Social Influence: From Contagion to a Richer Causal Understanding

Dimitra Liotsiou(✉), Luc Moreau, and Susan Halford

University of Southampton, Southampton, UK
{dl1g13,l.moreau}@ecs.soton.ac.uk, susan.halford@soton.ac.uk

Abstract. A central problem in the analysis of observational data is inferring causal relationships - what are the underlying causes of the observed behaviors? With the recent proliferation of Big Data from online social networks, it has become important to determine to what extent social influence causes certain messages to 'go viral', and to what extent other causes also play a role. In this paper, we present a causal framework showing that social influence is confounded with personal similarity, traits of the focal item, and external circumstances. Combined with a set of qualitative considerations on the combination of these sources of causation, we show how this framework can enable investigators to systematically evaluate, strengthen and qualify causal claims about social influence, and we demonstrate its usefulness and versatility by applying it to a variety of common online social datasets.

Keywords: Social influence · Contagion · Causal inference · Graphical causal models · Confounding · Computational social science

1 Introduction: Social Influence and Confounded Causes Behind Observed Actions

Social influence has long been an important research topic in the social sciences. With the emergence of online social network platforms like Facebook and Twitter over the last decade, Big Data from social interactions has been produced at an unprecedented volume and detail, offering scientists new kinds of 'found' observational data through which to examine social processes. This has led to social influence becoming an increasingly prominent topic of study in the field of computer science, as well as to the birth of the interdisciplinary field of computational social science [33] for which methods need to be developed for systematically combining the social and the computational sciences [18,34,48].

Understanding social influence is pivotal since it has been claimed that social influence drives the spread of behaviors and attitudes as diverse as smoking, obesity, happiness, and political participation along social ties, in a process analogous to the contagious spread of viruses [2,5,17,29,31,36], to the extent that ensuring a select few trend-setting individuals (the so-called 'influentials') adopt a behavior would suffice to lead a large population to follow their example and

E. Spiro and Y.-Y. Ahn (Eds.): SocInfo 2016, Part II, LNCS 10047, pp. 116–132, 2016.
DOI: 10.1007/978-3-319-47874-6_9

also adopt this behavior. If social influence does operate in this manner, then harnessing its power would bring immense benefits to marketing, public policy, and public health interventions.

This type of contagion-based paradigm for social influence has been extensively applied to theoretical and observational studies of online social networks like Twitter and Flickr [3,6,8,25,26]. Here, if user j's social connection i mentions the same entity as them (e.g. a URL or a hashtag), within a narrow time window, or if i re-shares or up-votes j's post, or chooses to follow j, or mentions j's username [14,24], then i's action is assumed to be due to social influence from j. One may say that such measures of online activity represent the levels of attention or interest that a given piece of content has generated [1,49]. However, beyond indicating some degree of attention, it is far from straightforward to infer the *meaning* or the *causes* behind such measures of observed actions, and indeed [3,6] recognize that this approach yields an overestimate of social influence. Moreover, it has been acknowledged that the ideal way to make causal claims in empirical settings is to use controlled experiments, but this can often be difficult or infeasible in practice [3,43,45,47].

The difficulty in estimating the extent of social influence from non experimental, observational data is that social influence is only one of many possible causes behind a pair of observed actions. Rather than social influence, there may be other unobserved common causes (called *confounders* [38]) behind two observed actions. These other, often unobserved, causes are commonly grouped into the classes of: similarity of personal traits, intrinsic properties of the focal item, and external circumstances [3,10,15,43,44]. The *focal item* might be a message, behavior, action, or some other item involved in the observable actions (*outcomes*) that the investigators want to study. Observationally determining that a cause of a given action is social influence rather than any one of the other causes, or a mix of many of these causes, is known to be a very difficult problem [3,4,6,43–45].

Therefore, we focus on the questions of why does a person (or a group of people) take a given observed action -what are the underlying causes and the mechanism that determine whether this person (or group) takes this action? If one were to intervene upon a causal factor, e.g. recruiting an 'influential' to endorse a product or healthy behavior on social media [6], what might be the reaction of people exposed to this? These are questions typical of *causal* inference [47].[1]

In this paper, we propose a causal framework for social influence, expanding on [43], and use it to show that social influence is confounded with causes related to personal similarity, traits of the focal item, and external circumstances. We then describe how this framework enables an investigator to systematically evaluate, improve and qualify causal claims on social influence versus each of the other types of possible causes, focusing on observational ('found') data from online social settings. This framework merges computational methods with causal assumptions rooted in social science findings, offering a promising way

[1] As opposed to inference based on *statistical* prediction methods [9,20–23,28,47], which have been used elsewhere in the literature (e.g. [11,16,40]).

to address the need for interdisciplinary common methodological ground in the nascent field of computational social science [18,33,34,48]. We limit our focus here to building this theoretical framework, and to performing an initial evaluation using previous studies of online datasets. A full empirical application to, and validation of the framework on, an online dataset that can adequately capture the confounding causes (typically left at least partly unobserved in online social datasets) is in our future work plans.

The rest of this paper is structured as follows: We first present the necessary background on social influence and the other three classes of possible causes. We then describe our methodology, which is based on graphical causal models, and in the following section apply it to the context of social influence, and show how graphical causal models both make the causal confounding visible and indicate how it can be removed to yield an unbiased estimate of social influence. Following this, we discuss some important qualitative and meaning-related aspects of social influence. We then demonstrate and evaluate how applying our causal framework to well known online social interaction settings enables one to assess the adequacy of the datasets and methods used, and to strengthen one's causal claims. We finally discuss possible directions for future work and present concluding remarks.

2 Social Influence and Other Classes of Causes

This section lays out the necessary background on social influence and the other possible causes behind observed actions, namely similarity of personal traits, intrinsic traits of the focal item and external circumstances. In all cases, we note that each factor may cause two people to exhibit the same observed behavior regardless of whether there is a social tie between them or not [10,30].

Social Influence. Social influence can be defined as the phenomenon where a person's behavior (action, opinion, or belief) is caused by another person's observed behavior: a person i may perform an action that person j performed earlier because j's performing of the action was so inspirational, persuasive, or impressive (e.g. j making a persuasive argument based on domain expertise) that i was convinced or became inclined to also perform it [30,43]. We only consider cases where i has free choice, i.e. j cannot force i to perform the given action. For instance, [35] defines influence as a form of causation, occurring in a possibly covert, unclear, or unintentional way, that does not involve force or coercion. Similarly in [39], a seminal work from the communications literature, the term social influence is used in the sense of a person causing another person to change their behavior, through the use of appropriate incentives.

Similarity of Personal Traits. Two people i and j may *independently* adopt the same behavior because they share one or more personal traits, such as interests, values, beliefs, opinions, needs, desires, personality profile, or demographic characteristics, like age, race, gender, social class [5,7,46]. For instance, two people

may each independently post about political news on Twitter, because they each have an active interest in politics.

Intrinsic Properties of the Focal Item. In the social psychology, management, and marketing literature [10,32], it has been established that certain features can be 'engineered into' a *focal item*(e.g. a message, a product) that entice people to reshare it with others, making it 'go viral' and potentially increasing sales or adoption rates. An important type among them is features that invoke emotional arousal, specifically *activating* emotions such as excitement or anger, as these have been found to increase the chances that the viewer will then reshare, discuss or even adopt this message, behaviour or product. Hence, investigators should account for such relevant features, as well as other more general features (e.g. the price of a product; the effort or risk associated with a behavior [13]) that play a causal role in a person's reaction to a focal item.

External Circumstances. External circumstances may be the common cause why two people may *independently* take the same action, e.g. users i and j may post the same video or URL on social media because it relates to an important current news item, or a popular trend, that they both are aware of. External circumstances encompass factors from the external environment (e.g. a news item, a trend or a currently popular belief or attitude, a new law, a natural disaster), outside the personal traits of person i and j, and outside the traits of the focal item.

3 Methodology: Graphical Causal Models

It is an often-repeated cautionary phrase in statistics that 'correlation does not imply causation.' The field of causality theory, which saw rapid developments in the last thirty years, allows one to go beyond correlations and reason about causation in a rigorous, formal way, using tools like graphical causal models, which in turn are based on directed graphs and probability theory [38]. In this paper we will be using graphical causal models to reason about social influence versus the other possible causes of observed actions, expanding upon the work presented in [43]. We present the relevant theory here, based on [37,38,42].

A graphical causal model can be represented as a directed acyclic graph G, comprised of a set of nodes, N, and a set of directed edges, or arrows, E - that is, $G = \{N, E\}$. Nodes represent variables, and edges denote causal relationships. A directed edge from a node A to a node B denotes the *direct causal effect* of A on B, where A is a cause of B; A is called a *parent* or *ancestor* of B, and B a *child* or *descendant* of A. If a node has no arrow pointing to it, i.e. no parents, it is called *exogenous*, otherwise it is called *endogenous*. A *path* is a sequence of consecutive edges that do not all necessarily point in the same direction. Which nodes are connected to which depends on the modeller's causal assumptions, which should be well-justified and grounded in domain expertise [38]. The rules for manipulating graphical causal models then show what causal inferences can be made from these causal assumptions.

Graphical causal models are particularly useful for identifying latent (unobserved) variables that introduce *confounding bias* to the estimate of the causal effect of a variable X on another variable Y, and for then *adjusting for* those variables to obtain the unbiased causal effect of X on Y.

We illustrate this using the simple example causal model in Fig. 1, whose structure appears in our model of the causal effect of social influence and other factors on observed outcomes, as we shall see. Figure 1 represents a situation where the observed variable X is a cause of the observed variable Y, but variable U is a latent (unobserved) cause of both X and of Y. As causal graphs are governed by the Causal Markov Condition, whereby endogenous variables only depend on their parents [38], the joint probability distribution representing Fig. 1 is: $P(y, x, u) = P(u)P(x|u)P(y|x, u)$, where $P(w)$ is short for $P(W = w)$, since Y's parents are U and X, U is the parent of X, and U has no parents.

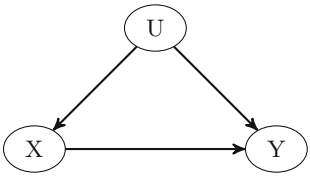


Fig. 1. Example graphical causal model

We want to estimate the causal effect of X on Y, which we write as $P(Y = y|do(X = x))$ in Pearl's do-notation, denoting the distribution of Y which would be generated, counterfactually, if X were set to the particular value x through experimental *manipulation* or *intervention*. In the causal graph this would mean deleting all arrows into X, setting X's value to x, and leaving the rest unchanged. This post-intervention distribution of Y is not in general the same as the ordinary conditional distribution $P(Y = y|X = x)$, as the latter represents taking the original, pre-intervention, population and *selecting* from it only the sub-population where $X = x$. The mechanisms that set X to that value may have also influenced Y through other channels, so the latter distribution would not typically really tell us what would happen if we externally manipulated X.

Figure 1 illustrates this point. If we consider the dependence of Y on X, in the form of the conditional $P(Y = y|X = x)$, we see there are two channels of information flow from cause to effect: one is the direct, causal path from X to Y, represented by $P(Y = y|do(X = x))$. However, there is also another, indirect path, between X and Y through their unobserved common cause U, where observing X gives information about its parent U, and U gives information about its child Y. If we just observe X and Y, we cannot distinguish the causal effect from the indirect inference -the causal effect is *confounded* with the indirect dependence between X and Y created by their common cause U. More generally, the effect of X on Y is confounded whenever $P(Y = y|do(X = x)) \neq P(Y = y|X = x)$. If

there is a way to write $P(Y = y|do(X = x))$ in terms of distributions of observables, then we say that the confounding can be removed by an *identification, or deconfoundng, strategy*, which renders the causal effect *identifiable*.

Formally, to test whether there is confounding, we must first test whether some variables "*block*" (stop the flow of information or dependency along) all paths from X to Y, using the so-called *d-separation criterion* (as per [38]): A set of nodes Z *block* or *d-separate* a path p if and only if (i) p contains a *chain* $i \rightarrow m \rightarrow j$ or a *fork* $i \leftarrow m \rightarrow j$ such that the middle node m is in Z, or (ii) p contains a *collider* $i \rightarrow m \leftarrow j$ such that neither the middle node m, nor any of its descendants, are in Z. Then, a set Z d-separates X from Y if and only if Z blocks *every* path from X to Y. Further, a set of variables Z satisfies the *back-door criterion* (as per [38]) relative to X and Y if: (i) no node in Z is a descendant of X, and (ii) Z blocks every path between X and Y that contains an arrow into X. Then the set Z is called a *sufficient, admissible* or *deconfounding* set. Finding this deconfounding set permits the confounding bias to be removed, thus rendering the causal effect X on Y identifiable from non-experimental data, using the *back-door adjustment* formula [37,38]:

$$P(Y = y|do(X = x)) = \sum\nolimits_z P(Y = y|X = x, Z = z)P(Z = z) \quad (1)$$

Since the right-hand side of Eq. 1 contains only probabilities which are estimable (e.g. by regression) from our observational, non-experimental data, the causal effect of X on Y can be estimated from such data without bias.

In the example of Fig. 1, we see that variable U satisfies the back-door criterion, and hence, to obtain the direct causal effect of X on Y, one should simultaneously measure X, Y and U for every member of the randomly-selected sample under study, and then obtain the causal effect by using the back-door adjustment formula (Eq. 1) for $Z = \{U\}$.

In summary, to remove confounding and obtain the unbiased causal effect of X on Y, our *deconfounding strategy* is: (1) select a large random sample from the population of interest, (2) for every individual in the sample, measure X, Y, and all variables in Z, and (3) adjust for Z by partitioning the sample into groups that are homogeneous relative to Z, assess the effect of X on Y in each homogeneous group, and then average the results, as per Eq. 1.

4 Application to Social Influence: Confounding with Other Possible Causes

In this section, we use graphical causal models to reason about possible confounding of social influence with other causes, when working with observational data. We begin with the framework presented in [43], which we then simplify slightly without affecting its results with respect to confounding. We then adjust this framework such that it can model confounding even in the absence of a social tie. We next construct similar causal frameworks which show how social influence is confounded with personal traits, with intrinsic traits of the focal item,

and with external circumstances. Finally, we put these separate models together into a single graphical causal model which shows the causal relations between causes and outcomes, and makes visible which variables should be measured and adjusted for to remove confounding.

4.1 Social Influence Is Confounded with Homophily

We begin by presenting the graphical causal model used in [43], which demonstrated that the phenomena of homophily (the tendency of people to form social ties with people similar to them) and of behaviour adoption due to social influence from friends are confounded in observational social network data. We follow their notation for continuity: Symbols X_k and Z_k denote sets of random variables representing, respectively, the unobserved and observed personal traits of person k. Each of those may be discrete or continuous, and is assumed to remain constant during the time period studied. $A_{k,l}$ is an observed variable, for simplicity in this case assumed to by binary, with value 1 if person k considers person l to be a 'friend', and with value 0 otherwise. $Y_{k,t}$ is an observed response variable, denoting whether person k performs action Y at a time t, and may be discrete or continuous. For simplicity, we assume time progresses in discrete steps (although this is not essential, as stated in [43]). It is also assumed that there is *latent* homophily in this system, hence whether two people are friends, i.e. whether $A_{i,j} = 1$, depends causally on their latent personality traits X_i and X_j. The model is shown in Fig. 2a.

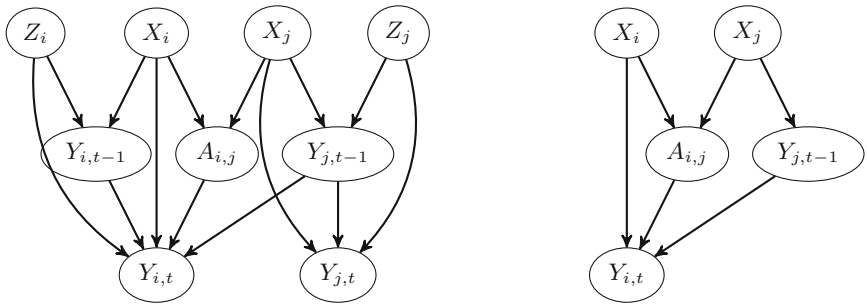
We are interested in estimating social influence, i.e. the *direct causal effect* of person j's performing of action Y, $Y_{j,t-1}$, on person i's subsequent performing of the same action, $Y_{i,t}$, represented by the arrow $Y_{j,t-1} \rightarrow Y_{i,t}$: person i performs action Y because person j's example inspired them to do the same.[2] Homophily introduces a backdoor path between $Y_{i,t}$ and $Y_{i,t-1}$ through the latent X_i and X_j: $Y_{i,t} \leftarrow X_i \rightarrow A_{i,j} \leftarrow X_j \rightarrow Y_{j,t-1}$, i.e. the latent X_i and X_j are in the deconfounding set, thus social influence (the direct causal effect of $Y_{j,t-1}$ on $Y_{i,t}$) is confounded with homophily. So X_i and X_j should be measured and adjusted for, to retrieve the pure causal effect of $Y_{j,t-1}$ on $Y_{i,t}$.

Before we move on to apply this type of modeling to show how influence is confounded with other causes, we first simplify the model for ease of examination of paths and of manipulation. As [43] say, the assumption that $Y_{i,t-1}$ has a direct causal effect on $Y_{i,t}$ can be dropped without affecting the results of the investigation. Therefore, we remove $Y_{i,t-1}$, and, similarly for j, we remove $Y_{j,t}$. Since we are interested in examining the causes behind why i did Y at time t, $Y_{j,t}$ is not relevant.[3] In addition, since the observed personal traits Z_i and Z_j do

[2] In [43], it is assumed that one can be directly socially influenced only by those people she considers her 'friends' ($A_{i,j} = 1$), and not by anyone else.

[3] We note that $Y_{i,t-1}$ might represent a plausible and relevant kind of cause, e.g. that i does Y at time t because i did Y at $t-1$ and was happy with the results, or out of habit from having done it previously at time $t-1$. However, this previous happiness or habit may best be included in X_i as a variable representing an interest in Y.

not play a role in either introducing or removing confounding in this model or in our next models, we also remove those, and assume that all personality traits are unobserved, hence represented by X_i and X_j - indeed, usually there is no, or insufficient, data on users' personal traits in observational online social network studies. This simplification yields the model in Fig. 2b.



(a) Causal model from [43] (b) Simplified version

Fig. 2. Graphical causal model from [43] (a), and simplified version (b)

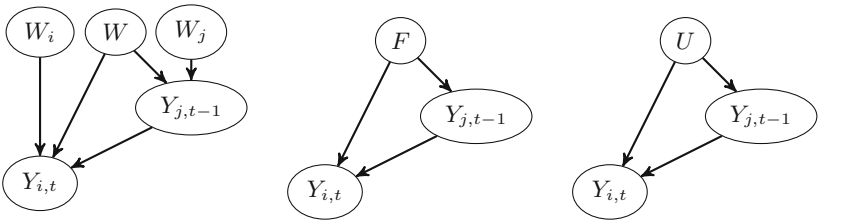
4.2 Social Influence Is Confounded with Similarity in Personality Traits, Focal Item Traits, and External Circumstances

In this section, we present the graphical causal models that show how social influence is confounded with each of the following types of causes: similarity in personality traits, focal item traits, and external circumstances. We note that all confounding cases are due to structurally equivalent back-door paths of the form presented in Fig. 1 - each could essentially be regarded as a common cause: person-internal (personal traits), item-internal, or external.

Confounding with Similarity in Personality Traits. To show how a shared personality trait may be a cause behind i and j independently performing the same action Y, we now replace the previous latent personal trait variables X_i and X_j with W, representing the latent shared traits between i and j (i.e. W is the intersection of sets X_i and X_j), and W_i, i's remaining latent traits that j does not share, and respectively W_j for j's latent traits that i does not share. This produces the model of Fig. 3a, which shows that $Z = \{W\}$ is the deconfounding set on which to perform back-door adjustment.

Confounding with Traits of Focal Item. Similarly to Fig. 3a, b shows that variable F, representing the focal item traits, lies on a backdoor path $Y_{i,t} \leftarrow F \rightarrow Y_{j,t-1}$. Hence, the deconfounding set to be back-door adjusted is $Z = \{F\}$.

Confounding with External Circumstances. Similarly to Fig. 3b, in Fig. 3c variable U represents the external common cause (e.g. a shocking news item), and the back-door path $Y_{i,t} \leftarrow U \rightarrow Y_{j,t-1}$ introduces confounding. Hence $Z = \{U\}$ is the deconfounding set that should be back-door adjusted.



(a) Common personality traits (b) Focal item traits (c) External common cause

Fig. 3. Graphical causal models for social influence versus similarity in personality traits (a), focal item traits (b), and external circumstances (c)

4.3 Putting It All Together: Social Influence, Personal Similarity, Focal Item Traits, External Circumstances

We now put together all the above graphical causal models, to show the full picture of all causes that affect person i's decision to perform action Y at time t, and how these, if left unobserved and unadjusted for, introduce confounding bias into our estimate of social influence from person j's action Y at time $t-1$.

Keeping the same notation, we present two models, one without a social tie variable $A_{i,j}$ in Fig. 4a, and one with that tie in Fig. 4b. Given our split of personal traits into those that both people have in common (W) and those they do not (W_i, W_j), we assume that the decision to consider someone a 'friend' depends on having enough things in common (W), and also on not having too many differences in personality (e.g. to the extent that one cannot tolerate or is offended by the other's value system) - hence, besides W, we assume that W_i and W_j also causally affect whether a social tie will be fostered.

Therefore, the minimal deconfounding set for Fig. 4a is $Z = \{F, U, W\}$, and for Fig. 4b it is $Z' = \{F, U, W, W_j\}$[4]. Therefore, in order to retrieve the pure direct causal effect of $Y_{j,t-1}$ on $Y_{i,t}$, an investigator must implement our deconfounding strategy - crucially, all variables in the appropriate minimal deconfounding set must be measured for *every* individual in the random sample, and adjusted for as per Eq. 1.

We see that this full model presents a complex picture, with many factors playing a role in i's decision to take action Y. Indeed, as we shall discuss in the following two sections, it is known in the social sciences that social influence alone is seldom enough to ensure $Y_{i,t}$ - rather, a specific combination of social influence and all the other causal factors is needed.

[4] W_i could be in Z' but it is redundant, due to the assumed asymmetry of A_{ij}; if there was an edge $A_{ij} \rightarrow Y_{j,t-1}$ then W_i would have to be in the minimal confounding set.

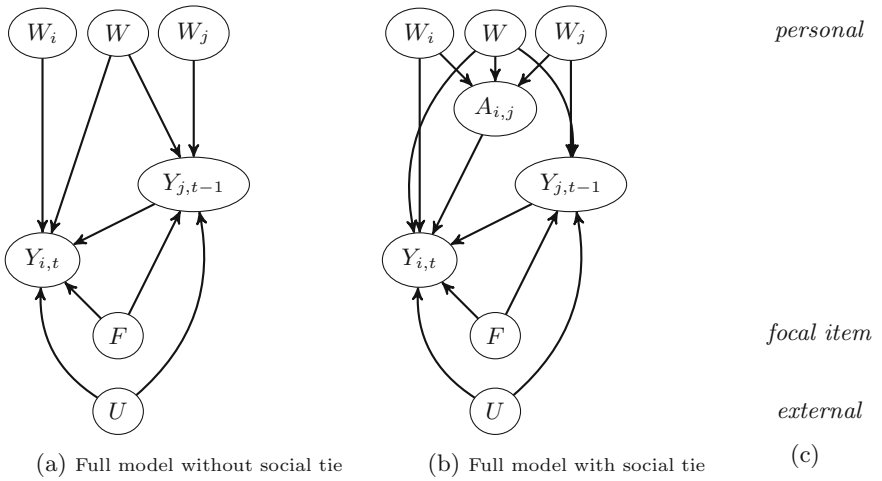


Fig. 4. Full graphical causal models for social influence versus other causes, without social ties (a), and with social ties (b), with the legend (c) on the right showing the context of each latent causal variable

5 The Impact of Causal Factor Characteristics on the Nature of Observed Outcomes

In this section, we aim to shed some further light on the question of what kinds of causal circumstances are needed for a person or group to take a given action. In the empirical and the theoretical literature [10,29,30,46,50] it has been widely acknowledged that no person is a clean slate, and no situation is 'neutral', therefore social influence does not operate in a vacuum, and on its own is rarely sufficient to ensure one or more individuals i take a specific action or commit to a new behavior ($Y_{i,t}$) (e.g. making some online content 'go viral', or a product sell out): a single well-connected person j alone is not enough to reliably influence others i to act a certain way; rather, a combination of compatible personal traits (W and W_j), a focal item with appropriate features F, and beneficial external conditions U are also needed.

Therefore, we next examine some important qualitative aspects of how different combinations of causal factors may lead to different qualities in the final observed outcomes. These qualitative aspects affect the extent and nature of claims one can make about social influence, and hence should be measured, e.g. by recording more details of the decision-making process than is common in observational social network datasets, or (e.g. to avoid making the process intrusive for participants) through interviews, or through a combination of methods.

Magnitude, Direction, and Duration. Instead of modeling the outcome $Y_{i,t}$ as binary, it could instead have a magnitude, duration, and direction. The magnitude would represent the intensity of i's engagement with Y from time t onwards,

whether this engagement is only superficial (small magnitude) or serious and incorporated into their value system (large magnitude), while duration would capture how short-lived or long-lasting this is [12,30]. The direction would capture whether i does the same as j with respect to Y (positive direction), or the opposite (negative), e.g. because j's way of engagement with Y was against i's values, or whether i does not take substantive action in relation to Y, e.g. out of loss of interest [2]. For instance, Facebook's addition of specific reaction buttons for love, anger, etc. to the Like button (which was previously used to express any type of reaction) [27], is one approach to capturing direction.

Normative versus Informational Social Influence. A person may change their behavior or take an action not because they find the traits of that behavior or action (F) inherently worthwhile, but rather because they want to please or feel accepted by someone they know (j, $A_{i,j}$) or by a wider social group (U). In [19], the former type is termed *informational influence*, and the latter *normative influence*, as discussed in [30]. Which type of social influence occurs in a given case depends on all the causal variables.

Generalizability of Observed Outcomes. Often, investigators use observational social media data capturing the levels of online interest in a product of behavior as proxies for estimating a different outcome like product sales or adoption of that behavior. However, it has been shown that the levels of interest on social media may not translate to actual purchases or behavior change [12] (e.g. the case of the popular Evian advert that did not increase sales, in [10]). That is because the causal factors in the two cases are very different: in the latter case, factors that do not apply in online discussions, like for instance the price, qualities, effort and/or risk associated with this product or behavior, F, and society's views of adopting it, U, come into play. Therefore, when using data from online social networks as proxies, the underlying causal factors should be adequately similar.

Changing Deep Rooted Behaviors: Identity, Effort and Risk. It has been claimed that social influence drives behaviors as diverse as sharing a message with friends, purchasing decisions, smoking habits, and happiness levels [10,17]. However, some behaviors (e.g. quitting or restarting smoking, or becoming happier) are much more deeply rooted in a person's identity, psychology or worldview (X_i plays a stronger role), are more difficult to change (F), and carry more risk in terms of social acceptance (U) [10,13], than other actions (e.g. re-sharing some information on social media, or choosing which brand of bottled water to buy).

6 Evaluation

In this section, we demonstrate how our causal framework and qualitative considerations might help investigators position their findings within the full causal picture for social influence, assess the extent and types of causal claims on influence their data allows them to make, and determine what causal variables should

next be measured and adjusted for in order to make more robust causal claims. We examine examples of studies that actively try to capture causal effects of influence by reducing the effects of confounders, using quantitative and/or qualitative methodologies, in research settings involving one or more of the disciplines of sociology, social psychology, marketing, and computer science. We use our framework to examine how these studies lay out potential avenues, as well as expose caveats, for future attempts at measuring and adjusting for confounders and at capturing the qualitative aspects of social influence processes.

In [5], a controlled experiment on Facebook is performed, with the focal item being a Facebook app about films. It is randomized which friends i of j see messages $Y_{j,t-1}$ declaring j's use of this app, aiming to measure social influence versus susceptibility (i's tendencies to adapt to $Y_{j,t-1}$ by also downloading the focal item). It is assumed that randomly choosing the subjects i who will be exposed to $Y_{j,t-1}$ will suffice to control for homophily (similarity W among friends i and j linked through $A_{i,j}$) and for exposure to common external causes (U). Hence, it is assumed that whenever an exposed person i also downloads the app ($Y_{i,t}$) the only cause must be social influence ($Y_{j,t-1} \rightarrow Y_{i,t}$). However, we note that since the alternative causes have not been measured, they may continue to introduce confounding, despite the random selection - for instance, it might have been that all people who also downloaded the app did so because they themselves had an interest (W) in films, and all the people who did not download it did so because they had no interest in films. Therefore, the cause might rather have been a common personal trait W - we cannot know whether the cause was social influence or another cause, until we have measured and adjusted for the confounders for every person i in the sample.

Taking steps to observationally measure personality traits for each participant, [4] use an observational dataset containing many personal traits (X_i, X_j) for each pair of users, in an attempt to disentangle homophily from social influence. Still, as explained in [43], due to the methods used, there may still remain some latent personal similarity (W) which affects behavior adoption ($Y_{i,t}$). Moreover, we note that the confounders relating to the focal item traits (F), and to external common cause (U) remain latent. Still, this study shows a way to observationally measure X_i and X_j to some extent.

In an online randomized experiment, [41] manage to measure some confounders and obtain a relatively close estimate of the causal effect of aggregate social influence on users' choices of whether to download a song (focal item). It is randomized which users i are exposed to aggregate social influence (total number of downloads a song has received, $\sum_j Y_{j,t-1}$, where the identities of users j are not displayed). To reduce the effect of external common cause U, special care is taken (including conducting surveys) to ensure the displayed songs and artists are virtually unknown. The songs are kept the same (F constant) while some participant groups see the number of downloads for each song and other groups do not. However, as W has not been measured, and neither has F (e.g. song genre), a small possibility remains that the same song might have been dowloaded more in a social influence group than in a neutral one not because of

social influence (from the displayed download count), but rather because that group contained more participants who were fans (W) of that song's genre (F). Therefore, some confounding due to latent W and F might remain, so these should be measured and adjusted for. Still, this study offers a good example of a significant and detailed effort to reduce U while experimentally controlling F.

In [45], observational data is used to study the causal effect of Amazon recommendations of the form 'Customers who bought this [product A] also bought [that product B]' on the views of product B (the focal item). Again, i cannot see the identities of customers j who bought both products. The investigators attempt to control for F to an extent, by studying many different product categories, and try to ensure that external causes U are held constant as much as possible. They also investigate the effect of the type of users i they have studied (X_i) on the causal effect of the recommendation. In qualitative terms, they recognize that a user's clicking on a recommendation might be due to convenience rather than the persuasive qualities of this particular recommendation. Overall, they caution that their results are still an upper bound for the causal effect of social influence, but a stricter one than under naive assumptions, and acknowledge that their results may not readily generalize to the average Amazon user, or to all Amazon product categories, or to other recommendation settings.

An example of how qualities of outcomes can be measured at a fine granularity and over time is presented in [2]. Here, the social influence from one participants' emotional state on another's (effect of $Y_{j,t-1}$ on $Y_{i,t}$), in the setting of face-to-face offline interactions, is measured using a mixed methodology of infrared sociometric sensors (badges) and questionnaires. The authors measure here many 'directions' of outcomes: not just mimetic (termed 'attraction'), but also neutral or negative (termed 'inertia, repulsion and push') at three points per day. They also measure participants' fixed personality traits X_i and X_j, but do not measure other confounders, and are careful to clarify that their social influence claims are correlational, not causal.

The offline controlled experiments in [30] offer useful examples of how to design experiments, control for some confounders, and use varied types of questionnaires, and how to measure the ways in which the combination of causal circumstances ($U, F, Y_{j,t-1}$) affect the nature of the resulting outcome $Y_{i,t}$. Here, the goal is to empirically evaluate how different combinations of causal circumstances (particularly $Y_{j,t-1}, U$) lead to different types of outcomes (termed 'compliance, identification and internalization'). Still, the broader external environment U (e.g. popular attitudes relevant to the topic of the focal message) and the participants' personal views (W) remain unmeasured and so may introduce confounding. Experimentally, the core of the argument (F) is kept the same, but the way it is framed ($Y_{j,t-1}$) is varied. To measure the 'magnitude' of the outcome, i.e. extent to which it was internalized and incorporated into i's worldview and value system, and its duration, questionnaires are used which include open-ended questions, both soon after exposure to $Y_{j,t-1}$ and some weeks after.

In summary, we have demonstrated how our causal framework and qualitative considerations can be used to help one position, assess and improve the

claims they can make on social influence by ensuring they measure all relevant confounders as much as possible and adjust for them. To demonstrate how this might be achieved in practice, we have assessed the merits of practical attempts at reducing confounding and at accounting for qualitative aspects, both in observational and experimental settings, whether online or in mixed online-offline setups, covering quantitative and qualitative methods.

7 Conclusion and Future Work

Overall, we have proposed a methodological framework for assessing the causal effect of social influence, covering the space of other types of causes that may lead to an observed action (outcome), namely similarity of personal traits, traits of the focal item, and external circumstances. We have shown that social influence is confounded with each of these types of causes, using the formal rules of graphical causal models and based on robust causal assumptions about what types of causes might directly affect one's actions, which stem from well-established results from the social sciences literature. In merging computational rules with social science-based causal assumptions, this framework offers a promising interdisciplinary methodology of the type that is much-needed in computational social science. Drawing from social and computational disciplines, we then presented some important characteristics of the observed outcomes and the causal variables, which affect the nature, form and extent of the claims one can make on social influence. We then demonstrated how our causal framework and qualitative considerations may be applied in practice, by using them to evaluate the robustness of social influence estimates (how much confounding has been successfully adjusted for, how much still remains, and what qualitative aspects have been examined) from a set of diverse social influence studies from the social science and computer science literature that employed a varied range of practical methods.

As discussed, typical online social datasets do not adequately capture all relevant confounding causes. So, in future work, in order to make robust causal claims about social influence, we plan to apply our proposed framework to our own online dataset, taking care to obtain data that is detailed enough in capturing all relevant causes as much as possible. Further, it would be worth investigating how to harness social science expertise to devise systematic methods for identifying which specific causal variables for each type of cause are relevant in a given setting and should be measured, and how this may vary across different settings. Moreover, since the observed outcome (whose causes we aim to estimate) reflects a possibly subjective decision made by a specific person, we note that this person's choice and interpretation of relevant causes might differ from the investigator's, so it may be worth accounting for this potential difference using social science expertise (e.g. from social psychology).

References

1. Ackland, R.: Web social science: Concepts, data and tools for social scientists in the digital age. Sage, London (2013)
2. Alshamsi, A., Pianesi, F., Lepri, B., Pentland, A., Rahwan, I.: Beyond contagion: Reality mining reveals complex patterns of social influence. PloS One **10**(8), e0135740 (2015)
3. Anagnostopoulos, A., Kumar, R., Mahdian, M.: Influence and correlation in social networks. In: Proceedings of the 14th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, pp. 7–15. ACM (2008)
4. Aral, S., Muchnik, L., Sundararajan, A.: Distinguishing influence-based contagion from homophily-driven diffusion in dynamic networks. Proc. Nat. Acad. Sci. **106**(51), 21544–21549 (2009)
5. Aral, S., Walker, D.: Identifying influential and susceptible members of social networks. Science **337**(6092), 337–341 (2012)
6. Bakshy, E., Hofman, J.M., Mason, W.A., Watts, D.J.: Everyone's an influencer: quantifying influence on twitter. In: Proceedings of the fourth ACM international conference on Web search and data mining. pp. 65–74. ACM (2011)
7. Bakshy, E., Rosenn, I., Marlow, C., Adamic, L.: The role of social networks in information diffusion. In: Proceedings of the 21st International Conference on World Wide Web, pp. 519–528. ACM (2012)
8. Barbieri, N., Bonchi, F., Manco, G.: Influence-based network-oblivious community detection. In: 2013 IEEE 13th International Conference on Data Mining (ICDM), pp. 955–960. IEEE (2013)
9. Barnett, L., Barrett, A.B., Seth, A.K.: Granger causality and transfer entropy are equivalent for gaussian variables. Phys. Rev. Lett. **103**(23), 238701 (2009)
10. Berger, J.: Contagious: Why Things catch on. Simon and Schuster, New York (2013)
11. Borge-Holthoefer, J., Perra, N., Gonçalves, B., González-Bailón, S., Arenas, A., Moreno, Y., Vespignani, A.: The dynamics of information-driven coordination phenomena: A transfer entropy analysis. Sci. Adv. **2**(4), e1501158 (2016)
12. Cebrian, M., Rahwan, I., Pentland, A.S.: Beyond viral. Commun. ACM **59**(4), 36–39 (2016)
13. Centola, D., Macy, M.: Complex contagions and the weakness of long ties1. Am. J. Soc. **113**(3), 702–734 (2007)
14. Cha, M., Haddadi, H., Benevenuto, F., Gummadi, P.K.: Measuring user influence in twitter: The million follower fallacy. ICWSM **10**, 10–17 (2010)
15. Cheng, J., Adamic, L., Dow, P.A., Kleinberg, J.M., Leskovec, J.: Can cascades be predicted? In: Proceedings of the 23rd International Conference on World Wide Web, pp. 925–936. International World Wide Web Conferences Steering Committee (2014)
16. Chikhaoui, B., Chiazzaro, M., Wang, S.: A new granger causal model for influence evolution in dynamic social networks: The case of dblp. In: Twenty-Ninth AAAI Conference on Artificial Intelligence (2015)
17. Christakis, N.A., Fowler, J.H.: Social contagion theory: examining dynamic social networks and human behavior. Statist. Med. **32**(4), 556–577 (2013)
18. Counts, S., De Choudhury, M., Diesner, J., Gilbert, E., Gonzalez, M., Keegan, B., Naaman, M., Wallach, H.: Computational social science: Cscw in the social media Era. In: Proceedings of the Companion Publication of the 17th ACM Conference on Computer Supported Cooperative Work and Social Computing, pp. 105–108. ACM (2014)

19. Deutsch, M., Gerard, H.B.: A study of normative and informational social influences upon individual judgment. J. Abnorm. Soc. Psychol. **51**(3), 629 (1955)
20. Diebold, F.X.: Elements of forecasting. Citeseer, Ohio (1998)
21. Diebold, F.X.: Forecasting. Department of Economics, University of Pennsylvania (2015). http://www.ssc.upenn.edu/~fdiebold/Textbooks.html
22. Eichler, M.: Graphical modelling of multivariate time series. Probab. Theor. Relat. Fields **153**(1–2), 233–268 (2012)
23. Eichler, M.: Causal inference with multiple time series: principles and problems. Philos. Trans. Royal Soc. London A Math. Phys. Eng. Sci. **371**(1997), 20110613 (2013)
24. Ghosh, R., Lerman, K.: Predicting influential users in online social networks. In: Proceedings of KDD Workshop on Social Network Analysis (SNA-KDD), July 2010
25. González-Bailón, S., Borge-Holthoefer, J., Rivero, A., Moreno, Y.: The dynamics of protest recruitment through an online network. Sci. Rep. **1**, 197 (2011)
26. Goyal, A., Bonchi, F., Lakshmanan, L.V.: Learning influence probabilities in social networks. In: Proceedings of the Third ACM International Conference on Web Search and Data Mining, pp. 241–250. ACM (2010)
27. Greenberg, J.: Advertisers don't like facebook's reactions. They love them. WIRED (2016). http://www.wired.com/2016/02/advertisers-feel-facebooks-new-reactions-%F0%9F%98%8D/
28. Hlaváčková-Schindler, K., Paluš, M., Vejmelka, M., Bhattacharya, J.: Causality detection based on information-theoretic approaches in time series analysis. Phys. Rep. **441**(1), 1–46 (2007)
29. Katz, E., Lazarsfeld, P.F.: Personal Influence, The Part Played by People in the Flow of Mass Communications. The Free Press, New York (1955)
30. Kelman, H.C.: Processes of opinion change. Public Opin. Q. **25**(1), 57–78 (1961)
31. Kempe, David, Kleinberg, Jon, Tardos, Éva: Influential nodes in a diffusion model for social networks. In: Caires, Luís, Italiano, Giuseppe, F., Monteiro, Luís, Palamidessi, Catuscia, Yung, Moti (eds.) ICALP 2005. LNCS, vol. 3580, pp. 1127–1138. Springer, Heidelberg (2005). doi:10.1007/11523468_91
32. Kilduff, M., Chiaburu, D.S., Menges, J.I.: Strategic use of emotional intelligence in organizational settings: Exploring the dark side. Res. Organ. Behav. **30**, 129–152 (2010)
33. Lazer, D., Pentland, A., Adamic, L., Aral, S., Barabsi, A.L., Brewer, D., Christakis, N., Contractor, N., Fowler, J., Gutmann, M., Jebara, T., King, G., Macy, M., Roy, D., Van Alstyne, M.: Computational social science. Science **323**(5915), 721–723 (2009). http://www.sciencemag.org/content/323/5915/721.short
34. Mason, W., Vaughan, J.W., Wallach, H.: Computational social science and social computing. Mach. Learn. **95**(3), 257 (2014)
35. Morriss, P.: Power: A Philosophical Analysis. Manchester University Press, Manchester (1987)
36. Nickerson, D.W.: Is voting contagious? evidence from two field experiments. Am. Polit. Sci. Rev. **102**(01), 49–57 (2008)
37. Pearl, J.: Causal inference in statistics: An overview. Stat. Surv. **3**, 96–146 (2009)
38. Pearl, J.: Causality. Cambridge University Press, Cambridge (2009)
39. Rogers, E.M.: Diffusion of Innovations. Simon and Schuster, New York (2003)
40. Runge, J.: Quantifying information transfer and mediation along causal pathways in complex systems. Phys. Rev. E **92**(6), 62829 (2015)
41. Salganik, M.J., Dodds, P.S., Watts, D.J.: Experimental study of inequality and unpredictability in an artificial cultural market. Science **311**(5762), 854–856 (2006)

42. Shalizi, C.: Advanced Data Analysis from an Elementary Point of View. Cambridge University Press, New York (2013)
43. Shalizi, C.R., Thomas, A.C.: Homophily and contagion are generically confounded in observational social network studies. Sociol. Methods Res. **40**(2), 211–239 (2011)
44. Sharma, A., Cosley, D.: Distinguishing between personal preferences and social influence in online activity feeds. In: Proceedings of the 19th ACM Conference on Computer-Supported Cooperative Work and Social Computing, pp. 1091–1103. CSCW 2016, NY, USA (2016). http://doi.acm.org/10.1145/2818048.2819982
45. Sharma, A., Hofman, J.M., Watts, D.J.: Estimating the causal impact of recommendation systems from observational data. In: Proceedings of the Sixteenth ACM Conference on Economics and Computation, pp. 453–470. ACM (2015)
46. Sperber, D.: Explaining culture: A naturalistic approach. Cambridge University Press, New York (1996)
47. Spirtes, P.: Introduction to causal inference. J. Mach. Learn. Res. **11**(May), 1643–1662 (2010)
48. Wallach, H.: Computational social science: Toward a collaborative future. In: Computational Social Science: Discovery and Prediction (2016)
49. Watts, D.: Challenging the influentials hypothesis. WOMMA Measuring Word Mouth **3**(4), 201–211 (2007)
50. Watts, D.J.: Everything is obvious: * Once you know the answer. Crown Business (2011)