Estimating the Causal Effect of Early ArXiving on Paper Acceptance Yanai Elazar*, Jiayao Zhang*, David Wadden*, Bo Zhang, Noah A. Smith



Research questions

- Does early arXiving affect likelihood of a paper's acceptance at a conference?
- If so, does this effect vary based on researcher characteristics (e.g., host institution, citation count)?

Dataset

ICLR papers between 2018-2022

Paper characteristics	Author / institute characteristics
Number of tables and figuresPaper topic	Author citation countNumber of non-male authors

Causal effect estimation

Matched sample Each paper in treatment group (A = 1) is matched on C with a paper in control group (A = 0).

Estimation target
$$ATET = \mathbb{E}[Y_{A=1} - Y_{A=0} \mid A=1]$$

Difference-in-differences

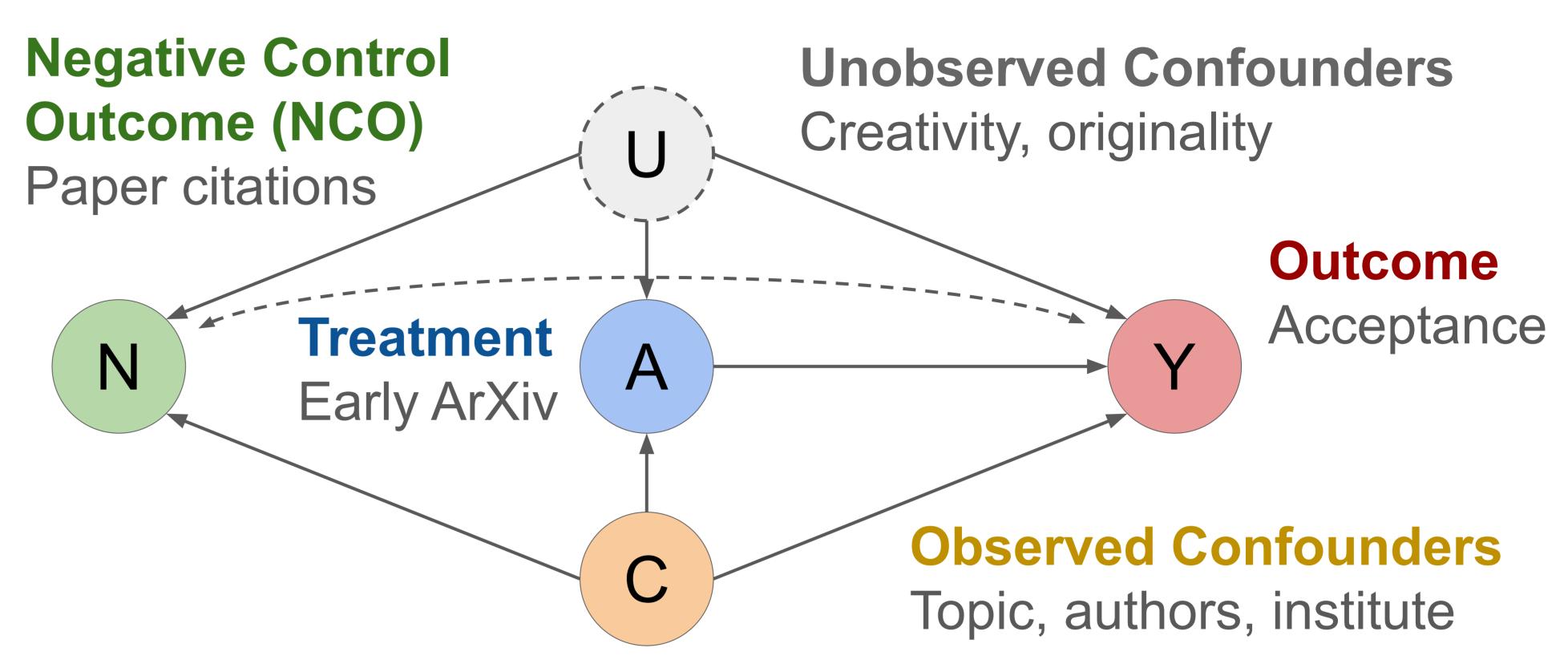
$$ATET = \mathbb{E}[Y_{A=1} - N_{A=1}] - \mathbb{E}[Y_{A=0} - N_{A=0}]$$

Expectation via logistic regression

$$\widehat{ATET} = \sum_{i:A_i=1} \left[\sigmaig(w_{m{Y}}^T c_iig) - \sigmaig(w_N^T c_iig)
ight] - \sum_{i:A_i=0} \left[\sigmaig(w_{m{Y}}^T c_iig) - \sigmaig(w_N^T c_iig)
ight]$$

Unadjusted estimate
$$\widetilde{ATET} = \sum_{i:A_i=1} \left[\sigmaig(w_{m{Y}}^T c_iig)
ight] - \sum_{i:A_{i=0}} \left[\sigmaig(w_{m{Y}}^T c_iig)
ight]$$

Causal Graph



$$N = egin{cases} 1 & ext{if citations above 90th quantile} \\ 0 & ext{otherwise} \end{cases}$$

Results

