Causal inference, from experimental and observational studies, is critical to answering important questions in natural, social and digital systems. Unfortunately, applying causal inference to large systems—such as markets, societies or even teams of people—presents critical challenges in causal inference due to network effects, feedback loops and other complications. While many causal methods have been introduced and are applicable to some of these problems, their use requires careful thought and adaptation by experts. But what if we could identify a (large) class of important questions that could be answered without repeated expert intervention? We identify such a broad class of simple questions about individual experiences—essentially, what happens after a person takes some action or has some experience—that can be answered through analysis of a large-scale corpus of individual-level social media timelines under ignorability and SUTVA assumptions. Our goal is to create a framework for data processing and causal inference methods that can best answer these action-outcome questions from social media timelines.

Providing answers for this class of causal questions using our simplified causal framework is of interest to individuals, scientists and policy makers. For example, individuals may look for ways to better understand the consequences of their decisions. Using our framework, they can effectively aggregate the experiences of hundreds of millions of people, many of whom have made similar decisions and reported on their experiences. The inferred causal outcomes can help people make better decisions, from selecting better products to making better career and life decisions. For scientists and policy makers, understanding various situations and their possible implications of taking actions provides an opportunity to better understand phenomena of social importance, e.g., bullying, planning for retirement, college graduation and unemployment, among many others. Advantages in using social media data for this purpose are as follows. First, results are grounded based on the real experiences of people who have taken an action which increases the reliability of the results. Second, while some goals are common and there are many web articles and advice about them, using social media platform increase the chance of finding an answer. And third, given the preponderance of data, we may provide personalized answers tailored to the asker.

We define our causal framework as follows: let $T$ be the set of experiences (i.e., treatments) we wish to consider and $X$ the set of users. Each user $x$ is characterized by a vector of covariates (e.g., textual features extracted from their posts) $x \in \mathbb{R}^n$. We are interested in the case of binary set of experiences, i.e., $T = \{0, 1\}$, e.g., $T = 1$ is taking a certain medicine. Therefore, for any given $T$, we find all social media users who have reported in their posts that they had that experience (i.e., took the medicine); The set of users who have the experience i.e., $T = 1$ is often known as "treated" group in causal inference literature and the set of users who do not have the experience, i.e., $T = 0$, is known as the "control" group. Let $Y$ be the set of possible outcomes. We consider both $X$ and $Y$ to be represented by textual features. To identify $Y$, we analyze all posts come after the one containing $T$ from the time-lines of the treated group. To represent $X$ and $Y$ of the control group, we randomly select a time-stamp $ts$ from the set of timestamps of the treatment posts and then split the time-lines of the control group to posts come before and after the selected time-stamp $ts$. Any given $y \in Y$ is either $y_1$ or $y_0$, where $y_1$ means a textual feature $y$ occurs and $y_0$ means a textual feature $y$ does not occur. Note that we can only observe one of the outcomes for each user $x$. There is a large literature on various approaches to deal with estimating average causal effects (i.e., $ATE_{x \sim D(X)} = E[Y_1(x) - Y_0(x)]$). Part of this work which employs propensity score analysis has been published in [1].

To identify the best inference methods for our scenario, we empirically evaluated several standard approaches, using both naturalistic data as well as synthetic and mixed naturalistic and synthetic datasets. We investigate four categories of these approaches, namely 1- matching (e.g., propensity score and Mahalanobis), 2- weighting (e.g., inverse propensity score), 3- regression adjustment, and 4- doubly robust methods. We gathered three months Twitter data with more than 69M tweets of
We borrow concepts from the causal inference literature, however it is important to note that our frame-