ROYAL STATISTICAL SOCIETY
DATA | EVIDENCE | DECISIONS

Journal of the Royal Statistical Society
Statistics in Society
Series A

J. R. Statist. Soc. A (2020)
183, *Part* 4, *pp.* 1659–1676

Causal inference, social networks and chain graphs

Elizabeth L. Ogburn and Ilya Shpitser

Johns Hopkins University, Baltimore, USA

and Youjin Lee

University of Pennsylvania, Philadelphia, USA

[Received December 2018. Revised June 2020]

Summary. Traditionally, statistical inference and causal inference on human subjects rely on the assumption that individuals are independently affected by treatments or exposures. However, recently there has been increasing interest in settings, such as social networks, where individuals may interact with one another such that treatments may spill over from the treated individual to their social contacts and outcomes may be contagious. Existing models proposed for causal inference using observational data from networks of interacting individuals have two major shortcomings. First, they often require a level of granularity in the data that is infeasible in practice to collect in most settings and, second, the models are high dimensional and often too big to fit to the available data. We illustrate and justify a parsimonious parameterization for network data with interference and contagion. Our parameterization corresponds to a particular family of graphical models known as chain graphs. We argue that, in some settings, chain graph models approximate the marginal distribution of a snapshot of a longitudinal data-generating process on interacting units. We illustrate the use of chain graphs for causal inference about collective decision making in social networks by using data from US Supreme Court decisions between 1994 and 2004 and in simulations.

Keywords: Causal inference; Chain graphs; Collective behaviour; Graphical models; Social networks

1. Introduction

Traditionally, statistical inference and causal inference on human subjects rely on the assumption that individuals are independently affected by treatments or exposures. This is sometimes referred to as the *no-interference* assumption; it is also part of the *stable unit treatment value* assumption. However, recently there has been increasing interest in settings where treatments 'spill over' from the treated individual to his or her social contacts, or where outcomes are contagious. Researchers who are interested in causal inference have developed methods for *interference*—when one individual's treatment or exposure affects not only his own outcome but also the outcomes of his contacts (Aronow and Samii, 2017; Athey *et al.*, 2018; Bowers *et al.*, 2013; Eckles *et al.*, 2016; Forastiere *et al.*, 2020; Graham *et al.*, 2010; Halloran and Hudgens, 2012; Hong and Raudenbush, 2006, 2008; Hudgens and Halloran, 2008; Jagadeesan *et al.*, 2020; Liu and Hudgens, 2014; Liu *et al.*, 2016; Ogburn and VanderWeele, 2014; Rosenbaum, 2007; Rubin, 1990; Sobel, 2006; Tchetgen Tchetgen and VanderWeele, 2012; VanderWeele, 2010). Researchers who are interested in social networks have attempted to model the spread of contagious outcomes across network ties (Christakis and Fowler, 2007, 2008, 2010; Ali and Dwyer, 2009; Cacioppo

Address for correspondence: Elizabeth L. Ogburn, Department of Biostatistics, Johns Hopkins Bloomberg School of Public Health, 615 North Wolfe Street, Baltimore, MD 21205, USA.
E-mail: eogburn@jhsph.edu

 0964–1998/20/1831659

et al., 2009; Lazer *et al.*, 2010; Rosenquist *et al.*, 2010), but existing methods for such modelling are either flawed (Cohen-Cole and Fletcher, 2008; Lyons, 2011; Shalizi and Thomas, 2011) or limited by strong assumptions and burdensome data requirements. One stumbling block for inference about networks of interacting individuals is dimensionality: in many settings, if n individuals can interfere with or transmit to one another, all n outcomes are dependent, resulting in a saturated likelihood with the number of parameters growing exponentially in n even before treatments and covariates have been included. In this paper we illustrate a parsimonious parameterization for such social network data and explore when this new parameterization might be justified. Our parameterization corresponds to a particular family of graphical models that are known as *chain graphs*. Different types of chain graph models have been proposed in the literature (Drton, 2009). In this paper we exclusively consider chain graph models under the Lauritzen, Wermuth and Frydenberg interpretation, because these models generalize graphical models that are associated with both directed acyclic graphs (DAGs) and undirected graphs, and form curved exponential families under many parameterizations (Lauritzen, 1996).

Lauritzen and Richardson (2002) defined causal models using chain graphs via structural equation semantics and Gibbs sampling. Shpitser *et al.* (2017) formulated causal chain graphs by using the potential outcomes semantics as a general approach to interference problems, and considered how causal mediation analysis ideas generalize to chain graph models to yield a principled approach to analysing contagion and infectiousness effects in interference contexts. Tchetgen Tchetgen *et al.* (2017), who assumed a conventional causal model for data drawn from units in a network, used chain graph Markov assumptions as a model on the observed data to make inference possible in the *full interference* setting, where the effective sample size is 1. Sherman and Shpitser (2018) gave a complete identification theory for causal effects in the presence of both network dependence and unobserved confounding in latent variable chain graph models, whereas Bhattacharya *et al.* (2019) gave a model selection algorithm in situations where the exact structure of a network given by a chain graph is not known. Pena (2019) described a general chain graph model which unified different types of chain graphs, allowing for different types of network relationships to be represented by different types of edges in the unified model. Lauritzen and Richardson (2002) argued against the cavalier use of chain graphs as causal models, but justified their use when interest is in a structured system which contains equilibrium distributions, as happens when outcomes of interest represent collective behaviour or collective decisions across interacting individuals. However, the true data-generating process where equilibrium-generating dynamics are absent are often better represented by a DAG model. Causal models of DAGs have been used for decades to guide inference and modelling, especially for causal inference (Pearl, 2000).

In this paper, we explore possible justifications for the use of chain graphs to model data on interacting units, where we assume a latent variable causal DAG model but use chain graphs as a tractable approximation of the resulting observed data likelihood. Although chain graph models are known to be incompatible with such DAG models in general, we show in simulations that, in certain settings, the conditional independences that are entailed by a chain graph model may approximate those from a DAG model with certain properties. In addition, we apply the chain graph approach to data on Supreme Court cases, where the underlying data-generating process may be viewed as containing equilibrium dynamics, as opinions about a case are gradually formed. Using chain graph models for causal inference with social network data extends a number of references that have used undirected graph models, or Markov random fields, to model social interactions, including the 'sociophysics' literature on Ising models for collective decision making.

The rest of the paper is organized as follows: Section 2 reviews required concepts, definitions

and notation pertaining to undirected, directed and chain graph models; readers who are familiar with graphical models may be able to skip this section. In Section 3, we describe previous work using DAGs for causal inference with social network data and in settings with contagion and interference, and previous work using undirected graphical models to study collective problem solving. We propose conditions under which chain graphs may approximate a true underlying DAG model for social network data and illustrate the relationship between the DAG model and a chain graph approximation in simulations. We explain how chain graphs can be used to analyse data with contagion and interference when DAG models are intractable. These chain graph models extend previously proposed models for collective problem solving in important and useful ways. In Section 4 we analyse data on Supreme Court decisions to illustrate how chain graph models can be used to estimate causal effects on collective outcomes. Only group level exposure variables are available in the real data, so in Section 5 we simulate data based on the Supreme Court decision data but with individual level exposures to illustrate the estimation of individual level causal effects by using chain graph models. Section 6 concludes the paper.

2. Graphs and graphical models

Graphical models use graphs—collections of vertices, representing random variables, and edges representing relationships between pairs of vertices to represent concisely conditional independences that hold between the random variables. At their most general, the graphical models that we shall consider in this paper are represented by mixed graphs containing directed ('$\rightarrow$'), and undirected ('$-$') edges, such that at most one edge connects two vertices. This section contains a brief, informal review of graphical models; a more comprehensive review is in the on-line appendix.

A graph with only undirected edges is an *undirected graph*, as in Fig. 1(a). Such graphs are often used to model networks of connected individuals; the absence of an edge between two nodes indicates, roughly, that the nodes are independent conditional on other intermediate nodes. When we analyse the US Supreme Court data, we shall use an undirected graphical model to represent the pairwise relationships between the justices, and the absence of an edge will indicate that there is no evidence of a direct connection between the two justices beyond what can be explained by individual level variables or by other pairwise connections in the network of justices.

A graph with only directed edges is called a *DAG*, as in Fig. 1(c). Such graphs underpin most causal inference methods, with the absence of an edge between two nodes indicating, roughly, that there is no causal effect of one on the other that is not mediated by other nodes on the graph. Fig. 1(c) depicts a setting in which A causes Y and C is a common cause, or a confounder, of the A–Y relationship.

A mixed graph containing both directed and undirected edges with no partially directed cycles is called a *chain graph* (Lauritzen and Wermuth, 1989; Frydenberg, 1990). A simple example is

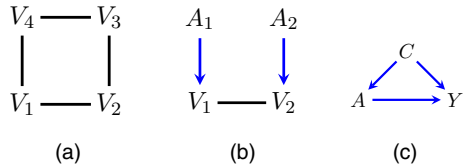

Fig. 1. (a) A simple undirected graph, (b) the simplest chain graph with an independence model not representable as either a DAG or an undirected graph and (c) a causal graph representing observed confounding of the treatment A and outcome Y by a set of covariates C

depicted in Fig. 1(b). As we describe in detail below, graphs of this type may be used to represent systems with interacting units. A chain graph forms a natural generalization of statistical models both associated with DAGs, and associated with undirected graphs. However, the combination of directed and undirected edges can result in conditional independences that would not be familiar or intuitive to practitioners who are familiar with DAG and undirected graph models. For example, in Fig. 1(b), A_1 is independent of V_2 given V_1 and A_2, as we might expect, but A_1 is *not* independent of V_2 given only V_1. The factorization that defines chain graph models and implies the constraints above in the special case of the model that is shown in Fig. 1(b) is given in the on-line appendix.

DAGs and chain graphs have both been used to define statistical and causal models. Statistical graphical models associate the observed data distribution $p(\mathbf{V})$ with a graph where vertices are associated with random variables in $\mathbf{V}$. Graphical models are often defined by means of a factorization, where $p(\mathbf{V})$ may be written as a product of smaller factors, with a rule for obtaining these factors given by the graph.

Causal inference from observational data is concerned with making inferences on *counterfactual* or *potential outcome* random variables of the form $Y(a)$ from the observed data distribution $p(\mathbf{V})$. Here, $Y(a)$ is taken to mean 'the variable Y if we had intervened on A, possibly contrary to fact, to set it to a'. Causal parameters of primary interest are usually low dimensional summaries or contrasts obtained from counterfactual distributions, rather than counterfactual distributions themselves. For example, the *average causal effect* is defined as the mean contrast $\mathbb{E}[Y(a)] - \mathbb{E}[Y(a')]$. Causal graphical models define a link between counterfactual targets of inference, such as the average causal effect, and the observed data distribution. Causal graphical models imply a factorization, and thus a statistical graphical model, on the observed data distribution. The link between this distribution and counterfactual parameters is provided by means of a modified factorization of the observed data distribution. In the case of causal models of a DAG, this modified factorization is known as the g-formula (Robins, 1986). In the case of causal models of a chain graph, this modified factorization is described in the on-line appendix, and originally in Lauritzen and Richardson (2002). Causal DAG models are powerful tools for causal inference using observational data and have gained widespread use in epidemiology, social sciences and other fields, because they can be used to display sources of bias such as confounding clearly, and can be used to derive identification theory for many counterfactual targets of inference in complex multivariate causal systems.

Chain graphs allow both directed and undirected edges, and can be used to define both statistical and causal graphical models that combine features of both undirected graphs and DAGs (Lauritzen, 1996). For simplicity, in this paper we shall restrict ourselves to chain graphs with an undirected component or block $\mathbf{Y}$, representing outcomes that are associated with nodes in a network, and with exposures $\mathbf{A}$ having directed edges into $\mathbf{Y}$. We shall consider only interventions on $\mathbf{A}$—here we shall not consider interventions on nodes in the undirected component of outcomes. A general treatment of chain graph models may be found in Lauritzen (1996), Lauritzen and Richardson (2002) and Sherman and Shpitser (2018).

3. Graphical models for social interactions

DAG models are assumed either implicitly or explicitly in almost all existing methods for learning about social interactions, social influence, interference, contagion and other causal effects from network data. Causal DAG models, or the mathematically equivalent causal structural equation models, correspond to a mechanistic view of the (macroscopic) world, which is espoused by most researchers across many disciplines and underpins almost all approaches to learning about

causal effects from data. Ogburn and VanderWeele (2014) have provided an overview of the use of DAGs to represent interference and contagion. New methods for learning about causal effects from social network data rely on assumptions that are consistent with DAG models, explicitly in the case of methods for observational data proposed by van der Laan (2014) and Ogburn *et al.* (2017) and implicitly in many of the methods that are based on randomized experiments (e.g. Aronow and Samii (2017), Athey *et al.* (2018), Bowers *et al.* (2013), Choi (2014), Eckles *et al.* (2016), Forastiere *et al.* (2020), Graham *et al.* (2010), Hong and Raudenbush (2006, 2008), Hudgens and Halloran (2008), Jagadeesan *et al.* (2020), Liu and Hudgens (2014), Liu *et al.* (2016), Rosenbaum (2007), Rubin (1990), Sobel (2006), Tchetgen Tchetgen and VanderWeele (2012) and VanderWeele (2010)). However, as we shall show in the next section and as has been acknowledged by some of the aforementioned researchers, DAGs in these settings can run into difficulties in practice.

Undirected graph models have also been used to model social networks (see, for example, West *et al.* (2014), Domingos and Richardson (2001), Ahmed and Xing (2009) and Kindermann and Snell (1980), and a niche literature uses a particular class of undirected graph models, namely *Ising models*, to model the collective behaviour of individual actors. The Ising model was originally developed by physicists to model spin states (up or down) of atoms of a metal arranged on a lattice (Ising, 1925); it is meant to represent the state of a physical system at a particular temperature. As temperature decreases but remains strictly greater than absolute zero, the system may transition from a relatively disordered state to a state where most spins are either up or down. Such a phenomenon is called a *phase transition*. Phase transitions do not occur for Ising models in a one-dimensional lattice, but they occur in all higher dimensional grids, starting with dimension 2 (Peierls, 1936). A literature has developed that uses the Ising model to represent the forming of consensus in groups of interacting individuals, each holding one of a pair of possible opinions, e.g. for or against a proposition (Galam *et al.* (1982), Galam (1997) and Sznajd-Weron and Sznajd (2000); see also Galam (2008) and references therein). A phase transition occurs if individuals converge to a consensus. The rectangular lattice underlying the model restricts each individual to interacting with a fixed number of other individuals, depending on the dimension of the lattice. Despite the fact that these models have not been fitted, validated or tested against real data, and despite the fact that social networks are usually nowhere close to being rectangular lattices, sociophysicists have made strong and empirically verifiable claims about social phenomena by using these models (e.g. Galam (2008)).

The Ising model is a special case of a log-linear model defined on an undirected graph. Chain graph models, which include undirected graphs as a special case, generalize Ising models for human behaviour in several directions: to arbitrary network structure, rather than lattices, to facilitate statistical inference and model fitting using real data, to introduce the idea of treatments with causal effects on nodes and to clarify the uses and limitations of such models for human behaviour. These chain graph models can also be seen to extend other undirected models for social networks.

Consider a social network of n individuals, or nodes. Node i is associated with a treatment or exposure A_i, an outcome Y_i and possibly covariates. For example, Y could represent opinions and A advertising campaigns; Y could represent behaviour and A encouragement interventions, or Y could represent an infectious disease and A vaccination. In our analysis of Supreme Court decisions below, Y_i is a binary variable representing whether justice i's decision was liberal or conservative, and A, which simultaneously treats all the justices, is an indicator of the issue area of the case. We specifically consider chain graphs in which the set of outcomes Y form a single undirected block, whereas treatments and covariates are represented by vertices with directed edges into Y-vertices. A simple example is depicted in Fig. 2(c).

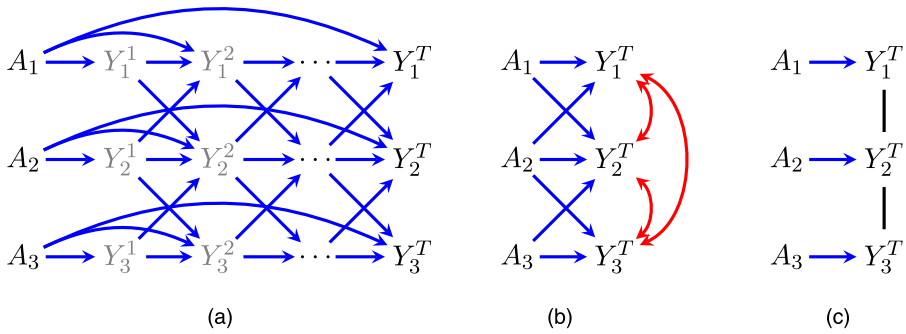

Fig. 2. (a) Causal DAG representing opinion formation among peers (A_i represents interventions meant to influence subject i; Y_i^k is the ith subject's opinion at time k), (b) a latent projection of the model in (a) onto variables A_1, A_2, A_3, Y_1^T, Y_2^T and Y_3^T, representing the distribution of opinion in (a) at time T, before equilibrium is reached (the bidirected arrows represent the fact that the outcomes at intermediate time points are unmeasured common causes of the observed outcomes) and (c) a chain graph model that approximates the distribution of opinion in (a) at time T under certain data-generating processes

When individuals' beliefs or opinions undergo phase transitions to orderly states, e.g. when there is external pressure to reach a unanimous consensus, or when it can be argued that the distribution of individuals' behaviours, beliefs, opinions or other outcomes attains an equilibrium across network ties, then a chain graph may be the correct model for the joint distributions of outcomes across a network and interventions on those outcomes. For example, in the Supreme Court data, outcomes represent decisions that are made under time constraints and with pressure for the nine justices to reach a unanimous decision; these may indeed be in equilibrium. More common in the existing literature are settings in which DAG models would be the most appropriate class of models, but they are not tractable given reasonable constraints on data collection.

Fig. 2(a) depicts a DAG model for a three-node network in which individuals 1 and 2 exhibit contagion, as do 2 and 3, where contagion is any causal effect of one individual's outcome at a particular time on their social contacts' future outcomes—or an arrow from Y_i^t to Y_j^s for $t < s$. In opinion formation, like the Supreme Court decision data, contagion could be the influence that one justice's way of thinking about a case can have on other justices through conversation or debate. The pairs of individuals who influence one another often correspond to pairs of individuals with ties in an observed social network. Fig. 2(a) does not include any direct interference; this would be a causal effect of one individual's treatment on another's outcome, i.e. an arrow from A_i to Y_j^t.

For this DAG to be valid, the units of time captured must be sufficiently small that any influence passing from individual 1 to 3 through 2 cannot occur in fewer than two time steps (Ogburn and VanderWeele, 2014). This will be so if influence can occur only during discrete interactions such as in-person or on-line encounters, and the unit of time is chosen to be the minimum time between encounters. This DAG model encodes several conditional independences and, if we can observe the outcome for all agents at all time steps, inference under these sorts of models may be possible (Ogburn *et al.*, 2017).

However, in most practical applications, with the exception of on-line social networks, it is possible to observe only the outcome at one or a few time points. If data are generated according to the DAG in Fig. 2(a) but the outcome is observed at only one time point (at which the outcome is not in a chain graph equilibrium), then the resulting model is represented by a mixed graph representing the *latent projection* (Verma and Pearl, 1990) of all of the variables in Fig. 2(a) onto the subset of those variables that are actually observed, with *bidirected* edges representing the

presence of one or more hidden common causes. A general construction algorithm for these latent projection mixed graphs was given by Pearl (2009), and the result for Fig. 2(a) is shown in Fig. 2(b).

Collecting or accessing the detailed temporal data that are required to use the models like Fig. 2(a) is often impractical or impossible, but the saturated model for the marginal in Fig. 2(b) quickly becomes unwieldy, as the number of parameters that are required to estimate and use the model grows exponentially with the number of nodes: the latent projection graph will generally not be sparse, even if the underlying social network governing opinion formation is. To see this, note that, after a single time step, an individual influences only neighbouring individuals but, after two time steps, also neighbours of neighbours. In the three-person network that is represented by Fig. 2, this is enough to render the latent projection of Fig. 2(b) fully saturated, with no conditional independences. After many time steps, the individual's influence would have time to reach most of the social network. This implies that any two outcomes at time t, for a sufficiently large t, will be related via a chain of hidden common causes, even if the corresponding individuals are far from each other in the social network. To represent these chains of hidden common causes, the latent projection graph would contain a clique of bidirected edges encompassing opinions of everyone in the network. Given a mixed graph containing such a clique, the number of parameters that are needed to specify the appropriate mixed graph likelihood will be exponential in the size of the clique in general (Evans and Richardson, 2014). These limitations are reflected in the literature, which posits DAG models (either explicitly or implicitly) but rarely includes applications to real data.

Unlike the model in Fig. 2(b), the chain graph model that is represented by Fig. 2(c) is not saturated and generally has a likelihood with the number of parameters polynomial in the size of the graph.

In general, local interactions between units leading to a causal model of the type that is shown in Fig. 2(a) cannot correctly be represented by undirected edges of a chain graph. Indeed, most seemingly reasonable uses of such edges lead to inconsistencies or Markov properties that are not represented by chain graphs, as discussed extensively in Lauritzen and Richardson (2002). Nevertheless, a chain graph model that is as sparse as the underlying social network may serve as a good approximation of the underlying latent variable causal model, even if the likelihood corresponding to its latent projection representation is intractable.

Consider chain graph models like that in Fig. 2(c), but with arbitrary undirected components corresponding to outcomes that are observed on social network nodes. These models imply that each node's outcome is independent of its non-neighbours' outcomes conditionally on its neighbours' outcomes and on any treatments or covariates with arrows pointing into the node. For chain graphs like Fig. 2(c), with a single treatment for each node, the conditional independences that are implied by the graph are of the form $Y_i^T \perp\!\!\!\perp A_j, Y_j^t | A_i, \{Y_l^T, \forall\, l \text{ adjacent to } i\}$. These conditional independences fail to hold in the corresponding DAG models because of two different types of paths, depicted by colour in Fig. 3.

Paths like that in Fig. 3(a) represent the fact that the past outcomes of mutual connections affect both Y_i^T and Y_j^T; this is just one of many such paths. All these paths can be blocked by conditioning on $\{Y_l^t,$ for all l adjacent to i and for $1 \leqslant t \leqslant T-1\}$ (Ogburn and VanderWeele, 2017). If the outcome evolves slowly over time, Y_l^t and Y_l^T will be highly correlated and conditioning on $\{Y_l^T,$ for all l adjacent to $i\}$ will mostly block these paths. We expect the paths through Y_l^t to be weaker for smaller t than for t close to T. If paths through Y_l^t are weaker for earlier times t, then the relationship between Y_l^t and Y_l^T can also weaken for decreasing t—as long as it remains sufficiently strong to allow conditioning on Y_l^T to block paths through Y_l^t approximately.

However, conditioning on $\left\{Y_l^T,\right.$ for all l adjacent to $\left.i\right\}$ opens paths through colliders, like

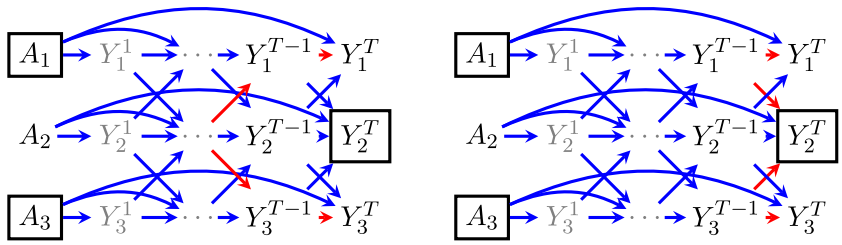

Fig. 3. Paths that connect Y_1^T and Y_3^T even when conditioning on Y_2^T and A_1 and/or A_3: □, variables that are conditioned on

that depicted in Fig. 3(b). M-shaped collider paths like these are known often to induce weak dependence (Greenland, 2003), and the magnitude can be bounded more precisely by using knowledge of the partial correlation structure of the variables along the path (Chaudhuri and Richardson, 2002). Informally, if the dependence of Y_l^T on Y_l^{T-1} is stronger than that of Y_l^T on Y_i^{T-1} and Y_j^{T-1}, as it will be if the outcome evolves slowly over time, then the dependence that is induced by paths through colliders may be negligible.

Although chain graph models exist in which the relationships along undirected edges are not symmetric, we found in simulations that DAGs with symmetric relationships for connected pairs of individuals were better approximated by chain graphs. It might be reasonable to assume this kind of symmetry if, for example, the outcome is a behaviour or belief and the subjects are peers with no imbalance of power or influence, or if the outcome is an infectious disease and the subjects have similar underlying health and susceptibility statuses.

On the basis of the arguments above, we make the following conjecture. Assume that data on interacting units are generated from a process given by a temporal DAG model containing baseline factors, exposures and unit outcomes evolving over time, where

(a) there is contagion (a causal link from a network neighbour's outcome at a previous time point to the unit's present outcome) but no direct interference (causal links between neighbouring units at the same time point),
(b) outcomes evolve slowly over time and
(c) causal relationships are symmetric between pairs of units that share a network tie.

Then a marginal distribution containing a set of baseline factors, exposures and outcomes at the final observed time point generated from such a process is well approximated by a chain graph model that links outcomes of units that share a network tie by an undirected edge while maintaining directed edges between variables within every unit.

Formal results, e.g. quantifying the distance between a chain graph approximation and a true DAG model, are beyond the scope of this paper and may not be possible in full generality. However, we verified our conjecture in simulations. We simulated 10 random nine-node social networks with edge probability $p = 0.3$ for every pair of nodes. For each random network, we generated outcomes for the nodes 1000 times according to a DAG model like that in Fig. 2(a), with symmetric causal effects for every edge in the underlying network. For all non-adjacent pairs (i, j), we tested three (conditional) independence hypotheses:

(a) the null hypothesis of marginal independence $Y_i^t \perp\!\!\!\perp Y_m^t$,
(b) the null hypothesis of conditional independence $Y_i^t \perp\!\!\!\perp Y_m^t | \{Y_l^t, \forall\, l \text{ adjacent to } i\}$ and
(c) the null hypothesis of conditional independence $Y_i^t \perp\!\!\!\perp Y_m^t | A_i, \{Y_l^t, \forall\, l \text{ adjacent to } i\}$.

The conditional independences in hypotheses (b) and (c) are implied by the chain graph model (for the simple case of three units, the associated chain graph is shown in Fig. 2(c)) but do not hold in the true model (for the simple case of three units, the associated hidden variable DAG is shown in Fig. 2(a)). We also tested hypothesis (a) to ensure that we had not inadvertently generated data with such weak dependences between individuals that (b) and (c) would hold because of marginal independences. We found that the conditional independence nulls were rejected at close to the nominal rate of 5% expected under the null. In contrast, the marginal independence null was rejected more frequently, suggesting that conditioning on neighbours' outcomes may recover approximate independence under at least some data-generating processes, and that the chain graph model may in those cases be a reasonable parsimonious approximation to the true underlying conditional independences. In another set of simulations with three agents we found that a chain graph model could approximately estimate causal effects from data generated under a DAG model. For details and full results see the on-line supplementary materials.

In the next sections we illustrate the use of chain graph models for causal inference about social interactions in real and simulated data on US Supreme Court decisions.

4. Using chain graph models to analyse US Supreme Court decisions

The US Supreme Court is comprised of nine justices, one of whom is the Chief Justice, tasked with presiding over oral arguments, serving as the spokesperson for the court and other administrative roles. After a case has been heard by the Supreme Court, the justices discuss and decide the case over a period of several weeks or months. The final outcome is decided by majority vote; the majority and, when the decision is not unanimous, the minority write opinions justifying their decisions. The oral and written arguments that are presented to the court and the judicial opinions are public resources; however, we have no access to the debates and discussions that lead the justices to their decisions. This precludes the use of a DAG model for the evolution of justices' opinions over time but is amenable to a chain graph model with Y_i defined as Justice i's final opinion. Indeed, such a chain graph model may be appropriate here not just as an approximation. This is an example of collective opinion formation with pressure for unanimity, and it may attain a degenerate equilibrium in which each justice is unwilling or unable to be swayed from a fixed decision, even with further discussions. In what follows we are agnostic about whether the data were truly generated under a chain graph model representing opinion formation as an equilibrium process or a DAG model; we simply assume that the chain graph model that we use is a good approximation for the observed data distribution that is generated by the true data-generating process. However, we also give the *caveat* that this data analysis is meant to serve as an example illustrating the use of chain graph models for simultaneous outcomes across network nodes, rather than to draw substantive conclusions about the operations of the US Supreme Court.

Data on all Supreme Court decisions since 1946, along with rich information on the nature of the cases and the opinions, is maintained by Washington University Law School's Supreme Court database (`http://scdb.wustl.edu/data.php`). We used the subset of these data corresponding to the Second Rehnquist Court, a period of 10 years (1994–2004) during which the same nine justices served together: William Rehnquist (Chief Justice), John Paul Stevens, Sandra Day O'Connor, Antonin Scalia, Anthony Kennedy, David Souter, Clarence Thomas, Ruth Bader Ginsburg and Stephen Breyer. Over these 10 years the court decided 893 cases; each case serves as a data point for our chain graph model.

The Supreme Court database has classified each case into one of 14 issue areas, such as criminal procedure and civil rights. Table 1 presents the number of cases that are associated with

Table 1. Number of cases decided during 1994–2004†

Issue area	*Number of cases*
Criminal procedure	231
Civil rights	161
First amendment	59
Due process	43
Privacy	21
Attorneys	5
Unions	18
Economic activity	145
Judicial power	133
Federalism	57
Federal taxation	20
Total	893

†No cases were identified for interstate relations, miscellaneous or private action.

each issue area; cases were heard in only 11 out of 14 areas during the Second Rehnquist Court. The database has also classified the decisions for each case as either liberal or conservative. For each case, each justice has an outcome Y, which denotes whether the justice's personal opinion was liberal ($Y = 1$) or conservative ($Y = -1$); the ruling in the case is liberal if at least five of the justices form liberal opinions and conservative otherwise.

Using data on the case issue area and binary outcomes of each justice for $n = 893$ cases, we examined the effect of issue area on whether the individual justices reached conservative or liberal opinions, whether the court reached a conservative *versus* a liberal ruling and whether the decision was unanimous or divided. (Only group level exposure variables are available in the database; in the next section we simulate individual level treatments to illustrate the use of chain graph models in more general settings.) There is strong evidence (including self-report by the justices) that the Court works hard to come to unanimous decisions, but 5-to-4 decisions are frequent (Sunstein, 2014; Riggs, 1992), and we found that there is an effect of issue area on this outcome. There is also considerable academic interest in each justice's personal orientation (Songer and Lindquist, 1996; Tate, 1981). During the Rehnquist Court, 56% of the decisions were conservative. Clarence Thomas was the most conservative justice, signing the conservative opinion in 72% of cases, whereas Ruth Bader Ginsburg was the most liberal, signing the liberal opinion in 60% of cases. We found that issue area had a strong effect on both individual outcomes and on overall court decisions, which is consistent with literature on the effect of issue areas on the ideology of each justice and on the final decision of the Supreme Court (Tate, 1981; Lu and Wang, 2011).

We considered the effects of

(a) criminal procedure,
(b) civil rights,
(c) economic activity and
(d) judicial power on conservative *versus* liberal opinions.

To illustrate our method for binary treatment variables, and because considering a higher dimensional treatment variable would require a larger sample size, we fitted separate chain graph models for each issue area, coded $a = 1$ for the issue area of interest and $a = 0$ for all other issue

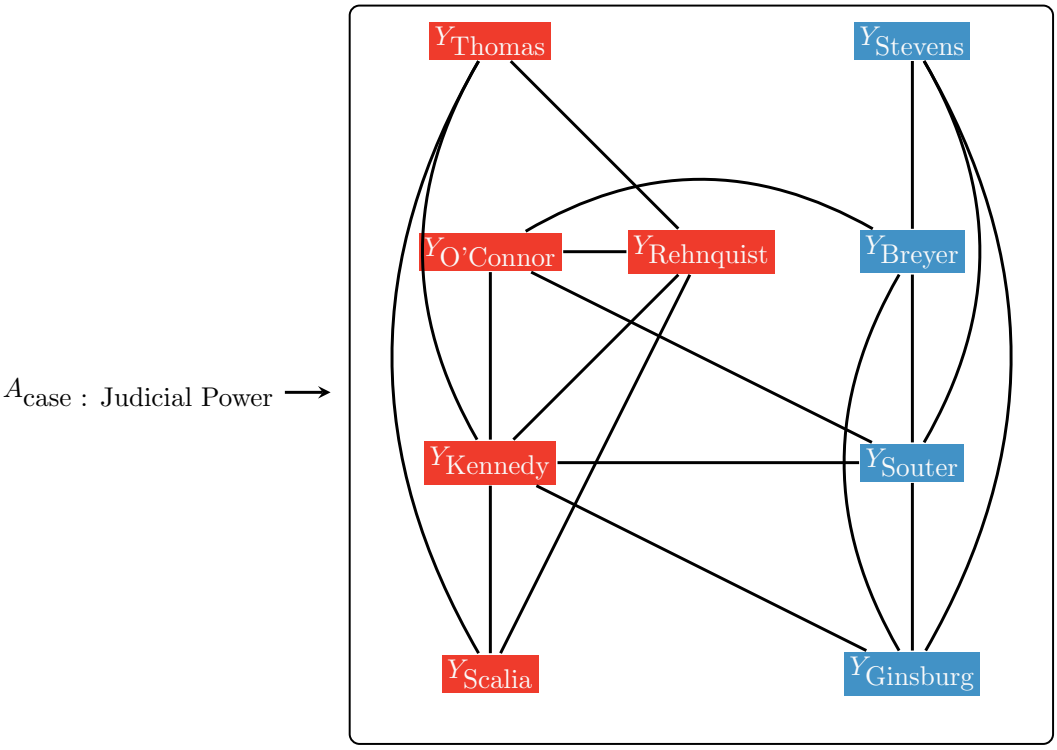

Fig. 4. Estimated underlying network among the nine justices for cases involving judicial power: the colour of each node indicates whether the justice is commonly considered to be liberal (right) or conservative (left)

areas. Although issue area is not manipulable in practice, these causal effects are of explanatory interest: the effect of a hypothetical intervention that holds all relevant features of a case constant while changing the issue area can elucidate the role that issue area plays in the collective decision making of the justices, even if it cannot inform policy.

So far, we have assumed that the underlying network according to which individuals interact is known. In this case, however, we have only anecdotal evidence about the relationships between the justices. Therefore, as a first step, we estimated the undirected component of the chain graph model by using data on the justices' opinions for a particular issue area. Estimation was implemented with the `XMRF` R package (Wan *et al.*, 2016), which used gradient descent to maximize penalized node conditional likelihoods and fitted an exponential family Markov random-field model to the data. Incorporating uncertainty about the underlying network into causal inference is beyond the scope of this paper but is a problem on which we are actively working.

To complete the chain graph of interest, we added a single treatment variable, i.e. issue area, that jointly affects each justice's outcome. Our causal chain graph models and resulting analyses assume that there are no confounders of the relationship between issue area and justices' decisions. We think that this is a plausible assumption, but it is possible that there could be (unmeasured) confounding, e.g. due to the court's selection of cases to hear, or due to time trends in both issue areas and justices' behaviour. We demonstrate how to use chain graph models with individual level treatments and confounders in Section 5.

The resulting chain graph for the judicial power issue area is displayed in Fig. 4. We found that justices interact with one another on the basis of their shared liberal or conservative leanings,

as would be expected, but also across that divide, in some cases through relationships that are supported by anecdotal evidence. For example, the tie between Breyer and O'Connor could be explained by their social connections (`http://blogs.findlaw.com/supreme_court/2017/03/supreme-court-shutters-justice-oconnors-workout-class.html`) or their shared views on judicial independence (`http://www.pbs.org/newshour/bb/law-july-dec06-independence_09-26/`).

Separately for each of the four issue areas, we estimated the parameters of the following chain graph model:

$$p\{\mathbf{Y}(a)=(y_1,y_2,\ldots,y_9)\}=\frac{1}{Z(a)}\exp\left(\sum_{i=1}^{9}h_i y_i+\sum_{i,j=1,e_{ij}=1}^{9}k_{ij}y_i y_j+\sum_{i=1}^{9}\gamma_i a y_i\right), \qquad (1)$$

where $\mathbf{Y}(a)$ is the vector of counterfactual outcomes under $A=a$, $e_{ij}=1$ implies that justices i and j share an undirected edge in the chain graph and $Z(a)$ is a normalizing constant. The parameters h, k and γ all relate to conditional log-odds (in this particular model they must first be multiplied by 2, since Y is coded 1 or -1 instead of 1 or 0). The parameter h_i represents the conservative or liberal leaning of Justice i, with a positive h_i indicating bias towards liberal opinions. Specifically, $2h_i$ is the conditional log-odds of $y_i=1$ given that all the parameters that are associated with relational dependence of y_i ($\{k_{ij}:e_{ij}=1\}$) and with the treatment (γ_i) are 0. The interaction parameter k_{ij} captures the tendency of Justices i and j to agree beyond what can be explained by their individual leanings, with a positive k_{ij} indicating a tendency to agree whereas a negative k_{ij} indicates a tendency to disagree ($2k_{ij}$ is the conditional log-odds-ratio for the association between Y_i and Y_j). The parameter γ_i gives the causal effect of issue area A on Justice i's opinions, with positive γ_i indicating a tendency towards liberal opinions above and beyond what can be explained by the Justice's independent leaning or by the interactions with the other justices ($2\gamma_i$ is the conditional log-odds-ratio for $Y_i=1$ when $A=1$ compared with $A=0$). In principle, higher order interactions could be added to the model to capture tendencies of larger groups of justices to agree or disagree beyond what the pairwise interactions explain, but there was no empirical evidence to support the inclusion of these interactions (see the on-line supporting material). We bootstrapped the standard errors to calculate 95% confidence intervals for the parameters of interest, with $nb=500$ bootstrap samples for each model.

Fig. 5 illustrates the estimated parameter values for model 1, when A is an indicator of judicial power *versus* other issue areas. The shade of the node reflects the estimated main effect $\hat{h}_i$, the width of the undirected edges reflects the magnitude and sign of the estimated interaction between justices i and j, $\hat{k}_{ij}$, the width and colour of a directed edge reflect the estimated effect of A on Justice i's decisions, $\hat{\gamma}_i$, when A represents judicial power *versus* other issue areas.

Using the model that is given in equation (1), we estimated the causal effects of each issue area on the majority-based decisions of the nine justices. As an example, the causal effect (on the risk difference scale) of the judicial power issue area on the probability of a unanimous liberal decision (i.e. $\mathbf{Y}=\mathbf{1}$) is

$$\begin{aligned}p\{\mathbf{Y}(1)=\mathbf{1}\}-p\{\mathbf{Y}(0)=\mathbf{1}\}&=\frac{1}{Z(1)}\exp\left(\sum_{i=1}^{9}h_i+\sum_{i,j=1,e_{ij}=1}^{9}k_{ij}+\sum_{i=1}^{9}\gamma_i\right)\\&\quad-\frac{1}{Z(0)}\exp\left(\sum_{i=1}^{9}h_i+\sum_{i,j=1,e_{ij}=1}^{9}k_{ij}\right)\\&=\exp\left(\sum_{i=1}^{9}h_i+\sum_{i,j=1,e_{ij}=1}^{9}k_{ij}\right)\left\{\frac{\exp(\sum_{i=1}^{9}\gamma_i)}{Z(1)}-\frac{1}{Z(0)}\right\}.\end{aligned}$$

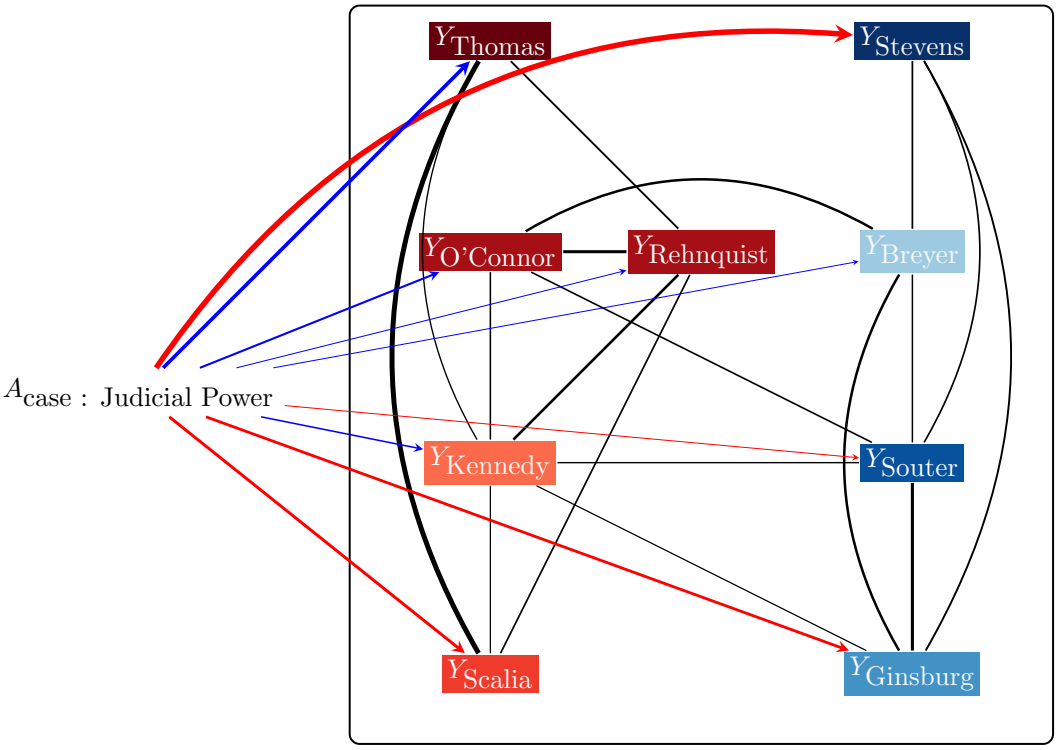

Fig. 5. The colour of each node represents the justice's tendency towards liberal or conservative decisions, estimated by $\hat{h}_i$, with darker red representing more conservative and darker blue more liberal decisions: the width of the undirected edges represents the strength of pairwise interactions between justices, estimated by $\hat{k}_{ij}$; the colour and width of the directed edges represent the direction and magnitude of the individual level causal effects of an intervention on whether or not a case concerns judicial power, estimated by $\hat{\gamma}_i$, with red for negative and blue for positive effects

When $\Sigma_{i=1}^{9}\gamma_i = 0$, i.e. when the sum of the log-odds-ratios is equal to 0, the causal effect is also equal to 0. This causal effect, which aggregates across all the justices, depends not only on the γ_i-parameters but also on the main effects h_i and interactions k_{ij}. We can similarly derive the effect of issue area on the probability of a unanimous conservative decision, i.e. $\mathbf{Y} = \mathbf{0}$, or on split 5–4 decisions, where $\Sigma Y_i = 5$ represents a 5-to-4 liberal decision and $\Sigma Y_i = 4$ represents a 5-to-4 conservative decision. We could also derive causal effect estimates on the risk ratio or odds ratio scale. We found that judicial power had the strongest causal effect on the probability of unanimous conservative decisions (Table 2). Economic activity caused an increase in the probability of liberal unanimous decisions (on-line supporting material Table 4). Criminal procedures and civil rights both caused an increase in the probability of 5-to-4 conservative decisions (supporting material Tables 2 and 3).

5. Simulations with individual level treatments and covariates

In the application above, we considered a single treatment variable that jointly affects the justices' opinions, and we assumed no confounding. To illustrate how chain graphs can be used to estimate causal effects with individual level treatments and with confounders, we simulated data from the undirected component of the graph in Fig. 4 with the addition of individual

Table 2. Estimated potential outcomes and causal effects (with standard errors in parentheses) comparing judicial power ($a = 1$) with other issue areas ($a = 0$), where the outcomes of interest are the probabilities of unanimous decisions

	$p\{\mathbf{Y}(a) = \mathbf{0}\}$, *probability of a unanimous conservative decision*	$p\{\mathbf{Y}(a) = \mathbf{1}\}$, *probability of a unanimous liberal decision*
Judicial power ($a = 1$)	0.33 (0.03)	0.16 (0.02)
Other issues ($a = 0$)	0.20 (0.01)	0.17 (0.01)
Causal effect	0.13 (0.03)	−0.01 (0.03)

Table 3. Probability of having a unanimous liberal decision ($\Sigma\mathbf{y} = 9$), unanimous conservative decision ($\Sigma\mathbf{y} = 0$), five liberal votes ($\Sigma\mathbf{y} = 5$) and five conservative votes ($\Sigma\mathbf{y} = 4$) under six treatment assignments

Intervened justice ($\mathbf{a}$)	$p\{\mathbf{Y}(\mathbf{a}) = \mathbf{y}; \Sigma\mathbf{y} = 9\}$	$p\{\mathbf{Y}(\mathbf{a}) = \mathbf{y}; \Sigma\mathbf{y} = 0\}$	$p\{\mathbf{Y}(\mathbf{a}) = \mathbf{y}; \Sigma\mathbf{y} = 5\}$	$p\{\mathbf{Y}(\mathbf{a}) = \mathbf{y}; \Sigma\mathbf{y} = 4\}$
O'Connor, Scalia, Kennedy, Thomas	0.37	0.11	0.06	0.06
Stevens, Souter, Ginsburg, Breyer	0.21	0.06	0.13	0.19
Rehnquist	0.16	0.23	0.07	0.10
Thomas	0.17	0.24	0.07	0.10
Stevens	0.13	0.19	0.09	0.13
Scalia	0.16	0.23	0.07	0.10

level treatments A, and individual level binary confounders C that are dependent across justices and have direct casual effects on A and Y. Treatment $A_i \in \{0, 1\}$ nudges Justice i towards a liberal ($A = 1$) or conservative ($A = 0$) decision. For an initial simulation, we set the main effect and pairwise interaction parameters to their estimates from a log-linear model fit to the actual Supreme Court data. We then varied the magnitude of the main effects and two-way interaction terms by multiplying those parameters by $\alpha, \beta \in \{0.5, 1\}$ respectively. Specifically, we generated the outcomes $\mathbf{Y}$ from the chain graph model below, where h_i and k_{ij} were estimated from the real data, and we varied α and β across simulation settings:

$$p\{\mathbf{Y} = (y_1, y_2, \ldots, y_9) | \mathbf{A} = \mathbf{a}, \mathbf{C} = \mathbf{c}\} = \frac{1}{Z(\mathbf{a}, \mathbf{c})} \exp\left(\alpha \sum_{i=1}^{9} h_i y_i + \beta \sum_{i,j=1, e_{ij}=1}^{9} k_{ij} y_i y_j + \gamma \sum_{i=1}^{9} a_i y_i + \kappa \sum_{i=1}^{9} c_i y_i \right). \quad (2)$$

For each combination of α and β, we generated 500 simulated data sets from the chain graph model, each of which used Gibbs sampling to produce 2000 observations of $(\mathbf{Y}, \mathbf{A}, \mathbf{C})$. Additional details of the simulation are in the on-line supporting material.

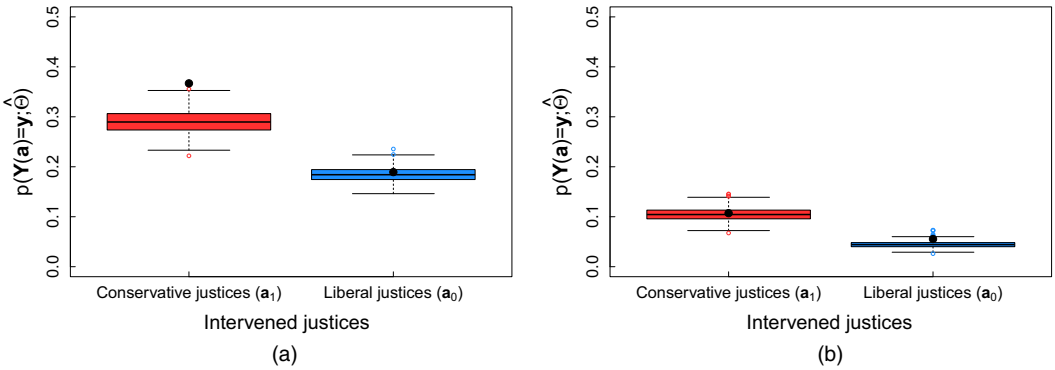

Fig. 6. Effects of intervening on the four liberal ($\mathbf{A} = \mathbf{a}_1$) or the four conservative justices ($\mathbf{A} = \mathbf{a}_0$) on the probabilities of (a) unanimous liberal ($\mathbf{y} = (y_1, y_2, \ldots, y_9) = \mathbf{1}_9$) and (b) unanimous conservative decisions ($\mathbf{y} = (y_1, y_2, \ldots, y_9) = \mathbf{0}_9$) : •, $p\{\mathbf{Y}(\mathbf{a}) = \mathbf{y}; \Theta\}$

Causal effects comparing two different treatment vectors $\mathbf{a}_1 = (a_{1,1}, a_{2,1}, \ldots, a_{1,9})$ and $\mathbf{a}_0 = (a_{0,1}, a_{0,2}, \ldots, a_{0,9})$ are functions of the parameters of the chain graph model in equation (2). For example, the causal effect on the probability of unanimous liberal decisions (on the risk difference scale) is given by

$$
\begin{aligned}
p\{\mathbf{Y}(\mathbf{a}_1) = \mathbf{1}\} - p\{\mathbf{Y}(\mathbf{a}_0) = \mathbf{1}\} &= \sum_{\mathbf{C}} \{p(\mathbf{Y} = \mathbf{1} | \mathbf{A} = \mathbf{a}_1, \mathbf{C} = \mathbf{c}) - p(\mathbf{Y} = \mathbf{1} | \mathbf{A} = \mathbf{a}_0, \mathbf{C} = \mathbf{c})\} p(\mathbf{C} = \mathbf{c}) \\
&= \sum_{\mathbf{C}} \exp\left(\alpha \sum_{i=1}^{9} \hat{h}_i + \beta \sum_{i,j=1, e_{ij}=1}^{9} \hat{k}_{ij} + \kappa \sum_{i=1}^{9} c_i \right) \\
&\quad \times \left\{ \frac{\gamma \sum_{i=1}^{9} a_{1,i}}{Z(\mathbf{a}_1, \mathbf{c})} - \frac{\gamma \sum_{i=1}^{9} a_{0,i}}{Z(\mathbf{a}_0, \mathbf{c})} \right\} p(\mathbf{c}).
\end{aligned}
$$

Table 3 presents the true probability of four counterfactual outcomes when $\alpha = \beta = 1$, under six treatment assignments (treating four conservative justices; treating four liberal justices; treating chief justice Rehnquist; treating Justice Thomas; treating Justice Stevens; treating Justice Scalia). Table 3 shows that, in our simulated data, treating the four conservative justices results in a higher probability of unanimous liberal decisions (0.37) than treating the four liberal justices (0.21). Similarly, treating the most liberal justice (Stevens) has the smallest effect on the probability of unanimous liberal decisions compared with treating Justice Rehnquist, Thomas or Scalia. Treating Justice Stevens has the greatest effect on the probabilities of 5-to-4 or 4-to-5 decisions. Treating the four liberal justices together results in a relatively high probability of 4 (liberal)-to-5 (conservative) decisions (0.19). When we estimated each of these probabilities in 500 simulations, the average absolute bias ranged from 0.002 to 0.055 and the average standard error ranged from 0.003 to 0.022; full results are in the on-line appendix.

As an example, Fig. 6 shows the true and estimated potential outcomes—in this case the probability of unanimous decisions—under two treatment vectors. Treatment $\mathbf{a}_1$ treats (nudges towards a liberal decision) the four most conservative justices (Justice O'Connor, Scalia, Kennedy and Thomas), and treatment $\mathbf{a}_0$ treats the four most liberal justices (Justice Stevens, Souter, Ginsberg and Breyer). We found that treating the conservative justices had a significant positive effect on the probability of a unanimous liberal decision compared with treating the liberal justices. In contrast, there was no significant causal effect on the probability of unanimous conservative decisions. In the on-line supporting materials, we also present similar results under different

simulated magnitudes of the main effect parameters for each justice (controlled by α) and pairwise interactions for connected justices (controlled by β), and for different treatment vectors. All the code and accompanying data can be found at `https://github.com/youjin1207/Chain`.

6. Conclusion and next steps

We have described a chain graph model for outcomes that are associated with nodes in a social network, with dependence along network ties induced by social interactions, contagion or interference. Although our Supreme Court example afforded multiple, ostensibly independent and identically distributed observations from the same chain graph, statistical inference for a single realization of a large chain graph was developed in Tchetgen Tchetgen *et al.* (2017). The chain graph model can represent only the true data-generating distribution if the outcomes are in very specific kinds of equilibria, which may be plausible if the outcomes represent collective beliefs or decisions (as in the Supreme Court example) but are often implausible. However, we showed that data that are generated from a causal DAG may be well approximated by a chain graph model, clarifying the conditions under which it may be reasonable to use these models in the literature on Ising models for collective behaviour, undirected models for social networks and chain graph models for causal inference in social networks. This approximation has two major limitations: first, it requires that the outcome evolve slowly over time and, second, it requires that the only source of dependence across nodes in the network be due to the causal effects from one node's outcome to another. This rules out *latent variable dependence*, where outcomes from nodes that are connected or close in the social network may be dependent because of shared environment, genetics or other characteristics (Ogburn *et al.*, 2017; Ogburn, 2018). Future work is needed to develop tractable statistical models that can handle these more general kinds of dependence, and to test these models against real world data.

Acknowledgements

ELO and YL were supported by Office of Naval Research grant N000141512343. ELO, YL and IS were supported by Office of Naval Research grant N000141812760.

The authors are grateful for helpful feedback from Henry Farrell, Cosma Shalizi and other participants in a workshop sponsored by the MacArthur Foundation Research Network on Opening Governance.

References

Ahmed, A. and Xing, E. P. (2009) Recovering time-varying networks of dependencies in social and biological studies. *Proc. Natn. Acad. Sci. USA*, **106**, 11878–11883.

Ali, M. M. and Dwyer, D. S. (2009) Estimating peer effects in adolescent smoking behavior: a longitudinal analysis. *J. Adolesc. Hlth*, **45**, 402–408.

Aronow, P. M. and Samii, C. (2017) Estimating average causal effects under general interference, with application to a social network experiment. *Ann. Appl. Statist.*, **11**, 1912–1947.

Athey, S., Eckles, D. and Imbens, G. W. (2018) Exact p-values for network interference. *J. Am. Statist. Ass.*, **113**, 230–240.

Bhattacharya, R., Malinsky, D. and Shpitser, I. (2019) Causal inference under interference and network uncertainty. In *Proc. 35th Conf. Uncertainty in Artificial Intelligence*. Association for Uncertainty in Artificial Intelligence Press.

Bowers, J., Frederickson, M. M. and Panagopoulos, C. (2013) Reasoning about interference between units: a general framework. *Polit. Anal.*, **21**, 97–124.

Cacioppo, J. T., Fowler, J. H. and Christakis, N. A. (2009) Alone in the crowd: the structure and spread of loneliness in a large social network. *J. Personlty Socl Psychol.*, **97**, 977–991.

Chaudhuri, S. and Richardson, T. (2002) Using the structure of d-connecting paths as a qualitative measure of the strength of dependence. In *Proc. 19th Conf. Uncertainty in Artificial Intelligence*, pp. 116–123. San Francisco: Morgan Kaufmann.
Choi, D. S. (2014) Estimation of monotone treatment effects in network experiments. *Preprint arXiv:1408.4102*.
Christakis, N. and Fowler, J. (2007) The spread of obesity in a large social network over 32 years. *New Engl. J. Med.*, **357**, 370–379.
Christakis, N. and Fowler, J. (2008) The collective dynamics of smoking in a large social network. *New Engl. J. Med.*, **358**, 2249–2258.
Christakis, N. and Fowler, J. (2010) Social network sensors for early detection of contagious outbreaks. *PLOS One*, **5**, no. 9, article e12948.
Cohen-Cole, E. and Fletcher, J. (2008) Is obesity contagious?: Social networks vs. environmental factors in the obesity epidemic. *J. Hlth Econ.*, **27**, 1382–1387.
Domingos, P. and Richardson, M. (2001) Mining the network value of customers. In *Proc. 7th Int. Conf. Knowledge Discovery and Data Mining*, pp. 57–66. New York: Association for Computing Machinery.
Drton, M. (2009) Discrete chain graph models. *Bernoulli*, **15**, 736–753.
Eckles, D., Karrer, B. and Ugander, J. (2016) Design and analysis of experiments in networks: reducing bias from interference. *J. Causl Inf.*, **5**, no. 1.
Evans, R. J. and Richardson, T. S. (2014) Markovian acyclic directed mixed graphs for discrete data. *Ann. Statist.*, 1–30.
Forastiere, L., Airoldi, E. M. and Mealli, F. (2020) Identification and estimation of treatment and interference effects in observational studies on networks. *J. Am. Statist. Ass.*, to be published.
Frydenberg, M. (1990) The chain graph Markov property. *Scand. J. Statist.*, **17**, 333–353.
Galam, S. (1997) Rational group decision making: a random field Ising model at t = 0. *Physica* A, **238**, 66–80.
Galam, S. (2008) Sociophysics: a review of Galam models. *Int. J. Mod. Phys.* C, **19**, 409–440.
Galam, S., Gefen, Y. and Shapir, Y. (1982) Sociophysics: a new approach of sociological collective behaviour: I, mean-behaviour description of a strike. *J. Math. Sociol.*, **9**, 1–13.
Graham, B., Imbens, G. and Ridder, G. (2010) Measuring the effects of segregation in the presence of social spillovers: a nonparametric approach. *Technical Report*. National Bureau of Economic Research, Cambridge.
Greenland, S. (2003) Quantifying biases in causal models: classical confounding vs collider-stratification bias. *Epidemiology*, **14**, 300–306.
Halloran, M. and Hudgens, M. (2012) Causal inference for vaccine effects on infectiousness, *Int. J. Biostatist.*, **8**, no 2.
Hong, G. and Raudenbush, S. (2006) Evaluating kindergarten retention policy. *J. Am. Statist. Ass.*, **101**, 901–910.
Hong, G. and Raudenbush, S. (2008) Causal inference for time-varying instructional treatments. *J. Educ. Behav. Statist.*, **33**, 333–362.
Hudgens, M. G. and Halloran, M. E. (2008) Toward causal inference with interference. *J. Am. Statist. Ass.*, **103**, 832–842.
Ising, E. (1925) Beitrag zur Theorie des Ferromagnetismus. *Z. Phys.*, **31**, 253–258.
Jagadeesan, R., Pillai, N. and Volfovsky, A. (2020) Designs for estimating the treatment effect in networks with interference. *Ann. Statist.*, **48**, 679–712.
Kindermann, R. P. and Snell, J. L. (1980) On the relation between Markov random fields and social networks. *J. Math. Soc.*, **7**, 1–13.
van der Laan, M. J. (2014) Causal inference for a population of causally connected units. *J. Causl Inf.*, **2**, 13–74.
Lauritzen, S. L. (1996) *Graphical Models*. Oxford: Clarendon.
Lauritzen, S. L. and Richardson, T. S. (2002) Chain graph models and their causal interpretations. *J. R. Statist. Soc.* B, **64**, 321–348.
Lauritzen, S. L. and Wermuth, N. (1989) Graphical models for associations between variables, some of which are qualitative and some quantitative. *Ann. Statist.*, **17**, 31–57.
Lazer, D., Rubineau, B., Chetkovich, C., Katz, N. and Neblo, M. (2010) The coevolution of networks and political attitudes. *Polit. Commun.*, **27**, 248–274.
Liu, L., Hudgens, M. and Becker-Dreps, S. (2016) On inverse probability-weighted estimators in the presence of interference. *Biometrika*, **103**, 829–842.
Liu, L. and Hudgens, M. G. (2014) Large sample randomization inference of causal effects in the presence of interference. *J. Am. Statist. Ass.*, **109**, 288–301.
Lu, Y. and Wang, X. (2011) Understanding complex legislative and judicial behaviour via hierarchical ideal point estimation. *Appl. Statist.*, **60**, 93–107.
Lyons, R. (2011) The spread of evidence—poor medicine via flawed social-network analysis. *Statist. Polit. Poly*, **2**, 1–26.
Ogburn, E. L. (2018) *Challenges to Estimating Contagion Effects from Observational Data*, pp. 47–64. Cham: Springer.
Ogburn, E. L., Sofrygin, O., Diaz, I. and van der Laan, M. J. (2017) Causal inference for social network data. *Preprint arXiv:1705.08527*.

Ogburn, E. L. and VanderWeele, T. J. (2014) Causal diagrams for interference. *Statist. Sci.*, **29**, 559–578.
Ogburn, E. L. and VanderWeele, T. J. (2017) Vaccines, contagion, and social networks. *Ann. Appl. Statist.*, **11**, 919–948.
Pearl, J. (2000) *Causality: Models, Reasoning and Inference*. New York: Cambridge University Press.
Pearl, J. (2009) *Causality: Models, Reasoning, and Inference*, 2nd edn. New York: Cambridge University Press.
Peierls, R. (1936) On Ising's model of ferromagnetism. *Math. Proc. Camb. Phil. Soc.* **32**, 477–481.
Pena, J. M. (2019) Unifying Gaussian LWF and AMP chain graphs to model interference. *J. Causl Inf.*, **8**, no. 1.
Riggs, R. E. (1992) When every vote counts: 5-4 decisions in the United States Supreme Court, 1900-90. *Hofstra Law Rev.*, **21**, 667–724.
Robins, J. M. (1986) A new approach to causal inference in mortality studies with sustained exposure periods—application to control of the healthy worker survivor effect. *Math. Modlng* **7**, 1393–1512.
Rosenbaum, P. (2007) Interference between units in randomized experiments. *J. Am. Statist. Ass.*, **102**, 191–200.
Rosenquist, J. N., Murabito, J., Fowler, J. H. and Christakis, N. A. (2010) The spread of alcohol consumption behavior in a large social network. *Ann. Intern. Med.*, **152**, 426–433.
Rubin, D. (1990) On the application of probability theory to agricultural experiments: essay on principles, section 9, comment: Neyman (1923) and causal inference in experiments and observational studies. *Statist. Sci.*, **5**, 472–480.
Shalizi, C. and Thomas, A. (2011) Homophily and contagion are generically confounded in observational social network studies. *Sociol. Meth. Res.*, **40**, 211–239.
Sherman, E. and Shpitser, I. (2018) Identification and estimation of causal effects from dependent data. In *Advances in Neural Information Processing Systems 31*.
Shpitser, I., Tchetgen, E. T. and Andrews, R. (2017) Modeling interference via symmetric treatment decomposition. *Preprint arXiv:1709.01050*.
Sobel, M. (2006) What do randomized studies of housing mobility demonstrate? *J. Am. Statist. Ass.*, **101**, 1398–1407.
Songer, D. R. and Lindquist, S. A. (1996) Not the whole story: the impact of justices' values on Supreme Court decision making. *Am. J. Polit. Sci.*, **40**, 1049–1063.
Sunstein, C. R. (2014) Unanimity and disagreement on the supreme court. *Cornell Law Rev.*, **100**, 769–824.
Sznajd-Weron, K. and Sznajd, J. (2000) Opinion evolution in closed community. *Int. J. Mod. Phys.* C, **11**, 1157–1165.
Tate, C. N. (1981) Personal attribute models of the voting behaviour of US Supreme Court justices: liberalism in civil liberties and economics decisions, 1946–1978. *Am. Polit. Sci. Rev.*, **75**, 355–367.
Tchetgen Tchetgen, E. J., Fulcher, I. and Shpitser, I. (2017) Auto-g-computation of causal effects on a network. *Preprint arXiv:1709.01577*. University of Pennsylvania, Philadelphia.
Tchetgen Tchetgen, E. J. and VanderWeele, T. (2012) On causal inference in the presence of interference. *Statist. Meth. Med. Res.*, **21**, 55–75.
VanderWeele, T. (2010) Direct and indirect effects for neighborhood-based clustered and longitudinal data. *Sociol. Meth. Res.*, **38**, 515–544.
Verma, T. S. and Pearl, J. (1990) Equivalence and synthesis of causal models. *Technical Report R-150*. Department of Computer Science, University of California, Los Angeles.
Wan, Y.-W., Allen, G. I., Baker, Y., Yang, E., Ravikumar, P., Anderson, M. and Liu, Z. (2016) Xmrf: an R package to fit Markov networks to high-throughput genetics data. *BMC Syst. Biol.*, **10**, no. 3, article 69.
West, R., Paskov, H. S., Leskovec, J. and Potts, C. (2014) Exploiting social network structure for person-to-person sentiment analysis. *Trans. Ass. Computnl Ling.*, **2**, 297–310.

Supporting information

Additional 'supporting information' may be found in the on-line version of this article:

'Supporting material for Collective problem solving, causal inference, and chain graphs'.