

The Expert Effect on Network Formation: An Application to Genome Editing Opinions on Twitter

Joseph P. Navelski*

School of Economic Sciences

Washington State University

November 16, 2022

Abstract

I develop an alternative method to estimate the structure and influence in a social network within a learning model. Agents build a network by comparing the experts they follow in a particular field (i.e, politicians, media outlets, academics). I use a latent variable spatial following model to explain why agents follow these experts. The model estimates the underlying individual parameters that explain the decision to follow in a network, including the followers' and experts' ideological positions. I then use these estimates to derive each agent's level of influence in the social network and learning process. Using Twitter data, I apply this method to experts in the field of genome editing in domestic livestock (GEDL) showing that the anti-GEDL followers own 69% of the social influence in any conversation. In a post hoc analysis, I find that the consensus on Twitter about GEDL is anti-GEDL leaning. Implications are that any conversations about GEDL on Twitter will be heavily influenced by anti-GEDL followers, making it difficult for pro-GEDL opinions to be accepted.

Keywords: Latent variable MCMC estimation, network formation, opinion formation, social influence, social network analysis.

JEL Classification: C11 Bayesian Analysis: General; C15 Statistical Simulation Methods: General; C21 - Cross-Sectional Models; Spatial Models; Treatment Effect Models; Quantile Regressions; C49 Other; D82 - Information, Knowledge, and Uncertainty; Z13 Economic Sociology;

*Address: 301H Hulbert Hall, Washington State University, Pullman, WA 99164. E-mail: joseph.navelski@wsu.edu.

1 Introduction

Social networks facilitate information transmission between agents, helping agents gain insight from each other through repeated interaction. This process is known as social learning (Mobius and Rosenblat, 2014). In his seminal work, DeGroot (1974) proposes that agents learn over time by repeatedly updating their own beliefs (or opinions) through the averaging of all agents’ previous period beliefs, including their own. This approach is used in most non-Bayesian social learning and opinion dynamics research (Banerjee et al., 2021; Chandrasekhar et al., 2020; DeMarzo et al., 2003; Golub and Jackson, 2010).¹

A contribution from DeGroot (1974) is the notion of “learning convergence,” where agents, by repeated interaction in a network, eventually reach a consensus in their updated beliefs (i.e., a steady state). Reaching a consensus requires the social network to have certain graphical properties. If these properties are satisfied, each agent’s relative level of influence in the learning process is realized (DeGroot, 1974; DeMarzo et al., 2003; Golub and Jackson, 2010). This implies that learning, reaching a consensus, and an agent’s social influence are all depend on the social network’s geometric structure.

Estimating a network’s structure is not trivial, primarily because the empirical network formation literature in the learning framework is still at its infancy. Chandrasekhar (2016) provides a summary of utility-based models, but many lack the ability to estimate a network whose properties satisfy those required by DeGroot (1974). For example, random geometric graph models incorporate individual heterogeneity (Erdős et al., 1960; Penrose, 2003), but they are governed by thresholds chosen by the researcher. Additionally, parameter estimates are rarely used to interpret why the geometric structure exists. Graham (2017) provides an alternative model, where he uses unobserved variables (i.e., fixed effects in panel data) to uncover individual degree heterogeneity, but his model does not estimate the weight each individual puts on their own belief.

Another issue in the empirical network formation literature is the availability of data. Often times researchers need to have data on the connections between agents and the observable characteristics of each agent. This data is usually expensive to collect, and many times researchers

¹Other economist model learning in a Bayesian setting where opinions update as new evidence or information becomes available (Acemoglu and Ozdaglar, 2011; Banerjee, 1992; Bikhchandani et al., 1992; Glass and Glass, 2021; Smith and Sørensen, 2000). In this paper, I focus on a non-Bayesian setting.

need to define restrictive dimensions on the network because its size can become vast. Additionally, researchers may only have access to partitions or subsets of a network. This data can be rich with information, but difficult to estimate using the utility-based network formation models mentioned above. With these hurdles, it behooves researchers to find other ways to estimate networks for the non-Bayesian social learning environment.

Assuming social networks are homophilic (i.e., agents have a higher probability of connecting when they are similar (McPherson et al., 2001)), I develop an alternative methodology to estimate a social network to use in the DeGroot (1974) learning model. In particular, I assume agents form their own social network by comparing the experts they follow outside of their network. The network of followers is seen as a partition of a larger network. I estimate this network partition using a latent variable spatial following model developed from item-response theory (Bafumi et al., 2005; Barberá, 2015; Barberá et al., 2015; Hoff, 2003; Hoff et al., 2002; Navelski and Pascual, 2022; Rasch, 1993). This method uncovers each individual’s ideological position and social influence in the network. I compare these estimates to see which ideological positions have the most social influence in the network.

I apply this method to a Twitter dataset where I focus on agents comparing the expert accounts they follow in the field of genome editing in domestic livestock (GEDL). I find that individuals with anti-GEDL ideologies have 69% of the social influence on Twitter, indicating that any consensus reached will be heavily influenced by individuals who are against GEDL. I support this result with a post hoc analysis where I find that the consensus on GEDL is anti-GEDL leaning on Twitter. To my knowledge, this is one of the first non-Bayesian analyses to show that anti-GEDL individuals have the most influence in social media, indicating that pro-GEDL opinions will be difficult to adopt.

1.1 Related Literature

DeGroot (1974) was one of the first to introduce a model of learning and social influence. A finite amount 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 agent 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 (Golub and Jackson, 2010; Jackson, 2010). This convergence is the equivalent to reaching a consensus about the common idea or thought. When a consensus is reached the relative social influence weight each agent has in the learning process is realized.² 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.

Economists have recently started to empirically investigate how the structure of social networks affects learning and convergence in the DeGroot (1974) model. Chandrasekhar et al. (2020), for instance, propose methods to determine if agents use a mixture of learning types, Bayesian and non-Bayesian, and investigate how a sparse network (i.e., few connections between agents) leads to failures in asymptotic learning (i.e., drawing a consensus). They estimate the social network using a random utility framework and a mixture of two methods summarized by Chandrasekhar (2016) in Bramoullé et al. (2016). They use Penrose (2003)’s and Erdős et al. (1960)’s random geometric graph methods to model the presence of “clans” within the network. Penrose (2003)’s 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 more sparse network increases the chances that agents become “stuck” in their learning process, leading to an asymptotic learning failure. This research makes a seminal 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 similar methods to Chandrasekhar et al. (2020), proposing a generalization of DeGroot (1974)’s model where there is a mixture of informed and uninformed agents at time $t = 0$. 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 changes for agents in a sparse network using Penrose (2003)’s and Erdős et al. (1960)’s random

²I provide a complete review of the DeGroot (1974) model in the Appendix Section 6.1 for reference.

graph methods. They find that sparse social networks and signals (i.e., few connections and a small amount 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 give strong insights as to why true initial signals become construed or altered over time, but do not investigate how individual characteristics govern the formation of the social network affect learning and influence.

The empirical network formation literature is limited in the social learning setting. [Chandrasekhar \(2016\)](#) provides a summary of the utility-based network formation models where many have limitations in their application to [DeGroot \(1974\)](#)’s model. For example, both [Chandrasekhar et al. \(2020\)](#) and [Banerjee et al. \(2021\)](#) use random geometric graphs from [Erdős et al. \(1960\)](#) and [Penrose \(2003\)](#), but neither of them focus on individual (node) parameter estimates and how these estimates dictate a networks 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 \(2017\)](#)’s recent econometric work established 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 why [Graham \(2017\)](#)’s method has not yet been used in learning models is because 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.

Another reason why the utility-based network formation models have been unexplored in the context of learning is because it is difficult to collect network data ([Banerjee et al. \(2013\)](#); [Chandrasekhar \(2016\)](#)). Most network models use data that indicates how individuals are connected (i.e., an adjacency matrix), and if available, data on individual characteristics (e.g., gender, income, and race). Most of the time data is acquired through surveys that ask agents about their connections with other agents ([Sampson \(1968\)](#), [Banerjee et al. \(2013\)](#) and [Krackhardt \(1987\)](#)). Collecting this data is usually costly and many times researchers can only collect data on small portions of large networks. An additional difficulty is that many times researchers need to decide the bounds on a network. This decision may affect the interpretation of results and alter the research questions. For example, if a researcher wants to investigate how individuals learn and influence each other about a new technology, how do they gather data on a network when the world is highly connected?

An interesting 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 they are interested in. This implies data on the entire network does not exist, but there is data that can still provide insights. In particular, this data can provide insights into what led to these following decisions, and how these individuals might learn and influence each other in their own social network.

This paper utilizes the above scenario and proposes an alternative way to estimate a social network in [DeGroot \(1974\)](#). I assume social networks are homophilic and that agents build networks by comparing the connections they have with experts outside of the network. I estimate the unobserved individual parameters that explain the following behaviors between agents and experts. This method uncovers each individual’s ideological position and social influence in the network using a latent variable spatial following model ([Barberá, 2015](#); [Barberá et al., 2015](#); [Hoff, 2003](#); [Hoff et al., 2002](#); [Navelski and Pascual, 2022](#)). I compare these estimates to see which ideological positions have the most social influence in the network.

I apply this method to a Twitter dataset where I focus on agents comparing the expert accounts they follow in the field of genome editing in domestic livestock (GEDL). I find that individuals with anti-GEDL ideologies have 69% of the social influence on Twitter indicating that any consensus reached will be heavily influenced by individuals that are against GEDL. I conduct a post hoc analysis that supports this result, where I find that the consensus on GEDL is anti-GEDL leaning on Twitter. To my knowledge, this is one of the first non-Bayesian analyses to show that anti-GEDL individuals have the most influence in social media. Implications are that pro-GEDL opinions will be difficult to adopt in social media.

This paper is organized as follows: Section 2 presents the theoretical model, empirical model and the network estimation methodology. Section 3 surveys the Twitter data used in the application, and reviews the estimation diagnostics. Section 4 discusses the estimation, social influence and post hoc consensus results, and Section 5 concludes.

2 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.

2.1 A Social Network based on Common Connections

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). Let the social network g 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 $\mathbf{U}_{m \times m}$, $\mathbf{T}_{n \times n}$ and $\mathbf{W}_{m \times n}$ are not observed by the researcher, but matrix $\mathbf{A}_{n \times m}$ is, which takes the form

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

and it is a matrix of dyadic links. These links represent the directed one-way relationship between the i^{th} agent in n choosing to link with the j^{th} agent in m .³ Hence forth, \mathbf{A} will be denoted as the “connections matrix,” the m agents as “experts,” and the n agents as “followers.” Experts are knowledgeable and informed about their specializations, and followers link with experts to gain information about their specializations and because their positions or beliefs align (i.e., the

³If $\mathbf{W}_{m \times n} = (\mathbf{A}_{n \times m})^T$ then $\mathbf{A}_{n \times m}$ can be seen as an undirected network, but for this application I only consider $\mathbf{A}_{n \times m}$ as a directed partition of $\mathbf{G}_{N \times N}$.

connections are assumed to be homophilic).⁴

Let \mathbf{A} represent the disjoint graphical partition of \mathbf{G} , and the goal is to uncover the social network of followers represented by \mathbf{T} . Assume all followers form a social network by comparing each expert they follow in m . 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 $t_{ik}^* = t_{ik} / \sum_{i=1}^n t_{ik}$. 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 (1974)'s 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. Intuitively, the n followers compare the experts they follow outside of the network and this dictates the weight or trust they put on each others beliefs. Followers then reevaluate each connection relative to all of their connections.

2.2 An Estimated Social Network Based on Common Connections

Consider the same connections matrix \mathbf{A} , but where each dichotomous choice can be modeled with the logistic regression model, a specific form of a generalized linear mixed model (GLMM), given by

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

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

⁴McPherson et al. (2001) makes a strong and 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).

The specification in Equation 1 follows Hoff et al. (2002), Barberá (2015), and Navelski and Pascual (2022), who all propose models that 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 (i.e., the number of total followers for expert j and the number of total follows for follower i) These parameters can be interpreted as expert j ’s popularity and follower i ’s engagement, respectively (Barberá, 2015; Navelski and Pascual, 2022).

Parameters ϕ_j and θ_i are assumed to be in a one-dimensional space, and the absolute distance specification $-\gamma|\theta_i - \phi_j|$ follows the “homophilic” assumption (Hoff, 2003). These parameters are interpreted as “ideal points” (Bafumi et al., 2005; Curtis, 2010; Poole and Rosenthal, 2000; Rasch, 1993).⁵ An agent’s “ideal point” is his or her preference or position within a spatial framework, and the simplest spatial framework is characterized in a single dimension. In the political context, an ideal point represents an individual’s position on a scale from extremely liberal ($-\infty$) to extremely conservative ($\infty+$). I use Navelski and Pascual (2022)’s inverse arc-tangent method to map these latent positions on a spectrum between -1 and 1 , -1 is the most extreme “anti” point of view and 1 is the most “pro” point of view. I map to this scale for ease of interpretation and comparison.

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

$$\hat{a}_{ij} = [1 + \exp(-\hat{\pi}_{ij})]^{-1},$$

and the initial social interaction rule is defined as

$$\hat{t}_{ik} = \sum_{j=1}^m \hat{a}_{ij} \hat{a}_{kj}.$$

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

⁵This interpretation is further supported by the probabilistic voting model literature where agents vote for candidates based on how close their “ideology” or “reputations” align (Coughlin and Nitzan, 1981; Coughlin, 1992; Enelow and Hinich, 1984, 1989). 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.

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 $t_{ik}^* = t_{ik} / \sum_{i=1}^n t_{ik}$ to form a row vector corresponding to follower i in \mathbf{T}^* . Followers then learn from each other based on \mathbf{T}^* . Assuming \mathbf{T}^* is strongly connected, beliefs converge to a common belief, and a social influence vector \mathbf{s} is realized.

Intuitively, this social interaction matrix \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 . I 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.

2.2.1 Markov-Chain Monte Carlo (MC-MC) Estimation

The model parameters are unknown and the statistical problem is to perform inference on $\boldsymbol{\alpha} = (\alpha_1, \dots, \alpha_m)'$, $\boldsymbol{\beta} = (\beta_1, \dots, \beta_n)'$, $\boldsymbol{\phi} = (\phi_1, \dots, \phi_m)'$, $\boldsymbol{\theta} = (\theta_1, \dots, \theta_n)'$, 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

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

where $\pi_{ij} = \mu + \alpha_j + \beta_i - \gamma|\theta_i - \phi_j|$ 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 $|\theta_i - \phi_j|$. 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 becomes even more difficult as datasets become larger.

To overcome the tractability problem, I follow [Navelski and Pascual \(2022\)](#)’s 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. I provide more details on how to implement this estimation method in the Appendix Section [6.2](#).

3 Data: Genome Editing in Domestic Livestock (GEDL)

Social media is a natural setting for this framework because it has network data on followers who choose to connect with, or “follow,” an expert in the field they are interested in. I use data from Twitter because about one in four US adults (23%) say they use this social network and Twitter allows academics to conduct research using information from their public accounts ([Pew Research Center, 2021](#); [Twitter Inc., 2022a](#)). I apply my model to a panel of $m = 46$ experts, and $n = 3,383$ of their most informed followers, from the genome editing in domestic livestock (GEDL) field.⁶

3.1 Choosing the GEDL Experts and Procuring the Data from Twitter

I first defined a list of GEDL experts on Twitter. The list consists of academics, organizations, journals, politicians and companies that have an active presence in the GEDL industry, or in an industry that is closely related to GEDL (e.g., the genome editing industry in general). This list was assembled in two steps.

The first step was to include accounts that appear when searching for terms related to GEDL on Twitter. The terms I searched for were *animal welfare*, *biotechnology*, *crispr*, *dairy*, *dehorning*, *gene editing*, *genetically modified*, *genome editing*, *genome engineer*, *GMO*, and *organic*, and these terms were gathered from [Social Mention](#). [Social Mention](#) is an online software that identifies the key terms that are closely related to a topic people are posting about online. I selected accounts that had more than 1,000 followers and clearly had a position in genome editing. I used these thresholds because I wanted to gather accounts that were reasonably popular but also not too general where their position on GEDL was ambiguous or unclear. Examples of ambiguous accounts

⁶[Barberá \(2015\)](#) and [Navelski and Pascual \(2022\)](#) 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.

are “Biotechnology” or “NatureBiotech” where they may promote genome editing technologies, but also other biotechnology technologies. I also chose to include five producers from the domestic livestock industry because their position on GEDL could dictate their following on Twitter. I also was interested to see how their ideological positions compared to other accounts.

In my second step, I did a web-search on “Twitter accounts to follow for information about genome editing.” The top search result was a blog post by Synthego, a large player in the genome editing industry. The blog post provided a list of the “top 20 Twitter accounts to follow for the latest CRISPR news,” and some of these accounts are verified on Twitter as being authentic, notable, and active (Prabhune, M., 2019; Synthego, 2022; Twitter Inc., 2022b). I included all of these accounts on the list since this blog post would lead a user to follow these accounts when investigating GEDL. All of these had more than 1,000 followers, and the blog post clearly states how each account is related to the genome editing industry. Most of these accounts actively disseminate information about GEDL and genome editing in general, whether the information is positive or negative, and many of them have amassed a following based on their informativeness and position about GEDL. For example, NonGMOProject and CRISPR_News both have over 30,000 followers because they are seen to be a valid source of information about GEDL and/or genome editing in general.

These two steps led to a final list of 46 experts in the field of GEDL, and I provide a detailed list of the experts accounts and their account characteristics in Figure 6 in the Appendix. I downloaded all of the accounts that follow each expert, and merged the datasets by each follower’s unique ID.⁷ This produced a connections matrix \mathbf{A} that has $n = 187,209$ followers.

The experts are initially labeled as being “anti-” or “pro-” GEDL based on the message they are trying to disseminate, but this labeling is only used for preliminary analyses to motivate the data structure.⁸ There are 12 anti- and 34 pro-GEDL accounts. OrganicConsumer and NonGMOProject have 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 shows that the average number of followers per anti-GEDL 59,596, account is greater than pro-GEDL 15,223 followers. Additionally, the anti-GEDL experts are more active on Twitter than the pro-GEDL experts because they tend to follow

⁷I 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).

⁸This 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-GEDL vs. the most pro-GEDL.

more accounts on Twitter and tweet more on average. These results imply the average anti-GEDL experts are more connected and engaged to the Twitter network than the pro-GEDL experts.

Table 1: Summary Statistics of the Experts’ Account Characteristics by GEDL Viewpoint
Means and Standard Deviations (in Parentheses)

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

3.2 Data Reduction

I reduce the dataset to include only those followers that follow at least 9 out of the 46 accounts. I perform this reduction to focus the analysis on the most informed followers in GEDL, reduce the amount of potential Twitter “bots” in the data, to help with estimation tractability, and to ensure matrix $\mathbf{T}_{n \times n}^*$ is strongly connected. This reduces the set of followers to $n = 3,383$, and this reduction slightly alters the research question to be focused on the “more informed” followers in regards to GEDL. The number of followers per account in the reduced connections matrix is presented in Table 2, and Table 3 presents the anti-GEDL and pro-GEDL group averages where the anti-GEDL accounts still have, as a group, more followers (1,456) on average than the pro-GEDL accounts (517).

Table 2: Total Number of Followers per Expert Account for the Reduced (More Informed) Dataset

Screen Name	Followers	Position	Screen Name	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	KevinADavies	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

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 “dash 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 are the anti-GEDL experts, which are sorted in decreasing order by the number of followers they have. The subsequent 34 columns are the pro-GEDL 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-GEDL accounts they follow less the amount of pro-GEDL accounts they follow. Intuitively, the first follower is the most anti-GEDL in terms of following, while the last follower is the most pro-GEDL.

Table 3: The Experts’ Number of Followers by GEDL Viewpoint
for the Reduced (More Informed) Dataset
Means and Standard Deviations (in Parentheses)

Group	Followers
Anti-GEDL	1456 (558)
Pro-GEDL	517 (302)

Many of the followers in the northwest quadrant follow a large portion of the anti-GEDL accounts, indicated by the dark black mass. The followers in the southeast quadrant are more sparse and not following a large proportion of the pro-GEDL accounts, indicated by the patchy black and white area. This indicates anti-GEDL account followers tend to be more “loyal” whereas pro-GEDL account followers tend to have more of a “smattering,” or mixed, in their following structure. Another key observation about the connections matrix is the lack of extremely informed followers that follow many of the 46 accounts. These followers are presented in the middle of the matrix, and it is apparent that out of these followers many of them follow the anti-GEDL accounts while following some of the pro-GEDL accounts. This indicates that even though some followers are extremely informed about GEDL, most of them are more informed about the anti-GEDL experts rather than the pro-GEDL experts.

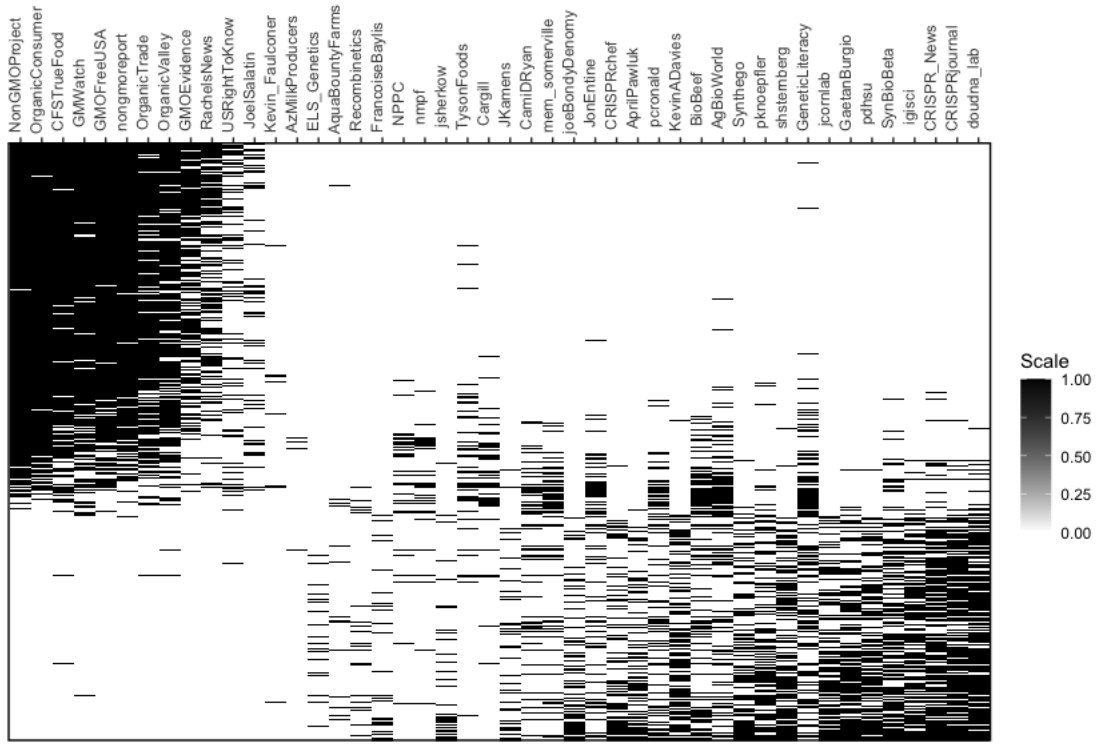


Figure 1: Heat Map of Connections Matrix **A**

3.3 Estimating the Data Structure

I estimate the connection matrix **A** from Figure 1 using the model and methodology outlined in Section 2.2. All MC-MC (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 MC-MC estimation process. Convergence also implies the likelihood function is in the same form as Equation 2, estimates are consistent, and hypothesis testing can be conducted. The MC-MC 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.⁹ To further motivate estimation results, Figure 2 shows a heat-map of the estimated probabilities of following, which closely resembles the raw data structure presented in Figure 1.

⁹I provide a more detailed explanation about prediction diagnostics in Appendix Section 6.5.

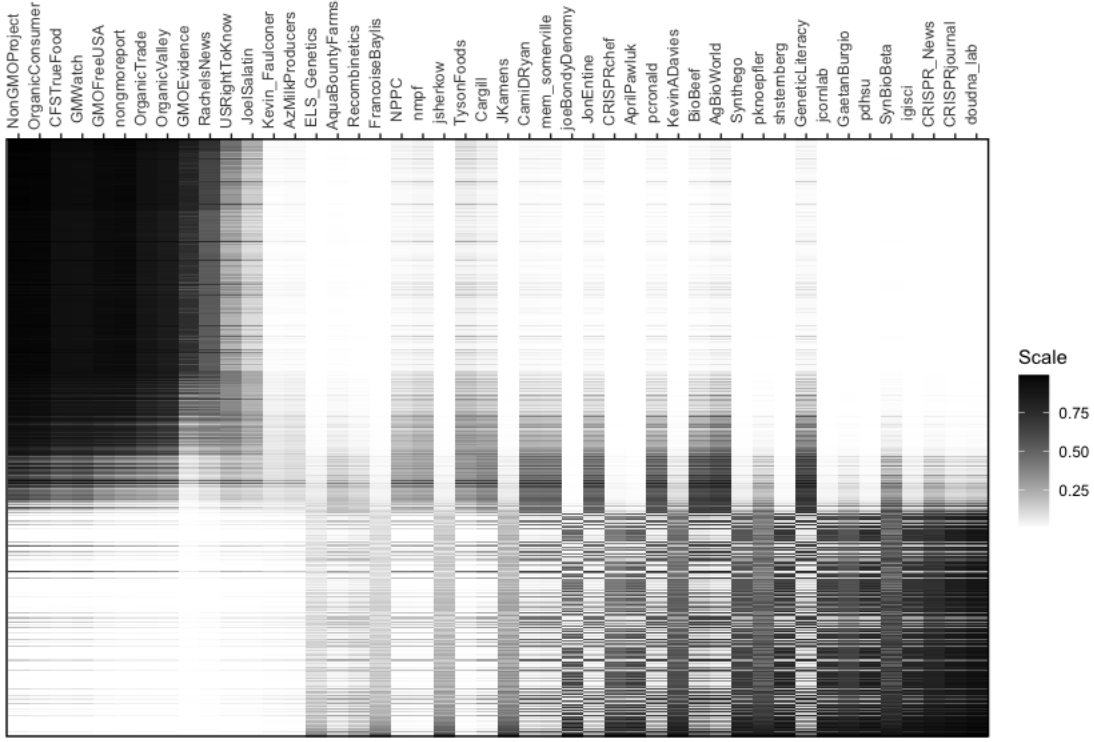


Figure 2: Heat Map of the Estimated Connections Matrix $\hat{\mathbf{A}}$

Figure 1 and 2 are used as the initial connections matrices \mathbf{A} to derive the weighted “trust” matrix \mathbf{T}^* used in the learning process. Figures 11 and 12 in the Appendix Section 6.6 are heat-maps of these matrices. 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. I next use these matrices are used in a DeGroot (1974) model to derive the social influence vectors $\mathbf{s}_{1 \times n}$.

4 Results

4.1 Estimation Results

Results show that @NonGMOProject is the most popular expert with an estimate of $\hat{\alpha}_2 = 3.78$, and that @Kevin.Faulconer is the least popular expert with an estimate of $\hat{\alpha}_{15} = -3.30$. These results are expected since @NonGMOProject and @Kevin.Faulconer have the most and least amount of followers in the data. The estimated ideal positions show that the most anti-GEDL expert is

@GMOEvidence ($\hat{\phi}_{12} = -0.687$) and the most pro-GEDL expert is @joeBondyDenomy ($\hat{\phi}_{42} = 0.847$). Figure 3 plots all of the expert’s estimated ideologies on a scale ranging from -1 (anti) to 1 (pro) with hash mark in the middle representing a zero line. I highlight some experts to show how their estimates align with the detailed information in their profile.

As expected, results show many expert accounts that are thought of to be anti-GEDL are since most of their ideal points are closer to -1 . Additionally, @doudna_lab is the official Twitter account for Dr. 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 discovery of CRISPR, and it is not surprising that her lab’s ideal point is positioned on the more pro-GEDL side of the spectrum at 0.72 .¹⁰ One unexpected result is that @NPPC, @Cargill and @TysonFoods are all representatives of the meat producing industry and are expected to be more pro-GEDL to reduce production costs, but they are seen to have ideologies that are more moderate since their ideal point estimates are closer to zero.

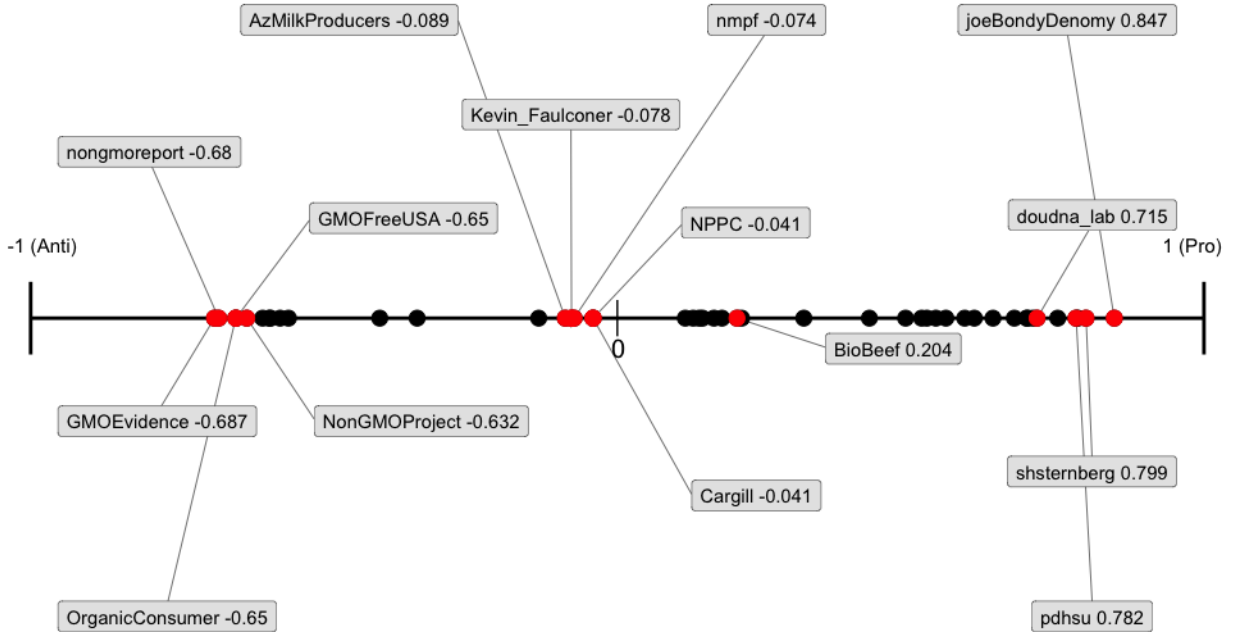


Figure 3: Estimated Ideology of Experts $\hat{\phi}$ for the GEDL Data

¹⁰A summary and discussion about expert popularity and ideology is presented in the Appendix Section 6.4.

Figure 4 (a) is a histogram of the estimated engagement parameters $\hat{\beta}_i$ 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 GEDL is high for some followers, but low for most. Figure 4 (b) is a histogram of the 3,383 followers’ ideal points $\hat{\theta}_i$ where many individuals are polarized about GEDL. 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 GEDL will have an anti-GEDL ideology. These metrics are represented by the solid (mean) and dotted (median) lines in the middle of Figure 4 (b), and this result is even more apparent when observing the large “spike” on the anti-GEDL 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 at 56.73% of the followers align more with the anti-GEDL expert accounts.

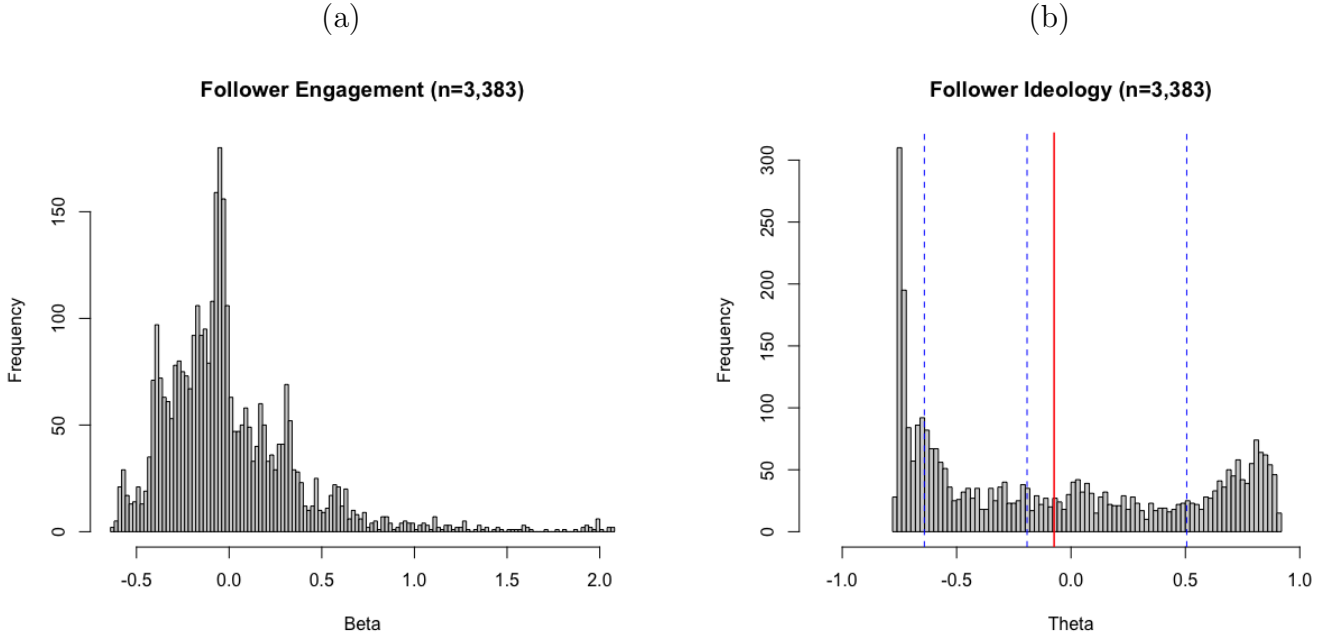


Figure 4: Ideology $\hat{\beta}_i$ (a) and Engagement $\hat{\theta}_i$ (b) for the 3,383 Followers

4.2 Social Influence

Table 4 shows that the social influence distribution is left-skewed when agents interact via the base-case social network (Section 2.1) and the estimated social network (Section 2.2). 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. This also implies that the individuals with low social influence have more intense levels of low influence than those with high influence.

Table 4: Summary Statistics for the Social Influence Vectors \mathbf{s}

Method	Min.	1st Qu.	Median	Mean	SD	3rd Qu.	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%

I compare the followers’ estimated ideologies with their social influence estimates and summarize the results in Table 5. I find that individuals with negative ideology 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-GEDL individuals have 31% of the total influence while the more pro-GEDL individuals have 18% of the social influence. This implies that the most extreme anti-GEDL 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-GEDL followers have the majority of the influence in the learning process.

Table 5: Social Influence By Ideology θ

$\theta_i < Q1$	$\theta_i < 0$	$\theta_i > 0$	$\theta_i > Q3$
31%	69%	31%	18%

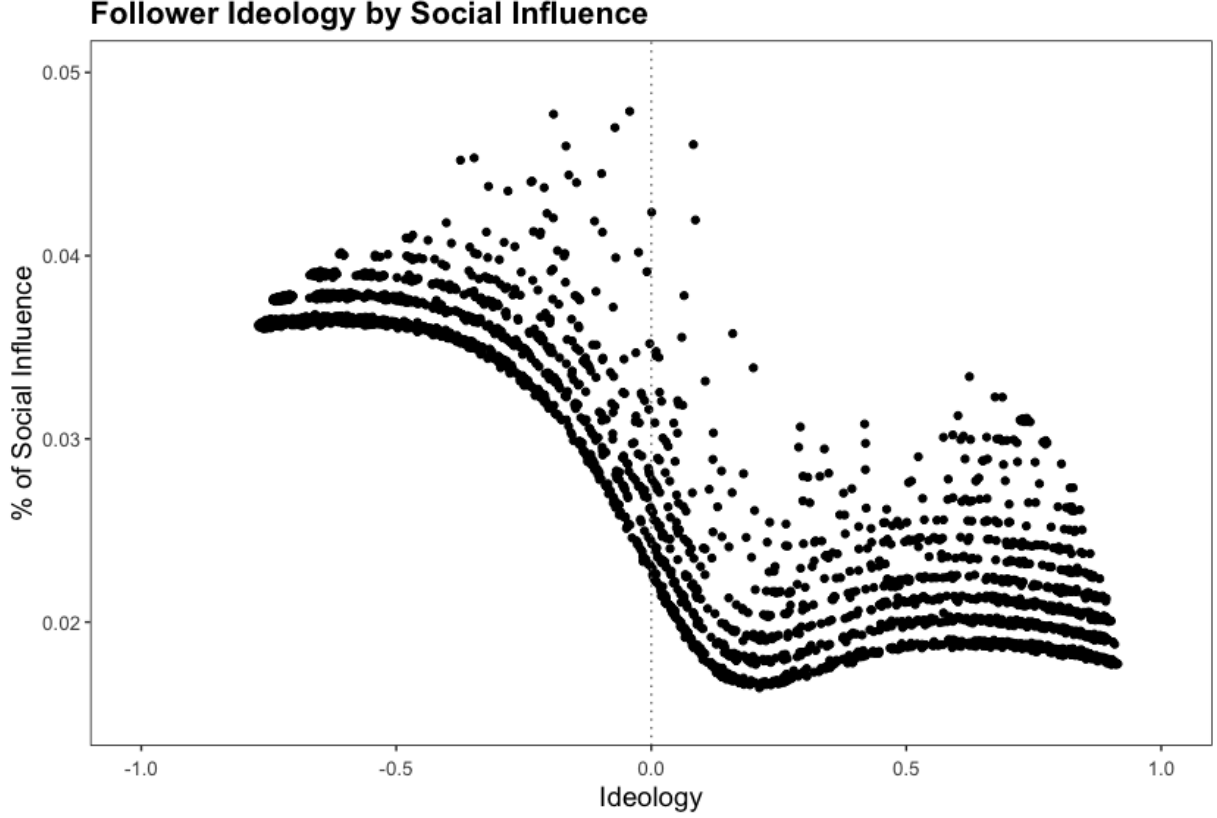


Figure 5: Estimated Ideology of Followers $\hat{\theta}_i$ by Social Influence s_i for the GEDL Data

4.3 Post Hoc Analysis: A Consensus Belief Based on Ideologies

One by-product of my social network estimation method is that agents' initial beliefs are uncovered in the form of ideologies. I 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 negative one to one $[-1, 1]$, 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 GEDL (extremely anti-GEDL), while a one is the most in favor of GEDL (extremely pro-GEDL). In the context of GEDL, I find that agents converge to a consensus belief of .395 using the equation $P^{(\infty)} = \mathbf{s}P^{(0)}$ from [DeGroot \(1974\)](#). This indicates that the consensus belief on Twitter about GEDL is anti-GEDL given that initial beliefs are standardized ideology estimates.

5 Conclusion

There are two contributions of this paper. The first contribution is that I develop an alternative method to estimate the structure and influence of a social network in a learning model. I 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 I 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, I 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 my model with large datasets to target agents who have the most influence in a social network and align with their viewpoint.

The second contribution I make is that I apply this method to a social media dataset from Twitter with 46 experts from the genome editing in domestic livestock (GEDL) industry and 3,383 of their most informed followers. The main results are that 56.73% of the followers have ideologies that align with anti-GEDL experts and that these followers own 69% of the social influence on Twitter. The post hoc analysis shows that the consensus on GEDL is .395 on a scale where zero is the most anti-GEDL and one is the most pro-GEDL. To my knowledge, this is one of the first papers to show that individuals who are anti-GEDL have the most 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-GEDL followers. This means that anti-GEDL will dominate the conversation on Twitter, and this could make it difficult for positive opinions about GEDL to be accepted. The post hoc analysis supports this intuition where the consensus on Twitter is not in favor of GEDL 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 GEDL on social media is negative, and policy makers should target anti-GEDL followers with high influence in an attempt to change their position and message on GEDL.

References

- Acemoglu, D. and Ozdaglar, A. (2011). Opinion dynamics and learning in social networks. *Dynamic Games and Applications*, 1(1):3–49.
- Bafumi, J., Gelman, A., Park, D. K., and Kaplan, N. (2005). Practical issues in implementing and understanding bayesian ideal point estimation. *Political Analysis*, 13(2):171–187.
- Banerjee, A., Breza, E., Chandrasekhar, A. G., and Mobius, M. (2021). Naive learning with uninformed agents. *American Economic Review*, 111(11):3540–74.
- Banerjee, A., Chandrasekhar, A. G., Duflo, E., and Jackson, M. O. (2013). The diffusion of microfinance. *Science*, 341(6144):1236498.
- Banerjee, A. V. (1992). A simple model of herd behavior. *The quarterly journal of economics*, 107(3):797–817.
- Barberá, P. (2015). Birds of the Same Feather Tweet Together: Bayesian Ideal Point Estimation Using Twitter Data. *Political Analysis*, 23:76–91.
- Barberá, P., Jost, J. T., Nagler, J., Tucker, J. A., and Bonneau, R. (2015). Tweeting from left to right: Is online political communication more than an echo chamber? *Psychological Science*, 26(10):1531–1542. PMID: 26297377.
- Bikhchandani, S., Hirshleifer, D., and Welch, I. (1992). A theory of fads, fashion, custom, and cultural change as informational cascades. *Journal of political Economy*, 100(5):992–1026.
- Bramoullé, Y., Galeotti, A., and Rogers, B. W. (2016). *The Oxford handbook of the economics of networks*. Oxford University Press.
- Chandrasekhar, A. (2016). Econometrics of network formation. *The Oxford handbook of the economics of networks*, pages 303–357.
- Chandrasekhar, A. G., Larreguy, H., and Xandri, J. P. (2020). Testing models of social learning on networks: Evidence from two experiments. *Econometrica*, 88(1):1–32.
- Coughlin, P. and Nitzan, S. (1981). Electoral outcomes with probabilistic voting and nash social welfare maxima. *Journal of Public Economics*, 15(1):113–121.
- Coughlin, P. J. (1992). *Probabilistic voting theory*. Cambridge University Press.
- Csardi, G. and Nepusz, T. (2006). The igraph software package for complex network research. *InterJournal, Complex Systems*:1695.
- Curtis, S. M. (2010). Bugs code for item response theory. *Journal of Statistical Software, Code Snippets*, 36(1):1–34.
- DeGroot, M. H. (1974). Reaching a consensus. *Journal of the American Statistical Association*, 69(345):118–121.

- DeMarzo, P. M., Vayanos, D., and Zwiebel, J. (2003). Persuasion bias, social influence, and unidimensional opinions. *The Quarterly journal of economics*, 118(3):909–968.
- Enelow, J. M. and Hinich, M. J. (1984). *The spatial theory of voting: An introduction*. CUP Archive.
- Enelow, J. M. and Hinich, M. J. (1989). A general probabilistic spatial theory of elections. *Public Choice*, 61(2):101–113.
- Erdős, 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, J. B., Stern, H. S., and Rubin, D. B. (1995). *Bayesian data analysis*. Chapman and Hall/CRC.
- Gelman, A. and Rubin, D. B. (1992). Inference from iterative simulation using multiple sequences. *Statistical science*, 7(4):457–472.
- Glass, C. A. and Glass, D. H. (2021). Social influence of competing groups and leaders in opinion dynamics. *Computational Economics*, 58(3):799–823.
- Golub, B. and Jackson, M. O. (2010). Naive learning in social networks and the wisdom of crowds. *American Economic Journal: Microeconomics*, 2(1):112–49.
- Graham, B. S. (2017). An econometric model of network formation with degree heterogeneity. *Econometrica*, 85(4):1033–1063.
- Hoff, P. D. (2003). Random effects models for network data.
- Hoff, P. D., Raftery, A. E., and Handcock, M. S. (2002). Latent space approaches to social network analysis. *Journal of the American Statistical Association*, 97(460):1090–1098.
- Jackson, M. O. (2010). *Social and economic networks*. Princeton university press.
- Jadbabaie, A., Molavi, P., Sandroni, A., and Tahbaz-Salehi, A. (2012). Non-bayesian social learning. *Games and Economic Behavior*, 76(1):210–225.
- Krackhardt, D. (1987). Cognitive social structures. *Social networks*, 9(2):109–134.
- McPherson, M., Smith-Lovin, L., and Cook, J. M. (2001). Birds of a feather: Homophily in social networks. *Annual Review of Sociology*, 27(1):415–444.
- Merriam-Webster (2021). Definition of “Influencer”. Retrieved from: <https://www.merriam-webster.com/dictionary/influencer>.
- Mobius, M. and Rosenblat, T. (2014). Social learning in economics. *Annual review of economics*, 6(1):827–847.
- Navelski, J. and Pascual, J. (2022). New methods in estimating the latent space network model with an application to the us media industry on twitter. *Working Paper*.

- Penrose, M. (2003). *Random geometric graphs*, volume 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, K. T. and 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, S. F. (1968). *A novitiate in a period of change: An experimental and case study of social relationships*. Cornell University.
- Smith, L. and 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.
- Synthego (2022). 60 Best CRISPR News Sources: Podcasts, Blogs, Journals, Twitter, and More. Retrieved from: <https://www.synthego.com/learn/best-crispr-news-sources>.
- Twitter Inc. (2022a). Twitter Academic Research API. Accessed through Twitter API v2: <https://developer.twitter.com/en/products/twitter-api/academic-research/>.
- Twitter Inc. (2022b). Twitter Documentation. Retrieved from: <https://help.twitter.com/en/managing-your-account/about-twitter-verified-accounts>.

6 Appendix

6.1 A Review of DeGroot (1974)’s Social Learning Model

The model considers a finite set of $N = \{1, \dots, n\}$ *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 .¹¹ 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.¹²

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

$$\mathbf{p}^{(t)} = \mathbf{T}^t \mathbf{p}^{(0)} \quad (3)$$

where,

$$\mathbf{T}_{n \times n} = \begin{bmatrix} t_{11} & t_{12} & \dots & t_{1n} \\ t_{21} & t_{22} & \dots & t_{2n} \\ \vdots & \vdots & \dots & \vdots \\ t_{n1} & t_{n2} & \dots & t_{nn} \end{bmatrix} \quad \text{and} \quad \mathbf{p}_{n \times 1}^{(t)} = \begin{bmatrix} p_1^{(t)} \\ p_2^{(t)} \\ \vdots \\ p_n^{(t)} \end{bmatrix}.$$

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.¹³ Intuitively, one can think of a strongly connected network as a network where there are no partitions that are completely isolated from the other parts of the network. These two properties guarantee a consensus is reached where $\mathbf{p}_{n \times 1}^{(\infty)} = \lim_{t \rightarrow \infty} \mathbf{T}^t \mathbf{p}^{(0)}$ 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 $p_1^{(\infty)} = \dots = p_n^{(\infty)}$.

¹¹This 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, and it can be directed or undirected.

¹²DeMarzo et al. (2003) refer to $\mathbf{T}_{n \times n}$ as the “listening” matrix where each element t_{ik} represents how much agent i listens to agent k ’s opinion, Golub and Jackson (2010) refer to t_{ik} as how much precision agent i puts k ’s opinion, Jadbabaie et al. (2012) refer $\mathbf{T}_{n \times n}$ as the social interaction matrix where t_{ik} represents the “influence” or “persuasion power” agent i 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 i th agent has on the current belief of agent k in forming its own belief for the next period. In this paper, I will be referring to t_{ik} as the “trust” agent i puts on agent k ’s opinion.

¹³I 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.

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 $\lim_{t \rightarrow \infty} \mathbf{T}^t \mathbf{p}^{(0)} = \mathbf{s} \mathbf{p}^{(0)}$. Each element s_i in vector $\mathbf{s}_{1 \times n}$ represents the amount of influence each agent has in 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 $p^{(\infty)} = \mathbf{s} \mathbf{p}^{(0)} = \sum_{i=1}^n s_i p_i^{(0)}$ for any vector of initial beliefs (DeGroot, 1974; DeMarzo et al., 2003; Golub and Jackson, 2010; Jackson, 2010). This implies that the structure of social networks have an important effect on learning, drawing a consensus, and influence in the DeGroot (1974) model.

View	Username	Location	Created At	Name	Description	Verified	Follower Count	Following Count	Tweet Count	Linked Scraped	Followers Scraped	Diff
Anti	GMOEvidence	United Kingdom	9/9/12 11:10	GMO Evidence	GMO Evidence is a worldwide user friendly library of evidence of harm caused by GMOs to animals and humans. Investigate public health groups working globally to expose corporate wrongdoing and government failures threatening our food, environment and health.	FALSE	6785	2516	1609	138	6000	785
Anti	USRightToKnow	United States	8/4/13 3:43	U.S. Right to Know	Official TWITTER: Healing the earth with good food. Author of 8 books. Featured in National Geographic, Smithsonian, Oritmore, A&E, Food Inc.	FALSE	10151	1910	6118	231	10000	151
Anti	JoelSalatin	Swoope, VA	7/12/11 17:05	Joel Salatin	Monthly news magazine focusing on the risks of genetically modified foods and the non-GMO food trend.	FALSE	29272	12	35	876	29000	272
Anti	nonreport	NA	3/25/09 22:28	Non-GMO Report		FALSE	31293	842	21014	460	37000	5707
Anti	GMAWatch	NA	8/17/09 16:46	GMAWatch	Connecting the propagandists of the biotech industry. Subscribe to free email newsletters: https://t.co/GKqRNv4dL , https://t.co/8WpP3CtY and https://t.co/8WpP3CtY (related TV hour @CRChips). https://t.co/8WpP3CtY @Tweets https://t.co/8WpP3CtY Organic https://t.co/8WpP3CtY @CenterforH	FALSE	42109	2334	96161	1423	41997	112
Anti	BioethicsNews	Made in Canada	9/14/11 16:58	bioethics Parent		FALSE	44885	29232	83899	825	43999	886
Anti	CFSTFood	DC CA OR HI	1/6/09 20:56	Center f Food Safety	Protecting our food, farms & environment (Get the latest: https://t.co/959q2vDnH5	FALSE	44881	9267	16007	1264	44000	891
Anti	GMOFreeUSA	NA	7/16/12 2:14	GMO Free USA	We are a national group, educating consumers about the potential hazards of Genetically Engineered foods. The Organic Trade Association (OTA) is the membership-based business association for organic agriculture and products in North America.	FALSE	52621	14613	41595	715	55998	3377
Anti	OrganicTrade	Washington, D.C.	4/8/09 15:52	Organic Trade		FALSE	59563	4214	20771	1160	58968	565
Anti	OrganicValley	La Farge, WI	7/17/08 14:29	Organic Valley	We like a cooperative of family farmers on a mission to produce the world's best organic food for you and your family. A nonprofit organization committed to preserving and building sources of free-from-GMO products, educating consumers, and providing verified non-GMO choices.	TRUE	80863	8094	16402	1394	79999	864
Anti	NonGMOProject	NA	6/11/09 23:20	Non-GMO Project	Want to take action on the issue you care about: Ag pesticides, factory farms and GMO's? Subscribe to our newsletter: https://t.co/auYpYp1	FALSE	125516	1100	7274	1676	127995	2479
Anti	OrganicConsumer	Finland, Minnesota	2/3/09 16:42	Organic Consumers Association	Booth leader in precision breeding (Acceligen), borrowed some models for preclinical research (Surragen) & enablers for regenerative medicine (Regenoid)	FALSE	187209	2939	27686	3288	186995	214
Pro	Recombinetics	Saint Paul, MN USA	1/23/11 15:37	Recombinetics	Official Twitter account for AZ Milk Producers and Dairy Council of AZ. Encouraging everyone to get their 3 servings of delicious dairy everyday.	FALSE	1218	1080	1524	46	1000	238
Pro	AZMilkProducers	Arizona	11/12/09 17:23	AZ Milk Producers	Dr. David Nisbet and professor of food science at @NCSU. Editor in Chief of The @GISIPJournal.	FALSE	2469	1175	6530	75	2000	459
Pro	CRISPRedit	NC State University	11/10/17 20:43	rodolphe Barrangou	We provide fresh, nutritious Atlantic Salmon, genetically engineered once 30 years ago and raised in safe, secure & sustainable bird-based aquaculture farms.	FALSE	2511	132	190	50	2000	511
Pro	AquaBountyFarm	Maynard, Massachusetts	12/26/12 16:24	AquaBounty	BLACK LIVES MATTER. Associate Prof. @JUCSF, Dept. of Micro & Imm., Toronto and Waterloo alum. Acigen BioSciences co-founder.	FALSE	3341	638	1634	54	3000	341
Pro	JoeBonnyGenomy	San Francisco, CA	12/25/12 1:54	Joe Bonny-Genomy		FALSE	3851	2120	2919	41	3000	851
Pro	FranciscoBaylis	Hillfax, Canada	11/7/12 14:57	Francisco Baylis, PhD	Author "Altered Inheritance" (Philosophy, University Research Professor, Dalhousie University)	FALSE	3919	656	12963	112	3000	919
Pro	Isabelle	Zürich, Switzerland	2/23/14 21:47	Isabelle Com	Genome editing, functional genomics, and cell figuring out how to eat healthier without dying. Professor of Genome Biology at ETH Zurich.	FALSE	4248	438	2505	78	4000	248
Pro	AgriPawuk	Toronto, Ontario	5/2/01 1:56	Dr. Agri Pawuk	Strategic Program Manager at Harvard Chan Microbiome in Public Health Center. Full-time PhD. Former CalPines editor. Regularly featured by microbios. (Dr./he/hu)	FALSE	4262	1799	7554	62	4000	262
Pro	jbarkow	Champaign, IL	8/17/12 0:12	Jacob S. Barkow @Jbarkow110	Prof @Jbarkow110 and @Giblinho; also @Cell, Center apatens, HIP, Abduct, #Pharma, #Bioethics, Ducks, drugs, and risks that roll in Creation"	FALSE	4840	299	18887	173	4000	840
Pro	ashenberg	New York, NY	4/23/15 4:04	Sam Sternberg	Assistant professor at Columbia University / CRISPR aficionado / Protein/Albina biochemist / Co-author with Jennifer Doudna of "A Crack in Creation"	FALSE	4946	495	1788	64	4000	946
Pro	Joanens	Boston, MA	5/25/11 14:05	Dr. Joanne Joanens	Dr. Joanne Joanens, PhD, Director of the Center for Food Safety and Food Security, USDA	TRUE	5266	4808	19086	175	5000	826
Pro	mem_cornville	NA	6/22/08 21:36	Dr. mem_cornville Wossamotta U	The kind of food we eat is what people claim they want to have from, and they dislike the conclusions they hear. Spectacularly combustible... "a real person?"	FALSE	6279	4485	83965	259	5999	280
Pro	Synthgo	California, USA	8/6/12 1:17	Synthgo	The genome engineering platform company, @USF, @CRISPR, @GenomeEngineer	TRUE	6137	2178	3876	145	6000	327
Pro	pcornald	In the lab or garden in Davis	5/20/08 23:15	@pcornald	UC Davis white geneticist & author of Tomorrow's Table: Organic Farming Genetics & the Future of Food. TED talk. My lab studies immunity & stressors. #SciComm	FALSE	7096	2024	9108	321	6999	97
Pro	lorietrine	Cincinnati	3/25/09 15:44	Jon Eirine	Dad, journalist, iconoclast, founder https://t.co/VaCwqg4eoa - #biotech/#biomedicine/sustainability/ethics. Emmy-winning TV news producer, book author	FALSE	7569	5961	22307	190	6004	575
Pro	ELIS_Genetics	Amsterdam, Boston, Oxford	10/11/11 18:59	Genetics & Genomics	Genetics & Genomics books & journals. General genetics & genomics news, call for papers, special issues, conference abstracts & more!	FALSE	8166	7976	3939	140	8000	166
Pro	KevinDavies	Washington DC	4/17/09 22:22	Kevin Davies	Author "ERRING HUMANNITY: The CRISPR Revolution and the New Era of Genome Editing"; Exec Ed @CRISPRJournal, @GEMbiotechIntl, @naspace4ers	FALSE	8419	1675	10122	312	8000	419
Pro	BioBeef	Davis, California	9/30/10 19:24	Prof Alison Van Eenennaam (@BioBeef)	Animal #iGen biotech/bio/genetics, Univ of CA Davis, Mom, Consumer, Ag Science, https://t.co/15Ua1TtUa1T Melbourne Uni, Food Evolution https://t.co/0cm6gDvXvX	FALSE	10059	1124	17188	236	10000	59
Pro	igsci	Berkeley, CA	2/24/16 20:33	Innovative Genomics Institute	The Innovative Genomics Institute (IGI) is dedicated to advancing genome research for a better world. Partnership between @UCBerkeley and @UCSF	FALSE	10655	750	3974	164	10000	655
Pro	CantidHyam	St Louis County, Missouri	3/6/09 23:58	Cami Ryan, PhD	information "And I'm waving my flag!" Social & behavioral health sciences @UW • @BayAreaCrops @GenomeCanada @agbiourask and @UW more interesting on IG and Facebook	FALSE	10797	5012	57347	307	10000	797
Pro	imgf	Arlington, VA, United States	7/31/09 19:02	National Milk Producers Federation	Representing farmer-owned dairy cooperatives on Capitol Hill since 1916.	FALSE	11824	1471	1678	251	12000	176
Pro	pbhu	hhu.berkeley.edu	3/17/09 3:05	Patrick Hsu	Assistant Professor @BerkeleyPhD, Co-founder @Architecture CRISPR & biotech to improve human health https://t.co/15Ua1TtUa1T to @harvard @coastaluniv @edelmanw @JatGrants	FALSE	12272	1554	1538	148	12000	272
Pro	NFPC	Washington, DC	1/13/09 16:53	NFPC	The National Pork Producers Council is the global voice of the U.S. pork industry. #FFacebook #FFTwitter	FALSE	12422	2409	7423	272	12000	422
Pro	CanberraBio	Canberra, Australia	8/14/14 10:50	Dr. Graeme Burgo, PhD.	Group leader, Australian National University-ANU, Australia, Genetecist working on infections & #CRISPR. Opinions and views ...	FALSE	12430	1589	22856	415	12000	430
Pro	CRISPRJournal	New Rochelle, NY	5/18/17 20:15	The CRISPR Journal	Genetics & Genomics books & journals. General genetics & genomics news, call for papers, special issues, conference abstracts & more!	FALSE	18303	912	4967	220	18000	303
Pro	GeneticLiteracy	NA	1/20/12 17:34	Genetic Literacy Project	@GeneticLiteracy, Publisher @biotechpub	FALSE	18880	7412	36732	495	17999	881
Pro	Knoppler	Davis, CA	2/16/10 20:49	Paul Knoppler	The Genetic Literacy Project fosters dialogue about the scientific, social and ethical implications of human and agricultural genetics.	TRUE	21595	4750	26285	537	20999	596
Pro	AgBioWorld	Alabama, USA	4/14/11 14:09	Chama Prakash	Prof @UCDavis #Genetics & Kids Brain Cancer Epigenetics @CRISPR TED talk https://t.co/Qm5UJ0S4S Skeeet. Author of The Niche	FALSE	25412	2440	54085	514	24999	413
Pro	Synbioetica	San Francisco, CA	8/29/12 2:06	Synbioetica	https://t.co/8WpP3CtY YouTube channel https://t.co/8WpP3CtY @t.co/8WpP3CtY666	FALSE	31192	4210	27606	666	30998	194
Pro	CRISPR_News	San Francisco, CA	6/8/15 5:42	CRISPR_News	Professor, biotech guru, I am curious about science, farming, food security, innovation, plants, history. Come for the dog/cat videos, but stay for science.	FALSE	32463	539	1479	391	32000	463
Pro	doudna_lab	Berkeley, CA	12/3/15 21:01	Doudna Lab	The leading community of innovators in #synbio. Our Twitter Spaces show #BioEthicsSpaces goes LIVE every Weds 8pm PDT: https://t.co/8WpP3CtY	FALSE	42945	55	304	430	41999	946
Pro	Kevin_Faulconer	San Diego	6/18/09 23:35	Kevin Falconer	The latest CRISPR biotechnology... Also at https://t.co/8WpP3CtY @biotechnology #CRISPR #Genetics #Paradise #Genomics #Regenerative News from Jennifer Doudna's lab at @UCBerkeley. Tweets from lab members and Dr Jennifer Doudna unless signed JD. Tweets represent personal views only.	TRUE	45607	1194	6666	614	44998	609
Pro	Carlgil	Minneapolis, Minnesota (MN)	4/28/09 20:48	Carlgil	Candidate for California Governor. #CaliforniaComback Husband, father, former Mayor of San Diego, and visiting professor at @speakeasyapp	TRUE	57468	258	7497	948	58000	532
Pro	tysonfoods	Gibbst - HQ in Arkansas	8/13/08 12:42	Tyson Foods	Official Twitter account for Carlgil Inc.	TRUE	57592	18878	12130	913	57999	47

Figure 6: Expert Accounts Demographics (Raw Data)

6.2 Estimation Methodology and Prior Distributions for Parameters

I use RStan’s No U-Turn Sampling algorithm, developed by [Gelman et al. \(1995\)](#), and simulate two chains with 1,000 draws and a burn-in of 500 samples ([Stan Development Team, 2021](#)). The assumed prior distributions for the population are

$$\begin{aligned}\mu &\sim \mathcal{N}(\mu_\mu, \sigma_\mu) & \gamma &\sim \mathcal{N}(\mu_\gamma, \sigma_\gamma) \\ \alpha_j &\sim \mathcal{N}(\mu_\alpha, \sigma_\alpha) & \beta_i &\sim \mathcal{N}(\mu_\beta, \sigma_\beta) \\ \theta_i &\sim \mathcal{N}(\mu_\theta, \sigma_\theta) & \phi_j &\sim \mathcal{N}(\mu_\phi, \sigma_\phi),\end{aligned}$$

and the full joint posterior distribution is thus defined as:

$$\begin{aligned}p(\mu, \alpha, \beta, \gamma, \theta, \phi | \mathbf{y}) &\propto p(\mu, \alpha, \beta, \gamma, \theta, \phi, \boldsymbol{\mu}, \boldsymbol{\sigma}) \\ &\propto \prod_{i=1}^n \prod_{j=1}^m \text{logit}^{-1}(\pi_{ij})^{y_{ij}} (1 - \text{logit}^{-1}(\pi_{ij}))^{1-y_{ij}} \\ &\quad \prod_{j=1}^{m \times n} \mathcal{N}(\mu | \mu_\mu, \sigma_\mu) \prod_{i=1}^{m \times n} \mathcal{N}(\gamma | \mu_\gamma, \sigma_\gamma) \\ &\quad \prod_{j=1}^m \mathcal{N}(\alpha_j | \mu_\alpha, \sigma_\alpha) \prod_{i=1}^n \mathcal{N}(\beta_i | \mu_\beta, \sigma_\beta) \\ &\quad \prod_{i=1}^n \mathcal{N}(\theta_i | \mu_\theta, \sigma_\theta) \prod_{j=1}^m \mathcal{N}(\phi_j | \mu_\phi, \sigma_\phi)\end{aligned} \tag{4}$$

where $\pi_{ij} = \mu + \alpha_j + \beta_i - \gamma|\theta_i - \phi_j|$, and the latent prior parameters are $\boldsymbol{\mu} = (\mu_\theta, \mu_\phi)'$ and $\boldsymbol{\sigma} = (\sigma_\theta, \sigma_\phi)'$. While [Navelski and Pascual \(2022\)](#) develop a new set of Jeffery’s priors, I use the prior specification from [Barberá et al. \(2015\)](#), which is supported by [Hoff \(2003\)](#).

6.3 Identification Strategy

The model in Equation (1) is still unidentified due to “additive aliasing” and “scaling 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 $\mu = 0, \alpha_j = 1, \beta_i = -1, \phi_j = -1, \theta_i = 1$, and an example of scaling invariance is multiplying the distance $-\gamma|\theta_i - \phi_j|$ by any constant k where γ will absorb part of the constant $-\frac{\gamma}{k}|(\theta_i - \phi_j)k|$ ([Barberá \(2015\)](#)).¹⁴ These problems are usually solved by restricting one of the j^{th} or i^{th} parameters in each parameter set, but becomes difficult to do when working with the distance between two latent parameters. [Navelski and Pascual \(2022\)](#) suggest an alternative identification strategy where all priors are treated equally for ϕ_j and θ_i , and are transformed using an invariant transformation.

¹⁴[Barberá \(2015\)](#) does a great job at explain this in the Supplementary Materials if a deeper explanation is of interest.

6.3.1 Latent Positions (ϕ_j)

I employ a modification of Fisher’s inverse arc-tangent transformation, which was developed and applied by [Navelski and Pascual \(2022\)](#), and used in this setting. The general formulation is:

$$x'_1 = \arctan \left(e^{(x_1 + x_{center}) \frac{2}{\pi}} - x_{cut} \right) \quad x'_2 = \arctan \left(x_{cut} - e^{(x_2 + x_{center}) \frac{2}{\pi}} \right) \quad (5)$$

where, $x_1 \in \mathbb{R}$ and $x_2 \in \mathbb{R}$ are mapped to $x'_1 \in (x_{cut}, 1)$ and $x'_2 \in (-1, x_{cut})$, and

- x_1 and x_2 are the initially estimated values
- x'_1 and x'_2 are the transformed values
- x_{cut} is hyperparameter and a constraint put on the lower bound for each side of the estimation
- $x_{center} = \ln(\tan(\frac{\pi}{2} * 0) + x_{cut})$ and is a value that centers the transformation.

I apply this transformation to the expert latent parameters to constrain the parameter estimates and to map estimates to an intuitive scale. More specifically, I map all samples drawn from the priors to a constrained parameter space where $\phi_j \in (-1, 1)$ for all $j \in \{1, \dots, m\}$. This transformation is intuitive because it allows researchers to analyze relative latent positions on a -1 (anti) to 1 (pro) scale. In Equation 5 x_{cut} is a hyperparameter that 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 GEDL application the @AzMilkProducers expert was originally classified as pro-GEDL and were given an initial value of .8 to initiate the MC-MC estimation, and the mean of their latent posterior distribution converged to -0.089 , which is more anti-GEDL leaning than pro-GEDL. This intuitive transformation

An additional difficulty when using a distance model is “reflection invariance” where the resulting scale between θ_i and ϕ_j could lead to estimates that are misinterpreted since their signs could be flipped. To combat “reflection invariance,” I assume the aforementioned transformation in Equation 5 for ϕ , and use $\phi_{cut} = .6$ to guide the parameter estimates to their anticipated estimates. To aid this transformation, I assume -0.8 as the starting value for ϕ_j for the anti-GEDL experts and $+0.8$ for the pro-GEDL experts. These assumptions are not strong as Bayesian theory allows posterior draws to converge to their theoretically correct distribution. If estimates were to diverge from their anticipated underlying distribution, I would see estimates trying to converge to the posterior distribution on the other side of the scale. This would also lead to unsatisfactory conversion diagnostics and high prediction errors. In this analysis, I use the standard practice of constraining the m random effects to sum up to zero (i.e., $\sum_{j=1}^m \alpha_j = 0$) and the same for the n random effects (i.e., $\sum_{i=1}^n \beta = 0$). This allows all estimates to be identified relative to their expectation, which is zero.

6.4 Summary of Estimation Results

Figures 7 (a) and (b) show the distribution of the posterior means of expert popularity and follower engagement estimates, respectively. Figures 8 (a) and (b) show the distribution of the posterior means for the expert and follower ideology estimates. Expert popularity is centered at zero (i.e., $E(\hat{\alpha}_j) = 0$), and even though the distribution seems to be 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.

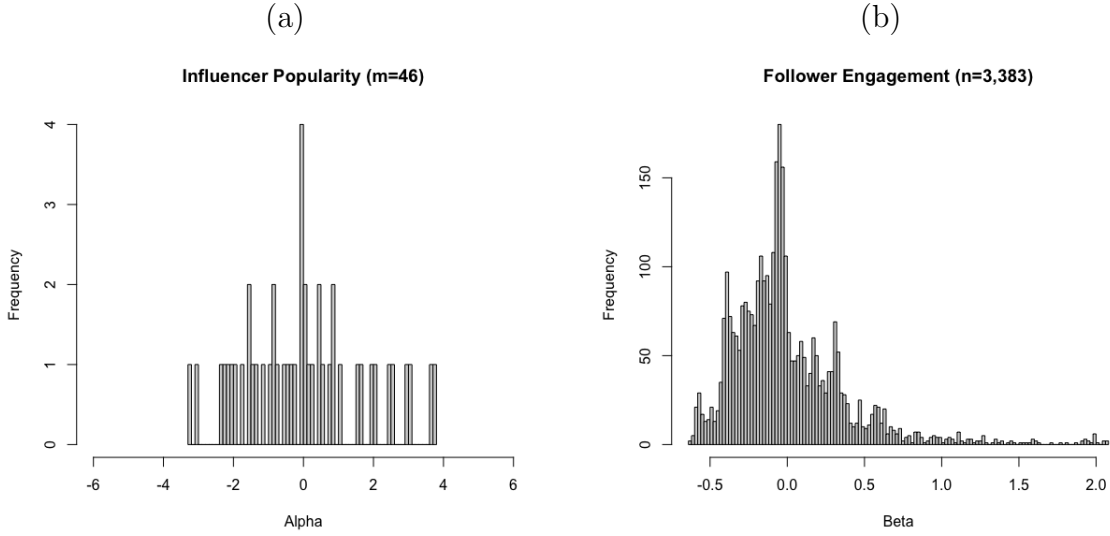


Figure 7: Popularity for Experts $\hat{\alpha}_j$ (a) and Engagement of Followers $\hat{\beta}_i$ (b).

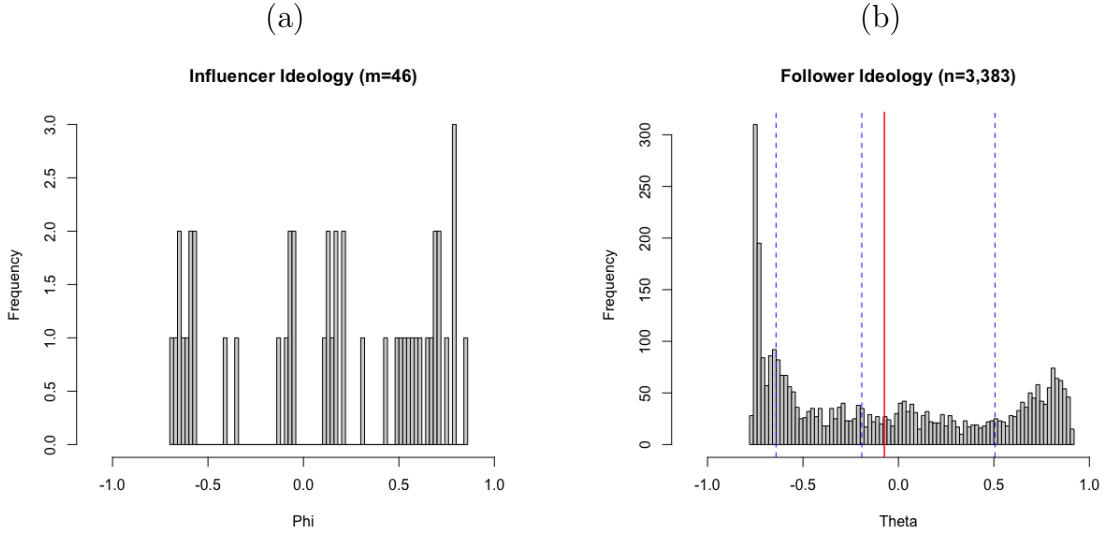


Figure 8: Ideology for Experts $\hat{\phi}_j$ (a) and Followers $\hat{\theta}_i$ (b).

Table 6: Examples of Popularity and Ideology Extremes

Parameter (Name - Initializing View)	\hat{R}	Mean	SD	2.5%	97.5%
Most Popular					
$\hat{\alpha}_2$ (@NonGMOProject - Anti)	1.00	3.78	0.16	3.49	4.11
$\hat{\alpha}_1$ (@OrganicConsumer - Anti)	1.00	3.69	0.16	3.38	4.02
Least Popular					
$\hat{\alpha}_{15}$ (@Kevin.Faulconer - Pro)	1.00	-3.30	0.12	-3.51	-3.06
$\hat{\alpha}_{45}$ (@AzMilkProducers - Pro)	1.00	-3.07	0.12	-3.29	-2.84
Most Extreme Anti-GEDL Ideology					
$\hat{\phi}_{12}$ (@GMOEvidence - Anti)	1.03	-0.69	0.02	-0.72	-0.66
$\hat{\phi}_4$ (@nongmoreport - Anti)	1.01	-0.68	0.02	-0.71	-0.65
$\hat{\phi}_1$ (@OrganicConsumer - Anti)	1.01	-0.65	0.02	-0.68	-0.62
$\hat{\phi}_5$ (@GMOFreeUSA - Anti)	1.03	-0.65	0.01	-0.68	-0.62
Most Extreme Pro-GEDL Ideology					
$\hat{\phi}_{42}$ (@joeBondyDenomy - Pro)	1.03	0.85	0.03	0.79	0.91
$\hat{\phi}_{37}$ (@shsternberg - Pro)	1.00	0.80	0.02	0.77	0.84
$\hat{\phi}_{25}$ (@pdhsu - Pro)	1.01	0.78	0.02	0.75	0.81
$\hat{\phi}_{39}$ (@AprilPawluk - Pro)	1.01	0.78	0.02	0.74	0.83
Moderate GEDL Ideology					
$\hat{\phi}_{24}$ (@NPPC - Pro)	1.01	-0.04	0.02	-0.08	-0.00
$\hat{\phi}_{14}$ (@Cargill - Pro)	1.01	-0.04	0.02	-0.07	-0.01
$\hat{\phi}_{26}$ (@nmpf - Pro)	1.01	-0.07	0.02	-0.11	-0.04
$\hat{\phi}_{15}$ (@Kevin.Faulconer - Pro)	1.00	-0.08	0.04	-0.15	0.00

Ideologies of the experts and followers exhibit opposite distributional patterns. The negative side of distribution in Figure 8 (a) and (b) represent those that are anti-GEDL, and the positive side are those that are pro-GEDL. For clarity, an ideology value of -1 indicates the most extreme anti-GEDL ideology, while a value of 1 indicates the most extreme pro-GEDL ideology. Both the ideology of the experts and followers tend to be polarized since 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 @AprilPawluk 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-GEDL accounts are very extreme while the pro-GEDL accounts range from extreme to moderate. For example, @NPPC, @Cargill, @nmpf and @Kevin.Faulconer were all initially considered experts on the pro-GEDL 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-GEDL experts while their ideological estimates are moderate to anti-moderate GEDL. This type of result could imply that these accounts have more moderate ideologies.

6.5 Estimation Diagnostics

Gelman and Rubin (1992) recommend a \hat{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, Figure 9 plots all \hat{R} values, showing that all values are below the 1.1 line. The optimal classification threshold was derived by maximizing the area under the ROC curve (i.e., maximizing the sensitivity and specificity of the prediction diagnostics), and Figure 10 (a) and (b) show the ROC curve and confusion matrix, respectively.

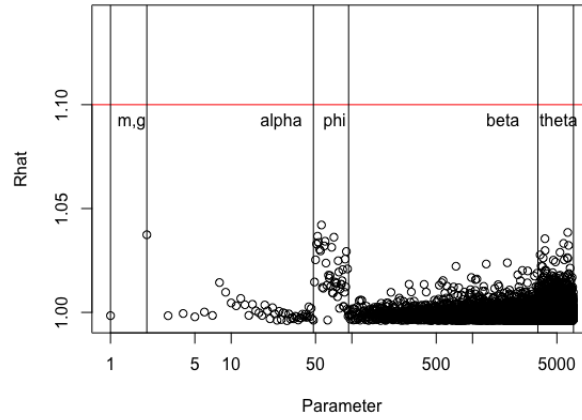


Figure 9: \hat{R} Plot of All 6,860 Parameters (MCMC Convergence Diagnostics)

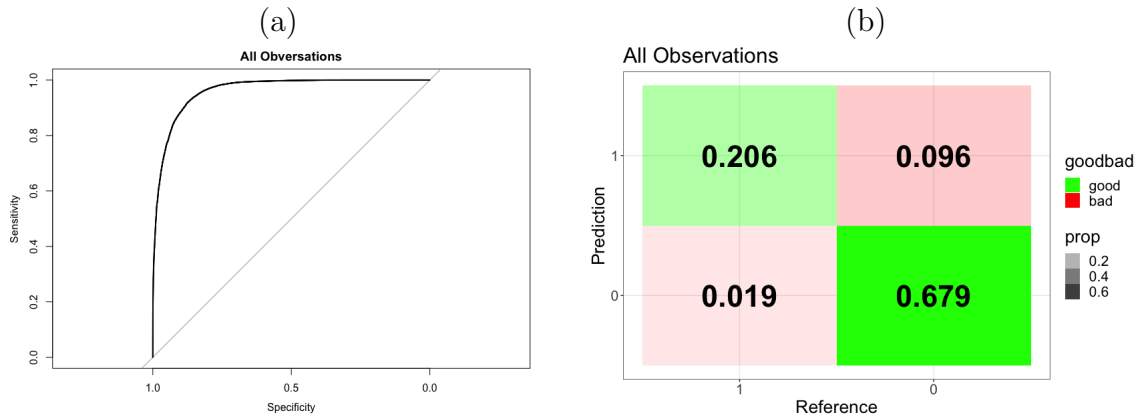


Figure 10: ROC Curve (a) and Confusion Matrix (b) for All Observations

6.6 T^* Matrix for All Examples

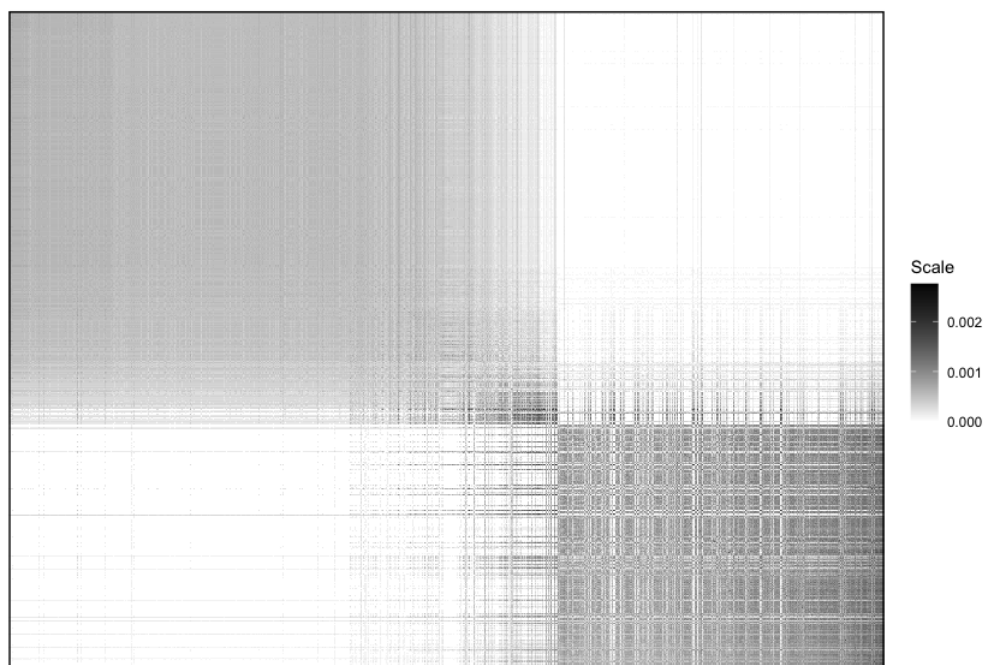


Figure 11: A Social Network Based on Common Connections

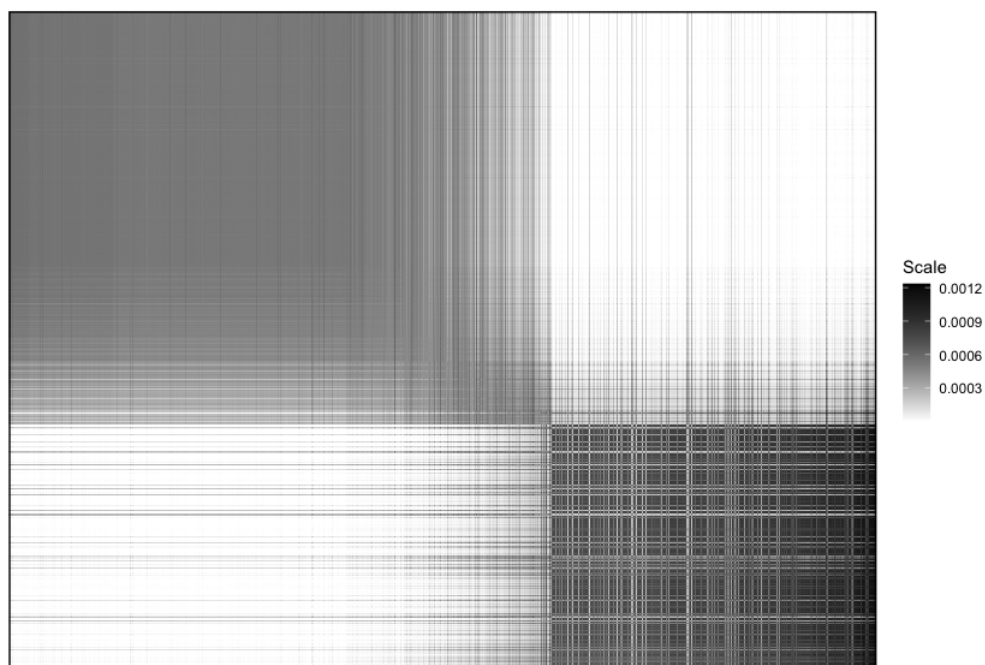


Figure 12: An Estimated Social Network Based on Common Connections