Applied Network Science

RESEARCH Open Access

Estimating ideology and influence in social media networks: opinions on genome editing of livestock

Joseph Navelski[1,4*], Syed Badruddoza[2], Jill J. McCluskey[1] and Francis G. Pascual[3]

Electronic Supplementary Materials: This article contains supplementary materials including replication code. The supplementary materials can be found at the corresponding author's GitHub (https://github.com/JNavelski), and they are available to all users.

*Correspondence:
Joseph Navelski
joseph.navelski@wsu.edu

Full list of author information is available at the end of the article

Abstract

We develop a model to infer the ideologies and influence of agents using social media. The novelty of our model is that it only requires a subset of connections in the network structure, where most models require the entire network structure and detailed socio-demographic characteristics of its agents. We apply the model to Twitter data to evaluate the social network between experts in the field of genome editing in livestock (GEL) and their followers. The model generates an ideological position score for each expert and follower, and their relative social influence. Estimates show that followers opposed to genome editing have more influence than those in favor, e.g., anti-GEL followers own 69% of the social influence in a typical conversation. A post hoc analysis shows that the projected consensus realized through interaction and learning is opposed to GEL. Our model can be used to infer positions, perceptions, and influences of social media users for any topic, and the empirical results can help inform investment, marketing, and policymaking decisions within the livestock industry.

Keywords Latent variable MCMC estimation, Network formation, Opinion formation, Social influence, Social network analysis

JEL Classification D85, Q16, O33

Introduction

Social networks enable the process of social learning, where information is transmitted and agents form opinions through repeated interaction, a concept modeled in both Bayesian and non-Bayesian settings (Mobius and Rosenblat 2014). In a non-Bayesian framework, DeGroot (1974) introduced the concept of agents updating their beliefs through weighted averages, leading to learning convergence. This approach depends on the network's geometric structure and allows agents to reach consensus, revealing influence levels (DeMarzo et al. 2003; Golub and Jackson 2010; DeGroot 1974). Estimating a network's structure is complex, and existing models often lack compatibility with DeGroot's (1974) requirements (Chandrasekhar 2016; Penrose 2003; Erdös et al. 1960;

Graham 2017). Additionally, the data required is expensive and often difficult to access, further complicating the empirical network formation.

In this study, we propose an alternative model, assuming homophilic social networks—which means that agents have a higher probability of connecting when they are similar (McPherson et al. 2001; Currarini et al. 2009). We use a latent variable spatial following model to estimate the social network for DeGroot's (1974) learning model (Hoff 2003; Hoff et al. 2002; Barberá 2015; Barberá et al. 2015; Navelski and Pascual 2025; Bafumi et al. 2005; Rasch 1993). This model uncovers individual ideological positions and influence within the network without requiring detailed sociodemographic characteristics of individuals. Our model can be used to infer positions, perceptions, and influences of social media users for any topic, and the empirical results can help inform investment, marketing, and policymaking decisions within the livestock industry.

The motivation of this study is two-fold. First, modeling social networks has been gaining importance in recent years, as it provides critical insights into human interactions on social media, influencing areas ranging from marketing to political dynamics. Its application spans various fields, enabling more targeted strategies and informed decision-making, reflecting the intricate connections and influence patterns within communities and societies. Second, social media analysis provides ex-ante insights into advanced technologies that have not been implemented yet but can be implemented in the future. The raw perceptions, influence, and ideological positions of social media users can be a microcosm of future perceptions and acceptance of novel technologies, including food technologies.

Several studies have approached network modeling. DeGroot (1974) was one of the first to introduce a model of learning and social influence. A finite number of agents learn and interact with each other through a weighted and possibly directed network over time. Each agent is endowed with some initial belief about a common idea or thought, such as the probability of an event happening or the perceived level of quality in a new technology. Agents repeatedly discuss and share beliefs over time. An agent's updated belief is the weighted average of all agents' beliefs from the previous period, including their own. Over time, provided the social network is row-stochastic (i.e., all row entries sum to one for each agent) and strongly connected (i.e., there is a path from any agent to every other agent, even if it is indirect) the learning process will converge to a common belief (Jackson 2010; Golub and Jackson 2010). This convergence is the equivalent of reaching a consensus about the common idea or thought. When a consensus is reached, the relative social influence weight that each agent has in the learning process is realized. This implies that the social network's structure has an important effect on learning, drawing a consensus, and influence, and that this effect should be investigated empirically to support its theory. For readers unfamiliar with the DeGroot model, we provide a detailed review of its assumptions, consensus conditions, and influence vector calculation in the Appendix.

Economists have recently started to empirically investigate how the structure of social networks affects learning and convergence in the DeGroot (1974) model. A key feature of the DeGroot framework is that agents are boundedly rational: rather than updating beliefs optimally in a Bayesian sense, they simply form weighted averages of their neighbors' beliefs. This stylized assumption provides analytical tractability and has motivated a wide range of empirical and theoretical extensions. Chandrasekhar et al. (2020), for

instance, extend DeGroot by modeling agents as a mixture of DeGroot updaters and Bayesian optimizers, reflecting bounded rationality in heterogeneous agent behavior. The authors remark that departures from full information or variation in agent type still yield outcomes interpretable within a DeGroot-style learning framework. These methods estimate the social network using a random utility framework and a mixture of Penrose's (2003) and Erdös et al.'s (1960) random geometric graph methods to model the presence of "clans" within the network (Bramoullé et al. 2016).

Penrose's (2003) theory assumes individuals connect if the distance between their latent parameters is less than some radial threshold. They test their method using real-life social network data from two different settings and find that a sparser network increases the chances that agents become "stuck" in their learning process, leading to an asymptotic learning failure. This research makes an important contribution to the learning literature, but it does not focus on how individual characteristics can explain or change asymptotic learning in the network.

Banerjee et al. (2021) use methods similar to Chandrasekhar et al. (2020), proposing an extension of DeGroot's (1974) model where there is a mixture of informed and uninformed agents. Agents update beliefs by "naively" adopting the beliefs of informed agents and ignoring the beliefs of uninformed agents. They also demonstrate how beliefs and social influence change for agents in a sparse network using Penrose's (2003) and Erdös et al.'s (1960) random graph methods. They find that sparse social networks and signals (i.e., few connections and a small number of informed agents) can lead to a consensus where only the most-informed agents' beliefs are accepted. They call this "belief dictatorship." Their proposed methods provide insights as to why true initial signals become construed or altered over time but do not investigate how individual characteristics governing the formation of the social network affect learning and influence.

The empirical network-formation literature in the social-learning setting is limited. Chandrasekhar (2016) provides a summary of the utility-based network formation models where many have limitations in their application to DeGroot's (1974) model. For example, both Chandrasekhar et al. (2020) and Banerjee et al. (2021) use random geometric graphs from Erdös et al. (1960) and Penrose (2003), but neither of them focuses on individual (node) parameter estimates and how these estimates dictate a network's structure and asymptotic learning.

Another limitation in the utility-based models is that only a few models estimate the individual heterogeneous effects that contribute to the formation of a network. Graham (2017) establishes a base for how to estimate and identify the individual unobserved parameters that govern the formation of a network. He also characterizes the marginal effects of these parameters and provides details on how these parameters can alter the geometric structure of a network. One reason Graham's (2017) method has not yet been used in learning models is that it assumes agents do not build a connection with themselves. This means that agents do not weigh their own beliefs relative to others, which DeGroot (1974) requires to fulfil the aperiodicity condition that allows for a consensus to be reached.

Another reason the utility-based network formation models have been underexplored in the context of learning is that it is difficult to collect network data. Most early network models relied on data from surveys that ask agents to list their connections with other agents (Sampson 1968; Banerjee et al. 2013, 2021; Krackhardt 1987; Chandrasekhar et

al. 2020). While the rise of large-scale social media platforms has made digital trace data more available, such data often provide information only on observed links and not on detailed individual characteristics (e.g., gender, income, or attitudes). Moreover, academic researchers frequently face restrictions on accessing comprehensive social media datasets due to proprietary and privacy concerns. This helps explain why even recent work in the DeGroot tradition continues to use survey-based data collection—for example, Banerjee et al. (2021) in India and Chandrasekhar et al. (2020) in Mexico. Collecting survey data is also costly and often limited to small subsets of larger networks. A further challenge is that researchers must decide on the bounds of the network, a choice that can shape interpretation and alter research questions. For example, if the goal is to study how individuals learn and influence each other about a new technology, how can one define the relevant "network" in today's highly interconnected world?

A case researchers may face when working with network data is that they only observe individual links outside of a network. For example, researchers may only observe a partition of a network where n individuals follow m experts in a field in which they are interested. This implies data on the entire network does not exist, but there is enough data that can still provide insights. In particular, it can provide insights into what led to the decision to follow a particular expert, and how these individuals might learn and influence each other in their own social networks.

This study utilizes the above scenario and proposes an alternative way to estimate a social network proposed by DeGroot (1974). The main assumption is that social networks are "homophilic." Homophily is a term first developed by Lazarsfeld and Merton (1954), and the etymology of the term is self-love. It is well-documented that social networks tend to be homophilic, meaning that individuals tend to associate with others that have similar traits (McPherson et al. 2001). For example, high school friendships in the United States exhibit homophily in ethnicity, where 85% of white students tend to be friends with other white students, while 85% of black students tend to be friends with other black students (Currarini et al. 2009). These results exhibit similar patterns for students in Dutch high schools (Baerveldt et al. 2004), and these homophilic patterns also tend to be associated with age, race, gender, religion, and profession (McPherson et al. 2001).

With the assumption that networks are homophilic, we develop a learning model in which agents called "followers" build a social network among themselves (a follower-to-follower network) by developing affinity through experts they commonly follow. The experts represent a knowledgeable source of information for the followers on a particular topic. The more experts two followers follow, the stronger the affinity between these followers. The social network is used in a learning model where each follower's social influence is realized in the learning process, and a consensus is reached. We then estimate the unobserved individual parameters that explain the following behaviors between the agents who follow the experts.

The method we propose uncovers each follower's ideological position relative to the experts' positions using a latent-variable spatial-following model (Hoff et al. 2002; Hoff 2003; Barberá 2015; Barberá et al. 2015; Navelski and Pascual 2025). These estimates are used to predict the social network geometric structure and derive each follower's level of social influence in the network. We compare these estimates to see which ideological

positions have the most social influence in the network, and then use these estimates in a post hoc analysis to get a consensus on the topic the experts represent.

We apply this method to a Twitter, currently rebranded as X, dataset where we focus on agents comparing the expert accounts they follow in the field of genome editing in livestock (GEL). We find that individuals with anti-GEL ideologies have 69% of the social influence on Twitter indicating that any consensus reached will be heavily influenced by individuals that are against GEL. We conduct a post hoc analysis and find that the consensus on GEL is more anti-GEL on Twitter. To our knowledge, this is the first paper to show that individuals who are anti-GEL hold disproportionate influence on social media. While prior work has examined public sentiment toward gene editing in food and agriculture (Tabei et al. 2020; Middelveld and Macnaghten 2021; McCluskey et al. 2025), responsible innovation frameworks for livestock applications (Bruce & Bruce 2019), and public narratives in focus groups (Middelveld et al. 2023), none have quantified relative influence shares across pro- and anti-GEL actors in online networks. Implications are that pro-GEL opinions will be difficult to adopt in social media.

Model

Experts often amass a following due to their knowledge and beliefs about their field of work. For example, political elites attract voters that support their ideals, media sources target certain types of viewers, and academic researchers attract followers eager to align with their next groundbreaking discovery. This behavior can be seen as a network partition where n different agents (or nodes), also called followers, connect with or follow m different expert agents. This "following" behavior can provide insights into the underlying characteristics that describe each agent, and how the n followers interact and learn from each other in their own social networks.

We use this following behavior, which is captured in observed data, to construct a follower-to-follower network that can be used in the DeGroot (1974) model. The observed data is $n \times m$ follower to expert connections matrix $\mathbf{A}$, and we propose using the data and the information in the data to construct a follower-to-follower interaction matrix $\mathbf{T}$ by comparing shared expert connections across all pairs of followers. This allows us to derive a row-stochastic matrix $\mathbf{T}^*$ that can be used in DeGroot (1974) to uncover each follower's social influence when engaging about the topic the experts specialize in. We call this modeling process *A Social Network Based on Common Connections*.

We then enrich this modeling process by using the information in the $n \times m$ follower-to-expert connections matrix $\mathbf{A}$ with a statistical model to derive an estimated matrix $\widehat{\mathbf{A}}$. This estimation provides fitted probabilities $\widehat{a}_{ij}$ that incorporate individual heterogeneity in expert popularity $\widehat{\alpha}_j$, follower engagement $\widehat{\beta}_i$, and latent ideological positions of each expert $\widehat{\phi}_j$ and follower $\widehat{\theta}_i$. We then use $\widehat{\mathbf{A}}$ to construct the follower-to-follower interaction matrices $\widehat{\mathbf{T}}$ and $\widehat{\mathbf{T}}^*$ to use in DeGroot (1974). We call this modeling process *An Estimated Social Network Based on Common Connections*, and this process allows us to capture unobserved preferences, avoid the rigidity of purely binary data, and generate continuous measures of ideological proximity between followers and experts. For example, we use this process to compare follower ideology to their social influence in the learning process, and to use their ideology as an initial belief in the social learning process. We formally present both modeling processes in the next subsections, and we

provide a review of the DeGroot (1974) social learning model and some background on the theoretical origin of matrices **A** and **T** in the Appendix.

A social network based on common connections

Consider a social network (or graph) $\mathbf{A}_{n\times m}$ where there are n agents that choose to connect with m other agents. Matrix $\mathbf{A}_{n\times m}$, which takes the form

$$\mathbf{A}_{n\times m} = \begin{bmatrix} a_{11} & \cdots & a_{1j} & \cdots & a_{1m} \\ \vdots & \ddots & \vdots & \ddots & \vdots \\ a_{i1} & \cdots & a_{ij} & \cdots & a_{im} \\ \vdots & \ddots & \vdots & \ddots & \vdots \\ a_{n1} & \cdots & a_{nj} & \cdots & a_{nm} \end{bmatrix},$$

is a matrix of $n \times m$ dyadic links, and the links represent the directed one-way relationship between the *i*th agent in *n* choosing to link with the *j*th agent in *m*. All connections from the *i*th agent to the *j*th agent are observed, while no connections from any of the *j*th agents to the *i*th agents are observed.

Henceforth, **A** will be denoted as the "connections matrix," the *m* agents as "experts," and the *n* agents as "followers." Follower decisions to connect with the experts are observed, but none of the experts' connections are observed.[1] Experts are knowledgeable and informed about their specializations, and followers link with experts to gain information about their specializations because their positions or beliefs align (i.e., the connections are assumed to be homophilic).[2] The goal is to use the information in the connections matrix **A** to uncover a social network of followers **T** for the DeGroot (1974) social learning model.

Assume all followers form a social network by comparing each of the *m* experts they follow, and model this behavior as follows. The mathematical representation of this behavior is $\mathbf{T} = \mathbf{A}\cdot(\mathbf{A})^T$, where each element in this matrix is $t_{ik} = \sum_{j=1}^{m} a_{ij}a_{kj}$. All followers then reevaluate each interaction and weigh them relative to all other interactions $t^*_{ik} = \frac{t_{ik}}{\sum_{j=1}^{n} t_{ik}}$. This process yields a row-stochastic transition matrix

$$\mathbf{T}^*_{n\times n} = \begin{bmatrix} t^*_{11} & \cdots & t^*_{1k} & \cdots & t^*_{1n} \\ \vdots & \ddots & \vdots & \ddots & \vdots \\ t^*_{i1} & \cdots & t^*_{ik} & \cdots & t^*_{in} \\ \vdots & \ddots & \vdots & \ddots & \vdots \\ t^*_{n1} & \cdots & t^*_{nk} & \cdots & t^*_{nn} \end{bmatrix}$$

that satisfies the properties of DeGroot's (1974) interaction matrix.

All followers then learn via the DeGroot (1974) process. If $\mathbf{T}^*$ is *strongly connected*, a consensus is reached yielding a social influence weight s_i for each follower. The social influence weight s_i is calculated by solving for the left-hand eigen vector of the social interactions matrix. Intuitively, the *n* followers compare the experts they follow outside

[1] An example of this behavior is when users in social media choose to follow or not follow an account, while the other account's decision to follow or not follow the other users is not observed.

[2] McPherson et al. (2001) makes the well documented case that this behavior is true in practice for social networks, and this assumption has been applied in other empirical network formation models (Graham 2017; Chandrasekhar et al. 2020).

of the network, and this dictates the weight or trust they put on each other's beliefs. Followers then reevaluate each connection relative to all of their connections.

An estimated social network based on common connections

Consider the same connections matrix **A**, but one in which each dichotomous choice can be modeled with the logistic regression model, a specific form of a generalized linear mixed model given by

$$\Pr\left(a_{ij}=1|\mu,\gamma,\alpha_j,\beta_i,\theta_i,\phi_j\right)=\left[1+\exp\left(-\pi_{ij}\right)\right]^{-1} \tag{1}$$

where $\pi_{ij}=\mu+\alpha_j+\beta_i-\gamma(\theta_i-\phi_j)^2$. The intercept $\mu\in\mathbb{R}$ is a fixed effect, $\alpha_j\in\mathbb{R}$ and $\beta_j\in\mathbb{R}$ represent individual random effects observed in the connections matrix, $\phi_j\in\mathbb{R}$ and $\theta_j\in\mathbb{R}$ are latent variables representing spatial positions in the network structure, and γ is a weighting parameter.

The specification in Eq. 1 follows the specifications from Hoff et al. (2002), Barberá (2015), and Navelski and Pascual (2025), whose models assume links are formed based on how closely related their latent positions are in space. The random effects α_j and β_i are correlated with the observed individual effects that explain the connection between expert *j* and follower *i*. These parameters can be interpreted as expert *j*'s popularity and follower *i*'s engagement, respectively (Barberá 2015; Navelski and Pascual 2025).

Parameters ϕ_j and θ_i are assumed to be in a one-dimensional space, and the square distance specification – $\gamma(\theta_i-\phi_j)^2$ follows the "homophilic" assumption (Hoff 2003). These parameters are interpreted as "ideal points" (Curtis 2010; Rasch 1993; Poole and Rosenthal 2000; Bafumi et al. 2005).[3] An agent's "ideal point" is his or her preference or position within a spatial framework. We use Navelski and Pascual's (2025) method to map these latent positions on to an interpretative spectrum between – 1 and 1, where – 1 is the most extreme "anti" point of view, and 1 is the most "pro" point of view.

Data is used to fit the model and each predictive element is derived where

$$\widehat{a}_{ij}=\left[1+\exp\left(-\widehat{\pi}_{ij}\right)\right]^{-1}$$

and the initial social interaction rule is defined as
$\widehat{t}_{ik}=\sum_{j=1}^{m}\widehat{a}_{ij}\widehat{a}_{kj}$.

Each $\hat{a}_{ij}$ is an index on how similar follower *i* is to expert *j* and indicates the probability of follower *i* following expert *j*. $\hat{t}_{ik}$ defines a weighted relationship between followers *i* and *k*, and this weight increases when they are more connected and have similar following profiles. Similar to the base-case example, each follower *i* reevaluates their relationship with follower *k* relative to all others $\hat{t}^{*}_{ik}=\frac{\hat{t}_{ik}}{\sum_{j=1}^{n}\hat{t}_{ik}}$ to form a row vector corresponding to follower *i* in $\hat{\mathbf{T}}^{*}$. Followers then learn from each other based on $\hat{\mathbf{T}}^{*}$. Assuming $\hat{\mathbf{T}}^{*}$ is strongly connected, beliefs converge to a common belief, and a social-influence vector **s** is realized.

[3] This interpretation is further supported by the probabilistic voting model literature where agents vote for candidates based on how close their ideology or reputations align (Enelow and Hinich 1984; Coughlin and Nitzan 1981; Enelow and Hinich 1989; Coughlin 1992). This literature also proposes that the ideology parameters can be functions of many other parameters that explain patterns in ideology. For example, the ideological position of a candidate can be a function of an array of political positions on certain issues and/or it can be a function of other individual characteristics such as charisma or wealth.

Intuitively, this social-interaction matrix $\hat{\mathbf{T}}^*$ represents followers comparing their weighted indices $\hat{a}_{ij}$ and $\hat{a}_{kj}$ for each expert *j*. These indices are functions of the unobserved individual characteristics of each agent, including the latent variables whose distance dictates the probability of follower *i* linking with expert *j*. We map the latent variables onto a scale between – 1 and 1 that represents the relative position individuals have on the specialization the experts represent. Both expert and follower positions are mapped onto their own scale, giving an overall distribution of ideological positions on a topic.

Markov-Chain Monte Carlo (MCMC) estimation

The model parameters are unknown, and the statistical problem is to perform inference on $\alpha = (\alpha_1, \ldots, \alpha_m)'$, $\beta = (\beta_1, \ldots, \beta_n)'$, $\phi = (\phi_1, \ldots, \phi_m)'$, $\theta = (\theta_1, \ldots, \theta_n)'$, γ, and μ. Under the assumption of logical independence (i.e., individual following decisions are independent across all users *n* and *m* given the parameters), the likelihood function to maximize is given by

$$\Pr(\boldsymbol{\mu}, \boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\gamma}, \boldsymbol{\theta}, \boldsymbol{\phi} | \mathbf{A}) = \prod_{i=1}^{n} \prod_{j=1}^{m} \text{logit}^{-1}(\pi_{ij})^{a_{ij}} \left[1 - \text{logit}^{-1}(\pi_{ij})\right]^{1-a_{ij}} \quad (2)$$

where $\pi_{ij} = \mu + \alpha_j + \beta_i - \gamma(\theta_i - \phi_j)^2$ and $\text{logit}^{-1}(x) = [1 + \exp(-x)]^{-1}$ for $x \in \mathbb{R}$.

Without additional assumptions regarding the parameters, this model is not identifiable. For example, there are an infinite number of θ_i and ϕ_j combinations that will produce the same distance $(\theta_i - \phi_j)^2$. Even when the identifiability issues are addressed, the complexity of this equation makes direct estimation using maximum likelihood highly intractable because there is no analytical solution to the maximization problem. There are *m* α_j's, *m* ϕ_j's, *n* β_i's, *n* θ_i's, one intercept μ, and one weighting constant γ, implying that the total number of parameters to estimate is $2 \times (m + n) + 2$. Thus, maximum likelihood estimation becomes more difficult as datasets become larger.

To overcome the identification and tractability problems, we follow Navelski and Pascual's (2025) Bayesian estimation approach to generate samples from the posterior distribution where each of the parameters μ, α_j, β_i, γ, θ_i, and ϕ_j are assumed to be drawn from independent prior population distributions. We provide more details about the implementation of this identification and estimation method in the Appendix.

Data: genome editing in livestock (GEL)

Social media is a natural setting to study how individuals interact about the topic of GEL because it has network data on followers who choose to connect with, or "follow," an expert in the GEL field in which they are interested. We use data from Twitter because about 23% of U.S. adults say they use this social network, and Twitter allows academics to conduct research using information from their public accounts (Pew Research Center 2021; Twitter, 2022). We apply the model to a panel of experts on the GEL field and their most informed/active followers.[4]

[4] Barber´a (2015) and Navelski and Pascual (2025) also use Twitter data in their analyses namely, one dataset based on political elites and their followers in the former and one based on US media influencers and their followers in the latter. Merriam-Webster (2021) defines an "influencer" as "a person who is able to generate interest in something (such as a consumer product) by posting about it on social media" and a "follower" is an individual that follows that influencer because they have similar interests or ideologies.

Choosing the GEL experts and procuring the data from Twitter

We first define a list of GEL experts on Twitter. The list consists of academics, organizations, journalists, politicians, and companies who communicate about GEL. As a first step, we only include accounts that appear when searching for terms related to GEL on Twitter. The terms included *animal welfare, biotechnology, CRISPR, dairy, dehorning, gene editing, genetically modified, genome editing, genome engineering, GMO,* and *organic*. These terms were gathered from www.socialmention.com, which is an online software that identifies the key terms that are closely related to a topic people are posting about online. Next, to ensure that the accounts have at least a minimum level of impact, we limited our experts to accounts that had more than 1,000 followers and have a focus and position on GEL. The accounts should actively disseminate information about GEL and genome editing, in general, whether the information is positive or negative.

These criteria led to a final list of 46 experts in the field of GEL. Many of them have amassed followings based on their informativeness and position about GEL. For example, NonGMOProject and CRISPR News both have over 30,000 followers because they are seen to be valid sources of information about GEL and/or genome editing in general. We downloaded all of the accounts that follow each expert and merged the datasets by each follower's unique identification.[5] This produced a connections matrix **A** that has $n = 1{,}224{,}964$ followers.

The experts are initially labeled as being "anti-" or "pro-" GEL, based on the messages they disseminate, but this labeling is only used for preliminary analyses to motivate the data structure.[6] There are 12 anti- and 34 pro-GEL accounts. OrganicConsumer and NonGMOProject had the largest number of followers at 187,209 and 125,516, respectively, and Recombinetics has the smallest number of followers at 1,238. Table 1 presents that the average number of followers per anti-GEL account is greater than pro-GEL followers.

Data reduction

We reduce the dataset to include only "super followers," those who follow at least 9 out of the 46 accounts.[7] We perform this reduction to focus the analysis on the most-informed/active followers on the topic of GEL, reduce the amount of potential Twitter "bots" in the data, help with estimation tractability, and ensure matrix $\mathrm{T}^{*}_{n\times n}$ and $\hat{\mathrm{T}}^{*}_{n\times n}$ are strongly connected. This reduces the set of followers to $n = 3{,}383$ super followers, and this reduction slightly alters the research question to be focused on the more-informed or more-active followers or "super followers" regarding GEL. The number of super followers per account in the reduced connections matrix is presented in Table 2, and 3 presents

Table 1 Summary statistics of the experts' account characteristics by GEL viewpoint means and standard deviations (in parentheses)

Viewpoint	Followers	Following	Tweets
Anti-GEL	59,596 (51,190)	6,373 (8,381)	28,214 (31,220)
Pro-GEL	15,223 (15,536)	2,665 (3,274)	15,041 (18,700)

[5] We use Twitter's Academic Research's application programming interface (API) to obtain each expert's list of followers. The Twitter API query was conducted January 2022 (Twitter Inc., 2022a, b)

[6] This labeling will not affect future analyses as the estimation technique allows individuals to move on a continuous spectrum from − 1 to 1, which represents being the most anti-GEL to the most pro-GEL.

[7] We chose the following of 9 experts as the threshold because any dataset larger than this loses the ability to converge to the correct posterior distribution and we observe unfavorable convergence diagnostics. At the same time, we estimate the largest dataset we can so that we capture the underlying parameters, and thus the behaviors, of the larger more generalized population.

Table 2 Number of super followers per expert account

Twitter handle	Super followers	Position	Twitter handle	Super followers	Position
GMOEvidence	1184	Anti	Recombinetics	139	Pro
USRightToKnow	582	Anti	AzMilkProducers	62	Pro
JoelSalatin	330	Anti	CRISPRchef	474	Pro
nongmoreport	1795	Anti	AquaBountyFarms	131	Pro
GMWatch	1824	Anti	joeBondyDenomy	431	Pro
RachelsNews	959	Anti	FrancoiseBaylis	158	Pro
CFSTrueFood	1825	Anti	jcornlab	746	Pro
GMOFreeUSA	1811	Anti	AprilPawluk	514	Pro
OrganicTrade	1665	Anti	jsherkow	236	Pro
OrganicValley	1583	Anti	shsternberg	735	Pro
NonGMOProject	1978	Anti	JKamens	326	Pro
OrganicConsumer	1936	Anti	mem_somerville	402	Pro
TysonFoods	305	Pro	Synthego	708	Pro
Cargill	309	Pro	pcronald	523	Pro
Kevin_Faulconer	45	Pro	JonEntine	441	Pro
doudna_lab	1119	Pro	ELS_Genetics	119	Pro
CRISPR_News	1008	Pro	KevinADavis	659	Pro
SynBioBeta	873	Pro	BioBeef	665	Pro
AgBioWorld	675	Pro	igisci	881	Pro
pknoepfler	718	Pro	CamiDRyan	395	Pro
GeneticLiteracy	736	Pro	nmpf	231	Pro
CRISPRjournal	1038	Pro	pdhsu	814	Pro
GaetanBurgio	760	Pro	NPPC	196	Pro

Table 3 Experts' number of super followers by GEL viewpoint Means and Standard Deviations (in Parentheses)

Group	Super followers
Anti-GEL	1456 (558)
Pro-GEL	517 (302)

the anti-GEL and pro-GEL group averages where the anti-GEL accounts still have, as a group, more super followers (1,456) on average than the pro-GEL accounts (517).

The reduced dataset is the connections matrix **A** used in all subsequent analyses. Figure 1 depicts a heat-map of matrix **A**, where the columns are the 46 expert accounts, and the rows are the 3,383 followers. A black mark indicates users following expert accounts, which corresponds to 1's in the connections matrix. The white space indicates users not following and corresponds to 0's. To reveal patterns in the data, the first 12 columns of A are the anti-GEL experts, which are sorted in decreasing order by the number of followers they have. The subsequent 34 columns are the pro-GEL experts, which are sorted in increasing order by the number of followers they have. The followers are sorted with respect to the amount of anti-GEL accounts they follow, less the amount of pro-GEL accounts they follow. Intuitively, the first follower is the most anti-GEL in terms of following, while the last follower is the most pro-GEL.

Many of the followers in the northwest quadrant follow a large proportion of the anti-GEL accounts, indicated by the dark black mass. The followers in the southeast quadrant are sparser and not following a large proportion of the pro-GEL accounts, indicated by the patchy black and white area. This indicates anti-GEL account followers tend to concentrate on a group of anti-GEL experts, whereas pro-GEL account followers tend to follow fewer and varied pro-GEL experts in their following structure.

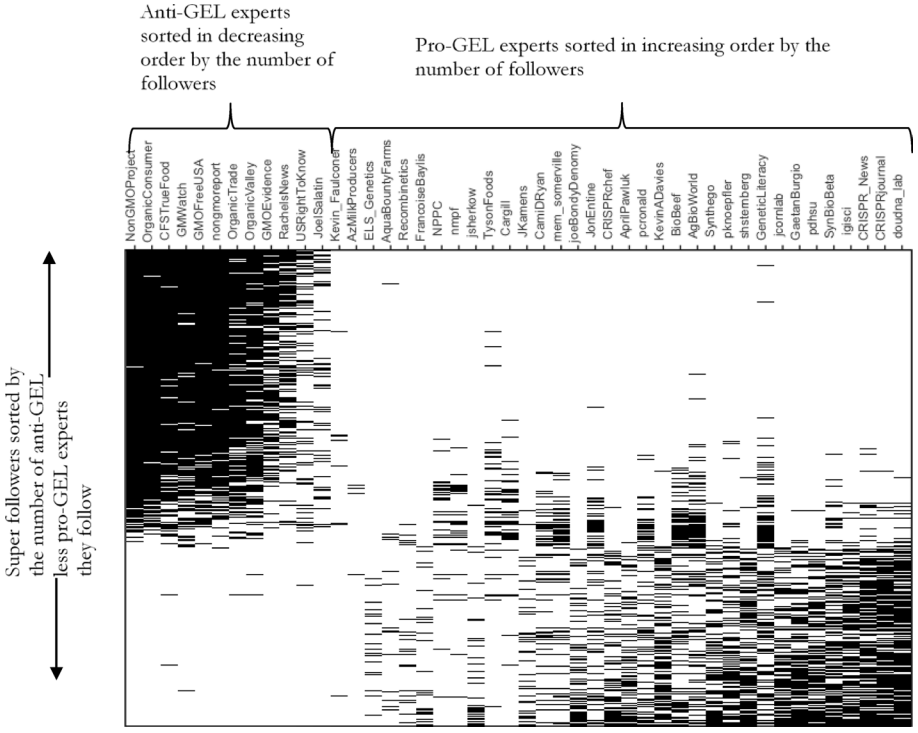

Fig. 1 Heat map of connections Matrix **A**

Estimating the data structure

Next, we estimate the connection matrix **A**. All MCMC (Bayesian) diagnostics yield expected results. All $\hat{R}$ values are less than 1.1, which is the standard recommendation in practice, implying that all chains have converged to the same posterior distribution, and thus, there is no divergence in the MCMC estimation process.[8] Convergence also implies the likelihood function is in the same form as Eq. 2, estimates are consistent, and hypothesis testing can be conducted. The MCMC diagnostics are supported by model fit diagnostics where, using all 155,618 individual decisions as observations in cross validation ($n \times m$), the prediction rate is 88.5% accurate. To further motivate estimation results, Fig. 2 shows a heat-map of the estimated probabilities of following, which closely resembles the raw data structure presented in Fig. 1.

Figures 1 and 2 are $\mathbf{A}_{n\times m}$ and $\hat{\mathbf{A}}_{\mathbf{n}\times\mathbf{m}}$ respectively, and they are used as the initial connections matrices to derive the weighted "trust" matrices $\mathrm{T}^{*}_{n\times n}$ and $\hat{\mathbf{T}}^{*}_{\mathbf{n}\times\mathbf{n}}$ used in the learning process. It is clear that individuals in the northwest and southeast quadrants have strong connections. These strong connections are a result of them having similar expert connections, which leads to high probabilities of interacting. We next use these matrices to derive the social influence vectors $\mathbf{s}_{1\times n}$.

Results

The estimation results show that @NonGMOProject is the most popular expert with an estimate of $\alpha_2 = 3.78$, and that @Kevin Faulconer is the least popular expert with an estimate of $\alpha_{15} = -3.30$. These results are expected since @NonGMOProject and @Kevin Faulconer have the most and least amount of super followers in the data, respectively.

[8] We provide a more detailed explanation about prediction diagnostics in Appendix.

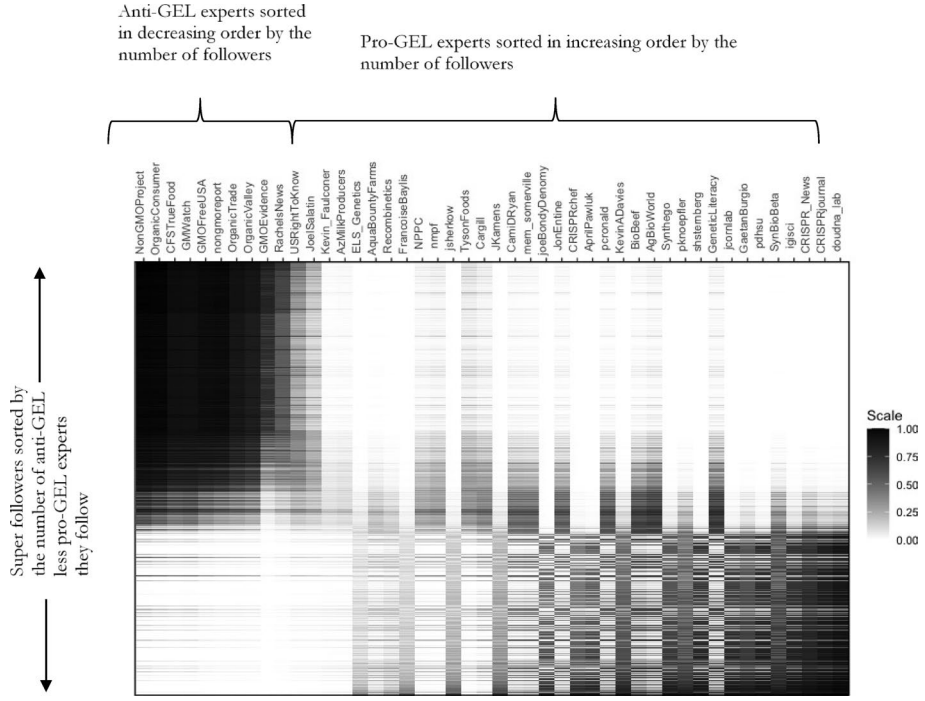

Fig. 2 Heat map of the estimated connections Matrix $\hat{A}$

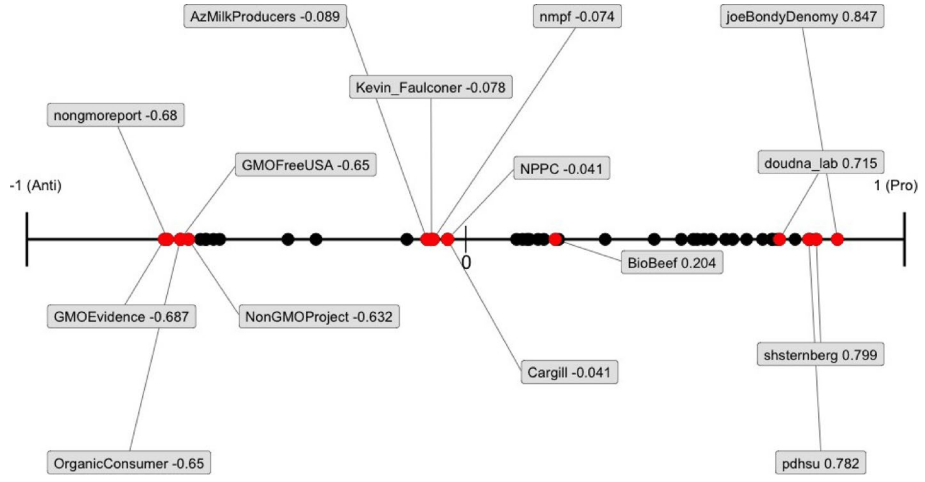

Fig. 3 Estimated ideology of experts for the GEL Data

The estimated ideal positions show that the most anti-GEL expert is @GMOEvidence ($\phi_j = -0.687$) and the most pro-GEL expert is @joeBondyDenomy ($\phi_j = 0.847$). Figure 3 plots the expert's estimated ideologies on a scale ranging from – 1 (anti) to 1 (pro) with a hash mark in the middle representing a zero line. We highlight some experts to show how their estimates align with the detailed information in their profiles.

As expected, results show many expert accounts lean anti-GEL, since most of their ideal points are closer to – 1 (the anti-GEL extreme). Additionally, @doudna_lab is the official Twitter account for Jennifer Doudna's lab. Dr. Doudna was awarded, with Dr. Emmanuelle Charpentier, the 2020 Nobel Prize in Chemistry for their methodological developments in genome editing. These developments were essentially, the first

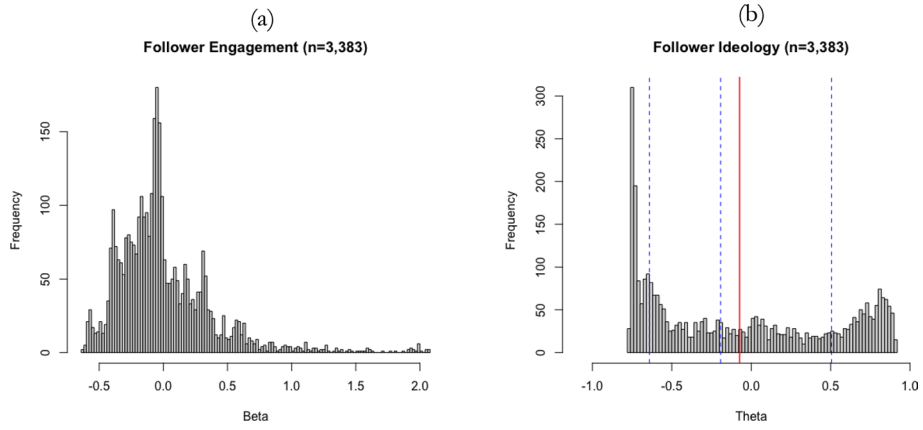

Fig. 4 Engagement $\hat{\beta}_i$ (a) and Ideology $\hat{\theta}_i$ (b) for the 3,383 super followers

Table 4 Summary statistics for the social influence vectors

Method	Min	1st Qtr	Median	Mean	SD	3rd Qtr	Max
Base-case	0.00871%	0.0203%	0.0328%	0.0295%	0.0093%	0.0381%	0.0577%
Estimated	0.0164%	0.0207%	0.0337%	0.0295%	0.008%	0.0364%	0.0479%

discovery of CRISPR, and it is not surprising that her lab's ideal point is positioned on the more pro-GEL side of the spectrum at 0.72.[9] One unexpected result is that @NPPC, @Cargill, and @TysonFoods are all representatives of the meat-producing industry and are expected to be more pro-GEL to reduce production costs, but they are seen to have ideologies that are more moderate since their ideal point estimates are closer to zero.

Figure 4(a) is a histogram of the estimated engagement parameters β_i for all 3,383 followers. It has a short "fat" tail on the negative side of the distribution and a long "thin" tail on the positive side indicating that overall engagement in GEL is high for some followers, but low for most. Figure 4(b) is a histogram of the 3,383 followers' ideal points θ_i where many individuals are polarized about GEL. The mean and median ideology estimates are – 0.074 and – 0.19, respectively, implying that the distribution is right-skewed and that the average informed follower about GEL will have an anti-GEL ideology. These metrics are represented by the solid (mean) and dotted (median) lines in the middle of Fig. 4(b), and this result is even more apparent when observing the large "spike" on the anti-GEL side. The zero-line is the theoretical center of the ideological distribution, and 56.73% of the followers are below this center line. This implies that 56.73% of the followers align more with the anti-GEL expert accounts.

Social influence

Table 4 shows that the social influence distribution is left-skewed when agents interact via the base-case social network and the estimated social network. The median percentage of social influence is 0.0328% and 0.0337% for the base case and estimated social network, respectively. These are both higher than the mean percentage at 0.0295%. This implies that there are more individuals with high social influence than those with low.

We compare the followers' estimated ideologies with their social influence estimates and summarize the results in Table 5. We find that individuals with anti-GEL ideology

[9] A summary and discussion about expert popularity and ideology is presented in the Appendix.

Table 5 Social Influence (SI) By Ideology θ

	$\theta_i < 0$	$\theta_i > 0$
SI	69%	31%

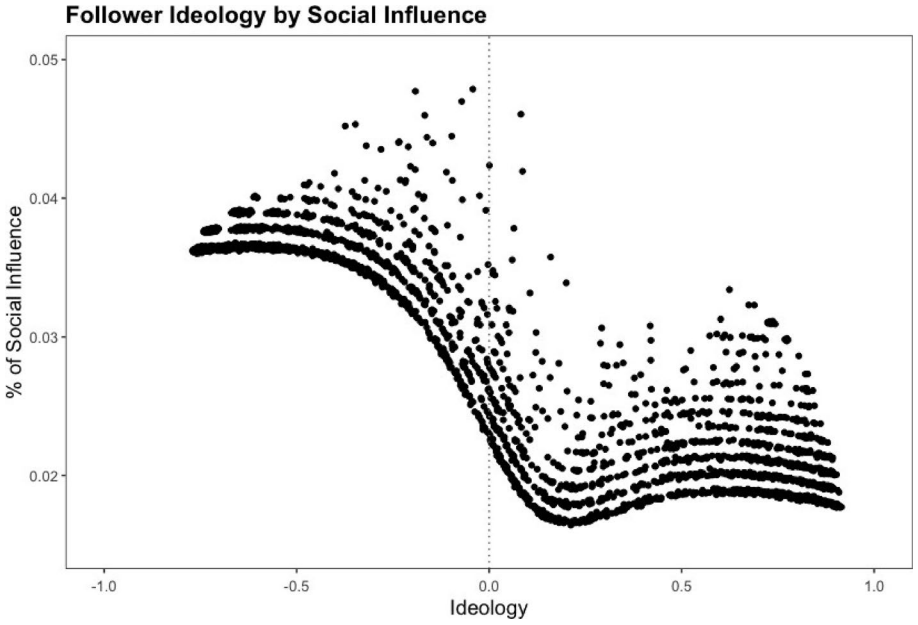

Fig. 5 Estimated Ideology of Followers θ_i by Social Influence s_i for the GEL Data

estimates have more social influence than individuals with positive ideology estimates at 69% and 31%, respectively. Furthermore, the individuals with more extreme ideologies, those with ideologies less than the first quartile and greater than the third quartile, have a similar pattern of influence. The more extreme anti-GEL individuals have 31% of the total influence while the more pro-GEL individuals have 18% of the social influence. This implies that the most extreme anti-GEL individuals have a little under a third of the total influence. Figure 5 plots follower ideology by their social influence and it is clear that anti-GEL followers have the majority of the influence in the learning process.

Post Hoc analysis: a consensus belief based on ideologies

One by-product of the social network estimation method is that agents' initial beliefs are uncovered in the form of ideologies. We standardize follower ideology estimates $\hat{\theta}_i$ to be on a scale of zero to one [0, 1], using $P_i^{(0)} = \frac{\hat{\theta}_i + 1}{2}$, instead of [– 1, 1], and agents use these estimates as endowed relative beliefs in the learning process. In this setting, a belief of zero is the least in favor of GEL (extremely anti-GEL), while a belief of one is the most in favor of GEL (extremely pro-GEL). In the context of GEL, we find that agents converge to a consensus belief of 0.395 using the equation $\mathrm{P}^{(\infty)} = \mathbf{sP}^{(0)}$ from DeGroot (1974). This indicates that the consensus belief on Twitter about GEL is anti-GEL, given that initial beliefs are standardized ideology estimates.

Conclusions

In this article, we developed an alternative method to estimate the structure and influence of a social network in a learning model. We assume that agents build connections based on their similarities (i.e., social networks are homophilic). Agents build connections by comparing the experts they follow in a particular field, and we estimate the underlying parameters that explain why followers link with experts. This estimation process uncovers the relative ideology of all experts and followers. Followers then learn in their own social network until convergence, and a social influence weight for each individual is realized. In a post hoc analysis, we derive the consensus belief of the followers assuming they use their ideological positions as initial beliefs in the learning process. Policy makers and companies can use the proposed model with large datasets to target agents who have the most influence in a social network and to align with their viewpoint.

We applied this method to a social media data set from Twitter with experts in genome editing in livestock (GEL) and their most informed/active followers. The main results are that 56.73% of the followers have ideologies that align with anti-GEL experts and that these followers own 69% of the social influence on Twitter. The post hoc analysis shows that the consensus on GEL is 0.395 on a scale where zero is the most anti-GEL and one is the most pro-GEL. To our knowledge, this is the first paper to show that individuals who are anti-GEL hold disproportionate influence on social media.

These results imply that any initial belief these followers receive will lead to a social learning consensus that is heavily influenced by anti-GEL followers. This means that anti-GEL will dominate the conversation on Twitter, and this could make it difficult for positive opinions about GEL to be accepted. The post hoc analysis supports this intuition where the consensus on Twitter is not in favor of GEL when ideologies are used as initial beliefs. Policy makers who promote and educate the public about new genome editing technologies need to realize their efforts could be squandered by individuals who perceive these technologies as something negative. The public perception of GEL on social media is negative, and policy makers should target anti-GEL followers with high influence in an attempt to change their position and message on GEL.

The broader implication of our findings is that new food technologies, particularly those related to genome editing in livestock, are viewed skeptically on social media platforms such as Twitter. The dominance of negative influence makes it challenging for positive opinions about GEL to gain traction and be accepted within the social media landscape. It also implies that efforts to promote and educate the public about new genome editing technologies might be hindered by this widespread negative perception, necessitating targeted strategies to reach and possibly sway those with significant influence who hold skeptical or negative views on GEL.

Appendix

A review of DeGroot's 1974 social learning model

The model considers a finite set of $N = \{1, \ldots, n\}$ agents that interact through links in a social network. The social network is defined by an $n \times n$ non-negative row-stochastic interaction matrix $\mathbf{T}_{n\times n}$ where each element t_{ik} represents the weight or trust agent

i puts on the belief (or opinion) of agent k.[10] Matrix $\mathbf{T}$ does not need to be symmetric implying that agent i can put a high weight on agent k's belief, but agent k need not put a high weight on agent i's belief.[11]

Each agent is endowed with some initial subjective belief $p_i^{(0)} \in [0,1]$ at time $t = 0$, and the vector of all n initial beliefs is represented by $\mathbf{p}_{n\times 1}^{(0)}$. Beliefs can represent things like the perceived level of product quality or the probability that a given statement is true. Agent i's belief at time t is $p_i^{(t)} \in [0,1]$, and the vector of all n beliefs is $\mathbf{p}_{n\times 1}^{(t)}$. The belief updating rule is $\mathbf{p}^{(t)} = \mathbf{T}\mathbf{p}^{(t-1)}$ which implies

$$\mathbf{p}^{(t)} = \mathbf{T}^t\mathbf{p}^{(0)}$$

where,

$$\mathbf{T}_{n\times n} = \begin{bmatrix} t_{11} & \cdots & t_{1n} \\ \vdots & \ddots & \vdots \\ t_{n1} & \cdots & t_{nn} \end{bmatrix} \text{ and } \mathbf{p}_{n\times 1}^{(t)} = \begin{bmatrix} p_1^{(t)} \\ \vdots \\ p_n^{(t)} \end{bmatrix}.$$

Intuitively, each agent's belief at time t is the weighted average of all agent's beliefs including their own $p_i^{(t)} = \sum_{k=1}^{n} t_{ik} p_k^{(t-1)}$. The interaction process continuously updates and reaches a consensus if and only if $\mathbf{T}$ is convergent.

A matrix $\mathbf{T}$ is convergent if it is *row stochastic* and *strongly connected* (Golub and Jackson 2010; Jackson 2010). A matrix $\mathbf{T}$ is *row stochastic* if all rows in the matrix sum to 1, and it is *strongly connected* if there is a path from any node i to every other node k, even if it is indirect.[12] Intuitively, one can think of a strongly connected network as a network where there are no partitions that are completely isolated from the other partitions of the network.

These two properties guarantee a consensus is reached where $\mathbf{p}_{n\times t}^{(\infty)} = \lim_{t\to\infty} \mathbf{T}^t\mathbf{p}^{(0)}$ for *any initial* vector $\mathbf{p}^{(0)}$. This implies that for any initial belief vector $\mathbf{p}^{(0)}$, the learning process will reach a consensus where all beliefs in the limit converge to a common and constant belief where each element in $\mathbf{p}^{(\infty)}$ is the same $p_1^{(\infty)} = \cdots = p_n^{(\infty)}$.

Reaching a consensus also implies there is a unique left-hand unit eigenvector $\mathbf{s}_{1\times n}$ of $\mathbf{T}$ that solves the limiting equation of $\lim_{t\to\infty} \mathbf{T}^t\mathbf{p}^{(0)} = \mathbf{s}\mathbf{p}^{(0)}$. Each element s_i in vector $\mathbf{s}_{1\times n}$ represents the amount of influence each agent has on the learning process. All elements sum to one $\sum_{i=1}^{n} s_i = 1$, and $\mathbf{s}_{1\times n}$ can be used to calculate the limiting beliefs in a consensus of $p^{(\infty)} = \mathbf{s}\mathbf{p}^{(0)} = \sum_{i=1}^{n} s_i p_i^{(0)}$ for any vector of initial beliefs (DeGroot 1974; DeMarzo et al. 2003; Golub and Jackson 2010; Jackson 2010). This implies that the struc-

[10] This social network can be represented as a graph g where agents are nodes and the links are edges. The $\mathbf{T}_{n\times n}$ matrix is the graphical representation of the social network in matrix form, and this graph can be weighted or unweighted, or it can be directed or undirected.

[11] DeMarzo et al. (2003) refer to $\mathbf{T}_{n\times n}$ as the "listening" matrix where each element t_{ik} represents how much agent listens to agent k's opinion, Golub and Jackson (2010) refer to t_{ik} as how much precision agent puts k's opinion, Jadbabaie et al. (2012) refer to $\mathbf{T}_{n\times n}$ as the social interaction matrix where t_{ik} represents the "influence" or "persuasion power" agent gets from agent k, and DeGroot (1974) and Jackson (2010) refers to $\mathbf{T}_{n\times n}$ as the "weight" or "trust" matrix where t_{ik} represents the weight or trust the ith agent has on the current belief of agent k in forming its own belief for the next period. In this paper, we will be referring to t_{ik} as the "trust" agent puts on k's opinion.

[12] We verify that all $\mathbf{T}_{n\times n}$ matrices are *strongly connected* in this paper using a Depth First Search (DFS) algorithm (Csardi and Nepusz 2006). The Depth First Search (DFS) algorithm checks to see if any node in a matrix can be reached starting from every other node in the matrix.

ture of social networks has an important effect on learning, drawing a consensus, and influence in the DeGroot (1974) model.

Theoretical origin of the connections matrix $\mathbf{A}_{n\times m}$ and interaction matrix $\mathbf{T}_{n\times n}$

In our analysis, we call $\mathbf{A}_{n\times m}$ the "connections matrix" and $\mathbf{T}_{n\times n}$ the "interaction matrix" or "social interaction matrix." In our model, we use the following structure of $\mathbf{A}$ to derive $\mathbf{T}$ but these matrices also have theoretical origins in a larger social network. For example, consider a social network (or graph) g where there are $n+m=N$ agents (or nodes), and $(n+m)\times(n+m)$ edges (or links). The social network g can be represented as an adjacency matrix

$$\mathbf{G}_{N\times N}=\begin{bmatrix}\mathbf{U}_{m\times m} & \mathbf{W}_{m\times n}\\ \mathbf{A}_{n\times m} & \mathbf{T}_{n\times n}\end{bmatrix}$$

where $\mathbf{U}_{m\times m}$, $\mathbf{W}_{m\times n}$, $\mathbf{A}_{n\times m}$ and $\mathbf{T}_{n\times n}$, are block matrices that partition $\mathbf{G}_{N\times N}$. All of these matrices can be weighted or unweighted, but our proposed model only needs to observe the connections matrix $\mathbf{A}$. The links in $\mathbf{A}$ represent the directed one-way relationship between the ith agent in n choosing to link with the j^{th} agent in m.[13] We call the m agents "experts" and the n agents as "followers" in our analysis for simplicity and because it closely resembles the following structure of followers to experts and/or influencers in social media.

Identification strategy

The model in Eq. (1) is still unidentified due to "additive aliasing," "multiplicative aliasing" and "reflection invariance," since there are an infinite number of combinations between the parameters that will give the same probability of following. An example of additive aliasing is $\mu=0$, $\alpha_j=-1$, $\beta_i=1$, $\phi_j=1$, $\theta_i=-1$, which gives the same probability of $\mu=0$, $\alpha_j=1$, $\beta_i=-1$, $\phi_j=-1$, $\theta_i=1$. An example of multiplicative aliasing and reflection invariance is multiplying the distance $\gamma(\theta_i-\phi_j)^2$ by any constant k where γ will absorb part of the constant $\frac{\gamma}{k}(\theta_i-\phi_j)^2$ (Barberá 2015). Reflection invariance occurs when the constant k has a different sign, and the probability is the same. Reflection invariance is usually a problem when the model specification includes some sort of "distance" specification between parameters (i.e., the absolute value between $\gamma(\theta_i-\phi_j)^2$). These problems are usually solved by restricting one of the jth or ith parameters in each parameter set but becomes difficult to do when working with the distance between two latent parameters. Navelski and Pascual (2025) suggest an alternative identification strategy where all priors are treated equally for ϕ_j and θ_i.

Additive aliasing

To address the additive aliasing issue, we transform all α_j's and β_is by subtracting the average of all the parameters from each parameter $\alpha_j^*=\alpha_j-\overline{\alpha}$ and $\beta_{i^*}=\beta_i-\overline{\beta}$, where $\overline{\alpha}=\sum_{j=1}^{m}\frac{\alpha_j}{m}$ and $\overline{\beta}=\sum_{j=1}^{n}\frac{\beta_j}{n}$. μ and γ face a similar issue, which is easily fixed by combining these two terms. This is a well-known method used when estimating two-way fixed effects models, and we account for $\overline{\alpha}$ and $\overline{\beta}$ by combining them with the constant $\mu^*=\mu+\overline{\alpha}+\overline{\beta}+\gamma$ (Bafumi et al. 2005; Gelman et al. 2004).

[13] If $\mathbf{W}_{m\times n}=(\mathbf{A}_{n\times m})^T$, then $\mathbf{A}_{n\times m}$ can be seen as an undirected network, but for this application we only consider $\mathbf{A}_{n\times m}$ as a directed partition of $\mathbf{G}_{N\times N}$.

Multiplicative aliasing

We address the multiplicative aliasing identification issue by setting $\phi_j^* = \frac{(\phi_j - \overline{\phi})}{s_\phi}$ and $\theta_j^* = \frac{(\theta_j - \overline{\phi})}{s_\phi}$ for all $j = 1, \ldots, m$ and $i = 1, \ldots, n$ in the model. s_ϕ is the sample standard deviation for $\phi_1, \ldots, \phi_m$. This method is explained in detail by Bafumi et al. (2005), and it standardizes the latent parameters relative to $\overline{\phi}$. This prevents the latent parameter from converging to the incorrect location and spread of their distribution. This also implies the transformed parameter $\gamma^* = \gamma s_\theta^2$.

Reflection invariance

The last identification issue is called reflection invariance, which is when the latent parameters tend to "flip" and converge to the wrong side of their distribution. A traditional method to combat reflection invariance is to "anchor" two parameters on the opposite side of the spectrum, anticipating they will converge to that point (i.e., let one of the $\phi_j^{anti} = -1$ and one of the $\phi_j^{pro} = 1$). It is also recommended to "pre-code" the initial values and priors to the parameters so that they operate on the side of the spectrum that the parameters are anticipated to converge to. We use the latter method to help with identifiability.

We employ a "soft constraint" that "guides" the parameters to converge to the correct side of the spectrum. We define a hyperparameter ϕ_{cut} that acts as an upper bound for the Anti-GEL Experts $\phi_j^{Anti} \in (-1, \phi_{cut})$ and a lower bound for the Pro-GEL Experts $\phi_j^{Pro} \in (-\phi_{cut}, 1)$. The estimation is then conducted, and we evaluate if any of the $\widehat{\phi}_j$ estimates are against the bound. If any of them are, then we increase the ϕ_{cut} value until none of the $\widehat{\phi}_j$ estimates are against the bound. We only need to apply this soft constraint to the ϕ_j's because once they are identified, the distance specification between the latent parameters then guides the θ_i's to the appropriate intuitive scale.

We call this method a "soft constraint" because the ϕ_{cut} hyperparameter still allows latent parameters to "switch" to the other side of the spectrum if that is the true location of the latent parameter. For example, in the GEL application, the @AzMilkProducers expert was originally classified as pro-GEL and was given an initial value of 0.8 to initiate the MCMC estimation but the mean of their latent posterior distribution converged to – 0.089, which is more anti-GEL leaning than pro-GEL. Because we specify $\phi_{cut} = 0.6$, it is allowed to cross over to the other side of the spectrum without forcing it to be positive and without hitting the $-\phi_{cut}$ constraint value.

The identified model

The identified model with the adjusted parameters used in the estimation process is defined as

$$\Pr(A_{ij} = 1 | \mu, \alpha_j, \beta_i, \gamma, \theta_i) = [1 + \exp(-\pi_{ij}^*)]^{-1} \tag{A1}$$

where $\pi_{ij}^* = \mu^* + \alpha_j^* + \beta_i^* - \gamma^*(\theta_i^* - \phi_j^*)^2$. This model is specified in the estimation process as a transformation where both the original and adjusted parameters are estimated at the same time. We use the original parameters for interpretation because the transformation is a one-to-one transformation between the original and adjusted parameters, implying that both sets of parameters are identified.

Summary of estimation results

Figure 6(a) and (b) show the distribution of the posterior means of expert popularity and follower engagement estimates, respectively. Figure 7 show the distribution of the posterior means for the expert and follower ideology estimates. Expert popularity is centered at zero (i.e., E $\alpha_j = 0$), and even though the distribution appears symmetric overall, it is clear that the more popular experts, experts with values greater than zero, are more "intensely" popular than those on the negative side. This indicates that a popular expert has more of an effect on a follower's decision to follow than an unpopular expert since the change in the probability of following has a greater increase for a popular expert than a decrease in an unpopular expert. This result is motivated in Table 6 where @NonGMOProject and @OrganicConsumer both have popularity estimates of 3.78 and 3.69, respectively, and @Kevin Faulconer and @AzMilkProducers have popularity estimates of – 3.30 and – 3.07, respectively.

Ideologies of the experts and followers exhibit opposite distributional patterns. The negative side of the distributions in Figs. 7(a) and (b) represents those that are anti-GEL, and the positive side represents those that are pro-GEL. For clarity, an ideology value of – 1 indicates the most extreme anti-GEL ideology, while a value of 1 indicates the most extreme pro-GEL ideology. Both the ideology of the experts and followers tend to be polarized since

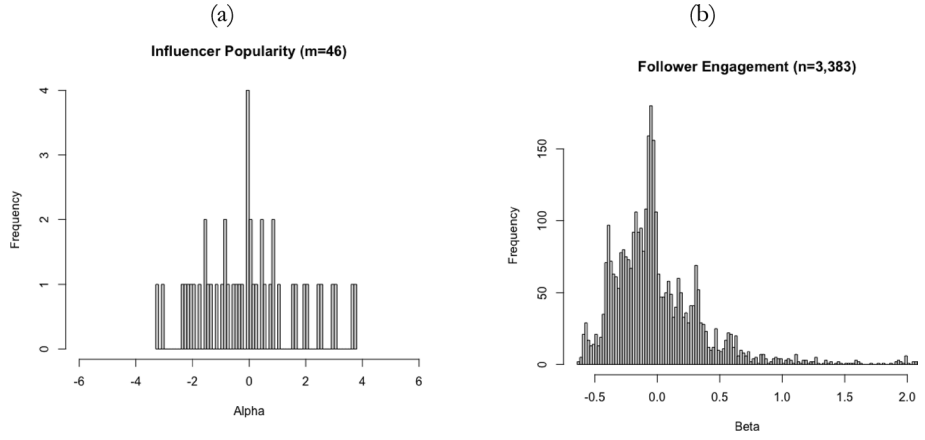

Fig. 6 Popularity for Experts $\hat{\alpha}_j$ (a) and Engagement of Followers $\hat{\beta}_i$ (b)

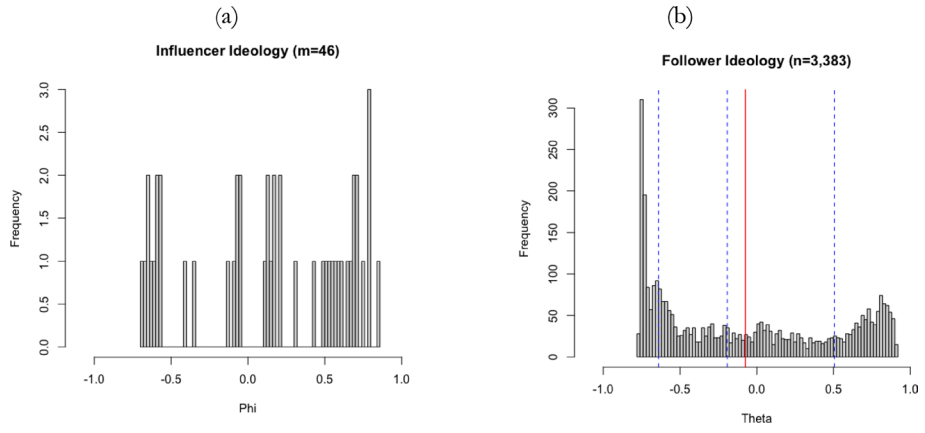

Fig. 7 Ideology for Experts $\hat{\phi}_j$ (a) and Followers $\hat{\theta}_i$ (b)

Table 6 Examples of popularity and ideology extremes

Name—Initializing View	$\hat{R}$	Mean	SD	2.5%	97.5%
Most popular					
@NonGMOProject—Anti	1.00	3.78	0.16	3.49	4.11
@OrganicConsumer—Anti	1.00	3.69	0.16	3.38	4.02
Least popular					
@Kevin Faulconer—Pro	1.00	− 3.30	0.12	− 3.51	− 3.06
@AzMilkProducers—Pro	1.00	− 3.07	0.12	− 3.29	− 2.84
Most extreme anti-GEL ideology					
@GMOEvidence—Anti	1.03	− 0.69	0.02	− 0.72	− 0.66
@nongmoreport—Anti	1.01	− 0.68	0.02	− 0.71	− 0.65
@OrganicConsumer—Anti	1.01	− 0.65	0.02	− 0.68	− 0.62
@GMOFreeUSA—Anti	1.03	− 0.65	0.01	− 0.68	− 0.62
Most extreme pro-GEL IDEOLOGY					
@joeBondyDenomy—Pro	1.03	0.85	0.03	0.79	0.91
@shsternberg—Pro	1.00	0.80	0.02	0.77	0.84
@pdhsu—Pro	1.01	0.78	0.02	0.75	0.81
@AprilPawluk—Pro	1.01	0.78	0.02	0.74	0.83
Moderate GEL Ideology					
@NPPC—Pro	1.01	− 0.04	0.02	− 0.08	− 0.00
@Cargill—Pro	1.01	− 0.04	0.02	− 0.07	− 0.01
@nmpf—Pro	1.01	− 0.07	0.02	− 0.11	− 0.04
@Kevin Faulconer—Pro	1.00	− 0.08	0.04	− 0.15	0.00

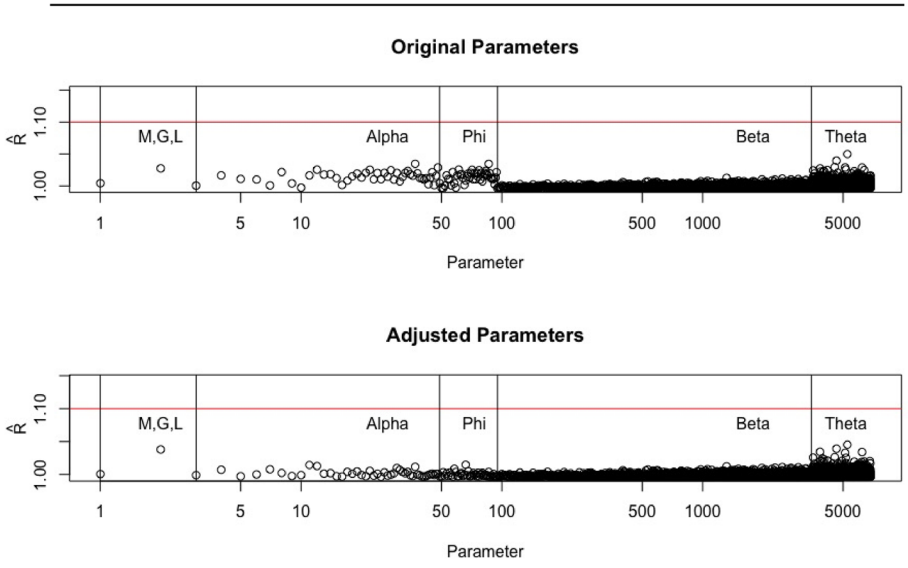

Fig. 8 $\widehat{R}$ Plot of All 6,860 Parameters (MCMC Convergence Diagnostics—Original vs. Adjusted)

a large majority of the estimates are concentrated at the end of the spectrum (− 1, 1). The experts with the most extreme ideologies are presented in Table 6 where @GMOEvidence, @nongmoreport, @OrganicConsumer, and @GMOFreeUSA all have the lowest ideal points at − 0.69, − 0.68, − 0.65, and − 0.65, respectively, and @joeBondyDenomy, @shsternberg and @pdhsu, and @April-Pawluk have the highest ideal points at 0.85, 0.80, 0.78 and 0.78, respectively. The polarization between experts' ideology is interesting because the anti-GEL accounts are very extreme while the pro-GEL accounts range from extreme to moderate. For example, @NPPC, @Cargill, @nmpf, and @Kevin Faulconer were all ini-

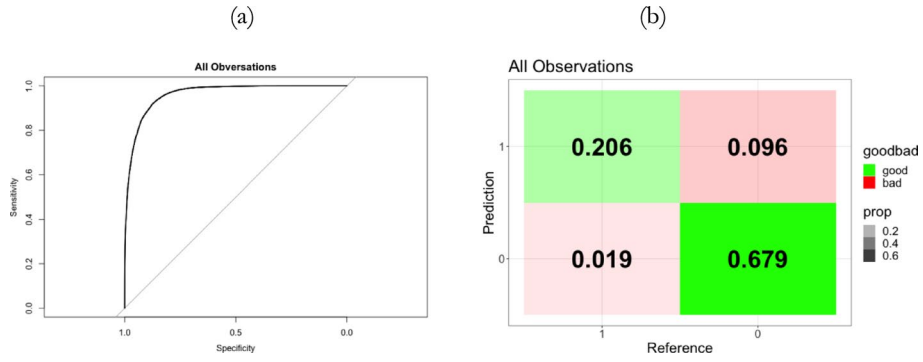

Fig. 9 ROC Curve (**a**) and Confusion Matrix (**b**) for All Observations

tially considered experts on the pro-GEL side, but in reality, they have ideologies that are more moderate at – 0.04, – 0.04, – 0.07 and – 0.08, respectively. These point estimates are presented in Table 6, and this is an interesting result because these experts initially started as pro-GEL experts while their ideological estimates are moderate to anti-moderate GEL. This type of result could imply that these accounts have more moderate ideologies.

Estimation diagnostics

Gelman and Rubin (1992) recommend an $\widehat{R}$ statistic at 1.1, implying that there are no divergent transitions in the estimation process, and this is the benchmark most researchers follow in practice. To support this intuition, Fig. 8 plots all $\widehat{R}$ values, showing that all values are below the 1.1 line. The optimal classification threshold was derived by maximizing the area under the receiver operating characteristic (ROC) curve (i.e., maximizing the sensitivity and specificity of the prediction diagnostics), and Fig. 9(a) and (b) show the ROC curve and confusion matrix, respectively.

Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1007/s41109-025-00759-y.

Additional file 1.

Author contributions

J. N. conceived and designed the analysis, collected the data, contributed to the development of the data analysis and tools, performed the analysis, produced the tables and figures, and wrote the paper. S. B. conceived and designed the analysis, contributed to the development of the data analysis and tools, produced the tables and figures, and wrote the paper. J. M. conceived and designed the analysis, contributed to the development of the data analysis and tools, produced the tables and figures, and wrote the paper. F. P. conceived and designed the analysis, collected the data, contributed to the development of the data analysis and tools, performed the analysis, produced the tables and figures, and wrote the paper.

Funding

This research was supported in part by the U.S. Department of Agriculture, National Institute of Food and Agriculture, Grant: 2020–67023-31637. The views expressed here cannot be attributed to USDA-NIFA. The findings and conclusions in this paper are those of the authors and should not be construed to represent any official USDA or U.S. Government determination or policy.

Declarations

Competing interests

All authors have no competing interests to disclose.

Author details

[1]School of Economic Sciences, Washington State University, Pullman, WA 99164, USA
[2]Department of Agricultural and Applied Economics, Texas Tech University, Lubbock, TX 79409, USA
[3]Department of Mathematics and Statistics, Washington State University, Pullman, WA 99164, USA
[4]Present address: 415 W 8 St, Unit F, Charlotte, NC 28202, USA

Received: 12 June 2025 / Accepted: 10 November 2025
Published online: 25 March 2026

References

Acemoglu D, Ozdaglar A (2011) Opinion dynamics and learning in social networks. Dyn Games Appl 1(1):3–49
Baerveldt C, Van Duijn MA, Vermeij L, Van Hemert DA (2004) Ethnic boundaries and personal choice. Assessing the influence of individual inclinations to choose intra-ethnic relationships on pupils' networks. Soc Netw 26(1):55–74
Bafumi J, Gelman A, Park DK, Kaplan N (2005) Practical issues in implementing and understanding Bayesian ideal point estimation. Polit Anal 13(2):171–187
Banerjee A, Breza E, Chandrasekhar AG, Mobius M (2021) Naive learning with uninformed agents. Am Econ Rev 111(11):3540–3574
Banerjee A, Chandrasekhar AG, Duflo E, Jackson MO (2013) The diffusion of microfinance. Science 341(6144):1236498
Banerjee AV (1992) A simple model of herd behavior. Q J Econ 107(3):797–817
Barberá P (2015) Birds of the Same Feather Tweet Together: Bayesian Ideal Point EstimationUsing Twitter Data. Polit Anal 23:76–91
Barberá P, Jost JT, Nagler J, Tucker JA, Bonneau R (2015) Tweeting from left toright: Is online political communication more than an echo chamber? Psychol Sci 26(10):1531–1542. PMID: 26297377
Bikhchandani S, Hirshleifer D, Welch I (1992) A theory of fads, fashion, custom, and cultural change as informational cascades. J Polit Econ 100(5):992–1026
Bird S, Klein E, Loper E (2009) Natural language processing with Python: analyzing text with the natural language toolkit. O'Reilly Media, Inc
Bramoullé Y, Galeotti A, Rogers BW (2016) Oxford handbook of the economics of networks. Oxford University Press
Bruce A, Bruce D (2019) Genome editing and responsible innovation, can they be reconciled? J Agric Environ Ethics 32(5):769–788
Chandrasekhar A (2016) Econometrics of network formation. Oxford Handbook of the Economics of Networks pp 303–357
Chandrasekhar AG, Larreguy H, Xandri JP (2020) Testing models of social learning on networks: evidence from two experiments. Econometrica 88(1):1–32
Coughlin PJ (1992) Probabilistic voting theory. Cambridge University Press
Coughlin P, Nitzan S (1981) Electoral outcomes with probabilistic voting and Nash social welfare maxima. J Public Econ 15(1):113–121
Csardi G, Nepusz T (2006) The igraph software package for complex network research. InterJournal Complex Syst 1695
Currarini S, Jackson MO, Pin P (2009) An economic model of friendship: homophily, minorities, and segregation. Econometrica 77(4):1003–1045
Curtis SM (2010) Bugs code for item response theory. J Stat Softw Code Snippets 36(1):1–34
Dauphin YN, Pascanu R, Gulcehre C, Cho K, Ganguli S, Bengio Y (2014) Identifying and attacking the saddle point problem in high-dimensional non-convex optimization. Adv Neural Inf Process Syst 27
DeGroot MH (1974) Reaching a consensus. J Am Stat Assoc 69(345):118–121
DeMarzo PM, Vayanos D, Zwiebel J (2003) Persuasion bias, social influence, and unidimensional opinions. Q J Econ 118(3):909–968
Enelow JM, Hinich MJ (1984) The spatial theory of voting: an introduction. Cambridge University Press Archive
Enelow JM, Hinich MJ (1989) A general probabilistic spatial theory of elections. Public Choice 61(2):101–113
Erdös P, Rényi A et al (1960) On the evolution of random graphs. Publ Math Inst Hung Acad Sci 5(1):17–60
Gelman A, Carlin JB, Stern HS, Rubin DB (2004) Bayesian data analysis, 2nd edn. Chapman and Hall/CRC
Gelman A, Rubin DB (1992) Inference from iterative simulation using multiple sequences. Stat Sci 7(4):457–472
Glass CA, Glass DH (2021) Social influence of competing groups and leaders in opinion dynamics. Comput Econ 58(3):799–823
Golub B, Jackson MO (2010) Naive learning in social networks and the wisdom of crowds. Am Econ J Microecon 2(1):112–149
Graham BS (2017) An econometric model of network formation with degree heterogeneity. Econometrica 85(4):1033–1063
Hoff PD (2003) Random effects models for network data. University of Washington
Hoff PD, Raftery AE, Handcock MS (2002) Latent space approaches to social network analysis. J Am Stat Assoc 97(460):1090–1098
Jackson MO (2010) Social and economic networks. Princeton University Press
Jadbabaie A, Molavi P, Sandroni A, Tahbaz-Salehi A (2012) Non-Bayesian social learning. Games Econ Behav 76(1):210–225
Krackhardt D (1987) Cognitive social structures. Soc Networks 9(2):109–134
Lazarsfeld P, Merton RK (1954) Friendship as a social process: a substantive and methodological analysis. Freedom and Control in Modern Society, pp 18–66.
McCluskey JJ, Gallardo RK, Ma X (2025) Is ignorance bliss? Milk from gene-edited cows and animal welfare considerations. Agric Resource Econ Rev, pp 1–32.
McPherson M, Smith-Lovin L, Cook JM (2001) Birds of a feather: homophily in social networks. Annu Rev Sociol 27(1):415–444
Merriam-Webster (2021) Definition of "Influencer". Retrieved from: https://www.merriam-webster.com/dictionary/influencer
Middelveld S, Macnaghten P (2021) Gene editing of livestock: sociotechnical imaginaries of scientists and breeding companies in the Netherlands. Elementa Sci Anthropocene 9:1
Middelveld S, Macnaghten P, Meijboom F (2023) Imagined futures for livestock gene editing: public engagement in the Netherlands. Public Underst Sci 32(2):143–158
Mobius M, Rosenblat T (2014) Social learning in economics. Annu Rev Econ 6(1):827–847
Navelski J, Pascual F (2025) New methods in estimating the latent space network model: An application to the U.S. media industry influencers and followers on Twitter. Working Paper
Penrose M (2003) Random geometric graphs, vol v. 5. OUP Oxford
Pew Research Center (2021). Social Media Use in 2021. Retrieved from: https://www.pewresearch.org/internet/2021/04/07/social-media-use-in-2021/.
Poole KT, Rosenthal H (2000) Congress: a political-economic history of roll call voting. Oxford University Press on Demand

Prabhune M (2019) Top 20 twitter accounts to follow for the latest CRISPR News. Retrieved from: https://www.synthego.com/blog/10-crispr-accounts-twitter.
Rasch G (1993) Probabilistic models for some intelligence and attainment tests. ERIC
Sampson SF (1968) A novitiate in a period of change: an experimental and case study of social relationships. Cornell University
Smith L, Sørensen P (2000) Pathological outcomes of observational learning. Econometrica 68(2):371–398
Stan Development Team (2021). RStan: the R interface to Stan. R package version 2.21.3
Statista (2020) Advertising revenue generated by selected media companies in 2020 (in billion U.S. dollars). Retrieved from: https://www.statista.com/statistics/554103/media-companies-ad-revenue/.
Synthego (2022). 60 Best CRISPR News Sources: Podcasts, Blogs, Journals, Twitter, and More. Retrieved from: https://www.synthego.com/learn/best-crispr-news-sources
Tabei Y, Shimura S, Kuan Y, Itaka S and Fukino N (2020) Analyzing Twitter conversation on genomeedited foods and their labeling in Japan. Front Plant Sci 11.
Twitter Inc (2022a) Twitter documentation. Retrieved from: https://help.twitter.com/en/managing-your-account/about-twitter-verified-accounts.. Internet Live Stats (2021). Internet Live Stats 1 Second. Retrieved from: https://www.internetlivestats.com/one-second/.
Twitter Inc (2022b) Twitter Academic Research API. Accessed through Twitter API v2: https://developer.twitter.com/en/products/twitter-api/academic-research/
Twitter Inc. (2022). Twitter Documentation. Retrieved from: https://help.twitter.com/en/managing-your-account/about-twitter-verified-accounts

Publisher's Note