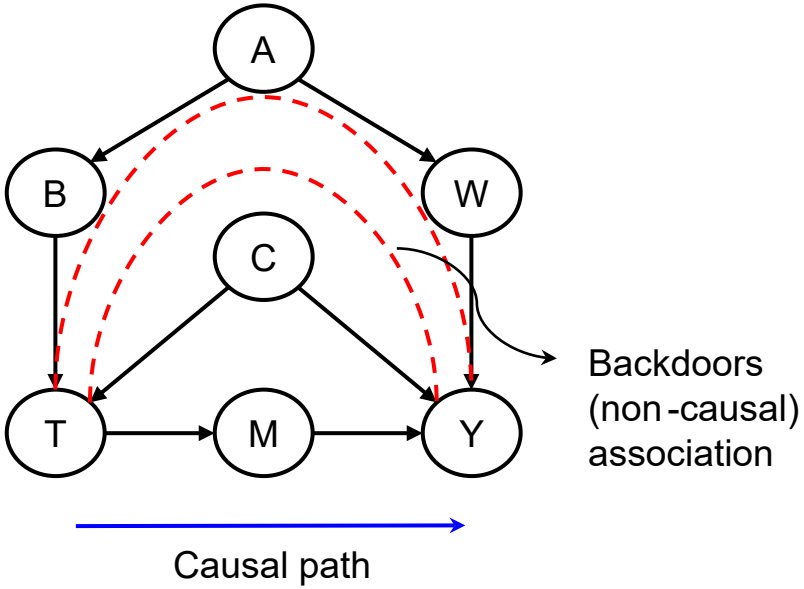
Blocked path

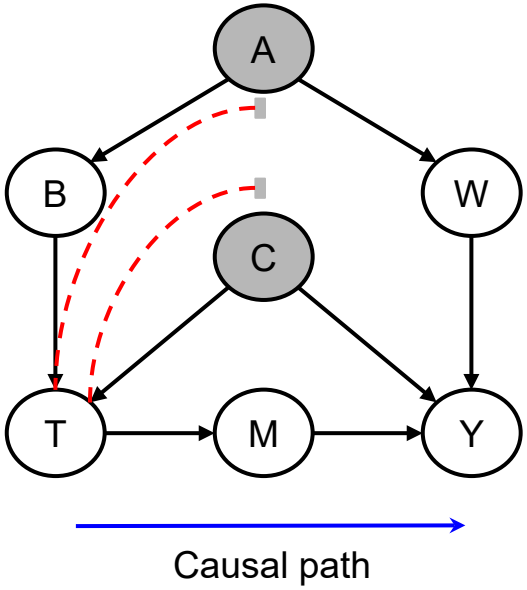


Unblocked path

- Conditioning on the collider or any of its descendants unblocks the path
- D-separation in probabilistic graphical models

Graphical Models: Backdoor adjustments





Structural Causal Models

Structural equation for A as a cause of B

$$B := f(A)$$

Equality does not convey any causal information

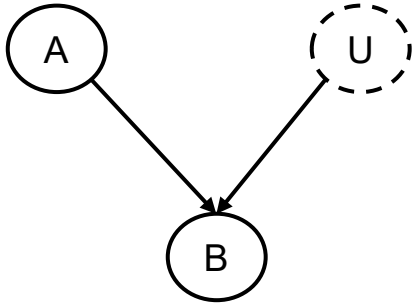
Unobserved characteristics: Incorporates stochasticity

$$B := f(A, U)$$

Causal mechanism :

$$X_i := f(A, B, \ldots)$$

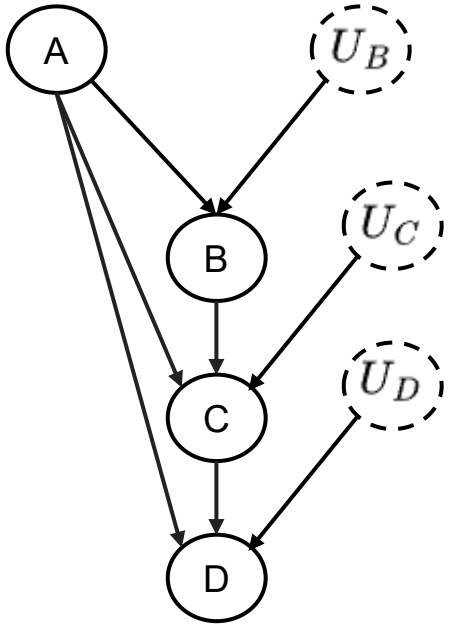
Parents of X_i



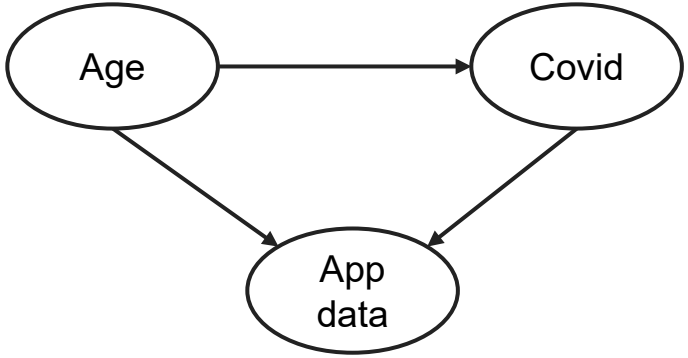
Structural Causal Models

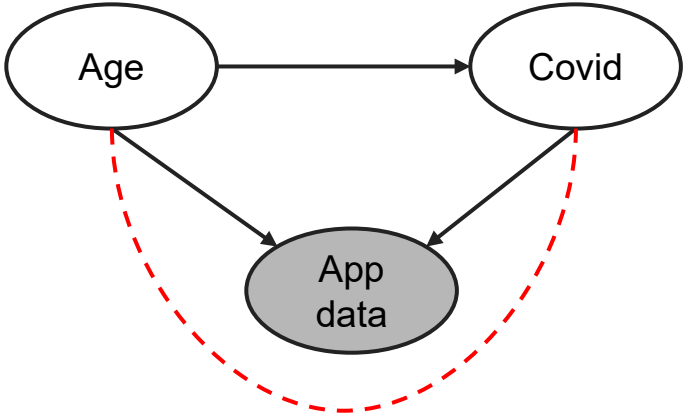
$$
\begin{aligned}
B &:= f_B(A, U_B) \\
M: \; C &:= f_C(A, B, U_C) \\
D &:= f_D(A, C, U_D)
\end{aligned}
$$

- Set of endogenous variables
- Set of exogenous variables
- A set of functions, each to generate a endogenous variable from other variables



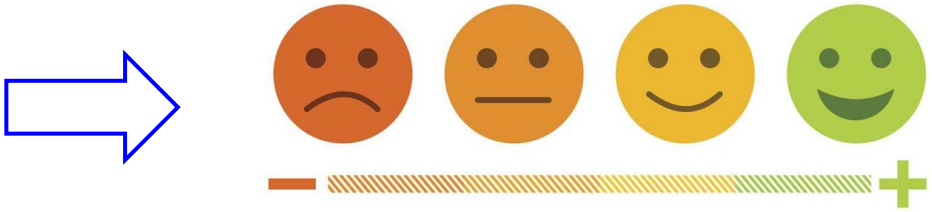
Structural Causal Models: Collider Bias





- Conditioning on the collider (App data – collected from mobile applications) opens up an unintended association between treatment and outcome variables
- Graphical models can help in identifying such biases

Causality and Web



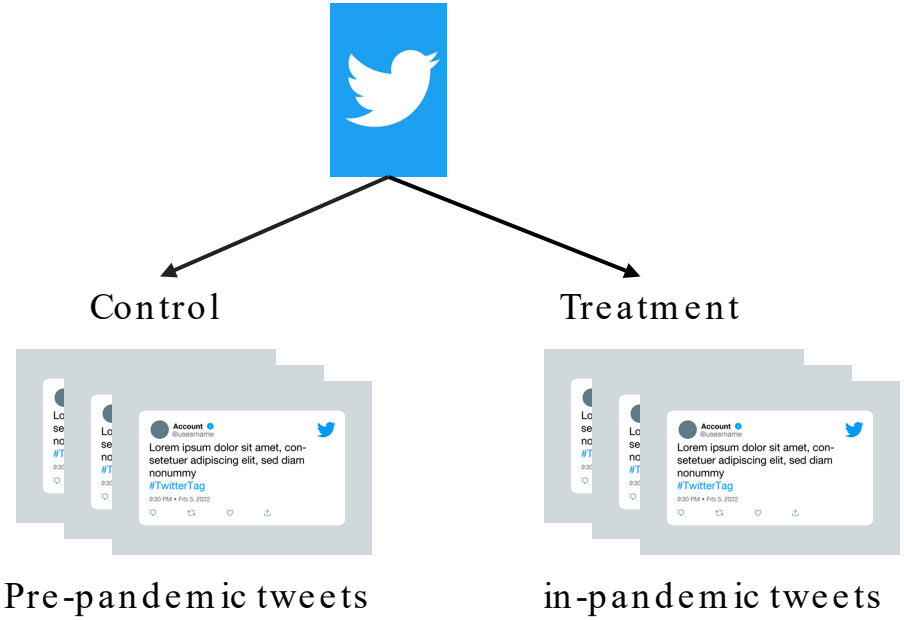
- Social media activity representative of emotional state
- Can we utilize social media data to study mental wellbeing?
- **Can causality help?**

https://www.thescienceofpsychotherapy.com/emotions-feel-may-shape-see/

https://www.techtimes.com/articles/262044/20210626/facebook-social-media-companies-held-liable-used-criminal-activity.htm

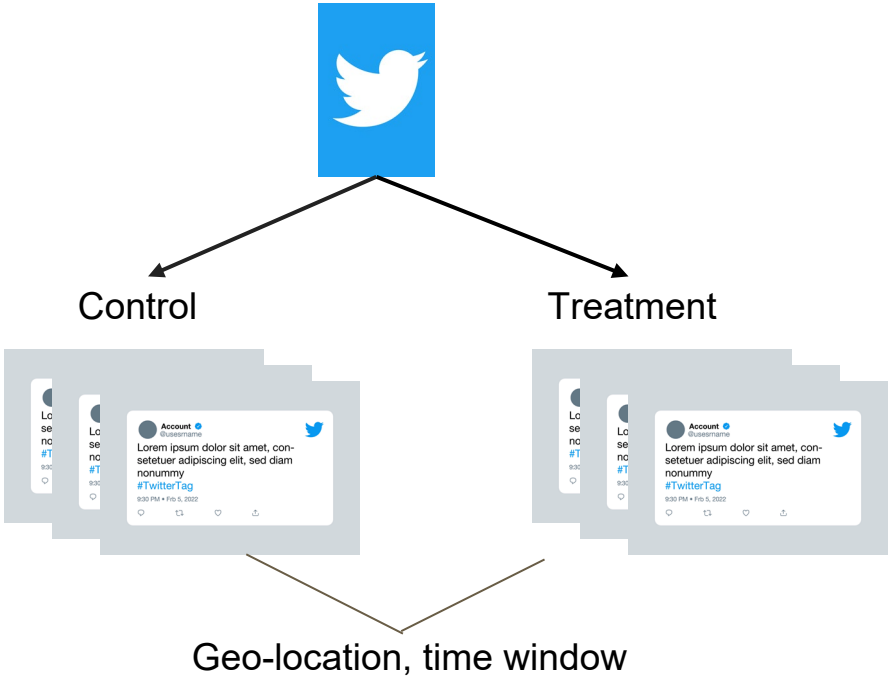
Psychological effects of COVID-19 pandemic

Objective : To study the temporal and linguistic changes in symptomatic mental health and support expressions in the pandemic context.



Saha et. al. (2020). Psychosocial effects of the COVID-19 pandemic: large-scale quasi-experimental study on social media. *Journal of medical internet research*, *22*(11), e22600.

Psychological effects of COVID-19 pandemic



Saha et. al. (2020). Psychosocial effects of the COVID-19 pandemic: large-scale quasi-experimental study on social media. *Journal of medical internet research*, *22*(11), e22600.

Psychological effects of COVID-19 pandemic

Expression	Treatment (2020), mean (SD)	Control (2019), mean (SD)	Δ (%)	Cohen *d*	*t* test (*df*)	*P* value
Symptomatic mental health expressions						
Anxiety	1.65 (0.20)	1.35 (0.08)	21.32	1.96	12.60 (151)	<.001
Depression	9.00 (0.60)	8.17 (0.35)	10.18	1.71	10.72 (151)	<.001
Stress	19.31 (0.77)	18.61 (0.43)	3.76	0.81	3.65 (151)	.009
Suicidal ideation	3.14 (0.31)	2.62 (0.13)	19.73	2.14	13.54 (151)	<.001

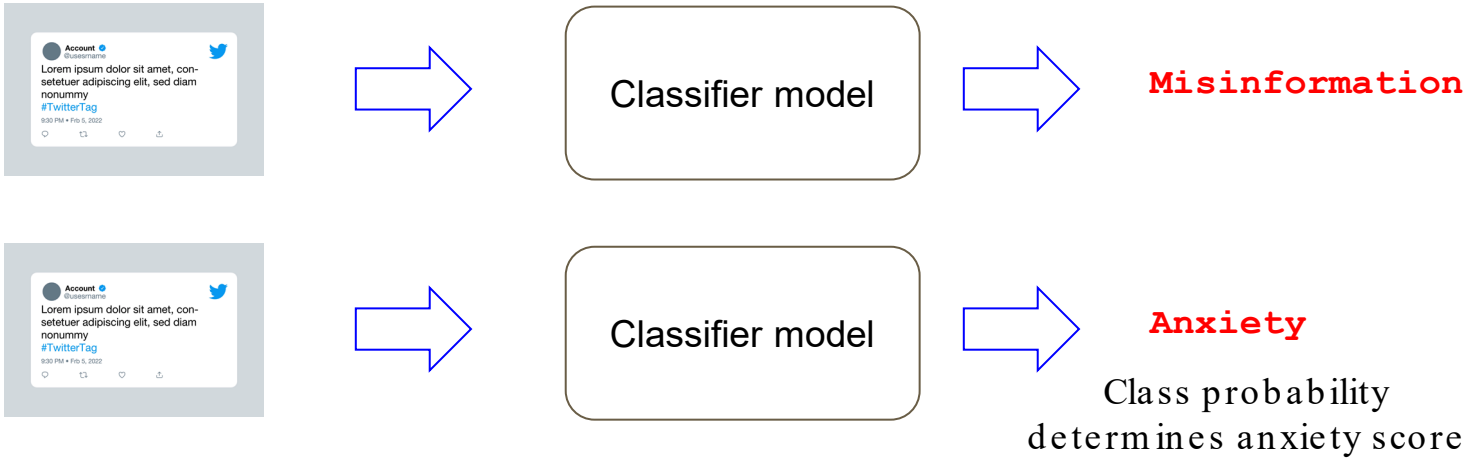
Significant differences in social media expressions from pre - to in -pandemic period

Saha et. al. (2020). Psychosocial effects of the COVID-19 pandemic: large-scale quasi-experimental (2019 / 2020 same time / same region datasets from Twitter API) study on social media. *Journal of medical internet research*, *22*(11), e22600.

Impact of sharing COVID-19 misinformation

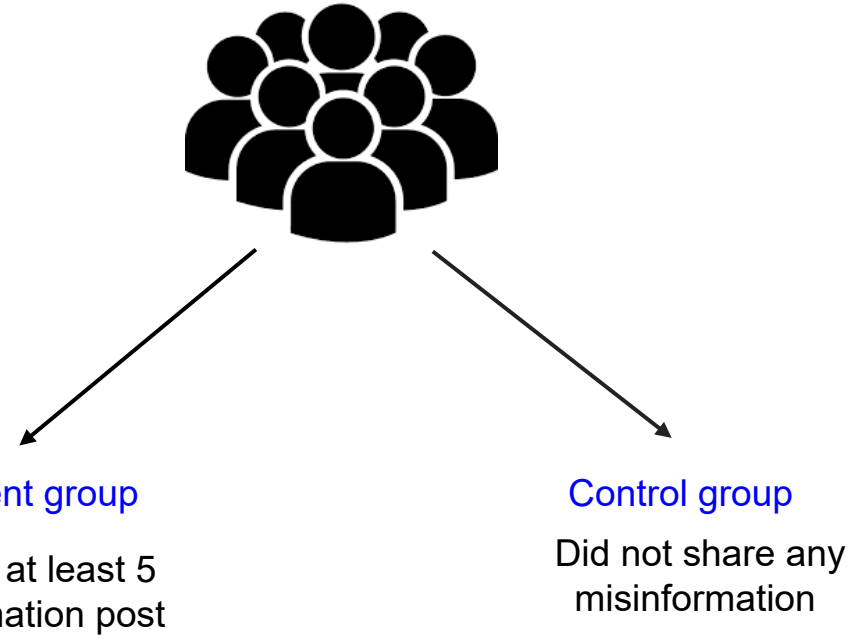
Hypothesis : Consuming misinformation online can potentially worsen the mental health of individuals

Data: 80 million Twitter posts made by 76,985 Twitter users during an 18.5 month period.



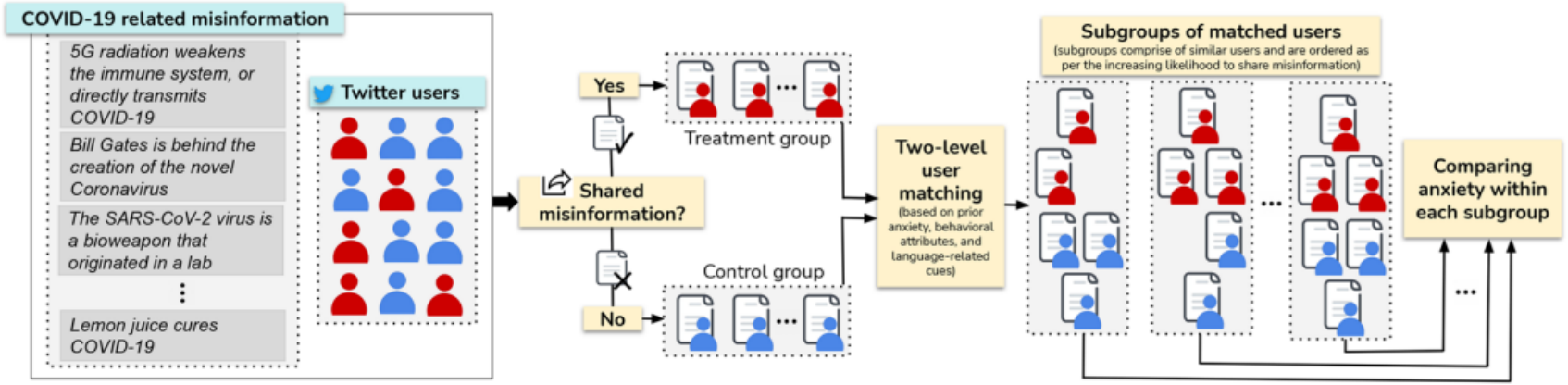
Verma, G., Bhardwaj, A., Aledavood, T., De Choudhury, M., & Kumar, S. (2022). Examining the impact of sharing COVID-19 misinformation online on mental health. *Scientific Reports*, *12*(1), 1-9.

Impact of sharing COVID-19 misinformation



Verma, G., Bhardwaj, A., Aledavood, T., De Choudhury, M., & Kumar, S. (2022). Examining the impact of sharing COVID-19 misinformation online on mental health. *Scientific Reports*, *12*(1), 1-9.

Impact of sharing COVID-19 misinformation



$$TE_i^{rel} = \frac{(A_{after}^{trt} - A_{before}^{trt}) - (A_{after}^{ctrl} - A_{before}^{ctrl})}{A_{after}^{ctrl} - A_{before}^{ctrl}}.$$

Users matched based on prior anxiety, twitter interaction and linguistic cues

Verma, G., Bhardwaj, A., Aledavood, T., De Choudhury, M., & Kumar, S. (2022). Examining the impact of sharing COVID-19 misinformation online on mental health. *Scientific Reports*, *12*(1), 1-9.

Effects of early alcohol use on college success

Hypothesis : Early alcohol use has an effect on college success

Data: Twitter data of 63k college students over 5 years

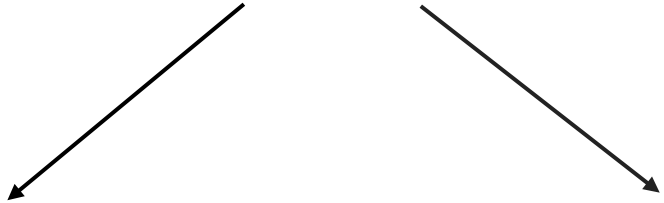
Identifying students entering college : Tweets with phrases such as "*Can't wait to start college next week*"

Identifying tweets mentioning alcohol usage : Tweets with phrases such as "*beer in the fridge n I'm ready to go*"

Topics related to college success : "*Peer group interaction*", "study habits", "legal/criminal challenges" (identifies through specific keywords)

Kiciman, E., Counts, S., & Gasser, M. (2018, June). Using longitudinal social media analysis to understand the effects of early college alcohol use. In *Twelfth International AAAI Conference on Web and Social Media*.

Effects of early alcohol use on college success



Treatment group

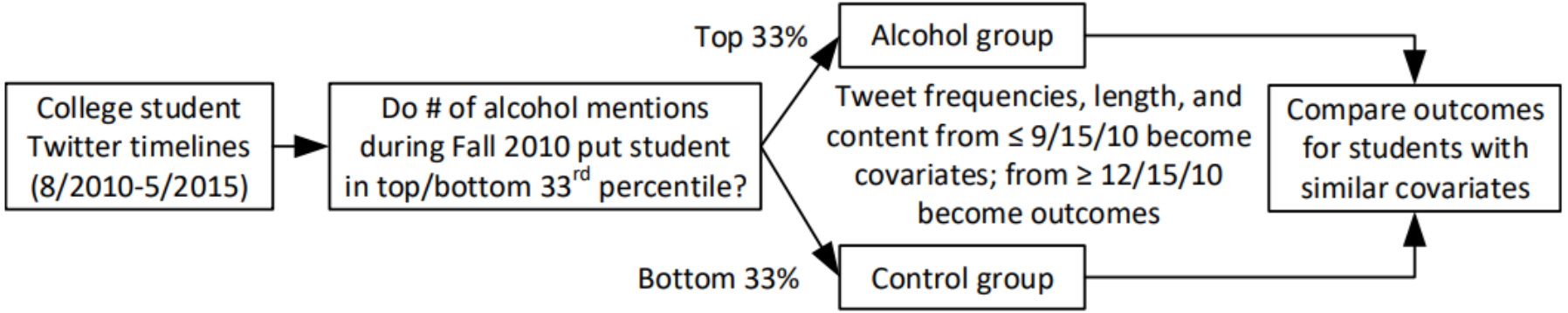
top 33% of the population, as measured by the number of alcohol -related tweets during first 3 months of college

Control group

Did not tweet about alcohol

Kiciman, E., Counts, S., & Gasser, M. (2018, June). Using longitudinal social media analysis to understand the effects of early college alcohol use. In *Twelfth International AAAI Conference on Web and Social Media*.

Effects of early alcohol use on college success



Relative treatment effect (RTE) : Relative treatment effects are the ratio of the topical word likelihood for the Control group divided by that for the Alcohol group.

Covariates: Tweet frequency, tweet length and word distribution

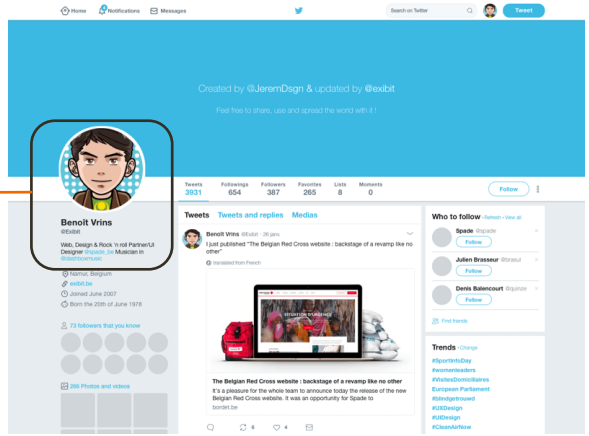
Kiciman, E., Counts, S., & Gasser, M. (2018, June). Using longitudinal social media analysis to understand the effects of early college alcohol use. In *Twelfth International AAAI Conference on Web and Social Media*.

Opinion mining of migration related tweets

Objective : Examine whether twitter opinion changes depending on the country

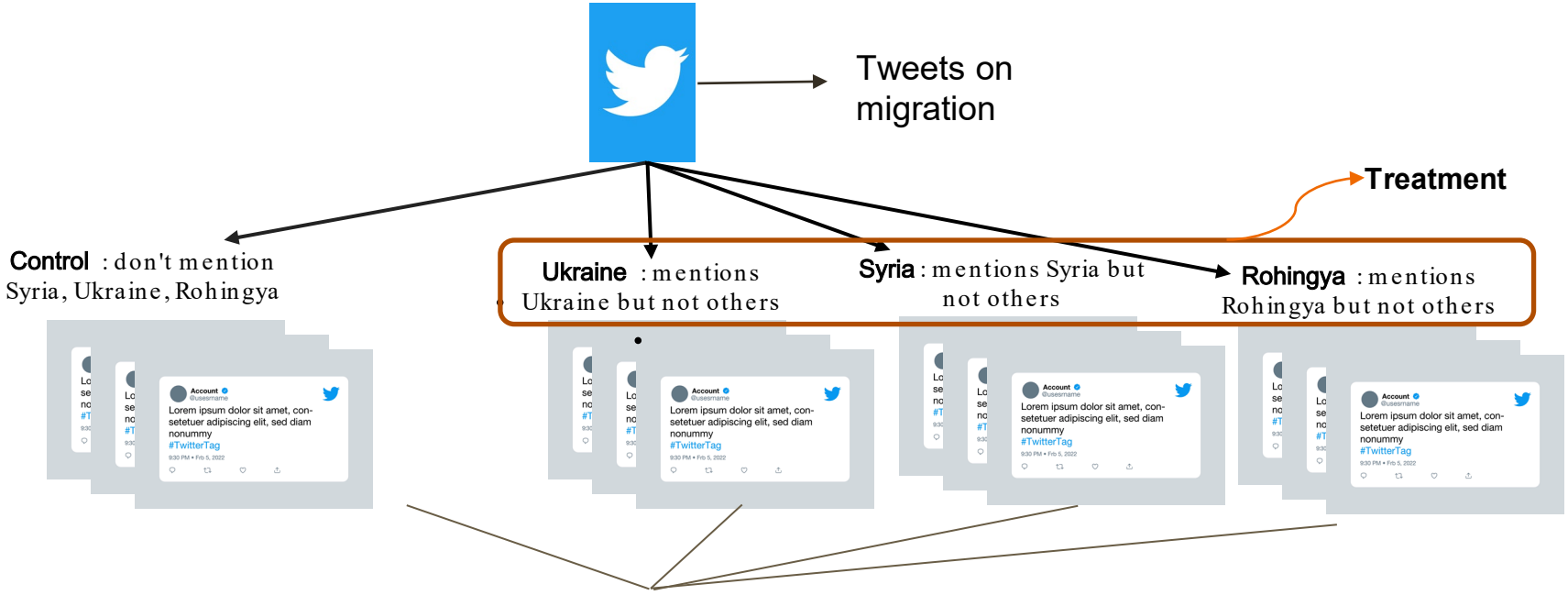
Data: ~4M English tweets with keywords like migrant, refugee, immigrant, asylum seeker, migration, migration policy

Utilize profile picture, name and twitter description to obtain demographic information (Gender and Age) - Wang et. al (2019)



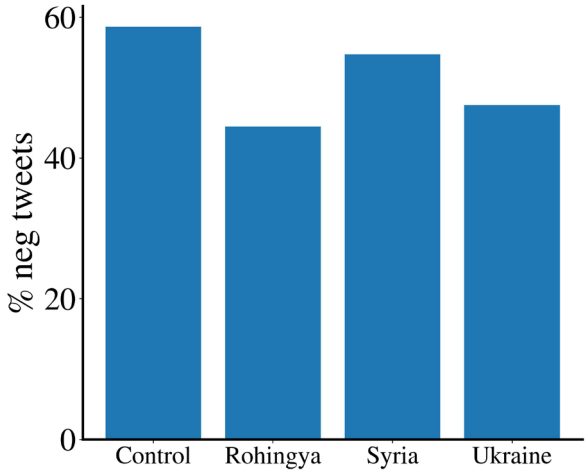
https://www.sketchappsources.com/free-source/3114-twitter-profile-template-sketch-freebie-resource.html

Opinion mining of migration related tweets



Opinion mining of migration related tweets

Overall effect

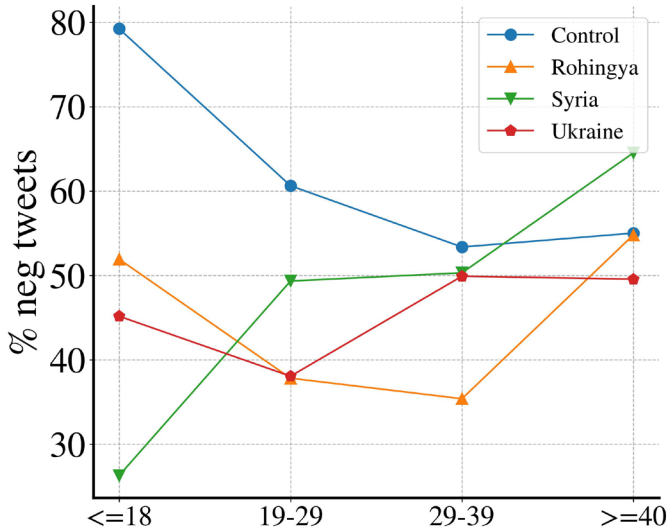


% of negative tweets are lower for Rohingya and Ukraine

CAVEAT: do not trust our sentiment classifier, example used only for illustration purposes!

Opinion mining of migration related tweets

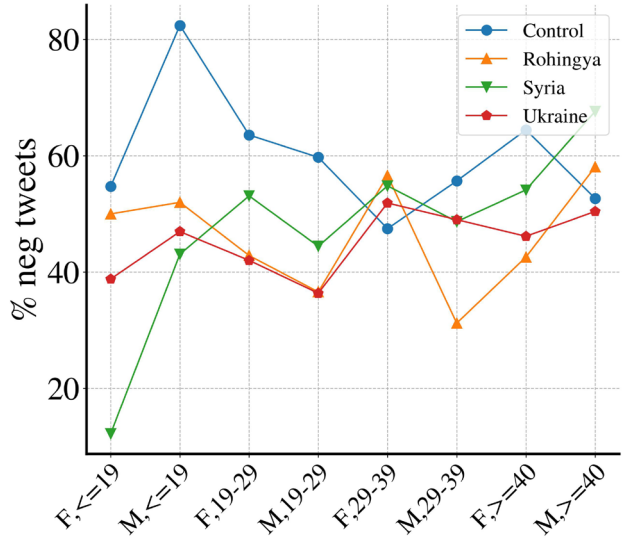
What if we condition on age?

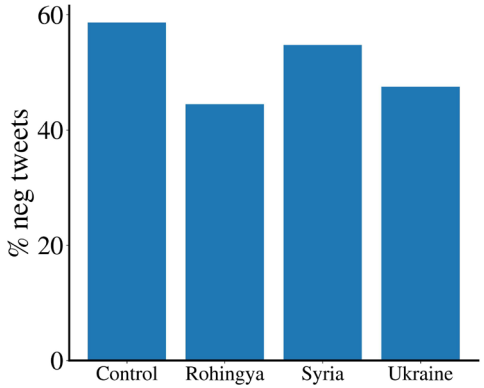


- We observe an opposite trend for Syria and Ukraine than control
- Rohingya roughly follows a similar trend
- People <=18 are much less negative about Syria than in general about migration

Opinion mining of migration related tweets

What if we condition on both age and gender?





Significant differences from control are observed when considering different demographic subgroups.

Causality in Social Media

- Olteanu , Alexandra, Onur Varol, and Emre Kiciman . "Distilling the outcomes of personal experiences : A propensity -scored analysis of social media ." Proceedings of the 2017 ACM Conference on Computer Supported Cooperative Work and Social Computing. ACM, 2017.
- Wager, Stefan, and Susan Athey . "Estimation and inference of heterogeneous treatment effects using random forests ." Journal of the American Statistical Association (2017)
- Sharma, Amit, Jake M. Hofman, and Duncan J. Watts. " Estimating the causal impact of recommendation systems from observational data ." Proceedings of the Sixteenth ACM Conference on Economics and Computation . 2015.
- Shalizi, Cosma Rohilla , and Andrew C. Thomas. " Homophily and contagion are generically confounded in observational social network studies ." Sociological methods & research 40.2 (2011): 211-239.
- Kıcıman, E., & Thelin , J. (2018). Answering what if , should i and other expectation exploration queries using causal inference over longitudinal data . In Conference on Design of Experimental Search and Information Retrieval Systems (DESIRES).

Causality in Social Media

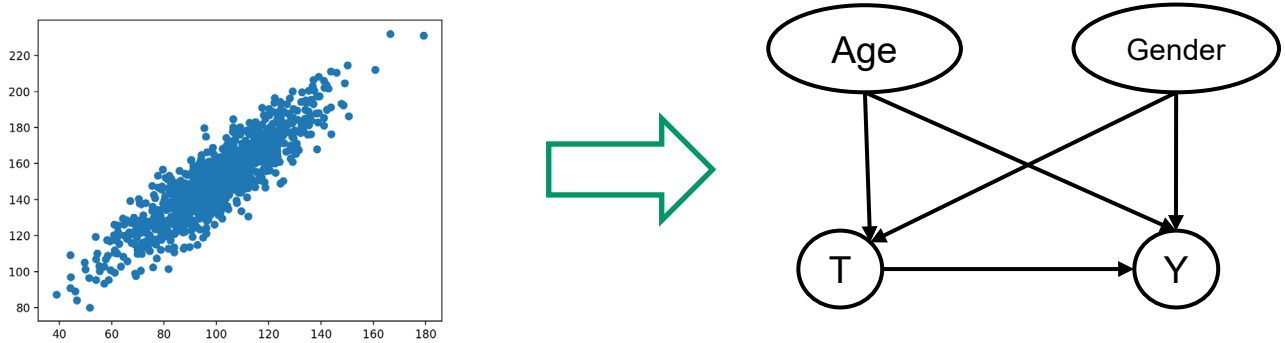
- Weld, G., West, P., Glenski, M., Arbour, D., Rossi, R. A., & Althoff, T. (2022, May). Adjusting for confounders with text: Challenges and an empirical evaluation framework for causal inference. In Proceedings of the International AAAI Conference on Web and Social Media (Vol. 16, pp. 1109-1120).
- Saha, Koustuv, and Amit Sharma. "Causal factors of effective psychosocial outcomes in online mental health communities." In Proceedings of the International AAAI Conference on Web and Social Media, vol. 14, pp. 590-601. 2020.
- Saha, Koustuv, John Torous, Eric D. Caine, and Munmun De Choudhury. "Psychosocial effects of the COVID-19 pandemic: large-scale quasi-experimental study on social media." *Journal of medical internet research* 22, no. 11 (2020): e22600.
- Zhang, Justine, Sendhil Mullainathan, and Cristian Danescu-Niculescu-Mizil. "Quantifying the causal effects of conversational tendencies." Proceedings of the ACM on Human-Computer Interaction 4, no. CSCW2 (2020): 1-24.
- Saha, Koustuv, and Munmun De Choudhury. "Modeling stress with social media around incidents of gun violence on college campuses." Proceedings of the ACM on Human-Computer Interaction 1, no. CSCW (2017): 1-27.

Causality in Social Media

- Saha, Koustuv, Ingmar Weber, and Munmun De Choudhury. "A social media based examination of the effects of counseling recommendations after student deaths on college campuses." In Proceedings of the International AAAI Conference on Web and Social Media, vol. 12, no. 1. 2018.
- Gligorić, Kristina, Ryen W. White, Emre Kiciman, Eric Horvitz, Arnaud Chiolero, and Robert West. "Formation of social ties influences food choice: A campus-wide longitudinal study." Proceedings of the ACM on Human-Computer Interaction 5, no. CSCW1 (2021): 1-25.
- Kiciman, Emre, Scott Counts, and Melissa Gasser. "Using longitudinal social media analysis to understand the effects of early college alcohol use." In Twelfth International AAAI Conference on Web and Social Media. 2018.
- De Choudhury, Munmun, and Emre Kiciman. "The language of social support in social media and its effect on suicidal ideation risk." In Eleventh International AAAI Conference on Web and Social Media. 2017.
- De Choudhury, Munmun, Emre Kiciman, Mark Dredze, Glen Coppersmith, and Mrinal Kumar. "Discovering shifts to suicidal ideation from mental health content in social media." In Proceedings of the 2016 CHI conference on human factors in computing systems, pp. 2098-2110. 2016.

Causality in Social Media

We will see more often the use of causal methods for social media data analysis



see also our
Survey on causality and trustworthy AI. Badar et. al, under preparation