# Social Learning in the Media: The Importance of Bias for Trust Between Media Firms

Bryan Kelly, Joseph Leland Bybee

July 5, 2014

## 1 Introduction

The problem. The media is important to determine what people know about the world. How the media gets its information is important as a result. When media firms aren't getting information from their reporters they are getting information from other news firms.

Our goal is to develop a model of media trust that accounts for trust in the media and how different media firms learn from each other. Our model allows us to develop a network of trust between different media firms that allows for information diffusion between media firms.

We examine two novel data sets to investigate this model, the first is a data set of tweets for a list of 3000 media firms in the United States from June to August of 2014. Additionally, we consider a data set of articles taken from the free archive ArXiv. We build a trust network for both of these data sets based on a number of different measures of bias and use this to investigate our model. NOTE: Obviously, if we are going to include this second dataset the problem will be generalized beyond media firms to some idea of "information agents".

We specify our model using a topic model comparable to the popular topic modeling approach, Latent Dirichlet Allocation (LDA). We uses this to estimate the amount of information that different firms get from each other and uses this as an estimate of trust.

We consider 5 predictions. First, firms that share common beliefs are more likely to trust each other. Second, firms that share common beliefs will communicate information faster. Third, higher quality firms are more likely to be trusted by all firms. Fourth, information that originates from multiple distance media sources is more likely to be accurate. Fifth, neutral media sources are important in order to connect different biased groups.

The paper is grouped as follows. Section 2 presents the prior literature. Section 3 develops the model used throughout. Section 4 covers the data sets used. Section 5 covers the implementation of the model. Section 6 presents the results. Finally, Section 7 summarizes and concludes.

# 2 Literature Review

We draw heavily from Gentzkow and Shapiro's 2006 paper *Media Bias and Reputation*[1]. Gentzkow and Shapiro develop a model where media firms bias their reporting to align with the beliefs of their market to maximize profits. We generalize this model to a network of media firms, where each firm learns from the actions of other firms in the network.

To estimate media bias we consider a number of different measures drawn from several different sources. Groseclose and Milyo develop a measure that looks at the citations made by different media firms and how these coincide with the citations made by members of Congress[2]. Similarlly, Gentzkow and Shapiro develop a measure of media slant that uses the similarity between words used by different media outlets and members of Congress[3]. Finally, Chiang and Knight develop a model of media bias that looks at endorsements made by media firms[4].

The literature on slant in academia is more limited than in the media case, perhaps because the impact is less apparent. Much of the literature focuses on divisions of citation networks, though some look at text clustering including one example presented below. Onder and Tervio look at the division of citation networks into "freshwater" and "saltwater" departments for economics. They find that the division is very real and significant in a number of cases[5].

The key work we draw from in the social learning literature is Acemoglu et. al's 2011 paper *Bayesian Learning in Social Networks*[6]. They develop a model of Bayesian social learning the is built around a stochastic network similar to the one we develop here. They derive a number of general conclusions that are significant to this type of model. Acemoglu et. al's work on social learning in endogenous social networks is also significant as it mirrors the network structure we develop here[7]. Non Bayesian approaches have also been considered in the literature including Jadbabaie et. al's work[8].

There is a large literature on more general information diffusion. Young provides a good overview of how the social learning methods popular in economics compare to those used in other fields[9]. Golub and Jackson study how homophily affects diffusion and learning in networks. Their results are directly relevant to the work done here[10]. Yang and Leskovec develop a general model of information diffusion in implicit networks. This is significant here because the network structure for our model is implicit[11].

The method we use to estimate our model has its roots in the topic mod-

---

[1] Gentzkow and Shapiro. Media Bias and Reputation. 2006

[2] Groseclose and Milyo. A Measure of Media Bias. 2005.

[3] Gentzkow and Shapiro. What Drives Media Slant? Evidence from U.S. Daily Newspaper. 2006.

[4] Chiang and Knight. Media Bias and Influence: Evidence from Newspaper Endorsements. 2011.

[5] Onder and Tervio. Is Economics a House Divided? Analysis of Citation Networks. 2014.

[6] Acemoglu, Dahleh, Lobel and Ozdaglar. Bayesian Learning in Social Networks. 2011.

[7] Acemoglu, Bimpikis and Ozdaglar. Dynamic of Information Exchange in Endogenous Social Networks. 2013.

[8] Jadbabaie, Molavi, Sandroni and Tahbaz-Salehi. Non-Bayesian Social Learning. 2012.

[9] Young. Innovation Diffusion in Heterogeneous Populations: Contagion, Social Influence and Social Learning. 2014.

[10] Golub and Jackson. How Homophily Affects Learning and Diffusion in Networks. 2009.

[11] Yang and Leskovec. Modeling Information Diffusion in Implicit Networks. 2010.

eling literature, in particular the idea of LDA, first proposed by Blei, Ng and Jordan[12]. LDA is essentially a hierarchical model where it is assumed that each word in a document is drawn from some topic distribution that is drawn for each document. Using this hierarchical model Blei, Ng and Jordan proposed a method to estimate the underlying topic distribution for a set of documents. Since their seminal paper the literature has exploded with examples of topic models. Blei provides a good survey of the methodology[13].

Two key variants that are important here are dynamic topic models and author topic models. A combination of these ideas form the basis for the estimation procedure used in this paper. Blei and Lafferty introduce the idea of dynamic topic models to represent how topic may change over time[14]. They assume that the underlying topic proportions are drawn from some multivariate normal distribution centered around the prior topic distribution. While we don't use the same implementation here, the idea a dynamic topic is essential to our work as our estimates for the trust edges represent how much a news story influences the underlying topic distribution for other media firms. Rosen-Zvi et. al present another important variant in the idea of author topic models[15]. Their model assumes that each author has some underlying preference for certain topics that can be used to cluster documents. While we don't directly implement this procedure the idea of a biased view of topics tied to each author is important to our model.

A number of studies have used similar methods to investigate similar datasets to ours. Hong and Davison investigate and author topic model in the Twitter context to determine how Twitter users cluster and share information[16]. Grant et. al develop an online approach to estimating topics in the Twitter search context[17]. Zhao et. al use a topic model to compare Twitter media to traditional media sources[18]. Finally, Blei and Lafferty use a correlated topic model to estimate correlation within the scientific community using an academic article data set similar to the one used here[19].

# 3   Model

The roots of our model lie the Gentzkow and Shapiro paper *Media Bias and Reputation*. As with the Gentzkow and Shapiro model, we assume that there are two states of the world

$$S \in [L, R].$$

There are $N$ distinct populations that each make a decision

$$A_i \in [L, R].$$

---

[12]Blei, Ng, Jordan. Latent Dirichlet Allocation. 2003.

[13]Blei. Probabilitic Topic Models. 2012

[14]Blei and Lafferty. Dynamic Topic Models. 2006

[15]Rosen-Zvi, Griffiths, Steyvers, Smyth. The Author-Topic Model for Authors and Documents. 2004.

[16]Hong and Davison. Empirical Study of Topic Modeling in Twitter. 2010.

[17]Grant, George, Jenneisch and Wilson. Online Topic Modeling for Real-time Twitter Search. 2011.

[18]Zhao, Jiang, Weng, He, Lim, Yan and Li. 2011.

[19]Blei and Lafferty. A Correlated Topic Model of Science. 2007.

The $i$th population will receive a positive payoff in the case that $A_i = S$ and a payoff of zero otherwise. Each population has a prior belief about the state of the world

$$P_i^c(R) = \theta_i^c.$$

For each population there is a monopolist news firm that has a prior belief about the state of the world

$$P_i(R) = \theta_i.$$

Each news firm receives a signal about the state of the world

$$s_i \in [l, r]$$

and can report their signal (or a variation of their signal) to their population. There is some probability $\lambda$ that a given news firm is high quality. A high quality news firm is a news firm that receives a perfect signal of the world in each period, that is

$$p_i(l|L) = p_i(r|R) = 1.$$

Alternatively, a firm can be a normal quality firm. If this is the case than there is a probability $\pi_i$ that the firms receives a signal that aligns with the state of the world

$$p_i(l|L) = p_i(r|R) = \pi_i.$$

Each firm can publish a report based on their signal

$$\hat{s}_i \in [\hat{l}, \hat{r}].$$

Gentzkow and Shapiro show that the optimal strategy for a high quality firm is to always report the truth, that is

$$\hat{s}_i = s_i.$$

However, for the normal quality firm it may be better to lie. Each firm can chose a strategy based on their signal

$$\sigma_{is}(\hat{s}) = p_i(\hat{s}|s).$$

For the remainder of this paper we constrain ourselves to the case where

$$\sigma_{ir}(\hat{r}) > \sigma_{il}(\hat{l}).$$

Gentzkow and Shapiro show that, in the case where a consumer receives no exogenous signal about the state of the world, their posterior beliefs about the quality of their monopolist news provider is a strictly increasing function of the likelihood ratio

$$\frac{p(\hat{r}|\text{high}_i)}{p(\hat{r}|\text{normal}_i)} = \frac{\theta_i^c}{\theta_i^c[\sigma_r(\hat{r}_i)\pi_i + \sigma_l(\hat{r}_i)(1 - \pi_i)] + (1 - \theta_c^i)[\sigma_r(\hat{r}_i)(1 - \pi_i) + \sigma_l(\hat{l}_i)\pi_i]}$$

This suggests that the higher $\theta_i^c$, the more likely that a consumer will believe the $i$th firm is of high quality. In equilibrium they show that the expected gain for publishing $\hat{r}$ is

$$\Delta(s)_i = (1 - \mu)\Delta_i^{nf} + \mu\Delta_i^f(s)$$

where

$$\Delta_i^{nf} = f(\lambda_i(\hat{r}, 0)) - f(\lambda_i(\hat{r}, 1)$$

and

$$\Delta_i^f(s) = \theta_i(s)[f(\lambda_i(\hat{r}, R)) - f(0)] - (1 - \theta_i(s))[f(\lambda_i(\hat{l}, L) - f(0)]$$

In the above, $\lambda_i(\hat{s}, 0)$ represents the consumers posterior belief about the quality of the $i$th firm in the case of no feedback and $\lambda_i(\hat{s}, S)$ represents the consumers posterior belief about the quality of the $i$th firm in the case of feedback. $\theta_i(s)$ is the $i$th firms posterior belief about the state of the world after receiving their signal $s_i$. Gentzkow and Shapiro show that under equilibrium conditions it is in a normal media firms interests to skew their reporting to align with the consumers priors about the state of the world.

## 3.1 Inter-Firm Trust

We now consider the situations in which firms can learn from each other. We assume that in equilibrium the $i$th firm generally publish articles that align with their consumers priors, $\theta_i^c$ and that the firms priors should reflect this, so $\theta_i \approx \theta_i^c$. $\theta_i$ can then be thought of as the proportion of news stories published by the $i$th firm that correspond to the $R$ state of the world. So then over $T$ periods we would expect $\theta_i T$ reports to be $\hat{r}$. Let's now turn to the $j$th firm's beliefs about the quality of the $i$th firm over $T$ periods. In each period, the $j$th firms beliefs should be

$$p_j(\text{high}_i | \hat{\mathbf{s}}_{it}) = \frac{p_j(\hat{\mathbf{s}}_{it} | \text{high}_i)p_j(\text{high}_i)}{p_j(\hat{\mathbf{s}}_{it})}$$

where $\hat{\mathbf{s}}_{it}$ is a vector of the reports made by the $i$th firm. Assuming that the beliefs follow the monotone likelihood ratio property we can show that

$$p_j(\text{high}_i | \hat{\mathbf{s}}_{it}) = \frac{p_j(\text{high}_i)}{1 + \frac{p_j(\hat{\mathbf{s}}_{it} | \text{normal}_i)}{p_j(\hat{\mathbf{s}}_{it} | \text{high}_i)}}$$

This gives us the Gentzkow and Shapiro result that $p_j(\text{high}_i | \hat{\mathbf{s}}_{it})$ is a strictly increasing function of $\frac{p_j(\hat{\mathbf{s}}_{it} | \text{high}_i)}{p_j(\hat{\mathbf{s}}_{it} | \text{normal}_i)}$. Since the proportion of $\hat{\mathbf{s}}_{it}$ that is $\hat{r}$ is an increasing function of $\theta_j$, $p_j(\text{high}_i | \hat{\mathbf{s}}_{it})$ is an decreasing function of $\theta_j - \theta_i$, that is, the more similar the $i$th firms beliefs are to those of the $j$th firm the more likely the $j$th firm is to believe the $i$th firm is high quality.

We can use this idea to represent how much different news firms trust each other. We use this to help derive a number of bias-free measures of truth and quality that could be used to help understand how the media market works. Throughout the remainder of this paper we will think of $p_j(\text{high}_i | \hat{\mathbf{s}}_{it})$ as the

probability that the $j$th firm will incorporate the $i$th firms report into their own report.

NOTE: I think the next two sections may be the most interesting but they still need a good deal of development. I don't really know what to do with them.

## 3.2 Estimating Bias-Free Truth

Consider a case where you have $N$ firms, as before, however, each firm cannot observe the reports of the other firms. Additionally, assume that you are a Bayesian agent observing all $N$ firms and you want to get some estimate of the actual state of the world. You may have your prior belief about the state of the world

$$p(R) = \theta$$

Additionally, you have the reports for each of the $N$ firms as evidence. Then what you might do is update your beliefs following Bayes rule

$$p(R|\hat{s}) \propto p(\hat{s}_1|R)...p(\hat{s}_N|R)p(R)$$

Unfortunately, this does not work in the case considered above because each firm can observe the reports of their fellow firms, so it is possible you may observe duplicate information. To account for this we need to account for how much firms trust each other and how likely they are to borrow information from other firms. Fortunately, you have a measure of how much likely it is that each firm will incorporate the reports of another firm into their report, $p_j(\text{high}_i|\hat{s}_{it})$. We could modify the above posterior to account for this, let $\gamma_i j$ represent $p_j(\text{high}_i|\hat{s}_{it})$ x

$$p(R|\hat{s}) \propto (1 - \gamma_{12})(1 - \gamma_{13})...(1 - \gamma_{1N})p(\hat{s}_1|R)...$$

$$(1 - \gamma_{N1})(1 - \gamma_{N2})...(1 - \gamma_{NN-1})p(\hat{s}_N|R)p(R)$$

We propose that this may be used as a measure to estimate the actual truth of a given state of the world. One implication of this proposal is that we are more certain about the state of the world the more dispersed a set of signals are. To see this consider a simple two agent case.

$$p(R|\hat{s}) \propto (1 - \gamma_{12})p(\hat{s}_1|R)(1 - \gamma_{21})p(\hat{s}_2|R)p(R)$$

In this case the more likely either firm trust the other ($\gamma_{12}$ is higher or $\gamma_{21}$ is higher), the lower $p(R|\hat{s})$. That is that if two firms share a similar belief about the state of the world and report the same signal, this says less than if two firms that have different beliefs report the same signal.

## 3.3 Estimating Bias-Free Quality

Next we propose a measure of bias-free quality for a given agent. One aspect of our model that has so far been ignored is the $j$th firm's prior beliefs about the quality of the $i$th firm. We could think of this prior as the equilibrium belief about the quality of the $i$th firm. Since $\theta_i$ and $\theta_j$ are the equilibrium beliefs for

each firm about the state of the world, a reasonable choice for $p_j(H_i)$ might be $\theta_{ij}$, where $\theta_{ij} = \theta_i - \theta_j$. If we accept this prior we can then get an estimate for the likelihood function

$$p_j(\hat{\mathbf{s}}_i|\text{high}_i) = \frac{p_j(\text{high}_i|\hat{\mathbf{s}}_i)}{p_j(\text{high}_i)} = \frac{\gamma_i j}{\theta_i j}$$

Using this measure we could then think like the Bayesian agent from the sub section above

$$p(\text{high}_i|\hat{\mathbf{s}}_i) \propto p(\hat{\mathbf{s}}_i|\text{high}_i)p(\text{high}_i)$$

$$= p_1(\hat{\mathbf{s}}_i|\text{high}_i)p_2(\hat{\mathbf{s}}_i|\text{high}_i)...p_N(\hat{\mathbf{s}}_i|\text{high}_i)p(\text{high}_i)$$

$$= \frac{\gamma_{1i}}{\theta_{1i}}\frac{\gamma_{2i}}{\theta_{2i}}...\frac{\gamma_{Ni}}{\theta_{Ni}}p(\text{high}_i)$$

This would then suggest that our bias free estimate of quality would be an increasing function of the total trust in the $i$th firm, and a decreasing function of the shared beliefs with the $i$th firm. The intuition behind this measure is that higher quality firms should be more trusted by all firms controlling for bias.

## 3.4 The Significance of Border Cases

Next we consider the possibility for border agents. The intuition behind the idea for border agents is that they are agents who have a high enough $\gamma$ value for two different agents with rather low $\gamma$ values that they allow the communication of ideas between groups who otherwise would not communicate. More formally, consider a network with 3 agents, $i$, $j$, and $k$. If

$$\gamma_{ij}\gamma_{jk} > \gamma_{ik}$$

the $i$th agent is a border agent. Border agents are important because they mean that given a lag two groups that otherwise would not receive the same information can receive the same information.

We consider what proportion of the population of firms are border agents and how much information passes through the border agents. NOTE: This sub section may belong in a different section.

## 3.5 Risk Averse News Firms

Next we consider the possibility that there are risk averse media firms. These are firms that are willing to pay a premium to avoid poor quality information. This premium may involve waiting longer to verify a report or only trusting higher quality firms. We say that these firms are more likely to trust firms that have a higher quality level controlling for bias than less risk averse firms. To estimate who the risk averse firms are we could assume the following relationship between trust, bias and bias-free quality

$$\gamma_{ij} = \beta_1\theta_{ij} + \beta_2 p(\text{high}_j|\hat{\mathbf{s}}_j)$$

Firms that are more risk averse should have a higher $\beta_2$ relative to the rest of the population of firms. NOTE: This may also fit better in a different section.

7

# 4 Implementation

In order to estimate the trust between agents we specify a variation on the Dynamic Topic Model proposed by Blei and Lafferty[20]. Topic models are a set of hierarchical models that assume a latent "topic" underlying the choice of words in a collection of documents. The classical model is Latent Dirichlet Allocation (LDA), proposed by Blei et al.[21]. Assume that you have a collection of documents $D$ where $D = \{\mathbf{w}_1, \mathbf{w}_2, ... \mathbf{w}_m\}$. Each $\mathbf{w}_i$ is a document consisting of a set of words $w = \{w_1, w_2, ..., w_n\}$. LDA specifies the following generating process for the collection of documents:

$$\theta_m \sim \text{Dir}(\alpha)$$

$$z_n \sim Mult(\theta_m)$$

$$w_n \sim p(w_n | z_n, \beta) = Mult(\beta_z)$$

That is, $\theta$ is drawn from a Dirichlet distribution for each document. Then $z_n$ is drawn for each word in the document, and the word is from a Multinomial disttribution contional on the specific $z_n$ drawn and $\beta$. $\beta$ is a probability matrix where $\beta_{ij} = p(w^j = 1 | z^i = 1)$. What makes LDA particularly interesting among the class of clustering models is that topic parameter is sampled for each individual word, allowing for a mixture of topics within a document. Ultimately LDA allows for the estimation of $\theta$ and $\mathbf{z}$ which estimate the distribution of topics within a collection of documents.

Variants of LDA have seen a great deal of success for modeling the popularity of certain ideas within a collection of documents, in particular we focus on the Dynamic Topic Model. Blei and Lafferty developed the Dynamic Topic Model to account for the changing distribution of topics over time. It allowed for more accurate estimation when faced with a time series of documents. The Dynamic Topic Model has a similar structure to LDA, however, it assumes time dependence for the $\alpha$ and $\beta$ parameters.

$$\alpha_t | \alpha_{t-1} \sim N(\alpha_{t-1}, \delta^2 I)$$

$$\beta_t | \beta_{t-1} \sim N(\beta_{t-1}, \sigma^2 I)$$

$$\theta \sim N(\alpha_t, a^2 I)$$

$$z_{tn} \sim Mult(\pi(\theta))$$

$$w_{tn} \sim Mult(\pi(\beta_{t,z}))$$

This generation procedure allows the distribution of topics to change over time, accounting for shifts in the content of documents. Such a model is important for the dataset we wish to use where topics are changing regularly. We

---

[20]Blei and Lafferty. Dynamic Topic Model. 2006.
[21]Blei, Ng and Jordan. Latent Dirichlet Allocation. 2003

use a variant of the Dynamic Topic Model that allows for the distribution of $\alpha_{it}$ to be impacted by $\alpha_{-it-1}$. That is we assume that the underlying topic distribution is ultimately effected by the topic distributions for other agents in a network of agents. More formally we say that,

$$\alpha_{it} \sim N(\alpha_{t-1}^T \gamma, \delta^2 I)$$

$\alpha$ is a vector of the $\alpha$ parameter for each of the $N$ agents, and $\gamma$ is an $N$ dimensional vector of weights for the $N$ agents. That is, $\alpha_{it}$ is drawn from a weighted sum of the $\alpha$ values for each agent in the prior period. $\gamma$ gives an estimate of the impact that each agent has on the topic distribution of the $i$th agent. We use $\gamma$ to represent our trust value discussed above. Unfortunately, this generating process does not have a conjugate form so we cannot using Gibbs sampling as used commonly with LDA. Instead we must use a more complicated Metropolis Hastings algorithm to estimate $\gamma$.

# A    Full Conditionals

$$p(\gamma_i|\mu, \xi^2) = N(\mu, \xi^2 I)$$

$$p(\alpha_{it}|\boldsymbol{\alpha}_{t-1}.\gamma_i, \delta^2) = N(\boldsymbol{\alpha}_{t-1}^T \gamma_i, \delta^2 I)$$

$$p(\beta_t|\beta_{t-1}, \sigma^2) = N(\beta_{t-1}, \sigma^2 I)$$

$$p(\theta_{itd}|\alpha_{it}, a^2) = N(\alpha_{it}, a^2)$$

$$p(z_{itnd}|\theta_{itd}) = Mult(\pi(\theta_{itd}))$$

$$p(w_{itnd}|\beta_t, z_{itnd}) = Mult(\pi(\beta_t^{z_{itnd}}))$$

$$p(\theta, \boldsymbol{\gamma}, \boldsymbol{\alpha}, \boldsymbol{\beta}, \mathbf{z}|\mathbf{w}) \propto \prod_{i=1}^{M} \prod_{t=1}^{T} \prod_{n=1}^{N} \prod_{d=1}^{D} p(w_{itnd}|z_{itnd}, \beta_t) p(\beta_t|\beta_{t-1}, \sigma^2) p(z_{itnd}|\theta_{itd})$$

$$p(\theta_{itd}|\alpha_{it}, a^2) p(\alpha_{it}|\boldsymbol{\alpha}_{t-1}, \gamma_i, \delta^2) p(\gamma_i|\mu, \xi^2)$$

## A.1    Posterior Specification for $z_i$